Volume 233, number 1, 87–94 FEB 05920 June 1988

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

Structure of the gene for the catalytic subunit (αIII-form) and its relationship with structural features of the protein

Yu.A. Ovchinnikov†, G.S. Monastyrskaya, N.E. Broude, Yu.A. Ushkaryov, A.M. Melkov, Yu.V. Smirnov, I.V. Malyshev, R.L. Allikmets, M.B. Kostina, I.E. Dulubova, N.I. Kiyatkin, A.V. Grishin, N.N. Modyanov and E.D. Sverdlov

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

Received 1 April 1988

The primary structure of a gene of the Na^+,K^+-ATPase multigenic family in the human genome has been determined. The gene corresponds to a hypothetical αIII-form of the enzyme catalytic subunit. The gene comprises over 25000 bp, and its protein coding region includes 23 exons and 22 introns. Possible correlation between structural features of the protein and location of introns in the gene are discussed.

Na^+,K^+-ATPase; Catalytic subunit; Gene sequence; Exon/intron structure; Protein spatial organization

1. INTRODUCTION

Integral membrane protein Na^+,K^+-ATPase performs the active transport of Na^+ and K^+ across the cellular membranes. We determined complete primary structures of cDNA of both (α and β) enzyme subunits from pig kidney [1–3]. While studying the structure of the human genome locus encoding the enzyme, we discovered the family of related genes coding for the proteins homologous to the enzyme catalytic subunit [4,5] and showed a tissue-specific character of regulation of the expression level of these genes [6]. The paper presents the primary structure of one of the genes of this family that correlates to αIII-form according to the nomenclature proposed in [7]. Parts of the present data have been published earlier [4,5,8,9].

Correspondence address: E.D. Sverdlov, Shemyakin Institute of Bioorganic Chemistry, USSR Academy of Sciences, ul. Miklukho-Maklaya 16/10, 117871 GSP Moscow V-437, USSR

† Deceased

2. MATERIALS AND METHODS

Recombinant phages λNKαR10-3 and λNKαR3-2 (fig.1) were identified and analysed as described [4]. A new genomic library was constructed to find out the N-terminal part of the gene. DNA was isolated from human brain as in [10]. After partial cleavage of the genomic DNA with *Sau*3A endonuclease and centrifugation in the sucrose density gradient, the DNA fragments ranging between 10000 and 20000 bp were cloned into the *Bam*HI site of vector λEMBL3 [11]. The library (3×10^6 independent clones) was screened by hybridization to the exon fragment (coordinates 361–441) from the genomic insert of λNKαR3-2 recloned into phage M13.

Restriction mapping of the positive clones revealed that the genomic insert of λNKαRM1 overlapped with that of λNKαR3-2 [4]. Random fragments of the λNKαRM1 obtained by sonication were cloned into phage M13mp18 and sequenced according to Sanger [12] in both directions. Determination of the primary structure confirmed the complete correlation of nucleotide sequences of the λNKαRM1 and λNKαR3-2 inserts in their overlapping region. The human brain cDNA library in pBR322 was screened by hybridization to the probes: (i) nick-translated *Pst*I fragment of plasmid pB2801 (which codes for

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

the pig kidney enzyme α-subunit with coordinates 1545–2880) [3,4]; (ii) synthetic oligonucleotide corresponding to coordinates 7–31 of the studied gene.

3. RESULTS AND DISCUSSION

The physical map of the gene of the human Na^+,K^+-ATPase catalytic subunit is given in fig.1; its sequence in fig.2.

The gene comprises more than 25000 bp in length and its protein-coding region consists of 23 exons and 22 introns. The primary structure of exons 2–4 and 14–23 is completely identical to the sequences of cDNA fragments isolated from the human brain cDNA library. Comparison of the sequences of the gene and cDNA shows that the intron preceding exon 2 separates it from the 5′-untranslated region and 6 nucleotides encoding the first two amino acid residues (Met-Gly) of the protein-precursor. Since we have not identified the genomic sequence corresponding to this cDNA part, the exact size of exon 1 is unknown. It is noteworthy, in the gene sequence established by us that codes for the $\alpha+$-form of the Na^+,K^+-ATPase catalytic subunit exon 1 includes the extended part of the 5′-untranslated region and first 4 codons of the protein-precursor [9]. Except for exon 1 and exon 23, which includes 26 protein coding base pairs, the stop codon (TGA) and the entire 3′-untranslated region (548 bp), all other exons contain only protein-coding sequences ranging between 60 and 269 bp with a medium size of 143 bp. This accords well with the Blake correlation [13] establishing 2–3 exons per 100 amino acid residues on an average. The intron lengths range between 70 and 5000 bp. Among the introns, 12 occur between codons, 6 interrupt at frame 1, and 4 at frame 2. Of other gene characteristics, the presence of a polyadenylation noncanonic site AATATA located 15–17 nucleotides upstream from the poly(A) attachment site (see legend to fig.2) deserves special attention. The site of the similar structure was observed in one of the SV-40 mutants with the markedly decreased ability of polyadenylation of mRNA [14]. Several Alu repeats were revealed in intron regions. The detailed analysis of these and other peculiarities will be published elsewhere. Table 1 shows consensus sequences of 5′- and 3′-splice sites determined for αIII-form of the human enzyme gene. The consensus derived coincides with that described in literature [15].

