Polymorphism and structure of the gene coding for the α1 subunit of the *Artemia franciscana* Na/K-ATPase

Alberto GARCÍA-SÁEZ, Rosario PERONA and Leandro SASTRE*

Instituto de Investigaciones Biomédicas del CSIC, C/Arturo Duperier, 4, 28029 Madrid, Spain

Genomic clones coding for one of the two identified *Artemia franciscana* Na/K-ATPase α subunits, the α1 subunit, have been isolated. Several overlapping clones were obtained, although their restriction maps showed a large heterogeneity. Sequencing of their exons showed that they differ in up to 3.46% of their nucleotides in translated regions and 8.18% in untranslated regions. Southern blot analysis of DNA purified from different lots of *A. franciscana* cysts and from isolated individuals suggests that the variation is due to the existence of multiple Na/K-ATPase α1 subunit alleles in *A. franciscana*. The Na/K-ATPase α1 subunit gene is divided into 15 exons. Ten of the 14 introns are located in identical positions in this gene as in the human Na/K-ATPase α3 subunit gene. Analysis of the 5′ flanking region of the gene has allowed identification of the transcription-initiation sites. The adjacent upstream region has been shown to have functional promoter activity in cultured mammalian cells, suggesting the evolutionary conservation of some of the promoter regulatory sequences.

INTRODUCTION

Na/K-ATPase is the enzyme responsible for the exclusion of three sodium ions from the cell in exchange for two potassium ions at the expense of the hydrolysis of one ATP molecule. This enzymic activity is necessary for the maintenance of cell-membrane potential and mediates important functions of the animal cell, such as transmembrane transport and the maintenance of osmotic equilibrium and cell volume. In multicellular organisms the enzyme is also involved in basic functions, such as osmoregulation and transmission of the nervous signal [1].

The importance of this enzyme has stimulated its detailed study in several organisms. The enzyme is composed of two subunits, the α subunit, of about 110 kDa, and the β subunit, of 45–60 kDa [1]. The α subunit contains most of the active centres of the enzyme [2,3] and is homologous to other ion-transporting ATPases, such as the sarco/endoplasmic reticulum Ca-ATPase, the plasma membrane Ca-ATPase, the H/K-ATPase or the plant and yeast H-ATPase [4]. The β subunit is specific for the Na/K-ATPase and the highly related H/K-ATPase [5] and is required for enzyme assembly and transport to the plasma membrane [6,7]. In addition, part of the potassium-binding region has been located in the β subunit [8].

The Na/K-ATPase has been extensively studied in the crustacean *Artemia* (brine shrimp), because of its capacity to live under conditions of high salinity [9]. In extreme hyperosmotic conditions these animals are able to maintain an internal isosmotic medium through the activity of specialized organs that actively secrete ions to the environment [10]. The main osmoregulatory organs are the salt gland, in the larvae, and appendage glands in larvae and adults [10,11]. These organs express high levels of Na/K-ATPase activity that increase with the concentration of salt present in the external medium [10,12]. High levels of Na/K-ATPase activity are also expressed in the *Artemia* midgut and seem to be required for water uptake [10].

Biochemical studies allowed the identification in *Artemia* of two different Na/K-ATPase α subunits and one β subunit [13]. Later studies also led to the isolation of cDNA clones coding for two α subunits, that are 74% identical at the amino acid level, and one β subunit [14–16]. The existence of more than one α subunit gene has also been shown in vertebrates, where three functional genes have been characterized [17]. In contrast, another arthropod, *Drosophila melanogaster*, only has one α subunit gene [18]. Comparison of the amino acid sequences of these proteins suggests that vertebrate and invertebrate genes are derived from a unique ancestor gene and that gene duplication events occurred independently in *Artemia* and in vertebrates [5,15].

In *Artemia*, *in situ* hybridization experiments have shown that one of the α subunit clones, named pArATNa136, is expressed in the main larval osmoregulatory organs, salt gland, antennal gland and midgut, suggesting that the encoded protein plays an important role in salt tolerance [19]. This cDNA has been proposed to code for the α1 protein isoform because their tissue specificity and temporal pattern of expression during development are coincident. For similar reasons, the other identified cDNA clone, which is specifically expressed in the salt gland, has been proposed to code for the α2 protein isoform [20], although there are some discrepancies between patterns of expression of the protein and mRNA [19].

The suggested relevance of the α1 subunit in osmoregulation has prompted us to continue its study through the isolation of genomic clones coding for this gene. The establishment of the exon/intron structure might provide interesting data to further our understanding of the evolution of the gene. Furthermore, the isolation of genomic clones is necessary for the characterization of the regulatory regions that direct its tissue-specific and developmentally regulated expression.

MATERIALS AND METHODS

Isolation of genomic clones

A total of 7.5×10^5 independent clones from an *A. franciscana* genomic library [21] were screened using as probe the insert of

Abbreviation used: TFD, transcription-factor domain database.

* To whom correspondence should be addressed.

The nucleotide sequence data reported in this paper appear in the EMBL, Gen Bank and DDBJ Nucleotide Sequence Databases under the accession numbers X92862 and X92863.

the cDNA clone pArATNa136, coding for the α1 subunit of the Na/K-ATPase [15]. Filters were hybridized in 6 × SSC (SSC is 0.15 M NaCl/0.015 M sodium citrate), 50 % (v/v) formamide, 1 % (w/v) SDS, 5 × Denhardt's (0.02 % Ficoll 400/0.02 % polyvinylpyrrolidone/0.002 % BSA), 100 μg/ml calf thymus DNA and 10^6 c.p.m./ml of the probe at 42 °C for 14 h. The filters were washed twice in 2 × SSC/1 % SDS at 65 °C for 30 min and twice in 0.1 × SSC at room temperature, for 30 min each time.

Southern blot analysis

DNA was obtained from dechorionized *A. franciscana* cysts of the following San Francisco Bay Brand (Newark, CA, U.S.A.) lots: lot number 3556 (named L1 in this paper), lot 3018 (L2), lot 1250 (L3) and lot 1808 (L4). Cysts from lot L1 were also grown to adult animals in axenic media [22] or in reconstituted sea water, using baking yeast as food. DNA was purified according to the method of Cruces et al. [23]. A 15 μg amount of DNA from cysts from each lot, or variable but small amounts (< 2 μg) of DNA from the adults, were digested with the restriction enzymes indicated in each experiment, analysed on 0.8 % agarose gels and transferred to nylon membranes. Filters were hybridized with 10^6 c.p.m./ml of the different probes in 7 % SDS/500 mM sodium phosphate/1 mM EDTA, pH 7.2, at 65 °C for 15 h, and then washed three times in 0.1 × SSC/1 % SDS at 65 °C for 30 min.

Nucleotide sequencing

The nucleotide sequence (both sense and anti-sense) of the gene exons was determined after cloning the different genomic fragments in the plasmid vectors pUC18 [24] or pBluescript (Stratagene). The sequence was determined by the dideoxy chain-termination method using the Taq dye deoxy terminator cycle sequencing kit and the 373A sequencer from Applied Biosystems. Primers derived from plasmid vector sequences, or internal primers derived from cDNA or genomic sequences, were utilized in the different experiments. Sequencing products used as markers in primer extension and Nuclease S1 protection experiments were prepared using [α-^{35}S]dATP as substrate and the ^{T7}Sequencing kit from Pharmacia.

Primer extension

Total RNA was isolated from *A. franciscana* cysts (lot L1) cultured for 10 h at 30 °C as previously described [25]. RNA (100 μg) was incubated with 5×10^5 c.p.m. of labelled oligonucleotide OliNa-1 (5′-CCCGAAGTTTTTCCTAACTGC-3′) [15] overnight at 30 °C, and the extension reaction was carried out with 50 units of avian myeloblastosis virus reverse transcriptase for 1 h at 42 °C [26]. Extension products were analysed in 8 % polyacrylamide/7 M urea sequencing gels. Sequencing products of the genomic clones primed with the same oligonucleotide used in primer extension were used as size markers.

