THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 257, No. 6, Issue of March 25, pp. 3018-3025, 1982
Printed in U.S.A.

The Primary Structure of the *Saccharomyces cerevisiae* Gene for Alcohol Dehydrogenase I*

(Received for publication, July 14, 1981)

Jeffrey L. Bennetzen‡ and Benjamin D. Hall

From the Department of Genetics, SK-50, Department of Biochemistry, SJ-70, University of Washington, Seattle, Washington 98195

The DNA sequence of the gene for the fermentative yeast alcohol dehydrogenase has been determined. The structural gene contains no introns. The amino acid sequence of the protein as determined from the nucleotide sequence disagrees with the published alcohol dehydrogenase isozyme I (ADH-I) sequence for 5 of the 347 amino acid residues. At least one, and perhaps as many as four, of these differences is probably due to ADH-I protein heterogeneity in different yeast strains and not to sequencing errors. S1 nuclease was used to map the 5′ and 3′ ends of the ADH-I mRNA. There are two discrete, mature 5′ ends of the mRNA, mapping 27 and 37 nucleotides upstream of the translation initiating ATG. These two equally prevalent termini are 101 and 91 nucleotides, respectively, downstream from a TATAAA sequence. Analysis of the 3′ end of ADH-I mRNA disclosed two minor ends upstream of the major poly(A) addition site. These three ends map 24, 67, and 83 nucleotides, respectively, downstream from the translation-terminating TAA triplet. The sequence AATAAG is found 28 to 34 nucleotides upstream of each ADH-I mRNA poly(A) addition site. Sequence comparisons of these three 3′ ends with those for four other yeast mRNAs yielded a 13-nucleotide consensus sequence to which TAAATAA$\begin{matrix}A\\G\end{matrix}$ is central. All of the known yeast poly(A) addition sites map at or near the A residue of a $\begin{matrix}T\\C\end{matrix}$A site 25 to 40 nucleotides downstream from this consensus octamer.

Among the enzymes in yeast which are involved in glycolysis and alcoholic fermentation, alcohol dehydrogenase is the most extensively characterized by genetic means (1-5). Yeast mutants lacking functional alcohol dehydrogenase are readily selected on plates containing allyl alcohol which is converted by alcohol dehydrogenase to the toxic compound acrolein (3, 4). Combined enzymological and genetic studies have established that yeast cells possess three isozymes of alcohol dehydrogenase coded by different structural genes (3, 4). The two cytoplasmic enzymes ADH-I[1] and ADH-II are closely related in amino acid sequence (6) and are active as homo or heterotetramers. They have significant catalytic differences, however, including different K_m values and cooperativity for ethanol metabolism (1, 2, 7). The other enzyme, m-ADH, is associated with mitochondria (8) and does not form heterotetramers with either cytoplasmic isozyme (3). ADH-I, the enzyme responsible for ethanol production from acetaldehyde and NADH, is produced in large amounts (1% or more of total cell protein) in glucose-grown cells. An opposite physiological role, oxidation of ethanol to acetaldehyde, is believed to be carried out by the glucose-repressible enzyme ADH-II. Although ADH-I was originally considered to be a constitutively produced enzyme (9), recent studies by C. Denis and E. T. Young[2] indicate that ADH-I synthesis greatly diminishes as yeast cells adapt from glucose to ethanol as carbon source. The physiology and genetics of the yeast mitochondrial alcohol dehydrogenase are relatively unexamined and largely unknown.

The availability of mutant yeast strains in which all alcohol dehydrogenase activities are lacking (3, 4) has made it possible to isolate the genes coding for these enzymes. From a recombinant plasmid pool containing random fragments of yeast genomic DNA, a segment of DNA was isolated which complemented the mutant defects of a yeast recipient strain lacking both ADH-I and ADH-II (10). Confirmation that the sequence cloned actually codes for the ADH-I protein was obtained by sequencing the region around two closely spaced *Hin*dIII sites. There is a nearly exact correlation between a DNA sequence 360 base pairs in length and that predicted from the amino acid sequence of ADH-I between residues 213 and 332. In this paper, we extend the sequence analysis of the structural gene for ADH-I to the entire protein-coding region as well as 1000 base pairs of the flanking DNA. Together with mapping data on the positions of the 5′ and 3′ termini of ADH-I mRNA, this sequence information provides a preliminary description of the transcription initiation and termination regions of this gene as well as a basis for future experiments to precisely map the sequences essential for initiating and terminating ADH-I mRNA transcription.

EXPERIMENTAL PROCEDURES

Materials—[γ-^{32}P]ATP (3000 Ci/mmol) and some [α-^{32}P]dATP (2000-3000 Ci/mmol) were obtained from New England Nuclear. Most of the [α-^{32}P]dCTP and [α-^{32}P]dATP employed was purchased from Amersham Corp. Restriction endonucleases *Alu* I, *Hha* I, *Hinf* I, *Hpa* II, *Taq* I, and *S1 nuclease* were acquired from Bethesda Research Laboratories; *Ava* II, *Hae* III, *Msp* I, and *Sau* 3A were from New England Biolabs; *Bam* HI and *Hin*dIII were supplied by Boehringer Mannheim. *Acc* I was a gift from R. Roberts (Cold Spring Harbor Laboratories) and R. Gelinas (Fred Hutchison Cancer Research Center). *Escherichia coli* DNA polymerase I (Klenow) was also pur-

* This research was supported by a training grant in Molecular and Cell Biology and by a research grant from the National Institute of General Medical Sciences. 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.

‡ Present address, Department of Biological Sciences, Stanford University, Stanford, California 94305.

[1] The abbreviations used are: ADH-I, alcohol dehydrogenase isozyme I; SSC, standard saline citrate; kb, kilobase; SDS, sodium dodecyl sulfate; ADH-II, alcohol dehydrogenase isozyme II; m-ADH, mitochondrial alcohol dehydrogenase.

[2] C. Denis and E. T. Young, personal communication.

chased from Boehringer Mannheim. T4 Polynucleotide kinase was from Bethesda Research Laboratories and bacterial alkaline phosphatase was from Worthington Biochemicals. pJD14 Plasmid DNA was prepared by a modification of the "clear lysis" procedure (11). pJD14 was constructed by inserting the 3.45-kilobase (kb) *Bam* HI fragment containing the ADH-I structural gene from pY9T6 (10, 12) into the tetracycline resistance gene of pBR322.

Restriction Endonuclease Digestion—Digestions with *Ava* II, *Bam* HI, *Hinc*II, *Hin*dIII, and *Hin*fI were performed in 6.6 mM $MgCl_2$, 66 mM NaCl, and 6.6 mM Tris-HCl, pH 7.4. *Acc* I, *Alu* I, *Hha* I, and *Hpa* II cleavages were done in 6 mM $MgCl_2$, 6 mM NaCl, and 6 mM Tris-HCl, pH 7.5. *Hae* III digestions were in buffer containing 5 mM $MgCl_2$, 50 mM NaCl, and 4 mM Tris-HCl, pH 7.5. *Sau* 3A digestions were performed in 15 mM $MgCl_2$, 60 mM NaCl, and 6 mM Tris-HCl, pH 7.5. *Taq* I digestions were done at 65 °C in 6 mM $MgCl_2$ and 10 mM Tris-HCl, pH 8.4. All digestions also contained 6 mM 2-mercaptoethanol and 100 μg/ml of bovine serum albumin. All cleavage reactions, except *Taq* I digestions, were done at 37 °C for 1–2 h.

