# β-Sarcoglycan: genomic analysis and identification of a novel missense mutation in the LGMD2E Amish isolate

F. Duclos[1a], O. Broux[1b], N. Bourg[1b], V. Straub[a], G.L. Feldman[c], Y. Sunada[a], L.E. Lim[a],
F. Piccolo[d], S. Cutshall[a], F. Gary[b], F. Quetier[b], J.-C. Kaplan[d],
C.E. Jackson[c], J.S. Beckmann[b,e], K.P. Campbell[a,*]

[a]*Howard Hughes Medical Institute, Department of Physiology and Biophysics and Department of Neurology,
University of Iowa College of Medicine, Iowa City, IA 52242, USA*
[b]*CNRS URA 1922, France, Généthon, 91000 Evry, France*
[c]*Henry Ford Hospital, Department of Medical Genetics, Detroit, MI 48202, USA*
[d]*Hopital Cochin Maternités, INSERM U129, 75014 Paris, France*
[e]*Fondation Jean Dausset-CEPH, 75010 Paris, France*

## Abstract

The sarcoglycan complex is involved in the etiology of four autosomal recessive limb-girdle muscular dystrophies (LGMD2C–F). A missense mutation (T151R) in the β-sarcoglycan gene on chromosome 4q12 has been shown to cause a mild form of LGMD2E in 11 families from a Southern Indiana Amish community sharing a common haplotype. We now report that two sibs from another Amish family with mild LGMD2E are compound heterozygotes for chromosome 4q12 markers. In order to characterize the genetic defect in this new family, we determined the genomic organization of the β-sarcoglycan gene. A second missense mutation (R91C) has now been identified in this LGMD2E Amish family. This mutation is also present in the homozygous state in another family of probable Amish ancestry. Finally, analysis of all the components of the dystrophin-glycoprotein complex was performed for the first time on a biopsy from a patient homozygous for the β-sarcoglycan mutation (T151R). Interestingly, in addition to the loss of the entire sarcoglycan complex, we detected a reduction of α-dystroglycan which suggests a role for the sarcoglycan complex in stabilizing α-dystroglycan at the sarcolemma. © 1998 Elsevier Science B.V.

*Keywords:* Limb-girdle muscular dystrophy; β-Sarcoglycan; Sarcoglycan complex; α-Dystroglycan

## 1. Introduction

In the dystrophin-glycoprotein complex (DGC) [1–4], β-sarcoglycan, a 43 kDa glycoprotein, is closely associated with α-, γ-, and δ-sarcoglycan [5–7]. This subcomplex can be distinguished from α-, and β-dystroglycan [8], the syntrophins [9,10] and from dystrophin [11] by differential solubilization [6] or by immunoprecipitation [12]. Together, the DGC confers a structural link between laminin 2 in the extracellular matrix, via α-dystroglycan, and the cytoskeleton, via dystrophin, and is believed to protect muscle cells from contraction-induced damage [13,14].

The DGC has been implicated in several forms of muscular dystrophy. In Duchenne muscular dystrophy (DMD), mutations in the dystrophin gene result in a severe dystrophic phenotype [2,11], whereas in Becker muscular dystrophy, dystrophin mutations result in a dysfunctional dystrophin and are associated with a milder phenotype [15]. Severe and mild forms of congenital muscular dystrophy have recently been characterized by mutations in the laminin α2-chain gene [16–18].

Limb-girdle muscular dystrophies (LGMDs) represent a clinically and genetically heterogeneous group of diseases which are characterized by progressive weakness of the pelvic and shoulder girdle muscles [19–21]. The genes

---

* Corresponding author. Tel.:+1 319 3357867; fax: +1 319 3356957; e-mail: kevin-campbell@uiowa.edu
[1] These authors contributed equally to this work.

responsible for LGMD2C–F have been identified and encode for γ-sarcoglycan [22], α-sarcoglycan [23], β-sarcoglycan [24,25], and δ-sarcoglycan [26,27], respectively. In LGMD2C–F caused by sarcoglycan mutations, a mutation in one of the sarcoglycan proteins leads to the concomitant loss or reduction of the other sarcoglycans at the sarcolemma [7,22–32].

To facilitate the analysis of LGMD2E patients and to characterize the β-sarcoglycan gene, we performed an extensive analysis of 36 kb of genomic sequences overlapping the β-sarcoglycan gene on chromosome 4q12. Using various predictive algorithms, we analyzed the entire genomic sequence and detected two putative exons that could be part of tissue specific isoforms of β-sarcoglycan. We have also found several additional polyadenylation sites that may account for the two transcripts of 3.0 and 4.4 kb which have not been yet characterized. A second LGMD2E mutation was identified in an Amish family with two affected sibs carrying the common haplotype [24] at the heterozygous state. This second mutation was also present in the homozygous state in patients belonging to another family with probable Amish origins. In addition, we screened for α-sarcoglycan deficiency 107 cases diagnosed with myopathy, primarily characterized with normal immunohistochemical findings for dystrophin. We identified 10 sporadic cases with these criteria and screened them for β-sarcoglycan mutations. A novel missense mutation I119F was identified in the heterozygous state in the β-sarcoglycan gene of one of these patients. Of particular interest, we demonstrated with immunofluorescence analysis, a reduction in the expression level of both the sarcoglycan complex and of α-dystroglycan at the sarcolemma of one individual carrying a homozygous β-sarcoglycan T151R mutation.

## 2. Materials and methods

### 2.1. Subcloning and sequencing of cosmid Cos4

DNA of cosmid Cos4 (cloning vector, Supercos 1) was sonically sheared and subcloned into a M13 vector. Insert DNA was amplified by PCR under standard conditions using 1 μl of overnight culture in a 50 μl total volume containing the M13 forward and M13 reverse primers. After a 5 min hot start at 94°C, 2 units of Taq Polymerase (Perkin-Elmer) were added, followed by 35 cycles of 60 s denaturation at 94°C, 60 s annealing at 55°C, and 90 s extension at 72°C. The PCR amplification products were analyzed by agarose gel electrophoresis and only those displaying an insert size higher than 150 bp were sequenced on one strand of DNA for sizes <600 bp or on both DNA strands for size >600 bp.

