European Journal of Human Genetics (1998) 6, 325–330

http://www.stockton-press.co.uk/ejhg

ORIGINAL PAPER

Genomic organization of the *MTM1* gene implicated in X-linked myotubular myopathy

Jocelyn Laporte[1], Christophe Guiraud-Chaumeil[1], Stephan M Tanner[2], François Blondeau[1], Ling-Jia Hu[1], Serge Vicaire[1], Sabina Liechti-Gallati[2] and Jean-Louis Mandel[1]

[1]*Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS, Illkirch Cedex, France*
[2]*Department of Clinical Research & Department of Pediatrics, University of Berne–Inselspital, Berne, Switzerland*

X-linked recessive myotubular myopathy (XLMTM) is a very severe congenital muscular disease characterised by an impaired maturation of muscle fibres, and caused by defects in the *MTM1* gene. This gene defines a new family of putative tyrosine phosphatases conserved through evolution. We have determined intronic flanking sequences for all the 15 exons to facilitate the detection of mutations in patients and genetic counselling. We characterised a new polymorphic marker in the immediate vicinity of the gene, which might prove useful for linkage analysis. Sequencing of the TATA-less predicted promoter provides the basis for transcriptional regulatory studies.

Keywords: myotubular myopathy; promoter; genomic structure; linkage; *MTM1*; tyrosine phosphatase

Introduction

X-linked recessive myotubular myopathy (XLMTM OMIM31040) is a very severe congenital muscular disease characterised by hypotonia and generalised muscle weakness in newborn males. Most patients die within the first months of life. Characteristic muscle histopathology consists of small rounded muscle cells with centrally located nuclei resembling foetal myotubes and it has been suggested that the disorder results from an arrest in the normal development of muscle fibres, probably during late myogenesis.[1] The locus responsible for the disease (*MTM1*) was mapped to proximal Xq28 by linkage analysis and the candidate region was restricted by analysing recombination events and patients carrying deletions.[2,3] We have recently reported the identification of the *MTM1* gene by positional cloning.[4] It encodes a protein (myotubularin) with a putative tyrosine phosphatase domain (PTP). Tyrosine phosphatases have been shown to be involved in transduction pathways controlling cell growth and differentiation.[5] Myotubularin is highly conserved through evolution and several homologous human genes were found, which define a new family of putative tyrosine phosphatases in man.

In this study we have determined the complete exon–intron structure of the human *MTM1* gene and 2.7 kb of upstream sequences in 5'. This information is very useful for the characterisation of mutations in XLMTM patients, a prerequisite for reliable genetic counselling and determination of carrier status.[6] A new polymorphic marker, located less than 3 kb upstream from the *MTM1* gene, may allow a more reliable linkage analysis in familial cases.

Correspondence: Jean-Louis Mandel, IGBMC, 1 rue Laurent Fries, BP163, 67404 Illkirch Cedex, France. Tel: (33) 388 65 32 44. Fax: (33) 388 65 32 46. E-mail: mtm@igbmc.u-strasbg.fr
Received 26 September 1997; revised 22 December 1997; accepted 24 December 1997

Materials and Methods

Four cosmids (ICRF C104-H05147, C104-E03147, C104-B03151 and a Xq28 specific library cosmid Qc3A3) were subcloned as described.[3] Subclones were screened by hybridisation with radioactively labelled cDNAs or oligonucleotides, and by PCR, for the presence of MTM1 exons and further sequenced with exon specific primers. As exons 5 to 8 were not included in our cosmid contig, PCR and long-range PCR from genomic DNA was performed with exonic primers, to clone introns 4, 5, 6, 7 and 8. For long-range PCR analysis of intron sizes, 300 ng of genomic DNA was amplified using the Expand Long Template PCR System kit (Boehringer Mannheim, Meylan, France) for 30 cycles with an elongation time ranging from 4 to 15 minutes, and to assess their size more precisely each product was digested with Bam HI, Cla I, Hpa I, Pvu II and Sac I and compared to size marker (from Boehringer Mannheim, Meylan, France) (Figure 1). Long-range PCR was unable to amplify intron 1 and its size was estimated from hybridisation experiments using specific probes from exons 1 and 2 on the previously characterised cosmid contig, digested with Bam HI and BglII or EcoRI. Exon 1 is present in cosmid Qc3A3 from a Xq28 library, while exon 2 is present in the overlapping cosmid C104-B03151 from Lawrence Livermore National Laboratories.[3]

