Reprinted from
Cold Spring Harbor Symposia on Quantitative Biology, Volume L

Constitutive and Coordinately Regulated Transcription of Yeast Genes: Promoter Elements, Positive and Negative Regulatory Sites, and DNA Binding Proteins

K. STRUHL, W. CHEN, D.E. HILL, I.A. HOPE, AND M.A. OETTINGER
Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02115

The yeast genome specifies approximately 5000 protein coding genes that are densely clustered on 16 linear chromosomes. In the form of nuclear chromatin, these genes are transcribed by RNA polymerase II from discrete initiation sites. The average yeast gene is transcribed about five to ten times during each cell cycle which results in a steady-state mRNA level of one to two molecules per cell. However, some genes are transcribed constitutively at considerably different rates, whereas other genes are transcribed at variable rates depending on the physiological circumstances. Such regulated expression is achieved either by positive factors that increase transcription above a basal level or by negative factors that repress transcription below a basal level (for a general review of yeast promoters, see Struhl 1985a).

The transcriptional properties of the yeast genes investigated in these experiments are illustrated in Figure 1. *pet56*, *his3*, and *ded1* are adjacent but unrelated genes located on chromosome XV (Struhl and Davis 1981; Struhl 1985b). Under normal growth conditions, *his3* and *pet56* are transcribed at average levels, whereas *ded1* is transcribed at a fivefold higher level. Under conditions of amino acid starvation, *his3* transcription increases by a factor of 5, whereas the other genes are not affected (Struhl and Davis 1981). This physiological circumstance also results in the coordinate increase in the expression of many amino acid biosynthetic genes which are scattered around the genome (for review, see Jones and Fink 1982). The coordinately regulated *gal1-gal10* genes are closely linked on chromosome II, and they are transcribed divergently (St. John and Davis 1981). In glucose medium, the *gal* genes are not expressed, whereas in galactose medium, they each account for about 1% of the steady-state mRNA (40 molecules/cell). Thus, these genes provide representative examples of different kinds of yeast promoters.

Experimentally, mutant and hybrid promoters are created by a variety in vitro manipulations of cloned DNA, structurally characterized by DNA sequencing, and introduced back into yeast cells such that there is one copy per cell exactly at the normal chromosomal location; in many experiments, the DNAs to be tested directly replace the normal chromosomal sequence. In this way, mutations constructed in vitro are examined under true in vivo conditions for their effects on transcription and chromatin structure (for a review of this methodology, see Struhl 1983b). The transcriptional properties of these promoters are assayed by a quantitative 5′ mapping procedure which provides information concerning the level and the initiation sites.

This paper describes the properties of *cis*-acting elements that determine the accuracy, level, and regulation of transcription as well as the DNA binding properties of the *gcn4* positive regulatory protein. It also provides evidence suggesting that constitutive and regulatory promoters may operate by different molecular mechanisms.

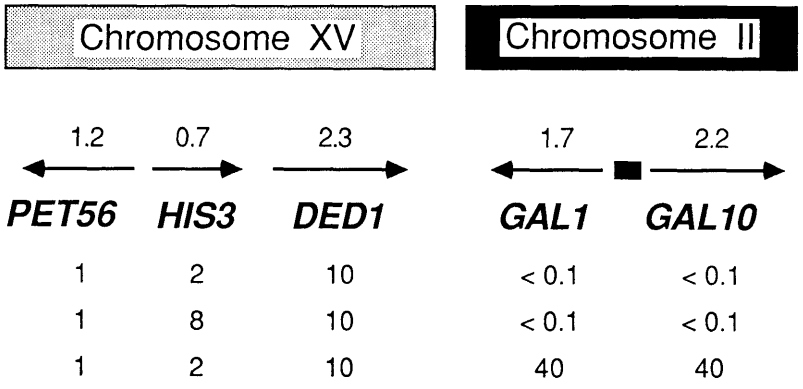

Figure 1. Structural and transcriptional properties of the *pet56-his3-ded1* and *gal1-gal10* genes regions. The *pet56-his3-ded1* region is located on chromosome XV (gray bar) and the *gal1-gal10* region is located on chromosome II (black bar). The size (kb), directionality, and relative order of the transcripts are indicated. The transcription levels observed during growth in normal conditions, amino acid starvation conditions, or growth using galactose instead of glucose as a sole source of carbon are indicated (in mRNA molecules/cell at the steady state).

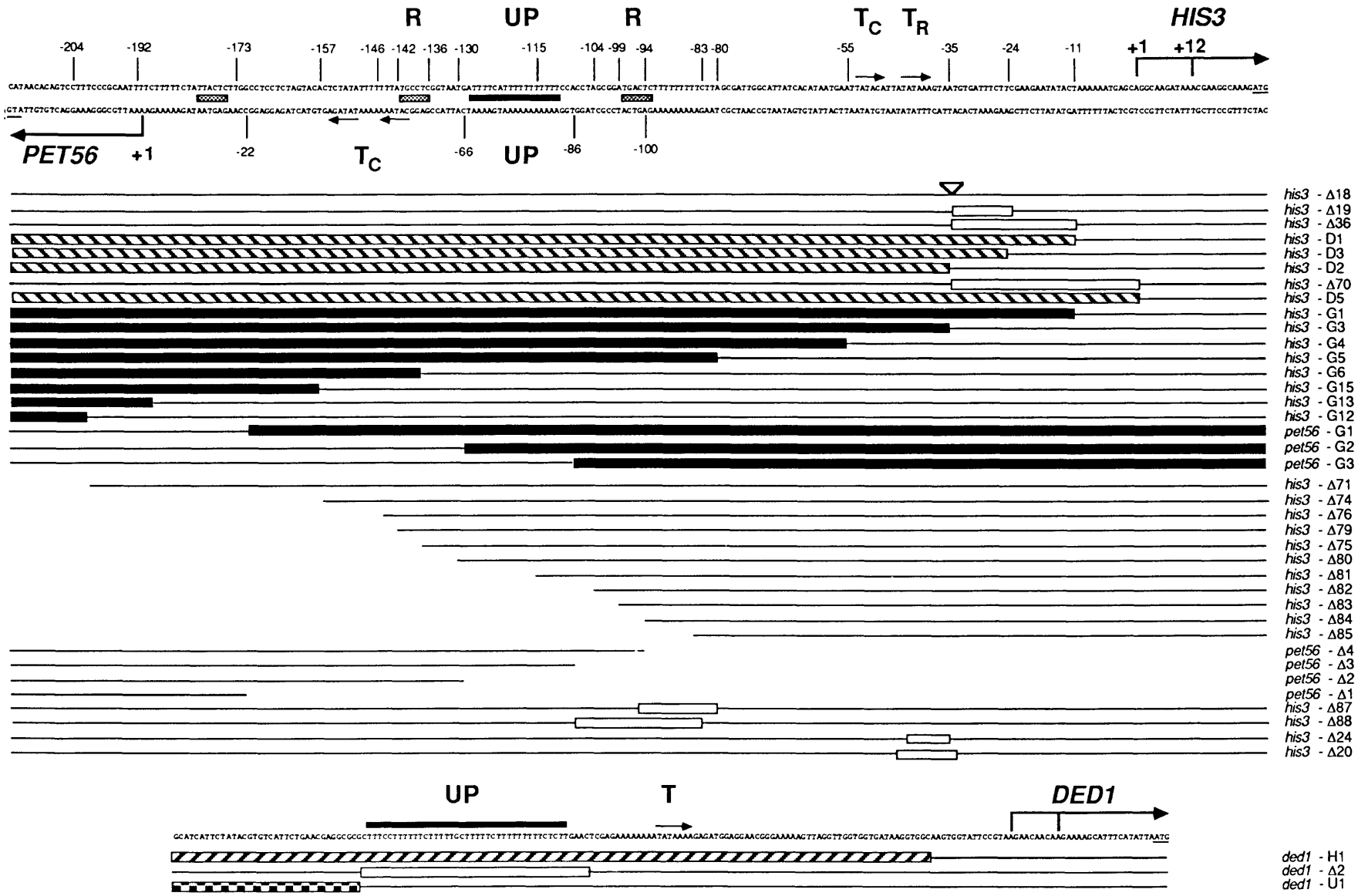

Figure 2. (*See facing page for legend.*)

ACCURATE INITIATION

Transcription of the wild-type *his3* gene is initiated with equal efficiency at two sites (defined as +1 and +12). At least two distinct elements, called the upstream and TATA sequences, are necessary for wild-type levels of transcription (Struhl 1981, 1982a). Deletion of either promoter element markedly reduces the level of expression. On the other hand, deletions that remove sequences between or adjacent to these elements do not significantly alter the basal expression level.

Which, if any, of these promoter elements determines where transcription begins? Analysis of mutant and hybrid promoters strongly suggests that unlike the situation in higher eukaryotes, accurate yeast mRNA initiation is determined primarily by an "initiator" element, not by the distance from the TATA sequence (Chen and Struhl 1985; Fig. 2; Table 1).

First, when the spacing between the *his3* TATA element and initiation sites is altered by an 8-bp insertion (Δ18), or by 3- and 16-bp deletions (Δ19, Δ36), *his3* transcription is indistinguishable from the wild-type gene. This indicates that the precise distance is relatively unimportant, and that something else must impart initiation specificity.

