Cell, Vol. 46, 885–894, September 12, 1986, Copyright © 1986 by Cell Press

Functional Dissection of a Eukaryotic Transcriptional Activator Protein, GCN4 of Yeast

Ian A. Hope and Kevin Struhl
Department of Biological Chemistry
Harvard Medical School
Boston, Massachusetts 02115

Summary

Yeast GCN4 protein binds specifically to the promoters of amino acid biosynthetic genes and coordinately induces their transcription. Serially deleted GCN4 and hybrid LexA-GCN4 proteins were assayed for specific DNA binding activity in vitro, and for stimulation of transcription in vivo. The specific DNA binding activity resides in the 60 C-terminal amino acids, a basic region of GCN4. However, certain deletions containing the entire DNA binding region are unable to activate transcription and instead act as repressors in vivo. The activation function appears to critically involve just 19 amino acids that are centrally located in an acidic region of GCN4. In addition to their functional separation, the DNA binding and transcriptional activation regions of the protein can be separated physically by elastase cleavage. The implications of these results for the mechanisms of DNA sequence recognition and transcription activation are discussed.

Introduction

Transcription of E. coli genes is mediated directly by the binding of RNA polymerase to promoter sequences. Transcriptional induction above the basal level requires activator proteins that recognize specific regulatory DNA sequences that are upstream of, but close to, the promoter sequences (reviewed by Von Hippel et al., 1984). However, there are λ*cI* and P22*c2* mutant proteins that are unable to induce transcription although they bind to their recognition sites as well as their wild-type counterparts (Guarente et al., 1982; Hochschild et al., 1983). These positive control mutants indicate that DNA binding and transcriptional activation are separate functions. Although the mechanism of activation is not understood in detail, it is believed to involve specific interactions between the activator and RNA polymerase. Single amino acid substitutions in the *cI* gene product that affect the activation function map close together on the surface of the DNA binding domain where they would probably be in close contact with a bound RNA polymerase that is ready to initiate transcription. Catabolite activator protein (CAP) seems to stimulate transcription by increasing the affinity of RNA polymerase for the promoters (Malan et al., 1984).

The mechanism of eukaryotic gene transcription, although less understood, is clearly different. Eukaryotic promoters are composed of TATA and upstream elements. The upstream elements (sometimes called enhancers or UAS) are of particular interest because they can act at long and variable distances with respect to the transcriptional initiation site. These promoter elements are not recognized by RNA polymerase II but rather by specific transcription factors. Several eukaryotic activator proteins have been identified by biochemical purification of specific DNA binding proteins from crude nuclear extracts (Dynan and Tjian, 1983; Parker and Topol, 1984; Heberlein et al., 1985; Sawadogo and Roeder, 1985; Carthew et al., 1985). However, little is known about the structure of these proteins or their role in the transcription process in vivo.

Yeast activator proteins were initially identified by mutations that prevented the coordinate induction of specific classes of genes. Subsequently, proteins encoded by the *GAL4* and *GCN4* genes were shown to bind specifically to upstream regulatory sites of the relevant coregulated genes (Giniger et al., 1985; Bram and Kornberg, 1985; Hope and Struhl, 1985). Analysis of GAL4 indicates that DNA binding and transcriptional activation are separable functions. First, the N-terminal 73 amino acids specifically binds DNA but does not activate transcription (Keegan et al., 1986). Second, a LexA-GAL4 hybrid protein lacking the GAL4 DNA binding region can activate transcription of promoters containing the *lexA* operator as an upstream element (Brent and Ptashne, 1985). The hybrid protein binds DNA via the *lexA* repressor–operator interaction and stimulates transcription via the GAL4 activation function. These experiments did not localize the GAL4 activation region.

The yeast *GCN4* gene product is necessary for the coordinate induction of 30–50 genes encoding amino acid biosynthetic enzymes (reviewed by Jones and Fink, 1982). GCN4 protein, synthesized from the cloned gene by transcription and translation in vitro, binds specifically to the promoter regions of *HIS3* and three other coregulated genes, but not to analogous regions of genes not regulated by general control (Hope and Struhl, 1985). Using deletions of the *HIS3* promoter region and direct DNAasel footprinting, we demonstrated that GCN4 binds specifically to a 12 bp region that is critical for transcriptional induction in vivo (Hope and Struhl, 1985). Analysis of numerous point mutations as well as comparisons of coregulated genes indicates that GCN4 recognizes sequences whose consensus is rTGACTCatt (capital letters indicate bases that are absolutely specified) (Hill et al., 1986). The DNA sequence requirements for protein recognition and for transcriptional activation could not be distinguished. Without exception, mutations that abolish *his3* induction also fail to bind GCN4, and derivatives that confer *his3* inducibility are able to bind GCN4. These observations strongly suggest that GCN4 binding to the *HIS3* regulatory site directly mediates induction in vivo (Hill et al., 1986).

Here we describe the functional dissection of GCN4. Specifically, we generated serial deletions of the *GCN4* gene and determined the properties of the encoded proteins in three different ways. First, mutant proteins synthesized by transcription and translation in vitro were tested

Table 1. N-Terminally Deleted GCN4 Proteins

GCN4 Derivative	N-Terminal Sequence	GCN4 Amino Acids Present	Specific DNA Binding Activity	*GCN4* Activity In Vivo
GCN4-wt	–	1 to 281	+	+ + +
gcn4-C267	MSDPL-	15 to 281	+	+ + +
gcn4-C210	MSDP-	72 to 281	+	+ + +
gcn4-C197	MSDPRRIP-	85 to 281	+	+ + +
gcn4-C186	MSDPQ-	96 to 281	+	+ + +
gcn4-C163	MSDPQ-	119 to 281	+	+
gcn4-C141	MSDPRRIP-	141 to 281	+	–
gcn4-C131	MSDPR-	151 to 281	+	–
gcn4-C114	MSDPRRIP-	168 to 281	+	–
gcn4-C83	MSDPRRIP-	199 to 281	+	–
gcn4-C60	MSDP-	222 to 281	+	–
gcn4-C37	MSDPP-	245 to 281	–	–

The amino acid sequence at the N-terminus of the GCN4 derivatives is given in the one letter code. The ability (+) or inability (–) to bind specifically to the upstream region of *HIS3* in vitro is indicated. In the assay for total *GCN4* activity in yeast, each derivative conferred either full resistance (+ + +), partial resistance (+), or no resistance (–) to aminotriazole.

