A COMPLEX ARRANGEMENT OF PROMOTER ELEMENTS MEDIATES INDEPENDENT REGULATION OF THE DIVERGENTLY TRANSCRIBED *HIS3* AND *PET56* GENES IN YEAST

KEVIN STRUHL
Department of Biological Chemistry, Harvard Medical School, Boston, Mass. 02115

ABSTRACT

His3 and *pet56* are adjacent yeast genes that are transcribed in opposite directions from initiation sites that are separated by 191 bp. Under normal growth conditions where *his3* and *pet56* are transcribed at similar basal levels, a poly (dA-dT) sequence located between the genes serves as the upstream promoter element for both. In contrast, *his3* but not *pet56* transcription is induced during conditions of amino acid starvation, even though the critical regulatory site is located upstream of both respective TATA regions. Similarly, when the upstream *his3-pet56* regulatory region is replaced by the analogous *gal* regulatory sequences, *his3* but not *pet56* transcription is induced when cells are grown in galactose medium. Although constitutive *his3* transcription is initiated equally from two initiation sites (+1 and +12), induction mediated by the *his3* or *gal* upstream regulatory elements or by *ope* suppressor mutations is associated with preferential utilization of the +12 site. This selectivity is determined by the distance of the mRNA initiation sites to a specific sequence in the *his3* TATA region. Analysis of chromatin indicates that micrococcal nuclease sensitivity of the *his3* TATA region is associated with the constitutive but not the inducible mode of transcription. From these observations, I suggest that the *his3-pet56* intergenic region contains constitutive and inducible promoters with different properties. In particular, two classes of TATA elements, constitutive (T_c) and regulatory (T_r), can be distinguished by their ability to respond to upstream regulatory elements, by their effects on the selection of initiation sites, and by their physical structure in nuclear chromatin. Molecular mechanisms for these different kinds of yeast promoters are proposed.

INTRODUCTION

Yeast promoters are composed of upstream, TATA, and initiator elements that are necessary for the accuracy, amount, and regulation of transcriptional initiation [reviewed in 11,44]. Upstream elements, which resemble mammalian enhancers [1,2], are required for transcription, and they usually determine a promoter's particular regulatory properties. Upstream elements for coregulated genes are similar in DNA sequence whereas those of unrelated genes are different. In several cases, these elements have been shown to be specific DNA binding sites for transcriptional regulatory proteins [3-5]. TATA elements (consensus sequence TATAAA) are present in essentially all yeast promoters although the distances from their respective mRNA initiation sites range between 40 and 120 bp. These elements are required for gene expression and have been presumed to have a general role in the process of transcription. The initiator element, located near the mRNA start site, has little effect on the overall RNA level but it determines where transcription begins.

In the natural yeast genome, *his3* and *pet56* are adjacent yeast genes that perform unrelated functions [6]. *His3* encodes imidazoleglycerolphosphate dehydratase, a histidine biosynthetic enzyme, and *pet56* is essential for mitochondrial function. These genes are expressed at similar basal levels under normal growth conditions [7], and they are transcribed divergently from initiation sites that are separated by 191 bp [6; Fig.1]. Although each gene has its own TATA element, a 17 bp region of poly (dA-dT) located between the genes serves as the upstream promoter element for both [8]. Thus, this constitutive element acts bidirectionally to activate transcription of two unrelated genes.

