Biochemical Genetics, Vol. 30, Nos. 5/6, 1992

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

Nils C. H. Bergenhem,[1] Patrick J. Venta,[1,2] Penelope J. Hopkins,[1,3] and Richard E. Tashian[1,4]

Received 12 Nov. 1991—Final 30 Jan. 1992

Two, electrophoretically distinct, forms of carbonic anhydrase I (CA Ia and CA Ib) are found at high polymorphic frequencies in red cells of natural populations of pigtail macaques, Macaca nemestrina, *from southeast Asia. By use of the polymerase chain reaction, exons of the CA I gene were amplified from homozygous* (a/a, b/b) *and heterozygous* (a/b) *animals. Direct sequencing of the amplified DNA from four animals revealed differences between the a and the b electrophoretic alleles ranging from three to six nucleotides, and from one to three differences within each allele. These results indicate a greater genetic variability at the CA I locus in this macaque species than previously realized.*

KEY WORDS: *Macaca nemestrina;* carbonic anhydrase I locus; polymerase chain reaction; DNA sequencing; allelic variation.

We thank the personnel of the University of Washington, Regional Primate Research Center, Seattle, for sending us blood and tissue samples from pigtail macaques (supported by NIH Grant RR 00166). Our work was supported by NIH Grant GM 24681.

[1] Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan 48109-0618.

[2] Small Animal Clinical Sciences, College of Veterinary Medicine, Michigan State University, East Lansing, Michigan 48824-1314.

[3] Present address: Department of Molecular and Experimental Medicine, Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, California 92037.

[4] To whom correspondence should be addressed.

0006-2928/92/0600-0279$06.50/0 © 1992 Plenum Publishing Corporation

INTRODUCTION

Alleles are frequently typed on the basis of the characteristic electrophoretic migration patterns of their protein products. However, little is known about the extent to which nucleotide differences are present in such electrophoretic alleles other than the one primarily responsible for its characteristic electrophoretic type or electromorph. Previous approaches to this problem include (a) sequencing the electrophoretic alleles of horse carbonic anhydrase I, CA I (Jabusch *et al.*, 1980), (b) subtyping electromorphs of phosphoglucomutase in human populations by isoelectric focusing (Takahashi *et al.*, 1982), and (c) detecting nucleotide variability in electrophoretic and nonelectrophoretic alleles by thermolability testing and restriction enzyme analyses (Neel, 1984).

We report here comparative sequence data for two electrophoretic alleles at the carbonic anhydrase I locus of the pigtail macaque, *Macaca nemestrina,* the complete gene structure of which has recently been determined by one of us (P.J.H.) (Nicewander, 1990). We chose to study this species of macaque because of the considerable genetic variability we had previously observed in its red-cell carbonic anhydrase (CA) isozymes, CA I and CA II (Tashian *et al.*, 1971). Four electrophoretic variants of CA I (CA Ia, Ib, Ic, and Id) are found at polymorphic frequencies (i.e., >0.02) in natural populations from Thailand and Malaysia (Table I, Fig. 1). Further genetic variability is also present in the form of a red cell-specific CA I-deficiency polymorphism, where the frequencies for the deficiency gene range from 0.55 in Thailand and 0.57 in Malaysia to 0.70 in Sumatra (Tashian *et al.*, 1971; Darga *et al.*, 1975; Tashian and Carter, 1976). The CA I-deficient phenotype has only trace levels of CA Ia and/or CA Ib in the red cells but normal levels in the colon and probably other nonerythroid tissues where CA I is normally expressed (DeSimone *et al.*, 1973; Tashian *et al.*, 1990).

MATERIALS AND METHODS

DNA Preparation

DNA was prepared from autopsied pigtail macaque tissue kindly provided by the University of Washington Regional Primate Research Center, Seattle.