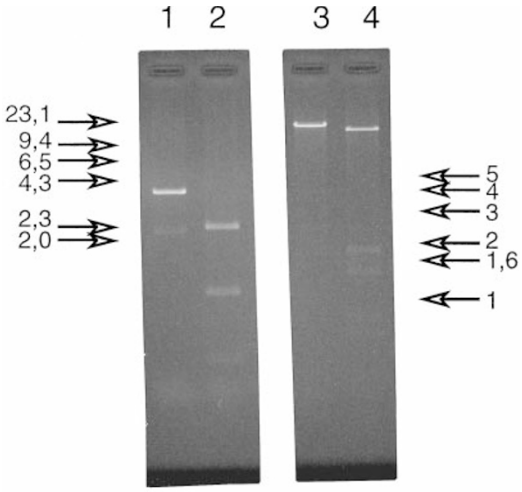

Figure 1 *Long range PCR analysis of the size of introns 2 and 6. Long range PCR was performed as described in methods using exon specific primers VN14 (5' A G A A C C T G TA A A G T A G T A C C 3') and VC113 (5'AGACTTCTCCTCAAGTTATGC3') for intron 2 (annealing at 52°C with a 4 min elongation step) and GT003 (5'CTCACGAGATACGCGTTTC3') and TW57 (5'CTGCCTCCTGTATTCTTCC3') for intron 6 (annealing at 56°C with a 15 min elongation step). PCR fragment sizes were estimated on a 0.8% agarose gel either undigested and after digestion with HpaI, PvuII, and SacI (the latter two are not shown). Lane 1 and 2 correspond to intron 2 (undigested and HpaI digested respectively) and lane 3 and 4 the corresponding products for intron 6. Size markers are in the exterior lane. Size of intron 2 and intron 6 were estimated at 3.8 kb and 20.4 kb respectively.*

A polymorphic CA repeat was characterised in the putative promoter sequence at position –2624 to –2587 of the sequence in Figure 2. Briefly, 100 ng of patient DNA were amplified in 10 mM tris-HCl pH 8.3, 50 mM KCl, 200 μM dNTP, 1.5 mM $MgCl_2$, with 10 pmoles of each primers, 1 nmole (3 μCi) of ^{32}P dCTP and 1 unit of Taq polymerase in a total volume of 25 μl. Thermocycling was carried out in the PTC-100 Programmable Thermal Controller (MJ Research, Inc., USA) at 94°C for 5 min followed by 30 cycles at 94°C for 10 s, 53°C for 10 s, 72°C for 10 s, and a final elongation step at 72°C for 5 min. After PCR, the products were denaturated at 94°C for 10 min with 6 μl of SSCP loading buffer (0.2 M NaOH, 0.25% w/v each bromophenol blue and xylene cyanol in formamide) and separated on a 6% native polyacrylamide gel. The sequence data described in this paper have been submitted to GenBank under accession numbers AF020663–AF020676.

