MOLECULAR AND CELLULAR BIOLOGY, Dec. 1986, p. 4281–4294
0270-7306/86/114281-14$02.00/0

Vol. 6, No. 12

Structure of the *Saccharomyces cerevisiae HO* Gene and Analysis of Its Upstream Regulatory Region

DAVID W. RUSSELL,[1]† ROB JENSEN,[2]‡ MARK J. ZOLLER,[1]§ JANICE BURKE,[1] BEVERLY ERREDE,[3] MICHAEL SMITH,[1] AND IRA HERSKOWITZ[2]*

Department of Biochemistry, University of British Columbia, Vancouver, British Columbia V6T 1W5, Canada[1]; Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143[2]; and Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27514[3]

Received 19 November 1985/Accepted 1 August 1986

The *HO* gene product of *Saccharomyces cerevisiae* is a site-specific endonuclease that initiates mating type interconversion. We have determined the nucleotide sequence of a 3,129-base-pair (bp) segment containing *HO*. The segment contains a single long open reading frame encoding a polypeptide of 586 amino acids, which has unusual (unbiased) codon usage and is preceded by 762 bp of upstream region. The predicted *HO* protein is basic (16% lysine and arginine) and is calculated to have a secondary structure that is 30% helical. The corresponding transcript is initiated approximately 50 nucleotides prior to the presumed initiation codon. Insertion of an *Escherichia coli lacZ* gene fragment into the putative *HO* coding segment inactivated *HO* and formed a hybrid *HO-lacZ* gene whose beta-galactosidase activity was regulated by the mating type locus in the same manner as *HO* (repressed by a1-α2). Upstream regions of 1,360 and 762 bp conferred strong repression; 436 bp led to partial constitutivity and 301 bp to full constitutivity. Thus, DNA sequences that confer repression of *HO* by a1-α2 are at least 250 nucleotides upstream of the transcription start point and are within 436 nucleotides of the *HO* initiation codon. The progressive loss of repression suggests that both the −762 to −436 and the −436 to −301 intervals contain sites for regulation by a1-α2. The *HO* gene contains two distinct regions that promote autonomous replication of plasmids in *S. cerevisiae*. These regions contain sequences that are homologous to the two conserved sequences that are associated with ARS activity.

Saccharomyces cerevisiae strains carrying the *HO* (homothallism) gene exhibit a homothallic life cycle—individual haploid spores grow into clones that contain diploid a/α cells. Strains carrying the *ho* allele, which is recessive to *HO*, exhibit a heterothallic life cycle—diploid a/α cells are formed only by mating between haploid cells from separate clones. Thus, the *HO* gene is responsibe for production of diploid cells from haploid cells, that is, for diploidization (66). The action of *HO* leads to high-efficiency switching between a and α cell types, which occurs by transposition of a block of information from a genomic storage position to the mating type locus, where this information is expressed and determines cell type (reviewed in reference 25). Strains that carry a functional *HO* gene exhibit a double-strand break at the mating type locus that is essential for initiating transposition (61). Recently, it has been shown that the *HO* gene codes for this double-strand endonuclease: *Escherichia coli* strains that express *HO* produce this activity (33).

Mating type interconversion is under several different types of control, many of which occur via regulation of the *HO* gene itself. First, mating type interconversion occurs in a and α cells but not in a/α cells, in which the *HO* transcript is absent (29). This inhibition requires both the α2 product of *MAT*α and the a1 product of *MAT*a. In addition to *HO*, there are several other yeast genes whose expression is inhibited in a/α diploid cells by the a1 and α2 products (Fig. 1). This set of genes includes *MAT*α*1* (32, 46), *STE5* (36), and the transposable element Ty1 (14). In addition, there are several cases in which Ty1 has inserted adjacent to a locus (for example, next to *CYC7*, *ADR2*, or *HIS4*) and placed that gene under negative regulation by a1-α2 (15, 16, 49, 64). It has been suggested that Ty1 elements might be responsible for regulation of some wild-type *S. cerevisiae* genes by a1-α2 (15).

Additional controls of *HO* expression have been invoked to explain the characteristic pattern of cell type switching observed in lineages of cells undergoing mating type interconversion (60). This specific cell lineage is governed in large part by regulation of the *HO* gene (28, 42; R. Jensen, Ph.D. thesis, University of Oregon, Eugene, 1983). In particular, *HO* is expressed only at a specific phase of the cell cycle and is not expressed in daughter cells (28, 42). Five genes (*SWI1-SWI5*) are required for expression of *HO* (58) and may play a role in such regulation.

To provide information on the *HO* gene product and on different ways in which it is regulated, we have sequenced a DNA segment that contains *HO*, determined the start point for the *HO* transcript, and identified regions necessary for regulation by the mating type locus.

MATERIALS AND METHODS

Strains and relevant genetic markers. 3B54 is a UV-induced *ste14* mutant derived from X10-1b (L. C. Blair, Ph.D. thesis, University of Oregon, Eugene, 1979). DHR125 (*ho/ho MAT*a*/MAT*α *ura3-52/ura3-52*) was obtained by transforming HR125-5d with a plasmid that contains the *HO* gene (pHO-c12), isolating diploid colonies that formed as a result of mating type interconversion, and then screening for colonies that lost the *HO* plasmid on nonselective medium. Strain 1369 (*ho MAT*α *ura3-52*) is a haploid segregant iso-

* Corresponding author.
† Present address: Department of Molecular Genetics, University of Texas Health Science Center, Dallas, TX 75235.
‡ Present address: Department of Biology, University of California, La Jolla, CA 92093.
§ Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724.

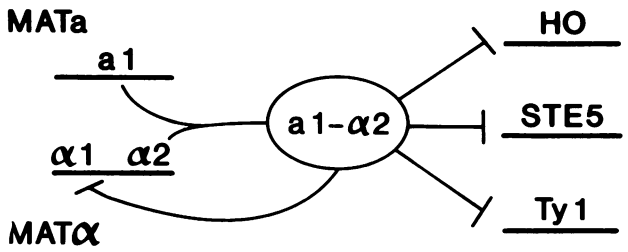


FIG. 1. Repression of various genes in a/α cells by a1-α2. The two *MAT* alleles of an a/α cell are drawn to the left. They code for the a1-α2 regulatory activity, which is shown inhibiting RNA synthesis from *MATα1*, *HO*, *STE5*, and the yeast transposable element Ty1. Genes such as *HO*, *STE5*, and Ty1, which are transcribed at high level in **a** and α cells but not in a/α cells, are termed haploid-specific genes.

lated from this diploid. Strains HR125-5d, 1369, and DHR125 are an isogenic series of *MAT***a**, *MAT*α, and *MAT***a**/*MAT*α strains. YCp50 (obtained from C. Mann) is a derivative of YIp5 (62) that contains the yeast *URA3* gene and a centromere (*CEN4*). Plasmid YEp13 (7) was obtained from K. Nasmyth. Other information on strains is given in the text and in Table 1.

Media and genetic methods. Standard yeast genetic techniques and media were as described (26). The activity of the *HO* gene in HR145-4c and derivatives was determined by an α-factor assay. *HO MAT*α *HML*α *HMR*α cells produce low levels of α-factor, whereas strains that are otherwise isogenic and that carry *ho* or mutations in *HO* produce normal levels of α-factor (58; Blair, Ph.D. thesis). α-Factor was assayed by the halo assay of Fink and Styles (19), as described by Sprague and Herskowitz (57), in which strains to be tested are replica-plated to a lawn of **a** cells. α-Factor secreted by the α cells inhibits growth of the **a** cells and produces a zone of inhibition, the halo.

For mapping the position of the *HO-lacZ* mutation introduced into *HO MAT*α strain HR145-4c by transformation, an Ho^- transformant (HR180) was mated to *ho MAT***a** strain HR125-5d to form diploid HR181, and meiotic products were analyzed. A 4 Ho^-:0 Ho^+ segregation was observed in all tetrads. In addition, beta-galactosidase activity was shown to segregate 2 $LacZ^+$:2 $LacZ^-$ (see below). Beta-galactosidase activities of two segregants from this diploid, HR181-1c (*HO-lacZ MAT***a**) and HR181-1d (*HO-lacZ MAT*α), are shown in Table 2.

DNA sequence determinations. Isolation of the DNA containing the *HO* gene and its regulatory region has been described (29). The materials and procedures used for sequence determination of the 3,129-base-pair (bp) *Bam*HI-*Hin*dIII segment containing *HO* were those described by Russell et al. (52). The sequencing strategy and additional details are given in Figure 2.

Beta-galactosidase assays. The beta-galactosidase activity of the *HO-lacZ* fusion in yeast cells was determined as described (24): pellets from 1.5 ml of log-phase yeast cells were suspended in 0.15 ml of Z buffer, treated with 0.05 ml of chloroform and 0.02 ml of 0.1% sodium dodecyl sulfate (SDS), vortexed for 30 s, and assayed (41). Beta-galactosidase activity of yeast cells grown on plates was assayed as described (50), except that X-Gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside) was used at 0.2 M. Beta-galactosidase activity of *E. coli* cells that carry *HO-lacZ* plasmids was detected on plates as described (41).

Construction of the *HO-lacZ* fusion. pMC1871, a plasmid containing the carboxy-terminal coding region of the beta-galactosidase gene on a 3.0-kilobase-pair (kbp) *Pst*I fragment, was obtained from M. Casadaban (10). This plasmid was digested with *Pst*I, and the *lacZ* fragment was inserted into the *Pst*I site within the putative *HO* reading frame present in plasmid pHO-D6 (see Fig. 4). Because pHO-D6 has three *Pst*I sites, this plasmid was incompletely digested with *Pst*I, added to the complete *Pst*I digest of pMC1871, and ligated with T4 DNA ligase. A plasmid containing the *lacZ* fragment inserted into the correct *Pst*I site of pHO-D6 (at position 245, see Fig. 3) and in the correct orientation was identified in *E. coli* as an Amp^r Lac^+ transformant. The structure of one plasmid, pHO-lac-6, was confirmed by restriction enzyme analysis. The procedures for plasmid preparations from *E. coli*, transformation of yeast and *E. coli*, and isolation of yeast DNA have been described (29, 37).

