Transcriptional Activation by Yeast GCN4, a Functional Homolog to the *jun* Oncoprotein

K. STRUHL, C.J. BRANDL, W. CHEN, P.A.B. HARBURY, I.A. HOPE,* AND S. MAHADEVAN
Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02115

In the yeast *Saccharomyces cerevisiae*, the 50–100 genes involved in amino acid biosynthesis are regulated coordinately (see Fig. 1) (for review, see Hinnebusch 1986). Under normal growth conditions, these genes are transcribed at a basal level that is typical for yeast genes in general, approximately 1–2 mRNA molecules per cell (Struhl and Davis 1981). However, when yeast cells are starved for amino acids, they respond by inducing the transcription of all the coregulated genes by a factor of 3–5. The starvation response can be elicited by metabolic poisons or by mutations that inhibit the synthesis of any single amino acid or the level of tRNA charging. This general control mechanism differs from the situation in *Escherichia coli*, where regulation occurs at the level of individual biosynthetic pathways.

General control is best viewed as a mechanism to regulate protein synthesis by controlling the amount of amino acid precursors. This regulatory mechanism is important for controlling cell growth because yeast cells initiate new cell division cycles only if they have sufficient amino acids to complete the cycle; amino acid auxotrophs arrest at the start of the cycle upon being switched to medium lacking the required amino acid. Moreover, yeast strains containing *gcd* mutations do not properly regulate transcription of the amino acid biosynthetic genes, and they also arrest at the start of the cell cycle when grown at high temperatures (Wolfner et al. 1975). Thus, transcriptional regulation of amino acid biosynthetic genes represents part of a more global mechanism that regulates cell growth and the decision to initiate new cell division cycles.

Transcriptional induction of the amino acid biosynthetic genes is mediated by GCN4, a protein that binds specifically to the promoter regions of these genes (Hope and Struhl 1985; Arndt and Fink 1986). Strains lacking a functional *GCN4* gene product are unable to coordinately induce the transcription of these genes in response to amino acid starvation. GCN4 protein is made only during conditions of amino acid starvation, even though *GCN4* mRNA is made at all times (Hinnebusch 1984; Thireos et al. 1984). This novel translational control mechanism, which involves a 600-base RNA leader containing four AUG codons, explains why the amino acid biosynthetic genes are transcriptionally induced only during starvation conditions. In terms of signal transduction, it is sensible that GCN4, a regulator of genes involved in protein synthesis, is itself controlled by the translation process.

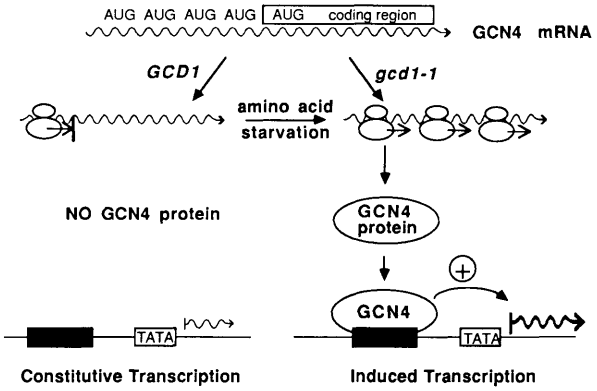

Figure 1. Molecular model for the general control of amino acid biosynthesis by GCN4. See text for details.

GCN4 binds as a dimer (Hope and Struhl 1987) to target sites whose consensus is the 9-bp palindrome ATGA(C/G)TCAT; this consensus sequence also represents the optimal GCN4-binding site (Hill et al. 1986). Extensive deletion analysis of GCN4 indicates that the 60 carboxy-terminal amino acids are sufficient both for specific DNA binding (Hope and Struhl 1986) and for dimerization (see Fig. 2) (Hope and Struhl 1987). These results, as well as the sequence of the DNA-binding domain, make it very unlikely that either of the major structural motifs, helix-turn-helix or zinc-finger, are involved in DNA recognition.

Although the GCN4 DNA-binding domain is neces-

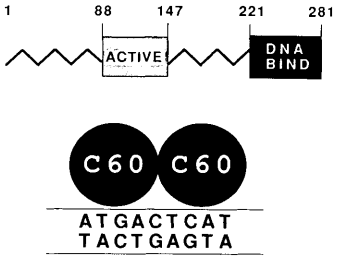

Figure 2. DNA-binding and transcriptional activation functions of GCN4 protein. (*Top*) DNA-binding domain (black box), transcriptional activation region (gray box), and nonessential regions (wavy lines) of GCN4 (281 amino acids; important residues indicated). (*Bottom*) Dimeric DNA-binding domain (the 60 carboxy-terminal amino acids) interacting with the optimal 9-bp dyad symmetric sequence.

*Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge, United Kingdom CB2 2QH.

sary for recognizing the appropriate promoters, a short acidic region in the center of the protein is required for transcriptional activation (Hope and Struhl 1986). Derivatives containing only the DNA-binding domain do not activate transcription in vivo, and indeed can actually repress transcription in certain promoters. The transcriptional activation region of GCN4 stimulates transcription when fused to a heterologous DNA-binding domain, the *E. coli* LexA repressor. The resulting LexA hybrid proteins activate transcription of promoters that contain LexA-binding sites as upstream elements.

