Characterization of the Human Na,K-ATPase α2 Gene and Identification of Intragenic Restriction Fragment Length Polymorphisms*

Marcia M. Shull‡, Diana G. Pugh, and Jerry B. Lingrel§

From the Department of Molecular Genetics, Biochemistry, and Microbiology, University of Cincinnati College of Medicine, Cincinnati, Ohio 45267-0524

We have determined the structure of the gene that encodes the α2 isoform of the human Na,K-ATPase. The gene contains 23 exons and spans approximately 25 kilobases. The amino acid sequence of the human α2 isoform deduced from the genomic sequence exhibits 99% identity to the rat α2 isoform. One of the nine amino acid differences between the human and rat sequences occurs at an amino acid position which is known to be involved in species differences in sensitivity of the α1 isoform to cardiac glycosides. Approximately 1500 base pairs of sequence flanking the 5' end of the α2 gene have been determined. This region contains numerous potential AP-1, AP-2, and NF-1-binding sites, a potential Sp1 recognition site, and several sequences that are similar to the glucocorticoid receptor-binding site. The transcription start site was mapped by primer extension and S1 nuclease protection analyses of RNA from human brain, skeletal muscle, and heart. Multiple transcription initiation sites are clustered between residues −104 to −99 relative to the translation initiation codon. A potential TATA box is located 29 base pairs upstream of the first transcription initiation site. Immediately 5' to the apparent TATA box is a 35-base pair polyurine-polypyrimidine tract containing an imperfect mirror repeat which resembles sequences that form triple-stranded structures. Two intragenic DNA probes which detect restriction fragment length polymorphisms associated with the α2 gene have been identified. These probes will be useful in genetic linkage analyses designed to define the possible role of the Na,K-ATPase in certain hereditary disorders.

The Na,K-transporting ATPase (EC 3.6.1.37), an integral membrane protein present in all animal cells, is responsible for maintaining Na⁺ and K⁺ gradients across the plasma membrane. The enzyme consists of two subunits, a large (Mr ~112,000) catalytic subunit (α) and a smaller (protein component, Mr ~35,000) glycoprotein subunit (β) whose function is unknown. Multiple isoforms of the α subunit (α1, α2, and α3) have been identified using biochemical (Sweadner, 1979; Lytton, 1985a; Hsu and Guidotti, 1989) and cDNA cloning (Shull et al., 1986; Hara et al., 1987) techniques. These isoforms exhibit differences in tissue specificity, developmental patterns of expression, hormonal regulation (reviewed in Lingrel et al., 1989), sensitivity to cardiac glycosides (Sweadner, 1979), and affinity for Na⁺ (Lytton, 1985b). The α1 isoform, although most abundant in kidney, is expressed in a broad range of tissues, while the α2 and α3 isoforms exhibit a more limited tissue distribution. Studies of rat Na,K-ATPase mRNA abundance (Young and Lingrel, 1987; Herrera et al., 1987; Orlowski and Lingrel, 1988; Schneider et al., 1988) indicate that the α2 isoform is found predominantly in brain, skeletal muscle, and heart, increasing in all three tissues during development. The α3 isoform is found primarily in brain and to a lesser extent in skeletal muscle and heart. A developmentally regulated transition in the expression of α2 and α3 mRNAs occurs in heart. The α3 and α1 mRNAs are predominant in fetal and neonatal heart, while α2 and α1 predominate in juvenile and adult heart (Orlowski and Lingrel, 1988; Schneider et al., 1988).

The functioning of the Na,K-ATPase is essential for a variety of cell homeostatic processes including osmotic balance and cell volume regulation, Na⁺-coupled transport of nutrients including glucose and amino acids, and maintenance of the resting membrane potential (reviewed in DeWeer, 1985). In addition, certain specialized functions including electrical excitability of nerve and muscle and fluid movement across kidney and intestinal transport epithelia require Na,K-ATPase activity (DeWeer, 1985). The enzyme may also facilitate fluid absorption from the lung at birth (Bland and Boyd, 1986) and may play a role in determining the ionic composition of the cerebrospinal fluid (Vates et al., 1964) and aqueous humor (Cole, 1961).

The functional diversity of Na,K-ATPase activity is due in part to the existence of multiple α isoforms that are encoded by separate genes. Human genes encoding the α1, α2, and α3 isoforms have been isolated (Shull and Lingrel, 1987; Sverdlov et al., 1987), and one of these, α3, has been partially sequenced (Ovchinnikov et al., 1988). Two additional genomic sequences exhibiting nucleotide and deduced amino acid similarity to the α isoforms have also been identified (Shull and Lingrel, 1987; Sverdlov et al., 1987). The members of this gene family

*This work was supported in part by National Institutes of Health Grant HL28573. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankEMBL Data Bank with accession number(s) J05096.

‡ Supported by a predoctoral fellowship from the Albert J. Ryan Foundation.

§ To whom correspondence should be addressed.

1 The nomenclature is that used by a number of investigators (Felsenfeld and Sweadner, 1986; Hsu and Guidotti, 1989; Ismail-Beigi et al., 1988). α1 or α refers to the predominant kidney isoform with NH₂ terminus Met-Gly-Lys-Gly-Val; α2 or αI, previously referred to as α(+)) (Sweadner, 1973; Lytton, 1985a), has NH₂ terminus Met-Gly-Arg-Gly-Ala; α3, identified initially by cDNA cloning and referred to as αIII (Shull et al., 1986), has NH₂ terminus Met-Gly-Asp-Lys-Lys.
Human Na,K-ATPase α2 Gene Structure and RFLPs

17533

are dispersed in the human genome. The α1 and α2 genes, designated ATP1A1 and ATP1A2, are located on the short and long arms, respectively, of chromosome 1, while the β gene, ATP1B, is located on chromosome 19 (Yang-Feng et al., 1988). One of the α-like genomic sequences, ATP1AL2, is physically linked to the α2 gene (Shull and Lingrel, 1987). A fifth α-like sequence, ATP1AL1L, is located on chromosome 13 (Yang-Feng et al., 1988). Sequences of a human β subunit gene and β pseudogene have also been determined (Lan et al., 1989) and have been localized to human chromosomes 1 and 4, respectively (Yang-Feng et al., 1988).

