MOLECULAR AND CELLULAR BIOLOGY, July 1988, p. 2964–2975
0270-7306/88/072964-12$02.00/0

Vol. 8, No. 7

Mutational Analysis of the *HIS4* Translational Initiator Region in *Saccharomyces cerevisiae*

A. MARK CIGAN, EDWARD K. PABICH, AND THOMAS F. DONAHUE*

Department of Molecular Biology, Northwestern University Medical School, 303 East Chicago Avenue, Chicago, Illinois 60611

Received 19 February 1988/Accepted 12 April 1988

We have mutated various features of the 5′ noncoding region of the *HIS4* mRNA in light of established *Saccharomyces cerevisiae* and mammalian consensus translational initiator regions. Our analysis indicates that insertion mutations that introduce G+C-rich sequences in the leader, particularly those that result in stable stem-loop structures in the 5′ noncoding region of the *HIS4* message, severely affect translation initiation. Mutations that alter the length of the *HIS4* leader from 115 to 39 nucleotides had no effect on expression, and sequence context changes both 5′ and 3′ to the *HIS4* AUG start codon resulted in no more than a twofold decrease of expression. Changing the normal context at *HIS4* 5′-AAUAAUGG-3′ to the optimal sequence context proposed for mammalian initiator regions 5′-CACCAUGG-3′ did not result in stimulation of *HIS4* expression. These studies, in conjunction with comparative and genetic studies in *S. cerevisiae*, support a general mechanism of initiation of protein synthesis as proposed by the ribosomal scanning model.

Comparative analysis of 131 *Saccharomyces cerevisiae* genes (7) indicates that basic sequence and structural features associated with yeast translational initiator regions parallel those features associated with translational initiator regions from higher eucaryotic organisms that are proposed to initiate translation by a ribosomal scanning mechanism (16, 17). Namely, the first AUG codon nearest the 5′ end of the message serves as the start site of translation at 95% of yeast genes, 70% have leader regions with distances in the range of 20 to 80 nucleotides having an average length of 52 nucleotides, and sequence context both 5′ and 3′ to the AUG start codon is biased in base composition. However, yeast initiator regions do differ from mammalian initiator regions. Specifically, the base composition of yeast initiator regions is biased toward A nucleotides (46% A, 21% C, 21% U, and 12% G; from position −25 to −1 relative to the A of AUG) and sequences immediately 3′ to the AUG reflect a nonrandom distribution of amino acids at the +2 amino acid position, as well as the preferred codon usage observed at yeast genes (7). These properties of yeast initiator regions are reflected in its consensus, 5′-$\hat{A}_{Y}\hat{A}_{U}$AAUGUCU-3′, which contrasts the mammalian consensus, 5′-CCACCAUGG-3′ (20), with the exception of a similar high bias for A nucleotides at the −3 position (75% and 79% for the yeast and mammalian consensus, respectively).

The importance of the higher eucaryotic consensus to the mechanism of ribosomal recognition of an AUG start codon has been suggested by mutational studies (15, 18, 21, 23, 25). Alteration of this optimal sequence context at the rat preproinsulin gene can result in 20-fold differences in translational efficiency at the AUG start codon. As a result of inefficient initiation, the ribosome can bypass this region to initiate at a subsequent AUG (23). The significance of these effects would appear to relate to a functional role for sequence context in establishing different rates of initiation at different genes and an ability to initiate at a second AUG that corresponds to the start site of the primary gene product in a subset of higher eucaryotic mRNAs. Genetic evidence at the *CYC1* (37) and *HIS4* (8) genes supports a mechanism of translation initiation in *S. cerevisiae* that is dependent on the first AUG codon nearest the 5′ end of the message, consistent with the ribosomal scanning model. However, these studies do not provide any insight into how yeast genes might arrive at alternative rates of initiation, nor do they explain how a similar subset of yeast genes is capable of bypassing upstream AUGs to initiate at subsequent downstream AUG codons (7). Reversion of initiator codon mutant alleles at *CYC1* shows that translation can be restored by the presence of an AUG codon at various positions within a 37-base-pair (bp) region, suggesting that flanking nucleotides are not important to efficient initiation at an AUG codon. Likewise, converse genetic studies at *HIS4* that directly selected for mutations that reduce or abolish ribosomal recognition of the normal initiator region only identified mutations in the AUG start codon, suggesting that simple base changes in the sequence context cannot afford ribosomal bypass of an AUG in *S. cerevisiae*. Therefore, despite parallels between features of yeast and mammalian initiator regions, functional differences may exist in the mechanism of initiation of protein synthesis.

To gain insight into signals that might mediate or control ribosomal recognition of the initiator region, we have altered the sequence context both 5′ and 3′ to the *HIS4* AUG start codon, addressing features common to the majority of yeast or mammalian initiator regions. The effects of these mutations on gene expression were quantitated in *HIS4-lacZ* fusion strains and as alleles of the *HIS4* locus to test the contribution of sequence context to the initiation process in *S. cerevisiae*. This analysis indicates that sequence context does not significantly contribute to recognition and utilization of the *HIS4* AUG start codon in *S. cerevisiae* and that simple base changes of these sequences do not support ribosomal bypass of the AUG, in agreement with genetic studies at *HIS4* described in a preceding report (8). In light of differences between yeast and mammalian initiator regions and the limited experimental evidence in *S. cerevisiae* that supports a ribosomal scanning mechanism, we have also altered structural features of the 5′ noncoding region of the *HIS4* message to gain a better mechanistic perspective of the initiation process in *S. cerevisiae*. One structural feature we

* Corresponding author.

have altered is the length of the *HIS4* leader to determine whether general variability of yeast leader lengths can affect the efficiency of initiation at an AUG codon. In addition, hairpin-loop structures were introduced in the 5′ noncoding region of the *HIS4* message, since similar studies at the herpes thymidine kinase (32) and rat preproinsulin (22) genes suggest that secondary structure inhibits the initial steps of the scanning process by blocking the ability of the ribosome either to bind a free 5′ end or to migrate toward the AUG. Our analysis demonstrates that changes in leader length have little or no effect on translation initiation, whereas secondary structure strongly inhibits expression, further indicating that *S. cerevisiae* uses a scanning-type mechanism for initiation of protein synthesis.

MATERIALS AND METHODS

DNA and RNA methods. Oligonucleotides were synthesized by the Northwestern University Biotechnology Facilities (Evanston, Ill.) and routinely purified by electrophoresing 2.5 optical density (260 nm) units through a 20% acrylamide gel (40 by 20 by 0.7 cm) containing 8 M urea. The DNA band corresponding to the full-length oligomer was identified by UV shadowing, extracted, crushed, and soaked overnight at 37°C in 1 ml of TE buffer (10 mM Tris hydrochloride [pH 7.4], 1 mM EDTA). The cleared supernatant was passed over 25 μl of TE-equilibrated DE-52 (Whatman, Inc., Clifton, N.J.), washed with 1 ml of double-distilled water, and eluted with 4 M ammonium acetate (225 μl). An equal volume of double-distilled water was added, and the sample was frozen and lyophilized. Typically, 1.0 to 1.5 optical density (260 nm) units was recovered.

The 5′ mapping of *HIS4* transcripts was determined by primer extension (30) by using [γ-^{32}P]ATP 5′-end-labeled oligonucleotides and Klenow polymerase. Total RNA was isolated from cells grown to log phase (2×10^8 cells per ml) in YEPD by the method of Carlson and Botstein (4). The 5′ start position of the *HIS4* wild-type transcript has been previously determined by both S1 nuclease and primer extension methods (10, 29) and is located 60 nucleotides 5′ from the AUG start codon (A of AUG is the +1 position).

The procedures used for preparation of plasmid DNA, restriction mapping, subcloning and Southern analysis, as well as bacterial strains used for transformation and bacteriophage isolation, have been described in previous reports (6, 10, 45).

Construction of *HIS4* leader regions with deletion, duplication, and insertion mutations. The length of the *HIS4* leader region was altered by deletion or duplication of DNA that corresponds to the 5′ noncoding portion of the *HIS4* message. To achieve these constructions, a series of deletions were generated in both the 5′-to-3′ and 3′-to-5′ directions with mapped endpoints in the *HIS4* leader region by using BAL 31 nuclease. The starting plasmid of these deletions, p176 (Fig. 1), contains the 1,578-bp *Sal*I fragment of the proximal *HIS4* region (10) inserted at the unique *Eco*RI site of plasmid B142. The 1,578-bp *Sal*I restriction DNA fragment contains the promoter region of *HIS4*, the 60-bp 5′ noncoding region of the *HIS4* message, and 474 bp of the amino terminus of the *HIS4* coding region. The complete DNA sequence of *HIS4* has been previously reported (10). Plasmid B142 is identical to the integrating vector YIp5 that contains the yeast selectable marker *URA3* (3), except that the pBR322-derived restriction site *Pvu*II has been deleted (9). Both *Sal*I ends of the *HIS4* fragment and the *Eco*RI site of B142 were filled in with reverse transcriptase, and the *Sal*I fragment was blunt-end ligated into the filled-in *Eco*RI site. As a result of this ligation event, no *Eco*RI sites are present in p176 (Fig. 1).