To prepare DNA suitable for polymerase chain reaction (PCR) studies, 100 mg of tissue (muscle or liver) was cut in pieces with a razor blade and incubated overnight in 1.0 ml of 6.0 *M* guanidine–HC1, containing 0.1 *M* sodium acetate. The supernatant was withdrawn and the DNA precipitated by the addition of an equal volume of 4.0 *M* ammonium acetate and 2 vol of

Table I. Allelic Frequencies for Electrophoretic Variants of CA I in *Macaca nemestrina*[a]

Subspecies	Location	No.	Type *a*	*b*	*c*	*d*
M. n. leonia	Thailand	63	0.109	0.829	0	0.072
M. n. nemestrina	W. Malaysia	165	0.556	0.416	0.028	0

[a]Data from Tashian *et al.* (1971) and Fooden (1975).

ethanol. After pelleting the DNA in a microtube by centrifugation, the pellet was washed with 95% ethanol and briefly dried. The DNA was then dissolved in 100 μl of 1× TE (10 m*M* Tris–HCl, 1 m*M* EDTA, pH 8) by placing the tube in a boiling water bath for 5 min.

PCR Amplification

Seven sets of primers (Table II) were used at a concentration of 0.1 μg/100 μl to amplify the seven coding exons of the CA I gene according to the protocol provided with the Perkin Elmer Cetus PCR kit. Thermal cycling was done in a Bellco DNA PaCeR with the following parameters: 94°C, 40 sec; 54°C, 1 min; 71°C, 1 min; 40 cycles.

Sequencing of PCR-Amplified DNA

After amplification, DNA fragments were purified from excess amplification primers, and any possible nonspecific fragments by electrophoresis in a 4%, 19:1 acrylamide:bisacrylamide gel using 1× TBE (90 m*M* Tris, 90 m*M* boric acid, 2 m*M* EDTA). The DNA fragments were visualized by ethidium

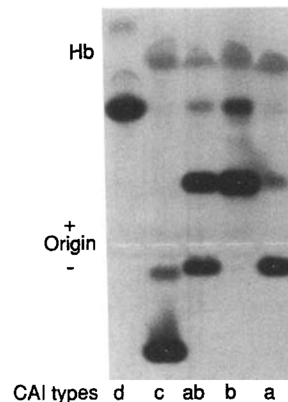

Fig. 1. Electrophoretic patterns of red-cell CA I in the pigtail macaque, *Macaca nemestrina.* Double dots indicate major CA I isozyme band; single dots, the secondary isozymes. See Tashian *et al.* (1971) for procedural details.

Table II. The Nucleotide Sequence of the Primers Used for PCR Amplification and Sequencing[a]

Exon		
1a	GGAATTCAATCCACACCCCAAC-CACTTC	GGAATTCACAGCTCTGAATGAGA-GAAGG
1	GTTGGAATCTTGAGTGTACAAG	GGGATCCGCAGACAGTTCAACAA-TTAACC
2	CAAACAGGTAACTACACTCCT	AATGGGTGTCATGTTTCTCG
3	GGAATTCGCAAAGATAAGCTAG-AGTTTG	GGGATCCAGGGTAATTATCTCTCA-CTTAC
4	GGAATTCCACTGGATAAAGGTT-CACATA	CCTTCTATTTTGAGGTCTAATTGG
5	GCAGTGTTTGATTGACAATAATC	GGAATTCACCCCCAGTTTTAATAC-TTCA
6	GGAATTCAATGACTCTTAGCTAA-AATCTC	AATATTCCTGCTACTATATTCCC
7	TCAGTGCGTTAGTAATCCTGTAA	AAAGCTTGGGCTGTGTTCTTGAG-GAAGG

[a]For each exon the upstream primer is given first. The primer sequences are written in the 5′-to-3′ direction.

