Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 9571–9575, December 1986
Genetics

Nucleotide sequence, tissue-specific expression, and chromosome location of human carbonic anhydrase III: The human *CAIII* gene is located on the same chromosome as the closely linked *CAI* and *CAII* genes

(multigene families/muscle gene expression)

ROBERT WADE*, PETER GUNNING*, ROGER EDDY†, THOMAS SHOWS†, AND LARRY KEDES*

*The Medigen Project, Department of Medicine, Stanford Medical School and Veterans Administration Medical Center, Palo Alto, CA 94304; and †Department of Human Genetics, Roswell Park Memorial Institute, New York State Department of Health, Buffalo, NY 14263

Communicated by Joseph G. Gall, August 21, 1986

ABSTRACT **The carbonic anhydrases (CA) are a class of metalloenzymes that catalyze the reversible hydration of carbon dioxide. The genes for the carbonic anhydrase isozymes are members of a multigene family that are differentially expressed in a number of cell types. We have isolated a full-length representative of a CAIII mRNA transcript from an adult human muscle cDNA library, and we present the complete nucleotide sequence of this cDNA clone. RNA blots demonstrate that CAIII messages can be detected in a variety of cell types but that high-level expression is limited to human fetal and adult skeletal muscle and to rodent slow skeletal muscle and liver. In addition, we have used a panel of human–mouse cell hybrids to localize the human *CAIII* gene to chromosome 8. Previous reports have established the *CAI* and *CAII* isozyme genes to be closely linked on chromosome 8, and the assignment of the *CAIII* gene to the same chromosome raises the possibility that these genes may all be linked at a single complex locus.**

The carbonic anhydrases (CAs) are a class of zinc metalloenzymes that catalyze the reversible hydration of carbon dioxide (reviewed in refs. 1 and 2). In mammals, three isozymes have been well-characterized (CAI, CAII, and CAIII), and there is evidence that at least two other isozymes may also exist (3–5). CAI is a major protein constituent of erythrocytes, whereas CAII is found in a wide variety of tissues. Genetic deficiencies of CAI with no notable clinical consequences have been identified in humans (6). However, CAII deficiency has been directly correlated with an inherited syndrome marked by osteopetrosis with renal tubular acidosis and cerebral calcification (7). Genetic studies have shown that the genes encoding CAI and CAII are closely linked in mammals (8–10) and the availability of cloned probes of the mouse *CAII* gene has led to the assignment of these genes to chromosome 8 in humans (11).

CAIII differs markedly from the CAI and CAII isozymes by its comparatively low CO_2 hydrase activity, high K_m value, and resistance to inhibition by sulfonamides (1, 2). CAIII is largely restricted to skeletal muscles characterized by slow twitch myofibers, where it comprises up to 50% of the soluble protein (12). High levels of CAIII have also been found in rodent liver, where it appears that its expression level is at least partially subject to hormonal regulation (13). The linkage relationship of this isozyme with that of the other CAs is not yet known.

We have recently described the isolation of a large number of full-length cDNA clones from a human skeletal muscle cDNA library that correspond to abundant muscle-specific transcripts (14). One of these cDNAs, designated H16a in the initial study, was found to represent an RNA transcript present at high levels in adult human skeletal muscle but absent from both human heart and fibroblast RNA. We report here the identification of this cDNA as a full-length representative of a human CAIII transcript, its complete nucleotide sequence, and its expression in human and rodent tissues. In addition, we have identified a subsegment of the CAIII cDNA whose sequence is human specific. We have used this subsegment probe and a panel of mouse–human somatic cell hybrids to assign the structural gene for CAIII to chromosome 8. This result raises the possibility that all three CA genes may be linked within the genome.

MATERIALS AND METHODS

Sequencing and Clone Identification. DNA sequencing was conducted by directional subcloning of intact and BAL-31 exonuclease-treated restriction fragments into the vectors M13mp10 and M13mp11. Each region of the cDNA was sequenced at least twice along both strands by the method of Sanger *et al.* (15). The predicted translation product of the cDNA was determined by using the SEQ computer program for DNA sequence analysis. The predicted translation product was compared with the National Biomedical Research Foundation protein sequence data base by using the IFIND data base searching program. Additional protein sequences were entered by use of the PEP sequence program. SEQ, IFIND, and PEP are software products of IntelliGenetics (Mountainview, CA).

Preparation of RNA. Total RNA was isolated from human and rodent skeletal and heart muscle by the phenol/chloroform method (16). RNA from other rodent tissues was prepared by a modification of the guanidine hydrochloride method (17).

