Volume 244, number 2, 481-483 FEB 06869 February 1989

Family of human Na^+,K^+-ATPase genes

Structure of the putative regulatory region of the α^+-gene

E.D. Sverdlov, D.A. Bessarab, I.V. Malyshev, K.E. Petrukhin, Yu.V. Smirnov, Yu.A. Ushkaryov, G.S. Monastyrskaya, N.E. Broude and N.N. Modyanov

Shemyakin Institute of Bioorganic Chemistry, USSR Academy of Sciences, 117871 Moscow, USSR

Received 26 December 1988

The primary structure of the putative regulatory region of a gene of the Na^+,K^+-ATPase multigene family in the human genome has been determined. This region includes the first exon with all of the untranslatable sequence of mRNA and a dozen nucleotides, coding for the first four amino acids of the hypothetic precursor of the α^+-subunit. The entire region comprises over 1400 bp. The possible role of specific nucleotide blocks within this region in comparison with other genes is discussed.

ATPase, Na^+,K^+-; Subunit, α^+-catalytic; Gene sequence; Regulatory region

1. INTRODUCTION

Na^+,K^+-ATPase of animal cells is an integral membrane protein, containing a catalytic α-subunit and a β-subunit of unknown function (review [1]). As demonstrated previously, there are two distinct isoforms of the α-subunit, called α and α^+, and their localization is tissue-specific (review [2]).

At least five genes and/or pseudogenes, highly homologous to the α-subunit of Na^+,K^+-ATPase [3], have been shown to exist in the human genome, the expression of which is tissue-specific [4]. The tissue specificity of their expression can be regulated at different levels including the regulation of transcription [5]. The identification of structural elements of DNA that are responsible for tissue-specific transcription is essential for understanding the mechanisms of their interaction with cellular transacting factors, probably forming the background of cell differentiation and individual development. This paper presents the primary structure of the region upstream of the coding sequence of the human α^+-isoform catalytic subunit gene, which has been shown to be expressed in brain tissue [2].

The clone RC10-16 was isolated from a human genomic library in the cosmid pHC79 [6] after screening with a pig kidney Na^+,K^+-ATPase α-subunit hybridization probe [7]. The restriction map of the RC10-16 clone is presented in fig.1. Analysis of the restriction map of this clone and the results of sequencing demonstrate that it contains a gene coding for the α^+-isoform of the Na^+,K^+-ATPase catalytic subunit and partially

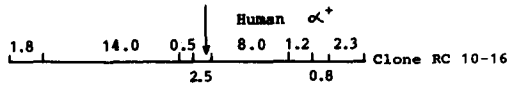

Fig.1. *Eco*RI restriction map of cosmid RC10-16. Sizes of fragments are given in kb. The arrow indicates the position of the ATG codon.

Correspondence address: E.D. Sverdlov, Shemyakin Institute of Bioorganic Chemistry, USSR Academy of Sciences, 117871 Moscow, USSR

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession no.Y07494

Published by Elsevier Science Publishers B.V. (Biomedical Division)

overlaps with the clone λNKαRD-16 found earlier in the human genomic library in λEMBL3 vector [7].

This paper describes the nucleotide sequence of the 1.4 kb fragment lying upstream from the inner *Xba*I site of the 2.5 kb *Eco*RI fragment, containing the ATG initiation codon.

2. MATERIALS AND METHODS

The nucleotide sequence was determined using the dideoxy method of Sanger et al. [8] as modified by McGraw [9] and Henikoff [10]. This method allows one to determine the nonrandom primary structure of large DNA fragments.

3. RESULTS AND DISCUSSION

The sequence in fig.2 includes the first exon of the gene with all of the 5′-untranslatable region of mRNA as is evident from a comparison with the sequence of mRNA coding for rat α^+-subunit [11], and a dozen nucleotides, coding for the first four amino acids of the hypothetic α^+-subunit precursor. The coding sequence is followed by the first intron, over 3500 bp long. The putative site of initiation of transcription is pyrimidine-rich and situated 100 bp upstream from the ATG initiation codon. The total sequenced fragment covers about 1400 bp and in analogy with other genes [5] possibly contains the regulatory elements. 30 bp upstream from the pyrimidine-rich cluster, an imperfect TATA box is localized (fig.2) of which the structure TATTTAAAG closely resembles that of the corresponding regions of some other genes coding for mouse kallikrein, human α-interferon, Raus sarcoma virus and AAV2 virus RNA (review [12]). The homology of the sequenced fragment of the α^+ gene with the regulatory region of the kallikrein gene is not only restricted by the TATA box (fig.3). In particular, there is an octanucleotide repeat GGGGGAGA upstream of the TATA box in the α^+-subunit gene. The same sequence GGGGGAGA is also found in a similar position of the kallikrein gene. A GC-rich region is also situated in the vicinity of the TATA box in the rabbit Ca^{2+}-ATPase gene, which has been sequenced recently [13]. The heptanucleotide GGGGGAG repeated twice is localized there (fig.3). A highly homologous sequence also repeated twice in the opposite orientation was found in a similar position of a gene coding for human serine protease from bone marrow [14]. The role of this element is not yet clear, but it is tempting to speculate that there is some similarity in the mechanisms of expression of these genes.