### 2.1.1. EcoRI/NotI restriction map of Cos4

Restriction fragments of Cos4 obtained upon EcoRI and NotI digestion were separated by agarose gel electrophoresis and capillary transferred to (Amersham) nylon membrane and hybridized with T3 and T7 primers for mapping.

### 2.1.2. Sequence analysis

The sequences were analyzed and alignments performed using XBAP software of the STADEN package (version 94.2). Gaps between sequence contigs were filled by walking with internal primers and by sequencing both ends of several EcoRI-EcoRI, NotI-NotI and EcoRI-NotI restriction fragments. The total consensus sequence was scanned using the following software programs: Grail2a and XGrail2a (version 1.3b); FEXH, HEXON, FGENEH, TSSG, TSSW of the BCM Gene finder package (version 2.10.5.94); and Promoter Scan program (version 1.7). Accession number EMBL# Y09781.

### 2.2. Immunofluorescence analysis

Amish patients were ascertained through the Henry Ford Hospital, at the Department of Medical Genetics, Detroit, MI, USA. Skeletal muscle tissue from LGMD patients were obtained from diagnostic biopsies. Human control muscle was received from patients undergoing orthopedic surgery not related to neuromuscular disorders. Antibody staining on 7 μm transverse cryosections was performed at room temperature. Sections were treated with AB blocking solutions (Vector), blocked with 3% bovine serum albumin (BSA) in phosphate-buffered saline (PBS) for 30 min, and then incubated with a dilution of the primary antibody for 90 min. Antibodies against the following components of the DGC were tested: dystrophin, laminin α2-chain, α-, β-, γ-, δ-sarcoglycan, and α- and β-dystroglycan. Sections were then treated and analyzed as previously described [24]. Monoclonal antibodies VIA4$_2$ against dystrophin and IVD3$_1$ against α-sarcoglycan were previously characterized [2,33]. Monoclonal antibody 5B1 against β-sarcoglycan and 21B5 against γ-sarcoglycan were produced in collaboration with Dr. Louise Anderson (Muscular Dystrophy Research Labs, Newcastle General Hospital, Newcastle upon Tyne, NE4 6BE, UK). Monoclonal antibody 8D5 against β-dystroglycan was kindly provided by Dr. Louise Anderson. An affinity-purified δ-sarcoglycan antibody (rabbit 215) was produced against the N-terminal sequence (MMPQE-QYTHHRSTMPGM) of human δ-sarcoglycan. This synthetic peptide (Research Genetics) with an additional cysteine at the N terminus was conjugated with the N-terminal cysteine of keyhole limpet hemocyanin (Pierce) using m-maleimidobenzoic acid-N-hydroxysuccinimide ester (Pierce), mixed with Freund's adjuvant (Sigma) and injected into rabbit 215. Polyclonal antibody against the peptide was affinity purified from crude sera using BSA-conjugated peptide as described previously [12]. Antibodies against different components of the dystrophin-glycoprotein complex were produced in a goat (Goat 20) using the purified rabbit skeletal muscle DGC [2] as previously described [27]. A specific antibody against α-dystroglycan was affi-

Table 1

Primers used for RT-PCR and genomic analysis of β-sarcoglycan

| Gene/fragment | Sequences of primers | Length (bp) | Annealing T°C |
|---|---|---|---|
| Exon 1–3 | BF0: GGA.CAG.TCG.GGC.GGG.GA | 365 | 58 |
| (RTPCR) | BR0: CTT. TGT.TGT.CCC.TTG CTG.A | | |
| Exon 2–6 | BF9: AGT.TCC.AAT.GGT.CCT.GTA.AAG.A | 1085 | 58 |
| (RTPCR) | BR8: TTA.GGC.TCT.CTG.AGA.AGA.TT | | |
| Exon 2 | BF9: AGT.TCC.AAT.GGT.CCT.GTA.AAG.A | 286 | 58 |
| | BR1: AAT.AAA.GTC.TAT.GAT.TTC.ACT.ATA | | |
| Exon 3 | BF3: GGC.CTA.CCA.ACA.GCG.GAC | 316 | 55 |
| | BR3: AAT.ATC.TCT.TTC.CAA.TAA.TCA.ACT.AG | | |
| Exon 4 | BF4: TTC.AGG.AAT.TTG.TTT.GCA.GTC.T | 205 | 54 |
| | BR4: ACT.CTA.GAG.AAT.AAT.TCT.CTC | | |
| Exon 5 | BF5: TCT.GAT.AAC.AAT.TCT.CAC.ATT.AAT | 254 | 54 |
| | BR5: TAT.GGA.TTT.ATG.TAC.CCA.AGA.A | | |
| Exon 6 | BF6: CCT.CAT.GGA.GGT.AAA.TAA.ATC | 291 | 55 |
| | BR6: AAG.TCA.AGT.CAA.GAT.ATA.AA | | |

Primers designed to amplify β-sarcoglycan cDNA or β-sarcoglycan exons are indicated with the annealing temperature used for PCR amplification of the corresponding fragment. Positions are indicated according to the sequence reported for β-sarcoglycan cDNA (GenBank U29586).

nity-purified from Goat 20 antiserum using Immobilon P (Millipore) strips of α-dystroglycan fusion protein-D [8]. Monoclonal anti-human laminin α2 chain antibody (clone 5H2) was purchased from Chemicon.

### 2.3. Mutation screening of β-sarcoglycan gene

For β-sarcoglycan cDNA fragments were amplified following reverse-transcription from total RNA prepared from biceps brachii muscle biopsies of patients [24]. Primer sequences used for these experiments are reported in Table 1. The RT-PCR products were directly sequenced and mutations confirmed by allele specific oligonucleotide hybridization or restriction digest using Hinf1 due to the R91C mutation leading to the loss of a restriction site for

this enzyme. Exons 2-6 of the β-sarcoglycan gene including flanking intronic sequences were amplified from patients' DNA and sequenced. The absence of variants in 100 control chromosomes from the normal population was tested on DNA from the CEPH reference families.