Second, hybrids between the *his3* and *ded1* promoter regions indicate that accurate initiation occurs even when transcription depends on "foreign" elements that are located at different positions from the "normal" elements. When the *ded1* initiation region downstream from −15 is fused to the *his3* TATA and upstream elements (*ded1*-H1), the initiation pattern is indistinguishable from the wild-type *ded1* gene. In this case, the *his3* TATA element is only 33 bp from the *ded1* +1 site as compared with the 60-bp separation in the wild-type *ded1* gene. Conversely, hybrid promoters (*his3*-D1 through D4) consisting of the *ded1* upstream and TATA elements fused to a series of *his3* segments containing the initiation region all initiate transcription at the +1 site. The distances between the *ded1* TATA element and *his3* initiation site range from 69 to 93 bp whereas the spacing in the wild-type *his3* promoter is 45 bp. These experiments indicate that the signal for accurate initiation depends on specific sequences downstream of the TATA element. In particular *his3* sequences upstream of −11 and *ded1* sequences upstream of −15 are unnecessary for accurate initiation.

Third, when the *his3* or *ded1* promoter elements are fused to *his3* nucleotide +2 (*his3*-Δ70 and *his3*-D5), transcription initiated from a position that is equivalent to +1 is not observed. This cannot be explained by a "closeness" effect because in *his3*-D5, the TATA element is 51 bp from the +1 equivalent site. Moreover, both derivatives produce the same initiation pattern, even though the TATA elements are located at different distances from the observed +12 and +22 mRNA start sites.

Although there is no precise spacing relationship between the TATA element and RNA start sites, there are distance limits over which a TATA element can act. The maximum distance appears to be about 90–100 bp. The best evidence for this comes from hybrid promoters (*his3*-D2, D3, D4) in which transcription from +12 is reduced compared with that from +1, and aberrant initiation from −3 to −6 is observed. The simplest interpretation for these observations is that the *ded1* TATA element is too far from the +12 site for efficient transcription. In *his3*-D2, the distance is 103 bp from the +12 site whereas it is only 92 bp from the +1 site, which is efficiently used. In *his3*-D4, the relevant distances are 92 and 81 bp. The minimal distance is more difficult to define. In *his3*-Δ36, the distance from the "best" sequence (TATAAA) is only 28 bp, similar sequences (TATATA and TATACA) are 30 and 37 bp away and lie within the region defined by functional criteria to be the TATA element (Struhl 1982a, 1984). Thus, it seems that the TATA element can be located anywhere between 30 and 90 bp from the initiation site with minimal functional consequences.

POSITIVE AND NEGATIVE REGULATORY ELEMENTS MEDIATE TRANSCRIPTIONAL CONTROL OF THE *gal1,10* GENES

The basic experiment is to fuse a 365-bp DNA fragment encoding the *gal* upstream regulatory region (Guarente et al. 1982; Guarente and Mason 1983) to a set of *his3* derivatives that systematically delete sequences upstream from the mRNA initiation sites (Struhl 1984, 1985c; Fig. 2). This DNA segment is responsible both for extremely high transcription in galactose medium as well as lowered levels in medium containing glucose and galactose (catabolite repression). Because the wild-type *his3* gene is always expressed but is not regulated as a function of growth on glucose or galactose, these *gal-his3* fusions are useful for distinguishing between positive and negative regulatory control mechanisms. The phenotypes of these fusions are summarized in Table 2.

Figure 2. The *his3-pet56* and *ded1* promoter regions. The nucleotide sequence of both strands of the divergent *his3-pet56* promoter region is shown at the top of the figure. *his3* deletion breakpoints are indicated as vertical lines above the top strand and *pet56* deletion breakpoints are indicated below the bottom strand. The positions of the poly(dA:dT) sequence for the constitutive expression of both genes (UP, black bar), the TGACTC regulatory sequences (R, gray bars), TATA elements (T_C, T_R, thin arrows), RNA initiation sites (vertical lines and thick arrows), and AUG initiation codons (underlined) are indicated. The structures of many of the mutant derivatives are shown below the DNA sequence. For each derivative, the horizontal line indicates the normal DNA sequences that are present, open boxes indicate internally deleted regions, striped boxes indicate the *ded1* promoter region (from −447 to −15), and the black boxes indicate the 365-bp *gal* regulatory region. The nucleotide sequence of the *ded1* promoter region is shown at the bottom of the figure. Symbols are as described above except that the striped bar indicates the *his3* promoter region (from −2500 to −35) and the checkered bar indicates *ura3* sequences.

Table 1. Requirements for Accurate Initiation

Alleles	End points		Spacing				Patterns of initiation			
			his3		*ded1*		*his3*		*ded1*	
			+1	+12	+1	+10	+1	+12	+1	+10
Wild type	–	–	45	56	65	74	++	++	++	++
Spacing mutants										
his3-Δ18	−35	−35	53	64	65	74	++	++	++	++
his3-Δ19	−35	−24	42	53	65	74	++	++	++	++
his3-Δ36	−35	−11	29	40	65	74	++	++	++	++
"Swap" mutants	*ded1*	*his3*								
his3-D1	−15	−11	69	80	65	74	++	++	++	++
his3-D3	−15	−24	82	93	65	74	++	+[a]	++	++
his3-D4	−15	−31	89	100	65	74	++	+[a]	++	++
his3-D2	−15	−35	93	104	65	74	++	+[a]	++	++
	his3	*ded1*								
ded1-H1	−35	−15	–	–	33	42	–	–	++	++
"Initiator" mutants										
his3-Δ70	−35	+2	18[b]	29	65	74	–	++[c]	++	++
	ded1	*his3*								
his3-D5	−15	+2	58[b]	69	65	74	–	++[c]	++	++

Spacing mutants represent deletions and insertions between the *his3* TATA element and the normal initiation sites; swap mutants represent hybrids between *his3* and *ded1* promoter and initiation regions; initiator mutants represent deletions that remove part of the initiation region (Chen and Struhl 1985; see Fig. 2). The end points are measured with respect to the normal *his3* or *ded1* +1 initiation site, and they refer to the last nucleotide that is present before the *Eco*RI linker joint. The spacing indicates the distance between the most upstream T of the relevant TATA element and the *his3* or *ded1* initiation sites. For each normal initiation site, transcriptional levels are indicated as follows: (++) the level that is expected from the upstream promoter element; (+) detectable, but lower than expected levels; and (−) very low or undetectable RNA levels.

[a]There are also aberrant initiations at −3 to −6.

[b]Since the +1 nucleotide is removed in these two derivatives, these numbers refer to the distances corresponding to the equivalent +1 position (i.e., to the C residue of the *Eco*RI site).

[c]The +22 site, normally a minor *his3* initiation site, is utilized at equal or nearly equal efficiency as the +12 site.

Positive Control is Mediated by an Enhancer-like Sequence

The *gal* regulatory region possesses many properties of enhancer elements (Struhl 1984). When appropriately fused to the *his3* mRNA coding region, it confers extremely high transcriptional activity, but only in galactose medium. In terms of mRNA molecules per cell, *his3* expression in such *gal-his3* fusion strains occurs at equivalent levels to those observed for the wild-type *gal1* and *gal10* genes. Although the *gal* region is clearly necessary for this activation, the region between −35 and −55, which coincides precisely with the *his3* TATA element, is also required (compare *his3*-G3 and *his3*-G4). The *gal* element activates transcription at various distances from the *his3* initiation site, even as far as 600 bp away (*his3*-G11). Finally, the *gal* element functions equally well in either orientation with respect to the *his3* coding sequences (compare *his3*-G4 and *his3*-G16), and it initiates transcription in an indistinguishable manner. Fine-scale deletion analysis and inversion of smaller fragments indicate that the *gal* segment does not contain two separate sites that independently regulate *gal1* and *gal10* expression (Johnston and Davis 1984; West et al. 1984; Giniger et al. 1985). Thus, a bidirectional element coordinately regulates the divergently transcribed *gal1* and *gal10* genes.

One way in which the *gal* regulatory fragment apparently differs from enhancer sequences is that it does not activate transcription even when placed only 82 bp downstream from the *his3* initiation region (Struhl 1984). Trivial explanations for this result, such as distance effects and unstable RNA species due to the insertion of the *gal* element into the *his3* structural gene, have been excluded. However, one possible explanation is that the region including the TATA and initiator elements may "block" galactose activation in a manner analogous to effects mediated by the bacterial *lexA* repressor protein binding to its cognate operator site (Brent and Ptashne 1984). Indeed, in both experiments that demonstrated the inability of upstream elements to function in a downstream location, this possibility was addressed and blocking effects were observed (Guarente and Hoar 1984; Struhl 1984). Thus, it may be that the properties of yeast upstream elements are not inherently different from enhancer sequences, but rather that these effects are due to apparent differences in the mechanism of transcriptional initiation, such as described in the previous section.