The sequences we established for the gene and cDNA are highly homologous to cDNA of αIII-form of the catalytic subunit from rat brain [7]. Homology between the derived amino acid sequences is even more evident, while homology of the primary structure of the gene and cDNAs corresponding to other isoforms of the rat α-subunit is less. Therefore the sequence published most probably represents the gene coding for αIII-form of the human Na^+,K^+-ATPase catalytic subunit. Though the sequence of the corresponding mRNA possesses necessary translation start and stop signals, it is still unclear whether this protein exists in the cell. If this is the case, it may be a new form of the Na^+,K^+-ATPase catalytic subunit. One cannot exclude the possibility that this subunit principally belongs to another ion-transporting ATPase.

As mentioned previously, expression of the studied gene is tissue-specific and more intense in human kidney and brain [6]. (The same paper indicates the differences in hybridization patterns of Southern blots of DNA isolated from different tissues. Further analysis showed that this phenomenon is probably not universal, and its origin remains obscure.)

The derived amino acid sequence of the primary product of the gene translation consists of 1013 amino acid residues (molecular mass 111688 Da). Its comparison with the known structures of analogous subunits of other Na^+,K^+-ATPases unravels the following specific features. It is ten residues shorter than the precursor of the catalytic subunit of the enzyme (α-form) from HeLa cells [16]. The total number of amino acid distinctions in these proteins is 139, of them 32 are concentrated in the N-terminus, the most variable part of E_1E_2 ion-transporting ATPases (residues 1–59), the sequence of the first 11 residues completely differs from the corresponding regions of other forms. The second cluster of variability (residues 478–487) directly precedes the lysine residue modified with fluorescein isothiocyanate [17,18] and probably pertains to the ATP-hydrolysing site. The homology degree of the remaining regions of the compared polypeptides is ~90%.

The number of the differences between the described structure and rat αIII sequence [7,19] is

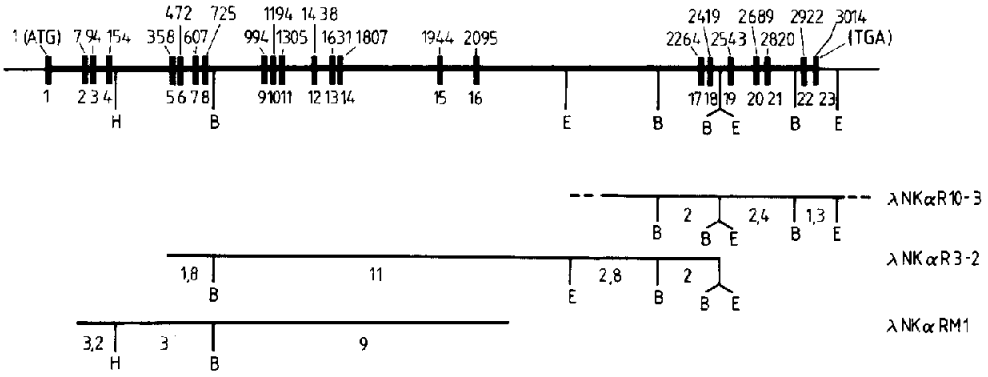

Fig.1. Structural organization of the gene for αIII-form of the human Na^+,K^+-ATPase catalytic subunit. Exons 1–23 are designated by vertical bars. The figures above indicate the ordinal number of the first nucleotide in exons. Restriction maps of overlapping clones are given below, figures indicate fragment lengths in kb. E, *Eco*RI; B, *Bam*HI; H, *Hin*dIII.

10. This demonstrates that the gene coding for the αIII protein appeared before the divergence of the primate and rodent lineages during evolution.

The possible relation of the exon-intron organization of the gene and structure-functional features of the Na^+,K^+-ATPase catalytic subunit can be considered in the light of the hypothesis on correspondence between the exons and functional or structural protein domains [20–22].

At present, the structures of several animal ion-transporting ATPases of the E_1E_2 type – Na^+,K^+-ATPases [1,3,7,9,16,19,23,24], H^+,K^+-ATPase [25], and Ca^{2+}-ATPases [26] – are defined by analysis of cDNA nucleotide sequences. Comparison of the known primary structures including that described above revealed several structural regions highly conservative among all these related enzymes. These regions designated in fig.2 have the following coordinates: 168–195 (exon 6), 200–210 (exon 7), 223–282 (exons 7–8), 313–379 (exons 8–9), 437–444 (exon 11), 495–524 (exon 12), 574–635 (exons 13–14), 678–775 (exons 15–17), 793–828 (exons 17–18) and 971–998 (exon 22). Except for the two latter portions, others are in hydrophilic regions linking membrane segments II and III, IV and V. According to the structural model of the Na^+,K^+-ATPase catalytic subunit they are located in the cytoplasmic region of the enzyme [27,28]. Homologous portions 313–379, 495–524 and 678–775 include all the known components of the enzyme ATP-hydrolyzing site which are Asp-364, Lys-496, Asp-705, Asp-709 and Lys-714 in this sequence. Here the homology extent is maximal. Undoubtedly, all the homologous regions (among them those not ascribed to any function yet) form the spatially interrelated structural basis of E_1E_2-ATPases activity [28].

