MOLECULAR AND CELLULAR BIOLOGY, Dec. 1985, p. 3545–3551
0270-7306/85/123545-07$02.00/0

Vol. 5, No. 12

Saccharomyces cerevisiae CYC1 mRNA 5'-End Positioning: Analysis by In Vitro Mutagenesis, Using Synthetic Duplexes with Random Mismatch Base Pairs

J. BRYAN McNEIL† AND MICHAEL SMITH*

Department of Biochemistry, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia V6T 1W5, Canada

Received 20 May 1985/Accepted 17 August 1985

Expression of the *Saccharomyces cerevisiae CYC1* gene produces mRNA with more than 20 different 5' ends. A derivative of the *CYC1* gene (*CYC1-157*) was constructed with a deletion of a portion of the *CYC1* 5'-noncoding region, which includes the sites at which many of the *CYC1* mRNAs 5' ends map. A 54-mer double-stranded oligonucleotide homologous with the deleted sequence of *CYC1-157* and which included a low level of random base pair mismatches (an average of two mismatches per duplex) was used to construct mutants of the *CYC1* gene and examine the role of the DNA sequence at and immediately adjacent to the mRNA 5' ends in specifying their locations. The effect of these mutations on the site selection of mRNA 5' ends was examined by primer extension. Results indicate that there is a strong preference for 5' ends which align with an A residue (T in the template DNA strand) preceded by a short tract of pyrimidine residues.

In higher eucaryotes, in vivo studies suggest that A-T-rich (TATA) sequences are required for proper transcription initiation and are positioned 25 to 30 base pairs (bp) upstream from the 5' end of the corresponding mRNA (3, 20, 21, 25). Examination of transcription start sites indicates that mRNA 5' ends usually begin with an A residue, aligned with a T in the template strand, which has a short region of purines on both sides.

In *Saccharomyces cerevisiae* , the relationship between the position of mRNA 5' ends and promoter elements is not well understood. TATA sequences have been observed within the 5'-noncoding region of most RNA polymerase II-transcribed *S. cerevisiae* genes (26). Studies suggest that TATAs are required for wild-type transcription initiation of the yeast *HIS3* (29) and the *MATα1* and *MATα2* (27) genes. Many *S. cerevisiae* genes transcribe mRNAs with a multiplicity of 5' ends (7, 14, 15, 30). The large number of 5' ends suggests a lack of the stringent spatial relationship between TATA sequences and mRNA 5' ends observed in higher eucaryotic systems. Expression of the *S. cerevisiae CYC1* gene involves the initiation of more than 20 mRNAs with different 5' termini spaced over a 100-bp region at variable efficiencies. To obtain more information concerning the criteria for the positioning of *CYC1* mRNA 5' ends and the efficiency of initiation, an in vitro mutagenesis analysis of the DNA sequence immediately adjacent to *CYC1* mRNA start sites was carried out. The *CYC1* gene is an attractive target for this type of study because the expression and regulation of *CYC1* has been (7–13, 16, 23), and is, under extensive examination. A strategy for creating mutations in the regions of mRNA start sites was developed for this study, in which the start region of the *CYC1* gene, in a recombinant DNA clone, was replaced by a chemically synthesized duplex containing a low level of base pair mismatches. After transformation of *Escherichia coli* with the semisynthetic plasmid, mutants containing 1 or 2 bp changes, in the desired target region, could be identified at a high frequency by screening and sequence determination. The effect of the mutations on the sites of *CYC1* mRNA 5' ends was analyzed by transformation of *S. cerevisiae* and mRNA 5' mapping by primer extension.

MATERIALS AND METHODS

Construction of *CYC1-157*. A derivative of the *S. cerevisiae CYC1* gene was constructed with a 58-bp deletion of the transcription start site region. The deleted segment was replaced with tandem *Pst*I linkers. The *CYC1* mutant was constructed as described by McKnight and Kingsbury (17). Two fractions of pBM142 (which includes the *S. cerevisiae CYC1* gene contained on a 2.5-kilobase *Hin*dIII-*Bam*HI fragment cloned into the *Hin*dIII-*Bam*HI site of an *Eco*RI-resistant derivative of pEMBL8(+) [5]), was digested with *Eco*RI or *Xho*I. Subsequent exonuclease III-endonuclease S1 digestions were carried out so that a spectrum of deletion endpoints were produced, and *Pst*I linkers (GCTGCAGC) were ligated onto these endpoints. A single clone was retrieved from each of the two families of deletions with a deletion endpoint near the distal boundary of the *CYC1* transcription start site region with respect to the original restriction site (*Eco*RI or *Xho*I). These two DNAs were used to reconstruct the *CYC1* gene with a 58-bp deletion by combined digestion of both plasmids with *Pst*I and *Bam*HI, followed by ligation to yield pBM157. The DNA sequence of the mutant *CYC1-157* within pBM157 in the region of the deletion is illustrated in Fig. 1a.

