

Cloning and characterization of the entire cDNA encoded by ATP1AL1 – a member of the human Na,K/H,K-ATPase gene family

Alexander V. Grishin*, Vladimir E. Sverdlov, Mariya B. Kostina, Nikolai N. Modyanov

Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow 117871, Russian Federation

Received 17 May 1994

Abstract

The cDNA for ATP1AL1 – the fifth member of the human Na,K/H,K-ATPase gene family – was cloned and sequenced. The deduced primary ATP1AL1 translation product is 1,039 amino acids in length and has M_r of 114,543. The encoded protein has all of the structural features common to known catalytic subunits of P-type membrane ion-transporting ATPases and is equally distant (63–64% of identity) from the Na,K-ATPase isoforms and the gastric H,K-ATPase. The ATP1AL1 encoded protein was proposed to represent a new separate group within the family of human potassium-dependent ion pumps.

Key words: Na,K-ATPase; H,K-ATPase; Gene family; cDNA sequence

1. Introduction

Among various cation-transporting ATPases which form phosphorylated intermediates during the catalytic cycle (P-ATPases), two groups possess the most similarity in structural and functional properties. These are the Na,K-ATPase, a universal component of animal cell plasma membranes, and the H,K-ATPase, a pump specific to the parietal cells of the gastric mucosa. These enzymes can be thought of as belonging to one group designated as the X,K-ATPases. A gene family, composed of six closely related genes was identified in the human genome [1,2]. Three of them were shown to encode the $\alpha 1$, $\alpha 2$ and $\alpha 3$ isoforms of the Na,K-ATPase catalytic subunit while a fourth encodes the catalytic subunit of the gastric H,K-ATPase. The functional status of the remaining two genes designated as ATP1AL1 (described also as SW 3.2 [1] and αD [2]) and ATP1AL2 (αC [2]) was unknown [3]. Further studies resulted in the isolation, mapping and partial sequencing of the ATP1AL1 gene and further demonstrated that it is expressed in human brain and kidney, at least [4]. The limited sequence information allowed us to suggest that the gene probably encodes a novel type of ion pump rather than an isoform of known ones [4]. Here, we

report the nucleotide and deduced amino acid sequences corresponding to the ATP1AL1 gene coding region. The protein sequence comparison demonstrates that ATP1AL1 protein – the fifth member of the human X,K-ATPase family – is equally distant from the Na,K-ATPase isoforms and gastric H,K-ATPase, and together with the putative rat distal colon and toad bladder H,K-pumps [5,6] belongs to the third separate group within the family of the X,K-ATPases.

2. Experimental

2.1. Total and poly(A)⁺ RNA preparation, cDNA library screening and cDNA sequencing

Total cellular RNA from frozen human tissue (except skin) was isolated by the guanidine thiocyanate method according to a protocol in [7]. Total cellular RNA from the skin samples was purified by the acid phenol method [8]. The poly(A)⁺-RNA was isolated as previously described [4].

The oligo(dT) and random-primed cDNA was generated following the protocol in [9,10]. Double stranded cDNA was inserted into plasmid pBR322 cleaved with *Pst*I by the dG:dC tailing technique [7], or was cloned into the *Eco*RI site of the phage vector λ gt10 or plasmid pSP64 (#P1091, Promega Corp., Madison, WI, USA) according to procedures supplied by the manufacturer (λ gt10 cDNA cloning kit, #RPN1713, Amersham Corp., Arlington Heights, IL). cDNA libraries were screened by hybridization with ³²P-labeled oligonucleotides. The oligonucleotide radiolabeling, hybridization, restriction analysis, subcloning and purification of DNA were carried out as in [7]. The nucleotide sequencing of all of the cDNA fragments was performed on both strands by the dideoxy chain termination method [11] using Sequenase Version 2.0 DNA Sequencing Kit (#70770, United States Biochem. Corp., Cleveland, OH, USA).

Computer analyses of the nucleotide and protein sequences were performed using DNASTAR (DNASTar Inc., Madison, WI, USA) and UWGCG package (Genetics Computer Group, Madison, WI, USA) program pockets.

2.2. PCR-amplification and oligonucleotides

For PCR amplification *Taq* and *Tth* (NMDCentr, Moscow, RF), *Vent* (New England Biolabs, Inc., Beverly, MA, USA) and *Pfu*

* Corresponding author. *Present address:* Department of Cellular and Molecular Physiology, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA. Fax: (1) (203) 785 4951. E-mail: Grishin@Biomed.Med.Yale.Edu

Abbreviations: FITC, fluorescein isothiocyanate; ClATP, α -[4-(*N*-2-chloroethyl-*N*-methylamino)]benzylamide-ATP.

The nucleotide sequence presented here is available in the GenBank/EMBL Data Bank with Accession Number U02076.

(STRATAGENE Cloning Systems, La Jolla, CA, USA) DNA polymerases were used. The reaction was performed as previously described [4] with modifications. Briefly, random-primed cDNA was generated from 1 µg of the total or 0.5 µg of poly(A)⁺ RNA following a standard technique [7] and one tenth of the synthesized cDNA was subjected to 30 cycles of enzymatic amplification at 95°C, 0.5 min; 55–60°C, 0.5 min; and 72°C, 2 min.

For the race amplification of the 5' region of the mRNA of interest (5'-RACE) cDNA was generated from 0.5–1 µg of the total or of poly(A)⁺ RNA with oligonucleotide 830, as a primer, as described above [7]. After this step the protocol described in [12] was followed, except the low melting agarose electrophoresis and anion-exchanging chromatography stages for the separation of the cDNA first chain from the excess of primer were omitted, and oligonucleotide 825 was used as a gene-specific primer in PCR amplification reaction.

The PCR-generated fragments were blunted with T4 DNA polymerase [7] and cloned into *Sma*I-cut pUC18 and sequenced by the dideoxy method [11].

