MOLECULAR AND CELLULAR BIOLOGY, Jan. 1983, p. 20–31
0270-7306/83/010020-12$02.00/0

Vol. 3, No. 1

Characterization of Transposable Element-Associated Mutations that Alter Yeast Alcohol Dehydrogenase II Expression

VALERIE M. WILLIAMSON,[1]†* DAVID COX,[1] ELTON T. YOUNG,[1] DAVID W. RUSSELL,[2]‡ AND MICHAEL SMITH[2]

Department of Biochemistry, University of Washington, Seattle, Washington 98195,[1] and Department of Biochemistry, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5[2]

Received 21 May 1982/Accepted 8 October 1982

Seven *cis*-dominant, constitutively expressed mutations of the normally glucose-repressible isozyme of alcohol dehydrogenase (ADHII) from the yeast *Saccharomyces cerevisiae* are caused by insertion of transposable elements from the Ty1 family in front of the ADHII structural gene (*ADR2*) (V. M. Williamson, E. T. Young, and M. Ciriacy, Cell 23:605–614, 1981). We cloned *ADR2* with its associated Ty1 element from five *S. cerevisiae* strains carrying these mutations. Comparison of the Ty1 elements by heteroduplex studies and restriction enzyme analyses indicated that four were very similar; the fifth, although the same size as the others (about 5.6 kilobases), differed by the presence of two large substitutions of approximately 1 and 2 kilobases. The DNA sequences of the terminal direct repeats (deltas) were very homologous but not identical and were similar to previously reported Ty1 element direct repeats. We determined the 5′-flanking sequences of the *ADR2* gene isolated from a wild-type strain and from five Ty1-associated mutations. The 5-base pair target sequence at the site of Ty1 insertion was present at both ends of each Ty1 element. The sites of insertion of the elements were all different and occurred from 125 to 210 base pairs in front of the coding region of *ADR2*. The 5′ end of the major transcript as determined by S1 mapping was the same in wild-type cells and in Ty1-associated constitutive mutants and was approximately 54 base pairs upstream from the coding region. *ADR2* transcripts were not detected when a solo delta sequence was present in the 5′-flanking region of this gene.

In eucaryotic organisms the insertion of a mobile DNA sequence in the vicinity of a structural gene is frequently associated with altered expression of that gene. Examples of this phenomenon include the association of a *copia* transposable element with the *white* locus in the *white-apricot* mutation of *Drosophila melanogaster* (4, 5, 22) and the increased expression of a putative cellular oncogene (c-*myc*) adjacent to inserted proviruses of avian leukosis virus in mice (24, 32). Several mutants in *Saccharomyces cerevisiae* have been characterized which are altered in gene expression due to the insertion of a 5.6-kilobase (kb) transposable element, Ty1, upstream from the structural gene. A variety of phenotypes have been associated with this type of insertion. For instance, transposition of Ty1 to a location upstream from the *HIS4* gene can produce a His$^-$ phenotype (33), but insertion of Ty1 into the 5′-flanking region of the gene encoding iso-2-cytochrome *c* results in a 20-fold increase in expression of this gene (18). We have presented evidence that constitutive expression of the glucose-repressible isozyme of alcohol dehydrogenase (ADHII) can be caused by insertion of a Ty1 element into the 5′-flanking region of the gene (42).

S. cerevisiae contains three electrophoretically distinct isozymes of alcohol dehydrogenase, ADHI, ADHII, and mADH (10). ADHI, the classical fermentative isozyme, is present when yeast cells are grown in the presence of glucose; ADHII is repressed by glucose and expressed when glucose is absent from the medium; and mADH is localized in the mitochondria. ADHI, ADHII, and mADH are coded by three unlinked genetic loci, *ADC1*, *ADR2*, and *ADM*, respectively (10, 41). Mutants which express ADHII in the presence of glucose due to alterations at a locus (*ADR3*) which is tightly linked to the structural gene have been described (10, 11). *S. cerevisiae* strains with this type of mutation

† Present address: ARCO Plant Cell Research Institute, Dublin, CA 94568.
‡ Present address: Department of Molecular Genetics, University of Texas Health Science Center, Dallas, TX 75235.

(*ADR3*^{c} mutations) differ in the level and regulation of ADHII activity (10, 11). We have presented evidence that in seven of nine cases analyzed this alteration in regulation is caused by insertion of a Ty1 element into the 5'-flanking region of the gene (42).

Ty1 elements consist of an internal 5.3-kb fragment of DNA (epsilon DNA) flanked by direct repeats (deltas) of about 0.3 kb (8, 21). There are approximately 30 copies of the complete element and over 100 copies of Ty1-associated and solo delta sequences in the haploid *S. cerevisiae* genome of laboratory strains, although the number and distribution vary considerably from strain to strain (8, 15, 20). These elements are transcribed into polyadenylated RNA species which are initiated in one direct repeat and terminated in the other (R. T. Elder, personal communication). Structurally similar transcripts are observed from *copia* elements of *Drosophila* and from the integrated proviral forms of various retroviruses. There is some sequence homology among the terminal direct repeats from Ty1 elements, *copia* elements, and retroviruses, and a common origin for these DNA sequences has been proposed (15, 26, 28, 40).

Little is presently known about the mechanism by which these insertions alter the regulation or expression of adjoining genes. In the case of increased production of c-*myc* associated with insertion of the avian leukosis virus provirus, it has been shown that in several, but not all, cases the c-*myc* mRNA contains retroviral sequences, implying that one mechanism for increased transcription is read-through from retroviral sequences (24, 32). Although yeast Ty1 elements are transcribed, their major transcripts start in the delta sequence adjacent to the affected *ADR2* gene and proceed in the opposite direction (Elder, personal communication). Such a conformation also occurs in some avian leukosis virus-induced lymphomas (32). The altered gene expression in these cases is likely to be due to a different mechanism.

The existence of several Ty1 insertion mutations of *ADR2* provides an opportunity to study the mechanism of Ty1-altered gene expression by relating the site of insertion and the type of Ty1 element inserted to the effects on ADHII levels. Previous work showed that both the sites of insertion and the restriction sites of the elements themselves vary (42). All Ty1 elements are inserted in the same orientation with respect to the adjacent *ADR2*, and this same orientation is found for the Ty1 element associated with the iso-2-cytochrome *c* mutation *CYC7-H2* (17, 18). Mutants which have lost ADHII activity have been isolated from Ty1 insertion mutants (13) and yield further insights into mechanisms involved in Ty1 regulation. Most of these mutants have lost the Ty1 element adjacent to *ADR2* but have retained a single delta sequence in its place. Other secondary mutants contain chromosomal rearrangements or mutations at unlinked sites. The His$^-$ phenotype caused by Ty1 insertion is also altered by loss of the Ty1 element with retention of a single delta sequence, by DNA rearrangements which appear to be associated with the Ty1 element, and by mutations at unlinked loci (9, 33).

Previously we cloned the *ADR2* gene and the adjoining Ty1 sequence from one constitutive mutant (mutation *ADR3-2*^{c}) (42). In this paper we describe the cloning of four other mutant loci (*ADR3-3*^{c}, *ADR3-6*^{c}, *ADR3-7*^{c}, and *ADR3-8*^{c}) and compare the cloned DNA sequences by electron microscopy and restriction enzyme mapping. We determine the DNA sequence of the 5'-flanking region of the wild-type *ADR2* gene, the insertion sites of the five cloned Ty1 elements into this DNA, and the terminal direct repeats of each of these Ty1 elements. In addition, we analyze *ADR2* transcripts from wild-type and mutant strains.

