31

Yeast GCN4 Transcriptional Activator Protein

Kevin Struhl
Department of Biological Chemistry and Molecular Pharmacology
Harvard Medical School
Boston, Massachusetts 02115

OVERVIEW

Yeast cells respond to conditions of amino acid starvation by synthesizing GCN4, a protein that binds to the promoters of many amino acid biosynthetic genes and coordinately activates their transcription. GCN4 belongs to the AP-1 transcription factor family, which is highly conserved throughout the eukaryotic kingdom and includes the Jun and Fos oncoproteins. In this review, I discuss the following topics: the GCN4 recognition sequence and its relationship to ATF/CREB sites; the leucine zipper and basic region subdomains that respectively mediate dimerization and specific DNA-binding; the acidic transcriptional activation domain; evidence for functional interactions between GCN4 and TFIID; GCN4 activation in the absence of a TATA element; and potential mechanisms for synergistic activation.

INTRODUCTION

Yeast cells respond to starvation for any single amino acid by coordinately activating at least 40 genes from a wide variety of amino acid biosynthetic pathways (for review, see Hinnebusch 1988). This coordinate induction is mediated by GCN4, a protein that binds specifically to the promoters of the amino acid biosynthetic genes (Hope and Struhl 1985; Arndt and Fink 1986). GCN4 is synthesized only when cells are starved for amino acids, thus explaining why the amino acid biosynthetic genes are transcriptionally induced during starvation conditions (Fig. 1) (Hinnebusch 1984; Thireos et al. 1984). Moreover, the response to starvation is unlikely to involve a regulated modification of GCN4 or interaction with a coregulatory molecule, because GCN4 is a fully functional transcription factor when artificially expressed under normal growth con-

Transcriptional Regulation.
Copyright 1992 Cold Spring Harbor Laboratory Press 0-87969-410-6/92 $3 + 00

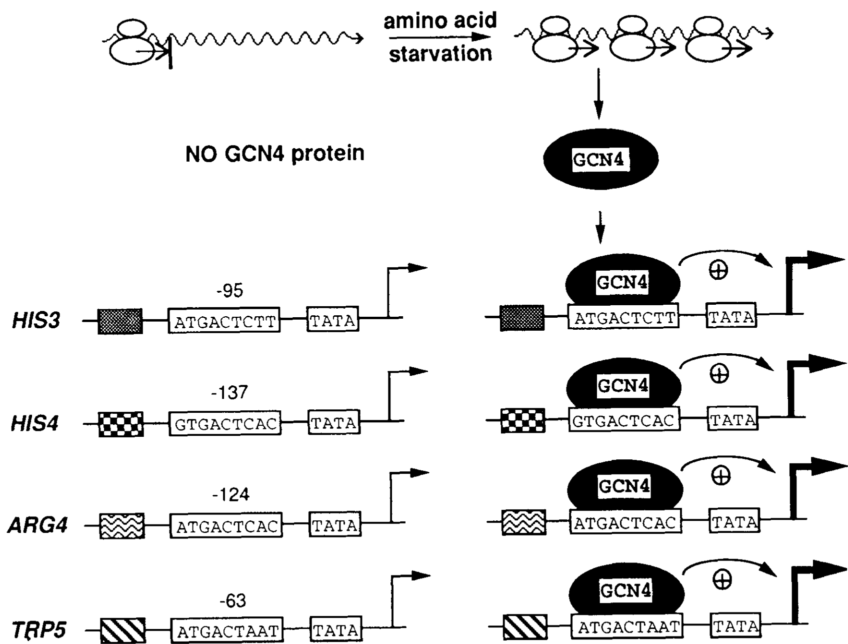

Figure 1 General control of amino acid biosynthesis by GCN4. (*Top*) Under normal growth conditions, translation of GCN4 mRNA is restricted to an upstream open reading frame(s), thus resulting in very low levels of GCN4 protein. Under starvation conditions, ribosomes can reinitiate at the correct AUG codon, which occurs 600 bp from the mRNA start site, thus generating GCN4 protein. (*Bottom*) The promoters of four amino acid biosynthetic genes from three different pathways that are transcriptionally activated by GCN4. The precise sequences of the GCN4-binding sites, and the upstream promoter elements necessary for constitutive transcription (boxes with various patterns), differ among individual promoters. For a thorough review of general control, see Hinnebusch (1988).

ditions. The general control of amino acid biosynthetic genes represents a mechanism to regulate protein synthesis by controlling the amount of amino acid precursors; hence, it is sensible that GCN4, the crucial regulator, is itself controlled by the translation process. As a consequence of its role in regulating protein synthesis, GCN4 is part of the global mechanism that controls cell growth and the decision to initiate new cell division cycles.