bromide staining and placement of the gel on a 360-nm ultraviolet light box. The sequences obtained when a 300-nm source was used were unreadable, presumably because of pyrimidine dimer formation. The bands were then cut from the gel and soaked in 0.5 ml of 0.5 *M* ammonium acetate for 10–15 hr. After removal of the gel piece, and addition of 1 ml of 95% ethanol, the DNA was pelleted, washed once with 95% ethanol, briefly dried, and dissolved in 20 μl of TE. This method of purification removes primers and extraneous PCR products at the same time, allowing multiple samples to be processed with a minimum of labor and expense. The use of ammonium acetate is necessary to prevent the borate from precipitating. Seven microliters of this DNA solution was mixed with 1 μl 5% Nonidet P-40 (Bachman *et al.,* 1990), 1 μl (1 μg/μl) primer, and 2 μl sequencing buffer [United States Biochemical Corp (USBC)]. The same primers that were used for amplification were used for sequencing. Although in the original protocol NP-40 was added to each step, we have found it necessary only in this first step. The primer–template annealing was performed by first boiling the samples contained in microtubes for 3–5 min, followed by rapidly transferring the tubes to a dry ice/ethanol bath for 30–60 sec (Casanova *et al.,* 1990) and then placing the tubes on wet ice for 10 min. Snap-cooling of the sample appears to be critical for obtaining strong readable sequencing ladders. The rest of the sequencing reaction was performed according to the protocol supplied with the Sequenase sequencing kit (USBC). Both strands were sequenced in all cases.

Phenotyping

Starch gel electrophoresis of hemolysates, followed by staining for CA activity, was used to type the electrophoretic alleles (Tashian *et al.*, 1971). The electrophoretic phenotype of the red cell CA I-deficient animal was deduced from the sequence as discussed under Results. A *Bsu*36 I restriction site in a PCR product upstream of the noncoding exon 1a (Table II), which is present only in red cell CA I-deficient animals, was used to show that the *a*/*a* and *b*/*b* animals were not heterozygous for the CA I deficiency (data not shown).

RESULTS AND DISCUSSION

The PCR technique we used for direct sequencing of double-stranded DNA is relatively fast, and the sequences of the exons for both electrophoretic alleles can be read at the same time (Fig. 2). We have found this method to be highly reliable, with little need for optimization from primer to primer. To determine if the coding region of the CA I gene from a red cell CA I-deficient animal differs from that of a normal animal, all seven coding exons for both alleles, with flanking regions, were sequenced from one red cell CA I-deficient animal (No. 1 in Table III). From this sequence, the sequence of a cloned *a*/*b* animal (No. 5 in Table III), and the relative electrophoretic mobilities of the four electrophoretic alleles (Fig. 1), it can be deduced that the animal in question must be an *a*/*b* type. As can be seen, the change responsible for the electrophoretic difference between the *a* and the *b* types is located at amino acid position 242 (nucleotide 727), which is Gln (CAG) in CA Ia and Glu (GAG) in CA Ib. The only other amino acid difference is at position 241 (nucleotide 725), which is Ser (AGT) for CA Ia and Ile (ATT) for CA Ib. In addition, four "silent" nucleotide differences were detected at positions 291, 309, 717, and 750. The locations of the two differences in exon 3 and the four in exon 7 are shown in Fig. 3.

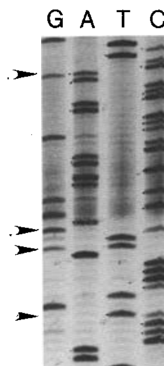

Fig. 2. Part of the sequence of exon 7 from the CA I gene of *Macaca nemestrina*. The sequencing was performed as described under Materials and Methods. The arrows indicate positions where the *a* and *b* alleles differ. Two bands at these positions were confirmed in the sequence of the complementary DNA strand.