Nuclease S1 protection experiments

DNA probes were prepared by extension of the radioactively labelled oligonucleotide OliNa-1 by the Klenow fragment of the DNA polymerase I, using fragments of genomic clones 36 and 72 as substrate. Extension products were digested with *Spe*I and the labelled fragment was purified using 6 % polyacrylamide/7 M urea gels. Nuclease S1 protection experiments were carried out as previously described [27]. Briefly, 100 μg of total RNA from cysts cultured for 10 h or from *Torula* yeast RNA were incubated with 10^5 c.p.m. of the probe overnight at 55 °C. The hybridization reactions were then digested with 150 units of Nuclease S1, or without Nuclease, for 45 min at 37 °C. Digested products were analysed in 8 % polyacrylamide/7 M urea gels. Sequencing reactions of the same genomic fragments, using the oligonucleotide OliNa1 as primer were used as size markers.

Transient expression experiments

A 1.4 kb fragment from genomic clone 72, from the transcription-initiation site to the closer upstream *Eco*RI site, was amplified by PCR, using the oligonucleotide OriNa (5′-CGAGATCTGAGT-GAAGAGGCCAAAGTGC- 3′), that included at its 5′-end a *Bgl*II site, and the −20 universal sequencing primer. A similar fragment of 1.8 kb was amplified from genomic clone 36, from the same oligonucleotide OriNa to an immediately upstream *Xba*I site. Both fragments were digested with *Bgl*II and either *Eco*RI or *Xba*I, and cloned in the plasmid vector pXP1 containing the luciferase reporter gene [28]. The absence of mutations in the amplified products was confirmed by nucleotide sequencing. Purified DNAs from these constructs were transfected into the indicated cultured cell lines by the calcium phosphate method [29]. After transfection, cells were cultured for 24 h in growth media containing 90 % Dulbecco's modified Eagle's medium and 10 % (v/v) calf serum (GIBCO). Cell extracts were prepared as described by Murguía et al. [30]. Luciferase activity was determined with a commercial kit from Promega, according to the manufacturer's instructions. The level of induction was determined by calculating the ratio of the 36 or 72 plasmid vector to the empty vector.

RESULTS

Isolation of genomic clones coding for the α1 subunit of the *A. franciscana* Na/K-ATPase

Genomic clones were isolated by screening 7.5×10^5 independent clones of a genomic library made from *A. franciscana* DNA (lot L1) in the λEMBL-3 phage vector [21], using the α1 subunit cDNA clone pArATNa136 [15] as probe. Several positive clones were isolated and 15 of them were shown to hybridize to different fragments of the cDNA clone and characterized further. The restriction maps of these 15 clones were established by single and double digestions with the enzymes *Eco*RI, *Hin*dIII and *Sal*I and by the end labelling method of Rackwitz et al. [31], and are shown in Figure 1. Except for some groups, there were few restriction patterns conserved between the different clones. To establish the relation between them it was necessary to determine the exons contained in each clone.

The restriction fragments of the genomic clones containing coding regions were identified by hybridization to cDNA probes, and many of them were cloned in plasmid vectors. The different exons were identified by hybridization with oligonucleotide probes derived from the cDNA sequence, and their nucleotide sequences were determined using either universal primers or internal primers synthesized according to the cDNA or to adjacent genomic sequences. The position of the exon on the genomic fragment was determined either from the nucleotide sequence data or by PCR, utilizing primers complementary to one end of the plasmid vector and to the exon being studied. The exons that have been located are indicated as boxes in Figure 1. The results obtained showed that each exon had the same limits in all the phages where it was studied. The position on the cDNA sequence of the 14 introns identified from these clones is indicated in Figure 2. The sequences of intron/exon borders present a high similarity to the consensus donor or acceptor splicing signals

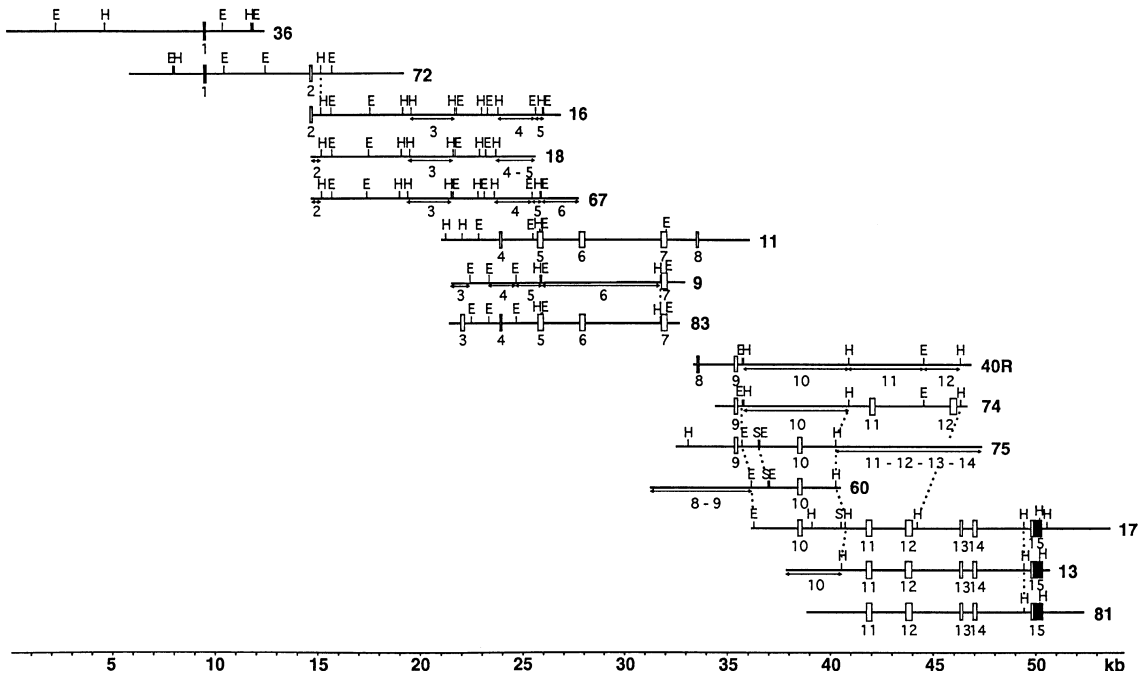


Figure 1 Restriction map of the genomic clones coding for the α1 subunit of the *A. franciscana* Na/K-ATPase

The restriction map of 15 of the genomic clones isolated coding for the α1 subunit of the Na/K-ATPase was determined using the enzymes *Eco*RI (E), *Hin*dIII (H) and *Sal*I (S). The number assigned to each clone is shown to the right of the corresponding map. The different clones presented large differences in their restriction maps and have been aligned by the exons they contain, from the 5′-end of the mRNA (left) to the 3′-end (right). The exons contained in each restriction fragment were determined by hybridization to fragments of the cDNA clones, by nucleotide sequencing and by PCR experiments, and are indicated as boxes on the restriction maps of the genomic clones where they have been found. Black boxes indicate exons coding for untranslated regions of the mRNA and open boxes exons coding for translated regions. The number of each exon is indicated under the corresponding box. Exons that have been assigned to a restriction fragment but have not been exactly localized are indicated as horizontal lines with two arrow heads. Vertical broken lines join restriction sites that have been found to be identical in two or more genomic clones. The lower scale indicates the size of the clones in kb.

[32]. Intron sizes are also shown in Figure 2. In general, intron sizes are conserved between clones, but there are some exceptions, such as intron 11 whose size has been estimated as 1.6 kb in clone 17 and 3.6 kb in clone 74.