## 3. Results

### 3.1. β-Sarcoglycan gene: exon-intron structure and genomic sequence

To facilitate the mutation search and the identification of regulatory domains in the β-sarcoglycan gene, we have determined its genomic sequence and structure. We had

Table 2

Analyses of the human β-sarcoglycan genomic sequence

| Splice acceptor | Score (%)[a] | Exon (size, bp) | Exon~position | Splice donor | Score (%)[a] | Intron (size, bp) | Power of prediction algorithms[c] | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Grail2a | FEXH | HEXON |
| *Exon-intron organization of the β-sarcoglycan gene corresponding to the cDNA* | | | | | | | | | |
| | | 1 (75) | 6958–7032 | AACAGgtctgt | 70.43 | 1 (4569) | Med | Good | – |
| tatttcatttttagCAAAG | 84.00 | 2 (210) | 11602–11811 | TAAATgtgagt | 76.46 | 2 (3546) | Exc | Exc | Exc |
| GtgttcttttcagATAAC | 93.34 | 3 (186) | 15358–15543 | AGCCTgtaagt | 79.00 | 3 (756) | Exc | Med | Med |
| TgttttttttgaagATTGT | 84.79 | 4 (192) | 16300–16491 | AAAGGgtatga | 75.18 | 4 (629) | Good | – | – |
| ttttaaatttttcagATTAC | 90.54 | 5 (132) | 17121–17252 | AGGCGgtgagt | 88.50 | 5 (3800) | Exc | Exc | Exc |
| TattctcctaatagGAAAA | 88.70 | 6 (201) | 21053–21256 | Exc | Good | – | | | |
| | | | | | | | | | |
| *Additional predicted exons*[b] | | | | | | | | | |
| actgtattgtttagGGAAT | 75.10 | A (207) | 10829–11035 | TACAGgtgcgt | 85.22 | – | – | Good | Good |
| tttccttttccagGCATT | 97.81 | B (501) | 27595–28095 | CAAGGgtgagt | 88.86 | – | Exc | Exc | Exc |

~ Position 1 corresponds to the first nucleotide on the sense strand of the sequenced cosmid Cos4 (Accession # Y03781).

[a] A score expressing adherence to the consensus was calculated for each size according to Shapiro and Senapathy [34].

[b] Only those exons showing at least one Excellent or two Good scores are listed.

[c] Med, Good, Exc and – refer to mediocre, good, excellent and no detected scores, respectively.

The values given by BCM were arbitrarily translated into qualitative scores as follows; exc ≥ 9; 5.5 > good > 9; med ≤ 5.5. The scores listed for Grail2a correspond to the best score obtained using either Grail2a and XGrail2a (version 1.3b).
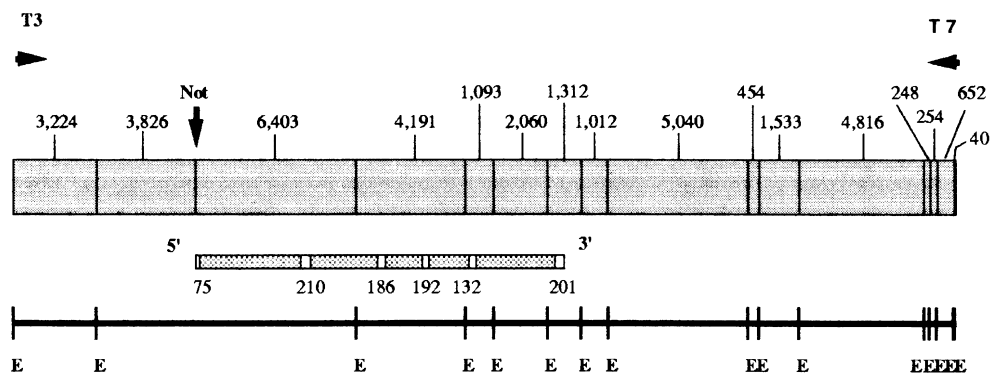
Fig. 1. Exon-intron structure of the β-sarcoglycan gene. EcoRI restriction map of Cos4 DNA was determined. Size in bases of EcoRI fragment is indicated at the top of the cosmid DNA box. Below is the exon-intron organization of the β-sarcoglycan gene. The genomic structure was inferred from comparison with the published cDNA sequences [24,25] and upon analyses of the genomic sequence with several algorithms. The human β-sarcoglycan gene is organized in 6 exons (none of the programs identified the six) whose size range from 75 to 210 bp, separated by 5 introns (from 629 to 4569 bp).

previously isolated a cosmid (Cos4) for the β-sarcoglycan gene [24]. The Cos4 insert (nucleotides 1–36 159; Accession EMBL# Y09781) was entirely sequenced and it will be referred to hereafter as the sense strand. The consensus sequence has been determined with an average 9× coverage value (minimal local value 3×). The genomic organization of the β-sarcoglycan gene was determined from comparison of the published cDNA sequences [24,25] with the genomic sequence using several prediction algorithms. A comparison of the relative power of the latter demonstrated the benefits that can be obtained in the absence of prior exon information by the use of all existing algorithms: individually, none of these programs were able to identify all six exons. Since the 5′ stretch of the cDNA published by Bönnemann et al. [25] extended upstream of that of Lim et al. [24], the former was taken into account. Furthermore, local discrepancies in

the translated parts were easily ascertained by genomic sequencing (6× coverage in these regions).

