

## Characterization of the Human Na,K-ATPase $\alpha 2$ Gene and Identification of Intragenic Restriction Fragment Length Polymorphisms\*

(Received for publication, April 18, 1989)

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  $\alpha 2$  isoform of the human Na,K-ATPase. The gene contains 23 exons and spans approximately 25 kilobases. The amino acid sequence of the human  $\alpha 2$  isoform deduced from the genomic sequence exhibits 99% identity to the rat  $\alpha 2$  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  $\alpha 1$  isoform to cardiac glycosides. Approximately 1500 base pairs of sequence flanking the 5' end of the  $\alpha 2$  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 polypurine-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  $\alpha 2$  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  $\text{Na}^+$  and  $\text{K}^+$  gradients across the plasma membrane. The enzyme consists of two subunits, a large ( $M_r \sim 112,000$ ) catalytic subunit ( $\alpha$ ) and a smaller (protein component,  $M_r \sim 35,000$ ) glycoprotein subunit ( $\beta$ ) whose function is unknown. Multiple isoforms of the  $\alpha$  subunit ( $\alpha 1$ ,  $\alpha 2$ , and

$\alpha 3$ )<sup>1</sup> 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  $\text{Na}^+$  (Lytton, 1985b). The  $\alpha 1$  isoform, although most abundant in kidney, is expressed in a broad range of tissues, while the  $\alpha 2$  and  $\alpha 3$  isoforms exhibit a more limited tissue distribution. Studies of rat Na,K-ATPase mRNA abundance (Young and Lingrel, 1987; Herrera *et al.*, 1987; Orłowski and Lingrel, 1988; Schneider *et al.*, 1988) indicate that the  $\alpha 2$  isoform is found predominantly in brain, skeletal muscle, and heart, increasing in all three tissues during development. The  $\alpha 3$  isoform is found primarily in brain and to a lesser extent in skeletal muscle and heart. A developmentally regulated transition in the expression of  $\alpha 2$  and  $\alpha 3$  mRNAs occurs in heart. The  $\alpha 3$  and  $\alpha 1$  mRNAs are predominant in fetal and neonatal heart, while  $\alpha 2$  and  $\alpha 1$  predominate in juvenile and adult heart (Orłowski 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,  $\text{Na}^+$ -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  $\alpha$  isoforms that are encoded by separate genes. Human genes encoding the  $\alpha 1$ ,  $\alpha 2$ , and  $\alpha 3$  isoforms have been isolated (Shull and Lingrel, 1987; Sverdlov *et al.*, 1987), and one of these,  $\alpha 3$ , has been partially sequenced (Ovchinnikov *et al.*, 1988). Two additional genomic sequences exhibiting nucleotide and deduced amino acid similarity to the  $\alpha$  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 GenBank™/EMBL Data Bank with accession number(s) J05096.

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

§ To whom correspondence should be addressed.

<sup>1</sup> The nomenclature is that used by a number of investigators (Felsenfeld and Sweadner, 1988; Hsu and Guidotti, 1989; Ismail-Beigi *et al.*, 1988).  $\alpha 1$  or  $\alpha I$  refers to the predominant kidney isoform with  $\text{NH}_2$  terminus Met-Gly-Lys-Gly-Val;  $\alpha 2$  or  $\alpha II$ , previously referred to as  $\alpha (+)$  (Sweadner, 1979; Lytton, 1985a), has  $\text{NH}_2$  terminus Met-Gly-Arg-Gly-Ala;  $\alpha 3$ , identified initially by cDNA cloning and referred to as  $\alpha III$  (Shull *et al.*, 1986), has  $\text{NH}_2$  terminus Met-Gly-Asp-Lys-Lys.

are dispersed in the human genome. The  $\alpha 1$  and  $\alpha 2$  genes, designated ATP1A1 and ATP1A2, are located on the short and long arms, respectively, of chromosome 1, while the  $\alpha 3$  gene, ATP1A3, is located on chromosome 19 (Yang-Feng *et al.*, 1988). One of the  $\alpha$ -like genomic sequences, ATP1AL2, is physically linked to the  $\alpha 2$  gene (Shull and Lingrel, 1987). A fifth  $\alpha$ -like sequence, ATP1AL1, is located on chromosome 13 (Yang-Feng *et al.*, 1988). Sequences of a human  $\beta$  subunit gene and  $\beta$  pseudogene have also been determined (Lane *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  $\alpha$  isoforms and to investigate the role of the enzyme in human disease processes, we have characterized the gene encoding the human Na,K-ATPase  $\alpha 2$  isoform and its 5'-flanking sequences. Based on the genomic sequence, we have deduced the amino acid sequence of the human  $\alpha 2$  protein. We have also identified intragenic probes which detect 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 pIB131 (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 [ $\alpha$ - $^{32}$ S]dATP and DNA Sequencing kits from Amersham Corp., Pharmacia LKB Biotechnology, Inc., or United States Biochemical Corp. Custom-designed oligonucleotides synthesized on an Applied Biosystems DNA synthesizer model 380A were occasionally used as sequencing primers. Deoxyinosine triphosphate was frequently substituted for dGTP in order to relieve compressions. All restriction sites were sequenced across except the *Eco*RI sites in introns 1, 6, 7, 10, and 20 and a *Hind*III site in intron 13. With the exception of small *Eco*RI and *Hind*III fragments which may have been present at these restriction sites, the gene was sequenced in its entirety. Seventy-five % of the sequence indicated in Fig. 1 was obtained in both strands. Computer analyses were performed using the program DNANALYZE, version Z2.1 (Wernke and Thompson, 1989).

**Protein Structure Analysis**—The predicted secondary structure of the human  $\alpha 2$  protein was determined using the methods of Chou and Fasman (1974) and Garnier *et al.* (1978). Hydrophobicity 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)<sup>+</sup> 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 (Kingston, 1987). Two synthetic oligonucleotides complementary to the human  $\alpha 2$  gene, nucleotides -8 to -35 and -36 to -65 relative to the translation initiation site, and two oligonucleotides complementary to the rat  $\alpha 2$  cDNA, nucleotides -8 to -36 and -37 to -66 relative to the translation initiation site, were end-labeled using [ $\gamma$ - $^{32}$ P]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/

HCl, pH 8.0, 5 mM MgCl<sub>2</sub>, 5 mM dithiothreitol, 50 mM KCl) containing 0.55 mM of each deoxynucleotide triphosphate. The samples were precipitated by addition of ethanol, denatured, and analyzed on an 8% polyacrylamide sequencing gel.

