ELSEVIER

Biochimica et Biophysica Acta 1240 (1995) 119–124

BBA Biochimica et Biophysica Acta

Short sequence-paper

Analysis of the 5′ end of the rat plasma membrane Ca^{2+}-ATPase isoform 3 gene and identification of extensive trinucleotide repeat sequences in the 5′ untranslated region [1,2]

Scott E. Burk, Anil G. Menon, Gary E. Shull *

Department of Molecular Genetics, Biochemistry and Microbiology, University of Cincinnati College of Medicine, 231 Bethesda Avenue, ML 524, Cincinnati, OH 45267-0524, USA

Received 24 July 1995; accepted 30 August 1995

Abstract

We have characterized the 5′ end of the rat gene encoding isoform 3 of the plasma membrane Ca^{2+}-ATPase using S1 nuclease protection and DNA sequence analysis. The 5′-untranslated region consists of over 900 nucleotides and includes a 217-nucleotide sequence composed of alternating tracts of TCC and ACC trinucleotides. Analysis of genomic sequences 5′ to the transcription initiation site revealed potential binding sites for transcription factors that are active in muscle and brain.

Keywords: Calcium pump; PMCA3; Trinucleotide repeat

Calmodulin-sensitive plasma membrane Ca^{2+}-transporting ATPases are encoded by four distinct genes [1–6], each of which gives rise to multiple PMCA variants by alternative splicing of one or more exons [7–14]. PMCA1 and PMCA4 are expressed in most rat tissues [3,6], whereas PMCA2 is expressed primarily in brain and heart [3], and PMCA3 is expressed primarily in brain and skeletal muscle [3,12]. To begin the analysis of its genetic regulation we recently characterized the structure of the PMCA3 gene [12], however, our attempts to identify the transcription initiation site were unsuccessful due to the presence of an extensive trinucleotide repeat sequence at the 5′ end of the gene.

Interestingly, trinucleotide repeat sequences located in both coding and untranslated regions have been shown to be involved in a number of human genetic diseases, such as Fragile X Syndrome [15], Myotonic Dystrophy [16], and Huntington's Disease [17]. The human PMCA3 gene has been mapped to the Xq28 region of the X chromosome [18]. The disease genes for a number of X-linked neurological and skeletal muscle disorders have been mapped to this region [19], and if PMCA3 is involved in any of these diseases it is conceivable that the trinucleotide repeat sequence could be involved. Thus, one objective of the current study was to characterize the repeat sequence and determine whether it was included within the transcribed region of the gene and within the mature mRNA.

Because the plasma membrane calcium pumps play a critical role in cellular calcium homeostasis the regulation of their expression is of considerable interest. However, identification and analysis of the transcriptional control regions of the PMCA genes have been hampered by the presence of large introns in the untranslated regions [12,13], by GC-rich sequences in the case of the PMCA1 gene [13], and by the trinucleotide repeat sequence in the case of PMCA3 [12]. Thus, a second objective of this study was to identify the promoter region of the PMCA3 gene.

Rat brain cDNA clone RB 7-2, the longest PMCA3 cDNA identified in one of our previous studies [3], contained a lengthy 5′-untranslated sequence that extended to nucleotide −695 relative to the translation initiation site.

Abbreviations: PMCA1, 2, 3, and 4, plasma membrane calcium ATPase isoforms 1, 2, 3, and 4; PCR, polymerase chain reaction.

* Corresponding author. Fax: +1 (513) 5588474.

[1] The complete nucleotide sequence reported in this paper has been submitted to the EMBL/GenBank Data libraries under the accession number U29397.

[2] This work was supported by National institutes of Health Grants HL41558 and HL41496.

SSDI 0005-2736(95)00217-0

When we characterized the organization of the PMCA3 gene [12] we determined that the 5′-untranslated sequence was distributed over at least 3 exons separated by introns of 3.5 and 16.5 kb and that the 5′ end of the PMCA3 cDNA began at the end of a complex trinucleotide repeat sequence of over 200 nucleotides. Based on data from these earlier studies it was clear that at least a portion of the trinucleotide sequence was included in the transcribed region of the gene, and it seemed likely that transcription would begin upstream of the repeat element rather than within it.