Deletions in the 5′-to-3′ direction were generated from the unique *Pvu*II restriction site located at position −588 from the 5′ mapped position of the *HIS4* transcript (Fig. 1). Deletions in the 3′-to-5′ direction were constructed from the unique *Xho*I site located at position +100 in the *HIS4* coding region (Fig. 1). Conditions for BAL 31 digestions and sizing the extent of deletions were identical to those previously described for the construction of BAL 31 deletions in the *HIS4* promoter region (9), with the exception that ligations were performed in the presence of the *Eco*RI DNA linker, 5′-GGAATTCC-3′. The endpoints of the deletions were determined by the Maxam and Gilbert DNA-sequencing method (27). The *Eco*RI restriction sites was 5′-end labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. A subsequent restriction with *Pvu*II or *Xho*I (Fig. 1) enabled the isolation of DNA fragments labeled at a unique end from the 3′-to-5′ and 5′-to-3′ deletion constructions, respectively. DNA sequences were resolved on a 20% polyacrylamide gel containing 8 M urea. By combining 5′-to-3′ and 3′-to-5′ deletion fragments through their common and unique *Eco*RI restriction sites, a deletion or duplication of the *HIS4* leader region with precise mapped endpoints was constructed on plasmids (Fig. 1).

The *HIS4* leader regions containing deletion or duplication mutations were also fused in frame with the *lacZ* coding region (Fig. 1). The *Sau*3A restriction fragment from each *HIS4* region was ligated into the *Bam*HI site of plasmid p596. This plasmid is identical to p349 previously described for other *HIS4-lacZ* fusion constructions (8), with the exception that the two *Eco*RI sites in p349 have been destroyed. The *Sau*3A restriction fragment from *HIS4* wild-type DNA, when fused to *lacZ*, allows expression of β-galactosidase in *S. cerevisiae*; this β-galactosidase expression is dependent on proper transcriptional and translational signals present in the upstream *HIS4* region (38).

The unique *Eco*RI restriction site in the *HIS4* leader region of these plasmid constructions served as the point of insertion of DNA synthesized as complementary oligonucleotides with *Eco*RI ends. The different *HIS4* leader constructions used, as well as the oligonucleotides inserted, are described in the text and legends for each specific experiment. Oligonucleotides were ligated into the *Eco*RI restriction site present in each respective p176- and p596-derived *HIS4* construction (Fig. 1). The presence and number of inserts in these *HIS4* plasmids were confirmed either by Maxam and Gilbert DNA sequencing or by the double-stranded chain termination method described by U.S. Biochemical Inc. For Maxam and Gilbert sequencing, the *Xho*I restriction site at position +100 in the *HIS4* coding region was end labeled and sequences were extended toward the *Pvu*II restriction site in the 5′ noncoding region of *HIS4*. For the chain termination method, a 17-nucleotide primer complementary to the upstream *HIS4* 5′ noncoding region (positions −116 to −102 from AUG) was used to synthesize DNA from the *HIS4* sense strand. For some of the larger oligonucleotides that introduce strong secondary structure in the *HIS4* 5′ noncoding region, the complete DNA sequence of the insert could not be fully resolved because of strong compressions and premature termination of the reaction in the region corresponding to the insert even when dITP was substituted for dGTP. However, by comparing the DNA sequences of these constructions with that of the original starting plasmid we could identify the presence of an insert

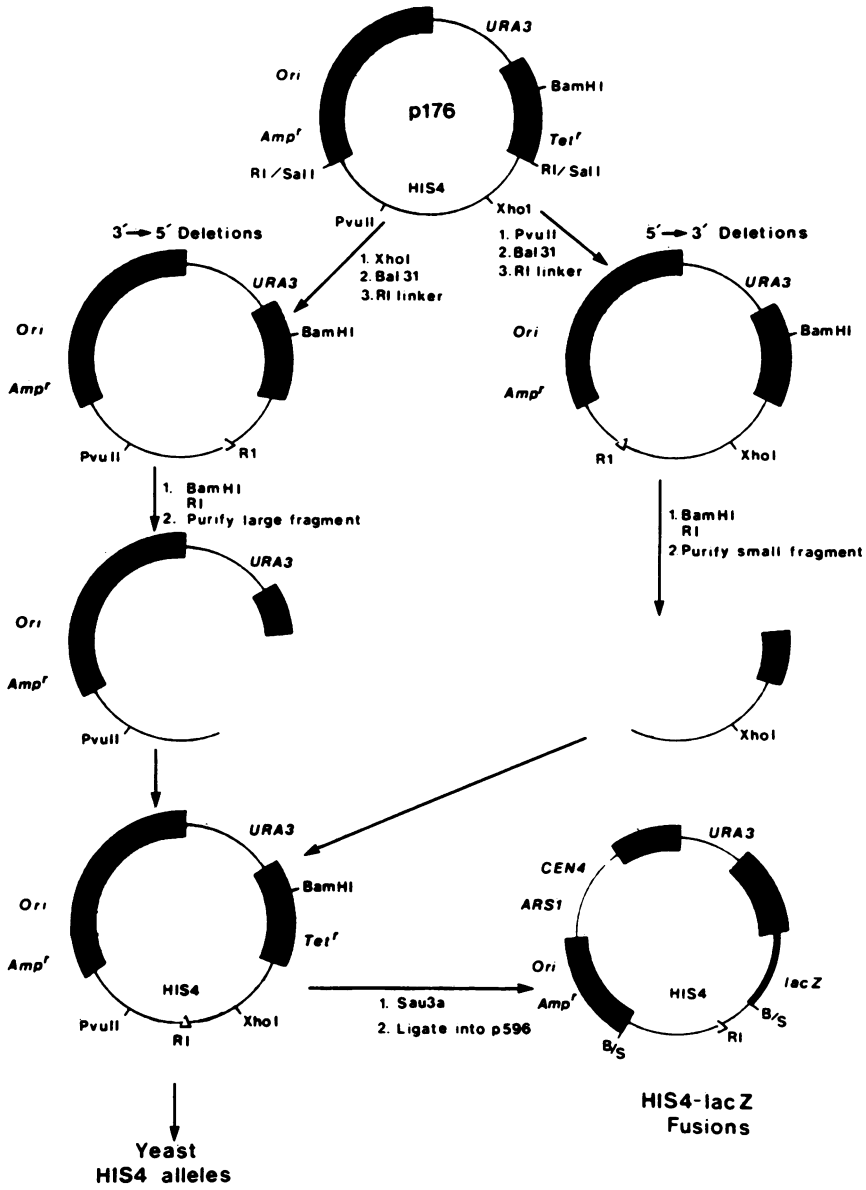


FIG. 1. Construction of plasmids containing deletions or duplications of *HIS4* leader sequences. Plasmid p176 contains the 1,578-bp *Sal*I fragment from the proximal region of *HIS4* blunt-end ligated into the *Eco*RI restriction site (RI) of plasmid B142 (8), a derivative of the integrating vector YIp5. Thick lines, pBR322 sequences; thin lines, yeast sequences. Deletions in the *HIS4* leader region were generated in both the 3'-to-5' and 5'-to-3' directions with BAL 31 nuclease from the *Xho*I (+100 position of *HIS4* coding region) and *Pvu*II (588 bp from the start of transcription) restriction sites, respectively. An *Eco*RI restriction site, 5'-GGAATTCC-3', was introduced at the novel joint in each set of deletion plasmids, and the extent of each deletion was determined by DNA sequencing. Plasmids which recreate *HIS4* promoter and coding sequences but have altered leader regions were constructed by ligating the large and small *Bam*HI-*Eco*RI restriction fragments from 3'-to-5' and 5'-to-3' deletion plasmids, respectively. The 767-bp *Sau*3A DNA fragments from these plasmids were subcloned into the *Bam*HI site of plasmid p596 to generate in-frame *HIS4-lacZ* fusions. The *Eco*RI site in both plasmids served as the site of ligation of oligonucleotides for insertion mutagenesis studies. Parent and insertion plasmids were used to transform yeast strain TD28 to determine the effects of altered *HIS4* leader regions on growth and β-galactosidase activity. Ori, Origin; B/S, *Bam*HI/*Sau*3A ligation junction.

at *HIS4* that corresponded in length and partial sequence to the starting oligonucleotide, as well as the integrity of the 5' and 3' *HIS4* sequence flanking the site of insertion.

Construction of mutations flanking the *HIS4* translational start codon. Specific base changes within the initiator region of the *HIS4* gene were introduced by the two-primer site-directed method of Zoller and Smith (45). Oligonucleotides containing single or multiple base changes ranged from 17 to 23 nucleotides, and altered sequences were flanked by a minimum of 8 nucleotides. The template used for the mutagenesis was the phage vector M13mp10 that contained the 1,578-bp *Sal*I DNA restriction fragment from the proximal region of *HIS4* as previously described for the in vitro construction of *HIS4* initiator codon mutations (8). The presence of each mutation at *HIS4* was confirmed in at least two independent hybridization positive phage plaques by the

DNA sequence methods described above, with the exception that a single-stranded template was used for the chain termination method.

The 1,578-bp *Sal*I DNA fragment containing each site-directed mutation at *HIS4* was isolated as double-stranded DNA from each of the two independent phage isolates and subcloned into the *Sal*I site of plasmid B142. In addition, the mutated *HIS4* region contained within the 767-bp *Sau*3A DNA restriction fragment (as described above) was subcloned into the *Bam*HI site of the plasmids p349 and p451. Details of the construction of both p349 and p451 have been reported elsewhere (8). Briefly, p349 contains the *lacZ* coding region, and fusion with the *HIS4 Sau*3A fragment in the proper orientation puts the normal initiator codon at *HIS4* in frame with *lacZ*. Plasmid p451 also contains the *lacZ* coding sequence; however, the *Bam*HI site at the amino terminus is in a different reading frame. Subcloning the *HIS4 Sau*3A fragment in the proper orientation at this *Bam*HI site enables the AUG codon at position +29 to be in frame with *lacZ*. Both p349 and p451 are derived from the yeast centromere-containing vector YCp50 (31).