Positions of introns 5,6,9–12,14,21,22 coincide with the boundaries of a series of the above mentioned homologous regions of ion-transporting ATPases. However, other introns (7,8,13,15–17) appear to fall into the conservative sequences.

Comparison of the structures of the catalytic subunits of various Na^+,K^+-ATPases shows that position of intron 3 correlates with the boundary of the most variable N-terminal region encoded by the first exons. In the same way, introns 9 and 10 clearly mark the variable part of the structures coded by exon 10. A number of clusters of non-conservative residues are localized at the beginning of exons 12,13,15. On the contrary, in case of exon 19 the variable part is in its middle.

Positions of the identified introns, except for intron 1 [9], in different genes of the multigenic family coincide [5]. This fact implies rather similar intron-exon structural organization for all the known genes in the multigenic family including those for Na^+,K^+- and H^+,K^+-ATPases [5,29] and demonstrates their origin to be the common primordial gene.

Certain relations are traced when comparing the exon-intron structure of the analysed gene and the model for the transmembrane organization of the α-form of the Na^+,K^+-ATPase catalytic subunit based upon the data of limited proteolysis [27,28], immunochemical analysis [30] and differentiated

```
        AACGGACGGACGGACGGACGGCGCACCTACCGACGGCGCGGGCGCTGCAGAGGCTCCCAGCCCAAGCCTGAGCCTGAGCCCGCCCCGAGGTCCCCGCCCCG

1-18    CCCGCCTGGCTCTCTCGCCGGAGCCGCCAAG ATG GGG.....intron 1.....ATCGCCACCTCGCAG GAC AAG AAA GAT
                                        Met-Gly                                  Asp-Lys-Lys-Asp-

19-93   GAC AAG GAC TCA CCC AAG AAG AAC AAG GGC AAG GAG CGC CGG GAC CTG GAT GAC CTC AAG AAG GAG GTG GCT ATG
        Asp-Lys-Asp-Ser-Pro-Lys-Lys-Asn-Lys-Gly-Lys-Glu-Arg-Arg-Asp-Leu-Asp-Asp-Leu-Lys-Lys-Glu-Val-Ala-Met

94-117  GTAAGCCTCGCCCTC.....intron 2 (99b.p.).....ACTCCTGTTCCTCAG ACA GAG CAC AAG ATG TCA GTG GAA
                                                                 Thr-Glu-His-Lys-Met-Ser-Val-Glu-

118-153 GAG GTC TGC CGG AAA TAC AAC ACA GAC TGT GTG CAG GTGTGGCCAGGCTGT.....intron 3.....GTCCTCTGTC
        Glu-Val-Cys-Arg-Lys-Tyr-Asn-Thr-Asp-Cys-Val-Gln

154-225 CCTAG GGT TTG ACC CAC AGC AAA GCC CAG GAG ATC CTG GCC CGG GAT GGG CCT AAC GCA CTC ACG CCA CCG CCT ACC
              Gly-Leu-Thr-His-Ser-Lys-Ala-Gln-Glu-Ile-Leu-Ala-Arg-Asp-Gly-Pro-Asn-Ala-Leu-Thr-Pro-Pro-Pro-Thr-

226-300 ACC CCA GAG TGG GTC AAG TTT TGC CGG CAG CTC TTC GGG GGC TTC TCC ATC CTG CTG TGG ATC GGG GCT ATC CTC
        Thr-Pro-Glu-Trp-Val-Lys-Phe-Cys-Arg-Gln-Leu-Phe-Gly-Gly-Phe-Ser-Ile-Leu-Leu-Trp-Ile-Gly-Ala-Ile-Leu-

301-357 TGC TTC CTG GCC TAC GGT ATC CAG GCG GGC ACCGAG GAC GAC CCC TCT GGT GAC AAC GTGAGTGCCTGGACC.....
        Cys-Phe-Leu-Ala-Tyr-Gly-Ile-Gln-Ala-Gly-ThrGlu-Asp-Asp-Pro-Ser-Gly-Asp-Asn

