Volume 243, number 1, 57–60 FEB 06697 January 1989

Molecular structure of the human alcohol dehydrogenase 1 gene

Yoshinori Matsuo and Shozo Yokoyama

Department of Ecology, Ethology, and Evolution, University of Illinois at Urbana-Champaign, Shelford Vivarium, 606 East Healey Street, Champaign, IL 61820, USA

Received 21 November 1988

The structure and nucleotide sequence of an allele at the ADH1 locus have been determined. The nucleotide sequence of this allele is identical to that of a cDNA clone [(1986) Biochemistry 25, 2465–2470] and the intron positions of the ADH1 gene are identical to that of the ADH2 gene [(1986) J. Biol. Chem. 261, 2027–2033].

Alcohol dehydrogenase; DNA, genomic; Nucleotide diversity; (Human)

1. INTRODUCTION

Human class I alcohol dehydrogenase (ADH) consists of the subunits α, β, and γ, which are encoded by three separate loci ADH1, ADH2, and ADH3, respectively [1]. Molecular characterization of the ADH2 gene shows that 9 exons are stretched over 15 kilobases (kb) in length [2,3]. However, the molecular structures of the ADH1 and ADH3 loci have not been determined.

The complementary DNA sequences of the ADH1 alleles isolated from two different individuals were identical [4,5]. To study the genomic structure and level of DNA polymorphism at the ADH1 locus, we determined the DNA sequence of all nine exons and the intron/exon structure of a human ADH1.

2. MATERIALS AND METHODS

Two sets of genomic libraries were constructed by using genomic DNA partially digested either with *Mbo*I or *Eco*RI and ligating with λEMBL3 or λEMBL4 DNA, respectively. About 120 μg of the genomic DNA from one of us (S.Y.), obtained from peripheral blood leukocytes [6], was partially digested and fractionated on an agarose gel. The DNA in the size range of 9–23 kb was electroeluted from the gel and ligated with λEMBL vector DNA which had been double digested with *Bam*HI and *Eco*RI. The ligated DNA was packaged in vitro into phage particles by using Gigapack packaging extract (purchased from Stratagene) and plated on the nonpermissive *E. coli* host NM539.

Plaque hybridization was carried out using the method of Benton and Davis [7] for 24–36 h at 68°C in 4 × SETDS with nick-translated [8] cDNA probe of ADHβ [9] and 50 μg/ml of heat denatured herring sperm DNA (4 × SETDS: 0.6 M NaCl (pH 7.5), 8 mM EDTA, 10 × Denhardt, 0.1% SDS). From the two human genomic libraries, a total of one million recombinant plaques were screened.

Oligonucleotides which are specific for the subunits α, β, and γ between amino acids 312 and 320 were synthesized and used for locus specific hybridization. Hybridization was carried out with the γ-^{32}P-end-labelled oligomers and heat denatured herring sperm DNA at 37°C for 12 h in 4 × SETDS [3].

Subcloning was conducted using restriction enzymes *Bam*HI, *Eco*RI, *Hin*dIII, *Kpn*I, *Nsi*I, *Pst*I, *Sau*3A, *Spe*I, *Sst*I and *Xba*I. Digested DNA fragments were ligated into the plasmid Bluescript vector from Stratagene. They were sequenced by using the dideoxy-chain-termination method [10–12].

Correspondence address: Y. Matsuo, Department of Ecology, Ethology, and Evolution, University of Illinois at Urbana-Champaign, Shelford Vivarium, 606 East Healey Street, Champaign, IL 61820, USA

3. RESULTS AND DISCUSSION

For 63 positive clones, restriction mapping and locus specific oligonucleotide hybridization were conducted. The numbers of clones which belong to