Replacement of the chromosomal *HO* gene with *HO-lacZ*. Strain HR180 (*HO-lacZ MAT*α *leu2-3 leu2-112*) was constructed by replacing the *HO* gene of HR145-4c with *HO-lacZ* by the one-step gene replacement procedure of Rothstein (51). HR145-4c was cotransformed with a 5.5-kbp *Hin*dIII fragment that contained the *lacZ* insert in *HO* and circular plasmid DNA that contained the yeast *LEU2* gene (YEp13) with a 10- to 100-fold excess of linear molecules. Leu^+ transformants were isolated and screened for Ho^- as described above. Approximately 3% of the Leu^+ transformants were Ho^-. In one candidate, HR180, the *HO-lacZ* fusion was shown to segregate as an allele of *HO*. In addition, Southern analysis (see below) indicated that the 2.5-kbp *Hin*dIII fragment of *HO* in the recipient was re-

TABLE 1. Strains of *S. cerevisiae* used

Strain	Relevant genotype	Reference or comment
1369	*ho MAT*α *ura3-52*	Segregant from DHR125
AB320	*HO*/*HO MAT***a**/*MAT*α	45
DHR125	*ho*/*ho MAT***a**/*MAT*α *ura3-52*/*ura3-52*	Diploidization of HR125-5d
HR125-5d	*ho MAT***a** *ura3-52*	This work
HR145-4c	*HO MAT*α *HML*α *HMR*α *leu2-3 leu2-112*	This work
HR180	*HO-lacZ MAT*α	Transformant of HR145-4c
HR181	*HO-lacZ*/*ho MAT***a**/*MAT*α	HR180 × HR125-5d
HR181-1c	*HO-lacZ MAT***a**	Segregant from HR181
HR181-1d	*HO-lacZ MAT*α	Segregant from HR181
X10-1b	*HO*/*HO MAT***a**/*MAT*α	26
3B54	*ste14* derivative of X10-1b	29

TABLE 2. Regulation of β-galactosidase activity produced by an *HO-lacZ* gene at the *HO* locus[a]

Strain	Cell type	β-Galactosidase activity (U)[b]
HR180 (*HO-lacZ MAT*α)	α	8.2
HR181 (*HO-lacZ*/*ho MAT***a**/*MAT*α)	a/α	<0.01
HR181-1c (*HO-lacZ MAT***a**)	**a**	8.0
HR181-1d (*HO-lacZ MAT*α)	α	8.4

[a] The *HO-lacZ* hybrid gene is present at the *HO* locus as a replacement of the wild-type *HO* allele (see text).

[b] Specific activity of β-galactosidase is as described in reference 41. Activities are averages of at least three independent assays.

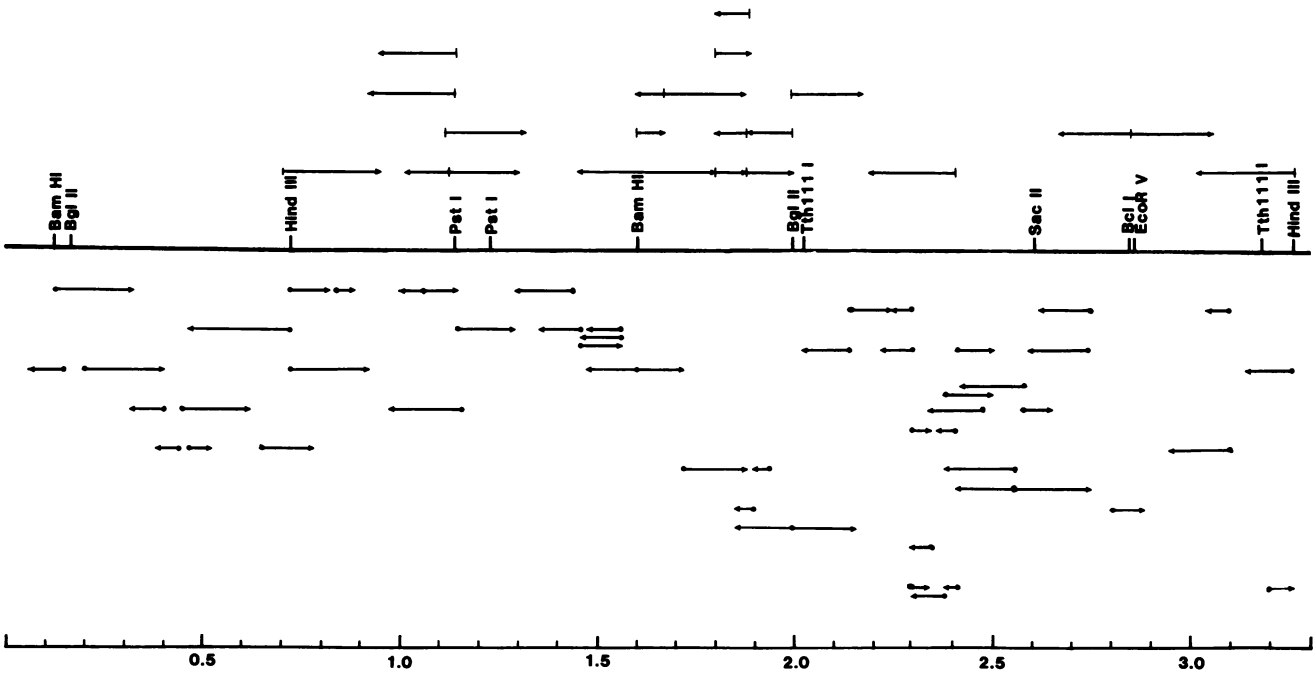


FIG. 2. Map of restriction endonuclease cleavage sites and DNA sequencing strategy for the *HO* gene. The scale is given in kilobase pairs. Arrows above the map indicate sequences determined by the Sanger et al. method (39, 53) after subcloning in a bacteriophage M13 vector. Arrows below the map indicate sequences determined by the method of Maxam and Gilbert (38). The Sanger sequencing was carried out on subcloned *Hin*dIII, *Pst*I, and *Rsa*I fragments with synthetic d(T_8GT) priming at nucleotide 2195. Maxam and Gilbert sequence determination was carried out on *Alu*I, *Bam*HI, *Bgl*II, *Dde*I, *Hin*fI, *Fnu*4HI, *Msp*I, *Pst*I, *Rsa*I, *Sau*96I, and *Taq*I fragments. Sites for enzymes recognizing 6 bp are shown.

placed by the 5.5-kbp *Hin*dIII fragment of *HO-lacZ* in HR180 (see Fig. 5).

Isolation of pHO-12 and pHO-15. The *HO* gene was isolated by complementation on a plasmid designated YEpHO (29). Based on the results presented here, YEpHO contains 436 bp of 5'-flanking DNA. To isolate plasmids carrying additional upstream *HO* sequences, a yeast genomic library in vector YEp13 (45) was screened by colony hybridization (21). The 875-bp *Hin*dIII-*Bam*HI fragment (Fig. 3), which contains 171 bp of *HO* 5'-flanking DNA, was used as a probe following nick translation in the presence of deoxyribonucleoside triphosphates. Two plasmids, YEpHO-12 and YEpHO-15, were isolated in this manner.

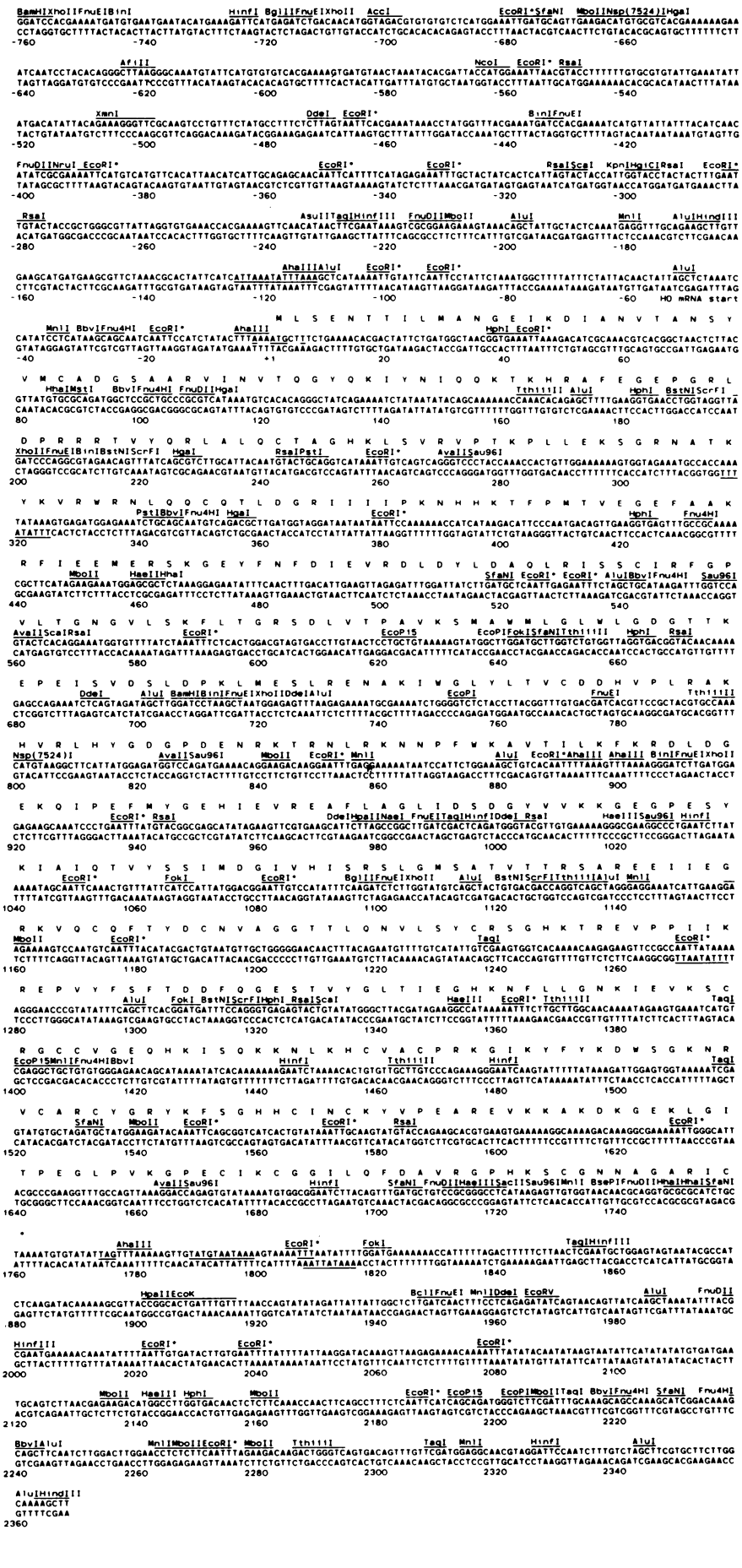
Construction of YCp50 derivatives carrying *HO* and *HO-lacZ*. Plasmids YEpHO-6 (the original YEpHO plasmid of Jensen et al. [29]), YEpHO-12, and YEpHO-15 all carry the *HO* gene on a DNA fragment inserted into the *Bam*HI site of the vector YEp13. The *HO*-containing inserts are all in the same orientation in YEp13 and contain different amounts of yeast sequences upstream of the *HO* coding region (see Fig. 7). The *HO* gene and its upstream regions were transferred into the centromere-containing vector YCp50 on the *Sal*I-*Eco*RI fragment shown in Fig. 7. Each donor plasmid was digested with *Eco*RI and *Sal*I and mixed with an *Eco*RI-*Sal*I digest of plasmid YCp50, ligated with DNA ligase, and transformed into *E. coli*. Plasmids pHO-c6, pHO-c12, and pHO-c15 carry the *Eco*RI-*Sal*I *HO* fragments from YEpHO-6, YEpHO-12, and YEpHO-15, respectively, in YCp50.