^{32}P End Labeling—Terminal DNA labeling of extended 5′ ends produced by restriction enzyme cleavage was performed using [γ-^{32}P] ATP and T4 polynucleotide kinase as described by Richardson (13). Deoxyribonucleoside-5′-[α-^{32}P]triphosphate and *E. coli* DNA polymerase (Klenow) were used to label 3′ DNA termini. DNA (5 μg) restricted with an enzyme leaving 5′ extended sticky ends was incubated in a 50-μl solution of 60 mM NaCl, 7 mM $MgCl_2$, 7 mM Tris-HCl (pH 7.4), 6 mM 2-mercaptoethanol, 100 μg/ml of bovine serum albumin containing 50 μCi (25 pmol) of [α-^{32}P]dCTP or [α-^{32}P]dATP, 1 nmol of the appropriate cold dNTP (if required), and 2.5 units of the *E. coli* DNA polymerase (Klenow). Incubation was for 30 min at 20 °C. Labeling at *Hpa* II and *Taq* I sites was performed using only [α-^{32}P]dCTP and no cold nucleotides. *Hin*dIII and *Hin*fI 3′ ends were labeled in the same manner, but using only [α-^{32}P]dATP, while *Sau* 3A ends were labeled using cold dGTP and [α-^{32}P]dATP. Incorporation of label using Klenow DNA polymerase was consistently 20–70% of the labeled triphosphate added.

Restriction Site Mapping—Cleavage sites for *Alu* I, *Ava* II, *Hae* III, *Hha* I, *Hin*fI, *Hpa* II (*Msp* I), *Sau* 3A, and *Taq* I were visualized on the 3.45-kb *Bam* HI fragment by the partial digestion procedure of Smith and Birnstiel (14). Two nearby *Hin*dIII sites within the ADH-I structural gene were 5′ labeled with [α-^{32}P]dATP and T4 polynucleotide kinase, the enzyme was inactivated by incubation at 65 °C, and the labeled DNA was digested with *Bam* HI. The 2.8-kilobase and 0.5-kilobase fragments containing the ADH-I gene were resolved on an 0.7% horizontal agarose slab gel, electroeluted overnight, purified over a DE52 column, and ethanol precipitated. A small aliquot of the labeled fragment was added to 1 μg of unlabeled pBR322 DNA and cut for 10 min with 1 unit of restriction enzyme. One-half of each sample was analyzed by electrophoresis on a 5% acrylamide gel and the other half on a 1.5% agarose gel. Dried gels were used to expose RP Royal X-omat film.

DNA Sequence Analysis—A 2.55-kb *Bam* HI-*Ava* II fragment of pJD14 containing the entire ADH-I structural gene was prepared by electrophoresis on a 0.7% agarose gel, electroelution, DE52 column purification, and ethanol precipitation. Virtually all sequence analyses, except those done previously (10) from the *Hin*dIII sites, started with this fragment. Generally 5 μg of the 2.55-kb *Bam* HI-*Ava* II fragment were cleaved with a single restriction enzyme, end labeled, and individual fragments resolved on a low percentage (5–8%) acrylamide gel or 1.5% agarose gel. The positions of labeled fragments were identified by autoradiography and these bands were cut out and electroeluted. Electroeluted fragments were chromatographed over DE52 columns, ethanol precipitated, and suspended in 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA. Individual fragments had their two labeled ends separated by gel electrophoresis after cutting with a second restriction enzyme or by denaturation and strand separation on a native acrylamide gel (15). Separated labeled ends were once again purified as described above. Sequencing was done according to Maxam and Gilbert (15) and the products analyzed on acrylamide gels (0.7- or 0.4-mm thick). Standard gels were 39 cm in length. 20% acrylamide gels were electrophoresed at 2000 volts for 3½ or 13½ h. Longer distances from the labeled ends were analyzed on 39-cm 8% acrylamide gels run for 3 h at 2000 volts and on 88-cm 8% gels run for 16 h at 2500 volts.

Transcript Mapping—For mapping the ends of the mature mRNA, a modification of the S1 mapping procedure of Berk and Sharp (16) was employed. A specific, end-labeled restriction fragment encompassing either the 3′ or 5′ end of the mature mRNA was loaded onto a nitrocellulose filter disc. The filter-bound DNA was then hybridized with excess yeast poly(A) RNA under standard conditions (2 h at 65 °C in 4 x SSC and 0.5% SDS with 60 μg of yeast poly(A) RNA in 2 ml). The filter was then washed 4 times with 2 ml of 2 x SSC and added to 0.5 ml of 30 mM sodium acetate (pH 4.6), 0.15 M NaCl, and 1 mM $ZnSO_4$. The solution was brought to 1.5 units/100 μl in S1 nuclease and incubated for 10 min at 37 °C. 100 μl of the S1 solution were then removed into 5 μl of 0.2 M EDTA and ethanol precipitated with 5 mg of carrier yeast tRNA. The rest of the sample was brought up 6 units/100 μl in S1 and incubated a further 10 min at 37 °C. A 100-μl sample was then removed and this process repeated, increasing the S1 concentration 4-fold at each step until no incubation mixture remained.

The ethanol-precipitated samples were resuspended in 0.1 M NaOH and 1 mM EDTA and heated 10 min at 90 °C prior to electrophoretic analysis on denaturing (7 M urea) gels.

The 3′ end was mapped using an approximately 900-nucleotide *Hin*dIII-*Hin*dIII fragment from pJD14 which was 3′ end labeled with Klenow DNA polymerase. The 5′ end of the mRNA was positioned with a kinase-labeled 687-nucleotide *Hin*fI-*Hin*fI fragment.

Containment—All work involving recombinant DNA was carried out according to the National Institutes of Health guidelines for recombinant DNA. Recombinant plasmids were constructed under P2 physical and EK1 biological containment (pBR322) (National Institutes of Health guidelines of April 11, 1979).

RESULTS

Restriction Endonuclease Mapping—The 3.45-kb *Bam* HI fragment from pY9T6 (10) containing the ADH-I gene was inserted into the *Bam* HI site of plasmid pBR322 to form recombinant plasmid pJD14. The same 3.45-kb *Bam* HI fragment inserted into the vector YRp7 (named pJC1) was found to fully complement alcohol dehydrogenase-deficient yeast cells (as in Ref. 10), suggesting no loss of functional material in the subcloning. Northern gel analysis (data not shown) and R-looping experiments[3] disclosed no other major RNA species, besides ADH-I mRNA, coded for by the 3.45-kb *Bam* HI fragment.

