Reprinted from
*Cold Spring Harbor Symposia on Quantitative Biology*, Volume XLVII


# Promoter Elements, Regulatory Elements, and Chromatin Structure of the Yeast *his3* Gene

K. STRUHL
*Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02115*

This paper deals with the relationship between the structure and function of a simple eukaryotic gene, the *his3* gene of the baker's yeast *Saccharomyces cerevisiae*. First, I describe genetic experiments that define elements necessary for *his3* expression (promoter elements) as well as elements responsible for regulating the level of expression as a function of the cell's physiological state (regulatory elements). Second, I discuss the relationship between these genetic elements and the chromatin structure of the *his3* gene.

## A BRIEF INTRODUCTION TO YEAST MOLECULAR BIOLOGY

Although yeast is a unicellular microorganism, it has most of the molecular characteristics of higher eukaryotic cells. Here it suffices to note that the chromosomal DNA is organized into a discrete chromatin structure, mRNAs are synthesized by RNA polymerase II, and genes transcribed by this polymerase contain the sequence TATAAA or some close variant just upstream from the site of transcriptional initiation (the TATA box).

The reason for studying yeast genes is that the sophistication of yeast molecular genetics equals that of prokaryotic organisms such as *Escherichia coli*. In particular, it is possible to isolate essentially any defined gene as cloned DNA, to alter its DNA sequence at will, and to introduce these mutated derivatives back into yeast cells such that there is one copy per cell exactly at the normal chromosomal location (see Fig. 1). In other words, one can obtain a limitless number of yeast strains that differ from the wild type only in that a normal gene has been replaced by a mutated one. Therefore, mutations constructed in vitro can be examined under true in vivo conditions for their effects on gene expression, gene regulation, and chromatin structure.

## THE YEAST *his3* GENE

The *his3* gene (located on chromosome XV) is a simple regulated gene that encodes the structural gene for imidazoleglycerolphosphate (IGP) dehydratase, one of the ten enzymes composing the pathway for histidine biosynthesis (Fink 1964).

*his3* is essential for cell growth only under certain defined conditions. Yeast strains that lack the entire *his3* gene grow just as well as wild-type strains when histidine is present in the growth medium. However, unlike

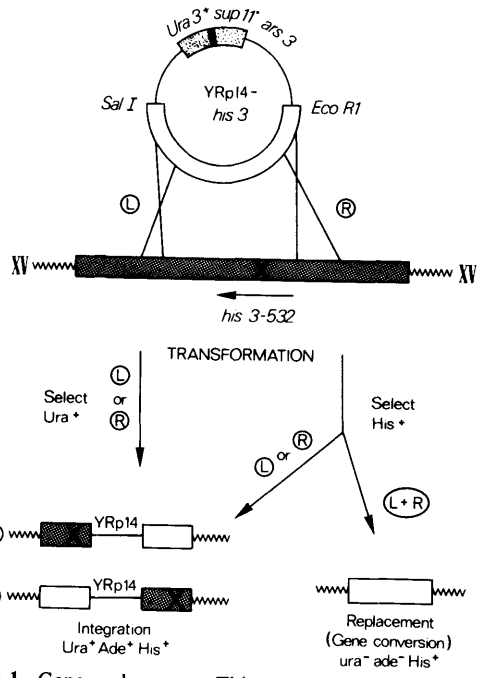


**Figure 1.** Gene replacement. This procedure was used for the analysis of many of the mutations described in this paper (specifically, those listed in Tables 2–4). *his3* DNA fragments (all derived from the 6.1-kb fragment, Sc2605; see Fig. 2) were cloned into YRp14, a vector containing the *ura3* gene, the ochre-suppressing version of a tyrosine-inserting tRNA (*sup11*), and *ars3*, a sequence that inefficiently causes hybrid molecules to replicate autonomously. Each hybrid DNA was introduced into KY137, a strain containing the ochre suppressible allele *ade2*-1, nonrevertible alleles *ura3*-1,2, and *his3*-532. The chromosomal counterpart of the cloned *his3* DNA fragment is shown as a stippled box with an × that marks the spot of the *his3*-532 mutation; the flanking regions of chromosomes XV are indicated by a wavy line. Mitotically stable Ura⁺ or His⁺ transformants were selected. These arise by homologous recombination between the cloned *his3* DNA and yeast chromosome XV. The L crossover occurs to the left of *his3*-532, and the R crossover occurs to the right of *his3*-532. The resulting structures are shown at the bottom of the figure. Almost all of the Ura⁺ transformants were also His⁺ Ade⁺, and the His⁺ transformants fell into two classes. The class of His⁺ colonies that was also Ura⁺ Ade⁺ is equivalent to the Ura⁺ transformants. In these cases (represented by a single L or R crossover), the entire YRp14 hybrid molecule has integrated into the genome, almost always at the *his3* locus. The other class of His⁺ transformant is Ade⁻ Ura⁻ and results from replacement of the *his3*-532 allele by the transforming allele, probably via gene conversion. This class, which occurs only for cloned His⁺ derivatives, is equivalent to the double (L + R) crossover. Both classes of transformants have one copy of the transforming *his3* allele per cell. Technical details and possible problems involved in these procedures are described elsewhere (Hinnen et al. 1978; Struhl et al. 1979; Scherer and Davis 1979; Struhl 1982b).

the wild type, His⁻ strains cannot grow when histidine is omitted from the medium; this property provides both a qualitative assay for *his3* expression and a powerful genetic selection. Nevertheless, the level of *his3* expression is the same regardless of the presence or the absence of histidine in the medium. This basal level (about one mRNA molecule per cell) is sufficient for cells to produce enough histidine such that they grow at wild-type rates even when histidine is omitted from the medium (Struhl and Davis 1981a).

Although *his3* expression is not regulated as a function of exogenously added histidine, it is increased in response to amino acid starvation. Under these conditions, the levels of many amino acid biosynthetic genes are coordinately increased (Wolfner et al. 1975). Of relevance here is the fact that *his3* mRNA levels increase about fivefold to tenfold (Struhl and Davis 1981a) and IGP dehydratase activity increases threefold (Struhl 1982a); the basis of the apparent discrepancy is not known. Current evidence suggests that *his3* expression is regulated solely by this general control mechanism.

Molecular analysis of the *his3* gene and its expression began with its isolation as cloned DNA (Struhl et al. 1976; Struhl and Davis 1977). A structural description of the gene region (Struhl and Davis 1980) and its mRNA transcripts (Struhl and Davis 1981a,b) is shown in Figure 2. *his3* is located very near five apparently unrelated (i.e., not co-regulated) genes. The gene itself does not contain intervening sequences and there is no evidence for a precursor to the mRNA (Struhl and Davis 1981a). The mRNA contains a nontranslated "leader" 41 nucleotides in length. As expected for a eukaryotic gene, translation begins at the 5′ proximal AUG codon. The *his3* DNA sequence contains some notable features (Struhl and Davis 1981b). Among them are a TATA box region located 37–51 bp upstream from the 5′ end of the mRNA (nucleotides −37 to −51), a perfect inverted repeat between positions −26 and −15, and an 8-bp sequence (ATGCCTCT) that with minor modifications is repeated around nucleotides −95, −135, −175, −215, and −235.

