GENOMICS **54,** 256–266 (1998)
ARTICLE NO. GE985560

Genomic Organization of a 225-kb Region in Xq28 Containing the Gene for X-Linked Myotubular Myopathy (*MTM1*) and a Related Gene (*MTMR1*)

Group 1: Petra Kioschis,* Stefan Wiemann,* Nina S. Heiss,* Fiona Francis,†,[1] Carmen Götz,* and Annemarie Poustka*,[2]
Group 2: Stefan Taudien,‡,[3] Matthias Platzer,‡,[3] Thomas Wiehe,‡ Georg Beckmann,§ Jacqueline Weber,‡ Gabriele Nordsiek,‡ and André Rosenthal‡,¶,[2]

**Deutsches Krebsforschungszentrum, Molekulare Genomanalyse, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany; †Max-Planck-Institut für Molekulare Genetik, Ihnestrasse 73, 14193 Berlin, Germany; ‡Institut für Molekulare Biotechnologie, Genomanalyse, Beutenbergstrasse 11, 07745 Jena, Germany; §Max-Delbrück-Centrum für Molekulare Medizin, Bioinformatik, Robert-Rössle-Strasse 10, 13122 Berlin-Buch, Germany; and ¶Friedrich-Schiller-Universität, 07743 Jena, Germany*

Received June 8, 1998; accepted September 9, 1998

MTM1* is responsible for X-linked recessive myotubular myopathy, which is a congenital muscle disorder linked to Xq28. *MTM1* is highly conserved from yeast to humans. A number of related genes also exist. The *MTM1* gene family contains a consensus sequence consisting of the active enzyme site of protein tyrosine phosphatases (PTPs), suggesting that they belong to a new family of PTPs. Database searches revealed homology of myotubularin and all related peptides to the cisplatin resistance-associated α protein, which implicates an as yet unknown function. In addition, homology to the Sbf1 protein (SET binding factor 1), involved in the oncogenic transformation of fibroblasts and differentiation of myoblasts, was also evident. We describe 225 kb of genomic sequence containing *MTM1* and the related gene, *MTMR1,* which lies 20 kb distal to *MTM1.* Although there is only moderate conservation of the exons, the striking similarity in the gene structures indicates that these two genes arose by duplication. Calculations suggest that this event occurred early in evolution long before separation of the human and mouse lineages. So far, mutations have been identified in the coding sequence of only 65% of the patients analyzed, indicating that the remaining mutations may lie in noncoding regions of *MTM1* or possibly in *MTMR1.* Knowledge of the genomic sequence will facilitate mutation analyses of the coding and noncoding sequences of *MTM1* and *MTMR1.

INTRODUCTION

X-linked myotubular myopathy (XLMTM; OMIM 310400) is a congenital muscle disorder linked to Xq28. The disease is characterized by severe hypotonia, generalized muscle weakness, and respiratory failure and usually leads to death within the first months of life (Wallgren-Peterson and Thomas, 1994). Characteristic histopathological features point to an impaired maturation of muscle fibers. The *MTM1* gene is responsible for XLMTM and is highly conserved throughout evolution (Kioschis *et al.,* 1996; Laporte *et al.,* 1996, 1997).

The three related genes *MTMR1, MTMR2,* and *MTMR3* show varying degrees of homology to *MTM1* (Laporte *et al.,* 1996). At the peptide level they all contain the consensus sequence for the active site of protein tyrosine phosphatases (PTPs). This conserved catalytic domain consists of 200–300 residues and contains the PTP consensus sequence (I/V)HCXAGXXR(S/T)G (Denu *et al.,* 1996; Fischer *et al.,* 1991). The homology of myotubularin and the related genes to other PTPs is restricted to the catalytic signature of PTPs and has led to the assumption that they represent a new family of PTPs. PTPs are divided into two classes, the transmembrane receptor-like phosphatases and the nonreceptor cytoplasmic phosphatases. The receptor-like PTPs possess an extracellular domain, a transmembrane segment, and one or two intracellular catalytic domains, a structure which is typical of transmembrane receptors. The intracellular PTPs form part of the receptor signaling cascade in a more indirect sense by associating with specific growth factor receptors (Dixon, 1996; Fauman and Saper, 1996; Jia, 1997). Since none of the receptor-like domains are present in myotubularin, it is probably an intracellular PTP.

Mutation analyses of *MTM1* have revealed a high degree of heterogeneity in the type and distribution of mutations, although a clustering of mutations does appear to occur in exons 4, 8, 9, 11, and 12. So far, mutation analyses have been restricted to the coding

[1] Present address: CHU Cochin Port-Royal, Institut Cochin de Genetique Moleculaire, INSERM U.129, Paris, France.
[2] To whom correspondence should be addressed.
[3] S.T. and M.P. contributed equally to this work.

0888-7543/98 $25.00

sequence of *MTM1*, which may explain why mutations have been detected in only 65% of the patients (de Gouyon *et al.*, 1997; Laporte *et al.*, 1997). Recently, an intronic mutation in *MTM1* was identified (Tanner *et al.*, 1998). It is thus conceivable that some of the mutations are located in noncoding regions of *MTM1*. Furthermore, it cannot be excluded that mutations in the related *MTMR1* gene, which lies distal to *MTM1*, may also cause XLMTM.

In this paper we describe 225 kb of genomic sequence that contains the *MTM1* and *MTMR1* genes, although a gap remains in the first intron of *MTM1*. A comparison of the cDNAs and of the gene structures of *MTM1* and *MTMR1* shows that these two genes arose by intrachromosomal duplication. The genomic sequence will be very useful for the identification of additional mutations in myotubular myopathy patients.

MATERIALS AND METHODS

Libraries. Cosmids indicated by "Qc" were isolated from a Xq28-specific cosmid library constructed from the hamster/human cell hybrid QIZ (Kioschis *et al.*, 1996; Warren *et al.*, 1990). Clones isolated from the Lawrence Livermore National Laboratory chromosome X cosmid library (LLOXNCO1) are indicated with "L". A pool of cosmids binned to Xq28-specific YACs was kindly provided by D. L. Nelson (Baylor College of Medicine, Institute for Molecular Genetics, Houston, TX). Plasmid clones were isolated from a human λPS genomic library (Mobitec; PS01; Nehls *et al.*, 1994). Plating and excision rescue into plasmids were performed according to the manufacturer's instructions.

cDNA clones were obtained from the following libraries: (i) cDNA selection library from the *MTM1* region (Kioschis *et al.*, 1996), (ii) fetal brain (DKFZphfbr2 oligo(dT), in pAMP1, Library No. 564; Stefan Wiemann, Resource Center, http://www.rzpd.de), (iii) human placenta (Stratagene 936203, oligo(dT)), human fetal brain (Stratagene 937226, oligo(dT)), and (iv) human fetal heart (Clontech HL3018b, oligo(dT) and random).

DNA labeling and hybridization. PCR products and cDNA inserts were gel-purified using the Qiagen gel extraction kit and labeled by random hexamer priming (Feinberg and Vogelstein, 1983). Hybridizations were performed for 16 h in 7% SDS, 0.5 M sodium phosphate, and 1 mM EDTA at 65°C (Church and Gilbert, 1984) and washed twice at 65°C in 40 mM sodium phosphate, pH 7.2, 1% SDS for 30 min each time. Filters were exposed overnight to X-ray film (XAR-5; Kodak).