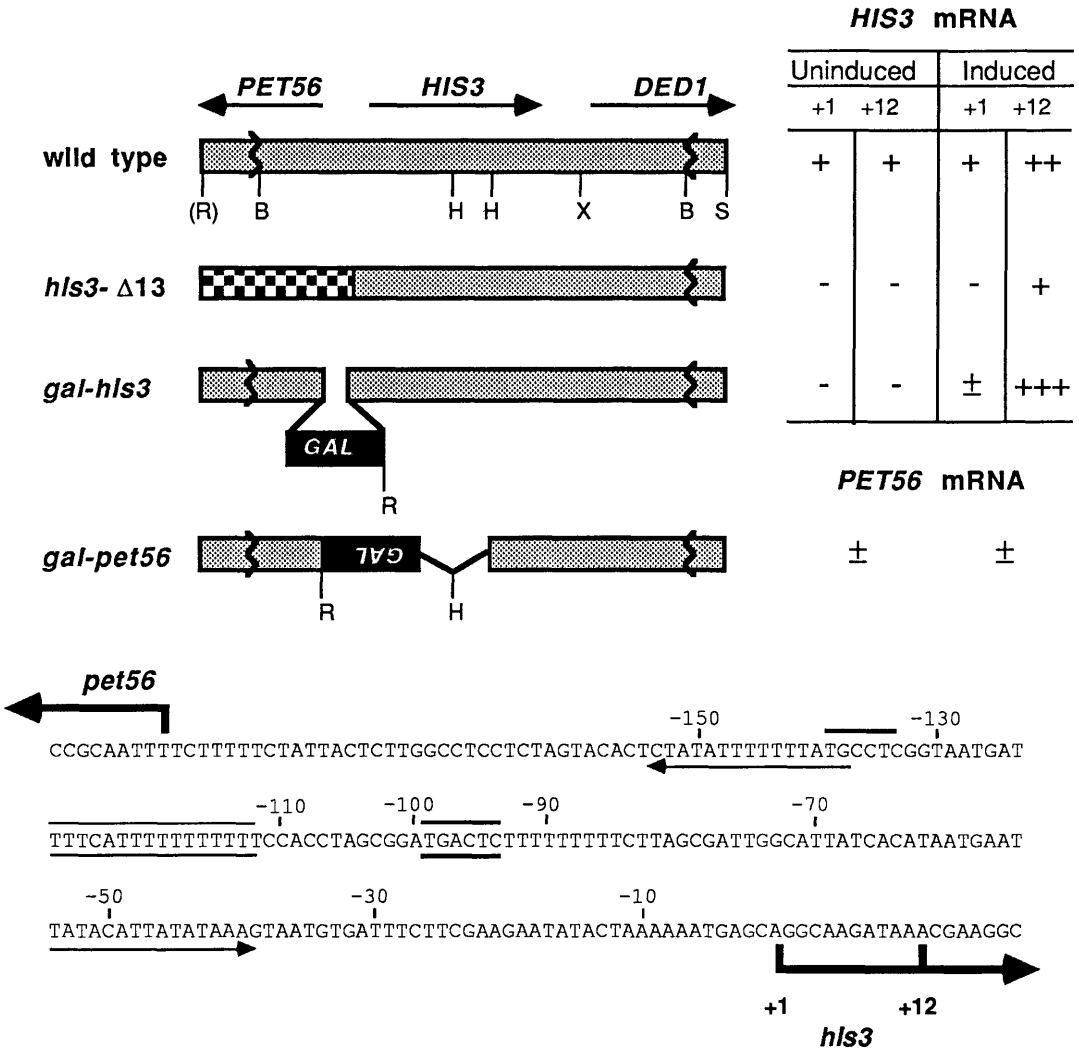

Figure 1: Structure of the *pet56-his3-ded1* region in wild-type and mutant strains and DNA sequences of the *his3-pet56* promoter region.

The shaded bar at the top represents a 6.1 kb *Eco*RI-*Sal*I fragment (with a mutated *Eco*RI site) that contains the intact *pet56*, *his3*, and *ded1* genes (6). The location and orientation of the transcripts are indicated by arrows above the shaded bar, and restriction endonuclease cleavage sites are indicated as vertical lines below the bar (R, *Eco*RI; B, *Bam*HI; H, *Hin*dIII; X, *Xho*I; S, *Sal*I). The drawing is to scale for the 1765 base pairs between the *Bam*HI sites. The structures of *his3*-Δ13 as well as *gal-his3* and *gal-pet56* fusions are shown with the solid bar indicating the 365 base pair *gal* segment (14), and with the checkered bar indicating the IS1 and bacteriophage λ sequences fused to position -66 of the *his3* gene (16). For each class of molecules, the levels of *his3* or *pet56* mRNA under uninducing or inducing conditions are indicated as follows: very high levels (+++); normal induced levels (++); normal basal levels (+); low, but detectable levels (±); and undetectable levels (-). Normal and inducing conditions for these derivatives were achieved as described in the legend to Fig.2. The bottom half of the figure shows the nucleotide sequence of the *his3* coding strand of the *his3-pet56* promoter region. The coordinates above the sequence are determined with respect to the upstream-most *his3* initiation site which is defined as +1. Features of the DNA sequence include the *his3* and *pet56* initiation sites (thick arrows)(8), the poly (dA-dT) upstream element for constitutive expression (thin lines above and below the sequence)(8), the TGACTC sequence that is critical for *his3* induction (thick lines above and below the sequence)(9), the related sequence that affects the maximal level of *his3* induction (thick line above the sequence)(10), and the *his3* and *pet56* TATA regions (thin, directional arrows)(8,20).

Under conditions of amino acid starvation, transcription of the *his3* gene is induced three-fold over the basal level, whereas *pet56* transcription remains at its basal level [7]. Maximal induction depends upon two copies of an upstream regulatory site that are located upstream of the *his3* and *pet56* TATA sequences and on either side of the poly (dA-dT) sequence necessary for constitutive transcription [9,10; see Fig.1]. The *his3* proximal copy is absolutely required for induction, and by itself is sufficient to confer partial induction. The distal copy is necessary only for full induction, and it is inactive in the absence of the proximal copy. This regulatory sequence (consensus TGACTC) is also present several times at similar locations in promoters of many amino acid biosynthetic genes that are subject to the same general control mechanism [9,11]. Analysis of the *his4* promoter indicates that a single copy of the regulatory site acts bidirectionally to induce transcription under starvation conditions [11,12]. Coordinate induction of amino acid biosynthetic genes is mediated by the *gcn4* protein, which binds specifically to the *his3* proximal TGACTC sequence and to promoter regions of other coregulated genes [5].

Constitutive and inducible *his3* expression is most simply explained by postulating that a core promoter is activated by independent upstream elements with different transcriptional specificity. However, this view does not explain why *pet56* transcription is not induced along with *his3* transcription during amino acid starvation [7] even though the TGACTC regulatory site, which is located upstream of both genes, functions bidirectionally [12] and at variable distances from mRNA start sites [9,12]. Here, I review genetic and biochemical observations that define the properties of the promoters that are responsible for transcription of the *his3* and *pet56* genes. I suggest that yeast cells contain two distinct classes of promoters, constitutive and inducible, that operate by different molecular mechanisms.