MATERIALS AND METHODS

Yeast strains and plasmids. The *S. cerevisiae* strains used in this work were obtained from M. Ciriacy and have been previously described (10, 11, 13). Strain 43-2B, which is deficient in ADHI and mADH activities (mutations *adc1-11* and *adm*, respectively), was originally derived from strain 7972C (Seattle stock collection). Strains containing the *ADR3*^{c} mutations *ADR3-2*^{c}, *ADR3-3*^{c}, *ADR3-6*^{c}, *ADR3-7*^{c}, and *ADR3-8*^{c} were isolated from strain 43-2B or a related strain by M. Ciriacy after treatment of haploid cells with 1% ethyl methane sulfonate or UV irradiation (10, 11). They were selected by their ability to grow on antimycin A-containing medium (11). Mutants M203 and M601 are derivatives of strains containing mutations *ADR3-2*^{c} and *ADR3-6*^{c}, respectively, which have little or no ADHII activity and appear to have a single delta sequence in front of the *ADR2* gene (13).

Plasmid A2M, which contains the wild-type *ADR2* gene, was isolated by complementation of function by M. Ciriacy from a pool of yeast DNA fragments in plasmid YRp7 (39). This pool was constructed by C. Denis, using DNA from a yeast strain derived from 7972c, the parent of the constitutive strains described here (C. L. Denis, Ph.D. thesis, University of Washington, Seattle, 1982).

Cloning and restriction enzyme analyses of DNA fragments. Isolation of the DNA restriction fragments containing *ADR2* and the associated Ty1 element from *ADR3-3*^{c}, *-6*^{c}, *-7*^{c}, and *-8*^{c} was carried out as previously described for mutation *ADR3-2*^{c} (42). To clone *ADR3-6*^{c}, *-7*^{c}, and *-8*^{c}, *Bam*HI restriction enzyme fragments of the appropriate size were isolated from a preparative agarose gel of genomic DNA fragments. These fragments were then ligated into the *Bam*HI site of plasmid YRp7 (39), used to transform *Escherichia*

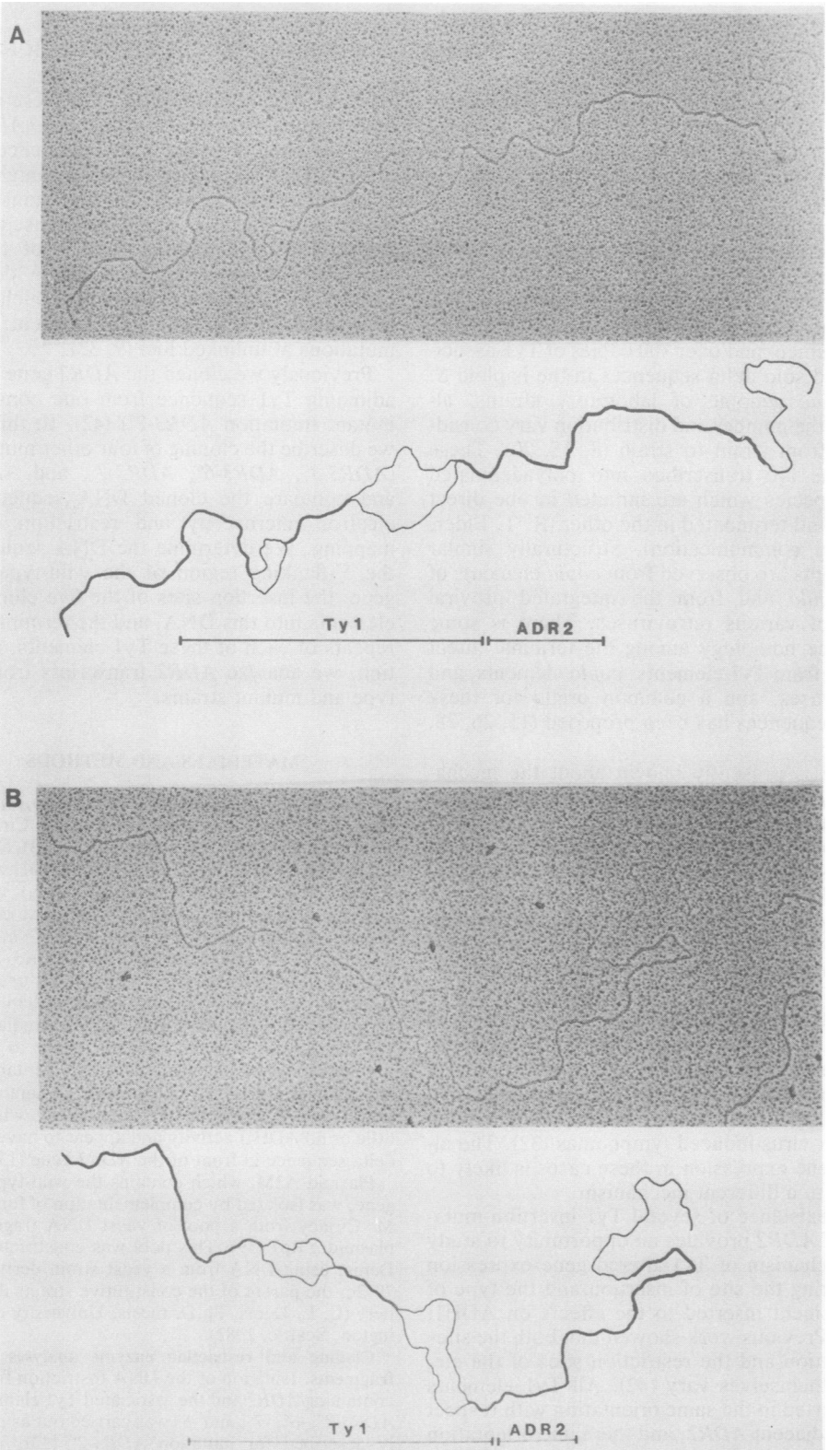

FIG. 1. Heteroduplex molecules formed between Ty1-containing *ADR2* clones. Heteroduplexes are between restriction fragments containing *ADR3-3*^{c} and *ADR3-6*^{c} (A) and *ADR3-3*^{c} and *ADR3-7*^{c} (B). Plasmid-containing *ADR3-3*^{c} was digested with *Eco*RI and plasmid containing *ADR3-6*^{c} (A) or *ADR3-7*^{c} (B) was digested with *Bam*HI. Samples were prepared as previously described (42). The approximate locations of Ty1 and *ADR2* sequences are indicated below the line diagrams.

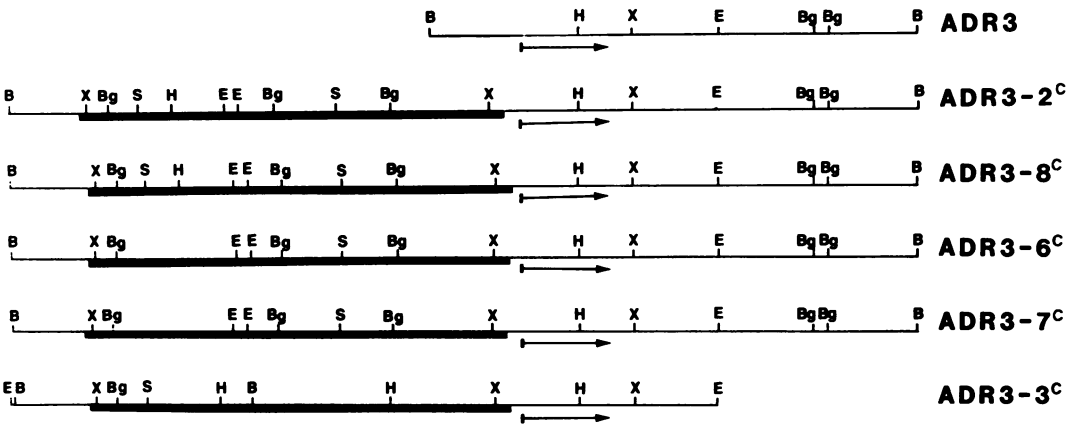

FIG. 2. Comparison of restriction endonuclease cleavage sites of *ADR2* regions from wild-type and mutant yeast strains. Plasmids containing the cloned *ADR2* regions from wild-type and mutant cells were mapped by cleavage of purified plasmids with *Bam*HI (B), *Hin*dIII (H), *Xho*I (X), *Eco*RI (E), *Bgl*II (Bg), and *Sal*I (S). Heavy lines represent Ty1 sequences. Arrows indicate position and direction of *ADR2* transcription and size of the transcript (1.3 kb).

