UC Irvine
UC Irvine Previously Published Works

Title
Molecular analysis of the human class I alcohol dehydrogenase gene family and nucleotide sequence of the gene encoding the beta subunit.

Permalink
https://escholarship.org/uc/item/9z54n8fr

Journal
Journal of Biological Chemistry, 261(5)

ISSN
0021-9258

Authors
Duester, G
Smith, M
Bilanchone, V
et al.

Publication Date
1986-02-01

DOI
10.1016/s0021-9258(17)35892-1

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 261, No. 5, Issue of February 15, pp. 2027-2033, 1986
Printed in U.S.A.
© 1986 by The American Society of Biological Chemists, Inc.

Molecular Analysis of the Human Class I Alcohol Dehydrogenase Gene Family and Nucleotide Sequence of the Gene Encoding the β Subunit*

(Received for publication, June 10, 1985)

Gregg Duester‡§, Moyra Smith‡¶, Virginia Bilanchone¶, and G. Wesley Hatfield‡

From the ‡Department of Microbiology and Molecular Genetics and the ¶Department of Pediatrics, University of California, Irvine, Irvine, California 92717

Human alcohol dehydrogenase (ADH) exists as a heterogeneous group of isozymes capable of oxidizing a wide variety of aliphatic and aromatic alcohols. The five distinct human ADH subunits, each encoded by a separate gene, are differentially expressed during development and are subject to tissue-specific regulation. To analyze the organization and regulation of human ADH genes we first isolated a cDNA clone (pADH12) encoding the 3′ portion of the β ADH gene. In the current study pADH12 was used to screen a human genomic library, and several overlapping and nonoverlapping clones were selected. Hybridization and partial nucleotide sequence analyses of the clones indicated that three full-length human ADH genes encoding the α, β, and γ subunits were isolated. Human genomic DNA hybridization results indicate that the α, β, and γ ADH genes form a closely related gene family and suggest that the other known human ADH genes (*i.e.* those encoding the π and χ subunits) share a more distant evolutionary relationship. Nucleotide sequence analysis of the β ADH gene reveals that the coding region is interrupted by eight introns and spans approximately 15 kilobases. A presumptive transcription initiation site for the β ADH gene was located by S1 nuclease mapping at a position 70 base pairs upstream of the start codon. The 5′ flanking region possesses a TATA box promoter element as well as two tandem DNA sequences which display homology to previously examined glucocorticoid-responsive elements.

Alcohol dehydrogenase (ADH[1]; alcohol:NAD^+ oxidoreductase, EC 1.1.1.1) is an enzyme well characterized in many species. The amino acid sequence (1) and tertiary structure (2) of horse liver ADH are known, as well as catalytic mechanisms and evolutionary divergence (3), ADH amino acid and DNA sequences as well as patterns of gene expression have been analyzed in yeast (4–7), *Drosophila* (8–10), and maize (11, 12).

Previous analyses of human ADH have laid the groundwork for in-depth analysis of gene regulation and evolution in this system. Human liver ADH exists as a heterogeneous group of isozymes (13) which can be placed into three classes based on structural and functional distinctions (14). The Class I ADH isozymes include homodimers and heterodimers of the α, β, and γ protein chains which are encoded by *ADH*, *ADH2*, and *ADH3*, respectively (15). The π isozyme encoded by *ADH4* has been designated Class II ADH (16), and the χ isozyme encoded by *ADH5* is referred to as Class III ADH (17). The various ADH isozymes can oxidize a wide variety of aliphatic and aromatic alcohols (18). For many isozymes it has been determined that long-chain aliphatic and aromatic alcohols are better substrates than ethanol, suggesting that ADH is involved in the metabolism of a wide variety of alcohols in the liver (19).

Tissue-specific expression of human ADH is evident. Liver is by far the richest source of Class I and Class II ADH, though low levels of activity have also been found in lung, kidney, and the gastrointestinal tract (15). Class III ADH, however, is easily detected in all human tissues examined (20). The distribution of Class I ADH subunits in human tissue has been examined in detail. Adult liver contains large amounts of α, β, and γ subunits; adult stomach, intestine, and kidney exhibit mostly γ subunits with a small amount of β subunits; adult lung has only β subunits (18, 21). Developmental regulation of Class I ADH gene expression is also evident. The α subunit is solely expressed in the early fetal liver with the additional expression of β during the second trimester, and γ shortly after birth (18).

Allelic variants at both the *ADH2* and *ADH3* loci encode the β_1 and β_2 polypeptides, and the γ_1 and γ_2 polypeptides, respectively (22). In Caucasians the β_1 allele is predominant over the β_2 allele, which is observed in only 10% of those examined (23). Conversely, β_2 is the prominent allele in Oriental populations, since it occurs in 85–90% of those examined (24, 25). The polymorphic variants of γ ADH also occur in racially distinct gene patterns. Caucasians possess the γ_1 and γ_2 alleles at gene frequencies of 0.6 and 0.4, respectively (21), but in Orientals γ_1 predominates over the γ_2 allele with a gene frequency of 0.91 (24). Since the allelic forms of β ADH (26) and γ ADH (19) differ in how efficiently they utilize ethanol as a substrate, it will be of interest to determine if certain alleles correlate with individuals who suffer from alcoholism (27) or fetal alcohol syndrome (28). The recent identification of restriction fragment length polymorphisms specific for human Class I ADH genes should facilitate such an analysis (29, 30).

Amino acid sequence analyses of human β_1 and β_2 ADH have indicated a single amino acid substitution at residue 47 (31). The complete amino acid sequences for human β_1 ADH (32), γ_1 ADH (33), and α ADH (34) indicate a very close relationship between all three, *i.e.* about 95% identity. This structural information coupled with the availability of a cDNA for human β ADH (35) has enabled us to initiate an in-depth molecular analysis of the Class I ADH gene family.

* This research was supported by Grants GM-08733 and AA-06622 (G. D.), GM-32112 (G. W. H.), and AA-05781 (M. S.) from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed.

[1] The abbreviations used are: ADH, alcohol dehydrogenase; kb, kilobase(s); bp, base pair(s); PIPES, 1,4-piperazinediethanesulfonic acid; GRE, glucocorticoid responsive elements.

Here we describe the isolation and characterization of genomic clones encoding human α, β, and γ ADH. Our results indicate that these three genes are evolutionarily very closely related to each other, but only distantly related to the π or χ ADH genes. The nucleotide sequence of the β ADH gene has revealed that the amino acid-coding region spans 15 kilobases (kb) and is divided into nine exons. Conserved sequences upstream of a putative β ADH transcription initiation site are discussed.