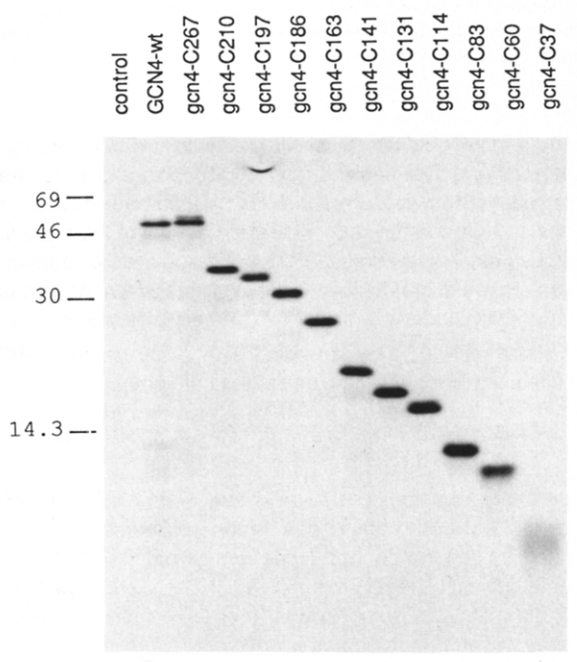

Figure 1. N-Terminally Deleted GCN4 Derivatives

GCN4 protein derivatives generated by transcription and translation in vitro (20,000 cpm of acid-precipitable counts) were examined by SDS-PAGE on a 15% gel and autoradiography. The GCN4 derivatives are identified above each lane, and the control represents the translation products generated from a transcript of the pSP64 vector. Molecular weight markers are in kilodaltons. As shown previously, the full-length protein (predicted molecular weight of 31,000) migrates very anomolously (apparent molecular weight approximately 45,000). The migration of the deleted proteins are approximately but not strictly related to their proportion of the full-length protein.

for their ability to bind specifically to *HIS3* DNA. Second, mutant proteins were expressed in yeast cells and examined for their ability to induce *HIS3* expression in response to conditions of amino acid starvation. Third, by analogy to the experiments of Brent and Ptashne (1985), deleted derivatives of a LexA-GCN4 hybrid protein were assayed for their ability to induce transcription of a promoter containing the *lexA* operator as an upstream element. From these experiments, we were able to separate the DNA binding and transcriptional activation functions of GCN4, and to localize these activities to small regions of the protein.

Results

The DNA Binding Domain Is Localized to the 60 C-Terminal Amino Acids

In previous work, we showed that a synthetic protein lacking the 42 C-terminal amino acids was unable to bind DNA, thus indicating that the C-terminus of GCN4 is necessary for binding (Hope and Struhl, 1985). To determine the region that is sufficient for binding, a series of N-terminally deleted proteins were synthesized by transcription in vitro using SP6 RNA polymerase and translation in vitro using a wheat germ extract. The templates used in the transcription reaction were derivatives of pSP64-*GCN4* (Hope and Struhl, 1985) in which DNA was deleted from the second codon of the *GCN4* gene to various positions throughout the coding region, ensuring that the correct reading frame was maintained. The series of proteins contained from 37 to 267 C-terminal amino acids out of the 281 amino acids in the full-size GCN4 protein (Table 1). The protein preparations used for DNA binding experiments were analyzed by electrophoresis in a denaturing polyacrylamide gel (Figure 1).

The series of N-terminally deleted, in vitro synthesized proteins were assayed for ability to bind specifically to the upstream region of *HIS3*, a gene that is subject to general control and whose upstream region is recognized by GCN4 protein. DNA–protein complex formation was detected as a shift in the mobility of the ^{35}S-methionine labeled protein upon native PAGE in the presence of particular DNA fragments. Proteins with specific DNA binding activity produced an intense band in the presence of *HIS3* DNA fragments, but did not with control vector DNA fragments or in the absence of DNA (Figure 2). All of the

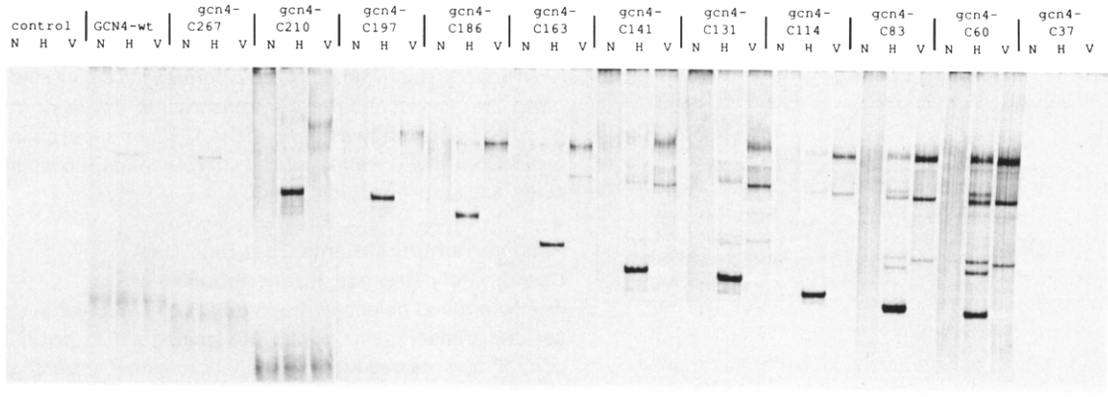

Figure 2. Assay for Specific DNA Binding Activity

^{35}S-methionine labeled wild-type (GCN4), N-terminally deleted (gcn4-Cxxx), or control translation products (at 0.2 nM) were incubated with TaqI-digested pUC8-*HIS3* (H) or pUC9 (V) DNAs (at 4 nM), or with no DNA (N) and analyzed by native PAGE on a 5% gel and autoradiography. From titration experiments using a constant amount of protein and varying amounts of DNA (data not shown), and from the dissociation constant (approximately 10^{-10} M as determined by DNAase I footprinting) (see Figure 6 of Hope and Struhl, 1985), these conditions are only sufficient to drive approximately 30%–70% of the protein into specific complex formation (the strong band in the H lanes) even though the DNA was in excess. Thus a small decrease in binding affinity (e.g., a factor of 10) would cause a noticeable decrease in the amount of protein complexing with the DNA. The faint bands observed in the H and V lanes are due to nonspecific binding interactions with the other DNA fragments in the reaction mixture. The analyses with the wild-type GCN4 and the gcn4-C267 products were performed with 2- 3-fold less protein.