The coding sequences derived from the different genomic clones were compared with each other and with the cDNA clone pArATNa136; the differences observed are also shown in Figure 2. Only two genomic clones were identical in the regions compared, clones 13 and 81, although their restriction maps differed at a *Hin*dIII site. The nucleotide sequence of exon 10 was also identical in genomic clones 60 and 75. Except for these regions, the genomic clones studied differed from each other and from cDNA clone 136 in between 0.41 and 3.46% of their nucleotide sequences in translated regions and between 1.57 and 8.18% in untranslated regions. The changes in nucleotide sequence result in 12 changes in the predicted amino acid sequences (Figure 2). All these changes are conservative, except for a Ser/Asn change at amino acid 419 and a Cys/Ser change at amino acid 857.

The large differences observed between the genomic clones raised the possibility that they might code for different, although very similar, genes. Alternatively, the different clones could code for several alleles of the same gene. To study these possibilities, DNAs from different commercial lots of *A. franciscana* cysts were analysed by Southern blotting, using fragments of the genomic clones as probes. A representative example of the results obtained is shown in Figure 3(A), where DNAs obtained from four different lots of cysts were hybridized to the *Hin*dIII–*Hin*dIII fragment of genomic clone 17, which contains the 5′ region of exon 15. Even if the probe was a small *Hin*dIII–*Hin*dIII fragment, two or three different *Hin*dIII–*Hin*dIII fragments and two to four fragments obtained after *Eco*RI–*Hin*dIII digestion hybridized in each DNA sample. These results are in agreement with the heterogeneity found in the restriction maps of the genomic clones. The hybridization patterns observed were different in the analysed cyst lots, so that *Hin*dIII–*Hin*dIII and *Eco*RI–*Hin*dIII fragments present in lots L2 and L3 were absent from L1 and L4 (indicated by arrows in Figure 3A). These results suggest that the heterogeneity observed is due to allelic variation rather than to the existence of several Na/K-ATPase α1 subunit genes, which would be expected to be the same in all *A. franciscana* populations. As an internal control, the same blot was washed and hybridized with a probe from a different *A. franciscana* gene, the 1 kb *Hin*dIII–*Hin*dIII fragment that contains exon 7 of the sarco/endoplasmic reticulum Ca-ATPase gene, isolated from the previously characterized genomic clone gArATCa23 [33]. The results obtained are shown in Figure 3(B): only one restriction fragment hybridized to the probe in each lane and the hybridization pattern was the same in the four cyst lots analysed.

The variability of the gene was further studied by analysing the restriction patterns of DNAs isolated from individual *A. franciscana* adults cultured from lot 1 cysts. DNAs obtained from 13 individuals were digested with *Eco*RI and *Hin*dIII and analysed by Southern blotting using the same *Hin*dIII–*Hin*dIII probe from clone 17, containing part of exon 15, described above. The results obtained are shown in Figure 4. Some individual DNAs (lanes 1, 2, 3, 6, 7, 10 and 11 of Figure 4) contain the same two restriction fragments of 2.7 and 0.8 kb as

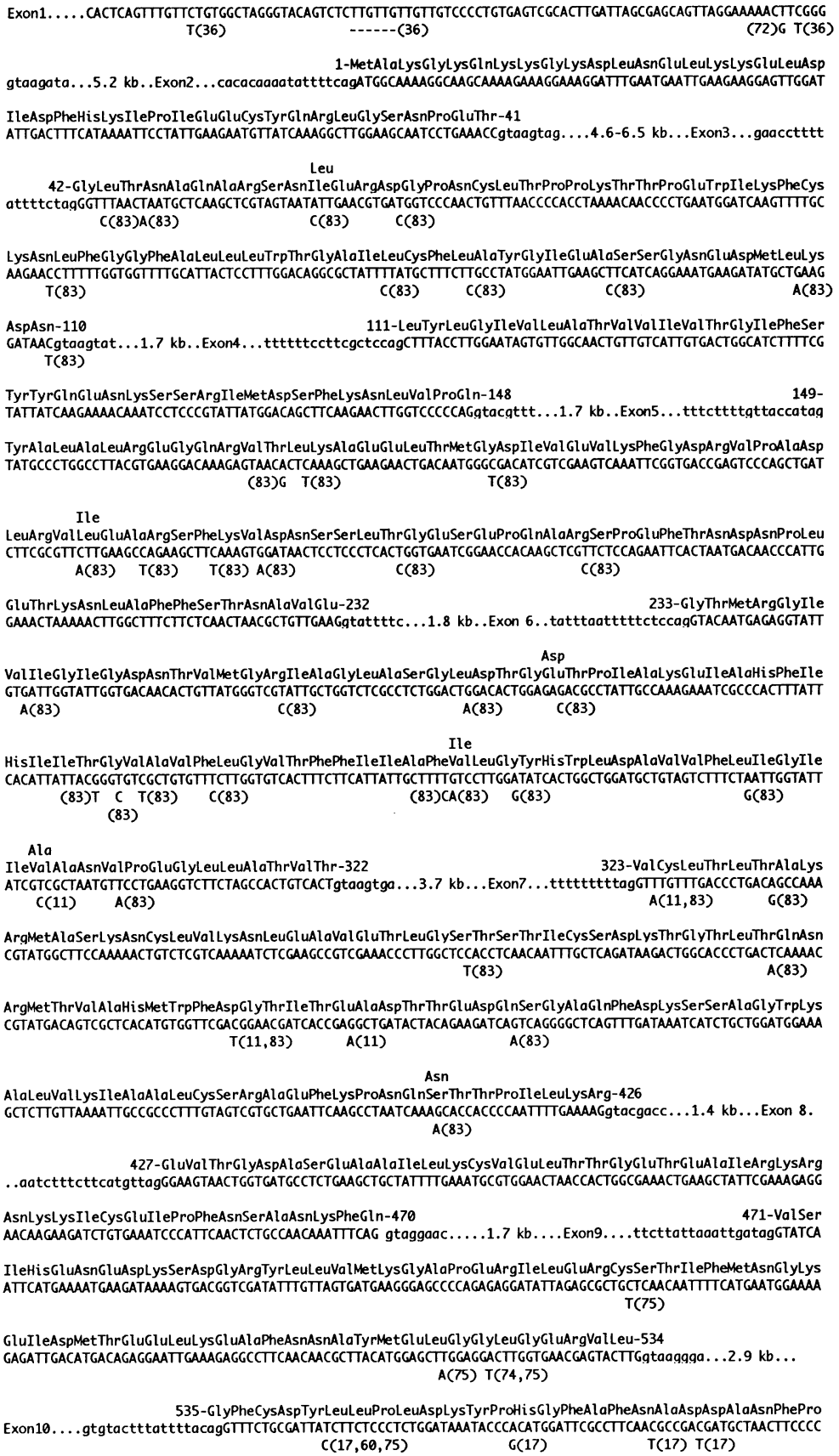
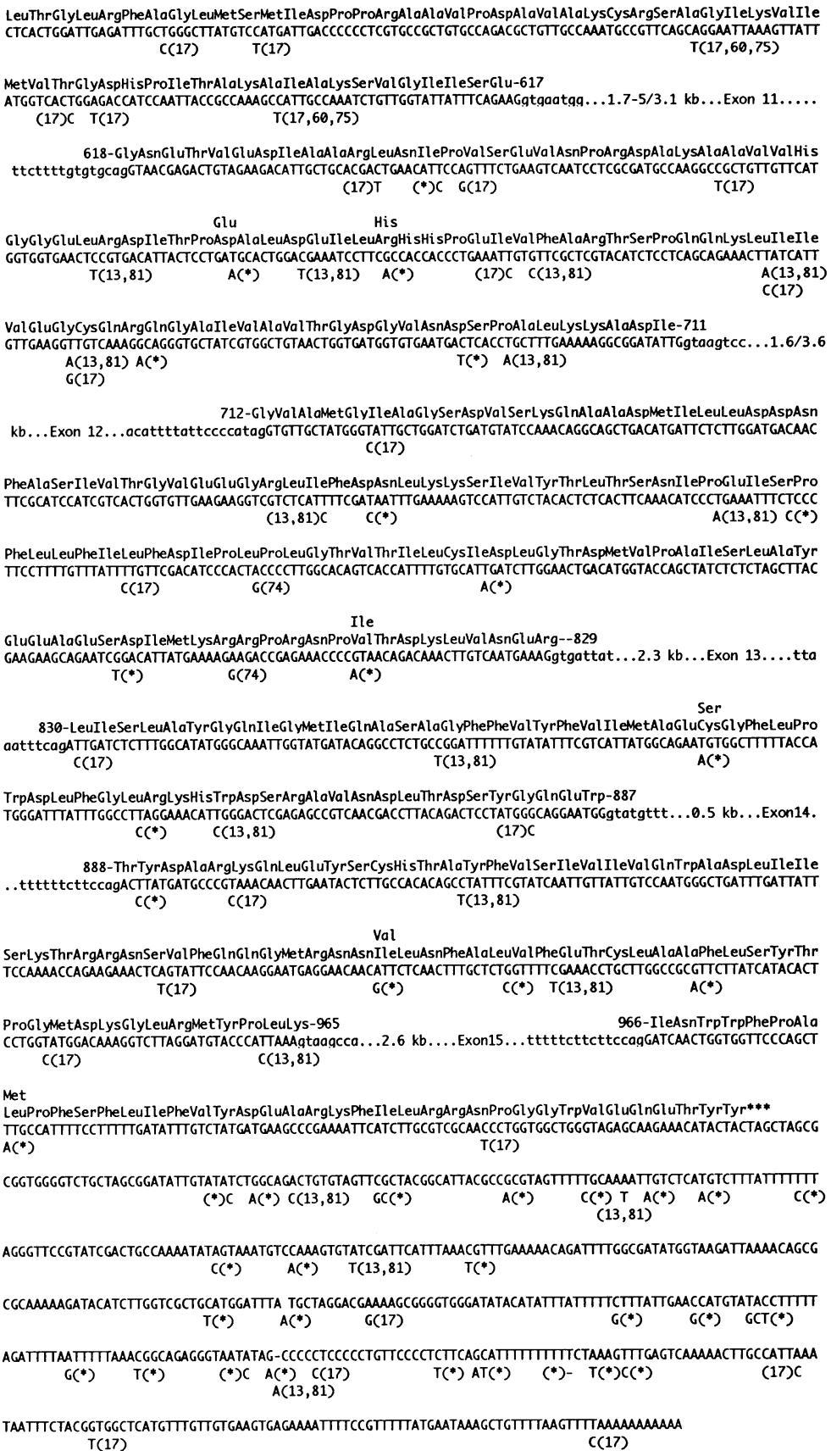