The oligonucleotide used in this study listed below with coordinating cDNA positions in parenthesis: pU69, 5'-AGATTCCGAGAAGAA-GACCA-3' (464–483); pU70, 5'-GCTGGGGCTCAGACTCCCCCG-TGAGA-3' (696–721); 825, 5'-CAAGCACACAGCCCAAGTAC-3' (435–454); 826, 5'-AAGGTACCAGAGGGAATACCT-3' (2,766–2,786); 827, 5'-GTAGCCCGTCCATTCTAGGTAT-3' (2,781–2,802); 828, 5'-GACCCTGTCGCTGACAGCA-3' (1,071–1,089); 829, (complementary to 828) 5'-TGCTGTCAGCGACAGGGTC-3'; 830, 5'-CCGTAAAATGACCACCAGACC-3' (454–475); 893, GGATGCC-AAGGCCGCTGTGGTGAC-3' (2,019–2,042); 894, 5'-CAGTCCA-TGCCAGTCACCACAGC-3' (2,032–2,055); 983, 5'-TGGTGATGA-AGGGCGCCC-3' (1,559–1,577); 984, (complementary to 983) 5'-GGGGCGCCCTTCATCACCA-3'. Information for the synthesis of oligonucleotides used in present experiments was obtained by comparison of genomic sequence data obtained in previous studies [1,4] with other human P-ATPases. The sequence for the synthesis of the oligonu-

cleotides 983 and 984 was chosen based on the codon preference usage [13] (see section 3.1). The 'GGC' codon for Gly⁵²⁴ was chosen as the most frequently used, and the cytidine at third position is substituted for G in the cloned cDNA sequence.

3. Results and discussion

3.1. Isolation and sequence analysis of the ATP1A1 cDNA from human kidney and skin

In order to isolate the ATP1A1 cDNA we prepared and screened two human kidney libraries (randomly and oligo(dT) primed, about 450,000 colonies in sum) with gene-specific oligonucleotides (826 and 830 – see Fig. 1). Two individual clones (pHK 14 and pHK12, Fig. 1) that gave reproducible positive hybridization signals with oligonucleotide 826 were isolated. Sequence analysis showed that these cDNAs included 596 bp from 3'-end of the coding region plus 290 nucleotides from the 3'-untranslated region, including the poly(A) tail. Several attempts to isolate other positive clones with different gene-specific oligonucleotides from this and two others kidney libraries (in γ gt10 vector, $\sim 2.5 \times 10^6$ clones in total) were unsuccessful.

To overcome this problem we undertook PCR-based amplification employing gene-specific oligonucleotides.

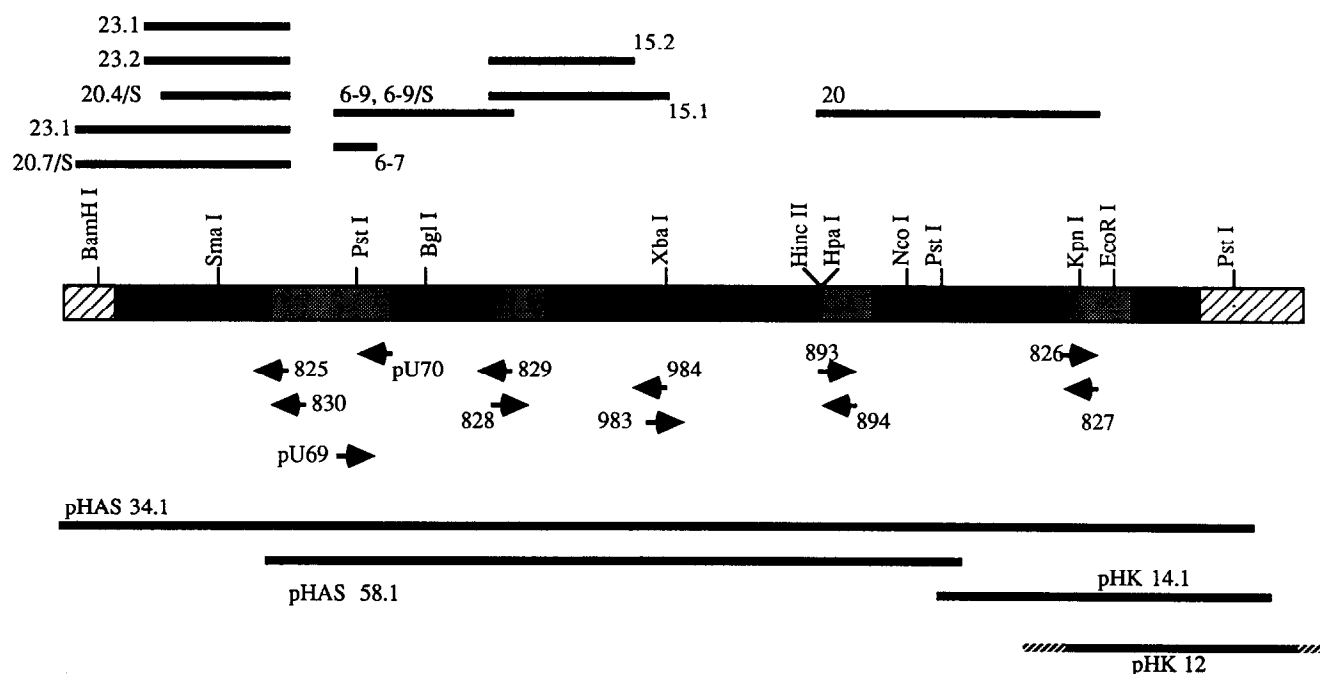


Fig. 1. Restriction map and cloning and sequencing strategy for the human ATP1A1 cDNA. The black and shaded rectangles represent coding regions and hatched rectangles represent untranslated regions. The shaded rectangles correspond to the genomic sequence of the ATP1A1 exons, determined earlier [4]. The black bars represent cDNA fragments obtained in PCR-amplification experiments (above colour-coded rectangle) or cloned from non-amplified human cDNA libraries (below). PCR-fragments with numbers were amplified from poly(A)⁺ RNA fraction from kidney and with numbers plus 'S' – from the total cellular RNA from armpit skin (see text). The gene-specific oligonucleotides that were used as PCR and sequencing primers are shown by arrowheads. The direction and extent of sequencing are indicated by arrows. Nucleotide sequence of the pHK12 cDNA was determined only for the corresponding PCR fragment obtained in the amplification reaction of the original plasmid with 826 and oligo(dT) primers, since the cDNA insert could not be excised from the cloning vector.