The *gal* element differs from the SV40 (Jongstra et al. 1984) and glucocorticoid receptor (Zaret and Yamamoto 1984) enhancers in terms of nuclease hypersensitivity of chromatin. Although the *gal* regulatory region is hypersensitive to DNase I digestion, this structural feature does not correlate with transcriptional activation (Struhl 1984). It is observed both in glucose and in galactose medium, and in the presence or absence of a TATA element. These results implicate

Table 2. Analysis of *gal-his3* Fusions

Allele	End point	Glucose	Galactose
his3-G1	−8	−	−
his3-G2	−23	−	−
his3-G3	−35	−	−
his3-G4	−55	−	+ + +
his3-G16	−55	−	+ + +
his3-G5	−80	−	+ + +
his3-G6	−136	±	+ + +
his3-G15	−157	±	+ + +
his3-G14	−173	±	+ + +
his3-G13	−192	±	+ + +
his3-G12	−204	±	+ + +
his3-G7	−253	+	+ +
his3-G8	−330	+	+ +
his3-G9	−357	+	+ +
his3-G11	−389	+	+ +
pet56-G1	−86	±	±
pet56-G2	−66	±	±
pet56-G3	−22	−	−

For each allele, the 365-bp DNA segment containing the *gal* upstream regulatory sequences (indicated as black boxes in Figs. 1 and 2) was fused to the *his3* or to the *pet56* promoter region at the positions listed (Struhl 1984; 1985c; see also Fig. 2). The difference between *his3*-G4 and *his3*-G16 is the orientation of the *gal* segment with respect to the *his3* transcription unit. For each allele, appropriate strains were grown in broth containing 2% glucose or galactose as the sole source of carbon. *his3* or *pet56* transcription levels were assayed by hybridizing to completion 50 μg of RNA with an excess of the relevant ^{32}P-end-labeled, single-stranded DNA probe, digesting the reaction mixture with S1 nuclease, and separating the denatured products by polyacrylamide gel electrophoresis (see Fig. 3). The transcription levels are indicated as follows: (+) normal constitutive levels of *his3* or *pet56* RNA; (−) no detectable RNA; (±) low RNA levels, roughtly 20% of normal; (+ + +) RNA levels that are equivalent to those of the normal *gal1-gal10* genes (i.e., complete induction); and (+ +) RNA levels that are approximately fivefold above the normal *his3* level but below the completely induced *gal* levels (Struhl 1984, 1985b).

a protein bound to this region which requires a conformational change or another protein for galactose induction. Presumably, this effect is not due to the *gal4* regulatory protein which interacts directly with four related 17-bp sequences located within the 365-bp DNA fragment (Bram and Kornberg 1985; Giniger et al. 1985).

The Catabolite Repression Regulatory Site Exerts Its Effects When Upstream of an Intact Promoter Region

Catabolite repression is a phenomenon whereby the expression of genes involved in carbon metabolism is reduced when glucose is present in the growth medium (Magasanik 1962). In principle, the molecular mechanism for catabolite repression could involve positive and/or negative control. In other words, the lowered levels in glucose medium could reflect either a failure to activate transcription (positive control) or actual repression of transcription that would otherwise occur (negative control). In *Escherichia coli*, catabolite repression, despite its name is a positive control mechanism because it depends on transcriptional activation by the cAMP:CAP protein complex (for review, see deCrombrugghe et al. 1984).

The properties of several *gal-his3* fusions strongly suggest that catabolite repression in yeast occurs by a true repression mechanism (Struhl 1985c; Table 2). When the 365-bp *gal* regulatory site is fused to *his3* derivatives that contain the entire promoter region (*his3*-G12 through G15), transcription in glucose medium is repressed below the wild-type basal level. This effect is clearly due to catabolite repression because it is abolished in raffinose medium, and the observed repression affects both the normal and the induced *his3* levels (Struhl 1985c). Therefore, the behavior of alleles *his3*-G12 through G15 in glucose medium cannot be explained by the lack of a functional activator protein, because if this were the case, wild-type *his3* expression levels should have been observed.

Two other points are worth noting. First, catabolite repression and galactose activation are separable properties of the *gal* regulatory region. In particular, catabolite repression of *his3* expression is observed in the absence of galactose. This strongly suggests that these two regulatory effects are mediated, at least in part, by different DNA sequences, a conclusion supported by deletion mutants of the *gal* segment described by West et al. (1984). Second, the ability of the *gal* region to exert its effects even when located at variable positions upstream from an intact promoter is markedly different from standard repression sites. In *E. coli*, the repression sites overlap or are between the required promoter elements (for review, see Reznikoff and McClure 1985), and in yeast, *lexA* repression of galactose induction occurs only if the operator site is between the *gal* upstream regulatory site and the TATA element (Brent and Ptashne 1984). Thus, it seems probable that the repression mechanism does not involve steric competition between the presumptive regulatory protein and the transcriptional apparatus. Instead, it seems possible that repression and enhancer-like activation represent opposite sides of the same basic mechanism.

POLY(dA:dT) SEQUENCES ACT AS UPSTREAM ELEMENTS FOR CONSTITUTIVE EXPRESSION

The *his3*, *pet56*, and *ded1* upstream promoter elements necessary for wild-type levels of transcription were determined by analyzing the phenotypes of deletion mutants that successively remove DNA sequences upstream from the respective structural genes (Struhl 1985d; Fig. 2; Table 3). Such sequential 5′ deletion analysis defines the minimum contiguous sequence necessary for wild-type transcription levels. For all three genes, the critical sequences are naturally occurring stretches of poly(dA:dT) located upstream of the TATA element. Deletion mutants that retain a particular stretch of dA:dT residues are transcribed equally efficiently as the wild-type gene, whereas related derivatives that lack this region are transcribed significantly below the normal level.

Table 3. Deletions of the *his3*, *pet56*, and *ded1* Promoter Regions

Allele	End point	Normal	Starvation
HIS3$^+$	–	+	+ + +
his3-Δ71	−204	+	+ + +
his3-Δ72	−192	+	+ + +
his3-Δ73	−173	+	+ + +
his3-Δ74	−157	+	+ + +
his3-Δ76	−146	+	+ + +
his3-Δ79	−142	+	+ + +
his3-Δ75	−136	+	+ +
his3-Δ80	−130	+	+ +
his3-Δ81	−115	±	+ +
his3-Δ82	−104	±	+ +
his3-Δ83	− 99	±	+ +
his3-Δ84	− 94	−	−
his3-Δ85	− 83	−	−
his3-Δ87	− 95/−80	+	+
his3-Δ88	−109/−83	+	+
his3-Δ24	− 44/−35	+	+
his3-Δ20	− 46/−34	+	+
his3-142	−100/−98	+	+
PET56$^+$	–	+	+
pet56-Δ4	−100	+	+
pet56-Δ3	− 86	+	+
pet56-Δ2	− 66	±	NT
pet56-Δ1	− 22	−	NT
DED1$^+$	–	+ + +	+ + +
ded1-U1	−123	+ + +	+ + +
ded1-Δ2	−123/−78	+	NT

Alleles with one listed end point represent deletions that remove all sequences upstream of that position; those with two end points indicate internal deletions that remove sequences between these positions (Struhl, 1982b, 1985c,d; D.E. Hill and K. Struhl, in prep.; Fig. 2). For each allele, wild-type and *gcd1*$^-$ strains were grown in broth containing 2% glucose as the sole source of carbon; *gcd1*$^-$ strains cause induced levels of amino acid biosynthetic genes under all conditions, and hence provide a convenient method to simulate amino acid starvation (Wolfner et al. 1975; Struhl 1982b; Donahue et al. 1983). *his3* transcription levels were assayed as described in Table 2 and Fig. 3; a *ded1* hybridization probe was included to provide an internal standard (Struhl 1985d; D.E. Hill and K. Struhl, in prep.). The transcription levels are indicated as follows: (+) normal constitutive levels of *his3* RNA; (−) no detectable RNA; (±) low RNA levels, roughly 20% of normal; (+ + +) levels equivalent to complete induction of the wild-type *his3* gene; and (+ +) partial induction (see text).

For the *his3* gene, deletions that retain as few as 130 bp upstream from the mRNA coding region behave indistinguishably from the wild-type gene, whereas related deletions that extend to −115 or further are transcribed at about 20% of the wild-type level. Earlier analyses of the *his3* promoter region indicate that besides the TATA element, sequences downstream from −115 are not important for normal expression levels (Struhl 1982a). Thus, the region between −115 and −130, which coincides with a 17-bp region containing 15 dT residues in the coding strand, is necessary and sufficient for constitutive expression.

Similar analysis of the *pet56* promoter indicates that this same 17-bp poly(dA:dT) region defines the upstream promoter element. In particular, sequences more than 86 bp upstream from the *pet56* initiation site are unimportant for full promoter function, whereas the region between −66 and −86 is necessary. In other words, although *his3* and *pet56* encode genes of unrelated function, the same poly(dA:dT) region located between these divergently transcribed genes is responsible for the constitutive transcription of both. This suggests that poly(dA:dT) sequences activate transcription in a bidirectional manner.

The *ded1* gene contains a typical TATA sequence located between −60 and −65 as well as a 34-bp region between −88 and −121 that contains 28 dT residues in the coding strand. Deletion of this region (*ded1*-Δ2) reduces *ded1* transcription to about 20% of the normal level. On the other hand, insertion of a *ura3* DNA fragment at position −121 (*ded1*-U1) does not affect *ded1* RNA levels, suggesting that sequences upstream of −125 are relatively unimportant. The fact that the *ded1* poly(dA:dT) region is longer than the *his3-pet56* region may explain why constitutive *ded1* RNA levels are roughly five times higher than *his3* or *pet56* levels.