Filter Hybridizations. Southern blots and RNA blots were prepared (18) and the nitrocellulose filters were hybridized with nick-translated probes (19) as described (14). Filters bearing human DNA or RNA were washed in 0.5× SSC/0.1% $NaDodSO_4$ at 65°C (1× SSC is 0.15 M NaCl/0.015 M sodium citrate). Filters bearing DNA or RNA from either rabbit or mouse tissues were washed in 0.5× SSC/0.1% $NaDodSO_4$ at 55°C. After autoradiography, the levels of

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: CA, carbonic anhydrase; bp, base pair(s); kb, kilobase(s); UTR, untranslated region.

‡National Biomedical Research Foundation (1985) Protein Sequence Database, Protein Identification Resource (Natl. Biomed. Res. Found., Washington, DC 20007), Release No. 4.0.

hybridizing transcripts were quantitated by scanning densitometry.

RESULTS

Sequence and Identification of a Full-Length Human CAIII cDNA Clone. We have recently reported the isolation of a number of full-length cDNA clones representing abundant muscle mRNA species (14). Fig. 1 depicts key restriction endonuclease sites on a map of the cDNA clone that was designated H16a in our previous study. In an attempt to identify this clone, we initially sequenced a 200-base-pair (bp) segment of the cDNA from the *Pst* I site indicated in Fig. 1 and identified a single open reading frame. The predicted translation product was used in a computer-assisted search of the National Biomedical Research Foundation protein sequence data base and was found to be related to the CA protein sequences contained in the data base. Upon sequencing the entire insert, the H16a cDNA clone was found to contain an ≈1.8 kilobase-pair (kbp) insert [including poly(G) and poly(A) tails] encoding a protein of 260 amino acids with 5′ and 3′ untranslated regions of 53 and 893 bases, respectively (Fig. 2).

The complete amino acid sequences of equine and bovine carbonic anhydrase III are known (21–23). Alignment of the deduced human amino acid sequence with that of the bovine and equine proteins revealed 86% and 88% similarity, respectively. A three-way comparison of the predicted human protein with equine and bovine CAIII showed >90% similarity of the human protein with either one or both of the ungulate amino acid sequences. Comparison of complete and partial amino acid sequences of CA isozymes from a number of species has identified 40 invariant amino acid residues at specific positions that distinguish the CAIII protein from that of CAI and CAII (1). The deduced sequence of the human CA protein shown in Fig. 2 contains all 40 of these invariant amino acids at their appropriate positions, firmly establishing this clone to be a human CAIII cDNA. In addition, the amino acid sequence deduced from this cDNA is in good agreement with partial amino acid sequence data previously available for human CAIII (1, 20). The partial amino acid sequencing of human CAIII has revealed a widespread isoleucine/valine polymorphism at residue 31 (20). Position 31 of the derived protein sequence in Fig. 2 is occupied by a valine, which is underlined since it differs from the isoleucine found at this position in both the bovine and equine sequences.

Expression of CAIII in Human and Mouse Tissues. We used the pHMCAIII clone to examine the tissue-specific distribution of CAIII transcripts. We have previously shown that the

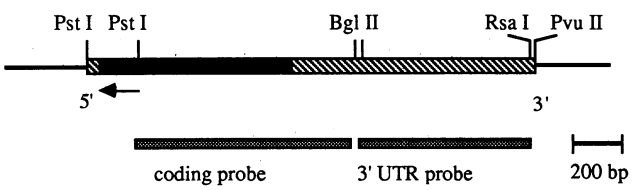

FIG. 1. Restriction endonuclease map of the cDNA insert of pHMCAIII and localization of restriction fragments used as coding and 3′ UTR probes. Vector sequence is indicated by the thin line; cDNA insert is represented as a thick line; solid bar indicates coding region; hatched bar corresponds to the 5′ and 3′ UTRs. Arrow denotes the region of the cDNA that was initially sequenced from the *Pst* I site to identify the clone. Stippled bars beneath the map represent the restriction endonuclease-digested DNA fragments that were used as probes. Coding region probe is a *Pst* I/*Bgl* II fragment containing the entire coding segment from amino acid 16 and extending 140 bases into the 3′ UTR. The UTR probe consists of a 671-bp *Bgl* II/*Rsa* I fragment that begins 183 bases within the 3′ UTR and extends to base 853 of the 893-bp 3′ UTR, thus encompassing most of the noncoding segment.

species of RNA represented by pHMCAIII is present in adult human skeletal muscle at levels comparable to that of actin mRNA, yet it is virtually absent in human heart muscle and cultured human fibroblasts (14). We expanded our examination of the tissue-specific expression of CAIII to additional human and rodent tissues. A nick-translated coding region probe was prepared from the 830-bp *Pst* I/*Bgl* II fragment indicated in Fig. 1. The probe was hybridized with RNA blots of total RNA isolated from various human and rodent tissues. This CAIII coding region probe presumably detects only CAIII transcripts, and not the other CA isozyme transcripts, in both human and rodent tissues since (*i*) CAI and CAII protein sequences are widely diverged from that of CAIII (1), (*ii*) the CAIII coding region probe detects only a single copy gene in human DNA, and (*iii*) CAI and CAII transcripts differ in size from the transcripts that hybridize with the CAIII probe (24–26).