Table III. Nucleotide Changes Found in the *a* and *b* Electrophoretic Alleles at the Carbonic Anhydrase I Locus of the Pigtail Macaque, *Macaca Nemestrina*

Animal No.	Type	Exons sequenced	Nucleotide position[a] 291	309	717	725	727	750
1	ab[b]	1–7	**C/T**	**A/G**	**C/T**	**T/G**[c]	**C/G**[d]	**G/A**
2	ab	3, 7	T/T	G/G	T/T	**T/G**	**C/G**	**G/A**
3	bb	3, 7	**C/T**	G/G	T/T	T/T	G/G	A/A
4	aa	3, 7	C/C	**A/G**	**C/T**	G/G	C/C	G/G
5	ab[e]	1–7	T	G	T	G	C	G
Other species[f]								
Human CA I			C	G	T	T	C	A
Mouse CA I			C	C	A	T	C	A
Rabbit CA I			T	C	T	T	C	G

[a]Numbering from translational start site.
[b]Typed from a red-cell CA I-deficient animal.
[c]T = Ile-241 (ATT), G = Ser-241 (AGT).
[d]C = Gln-242 (CAG), G = Glu-242 (GAG).
[e]Cloned CA I gene; only one allele was sequenced (Nicewander, 1990).
[f]The nucleotides at these positions from single human, mouse, and rabbit CA I sequences are given for comparison (Lowe *et al.,* 1990; Fraser *et al.,* 1986; Konialis *et al.,* 1985)

In order to determine if this variability was limited to red-cell CA I-deficient animals, the sequence of exons 3 and 7 for both alleles was determined in three pigtail macaques: one heterozygote (*a*/*b*) and two homozygotes (*a*/*a* and *b*/*b*). The nucleotides at the positions where differences were found are listed in Table III (Nos. 2–4). In addition, the sequence of one allele of an *a*/*b* animal (No. 6, Table III) was examined by sequencing the cloned genomic DNA (Nicewander, 1990). As can be seen, the variability of the *a* or *b* alleles of the red-cell CA I-deficient animal (a difference of six nucleotides) is present in the *a* or *b* alleles of the normal animals as well (Nos. 2–4, Table III). In addition, the *a* and *b* alleles are found to differ by up to three nucleotides within each electrophoretic allele. It is of interest that the nucleotide variation seen in this study is not randomly distributed but found in two clusters. The two positions in exon 3 (nucleotides 291–309) are within 19 bp of each other, and the four in exon 7 (nucleotides 717–750) are within 34 bp (Fig. 3). Electrophoretic alleles at the CA I locus, seemingly at polymorphic frequencies, have also been reported for the orangutan (Tashian, 1965) and the horse (Sandberg, 1968; Jabusch *et al.,* 1980). In the study by Jabusch *et al.* (1980), the amino acid sequences of five electromorphs of CA I were analyzed. One variant had five amino acid substitutions, and the other four showed one, two, two, and three changes, respectively. These 11 changes are more randomly distributed than those in the pigtail macaque. Three changes in one cluster that spans exons 6