Standard restriction mapping of pJD14 disclosed that the 3.45-kb *Bam* HI fragment contained two *Hin*dIII sites, two *Hinc*II sites, and one *Ava* II site. No sites were observed for *Ava* I, *Bgl* II, *Bst* EII, *Cla* I, *Eco* RI, *Hpa* I, *Pst* I, *Sal* I, *Sst* I, *Sst* II, *Xba* I, or *Xho* I. The mapping of cleavage sites for restriction enzymes with four base recognition sequences was done by the partial digestion procedure of Smith and Birnstiel (14) (Fig. 1). Later sequencing results confirmed the presence of all predicted restriction enzyme cleavage sites.

DNA Sequence Determination—Previous partial sequencing of the gene coding for ADH-I (10) had disclosed that the two *Hin*dIII sites in the 3.45-kb *Bam* HI fragment were located in the part of the ADH-I gene which coded for the COOH terminus of the ADH-I protein. By this analysis, the *Ava* II site in the 3.45-kb *Bam* HI fragment was calculated to be over 1.0 kb upstream of the structural gene, assuming there were no introns in the ADH-I gene. This assumption has since been proven correct by transcription mapping (see below). Therefore, the 2.55-kb *Bam* HI-*Ava* II fragment of pJD14 was preparatively purified on an agarose gel and used as the starting material for all subsequent DNA sequence determinations.

Restriction sites used for sequencing and the extent of sequence read from each site are diagrammed in Fig. 1. Whenever possible, both strands of the DNA were sequenced. More than 50% of the DNA sequence presented was read off both strands and more than 80% of the DNA was independently sequenced two or more times. Efforts were made to overlap all restriction sites, especially those outside the gene. The only such sites not overlapped are the two *Taq* I recognition sequences in the *Bam* HI-*Ava* II fragment which are

[3] K. Nasmyth, personal communication.

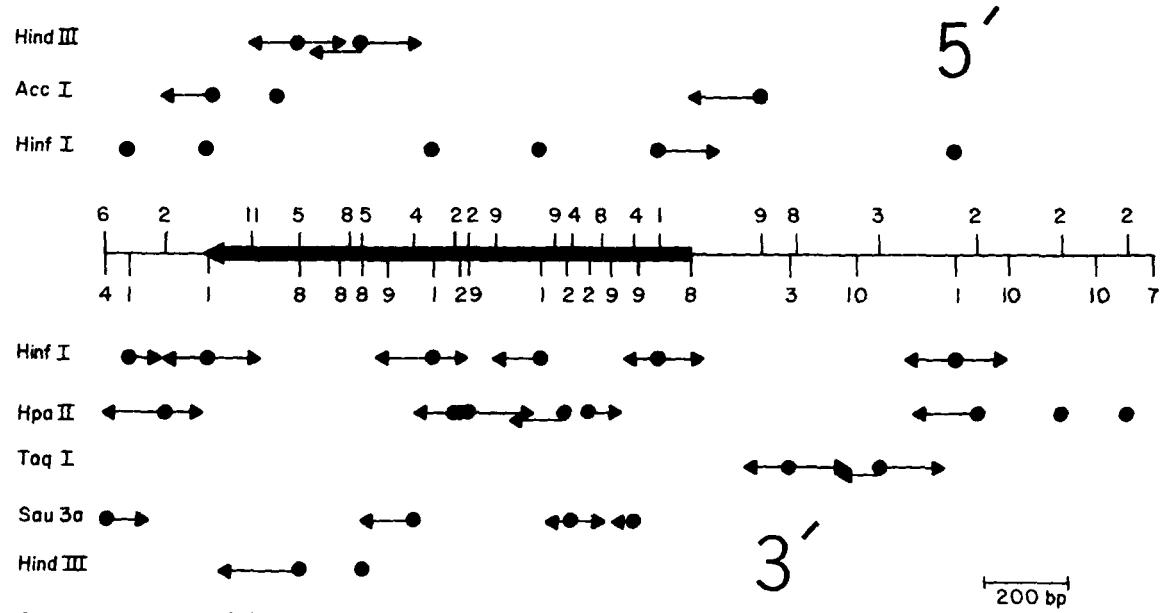

FIG. 1. **Sites and extent of DNA-sequencing experiments.** The *bar graphs* at the center gives a detailed map of all of the restriction sites for 11 of the enzymes on the 2.55-kb *Bam* HI-*Ava* II fragment containing the ADH-I structural gene. The placement of most of these sites has been confirmed by DNA sequencing. *1*, *Hin*fI; *2*, *Hpa* II; *3*, *Taq* I; *4*, *Sau* 3A; *5*, *Hin*dIII; *6*, *Bam* HI; *7*, *Ava* II; *8*, *Alu* I; *9*, *Hae* III; *10*, *Hha* I; *11*, *Hin*cII. Those sites above the bar graph were 5′ end-labeled with kinase and those below were 3′ end-labeled with Klenow DNA polymerase for sequence analysis. ●, Each of the restriction sites for the given enzyme of this 2.55-kb *Bam* HI-*Ava* II fragment; ←●→, direction and extent of DNA sequence read. The *bold arrow* designates the orientation and extent of the mature ADH-I mRNA synthesized in logarithmic, glucose-grown yeast cells.

nearest the ADH-I structural gene and the *Hpa* II site 3′ distal to the gene. If a small *Taq* I-*Taq* I or *Hpa* II-*Hpa* II fragment was missed at any of these sites, complete restriction digestion would indicate that it must be smaller than 30 base pairs (data not shown). Four restriction sites within the gene were not overlapped, the *Sau* 3A site at the sequence coding for amino acids 83 and 84, the *Hin*fI site at nucleotide +40 through +44, and the two *Hin*dIII sites (Fig. 2). The sequence of the region of +40 through +44 is known to be GAATC since labeling of this *Hin*fI site with [α-^{32}P]dATP leads to a doublet pattern on sequencing gels. Multiple sequencing experiments were done from *Taq* I sites and from the *Hin*fI site at the 5′ proximal end of the gene.

A contiguous sequence of 2150 nucleotides was determined, 1041 nucleotides of which code for the 347 amino acids of the mature ADH-I protein (Fig. 2). The nucleotide sequence predicts five amino acid differences from the reported sequence of ADH-I from a laboratory *Saccharomyces cerevisiae* strain (6). The DNA sequence predicts that amino acid 20 is a histidine rather than a tyrosine, 58 is valine rather than threonine, 147 is glutamine instead of glutamic acid, 151 is isoleucine rather than valine, and 236 is aspartic acid instead of asparagine.

There is a severe bias in the choice of amino acid codons in the ADH-I gene. Only 33 of the 61 possible triplets are used, and only 25 of these codons are used extensively (Table I). The proposed basis of this extreme codon bias is discussed in detail in the accompanying paper (17).

The DNA sequence of the regions flanking the 5′ and 3′ ends of the structural gene are decidedly nonrandom.

The 5′ flanking region of the ADH-I gene has a highly purine-rich region (67.5%) at −240 to −350 surrounded by two very long all pyrimidine stretches at −224 to −239 and −373 to −407. Several smaller polypurine and polypyrimidine tracts are seen throughout the promoter proximal region. The first ATG upstream of the actual translation initiating ATG codon is found at −202 to −200. The area between −200 and the start of the protein-coding region is AT rich (65%) but, oddly, has a very long stretch (−104 through −49) with only a single A in the mRNA identical strand. The sequence TATAAA, a proposed transcription initiation signal (Ref. 18), is located at −128 through −123.