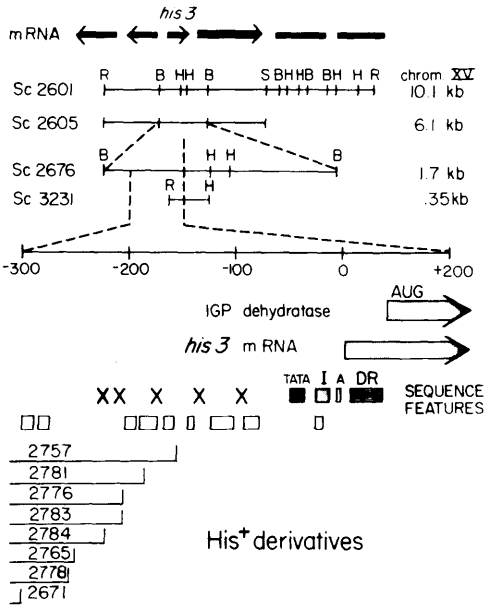


**Figure 2.** The *his3* gene. The top line indicates the location of mRNA species with respect to a 10.1-kb fragment (Sc2601) of chromosome XV containing the *his3* gene and with respect to various subcloned fragments. An expanded view of the region between positions −300 and +100 (measured with respect to the start of transcription) is shown in the middle of the figure. The nucleotide sequence of this region has been reported elsewhere (Struhl and Davis 1981b; K. Struhl, in prep.); part of it is shown in Fig. 3. Features of the gene are the translated region (coding for IGP dehydratase); the transcribed region; the direct repeat located in the RNA leader (DR); the run of dA residues in the coding strand just prior to the mRNA start (A⌄); the 12-bp inverted repeat (I); the TATA box, an 8-bp sequence that with minor modifications is repeated five times (×); and the regions that contain at least 90% pyrimidine residues in the coding strand (□). Shown below the sequence features are the mapping positions of various His⁺ derivatives.

## PROMOTER AND REGULATORY ELEMENTS

More than 100 derivatives of the cloned *his3* gene have been physically characterized (usually by DNA sequence analysis) and introduced back into yeast cells, as illustrated in Figure 1 (Struhl 1981a,b, 1982a,b; Struhl and Davis 1981b). Every derivative described here has the entire mRNA coding sequences; thus, functional defects caused by any particular mutation must involve DNA sequences outside the structural gene. Before the results are presented, it will be useful to define terms.

Promoter elements are essential for gene expression; thus, promoter mutations eliminate (or severely reduce) *his3* expression. Qualitatively, such mutations prevent (or slow) cell growth in the absence of histidine. When histidine is added, these mutations have no effect on cell growth, but IGP dehydratase levels are significantly below the basal level, if detectable.

Regulatory elements are responsible for changing the level of *his3* expression in response to amino acid starvation; thus, regulatory mutations cause IGP dehydratase to be produced at the same level under all growth conditions. In the experiments described here, strains harboring mutant genes are always grown in medium containing histidine to ensure that *his3* expression is gratuitous for cell growth. Inducing conditions were achieved by starving the cells for tryptophan.

The phenotypes of the mutations fall into clear patterns that lead to a simple and internally consistent picture of *his3* promoter and regulatory elements. However, it should be stressed that the conclusions represent the simplest formal description of the data. It is extremely difficult to disprove alternative explanations for the phenotype of any given mutation, and more complex models can be advanced to fit the data. Furthermore, genetic experiments are very useful for drawing inferences concerning molecular mechanisms, but proof of any particular model depends on knowing the biochemical properties of purified proteins and their target DNAs.