coli cells to ampicillin resistance, and identified by colony hybridization, using the previously cloned wild-type *ADR2* gene as a probe. *ADR3-3*^{c} was cloned by isolating a genomic *Eco*RI restriction fragment and cloning into the *Eco*RI site of pBR322. Techniques used to screen for and to analyze the *ADR3*^{c}-containing plasmids have been described before (42).

DNA sequence analysis. The DNA sequence of the 5'-flanking region of the *ADR2* gene was determined by using the following techniques. M13 cloning and sequencing were performed as described by Messing et al. (30). Chemical sequencing was carried out as described by Maxam and Gilbert (29). The oligonucleotide T_8CA, a kind gift of Shirley Gillam, was used as a primer on linear double-stranded DNA as follows: 1.5 μg (0.25 pmol) of plasmid A2M was digested with the restriction enzyme *Eco*RI to linearize the DNA. The digested plasmid (10 μl) was combined with 0.002 optical density unit of primer (2 μl), sealed in a capillary, and boiled in a water bath for 3 min. After this denaturation step, the capillary was submerged in a beaker of ice water. After 1 min, the primer-template solution was divided evenly among four capillaries, and Sanger "terminator" sequencing (35) was performed at room temperature or at 39°C. The DNA sequence of the direct repeats was determined by 3'-end labeling of the internal *Xho*I site (37) and subsequent chemical sequencing. To confirm this sequence, an *Fnu*EI fragment beginning in the right delta and extending into the coding sequence of the *ADR2* gene was cloned and sequenced in M13 mp7.

S1 analysis of transcripts. S1 mapping of the 5'-endpoints of the *ADR2* mRNA was carried out with techniques described by Nasmyth et al. (31) as modified below to map the endpoints of the transcripts at the *MAT* loci. The DNA probe used was prepared by isolating the 214-base pair (bp) *Sph*I-*Hin*fI fragment which contains the translation initiation codon. This fragment was treated with bacterial alkaline phosphatase, labeled with ^{32}P at its 5' end, and strand separated by techniques described by Maxam and Gilbert (29). The single-stranded DNA fragment which was labeled at the *Hin*fI end, 44 nucleotides within the coding region, was used as a hybridization probe. The RNA used was total RNA extracted from yeast cells essentially as described by Shultz (36). Preliminary assays done with increasing amounts of RNA demonstrated that, in RNA excess, hybridization to *ADR2* mRNA was favored over hybridization to *ADC1* mRNA because of the shorter stretch of homology of the probe with *ADC1* mRNA (data not shown). Therefore, to allow estimation of the amounts of both *ADR2* and *ADC1* mRNA and to reduce the possibility of failing to detect shorter *ADR2* mRNA species, experiments presented in this paper were done in DNA excess. The S1-digested DNA-RNA duplex was denatured in 90% formamide rather than in base before electrophoresis on DNA sequencing gels (30).

RESULTS

Comparison of cloned *ADR3*^{c} mutants. DNA restriction enzyme fragments containing *ADR2* and the associated Ty1 element were isolated from *S. cerevisiae* strains containing the mutations *ADR3-3*^{c}, *-6*^{c}, *-7*^{c}, and *-8*^{c} essentially as previously described for *ADR3-2*^{c} (42; see Materials and Methods). To investigate the extent of homology between Ty1 elements, heteroduplex analysis was carried out between the *ADR2*-containing fragments of *ADR3-3*^{c} and *-6*^{c}, *ADR3-3*^{c} and *-7*^{c}, *ADR3-2*^{c} and *-6*^{c}, and *ADR3-6*^{c} and *-7*^{c}. In the last two cases no loops or bubbles were seen (data not shown). Minor sequence divergence would not be observable, however, under the hybridization conditions which we used. Heteroduplexes formed between *ADR3-3*^{c} and *ADR3-6*^{c} and between *ADR3-3*^{c} and *ADR3-7*^{c} showed the presence of two large bubbles, indicating that two large regions of nonhomology existed between these elements (Fig. 1). Measurements of the heteroduplexes showed

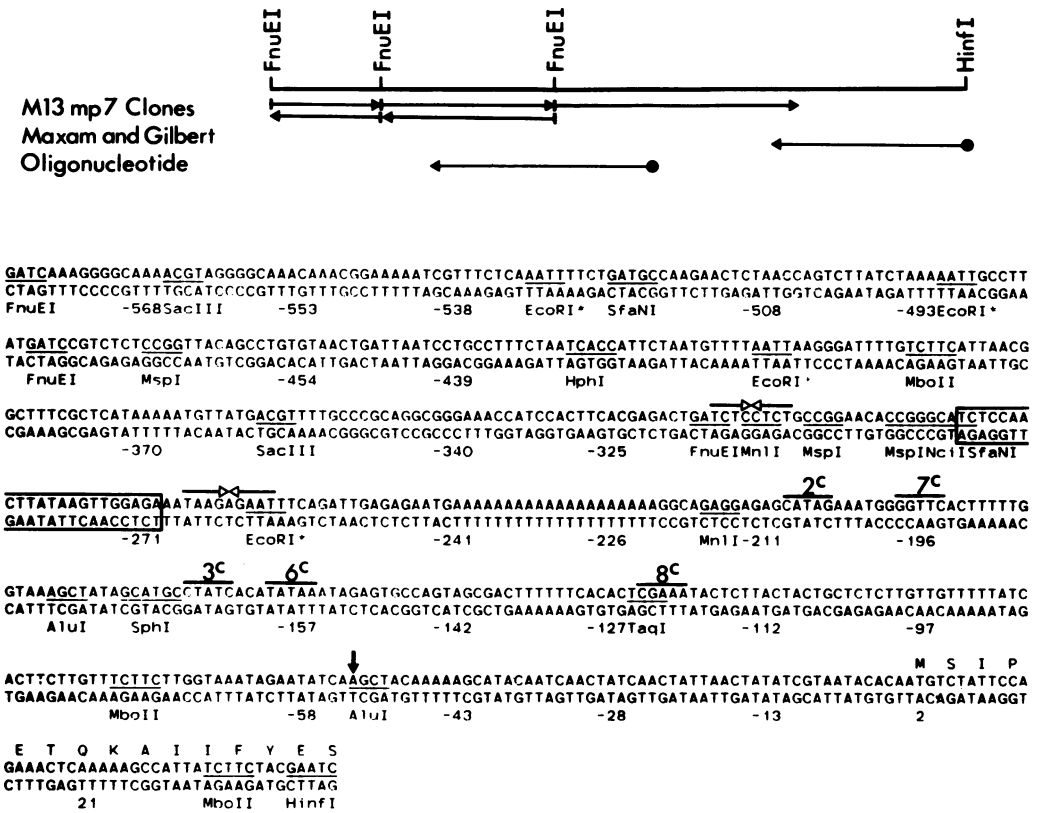

FIG. 3. DNA sequencing strategy and primary structure of the 5'-flanking region of the *ADR2* gene. Three methods of sequence analysis were used to determine the DNA sequence as described in the text. The methods, individual experiments, and the extent of sequence determined from each are indicated at the top. The bottom depicts the sequence of both strands of the DNA. Numbering is centered on the first nucleotide of the coding sequence. A 22-bp region of dyad symmetry in the DNA is boxed. Converging arrows indicate inverted repeats. Overlined sequences represent the sites of insertion of the five transposable elements associated with constitutive expression of the gene. A vertical arrow indicates the major 5' end of the mRNA detected by nuclease S1 mapping. Key restriction sites are labeled.

that the approximate sizes (1 and 2 kb) and positions (approximately 2.0 and 4.0 kb from the left end of the Ty elements) of these bubbles corresponded to those observed by Kingsman et al. (27) for heteroduplex analysis of elements Ty1 and Ty1-17.