Sequencing. cDNAs were sequenced either by primer walking or by preparing deletion clones. The cosmids and λPS DNA were sequenced as described previously (Craxton, 1993) with several modifications. M13 templates were prepared by the Triton method (Mardis, 1994) and sequenced using Thermo Sequenase (Amersham) and dye terminator chemistry (Perkin–Elmer). Data were collected using the ABI377 automated sequencers, and sequences were assembled by using the GAP program (Dear and Staden, 1991). To obtain a genomic sequence with the highest possible reliability the final consensus sequence was obtained from more than one clone wherever possible. The actual number of sequencing reads depended on the degree and number of overlapping clones. Data from overlapping clones were assembled individually and joined subsequently. Gaps were closed by resequencing the M13 templates with ET dye primers (Amersham) and/or using custom-made primers on M13 templates, PCR products, or cosmid/λPS DNA in combination with dye terminators. Standard PCR conditions for amplification of selected regions of genomic DNA were as follows: 1 min at 94°C, 30 cycles (30 s at 94°C, 1 min at 55°C, 2 min at 72°C), and 4 min at 72°C. PCR products appearing as a distinct band on agarose gels were purified by PEG precipitation (Rosenthal *et al.*, 1993). Otherwise, the appropriate bands were cut out of the agarose gel and purified using the Qiaex Kit (Qiagen). Finishing was performed to achieve a sequencing precision of at least 99.99% (Platzer *et al.*, 1997).

Nested RT-PCR. For the amplification of *MTMR1* exons, cDNAs of different tissues (brain, heart, kidney, liver, lung, skeletal muscle, pancreas, and placenta; Clontech) were used. The primers were A, 5′-TTCCTTCAGCATTCGAG-3′; B, 5′-GTCTCATCCGTGGATTC-3′; C, 5′-GGATCACCTTTATAGCTGTC-3′; D, 5′-TGACTCAGACTAGCAAC-3′; E, 5′-GCGGAGACTACATGAAC-3′; F, 5′-CATCCTCTGTGGTGTTG-3′; G, 5′-CGGTTTTGG AGTTGGTG-3′; and H, 5′-GGAAATGGAGTCTGCTG-3′. PCR was performed under the following cycle conditions: 60 s at 94°C and 30 cycles of 30 s at 50°C, 120 s at 72°C, 30 s at 94°C followed by 10 min at 72°C. For the second round of nested/seminested PCR, 1 μl of a 1:5 dilution of the first amplification was used. Several primer combinations were used: (i) primary PCR C/B, secondary PCR C/D; (ii) first A/G; second C/E; and (iii) first F/B, second H/D.

Computer analysis. Analysis and assembly of cDNA sequences were carried out using CLUSTAL V (Higgins *et al.*, 1992), BESTFIT (Devereux *et al.*, 1984), and the HUSAR 4.0 package (DKFZ Heidelberg). Genomic sequences were analyzed as described by Platzer *et al.* (1997). Furthermore, the programs GCG DIVERGE (Li *et al.*, 1985), GCG DOTPLOT (Maizel and Lenk, 1981), SIM (Huang *et al.*, 1990), and PSI-BLAST (Altschul *et al.*, 1997) were used for sequence analysis.

RESULTS

Genomic Mapping, Sequencing, and Analysis of the MTM1 and MTMR1 Loci

The original cosmid map (Kioschis *et al.*, 1996) entirely spanned *MTMR1* but left a gap in *MTM1* (Fig. 1). To close the gap, PCR probes derived from the flanking genomic sequence were used to screen a human PAC library (7×10^4 clones), an X-chromosome-specific cosmid library (6×10^4 clones), a pool of Xq28-specific cosmids binned to YACs, and a total human genomic λPS library (3×10^6). This resulted in the isolation of one additional cosmid (LP162H3) and eight plasmid clones (λPS1 to λPS8) with inserts ranging from 10 to 20 kb (Fig. 1). Nevertheless, a gap that is located in intron 1 of *MTM1* remains between clones Qc4F1 and λPS6. From the available sequence 17 kb of this intron was determined. However, it is difficult to estimate the size of the entire intron because the end probes from clones Qc4F1 and λPS6 do not recognize the same digestion fragment of overlapping YAC clones (Rogner *et al.*, 1994) on hybridization.

In total, the 10 cosmids and 8 λPS clones span 225 kb of genomic sequence (Fig. 1; Accession Nos. Y15994 and AF002223). Two cosmid clones (Qc3F12, Qc12E11) were found to contain chimeric segments at one end and therefore this sequence was disregarded (Fig. 1).

Of the 225 kb, 75% is covered by at least two independent clones. About 87 kb are represented by clones derived from the same library (λPS6, λPS7, λPS8 15 kb; Qc8D11 to Qc12E11 55 kb; LP0142, LD0132 17 kb). These clones differ by only three mono- and one dinucleotide repeat stretches. In contrast, the 82 kb covered by clones from different libraries (λPS1-5/L162H3, Qc8D11 33 kb; Qc12E11 to LJ2236 27 kb; LJ2236/

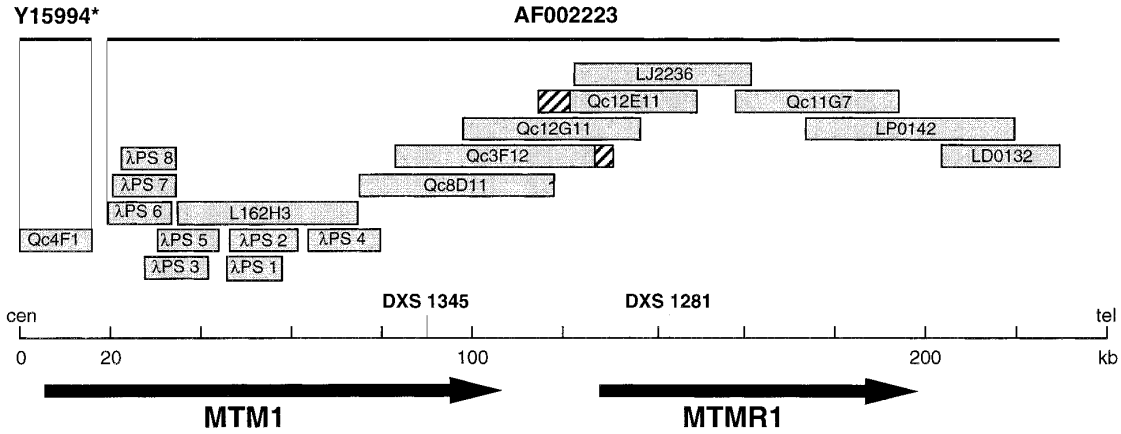


FIG. 1. Map of the sequenced cosmids and λPS clones in the *MTM1/MTMR1* region of human Xq28. Horizontal arrows indicate orientation and size of *MTM1* and *MTMR1*. Hatched bars indicate rearranged clone parts. Accession numbers of the sequence segments are given at the top. The marker DXS1345, used for mapping the region, is located in intron 13 of *MTM1* (nucleotides 77,574–77,753 in AF002223). The marker DXS1281 is located in intron 2 of *MTMR1* (nucleotides 127,944–128,103 in AF002223). Asterisk marks positions 32,301–47,300.

Qc11G7 3 kb; Qc11G7/LP0142 19 kb) contained 53 sequence differences that were not the result of sequencing errors. Among these, 27 (51%) are transitions, 9 (17%) are transversions, and 16 (30%) are insertions/deletions in mononucleotide or dinucleotide repeat stretches. In addition, a 20-bp insertion/deletion at nucleotide 43,103 was detected. Although it cannot be excluded that these differences are due to cloning artifacts, they probably represent naturally occurring allelic sequence variations.

The entire sequence was screened for the presence of repeats. Sixteen different microsatellites were identified (Table 1). Three of these were shown to represent different haplotypes because of variation between different clones. Further analysis of repetitive sequences showed that genome-wide repeats constitute 34.1% of the *MTM1* and *MTMR1* locus with LINE (18.7%) and SINE (8.6%) sequences being the major contributors (Table 2).

Predictions carried out to identify putative promoters and transcription factor binding sites resulted in high-score predictions of a TATA-less promoter in the 5′ parts of the *MTM1* and *MTMR1* genes and the identification of several regulatory elements, including SP1 binding sites (GC-box) and E box elements (Lassar and Munsterberg, 1994; Lassar *et al.*, 1991) in both the *MTM1* and the *MTMR1* genes.