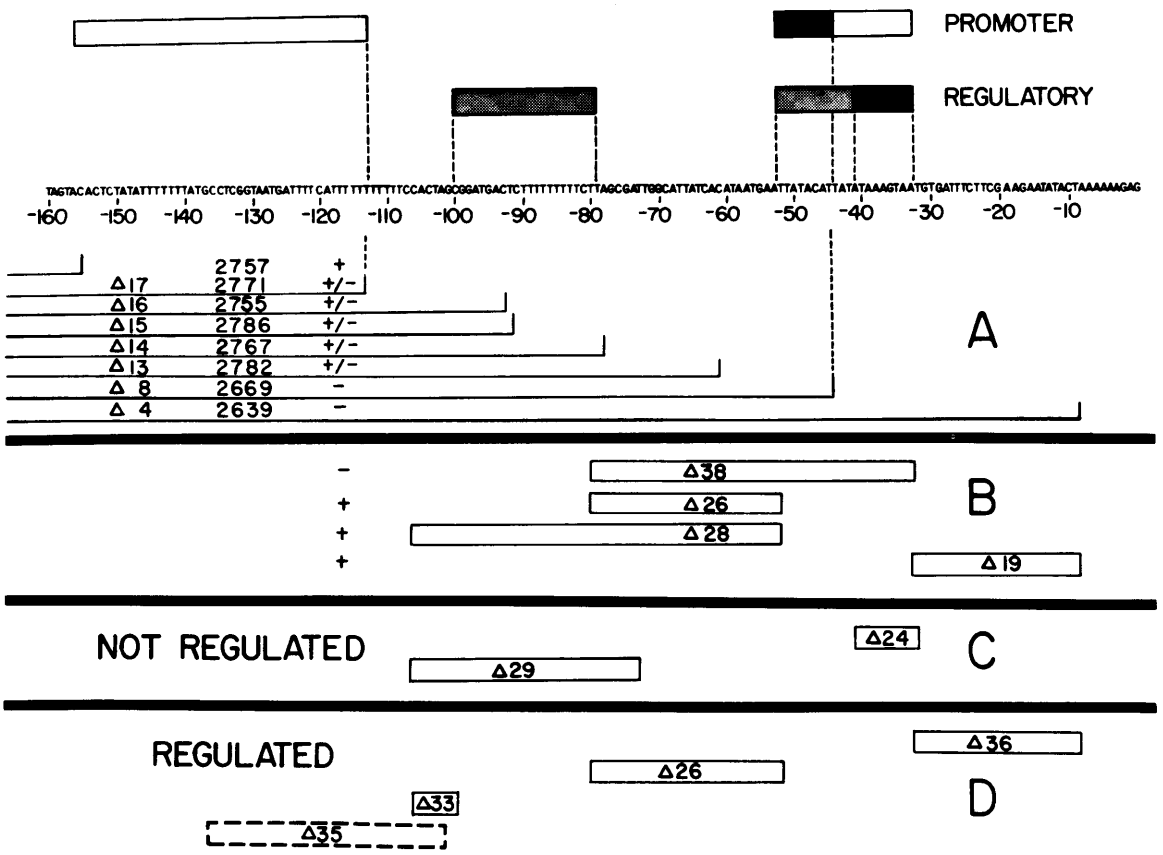


**Figure 3.** Promoter and regulatory elements. Open boxes on the top line indicate the extent of promoter elements, and shaded boxes on the second line show the locations of regulatory elements. Black regions within these boxes indicate sequences that are essential, but not necessarily sufficient, for function. These elements are placed with respect to the nucleotide sequence of the coding strand. (*A–D*) Mapping positions of key deletions used to locate the genetic elements (+ indicates wild-type phenotype, +/− indicates slow growth in the absence of histidine, and − indicates no detectable *his3* expression). Deletions in *A* come from the sequential deletion analysis used to define the upstream promoter element (Table 1), and those in *B* come from a matrix of internal deletions (Table 2). Mutations in *C* and *D* were used to map regulatory sites (Table 3); mutations in *C* do not induce *his3* expression in response to amino acid starvation, whereas those in *D* do.

### The Upstream Promoter Element

The first experiments were designed to determine the minimum contiguous DNA sequence necessary for a fully functional *his3* promoter (Struhl 1981a). A series of 24 deletion mutations that successively remove DNA sequences adjacent to the 5′ end of the mRNA coding region were isolated (Fig. 3A; Table 1). These mutations describe three distinct phenotypic classes, as evidenced by the growth properties they confer in the absence of histidine. Derivatives that retain more than 155 bp upstream from the start of *his3* transcription are phenotypically indistinguishable from the wild type; strains containing them grow at the normal rates and express *his3* at the usual levels. However, the more extensive deletions representing the other two classes do not show wild-type phenotypes. Those that retain between 60 and 115 upstream base pairs allow cells to grow slowly and have IGP dehydratase levels approximating 5–10% of the normal basal level. Those that contain less than 45 upstream base pairs prevent cell growth in the absence of histidine, and enzyme activity levels are not detectable (less than 3% of the wild-type level).

It is important to remember that deletion mutations cannot be viewed simply as "holes" in DNA but rather as fusions between two previously separated regions. Mapping the *his3* promoter by sequential deletion analysis depends on the assumption that the fused sequences (in this case from bacteriophage λ) do not have variable effects on expression. Here, such artifacts are unlikely because sequential deletion into the promoter region decreases expression in a monotonic fashion. Thus, even though each derivative has a unique sequence at the novel joint, and many have completely different fusion sequences, the phenotypes produce a simple and consistent pattern.

Two strong conclusions can be drawn from these results. First, the entire *his3* promoter is located within the 155-bp region adjacent to the 5′ end of the structural gene. Second, the most-upstream promoter element includes a sequence between positions −113 and −155.

### The *his3* Promoter Contains at Least Two Elements

The distinction between derivatives that express *his3* at reduced levels and those that fail to express the gene (see above) suggests that the promoter may contain two elements. To prove this, it is necessary to show that the

**Table 1.** Sequential Deletion Analysis

| *his3* fragment | Deletion allele | Endpoint | *his3* expression |
|---|---|---|---|
| Sc2773 | 9 | +6±2 | − |
| Sc2787 | 10 | +4±2 | − |
| Sc2763 | 11 | +2±2 | − |
| Sc2694 | 2 | −4±2 | − |
| Sc2695 | 3 | −4±2 | − |
| Sc2639 | 4 | −8 | − |
| Sc2670 | 5 | −8 | − |
| Sc2779 | 12 | −29 | − |
| Sc2667 | 6 | −35 | − |
| Sc2666 | 7 | −39 | − |
| Sc2669 | 8 | −43 | − |
| Sc2782 | 13 | −60±2 | +/− |
| Sc2767 | 14 | −78±2 | +/− |
| Sc2786 | 15 | −90±2 | +/− |
| Sc2755 | 16 | −92±2 | +/− |
| Sc2771 | 17 | −113±3 | +/− |
| Sc2757 | | −155±3 | + |
| Sc2781 | | −185±5 | + |
| Sc2776 | | −205±5 | + |
| Sc2783 | | −205±5 | + |
| Sc2784 | | −220±5 | + |
| Sc2765 | | −250±5 | + |
| Sc2778 | | −255±5 | + |

The cloned fragment numbers and deletion alleles (for non-wild-type derivatives) are listed. The *his3* endpoints are located with respect to the normal start of transcription (see Fig. 3); the other deletion endpoints map at various positions in bacteriophage λ. *his3* expression is determined by growth in the absence of histidine (+ indicates wild-type growth; +/− indicates slow growth; − indicates no growth). All wild-type derivatives produce wild-type levels of IGP dehydratase and regulate the expression properly. Data from Struhl (1981a).

**Table 2.** Deletion Matrix Analysis

| *his3* fragment | Deletion allele | Endpoints | *his3* expression |
|---|---|---|---|
| Sc2857 | 19 | −32 −32 | + |
| Sc2854 | 18 | −32 −20 | + |
| Sc2855 | 36 | −32 −8 | + |
| Sc2883 | 39 | −80 −8 | − |
| Sc2889 | 40 | −106 −8 | − |
| Sc2886 | 41 | −112 −8 | − |
| Sc2895 | 42 | −119 −8 | − |
| Sc2882 | 43 | −80 −20 | − |
| Sc2888 | 44 | −106 −20 | − |
| Sc2885 | 45 | −112 −20 | − |
| Sc2894 | 46 | −119 −20 | − |
| Sc2884 | 38 | −80 −32 | − |
| Sc2890 | 47 | −106 −32 | − |
| Sc2887 | 48 | −112 −32 | − |
| Sc2896 | 49 | −119 −32 | − |
| Sc3121 | 26 | −80 −53 | + |
| Sc3125 | 28 | −106 −53 | + |
| Sc3129 | 30 | −119 −53 | + |
| Sc3122 | 27 | −80 −73 | + |
| Sc3126 | 29 | −106 −73 | + |
| Sc3130 | 37 | −119 −73 | +/− |

For each cloned fragment, the upstream and downstream endpoints are listed from left to right. At the deletion break points of all these mutations is the *Eco*RI octanucleotide linker GGAATTCC. The measure of *his3* expression is the ability to grow in the absence of histidine. Strains that grow at the wild-type rate have an IGP dehydratase level of at least 20% of the wild type. Strains that fail to grow have levels of enzyme that are less than 3% of the wild type. Data from Struhl (1982a).

upstream element is not sufficient for wild-type levels of gene expression. Table 2 lists nine small deletion mutations, each of which retains the intact upstream element but fails to express the gene. Thus, it is possible to delete either of two separate regions of DNA and greatly reduce *his3* promoter function. This indicates that the promoter contains at least two elements, both of which are necessary, but neither of which is sufficient individually. This analysis does not eliminate the possibility of additional promoter elements.