By restriction enzyme analysis the inserted elements in *ADR3-2^c*, *-6^c*, *-7^c*, and *-8^c* were quite homologous (Fig. 2). No restriction site differences were seen between the elements in *ADR3-2^c* and *-8^c* or between those in *ADR3-6^c* and *-7^c* with the restriction enzymes shown in Fig. 2. The restriction sites for *Sal*I and *Hin*dIII which were present near the *ADR2*-distal end of the element in the *ADR3-2^c* and *-8^c* clones were not present in the *ADR3-6^c* or *-7^c* clones. As expected from heteroduplex analysis, there were several differences in restriction enzyme cleavage sites between the Ty1 insert in mutation *ADR3-3^c* and the others. Comparison of the positions of the substitutions in the heteroduplex (Fig. 1) to the restriction maps (Fig. 2) indicated that the small substitution bubble which was about 1 kb long in the heteroduplex contained the *Eco*RI sites of the elements in *ADR3-2^c*, *-6^c*, *-7^c*, and *-8^c* and the *Hin*dIII and *Bam*HI sites in the *ADR3-3^c*-associated element. The larger bubble, which was approximately 2 kb long, included the *Bgl*II site most proximal to *ADR2* in the first class of elements and the *ADR2*-proximal *Hin*dIII site in the second class.

DNA sequence analysis of the wild-type *ADR2* 5'-flanking sequences. The strategy used to determine the DNA sequence of the 5'-flanking region of the wild-type *ADR2* gene (clone A2M) is outlined in Fig. 3 and detailed in Materials and Methods. Individual *Fnu*EI fragments were isolated, cloned in both orientations, and sequenced in the single-stranded M13 mp7 vector as described by Messing et al. (30). Chemical sequencing methods (29) were used to obtain an overlap in the region immediately adjacent to the

coding sequence. To align the various *Fnu*EI fragments and to confirm the indicated DNA sequence, the synthetic oligodeoxyribonucleotide T_8CA was used as a primer, with plasmid A2M as template (23, 37). The sequence shown in Fig. 3 extends from an *Fnu*EI site at position −582 relative to the initiation codon of *ADR2* to a *Hin*fI site at position +44 within the coding sequence.

A number of interesting features were noted in the 5′-flanking DNA sequence of the *ADR2* gene. There was a sequence TATAAA identical to the "Goldberg-Hogness TATA box" at position-160 (M. Goldberg, Ph.D. thesis, Stanford University, Stanford, Calif., 1979). The DNA sequence to the right of this region and up to the coding sequence was 50% homologous with the gene (*ADC1*) encoding the fermentative ADH isozyme ADHI (2; D. W. Russell, M. Smith, V. M. Williamson, and E. T. Young, J. Biol. Chem, in press). A block of 20 consecutive adenine residues occurred from positions −242 to −223. This sequence was centered in a region of DNA extending from −275 to −198 in which the distribution of purines and pyrimidines was decidedly biased; 85% of the bases in the mRNA identical strand were purines. The converse was true for the sequence from −118 to −71, in which 83% of the bases were pyrimidines. The DNA sequence to the left of the block of A's contained a 22-bp region of perfect dyad symmetry from positions −292 to −271. Inverted repeats of 8 and 9 bp flanked this sequence extending 5′ to 3′ from positions −316 and −268, respectively.

DNA sequence analysis of transposable element-induced constitutive mutants. Southern blotting analysis of genomic DNA isolated from various Ty1-associated constitutive mutants of *ADR2* predicted slight differences in the site of Ty1 insertion (42). We determined the precise integration point of the five cloned elements, using the DNA sequencing strategy outlined in Materials and Methods. As has been seen at other Ty1 element insertion sites (19, 21), 5-bp duplications of the target DNA were generated. The duplicated sequences are listed in Table 1, and the sites of the insertion into the wild-type DNA sequence are indicated by overlines in Fig. 3. The adenine-plus-thymine (A+T) content of

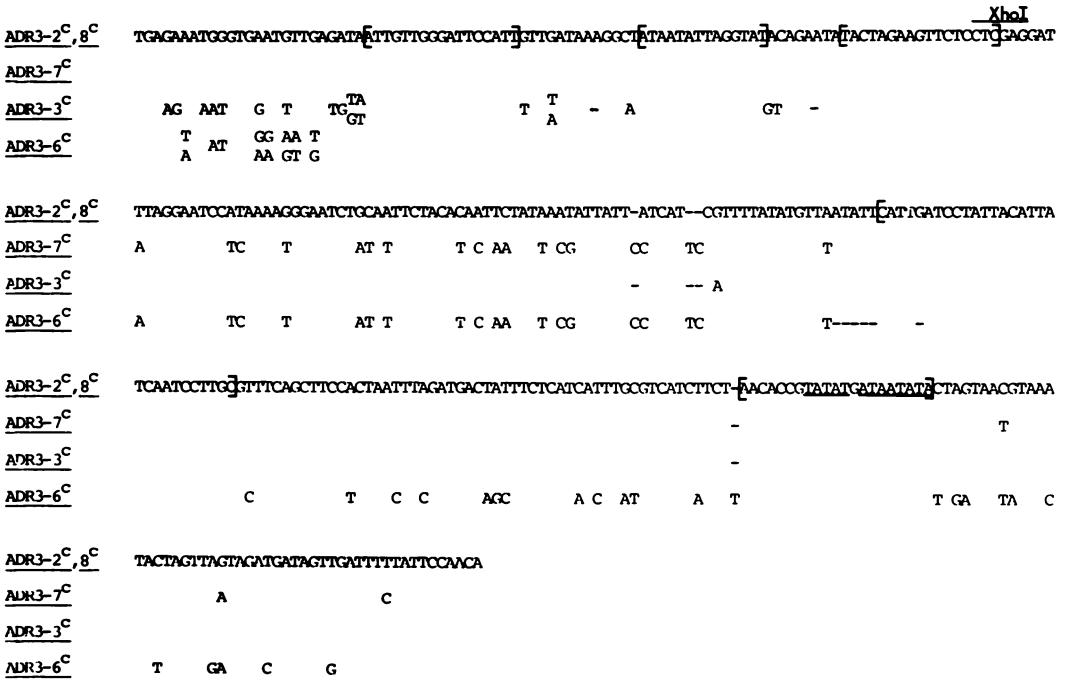

FIG. 4. Primary structure of the direct repeats flanking the cloned transposable elements. The complete sequence of the *ADR3-2c*, *-8c* delta is shown. Nucleotides which differ from this in the other delta sequences determined are indicated. Gaps maximize homology between direct repeats. Where differences exist between the deltas of a transposable element, the raised letter indicates the nucleotide in the left delta as represented in Fig. 2 and the lowered letter indicates the nucleotide in the right delta. The position of the *Xho*I site present in all of these DNAs is indicated. Brackets indicate longer conserved sequences among all Ty1-associated repeat sequences published to date. The underlined sequences contain potential TATA-box sequences for *ADR3-8c* (see text).

the 5-bp duplication ranges from 40% for the *ADR3-7*^{c} mutation to 100% for *ADR3-6*^{c}. The elements were inserted in the same orientation within a region lying from 210 to 125 bp 5′ to the coding sequence. Three of the five elements inserted to the left of the sequence TATAAA at position −160, whereas one Ty1 (mutation *ADR3-6*^{c}) inserted directly into this region of the DNA, resulting in duplication of the sequence TATAA at the left end of the Ty1 element. In a fifth mutation, *ADR3-8*^{c}, Ty1 transposed to a site to the right of the sequence TATAAA, at position −125.