This paper summarizes several aspects of our knowledge about GCN4 protein. These include its functional relationship to the jun oncoprotein, the structure and function of the transcriptional activation region, its ability to replace the TATA function, and its interaction with RNA polymerase II. These observations are discussed with respect to molecular mechanisms of transcriptional activation and to regulation of amino acid biosynthesis in response to starvation.

jun Oncoprotein Is Homologous to GCN4, Binds the Same DNA Sequences, and Activates Transcription in Yeast

The *jun* oncogene was originally derived from a defective avian sarcoma virus that causes fibrosarcomas in chickens and transforms chick embryo fibroblasts in cell culture. Surprisingly, the amino acid sequence of the carboxyl terminus of the predicted jun protein is 45% identical to the GCN4 DNA-binding domain, including a 30-amino-acid region in which there are 17 identical residues and 4 conservative differences (Fig. 3) (Vogt et al. 1987). This striking similarity suggested that GCN4 and jun might bind the same DNA sequences, a hypothesis supported by the similarity of the recognition sequences for GCN4 (Hill et al. 1986) and AP-1, a mammalian transcription factor that interacts with phorbol-ester-inducible promoter elements (Angel et al. 1987; Lee et al. 1987).

DNA binding. To investigate whether the jun oncoprotein and GCN4 bind to the same DNA sequences, molecules were constructed in which the region encoding the GCN4 DNA-binding domain was replaced by the homologous jun region (Fig. 3) (Struhl 1987a). Specifically, the molecules encoded LexA-GCN4-jun hybrid proteins containing the LexA DNA-binding domain, the GCN4 transcriptional activation region, and the presumptive jun DNA-binding domain. Since the GCN4 transcriptional activation region functions even when fused to a heterologous DNA-binding domain, the ability of the hybrid proteins to complement a *gcn4* mutation constitutes an in vivo DNA-binding assay for the *jun*-coding region.

A LexA-GCN4-jun protein containing the carboxy-terminal 112 amino acids of jun functionally substitutes for GCN4 in its ability to induce the expression of *HIS3* and other amino acid biosynthetic genes. A related protein containing only 99 carboxy-terminal amino acids of jun also functionally substitutes for GCN4, but with slightly less efficiency. Surprisingly, derivatives lacking the LexA DNA-binding domain do not complement the *gcn4* deletion. The requirement for the LexA DNA-binding domain cannot be due to its specific DNA-binding properties because the *HIS3* promoter region does not contain a Lex DNA-binding site. Instead, it has been suggested that the LexA domain facilitates dimerization (or formation of higher-order structures) of the jun domain.

To prove that GCN4 and jun recognize the same DNA sequences, plasmids expressing the LexA-GCN4-jun protein were introduced into a set of isogenic *gcn4* deletion strains that differ solely at the GCN4-binding site within the *HIS3* promoter (Struhl 1987a). The jun DNA-binding domain activates *HIS3* transcription only if the promoter region contains a functional GCN4-binding site. Moreover, maximal *HIS3* induction

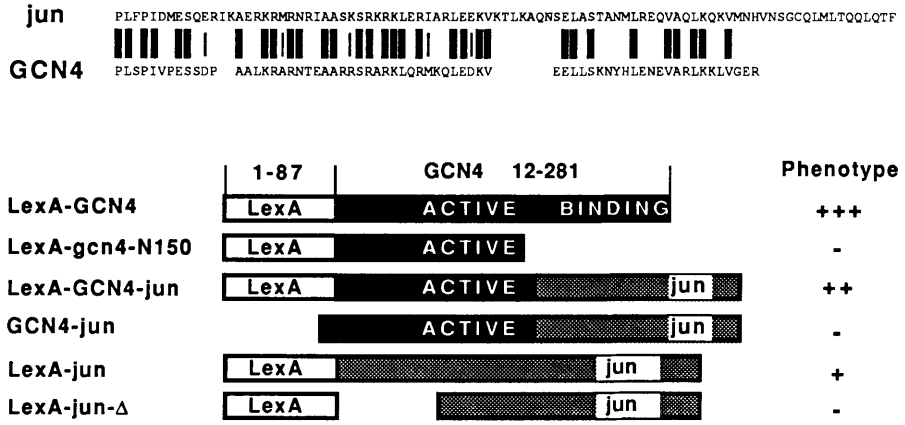

Figure 3. Structural and functional relationships between GCN4 and the jun oncoprotein. (*Top*) Comparison of the DNA-binding domains of jun and GCN4 with identical and similar residues indicated, respectively, by thick and thin lines. (*Bottom*) Structures of hybrid proteins containing the LexA DNA-binding domain (open box), GCN4 (black box), and jun (shaded box). The GCN4 phenotypes were determined by complementation of a *gcn4* deletion strain as described by Struhl (1987a, 1988).

mediated by the jun hybrid protein occurs in combination with the optimal GCN4-binding site. Thus, the GCN4 and jun DNA-binding regions behave homologously on a set of target sites, thereby indicating that they recognize very similar DNA sequences by a common structural motif.

Transcriptional activation. The above experiment does not indicate whether jun itself can activate transcription in yeast, because LexA-GCN4-jun contains the intact GCN4 activation region. However, a LexA-jun hybrid protein in which the entire *jun*-coding region is fused directly to the LexA DNA-binding domain can functionally replace GCN4, although with a slightly reduced efficiency in comparison to LexA-GCN4-jun (Fig. 3) (Struhl 1988). In addition, LexA-jun activates transcription through the heterologous LexA DNA-binding domain at a level comparable to that achieved by LexA-GCN4-jun or LexA-GCN4. These observations indicate that the *jun* oncogene contains a sequence(s) that functions as an efficient transcriptional activation region in yeast.

The jun protein contains a region between residues 15 and 59 with a net negative charge of −7 and a region between residues 87 and 102 with a net charge of −4. Deletions that remove more than 100 amino-terminal residues of jun lack both acidic regions and confer extremely low levels of activation (Struhl 1988). Deletions with end-points between residues 54 and 71, which remove one of the acidic regions, show a two- to fivefold decrease in the level of expression. Thus, as is the case for the yeast GCN4 and GAL4 activator proteins (Hope and Struhl 1986; Ma and Ptashne 1987a), the acidic regions of jun appear to be important for transcriptional activation in yeast.

