McLennan, Eds. (Oxford Univ. Press, New York, 1985), pp. 160–186.
24. J. Schumpeter, *The Theory of Economic Development* (Harvard Univ. Press, Cambridge, MA, 1934).
25. D. W. Jorgensen [in *The Positive Sum Strategy*, R. Landau and N. Rosenberg, Eds. (National Academy of Sciences, Washington, DC, 1985), pp. 55–76] argued that high energy prices discouraged investment and has also suggested that the energy crisis shifted the direction of innovation and reduced its productivity impact.
26. National Science Board, *Science Indicators: The 1985 Report* (Government Printing Office, Washington, DC, 1986), p. 218.
27. F. M. Scherer, "The world productivity growth slump," mimeograph copy (Swarthmore College, Swarthmore, PA, 1984).
28. M. N. Baily and A. K. Chakrabarti, *Brookings Pap. Econ. Act.* **2**, 609 (1985).
29. The innovation data reflect the diffusion of basic technical change more than its origination. A breakthrough in technology tends to be followed by a high rate of patenting and new product or process introduction as the idea is applied in different ways. A consequence of this timing pattern is that innovation, as we measure it, is coincident with productivity growth. We are catching technical change as it comes into use.
30. The data in Fig. 5 are raw counts of innovations. The rankings by importance did not change the conclusions given in the text. For example, the slowdown in chemical innovations occurred both in those categories classified by chemical engineers as major, and also in those classified as minor.
31. Department of Labor, *Employment and Training Report of the President* (Government Printing Office, Washington, DC, 1982), table A-19, p. 178.
32. ———, *ibid.*, table C-2, p. 241.
33. An executive in a leading corporation provided this figure.
34. R. H. Hayes and W. J. Abernathy, *Har. Bus. Rev.* **58** (July–August 1980), p. 67.
35. The United Kingdom has improved productivity in manufacturing since 1979, largely as a result of shake-outs induced by Prime Minister Thatcher in major nationalized industries such as automobile and steel.
36. Productivity declined in other types of mining also, partly because of comparable depletion of reserves, and also because the Mine Safety Act mandated substantial changes in work practices that lowered productivity. Oil and gas mining is the largest part of the industry.
37. The views expressed here are those of the author and should not be ascribed to the trustees or other staff members of the Brookings Institution.

Research Articles

Saturation Mutagenesis of the Yeast *his3* Regulatory Site: Requirements for Transcriptional Induction and for Binding by GCN4 Activator Protein

DAVID E. HILL, IAN A. HOPE, JENNIFER P. MACKE, KEVIN STRUHL

Expression of the yeast *his3* and other amino acid biosynthetic genes is induced during conditions of amino acid starvation. The coordination of this response is mediated by a positive regulatory protein called GCN4, which binds specifically to regulatory sites upstream of all coregulated genes and stimulates their transcription. The nucleotide sequence requirements of the *his3* regulatory site were determined by analysis of numerous point mutations obtained by a novel method of cloning oligonucleotides. Almost all single base pair mutations within the nine base pair sequence ATGACTCTT significantly reduce *his3* induction in vivo and GCN4 binding in vitro, whereas changes outside this region have minimal effects. One mutation, which generates a sequence that most closely resembles the consensus for 15 coregulated genes, increases both the level of induction and the affinity for GCN4 protein. The palindromic nature of the optimal sequence, ATGACTCAT, suggests that GCN4 protein binds as a dimer to adjacent half-sites that possibly overlap.

TRANSCRIPTIONAL REGULATION OF GENE EXPRESSION IS mediated by activator or repressor proteins that bind specifically to regulatory DNA sequences. The expression of unlinked genes can be regulated in unison if they contain similar regulatory sequences that are recognized by a common activator or repressor. In prokaryotic organisms, regulatory sites are associated with palindromic sequences whose "half-sites" are seven to nine bases in length (*1–4*). This symmetry is believed to be important because these regulatory sites are recognized by dimers of the cognate DNA binding proteins (*5–9*). However, even in these well-studied regulatory sites, the structural requirements are poorly defined because they are based primarily on a relatively small number of point mutations. Positions of transcriptional control elements for some eukaryotic genes have been established but, although some inferences have been drawn from DNA sequence comparisons, the functional determinants have yet to be defined.

In bakers' yeast, *Saccharomyces cerevisiae*, the general control system coordinately regulates the expression of unlinked genes encoding amino acid biosynthetic enzymes from different pathways (*10*). That is, all genes regulated by general control are transcriptionally induced in unison above the basal level (two- to tenfold, depending on the gene) during conditions of amino acid starvation. Extensive deletion analysis of *his3*, a coregulated gene, has defined the cis-acting elements necessary and sufficient for promoter function and for regulation by general control (Fig. 1). Basal level *his3* expression requires a TATA element located 35 to 55 nucleotides upstream of the transcription initiation site and a poly(dA-dT) sequence located between −115 and −129 (*11, 12*). A separate genetic element, located between −84 and −104 is necessary for positive regulation of *his3* in response to amino acid starvation (*13*).

The authors are in the Department of Biological Chemistry, Harvard Medical School, Boston, MA 02115. The present address of D. E. Hill is Genetics Institute, 87 Cambridge Park Drive, Cambridge, MA 02140.

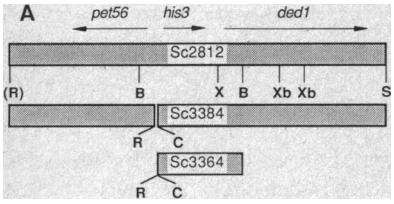
This regulatory region contains the sequence TGACTC, which is repeated, with minor variation, five additional times between −130 and −260.

The TGACTC sequence is presumed to be the critical cis-acting element for genes under general control. It is present within the promoter region of all coordinately regulated amino acid biosynthesis genes examined (*13–26*). Deletion analyses of *his3* (*13*) and *his4* (*15*) indicate that at least one copy of TGACTC is required for regulation by general control. When a synthetic oligonucleotide containing the TGACTC sequence is inserted into the *cyc1* promoter region, *cyc1* expression, which is ordinarily regulated by glucose and heme, becomes subject to general control (*27*). Finally, GCN4, a positive regulatory protein that is critical for coordinate induction (*28, 29*), interacts directly with the *his3* TGACTC sequence and binds specifically to the promoter regions of three other coregulated genes (*30*).

We now report the nucleotide requirements for positive regulation of *his3* by general control. Base pair substitution mutations throughout the *his3* regulatory region were obtained and then analyzed for their effects on *his3* induction in vivo and on binding by GCN4 protein in vitro.