Of the first 75 nucleotides after the last amino acid codon, 85% are either A or T (Fig. 2). There are two UGA (umber) nonsense codons, in different reading frames, within the first 22 nucleotides downstream of the ochre termination triplet at the end of the structural gene. Between 10 and 24 nucleotides further downstream, there are three additional ochre terminators, one in each reading frame. All told, there are 10 nonsense triplets between +1060 and +1139. Three repeats of the sequence AATAAG are seen at +1043 through +1048, +1082 to +1087, and +1098 through +1103. This resembles the 5′-AAUAAA-3′ sequence seen toward the 3′ end of a number of eukaryotic mRNAs (18, 19). Three poly(dT) stretches are seen beyond the translation termination site on the 3′ end of the mRNA identical strand. These are located at +1070 through +1074, +1114 to +1117, and +1133 through +1136. Stretches of four or more T's in the RNA identical strand have been found to be a prerequisite for termination by eukaryotic RNA polymerase III (20, 21).

Transcription Mapping—Hybridization to the 3′ end-labeled 0.5-kb *Bam* HI-*Hin*dIII fragment was used to determine the 3′ end of the mRNA. The fragment was denatured, loaded onto nitrocellulose filters, and hybridized to total poly(A) RNA from a logarithmic yeast culture, and the hybrids were digested with S1 nuclease. This modification of the Berk and Sharp (16) S1 mapping method allows the hybridization to be done under standard rather than R-looping conditions. The results demonstrate a major termination site 240 nucleotides downstream from the *Hin*dIII site or 83 nucleotides beyond the translation termination site (Figs. 3 and 4). Two other very minor sites, 24 and 67 nucleotides beyond the ochre stop codon, are also seen. It is not clear whether these minor bands are minor terminators or S1 artifacts, but they are seen even at the lowest S1 concentrations. High S1 nuclease concentrations do produce a large number of artifactual bands by

```
-750                          -720                          -690
  CGATTTTTTTCTAAACCGTGGAATATTTCGGATATCCTTTTGTTGTTTCCGGGTGTACAATAT

                         -660                          -630
  GGACTTCCTCTTTTCTGGCAACCAAACCCATACATCGGGATTCCTATAATACCTTCGTTGGTC

                      -600                          -570
  TCCCTAACATGTAGGTGGCGGAGGGGAGATATACAATAGAACAGATACCAGACAAGACATAAT

                   -540                          -510
  GGGCTAAACAAGACTACACCAATTACACTGCCTCATTGATGGTGGTACATAACGAACTAATAC

                -480                          -450
  TGTAGCCCTAGACTTGATAGCCATCATCATATCGAAGTTTCACTACCCTTTTTCCATTTGCCA

             -420                          -390
  TCTATTGAAGTAATAATAGGCGCATGCAACTTCTTTTCTTTTTTTTTCTTTTCTCTCTCCCCC

          -360                          -330
  GTTGTTGTCTCACCATATCCGCAATGACAAAAAAAAATGATGGAAGACACTAAAGGAAAAAATT

       -300                          -270
  AACGACAAAGACAGCACCAACAGATGTCGTTGTTCCAGAGCTGATGAGGGGTATCTTCGAACA

    -240                          -210
  CACGAAACTTTTTCCTTCCTTCATTCACGCACACTACTCTCTAATGAGCAACGGTATACGGCC

  -180                          -150
  TTCCTTCCAGTTACTTGAATTTGAAATAAAAAAAAGTTTGCCGCTTTGCTATCAAGTATAAATA

-120                           -90                           -60
  GACCTGCAATTATTAATCTTTTGTTTCCTCGTCATTGTTCTCGTTCCCTTTCTTCCTTGTTTC

                             -30                           -1
  TTTTTCTGCACAATATTTCAAGCTATACCAAGCATACAATCAACTATCTCATATACA

  ATG TCT ATC CCA GAA ACT CAA AAA GGT GTT ATC TTC TAC GAA TCC CAC
  Met Ser Ile Pro Glu Thr Gln Lys Gly Val Ile Phe Tyr Glu Ser His
   0   1                                   10

  GGT AAA TTG GAA CAC AAG GAT ATT CCA GTT CCA AAG CCA AAG GCC AAC
  Gly Lys Leu Glu His Lys Asp Ile Pro Val Pro Lys Pro Lys Ala Asn
                 20                                            30

  GAA TTG TTG ATC AAC GTT AAG TAC TCT GGT GTC TGT CAC ACT GAC TTG
  Glu Leu Leu Ile Asn Val Lys Tyr Ser Gly Val Cys His Thr Asp Leu
                                  40

  CAC GCT TGG CAC GGT GAC TGG CCA TTG CCA GTT AAG CTA CCA TTA GTC
  His Ala Trp His Gly Asp Trp Pro Leu Pro Val Lys Leu Pro Leu Val
          50                                            60
```