**S1 Nuclease Mapping**—S1 nuclease protection of the 5' end of the  $\alpha 2$  transcript was performed as described (Greene, 1987). Briefly, a  $^{32}$ P 5'-end-labeled synthetic oligonucleotide complementary to nucleotide -8 to -35 was annealed to a single-stranded M13mp18 template containing a 2.5-kb *Eco*RI insert covering the putative transcription initiation site. The primer was extended using the large fragment of DNA polymerase I and the resulting double-stranded product cleaved with *Sph*I. After denaturing alkaline agarose gel electrophoresis, a radiolabeled single-stranded fragment, 230 nucleotides in length, was eluted from the gel.  $5 \times 10^4$  cpm of probe and 50  $\mu$ g of total RNA or 15  $\mu$ g of poly(A)<sup>+</sup> RNA were incubated in 40  $\mu$ l of S1 hybridization solution at 65 °C for 10 min, then at 30 °C for 18 h. The samples were digested with 50, 100, or 300 units of S1 nuclease by incubating for 1 h at 30 °C in 300  $\mu$ l of S1 buffer (280 mM NaCl, 50 mM sodium acetate, pH 4.5, and 4.5 mM ZnSO<sub>4</sub>) containing 6  $\mu$ g of denatured salmon sperm DNA and S1 nuclease. The protected fragments were precipitated by addition of ethanol, denatured, and analyzed on an 8% polyacrylamide sequencing gel.

**Preparation of Repeat-free, Gene-specific Probes**—DNA was isolated from recombinant phage clones containing human Na,K-ATPase  $\alpha 2$  sequences, digested with restriction endonucleases, separated by agarose gel electrophoresis, and transferred to nylon membranes. Total human genomic DNA, labeled with  $^{32}$ P by nick translation, was hybridized to the Southern blots using conditions described previously (Shull and Lingrel, 1987). Fragments which appeared to lack repetitive elements, based on the absence of an autoradiographic signal, were isolated by preparative gel electrophoresis and subcloned into the plasmid vector pIB131. The insert-containing plasmids were labeled with  $^{32}$ P by nick translation and used as probes on Southern blots of restriction enzyme-digested DNAs from phage clones containing inserts representing each of the  $\alpha$  genes ATP1A1, ATP1A2, ATP1AL1, and ATP1AL2 in order to confirm gene specificity. The same probes were also used on restriction enzyme-digested human genomic DNA to confirm gene specificity and absence of repetitive elements.

#### RESULTS

**Structure of the Human Na,K-ATPase  $\alpha 2$  Gene**—The nucleotide sequence of the human Na,K-ATPase  $\alpha 2$  gene was determined by sequencing genomic DNA inserts from phage clones CL6-2, CL23-1, and CL30-2 which were reported previously (Shull and Lingrel, 1987). Seventy-five % of the sequence, including the entire 5'-flanking sequence, the 5'-untranslated region, and all coding exons, was determined in both strands. The gene contains 23 exons and spans approximately 25 kb of genomic DNA. As illustrated in Fig. 1, the exons are clustered in three regions. The first exon, containing 5'-untranslated sequences and nucleotides encoding the first four amino acids of the primary translation product, is separated from the remaining exons by a 5.0-kb intron. Exons 2 through 13 occupy the central portion of the gene, while exons 14 through 23 are clustered at the 3' end. The protein coding exons range in size from 60 to 269 bp, while introns vary in length from 129 bp to 5.0 kb. As shown in Table I, the intron/exon boundary sequences conform to published consensus sequences (Breathnach and Chambon, 1981; Mount, 1982). Each splice donor site begins with GT, while each splice acceptor site ends with an AG and is preceded by a polypyrimidine tract. The locations of introns within the protein coding sequence are illustrated in Fig. 2.

**Deduced Sequence of the Human Na,K-ATPase  $\alpha 2$  mRNA and Protein**—A composite sequence representing the human  $\alpha 2$  mRNA is shown in Fig. 2. Portions of the genomic sequence corresponding to coding regions could be unambiguously determined because of the high similarity between the human and rat  $\alpha 2$  amino acid sequences. Also, sequences at predicted intron-exon boundary positions agree with published consensus sequences (Table I). As described below, the 5' end of the

<sup>2</sup> The abbreviations used are: PIPES, 1,4-piperazinediethanesulfonic acid; kb, kilobase pair(s); bp, base pair(s); C18ATP,  $\gamma$ -[4-N-(2-chloroethyl-N-methylamino)]benzylamide ATP].

