The EMBO Journal vol.8 no.1 pp.261 – 268, 1989

Yeast upstream activator protein GCN4 can stimulate transcription when its binding site replaces the TATA element

Wei Chen and Kevin Struhl

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

Communicated by A.A.Travers

We replaced the required TATA element of a yeast *gal–his3* promoter by a binding site for GCN4, a protein that normally activates transcription when bound upstream of a TATA element. Surprisingly, GCN4 efficiently activates *his3* transcription from wild-type initiation sites, though in a pattern associated with constitutive *his3* transcription rather than GCN4 upstream activation through a TATA element. Transcriptional stimulation by GCN4 requires both the DNA-binding domain and the acidic activation function but is not affected by changing the spacing or helical relationship between the GCN4 binding site and the mRNA start sites. GCN4 is not sufficient for this TATA-independent activation; a sequence in the *gal* fragment distinct from the GAL4 binding sites is also required. Thus, GCN4 functions both when bound upstream of a TATA element and also when bound at the position of a TATA element. In the latter case, we suggest the possibility that GCN4 might be able to stimulate transcription by an alternate mechanism that does not involve a conventional TATA-binding transcription factor.

Key words: eukaryotic promoters/RNA polymerase II/transcriptional activation/TATA-binding protein/upstream activator protein

Introduction

A typical eukaryotic RNA polymerase II promoter consists of at least two kinds of *cis*-acting DNA sequences, upstream elements and TATA elements (reviewed by McKnight and Tjian, 1986; Struhl, 1987a; Guarente, 1988). TATA elements, located near mRNA initiation sites and resembling the sequence TATAAA, are important for transcription and are believed to interact with a component of the basic transcription machinery. Upstream elements (called enhancers or UASs) are usually located relatively far upstream of the transcription initiation sites, behave in an orientation and distance-independent manner, and are target sites for various activator proteins. Yeast activators stimulate transcription in mammalian cells (Kakidani and Ptashne, 1988; Webster *et al.*, 1988) and vertebrate activators stimulate transcription in yeast cells (Lech *et al.*, 1988; Struhl, 1988), thus suggesting a common molecular mechanism for transcriptional activation in eukaryotic organisms.

The GCN4 activator protein of the yeast *Saccharomyces cerevisiae* binds specifically to many genes involved in amino acid biosynthesis and coordinately stimulates their transcription (Hope and Struhl, 1985; Arndt and Fink, 1986). GCN4 binds as a dimer to the target site [consensus ATGA(C/G)TCAT], and both the dimerization and specific DNA-binding functions are localized within the 60 C-terminal amino acids (the full length protein is 281 amino acids) (Hill *et al.*, 1986; Hope and Struhl, 1986; 1987). In addition to the DNA-binding domain, transcriptional stimulation *in vivo* requires a separate activation function that is localized to a short region of 35–40 amino acids rich in acidic residues (Hope and Struhl 1986; Hope *et al.*, 1988). Moreover, different acidic portions of GCN4 are equally functional, suggesting that transcriptional activation functions consist of acidic regions with minimal primary requirements. Similar conclusions have been reached from mutational analyses of the GAL4 activator protein and from selection of functional activation regions from *Escherichia coli* DNA sequences (Ma and Ptashne, 1987a,b; Gill and Ptashne, 1987).

A crucial question is how short acidic regions activate transcription and how this mechanism is integrated into the overall pattern of gene regulation given that activator proteins such as GCN4 and GAL4 all influence the same RNA polymerase. From the short size and the non-stringent primary sequence requirements, it has been suggested that the activation region does not possess a catalytic function, but rather serves as a surface for interacting with a component(s) of the general transcription machinery (Hope and Struhl, 1986). Experimental support for a specific interaction mechanism comes from the observation that GAL4 is unable to activate transcription by T7 RNA polymerase in yeast cells; in contrast, a poly(dA–dT) element does stimulate transcription by T7 RNA polymerase (Chen *et al.*, 1987). From these observations, we argued against a model in which acidic activation regions interact with basic histones and stimulate transcription by affecting chromatin structure. Thus, TATA-binding proteins (Davison *et al.*, 1983; Parker and Topol, 1984; Sawadogo and Roeder, 1985) and RNA polymerase II are the two obvious candidates for targets of interaction by GCN4 and other upstream activator proteins.

In one popular model, the TATA-binding protein is viewed as a component of the basic RNA polymerase II machinery and upstream activator proteins stimulate this otherwise inefficient machinery by contacting the TATA-binding protein. Such an interaction is supported by the fact that a mammalian upstream activator and TATA-binding protein bind cooperatively to DNA (Sawadogo and Roeder, 1985). In addition, two functionally different TATA elements have been identified in the *his3* promoter region, T_R and T_C (Struhl, 1986), and saturation mutagenesis of the T_R promoter element has strongly suggested that T_R and T_C are binding sites for distinct proteins (Chen and Struhl, 1988). It has been suggested that GCN4 might function by inter-

acting with a T_R binding protein (Struhl, 1987a) since GCN4 activates transcription in combination with T_R but not with T_C.

Alternatively, GCN4 and other activator proteins might stimulate transcription by contacting RNA polymerase II directly. In this model, interactions between upstream activators and TATA-binding proteins are not crucial, although they may occur. A number of eukaryotic promoters do not appear to contain TATA sequences, thus suggesting that factors other than the conventional TATA-binding protein can interact with RNA polymerase II. For example, the yeast *TRP3* promoter apparently lacks a TATA sequence, but contains a GCN4 binding site at position −32 where a TATA element would normally be expected (Aebi *et al.*, 1984; Zalkin *et al.*, 1984). In addition, the largest subunit of RNA polymerase II from yeast to man contains a conserved seven amino acid sequence repeated many times at the C-terminus (Allison *et al.*, 1985; Corden *et al.*, 1985). This 'tail' is required for transcriptional activation *in vitro* (Dahmus and Kedinger, 1983) and *in vivo* (Nonet *et al.*, 1987; Allison *et al.*, 1988; Bartolomei *et al.*,1988), although it is not important for the basic RNA polymerizing activity. It has been proposed that the tail might interact with upstream activator proteins, an idea that could provide an additional explanation beyond cooperative DNA binding for the synergism often observed in activation.