As a result, several ATP1AL1 cDNA fragments that overlapped 1,837 nucleotides of the mRNA of interest, including 71 bp of the predicted 5'-untranslated region (see Figs. 1 and 2), were cloned and sequenced. The sequence corresponding to the region between nucleotides 1,256 and 2,135 was not recovered, despite the use of different DNA polymerases and attempts to optimize the PCR-reaction conditions (see section 2). The PCR-fragment for the 5'-half of this region, but not for the 3'-half (Fig. 1), was isolated using two oligonucleotides (983 and 984) corresponding to the FITC binding site which is characteristic of the ion-motive P-ATPase family [14] (Fig. 3). Presumably, the region between nucleotides 1,570 and 2,135 contained some kind of 'stop'-sequence(s) for the polymerases used in these experiments.

A search for the ATP1AL1 transcript in different human tissues and organs revealed the presence of the mRNA of interest in skin samples from several specific areas, including axilla (Akopyanz N.S., Grishin A.V. et al., manuscript in preparation). Presumably, the ATP1AL1 protein is a component of sweat glands (ibid.). The relative abundance of this mRNA was higher in this tissue than in all others examined. Therefore, RNA of this tissue was chosen for further cloning experiments. In spite of the low content of the poly(A)⁺-fraction in this tissue (less than 0.3%), we were able to obtain a cDNA library of 2.4×10^5 independent clones which was screened with two oligonucleotides (828 and 894, Fig. 1). As a result, two clones (pHAS 34.1 and pHAS 58.1, Fig. 1) were isolated and sequenced. The cDNA insert from the pHAS 34.1 plasmid was of 3,403 base pairs and contained the entire coding part of the gene as well as 5'- and 3'-untranslated regions (Fig. 1). It allowed us to bring together all of the sequence information that had been obtained earlier. No differences between skin and kidney cDNAs and the known genomic sequences [4], ad Sverdlov et al., unpublished) were revealed, and

there was no indication of any alternative splicing of the ATP1AL transcript.

At present, the actual molecular size of the ATP1AL1 mRNA remains to be determined. Possibly due to actual low number of the mRNA copies in poly(A)⁺ RNA from brain and kidney and nontypical low relative content of poly(A)⁺ fraction in the skin total RNA, no positive signal could be detected by Northern blot analysis using different specific probes (data not shown).

The composite restriction map and cloning and DNA sequencing strategy for the human ATP1AL1 gene encoded cDNA are presented in Fig. 1. The nucleotide sequence of the 5'- and 3'-untranslated regions and deduced amino acid sequences are shown in Fig. 2 and Fig. 3, respectively. The cDNA consists of 3,574 nucleotides and includes a protein coding region of 3,117 nucleotides that is preceded by a 168-nucleotide 5'-untranslated sequence and followed by a 3'-untranslated region of 290 nucleotides. The predicted initiation methionine codon is the first ATG-codon and is preceded by non-interrupted open reading frame which we believe to represent mRNA 5'-untranslated sequence. The 3'-untranslated sequence includes a poly(A)-tail which is preceded by the AAAAAA sequence that has been shown to be a very rarely used but functional polyadenylation signal [15].

3.2. Structural features of the ATP1AL1 protein and its relationship with other X,K-ATPases

The primary ATP1AL1 gene translation product deduced from the cDNA sequence (Fig. 3) consists of 1,039 amino acid residues and has a M_r of 114,543. This new human protein has all structural features of an ion-transporting P-ATPase. These include several conserved motifs such as a catalytic phosphorylation site (Asp³⁹¹*[17]), components of the ATP-binding site (Asp⁷³² and Asp⁷³⁶ [18]), the target of FITC modification (Lys⁵²³ [14]), and the very characteristic disposition of potential transmembrane domains [14]). The alignment of the deduced protein sequences of the catalytic subunits of all five human X,K-ATPases and two others which are homologous to the ATP1AL1 protein (putative rat colon and toad bladder H,K-ATPases) is shown in Fig. 3.

ATP1AL1 exhibits 63.1–64.2% amino acid identity to the catalytic subunit of human gastric H,K-ATPase and the three α -isoforms of human Na,K-ATPase. Thus, the ATP1AL1 protein is equally distant and related to both of the above groups and can not be assigned as an isoform of any of them. The same level of sequence similarity was demonstrated for the related family of five rat ATPases [5]. The toad bladder H,K-ATPase α is 67% identical to the toad Na,K-ATPase α -subunit [6]. The sequence comparison of new human, rat and toad

5'-UTR

```
CGGCCGCGGAGGTGCGTGCAGGGCCCGCCGCGCGGTATCTCCACCGCCAAACACCTC -108
AGCCACTGCCACTGCCACAGCCACAGGCCCCCACCCTGCGCTCCGCGCTGCGGTC -48
CCGGATCCGCGCTCCACGCCCGCAGCCCGCGCGGCCACAGCCAGCATG 3
```

3'-UTR

```
TAAAGACCACCTCCCTTCTATGTCTCTCAGCAGCAGTGGGGCAGCACTTGTCATCTTC 3177
TGACCGTTTCTGGGCTATTCCTCTGAGTGCAGACATCGTCAAAATTCATAAGAGGA 3237
AATTTTCATGCAGAAAGCTGATGCAGGATGCTCACTGATGTTTGCACCTTAAACTGA 3297
AATTCACCTCTTTATAGGATTTCTTTCTATCTCCATCTCTCATTAATAAAACAGT 3357
ACATTTCCAGGTAATGGTATATAAAAAAAAAAAAAAAAAAAAAAAAAA 3407
```

Fig. 2. Nucleotide sequence of the 5'- and 3'-untranslated regions (5'-UTR and 3'-UTR, respectively) of the human ATP1AL1 cDNA. The predicted initiation ATG and TAA-stop codons, and potential polyadenylation site are underlined. Nucleotides numbered at the right of the sequences, and coordinates are indicated according to the complete cDNA sequence, which is available in the GenBank/EMBL Data Bank with Accession Number U02076).

* Here and throughout, all amino acid numbering is indicated according to the ATP1AL1 encoded protein sequence (Fig. 3).