EXPERIMENTAL PROCEDURES

Cloning—The human total genomic library constructed by Lawn *et al.* (36) was screened by the plaque hybridization method of Benton and Davis (37). The 1.0-kb β ADH cDNA insert of pADH12 (35) was electrophoretically purified, then ^{32}P-labeled by nick translation and hybridized to the human library as described elsewhere (38). Chromosome walking was accomplished by rescreening the library with an electrophoretically purified 1.3-kb *Eco*RI fragment from λ ADH5 (Fig. 1) which lacks repetitive sequences. Positively hybridizing phage were plaque purified. Recombinant phage were grown, and the DNA was isolated as previously described (38). Restriction endonuclease mapping was performed using standard single and double digestions electrophoresed in one dimension in agarose gels. In addition, analysis of double digests by two-dimensional electrophoresis in low melting temperature agarose gels was performed (36).

DNA Sequence Analysis—*Eco*RI restriction fragments of several recombinant phage were subcloned into pUC8 (39). Plasmid DNA was isolated by CsCl/ethidium bromide density gradient centrifugation (38). The human DNA inserts were released by *Eco*RI digestion and purified electrophoretically. Digestion of this DNA with *Sau*3AI, *Hae*III, *Alu*I, *Rsa*I, etc. generated restriction fragments which were cloned in either orientation into M13mp10 or M13mp11 vectors (40). Recombinant M13 clones were screened for the presence of ADH exon sequences by dot blot anlaysis using various hybridization probes. The ^{32}P-labeled nick-translated β ADH cDNA insert was used to detect M13 clones containing exons 7, 8, and 9. Knowledge of the entire amino acid sequence of β ADH (32) allowed the construction of synthetic 14-mer oligonucleotides specific for amino acid residues 2–6, 7–11, and 86–100 as well as a 17-mer specific for residues 15–20. The partially purified oligonucleotides were purchased from the Department of Chemistry (University of California, San Diego) and were further purified by polyacrylamide gel electrophoresis. They were labeled at their 5′ termini by T4 polynucleotide kinase (41) and used as hybridization probes to detect M13 clones containing exons 1, 2, and 4. M13 clones showing positive hybridization to any of these probes were subjected to DNA sequence analysis carried out by the dideoxynucleotide chain termination method of Sanger *et al.* (42). The remaining exons (3, 5, and 6) were located by random sequencing. For some clones sequencing was carried out using the aforementioned oligonucleotides as the primer, as well as an additional one encoding β ADH amino acid residues 333–337 (purchased from Applied Biosystems, Foster City, CA).

Genomic DNA Analysis—Human genomic DNA was isolated from fresh leukocytes as described (43). 10 μg of DNA was restricted with *Eco*RI or *Hin*dIII, subjected to agarose gel electrophoresis, and transferred to nitrocellulose filters by the method of Southern (44). The 1.0-kb β ADH cDNA insert of pADH12 (35), the 2.2-kb *Eco*RI fragment of λ ADH7 (α ADH gene, Fig. 1), and the 0.5-kb *Eco*RI-*Pst*I fragment of λ ADK15 (β ADH gene exon 1 probe, Fig. 4) were electrophoretically purified, ^{32}P-labeled by nick translation, and used as hybridization probes. A low stringency wash (38) of 1 × SSC (SSC = 0.15 M NaCl, 0.015 M Na citrate) at 65 °C was performed in order to allow cross-hybridization of the α and β ADH probes to other ADH genes. A *Hin*dIII digest of λ bacteriophage DNA served as a molecular size standard.

S1 Nuclease Mapping—A 200-bp *Xba*I-*Hin*fI DNA fragment and a 239-bp *Xba*I-HgiAI DNA fragment containing portions of the first exon as well as 166 bp of 5′-flanking sequence were radiolabeled at their 5′ termini using T4 polynucleotide kinase and used for S1 nuclease protection experiments (46). Total cellular RNA was isolated from post-mortem human adult liver by the method of Chirgwin *et al.* (45). Either 100 μg of human adult liver total RNA or 100 μg of yeast tRNA as a control were incubated with the labeled probes under stringent hybridization conditions at 52 °C for 3 h in 40 μl of 80% formamide, 1 mM EDTA, 0.4 M NaCl, and 40 mM PIPES, pH 6.4.

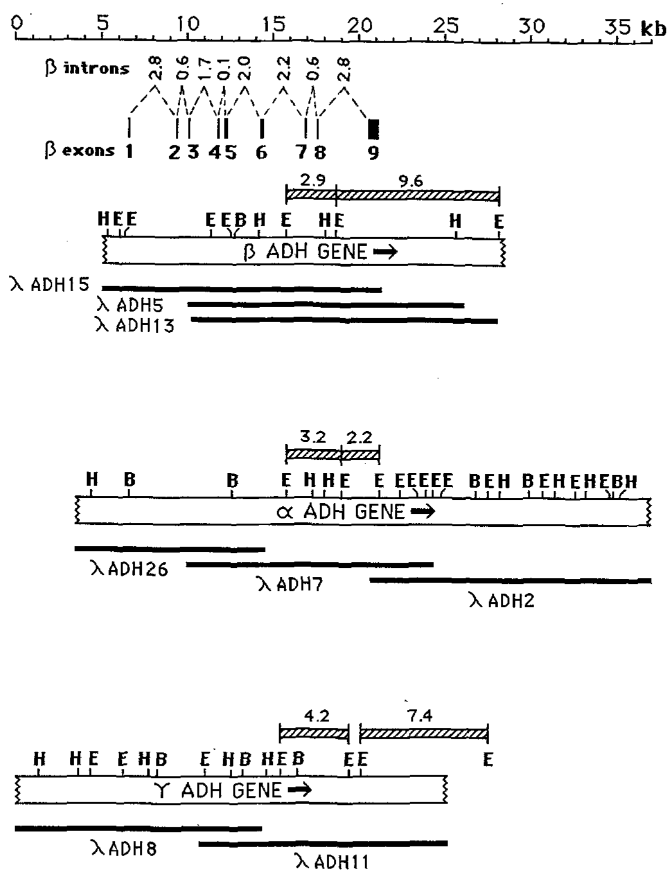

FIG. 1. **Restriction maps of human Class I ADH genes.** The λ genomic clones encoding the α, β, and γ ADH genes are indicated as *solid bars* below the *open bars* which indicate three regions in the human genome containing the genes. The locations of *Eco*RI (*E*), *Hin*dIII (*H*), and *Bam*HI (*B*) sites are shown as well as *arrows* which indicate the direction of transcription which is inferred from the nucleotide sequence data shown in Fig. 2. The locations of the nine exons of the β ADH gene are shown, as are the sizes of the eight introns. The *cross-hatched bars* indicate *Eco*RI fragments (sizes in kb) which hybridize to a β ADH cDNA encoding exons 7, 8, and 9; the 7.4-kb *Eco*RI fragment from the γ ADH gene was detected by Southern blot hybridization of human genomic DNA, and we presume that it is truncated to 5.0 kb in λ ADH11 which does not extend far enough in the 3′ direction to contain all 7.4 kb. An exon 1 probe derived from the β ADH gene hybridizes to λ ADH26 and λ ADH8 (data not shown), indicating that these genomic clones contain the 5′-flanking regions of the α and γ ADH genes, respectively. The size scale is in kb and applies to all three genes.

