cDNA clones coding for the β-subunit of human liver alcohol dehydrogenase have differently sized 3′-non-coding regions

Lars-Olof Hedén, Jan-Olov Höög, Kerstin Larsson, Mats Lake, Eva Lagerholm, Arne Holmgren, Bert L. Vallee, Hans Jörnvall and Hedvig von Bahr-Lindström

Department of Chemistry I, Karolinska Institutet, S-104 01 Stockholm, KabiGen AB, S-112 87 Stockholm, Sweden and *Center for Biochemical and Biophysical Sciences and Medicine, Harvard Medical School, Boston, MA 02115, USA*

Received 21 October 1985

Three different size classes of cDNA clones coding for the β_1-subunit of human alcohol dehydrogenase (ADH) were characterized from a human liver cDNA library. Clones were identified by hybridization with synthetic oligodeoxyribonucleotides. A total of 2530 nucleotides were determined, covering an ADH-coding region of 1122 nucleotides, a preceding 72-nucleotide segment and 3 types of 3′-non-coding region. The coding nucleotide sequence is in full agreement with the amino acid sequence of the β_1-subunit. Of 8 clones identified, 6 had a short, 213-nucleotide 3′-non-coding region; 1 an intermediate, 590-nucleotide 3′-region; and 1 a long, 1330-nucleotide 3′-region. In addition, 2 unused polyadenylation signals were found. These results suggest that human liver β-ADH mRNAs occur in several size classes, and that in addition to the consensus sequence AATAAA further signals are important for 3′-end formation.

Alcohol dehydrogenase *cDNA* *Oligonucleotide probe* *DNA sequence*

1. INTRODUCTION

Human alcohol dehydrogenase (ADH, EC 1.1.1.1) is a complex enzyme system of different dimeric isozymes, which have been divided into 3 classes, I, II and III, based on electrophoretic mobility and sensitivity to inhibition by pyrazole [1]. The subunits of class I, α, β and γ, are clearly related polypeptide chains [2], coded for by 3 separate gene loci [3]. Alleles giving rise to 3 types of β chain and 2 types of γ chain have been identified [3,4].

The amino acid sequence is known for the class I subunits α, β_1, β_2 and γ [2,5–8]. At the nucleotide level, 3 cDNA sequences for human liver alcohol dehydrogenase have been determined [8–10]. However, one is only a partial sequence [9], another [10] contains a few differences in relation to the protein structure reported [5], and a third suggests a cDNA with the presence of a deletion [8].

Here we present a full-length cDNA sequence derived from the ADH_2^1 allele, prove complete agreement with the protein structure of the β_1-subunit [2,5] and show that there is a variation in the length of the 3′-non-coding region between different β-cDNA clones.

2. MATERIALS AND METHODS

2.1. *RNA preparation and construction of a cDNA library*

Total RNA was prepared from healthy adult human liver of Caucasian origin by the guanidine thiocyanate method [11]. The preparation was enriched in poly(A) RNA by passage twice over oligo(dT)-cellulose (Collaborative Research, type 3).

A cDNA library was constructed from mRNA (Hedén et al., in preparation) essentially according to Okayama and Berg [12]. Vector pT4 (kindly provided by G. Gross, Gesellschaft für Biotech-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

nologische Forschung, Braunschweig, FRG), used as primer and for linker fragment generation, was a derivative of pBR322 where the *Pst*I-*Eco*RI fragment had been replaced by a 168-nucleotide fragment containing a unique *Kpn*I site. The vector-primer was made by T-tailing at the *Kpn*I site followed by *Hin*dIII digestion, separation in 1.0% agarose gel, electro-elution and final purification on oligo(dA)-cellulose (Collaborative Research). The linker fragment was made from the same vector by C-tailing of *Kpn*I digested vector, followed by *Hin*dIII digestion, purification in 1.8% agarose gel, electro-elution and ethanol precipitation. For cDNA synthesis, amounts used were 1 μg of the vector-primer and 5 μg mRNA. *Escherichia coli* strain 5K (*thr thia hsdR hsdM str*^{r}) [13] was used as recipient in transformations [14]. The transformants were pooled and frozen at $-70°C$ as aliquots in glycerol medium [15].