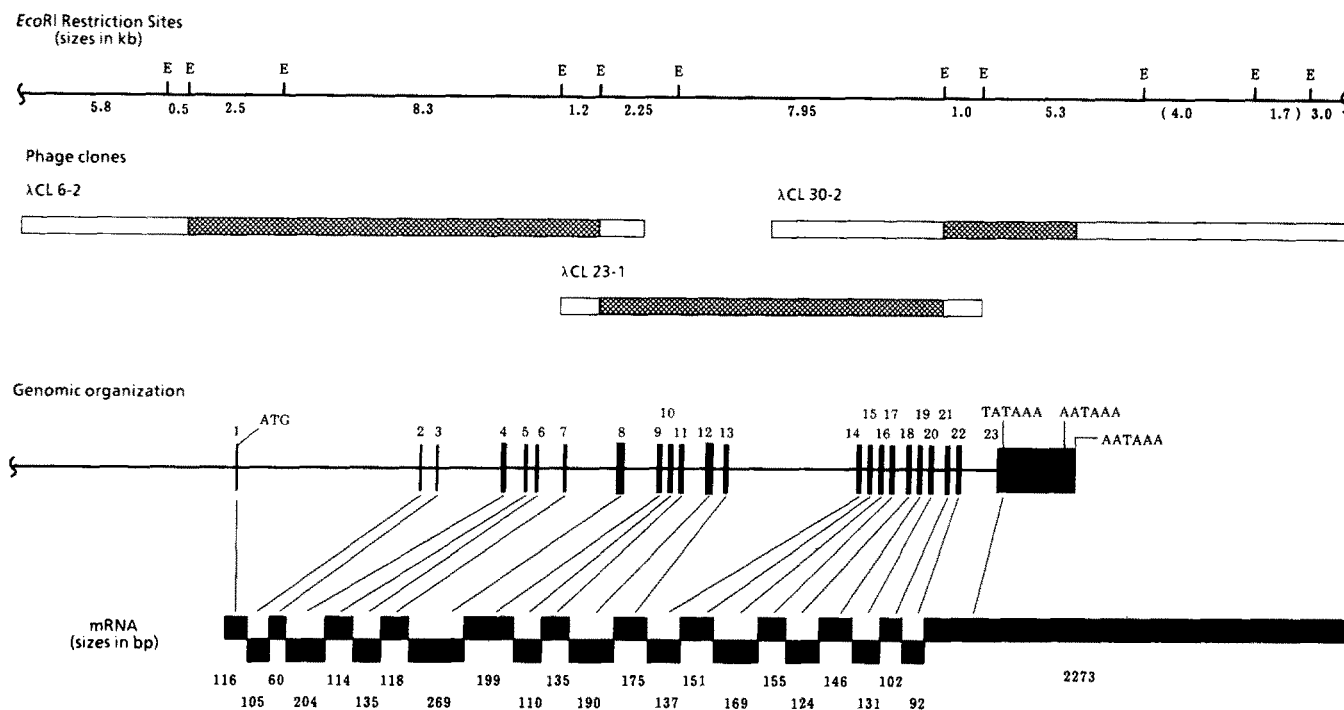


FIG. 1. Organization of the human Na,K-ATPase  $\alpha 2$  gene. Line 1, EcoRI restriction map of the  $\alpha 2$  genomic locus. Lines 2 and 3, overlapping genomic inserts from recombinant phage clones are shown relative to the restriction map. Regions that were sequenced are indicated by hatched bars. Line 4, locations of  $\alpha 2$  exons, depicted by solid bars, are shown relative to the restriction map. The start of translation, ATG, and the three potential polyadenylation signals are marked. Line 5, sizes of exons (bp) are indicated.

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  $\alpha 2$  protein coding sequence exhibits 90% nucleotide identity to that of the rat  $\alpha 2$  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  $\alpha 2$  cDNA.

The deduced human  $\alpha 2$  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  $\rightarrow$  rat): Val-64  $\rightarrow$  Ile, Ile-102  $\rightarrow$  Leu, Gln-111  $\rightarrow$  Leu, Val-127  $\rightarrow$  Ile, Arg-274  $\rightarrow$  Gln, Met-644  $\rightarrow$  Val, Lys-675  $\rightarrow$  Arg, Asn-676  $\rightarrow$  Asp, and Met-887  $\rightarrow$  Thr. One of these amino acid differences, Gln-111, occurs at a position which is involved in species differences in sensitivity of the  $\alpha 1$  isoform to cardiac glycosides (Price and Lingrel, 1988). The human  $\alpha 2$  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  $\alpha 2$ , then the mature protein would consist of 1015 amino acids and have a  $M_r$  of 111,857.

**Location of the Transcription Initiation Site and Nucleotide Sequence of the 5'-Flanking Region**—To map the  $\alpha 2$  transcrip-

tion 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)<sup>+</sup> 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)<sup>+</sup> 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  $\alpha 2$  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  $\alpha 2$  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  $\alpha 2$  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,

TABLE I  
Exon/Intron boundaries in the human Na,K-ATPase  
 $\alpha 2$  subunit gene

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 NH<sub>2</sub>-terminal amino acids cleaved to yield the mature protein are assigned negative numbers. Numbers above codons 5' of the splice sites enumerate the nucleotides within the mRNA after which splicing occurs. The first base of the ATG initiation codon is designated +1. Intron sizes (bp) are shown in parentheses.

| EXON                      | INTRON                                 | EXON               |
|---------------------------|--|--------------------|
| 12                        |  |                    |
| 1 CGT GGG<br>Arg Gly -2   | /gtgagt----- (5020) ---ctccctccctcag/  | GCT GGC<br>Ala Gly |
| 117                       |  |                    |
| 2 GCA ATG<br>Ala Met 34   | /gtgagg----- (181) ---ctcccttccccag/   | GAT GAC<br>Asp Asp |
| 177                       |  |                    |
| 3 TCC AAG<br>Ser Lys 54   | /gtgagt----- (1961) ---cttgctccctcag/  | GGC CTC<br>Gly Leu |
| 381                       |  |                    |
| 4 GAC AAT<br>Asp Asn 122  | /gtgagc----- (526) ---ttctccttaccag/   | CTA TAT<br>Leu Tyr |
| 495                       |  |                    |
| 5 CCT CAG<br>Pro Gln 160  | /gtaaga----- (239) ---tggtctccccacag/  | CAA GCC<br>Gln Ala |
| 630                       |  |                    |
| 6 TGT AAG<br>Cys Lys 205  | /gtgagg----- (705) ---ccttctctctcag/   | GTG GAT<br>Val Asp |
| 748                       |  |                    |
| 7 GAA G<br>Glu Gly 245    | /gtgaga----- (1419) ---caccatgttgcag/  | GC ACT<br>Gly Thr  |
| 1017                      |  |                    |
| 8 GTC ACT<br>Val Thr 334  | /gtgagt----- (831) ---ctttctctaccag/   | GTG TGC<br>Val Cys |
| 1216                      |  |                    |
| 9 TCT G<br>Ser Gly 401    | /gtgatt----- (129) ---catttctctccag/   | GG GCC<br>Gly Ala  |
| 1326                      |  |                    |
| 10 TCT AAG<br>Ser Lys 437 | /gtaggg----- (176) ---tctcttctggcag/   | CGG GAC<br>Arg Asp |
| 1461                      |  |                    |
| 11 TAC CAG<br>Tyr Gln 482 | /gtctgc----- (701) ---gttgctctctccag/  | CTG TCT<br>Leu Ser |
| 1651                      |  |                    |
| 12 CTG G<br>Leu Gly 546   | /gtgaga----- (130) ---tgcccccttttag/   | GA TTC<br>Gly Phe  |
| 1827                      |  |                    |
| 13 ATC AAG<br>Ile Lys 604 | /gtactg----- (3885) ---tccttcccactag/  | GTG ATC<br>Val Ile |
| 1964                      |  |                    |
| 14 CCC AG<br>Pro Arg 650  | /gtgagg----- (524) ---accacccctccag/   | A GAA<br>Arg Glu   |
| 2115                      |  |                    |
| 15 AGG CAG<br>Arg Gln 700 | /gtgagc----- (138) ---ccgtcttccctccag/ | GGA GCC<br>Gly Ala |
| 2284                      |  |                    |
| 16 GAG G<br>Glu Gly 757   | /gtgagg----- (236) ---aatctccccacag/   | GC CGC<br>Gly Arg  |
| 2439                      |  |                    |
| 17 GAT ATG<br>Asp Met 808 | /gtgagc----- (253) ---ctacccccacag/    | GTC CCT<br>Val Pro |
| 2563                      |  |                    |
| 18 ATC G<br>Ile Gly 850   | /gtgcgc----- (199) ---cctctaccacag/    | GG ATG<br>Gly Met  |
| 2709                      |  |                    |
| 19 GAG TGG<br>Glu Trp 898 | /gtgagt----- (185) ---accctccctccag/   | ACC TAT<br>Thr Tyr |
| 2840                      |  |                    |
| 20 ATG AA<br>Met Lys 942  | /gtgagt----- (688) ---ctttgcctttcag/   | G AAC<br>Lys Asn   |
| 2942                      |  |                    |
| 21 CTC AA<br>Leu Lys 976  | /gtgagt----- (151) ---ccattgctttcag/   | A GTC<br>Lys Val   |
| 3034                      |  |                    |
| 22 GGT G<br>Gly Gly 1007  | /gtaagc----- (1304) ---atctctctaacag/  | GC TGG<br>Gly Trp  |