In addition, it should be noted that the closely homologous sequence in the opposite orientation

Fig.2. Nucleotide sequence of a putative regulatory region of α^+ gene. Sequences of the SP-1-binding site, putative CCAAT site and TATA box are boxed. The 5′-untranslatable region of mRNA, as is evident from comparison with the sequence of rat α^+-subunit cDNA, is underlined.

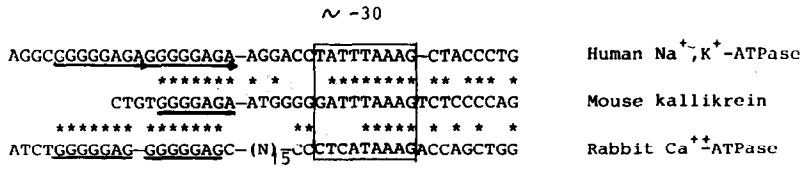

Fig.3. Comparison of sequences preceding the transcription initiation site of genes coding for the α^+-subunit of human Na^+,K^+-ATPase, mouse kallikrein [12] and rabbit muscle Ca^{2+}-ATPase [13]. TATA sequences are boxed. Repeated heptanucleotide is underlined. Homologous nucleotides are denoted by asterisks.

(so called CT motif) was found in the SV-40 enhancer [15], which was shown to be important for enhancer activity. Finally, a homologous motif of polyoma virus enhancer exists as the palindrome ACTGCCCTCCAGAGGGCAGT, parts of which interact with cell proteins, as shown by chemical modification [16]. This preliminary analysis shows the nonrandom character of the sequence and poses the question of its function.

In the −80 region, where the conserved sequence CCAAT has been established for a number of eukaryotic promoters, we have found the sequence CAACAAAC which may have the functions of a CCAAT box. This sequence is flanked by the tetranucleotide GTTT at both 5′- and 3′-termini. In the −120 region is the sequence, GGGGCGG, that interacts with SP1 factor [5]. Another structural peculiarity of the region upstream of the transcription initiation site is the large number of oligo(G) blocks, varying in size from 3 to 7 bp and localized in an interval of ~(−30)–(−120).

More detailed analysis of the region under discussion will be published elsewhere.

Acknowledgement: The authors are grateful to Dr R. Allikmets for construction of the genomic library.

REFERENCES

[1] Glynn, I.M. (1985) in: The Enzymes of Biological Membranes (Martonosi, A.N. ed.) pp.35–114, Plenum, New York.
[2] Rossier, B.C., Geering, K. and Kraehenbuhl, J.P. (1987) Trends Biochem. Sci. 12, 483–487.
[3] Sverdlov, E.D., Monastyrskaya, G.S., Broude, N.E., Ushkaryov, Yu.A., Allikmets, R.L., Melkov, A.M., Smirnov, Yu.V., Malyshev, I.V., Dulubova, I.E., Petrukhin, K.E., Grishin, A.V., Kiyatkin, N.I., Kostina, M.B., Sverdlov, V.E., Modyanov, N.N. and Ovchinnikov, Yu.A. (1987) FEBS Lett. 217, 275–278.
[4] Sverdlov, E.D., Broude, N.E., Sverdlov, V.E., Monastyrskaya, G.S., Grishin, A.V., Petrukhin, K.E., Akopyanz, N.S., Modyanov, N.N. and Ovchinnikov, Yu.A. (1987) FEBS Lett. 221, 129–133.
[5] Maniatis, T., Goodbourn, S. and Fisher, J.A. (1987) Science 236, 1237–1245.
[6] Hohn, B. and Collins, J. (1980) Gene 11, 291–298.
[7] Ovchinnikov, Yu.A., Monastyrskaya, G.S., Broude, N.E., Allikmets, R.L., Ushkaryov, Yu.A., Melkov, A.M., Smirnov, Yu.V., Malyshev, I.V., Dulubova, I.E., Petrukhin, K.E., Grishin, A.V., Sverdlov, V.E., Kiyatkin, N.I., Kostina, M.B., Modyanov, N.N. and Sverdlov, E.D. (1987) FEBS Lett. 213, 73–80.
[8] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[9] McGraw, R.A. (1984) Anal. Biochem. 143, 298–303.
[10] Henikoff, S. (1984) Gene 28, 351–359.
[11] Shull, G.E., Greeb, J. and Lingrel, J.B. (1986) Biochemistry 25, 8125–8132.
[12] Bucher, P. and Trifonov, E.N. (1986) Nucleic Acids Res. 14, 10009–10026.
[13] Korzak, B., Zarain-Herzberg, A., Brandl, C.J., Ingles, C.J., Green, N.M. and MacLennan, D.H. (1988) J. Biol. Chem. 263, 4813–4819.
[14] Nakamura, H., Okano, K., Aoki, Y., Shimizu, H. and Naruto, M. (1987) Nucleic Acids Res. 15, 9601–9602.
[15] Zenke, M., Grundstrom, T., Matthes, H., Wintzerith, M., Schatz, C., Wildeman, A. and Chambon, P. (1986) EMBO J. 5, 387–397.
[16] Piette, J. and Yaniv, M. (1986) Nucleic Acids Res. 14, 9595–9611.