Construction of *HIS4* alleles. Most plasmids derived from the integrating B142 vector containing deletion, duplication, or insertion mutations and base changes in the *HIS4* region were introduced at the *HIS4* locus on chromosome III by transforming the S288C-derived yeast strain TD28 (*MAT***a** *ura3-52 ino1*-13) to Ura3$^+$ (13). Enrichment for loss of the Ura3$^+$ vector was performed by the 5′-fluoro-orotic acid positive selection procedure (2). Ura3$^+$ His$^-$ strains identified by this procedure were purified and analyzed genetically in crosses to the *HIS4*$^+$ strains S288C (*MAT*α) or 6288-8C (*MAT*α *leu2-3, -112*). Standard genetic techniques and media used for these studies have been previously described (35).

The inability to identify His$^-$ yeast strains among TD28 transformants indicated that some mutations constructed at *HIS4* did not confer a His$^-$ phenotype. These plasmids, as well as control plasmids, were used to transform the TD28-derived yeast strains 358-8C (*MAT*α *his4-308 ura3-52 ino1-13*) and 361-1B (*MAT*α *his4-310 ura3-52 ino1-13 leu2-3, -112*) to Ura3$^+$. Both of these strains are His$^-$ because of the presence of deletions of the initiator region at *HIS4*, positions −4 to +9 and −4 to +16, respectively, that were constructed from our BAL 31 deletion studies. Ura3$^+$ His$^+$ transformants were purified and subjected to 5′-fluoro-orotic acid to select for vector loss. All His$^+$ strains constructed in this manner were analyzed genetically in crosses to the *his4* mutant yeast strains 44-4D (*MAT***a** *his4-401 leu2-3,-112 ino1-13*) and 492-3A (*MAT***a** *his4-401 ura3-52 ino1-13 gcn4-103*).

Construction of *HIS4-lacZ* fusion strains. The effects of in vitro-constructed mutations on *HIS4* gene expression were quantitated in *HIS4-lacZ* fusion strains. All *HIS4-lacZ* constructions were used to transform yeast strain TD28 by the lithium acetate procedure (14). Preparation of yeast extracts and assays of β-galactosidase activity were identical to previously described methods (8). At least two independent yeast strains containing each *HIS4-lacZ* construction were assayed in two independent sets of experiments. β-Galactosidase specific activities (nanomoles of *o*-nitrophenyl galactoside cleaved per minute per milligram of protein) are expressed as a percentage of the *HIS4-lacZ* wild-type activity (typically 350 to 450 U).

RESULTS

Effects of mutated sequence context 5′ to the *HIS4* AUG start codon. The proximal region from *HIS4* wild-type DNA was subcloned into the phage vector M13mp10, and mutations at various positions flanking the AUG start codon were introduced by site-directed mutagenesis in light of preferred features observed at the majority of yeast initiator regions as well as functional components of mammalian initiator regions. These mutated *HIS4* regions were sequenced to confirm the presence of the mutation, subcloned in frame with the *lacZ* coding region present on the centromere-containing yeast vector p349 (8), and introduced into strain TD28. Some mutated regions were also subcloned into the integrating vector B142 and substituted for the *HIS4* initiator region to assay their effects on *HIS4*-related growth properties (8).

TABLE 1. Effects of changing the sequence context 5′ and 3′ to the *HIS4* start codon

HIS4 sequence context[a]	Result of assay: β-Galactosidase sp act (% of wild type)[b]		*HIS4* growth[c]
	% Initiation at +1 AUG	% Initiation at +29 AUG	
−3 +4			
AAUAAUGGUU	100	<0.1	+
C	97	<0.1	
G	77		+
U	61	<0.2	+
CACC	80		+
+4			
A	88		+
C	102	<0.1	
U	92	<0.1	+
UCG	100		+
UCU	68		+

[a] Base changes are underlined and correspond in position to the presented wild-type mRNA sequence.

[b] p349-derived in-frame *HIS4-lacZ* fusion in yeast strain TD28 enables expression of β-galactosidase from the normal *HIS4* initiator region (+1 AUG). Out-of-frame *HIS4-lacZ* fusions derived from p451 put the next available AUG codon at *HIS4* in frame with *lacZ* (+29 AUG). β-Galactosidase specific activity is measured in nanomoles of *o*-nitrophenyl galactoside cleaved per minute per milligram of total protein. β-Galactosidase is expressed as a percentage of the wild-type in-frame fusion level to normalize independent sets of assays. Typical in-frame wild-type fusion-specific activities ranged from 350 to 450 U.

[c] Altered initiator regions were substituted for the *HIS4* locus on chromosome III by transformation/transplacement procedures by using the TD28-derived strain 358-8C or 361-1B.

The *HIS4* gene, like 75% of yeast genes, has an A nucleotide at the −3 position relative to the AUG codon that corresponds to the start site of translation. This bias is identical to that observed in higher eucaryotes (7, 20), and point mutations at this position have a more pronounced effect on expression at the preproinsulin gene than any other base change in conserved flanking nucleotides (23).

As shown in Table 1, changing the A nucleotide at the −3 position at *HIS4* to C, G, or U resulted in 97, 77, and 61% of wild-type levels as measured in *HIS4-lacZ* fusion strains. This lower level of expression associated with U at the −3 position reflects comparative studies (7, 11) whereby U is the least represented nucleotide at the −3 position (3%). However, this reduction, as measured by β-galactosidase levels, does not appear to reflect any physiological consequence on in vivo *HIS4* expression. The corresponding mutated allele, when present at *HIS4*, conferred growth properties to yeast strains on histidine-free SD medium plates which were indistinguishable from the growth properties of wild-type strains.

We also introduced two complex sequences 5′ to the AUG start codon at *HIS4* to test their effects on expression. One sequence, 5′-ACACACAC-3′, is present at various positions 5′ to the start codon in 10% of 131 yeast genes (7) and has been implicated, by mutation at the *CYC1* locus in yeast, to have a potential role in initiation (40). Introducing this sequence 27 nucleotides upstream of the AUG start codon at *HIS4* did not result in any significant change in *HIS4* expression as measured in *HIS4-lacZ* fusion strains (108% of wild-type levels; data not shown). Likewise, changing the normal *HIS4* initiator region by site-directed mutagenesis from AAUAAUGG to CACCAUGG, to resemble the optimal sequence context of higher eucaryotic initiator regions as established by comparative and mutational studies (20, 23), did not increase the efficiency of initiation at this AUG as measured by β-galactosidase and growth assays (Table 1).

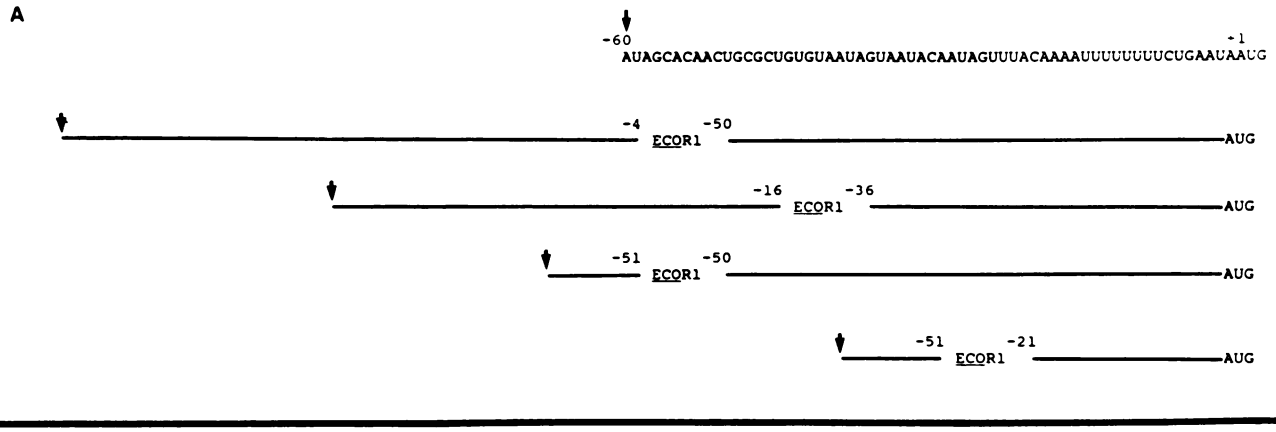
Effects of mutated sequence context 3′ to the *HIS4* AUG start codon. We also tested the effects of point mutations at the +4 position, since *HIS4* maintains a G at this position as observed for 39% of higher eucaryotic genes (20) whereas yeast genes (7, 11) predominantly exhibit U at the +4 position (39%) with C at +4 being least represented (8%). As assayed in *his4-lacZ* fusions, 102, 92, or 88% of wild-type β-galactosidase levels were observed when C, U, or A was present at the +4 position, respectively, and none of these alleles affected the phenotypic expression of *HIS4* (Table 1).

Comparative studies also indicated that the +5 and +6 positions are biased at yeast genes with C (49%) and U (47%), respectively, being the most represented. Inspection of the amino acids encoded at the +2 amino acid position of 131 yeast genes indicated that the most represented amino acid is serine (10%) encoded by UCU (7). Because this codon was prevalent, suggestive of its potential importance to some aspect of the initial steps of translation, we changed the valine GUU codon at the +2 amino acid position at *HIS4* for the preferred and less preferred serine codons UCU and UCG, respectively, and measured their effects in *HIS4-lacZ* fusions and as alleles of the *HIS4* locus. As shown in Table 1, the β-galactosidase activity associated with the UCG triplet was comparable to wild-type activity, while the UCU triplet conferred β-galactosidase levels that are 68% of the wild-type level. Neither of the corresponding UCG or UCU *HIS4* alleles had any effect on growth when they were present in yeast strains (Table 1). Because we anticipated that the more preferred UCU serine codon would not have resulted in lower β-galactosidase expression than the least preferred UCG codon at *HIS4*, we isolated each *HIS4-lacZ* construction from strains that contained the two independent in vitro constructs of UCU and UCG and sequenced the corresponding *HIS4* regions. In each case, the respective mutations at the +4, +5, and +6 positions were present as initially constructed.