2.2. *Screening for transformants containing ADH nucleotide sequences*

Two oligodeoxyribonucleotides (probes I and II in table 1) were synthesized by the solid-phase phosphoamidite method [16] and labelled at the 5′-end by transfer from [γ-^{32}P]ATP (Amersham, 3000 Ci/mmol) using T_4 polynucleotide kinase (Amersham). The labelled oligodeoxyribonucleotides were used as probes in screening of the cDNA library by colony hybridization and in blotting analysis of restriction enzyme fragments [17].

2.3. *DNA sequence analysis*

Plasmid DNA from hybridization positive clones was cleaved with restriction enzymes and separated by agarose gel electrophoresis (0.8%, low gel temperature agarose; Biorad). Restriction enzymes were used under the conditions recommended by the manufacturers. Nucleotide fragments in agarose slices were ligated into M13 vectors [18]. Sequence analysis was carried out by the dideoxy chain termination method [19] using as primer an M13 specific 17-mer (Amersham) or synthetic oligodeoxyribonucleotides (probes III, IV and V in table 1) [16] corresponding to the then analyzed parts of the nucleotide sequence. The labelled nucleotide was [α-^{35}S]dATP (Amersham, 600 Ci/mmol) and the sequence mixture was separated on ultrathin (0.2 mm) 7 M urea-6% polyacrylamide gels.

3. RESULTS

3.1. *Characterization of plasmids*

From a frozen stock of a human liver cDNA library, containing 1.5×10^5 transformants with inserts up to 3 kb, approx. 10000 colonies were screened for ADH sequences by hybridization. Thirteen clones hybridized with probe I (table 1), and partial sequence analysis revealed 8 of them to correspond to the β-subunit of human liver alcohol dehydrogenase. Plasmid DNA was isolated from these 8 clones and digested with *Kpn*I and *Pst*I. The fragments were separated by agarose gel electrophoresis, hybridized to probe II (table 1), and submitted to partial sequence determination. The results revealed that there is a variation in length of the 3′-non-coding region of the cDNA clones.

Table 1

Synthetic oligodeoxyribonucleotides used as hybridization probes and as primers in the sequence analysis

Probe	Size	Nucleotide sequence	Corresponding to
I	18-mer	AGC AGG AGG TAC CCC TAC	Val 292–Ala 297
II	21-mer	TTA ATA ACC CAT GTT TTA CCT	Leu 345–Pro 351
III	18-mer	TTA GGA TGA TGA CTG AAA	nucleotides 1942–1960
IV	18-mer	CTG TGT TAA TGC TTT GTT	nucleotides 2241–2259
V	18-mer	GAA TGT AGA TAT TGC AAC	nucleotides 2418–2436

Oligodeoxyribonucleotides I and II were synthesized to correspond to the partial cDNA sequence [9] of the β-subunit of human liver alcohol dehydrogenase. III, IV and V were synthesized to correspond to the presently determined sequence

Thus, 3 classes in size were observed (fig.1). Six clones (pADH1, 2, 5, 9, 10 and 13) have an identical, short 3′-non-coding region, one (pADH3) contains an intermediate-sized 3′-non-coding region, and one (pADH7) a long 3′-non-coding region. The restriction enzyme analysis also indicated that pADH2 contained the longest 5′-non-coding region. From these data, pADH2, 3 and 7 were chosen for complete sequence analysis.

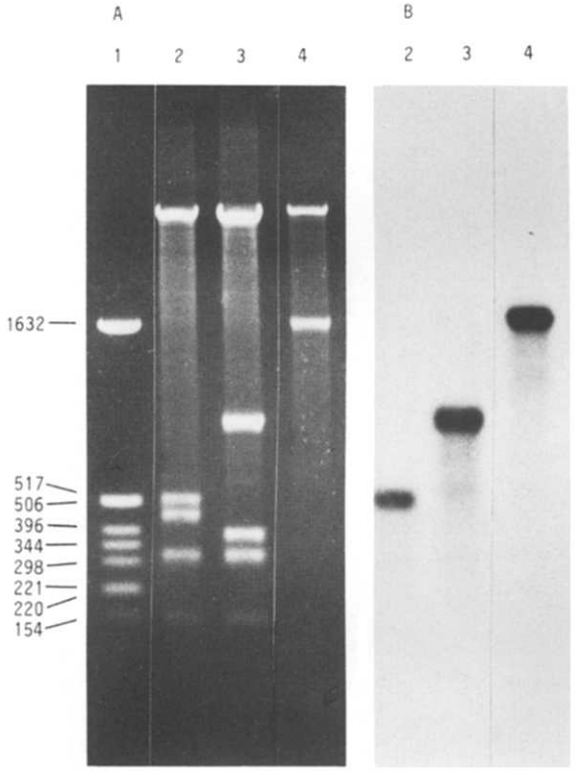


Fig.1. Restriction enzyme analysis and Southern hybridization. pADH plasmid DNA was digested with *Kpn*I plus *Pst*I and separated in 0.8% agarose. Patterns obtained are shown in A. Autoradiography after transfer to nitrocellulose and hybridization with probe II (table 1) is shown in B. To the left is a standard of pBR322 digested with *Hin*fI. Lanes: 2, pADH2; 3, pADH3; 4, pADH7.