The hybrids were digested with 500 units/ml S1 nuclease for 30 min at 37 °C in a 400-μl reaction containing 50 mM sodium acetate (pH 4.6), 150 mM NaCl, and 1 mM $ZnCl_2$. Nuclease-resistant DNA was recovered by ethanol precipitation, fractionated on an 8% polyacrylamide, 8 M urea gel, and visualized by autoradiography.

RESULTS AND DISCUSSION

Cloning the Class I ADH Multigene Family—We previously reported the characterization of a cDNA (pADH 12) encoding the carboxyl-terminal 91 amino acids of human β ADH (35). Using pADH 12 as a hybridization probe, we screened a human genomic library (36) and selected three clones. Five additional genomic clones were selected in a chromosomal walk using a 1.3-kb *Eco*RI fragment from the left end of λ ADH5 (Fig. 1). One of these clones, λ ADH26, was isolated from a genomic library prepared from a human-hamster cell line containing human chromosome 4 which is known to contain the human Class I ADH genes (30, 47). These clones

ADH	λ	
		321 ... 330 ... 14-mer
β	ADH5	Gly Phe Lys Ser Lys Glu [Gly Ile] Pro Lys Leu Val Ala Asp Phe Met Ala 5'-TCTTCTTTTCAGGC TTT AAG AGT AAA GAA [GGT ATC] CCA AAA CTT GTG GCT GAT TTT ATG GCT * ** *
α	ADH7	Gly Phe Lys Ser Lys Glu [Cys Val] Pro Lys Leu Val Ala Asp Phe Met Ala 5'-TCTTCTTTTCAGGC TTT AAA AGT AAA GAA [TGT GTC] CCA AAA CTT GTG GCT GAT TTT ATG GCT * ** *
γ	ADH11	Gly Phe Lys Ser Lys Glu [Ser Val] Pro Lys Leu Val Ala Asp Phe Met Ala 5'-TCTTCTTTTCAGGC TTT AAG AGT AAA GAA [TCT GTC] CCA AAA CTT GTG GCT GAT TTT ATG GCT * ** *

FIG. 2. **Positive identification of the three Class I ADH genes.** The 2.9-, 3.2-, and 4.2-kb *Eco*RI fragments of λ ADH5, λ ADH7, and λ ADH11, respectively (see Fig. 1), were cloned into M13mp11 or pUC8, and short DNA sequences were obtained using a 14-mer encoding amino acid residues 333–337 as the primer. All three genes possess an intron at residue 321, and the conserved acceptor AG sequence is *underlined*. The nucleotides found to differ among the three genes are indicated with *asterisks*. The two amino acid residues found to differ are *boxed*. The amino acid sequences for α, β, and γ ADH predicted from these DNA sequences match perfectly with the corresponding sequences obtained from the proteins (32–34).

were mapped with several restriction enzymes and found to span three distinct loci in the human genome (Fig. 1). The β ADH probes were, therefore, able to cross-hybridize to two other ADH genes. Limited DNA sequence analyses performed on λ ADH5, λ ADH7, and λ ADH11 indicated that all three ADH genes possess an intron at amino acid residue 321 and that amino acid differences occur at residues 327 and 328 (Fig. 2), where substitutions peculiar for α, β, and γ ADH have previously been ascertained by amino acid sequence analyses (32–34). Thus, cloned DNA fragments spanning 23 kb around the β ADH gene, 33 kb around the α ADH gene, and 25 kb around the γ ADH gene were obtained in the genomic screenings (Fig. 1). The restriction maps show no overlap between the three sets of ADH clones (Fig. 1) indicating that the genes, if linked, are greater than 10 kb apart.

Southern blot analysis of human genomic DNA digested with *Eco*RI and hybridized to pADH12 (β ADH cDNA) showed six fragments (Fig. 3*A*). The 2.2-kb *Eco*RI fragment of λ ADH 7 (α ADH gene) was purified, and when used as a hybridization probe it detected three of the *Eco*RI fragments previously detected by the β ADH probe (Fig. 3*B*). Five of these fragments are identical in size to *Eco*RI fragments mapped in the three sets of genomic clones, and the sixth (7.4 kb) corresponds to the rightward edge of λ ADH11, where a truncated 5.0-kb *Eco*RI fragment is seen (Fig. 1). Confirmation that the *Eco*RI fragments seen in the genome and in the clones match was obtained by Southern blot hybridization of the *Eco*RI-digested clones using the β and α ADH probes (data not shown). The α, β, and γ ADH genes each contain two of the six *Eco*RI fragments detected by the β cDNA probe (and one of the three detected by the α probe), thus accounting for all fragments detected in the genome. Also, *Eco*RI (Fig. 3*C*) and *Hin*dIII (Fig. 3*D*) digests of human genomic DNA probed with a DNA fragment derived from exon 1 of the β ADH gene exhibit in each case only three hybridizing fragments. The sizes of these *Eco*RI and *Hin*dIII fragments match those seen in Southern blots of *Eco*RI or *Hin*dIII digests of the α, β, and γ genomic clones hybridized to the exon 1 probe (data not shown). Since all hybridizing fragments seen in the genomic DNA blots were accounted for in the genomic clones, there exist only three Class I ADH genes in the human genome, and the existence of pseudogenes is unlikely unless they failed to cross-hybridize under these conditions. Also, there was evidently no cross-hybridization of the Class I ADH genes to either Class II or III ADH genes under these conditions.