Compositional analysis of the 225-kb sequence revealed a G+C content of 42.5%. Using a 4-kb window in steps of 1 kb, two definite G+C peaks containing the CpG islands and exons 1 of the *MTM1* and *MTMR1* genes were observed (Fig. 2A). In the distal end of *MTMR1*, the G+C content rises to above 55%. By removing regions of GC-rich genome-wide repeats and repeating the analysis, possible distortion of the results by such elements was excluded. According to these analyses, *MTM1* has a uniform G+C distribution of 38.5 ± 2.3%, whereas that of *MTMR1* is nonuniform, ranging from 37.0 to 58.2% with an average of 46.2 ± 7.0%. Overall, the G+C content of the *MTM1* and *MTMR1* coding regions shows a difference of only 3.8% (Fig. 2B).

On screening the genomic sequence against the EST database, 46 and 27 ESTs representing the *MTM1* and *MTMR1* genes, respectively, were identified. In addition, a cluster of 81 ESTs overlapped with the previously identified cDNA selection clone XAP89 (Kioschis *et al.*, 1996), indicating that there is a third gene in this region. The ESTs form a consensus sequence of 2167 bp that is colinear with the genomic sequence. The absence of an ORF suggests that the sequence represents the 3′ UTR of this gene. Furthermore, 10 of the ESTs carry polyadenylation tails that lie 1411 bp distal of the

TABLE 1

Distribution of Di-, Tri-, and Tetranucleotide Repeats in the 225-kb Genomic Sequence

Sequence	Copies	Position	Localization	Rel. pos.
CA	19	34529[b]	5′ of *MTM1* exon 1	−2612
TG	6	8666	*MTM1* intron 2	+1408
AG	7	10144	*MTM1* intron 2	−936
TA	7*	61776	*MTM1* intron 9	+1324
TA	6	67172	*MTM1* intron 10	+2691
TA	5	67228	*MTM1* intron 10	+2747
CA	18	67240	*MTM1* intron 10	+2759
CA	13[a]	128497	*MTMR1* intron 2	−4707
TA	5	128529	*MTMR1* intron 2	−4675
CA	8	132652	*MTMR1* intron 2	−552
CA	14	143799	*MTMR1* intron 5	−889
GAT	7[a]	8539	*MTM1* intron 2	+1281
CAA	5	100153	5′ of *MTMR1* exon 1	−7967
CAT	7	116180	*MTMR1* intron 2	+2300
TAAG	7	179697	3′ of *MTMR1* exon 16	+179
ATTG	6	179747	3′ *MTMR1* exon 16	+247

[a] Repeats for which the copy number was different between different haplotypes; the numbers in the third column refer to positions in sequence AF002223 with the exception of *b*, which refers to the position in sequence Y15994.

TABLE 2

Distribution of Genome-wide Repeats in the 225-kb Genomic Sequence

Type	Copies	Fraction of locus (%)
SINE	80	8.60
LINE	85	18.71
LTR	29	4.53
DNA elements	24	1.98
Unclassified	3	0.30
Total	221	34.12

MTMR1 polyadenylation site, indicating that the novel gene is transcribed in the opposite orientation compared to *MTM1* and *MTMR1*.

Analysis of the MTM1 and MTMR1 cDNAs

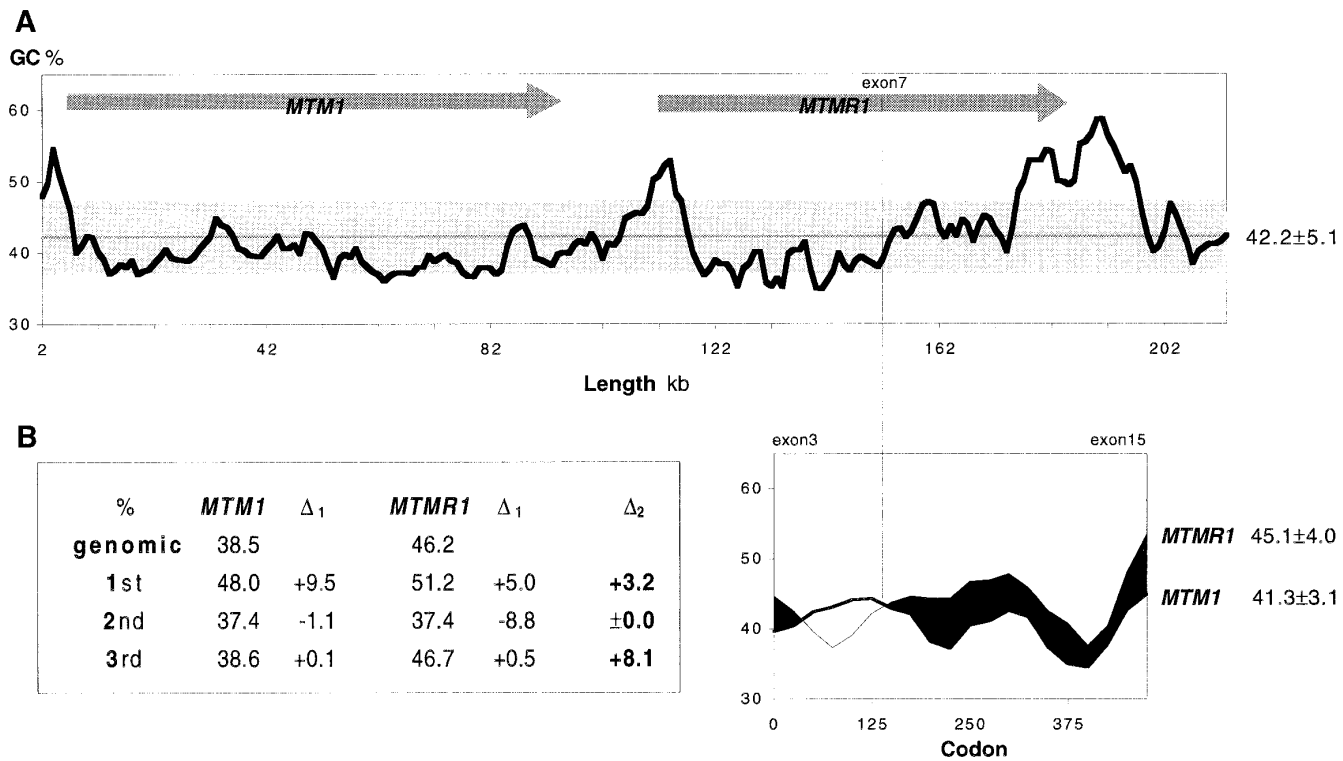
The previously described *MTM1* cDNA sequence has a length of 3.4 kb and is close to complete (U46024 in Laporte *et al.*, 1996). An unambiguous ATG start codon could not be identified, however, because the ORF continues beyond the 5′-most methionine. On screening of a human fetal brain cDNA library enriched for full-length cDNAs (DKFZphfbr2 oligo(dT), Library No. 564) a clone that extends the 5′ end of the original sequence by only 13 bp was isolated.

Several human cDNA libraries were screened with previously isolated partial cDNA clones (XAP86, XAP90, Kioschis *et al.*, 1996) to obtain the full-length *MTMR1* cDNA. The sequences of nine partial cDNA clones overlapped to yield 2581 bp of sequence (Accession No. AJ224979). One cDNA clone contained an insertion of 211 bp at nucleotide 1647 (equivalent to the 3′ end of exon 13) and suggested that *MTMR1* undergoes alternative splicing. We have designated this transcript *MTMR1a* (Figs. 3 and 4).

Comparison of the Genomic Structure of MTM1 and MTMR1

Alignment of the cDNA sequences with the genomic sequence showed that *MTM1* spans more than 98 kb while *MTMR1* spans 71 kb. *MTMR1* lies 20,399 bp distal to *MTM1* and is transcribed in the same direction.