The influence of long poly(dA:dT) sequences was initially observed in constitutive up-promoter mutants of the *adr2* gene (Russell et al. 1983). In two such mutants, the expansion of a normal 20-bp dA:dT sequence to a 54- or 55-bp homopolymer stretch presumably causes high levels of constitutive expression. Presumably, these rare mutants reflect functions of poly(dA:dT) tracts that are actually used in wild-type genes.

These are two mechanisms by which poly(dA:dT) sequences might confer their effects. One possibility is that a transcription factor binds to such sequences and activates transcription. A second and particularly attractive suggestion is that the RNA polymerase II transcription machinery recognizes the unusual structure of poly(dA:dT). Such sequences have a helix repeat of 10.0 bp instead of the normal 10.6 (Peck and Wang 1981; Rhodes and Klug 1981), and they are associated with kinks in DNA (Marini et al. 1983). Of particular interest is the observation that dA:dT regions prevent nucleosome formation in vitro (Kunkel and Martinson 1981; Prunell 1982). An attractive feature of this second suggestion is that because specific transcription factors are not invoked, it explains why poly(dA:dT) sequences behave as constitutive upstream elements.

his3 UPSTREAM REGULATORY ELEMENTS

When yeast cells are subjected to conditions of amino acid starvation, transcription of *his3* and other amino acid biosynthetic genes is induced over the basal level. Deletion mutations analyzed previously indicate that this regulation of *his3* expression depends on a sequence between −83 and −103 (Struhl 1982b). Deletion mutations that lack this region express *his3* at the normal basal level, but are unable to induce transcription in response to starvation conditions. Thus, the −83 to −103 region encodes a positive regulatory site necessary for transcriptional induction.

The TGACTC sequence (−99 to −94) within the *his3* regulatory region is of particular interest because it is repeated, with minor variations, several times in pro-

moters of coregulated genes (Struhl 1982b; Donahue et al. 1983; Hinnebusch and Fink 1983a). In the *his3* gene, perfect TGACTC sequences are found at −99 to −94 and −258 to −263, and imperfect variants containing five out of six matches are found at −142 to −137, −181 to −176, −216 to −221, and −225 to −230 (Struhl 1985b).

Two TGACTC Sequences Are Necessary for Full *his3* Induction

Sequential 5′-deletion mutants similar to those described in the section on constitutive expression were analyzed for their abilities to induce *his3* expression to the level observed for the wild-type gene (D.E. Hill and K. Struhl, in prep.). A fortuitous feature of *his3* transcription provides a very sensitive method for distinguishing between constitutive and regulated expression. As will be discussed in detail in a later section, when wild-type strains are subjected to conditions of amino acid starvation, transcription from the +12 site is induced fivefold whereas transcription from the +1 site is unaffected. Since the relative levels of the +1 and +12 transcripts are directly determined, even twofold changes in the ratio (which correspond to a total induction of only 50%) are easily observed.

Deletion mutants that encroach as far downstream as −142 behave indistinguishably from the wild-type gene (Table 3). Thus the four copies of the TGACTC repeat that lie upstream from this point are unimportant for regulation. However, deletions that extend to −136 or as far downstream as −99 fail to induce *his3* expression to the maximal level, although partial induction is observed. Because the only difference between the −142 and −136 derivatives is the imperfect TGCCTC sequence, this result indicates that two copies of the conserved sequence are necessary for full induction.

Several lines of evidence indicate that the proximal TGACTC sequence located between −99 and −94 is essential for induction (D.E. Hill and K. Struhl, in prep.). First, small deletions (*his3*-Δ87, *his3*-Δ88) that remove this region but retain all the other repeat sequences are uninducible (Table 3). Second, *his3*-Δ83, which removes all nucleotides upstream from the proximal TGACTC sequence shows the typical partial induction phenotype, whereas a derivative that extends 5 bp further and hence retains only the last C residue is transcriptionally defective under all conditions. This result also indicates that the proximal regulatory site is sufficient to confer partial levels of induction. Third and most compelling, a single base pair deletion of the dT residue at position −99 (*his3*-142) abolishes *his3* inducibility (Table 3).

Nucleotide Requirements of the TGACTC Regulatory Site

By using a set of synthetic oligonucleotides corresponding to the region between −102 and −91, mutants that differ by single base pair changes in the TGACTC regulatory site have been analyzed (D.E. Hill and K. Struhl, in prep.). Specifically, these oligonucleotides were joined via a *Sac*I linker to a *his3* test fragment that contains all sequences downstream of nucleotide −83. Yeast strains containing the test fragment are defective in *his3* transcription and consequently are unable to grow in the absence of histidine. However, when the wild-type oligonucleotide is fused to the test fragment, the resulting strains do grow. Thus, this growth phenotype provides a simple and qualitative assay for the effects of such point mutations with respect to *his3* transcription.

There are 18 possible base pair substitutions of the TGACTC regulatory sequences, of which 13 have been examined. As shown in Table 4, 12 out of these 13 do not permit growth in the absence of histidine, and hence are functionally defective. The sole exception is a C:T change resulting in the sequence TGACTT (*his3*-158). The fact that each nucleotide of the consensus can be mutated to produce a nonfunctional regulatory site indicates that each nucleotide is of some importance.

Although the TGACTC sequence is clearly essential, several lines of evidence suggest that the nine dT residues immediately downstream from the regulatory site are also important. First and most important, when the wild-type oligonucleotide is joined to the test fragment (*his3*-145), the resulting strains do not grow in the presence of aminotriazole, a competitive inhibitor of IGP dehydratase, the *his3* gene product. In contrast, a deletion mutant that simply removes all sequences upstream of −102 (listed as *HIS3*$^+$ in Table 4) permits growth under these circumstances. The only difference between these two derivatives is that the six downstream-most dT residues after the TGACTC sequence are replaced by a *Sac*I linker (GAGCTC). A derivative containing six dT residues (*his3*-143) is fully functional, whereas a derivative containing four dT residues (*his3*-144) is only partially functional. Second, the only functional point mutation (*his3*-158) produces a stretch of five dT residues; in comparison, the wild type and all the other derivatives contain only three dT residues. Third, other coregulated genes such as *his4*, *arg5*, and *trp5* all have dT tracts just downstream of the TGACTC sequence. Fourth, enhanced DNase I cleavage at the dT residues is observed when the GCN4 protein is bound to the regulatory sequences (Hope and Struhl 1985; see below).

gcn4 PROTEIN BINDS TO THE *his3* REGULATORY SEQUENCES

his3 transcription is regulated coordinately with many other genes involved in amino acid biosynthesis. *Trans*-acting factors involved in this general control phenomenon have been identified by mutations that fail to regulate properly the biosynthetic genes (Schurch et al. 1974; Wolfner et al. 1975; Penn et al. 1983; Hinnebusch and Fink 1983b). Epistatic relationships

Table 4. Point Mutations of the *his3* Regulatory Site

Allele	Sequence	Expression
HIS3$^+$	GGATGACTCTTTTTTTTC	+
his3-143	GGATGACTCTTTTTTGAGCTC	+
his3-144	GGATGACTCTTTTGAGCTC	±
his3-145	GGATGACTCTTTGAGCTC	±
his3-146	GGATGACACTTTGAGCTC	−
his3-147	GGATGACCCTTTGAGCTC	−
his3-148	GGATGACGGCTTTGAGCTC	−
his3-149	GGAAGACTCTTTGAGCTC	−
his3-150	GGACGACTCTTTGAGCTC	−
his3-151	GGATGAATCTTTGAGCTC	−
his3-152	GGATGATTCTTTGAGCTC	−
his3-153	GGATCACTCTTTGAGCTC	−
his3-154	GGATAAACTCTTTGAGCTC	−
his3-155	GGATTTACTCTTTGAGCTC	−
his3-156	GGATGTCTCTTTGAGCTC	−
his3-157	GGATGCCTCTTTGAGCTC	−
his3-158	GGATGACTTTTTTGAGCTC	+

For each allele, the DNA sequence between positions − 102 and − 83 is shown; in all cases, sequences between − 102 and − 447 are deleted (D.E. Hill and K. Struhl, in prep.; see text). For the wild-type derivative, the TGACTC sequence is underlined; in the point mutations, the altered base is underlined. *his3* expression was determined by the ability of strains to grow in the absence of histidine. (+) Growth equivalent to the wild-type derivative; (±) slower growth; and (−) no growth.

among these mutations suggest that the *gcn4* gene product has the most direct role in the transcriptional regulation process (Hinnebusch and Fink 1983b).

Recent experiments demonstrate that the *gcn4* gene encodes a specific DNA binding protein. This was accomplished by using a new and general method for analyzing protein–DNA interactions (Hope and Struhl 1985). Specifically, the *gcn4* protein-coding sequences were cloned into a vector containing a promoter for SP6 RNA polymerase, mRNA was synthesized by transcribing the template with this enzyme, and *gcn4* protein was synthesized as a pure ^{35}S-labeled species by in vitro translation of this mRNA. DNA binding activity was detected by incubating the labeled protein with specific DNA fragments, and separating protein–DNA complexes from free protein by electrophoresis in native acrylamide gels. By this assay, protein–DNA complexes are observed when *gcn4* protein is incubated with *Taq*I-cleaved pUC8-*his3* DNA (contains the 1.7-kb *Bam*HI fragment) (Hope and Struhl 1985). Such specific complexes are not observed with the vector DNA, although faint bands on the autoradiogram indicate that the protein does possess some nonspecific binding activity.