Even though the Class I ADH genes failed under these hybridization conditions to cross-hybridize to the Class II or III ADH genes, the three classes of ADH are related. *ADH1*, *ADH2*, and *ADH3* (encoding Class I ADH) and *ADH5* (encoding Class III ADH) are all located on the long arm of chromosome 4 in the region 4q21 to 4q24 (30, 47) suggesting that the Class I and Class III ADH genes form a closely linked multigene family. In addition, monoclonal antibodies which react with all three classes of ADH have been identified (47). Thus, it seems likely that the three classes of human ADH were all derived from a common ancestral gene. Perhaps lower stringency hybridization conditions would allow the detection of genomic DNA fragments bearing Class II or III ADH genes using Class I ADH gene probes.

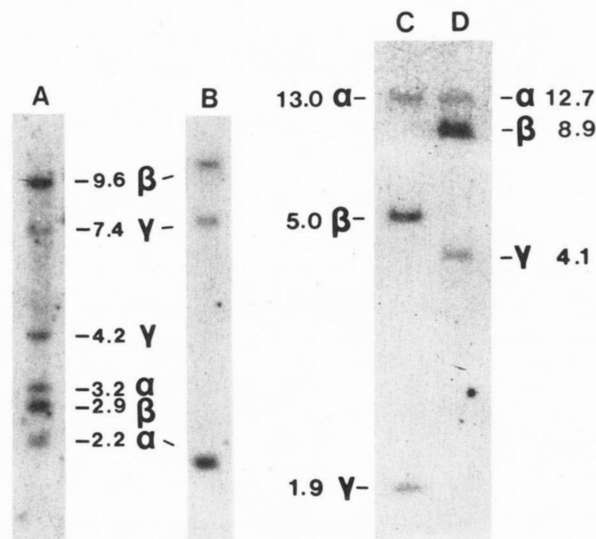

FIG. 3. **Analysis of ADH genomic organization.** *Lanes A, B,* and *C* contain 10 μg of human genomic DNA digested with *Eco*RI; *lane D* contains 10 μg of *Hin*dIII-digested human genomic DNA. Electrophoresis on a 1% agarose gel was carried out, followed by Southern blot hybridization to the β ADH cDNA insert of pADH12 (*lane A*), the 2.2-kb *Eco*RI fragment of the α ADH gene (*lane B*), or a 0.5-kb *Eco*RI/*Pst*I fragment from exon 1 of the β ADH gene (*lanes C* and *D*). A low stringency wash was performed (1 × SSC at 65 °C) to allow cross-hybridization of these probes to other ADH genes. The numbers refer to lengths in kb. *Lane B* was electrophoresed for a greater length of time than *lane A*. The bands are labeled with α, β, or γ to indicate the ADH gene from which the fragment is derived as deduced by inspection of the *Eco*RI and *Hin*dIII maps of the ADH genomic clones (Fig. 1). The most intense bands in each blot correspond to the gene from which the probe was derived whereas the less intense bands correspond to cross-hybridizing ADH genes.

Fine Structure of the β ADH Gene—To determine the intron/exon structure of a human ADH gene and localize the promoter regulatory region, the genomic clone λ ADH15,

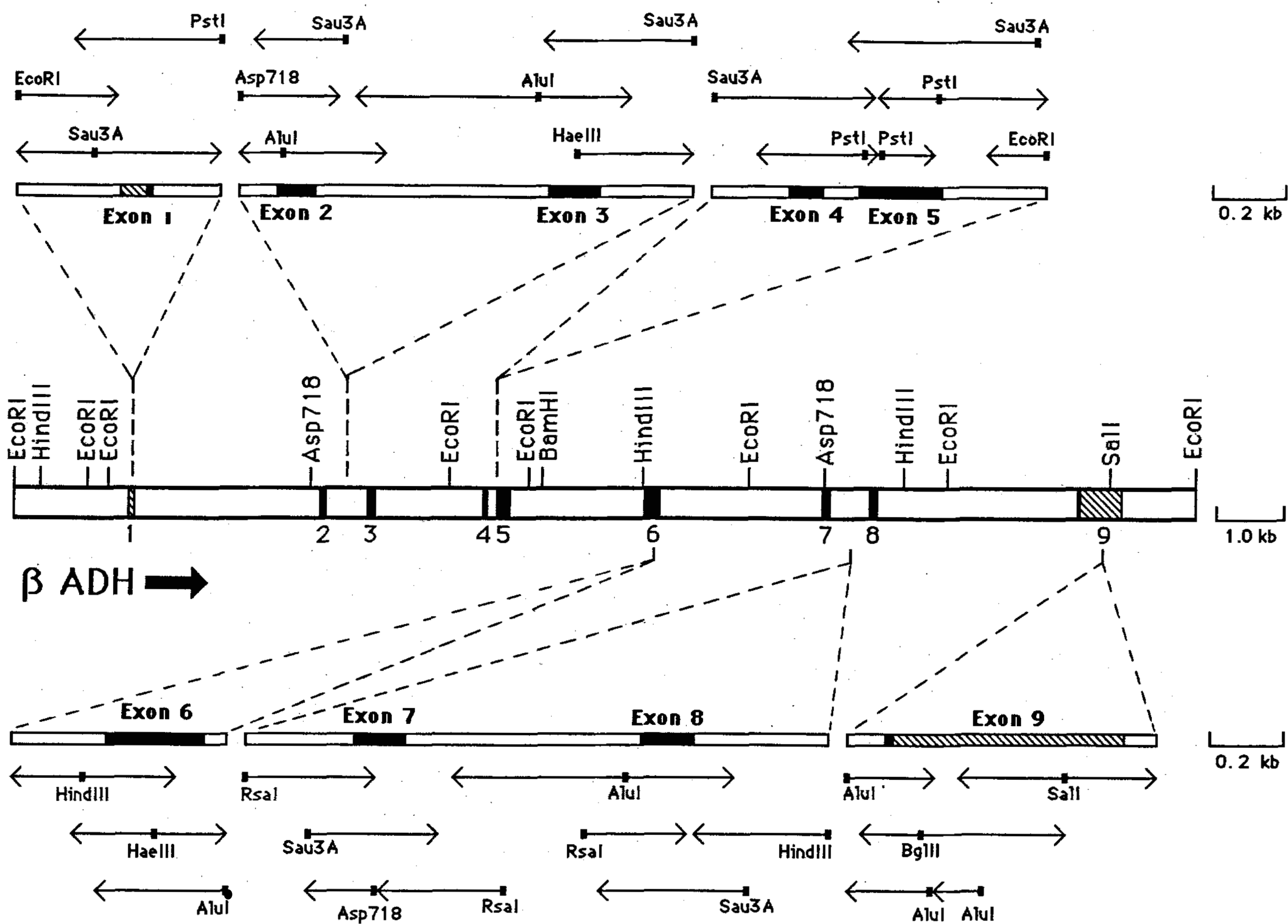

FIG. 4. **Restriction map and sequencing strategy for the β ADH gene.** The *thick bar* (1.0-kb scale) indicates the restriction map of λ ADH15 which encodes human β ADH. The *thin bars* (0.2-kb scale) indicate expanded views of the regions near the nine exons, and the *arrows* indicate the sequencing strategy. The *solid bars* indicate the translated regions of the nine exons, and the *cross-hatched bars* indicate 5′- and 3′-untranslated regions.