CAACAAAC, resembling CCAAT is located 83 bp 5' of the transcription initiation site. A potential Sp1-binding site, GGGGCGGG (Dyran *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  $\alpha 2$  sequence was examined for the presence of human 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  $\alpha 2$  and  $\alpha 3$  Gene Structure**—Both the  $\alpha 2$  and  $\alpha 3$  (Ovchinnikov *et al.*, 1988) genes contain 23 exons and 22 introns. With the exception of introns 1 and 10, the introns in the  $\alpha 2$  and  $\alpha 3$  genes occur in exactly the same positions. Exon 1 in  $\alpha 2$  includes the 5'-untranslated sequence and encodes the first four amino acids that are apparently removed posttranslationally, while exon 1 in  $\alpha 3$  contains the 5'-untranslated sequence and encodes the first two amino acids. In the  $\alpha 2$  gene, intron 10 occurs between the codons for Lys-437 and Arg-438, while in  $\alpha 3$  it interrupts the Arg codon (AG-G). Although the positions of introns in the  $\alpha 2$  and  $\alpha 3$  genes are basically the same, there appears to be little similarity in terms of intron size or sequence. The size of nine  $\alpha 3$  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  $\alpha 2$  and  $\alpha 3$  introns. For example, intron 13, which is the second largest (3.9 kb) intron in the  $\alpha 2$  gene, is the smallest (70 bp) reported  $\alpha 3$  intron. There is also no apparent sequence similarity between corresponding  $\alpha 2$  and  $\alpha 3$  introns. In addition, the occurrence of repetitive elements is not conserved. Whereas  $\alpha 2$  has five *Alu* repeats, one copy each in introns 1, 3, 13, 22, and in the 3'-untranslated region, an analysis of the published  $\alpha 3$  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 (Craik *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  $\alpha 2$  mRNA.** The sequence corresponding to the human  $\alpha 2$  mRNA was deduced by comparing the  $\alpha 2$  gene sequence to the rat  $\alpha 2$  cDNA and by locating intron/exon boundaries conforming to splice junction consensus sequences. The deduced amino acid sequence is *numbered* starting with the NH<sub>2</sub> terminus of the mature protein (Lyttton, 1985a) and is preceded by five amino acids that occur in the primary translation product. Positions of introns are indicated by *arrows*. The phosphorylation site (Asp-369) is indicated by an *asterisk*.



FIG. 3. Nucleotide sequence of the 5' end of the human Na,K-ATPase  $\alpha 2$  gene. 5'-Flanking sequences are indicated by *lower case letters*; transcribed sequences occurring in the first exon are indicated by *upper case letters*. Amino acids encoded within the first exon are indicated below the appropriate codons. A *diagonal line* indicates the beginning of the first intron. Oligonucleotides used for primer extension analysis are *underlined*. An *SphI* site used in generating the 230-nucleotide S1 nuclease protection probe is indicated. Transcription initiation sites are indicated by *large arrows*. An apparent TATA sequence is *boxed*. A polypurine-polypyrimidine region adjacent to the apparent TATA box is indicated by *brackets*. An imperfect mirror symmetry repeat sequence within this region is *underlined*; the center of symmetry is indicated by an *asterisk*. Sequences exhibiting similarity to transcription factor (AP-1, AP-2, Sp1, CP1/CP2, NF-1, and CACCC factors) or hormone receptor binding site (glucocorticoid response element) consensus sequences are *underlined*. Numbers to the right of the figure refer to nucleotide positions relative to the first transcription start site.