In order to understand the molecular mechanisms involved in the regulated developmental and tissue-specific expression of the Na,K-ATPase α isoforms and to investigate the role of the enzyme in human disease processes, we have characterized the gene encoding the human Na,K-ATPase α2 isoform and its 5′-flanking sequences. Based on the genomic sequence, we have deduced the amino acid sequence of the human α2 protein. We have also identified intragenic probes which isolate restriction fragment length polymorphisms. These probes will be useful in investigating the genetic linkage of the Na,K-ATPase to hereditary diseases potentially involving the sodium pump.

EXPERIMENTAL PROCEDURES

DNA Sequencing—Restriction fragments from phage clones CL6-2, CL23-1, and CL30-2 were subcloned into either m13mp18 and m13mp19 or pBluescript (International Biotechnologies, Inc.), and a series of nested deletions were prepared using Cyclone or Tornado deletion subcloning kits (International Biotechnologies, Inc.). Single-stranded templates were sequenced by the dideoxy method of Sanger et al., (1977) using [α-35S]dATP and DNA Sequencing kits from Amersham Corp., Pharmacia LKB Biotecnology, Inc., or United States Biochemical Corp. Custom-designed oligonucleotides synthesized on an Applied Biosystems DNA synthesizer model 380A were occasionally used as sequencing primers. Deoxynucleoside triphosphate was frequently substituted for dGTP in order to relieve compressions. Deoxyinosine triphosphate was obtained in both strands. Computer analyses were performed using the program DNAANALYZE, version Z2.1 (Weresk and Thompson, 1989).

Protein Structure Analysis—The predicted secondary structure of the human α2 protein was determined using the methods of Chou and Fasman (1974) and Garnier et al. (1978). Hydropathicity and potential membrane-associated helical regions were predicted using the algorithms of Kyte and Doolittle (1982) and Eisenberg et al. (1984).

RNA Isolation—Total RNA was isolated from adult human brain, skeletal muscle, and heart, and from rat brain and heart using the procedure of Chirgwin et al. (1979). Poly(A)+ RNA from human heart and rat brain and heart was obtained in both strands. Computer analyses were performed using the program DNAANALYZE, version Z2.1 (Weresk and Thompson, 1989).

RNA Isolation—Total RNA was isolated from adult human brain, skeletal muscle, and heart, and from rat brain and heart using the procedure of Chirgwin et al. (1979). Poly(A)+ RNA from human heart and rat brain and heart was obtained by affinity chromatography on oligo(dT)-cellulose.

Primer Extension Analysis—Primer extension analysis was performed according to the method described (Kingsley, 1987). Two synthetic oligonucleotides complementary to the human α2 gene, nucleotides -8 ~ -35 and -36 ~ -66 relative to the transcription initiation site, and two oligonucleotides complementary to the rat α2 cDNA, nucleotides -8 ~ -36 and -37 ~ -66 relative to the transcription initiation site, were end-labeled using γ-32P]-ATP and T4 polynucleotide kinase. The labeled primers were hybridized to RNA by incubating overnight at 30 °C in S1 hybridization solution (80% formamide, 40 mM Pipes, pH 6.4, 400 mM NaCl, and 1 mM EDTA). After precipitation of the annealed primer and template, the hybridized primers were extended by incubating with 40 units of AMV reverse transcriptase in reverse transcriptase buffer (50 mM Tris/

7 The abbreviations used are: Pipes, 1,4-piperazinediethanesulfonic acid; kb, kilobase pair(s); sp, base pair(s); C1RATP, γ-[N-2-chloroethyl-N-methyamino]benzylamide ATP).
transcript was determined by S1 nuclease protection and primer extension analyses. The 3' sequence shown in Fig. 2 includes three potential polyadenylation signals analogous to those used in the rat. Based on homology to the consensus, AATAAA, and to the rat sequence around the polyadenylation signals, it is expected that the second and third sites would be used in the human. However, it is not clear if the first site is used. The human a2 protein coding sequence exhibits 90% nucleotide identity to that of the rat a2 cDNA (Shull et al., 1986). The sequence of the 5'-untranslated and 3'-untranslated regions (excluding the Alu element in the 3'-untranslated region) exhibit 89 and 66% identity, respectively, to the corresponding regions (excluding the AC repeat in the 3'-untranslated region) of the rat a2 cDNA.

The deduced human a2 amino acid sequence exhibits 99% identity to the rat sequence. There are no insertions or deletions in the human sequence relative to that of rat and only nine amino acid differences occur between the two sequences, most of which are conservative changes. These are (human amino acid → rat): Val-64 → Ile, Ile-102 → Leu, Gln-111 → Leu, Val-127 → Ile, Arg-274 → Gln, Met-644 → Val, Lys-675 → Arg, Asn-676 → Asp, and Met-887 → Thr. One of these amino acid differences, Gln-111, occurs at a position which is involved in species differences in sensitivity of the α1 form to cardiac glycosides (Price and Lingel, 1988). The human a2 primary translation product, like that of the rat, is 1020 amino acids in length. The first five amino acids, numbered −5 to −1 in Fig. 2, are identical to those of the rat which apparently are removed by posttranslational processing (Lytton, 1985a; Shull et al., 1986). If this is also true for human α2, then the mature protein would consist of 1015 amino acids and have a M₀ of 111,857.