Figure 2 Comparison of the nucleotide sequence of the Na/K-ATPase α1 subunit cDNA and genomic clones, with indication of intron positions and sizes

The coding nucleotide sequence shown in uppercase letters corresponds to the cDNA clone pArATNa136 [15], except for the first 40 nucleotides of exon 1, which were determined from genomic clone 72. The deduced amino acid sequence of the cDNA clone is shown above the nucleotide sequence. The differences in nucleotide sequence found in the genomic clones are indicated on the lower



line. The number of the genomic clone where each difference was found is indicated in parentheses. (*) indicates differences found between the cDNA and genomic clones 13, 17 and 81. Differences in the deduced amino acid sequences are indicated above the cDNA amino acid sequence. The intron sequences shown (lowercase letters) correspond to genomic clones 36 (intron 1), 72 (introns 1 and 2), 16 (introns 1 and 2), 83 (introns 2–7), 11 (introns 3–8), 75 (introns 8–10), 17 (introns 9–15), 13 (introns 10–15) and 81 (introns 10–15). Small differences in intron sequences between genomic clones are not indicated. Intron sizes were determined by nucleotide sequencing, PCR or from the restriction maps. The different sizes estimated for intron 11 from clone 17 (1.6 kb) and 74 (3.6 kb) are indicated.

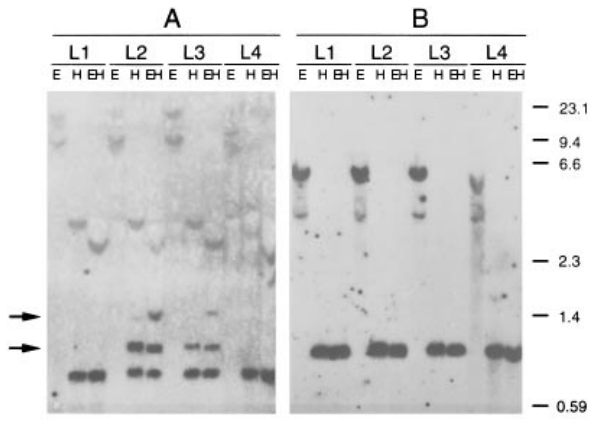


Figure 3 Southern blot analysis of DNAs obtained from four different *A. franciscana* cyst lots

Four different commercial lots of *A. franciscana* cysts (lanes L1, L2, L3 and L4) were used to purify DNAs that were digested with the restriction enzymes *Eco*RI (lanes E), *Hin*dIII (lanes H) or both (lanes EH). The digested DNAs (15 μg) were analysed on 0.8% agarose gels and transferred to nylon membranes. The membrane shown in (**A**) was hybridized to the *Hin*dIII–*Hin*dIII fragment from genomic clone 17, which contains exon 15 of the α1 subunit of the Na/K-ATPase. The arrows indicate two restriction fragments that are present in lots L2 and L3 and not in L1 and L4. The membrane shown in (**B**) was hybridized to a *Hin*dIII–*Hin*dIII fragment that contains exon 7 of the *A. franciscana* sarco/endoplasmic reticulum Ca-ATPase gene. Migration of the size markers, indicated in kb, is shown at the right of the Figure.

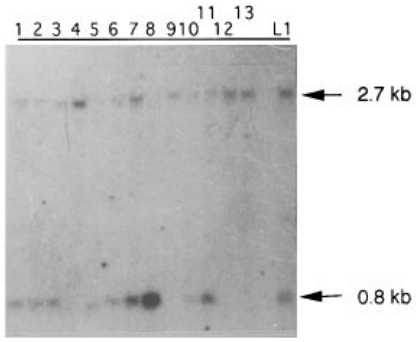


Figure 4 Southern blot analysis of DNAs isolated from individual *A. franciscana* adult animals

DNAs were prepared from 13 *A. franciscana* adult animals (lanes 1–13) or from the original cyst population (lane L1). Purified DNAs were digested with *Eco*RI and *Hin*dIII, analysed on 0.8% agarose gels, transferred to nylon membranes and hybridized with the *Hin*dIII-*Hin*dIII fragment from genomic clone 17, which contains exon 15 of the α1 subunit of the Na/K-ATPase. The hybridizing fragments of 2.7 and 0.8 kb are indicated.

the original cyst lot (lane L1 in Figure 4). Other individual DNAs exclusively contain the 2.7 kb fragment (lanes 4, 9, 12 and 13 of Figure 4) or the 0.8 kb fragment (lanes 5 and 8 of Figure 4). Similar results were obtained using other fragments of the Na/K-ATPase genomic clones as probes, except that some of them detected more than two polymorphic restriction fragments (results not shown). In contrast, the hybridization of a similar blot with the sarco/endoplasmic reticulum Ca-ATPase probe described above, showed the same pattern of hybridization for all the 15 individual DNAs analysed (results not shown). The results obtained with the α1 subunit Na/K-ATPase probes are in agreement with the distribution expected for different alleles of a single gene, where homozygous individuals for each of the alleles and heterozygous individuals are observed. These results are, however, incompatible with the existence of several independent genes in each individual.