Ribosomal bypass of *HIS4* sequence context mutants. Aside from decreasing the efficiency of initiation at an AUG start codon, mutating the preferred sequence context surrounding an AUG in mammalian genes can result in an increase in translation initiation at the next available downstream AUG (15, 23), presumably as a result of reducing ribosomal recognition of the mutant upstream initiator region. We, therefore, tested some of our mutated sequence contexts at *HIS4* for their ability to confer similar effects. The proximal *Sau*3A restriction fragments from these mutated DNAs were subcloned into the centromere-containing plasmid p451 (8). As a result of this subclone, the normal initiator region was out of frame with *lacZ*. However, the next available AUG codon in the *HIS4* message (position +29) would be fused in frame with the *lacZ* coding region. This construction enabled the detection of initiation events at this downstream AUG as a result of ribosomal bypass of the mutated upstream initiator region, when present, in *S. cerevisiae* TD28. As shown in Table 1, none of the *HIS4* sequence context mutants tested supported any appreciable β-galactosidase expression. Specifically, the A-to-U change at the −3 position, which presumably confers an approximate 40% decrease in the efficiency of initiation at the normal +1 start position, did not result in any significant supplementary increases in initiation at the +29 position when it was fused in the +1 frame.

Alternating the length of *HIS4* leader and the effects on *HIS4-lacZ* expression. The proximal *HIS4* region was altered on an integrating YIp5-derived plasmid by the generation of a series of BAL *31* deletions in both the 5′-to-3′ and 3′-to-5′ direction with mapped endpoints in the DNA region corresponding to the 5′ noncoding region of the *HIS4* message (Fig. 1; described in Materials and Methods). By combining 5′-to-3′ and 3′-to-5′ deletion fragments through the common *Eco*RI linker present at the novel joint, we could duplicate or delete sequence information between the 5′ end of the message and the normal AUG start codon (Fig. 1). The constructions chosen for our analysis are diagrammed in Fig. 2. These constructions vary the normal 60-nucleotide distance of the *HIS4* leader over a range of 115 to 39 nucleotides in length. Plasmids p−4/−50 and p−16/−36 duplicate positions −4 to −50 and −16 to −36, respectively, in the *HIS4* region; plasmid p−51/−50 maintains the normal *HIS4* sequence content; and plasmid p−51/−21 contains a deletion between positions −51 to −21. A unique *Eco*RI restriction linker is present at the novel joint in these altered *HIS4* leader regions. These altered *HIS4* regions were subcloned as a *Sau*3A restriction DNA fragment and ligated in frame with the *lacZ* coding region present on plasmid p596 (Fig. 1), which is a centromere-containing yeast vector derived from YCp50. As shown in Table 2, each region, when in frame with the *lacZ* coding region (p−4/−50Z, p−16/−36Z, p−51/−50Z, and p−51/−21Z) and present in yeast strain TD28, expressed β-galactosidase from the normal AUG start codon at *HIS4* at levels comparable to that of a control wild-type *HIS4-lacZ* fusion strain. This demonstrates that despite a threefold change in the length of the *HIS4* leader region, little or no variation was observed for expression.

Effects of G+C-rich inserts with dyad symmetry on *HIS4-lacZ* expression. Three oligonucleotides, B1, B2, and B3, of increasing length and dyad symmetry were each ligated into the *Eco*RI restriction site of *HIS4-lacZ* plasmids p−16/−36Z, p−51/−50Z, and p−51/−21Z (Fig. 2), and the presence of each insert was confirmed by sizing on acrylamide gels and DNA sequencing. These oligonucleotides, when transcribed at *HIS4*, should introduce stable hairpin-loop structures in the 5′ noncoding region of the *HIS4* message that differ in size and strength as predicted by the Zuker RNA-folding program (46) and calculation of the theoretical standard free energies (41). One set of constructions, p−16/−36ZB1, p−16/−36ZB2, and p−16/−36ZB3, would result in increasing amounts of secondary structure at a central location in the *HIS4* leader (46 nucleotides from the 5′ end of the message and 36 nucleotides from the AUG start codon). Another set of constructions, p−51/−50ZB1, p−51/−50ZB2, and p−51/−50ZB3, introduces secondary structure nearer the 5′ end of the transcriptional start point (9 nucleotides 3′) but further 5′ from the AUG (50 nucleotides). The third set of constructions, p−51/−21ZB1, p−51/−21ZB2, and p−51/−21ZB3, provides controls for changes in the natural length of the *HIS4* leader region, as insertion of B1 would not



B

OLIGONUCLEOTIDE	SEQUENCE	CALCULATED ΔG° OF INSERT (kcal/mole)[a]
B1	AATTCCCGGATCCGGG	-21.3
B2	AATTCCCGGATCCGGGCCCGGATCCGGG	-39.9
B2M1	AATTGGCCGCGCCGGATCCCGGGGACTC	-21.8
B2M2	AATTGAGTCCCCGGGATCCGGCGCGGCC	-20.1
B3	AATTCCCGGATCCGGGCCCGGATCCGGGCCCGGATCCGGG	-66.4

FIG. 2. Diagrammatic representation of altered *HIS4* leaders in *HIS4-lacZ* or *HIS4* allele constructions. Deletion plasmids generated in the 3'-to-5' direction (Fig. 1) with mapped endpoints at positions −4, −16, and −51 in the *HIS4* leader region were combined with 5'-to-3' deletion plasmids (Fig. 1) that have mapped endpoints at position −50, −36, and −21 to produce four *HIS4* leader constructions designated −4/−50, −16/−36, −51/−50, and −51/−21, as shown above in comparison to the wild-type leader sequence. When these leaders are present as part of plasmid p176, they are referred to in the text as p−4/−50, p−16/−36, p−51/−50, and p−51/−21. The p is dropped when referring to yeast strains containing the corresponding *his4* allele construction. On the basis of our DNA constructions, the distance between the 5' end of the *HIS4* message and the AUG would be varied from 115 to 39 nucleotides by duplication or deletion of *HIS4* leader sequences, and this has been confirmed for some of these leaders by transcript analysis (see text and Fig. 4). When these leaders are fused in frame to the *lacZ* plasmid p596 (Fig. 1), they are referred to in the text as p−4/−50Z, p−16/−36Z, p−51/−50Z, and p−51/−21Z. The p is dropped when referring to yeast strains containing the corresponding *HIS4-lacZ* plasmid. (B) Oligonucleotides B1, B2, and B3 are of increasing length and dyad symmetry (arrows) and have 1, 2, and 3 copies of the *Bam*HI linker 5'-CCCGGATCCGG-3', respectively. Each oligonucleotide is self-complementary, and when each is self-annealed, 5' protruding ends can be ligated in the *Eco*RI site. The position of the *Eco*RI restriction site at the novel joint of these leaders is presented to emphasize the position of hairpin-loop structures in the mRNA relative to the 5' end of the message and AUG as a result of inserting these oligonucleotides. Oligonucleotides B2M1 and B2M2 are complementary with compatible *Eco*RI ends that are identical in length and base compositions to B2 but have reduced dyad symmetry. Gibbs free-energy constants were calculated by a Zuker RNA-folding program (46) over the entire length of the oligonucleotide inclusive of *Eco*RI restriction site ends. When plasmids or yeast strains contain altered leaders with an insert, this is designated by the oligonucleotide code (i.e., for plasmids, p−51/−50B1 and p−51/−50ZB1; for yeast strains, −51/−50B1 and −51/−50ZB1).

increase the *HIS4* leader beyond the normal 60-nucleotide length and B2 only extends this length by a few nucleotides. All plasmids were introduced into yeast strain TD28, and β-galactosidase activities were compared with those of parent and wild-type *HIS4-lacZ* fusion strains. As shown in Table 2, insertion of oligonucleotides B1, B2, or B3 that contribute ΔG values less than −21.3 kcal (1 kcal = 4,184 J) (Fig. 2B) virtually abolished β-galactosidase activity in all three different leader constructions. As the length of the insert was reduced (and, therefore, the secondary structure as noted by B1- versus B2- or B3-related constructions), the level of β-galactosidase activity increased. Nevertheless, β-galactosidase levels reflected extreme sensitivity to the short G+C-rich insert, B1, in each leader construction, being 20- to 100-fold lower than wild-type or parent control levels.

Stability of *HIS4-lacZ* plasmids. One interpretation of the lowered β-galactosidase activities associated with insertion mutations is that stable hairpin loop structures in the DNA confer instability that results in loss or rearrangement of the plasmids. To address this possibility, we grew representative parent and B2 insertion *HIS4-lacZ* fusion strains in SD medium lacking uracil, and total DNA was isolated, restricted with *Eco*RI, and probed by Southern analysis with ^{32}P-labeled DNA from the proximal region of *HIS4*. As shown in Fig. 3, two DNA fragments hybridized to the *HIS4* probe in each *HIS4-lacZ* fusion strain. The higher-molecular-weight fragment corresponded in size (14 kilobases [kb]) to the starting plasmid p−51/−50Z, which was linearized by restriction with *Eco*RI. This is consistent with each *HIS4-lacZ* construction being present in yeast as part of the autonomously replicating YCp50 vector. The lower-molecular-weight DNA fragment shown in Fig. 3 corresponded to the 3-kb proximal *Eco*RI fragment derived from the chromosomal wild-type *HIS4* gene (10) in yeast strain TD28. The relative intensity of the 3-kb *HIS4* fragment to the 14-kb plasmid fragment in each strain suggests that the *HIS4-lacZ* construction, whether from parent or insertion plasmids, was present at a similar copy number.

