The Genomic Organization of a Human Creatine Transporter (CRTR) Gene Located in Xq28

NATALIA SANDOVAL,* DAVID BAUER,* VOLKER BRENNER,* JOHANNES F. COY,† BERND DRESCHER,* PETRA KIOSCHIS,† BERND KORN,† GERALD NYAKATURA,* ANNEMARIE POUSTKA,† KATHRIN REICHWALD,* ANDRÉ ROSENTHAL,*,[1] AND MATTHIAS PLATZER*

*Institut für Molekulare Biotechnologie, Genomanalyse, Beutenbergstrasse 11, 07745 Jena, Germany; and †Deutsches Krebsforschungszentrum Heidelberg, Molekulare Genomanalyse, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

Received January 12, 1996; accepted May 3, 1996

During the course of a large-scale sequencing project in Xq28, a human creatine transporter (CRTR) gene was discovered. The gene is located approximately 36 kb centromeric to ALD. The gene contains 13 exons and spans about 8.5 kb of genomic DNA. Since the creatine transporter has a prominent function in muscular physiology, it is a candidate gene for Barth syndrome and infantile cardiomyopathy mapped to Xq28.

Large-scale genomic sequencing is a fast and efficient approach to identify genes in chromosomal regions between several hundred kb and 1–2 Mb. With this aim, we are currently sequencing a 360-kb region of Xq28 around the locus of the neural cell adhesion molecule L1CAM (15). This region is contained in a cosmid contig comprising 15 overlapping cosmids (11; Kioschis *et al.*, in preparation). By sequencing, we have shown that this region contains a number of genes including ALD, L1CAM, AVPR2, and MeCP2 (Platzer *et al.*, in preparation). Cosmid 8B7 is the most centromeric of the cosmid contig, and it contains the entire creatine transporter (CRTR) gene.

The CRTR has a prominent role in muscle physiology, as it is required for the uptake of creatine. Interconversion between creatine and phosphocreatine, catalyzed by creatine kinase, provides a dynamic reservoir of high-energy phosphate in muscle and brain (3). It has been shown that inhibition of creatine transport in experimental animals causes muscle weakness (14). Due to its importance in muscular physiology, the CRTR gene can be considered a candidate gene for neuromuscular disorders mapped to Xq28.

In the course of sequencing the Xq28-specific cosmid 8B7, an 11-kb finished sequence containing the CRTR gene was obtained and submitted to the EMBL database (Accession No. Z66539). A homology search against the EMBL database revealed two major hits, both representing human CRTR cDNAs (Accession Nos. L31409 and S74039). Alignment analysis together with gene prediction identified a human CRTR gene composed of 13 exons (Fig. 1). With the exception of exons 1 and 13, they range from 100 to 250 bp, with an average of 137 bp (Table 1). The introns vary in size, from 76 to 1538 bp. The consensus dinucleotides GT and AG were found at the donor and acceptor splice sites of all introns. The first methionine of the open reading frame is located in exon 1 at position 2675 of the database entry, and the stop codon is located in exon 13 at position 9312. Thus, the first exon contains a 5′-untranslated region (UTR) of at least 472 bp, the last exon a 3′ UTR of 1511 bp. In the translated region, the genomic sequence is identical to cDNA L31409 isolated from human kidney library (12) with the exception of one base at position 7215. In contrast, the homology of the genomic sequence to cDNA S74039 isolated from human brainstem/spinal cord (16) is much lower. A 94% identity on the nucleic acid level and a 98% identity on the amino acid level was found. The UTRs differ significantly. These differences cannot be explained through sequencing errors or polymorphisms but instead indicate different origins of the cDNAs. cDNA L31409 appears to be a transcript of the Xq28 CRTR gene reported here. cDNA S74039 is obviously expressed from another locus. A second CRTR is located on chromosome 16 (8), but it is not yet established whether this second locus is expressed.

Sequence data from this article have been deposited with the GenBank/EMBL Data Libraries under Accession No. Z66539.

[1] To whom correspondence should be addressed. Telephone: 0049-3641-656242. Fax: 0049-3641-656255.