The data described above suggest that the genomic clones isolated are derived from different alleles of a single gene and that the restriction maps shown in Figure 1 and the nucleotide sequences of Figure 2 do not correspond to any particular allele and are a combination of parts of several. Summarizing the data from all the genomic clones, the gene is composed of 15 exons, as shown in Figures 1 and 2. The size of the gene can be estimated from these data as 40–45 kb, although the detected differences in intron size between alleles indicates that the size of this gene is variable in the population studied.

The structure of the *A. franciscana* α1 subunit gene has been compared in Figure 5 with that of the human α3 gene [34], which is almost identical to that of horse α1 and human α2 genes [35,36]. The positions of the introns in the deduced amino acid sequences are indicated by arrows. Ten of the 14 *A. franciscana* introns interrupt the gene in the same position as introns in the human gene. It is interesting that most of the intron differences are found in the region between *A. franciscana* exons 10 and 13 (exons 13 and 19 in the human gene), where both genes have a very different intron/exon structure.

The positions of the introns in relation to suggested structural or functional regions of the protein [37] are also indicated in Figure 5. Two types of regions have been considered, the proposed transmembrane regions (H1–H10 in Figure 5) and regions conserved between different ion-transporting ATPases that have been proposed to have functional relevance [37] (a–j in Figure 5). These domains are coincident for *Artemia* α1 and human α3 subunits. Intron positions that interrupt transmembrane or conserved domains, marginally or centrally, are indicated by one or two asterisks respectively. Only one of the three introns that are specific for the *Artemia* α1 gene, intron 11, interrupts marginally one conserved domain (domain i). In contrast, three of the human α3 specific introns interrupt centrally two conserved domains (domains h and j) and one proposed transmembrane domain (H7), and one other intron interrupts marginally the conserved domain j.

Structural and functional analysis of the 5′ region of the gene

Characterization of the gene was continued by the determination of the transcription-initiation site. Primer extension analysis was carried out utilizing total RNA from *A. franciscana* cysts cultured for 10 h as substrate, and the oligonucleotide OliNa-1, complementary to the 5′ untranslated region of the mRNA, as primer (Figure 6A). Several fragments were generated; the longest ones correspond to initiation at the nucleotides indicated by arrow heads on the nucleotide sequences of clones 36 and 72. These sequences are not identical in both clones because of the existence of six additional nucleotides in the 5′ untranslated region of clone 72 with respect to clone 36.

Nuclease S1 protection experiments with probes derived from clones 36 and 72 were carried out to solve this uncertainty. The results obtained, shown in Figure 6(B), indicate that both probes specifically protected several fragments from digestion. The largest fragments correspond to initiation at identical nucleotides in both genomic clones, and are indicated by arrows in Figure 6(B). Three of the six initiation sites determined by Nuclease S1 protection are coincident with those determined by primer extension when assigned according to the nucleotide sequence of clone 72. These results suggest that the mRNA of the allele represented by clone 72 is more abundant in the population studied than that of the allele represented by clone 36. The small

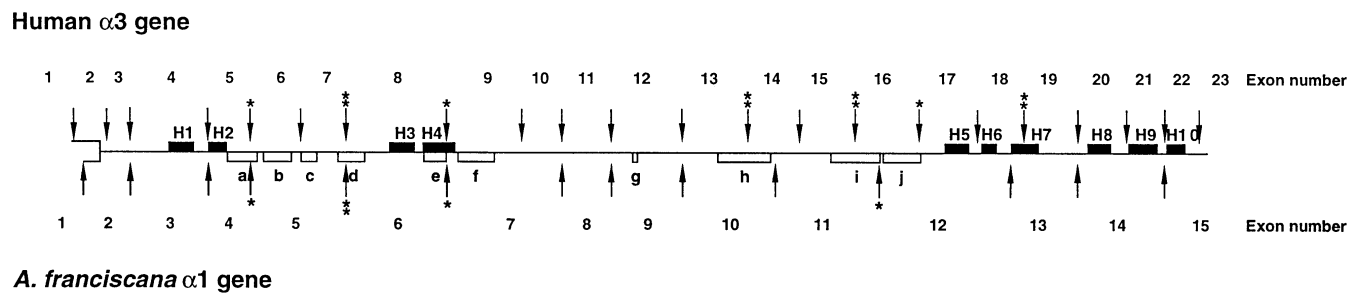


Figure 5 Comparison of the exon/intron structure of the human Na/K-ATPase α3 gene and the *A. franciscana* α1 gene

The positions of the introns in the amino acid sequence of the human Na/K-ATPase α3 subunit (upper diagram) and the *A. franciscana* α1 subunit (lower diagram) are represented by arrows. The locations of the proposed transmembrane regions (black boxes, named H1–H10) and putative functional regions conserved between ion-transporting ATPases (white boxes, named a–j), which are similar in both proteins, are shown. Introns that interrupt some of the structural or functional regions are indicated by asterisks. One asterisk denotes introns located in the margin of the domain and two asterisks introns that interrupt the domain centrally. The poorly conserved N-terminal part of the protein is indicated as two parallel lines; it is longer in the human α3 gene than in the *A. franciscana* α1 gene.

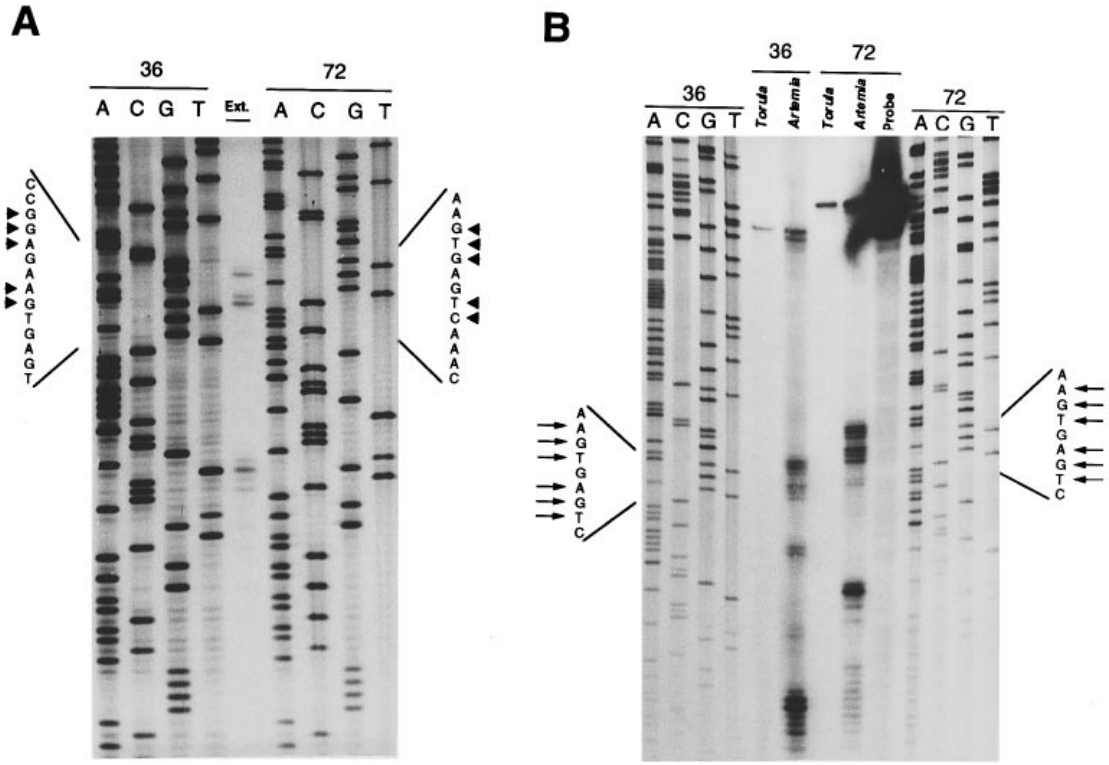


Figure 6 Analysis of the transcription-initiation site of the *A. franciscana* Na/K-ATPase α1 subunit gene