358-405 ...intron 4.....CACATCTCCCCACAG CTG TAC CTG GGC ATC GTC CTG GCG GCC GTG GTG ATC ATC ACT GGC TGC
                                     Leu-Tyr-Leu-Gly-Ile-Val-Leu-Ala-Ala-Val-Val-Ile-Ile-Thr-Gly-Cys-

406-471 TTC TCC TAC TAC CAG GAG GCC AAG AGC TCC AAG ATC ATG GAG TCC TTC AAG AAC ATG GTG CCC CAG GTGAAGGGTGCC
        Phe-Ser-Tyr-Tyr-Gln-Glu-Ala-Lys-Ser-Ser-Lys-Ile-Met-Glu-Ser-Phe-Lys-Asn-Met-Val-Pro-Gln

472-504 CAG.....intron 5 (113b.p.).....TCCATCCCACCTCAG CAA GCC CTG GTG ATC CGG GAA GGT GAG AAG ATG
                                                    Gln-Ala-Leu-Val-Ile-Arg-Glu-Gly-Glu-Lys-Met-

505-579 CAG GTG AAC GCT GAG GAG GTG GTG GTC GGG GAC CTG GTG GAG ATC AAG GGT GGA GAC CGA GTG CCA GCT GAC CTG
        Gln-Val-Asn-Ala-Glu-Glu-Val-Val-Val-Gly-Asp-Leu-Val-Glu-Ile-Lys-Gly-Gly-Asp-Arg-Val-Pro-Ala-Asp-Leu-

580-606 CGG ATC ATC TCA GCC CAC GGC TGC AAG GTGGGCCTGTAGGGC.....intron 6 (433b.p.).....TGCACCTACCC
        Arg-Ile-Ile-Ser-Ala-His-Gly-Cys-Lys

607-678 CCAG GTG GAC AAC TCC TCC CTG ACT GGC GAA TCC GAG CCC CAG ACT CGC TCT CCC GAC TGC ACG CAC GAC AAC CCC
              Val-Asp-Asn-Ser-Ser-Leu-Thr-Gly-Glu-Ser-Glu-Pro-Gln-Thr-Arg-Ser-Pro-Asp-Cys-Thr-His-Asp-Asn-Pro-

679-724 TTG GAG ACT CGG AAC ATC ACC TTC TTT TCC ACC AAC TGT GTG GAA G GTGAGGCGGGTGCAG.....intron 7 (115 b.p.)
        Leu-Glu-Thr-Arg-Asn-Ile-Thr-Phe-Phe-Ser-Thr-Asn-Cys-Val-Glu-G(ly)

725-780 .....CCCTCTGGCCTGCAG GC ACG GCT CGG GGC GTG GTG GTG GCC ACG GGC GAC CGC ACT GTC ATG GGC CGT ATC
                              (G)ly-Thr-Ala-Arg-Gly-Val-Val-Val-Ala-Thr-Gly-Asp-Arg-Thr-Val-Met-Gly-Arg-Ile-

781-855 GCC ACC CTG GCA TCA GGG CTG GAG GTG GGC AAG ACG CCC ATC GCC ATC GAG ATT GAG CAC TTC ATC CAG CTC ATC
        Ala-Thr-Leu-Ala-Ser-Gly-Leu-Glu-Val-Gly-Lys-Thr-Pro-Ile-Ala-Ile-Glu-Ile-Glu-His-Phe-Ile-Gln-Leu-Ile-

856-930 ACC GGC GTG GCT GTC TTC CTG GGT GTC TCC TTC TTC ATC CTC TCC CTC ATT CTC GGA TAC ACC TGG CTT GAG GCT
        Thr-Gly-Val-Ala-Val-Phe-Leu-Gly-Val-Ser-Phe-Phe-Ile-Leu-Ser-Leu-Ile-Leu-Gly-Tyr-Thr-Trp-Leu-Glu-Ala-

931-993 GTC ATC TTC CTC ATC GGC ATC ATC GTG GCC AAT GTC CCA GAG GGT CTG CTG GCC ACT GTC ACT GTAAGGCCAGGCTCC....
        Val-Ile-Phe-Leu-Ile-Gly-Ile-Ile-Val-Ala-Asn-Val-Pro-Glu-Gly-Leu-Leu-Ala-Thr-Val-Thr

994-1038 ........intron 8.....TTGCTCGTCCTCCAG GTG TGT CTG ACC GTG ACC GCC AAG CGC ATG GCC CGG AAG AAC TGC
                                              Val-Cys-Leu-Thr-Val-Thr-Ala-Lys-Arg-Met-Ala-Arg-Lys-Asn-Cys-

1039-1113 CTG GTG AAG AAC CTG GAG GCT GTA GAG ACC CTG GGC TCC ACG TCC ACC ATC TGC TCA GAT AAG ACA GGG ACC CTC
          Leu-Val-Lys-Asn-Leu-Glu-Ala-Val-Glu-Thr-Leu-Gly-Ser-Thr-Ser-Thr-Ile-Cys-Ser-Asp-Lys-Thr-Gly-Thr-Leu-

1114-1188 ACT CAG AAC CGC ATG ACA GTC GCC CAC ATG TGG TTT GAC AAC CAG ATC CAC GAG GCT GAC ACC ACT GAG GAC CAG
          Thr-Gln-Asn-Arg-Met-Thr-Val-Ala-His-Met-Trp-Phe-Asp-Asn-Gln-Ile-His-Glu-Ala-Asp-Thr-Thr-Glu-Asp-Gln-