Isolation and phenotypic characterization of *his3* point mutations. The nucleotide sequence of the *his3* promoter region (*31*) (including the regulatory site) is shown in Fig. 1 along with the structures of several *his3* derivatives. In order to generate a large number of single base pair substitutions throughout the desired region, we used a local mutagenesis procedure in which degenerate oligonucleotides are synthesized and then cloned (*32*) (Figs. 1 and 2). A degenerate oligonucleotide is a synthetically derived mixture of oligonucleotides whose heterogeneous central portions are bounded at their 5′ and 3′ ends by common sequences recognized by restriction endonucleases. In the cases described below, the central portions were mutagenized versions of the *his3* regulatory region. These degenerate oligonucleotides were converted to the double-stranded form by the method of mutually primed synthesis (*33*) (Fig. 2) and then cloned into appropriate *his3* derivatives in order to replace the DNA between −85 and −102 (GGA*T-GACTC*TTTTTTTTT). In this way, we were able to obtain a diverse collection of base pair substitution mutations from the products of a single DNA synthesis.

Mutations of the *his3* regulatory region were obtained in four different ways. (i) We decided to saturate the conserved TGACTC sequence with single base changes. Six oligonucleotide mixtures were synthesized, each containing the three sequences corresponding to all possible mutations of one of the TGACTC nucleotides. The degenerate oligonucleotides and a wild-type control oligonucleotide were cloned as Eco RI–Sac I fragments (Fig. 1 and Table 1). (ii) In order to determine whether the nine dT residues immediately following the TGACTC core affect *his3* induction, Eco RI–Sac I oligonucleotides containing the core and 0, 3, 4, and 6 dT residues were introduced into the *his3* promoter region. The resulting derivatives differ not only in the number of dT residues, but also in the spacing between the core and the messenger RNA (mRNA) start sites (Table 2). Similarly, we analyzed a derivative in which the GGA residues immediately preceding the core were removed. (iii) We used a single degenerate oligonucleotide to create a variety of



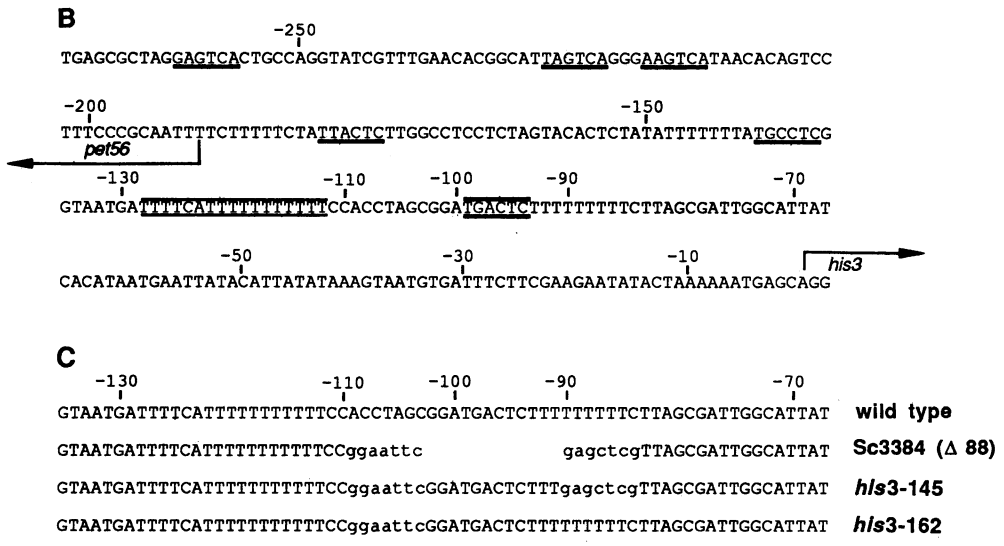


Fig. 1. Structures of *his3* derivatives. (A) Structures (drawn to scale) of cloned *his3* segments. Sc2812 is a 6.1-kb DNA fragment containing the intact *pet56*, *his3*, and *ded1* genes (locations and orientations are defined by arrows). Restriction endonuclease cleavage sites for Eco RI (R), Bam HI (B), Xho I (X), Xba I (Xb), Sal I (S), and Sac I (C) are shown (the parentheses around the leftmost Eco RI site indicates that it has been mutated). (B) Nucleotide sequence of the divergent *his3-pet56* promoter region corresponding to *his3* nucleotides −273 to +3. The locations of the *his3* regulatory site (bold lines under and over the sequence), related sequences (bold underline), upstream elements for constitutive expression (thin lines under and over the sequence), and *his3* and *pet56* mRNA initiation sites (arrows) are indicated. (C) Nucleotide sequence between −135 and −67 of various derivatives. Wild-type nucleotides are indicated with capital letters and nucleotides corresponding to restriction sites are shown as small letters. To create point mutations in the core of the *his3* regulatory region, we synthesized seven different oligonucleotides containing recognition sequences for Eco RI at the 5′ end and Sac I at the 3′ end. One of these oligonucleotides contained the wild-type sequence between −91 and −103 and corresponds to *his3*-145. At the appropriate step during each of the remaining six DNA syntheses, the wild-type nucleotide precursor at one predetermined position in the TGACTC core was replaced by an equimolar mixture of the three "mutant" precursors. Thus, each DNA synthesis yields a mixture of three oligonucleotides that represent all three point mutations at a specific position within the core. The oligonucleotides were converted to the double-stranded form by mutually primed synthesis (*33*) (Fig. 2), cleaved with Eco RI and Sac I, and inserted into M13mp19-Sc3364. The DNA sequence of each phage was determined by the chain termination method (*39*). Eco RI–Xho I fragments from DNA derivatives containing point mutations within the core were subcloned into YIp55-Sc3384 (*36*). Variants of the *his3* regulatory region containing 0, 4, and 6 T residues downstream from the core were synthesized individually and cloned in an identical fashion. YIp55 hybrid DNA's containing the *his3* mutations were cleaved with Xba I and introduced into yeast strains KY117 (*40*) or KY603 (a *gcd1*-101 derivative of KY117) by selecting for uracil prototrophy. The resulting strains, which contain a single copy of the transforming DNA integrated at the *his3* locus, were grown in nonselective medium for ten generations, and ura$^-$ segregants were selected by their ability to grow on plates containing 5-fluoroorotic acid (*41*). Approximately 50 percent of these segregants were His$^+$, thus indicating that the original *his3*-Δ200 allele was replaced by the *his3* point mutation of interest. It was expected that point mutants in the regulatory region would be His$^+$ because *his3*-Δ88, which deletes the entire regulatory site, does not affect the basal, uninduced level of *his3* expression and hence permits growth in the absence of histidine (*37*).

single and multiple base pair changes throughout the region between −85 and −101. Each base within this region was mutated at a rate of 10 percent by programming the synthesis reaction with four mixtures, each composed of one major (90 percent) nucleotide precursor and equal amounts of the three remaining precursors (10 percent of the total). The degenerate oligonucleotide was cloned as an Eco RI–Dde I fragment (Fig. 2), and a collection of single and multiple base substitutions as well as the wild-type sequence were obtained (Table 3). (iv) We designed four degenerate Eco RI–Dde I oligonucleotides to produce almost all of the remaining mutations of interest (Table 4).