The probe detects a single ≈1.8 kb transcript in both 24-week-old fetal and adult human skeletal muscle RNA samples (Fig. 3*A*), but the transcript is present at 2.5-fold higher levels in adult skeletal muscle, as measured by densitometer tracing. This result correlates well with the fact that CAIII protein concentration in adult skeletal muscle is severalfold higher than in fetal skeletal muscle (27).

Human skeletal muscles comprise a mixture of fast- and slow-twitch types of myofibers, and it is not known whether expression of CAIII in humans is predominantly limited to the slow muscle fibers. In rodents and lagamorphs, such as rabbits, the distribution of CAIII protein among fast- and slow-twitch fibers can be clearly established due to the limited mixing of fiber types in individual muscles. CAIII accounts for nearly 4% of the total protein of rat soleus muscle (a muscle that comprises almost exclusively red, or slow-twitch, fibers), but in muscles that comprise predominantly white, or fast-twitch, fibers CAIII has been detected only by immunological techniques and at much lower concentrations (12). We examined the level of CAIII transcripts in rabbit soleus (predominantly slow-twitch fibers), ventrical (cardiac fibers), and psoas (predominantly fast-twitch fibers) by hybridizing an RNA blot of total RNA isolated from these muscle tissues with the nick-translated coding region probe. Fig. 3*B* shows the autoradiograph of this blot, which clearly demonstrates that slow-twitch fiber rabbit soleus muscle (lane 1) contains high levels of a single RNA species that cross-hybridizes with the human CAIII probe, whereas no signal is detected in rabbit ventricle (lane 2) or fast fiber psoas RNA (lane 3). We conclude that the disparity in levels of CAIII protein between different muscle types is correlated with the relative levels of CAIII transcripts (12).