FIG. 2. **Sequence of the ADH-I gene.** Presented is the complete sequence of a contiguous 2116-nucleotide region encompassing the ADH-I structural gene and over a thousand nucleotides of flanking DNA. The sequence is shown only for the mRNA identical (*i.e.* plus) strand of the gene. The *numbers above the sequence* indicate the number of nucleotides, in each direction, from the A in the translation-initiating ATG. The *numbers below the sequence* denote the amino acid number, starting with 1 as the serine which is the NH_2-terminal amino acid of the mature ADH-I protein (2). The various homopolymer clusters and lack of restriction sites between −50 and −160 made this region exceedingly difficult to sequence accurately. There is a possibility that an extra T or C may have been added to or omitted from this region.

```
  GGT GGT CAC GAA GGT GCC GGT GTC GTT GTC GGC ATG GGT GAA AAC GTT
  Gly Gly His Glu Gly Ala Gly Val Val Val Gly Met Gly Glu Asn Val
                                  70

  AAG GGC TGG AAG ATC GGT GAC TAC GCC GGT ATC AAA TGG TTG AAC GGT
  Lys Gly Trp Lys Ile Gly Asp Tyr Ala Gly Ile Lys Trp Leu Asn Gly
   80                                            90

  TCT TGT ATG GCC TGT GAA TAC TGT GAA TTG GGT AAC GAA TCC AAC TGT
  Ser Cys Met Ala Cys Glu Tyr Cys Glu Leu Gly Asn Glu Ser Asn Cys
                 100                                           110

  CCT CAC GCT GAC TTG TCT GGT TAC ACC CAC GAC GGT TCT TTC CAA CAA
  Pro His Ala Asp Leu Ser Gly Tyr Thr His Asp Gly Ser Phe Gln Gln
                                 120

  TAC GCT ACC GCT GAC GCT GTT CAA GCC GCT CAC ATT CCT CAA GGT ACC
  Tyr Ala Thr Ala Asp Ala Val Gln Ala Ala His Ile Pro Gln Gly Thr
         130                                           140

  GAC TTG GCC CAA GTC GCC CCC ATC TTG TGT GCT GGT ATC ACC GTC TAC
  Asp Leu Ala Gln Val Ala Pro Ile Leu Cys Ala Gly Ile Thr Val Tyr
                                 150

  AAG GCT TTG AAG TCT GCT AAC TTG ATG GCC GGT CAT TGG GTT GCC ATT
  Lys Ala Leu Lys Ser Ala Asn Leu Met Ala Gly His Trp Val Ala Ile
  160                                           170

  TCC GGT GCT GCC GGT GGT CTA GGT TCT TTG GCT GTT CAA TAC GCC AAG
  Ser Gly Ala Ala Gly Gly Leu Gly Ser Leu Ala Val Gln Tyr Ala Lys
                 180                                           190

  GCT ATG GGT TAC AGA GTC TTG GGT ATT GAC GGT GGT GAA GGT AAG GAA
  Ala Met Gly Tyr Arg Val Leu Gly Ile Asp Gly Gly Glu Gly Lys Glu
                                 200

  GAA TTA TTC AGA TCC ATC GGT GGT GAA GTC TTC ATT GAC TTC ACT AAG
  Glu Leu Phe Arg Ser Ile Gly Gly Glu Val Phe Ile Asp Phe Thr Lys
         210                                           220

  GAA AAG GAC ATT GTC GGT GCT GTT CTA AAG GCC ACT GAC GGT GGT GCT
  Glu Lys Asp Ile Val Gly Ala Val Leu Lys Ala Thr Asp Gly Gly Ala
                                 230

  CAC GGT GTC ATC AAC GTT TCC GTT TCC GAA GCC GCT ATT GAA GCT TCT
  His Gly Val Ile Asn Val Ser Val Ser Glu Ala Ala Ile Glu Ala Ser
  240                                           250

  ACC AGA TAC GTT AGA GCT AAC GGT ACC ACC GTT TTG GTC GGT ATG CCA
  Thr Arg Tyr Val Arg Ala Asn Gly Thr Thr Val Leu Val Gly Met Pro
                 260                                           270

  GCT GGT GCC AAG TGT TGT TCT GAT GTC TTC AAC CAA GTC GTC AAG TCC
  Ala Gly Ala Lys Cys Cys Ser Asp Val Phe Asn Gln Val Val Lys Ser
                                 280

  ATC TCT ATT GTT GGT TCT TAC GTC GGT AAC AGA GCC GAC ACC AGA GAA
  Ile Ser Ile Val Gly Ser Tyr Val Gly Asn Arg Ala Asp Thr Arg Glu
         290                                           300

  GCT TTG GAC TTC TTC GCC AGA GGT TTG GTC AAG TCT CCA ATC AAG GTT
  Ala Leu Asp Phe Phe Ala Arg Gly Leu Val Lys Ser Pro Ile Lys Val
                                 310

  GTC GGC TTG TCT ACC TTG CCA GAA ATT TAC GAA AAG ATG GAA AAG GGT
  Val Gly Leu Ser Thr Leu Pro Glu Ile Tyr Glu Lys Met Glu Lys Gly
  320                                           330

                                                          1050
  CAA ATC GTT GGT AGA TAC GTT GTT GAC ACT TCT AAA TAA GCGAATTTCTT
  Gln Ile Val Gly Arg Tyr Val Val Asp Thr Ser Lys
                 340
                    1080                          1110
ATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAAATAAGTGTATACAAATTTTAAAG

                 1140                          1170
TGACTCTTAGGTTTTAAAACGAAAATTCTTGTTCTTGAGTAACTCTTTCCTGTAGGTCAGGTT

            1200                          1230
GCTTTCTCAGGTATAGCATGAGGTCGCTCTTATTGACCACACCTCTACCGGCATGCCGAGCAA

          1260                          1290
ATGCCTGCAAATCGCTCCCCATTTCACCCAATTGTAGATATGCTAACTCCAGCAATGAGTTGA

       1320                          1350
TGAATCTCGGTGTGTATTTTATGTCCTCAGAGGACAATACCTGTTGTAATCCGTCC
```

TABLE I

Codon usage in the ADH-I gene

The data presented were collected from the ADH-I gene sequence presented in Fig. 2.

	U			C			A			G		
U	Phe	UUU	0	Ser	UCU	14	Tyr	UAU	0	Cys	UGU	8
	Phe	UUC	8	Ser	UCC	7	Tyr	UAC	13	Cys	UGC	0
	Leu	UUA	2	Ser	UCA	0						
	Leu	UUG	19	Ser	UCG	0				Trp	UGG	5
C	Leu	CUU	0	Pro	CCU	2	His	CAU	1	Arg	CGU	0
	Leu	CUC	0	Pro	CCC	1	His	CAC	10	Arg	CGC	0
	Leu	CUA	3	Pro	CCA	10	Gln	CAA	9	Arg	CGA	0
	Leu	CUG	0	Pro	CCG	0	Gln	CAG	0	Arg	CGG	0
												0
A	Ile	AUU	9	Thr	ACU	5	Asn	AAU	0	Ser	AGU	0
	Ile	AUC	12	Thr	ACC	9	Asn	AAC	11	Ser	AGC	0
	Ile	AUA	0	Thr	ACA	0	Lys	AAA	4	Arg	AGA	8
	Met	AUG	7	Thr	ACG	0	Lys	AAG	20	Arg	AGG	0
G	Val	GUU	19	Ala	GCU	19	Asp	GAU	2	Gly	GGU	41
	Val	GUC	17	Ala	GCC	16	Asp	GAC	14	Gly	GGC	3
	Val	GUA	0	Ala	GCA	0	Glu	GAA	20	Gly	GGA	0
	Val	GUG	0	Ala	GCG	0	Glu	GAG	0	Gly	GGG	0