In many respects, GCN4 is a typical eukaryotic transcriptional activator protein. First, it contains functionally distinct and physically separate domains for specific DNA binding and for transcriptional activation

(Hope and Struhl 1986). Second, the protein binds as a dimer to a dyad-symmetric sequence which is conserved among binding sites in native yeast promoters (Hill et al. 1986; Hope and Struhl 1987). Third, GCN4 belongs to the bZIP class of eukaryotic transcription factors which is defined by novel structural motifs that mediate dimerization and specific DNA binding (Landschulz et al. 1988). Fourth, it contains a short acidic activation region that is necessary and sufficient for stimulating transcription by RNA polymerase II (Hope and Struhl 1986). Fifth, GCN4 is structurally similar to the Jun oncoprotein (Vogt et al. 1987), and both proteins recognize the same DNA sequences from which they activate transcription in yeast cells (Struhl 1987a, 1988); thus GCN4 is a member of the eukaryotic AP-1 transcription factor family (for review, see Curran and Franza 1988). In this chapter, I review studies that have elucidated the molecular mechanisms by which GCN4 binds its target sites and activates transcription.

NATURE OF THE GCN4 RECOGNITION SITE

The GCN4 recognition sequence has been investigated by saturation mutagenesis of the binding site in the *his3* promoter (Hill et al. 1986) and by selection of binding sites from random-sequence DNA (Oliphant et al. 1989). Both approaches indicate that a 9-bp dyad-symmetric sequence, ATGA(C/G)TCAT, is optimal for DNA binding, with the central 7 bp being most important. Furthermore, uracil interference experiments indicate that the thymine methyl groups at positions ±1 and ±3 are critical for high-affinity binding (Pu and Struhl 1992). The DNA sequence requirements for GCN4 binding in vitro and for transcriptional induction in vivo appear indistinguishable (Hill et al. 1986). Although the optimal recognition sequence strongly resembles the consensus of binding sites from GCN4-regulated promoters, none of the naturally occurring sites is identical to the consensus (Hill et al. 1986). Thus, yeast has evolved a coordinate regulatory system in which the individual promoters contain good, but not optimal, binding sites. Presumably, this permits GCN4 to interact efficiently with a wider variety of sequences, which may allow for regulatory and evolutionary flexibility.

The dyad-symmetric recognition site is recognized by a GCN4 dimer, indicating that the complex consists of two protein monomers interacting with adjacent DNA half-sites (Hope and Struhl 1987). However, the GCN4-binding site is unusually short, with the crucial positions being contiguous and within a single turn of the DNA helix. The compact nature of the target sequence and several other observations suggest that GCN4 dimers bind to overlapping, nonequivalent half-sites (Fig. 2). The optimal binding site is inherently asymmetric because it contains an odd

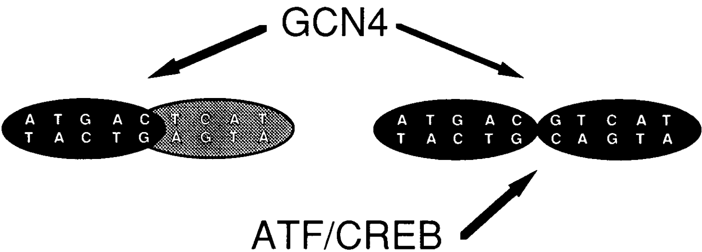

Figure 2 Relationship between the GCN4 and ATF/CREB recognition sequences. The optimal GCN4-binding site is depicted as containing nonequivalent. overlapping half-sites, with the left half-site (black oval) contributing more to the overall affinity than the right half-site (gray oval). The ATF/CREB site is shown as a pair of adjacent, and hence differently spaced, left half-sites. GCN4 prefers the overlapping half-site arrangement to the adjacent arrangement, whereas ATF/CREB proteins have the opposite preference for half-site spacing (Sellers et al. 1990).

number of base pairs and because mutation of the central C:G base pair strongly reduces specific DNA binding (Hill et al. 1986). Moreover, the collection of GCN4-binding sites selected from random-sequence oligonucleotides show nonidentical sequence preferences at symmetrically equivalent positions (Oliphant et al. 1989). These observations indicate that the central C:G base pair is specifically recognized by GCN4 and that the adjacent half-sites contain distinct DNA sequences, ATGAC and ATGAG, and contact the GCN4 dimer with different affinities.

