The complete nucleotide sequence of the glucose-repressed alcohol dehydrogenase II gene (ADR2) from yeast has been established together with its 5'- and 3'-flanking regions. The limits of the gene have been determined by reverse transcriptase sequencing of the 5'-ends and by S1 nuclease mapping of the 3'-ends of the mature ADR2 mRNA. Comparison of the alcohol dehydrogenase I gene (ADCl) sequence (Bennetzen, J. L., and Hall, B. D. (1982) J. Biol. Chem. 257, 3018-3023) with that of ADR2 indicated four regions of sequence conservation in the 5'-flanking DNAs. One of these conserved regions contains the sequence TCAAG which may function as a yeast cap sequence. The coding sequence of ADR2 is 89% homologous with that of ADCl and exhibits a bias in its codon utilization. Evidence is presented that the intergenic region at the 3'-end of the ADR2 gene is less than 550 base pairs.

The alcohol dehydrogenase isozymes of Saccharomyces cerevisiae are responsible for the interconversion of acetaldehyde and ethanol. When grown on fermentable carbon sources such as glucose, yeast expresses an isozyme which preferentially catalyzes the conversion of the penultimate intermediate of the fermentative pathway, acetaldehyde, to ethanol. Alcohol dehydrogenase I is the classical fermentative enzyme, first isolated and characterized by Racker (1956), and studied by Lutstorf and Megnet (1968). Exhaustion of glucose in the media, or growth on nonfermentable carbon sources such as ethanol or glycerol, results in the expression of the glucose-repressed alcohol dehydrogenase of yeast. A third alcohol dehydrogenase isozyme, whose role in intermediary metabolism is unknown, is found associated with the mitochondria.

Due to their preponderance in the cell, the alcohol dehydrogenase I and II isozymes are readily purifiable (Wills and Jornvall, 1979a), and much is known regarding their structure and mechanisms of action (Branden et al., 1975). Both enzymes are composed of four identical subunits, each having $M_r \approx 37,000$. The subunits of a given isozyme can interact with those of the other isozyme and form active hybrid tetrameric alcohol dehydrogenases; however, the physiological relevance of these hybrid isozymes is not known (Ciriacy, 1975a). Preferred reactions catalyzed by the homomeric isozymes are thought to be brought about by differences in their substrate and cofactor $K_m$ values (Wills, 1976). Protein sequence data obtained from both enzymes have identified several amino acid differences which may be responsible for these altered kinetic parameters (Wills and Jornvall, 1979a).

Although none of the yeast alcohol dehydrogenase isozymes has yet been crystallized, much insight into their predicted tertiary structures has been gained by comparison to that of the horse liver enzyme (Eklof et al., 1976; Jornvall et al., 1978).

The regulation and expression of the yeast alcohol dehydrogenase genes have been the subject of a number of studies. Extensive genetic analysis by Ciriacy (1975a) has resulted in the identification of genes which encode structural information for the alcohol dehydrogenase I isozyme (ADCl) and the alcohol dehydrogenase II isozyme (ADR2). Moreover, genes which specifically regulate the expression of the ADR2 locus have been defined. These include the trans-acting regulatory elements, ADR1 and ADR4 (Ciriacy, 1975b; 1979; Denis et al., 1983), and a tightly linked cis-acting regulatory locus, ADR3 (Ciriacy, 1976). A number of mutants at the cis-acting regulatory locus (genetic designation, ADR3') have been isolated which express the alcohol dehydrogenase II isozyme constitutively when cells are grown with glucose as a carbon source (Ciriacy, 1976; 1979).

The development of transformation in yeast (Beggs, 1978; Hinoen et al., 1978) has allowed the cloning by complementation of function of a number of yeast genes (Olsen, 1981). These techniques have been used to clone the ADCl gene (Williamson et al., 1980), the ADR2 gene (Williamson et al., 1981), and the ADR1 gene. Using the cloned ADR2 gene as a probe, several mutants at the ADR3 locus have been characterized in which the alcohol dehydrogenase II isozyme is expressed on glucose-containing media (Williamson et al., 1981; 1983; Ciriacy and Williamson, 1981). These ADR3' mutants may be divided into two classes; those brought about by the insertion of a yeast transposable element in the 5'-flanking sequences of the ADR2 gene, and those which result from a more complex DNA rearrangement in these sequences.

* 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.

† Supported by Fellowship DRC-455 of the Damon Runyon-Walter Winchell Cancer Fund. Present address, Department of Molecular Genetics, University of Texas Health Science Center, 5223 Harry Hines Blvd., Dallas, TX 75235.

‡ Supported by a grant from the Medical Research Council of Canada, and Career Investigator of the Medical Research Council of Canada.

§ Supported by a National Institutes of Health postdoctoral fellowship.

¶ Supported by ARCO Plant Cell Research Institute, 4560 Trinity Court, Dublin, CA 94566.

