Gene, 25 (1983) 325–332
Elsevier

GENE 884

The nucleotide sequence and derived amino acid sequence of cDNA coding for mouse carbonic anhydrase II

(Comparison with mammalian carbonic anhydrase isozymes; evolution; unique residues; long 3′ noncoding region; recombinant DNA)

Peter J. Curtis *, Elizabeth Withers, Donald Demuth, Rosemary Watt, Patrick J. Venta **, and Richard E. Tashian **

*The Wistar Institute of Anatomy and Biology, 36th Street at Spruce, Philadelphia, PA 19104 (U.S.A.) Tel. (215) 898-3859, and ** Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109 (U.S.A.) Tel. (313) 764-1359*

(Received May 19th, 1983)
(Revision received July 28th, 1983)
(Accepted August 3rd, 1983)

SUMMARY

The nucleotide sequence of a clone containing mouse carbonic anhydrase (CA) cDNA in pBR322 has been determined. The cloned cDNA contains all of the coding region except for nucleotides specifying the first eight amino acids, and all of the 3′ noncoding region, which consists of 700 nucleotides. A cDNA clone was identified which contains an additional 54 bp at the 5′ end, so that the complete amino acid sequence of mouse CA could be deduced. This sequence showed a 73–81% homology with other mammalian CA form II isozymes, 56–63% with form I isozymes, and 52–56% with form III isozymes. By examination of the amino acids which are unique and invariant for each isozyme, the mouse amino acid sequence was found to contain 16 of the 23 residues that are unique and invariant to mammalian CA form II isozymes, but only one or no residue for forms I and III, respectively.

INTRODUCTION

The carbonic anhydrases (EC 4.2.1.1) are zinc metalloenzymes of apparently great evolutionary antiquity which catalyze the reversible hydration of CO_2 (for recent reviews see Pocker and Sarkanen, 1978; Lindskog, 1982; Tashian et al., 1983). At least three isozymes of CA, termed CA I, CA II and CA III, appear to be characteristically present in amniotes (reptiles, birds and mammals) where they are under the control of separate genetic loci. The CA II genes are expressed in a wide variety of cells and tissues, whereas the CA I and CA III genes seem to have a more limited tissue expression (Tashian, 1977; Tashian et al., 1983).

The complete, or partially (>60%) complete, amino acid sequences of six CA I, seven CA II and two CA III isozymes have now been determined from 11 species of mammals (Tashian et al., 1983). These are all monomeric enzymes 259 or 260 amino acid residues in length (M_r approx. 29000). When

* To whom correspondence and reprint requests should be sent at the first address.

Abbreviations: bp, base pairs; CA, carbonic anhydrase.

the sequences of all of these CA isozymes are compared, about 30% of the residues were found to be identical. Despite this relatively high degree of evolutionary homology, however, the often disparate physicochemical and kinetic properties, and considerable variation in tissue distribution of the CA isozymes, suggest different physiological roles for each isozyme.

Although some limited sequence data have been obtained for the CA III isozyme of the rat (Carter et al., 1981), and what appears to be CA II cDNA from the mouse (Curtis, 1983), no extensive nucleotide or amino acid sequences are as yet available for a CA isozyme from any rodent species. In this respect, it will be very useful to determine the sequences of the CA isozyme genes of mice, not only because of their obvious relevance to evolutionary and genetic studies, but also because of their potential use in developmental studies where it is important to identify the CA isozymes, e.g., during normal red cell development, or in transformed mouse erythroleukemia cells (Curtis, 1983). In the present report, the complete nucleotide sequence which includes the entire coding region, of mouse CA cDNA cloned from the mRNA purified from mouse anemic spleen has been determined, and identified as CA II by comparison with the known sequences of other mammalian CA isozymes.

MATERIALS AND METHODS

(a) Isolation of plasmid clones

The characterization of a clone which contained CA mRNA derived from Balb/c mice, p6-69, has been described (Curtis, 1983) and shall be designated as pMCAII. Plasmid was isolated from *Escherichia coli* containing pMCAII by the procedure of Ish-Horowitz and Burke (1981).