To determine the sequence of both the trinucleotide repeat region, as well as the upstream flanking region that seemed likely to contain the promoter, we analyzed a cosmid clone containing genomic sequences upstream of the 5′-most exon identified previously [12]. A *Pst*I fragment extending 2.8 kb upstream from the known sequences of exon 1 was subcloned into a plasmid vector and most of the sequence was determined in both strands by the dideoxy chain-termination method [20] using T7 DNA polymerase. However, because the trinucleotide repeat sequence was refractory to analysis using the chain termination procedure the chemical cleavage procedure [21] was also used.

As illustrated in Fig. 1, sequence analysis by chemical cleavage revealed a striking pattern of alternating trinucleotide repeat sequences extending from position −905 to −687 relative to the translation initiation site (also see Fig. 4). It begins with a 57-nucleotide sequence consisting of TCC trinucleotides, which is followed by a 30-nucleotide sequence consisting of ACC trinucleotides. The ACC repeat is followed by a 28-nucleotide sequence consisting primarily of TCC trinucleotides with a mirror repeat in the middle of the pattern having the sequence CCCTCTTCTCCC. Finally, this sequence is followed by a nearly perfect 102-nucleotide ACC repeat containing a single T for C substitution.

We next performed experiments to identify the promoter region of the gene. One of the PMCA3 mRNAs identified previously is approximately 4.5 kb in length and contains coding sequence and 3′-untranslated sequence totaling 3.6 kb, excluding the poly(A) tract [12]. Thus, the 5′-untranslated sequence would not be expected to exceed approximately 900 nucleotides, leaving little room for additional sequence beyond the repeat element which extends to nucleotide position −905. For this reason it seemed likely that the transcription start site would be located just 5′ to the repeat sequence. S1 nuclease protection experiments were performed using a 78-base synthetic oligonucleotide complementary to nucleotides −978 to −901, which are immediately 5′ of the trinucleotide repeat element (probe B, see Fig. 4). As shown in Fig. 2, multiple protected fragments were detected in the region between nucleotides −934 and −919, with the sites around position −934 corresponding to a consensus initiator site ([22], labeled INR in Fig. 4). Because these bands were very