‖ Supported by National Institutes of Health Grant GM26079.
In this paper, the DNA sequence of the wild type gene (ADR2) encoding the structural information for the alcohol dehydrogenase II isoform is presented. Several compelling reasons led to the determination of the primary structure of this gene. First, the complete DNA sequence of the gene encoding the alcohol dehydrogenase I isoform (ADR1) has been determined (Bennetzen and Hall, 1982a), allowing a comparison of two highly conserved genes which differ in the regulation of their expression. Second, insight into the preferred metabolic directions of the two isozymes could be gained by comparing the coding sequences of these two genes. Third, knowledge of the wild type ADR2 sequence is necessary to interpret the sequence changes occurring in the various constitutive mutants. And fourth, it was desirable to elucidate the arrangement of genes flanking ADR2 and to determine if these genes are also regulated by glucose.

**EXPERIMENTAL PROCEDURES**

**Materials—**Deoxyribonucleoside $\alpha$-$\beta$-triphosphates were obtained from Amersharm Corporation at a specific activity of 2000–3000 Ci/mmol. Restriction endonucleases were purchased from New England Biolabs, Boehringer Mannheim, or Bethesda Research Laboratories. Reverse transcriptase was from Life Sciences. DeoxyNTPs and 2',3'-dideoxyNTPs were purchased from P-L-Biochemicals.

**Isolation of Plasmid DNA—**Plasmid DNAs containing the cloned ADR2 gene (Williamson et al., 1981) were purified by a modification of the method of Clewell and Helinski (1970). Restriction endonuclease-derived plasmid DNA fragments were separated on preparative acrylamide gels and subsequently isolated by electroelution as described by McDonell et al. (1977).

**Isolation of Yeast Poly(A)-containing mRNA—**RNA was extracted from yeast as described by Schultz (1978). Poly(A)-containing RNA was purified on oligo(dT)-cellulose as described by Aviv and Ledler (1975). Poly(A)-containing RNA employed as a template in sequencing reactions was isolated from yeast strain R234 (adel ADR1-5 ADR2) (Ciriacy, 1979) 2 h after transfer to a media containing ethanol as the carbon source.

**DNA Sequence Analysis of ADR2—**DNA fragments were 3'-end labeled with deoxyribonucleoside $\alpha$-$\beta$-triphosphates and the large fragment of E. coli polymerase I exactly as described by Smith et al., 1981; a kind gift of Dr. David Leung, Genentech Corporation. For both sequencing methods, gel electrophoresis was performed as described by Sanger and Coulson (1978). The oligodeoxyribonucleotide d(pTCA) (Gillam and Smith, 1980), was used as a primer on a denatured plasmid DNA template as described previously (Smith et al., 1979, Williamson et al., 1983).

**Sequence Analysis of ADR2 mRNA—**The sequences of the 5'-ends of the ADR2 mRNA were determined by the methods of Levy et al. (1979) and Sures et al. (1980). A 77-bp restriction endonuclease fragment corresponding to nucleotides 1245 to 1319 (Fig. 2) was isolated and strand separated. 0.75 pmol of the mRNA-complement strand was mixed with 60 pmol of poly(A)-containing mRNA isolated from cells derepressed for ADR2. The mixture, 10 pmol in Buffer A (50 mM Tris-HCl, pH 7.3, 50 mM KCl, 5 mM MgCl2, and 20 mM dithiothreitol), was sealed in a capillary tube and incubated at 65 °C for 45 min to allow formation of the primer-template complex. Following annealing, 20 pmol of [α-32P]dATP was added, together with 0.5 pmol of 0.1 nm dATP. The mixture was divided equally among four tubes containing 4 pmol of a deoxyNTP/dideoxyNTP mixture ( 5°C = 0.1 M dCTP, 0.1 M dTTP, 0.005 M dGTP, and 0.005 M ddGTP; "T" = 0.1 M dGTP, 0.1 M dTTP, 0.005 M dCTP, and 0.005 M ddCTP; "A" = 0.1 M dGTP, 0.005 M dCTP, 0.1 M dTTP, and 0.005 M ddATP; "G" = 0.1 M dCTP, 0.1 M dTTP, 0.005 M dGTP, and 0.005 M ddGTP; all in Buffer A). Two units of avian myeloidosis virus reverse transcriptase were added, and the reactions were incubated at 37 °C for 10 min. At this time, 0.25 pmol of a solution containing 2.5 M each of the four deoxyribonucleotide triphosphates was added and the reaction mixture was incubated a further 10 min at 37 °C. Reactions were terminated by the addition of 4 μl of 90% formamide, 25 mM EDTA, 0.02% xylene cyanol, and 0.02% bromophenol blue, and then quick cooled in an ice-bath. Electrophoresis was carried out on thin gels described above, in a buffer containing 90 mM Tris-borate, pH 8.3, 2.5 mM EDTA.

**5' Nucleotide Mapping of ADR2 mRNA—**The 3'-ends of points mapping ADR2 mRNA were determined by S1 nuclease protection experiments. Experimental conditions have been described elsewhere (Williamson et al., 1983). The FnuEI-HindII DNA fragment extending from nucleotide 2252 to 2526 (Fig. 2) was isolated and 3'-end labeled with [α-32P]dATP using the large fragment of E. coli DNA polymerase I. This fragment was strand separated and used as the hybridization probe.

**Computer Analysis of DNA Sequences—**Computer-assisted sequence analysis was carried out on the University of British Columbia computer, using the programs of Delaney (1982).