Additional plasmids containing CA sequences were identified in a plasmid library containing cDNA sequences from mouse anemic spleen. To make the plasmid cDNA library, double-stranded DNA was synthesized from poly(A)$^+$RNA of mouse anemic spleen, treated with S1 nuclease, tailed with dCMP residues and selected for a size greater than 1000 bp (Curtis, 1983). The tailed DNA was annealed with dG-tailed pBR322 and the hybrid DNA used to transfect *E. coli* HB101. One million tetracycline-resistant colonies were collected from the agar plates and plasmid extracted as described above. The pooled plasmid DNA was used to transfect *E. coli*. To screen the colonies, pMCAII was cut with *Pst*I and fragments derived from the cloned sequences were purified by agarose gel electrophoresis and labeled using calf thymus primers and *E. coli* DNA polymerase I (Summers, 1975).

(b) DNA sequencing

Restriction fragments derived from pMCAII by digestion with *Pst*I, *Bam*HI and *Hin*dIII, were subcloned into M13mp8 and sequenced by the dideoxy chain termination method (Sanger et al., 1977; Messing and Vieira, 1982). Additional sequences were determined by the procedure of Maxam and Gilbert (1977) using 3′ end-labeled restriction fragments.

RESULTS AND DISCUSSION

(a) Strategy of DNA sequencing

The plasmid pMCAII for mouse CA contains an insert of about 1500 bp with two internal *Pst*I sites resulting in three fragments of 700, 450, 320 bp (Fig. 1). These fragments were cloned into M13mp8 cut with *Pst*I. The resultant six clones were sequenced by the dideoxy chain termination method. The two smaller fragments were sequenced completely. For the larger fragment, the middle section of the sequence was obtained from M13mp8 clones containing the *Bam*HI-*Bgl*II fragment as well as the *Hin*dIII-*Pst*I. Additional segments were sequenced by the Maxam and Gilbert method (1977). After digestion separately with *Bam*HI and *Hin*dIII, the 3′ ends were labeled, and singly labeled fragments produced by further digestion with restriction enzymes were purified by agarose gel electrophoresis.

The sequence derived from pMCAII could be translated into an amino acid sequence that shared extensive homology with the other mammalian CA II isozyme sequences. When these amino acid sequences were aligned, the sequence for the first

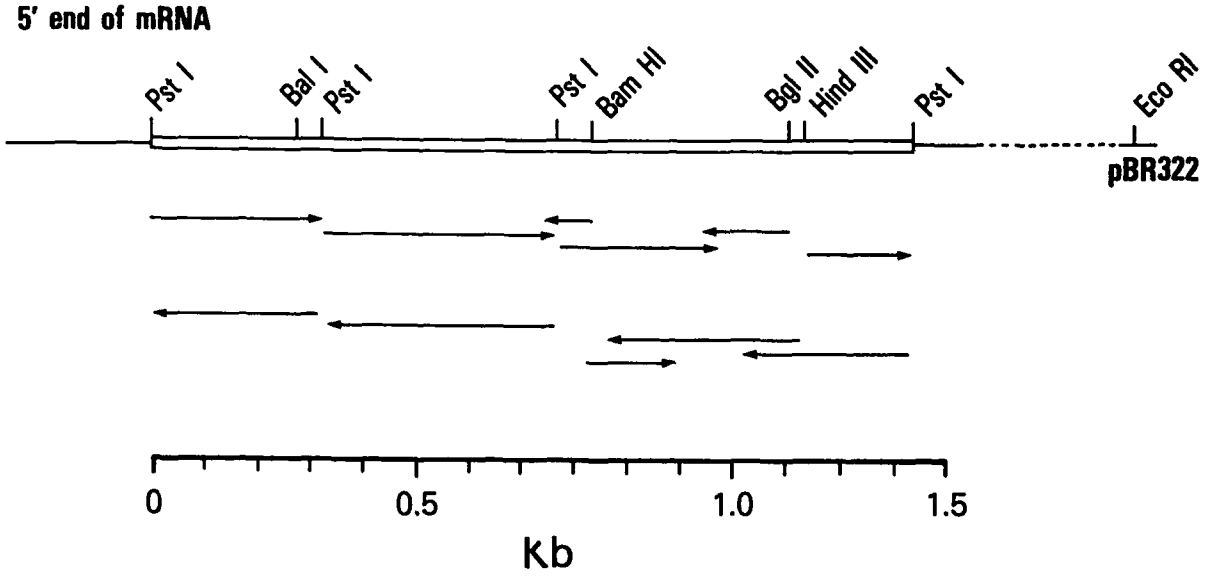

Fig. 1. Restriction map of pMCAII. The arrows indicate the direction and extent of the sequence that was determined by the chain termination method and by the Maxam and Gilbert (1977) procedure.