3.2. *Nucleotide sequence analysis*

The inserts of the 3 plasmids chosen were subjected to sequence analysis as shown in fig.2. Plasmid DNA was cleaved with restriction enzymes, the digests were separated by agarose gel electrophoresis and large fragments were recleaved (fig.2). All fragments were ligated in both directions in appropriately digested M13 vectors and their nucleotide sequences were determined with the dideoxy chain termination method on both strands except for a short segment of the 3′-non-coding region where probes III, IV and V (table 1 and fig.2) were utilized as sequencing primers.

The short variant of the 3′-non-coding region was shown to contain 213 nucleotides, the intermediate-sized variant 590 nucleotides, and the long variant 1330 nucleotides. pADH2 covers the region from −72 to 1400, pADH3 that from −7 to 1990, and pADH7 that from 600 to 2458 (fig.3). In total, 2530 nucleotides were determined. The 1122-nucleotide coding region is initiated by an ATG codon, and terminated by a TGA codon. The translated nucleotide sequence is in full agreement with the amino acid sequence of the β_1-subunit of human alcohol dehydrogenase [2,5].

4. DISCUSSION

4.1. *Characterization of cDNA clones*

One of the clones, pADH2, corresponding to the β-subunit of alcohol dehydrogenase has a 5′-non-

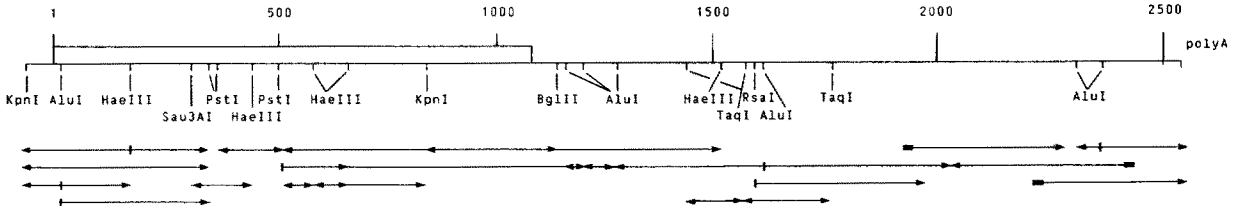
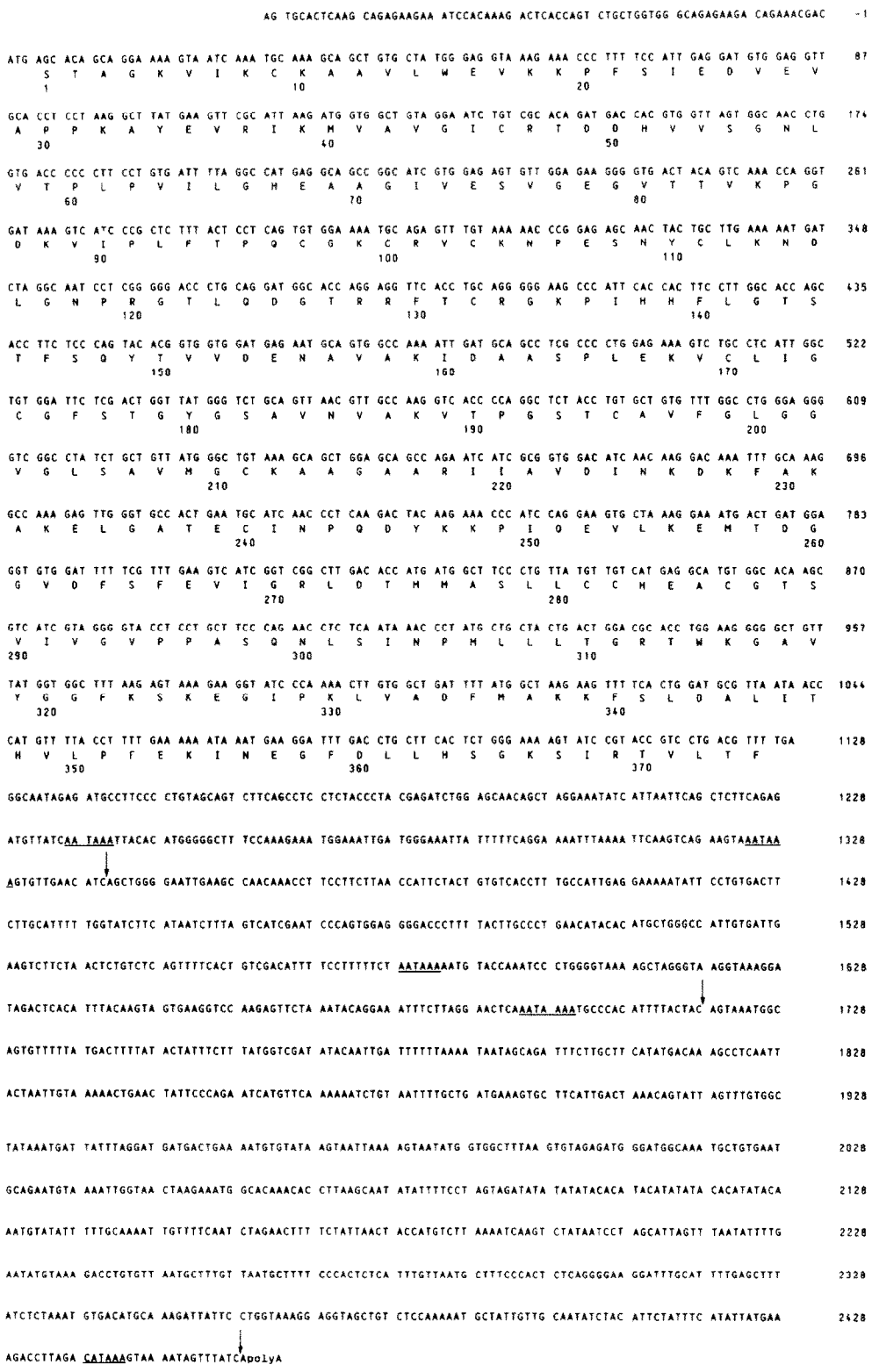