GENOMICS 35, 383–385 (1996)
ARTICLE NO. 0373
0888-7543/96 $18.00

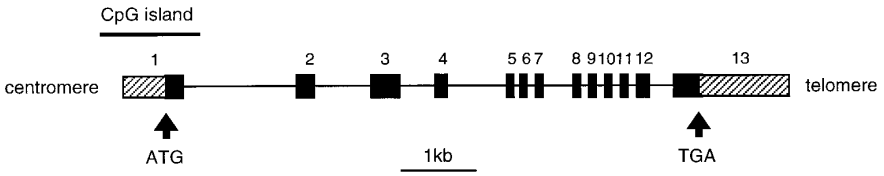

FIG. 1. Genomic map of the human creatine transporter located on Xq28-specific cosmid 8B7 showing the positions of exons and introns. The shaded areas are 5′- and 3′-untranslated regions, and the solid areas are the regions of the exons containing the open reading frame. Cosmid 8B7 DNA was prepared and sequenced as previously described by Craxton (6). In brief, after sonication of the DNA, a 0.8- to 1.4-kb fraction was subcloned into the *Sma*I site of M13mp18. M13 templates were prepared through magnetic bead technology and sequenced using dye-primer chemistry. The raw data were collected using ABI 373 A automated sequencers and assembled with the XBAP program (7). Gaps were closed using custom-made primers on M13 templates, PCR products, or cosmid DNA in combination with *Taq* dye terminator chemistry (Perkin–Elmer) or internal labeling (Pharmacia). Homology searches against the EMBL database were performed using BLAST (Version 1.4) (1) and FASTA (Version 2.0) (13). Gene prediction programs GRAIL (18) and XPOUND (17) were used. Sequence alignments were performed by "Global Alignment Program" (GAP) (10). The computer program "Transcription Start Site" using both Ghosh/Prestridge (TSSG) motif data and the Wingender (TSSW) motif database (V. V. Solovyev, A. A. Salamov, and C. B. Lawrence, in preparation) was applied for promoter prediction.

Promoter prediction indicated a TATA box at position 1996 and a transcription start site at position 2032. Numerous transcription factor binding sites were predicted using TSSG and TSSW. These included four GC boxes, putative binding sites for the transcription factor Sp1 (GGGCGG), which were present at positions 1781, 1820, 1889, and 1908. Ten putative binding sites for AP-2 (CCCMNSSS) are found in the same region (1776, 1783, 1934, 1940, 1946, 1963, 1964, 1969, 1970, and 2020).

A CpG island spans positions 1751 to 3136, entirely covering exon 1. The GC content of the CpG island upstream of the putative transcription initiation site is 75%. In the downstream region, the GC portion rises to 81%. CpG islands are known to be a marked feature of housekeeping genes (4).

Mapping and sequencing data obtained in the course of our large-scale effort around the L1CAM locus have shown that the CRTR gene is located 36 kb centromeric to ALD and is transcribed from centromere to telomere (Platzer *et al.*, in preparation). This finding is contradictory to an earlier published report localizing CRTR distal to G6PD (9). There are no indications available that there is a second CRTR locus in Xq28.

Since the CRTR plays a major role in muscular physiology, it is appropriate to consider the CRTR as a candidate gene for Barth syndrome (2, 5) and infantile cardiomyopathy (19)–neuromuscular disorders that have been mapped to Xq28. Barth syndrome is characterized by cardioskeletal myopathy with short stature, neutropenia, and abnormal mitochondria. It has been speculated that infantile cardiomyopathy is a more severe allelic form of Barth syn-

TABLE 1

Exon–Intron Organization of the Human Creatine Transporter Gene.

No.	Exon length (bp)	Position[a]	Splice acceptor	Splice donor	Intron length (bp)
1	734	2202–2936		GCGGAG**gt**gagttcccccgccgccg	1538
2	132	4475–4606	cgctgtgcctccaccccc**ag**GTGTGT	TCAAAG**gt**gagcagcccttggccagc	797
3	250	5404–5653	caaggacttcccggcccc**ag**GCCTGG	CTGGGA**gt**gagtccggcacctctggg	421
4	133	6075–6207	gacctccctccctcccct**ag**GAACAA	GGAAAG**gt**accactagaggcatgcag	933
5	135	7141–7275	tgagcagcctggccccccc**ag**ATCGTG	CCTCAG**gt**gaggtggaggtggagagg	87
6	104	7363–7466	gctctccggcccttctct**ag**GTGTGG	CTACAA**gt**aagcaccgccgccctgcc	95
7	125	7562–7686	ctggcccctccaccccctc**ag**GGACGC	AGTCAG**gt**agggccctacccccagcc	318
8	113	8005–8117	agcctgcacctttcccac**ag**GGCCGG	AGCCAG**gt**ttgcatggggctctggga	112
9	138	8230–8367	ctgagctgccctggccac**ag**TTTGTA	ACTGAT**gt**gagtggggtgggggtct	76
10	103	8444–8546	gactgggctctgtcccccc**ag**GGCGGG	TGTACG**gt**aggtcatggctgagggct	86
11	101	8633–8733	cccgcctcacctcgccg**ag**GAGCTG	TGCATG**gt**aagggctgggggaggtgg	85
12	171	8819–8989	cgcagcattctggtccgt**ag**GGCATC	GCTGAG**gt**aaggctcccgcccggccc	184
13	1511	9174–10684	catgtcctcctctcctgc**ag**CGCTGG		

Note. The capital letters represent coding sequences and lower case letters represent intron sequences.
[a] Refers to database entry Z66539.

drome (19). We are currently performing mutation analysis in Barth syndrome families.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft Grant Pl173/2-1 and the EC Genome Analysis Program.