The human β-sarcoglycan gene (Table 2 and Fig. 1) is organized into 6 exons, which range in size from 75 to 210 bp, separated by 5 introns (from 629 to 4569 bp). All donor and acceptor splicing sequences display the GT/AG consensus. The translation start codon is the ATG at position 7000, and the stop codon at position 21 254 is followed by several putative polyadenylation sites (Fig. 2) which could account for several transcript isoforms. The possible biological roles of these isoforms either in regulating the stability or addressing of the mRNA [35] remain to be elucidated. Analysis of the public domain expressed sequence tag (EST) databases and of the Northern blot expression patterns [24] bring further credence to the biological relevance of these alternative polyadenylation sites. Indeed, we reported the pre-
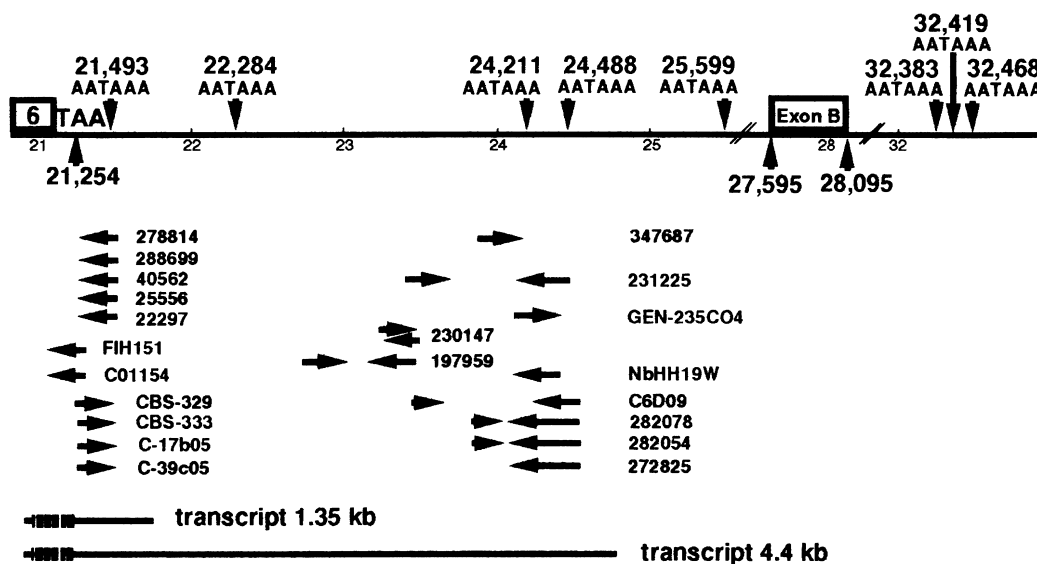


Fig. 2. Analysis of the 3′ untranslated region of the β-sarcoglycan gene. The top lane represents the Cos4 DNA region corresponding to the 3′-UTR of the β-sarcoglycan gene. The scale is in kilobase. The different putative polyadenylation sites AATAAA, the stop codon in exon 6 and predicted exon B are indicated with their position. The arrows represent ESTs and their orientation with the name of the cDNA clones from which they were obtained. The position of the 4.4 and 1.35 kb transcripts detected by Northern blot [24] are shown.

sence of mRNAs of 4.4, 3.0 and 1.35 kb, respectively [24]. The 4.4 and 1.35 kb molecules could result from the use of these alternative polyadenylation sites (Fig. 2), whereas the 3.0 kb transcript still remains unaccounted for. The latter may correspond to the 3′ end EST of the cDNA clones 197959 and 230147 (see Fig. 2), although no polyadenylation site was identified in the corresponding genomic sequence.

Use of the Promoter Scan algorithm predicted a promoter-like structure (from nucleotide 6946 to 7195; score 18.18) upstream of the gene, containing a GC box (position −41 with respect to the initiator ATG), and spanning a 250 bp long CpG rich island (82%) covering exon 1 and part of intron 1. These deductions were in agreement with the predictions of the TSSW program of BCM Gene finder (Human Genome Center, Baylor College of Medicine, Houston, TX, USA).

Three prediction algorithms, Grail 2A, FEXH BCM and HEXON BCM were applied to the genomic DNA sequences; Table 2 shows that none of them detected all known exons with at least a good score, and that several exons were not detected by one or two of these algorithms. In addition, these programs also revealed additional potential exons, two of which had good or excellent scores, respectively. Exon A was ranked good by two algorithms and mapped between exon 1 and exon 2 and would lead, if authentic, to another isoform of this gene. However, its location within an Alu area and its content in repetitive DNA diminished the likelihood that it is translated. In contrast, exon B was scored excellent by the three algorithms. This 501 bp long exon B lies downstream of the recognized β-sarcoglycan cDNA sequence on the same strand (Fig. 2). It could therefore correspond to an as yet undocumented isoform. Alternatively, it could be part of another gene, although no promoter or initiating ATG was significantly identified in silicio. To test these hypotheses, Northern blot as well as RT-PCR experiments on human skeletal, cardiac and brain mRNA were conducted. All tests performed to visualize the corresponding transcripts were negative (data not shown). Either the corresponding transcript was present at too low levels to be detected, or its restricted spatio-temporal pattern of expression prevented its detection. Whereas no homologous amino-acid sequence was recognized upon screening public domain (TIGR, Genbank, EST, *C.elegans* and yeast) databases, two regions showed amino-acid homology with sequences internal to the β-sarcoglycan exon 4. The most significant was the RxxNILF motif which is also found in Mycoplasma (Accession: U02188); *Escherichia coli* (Accession: AE000317), and Chloroplast 30S ribosomal protein (e.g. Accession: P36471). Although this homology does not imply an actual functional role for exon B, it seems likely that the latter is somehow related to the β-sarcoglycan gene. Given the difficulty to assess whether exon B is an authentic exon, it may be worthwhile to investigate the corresponding sequence in other mammals. Additionally, one exon was scored excellent with pro-

grams FEXH, HEXON and FGENEH of BCM Gene Finder and good with Grail 2A (reverse position: 25 750–25 805, data not shown), on the opposite strand of the cosmid insert. This putative 'orphan' exon showed no hit with any published sequence and hence was not integrated onto the map (Table 2).