The degenerate oligonucleotides were cloned so that the resulting molecules were identical to the wild-type *his3* gene except for the region derived from the oligonucleotide. The mutant DNA's were introduced into yeast cells in such a way that they precisely replaced the normal *his3* allele on chromosome XV (see legend to Fig. 1). The phenotypes conferred by the point mutations were assessed by the growth properties of strains and by direct measurements of *his3* mRNA. The strains were tested initially for their ability to grow on medium containing aminotriazole (NH_2T), a drug that causes histidine starvation because it is a competitive inhibitor of the *his3* gene product (*34, 35*). Mutants containing a defective *his3* regulatory site grow slowly as compared to those containing a functional site because they are unable to induce *his3* expression in response to starvation conditions. The relative levels of *his3* transcription were determined by quantitative S1 nuclease mapping with the *ded1* RNA as an internal control (*12, 31*). Under normal growth conditions, *his3* transcription is initiated equally from positions +1 and +12. However, under conditions of amino acid starvation, the amount of the +12 transcript (and a minor transcript at +22) is increased by a factor of 5, whereas the +1 transcript is unaffected (*36*) (Fig. 3). Thus, the relative levels of the +12 and +1 transcripts during conditions of amino acid starvation provides a sensitive measurement for the ability of any particular mutation to induce *his3* expression.

Each nucleotide of the TGACTC core is essential for *his3* induction in vivo. The growth properties of strains containing Eco RI–Sac I oligonucleotides that represent 15 out of the 18 possible point mutations within the TGACTC core (*his3*-146 to *his3*-160) are listed in Table 1. Each of the 15 point mutations within the TGACTC core confers sensitivity to aminotriazole. Thirteen of these strains are extremely sensitive to aminotriazole, a phenotype which is similar to strains in which the proximal TGACTC element is deleted (*37*). Strains containing *his3*-158 or *his3*-159 (in which the final C is changed, respectively, to T or A) are slightly less inhibited. In contrast, strains containing a wild-type oligonucleotide (*his3*-144 and *his3*-145) grow in the presence of aminotriazole.

These results indicate that any point mutation within the core inactivates the *his3* regulatory site, and therefore prevents induction in response to amino acid starvation. This was confirmed by directly measuring the levels of the +1 and +12 transcripts (Fig. 3). The *his3*-145 strain, which contains the wild-type oligonucleotide, shows a clear induction of the +12 transcript, whereas strains containing any of the 15 point mutants (*his3*-146 to *his3*-160) show equal levels of both the +1 and +12 transcripts. For some of these mutations, we also analyzed RNA from cells that were not starved for amino acids (Fig. 3). In all cases, basal expression levels for the wild-type and the *his3* point mutants were indistinguishable. This indicates that the point mutations do not affect constitutive, basal *his3* expression.

Sequences on both sides of the TGACTC core are essential for *his3* induction. The core, although essential, is not sufficient to constitute a functional *his3* regulatory site (Table 2). First, *his3*-161, which removes all nine T residues downstream from the core, is functionally defective as assayed by aminotriazole sensitivity or by mRNA quantitation. This indicates that at least some of the T residues downstream from the core are necessary for the *his3* regulatory site. Derivatives *his3*-145 and *his3*-144, which contain three or four T residues, are functional in both assays, although *his3* induction occurs only to approximately 60 to 70 percent that of wild type. Transcriptional induction of *his3* alleles containing six or nine T residues (*his3*-143, *his3*-162) is indistinguishable from that of the wild-type allele. Second, *his3*-Δ83, with a deletion which replaces the three purine residues upstream of TGACTC by the three pyrimidine residues from the Eco RI site, is unable to induce *his3* transcription during amino acid deprivation (*37*). As *his3* sequences upstream of −102 and the Eco RI linker between −103 and −109 do not influence *his3* induction, at least some of these purine residues are critical for a functional *his3* regulatory site.

Additional information concerning the importance of nucleotides flanking the core comes from the phenotypes of mutations produced by Eco RI–Dde I oligonucleotides (Table 3 and Fig. 3). As would be expected, a strain containing the wild-type oligonucleotide (*his3*-162) cannot be distinguished from the wild-type strain, whereas strains containing mutations within the core are sensitive to aminotriazole and fail to induce *his3* transcription. However, *his3*-171 and *his3*-172 (in which the second T residue after the core is changed to C and G, respectively) induce *his3* transcription very poorly, and

Table 1. Saturation mutagenesis of the core. For each *his3* allele, the nucleotide sequence of the regulatory region cloned between the Eco RI and Sac I sites is shown, and mutations are underlined. Aminotriazole (NH_2T) sensitivity is represented as follows: ++ indicates growth at wild-type rates in 10 to 20 m*M* NH_2T and slow growth at 40 m*M* NH_2T, + indicates growth at all concentrations but at slightly slower rates than wild-type at 20 to 40 m*M* NH_2T, ± indicates no growth above 20 m*M* NH_2T and slow growth in 10 m*M* NH_2T, and − indicates no growth at 10 m*M* and slow growth at 4 m*M* NH_2T. The mRNA levels were determined from the autoradiogram of Fig. 3. The entries represent the ratio of the +12 transcript to the +1 transcript. For binding by GCN4 protein, entries are represented as ++ for wild-type binding ability, + for binding that is detectably below the maximal level, ± for weak, but possibly detectable binding, and − for no detectable binding. ND indicates not determined.

his3 allele	Sequence	NH_2T	RNA	GCN4
wild type	GGATGACTCTTTT*	++	5.1	++
his3-Δ88	deleted	-	1.0†	-
his3-144	GGATGACTCTTTT	+	2.9	+
his3-145	GGATGACTCTTT	+	3.2	+
his3-146	GGATGAC**A**CTTT	-	0.8	-
his3-147	GGATGAC**C**CTTT	-	1.0	-
his3-148	GGATGAC**G**CTTT	-	1.1	-
his3-149	GGA**A**GACTCTTT	-	0.8	-
his3-150	GGA**C**GACTCTTT	-	1.0	-
his3-151	GGATGA**A**TCTTT	-	0.7	-
his3-152	GGATGA**T**TCTTT	-	0.7	-
his3-153	GGAT**C**ACTCTTT	-	0.7	-
his3-154	GGAT**A**ACTCTTT	-	0.9	-
his3-155	GGATT**A**CTCTTT	-	0.8	-
his3-156	GGATG**T**CTCTTT	-	0.6	-
his3-157	GGATG**C**CTCTTT	-	0.8	-
his3-158	GGATGACT**T**TTTT	±	1.1	ND‡
his3-159	GGATGACT**A**TTTT	±	1.2	ND‡
his3-160	GGATGACT**G**TTTT	-	0.8	ND‡

*The wild-type sequence does not contain the Eco RI or Sac I sites flanking the core. † Data from (*37*). ‡Identical point mutations (cloned as Eco RI–Dde I oligonucleotides) were tested as shown in Table 3.