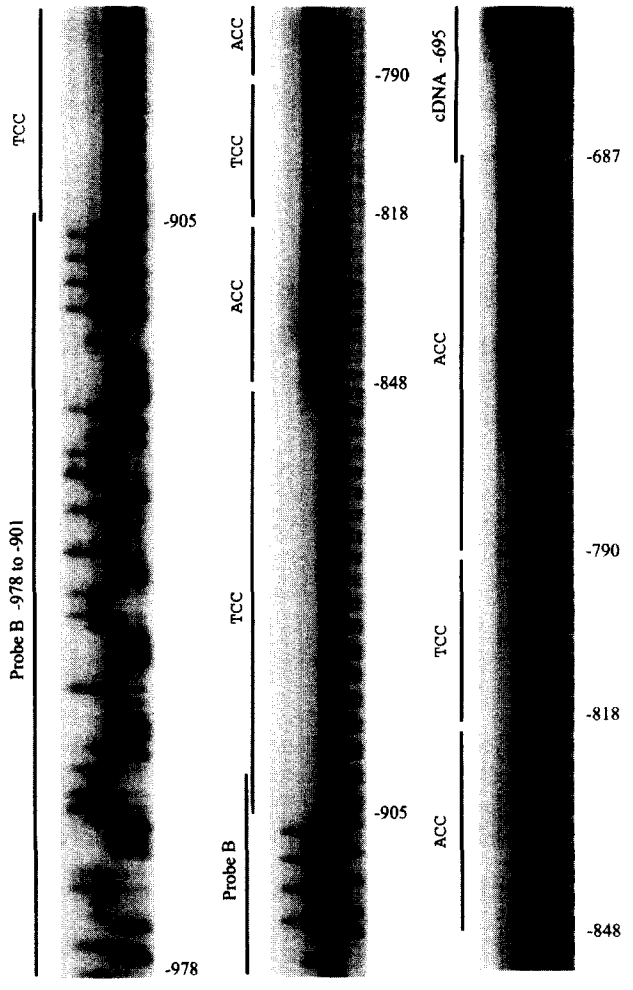

Fig. 1. Sequence analysis of the 5′ end of the rat PMCA 3 gene. The three panels show overlapping sequence determined using the chemical cleavage procedure [21]. For each panel: lane 1, G reaction; lane 2, G+A reaction; lane 3, C+T reaction; lane 4, C reaction. The nucleotide sequence is numbered relative to the translation initiation site of the mature mRNA and extends from nucleotide −978 (lower left) in the 5′-flanking sequence to approximately nucleotide −610 (upper right) in the first exon. Vertical lines delineate sequences corresponding Probe B (used for S1 nuclease protection and Northern blots), TCC and ACC trinucleotide tracts located between residues −901 and −688, and the 5′ end of rat PMCA3 cDNA clone RB7-2 [3] beginning at nucleotide −695. Note the striking pattern of alternating TCC and ACC trinucleotide repeats that occur at the beginning of the 5′-untranslated sequence.

faint we attempted to confirm the results using a 5′ Rapid Amplification of cDNA Ends protocol (Clontech), in which brain mRNA was reversed transcribed, the first strand cDNA ligated to an anchor sequence, and the products amplified by PCR. However, these experiments were unsuccessful, possibly due to difficulties in reverse transcribing the trinucleotide sequences and to the lack of a significant stretch of unique sequence preceding the trinucleotides. It should be noted that the first major TCC repeat is preceded by four GCT trinucleotides, and that the consensus initiator sequence and potential start sites identified by S1 nuclease protection are within 5 to 20 nucleotides of the GCT trinucleotides.

Because the nuclease protection experiments provided only one piece of evidence regarding the location of the transcription initiation site, Northern blot hybridization was carried out to further examine the possibility that transcription initiation occurs within the region immediately preceding the trinucleotide repeat sequence. The 78-nucleotide S1 probe (probe B) was PCR amplified and the antisense strand was uniformly labeled and used as a probe to analyze brain and skeletal muscle mRNA. Although the signals were very faint, as anticipated based on the small amount of sequence with which the probe was expected to hybridize (35 nucleotides or less), the more abundant 7.5 kb PMCA3 mRNA was clearly detected in brain (Fig. 3, panel B), which contains the highest levels of PMCA3 mRNA. This confirmed that the mature 7.5 kb brain mRNA contains at least a small amount of sequence 5′ to the repetitive element. A trace signal was detected in the skeletal muscle mRNA sample as well, but it was not as evident as in the brain mRNA sample. The blot was then stripped and hybridized with probe A (see Fig. 4), which overlaps the 5′ end of probe B but does not include the region containing the apparent transcription initiation sites (Fig. 3, panel A). PMCA3 mRNAs were not detected with this probe, even after an autoradiographic exposure time of 17 days, indicating that sequences immediately upstream of the sites identified by S1 nuclease protection are not included in the PMCA3 mRNA. The blot was then hybridized with probe C, which corresponds to the 3′ end of exon 1 and spans nucleotides −585 to −411 (see Fig. 4). This 175-nucleotide probe yielded strong signals (Fig. 3, panel C) with all of the PMCA3 mRNAs identified previously [3].