## The TATA Box Region Can Serve as the Downstream Promoter Element

The smallest deletion mutation that retains the upstream element but eliminates promoter function (deletion 38; see Fig. 3B) indicates that the downstream region includes a sequence 32–80 bp prior to the transcriptional initiation site. The region between positions −37 and −51 includes three sequences that closely resemble TATAAA. These include TATACA (between −51 and −46), TATATA (between −44 and −39), and TATAAA (between −42 and −37). In this section, I show that something within the *his3* TATA box region is sufficient to constitute the downstream promoter element.

A matrix of deletion mutations was generated by pairwise joining of upstream and downstream promoter fragments (Struhl 1982b) (see Table 2). The downstream deletion endpoints map at positions −8, −20, −32, −53, and −73, and the upstream deletion endpoints map at positions −32, −80, −106, −112, and −119. The 21 mutations examined produce a simple phenotypic pattern. All 12 derivatives that lack the region between −32 and −52 fail to express the gene, whereas all 9 mutations that retain this region express the gene at some level. Furthermore, deletion 19 (which removes sequences between −31 and −8) and deletion 28 (which removes sequences −106 to −53) express *his3* at the wild-type level. Thus, the region between −32 and −52 suffices as the downstream promoter element (see Fig. 3B); henceforth, it will be called the TATA box element.

This localization of the downstream promoter element fits nicely with the results from the sequential deletion analysis (compare A and B in Fig. 3). There, the position was inferred to include a sequence between positions −45 and −60 (the endpoints of the relevant nonexpressing and partially expressing deletions). The combined results therefore indicate that the downstream element includes a sequence between −45 and −52 and it may include sequences as far downstream as −32. It is too early to determine which, if any, of the three TATA boxes is the key sequence.

There is one complication to this simple picture. Deletion mutations that destroy the downstream promoter element are relatively large (the smallest is deletion 38, which lacks nucleotides −32 to −80). Less extensive lesions that remove most or all of the TATA box region do not abolish *his3* expression. In fact, deletion

22 (missing nucleotides −31 to −50) expresses *his3* at the wild-type level, whereas deletion 25 (missing nucleotides −32 to −58) reduces expression only fivefold. This fivefold decrease is a qualitatively minor effect; such cells grow at wild-type rates in the absence of histidine.

The conclusions from these experiments are as follows. The downstream promoter element is absolutely required for gene expression. The TATA box region suffices as the downstream promoter element, but it is not absolutely required. In the absence of the TATA box region, other sequences (probably located somewhere between positions −59 and −80) can serve as the downstream element.

### *his3* Regulation Requires at Least Two Distinct Regions

*his3* regulatory sites are defined by comparing the DNA sequences of mutations that either regulate or fail to regulate gene expression as a function of amino acid starvation (Struhl 1982a) (Fig. 3C,D; Table 3). Obviously, this can only be performed for derivatives that express the gene at some level (i.e., have some promoter function).

The extent of deletion 24 indicates that a sequence between positions −32 and −41 is implicated in regulation. Deletion of this region can account for the regulatory defects of five other derivatives (deletions 20, 21, 22, 23, and 25). The structures of other mutations suggest a second regulatory element. All derivatives that delete the region between −80 and −100 also fail to regulate the gene properly.

**Table 3.** Analysis of Regulatory Sites

| *his3* fragment | Deletion allele | Endpoints | Basal | Induced |
|---|---|---|---|---|
| Sc2857 | 18 | −32 −32 | 0.7 | 2.0 |
| Sc2854 | 19 | −32 −21 | 0.5 | 1.5 |
| Sc3101 | 20 | −43 −31 | 1.1 | 1.0 |
| Sc3102 | 21 | −56 −6 | 0.2 | 0.2 |
| Sc3110 | 22 | −50 −31 | 1.2 | 1.3 |
| Sc3111 | 23 | −47 −24 | 0.8 | 1.0 |
| Sc3112 | 24 | −41 −32 | 1.1 | 1.4 |
| Sc3113 | 24 | −58 −32 | 0.2 | 0.2 |
| Sc3121 | 26 | −80 −53 | 0.9 | 2.7 |
| Sc3122 | 27 | −80 −73 | 0.6 | 1.5 |
| Sc3125 | 28 | −106 −53 | 1.1 | 1.0 |
| Sc3126 | 29 | −106 −73 | 0.2 | 0.3 |
| Sc3129 | 30 | −119 −53 | 0.2 | 0.1 |
| Sc3130 | 37 | −119 −73 | 0.05 | 0.05 |
| Sc3138 | 31 | −106 −109 | 1.0 | 3.1 |
| Sc3159 | 32 | −106 −102 | 0.8 | 2.9 |
| Sc3160 | 33 | −106 −100 | 1.2 | 3.2 |
| Sc3161 | 34 | −118 −107 | 0.2 | 0.6 |
| Sc3165 | 35 | −136 −101 | 0.5 | 1.2 |
| Sc2855 | 36 | −32 −8 | 0.6 | 1.8 |
| Wild type | | | 1.0 | 3.0 |

To examine mutations for their ability to regulate *his3* expression as a function of amino acid starvation, cells were grown to the middle of exponential phase in minimal medium containing uracil, histidine, tryptophan, and adenine (the requirements for KY137, see Fig. 1), harvested, and washed twice with water. One half of these cells were assayed for IGP dehydratase (Struhl and Davis 1977); those remaining were resuspended in growth medium lacking tryptophan and incubated for an additional 6–9 hr. Under these conditions, the cells undergo an average of one additional cell division before their growth stops as a consequence of tryptophan starvation. After the incubation, cells are harvested as above and assayed. The IGP dehydratase levels under basal and induced conditions are shown. All entries are normalized to the wild-type basal level of expression (defined as 1.0) and are accurate to ±10%. Data from Struhl (1982b).

Deletions do not indiscriminately abolish proper regulation. Sequences between −137 and −101, −80 and −53, and −32 and −21 can be removed without altering the ability of cells to induce the usual threefold higher levels of IGP dehydratase activity during amino acid starvation. Some of these derivatives have reduced basal levels of expression (as much as fivefold) (Struhl 1982a).

Thus, there seem to be at least two regulatory elements for *his3* expression that are separated by at least 30 bp. The downstream region maps between positions −32 and −52 and includes a sequence between positions −32 and −41. The upstream element maps somewhere between positions −80 and −100.