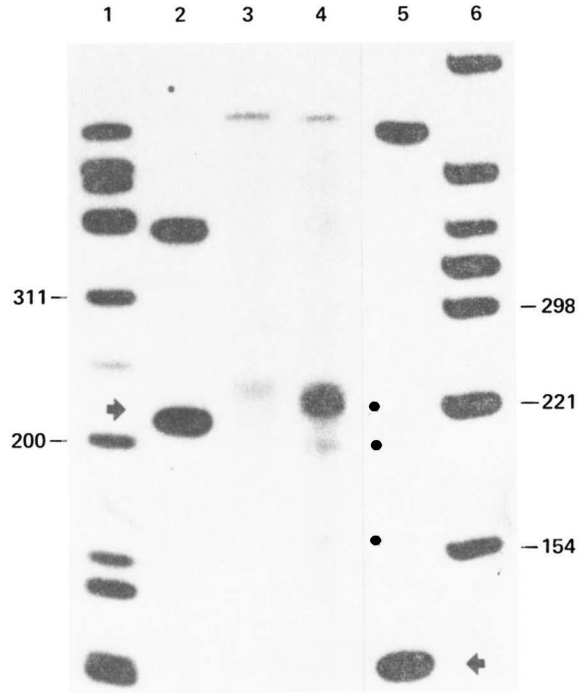

FIG. 3. **Mapping of the 3′ end of ADH-I mRNA.** S1 mapping of the 3′ end of ADH-I mRNA from the *Hin*dIII site at nucleotides 911 through 917 was performed as described under "Experimental Procedures." The gel pictured is 5% acrylamide (20:1, acrylamide/bisacrylamide) containing 7 M urea. All samples were loaded in 0.1 M NaOH and 1 mM EDTA and the gel run at high temperature. *Lane 1*, φX174 DNA restricted with *Hin*fI and end-labeled with Klenow DNA polymerase. The fragment sizes indicated are in base pairs. *Lane 2*, the 900-nucleotide *Hin*dIII-*Hin*dIII fragment used for this S1 mappping cut with *Hin*fI. The *arrow* marks the 210-nucleotide [^{32}P] *Hin*dIII/*Hin*dI fragment which is complementary to ADH-I mRNA. *Lane 3*, 900-nucleotide *Hin*dIII-*Hin*dIII fragment which has been subjected to S1 mapping analysis with a low S1 concentration (15 μ/ml). *Lane 4*, same as *lane 3* but with a higher S1 concentration (60 μ/ml). *Asterisks* mark the three bands which denote ADH-I mRNA 3′ ends. Fragment sizes are given in base pairs. *Lane 5*, 900-base pair *Hin*dIII-*Hin*dIII fragment used for S1 mapping cleaved with *Hin*cII. An *arrow* marks the 119-nucleotide [^{32}P]*Hin*dIII-*Hin*cII fragment complementary to the 3′ end of ADH-I mRNA. *Lane 6*, pBR322 DNA restricted with *Hin*fI and end labeled with Klenow DNA polymerase. Fragment sizes are given in base pairs.

digesting labile regions of RNA/DNA hybrids and DNA/DNA duplexes (data not shown).

Preliminary mapping of the 5′ end of the ADH-I mRNA (data not shown)[4] disclosed two mature ends. Analysis next to a DNA-sequencing gel indicated that these two, approximately equimolar, ends were at nucleotides −27 and −37. Both ends are at the central A of the repeated pentamer 5′-CAAGC-3′ (Fig. 4).

Alignment of seven mapped yeast poly(A) addition sites was performed to try and uncover any common sequences which might signal 3′ end maturation (Fig. 5). For the four lower sequences in Fig. 5, 3′ transcripts have not been mapped; these sequences have been positioned by their homology to the others. An eight-nucleotide block (TAAATAA$\frac{A}{G}$) of highly homologous sequence (at least 5/7 matches) begins 25 to 40 nucleotides upstream of the mapped ends. Included in this region is an AATAAA-like sequence similar to that seen upstream of animal mRNA poly(A) addition sites (18, 19). The exact position of these ends all appear to map at or near the A of a 5′-$\frac{T}{C}$ A-3′ sequence (Fig. 5).

DISCUSSION

The Coding Region—The sequence of the ADH-I structural gene and over a thousand nucleotides of the flanking regions have been determined. Of the 347 amino acid-coding triplets, only five disagreements were found with the ADH-I protein sequence (6). The presence of valine rather than threonine at position 58 would be expected to have the most profound effect on the tertiary structure and function of the enzyme of the differences noted. This valine residue is situated within the hydrophobic core of the protein (22) and may function as well as a threonine in this area. The isoleucine and valine substitution at position 151 is fully conservative and, in fact, improves the homology between yeast ADH-I and several other eukaryotic alcohol dehydrogenases (22). Although these residues may have been misassigned in the original protein sequencing (6), the differences observed may also be a function of strain heterogeneity. Differences between yeast strain have been observed in the use of valine or isoleucine at positions 313 and 338 in ADH-II (6). In this same vein,

[4] G. Ammerer, personal communication.

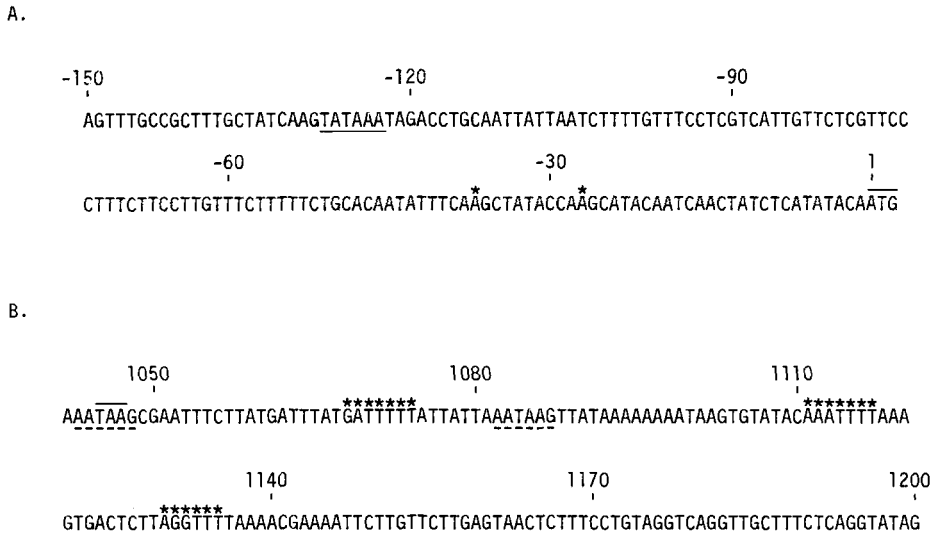

FIG. 4. **Sequence of the 5′ and 3′ ends of the ADH-I structural gene.** The sequence presented encompasses the 150 nucleotides upstream of the TCT triplet which codes for the first amino acid of the mature ADH-I protein (*A*) and the 150 nucleotides downstream from the nucleotides encoding the COOH-terminal lysine in the mature ADH-I protein (*B*). The sequence TATAAA at the 5′ end of the gene is underlined. *Dashes* also underline the repeated sequence AATAAG found at the 3′ end of the gene. The initiation and termination of translation triplets are overlined. *, Observed 5′ and 3′ ends of the ADH-I mRNA (neglecting 5′ capping and 3′ polyadenylation). Sequence shown is the mRNA identical (*i.e.* plus) strand.