The experiments in this paper were designed to address the issue of whether GCN4 activates transcription by contacting the TATA-binding protein or RNA polymerase. We reasoned that if GCN4 stimulated transcription by contacting RNA polymerase, then perhaps it would activate transcription in the absence of a TATA element when a GCN4 binding site was located at the position normally occupied by the TATA sequence. The T_R element (TATAAA) of a *gal*−*his3* promoter containing the UAS_G element for galactose induction was replaced by the GCN4 binding site to create a novel promoter with two conventional upstream elements and no TATA element. We find that the GCN4 binding site can substitute for the T_R element and then characterize the requirements for transcription from this novel promoter. The results are discussed with respect to molecular models of transcriptional activation and to strategies of promoter organization.

Results

A GCN4 binding site can functionally replace the T_R (TATAAA) element

We have previously described a *gal*−*his3* hybrid promoter, *his3*-G17, in which transcription is totally dependent on an oligonucleotide containing the *his3* T_R element (Chen and Struhl, 1988; see Figure 1). In the absence of the T_R oligonucleotide, the 365 bp *gal* segment, which contains four binding sites for the GAL4 activator protein (Bram and Kornberg, 1985; Giniger *et al.*, 1985), is unable to stimulate *his3* transcription. With the T_R oligonucleotide inserted between this *gal* upstream element and the *his3* mRNA initiation sites, *his3* transcription is strongly induced when cells are grown in medium containing galactose but not when grown in medium containing glucose. As expected from T_R-mediated transcription, initiation from the *his3* +12 site is strongly preferred over initiation from the +1 site. Almost all single base-pair substitutions of the TATAAA sequence in the T_R oligonucleotide abolish *his3* transcription.

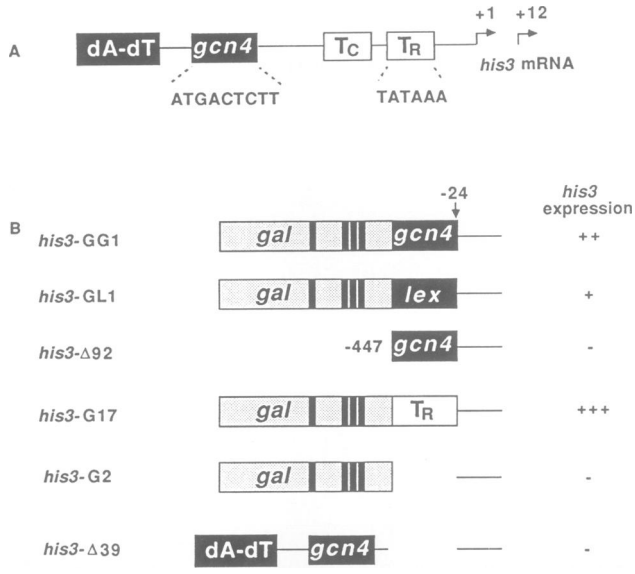

Fig. 1. DNA structures. (**A**) Structure of the wild-type *his3* promoter region as determined by extensive mutational analysis (Struhl, 1986; Struhl *et al.*, 1985). Constitutive *his3* transcription depends on a poly(dA−dT) upstream element and the T_C TATA element and is initiated with nearly equal frequency from the +1 and +12 sites. GCN4 activated transcription depends on the GCN4 binding site and the T_4 TATA element, and it is initiated with a strong preference from the +12 site. The important sequences for GCN4 and T_R function, determined by saturation mutageneses (Hill *et al.*, 1986; Chen and Struhl, 1988), are shown below their respective elements. (**B**) Some of the promoters used in this investigation; these are derived from previous *gal*−*his3* fusions or *his3* deletion mutants. The *gal* box represents the 365 bp *DdeI*−*Sau*3A DNA fragment from the *gal* 1−10 promoter region containing four GAL4 binding sites, and the *gcn*4, T_R and *lex* boxes are *Eco*RI or *Eco*RI−*Sac*I oligonucleotides (see Materials and methods). The levels of *his3* expression were determined by growth in the presence of aminotriazole (Hill *et al.*, 1986; Hope and Struhl, 1986; Hope *et al.*, 1988) and are indicated as follows: +++, wild-type growth in 20 mM aminotriazole (corresponds to maximal GCN4 induction of the wild-type *his3* gene); ++, slow growth in 20 mM but wild-type growth in 10 mM aminotriazole (corresponds to normal basal level of *his3* expression); +, slow growth in 10 mM aminotriazole (~20−50% of the normal basal level); −, no growth at 5 mM aminotriazole (<5−10% of the normal basal level).

An oligonucleotide containing the GCN4 binding site from the wild-type *his3* gene GGATGACTCTTTTTT was inserted in place of the T_R element in *his3*-G17; i.e. between the *gal* fragment and position −24 with respect to the *his3* RNA coding region (Figure 1). This novel promoter, designated *his3*-GG1, contains two conventional upstream elements (the GAL4 and GCN4 binding sites) and no TATA elements. DNA containing *his3*-GG1 was introduced into yeast strain KY320 in such a way that it replaced the original chromosomal *his3* gene in single copy at the wild-type locus.

Surprisingly, this novel promoter efficiently activates *his3* transcription. In the presence of 20 mM aminotriazole, conditions that require significant expression of the *his3* gene, the strain containing *his3*-GG1 grows almost as well as a strain containing the wild-type *his3* gene. This indicates that the level of transcription from *his3*-GG1 is nearly as high as that obtained when GCN4 activates transcription from its normal position upstream of the TATA element. Quantitative S1 nuclease analysis using *ded1* RNA as an internal control (Figure 2) indicates that the amount of *his3* RNA in the *his3*-GG1 strain (lanes 11 and 12) is similar (less than a factor of 2) to that obtained from the constitutive *his3* promoter