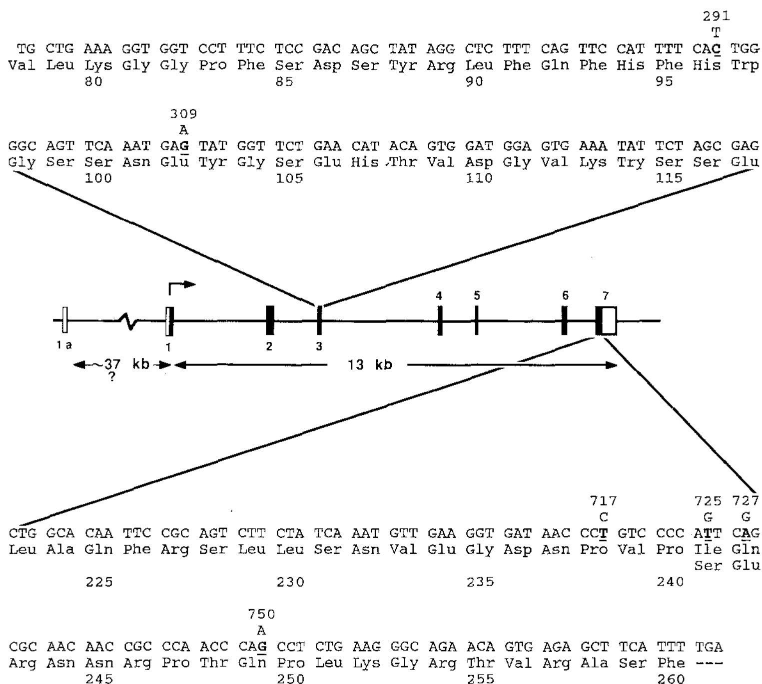

Fig. 3. Gene structure of CA I from *Macaca nemestrina* (Nicewander, 1990) and complete sequences of exons 3 and 7 showing positions of nucleotide variants. The translated regions of exons are filled in. The size of infron la (~ 37 kb) is based on the human CAI gene (Lowe *et al.,* 1990).

and 7 are within 34 bp, and three changes in exon 3 are only 6 bp apart. If the PCR techniques used in the present study were applied to the horse study, one would expect to see an even greater number of variant nucleotides. Polymorphic variability in noncoding regions of human DNA has been estimated to be about one in 400 nucleotides (Neel, 1984). We would, however, expect polymorphic variability in coding regions to be lower due to selective pressure. The macaque and horse CA I genes have 6 and 11 polymorphic sites, respectively, of the 780 nucleotides in the coding regions. It is of further interest that CA I variants in humans are found mainly in exons 3 and 7. Of the 10 mutations of CA I known in humans, 3 are in exon 3 at residue positions 86, 100, and 102, and 5 are in exon 7 at positions 225, 236, 246, 253, and 255 (Tashian *et al.,* 1983; Wagner *et al.,* 1991). Thus, in a total of 28 variants found in the coding regions of the CA I genes in the pigtail macaque, human, and horse, 10 were found in exon 3 (117 bp) and 10 in exon 7 (111 bp). The remaining eight variants ranged from none in exon 4 (93 bp) to three in exon 6 (153 bp). It is possible that greater mutability or lower selection in these regions might account for this distribution. An

example of the clustering of seven rare point mutations (resulting in charge changes) has been reported in the human serum albumin gene within an 87-bp sequence coding for amino acid residues 354–382 (Arai *et al.,* 1989). However, these changes are found in separate alleles. Clustering of changes also has been seen over long evolutionary periods (Kafatos *et al.,* 1977), but it is generally believed that these mutations occur sequentially after each has been fixed in the population, rather than occurring simultaneously in polymorphic proportions before fixation.

Some insight concerning the evolution of nucleotide changes at the variant sites can be obtained by comparing the same six variable nucleotide positions in other CA I genes. These are shown in Table III as a C/T transition at positions 291 and 717, a G/A transition at positions 309 and 750, a G/T transversion at position 725, and a C/G transversion at position 727. It is also noteworthy that two of the positions are in the codons for His-96 and Gln-249, which are invariant in all carbonic anhydrases that have been sequenced to date (Hewett-Emmett and Tashian, 1991). The other four are in codons for Glu-102, Pro-238, Ile/Ser-241, and Gln/Glu-242. Because Ser and Glu have not been observed at positions 241 and 242, respectively, in any other carbonic anhydrase (Hewett-Emmett and Tashian, 1991), the nucleotide changes from the ancestral sequence are probably T → G at 725 and C → G at 727.

The results presented here, although limited, provide further evidence of nucleotide variability at certain genetic loci. Obviously, many more sequences of the *a* and *b* electrophoretic types, and also the *c* and *d* types, must be analyzed, as well as the CA I genes of other species of macaque monkeys, in order to achieve a fuller understanding of the extent of nucleotide variability within various alleles and how they may have evolved. Whether the polymorphic variability described in this study is unique to the CA I loci of mammals, or whether it is a relatively common feature of other mammalian genes, will likely soon come to light given the speed and ease of direct sequencing of PCR products.

REFERENCES

Arai, K., Madison, J., Huss, K., Ishioka, N., Satoh, C., Fujita, M., Neel, J. V., Sakurabayashi, I., and Putnam, F. W. (1989). Point substitutions in Japanese alloalbumins. *Proc. Natl. Acad. Sci. USA* **86**:6092.