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  $\alpha 1$  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  $\alpha 2$  gene and the fast-twitch skeletal muscle Ca-ATPase (Korczak *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

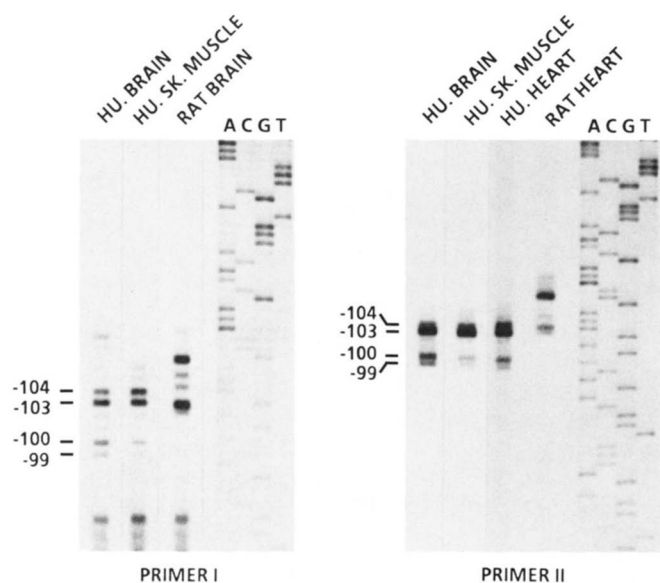


FIG. 4. **Primer extension analysis of the 5' end of the human Na,K-ATPase  $\alpha 2$  gene.** Two synthetic oligonucleotide primers, 30 and 28 nucleotides in length (*primers I and II*, Fig. 3), complementary to the 5'-untranslated region of the human  $\alpha 2$  mRNA, were annealed to 50  $\mu$ g of total RNA from human brain or skeletal muscle or to 15  $\mu$ g of poly(A)<sup>+</sup> RNA from human heart. Two oligonucleotide primers, 30 and 29 nucleotides in length and complementary to the same region of the rat  $\alpha 2$  cDNA (Shull *et al.*, 1986), were annealed to 15  $\mu$ g of poly(A)<sup>+</sup> RNA from rat brain or heart. The products resulting from extension with avian myeloblastosis virus reverse transcriptase were analyzed by electrophoresis on 8% denaturing polyacrylamide gels. The human-derived synthetic oligonucleotides were also used as sequencing primers for Sanger sequencing of a fragment containing the 5' end of the gene. The sequence read off the gel is the antisense strand. *Left panel*, primer extension and sequencing with primer I. *Right panel*, primer extension and sequencing with primer II.

aspartyl-phosphate transport ATPases. Mutation of this residue to aspartate confers vanadate-resistant ATPase activity to the fungal H<sup>+</sup>-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 Greeb, 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  $\alpha 2$  gene (amino acids 497–519), contains the conserved lysine (Lys-500 in  $\alpha 2$ ) 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 NH<sub>2</sub>-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  $\alpha 2$ , in the sarcoplasmic reticulum fast-twitch Ca-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.

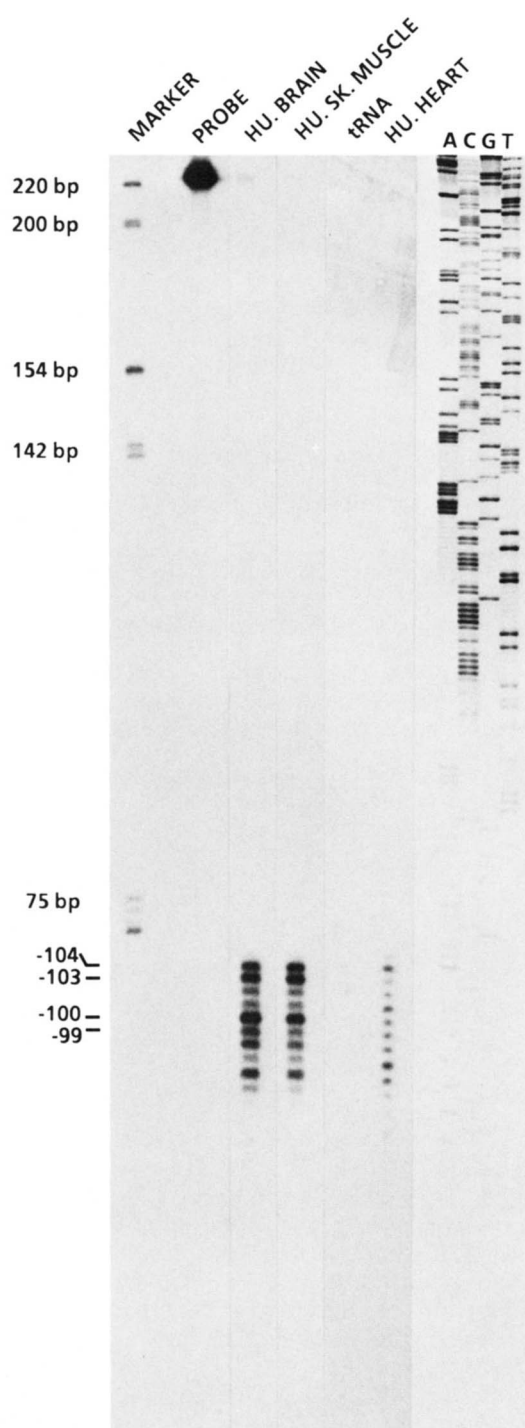


FIG. 5. **S1 nuclease mapping of the 5' end of the human Na,K-ATPase  $\alpha 2$  gene.** A single-stranded <sup>32</sup>P-labeled fragment beginning 36 bp 5' of the translation start site and extending 230 bp 5' was hybridized to 50  $\mu$ g of total RNA from human brain or skeletal muscle or to 15  $\mu$ g of poly(A)<sup>+</sup> RNA from human heart. S1 nuclease protection analysis was performed as described in "Experimental Procedures." The protected fragments were analyzed on 8% denaturing polyacrylamide gels. *Lane 1*, size markers; *lane 2*, undigested probe; *lane 3*, 50  $\mu$ g of total RNA from human brain; *lane 4*, 50  $\mu$ g of total RNA from human skeletal muscle; *lane 5*, 10  $\mu$ g of yeast tRNA; *lane 6*, 15  $\mu$ g of poly(A)<sup>+</sup> RNA from human heart; *lanes 6–9*, sequencing reaction using primer I, which was used in generating the S1 probe (see "Experimental Procedures").