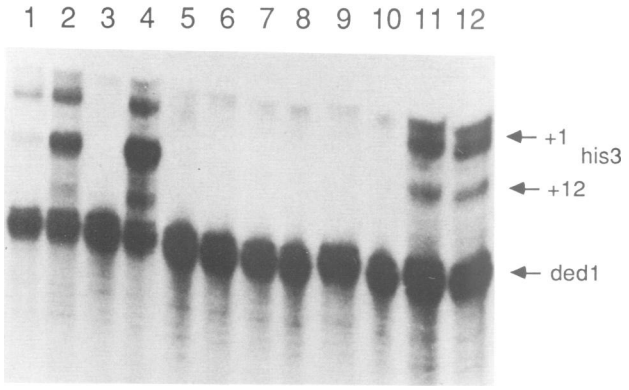

Fig. 2. Quantitation of *his3* RNA levels. Yeast strains containing the indicated *his3* alleles were grown in broth (non-inducing conditions) or in minimal medium containing 10 mM aminotriazole (inducing conditions) with glucose or galactose as carbon sources. RNAs from these strains were hybridized to completion with an excess of ^{32}P-labeled oligonucleotides for *his3* and the *ded1* internal control, and the products were treated with S1 nuclease: wild-type *his3* in glucose medium in non-inducing (**lane 1**) or inducing (**lane 2**) conditions; *his3*-G17 (contains T_R) in broth containing glucose (**lane 3**) or galactose (**lane 4**); *his3*-G2 (lacks any TATA element) in medium containing glucose (**lanes 5** and **7**) or galactose (**lanes 6** and **8**) in non-inducing (**lanes 5** and **6**) or inducing (**lanes 7** and **8**) conditions; *his3*-GG1 (contains the GCN4 binding site in place of the T_R element) in broth containing glucose (**lane 9**) or galactose (**lane 10**), or inducing conditions containing glucose (**lane 11**) or galactose (**lane 12**). The positions of *his3* RNAs initiated at positions +1 and +12 and the *ded1* control RNA are indicated.

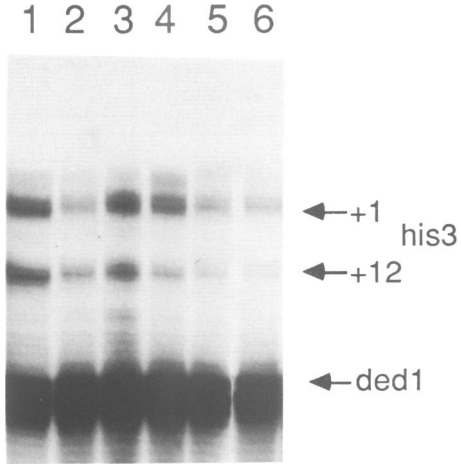

Fig. 3. *His3* RNA in strains containing varied helical turns between the *gcn4* and initiation sites. Yeast strains containing the following *his3* alleles were grown in glucose medium with 10 mM aminotriazole except for the wild-type strain (grown in the absence of aminotriazole) and RNAs were analysed as described in Figure 2: *his3* wild type (**lane 1**); *his3*-GG2 (**lane 2**); *his3*-GG3 (**lane 3**); *his3*-GG4 (**lane 4**); *his3*-GG5 (**lane 5**); *his3*-GG1 (**lane 6**). Densitometric scanning indicates that when normalized to the amount of *ded1* RNA in each lane, *his3* RNA levels and the ratio of the +1 to +12 transcripts in lanes 2–6 differ by less than a factor 50% (about twice as much RNA was loaded in lanes 3 and 4). The wild-type strain contains about twice as much *his3* RNA as the strains containing the *his3*-GG alleles.

that is GCN4-independent (lane 1); this corresponds to 20–40% of the maximal GCN4-induced level of the wild-type *his3* gene (lane 2).

The following lines of evidence indicate that *his3* transcription is due to activation by GCN4 protein. First, transcription from *his3*-GG1 depends on the GCN4 binding site, but not on hypothetical TATA-like elements within the *gal* fragment, because transcription is not detectable in the absence of the GCN4 binding site (lanes 5–8). Second, activation depends on GCN4 rather than some fortuitous function interacting with the GCN4 binding site because transcription from the *his3*-GG1 promoter is only observed in the presence of aminotriazole (lanes 11 and 12), not in its absence (lanes 9 and 10). Aminotriazole causes amino acid starvation, which results in the synthesis of GCN4; due to a translational control mechanism, GCN4 is not synthesized under normal growth conditions (Hinnebusch, 1984; Thireos *et al.*, 1984). Third, replacement of the GCN4 binding site with an oligonucleotide containing the *E.coli* LexA binding site, allele *his3*-GL1, eliminates *his3* expression except in the presence of a plasmid capable of expressing the LexA–GCN4 hybrid protein (Figure 1). LexA–GCN4 activates transcription from *his3*-GL1 somewhat less efficiently than it activates transcription from *his3*-GG1. This difference almost certainly reflects the higher affinity of GCN4 to its target site as compared to the LexA DNA-binding domain to its operator (Hope and Struhl, 1985; Hurstel *et al.*, 1988).

Interestingly, transcription from *his3*-GG1 is initiated from both the +1 and +12 sites with a slight preference for the +1 site (lanes 11 and 12). In contrast, when GCN4 induces transcription from the wild-type *his3* promoter in combination with the T_R element, initiation occurs primarily at the +12 (lane 2), a preference also observed in galactose induction through the T_R element (lane 4). The *his3*-GG1 initiation pattern, resembles that mediated by the constitutive *his3* promoter (see Discussion).

Varying the spacing and helical relationship between the GCN4 of LexA binding sites and the his3 initiation sites does not affect transcription

In *his3*-GG1, the center of the GCN4 binding site is close (41 bp) to the +1 initiation site, suggesting that GCN4 might interact with proteins around the initiation sites such as RNA polymerase II. Such an interaction might be influenced by the distance between the GCN4 binding site and mRNA initiation site and/or the helical relationship between these two sites. To test this possibility, the distance between the GCN4 binding site and the mRNA initiation site was varied by +11, +7, +5 and −13 with respect to the original *his3*-GG1 allele by altering the downstream insertion point of the oligonucleotide to −35, −31, −29 and −11. Strains containing the resulting derivatives (*his3*-GG2, GG3, GG4, GG5) grew at equivalent rates as the original *his3*-GG1 strain in medium containing 20 mM aminotriazole. Quantitative RNA analysis indicates that these strains have similar *his3* RNA levels and that the pattern of initiation site selection is not affected (Figure 3), Thus, *his3* expression is not significantly affected by variations in distance, even when the GCN4 binding site is moved to −11; i.e. only 28 bp from the +1 initiation site. We also changed the distances between the LexA binding site and the *his3* initiation site by inserting the *lexA* oligonucleotide at −11, −24, −31 and −35 (alleles *his3*-GL2, GL3, GL4, GL5). As assayed by growth in the presence of aminotriazole, *his3* expression is still dependent on the presence of the LexA–GCN4 hybrid protein, but is not influenced by the spacing or by the variation of the helical turns.