eight amino acids from the N terminus was missing. In order to obtain this sequence, a library of mouse cDNA sequences derived from poly(A)$^+$ RNA of anemic spleen was screened using the inserted sequences of pMCAII as a labeled probe. Approx. 100 colonies were identified, of which 20 were selected with the strongest response. Plasmid was prepared from these colonies, digested with *Pst*I and analyzed by agarose gel electrophoresis (Fig. 2). From the sequencing it had been determined that the 320-bp fragment contained sequences of the 5′ end of the mRNA, the 450-bp fragment contained sequences of the 3′ end of the coding region of the mRNA, while the 700-bp fragment contained almost all of the 3′ noncoding region. Two plasmids were identified which contained the 450-bp fragment, but had a fragment of approx. 350–400 bp instead of a 320-bp fragment. They also had a third fragment somewhat smaller than the 700-bp fragment of pMCAII. Presumably these two plasmids contain extra sequences at the 5′ end, but lack sequences from the 3′ end. The shortened 3′ end probably arose from incomplete synthesis of the double-stranded cDNA. Only one RNA species was detected by hybridization of pMCAII to RNA fractionated by gel electrophoresis and transferred to aminophenylthioether paper (not shown). In addition the size of the 3′ end fragment was variable in several different plasmids and there is only one poly(A) addition site in the 3′ noncoding sequence. The two plasmids were digested with *Pst*I and *Bal*I and the digest ligated to M13mp8 digested with *Pst*I and *Sma*I. The appropriate clones of M13 were identified by a hybridization test with a M13 clone containing the 320-bp fragment of pMCAII. The M13 clones derived from the two plasmids were sequenced as above.

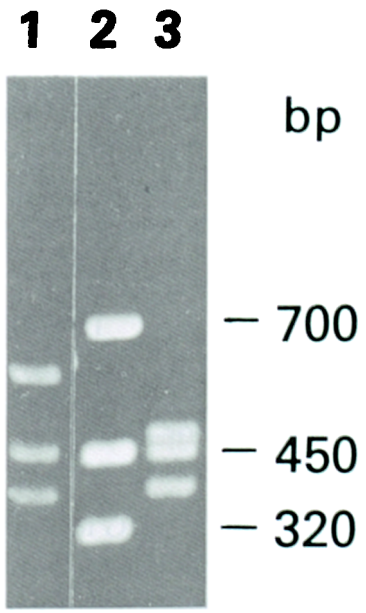

Fig. 2. Agarose gel electrophoresis of plasmids containing mouse CA cDNA. Plasmids were digested with *Pst*I and electrophoresed in a 1.5% agarose gel. Lanes 1, 3: two plasmids containing CA II cDNA. Lane 2: pMCAII.