Plasmid pHO-c20 is a deletion derivative of pHO-c12 (Fig. 7) and was constructed as follows. pHO-c12 was digested with *Kpn*I, and the DNA ends were rendered flush with T4 DNA polymerase as described (37). YCp50 was digested with *Bam*HI and the DNA ends were filled in. Both of these plasmids were digested with *Eco*RI, mixed, and ligated with T4 DNA ligase. Blunt-end ligation of a flush *Kpn*I site and a filled-in *Bam*HI site regenerated a *Bam*HI site.

The *HO-lacZ* insert of pHO-lac-6 (see above) was used to replace the *HO* gene of plasmids pHO-c6, -12, -15, and -20 to produce the corresponding plasmids pHO-lac-c6, pHO-lac-c12, pHO-lac-c15, and pHO-lac-c20. This substitution was accomplished by replacing the 2.5-kbp *Hin*dIII fragment that carries the *HO* gene with the 5.5-kbp *Hin*dIII fragment that carries the *HO-lacZ* gene. pHO-lac-c1 and pHO-lac-c2 were constructed by inserting the 5.5-kbp *Hin*dIII fragment into the *Hin*dIII site of YCp50. pHO-lac-c1 and pHO-lac-c2 have the *HO-lacZ* fragment in opposite orientations (see Fig. 7).

S1 nuclease mapping of the 5' end of the *HO* transcript. M13 HO-2 and M13 HO-11, which contain the 411-bp *Hin*dIII-*Pst*I fragment (Fig. 2) of *HO* inserted in the M13 vectors mp9 and mp8, respectively, were constructed by standard methods (37). With single-stranded phage DNA isolated from M13 HO-2 and M13 HO-11 as a template, DNA polymerase was used to extend a 17-nucleotide M13 primer (kindly supplied by Stan Fields) across the inserted *HO* sequences in the presence of radioactively labeled dCTP. The resulting double-stranded DNA was cleaved with *Hin*dIII (for M13 HO-2) or with *Pst*I (for M13 HO-11), and the radioactively labeled 428-nucleotide fragment (411 bases of *HO* and 17 bases of primer) was isolated by electrophoresis through a 7 M urea–4% acrylamide gel. The methods used for hybridization of these DNA fragments to yeast RNA, S1 nuclease digestion, electrophoresis, and autoradiography have been described (46). *HO* RNA for this analysis was isolated from a culture of strain 3B54, which is a mixed population of *MAT*α *HO ste14* and *MAT***a** *HO ste14* cells that produces *HO* RNA (29).

Hybridization analysis. DNA was digested and transferred to nitrocellulose as described by Southern (56) and hybrid-



ized as described previously (37). The *HO*-containing probe (YIp5-BH2 [29]) and the Ty1-containing probe (pBR322-S13 [14]) were radioactively labeled by nick translation as described previously (29).

RESULTS

Sequence of a segment containing the *HO* gene and adjacent region. A 3.1-kilobase (kb) *Bam*HI-*Hin*dIII DNA fragment has been shown previously (29) to contain the *HO* gene by two criteria: it provides *HO* gene activity to *ho* strains, and it directs integration by homologous recombination at *HO*. The map of key 6-bp restriction endonuclease sites is shown in Fig. 2, together with the sequencing strategy. The endmost *Bam*HI site of the *HO* segment originally cloned (29) was formed by joining the *Sau*3A site at position −436 to the *Bam*HI site of the vector. Overlapping sequence determinations were carried out for all sequencing origins. Most of the sequence was determined at least twice, and both strands were sequenced in regions of uncertainty or ambiguity. A majority of the DNA was sequenced by both chemical and enzymatic methods.

The complete 3,129-bp sequence is shown in Fig. 3. Analysis of the distribution of translation termination codons in the six reading frames (not shown) revealed an extended open reading frame of 586 amino acids in one orientation (frame 3) and open reading frames of 150 and 200 amino acids in the other orientation (frame 6). Earlier work (R. Jensen, unpublished observations) has shown that the *HO* gene is transcribed with the same polarity as reading frames 1, 2, and 3. Analyses described below show that the open reading frame of 586 amino acids is the *HO* amino acid sequence.

Identification of the *HO* coding sequence. To determine whether the longest open reading frame was that of *HO*, we constructed and analyzed an insertion mutation at the *Pst*I site at position 245. As shown in Fig. 4, we inserted a *Pst*I fragment containing a truncated *E. coli lacZ* gene (10) lacking the first seven amino acids into this *Pst*I site. Production of beta-galactosidase requires that the translation start codon and other signals missing from the *lacZ* gene be supplied by the upstream *S. cerevisiae* DNA. This insertion should result in a polypeptide chain in which the *lacZ* segment has been inserted into the long open reading frame. The hybrid polypeptide should contain 82 amino acids from the putative *HO* protein and a segment of 10 amino acids from the polylinker, followed by amino acids 8–1023 of beta-galactosidase. If the long open reading frame is that of *HO*, then we have three expectations. (i) The insertion should inactivate *HO*. (ii) The insertion should provide proper transcription and translation signals for synthesis of a hybrid protein having beta-galactosidase activity. (iii) The beta-galactosidase activity of the hybrid protein should be regulated in the manner of the *HO* gene; in particular, it should be under control by the mating type locus.

To test the first and second expectations, the *lacZ* insertion mutation was used to replace an *HO* allele by the method of Rothstein (51). A 5.5-kbp *Hin*dIII fragment containing the *lacZ* insert (Fig. 4) was cotransformed with plasmid YEp13, which carries the yeast *LEU2* gene, into

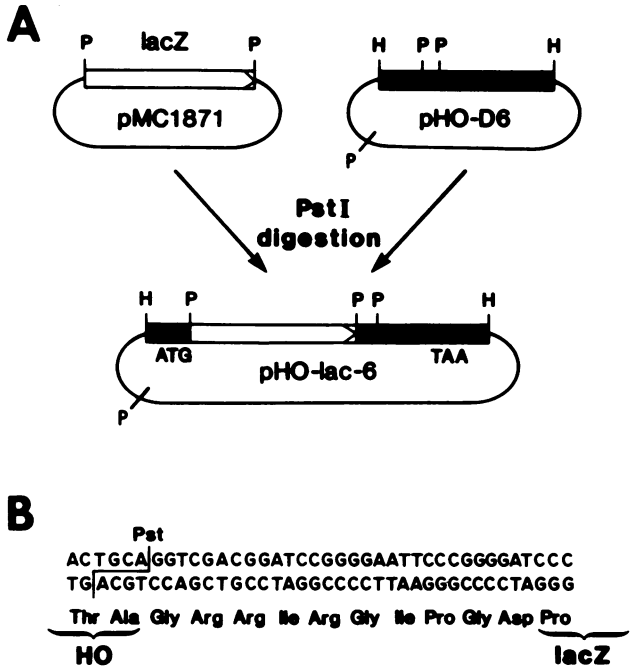


FIG. 4. Construction of an *HO-lacZ* fusion. (A) A 3.0-kbp *Pst*I fragment that carries the *lacZ* gene was inserted into the *Pst*I site at position 245 in the *HO* coding region after partial digestion of plasmid pHO-D6, as described in Materials and Methods. The positions of presumed initator and terminator codons of *HO* are shown. (B) Junction between the *HO* and *lacZ* coding sequences in the *HO-lacZ* hybrid is shown. Residues Thr-81 and Ala-82 of *HO* and Pro-8 of β-galactosidase are indicated. The predicted hybrid protein contains the amino-terminal 82 amino acids of *HO* protein attached to β-galactosidase lacking its amino-terminal seven residues. The amino acids between *HO* and *lacZ* are from the synthetic oligonucleotide sequence adjacent to *lacZ* (10).

strain HR145-4c (*HO MATα HMLα HMRα leu2*), and Leu$^+$ transformants were selected. The Ho activity of these colonies was determined by screening colonies for the production of α-factor as described in Materials and Methods; approximately 3% of the Leu$^+$ transformants were Ho$^-$. As expected, all five Ho$^-$ transformants tested formed blue (LacZ$^+$) colonies on indicator plates. One such transformant (HR180) was chosen for further analysis. By standard genetic analysis (see Materials and Methods), the *lacZ* gene in this strain was shown to be located at the *HO* locus. In addition, Southern analysis showed that the normal 2.5-kbp genomic *Hin*dIII fragment containing the wild-type *HO* gene was replaced in HR180 by the expected 5.5-kbp *Hin*dIII fragment of the *lacZ* insertion mutation (Fig. 5). These results provide evidence in support of the first two predictions.

As noted above, we anticipated that the beta-galactosidase activity of the *HO-lacZ* gene fusion would be regulated by the mating type locus. Indeed, we observed that the beta-galactosidase activity of *MATa/MATα* diploid HR181 (*HO-lacZ/ho*) was unmeasurable, in contrast to HR180, which exhibited 8.2 U (Table 2). Both **a** and α segregants from HR181 exhibited high levels of beta-galactosidase activity

FIG. 3. Sequence of the *HO* locus. The putative *HO* protein-coding region is from nucleotide 1 to nucleotide 1758. The position of the mRNA start is near nucleotide −46, as indicated. The following sequences are interlined: the TATA sequence of an A/T-rich region at −126; the AATAAA at 1793; and TAG, TATGT, and TTT at 1773, 1788, and 1806. The putative ARS sequences at 316 and 1807 are underlined (see Fig. 14). Positions of the restriction endonuclease recognition sites (48) were determined by computer search (12). (The sequence information shown in this figure is available on diskette.)