Note added in proof. While the manuscript was in publication, a paper was published by Bione *et al.* (1996, *Nature Genet.* 12: 385–389) linking Barth Syndrome to the novel gene G4.5.

REFERENCES

1. Altschul, S. F., Warren, G., Miller, W., Myers, E. M., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215: 403–410.
2. Barth, P. (1983). An X-linked mitochondrial disease affecting cardiac muscle, skeletal muscle and neutrophil leucocytes. *J. Neurol. Sci.* 62: 327–355.
3. Bessman, S. P. (1985). The creatine–creatine phosphate energy shuttle. *Annu. Rev. Biochem.* 54: 831–862.
4. Bird, A. (1986). CpG-rich islands and the function of DNA methylation. *Nature* 321: 209–213.
5. Bolhuis, P., Hensels, G. W., Hulsebos, T. J., and Bass, F. (1991). Mapping of the locus for X-linked cardioskeletal myopathy with neutropenia and abnormal mitochondria (Barth syndrome) to Xq28. *Am. J. Med. Genet.* 48: 481–485.
6. Craxton, M. (1993). Cosmid sequencing. *In* "Methods in Molecular Biology (DNA Sequencing Protocols)" (H. and A. Griffin, Eds.), Vol. 23, pp. 149–167.
7. Dear, S., and Staden, R. (1991). A sequence assembly and editing for efficient management of large projects. *Nucleic Acids Res.* 19: 3907–3911.
8. Eichler, E. E., Lu, F., Antonacci, R., Doggett, N., Moyzis, R., Baldini, A., Lee, C. C., Gibbs, R. A., and Nelson, D. L. (1995). A gene-rich duplication between Xq28 and 16p11.1 suggests a novel mechanism for genome evolution. 6th X Chromosome Workshop, Banff, Canada.
9. Gregor, P., Nash, S. R., Caron, M. F., and Warren, S. T. (1995). Assignment of the creatine transporter gene (SLC6A8) to human Chromosome Xq28 telomeric to G6PD. *Genomics* 25: 332–333.
10. Huang, X. (1994). On global sequence alignment. *Comput. Appl. Biosci.* 10: 227–235.
11. Kioschis, P., Gong, W., Rogener, U. C., Wilke, K., Manca, A., Coy, J. F., and Poustka, A. (1994). Cosmid contigs in Xq27.3–Xqter. *Cytogenet. Cell Genet.* 67: 355.
12. Nash, S., Giros, B., Kingsomore, S., Rochelle, J., Suter, S., Gregor, P., Seldin, M., and Caron, M. (1994). Cloning, pharmacological characterization, and genomic localization of the human creatine transporter. *Receptors Channels* 2: 165–174.
13. Pearson, W. R., and Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* 85: 2444–2448.
14. Petrofsky, J., and Fitch, C. (1980). Contractile characteristics of skeletal muscles depleted of phophocreatine. *Pfluegers Arch.* 384: 123–129.
15. Platzer, M., Bauer, D., Brenner, V., Coy, J. F., Drescher, B., Kioschis, P., Korn, B., Nyakatura, G., Poustka, A., Reichwald, K., Sandoval, N., and Rosenthal, A. (1995). Sequencing and analysis of 360 kb of human Xq28 genomic DNA in the region of the L1CAM locus. 6th X Chromosome Workshop, Banff, Canada.
16. Sora, I., Richman, J., Santoro, G., Wei, H., Wang, Y., and Vanderah, T. (1994). The cloning and experession of a human creatine transporter. *Biochem. Biophys. Res. Commun.* 204: 419–427.
17. Thomas, A., and Skolnick, A. (1994). Probablistic model for detecting coding regions in DNA sequences. *IMA J. Math. Appl. Med. Biol.* 11: 149–160.
18. Uberbacher, E. C., and Mural, R. J. (1991). Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. USA* 88: 11261–11265.
19. Wilson, M. J., Gedeon, A. K., Colley, A. C., Mulley, J., and Sillence, D. O. (1993). Infantile X-linked cardiomyopathy linked to Xq28. 4th International Workshop on the X Chromosome, St. Louis.