Four lines of evidence indicate that *gcn4* protein binds specifically to *his3* regulatory sequences and to promoter regions of other coregulated genes (Hope and Struhl 1985). First, analysis of deletion DNAs described previously indicates that the 20 bp region between − 85 and − 104 is necessary and sufficient for *gcn4* binding. This corresponds precisely with the proximal TGACTC regulatory sequences. Second, *gcn4* protein protects a 10-bp region including the TGACTC nucleotides from DNase I cleavage. This directly demonstrates an interaction between the critical regulatory sequences and the *gcn4* protein. Third, *gcn4* mutant proteins lacking various amounts of the carboxyl terminus, which were generated after restriction endonuclease cleavage of the original template, have no specific or nonspecific DNA binding activity. In particular, the mutant protein lacking the carboxyterminal 40 amino acids is almost identical to the predicted protein produced by strains carrying the *gcn4*-Δ1306 allele of Hinnebusch (1984). Thus, the fact that such strains are unable to induce the transcription of the coregulated genes is probably due to the inability of the mutant *gcn4* protein to bind DNA. Fourth, *gcn4* protein binds to the promoter regions of three other coregulated genes (*his4*, *trp5*, *arg4*), whereas it does not bind to analogous regions of four unregulated genes (*ded1*, *gal1*, *10*, *ura3*, *trp1*). All these results indicate that at the level of protein and DNA sequences, there is a direct correlation between DNA binding in vitro and transcriptional activation in vivo.

CONSTITUTIVE AND REGULATORY *his3* EXPRESSION

The results presented in previous sections indicate that the *his3* gene contains two distinct upstream promoter elements. Constitutive expression depends on the poly(dA : dT) sequence between − 130 and − 115, whereas induced expression in response to conditions of amino acid starvation depends on the TGACTC regulatory sequences. Thus, constitutive and regulated *his3* expression is most simply explained by a core promoter being activated by two independent upstream elements with different transcriptional specificity. However, the observations reported below strongly suggest that the

constitutive and regulatory promoters for *his3* expression are qualitatively different.

Regulated *his3* Transcription Is Initiated at the +12 Site

Under normal growth conditions, *his3* transcription is initiated at equal efficiency from positions +1 and +12. Surprisingly, under conditions of amino acid starvation, transcription from +1 remains at the normal basal level, whereas transcription from +12 is induced about fivefold (K. Struhl, in prep.; see Fig. 3). In addition, transcription from +22, normally a minor initiation site, is also induced. Thus, constitutive and regulated modes of *his3* expression are distinguished not only by their required upstream sequences, but also by their utilization of transcriptional initiation sites.

The same selectivity is observed in strains containing *gal-his3* fusion promoters in which the *his3* upstream region is replaced by the *gal1,10* enhancer-like sequence (Struhl 1984; Fig. 3). When such strains are grown in galactose medium, essentially all the transcripts are initiated at +12 and +22. This result does not depend on the location of the *gal* regulatory site because the same transcription pattern is observed in many different *gal-his3* fusions (Fig. 3).

A third example of this phenomenon is represented by revertants of *his3*-Δ13, a promoter mutation that contains the TATA element but lacks the entire upstream promoter region (Oettinger and Struhl 1985; Fig. 3). These revertants are all due to recessive suppressor mutations in three different genes, *ope1*, *ope2*, and *ope3*. In strains containing *his3*-Δ13 and any one of the *ope* suppressor mutations, *his3* transcripts initiated preferentially from the +12 site are observed in minimal medium, whereas no transcripts from +1 or +12 are detected in rich broth. The suppressor mutations presumably cause their transcriptional effects by activating cryptic upstream promoter elements.

Thus, in these three examples, the same initiation pattern is observed even though *his3* transcription depends on different upstream regulatory elements. This

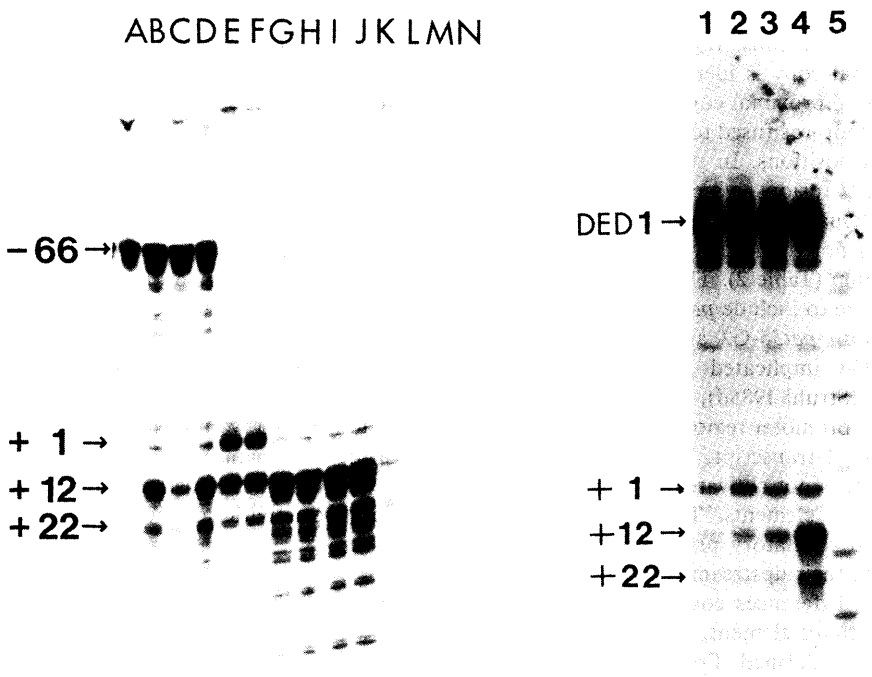

Figure 3. 5′ mapping of RNA transcripts. Total RNA (50 μg except for lanes G–J where 10 μg was tested) was hybridized to completion with an excess of a single-stranded *his3* probe (end-labeled with ^{32}P at position +174 and extending to −136), and in some cases with a *ded1* probe (end-labeled at position +262 and extending to −75). The reaction products were treated with S1 nuclease, denatured, and subjected to electrophoresis in a 6% polyacrylamide gel containing 7 M urea (Chen and Struhl 1985; Struhl 1985b). In lanes *A–D*, RNAs were prepared from cells grown in minimal medium from strains of the following genotypes: *his3*-Δ13 (lane *A*); *his3*-Δ13, *ope1* (lane *B*); *his3*-Δ13, *ope2* (lane *C*), and *his3*-Δ13, *ope3* (lane *D*). The *his3*-Δ13 allele, which deletes all *his3* sequences upstream of −66, does not express *his3*; the recessive *ope* suppressor mutations restore wild-type levels of expression (Oettinger and Struhl 1985). Lanes *E* and *F* contain RNA from a wild-type strain grown in broth containing 2% glucose (lane *E*) or 2% galactose (lane *F*) as the sole carbon source. RNAs from the following *gal-his3* fusions are analyzed in lanes *G–J* (galactose medium) and *K–N* (glucose medium): *his3*-G13 (lanes *G,K*); *his3*-G15 (lanes *H,L*); *his3*-G6 (lanes *I,M*); and *his3*-G4 (lanes *J,N*). In the autoradiogram on the right side of the figure, RNAs were examined from strains of the following genotypes: *gcn4*-1 (lane *1*); *gcn4*-2 (lane *2*); wild type (lane *3*); *gcd1*-1 (lane *4*); molecular weight standards (lane *5*). As described in the text, *gcn4* encodes the positive regulatory protein that binds the *his3* regulatory sequences; *gcd1*-1 is a mutation that causes induced *his3* transcription under all conditions, and hence is equivalent to amino acid starvation. The positions representing *ded1* transcripts and *his3* transcripts that initiate from nucleotides +1, +12, and +22 as well as read-through transcripts (RT) initiating upstream of nucleotide −66 (only relevant for strains containing *his3*-Δ13) are indicated.

suggests that although different proteins interact with these different regulatory sequences, the basic mechanism of transcriptional activation is similar. In contrast, the observed initiation pattern during constitutive *his3* expression is qualitatively different. This suggests that the poly(dA:dT) upstream element behaves in a functionally distinct manner from the regulatory elements.

pet56 Transcription Is Uninducible by the *his3* or *gal* Elements

The promoter region for the divergently transcribed *his3* and *pet56* genes appears symmetrical. Constitutive transcription of both genes is mediated by a centrally located dA:dT region that acts as a bidirectional upstream promoter element. Copies of the TGACTC regulatory elements are located between this common dA:dT region and the *his3* and *pet56* TATA elements. However, even though the TGACTC regulatory sequence functions in both orientations (Hinnebusch et al. 1985), *pet56* transcription is not increased when wild-type cells are subject to amino acid starvation (Struhl and Davis 1981).