Location of the Transcription Initiation Site and Nucleotide Sequence of the 5'-Flanking Region—To map the α2 transcription start sites, primer extension and S1 nuclease protection analyses were performed. For primer extension analysis, two primers complementary to the 5'-untranslated sequence, depicted in Fig. 3, were used to analyze human total RNA isolated from adult brain and skeletal muscle and poly(A)+ RNA from heart. As shown in Fig. 4, a cluster of apparent transcription initiation sites located within a single 6-bp region were observed in all three tissues with both primers. Four predominant bands were observed at positions −104, −103, −100, and −99 relative to the start of translation. For comparison, primer extension analysis was also performed using rat brain and heart poly(A)+ RNA. In rat brain and heart samples, a cluster of apparent initiation sites was also observed, with the major bands located at positions −108 and −104 relative to the translation start site. The site at position −104 site is only 6 bp beyond the 5'-terminal end of the rat brain a2 cDNA described previously (Shull et al., 1986).

To confirm the location of the transcription initiation sites indicated by primer extension analyses, S1 nuclease protection studies were performed (Fig. 5). The S1 probe consisted of a 230-nucleotide single-stranded fragment extending from position −36 to −265 relative to the translation start site. Again, a cluster of initiation sites was identified in all tissues, with major sites at −104, −103, −100, and −99. Although in many eukaryotic genes the site of initiation is frequently an adenine (Breathnach and Chambon, 1981), this is not the case for the Na,K-ATPase α2 gene, where the apparent initiation sites occur at T, C, and G residues.

The nucleotide sequence of the 5' end of the Na,K-ATPase α2 gene is shown in Fig. 3. A potential TATA sequence (Breathnach and Chambon, 1981), TATTTAAA, is located 29 bp upstream of the 5'-most transcription initiation site. No obvious CCAAT consensus sequence is observed within 100 bp of the transcription start site. However a sequence,
Nucleotide sequences around the exon/intron boundaries are presented. Exon sequences are in upper case letters and intron sequences in lower case. Amino acids encoded by codons bordering the splice junctions are shown and the number of the amino acid immediately 5' of the splice site is indicated. The NH2-terminal amino acids cleaved to yield the mature protein are assigned negative numbers. Numbers above codons 5' of the splice sites enumerate the nucleotides flanking the splice sites. The first base of the ATG initiation codon is designated +1. Intron sizes (bp) are shown in parentheses.

**TABLE I**

Exon/Intron boundaries in the human Na,K-ATPase α2 subunit gene

| Position | Exon | INTRON | Exon |
|----------|------|--------|------|
| 12       | 1    | GGT    | GCT  |
| 117      | 2    | GAG    | GAC  |
| 177      | 3    | TCC    | GCC  |
| 381      | 4    | GAC    | GGT  |
| 495      | 5    | CCT    | Pro  |
| 748      | 7    | GAA    | CAC  |
| 1017     | 8    | GTG    | Val  |
| 1216     | 9    | TCT    | Ser  |
| 1326     | 10   | TCA    | Ala  |
| 1461     | 11   | TAC    | Tyr  |
| 1561     | 12   | CTG    | Leu  |
| 1887     | 13   | ACT    | Ile  |
| 1954     | 14   | CCC    | Val  |
| 2115     | 15   | AAG    | Arg  |
| 2284     | 16   | GAG    | Gly  |
| 2439     | 17   | GAT    | Thr  |
| 2840     | 18   | AGA    | Met  |
| 2942     | 19   | GAG    | Leu  |
| 3034     | 20   | GAT    | Gly  |

CAACACAC, resembling CCAAT is located 83 bp 5' of the transcription initiation site. A potential Sp1-binding site, GCGCGCGCG (Dyman et al., 1986), is located 119 bp upstream of the apparent cap site. The entire 1.6-kb 5'-flanking region exhibits 57% GC content. The 159-bp region immediately preceding the apparent TATA box and continuous with the strand corresponding to the mRNA exhibits 71% AG content and includes a polypurine-rich region, consisting of 34 purines out of 35 bases, that may have regulatory significance (see “Discussion”).

**Repetitive DNA**—The α2 sequence was examined for the presence of repetitive elements including Alu (Britten et al., 1988; Jurka and Smith, 1988), O (Sun et al., 1984), K (Sun et al., 1984), and L1 (Scott et al., 1987) repeats. Five complete Alu repeats were identified, each exhibiting 72–87% nucleotide identity to published Alu consensus sequences. The Alu elements are present in introns 1, 3, 13, and 22, and in the 3'-untranslated region of exon 23. The Alu repeat in intron 1 is followed by an 11-nucleotide poly(A) tract and is flanked on both sides by a perfect 6-bp direct repeat. Similarly, the Alu element in intron 13 possesses a 15-bp poly(A) tract and is flanked by perfect 9-bp direct repeats. The Alu sequence in intron 3, although not followed by a poly(A) tract, is followed by a 23-nucleotide A-rich region (70% A). However, it is not flanked by short repeats. The Alu element in intron 22 is present in the reverse orientation relative to the gene. It is flanked on both sides by perfect 6-bp inverted repeats, rather than direct repeats, and is followed by a 30-nucleotide poly(A) tract. The Alu element in the 3'-untranslated region is followed by a 14-nucleotide poly(A) tract and flanked by perfect 12-bp direct repeats.