The contributions of the individual half-sites were determined by analyzing symmetrical derivatives of the optimal binding sequence that delete or insert a single base pair at the center of the site (Fig. 2) (Sellers et al. 1990). GCN4 binds efficiently to the sequence ATGACGCAT but not to ATGAGCTCAT or ATGATCAT, thus indicating that the optimal half-site for GCN4 binding is ATGAC, not ATGAG. When GCN4 interacts with the optimal 9-bp target sequence, the left half-site (ATGAC) contributes more to the overall affinity than the right half-site (ATGAG), presumably because the central base pair is contacted only by the monomer interacting with the left half-site. Because alterations in the right half-site are tolerated better than symmetrically equivalent alterations in the left half-site (Oliphant et al. 1989), GCN4 clearly prefers to bind a sequence with one optimal and one weak half-site rather than a sequence with two moderate half-sites; this probably reflects cooperative binding to adjacent half-sites.

The DNA-binding domain of GCN4 is surprisingly flexible because it

can accommodate a major structural disruption, the insertion of a single base pair, in the center of its compact binding site. Although many DNA-binding proteins are highly sensitive to spacing changes in the target site, some proteins tolerate or even prefer different spacings between half-sites (Sadler et al. 1983; Falvey and Grindley 1987; Sauer et al. 1988). However, in all these cases of flexibility, the sequence at the center of the binding site is relatively unimportant, and the protein dimerization region resides in a distinct structural domain from the region required for DNA contacts; thus, the DNA interaction surfaces of the two monomers are structurally independent. In contrast, the dimerization and DNA-binding functions of GCN4 are localized to the 60 carboxy-terminal residues (Hope and Struhl 1986, 1987), a region that appears to be a single structural domain as determined by proteolytic mapping (Hope et al. 1988).

The ATGACGTCAT sequence recognized by GCN4 strongly resembles sites bound by the yeast and mammalian ATF/CREB family of proteins (Hai et al. 1988; Roesler et al. 1988). Like GCN4, these proteins bind as dimers, and they contain leucine zipper motifs and adjacent basic regions (Hoeffler et al. 1988; Gonzalez et al. 1989; Hai et al. 1989). Thus, GCN4 and the ATF/CREB protein family recognize adjacent ATGAC half-sites but have different spacing requirements. In support of this idea, the mammalian AP-1 protein family, which recognizes the same sequences as GCN4 (Bohmann et al. 1987; Struhl 1987a), is immunologically related to the ATF/CREB protein family (Hai et al. 1988). Thus, the GCN4/AP-1 and ATF/CREB classes of proteins likely belong to the same evolutionarily conserved superfamily of proteins that recognize essentially identical half-sites (Fig. 2). This situation resembles that of the estrogen and thyroid hormone receptors, which recognize similar half-sites with distinct spatial constraints (Glass et al. 1988; Umesono and Evans 1989).

Although DNA-binding specificity is defined largely by the ATGAC half-sites, the region of DNA covered by the bound protein is considerably more extensive (Gartenberg et al. 1990). When GCN4 is incubated with a collection of different-sized oligonucleotides containing a given target site, optimal binding is observed only with DNAs containing at least an 18-bp region encompassing the half-sites. Chemical modification experiments reveal that GCN4 contacts essentially all nucleosides and phosphates over a region spanning one and a half turns of the DNA helix. GCN4 interacts primarily, and possibly exclusively, with the major groove of DNA. The protein yields no detectable footprint with hydroxyl radical (a reagent specific for minor groove interactions) (Gartenberg et al. 1990), and affinity cleavage experiments involving an iron-EDTA

modified protein indicate that the amino termini of the dimeric DNA-binding domain lie in the major groove 9–10 bp apart (Oakley and Dervan 1990).