To determine whether this lack of inducibility was due to properties of the *his3* regulatory sequences (e.g., the *his3* proximal copy is identical to the consensus, whereas the *pet56* proximal copy is imperfect), the *gal* regulatory element was fused to the *pet56* promoter region at several positions. In striking contrast to *gal-his3* and *gal-cyc1* fusions, which used the identical *gal* DNA segment (Guarente et al. 1982; Struhl 1984), *pet56* transcription is not induced when cells are grown in galactose medium (Table 2). This result cannot be explained by failure to include *pet56* TATA sequences in the fusions. Both *pet56*-G2 and G3 contain the sequence TATACA implicated as a functional *pet56* TATA element (Struhl 1985d), and *pet56*-G3 contains the entire *pet56* promoter region.

Thus, the *pet56* transcript, like the *his3* transcript initiating at +1, is not activated by two different upstream regulatory elements. This indicates that the constitutive and regulatory promoters must differ in ways other than their upstream promoter elements. In principle, such differences could occur at the TATA element, the initiator element, or other sequences that have not yet been defined. From the experiments reported below, I suggest that there are two distinct classes of TATA elements.

Evidence for Two Classes of TATA Elements

The basic model is that TATA elements (overall consensus sequence TATAAA) can be divided into two classes, constitutive and regulatory. The functional distinction between these hypothetical classes is that regulatory TATA elements (T_R) are active in the presence of any upstream regulatory site, whereas constitutive TATA elements (T_C) are not. By this proposal, the *his3* promoter contains both kinds of TATA elements, whereas the *pet56* promoter contains only the constitutive type (see Fig. 2). Two independent lines of genetic evidence support this model.

First, derivatives with small deletions in the TATA region (*his3*-Δ24 and *his3*-Δ20) do not affect the basal level of *his3* expression, but they do prevent induction (Struhl 1982b; Table 3). Moreover, in contrast to the wild-type gene, equal levels of both the +1 and +12 transcripts are observed even during starvation conditions (K. Struhl, in prep.). Formally, this result indicates that differences in the TATA region can account for differences between constitutive and regulatory expression. Presumably, these deletion mutants remove the T_R element without affecting the T_C element. Both of the mutations that delete the *his3* T_R element destroy the only perfect TATAAA sequence in the promoter region (nucleotides −45 to −40). However, they retain the sequence TATACA (nucleotides −54 to −49) that is also located in the functionally defined TATA region. The *pet56* gene does not contain any perfect TATAAA sequences, although both TATAGA (nucleotides −40 to −35) and CATAAA (nucleotides −50 to −45) are found in the region implicated as being functionally important. Thus the consensus sequence may act as a T_R element, while imperfect sequences may constitute T_C elements.

Second, when the spacing between the TATA and initiation region is increased by 8 bp (*his3*-Δ19), transcriptional induction is observed equally at the +1 and +12 sites (K. Struhl, in prep). This shows directly that transcription from the +1 site has the potential to be induced and thus suggests that the initiation region itself does not confer any specificity with regard to constitutive versus regulatory expression. In addition, this result suggests that the preferential utilization of the +12 initiation site occurs simply because T_R is too close to the +1 initiation site. The proposed T_C element is located about 10 bp further away from the initiation region, which would explain why it is capable of activating transcription from both sites with equal efficiency. Thus, the distinct initiation patterns may reflect the activities of the different TATA elements.

Chromatin Structural Changes in the TATA Region Associated with Constitutive and Regulatory Expression

In nuclear chromatin, the TATA region of the wild-type *his3* gene is preferentially cleaved by micrococcal nuclease (Struhl 1983a). This hypersensitivity is seen in *his3* derivatives lacking all sequences upstream of −155, whereas it is not observed in deletion mutants that lack the upstream promoter region (Struhl 1983a). For example, *his3*-Δ13 strains, which contain the TATA element but lack all sequences upstream of −66, do not show nuclease hypersensitivity and they are transcriptionally defective. However, in the presence of *ope* suppressor mutations, the *his3*-Δ13 allele is transcribed at wild-type levels but only under certain growth conditions (Oettinger and Struhl 1985). Nevertheless, these strains do not have a wild-type chromatin structure un-

der transcriptionally active or inactive conditions (Oettinger and Struhl 1985). Similarly, when the *gal* upstream regulatory site replaces the *his3* upstream promoter region (*his3*-G4), nuclease sensitivity is not observed in galactose medium, conditions causing extremely high levels of *his3* expression (K. Struhl, in prep.).

Thus, nuclease sensitivity at the TATA region is not correlated with transcription per se, but rather is associated with normal levels of transcription initiating at +1. In other words, these experiments provide direct evidence for a structural change at the TATA region that distinguishes constitutive expression from regulated expression; thus, they support the proposal that there are two different classes of TATA elements.

MOLECULAR MECHANISMS: INFERENCES AND SPECULATIONS

The properties of yeast promoter/regulatory elements are summarized schematically in Figure 4, and the conclusions derived from them are listed below.

1. Transcriptional initiation is not a simple enzyme–substrate interaction between RNA polymerase and DNA because there is no precise spacing arrangement of any of the promoter elements.
2. The facts that the distance between the TATA element and the initiation site is variable and that purified RNA polymerase II does not initiate at specific sequences in vitro suggest that a protein distinct from RNA polymerase II recognizes the TATA element. This is supported by TATA binding proteins isolated from higher eukaryotic organisms (Davison et al. 1984; Parker and Topol 1984).
3. Promoter specificity is determined primarily by the upstream element. The regulatory properties of the *his3* and *gal1,10* genes are determined by different DNA binding proteins that recognize different DNA sequences. The constitutive expression of the apparently unrelated *pet56*, *his3*, and *ded1* genes is mediated by poly(dA:dT) sequences, and the length of such homopolymer tracts influences the level of transcription.
4. It seems unlikely that transcriptional activation by upstream promoter elements is mediated by specific protein–protein interactions between the activator proteins and the transcriptional machinery. All these elements function bidirectionally, and in the case of the *gal* enhancer-like element, act at long and variable distances from the TATA element and initiation site. In addition, a given TATA element can function with different upstream elements, and presumably the cognate activator proteins.
5. Negative control of transcription cannot occur simply by steric competition between a repressor protein and the transcription apparatus. In contrast to prokaryotic repression models, the *gal* catabolite repression site exerts its effects when upstream from an intact promoter region.
6. The full induction of *his3* expression in response to conditions of amino acid starvation requires two copies of the TGACTC regulatory sequence, although the promoter-proximal copy is sufficient to confer partial induction. Each nucleotide of the TGACTC sequence is important for induction, as is the stretch of dT residues immediately downstream. *gcn4* protein, which is required for induction in vivo, binds specifically to the *his3* proximal TGACTC regulatory sequences and to promoter regions of other coregulated genes.
7. Two classes of TATA elements, constitutive (T_C) and regulatory (T_R), can be distinguished by their ability to respond to upstream regulatory elements and by their physical structure in nuclear chromatin. The arrangement of such elements in the divergent *his3-pet56* promoter region explains why constitutive expression of both genes depends on a common dA:dT region whereas only the *his3* gene is inducible.
8. Constitutive and regulated *his3* expression is mediated by interlaced promoters. Regulated transcription is initiated primarily at the +12 site, and it depends on the TGACTC upstream sequences and

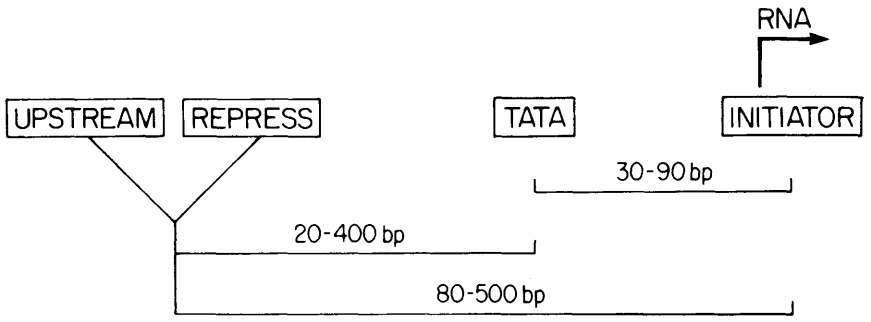

Figure 4. Promoter/regulatory elements in yeast. *cis*-acting elements of a hypothetical yeast gene are indicated as boxes. The initiator element, which is located near the RNA start (arrow) is important for determining where transcription begins. The TATA element, located 30–90 bp away from the RNA start, is required for transcription. The upstream promoter element, which can be located at variable distances away from the other elements, is important for transcription and also for regulation. Repressor sites, which are important for negative control, are also located at variable positions upstream of the TATA element. (See text for details.)

the T_R element. On the other hand, constitutive transcription is initiated equally at the +1 and +12 sites, and it depends on the poly (dA:dT) tract and the T_C element.

Transcriptional Initiation

A highly speculative model for constitutive and regulated transcription is shown in Figure 5. The basic assumption is that standard chromatin structure represents an inert form of DNA which is not recognized by RNA polymerase II. Thus, this structure must be disrupted for transcription to occur. Two mechanisms are proposed, both of which involve the upstream promoter element. Nucleosomes could be excluded from the promoter region either by the binding of specific activator proteins to their cognate regulatory sequences, or alternatively by the unusual structure of poly(dA:dT) sequences implicated as constitutive upstream elements. It seems less likely that the initial structural change would occur at the TATA element, because nuclease sensitivity of the *his3* region, which presumably measures a specific protein–DNA interaction, requires an upstream element.