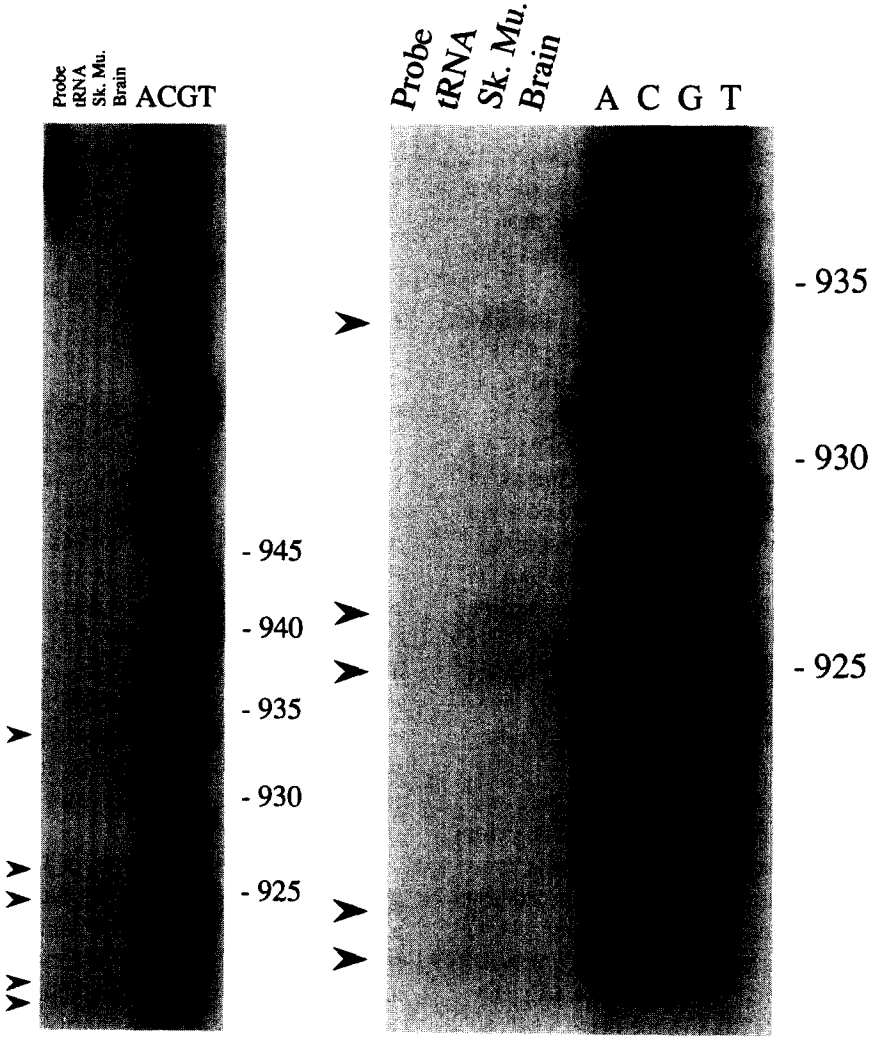

Fig. 2. S1 nuclease protection analysis of the transcription initiation site. Probe B, complementary to nucleotides −978 to −901 (underlined in Fig. 4) was 5′ end-labeled with ^{32}P, annealed with 50 μg of tRNA, 50 μg of rat skeletal muscle total RNA, or 50 μg of rat brain total RNA, and digested with S1 nuclease as described [11]. The marker lanes (labeled according to the sense strand) are sequencing reactions in which a labeled primer complementary to nucleotides −901 to −925 was extended on a template consisting of a subcloned *Pst*I fragment from the 5′ end of the gene. The samples were fractionated on a DNA sequencing gel and analyzed by autoradiography. The autoradiogram (left panel) revealed five protected fragments between nucleotide positions −934 and −919. Because the signals were faint and did not reproduce well, the region of the autoradiogram containing the signals was scanned using NIH Image version 1.57, background was subtracted, and contrast was enhanced (right panel).

The results of the S1 nuclease protection (Fig. 2) and Northern blot (Fig. 3) analyses strongly suggest that initiation of PMCA3 gene transcription occurs at a cluster of sites located within 30 nucleotides of the trinucleotide repeat sequence, at least in the case of the 7.5 kb brain mRNA. Because the signal for the 7.5 kb skeletal muscle mRNA was very faint and a signal was not apparent for the 4.5 kb mRNA (possibly obscured by background), we cannot rule out the possibility that transcription of the skeletal muscle mRNA (particularly the 4.5 kb mRNA) also initiates at a site 3′ to the trinucleotide repeat sequence. Also, the possibility that transcription initiation occurs further upstream for one or more of the PMCA3 mRNAs, with an additional exon being spliced to the region identified by probe B, cannot be entirely ruled out. This seems unlikely, however, due to the lack of good potential splice acceptor sites within this region. Although there are numerous AG dinucleotides in the upstream region, they either lack the polypyrimidine tract usually found before the splice acceptor site or are closely preceded by another AG dinucleotide, which has not been observed within 15 nucleotides of known acceptor sites [23]. The presence of a good consensus initiator sequence [22] around position −935, which corresponds to the 5′-most protected fragments identified by nuclease protection, lends additional support to the likelihood that this region contains the PMCA3 promoter.