NATURE OF THE GCN4 DNA-BINDING DOMAIN

Extensive deletion analysis of the 281-amino-acid GCN4 protein indicates that the 56 carboxy-terminal amino acids are sufficient both for dimerization and for specific DNA binding (Hope and Struhl 1986, 1987; Weiss et al. 1990). The DNA-binding domain can be isolated from the full-length protein as a proteolytically stable fragment, indicating that it folds independently of the remainder of the protein (Hope et al. 1988). Moreover, GCN4 and the Jun oncoprotein bind the same DNA sequences (Struhl 1987a), yet amino acid sequence conservation between these proteins is restricted to the 65 carboxy-terminal residues (Fig. 3) (Vogt et al. 1987).

The GCN4 DNA-binding domain contains the bZIP structural motif found in a class of eukaryotic transcription factors that includes C/EBP and the Jun and Fos oncoproteins (Landschulz et al. 1988). Within the bZIP domain is a leucine zipper that consists of four or five leucines spaced exactly seven amino acids apart, embedded within a region whose sequence is consistent with the formation of an amphipathic α-helix. Adjacent to the leucine zipper is a conserved region that is rich in basic

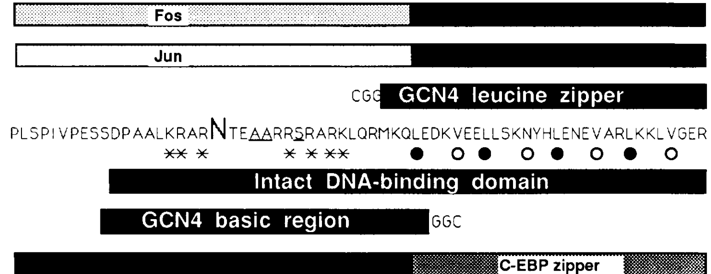

Figure 3 The GCN4 DNA-binding domain. The sequence of the 65 carboxy-terminal residues is shown with conserved features highlighted as follows: leucine residues that define the leucine zipper (closed circles); alternate hydrophobic residues that form the coiled-coil interface (open circles); positively charged residues that define the basic region (asterisks); other conserved residues in the basic region (underlined); the invariant asparagine (large bold N). Shown above the sequence are a leucine zipper peptide and Fos-GCN4 or Jun-GCN4 chimeric proteins that all display GCN4 dimerization specificity. Shown below are the intact bZIP domain, a basic region peptide, and a GCN4-C/EBP chimeric protein, all of which display GCN4 DNA-binding specificity.

residues and contains a quartet of uncharged residues including two alanines, a serine/cysteine, and an invariant asparagine (Fig. 3). The spacing between the leucine zipper and the basic region is precisely maintained in this family of DNA-binding proteins.

Distinct Subdomains for Dimerization and DNA Binding

Chimeric proteins have been used to prove that the GCN4 leucine zipper confers the specific dimerization properties of the intact protein and that the adjacent basic region is sufficient for specific DNA binding (Fig. 3). The basis of such experiments is that the various bZIP proteins have distinct dimerization and DNA-binding properties, despite having common sequence motifs. In the case of the dimerization, GCN4, Jun, and Fos contain the conserved leucines in the zipper and interact with the same DNA sites, yet the only functional species are GCN4 homodimers, Jun homodimers, and Fos-Jun heterodimers. However, precise replacement of the Fos zipper by the GCN4 zipper generates a Fos-GCN4 chimeric protein with GCN4 dimerization specificity; it binds DNA as a homodimer or as a heterodimer with GCN4, but not as a heterodimer with Jun (Kouzarides and Ziff 1989; Sellers and Struhl 1989). Conversely, GCN4 and C/EBP recognize different DNA sequences, and analysis of similar zipper-basic region chimeric proteins indicates that DNA-binding specificity tracks with the basic region (Agre et al. 1989).

The fact that leucine zipper and basic regions can be interchanged between different family members to generate chimeric proteins with predicted dimerization and DNA-binding specificities indicates that these conserved motifs encode distinct structural subdomains. More compellingly, synthetic peptides corresponding to the isolated subdomains are functionally active (Fig. 3). Synthetic leucine zippers form dimers of appropriate specificity (O'Shea et al. 1989a,b), and a synthetic basic region (dimerized via a disulfide bond) can specifically interact with the correct target sequences, although with reduced affinity (Talanian et al. 1990).