The sizes of exons 4 and 6 through to exon 14 are identical in both genes (Table 3). An exon by exon comparison of *MTM1* and *MTMR1* identified different degrees of homology. Exons 11 to 13 are highly homologous (nucleotide 71–80%, amino acid 74–77%), exons 5–10 and 14 show medium homology (64–68%, 55–67%), and exons 3–4 and the coding part of exon 15 show low homology (40–58%, 36–41%). Exons 1 and 2 and the non-



%	*MTM1*	Δ_1	*MTMR1*	Δ_1	Δ_2
genomic	38.5		46.2		
1st	48.0	+9.5	51.2	+5.0	**+3.2**
2nd	37.4	-1.1	37.4	-8.8	**±0.0**
3rd	38.6	+0.1	46.7	+0.5	**+8.1**

FIG. 2. Compositional analysis of the human *MTM1* locus. (**A**) G+C content was determined along the 225 kb sequence, although only 215 kb of the sequence starting at the CpG island of *MTM1* are shown here. Gray lines and shaded areas show the average values ± standard deviations. Distributions were calculated with a moving window of 4 kb, step 1 kb. (**B**) Composition of the coding regions of *MTM1* and *MTMR1* along the homologous exons 3 to 15. Black areas of the graph indicate regions of G+C content higher in *MTMR1* than in *MTM1*. Distributions were calculated with a moving window of 50 kb, step 25 kb. The table compares average G+C values of the respective local genomic regions and of the first, second, and third codon positions. Δ_1, difference between codon position and local genomic environment; Δ_2, difference between codon positions of *MTMR1* and *MTM1*.

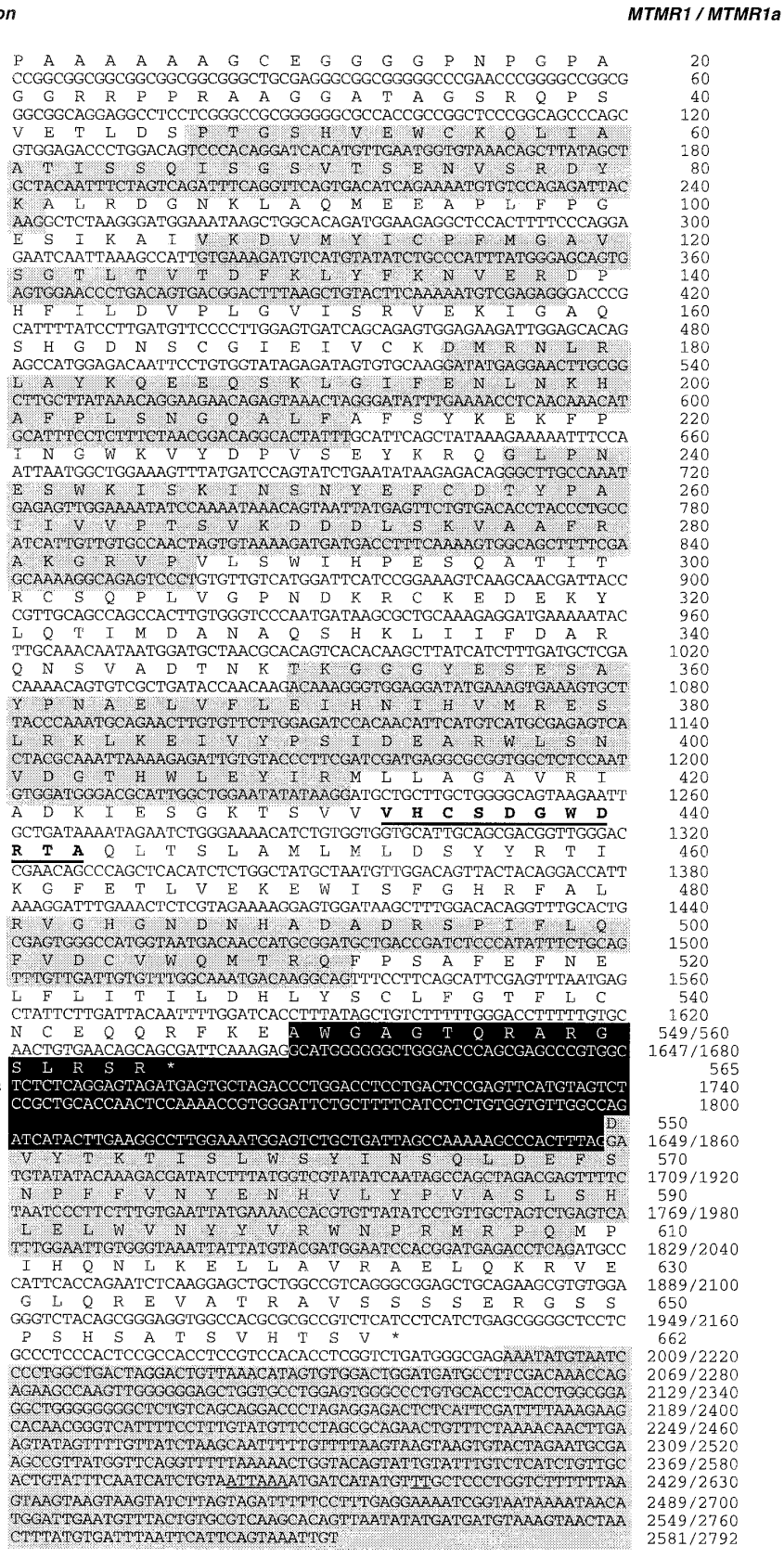


FIG. 3. Human *MTMR1* and *MTMR1a* cDNAs with the predicted amino acid sequences. The sizes of exons and exon–exon boundaries are shown by alternating light gray shading; black shading shows the alternative *MTMR1a* exon; the PTP signature is shown bold and underlined; the polyadenylation signal and putative polyadenylation site are underlined.