The similar DNA-binding properties of GCN4 to both the mammalian transcription factor AP-1 and to the jun oncoprotein led to the demonstration that jun represents an oncogenic version of a normal cellular transcription factor (Bohmann et al. 1987; Angel et al. 1988). The fact that jun activates transcription in yeast indicates that this oncoprotein can interact functionally with the basic transcription machinery of yeast. The obvious implication is that the basic transcription machineries of eukaryotic organisms from yeast to man are evolutionarily conserved. In support of this idea, GAL4 activates transcription in mammalian cells (Kakidani and Ptashne 1988; Webster et al. 1988), and the fos oncoprotein, a hypothetical transcription factor, activates transcription in yeast cells (Lech et al. 1988). Thus, it appears that mRNA transcriptional initiation in all eukaryotes may occur by a common molecular mechanism.

Structural and Functional Characterization of the GCN4 Transcriptional Activation Region

As mentioned earlier, deletion analyses localized the GCN4 transcriptional activation to a short acidic region in the center of the protein (Hope and Struhl 1986; Hope et al. 1988). Derivatives retaining only 35–40 amino acids from this acidic region are sufficient for wild-type levels of transcriptional activation when fused directly to the GCN4 DNA-binding domain. Moreover, the distance and orientation of the activation region with respect to the DNA-binding domain are functionally unimportant. This indicates that the activation region is an autonomous function and that there is no requirement for a spacer between the activation region and DNA-binding domain.

Several lines of evidence indicate that yeast transcriptional activation functions are defined by short acidic regions with little sequence homology. First, different portions of the GCN4 acidic region are equally capable of activating transcription, even though their primary sequences are dissimilar (Hope and Struhl 1986). Second, the transcriptional activation regions of GCN4 and jun are acidic, yet are not homologous to the GCN4 region. Third, acidic character is the common feature of transcriptional activation regions selected from *E. coli* DNA segments (Ma and Ptashne 1987b). Fourth, GAL4 mutations that increase or decrease activation usually increase or decrease negative charge, respectively (Gill and Ptashne 1987). These results strongly suggest that transcriptional activation regions are not defined by a specific primary sequence but rather by a more general structure feature, presumably involving net negative charge.

High-resolution deletions. The nature of the GCN4 transcriptional activation region was investigated in more detail by a high-resolution deletion analysis (Hope et al. 1988). A series of amino-terminal segments ranging in size from 3 to 41 residues was fused to the 100 carboxy-terminal amino acid DNA-binding domain to generate three classes of fusions representing the different reading frames (Fig. 4). Since this affects the amino acid sequence at the junction between the activation peptide and the DNA-binding domain, derivatives containing the same number of GCN4 residues from the activation region can have a different number of acidic residues.

Strikingly, progressive deletion did not reveal a position where there was a sudden complete loss in activity but instead yielded a series of small, stepwise reductions in activity. Seven discrete phenotypes were distinguished, ranging from high GCN4 activity for the largest derivatives to the repression phenotype when only the GCN4 DNA-binding domain is present. The fact that all 26 derivatives fit a simple pattern strongly argues that the different levels of activity reflect differences in amount of transcriptional activation function.

GCN4 activity appears to be directly related to the size of the transcriptional activation region remaining. There is not a single case where a shorter region activates transcription more efficiently than a longer region. In contrast, there is no such precise relationship of transcriptional activity to the number of acidic residues. In particular, there are several examples in which derivatives with fewer acidic residues activate transcription better than derivatives with more acidic residues.

	116 123 130 137 144 - - - - | + - | - - | - - | - + - | - -	amino acids	net charge	GCN4 activity
Δ35	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLADKAIESTEEVSL	wt	-11	
Δ36	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLADKAIESTEEgstk	41	-10	
Δ37	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLADKAIESTEgstk	40	-9	4
Δ38	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLADKAIESTEvdq	40	-11	
Δ39	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLADKAIESgrpvk	38	-7	
Δ40	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLADKAvdq	35	-9	
Δ41	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLADKAgrpvk	35	-6	
Δ42	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLADKArstk	35	-6	3
Δ43	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLAgstk	32	-7	
Δ44	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSLgstk	31	-7	
Δ45	MFEYENLEDNSKEWTSLFDNDIPVTTDDVSgrpvk	30	-6	2
Δ46	MFEYENLEDNSKEWTSLFDNDIPgstk	23	-5	
Δ47	MFEYENLEDNSKEWTSLFDNgrpvk	20	-3	1
Δ48	MFEYENLEDNSKEWTSLFDrstk	19	-3	
Δ49	MFEYENLEDNSKEWTSLFDgrpvk	19	-3	
Δ50	MFEYENLEDNSKEWTSLFgstk	18	-3	0
Δ51	MFEYENLEDNSKEWTSLwstk	17	-3	
Δ52	MFEYENLEDNSKEWTgrpvk	15	-2	
Δ53	MFEYENLEDNSKEWgrpvk	14	-2	-1
Δ54	MFEYENLEDNSKgstk	12	-2	
Δ55	MFEYENLEDNSvdq	11	-5	
Δ56	MFEYENLEDgrpvk	9	-2	
Δ57	MFEYENLEDrstk	9	-2	
Δ58	MFEYENLEgstk	8	-2	
Δ59	MFEYENLvdq	7	-3	-2
Δ60	MFEgrpvk	3	+1	

Figure 4. High-resolution deletions of the GCN4 transcriptional activation region. For each derivative, the indicated sequence was fused to the 100-residue GCN4 DNA-binding domain. Shown are the number of GCN4 residues (uppercase letters) excluding joint sequences (lowercase letters), the net charge, and the level of GCN4 activation as determined by Hope et al. (1988). The location of charged residues and the probable boundaries that distinguish the phenotypic classes are shown above the sequences.