which contains the entire β ADH gene, was examined in detail. Exon-containing M13 clones were sequenced using the dideoxynucleotide chain termination method of Sanger *et al.* (42) as indicated in Fig. 4. The DNA sequence analysis revealed that the β ADH gene spans 15 kb and is broken into 9 exons separated by 8 introns, ranging in size from 97 bp to 2.8 kb (Fig. 1). This procedure yielded DNA sequences for all the exons, the intron/exon junctions, and the 5′- plus 3′-flanking regions (Fig. 5).

The region excoding β ADH was found to contain 374 codons preceded by an initiation methionine codon, thus indicating that human and horse ADH share an equal number of amino acids. This is in contradiction to a previous report (32), which indicated that the human β ADH protein subunit contained 373 amino acids. However, this situation was resolved by further amino acid sequence analysis of β ADH (48) which indicates (as we do) that there exists tandem arginine residues at positions 128 and 129, rather than a single arginine as determined previously. Other than the single arginine initially missed by amino acid sequence analysis, the β ADH amino acid sequences determined from the protein and gene match perfectly. The genomic DNA sequence also perfectly matches the β ADH cDNA sequence (35) with the exception of an additional cytosine located 13 bp upstream of the polyadenylation site in the genomic sequence. The presence of an arginine residue at position 47 indicates that the gene encodes the β_1 allele; β_1 and β_2 were previously determined to differ only at position 47, where arginine and histidine, respectively, were found (31).

The nucleotide sequence around the initiation methionine codon, GACATGA (Fig. 5), is similar to that found at most functional initiator codons (49). A polyadenylation signal AATAAA (50) is located 25 bp upstream from the site of polyadenylation as deduced from the β ADH cDNA sequence