The Leucine Zipper

The original structural concept of the leucine zipper invoked an α-helical dimer formed primarily by interdigitation of leucine residues within the hydrophobic interface (Landschulz et al. 1988). In support of this idea, a GCN4 leucine zipper peptide (the 33 carboxy-terminal residues) forms stable α-helical dimers in solution (O'Shea et al. 1989a). The same region exists as a dimeric α-helical structure in the context of a func-

tional DNA-binding domain, although it is considerably less stable than the zipper peptide, probably due to repulsion of the basic regions (O'Neil et al. 1990; Weiss et al. 1990). However, in contrast to the prediction of the initial interdigitation model, the α-helices associate in the parallel rather than antiparallel arrangement (O'Shea et al. 1989a). Moreover, X-ray scattering studies demonstrate that the GCN4 leucine zipper is similar to the coiled-coil structure found in muscle filament proteins (Rasmussen et al. 1991), and DNA binding by GCN4 is maintained when the leucine zipper is replaced by an artificial coiled coil (O'Neil et al. 1990). In the coiled coil, the dimerization interface is not formed by leucine interdigitation, but rather by interaction of the leucines with hydrophobic residues predicted to lie on the same side of the α-helix.

Because the canonical leucine residues are common to all zipper proteins, nonconserved residues in the various zipper regions must have critical roles in generating distinct dimerization specificities and hence zipper association properties. In fact, the GCN4 leucine zipper is surprisingly tolerant of mutations in the leucine residues (Struhl 1989; Hu et al. 1990; vanHeeckeren et al. 1992). A wide variety of single substitutions at any of the four leucines including basic (Arg-267 and Arg-274) and acidic (Glu-260) amino acids behave indistinguishably from wild-type GCN4, and some derivatives containing two leucine substitutions display detectable but reduced function. The observations do not imply that the leucines are functionally unimportant, but rather indicate that numerous other interactions within the coiled coil are crucial for efficient dimerization.

Nuclear magnetic resonance experiments carried out on a 58-residue DNA-binding domain yield an estimated lifetime between 10 and 1000 milliseconds for GCN4 dimers (Weiss et al. 1990). In conjunction with the dissociation constant of GCN4 for DNA, this suggests that unfolding and reassembly of GCN4 occurs easily, thereby facilitating subunit exchange. This property, which is also observed for C/EBP (Shuman et al. 1990) and is probably generally true for bZIP proteins, makes it possible for organisms to rapidly change the spectra of homodimeric and heterodimeric species in response to environmental and developmental signals.

Spacing between the Leucine Zipper and Basic Region

The precisely conserved spacing relationship between the subdomains led to the suggestion that the leucine zipper positions the basic region for specific DNA binding (Landschulz et al. 1988; Vinson et al. 1989). In support of this idea, disruption of this spacing by insertion of two, four,

five, or six amino acids between the GCN4 leucine zipper and basic region abolishes GCN4 function (Agre et al. 1989; Pu and Struhl 1991a). More convincingly, insertion of a surprisingly wide variety of seven-amino-acid sequences results in proteins displaying weak to wild-type levels of GCN4 activity (Fig. 4) (Pu and Struhl 1991a). Thus, the correct spatial relationship is retained upon the insertion of an integral number of α-helical turns (7 residues) between the zipper and basic region. Interestingly, heterodimers between GCN4 and several heptapeptide insertion proteins fail to bind DNA; i.e., both proteins contain an acceptable spacing between the leucine zipper and basic region, but the distinct spacings are not mutually compatible (Fig. 4). These results strongly suggest that the leucine zipper symmetrically orients the two basic regions along the adjacent half-sites and that the region between the two subdomains is α-helical. In addition, they suggest that GCN4 homodimers are the primary, and possibly the sole, mediators of GCN4 function in yeast cells (because it is extremely unlikely that GCN4 and the heptapeptide insertions can form DNA-binding heterodimers with a common set of other leucine zipper proteins).

