Characterization of the Human Gene for a Newly Discovered Carbonic Anhydrase, CA VII, and Its Localization to Chromosome 16

JEFFRY C. MONTGOMERY,[*,1] PATRICK J. VENTA,[*] ROGER L. EDDY,[†] YAO-SHAN FUKUSHIMA,[†] THOMAS B. SHOWS,[†] AND RICHARD E. TASHIAN[*,2]

*Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan 48109-0618; and †Department of Human Genetics, Roswell Park Memorial Institute, New York State Department of Health, Buffalo, New York 14263

Received May 9, 1991; revised July 19, 1991

Six carbonic anhydrase (CA) isozymes (CA I–VI) in mammals and other amniotes have been described. We have isolated an additional CA gene from a human genomic library and designated its putative product carbonic anhydrase VII (CA VII). The gene is approximately 10 kb long and contains seven exons and six introns found at positions identical to those determined for the previously described CA I, CA II, and CA III genes. The finding of a 17-bp GT-rich segment in a position 28 bp downstream of the poly(A)$^+$ signal and the high correspondence of the 5' and 3' splice sites of the six introns with consensus junction sequences are consistent with the gene being functional. The 5' flanking regions of the CA VII gene do not contain the TATA and CAAT promoter elements usually found within 100 bp upstream of transcription initiation, but do contain a TTTAA sequence 102 nucleotides upstream of the initiation codon. The 5' region of the gene (−243 to +551) is GC-rich and contains 80 CpG dinucleotides and four possible Sp1 (GGGCGG or CCGCCC) binding sites. Northern analysis has identified the salivary gland as a major site of expression. The derived amino acid sequence of the CA VII gene is 263 amino acids long and has 50, 56, and 49% identity with human CA I, CA II, and CA III, respectively. No differences were found at any of the 39 positions that have remained invariant in all mammalian CA isozymes sequenced to date. Based on analysis of interspecific somatic cell hybrids, the human CA VII gene, CA7, was assigned to chromosome 16, with localization to the long arm at the q21–23 region by in situ hybridization. This is in contrast to the location of the CA I, CA II, and CA III gene cluster on human chromosome 8 and that of the human CA VI gene on chromosome 1. © 1991 Academic Press, Inc.

INTRODUCTION

The carbonic anhydrases (CA; EC 4.2.1.1) are unusually diverse in their distribution for a multilocus enzyme family (Tashian, 1989). In mammals, six CA isozymes (CA I–CA VI) and two other partially characterized CA or CA-like isozymes, CA Y (= CA V?) and a CA-related protein (CARP), have been described (cf. Amor-Gueret and Levi-Strauss, 1990; Kato, 1990; Hewett-Emmett and Tashian, 1991). The expression of the CA genes ranges from their presence in certain cells of virtually all tissues (e.g., CA II) to their seemingly limited presence in one cell type of a specific tissue, e.g., CA VI in the serous accinar cells of the partoid and submandibular glands (Aldred et al., 1991) and CARP in the Purkinje cells of the cerebellum (Kato, 1990). Further diversity may be seen in their intracellular distribution where these CA isozymes may be cytosolic (CA I, CA II, CA III), membrane-bound (CA IV), mitochondrial (CA V), or present in secretory granules (CA VI) (cf. Tashian, 1989).

The CA isozyme genes vary in length from about 10 kb in human CA III to about 50 kb in human CA I, and they all contain seven coding exons that are interrupted by six introns at identical positions (cf. Hewett-Emmett and Tashian, 1991). The CA I gene, however, is unique in that it not only has an additional, noncoding exon, upstream of exon 1 (Fraser et al., 1989; Lowe et al., 1990), but also is under the control of two promoters that are apparently responsible for the expression of CA I either in red cells or in tissues other than red cells (Fraser et al., 1989; Tash-

ian *et al.*, 1990). The CA I, CA II, and CA III genes are closely linked on chromosome 3 in the mouse (Beechey *et al.*, 1990) and on the long arm of chromosome 8 in humans (Davis *et al.*, 1987; Nakai *et al.*, 1987). The human CA VI gene is located on the tip of the short arm on chromosome 1 (Sutherland *et al.*, 1989).

During the course of screening a human genomic library with a mouse CA II cDNA probe, we isolated a CA-like gene (originally designated by us as CA Z), which on preliminary analysis appeared to code for a new CA isozyme. Here, we report the structure and chromosomal location of this new CA gene, the product of which we designate carbonic anhydrase VII (CA VII).

MATERIALS AND METHODS

Restriction Mapping

Physical maps of inserts from clones H10 and H11 were determined by single, double, and triple restriction digest analyses. Positioning of ambiguous sites was facilitated by hybridizing radiolabeled *Eco*RI fragments from the inserts to Southern blots of digested insert DNA. Restriction fragments from clones H21, 9-1B, and 2-4 were ordered and aligned by comparison to maps from clones H10 and H11. Certain subregions of some of the clones were mapped in greater detail by subcloning the fragment of interest prior to analysis of multiple digests with several additional enzymes.

DNA Blots

Restriction digests of genomic or cloned DNA were transferred from agarose gels to GeneScreen Plus (Dupont/NEN) by the alkaline method (Chomczynski and Qasba, 1984). Prehybridizations were carried out for 1 to 18 h at 42°C in 10–45% formamide, 5× SSPE (1× SSPE is 0.18 M NaCl, 10 mM NaH$_2$PO$_4$ (pH 7.4), 1.0 mM EDTA), 1.0% SDS, and 10% dextran sulfate. DNA probes were labeled by the mixed oligonucleotide random-priming technique (Hodgson and Fisk, 1987) and RNA probes were generated by incorporation of [α-^{32}P]UTP during SP6 transcription from linearized subclones in pSP64 or pSP65 vectors (Promega). Probes were added directly to the prehybridization solution for 18 to 48 h and washed initially at room temperature in two changes of 2× SSC, 0.1% SDS (1× SSC is 0.15 M NaCl, 0.015 M sodium citrate). Further washes with the same solution were performed at 55°C for low-stringency hybridization experiments, while higher-stringency hybridization filters were washed with at least two addi-

tional changes of 0.1× SSC, 0.1% SDS at 55–65°C. The filters were exposed to Kodak XAR-5 film at −70°C with intensifying screens. Blots were stripped for rehybridization by washing in 0.4 M NaOH and neutralizing with 0.2 M Tris–Cl (pH 7.5), 0.1× SSC, 0.1% SDS each for 1 h at 53°C.

Isolation of Genomic Clones

A chromosomal walk to isolate upstream clones overlapping the CA VII gene locus was carried out by screening 1 × 10^6 λ EMBL-3 recombinants (prepared by T. M. Johnson, this laboratory) made with partially digested DNA from an individual deficient in CA II (Sly *et al.*, 1985). Standard plaque lifts (Maniatis *et al.*, 1982) were hybridized to a probe derived from the most 5' *Eco*RI fragment of clones H10 and H11. Positive plaques were purified by two additional rounds of replating and screening. Phage DNA was prepared by polyethylene glycol precipitation of bacteriophage particles from 1 liter of liquid lysates generated by infection of *Escherichia coli* strain LE392 with 5 to 50 × 10^8 phage (Maniatis *et al.*, 1982).

Subcloning and Sequencing

Fragments of interest were routinely subcloned into the pUC series of plasmid vectors [pUC8, pUC9 (Viera and Messing, 1982; Hanna *et al.*, 1984); pUC18, pUC19 (Norrander *et al.*, 1983)] or pBSM13+ or pBSM13− (Stratagene) using standard methods. A series of nested deletion subclones was created to facilitate sequencing the regions containing exons 5, 6, and 7 (Pontz *et al.*, 1982; Henikoff, 1984). Fragments to be sequenced by the chemical cleavage method were end-labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP, digested to create fragments with only one labeled end, and separated on polyacrylamide gels prior to base modification and cleavage reactions (Maxam and Gilbert, 1980). Double-stranded, supercoiled plasmids, prepared by the boiling method (Maniatis *et al.*, 1982), were utilized for dideoxy chain termination sequencing (Chen and Seeburg, 1985; Hattori and Sakaki, 1986) with [α-^{35}S]dATP (Biggin *et al.*, 1983) and the Sequenase T7 DNA polymerase (Tabor and Richardson, 1987). Ladders were resolved on 6% polyacrylamide/7.6 M urea buffer gradient gels (Biggin *et al.*, 1983), which were fixed with 10% acetic acid and dried onto a glass plate that had been pretreated with γ-(methacryloxy)propyltrimethoxysilane (Garoff and Ansorge, 1981). Attempts were made to obtain sequence data from both strands, particularly within the exons. When this was not achieved, the same strand was sequenced at least twice. Sequences were compiled and analyzed with the aid of

the Pustell sequence analysis software (Pustell and Kafatos, 1982a,b, 1984).

Chromosomal Mapping

The 40 human–mouse somatic cell hybrid lines, which were derived from fusions of 14 unrelated human cell lines and 4 mouse cell lines, came from previously described hybrid sets (Shows et al., 1982, 1984; Sakaguchi and Shows, 1982). Each hybrid was analyzed for the presence or absence of the human CA VII fragment, as well as for the presence of individual human chromosomes, mapped enzyme markers, and in some cases mapped DNA probes (Shows et al., 1982, 1984, 1978; Shows, 1983). Southern blots of XbaI-digested DNA from the hybrids were hybridized to the ^{32}P-radiolabeled 3.0-kb EcoRI fragment from clone H11 as described (Shows, 1983; Shows et al., 1984). Filters were washed twice for 30 min at 65°C with 0.1× SSC, 0.1% SDS and exposed to film for 1–3 days at −70°C with intensifying screens. Chromosomes for in situ hybridization were prepared from human lymphocytes by previously detailed methods (Dutrillaux and Viegas-Pequignot, 1981; Zabel et al., 1983) for obtaining long chromosomes. Hybridization to the same probe used above (^{3}H-labeled), autoradiography, and G-banding were also as described (Zabel et al., 1983).

RESULTS

Isolation and Identification of the Carbonic Anhydrase VII Gene

The screening of the bacteriophage λ library has been described previously (Venta et al., 1984). Briefly, the insert from mouse CA II cDNA clone pMCAII (Curtis, 1983) was hybridized under reduced stringency conditions (35% formamide) to 3 × 10^5 recombinant phage. Six recombinants were plaque-purified from a total of 25 potential positive signals. Clones H10 and H11, isolated from adjacent, weakly hybridizing plaques, had inserts with restriction maps unlike those of the other clones which were subsequently shown to contain the CA I or CA II genes (Butterworth et al., 1991; Venta et al., 1985). Although the inserts of these clones differ in length (15.1 kb for H10 vs 12.1 kb for H11), the configuration of restriction sites at the ends of the inserts is identical in both clones, suggesting that the H11 insert differs from H10 only by a 3-kb internal deletion (Fig. 1). The segment of DNA that cross-hybridized to the pMCAII probe was mapped to the center of the H10 insert within 7.1-kb BamHI, 7.3-kb EcoRI, and 6.9-kb HindIII fragments (Fig. 1). Fragments of identical size were detected in a Southern blot of human genomic DNA that was probed with a subfragment of the 7.3 kb EcoRI fragment from H10 (Fig. 2). This demonstrates that no major rearrangements had taken place during cloning within a minimum region of 12.0 kb in H10 that extends from the 5′ end of the 7.1-kb BamHI fragment to the 3′ end of the 6.9-kb HindIII fragment, and that H10, not H11, faithfully represents the arrangement of DNA in this region of the genome. The simplicity of the banding pattern indicates that a gene located here is probably present as a single copy in the haploid genome.