### 3.2. β-Sarcoglycan gene mutations in the LGMD2E Amish community

In agreement with the common haplotype displayed by the first 11 families analyzed, a missense mutation (T151R) common to all Southern Indiana Amish patients tested has been found and was associated with a mild form of muscular dystrophy [24]. No evidence of cardiomyopathy has been found, and joint contractures have been infrequent in these patients. Since then, two sibs from an Amish family exhibiting heterozygosity for the common haplotype with chromosome 4q12 markers and a second distinct haplotype were identified among the community of Southern Indiana. Subsequently, mutation screening of these patients revealed in addition to the common T151R mutation, a C to T transition corresponding to the substitution of arginine 91 by a cysteine. These two patients (2a and 2b) had a late onset of the disease (Table 3), characterized by muscle weakness and enlargement of calves. At age 52 and 57 years, respectively, both individuals were still able to carry on work requiring physical exertion and were still able to walk. Five other children in the family have not been found to be affected. It is unclear whether the unaffected members of this family are carriers of mutations (i.e. a penetrance issue) or just unaffected genetically (i.e. carrying one or two copies of the normal allele). Surprisingly, two other familial cases not belonging to the Amish community carried this mutation in the homozygous state (Table 3, cases 3a and 3b). No relationship was ascertained between the father and mother, however Amish names were reported in their ancestry. Patient 3a and her affected sister (3b) presented leg cramps and weakness after exercise at age 4 and 11 years, respectively. Calf hypertrophy was noted in early childhood in both patients. Patient 3a was still able to ambulate at age 40 years whereas her affected sister has used a walker since age 37 years. They have two unaffected brothers. The status of the R91C mutation was analyzed by restriction digest of PCR amplified genomic DNA using Hinf1, which is ablated as a result of the mutation. Analysis of 100 chromosomes from the CEPH reference families confirmed that the R91C was a mutation and not a polymorphism.

### 3.3. Analysis of the expression of the components of the DGC

Since the original characterization of the β-sarcoglycan gene defect in 11 LGMD2E Amish families from Southern Indiana [24], two additional sarcoglycan genes have been

cloned and new antibodies developed against these molecules. We examined the expression of the sarcoglycan and dystroglycan complexes, dystrophin, and the α2 chain of laminin 2 in a biopsy from an Amish patient homozygote for the T151R mutation. This analysis extended previous examinations [24,25,30] by using a monoclonal antibody against γ-, and δ-sarcoglycan. In addition, a newly affinity purified anti-α-dystroglycan antibody was developed. The antibody against α-dystroglycan allowed us for the first time to specifically look at the expression of one of the key proteins of the DGC on skeletal muscle cryosections of LGMD patients. The expression level of dystrophin, β-dystroglycan, and the laminin α2 chain in the Amish patient was comparable to the levels found in normal human control muscle. On the other hand, expression of the sarcoglycan complex was strongly reduced in the patient's biopsy compared to control muscle. All sarcoglycan proteins showed a similar level of reduction at the sarcolemma. In addition to the secondary loss of α-, γ-, and δ-sarcoglycan, we also found a reduced expression of the α-subunit of the dystroglycan complex. This unexpected and interesting finding may unveil for the first time a direct or indirect link between the sarcoglycan complex and α-dystroglycan.

### 3.4. Patient screening for β-sarcoglycan gene mutations

In order to identify additional patients carrying a primary β-sarcoglycan deficiency, we examined a total of 107 new independent cases of autosomal recessive myopathy characterized by normal immunostaining of dystrophin and markedly elevated serum creatine kinase level. We primarily screened these patients by immunostaining of skeletal muscle cryosections for α-sarcoglycan deficiency and identified 10 new cases. All these sporadic cases originate in the USA. In all patients, the serum creatine kinase levels were elevated but the clinical pattern was of variable severity, ranging from severe childhood to late onset progressive muscular dystrophy (data not shown). There were no intellectual impairments, nor cardiomyopathy. Skeletal muscle sections of these patients were further analyzed by immunofluorescence staining for the other components of the DGC and a concomitant loss or a drastic reduction of β, γ and δ-sarcoglycan was observed (data not shown). The immunofluorescence data did not indicate the primary defect causing the observed phenotype.

All sporadic LGMD cases were analyzed for mutations in the β-sarcoglycan gene as well as in the other known sarcoglycan genes (data not shown), due to the absence of linkage data. For one of the sporadic cases (T044), only one mutant allele corresponding to the substitution of an isoleucine by a phenylalanine at position 119 (I119F), was found in the β-sarcoglycan gene although all coding sequences were investigated. This patient developed a proximal muscle weakness as well as progressive scoliosis with onset in early childhood. At age 13 years, he was confined to a wheelchair and presented more severe clinical symptoms

as compared to the Amish patients. In this isolated case, the second mutation could be in the unexplored untranslated region or possibly in one of the alternative exons. Allele specific hybridization with either a control oligonucleotide or an oligonucleotide including this latter mutation, revealed that this substitution was not found in 100 independent chromosomes from CEPH reference families (data not shown) and that the sporadic LGMD patient was thus in all likelihood a compound heterozygote for the I119F and a second, as yet unidentified mutation. We also determined the primary sarcoglycan gene defect in 7 out of 9 remaining isolated cases: four patients have a mutated α-sarcoglycan gene and three have a mutated γ-sarcoglycan gene (data not shown).

## 4. Discussion

In order to facilitate further gene expression analysis and the identification of functional regulatory elements, we performed sequence analysis of a 36 kb genomic fragment containing the β-sarcoglycan gene. Interestingly, several putative additional exons were predicted by the computer programs XGrail and/or BCM in the 3′-UTR of the β-sarcoglycan gene, which may be part of different tissue-specific transcripts. One GC box was detected at position 6952 using Promoter scan, and several polyadenylation sites were found (Table 2 and Fig. 2). The relationship between these putative polyadenylation sites and the different transcripts detected by Northern hybridizations at 4.4, 3.3, 1.3–1.4 kb requires further probing with specific 3′-UTR fragments to be ascertained. The tissue-specificity of expression should also be determined. Although experiments to confirm the existence of the two predicted exons gave negative results it was not possible to conclude whether these two exons form part of real mRNA. This study constitutes a good example of what to expect from the human genome sequencing project, since it illustrates how far one can get inferring relevant genomic information just from sequencing data. On the one hand it highlights some of the limits of the current prediction algorithms, and on the other hand it points also to the presence of potential false positives.

Although a common chromosome 4q11–q12 haplotype is displayed by most of the Southern Indiana Amish LGMD2E families around the disease locus [24], two patients from one family were compound heterozygotes in this region. Mutation screening revealed that these patients carried the common T151R mutation [24] together with another missense mutation, R91C. In addition, we identified the same mutation (R91C) in an unrelated familial case (Table 3, Amish 3) who does not belong to the Southern Indiana Amish community, but whose ancestors have Amish names on both paternal and maternal sides. The presence of a second mutation in the Amish community was unexpected, although the presence of multiple mutations in a small ethnic community has already been observed before.