The combined nucleotide sequences are shown in

```
                                                                                 30
                                                          CTGCTCTGCCCAATCACCGGCGTGACCATG
                                                                                MET

                                                                                                                       120
TCC CAC CAC TGG GGA TAC AGC AAG CAC AAC GGA CCA GAG AAC TGG CAC AAG GAC TTC CCC ATT GCC AAT GGA GAC CGG CAG TCC CCT GTG
Ser His His Trp Gly Tyr Ser Lys His Asn Gly Pro Glu Asn Trp His Lys Asp Phe Pro Ile Ala Asn Gly Asp Arg Gln Ser Pro Val
1                                                    15                                                                 30

                                                                                                                       210
GAT ATT GAC ACA GCA ACT GCC CAC CAT GAC CCT GCC CTA CAG CCT CTG CTC ATA TCT TAT GAT AAA GCT GCG TCC AAG AGC ATT GTC AAC
Asp Ile Asp Thr Ala Thr Ala His His Asp Pro Ala Leu Gln Pro Leu Leu Ile Ser Tyr Asp Lys Ala Ala Ser Lys Ser Ile Val Asn
                                                     45                                                                 60

                                                                                                                       300
AAC GGC CAC TCC TTT AAC GTT GAG TTT GAT GAC TCT CAG GAC AAT GCA GTG CTG AAA GGA GGA CCC CTC AGT GAC TCC TAC AGA TTG ATC
Asn Gly His Ser Phe Asn Val Glu Phe Asp Asp Ser Gln Asp Asn Ala Val Leu Lys Gly Gly Pro Leu Ser Asp Ser Tyr Arg Leu Ile
                                                     75                                                                 90

                                                                                                                       390
CAG TTT CAC TTT CAC TGG GGC TCA TCT GAT GGC CAG GGC TCT GAG CAC ACT GTG AAC AAA AAA AAA TAT GCT GCA GAG CTT CAC TTG GTT
Gln Phe His Phe His Trp Gly Ser Ser Asp Gly Gln Gly Ser Glu His Thr Val Asn Lys Lys Lys Tyr Ala Ala Glu Leu His Leu Val
                                                    105                                                                120

                                                                                                                       480
CAC TGG AAC ACC AAA TAT GGG GAC TTT GGA AAA GCT GTG CAG CAA CCG GAT GGA TTG GCT GTT TTG GGC TAT TTT TTG AAG ATT GGA CCT
His Trp Asn Thr Lys Tyr Gly Asp Phe Gly Lys Ala Val Gln Gln Pro Asp Gly Leu Ala Val Leu Gly Tyr Phe Leu Lys Ile Gly Pro
                                                    135                                                                150

                                                                                                                       570
GCC TCA CAA GGC CTT CAG AAA GTC CTT GAA GCA CTG CAT TCC ATT AAA ACA AAG GGG AAG CGT GCG GCC TTT GCT AAC TTC GAC CCT TGC
Ala Ser Gln Gly Leu Gln Lys Val Leu Glu Ala Leu His Ser Ile Lys Thr Lys Gly Lys Arg Ala Ala Phe Ala Asn Phe Asp Pro Cys
                                                    165                                                                180

                                                                                                                       660
TCC CTT CTT CCT GGA AAC TTG GAC TAC TGG ACA TAC CCT GGC TCT CTG ACC ACT CCG CCT CTG CTG GAA TGT GTG ACC TGG ATC GTG CTC
Ser Leu Leu Pro Gly Asn Leu Asp Tyr Trp Thr Tyr Pro Gly Ser Leu Thr Thr Pro Pro Leu Leu Glu Cys Val Thr Trp Ile Val leu
                                                    195                                                                210

                                                                                                                       750
AGG GAG CCC ATT ACT GTC AGC AGC GAG CAG ATG TCT CAT TTC CGT ACG CTG AAC TTC AAT GAG GAG GGG GAT GCT GAA GAA GCG ATG GTG
Arg Glu Pro Ile Thr Val Ser Ser Glu Gln Met Ser His Phe Arg Thr Leu Asn Phe Asn Glu Glu Gly Asp Ala Glu Glu Ala Met Val
                                                    225                                                                240

                                                                                 810                        840
GAC AAC TGG CGT CCA GCT CAG CCG CTA AAG AAT AGA AAG ATC AAA GCG TCC TTT AAG TAAAACAACCCTGCAGCAGGGGTCCGAAGGCACAAGTGT
Asp Asn Trp Arg Pro Ala Gln Pro Leu Lys Asn Arg Lys Ile Lys Ala Ser Phe Lys
                                                    255

                                880                                                        930                        960
GACCGCCTCTCTGTAGCTAAGCGCAGTTACGGCTGGGTGATTTGGATCCCGACTCGCATCTGGTATTGTAGACCTTTTACCTCTCATCCGTTGTGCTTACTAACAAAAGCAAGACCCAGG

                               1000                                                       1050                       1080
TGTCTCATGTGGTGGCAGCACAGTGGCAGGCCAGTGGTCAACTTAGGGCATCTTTTCTCTGCCACGGCAGCGCAATGCAAAGAGCAGACATGGCCTCTTGCTTCTCTTCACAGCCATAGG

                               1120                                                       1170                       1200
ATAATGAATACTCAGGCCTGTTTGTTAAAATGCTATTTTAAAACCATATGAAGGTAGGATAATTAATTACAAGTCCACATCATGAGACAAACTGAAGTAACTTAGGCAAAACAGGTAAAC

                               1240                                                       1290                       1320
AGTCATAGTTTTGTGATTATAAATGAGATGAATGTTCACCCTTCCAAGATCTTATATTAAAGAAAAAATTTTAAAAAGCTTATATATTTGTAGCAAAGTTATTCTTAAATATGAATTATG

                               1360                                                       1410                       1440
TTATAACTTAGTGACTTTTGATTTCTAGAGGTGTAAATGAGGATGTAAAAATTGATATAGTTGTGATACAGAGTATATTTCCCTTCAGATAACATACCACAACACAATGGATAATGTAT

                               1480
TTTAGATATATTCTCTAATAAAATTGAGAACTCT
```