GCN4 structure. The nonstringent primary sequence requirements and the observation that progressive deletion of the activation region causes small stepwise reductions in activity strongly indicate that transcriptional activation regions do not have a defined tertiary structure such as is found in active sites or domains in a protein. This was confirmed by treating wild-type and deletion proteins with proteases that preferentially cleave unstructured protein regions (Hope et al. 1988). A stable carboxy-terminal domain was obtained by treatment with elastase, subtilisin, or papain under conditions that completely cleaved the amino-terminal portion of the protein. Thus, GCN4 contains an independently structured DNA-binding domain, with the rest of the protein being relatively unstructured.

An interesting structural feature of GCN4 is revealed by chymotrypsin digestion. Specifically, chymotrypsin generates two equally stable intermediates both for GCN4 and for many amino-terminally deleted derivatives (Hope et al. 1988). In other words, large amino-terminal segments of GCN4 are resistant to protease digestion as if part of a structured domain, yet can be removed without destroying the integrity of that domain. However, a stable amino-terminal fragment is not generated for derivatives lacking the transcriptional activation region. Thus, the transcriptional activation region influences GCN4 structure even though it resides in a part of the protein that lacks tertiary structure.

Molecular implications. The strong correlation between the length of the GCN4 activation region and level of transcriptional activity is strongly suggestive of a repeating structure consisting of units that act additively. Moreover, the unusual chymotrypsin digestion pattern strongly supports the view that the activation region has some kind of structure. Acidic character, although important, cannot be the only important structural feature because the level of transcriptional activation is only moderately correlated with net negative charge. One suggestive clue to the structure of the activation region is that the boundaries defining the stepwise levels of activation may occur every seven amino acid residues (116, 123, 130, 137, and 144), a repeat unit consistent with two turns of an α helix.

The acidic region of GCN4 could form three α helices, each containing five to six turns, that are separated by proline residues 87, 106, 129, and 152. Interestingly, these helices have amphipathic character, i.e., acidic and hydrophobic residues tend to be clustered along separate surfaces. In this regard, the anomalously high activity of LexA-GCN4-N125 (Hope and Struhl 1986) may be due to extraneous residues at the carboxyl terminus that would fortuitously improve the length and amphipathic quality of the hypothetical central helix. In addition, it has been shown recently that a synthetic 15-amino-acid region whose sequence is consistent with forming two turns of an amphipathic helix could confer some transcriptional activity in yeast when fused to the GAL4 DNA-binding domain (Giniger and Ptashne 1987). However, as the major GAL4 and other activation regions are unlikely to form amphipathic helices, a simple relationship between this structure and function appears unlikely.

We have proposed that the activation region is a dimer of intertwined α helices, one helix from each GCN4 monomer (Fig. 5) (Hope et al. 1988). Chymotrypsin reveals a structure of GCN4 that depends on the acidic activation region. The formation/stability of this structure should be facilitated by the stability of the dimeric DNA-binding domain even in the absence of target DNA (Hope and Struhl 1987). Moreover, since the LexA domain binds very poorly to its operator because of weak dimerization, it is likely that transcriptional stimulation through the LexA domain also requires that the activation region facilitate dimerization (Hope and Struhl 1986; Ma and Ptashne 1987a; Struhl 1987a). The dimerization model also explains why amphipathic helices should form functional transcriptional activation regions, since it would easily permit a structure involving interacting hydrophobic residues that are protected from solvent and exposed acidic residues.

Multiple Proteins Can Perform the TATA Function

It is commonly assumed that TATA sequences are general promoter elements that are recognized by a common transcription factor that is part of the basic transcriptional machinery. The yeast *his3* promoter region contains two distinct classes of TATA elements, constitutive (T_C) and regulatory (T_R), that are defined by their interactions with upstream promoter elements, selectivity of initiation sites, and chromatin structure (Struhl 1986). Transcription dependent on T_C is initiated equally from two sites, +1 and +12, whereas transcription dependent on T_R initiates preferentially from the +12 site; this selectivity is determined primarily by the distance between T_R and the initiation sites. Transcriptional activation by GCN4 and GAL4 occurs only in combination with the T_R element, not T_C. The T_C element maps between −83 and −53, and T_R maps between −55 and −35, regions in which there are several TATA-like sequences.

Genetic evidence for a specific T_R-binding protein. To determine if T_R and T_C represent binding sites for distinct proteins, the structural requirements of the T_R element were determined by saturation mutagenesis (Chen and Struhl 1988a). To avoid the complications of redundant elements, the phenotypic analysis was carried out by using a *gal-his3* hybrid promoter that could not support transcription because it lacked any functional TATA element. Specifically, oligonucleotides containing prospective T_R elements were cloned between UAS_G and the *his3* mRNA-coding sequences to identify a minimal sequence that was sufficient for full T_R function. Once identified, all possible base-pair substitutions of this minimal functional sequence were tested for their ability to activate transcription (Fig. 6).

The sequence TATAAA is highly conserved among eukaryotic promoters, and it is located within the T_R element. Insertion of an oligonucleotide containing TATAAA into the *gal-his3* promoter at the proper distance from *his3* mRNA start sites yields a promoter that behaves indistinguishably from a fully functional *gal-his3* fusion. Upon galactose induction, this promoter confers equal levels of mRNA transcripts that are initiated with the characteristic preference at the +12 site. Thus, TATAAA is functionally equivalent to the *his3* T_R element.