Results and Discussion

The human *MTM1* gene is composed of 15 exons ranging from 44 to 207 bp (Table 1), the last exon been 223 or 1713 bp in length, depending on the polyadenylation signal used.[4] The predicted sizes of polyadenylated transcripts are about 3.6 and 2.1 kb, compared to the previously reported sizes of 3.9 and 2.4 kb observed on northern blots. This difference may result from experimental errors in sizing and/or reflects a 5' extension of the transcript with respect to the cloned cDNAs. This gene covers around 100 kb (between 96 and 104 kb) in the Xq28 proximal region. Exon 11 encodes the predicted tyrosine phosphatase active site and exon 1 is non-coding. Phase 0 introns (uninterrupted codon) are highly predominant (12 out of 13), whilst phase 1 (splice site after the first codon position) accounts for only one intron and phase 2 (splice site after the second nucleotide) is not present. Alternative spliced transcripts have not been reported so far, but if they exist, they will have a high chance of keeping the same open reading frame as the initial transcripts. Moreover, nucleotide changes affecting splice sites may still give a coding transcript lacking a portion of the protein if they lead to exon-skipping.

All of the exon-intron boundaries conform to the AG/GT rule, except for the donor site of exon 10 where a GC is found instead of GT in every patient and normal DNA analysed. This has already been described for a few other examples.[7] A polymorphism at position 1314 + 3 in the donor site of intron 11 was found, where A or G were equally present in the population,[6] and both nucleotides fit with the consensus sequence. The intronic sequences flanking the *MTM1* exons and the sequence of the predicted promoter can now be used to derive primers and screen for mutations in patients,

especially for patients for whom only genomic DNA is available. This has very recently allowed the identification of 79 mutations in 124 patients analysed and confirmed the highly heterogeneous nature of such mutations.[6,8]

We have recently sequenced a full length cDNA clone corresponding to one of the two homologous genes present in *Caenorhabditis elegans* (unpublished results, Laporte J, 1997). Comparison with the genomic sequence in the ACeDB database[9] showed that the