Table 2. Phenotypic analysis of deletions. The nucleotide sequence between −109 and −80 is shown for each *his3* allele. The TGACTC core is underlined, capital letters indicate wild-type nucleotides, and small letters indicate nucleotides from the Eco RI and Sac I sites. The upstream deletion endpoints are listed for derivatives above the wild type, and the downstream deletion endpoints are shown for derivatives below the wild type. The phenotypes are listed as described in Table 1.

his3 allele	Sequence	Endpoint	NH_2T	RNA	GCN4
his3-143	ggaattcGGATGACTCTTTTTTgagctcGTTAG	-88	++	4.8	++
his3-144	ggaattcGGATGACTCTTTT gagctcGTTAG	-90	+	2.9	+
his3-145	ggaattcGGATGACTCTTT gagctcGTTAG	-91	+	3.8	+
his3-161	ggaattcGGATGACTC gagctcGTTAG	-94	±	0.9	±
wild-type	ACCTAGCGGATGACTCTTTTTTTTCTTAG	None	++	5.1	++
his3-Δ83	ggaattc TGACTCTTTTTTTTCTTAG	-99	±	1.0*	±
his3-162	ggaattcGGATGACTCTTTTTTTTCTTAG	-102	++	5.9	++
his3-Δ82	aattcTCGGATGACTCTTTTTTTTCTTAG	-104	++	ND	++
his3-Δ81	ACCTAGCGGATGACTCTTTTTTTTCTTAG	-115	++	ND	++

*Data from (*37*).

they are strongly inhibited by aminotriazole although not quite to the extent of point mutations within the core. A change of A to G just upstream of the core, *his3*-164, appears to confer an induction similar to that of wild type. Changes in any of the distal seven T residues 3′ to the core have no effect on induction.

Similar nucleotide requirements for *his3* induction and GCN4 binding. GCN4 protein, synthesized by transcription and translation in vitro, binds directly to the core of the *his3* regulatory site (*30*). Moreover, analysis of sequentially deleted *his3* DNA's indicates that the ability to activate transcription in vivo directly correlates with the ability to bind GCN4 protein in vitro (*30*). We extended these results by analyzing the mutations described in the previous sections. For each derivative, *his3* DNA fragments were incubated with ^{35}S-labeled GCN4 protein, and the products were analyzed by electrophoresis in nondenaturing polyacrylamide gels to assay the formation of GCN4 protein–DNA complexes. The concentrations of DNA and protein in these experiments were chosen such that a small decrease in the binding constant would cause a clear reduction in the intensity of the band due to protein-DNA complex formation (*30*) (Fig. 4).

Analysis of the deletion templates (Fig. 4 and Table 2) gave the following results. With respect to sequences upstream of the core, GCN4 binding is normal when there are three or more wild-type nucleotides (for example, *his3*-162, *his3*-Δ82, and *his3*-Δ81). However, a derivative containing the core but no upstream nucleotides (*his3*-Δ83) binds GCN4 protein very weakly. With regard to the T residues downstream from the core, derivatives containing six (*his3*-143) or nine (*his3*-162) T residues bind GCN4 protein indistinguishably from the native gene, derivatives containing three or four T residues (*his3*-145 and *his3*-144) bind somewhat less well, and the derivative with no T residues (*his3*-161) binds weakly if at all. These results indicate that DNA sequences flanking both sides of the core are necessary for GCN4 binding.

Point mutants of the TGACTC sequence are unable to bind GCN4 protein (Fig. 4). This is true for all derivatives originally obtained with the Eco RI–Sac I (Table 1) or the Eco RI–Dde I (Table 3) oligonucleotides. Point mutants upstream of the core appear unaffected in binding ability as are mutations in the last seven T residues. However, both point mutations of the second T residue appear to result in reduced binding capability.

The results indicate that, even at nucleotide resolution, the structural requirements for transcriptional activation and DNA binding are very similar. Without exception, mutations that abolish *his3* induction also fail to bind GCN4 protein whereas mutations that confer *his3* inducibility are able to bind GCN4 protein. These observations suggest that binding of GCN4 protein to the *his3* regulatory site directly mediates transcriptional regulation by general control.