**RESULTS**

**DNA Fragment Isolation and Restriction Endonuclease Mapping of ADR2—**Three plasmids were used as a source of DNA in determining the sequence of the ADR2 gene. A majority of the coding and 3'-flanking sequences were derived from the DNA of the ADR2 plasmid (Williamson et al., 1981). Yeast DNA cloned in this plasmid was isolated from a mutant constitutive for ADR2 expression, brought about by the insertional of a yeast transposable element in the 5'-flanking sequences of the gene. Partial sequence data were obtained.

**The abbreviations used are: bp, base pair; kb, kilobase.**
Sequence of Yeast Alcohol Dehydrogenase II Gene

A majority of the 5'-flanking sequence was derived from DNA isolated from a plasmid (A2M) carrying the wild type ADR2 gene. In regions of the DNA sequence established from more than one plasmid (over 30% of the total), no differences were detected at the nucleotide level excepting those in the 5'-noncoding region associated with the constitutive phenotypes.

A restriction endonuclease map of the ADR2 gene region derived from these three plasmids is shown in Fig. 1. The BalI, AccI, HindIII, and HincII sites depicted in the coding region are conserved between both ADR2 and ADC1 (Bennetzen and Hall, 1982a). However, none of the restriction endonuclease cleavage sites occurring in the 5'- and 3'-flanking regions of Fig. 1 is conserved between these two genes. More detailed restriction maps were constructed for ADR2 by the method of Smith and Birnstiel (1976), and were used in determining the DNA sequence of the gene.

DNA Sequencing Strategies—Three methods of DNA sequence analysis were employed in determining the primary structure of the ADR2 gene. Approximately one-half of the sequence was established using the enzymatic terminator method (Sanger et al., 1977) and a universal primer on single-stranded DNA clones obtained in the bacteriophage M13mp7 vector (Messing et al., 1981). Individual clones isolated and sequenced in this manner are indicated above the restriction endonuclease map in Fig. 1. One segment of the 5'-flanking sequence of the gene was determined by the terminator method directly on the denatured plasmid DNA A2M using the synthetic oligodeoxyribonucleotide, d(pT&CA) (Gillam and Smith, 1980), as a primer. This DNA sequencing strategy was first used in the determination of the yeast iso-1-cytochrome c gene sequence (Smith et al., 1979), and has been used extensively in the analysis of mutant DNAs derived from the SUP-40 locus (Kurjan et al., 1980; Koski et al., 1980). ADR2 sequence obtained by this method allowed the ordering of the adjacent M13 clones S5, S13, and S23 shown in Fig. 1. The remainder of the DNA sequence was established using the base-specific cleavage reactions of Maxam and Gilbert (1980). The individual experiments performed and the extent of the DNA sequence established in each are indicated below the restriction endonuclease map in Fig. 1.

The complete sequence of the ADR2 gene and its 5'- and 3'-flanking regions is shown in Fig. 2. Overlapping sequences were obtained throughout, and over 80% of the sequence was determined for both strands of the DNA, often by two different sequence determination methods. Errors in recording the data were minimized by two independent readings of all sequencing gels and rechecking these readings directly against a computer record of the sequence. The determination of the DNA sequence in the coding region of ADR2 was facilitated by comparison with the ADHII protein sequence (Wills and Jornvall, 1979a) and with the DNA sequence of the coding region of the ADHI gene, ADC1 (Bennetzen and Hall, 1982a). Any differences between the sequence presented in Fig. 2 and these other sequences were carefully scrutinized.

To further check the data indicated in Fig. 2, the appropriate DNA was digested separately with 11 restriction endonucleases, radioactively end labeled, and analyzed by polyacryl-

![Fig. 2. The DNA sequence of the ADR2 gene and its 5'- and 3'-flanking regions. The sequence of part of the DNA is indicated together with a detailed restriction endonuclease map for this DNA. The sequence begins at a FnuEI recognition sequence and extends for 3627 bp. Numbers below the sequence indicate the distance in nucleotide pairs from the FnuEI cleavage site. The coding sequence of the ADR2 gene begins at position 1202 and extends to position 2248. The resulting protein sequence for the alcohol dehydrogenase II isozyme is indicated above the DNA sequence. A second long open translation reading frame beginning at position 3006 is also indicated above the DNA sequence.]
amide gel electrophoresis (data not shown). The results obtained from these experiments agreed completely with the indicated sequence.

Sequence Analysis of ADR2 mRNA—The 5'-ends of the mature ADR2 mRNA were determined by reverse transcriptase sequencing. Poly(A)-containing mRNA was isolated from yeast cells grown on ethanol as a carbon source and used as a template in a dideoxyribonucleotide-sequencing reaction (Sanger et al., 1977). A 77-base pair HphI-RsaI restriction endonuclease fragment corresponding to nucleotides 1245 to 1319 in Fig. 2 was isolated and strand separated. The mRNA-complementary strand was annealed to poly(A)-mRNA and extended with the enzyme reverse transcriptase in the sequence reaction (Levy et al., 1979; Sures et al., 1980). The resulting fragments were electrophoresed on a denaturing polyacrylamide gel and, following autoradiography, the pattern shown in Fig. 3 was obtained. Four distinct DNA transcript stops corresponding to ADR2 mRNA 5'-ends were resolved by the gel. With the exception of a single G band, sequences beyond the transcription initiation point 56 nucleotides 5' to the initiation codon could not be read and are indicated by N. However, the use of calipers to accurately space the more distal bands, together with the location of an above G band, allowed an accurate assignment of the transcript initiation points to those shown in Fig. 3. Based on their intensity on the autoradiograph, two of the transcripts (closed stars, Fig. 3) account for a majority of the ADR2 mRNA molecules. All four of the transcripts initiated in a region of conserved DNA sequence flanking the 5'-ends of the ADR2 and ADC1 genes (see "Discussion").