Sequence analysis of this region revealed an open reading frame, bounded by consensus acceptor and donor splice sites, that predicted an amino acid sequence which was 56, 65, and 52% identical to the corresponding regions of the human CA I, CA II, and CA III proteins, respectively. This region of the carbonic anhydrases is precisely encoded by exon 6 of all other characterized vertebrate CA genes (Tashian et al., 1990; Hewett-Emmett and Tashian, 1991). Additional potential coding material, which corresponded to a CA-like exon 3, was identified by reduced stringency hybridization (10% formamide) of H11 subfragments to a single-stranded RNA probe derived from the first five exons of the mouse CA II cDNA (data not shown). While no additional independent regions of cross-hybridization were detected with this probe or with a probe derived from exon 7 and the 3′ untranslated region (UTR) of the mouse CA II gene, genomic or cDNA probes containing the first two exons of the human CA II gene did detect, under similar low-stringency conditions, a CA-like exon 2 in the 5.9-kb EcoRI and 7.2-kb HindIII fragments of 9-1B and 2-4, two clones obtained by genomic walking from H10 (data not shown). The inability of the mouse probe to hybridize to this region, which is present in H10 and H11, is probably a result of its slightly lower level of sequence identity with the identified exon compared to that of the human CA II gene (56% vs 60%) and a lower number of hydrogen bonds per base pair in the region of hybridization. The locations of exons 4, 5, and 7 were easily determined by directly sequencing the DNA downstream of exons 3 and 6, and a cluster of seven GC-dinucleotide-containing restriction sites (one NarI, three SmaI, one BssHII, and two SacII; data not shown) within a 0.5-kb region of an additional overlapping clone, H21, led to the discovery of exon 1 within a suspected "HTF island" (Lindsay and Bird, 1987), which is often a landmark for the 5′ end of transcribed genes (Bird, 1987).

Nucleotide Sequence of the CA VII Gene

The nucleotide sequence encompassing the CA VII gene region from 510 bp upstream of the putative
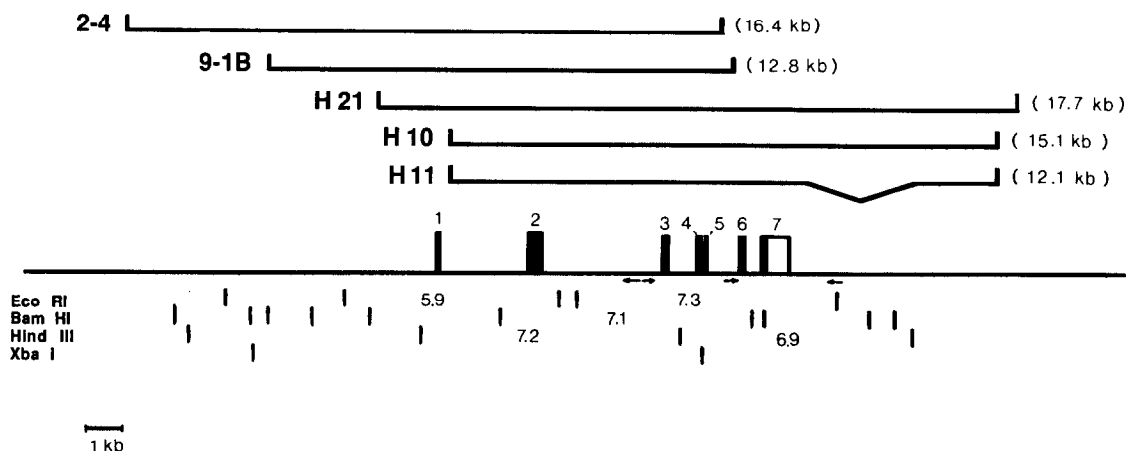
FIG. 1. Restriction endonuclease map of the human CA VII gene locus. Overlapping genomic clones are represented above the physical map of the CA VII gene. The break in clone H11 represents an internal deletion of 3.0 kb in this clone. Numbered exons are indicated by boxes in which the coding regions are filled in black. Sizes of restriction fragments discussed in the text are given in kb. Arrows indicate positions and orientations of *Alu* repeat elements.

translation initiation codon in exon 1 to approximately 1.3 kb downstream of a single poly(A)$^+$ addition signal is shown in Fig. 3. The seven exons and six
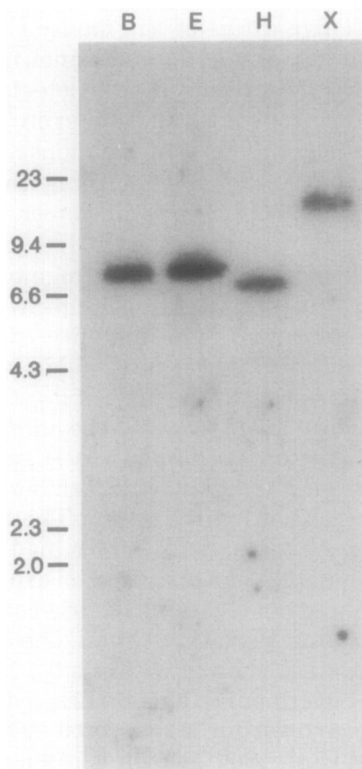
FIG. 2. Southern analysis of the CA VII gene locus. A blot of human genomic DNA was probed with a fragment containing the 3' half of exon 6. The sizes (kb) and positions of *Hind*III-digested λ marker fragments are given on the left. B, *Bam*HI; E, *Eco*RI; H, *Hind*III; X, *Xba*I.

introns define a gene of approximately 10 kb, assuming a 5' untranslated region (UTR) of about 70 nucleotides (see below) and that polyadenylation occurs near the addition signal (nucleotide 7152). The presence of a 17-nucleotide GT-rich segment (7185–7201) at a position (28 nucleotides downstream of the poly(A)$^+$ signal) where similar sequences have been found to be necessary for the formation of proper 3' termini in various mRNAs (McLauchlan *et al.*, 1985; Birnsteil *et al.*, 1985) is consistent with the signal being functional. This suggests that the putative CA VII mRNA would be 1.5–1.6 kb long and would contain about 650 untranslated nucleotides at the 3' end.

The first six exons are interrupted by intervening sequences at positions that are identical to those seen in the CA I, CA II, and CA III genes (cf. Hewett-Emmett and Tashian, 1991). These introns contain invariant GT and AG dinucleotides at their 5' and 3' ends (Breathnach and Chambon, 1981) that are part of larger consensus donor and acceptor splice junctions (Mount, 1982). The high degree of sequence identity between these splice junctions and the consensus sequences shown in Table 1 is consistent with their being functional. A pentameric consensus sequence, CT(CG)A(GTC), which has been identified as the point at which the 5' end of many spliced introns forms a lariat branch (Ruskin *et al.*, 1984), was found at the 3' end of introns 1, 2, 4, 5, and 6 within 21–29 nucleotides of the splice junction. Computer searches of a number of introns have led to the discovery of similar sequences that have been postulated to bind via hydrogen bonding with a conserved region of the U2 snRNA during the splicing process (Keller and Noon, 1984).

Four *Alu* repeat elements are located in the CA VII gene region (arrows in Fig. 1; boxed sequences in Fig. 3). One element within intron 5 (nucleotides 5208–5687) has been sequenced in its entirety and found to be flanked by identical 12-bp direct repeats. Three other *Alu* repeats have been partially sequenced. Two of these were found in intron 2 and represent the 3' end of a repeat oriented in the same direction as the CA VII gene (nucleotides 3638–3689) and the middle of a repeat oriented in the opposite direction (nucleotides 3439–3637). Another oppositely oriented element was found in the 3' UTR (nucleotides 7911–8204). A poly(GT) sequence, $(GT)_{14}$, was found in the first intron (nucleotides 2151–2178).

The region of the CA VII gene upstream of the putative translation initiation codon does not contain a consensus TATA RNA polymerase II promoter element or a CAAT box element, both of which are commonly found approximately 20–30 and 40–100 bp, respectively, upstream of the start site of transcription for many eukaryotic genes (Breathnach and Chambon, 1981). Located at the −102 position relative to the initiation codon, however, is the sequence TTTAA, which closely resembles the TTAA sequence found at the −27 position relative to the transcriptional start site of the human adenosine deaminase gene (Valerio et al., 1985), and the TTAAAA found at the −23 position relative to the start site of the human thrombospondin gene (Laherty et al., 1989). The former sequence is located in a GC-rich region upstream of the structural gene. The GC content of the 5' end of the CA VII gene is also high. Calculated over ranges of 100 bases centered around each nucleotide from −243 to +551 relative to the putative start codon, GC content ranges from 65 to 84%. Eighty CpG dinucleotides, or an average of about 1 in 10 bp, are found in this region. Not unexpectedly, several GC boxes or potential Sp1 transcription factor binding sites (Kadonaga et al., 1986) were discovered in this region, including strict consensus sequences (GGGCGG or CCGCCC) at positions −185, −143, +155, and +398 relative to the start codon.

The Predicted CA VII Protein

The protein predicted by the seven exons of the CA VII gene is 263 amino acids long. An alignment with the cytosolic human CA isozymes (CA I, II, and III) in Fig. 4 shows an extra amino acid at the N-terminus relative to the CA I protein and two extra relative to CA II and CA III; at the C-terminus, CA VII has two extra amino acids compared to CA I and one more than CA II and CA III. The only other gap needed to align the CA VII protein is found at residue 126 of CA II and CA III, where an extra amino acid is found in CA I and CA VII. In contrast, to maximally align the CA VII sequence with the predicted protein sequence of the human CA VI secreted salivary isozyme, a total of 11 gaps need to be introduced into the two sequences.

Overall, the predicted CA VII protein shares 50, 56, 49, and 37% sequence identity with human CA I, CA II, CA III, and CA VI, respectively. Sequence identity is not evenly distributed throughout the length of the protein, however. The alignment in Fig. 4 shows that there are distinct clusters of conservation and divergence between the CA VII protein and the other CAs. For example, residues 92–98, 104–111, 117–124, and 191–203 are strongly conserved regions among the CA I, CA II, CA III, and CA VI isozymes, and also remained unchanged for the most part in CA VII. In contrast, amino acids 150–170 are poorly conserved. The conserved regions of CA VII tend to fall into one or both of two categories. These residues usually either occupy positions that correspond to active site residues in human CA I and CA II (asterisks in Fig. 4) or are in positions where the residues are conserved in all mammalian carbonic anhydrase isozymes that have been sequenced to date (underlined in Fig. 4). Of 36 active site residues of CA VII, 33 are conserved in at least one other isozyme, including the key residues involved in the zinc ion ligands, His-94, His-96, and His-119, as well as Thr-199, the residue hydrogen-bonded to the zinc-bound solvent molecule (cf. Lindskog, 1982; Silverman and Vincent, 1983). Only the active site residues at positions 69, 91, and 204 of CA VII are not found in the corresponding position in at least one of the other human isozymes. In each of these cases, however, the position is poorly conserved, as each human isozyme has its own unique amino acid. The active site changes in CA VII do represent conservative replacements with respect to one of the other isozymes: Asp for Glu at codon 69 of CA II and Lys for Arg at codon 91 of CA III. While a Ser at active site residue 204 is not found in any of the other human CAs, its presence in turtle CA I (Hewett-Emmett and Tashian, 1991) suggests that this residue is not incompatible with enzyme function. There are no differences in the predicted protein sequence of CA VII at any of the 39 positions that are conserved in all vertebrate CA isozymes for which sequence information is currently available.