**Comparison of α2 and α3 Gene Structure**—Both the α2 and α3 (Ovchinnikov et al., 1988) genes contain 23 exons and 22 introns. With the exception of introns 1 and 10, the introns in the α2 and α3 genes occur in exactly the same positions. Exon 1 in α2 includes the 5'-untranslated sequence and encodes the first four amino acids that are apparently removed posttranslationally, while exon 1 in α3 contains the 5'-untranslated sequence and encodes the first two amino acids. In the α2 gene, intron 10 occurs between the codons for Lys-437 and Arg-438, while in α3 it interrupts the Arg codon (AG-G). Although the positions of introns in the α2 and α3 genes are basically the same, there appears to be little similarity in terms of intron size or sequence. The size of nine α3 introns has been determined and the sequence of five of these has been reported (Ovchinnikov et al., 1987a). There is no correlation between the sizes of the corresponding α2 and α3 introns. For example, intron 13, which is the second largest (3.9 kb) intron in the α2 gene, is the smallest (70 bp) reported in the α3 intron. There is also no apparent sequence similarity between corresponding α2 and α3 introns. In addition, the occurrence of repetitive elements is not conserved. Whereas α2 has five Alu repeats, one copy each in introns 1, 3, 13, 22, and in the 3'-untranslated region, an analysis of the published α3 sequence reveals four Alu repeats, all present in intron 16.

**Position of Intron/Exon Boundaries Relative to Structural or Functional Domains**—A number of investigators have proposed that exons may represent structural units (Go, 1981), elements of sequential supersecondary structure (Blake, 1978; Lonberg and Gilbert, 1985; Blake, 1985), or structural and functional domains (Gilbert, 1978; Gilbert, 1985; Sakano et al., 1979). In addition, intron/exon splice junctions frequently map to predicted surface segments of regions predicted not to be membrane-embedded (Craig et al., 1982; Argos and Rao, 1985). Because the structural and functional domains of the Na,K-ATPase have not been fully characterized, only limited
FIG. 2. Deduced nucleotide sequence of the human Na,K-ATPase a2 mRNA. The sequence corresponding to the human a2 mRNA was deduced by comparing the a2 gene sequence to the rat a2 cDNA and by locating intron/exon boundaries conforming to splice junction consensus sequences. The deduced amino acid sequence is numbered starting with the NH$_2$ terminus of the mature protein (Lytton, 1985a) and is preceded by five amino acids that occur in the primary translation product. Positions of introns are indicated by an asterisk. The phosphorylation site (Asp-369) is indicated by an asterisk.
predicted membrane-spanning domains that is apparently removed to yield the NH₂ terminus of the plasmic region predicted to be α-helical in structure. Within intron positions generally do not interrupt sequences encoding phobic domains, are also shown. As indicated on the graph, residues which may be involved in functional properties of the enzyme are also indicated. The first intron interrupts the five-amino acid sequence MetGyArgGly.

The predominant feature of exon 4 is that it encodes the first transmembrane domain as well as the first extracellular domain. Located at the borders of this extracellular sequence are two amino acids which are involved in species differences in sensitivity of the α₁ isoform to inhibition by ouabain (Price and Lingrel, 1988). Exon 5 encodes the second transmembrane domain. The cytoplasmic region between the second and third transmembrane domains is encoded by the 3'-portion of exon 5 and by exons 6, 7, and the 5' portion of exon 8. Within this cytoplasmic region are two segments which are conserved among aspartyl-phosphate transport ATPases. Introns 6 and 7 fall within the sequences encoding these conserved regions in both the Na,K-ATPase α₂ gene and the fast-twitch skeletal muscle Ca-ATPase (Korzak et al., 1988). The significance of these conserved domains is unknown. However, within the second conserved region is a glycine residue (Gly-261) that is present in all eukaryotic Na,K-ATPase isoforms and species. The function of this domain is unknown but it has been suggested that it may serve as an ion-selective gate controlling passage of Na⁺ and K⁺ ions to and from binding sites (Shull et al., 1985; Shull et al., 1986).

The graph on the right of the figure refers to nucleotide positions relative to the first transcription start site.
aspartyl-phosphate transport ATPases. Mutation of this residue to aspartate confers vanadate-resistant ATPase activity to the fungal H+-ATPase (Ghislain et al., 1987). The third and fourth hydrophobic membrane-spanning domains are encoded by the 3' portion of exon 8.

The large cytoplasmic portion of the enzyme located between the fourth and fifth predicted transmembrane domains is encoded by exons 9–15 and part of exon 16. Exon 9 contains the highly conserved region surrounding the aspartate residue which is phosphorylated during the catalytic cycle. Three additional regions of similarity among cation transport ATPases are located in this large cytoplasmic region (Green et al., 1988; Korczak et al., 1988; Serrano, 1988; Shull and Greed, 1988). Based on chemical labeling studies (reviewed in Lingrel et al., 1989) and predicted secondary structure analysis (Taylor and Green, 1989), these regions appear to be at or near the nucleotide-binding domain. The first of these homologous regions, located entirely within exon 12 in the α2 gene (amino acids 497–519), contains the conserved lysine (Lys-500 in α2) which is labeled by fluorescein 5'-isothiocyanate (Farley et al., 1984; Kirley et al., 1984). The labeled Lys and the Gly-Ala following it are conserved in all eukaryotic aspartyl-phosphate transport ATPases and may be analogous to the NH2-terminal ATP-binding loop of adenylate kinase (Taylor and Green, 1989; Walker et al., 1982). Interestingly, although the entire region of homology around the fluorescein 5'-isothiocyanate site (amino acids 497–519) is within a single exon in α2, in the sarcoplasmic reticulum fast-twitch Ca2+-ATPase gene (Korczak et al., 1988), intron 13 occurs between the highly conserved Lys and the Gly following it, interrupting the predicted ATP-binding loop.