H,Kb	MSTENS-DV..TAIN.DNE.GYP-.EKEEDIGLEVOKK.EKKKK..GNNVDLQK.D.E...IE...A.E.SL-Q..T.A...I.....T....G....	107
H,Kc	MRR-...N...-PAQR...D..FK.A.KD..PN.SHE...LK...D...R...T...Q...N...Q...V...T...T...S...T...T...T...	102
X,Kh	MHQK-TPETYSVELSGTDIVKT-DKG-DGKGYKGLNKNCLKQKHKKEPKHLDHDKLSNRELEEKYGDIDIMGLSSTRAAELLARQGNLSITPPKQTEI	104
H,Kg	MKAEN-Y.L...GGPGGGMARMSKK-KAGGGGGK...R.E.-L.NMK..MEIN.Q.VA...Q.Q.SATK..ASL...L...A.R..RG...Y	100
N,K1	MKGAGR-EYSEPAITAENGKK-KQ...E-ELD.LK..VAM...LD..GR..QV.LSK..TNQ..QDV...A...PT...W	89
N,K2	MKGAGR-EYSEPAITAENGKK-KQ...E-ELD.LK..VAM...LD..GR..QV.LSK..TNQ..QDV...A...PT...W	87
N,K3	MKGAGR-EYSEPAITAENGKK-KQ...E-ELD.LK..VAM...LD..GR..QV.LSK..TNQ..QDV...A...PT...W	79
* * *		
H,Kb	I...I...L...A..I...L.AQ.HNT.RD.L..I..AV...C...F...A...Q...V.T.NQ.Q.L.E.GPKD...L.D...R.	217
H,Kc	I...I...L...A..I...L.AQ.HNT.RD.L..I..AV...C...F...A...Q...V.T.NQ.Q.L.E.GPKD...L.D...R.	211
X,Kh	VFILQMGSGFSLDWGAPLWLAIGIYSSSKASLNWVLCVILGLVILTGIPAYQKASTNDSFNMKIPQQAIVRDESKTIPSEQLVWGVIVKGGQDI	214
H,Kg	AR.LA..LQC.M..A.AI.L..FA..A.EGLITTD.L..AIA..IA..VV..C.G...F...LA..KLV...T...GD.FQ.NND...L..M...RV	210
N,K1	CR.LF...I..I..FL..S..AAIEEPQND.L..V.SA..I..C.S...SK..E..KN.V...NG..MS.NA.EV...L...R.	199
N,K2	CR.LF...I..I..FL..S..AAIEEPQND.L..V.SA..I..C.S...SK..E..KN.V...NG..MS.NA.EV...L...R.	197
N,K3	CR.LF...I..I..FL..S..AAIEEPQND.L..V.SA..I..C.S...SK..E..KN.V...NG..MS.NA.EV...L...R.	189
H-1 H-2		
H,Kb	L.IIFA..K...SL...A...A...AR.F...Q...R...V...VGV...	327
H,Kc	L.IIFA..K...SL...A...A...AR.F...Q...R...V...VGV...	321
X,Kh	PADIRVLSQSGRQVNSLGESEPSPRSSEPTHEPLETNNICFYSTLLEGVIGMINTDRTIIGHLASLASVSGNEKTPDAIEDEHVLHAGVAVSGLIEFII	324
H,Kg	I.AA..K...T..P.C...S...R.A.F..M...AQ.L.V...R...E...D..I..L..L..L..AT...V	320
N,K1	L.II.AN..K...T..P.D...N...R.A.F..N.V...AR.I.VI...VM.R..T...LEGGQ...A...I..IT...FL.VS...L	309
N,K2	L.II.AN..K...T..P.D...N...R.A.F..N.V...AR.I.VI...VM.R..T...LEGGQ...A...I..IT...FL.VS...L	307
N,K3	L.II.AN..K...T..P.D...N...R.A.F..N.V...AR.I.VI...VM.R..T...LEGGQ...A...I..IT...FL.VS...L	299
H-3		
H,Kb	ICMG.SA.N...DH.HI...Q.HHS.E.TPE..N	437
H,Kc	ICMG.SA.N...DH.HI...Q.HHS.E.TPE..N	431
X,Kh	AVSLKQVDSIIFLIGIVANVEGLLAVIVELSLTAKRPAKRCIVNLEAVETGTSIICSQKTLITQNMIVAHLEFMDQIFVADTSDHSNVEFDQSSRTWA	434
H,Kg	MCIG.TF.RAMV.FMA.V..Y...C...L.S..V...V...S...H.HT...T..Q.G.T...E..R	430
N,K1	SLI.E.TW.EAV...C.T...R...T...M...HE..T.NQ.GVS..KT.A..L	419
N,K2	SLI.E.TW.EAV...C.T...R...T...M...HE..T.NQ.GVS..KT.A..L	417
N,K3	SLI.E.TW.EAV...C.T...R...T...M...HE..T.NQ.GVS..KT.A..L	409
H-4		
H,Kb	A.C..VS...A..DD...VAV...T.N..N..SQ...C...T...QDQ.L.L...G.K.L.	547
H,Kc	A.C..VS...A..DD...VAV...T.N..N..SQ...C...T...QDQ.L.L...G.K.L.	541
X,Kh	SLSKITITLQRAEFPKQENVPMKAVIGDASETALKFSEVILGDVMEIRKRVKVAEIPFNSTNGFQSLSHEDDPHGKRFIMMKGAPERILEKSTIMINGEHP	544
H,Kg	A.CRVL...A..S..DA..VP.RI...LT..NA.GY.D.F..C...TLE..RDP.H.L...V..R..S.L.K.Q.L.	540
N,K1	A..R.AG...V.QAN..NL..L.R.A...S...CI.LCC.S.K.M.E.YA.IV...Y...KNFNTSEPH.L...DR..S.LH.K.Q.	529