Tissue-Specific Expression of the CA VII Gene

Figure 5A is an autoradiogram of a Northern blot of total RNA from various baboon tissues and several mammalian cell lines that was probed with a frag-

```
ACAGCCTCTGGCAAACAGTGGAAGCTCCAGTTTTTGGACACTTGATGGTCAGTGTAGTTTGCAGTTTAAAGGGGACTTTCCAACCCCAAAAGATTCAGGACCCAGAGCTCAGGTCTCTGGT      120
GGTGGCATTTGCAGGTCTGGGGCAAGCCTAGGGAAGGGACGCCATGTGCGGCACGGGCATCTAGTGTCCCAGAGTCCGCACGGGTTGCTGGGGGGAAAAAGAGGGATAGTCGCCCAGCCAG      240
GGCAGGGGTGGGAGTTGGGGGGGAGAGGGTTACTTCCTCCTGCCCTGAATCTCCGCTCGGGGTTCAACTGGCTCGCCGGCGGGGCGAGGGCGGGGCAGGCCGGCCTCTCTGGCCCTGCTGGG      360
GGGTGGGGGGCGGGTCCCCCGGGAACTGCTCCTCCCGCCCAGCCGCTGCTTTAAGAGGCTGCTCCGCGGGTAGCGAGCGGGGGCCGGAGCCGCAGCCCGAACGAGCGGACCGAGCCGACCGGG      480
CAGGTGCACGGCTGCGGGGGACGGCAGCGGCATGACCGGCCACCACGGCTGGGGCTACGGCCAGGACGACGGGTAGGCACAGACCTCCCCCCACCGCCTCTGGCTCGGACCCCCGACCCAACT      600
                                     (1)MetThrGlyHisHisGlyTrpGlyTyrGlyGlnAspAspG
                                        1                         10
GCACTCGCCCCAACCCTCTTGCCCAGCTGGTTTCCCAGACACGGAAAGCTCCCACACTCCAGGCTGGGCGGGACCCCTCTTCCCATCCCGGCCCCGAACTTGGCCCAAGCCTTGGTCCCC      720
CGCGCCAGACGCAAAGTCCGCGCTGCGCCCCGGGTGGCCCCGCCGCCAGCAAGCGACCTGGCnnnnGCTGTGTCCTGCCGCCTCCTCCCGAAGCCCTGATTGCCCAGCCCGCTCAGACC      840
CAGGCCCAGACTCTCACTCACTCGCCCCGGGCGTGCGCACGCCTGGGGGTGGCCCAGGGCAGCGGGGGGCGGGCGCCGCGGAGCGCTGTGCGCGGGTGTCTGCGGGGAGCGCGCACAGCC      960
AAGTGAGCGGGGACCTCGGGGGGACGCGGGGCTCCGCGTGGCAAGGGCGCAGTGTCCCCCGGAGAGAATTTCCCCGCAGAAGTGGCCGAGCTGCCTTCTCGGCCGGTACGTGAGGGAAGGGG      1080
GCACTTGGGGGGCGCAAGCATGTGTCTGGGTGTCTACACCGAGCAGG----------approx. 0.1 kb----------CCCACAAGCCCTCCTnnnnnACGCGGAAGAGAGCTCCTTC      1166
ATCCTGCCTCAGTTCTTAGGACTCAGAGAAGGAGAAGAGGCAGAAGGTGAGACAGACGGCTGGAGTGTGGGTCTGGGGCAGTCAATTCCCTTTGGGGGCCATTTCTTACCAAGCAGGGGTG      1286
GGGGTGTGGGGTAAGGGGTTCCCTATTTCCCTGGGGAGCCAGTGCAAGACAGGTGGGGTGCGGGAAGAAAAAGTAATGATGGGGGGAACATCCCCCAGCCCCACTCCTCTGCGGAGAGG      1406
CAGTCACTTTGTGAGGTTGGGTGGGGGACCACAGTCTGGGAAGCCAGAAAGTTGCAGTAGACCAGAGGGAGGGTGCTGTGGAGTCTTTGGCTCCCTCTGGGGCCCCAAGAACAAAGGAGC      1526
ATATTTTGGAAGAGAACCCCCAGATCTGGGTATGCAAGAGCTGGAAAAGAGACATTCAGAAAAAGATGATATAGACAAGCAGAGGGGAGAGACAGGCAGACATGGGGCAGATGGACAGGA      1646
CCCATAGGGATCATCTTCTCCCAGCACCTGCTCTGAGCCACCTTTTGGGTGCTTCCCCCAGAGCCAGAGTCCCCTGCCTCCAGCTCTTCTTGCTAGTCATAGCCCACCAGCTCTGAGCCACC      1766
CAGTGTCCTGTGTAGCCTGCTCCATGGGCCTGCTCCCCCAAAGCAGCCACCCCATCACCCAGGCAGAGCTGGATTGTCCAGTCGAGGAAGGGAAGGACAGAGGACAGAGGTGCCGGCATG      1886
CAGCAGTGTTCACAGGGCTTCAAAAATAACCCAGAGGGCCGGGCCACGATTCTGAGTCATCCATTGGGGACACGGCCTCTGTCGGCACCTGCTGCGTGTTTTCTCAGGAAAACTTAAGAAG      2006
CCAAAAAGGGGGGAGAAGCACCCAGCCGGCAGTAGCTCTAAAAATAGGCTGCATCTCAGGATGTTCCCCAGGGAGATGCTGAGCCTCTCCGGCTGCAGATGGATCCTCTGGGAGGCAGGTGGT      2126
GGGCGCCTGTGGGTGTGTATGCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGCATGGGCATGTGTTTTGCTGTGGTCCTTTATGCAAACTTGGGTCTCCAACAGAGGGGTTTGTGTG      2246
AGAGAGCATATGCGTGTGTCAGATGTAGTCTGCAGGGCTGTGTGTGTTTGTACACATGAATGTTCTGGGCCTGGCTGCTCTGTGTGCACAATGGAAGATCATAGAAAGATTAATGTGT      2366
TCATAGGAAGATCTCTATGGACTTGCAGGCACACACATTGGGAGGGCAGTCGTGGGGGCAGTGCATGCACATGCTGTATTTATATTAGTACATATGCATGCACACATGTGTGAGTGGCCCTC      2486
AGTTCCCTCATCTGTGAAATGGGGATCTTGCTGGTTTCTATTTCACAAGGTGGTTGTGAGAGGCAAATACATGTCACAGAGTCAAAGCTAGGTTGATGTGAGTTTTTGTTGTTATTTATC      2606
ACAAACACCAACACCAAATTGCCCAAGGCACAGAAAGAGGTGAGATACTGATGGGGTCAGGTCAGGACTTCAGATGTCTGGAGGTCTCCCTGTTGGGGAGACTTTATTACTCTGAGCCAC      2726
CCCGAGCCTGTAGCTGTTATGTTCCCTTCCATACAGTACCTCTCTGGGCCTCCTATCAGCACTGGAAGGTGAGTTTTAGTGAGACCATTTCACAGACAAAGAAATGGAGGCTGAAAGGGG      2846
AAATGGCAAAGCTGGGATTGGAACTCCAAGCTGTGCAGCTCCAATGCAGGGGCTCACCTCTGAAAGGATCCTGAAGAGACCATTCCCCTATGGGTGTCCTCGGCCTGAAGCCCAGAGATC      2966
CTGGCTGGCCTGAGTCCTTGCnnTCCTCCCCCACCTGCCAGGCCCCTCGCATTGGCACAAGCTGTATCCCATTGCCCCAGGGAGATCGCCAATCACCCATCAATATCATCTCCAGCCAGGCTG      3086
                    (2)lyProSerHisTrpHisLysLeuTyrProIleAlaGlnGlyAspArgGlnSerProIleAsnIleIleSerSerGlnAlaV
                                     20                    30
TGTACTCTCCCAGCCTGCAACCACTGGAGCTTTCCTATGAGGCCTGCATGTCCCTCAGCATCACCAACAATGGCCACTCTGTCCAGGTAGACTTCAATGACAGCGATGACCGAACCGGTA      3206
alTyrSerProSerLeuGlnProLeuGluLeuSerTyrGluAlaCysMetSerLeuSerIleThrAsnAsnGlyHisSerValGlnValAspPheAsnAspSerAspAspArgThrV
     40                    50                    60                    70
AGTGGCCCCTGCCAAAGCCTGGCACCTGGCCCCTGGCCCCTTAGGGTCCTAACCTTGTGTTGGGCAGTCTCAGGAAGCATGCTATGGTGGGTAAGAAAGGTTCCTCCTCTCACCAGCTGT      3326
GAAGCCTTGGGCTGATCTTCAGTCTCCTTATGTATAAAGTGAGAGCAATAAGACCTACCCTCTTAGGTCATAGAGGGGAAAAATGAAAAAACCAAGTAAGGTCGATAGATCT--------      3438
approx 2.4 kb----------------------------------------------------------------------------------------------------------
TAGCCTCAACCTCTTGGCCTCAAGCCATCCTCCTGTCTTGGCCCCCCCGAGTAGCTGGGACTATAGGCATGTGGCACCACACCTAGCTAATTTGTTTATTTTTTAGAGATGGGGTCTC      3558
TCTATGTTGGCCAGGCTGGTCTTGAACTCATGGGCTCAAGCAATCCTCCTGCCTTGGCCTCCCAAAATGCTGGGATTAC-----------approx. 0.4 kb--------CATGGCA      3644
CTCCAGCCTGGGCAACAGAGCGAGACCCTGACTCAAAAAAAAAAATGTCCCTTTCCTGGGCTGCTGCCTGGTCCTGGCACTGGGCTGTCGGGTCCTCTCCTACTCATTGCCACTCGCCCCA      3764
CTTCTACCCACAGTGGTGACTGGGGGCCCCCCCTGGAAGGGCCCTACCGCCTCAAGCAGTTTCACTTCCACTGGGGCAAGAAGCACGATGTGGGTTCTGAGCACACGGTGGACGGCAAGTCT      3884
                     (3)alValThrGlyGlyProLeuGluGlyProTyrArgLeuLysGlnPheHisPheHisTrpGlyLysLysHisAspValGlySerGluHisThrValAspGlyLysSer
                           80                    90                   100                   110
TTCCCCAGCGAGGTACGGGCCCTCCTCCACTTGAATCCCTCTGCTACATGGGAAGGAACGCAGGCCTGGGTAGTCAGGGGCTTGAGGGATGCCTGGGGTCGGGCCTTGGAGAGGGGGAAGA      4004
PheProSerGlu
GGAGCACCCGGGCCTTCGGGAGAGACTATTTTTCTACCCACCCCCAGCACAGGGTCTTCCATCTGACCCTGTCCACTGCTATCCTTCTTGGTCTGTGCGTCGTGGAGTTTTTCCGAAAGGG      4124
CTACCCCAGTTCCCCACTGCAGCCTTGCGCGCCCTTTAAAGGGTTGAGCCCAGTATTCAAGACTGTGCATGAACTGATCTTGCCTAGCCTCATCCTTCCCCTCACCTCTTCATACACTGA      4244
GCTGCATGTGACGAATCCCCTGTCCACACCTAGCTCCTTCATACCACCAAGCTTTGTCTGTGCTGTTCTCTCTGCTTGGAATGCCTTTCCCCAAGGCTTCCCTCTGATAACCCTCTGCTT      4364
ATTCTCTGGACTCAGCTGTAGGCACCACCTCCTCAAGGAGGTGTCACGCTGGATGTTCATGATCGATTTCACAGCTCGTTCCCTGCAAGGC------------------approx. 0.2      4455