1189-1212 TCA G GTGAGCGCAGGCCCC.....intron 9 (74b.p.).....ACATGCCTCCCCCAG GG ACC TCA TTT GAC AAG AGT
          Ser-G(ly)                                                    (G)ly-Thr-Ser-Phe-Asp-Lys-Ser-

1213-1287 TCG CAC ACC TGG GTG GCC CTG TCT CAC ATC GCT GGG CTC TGC AAT CGC GCT GTC TTC AAG GGT GGT CAG GAC AAC
          Ser-His-Thr-Trp-Val-Ala-Leu-Ser-His-Ile-Ala-Gly-Leu-Cys-Asn-Arg-Ala-Val-Phe-Lys-Gly-Gly-Gln-Asp-Asn-

1288-1311 ATC CCT GTG CTC AAG AG GTGGGTTAGCTACTG.....intron 10 (84b.p.).....CCTTTCTCCTTGCAG G GAT GTG
          Ile-Pro-Val-Leu-Lys-Ar(g)                                                  (Ar)g-Asp-Val-

1312-1386 GCT GGG GAT GCG TCT GAG TCT GCC CTG CTC AAG TGC ATC GAG CTG TCC TCT GGC TCC GTG AAG CTG ATG CGT GAA
          Ala-Gly-Asp-Ala-Ser-Glu-Ser-Ala-Leu-Leu-Lys-Cys-Ile-Glu-Leu-Ser-Ser-Gly-Ser-Val-Lys-Leu-Met-Arg-Glu-

1387-1437 CGA AAC AAG AAA GTG GCT GAG ATT CCC TTC AAT TCC ACC AAC AAA TAC CAG GTACTCTGGCTTTCC.....intron 11..
          Arg-Asn-Lys-Lys-Val-Ala-Glu-Ile-Pro-Phe-Asn-Ser-Thr-Asn-Lys-Tyr-Gln

1438-1497 ........CTGATCGGTCCCCAG CTC TCC ATC CAT GAG ACC GAG GAC CCC AAC GAC AAC CGA TAC CTG CTG GTG ATG AAG GGT
                                  Leu-Ser-Ile-His-Glu-Thr-Glu-Asp-Pro-Asn-Asp-Asn-Arg-Tyr-Leu-Leu-Val-Met-Lys-Gly-

1498-1572 GCC CCC GAG CGC ATC CTG GAC CGC TGC TCC ACC ATC CTC CTA CAG GGC AAG GAG CAG CCT CTG GAC GAG GAA ATG
          Ala-Pro-Glu-Arg-Ile-Leu-Asp-Arg-Cys-Ser-Thr-Ile-Leu-Leu-Gln-Gly-Lys-Glu-Gln-Pro-Leu-Asp-Glu-Glu-Met-

1573-1630 AAG GAG GCC TTT CAG AAT GCC TAC CTT GAG CTC GGT GGC CTG GGC GAG CGC GTG CTT G GTGCGAGGTGC.....
          Lys-Glu-Ala-Phe-Gln-Asn-Ala-Tyr-Leu-Glu-Leu-Gly-Gly-Leu-Gly-Glu-Arg-Val-Leu-G(ly)

1631-1677 ..intron 12.....CTTCCCTGCCACTAG GT TTC TGC CAT TAT TAC CTG CCC GAG GAG CAG TAT CCC CAA GGC TTT
                                    (G)ly-Phe-Cys-His-Tyr-Tyr-Leu-Pro-Glu-Glu-Gln-Tyr-Pro-Gln-Gly-Phe-

1678-1752 GCC TTC GAC TGT GAT GAC GTG AAC TTC ACC ACG GAC AAC CTC TGC TTT GTG GGC CTC ATG TCC ATG ATC GGC CCA
          Ala-Phe-Asp-Cys-Asp-Asp-Val-Asn-Phe-Thr-Thr-Asp-Asn-Leu-Cys-Phe-Val-Gly-Leu-Met-Ser-Met-Ile-Gly-Pro-

1753-1806 CCC CGC GCA GCC GTC CCT GAC GCG GTG GGC AAG TCT CGC AGC GCA GGC ATC AAG GTGTGGCTTGGGTGC.....intron
          Pro-Arg-Ala-Ala-Val-Pro-Asp-Ala-Val-Gly-Lys-Cys-Arg-Ser-Ala-Gly-Ile-Lys

1807-1854 13 (70b.p.).....AAACAATGCCTGCAG GTC ATC ATG GTC ACC GGC GAT CAC CCC ATC ACG GCC AAG GCC ATT GCC
                                          Val-Ile-Met-Val-Thr-Gly-Asp-His-Pro-Ile-Thr-Ala-Lys-Ala-Ile-Ala-

1855-1929 AAG GGT GTG GGC ATC ATC TCT GAG GGC AAC GAG ACT GTG GAG GAC ATC GCC GCC CGG CTC AAC ATT CCC GTC AGC
          Lys-Gly-Val-Gly-Ile-Ile-Ser-Glu-Gly-Asn-Glu-Thr-Val-Glu-Asp-Ile-Ala-Ala-Arg-Leu-Asn-Ile-Pro-Val-Ser-

1930-1959 CAG GTT AAC CCC CG GTGAGCCACCCATTC.....intron 14.....CTCTGCTCTTCCCAG G GAT GCC AAG GCC TGC
          Gln-Val-Asn-Pro-Ar(g)                                                  (Ar)g-Asp-Ala-Lys-Ala-Cys-
```