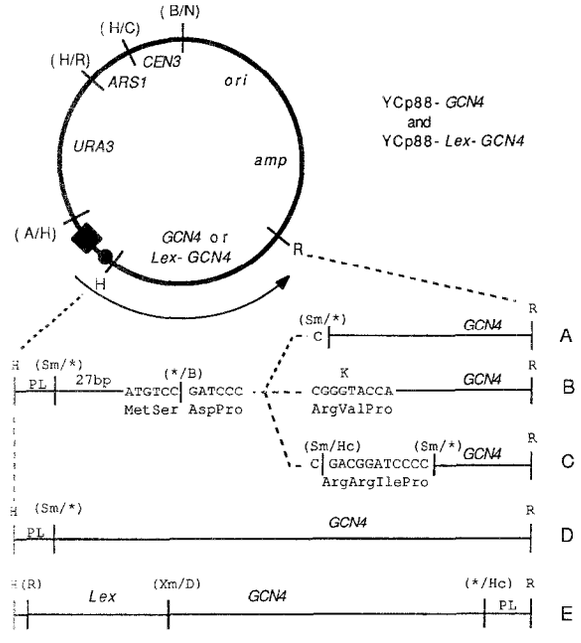

Figure 3. Plasmids for Expression of GCN4 and LexA-GCN4 Derivatives in Yeast

The vector YCp88 consists of the following: (1) an EcoRI–NdeI fragment of pBR322 (Sutcliffe, 1978) containing the ampicillin resistance gene and the origin of replication (from pUC9); (2) a 1.1 kb BamHI–ClaI fragment containing *CEN3* (Fitzgerald-Hayes et al., 1982), an 850 bp HindIII–EcoRI fragment containing *ARS1* (Tschumper and Carbon, 1980), and a 1.1 kb HindIII fragment containing *URA3* (Bach et al., 1979) all for maintenance in yeast; (3) a promoter fragment consisting of a 260 bp fragment containing the upstream and TATA promoter elements of *DED1* (from positions 646 to 903 as defined by Struhl, 1985) fused to a 66 bp FnuDII–HindIII fragment from pSP64 (Melton et al., 1984) containing the bacteriophage SP6 promoter from which transcription of GCN4 and derivatives can be initiated in vitro and in vivo (arrow); (4) a HindIII–EcoRI polylinker from M13mp18 (Messing, 1983) into which fragments encoding GCN4 and lexA-GCN4 and derivatives can be cloned. The fragment encoding GCN4 is derived from pSP64-*GCN4* (Hope and Struhl, 1985) except that the region following the coding sequences between the EcoRI and PvuII sites has been deleted with regeneration of the EcoRI site; this molecule pSP64-Sc4342 was constructed by Joan Sellers. The fragments encoding N-terminally deleted proteins are depicted on lines A (C267, C210, C186, C163, C60, and C37), B (C131), C (C197, C141, C83), and D (C268, C243, C175, and C32). Some DNA sequences and the encoded amino acids are shown. The fragment encoding LexA-GCN4 is derived from pRB1280 (constructed by and obtained from Roger Brent) except that an EcoRI site adjacent to the HindIII site was destroyed. The fragments encoding C-terminally deleted Lex-GCN4 derivatives are depicted on line E. YCp86 is the same as YCp88 except that it does not contain the *DED1* or SP6 promoter fragments. Restriction endonuclease cleavage sites are indicated as A, Asp718; B, BamHI; C, ClaI; D, DraI; H, HindIII; Hc, HincII; K, KpnI; N, NdeI; R, EcoRI; Sm, SmaI; Xm, XmnI. The asterisk indicates that the end of the fragment was generated using the exonuclease Bal31, the diagonal line indicates that the fragment ends were treated with the large fragment of E. coli DNA polymerase I and the blunt ends ligated together, the parentheses indicate restriction sites that have been destroyed, and PL indicates the polylinker sequences. The diagram is not drawn to scale.

N-terminally deleted proteins bound specifically and apparently with the same affinity to *HIS3* DNA except for the smallest derivative, gcn4-C37. Interestingly, the mobility of the complex is inversely related to the size of the protein. The specific binding observed for gcn4-C60 demonstrates that the DNA binding domain of GCN4 protein is contained entirely within the 60 C-terminal amino acids and can fold into a functional unit in the absence of the rest of the protein.

DNA Binding Is Not Sufficient for Transcriptional Activation In Vivo

The same N-terminally deleted GCN4 proteins were expressed in yeast cells and tested for their ability to induce *HIS3* transcription. The abbreviated *GCN4* coding regions were recloned into YCp88 (Figure 3) such that they would be transcribed from the *DED1* promoter. The native *DED1*

Table 2. N-Terminally Deleted GCN4 Proteins—Series 2

GCN4 Derivative	GCN Amino Acids Present	*GCN4* Activity In Vivo
GCN4-wt	1 to 281	+ + +
gcn4-C268	14 to 281	+ + +
gcn4-C243	39 to 281	+ + +
gcn4-C175	107 to 281	+ + +
gcn4-C32	250 to 281	–

In the assay for total *GCN4* activity in yeast each derivative conferred either full resistance (+ + +) or no resistance (–) to aminotriazole.

gene is transcribed constitutively to generate approximately 5–10 mRNA molecules per cell (Struhl, 1985), comparable to the level of the wild-type *GCN4* transcript. Upon transformation into KY803, a yeast strain carrying a genomic *gcn4* deletion, GCN4 function was assayed by challenging with aminotriazole, a competitive inhibitor of the *HIS3* gene product. In the absence of exogenous histidine, aminotriazole will completely inhibit cell growth unless a functional *GCN4* gene product induces transcription of *HIS3* and other coregulated genes.

Strains containing gcn4-C186 and larger GCN4 derivatives were completely resistant to aminotriazole and grew as well as the control strain containing the full-length GCN4 protein (Table 1). Transformants harboring gcn4-C163 were partially resistant to aminotriazole in that they grew much more slowly. Plasmids encoding smaller GCN4 derivatives conferred no resistance to aminotriazole. These results suggest that the 186 C-terminal amino acids of GCN4 can both bind DNA specifically and activate transcription of genes regulated under general control. As the smaller gcn4 derivatives appear to bind in vivo (see next section), their failure to confer full *GCN4* activity is likely to be due to the removal of part (or all) of a distinct region necessary for transcriptional activation. This transcriptional activation region must be located within the C-terminal 186 amino acids of the protein.

In the deletions described above, translation initiates at the normal *GCN4* initiation codon. To determine whether the four internal methionine codons could function as translational initiation sites, a second series of deletion mutants was generated and cloned into YCp88 and analyzed as described above (Figure 3). Specifically, the entire *GCN4* coding region including the normal initiator codon was deleted up to just before each of the four internal methionine codons. Upon transformation into KY803, full *GCN4* activity was observed for gcn4-C268, gcn4-C242, and gcn4-C175 (Table 2). The extensively deleted gcn4-C32, which presumably lacks the transcription activation region and part of the DNA binding domain, was inactive as expected. Thus, there are four methionine codons in the wild-type *GCN4* mRNA at which translation can be initiated efficiently to generate a fully functional protein. This result explains why certain *gcn4* deletion mutations produce functional proteins even though they remove sequences encoding the entire promoter region and N-terminal portions of the protein (Hinnebusch, 1984). In addition, these observations may be relevant to the mechanism by which GCN4 activity is translationally regulated (Thireos et al., 1984; Hinnebusch, 1984).

The observation that gcn4-C175 confers full activity indicates that the location of the transcription activation region has been narrowed down to the 175 C-terminal amino acids. Conversely, it appears that the N-terminal 106 amino acids are functionally dispensible.

Positive Control Mutants That Bind DNA Coordinately Repress Transcription

The N-terminal deletions that bind DNA in vitro yet fail to activate transcription in vivo are analogous to positive control mutants of several bacterial proteins (Guarente et al., 1982; Hochschild et al., 1983) and the yeast GAL4 protein (Keegan et al., 1986). However, because these derivatives might fail to activate transcription simply because the encoded proteins are unstable, it is important to show that they are able to bind DNA in vivo. From the results of Brent and Ptashne (1984) and Keegan et al. (1986), it is expected that a protein that binds to DNA sequences between an upstream activation sequence and a TATA element will repress transcription. In this regard, the coregulated amino acid biosynthetic genes are ideal for testing whether particular *gcn4* deletion mutations will act as repressors because they all are capable of constitutive transcription in the absence of GCN4, yet contain a GCN4 binding site. Thus, positive control mutants of GCN4 should act as coordinate repressors of amino acid biosynthetic genes.