A comparison of the *his3* regulatory sequences with a co-regulated gene (*his4*) (Donoghue et al. 1982) indicates a striking resemblance. The *his3* sequence between positions −44 and −31, TATATAAAGTAATG, contains 12 out of 14 residues identical (with a single-base "loop-out") with the *his4* sequence TATATAATAGATATG, which is located 46–60 bp prior to the mRNA start point. For the upstream region, the *his3* sequence between positions −97 and −90, ATGACTCT, shares 7 out of 8 bp with the *his4* sequence ATGACTAT located between positions −86 and −79. Furthermore, highly conserved variants of this sequence are found at *his3* positions −139 to −132 (ATGCCTCG) and −178 to −171 (ATTACTCT) and *his4* positions −137 to −130 (GTGACTCA) and −181 to −174 (GTGACTCC). This striking sequence homology adds support to the view that these regions play an important role in *his3* regulation.

### Evidence for the Positive Regulation of *his3* Expression

Either of two simple models could account for the increased *his3* expression during conditions of amino acid starvation. One model posits that the gene interacts with positive regulatory factors that stimulate expression above the basal level. The second model posits that *his3* normally interacts with negative regulatory factors that must be removed during conditions of amino acid starvation.

Although genetic experiments can never definitively eliminate either of these models, the present evidence favors positive regulation of *his3* expression. By definition, mutations of a positive regulatory site would result in basal levels of expression under all growth conditions. Indeed, five mutations (deletions 20, 22, 23, 24, and 28) confer such a phenotype. Because the other four regulatory mutations result in constitutive expression below the basal level, their phenotypes cannot be assessed in terms of positive or negative control.

On the other hand, there is no evidence that supports the negative-control model. Mutations of a negative-control site would result in "induced" *his3* levels under all growth conditions; these have yet to be isolated. Such mutations might be difficult to isolate if the postulated negative-control element maps close to a promoter element. For example, mutations that always express *his3* at the basal level could result from elimination of a negative-control site and a simultaneous threefold reduction in the IGP dehydratase basal level. However, it seems very unlikely that five deletion mutations fortuitously result in the same quantitatively minor promoter defect, especially considering that nine deletions with full or partially reduced *his3* levels under normal conditions regulate the gene properly.

### Evidence That Regulatory Elements Can Be Separated from Promoter Elements

As described earlier, the distinction between promoter and regulatory elements is conceptual. In physical terms, they can be inextricably linked, partially overlapping, or totally separate.

The derivatives described above define two regulatory regions. At least one alteration of each region expresses *his3* at the basal level under all conditions. Thus, it seems possible to destroy either regulatory site without inactivating promoter function. Conversely, in a number of mutations, promoter function is reduced as much as fivefold while regulatory function is apparently unaffected. Thus, promoter and regulatory functions can be separated by mutation.

It is possible to explain this functional separation in terms of the locations of the promoter and regulatory elements along the *his3* gene (see Fig. 3). Nevertheless, it should be stressed that the relationship between all of these genetic elements (particularly those at the TATA box) remains to be clarified.

## PROPERTIES OF *his3* PROMOTER ELEMENTS

In the previous sections, I identified two promoter elements and two regulatory elements. Here, I describe experiments that further characterize the TATA promoter element and its relationship to the upstream element. The results strongly suggest that this eukaryotic promoter is radically different from prokaryotic counterparts.

### The TATA Box Element Requires Specific Sequences

Deletion of the TATA promoter element could eliminate gene expression for either of two reasons. The simplest interpretation is that the mutations delete critical sequences. Another explanation is that they alter important spacing relationships between other sequences. These possibilities have been distinguished by analyzing mutations in which *his3* sequences have been replaced by "foreign" DNA (Struhl 1982b). All of the deletion mutants listed in Table 2 conveniently have an *Eco*RI site at the deletion break points. A 31-bp DNA fragment (or multiple tandem copies of it) from coliphage M13 was inserted into the "space" of four different deletion mutants (Fig. 4). Two of the original deletions (38 and 47) lack the TATA promoter element, whereas the analogous pair (deletions 26 and 28) contain this region. The phenotypes of the derivatives analyzed in Table 4 give the following clear pattern. All derivatives containing the M13 fragment in the + orientation express *his3*. Of derivatives containing the M13 fragment in the − orientation, only those with the *his3* TATA box region express the gene. Thus, it seems that the + orientation can serve as the downstream promoter element, whereas the − orientation cannot. In this regard it is interesting to note that the + orientation contains a sequence resembling the TATA box (TGTAAA) and the − orientation does not. Of more importance, the − orientation does not qualitatively affect promoter activity when present in derivatives containing the *his3* TATA box. Thus, it appears to behave as "neutral" DNA.

Therefore, the absolute requirement for the downstream promoter element can be satisfied by the *his3* TATA box region or by a sequence that resembles it. Derivatives differing only by the orientation of the M13 fragment have identical spacing relationships but opposite phenotypes. Thus, deletions lacking the downstream promoter element do not express the gene because they lack specific sequences.

### The Spacing between Promoter Elements Is Not Critical

In the wild-type *his3* gene, the two promoter elements are located approximately 90 bp apart. In the 13 deriva-

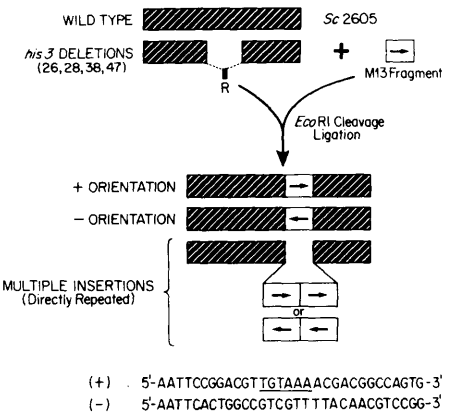


**Figure 4.** Replacement of *his3* sequences by M13 DNA. A sample original *his3* deletion mutant is indicated by the solid bars (wild-type sequences), empty space (deleted DNA), and the *Eco*RI site (the sequence GGAATTCC) at the site of deletion. Ligation of *Eco*RI-cleaved *his3* mutant DNA and the M13 fragment results in the structures below the arrow. The 31-bp M13 fragment is drawn as an open box with an arrow (pointing rightward for the + orientation and leftward for the − orientation). Many derivatives have multiple tandem insertions, all oriented in the same direction. The sequences of the sense strands of the + and − orientations are shown, and a possible TATA box is underlined.