Significant levels of CAIII (up to 1% of total protein) are found in adult male rodent liver (12). In addition, trace quantities of CAIII have been detected immunologically in rodent heart, kidney, brain, and testis. We next sought to examine the distribution of CAIII mRNA in a variety of mouse tissues. Fig. 3*C* represents an autoradiograph of an RNA blot of total RNA isolated from male mouse heart, diaphragm, and liver. The blot was hybridized with the radiolabeled coding region probe from the CAIII cDNA. Comparison of signal intensity shows that the mouse diaphragm, a muscle with a significant fraction of slow-twitch muscle fibers, contains the highest level of the 1.8-kb long hybridizing transcripts (lane 1). The mouse liver also contains significant quantities of 1.8-kb long hybridizing transcripts (lane 3) but at one-third the levels of diaphragm RNA. Only about one-third of the muscle fibers of the rat diaphragm are of the slow-twitch type (28). If slow-twitch fibers are also the minor type of fibers in the mouse diaphragm, the 3-fold difference between CAIII message levels in mouse slow-twitch muscle fibers and male liver would be underestimated in this experiment. We have not detected CAIII transcripts in

```
G19ACCACGCAGGGGAAGAGAAAGCAGGAGCCGTCCAGCACGGAGGAAGGCAGACC ATG GCC AAG GAG TGG GGC TAC GCC AGT CAC AAC GGT CCT GAC (114)
                                                          MET Ala Lys Glu Trp Gly Tyr Ala Ser His Asn Gly Pro Asp (14)

CAC TGG CAT GAA CTT TTC CCA AAT GCC AAG GGG GAA AAC CAG TCG CCC GTT GAG CTG CAT ACT AAA GAC ATC AGG CAT GAC CCT (198)
His Trp His Glu Leu Phe Pro Asn Ala Lys Gly Glu Asn Gln Ser Pro Val Glu Leu His Thr Lys Asp Ile Arg His Asp Pro  (42)
                                                                (31)
TCT CTG CAG CCA TGG TCT GTG TCT TAT GAT GGT GGC TCT GCC AAG ACC ATC CTG AAT AAT GGG AAG ACC TGC CGA GTT GTA TTT (282)
Ser Leu Gln Pro Trp Ser Val Ser Tyr Asp Gly Gly Ser Ala Lys Thr Ile Leu Asn Asn Gly Lys Thr Cys Arg Val Val Phe  (70)

GAT GAT ACT TAT GAT AGG TCA ATG CTG AGA GGG GGT CCT CTC CCT GGA CCC TAC CGA CTT CGC CAG TTT CAT CTT CAC TGG GGC (366)
Asp Asp Thr Tyr Asp Arg Ser Met Leu Arg Gly Gly Pro Leu Pro Gly Pro Tyr Arg Leu Arg Gln Phe His Leu His Trp Gly  (98)

TCT TCG GAT GAT CAT GGC TCT GAG CAC ACC GTG GAT GGA GTC AAG TAT GCA GCG GAG CTT CAT TTG GTT CAC TGG AAC CCG AAG (450)
Ser Ser Asp Asp His Gly Ser Glu His Thr Val Asp Gly Val Lys Tyr Ala Ala Glu Leu His Leu Val His Trp Asn Pro Lys (126)

TAT AAC ACT TTT AAA GAA GCC CTG AAG CAG CGC GAT GGG ATC GCT GTG ATT GGC ATT TTT CTG AAG ATA GGA CAT GAG AAT GGC (534)
Tyr Asn Thr Phe Lys Glu Ala Leu Lys Gln Arg Asp Gly Ile Ala Val Ile Gly Ile Phe Leu Lys Ile Gly His Glu Asn Gly (154)

GAG TTC CAG ATT TTC CTT GAT GCA TTG GAC AAG ATT AAG ACA AAG GGC AAG GAG GCG CCC TTC ACA AAG TTT GAC CCA TCC TGC (618)
Glu Phe Gln Ile Phe Leu Asp Ala Leu Asp Lys Ile Lys Thr Lys Gly Lys Glu Ala Pro Phe Thr Lys Phe Asp Pro Ser Cys (182)

CTG TTC CCG GCA TGC CGG GAC TAC TGG ACC TAC CAG GGC TCA TTC ACC ACG CCG CCC TGC GAG GAA TGC ATT GTG TGG CTG CTG (702)
Leu Phe Pro Ala Cys Arg Asp Tyr Trp Thr Tyr Gln Gly Ser Phe Thr Thr Pro Pro Cys Glu Glu Cys Ile Val Trp Leu Leu (210)

CTG AAG GAG CCC ATG ACC GTG AGC TCT GAC CAG ATG GCC AAG CTG CGG AGC CTC CTC TCC AGT GCT GAG AAC GAG CCC CCA GTG (786)
Leu Lys Glu Pro Met Thr Val Ser Ser Asp Gln Met Ala Lys Leu Arg Ser Leu Leu Ser Ser Ala Glu Asn Glu Pro Pro Val (238)

CCT CTT GTG AGC AAC TGG CGA CCT CCA CAG CCT ATC AAT AAC AGG GTG GTG AGA GCT TCC TTC AAA TGA GGCTGCTGGATCTTGCCCT (874)
Pro Leu Val Ser Asn Trp Arg Pro Pro Gln Pro Ile Asn Asn Arg Val Val Arg Ala Ser Phe Lys                          (260)

CTTCAGGAAAGGAAACCTACCATTGGAGAGCTTGGTTCCTTGCCTCCTTCTGGTGCTCTTACTCCAAGTCTATTTCATTTTTCCACACTGAGCAATGAATGTGAGAGATG  (984)
TGGTCACCAAGATCTAAGTTACTTGTTGAAAGAAAGTTACTTTCGACAAGATCTAATATGAAAGCATAGATTTCACATTTGATCTCTGTAATAATCATCTTTCCTATAAA (1094)
AGTAGCATTTTTGGTAAAGTTTCAAAGAAGAAGAAACAGAGATGGAAGAGTAAAGATATTTTTAAAATGGCTAGCTATTGGGCACCAGTTTTTCTGTTATCTAAAATTTC (1204)
ACACAACTTCATGTTTTTATTTTTATATTATGAGTTGTCCATCTTAAAGAAATATGAGTAATTCTACATGTAGTAGAGGTGTATGAAGATCATATAACAATTAAACATAA (1314)
GCCAGAAATTAAAATGACTATAGACAGCAAGAATTGAGCTAATAATATGTTTTAACTCTTAACACCAGCAAGAAGTCAGTCATTTATTGAAGTTTTAGCTACTAAGATTA (1424)
CTTGGTTTTGATTACCAGTGAAAAGAAAACACAATACAATCAGGAGTTTTCAAATTTTTGATTCAGTATTTGAATTTCTTCTTCATAAATGTAGTTGAATTTATCCTAGT (1534)
ATTTTTCTTTACCTGAAGGAGGGCCATTTATTTTTAATTTCACTACATTTTTCTTTGCATGATTATTAAAATAAAAACTGCCTCTGTTGTGTTTCTCACTGGAGGCTGGA (1644)
ATGAATGATCACTAGAACACAAAAGAGTGAATGATGACACTTGAAGTCAAAGCAGTTGTACTGATCACCAGAACCAATAAAGACATAAATGGAAAACGTTG(A)n      (1775)
```

FIG. 2. Nucleotide sequence and protein translation of pHMCAIII. The nucleotide sequence of the human CAIII cDNA [including the poly(G) linker segment used in construction of the cDNA library] is presented along with the deduced amino acid sequence. Underlined amino acids correspond to residues found to be unique to the human sequence as compared to bovine and equine CAIII. The valine residue at position 31 indicates the cryptic polymorphic site that has been identified (20). The polyadenylylation signal AATAAA is also underlined.