To ensure that no small rearrangements had occurred at

TABLE 2. Effects of altering the structure of the leader region on *HIS4*-related expression

Leader			Result of assay		
			β-Galactosidase sp act (% of wild type)[c]	Growth[d] for genetic background:	
Name	Length[a]	ΔG value[b]		*GCN4*$^+$	*gcn4* mutant
Wild type	60	−4.7	100	+	+
−4/−50	115	ND[e]	95	ND	ND
−16/−36	89	−11.5	133	+	+
B1	105	−31.0	1.3	−	−
B2	117	−48.0	<0.2	−	−
B3	129	−76.1	<0.1	−	ND
B2M1	117	−29.8	50	+	+
B2M2	117	−31.5	37	+	+/−
−51/−50	68	−6.7	118	+	+
B1	84	−27.9	5.6	+/−	−
B2	96	−46.5	1.0	−	−
B3	108	−73.0	0.3	−	−
B2M1	96	−26.7	21	+	−/+
B2M2	96	−28.4	3.5	+	−
−51/−21	39	−1.7	95	+	+
B1	55	−21.3	2.6	ND	ND
B2	67	−39.9	0.8	−	ND
B3	79	−66.4	0.3	ND	ND
B2M1	67	−20.1	27	+	−/+
B2M2	67	−21.8	5.0	+	−

[a] Number of nucleotides from the A of the AUG codon to the determined or predicted transcriptional start site, including the insertion when appropriate.
[b] Calculated by the Zuker RNA-folding program (46) for the predicted leader length, including the AUG start codon.
[c] p596-derived in-frame *HIS4-lacZ* fusions in yeast strain TD28 enable β-galactosidase expression from the +1 AUG (Fig. 1). See Table 1, footnote *c*, for β-galactosidase specific activity.
[d] *HIS4* leader constructions were substituted for the *HIS4* locus on chromosome III in the *GCN4*$^+$ yeast strain TD28 or TD28-derived yeast strains (358-8C or 361-1B). These strains were compared in relative growth tests, on SD medium lacking histidine, to the parent strain TD28. The *gcn4* mutant strains containing *HIS4* alleles with altered leaders were isolated as meiotic segregants from crosses with the *gcn4* mutant yeast strain 492-3A and compared with *HIS4*$^+$ *gcn4* yeast strains for their ability to grow on SD plates lacking histidine. Growth was scored after 1 day at 30°C.
[e] ND, Not determined.

the *HIS4* region, we isolated all of the *HIS4-lacZ* plasmids from *S. cerevisiae* and sequenced the regions corresponding to the insertion mutations (data not shown). In each case, the presence of the proper insert was confirmed and the integrity of the *HIS4* DNA region was established. Therefore, decreased levels of β-galactosidase activity associated with insertion mutagenesis must be a result of impairing gene expression as it cannot be attributed to plasmid loss nor to rearrangement of the *HIS4-lacZ* DNA.

Effects of G+C-rich inserts with dyad symmetry on *HIS4*-related growth properties. The three oligonucleotides B1, B2, and B3 were also inserted into the *Eco*RI site present in the *HIS4* leader region of the integrating plasmids p−16/−36 and p−51/−50, and B2 alone was inserted into plasmid p−51/−21 (Fig. 2). These regions, as well as the corresponding parent constructions, were then substituted for the normal leader region at *HIS4* by the transformation/transplacement procedure (34). These strains were tested for their ability to grow in the presence or absence of histidine on plates. As shown in Table 2, all alleles constructed to contain either oligonucleotide B2 or oligonucleotide B3 conferred a His$^-$ phenotype after 1 day of growth, in contrast to parent alleles, which, in the absence of histidine, supported growth indistinguishable from that of wild-type alleles. The −16/−36B1 allele also conferred a His$^-$ phenotype after 1 day of growth, but its counterpart, −51/−50B1, conferred weak growth properties which were reduced in comparison to those of wild-type and parent strains. Upon prolonged incubation (2 to 3 days), the −16/−36B1 and −51/−50B2 strains exhibited a weak His$^+$ phenotype.

Our studies indicate that all insertions are deleterious to β-galactosidase expression yet, in some cases, still support growth (albeit weak) when present as alleles of *HIS4*. One difference between the β-galactosidase and growth assays is that the strains in the latter assays are grown in the absence of histidine. As a result of poor *HIS4* expression, these strains will starve for histidine and derepress *HIS4* transcription four- to fivefold (9) by the general amino acid control system (43). Therefore, some strains that have slightly higher levels of *HIS4* expression, as reflected by higher β-galactosidase activities, may be able to derepress to a level high enough to permit growth on histidine-free SD medium plates. To test this, all insertion alleles were introduced into a yeast strain that contains a deletion of the *GCN4* locus (*gcn4-103*). The *gcn4* mutation prevents derepression of *HIS4* transcription in response to the general amino acid control system (12). As shown in Table 2, the weak His$^+$ phenotype associated with insertion strains could be reduced to His$^-$ when these alleles were present in a *gcn4* mutant yeast strain. This demonstrates that the ability to derepress *HIS4* transcription in the *GCN4*$^+$ strain compensates for the inhibitory effects of these insertion mutations, allowing residual growth on SD medium lacking histidine. These results at the *HIS4* locus, therefore, parallel our observations in *HIS4-lacZ* fusion strains, thereby independently

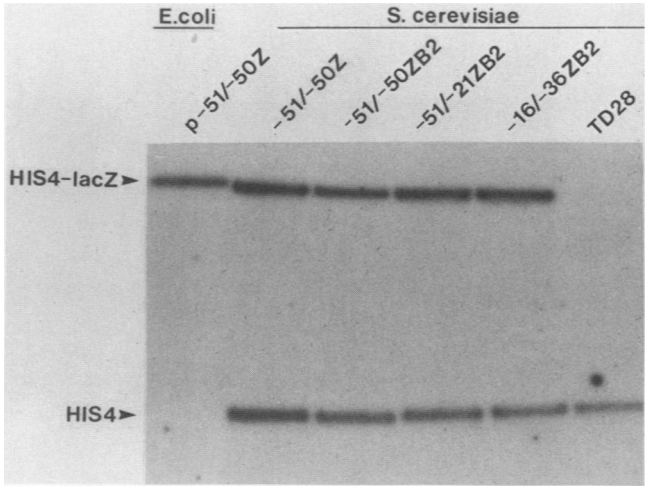


FIG. 3. Southern analysis of *HIS4-lacZ* fusion strains. Total DNA was isolated from the TD28-derived yeast strains −51/−50Z, −51/−50ZB2, −51/−21ZB2, and −16/−36ZB2 that were transformed with the *HIS4-lacZ* plasmids p−51/−50Z, p−51/−50ZB2, p−51/−21ZB2, and p−16/−36ZB2, respectively, and analyzed by Southern analysis in comparison to yeast strain TD28 and plasmid p−51/−50Z purified from *Escherichia coli*. Genomic DNA (10 μg) and p−51/−50Z (2 ng) were restricted with *Eco*RI, coelectrophoresed in a 0.6% agarose gel, transferred to nitrocellulose, and probed with nick-translated *HIS4* DNA (wild-type 767-bp *Sau*3A DNA fragment). The higher-molecular-weight fragment (*HIS4-lacZ*) corresponds in size to the starting plasmids (14 kb). The lower-molecular-weight fragment (*HIS4*) corresponds in size to the 3.0-kb *Eco*RI DNA fragment from the proximal region of the *HIS4* locus on chromosome III.

demonstrating the inhibitory effects of insertion mutations and their physiological relevance to in vivo gene expression.

Effects of G+C-rich inserts with reduced dyad symmetry. In light of the dyad symmetry present in oligonucleotides B1, B2, and B3, a simple interpretation of these insertion experiments is that hairpin-loop structures in the upstream *HIS4* region are inhibitory to expression. One prediction of this interpretation is that if the base composition and length of the oligonucleotide is maintained but the sequence is rearranged to reduce secondary structure, then a greater level of *HIS4* expression should be conferred. We therefore synthesized two complementary oligonucleotides, B2M1 and B2M2 (Fig. 2), that are identical in base composition and length to the oligonucleotide B2 but have a different order of nucleotides to reduce the amount of secondary structure. These two oligonucleotides were annealed and inserted into *HIS4* leader plasmids. Both orientations that would result in either B2M1 or B2M2 sequences being present in the *HIS4* message (referred to as B2M1 and B2M2 insertions, respectively) were identified at *HIS4*. The contributed ΔG values (Fig. 2) of the B2M1 and B2M2 inserts (−21.8 and −20.1 kcal, respectively) more closely parallel the shorter B1 insert (−21.3 kcal). These constructions were introduced into yeast strains, and their effects on expression were analyzed by β-galactosidase and growth assays in conjunction with B2, parent, and wild-type control strains. As shown in Table 2, B2M1 inserts at all *HIS4-lacZ* leader constructions supported 20- to 50-fold-higher levels of β-galactosidase expression than the B2-related strains and enabled growth on SD medium lacking histidine in both *GCN4*$^+$ and *gcn4* strains. On the other hand, the effects of B2M2 insertions were variable. When present at −16/−36, the effects of insertions on expression were consistent with B2M1 effects; that is, the insertions also exhibited higher levels of β-galactosidase expression and a His$^+$ phenotype. However, when such insertions were present in either −51/−50- or −51/−21-related constructions, β-galactosidase levels were low and *HIS4*-related growth was poor. The inhibitory effects of B2M2 insertions in these constructions more closely paralleled predictions based on ΔG values, being as inhibitory as B1 inserts but less inhibitory than the same-size B2 insert that had greater dyad symmetry. Either residual secondary structure or the sequence per se in B2M1 and B2M2 must account for the variability in expression when either B2M1 or B2M2 was present at identical or different positions in the *HIS4* leader region. Despite this variability, these data suggest that reduction of dyad symmetry in the B2M1 and B2M2 inserts confers better expression properties to *HIS4* in our assays than the B2 insert does.