This expectation is strongly supported by the variation in growth properties of the yeast transformants. When grown on medium lacking amino acids, colonies of transformants containing YCp86, a control plasmid with no *GCN4* coding region, appeared in three days. Plasmids encoding a fully functional GCN4 protein (i.e., gcn4-C175 and larger derivatives) produced colonies in only two days, presumably because GCN4 improves growth by stimulating expression of the amino acid biosynthetic genes. More interestingly, plasmids encoding nonfunctional gcn4 proteins (gcn4-C141 and smaller derivatives) conferred a very low growth rate, with small colonies only detected after four days. Presumably these protein derivatives contain the DNA binding but not the transcription activation function, and so they actually inhibit expression of genes encoding amino acid biosynthetic enzymes.

This repression effect is specific because the growth inhibition is eliminated by the addition of amino acids to the medium. Moreover, gcn4-C37, a derivative that showed no specific DNA binding activity in vitro, also failed to inhibit growth in vivo. Transformants expressing gcn4-C83 also gave colonies in three days, even though this derivative did demonstrate specific DNA binding activity in vitro. This one apparent exception may be a consequence of instability of gcn4-C83 in yeast. These observations indicate that gcn4-C141 and smaller derivatives fail to activate transcription in vivo not because of protein instability but because the region of GCN4 protein responsible for activation has either been structurally altered or deleted. In addition, the growth inhibition that is a result of these positive control mutant proteins is sufficiently severe to allow

Table 3. C-Terminally Deleted Lex-GCN4 Proteins

Lex-GCN4 Derivatives	Amino Acids Present			β-Galactosidase Activity	
	Lex	GCN4	C-Terminus	Liquid	Plate
Lex-GCN4-wt	1 to 87	12 to 281	—	420	+ +
Lex-gcn4-N281	1 to 87	12 to 281	—	440	+ +
Lex-gcn4-N270	1 to 87	12 to 270	L	80	+
Lex-gcn4-N249	1 to 87	12 to 249	L	40	+
Lex-gcn4-N222	1 to 87	12 to 222	2	70	+
Lex-gcn4-N194	1 to 87	12 to 194	3	120	+
Lex-gcn4-N187	1 to 87	12 to 187	3	60	+
Lex-gcn4-N181	1 to 87	12 to 181	3	80	+
Lex-gcn4-N150	1 to 87	12 to 150	2	70	+
Lex-gcn4-N127	1 to 87	12 to 127	2	70	+
Lex-gcn4-N125	1 to 87	12 to 125	K-3	220	+
Lex-gcn4-N120	1 to 87	12 to 120	2	30	+/−
Lex-gcn4-N77	1 to 87	12 to 77	3	<1	−
Lex-gcn4-N71	1 to 87	12 to 71	3	<1	−
Lex-gcn4-N43	1 to 87	12 to 43	2	<1	−
Lex-N84	1 to 84	—	2	<1	−
Lex-N82	1 to 82	—	3	<1	−
GCN4-wt	—	1 to 281	—	<1	−

The C-terminal amino acid residues are indicated by the one letter code with 2 = DSRGSPGTELEFLKTKGPRDTPIFIG and 3 = TLEDPRVPSSNS. The plasmid expressing Lex-gcn4-N281 differs from that expressing Lex-GCN4-wt in lacking all yeast DNA after the GCN4 stop codon. The units of the β-galactosidase activities determined for liquid cultures have been defined as 1000 × change in OD_{420} due to hydrolysis of o-nitrophenol-β-D-galactoside/(assay time in minutes × volume of culture assayed × OD_{600} of culture at time of harvesting) (Guarente, 1983). For each derivative β-galactosidase activities were determined in triplicate on two separate occasions, with each assay performed on an independent transformant, and results were consistent. For the plate assay of β-galactosidase activity, hydrolysis of X-gal generated colonies that were strongly blue (+ +), blue (+), weakly blue (+/−), or white (−).

for the selection of derivatives that fail to bind DNA specifically.

Localization of the GCN4 Activation Function

Deletion analysis from the C-terminus of GCN4 was used to map the C-terminal boundaries of the DNA-binding domain and the transcriptional activation region. However, such sequential deletions would disrupt the DNA binding function first. Therefore, to define the C-terminal boundary of the activation domain in this way, another specific DNA-binding domain had to be provided at the N-terminus. Such a construction had already been prepared and proven functional by Brent and Ptashne (1985). They fused the DNA binding domain of the prokaryotic LexA repressor protein close to the N-terminus of GCN4. This fusion protein activated transcription in yeast of a *lacZ* gene upstream of which had been placed a lexA DNA binding site.

A DNA fragment containing the fused *LexA-GCN4* gene, kindly provided by Roger Brent, was deleted to various positions in the coding region from the C-terminus (Table 3) and cloned into YCp88 (Figure 3). Translation of these truncated coding regions terminated in adjacent DNA and so the C-termini of the proteins encoded varied depending upon the reading frame in which the deletion endpoints fell. However, as shown in the results below, this C-terminal sequence variation appeared to have no detrimental effect on the experimental observations.

First, we examined these C-terminally deleted Lex-GCN4 proteins for their ability to confer normal *GCN4* function. The constructions for expressing Lex-GCN4-wt, Lex-gcn4-N281, Lex-gcn4-N270, and Lex-gcn4-N249 were transformed into KY803, and the transformants were assayed for resistance to aminotriazole. The two plasmids expressing the full-size fusion protein conferred aminotriazole resistance while the other two plasmids did not. Therefore, the DNA binding domain of GCN4 extends to within 11 amino acids of the C-terminus. The extent of the *GCN4* DNA binding domain could not be refined further from these results.

All of the C-terminally deleted Lex-GCN4 fusion proteins were assayed for ability to activate transcription through the LexA DNA binding domain. The target for this activation, a *lacZ* gene with the lexA DNA binding site placed upstream, was used originally by Brent and Ptashne (1985). A DNA fragment from a plasmid kindly provided by Roger Brent was used to construct YEp21-Sc3423, which also contains 2μ plasmid sequences for replication in yeast and *LEU2* as a selectable marker (Figure 4). KY803 was transformed sequentially with YEp21-Sc3423 and YCp88 plasmids containing *Lex-GCN4* gene derivatives encoding C-terminally deleted fusion proteins. The transformants were assayed for β-galactosidase activity both qualitatively on X-gal indicator plates and quantitatively by spectrometric assays of liquid cultures. Results from the two assays were completely consistent (Table 3).

As previously reported (Brent and Ptashne, 1985), the Lex-GCN4 fusion protein activated transcription of *LacZ*. As the *Lex-GCN4* coding region was deleted, two stepwise drops in activation ability could be distinguished. The first drop, to approximately 20% of maximum activity, occurs as soon as any of the *Lex-GCN4* coding region is deleted. There are several possible explanations for this (see Discussion). However, considerably more of *Lex-GCN4* may be deleted without further affects on *LacZ* activity. In particular, a high level of activity is observed with Lex-gcn4-N125. Lex-gcn4-N120 produces detectable activity, but less than Lex-gcn4-N125 and so may be partially defective in its activation function. More extensively deleted proteins are completely unable to activate *LacZ* expression. The high level of activity observed with Lex-gcn4-N125 suggests that a fully functional *GCN4* transcription activation region is located within the 125 N-terminal amino acids of the native GCN4 protein.