The DNA sequences of the Ty1 direct repeats (delta sequences) from the five cloned *ADR3*^{c} mutations were determined as described in Materials and Methods and are presented in Fig. 4. The delta sequences averaged 70% (A+T), varied in length from 332 to 336 bp, and were very homologous. Sequence differences among the family of deltas were most abundant at the left end (Fig. 4), but some differences are distributed throughout the remainder of the repeat. The regions in brackets represent DNA sequences which were conserved among all published delta sequences flanking Ty1 elements (19, 21). The underlined conserved sequence contains several alternating TA arrangements which resemble a TATAAA box (see Discussion). The transposable elements of the *ADR3-2*^{c} and *ADR3-8*^{c} mutations which had identical restriction maps also had identical sequences in their direct repeats. However, the *ADR3-6*^{c} and *ADR3-7*^{c} elements had the same restriction maps, but differences existed in their direct repeat sequences. Three of the five pairs of direct repeat sequences had no differences between their left and right members. The left delta of the *ADR3-3*^{c} mutation differed in three positions from the right delta, and six differences occurred between the deltas flanking the *ADR3-6*^{c} element.

TABLE 1. ADH activity, integration sites, and duplicated sequences of Ty1-associated *ADR3*^{c} mutations

Mutant	Insertion site[b]	5-bp duplications[c]	ADHII activity (μg/mg of protein)[a]	
			Glucose	Ethanol
Wild type			15	1,850
ADR3-2^c	−210	CATAG	630	950
ADR3-7^c	−199	GGTTC	522	746
ADR3-3^c	−169	CTATC	468	220
ADR3-6^c	−161	TATAA	738	1,590
ADR3-8^c	−125	CGAAA	184	183

[a] Taken from Williamson et al. (42).
[b] Base pairs upstream from coding region.
[c] Sequences which were duplicated at the site of Ty1 integration.

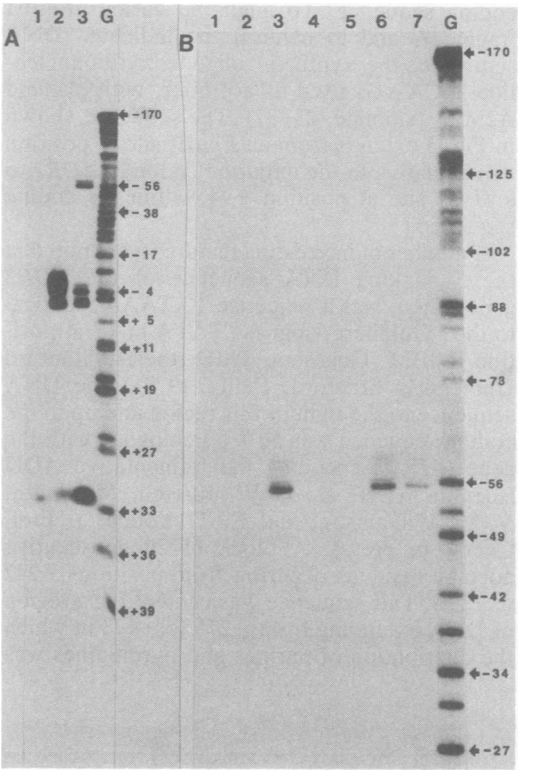

FIG. 5. Analysis of the 5′ end of *ADR2* transcripts by S1 mapping. A single-stranded DNA fragment with its 5′ end 44 nucleotides within the structural gene for *ADR2* and its 3′ end 170 bp upstream from the *ADR2* coding region was labeled at its 5′ end with ^{32}P and hybridized to total yeast RNA. The hybrids were treated with S1 nuclease as described by Nasmyth et al. (31), denatured in 90% formamide, and analyzed on an 8% polyacrylamide sequencing gel. The standard (G) contains fragments after guanine-specific cleavage (29) of the same DNA fragment used above. Numbers on the right indicate nucleotides from the transcription initiation codon. (A) DNA was electrophoresed for 1 hour before autoradiography. Lane 1, no RNA control; lane 2, RNA isolated from a wild-type strain (7972C; *ADC1 ADR2 ADR3*) grown logarithmically in the presence of glucose; lane 3, RNA isolated from wild-type cells which had been shifted to ethanol-containing medium for 2 h. (B) DNA was electrophoresed for 3 h before autoradiography. RNA was extracted from the following mutant and wild-type strains grown as indicated below: lane 1, ethanol-grown cells from mutant M203, a delta-only derivative of *ADR3-2*^{c}; lane 2, ethanol-grown cells from mutant M601, a delta-only derivative of *ADR3-6*^{c}; lane 3, glucose-grown cells from a yeast strain containing mutation *ADR3-9*^{c} which contains a Ty1 element resembling that found in *ADR3-2*^{c}; lanes 5, 6, and 7 contain RNA from strain 7972C grown on glucose (lane 5), transferred to ethanol-containing medium for 2 h (lane 7) or 6 h (lane 8). Lane 4 is a no-RNA control. All lanes used approximately the same amounts of total RNA (1 μg) except lane 8, which used about 0.5 μg.

A 5-bp sequence, TACCA, was present at all epsilon-delta junctions of the five cloned *ADR2*-associated Ty1 elements. This same junction sequence is seen for the Ty1 elements isolated by Farabaugh and Fink (19) and Gafner and Philippsen (21).

mRNA transcript mapping of *ADR3*^{c} mutants. The 5′ ends of *ADR2* mRNA isolated from wild-type and mutant strains were determined by S1 nuclease mapping (3, 31) as described in Materials and Methods. The hybridization probe used, a 218-nucleotide, single-stranded DNA fragment whose labeled 5′ end is 44 nucleotides within the coding region of *ADR2*, was prepared as described in Materials and Methods. The lengths of the DNA fragments protected from S1 nuclease digestion were determined on sequencing gels (Fig. 5). By this method the 5′ end of the major mRNA coding for *ADR2* was located approximately 54 bp upstream from the initiation codon (Fig. 5). Four minor transcription start sites were also seen upstream from this region. S1 analysis indicated that *ADR2* mRNA is not present in glucose-grown cells (Fig. 5A, lane 2), but is present in yeast cells when ethanol was used as the carbon source (Fig. 5A, lane 3). These data support the results obtained by in vitro translation by Denis et al. (14), i.e., that ADHII activity reflects mRNA levels.