**Table 4.** Replacement of *his3* Sequences by M13 DNA

| Allele | Endpoints | M13 inserts | Spacing change | *his3* expression |
|---|---|---|---|---|
| 26 | −80 −53 | 0 | −20 | + |
| 50 | −80 −53 | 1+ | +11 | + |
| 51 | −80 −53 | 1− | +11 | + |
| 52 | −80 −53 | 2+ | +42 | + |
| 53 | −80 −53 | 3− | +73 | + |
| 54 | −80 −53 | 5+ | +135 | + |
| 38 | −80 −32 | 0 | −41 | − |
| 55 | −80 −32 | 1+ | −10 | + |
| 56 | −80 −32 | 1− | −10 | − |
| 57 | −80 −32 | 2+ | +21 | + |
| 58 | −80 −32 | 2− | +21 | − |
| 59 | −80 −32 | 3+ | +52 | + |
| 60 | −80 −32 | 3− | +52 | − |
| 61 | −80 −32 | 4+ | +83 | + |
| 28 | −106 −53 | 0 | −46 | + |
| 62 | −106 −53 | 1+ | −15 | + |
| 63 | −106 −53 | 1− | −15 | + |
| 64 | −106 −53 | 2+ | +16 | + |
| 65 | −106 −53 | 3+ | +47 | + |
| 47 | −106 −32 | 0 | −67 | − |
| 66 | −106 −32 | 1+ | −36 | + |
| 67 | −106 −32 | 1− | −36 | − |
| 68 | −106 −32 | 2+ | −5 | + |

The principles by which these mutations were constructed are shown in Fig. 4. For each allele, the original endpoints and the number and orientation of M13 fragments are indicated (see text for definition of + and − orientations). The spacing change refers to the number of nucleotides that are added to, or subtracted from, the wild-type gene.

tives that contain both promoter elements (and hence express the gene), this distance varies from about 40 bp to 160 bp. Most importantly, the insertion of apparently neutral DNA (the − orientation of the M13 fragment) between the *his3* elements has little, if any, effect on function.

The presence of both elements is therefore necessary and sufficient for promoter function, whereas the spacing between them is apparently unimportant. Deletion of either region results in severe defects because specific sequences are missing, and not because the relative positions of any two regions of DNA are changed. These conclusions are completely consistent with the phenotypes of all derivatives analyzed to date.

### Upstream Promoter Elements and Long-range Effects on Gene Expression

The ability of the upstream promoter element to act at a seemingly long distance is reminiscent of position effects. Previously, I showed that sequences located more than 300 bp from the structural gene can affect gene expression (Struhl 1981b). Specifically, inversion of certain DNA fragments simultaneously altered both *his3* and *trp1* expression, even though the inversion break points are located more than 300 bp from the genes. In these cases, the *trp1* and *his3* derivatives lack the normal upstream promoter element. Thus, these position effects could be explained by the presence or absence of an upstream element located far from the gene.

### Inferences Concerning the Molecular Nature of the *his3* Promoter

Our current understanding of promoters derives from studies of prokaryotic genes. Extensive genetic, biochemical, and physical analyses indicate that a prokaryotic promoter is a transcriptionally competent, RNA-polymerase-binding site. First, the two regions defined by genetic means as promoter elements (the −10 and −35 regions) are required for RNA polymerase binding and subsequent transcription in vitro. Second, these regions and the actual site of transcriptional initiation are in close physical contact with the enzyme. Third, mutations that change the spacing between the two promoter elements even by a single base pair can have major effects on promoter function.

Although there is no direct evidence, the following observations lead to the general assumption that eukaryotic RNA polymerase II interacts specifically with the TATA box. First, the TATA box is highly conserved in front of essentially all eukaryotic genes (for review, see Breathnach and Chambon 1981). Second, its DNA sequence is similar to the prokaryotic −10 region, the key determinant of *E. coli* RNA polymerase binding. Third, it is essential for transcription in crude in vitro systems (Mathis and Chambon 1981; Grosveld et al. 1981). Finally, the distance between the TATA box and the start of transcription is relatively constant (Gluzman et al. 1980; Grosveld et al. 1981).

The properties of the *his3* downstream promoter element provide additional support for this view: It is absolutely required for gene expression, it requires orientation-specific DNA sequences, and it can be defined by the *his3* TATA box or by a related M13 sequence.

However, in striking contrast to the properties of prokaryotic promoters, the spacing between the *his3* promoter elements can be varied with little effect on gene expression. This is important because it is extremely difficult to imagine how any enzyme could interact simultaneously with two DNA regions that have no fixed positional relationship to each other. Thus, the *his3* promoter elements probably interact with different proteins (or different conformations of the same protein). If the TATA box is part of a transcriptionally competent RNA-polymerase-binding site (analogous to the prokaryotic −10 region), then it follows that the upstream element is not part of this site.

## RELATIONSHIP BETWEEN CHROMATIN STRUCTURE AND GENE EXPRESSION

How could the upstream promoter element act at a relatively far and variable distance? There are two classes of models: One possibility is that the upstream region serves as an entry site for RNA polymerase from which the enzyme moves to its transcriptionally competent binding site (the TATA box?). The other possibility is that proteins interact with the upstream element such that polymerase can bind to the TATA box. One specific example of the latter class is that such proteins create an

open domain of chromatin. Both models explain the unusual property of the upstream promoter element, and it is difficult to distinguish between them.

In their simplest forms, the models predict that different regions of DNA in chromatin should be accessible to RNA polymerase. The entry-site model, by definition, predicts that the upstream region is preferentially accessible. The chromatin-domain model implies that the TATA box should be accessible, but only if the upstream promoter region is present. To address this issue, I examined the accessibility of *his3* DNA sequences in nuclear chromatin by using micrococcal nuclease as a structural probe (K. Struhl, in prep.).

### The TATA Box Region Is Preferentially Cleaved by Micrococcal Nuclease

Micrococcal nuclease preferentially digests DNA located in the spacer regions between nucleosome cores (for review, see Kornberg 1977). Treating yeast nuclei with this enzyme demonstrates that the chromatin structure at the *his3* locus is indistinguishable from that of total genomic DNA. The size and shape of hybridization bands shown in Figure 5 (lanes A–D) indicate that the spacing between nucleosome cores averages 160 bp with a deviation of about 5 bp.

Specific sites of cleavage by micrococcal nuclease are mapped relative to a *Hin*dIII site within the structural gene located 330 bp from the point of transcriptional initiation (see Fig. 1). From the results presented in Figure 5 (lanes F–J), four (possibly five) preferentially cleaved regions are apparent; these map 210, 370, (530), 700, and 1900 bp from the *Hin*dIII marker. The sites that are 700 bp and 1900 bp from the *Hin*dIII marker map near sequences encoding the 5′ ends of genes adjacent to *his3*. Those located 210 bp and 370 bp from the marker correspond to *his3* mRNA coding sequences and to the *his3* TATA box region, respectively. Fine-structure mapping (K. Struhl, in prep.) indicates that the latter region maps between nucleotides −47 and −37, i.e., entirely within the TATA box.

It is very likely that these cleavage sites reflect features of chromatin structure and not simply the enzyme's known preference for particular DNA sequences (Dingwall et al. 1981). The identical mapping technique applied to purified DNA reveals 15 sites located 200–2000 bp from the *Hin*dIII marker (Fig. 5, lanes L,M). Only three of these (370, 700, and 1900) are recognized in chromatin. Thus, although DNA sequences at these locations are recognized preferentially by micrococcal nuclease per se, they are selectively cleaved in the chromatin structure. Moreover, the cleavage site in the *his3* structural gene is observed in chromatin but not in naked DNA.

