#### **FEBS 14508**

# Molecular characterization of a class IV human alcohol dehydrogenase gene (ADH7)

# Shozo Yokoyama\*, Yoshinori Matsuo, Rebecca Ramsbotham, Ruth Yokoyama

Department of Biology, Syracuse University, 130 College Place, Syracuse, NY 13244, USA

Received 4 August 1994

Abstract Class IV alcohol dehydrogenase (ADH) is a form preferentially expressed in stomach. We report here the isolation and sequence determination of a novel human ADH gene (ADH7). Phylogenetic analysis strongly suggests that ADH7 is a functional class IV ADH gene.

Key words: Alcohol dehydrogenase; DNA; Genomic; Molecular evolution; Human

#### 1. Introduction

Human alcohol dehydrogenases (ADHs) are dimeric metalloenzymes and had been grouped early into three classes, I-III, according to different electrophoretic properties and different substrate specificities (e.g., see [1]). These ADHs have been conserved among diverse species such as mammals, plants and yeasts [2,3]. Additional ADH classes IV [4] and V [5] have since then discovered. Phylogenetic analysis of available amino acid sequences has shown that class I and V are most closely related, their common ancestor diverged from the ancestor of class II, and their common ancestor diverged from the ancestor of class III before that ([6]; see also [7]).

At present, mammalian class IV and V ADHs are relatively poorly understood. The class V protein has been deduced from genomic and cDNA clones [5], but the protein has not been isolated from natural sources. The class IV protein appears to be expressed mainly in the stomach mucosa [4, 8]. The molecular structure of a class IV protein from rat has been characterized only recently [8], but that for human is not known. We report here the cloning and sequencing of a novel human ADH gene (ADH7). Phylogenetic analysis of the amino acid sequence deduced from ADH7 and other vertebrate ADHs strongly suggests that ADH7 is a functional class IV ADH gene.

#### 2. Materials and methods

## 2.1. \( \lambda \) clone characterization and DNA sequencing

Two sets of human genomic DNA libraries were constructed using genomic DNA partially digested with Sau3A or EcoRI and by ligating it with  $\lambda EMBL3$  or  $\lambda EMBL4$  DNA, respectively (e.g. see [9]). To characterize a novel ADH gene, ADH7, two overlapping clones  $\lambda 1.1$  and  $\lambda 18$  have been obtained.  $\lambda 18$  was isolated from the  $\lambda EMBL4$  library using bADH cDNA probe [10].  $\lambda 1.1$  was isolated from the  $\lambda EMBL3$  library using a 0.5 kb fragment encompassing exons 3 and 4 of  $\lambda 18$  as a probe.  $\lambda 1.1$  contained exons 2–8, while  $\lambda 18$  contained exons 1–5 (Fig. 1). The two clones were verified to be overlapping by restriction enzyme analysis and sequencing. By amplifying genomic DNA using a primer 100 bp from the end of  $\lambda 1.1$  and a degenerate class 1 exon 9 primer, it became clear that exon 9 is located approximately 1.2 kb from the end of  $\lambda 1.1$ . Furthermore, genomic Southern hybridization indicated that exon 9 was contained on a 2.4 kb HindIII fragment (data not shown). The sequence of exon 9 was obtained by using the

\*Corresponding author. Fax: (1) (315) 443 2012. INTERNET: syokoyam@suvm.syr.edu BITNET: SYOKOYAM@SUVM.BITNET

inverse PCR method [11]. To perform inverse PCR, 2 oligonucleotides (int8A and int8B) were designed from intron 8 sequence to flank and extend into the unknown genomic region upon circularization of the HindIII-digested genomic DNA (Fig. 1). Int8A (5'-GGCCCAAT-GAGTGAAAGTCATTTC-3') corresponds to sequence 100 bp from the end of  $\lambda 1.1$  and was oriented to amplify towards exon 8. Int8B (5'-CTCAGATACCCTTATCAGAGTTGCG-3') corresponds to sequence about 200 bp upstream from exon 9 and was oriented to amplify towards exon 9. Ligation of 1 µg HindIII-digested genomic DNA at a concentration of 2 µg/ml was done to promote monomeric circularization. The ligated product was amplified using int8A and int8B for 30 cycles (94°C for 45 s, 55°C for 90 s, 72°C for 2 min). The expected 1.2 kb product was gel isolated, blunt-ended by Klenow treatment and subcloned into EcoRV-digested Bluescript. Exon 9 sequence was obtained using int8B primer. Subcloning and double-stranded sequencing of the  $\lambda 1.1$  and  $\lambda 18$  were conducted as previously described [9,12,13].