A series of bands is seen in Fig. 5A, lanes 2 and 3 (and also in Fig. 6), representing DNA fragments ending 2 to 10 nucleotides 5′ to the coding region. These bands are due to hybridization of the *ADR2* probe to mRNA coding for the other major isozyme of alcohol dehydrogenase (ADHI, genetic locus *ADC1*; 2, 10, 41). Evidence for this assertion is that these bands are absent when mRNA from a yeast strain containing a deletion of the 5′ end of *ADC1* is analyzed by S1 mapping (V. M. Williamson, D. Beier, and E. T. Young, *in* P. F. Lurquin and A. Kleinhofs, ed., *Genetic Engineering in Eukaryotes*, in press). The amount of *ADC1* mRNA, as deduced from the intensity of the lower bands in lanes 2 and 3 of Fig. 5A (and also in Fig. 6, lane 7), is decreased in ethanol-grown cells compared with glucose-grown cells. This is consistent with the 5- to 10-fold decrease in the level of translatable *ADC1* mRNA which occurs when cells are transferred from glucose- to ethanol-containing medium (C. L. Denis, J. Ferguson, and E. T. Young, J. Biol. Chem., in press).

S1 analysis was also used to locate the 5′ ends of *ADR2* mRNA in mutant strains and to estimate the relative mRNA levels in these strains (Fig. 5B and 6). S1 mapping of RNA isolated from Ty1-associated strains indicated that the 5′ ends of the *ADR2* mRNAs were identical to each other and to those from wild-type cells (Fig. 6). The relative intensities of the bands correlated with the ADHII activities found in these mutants (Fig. 6; Table 1). Secondary mutants derived from strains containing the Ty1 element-associated constitutive mutations *ADR3-2*^{c} and *ADR3-6*^{c} have little or no *ADHII* activity under any growth conditions (13). The majority of these mutants appeared to retain a single delta sequence in front of the *ADR2* gene. S1 analysis failed to detect *ADR2* transcript in the two delta-only mutant strains tested (Fig. 5B, lanes 1 and 2).

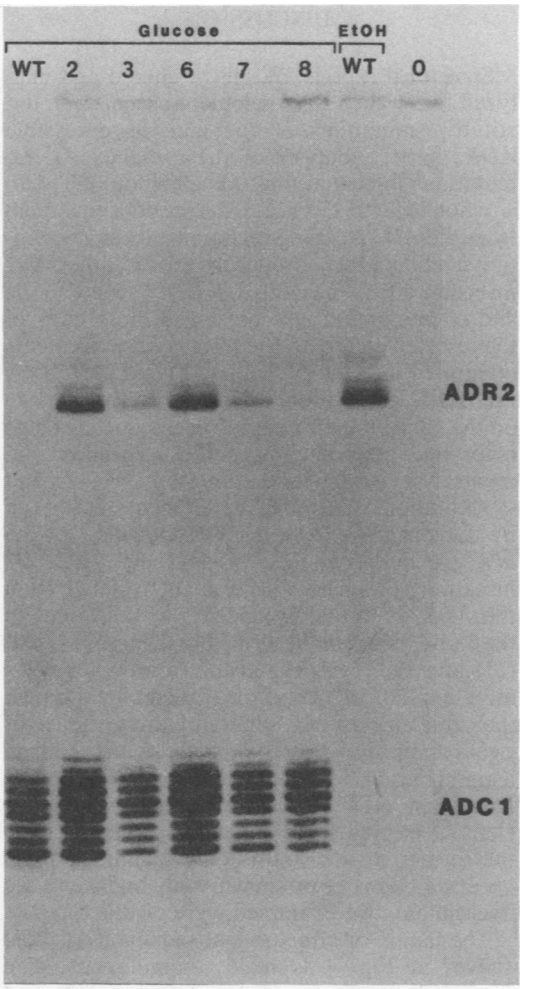

FIG. 6. Determination of the 5′ ends of *ADR2* from Ty1-caused constitutive mutations by S1 mapping. Total RNA was isolated from wild-type and mutant yeast cells grown in the presence of glucose and from wild-type cells grown for 2 days on yeast medium plus ethanol (EtOH) as indicated. The 5′ ends of the RNA were determined as described in the legend to Fig. 5. Wells marked WT contain RNA from strain 43-2B (*adc1 ADR2 ADR3*), a strain closely related to the mutants. Wells 2, 3, 6, 7, and 8 analyze RNA from mutants containing *ADR3-2*^{c}, *-3*^{c}, *-6*^{c}, *-7*^{c}, and *-8*^{c}, respectively. 0 is a no-RNA control.

DISCUSSION

Structural aspects. We have cloned and analyzed five transposable element-associated constitutive mutations of the glucose-repressible *ADR2* gene. Analysis of these cloned *ADR3*^{c} mutations indicated that transposition of Ty1 to a region located 125 to 210 nucleotides upstream from the *ADR2* coding region results in a constitutive phenotype. As described for other Ty1 insertion sites, the 5-bp DNA sequence at the site of integration was present at both ends of the inserted element. The transposable elements integrated into the 5'-flanking region (*ADR3*) of the *ADR2* gene can be divided into two classes on the basis of their restriction enzyme cleavage maps and heteroduplex studies. The Ty1 elements associated with the *ADR3-2*^{c}, *-6*^{c}, *-7*^{c}, and *-8*^{c} mutations resemble the elements described by Cameron et al. (8), those responsible for the *his4-912* mutation (19, 33) and the *CYC7-H2* mutation (17). The Ty1 element isolated from the *ADR3-3*^{c} mutation appears to be similar to the Ty1-17 element described by Kingsman et al. (27) and the Ty1 responsible for the *his4-917* mutation (20, 33). This type of element, although the same size as Ty1, differs by the presence of two substitutions approximately 1 and 2 kb in length (Fig. 1). Results of Roeder et al. (33) and Kingsman et al. (27) indicate that the latter element is present in lower copy number than the former in the strains they examined. This observation may explain why only one out of the five mutants we examined were of this type.

The family of direct repeat sequences (deltas) shown in Fig. 4 is highly homologous. The extended sequences enclosed in brackets and some shorter sequences are conserved among all of the Ty1-associated direct repeats studied to date (this work; 21, 33). The sequences found at the end of all other sequenced Ty1-associated deltas, 5' TGA. . . . .TATTCCAACA3', were also seen here. Although the function of these conserved sequences is not known at present, they are likely to be involved in transcription from the Ty1 element or transposition or both. The direct repeat sequences of the *ADR3-2*^{c}- and *ADR3-8*^{c}-associated elements differ at only eight positions from those flanking the *his4-912*-associated element (19), whereas the *ADR3-6*^{c} deltas more closely resemble (29 differences) the direct repeats at the ends of the Ty1 elements B10 and D15 described by Gafner and Philippsen (21). It is interesting that the *ADR3-3*^{c} Ty element and element Ty-917 (33), which have very similar restriction sites in their epsilon DNA, also have very homologous delta sequences (only 12 differences between the right delta sequences of these two elements) (19). The terminal repeats flanking the *ADR3-7*^{c} element are homologous to the *ADR3-2*^{c} direct repeats in certain stretches and to *ADR3-6*^{c} direct repeats in other regions. Very few differences (19 of 6,026 bp) are unique to a given direct repeat. The patchwork homologies between direct repeats flanking different Ty1 elements and the paucity of unique base pair changes suggests that these DNAs may interact through gene conversion (1).

In procaryotes, two characteristics of DNA are associated with an increased frequency of insertion of movable genetic elements into a given sequence. These properties are homology between the inserting element and the target DNA and a high A+T content in the target DNA (7). A computer search of the 5'-flanking sequence of the *ADR2* gene did not reveal any extensive homology with the delta sequences. All of the Ty1 integrations reported to data occur in A+T-rich regions of DNA. It is also of note that all of the characterized integration sites of Ty1 elements are between yeast structural genes rather than within them. One might postulate that some feature of DNA structure, for example, the adenine nucleotide tract beginning at position −242 of the *ADR2* gene (Fig. 3), could promote Ty1 integration. Because we selected for a specific phenotype, our sample of integration sites is biased for insertion into the regulatory region of *ADR2*.