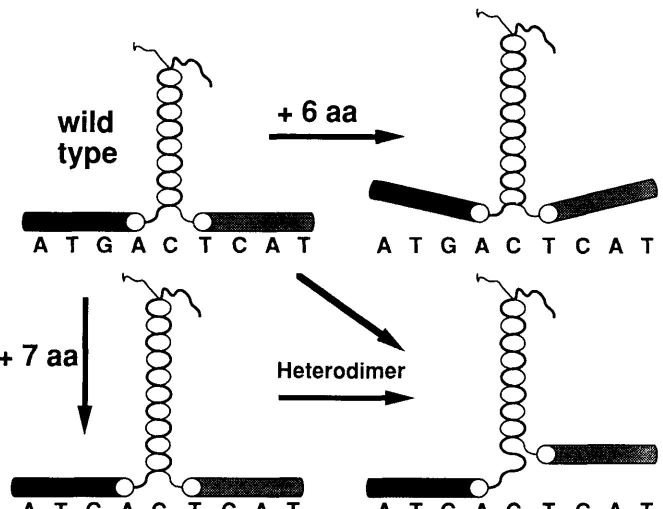

Figure 4 The leucine zipper symmetrically positions the adjacent basic regions for DNA binding. The spacing between the leucine zipper (intertwined wavy lines perpendicular to the DNA) and both adjacent basic regions (cylinders; black representing the left monomer and gray representing the right monomer) must be correct for high-affinity binding (shown by a close, parallel arrangement of both basic regions to the DNA sequence). The region between the zipper and basic region is shown as α-helical because this spacing can be altered by the insertion of an integral number of α-helical turns (Pu and Struhl 1991a).

GCN4 Undergoes a Global Folding Transition upon Specific DNA Binding

DNA-binding domains are generally prefolded structures that specifically interact with the DNA helix by virtue of complementary surfaces. In striking contrast, the GCN4 DNA-binding domain undergoes a global folding transition upon specific interaction with DNA (Fig. 5) (O'Neil et al. 1990; Talanian et al. 1990; Weiss et al. 1990). In the absence of DNA, the dimeric DNA-binding domain (56 carboxy-terminal residues) is approximately 70% α-helical, as determined by circular dichroism spectroscopy. This α-helicity is accounted for by the leucine zipper, thereby implying that the adjoining basic region is largely unstructured in the absence of DNA. However, addition of a GCN4-binding site increases the α-helix content to at least 95%, indicating that the basic region acquires substantial α-helical structure when it specifically binds to DNA. These observations are consistent with, but not specific to, the scissors-grip (Vinson et al. 1989) and induced fork (O'Neil et al. 1990) models, which predict that the leucine zipper symmetrically positions the diverging pair of α-helical basic regions for specific DNA binding to abutting half-sites.

Although the basic region is largely unstructured in the absence of DNA, the α-helical content of the GCN4 DNA-binding domain increases to about 80% at lower temperatures (Weiss et al. 1990). This partial α-helical transition is also observed with a 26-residue peptide correspond-

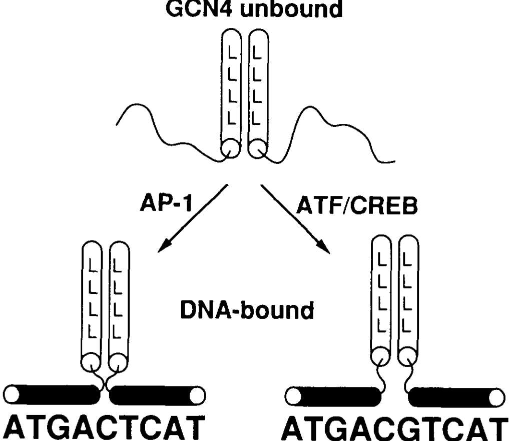

Figure 5 GCN4 undergoes a global folding transition upon specific binding to DNA (Weiss et al. 1990). Unbound GCN4 is shown as dimerized leucine zipper (cylinders with leucine residues indicated) and unstructured basic regions (wavy lines). Upon binding to the AP-1 or ATF/CREB sites, the basic regions become almost completely α-helical (black cylinders). In order to accommodate the different half-site spacing of these binding sites, the region between the leucine zipper and basic region is shown as being flexible.

ing to the basic region, suggesting that these conformational properties are locally determined. These observations suggest that in the absence of DNA, the basic region of GCN4 exists as an ensemble of structures, with the folded state being significantly populated only at low temperature. More importantly, specific target sequences stabilize the α-helical conformation of the basic region, thus inducing the fit between protein and DNA.