The sequence of the first exon, which includes the trinucleotide repeat sequence, and the sequence of the 5′-flanking region of the PMCA3 gene are shown in Fig. 4. The region characterized contains 2839 base pairs of sequence beyond that which was present in the PMCA3 cDNA [3]. With regard to basic promoter elements, a consensus initiation site (INR) [22] is present at position −935, consistent with the identified region of transcription initiation. The sequence does not contain a TATA box in close proximity to the putative initiator, but does contain a CCAAT motif as well as two potential SP1 binding sites within 250 base pairs of the putative initiator. The sequence also contains two additional potential SP1 sites (KRGGCKRRK) [3] [24] as well as a potential AP1 binding site (TGASTMA) [24].

Because PMCA3 is expressed in skeletal muscle and brain it was of interest to examine the sequence for potential binding sites for muscle and brain specific transcription factors. Potential muscle specific transcription factor sites that were identified include: (1) four CArG sites that match the consensus sequence (CCWWWWWWGG)[3] [25] at 9/10 positions, (2) fourteen perfect consensus E-Box sites (CANNTG) [26], (3) one perfect MCAT site (CATTCCT) [27] and ten sites that match the MCAT consensus sequence at 6/7 positions, and (4) two MADS-Box sites (YTAWWWTAR) [28] that match the consensus sequence at 9/10 positions. Several potential brain transcription factor binding sites were identified including a perfect Octamer consensus site (ATGCWAAT) [29] and two perfect POU consensus sites (GMATN$^{=0,2,3}$ WAAT). Particularly interesting was the presence of a potential BSF1 site that matched the consensus sequence (GAGAGGGGAGAGGRGRGAGRRG) [30] at 19/22 positions. This site was located in the antisense strand within a mirror repeat occurring in the second TCC trinucleotide element of the 5′-untranslated region. Further studies will be needed to confirm which, if any, of these potential transcription factor binding sites are involved in PMCA3 gene regulation.

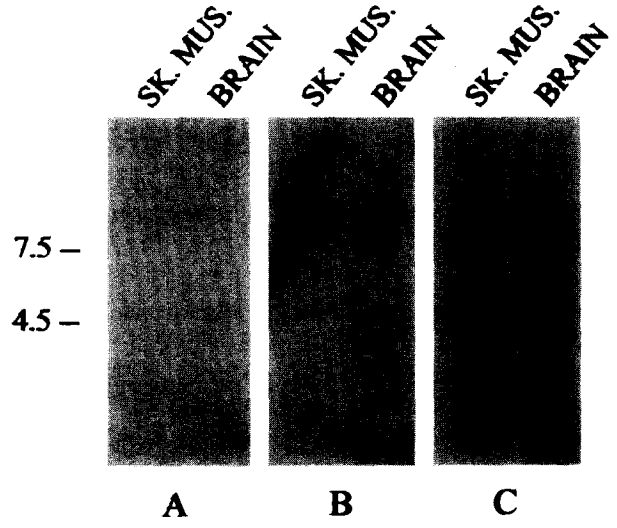

Fig. 3. Northern blot analysis of PMCA3 mRNAs. 15 μg of poly(A)$^+$ RNA from rat skeletal muscle and brain were analyzed as described [3]. The samples were hybridized first with a probe complementary to sequence B, a 78-nucleotide sequence spanning the transcription initiation site that was also used for S1 nuclease analysis. The blot was then stripped and hybridized with a probe complementary to sequence A, a 292-nucleotide sequence that overlaps probe B and is 5′ to the transcription initiation site. The blot was then stripped and hybridized with a probe complementary to sequence C, a 175-nucleotide sequence from the 3′ end of exon 1 that was present in the previously cloned PMCA3 cDNA. Panels are labeled according to the probe used. Sequences corresponding to each probe are underlined in Fig. 4. The autoradiographic exposure time was 17 days for the probe A, 10 days for probe B, and 4 days for probe C.