kb---GTTCCCAGCCTCTGGGAGGCACGAGCATTGGCCTGTTGTTGCCCCATAGCTGCATCTGGTTCACTGGAAGTGCCAAGAAGTACAGCACTTTTGGGGAGGCGGCCTCAGCACCTGAT      4570
                                                          (4)LeuHisLeuValHisTrpAsnAlaLysLysTyrSerThrPheGlyGluAlaAlaSerAlaProAspAsp
                                                             120                   130
GGCCTGGCTGTGGTTGGTGTTTTTTTGGAGGTGAGTGGTGGTTTGCTGGGGCCTGGAGGGGCATGGGAGGCCTGCACTGCCCCTAGTTCTAGAACACAGTGGGCTCTGGGCTCACACTGC      4690
GlyLeuAlaValValGlyValPheLeuGlu
140
CCTCTCCCTGACAGACAGGAGACGAGCACCCCAGCATGAATCGTCTGACAGATGCGCTCTACATGGTCCGGTTCAAGGTAAAGTCCCTGCCCCTGACCCAAGCAGCCCGATGGGACAGAGA      4810
                    (5)ThrGlyAspGluHisProSerMetAsnArgLeuThrAspAlaLeuTyrMetValArgPheLys
                       150                   160                   170
GGAAGAGGCTGGCTCACAAGTTGGGCCCTTCCAGTCCTTCATTCTGGGAGTAAACCCTTGCTGAGACCAAGGAGGAGGGTCAGCATCAGCAGGGCCCTCACCTGATTCCTCATCGGGGTC      4930
AGCAGCCCCCTTGGTGTCCAGCATGAAGCCCGGCTAGGTATAAAATTGATAAGATGCTTGTCACATTAATGGTGGTCTAATAGGCAATGTCATTGAGGGCAGCTGAAGGCAAGGGAGGAGGGA      5050
AGATCGTGGCCAAATACCTCCAGGATGCCTTTACATACAGACAAACGGATGAGAGGGCCCGGCTCAGGCTGGAGATGGCATCTGGGATCACAGACATTTGCTCAGAATTTGCAT      5170
CAGTGGACACAGAGAGGCCTAAATCCTGCTGATAGAAATGGGTGGAGAATGCCGGGCATGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCAGGCGGATCGACTTGAGGGT      5290
CAGGAGTTCAAGACCAGCCTGGCTAACATGGTGAAACCCTGTCTCTACTAAACATACAAAAATTAGCCAGGTATGGTGGTGCATGCCTACAATCTAGCTACTTGGGAGGCTGAGAGAGGGA      5410
GAATTGCTTGAACCCCAAGAGGCAGAGGTTGCAGTGAGCCGAGATCGTGCCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCCACCTCAAAAAAAAAGAAAAAGAAAAAAGAAAAGAAA      5530
GAAAGAGAGAGAGGGGGAAGGAAGGAAGGGAAAGGAAGGAAGAGAAAGAAAGAAAGAAAAAGAAGAAAAAAGAAAAAAGAAAAGAAAAGAAAA      5650
AAGAAAAGAAAAAAAGAAAGAAAGAGATGGGTGGAGAGTTGCTGGTTTGGTCCCAGTGGGAGGTCTATGCTGATTCCTGGGTTCCTCCACCTTGTTCCAGGGCCACCAAGCCCAGTTCAG      5770
                                                     (6)GlyThrLysAlaGlnPheSe
CTGCTTCAACCCCAAGTGCCTCCTGCCTGCCAGCCGGCACTACTGGACCTACCCGGGCTCTCTGACGACTCCCCCACTCAGTGAGAGTGTCACCTGATTGTGCTCCGGGAGCCCATCTG      5890
rCysPheAsnProLysCysLeuLeuProAlaSerArgHisTyrTrpThrTyrProGlySerLeuThrThrProProLeuSerGluSerValThrTrpIleValLeuArgGluProIleCy
           180                   190                   200                   210
CATCTCTGAAAGGCAGGTGAGTCCTCTCAGAGGACCAGATGGAGGGACATGGCACTCAGGGCTCACCCCAGGCAGTGTGACTCTCCAACTAGCACCCCACACCCCGTGATCCCACCTTCA      6010
sIleSerGlyArgGln
220
AGGTTCCTCCCCATTTTTTTACTAATAAATGTATAGATTCTTGGTTGGGAGTTAGCTTGATTGTTAATCACCAGAACCTAGCACTGGCTGGGAGCGGGGATACCTGGATCCTGGGACCCC      6130
CACCCCCGCTACTGTGTTGGAATGTCGAATGCTACATnTGCTGATAAGTCTTAGCTAGATGGGCTTGGAACTAGAACTAGGTAAAGCAGAGCCCAGACCCCTTTAGTGAGTAAATAAATGAG      6250
GGTCCTGCAGAGAAAAATGAGTGGGGAAGATAGGGACAGACCCTAAGGAAGGAGGGAGGGGGGTAGAGCTGGCTGGGCAGGTATGCAGATGATGCATCTGGGGTCATAGAGGACCACA      6370
GCACCCCCAATCTGCCCTGAGCATTCCTTCTGCTTCCTCAAGATGGGGAAGTTCCGGAGCCTGCTTTTTACCTCGGAGGACGATGAGAGGATCCACATGGTGAACAACTTCCGGCCACCA      6490
                    (7)MetGlyLysPheArgSerLeuLeuPheThrSerGluAspAspGluArgIleHisMetValAsnAsnPheArgProPro
                                     230                   240
CAGCCACTGAAGGGCCGCGTGGTAAAAGGCCTCCTTCCGGGCGTGAGCTGCCCATCTGCCTAGCCGGCCACTAGGGCACCATCTTCTCAAGGGCTTCCATGTCAGCAGACACCAAACCATC      6610
GlnProLeuLysGlyArgValValLysAlaSerPheArgAla
     250                   260
TGAGGCTTCCTCCCTGGGGGGTGCTGGGGAGACCCTCCTTCAGCCAGTTTGCTCCTTGGTCACCCTGGAGGCTTCTGGATGGGACCCTTGAGTCTGGCGGCAGCCCTTCCAGCTGCCCTGGG      6730
GACAGGAAGGAGCAGGAGCTAAGCAGGGTCCAAGCCTGGGGCTGCCTCTGCTCTCCCAGACCCCAAAGACCCTGGGAACCTCCTCTGGTCTTCCCACCTGGCAGTGCAGCGAGGCCACCGCCA      6850
AGCGCACACTGTGATGGAGGAGACTGAGCTCCCTGGGGCGGGCAGCTGACTACTACCAGAGAGACTCAAGCAATAATTAGAGGTGGGCAGAGCTGCCCTCTCGGCATTACCTCTTCTGCAG      6970
GCTCTGCCATGCACGCACCTCACTGCCAGGCCATTAAAATCAGCACCCAGCATGCTGGAGGTGACGTGCTGCCTTCTCCTCCAGCACTGTGCCAGGCACCCTGGCTATAGCTTATACAGTAT      7090
CTCCCCTTGTCCCCACCCCAGTCACCAAAGCCACCTACATGACAGTCCATCCCTGTTGAATTAATAAAATTAATGTATCCATGCAACAAAATACCAATTGAGATGCTGTGTGGCCTGTCTCTCTAGTAC      7210
CGTCCTAGGCTTTGGGGAGCAATGGTGGTAGTTAAGCACTATGTGTTGGGAGGGGATATGAGTGGTGAGCAGGTGAGGGTTACACAAAGGGCCAGGGAAGGGGCTATGGAGCAnnTAGAGAG      7330
GGGAGACCATGTCACTCAGCAAAGAAATGCAATTCTGGTCAGGCGTGGTGGCTCACGCTTACAGTCCCAGCACTTTGGGAGGCCAAAGCAGGTGGATCACTTGAGCTCAGGAGTTCGAG      7450
AGCAGCCTGGCCAACTGCTGAAAACCCCGTCTCTACTAAAATTACAAAAATTAGCTGGGTGATAGTGTGTGCACCTGTAATCCCAGCTGCTCAGGAGCCTGAGGCAGGAGAATCGCTTGAA      7570
CCCAGGAGGCAGAGGTTGCAGTGAGCTGAGGTCGCACCACTGTACCCAGGCTGGGCGACAGAGCGAGACTCCGTCTCAAACAGAGAAGAAAGGGAAGAAATGAAATTCCAACCAGTATTCCAAA      7690
TCAGATAGAAGGGACAACATGTTTGAGGCCAAGGGAAGGGCAAGTGTCACACGTGGTCAGACTGAGAGAGACCCAGTGTGGTTGTGGGGCTCAAAAGGCTCTAGAGCCTGGGGG      7810
AAACCTGGGACCCCCTCCCCAGGCTGACGGGGAGCTACCAAGGGCTTATAGTAGAGGAGGGGCAGGCCAGGCCTCGAAAATGGGGAACATCCCTGAGGGAAT---------approx. 150      7910
bp-------------------------ATTTATTTATTTTTGAGACGGAGTCTTGCTCTGTGCCCAGGCTGAAGTGAGTGGCATGATCTCAGCTCACTGCAACCTCCGCCTCCCAAGTTCA      8004
AGCAATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGTGCGCACCCACCGTGCCCGGCCAATTTTTGTGTTTTTAGTAGAGATGGGGTTTCACCATGTTGGTCAAGCTGGTCTC      8124
GAATTCCCTGACTCCAAGAATCCACCCACTTGGGCCTCCTAAAAGTGCTGGGATTACAGGTGTGAGCCACCGTGCTTGGCCTAGAGTGCTTTTTTTTAAACGTGCATCCTCACCCTAGTC      8244
CCTAGTCCAAGTCCAAAAAAAGTGCCCAAGGGCAAGTTCACAGTTGGGCTAGAGACCAAGTTGATTGATTGTTGACCCAGATT      8326
```

ment containing exon 3 from the CA VII gene. A single mRNA of approximately 1.6 kb migrates between the 18 S and 16 S markers in the salivary gland RNA (lane 8). When the same blot was stripped and reprobed with a human CA II cDNA insert (Fig. 5B), a similar sized mRNA was detected in the expected tissues: brain, kidney, liver, salivary gland, testes, and liver poly(A)$^+$ RNA (lanes 2, 4, 5, 8, 10, and 16). The absence of a CA II mRNA in spleen (lane 9), lung (lane 6), and thyroid (lane 11) was unexpected but may be accounted for, at least in the latter two cases, by reduced levels of RNA on the membrane, since the background hybridization found in all the other lanes is absent in lanes 6 and 11 of Fig. 5A. The nature of the high-molecular-weight RNA hybridizing to the CA VII probe at the top of the HT1080 fibrosarcoma cell line lane (lane 13) is not known.