A second region of similarity in cation-transporting ATPases occurs between residues 574 and 625 in α2. This segment contains a 19-amino acid sequence (amino acids 594–
precedes the Thr of the Thr-Gly-Asp sequence (Korczak et al., 1987b). This region may form part of the nucleotide-binding domain (Ohta et al., 1986) or may form a hinge region contacting the phosphorylation and nucleotide-binding domains (Taylor and Green, 1989). In α2, this region is located almost entirely within exon 16. In the Ca-ATPase, intron 15 immediately precedes the Thr of the Thr-Gly-Asp sequence (Korczak et al., 1988).

Exon 17 encodes a large hydrophobic region which may represent one or two transmembrane passes. Exons 18 encodes a hydrophilic region which is conserved among mammalian aspartyl-phosphate transport ATPases. The significance of this conservation is not clear. Exons 19–22 encode hydrophobic domains which may represent transmembrane domains. However, the number of transmembrane passes in the COOH-terminal half of the protein is uncertain. The 3′ portion of exon 20 contains a potential cAMP-dependent phosphorylation site (Shull et al., 1986). Exon 23 encodes the hydrophilic COOH-terminal region which is probably located in the cytoplasm.

Detection of Restriction Fragment Length Polymorphisms with Na,K-ATPase α2 Probes—Several restriction fragments from within the α2 gene were identified which lacked repetitive elements and were gene specific. These probes were tested for their ability to detect restriction fragment length polymorphisms in human genomic DNA that had been digested with a panel of restriction endonucleases (BamHI, BglII, EcoRI, HindIII, KpnI, MspI, PstI, PvuII, SacI, and TaqI). Two probes which detected polymorphisms are shown in Fig. 7. Probe ATP1A2 (6–2–3) is a 2.5-kb EcoRI fragment representing the 5′ end of the α2 gene. Probe ATP1A2 (30–2–6) is a 0.6-kb fragment from the 3′ portion of the α2 gene. Allele sizes are indicated.
human genomic DNA. The polymorphic restriction site generates two fragments of 8.0 and 3.3 kb in heterozygous individuals. Allele frequencies are 0.85 for the 8.0-kb allele and 0.15 for the 3.3-kb allele. Probe ATP1A2 30–2–6 is a 1.0-kb EcoRI fragment from the 3′ portion of the α2 gene and includes exons 21 and 22. It exhibits a polymorphism only with BglII, yielding bands of 10.5 and 4.3 kb in heterozygotes. The 4.3-kb allele exhibits a frequency of 0.8, while the 10.5-kb allele exhibits a frequency of 0.2. Codominant Mendelian inheritance of these alleles has been demonstrated in several families, two of which are shown in Fig. 8. The genotype distribution for each probe does not deviate significantly from that expected for Hardy-Weinberg equilibrium.

### DISCUSSION

The objective of this study was to analyze the structure of the Na,K-ATPase α2 gene, to determine the 5′-flanking sequence, and to identify intragenic probes which can be used in genetic linkage studies. The human Na,K-ATPase α2 gene spans approximately 25 kb and, like the human Na,K-ATPase α3 gene (Ovchinnikov et al., 1988) and the rabbit fast-twitch skeletal muscle sarcoplasmic reticulum Ca-ATPase gene (Korczał et al., 1988), contains 23 exons. The gene structure appears typical for a eukaryotic protein-coding gene. The median exon size, 135 bp, falls within the most abundant exon size class observed in higher eukaryotic protein-coding genes (Naora and Deacon, 1982). Similarly, the median intron size, 253 bp, is similar to the size of the most abundant eukaryotic intron class (Naora and Deacon, 1982). Although the position of introns in the α2 and α3 genes are basically the same, the intron size and sequence and the position of Alu repeats are not conserved.

As the α2 isoform is one of the Na,K-ATPase isoforms expressed in adult heart, it is a receptor for cardiac glycosides used in the treatment of congestive heart failure. Therefore, the sequence of human α2 in the regions involved in cardiac glycoside binding and sensitivity is of considerable interest. Several sites that may be involved in the interaction of cardiac glycosides with the α subunit have been identified, including the extracellular junctions between the first and second transmembrane domains (H1 and H2) (Shull et al., 1986; Price and Lingrel, 1988). Two residues that account for species differences in ouabain sensitivity of the α1 isoform, located at the boundaries of the H1–H2 junction have been determined by site-directed mutagenesis. Conversion of Gln-111 and Asn-122 of the sensitive sheep enzyme to Arg and Asp, respectively, which occur in the insensitive rat enzyme, results in a fully resistant enzyme (Price and Lingrel, 1988). One of the amino acid differences between the human and rat α2 sequences occurs at residue 111 (human, Gln; rat, Leu). Thus, there may be differences in ouabain binding between the human and rat α2 isoforms. The Gin-111 and Asn-122 residues are also found in human (Kawakami et al., 1986) and pig (Ovchinnikov et al., 1988) α1 and in human α3 (Ovchinnikov et al., 1988). The codon for Asn-122 (AAT) in the α2 gene borders intron 4. The last nucleotide of the exon 4 bordering the splice donor site is usually a G. However, in this case the nucleotide is a T. In the α3 gene, the corresponding Asn codon is AAC (Ovchinnikov et al., 1988). If a G were present in this position in either gene, the codon would represent Lys, a charged amino acid, and the resulting enzyme might be expected to be more resistant to ouabain. This suggests the possible existence of evolutionary pressure for maintaining sensitivity to ouabain-like compounds including putative endogenous inhibitors of the Na,K-ATPase.