Fig. 3. Nucleotide sequence of cloned mouse CA II cDNA and the deduced amino acid sequence. The underlined amino acid residues are unique to the mouse CA II sequence as compared to the CA II sequence of human, sheep, ox, rabbit and horse.

Fig. 3. There are 27 nucleotides preceding the initiation codon in the 5′ noncoding region. No homology was found in the 5′ noncoding region with globin mRNA. Examination of the coding region shows that the use of triplets does not differ significantly from the distribution found in globin mRNA (Kafatos et al., 1977; Heindell et al., 1978; Curtis, 1980), except that in CA II the triplet CUG for leucine is not used as extensively as in globin, while for glycine, GGA is used frequently and GGU is not used at all. However, since CA II constitutes <1% of total protein in most mammalian erythrocytes as well as being present at low levels in almost all tissues (Tashian, 1977), it is less likely to be affected by the

relative concentrations of tRNAs which in reticulocytes correlates with the amino acid composition of globin (Smith, 1975). The growing number of processed transcripts characterized by long (> 500 bp) noncoding regions (e.g., dihydrofolate reductase mRNA, porcine β-neo-endorphin mRNA, seal myoglobin gene, human plasminogen activator mRNA, and human hypoxanthine phosphoribosyltransferase mRNA; Setzer et al., 1980; Kakidani et al., 1982; Blancetot et al., 1983; Pennica et al., 1983; Jolly et al., 1983) suggests that such a phenomenon may not be as uncommon as previously thought. However, no significance has as yet been associated with large 3′-nontranslated regions. In CA II mRNA following the termination signal the same reading frame remains open for 324 nucleotides before several stop codons are encountered. It would be of interest to know if read-through of the normal termination signal occurs at all. Located 18 nucleotides from the beginning of the poly(A) region at the 3′ end is the sequence AATAAA which is found in all poly(A)$^+$-containing mRNAs and acts as a signal either for cleavage of the pre-mRNA or as a recognition site for polyadenylation.

(b) Identification of the mouse CA cDNA as a CA II isozyme

From the nucleotide sequence, the complete amino acid sequence of a mouse CA was deduced. One of the most effective methods that can be used to assign an unknown CA to a specific isozyme type (i.e., CA I, CA II or CA III) is to compare the residues in its sequence to those residues which are invariant in all examples of that isozyme sequenced to date, but have a different residue at this position in the other two isozymes. For example, all of the CA II isozymes have His residues at positions 2 and 3 (as does the mouse CA isozyme), whereas the His does not appear at these positions in any of the CA I or CA III isozymes. This information is summarized in Table I. As can be seen, the mouse CA sequence shows only one residue in common with the 21 residues that are unique and invariant in the CA I isozymes, and no residues in common with the 31 residues unique to the CA III isozymes. However, the fact that 16 of the 23 residues that were considered previously to be unique and invariant to the CA II isozymes were identical to the mouse sequence, strongly suggests that the mouse CA sequence is most closely related to the CA II isozymes of mammals. The fact that the mouse CA II has different amino acids in seven positions, which are invariant in all other CA IIs, may be related to the unusual properties of mouse CA II, as discussed below.