Fig.2. Strategy for sequence determination of cDNA clones corresponding to the β_1-subunit. Only restriction sites used for the dideoxy sequence method are shown. The coding region is boxed. Arrows indicate the direction of sequencing and the scale at the top refers to base pairs. Filled boxes show location of sequencing primers (table 1).


AG TGCACTCAAG CAGAGAAGAA ATCCACAAAG ACTCACCAGT CTGCTGGTGG GCAGAGAAGA CAGAAACGAC -1
ATG AGC ACA GCA GGA AAA GTA ATC AAA TGC AAA GCA GCT GTG CTA TGG GAG GTA AAG AAA CCC TTT TCC ATT GAG GAT GTG GAG GTT 87
S T A G K V I K C K A A V L W E V K K P F S I E D V E V
1 10 20
GCA CCT CCT AAG GCT TAT GAA GTT CGC ATT AAG ATG GTG GCT GTA GGA ATC TGT CGC ACA GAT GAC CAC GTG GTT AGT GGC AAC CTG 174
A P P K A Y E V R I K M V A V G I C R T D D H V V S G N L
30 40 50
GTG ACC CCC CTT CCT GTG ATT TTA GGC CAT GAG GCA GCC GGC ATC GTG GAG AGT GTT GGA GAA GGG GTG ACT ACA GTC AAA CCA GGT 261
V T P L P V I L G H E A A G I V E S V G E G V T T V K P G
60 70 80
GAT AAA GTC ATC CCG CTC TTT ACT CCT CAG TGT GGA AAA TGC AGA GTT TGT AAA AAC CCG GAG AGC AAC TAC TGC TTG AAA AAT GAT 348
D K V I P L F T P Q C G K C R V C K N P E S N Y C L K N D
90 100 110
CTA GGC AAT CCT CGG GGG ACC CTG CAG GAT GGC ACC AGG AGG TTC ACC TGC AGG GGG AAG CCC ATT CAC CAC TTC CTT GGC ACC AGC 435
L G N P R G T L Q D G T R R F T C R G K P I H H F L G T S
120 130 140
ACC TTC TCC CAG TAC ACG GTG GTG GAT GAG AAT GCA GTG GCC AAA ATT GAT GCA GCC TCG CCC CTG GAG AAA GTC TGC CTC ATT GGC 522
T F S Q Y T V V D E N A V A K I D A A S P L E K V C L I G
150 160 170
TGT GGA TTC TCG ACT GGT TAT GGG TCT GCA GTT AAC GTT GCC AAG GTC ACC CCA GGC TCT ACC TGT GCT GTG TTT GGC CTG GGA GGG 609
C G F S T G Y G S A V N V A K V T P G S T C A V F G L G G
180 190 200
GTC GGC CTA TCT GCT GTT ATG GGC TGT AAA GCA GCT GGA GCA GCC AGA ATC ATC GCG GTG GAC ATC AAC AAG GAC AAA TTT GCA AAG 696
V G L S A V M G C K A A G A A R I I A V D I N K D K F A K
210 220 230
GCC AAA GAG TTG GGT GCC ACT GAA TGC ATC AAC CCT CAA GAC TAC AAG AAA CCC ATC CAG GAA GTG CTA AAG GAA ATG ACT GAT GGA 783
A K E L G A T E C I N P Q D Y K K P I Q E V L K E M T D G
240 250 260
GGT GTG GAT TTT TCG TTT GAA GTC ATC GGT CGG CTT GAC ACC ATG ATG GCT TCC CTG TTA TGT TGT CAT GAG GCA TGT GGC ACA AGC 870