RNA from cardiac muscle. However, trace amounts of CAIII protein have been detected immunologically in rodent cardiac tissue (12). Our nucleic acid hybridization experiments with small quantities of total RNA may not have been sensitive enough to detect an amount of mRNA that produces only trace quantities of CAIII in cardiac muscle.

Fig. 3*D* shows an RNA blot of total RNA from a variety of mouse muscle and nonmuscle tissues that was hybridized with the human CAIII coding region probe. Most of the tissues examined were found to contain CAIII transcripts, but at much lower levels than that of diaphragm or male liver. We quantitated the relative levels of CAIII message in the various RNA samples by densitometer tracing. Total hind limb muscle (comprised of a mixture of individual leg muscles) contained 12% of the level of CAIII transcripts seen in male liver RNA. Lung and kidney RNA contained 5% and 4%, respectively, of the level in liver. Trace signals (<2% of that in the liver) were found in mouse testis and brain. No CAIII mRNA was detected with RNA prepared from stomach or heart. The relatively low amount of CAIII-related transcripts in RNA prepared from total hind limb muscles is most probably a reflection of the abundance in this sample of muscles rich in fast-twitch fibers. Slow-twitch fiber-rich muscles such as the soleus only account for a very small fraction of the mass of total hind limb muscles. Thus, as is the case in rabbit fast-twitch muscles, mouse muscles rich in fast-twitch myofibers probably contain little or no CAIII mRNA.

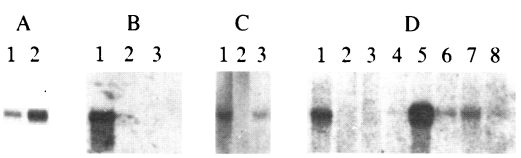

FIG. 3. Distribution of CAIII transcripts in human and rodent tissues. A series of autoradiograms is shown of RNA blots prepared with 2-μg quantities of total RNA from a variety of human and mouse tissues, which were then hybridized with the radiolabeled 830-bp *Pst* I/*Bgl* II coding region fragment of the cDNA insert of pHMCAIII (see Fig. 1). The tissues or organs from which the RNAs in each panel were derived are as follows: (*A*) Lane 1, 24-week-old fetus skeletal muscle; lane 2, adult human skeletal muscle. (*B*) Lane 1, rabbit soleus muscle; lane 2, heart ventrical; lane 3, psoas muscle. (*C*) Lane 1, mouse diaphragm; lane 2, heart; lane 3, liver. (*D*) Lane 1, mouse hind limb skeletal muscle; lane 2, heart; lane 3, stomach; lane 4, brain; lane 5, liver; lane 6, kidney; lane 7, lung; lane 8, testis. Blot was deliberately over-exposed to reveal the weak signals seen with RNA from several of the mouse tissues [e.g., compare the relative intensity of signal between the liver RNA samples in lane 3 (*C*), with lane 5 (*D*)]. A trace signal from brain and testis, visible in the original autoradiograph of *D*, did not reproduce well.

CAIII transcripts are found in a variety of tissues, but the steady-state levels of the message in different tissues appears to be modulated over 2 or 3 orders of magnitude. In the tissues that contained detectable levels of CAIII message, the relative level of transcript correlated well with the level of CAIII protein that have been measured in rodent tissues (12). It appears likely that the steady-state level of CAIII message is the primary means of regulating the level of CAIII protein.

We also used a segment of the 3′ untranslated region (UTR) of the CAIII cDNA (indicated in Fig. 1) as probe against an RNA blot panel of rodent RNAs but we detected no signal, even under reduced hybridization criteria (data not shown). This result indicated that the 3′ UTRs of human and rodent CAIII messages have undergone extensive sequence divergence and allowed us to use the 3′ UTR segment of pHMCAIII as a human-specific probe.