Chromosomal Location of the CA VII Gene

The 3.0-kb 5' EcoRI junction fragment from clone H11 was used as a hybridization probe to detect the presence of the human CA VII gene in human–mouse somatic cell hybrids. Figure 6 shows a representative Southern blot of XbaI-digested DNA from several hybrid cell lines. The probe detects a 12.0-kb fragment in human genomic DNA and an 8.6-kb fragment in mouse genomic DNA. Cross-hybridization of the mouse fragment was not due to the presence of exon 2 in the probe, but rather to a region that was found to reside in a 0.6-kb BglII/BamHI subfragment within the first intron (underlined sequence between nucleotides 1548 and 2109 in Fig. 3). In addition to human and mouse, this subfragment detects a single band in genomic DNA from macaque and rabbit (data not shown). The nature of this conserved region and its relationship to the CA VII gene are unclear. Computer analysis failed to show significant sequence similarity between the translated sequence of this fragment and the other carbonic anhydrase isozymes and searches of the GenBank and EMBL databases revealed no identity with known genes. The presence of a pair of consensus acceptor and donor splice junction

sequences, which would suggest an additional conserved exon in this region, was also not detected.

DNA from the somatic cell hybrids either had the mouse fragment only, and therefore lacked the human chromosome containing the human CA VII gene (Fig. 6, lanes 1–3), or had both human and mouse bands and therefore carried the human chromosome containing the CA VII gene (Fig. 6, lanes 4–6). The segregation analysis of the CA VII marker and various human chromosomes found in 40 somatic cell hybrids is shown in Table 2 and establishes the location of the CA VII gene on human chromosome 16. For each of the other human chromosomes, at least 21% of the hybrids were discordant.

In situ hybridization of the same probe to metaphase spreads of human lymphocyte chromosomes was performed to confirm the above result and to regionally assign the position of the CA VII gene. Based on the distribution of 331 grains on 200 metaphases examined, the CA VII gene was localized to the long arm of chromosomes 16 at the q21–23 region (Fig. 7B). Figure 7A shows that only this region has a signal above the background. The percentage of metaphases with a grain at 16q21–23 was 14%, while 14.2% of the total grains were on chromosome 16, with 68.1% of these grains located at 16q21–23 (Fig. 7A).

DISCUSSION

The overall structural organization of the human CA VII gene, consisting of seven exons with six introns interrupting putative coding sequences at the same nucleotides that mark exon boundaries in other CA genes, is similar to that seen for mammalian CA II and CA III genes. The identification of a 1.6-kb transcript is consistent with the predicted gene size of 10 kb, which is based on two assumptions. The first is the inferred position of the 3' terminus downstream of the poly(A)$^+$ addition signal, which defines a rather large 3' untranslated region of greater than 600 nucleotides, similar to sizes seen in CA II genes. The second is the likely location of the transcriptional initiation site in the vicinity of the TTTAA pentamer,

---

FIG. 3. Nucleotide sequence of the CA VII gene. The sequence begins just downstream of the HindIII site with the 5.9-kb EcoRI fragment of clones 2-4 and 9-1B (c.f. Fig. 1). The nucleotides are numbered from the first base sequenced. The numbering does not take into account gaps in the sequence in intron 1, intron 2 (two gaps), and the 3' flanking region of the gene. Each exon is designated by encircled numbers at the beginning of the exon and the translated protein sequence appears below the nucleotide sequence. Amino acids are numbered below the protein sequence according to the CA I numbering system (cf. Ref. (25)). The putative ATG start and TGA stop codons are boxed at nucleotides 511 and 6533, respectively. A poly(A)$^+$ addition signal is also boxed at nucleotide 7152. A GT-rich segment, similar to those found to be necessary for 3' end formation of other mRNAs, is underlined 28 bp downstream of the poly(A)$^+$ signal. A conserved region in intron 1 is underlined and described in the text. Potential lariat branch points are doubly underlined upstream of the beginning of exons 2, 3, 5, 6, and 7. Located between nucleotides 2151 and 2178 is a (GT)$_{14}$ repeat element that is underlined with dashes. Alu repeats are boxed in intron 2 (two repeats), intron 5, and the 3' flanking region. The repeat in intron 5 is flanked by 12-bp direct repeats (dashed boxes). Sequence data for nucleotides 783–786 could not be determined.

TABLE 1

Nucleotide Sequences at Exon/Intron Splice Junctions[a]

Exon	Donor	Intron	Acceptor
1 (117 bp)	AcG/GTAgGc	1 (2.5 kb)	CCTCCCCaCCTGCAG/G
2 (198 bp)	CcG/GTAAGT	2 (3.4 kb)	CaCTTCTaCCCACAG/t
3 (118 bp)	gAC/GTAcGg	3 (0.8 kb)	gTTgTTgCCCCATAG/c
4 (92 bp)	gAG/GTGAGT	4 (0.1 kb)	CCCTCTCCCTgACAG/a
5 (61 bp)	AAG/GTAAag	5 (1.0 kb)	TCCaCCTTgTTCCAG/G
6 (155 bp)	CAG/GTGAGT	6 (0.5 kb)	TTCTgCTTCCTCaAG/a
	C G		CCCCCCCCCCC C
Consensus[b]	AG/GT AGT		N AG/G
	A A		TTTTTTTTTTT T

[a] Capital letters indicate nucleotide matches with the consensus sequence. Lowercase letters indicate mismatches. Invariant GT and AG dinucleotides at the beginning and end of each intron are underlined.
[b] Consensus sequences are from Ref. (44).

located 102 bp upstream of the putative start codon in exon 1. This is the only sequence motif in a very GC-rich region that resembles a TATA promoter. If the cap site is located 20–30 bp downstream, an untranslated region of about 70–80 nucleotides, consistent with the sizes of the 5' UTRs of CA II and CA III genes, would precede the start codon. Although the methylation status of the CA VII gene has not been assessed, a high concentration of CpG dinucleotides in the region of the first exon suggests the presence of an HTF island (Bird, 1987). Consistent with this is the presence of four GC-box hexamers which are potential binding sites for the general transcription factor Sp1, an activator of transcription for a number of viral and cellular promoters (Kadonaga et al., 1986; Jones et al., 1986; Briggs et al., 1986). While the relevance of the Sp1 factor to regulation of CA VII gene expression remains to be established, the presence of these potential binding sites may be further evidence that this region contains the 5' end of the gene.

On one hand, the predicted sequence of the putative CA VII protein shows a relatively high degree of divergence relative to that of the other human CA isozymes, but on the other hand it demonstrates the existence of a selective pressure for conservation of regions expected to be critical for proper enzyme activity. The similar pattern of conservation and divergence in CA II correlates well with the idea that the tertiary structure of CA VII will be similar to that of the human CA II molecule. Figure 4 shows that the region encoded by the 3' half of exon 6, amino acids 191–222, is relatively conserved among the human CA isozymes, whereas the part of the protein encoded by the 5' half of exon 6, amino acids 171–190, is much less conserved. Examination of the three-dimensional structures of human CA I, CA II, and bovine CA III

(Eriksson, 1988) reveals that the former region is buried in the center of the molecule and forms part of the active site cavity, where substitutions would tend to be tolerated less in terms of maintenance of enzyme function. The latter region, however, is found at the external surface of the molecule in an area that can probably diverge much more freely without affecting function. Similar observations hold true for other areas of divergence. Poorly conserved regions of CA VII in exon 2 (residues 32–60), exon 3 (residues 99–103 and 112–116), exon 5 (residues 149–169), and exon 7 (residues 232–248) are all predicted to be located toward the surface of the CA VII protein and away from the active site based on their locations in the other CAs. Interestingly, all key active site residues, including the zinc-ion liganding histidines (His-94, His-96, and His-119) and the residues (Ser-29, Gln-92, Glu-106, His-107, Glu-117, Tyr-194, Thr-119, Trp-209, and Asn-244) thought to be involved in a hydrogen-bonding network with the zinc-bound solvent molecule (cf. Lindskog, 1982; Silverman and Vincent, 1983), are strictly conserved in CA VII, as they are in all CA isozymes. It would thus appear that conservation within regions of the CA VII protein is nonrandom, such that residues that are critical for CA enzyme activity are maintained. This is a strong indication that the CA VII gene encodes a functional protein.

Because it was immediately obvious that the CA VII gene could not encode the well-characterized cytosolic CA I, CA II, or CA III proteins, it was initially thought that its product could be either one of the less well described CA isozymes or a completely novel form of the enzyme. The latter possibility appears now to be most likely. The recently published cDNA sequence (Aldred et al., 1991) of the human secreted
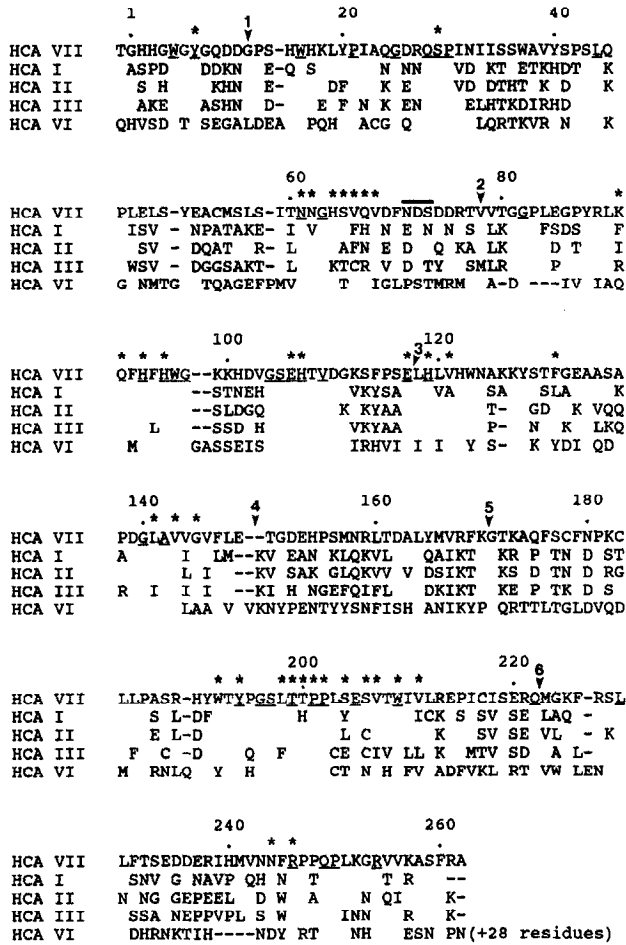
```
             1        *   1      20              40
             .        *   ▼       .         *    .
HCA VII  TGHHGWGYGQDDGPS-HWHKLYPIAQGDRQSPINIISSWAVYSPSLQ
HCA I    ASPD   DDKN  E-Q S      N NN   VD KT ETKHDT  K
HCA II   S H    KHN   E-   DF  K E      VD DTHT K D    K
HCA III  AKE    ASHN  D-   E F N K EN    ELHTKDIRHD
HCA VI   QHVSD T SEGALDEA  PQH  ACG Q     LQRTKVR N    K

                 60               2  80
              . ** *****           ▼  .              *
HCA VII  PLELS-YEACMSLS-ITNNGHSVQVDFNDSDDRTVVTGGPLEGPYRLK
HCA I    ISV - NPATAKE- I V   FH N E N N S LK   FSDS   F
HCA II   SV - DQAT R- L    AFN E D  Q KA LK    D T    I
HCA III  WSV - DGGSAKT- L   KTCR V D TY  SMLR     P     R
HCA VI   G NMTG  TQAGEFPMV    T  IGLPSTMRM  A-D ---IV IAQ

            *  *  *    100         **        120
            * * *      .         **     *3*.*         *
HCA VII  QFHFHWG--KKHDVGSEHTVDGKSFPSELHLVHWNAKKYSTFGEAASA
HCA I      --STNEH      VKYSA   VA    SA   SLA      K
HCA II     --SLDGQ    K KYAA         T-  GD K VQQ
HCA III  L   --SSD H      VKYAA       P-  N K LKQ
HCA VI   M   GASSEIS     IRHVI I I  Y S-  K YDI QD

            140       4       160           5    180
            .* * *    ▼       .            ▼    .
HCA VII  PDGLAVVGVFLE--TGDEHPSMNRLTDALYMVRFKGTKAQFSCFNPKC
HCA I    A     I  LM--KV EAN KLQKVL   QAIKT  KR P TN D ST
HCA II        L I  --KV SAK GLQKVV V DSIKT  KS D TN D RG
HCA III  R I   I I --KI H NGEFQIFL   DKIKT  KE P TK D S
HCA VI   LAA  V VKNYPENTYYSNFISH ANIKYP  QRTTLTGLDVQD

                      200             220 6
              * *    ***** *  ** * *    .▼
HCA VII  LLPASR-HYWTYPGSLTTPPLSESVTWIVLREPICISERQMGKF-RSL
HCA I      S L-DF     H    Y    ICK S SV SE LAQ  -
HCA II     E L-D           L C    K  SV SE VL  - K
HCA III  F  C -D    Q  F    CE CIV LL K  MTV SD   A L-
HCA VI   M  RNLQ  Y  H       CT N H FV ADFVKL RT VW LEN

                      240             260
                        * *
HCA VII  LFTSEDDERIHMVNNFRPPQPLKGRVVKASFRA
HCA I    SNV G NAVP QH N  T      T R   --
HCA II   N NG GEPEEL  D W A    N QI    K-
HCA III  SSA NEPPVPL S W    INN    R   K-
HCA VI   DHRNKTIH----NDY RT    NH    ESN PN(+28 residues)
```