EXPERIMENTAL OBSERVATIONS

Induced *his3* transcription occurs at a subset of normal initiation sites

Under normal growth conditions, *his3* transcription is initiated at equal frequency from two major sites defined as +1 and +12 [6; Fig.2]. This pattern of initiation sites is observed even when the *his3* upstream and TATA elements are replaced by the analogous *ded1* promoter sequences [13]. The *his3* initiation pattern is similar in a *gcn4* mutant strain grown in normal or starvation conditions, although the level of the +12 transcript may be slightly reduced with respect to the +1 transcript (Fig.2). In this strain, *his3* transcription is uninducible because the *gcn4* mutation inactivates the positive regulatory protein that binds to the TGACTC sequence [5]. However, in a wild type strain when *his3* transcription is induced three-fold above the basal level, the normal initiation pattern is not observed. Instead, transcription from +1 remains at the normal basal level, whereas transcription from +12 is induced about 5 fold (Fig.2). In addition, transcription from +22, normally a minor initiation site, is also induced such that its level equals that of the +1 transcript. Thus, constitutive and induced modes of *his3* expression are distinguished not only by their required upstream sequences, but also by their utilization of initiation sites.

The same selectivity of *his3* initiation sites has been observed in two other situations. First, in strains where the *gal1,10* enhancer-like sequence is fused to the *his3* promoter region at various positions (Fig.1), *his3* transcription is subject to *gal* control; i.e. extremely high levels in galactose medium and catabolite repressed levels in glucose medium [14,15]. However, when any of these *gal-his3* fusion strains are grown in galactose medium, essentially all the transcripts are initiated at +12, +22 and further downstream [14; Fig.2]. This indicates that the selectivity of initiation sites is not determined by the distance between the upstream and the TATA elements. In glucose medium where transcription is repressed below the *his3* basal level even in the presence of the poly (dA-dT) upstream element [15], the level of the +1 and +12 transcripts are reduced equivalently (Fig.2). Second, similar selectivity of *his3* initiation sites is seen in revertants of *his3*-Δ13, a promoter mutation that normally is transcriptionally inactive because it lacks the entire upstream promoter region [16; Fig.2]. These revertants, which are due to recessive suppressor mutations in three different *ope* genes, confer wild type *his3* levels in minimal medium, but undetectable levels in rich broth. The suppression and the novel form of regulation have been explained by proposing that the *ope* mutations activate a cryptic upstream promoter element [16]. However, when *his3* transcription is observed in minimal medium, it is initiated preferentially from the +12 and +22 sites as compared to the +1 site.

Thus, the same initiation pattern is observed when *his3* transcription depends on the normal *his3* regulatory sequences, the *gal1,10* regulatory element, or the cryptic element associated with the *his3*-Δ13 mutation. In all these cases, *his3* transcription is subject to some form of regulation. In contrast, the observed initiation pattern during constitutive *his3* expression depending on the poly (dA-dT) sequence is qualitatively different.

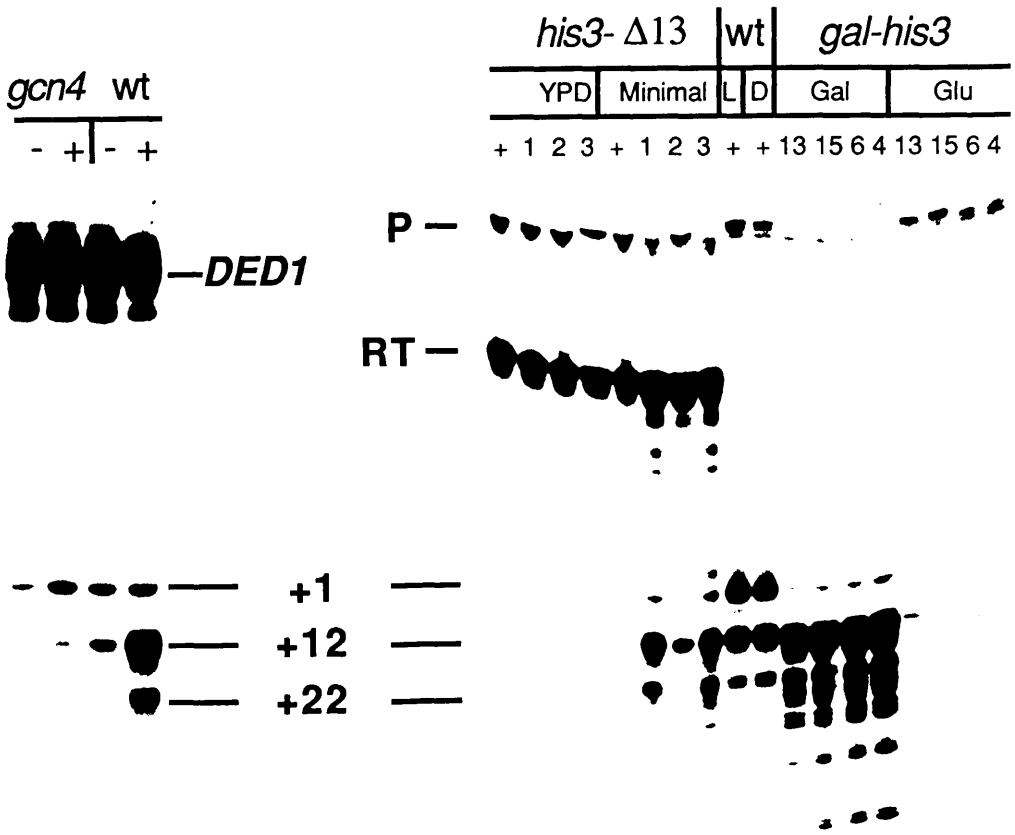

Figure 2: Selectivity of *his3* initiation sites.