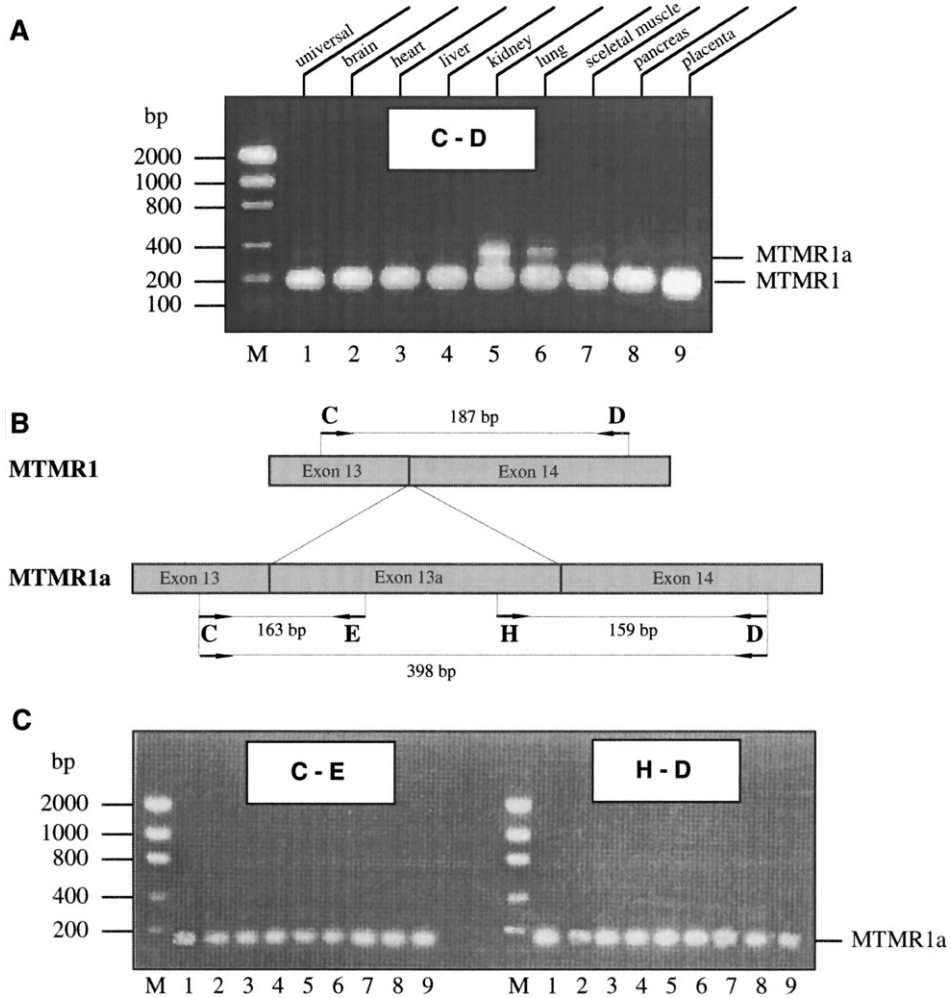


FIG. 4. Schematic representation of the RT-PCR approach (**B**) providing evidence of expression of the two alternatively spliced transcripts *MTMR1* (**A**) when using primer pair C and D (C–D) and *MTMR1a* (**C**) when using primer pairs C and E (C–E) and H and D (H–D) in all eight tissues examined. **M,** marker lane. All PCR products were sequenced and confirmed the boundaries of exons 13/14, 13/13a, and 13a/14. A stop codon in exon 13a indicates that *MTMR1a* encodes a truncated peptide with the ORF terminating 48 bp downstream of the alternative splice site.

coding part of exon 15 show no homology. Furthermore, the structures of the 5′ and 3′ untranslated regions of *MTM1* and *MTMR1* are different (Fig. 5).

Although all intron phases are identical in both genes, extensive searches using computer programs LFASTA, DOTPLOT, and SIM did not reveal significant homologies of introns or 5′ and 3′ flanking sequences between the two genes.

Evolutionary Analysis of the MTM1 and MTMR1 Loci

Similarities in the gene structure of *MTM1* and *MTMR1* suggest that they arose by the duplication of a common ancestral gene. To date the duplication event, the coding sequences of the human *MTM1* and *MTMR1* genes were compared to the mouse homologue *Mtm1* (P. Kioschis, Heidelberg, pers. comm., June 1998.) using the program DIVERGE. The number of substitutions per basepair was computed for synonymous (s) and nonsynonymous (n) sites for all possible pairwise comparisons (*MTM1* with *Mtm1* (K_1(s)=0.401, K_1(n)=0.047), *MTM1* with *MTMR1* (K_2(s)=2.159, K_2(n)=0.330), and *Mtm1* with *MTMR1* (K_3(s)=2.216, K_3(n)=0.342). Assuming a divergence time of primates and rodents of about 75 Myr B.P. (Li *et al.*, 1987), the time of the duplication event was calculated separately for synonymous and nonsynonymous sites. Values of T_D(s)=409 Myr and T_D(n)=537 Myr were calculated for synonymous and nonsynonymous sites, respectively. On taking the average as a final estimate for the duplication time (Li and Graur, 1991), the gene duplication can be dated to 470 Myr B.P., which is long before divergence of the primate and rodent lineages.

Sequence Comparison of Myotubularin to Related Proteins

For the human *MTM1* and related cDNA sequences, *MTM1, MTMR1* (this report), and *MTMR3* are full-length (Accession No. AB002369; Nagase *et al.*, 1997). Database searches identified full-length homologous

TABLE 3

Exon–Intron Organization of the *MTM1* and *MTMR1* Genes

	Exon size (bp)		Intron size (bp)		
Number	*MTM1*	*MTMR1*	*MTM1*	*MTMR1*	Phase
1	57	137	>17289	5518	
2	10, **63**	106	3821	19323	0
3	**73**	10, **66**[a]	2021	8520	1
4	**95**	**95**	15902	374	0
5	**111**	**108**	4336	2317	0
6	**102**	**102**	19803	314	0
7	**84**	**84**	2242	861	0
8	**150**	**150**	4264	922	0
9	**189**	**189**	3844	3864	0
10	**186**	**186**	7919	461	0
11	**207**	**207**	1636	6912	0
12	**93**	**93**	610	6276	0
13	**114**	**114**	2948	186	0
13a		**48**		4448	
		163			
14	**177**	**177**	7818	6696	0
15	**165**	**162**		1774	
	1549	11			
16		413			

Note. Exons sizes shown in bold are coding regions. Due to the clone gap, the complete sequence of intron 1 of *MTM1* was not determined but is estimated to be larger than 17,289 bp.

[a] Translation start site unclear.

cDNAs from other species; two are derived from *Caenorhabditis elegans* (CeT24A11.1, Accession No. Z49072; CeF53A2.f, Accession No. Z81546), one from *Saccharomyces cerevisiae* (ScYJR110w, Accession No. Z49610), and one from *Schizosaccharomyces pombe* (SpPPase, Accession No. Z98974). The entire peptide sequences were predicted for these seven full-length cDNAs, and the conserved sequence motifs were compared (Fig. 6). The amino-terminal 150 amino acid residues show low sequence similarity, although two short regions of high similarity are evident. A high degree of conservation is restricted to a region of about 350 amino acid residues that contains the active site of PTPs. The carboxy-terminal ends of all peptides vary considerably in length and do not exhibit significant homology. The *MTMR3* and *C. elegans* peptides possess a FYVE-type zinc finger domain at the C-terminus, which is not present in the other peptides and has been implicated to play a role in endosomal targeting (Stenmark *et al.*, 1996).

Using automated database searches and construction of consensus strings (Beckmann *et al.*, 1998), a low degree of homology of myotubularin and related proteins to the human cisplatin –resistance-associated α protein was detected (CRAα; Accession No. U78556). As a cross-reference test, CRAα was subjected to a PSI-BLAST search and again myotubularin and related peptides were detected. The *P* values of $P < 10^{-10}$ were significant and the *P* value of the next best sequence was close to 1. Identity of CRAα to myotubularin and related proteins is restricted to the N-terminal part and does not include the PTP active site (Fig. 6). Moreover, the 3′ UTR of the CRAα peptide shows homology to myotubularin and related peptides, suggesting that the predicted CRAα peptide may be truncated as the result of sequencing errors. By a search for protein motifs using BLASTP and BEAUTY programs, significant homologies to the Sbf1 protein (SET binding factor 1; Accession No. U93181; Cui *et al.*, 1998) were identified ($P < 10^{-38}$).