N,K2	A..R.AG...V.QAN..NL..L.R.A...S...CI.LCC.S.K.M.E.YA.IV...Y...KNFNTSEPH.L...DR..S.LH.K.Q.	528
N,K3	A..R.AG...V.QAN..NL..L.R.A...S...CI.LCC.S.K.M.E.YA.IV...Y...KNFNTSEPH.L...DR..S.LH.K.Q.	519
H,Kb	E.MDS.Q...EE.Y.SS.A..IES...R...G..	657
H,Kc	E.MDS.Q...EE.Y.SS.A..IES...R...G..	651
X,Kh	LDKSTAKIFHTAMGLGLGERVLPCHLYLPADEFPETYSFDIDAMFPTSNLCFVGLLSMIDPPRSTVPDAVTKRSAGIKVIMVGDHPITAKALAKSVGISANSE	654
H,Kg	EQMREA.Q...LS...Q...NEKDY.PGYA..VE...S.G..A.V...A...L..T...R...A...EG..	650
N,K1	EEKQDA.QN..L...P..DEQ..GFQ..T.DV...ID...I...AA...G...EGN.	639
N,K2	EEKQDA.QN..L...P..DEQ..GFQ..T.DV...ID...I...AA...G...EGN.	636
N,K3	EEKQDA.QN..L...P..DEQ..GFQ..T.DV...ID...I...AA...G...EGN.	629
H,Kb	D..K...P...N..E...N.G...E..D..T.HA...NY...	767
H,Kc	D..K...P...N..E...N.G...E..D..T.HA...NY...	760
X,Kh	TVEDIARLNIAVEQV-NKRDAAVAVTGMELKMSSEQLDELLANYQEVFARTSEPOKLIIVEGQFQDAVAVTGGVNDSPALKKADIGIAMGAGSDAANKAADM	763
H,Kg	A..RVE.D...-RK..R.C.IN..Q..DPSE.V.A.RHE.M...V..S..LG.I...V...V...VS.Q...	759
N,K1	A..RVE.D...-RK..R.C.IN..Q..DPSE.V.A.RHE.M...V..S..LG.I...V...V...VS.Q...	748
N,K2	A..RVE.D...-RK..R.C.IN..Q..DPSE.V.A.RHE.M...V..S..LG.I...V...V...VS.Q...	746
N,K3	A..RVE.D...-RK..R.C.IN..Q..DPSE.V.A.RHE.M...V..S..LG.I...V...V...VS.Q...	738
H-5, H-6		
H,Kb	I...I...S.G.T...V...ADI...VSEF...R...R..V...Q..L.A.Q...	877
H,Kc	I...I...S.G.T...V...ADI...VSEF...R...R..V...Q..L.A.Q...	870
X,Kh	VLLDNFASIVIGVEEGRLLFNKKTAYSLTKNLAELPFLIYIVGLFLPFGITITLFDLGDIDILPSIALAYEKAESDMMKPKPKKDRLVNQLAVISYIHIG	873
H,Kg	I...I...Q...S...T...P.T.Y...T.SV...L.C...E.C...F..VS...HLR..NEKR...E..A..FQ...	869
N,K1	I...I...S...T...S.P.IT...F..ANI..L.V..C...MV.A.S...Q...K.Q..NKPT.K..ER.ISKA.GQ..	858
N,K2	I...I...S...T...S.P.IT...F..ANI..L.V..C...MV.A.S...Q...K.Q..NKPT.K..ER.ISKA.GQ..	855
N,K3	I...I...S...T...S.P.IT...F..ANI..L.V..C...MV.A.S...Q...K.Q..NKPT.K..ER.ISKA.GQ..	848
H-7 H-8 H-9		
H,Kb	II.SV...N...M.EQ...H..VGI.ID...INNO..E...FS..CF...S.V.E.L...V...FLM.LA..V..AAF	987
H,Kc	II.SV...N...M.EQ...H..VGI.ID...INNO..E...FS..CF...S.V.E.L...V...FLM.LA..V..AAF	980
X,Kh	IMQAGFLNYFTVYQGEFLPRLINLRVENEKDYVNDKDSYGQEWIRYOREYLEWIGYTAFFVGILVQDIADLIIRKTRNSIPQGLFRNKVWIGITQIIIGLI	983
H,Kg	AI.SFAG.TD..AM..WE.LLCVG..AQ..DHLHQ..Q...FG..L.QOY.C..V..IS.E.C...VL...L.A...F...ILVIA.VF.VC..CF	979
N,K1	MI...G.FT..VIL.EN...SR.LGI.LD.DDRIM..E...YE..KVF.F.CH...AS.V.V.W...C...V...MK..ILIF.LFEETALAAP	967
N,K2	MI...G.FT..VIL.EN...SR.LGI.LD.DDRIM..E...YE..KVF.F.CH...AS.V.V.W...C...V...MK..ILIF.LFEETALAAP	964
N,K3	MI...G.FT..VIL.EN...SR.LGI.LD.DDRIM..E...YE..KVF.F.CH...AS.V.V.W...C...V...MK..ILIF.LFEETALAAP	957
H-7 H-8 H-9		
H,Kb	CPMPY..K..P...A.Y.L..F...I...R...	1043
H,Kc	CPMPY..K..P...A.Y.L..F...I...R...	1036
X,Kh	LSYGLGSVTSALSTMLRAQVAVPHAILIWDVDEKFLIRLYPGWMDKNMY	1039
H,Kg	CP.MENIEN.MPI.F.W.L.PL.YG..F...I...GV.CC...QEL..	1035
N,K1	CP.MGV..RMP.KVW..C.F.YSL..FI...I..RR..G.VE.ET..	1023
N,K2	CP.MGV..RMP.KVW..C.F.YSL..FI...I..RR..G.VE.ET..	1020
N,K3	CP.MGV..RMP.KVW..C.F.YSL..FI...I..RR..G.VE.ET..	1013
H-10		