Transcript analysis of *HIS4* insertion alleles. Total RNA was isolated from yeast strains that contain insertion mutations in the −16/−36 *HIS4* leader construction and compared by primer extension with control strains to determine the relative levels of *HIS4* mRNA and structure of the 5′ noncoding region. Figure 4 presents the results of this analysis, as well as a diagram relating extended products to the *HIS4* transcripts. When a 5′-end-labeled oligonucleotide complementary to a region 3′ to the AUG codon (Fig. 4A) was used, the primer extension product from wild-type RNA (W) mapped to the normal start position 60 nucleotides 5′ to the *HIS4* AUG start codon. The same reaction with RNA isolated from the −16/−36 parent strain (P) identified a higher-molecular-weight product (Fig. 4A) that corresponded to the extended length that we predicted on the basis of our DNA constructions (Fig. 2), being 29 nucleotides larger than the wild type as a result of the 21-bp duplication of *HIS4* sequences and the 8-bp *Eco*RI linker in the leader region. In contrast, extended products primed from RNA which was extracted from insertion strains (I) did not identify a product that would correspond to the expected full-length leader. Instead, these products mapped to the site of the insertions 36 nucleotides 5′ to the AUG start codon at *HIS4* (Fig. 4A). Interestingly, the relative levels of *HIS4* mRNA produced from the insertion strains are comparable to control strain levels. We believe that the decreased intensity of the B2M1 products is a result of reduced dyad symmetry which affords priming beyond the site of insertion, thus generating a heterogenous banding pattern (Fig. 4A). Nevertheless, no product could be identified from any of these insertion strains that would represent the predicted full length of the leader that extends to the known start site of transcription at *HIS4*. Similar results were obtained with strains containing −51/−50 and −51/−21 parent and insertion alleles (data not shown). The *HIS4* 5′ noncoding region corresponded to the predicted length and normal start position in the parent allele strains, but the extended products stopped at positions −50 and −21 when RNA from the −51/−50 and −51/−21 insertion allele strains was used, respectively. Our inability to prime beyond the site of insertion in three different leader constructions suggests that DNA cannot be synthesized to the 5′ end of these mRNA templates because secondary structure in the message inhibits the primer extension reaction.

Attempts to identify the 5′ transcriptional start positions in our *HIS4* insertion strains by using B1, B2, or B3 as primers were unsuccessful under a number of different reaction conditions. As an alternative approach to map the 5′ transcriptional start point in some of our insertion strains, we

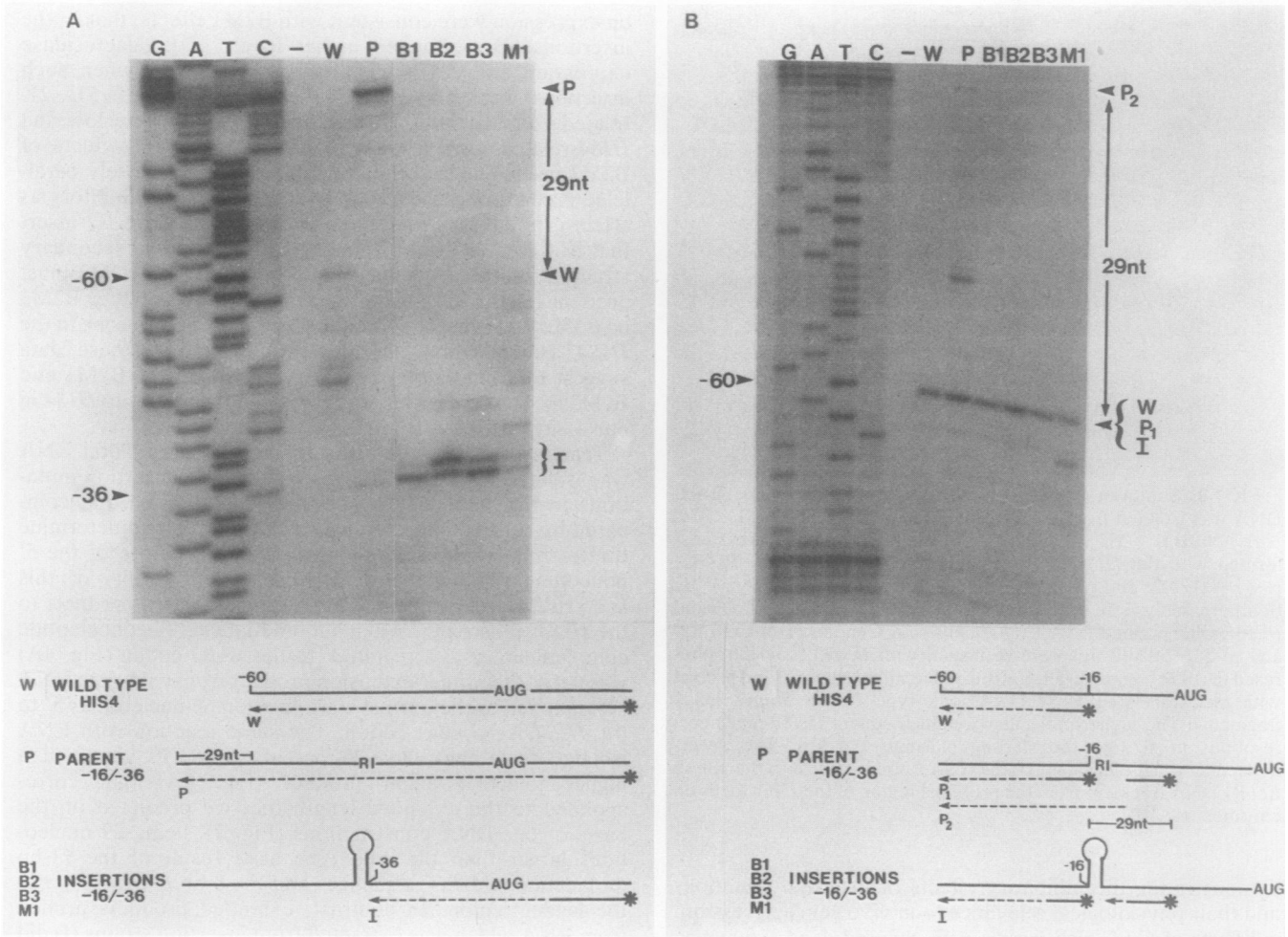


FIG. 4. Primer extension analysis. Oligonucleotides B1, B2, B3, and B2M1 were inserted into the −16/−36 leader construction (Fig. 2) present on plasmid p176 (Fig. 1), and each of the five constructions was substituted for the *HIS4* wild-type region in yeast strain TD28. Total RNA was isolated from the parent (P) and insertion mutant strains (B1, B2, B3, and M1) and compared by primer extension analysis with RNA isolated from the *his4* deletion strain 45-3B (−) and the *HIS4* wild-type strain TD28 (W). (A) RNA (50 μg) from each strain was primed with 0.5 pmol of an oligonucleotide, 5′-CATCAATTAACGGTAGAATCGG-3′, complementary to positions +10 to +31 in the *HIS4* coding region. A DNA sequencing ladder from wild-type DNA, generated from the same oligonucleotide used for priming RNA, is presented. The products generated from wild-type (W), parent (P), and insertion (I) RNAs are noted, and their relationships to the 5′ end of the *HIS4* message (−60) and site of insertion (−36) relative to the AUG are indicated on the sequence ladder. (B) The same experiment is shown, this time with priming by the oligonucleotide 5′-TTTTGTAAACTATTGTATTAC-3′ complementary to *HIS4* leader sequences −36 to −16 which are duplicated in the leader region of −16/−36-related constructions. A DNA sequence ladder from wild-type DNA is provided which was generated from this oligonucleotide. The extended products primed from the wild-type (W), parent (P_1 and P_2), and insertion (I) RNAs are noted in relationship to the 5′ end of the *HIS4* message. The results of these experiments (A and B) are interpreted schematically for each *HIS4* mRNA below each autoradiograph. nt, Nucleotides.

synthesized an oligonucleotide complementary to the *HIS4* message that would enable us to prime from a position 5′ to the site of insertion in the −16/−36-derived yeast strains. Although the oligonucleotide is complementary to the region duplicated in these −16/−36 *HIS4* leaders and, in theory, would give rise to multiple extended products, it should enable the identification of extended products that correspond to the normal transcriptional start point. As shown in Fig. 4B, the use of this oligonucleotide to prime from total RNA isolated from a *HIS4*$^+$ wild-type strain identified the predicted 5′ end of the message (W). The same experiment with the −16/−36 parent strain identified a number of different products, two of which (P_1 and P_2) are relevant to the predicted transcriptional start point. The smallest product (P_1) was identical in length to the wild-type product and was a result of priming from the oligonucleotide which hybridizes to complementary *HIS4* sequences located at the same position relative to the 5′ end of the wild-type and −16/−36 *HIS4* mRNAs (Fig. 4B). The highest-molecular-weight product (P_2) corresponded to an expected product that is 29 nucleotides larger than wild-type and P_1 products, a result which is consistent with priming and extension from the duplicated complementary region located 3′ to the *Eco*RI restriction site (Fig. 4B). The reduced intensity of the highest-molecular-weight product and the presence of intermediate products between P_1 and P_2 are consequences of priming from both complementary regions that are duplicated in the −16/−36 mRNA, since these intermediate products are not observed when priming is performed from a unique region in this *HIS4* message (Fig. 4A).