Table 3

Clinical and biochemical data of patients with LGMD

| Case no. | Age (years) | Sex | Onset (years) | CK value at (age) | Calf muscle hypertrophy | Activity | β-Sarcoglycan gene defect |
|---|---|---|---|---|---|---|---|
| Amish 1 | 9 | F | Presymptomatic | 12480 (5) | + | Ambulant | T151R |
| Amish 2a | 54 | F | 24 | 1160 (36) | + | Ambulant | T151R–R91C |
| Amish 2b | 57 | F | 43 | ? | ? | Ambulant | T151R–R91C |
| Amish 3a | 39 | F | 11 | ? | + | Walker | R91C |
| Amish 3b | 29 | F | 4.5 | 6140 (26) | + | Ambulant | R91C |
| T044 | 15 | M | ? | ? | ? | Wheelchair | I119F–? |

Creatine kinase levels (CK) measured at (age) of the patients are indicated. Normal CK values are estimated to be lower than 100. Severity of the condition of the patients at present is indicated by their activity. No cardiomyopathy nor intellectual impairment have been reported in these cases. The second allele mutation in the sporadic case T044 was not determined despite the analysis of the whole coding sequence of the β-sarcoglycan gene. Amish 1 refers to one patient from one of the 11 families originally characterized with the common mutation T151R for which we had a biopsy available. Amish 2 refers to the two LGMD2E patients from an Amish family heterozygote for the common T151R and the R91C mutations. Amish 3 refers to two affected sibs from a LGMD2E family of probable Amish descent which was identified independently, and shown to carry an homozygous R91C mutation.

The presence of several mutations in the calpain 3 gene among LGMD2A patients from La Réunion Island was puzzling [36]. A hypothesis that could account for this and other related observations has been reported [36–38]. The identification of a second mutation in another family of probable Amish origin suggests a relatively ancient occurrence of this mutation, in contrast to the possibility of a high mutation rate coupled with a recent occurrence of these mutations in this isolate as proposed by Zlotogora et al. [39] for Hurler syndrome, metachromatic leukodystrophy and LGMD2A. In such highly inbred populations, one mutation associated with a founder haplotype may become frequent by result of a genetic drift. This frequent haplotype/mutation can now uncover other rare events leading possibly to these multiple disease alleles we observe now in the Amish isolate from Southern Indiana. These homozygotes (R91C; codon CGC mutated to TGC) displayed a mild muscular dystrophy phenotype (Table 3, Amish 3a and 3b). Surprisingly, Bönnemann et al. [30] reported a R91P mutation (codon CGC changed to CCC) in three homozygous sibs from a Brazilian family associated with a severe form of LGMD. Secondary structure prediction analysis suggested that the mutation R91C did not affect the protein structure whereas the proline at position 91 did. These findings suggest that the arginine 91 in β-sarcoglycan may constitute a critical site of interaction in the complex. These homozygote cases may be helpful in unraveling some genotype/phenotype correlations in LGMD2E and in providing clues on the sarcoglycan/dystroglycan interaction.

We have also analyzed 107 patients with normal dystrophin and identified 10 sporadic cases with a sarcoglycan complex deficiency. These cases represent 10% of the muscular dystrophy patients tested and this correlates with the ratio obtained by a recent study [31]. These patients were investigated for sarcoglycan mutations in the α-, β-, γ- and δ-sarcoglycan genes and a sarcoglycan molecular defect was revealed in 8 out of 10 (data not shown). These two latter may carry mutations in another yet unidentified sar-

coglycan gene. We identified one LGMD patient (Table 3, case T044) carrying on one chromosome a novel missense mutation (I119F) in the β-sarcoglycan gene, the second allele remaining unidentified despite the fact that all coding sequences were investigated. Interestingly, many of the T151R and R91C Amish cases are associated with a milder course of the disease in contrast to the severe phenotype of LGMD2E patients reported [25,30] and one LGMD2E case (T044) reported in this study. All mutations mapped to the extracellular domain of the protein. The β-sarcoglycan gene mutation R91C (**CGC** to **TGC**) abolished a CpG site. The functional importance of the region in which the mutation occurred has not yet been defined. On the other hand, missense mutations can cause abnormal protein processing or stability, as was seen in the majority of cases with LGMD2 [22–32]. In LGMD2D, a broad spectrum of missense mutations in the α-sarcoglycan gene have been reported which affected protein targeting to the sarcolemma and resulted in muscular dystrophy of varying degrees of severity [29,32]. This phenomenon has already been observed with other transmembrane proteins, including CFTR [40].