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

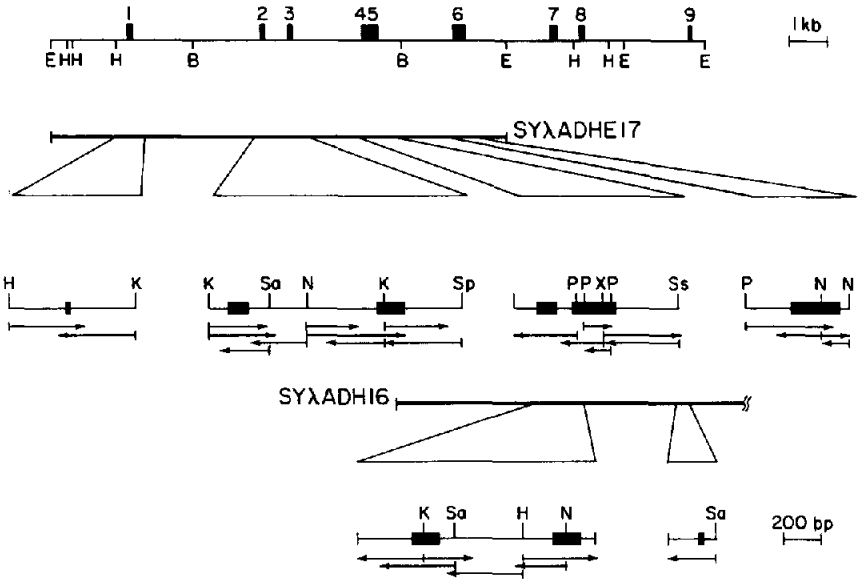

Fig.1. Restriction map and sequencing strategy of the ADH1 gene. B, E, H, K, N, P, S_a, S_p, S_s, X denote *Bam*HI, *Eco*RI, *Hin*dIII, *Kpn*I, *Mbo*I, *Nsi*I, *Pst*I, *Sau*3A, *Spe*I, *Sst*I and *Xba*I sites, respectively. The complete nucleotide sequences of all nine exons were determined by using two overlapping clones SYλADHE17 and SYλADH16.

the ADH1, ADH2, and ADH3 loci were 17, 14, and 13, respectively. The remaining 19 positive clones have not been characterized yet.

To determine the complete nucleotide sequence of all nine exons of the ADH1 gene, two overlapping clones SYλADHE17 and SYλADH16 were used (fig.1). Hybridization experiments showed that 10.7 kb SYλADHE17 had exons 1–6, whereas 16 kb SYλADH16 had exons 6–9. Evidence that the two clones were derived from the same allele comes from a restriction site polymorphism within intron 6. Four out of ten overlapping clones studied have a *Taq*I site about 350 bp downstream from exon 6, but the remaining six clones, including SYλADHE17 and SYλADH16, did not have that site.

The DNA sequence of all nine exons of the ADH1 allele is shown in fig.2. When this sequence was compared to the published cDNA sequences of two different ADH1 alleles [4,5], no nucleotide difference in the coding region was found. Similarly, a very low level of genetic variability has been found between the two different electrophoretic alleles at the ADH2 locus [3,13].

Fig.2 also shows that the ADH1 gene is divided by eight introns (see also fig.1). The positions of the introns are identical to those of the ADH2 gene. Nine exons of the ADH1 gene are stretched over about 15 kb in length and the approximate sizes of the intron 1–8 of the ADH1 gene are 3.2, 0.6, 1.8, 0.1, 1.9, 2.2, 0.6 and 2.8 kb, respectively. The corresponding introns of the ADH2 gene are 2.8, 0.6, 1.7, 0.1, 2.0, 2.2, 0.6 and 2.8 kb, respectively [2,3]. These sizes are very similar and the intron/exon structure of the ADH1 and ADH2 loci are well conserved since their divergence.

Acknowledgements: This work was partially supported by grants from the National Institutes of Health and National Science Foundation.