Synthesis of the 54-mer oligonucleotides. Both A and B strands (BM54A and BM54B) of the 54-mer oligonucleotide were synthesized on an Applied Biosystems model 380A DNA synthesizer. Strand A is homologous to *CYC1* sequences except for the addition of a *Pst*I sticky end at the 3' end (Fig. 1b). The B strand of the 54-mer oligonucleotide is complementary to the A strand, which has a *Pst*I sticky end at its 3' end. At residues −43 to −49 (position of residues from the ATG codon of the iso-1-cytochrome *c*) of the B strand of the 54-mer oligonucleotide, a mixed synthesis was

* Corresponding author.
† Present address: Department of Microbiology, FitzGerald Building, University of Toronto, Toronto, Ontario M5S 1A8, Canada.

performed with a ratio of 0.71 wild-type base to 0.097 of each of the other three bases. This ratio is that required to produce, on average, 2 bp changes over the −43 to −49 region (7 bp). To determine the proper fraction of wild-type phosphoramidite, the number of residues that remained that of the wild type (5) were divided by the total number of residues over the target region (7). Accordingly, the fraction of phosphoramidite that should be wild type is 0.71, and the fraction to constitute the other phosphoramidite is 0.29. To illustrate that these numbers are correct the binomial distribution equation was used

$$P(p) = n!/[(n - r)! \times (r!)] \times p^{n-r} \times (1 - p)^r \quad (1)$$

In the present context $P(p)$, the probability of observing $n - r$ of n item, yields the fraction of retrieved plasmids with 2-bp changes (r) over the 7-bp region (n). To determine the fraction of wild-type phosphoramidite (p) during the synthesis of each residue, it is necessary to determine the maximum valve of $P(p)$ for an average of 2 bp changes ($r = 2$). This is accomplished by taking the first derivative of $P(p)$ with respect to $p(dP/dp)$ and then setting the resulting expression equal to zero and solving for p. The first derivative of equation 1, after substituting coefficients n and r for 7 and 2, respectively, is

$$dP/dp = -14p^5(1 - p) + 35p^4(1 - p)^2 \quad (2)$$

After simplifying and setting equation 2 equal to zero, the equation becomes

$$0 = 7p^4(1 - p)(5 - 7p) \quad (3)$$

The values of p that set the right side equal to zero are 0, 1, and 5/7. p equal to 0 and 1 are minimums; however, p equal to 5/7 is a maximum. The fraction of wild-type phosphoramidite that yields an average of 2 bp changes is 5/7, and the fraction of the other three phosphoramidites is correspondingly 2/7. These values are identical to those determined by dividing the number of residues which on average remain wild type by the number of residues over the target region, shown above. Substituting 0.71 and 0.29 for p and $1 - p$, respectively, into the binomial distribution equation the expected mutation yields are 9, 26, and 32% for 0 bp ($n = 7$, $r = 0$), 1 bp ($n = 6$, $r = 1$), and 2 bp ($n = 5$, $r = 2$) changes, respectively. However, since only one DNA strand contains mutants, 55% of the retrieved plasmids should be wild type over the 7-bp region, and 29% should have 1 or 2 bp changes.

Construction and isolation of *CYC1* point mutants. The 5′-unphosphorylated A and B strands of the 54-mer oligonucleotide (BM54A and BM54B) were mixed in equal molar ratios and ligated with a *Pst*I digest of pBM157 (Fig. 1b). The ligation mixture was heated at 65°C for 20 min to inactivate the ligase, NaCl was added to a final concentration of 50 mM, and the mixture was digested with *Pst*I to linearize any recircularized pBM157. This DNA was used to transform *E. coli* JM101 (19) to ampicillin resistance. Half of the cells were plated to determine the transformation efficiency, and the other half were used to innoculate a 10-ml culture. Plasmid DNA was isolated from the 10-ml culture and used to transform competent *E. coli* JM101. Single-stranded DNA was isolated from 1.5-ml cultures of single-colony transformants after phage superinfection as described previously (5). Dot blot hybridization was carried out (31) on single-stranded plasmid DNA, with the ^{32}P-labeled B strand of the 54-mer oligonucleotide used as a probe to identify plasmids with the double-stranded oligonucleotide in the desired orientation. Those plasmids identified as such were subjected to sequencing by the method of Sanger et al. (24). The *Bam*HI-*Hin*dIII fragment containing each mutant *CYC1* gene was transferred from the pEMBL vector to the yeast vector YEp13 (4), and the latter was used to transform *S. cerevisiae* GM-3C-2 (7) which lacks a functional *CYC1* locus.

Isolation of total *S. cerevisiae* RNA. RNA was isolated as described by A. Spence and M. Smith (manuscript in preparation). A 50-ml culture of *S. cerevisiae* transformed with a YEp13 derivative, which includes a mutant *CYC1* gene, were grown to a cell density of 2×10^7 cells per ml under selective conditions. The cells were collected by centrifugation and suspended in 0.5 ml of 80 mM Tris hydrochloride (pH 7.5)–10 mM $CaCl_2$–10 mM β-mercaptoethanol–10 mM vanadyl ribonucleoside complex (Bethesda Research Laboratories, Inc., Gaithersburg, Md.). Acid-washed and sterilized glass beads were added (0.5 ml). The mixture was kept at 0°C between 10-s periods of vortexing and was continued until lysis was complete. The extract was removed to an Eppendorf tube and centrifuged for 5 min, and the supernatant was recovered. Sodium dodecyl sulfate, vanadyl ribonucleoside complex, and proteinase K (Boehringer GmbH, Mannheim, Federal Republic of Germany) were added to a final concentration of 0.5%, 10 mM, and 0.6 mg/ml, respectively. The mixture was incubated at 37°C for 1 h, and then 3 M sodium acetate was added to a final concentration of 0.3 M, and the RNA was precipitated with 2 volumes of 95% ethanol. The RNA pellet was rinsed with 70% ethanol and suspended in 0.4 ml of 10 mM EDTA, and then 0.1 ml of 10 M LiCl was added and the mixture was kept overnight at 0°C. The RNA precipitate was collected by centrifugation and suspended in 10 mM EDTA, and a second ethanol precipitation and rinse was performed. The pellet was suspended in water, and the RNA concentration was determined (A_{260}).