Fig. 3. Comparison of the amino acid sequences of the human ATP1A1 encoded protein (X,Kh) and of the catalytic subunits of the putative rat distal colon H,K-ATPase (H,Kc) [5], toad bladder H,K-ATPase (H,Kb) [6], human gastric H,K-ATPase (H,Kg) [19,20], and the human α_1 -, α_2 -, and α_3 -isoforms of Na,K-ATPase (N,K1, N,K2, and N,K3, respectively) [21–23]. Gaps (represented by dashes) were introduced to maintain the alignment. Dots indicate identity to the corresponding residues of the human ATP1A1 putative protein. Hydrophobic sequences that meet the criteria for consideration as potential transmembrane domains (determined by the method of [16]) are underlined and labeled. The phosphorylation site, the binding site for FITC and CIATP (Asp⁷³² and Asp⁷³⁶) [18] are overlined. Amino acid residues which were shown to be important for binding of ouabain, Sch 28080 and omeprazole indicated by an asterisk (*). The charged amino acids residues located within the predicted transmembrane segments are indicated by filled diamonds (◆ – which are preserved, or conservatively substituted, in all proteins) or by filled triangles (▲ – which are not preserved in all sequences). The amino acids are numbered at the right of the sequences.

ATPases revealed that ATP1AL1 exhibits 86.1% of overall identity with rat colon pump and both mammalian enzymes are about 75% identical to the toad bladder H,K-ATPase. In contrast, the human and rat Na-pump α -isoforms and human and rat gastric H,K-ATPases each exhibit 96.7–99.1% homology to one another. About 93% of the toad Na,K-ATPase α -subunit amino acid sequence is identical to the rat and human $\alpha 1$ -isoform. On the other hand, the Na,K-ATPase $\alpha 1$, $\alpha 2$ and $\alpha 3$ from the same organism share 86–87% identical amino acids. The relationships of these sequences are presented as dendrogram in Fig. 4. Obviously, the ATP1AL1-encoded protein, the rat colon pump and the toad bladder H,K-ATPase represent one class of homology which is distant both from the Na,K- and gastric H,K-ATPases. However, members of this third group are more divergent than related enzymes. This may indicate that sequence requirements for the ATP1AL1 protein, putative rat colon and toad bladder H,K-ATPases were not as rigid during evolution, as was the case for the extremely well-conserved H,K- and Na,K-ATPases. At present, the possibility that these proteins [5,6] are products of closely related but distinct genes can also not be excluded.

The availability of three closely related sequences from different species allows us to dwell on the structural features specific to this group of ion-pumps (see also [5,6]). Thus, the alignment in Fig. 3 depicts a number of regions which are extremely conserved in all of the X,K-ATPases. The domains can be defined in coordinates of the ATP1AL1 amino acid sequence as 225–240, 337–412, 491–508, 521–538, 556–576, 591–650 and 702–796 (Fig. 3). In addition, fragments 94–132, 144–189, 200–221, 804–852, 947–968, 1,012–1,023 form a group of less conserved but still homologous sequences. Most of these sequences are located in major cytoplasmic domains (H2–H3 and H4–H5), that were shown to be involved in ATP-binding or catalytic activity (see Fig. 3 and above), but cover also several transmembrane segments, including H1, H2, H4 and part of the H5–H6 pair. The essential sequence peculiarity of the third group of pumps within the above regions is that in certain positions they all have amino acids residues characteristic of either the gastric H,K-ATPase or the sodium pump α -isoforms.

Large differences in the distribution of charged amino acid residues can be revealed upon comparison of hydrophobic domains forming the putative cation-conducting pathway (Fig. 3). It was suggested that charged aspartate and glutamate residues could be involved in cation occlusion [28,29]. This appears to be the case for Glu⁸⁰¹, Asp⁸²⁶, and Asp⁸³⁰ that have been shown to be crucial for the Na,K-ATPase transporting activity [30,31]. All of them are preserved in all of the sequences in Fig. 3 except Asp⁸²⁶/Glu substitution in the gastric pump. The functional role of Glu³⁴⁹ and Asp⁹⁴⁸ which are conserved in all proteins, remains unclear [30]. In addition, several

charged residues can be pointed out as characteristic of one of the groups – Asp⁸⁸³ and Glu⁹⁴² (for gastric H,K-ATPase) and Glu⁹⁷⁶ and Glu⁹⁷⁷ – for Na,K-ATPase (but see [32]). On the other hand, the ATP1AL1 protein, rat colon pump and toad bladder H,K-ATPase bear different amino acid residues in each position where any of them has charged amino acid and there is no counterpart in the other sequence groups. For example, both the positions 318 and 812 are occupied either by Asp or Gly, and 944 – by Glu or Gln (Fig. 3). (This type of divergence is very characteristic also for extracytoplasmic part of polypeptide chains (Fig. 3)). This argues against the presumed functional significance for these particular amino acid residues. There is only one common charged amino acid residue in the membrane portion of the gastric H,K-ATPase, the ATP1AL1 protein and the colon and bladder pumps, namely Lys⁷⁹⁷. The toad bladder enzyme, when expressed in *Xenopus* oocytes in association with β -subunit was demonstrated to pump protons outside and potassium inside the cell [14]. Whether this amino acid participates in proton transport remains to be shown.

The extracellular protein domain contributes to ouabain binding in the Na,K-ATPase [33–39] and the docking of Sch 28080 and omeprazole in the gastric H,K-ATPase [40]. It can be seen that the ATP1AL1 protein, rat colon and the toad bladder pumps share many amino acid residues that were shown to be important for ouabain binding – Cys¹²⁶ [33,34], Tyr/Phe¹³⁰ [33], Tyr³³⁰ [35], Thr⁸¹⁹ [36] and Arg⁹⁰² [37]. The non-charged constituents of the H1–H2 extracytoplasmic border residues [38,39] are also conserved (Fig. 3). One of those residues, the Phe¹³⁰, in concert with Asp¹³⁸ (which presents only in the gastric H,K-ATPase), was predicted to participate in the binding of Sch 28080 [40]. Interestingly, the toad bladder pump was inhibited by both ouabain and Sch 28080 with K_i values of 25 μ M and 230 μ M, respectively [6]. The ouabain inhibition constants within the same order of magnitude were shown for the Na,K-ATPase α -subunit mutants which possessed at least one of the residues in positions 126, 130, 819 or 902 substituted [33,34,36,37] (see above), while in toad bladder enzyme they are all preserved. It demonstrates that many aspects of the inhibitor pump interaction remain unknown.

Sensitivity to omeprazole within this family seems to be a unique feature of the gastric H,K-ATPase, since only this enzyme possesses three cysteine residues at positions 819, 828 and 898 within the H5–H6 and H7–H8 extracytoplasmic domains that were shown to bind this drug covalently [40]. Interestingly, in all other proteins Cys⁸¹⁹ is substituted by a Thr involved in ouabain binding [36].