Surprisingly, 17 out of the 18 possible single mutations and 9 out of the 10 double mutations of the TATAAA sequence abolish T_R function (Chen and Struhl 1988a). The phenotypes are similar to that observed with a control molecule lacking the TATAAA oligonucleotide. The only functional sequence out of the single substitutions is TATA*T*A, which produces half of the *his3* RNA level attained by TATAAA. The high sequence specificity of the *his3* T_R element provides strong genetic evidence for a T_R-binding protein; indeed, this protein is defined by its DNA sequence requirements.

The proposal of a specific T_R-binding protein with a high sequence specificity has important implications for yeast promoters that lack sequences compatible with

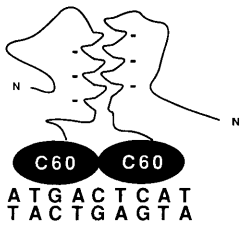

Figure 5. Structural model for GCN4 protein. The dimeric GCN4 DNA-binding domain is depicted as black ovals, the transcriptional activation region is shown as a dimer of interacting α helices (wavy line) with acidic residues (−) exposed, and the remainder of the amino terminus is shown as unstructured.

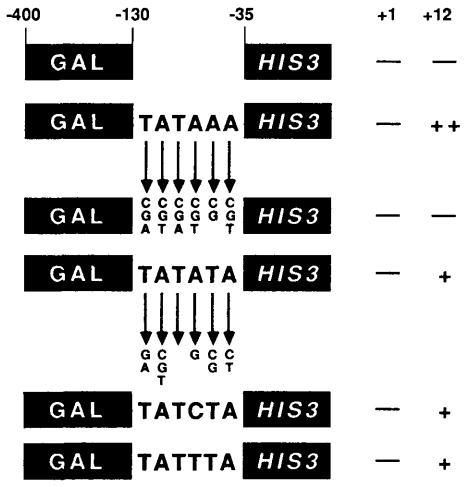

Figure 6. Saturation mutagenesis of the *his3* T_R element. All *gal-his3* hybrid promoters contain a 365-bp *gal*-enhancer fragment, *his3* sequences downstream from −35, and the indicated oligonucleotides containing various TATA sequences. The ability to activate *his3* transcription from the +1 and +12 initiation sites is indicated (see Chen and Struhl 1988a).

T_R function. Assuming such promoters require "downstream elements" for transcription, it follows that these elements cannot be recognized by the T_R protein but rather are targets for another DNA-binding protein. For example, the T_C element necessary for constitutive *his3* expression does not have a sequence that fits the T_R rules. Thus, it is almost certain that the T_R and T_C elements interact with different proteins, an explanation that easily accounts for the functional distinctions between T_R and T_C elements with regard to their interactions with upstream activator proteins (Struhl 1986). By analogy with bacterial σ factors that interact with the core RNA polymerase to generate distinct holoenzymes that recognize different promoter sequences, yeast cells may contain multiple proteins that carry out a related "downstream element function" but have different specificities for DNA binding. Two additional lines of evidence discussed below strongly support the concept of multiple TATA-binding proteins.

Certain double mutations of TATAAA are functional TATA elements. Since 17 out of 18 mutations of TATAAA abolish transcription from the *gal-his3* promoter, it was expected that double mutations of this sequence would also be nonfunctional. Indeed, the initial experiments revealed that 9 out of 10 double mutations were defective in supporting transcription. Interestingly, however, the TAT*CT*A double mutation behaves as a functional TATA element, whereas the related single mutation TAT*C*AA does not (Fig. 6) (Chen and Struhl 1988a). This suggests the possibility that a protein distinct from T_R binds to the double mutant DNA and activates transcription.

To eliminate the alternative explanation that the consensus T_R element is actually TATATA and that the TATCTA double mutation actually represents a single mutation, 11 single substitutions of the sequence TATATA were examined for T_R function in the *gal-his3* promoter (Fig. 6) (P.A.B. Harbury et al., in prep.). Ten of these mutations prevent *his3* transcription, thus suggesting that TATATA is not an appropriate consensus sequence for the T_R element. Interestingly, the sole exception, TAT*TT*A, is related by a single base-pair change to the other exceptional double mutant TAT*CT*A. The observations that two related double mutants but not the corresponding single mutants act as downstream elements provide genetic evidence for two distinct proteins that recognize related TATA-like DNA sequences.

TATA elements that function in combination with GCN4 but not GAL4. In attempting to find functional distinctions between TATA elements, DNA fragments containing the various mutated TATAAA sequences and the *his3* structural gene were cloned downstream from the GCN4-binding site in the native *his3* promoter (P.A.B. Harbury et al., in prep.). In almost all cases, mutations that prevent activation of the *gal-his3* promoter by GAL4 protein also are inactive in combination with the GCN4-binding site. However, two mutations, TATAA*G* and TATA*TA*C have the novel property of activating transcription in combination with GCN4 but not with GAL4. The level of GCN4 activation through these TATA derivatives is about 50% the level achieved with the TATAAA sequence. The simplest interpretation of these results is that there are two "TATA-binding" proteins: T_R, which recognizes TATAAA and interacts functionally with both GCN4 and GAL4, and a distinct protein that recognizes a related sequence but interacts only with GCN4.

Evidence for an Interaction between GCN4 and RNA Polymerase

As discussed earlier, it is likely that the GCN4 transcriptional activation region does not encode a catalytic activity but rather stimulates transcription by interacting with another protein. From experiments involving transcription by bacteriophage T7 RNA polymerase in yeast cells, it appears unlikely that acidic activation regions interact with histones and activate transcription by affecting chromatin structure (Chen et al. 1987). Instead, TATA-binding proteins and RNA polymerase II are the two obvious candidates for targets of interaction by GCN4 and other upstream activator proteins.