FIG. 4. Comparison of the derived CA VII protein sequence with human CA I, CA II, CA III, and CA VI (cf. Refs. (1, 26)). Residues are shown for CA I, II, III, and VI that differ from CA VII in the corresponding position. Amino acids are numbered according to the CA I numbering system (cf. Ref. (26)). Gaps are introduced into the sequences to maintain the alignment of the residues and are indicated by dashes. Positions corresponding to active site residues in human CA I and CA II and bovine CA III are indicated by asterisks (cf. Ref. (65)), and positions that are conserved in all mammalian CA isozymes sequenced to date are underlined in the CA VII sequence (Refs. (2, 26, 33)). A potential N-linked glycosylation site at residues 71–73 is overlined. Intron positions for CA VII are indicated by numbered arrowheads. The single-letter amino acid code is used. Human CA VI has a leader sequence of 18 residues that is not shown (Ref. (1)).

salivary isozyme, CA VI, has eliminated this isozyme as a candidate for the CA VII gene product. Similarly, a comparison of the CA VII sequence with the partial protein sequence for the human CA IV membrane-bound CA isozyme (Zhu and Sly, 1990) shows only 33% identity (Tashian, 1989). No sequence data exist for the human CA V mitochondrial isozyme, but a comparison with the partial guinea pig CA V sequence (Hewett-Emmett et al., 1986) shows that the N-ter-
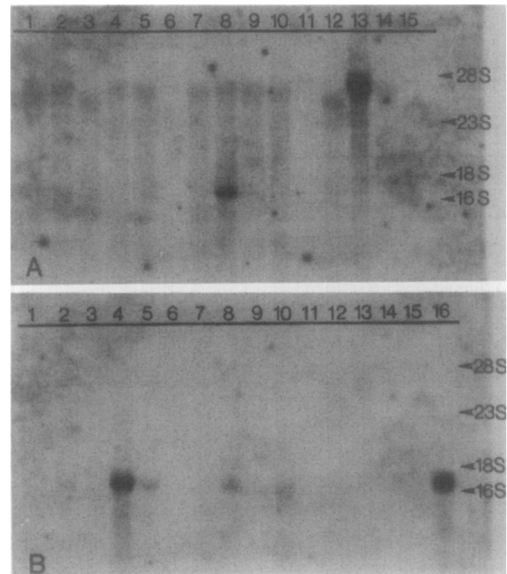
FIG. 5. Northern blot analysis of baboon and human RNAs with a CA VII probe in A containing exon 3 plus 43 bp of intron 2 and 116 bp of intron 3. The probe in B was the insert from a human CA II cDNA clone containing the entire cDNA except for the last 172 nucleotides of the 3′ untranslated region (Ref. (43)). After hybridization, the blot was washed several times under low-stringency conditions, with a final high-stringency wash of 0.1× SSC, 0.1% SDS at 55°C. The RNAs in each lane are: 1, adrenal; 2, brain; 3, heart; 4, kidney; 5, liver; 6, lung; 7, pancreas; 8, salivary gland; 9, spleen; 10, testis; 11, thyroid; 12, thymus; 13, human fibrosarcoma HT 1080 cell line; 14, rat hepatoma HPC PAI cell line; 15, human hepatoma HEP-G2 tube (10 mg); 16, human liver poly(A)⁺ RNA (2 μg).

minus of CA V aligns with amino acid 22 of other CA isozymes, implying that cleavage of an upstream mitochondrial import signal has occurred. It is unlikely
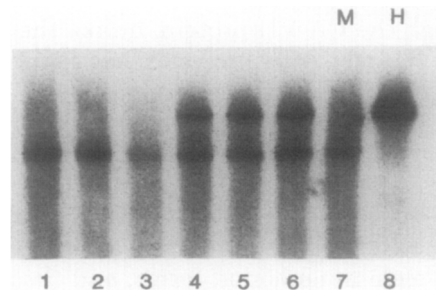
FIG. 6. Representative Southern blot of human–mouse somatic cell hybrid DNA. The 3.0-kb EcoRI fragment from clone H11 hybridized to a single unique fragment in XbaI-digested DNA from mouse (8.6 kb, lane 7) and human (12.0 kb, lane 8). Cell hybrids missing the human specific CA VII fragment are in lanes 1–3 and were scored negative for the presence of the human chromosome containing the CA VII gene. Hybrids in lanes 4–6 contain the human fragment and were scored as positives.

MONTGOMERY ET AL.

TABLE 2

Segregation of the Carbonic Anhydrase VII Gene with Human Chromosomes in Human–Mouse Hybrids

	Chromosome																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
No. of concordant hybrids[a]																							
+/+	7	7	11	10	10	7	10	9	5	11	8	12	8	11	9	13	11	10	6	11	11	8	6
−/−	24	18	13	17	15	19	16	11	25	13	18	14	15	12	19	27	6	15	22	12	5	19	9
No. of discordant hybrids[b]																							
+/−	5	6	2	3	3	6	3	4	6	2	4	1	5	2	3	0	2	3	7	2	2	5	3
−/+	3	9	11	10	12	8	11	16	2	14	9	13	12	15	8	0	20	12	5	15	22	8	16
% Discordancy	21	38	35	33	38	35	35	50	21	40	33	35	43	43	28	0[c]	56	38	30	43	60	32	56

[a] Concordant hybrids have either retained (+/+) or lost (−/−) the human CA VII signal together with a specific human chromosome.

[b] Discordant hybrids either retained the gene, but not a specific chromosome (+/−), or the reverse (−/+).

[c] A 0% discordancy is the basis for chromosomal assignment.

that the N-terminus of the CA VII protein acts as such as signal because the first 22 amino acids of the CA VII protein align well with those of other CAs, and because the absence of positive charges in the first 18 residues and the presence of negatively charged residues at positions 10 and 11 are features not generally found in known import signals (Douglas et al., 1986). In addition, the recently reported cDNA sequences of two distinct CA or CA-like proteins from mouse liver (Amor-Gueret and Levi-Strauss, 1991) and cerebellum (Kato, 1991) showed only low sequence identities (i.e., 52 and 37%, respectively) to the human CA VII sequence. This degree of conservation is in contrast to the much higher percentage identities (70–85%) seen when the orthologous gene products for other carbonic anhydrases, such as CA I, CA II, and CA III, are compared between these two species (Hewett-Emmett and Tashian, 1991) and suggests, therefore, that CA VII is not the human counterpart to either of these two mouse proteins.

Although the CA VII protein, unlike the cytosolic CA I, CA II, and CA III isozymes, has a single potential N-linked glycosylation site at codons 71–73, Asn–Asp–Ser, the above analysis suggests that it may be nonfunctional. The similar sizes of the predicted CA VII protein and the cytosolic enzymes, the higher degree of sequence conservation between it and these isozymes compared to that between it and the secreted or membrane-bound forms, and the apparent lack of a signal peptide all point to a probable cytosolic localization for CA VII, in which case it would not be glycosylated.

The detection of a band of the predicted size for a CA VII transcript (1.6 kb) in salivary gland RNA provides clear evidence that this locus is expressed, thus confirming predictions that the gene encodes a func-

tional molecule. These predictions were based on findings of a high degree of widespread divergence in CA VII, except for a remarkably strict conservation of residues necessary for enzyme function. This high degree of divergence, which prevents other CA gene sequences from readily cross-hybridizing to CA VII sequences and vice versa (see Fig. 2), coupled with the detection of a band in RNA from only a single tissue, eliminates the probability that the CA VII probe is detecting a transcript from one of the other isozymes. Carbonic anhydrases I, II, and III have been detected in subcellular locations within various cell types of human, rat, and mouse salivary glands (Hennigar et al., 1983; Ikejima and Ito, 1984; Carpentier et al., 1985; Noda et al., 1986; Parkkila et al., 1990; Spicer et al., 1990; Aldred et al., 1991; Fernley et al., 1991). However, since activity assays only detect CA activity and immunological methods can be subject to cross-reactivity, the presence of novel CA isozymes can be missed. For example, it is now clear that CA activity found in secretory granules of these glands is actually the CA VI isozyme. The additional presence of the CA VII isozyme is thus possible. Further work will be necessary to confirm this and to elucidate the roles of possibly five forms of carbonic anhydrase (CA I, CA II, CA III, CA VI, and CA VII) in salivary glands.