In summary, our analysis of the 5′ end of the PMCA3 gene provides evidence that transcription initiation occurs at a cluster of sites between nucleotides −935 and −920 relative to the translation initiation site (excluding intron sequences). The beginning of this cluster corresponds to a sequence with strong similarity to a consensus initiator sequence [19]. The PMCA3 promoter occurs within a relatively GC-rich sequence although, unlike the PMCA1 promoter which is embedded in a CpG island [13], the PMCA3 promoter region contains very few CpG dinucleotides. The promoter lacks a TATA box but contains several potential SP1 sites and a CCAAT sequence, which

[3] Standard IUPAC abbreviations: K = G or T, M = A or C, R = A or G, S = G or C, W = A or T, Y = C or T.

may serve as basic promoter elements. Also present are numerous sites that have the potential to be involved in the skeletal muscle and brain specificity of PMCA3 expression.

One of the most striking features of the PMCA3 gene is the presence of a complex trinucleotide repeat sequence of over 200 nucleotides. It is contained in the mature PMCA3 mRNA and is located near the beginning of an unusually long 5′-untranslated sequence that is distributed over three exons spanning 20 kb of genomic DNA [12]. The presence of a potential binding site for the brain specific transcription factor BSF1 [30] within the TCC trinucleotide tract located approximately 120 nucleotides downstream of the transcription start site is intriguing, particularly as it occurs within a subtle alteration in the trinucleotide pattern that forms the center of a perfect 80-nucleotide mirror-repeat. With regard to the possible regulatory significance of the trinucleotide repeats and the potential BSF1 site, it should be noted that sequences exhibiting mirror-repeat symmetry and containing primarily purines in one strand and pyrimidines in the other strand tend to form unusual DNA structures and are often found in or near the promoter regions of eukaryotic genes [31].

Finally, it is worth noting that trinucleotide repeats have been implicated in several disease states including Fragile X Syndrome [15], Myotonic Dystrophy [16], and Huntington's Disease [17]. While additional studies will be necessary to determine whether the trinucleotide repeat se-