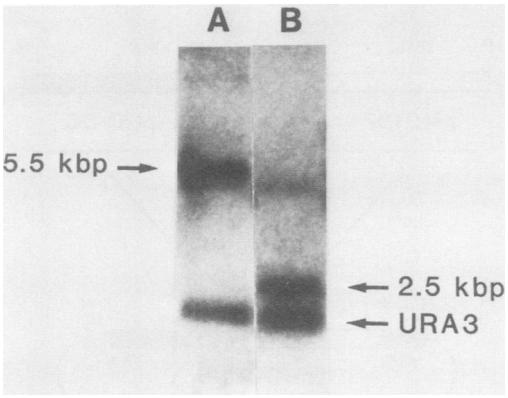


FIG. 5. Characterization of the *HO-lacZ* gene located in the chromosome. Autoradiograph from the Southern analysis of DNA isolated from HR145-4c (*HO MAT*α, lane B) and HR180 (*HO-lacZ MAT*α, lane A). DNA was digested with *Hin*dIII, separated by size, and transferred to nitrocellulose as described in Materials and Methods. The filter was hybridized to a radioactively labeled probe that carries *HO* and *URA3* sequences (YIp5-BH2, containing *HO* sequences between −171 and 705). The 2.5-kbp fragment (lane B) and 5.5-kbp fragment (lane A) are specific to *HO*. The 2.2-kbp fragment is specific to *URA3*. DNA lengths were determined from DNA standards of known size.

(Table 2). The observation that beta-galactosidase activity was inhibited at least 800-fold in **a**/α cells supports the third prediction.

Location of the *HO* RNA start site. The 5′ end of the mature *HO* mRNA was identified by the S1 nuclease mapping procedure of Berk and Sharp (4). Because RNA start sites in yeast are frequently found close to the start of translation (3), it was expected that the *HO* transcript would initiate within the 411-bp *Hin*dIII-*Pst*I fragment (Fig. 2) that contains the putative translation start of *HO* (Fig. 3). Therefore, a single-stranded, radioactively labeled probe complementary to the *HO* RNA (HP1) was isolated from the *Hin*dIII-*Pst*I fragment as described in Materials and Methods. As shown in Fig. 6, when the HP1 probe was hybridized to RNA isolated from cells expressing the *HO* gene (strain 3B54; lanes A, B, and F), a single protected fragment of approximately 290 bases was observed. No protected fragment was found when the RNA used in these experiments was isolated from *MAT***a**/*MAT*α diploid cells (strain X10-1b; lane G), in which *HO* is not expressed, or when the mRNA-identical strand of the *Hin*dIII-*Pst*I fragment was used (lanes D and E). Because the *Pst*I site of HP1 represents the 3′ end of the DNA, the 5′ end of the *HO* RNA is located 290 (± 3) nucleotides from the *Pst*I site, as shown in Fig. 3. Hence, the *HO* RNA start is approximately 45 to 50 bases upstream of the predicted *HO* protein start site.

Use of the *HO-lacZ* fusion to identify sequences required for *HO* regulation. To define regions upstream of the *HO* gene necessary for regulation by the mating type locus, we constructed a set of plasmids that have the *HO-lacZ* fusion and different amounts of the 5′-flanking sequence of *HO* (Fig. 7). The inserts of all of these plasmids (except pHO-lac-c1 and pHO-lac-c2) are joined at their 3′ and 5′ ends to the same sites of the vector (*Eco*RI and *Bam*HI sites, respectively). pHO-lac-c1 carries the *HO-lacZ* fragment inserted into the *Hin*dIII site of the vector; pHO-lac-c2 carries the same fragment in the opposite orientation. The vector used in this analysis was YCp50, which carries the yeast *URA3* gene and a yeast centromere, replicates autonomously in yeast, and is present in only one or two copies per cell (11; C. Mann, personal communication).

These plasmids were introduced into an isogenic set of *MAT***a**, *MAT*α, and *MAT***a**/*MAT*α cells and assayed for beta-galactosidase activity. As shown in Table 3, cells with plasmids pHO-lac-c15 or pHO-lac-c12 behaved like cells with the *HO-lacZ* gene in its normal chromosomal position: beta-galactosidase activity was substantial in haploid **a** or α cells (4 U) but was very low in *MAT***a**/*MAT*α cells (approximately 0.08 U). These plasmids contain approximately 1,360 and 762 bp of DNA upstream of the *HO* initiation codon, respectively.

Plasmid pHO-lac-c6, which has 436 bp of upstream sequences, showed only partial control by the mating type locus: 4.5 and 4.2 U in **a** and α haploid cells, respectively, and 0.7 U in **a**/α diploid cells (Table 3). Thus, expression of this *HO-lacZ* gene was reduced only sixfold in **a**/α cells rather than the 200-fold reduction seen for genes with more 5′-flanking DNA. These results indicate that the region between −762 and −436 is necessary for full repression and that 436 nucleotides of 5′-flanking DNA are sufficient for partial repression by the mating type locus.

Further removal of upstream sequences led to fully con-

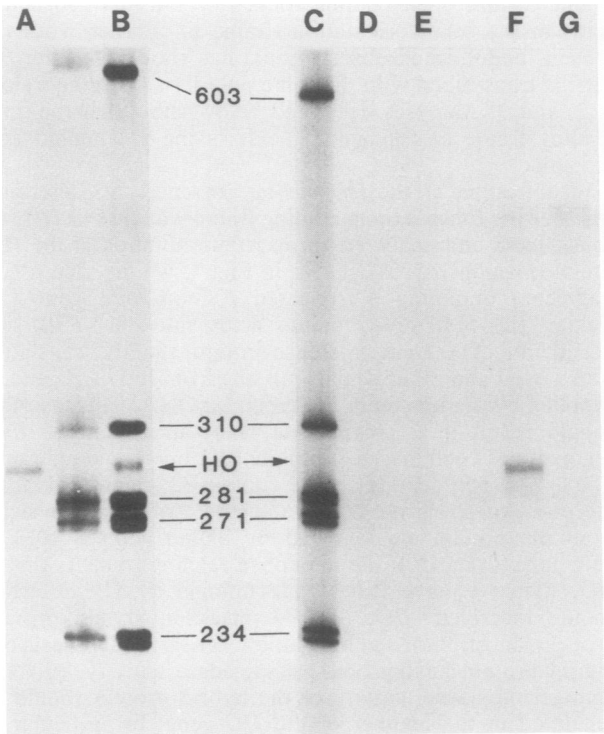


FIG. 6. Nuclease S1 mapping of the 5′ end of the *HO* transcript. RNA was isolated from strain 3B54, in which the *HO* gene is expressed, and hybridized to the HP1 fragment. After treatment with S1 nuclease, a fragment approximately 290 bases long was protected (lanes A, B, and F). HP1 is a single-stranded, radioactively labeled *Hin*dIII-*Pst*I fragment of *HO* predicted to be complementary to *HO* RNA. RNA isolated from **a**/α diploid cells (strain X10-1b) did not yield a protected fragment (lanes E and G). A fragment complementary to HP1 (HP2) was not protected by RNA from 3B54 (lane D) or X10-1b (lane E). Size markers (lanes B and C) are end-labeled *Hae*III fragments of ϕX174 DNA. Lane B is a mixture of the protected *HO* fragment and the ϕX174 markers. Sizes are shown in bases.

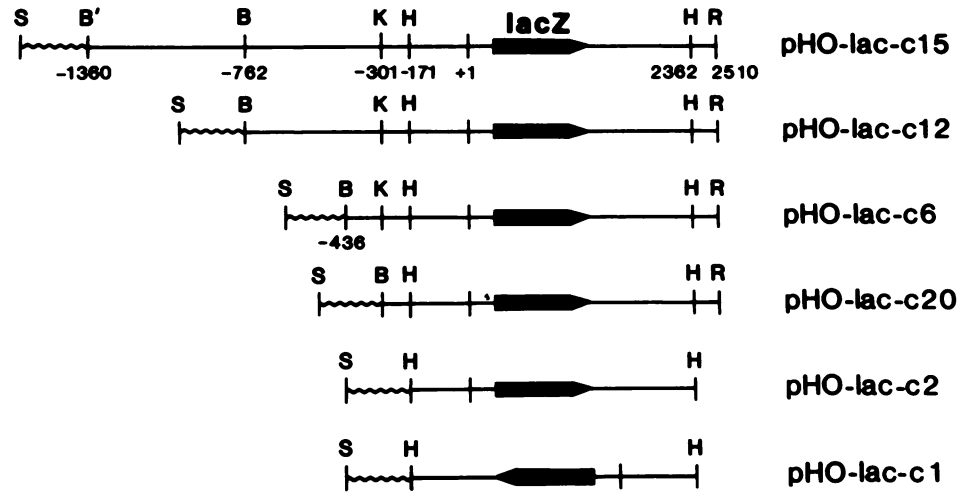


FIG. 7. Structure of *HO-lacZ* plasmids with different amounts of *HO* upstream region. The *HO-lacZ* segments are carried in YCp50 (described in Materials and Methods). Wavy lines represent vector sequences; straight lines represent *HO* sequences (not drawn to scale). B' represents the junction between a *Sau*3A site present in *HO* and a *Bam*HI site of YCp50 that does not regenerate an intact *Bam*HI site. The *Bam*HI site of pHO-lac-c20 is an alteration of the *Kpn*I site at position −301 that has been converted to a *Bam*HI site in vitro as described in Materials and Methods. The *Bam*HI site of pHO-lac-c6 was formed by joining the *Sau*3A site at position −436 to the *Bam*HI site of the vector. The *lacZ* segment is inserted at position 275 (Fig. 4). Abbreviations: S, *Sal*I; B, *Bam*HI; K, *Kpn*I; H, *Hin*dIII; R, *Eco*RI.

stitutive expression of the *HO-lacZ* gene: pHO-lac-c20, which has 301 bp of upstream sequences, directed similar levels of beta-galactosidase activity in haploid cells (4.0 and 3.9 U) and in a/α diploid cells (4.1 U). These results suggest that a site responsible for negative regulation by the mating type locus is located between −436 and −301.

A plasmid that contains only 171 bp of 5'-flanking DNA (pHO-lac-c2) also exhibited constitutive expression of beta-galactosidase activity, at a level approximately 25% of the other plasmids: approximately 1 U of enzyme activity was detected in all three cell types (Table 3). A plasmid containing this same segment in inverted orientation with respect to vector sequences (pHO-lac-c1) did not produce detectable beta-galactosidase activity. These results suggest that deletion to −171 removes a site necessary for full expression of *HO*.

A Ty1 element is not located near *HO*. Errede et al. (15) have suggested that the Ty1 element might be naturally associated with some yeast genes and therefore be responsible for such genes being under the control of the mating type locus. To determine whether *HO* is adjacent to a Ty1 element, we carried out two Southern blotting analyses.

TABLE 3. Influence of upstream region on expression of *HO*

Plasmid	5'-flanking DNA (bp)	β-Galactosidase activity[a] (U) in HR125-5d derivative:		
		*MAT***a**	*MATα*	*MAT***a**/*MATα*
pHO-lac-c15	1,360	4.2	4.0	0.08
pHO-lac-c12	762	4.0	3.8	0.04
pHO-lac-c6	436	4.5	4.2	0.7
pHO-lac-c20	301	4.0	3.9	4.1
pHO-lac-c2	171	1.1	0.9	1.2
pHO-lac-c1	171	0.02	0.01	0.03
YCp50		<0.01	<0.02	<0.01

[a] β-Galactosidase activity of **a**, α, and **a**/α stains carrying the indicated plasmids. Yeast strains are HR125-5d and isogenic derivatives. Values are averages of at least three independent assays. The size of the 5' flanking region of pHO-lac-c15 was estimated from DNA fragment lengths; other sizes are from the nucleotide sequence (Fig. 3).