In order to begin addressing questions concerning the regulation of Na,K-ATPase expression at the transcriptional level, we determined the transcription initiation sites and examined the 5′-flanking sequence of the α2 gene for potential promoter elements, transcription factor-binding sites (Sp1, AP-1, AP-2, CP1/CP2, NF-1, and CACCC factor-binding sites), hormone response elements (thyroid hormone receptor and glucocorticoid receptor-binding sites), and for regions of unusual DNA structure. Primer extension and S1 nuclease protection analyses of RNA from human brain, skeletal muscle, and heart, the three major tissues in which the α2 isoform is expressed, demonstrate that there are four transcription initiation sites clustered between −194 and −99 relative to the translation start site. All four sites appear to be used in each tissue. However, in skeletal muscle, the −100 and −99 sites appear to be used less frequently than in brain or heart.

The α2 5′-flanking region contains two sequences exhibiting similarity to the CCAAT element consensus sequence or its reverse complement (Chodosh et al., 1988) (Fig. 3 and Table II). This pentanucleotide is the core-binding site for the transcription factors CP1 and CP2 and is usually located within 100 bp of the transcription start site. The sequences in α2 which exhibit similarity to the consensus are located much further from the transcription start site, at −430 and −1031. Another CCAAT element, with the consensus TTGGCTNNNAGCCAA, is recognized by nuclear factor I (NF-I/CTF) (Jones et al., 1987). There are five potential NF-1-binding sites in the α2 5′-flanking region which maintain at least 4 of the 6 residues in the NF-1-binding site that, based on methylation interference analysis, seem to contact the NF-1 protein (Chodosh et al., 1988). One of these, at position −1457, has all six essential contact points, resulting in a good match to both half-sites of the palindromic CTF/NF-1-binding site. A site at −183 exhibits 100% identity to the core AGCCCA hexanucleotide, the minimal recognition site for CTF/NF-1 (Jones et al., 1987).

The CACCC element, GCCCACCC (Dierks et al., 1983), is an upstream promoter element that seems to act synergistically with glucocorticoid/progesterone receptor-binding sites in mediating hormone induction probably via interactions of the hormone receptor and CACCC-binding factor (Schüle et al., 1988). There are three sequences in the α2 5′-flanking region that exhibit a five out of five match to the core pentanucleotide CACCC sequence or its reverse complement. One of these (at −1190) overlaps a potential glucocorticoid response element.
**Table II**

Consensus binding sites for the glucocorticoid receptor (GRE) and for transcription factors AP-1, AP-2, CP1/CP2, NF-1, and the CACCC box factor are indicated. Sequences within the 5′-flanking region of the human \( \alpha_2 \) gene which exhibit similarity to these consensus sequences are shown. For the glucocorticoid response element and CP1/CP2 elements, matches to the reverse complements were observed and are also shown. Numbers indicate the distance of the element from the first transcription start site.

| AP-1 | AP-2 | CACCC FACTOR | GRE |
|------|------|--------------|-----|
| TAGCTCA -1211 | CCCCCAGGC | aGACACC -1190 | GGGgacAtcTGTcCT -1321 |
| TGCCTCA -1560 | aCCCAGGC -241 | GtcCcACCC -723 | tcctCtcgcTGTcCC -1310 |
| TGCCTT -1064 | CccCaCAGGC -199 | aGcAgacctGTCCT -1276 |
| TGCCTCC -573 | | aGctgctgcTGTcCT -1174 |
| TGCTCTA -546 | | aGagaccgcTGTTC -1007 |
| TGAGCTCA -297 | | cAtctctagpGTCCT -874 |

| CP1/CP2 | NF-1 | CACCC FACTOR |
|---------|------|--------------|
| GGCCTTCGAGGTGACGTGA -15 bp upstream of the transcription initiation site. All matches of at least five out of six to the core glucocorticoid receptor-binding site are shown. | TTGCTNNNAGCCAA |

| GGCCTTCGAGGTGACGTGA -15 bp upstream of the transcription initiation site. All matches of at least five out of six to the core glucocorticoid receptor-binding site are shown. | TTGCTNNNAGCCAA |

The octamer motif, ATGCAAT, is found within SV40 and immunoglobulin enhancers and in the upstream promoter regions of a variety of other genes (see Fletcher et al., 1987). A sequence (ATTCACAAT) that exhibits a seven out of eight nucleotide match to this motif is located in the \( \alpha_2 \) gene at position -384 relative to the transcription start site.

AP-1 is a mammalian transcription factor that influences basal transcription levels and is also required for induction of transcription by phorbol ester tumor promoters (reviewed in Curran and Franzia, 1988). Recently, protein products of the jun and fos proto-oncogenes have been shown to contribute to the AP-1 protein complex and to bind to the AP-1 consensus recognition sequence, TGA(G,C)TCA (see Curran and Franzia, 1988). The first 5′ end of the \( \alpha_2 \) gene contains six sequences which exhibit at least a six out of seven match to the AP-1 consensus sequence. One of these, located at -297, exhibits 100% identity to the AP-1 consensus. Since the fos and jun proto-oncogene products have been associated with cellular processes involved in development, differentiation, and neuronal function (Curran and Franzia, 1988), the possibility that the \( \alpha_2 \)-ATPase may be regulated by fos and jun products is intriguing. In addition, the potential regulation of Na,K-ATPase expression by phorbol esters deserves investigation.