DNA sequence comparisons to other coregulated genes. To

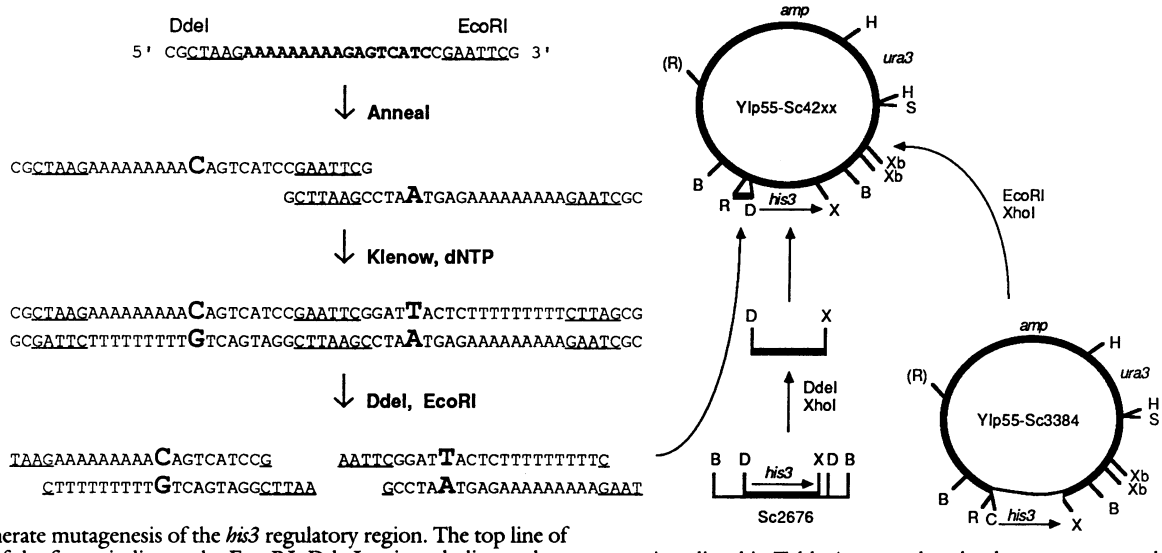


Fig. 2. Degenerate mutagenesis of the *his3* regulatory region. The top line of the left part of the figure indicates the Eco RI–Dde I activated oligonucleotide containing 17 central bases (bold) that were synthesized with 90 percent of the base indicated and 3.3 percent of the other three bases. Shown below is the mutually primed synthesis procedure (*33*) used to convert the oligonucleotide mixture to the double-stranded form suitable for cloning. The large, boldfaced residues indicate specific base pair substitution mutations of the regulatory region. The Eco RI–Dde I oligonucleotide mixture was combined with the 9-kb Eco RI–Xho I fragment of YIp55-Sc3384 and the 0.9-kb Dde I–Xho I fragment of Sc2676 (the relevant parts of these molecules are indicated in boldface) to produce the desired molecules (indicated as YIp55-Sc42xx). The same procedure was used to generate the mutations listed in Table 4 except that the degenerate, central regions of the four oligonucleotides were different. YIp55 was constructed by cloning the 1.1-kb Hind III *ura3* fragment (*42*) into pUC8 (*43*). The locations of the *amp*, *ura3*, and *his3* genes as well as various restriction endonuclease sites [abbreviated as in Fig. 1 except for Dde I (D) and Hind III (H)]. The positions of Dde I sites located outside of the 1.8-kb *his3* Bam HI fragment are not shown. For each derivative, the nucleotide sequence of the *his3* regulatory region was determined with double-stranded plasmid DNA as a template (*44*). DNA's were introduced into yeast cells such that the cloned *his3* sequences replaced the chromosomal *his3* allele as described in the legend to Fig. 1.

obtain more information on the importance of the flanking nucleotides, we compared promoter sequences of other genes subject to general control. Previous comparisons (*13, 15, 16*) were complicated by the inability to distinguish functional regulatory sites from defective ones. For example, although the *his3* promoter region contains five additional sequences that resemble the *his3* regulatory site, none of these are bound by GCN4 protein (*30*).

Our results indicate that each of the TGACTC nucleotides is critical for a functional *his3* regulatory site. We therefore searched for perfect core sequences on both strands between 50 and 350 nucleotides 5′ to transcription start sites in 15 coregulated genes (Fig. 5). Although many of these genes have multiple imperfect copies, a perfect core sequence was found in 13 genes, 11 of which contain only a single perfect core. The *his1* promoter contains two perfect core sequences (both listed in Fig. 5), and the *his4* promoter contains three perfect core sequences. However, only the proximal *his4* copy is listed because induction mediated by the other two copies is 50- to 100-fold below the maximum level (*27*). Thus, unlike previous comparisons, it is likely that the DNA sequences listed in Fig. 5 represent the regulatory sites for induction during amino acid starvation.

This list of presumptive functional regulatory sites makes it clear that the nucleotides flanking the core on both sides show considerable sequence conservation (Fig. 5). The two nucleotides before the core are usually purines (13 out of 16 cases for each position), and an A residue immediately precedes the core in 8 out of 16 cases. In the four positions just after the core, the most common nucleotides are ATTT (at least ten matches at each position). In addition, there is a slight preference for T residues even further downstream, especially the residue corresponding to position −86 of the *his3* site. Interestingly, although the *trp5* and *ilv2* promoters do not contain a perfect core sequence (the final C is an A for *trp5* and the central C is a T for *ilv2*), the flanking nucleotides are an excellent match to the consensus. This suggests that a regulatory site with an imperfect core can be functional if all the other residues are favorable. Thus, we propose that the expanded consensus for the regulatory site for general control is RR*TGACTC*ATTT (R is a purine). Interestingly, none of the 16 sites listed in Fig. 5 are identical to the consensus sequence. The *his3* site differs only at position −93, which contains a T residue instead of the more conserved A.

A point mutation that increases *his3* induction and GCN4 binding. To test the validity of our proposed consensus more explicitly, we cloned four degenerate Eco RI–Dde I oligonucleotides and obtained almost all the remaining mutations within the critical region. The results of the phenotypic analyses (Table 4 and Figs. 3 and 4) confirm the strong correlation between *his3* induction in vivo and GCN4 binding in vitro and are summarized below. The three remaining changes in the TGACTC core are defective in *his3* induction and GCN4 binding, although the phenotype of *his3*-190 may be slightly less severe. Upstream of the core, the two new mutations of the G at −101 behave indistinguishably from the wild-type gene, whereas the A to C change at −100 produces a somewhat defective regulatory site. Double mutations at positions −101 and −100 tend to have more deleterious effects than the analogous single mutations. Mutations in any of the three proximal T residues downstream of the core, *his3*-186, *his3*-187, and *his3*-188, clearly show defects in binding and induction, but these are less extreme than those conferred by mutations in the core.

The most interesting mutation is *his3*-189, which generates a sequence identical to the consensus shown in Fig. 5. *His3*-189 is the only one among the 57 mutations examined that increases the level of *his3* induction. A strain containing *his3*-189 grows extremely well