Elastase Preferentially Cleaves GCN4 between the Binding and Activation Regions

Proteolytic cleavage in a particular region of a protein commonly identifies unstructured spacer segments which connect highly structured and independently functional domains. Examples of proteins examined in this way include immunoglobulins (Porter, 1959), E. coli DNA polymerase I (Jacobsen et al., 1974), and λcI (Pabo et al., 1979). In contrast to these studies, which require pure pro-

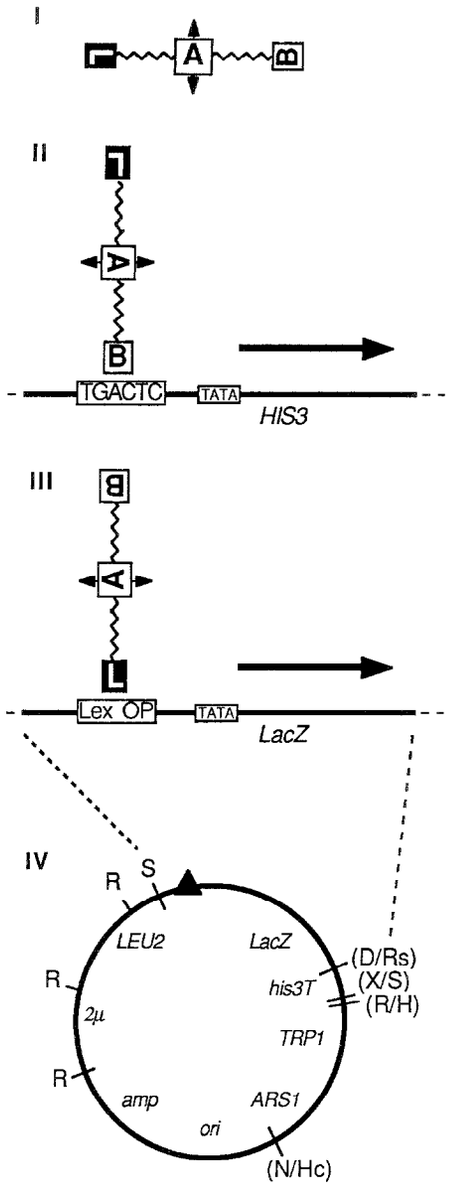

Figure 4. Transcriptional Activation by Lex-GCN4

I. Lex-GCN4 is depicted as a LexA DNA binding domain (black box marked L), a GCN4 transcriptional activation region (white box marked A), and a GCN4 DNA binding domain (white box marked B), each connected by other GCN4 segments (zig-zag lines). The arrowheads on the activation domain indicate bidirectional activation capability.

II. Lex-GCN4 can activate *HIS3* transcription (arrow) upon binding to the critical regulatory sequence whose core is TGACTC through the GCN4 binding domain.

III. Lex-GCN4 can activate *LacZ* transcription (arrow) upon binding to the *lexA* operator (LexOP) through the LexA DNA binding domain.

IV. YEp21-Sc3423, the plasmid carrying the *LacZ* target gene, consists of the following segments: (1) a SalI–EcoRI fragment from YEp21 (Botstein et al., 1979) containing the *LEU2* gene and 2μ plasmid sequences for maintenance in yeast; (2) the EcoRI–NdeI fragment containing the ampicillin resistance gene and E. coli origin of replication as described for YCp88 (see Figure 3); (3) a HincII–EcoRI fragment containing *ARS1* and *TRP1* (Tschumper and Carbon, 1979) originally for maintenance in yeast but found to be insufficient in the presence of YCp88; (4) a SalI–EcoRI polylinker fragment from pUC18 (Messing, 1983); (5) a RsaI–XhoI fragment containing the *HIS3* transcriptional terminator (Struhl, 1985); (6) a SalI–DraI fragment from pRB1155 (Brent and Ptashne, 1985) containing the *lexA* operator-*CYC1* TATA-*lacZ* fusion gene prepared and obtained from Brent and Ptashne (except that part of the *lacZ* portion came from Malcolm Casadaban so that the internal SacI site was no longer present). The closed triangle indicates the *lexA* operator. Restriction endonuclease cleavage sites are indicated as in Figure 3 except for Rs, RsaI; S, SalI, X, XhoI. The figure is not drawn to scale.

teins, the radioactively pure proteins synthesized in vitro can be examined without further purification. Moreover, by cleaving a series of terminally deleted proteins, the sites of cleavage can be easily determined.

Elastase cleaves GCN4, which has an apparent mobility of 46 kd on a 15% SDS-polyacrylamide gel, into two major fragments of 30 kd and 16 kd (Figure 5). The smaller fragment is derived from the C-terminus because it is also generated from the N-terminally deleted proteins gcn4-C197 (36 kd) and gcn4-C131 (18.5 kd). The N-terminal fragment produced by elastase cleavage is 20 kd for gcn4-C197 and is too small to be detected for gcn4-C131.

The site of preferential cleavage by elastase maps at amino acid 181 ± 10. Thus, elastase physically separates the regions of GCN4 defined functionally as important for specific DNA binding and for transcriptional activation. Further studies using other proteases may reveal more structural details about GCN4, in particular whether the transcription activation region exists in a structurally ordered domain.

Discussion

Functional Arrangement of GCN4 Protein

GCN4 protein has two separable activities. Genes to be regulated under general control are recognized by GCN4 binding specifically to their upstream regions (Hope and Struhl, 1985). GCN4 then stimulates transcription of the genes it has recognized. Separate regions responsible for these two functions have now been defined and located along the linear amino acid sequence of GCN4 protein (Figure 6).

The 221 N-terminal amino acids of GCN4 can be deleted without any detectable effect on the specific DNA-binding activity as assayed in vitro. This shows that the remaining 60 C-terminal amino acids can fold, independent of the rest of the protein, into a discrete, fully-functional, DNA-binding domain. Previous DNA-binding studies using truncated GCN4 protein derivatives generated in vitro from a *GCN4* coding region DNA template cleaved with restriction enzymes had already suggested that the C-terminus of GCN4 was involved in the DNA-binding activity (Hope and Struhl, 1985).

The two key observations that locate the activation region are that gcn4-C175 has full *GCN4* activity and Lex-gcn4-N125 can activate transcription through the LexA DNA-binding domain. The obvious interpretation is that the 19 amino acids common to these two recombinant proteins (GCN4 amino acids 107 to 125) contain the region responsible for transcriptional activation. However, other models are also possible (see below).