cDNA synthesis. Total yeast RNA (50 μg) was added to 2 pmol of 5′-end, ^{32}P-labeled (31) P6 (5′-CTTAGCAGAACCAACCTTGA-3′) oligonucleotide (specific activity, ~10^6 cpm/pmol). The mixture was ethanol precipitated, and the pellet was washed with 70% ethanol. The pellet was dried and suspended in 10 μl of primer extension buffer (50 mM Tris hydrochloride [pH 8.0], 0.4 M NaCl, 1 mM EDTA) and left at room temperature for 1 h. The reaction mixture was adjusted to 100 mM Tris hydrochloride (pH 8.3), 10 mM $MgCl_2$, 50 mM KCl, 10 mM dithiothreitol, 100 μg of actinomycin D (Sigma Chemical Co., St. Louis, Mo.) per ml, of each of the four deoxyribonucleotide triphosphates per ml to a final volume of 25 μl. After the addition of 10 U of avian myeloblastosis virus reverse transcriptase, the mixture was incubated for 1 h at 42°C. The cDNA products were ethanol precipitated and analyzed on polyacrylamide sequencing gels.

Recombinant DNA techniques. *S. cerevisiae* transformations (18), the isolation of plasmid DNA from *E. coli* (2), and DNA sequencing (24) were carried out as described previously.

RESULTS

The *CYC1* transcription start sites have been mapped by both S1 nuclease digestion of the mRNA-DNA heteroduplex (7) and by primer extension (J. B. McNeil and M. Smith, J. Mol. Biol., in press). There were a large number of mRNA start sites spaced over a 100-bp region, and the relative efficiency of transcription initiation was variable (Fig. 1). There was no obvious consensus sequence surrounding these start sites; however, the mRNA 5′ ends appeared to begin with a purine residue, usually an A residue. This study