S1 Nuclease Mapping of the 3'-Ends of ADR2 mRNA—The 3'-ends of the mature ADR2 mRNA were determined by the S1 nuclease-mapping procedure of Berk and Sharp (1978). A FnuEI-HphI restriction endonuclease fragment corresponding to nucleotides 2252 to 2526 in Fig. 2 was isolated, 3'-end labeled, and strand separated. The mRNA-complementary strand was annealed to total RNA isolated from cells repressed or derepressed for ADR2, and the resulting DNA-mRNA hybrid was digested with S1 nuclease. When electrophoresed on a denaturing polyacrylamide gel adjacent to a DNA sequence ladder generated by the G-specific cleavage reaction of Maxam and Gilbert (1980) obtained. Four major 3'-ends for the mature ADR2 mRNA were resolved by the gel, corresponding to nucleotides 2367 through 2370 of Fig. 2. These ends occur 21 to 24 nucleotides 3' to the unique sequence AAATAA (position 2346; Fig. 2), a sequence characteristic of 3'-untranslated regions of many eukaryotic mRNAs (Proudfoot and Brownlee, 1976). This sequence, or a variant thereof, occurs three times in the 3'-flanking region of the ADC1 gene and three mature ends for the ADC1 mRNA have been identified (Bennetzen and Hall, 1982a). A computer search of the 3'-flanking regions of ADC1 and ADR2 did not reveal any other extensive nucleotide sequence homologies.

Codon Usage in the ADR2 Gene—Table I summarizes the codon usage frequencies of the ADR2 gene. As with other highly expressed yeast genes (Holland and Holland, 1979; 1980; Holland et al., 1981; Wallis et al., 1980; Bennetzen and Hall, 1982a), there is an obvious codon utilization bias within a given family. Codons ending in a pyrimidine are used preferentially, and the arginine codons CGN do not occur at all in the DNA. Overall, however, the bias observed in the ADR2 gene (Bennetzen and Hall, 1982b) is not a strong in the ADR2 gene. Thus, only glutamine is encoded by a single codon and threonine by only two out of its possible four codons.

The Location of Putative Genes in the ADR2 Flanking Sequences—Previous studies have indicated that approximately 40% of the yeast genome is transcribed (Hereford and Rosbash, 1977; Kaback et al., 1979), and that the distances between transcripts in the histone H2A, H2B region (Hereford et al., 1979), the HIS3 gene region (Struhl and Davis, 1981a), and the GAL7, GAL10, GAL1 regions (St. John and Davis, 1981) vary. A search for open translation reading frames in the 5'-flanking sequences of the ADR2 gene (Fig. 2) did not reveal any putative genes. However, a long open reading frame of at least 207 amino acids, extending from nucleotide 3006 (Fig. 2) to the end of the determined sequence, was detected in the 3'-flanking DNA of ADR2. This reading frame begins with a threonine residue and does not encode any methionine residues. To determine if this region of the DNA was transcribed, whole cell RNA was isolated from yeast grown on glucose or ethanol and analyzed by Northern blotting. RNA samples were size fractionated on agarose gels, transferred to nitrocellulose paper (Thomas, 1980), and hybridized with a 32P-labeled DNA fragment which included the above open

![Fig. 3. The sequence of the 5'-ends of the ADR2 mRNAs. An appropriate DNA fragment was isolated and used as a primer on a population of mRNA molecules containing the ADR2 species as described in the text. Depicted is an autoradiogram obtained from a sequencing gel. The nucleotide sequence that can be read from the film is indicated on the right. Unresolved nucleotides are indicated by N. The four 5'-ends of the ADR2 mRNA are aligned with the DNA sequence in the lower portion of the figure. Closed stars indicate more abundant transcripts; open stars indicate less prevalent transcripts. Numbers indicate the distance from the A of the ATG initiation codon.](http://www.jbc.org/Downloaded from http://www.jbc.org/)
Fig. 4. S1 nuclease mapping of the 3'-ends of the ADR2 mRNA. Total yeast RNA was hybridized to a radiolabeled single-stranded DNA fragment extending from nucleotide 22252 to 2516 as designated in Fig. 2, treated with S1 nuclease, denatured in 90% formamide, and analyzed on an 8% polyacrylamide gel containing 7 M urea. Lane 1 contains the untreated, labeled fragment. Lane 2 is a no RNA control. For lane 3, 9 μg of total RNA were isolated from glucose-grown cells which express alcohol dehydrogenase I constitutively (mutant ADH2+2'). Lane 4 used 9 pg of RNA isolated from glucose-grown cells with wild type alcohol dehydrogenase II regulation. Lane 5 used 3 pg, and lane 6 9 pg of RNA isolated from wild type cells grown on a media containing ethanol as the carbon source. Lane G contains fragments obtained after guanine-specific cleavage of the same DNA fragment as described above. Numbers on the right indicate positions as indicated in Fig. 2.