Assignments of the two functions of GCN4 to separate and small regions of the protein, leaves two large regions (amino acids 1 to 106 and 126 to 220) without apparent function. The amino acids between the activation and

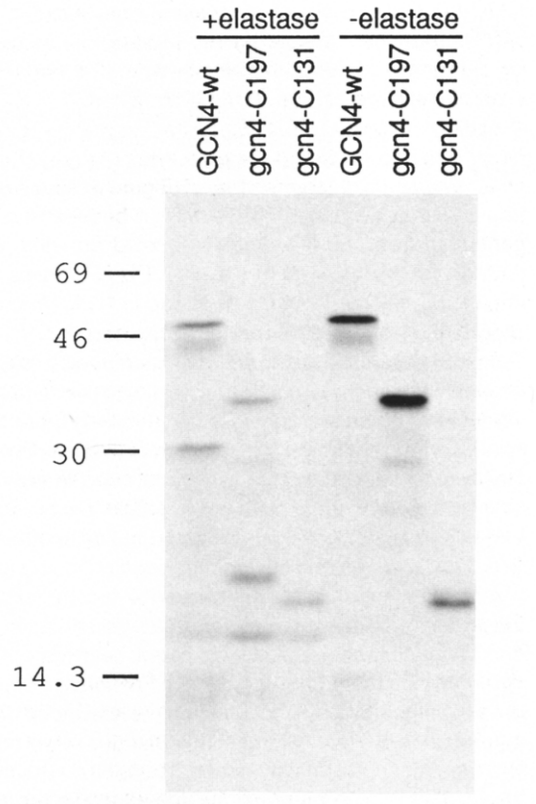

Figure 5. Elastase Digestion of GCN4

GCN4-wt, gcn4-C197, and gcn4-C131 proteins synthesized in vitro were either treated (+) or not treated (−) with elastase and examined by SDS-PAGE on a 15% gel. Molecular weight markers are in kilodaltons.

DNA-binding regions could have an important spacer function. The spacer may provide a required distance between the two functional regions and/or allow for the flexibility, which has been observed, in both position and orientation of the DNA sequence element recognized by GCN4 protein in relation to other components of yeast promoters (Struhl, 1982; Hinnebusch et al., 1985). In this regard, note that the GCN4 activation domain is functional with a DNA-binding domain located either at its natural C-terminal position or at the inverted N-terminal position (the LexA hybrid molecules). Alternatively, either of these two regions may have a role in subtle, as yet undetected, posttranslational regulation of GCN4 activity.

DNA-Binding Domain

Studies of specific DNA-binding proteins have revealed two distinct structural motifs, the "helix-turn-helix" of prokaryotic repressors and activators (reviewed by Pabo and Sauer, 1984), and the "zinc-finger" of the eukaryotic transcription factor TFIIIA (Miller et al., 1985) and other potential activators (Hartshorne et al., 1986; Rosenberg et al., 1986). When the Chou-Fasman (1978) rules for predicting protein secondary structure were applied to GCN4, the strongest predictions were obtained for the DNA-binding domain as defined here. Three α-helices are predicted: amino acids 228 to 233, 244 to 261, and 266 to 278, with a β-turn at 262 to 265. However, sequences in this region fail to show any of the homology normally observed in the helix-turn-helix motifs of other DNA-binding proteins, and the helices are longer. The zinc-finger motif is clearly not applicable to GCN4 because there are no cysteine residues in the protein. X-ray crystallographic and mutational studies will be necessary to determine the molecular basis of this particular protein–DNA interaction.

In many cases, proteins that bind DNA are dimeric and recognize a palindromic DNA sequence (Pabo and Sauer, 1984). The sequence recognized by GCN4 appears to be palindromic (Hill et al., 1986) and recent evidence strongly suggests that GCN4 binds as a dimer (Hope and Struhl, unpublished results). In E. coli, LexA dimerizes through a domain that is distinct from the DNA-binding domain, and only the dimeric form binds DNA (Little and Mount, 1984). The observation that Lex-GCN4 can activate transcription in yeast through the LexA DNA-binding domain could mean that the GCN4 part of the fusion protein provides the dimerization function. However, the monomeric LexA DNA-binding domain can bind DNA specifically but with a much reduced affinity (Little and Mount, 1982), and the concentration of Lex-GCN4 protein generated in yeast in these experiments is not known.

The activation of *lacZ* by Lex-GCN4 dropped slightly upon deletion of any of the GCN4 C-terminal coding region. Perhaps the loss of a dimerization function provided by an intact GCN4 DNA-binding domain causes a drop in the affinity of the fusion protein for the LexA DNA-binding site. However, other explanations of this observation are also possible. The presence of another DNA-binding activity on the same polypeptide chain may increase the affinity of the fusion protein for DNA in general and thus the LexA DNA-binding site in particular. Alternatively, loss of the GCN4 C-terminus normally present in yeast may decrease protein stability.

The 60 C-terminal amino acids of GCN4 protein contain 17 basic residues (lysine, arginine, and histidine) which is consistent with the DNA-binding activity demonstrated for this region. However, this strongly basic character extends further into the protein with 14 of the next 45 residues also being basic. This segment of the protein does not appear to contribute to the DNA-binding affinity because the loss of this contribution would have been detected in the in vitro DNA-binding assays presented here. However, this segment of the protein may contribute to the specificity of the interaction because gcn4-C83 and gcn4-C60 may have slightly more nonspecific DNA-binding activity than the larger derivatives.

Transcriptional Activation Region

The results presented here suggest that a 19 amino acid segment of GCN4 protein (amino acids 107 to 125) is critically involved in transcription activation. This indicates either that the 19 amino acid region is sufficient for activation, or that the protein contains redundant activities. For example, there could be two activation domains, both of which could activate transcription, but with both requiring some part of the 19 amino acid segment highlighted here. However, this latter possibility seems unlikely because

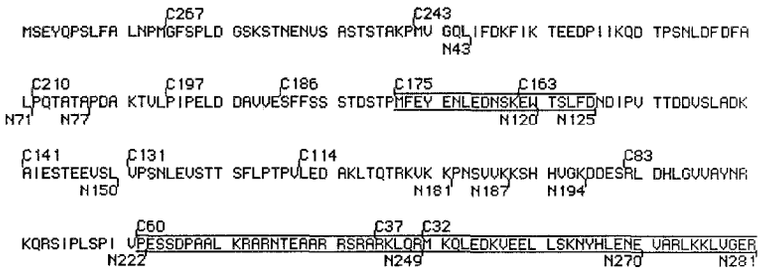

Figure 6. Structure of GCN4 Protein

The GCN4 amino acid sequence derived from the nucleotide sequence of the gene (Thireos et al., 1984; Hinnebusch, 1984) is depicted, and the region responsible for the DNA binding and transcriptional activation functions are over and underlined. The endpoints of the protein derivatives used to define these functionally important regions are indicated.

computer analysis of the GCN4 protein sequence has failed to reveal any evidence of homology suggestive of two structurally similar domains.

The 19 amino acid segment is at the center of a very acidic region of GCN4. The 60 amino acids between 88 and 147 contain 18 acidic residues and just two basic residues. Perhaps all that is required for transcription activation is an acidic region of the protein. The acidic region of GCN4 may be more than sufficient for transcription activation and so large portions of this region may be deleted from either direction, without detectable effect, as observed here. In support of this model, PHO4, a good candidate for an activator (Legrain et al., 1986) contains an extraordinarily acidic region towards its N-terminus. The GAL4 activator (Laughon and Gesteland, 1984) and the presumptive ADR1 activator (Hartshorne et al., 1986) also contain acidic regions. In addition, positive control mutations of bacterial proteins are associated with a decrease in acidic character, whereas revertants are associated with an increase in acidic character (Hochschild et al., 1983).