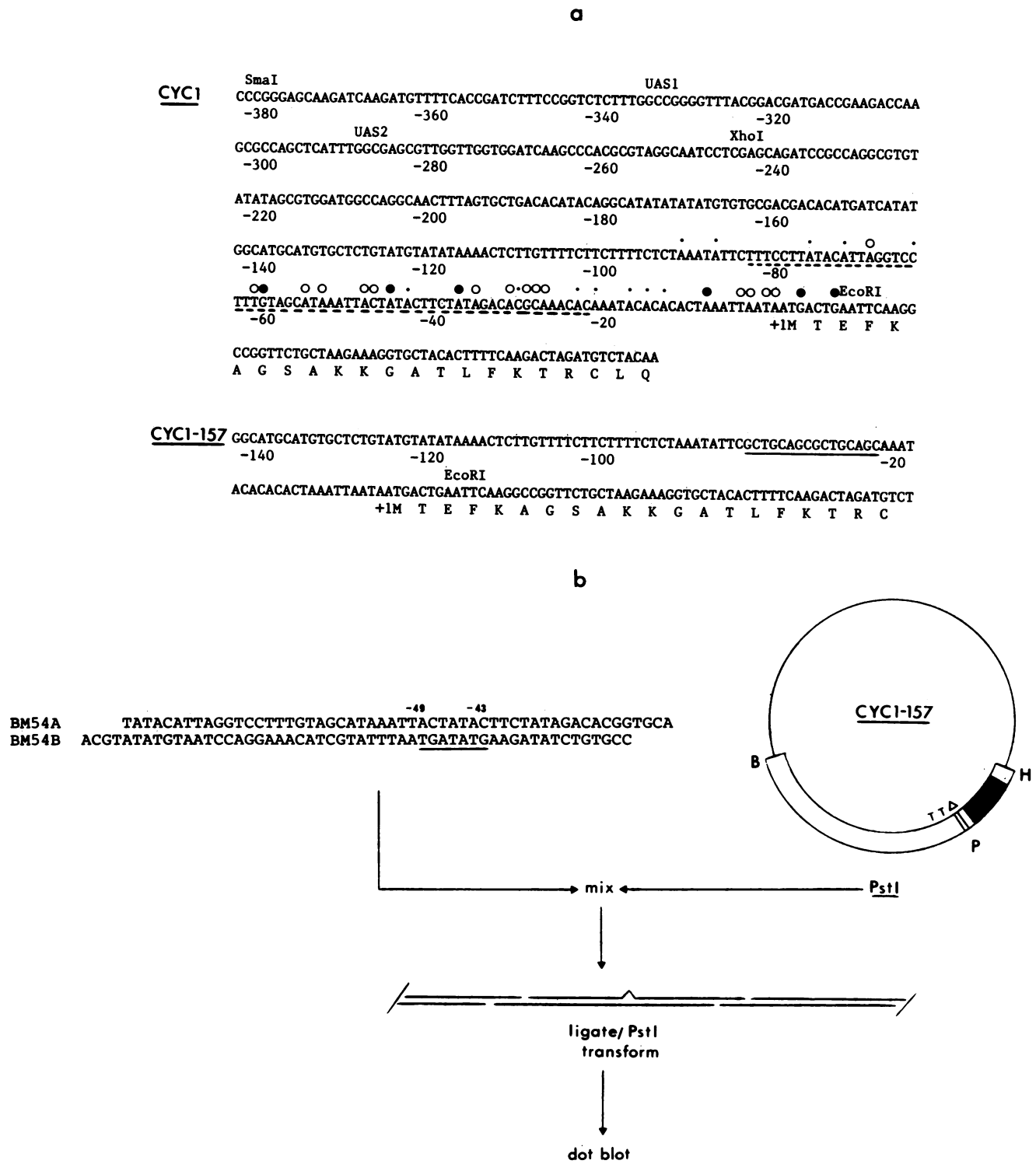

FIG. 1. (a) Schematic representation of the DNA sequence of the 5′-noncoding region of the *CYC1* and *CYC1-157* genes. The +1 position is the A of the *CYC1* ATG codon. The dashed underlined sequence of the *CYC1* gene was deleted to yield *CYC1-157*, as described in the text. The underlined sequence indicates the presence of two *Pst*I linkers. ●, ○, and • indicate the sites of high, moderate, and low transcription initiation, respectively. (b) Procedure used to clone the 54-bp duplex into the *Pst*I site of *CYC1-157*. The DNA sequences of the A and B strands of the 54-mer oligonucleotide (BM54A and BM54B, respectively) are shown. The region of 54B (−43 to −49) subjected to mixed synthesis is underlined. 54A and 54B were allowed to anneal to yield a double-stranded 54-mer oligonucleotide with *Pst*I sticky ends. The 54-mer was cloned into the *Pst*I site of *CYC1-157* as described in the text. The DNA mixture was used to transform competent strain JM101, and single-stranded DNA was isolated and characterized. T indicates the positions of TATA sequences and the deletion (Δ) with *Pst*I linker inserts. The coding region of *CYC1* is the filled-in box. Abbreviations: H, *Hin*dIII; B, *Bam*HI; P, *Pst*I.

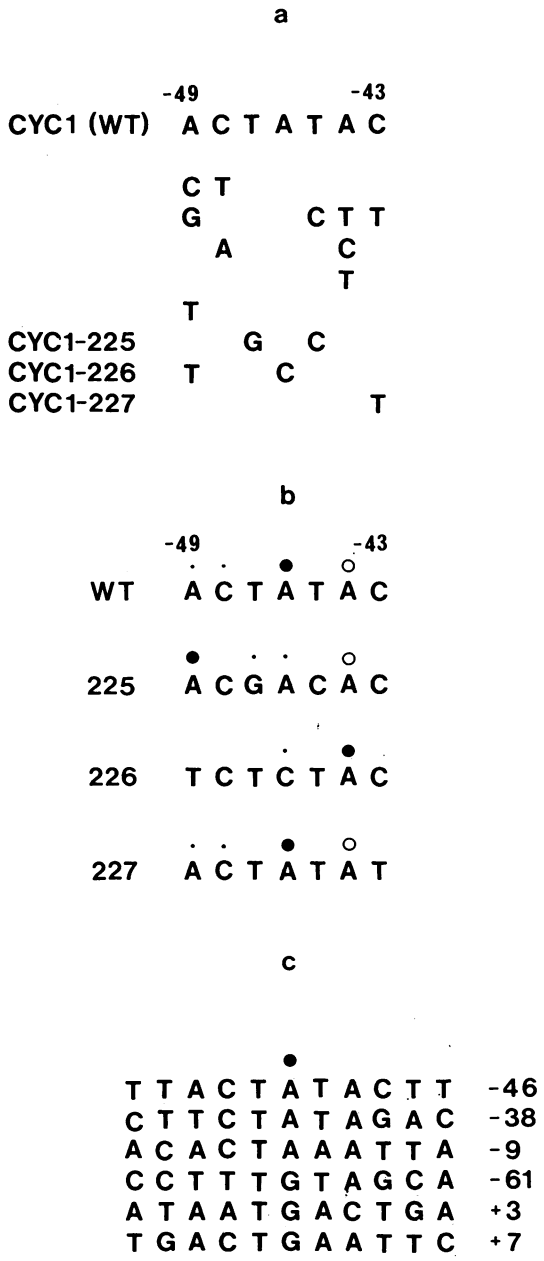

FIG. 2. (a) DNA sequence of the −43 to −49 region of the sense strand of the wild-type (WT) *CYC1* gene and for 8 of the 15 retrieved *CYC1* genes that possess changes over the region. Variant nucleotides over this region are indicated. *CYC1-225*, *CYC1-226*, and *CYC1-227* were selected for further study. The sites of *CYC1* transcription initiation are for the various *CYC1* mutants (*CYC1-224* to *CYC1-227*) (b) and the six most efficient *CYC1* start sites (●), (c). The numbers indicate the position of the start site in the *CYC1* gene. Other symbols are as described in the legend to Fig. 1.

examines the effects that base pair changes at and surrounding *CYC1* transcription start sites have on the initiation of transcription and the efficiency of initiation at these sites.

Synthesis of the mutagenic double-stranded 54-mer. The target region for the site-directed mutagenesis was a 7-bp region from −43 to −49 (+1 is the A of the *CYC1* ATG codon which codes for the translation start codon) of the *CYC1* 5′-noncoding region. This region was chosen because it includes one of the most efficient transcription initiation sites at −46 (Fig. 1; see Fig. 3, lanes 1 and 7) and several other start sites. If the DNA sequence at and around a transcription start site does effect initiation, then base pair changes within this region should exert an observable effect.

The chemical synthesis of an oligonucleotide by the phosphoramidite method (1) with a degeneracy over a 7-bp region requires the appropriate ratio of wild-type to mutant deoxynucleotide phosphoramidite to maximize the percent yield of an average number of specified base pair changes. An average of 2 bp changes was chosen for further examination, and the proper fraction of wild-type and mutant phosphoramidite used during synthesis at each of the residues over the −43 to −49 target region was determined as described above.

Isolation of *CYC1* point mutants. pBM157 (constructed as described above) includes a 58-bp deletion from −24 to −86 with a tandem *Pst*I linker insert. The deleted region includes many of the *CYC1* mRNA start sites (Fig. 1). Point mutants within the *CYC1* transcription initiation sites were created by ligating the double-stranded duplex of 54A and 54B into the *Pst*I site of pBM157 as described above (Fig. 1). Of the 96 DNA preparations isolated, 41 contained the insert in the correct orientation (data not shown).