When used as probe in blot hybridization, Ty1 hybridizes to approximately 30 copies of this element in the yeast genome (9). If *HO* contains a segment of Ty1, then we would anticipate a similar result—that an *HO* probe would hybridize to many yeast restriction fragments. This result was not observed when a probe that included the region between −436 and −171 was used (R. Jensen, unpublished observations). In another analysis, plasmid DNA containing the *ho* allele of *HO* (pBR328-*ho*) was probed with a radioactively labeled Ty1 fragment. This plasmid contains all of the signals required for regulation of *ho* (or *HO*) by the mating type locus. No hybridization was detected (data not shown). These data indicate that a large block of Ty1 sequences is not present at the *HO* locus.

DISCUSSION

Nucleotide sequence of the *HO* gene. In the *HO* DNA sequence described here (Fig. 3), the longest open reading frame predicts a 586-amino-acid protein with a molecular

Phe	UUU	10	Ser	UCU	7	Tyr	UAU	16	Cys	UGU	15
Phe	UUC	10	Ser	UCC	2	Tyr	UAG	6	Cys	UGC	6
Leu	UUA	9	Ser	UCA	10	—	UAA	1	—	UGA	0
Leu	UUG	11	Ser	UCG	0	—	UAG	0	Trp	UGG	6
Leu	CUU	12	Pro	CCU	7	His	CAU	11	Arg	CGU	6
Leu	CUC	3	Pro	CCC	4	His	CAC	5	Arg	CGC	5
Leu	CUA	3	Pro	CCA	11	Gln	CAA	9	Arg	CGA	3
Leu	CUG	4	Pro	CCG	2	Gln	CAG	10	Arg	CGG	0
Ile	AUU	13	Thr	ACU	8	Asn	AAU	15	Ser	AGU	8
Ile	AUC	10	Thr	ACC	4	Asn	AAC	10	Ser	AGC	4
Ile	AUA	15	Thr	ACA	12	Lys	AAA	40	Arg	AGA	15
Met	AUG	11	Thr	ACG	8	Lys	AAG	12	Arg	AGG	12
Val	GUU	15	Ala	GCU	14	Asp	GAU	18	Gly	GGU	22
Val	GUC	9	Ala	GCC	5	Asp	GAC	9	Gly	GGC	12
Val	GUA	9	Ala	GCA	10	Glu	GAA	29	Gly	GGA	13
Val	GUG	6	Ala	GCG	2	Glu	GAG	9	Gly	GGG	4

FIG. 8. Codon usage in the *HO* gene. The *HO* coding sequence is 59.2% dA,dT.

weight of 66,146. As yet, the *HO* gene product has not been isolated in sufficient quantities to allow direct protein sequencing and thus unambiguous identification of the 66,146-M_r protein as the *HO* product. However, construction of an *HO-lacZ* fusion based on this reading frame results in loss of *HO* activity and generation of a beta-galactosidase activity whose expression is regulated in a manner identical to *HO* (Table 2). Similarly, insertions of the *E. coli* transposon Tn*5* within this long open reading frame (at approximate positions 300 and 1700) also inactivate *HO* (R. Jensen, unpublished observations). Finally, Kostriken and Heffron (33) have recently expressed this reading frame in *E. coli* based on our sequence data and observed production of the *HO* endonuclease activity. These results demonstrate that the long open reading frame encodes the *HO* protein.

The deduced *HO* protein is highly basic, containing 52 Lys and 41 Arg residues (16% of total amino acids), and resembles other nuclear proteins of *S. cerevisiae* in this respect. It is also rich in acidic residues (11% Glu and Asp). The sequence predicts a secondary structure that is 30% helical (20).

An interesting feature of the sequence encoding the *HO* protein is the codon usage (Fig. 8). In general, yeast genes have a biased codon usage; most amino acids are encoded by 25 of the 61 possible sense triplets (3, 8). In the *HO* gene, however, all codons except two (Ser UCG and Arg CGG) are used, and the distribution within a given codon family is broader than for other genes. Genes that are normally expressed at high levels in yeast, such as those encoding glycolytic enzymes, often exhibit very biased codon distribution. This bias is thought to reflect the efficiency with which a given tRNA is utilized by the translation machinery and its abundance in the cell (3). The distinctive codon usage of *HO* might reflect a requirement for only a very low level of the *HO* protein, since it is an endonuclease that needs to cleave *MAT* DNA only once per cell cycle.

Two features of the amino acid sequence of the putative *HO* protein may have functional significance. The first is that the carboxy-terminal 100 amino acid residues of the *HO* protein contain five sequences of the form Cys-X-X-Cys (beginning at positions 466, 486, 508, 522, and 558; Fig. 9). Related sequences of the form Cys-X_2-His, Cys-X_3-His, Cys-X_4-His, His-X_3-His, and His-X_2-Cys are also present. These sequences are potential metal (Zn^{2+})-binding domains (reviewed in reference 3a) and may play a role in the binding of *HO* protein to DNA by formation of finger structures (40a) or participate in the loss of *HO* activity observed in vitro in crude yeast extracts upon addition of Zn^{2+} (R. Kostriken and F. Heffron, personal communication). The second feature is that the *HO* open reading frame contains nine pairs of basic residues (three each of Lys-Arg, Arg-Lys, and Lys-Lys) as well as one Arg-Arg-Arg sequence. These may provide signals for inactivation of the *HO* endonuclease activity (42) by cleavage via a *KEX2*-like protease (29a).

The *HO* transcript. Based on the nucleotide sequence, the transcript must be at least 1,850 nucleotides in length. The length of the *HO* mRNA can be estimated from RNA blotting analysis to be 1.8 to 2.0 kb (29). Nuclease S1 mapping analysis (Fig. 6) indicates that there are no introns in the first 290 nucleotides of the transcript. Furthermore, sequences indicative of RNA splicing (GTAYGT . . . TACTAAC [34]) are absent. (Although GTAYGT sequences are present beginning at positions 1002, 1106, and 1519, TACTAAC sequences are not found distally.) It is thus likely that the *HO* transcript does not contain an intervening sequence. The *HO* transcript corresponds in sequence to the upper strand of the DNA duplex (Fig. 3), with its 5′ end mapping approximately at position −45 to −50. Thus, there is an untranslated leader of approximately 46 nucleotides preceding the first AUG codon. Since this AUG begins the long open reading frame, it is presumed to define the N terminus of the *HO* protein.

The 5′ end of the mRNA maps close to the sequence TAAG (at position −29 to −32); the sequence YAAG is present at the 5′ end of a number of yeast genes (8, 13, 52). The sequence CACACA, which is present in approximately 50% of yeast mRNAs close to the initiation codon (13), is absent. On the other hand, the sequence A--AUG--U, which is common at yeast initiation codons (13), is present at the presumptive *HO* initiation codon.

The precise location of the 3′ end of the transcripts has not been defined. From the size of the mRNA, approximately 1.8 to 2.0 kb, we can only suggest where it might end. The region downstream of the *HO* termination codon contains various nucleotide sequences that are found in analogous positions in other eucaryotic genes. The common eucaryotic 3′-end signal AAUAAA is present at nucleotide 1793; its significance for transcription in yeast is yet to be established. The tripartite sequence UAG . . . UAUGU . . . UUU also has been found at the 3′ ends of many yeast mRNAs (67); its components are present in the 3′-flanking sequence of *HO* at nucleotides 1773, 1788, and 1806.

Nucleotide sequence comparisons of the 5′ region. Comparisons of DNA sequences upstream of yeast mRNA coding regions has identified several sequences that are often conserved and which, when removed by deletion, lead to changes in the control of transcription or in the position of the 5′ ends of the transcripts (2, 13, 18, 23, 27, 35, 52; reviewed in reference 22). Most prominent are sequences

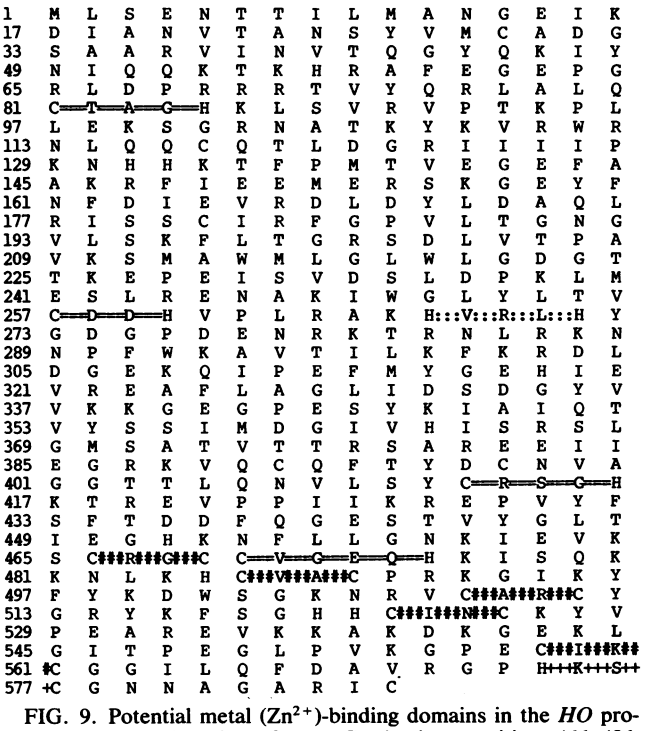


FIG. 9. Potential metal (Zn^{2+})-binding domains in the *HO* protein. Sequences of the form Cys-X_2-Cys begin at positions 466, 486, 508, 522, and 558 as indicated. Sequences of the form Cys-X_{2-4}-His begin at positions 81, 257, 412, and 470; His-X_3-His begins at position 267; and His-X_2-Cys begins at position 574.