The acidic activation region of GCN4 is required

Does transcription from *his3*-GG1 require the normal transcriptional activation function that is necessary when

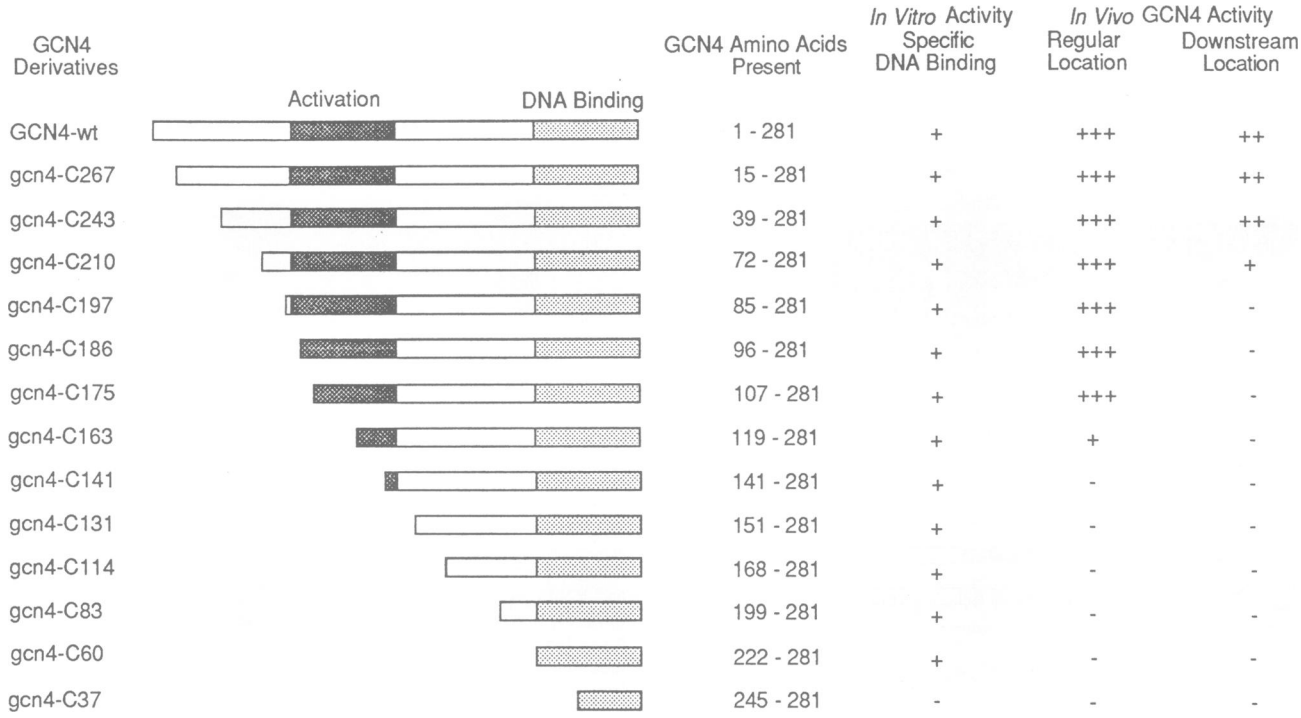

Fig. 4. Activation of N-terminally deleted GCN4 proteins. The structures of various N-terminally deleted GCN4 proteins are shown with the black bar representing the acidic transcriptional activation region and the shaded bar representing the C-terminal DNA-binding and dimerization domain. The abilities of these derivatives to activate *his3* transcription when the GCN4 binding site is at its regular location is the wild-type *his3* promoter (data of Hope and Struhl, 1986) or at the downstream location in place of the TATA element as in *his3*-GG1 are indicated as described in the legend to Figure 1.

GCN4 is bound upstream of a TATA element, or is a different part of the protein involved? To investigate whether similar regions of GCN4 are required for activating *his3*-GG1 and the wild-type promoter, we tested the properties of N-terminal deletion proteins that had been used previously to define the transcriptional activation region (Hope and Struhl, 1986). The GCN4 DNA binding function is located in the C-terminal 60 amino acids and the transcriptional activation function is fully retained in the 175 C-terminal amino acids (GCN4 contains a total of 281 amino acids). Deletions containing the DNA-binding domain but lacking the transcriptional activation region do not stimulate transcription and indeed can actually repress transcription in certain promoters.

A strain containing the *his3*-GG1 allele but lacking the entire *GCN4* structural gene was transformed with 13 plasmids carrying serial N-terminal deletions of GCN4, and the level of *his3* transcriptional activation was assessed by growth in the presence of aminotriazole (Figure 4). As expected, the full-length GCN4 protein activates *his3* transcription from *his3*-GG1, although to a slightly less extent compared to the wild-type *his3* promoter. This directly confirms transcription from *his3*-GG1 depends on GCN4. All derivatives lacking the GCN4 acidic transcriptional activation region fail to stimulate *his3* transcription. Three derivatives (gcn4-C197, C186, C175) do not stimulate transcription from *his3*-GG1 although they appear fully functional for activation of the wild-type *his3* promoter.

We performed a similar analysis using previously described C-terminal deletions of LexA−GCN4 (Hope and Struhl, 1986; see Figure 5). Although LexA−GCN4 activates transcription from *his3*-GL1 (Figure 1), the LexA binding site version of the novel promoter, almost all the deletion mutants appear to be inactive in this situation. Some of these deletions can activate transcription from a promoter containing the *lexA* operator upstream of a normal TATA element, but in that situation they are only 20% as efficient as LexA−GCN4 (Hope and Struhl, 1986; Figure 5). For this reason the apparent failure of these derivatives to activate *his3*-GL1 may simply reflect a problem of detection. In this regard, LexA-gcn4-N125, the only deleted derivative that activates transcription from *his3*-GL1, shows an anomalously high activity when the *lexA* operator is upstream of the TATA element (Hope and Struhl, 1986; Hope *et al.*, 1988). Thus, although LexA hybrid proteins activate *his3*-GL1 less efficiently than when the *lexA* site is upstream of a TATA element, the relative activities of these proteins follow the same pattern on the two promoters. These results corroborate those obtained with the N-terminally deleted GCN4 proteins and they suggest that similar, but possibly not identical, regions of GCN4 are required for activation in our construct and wild-type promoter.