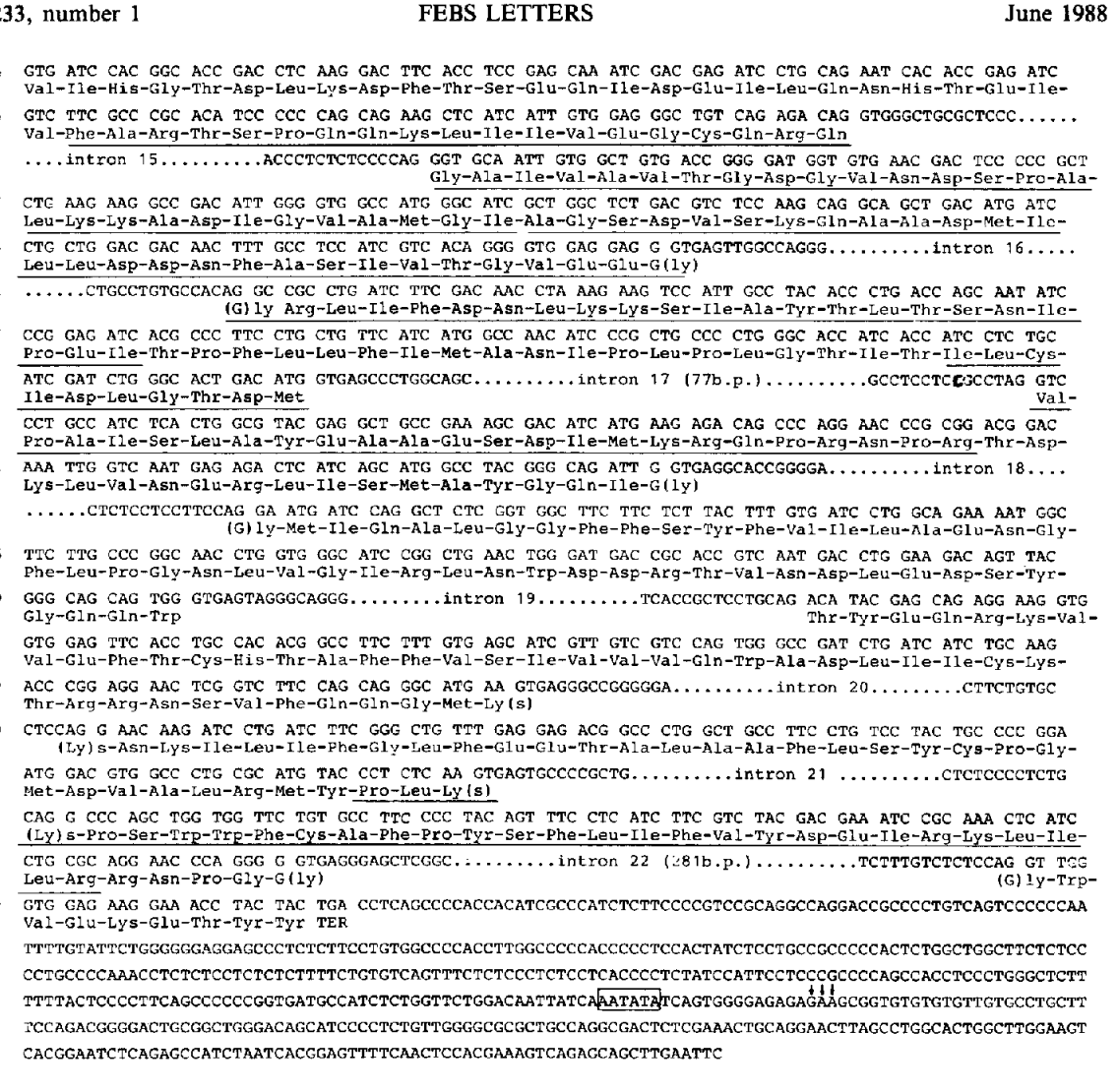

Fig.2. Nucleotide sequence and derived amino acid sequence of αIII-form of the human Na^+,K^+-ATPase catalytic subunit. Lengths of completed introns are in parentheses. On the left, numeration of nucleotides of the coding gene portion. The nucleotide sequence of 5′-untranslated region and six nucleotides coding for the first two amino acids are shown on cDNA structure. Exon sequences 2–4 and 14–23 of the gene are identical to sequences of cDNA independently determined. The 3′-untranslated region is completed with the poly(A) tail attached to the site indicated with an arrow at G. Since AA follows this G in the genome one cannot exclude that cleavage of the RNA-precursor occurs after any A. The regions of the amino acid sequence homologous to all the known ion-transporting ATPases are underlined. The proposed polyadenylation consensus sequence is boxed.

analysis of the secondary structure of hydrophilic and hydrophobic regions by Raman spectroscopy combined with predictional calculations [31] (fig.3). This comparison is correct due to the high homology of the primary structures and identity of the calculated secondary structures of variable portions of the α and αIII polypeptides. The site of intron 1 falls into the polypeptide region, that is probably cleaved off during protein maturation as in the case of $\alpha+$-form [9] and therefore not shown in fig.3.

It should be noted that as a rule introns are situated between the regions of the regular structure or at their boundaries; however, e.g. introns 11, 12 are in the β-structural stretch and intron 18 is localized within the membrane α-helical rod V. Exon 8 encodes the hairpin structure including the III and IV transmembrane rods, other membrane

Table 1

Consensus sequences of 5′- and 3′-splice sites in Na^+,K^+-ATPases gene (general consensus is from [15])

	5′-splice site											
	Exon				Intron							
G (%)	24	14	14	81	100	0	86	14	90	38	33	33
A (%)	14	38	57	9	0	0	14	67	5	5	5	19
U (%)	19	14	19	5	0	100	0	9	5	24	19	14
C (%)	43	33	9	5	0	0	0	9	0	33	43	33
Consensus	–	–	A	G	G	U	G	A	G	–	–	–
General consensus	–	–	A	G	G	U	A	A	G	U	–	–

	3′-splice site																				
	Intron																		Exon		
G (%)	9	23	14	8	4	14	4	8	14	23	27	8	4	0	27	0	0	100	77	23	14
A (%)	23	27	14	23	8	14	18	4	4	4	4	4	0	4	9	0	100	0	9	18	41
U (%)	23	9	27	23	46	23	36	41	14	55	23	32	18	46	9	14	0	0	0	36	18
C (%)	45	41	46	46	41	50	41	46	68	18	46	56	78	50	55	86	0	0	14	23	27
Consensus	–	–	Y	Y	Y	C	Y	Y	C	U	Y	C	C	Y	C	C	A	G	G	–	–
General consensus	–	–	–	–	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	C	A	G	G	–	–

segments are coded by exons 4,5,17,21 each. Positions of some introns (4,8,17,20,21), which mark boundaries of these exons, coincide with those of hydrophobic sequences. The analogous situation was first observed for the visual rhodopsin gene [32].