A second region of similarity in cation-transporting ATPases occurs between residues 574 and 625 in  $\alpha 2$ . This segment contains a 19-amino acid sequence (amino acids 594–

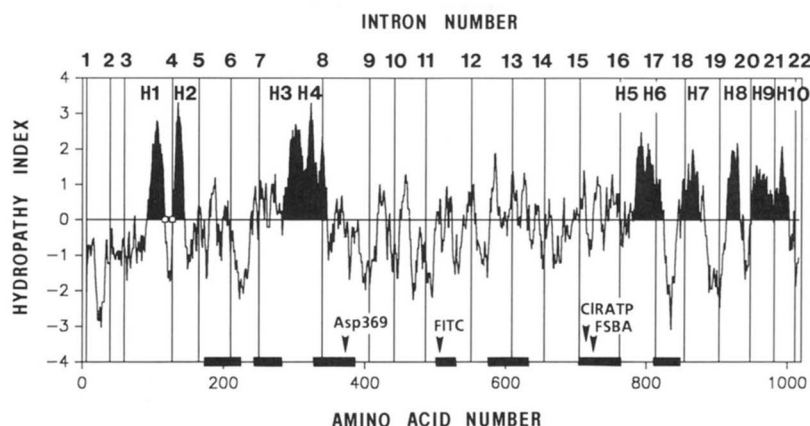


FIG. 6. **Correlation of intron locations with structural features.** Positions of introns are depicted as vertical lines on a hydropathy plot of the predicted Na,K-ATPase  $\alpha 2$  protein sequence. The hydropathy plot was generated using the algorithm of Kyte and Doolittle (1982) using a window of 13 residues. Hydrophobic residues are shown above and hydrophilic residues below the horizontal line. Regions which, based on the algorithms of Kyte and Doolittle and Eisenberg *et al.* (1984), may span the membrane are filled with black and denoted H1-H10. Regions of the  $\alpha 2$  protein sequence which exhibit homology among mammalian cation transporting ATPases (Na,K-ATPase, Ca-ATPase, H,K-ATPase) are indicated by black bars on the lower line of the figure. Asp-369 indicates the aspartic residue which is phosphorylated during the reaction cycle. The binding sites for fluorescein 5'-isothiocyanate, CIRATP, and 5'-(*p*-fluorosulfonyl)benzoyl adenosine are indicated. The amino acid residues at the H1-H2 boundary which are involved in differences in species sensitivity to inhibition by ouabain are indicated by white boxes.

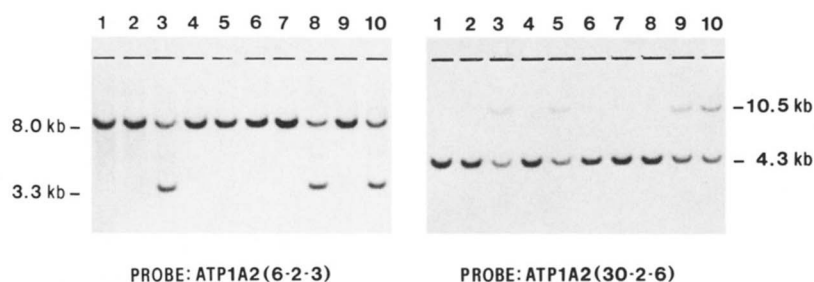


FIG. 7. **Restriction fragment length polymorphisms detected with human Na,K-ATPase gene probes.** DNA was isolated from peripheral leukocytes of 10 unrelated individuals, digested with *Bgl*II, size fractionated by agarose gel electrophoresis, transferred to nylon membranes and hybridized with Na,K-ATPase  $\alpha 2$  gene probes. Probe ATP1A2 (6-2-3) is a 2.5-kb *Eco*RI fragment representing the 5' end of the  $\alpha 2$  gene. Probe ATP1A2 (30-2-6) is a 0.6-kb fragment from the 3' portion of the  $\alpha 2$  gene. Allele sizes are indicated.

612) within which occurs a conserved Thr-Gly-Asp which may correspond to the COOH-terminal ATP-binding loop of adenylate kinase (Taylor and Green, 1989; Walker *et al.*, 1982). The aspartic acid residue (Asp-611) may correspond to the aspartate found in other adenine nucleotide-binding proteins which has been suggested to be involved in binding the magnesium ion in Mg-ATP (Walker *et al.*, 1982) or alternatively, in forming hydrogen bonds interacting with the transition state to facilitate catalysis (Al-Shawi *et al.*, 1988). In  $\alpha 2$ , intron 13 falls rather centrally (between amino acids 604 and 605) within this 19-residue sequence. A third region highly conserved among aspartyl-phosphate transport ATPases occurs between residues 704 and 758 of  $\alpha 2$  and contains a peptide fragment which is labeled with the ATP analog, 5'-(*p*-fluorosulfonyl)benzoyl adenosine (Ohta *et al.*, 1986). A second conserved Thr-Gly-Asp sequence is located at residues 707-709. The aspartate (Asp-709) is labeled with the ATP derivative CIRATP (Ovchinnikov *et al.*, 1987b). This region may form part of the nucleotide-binding domain (Ohta *et al.*, 1986) or it may form a hinge region contacting the phosphorylation and nucleotide-binding domains (Taylor and Green, 1989). In  $\alpha 2$ , 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  $\alpha 2$  Probes**—Several restriction fragments from within the  $\alpha 2$  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 (*Bam*HI, *Bgl*II, *Eco*RI, *Hind*III, *Kpn*I, *Msp*I, *Pst*I, *Pvu*II, *Sac*I, and *Taq*I). Two probes which detected polymorphisms are shown in Fig. 7. Probe ATP1A2 6-2-3 consists of a 2.5-kb *Eco*RI fragment from the 5' end of the  $\alpha 2$  gene and includes exon 1. This probe detects a two-allele polymorphism in *Bgl*II-digested



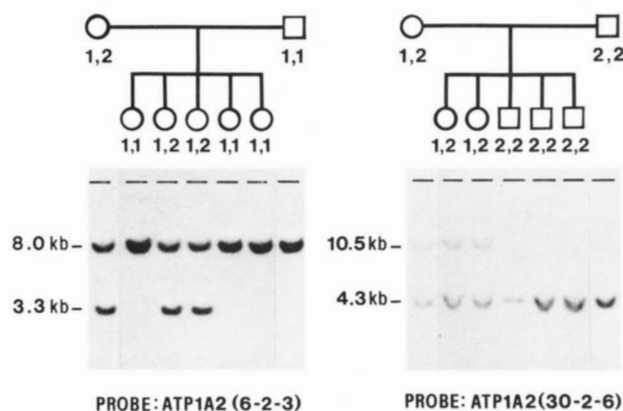


FIG. 8. Mendelian inheritance of Na,K-ATPase restriction fragment length polymorphisms. Genomic DNAs from members of two families were digested with *Bgl*II, size-fractionated by gel electrophoresis, transferred to nylon membranes, and hybridized with the probes as indicated. The digestion pattern for each member of the family is shown below that individual's position in the pedigree.