Transcription depends on the gal fragment but not on GAL4 protein

Unlike all previously described *gal*−*his3* promoters (Struhl, 1984; Chen and Struhl, 1988), *his3* transcription from both *his3*-GG1 and *his3*-GL1 is efficient when cells are grown in glucose medium (Figures 1 and 2). One possibility of this observation is that GCN4 alone is sufficient to activate transcription when its binding site is at the position of the TATA element. However, when the *gal* fragment is removed as in *his3*-Δ92 (Figure 1), *his3* transcription is abolished. Previously described deletions of the native *his3* promoter (*his3*-Δ38, 39, 43) lacking the TATA region but containing

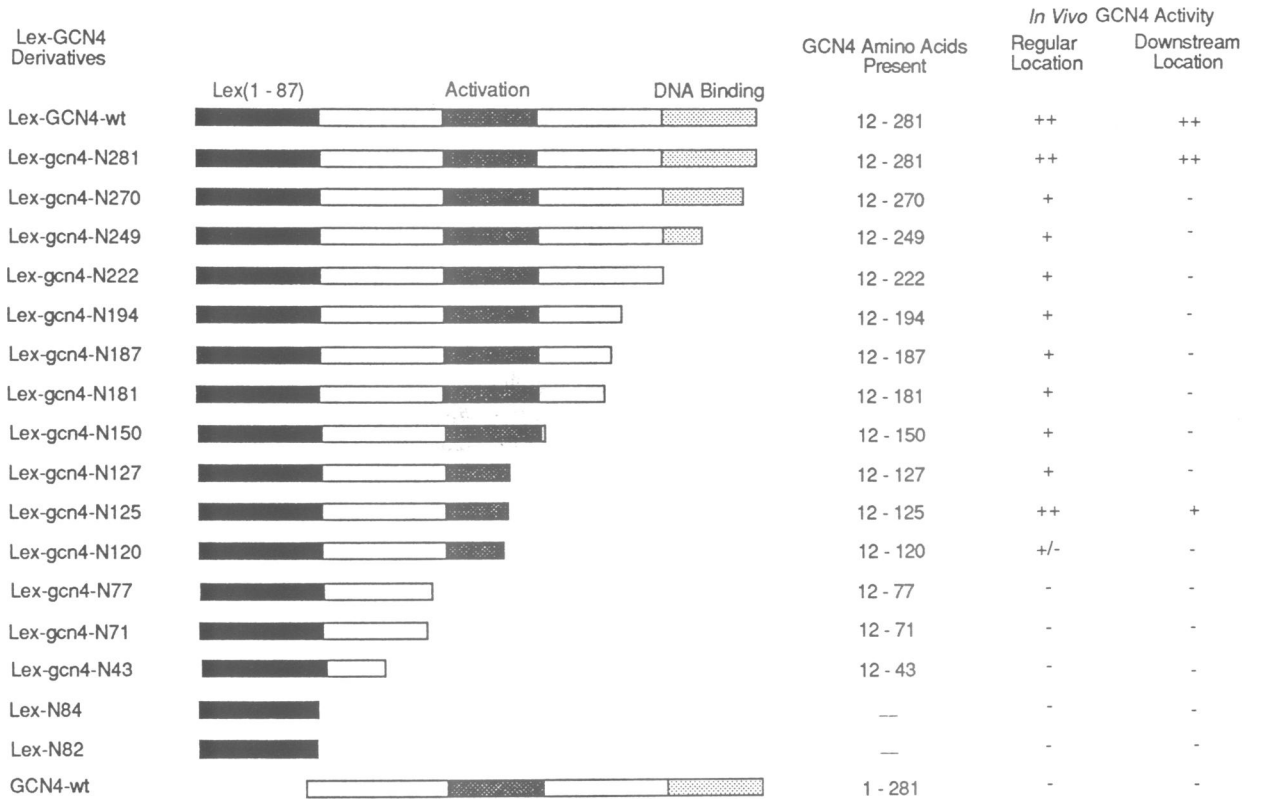

Fig. 5. Activation by C-terminally deleted Lex−GCN4 proteins. The structures of various C-terminal deletions of LexA−GCN4 are shown with the black bar representing the acidic transcriptional activation region and the shaded bar representing the C-terminal DNA-binding and dimerization domain. The abilities of these derivatives to activate *his3* transcription when the *lexA* binding site is upstream of a TATA element (listed as β-galactosidase activities as determined by Hope and Struhl, 1986) or at the downstream location in place of the TATA element as in *his3*-GL1 are indicated as described in the legend to Figure 1 for description of phenotypes (the designation ± indicates very poor growth in 10 mM aminotriazole).

the GCN4 binding site close to the initiation sites also do not activate transcription (Struhl, 1982). Thus, some genetic element within the 365 bp *gal* fragment is also required for GCN4 to activate transcription from *his3*-GG1.

The fact that GAL4 does not bind to UAS_G in glucose medium suggests that something other than GAL4 is required for transcription from *his3*-GG1. To prove this, we generated a deletion of the *GAL4* structural gene in the *his3*-GG1 strain by replacing the intact *GAL4* gene with a DNA fragment that has the LEU2 gene inserted in the middle of the *GAL4* coding sequence. The resulting *gal4* deletion strain grew well in the presence of 20 mM aminotriazole, thus indicating that GAL4 is not involved in transcription from the *his3*-GG1 promoter and that a previously unidentified functional element(s) in the *gal* fragment plays an important role.

Discussion

GCN4 can activate transcription when bound to upstream or downstream elements

Although GCN4 normally activates transcription when bound upstream of a TATA element, the major result of this paper is that a GCN4 binding site can functionally replace the T_R element to permit transcriptional activation of a *gal−his3* promoter. Several lines of evidence indicate that transcription from this novel promoter is not artifactual. First, transcription depends on GCN4 (or LexA−GCN4 when T_R is replaced by a LexA binding site), and it requires sequences in addition to the GCN4 DNA-binding domain. Second, it is initiated exclusively from the proper *his3* mRNA start sites (the significance of the initiation pattern will be discussed below). Third, transcription is efficient in that it occurs at the wild-type *his3* basal level, an amount typical for an average yeast promoter, and only ~2- to 3-fold below that observed for maximal GCN4 activation of the wild-type *his3* promoter. Fourth, transcription is not due to a fortuitous TATA element in the *gal* fragment or the GCN4 oligonucleotide because *his3* RNA was not observed in the absence of the oligonucleotide nor in galactose medium in the absence of GCN4. Finally, although *his3*-GG1 is an artificially constructed promoter, it appears to be organized like the native *TRP3* promoter which has a GCN4 binding site very close to the mRNA start sites and no conventional TATA sequences.