G V D F S F E V I G R L D T M M A S L L C C H E A C G T S
270 280
GTC ATC GTA GGG GTA CCT CCT GCT TCC CAG AAC CTC TCA ATA AAC CCT ATG CTG CTA CTG ACT GGA CGC ACC TGG AAG GGG GCT GTT 957
V I V G V P P A S Q N L S I N P M L L L T G R T W K G G A V
290 300 310
TAT GGT GGC TTT AAG AGT AAA GAA GGT ATC CCA AAA CTT GTG GCT GAT TTT ATG GCT AAG AAG TTT TCA CTG GAT GCG TTA ATA ACC 1044
Y G G F K S K E G I P K L V A D F M A K K F S L D A L I T
320 330 340
CAT GTT TTA CCT TTT GAA AAA ATA AAT GAA GGA TTT GAC CTG CTT CAC TCT GGG AAA AGT ATC CGT ACC GTC CTG ACG TTT TGA 1128
H V L P F E K I N E G F D L L H S G K S I R T V L T F
350 360 370
GGCAATAGAG ATGCCTTCCC CTGTAGCAGT CTTCAGCCTC CTCTACCCTA CGAGATCTGG AGCAACAGCT AGGAAATATC ATTAATTCAG CTCTTCAGAG 1228
ATGTTATCAA TAAATTACAC ATGGGGGCTT TCCAAAGAAA TGGAAATTGA TGGGAAATTA TTTTTCAGGA AAATTTAAAA TTCAAGTCAG AAGTAAATAA 1328
AGTGTTGAAC ATCAGCTGGG GAATTGAAGC CAACAAACCT TCCTTCTTAA CCATTCTACT GTGTCACCTT TGCCATTGAG GAAAAATATT CCTGTGACTT 1428
CTTGCATTTT TGGTATCTTC ATAATCTTTA GTCATCGAAT CCCAGTGGAG GGGACCCTTT TACTTGCCCT GAACATACAC ATGCTGGGCC ATTGTGATTG 1528
AAGTCTTCTA ACTCTGTCTC AGTTTTCACT GTCGACATTT TCCTTTTTCT AATAAAAATG TACCAAATCC CTGGGGTAAA AGCTAGGGTA AGGTAAAGGA 1628
TAGACTCACA TTTACAAGTA GTGAAGGTCC AAGAGTTCTA AATACAGGAA ATTTCTTAGG AACTCAAATA AAATGCCCAC ATTTTACTAC AGTAAATGGC 1728
AGTGTTTTTA TGACTTTTAT ACTATTTCTT TATGGTCGAT ATACAATTGA TTTTTTAAAA TAATAGCAGA TTTCTTGCTT CATATGACAA AGCCTCAATT 1828
ACTAATTGTA AAAACTGAAC TATTCCCAGA ATCATGTTCA AAAAATCTGT AATTTTGCTG ATGAAAGTGC TTCATTGACT AAACAGTATT AGTTTGTGGC 1928
TATAAATGAT TATTTAGGAT GATGACTGAA AATGTGTATA AGTAATTAAA AGTAATATG GTGGCTTTAA GTGTAGAGATG GGATGGCAAA TGCTGTGAAT 2028
GCAGAATGTA AAATTGGTAA CTAAGAAATG GCACAAACAC CTTAAGCAAT ATATTTTCCT AGTAGATATA TATATACACA TACATATATA CACATATACA 2128
AATGTATATT TTTGCAAAAT TGTTTTCAAT CTAGAACTTT TCTATTAACT ACCATGTCTT AAAATCAAGT CTATAATCCT AGCATTAGTT TAATATTTTG 2228
AATATGTAAA GACCTGTGTT AATGCTTTGT TAATGCTTTT CCCACTCTCA TTTGTTAATG CTTTCCCACT CTCAGGGGAA GGATTTGCAT TTTGAGCTTT 2328
ATCTCTAAAT GTGACATGCA AAGATTATTC CTGGTAAAGG AGGTAGCTGT CTCCAAAAAT GCTATTGTTG CAATATCTAC ATTCTATTTC ATATTATGAA 2428
AGACCTTAGA CATAAAGTAA AATAGTTTATCApolyA