Sequencing of the −43 to −49 region for these DNAs was undertaken by the method of Sanger et al. (24). Figure 2 illustrates the nature of the base pair changes retrieved. A total of 15 clones were sequenced, and 8 possessed changes.

5′-Mapping of the wild-type and mutant *CYC1* mRNAs. *CYC1-225*, *CYC1-226*, and *CYC1-227* were chosen for further study, because they involve changes at two transcription initiation sites (−46 and −49, *CYC1-226* and at the +1 and −1 positions with respect to the −46 position (*CYC1-225*). *CYC1-227* possesses a C to T change at −43. Two other plasmids were constructed: *CYC1-224*, which possesses the wild-type sequence at residues −43 to −49, and *CYC1-223*, which includes the wild-type 54-mer duplex cloned in the reverse orientation (see Fig. 4). These various *CYC1* mutant and wild-type genes were cloned into the *S. cerevisiae* shuttle vector YEp13 (4) and used to transform the *CYC1* deletion and *CYC7* point mutant strain GM-3C-2 (7). This strain is unable to grow on media which includes glycerol (or any nonfermentable carbon source) as the only carbon source. The ability of transformed GM-3C-2 to grow on a nonfermentable carbon source would indicate the presence of a functional *CYC1* gene on the transforming plasmid.

The 5′ ends of the mutant and wild-type *CYC1* mRNAs were mapped by primer extension as described above (Fig. 3). The wild-type transcription pattern is shown in lanes 1 and 7. Adjacent to lane 7 is the *CYC1* sequence subjected to the site-directed mutagenesis. Lane 3 illustrates the initiation pattern of *CYC1-224*-transformed GM-3C-2. The DNA sequence of *CYC1-224* is the same as that of the wild type, except at the junction between the 54-mer duplex and *CYC1* sequence. These changes (−83 to −78 and −28 to −23) appear to significantly alter the start site pattern within these regions and yield more subtle changes in transcription efficiencies at the wild-type start sites. Because of this modified pattern of *CYC1-224*, the analysis of *CYC1* mutants over the −43 to −49 region involves a comparison of their transcription initiation pattern with *CYC1-224* (Fig. 3).

CYC1-225 includes a T to C and T to G change at −45 and −47, respectively. These changes nearly eliminate transcription initiation at the −46 A nucleotide (Fig. 3, lane 4). A corresponding increase in transcription occurs at −49. The transversion mutation at −46 to a C residue and at −49 to a

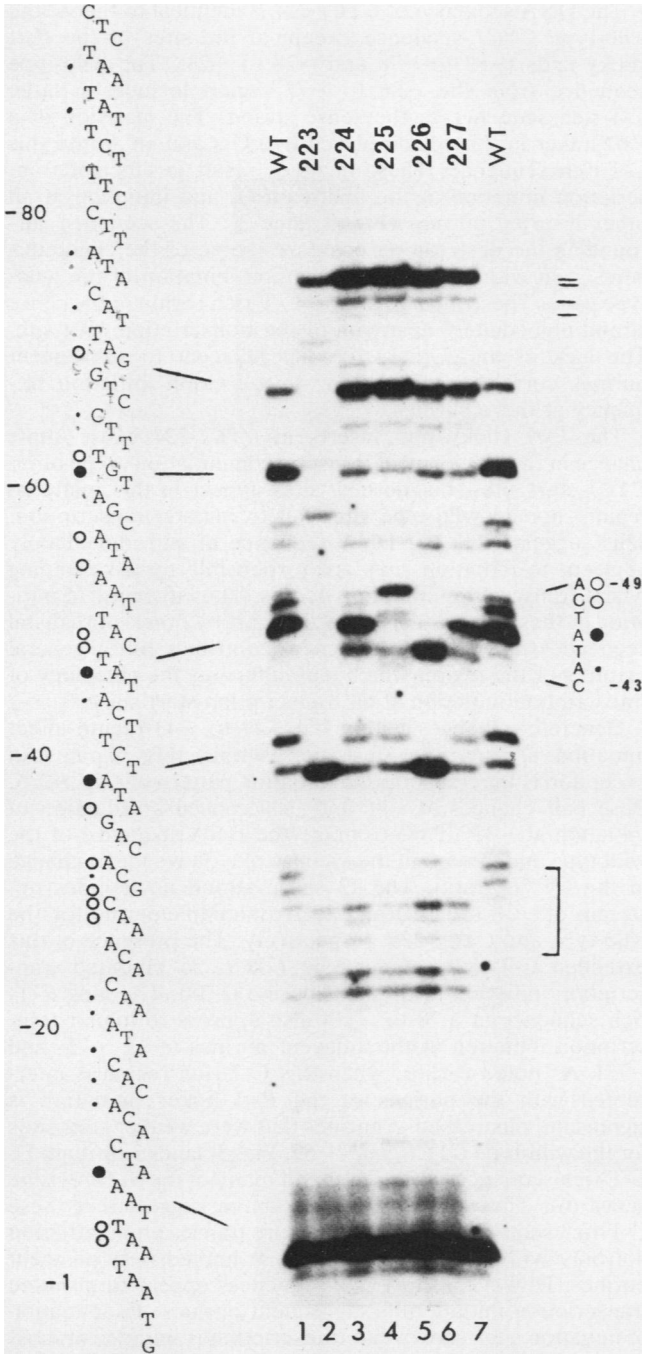

FIG. 3. Mapping of the 5′ termini of the *CYC1* wild type and various *CYC1* mutants by primer extension. Oligonucleotide P6 was annealed to total RNA isolated from GM-3C-2 transformed with *CYC1-223* (lane 2), *CYC1-224* (lane 3), *CYC1-225* (lane 4), *CYC1-226* (lane 5), *CYC1-227* (lane 6), and CYC1 (lanes 1 and 7) to prime cDNA synthesis. The cDNA products were electrophoresed on an 8% polyacrylamide sequencing gel. The 7-bp −43 to −49 region and the corresponding transcription initiation pattern are illustrated. The sites of sequence variability between the wild type (WT) and *CYC1-224* is indicated by a bracket over the −23 to −28 region, and changes over the −79 to −84 region are indicated by dashes. Other symbols are as described in the legend to Fig. 1.