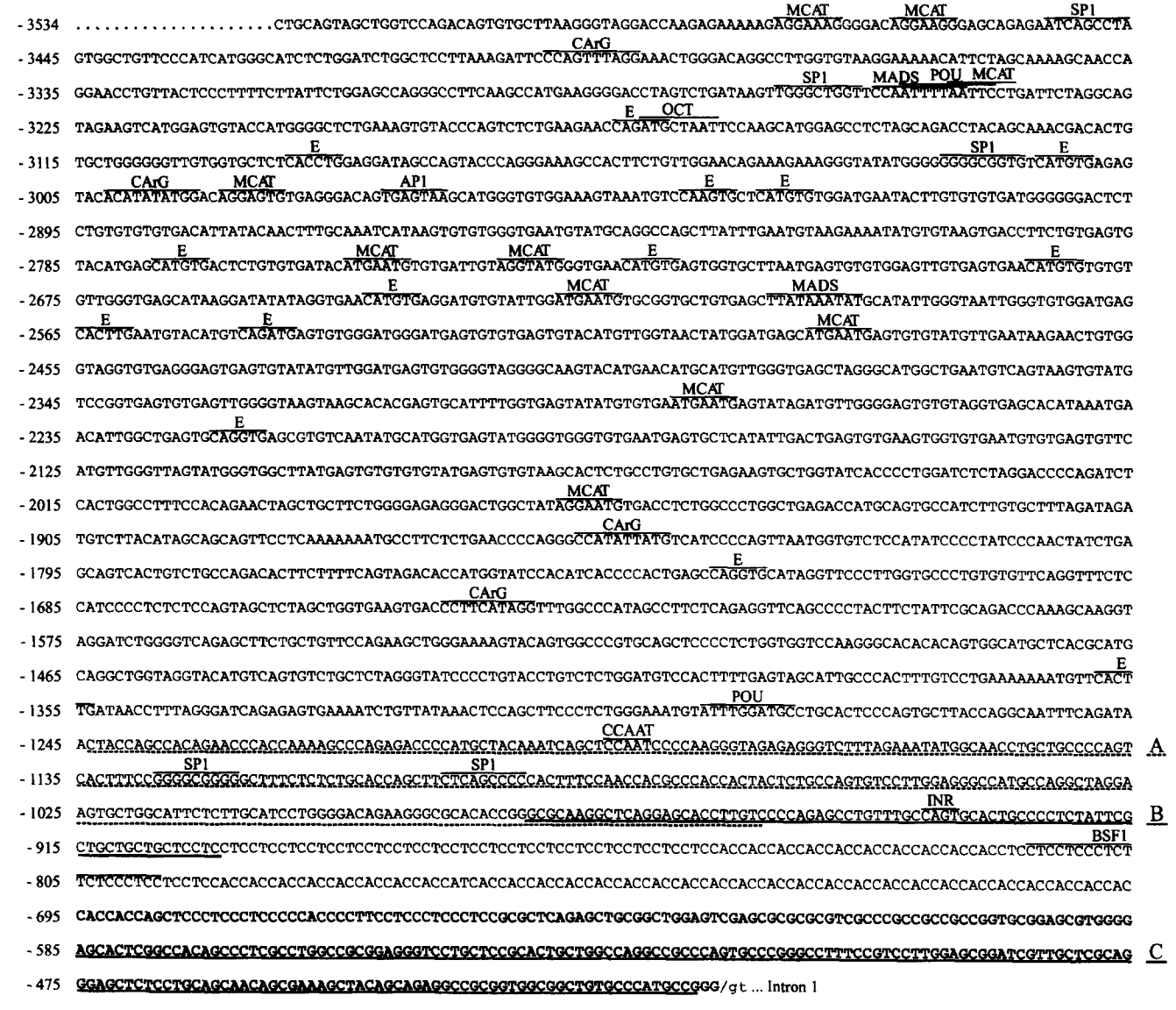

Fig. 4. Nucleotide sequence of the 5′ end of the PMCA3 gene. Nucleotide sequence of the 5′ flanking region and part of the first exon is shown, numbered with respect to the translation start site of the mature mRNA. The nucleotide sequences of probes A and B are underlined by dashed and solid lines, respectively. The nucleotide sequence of the portion of exon 1 present in the RB 7-2 cDNA is in bold type, while the portion of exon 1 used for probe C is underlined. Potential transcription factor binding sites (see text) are over-lined and labeled.

quence in the 5′-untranslated sequence of the PMCA3 mRNA is involved in a disease process, the mapping of human PMCA3 to chromosome Xq28 [18] and its expression in skeletal muscle and brain suggest that PMCA3 is a good candidate gene for skeletal muscle diseases such as X-linked myotubular myopathy [32,33] or some of the neurological diseases that map to this region [19].