AP-2 is a stimulatory transcription factor that binds to control regions of a number of eukaryotic genes (Mitchell et al., 1987). Within the 5′-flanking region of \( \alpha_2 \), there are three sequences which exhibit at least a seven out of eight identity to the AP-2 consensus binding site, CCCCCAGGC, or its reverse complement. One of these, at -1101, exhibits 100% identity to this recognition sequence.

Since Na,K-ATPase activity is modulated by hormones including glucocorticoids, mineralocorticoids, and triiodothyronine (reviewed in Lingrel et al., 1989), we examined the 5′-flanking region of the \( \alpha_2 \) gene for potential hormone response elements. A consensus recognition sequence for the mineralocorticoid receptor has not been reported. However, the mineralocorticoid receptor and glucocorticoid receptor may recognize similar regulatory elements (Arriza et al., 1987; Cato and Weinman, 1988). The glucocorticoid response element consists of a hexanucleotide core, TGTTCCT, which forms part of a 15-bp imperfect dyad symmetry element, GGTCANNNNTTCT (Karim et al., 1984; Jantzen et al., 1987). There are 14 sequences in the 5′-flanking region of \( \alpha_2 \) that exhibit at least five out of six match to the core glucocorticoid response element or its reverse complement. One of these at -1007 exhibits 100% identity to the hexanucleotide core. In most cases, homology to the entire 15-bp palindromic sequence is lost. However, a region at -468 exhibits a 10 out of 15 match to the complete sequence, having a five out of six match to the first half of the dyad and a five out of six match to the second half. The first intron also contains numerous matches of at least five out of six to the core element.

A consensus thyroid hormone receptor-binding site has not been reported for human genes. However, for the rat growth hormone gene, the minimal sequence requirement for thyroid hormone induction of the rat growth hormone promoter (reviewed by Samuels et al., 1988) or heterologous thymidine kinase promoter (Glass et al., 1987; Brent et al., 1989) is a 23-bp region, AGGTAAGATCGGGAGCTGCC, located ~160 bp upstream of the transcription initiation site. Although there are several sequences within the 5′ end of the \( \alpha_2 \) gene which exhibit partial identity to this consensus, none of these appears to be in particularly good agreement.

Because non-B DNA structures have been suggested to be involved in gene regulation, we examined the 5′ end of the \( \alpha_2 \) gene for sequences potentially capable of adopting unusual conformations. A striking sequence, located immediately 5′
of the TATA box, consists of a homopurine-homopyrimidine tract containing an imperfect mirror repeat sequence (see Fig. 3). Homopurine-homopyrimidine sequences are frequently found in promoter regions of eukaryotic genes and often exhibit S1 nuclease hypersensitivity (reviewed in Wells et al., 1988). Chemical modification and single-strand-specific nuclease sensitivity studies indicate that such sequences, particularly those possessing mirror repeat symmetry, may form unusual DNA structures, called H-DNA, which contain both single-stranded and triple-stranded regions (see Wells et al., 1988). These structures consist of a core triple helix and a fourth strand, which although closely associated with the triplex, is exposed. It has been suggested that such structures may provide access to transcription factors, facilitate protein-protein interactions, or inhibit stable chromatin assembly, thus functioning in the regulation of gene expression (Christophe et al., 1985; Htun and Dahlberg, 1988; Wells et al., 1988).

The Na,K-ATPase, because of its pivotal role in maintaining Na⁺ and K⁺ balance, has been implicated in a number of disease processes including essential hypertension, familial obesity, various kidney transport defects such as Liddle's syndrome and pseudohypoaldosteronism, and neuromuscular disorders such as periodic paralysis (Hilton, 1986; Layzer, 1982; Schwartz and Spitzer, 1979). One method for investigating the etiology of hereditary diseases is to examine the linkage of candidate genes to the disease. In this approach, polymorphic markers for a gene which is under consideration as the possible disease locus are used in genetic linkage studies (Gusella, 1986). We have identified two DNA probes from within the α2 gene which reveal restriction fragment length polymorphisms and thus should be useful in testing the potential role of a defective Na,K-ATPase α2 gene in the etiology of certain hereditary diseases.

Acknowledgments—We thank Drs. Jay Degen, Sandra Degen, Don Luse, and Richard Sinden for useful discussions, Drs. Lois Lane, John Orlowski, and Gary Shull for reading of the manuscript, Mike Hughes for synthesizing the oligonucleotides, Jennifer Schroeder for typing the manuscript, and Greg Wernke for help with computer analyses. We also thank Dr. Don Unverferth, Ohio State University, for the human heart tissue and Dr. Frank Zemlan for human brain tissue.

Note Added in Proof—The sequence of the 5' end of the human Na,K-ATPase α2 gene has also recently been determined by Sverdlov et al. (1989).

REFERENCES

Al-Shawi, M. K., Parsonage, D., and Senior, A. E. (1988) J. Biol. Chem. 263, 19633-19639
Argos, P., and Rao, J. K. M. (1985) Biochim. Biophys. Acta 827, 283-287
Arriza, J. L., Weinberger, C., Cerelli, G., Glaser, T. M., Handelin, B., Luse, and Richard Sinden for useful discussions, Drs. Lois Lane, John Orlowski, and Gary Shull for reading of the manuscript, Mike Hughes for synthesizing the oligonucleotides, Jennifer Schroeder for typing the manuscript, and Greg Wernke for help with computer analyses. We also thank Dr. Don Unverferth, Ohio State University, for the human heart tissue and Dr. Frank Zemlan for human brain tissue.