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 *Eco*RI fragment from the 3' portion of the  $\alpha 2$  gene and includes exons 21 and 22. It exhibits a polymorphism only with *Bgl*II, 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  $\alpha 2$  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  $\alpha 2$  gene spans approximately 25 kb and, like the human Na,K-ATPase  $\alpha 3$  gene (Ovchinnikov *et al.*, 1988) and the rabbit fast-twitch skeletal muscle sarcoplasmic reticulum Ca-ATPase gene (Korczak *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  $\alpha 2$  and  $\alpha 3$  genes are basically the same, the intron size and sequence and the position of *Alu* repeats are not conserved.

As the  $\alpha 2$  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  $\alpha 2$  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  $\alpha$  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  $\alpha 1$  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  $\alpha 2$  sequences occurs at residue 111 (human, Gln; rat, Leu). Thus, there may be differences in ouabain binding between the human and rat  $\alpha 2$  isoforms. The Gln-111 and Asn-122 residues are also found in human (Kawakami *et al.*, 1986) and pig (Ovchinnikov *et al.*, 1988)  $\alpha 1$  and in human  $\alpha 3$  (Ovchinnikov *et al.*, 1988). The codon for Asn-122 (AAT) in the  $\alpha 2$  gene borders intron 4. The last nucleotide of the exon bordering the splice donor site is usually a G. However, in this case the nucleotide is a T. In the  $\alpha 3$  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  $\alpha 2$  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  $\alpha 2$  isoform is expressed, demonstrate that there are four transcription initiation sites clustered between -104 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  $\alpha 2$  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  $\alpha 2$  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  $\alpha 2$  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 AGCCAA hexanucleotide, the minimal recognition site for CTF/NF-1 (Jones *et al.*, 1987).

The CACCC element, GCCACACCC (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  $\alpha 2$  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

Analysis of the 5'-flanking region of the human Na,K-ATPase  $\alpha 2$  gene for potential transcription factor and hormone receptor-binding sites

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<br>TGA <sub>C</sub> TCA  | AP-2<br>CCCCAGGC  | CACCC FACTOR<br>GCCACACCC                           | GRE<br>GGTACANNNTGTTCT   |
|---|---|---|--|
| TtACTCA -1211<br>TgGTCA -1160<br>TGACTct -1064<br>TGACTCc -573<br>TgCTCA -546<br>TGAGTCA -297 | CCCCAGGC -1101<br>aCCCCAGGC -241<br>CCaCAGGC -199   | aCaACACCC -1190<br>GtCcCACCC -723<br>GggACACCC -627 | GGggaAtccTGtCT -1321<br>tcctCtgacGTTcC -1310<br>aGcAgAattTtTCT -1281<br>aaTtttttCtTtTCT -1276<br>aGctgggtccTGtCT -1174<br>aGagaggccGTTCT -1007<br>caTcttagagGTTCT -874<br>GGaACAgagTGtCT -468<br>GGagacctgaGTTCT -414<br>taggCcacaTGtCT -329 |
| CP1/CP2<br>AACCAAT<br>GG  | NF-1<br>TTGGCTNNAGCCAA  |   | GRE-RC<br>AGAACANNNTGTACC  |
| ttCCAAT -1031<br>CP1/CP2-RC<br>ATTGG <sup>TT</sup> <sub>CC</sub><br>ATTGgAT -430              | TTaGCcttGtGCCct -1565<br>TaGGCctctgGCCAA -1457<br>gTtGagaggAGCCAA -183<br>agGGCcagtgGcTAA -135<br>TTctTgtctGCCAg -1 |   | AcAACAcccctagaC -1190<br>AGAcCagctgGtCct -1179<br>gGAACAgagTGgtCt -468<br>AGAAgAgttaGgcCa -337   |