(**A**) Primer extension experiments made using 100 μg of total RNA isolated from *A. franciscana* cysts cultured for 10 h as substrate with the oligonucleotide OliNa-1 as primer (lane Ext). Lanes 36 and 72 show a sequencing reaction of the fragments of these clones that contain the first exon of the gene, primed with the oligonucleotide OliNa-1. The position of the longest extension products on the sequences of clones 36 and 72 are indicated by arrow heads at both sides of the Figure. (**B**) The results obtained in Nuclease S1 protection experiments using as substrate 100 μg of total RNA obtained from *A. franciscana* cysts cultured for 10 h (lanes *Artemia*) or from *Torula* yeast (lanes *Torula*). The probes generated from genomic clone 36 (lanes 36) or 72 (lanes 72) were hybridized with the RNA, incubated with 150 units of Nuclease S1 and analysed on 8% polyacrylamide/7 M urea gels. The lane probe shows the migration of the clone 72 probe after hybridization to *A. franciscana* RNA and incubation without Nuclease S1. Sequencing reactions of clones 36 and 72 (lanes A, C, G, T), using OliNa1 as primer, were used as size markers. The positions of the largest protection products are indicated with arrows on the nucleotide sequence of the corresponding genomic clone.

differences found between the transcription-initiation sites determined by primer extension and Nuclease S1 protection techniques could be due to experimental artifacts, since the results differ by only one or two nucleotides. Alternatively, these differences could also be due to the existence of other major α1 subunit alleles slightly shorter in their 5′ untranslated sequence than that represented by clone 72.

The transcription-initiation sites determined are located in a directly repeated pentanucleotide TCACT/TCAGT. Initiation at the adenosine of the pentanucleotide has been detected in both

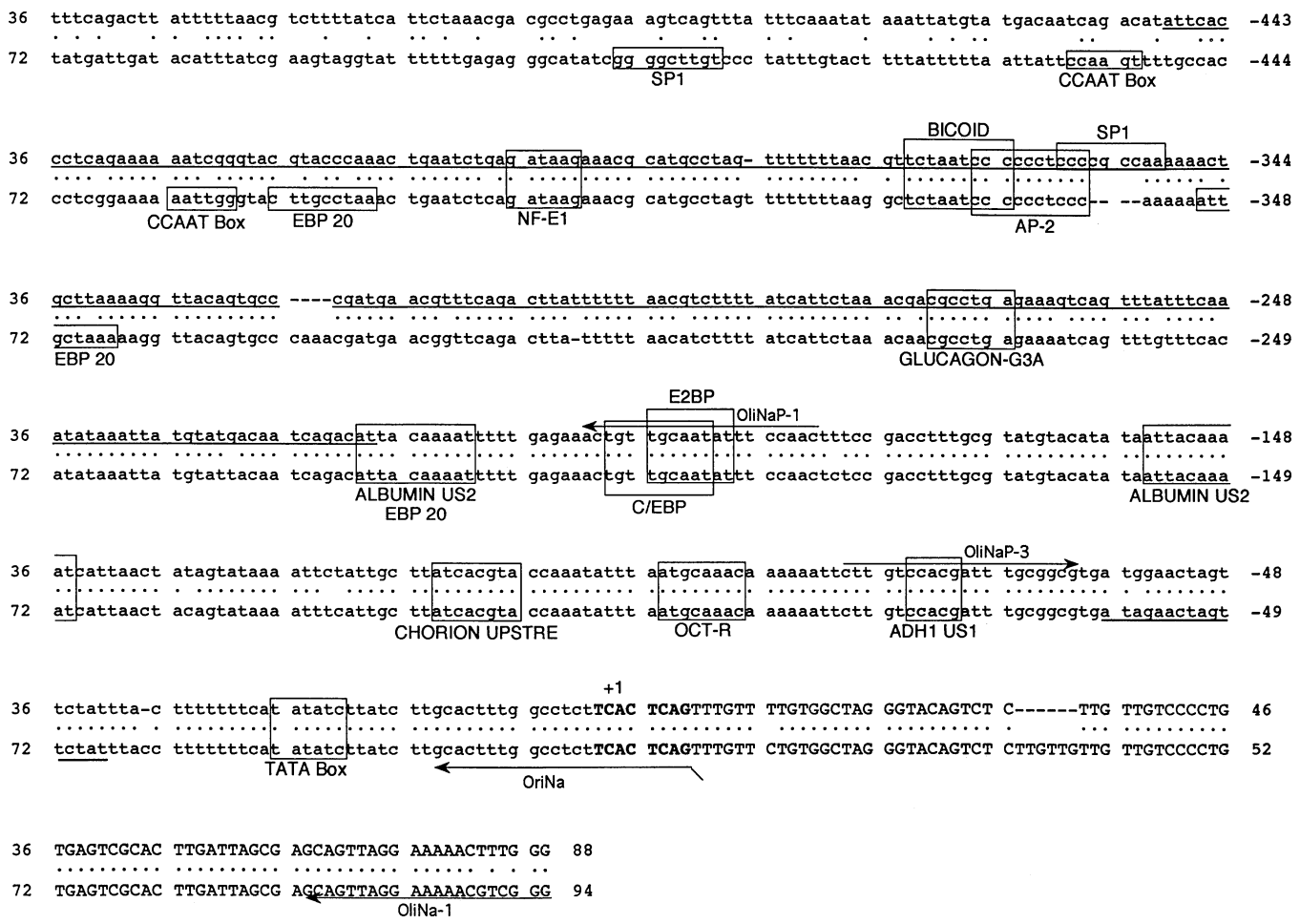


Figure 7 Nucleotide sequence of the 5′ flanking region of the *A. franciscana* Na/K-ATPase α1 subunit gene

The nucleotide sequence of the 5′ flanking region of the gene was determined for genomic clones 36 (upper line) and 72 (lower line). The nucleotide sequences have been numbered from one of the determined transcription-initiation sites (indicated as +1). Nucleotides identical for both clones are indicated by dots. Dashes indicate gaps introduced on the nucleotide sequences to increase their similarity. Oligonucleotides utilized in primer extension and Nuclease S1 protection experiments (OliNa-1) in the cloning of these regions (OriNa), or as primer to establish the nucleotide sequences (OliNaP-1, OliNaP-3) and their orientation, are indicated by arrows. The fragment that is repeated in the upstream region of clone 36 is underlined. The sequence underlined between nucleotides −44 and −59 contains an inverted repeat. Nucleotide sequences were compared with those contained in the TFD, and the similarities considered most significant are indicated by boxing the DNA domain and indicating the name of the consensus sequence or the binding transcription factor below or above the box.

repeats by primer extension and Nuclease S1 protection. Initiation at the first cytosine of the first repeat has also been shown by both techniques. The repeated oligonucleotide has some similarity to the two more general initiation consensus sites, the CAP site (CANC/T) [38] and the initiator site (G/A/T, C/T, A, G/T, T) [39].

The nucleotide sequence of a fragment of about 650 nt upstream of intron 1 from clones 36 and 72 is shown in Figure 7. The nucleotide sequences of the two clones are very similar up to nucleotide −445 (in clone 36) and −446 (in clone 72), but diverge in more upstream regions. This divergence is due to the repetition in clone 36 of a 230 nt fragment, which is underlined in Figure 7. Sequencing of more upstream regions of clone 36 has shown that there are five consecutive direct repetitions of this fragment. The nucleotide sequence of clone 36 upstream of these repetitions is similar to the nucleotide sequence present in clone 72 upstream of this fragment. Comparison of these sequences with the transcription regulatory sequences contained in the transcription-factor domain database (TFD) [40], using the MacPattern program [41], detected several putative regulatory sites that have been boxed in Figure 7. About 25–30 nt upstream of the initiation sites there are putative TATA boxes in both sequences. Other putative regulatory sites present in the two sequences are binding sites for general transcription factors such as SP-1, AP-2, OctR, EBP 20 and C/EBP. Some other regions have similarity with tissue-specific regulatory elements, such as glucagon G3A, Albumin US-2, NF-E1, ADH1 US1 or Chorion Upstream. Finally, there is a developmental regulatory site (Bicoid). There are three more general regulatory sites in the sequence of clone 72 that are not present in the sequence shown for clone 36, one SP-1 site, two CCAAT boxes and one EBP 20 site, while in 36 there is an SP-1 consensus site that is absent in clone 72.