DISCUSSION

The *MTM1* gene is responsible for X-linked myotubular myopathy and is highly conserved in many species (Laporte *et al.*, 1996). At least three *MTM1*-related genes exist in humans, indicating that a novel gene family has been identified. Construction of the transcript map surrounding the *MTM1* gene previously resulted in the isolation of one related gene, *MTMR1*, which maps immediately distal to *MTM1* (Kioschis *et al.*, 1996; Laporte *et al.*, 1996). In the present report we provide a detailed analysis of 225 kb of genomic sequence containing the *MTM1* and *MTMR1* genes. The genomic sequence of the *MTM1*/*MTMR1* region has

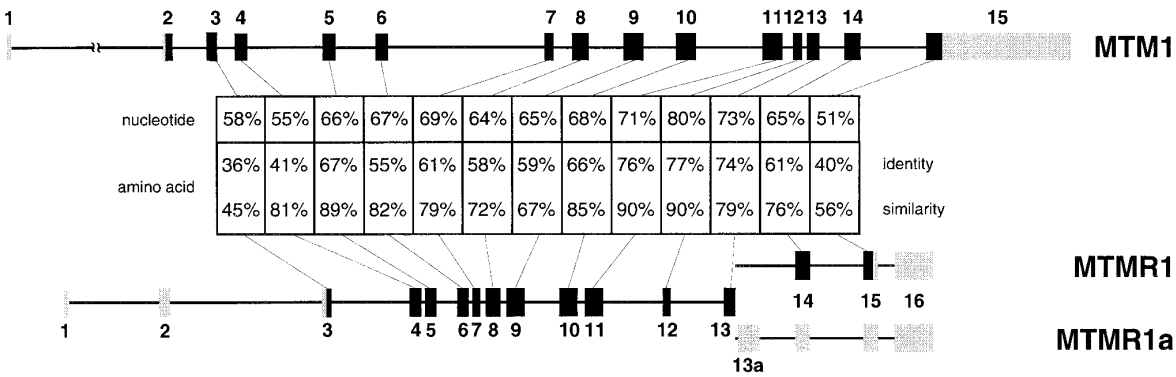


FIG. 5. Genomic organization of the human *MTM1* and *MTMR1* genes. The start codon of *MTM1* is situated in exon 2 (Laporte *et al.*, 1997), whereas in *MTMR1* the first potential start codon lies in exon 3. One large last exon, exon 15, forms the 3′ UTR of *MTM1*. In comparison, the 3′ UTR of *MTMR1* is interrupted by an intron. Alternative splicing of exon 13a of *MTMR1* generates transcript *MTMR1a*. Exons are represented by boxes, introns by lines. Black boxes, translated; gray boxes, untranslated; scale exon:intron = 10:1.

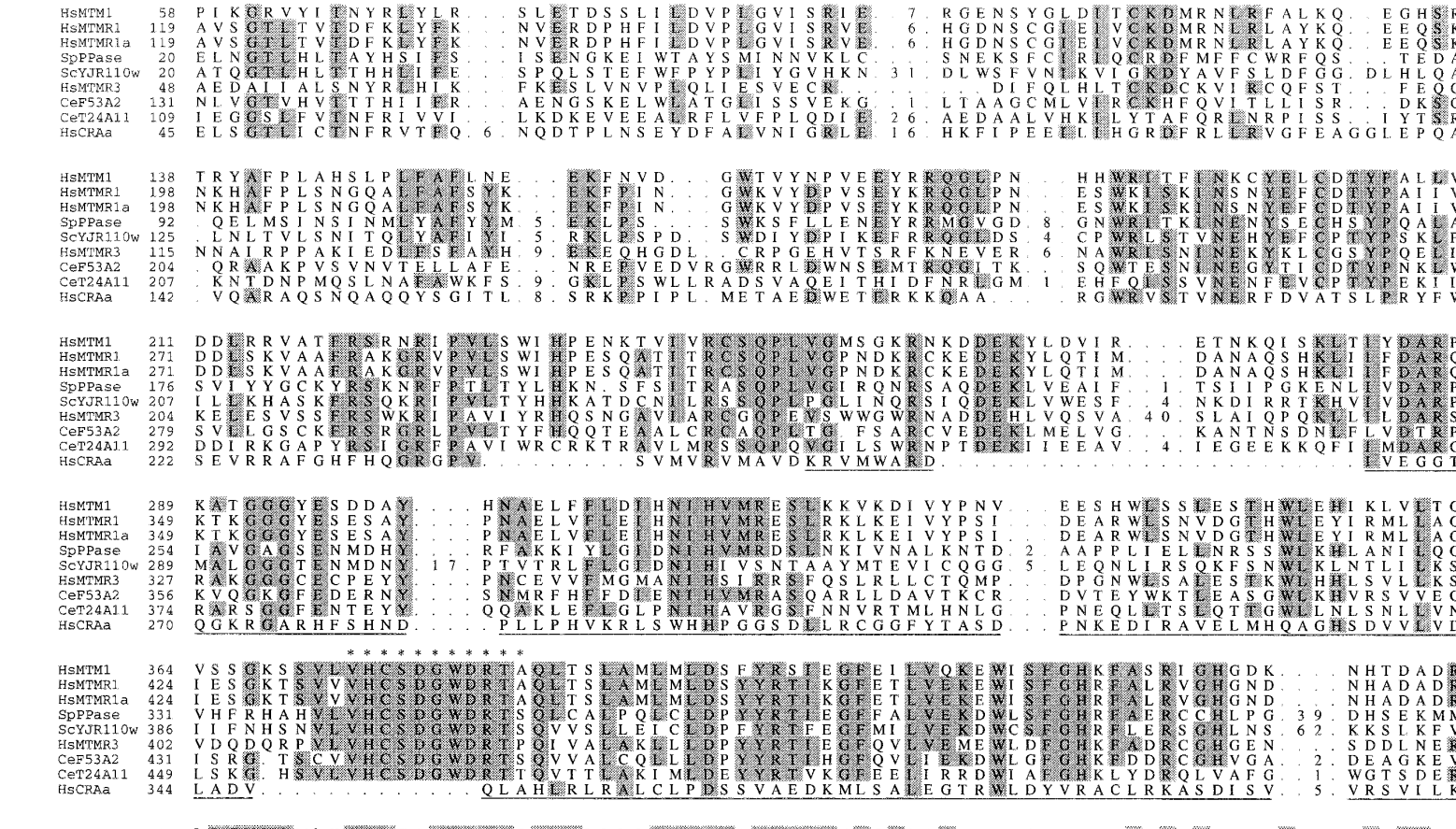


FIG. 6. Multiple sequence alignment of myotubularin with related peptides and the cisplatin resistance associated protein α. Hs, *Homo sapiens;* Sc, *S. cerevisiae;* Sp, *S. pombe;* Ce, *C. elegans;* shaded regions show identical amino acids in five or more of the seven sequences; similar but conserved amino acids are not highlighted; asterisks above the sequence block mark the PTP signature; numbers within gaps indicate stretches of no homology. The protein sequence CRAa in the alignment differs from the conceptual translation of the database entry U78556 by four changes (2 substitutions, 1 deletion, 1 insertion), leading to an extension of the ORF which is shown underlined.

also allowed a regional comparison of structural differences in Xq28 at a high resolution.

In agreement with the variation in GC richness and regional clustering of CpG islands in Xq28 (Pilia *et al.*, 1993; Palmieri *et al.*, 1994; De Sario *et al.*, 1996), construction of transcript maps have confirmed regional differences in gene density (Sedlacek *et al.*, 1993; Kioschis *et al.*, 1996; Heiss *et al.*, 1996; Rogner *et al.*, 1996). The 225-kb interval described here contains three genes with an average gene density of 1 gene in 75 kb. By contrast, the 220-kb region between the color vision genes (RCP/GCP) and G6PD contains 19 genes with an average density of 1 gene in 11.6 kb (Chen *et al.*, 1996). Compared to the average gene size of 7 kb in the G6PD region, the *MTM1* (>98 kb) and *MTMR1* (71 kb) genes are considerably larger.

The *MTM1/MTMR1* region also differs from the G6PD region in the composition of repetitive elements. The predominant type of repeats in the *MTM1/MTMR1* region belong to the long GC-poor LINE elements totaling 18.7% within 225 kb, whereas the short GC-rich SINEs contribute only 8.6%. In the G6PD region the GC-rich SINEs are far more abundant, reaching 20.4% in 220 kb.

The sequence homology between the *MTM1* and the *MTMR1* cDNAs and the similar gene structures show that these genes are related. The difference in intron length and the complete absence of homologies of introns and flanking sequences suggest that the ancestor of *MTM1* or *MTMR1* underwent intrachromosomal duplication followed by sequence divergence. As both genes are expressed, this mechanism points to a gain of function of one of the genes. Computer-based evolutionary analysis of *MTM1* and *MTMR1* together with the murine orthologue suggest that duplication of the ancestral gene occurred long before separation of the human and mouse lineages. Since the two genes map within 20 kb of each other, unequal crossing over might have been the mechanism of intrachromosomal duplication.