Although removal of the previously described acidic activation region prevents transcription from *his3*-GG1 or *his3*-GL1, certain derivatives that retain this region also fail to activate these promoters (Figures 4 and 5). While it is possible that different regions of GCN4 are required for activation from normal or novel promoters, we disfavour this explanation because two separate regions of GCN4 would have to be invoked to explain the results with the N- and C-terminal deletions. Alternatively, the apparent discrepancies might simply reflect a problem related to the fact that GCN4 activates *his3*-GG1 and *his3*-GL1 less efficiently than the wild-type *his3* promoter. In this view, the requirements for an acidic activation region would be qualitatively similar but quantitatively more stringent for

transcription from *his3*-GG1 or *his3*-GL1. This explanation completely accounts for the phenotypes of the C-terminal deletions of LexA−GCN4, including the high activity of LexA-gcn4-N125. To account for the phenotypes of N-terminal deletions such as gcn4-C175, C186, and C197, it is necessary to suggest that a fully functional GCN4 activation region requires additional residues beyond those defined from previous studies (Hope and Struhl, 1986; Hope *et al.*, 1988).

Multiple proteins can activate transcription from the downstream promoter element

Due to its presence as a downstream element in most eukaryotic promoters, the TATA element is often viewed as a target site for a common factor that is part of the basic transcription machinery. However, we have suggested that the *his3* T_R and T_C elements interact with distinct proteins with different sequence specificities for DNA binding that nevertheless perform a related 'downstream element' function (Struhl, 1986; Chen and Struhl, 1988). The work here supports this hypothesis because a different sequence can functionally replace the TATAAA oligonucleotide at the same location, and activation through this sequence requires a defined activator protein.

The properties of *his3*-GG1 indicate that the GCN4 binding site behaves more like the *his3* T_C element than the T_R element (Struhl, 1986). In *gal*−*his3* promoters, activation through T_R depends on GAL4 and occurs primarily at the +12 initiation site. In contrast, activation through the GCN4 binding site in *his3*-GG1 requires a different upstream sequence in the *gal* fragment and *his3* transcripts are initiated at both the +1 and +12 sites with a slight preference for the +1 site, an initiation pattern characteristic of constitutive *his3* transcription mediated through T_C. However, the GCN4 binding site is probably not equivalent to T_C because transcription is not observed in several derivatives such as *his3*-Δ39 (Struhl, 1982; Figure 1) in which the GCN4 binding site is downstream from the same sequences that stimulate constitutive *his3* transcription through T_C.

As yeast mRNA start sites are determined primarily by specific 'initiator' sequences near the mRNA start sites (Chen and Struhl, 1985; Hahn *et al.*, 1985; Nagawa and Fink, 1985; McNeil and Smith, 1986), it is not surprising that transcription from *his3*-GG1 is initiated from proper *his3* sites and that the initiation pattern is not very sensitive to the location of the GCN4 binding site. However, selectivity of initiation sites can be influenced by the distance from the TATA element; indeed, the strong preference for initiation at the +12 site when transcription depends on the *his3* T_R element is due to the closeness of T_R to the +1 site (Struhl, 1986). Nevertheless, activation by the putative T_R protein or by GCN4 yields a different initiation pattern even though both binding sites are at the same position. Most likely, each protein has its own requirements for the distance between its binding site and the mRNA initiation sites.

The idea that the downstream element function can be carried out by multiple proteins, including conventional upstream activator proteins, can easily account for transcriptional activity from promoters lacking the conserved TATAAA sequence. The yeast *TRP3* promoter lacks the conserved TATAAA sequence and instead contains a GCN4 binding site 28 bp upstream of the mRNA start site that presumably mediates induction in response to amino acid starvation (Aebi *et al.*, 1984; Zalkin *et al.*, 1984). The mammalian HMG-CoA reductase and HPRT promoters lack TATA-like sequences, but contain sequences resembling SP1 binding sites located close to initiation sites (Melton *et al.*, 1984; Reynolds *et al.*, 1984); perhaps SP1 a normal upstream activator (Gidoni *et al.*, 1985), is used as a downstream element in these promoters. Finally, *trans*-activation of certain viral promoters by ICP4 or E1A is mediated through a specific class of TATA elements that appears to contain sequences adjacent to the standard TATA homology (Homa *et al.*, 1988; Simon *et al.*, 1988).

At least two promoter elements are required, but not all combinations are functional

The GCN4 binding site alone does not constitute a functional promoter because *his3*-GG1 activates transcription only in the presence of the *gal* fragment. The 365-bp *gal* fragment contains several negative elements in addition to the four GAL4 binding sites (Struhl, 1985; West *et al.*, 1987). However, as activation does not depend on GAL4, there must be at least one other positive element in the *gal* fragment responsible for the transcription activity. From deletion analysis, this element maps ~30−50 bp downstream from the GAL4 binding sites (C.J.Brandl and K.Struhl, unpublished results), and for purpose of discussion will be termed Q. Thus, even when GCN4 functions from a downstream position, at least two promoter elements are necessary for transcription.

Although GCN4 and GAL4 can function with T_R, and Q can function with GCN4, other combinations of these proteins are nonfunctional (Figure 1). First, GAL4 and GCN4 cannot function together since *his3*-GG1 is not subject to galactose induction, and derivatives of *his3*-GG1 that lack the Q site but not the GAL4 sites are transcriptionally inactive (C.J.Brandl and K.Struhl, unpublished data). This is consistent with previous observations that multiple GCN4 or GAL4 binding sites do not activate transcription in the absence of a T_R element. Second, Q and T_R cannot function together because *his3*-G17 does not activate transcription in glucose medium. Third, GAL4 and Q cannot function together because *gal*−*his3* fusions lacking the T_R element are transcriptionally inactive. A comparison of *his3* promoters containing various combinations of elements does not reveal a simple rule for promoter element compatibility.

Implications for the molecular mechanism of GCN4 activation

We consider two models, both of which make the common and reasonable assumption that the conventional TATA-binding protein is part of the basic transcription machinery and presumably is associated with RNA polymerase II (Figure 6). In this discussion, RNA polymerase II includes auxillary initiation factors in addition to the core enzyme. In model 1, the interaction between GCN4 and the TATA-binding protein is the crucial step for activation, and GCN4 does not directly contact RNA polymerase II. In model 2, GCN4 and the TATA factor each interact with RNA polymerase; this model does not exclude potential interactions between GCN4 and the TATA-binding protein.