ADH1	TGACACTTCTAAATAAGCGAATTTCTTATGATTTATGATTTTTATTATTA
ADH1	TTTTATTATTAAATAAGTTATAAAAAAAATAAGTGTATACAAATTTTAAA
ADH1	GTTATAAAAAAAAATAAGTGTATACAAATTTTAAAGTGACTCTTAGGTTTT
MATa1	TATGAGATCTAAATAAATTCGTTTTCAATGATTAAAATAGCATAGTCGGG
MATα1	CTATGTATTTGTATAAAATATGATATTACTCAGACTCAAGCAAACAATCA
MATα2	TGTGTAGAATATATATATATATATTTCGCAAAAATACATAAACAATCAAC
CYC1	GTTATGTTAGTATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTT
Consensus	T TAAATAA(A/G) T A A T
CYC7	GTTATTTATTAATTAATTATTTTTATATGCATGCACATAAAAAGTCTATA
H3	AATAAAATATAAATCAATATATTTAGGTTTACTGGGTTCTTTAACAGTTG
H4	AATTATAAACATATAAGGGTATTTAATTTAATTGGGTTTAAACAGTTTGA
G3PDH	TCTTGCATTTAAATAAATTTTCTTTTTATAGCTTTATGACTTAGTTTCAA

FIG. 5. **The 3′ end homologies of several yeast genes.** The sequences presented are all for the mRNA identical strand of the DNA. The seven sequences above the consensus sequence are all cases where the poly(A) addition site has been mapped. These 3′ end sequences are *overlined* in the figure. The consensus sequence was determined from at least 5/7 hits for A, T, G, and C or 7/7 hits for $\frac{A}{G}$ or $\frac{T}{C}$. The sequences below the consensus sequence are from yeast genes whose mRNA 3′ ends have not been mapped. Four of these sequences were oriented to optimize homology with the consensus sequence. CYC7 (iso-2 cytochrome *c*) was positioned as shown here due to its homology with CYC1 (iso-1 cytochrome *c*) (27). The *underlined sequences* depict exact homology with the consensus sequence. The 3′ mRNA ends for ADH-I, MATa1 (37), and MATα1 (37) were all determined by S1 mapping. The CYC1 mRNA 3′ end was determined by primer extension (J. Boss, R. Zitomer, M. Smith and S. Gillam, personal communication). The DNA sequence for the yeast histone H3 and H4 genes is from the sequence of cloned genes (M. Smith, personal communication). The glyceraldehyde-3-phosphate dehydrogenase (*G3PDH*) sequence is of pGAP491 from Holland and Holland (25).

amino acid 20 in the ADH-II protein has been shown to be a histidine (6), as it is here in ADH-I. These results suggest that the ADH-I/ADH-II difference at amino acid 20 is merely a consequence of strain variability. The amino acids which are responsible for the different enzymatic properties of ADH-I and ADH-II (1, 2, 7) must, then, be looked for in the other 14 amino acid differences noted between the two isozymes (6). The asparagines at amino acids 236 and 244 had only been tentatively identified by Jörnvall (23) with the stipulation that one of them might be an aspartic acid. The DNA sequence demonstrates that amino acid 236 is aspartic acid, while position 244 is filled by asparagine. The glutamine identified as a glutamic acid in the ADH-I protein sequence (23) may have been missed due to an exceptional sensitivity to deamidation *in vivo* or *in vitro*.

Almost all of the amino acid differences observed between the published ADH-I protein sequences (6, 23) and the protein sequence derived here can be accounted for by single base changes. The one exception to this rule is the valine-threonine conflict at amino acid 58, which requires two transitions.

The amino acid next to the acylserine seen in the ADH-I protein predicted by the DNA sequence is a methionine. This suggests that there is no proteolytic maturation of the NH_2 terminus of the protein after translation other than removal of the initiation methionine.

The Function of Sequences 5′ to the ADH-I-coding Region—The sequence TATAAA is observed 128 nucleotides upstream of the translation-initiating triplet in the ADH-I gene. A sequence related to the model hexamer TATAAA is seen upstream of the translation-initiating codon in a number of eukaryotic genes (*cf.* Ref. 18), including those from yeast (24–28). None of the proposed TATAAA-like sequences in the various sequenced yeast genes (24–28) is surrounded by GC-rich regions as they are in multicellular eukaryotes (18). The best evidence to date suggests that this sequence is involved primarily in determining the exact positioning of initiation by RNA polymerase II *in vivo* (29, 30). In most multicellular eukaryotes, initiation is seen 32 nucleotides downstream from the TATAAA. The major 5′ ends of yeast ADH-I and iso-1 cytochrome *c* (29) mRNAs are 60 to 100 nucleotides downstream from TATAAA-like sequences. The possibility exists that these multiple mature ends may denote processing sites rather than true transcription initiation points. Both the multiplicity of observed 5′ ends of mRNAs in yeast (at least two for ADH-I and seven for iso-1-cytochrome *c*) and their greater distance from the TATAAA sequence could conceivably be explained by RNA processing. However, this would require that the concurrence between the capped nucleotide at the 5′ end of the mature mRNA and the transcription-initiating nucleotide seen in other eukaryotes (*cf.* Ref. 31) not hold true for a number of yeast genes. Further experiments will be necessary to determine the basis of the observed differences between the synthesis of mRNA 5′ ends in yeast and in multicellular eukaryotes.

Inserted between the TATAAA sequence in the ADH-I gene and the sequences coding for the mapped 5′ ends is a pyrimidine-rich cluster (60% T and 27% C) between 105 and

50 nucleotides upstream of the translation-initiation site on the mRNA identical strand. Only one A is found in this 56-nucleotide stretch. Montgomery and co-workers (27) proposed that long clusters of pyrimidines in this area may serve to signal a high level of transcription in yeast.

As is seen in other eukaryotic genes (32), the first ATG in the mRNA identical strand upstream of the protein-coding region is well removed (200 nucleotides in this case) from the actual translation-initiating ATG. The area between these two ATGs is enriched for AT base pairs (65% AT). This is a common property of regions preceding the start codons in several yeast genes (24–28). The alternating pyrimidine-purine-pyrimidine-rich stretches between nucleotides −223 and −406 in the promoter proximal region of the ADH-I gene are as extreme as any reported in eukaryotic DNA. Although this region may be too far upstream from the transcription initiation site to be involved in RNA polymerase recognition and initiation, it is not unlikely that this area might be involved in regulation of the transcription event. Alternatively, this highly nonrandom region may be involved in some function only peripherally related to the *ADC*1 locus.

The Function of Sequences 3′ to the ADH-I-coding Region—S1 nuclease mapping of the 3′ ends of ADH-I mRNA established that the vast majority of molecules end approximately 83 nucleotides beyond the translation termination site. In addition, minor 3′ end locations map to positions 24 and 67 nucleotides downstream from the coding region. Roughly paralleling these three transcription termini are three repeats of the sequence AATAAG located, in each case, 28 to 34 nucleotides upstream of one of the 3′ transcript ends. A related sequence, AAUAAA, which occurs in the 3′ trailer sequences of poly(A) mRNA molecules from higher cells (18, 19), has been implicated in specifying the 3′ ends of mRNAs transcribed from genes in animal cells (33). Typically, in these cases, AAUAAA occupies a position within the trailer sequence 21 to 26 nucleotides upstream of the poly(A) addition site (18, 19). In an effort to determine whether the 3′-proximal ADH-I AATAAG sequences represent an analogous element, we have aligned and compared the 3′-proximal DNA sequences for several yeast genes, including as a separate case each of the widely separated ADH-I termini (Fig. 5). Most of the highly conserved trailer sequences fall within a generally AT-rich region which begins 25 to 40 nucleotides before the poly(A) site. The model sequence in this region, $\mathrm{TAAATAA}\begin{matrix}\mathrm{A}\\\mathrm{G}\end{matrix}$, encompasses the AATAAG sequences of ADH-I and the AATAAA sequence of the MATa1 and glyceraldehyde-3-phosphate dehydrogenase genes.