reading frame, and extended from the Sphi restriction endonuclease cleavage site at nucleotide 2557 (Fig. 2) to an EcoRI site approximately 2 kb downstream. Two RNA transcripts of about 1.0 and 1.4 kb were detected in cells grown either on glucose or on ethanol (data not shown). As over one-half of the 2.0-kb DNA probe has been sequenced (Fig. 2), and the combined length of the complementary RNA transcripts is

Table 1

| Amino acid | Codon   | Frequency |
|------------|---------|-----------|
| Phc        | UUU     | 2         |
| Phc        | UUC     | 6         |
| Leu        | UUA     | 7         |
| Leu        | UUG     | 15        |
| Leu        | CUU     | 0         |
| Leu        | CUC     | 1         |
| Leu        | CUA     | 2         |
| Leu        | CUG     | 0         |
| Ile        | AUU     | 9         |
| Ile        | AUC     | 12        |
| Ile        | AUA     | 1         |
| Met        | AUG     | 5         |
| Val        | GGU     | 16        |
| Val        | GUC     | 12        |
| Val        | GUA     | 3         |
| Val        | GUG     | 1         |
| Ser        | UCU     | 13        |
| Ser        | UCC     | 7         |
| Ser        | UCA     | 1         |
| Ser        | UCG     | 1         |
| Pro        | CCU     | 2         |
| Pro        | CCC     | 1         |
| Pro        | CCA     | 12        |
| Pro        | CCG     | 0         |
| Thr        | ACU     | 5         |
| Thr        | ACC     | 9         |
| Thr        | ACA     | 0         |
| Thr        | ACG     | 20        |
| Ala        | GCU     | 18        |
| Ala        | GCC     | 12        |
| Ala        | GCA     | 3         |
| Ala        | GCG     | 4         |
| Tyr        | UAU     | 1         |
| Tyr        | UAC     | 12        |
| Ochr       | UAA     | 1         |
| Amber      | UAG     | 0         |
| His        | CAU     | 2         |
| His        | CAC     | 9         |
| Gin        | CAA     | 6         |
| Gin        | CAG     | 0         |
| Asn        | AAU     | 1         |
| Asn        | AAC     | 12        |
| Lys        | AAA     | 4         |
| Lys        | AAG     | 20        |
| Asp        | GAU     | 6         |
| Asp        | GAC     | 9         |
| Glu        | GAA     | 18        |
| Glu        | GAG     | 3         |
| Cys        | UGU     | 6         |
| Cys        | UGC     | 2         |
| Cys        | UGA     | 0         |
| Cys        | UGG     | 5         |
| Arg        | CGU     | 0         |
| Arg        | CGC     | 0         |
| Arg        | CGA     | 0         |
| Arg        | CCG     | 0         |
| Ser        | AGU     | 1         |
| Ser        | AGC     | 1         |
| Arg        | AGA     | 7         |
| Arg        | AGG     | 1         |
| Gly        | GGU     | 32        |
| Gly        | GGC     | 8         |
| Gly        | GGA     | 1         |
| Gly        | GGG     | 1         |
about 2.4 kb, it is probable that the long open translation reading frame shown in Fig. 2 does represent a gene. If this is the case, an NH₂-terminal methionine could be provided by the initiation of transcription upstream from the threonine codon at nucleotide 3006, followed by an appropriate splicing event to produce a mature mRNA containing an initiation codon. In agreement with this hypothesis, the nucleotide sequence adjacent to the threonine codon is TAG, which is homologous to the 3'-splice junctions of many eukaryotic mosaic mRNAs (Lerner et al., 1980). A computer-assisted search of sequences 5' to the putative acceptor splice site revealed a number of possible donor sites homologous to the consensus sequence described by Lerner et al. (1980). However, it has not yet proven possible to unambiguously define the correct site based on a search of this nature.

DISCUSSION

Published work on the DNA sequences in the 5'-flanking regions of non tandemly repeated yeast genes, such as those encoding glyceraldehyde-3-phosphate dehydrogenase (Hol-land and Holland, 1979; 1980), enolase (Holland et al., 1981), histone H2B and H2A (Wallis et al., 1980; Choe et al., 1981), and the iso-1- and iso-2-cytochrome c proteins (Smith et al., 1979; Montgomery et al., 1980), led to an a priori expectation of sequence homology in the 5'-flanking regions of the ADR2 and ADCl genes. These regions are aligned for maximum homology in Fig. 5. Four distinct sequences are conserved between these two highly expressed yeast genes. The first homologous sequence occurs approximately 160 bp distal to the coding sequence of ADR2. It includes the sequence TATAAA which was first identified by Goldberg (1979) and has subsequently been found in the 5'-flanking DNAs of most eukaryotic genes transcribed by RNA polymerase II (Breathnach and Chambon, 1981). A second region of homology is found in a pyrimidine-rich sequence in the nontranscribed strand. Both genes contain an inverted repeat within this sequence (Fig. 5). The third and longest region of homology (33 out of 39 bp) exists in sequences within which transcript initiation occurs. The sequence TPYdAACTA, separates the third and fourth regions of homology, occurring three times in direct repeat in the ADR2 gene and once in the ADCl gene (dashed arrows, Fig. 5). The fourth homologous region is located immediately adjacent to the coding DNA of the two genes and contains a variant of the sequence CACACA (Stiles et al., 1981). Conserved sequences similar to these four are also found in the 5'-flanking DNAs of other yeast genes encoding abundant mRNAs (Dobson et al., 1982), and direct repeats of 15 bp occur adjacent to the coding region of the HIS3 gene (Struhl and Davis, 1981a).