The configuration of the *MTM1* and *MTMR1* genes is an example of an intrachromosomal tandem gene duplication in Xq28. Other examples of intrachromosomal gene duplications in Xq28 include the IDS gene (Timms *et al.*, 1995) and the cluster of melanoma antigen genes (Rogner *et al.*, 1995; Muscatelli *et al.*, 1995; Heiss *et al.*, 1996). In addition, the adrenoleukodystrophy gene and a gene-rich cluster containing the human creatine transporter gene (SLC6A8) and the CDM gene (Eichler *et al.*, 1996, 1997) in Xq28 show interchromosomal duplications. By analogy, the *MTMR2* gene does not map to Xq28, and it will be interesting to determine its evolutionary relation to *MTM1* and *MTMR1*. In contrast, *MTMR3* (Accession Nos. AC003681, AC003071) probably did not arise by interchromosomal duplication as it shows a different gene structure compared to *MTM1* and *MTMR1*. Once specific functional roles have been associated with individual genes of the *MTM1* gene family, this should shed light on the evolution of the *MTM1* gene family.

The myotubularin family of proteins is highly conserved over a wide phylogenetic spectrum ranging from yeast to human. The most conserved domain common to all myotubularin peptides contains the consensus sequence for the active site of PTPs. There are no additional homologies to PTPs, suggesting that the myotubularins form a new family of PTPs (Laporte *et al.*, 1996), although tyrosine phosphatase activity still requires experimental evidence. The C-termini of the *C. elegans* and *MTMR3* peptide sequences contain a FYVE-type zinc finger domain (Stenmark *et al.*, 1996), which is absent in the other myotubularin proteins. This indicates that *MTMR3* belongs to an additional subfamily.

The classical PTPs contain other regulatory modules such as a fibronectin-type 3 domain, immunoglobulin-type domains, and src-2 domains. Although the myotubularins do not contain such domains, similarity to the CRAα is reminiscent of a regulatory module. This observation may help clarify the role myotubularin plays in the etiology of myotubular myopathy and could be an indication for a broader function of the myotubularins.

The myotubularins also show homology to the Sbf1 protein (SET binding factor 1), which is involved in the oncogenic transformation of fibroblasts and in the *in vitro* differentiation of myoblasts (Cui *et al.*, 1998). In addition, the promoter of *MTM1* was found to contain myogenic basic–loop–helix (bHLH) E protein heterodimer binding sites (E boxes). Such elements have been shown to be of functional significance for the specific expression of genes in skeletal muscle tissue (Lassar and Munsterberg, 1994), and some of these myogenic bHLH regulators initiate skeletal muscle differentiation (Lassar *et al.*, 1989; Tapscott *et al.*, 1992; Weintraub *et al.*, 1991). These findings support a role of myotubularin-type phosphatases in proliferation and differentiation pathways (Cui *et al.*, 1998).

Although there is no doubt that *MTM1* is involved in XLMTM, mutations were detected in only 65% of the patients (de Gouyon *et al.*, 1997; Laporte *et al.*, 1997). Mutation analyses were restricted to the coding sequence of *MTM1,* and this suggests that additional mutations may reside in the introns or in promoter sequences. It is also conceivable that mutations in *MTMR1* may cause XLMTM. Knowledge of the complete genomic structures of both genes will now permit mutation analyses of the coding and noncoding sequences in *MTM1* and *MTMR1* in XLMTM patients and in carriers.

ACKNOWLEDGMENTS

We thank Christiane Rutenberg, Christine Happes, Hella Ludewig, Evelyn Michaelis, and Diana Wiedemann for the excellent technical assistance. We also thank D. L. Nelson (Baylor College of Medicine, Institute for Molecular Genetics, Houston, TX) for provid-

ing cosmid clones. The work was supported by grants from German BMBF-BEO 0311108, Thuringian B0301-96121, and EU-BIOMED2 program BMH4-CT96-0338.

REFERENCES

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.

Beckmann, G., Hanke, J., Bork, P., and Reich, J. G. (1998). Merging extracellular domains: Fold prediction for laminin G-like and amino-terminal thrombospondin-like modules based on homology to pentraxins. *J. Mol. Biol.* **275:** 725–730.

Chen, E. Y., Zollo, M., Mazzarella, R., Ciccodicola, A., Chen, C., Zuo, L., Heiner, C., Burough, F., Repetto, M., Schlessinger, D., and D'Urso, M. (1996). Long-range sequence analysis in Xq28: Thirteen known and six candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and G6PD loci. *Hum. Mol. Genet.* **5:** 659–668.

Church, G. M., and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81:** 1991–1995.

Craxton, M. (1993). Cosmid sequencing. *In* "DNA Sequencing Protocols" (H. G. Griffin and A. M. Griffin, Eds.), pp. 149–167, Humana Press, Totowa, NJ.

Cui, X., De Vivo, I., Slany, R., Miyamoto, A., Firestein, R., and Cleary, M. L. (1998). Association of SET domain and myotubularin-related proteins modulates growth control. *Nat. Genet.* **18:** 331–337.

de Gouyon, B. M., Zhao, W., Laporte, J., Mandel, J. L., Metzenberg, A., and Herman, G. E. (1997). Characterization of mutations in the myotubularin gene in twenty six patients with X-linked myotubular myopathy. *Hum. Mol. Genet.* **6:** 1499–1504.

De Sario, A., Geigl, E. M., Palmieri, G., D'Urso, M., and Bernardi, G. (1996). A compositional map of human chromosome band Xq28. *Proc. Natl. Acad. Sci. USA* **93:** 1298–1302.

Dear, S., and Staden, R. (1991). A sequence assembly and editing program for efficient management of large projects. *Nucleic Acids Res.* **19:** 3907–3911.

Denu, J. M., Stuckey, J. A., Saper, M. A., and Dixon, J. E. (1996). Form and function in protein dephosphorylation. *Cell* **87:** 361–364.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12:** 387–395.

Dixon, J. E. (1996). Protein tyrosine phosphatases: Their roles in signal transduction. *Recent Prog. Horm. Res.* **51:** 405–414.

Eichler, E. E., Lu, F., Shen, Y., Antonacci, R., Jurecic, V., Doggett, N. A., Moyzis, R. K., Baldini, A., Gibbs, R. A., and Nelson, D. L. (1996). Duplication of a gene-rich cluster between 16p11.1 and Xq28: A novel pericentromeric-directed mechanism for paralogous genome evolution. *Hum. Mol. Genet.* **5:** 899–912.

Eichler, E. E., Budarf, M. L., Rocchi, M., Deaven, L. L., Doggett, N. A., Baldini, A., Nelson, D. L., and Mohrenweiser, H. W. (1997). Interchromosomal duplications of the adreno-leukodystrophy locus: A phenomenon of pericentromeric plasticity. *Hum. Mol. Genet.* **6:** 991–1002.

Fauman, E. B., and Saper, M. A. (1996). Structure and function of the protein tyrosine phosphatases. *Trends Biochem. Sci.* **21:** 413–417.

Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132:** 6–13.

Fischer, E. H., Charbonneau, H., and Tonks, N. K. (1991). Protein tyrosine phosphatases: A diverse family of intracellular and transmembrane enzymes. *Science* **253:** 401–406.

Heiss, N. S., Rogner, U. C., Kioschis, P., Korn, B., and Poustka, A. (1996). Transcriptional mapping in a 700 kb region around the DXS52 locus in Xq28: Isolation of six novel transcripts and a novel ATPase isoform (hPM CA5). *Genome Res.* **6:** 478–491.

Higgins, D. G., Bleasby, A. J., and Fuchs, R. (1992). CLUSTAL V: Improved software for multiple sequence alignment. *Comput. Appl. Biosci.* **8:** 189–191.