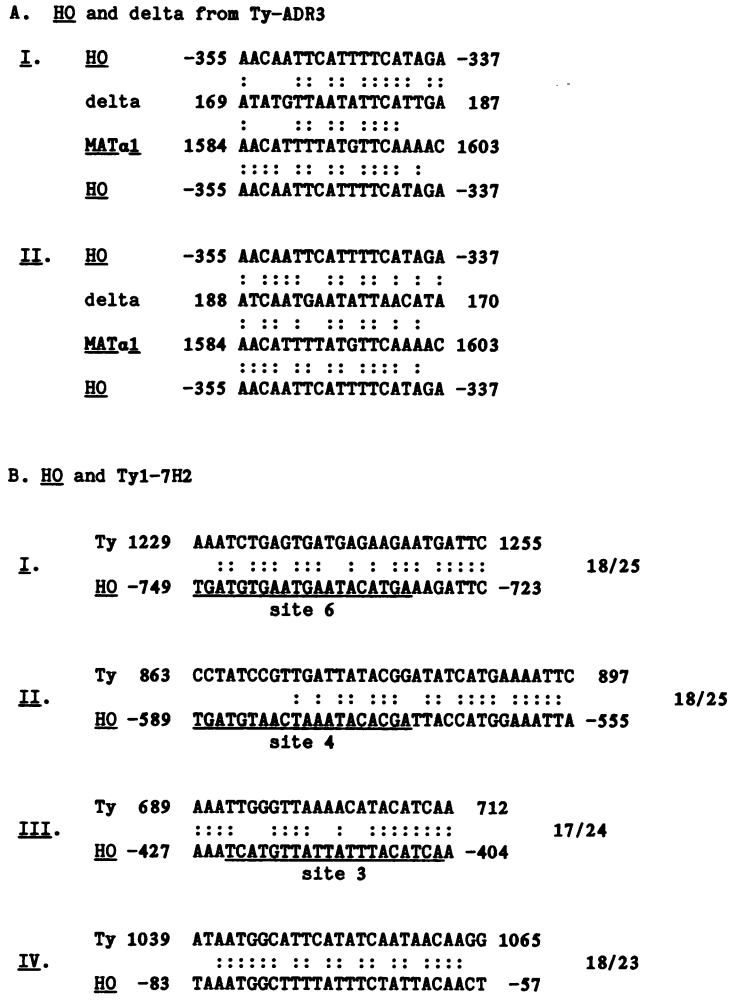


FIG. 10. (A) Comparison of *HO* and delta from Ty-*ADR3-2*^{c}. Numbers for *HO* are positions relative to the translation initiation site; for delta, the position from the end of the delta; and for *MATα1*, as described in reference 55. Sequence information is from references 1 (*MATα*) and 64 (Ty-*ADR3-2*^{c}). (B) Comparison of *HO* and delta from Ty1-*CYC7-H2*. Numbering of the Ty1 sequences is according to reference 16, in which 49 to 185 are delta sequences and 186 to 1556 are adjacent Ty1 (epsilon) sequences. Sequence comparisons were performed with a diagonal-traverse homology search (version 1.2 [63]). Sites 3, 4, and 6 are presumed a1-α2 sites (40) and are shown also in Fig. 13. A comparison of the *HO* sequence with the complement of the Ty1-*CYC7-H2* sequence identified four regions of homology that are not shown. They occur in *HO* and Ty1 as follows. Region 1: 17 of 25: 1372–1348 in Ty, −271 to −247 in *HO*. Region 2: 19 of 28: 1234–1207 in Ty, −198 to −171 in *HO*. Region 3: 17 of 25: 1015–991 in Ty, −80 to −56 in *HO*. Region 4: 18 of 25: 84–60 in Ty, −65 to −41 in *HO*.

related to TATAAATA, although they do not lie in the same fixed relationship to mRNA 5′ ends that is observed in genes of higher eucaryotes (5). Several sequences of this nature occur in the 5′-flanking region of *HO*, including an A/T stretch of 14 bp located 65 nucleotides upstream of the 5′ end of the mRNA (−113 to −126; interlined in Fig. 3).

Because the *HO* gene is transcribed in haploid cells and not in a/α cells (29), we compared sequences adjacent to other genes that are subject to this type of regulation. These include yeast Ty elements (14) and the *MATα1* and *MATα2* genes (32, 46). We carried out two types of nucleotide sequence comparisons, one between *HO* and sequences from Ty elements, the other between *HO* and *MATα*.

(i) *HO* and Ty. We compared the *HO* upstream region with sequences from two different Ty elements, Ty-*ADR3-2*^{c} (64) and Ty1-*CYC7-H2* (Ty1-7H2) (16). In the former case, we compared a 338-bp delta sequence of the Ty element with *HO*. The most extensive homologies are matches of 11 of 14 and 12 of 19 (Fig. 10A). The *HO* segment so identified (from −355 to −337) bears considerable homology (13 of 17 nucleotides) to a region of *MATα1* (1584–1603). The significance of these homologies is unknown. They are unrelated to the homologies observed by Errede et al. (16) (see below) and are unlikely to be involved in regulation by a1-α2 for several reasons: (i) the *HO* sequence is not related to the a1-α2 sites discussed below, and (ii) deletion of the *MATα* site does not affect regulation of *MATα1* and *MATα2* (54). Comparison of the upstream region of *HO* with the delta region of Ty1-7H2 and 1,371 nucleotides of its adjacent Ty sequences reveals four substantial homologies (Fig. 10B). Homologies I, II, and III overlap sites 6, 4, and 3, respectively, which have been proposed to be a1-α2 regulatory sites (40) (discussed below). A small segment from Ty1-7H2 that contains homology III (nucleotides 668–779) is sufficient

A. HO and MATα sequence sufficient for regulation by a1-α2

```
MATα 1643 GCTTCCCAATGTAGAAAAGTACCTCATA 1670
          :      :::::   :: ::::  ::
HO   -595 GAAAAGTGATGTAACTAAATACACGATT -568
                     site 4

MATα 1670 TATGATGTACTTTTCTACATTGGGAAGC 1643
           :: ::::  :: : :::::    :
HO   -425  ATCATGTTATTATTTACATCAACATA  -400
                  site 3
```

B. HO and other MATα sequences

```
MATα  1638 TTATTGCTTCCCAATGTAGAAAAGT 1662
           ::::::: : :::   ::
HO    -198 CTATTGCTACTCAAATGAG       -180
                                :::::::
HO    -246                      GAAAAGT -240
                                :::::::
HO    -595                      GAAAAGT -589
```

FIG. 11. Nucleotide sequence comparisons. (A) Numbers for *HO* are positions relative to the translation initiation site; for *MATα1*, as described in reference 55. The 28-bp sequence is sufficient to confer regulation by a1-α2 (55). This sequence and its complement were used to search for matches in the *HO* upstream region. Two matches of 11 of 15 or better were found with the diagonal-traverse homology search (63). Underlined sequences correspond to sites 3 and 4 of Miller et al. (40). (B) The upstream region of *HO* was searched for perfect matches to a 25-bp sequence of *MATα*, which contains most of the 28-bp region sufficient for a1-α2 regulation.

to place a test gene under control by a1-α2 (B. Errede, unpublished observations). These comparisons, together with our Southern hybridization results, indicate that although the homology between Ty1 and *HO* is not significant at a gross hybridization level, it is evident at the nucleotide level—Ty and *HO* share the presumed a1-α2 recognition sites. The fourth region of homology between *HO* and Ty1-7H2 (IV) is a strong one (18 of 23 nucleotides), but its functional role is unknown.

(ii) *HO* and *MATα*. Siliciano and Tatchell (55) have identified a 28-bp sequence from the region between *MATα1* and

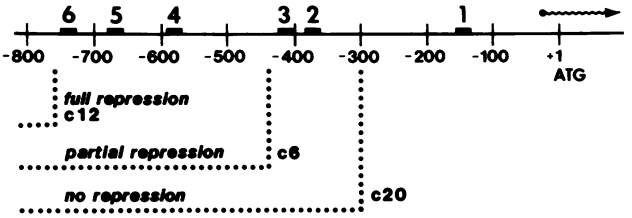


FIG. 13. Positions of a1-α2 consensus regulatory sequences with respect to deletion endpoints. The region upstream of *HO* is drawn to physical scale, showing the positions of the proposed a1-α2 regulatory sequences (40) (Fig. 12) and the deletion endpoints analyzed here. The positions of the a1-α2 regulatory sequences and (in parentheses) the extent to which they match the a1-α2 consensus sequence (Fig. 12, line 4) or its complement (Fig. 12, line 8) are as follows: site 6, −749 to −730 (19 of 20); site 5, −679 to −660 (18 of 20); site 4, −589 to −570 (20 of 20); site 3, −424 to −405 (20 of 20 to complement); site 2, −384 to −365 (19 of 20 to complement); site 1, −156 to −137 (16 of 20).

MATα2 genes that is sufficient to confer regulation by a1-α2. We examined the upstream region of *HO* for homologies to this *MATα* sequence and its complement and observed two strong homologies. These correspond to sites 3 and 4 of Miller et al. (40) (Fig. 11). Other homologies (of unknown functional significance) are also shown in Fig. 11.

The two sets of sequence comparisons described above (Fig. 10 and 11) have shown that *HO* sequences corresponding to sites 3 and 4 of Miller et al. (40) (positions −405 to −424 and −570 to −589) are homologous to both *MATα* and Ty1. Comparisons between the corresponding regions of *MATα* and Ty1 likewise reveal homologies (Fig. 12), which have been noted previously (16). We presume that these regions common to Ty, *HO*, and *MATα* confer regulation by a1-α2; this is known to be the case for the *MATα* sequence (55), *HO* site 3 (40), and the Ty segment 680–717 (B. Errede, unpublished observations). Errede et al. (16) and Roeder et al. (49) have noted that the segments of Ty described here contain the sequence CTTTCCA, which is present in mammalian enhancers, and suggest that a1-α2 might repress transcription of Ty1 by influencing this site. *HO* does not contain sequences that match CTTTCCA (or its complement) perfectly. It does, however, have two 6 of 7 matches (−686 to −680 and −563 to −558), which are adjacent to

```
Ty     680  TGACTTTCCAAATTGGGTTAAAACATACATCAAATTTT  717  enhancer core
            :    ::::   ::::  :  ::::::::  : :
HO    -436  GATCCACGAAAATCATGTTATTATTTACATCAACATAT -399  site 3
             :      : :: ::::  :: : :::::    :
MATα 1680  AGTTGTTTCATATGATGTACTTTTCTACATTGGGAAGC  1643  sufficient
a1-α2 con              TGATGTA—T—A—ACA—C—GA
                              T      T

Ty     857 AGACATCCTATCCGTTGATTATACGGATATCATGAAAATTCTTTCCAAA  905  enhancer core
            : :          : : :: :::  :: :::: :::::
HO    -595 GAAAAGTGATGTAACTAAATACACGATTACCATGGAAATTAACGTACCT -547  site 4
           :      :::::   :: ::::  ::    :   :  : :::
MATα 1643 GCTTCCCAATGTAGAAAAGTACATCATATGAAACAACTTAAACTCTTAA 1691  sufficient
a1-α2 con       TCATGT—A—A—TACATCA
                  G    T
```

FIG. 12. Nucleotide sequence comparisons of *HO*, Ty, and *MATα*. Conventions on numbering are given in the legend to Fig. 10. The enhancer core is a sequence identified in reference 16; site 3 and site 4 are described in reference 40 and in Fig. 10; "sufficient" indicates the 28-bp sequence that is sufficient for conferring regulation by a1-α2 (55).