GCN4 can replace the TATA function. If GCN4 stimulates transcription by contacting RNA polymerase, then perhaps it would activate transcription in the absence of a TATA element when a GCN4-binding site is located at the position normally occupied by the TATA sequence. Indeed, when an oligonucleotide containing a GCN4-binding site is inserted in place of the T_R element (allele *his3*-GG1), *his3* transcription is efficiently activated to 30–50% of the maximal GCN4-induced level of the wild-type *his3* gene (Fig. 7) (Chen

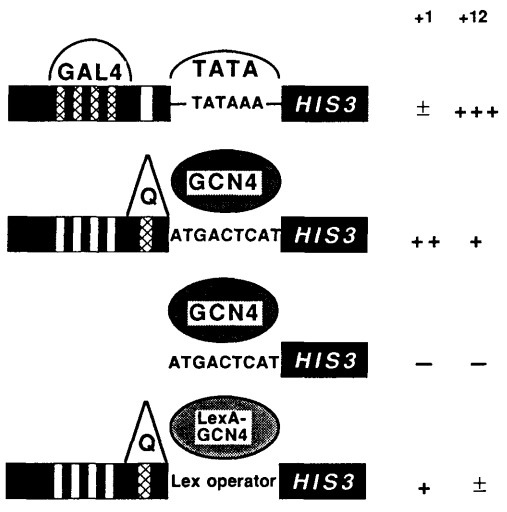

Figure 7. GCN4 activates transcription when its binding site replaces the *his3* TATA element. The *gal* enhancer fragment containing four GAL4-binding sites and a putative Q site (Chen and Struhl 1988b; C.J. Brandl and K. Struhl, in prep.; see text) is fused upstream of the T_R element, GCN4-binding site, or *lexA* operator. The ability to activate *his3* transcription from the +1 and +12 initiation sites is indicated.

and Struhl 1988b). Transcription from *his3*-GG1 depends on the GCN4-binding site and on the presence of GCN4, rather than some fortuitous function interacting with the GCN4-binding site. Moreover, replacement of the GCN4-binding site with the *E. coli* LexA operator, allele *his3*-GL1, eliminates *his3* expression except in the presence of a plasmid expressing the LexA-GCN4 hybrid protein.

Interestingly, transcription from *his3*-GG1 is initiated from both the +1 and +12 sites with a slight preference for the +1 site. Both the level of transcription and the selection of initiation sites are not affected when the distance between the GCN4-binding site and the mRNA initiation site is varied by +11, +7, +5, and −13; in the last case, the GCN4-binding site is only 28 bp from the +1 initiation site. Since *his3* mRNA start sites are determined primarily by specific "initiator" sequences, not by the distance to the TATA element (Chen and Struhl 1985), 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, the observed initiation pattern resembles that mediated by the T_C element, rather than GCN4 induction through the T_R element, where initiation occurs primarily at the +12 site (Struhl 1986).

Transcriptional stimulation by GCN4 requires both the DNA-binding domain and the acidic activation function (Chen and Struhl 1988b). Specifically, all amino- and carboxy-terminal deletions that lack the GCN4 acidic transcriptional activation region fail to stimulate *his3* transcription. A few derivatives do not stimulate transcription from *his3*-GG1, although they activate the wild-type *his3* promoter. Although distinct regions of GCN4 might be necessary for activation of these promoters, it is more likely that the requirements for an acidic activation region are qualitatively similar but quantitatively more stringent for transcription for *his3*-GG1.

Although previous *gal-his3* promoters activate *his3* transcription only when cells are grown in galactose medium, transcription from *his3*-GG1 is equally efficient in glucose and galactose media. However, GCN4 alone is not sufficient to activate transcription in this situation because removal of the *gal* fragment abolishes transcription. Similarly, other *his3* promoter deletions lacking the TATA region but containing the GCN4-binding site near the initiation sites do not activate transcription. Thus, some genetic element within the *gal* fragment is required for GCN4 to activate transcription from *his3*-GG1. This element, hereby termed Q, is distinct from the GAL4-binding sites because deletion of the *gal4* gene does not affect transcription and because it maps about 30 bp downstream from the GAL4-binding sites (Fig. 7) (C.J. Brandl and K. Struhl, in prep.). Thus, even when GCN4 replaces the TATA function, at least two distinct promoter elements are necessary for transcription.

The observation that a GCN4-binding site can functionally replace the T_R element strongly suggests that GCN4 is one of the multiple proteins that can carry out the TATA function. The different initiation patterns probably reflect distinct requirements of GCN4 and the T_R protein for the distance between its binding site and the mRNA initiation sites. Although *his3*-GG1 is an artificially constructed promoter, it resembles the yeast *TRP3* promoter, which lacks the conserved TATAAA sequence and instead contains a GCN4-binding site 28 bp upstream of the mRNA start site. Thus, GCN4 functions both upstream of a TATA element and in place of a TATA element, suggesting that there is no intrinsic difference between an upstream and TATA activator protein. The observation that GCN4 can substitute for the TATA function is inconsistent with the view that GCN4 stimulates transcription solely by contacting a TATA-binding factor, and is more suggestive of a direct interaction between GCN4 and RNA polymerase II.

Biochemical evidence for a GCN4-RNA polymerase II interaction. Affinity chromatography was employed to establish whether GCN4 and RNA polymerase interact directly (C.J. Brandl and K. Struhl, in prep.). In one approach, GCN4 was produced in *E. coli*, purified to apparent homogeneity, and covalently coupled to Sepharose. A highly purified preparation of yeast RNA polymerase II (estimated to be 10–50% pure) was mixed with about 10^5 cpm of ^{35}S-labeled total yeast proteins, applied to the column, and washed in 0.1 M NaCl. Under these conditions, 95% of the ^{35}S-labeled protein flowed through the column, whereas about 90% of the RNA polymerase II activity was retained. The RNA polymerase II was partially eluted from the column at 0.2 M NaCl and completely eluted at 0.3 M NaCl. In contrast, essentially all the RNA polymerase II flowed through columns containing covalently bound bovine serum albumin or no protein. Thus, RNA polymerase II preferentially associates with GCN4.