The final alternative is that the 19 amino acid segment is all that is required for activation. Although this segment seems unusually short to define a biochemical activity, similarly sized regions are sufficient to act as signal sequences in targeting proteins to particular locations around the cell (vonHeijne, 1983; van Loon et al., 1986). By analogy, the 19 amino acid region of GCN4 might serve as a recognition signal for one or more components of the transcription apparatus.

It has been suggested that activator proteins stimulate transcription by a common molecular mechanism even though individual activators recognize different upstream regulatory sequences (reviewed by Guarente, 1984; Struhl, 1986). In particular, when upstream regulatory elements are interchanged, transcription from the resulting hybrid promoters is initiated from the correct site and depends on the activator protein that binds to the upstream regulatory sequence. Thus, one might expect that the activation functions of yeast transcriptional activator proteins will share sequence homology. However, a computer comparison of the 19 amino acid segment to PHO4, GAL4, and ADR1 failed to reveal any good sequence homology. In this regard, there is no strict sequence requirement for signal sequences or for mitochondrial targeting sequences. In these cases, it has been proposed that functional similarity reflects structural similarities that are defined by parameters that are more complex than the primary sequence (Briggs et al., 1985; Roise et al., 1986; vonHeijne, 1986).

Mechanism of Transcriptional Activation

Firm conclusions about the mechanism by which GCN4 activates transcription cannot yet be drawn because of the possibility (which we consider very unlikely) that there could be two redundant transcription activation domains. However, either of the other two models for the activation domain are very suggestive of the GCN4 protein activation region acting as a ligand to be bound by another protein. The GCN4 protein polypeptide segment being recognized would not have a strict sequence requirement and would either be a protein region of acidic character or a short peptide region with, as yet, poorly defined composition. Perhaps the proportion of acidic residues or the degree of homology to the optimal peptide sequence has an influence on the level of transcriptional activation a positive regulatory protein can induce. Theories on the mechanism of activation must still include allosteric activation of a component of the transcription apparatus such as RNA polymerase II, interaction with histones to make the local chromatin structure more permissive to transcription, and provision of a target at the chromatin surface by which the transcription complex can more easily locate a gene to be transcribed. However the apparent small size required for transcriptional activation mediated by GCN4 suggests that activation does not involve a catalytic function such as a topoisomerase, nuclease, or methylase. Thus, these results support the idea developed from prokaryotic activators that transcriptional activation does not involve changes in the conformation of DNA, but rather involves specific protein–protein contacts.

Experimental Techniques

DNA Techniques

The DNA constructions used in this study are described fully in the legends to Figures 3 and 4 and were prepared as described previously (Struhl, 1983). DNA fragments deleted using Bal31 exonuclease were initially cloned into M13 derivatives (Messing, 1983) and sequenced by the dideoxy chain termination technique (Sanger et al., 1980).

Synthesis and Analysis of Proteins

N-terminally deleted GCN4 proteins were synthesized in vitro from CsCl purified pSP64 derivatives and analyzed on SDS-PAGE using techniques previously described (Hope and Struhl, 1985). The DNA binding assay for the in vitro synthesized proteins was also as previously described (Hope and Struhl, 1985) except that 20,000 TCA precipitable cpm of each protein was used per incubation. For the elastase digestion experiments GCN4 protein and derivatives, as synthesized in a wheat germ extract (1 μl, 20,000 TCA precipitable cpm), were treated with 0.001 U of elastase (Sigma) for 25 min at 37°C in 10 μl of 10 mM Tris (pH 8.8). The digestion products were diluted with 20 μl SDS sample buffer containing bovine serum albumin (0.5 μg/μl), incubated at 100°C for 5 min and examined by SDS-PAGE.

Analysis of GCN4 Proteins In Vivo

The yeast strain used throughout, KY803 (*trp1*-Δ1 *ura3*-52 *leu2*-P1 *gcn4*-Δ1), was constructed by Joan Sellers. The *gcn4* deletion extends from the *Bst*EII site preceding to the PvuII site following the GCN4 coding region (Penn, M. D. [1985]. Ph.D. thesis, Harvard University, Cambridge, MA). DNAs were introduced into yeast as described previously (Struhl, 1985) by selecting for strains able to grow in the absence of uracil or leucine. The assay for total GCN4 activity was carried out on minimal plates supplemented with 10 mM 3-amino-1,2,4-triazole, and many amino acids, but not histidine (2 mg/l tryptophan, methionine, arginine, and adenine, 3 mg/l tyrosine, lysine leucine, and isoleucine, 5 mg/l valine, phenylalanine, glutamate, and aspartate, 15 mg/l serine, and 20 mg/l threonine). The differential growth of yeast transformants containing N-terminally deleted GCN4 derivatives was observed on minimal plates supplemented only with adenine, histidine, tryptophan (all at 2 mg/l), leucine and lysine (both at 3 mg/l). Both plate and liquid culture assays of β-galactosidase activities were carried out as described previously (Guarante, 1983).

Acknowledgments

We thank Jennifer Macke for her expert technical assistance throughout many stages of the work, Roger Brent for his generosity with critical DNAs and advice that made it possible to perform the LexA-GCN4 experiments, Joan Sellers for the construction of yeast strain KY803, and David Botstein for YEp21 vector DNA. This work was supported by grants to K. S. from the National Institutes of Health (GM30186) and the Chicago Community Trust (Searle Scholars Program).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 28, 1986; revised July 3, 1986.

References

Bach, M. L., Lacroute, F., and Botstein, D. (1979). Evidence for transcriptional regulation of orotidine 5-phosphate decarboxylase in yeast by hybridization of mRNA to the yeast structural gene cloned in *Escherichia coli*. Proc. Natl. Acad. Sci. USA *76*, 386–390.

Botstein, D., Falco, S. C., Stewart, S. E., Brennan, M., Scherer, S., Stinchcomb, D. T., Struhl, K., and Davis, R. W. (1979). Sterile host yeasts (SHY): a eukaryotic system of biological containment for recombinant DNA experiments. Gene *8*, 17–24.

Bram, R., and Kornberg, R. (1985). Specific protein binding to far upstream activating sequences in polymerase II promoters. Proc. Natl. Acad. Sci. USA *82*, 43–47.

Brent, R., and Ptashne, M. (1984). A bacterial repressor protein or a yeast transcriptional terminator can block upstream activation of a yeast gene. Nature *312*, 612–615.

Brent, R., and Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell *43*, 729–736.

Briggs, M. S., Gierasch, L. M., Zlotnick, A., Lear, J. D., and DeGrado, W. F. (1985). *In vivo* function and membrane binding properties are correlated for *Escherichia coli* LamB signal peptides. Science *228*, 1096–1099.

Carthew, R. W., Chodosh, L. A., and Sharp, P. A. (1985). An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell *43*, 439–448.