It can be seen that the most divergent region in the X,K-ATPases family is the extreme N-terminal domain (Fig. 3) that was shown to be important for cation binding and occlusion in Na,K-ATPase [41–43]. The right-

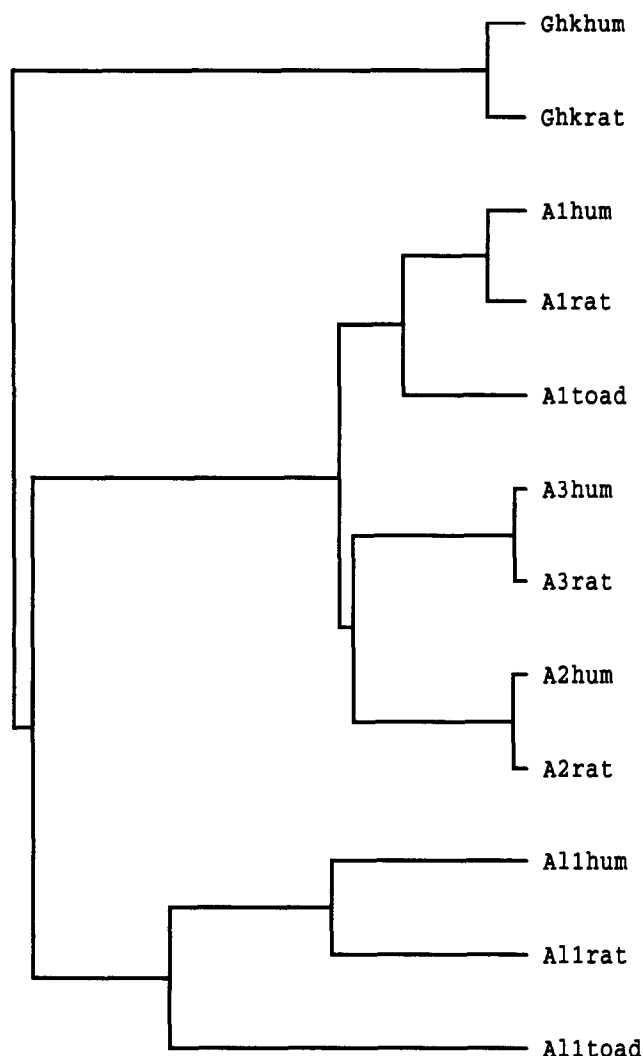


Fig. 4. Dendrogram of the amino acid sequences of the human ATP1A1 encoded protein (A11hum) and of the catalytic subunits of the putative rat distal colon H,K-ATPase (A11rat) [5], toad bladder H,K-ATPase (A11toad) [6], human (Ghkhum) [19,20] and rat (Ghkrat) gastric H,K-ATPases [24], toad Na,K-ATPase (A11toad) [25] and the α_1 -, α_2 -, and α_3 -isoforms of the Na,K-ATPase from human [21–23] and rat [26] (A1hum, A2hum, A3hum, and A1rat, A2rat, and A3rat, respectively), was created by PILEUP program (GCG Package) using a simplification of the progressive method of Feng and Doolittle [27]. Distance along the vertical axis is proportional to the difference between the sequences.

most boundary of this domain appears to lie between Leu⁵⁶ and Asp⁵⁷ that corresponds to the position of the second intron found in human genes encoding H,K-ATPase [19,20] and the Na,K-pump α_3 -isoform [23], as well as in the ATP1A1 gene (Sverdlov et al., unpublished observation). In this region, ATP1A1 exhibits only 55.4% identity with the rat colon enzyme, and these two are 28.5% and 33% identical to the bladder pump, respectively. Such a low level of similarity for homologous proteins is nontypical for the X,K-ATPase family. For example, all other human and rat homologs share about

90% identical amino acid residues with their corresponding rat homologs. Therefore, the prominent diversity between ATP1A1, rat colon and bladder pumps resemble the relations between isoforms or even different enzymes rather than species difference among the same enzyme. On the other hand, it may indicate that there is not the same significant role for the N-terminal domain in these new enzymes (if they are not isoforms but homologs) which was found for Na,K-ATPase [41–43].

All of the preceding observations indicate that the sequence mosaic within functionally important domains may be seen as the most essential structural feature of the new subfamily formed by ATP1A1, putative rat colon pump and toad bladder H,K-ATPase. However, there are several conservative motifs (914–968, 995–1,039) and amino acids (for example – Gly²⁹³, Glu⁴⁴⁷, Lys⁸⁵⁴ etc.) that could be considered as characteristic to this whole group.

In conclusion, we have isolated and sequenced the cDNA for the human ATP1A1 gene encoding a new member of the X,K-ATPases catalytic subunit family. The ATP1A1 protein, rat colon pump [5] and the toad bladder H,K-ATPase [6] form a distinct group within this family. The sequence relationships between members of this group are more divergent than that for the other related groups, and it remains to be shown whether these proteins are products of closely related but distinct genes [5,6], or if the structure–function relationships within this subfamily are quite different. Both the bladder pump [6] and the putative rat colon H,K-ATPase [5] were suggested to be involved in K⁺ homeostasis. One could thus propose that the ATP1A1 gene encodes the catalytic subunit of the human non-gastric H,K-ATPase. Verification of this proposal is the aim of our current research. The requirement for any distinct beta-subunit for the normal enzyme activity, possible physiological role of the ATP1A1 protein, and its pharmacological properties are also the goals of further study.

Acknowledgements: This work was supported in part by Russian Foundation for Fundamental Research. We thank Marina Orlova and Oksana Makarevich for technical assistance; Dr. N. Bystrov for oligonucleotide synthesis; Dr. E.D. Sverdlov and members of his laboratory for helpful discussion and support at initial stages of this study and members of our laboratory for encouragement. The authors are grateful to Dr. M.J. Caplan for supporting the completion of this work and for help in preparation of this manuscript. Dr. Kathi Geering is thanked for valuable comments on the manuscript.

References

- [1] Sverdlov, E.D., Monastyrskaya, G.S., Broude, N.E., Ushkarev, Y.A., Allikmets, R.L., Melkov, A.M., Smirnov, Y.V., Malyshev, I.V., Dulubova, I.E., Petrukhin, K.E., Grishin, A.V., Kijatkin, N.I., Kostina, M.B., Sverdlov, V.E., Modyanov, N.N. and Ovchinnikov, Y.A. (1987) FEBS Lett. 217, 275–278.
- [2] Shull, M.M. and Lingrel, J.B. (1987) Proc. Natl. Acad. Sci. USA 84, 4039–4043.