TABLE II

Comparison of mouse CA II amino acid sequence with amino acid sequences from selected mammalian CA isozymes

Species [a]	CA isozymes		
	CA I	CA II	CA III
	Residues identical to mouse CA II (%)		
Human	63	81	56
Rabbit	56	77	–
Ox	58	75	52
Horse	58	78	–

[a] See footnote to Table III for references.

TABLE I

Comparison of the putative mouse CA II amino acid sequence with the unique and invariant residues of mammalian CA I, CA II and CA III isozymes

CA isozyme	Number of species sequenced [a]	Number of unique and invariant residues	Number of mouse "CA II" residues identical to the unique and invariant residues of other CA isozymes (%)
CA I	6	21	1 (5)
CA II	7	23 [b]	16 (70)
CA III	2	31	0 (0)

[a] For references to these CA isozyme sequences, see Tashian et al. (1983) and citations therein.

[b] These are at positions 2, 3, 7, 18, 26, 55, 57, 66, 68, 75, 86, 111, 124, 127, 128, 130, 136, 164, 171, 173, 228, 229, 253 in Fig. 3.

(c) Evolutionary considerations

The fact that all three CA isozymes are probably present in all species of amniotes indicates that the duplications which gave rise to the CA isozyme genes occurred more than 300 million years ago (Tashian et al., 1983). Over this evolutionary period, it appears that the CA I and CA II genes have remained closely linked since they have been assigned to a region close to the centromere of chromosome 3 in the mouse (Eicher et al., 1976), and in all likelihood, are also linked on chromosome 8 in humans (Venta et al., 1983). Pairwise comparisons of the mouse CA II sequence with the sequences of the other CA isozymes (Table II) provide an approximation of the evolutionary distances among these isozymes. As can be seen in Table II, the mouse CA II shows a 73–81% homology with the CA II isozymes, a 56–63% homology with the CA I isozymes, and a 52 and 56% homology with the CA III isozymes.

(d) Residues unique to the mouse CA II sequence

The following discussion of the residue positions of the mouse CA II isozyme is based on the assumption that its three-dimensional structure is similar to that of human CA II. For references to the high-resolution molecular structure of human CA II, see Vaara (1974) and Notstrand et al. (1975). When the mouse CA II sequence is compared with the complete CA II sequences of human, sheep, ox, rabbit and horse (for references, see footnote to Table III), 33 residues were found to be unique to the mouse sequence and these are underlined in Fig. 3. All but five of the 33 residues are located on the surface of the human CA II molecule. It is of interest that 15 of the 33 amino acids are invariant in the non-mouse sequences. These residues which are listed in Table III, may contribute to the somewhat unusual properties that have been observed for the mouse CA II isozyme. These properties include the relatively rapid alterations (probably as a consequence of aging) in their high-pressure liquid chromatography elution patterns and electrophoretic patterns compared to the more stable patterns of the CA II isozymes of other mammals (D. Hewett-Emmett and Y.-S.L. Yu, unpublished observations). With the possible exception of Gln-153 which may have

TABLE III

Residues invariant in mammalian CA II sequences but different in the mouse CA II sequence

	Residues and position number [a]													
	7	14	44	57	75	85	86	153	160	171	173	222	223	235	253
Mouse	Ser	Asn	Gln	Ser	Asn	Asp	Ser	Gln	Glu	Arg	Ala	Ser	His	Ala	Lys
Other species [b]	Gly	His	Lys	Arg	Lys	Gly	Thr	Pro	Asp	Ser	Asp	Leu	Lys	Pro	Gln
Secondary structures [c]	–	–	–	β2	–	–	–	$E_1 3_{10}$	$E_1 3_{10}$	β1	–	α-G	α-G	–	–

[a] Based on actual mouse CA II sequence (Fig. 3); in contrast, evolutionary comparisons of CA isozyme sequences are usually based on the numbering of the human CA I sequence (Tashian, 1977; Tashian et al., 1980).

[b] Human (Henderson et al., 1976), sheep (Tanis et al., 1974), ox (Sciaky et al., 1976), rabbit (Ferrell et al., 1978), horse (Jabusch and Deutsch, 1981).