#### 2.2. Data analysis

The amino acid sequence, Hsa7, deduced from ADH7 was compared to other vertebrate ADHs. Vertebrate ADHs considered were class I ADH $\alpha$  (HsaA),  $\beta$  (HsaB),  $\gamma$  (HsaC) from human and those from baboon (PhaB), rhesus monkey (MacA), mouse (MmuA), rat (RnoA), quail (Cja), chicken (Gga), and frog (Rpe), and ADH-E (EcaE) and ADH-S (EcaS) from horse, one class II ADH from human (HsaP), class III ADHs from human (HsaX), mouse (MmuX), rat (RnoX) and horse (EcaX), and class V ADH (Hsa6) deduced from human ADH6 ([6] and references therein). Furthermore, ADH1 (PmaA) and ADH2 (PmaIV) from deer mouse (Peromyscus maniculatus) [14] and class IV ADH from rat (RnoIV) [8] were also included.

To construct a rooted phylogenetic tree for these ADHs, ADH1 and ADH2 from maize (see [6] and references therein) were used as an outgroup. These amino acid sequences can be aligned easily (see [6]). The number (K) of amino acid substitutions per site for two polypeptides was estimated by  $K = -\ln (1-p-p^2/5)$ , where p is the proportion of different amino acids between the two sequences [15]. Topology and branch lengths of the phylogenetic tree were evaluated by applying the neighbor-joining (NJ) method to the K values and reliability of its topology was evaluated by the bootstrap analysis, as previously described [6].

# 3. Results and discussion

#### 3.1. Nucleotide sequence of ADH7

ADH7 has 9 exons and 8 introns (Fig. 2), as in the case of ADH1 [13], ADH2 [12, 16], ADH3 [17], ADH4 [18], and ADH5 [19], all of which have identical intron positions. Compared to these genes, ADH6 lacks exon 9, but positions of its introns 1-7 are identical to those of the other ADH genes [5]. The nine exons of ADH7 span about 22 kb and the approximate sizes of the introns 1-8 of ADH7 are 5.3, 0.6, 0.3, 0.1, 7.2, 1.4, 3.4,

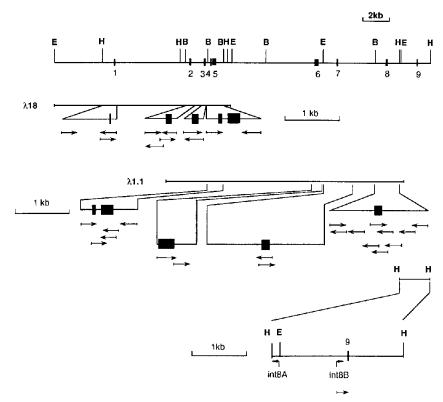


Fig. 1. The genomic composite of the human ADH7 gene determined by overlapping  $\lambda$  clones and PCR amplification. The nine coding exons are indicated by solid boxes with an enlarged scale being provided under each  $\lambda$  clone and genomic fragment used for inverse PCR amplification. (B) BamHI; (E) EcoRI; and (H) HindIII.

and 2.1 kb, respectively. The GT/AG consensus intron splice junctions are conserved.