Fig.2. DNA sequence of a human ADH1 gene. The DNA sequence of all nine exons is shown with the predicted 374 amino acids. The sizes of the eight introns are indicated.

```
          AAGCTTTATCACTTTAACTTAATATTTAACCTAATGAAAACAAAATCTTATTTGAAATTGGAAAAATCAATGTATTGATTGCTGGTTC

ATTGCCCTCTTCTTTATGATTTGACAGTCTGTGAATAATCTAATGGGTGTGGCTTAAAGACCTAGATCATGTGTGGAACTGGAATCGGGTGTTATCAAG

CAAAAAAAATAAATAAATACCTATGCAATACACCTGCTTTATGCACTTGAGCAGGGAAGAAATCCACAAGGACTCACCAGTCTCCTGGTCTGCAGAGAA

GACAGAATCAACATG AGC ACA GCA GGA AAA GTAAGCAAAAAATAT   3.2 kb     TGTTTGCTTTTCTAG GTA ATC AAA TGC AAA
            Met Ser Thr Ala Gly Lys                                              Val Ile Lys Cys Lys
                  1                                                                                10
GCA GCT GTG CTA TGG GAG TTA AAG AAA CCC TTT TCC ATT GAG GAG GTG GAG GTT GCA CCT CCT AAG GCC CAT GAA
Ala Ala Val Leu Trp Glu Leu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro Lys Ala His Glu
                                   20                                                  30
GTT CGT ATT AAG GTGAAATACATTTTT   0.6 kb     CTCTTCACTCTGTAG ATG GTG GCT GTA GGA ATC TGT GGC ACA GAT
Val Arg Ile Lys                                                Met Val Ala Val Gly Ile Cys Gly Thr Asp
                                                                40
GAC CAC GTG GTT AGT GGT ACC ATG GTG ACC CCA CTT CCT GTG ATT TTA GGC CAT GAG GCA GCC GGC ATC GTG GAG
Asp His Val Val Ser Gly Thr Met Val Thr Pro Leu Pro Val Ile Leu Gly His Glu Ala Ala Gly Ile Val Glu
 50                                        60                                                  70
AGT GTT GGA GAA GGG GTG ACT ACA GTC AAA CCA G GTACAGGATTCACAT   1.8 kb    TTTATCCCTCTCCAG GT GAT AAA
Ser Val Gly Glu Gly Val Thr Thr Val Lys Pro G                                              ly Asp Lys
                        80
GTC ATC CCA CTC GCT ATT CCT CAG TGT GGA AAA TGC AGA ATT TGT AAA AAC CCG GAG AGC AAC TAC TGC TTG AAA
Val Ile Pro Leu Ala Ile Pro Gln Cys Gly Lys Cys Arg Ile Cys Lys Asn Pro Glu Ser Asn Tyr Cys Leu Lys
     90                                  100                                            110
AAC GA GTAGGTTTCTGATGC   67 bp     TATTGCACTGTCCAG T GTA AGC AAT CCT CAG GGG ACC CTG CAG GAT GGC ACC
Asn As                                          p Val Ser Asn Pro Gln Gly Thr Leu Gln Asp Gly Thr
                                                                120
AGC AGG TTC ACC TGC AGG AGG AAG CCC ATC CAC CAC TTC CTT GGC ATC AGC ACC TTC TCA CAG TAC ACA GTG GTG
Ser Arg Phe Thr Cys Arg Arg Lys Pro Ile His His Phe Leu Gly Ile Ser Thr Phe Ser Gln Tyr Thr Val Val
        130                                 140                                            150
GAT GAA AAT GCA GTA GCC AAA ATT GAT GCA GCC TCG CCT CTA GAG AAA GTC TGT CTC ATT GGC TGT GGA TTT TCA
Asp Glu Asn Ala Val Ala Lys Ile Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly Phe Ser
                                160                                             170
ACT GGT TAT GGG TCT GCA GTC AAT GTT GCC AAG GTAAGAATGGCAATG     1.9 kb    TTTCTGAAAACACAG GTC ACC CCA
Thr Gly Tyr Gly Ser Ala Val Asn Val Ala Lys                                            Val Thr Pro
        180                                                                                 190
GGC TCT ACC TGT GCT GTG TTT GGC CTG GGA GGG GTC GGC CTA TCT GCT ATT ATG GGC TGT AAA GCA GCT GGG GCA
Gly Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Ala Ile Met Gly Cys Lys Ala Ala Gly Ala
                                200                                             210
GCC AGA ATC ATT GCG GTG GAC ATC AAC AAG GAC AAA TTT GCA AAG GCC AAA GAG TTG GGT GCC ACT GAA TGC ATC
Ala Arg Ile Ile Ala Val Asp Ile Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Lue Gly Ala Thr Glu Cys Ile
            220                                 230                                            240
AAC CCT CAA GAC TAC AAG AAA CCC ATC CAG GAG GTG CTA AAG GAA ATG ACT GAT GGA GGT GTG GAT TTT TCA TTT
Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Lys Glu Met Thr Asp Gly Gly Val Asp Phe Ser Phe
                                250                                             260
GAA GTC ATC GGT CGG CTT GAC ACC ATG GTATGTACCATGACA   2.2 kb     TTCACTTTATTCCAG ATG GCT TCC CTG TTA
Glu Val Ile Gly Arg Leu Asp Thr Met                                                Met Ala Ser Leu Leu
            270                                                                                 280
TGT TGT CAT GAG GCA TGT GGC ACA AGT GTC ATC GTA GGG GTA CCT CCT GAT TCC CAA AAC CTC TCA ATG AAC CCT
Cys Cys His Glu Ala Cys Gly Thr Ser Val Ile Val Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met Asn Pro
                                290                                             300
ATG CTG CTA CTG ACT GGA CGT ACC TGG AAG GGA GCT ATT CTT GGT G GTATGTAGTTAGGCT     0.6 kb     CCATCTT
Met Leu Leu Leu Thr Gly Arg Thr Trp Lys Gly Ala Ile Leu Gly G
                    310                                 320
CTTTTCAG GC TTT AAA AGT AAA GAA TGT GTC CCA AAA CTT GTG GCT GAT TTT ATG GCT AAG AAG TTT TCA TTG GAT
         ly Phe Lys Ser Lys Glu Cys Val Pro Lys Leu Val Ala Asp Phe Met Ala Lys Lys Phe Ser Leu Asp
                                    330                                             340
GCA TTA ATA ACC CAT GTT TTA CCT TTT GAA AAA ATA AAT GAA GGA TTT GAC CTG CTT CAC TCT GGG AAA AG GTAG
Ala Leu Ile Thr His Val Leu Pro Phe Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu His Ser Gly Lys Se
                        350                                 360
ATTTTTTACGT     2.8 kb     TCTTTCCTATTGCAG T ATC CGT ACC ATT CTG ATG TTT TGAGACAATACAGATGTTTTCCCTTGT
                                          r Ile Arg Thr Ile Leu Met Phe ***
                                                    370

GGCAGTCTTCAGCCTCCTCTACCCTACATGATC
```