Transcription initiates at four distinct positions in the 5'-flanking sequences of ADR2 (Fig. 3; Williamson et al., 1983). These sites are located 67, 63, 58, and 54 nucleotides in front of the coding sequence of the gene. As with other abundant yeast mRNAs (Sripati et al., 1976; De Kloet and Anderean, 1976), all four transcripts have an A at their 5'-terminus. Based on their autoradiographic intensity (Fig. 3), transcripts initiating 58 and 54 nucleotides 5' to the ATG appear to be the most prevalent. Several conclusions may be drawn from these data. First, all four transcripts initiate in a region of the DNA which is highly conserved between the ADR2 and ADCl genes (Region 3, Fig. 5). Similarly, transcripts initiating in the 5'-flanking sequences of the ADCl gene have been mapped to this region (Ammerer et al., 1981; Bennetzen and Hall, 1982a). The sequence TCAAG is found within this conserved region and, as has been noted by Dobson et al. (1982), this sequence is common to a number of yeast promoter regions. It may be postulated, based on the above transcription mapping data, that TCAAG represents a yeast cap sequence. Second, the four transcription initiation sites flanking the ADR2 gene occur in a region located 94 to 107 nucleotides 3' to a conserved TATAAA sequence (Region 1, Fig. 5). A similar situation characterizes the ADCl gene, in which transcripts initiate 91 and 101 nucleotides 3' to this sequence (Ammerer et al., 1981; Bennetzen and Hall, 1982a). These results indicate a fundamental difference between ADR2 and ADCl and other higher eukaryotic genes in which transcription generally initiates 25 to 30 nucleotides 3' to a TATAAA sequence (reviewed, Breathnach and Chambon, 1981). A varying distance separating a TATAAA sequence and the mRNA cap sequence may be a characteristic of yeast genes. Transcription initiates approximately 40 nucleotides 3' to a TATAAA sequence in the HIS3 gene (Struhl and Davis, 1981a), approximately 54 nucleotides 3' to a TATA sequence in the actin gene (Gallwitz et al., 1981), and about 70 nucleotides 3' to this sequence in the phosphoglycerate kinase gene (Dobson et al., 1982). Similarly, the CYCl mRNA initiates approximately 60 nucleotides 3' to the sequence TATAAA (Boss et al., 1981); however, there may be multiple TATA sequences and 5'-mRNA ends for this gene (Faye et al., 1981). Multiple 5'-ends for the mRNAs encoded by the mating type locus (MATa and MATa') also initiate approximately 36 to 59 nucleotides 3' to a TATA sequence (Nasmuth et al., 1981; Astell et al., 1981). These observations suggest that the yeast RNA polymerase II enzyme does not share the same rigid promoter requirements of the higher eukaryotic enzyme and can tolerate variable and in general longer distances between the TATA sequence and the mRNA cap site. The techniques of site-specific mutagenesis (Smith and Gillam, 1981) will be most useful in clarifying the role of the conserved TATA sequence in yeast gene expression.

S1 nuclease mapping indicated that there may be as many as four closely spaced mature 3'-ends for the ADR2 mRNA (Fig. 4). It is possible, however, that some of these multiple ends may be an artifact brought about by the increased susceptibility of A-U-rich hybrids to S1 nuclease digestion in an experiment of this design (Boss et al., 1981). Consistent with this observation, three of the four ends map in a tract of 5 U residues located at position 2366 in Fig. 2. With this caveat in mind, these ends occur approximately 20 nucleotides 3' to the consensus polyadenylation signal AAGAAA (Proudfoot and Brownlee, 1976), and result in an ADR2 mRNA containing about 120 nucleotides of 3'-untranslated sequence.

Fig. 5. A comparison of the 5'-flanking sequences of the ADR2 and ADCl genes. The four major regions of sequence homology are numbered above the DNA sequence. Solid arrows indicate direct repeats. Dashed arrows indicate direct repeats. The overlined sequence TCAAG is discussed in the text. Asterisks indicate transcription start points.
The 20-nucleotide distance separating the AAUAAA sequence from the 3'-end of the mature ADR2 mRNA is similar to that found for other yeast genes and for higher eukaryotic genes (Benoist et al., 1980). Three mature 3'-ends were detected for the ADCl mRNA, which mapped 28 to 34 nucleotides 3' to three different AUAAG sequences (Bennetzen and Hall, 1982a). Excepting homologies brought about by these sequences, no other extensively conserved regions were found in the 3'-flanking nucleotides of ADR2 and ADCl. A similar situation exists in the iso-1- and iso-2-cytochrome c genes (Montgomery et al., 1980). This finding is in contrast to other nontandemly repeated or highly homologous gene pairs in yeast, such as the glyceraldehyde-3-phosphate dehydrogenase genes (Holland and Holland, 1980) and enolase genes (Holland et al., 1981) in which extensive 3' homologies occur. The ADR2 and ADCl genes may have arisen primordially at an earlier time than did the glyceraldehyde-3-phosphate dehydrogenase and enolase genes, and as such, have undergone more divergent evolution in their 3'-flanking sequences.

