Gene, 126 (1993) 291–292

GENE 07007

The deduced amino acid sequence of human carbonic anhydrase-related protein (CARP) is 98% identical to the mouse homologue*

(PCR; polymerase chain reaction; evolution; isozymes)

Laura A. Skaggs, Nils C.H. Bergenhem, Patrick J. Venta** and Richard E. Tashian

Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109-0618, USA

Received by J.L. Slightom: 23 September 1992; Accepted: 17 November 1992; Received at publishers: 29 December 1992

SUMMARY

A recently reported mRNA, encoding 'carbonic anhydrase-related polypeptide' (CARP) from the Purkinje cells of mouse cerebellum, was shown to have a 30–40% deduced amino acid sequence identity with the carbonic anhydrases (CA) of mammals. In order to compare the mouse and human *CARP* sequences, we used the polymerase chain reaction (PCR) to amplify human *CARP* sequences from several cDNA libraries (salivary gland, testis and placenta). The sequence has an 89.3% sequence identity with mouse *CARP* at the nucleotide level and 97.9% at the amino acid level. This extremely high evolutionary conservation suggests an important function for the *CARP* gene product.

The genes that encode the seven mammalian carbonic anhydrase (CA) isozymes (CA I—CA VII) range in their expression from certain cells of virtually all tissues (e.g., *CA II*) to their expression in a single tissue, e.g., *CA VI* in salivary glands (cf. Tashian, 1989; 1992). Recently, a new member of the *CA* family was discovered in a mouse brain cDNA library and found to be expressed in the Purkinje cells of the cerebellum (Kato, 1990). In order to determine whether a similar CA-related protein (CARP) is found in humans, we used PCR to amplify human *CARP* (*HCARP*) from several human cDNA libraries (i.e., testis, salivary gland and placenta). We sequenced several amplified fragments (see Methods in Fig. 1 legend) containing the 5′ part of *HCARP* from both placenta and salivary gland clones in cDNA libraries. All clones showed cloning artifacts in the 5′ ends of the cDNA. When the human and mouse *CARP* sequences were compared, a non-homologous sequence was found to be spliced into the *HCARP*. This sequence was the same in all clones, but the splicing positions differed. Because of these artifacts, the clone with the most extensive sequence similarity with mouse *CARP* (*MCARP*) in the 5′ end was still found to lack two nt in the putative start codon. However, the unusually high degree of identity between the human and mouse *CARP* sequences (Fig. 1) strongly indicates that the AUG start codon is at the same position in both species. It is also possible that the mouse has the same TAG stop codon as in *HCARP*, and that the deletion of T^{867} in *MCARP* is due to a sequencing error or polymorphic variability. It is unlikely that the *HCARP* sequence is incorrect because *CARP* cDNAs were sequenced from several cDNA libraries.

When the sequences of genes encoding the CA I, II and III isozymes and CARP are compared between humans and mice (Table I), the *CARP* sequences are the

Correspondence to: Dr. R.E. Tashian, Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109-0618, USA. Tel. (313)764-1359; Fax (313)763-3784; e-mail: nils.bergenhem@-um.cc.umich.edu

*On request, the authors will supply detailed experimental evidence for the conclusions reached in this brief note.

**Present address: Small Animal Clinical Sciences, College of Veterinary Medicine, Michigan State University, East Lansing, MI 48824-1314 (USA) Tel. (517)336-2515.

Abbreviations: aa, amino acid(s); bp, base pair(s); CA, carbonic anhydrase; *CA*, gene encoding CA; CARP, CA-related protein; *CARP*, gene encoding CARP; kb, kilobase(s) or 1000 bp; H, human; M, mouse; nt, nucleotide(s); PCR, polymerase chain reaction.