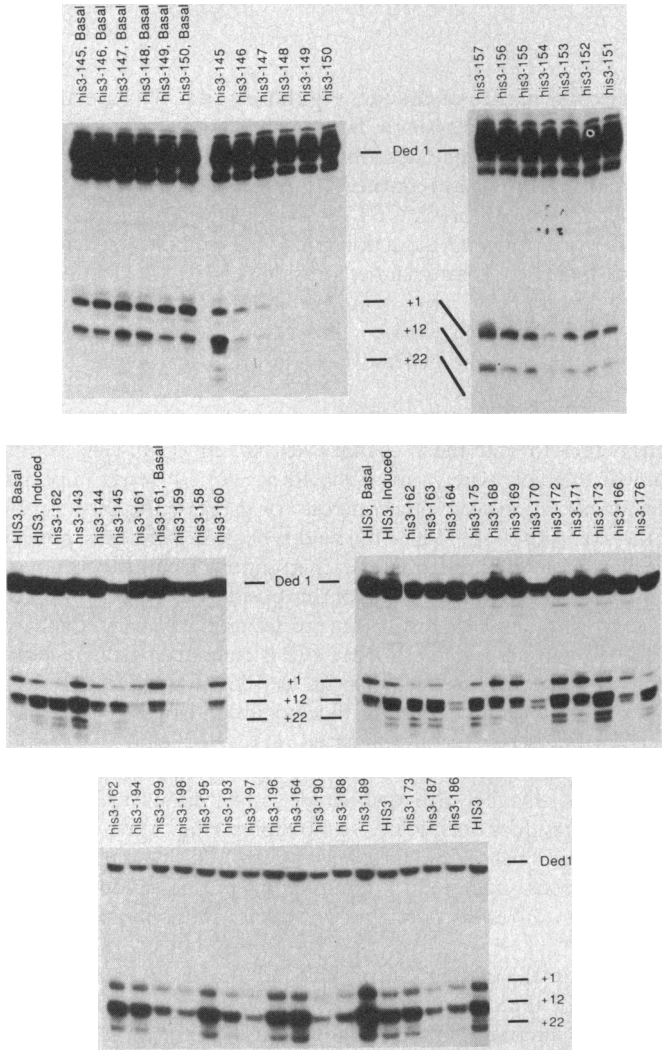


Fig. 3. RNA analysis. In strains containing a variety of mutations in the *his3* regulatory region, levels of *his3* mRNA produced during inducing conditions were determined by quantitative S1 nuclease mapping with the level of *ded1* RNA as an internal control (*12, 31*). The locations of the *ded1* RNA and the *his3* +1, +12, and +22 transcripts are indicated. Inducing conditions were achieved with strains containing the *gcd1*-101 allele. This allele abolishes the translational control of *gcn4* mRNA and results in a constitutively high level of GCN4 protein (*45, 46*). As this situation causes induced levels of all coregulated genes even during growth in broth, it is equivalent to conditions of amino acid starvation. For some of the *his3* mutations, the basal level of *his3* transcription was determined with the use of strains containing a wild-type *GCD1* allele. For each mutation, the inducibility, which is defined by ratio of the +12:+1 transcripts, was quantitated by scanning the autoradiogram with a spectrophotometer (Beckman DU-6). The ratio is internally controlled because it is determined with a single probe; hence, it is independent of amount of RNA placed in each lane. The error is approximately ±25 percent. The amount of RNA added in each lane, determined by the intensity of *ded1* band on a lighter exposure of this autoradiogram, was roughly equivalent except for lanes (i) representing *his3*-171, -172, -173, which contain approximately twice as much RNA; (ii) the *his3*-143 lane in the middle which contains about 50 percent more RNA; and (iii) lanes representing *his3*-158, -159, and -170, which contain approximately two to three times less RNA. Within experimental error, the levels of the +1 transcript (relative to *ded1*) are similar. The unusually low level of the *ded1* RNA in the *his3*-145 lane in the middle part of the figure is probably an artifact resulting from incomplete extraction of the larger 2.3-kb *ded1* RNA as compared to the 0.7-kb *his3* RNA's. In the experiment at the bottom of the figure, the specific activity of the *ded1* probe was roughly five times lower than the *his3* probe, thus accounting for the relatively low band intensities. The level of the +1 transcript produced by the *his3*-189 mutation may be slightly above that of the wild type. HIS3 refers to the wild-type *his3* allele.

in the presence of 40 m*M* aminotriazole, and it has a mRNA induction ratio of 8.6. Moreover, *his3*-189 DNA binds more tightly to GCN4 protein as evidenced by the increased intensity of the band representing protein-DNA complexes. These observations support the correlation between transcriptional induction and GCN4 binding, and they indicate that the optimal *his3* regulatory site is defined by the consensus sequence RRTGACTCATTT. The fact that increased GCN4 binding results in higher transcription levels also suggests that GCN4 protein levels are limiting for *his3* expression in vivo, even during conditions of amino acid starvation.

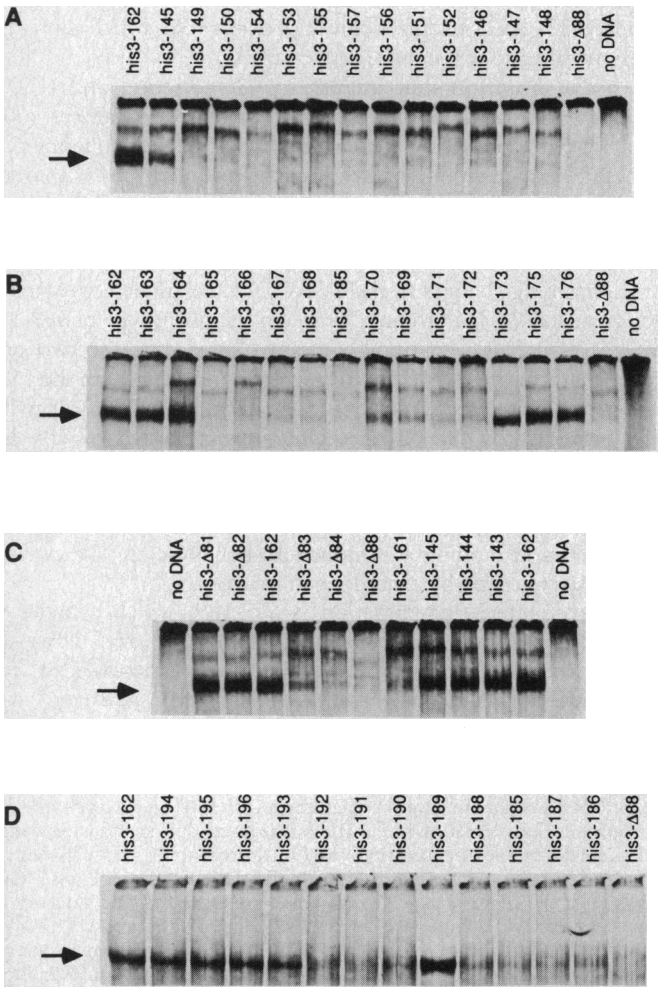


Fig. 4. GCN4 protein binding to *his3* mutant DNA's. The assay for complex formation between GCN4 protein and the mutated *his3* binding sites was performed as described (*30*). [^{35}S]Methionine-labeled GCN4 protein was generated by transcription in vitro of pSP64-Sc4313 (previously called pSP64-GCN4) with the use of SP6 RNA polymerase and subsequent translation in vitro in a wheat germ extract. For each *his3* mutation, the DNA fragment bounded by the Eco RI site at −102 and the Bam HI site at +1318 was cloned into pUC9 (*43*). The resulting molecules were prepared by a rapid lysate procedure and cleaved with Pvu II and Rsa I prior to incubation with GCN4 protein. In this way, the *his3* regulatory region is centrally located within a fragment of approximately 600 bp. On nondenaturing polyacrylamide gels, the electrophoretic mobility of the protein is increased by complex formation with the DNA (arrow indicates the position of the complex). From titration experiments and from the dissociation constant (approximately 10^{-10}*M*) (*30*), the concentrations of DNA (4 n*M*) and protein (0.2 n*M*) were chosen. The GCN4 protein used for the analysis in part D was prepared independently and had a two- to threefold lower specific activity. Although the DNA was in excess of the protein, it was still only just sufficient to drive approximately 30 to 70 percent of the protein into complex formation. Thus, a small change in binding affinity would cause a noticeable increase or decrease in the amount of protein complexing with the DNA.