The close association of the components of the sarcoglycan complex [6,12] and their concomitant loss in LGMD2C–F have been reported. In the BIO14.6 cardiomyopathic hamster, an animal model of δ-sarcoglycan deficient LGMD [41], a secondary reduction of α-dystroglycan expression had been reported [42]. To test whether the stability of α-dystroglycan is also affected in human sarcoglycan deficient skeletal muscle, we used a new anti-α-dystroglycan antibody and demonstrated for the first time that a missense mutation in the β-sarcoglycan gene (T151R) can modify α-dystroglycan expression (Fig. 3). This result suggests a possible link between the sarcoglycan complex and the peripheral membrane protein α-dystroglycan. This link could either be mediated through direct interaction or via an intermediate. The maintenance of β-dystroglycan expression in the sarcolemma of this patient supports the hypothesis that the sarcoglycan complex is required for high
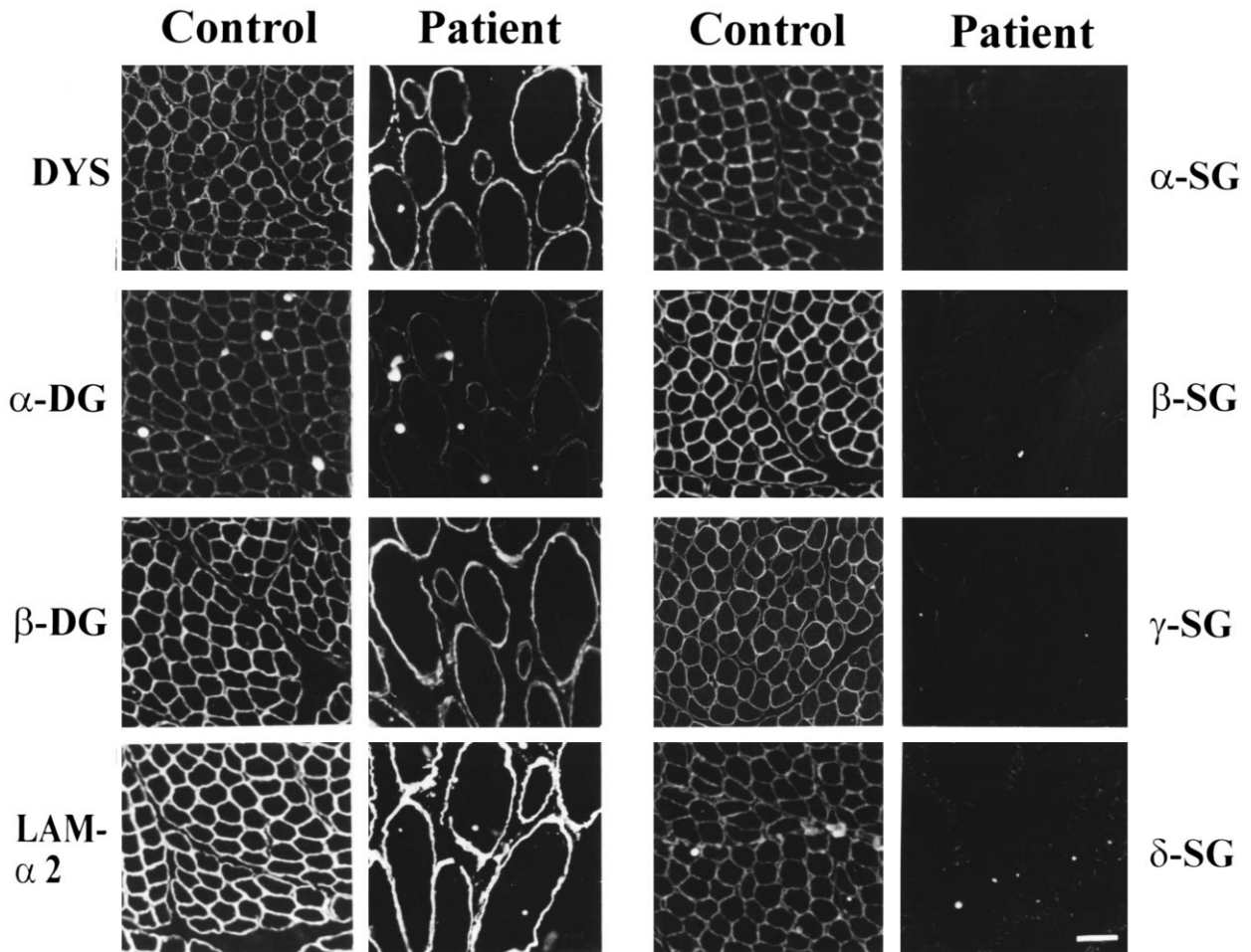
Fig. 3. Immunofluorescence analysis of DGC components in human control and Amish LGMD2E muscle. The LGMD2E patient (Patient) carried a homozygous T151R mutation. Seven micrometer cryosections were stained with dystrophin (Dys), α-dystroglycan (α-DG), β-dystroglycan (β-DG), laminin α2 (LAM-α2), α-, β-, γ-, and δ-sarcoglycan antibodies (α-SG, β-SG, γ-SG and δ-SG) as described in Section 2. The entire sarcoglycan complex is greatly reduced at the sarcolemma of skeletal muscle from the β-sarcoglycan mutated patient compared to control. Additionally, the sarcolemmal staining for α-dystroglycan is reduced. Bar, 50 $\mu$M.

affinity interaction of α-dystroglycan with the sarcolemma. Further investigation of muscle biopsies from sarcoglycan deficient patients are currently being performed.

The genetic and clinical heterogeneity of the limb-girdle muscular dystrophies has broad implications for the diagnosis and management of persons afflicted with this disease. Because defects in several distinct proteins can cause similar clinical symptoms and biochemical manifestations, a definitive diagnosis can so far only be achieved through genetic testing of all candidate genes. Analysis of additional cases will provide more insights in genotype/phenotype correlations and key domains of the proteins.

## Acknowledgements

## References

[1] Campbell KP, Kahl SD. Association of dystrophin and an integral membrane glycoprotein. Nature 1989;338:259–262.

[2] Ervasti JM, Ohlendieck K, Kahl SD. et al. Deficiency of a glycoprotein component of the dystrophin complex in dystrophic muscle. Nature 1990;345:315–319.

[3] Yoshida M, Ozawa E. Glycoprotein complex anchoring dystrophin to sarcolemma. J Biochem 1990;108:748–752.

[4] Ervasti JM, Kahl SD, Campbell KP. Purification of dystrophin from skeletal muscle. J Biol Chem 1991;266:9161–9165.

[5] Ervasti JM, Campbell KP. Membrane organization of the dystrophin glycoprotein complex. Cell 1991;66:1121–1131.

[6] Yoshida M, Suzuki A, Yamamoto H. et al. Dissociation of the complex of dystrophin and its associated proteins into several unique groups by n-octyl-D-glucoside. Eur J Biochem 1994;222:1055–1061.

[7] Mizuno Y, Noguchi S, Yamamoto H. et al. Selective defect of sarcoglycan complex in severe childhood autosomal recessive muscular dystrophy muscle. Biochem Biophys Res Commun 1994;203:979–983.

[8] Ibraghimov-Beskrovnaya O, Ervasti JM, Leveille CJ. et al. Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular matrix. Nature 1992;355:696–702.

[9] Adams ME, Butler MH, Dwyer TM. et al. Two forms of mouse syntrophin, a 58 kD dystrophin-associated protein, differ in primary structure and tissue distribution. Neuron 1993;11:531–540.

[10] Yang B, Ibraghimov-Beskrovnaya O, Moomaw CR. et al. Heterogeneity of the 59 kDa dystrophin-associated protein revealed by cDNA cloning and expression. J Biol Chem 1994;269:6040–6044.