References

[1] Shull, G.E. and Greeb, J. (1988) J. Biol. Chem. 263, 8646–8657.
[2] Verma, A.K., Filoteo, A.G., Stanford, D.R., Wieben, E.D., Penniston, J.T., Strehler, E.E., Fischer, R., Heim, R., Vogel, G., Mathews, S., Strehler-Page, M.-A., James, P., Vorherr, T., Krebs, J. and Carafoli, E. (1988) J. Biol. Chem. 263, 14152–14159.
[3] Greeb, J. and Shull, G.E. (1989) J. Biol. Chem. 264, 18569–18576.
[4] Strehler, E.E., James, P., Fischer, R., Heim, R., Vorherr, T., Filoteo, A.G., Penniston, J.T. and Carafoli, E. (1990) J. Biol. Chem. 265, 2835–2842.
[5] De Jaegere, S., Wuytack, F., Eggermont, J.A., Verboomen, H. and Casteels, R. (1990) Biochem. J. 271, 655–660.
[6] Keeton, T.P. and Shull, G.E. (1995) Biochem. J. 306, 779–785.
[7] Strehler, E.E., Strehler-Page, M.-A., Vogel, G. and Carafoli, E. (1989) Proc. Natl. Acad. Sci. USA 86, 6908–6912.
[8] Brandt, P., Neve, R.L., Kammesheidt, A., Rhoads, R.E. and Vanaman, T.C. (1992) J. Biol. Chem. 267, 4376–4385.
[9] Adamo, H.P. and Penniston, J.T. (1992) Biochem. J. 283, 355–359.
[10] Heim, R., Hug, M., Iwata, T., Strehler, E.E. and Carafoli, E. (1992) Eur. J. Biochem. 205, 333–340.
[11] Keeton, T.P., Burk, S.E. and Shull, G.E. (1993) J. Biol. Chem. 268, 2740–2748.
[12] Burk, S.E. and Shull, G.E. (1992) J. Biol. Chem. 267, 19683–19690.
[13] Hilfiker, H., Strehler-Page, M.-A., Stauffer, T.P., Carafoli, E. and Strehler, E.E. (1993) J. Biol. Chem. 268, 19717–19725.
[14] Stauffer, T.P., Hilfiker, H., Carafoli, E. and Strehler, E.E. (1993) J. Biol. Chem. 268, 25993–26003.
[15] Fu, Y.-H., Kulh, D.P.A., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J.M.H., Holden, J.J.A., Fenwick, R.G., Jr., Warren, S.T., Oostra, B.A., Nelson, D.L. and Caskey, C.T. (1991) Cell 67, 1047–1058.
[16] Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.-P., Hudson, T., Sohn, R., Zemelman, B., Snell, R.G., Rundle, S.A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P.S., Shaw, D.J. and Housman, D.E. (1992) Cell 68, 799–808.
[17] McDonald, M.E., Gusella, J.F., Lehrach, H., O'Donovan, M.C., Housman, D.E., Wasmuth, J.J., Collins, F.C., Harper, P.S. and the Huntingtion's Disease Collaborative Research Group (1993) Cell 72, 971–983.
[18] Wang, M.G., Yi, H., Hilfiker, H., Carafoli, E., Strehler, E.E. and McBride, O.W. (1994) Cytogenet. Cell. Genet. 67, 41–45.
[19] Mandel, J.L., Monaco, A.P., Nelson, D.L., Schlessinger, D. and Willard, H. (1992) Science 258, 103–109.
[20] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[21] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[22] Bucher, P. (1990) J. Mol. Biol. 212, 563–578.
[23] Mount, S.M. (1982) Nucleic Acids Res. 10, 459–472.
[24] Faisst, S. and Meyer, S. (1991) Nucleic Acids Res. 20, 3–26.
[25] Minty, A. and Kedes, L. (1986) Mol. Cell. Biol. 6, 2125–2136.
[26] Blackwell, T.K. and Weintraub, H. (1990) Science 250, 1104–1109.
[27] Mar, J.H. and Ordahl, C.P. (1990) Mol. Cell. Biol. 10, 4271–4283.
[28] Yu, Y.-T., Breitbart, R.E., Smoot, L.B., Lee, Y., Mahdavi, V. and Nadal-Ginard, B. (1992) Genes Dev. 6, 1783–1789.
[29] Li, P., He, X., Gerrero, M.R., Mok, M., Aggarwal, A. and Rosenfeld, M.G. (1993) Genes Dev. 7, 2483–2496.
[30] Motejlek, K., Hauselmann, R., Leitgeb, S. and Luscher, B. (1994) J. Biol. Chem. 269, 15265–15273.
[31] Wells, R.D., Collier, D.A., Hanvey, J.C., Shimizu, M. and Wohlrab, F. (1988) FASEB J. 2, 2939–2949.
[32] Thomas, N.S.T., Williams, H., Cole, G., Roberts, K., Clarke, A., Liechti-Gallati, S., Braga, S., Gerber, A., Meier, C., Moser, H. and Harper, P.S. (1990) J. Med. Genet. 27, 284–287.
[33] Dahl, N., Hu, L.J., Chery, M., Fardeau, M., Gilgenkrantz, S., Nivelon-Chevallier, A., Sidaner-Noisette, I., Mugneret, F., Gouyon, J.B., Gal, A., Kioschis, P., d'Urso, M. and Mandel, J.L. (1995) Am. J. Hum. Genet. 56, 1108–1115.