Table 3. Degenerate mutagenesis of the *his3* regulatory region. For each *his3* allele, the nucleotide sequence of the regulatory region cloned between the Eco RI and Dde I sites is shown with mutations underlined. Phenotypes are listed as described in Table 1.

his3 allele	Sequence	NH_2T	RNA	GCN4
his3-162	GGATGACTCTTTTTTTTT	++	5.9	++
his3-163	GTATGACTCTTTTTTTTT	++	5.9	++
his3-164	GGGTGACTCTTTTTTTTT	++	4.6	++
his3-165	GGAAGACTCTTTTTTTTT	-	1.0*	-
his3-166	GGATTACTCTTTTTTTTT	-	1.1	-
his3-167	GGATGACCCTTTTTTTTT	ND	ND	-
his3-168	GGATGACGCTTTTTTTTT	ND	1.5	-
his3-169	GGATGACTATTTTTTTTT	±	1.2	±
his3-170	GGATGACTGTTTTTTTTT	±	1.9	±
his3-185	GGATGACTTTTTTTTTTT	ND	ND	-
his3-171	GGATGACTCTCTTTTTTT	-	1.0	-
his3-172	GGATGACTCTGTTTTTTT	±	2.1	±
his3-173	GGATGACTCTTATTTTTT	++	5.0	++
his3-175	GGATGACTCTTTTTTATT	++	5.6	++
his3-176	GGATGACTCTTTTTTTTG	++	4.9	++
his3-177	GGTTGGCTCTTTTTTTTT	-	ND	ND
his3-178	GGAGGACTCTTATTTTTT	-	ND	ND
his3-179	GGAGGACTCTTTTTTTTG	-	ND	ND
his3-180	GGATAACACTTTTTTTTT	-	ND	ND
his3-181	GGATTACTCGTTTTTTTT	-	ND	ND
his3-182	GGATGCGTCTATTTTTTT	-	ND	ND
his3-183	GGATGAATCTTCTTATTT	-	ND	ND
his3-184	GGAGGACACTACTTTTTT	-	ND	ND

*Data from (*47*).

Molecular nature of the *his3* regulatory site. We have analyzed 30 of the 33 possible single base pair changes within an 11-bp region as well as some additional mutations. The major sequence determinant of the *his3* regulatory site is the 10-bp region between −100 and −91, A*TGACTC*TTT. Each position can be mutated to produce defects in *his3* induction and GCN4 binding, whereas any of the single base pair changes outside this region do not cause detectable effects in vivo or in vitro. Moreover, this sequence corresponds extremely well to the region that is protected from cleavage by deoxyribonuclease I (DNase I) upon GCN4 binding (*30*).

The core nucleotides TGACTC are critical because all changes are detrimental to and most changes abolish function. The residue immediately downstream from the core, −93, is clearly important because alterations can either increase or decrease the functionality of the site. Although −93 of the wild-type gene is the only position within the *his3* regulatory site that differs from the proposed consensus, a T residue is found at this position in 30 percent of the sequences in Fig. 5. Thus, both the mutational analysis and the proposed consensus indicate that a T is tolerated at this position although it is less optimal than an A. The T residue at −92 is relatively important because all three possible changes cause marked functional defects. The outermost residues, −100 and −91, appear to be less important because certain changes do not result in mutant phenotypes. Finally, although single substitutions in apparently conserved residues outside the 10-bp region (Fig. 5) do not appear to have functional effects, the phenotypes of double mutations and small deletions suggest that the upstream purine residues and the downstream T residues, can influence *his3* induction and GCN4 binding.

The optimal *his3* regulatory site and the proposed consensus sequence for the coregulated genes is a short palindrome centered around the internal C in the core (ATGACTCAT). By analogy with

Table 4. Phenotypic analysis of additional point mutations. For each *his3* allele, the nucleotide sequence of the regulatory region cloned between the Eco RI and Dde I sites is shown with mutations underlined. Phenotypes are listed as described in Table 1. The +++ entries indicate fast growth at 40 m*M* NH_2T and increased binding to GCN4 protein.

his3 allele	Sequence	NH_2T	RNA	GCN4
his3-186	GGATGACTCTTGTTTTTT	±	2.1	±
his3-187	GGATGACTCTATTTTTTT	±	1.9	±
his3-188	GGATGACTCGTTTTTTTT	±	1.5	±
his3-189	GGATGACTCATTTTTTTT	+++	8.6	+++
his3-190	GGATGAGTCTTTTTTTTT	±	1.6	±
his3-191	GGATGGCTCTTTTTTTTT	-	ND	-
his3-192	GGAGGACTCTTTTTTTTT	-	ND	-
his3-193	GGCTGACTCTTTTTTTTT	+	3.0	+
his3-194	GAATGACTCTTTTTTTTT	++	5.5	++
his3-195	GCATGACTCTTTTTTTTT	++	5.1	++
his3-196	GG TGACTCTTTTTTTTT	++	4.7	++
his3-197	GCTTGACTCTTTTTTTTT	±	1.9	ND
his3-198	GATTGACTCTTTTTTTTT	+	2.8	ND
his3-199	GACTGACTCTTTTTTTTT	+	3.1	ND

prokaryotic regulatory sites (*1–9*), this symmetry strongly suggests that GCN4 protein binds as a dimer to half-sites. Recently, we have demonstrated that GCN4 protein indeed binds to *his3* DNA as a dimer (*38*); using a mixture of the wild-type and a deleted GCN4 protein, we observed three distinct protein-DNA complexes corresponding to binding by both possible homodimers and the heterodimer. DNA sequence recognition by a dimer provides a simple explanation for the apparent bidirectionality of the regulatory site (*27*).

Gene	Sequence	Position	Ref.
his1	A G C G T G A C T C T T C C C G G A A	-171	16
his1	G A G G T G A C T C A C T T G G A A G*	-74	16
his3	C G G A T G A C T C T T T T T T T T T	-103	13
his4	A C A G T G A C T C A C G T T T T T T	-140	15
arg3	G T C G T G A C T C A T A T G C T T T	-296	21
arg4	T G A A T G A C T C A C T T T T T G G	-127	18
cpa1	T T C T T G A C T C G T C T T T T C T	-298†	24
cpa2	C G A A T G A C T C T T A T T G A T G	-297†	24
trp5	A G A A T G A C T A A T T T T A C T A	-66	14
trp3	T C G T T G A C T C A T T C T A A T C	-35	17
trp2	T T G C T G A C T C A T T A C G A T T	-72	19
ilv1	G A G A T G A C T C T T T T T C T T T*	-142†	23
ilv2	G C G A T G A T T C A T T T C T C T G	-358†	22
leu1	T A G A T G A C T C A G T T T A G T C	-209	20
leu4	T A A G T G A C T C A G T T C T T T C	-105	25
ils1	A T G A T G A C T C T T A A G C A T G	-80	26
G	4 5 8 5 0 16 0 0 0 0 1 2 1 0 3 4 2 1 5		
A	4 4 5 8 0 0 16 0 0 1 10 0 3 2 0 3 5 2 2	Nucleotide frequency	
T	6 4 0 2 16 0 0 1 16 0 5 11 10 12 9 6 7 12 6		
C	2 3 3 1 0 0 0 15 0 15 0 3 2 2 4 3 2 1 3		
	- - r r T G A C T C a t t t - - - t -	Consensus	

Fig. 5. Consensus for the general control regulatory site. DNA sequences corresponding to *his3* positions −103 to −85 for 16 prospective regulatory sites from 15 genes under general control. The TGACTC core is underlined and the position of the upstream most nucleotide with respect to the mRNA start site is indicated. The nucleotide frequency for each base pair is shown below with the most common residue underlined. Highly conserved nucleotides are shown as capital letters, conserved residues are shown as small letters, and nonconserved positions are indicated by dashes. (*) Positions with respect to the AUG translational initiation codon. (†) Sequence present in opposite orientation with respect to the mRNA start site.