REFERENCES

[1] Smith, M. (1986) Adv. Hum. Genet. 14, 249–290.
[2] Duester, G., Smith, M., Bilanchone, V. and Hatfield, G.W. (1986) J. Biol. Chem. 261, 2027–2033.
[3] Matsuo, M., Yokoyama, R. and Yokoyama, S., submitted.
[4] Ikuta, T., Szeto, S. and Yoshida, A. (1986) Proc. Natl. Acad. Sci. USA 83, 634–638.
[5] Von Bahr-Lindstrom, H., Hoog, J.O., Heden, L.O., Kaiser, R., Fleetwood, L., Larsson, K., Lake, M., Holmquist, B., Holmgren, A., Hempel, J., Vallee, B.L. and Jornvall, H. (1986) Biochemistry 25, 2465–2470.
[6] Kan, Y.W. and Dozy, A.M. (1978) Proc. Natl. Acad. Sci. USA 75, 5631–5635.
[7] Benton, W.D. and Davis, R.W. (1977) Science 196, 180–182.
[8] Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237–251.
[9] Yokoyama, S., Yokoyama, R. and Rotwein, P. (1987) Jpn. J. Genet. 62, 241–256.
[10] Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[11] Hattori, M., Hidaka, S. and Sakaki, Y. (1985) Nucleic Acids Res. 13, 7813–7827.
[12] Hattori, M. and Sakaki, Y. (1986) Anal. Biochem. 152, 232–238.
[13] Ehring, T., Von Wartburg, J.P. and Wermuth, B. (1988) FEBS Lett. 234, 53–55.