FIG. 5. **DNA sequence of the human β ADH gene.** The DNA sequence of all nine exons is shown with the predicted 374 amino acid residues of β ADH indicated. The approximate sizes of the eight introns are indicated, and the conserved GT and AG splice donor and acceptor sites are *underlined.* The ATG translation start codon and TGA stop codon are indicated. The presumptive promoter elements are *boxed,* and a GGAATT repeat is *overlined.* The polyadenylation signal AATAAA is *underlined,* and the site of poly(A) addition is indicated with an *arrow.*

```
                                -272                        enhancer core                 GRE - I               -200
                                GAATTCCAGAGGCCGGGGGGGGGTGGGAAGTGAGGAAAAGAGAAAGTGATTACAATTTATCACTTTAACTTAA
              GRE - II                   CAAT box                                                              -100
TATTTAAACTAATGAAAACAAAATCTTATCTAGAATTTGGAAGTCAATATTTTGATTGCTGGTTCAGTACCCTTTTATCTGTTTTGACAGTCTGGGAATA
                                                                                      TATA box                  +1
ATCCAGTGGGTGTGGCTTAAAGACATAGATCACGTGTGGAATTGGAATTGGATGTTACACAAGCAAACAAAATAAATATCTGTGCAATATATCTGCTTTA
                                                                        1
                                                                    Met Ser Thr Ala Gly Lys
TGCACTCAAGCAGAGAAGAAATCCACAAAGACTCACAGTCTGCTGGTGGGCAGAGAAGACAGAAACGACATG AGC ACA GCA GGA AAA GTAAGCA
                                                                    +++
                                                   10                                  20
                               Val Ile Lys Cys Lys Ala Ala Val Leu Trp Glu Val Lys Lys Pro
AAAAATAT....2.8 kb....TGTTTGATTTTCTAG GTA ATC AAA TGC AAA GCA GCT GTG CTA TGG GAG GTA AAG AAA CCC
                              30
Phe Ser Ile Glu Asp Val Glu Val Ala Pro Pro Lys Ala Tyr Glu Val Arg Ile Lys
TTT TCC ATT GAG GAT GTG GAG GTT GCA CCT CCT AAG GCT TAT GAA GTT CGC ATT AAG GTGAAATATTTTTCC..0.6 kb..
                40                                 50                                  60
                Met Val Ala Val Gly Ile Cys Arg Thr Asp Asp His Val Val Ser Gly Asn Leu Val Thr Pro
CTCTTTATTCTGTAG ATG GTG GCT GTA GGA ATC TGT CGC ACA GAT GAC CAC GTG GTT AGT GGC AAC CTG GTG ACC CCC
                              70                                      80
Leu Pro Val Ile Leu Gly His Glu Ala Ala Gly Ile Val Glu Ser Val Gly Glu Gly Val Thr Thr Val Lys Pro
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
                                                          90
G                                               ly Asp Lys Val Ile Pro Leu Phe Thr Pro Gln Cys Gly Lys
G GTACAGGATTCACAC...1.7 kb...TTTATTCCTCTCCAG GT GAT AAA GTC ATC CCG CTC TTT ACT CCT CAG TGT GGA AAA
100                                110
Cys Arg Val Cys Lys Asn Pro Glu Ser Asn Tyr Cys Leu Lys Asn As
TGC AGA GTT TGT AAA AAC CCG GAG AGC AAC TAC TGC TTG AAA AAT GA GTAGGTTTCTGATGC...67 bp...TATTGCACTG
                          120                                  130
      p Leu Gly Asn Pro Arg Gly Thr Leu Gln Asp Gly Thr Arg Arg Phe Thr Cys Arg Gly Lys Pro Ile His
CCCAG T CTA GGC AAT CCT CGG GGG ACC CTG CAG GAT GGC ACC AGG AGG TTC ACC TGC AGG GGG AAG CCC ATT CAC
    140                                150                                     160
His Phe Leu Gly Thr Ser Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Asn Ala Val Ala Lys Ile Asp Ala Ala
CAC TTC CTT GGC ACC AGC ACC TTC TCC CAG TAC ACG GTG GTG GAT GAG AAT GCA GTG GCC AAA ATT GAT GCA GCC
                            170                                 180
Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly Phe Ser Thr Gly Tyr Gly Ser Ala Val Asn Val Ala Lys
TCG CCC CTG GAG AAA GTC TGC CTC ATT GGC TGT GGA TTC TCG ACT GGT TAT GGG TCT GCA GTT AAC GTT GCC AAG
                                               190                                   200
                                           Val Thr Pro Gly Ser Thr Cys Ala Val Phe Gly Leu Gly Gly
GTGAGAATGACAATG... 2.0 kb ...TTTTTGAAAACACAG GTC ACC CCA GGC TCT ACC TGT GCT GTG TTT GGC CTG GGA GGG
                            210                                 220
Val Gly Leu Ser Ala Val Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Ala Val Asp Ile Asn Lys Asp
GTC GGC CTA TCT GCT GTT ATG GGC TGT AAA GCA GCT GGA GCA GCC AGA ATC ATT GCG GTG GAC ATC AAC AAG GAC
        230                                 240                                     250
Lys Phe Ala Lys Ala Lys Glu Leu Gly Ala Thr Glu Cys Ile Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu
AAA AAA GCA AAG GCC AAA GAG TTG GGT GCC ACT GAA TGC ATC AAC CCT CAA GAC TAC AAG AAA CCC ATC CAG GAA
                            260                                 270
Val Leu Lys Glu Met Thr Asp Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg Leu Asp Thr Met
GTG CTA AAG GAA ATG ACT GAT GGA GGT GTG GAT TTT TCG TTT GAA GTC ATC GGT CGG CTT GAC ACC ATG GTATGTA
                                              280                                       290
                                  Met Ala Ser Leu Leu Cys Cys His Glu Ala Cys Gly Thr Ser Val Ile
CCATGACA...2.2 kb...TTCACTTTATTCCAG ATG GCT TCC CTG TTA TGT TGT CAT GAG GCA TGT GGC ACA AGC GTC ATC
                                300                                 310
Val Gly Val Pro Pro Ala Ser Gln Asn Leu Ser Ile Asn Pro Met Leu Leu Leu Thr Gly Arg Thr Trp Lys Gly
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
        320                                                                                     330
Ala Val Tyr Gly G                                          ly Phe Lys Ser Lys Glu Gly Ile Pro Lys
GCT GTT TAT GGT G GTATGTAGTTGGGCT...0.6 kb...CTGTCTTCTTTTCAG GC TTT AAG AGT AAA GAA GGT ATC CCA AAA
                                    340                                 350
Leu Val Ala Asp Phe Met Ala Lys Lys Phe Ser Leu Asp Ala Leu Ile Thr His Val Leu Pro Phe Glu Lys Ile
CTT GTG GCT GAT TTT ATG GCT AAG AAG TTT TCA CTG GAT GCG TTA ATA ACC CAT GTT TTA CCT TTT GAA AAA ATA
                360
Asn Glu Gly Phe Asp Leu Leu His Ser Gly Lys Se                                              r Ile Arg
AAT GAA GGA TTT GAC CTG CTT CAC TCT GGG AAA AG GTAGATTTTAAGTTA...2.8 kb...TCTTTCCTATTGCAG T ATC CGT
370
Thr Val Leu Thr Phe
ACC GTC CTG ACG TTT TGAGGCAATAGAGATGCCTTCCCCTGTAGCAGTCTTCAGCCTCCTCTACCCTACAAGATCTGGAGCAACAGCTAGGAAA
                    ***
TATCATTAATTCAGCTCTTCAGAGATGTTATCAATAAATTACACATGGGGGCTTTCCAAAGAAATGGAAATTGATGGGAAATTATTTTTCAGGAAAATTT
AAAATTCAAGTGAGAAGTAAATAAAGTGTTGAACATCAGCTGGGGAATTGAAGCCAACAAACCTTCCTTCTTAACCATTCTACTGTGTCACCTTTGCCAT
TGAGGAAAAATATTCCTGTGACTTCTTGCATTTTTGGTATCTTCATAATCTTTAGTCATCGAATCCCAGTGGAGGGGACCCTTTTACTTGCCCTGAACAT
ACACATGCTGGGCCATTGTGATTGAAGTCTTCTAACTCTGTCTCAGTTTTCACTGTCGACATTTTCCTTTTTCTAATAAAAATGTACCAAATCCCTGGGG
TAAAAGCTAGGGTAAGGTAAAGGATAGACTCACATTTACAAGTAGTGAAGGTCCAAGAGTTCTAAATACAGGAAATTTCTTAGGAACTCAAATAAAATGC
CCCACATTTTACTACAGTAAATGGCAGTGTTTTTATGACTTTTATACTCATTTCTTTATGGTCGATATACAATTGATTTTTAAA
                ↑
```

FIG. 5

(35). In the genomic β ADH gene, an additional 70 bp of sequence beyond the polyadenylation site was obtained (Fig. 5). The 14 bp immediately following the poly(A) addition site was found to be similar to the consensus sequence for Class II polyadenylation signals which have been proposed by Berget (51). Each of the eight introns (Fig. 5) are bounded by the GT/AG sequences which are universally found at all intron/exon junctions (52).

Localization of the Promoter for the β ADH Gene—A putative promoter for the β ADH gene was identified by S1 mapping using two 5′-labeled DNA fragments (Fig. 6). The protected DNA fragments of 34 and 73 nucleotides in length obtained from the *Hin*fI and *Hgi*AI fragments, respectively, both suggest that the adenine labeled +1 (Fig. 5) is a putative transcription initiation site. This, indicates that there is a 70-bp 5′-untranslated region upstream from the translation initiation codon. It is very unlikely that the end point identified