Although the half-sites recognized by prokaryotic regulatory proteins are typically seven to nine bases in length (*1–4*), the half-sites inferred to be involved in GCN4 binding appear to consist of only four bases (5′-ATGA-3′) that are separated by a single nucleotide. However, it seems likely that the central C residue contacts GCN4 protein because it is critical for efficient binding and cannot be replaced by a G, its symmetrical counterpart. Although other interpretations cannot be excluded, we suggest that this central C is actually part of the half-site. In this view, GCN4 monomers would bind to overlapping, but not identical half-sites. This suggests the possibility that the protein-DNA interactions may not necessarily be identical at each of the two half-sites. Definitive proof of this model requires x-ray crystallographic analysis of protein-DNA complexes.

REFERENCES AND NOTES

1. T. Maniatis *et al.*, *Cell* **5**, 109 (1975).
2. A. Johnson, B. J. Meyer, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1783 (1978).
3. T. Taniguchi, M. O. O'Neill, B. deCrombrugghe, *ibid.* **76**, 5090 (1979).
4. D. S. Oppenheim, G. N. Bennett, C. Yanofsky, *J. Mol. Biol.* **144**, 133 (1980).
5. W. F. Anderson, D. H. Ohlendorf, Y. Takeda, B. W. Matthews, *Nature (London)* **290**, 754 (1981).
6. C. O. Pabo and M. Lewis, *ibid.* **298**, 443 (1982).
7. D. B. McKay and T. A. Steitz, *ibid.* **290**, 744 (1981).
8. A. J. Joachimiak, R. L. Kelley, R. P. Gunsalus, C. Yanofsky, P. B. Sigler, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 668 (1983).
9. Y. Takeda, D. H. Ohlendorf, W. F. Anderson, B. W. Matthews, *Science* **221**, 1020 (1983).
10. E. W. Jones and G. R. Fink, in *The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression*, J. N. Strathern, E. W. Jones, J. R. Broach, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982), pp. 181–299.
11. K. Struhl, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7385 (1982).
12. ———, *ibid.* **82**, 8419 (1985).
13. ———, *Nature (London)* **300**, 284 (1982).
14. H. Zalkin and C. Yanofsky, *J. Biol. Chem.* **257**, 1491 (1982).
15. T. F. Donahue, R. S. Daves, G. Lucchini, G. R. Fink, *Cell* **32**, 89 (1983).
16. A. G. Hinnebusch and G. R. Fink, *J. Biol. Chem.* **258**, 5238 (1983).
17. M. Aebi, R. Furtur, F. Prantl, P. Niederberger, R. Hutter, *Curr. Genet.* **8**, 165 (1984).
18. I. R. Beacham, B. W. Schweitzer, H. M. Warrick, J. Carbon, *Gene* **29**, 271 (1984).
19. H. Zalkin, J. L. Paluh, M. vanCleemput, W. S. Moye, C. Yanofsky, *J. Biol. Chem.* **259**, 3985 (1984).
20. Y.-P. Hsu and P. Schimmel, *ibid.*, p. 3714.
21. M. Crabeel *et al.*, *Mol. Cell. Biol.* **5**, 3139 (1985).
22. S. C. Falco, K. S. Dumas, K. J. Livak, *Nucleic Acids Res.* **13**, 4011 (1985).
23. S. Holmberg, M. C. Kielland-Brandt, T. Nilsson-Tillgren, J. G. L. Petersen, *Carlsberg Res. Commun.* **50**, 163 (1985).
24. M. Werner, A. Feller, A. Pierard, *Eur. J. Biochem.* **146**, 371 (1985).
25. J. P. Beltzer, L. L. Chang, A. E. Hinkkanen, G. B. Kohlaw, *J. Biol. Chem.* **261**, 5160 (1986).
26. Z. Altboum, S. Ludmerer, P. Schimmel, personal communication.
27. A. Hinnebusch, G. Lucchini, G. Fink, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 498 (1985).
28. M. D. Penn, B. Galgoci, H. Greer, *ibid.* **80**, 2704 (1983).
29. A. G. Hinnebusch and G. R. Fink, *ibid.*, p. 5374.
30. I. A. Hope and K. Struhl, *Cell* **43**, 177 (1985).
31. K. Struhl, *Nucleic Acids Res.* **13**, 8587 (1985).
32. D. E. Hill, A. R. Oliphant, K. Struhl, *Methods Enzymol.*, in press.
33. A. R. Oliphant A. L. Nussbaum, K. Struhl, *Gene* **44**, 177 (1986).
34. T. Klopotowski and A. Wiater, *Arch. Biochem. Biophys.* **112**, 562 (1965).
35. K. Struhl and R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5255 (1977).
36. K. Struhl, W. Chen, D. E. Hill, I. A. Hope, M. A. Oettinger, *Cold Spring Harbor Symp. Quant. Biol.* **50**, 489 (1985).
37. K. Struhl and D. Hill, *Mol. Cell. Biol.*, in press.
38. I. A. Hope and K. Struhl, in preparation.
39. F. Sanger, A. R. Coulson, B. G. Barell, A. J. Smith, B. A. Roe, *J. Mol. Biol.* **143**, 161 (1980).
40. K. Struhl, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7865 (1984).
41. J. D. Boeke, F. Lacroute, G. R. Fink, *Mol. Gen. Genet.* **197**, 345 (1984).
42. M. L. Bach, F. Lacroute, D. Botstein, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 386 (1979).
43. J. Vieira and J. Messing, *Gene* **19**, 259 (1982).
44. E. Y. Chen and P. H. Seeburg, *DNA* **4**, 165 (1985).
45. A. G. Hinnebusch, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6442 (1984).
46. G. Thireos, M. D. Penn, H. Greer, *ibid.*, p. 5096.
47. D. Hill and K. Struhl, unpublished observations.

48. We thank A. Nussbaum for the synthesis and purification of oligonucleotides, A. Oliphant and P. Youdarian for stimulating conversations, and Z. Altboum for communicating the *ils1* DNA sequence prior to publication. Supported by the Damon Runyon–Walter Winchell Cancer Foundation, DRG-670, and the Massachusetts Medical Foundation (D.E.H.), the Royal Society of Great Britain (I.A.H.), an NIH grant (GM 30186), and the Chicago Community Trust (Searle Scholars Program) (K.S.).

24 April 1986; accepted 8 September 1986