[c] Based on three-dimensional structure of human CA II (Vaara, 1974; Notstrand et al., 1975); β1 and β2 = β-structure segments 1 and 2; $E_1 3_{10}$ = 3_{10} helix E_1; α-G = α-helix G; –, not assigned to a defined structure.

the effect of extending α-helix E at its N-terminus, no obvious changes in secondary structures, aromatic clusters and hydrophobic core residues would appear to result from the amino acid differences at these 15 positions. Of the remaining 18 residues unique to mouse CA II but variable in the CA II isozymes of other mammals, two residues (Tyr-144 and Cys-180) are of interest. The residues at position 144 in the other CA II isozymes are Ile, Leu and Val. Residue 144 is located in the middle of β-structure segment 6 and has been assigned to the hydrophobic core of the human CA II molecule. The presence of Tyr at this position (since it is less hydrophobic than Ile, Val and Leu) might affect the stability of the mouse CA II molecule. The residue at position 180 is Arg, Ser or Gly in the other CA II isozymes and is located in 3_{10}-helix F of human CA II. Since this residue is found on the surface of the human CA II molecule, it is possible that Cys-180 in mouse CA II may form intermolecular dimers with other CA II molecules.

ACKNOWLEDGEMENTS

We thank Dr. G. Rovera and Dr. David Hewett-Emmett for their helpful comments during the preparation of the manuscript. This work was supported in part by U.S. Public Health Service Grant GM-24681 (R.T.) and National Institutes of Health Grants CA-30136, CA-10815 and CA-25875 (P.J.C.).

REFERENCES

Blanchetot, A., Wilson, V., Wood, D. and Jeffrey, A.J.: The seal myoglobin gene: an unusually long globin gene. Nature 301 (1983) 732–734.

Carter, N.D., Hewett-Emmett, D., Jeffery, S. and Tashian, R.E.: Testosterone-induced, sulfonamide-resistant carbonic anhydrase of rat liver is indistinguishable from skeletal muscle carbonic anhydrase III. FEBS Lett. 128 (1981) 114–118.

Curtis, P.J.: Globin mRNA in friend cells: its structure, function and synthesis. Biochim. Biophys. Acta 605 (1980) 347–364.

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

Eicher, E.M., Stern, R.H., Womack, J.E., Davisson, M.-T., Roderick, T.H. and Reynolds, S.C.: Evolution of mammalian carbonic anhydrase loci by tandem duplication; close linkage of Car-1 and Car-2 to the centromere region of chromosome 3 of the mouse. Biochem. Genet. 14 (1976) 651–660.

Ferrell, R.E., Stroup, S.K., Tanis, R.J. and Tashian, R.E.: Amino acid sequence of rabbit carbonic anhydrase II. Biochim. Biophys. Acta 533 (1978) 1–11.

Heindell, H.C., Liu, A., Paddock, G.V., Studnicka, G.M. and Salser, W.A.: The primary sequence of rabbit α-globin mRNA. Cell 15 (1978) 43–54.

Henderson, L.E., Henriksson, D. and Nyman, P.O.: Primary structure of human carbonic anhydrase C. J. Biol. Chem. 251 (1976) 5457–5463.

Ish-Horowicz, D. and Burke, J.F.: Rapid and efficient cosmid cloning. Nucl. Acids Res. 9 (1981) 2989–2998.

Jabusch, J.R. and Deutsch, H.F.: The sequence of the high activity equine erythrocyte carbonic anhydrase. Fed. Proc. 40 (1981) 1677.

Jolly, D.J., Okayama, H., Berg, P., Esty, A.C., Filpula, D., Bohlen, P., Johnson, G.G., Shively, J.E., Hunkapillar, T. and Friedmann, T.: Isolation and characterization of a full-length expressible cDNA for human hypoxanthine phosphoribosyltransferase. Proc. Natl. Acad. Sci. USA 80 (1983) 477–481.

Kafatos, F.C., Efstratiadis, A., Forget, B.G. and Weissman, S.M.: Molecular evolution of human and rabbit β-globin mRNAs. Proc. Natl. Acad. Sci. USA 74 (1977) 5618–5622.