In the converse experiment, RNA polymerase II was coupled to the column and ^{35}S-labeled GCN4 protein synthesized in vitro from the cloned gene (Hope and Struhl 1985) was applied. Under the conditions described above, most of the GCN4 bound to the RNA polymerase II column, whereas essentially none of the protein bound to the bovine serum albumin column. These reciprocal binding experiments demonstrate a direct interaction between RNA polymerase II and GCN4.

Interestingly, the interaction with RNA polymerase requires the intact GCN4 DNA-binding domain but not the transcriptional activation region. First, a column containing the 100 carboxy-terminal amino acids of GCN4 retained RNA polymerase II as well as a column containing the intact GCN4 protein. Second, the RNA polymerase column was equally efficient at retaining GCN4 and the derivative containing only the 60 carboxy-terminal residues. Third, a derivative lacking only the 11 carboxy-terminal residues of GCN4, which fails to bind DNA, is also incapable of binding to the RNA polymerase II column.

A Molecular Model for Transcriptional Activation

The short acidic transcriptional activation regions of GCN4 and other activator proteins are likely to be surfaces used for interactions with other proteins of the transcription machinery. It is very unlikely that short regions of limited homology could encode catalytic activities such as topoisomerases or nucleases. Experimental evidence in favor of a protein contact model and against a chromatin accessibility model comes from the observation that the yeast GAL4 upstream activator protein cannot stimulate transcription by bacteriophage T7 RNA polymerase in yeast (Chen et al. 1987).

Although protein interactions are usually imagined as involving highly specific contacts of complementary surfaces, the wide variety of activation sequences argues for a relatively low specificity, ionic association between the acidic activation region and a basic pocket of the contacted protein. However, the degree of transcriptional stimulation is not determined solely by the number of acidic residues, but is directly correlated with the length of the activation region (Fig. 3) (Hope et al. 1988). In light of the structural model for GCN4 (Fig. 4), the activation region may be viewed as an "acidic wand" that reaches out and touches a component of the basic transcription machinery.

Previously, we suggested that the acidic activation region interacts with a TATA-binding protein (Struhl 1987b) mainly because GCN4 and GAL4 activate *his3* transcription in combination with the T_R element but not the T_C element (Struhl 1986). The high sequence specificity of the T_R element and the absence of functional T_R sequences in the region containing the T_C element strongly implicate distinct proteins interacting with the T_R and T_C elements (Chen and Struhl 1988a). The simplest interpretation is that the GCN4 and GAL4 activation regions can associate with a T_R-binding protein to stimulate transcription, whereas they are unable to interact with a T_C protein. Similarly, the TATA mutations that permit activation by GCN4 but not GAL4 are most easily explained by a protein distinct from T_R that can interact functionally only with GCN4. Thus, these functional distinctions between TATA elements provide genetic evidence for an interaction between upstream activator and TATA-binding proteins. These functional interactions may be related to the cooperative binding of a mammalian upstream activator and TATA protein to DNA (Sawadogo and Roeder 1985).

It is generally assumed that the conventional TATA-binding protein is part of the basic transcription machinery and, presumably, is associated with RNA polymerase II. Thus, a simple model for transcriptional initiation is that the interaction between GCN4 and the TATA-binding protein is the crucial and sufficient step for activation. The interaction could either facilitate the binding of the TATA factor to its target site or allosterically affect the T_R protein such that it would be able to promote transcription more efficiently. However, the observation that GCN4 can replace the TATA function argues strongly against the model in which a contact between GCN4 and the T_R factor is essential for activation. The formal possibility that GCN4 and the T_R protein could interact in the absence of the T_R-binding site is unlikely because it cannot account for the drastic effects of point mutations in the T_R element nor for the T_C-like initiation pattern observed from the *his3*-GG1 promoter. Instead, the fact that GCN4 can activate transcription even when its binding site is extremely close to the mRNA initiation sites is more consistent with a direct contact between GCN4 and RNA polymerase. Of course, the biochemical experiments indicate that GCN4 and RNA polymerase II can physically interact (C.J. Brandl and K. Struhl, in prep.). However, it remains to be determined if this interaction observed in vitro is relevant for transcriptional activation in vivo.

From all these considerations, we propose the following model for the molecular mechanism of eukaryotic transcriptional initiation (Fig. 8). First, RNA polymerase II needs at least two qualitatively different contacts by promoter-binding proteins. This view explains why at least two promoter elements are necessary and why only certain combinations of two proteins can function together. In a typical promoter, transcriptional stimulation is achieved by an upstream activator protein that contains an acidic wand and a conventional TATA-binding protein. Second, upstream activator proteins, such as GCN4, contact proteins such as T_R and RNA polymerase II. The contact to T_R-like proteins involves the acidic wand, whereas the contact to RNA polymerase II involves a sequence located within the DNA-binding domain. One prediction of this model is that it might be possible to obtain mutations in the DNA-binding domain that fail to stimulate transcription yet retain DNA-binding activity. Thus, a functional transcription machinery is viewed as a complex involving a set of interactions between the various components. Complex transcriptional regulatory patterns are achieved by the specific proteins that bind to the promoter, by environmentally or develop-

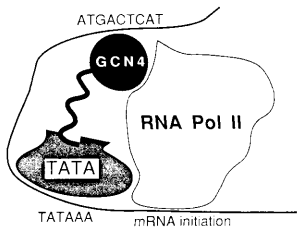

Figure 8. Molecular model for transcriptional activation. GCN4 (black) is shown as binding to its target DNA site, the TATA-binding protein (dark gray), and RNA polymerase II. The TATA-binding protein is shown as also interacting with its target sequence and RNA polymerase II. The interaction between GCN4 and the TATA-binding protein is proposed to occur through the acidic activation region (thick wavy line), and the DNA is illustrated as bending to allow for the protein-protein interactions.

mentally controlled cofactors that affect the synthesis or activity of the factors, and by compatibility between different protein factors.