←

Fig.3. Nucleotide sequence of the cDNA clones derived from the ADH_2^1 allele and the amino acid sequence of the β_1-subunit of human alcohol dehydrogenase. The nucleotides are numbered on the right-hand side in the 5′- to 3′-direction with 1 for A of the ATG initiator codon; negative numbers refer to the 5′-untranslated region. Amino acid residues are numbered below the sequence. The 3 proven and the 2 putative polyadenylation signals in the 3′-untranslated region are underlined. Arrows indicate where the 3′-non-coding regions are polyadenylated.

coding region of 72 nucleotides. As this region in eucaryotic mRNA usually ranges from 40 to 80 nucleotides [20], pADH2 probably represents a full-length cDNA copy.

All regions of the nucleotide sequence, except the part closest to the poly(A) tail, have been determined on both strands. The coding part shows total agreement with the amino acid sequence of the β_1-subunit of human alcohol dehydrogenase [2,5]. Consequently, the nucleotide sequence now determined corresponds to the mRNA from the ADH_2^1 locus coding for the β_1-subunit. The previously known partial cDNA sequence for the β-subunit [9] is identical to the region between nucleotide 853 and 1740 in the presently determined sequence (fig.3). However, a cDNA sequence previously suggested to correspond to the whole β_1-subunit [10] differs from the sequence given in fig.3 at 6 positions, and from the directly determined β_1-subunit. Five differences (positions 703–706 and 708: reported before as AGAG and T, respectively [10]) concern the coding region and correspond to 2 amino acid replacements and one concerns the 3′-non-coding region (position 1180: reported as A). The reason for the discrepancy is unknown, but the present data and the sequence directly determined from the protein agree.

4.2. *Size variation of the 3′-non-coding region*

The lengths of the 3′-non-coding regions differ between the ADH clones isolated. Three length variants have been found, containing 213, 590 and 1330 nucleotides, respectively, between the stop codon TGA, and the poly(A) tail. The shortest variant is found in 6 clones, whereas the medium-length and the long variants are only found in one clone each. Another cDNA clone reported [10] cannot be finally judged regarding poly(A) tailing but the partial cDNA clone previously isolated from yet another cDNA library [9] is a representative of the medium-length variant now defined. It thus appears as if different sizes of 3′-non-coding regions occur, with the shortest variant the most common cDNA in the present study. This is consistent with similar findings in other systems. In yeast alcohol dehydrogenase, however, more than 90% of the ADH-I mRNAs map to the poly(A) addition site furthest downstream from the structural gene [21].

The consensus sequence, AATAAA [22], is required for polyadenylation of mRNA. This polyadenylation signal is found at 4 positions in the cDNA sequence coding for human ADH. The presently defined short and medium forms of the 3′-non-coding region use signals at positions 1341 and 1718, respectively, and the long form uses a less common signal CATAAA [23] at position 2438. Clones using the remaining potential polyadenylation signals at positions 1579 and 1787 were not found, supporting the notion that AATAAA is necessary but not sufficient for polyadenylation [24]. One further type of consensus sequence, CAYTG, immediately up- or downstream from the polyadenylation site [25], is not found at any of the polyadenylation sites now used. Another signal sequence, YGTGTTYY, located about 30 nucleotides downstream from the AATAAA signal has been found in 67% of mammalian mRNAs [26]. A homolog of this sequence, AGTGTTTT, is now found only at position 1728 (fig.3), correctly positioned for use in the medium-length variant. It appears that still further signals are associated with 3′-end selection.