The octamer motif, ATGCAAAT, 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 (ATTCAAAT) 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 Franza, 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 Franza, 1988). The 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 Franza, 1988), the possibility that the Na,K-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, CCCCAGGC, 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, TGTTCT, which forms part of a 15-bp imperfect dyad symmetry element, GGTA-CANNNTGTTCT (Karin *et al.*, 1984; Jantzen *et al.*, 1987). There are 14 sequences in the 5'-flanking region of  $\alpha 2$  that exhibit at least a 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 low. 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, AGGTAAGATCAGGGACGTGACCG, 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  $\text{Na}^+$  and  $\text{K}^+$  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, 1978). 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  $\alpha 2$  gene which reveal restriction fragment length polymorphisms and thus should be useful in testing the potential role of a defective Na,K-ATPase  $\alpha 2$  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  $\alpha 2$  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**, 283–297.
- Arriza, J. L., Weinberger, C., Cerelli, G., Glaser, T. M., Handelin, B. L., Housman, D. E., and Evans, R. M. (1987) *Science* **237**, 268–275.
- Bastide, F., Meissner, G., Fleisher, S., and Post, R. L. (1973) *J. Biol. Chem.* **248**, 8385–8391.
- Blake, C. C. F. (1978) *Nature* **273**, 267.
- Blake, C. C. F. (1985) *Int. Rev. Cytol.* **93**, 149–185.
- Bland, R. D., and Boyd, C. A. R. (1986) *J. Appl. Physiol.* **61**, 507–515.
- Breathnach, R., and Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349–383.
- Brent, G. A., Larsen, P. R., Harney, J. W., Koenig, R. J., and Moore, D. D. (1989) *J. Biol. Chem.* **264**, 178–182.
- Britten, R. J., Baron, W. F., Stout, D. B., and Davidson, E. H. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 4770–4774.
- Cato, A. C. B., and Weinmann, J. (1988) *J. Cell Biol.* **106**, 2119–2125.
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
- Chodosh, L. A., Baldwin, A. S., Carthew, R. W., and Sharp, P. A. (1988) *Cell* **53**, 11–24.
- Chou, P. Y., and Fasman, G. D. (1974) *Biochemistry* **13**, 222–245.
- Christophe, D., Cabrer, B., Bacolla, A., Targovnik, H., Pohl, V., and Vassart, G. (1985) *Nucleic Acids Res.* **13**, 5127–5144.
- Cole, D. F. (1961) *Br. J. Ophthalmol.* **45**, 202–217.
- Craik, C. S., Sprang, S., Fletterick, R., and Rutter, W. J. (1982) *Nature* **299**, 180–182.
- Curran, J., and Fianza, B. R., Jr. (1988) *Cell* **55**, 395–397.
- DeWeer, P. (1985) in *The Kidney: Physiology and Pathophysiology* (Seldin, D. W., and Giebisch, G., eds) pp. 31–48, Raven Press, New York.
- Dierks, P., van Ooyen, A., Cochran, M. D., Dobkin, C., Reiser, J., and Weissmann, C. (1983) *Cell* **32**, 695–706.
- Dynan, W. S., Sazer, S., Tjian, R., and Schimke, R. T. (1986) *Nature* **319**, 246–248.
- Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984) *J. Mol. Biol.* **179**, 125–142.
- Farley, R. A., Tran, C. M., Carilli, C. T., Hawke, D., and Shively, J. E. (1984) *J. Biol. Chem.* **259**, 9532–9535.
- Felsenfeld, D. P., and Sweadner, K. J. (1988) *J. Biol. Chem.* **263**, 10932–10942.
- Fletcher, C., Heintz, N., and Roeder, R. G. (1987) *Cell* **51**, 773–781.
- Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120.
- Ghislain, M., Schlessner, A., and Geoffeau, A. (1987) *J. Biol. Chem.* **262**, 17549–17555.
- Gilbert, W. (1978) *Nature* **271**, 501.
- Gilbert, W. (1985) *Science* **228**, 823–824.
- Glass, C. K., Franco, R., Weinberger, C., Albert, V. R., Evans, R. M., and Rosenfeld, M. G. (1987) *Nature* **329**, 738–741.
- Go, M. (1981) *Nature* **291**, 90–92.
- Green, N. M., Taylor, W. R., and MacLennan, D. H. (1988) in *The Ion Pumps: Structure, Function and Regulation* (Stein, W. D., ed) pp. 15–24, Alan R. Liss, Inc., New York.
- Greene, J. M. (1987) 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. 4.6.1–4.6.9, John Wiley & Sons, New York.
- Gusella, J. F. (1986) *Annu. Rev. Biochem.* **55**, 831–854.
- Hara, Y., Urayama, O., Kawakami, K., Nojima, H., Nagamune, H., Kojima, T., Ohta, T., Nagano, K., and Nakao, M. (1987) *J. Biochem. (Tokyo)* **102**, 43–58.
- Herrera, V. L. M., Emanuel, J. R., Ruiz-Opazo, N., Levenson, R., and Nadal-Ginard, B. (1987) *J. Cell Biol.* **105**, 1855–1865.
- Hilton, P. J. (1986) *N. Engl. J. Med.* **314**, 222–229.
- Hsu, Y.-M., and Guidotti, G. (1989) *Biochemistry* **28**, 569–573.
- Htun, H., and Dahlberg, J. E. (1988) *Science* **241**, 1791–1796.
- Ismail-Beigi, F., Pressley, T. A., Haber, R. S., Gick, G. G., Loeb, J. N., and Edelman, I. S. (1988) *J. Biol. Chem.* **263**, 8162–8167.
- Jantzen, H.-M., Strähle, U., Gloss, B., Stewart, F., Schmid, W., Boshart, M., Miksicsek, R., and Schütz, G. (1987) *Cell* **49**, 29–38.
- Jones, K. A., Kadonaga, J. T., Rosenfeld, P. J., Kelly, T. J., and Tjian, R. (1987) *Cell* **48**, 79–89.
- Jurka, J., and Smith, T. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 4775–4778.
- Karin, M., Haslinger, A., Holtgreve, H., Richards, R. I., Krauter, P., Westphal, H. M., and Beato, M. (1984) *Nature* **308**, 513–519.
- Kawakami, K., Ohta, T., Nojima, H., and Nagano, K. (1986) *J. Biochem. (Tokyo)* **100**, 389–397.
- Kingston, R. E. (1987) 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. 4.8.1–4.8.3, John Wiley & Sons, New York.
- Kirley, T. L., Wallick, E. T., and Lane, L. K. (1984) *Biochem. Biophys. Res. Commun.* **125**, 767–773.
- Korczak, B., Zarain-Herzberg, A., Brandl, C. J., Ingles, C. J., Green, N. M., and MacLennan, D. H. (1988) *J. Biol. Chem.* **263**, 4813–4819.
- Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
- Lane, L. K., Shull, M. M., Whitmer, K. R., and Lingrel, J. B. (1989) *Genomics*, in press.
- Layzer, R. B. (1982) *Annu. Neurol.* **11**, 547–552.
- Lingrel, J. B., Orlowski, J., Shull, M. M., and Price, E. M. (1989) *Prog. Nucleic Acid. Res. Mol. Biol.*, in press.
- Lonberg, N., and Gilbert, W. (1985) *Cell* **40**, 81–90.
- Lytton, J. (1985a) *Biochem. Biophys. Res. Commun.* **132**, 764–769.
- Lytton, J. (1985b) *J. Biol. Chem.* **260**, 10075–10080.

- 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 Lett.* **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., Dzhandzugazyan, 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., Hüppi, K., Brack, C., Traunecker, A., 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 Margole, 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. M., and Lingrel, J. B. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 4039-4043
- Sun, L., Paulson, 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., Grishin, A. V., Kiyatkin, N. I., Kostina, M. B., Sverdlov, V. E., Modyanov, N. N., and Ovchinnikov, Yu. A. (1987) *FEBS Lett.* **217**, 275-278
- 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
- Sweadner, 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. D., Collier, D. A., Hanvey, 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., Schneider, 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