Left: Transcription of the wild type *his3* gene in *gcn4*$^-$ and *GCN4*$^+$ strains in normal (-) or in inducing (+) conditions. The locations of the *ded1* transcripts and the *his3* +1, +12, and +22 mRNA species are indicated. Right: The left-most 8 lanes represent strains containing the *his3*-Δ13 allele with or without various *ope* suppressor mutations (1,2,3 represent *ope1,2,3* and + indicates no suppressor) grown in YPD (uninducing conditions) or minimal medium (inducing conditions). The right-most 8 lanes are strains containing *gal-his3* fusions (G4, G6, G13, G15)(14), grown in glucose (glu) or galactose (gal). The *his3* end points of the fusions are -192 (G13), -157 (G15), 136 (G6), and -55 (G4). The central 2 lanes represent the wild type strain in glucose (D) or galactose (L) medium. Bands for readthrough transcription of *his3*-Δ13 (RT)(16) and undigested probe (P) are indicated.

Pet56 transcription is uninducible in *gal-pet56* fusion promoters

Pet56 transcription is not induced during conditions of amino acid starvation [7], even though the critical regulatory sites are located upstream of the TATA element [8] and the TGACTC regulatory sequence functions in both orientations [12]. To investigate the issue of *pet56* inducibility, the *gal* regulatory element was fused to the *pet56* promoter region at several positions [17]. Strikingly, *pet56* transcription is not induced when cells are grown in galactose medium [17]. In other words, these derivatives behave unlike numerous fusions between the identical *gal* DNA segment and the *cyc1* [18] or *his3* [14] promoters. This result

can not be explained by failure to include *pet56* TATA sequences in the fusions. Deletion analysis indicates that *pet56*-G2, which includes the sequences TATAGA and CATAAA, contains a functional TATA element, and that *pet56*-G3 contains the entire *pet56* promoter region [8]. Thus, the *pet56* transcript, like the *his3* +1 transcript, is not activated by the *his3* or *gal* upstream regulatory elements. Therefore, differences between the constitutive and inducible promoters can not be explained solely by the upstream elements.

Requirements for inducibility and selectivity of *his3* mRNA initiation

In the original description of mutations that fail to induce *his3* enzyme activity in response to starvation, two classes were obtained [9]. One class removed the TGACTC regulatory element, that is now known to bind the *gcn4* positive regulatory protein, whereas the other class deleted sequences in the TATA region (Fig.3). Transcriptional analysis of two small deletion mutations in the TATA region (*his3*-Δ24 which removes sequences between -35 and -44 and *his3*-Δ20 which removes the region between -34 and -46) confirms that they abolish induction but do not affect the constitutive level [17]. Moreover, the levels of both the +1 and +12 transcripts are similar during normal and starvation conditions. This result is not be due to a spacing effect because small deletions between the TATA and initiation region do not influence the choice of initiation sites [13], and because the +22 transcript is not inducible even though it is more than 50 bp from the TATA region.

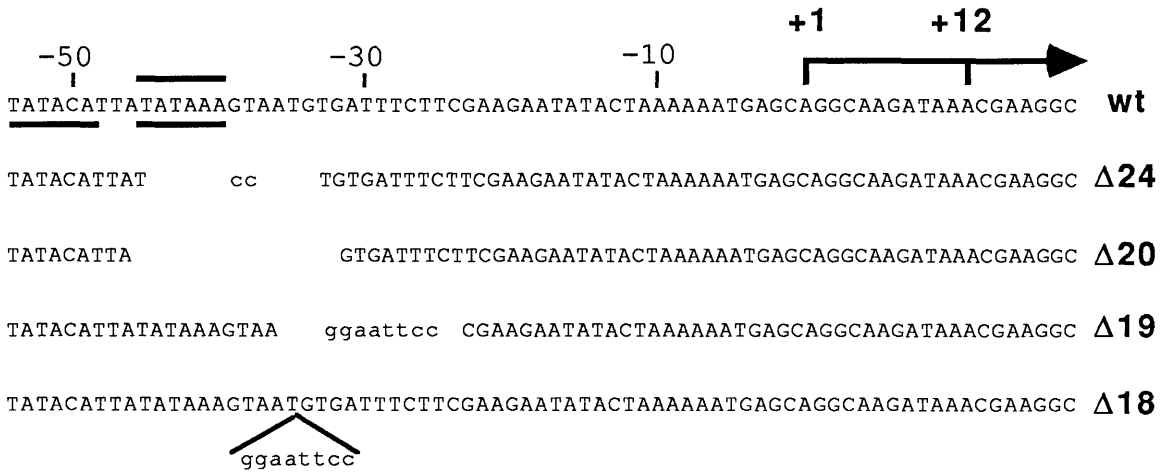

Figure 3: Sequences of DNAs with mutations in the *his3* TATA region.
The TATAAA sequence (-45 to -40) is over and underlined, and the TATACA sequence (-54 to -49) is underlined. Small letters indicate bases that derive from *Eco*RI linkers.