The amino acid sequence, Hsa7, deduced from *ADH7* consists of 374 residues (Fig. 2). This size is similar to other ADHs, i.e. 374 residues of HsaX, 375 residues of HsaA, B, and C, 380 residues of HsaP, and 368 residues of Hsa6. Hsa7 shows that

functionally important residues are conserved. They include amino acid residues Cys (47), His (68), and Cys (174), which provide ligands to one catalytic zinc atom, and amino acid residues Cys (98), Cys (101), Cys (104), and Cys (112), which provide ligands to another catalytic zinc atom [20]. The sequence conservation of *ADH7* is unlike the *ADH5* pseudogene,

Table 1 Class IV-specific amino acid polymorphism

	71	80	166	196	233	251	261	263	287	317	322	343
EcaX	A	G	L	Α	K	Q	G	D	G	T	W	D
HsaX	_	_	_	_	_	_	-	_	_	_	_	_
RnoX	_	_	_	_	_	_	_	_	_	_		_
MmuX	_	-	_	-	_	_	-	_	_	_	_	_
HsaP		_	_	_	_	_	_	_	_	_		
PmaV	T				_	_	_	_	_	S	-	_
Hsa6	-	_	_	-	Q		-	_	_	S	_	_
Hsa7	T	_	P	_	M	S	N	G	V	C	L	-
RnoIV	V	E	P	V	L	S	T	Q	_	C	_	G
Rpe	_	_	_	_	_	H	-	_	-	Α		_
Gga	_	K	_	-	-	Н	-	_	_	S	_	_
Cja	_	K	-	_	_	Н	-	-	-	S	_	_
PmaA						_	-	_	-	Α	F	E
RnoA	-	_	_	-	_	-	-	_	_	Α	F	E
MmuA		_								Α	F	-
EcaS	_	-	_	_	_	_	-	-	_	Α	F	_
EcaE	_	_	_	_	_	-	-	_	_	Α	F	
HsaC	_	-	_	_	-	-	-	-	-	Α	F	_
PhaB	_	***	_	-	-	-	-	-	-	Α	F	_
HsaB										Α	F	
HsaA	_	_	_	_	_	_	-		_	Α	F	_
MmuA	_	_	M	_	_	-	-	_	-	Α	F	_

```
CAAGACAAAGACAGG ATG GGC ACT GCT GGA AAA GTAAGTGGAA
                                                         5.3 kb
                                                                    TATTTCTTAG GTT ATT AAA TGC AAA
                 Met Gly Thr Ala Gly Lys
                                20
                                                                        30
GCA GCT GTG CTT TGG GAG CAG AAG CAA CCC TTC TEC ATT GAG GAA ATA GAA GTT GCC CCA ECA AAG ACT AAA GAA
Ala Ala Val Leu Trp Glu Gin Lys Gin Pro Phe Ser lie Glu Glu Ile Glu Val Ala Pro Pro Lys Thr Lys Glu
            40
GTT CGC ATT AAG GTAAGCGTGA
                                      TIGAACACAG ATT TIG GCC ACA GGA ATC TGT CGC ACA GAT GAC CAT
                            0.6 kb
                                                   He Leu Ala Thr Gly He Cys Arg Thr Asp Asp His
Val Arg Ile Lys
GTG ATA AAA GGA ACA ATG GTG TCC AAG TIT CCA GTG ATT GTG GGA CAT GAG GCA ACT GGG ATT GTA GAG AGC ATT
Val lie Lys Gly Thr Met Val Ser Lys Phe Pro Val lie Val Gly His Glu Ala Thr Gly Ile Val Glu Ser Ile
        80
GGA GAA GGA GTG ACT ACA GTG AAA CCA G GTATATGCAG
                                                  0.3 kb
                                                             TITCCTGTAG GT GAC AAA GTC ATC CCT CTC
Gly Glu Gly Val Thr Thr Val Lys Pro G
                                                                        ly Asp Lys Val Ite Pro Leu
                       100
                                                               110
TAY CTG CCA CAA TGT AGA GAA TGC AAT GCT TGT CGC AAC CCA GAT GGC AAC CTT TGC ATT AGG AGC GA GTAGGTTT
Phe Leu Pro Gin Cys Arg Glu Cys Asn Ala Cys arg Asn Pro Glu Gly Asn Leu Cys 1le Arg Ser As
                                       120
                                                                              130
     100 bp TATCAAACAG T ATT ACT GGT CGT GGA GTA CTG GCT GAT GGC ACC AGA TTT ACA TGC AAG GGG
                         pile Thr Gly Arg Gly Val Leu Ala Asp Gly Thr Thr Arg Phe Thr Cys Lys Gly
                    140
                                                          150