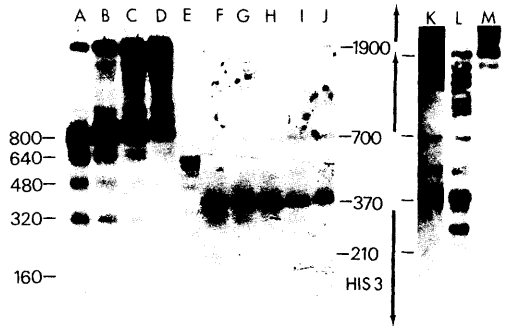


**Figure 5.** Chromatin structure at the *his3* locus. (*A–D*) Nuclei from wild-type yeast spheroplasts were incubated at 37°C for 10 min with various amounts of micrococcal nuclease, and DNA was purified, electrophoretically separated in 2% agarose, transferred to nitrocellulose, and challenged for hybridization with $^{32}$P-labeled Sc3231 DNA (K. Struhl, in prep.) (see Fig. 2). (*F–J*) Same procedure as in *A–D*, except that the purified DNA was cleaved to completion with *Hin*dIII prior to electrophoresis. (*E*) Size standard prepared by cleavage of pBR322 DNA with *Hae*III. *K–M* come from a different gel: Lane *K* is equivalent to lane *I*, whereas in lanes *L* and *M*, naked DNA from λgt4-Sc2601 was cleaved first with various concentrations of micrococcal nuclease and then with *Hin*dIII. Size markings at the left indicate the positions of nucleosome monomers, dimers, etc. Size markings at the right indicate bands corresponding to specific micrococcal nuclease cleavage sites. The locations of *his3* and adjacent mRNA transcripts are shown with respect to the size markings.

### Nuclease Cleavage at the TATA Box Depends on the Upstream Promoter Element

The chromatin structure at the *his3* locus was examined in yeast strains in which the wild-type *his3* gene is replaced by some of the derivatives shown in Figure 3A. All six strains have the entire TATA box region, but to various extents, they delete sequences further upstream. In the five cases in which the upstream element is deleted, nuclease cleavage at the TATA box is reduced significantly (Fig. 6). On the contrary, the derivative that retains the upstream element is indistinguishable from the wild-type gene with respect to chromatin structure at the TATA box. Because the TATA box region is identical in DNA sequence in all these strains, the differences in nuclease cleavage reflect structural dif-

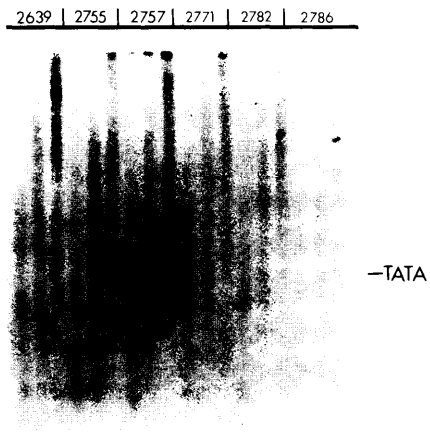


**Figure 6.** Chromatin structure of promoter mutants. The procedure described in Fig. 5 was used. For each deletion allele tested (listed at the top of the figure), three concentrations of micrococcal nuclease were used. The location of the band indicating cleavage at the TATA box is shown.

ferences. Thus, this experiment eliminates possible artifacts due to inherent specificity of micrococcal nuclease.

The results therefore constitute evidence for a correlation between chromatin structure and gene expression. Specifically, the presence of the upstream promoter element is necessary for nuclease action at the downstream promoter element.

### Chromatin Structure Is Not Correlated with Regulation of *his3* Expression

As stated in the introduction, *his3* expression is increased when yeast cells are starved for amino acids. These conditions, however, do not result in a detectable change in chromatin structure at the TATA box (Fig. 7). Furthermore, when proper *his3* regulation is prevented by mutation, normal chromatin structure is observed under conditions of feast and famine. This result was obtained for two *his3* regulatory mutations (deletions 22 and 24) as well as for mutations in other genes (*aas2* and *tra3*) that alter general amino acid control (and *his3* regulation as a consequence). *aas2* prevents *his3* induction and *tra3* causes induced levels of *his3* expression under all conditions (Struhl 1982a).

## SOME COMMENTS ON THE RELATIONSHIP BETWEEN GENE STRUCTURE AND FUNCTION

The standard interpretation of the results obtained with micrococcal nuclease is that the *his3* TATA box lies preferentially in spacer regions between nucleosome cores. If this interpretation is correct, this enzyme should cleave chromatin at positions displaced from the TATA box by the length of nucleosome units (160 bp; see Fig. 5). Indeed, the site in the structural gene is located the expected distance from the TATA box. When considering cleavage one nucleosome unit upstream from the TATA box (i.e., 530 bp from the *Hin*dIII marker), it is difficult to distinguish the desired band from the broad band corresponding to nucleosome trimers (480 bp on average). Nevertheless, in many experiments (Figs. 6 and 7), a band of 530 bp is seen when micrococcal nuclease is present at relatively low concentrations (conditions that produce low amounts of nucleosome trimers).

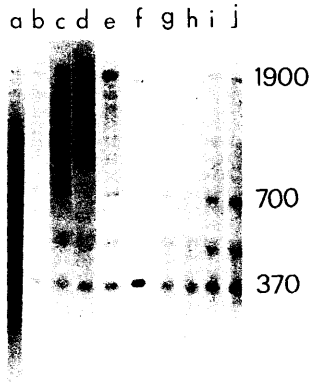


**Figure 7.** Chromatin structure and gene regulation. Each pair of lanes represents micrococcal nuclease treatment of chromatin from normally growing cells ("−" lanes *a,c,e,g,i*) and from starved cells ("+" lanes *b,d,f,h, j*). Cells were starved for histidine by addition to 10 mM aminotriazole, a competitive inhibitor of yeast IGP dehydratase (Klopotowski and Wiater 1965). The strains contained the following alleles: *tra3*-1 (*a,b*), *aas2*-5039 (*c,d*), wild type (*e,f*), *his3*-Δ22 (*g,h*), and *his3*-Δ24 (*i,j*).

However, even if this specific interpretation is not correct, preferential micrococcal nuclease cleavage should measure enzyme access to DNA in chromatin. Thus, in the intact cell, it seems likely that the TATA box should be exposed to the action of nuclear proteins, just as it is exposed to enzyme added exogenously to isolated nuclei. Because this region is important for promoter and regulatory functions, such proteins are likely to include RNA polymerase II and possible regulatory factors.