The results so far indicate that the pattern of *his3* transcriptional initiation does not depend on the sequences or the position of the upstream regulatory element, but can be affected by mutations in the TATA region. To exclude the possibility that selective utilization of the +1 and +12 sites depends on the properties of the initiation sites themselves, the transcriptional patterns of mutations that alter the spacing between the TATA and initiation region (Fig.3). In *his3*-Δ18, the distance is increased by 8 bp, and in *his3*-Δ19, the distance is decreased by 3 bp. When such strains are examined in normal growth conditions, the level of transcription and the initiation pattern are indistinguishable from the wild type strain [13]. In contrast, initiation during starvation conditions in *his3*-Δ18 (but not *his3*-19) is induced equally at the +1 and +12 sites although the overall level of induction is unchanged [17]. This indicates 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.

Two distinct chromatin structures in the *his3* TATA region

In nuclear chromatin, the TATA region of the wild type *his3* gene is preferentially cleaved by micrococcal nuclease [19]. Nuclease sensitivity is indistinguishable under normal or starvation conditions, and it is observed in a deletion mutant containing all the elements necessary for proper *his3* expression but lacking sequences upstream of -158. However, deletion mutants that retain the TATA region but lack the poly (dA-dT) upstream element do not show nuclease sensitivity. This observation reflects a feature of chromatin structure because control experiments indicate that the TATA region in purified *his3* mutant DNAs is equally sensitive to micrococcal nuclease [17]. Most importantly, *his3*-Δ13 strains never show TATA sensitivity even in the presence of *ope* suppressor mutations that confer wild type transcription levels [16], and nuclease sensitivity is not observed in several *gal-his3* fusions even when *his3* transcription is occurring at high levels [17].

These results indicate that nuclease sensitivity at the TATA region is correlated with the presence of the poly (dA-dT) sequence that serves as the upstream promoter element necessary for constitutive transcription. It is not correlated with transcription *per se*, and it is observed even when constitutive and inducible modes of *his3* expression are occuring simultaneously. Thus, these experiments provide evidence for a structural change at the TATA region that distinguishes constitutive expression from regulated expression.

CONCLUSIONS AND SPECULATIONS

Constitutive and inducible yeast promoters

This paper provides evidence for two classes of yeast promoters which are denoted as constitutive and inducible. In the *his3-pet56* intergenic region, the constitutive promoters are defined by the *his3* +1 and the *pet56* transcripts, and the inducible promoter(s) is defined by the +12 and +22 transcripts. The distinction between these classes is that the inducible promoters can be activated by different uptream elements representing unrelated forms of regulation, whereas the constitutive promoters are activated only by poly (dA-dT) sequences.

Induction of *his3* expression by the *his3* or *gal* upstream regulatory sequences or by *ope* mediated suppression clearly involves different upstream promoter sequences and presumably different proteins. Indeed, the *gal* regulatory site is bound by *gal4* regulatory protein [3,4], and the *his3* regulatory site interacts with *gcn4* protein [5]. Nevertheless, the fact that the same initiation pattern is observed suggests that the basic mechanism of transcriptional activation is similar. On the other hand, the clear difference in the initiation pattern observed during constitutive *his3* expression suggests that the poly (dA-dT) sequence behaves in a functionally distinct manner from the upstream regulatory elements.

Two classes of TATA elements

Although different upstream elements are associated with constitutive and inducible transcription in the *his3-pet56* region, their effects on the choice of initiation sites are unlikely to be direct. Three different upstream regulatory elements confer indistinguishable patterns of *his3* transcription. Moreover, systematic analysis of *gal-his3* fusions indicates that the initiation pattern does not depend on the orientation or the location of the *gal* element with respect to the rest of the *his3* promoter elements [14]. For the constitutive *his3* promoter, equal utilization of the +1 and +12 sites is observed with either the *his3* or the *ded1* [13] poly (dA-dT) upstream element, and it is independent of position. Thus, constitutive and inducible yeast promoters must differ in other ways besides the upstream elements.

Three separate lines of evidence strongly suggest that constitutive and inducible promoters contain different classes of TATA elements. First, small deletions in the *his3* TATA region abolish inducibility and selectivity of initiation without affecting constitutive transcription. This indicates that specific sequences in the TATA region are essential for the inducible but not the constitutive *his3* promoter. Second, the *his3* +1 transcript is inducible when the distance to the TATA element is increased by 8 bp. This indicates that differential utilization of the +1 and +12 transcripts does not depend upon specific sequences within the initation region, and it supports the view that the TATA region is important for the selective utilization of initiation sites. Third, the nuclease sensitivity experiments provides biochemical evidence that the constitutive and inducible *his3* promoters are associated with

different chromatin structures at the TATA region.

It should be noted that as the chromatin experiments measure the relative sensitivity of a particular sequence to micrococcal nuclease digestion, the information that is obtained applies only to that sequence. Moreover, the structural basis for micrococcal nuclease sensitivity of the *his3* TATA region is unknown. Some possibilities, which are not mutually exclusive, are accessibility to nuclear proteins, nucleosome phasing, tortional stress, or a specific protein DNA interaction. Nevertheless, although the molecular mechanism remains to be determined, the experiments indicate a clear difference between the two classes of promoters.