Huang, X. Q., Hardison, R. C., and Miller, W. (1990). A space-efficient algorithm for local similarities. *Comput. Appl. Biosci.* **6:** 373–381.

Jia, Z. (1997). Protein phosphatases: Structures and implications. *Biochem. Cell Biol.* **75:** 17– 26.

Kioschis, P., Rogner, U. C., Pick, E., Klauck, S. M., Heiss, N., Siebenhaar, R., Korn, B., Coy, J. F., Laporte, J., Liechti-Gallati, S., and Poustka, A. (1996). A 900-kb cosmid contig and 10 new transcripts within the candidate region for myotubular myopathy (MTM1). *Genomics* **33:** 365–373.

Laporte, J., Guiraud-Chaumeil, C., Vincent, M. C., Mandel, J. L., Tanner, S. M., Liechti-Gallati, S., Wallgren-Pettersson, C., Dahl, N., Kress, W., Bolhuis, P. A., Fardeau, M., Samson, F., and Bertini, E. (1997). Mutations in the MTM1 gene implicated in X-linked myotubular myopathy. ENMC International Consortium on Myotubular Myopathy. European Neuro-Muscular Center. *Hum. Mol. Genet.* **6:** 1505–1511.

Laporte, J., Hu, L. J., Kretz, C., Mandel, J. L. (group 1); Kioschis, P., Coy, J. F., Klauck, S. M., Poustka, A. (group 2); and Dahl, N. (group 3). (1996). A gene mutated in X-linked myotubular myopathy defines a new putative tyrosine phosphatase family conserved in yeast. *Nat. Genet.* **13:** 175–182.

Lassar, A., and Munsterberg, A. (1994). Wiring diagrams: Regulatory circuits and the control of skeletal myogenesis. *Curr. Opin. Cell Biol.* **6:** 432–442.

Lassar, A. B., Buskin, J. N., Lockshon, D., Davis, R. L., Apone, S., Hauschka, S. D., and Weintraub, H. (1989). MyoD is a sequence-specific DNA binding protein requiring a region of myc homology to bind to the muscle creatine kinase enhancer. *Cell* **58:** 823–831.

Lassar, A. B., Davis, R. L., Wright, W. E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D., and Weintraub, H. (1991). Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. *Cell* **66:** 305–315.

Li, W. H., and Graur, D. (1991). "Fundamentals of Molecular Evolution." Sinauer, Sunderland, MA.

Li, W. H., Tanimura, M., and Sharp, P. M. (1987). An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J. Mol. Evol.* **25:** 330–342.

Li, W. H., Wu, C. I., and Luo, C. C. (1985). A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **2:** 150–174.

Maizel, J. V., Jr., and Lenk, R. P. (1981). Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proc. Natl. Acad. Sci. USA* **78:** 7665–7669.

Mardis, E. R. (1994). High-throughput detergent extraction of M13 subclones for fluorescent DNA sequencing. *Nucleic Acids Res.* **22:** 2173–2175.

Muscatelli, F., Walker, A. P., De Plaen, E., Stafford, A. N., and Monaco, A. P. (1995). Isolation and characterization of a MAGE gene family in the Xp21.3 region. *Proc. Natl. Acad. Sci. USA* **92:** 4987–4991.

Nagase, T., Ishikawa, K., Nakajima, D., Ohira, M., Seki, N., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N., and Ohara, O. (1997). Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro. *DNA Res.* **4:** 141–150.

Nehls, M., Messerle, M., Sirulnik, A., Smith, A. J., and Boehm, T. (1994). Two large insert vectors, lambda PS and lambda KO, facilitate rapid mapping and targeted disruption of mammalian genes. *Biotechniques* **17:** 770–775.

Palmieri, G., Romano, G., Ciccodicola, A., Casamassimi, A., Campanile, C., Esposito, T., Cappa, V., Lania, A., Johnson, S., Reinbold,

R., *et al.* (1994). YAC contig organization and CpG island analysis in Xq28. *Genomics* **24:** 149–158.

Platzer, M., Rotman, G., Bauer, D., Uziel, T., Savitsky, K., Bar-Shira, A., Gilad, S., Shiloh, Y., and Rosenthal, A. (1997). Ataxia–telangiectasia locus: Sequence analysis of 184 kb of human genomic DNA containing the entire ATM gene. *Genome Res.* **7:** 592–605.

Pilia, G., Randall, D., Aissani, B., Bernardi, G., and Schlessinger, D. (1993). Isochores and CpG islands in YAC contigs in human Xq26.1–qter. *Genomics* **17:** 456–462.

Rogner, U. C., Heiss, N. S., Kioschis, P., Wiemann, S., Korn, B., and Poustka, A. (1996). Transcriptional analysis of the candidate region for incontinentia pigmenti (IP2) in Xq28. *Genome Res.* **6:** 922–934.

Rogner, U. C., Kioschis, P., Wilke, K., Gong, W., Pick, E., Dietrich, A., Zechner, U., Hameister, H., Pragliola, A., Herman, G. E., Yates, J. R. W., Lehrach, H., and Poustka, A. (1994). A YAC clone map spanning the human chromosomal band Xq28. *Hum. Mol. Gen.* **3:** 2137–2146.

Rogner, U. C., Wilke, K., Steck, E., Korn, B., and Poustka, A. (1995). The melanoma antigen gene (MAGE) family is clustered in the chromosomal band Xq28. *Genomics* **29:** 725–731.

Rosenthal, A., Coutelle, O., and Craxton, M. (1993). Large-scale production of DNA sequencing templates by microtitre format PCR. *Nucleic Acids Res.* **21:** 173–174.

Sedlacek, Z., Korn, B., Konecki, D. S., Siebenhaar, R., Coy, J. F., Kioschis, P., and Poustka, A. (1993). Construction of a transcription map of a 300 kb region around the human G6PD locus by direct cDNA selection. *Hum. Mol. Genet.* **11:** 1865–1869.

Stenmark, H., Aasland, R., Toh, B. H., and D'Arrigo, A. (1996). Endosomal localization of the autoantigen EEA1 is mediated by a zinc-binding FYVE finger. *J. Biol. Chem.* **271:** 24048–24054.

Tanner, S. M., Laporte, J., Guiraud-Chaumeil, C., and Liechti-Gallati, S. (1998). Confirmation of prenatal diagnosis results of X-linked recessive myotubular myopathy by mutational screening, and description of three new mutations in the MTM1 gene. *Hum. Mutat.* **11:** 62– 68.

Tapscott, S. J., Lassar, A. B., and Weintraub, H. (1992). A novel myoblast enhancer element mediates MyoD transcription. *Mol. Cell. Biol.* **12:** 4994–5003.

Timms, K. M., Lu, F., Shen, Y., Pierson, C. A., Muzny, D. M., Gu, Y., Nelson, D. L., and Gibbs, R. A. (1995). 130 kb of DNA sequence reveals two new genes and a regional duplication distal to the human iduronate-2-sulfate sulfatase locus. *Genome Res.* **5:** 71–78.

Wallgren-Peterson, C., and Thomas, N. (1994). Report on the 20th ENMC sponsored international workshop: Myotubular/centronuclear myopathy. *Neuromusc. Disord.* **4:** 71–74.

Warren, S. T., Knight, S. J., Peters, J. F., Stayton, C. L., Consalez, G. G., and Zhang, F. P. (1990). Isolation of the human chromosomal band Xq28 within somatic cell hybrids by fragile X site breakage. *Proc. Natl. Acad. Sci. USA* **87:** 3856–3860.

Weintraub, H., Dwarki, V. J., Verma, I., Davis, R., Hollenberg, S., Snider, L., Lassar, A., and Tapscott, S. J. (1991). Muscle-specific transcriptional activation by MyoD. *Genes Dev.* **5:** 1377–1386.