Chromosomal Assignment of the Human *CAIII* Gene. We hybridized the 3′ UTR probe with Southern blots of human DNA. As shown in Fig. 4, the 3′ UTR probe from the human CAIII cDNA hybridizes to a single ≈9-kbp *Eco*RI restriction fragment in human DNA. Single hybridizing bands were also

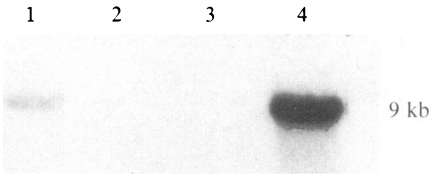

FIG. 4. Distribution of human CAIII sequences in representative cell hybrids. Portion of an autoradiograph resulting from hybridization of the species-specific human CAIII DNA probe to a Southern blot panel of *Eco*RI-digested DNA from human–mouse cell hybrids and their parental cell lines. The radiolabeled probe was prepared from the 3′ UTR region of pHMCAIII (Fig. 1). The probe hybridizes to a single 9-kb human DNA fragment. DNA from each of the following sources was loaded on the gel lanes: lane 1, WIL-14 (+), a hybrid cell line that has retained human chromosome 8; lane 2, wil-15 (−), a hybrid cell line that has lost human chromosome 8; lane 3, mouse Ltk$^-$ cells; lane 4, WI38 cells (human DNA control). The parental cell lines and the fusion, isolation, and characterization of the hybrid cell lines have been described (29–33). The size of the hybridizing human fragment was estimated by comparing its mobility with that of λ DNA *Hin*dIII molecular weight markers.

seen with blots of human DNA digested with seven other restriction enzymes (data not shown), and we concluded that the *CAIII* gene is present as a single copy in the human genome.

Thirty-one human–mouse somatic cell hybrids were tested for the presence or absence of the human CAIII sequence and scored against the presence or absence of specific human chromosomes in the hybrid lines. *Eco*RI-digested DNAs from the somatic cell hybrids as well as from the parental mouse and human cell lines were separated by gel electrophoresis, blotted onto nitrocellulose filters, and challenged with the radiolabeled human species-specific CAIII 3′UTR probe. An autoradiogram of a portion of such a Southern blot panel is shown in Fig. 4. The probe hybridizes to a single 9-kbp human *Eco*RI fragment but does not hybridize to mouse DNA. It was therefore possible to score which somatic cell hybrid lines contained the human gene. The results of the segregation analysis are presented in Table 1. The human CAIII signal segregated concordantly with chromosome 8. Discordant segregation was noted for all other chromosomes. We thus assign the human *CAIII* gene to chromosome 8. Chromosome 8 has previously been shown to contain the structural gene for CAII, and by virtue of its linkage to CAII in mammals, the gene encoding CAI (8–10). The assignment of the human *CAIII* gene to chromosome 8 establishes the possibility that all of the well-characterized members of the CA multigene family are linked and that together they may constitute a complex regulatory locus.

Table 1. CAIII segregation with human chromosomes in 31 human–mouse cell hybrids

	Probe/chromosome				
	Concordant		Discordant		%
Chromosome	+/+	−/−	+/−	−/+	discordancy
1	8	11	11	1	39
2	11	10	8	2	32
3	12	4	6	8	47
4	11	8	8	4	39
5	13	7	6	5	35
6	8	10	11	2	42
7	12	7	6	5	37
8	19	12	0	0	0
9	3	11	14	1	52
10	16	7	3	5	26
11	10	7	8	5	43
12	13	9	6	3	29
13	9	5	10	7	55
14	15	5	4	7	35
15	14	9	5	2	23
16	9	9	10	3	42
17	17	5	2	7	29
18	11	6	8	6	45
19	9	11	10	1	35
20	12	6	7	6	42
21	15	3	4	9	42
22	9	6	9	5	48
X	11	6	5	4	35

Compiled from 31 cell hybrids involving 12 unrelated human cell lines and 4 mouse cell lines (30–32). The hybrids were characterized by chromosome analysis and by mapped enzyme markers, and partly by mapped DNA probes (31–33). Concordant hybrids have either retained or lost the CAIII signal together with a specific human chromosome. Discordant hybrids either retained the gene, but not a specific chromosome, or the reverse. % discordancy indicates the degree of discordant segregation for the marker and a chromosome. A 0% discordancy is the basis for chromosome assignment.

DISCUSSION

We have described the identification and structural characterization of a full-length adult muscle cDNA representing human CAIII. Recently, Lloyd *et al.* (34) described the isolation of what appears to be a near full-length human CAIII cDNA, and the sequence reported here is in complete agreement with the 63 bp of preliminary sequence that they presented. However, in contrast to their inference that the human CAIII message may be present in much lower amounts than would be expected from the protein determinations, we have found the human CAIII to be among the most abundant messages in human skeletal muscle and among the most abundant classes of clones in our human muscle cDNA library (14). The results of dot blot quantitation experiments indicate that the human CAIII message is present in adult human skeletal muscle at a level 1.5- to 2.0-fold higher than the level of actin mRNA (P.G., unpublished data). RNA blot experiments demonstrate that high levels of CAIII message are present in rodent slow-twitch fiber muscle but are virtually absent in fast-twitch fiber muscle. The discrepancy in the estimates of CAIII message levels in human skeletal muscle between our study and that of Lloyd *et al.* (34) may reflect differing fiber type composition of the human muscle tissues that served as the source of RNA in the two studies.