T residue drastically reduces initiation at the −46 and −49 positions (*CYC1-226*; Fig 3, lane 5) with a corresponding increase at the −44 position. The increase of transcription initiation may be due to both a change of DNA sequence that promotes transcription initiation and a reduction of initiation at adjacent sites. The transition mutation of *CYC1-227* (T to C) at position −43 has little effect on transcription initiation (Fig. 3, lane 6). A summary of the transcription initiation pattern is illustrated in Fig. 4.

DISCUSSION

Double-stranded oligonucleotide-directed mutagenesis was used to isolate derivatives of the *CYC1* gene with changes within the transcription start site region. The region of interest is between −43 and −49. This 7-bp region was chosen because it is responsible for four different *CYC1* 5′ mRNA ends, including one of the major start sites. If the DNA sequence at or immediately adjacent to a start site does influence transcription initiation, a change anywhere within this region should exert some effect. The introduction of changes over this region and examination of the corresponding effects that these changes have on transcription initiation should allow the sequence prerequisites for efficient transcription initiation sites to be identified.

If one examines the DNA sequence surrounding the most efficient transcription start sites for the *CYC1* gene (Fig. 1 and 2c) a sequence bias is evident. The entire start site region is generally A-T rich. Transcription of *CYC1* mRNA can be initiated with any base; however, the most efficient starts align with a purine and, in particular, with an A nucleotide. The A nucleotide may be the favored purine because of a significant lack of G residues in the sense strand. The DNA sequence (sense strand) immediately 5′ to an efficient *CYC1* start site is predominantly C-T rich with very few purines. The sequence immediately 3′ to an efficient *CYC1* start site is A-T rich, but beyond the second nucleotide, any base may be present.

The mutations selected for examination were *CYC1-225*, which involves the replacement of a T at the −47 position with a G, disrupting the C-T-rich sequence immediately 5′ to the transcription start site at −46; *CYC1-226*, which includes a transversion at the site of initiation (−46) from an A to a C; and *CYC1-227*, which is altered from wild-type sequences by a transition at −43 from a C to a T. The *CYC1* mutants were cloned into YEp13 and used to transform the *CYC1* deletion strain GM-3C-2. Primer extension was carried out to map the 5′ termini of the *CYC1* mRNAs (Fig. 3).

A reduction of transcription initiation is indicated at −46 and −49 when these start sites are changed to a C and a T, respectively (*CYC1-226*; Fig. 3, lane 5). Based on analysis of these two changes and examination of the DNA sequence at the sites of efficient *CYC1* transcription initiation (Fig. 2c), it is evident that purines are the preferred site of transcription initiation for *CYC1*. Not all purines within the *CYC1* start site region are sites of initiation. The DNA context determines the potential for initiation at adjacent purines. Examination of the DNA sequence adjacent to efficient *CYC1* initiation sites indicates the absence of A's and G's immediately 5′ of transcription start sites and the presence of short pyrimidine-rich region 5′ of efficient start sites. The results for *CYC1-225* indicate that a purine (G) at −47 significantly reduces transcription initiation at −46 (Fig. 3, lane 4). The presence of C residues immediately 3′ of A starts (−49, −44, −32, −24, −18, −16, and −14) suggests that the T to C change in *CYC1-225* is not responsible for the reduction of initiation at −46. These observations suggest that the pres-