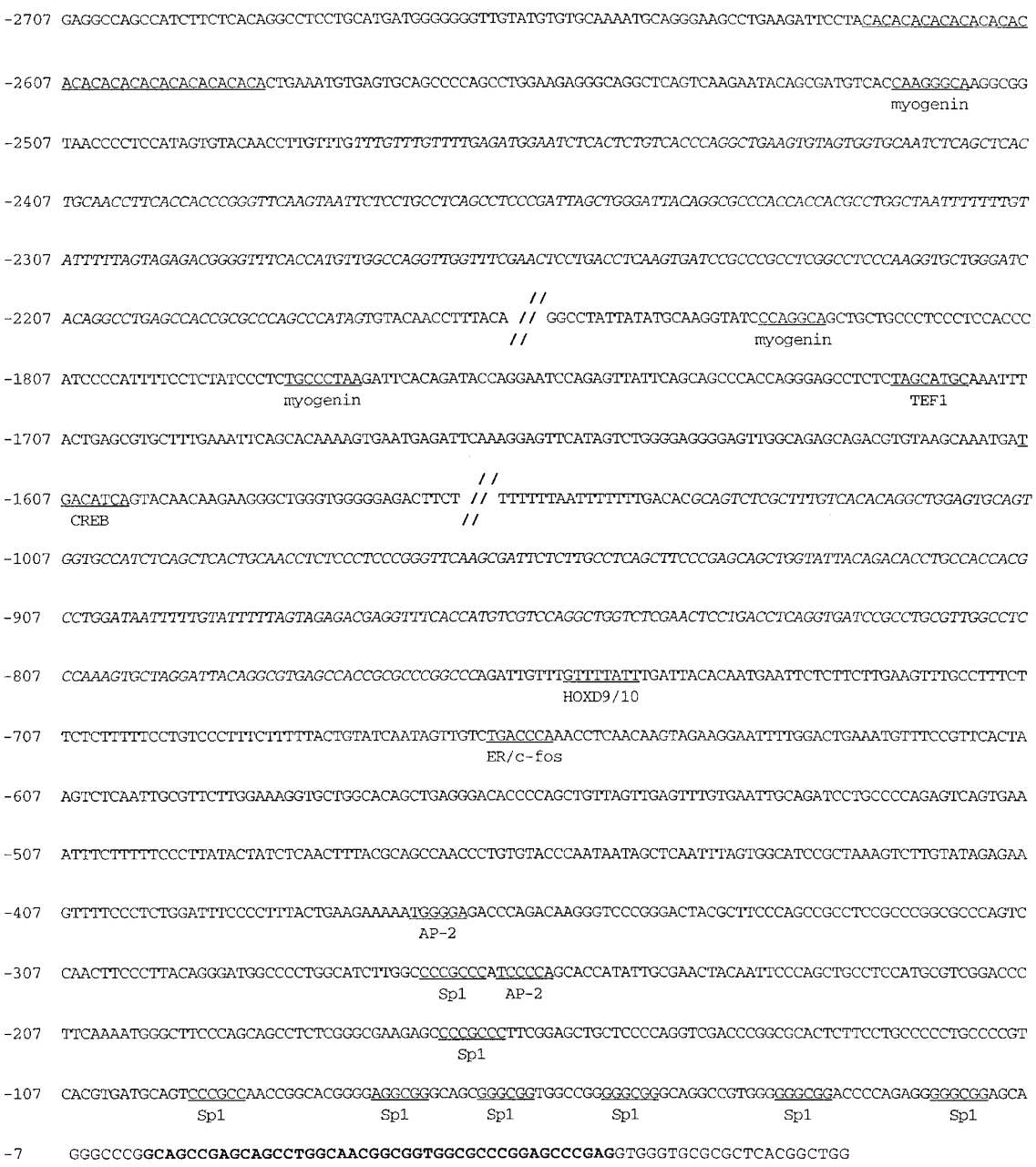

Figure 2 *Human MTM1 putative promoter. This sequence is to be found in GenBank under accession number AF020663. The exon 1 sequence is shown in bold and the negative numbering on the left is based on the first nucleotide of exon 1. Two Alu sequences are in italic and the polymorphic CA repeat (DXS9929) is underlined. The TESS database (http://agave.humgen.upenn.edu/tess/) was screened on a string-based search using the* homo sapiens *subset, and the predicted transcription factors binding sites are underlined with the name of the transcription factor written below. Part of the sequence that lacks obvious features was omitted.*

Table 1 Exon-intron organization of the *MTM1* gene

Exon[a]	*Exon size bp*	*3′ splice site*[c] *acceptor*	*5′ splice site*[c] *donor*	*Intron phase*	*Intron size kb*
1	44		GCCCGAG gtgggtgcgc	/	22–30
2	73	ttcttggtag AGTTTCC	– TAAGAGG gtaagttgaa	0	3.8
3	73	ttgtttctag ACGTCTC	– ATCACTG gtaaggacct	1	2.0
4	95	atctaatag ACAAAGA	– GGAAACG gtaagtagaa	0	14.2
5	111	tccattacag GATTCTT	– TTGTAAA gtaagagatt	0	1.3
6	102	ttgcccccag GACATGA	– CAGTCTG gtaaattcca	0	20.4
7	84	ccttttctag CCATTAT	– GAGGCAG gtaagatgtt	0	2.6
8	150	ttcctcacag GGCTTGC	– AATTCCA gtgagtactg	0	4.5
9	189	ttaaccatag GTGCTGT	– CAACAAG gtgagtggac	0	3.9
10	186	tgtttcttag GCAACAG	– TATCAAG gcaagtatat	0	7.9
11 (PTP)	207	tctttgtcag CTCGTTT	– TGCATCT gtragtaaac	0	1.7
12	93	atttattcag CGAATAG	– AAAACAG gtaaggaata	0	0.6
13	114	ttttgtttag TTCCCTA	– AAGACAG gtgagttaaa	0	3.1
14	177	ctctctgtag AAGGTTA	– GCAACAA gtaagtgaag	0	8.2
15[b]	1713	ttctctccag CAGCCGA			

[a]Exons 8, 9, 11, 14 and 15 correspond to exons named respectively a, b, c, d and e in reference 4; [b]In the additional 2.4 kb muscle specific transcript, exon 15 is only 223 bp long due to the use of a different polyadenylation signal;[4] [c]Exon and intron sequences are shown in upper and lower case, respectively; GenBank accession numbers for the sequence data: AF020663–AF020676.

genomic structure is not conserved compared with the human *MTM1* gene and that there is no preferential intron phase.