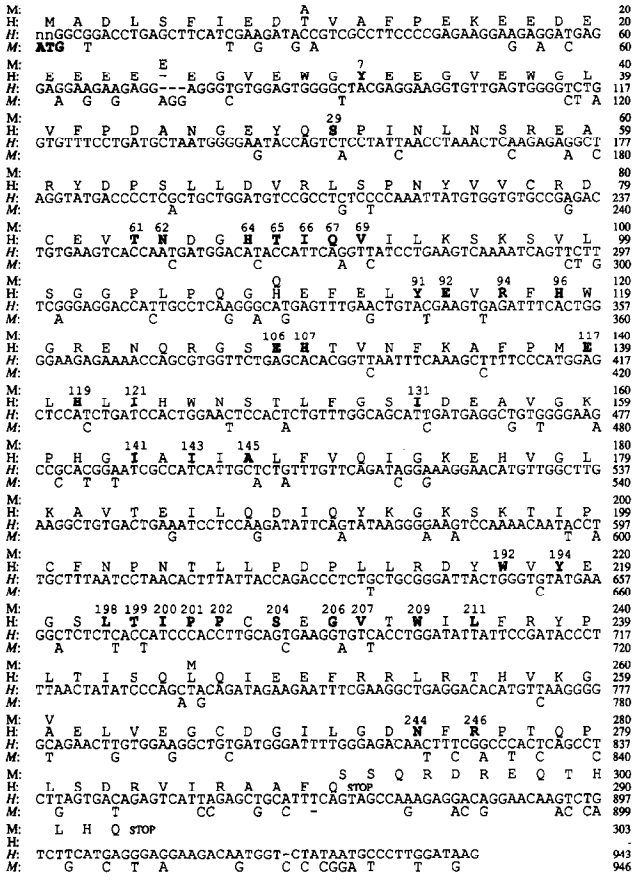

Fig. 1. The nt sequence of *HCARP* and the deduced aa sequence. The complete nt sequence for *HCARP* is presented (H), with only the differences between the *MCARP* and *HCARP* nt sequences shown. The deduced aa sequence of HCARP is given in one letter code (aligned with second nt of each codon). The nt are is numbered from the start codon. The aa that are identified as those corresponding to residues found in the active site of catalytically active carbonic anhydrase isozymes are typed in boldface, and the aa number (CA I numbering; cf. Hewett-Emmett and Tashian, 1991) is shown above the residues. Deletions are indicated by dashes. Residues that differ from those that are invariant in all CA isozymes sequenced to date are: 61 N/T, 92 Q/E and 94 H/R (cf. Tashian, 1992). GenBank accession No. for human CARP: L04656. **Methods:** PCR primers were made to match the *MCARP* sequence (Kato, 1990). These primers, and primers matching the λgt11-vector sequence, flanking the inserted cDNAs, were used to amplify fragments of *HCARP* cDNA. The amplified DNA was gel-purified and directly sequenced as described (Bachmann et al., 1990; Bergenhem et al., 1992). In order to be able to sequence as close to the amplification primers as possible, the 3′, and 5′ ends of the cDNA were amplified and cloned into a T-tailed pSK vector (Marchuck et al., 1991) before being sequenced with pSK primers.

most conserved, exhibiting remarkably high identities of 89.3% and 97.9% at the nt and deduced aa sequence levels, respectively. The highly conserved sequences of these genes indicate that there has been strong selective pressure to maintain an important function for this newly identified protein. In this respect, the acidic nature of the first 47 aa, i.e., 15–16 Glu, and four Asp residues (Fig. 1), suggests that this region may play a significant role in the function of CARP.

Because CARP lacks a His^{94} (see Fig. 1), which is one of the three His ligands (others are His^{96} and His^{119}) that bind the Zn ion which is essential for CA activity, it is possible that CARP has either no catalytic activity, or a completely new activity. If CARP has CA activity, it would be designated CA VIII.

This work was supported by NIH Grant GM24681.

TABLE I

Percent sequence identity between mouse and human cDNA and encoded CA isozymes and CARP[a]

	cDNA	Deduced aa
CA I	82.8	80.8
CA II	82.6	81.9
CA III	88.3	91.2
CARP	89.3	97.9

[a]The sequence of *MCARP* was obtained from Kato (1990). The other sequences used in the comparison are from Hewett-Emmett and Tashian (1991). Only the coding regions are compared; for *CARP* the comparison is limited to the coding region of the human sequence (see Fig. 1).

REFERENCES

Bachmann, B., Luke, W. and Hunsmann, G.: Improvement of PCR amplified DNA sequencing with the aid of detergents. Nucleic Acids Res. 18 (1990) 1309.

Bergenhem, N.C.H., Venta, P.J., Hopkins, P.J. and Tashian, R.E.: Variation in coding exons of two electrophoretic alleles at the pigtail macaque carbonic anhydrase I locus as determined by direct, double-stranded sequencing of polymerase chain reaction (PCR) products. Biochem. Genet. 30 (1992) 279–287.

Hewett-Emmett, D. and Tashian, R.E.: Structure and evolution of the carbonic anhydrase multigene family. In Dodgson, S.J., Tashian, R.E., Gros, G. and Carter, N.D. (Eds.), The Carbonic Anhydrases: Cellular Physiology and Molecular Genetics. Plenum Press, New York, 1991, pp. 15–32.

Marchuk, D., Drumm, M., Saulino, A. and Collins, F.S.: Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res. 19 (1991) 1154.

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

Tashian, R.E.: The carbonic anhydrases: widening perspective on their evolution, expression and function. BioEssays 10 (1989) 186–192.

Tashian, R.E.: Genetics of the mammalian carbonic anhydrases. Adv. Genet. 30 (1992) 321–356.