Recent evidence indicates that the signal for transcription termination and polyadenylation in some yeast genes may be tripartite in nature (Zaret and Sherman, 1982). The sequence TAG--TAGT--TGT or TAG--TATG--TGT, with variable distances between these groups, were found in the 3'-flanking sequences of a majority of yeast genes examined. Consistent with these findings are the sequences TAG-N1--TATG-N2--TGT and TAG-N1--TAGC-N2--TGT which precede the mature 3'-ends of the ADR2 mRNAs. However, a variant of this sequence does not appear to be present in the 3'-flanking DNA of the ADCl gene.

Approximately 760 bp separate the coding sequence of ADR2 and a long open translation reading frame in the 3'-flanking DNA (Fig. 2). This reading frame begins with a threonine residue and does not contain any methionine residues for at least 207 amino acids. By Northern blot analysis, this region of the DNA appears to be transcribed. One possible explanation for the lack of an initiation codon is that this putative gene represents a spliced gene. In support of this idea is the occurrence of a short sequence adjacent to the threonine encoding DNA which is homologous to the 3'-exon-intron junctions of many eukaryotic genes (Lerner et al., 1980). Although rare in yeast (Kaback et al., 1979), genes transcribed by RNA polymerase II and containing introns have been isolated (Gallwitz and Sures, 1980; Ng and Abelson, 1980; Rosbash et al., 1981); thus, precedence exists for this type of gene structure. The NH2-terminal coding portion of this putative gene has not yet been identified. Assuming that it is not an overlapping gene, the 3'-end must be located distal to the 3'-end of the ADR2 mRNA. As such, the intergenic distance separating ADR2 from its 3'-nearest neighbor could be up to 550 bp. This distance falls within a range (0.3 to 2.5 kb) estimated for other yeast intergenic regions at the histone H2A, H2B loci (Hereford et al., 1979), the actin gene (Gallwitz et al., 1981), the HIS3 locus (Struhl and Davis, 1981b), and the GAL7, GAL10, GAL1 genes (St. John and Davis, 1981). Finally, the fact that this segment of DNA is transcribed whether cells are grown on glucose (ADR2-repressed) or on ethanol (ADR2-derepressed) places an upper limit in the 3'-direction on the distance over which glucose repression can act in the ADR2 region of the chromosome. This limit corresponds to the above estimated intergenic distance of less than 550 bp.

**Fig. 6.** Alignment of the coding sequences of the two alcohol dehydrogenase genes. The ADR2 sequence is indicated in its entirety. The 117 nucleotides which are different in the ADCl coding region are indicated below the ADR2 sequence. Numbering begins with the A of the initiation codon. The ochre termination codon is indicated by asterisks.
Both protein sequence (Wills and Jornvall, 1979a) and DNA hybridization data (Williamson et al., 1981) indicated extensive sequence homology in the coding regions of the ADR2 and ADCl genes. A comparison of the coding regions of the alcohol dehydrogenase genes is shown in Fig. 6. The primary structure of the two genes is 89% homologous within these DNA sequences. Nucleotide changes at 117 positions result from 81 transition and 36 transversion substitutions. The majority, 80%, of the changes occur randomly spaced in the COOH-terminal two-thirds of the coding sequence. In contrast to this distribution, sequences between positions 193 and 431 (Fig. 6) in the NH2-terminal one-third of the DNAs differ by only two nucleotides. This high degree of sequence conservation, to the level of the third nucleotide in the codons involved, may be explained by a recent nonreciprocal interchromosomal gene conversion event between the ADR2 and ADCl loci. Conversion events of this nature between mutant his3 genes at different loci in yeast chromosomal DNA have been shown to occur at a frequency of $10^{-5}$ (Scherer and Davis, 1980). Similar, but less frequent events at the nonallelic CYCI and CYC7 genes have been defined at the DNA and protein sequence level (Ernst et al., 1981). Based on different substrate and cofactor requirements, the two alcohol dehydrogenase isozymes are thought to perform different metabolic functions (Wills, 1976); thus, conversion events between their genes would be predicted to occur at very low frequencies or to involve regions of the DNAs which did not greatly affect these kinetic parameters. By analogy, sequences lying outside the extended region of exact homology in Fig. 6 may be implicated in the different $K_m$ values which distinguish these two isozymes (Wills, 1976).