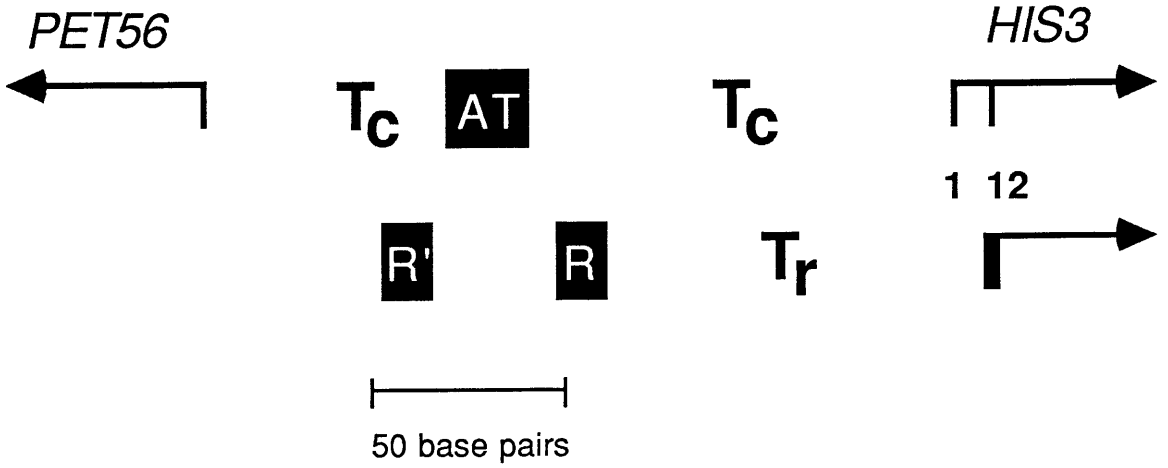

Figure 4: Elements for the constitutive and inducible expression of the *his3* and *pet56* genes.

The top line shows the poly (dA-dT) element for constitutive transcription of the *pet56* and *his3* gene, the T_c elements of the individual genes, and the *his3* +1 and +12 sites and the major *pet56* site. The bottom line indicates the *his3* upstream regulatory elements (R is essential for induction and R' is important to achieve the maximal level) and the T_r element necessary for transcriptional induction. Transcription from the inducible promoter occurs at a higher level (thick vertical line) and is initiated at the +12 site. The drawing is to scale.

The constitutive-inducible hypothesis

To account for all the experimental observations, I have defined two classes of TATA elements that correspond to the constitutive and inducible *his3* promoters [17]. The functional distinction is that regulatory TATA elements (T_R) are active in the presence of essentially any upstream regulatory sequence, whereas constitutive TATA elements (T_C) are not. With these definitions, the transcriptional initiation patterns in the *his3-pet56* region are explained as diagrammed in Fig.4. First, the *his3* promoter contains a T_R and a T_C element(s), whereas the *pet56* promoter contains only a T_C element. This explains why *his3* transcription is inducible by the *his3* and *gal* upstream regulatory elements, and why *pet56* transcription is uninducible by either element. Second, *his3*-Δ20 and *his3*-Δ24 delete the T_R element (thus preventing induction) but retain the T_C element(s) (thus permitting basal level expression). These mutations provide direct evidence for the existence and the location of the T_R element. Third, more extensive deletions of the TATA region (for example *his3*-Δ38 which removes *his3* sequences between -35 and -83) confer extremely low basal transcription levels [20], presumably because they lack all potential T_R and T_C elements. Fourth, preferential utilization of the +12 initiation site during induction occurs because T_R, which is required for induction, is too close to the +1 initiation site. This accounts for inducibility of the +1 transcript in *his3*-Δ18, the derivative where the distance between T_R and the +1 transcript, normally 45 bp, is increased to 53 bp. Although the minimal acceptable distance between TATA elements and initiation sites has not been determined precisely and may differ among promoters, the distances involved here are in excellent accord with previous experiments [13]. Fifth, the equal utilization of the +1 and +12

transcripts in the constitutive mode of expression is explained by the T_C element(s) being located upstream of T_R and hence far enough away from the +1 site to permit transcription. Thus, the distinct initiation patterns may reflect the activities of the different TATA elements.

From DNA sequence analysis of numerous eukaryotic promoters, the consensus sequence of the TATA element is TATAAA [21]. However, the functional distinction between T_R and T_C elements strongly suggests that they are defined by different DNA sequences. Although the precise nucleotide requirements remain to be determined, the results presented here provide some indications as to the sequences involved in these functions. 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). Moreover, additional deletion mutations indicate that T_R is entirely included between -35 and -45, a location that coincides with the TATAAA sequence (Chen and Struhl, unpublished results). However, at locations with appropriate spacings for initiation at +1, these T_R deletion mutants retain several sequences that resemble the canonical element, the best of which is TATACA (nucleotides -54 to -49). The *pet56* gene does not contain any perfect TATAAA sequences, although both TATAGA and CATAAA (nucleotides -40 to -35 and -50 to -45 with respect to the *pet56* initiation site) are found in the region implicated as being functionally important [8]. Thus the "consensus" sequence may act as a T_R element, while "imperfect" sequences may constitute T_C elements. The functional distinctions and the potential sequence differences between T_R and T_C suggests the possibility that these elements are targets for separate DNA binding proteins that are necessary for transcription.