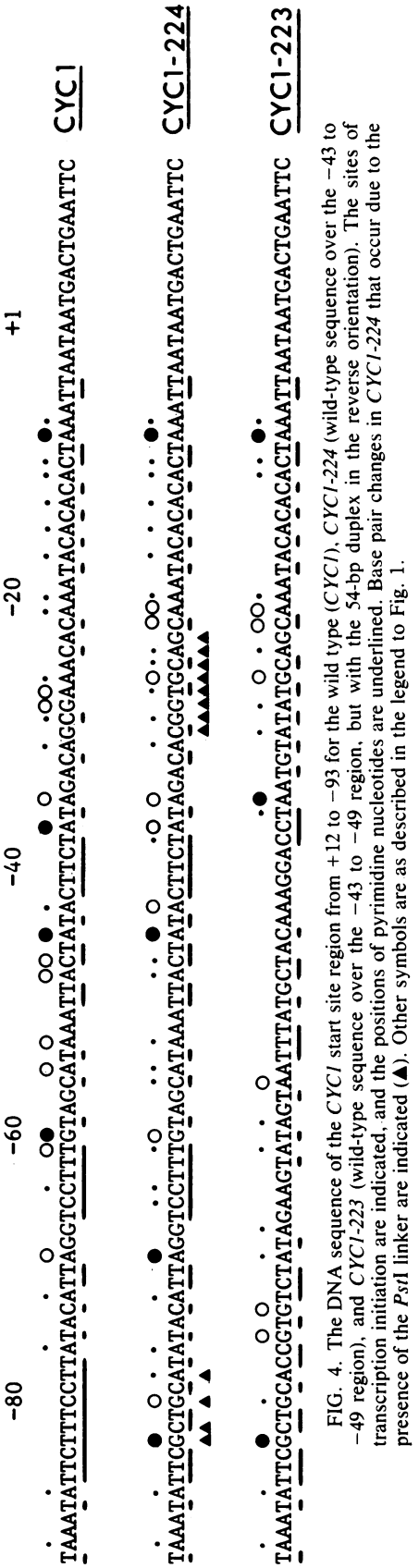

FIG. 4. The DNA sequence of the *CYC1* start site region from +12 to −93 for the wild type (*CYC1*), *CYC1-224* (wild-type sequence over the −43 to −49 region), and *CYC1-223* (wild-type sequence over the −43 to −49 region, but with the 54-bp duplex in the reverse orientation). The sites of transcription initiation are indicated, and the positions of pyrimidine nucleotides are underlined. Base pair changes in *CYC1-224* that occur due to the presence of the *Pst*I linker are indicated (▲). Other symbols are as described in the legend to Fig. 1.

ence of a pyrimidine immediately 5′ of a purine start site is essential for transcription initiation.

The DNA sequence of *CYC1-224* is identical to that of the wild-type *CYC1* sequence except at the sites of the *Pst*I sticky ends (−83 to −78 and −28 to −23). The wild-type sequence from the −86 to −77 region includes a large C-T-rich sequence in the sense strand. The insertion of a *Pst*I linker in this region placed two G's and an A into this C-T-rich sequence. These changes result in efficient transcription initiation at the upstream G, and initiation at all other inserted purines (Fig. 3, lane 3). The sequence surrounding the upstream G possesses some of the characteristics seen with efficient transcription initiation of the wild-type gene. These include a short C-T-rich region in the sense strand immediately upstream of the transcription start site. The lack of these characteristics adjacent to subsequent purines correlates with a low transcription initiation frequency at these purines.

The *Pst*I sticky end inserts in *CYC1-224* cause some change in the efficiency of transcription initiation at the other *CYC1* start sites, but no new sites appear in the unaltered region, nor do wild-type sites fail to initiate transcription. This suggests that the DNA sequence at and immediately adjacent to initiation sites are responsible for determining where transcription initiation occurs. The efficiency of initiation at these sites is determined in part by other more distal sequences. These distal sequences contribute to the general features of the region which may influence the efficiency of transcription initiation at all transcription start sites.

Generally, changes within the −49 to −43 region affect initiation at only the start sites within this region. An exception is the transcription initiation pattern of *CYC1-226*. Base pair changes at −46 and −49 caused a reduction of initiation at −38. If we examine the DNA sequence of the wild type and mutant in the vicinity of −38 we find a change in the CT/GA ratio. The 13 sense strand nucleotides upstream of −38 are 10/3 and 12/1 pyrimidine/purine for the wild type and *CYC1-226*, respectively. The presence of this extended C-T-rich sequence in *CYC1-226* inhibited transcription initiation at the next purine (−38). The long C-T-rich sequence at −78 to −114 also appears to inhibit transcription initiation at the adjacent purines (−77, −75, and −73). As noted earlier, when this C-T-rich region is interrupted with the purines of the *Pst*I linker, initiation is significantly increased at purines that were weaker start sites for the wild-type *CYC1* gene (−69; Fig. 3, lanes 3 through 7). C-T-rich sequences are a feature of many of the *S. cerevisiae* genes (6). The results discussed above suggest that these C-T-rich sequences inhibit premature transcription initiation not only within this sequence but at immediately adjacent purines. However, short C-T sequences appear to promote transcription initiation at subsequent purines. Examination of initiation sites shows that transcription is initiated at most purines that follow a pyrimidine (Fig. 4). When the 54-bp duplex is inserted in the reverse orientation, it yields a completely new DNA sequence over the transcription initiation region (*CYC1-223*; Fig. 3, lane 2), and, as expected, a new transcription initiation pattern is observed. Initiation does occur at the expected sites. All but one start site (−37) that appear in *CYC1-223* begin at a purine in the sense strand. Immediately preceding each of these start sites is a pyrimidine. Transcription initiation at unaltered DNA sequences (>−23 and <−84) shows the same initiation pattern as that observed with wild-type *CYC1* (Fig. 4).

The mechanism of transcription initiation of the *S. cerevisiae CYC1* gene is in some respects atypical of initiation of

higher eucaryotic genes. Transcription of higher eucaryotic cellular and viral genes involves initiation at a small number of 5' termini (3). Studies indicate that TATA sequences are required for proper transcription initiation 25 to 30 bp downstream from the TATA sequence (3). Transcription of the *CYC1* gene takes place over a 100-bp region. While the general location of these starts is determined by TATA sequences (McNeil and Smith, in press), their precise locations are determined by the DNA sequence at and immediately surrounding the sites of initiation. The mRNA start sites of the *CYC1* gene have some features in common with transcription start sites in higher eucaryotic systems. The function of DNA sequences over the simian virus 40 late RNA transcription start site region has been examined, and results of studies indicate that 5' ends of the approximately two dozen late RNAs of simian virus 40 are each determined by sequences within the immediate vicinity of the start site (22, 28). Transcription of a majority of higher eucaryotic genes examined indicate that mRNA 5' ends align at an A residue with adjacent pyrimidine-rich sequences 3' and 5' to the start sites (3). This is similar to the sequence at the site of an efficient *CYC1* start site.

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council of Canada and the National Cancer Institute of Canada. M.S. is a Career Investigator of the Medical Research Council of Canada. J.B.M. was a recipient of a National Sciences and Engineering Research Council Postdoctorial Fellowship.

The synthesis of oligonucleotides by T. Atkinson is gratefully acknowledged.