Note Added in Proof—The sequence of the 5' end of the human Na,K-ATPase α2 gene has also recently been determined by Sverdlov et al. (1989).

REFERENCES

Al-Shawi, M. K., Parsonage, D., and Senior, A. E. (1988) J. Biol. Chem. 263, 19633-19639
Argos, P., and Rao, J. K. M. (1985) Biochim. Biophys. Acta 827, 283-287
Arriza, J. L., Weinberger, C., Cerelli, G., Glaser, T. M., Handelin, B., Luse, and Richard Sinden for useful discussions, Drs. Lois Lane, John Orlowski, and Gary Shull for reading of the manuscript, Mike Hughes for synthesizing the oligonucleotides, Jennifer Schroeder for typing the manuscript, and Greg Wernke for help with computer analyses. We also thank Dr. Don Unverferth, Ohio State University, for the human heart tissue and Dr. Frank Zemlan for human brain tissue.

Note Added in Proof—The sequence of the 5' end of the human Na,K-ATPase α2 gene has also recently been determined by Sverdlov et al. (1989).
Mount, S. M. (1982) Nucleic Acids Res. 10, 847–861
Mitchell, P. J., Wang, C., and Tjian, R. (1987) Cell 50, 847–861
Naora, H., and Deacon, N. J. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 6196–6200
Ohta, T., Nagano, K., and Yoshida, M. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 2071–2075
Orlowski, J., and Lingrel, J. B. (1988) J. Biol. Chem. 263, 10436–10442
Ovchinnikov, Yu. A., Modyanov, N. N., Broude, N. E., Petrukhin, K. E., Grishin, A. V., Arzamazova, N. M., Aldanova, N. A., Monastyrskaya, G. S., and Sverdlov, E. D. (1986) FEBS Let. 201, 237–245
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., Gryshin, A. V., Sverdlov, V. E., Kiyatkin, N. I., Kostina, M. B., Modyanov, N. N., and Sverdlov, E. D. (1987a) FEBS Lett. 213, 73–80
Ovchinnikov, Yu. A., Dzhandzugaizyan, K. N., Lutsenko, S. V., Mustayev, A. A., and Modyanov, N. N. (1987b) FEBS Lett. 217, 111–116
Ovchinnikov, Yu. A., 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 Sverdlov, E. D. (1988) FEBS Lett. 233, 87–94
Price, E. M., and Lingrel, J. B. (1988) Biochemistry 27, 8400–8408
Sakano, H., Rogers, J. H., Maki, R., Wall, R., and Tonegawa, S. (1979) Nature 277, 627–633
Samuels, H. H., Forman, B. M., Horowitz, Z. D., and Ye, Z.-S. (1988) J. Clin. Invest. 81, 957–967
Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
Schneider, J. W., Mercer, R. W., Gilmore-Hebert, M., Utset, M. F., Lai, C., Greene, A., and Benz, E. J., Jr. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 284–288
Schüle, R., Müller, M., Otsuka-Murakami, H., and Renkawitz, R. (1988) Nature 332, 87–90
Schwartz, G. J., and Spitzer, A. (1978) in Pediatric Kidney Disease (Edelman, C. M., Jr., ed) pp. 1079–1093, Little, Brown and Co., Boston
Scott, A. F., Schmeckpeper, B. J., Abdelrazik, M., Comey, C. T., O’Hara, B., Rossiter, J. P., Cooley, T., Heath, P., Smith, K. D., and Margolet, L. (1987) Genomics 1, 113–125
Serrano, R. (1988) Biochim. Biophys. Acta 947, 1–28
Shull, G. E., Schwartz, A., and Lingrel, J. B. (1985) Nature 316, 691–695
Shull, G. E., Greeb, J., and Lingrel, J. B. (1986) Biochemistry 25, 8125–8132
Shull, G. E., and Greeb, J. (1988) J. Biol. Chem. 263, 8646–8657
Shull, M. E., and Lingrel, J. B. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 4039–4043
Sun, L., Paubson, K. E., Schmid, C. W., Kadyk, L., and Leinwand, L. (1984) Nucleic Acids Res. 12, 2669–2690
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., Gryshin, A. V., Sverdlov, V. E., Kiyatkin, N. I., Kostina, M. B., Modyanov, N. N., and Sverdlov, E. D. (1987) FEBS Lett. 213, 73–80
Sverdlov, E. D., Bessarab, D. A., Malyshev, I. V., Petrukhin, K. E., Smirnov, Yu. V., Ushkaryov, Yu. A., Monastyrskaya, G. S., Broude, N. E., and Modyanov, N. N. (1989) FEBS Lett. 244, 481–483
Sweedner, K. J. (1979) J. Biol. Chem. 254, 6060–6067
Taylor, W. R., and Green, N. M. (1989) Eur. J. Biochem. 179, 241–248
Vates, T. S., Jr., Bonting, S. L., and Oppelt, W. W. (1964) Am. J. Physiol. 206, 1165–1172
Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982) EMBO J. 1, 945–951
Wells, R. A., Collier, D. A., Harvey, J. C., Shimizu, M., and Wohlrab, F. (1988) FASEB J. 2, 2939–2949
Wernke, G. R., and Thompson, R. L. (1989) Biophys. J. 55, 390a
Yang-Feng, T. L., Schwartz, J. W., Lindgren, V., Shull, M. M., Benz, E. J., Jr., Lingrel, J. B., and Francke, U. (1988) Genomics 2, 128–138
Young, R. M., and Lingrel, J. B. (1977) Biochem. Biophys. Res. Commun. 145, 52–58