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

Casanova, J.-L., Pannetier, C., Jaulin, C., and Kourilsky, P. (1990). Optimal conditions for directly sequencing double-stranded PCR products with Sequenace. *Nucleic Acids Res.* **18**:4028.

Darga, L. L., Goodman, M., Weiss, M. L., Moore, G. W., Prychodko, W., Dene, H., Tashian, R., and Koen, A. (1975). Molecular systematics and clinal variation in macaques. In Markert, C. L. (ed.), *Isozymes: Genetics and Evolution, Vol. 4,* Academic Press, New York, pp. 797–812.

DeSimone, J., Magid, E., and Tashian, R. E. (1973). Genetic variation in the carbonic anhydrase isozymes of macaque monkeys. II. Inheritance of red cell carbonic anhydrase levels in different carbonic anhydrase I genotypes of the pig-tailed macaque, *Macaca nemestrina. Biochem. Genet.* **8**:165.

Fooden, J. (1975). Taxonomy and evolution of liontail and pigtail macaques. *Fieldiana Zool.* **67**:1.

Fraser, P. J., and Curtis, P. J. (1986). Molecular evolution of the carbonic anhydrase genes: Calculation of divergence time for mouse carbonic anhydrase II and III. *J. Mol. Evol.* **23**:294.

Hewett-Emmett, D., and Tashian, R. E. (1991). Structure and evolutionary origins 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, pp. 15–32.

Jabusch, J. R., Bray, R. P., and Deutsch, H. F. (1980). Sequence of the low activity equine erythrocyte carbonic anhydrase and delineation of the amino acid substitutions in various polymorphic forms. *J. Biol. Chem.* **255**:9196.

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

Konialis, C. P., Barlow, J. H., and Butterworth, P. H. W. (1985). Cloned cDNA for rabbit erythrocyte carbonic anhydrase I: A novel erythrocyte specific probe to study development in erythroid tissues. *Proc. Natl. Acad. Sci. USA* **82**:663.

Lowe, N., Hugh, J. M. B., Barlow, J. H., Sowden, J. C., Edwards, M., and Butterworth, P. H. W. (1990). Structure and methylation patterns of the gene encoding human carbonic anhydrase I. *Gene* **93**:277.

Neel, J. V. (1984). A revised estimate of the amount of genetic variation in human proteins. Implications for the distribution of DNA polymorphisms. *Am. J. Hum. Genet.* **36**:1135.

Nicewander, P. H. (1990). *Sequence and Organization of a* Macaca nemestrina *Carbonic Anhydrase I Gene,* Ph.D. thesis, University of Michigan, Ann Arbor.

Sandberg, K. (1968). Genetic polymorphism in carbonic anhydrase from horse erythrocytes. *Hereditas* **60**:411.

Takahashi, N., Neel, J. V., Satoh, C., Nishizaki, J., and Masunari, N. (1982). A phylogeny for the principal alleles of the human phosphoglucomutase-1 locus. *Proc. Natl. Acad. Sci. USA* **79**:6636.

Tashian, R. E. (1965). Genetic variation and evolution of the carboxylic esterases and carbonic anhydrases of primate erythrocytes. *Am. J. Hum. Genet.* **17**:257.

Tashian, R. E., and Carter, N. D. (1976). Biochemical genetics of carbonic anhydrase. In Harris, H., and Hirschhorn, K. (eds.), *Advances in Human Genetics,* Plenum Press, New York, pp. 1–56.

Tashian, R. E., Goodman, M., Headings, V. E., DeSimone, J., and Ward, R. H. (1971). Genetic variation and evolution in the red cell carbonic anhydrase isozymes of macaque monkeys. *Biochem. Genet.* **5**:183.

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

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.

Wagner, L. E., Venta, P. J., and Tashian, R. E. (1991). A human carbonic anhydrase I deficiency appears to be caused by a destabilizing amino acid substitution (246-Arg → His). *Isozyme Bull.* **24**:35.