LITERATURE CITED

1. **Adams, S. P., K. S. Kavka, E. J. Wykes, S. B. Holder, and G. R. Galluppi.** 1983. Hindered dialkylamino nucleoside phosphite reagents in the synthesis of two DNA 51-mers. J. Am. Chem. Soc. **105**:661–663.
2. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
3. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
4. **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.
5. **Dente, L., G. Cesareni, and R. Cortese.** 1983. pEMBL: a new family of single stranded plasmids. Nucleic Acids Res. **11**:1645–1655.
6. **Dobson, M. J., M. F. Tuite, N. A. Roberts, A. J. Kingsman, and S. M. Kingsman.** 1982. Conservation of high efficiency promoter sequences in *Saccharomyces cerevisiae*. Nucleic Acids Res. **10**:2625–2637.
7. **Faye, G., D. W. Leung, K. Tatchell, B. D. Hall, and M. Smith.** 1981. Deletion mapping of sequences essential for *in vivo* transcription of the iso-1-cytochrome *c* gene. Proc. Natl. Acad. Sci. USA **78**:2258–2262.
8. **Guarente, L., and E. Hoar.** 1984. Upstream activation sites of the *CYC1* gene of *Saccharomyces cerevisiae* are active when inverted but not when placed downstream of the "TATA box". Proc. Natl. Acad. Sci. USA **81**:7860–7864.
9. **Guarente, L., B. Lalonde, P. Gifford, and E. Alani.** 1984. Distinctly regulated tandem upstream activation sites mediate catabolite repression of the *CYC1* gene of *S. cerevisiae*. Cell **36**:503–511.
10. **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.
11. **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.
12. **Guarente, L., R. R. Yocum, and P. Gifford.** 1982. A *GAL10-CYC1* hybrid yeast promoter identifies the *GAL4* regulatory region as an upstream site. Proc. Natl. Acad. Sci. USA **79**:7410–7414.
13. **Gudenus, R., A. Spence, A. Hartig, M. Smith, and H. Ruis.** 1984. Regulation of transcription of the *Saccharomyces cerevisiae CYC1* gene: identification of a DNA region involved in heme control. Curr. Genet. **8**:45–48.
14. **Hsu, Y.-P., and P. Schimmel.** 1984. Yeast *LEU1*: repression of mRNA levels by leucine and relationship of 5'-noncoding region to that of *LEU2*. J. Biol. Chem. **259**:3714–3719.
15. **Johnston, M., and R. W. Davis.** 1984. Sequences that regulate the divergent *GAL1-GAL10* promoter in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **4**:1440–1448.
16. **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.
17. **McKnight, S. L., and R. Kingsbury.** 1982. Transcription control signals of a eukaryotic protein-coding gene. Science **217**:316–324.
18. **McNeil, J. B., R. K. Storms, and J. D. Friesen.** 1981. High frequency recombination and the expression of genes cloned on chimeric yeast plasmids: identification of a fragment of 2-μm circle essential for transformation. Curr. Genet. **2**:17–25.
19. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system of shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
20. **Orkin, S. H., J. P. Sexton, T. Cheng, S. C. Goff, P. J. V. Giardina, J. I. Lee, and H. H. Kazazian, Jr.** 1983. ATA box transcription mutation in β-thalassemia. Nucleic Acids Res. **11**:4727–4734.
21. **Osborne, T. F., R. B. Gaynor, and A. J. Berk.** 1982. The TATA homology and the mRNA 5' untranslated sequence are not required for expression of essential adenovirus E1A functions. Cell **29**:139–148.
22. **Piatak, M., P. K. Ghosh, L. C. Norkin, and S. M. Weissman.** 1983. Sequences locating the 5' ends of the major simian virus 40 late mRNA forms. J. Virol. **48**:503–520.
23. **Rymond, B. C., R. S. Zitomer, D. Schumperli, and M. Rosenberg.** 1983. The expression in yeast of the *Escherichia coli galK* gene on *CYC1::galK* fusion plasmids. Gene **25**:249–262.
24. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
25. **Schulz, T. Z., S. Y. Tsai, K. Itakura, X. Soberon, R. B. Wallace, M.-J. Tsai, S. L. C. Woo, and B. W. O'Malley.** 1982. Point mutagenesis of the ovalbumin gene promoter sequence and its effect on *in vitro* transcription. J. Biol. Chem. **257**:11070–11077.
26. **Sentenac, A., and B. Hall.** 1982. Yeast nuclear RNA polymerases and their role in transcription, p. 561–606. *In* J. N. Strathern, E. W. Jones, and J. R. Broach (ed.), The molecular biology of the yeast Saccharomyces: metabolism and gene expression. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Siliciano, P., and K. Tatchell.** 1984. Transcription and regulatory signals at the mating type locus in yeast. Cell **37**:969–978.
28. **Somasekhar, M. B., and J. Mertz.** 1985. Sequences involved in determining the locations of the 5' and of the late RNAs of simian virus 40. J. Virol. **56**:1002–1013.
29. **Struhl, K.** 1983. Promoter elements, regulatory elements, and chromatin structure of the yeast *his3* gene. Cold Spring Harbor Symp. Quant. Biol. **47**:901–910.
30. **Zalkin, H., J. L. Paluh, M. van Cleemput, W. S. Moye, and C. Yanofsky.** 1984. Nucleotide sequence of *Saccharomyces cerevisiae* genes *TRP2* and *TRP3* encoding bifunctional anthranilate synthase: indole-3-glycerol phosphate synthase. J. Biol. Chem. **259**:3985–3992.
31. **Zoller, M. J., and M. Smith.** 1982. Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations on any fragment of DNA. Nucleic Acids Res. **10**:6487–6500.