Ty1 elements and gene expression. Determination of the 5' end of *ADR2* mRNA from wild-type and constitutive mutants by S1 analysis indicated that they are the same in wild-type and *ADR3*^{c} mutant strains. However, since S1 analysis measures the length of continuous hybridization of the genomic DNA probe which we used, it is possible that the 5' end determined by this method is actually a processing site and not a transcription start site. Reverse transcriptase sequencing (Russell et al., in press) was also carried out with polyadenylated RNA isolated from derepressed cells, and the 5'-end sequence of the *ADR2* mRNA determined by this method agreed with results determined by S1 mapping. This indicates that the 5' end determined by S1 mapping is not a splice junction since, if this were the case, a longer and different sequence would be obtained with the reverse transcriptase sequencing. The possibility still remains that the original 5' end is rapidly cleaved off after synthesis.

Transposition of a Ty1 element into the region bounded by −210 and −125 results in varying degrees of constitutive expression of the *ADR2* gene (Table 1; 11, 12). These Ty1 insertion mutants differ from wild-type strains not only in the level of expression of the adjoining genes, but also in the regulation of that expression. In wild-type cells ADHII activity is repressed >100-fold when cells are grown on medium

containing glucose compared with levels on medium containing ethanol as the carbon source. However, in Ty1-associated mutants the levels of ADHII after growth on ethanol-containing medium are 0.5- to 2.0-fold those found after growth on glucose-containing medium and vary with the particular mutant (see Table 1). In comparing ADHII activity with Ty1 integration site, we noted that *ADR2* expression on glucose-containing medium is more than twofold lower for the mutation *ADR3-8*^{c}, in which the Ty1 is integrated to the right of the sequence TATAAA, than for any of the other *ADR3*^{c} mutations. The TATA sequence has been implicated in determining the site of transcription initiation in eucaryotic cells (6). For wild-type *ADR2* as well as *ADC1* the position of this sequence is approximately 100 bp to the left of the transcript start site instead of 30 bp to the left as is seen with other eucaryotic genes (2; Russell et al., in press). It is possible that although this sequence may not perform the same function as the Hogness-Goldberg box (Goldberg, Ph.D. thesis), it is important for efficient transcription. An examination of the *ADR3-8*^{c} delta sequence (Fig. 4) indicates that several "TATA" sequences are located approximately 130 bp 5′ to the site of *ADR2* transcript initiation. The 30-bp displacement of these sequences from that in wild type or differences in the exact sequence or both may be responsible for the lower *ADR2* expression in this mutant.

How Ty1 elements produce constitutive expression of *ADR2* is not clear. The transcript mapping described above indicates that the altered expression is not due to read-through from a transcription initiation site within the element. We previously postulated two general mechanisms by which such regulation could occur (42). One hypothesis was that sequences responsible for repression upstream from the gene are moved too far away to be effective, and the other hypothesis proposed that the inserted sequences themselves played a role in altering the regulation. In support of the former hypothesis, removal of sequences 5′ to the *Sph*I restriction enzyme site, which is at position −170 from the *ADR2* coding sequence (see Fig. 3), or displacement of these sequences to a site far upstream from the gene has been shown to allow constitutive expression of the *ADR2* gene product (D. Beier and E. T. Young, Nature [London], in press). This same region is moved 5.6 kb further away from the gene by Ty1 insertion. Displacement of the 5′-flanking sequence by only 300 bp, which is similar to the size of a solo delta mutant, has been shown to allow the adjacent gene to be fully repressed on glucose but to prevent efficient derepression. This result is similar to that seen when only a single delta sequence remains upstream from *ADR2*, where we detected little or no *ADR2* transcript by S1 mapping even under conditions in which the wild-type gene is expressed.

The displacement of a repressor sequence cannot fully explain the role of the Ty1 element in constitutive *ADR2* expression. Several lines of evidence suggest that the Ty1 elements themselves have an active role in regulation of the adjacent gene. For example, Errede et al. (18) observed that overproduction of iso-2-cytochrome *c* in the mutant *CYC7-H2*, which contains a Ty1 element 5′ to the structural gene, was affected by the mating competence of the cell. A 20-fold overproduction was seen in cells with the ability to mate, but only a 1- to 4-fold increase was seen in nonmating cells such as a/α diploid cells. The effect of mating competence is a characteristic of positively affected Ty1 mutations, including those at the *ADR2* locus (13, 44). Transcription of the Ty element itself is regulated by mating condition. Elder et al. (16) have observed that Ty1 transcripts are reduced in a/α diploid cells. All of the Ty1 insertions which we have observed to produce constitutive expression of *ADR2* are inserted in the same orientation with respect to *ADR2*, and the direction of Ty1 transcription is away from *ADR2* (Elder, personal communication), resulting in two divergent transcripts which may be coordinately regulated by the mating condition of the cells. Several examples of coordinately regulated divergent transcripts have been described in eucaryotes, for example, the mating loci (31) and galactose utilization genes of yeast (38) and developmentally regulated chorion genes of silk moth (25).

The hypothesis we favor to explain Ty1-mediated effects on *ADR2* expression is that Ty1 insertion displaces 5′-flanking sequences which are responsible for repressing *ADR2* during growth on glucose-containing medium. When interrupted by the 5.6-kb Ty1 element, these regulatory sequences are no longer effective. Displacement by a 330-bp solo delta sequence allows repression to remain effective. Efficient derepression does not occur in either case because sequences needed for positive activation are displaced or interrupted by the Ty1 insertion or delta sequence. In place of these regulatory signals, the *ADR2* gene now comes under the same control as the Ty1 element.

ACKNOWLEDGMENTS

We thank M. Ciriacy for providing the mutant yeast strains used and M. Ciriacy and C. Denis for providing us with plasmid A2M, which contains the wild-type *ADR2* sequences. We are grateful to C. R. Astell and S. Gillam for helpful suggestions.

V.M.W. was the recipient of a Public Health Service postdoctoral fellowship from the National Institutes of Health

and was supported during the writing of this manuscript by the ARCO Plant Cell Research Institute. D.W.R. is supported by fellowship DRG-455 of the Damon Runyon-Walter Winchell Cancer Fund. This research was supported by Public Health Service National Research Service Award 07270 from the National Institutes of Health at the University of Washington and by the Medical Research Council of Canada at the University of British Columbia. M.S. is a Career Investigator of the Medical Research Council of Canada.