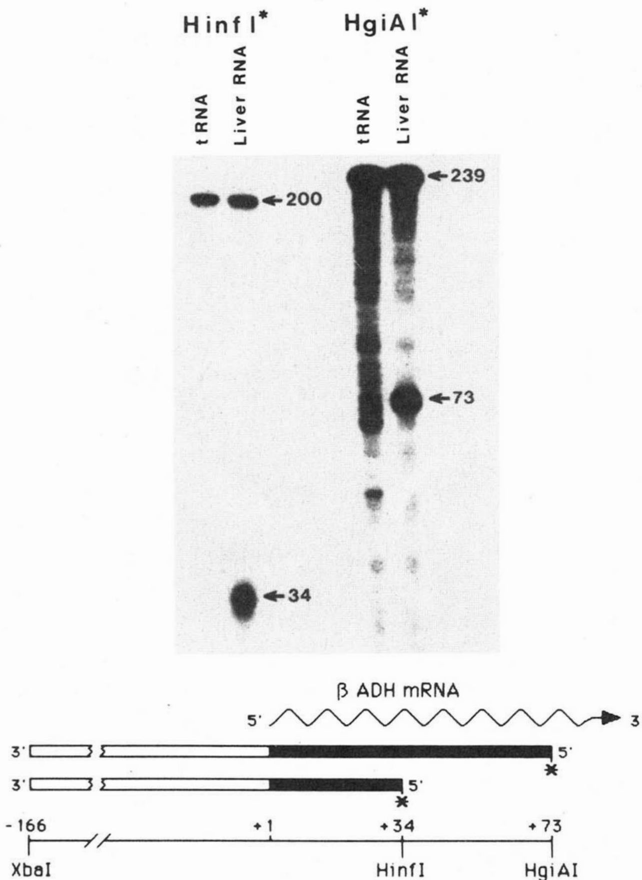

FIG. 6. **S1 mapping of the transcription start site.** The diagram shows the two ^{32}P-end-labeled probes used for S1 mapping (the numbering is the same as that in Fig. 5). The probes were hybridized to either 100 μg of human liver total RNA or 100 μg of yeast tRNA as a control. After S1 nuclease digestion the products were fractionated on an 8% polyacrylamide, 8 M urea gel. The liver RNA protected a 34-base region within the 200-base *Xba*I-*Hin*fI fragment and a 73-bpase region within the 239-base *Xba*I-*Hgi*AI fragment. A *Sau*3AI digest of M13mp10 (replicative form), radiolabeled by a Klenow fill-in reaction, served as a molecular size standard. For a more accurate size determination the products were run on 1-meter sequencing gels next to dideoxynucleotide sequence reactions of M13mp10. The S1 nuclease products of the *Hin*fI-labeled DNA fragment were resolved into a triplet with lengths of 32, 33, and 34 bases; the *Hgi*AI-labeled DNA fragment produced a triplet with lengths of 71, 72, and 73 bases (data not shown). This may represent a set of adjacent mRNA initiation sites, but degradation of the largest band cannot be ruled out.

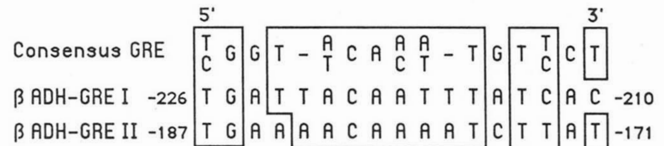

FIG. 7. **GRE sequences.** The consensus GRE is derived from DNA sequences in mouse mammary tumor virus and the human metallothionein gene which are known to bind a glucocorticoid hormone-receptor complex (60). The two putative β ADH GRE sequences (Fig. 5) are aligned with the consensus sequence and matches are *boxed.*

by S1 mapping represents an RNA splicing site for an intron in the 5′-untranslated region since it lacks homology to the consensus sequence for acceptor splice signals (52). The identification of this site as a putative transcription initiation site does not rule out the existence of additional initiation sites further upstream. The possibility of such additional sites has not been examined.

Located 27 bp upstream of the transcription initiation site is the sequence TAAATAT which matches closely with the consensus TATA box sequence found associated with most eukaryotic promoters (53). The sequence AGTCAATAT located at position −158 (Fig. 5) has a seven out of nine match with the consensus CCAAT box sequence found associated with some eukaryotic promoters (54). This element is usually found near positions −60 to −80, but the human γ globin genes contain two CCAAT boxes, one lying at position −120; natural mutations in this γ globin CCAAAT box lead to incorrect developmental expression of γ globin (55, 56). In addition, the sequence GTGGGAAG located at position −250 (Fig. 5) matches seven out of eight with the enhancer core consensus sequence common to most known viral and cellular enhancers (57) and is found embedded in a very purine-rich stretch of DNA (−265 to −227). Regions of purine-pyrimidine asymmetry are apparently required for correct function of the enhancer element of the mouse E_{β} major histocompatibility complex gene (58).

Located in the 5′-flanking region of the human β ADH gene are two tandem sequences with a high degree of homology to previously identified glucocorticoid responsive elements (GRE) found upstream of genes positively activated by a glucocorticoid receptor-hormone complex, *i.e.* the mouse mammary tumor virus long terminal repeat which contains tandem GRE sequences (59) and the human metallothionein-II_A gene (60). In the β ADH promoter these sequences (GRE I and GRE II) lie at positions −226 and −187 (Fig. 5). Both share 75% sequence identity (Fig. 7) with a consensus GRE derived from the mouse mammary tumor virus and metallothionein sequences (60). Since the metallothionein GRE shares only 75% sequence identity with the mouse mammary tumor virus GRE sequences, the sequence homology observed with the putative β ADH GRE sequences is significant. Also, the two putative β ADH GRE sequences are located at positions analogous to the two mouse mammary tumor virus GRE sequences which are located at nucleotides −186 and −129 from the start point of transcription (59). The function of the putative GRE sequences in ADH gene expression remains to be tested.

Acknowledgments—We are grateful to our colleagues Hans Jörnvall, Rolf Bühler, John Hempel, Hedwig von Bahr-Lindström, and Michael Felder for helpful discussions and to Elaine Ito for expert technical assistance. We thank Leon Carlock for providing a human-hamster genomic library used in these studies. We thank Eileen Rice, Terry White, and Lisa Catalfamo for help in preparing the manuscript.