ACKNOWLEDGEMENTS

This work was supported by grants from the Swedish Medical Research Council (projects 13X-3529, 13X-3532 and 13X-7148), the Magn. Bergvall Foundation, the Endowment for Research in Human Biology, Boston, and KabiGen AB.

REFERENCES

[1] Strydom, D.J. and Vallee, B.L. (1982) Anal. Biochem. 123, 422–429.

[2] Hempel, J., Holmquist, B., Fleetwood, L., Kaiser, R., Barros-Söderling, J., Bühler, R., Vallee, B.L. and Jörnvall, H. (1985) Biochemistry 24, 5303–5307.

[3] Smith, M., Hopkinson, D.A. and Harris, H. (1971) Ann. Hum. Genet. 34, 251–271.

[4] Bosron, W.F., Li, T.-K. and Vallee, B.L. (1980) Proc. Natl. Acad. Sci. USA 77, 5784–5788.

[5] Hempel, J., Bühler, R., Kaiser, R., Holmquist, B., De Zalenski, C., Von Wartburg, J.-P., Vallee, B.L. and Jörnvall, H. (1984) Eur. J. Biochem. 145, 437–445.

[6] Bühler, R., Hempel, J., Kaiser, R., De Zalenski, C., Von Wartburg, J.-P. and Jörnvall, H. (1984) Eur. J. Biochem. 145, 447–453.

[7] Jörnvall, H., Hempel, J., Vallee, B.L., Bosron, W.F. and Li, T.-K. (1984) Proc. Natl. Acad. Sci. USA 81, 3024–3028.

[8] Von Bahr-Lindström, H., Höög, J.-O., Hedén, L.-O., Kaiser, R., Fleetwood, L., Larsson, K., Lake, M., Holmquist, B., Holmgren, A., Hempel, J., Vallee, B.L. and Jörnvall, H. (1985) Biochemistry, in press.

[9] Duester, G., Hatfield, G.H., Bühler, R., Hempel, J., Jörnvall, H. and Smith, M. (1984) Proc. Natl. Acad. Sci. USA 81, 4055–4059.

[10] Ikuta, T., Fujiyoshi, T., Kurachi, K. and Yoshida, A. (1985) Proc. Natl. Acad. Sci. USA 82, 2703–2707.

[11] Chirgwin, J.M., Przybylska, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.

[12] Okayama, H. and Berg, P. (1982) Mol. Cell. Biol. 2, 161–170.

[13] Kreft, J., Berger, H., Härtlein, M., Müller, B., Weidinger, G. and Goebel, W. (1983) J. Bacteriol. 155, 681–689.

[14] Hanahan, D. (1983) J. Mol. Biol. 166, 557–580.

[15] Gergen, J.P., Stern, R.H. and Wensink, P.C. (1979) Nucleic Acids Res. 7, 2115–2136.

[16] Josephson, S., Lagerholm, E. and Palm, G. (1984) Acta Chem. Scand. B38, 539–545.

[17] Wallace, R.B., Schaffer, J., Murphy, R.F., Bonner, J., Hirose, T. and Itakura, K. (1979) Nucleic Acids Res. 6, 3543–3557.

[18] Norrander, J., Kempe, T. and Messing, J. (1983) Gene 26, 101–106.

[19] Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.

[20] Kozak, M. (1983) Microbiol. Rev. 47, 1–45.

[21] Bennetzen, J.L. and Hall, B.D. (1982) J. Biol. Chem. 257, 3018–3025.

[22] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211–214.

[23] Wickens, M. and Stephenson, P. (1984) Science 226, 1045–1051.

[24] McReynolds, L., O'Malley, B.W., Nisbet, A.D., Fothergill, J.E., Givol, D., Fields, S., Robertson, M. and Brownlee, G.G. (1978) Nature 273, 723–728.

[25] Berget, S.M. (1984) Nature 309, 179–181.

[26] McLauchlan, J., Gaffney, D., Whitton, J.L. and Clements, J.-B. (1985) Nucleic Acids Res. 13, 1347–1368.