The preferable disposition of introns in the gene moiety encoding the protein surface regions is proposed in [27]. This is also true for the catalytic subunit of Na^+,K^+-ATPase: by immunochemical methods some regions of its polypeptide chain are identified on the protein molecule surface

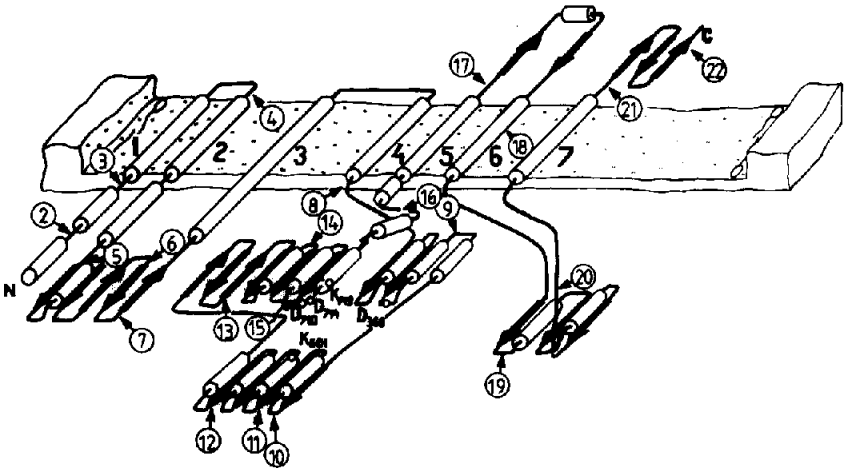

Fig.3. Model for spatial organization of Na^+,K^+-ATPase catalytic subunit (α-form) [31]. Cylinders and arrows correspond to α-helical segments and β-pleated sheets, respectively. Encircled numbers indicate intron positions in αIII-form [9]. The amino acid residues within the enzyme active sites are specially indicated [28].

[27,28,30,34] and introns 2,4,7,19,22 are located in the corresponding gene parts. Besides, introns 2,7 and 10 coincide well with characteristic points of primary limited proteolytic attacks of the native enzyme [35].

Numerous attempts have been undertaken to unravel the regularities underlying the structure of the nucleotide binding sites of different enzymes (review [36]). As mentioned above, in the animal E_1E_2-ATPases this site is evidently formed by conservative regions of the polypeptide chain of the catalytic subunits in the cytoplasmic portion of the enzyme molecules. Analysis of these regions for the presence of permissive consensus sequences [37] and characteristic elements of the secondary structure [38] may evidence that within the nucleotide-binding site of ion-transporting ATPases the β-α structure (residues 240–265) encoded by the initial part of exon 8 is present in addition to the structure coded by the initial part of exon 16 – a target for affinity modification with the ATP analogue [28,39–41]. The data concerning interrelations of intron positions and structural elements of the proposed nucleotide binding domains require thorough analysis and will be published. The study demonstrates certain correlations between structure-functional features of the protein and the exon-intron gene structure. Detailed knowledge of the Na^+,K^+-ATPase three-dimensional structure is necessary to prove whether the relationship exists between protein domains and gene exons.

REFERENCES

[1] Ovchinnikov, Yu.A., Arsenyan, S.G., Broude, N.E., Petrukhin, K.E., Grishin, A.V., Aldanova, N.A., Arzamazova, N.M., Arystarkhova, E.A., Melkov, A.M., Smirnov, Yu.V., Guryev, S.O., Monastyrskaya, G.S. and Modyanov, N.N. (1985) Dokl. Akad. Nauk SSSR 285, 1490–1495.

[2] Ovchinnikov, Yu.A., Broude, N.E., Petrukhin, K.E., Grishin, A.V., Kiyatkin, N.I., Arzamazova, N.M., Gevondyan, E.N., Melkov, A.M., Smirnov, Yu.V., Malyshev, I.V., Monastyrskaya, G.S. and Modyanov, N.N. (1986) Dokl. Akad. Nauk SSSR 287, 1491–1496.

[3] Ovchinnikov, Yu.A., Modyanov, N.N., Petrukhin, N.E., Broude, N.E., Grishin, A.V., Arzamazova, N.M., Aldanova, N.A., Monastyrskaya, G.S. and Sverdlov, E.D. (1986) FEBS Lett. 201, 237–245.

[4] 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.

[5] 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.

[6] 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.

[7] Shull, G.E., Greeb, J. and Lingrell, J.B. (1986) Biochemistry 25, 8125–8132.