- [3] Yang-Feng, T.L., Schneider, J.W., Lindgren, V., Shull, M.M., Benz, E.J., Lingrel, J.B. and Francke, U. (1988) *Genomics* 2, 128–138.
- [4] Modyanov, N.N., Petrukhin, K.E., Sverdlov, V.E., Grishin, A.V., Orlova, M.Y., Kostina, M.B., Makarevich, O.I., Broude, N.E., Monastyrskaya, G.S. and Sverdlov, E.D. (1991) *FEBS Lett.* 278, 91–94.
- [5] Crowson, M.S. and Shull, G.E. (1992) *J. Biol. Chem.* 267, 13740–13748.
- [6] Jaissner, F., Horisberger, J.-D., Geering, K. and Rossier, B.C. (1993) *J. Cell Biol.* 1421–1429.
- [7] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Gold Spring Harbor, NY.
- [8] Chomzynski, P. and Sacchi, N. (1987) *Anal. Biochem.* 162, 156–159.
- [9] Krug, M.S. and Berger, S.L. (1987) *Methods Enzymol.* 152, 317–325.
- [10] Ubler, U.G., *Methods Enzymol.* 152, 325–330.
- [11] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- [12] Belyavsky, A., Vinogradova, T. and Raewsky, K. (1989) *Nucleic Acid Res.* 17, 2919–2932.
- [13] Aota, S., Gojobri, T., Ishibashi, F., Uyama, T. and Ikemura, T. (1988) *Nucleic Acid Res.* 16, suppl. r315–402.
- [14] Horisberger, J.-D., Lemas, V., Kraehenbuhl, J.P. and Rossier, B.C. (1991) *Annu. Rev. Physiol.* 53, 565–584.
- [15] Sheets, M.D., Stephen, C.O. and Wickens, M.P. (1990) *Nucleic Acid Res.* 18, 5799–5805.
- [16] Eisneberg, D., Schwarz, E., Komaromy, M. and Wall, R. (1984) *J. Mol. Biol.* 179, 125–142.
- [17] Bastide, F., Meissner, G., Fleishner, S. and Post, R.L. (1973) *J. Biol. Chem.* 248, 8385–8391.
- [18] Ovchinnikov, Yu.A., Dzhandzugazyan, K.N., Lutsenko, S.V., Mustaev, A.A. and Modyanov, N.N. (1987) *FEBS Lett.* 217, 111–116.
- [19] Neuman, P.R., Greeb, J., Keeton, T.P., Reyes, A.A. and Shull, G.E. (1990) *DNA Cell Biol.* 9, 749–762.
- [20] Maeda, M., Oshiman, K.-I., Tamura, S. and Futai, M. (1990) *J. Biol. Chem.* 265, 9027–9032.
- [21] Kawakami, K., Ohta, T., Nojima, H. and Nagano, K. (1986) *J. Biochem.* 100, 389–394.
- [22] Shull, M.M., Pugh, D.G. and Lingrel, J.B. (1989) *J. Biol. Chem.* 264, 17532–17543.
- [23] Ovchinnikov, Yu., Monastyrskaya, G.S., Broude, N.E., Allikmets, R.L., Ushkarov, 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.
- [24] Shull, G.E. and Lingrel, J.B. (1986) *J. Biol. Chem.* 261, 16788–16791.
- [25] Jaissner, F., Caness, C.M., Horisberger, J.-D. and Rossier, B.C. (1992) *J. Biol. Chem.* 267, 16895–16903.
- [26] Schull, G.E., Greeb, J. and Lingrel, J.B. (1986) *Biochemistry* 25, 8125–8132.
- [27] Feng, D.F. and Doolittle, R.F. (1987) *J. Mol. Evolution* 25, 351–360.
- [28] Ovchinnikov, Yu.A., Modyanov, N.N., Broude, N.E., Petrukhin, K.E., Grishin, A.V., Arzamazova, N.N., Aldanova, N.A., Monastyrskaya, G.S. and Sverdlov, E.D. (1986) *FEBS Lett.* 210, 237–245.
- [29] Karlisch, S.J.D., Goldshleger, R., Tal, D.M., Capasso, J.M., Hoving, S. and Stein, W.D. (1992) *Acta Physiol. Scand.* 146, 69–76.
- [30] Jewell, E.A. and Lingrell, J.B. (1993) *Biochemistry* 32, 13523–13530.
- [31] Arguello, J.M. and Kaplan, J.H. (1994) *J. Biol. Chem.* 269, 6892–6899.
- [32] Van Huysse, J.W., Jewell, E.A. and Lingrell, J.B. (1993) *Biochemistry* 32, 819–826.
- [33] Canessa, C.M., Horisberger, J.D., Louvard, D. and Rossier, B.C. (1992) *EMBO J.* 11, 1681–1687.
- [34] Schultheis, P.J. and Lingrel, J.B. (1993) *Biochemistry* 32, 544–550.
- [35] Canessa, C.M., Horisberger, J.D. and Rossier, B.C. (1993) *J. Biol. Chem.* 268, 17772–17726.
- [36] Burns, E.L. and Price, E.M. (1993) *J. Biol. Chem.* 268, 25632–25635.
- [37] Schultheis, P.J., Wallick, E.T. and Lingrel, J.B. (1993) *J. Biol. Chem.* 268, 22686–22694.
- [38] Price, M. and Lingrell, J.B. (1988) *Biochemistry* 27, 8400–8408.
- [39] Price, M., Rice, A.D. and Lingrell, J.B. (1990) *J. Biol. Chem.* 265, 6638–6641.
- [40] Sachs, G., Shin, J.M., Besancon, M., Munson, K. and Hersey, S. (1992).
- [41] Burgener-Kairuz, P., Horisberger, J.D., Geering, K. and Rossier, B.C. (1991) *FEBS Lett.* 290, 83–86.
- [42] Weizberski, W. and Blostein, R. (1993) *Proc. Natl. Acad. Sci. USA* 90, 70–74.
- [43] Vasilets, L.A., Omay, H.S., Ohta, T., Noguchi, S., Kawamura, M. and Schwartz (1991) *J. Biol. Chem.* 266, 16285–16288.