A. ARS consensus Elements

Consensus		(A/T)	T	T	T	A	T	(A/C)	T	T	T	(A/T)		
I.	325	c*	T	T	T	A	T	A	T	T	T	g*	315	9/11
II.	1806	T	T	T	a	A	T	A	T	T	T	T	1816	10/11
Frequency[1]		13	19	14	16	16	18	19	19	16	17	19		

B. 3' Element Consensus

Consensus		C	T	T	T	T	A	G	C	(A/T)	(A/T)	(A/T)		
I. w	188	C	T	g	g*	T	A	G	g	T	T	A	198	(116)[2]
I. x	226	C	g*	T	c*	T	t*	G	C	A	T	T	236	(78)
I. y	286	C	T	g	T	T	g*	G	a*	A	A	A	296	(18)
I. z	299	C	T	T	T	T	t*	t*	C	c	A	A	289	(15)
II. x	1832	t	T	T	T	T	A	G	a*	c*	T	T	1842	(15)
II. y	1840	C	T	T	T	T	t*	c*	t*	T	A	A	1850	(23)
II. z	1894	C	T	T	T	T	t*	G	t*	A	T	c	1884	(67)
Frequency[3]		7	10	8	12	13	14	13	13	14	12	10		

FIG. 14. (A) Core element consensus (6) and (B) 3' element consensus (47). Matches to the consensus sequences are shown in uppercase letters. Asterisks indicate bases that were not previously observed at that position in reference 47. Sequence I contains the only match to nine consecutive bases of the core consensus. There are two additional sequences that match the core consensus in eight consecutive bases (936 to 943 and 2077 to 2084). Matches of 9 of 11 nucleotides to the 3' conserved element consensus begin at positions −475, −57, and 577 and to its complement at positions −87, 711, 736, 1605, 1754, and 2058. (A) The frequency of occurrence in a list of 19 known and putative ARS sequences (Table 1 in reference 47) is shown. (B) Distance from core element (in bases) is shown in parentheses. The frequency of occurrence in a list of 14 putative 3' conserved elements is also shown (Table 2 in reference 47). These searches were carried out with the probe match routine from the Mount-Conrad Computer Programs for DNA and Protein Sequence Analysis (version 3.7; D. M. Mount and B. Conrad, personal communication).

sites 5 and 4, respectively. The latter is part of a palindromic sequence ATTaCCATGGaAAT adjacent to site 4 (−570 to −547; Fig. 12). The significance of these homologies and whether Ty1 and *HO* share any common activator proteins are not known.

Functional analysis of the upstream region. Nasmyth (43, 44) has defined regions essential for expression of the *HO* gene by producing in vitro deletions in the upstream region and then assaying their effects after gene replacement. He found that the region between −175 and −931 (as in the deletion mutant 229–102) is not essential for expression, whereas sequences further upstream (between −1000 and −1400) are essential and define a region termed URS1. Our analysis of the regulation of *HO* by **a**1-α2 used several plasmids that lack this entire essential region: in fact, plasmids containing 762, 436, and 301 bp of upstream region all gave approximately the same level of expression as our plasmid with the largest upstream region (which had approximately 1,360 bp). We assume that the expression of the *HO* gene on our plasmids lacking URS1 results from a plasmid sequence that supplies a functional equivalent of URS1. The existence of a plasmid-borne sequence with "upstream activation site" activity has been strongly indicated from several studies (65). These data clearly demonstrate that there are major differences between the two deletion mapping strategies for analysis of essential regions, the plasmid method that we have used and the in situ method used by Nasmyth (43, 44).

To delimit sequences necessary for regulation of *HO* by **a**1-α2, we analyzed plasmids that contain different segments of the upstream region (Fig. 7, Table 3). A plasmid that contains 762 bp preceding the initiation codon is strongly regulated by the mating type locus. This plasmid (as well as one with approximately 1,360 bp of upstream region) did exhibit a measurable level of expression in **a**/α cells, a level greater than observed for the *HO-lacZ* gene when it was located at its chromosomal position. This residual expression might occur because the gene is carried on a plasmid or because it lacks sequences essential for full repression. A plasmid with only 436 bp preceding the initiation codon was also subject to regulation by the mating type locus but exhibited significant constitutivity in **a**/α cells (nearly 25% the level of haploid strains). Deletion of an additional 135 bases from the 5'-flanking region led to full constitutivity.

These results demonstrate that the *HO* gene can be regulated when carried on a plasmid and that targets for negative regulation by **a**1-α2 are present on plasmids that contain only 436 bp prior to the translation start point. We

found that the *HO* segment from −762 to −301 was sufficient for a1-α2 control. When this segment was placed upstream of the yeast galactokinase gene (*GAL1*), galactokinase activity was observed in haploid cells (or **a**/**a** diploids) but not in **a**/α diploids (R. Jensen, unpublished observations). The progressive loss of regulation by the mating type locus in deleting from −762 to −436 and from −436 to −301 suggests that both of these intervals contain sequences necessary for regulation by a1-α2. This conclusion is supported by studies of Miller et al. (40) (discussed below). In addition, it is supported by the observation that α2 protein binds in vitro to both of these regions (A. Johnson, personal communication). The precise sites of binding and their relationship to a1-α2 inhibition are under study.

Miller et al. (40) have proposed that regulation of *HO* by a1-α2 is mediated through a 20-bp sequence that is repeated in the *HO* upstream region (Fig. 12). The region from −1 to −762 contains six such sequences (Fig. 13). Four additional sequences are found in the 1,100 bp preceding position −762 (40). A key argument that these sites are responsible for regulation by a1-α2 is that fragments containing either site 6 (nucleotides −715 to −761) or site 3 (nucleotides −397 to −444) confer regulation by a1-α2 to a test gene (40). Our deletion analysis showed that the plasmid containing all six sites was strongly repressed, whereas a plasmid that retained sites 1, 2, and 3 was partially constitutive. The plasmid that retained only site 1 was fully constitutive. Thus, the magnitude of repression is greater with an increase in the number of sites. Our results indicate that site 1, which has the weakest match of these six sites to the consensus sequence (16 of 20; see legend to Fig. 13) is not sufficient for regulation by a1-α2, at least when carried on a plasmid. Two other sequences with 16 of 20 matches to the consensus are located within the coding region (822–841 and 1168–1187) and must also be ineffective in repression.

Homologies to ARS consensus sequences. An ARS element permits autonomous replication of plasmids in *S. cerevisiae* (17, 59). Two segments of *HO* and flanking regions are able to promote autonomous plasmid replication, one (BB) consisting of nucleotides −762 to 705 and the other (BH) of nucleotides 705 to 2362 (R. Jensen, unpublished observations). Nucleotide sequence comparisons of segments with ARS activity has led to identification of an 11-bp sequence that defines an essential "core element" (6) (Fig. 14). A second conserved sequence, also of 11 bp, has recently been found (the "3′ conserved element"), which is located 3′ to the T-rich strand of the core element and within 50 to 100 nucleotides of it (47). The BB segment contains nine sequences that match the core consensus in 9 of 11 positions. One of these (nucleotides 316–324; sequence I, Fig. 14) matches the core consensus in the central nine nucleotides. The BB segment contains four sequences that match the 3′ conserved element in 9 of 11 positions, but none of these is situated appropriately with respect to sequence I. Four sequences with an 8 of 11 match are situated appropriately and are shown in Fig. 14. The BH segment contains 22 sequences that match the core in 9 of 11 positions; one sequence (nucleotides 1806–1816) matches in 10 of 11 positions. Kearsey has found (30, 31) that a segment containing nucleotides 1796 to 1852 has ARS activity and that point mutations within the core homology inactivate the ARS. Hence, sequence II appears to be functional. Although the BH segment contains five sequences that match the 3′ conserved element in 9 of 11 positions, none are positioned appropriately with respect sequence II or contained on the 1796 to 1852 fragment. As noted by Palzkill et al. (47), the region adjacent to this core sequence contains homologies to the 3′ conserved element. It is not known whether sequence I (or other homologous sequences) are functional or whether ARS sequences play a role in any aspect of *HO* regulation.

ACKNOWLEDGMENTS

We thank Wendy Blankfort-Doyle and Prudence Bothen for help in preparation of this manuscript, Tim Palzkill and Carol Newlon for sending us their paper prior to publication and for discussion, and David Mount for introducing one of us (I.H.) to his sequence analysis programs.

This research was supported by research grants from the Public Health Service at the University of California, San Francisco, 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. D.W.R. was the recipient of a Damon Runyon-Walter Winchell Cancer Fund Postdoctoral Fellowship. Work by R.J. was submitted in partial fulfillment of the requirements for a Ph.D. (University of Oregon).

LITERATURE CITED

1. **Astell, C. R., L. Ahlstrom-Jonasson, M. Smith, K. Tatchell, K. A. Nasmyth, and B. D. Hall.** 1981. The sequence of the DNAs coding for the mating type loci of *Saccharomyces cerevisiae*. Cell **2**:15–23.
2. **Beier, D. R., and E. T. Young.** 1982. Characterization of a regulatory region upstream of the *ADR2* locus of *S. cerevisiae*. Nature (London) **300**:724–728.
3. **Bennetzen, J. L., and B. D. Hall.** 1982. Codon selection in yeast. J. Biol. Chem. **257**:3026–3031.

3a. **Berg, J. M.** 1986. Potential metal-binding domains in nucleic acid binding proteins. Science **232**:485–487.