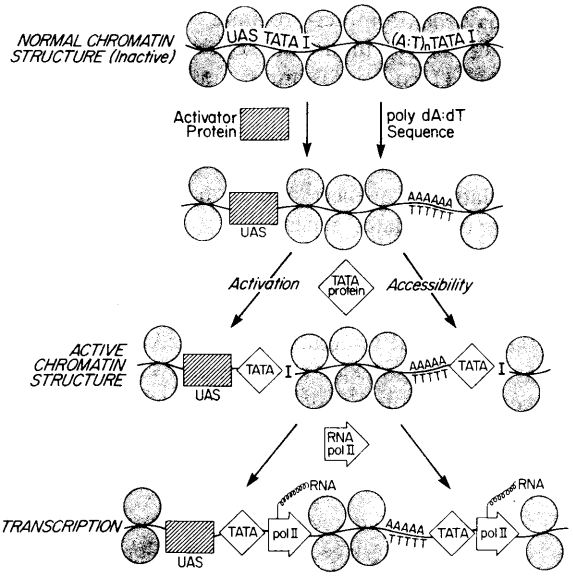

Figure 5. Molecular models for transcription. As described in the text, these models are highly speculative and are presented mainly to summarize the data. The top part of the figure shows a region of the yeast genome coated with nucleosomes (pairs of shaded circles). The promoter sequences of two genes are indicated. The gene on the left contains an upstream promoter element (UAS) typical for a regulated gene, whereas the gene on the right contains poly(dA:dT) tracts typical of a constitutively expressed gene. Both genes contain TATA and initiator (I) elements. The first step of transcriptional activation is diagramed as a disruption in chromatin structure mediated by an activator protein (striped box) or by the unusual properties of the poly(dA:dT) region. The second step involves interaction of the TATA protein (open diamond) with its cognate promoter element. This is pictured either as activation mediated by the particular protein that binds to the upstream element (for the regulated gene) or as accessibility due to nucleosome exclusion (for the constitutively expressed gene). By either of the proposed mechanisms, the result is an active chromatin structure. The final step is shown as the recognition of this active structure by RNA polymerase II followed by transcription initiation. The precise start point is mediated in some manner by the initiator element.

Such structural changes, although necessary, are insufficient for transcription because the TATA element is also required for transcription. In addition, R. Brent and M. Ptashne (pers. comm.) have provided direct evidence for an activation step that is distinct from DNA binding. Specifically, a *lexA-gal4* fusion protein, but not the *lexA* protein itself, activates transcription from a promoter containing the *lexA* operator as an upstream element. The fact that different regulatory elements can be functionally associated with a given TATA and initiation region suggests that this activation mechanism must be somewhat general.

The properties of the upstream elements described here and elsewhere suggest that a signal initiated at the upstream element must be transmitted downstream to the TATA element. This signal could be the movement of a protein, the obvious candidates being the activator protein, the TATA protein, or RNA polymerase II. Several considerations discussed in more detail elsewhere (Struhl 1985a) favor the TATA protein. Alternatively, the signal could represent a structural change induced at the upstream element and propagated in both directions. If activation represents nucleosome exclusion, the critical promoter sequences would be more accessible to the TATA protein and RNA polymerase II. Alternatively, if activation represents a change such as local supercoiling, the TATA protein could bind and/or be activated by recognizing such a change.

The basic proposal is that by any of these specific suggestions, an active chromatin structure is created which is recognized by RNA polymerase II such that the enzyme can bind nearby. The size of this binding region would correspond to the variability in spacing between the TATA element and initiation site, and the initiator element would correspond to the particular sequences within the binding region that are preferred by RNA polymerase II. This specificity could be due to intrinsic preferences of the polymerase or to an initiation factor that positions the enzyme.

This model provides a simple way to understand the basis of regulation. Regulated expression is mediated by specific binding proteins, whereas constitutive transcription depends on the unusual structural properties of poly(dA:dT) sequences. Positive control is achieved by transcription factors, which are functional only in association with cofactors that exist under specific environmental or developmental circumstances. In their active form, these proteins disrupt the normal chromatin structure, whereas in their inactive form, the chromatin is inert. Negative control is carried out by repressor proteins that also are affected by cofactors. Functional repressors could either alter the chromatin such that the transcription process does not begin, or

they could block the activation process that was begun by a positive factor. Thus, complex regulation can be viewed as a competition between activator and repressor proteins, each recognizing a specific DNA sequence and each subject to particular physiological controls, to determine the activity state of chromatin.

General Control of Amino Acid Biosynthesis

The *gcn4* gene, whose product is necessary for the coordinate induction of amino acid biosynthetic genes, encodes a protein that binds specifically to DNA sequences that are critical for this regulation (Hope and Struhl 1985). Thus, the simplest hypothesis is that *gcn4* protein is a specific transcription factor for genes that contain the TGACTC binding sites (Fig. 6). More complex hypotheses, such as *gcn4* protein interacting with a separate transcription factor, cannot be excluded.

This proposed mechanism, however, does not explain how yeast knows when to induce coordinately the transcription of amino acid biosynthetic genes. A cell must interpret the physiological state of starvation to produce a molecular signal, and it must transmit this signal to the effector molecule, *gcn4* protein, such that transcription of the coregulated genes is induced under appropriate conditions. The starvation signal is almost certainly produced during the process of translation because this is the only situation when all 20 amino acids are used for the same purpose, and because starvation for single amino acids results in the same transcriptional induction. Although the molecular nature of the starvation signal is unknown, the unusual structure and regulation of the *gcn4* gene (Hinnebusch 1984; Thireos et al. 1984) suggest how this signal is transmitted to the effector molecule, *gcn4* protein. The most revealing observation is that the level of a *gcn4-lacZ* fusion protein (and by inference the *gcn4* protein itself) is extremely low under normal growth conditions. Moreover, *gcn4* protein levels are increased dramatically under conditions of amino acid starvation, whereas the level of mRNA is minimally affected (Hinnebusch 1984; Thireos et al. 1984). Therefore, transcriptional induction of amino acid biosynthetic genes is mediated, at least in part, by the intracellular level of the specific DNA binding protein that recognizes the TGACTC regulatory sequences in the relevant promoter regions. Indeed, strains containing multiple copies of the *gcn4* gene cause high expression levels of amino acid biosynthetic genes even under normal growth conditions; presumably, this reflects an overproduction of the *gcn4* binding protein.

The structure of the *gcn4* gene strongly suggests a mechanism for its translational control (Hinnebusch 1984; Thireos et al. 1984). Unlike typical yeast mRNAs that contain short 5′ leaders before the AUG initiation codon, the leader for the *gcn4* mRNA is almost 600 bp in length and it includes four AUG codons that cannot initiate *gcn4* protein synthesis. Extensive studies of translation in yeast indicate that initiation begins at the 5′ proximal AUG codon and that it cannot be efficiently reinitiated at more downstream AUG codons (Sherman et al. 1980). Thus, the normal rules of translation preclude the synthesis of *gcn4* protein, thereby providing a simple explanation for why the coregulated genes are normally expressed at a basal level. However, under conditions of amino acid starvation, the basic translation rules must be altered in some manner such that *gcn4* protein is synthesized from its unusual mRNA.

An appealing feature of this model (Fig. 6) is that both the production of the molecular signal and its mode of transmission to the ultimate effector, *gcn4* binding protein, are associated with the basic translation process. It is tempting to speculate that these as-

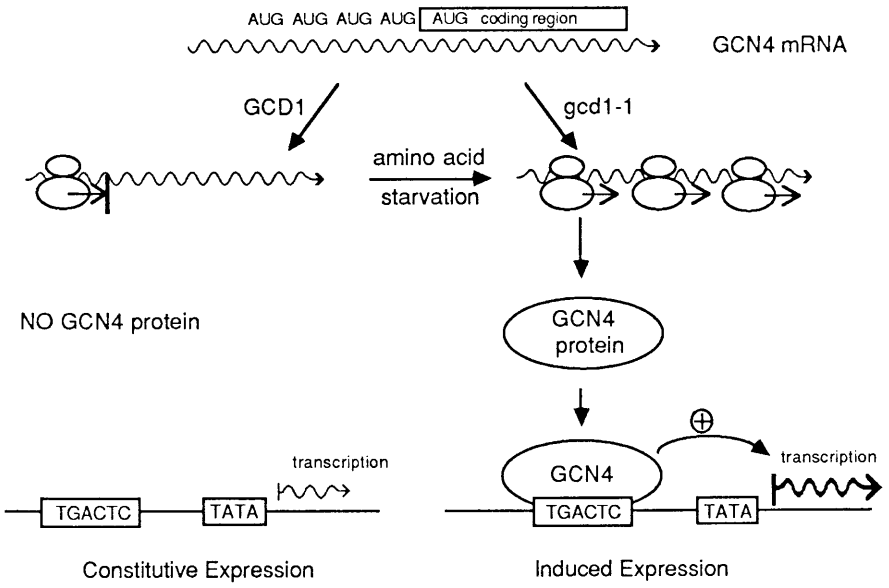

Figure 6. Molecular model for the general control of amino acid biosynthesis. (See text for details.)

pects of general control are related mechanistically and that the *gcd1* gene product is involved in both. *gcd1* mutant strains are defective in general control of amino acid biosynthetic genes (Wolfner et al. 1975), in protein synthesis (D.E. Hill and K. Struhl, in prep.), and in the translational control of *gcn4* protein levels (Hinnebusch 1984). For example, *gcd1* protein could sense the translationally stalled ribosomes that undoubtedly occur during starvation due to the lack of available charged tRNA species, and respond by releasing the block to translational reinitiation at the critical AUG codon of *gcn4* protein (Fig. 6). This provides a molecular mechanism for the translational specificity that ultimately determines the coordinate induction of amino acid biosynthetic genes in response to starvation conditions.