Surprisingly, less than one-half (780 bp) of the 1.8-kb CAIII transcript encodes protein. The bulk of the human CAIII mRNA (893 bp) is the 3′ UTR. RNA blots of mouse muscle RNAs hybridized with a coding region probe from the human CAIII cDNA indicated that this rodent species possesses CAIII transcripts nearly identical in size to that of the human mRNA. Since rodent and human CAIII proteins are of identical size, the rodent CAIII messages almost certainly contain unusually large 3′ UTRs as well. However, the rodent CAIII transcripts failed to hybridize, even under reduced stringency, with a probe containing most of the 3′ UTR from the human CAIII cDNA. Thus, despite the apparent conservation of the size of the 3′ UTR between humans and rodents, there has been no strong interspecies conservation of the sequence of the 3′ UTR. Since the human CAIII 3′ UTR probe did not cover the entire length of the UTR, it is possible that a short region of sequence conservation escaped our detection.

The human CAIII 3′ UTR served as a species-specific probe that was hybridized with Southern blot panels of DNA from human–mouse somatic cell hybrids to map the human *CAIII* gene to chromosome 8. Venta *et al.* (11) have used similar panels to localize the human *CAII* gene to chromo-

some 8, and genetic studies have established that the *CAI* and *CAII* genes are closely linked in mammals (8–10). Thus, the assignment of the *CAIII* gene to the same chromosome as the closely linked *CAI* and *CAII* genes suggests that all three of these *CA* genes may be clustered in a common locus. The gene duplication events that gave rise to the multiple CA isozymes appear to have occurred prior to the radiation of the amniotes, ≈300 million years ago (1). Strong evolutionary pressure must have been necessary to retain the proposed linkage arrangement of these three genes over such a long period of time. Genomic cloning and the generation of polymorphic markers will be necessary before the linkage of CAIII to the other CAs can be firmly established. The genes encoding the various CA isozymes are differentially expressed in a wide variety of cell types (1). In this report, RNA blots have detected CAIII transcripts in RNA isolated from human fetal and adult muscle, rodent slow-twitch skeletal muscle, liver, lung, kidney, brain, and testis. We have previously shown that little or no CAIII message is present in human heart tissue and cultured fibroblasts (14), and we extend that observation here to include rodent heart, fast-twitch skeletal muscle, and stomach. The wide range in steady-state levels of CAIII tanscripts in the various tissues examined suggests that the expression level of the active *CAIII* gene is subject to a high degree of modulation. However, we cannot presently exclude the possibility that the observed variability in CAIII message levels may reflect the fraction of the cell population expressing the gene in a given tissue.

While most genes encoding tissue-specific isozymes have not been found to be linked within the genome, there are other examples of such clustered multigene families. The α-fetoprotein and albumin genes are closely linked duplicated genes that are coactivated during mouse embryonic development. The intergenic region of these genes contains regulatory sequences that may play a role in their tissue-specific coactivation (35), and this may account for the selective pressure that would seem to have been necessary to maintain such a linkage arrangement over long evolutionary periods. If the *CAIII* gene is closely linked to the *CAI* and *CAII* genes, it should be possible to investigate whether similar mechanisms characterize this multigene family.

Note Added in Proof. Lloyd *et al.* (36) have recently reported the complete sequence of an independently isolated near full-length human CAIII cDNA that contains the isoleucine variant at position 31. They have also assigned the human CAIII gene to chromosome 8 (37).

This work was supported in part by grants from Veterans Administration (to L.K.) and the National Institutes of Health Public Health Service Grants HD17031 (to L.K.) and GM20454 and HD05196 (to T.S.). R.W. is a Fellow of the Muscular Dystrophy Association.