AAA CCA GTC CAC CAC TTC ATG AAC ACC AGT ACA TTT ACC GAG TAC ACA GTG GTG GAT GAA TCT TCT GTT GCT AAG
Lys Pro Val His His Phe Met Asn Thr Ser Thr Phe Thr Glu Tyr Thr Val Val Asp Glu Ser Ser Val Ala Lys
160
                                       170
                                                                              180
ATT GAT GAT GCA GCT CCT GCG GAG AAA GTC TGT TTA ATT GGC TGT GGG TTT TCC ACT GGA TAT GGC GCT GCT GTT
Ile Asp Asp Ala Ala Pro Pro Glu Lys Vai Cys Leu Ile Gly Cys Giy Phe Ser Thr Gly Tyr Gly Aia Ala Vai
                                                       190
AAA ACT GGC AAG GTAAGAAACA 7.2 kb
                                        CTCTCCACAG GTC AAA CCT GGT TCC ACT TGC GTC GTC TTT GGC CTG
                                                  Val Lys Pro Gly Ser Thr Cys Val Val Phe Gly Leu
                                   210
                                                                          220
GGA GGA GTT GGC CTG TCA GTC ATC ATC GGC TGT AAG TCA GCT GGT GCA TCT AGG ATC ATT GGG ATT GAC CTC AAC
Gly Gly Val Gly Leu Ser Val Ile Met Gly Cys Lys Ser Ala Gly Ala Ser Arg Ile Ile Gly Ile Asp Leu Asn
                                                      240
               230
AAA GAC AAA TIT GAG AAG GCC ATG GCT GTA GGT GCC ACT GAG TGT AYC AGT CCC AAG GAC TCT ACC AAA CCC ATC
Lys Asp Lys Phe Glu Lys Ala Met Ala Val Gly Ala Thr Glu Cys Ile Ser Pro Lys Asp Ser Thr Lys Pro 1le
                                   260
                                                                          270
AGT GAG GTG CTG TCA GAA ATG ACA GGC AAC AAC GTG GGA TAC ACG TTT GAA GTT ATT GGG CAT CTT GAA ACC ATG
Ser Glu Val Leu Ser Glu Met Thr Gly Asn Asn Val Gly Tyr Thr Phe Glu Val Ile Gly His Leu Asp Thr Met
                                                  280
                             CTCAG ATT GAT GCC CTG GCA TCC TGC CAC ATG AAC TAT GTG ACC AGC GTG GTT
GTAAGACCCA
                                   Ile Asp Ala Leu Ala Ser Cys His Met Asn Tyr Val Thr Ser Val Val
                               300
                                                                      310
GTA GGA GTT CCT CCA TCA GCC AAG ATG CTC ACC TAT GAC CCG ATG TTG CTC TTC ACT GGA CGC ACA TGG AAG GGA
Val Gly Val Pro Pro Ser Ala Lys Met Leu Thr Tyr Asp Pro Met Leu Leu Phe Thr Gly Arg Thr Trp Lys Gly
           320
                                                                                         330
TGT GTC TTT GGA G GTCAGGAAAG
                                 3.4 kb
                                          TTATTTGCAG GT TTG AAA AGC AGA GAT GAT GTC CCA AAA CTA
Cys Val Phe Gly G
                                                       ly Leu Lys Ser Arg Asp Asp Val Pro Lys Leu
                              340
                                                                      350
GTG ACT GAG TTC CTG GCA AAG AAA TTT GAC CTG GAC CAG TTG ATA ACT CAT GTT TTA CCA TTT AAA AAA ATC AGT
Val Thr Glu Phe Leu Ala Lys Lys Phe Asp Leu Asp Gln Leu Ile Thr His Val Leu Pro Phe Lys Lys Ile Ser
           360
                                                                                         370
GAA GGA TTT GAG CTG CTC AAT TCA GGA CAA AG GTAACTGTTT
                                                         2.1 kb
                                                                     TACATTICAG C ATT CGA ACG GTC
Glu Gly Phe Glu Leu Leu Asn Ser Gly Gln Se
                                                                                r Ite Arg Thr Val
CIG ACG TTT TGA GATCCAAAGT
Leu Thr Phe *
```