The promoter activity of the regions upstream from the transcription-initiation site was tested in transient expression experiments. A 1.4 kb fragment from clone 72 and a 1.8 kb fragment from clone 36 were amplified by PCR using the oligonucleotide OriNa (underlined in Figure 7), whose 5′-end was complementary to the transcription-initiation region, and oligonucleotides complementary to the plasmid vector. The

Table 1 Relative levels of reporter gene activity obtained after transfection of the indicated cell lines with plasmid vectors where transcription of the reporter gene was driven by the 5′ flanking region of the Na/K-ATPase α1 subunit clones 36 or 72 or in the absence of any regulatory region (None)

The data for the NIH 3T3 cell line summarize six independent experiments, and those on the MDCK and HEK 293 cell lines, two independent experiments; results given $\pm$S.D.

Regulatory region	Luciferase activity (fold induction)		
	NIH 3T3	MDCK	HEK 293
None	1	1	1
36	3.50 $\pm$ 1.21	2.75 $\pm$ 0.25	2.10 $\pm$ 0.10
72	8.56 $\pm$ 2.08	6.75 $\pm$ 1.25	3.85 $\pm$ 0.55

amplified fragments were cloned in the expression vector pXP1 [28] that contains luciferase as a reporter gene.

In the absence of cultured *Artemia* cell lines, three mammalian cell lines were transfected with the constructs described above, a mouse fibroblastoid cell line (NIH 3T3) and two cell lines derived from kidney (MDCK and HEK 293). The results obtained are shown in Table 1 where the activity of the reporter gene is expressed relative to the activity of the vector alone, without any promoter region.

The results shown in Table 1 indicate that the 5′-regions of clones 36 and 72 were able to promote transcription in the three mammalian cell lines tested. The cloned promoter region of clone 72 gave about twice the activity of the clone 36 promoter region in all the experiments, which might be related to the existence of the five direct repeats in clone 36.

DISCUSSION

Overlapping genomic clones containing the complete coding sequence of the α1 subunit of the *A. franciscana* Na/K-ATPase have been isolated and characterized. The gene is 40–45 kb long and is divided into 15 exons. A large heterogeneity has been found between both the different genomic clones isolated and the previously characterized cDNA clone pArATNa 136. The heterogeneity was reflected in the restriction map of the clones and made it impossible to align them by the criterion of restriction site overlap. Sequencing of the exons contained in the genomic clones also showed up to 3.5 % nucleotide sequence divergence in their protein coding regions. The results described in this paper strongly suggest that the variability observed is due to the existence of several alleles for this gene. This hypothesis would explain the different restriction patterns found in the cyst lots analysed. Furthermore, the presence of several alleles would also explain the variations observed in the restriction patterns in individuals from the same cyst lot.

Similar variability had not been observed in any other of the *Artemia* genes that have been analysed, such as the elongation factor 1α gene [42], the sarco/endoplasmic reticulum Ca-ATPase gene [33] and several actin genes [43], even though the genomic clones for the last two genes were obtained from the same library used in this study. Overlapping genomic clones isolated for these genes showed similar restriction maps, and the divergence observed between the nucleotide sequence of genomic and cDNA clones was smaller than 0.5 % for these genes. In fact, the level of nucleotide sequence divergence found between Na/K-ATPase α1 subunit alleles is similar to the level found between actin genes of two different *Artemia* species, *A. franciscana* and *A. parthenogenetica* [43].

As mentioned in the Introduction, the Na/K-ATPase α1 subunit is expressed in the main osmoregulatory organs of *Artemia* larvae, which suggests that it might play an important role in the osmoregulatory processes that allow *Artemia* to survive in saline media of very diverse composition, including high salt concentrations. It is tempting to speculate that the existence of so many diverse Na/K-ATPase α1 subunit alleles in an *A. franciscana* population might facilitate the survival of some of the individuals in any of these different environmental conditions. The existence of several amino acid differences between some of the clones, two of them in the proposed domains e (Val $\rightarrow$ Ala; exon 6) and H10 (Leu $\rightarrow$ Met; exon 15), could indicate the existence of functional differences between alleles, although there is no experimental evidence to support this suggestion.

As for the structure of the gene, it is divided into 15 exons with an average size of 240 nt, similar to the exon size found in other *Artemia* genes, such as the sarco/endoplasmic reticulum Ca-ATPase gene (230 nt), actin genes (233 nt) and the elongation factor EF-1α gene (325 nt) mentioned above. The average intron size is 2.6 kb, which is also similar to the size found in the other *Artemia* genes: 3.6 kb for the sarco/endoplasmic reticulum Ca-ATPase gene, 2.5 kb for actin genes and 1.9 kb for the elongation factor EF-1α gene.

The only other Na/K-ATPase α-subunit genes that have been characterized are the human α2 and α3 genes and the horse α1 gene, which have an almost identical structure [34–36]. The three vertebrate genes are divided into 23 exons although the size of these genes (25–30 kb) is smaller than that of the *Artemia* gene (40–45 kb) due to the larger size of the *Artemia* introns. A comparison of intron positions has shown that 10 of the 14 *Artemia* introns (71 %) are in identical positions as in the human α3 gene. It is interesting to note that intron 7 of the *Artemia* gene analysed and intron 10 of the human α3 gene are located in the same nucleotide position, but two nucleotides downstream from the equivalent horse α1 and human α2 introns. *Artemia* and vertebrate genes also have one intron in the 5′ untranslated region, but the lack of sequence similarity in these regions makes it impossible to know if their positions are related. The other three introns of the *Artemia* gene and ten introns of the human gene are located in different places. These results suggest than the ten introns that are similarly located in *Artemia* and humans and, perhaps, the one located in the 5′ untranslated region, might have existed in the ancestral Na/K-ATPase α subunit gene previous to protostoma/deuterostoma divergence. The other introns might have been either inserted or deleted in *Artemia* and vertebrate genes after their divergence. The three *Artemia* introns and six of the human introns that are different are consecutive in the gene, so that this gene region seems to have diverged more in its structure than the other regions of the gene. However, the amino acid sequence encoded by this gene region is very similar in *Artemia* and in vertebrates. Besides, two of the human-specific introns located in this region interrupt putative functional domains conserved between P-type ATPases, and another interrupts a proposed transmembrane region, which, according to the exon theory of genes [44], might favour the idea of a recent origin for these introns.

The results obtained in comparing the structure of *Artemia* and vertebrate Na/K-ATPase α subunit genes are very similar to those previously obtained for the structure of the sarco/endoplasmic reticulum Ca-ATPase gene. The *Artemia* sarco/endoplasmic reticulum Ca-ATPase gene is divided into 18 exons and the rabbit sarco/endoplasmic reticulum Ca-ATPase-1 gene contains 21 exons. The positions of 12 of the 18 *Artemia* introns (66 %) are the same as in vertebrates [33]. This high percentage of

intron-position identity suggests similar evolution of the two ATPase genes. It is not clear why, in both cases, the vertebrate genes have conserved or accumulated more introns than the *Artemia* genes, although some authors have related intron density to the complexity of the developmental programme [45].

The last part of the work deals with the characterization of the promoter region of the gene. Transcription-initiation sites were determined by primer extension and Nuclease S1 protection experiments, identifying several transcription-initiation sites in two consecutive small repeats (TCAC/GT) whose sequence presented similarity to the consensus CAP and initiator sites.