LITERATURE CITED

1. **Baltimore, D.** 1981. Gene conversion: some implications for immunoglobulin genes. Cell **24**:592–594.
2. **Bennetzen, J., and B. D. Hall.** 1982. The primary structure of the Saccharomyces cerevisiae gene for alcohol dehydrogenase I. J. Biol. Chem. **257**:3018–3025.
3. **Berk, A. J., and P. A. Sharp.** 1978. Spliced early mRNAs of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **75**:1274–1278.
4. **Bingham, P. M., and B. H. Judd.** 1981. A copy of the *copia* transposable element is very tightly linked to the w^a allele at the *white* locus of D. melanogaster. Cell **25**:705–711.
5. **Bingham, P. M., R. Levis, and G. M. Rubin.** 1981. Cloning of DNA sequences from the *white* locus of D. melanogaster by a novel and general method. Cell **25**:693–704.
6. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
7. **Calos, M. P., and J. H. Miller.** 1980. Transposable elements. Cell **20**:579–595.
8. **Cameron, J. R., E. Y. Loh, and R. Davis.** 1979. Evidence for transposition of dispersed repetitive DNA families in yeast. Cell **16**:739–751.
9. **Chaleff, D. T., and G. R. Fink.** 1980. Genetic events associated with an insertion mutation in yeast. Cell **21**:227–237.
10. **Ciriacy, M.** 1975. Genetics of alcohol dehydrogenase in Saccharomyces cerevisiae. I. Isolation and genetic analysis of adh mutants. Mutat. Res. **29**:315–326.
11. **Ciriacy, M.** 1976. Cis-dominant regulatory mutations affecting the formation of glucose-repressible alcohol dehydrogenase (*ADH*II) in Saccharomyces cerevisiae. Mol. Gen. Genet. **145**:327–333.
12. **Ciriacy, M.** 1979. Isolation and characterization of further cis- and trans-acting regulatory elements involved in the synthesis of glucose-repressible alcohol dehydrogenase (ADHII). Mol. Gen. Genet. **176**:427–431.
13. **Ciriacy, M., and V. M. Williamson.** 1981. Analysis of mutations affecting Ty-mediated gene expression in *Saccharomyces cerevisiae*. Mol. Gen. Genet. **182**:159–163.
14. **Denis, C. L., M. Ciriacy, and E. T. Young.** 1981. A positive regulatory gene is required for accumulation of functional messenger RNA for the glucose-repressible alcohol dehydrogenase from Saccharomyces cerevisiae. J. Mol. Biol. **148**:355–368.
15. **Eibel, H., J. Gafner, A. Stotz, and P. Philippsen.** 1980. Characterization of the yeast mobile element Ty1. Cold Spring Harbor Symp. Quant. Biol. **45**:609–617.
16. **Elder, R. T., T. P. St. John, D. T. Stinchcomb, and R. W. Davis.** 1980. Studies on the transposable element Ty1 of yeast. I. RNA homologous to Ty1. Cold Spring Harbor Symp. Quant. Biol. **45**:581–584.
17. **Errede, B., T. S. Cardello, F. Sherman, E. Dubois, J. Deschamps, and J.-M. Wiame.** 1980. Mating signals control expression of mutations resulting from insertion of a transposable repetitive element adjacent to diverse yeast genes. Cell **22**:427–436.
18. **Errede, B., T. S. Cardello, G. Wever, and F. Sherman.** 1980. Studies on transposable elements in yeast. I. ROAM mutations causing increased expression of yeast genes: their activation by signals directed toward conjugation functions and their formation by insertion of Ty1 repetitive elements. Cold Spring Harbor Symp. Quant. Biol. **45**:961–981.
19. **Farabaugh, P. J., and G. R. Fink.** 1980. Insertion of the eukaryotic transposable element Ty1 creates a 5-base pair duplication. Nature (London) **286**:352–356.
20. **Fink, G., P. Farabaugh, G. Roeder, and D. Chaleff.** 1980. Transposable elements (Ty) in yeast. Cold Spring Harbor Symp. Quant. Biol. **45**:575–580.
21. **Gafner, J., and P. Philippsen.** 1980. The yeast transposon Ty1 generates duplications of target DNA on insertion. Nature (London) **286**:414–418.
22. **Gehring, W. J., and R. Paro.** 1980. Isolation of a hybrid plasmid with homologous sequences to a transposing element of Drosophila melanogaster. Cell **19**:897–904.
23. **Gillam, S., and M. Smith.** 1980. Use of *E. coli* polynucleotide phosphorylase for the synthesis of oligodeoxyubonucleotide of defined sequence. Methods Enzymol. **65**:687–701.
24. **Hayward, W. S., B. G. Neel, and S. M. Astrin.** 1981. Activation of a cellular *onc* gene by promoter insertion in ALV-induced lymphoid leukosis. Nature (London) **290**:475–480.
25. **Jones, C. W., and F. C. Kafatos.** 1980. Structure, organization and evolution of developmentally regulated chorion genes in a silkmoth. Cell **22**:855–867.
26. **Ju, G., and A. M. Skalka.** 1980. Nucleotide sequence analysis of the long terminal repeat (LTR) of avian retroviruses: structural similarities with transposable elements. Cell **22**:379–386.
27. **Kingsman, A. J., R. L. Gimlick, L. Clark, A. C. Chinault, and J. Carbon.** 1981. Sequence variation in dispersed repetitive sequences in Saccharomyces cerevisiae. J. Mol. Biol. **145**:619–632.
28. **Levis, R., P. Dunsmuir, and G. M. Rubin.** 1980. Terminal repeats of the Drosophila transposable element *copia*: nucleotide sequence and genomic organization. Cell **21**:581–588.
29. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
30. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
31. **Nasmyth, K. A., K. Tatchell, B. D. Hall, C. Astell, and M. Smith.** 1980. Physical analysis of mating-type loci in Saccharomyces cerevisiae. Cold Spring Harbor Symp. Quant. Biol. **45**:961–981.
32. **Payne, G. S., J. M. Bishop, and H. E. Varmus.** 1982. Multiple arrangements of viral DNA and an activated last oncogene in bursal lymphomas. Nature (London) **295**:209–214.
33. **Roeder, G. S., P. J. Farabaugh, D. T. Chaleff, and G. R. Fink.** 1980. The origins of gene instability in yeast. Science **209**:1375–1380.
34. **Rubin, G. M., W. J. Brorein, Jr., P. Dunsmuir, A. J. Flavell, R. Levis, E. Strobel, J. J. Toole, and E. Young.** 1980. *copia*-like transposable elements in the *Drosophila* genome. Cold Spring Harbor Symp. Quant. Biol. **45**:619–628.
35. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
36. **Schultz, L. D.** 1978. Transcriptional role of yeast deoxyribonucleic acid dependent ribonucleic acid polymerase III. Biochemistry **17**:750–758.
37. **Smith, M., D. W. Leung, S. Gillam, C. Astell, D. L. Montgomery, and B. D. Hall.** 1979. Sequence of the gene for iso-1-cytochrome C in Saccharomyces cerevisiae. Cell **16**:753–761.
38. **St. John, T. P., and R. W. Davis.** 1981. The organization and transcription of the galactose gene cluster of *Saccharomyces*. J. Mol. Biol. **152**:285–316.
39. **Struhl, K., D. T. Stinchcomb, S. Scherer, and R. W. Davis.** 1979. High-frequency transformation of yeast: autono-

mous replication of hybrid DNA molecules. Proc. Natl. Acad. Sci. U.S.A. **76**:1035–1039.

40. **Temin, H. M.** 1980. Origin of retroviruses from cellular moveable genetic elements. Cell **21**:599–600.
41. **Williamson, V. M., J. Bennetzen, E. T. Young, K. Nasmyth, and B. D. Hall.** 1980. Isolation of the structural gene for alcohol dehydrogenase by genetic complementation in yeast. Nature (London) **283**:214–216.
42. **Williamson, V. M., E. T. Young, and M. Ciriacy.** 1981. Transposable elements associated with constitutive expression of yeast alcohol dehydrogenase II. Cell **23**:605–614.
43. **Young, T., V. Williamson, A. Taguchi, M. Smith, A. Sledziewski, D. Russell, J. Osterman, C. Denis, D. Cox, and D. Beier.** 1982. The alcohol dehydrogenase genes of the yeast, *Saccharomyces cerevisiae*: isolation, structure, and regulation, p. 335–361. *In* A. Hollaender, R. D. DeMoss, S. Kaplan, J. Konisky, D. Savage, and R. S. Wolfe (ed.), Genetic engineering of microorganisms or chemicals. Plenum Publishing Corp., New York.