Fig. 2. The nucleotide sequence of ADH7 and the deduced amino acid sequence.

which contains no introns and three frameshift mutations in the coding region with a total of 10 internal termination codons [9]. The gene duplication event between ADH5 and the pseudogene seems to have occurred only about 3.5 million years ago and the pseudogene has undergone a rapid change since then. Thus, possessing no apparent structural defect and conservation of functionally important codons, *ADH7* seems capable of being functional.

## 3.2. A phylogenetic tree

The phylogenetic tree for the Hsa7 and other vertebrate ADHs is shown in Fig. 3. Fig. 3 clearly shows that Hsa7 is most closely related to RnoIV and belongs to class IV. Bootstrap support of the two ADHs forming one cluster is 1.0 and is

highly reliable. Fig. 3 shows that class IV is most closely related to class I, their common ancestor diverged from the ancestor of class V, their common ancestor diverged from the ancestor of class II, and their common ancestor diverged from the ancestor of class III, which is consistent with Fig. 1 in Yokoyama and Harry [6]. Bootstrap analysis shows that clusterings of the mammalian classes I, III, IV and V ADHs are highly reliable. However, the relationship between the classes I and IV in Fig. 3 is not clear-cut and frog, chicken, and quail ADHs may be more closely related to the mammalian class I ADHs than to the class IV ADHs, reflecting the organismal relationship of the three species. Fig. 3 also clearly shows that recently characterized ADH1 (PmaA) and ADH2 (PmaV) from deer mouse belong to classes I and V, respectively.

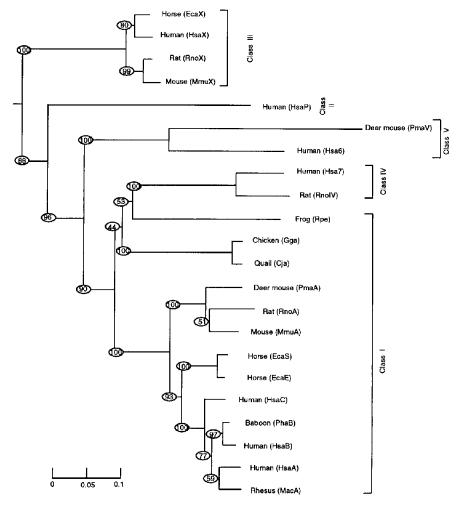


Fig. 3. The rooted phylogenetic tree for the vertebrate ADHs. Branch length was estimated by the number of amino acid replacements per site.

From Fig. 3, we can see that Hsa7 and RnoIV have similar branch lengths after their divergence from each other, showing that the two proteins are evolving with similar rates. The branch length for Hsa7 since its divergence from RnoIV is 0.07 per site. After their divergences from the corresponding murine ADHs, the branch lengths for HsaA, B, and C are 0.10, 0.09, and 0.08 per site, respectively, while that for HsaX is 0.03. Thus, class IV ADHs are evolving with rates somewhere between those of class I and class V ADHs (see also Fig. 3) Interestingly, the zinc-containing ADH from *Drosophila melanogaster* [21] has been found recently and is shown to belong to the evolutionarily most conserved class III. All of these observations strongly suggest that Hsa7 has had a strong functional constraint against amino acid changes and that *ADH7* is a functional gene.