Because *cis* elements of a promoter are generally arrayed within a few hundred base pairs of the mRNA initiation site, we determined 2.7 kb of sequence directly upstream from exon 1 (Figure 2). A polymorphism at position –110 was observed (T instead of C) in two patients (out of 60 patients tested) and in two normal X-chromosomes (out of 40 tested). One of the patients had already a frameshift mutation in exon 2 and the second had a mutation in exon 4 (patient N1-44 in ref 6 and unpublished data, Tanner SM, 1997, respectively).

In the original description of the cDNA sequence, the location of the translation initiation site had not been determined. Moreover, an open reading frame of 426 nt is present proximal to the first cDNA base. 5' RACE experiments and numerous cDNA libraries screening with 5' probes led to several clones starting at position +1 in Figure 2.[4] RT-PCR using exon 1 primers amplified the expected product while nothing was amplified using primers proximal to position +1 (not shown). More recent data using antibodies directed against myotubularin indicate that the endogenous protein is translated starting from the first methionine codon of the published cDNA sequence, corresponding to codon 19 (unpublished results, Laporte J, 1997).

The sequence between position –370 and the end of exon 1 is very GC-rich (73%) and exhibits rare CpG cutter sites (BssHII, EagI, NarI, NotI, SacII). It contains several predicted SP1 sites but no TATA-box or CAAT-box, as is often the case for a CpG island region upstream of a housekeeping gene.[10] The *MTM1* gene has been shown to be ubiquitously expressed with an additional shorter transcript that uses a different polyadenylation signal in muscle and testis.[4] No eukaryotic pol II promoter was predicted by the GRAIL program although the TSSG BCM program[11] predicted with a very high score (LDF-33.10) a TATA-less promoter in the neighbourhood of exon 1 in the expected orientation, with a putative transcription start site at position +37 of the *MTM1* cDNA. Other transcription factor sites were predicted by the TESS program,[12] especially three myogenin binding sites, which might be important for muscle-specific transactivation and differentiation. Fifteen putative E-box (CANNTG) elements are present in the predicted promoter sequence and the MyoD transcription factor has been shown to specifically activate muscle genes by binding to the E-box.[13] As myotubular myopathy seems to be due to a defect in late myogenesis, the regulation of the *MTM1* gene transcription or/and regulation of myotubularin activity might be an essential step toward the complete maturation of muscle fibres.

It is not uncommon that two genes start from the same CpG in opposite directions and several transcripts were reported in the neighbourhood of the *MTM1* gene,[14] especially XAP83 located proximal. The sequence of the putative promoter did not detect ESTs in dbEST and no exons were predicted by GRAIL[15] and BCM[11] programs, with the exception of exon 1

Table 2 Characterisation of a new polymorphic CA repeat (DXS9929)

Repeat	*(CA) 18–24*	*Alleles CA number*	*PCR size bp*	*Frequency %*
Primers	5′TGTCCTGGAGCCCTGGAG 3′			
	5′CGCTGTATTCTTGACTGAGC 3′	24	210	1
		23	208	20
Heterozygosity	65%	22	206	4
		21	204	7
		20	202	3
		19	200	54
		18	198	11

which was predicted, although not precisely, by FEXH BCM program with a score of 7.86.