Both alcohol dehydrogenases I and II have been the subject of intensive study regarding their protein chemistry, kinetics, and mechanisms of action (Branden et al., 1975). As described above, these enzymes have different substrate and cofactor $K_m$ values (Wills, 1976), and mutant proteins having altered kinetic parameters have been isolated (Wills and Phelps, 1975) and sequenced (Wills and Jornvall, 1979b). In addition, protein sequence data for these isozymes from a characterized yeast strain (Wills and Jornvall, 1979a) and an uncharacterized commercially available source (Jornvall, 1977) have been obtained. Table II summarizes the amino acid changes and their locations in the two isozymes for which DNA sequence data have been collected. Between these sequences, 117 nucleotide changes give rise to 23 amino acid differences, 10 of which represent neutral changes. The remaining 13 amino acid differences alter the charge environment as a result of a neutral $\Rightarrow$ charged or positive $\Rightarrow$ negative charge changes in the side chains of the amino acids. The distribution of changes is biased, being 1.5-fold more prevalent in the cofactor (NAD)-binding domain (residues 154 to 290) than in the catalytic domain (residues 1 to 153 and 291 to 347) (Jornvall et al., 1978). None of these amino acid changes alter highly conserved internal residues or requisite glycine residues (Jornvall et al., 1978). One interesting amino acid change which would have consequences for the tertiary structure of the enzyme occurs at position 204 in the ADR2-derived isozyme. The DNA sequence of the ADR2 gene in this region is GGT-GGT-CCA-GGA. Translation of this sequence gives rise to a change of residue 204 from a glutamic acid in all other sequenced yeast alcohol dehydrogenase isozymes to a proline in the ADR2-derived enzyme, and occurs in the peptide Gly-Gly-Pro-Gly. Protein sequences of this nature allow the peptide backbone to fold back on itself (Dickerson and Geis, 1969). Such a folding back would have a definite but unknown effect on the tertiary conformation of the alcohol dehydrogenase II isozyme in this region. Finally, a subset of changes occurring at positions 15, 168, 211, 229, 265, 270, and 283 represents alterations between the two isozymes which have been detected by both protein- and DNA-sequencing techniques between four different yeast strains. These seven nonidentities would be predicted to give rise to the altered substrate and cofactor $K_m$ values and, hence, the preferred metabolic reactions catalyzed by the two abundant alcohol dehydrogenase isozymes (Wills, 1976).

Acknowledgments—We thank J. Messing for answering many questions regarding the bacteriophage M13 cloning and sequencing system, C. Denis and M. Ciriacy for a plasmid containing the wild type ADR2 gene, C. R. Astell and S. Gillam for helpful discussions, A. D. Delaney and S. Hayashi for help with the computer-assisted analysis of data, and J. H. Harrison for discussions regarding the mechanisms of action of dehydrogenase enzymes.

REFERENCES

Ammerer, G., Hitzeman, R., Hagie, F., Barta, A., and Hall, B. D. (1981) in Recombinant DNA, Proceedings of the Third Cleveland Symposium on Macromolecules, Cleveland, Ohio (Walton, A. G., ed) pp. 185-197, Elsevier, Amsterdam.

Astell, C. R., Ablstrom-Jonasson, L., Smith, M., Tatchell, K., Nasmyth, K. A., and Hall, B. D. (1981) Cell 27, 15-23.

Aviv, H., and Leder, P. (1972) Proc. Natl. Acad. Sci. U. S. A. 69, 1408-1412.

Beggs, J. D. (1978) Nature (Lond.) 275, 104-109.

Bennetzen, J. L., and Hall, B. D. (1982a) J. Biol. Chem. 257, 3018-3025.

Bennetzen, J. L., and Hall, B. D. (1982b) J. Biol. Chem. 257, 3026-3031.

Bennoist, C., O'Hare, K., Breathnach, R., and Chambron, P. (1980) Nucleic Acids Res. 8, 127-142.

Berk, A. J., and Sharp, P. A. (1978) Proc. Natl. Acad. Sci. U. S. A. 75, 1274-1278.

Boss, J. M., Gillam, S., Zitomer, R. S., and Smith, M. (1981) J. Biol. Chem. 256, 12968-12961.

Branden, C. I., Jornvall, H., Ekland, H., and Furugren, B. (1975) in The Enzymes (Boyer, P. D., ed) Vol. 11, pp. 103-190, Academic Press, New York.

Breathnach, R., and Chambron, P. (1981) Annu. Rev. Biochem. 50, 349-384.

Choe, J., Kolodrubetz, D., and Grunstein, M. (1982) Proc. Natl. Acad.
Sequence of Yeast Alcohol Dehydrogenase II Gene

Jornvall, H., Koski, R. A., Clarkson, S. G., Kurjan, J., Hall, B. D., and Smith, M. (1980) J. Mol. Biol. 10, 733-737

I. Levy, S., Sures, I., and Kedes, L. H. (1980) Nature (Lond.) 283, 733-737

A. Nasmyth, K. A., Tatchell, K., Hall, B. D., Astell, C., and Smith, M. (1980) Nature (Lond.) 283, 733-737

R. W. Smith, M., and Gillam, S. (1980) J. Mol. Biol. 22, 733-737

J. W. Szostak, J. W., Young, A. T., Wu, R., Consaul, S., and Warner, J. (1981) Nature (Lond.) 289, 733-737

M. E. Sripati, C. E., Groner, Y., and Warner, J. R. (1976) B. Biol. Chem. 251, 2988-2994

R. A. Serjeantson, S. K. A., and Sripati, C. E. (1979) Nucleic Acids Res. 6, 933-944