The finding that the CA VII gene is located on chromosome 16 and a recent report localizing the CA VI gene to human chromosome 1 (Sutherland et al., 1989) are the first examples of CA isozyme genes not linked to the CA1/CA3/CA2 gene cluster on human chromosome 8 (Davis et al., 1987; Nakai et al., 1987). Phylogenetic analyses indicate that the duplications giving rise to the CA I, CA II, and CA III genes occurred after the duplication that formed the CA VII gene (Hewett-Emmett and Tashian, 1991), suggest-
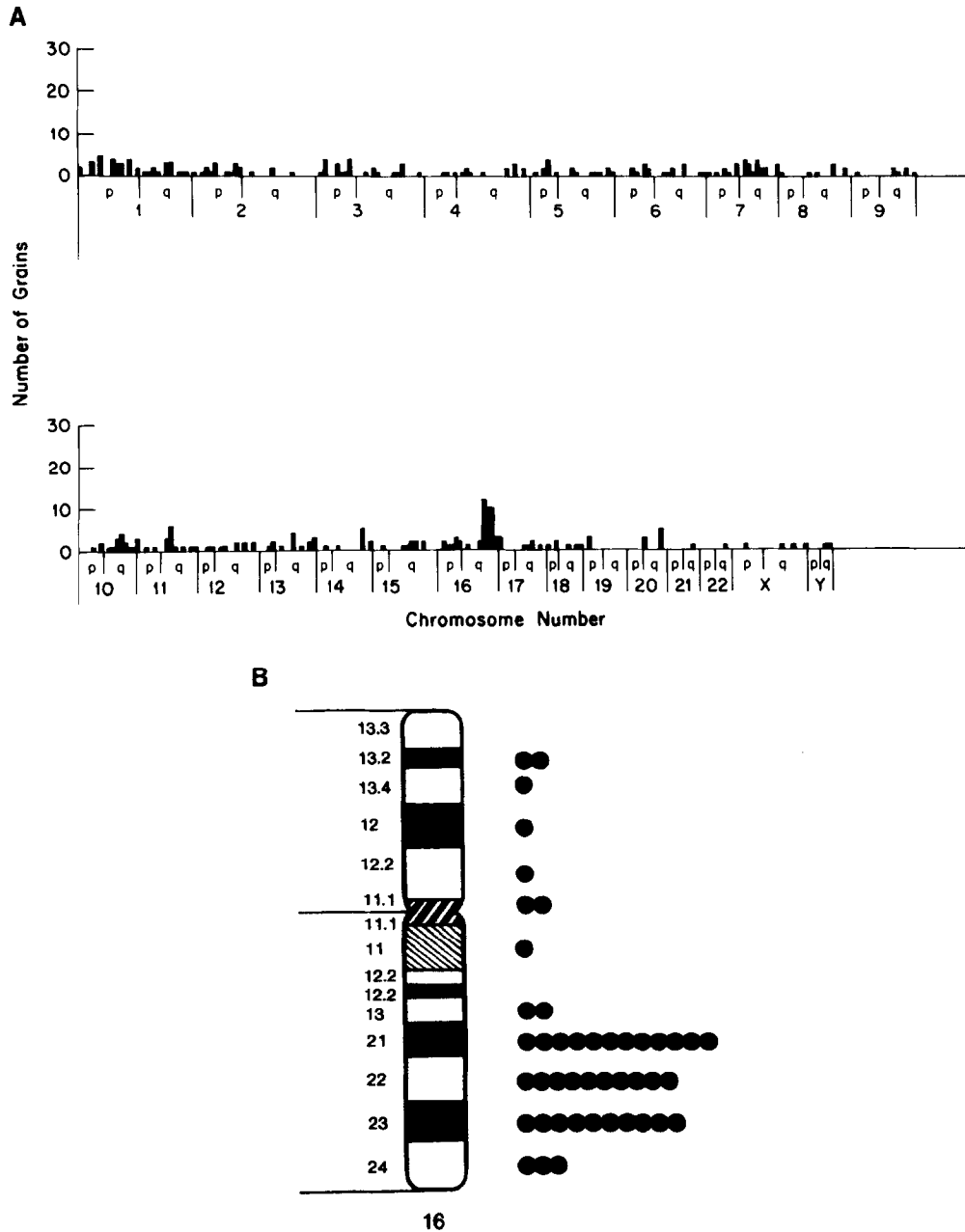
FIG. 7. Mapping of the CA VII gene to human chromosome 16 by *in situ* hybridization. (A) The total grain distribution after hybridization of the 3.0-kb *Eco*RI fragment from clone H11 to normal human lymphocyte metaphase chromosomes. Accumulation above background occurs only on the long arm of chromosome 16. (B) An idiogram localizing the majority of grains to 16q21–23.

ing that the latter gene may have dispersed prior to the formation of the former genes. The location of the CA VII gene on the long arm of chromosome 16 may be useful in further defining the linkage map of that portion of the genome (Keith *et al.,* 1987). The polymorphic nature of poly(GT) repeats (Weber and May, 1989) and the 3' ends of *Alu* repeat elements (Hobbs *et al.,* 1985) may facilitate the identification of informative CA VII probes that can be used for this purpose. The locations of these repeated elements may also be useful in identifying the presence of the CA

VII gene region in previously cloned segments of chromosome 16 (Hildebrand *et al.,* 1988).

The haptoglobin, adenine phosphoribosyltransferase, mitochondrial glutamate oxaloacetate transaminase, metallothionein I and II, lecithin–cholesterol acyltransferase, uvomorulin, and tyrosine aminotransferase genes, all located on the long arm of human chromosome 16, form part of a syntenic group that maps to mouse chromosome 8 (Nadeau, 1989; Lalley *et al.,* 1989). Because of its location among these genes on human chromosome 16, it is expected

that the CA VII gene will also be found on chromosome 8 of the mouse.

ACKNOWLEDGMENTS

REFERENCES

1. ALDRED, P., FU, P., BARRETT, G., PENSCHOW, J. D., WRIGHT, D., COGHLAN, J. P., AND FERNLEY, R. T. (1991). Human secreted carbonic anhydrase: cDNA cloning, nucleotide sequence, and hybridization histochemistry. *Biochemistry* **30:** 569–575.

2. AMOR-GUERET, M., AND LEVI-STRAUSS, M. (1990). Nucleotide and derived amino acid sequence of a cDNA encoding a new mouse carbonic anhydrase. *Nucleic Acids Res.* **18:** 1646.

3. BEECHEY, C., TWEEDIE, S., SPURR, N., PETERS, J., AND EDWARDS, Y. (1990). Mapping of mouse carbonic anhydrase-3, *Car-3:* A further locus in the homologous region of mouse chromosome 3 and human chromosome 8. *Genomics* **6:** 692–696.

4. BIGGIN, M. D., GIBSON, T. J., AND HONG, G. F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* **80:** 3963–3965.

5. BIRD, A. P. (1987). CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* **3:** 342–347.

6. BIRNSTEIL, M. L., BUSSLINGER, M., AND STRUB, K. (1985). Transcription termination and 3′ processing: The end is in site! *Cell* **41:** 349–383.

7. BREATHNACH, R., AND CHAMBON, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50:** 349–359.

8. BRIGGS, M. R., KADONAGA, J. T., BELL, S. P., AND TJIAN, R. (1986). Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. *Science* **234:** 47–52.

9. BUTTERWORTH, P. H. W., BARLOW, J. H., BARDY, H. J. M., EDWARDS, M., LOWE, N., AND SOWDEN, J. C. (1991). The structure and regulation of the human carbonic anhydrase I gene. *In* "The Carbonic Anhydrases: Cellular Physiology and Molecular Genetics" (S. J. Dodgson, R. E. Tashian, G. Gros, and N. D. Carter, Eds.), pp. 197–207, Plenum Press, New York.

10. CARPENTIER, P., CHETAIL, M., AND FOURNIE, J. (1985). Carbonic anhydrase in the rat salivary glands: Distribution and ultrastructural localization. *Biol. Cell* **54:** 241–249.

11. CHEN, E. Y., AND SEEBURG, P. H. (1985). Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. *DNA* **4:** 165–170.

12. CHOMCZYNSKI, P., AND QASBA, P. K. (1984). Alkaline transfer of DNA to plastic membrane. *Biochem. Biophys. Res. Commun.* **122:** 340–344.

13. CURTIS, P. J. (1983). Cloning of mouse carbonic anhydrase mRNA and its induction in mouse erythroleukemic cells. *J. Biol. Chem.* **258:** 4459–4463.

14. DAVIS, M. B., WEST, L. F., BARLOW, J. H., BUTTERWORTH, P. H. W., LLOYD, J. C., AND EDWARDS, Y. H. (1987). Regional localization of carbonic anhydrase genes CA1 and CA3 on human chromosome 8. *Somat. Cell Mol. Genet.* **13:** 173–178.

15. DOUGLAS, M. G., MCCAMMON, M. T. M., AND VASSAROTTI, A. (1986). Targeting proteins into mitochondria. *Microbiol. Rev.* **50:** 166–178.

16. DUTRILLAUX, B., AND VIEGAS-PEQUIGNOT, E. (1981). High resolution R- and G-banding on the same preparation. *Hum. Genet.* **57:** 93–95.

17. ERIKKSON, A. E. (1988). "Structural Differences between High and Low Activity Forms of Carbonic Anhydrases," Ph.D. thesis, Uppsala University, Sweden.

18. FERNLEY, R. T., WRIGHT, R. D., AND COGHLAN, J. P. (1991). Radioimmunoassay of carbonic anhydrase VI in saliva and sheep tissues. *Biochem. J.* **274:** 313–316.

19. FRASER, P., CUMMINGS, P., AND CURTIS, P. (1989). The mouse carbonic anhydrase I gene contains two tissue-specific promoters. *Mol. Cell. Biol.* **9:** 3308–3313.

20. GAROFF, H., AND ANSORGE, W. (1981). Improvements in DNA sequencing gels. *Anal. Biochem.* **115:** 450–457.

21. HANNA, Z., FREGEAU, C., PREFONTAINE, G., AND BROUSSEAU, R. (1984). Construction of a family of universal expression plasmid vectors. *Gene* **30:** 247–250.

22. HATTORI, M., AND SAKAKI, Y. (1986). Dideoxy sequencing method using denatured plasmid templates. *Anal. Biochem.* **152:** 232–238.

23. HENIKOFF, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28:** 351–359.

24. HENNIGAR, R. A., SCHULTE, B. A., AND SPICER, S. S. (1983). Immunolocalization of carbonic anhydrase in rat and mouse salivary and exorbital lacrimal glands. *Anat. Rec.* **207:** 605–614.

25. HEWETT-EMMETT, D., COOK, R. G., AND DODGSON, S. J. (1986). Novel gene encodes mitochondrial carbonic anhydrase (CA V) that lacks the active-site tyrosine of the aminoterminal region of CA I, II, and III. *Fed. Proc.* **45:** 1661.

26. HEWETT-EMMETT, D., AND TASHIAN, R. E. (1991). Structure and evolutionary origins of the carbonic anhydrase multigene family. *In* "The Carbonic Anhydrases: Cellular Physiology and Molecular Genetics" (S. J. Dodgson, R. E. Tashian, G. Gros, and N. D. Carter, Eds.), pp. 15–32, Plenum Press, New York.

27. HILDEBRAND, C. E., CLARK, L. H., MEYNE, J., BUCKINGHAM, J. M., JONES, M. D., BROWN, N. C., LOGMIRE, J. L., DEAVEN, L. O., CALLEN, D., STALLINGS, R. L., AND MOYZIS, R. K. (1988). Accelerated physical mapping of the human genome: Macro-restriction analysis of centromeric and telomeric regions. *Am. J. Hum. Genet.* **43:** A147.