At the beginning of this section, I described "chromatin domain" and "RNA polymerase entry site" models that would explain the unusual and striking features of the upstream promoter element. The observations (1) that micrococcal nuclease cleaves chromatin at the TATA box element but not at the upstream element and (2) that sensitivity at the TATA box depends on the presence of the upstream element are consistent with the naive predictions of the chromatin domain model. A simple picture of this model is that the upstream element is necessary for the "correct" organization of chromatin such that RNA polymerase II can "find" its favored binding site (the TATA box?).

However, it is important to state that the experimentally observed correlation between promoter function and chromatin structure does not establish any particular cause and effect relationship. For example, the correlation may not reflect the presence of an upstream element per se. It is entirely possible that the process of transcription itself (which depends on the upstream element) could alter the chromatin structure. The simplest relationship with transcription seems unlikely because induction of mRNA synthesis during amino acid starvation does not grossly alter the structure. To resolve this issue, it will be important to examine the chromatin structure of more mutants.

The observation that chromatin structure can be correlated with promoter function, but not with gene regulation, seems contrary to the results obtained with other genes. In particular, alterations in structure are associated typically with genes that are undergoing "developmentally regulated" changes in gene expression. However, such genes are controlled by an internal genetic program, whereas *his3* expression is responsive only to transient changes in the environmental conditions.

One way to explain these observations is to postulate that the "correct" chromatin structure is a necessary (pre)condition for transcription. Thus, the transcriptional inactivity of zygotes and early embryos may be due to an "incorrect" chromatin structure, and the developmentally regulated expression of genes may in-

volve a change to the "correct" structure in the appropriate cells at the appropriate times. Unlike multicellular embryos, yeast cell division requires transcription, and most yeast genes are transcribed during each cell cycle. For these reasons, it is possible that the vast majority of yeast genes are never in the "inactive" structure. Thus, inactive chromatin structures can be observed only in mutated derivatives of a given gene. According to the definitions used in this paper, chromatin structure would be correlated with promoter function, i.e., expression of the gene. On the other hand, regulation would involve differential interaction of proteins with already "correctly organized" chromatin. Such a regulatory scheme is ideal for rapid changes in transcription as a function of external stimuli. In this light, it is interesting that regulation of yeast mating-type genes is associated with changes in chromatin structure (K. Nasmyth, unpubl.; see Abraham et al., this volume). These genes, unlike *his3*, do not respond to environmental changes but rather constitute the key determinants of an internal genetic program, the control of cell type.

## CONCLUSIONS

This paper provides an initial description of the relationship between the chromatin structure and the promoter/regulatory elements of the yeast *his3* gene. The principal conclusions are listed below.

1. The *his3* promoter contains at least two distinct elements; these are located 113–155 bp and 32–52 bp from the start of transcription.
2. Regulation of *his3* expression depends on two regions (32–41 bp and 80–100 bp from the mRNA start) which have properties consistent with their being sites of positive regulation.
3. The spacing between the *his3* promoter elements is apparently unimportant for function, suggesting that the promoter, unlike prokaryotic counterparts, is not a transcriptionally competent RNA-polymerase-binding site.
4. In chromatin, the *his3* TATA box is preferentially sensitive to micrococcal nuclease cleavage; this cleavage depends on the presence of the upstream promoter element.

## ACKNOWLEDGMENT

This work was supported by grant GM-30186 from the National Institute of General Medical Sciences.

## REFERENCES

BREATHNACH, R. and P. CHAMBON. 1981. Organization and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50:** 349.

DINGWALL, C., G.P. LOMONOSOFF, and R. LASKEY. 1981. High sequence specificity of micrococcal nuclease. *Nucleic Acids Res.* **9:** 2659.

DONOGHUE, T., P. FARABAUGH, and G.R. FINK. 1982. The nucleotide sequence of the *his4* gene of yeast. *Gene* **18:** 47.

FINK, G.R. 1964. Gene-enzyme relationships in histidine biosynthesis in yeast. *Science* **146:** 525.

GLUZMAN, Y., J.F. SAMBROOK, and R.J. FRISQUE. 1980. Expression of early genes of origin-defective mutants of simian virus 40. *Proc. Natl. Acad. Sci.* **77:** 3898.

GROSVELD, G.C., C.K. SHEWMAKER, P. JAT, and R.A. FLAVELL. 1981. Location of DNA sequences necessary for transcription of the rabbit β-globin gene *in vitro*. *Cell* **25:** 215.

HINNEN, A., J.B. HICKS, and G.R. FINK. 1978. Transformation of yeast. *Proc. Natl. Acad. Sci.* **75:** 1929.

KLOPOTOWSKI, T. and A. WIATER. 1965. Synergism of aminotriazole and phosphate on the inhibition of yeast imidiazoleglycerolphosphate dehydratase. *Arch. Biochem. Biophys.* **112:** 562.

KORNBERG, R.D. 1977. Structure of chromatin. *Annu. Rev. Biochem.* **46:** 931.

MATHIS, D. and P. CHAMBON. 1981. The SV40 early region TATA box is required for accurate *in vitro* initiation of transcription. *Nature* **290:** 310.

SCHERER, S. and R.W. DAVIS. 1979. Replacement of chromosomal segments with altered DNA sequences constructed *in vitro*. *Proc. Natl. Acad. Sci.* **76:** 4951.

STRUHL, K. 1981a. Deletion mapping a eukaryotic promoter. *Proc. Natl. Acad. Sci.* **78:** 4461.

———. 1981b. Position effects in *Saccharomyces cerevisiae*. *J. Mol. Biol.* **152:** 569.

———. 1982a. Regulatory sites for *his3* gene expression in yeast. *Nature* **300:** 284.

———. 1982b. The yeast *his3* promoter contains at least two distinct elements. *Proc. Natl. Acad. Sci.* **79:** 7385.

STRUHL, K. and R.W. DAVIS. 1977. Production of functional eukaryotic enzyme in *Escherichia coli*: Cloning and expression of the yeast structural gene for imidiazoleglycerolphosphate dehydratase (*his3*). *Proc. Natl. Acad. Sci.* **74:** 5255.

———. 1980. A physical, genetic, and transcriptional map of the cloned *his3* gene region of *Saccharomyces cerevisiae*. *J. Mol. Biol.* **136:** 309.

———. 1981a. Transcription of the *his3* gene region in *Saccharomyces cerevisise*. *J. Mol. Biol.* **152:** 535.

———. 1981b. Promoter mutants of the yeast *his3* gene. *J. Mol. Biol.* **152:** 553.

STRUHL, K., J.R. CAMERON, and R.W. DAVIS. 1976. Functional genetic expression of eukaryotic DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **73:** 1471.

STRUHL, K., D.T. STINCHCOMB, S. SCHERER, and R.W. DAVIS. 1979. High frequency transformation of yeast: Autonomous replication of hybrid DNA molecules. *Proc. Natl. Acad. Sci.* **76:** 1035.

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.