The experiments described here do not distinguish between these models for the common situation where GCN4 is bound upstream of a TATA element. However, the observation that

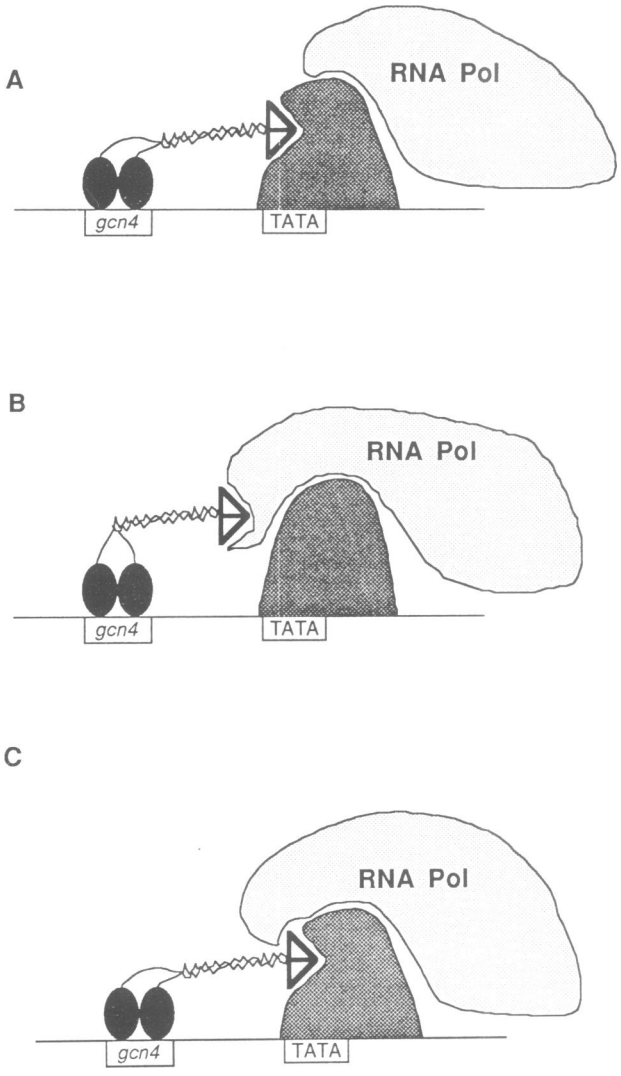

Fig. 6. Models for interactions among GCN4, the TATA-binding protein and RNA polymerase. In both models, GCN4 (two black ovals representing the dimeric DNA-binding domain) and the TATA-binding protein (dark shading) are shown binding to their target DNA sequences, and RNA polymerase II (light shading) is shown as interacting with the TATA-binding protein. In model 1, the crucial interaction (thick arrow) is between GCN4 and the TATA-binding protein. In model 2, the crucial interaction (thick arrow) is between GCN4 and RNA polymerase II; interactions between GCN4 and the TATA-binding protein (thin dashed arrow) are not essential for this model but may occur. See text for discussion.

GCN4 can activate transcription when its binding site replaces a conventional TATA element (T_R) suggests the possibility that a contact between GCN4 and the T_R factor might not be necessary in all circumstances. The possibility that GCN4 and the T_R protein could interact in the absence of the T_R binding site seems unlikely because it cannot account for the drastic effects of point mutations in the T_R element (Chen and Struhl, 1988) nor for the T_C-like initiation pattern observed from the *his3* GG1 promoter.

A modified version of model 1 is that the presumptive protein bound at the Q site carries out the TATA function, and the relative positions of the upstream and downstream promoter elements are reversed. To account for the data however, this interpretation requires at least two novel features. First, if Q were a TATA-like factor, it would have the property of being able to activate transcription in combination with GCN4 but not with GAL4. Moreover, GCN4 would be able to activate *his3*-GG1 transcription from the downstream position, whereas in the same promoter GAL4 would not activate from its normal upstream location. Second, there is no precedent for an upstream activator protein stimulating transcription when bound between the TATA element and the mRNA initiation sites. For these reasons, we disfavour but cannot exclude the modified version of model 1.

Instead, we suggest that the ability of GCN4 to activate transcription even when its binding site is extremely close to the mRNA initiation sites might reflect a direct contact between GCN4 and RNA polymerase (model 2). Our current model is that RNA polymerase II needs at least two qualitatively different contacts by promoter-binding proteins, a view that explains why at least two promoter elements are necessary and why only certain combinations of proteins function together. However, this simple picture is clearly insufficient to account for the transcriptional properties of all *his3* promoters that have been characterized here and elsewhere. These additional complexities might reflect interactions between the upstream and downstream activator proteins or they might be due to constraints about where particular elements can be placed in order to retain their functionality.

Materials and methods

The procedures for DNA manipulations, DNA sequencing, transformation of yeast cells, RNA preparation, and quantitative S1 nuclease analysis with oligonucleotide probes have been described previously (Chen *et al.*, 1987; Chen and Struhl, 1988). In all *gal*−*his3* promoter DNAs, the 365 bp *gal* fragment with four GAL4 binding sites is fused at various positions within the *his3* promoter region via an *Eco*RI linker (Struhl, 1984; Chen and Struhl, 1988). The DNA molecule containing the *his3*-GG1 promoter was derived from DNA containing *his3*-G17, the allele in which the *gal* segment is fused directly to an *Eco*RI−*Sac*I oligonucleotide containing the T_R element (Chen and Struhl, 1988). In *his3*-GG1, the T_R oligonucleotide was replaced by an *Eco*RI−*Sac*I oligonucleotide containing the wild-type GCN4 binding site from the *his3* promoter (GGATGACTCTTTTTT; Hill *et al.*, 1986). To generate the *his3*-GG2, GG3, GG4 and GG5 alleles, an oligonucleotide containing the same GCN4-binding site but flanked by *Eco*RI sites on both ends was inserted into *Eco* RI sites of appropriate *gal*−*his3* fusions (Struhl, 1984) such that the insertion points were at −35, −31, −29 and −35. The five *his3*-GL1 alleles were generated in a similar manner except that a self-complementary *Eco*RI oligonucleotide (AATTCGTACTGTATGT-ACATACAGTAC) containing the *lexA* operator was used. In *his3*-Δ92, the GCN4 binding site was fused directly to −447, a position within the *pet56* structural gene that does not contain any promoter elements (Struhl *et al.*, 1985). In most experiments, DNA molecules were introduced into yeast strain KY320 (Chen and Struhl, 1988) such that they replaced the normal *his3* locus. For testing the N-terminal GCN4 deletions, the *his3* locus of KY803 (Hope and Struhl, 1986) was replaced by the *his3* GG1 allele by successive gene replacement events (Struhl, 1987b); plasmids capable of expressing the GCN4 derivatives (Hope and Struhl, 1986) were then introduced. The level of *his3* expression in the resulting cells was determined by growth in the presence of aminotriazole as described previously (Hill *et al.*, 1986; Hope and Struhl, 1986; Hope *et al.*, 1988). The levels of *his3*, and *ded1* RNAs were quantitated by scanning the autoradiograms with a Beckman DU-6 spectrophotometer. For each determination, the intensity of the *his3* band(s) was normalized to that of the internal *ded1* control. In addition, the relative specific activities of the *his3* and *ded1* probes were determined using control RNA from wild-type cells (with probes of equal specific activity, the *ded1*:*his3* RNA ratio is 5).