Chou, P. Y., and Fasman, G. D. (1978). Empirical predictions of protein conformation. Ann. Rev. Biochem. *47*, 251–276.

Dynan, W. S., and Tjian, R. (1983). Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell *32*, 669–680.

Fitzgerald-Hayes, M., Clarke, L., and Carbon, J. (1982). Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. Cell *29*, 235–244.

Giniger, E., Varnum, S. M., and Ptashne, M. (1985). Specific DNA binding of GAL4, a positive regulatory protein of yeast. Cell *40*, 767–774.

Guarente, L. (1983). Yeast promoters and *lacZ* fusions designed to study expression of cloned genes in yeast. Meth. Enzymol. *101*, 181–191.

Guarente, L. (1984). Yeast promoters: positive and negative elements. Cell *36*, 799–800.

Guarente, L., Nye, J. S., Hochschild, A., and Ptashne, M. (1982). A mutant lambda repressor with a specific defect in its positive control function. Proc. Natl. Acad. Sci. USA *79*, 2236–2240.

Hartshorne, T. A., Blumberg, H., and Young, E. T. (1986). Sequence homology of the yeast regulatory protein *ADR1* with *Xenopus* transcription factor IIIA. Nature *320*, 283–287.

Heberlein, U., England, B., and Tjian, R. (1985). Characterization of Drosophila transcription factors that activate the tandem promoters of the alcohol dehydrogenase gene. Cell *41*, 965–977.

Hill, D. E., Hope, I. A., and Struhl, K. (1986). The yeast *HIS3* regulatory site: indistinguishable nucleotide requirements for transcriptional induction and GCN4 binding. Science, in press.

Hinnebusch, A. G. (1984). Evidence for translational regulation of the activator of general amino acid control in yeast. Proc. Natl. Acad. Sci. USA *81*, 6442–6446.

Hinnebusch, A., Lucchini, G., and Fink, G. (1985). A synthetic *HIS4* regulatory element confers general amino acid control on the cytochrome C gene (*CYC1*) of yeast. Proc. Natl. Acad. Sci. USA *82*, 498–502.

Hochschild, A., Irwin, N., and Ptashne, M. (1983). Repressor structure and the mechanism of positive control. Cell *32*, 319–325.

Hope, I. A., and Struhl, K. (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*, 177–188.

Jacobsen, H., Klenow, H., and Overgaard-Hansen, K. (1974). The N-terminal amino acid sequences of DNA polymerase I from *Escherichia coli* and of the large and the small fragments obtained by a limited proteolysis. Eur. J. Biochem. *45*, 623–627.

Jones, E. W., and Fink, G. R. (1982). Regulation of amino acid biosynthesis and nucleotide biosynthesis in yeast. In The Molecular Biology of the Yeast *Saccharomyces*: Metabolism and Gene Expression, J. N. Strathern, E. W. Jones, and J. R. Broach, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 181–299.

Keegan, L., Gill, G., and Ptashne, M. (1986). Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. Science *231*, 699–704.

Laughon, A., and Gesteland, R. (1984). Primary structure of the *Saccharomyces cerevisiae GAL4* gene. Mol. Cell. Biol. *4*, 260–267.

Legrain, M., De Wilde, M., and Hilger, F. (1986). Isolation, physical characterization and expression analysis of the *Saccharomyces cerevisiae* positive regulatory gene *PHO4*. Nucl. Acids Res. *14*, 3059–3073.

Little, J. W., and Mount, D. W. (1982). The SOS regulatory system of Escherichia coli. Cell *29*, 11–22.

Malan, T. P., Kolb, A., Buc, H., and McClure, W. (1984). Mechanism of CRP-cAMP activation of *lac* operon transcription initiation activation of the P1 promoter. J. Mol. Biol. *180*, 881–909.

Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K., and Green, M. R. (1984). Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucl. Acid Res. *12*, 7035–7056.

Messing, J. (1983). New M13 vectors for cloning. Meth. Enzymol. *101*, 20–79.

Miller, J., McLachlan, A. D., and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J. *4*, 1609–1614.

Pabo, C. O., and Sauer R. T. (1984). Protein–DNA recognition. Ann. Rev. Biochem. *53*, 293–321.

Pabo, C. O., Sauer, R. T., Sturtevant, J. M., and Ptashne, M. (1979). The λ repressor contains two domains. Proc. Natl. Acad. Sci. USA. *76*, 1608–1612.

Parker, C. S., and Topol, J. (1984). A Drosophila RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity. Cell *36*, 357–369.

Porter, R. R. (1959). The hydrolysis of rabbit γ-globulin and antibodies with crystalline papain. Biochem. J. *73*, 119–126.

Roise, D., Horvath, S. J., Tomich, J. M., Richards, J. H., and Schatz, G. (1986). A chemically synthesized pre-sequence of an imported mitochondrial protein can form an amphiphilic helix and perturb natural and artificial phopholipid bilayers. EMBO J. *5*, 1327–1334.

Rosenberg, U. B., Schroder, C., Preiss, A., Kienlin, A., Cote, S., Riede, I., and Jackle, H. (1986). Structural homology of the product of the *Drosophila Kruppel* gene with *Xenopus* transcription factor IIIA. Nature *319*, 336–339.

Sanger, F., Coulson, A. R., Barell, B. G., Smith, A. J., and Roe, B. A. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. *143*, 161–178.

Sawadogo, M., and Roeder, R. G. (1985). Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell *43*, 165–175.

Struhl, K. (1982). Regulatory sites for *his3* expression in yeast. Nature *300*, 284–287.

Struhl, K. (1983). Direct selection for gene replacement events in yeast. Gene *26*, 231–241.

Struhl, K. (1985). Nucleotide sequence and transcriptional mapping of the yeast *pet56-his3-ded1* gene region. Nucl. Acids Res. *13*, 8587–8601.

Struhl, K. (1986). Yeast promoters. In From Gene to Protein: Steps Dictating the Maximal Level of Gene Expression, W. Reznikoff and L. Gold, eds. (Boston: Butterworths), pp. 35–78.

Sutcliffe, J. G. (1978). Sequence of the plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. *43*, 77–90.

Thireos, G., Penn, M. D., and Greer, H. (1984). 5′ untranslated sequences are required for the translational control of a yeast regulatory gene. Proc. Natl. Acad. Sci. USA *81*, 5096–5100.

Tschumper, G., and Carbon, J. (1980). Sequence of a yeast DNA fragment containing a chromosomal replicator and the *TRP1* gene. Gene *10*, 157–166.

van Loon, A. P. G. M., Brändli, A. W., and Schatz, G. (1986). The presequences of two imported mitochondrial proteins contain information for intracellular and intramitochondrial sorting. Cell *44*, 801–812.

vonHeijne, G. (1983). Patterns of amino acids near signal-sequence cleavage sites. Eur. J. Biochem. *133*, 17–21.

vonHeijne, G. (1986). Mitochondrial targeting sequences may form amphiphilic helices. EMBO J. *5*, 1335–1342.

Von Hippel, P. H., Bear, D. G., Morgan, W. D., and McSwiggen, J. A. (1984). Protein–nucleic acid interactions in transcription: a molecular analysis. Ann. Rev. Biochem. *53*, 389–446.