Clarifications of the model

The basic proposal in this paper is that there are two different classes of yeast promoters that are exemplified in the *pet56-his3* region. Nevertheless, it is important to stress that while "constitutive" and "inducible" promoters and "T_R" and "T_C" elements are terms that usefully describe the *pet56-his3* promoter region, they are defined by specific functional criteria, not by the overall transcriptional properties of the genes. Although it appears that these terms may have more general application (see below), the knowledge that the expression of a particular gene is constitutive or regulated is insufficient to determine the kinds of promoters that are involved. Moreover, as the precise nucleotide requirements of T_R and T_C elements are unknown, simple inspection of the DNA sequence is insufficient to determine to which class a potential TATA element belongs. In fact, it is possible that a given sequence could function as both a T_R and a T_C element.

Some important clarifications to the T_R-T_C model should be mentioned. First, although T_C is defined to be unresponsive to upstream regulatory elements, there is no reason to exclude the possibility that T_R elements could interact with the poly (dA-dT) sequences. Thus, a constitutively transcribed gene could rely on either a T_R or a T_C element. Second, although poly (dA-dT) sequences can act as upstream elements for constitutive expression [8], it is certainly possible for constitutively transcribed genes to use other sequences as upstream elements. In this case, such constitutive upstream elements would presumably be recognized by specific DNA binding proteins. However, unlike regulatory proteins such as GCN4 or GAL4 whose level or activity varies with environmental conditions [4,22,23], these other proteins would presumably be equally active under all conditions. Third, the possibility of repression means that the presence of a poly (dA-dT) sequence upstream of a TATA element does not necessarily result in transcription. For example, in *gal-his3* fusions, the *gal* regulatory element represses both constitutive and inducible *his3* expression [15; Fig.2] even when it is located upstream of the entire *his3* promoter region. This also means that transcription of a negatively regulated gene could depend exclusively on a poly (dA-dT) upstream element.

Generality of the model

Selectivity of initiation sites has been observed previously during transcriptional induction of the yeast *suc2* [24] and *ura3* [25] genes and the *Neurospora crassa qa2* gene [26]. Although the promoter elements that are necessary for constitutive and inducible expression and for selectivity of initiation sites have not been determined in these cases, the transcriptional patterns of these other genes can be readily explained by the model proposed for *his3* and *pet56* expression. For example, at appropriate positions of both the *ura3* [27]

and *suc2* [28] promoters, there are poly (dA:dT) sequences of similar length and quality to the one between the *his3* and *pet56* genes, and there is a single presumptive T_R element as well as potential T_C elements.

Why does yeast have two classes of promoters? In the case of *suc2*, the selectivity has important biological consequences because the inducible transcript encodes the secreted form of invertase whereas the constitituve transcript encodes the intracellular form [24]. Although the *ura3* transcripts may encode separate proteins from different reading frames, the significance of the short non-*ura3* peptide is unknown [25]. For *qa2*, the significance of the selectivity is unclear because the different transcripts encode the identical protein [26]. However, the model provides a sensible rationale for the organization of the *his3-pet56* region as it permits closely packed and divergently transcribed genes to be regulated independently. This is particularly important for eukaryotic organisms because of their reliance on bidirectional upstream elements that can act at long and variable distances from the initiation site.

Possible molecular mechanisms

A common view of eukaryotic transcription is that RNA polymerase II, unlike *E.coli* RNA polymerase, does not bind to specific DNA sequences. Instead, it is believed that the enzyme recognizes an active chromatin structure that is created by the interaction of specific DNA binding proteins with their cognate promoter elements. In accord with this view, I propose the following molecular models to account for the differences between constitutive and inducible promoters.

For inducible promoters, the upstream element is recognized by transcriptional activator proteins. Specifically, the *his3* upstream regulatory element interacts with the GCN4 protein [5], whereas the *gal* element interacts with the GAL4 protein [3,4]. The TATA element is also presumed to interact with a specific protein. Although a yeast TATA protein has yet to be identified, analogous proteins have been characterized from higher eukaryotic cells [29,30].

Two lines of evidence indicate that the binding of upstream activator and TATA proteins is insufficient for transcription. First, there are *gal4* [31] and *gcn4* [32] mutations that prevent transcriptional induction without affecting the specific DNA binding activity of the protein. Second, hybrid proteins containing the DNA binding domain of the *E.coli* lexA repressor and the non-binding regions of GAL4 [33] or GCN4 [32] can function as transcriptional activators if an *E.coli lexA* operator serves as an upstream promoter element. Thus, it is likely that direct or indirect interactions of activator and TATA proteins are necessary for the generation of an active chromatin structure that can be recognized by RNA polymerase II. As *his3* induction mediated by the *gal* and *his3* upstream elements requires the same TATA sequences [14,20] and results in the same selectivity of initiation sites, it seems likely that the interactions between these different upstream activator and TATA proteins (and possibly RNA polymerase II) will share common features.

For constitutive promoters, I suggest that the poly (dA-dT) sequences are not recognized by specific proteins, but rather act as upstream elements by causing sequence-specific alterations in normal chromatin structure. In this regard, it is known that poly (dA-dT) tracts within natural DNA sequences inhibit nucleosome formation *in vitro* [34,35], possibly because of their unique helical repeat [36,37]. Thus, in contrast to the model for inducible promoters, transcriptional activation does not require specific interactions between upstream activator and TATA proteins. Instead, the poly (dA-dT) sequences disrupt the inactive chromatin structure such that the TATA element becomes accessible to the TATA protein. If, as discussed above, there are distinct TATA proteins, a T_R protein would contain a region that interacts with upstream activator proteins, whereas a T_C protein would lack such a region. In this way, these hypothetical TATA proteins would not only recognize different DNA sequences, but would respond differentially to upstream elements. As transcription can be induced from both the +1 and +12 initiation sites, it is presumed that both kinds of TATA proteins would interact directly or indirectly with RNA polymerase II in order to permit the initiation of transcription.