While the comparison made suggests that these common sequences may have a function in transcription termination and/or poly(A) addition, genetic experiments will be required in each case to establish such a functional role. For one of the yeast genes we have compared (MATα1) such evidence exists. A deleted MATα1 gene lacking the consensus sequence in Fig. 5, as well as 650 additional base pairs beyond the 3′ terminus, is unable to provide α1 function *in vivo* even though it retains the entire α1-coding sequence and the 5′ flanking sequence (34). This result suggests that the MATα1 sequences depicted have a functional role in determining the 3′ end of MATα1 mRNA.

Whether the yeast consensus sequence is viewed as a 20 to 30 nucleotides AT-rich block, as AATAAA, or as something intermediate, it is in any event located at a different distance from the 3′ mRNA terminus than is the higher eukaryote AATAAA signal (18, 19). Whereas for animal and animal virus mRNAs the initial A of this sequence is located 21 to 26 nucleotides from poly(A), the corresponding distances for the yeast AATAAA-like sequences are more commonly 28 to 33 nucleotides. An additional difference between yeast and animal 3′ termini is seen immediately proximal to the poly(A) site. The model sequence TTTTCACTGC, which Benoist and co-workers (18) discerned in animal cell and viral mRNAs, is not found at yeast mRNA 3′ termini.

The sequence $\begin{matrix}\mathrm{T}\\\mathrm{C}\end{matrix}$ A is found at or very near the eight mapped yeast poly(A) addition sites. An adenine nucleotide preceded by a G, T, or C has also been seen to be at the poly(A) sites in many animal genes (33). In yeast, as well as multicellular eukaryotes (33), this short sequence is probably one of the cues which signals and positions poly(A) addition.

There is no obvious reason, from the control regions so far postulated, why the three poly(A) addition sites in the ADH-I gene should be of such variable efficiency. Greater than 90% of the ADH-I mRNA molecules' 3′ ends map to the poly(A) addition site which is furthest downstream from the structural gene. Neither is there any apparent advantage obvious to the *ADC*1 locus in having two minor poly adenylation sites followed by a major one. If, as expected, termination of transcription occurs far beyond any of these three sites (35, 36), the predominance of the most 3′ distal site could be explained by the approach of the enzymes involved in mRNA 3′ end maturation from the 3′ direction (*i.e.* moving 3′ to 5′ along the RNA strand). In this case, the most 3′ distal signal would be seen first by the processing machinery. Alternatively, mRNA secondary structure or other cues not defined by our consensus sequence may be necessary accessory factors in determining the relative efficiency of these end maturation events.

Acknowledgments—We wish to thank Drs. C. Furlong, S. Bektesh, and K. Tatchell for assistance with the DNA sequencing.

REFERENCES

1. Ebisuzaki, K., and Guzman-Baron, E. S. (1957) *Arch. Biochem. Biophys.* **69,** 555–564
2. Lutstorf, U., and Megnet, R. (1968) *Arch. Biochem. Biophys.* **126,** 933–944
3. Ciriacy, M. (1975) *Mutat. Res.* **29,** 315–325
4. Ciriacy, M. (1975) *Mol. Gen. Genet.* **138,** 157–164
5. Ciriacy, M. (1976) *Mol. Gen. Genet.* **145,** 327–333
6. Wills, C., and Jörnvall, H. (1979) *Eur. J. Biochem.* **99,** 323–331
7. Wills, C. (1976) *Nature* **261,** 26–29
8. Heick, H. M. C., Willemot, J., and Begin-Heick, N. (1969) *Biochim. Biophys. Acta* **191,** 493–501
9. Fowler, P. W., Ball, A. J. S., and Griffiths, D. E. (1972) *Can. J. Biochem.* **50,** 35–43
10. Williamson, V. M., Bennetzen, J., Young, E. T., Nasmyth, K., and Hall, B. D. (1980) *Nature* **283,** 214–216
11. Elwell, L. P., deGraaf, J., Seiberg, P., and Falkow, S. (1975) *Infect. Immunol.* **12,** 404–410
12. Bennetzen, J. (1980) Ph.D. dissertation, University of Washington
13. Richardson, C. C. (1965) *Proc. Natl. Acad. Sci. U. S. A.* **54,** 158–165
14. Smith, H. O., and Birnstiel, M. L. (1976) *Nucleic Acids Res.* **3,** 2387–2398
15. Maxam, A. M., and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 560–564
16. Berk, A. J., and Sharp, P. A. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75,** 1274–1278
17. Bennetzen, J. L., and Hall, B. D. (1982) *J. Biol. Chem.* **257,** 3026–3031
18. Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980) *Nucleic Acids Res.* **8,** 127–142
19. Proudfoot, N. J., Cheng, C. C., and Brownlee, G. G. (1976) *Progr. Nucleic Acid Res. Mol. Biol.* **19,** 123–134
20. Bogenhagen, D. F., Sakonju, S., and Brown, D. D. (1980) *Cell* **19,** 27–35
21. Koski, R. A., Clarkson, S. G., Kurjan, J., Hall, B. D., and Smith, M. (1980) *Cell* **22,** 415–425

22. Eklund, H., Bränden, C.-I., and Jörnvall, H. (1976) *J. Mol. Biol.* **102,** 61–73
23. Jörnvall, H. (1977) *Eur. J. Biochem.* **72,** 425–442
24. Holland, J. P., and Holland, M. J. (1979) *J. Biol. Chem.* **254,** 9839–9845
25. Holland, J. P., and Holland, M. J. (1980) *J. Biol. Chem.* **255,** 2596–2605
26. Smith, M., Leung, D. W., Gillam, S., Astell, C. R., Montgomery, D. L., and Hall, B. D. (1979) *Cell* **16,** 753–761
27. Montgomery, D. L., Leung, D. W., Smith, M., Shalit, P., Faye, G., and Hall, B. D. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 541–545
28. Gallwitz, D., and Sures, I. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 2546–2550
29. Faye, G., Leung, D. W., Tatchell, K., Hall, B. D., and Smith, M. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 2258–2262
30. Grosschedl, R., and Birnstiel, M. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 1432–1436
31. Ziff, E. B., and Evans, R. M. (1978) *Cell* **15,** 1463–1475
32. Kozak, M. (1978) *Cell* **15,** 1109–1123
33. Fitzgerald, M., and Shenk, T. (1981) *Cell* **24,** 251–260
34. Nasmyth, K. A., Tatchell, K., Hall, B. D., Astell, C., and Smith, M. (1981) *Nature* **289,** 244–250
35. Fraser, N. W., and Hsu, M.-T. (1980) *Virology* **103,** 514–516
36. Nevins, J. R., Blanchard, J.-M., and Darnell, J. E., Jr. (1980) *J. Mol. Biol.* **144,** 377–386
37. Nasmyth, K. A., Tatchell, K., Hall, B. D., Astell, C., and Smith, M. (1981) *Cold Spring Harbor Symp. Quant. Biol.* **45,** 961–981

The primary structure of the Saccharomyces cerevisiae gene for alcohol dehydrogenase.

J L Bennetzen and B D Hall

J. Biol. Chem. 1982, 257:3018-3025.