When RNA isolated from yeast strains which contained insertion alleles derived from the −16/−36 leader region was used, an extended product was observed that was identical in length to the products (W and P_1) identified in experiments with wild-type and −16/−36 parent RNAs (Fig. 4B). This product was a result of priming from an oligonucleotide which hybridized to the complementary *HIS4* sequence located at the same position relative to the 5′ ends of the wild-type and −16/−36 *HIS4* mRNA (Fig. 4B) and therefore demonstrated that the same *HIS4* transcriptional start site is used in insertion strains as that established for parent and wild-type strains. The inability to identify an extended product of higher molecular weight with insertion strains further demonstrates that these inserts confer physical alterations to the *HIS4* mRNA which cannot be extended through by priming from the duplicated 3′ complementary region. The relative intensities of the signals identified with the insertion strains are comparable to levels of parent and wild-type strains, indicating that these insertions within the mRNA do not reduce the level of *HIS4* transcript in these strains as a means of conferring poor *HIS4* expression. Together, these data strongly suggest that the inhibitory effects of these insertions on *HIS4* expression are not a result of altering *HIS4* transcription and, therefore, must alter the ability of *HIS4* mRNA to be translated.

DISCUSSION

As previously described (8), genetic evidence at *HIS4* suggests a mechanism of translation initiation in *S. cerevisiae* that is dependent on the position of the AUG codon relative to the 5′ end of the message in accordance with the ribosomal scanning model (16). Our mutational analysis at *HIS4* of basic sequence and structural features associated with the majority of translational start regions in *S. cerevisiae* is consistent with these genetic studies, further demonstrating the 5′-to-3′ progressive nature of the translational initiation process. Deletion or duplication of the 5′ noncoding region of the *HIS4* message has virtually no effect on *HIS4* expression. This suggests that despite variability in the lengths of yeast leaders, ribosomes have the capability of migrating any distance and scanning the 5′ noncoding region of mRNA for an AUG start codon. This is further illustrated by insertion studies at *HIS4*. Interpretations of similar insertion mutagenesis studies at mammalian genes (22, 32) suggest that localized stem-loop structure in the 5′ noncoding region of mRNA blocks translation initiation either by preventing the ribosome from binding the 5′ end of the message or by preventing its migration toward the AUG start codon. Our observation, that secondary structure inhibits expression when present either 9 or 44 nucleotides 3′ from the *HIS4* transcriptional start point and only alters the physical structure of the *HIS4* mRNA, as shown with our −16/−36 insertion strains, strongly suggests that hairpin-loop structures within the message block either of these initial steps of the scanning process.

However, despite their similarities, functional differences appear to exist between yeast and mammalian translational start regions. Our data suggest that *HIS4* expression is particularly sensitive to secondary structure, showing significant inhibition of expression when DNA inserts have short regions of dyad symmetry that contribute a ΔG value of −20 kcal. In contrast, inserts of similar lengths and ΔG values (−20 to −30 kcal) have little or no effect when introduced into the leader of the herpes simplex thymidine kinase (32) and rat preproinsulin (22) genes, despite a starting ΔG value calculated for the former leader of −64 kcal in comparison to the wild-type ΔG value at *HIS4* of −4.7 kcal. Only when inserts of −50 and −60 kcal are used is significant inhibition of expression of these mammalian genes observed. These differences in the degree of sensitivity to insertion mutations could reflect differences in the stability of hairpin-loop structures as a result of intrinsic properties of each mRNA (i.e., primary sequence). Alternatively, the yeast translation initiation process may be more sensitive to intramolecular and, perhaps, intermolecular base pair interactions. However, our studies are still subject to other interpretations. Insertions at *HIS4* may be extremely inhibitory to expression because they reduce the level of "capped" *HIS4* message, or perhaps hairpin-loop structures inhibit transport from the nucleus, thereby decreasing the availability of *HIS4* message for translation. Some of these other factors could be reflected in our studies that show different effects on *HIS4* expression as a result of inserting identical oligonucleotides at different positions in the *HIS4* leader.

Our data also suggest that sequence context flanking the initiator codon may not function in an equivalent fashion to that believed to exist for mammalian genes. Studies at the preproinsulin gene demonstrate reductions in expression when an A nucleotide at the −3 position is changed to either C, G, or U, although the least represented nucleotides observed at the −3 position at mammalian genes C and U have the most significant effects (21, 23). Our studies indicate a reduction of expression when an unfavored U is present at the −3 position and little or no change when G or C is present. Therefore, it is unclear from our studies why yeast genes exhibit a higher-than-average A nucleotide bias at the −3 position, similar to those of higher eucaryotic genes, since mutations do not have similar marked effects. It has been speculated that the A at −3 functions in concert with 5′-CACC-3′, the optimal sequence context immediately 5′ to AUG start codons of mammalian genes, by mediating a complementary base pair interaction with 18S rRNA (23). Either of two regions within 18S rRNA has been proposed to form base pairs with 5′-CACC-3′; one is located 8 nucleotides from the 3′ end of 18S rRNA (33) and the other is located further 5′ (23). Our observation, that substitution of the *HIS4* initiator region 5′-AAUAAUGG-3′ with the mammalian consensus 5′-CACCAUGG-3′ does not result in increased *HIS4* expression, suggests that 5′-CACC-3′ immediately 5′ to an AUG is not significant for the utilization of an AUG start codon in *S. cerevisiae*. This agrees with comparative analysis which indicates that 5′-CACC-3′ is absent in the initiator region of 131 yeast genes that includes genes known to be highly expressed (7). However, it should be noted that the two different regions proposed to interact with 5′-CACC-3′ are both conserved in yeast (26) and mammalian (5) 18S rRNA. The inability of 5′-CACC-3′ to stimulate *HIS4* expression when adjacent to the AUG codon might then shed doubt as to whether this sequence interacts with complementary sequences in 18S rRNA in eucaryotic organisms.

The region immediately 3′ to an AUG start codon in yeast messages also does not appear to be functionally similar to the same region in mammalian genes. One interpretation of the preference for G at the +4 position in higher eucaryotic messages is that it may mediate a 4-nucleotide interaction with the initiator tRNA (24) that contains a C at position 33 immediately 5′ to the anticodon (19). In yeast $tRNA_i^{Met}$, a U is present at this position (39). Simple base changes of the G nucleotide at the +4 position at *HIS4*, however, have no significant effect on expression. This is in contrast to substi-

tutions at the +4 position of the preproinsulin gene that can confer twofold reductions in expression (23). These differences between yeast and mammalian systems suggest that the potential for a 4-nucleotide recognition process in *S. cerevisiae* does not exist.

Aside from the +4 position, the +5 and +6 positions of yeast initiator regions are biased as well (5'-AUGUCU-3'). This nucleotide bias reflects a nonrandom distribution of amino acids and the preferred codon usage for yeast genes (7). The major types of amino acids observed at this position would confer efficient posttranslational removal of the N-terminal methionine by aminopeptidase (36) and would stabilize yeast proteins as suggested by N-end rules (1). However, our mutational analysis, which alters the normal valine amino acid (GUU) adjacent to the *HIS4* start codon to four different amino acids (Ile [AUU], Leu [CUU], Phe [UUU], and Ser [UCU and UCG], which includes the preferred and less preferred serine codons) does not point to the physiological significance of the bias at positions +4, +5, and +6, as observed at the majority of yeast genes. Instead, our data suggest that the bias 3' to the AUG might reflect very complex functional constraints in this region. It is of interest that the +5 and +6 positions at mammalian genes are also biased in nucleotide distribution (7). Therefore, in light of our studies, mutational data at the preproinsulin gene could also reflect complex functional requirements 3' to an AUG start codon not directly related to initiation, as opposed to the interpretation that the bias at +4 is simply a reflection of a 4-nucleotide codon-anticodon interaction.

Comparative studies indicate that at least 5% of yeast mRNAs maintain one or more AUG codons upstream in the message that does not correspond to the start of translation (7). On the basis of mutational studies at mammalian genes, the modified scanning model (21) has been proposed to account for initiation at a subsequent downstream AUG in a similar subset of higher eucaryotic genes. Two features of leader regions are believed to contribute to these initiation events: (i) the consequence of mutations in conserved nucleotide positions that flank the AUG, allowing the ribosome to bypass and initiate at a downstream site; and (ii) reinitiation, whereby initiation events at an upstream AUG that terminate before or adjacent to the site of a downstream start codon support subsequent initiation events at the downstream start codon.

Among the minority of yeast messages that contain upstream AUG codons that do not correspond to the start site of the primary encoded gene products, either reinitiation features, a poor yeast consensus sequence context both 5' and 3' to the upstream AUG, or both are associated with their leaders (7). However, two of these genes, *CPA1* (42) and *GCN4* (28), are regulated at the level of translation initiation, and features common to the modified scanning process appear to be important to this regulation. Our studies suggest that in the absence of regulation, the role of sequence context surrounding the AUG would not appear to contribute substantially to allowing bypass of an AUG to accommodate downstream initiation. This is supported by our genetic studies at *HIS4* which suggest that only mutations in the AUG start codon at *HIS4* can enable ribosomes to bypass to initiate at a subsequent downstream AUG (8). Therefore, if other genes in *S. cerevisiae* that contain upstream AUGs are not subject to regulation at the level of initiation, this feature should confer very low rates of protein synthesis. However, one construction we have made at *HIS4* suggests that the initiation process in yeast might be more flexible, if not more complex, than our genetic and mutational studies at *HIS4* would indicate. We have introduced an AUG start codon in a normal initiator context 20 nucleotides from the 5' end of the message upstream from the normal +1 start site. This AUG should preclude downstream initiation events; however, we observe 10% activity at the downstream AUG as measured by *HIS4-lacZ* fusions. The corresponding allele construction at *HIS4* supports the growth of *S. cerevisiae* strains on SD medium lacking histidine. This downstream activity as measured by *HIS4-lacZ* fusions can be raised to 20% by multiple context mutations, one of which changes the −3 position relative to the upstream AUG from A to U. Although specific interpretations of this construction are limited because the upstream AUG is very close to the 5' end of the message and because there is an in-frame nonsense codon between the two AUG codons, the construction seems to illustrate that the efficiency of initiation at an AUG is mRNA specific and dependent on other features in the leader that have not been directly addressed in our analysis. This would agree with observations at the *CPA1* (42) and *CYC1* (44) genes which suggest that base changes at the −3 position under certain experimental conditions can allow bypass of a region that has initiation activity, although these changes cannot support very efficient downstream initiation events.