REFERENCES

1. Jörnvall, H. (1970) *Eur. J. Biochem.* **16,** 25–40
2. Eklund, H., Nordström, B., Zeppezauer, E., Söderlund, G., Ohlsson, I., Boiwe, T., Söderberg, B.-O., Tapia, O., Brändén, C.-I., and Åkeson, Å. (1976) *J. Mol. Biol.* **102,** 27–59
3. Brändén, C.-I., Jörnvall, H., Eklund, H., and Furugren, B. (1975) in *The Enzymes* (Boyer, P. D., ed) Vol. XI, pp. 103–190, Academic Press, New York
4. Jörnvall, H. (1977) *Eur. J. Biochem.* **72,** 443–452
5. Bennetzen, J. L., and Hall, B. D. (1982) *J. Biol. Chem.* **257,** 3018–3025
6. Russell, D. W., Smith, M., Williamson, V. M., and Young, E. T. (1983) *J. Biol. Chem.* **258,** 2674–2682
7. Russell, P. R., and Hall, B. D. (1983) *J. Biol. Chem.* **258,** 143–149
8. Thatcher, D. R. (1980) *Biochem. J.* **187,** 875–886
9. Benyajati, C., Place, A. R., Powers, D. A., and Sofer, W. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 2717–2721
10. Benyajati, C., Spoerel, N., Haymerle, H., and Ashburner, M. (1983) *Cell* **33,** 125–133
11. Dennis, E. S., Gerlach, W. L., Pryor, A. J., Bennetzen, J. L., Inglis, A., Llewellyn, D., Sachs, M. M., Ferl, R. J., and Peacock, W. J. (1984) *Nucleic Acids Res.* **12,** 3983–4000
12. Dennis, E. S., Sachs, M. M., Gerlach, W. L., Finnegan, E. J., and Peacock, W. J. (1985) *Nucleic Acids Res.* **13,** 727–743
13. Blair, A. H., and Vallee, B. L. (1966) *Biochemistry* **5,** 2026–2034
14. Strydom, D. J., and Vallee, B. L. (1982) *Anal. Biochem.* **123,** 422–429
15. Smith, M., Hopkinson, D. A., and Harris, H. (1971) *Ann. Hum. Genet.* **34,** 251–271
16. Li, T.-K., and Magnes, L. J. (1975) *Biochem. Biophys. Res. Commun.* **63,** 202–208
17. Pares, X., and Vallee, B. L. (1981) *Biochem. Biophys. Res. Commun.* **98,** 122–130
18. Smith, M., Hopkinson, D. A., and Harris, H. (1973) *Ann. Hum. Genet.* **37,** 49–67
19. Wagner, F. W., Burger, A. R., and Vallee, B. L. (1983) *Biochemistry* **22,** 1857–1863
20. Adinolfi, A., Adinolfi, M., and Hopkinson, D. A. (1984) *Ann. Hum. Genet.* **48,** 1–10
21. Smith, M., Hopkinson, D. A., and Harris, H. (1972) *Ann. Hum. Genet.* **35,** 243–253
22. Smith, M. (1982) in *Advances in Alcohol and Substance Abuse* (Stimmel, B., ed) Vol. 1, pp. 127–146, Haworth Press, New York
23. Stamatoyannopoulos, G., Chen, S. H., and Fukui, M. (1975) *Am. J. Hum. Genet.* **27,** 789–796
24. Teng, Y. S., Jehan, S., and Lie-Injo, L. E. (1979) *Hum. Genet.* **53,** 87–90
25. Harada, S., Misawa, S., Agarwal, D. P., and Goedde, H. W. (1980) *Am. J. Hum. Genet.* **32,** 8–15
26. von Wartburg, J.-P., Papenberg, J., and Aebi, H. (1965) *Can. J. Biochem.* **43,** 889–898
27. Li, T.-K. (1977) *Adv. Enzymol. Relat. Areas Mol. Biol.* **45,** 427–483
28. Streissguth, A. P., Landesman-Dwyer, S., Martin, J. C., and Smith, D. W. (1980) *Science* **209,** 353–361
29. Smith, M., Duester, G., Carlock, L., and Hatfield, G. W. (1984) in *NIDA-NIAAA Technical Review: Genetic and Biological Markers of Drug Abuse and Alcoholism* (Alcohol, Drug Abuse, and Mental Health Administration; Rockville, MD), in press
30. Duester, G., Hatfield, G. W., and Smith, M. (1985) *Alcohol.* **2,** 53–56
31. Jörnvall, H., Hempel, J., Vallee, B. L., Bosron, W. F., and Li, T.-K. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 3024–3028
32. Hempel, J., Bühler, R., Kaiser, R., Holmquist, B., de Zalenski, C., von Wartburg, J.-P., Vallee, B., and Jörnvall, H. (1984) *Eur. J. Biochem.* **145,** 437–445
33. Bühler, R., Hempel, J., Kaiser, R., de Zalenski, C., von Wartburg, J.-P., and Jörnvall, H. (1984) *Eur. J. Biochem.* **145,** 447–453
34. von Bahr-Lindström, H., Höög, J.-O., Heden, L.-O., Kaiser, R., Fleetwood, L., Larsson, K., Lake, M., Holmgren, A., Holmquist, B., Hempel, J., Vallee, B. L., and Jörnvall, H. (1986) *Biochemistry*, in press
35. Duester, G., Hatfield, G. W., Bühler, R., Hempel, J., Jörnvall, H., and Smith, M. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 4055–4059
36. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G., and Maniatis, T. (1978) *Cell* **15,** 1157–1174
37. Benton, W. D., and Davis, R. W. (1977) *Science* **196,** 180–182
38. Maniatis, T., Fritsch, E. F., and Sambrook, J. (eds) (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, New York
39. Vieira, J., and Messing, J. (1982) *Gene (Amst.)* **19,** 259–268
40. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78
41. Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560
42. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
43. Blin, N., and Stafford, D. W. (1976) *Nucleic Acids Res.* **3,** 2303–2308
44. Southern, E. (1975) *J. Mol. Biol.* **98,** 503–517
45. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299
46. Berk, A. J., and Sharp, P. A. (1977) *Cell* **12,** 721–732
47. Smith, M. (1985) in *Advances in Human Genetics* (Harris, H., and Hirschhorn, K., eds) Vol. 17, pp. 249–290, Plenum Press, New York
48. Hempel, J., Holmquist, B., Fleetwood, L., Kaiser, R., Barros-Soderling, J., Bühler, R., Vallee, B. L., and Jörnvall, H. (1985) *Biochemistry* **25,** 5303–5307
49. Kozak, M. (1981) *Nucleic Acids Res.* **9,** 5233–5252
50. Proudfoot, N. J., and Brownlee, G. G. (1976) *Nature* **263,** 211–214
51. Berget, S. M. (1984) *Nature* **309,** 179–182
52. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472
53. Breathnach, R., and Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383
54. Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980) *Nucleic Acids Res.* **8,** 127–142
55. Collins, F. S., Metherall, J. E., Yamakawa, M., Pan, J., Weissman, S. M., and Forget, B. G. (1985) *Nature* **313,** 325–326
56. Gelinas, R., Endlich, B., Pfeiffer, C., Yagi, M., and Stamatoyannopoulos, G. (1985) *Nature* **313,** 323–325
57. Weiher, H., Konig, M., and Gruss, P. (1983) *Science* **219,** 626–631
58. Gillies, S. D., Folsom, V., and Tonegawa, S. (1984) *Nature* **310,** 594–597
59. Scheidereit, C., Geisse, S., Westphal, H. M., and Beato, M. (1983) *Nature* **304,** 749–752
60. Karin, M., Haslinger, A., Holtgreve, H., Richards, R. I., Krauter, P., Westphal, H. M., and Beato, M. (1984) *Nature* **308,** 513–519