[11] Hoffman EP, Brown RH Jr, Kunkel LM. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. Cell 1987;51:919–928.

[12] Jung D, Leturcq F, Sunada Y. et al. Absence of γ-sarcoglycan (35DAG) in autosomal recessive muscular dystrophy linked to chromosome 13q12. FEBS Lett 1996;381:15–20.

[13] Ervasti JM, Campbell KP. A role for the dystrophin-glycoprotein complex as a transmembrane linker between laminin and actin. J Cell Biol 1993;122:809–823.

[14] Campbell KP. Three muscular dystrophies: loss of cytoskeleton-extracellular matrix linkage. Cell 1995;80:675–679.

[15] Hoffman E, Kunkel LM. Dystrophin abnormalities in Duchenne/Becker muscular dystrophy. Neuron 1989;2:1019–1029.

[16] Helbling-Leclerc A, Zhang X, Topaloglu H. et al. Mutations in the laminin α2chain gene (*LAMA2*) cause merosin-deficient congenital muscular dystrophy. Nature Genet 1995;11:216–218.

[17] Nissinen M, Helbling-Leclerc A, Zhang X. et al. Substitution of a conserved cysteine-996 in a cysteine-rich motif of the laminin α2-chain in congenital muscular dystrophy with partial deficiency of the protein. Am J Hum Genet 1996;58:1177–1184.

[18] Allamand V, Sunada Y, Salih MAM. et al. Mild congenital muscular dystrophy in two patients with an internally deleted laminin α2-chain. Hum Mol Genet 1997;6:747–752.

[19] Stevenson AC. Muscular dystrophy in Northern lreland. Ann Eugen 1953;18:50–91.

[20] Walton JN, Nattrass FJ. On the classification, natural history and treatment of the myopathies. Brain 1954;77:169–231.

[21] Jackson CE, Strehler DA. Limb-girdle muscular dystrophy: clinical manifestations and detection of preclinical disease. Pediatrics 1968;41:495–502.

[22] Noguchi S, McNally EM, Ben Othmane KB. et al. Mutations in the dystrophin associated protein γ-sarcoglycan in chromosome 13 muscular dystrophy. Science 1995;270:819–822.

[23] Roberds SL, Leturcq F, Allamand V. et al. Missense mutations in the adhalin gene linked to autosomal recessive muscular dystrophy. Cell 1994;78:625–633.

[24] Lim LE, Duclos F, Broux O. et al. β-Sarcoglycan: characterization and role in limb-girdle muscular dystrophy linked to 4q12. Nature Genet 1995;11:257–265.

[25] Bönnemann CG, Modi R, Noguchi S. et al. Mutations in the dystrophin-associated glycoprotein β-sarcoglycan (A3b) cause autosomal recessive muscular dystrophy with disintegration of the sarcoglycan complex. Nature Genet 1995;11:266–273.

[26] Nigro V, de Sa Moreira E, Piluso G. et al. Autosomal recessive limb-girdle muscular dystrophy, LGMD2F, is caused by a mutation in the δ-sarcoglycan gene. Nature Genet 1996;14:195–198.

[27] Jung D, Duclos F, Apostol B. et al. Characterization of δ-sarcoglycan, a novel component of the oligomeric sarcoglycan complex involved in limb-girdle muscular dystrophy. J Biol Chem 1996;271:32321–32329.

[28] Matsumura K, Tomé FMS, Collin H. et al. Deficiency of the 50K dystrophin-associated glycoprotein in severe childhood autosomal recessive muscular dystrophy. Nature 1992;359:320–322.

[29] Piccolo F, Roberds SL, Jeanpierre M. et al. Primary adhalinopathy: a common cause of autosomal recessive muscular dystrophy of varying severity. Nature Genet 1995;10:243–245.

[30] Bönnemann CG, Passos-Bueno MR, McNally EM. et al. Genomic screening for β-sarcoglycan gene mutations: missense mutations may cause severe limb-girdle muscular dystrophy type 2E (LGMD2E). Hum Mol Genet 1996;5:1953–1961.

[31] Duggan DJ, Gorospe JR, Fanin M. et al. Mutations in the sarcoglycan genes in patients with myopathy. N Engl J Med 1997;336:618–624.

[32] Carrie A, Piccolo F, Leturcq F. et al. Mutational diversity and hot spots in the α-sarcoglycan gene in autosomal recessive muscular dystrophy (LGMD2D). J Med Genet 1997;34:470–475.

[33] Jorgensen AO, Arnold W, Shen AC. et al. Identification of novel proteins unique to either transverse tubules (TS28) or the sarcolemma (SL50) in rabbit skeletal muscle. J Cell Biol 1990;110:1173–1185.

[34] Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res 1987;15:7155–7174.

[35] Kislauskis EH, Zhu X, Singer RH. Sequences responsible for intracellular localization of beta-actin messenger RNA also affect cell phenotype. J Cell Biol 1994;127:441–451.

[36] Richard I, Broux O, Allamand V. et al. Mutations in the proteolytic enzyme calpain 3, cause limb girdle muscular dystrophy type 2A. Cell 1995;87:27–40.

[37] Allamand V, Broux O, Bourg N, et al. Genetic heterogeneity of autosomal recessive limb-girdle muscular dystrophy in a genetic isolate (Amish) and evidence for a new locus. Hum Mol Genet 1995;4(3):459–464.

[38] Beckmann JS. The Réunion paradox and the digenic model. Am J Hum Genet 1996;59:1400–1402.

[39] Zlotogora J, Gieselmann V, Bach G. Multiple mutations in a specific gene in a small geographic area: a common phenomenon? Am J Hum Genet 1996;58:241–243.

[40] Welsh MJ, Smith AE. Molecular mechanisms of the CFTR chloride channel dysfunction in cystic fibrosis. Cell 1993;73:1251–1254.

[41] Nigro V, Okasaki Y, Belsito A. et al. Identification of the Syrian hamster cardiomyopathy gene. Hum Mol Genet 1997;6:601–607.

[42] Roberds SL, Ervasti JM, Anderson RD. et al. Disruption of the dystrophin-glycoprotein complex in the cardiomyopathic hamster. J Biol Chem 1993;268:11496–11499.