4. **Berk, A. J., and P. A. Sharp.** 1978. Spliced early mRNAs of simian virus 40. Proc. Natl. Acad. Sci. USA **75**:1274–1278.
5. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eukaryote split genes coding for protein. Annu. Rev. Biochem. **50**:349–383.
6. **Broach, J. R., Y.-Y. Li, J. Feldman, M. Jayaram, J. Abraham, K. A. Nasmyth, and J. B. Hicks.** 1983. Localization and sequence analysis of yeast origins of DNA replication. Cold Spring Harbor Symp. Quant. Biol. **47**:1165–1173.
7. **Broach, J. R., J. N. Strathern, and J. B. Hicks.** 1979. Transformation in yeast: development of a hybrid cloning vector and isolation of the *CAN1* gene. Gene **8**:121–133.
8. **Burke, R. L., P. Tekamp-Olson, and R. Najarian.** 1983. The isolation, characterization and sequence of the pyruvate kinase gene of *Saccharomyces cerevisiae*. J. Biol. Chem. **258**:2193–2201.
9. **Cameron, J. R., E. Y. Loh, and R. W. Davis.** 1979. Evidence for transposition of dispersed repetitive DNA families in yeast. Cell **16**:739–751.
10. **Casadaban, M. J., A. Martinez-Arias, S. K. Shapira, and J. Chou.** 1983. Beta-galactosidase gene fusions for analysing gene expression in *Escherichia coli* and yeast. Methods Enzymol. **100**:293–308.
11. **Clarke, L., and J. Carbon.** 1982. Isolation of a yeast centromere and construction of functional small circular chromosomes. Nature (London) **287**:504–509.
12. **Delaney, A. D.** 1982. A DNA sequence handling program. Nucleic Acids Res. **10**:61–67.
13. **Dobson, M. J., M. F. Tuite, N. A. Roberts, A. J. Kingsman, and S. M. Kingsman.** 1982. Conservation of high efficiency promoters in *Saccharomyces cerevisiae*. Nucleic Acids Res. **10**:2625–2637.
14. **Elder, R. T., T. P. St. John, D. T. Stinchcomb, and R. W. Davis.** 1981. Studies on the transposable element Ty1 of yeast. I. RNA homologous to Ty1. Cold Spring Harbor Symp. Quant. Biol. **45**:581–584.
15. **Errede, B., T. S. Cardillo, 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.
16. **Errede, B., M. Company, J. D. Ferchak, C. A. Hutchison III, and W. S. Yarnell.** 1985. Activation regions in a yeast transposon have homology to mating type control sequences and to mammalian enhancers. Proc. Natl. Acad. Sci. USA **82**:5423–5427.
17. **Fangman, W. L., and V. A. Zakian.** 1981. Genome structure and replication, p. 27–58. *In* J. N. Strathern, E. W. Jones, and J. R. Broach (ed.), The molecular biology of the yeast *Saccharomyces cerevisiae*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Faye, G., D. W. Leung, K. Tatchell, B. D. Hall, and M. Smith.** 1981. Deletion mapping of sequences essential for the *in vivo* transcription of the iso-1-cytochrome *c* gene. Proc. Natl. Acad. Sci. USA **78**:2258–2262.
19. **Fink, G. R., and C. A. Styles.** 1972. Curing of a killer factor in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **69**:2846–2849.
20. **Garnier, J., D. J. Osguthorpe, and B. Robson.** 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**:97–120.
21. **Grunstein, M., and D. S. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA **72**:3961–3965.
22. **Guarente, L.** 1984. Yeast promoters: positive and negative elements. Cell **36**:799–800.
23. **Guarente, L., and T. Mason.** 1983. Heme regulates transcription of the *CYC1* gene of *S. cerevisiae* via an upstream activation site. Cell **32**:1279–1286.
24. **Guarente, L., and M. Ptashne.** 1981. Fusion of *Escherichia coli lacZ* to the cytochrome *c* gene of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **78**:2199–2203.
25. **Herskowitz, I., and Y. Oshima.** 1981. Control of cell type in *Saccharomyces cerevisiae*: mating type and mating-type interconversion, p. 181–209. *In* J. N. Strathern, E. W. Jones, and J. R. Broach (ed.), The molecular biology of the yeast *Saccharomyces cerevisiae*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Hicks, J. B., and I. Herskowitz.** 1976. Interconversion of yeast mating types. I. Direct observation of the action of the homothallism (*HO*) gene. Genetics **83**:245–258.
27. **Hitzeman, R. A., F. E. Hagie, J. S. Hayflick, C. Y. Chen, P. H. Seeburg, and R. Derynck.** 1982. The primary structure of the *Saccharomyces cerevisiae* gene for 3-phosphoglycerate kinase. Nucleic Acids Res. **10**:7791–7808.
28. **Jensen, R. E., and I. Herskowitz.** 1984. Directionality and regulation of cassette substitution in yeast. Cold Spring Harbor Symp. Quant. Biol. **49**:97–104.
29. **Jensen, R. E., G. F. Sprague, Jr., and I. Herskowitz.** 1983. Regulation of yeast mating-type interconversion: feedback control of *HO* gene expression by the mating-type locus. Proc. Natl. Acad. Sci. USA **80**:3035–3039.
29a. **Julius, D., A. Brake, L. Blair, R. Kunisawa, and J. Thorner.** 1984. Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast prepro-alpha-factor. Cell **37**:1075–1089.
30. **Kearsey, S.** 1983. Analysis of sequences conferring autonomous replication in baker's yeast. EMBO J. **2**:1571–1575.
31. **Kearsey, S.** 1984. Structural requirements for the function of a yeast chromosomal replicator. Cell **37**:299–307.
32. **Klar, A. J. S., J. N. Strathern, J. R. Broach, and J. B. Hicks.** 1981. Regulation of transcription in expressed and unexpressed mating type cassettes of yeast. Nature (London) **289**:239–244.
33. **Kostriken, R., and F. Heffron.** 1984. The product of the *HO* gene is a nuclease: purification and characterization of the enzyme. Cold Spring Harbor Symp. Quant. Biol. **49**:89–96.
34. **Langford, C. J., and D. Gallwitz.** 1983. Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. Cell **33**:519–527.
35. **Lowry, C. V., J. L. Weiss, D. A. Walthall, and R. S. Zitomer.** 1983. Modulator sequences mediate oxygen regulation of *CYC1* and a neighboring gene in yeast. Proc. Natl. Acad. Sci. USA **80**:151–155.
36. **MacKay, V. L.** 1984. Cloning of yeast *STE* genes in 2 millimicron vectors. Methods Enzymol. **101**:325–343.
37. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
38. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
39. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–322.
40. **Miller, A. M., V. L. MacKay, and K. A. Nasmyth.** 1985. Identification and comparison of two sequence elements that confer cell-type specific transcription in yeast. Nature (London) **314**:598–603.
40a. **Miller, J., A. D. McLachlan, and A. Klug.** 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J. **4**:1609–1614.
41. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
42. **Nasmyth, K.** 1983. Molecular analysis of a cell lineage. Nature (London) **302**:670–676.
43. **Nasmyth, K.** 1985. At least 1400 base pairs of 5′-flanking DNA is required for the correct expression of the *HO* gene in yeast. Cell **42**:213–223.
44. **Nasmyth, K.** 1985. A repetitive DNA sequence that confers cell-cycle *START* (*CDC28*)-dependent transcription of the *HO* gene in yeast. Cell **42**:225–235.
45. **Nasmyth, K. A., and S. I. Reed.** 1980. Isolation of genes by complementation in yeast: molecular cloning of a cell-cycle gene. Proc. Natl. Acad. Sci. USA **77**:2119–2123.
46. **Nasmyth, K. A., K. Tatchell, B. D. Hall, C. Astell, and M. Smith.** 1981. A position effect in the control of transcription at yeast mating type loci. Nature (London) **289**:244–250.
47. **Palzkill, T. G., S. G. Oliver, and C. S. Newlon.** 1986. DNA sequence analysis of *ARS* elements from chromosome III of *Saccharomyces cerevisiae*: Identification of a new conserved element. Nucleic Acids Res. **14**:6247–6264.
48. **Roberts, R. J.** 1983. Restriction and modification enzymes and their recognition sequences. Nucleic Acids Res. **11**:r135–r167.
49. **Roeder, G. S., A. B. Rose, and R. E. Pearlman.** 1985. Transposable element sequences involved in the enhancement of yeast gene expression. Proc. Natl. Acad. Sci. USA **82**:5428–5432.
50. **Rose, M., M. J. Casadaban, and D. Bostein.** 1981. Yeast genes fused to beta-galactosidase in *Escherichia coli* can be expressed normally in yeast. Proc. Natl. Acad. Sci. USA **78**:2460–2464.
51. **Rothstein, R. J.** 1983. One-step gene disruption in yeast. Methods Enzymol. **101**:202–211.
52. **Russell, D. W., M. Smith, V. M. Williamson, and E. T. Young.** 1983. Nucleotide sequence of the yeast alcohol dehydrogenase II gene. J. Biol. Chem. **258**:2674–2682.
53. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
54. **Siliciano, P. G., and K. Tatchell.** 1984. Transcription and regulatory signals at the mating type locus in yeast. Cell **37**:969–978.
55. **Siliciano, P. G., and K. Tatchell.** 1986. Identification of the DNA sequences controlling the expression of the *MATα* locus of yeast. Proc. Natl. Acad. Sci. USA **83**:2320–2324.
56. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
57. **Sprague, G.F., Jr., and I. Herskowitz.** 1981. Control of yeast cell type by the mating type locus. I. Identification and control of expression of the a-specific gene, *BAR1*. J. Mol. Biol. **153**:305–321.
58. **Stern, M., R. Jensen, and I. Herskowitz.** 1984. Five *SWI* genes are required for expression of the *HO* gene in yeast. J. Mol. Biol. **178**:853–868.
59. **Stinchcomb, D. T., K. Struhl, and R. W. Davis.** 1979. Isolation and characterization of a yeast chromosomal replicator. Nature (London) **282**:39–43.
60. **Strathern, J. N., and I. Herskowitz.** 1979. Asymmetry and

directionality in production of new cell types during clonal growth: the switching pattern of homothallic yeast. Cell **17**:371–381.

61. **Strathern, J. N., A. J. S. Klar, J. B. Hicks, J. A. Abraham, J. M. Ivy, K. A. Nasmyth, and C. McGill.** 1982. Homothallic switching of yeast mating type cassettes is initiated by a double-stranded cut in the *MAT* locus. Cell **31**:183–192.
62. **Struhl, K., D. T. Stinchcomb, S. Scherer, and R. W. Davis.** 1979. High-frequency transformation of yeast: autonomous replication of hybrid DNA molecules. Proc. Natl. Acad. Sci. USA **76**:1035–1039.
63. **White, C. T., S. C. Hardies, C. A. Hutchison III, and M. H. Edgell.** 1984. The diagonal-traverse homology search algorithm for locating similarities between two sequences. Nucleic Acids Res. **12**:751–766.
64. **Williamson, V. M., D. Cox, E. T. Young, D. W. Russell, and M. Smith.** 1983. Characterization of transposable element-associated mutations that alter yeast alcohol dehydrogenase II expression. Mol. Cell. Biol. **3**:20–31.
65. **Wilson, K. L., and I. Herskowitz.** 1986. Sequences upstream of the *STE6* gene required for its expression and regulation by the mating type locus in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **83**:2536–2540.
66. **Winge, Ö., and C. Roberts.** 1949. A gene for diploidization in yeasts. C. R. Trav. Lab. Carlsberg, Ser. Physiol. **24**:341–346.
67. **Zaret, K. S., and F. Sherman.** 1982. DNA sequence required for efficient transcription termination in yeast. Cell **28**: 563–573.