28. HOBBS, H. H., LEHRMAN, M. A., YAMMAMOTO, T., AND RUSSELL, D. W. (1985). Polymorphism and evolution of Alu sequences in the human low density lipoprotein receptor gene. *Proc. Natl. Acad. Sci. USA* **82:** 7651–7655.

29. HODGSON, C. P., AND FISK, R. Z. (1987). Hybridization probe size control: optimized oligolabeling. *Nucleic Acids Res.* **15:** 6295.

30. IKEJIMA, T., AND ITO, S. (1984). Carbonic anhydrase in mouse

salivary glands and saliva: A histochemical, immunohisto-chemical, and enzyme activity study. *J. Histochem. Cytochem.* **32**: 625–635.

31. JONES, K. A., KADONAGA, J. T., LUCIW, P. A., AND TJIAN, R. (1986). Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. *Science* **232**: 755–759.

32. KADONAGA, J. T., JONES, K. A., AND TJIAN, R. (1986). Promoter-specific activation of polymerase II transcription by SP1. *Trends Biochem. Sci.* **11**: 20–33.

33. KATO, K. (1990). Sequence of a novel carbonic anhydrase-related polypeptide and its exclusive presence in Purkinje cells. *FEBS Lett.* **271**: 137–140.

34. KEITH, T., GREEN, P., REEDERS, S. T., BROWN, V. A., PHIPPS, P., BRICKER, A., KNOWLTON, R., NELSON, C., AND DONIS-KELLER, H. (1987). A linkage map of chromosome 16 with 41 RFLP markers. *Cytogenet. Cell Genet.* **46**: 638.

35. KELLER, E. B., AND NOON, W. A. (1984). Intron Splicing: A conserved internal signal in introns of animal pre-mRNAs. *Proc. Natl. Acad. Sci. USA* **81**: 7417–7420.

36. LAHERTY, C. D., GIERMAN, T. M., AND DIXIT, V. M. (1989). Characterization of the promoter region of the human thrombospondin gene. *J. Biol. Chem.* **264**: 11222–11227.

37. LALLEY, P. A., DAVISSON, M. T., GRAVES, J. A. M., O'BRIEN, S. J., WOMACK, J. E., RODERICK, T. H., CREAU-GOLDBERG, N., HILLYARD, A. L., DOOLITTLE, D. P., AND ROGERS, J. A. (1989). Report of the committee on comparative mapping. *Cytogenet. Cell Genet.* **51**: 503–532.

38. LINDSAY, S., AND BIRD, A. P. (1987). Use of restriction enzymes to detect potential gene sequences in mammalian DNA. *Nature* **327**: 336–338.

39. LINDSKOG, S. (1982). Carbonic anhydrase. *Adv. Inorg. Biochem.* **4**: 115–170.

40. LOWE, N., BRADY, H. J. M., BARLOW, J. H., SOWDEN, J. O., EDWARDS, M., AND BUTTERWORTH, P. H. W. (1990). Structure and methylation patterns of the gene encoding human carbonic anhydrase I. *Gene* **93**: 277–283.

41. MANIATIS, T., FRITSCH, E. F., AND SAMBROOK, J. (1982). "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

41a. MAXAM, A. M., AND GILBERT, W. (1980). Sequencing end-labeled DNA with base specific chemical cleavages. *In* "Methods in Enzymology" (L. Grossman and K. Moldave, Eds.), Vol. 65, pp. 499–560, Academic Press, New York.

42. MCLAUCHLAN, J., GAFFNEY, D., WHITTON, J. L., AND CLEMENTS, J. B. (1985). The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of 3' termini. *Nucleic Acids Res.* **13**: 1347–1368.

43. MONTGOMERY, J. C., VENTA, P. J., TASHIAN, R. E., AND HEWETT-EMMETT, D. (1987). Nucleotide sequence of human liver carbonic anhydrase II cDNA. *Nucleic Acids Res.* **15**: 4687.

44. MOUNT, S. M. (1982). A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**: 459–472.

45. NADEAU, J. H. (1989). Maps of linkage and synteny homologies between mouse and man. *Trends Genet.* **5**: 82–86.

46. NAKAI, H., BYERS, M. G., VENTA, P. J., TASHIAN, R. E., AND SHOWS, T. B. (1987). The gene for human carbonic anhydrase II (*CA2*) is located at chromosome 8q22. *Cytogenet. Cell Genet.* **44**: 234–235.

47. NODA, Y., SUMITOMO, S., HIKOSAKA, N., AND MORI, M.

(1986). Immunohistochemical localization of carbonic anhydrase I and II in the submandibular salivary glands of the mouse, rat, hamster, and guinea pig. *Arch. Oral Biol.* **31**: 795–800.

48. NORRANDER, J., KEMPE, T., AND MESSING, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide directed mutagenesis. *Gene* **26**: 101–106.

49. PARKKILA, S., KAUNISTO, K., RAJANIEMI, L., KUMPULAINEN, T., JOKINEN, K., AND RAJANIEMI, H. (1990). Immunohistochemical localization of carbonic anhydrase isoenzymes VI, II and I in human parotid and submandibular glands. *J. Histochem. Cytochem.* **38**: 941–947.

50. PONTZ, M., SOLOWIEJCZYK, D., BALLENTINE, M., SCHWARTZ, E., AND SURREY, S. (1982). "Nonrandom" DNA sequence analysis in bacteriophage M13 by the dideoxy chain-terminal method. *Proc. Natl. Acad. Sci. USA* **79**: 4298–4302.

51. PUSTELL, J., AND KAFATOS, F. C. (1982a). A high speed, high capacity homology matrix: zooming through SV40 and polyoma. *Nucleic Acids Res.* **10**: 4765–4782.

52. PUSTELL, J., AND KAFATOS, F. C. (1982b). A convenient and adaptable package of DNA sequence analysis programs for microcomputers. *Nucleic Acids Res.* **10**: 51–59.

53. PUSTELL, J., AND KAFATOS, F. C. (1984). A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis and homology determination. *Nucleic Acids Res.* **12**: 643–655.

54. RUSKIN, B., KRAINER, A. R., MANIATIS, T., AND GREEN, M. R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing *in vitro*. *Cell* **38**: 317–331.

55. SAKAGUCHI, A. Y., AND SHOWS, T. B. (1982). Cornavirus 229E susceptibility in man–mouse hybrids is located on human chromosome 15. *Somat. Cell Genet.* **8**: 83–94.

56. SHOWS, T., EDDY, R., HALEY, L., BYERS, M. G., HENRY, M., FUJITA, T., MATSU, H., AND TANIGUCHI, T. (1984). Interleukin 2 (IL2) is assigned to human chromosome 4. *Somat. Cell Mol. Genet.* **10**: 315–318.

57. SHOWS, T. B. (1983). *In* "Isozymes: Current Topics in Biological and Medical Research" (M. C. Rattazzi, J. G. Scandalios, and G. S. Whitt, Eds.), Vol. 10, pp. 323–339, A. R. Liss, New York.

58. SHOWS, T. B., BROWN, J. A., HALEY, L. L., BYERS, M. G., EDDY, R. L., COOPER, E. S., AND COGGIN, A. P. (1978). Assignment of the β-glucuronidase structural gene to the pter-q22 region of chromosome 7 in man. *Cytogenet. Cell Genet.* **21**: 99–104.

59. SHOWS, T. B., SAKAGUCHI, A. Y., AND NAYLOR, S. L. (1982). Mapping the human genome, cloned genes, DNA polymorphisms, and inherited disease. *In* "Advances in Human Genetics" (H. Harris and K. Hirschorn, Eds.), Vol. 12, pp. 341–452, Plenum Press, New York/London.

60. SILVERMAN, D. A., AND VINCENT, S. H. (1983). Proton transfer in the catalytic mechanism of carbonic anhydrase. *CRC Crit. Rev. Biochem.* **14**: 207–255.

61. SLY, W. S., WHYTE, M. P., SUNDARAM, V., TASHIAN, R. E., HEWETT-EMMETT, D., GUIBAUD, P., VAINSEL, M., BALVARTE, H. J., GRUSKIN, A., AL-MOSAWI, M., SAKATI, N., AND OHLSSON, A. (1985). Carbonic anhydrase II deficiency in 12 families with the autosomal recessive syndrome of osteopetrosis with renal tubular acidosis and cerebral calcification. *New Engl. J. Med.* **313**: 139–145.

62. SPICER, S. S., GE, Z.-H., TASHIAN, R. E., HAZEN-MARTEN, D. J., AND SCHULTE, B. A. (1990). Comparative distribution

of carbonic anhydrase isozyme III and II in rodent tissues. *Am. J. Anat.* **187**: 55–64.

63. SUTHERLAND, G. R., BAKER, E., FERNANDEZ, K. E. W., CALLEN, D. F., ALDRED, P., COGHLAN, J. P., WRIGHT, R. E., AND FERNLEY, R. T. (1989). The gene for human carbonic anhydrase VI (CA VI) is on the tip of the short arm of chromosome 1. *Cytogenet. Cell Genet.* **50**: 149–150.

64. TABOR, S., AND RICHARDSON, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84**: 4767–4771.

65. TASHIAN, R. E. (1989). The carbonic anhydrases: Widening perspectives on their evolution, expression and function. *BioEssays* **10**: 186–192.

66. TASHIAN, R. E., VENTA, P. J., NICEWANDER, P. H., AND HEWETT-EMMETT, D. (1990). Evolution, structure, and expression of the carbonic anhydrase multigene family. *Prog. Clin. Biol. Res.* **344**: 159–175.

67. VALERIO, D., DUYVESTEYN, M. G. C., DEKKER, B. M. M., WEEDA, G., BERKVENS, TH. M., VAN DER VOORN, L., VAN ORMONDT, H., AND VAN DER EB, A. J. (1985). Adenosine deaminase: Characterization and expression of a gene with a remarkable promoter. *EMBO J.* **4**: 437–443.

68. VENTA, P. J., MONTGOMERY, J. C., HEWETT-EMMETT, D.,

AND TASHIAN, R. E. (1985). Comparison of the 5′ regions of the human and mouse carbonic anhydrase II genes and identification of possible regulatory elements. *Biochim. Biophys. Acta* **826**: 195–201.

69. VENTA, P. J., MONTGOMERY, J. C., WEIBAUER, K., HEWETT-EMMETT, D., AND TASHIAN, R. E. (1984). Organization of the mouse and human carbonic anhydrase II genes. *Ann. NY Acad. Sci.* **429**: 309–323.

70. VIERRA, J., AND MESSING, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**: 259–268.

71. WEBER, J. L., AND MAY, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**: 388–396.

72. ZABEL, B. U., NAYLOR, S. L., SAKAGUCHI, A. Y., BELL, G. I., AND SHOWS, T. B. (1983). High resolution chromosomal localization of human genes for amylase, proopimeolanocortin, somatostatin, and a DNA fragment (D3S1) by *in situ* hybridization. *Proc. Natl. Acad. Sci. USA* **80**: 6932–6936.

73. ZHU, X. L., AND SLY, W. S. (1990). Carbonic anhydrase IV from human lung. Purification, characterization, and comparison with membrane carbonic anhydrase from human kidney. *J. Biol. Chem.* **265**: 8795–8801.