[8] Ovchinnikov, Yu.A., Monastyrskaya, G.S., Broude, N.E., Ushkaryov, Yu.A., Dolganov, G.M., Melkov, A.M., Smirnov, Yu.V., Akopyantz, N.S., Dulubova, N.E., Allikmets, R.L., Modyanov, N.N. and Sverdlov, E.D. (1986) Dokl. Akad. Nauk SSSR 287, 1251–1255.

[9] Sverdlov, E.D., Monastyrskaya, G.S., Broude, N.E., Ushkaryov, Yu.A., Melkov, A.M., Smirnov, Yu.V., Malyshev, I.V., Allikmets, R.L., Kostina, M.B., Dulubova, I.E., Kiyatkin, N.I., Grishin, A.V., Modyanov, N.N. and Ovchinnikov, Yu.A. (1987) Dokl. Akad. Nauk SSSR 297, 1488–1494.

[10] Blin, N. and Stafford, D.W. (1976) Nucleic Acids Res. 3, 2303–2308.

[11] DNA Cloning. A Practical Approach (Glover, D.M. ed.) pp.1–48, IRL Press, Washington.

[12] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.

[13] Blake, C.C.F. (1983) Nature 306, 535–537.

[14] Wickens, M. and Stephenson, P. (1984) Science 26, 1045–1051.

[15] Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986) Annu. Rev. Biochem. 55, 1119–1150.

[16] Kawakami, K., Ohta, T., Nojima, H. and Nagano, K. (1986) J. Biochem. 100, 389–397.

[17] Farley, R.A., Tran, C.M., Carilli, C.T., Hawke, D. and Shively, J.E. (1984) J. Biol. Chem. 259, 9532–9635.

[18] Kirley, T.L., Wallick, E.T. and Lane, L.K. (1984) Biochem. Biophys. Res. Commun. 125, 767–773.

[19] Hara, Y., Urayama, O., Kawakami, K., Hojima, H., Nagamune, H., Kojima, T., Ohta, T., Nagano, K. and Nakao, M. (1987) J. Biochem. 102, 43–58.

[20] Gilbert, W. (1978) Nature 271, 501.

[21] Blake, C.C.F. (1978) Nature 273, 267.

[22] Go, M. (1981) Nature 291, 90–92.

[23] Shull, G.E., Schwartz, A. and Lingrel, J.B. (1985) Nature 316, 691–695.

[24] Kawakami, K., Nogushi, S., Noda, M., Takahashi, H., Ohta, T., Kawamura, M., Nojima, H., Nagano, K., Hirose, T., Inayama, S., Hayashida, H., Miyata, T. and Numa, S. (1985) Nature 316, 733–736.

[25] Shull, G.E. and Lingrell, J.B. (1986) J. Biol. Chem. 261, 16788–16791.

[26] Brandl, C.J., Green, N.M., Korczak, B. and MacLennan, D.H. (1986) Cell 44, 597–607.
[27] Ovchinnikov, Yu.A., Arzamazova, N.M., Arystarkhova, E.A., Gevondyan, N.M., Aldanova, N.A. and Modyanov, N.N. (1987) FEBS Lett. 217, 269–274.
[28] Ovchinnikov, Yu.A. (1987) Trends Biochem. Sci. 12, 434–438.
[29] Shull, M.M. and Lingrell, J.B. (1987) Proc. Natl. Acad. Sci. USA 84, 4039–4043.
[30] Ovchinnikov, Yu.A., Luneva, N.M., Arystarkhova, E.A., Gevondyan, N.M., Arzamazova, N.M., Kozhich, A.T., Nesmyanov, V.A. and Modyanov, N.N. (1988) FEBS Lett. 227, 230–234.
[31] Ovchinnikov, Yu.A., Arystarkhova, E.A., Arzamazova, N.M., Dzhandzhugazyan, K.N., Efremov, R.G., Nabiev, I.R. and Modyanov, N.N. (1988) FEBS Lett. 227, 235–239.
[32] Nathans, J. and Hogness, D.S. (1984) Proc. Natl. Acad. Sci. USA 81, 4851–4855.
[33] Craik, G.S., Rutler, W.J. and Flefterick, R. (1983) Science 220, 1125–1129.
[32] Ball, W.J. and Loftice, C.D. (1987) Biochim. Biophys. Acta 916, 100–111.
[35] Jorgensen, P.L. and Collins, J.H. (1986) Biochim. Biophys. Acta 860, 570–576.
[36] Fothergill-Gilmore, L.A. (1986) in: Multidomain Proteins – Structure and Evolution (Hardie, D.G. and Coggins, J.R. eds) pp.85–174, Elsevier, Amsterdam.
[37] Walker, J.E., Saraste, M., Ruswick, M.J. and Gay, N.J. (1982) EMBO J. 1, 945–951.
[38] Rossman, M.G., Moras, D. and Olsen, K.W. (1974) Nature 250, 194–199.
[39] Dzhandzhugazyan, K.N., Lutsenko, S.V. and Modyanov, N.N. (1986) Biol. Membr. 3, 858–868.
[40] Ovchinnikov, Yu.A., Dzhandzhugazyan, K.N., Lutsenko, S.V., Mustayev, A.A. and Modyanov, N.N. (1987) FEBS Lett. 217, 111–116.
[41] Nagano, K., Ohta, T. and Yoshida, M. (1986) Proc. Natl. Acad. Sci. USA 83, 2071–2075.