In summary, our studies at *HIS4* provide strong genetic and mutational evidence that *S. cerevisiae* uses a scanning-type mechanism for initiation of protein synthesis, and they also indicate that no simple sequence can override the importance of the first AUG codon for determining the start site of translation. The observation that hairpin-loop structures can inhibit expression raises interesting speculation that differences in the base composition of yeast leader regions that promote different degrees of stable base-pairing interactions may reflect a means to arrive at alternative rates of initiation by a scanning mechanism, as has been postulated for mammalian genes (32).

ACKNOWLEDGMENTS

We thank J. Johansen for excellent technical assistance, Lynne Taylor for preparation of the paper, and A. Hinnebusch for the *gcn4-103* allele.

This work was supported by Public Health Service grant GM32263 from the National Institutes of Health and in part by the Searle Scholars Program of the Chicago Community Trust awarded to T.F.D.

LITERATURE CITED

1. **Bachmair, A., D. Finley, and A. Varshavsky.** 1986. *In vivo* half-life of a protein is a function of its amino-terminal residue. Science **234**:179–186.
2. **Boeke, J. D., F. LaCroute, and G. R. Fink.** 1984. A positive selection of mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol. Gen. Genet. **197**:345–346.
3. **Botstein, D., S. C. Falco, S. E. Stewart, M. Brennan, S. Scherer, D. T. Stinchcomb, K. Struhl, and R. W. Davis.** 1979. Sterile host yeasts (SHY): a eukaryotic system of biological containment for recombinant DNA experiments. Gene **8**:17–24.
4. **Carlson, M., and D. Botstein.** 1982. Two differentially regulated mRNAs with different 5' ends encode secreted and intracellular forms of yeast invertase. Cell **28**:145–154.
5. **Chan, Y.-L., R. Gutell, H. F. Noller, and I. G. Wool.** 1984. The nucleotide sequence of a rat 18S ribosomal RNA gene and a proposal for the secondary structure of 18S rRNA. J. Biol. Chem. **259**:224–230.
6. **Cigan, A. M., and T. F. Donahue.** 1986. The methionine initiator tRNA genes of yeast. Gene **41**:343–348.
7. **Cigan, A. M., and T. F. Donahue.** 1987. Sequence and structural

features associated with translational initiator regions in yeast—a review. Gene **59**:1–18.
8. **Donahue, T. F., and A. M. Cigan.** 1988. Genetic selection for mutations that reduce or abolish ribosomal recognition of the *HIS4* translational initiator region. Mol. Cell. Biol. **8**:2955–2963.
9. **Donahue, T. F., R. S. Daves, G. Lucchini, and G. R. Fink.** 1983. A short nucleotide sequence required for regulation of *HIS4* by the general control system of yeast. Cell **32**:89–98.
10. **Donahue, T. F., P. J. Farabaugh, and G. R. Fink.** 1982. The nucleotide sequence of the *HIS4* region of yeast. Gene **18**:47–59.
11. **Hamilton, R., C. K. Watanabe, and H. A. deBoer.** 1987. Compilation and comparison of the sequence context around AUG start codons in *Saccharomyces cerevisiae*. Nucleic Acids Res. **15**:3581–3593.
12. **Harashima, S., and A. G. Hinnebusch.** 1986. Multiple *GCD* genes required for repression of *GCN4*, a transcriptional activator of amino acid biosynthetic genes in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **6**:3990–3998.
13. **Hinnen, A., J. A. Hicks, and G. R. Fink.** 1978. Transformation of yeast. Proc. Natl. Acad. Sci. USA **75**:1929–1933.
14. **Ito, H., Y. Fukuda, K. Murata, and A. Kimura.** 1983. Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. **153**:163–168.
15. **Johansen, H., D. Schumperli, and M. Rosenberg.** 1984. Affecting gene expression by altering the length and sequence of the 5′ leader. Proc. Natl. Acad. Sci. USA **81**:7698–7702.
16. **Kozak, M.** 1978. How do eukaryotic ribosomes select initiation regions in messenger RNA? Cell **15**:1109–1123.
17. **Kozak, M.** 1980. Evaluation of the "scanning model" for initiation of protein synthesis in eucaryotes. Cell **22**:7–8.
18. **Kozak, M.** 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res. **9**:5233–5252.
19. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**:1–45.
20. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translation start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
21. **Kozak, M.** 1984. Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin *in vivo*. Nature (London) **308**:241–246.
22. **Kozak, M.** 1986. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. Proc. Natl. Acad. Sci. USA **83**:2850–2854.
23. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44**:283–292.
24. **Kozak, M., and A. J. Shatkin.** 1977. Sequences and properties of two ribosome binding sites from the small size class of reovirus mRNA. J. Biol. Chem. **252**:6895–6908.
25. **Lomedico, P. T., and S. J. McAndrews.** 1982. Eukaryotic ribosomes can recognize preproinsulin initiation codons irrespective of their position relative to the 5′-end of mRNA. Nature (London) **299**:221–226.
26. **Mankin, A. S., A. M. Kopylov, P. M. Rubtsov, and K. G. Skryabin.** 1981. 18S ribosomal RNA of eukaryotic ribosomes—the model of the secondary structure. Proc. Natl. Sci. USSR **256**:1006–1010.
27. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
28. **Mueller, P., and A. G. Hinnebusch.** 1986. Multiple upstream AUG codons mediate translational control of GCN4. Cell **45**: 201–207.
29. **Nagawa, F., and G. R. Fink.** 1985. The relationship between the "TATA" sequence and transcription initiation sites at the *HIS4* gene of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **82**:8557–8561.
30. **Natsoulis, G., F. Hilger, and G. R. Fink.** 1986. The *HSTI* gene encodes both the cytoplasmic and mitochondrial histidine tRNA synthetases of *S. cerevisiae*. Cell **46**:235–243.
31. **Parent, S. A., C. M. Fenimore, and K. A. Bostain.** 1985. Vector systems for the expression, analysis and cloning of DNA sequences in *S. cerevisiae*. Yeast **1**:83–138.
32. **Pelletier, J., and N. Sonenberg.** 1985. Insertion mutagenesis to increase secondary structure within the 5′ non-coding region of a eukaryotic mRNA reduces translational efficiency. Cell **40**: 515–526.
33. **Sargan, D. R., S. P. Gregory, and P. H. W. Butterworth.** 1982. A possible interaction between the 3′ end of the 18S rRNA and the 5′-leader sequence of many eukaryotic mRNAs. FEBS Lett. **147**:133–136.
34. **Scherer, S., and R. W. Davis.** 1979. Replacement of chromosome segments with altered DNA sequences constructed *in vitro*. Proc. Natl. Acad. Sci. USA **76**:4951–4955.
35. **Sherman, F., G. R. Fink, and C. W. Lawrence.** 1972. Methods in yeast genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
36. **Sherman, F., and J. W. Stewart.** 1982. Mutations altering initiation of translation of yeast iso-1-cytochrome c: contrasts between the eukaryotic and prokaryotic initiation process, p. 301–333. *In* J. N. Strathern, E. W. Jones, and J. R. Broach (ed.), The molecular biology of the yeast Saccharomyces. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
37. **Sherman, F., J. W. Stewart, and A. M. Schweingruber.** 1980. Mutants of yeast initiating translation of iso-1-cytochrome c within a region spanning 37 nucleotides. Cell **20**:215–222.
38. **Silverman, S. J., M. Rose, D. Botstein, and G. R. Fink.** 1982. Regulation of *HIS4-lacZ* fusions in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **2**:1212–1219.
39. **Simsek, M., and U. L. RajBhandary.** 1972. The primary structure of yeast initiator transfer ribonucleic acid. Biochem. Biophys. Res. Commun. **49**:508–515.
40. **Stiles, J. I., J. W. Stostak, A. T. Young, R. Wu, S. Consaul, and F. Sherman.** 1981. DNA sequence of a mutation in the leader region of the yeast iso-1-cytochrome c mRNA. Cell **25**:277–284.
41. **Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, M. D. Crothers, and J. Gralla.** 1973. Improved estimations of secondary structure in ribonucleic acids. Nature (London) New Biol. **246**:40–41.
42. **Werner, M., A. Feller, F. Messenguy, and A. Pierard.** 1987. The leader peptide of yeast gene *CPA1* is essential for the translation repression of its expression. Cell **49**:805–813.
43. **Wolfner, M., D. Yep, F. Messenguy, and G. R. Fink.** 1975. Integration of amino acid biosynthesis into the cell cycle of *Saccharomyces cerevisiae*. J. Mol. Biol. **96**:273–290.
44. **Zitomer, R. S., D. A. Walthall, B. C. Rymond, and C. P. Hollenberg.** 1984. *Saccharomyces cerevisiae* ribosomes recognize non-AUG initiation codons. Mol. Cell. Biol. **4**:1191–1197.
45. **Zoller, M. J., and M. Smith.** 1984. Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template. DNA **3**:479–488.
46. **Zuker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9**:133–148.