ACKNOWLEDGMENTS

This work was supported by grants from the Royal Society (I.A.H.), the Damon Runyon-Walter Winchell Cancer Fund (D.E.H.), the National Institute of Health (K.S.), and the Searle Scholars Program (K.S.).

REFERENCES

Bram, R. and R. Kornberg. 1985. Specific protein binding to far upstream activating sequences in polymerase II promoters. *Proc. Natl. Acad. Sci.* **82:** 43.

Brent, R. and M. Ptashne. 1984. A bacterial repressor protein or a yeast transcriptional terminator can block upstream activation of a yeast gene. *Nature* **312:** 612.

Chen, W. and K. Struhl. 1985. Yeast mRNA initiation sites are determined primarily by specific sequences, not by the distance from the TATA element. *EMBO J.* (in press).

Davison, B.L., J.M. Egly, E.R. Mulvihill, and P. Chambon. 1983. Formation of stable preinitiation complexes between eukaryotic class b transcription factors and promoter sequences. *Nature* **301:** 680.

deCrombrugghe, B., S. Busby, and H. Buc. 1984. Activation of transcription by the cyclic AMP receptor protein. In *Biological regulation and development* (ed. K. Yamamoto), vol. 3B, p. 129. Plenum Press, New York.

Donahue, T.F., R.S. Daves, G. Lucchini, and G.R. Fink. 1983. A short nucleotide sequence required for the regulation of *HIS4* by the general control system of yeast. *Cell* **32:** 89.

Giniger, E., S.M. Varnum, and M. Ptashne. 1985. Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell* **40:** 767.

Guarente, L. and E. Hoar. 1984. Upstream activation sites of the *cyc1* gene of *Saccharomyces cerevisiae* are active when inverted but not when placed downstream of the "TATA box." *Proc. Natl. Acad. Sci.* **81:** 7860.

Guarente, L. and T. Mason. 1983. Heme regulates transcription of the *cyc1* gene of *S. cerevisiae* via an upstream activation site. *Cell* **32:** 1279.

Guarente, L., R.R. Yocum, and P. Gifford. 1982. A *GAL10-CYC1* hybrid yeast promoter identifies the *GAL4* regulatory region as an upstream site. *Proc. Natl. Acad. Sci.* **79:** 7410.

Hinnebusch, A.G. 1984. Evidence for translational regulation of the activator of general amino acid control in yeast. *Proc. Natl. Acad. Sci.* **81:** 6442.

Hinnebusch, A.G. and G.R. Fink. 1983a. Repeated DNA sequences upstream from *HIS1* also occur at several other co-regulated genes in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **258:** 5238.

———. 1983b. Positive regulation in the general control of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* **80:** 5374.

Hinnebusch, A., G. Lucchini, and G.R. Fink. 1985. A synthetic *HIS4* regulatory element confers general amino acid control on the cytochrome C gene (*CYC1*) of yeast. *Proc. Natl. Acad. Sci.* **82:** 498.

Hope, I.A. and K. Struhl. 1985. GCN4 protein, synthesized in vitro, binds *HIS3* regulatory sequences: Implications for general control of amino acid biosynthetic genes in yeast. *Cell* (in press).

Johnston, M. and R.W. Davis. 1984. Sequences that regulate the divergent GAL1-GAL10 promoter in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **4:** 1440.

Jones, E.W. and G.R. Fink. 1982. Regulation of amino acid biosynthesis and nucleotide biosynthesis in yeast. In *The molecular biology of the yeast* Saccharomyces: *Metabolism and gene expression* (ed. J.N. Strathern et al.), p. 181. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Jongstra, J., T.L. Reudelhuber, P. Oudet, C. Benoist, C.B. Chae, J.M. Jeltsch, D.J. Mathis, and P. Chambon. 1984. Induction of altered chromatin structures by simian virus 40 enhancer and promoter elements. *Nature* **307:** 708.

Kunkel, G.R. and H.G. Martinson. 1981. Nucleosomes will not form on double-stranded RNA or over poly (dA):poly(dT) in recombinant DNA. *Nucleic Acids Res.* **9:** 6869.

Magasanik, B. 1962. Catabolite repression. *Cold Spring Harbor Symp. Quant. Biol.* **26:** 249.

Marini, J.C., S.D. Levene, D.M. Crothers, and P.T. Englund. 1983. A bent helix in kinetoplast DNA. *Cold Spring Harbor Symp. Quant. Biol.* **47:** 279.

Oettinger, M.A. and K. Struhl. 1985. Suppressors of *Saccharomyces cerevisiae* promoter mutations lacking the upstream element. *Mol. Cell. Biol.* **5:** 1901.

Parker, C.S. and J. Topol. 1984. A *Drosophila* RNA polymerase II transcription factor contains a promoter-region specific DNA binding activity. *Cell* **36:** 357.

Peck, L.J. and J.C. Wang. 1981. Sequence dependence of the helical repeat of DNA in solution. *Nature* **292:** 375.

Penn, M.D., B. Galgoci, and H. Greer. 1983. Identification of *AAS* genes and their regulatory role in general control of amino acid biosynthesis in yeast. *Proc. Natl. Acad. Sci.* **80:** 2704.

Prunell, A. 1982. Nucleosome reconstitution on plasmid inserted poly(dA):(dT). *EMBO J.* **1:** 173.

Reznikoff, W. and W. McClure. 1985. Prokaryotic promoters. In *From gene to protein: Steps dictating the maximal level of gene expression* (ed. W. Reznikoff and L. Gold). Benjamin/Cummings, California. (In press.)

Rhodes, D. and A. Klug. 1981. Sequence dependent helical periodicity of DNA. *Nature* **292:** 378.

Russell, D.W., M. Smith, D. Cox, V.M. Williamson, and E.T. Young. 1983. DNA sequences of two yeast promoter-up mutants. *Nature* **304:** 652.

Schurch, A., G. Miozzari, and R. Hutter. 1974. Regulation of tryptophan biosynthesis in *Saccharomyces cerevisiae*: Mode of action of 5-methyl-tryptophan and 5-methyltryptophan sensitive mutants. *J. Bacteriol.* **117:** 1131.

Sherman, F., J.W. Stewart, and A.M. Schweingruber. 1980. Mutants of yeast initiation translation of iso-1-cytochrome c within a region spanning 37 nucleotides. *Cell* **20:** 215.

St. John, T.P. and R.W. Davis. 1981. The organization and transcription of the galactose gene cluster of *Saccharomyces*. *J. Mol. Biol.* **152:** 285.

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

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

———. 1982b. Regulatory sites for *his3* expression in yeast. *Nature* **300:** 284.
———. 1983a. Promoter elements, regulatory elements, and chromatin structure of the yeast *his3* gene. *Cold Spring Harbor Symp. Quant. Biol.* **47:** 901.
———. 1983b. The new yeast genetics. *Nature* **305:** 391.
———. 1984. Genetic properties and chromatin structure of the yeast *gal* regulatory element: An enhancer-like sequence. *Proc. Natl. Acad. Sci.* **81:** 7865.
———. 1985a. Yeast promoters. In *From gene to protein: Steps dictating the maximal level of gene expression* (ed. W. Reznikoff and L. Gold). Benjamin/Cummings, California. (In press.)
———. 1985b. Nucleotide sequence and transcriptional mapping of the yeast *pet56-his3-ded 1* gene region. *Nucleic Acids Res.* (in press).
———. 1985c. The regulatory site that mediates catabolite repression in yeast exerts its effects when upstream of an intact promoter region. *Nature* (in press).
———. 1985d. Naturally occurring polydA:dT sequences are upstream promoter elements for constitutive transcription in yeast. *Proc. Natl. Acad. Sci.* (in press).
STRUHL, K. and R.W. DAVIS. 1981. Transcription of the *his3* gene region in *Saccharomyces cerevisiae. J. Mol. Biol.* **152:** 535.
THIREOS, G., M.D. PENN, and H. GREER. 1984. 5′ Untranslated sequences are required for the translational control of a yeast regulatory gene. *Proc. Natl. Acad. Sci.* **81:** 5096.
WEST, R.W., R.R. YOCUM, and M. PTASHNE. 1984. Yeast GAL1-GAL10 divergent promoter region: Location and function of the upstream activating sequence-UAS_G. *Mol. Cell. Biol.* **4:** 2467.
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.
ZARET, K.S. and K.R. YAMAMOTO. 1984. Reversible and persistent changes in chromatin structure accompany activation of a glucocorticoid-dependent enhancer element. *Cell* **38:** 29.