Acknowledgements

We thank Christopher Brandl for communicating unpublished results, Jun Ma for *gal4* DNA containing the *leu2* disruption. This work was supported by grant GM 30186 from the National Institutes of Health.

References

Abei,M., Furtur,R., Prantl,F., Niederberger,P. and Hutter,R. (1984) *Curr. Genet.*, **8**, 165–172.

Allison,L.A., Moyle,M., Shales,M. and Ingles,C.J. (1985) *Cell*, **42**, 599–610.

Allison,L.A., Wong,J.K., Fitzpatrick,V.D., Moyle,M. and Ingles,C.J. (1988) *Mol. Cell. Biol.*, **8**, 321–329.

Arndt,K. and Fink,G. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8516–8520.

Bartolomei,M.S., Halden,N.F., Cullen,C.R. and Corden,J.L. (1988) *Mol. Cell. Biol.*, **8**, 330–339.

Bram,R. and Kornberg,R. (1985) *Proc. Natl. Acad. Sci. USA.*, **82**, 43–47.

Chen,W. and Struhl,K. (1985) *EMBO J.*, **4**, 3273–3280.

Chen,W. and Struhl,K. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2691–2695.

Chen,W., Tabor,S. and Struhl,K. (1987) *Cell*, **50**, 1047–1055.

Corden,J.L., Cadena,D.L., Ahearn,J.M.Jr. and Dahmus,M.E. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7934–7938.

Dahmus,M.E. and Kedinger,C. (1983) *J. Biol. Chem.*, **258**, 2303–2307.

Davison,B.L., Egly,J.-M., Mulvihill,E.R. and Chambon,P. (1983) *Nature*, **301**, 680–686.

Gidoni,D., Kadonaga,J.T., Barrera-Saldana,H., Takahashi,K., Chambon,P. and Tjian,T. (1985) *Science*, **230**, 511–517.

Gill,G. and Ptashne,G. (1987) *Cell*, **51**, 121–126.

Giniger,E., Varnum,S.M. and Ptashne,M. (1985) *Cell*, **40**, 767–774.

Guarente,L. (1988) *Cell*, **52**, 303–305.

Han,S., Hoar,E.T. and Guarente,L. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8562–8566.

Hill,D.E., Hope,I.A., Macke,J.P. and Struhl,K. (1986) *Science*, **234**, 451–457.

Hinnebusch,A.G. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6442–6446.

Horna,F.L. Glorioso,J.C. and Levine,M. (1988) *Genes Dev.*, **2**, 40–53.

Hope,I.A. and Struhl,K. (1985) *Cell*, **43**, 177–188.

Hope,I.A. and Struhl,K. (1986) *Cell*, **46**, 885–894.

Hope,I.A. and Struhl,K. (1987) *EMBO J.*, **6**, 2781–2784.

Hope,I.A., Mahadevan,S. and Struhl,K. (1988) *Nature*, **333**, 635–640.

Hurstel,S., Granger-Schnarr,M. and Schnarr,M. (1988) *EMBO J.*, **7**, 269–275.

Kakidani,H. and Ptashne,M. (1988) *Cell*, **52**, 161–167.

Lech,K., Anderson,K. and Brent,R. (1988) *Cell*, **52**, 179–184.

Ma,J. and Ptashne,M. (1987a) *Cell*, **48**, 847–853.

Ma,J. and Ptashne,M. (1987b) *Cell*, **51**, 113–119.

McKnight,S.L. and Tjian,R. (1986) *Cell*, **46**, 795–805.

McNeil,J.B. and Smith,M. (1986) *J. Mol. Biol.*, **187**, 363–378.

Melton,D.W., Konecki,D.S., Brennand,J. and Caskey,C.T. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2147–2151.

Nagawa,F. and Fink,G.R. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8557–8561.

Nonet,M., Sweetser,D. and Young,R.A. (1987) *Cell*, **50**, 909–915.

Parker,C.S. and Topol,J. (1984) *Cell*, **36**, 357–369.

Reynolds,G.A., Basu,K., Osborne,T.F., Chin,D.J., Gil,G., Brown,M.S., Goldstein,J.L. and Luskey,K.L. (1984) *Cell*, **38**, 275–285.

Sawadogo,M. and Roeder,R.G. (1985) *Cell*, **43**, 165–175.

Simon,M.C., Fisch,T.M., Benecke,B.J., Nevins,J.R. and Heintz,N. (1988) *Cell*, **52**, 723–729.

Struhl,K. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7385–7389.

Struhl,K. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7865–7869.

Struhl,K. (1985) *Nature*, **317**, 822–824.

Struhl,K. (1986) *Mol. Cel. Biol.*, **6**, 3847–3853.

Struhl,K. (1987a) *Cell*, **49**, 295–297.

Struhl,K. (1987b) *Cell*, **50**, 841–846.

Struhl,K. (1988) *Nature*, **332**, 649–650.

Struhl,K., Chen,W., Hill,D.E., Hope,I.A. and Oettinger,M.A. (1985) *Cold Spring Harbor Symp. Quant. Biol.*, **50**, 489–503.

Thireos,G., Driscoll-Penn,M. and Greer,H. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5096–5100.

Webster,N., Jin,J.R., Green,S., Hollis,M. and Chambon,M. (1988) *Cell*, **52**, 169–178.

West,R.W.Jr., Chen,S., Peitz,H., Burler,G. and Banerjee,M. (1987) *Genes Dev.*, **1**, 1118–1131.

Zalkin,H., Paluh,J.L., vanCleemput,M., Moye,W.S. and Yanofsky,C. (1984) *J. Biol. Chem.*, **259**, 3985–3992.

Received on August 11, 1988; revised on October 24, 1988