ACKNOWLEDGMENT

This work was supported by grant GM-30186 from the National Institutes of Health to K.S.

REFERENCES

Angel, P., E.A. Allegretto, S.T. Okino, K. Hattori, W.J. Boyle, T. Hunter, and M. Karin. 1988. Oncogene *jun* encodes a sequence-specific *trans*-activator similar to AP-1. *Nature* **332:** 166.

Angel, P., M. Imagawa, R. Chiu, B. Stein, R.J. Imbra, H.J. Rahmsdorf, C. Jonat, P. Herrlich, and M. Karin. 1987. Phorbol ester-inducible genes contain a common *cis* element recognized by a TPA-modulated *trans*-acting factor. *Cell* **49:** 729.

Arndt, K. and G. Fink. 1986. GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5′ TCACTC 3′ sequences. *Proc. Natl. Acad. Sci.* **83:** 8516.

Bohmann, D. T.J. Bos, A. Admon, T. Nishimura, P.K. Vogt, and R. Tjian. 1987. Human proto-oncogene c-*jun* encodes a DNA binding protein with structural and functional properties of transcription factor AP-1. *Science* **238:** 1386.

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.* **4:** 3273.

———. 1988a. Saturation mutagenesis of a yeast *his3* TATA element: Genetic evidence for a specific TATA-binding protein. *Proc. Natl Acad. Sci.* **85:** 2691.

———. 1988b. Yeast upstream activator protein GCN4 can replace the TATA function required for transcriptional initiation. *EMBO J.* (in press).

Chen, W., S. Tabor, and K. Struhl. 1987. Distinguishing between mechanisms of the eukaryotic transcriptional activation with bacteriophage T7 RNA polymerase. *Cell* **50:** 1047.

Gill, G. and M. Ptashne. 1987. Mutants of GAL4 protein altered in an activation function. *Cell* **51:** 121.

Giniger, E. and M. Ptashne. 1987. Transcription in yeast activated by a putative amphipathic α helix linked to a DNA binding unit. *Nature* **330:** 670.

Hill, D.E., I.A. Hope, J.P. Macke, and K. Struhl. 1986. Saturation mutagenesis of the yeast *HIS3* regulatory site: Requirements for transcriptional induction and for binding by GCN4 activator protein. *Science* **234:** 451.

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

———. 1986. The general control of amino acid biosynthetic genes in the yeast *Saccharomyces cerevisiae*. *CRC Crit. Rev. Biochem.* **21:** 277.

Hope, I.A. and K. Struhl. 1985. GCN4 protein, synthesized in vitro, binds to *HIS3* regulatory sequences: Implications for the general control of amino acid biosynthetic genes in yeast. *Cell* **43:** 177.

———. 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* **46:** 885.

———. 1987. GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. *EMBO J.* **6:** 2781.

Hope, I.A., S. Mahadevan, and K. Struhl. 1988. Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein. *Nature* **333:** 635.

Kakidani, H. and M. Ptashne. 1988. GAL4 activates gene expression in mammalian cells. *Cell* **52:** 161.

Lech, K., K. Anderson, and R. Brent. 1988. DNA-bound fos proteins activate transcription in yeast. *Cell* **52:** 179.

Lee, W., P. Mitchell, and R. Tjian. 1987. Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell* **49:** 741.

Ma, J. and M. Ptashne. 1987a. Deletion analysis of GAL4 defines two transcriptional activating segments. *Cell* **48:** 847.

———. 1987b. A new class of yeast transcriptional activators. *Cell* **51:** 113.

Sawadogo, M. and R.G. Roeder. 1985. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. *Cell* **43:** 165.

Struhl, K. 1986. Constitutive and inducible *Saccharomyces cerevisiae* promoters: Evidence for two distinct molecular mechanisms. *Mol. Cell. Biol.* **6:** 3847.

———. 1987a. Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. *Cell* **49:** 295.

———. 1987b. The DNA-binding domains of the jun oncoprotein and the yeast GCN4 transcriptional activator are functionally homologous. *Cell* **50:** 841.

———. 1988. The JUN oncoprotein, a vertebrate transcription factor, activates transcription in yeast. *Nature* **332:** 649.

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 regulatory gene. *Proc. Natl. Acad. Sci.* **81:** 5096.

Vogt, P.K., T.J. Bos, and R.F. Doolittle. 1987. Homology between the DNA-binding domain of the GCN4 regulatory protein of yeast and the carboxy-terminal region of a protein coded for by the *onc* gene *jun*. *Proc. Natl. Acad. Sci.* **84:** 3316.

Webster, N., Jr., J.R. Jin, S. Green, M. Hollis, and P. Chambon. 1988. The yeast UAS_G is a transcriptional enhancer in human HeLa cell in the presence of the GAL4 *trans*-activator. *Cell* **52:** 169.

Wolfner, M., D. Yep, F. Messenguy, and G.R. Fink. 1987. Integration of amino acid biosynthesis into the cell cycle of *Saccharomyces cerevisiae*. *J. Mol. Biol.* **96:** 273.