A CA repeat is present at position –2624 to –2587 and contains 19 repeats. Analysis of 118 unrelated Caucasian X chromosomes with flanking primers revealed seven alleles ranging from 18 to 24 CA dinucleotides (Table 2). The predominant (CA)19 allele accounts for 54% of the chromosomes tested. The frequency of each allele is reported in Table 2 and the calculated heterozygosity is 65%. This new polymorphic marker, DXS9929, should prove useful for diagnosis by linkage analysis in familial cases, in combination with previously characterised DXS8377, DXS7423 and DXS1684.[16] DXS9929 is the closest characterised polymorphic marker from the *MTM1* gene and may be more reliable than for example, DXS8377 which is located around 200 kb proximal, and was already found to recombine in an XLMTM family.[16]

Acknowledgements

We wish to thank Christine Kretz and Vreni Schneider for excellent technical assistance. This study was supported by the Institut National de la Santé et de la Recherche Médicale, the Centre National de la Recherche Scientifique, the Centre Hospitalier Universitaire Régional, and by grants from the Association Française contre les Myopathies (AFM) and the Swiss National Research Foundation (No. 32-043631.95).

References

1 Fardeau M: Congenital myopathies. In Mastaglia FL & Walton of Detchant (eds). Skeletal Muscle Pathology. Churchill Livingstone: Edinburgh, 1992, pp 1504–1505.
2 Thomas N, Williams H, Cole G *et al*: X-linked centronuclear/myotubular myopathy: evidence for linkage to Xq28 DNA marker loci. *J Med Genet* 1990; **27**: 284–287.
3 Hu LJ, Laporte J, Kress W *et al*: Deletions in Xq28 in two boys with myotubular myopathy and abnormal genital development define a new contiguous gene syndrome in a 430 kb region. *Hum Mol Genet* 1996; **5**: 139–143.
4 Laporte J *et al* (Gp 1), Kioschis P *et al* (Gp 2), Dahl N (Gp 3): A gene mutated in X-linked myotubular myopathy defines a new putative tyrosine phosphatase family conserved in yeast. *Nat Genet* 1996; **13**: 175–182.
5 Fauman E, Saper MA: Structure and function of the protein tyrosine phosphatases. *Trends Biotech Sci* 1996; **21**: 413–417.
6 Laporte J *et al* (Gp 1), Tanner SM *et al* (Gp 2), Wallgren-Pettersson C *et al* (Gp 3): Mutations in the *MTM1* gene implicated in X-linked myotubular myopathy. *Hum Mol Genet* 1997; **6**: 1505–1512.
7 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.
8 de Gouyon B, Zhao W, Laporte J, Mandel JL, Metzenberg A, Herman G: Characterization of mutations in the recently identified myotubularin gene in twenty-six patients with X-linked myotubular myopathy. *Hum Mol Genet* 1997; **6**: 1499–1504.
9 Durbin R, Mieg JT: A *C. elegans* database. Documentation, code and data available from anonymous FTP servers at lirmm.lirmm.fr, cele.mrc-lmb.cam.ac.uk and ncbi.nlm.nih.gov..
10 Dynan WS: Promoters for housekeeping genes. *Trends Genet* 1986; **2**: 196–197.
11 Solovyev VV, Salamov AA, Lawrence CB: Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. *Nucleic Acids Res* 1994; **22**: 5156–5163.
12 Boulikas T: A compilation and classification of DNA binding sites for protein transcription factors from vertebrates. *Crit Rev Euk Gene Expression* 1994; **4**: 117–321.
13 Lassar AB, Buskin JN, Lockshon D *et al*: MyoD is a sequence-specific DNA binding protein requiring a region of *myc* homology to bind to the muscle creatine kinase enhancer. *Cell* 1989; **58**: 823–831.
14 Kioschis P, Rogner UC, Pick E *et al*: A 900-kb cosmid contig and 10 new transcripts within the candidate region for myotubular myopathy (MTM1). *Genomics* 1996; **33**: 365–373.

15 Xu Y, Mural R, Shah M, Uberbacher E: Recognizing exons in genomic sequence using GRAIL II. *Genet Eng* 1994; **16**: 241–253.

16 Hu LJ, Laporte J, Kioschis P *et al*: X-linked myotubular myopathy: refinement of the gene to a 280-kb region with new and highly informative microsatellite markers. *Hum Genet* 1996; **98**: 178–181.