Analysis of the nucleotide sequences upstream of the transcription-initiation sites in two genomic clones showed that both are very similar, except for the existence of five direct repeats of a 230 nt long fragment in clone 36. Comparison of the nucleotide sequences determined with the TFD showed the presence of several putative regulatory elements. However, the two regulatory elements that have been shown to have functional relevance in the human gene promoters, the ARE and E box regions [46,47], or those required for brain expression of the rat α3 gene [48], have not been identified in the *Artemia* 5′ flanking region.

The promoter activity of the *Artemia* 5′ flanking region has been tested by transfection into mammalian cells of constructs where the transcription of a reporter gene was driven by the 5′ flanking regions of clones 36 and 72. The two fragments gave transcriptional activities significantly higher than those obtained with the vector alone. These data demonstrate the promoter activity of the two 5′ flanking regions, and suggest the conservation of some of the transcription regulatory mechanisms between crustaceans and mammals.

We thank Maria Luisa Faraldo for donation of the plasmid vector pXP1, Jesús Cruces, Rafael Garesse and Claudio Fernandez de Heredia for donation of *Artemia* cysts, Ignacio Palmero, Ricardo Escalante and Maria-Asunción Ortega for their support and helpful discussions, and Antonio Fernandez and Ricardo Uña for preparation of the Figures. This work was supported by Grants PB92-0076 from the Dirección General de Investigación Científica y Técnica and 95/0882 from the Fondo de Investigaciones Sanitarias de la Seguridad Social.

REFERENCES

1 Mercer, R. W. (1993) in Molecular Biology of Receptors and Transporters (Friedlander, M. and Mueckler, M., eds.), pp. 139–168, Academic Press, New York
2 Vasilets, L. A. and Schwartz, W. (1993) Biochem. Biophys. Acta **1154**, 201–222
3 Lingrel, J. B. and Kuntsweiler, T. (1994) J. Biol. Chem. **269**, 19659–19662
4 Stokes, D. L., Taylor, W. R. and Green, N. M. (1994) FEBS Lett. **346**, 32–38
5 Fagan, M. J. and Saier, M. H. (1994) J. Mol. Evolut. **38**, 57–99
6 Fambrough, D. M., Lemas, M. V., Hamrick, M., Emerick, M., Renaud, K. J., Inman, E. M., Hwang, B. and Takeyasu, K. (1994) Am. J. Physiol. **266**, C579–C589
7 Eakle, K. A., Kabalin, M. A., Wang, S.-G. and Farley, R. A. (1994) J. Biol. Chem. **269**, 6550–6557
8 Chow, D. C. and Forte, J. G. (1995) J. Exp. Biol. **198**, 1–17
9 Browne, R. A. (1992) Tree **7**, 232–237
10 Holliday, C. W., Roye, D. B. and Roer, R. D. (1990) J. Exp. Biol. **151**, 279–296
11 Conte, F. P. (1984) Int. Rev. Cytol. **91**, 45–106
12 Cortas, N., Arnaout, M., Salon, J. and Edelman, I. S. (1989) J. Membr. Biol. **108**, 187–195
13 Peterson, G. L., Ewing, R. D., Hootman, S. R. and Conte, F. P. (1978) J. Biol. Chem. **253**, 4762–4770
14 Baxter-Lowe, L. A., Guo, J. Z., Bergstrom, E. E. and Hokin, L. E. (1989) FEBS Lett. **257**, 181–187
15 Macias, M.-T., Palmero, I. and Sastre, L. (1991) Gene **105**, 197–204
16 Bhattacharyya, K. K., Bergstrom, E. E. and Hokin, L. E. (1990) FEBS Lett. **269**, 233–238
17 Levenson, R. (1994) Rev. Physiol. Biochem. Pharmacol. **123**, 1–45
18 Levobitz, R. M., Takeyasu, K. and Fambrough, D. M. (1989) EMBO J. **8**, 193–202
19 Escalante, R., García-Sáez, A. and Sastre, L. (1995) J. Histochem. Cytochem. **43**, 391–399
20 Sun, D. Y., Guo, J. Z., Hartmann, H. A., Uno, H. and Hokin, L. E. (1992) J. Histochem. Cytochem. **40**, 555–562
21 Escalante, R. and Sastre, L. (1993) J. Biol. Chem. **268**, 14090–14095
22 Hernandorena, A. (1991) Reprod. Nutr. Dev. **31**, 57–63
23 Cruces, J., Sebastián, J. and Renart, J. (1981) Biochem. Biophys. Res. Commun. **98**, 404–409
24 Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene **33**, 103–119
25 Escalante, R., García-Sáez, A., Ortega, M.-A. and Sastre, L. (1994) Biochem. Cell Biol. **72**, 78–83
26 Triezenberg, S. J. (1993) in Current Protocols in Molecular Biology (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K., eds.), pp. 4.8.1–4.8.5, John Willey and Sons, Inc. New York.
27 Greene, J. M. and Struhl, K. (1993) in Current Protocols in Molecular Biology (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K., eds.), pp. 4.6.1–4.6.13, John Willey and Sons, Inc. New York
28 Nordeen, S. K. (1988) Biotechniques **6**, 454–457
29 Chen, C. and Okayama, H. (1987) Mol. Cell. Biol. **7**, 2745–2752
30 Murguía, J. R., De Vries, L., Gomez-García, L., Schönthal, A. and Perona, R. (1995) J. Cell. Biochem. **57**, 630–640
31 Rackwitz, H. R., Zehetner, G., Frischasuf, A.-M. and Lehrach, H. (1984) Gene **30**, 195–200
32 Shapiro, M. B. and Senepathy, P. (1987) Nucleic Acids Res. **15**, 7155–7174
33 Escalante, R. and Sastre, L. (1994) J. Biol. Chem. **269**, 13005–13012
34 Ovchinnikov, Y. A., Monastyrskaya, G. S., Broude, N. E., Ushkaryov, Y. A., Melkov, A. M., Smirnov, Y. V., Malyshev, I. V., Allikmets, R. L., Kostina, M. B., Dulubova, I. E. et al. (1988) FEBS Lett. **233**, 87–94
35 Kano, I., Nagai, F., Satoh, K., Ushiyama, K., Nakao, T. and Kano, K. (1989) FEBS Lett. **250**, 91–98
36 Shull, M. M., Pugh, D. G. and Lingrel, J. B. (1990) Genomics **6**, 451–460
37 Serrano, R. (1988) Biochim. Biophys. Acta **947**, 1–28
38 Bucher, P. and Trifonov, E. N. (1986) Nucleic Acids Res. **14**, 10009–10026
39 Purnell, B. A., Emanuel, P. A. and Gilmour, D. S. (1994) Genes Dev. **8**, 830–842
40 Ghosh, D. (1990) Nucleic Acids Res. **18**, 1749–1756
41 Fuch, R. (1991) Comput. Appl. Biosci. **7**, 105–106
42 Lenstra, J. A., Vliet, A. V., Arnberg, A. C., Hemert, F. J. V. and Möller, W. (1986) Eur. J. Biochem. **155**, 475–483
43 Ortega, M.-A., Díaz-Guerra, M. and Sastre, L. (1996) J. Mol. Evol. **43**, 224–235
44 Gilbert, W. and Glynias, M. (1993) Gene **135**, 137–144.
45 Mattick, J. S. (1994) Curr. Opin. Genet. Dev. **4**, 823–831
46 Suzuki-Yagawa, Y., Kawakami, K. and Nagano, K. (1992) Mol. Cel. Biol. **12**, 4046–4055
47 Ikeda, K., Nagano, K. and Kawakami, K. (1993) Eur. J. Biochem. **218**, 195–204
48 Pathak, B. G., Neumann, J. C., Croyle, M. L. and Lingrel, J. B. (1994) Nucleic Acids Res. **22**, 4748–4755

Received 22 April 1996/11 September 1996; accepted 18 September 1996