1. Hewett-Emmett, D., Hopkins, P. J., Tashian, R. E. & Czelusniak, J. (1984) *Ann. N.Y. Acad. Sci.* **429**, 338–358.
2. Tashian, R. E., Hewett-Emmett, D. & Goodman, M. (1982) in *Isozymes: Current Topics in Biological and Medical Research: Molecular Structure and Regulation*, eds. Rattazzi, M. C., Scandalios, J. G. & Whitt, G. S. (Liss, New York), Vol. 7, pp. 79–100.
3. Whitney, P. L. & Briggle, T. V. (1982) *J. Biol. Chem.* **257**, 12056–12059.
4. Dodgson, S. J., Forster, R. E., Storey, B. T. & Mela, L. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5562–5566.
5. Vincent, S. H. & Silverman, D. N. (1982) *J. Biol. Chem.* **257**, 6850–6855.
6. Kendall, A. G. & Tashian, R. E. (1977) *Science* **29**, 471–472.
7. Sly, W. S., Hewett-Emmett, D., Whyte, M. P., Ya-Shiou, Y. & Tashian, R. E. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2752–2756.
8. Carter, N. D. (1972) *Comp. Biochem. Physiol. B* **43**, 743–747.
9. DeSimone, J., Linde, M. & Tashian, R. E. (1973) *Nature (London) New Biol.* **242**, 55–56.
10. Eicher, E. M., Stern, R. H., Womack, J. E., Davison, M. T., Roderick, T. H. & Reynolds, S. C. (1976) *Biochem. Genet.* **14**, 651–660.
11. Venta, P. J., Shows, T. B., Curtis, P. J. & Tashian, R. E. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4437–4440.
12. Shiels, A., Jeffery, S., Wilson, C. & Carter, N. (1984) *Biochem. J.* **218**, 281–284.
13. Jeffery, S., Wilson, C. A., Mode, A., Gustafsson, J.-A. & Carter, N. D. (1986) *J. Endocrinol.* **110**, 123–126.
14. Garrison, J. C., Hardeman, E., Wade, R., Kedes, L. & Gunning, P. (1985) *Gene* **38**, 177–188.
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
16. Palmiter, R. D. (1974) *Biochemistry* **13**, 3606–3615.
17. Childs, G., Maxson, R. & Kedes, L. H. (1979) *Dev. Biol.* **73**, 153–173.
18. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1983) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
19. Rigby, P. W., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
20. Hewett-Emmett, D., Welty, R. J. & Tashian, R. E. (1983) *Genetics* **105**, 409–420.
21. Wendorff, K. M., Nishita, T., Jabusch, J. R. & Deutsch, H. F. (1985) *J. Biol. Chem.* **260**, 6129–6132.
22. Tashian, R. E., Hewett-Emmett, D., Stroup, S. K., Goodman, M. & Yu, Y.-S. L. (1980) in *Biophysics and Physiology of Carbon Dioxide*, eds. Bauer, C., Gros, G. & Bartels, H. (Springer, Berlin), pp. 165–176.
23. Enberg, P., Millquist, E., Pohl, G. & Lindskog, S. (1985) *Arch. Biochem. Biophys.* **2**, 628–638.
24. Curtis, P. J., Withers, E., Demuth, D., Watt, R., Venta, P. J. & Tashian, R. E. (1983) *Gene* **25**, 325–332.
25. Boyer, S. M., Ostrer, H., Smith, K. D., Young, K. E. & Noyes, A. N. (1984) *Ann. N.Y. Acad. Sci.* **429**, 324–331.
26. Konialis, C. P., Barlow, J. H. & Butterworth, P. H. W. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 663–667.
27. Jeffery, S., Edwards, Y. & Carter, N. (1980) *Biochem. Genet.* **18**, 843–849.
28. Barton, P. J. R., Cohen, A., Robert, B., Fiszman, M. Y., Bonhomme, F., Guenet, J.-L., Leader, D. P. & Buckingham, M. E. (1985) *J. Biol. Chem.* **260**, 8578–8584.
29. Shows, T. B. (1979) in *Genetic Mechanisms of Sexual Development*, eds. Vallet, H. L. & Porter, I. H. (Academic, New York), pp. 253–269.
30. Sakaguchi, A. Y. & Shows, T. B. (1982) *Somatic Cell Genet.* **8**, 83–94.
31. Shows, T., Eddy, R., Haley, L., Byers, M., Henry, M., Fujita, T., Matsui, H. & Taniguchi, T. (1984) *Somatic Cell Mol. Genet.* **10**, 315–318.
32. Shows, T. B., Sakaguchi, A. Y. & Naylor, S. L. (1982) in *Advances in Human Genetics*, eds. Harris, H. & Hirschhorn, K. (Plenum, New York), Vol. 12, pp. 341–452.
33. Shows, T. B. (1983) in *Isozymes: Current Topics in Biological and Medical Research: Genetics and Evolution*, eds. Rattazzi, M. C., Scandalios, J. G. & Whitt, G. S. (Liss, New York), Vol. 10, pp. 323–339.
34. Lloyd, J. C., Isenberg, H., Hopkinson, D. A. & Edwards, Y. H. (1985) *Ann. Hum. Genet.* **49**, 241–251.
35. Godbout, R., Ingram, I. & Tilghman, S. M. (1986) *Mol. Cell. Biol.* **6**, 477–487.
36. Lloyd, J., McMillan, S., Hopkinson, D. & Edwards, Y. H. (1986) *Gene* **41**, 233–239.
37. Edwards, Y. H., Barlow, J. H., Konialis, C. P., Povey, S. & Butterworth, P. H. W. (1986) *Ann. Hum. Genet.* **50**, 123–129.