Kakidani, H., Furutani, Y., Takahashi, H., Noda, M., Morioto, Y., Hirose, T., Asai, M., Inayama, S., Nakanishi, S. and Numa, S.: Cloning and sequence analysis of cDNA for porcine β-neo-endorphin/dynorphin precursor. Nature 298 (1982) 245–249.

Lindskog, S.: Carbonic anhydrase, in Eichhorn, G.L. and Marzilli, L.G. (Eds.), Advances in Inorganic Biochemistry, Vol. 4. Elsevier/North-Holland, New York, 1982, pp. 115–170.

Maxam, A.M. and Gilbert, W.: A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA 74 (1977) 560–564.

Messing, J. and Vieira, J.: A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene 19 (1982) 269–276.

Norstrand, B., Vaara, I. and Kannan, K.K.: Structural relationship of human erythrocyte carbonic anhydrases B and C, in Markert, C.L. (Ed.), Isozymes: Molecular Structure, Vol. 1. Academic Press, New York, 1975, pp. 575–599.

Pennica, D., Holmes, W.E., Kohr, W.J., Harkins, R.N., Vehar, G.A., Ward, C.A., Bennett, W.F., Yelverton, E., Seeburg, P.H., Heyneker, H.C. and Goeddel, D.V.: Cloning and expression of human tissue-type plasminogen activator cDNA in *E. coli*. Nature 301 (1983) 214–221.

Pocker, Y. and Sarkanen, S.L.: Carbonic anhydrase: structure, catalytic versatility and inhibition. Adv. Enzymol. 47 (1978) 149–274.

Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.

Sciaky, M., Limozin, N., Filippi-Foveau, D.,Gulian, J.-M. and Laurent-Tabusse, G.: Structure primaire de l'anhydrase carbonique érythrocytaire bovine CI, II. Séquence complète. Biochimie 58(9) (1976) 1071–1082.

Setzer, D.R., McGrogan, M., Nunberg, J.H. and Schimke, R.T.: Size heterogeneity in the 3' end of dihydrofolate reductase messenger RNAs in mouse cells. Cell 22 (1980) 361–370.

Smith, D.W.E.: Reticulocyte transfer RNA and hemoglobin synthesis. Science (1975) 529–533.

Summers, J.: Physical map of polyoma DNA fragments produced by cleavage with a restriction enzyme from *Haemophilus aegyptius* endonuclease R. *Hae*III. J. Virol. 15 (1975) 946–953.

Tanis, R.J., Ferrell, R.E. and Tashian, R.E.: Amino acid sequence of sheep carbonic anhydrase C. Biochim. Biophys. Acta 371 (1974) 534–548.

Tashian, R.E.: Evolution and regulation of the carbonic anhydrase isozymes, in Rattazzi, M.C., Scandalios, J.G. and Whitt, G.S. (Eds.), Isozymes, Current Topics in Biological and Medical Research, Vol. 2. Liss, New York, 1977, pp. 21–62.

Tashian, R.E., Hewett-Emmett, D. and Goodman, M.: On the evolution and genetics of carbonic anhydrase I, II, and III, in Rattazzi, M.C., Scandalios, J.G. and Whitt, G.S. (Eds.), Isozymes, Current Topics in Biological and Medical Research, Vol. 7. Liss, New York, 1983, pp. 79–100.

Tashian, R.E., Hewett-Emmett, D., Stroup, S.K., Goodman, M. and Yu, Y.-S.L.: Evolution of structure and function in the carbonic anhydrase isozymes of mammals, in Bauer, C., Gros, G. and Bartels, H. (Eds.), Biophysics and Physiology of Carbon Dioxide. Springer-Verlag, Berlin, 1980, pp. 165–176.

Vaara, I.: The molecular structure of human carbonic anhydrase, form C and inhibitor complexes. Ph.D. Dissertation, University of Uppsala, Sweden, 1974.

Venta, P.J., Shows, T.B., Curtis, P.J. and Tashian, R.E.: The polymorphic gene for human carbonic anhydrase II: a new molecular disease marker located on chromosome 8. Proc. Natl. Acad. Sci. USA 80 (1983) 4437–4440.

Communicated by A.-M. Skalka.