### 3.3. Characteristics of class IV proteins

By comparing RnoIV to other classes of ADHs, amino acid residues Gly (47), Ala (294), and Tyr (363) of RnoIV are suggested to be class IV-specific changes by affecting both substrate and coenzyme binding [8]. However, the corresponding amino acids for Hsa7 are Arg (48), Val (294), and Asn (363) and the three residues are not shared by the two class IV ADHs.

Having two class IV proteins, RnoIV and Hsa7, it is of

interest to identify class IV-specific amino acid changes. Table 1 shows polymorphic residues, where (i) amino acid changes of at least one of the two class IV proteins are not shared by others and (ii) amino acids among other classes of ADHs are either identical or belong to the same functional group. Comparing all vertebrate ADHs, Hsa7 and RnoIV are uniquely distinguished from others by Pro (166), Ser (251), and Cys (317) (Table 1). These three and additional amino acid changes in Hsa7 and RnoIV in Table 1 might have been important in its functional differentiation from class I ADH.

Acknowledgements: This work was supported by the National Institute of Health Grant AA-03539.

#### References

- [1] Smith, M. (1986) Adv. Hum. Genet. 14, 249-290.
- [2] Eklund, H., Branden, C.-I. and Jornvall, H. (1976) J. Mol. Biol. 102, 61-73.
- [3] Yokoyama, S., Yokoyama, R., Kinlaw, C.S. and Harry, D.E. (1990) Mol. Biol. Evol. 7, 143-154.
- [4] Moreno, A. and Pares, X. (1991) J. Biol. Chem. 266, 1128-1133,
- [5] Yasunami, M., Chen, C.-S. and Yoshida, A. (1991) Proc. Natl. Acad. Sci. USA 88, 7610–7614.
- [6] Yokoyama, S. and Harry, D.E. (1993) Mol. Biol. Evol. 10, 1215– 1226.

- [7] Sun, H.-W. and Plapp, B. (1992) J. Mol. Evol. 34, 522-535.
- [8] Pares, X., Cederlund, E., Moreno, A., Hjelmqvist, L., Farres, J. and Jornvall, H. (1994) Proc. Natl. Acad. Sci. USA 91, 1893-
- [9] Matsuo, Y. and Yokoyama, S. (1990) Am. J. Hum. Genet. 46,
- [10] Yokoyama, S., Yokoyama, R. and Rotwein, P. (1987) Jpn. J. Genet. 62, 241-256.
- [11] Ochman, H., Medhora, M.M., Garza, D. and Hartl, D.L. (1990) in: PCR protocols: a guide to methods and applications (Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J., Eds.) pp. 219–227, Academic Press, New York.
- [12] Matsuo, Y., Yokoyama, R. and Yokoyama, S. (1989) Eur. J. Biochem. 183, 317-320.
- [13] Matsuo, Y. and Yokoyama, S. (1989) FEBS Lett. 243, 57-60.
- [14] Zheng, Y.-W., Bey, M., Liu, H. and Felder, M.R. (1993) J. Biol. Chem. 268, 24933-24939.

- [15] Kimura, M. (1983) The Neutral Theory of Molecular Evolution, Cambridge University Press, Cambridge.
- [16] Duester, G., Smith, M., Bilanchone, V. and Hatfield, G. W. (1986) J. Biol. Chem. 261, 2027-2033.
- [17] Yokoyama, S., Matsuo, Y., Rajasekharan, S. and Yokoyama, R. (1992) Jpn. J. Genet. 67, 167-171.
- [18] von Bahr-Lindstrom, H., Jornvall, H. and Hoog, J.-O. (1991) Gene 103, 269-274.
- [19] Hur, M.-W. and Edenberg, H.J. (1992) Gene 121, 305-311.[20] Eklund, H., Nordstrom, B., Zeppezauer, E., Soderberg, B.-O., Tapia, O., Branden, C.-l. and Akeson, A. (1976) J. Mol. Biol. 102,
- [21] Danielsson, O., Atrian, S., Luque, T., Hjelmqvist, L., Gonzalez-Duarte, R. and Jornvall, H. (1994) Proc. Natl. Acad. Sci. USA 91, 4980-4984.