Though speculative, these models have several attractive features. First, transcriptional regulation can be achieved either by altering the synthesis or activity of a specific activator protein in response to environmental or developmental signals. A given activator protein will respond to a defined set of stimuli and co-regulatory molecules, and then influence the

transcription of all genes containing upstream elements to which it can bind. Second, although there are undoubtedly many activator proteins and intracellular signals, they appear to act through a common pathway involving T_R elements; this provides an evolutionary mechanism for altering gene regulation without affecting transcriptional competence. Third, constitutive transcription of a variety of yeast genes encoding unrelated functions can be achieved simply with the basic transcription machinery. Poly (dA-dT) sequences are observed upstream of many yeast transcription units [8], and their ability to inhibit nucleosome formation constitutes a plausible mechanism to activate transcription without gene-specific proteins. The level of constitutive transcription is influenced by the length and the quality of the poly (dA-dT) region [8]. Thus, the properties of constitutive and inducible promoters can account for the diverse transcriptional patterns that occur in yeast cells.

ACKNOWLEDGEMENTS

This work was supported by grants from the National Institutes of Health (GM30186) and from the Chicago Community Trust (Searle Scholars Program).

REFERENCES

1. L. Guarente, Cell 36, 799-800 (1984).
2. K. Struhl in: From gene to protein: Steps dictating the maximal level of gene expression, W. Reznikoff and L. Gold, eds.(Butterworths, Massachusetts, 1986) pp. 35-78
3. R. Bram and R.D. Kornberg, Proc. Natl. Acad. Sci. 82, 43-47 (1985)
4. E. Giniger, S.M. Varnum, and M. Ptashne, Cell 40, 767-774 (1985).
5. I.A. Hope and K. Struhl, Cell 43, 177-188 (1985).
6. K. Struhl, Nucl. Acids Res. 13, 8587-8601 (1985).
7. K. Struhl and R.W. Davis, J. Mol. Biol. 152, 535-552 (1981).
8. K. Struhl, Proc. Natl. Acad. Sci. 82, 8419-8423 (1985).
9. K. Struhl, Nature 300, 284-287 (1982).
10. K. Struhl and D.E. Hill, Mol. Cell. Biol., submitted.
11. T.F. Donahue, R.S. Daves, G. Lucchini, and G.R. Fink, Cell 32, 89-98 (1983).
12. A. Hinnebusch, G. Lucchini, and G.R. Fink, Proc. Natl. Acad. Sci. 82, 498-502 (1985).
13. W. Chen and K. Struhl, EMBO J. 4, 3273-3280 (1985)
14. K. Struhl, Proc. Natl. Acad. Sci. 81, 7865-7869 (1984).
15. K. Struhl, Nature 317, 822-824 (1985).
16. M.A. Oettinger, and K. Struhl, Mol. Cell. Biol. 5, 1901-1909 (1985).
17. K. Struhl, Mol. Cell. Biol., in press.
18. L. Guarente, R.R. Yocum, and P. Gifford, Proc. Natl. Acad. Sci. 79, 7410-7414 (1982).
19. K. Struhl, Cold Spring Harbor Symp. Quant. Biol. 47, 901-910 (1982).
20. K. Struhl, Proc. Natl. Acad. Sci. 79, 7385-7389 (1982).
21. R. Breathnach and P. Chambon, Ann. Rev. Biochem. 50, 349-383 (1981).
22. Hinnebusch, A.G., Proc. Natl. Acad. Sci. 81, 6442-6446, (1984).
23. G. Thireos, M.D. Penn, and H. Greer, Proc. Natl. Acad. Sci. 81, 5096-5100 (1984).
24. M. Carlson and D. Botstein, Cell 28, 145-154 (1982).
25. M. Rose and D. Botstein, J. Mol. Biol. 170, 883-904 (1983).
26. B.M. Tyler, R.F. Geever, M.E. Case, and N.H. Giles, Cell 36, 493-502 (1984).
27. M. Rose, P. Grisafi, and D. Botstein, Gene 29, 113-124 (1984).
28. L. Sarokin and M. Carlson, Mol. Cell. Biol. 4, 2750-2757 (1984).
29. B.L. Davison, J-M. Egly, E.R. Mulvihill, and P. Chambon, Nature 301, 680-686 (1983).
30. C.S. Parker, and J. Topol, Cell 36, 357-369 (1984).
31. L. Keegan, G. Gill, and M. Ptashne, Science 231, 699-704 (1986).
33. R. Brent and M. Ptashne, Cell 43, 729-736 (1985).
34. G.R. Kunkel and H.G. Martinson, Nucl. Acid Res. 9, 6869-6888 (1981).
35. A. Prunell, EMBO J. 1, 173-179 (1982).
36. L.J. Peck and J.C. Wang, Nature 292, 375-378 (1981).
37. D. Rhodes and A. Klug, Nature 292, 378-380 (1981).