In the protein-DNA complex, GCN4 is structurally quite rigid due to its almost completely α-helical nature. However, the protein undergoes the same global folding transition when bound to the ATF/CREB site that contains an additional base pair between the adjacent half-sites, suggesting some degree of flexibility in the protein-DNA complex. Since the DNA structure is not differentially affected by GCN4 binding, the alternative half-site spacings are accommodated by flexibility in the protein (Weiss et al. 1990). Flexibility in protein conformation, not a protein-induced bend in DNA, also appears to account for the anomalous electrophoretic behavior of protein-DNA complexes (Gartenberg et al. 1990). The most likely structural basis for this flexibility is at or near the bifurcation where the helices of the two basic regions split off from the dimeric coiled coil of the leucine zipper (Fig. 5).

DNA Binding

It has been proposed that the quartet of highly conserved, nonpolar residues in the basic region (corresponding to Asn-235, Ala-238, Ala-239, and Ser-242 in GCN4) lie on the face of the α-helix that docks against the DNA (O'Neil et al. 1990). In support of this idea, DNA-binding activity is retained in a derivative in which five nonconserved residues on the putative solvent-exposed surface are changed to alanine or glutamine (O'Neil et al. 1990). Surprisingly, however, these highly conserved residues are not essential for DNA binding by GCN4 (Pu and Struhl 1991b). At positions 238 and 239, a variety of nonpolar residues can functionally substitute for the conserved alanines; in general, increasing the size of the side chain results in decreased GCN4 function (Pu and Struhl 1991b). For the invariant asparagine (Asn-235), most substitutions abolish GCN4 DNA binding, but the Trp-235 protein displays nearly wild-type function, and the Gln-235 and Ala-235 proteins show detectable activity. The ability of an amino acid to functionally substitute for Asn-235 does not correlate with its preference for assuming the N-cap position of an α-helix (Richardson and Richardson 1988). This argues against a specific prediction of the scissors-grip model (Vinson et al. 1989) that the invariant asparagine functions primarily to form an N-cap

structure that permits the α-helical basic region to bend sharply so that it can wrap around the DNA. However, the more general feature of this model, that the basic region bends to maximize the protein-DNA interface, remains to be addressed.

Strong clues to the direct protein-DNA contacts involved in high-affinity binding have come from GCN4 derivatives that display altered DNA-binding specificity (Fig. 6). Such altered specificity mutants were identified by genetically selecting for proteins that can activate transcription from promoters containing symmetrically mutated binding sites. For example, wild-type GCN4 binds the optimal ATGACTCAT sequence much more efficiently than TTGACTCAA, whereas the Trp-235 protein binds these sites with similar affinity (Tzamarias et al. 1992). Moreover, the Trp-235, Ala-235, and Gln-235 proteins strongly discriminate against GTGACTCAC, a site efficiently bound by GCN4. These results strongly suggest a direct interaction between Asn-235 and the ±4 position of the target site. At the +3 position, a protein containing Tyr-238 instead of the conserved Ala-238 has the novel property of efficiently recognizing ACGACTCGT (J. Kim et al., unpubl.). Similarly, changing Ala-239 to Val-239 results in a protein that possesses near wild-type affinity

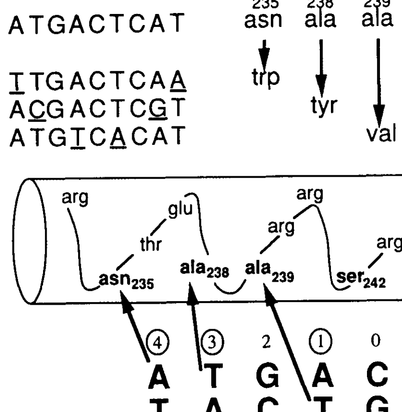

Figure 6 Potential alignment of GCN4 along the DNA as inferred by altered specificity mutants (Tzamarias et al. 1992; J. Kim et al., unpubl.). Shown below the optimal binding site and relevant wild-type GCN4 residues (Asn-235, Ala-238, and Ala-239) are the amino acid substitutions that display GCN4 activity on the indicated binding sites with symmetric mutations (underlined). These results suggest a model in which the crucial α-helix in the basic region (cylinder) specifically binds the target DNA (positions –4 to 0 corresponding to the left half-site indicated) via direct interactions (arrows) between the indicated amino acids (bold and numbered) and base pairs (circled).

for ATGTCACAT, a site not recognized by GCN4. Taken together, these observations strongly support a model in which the α-helical surface defined by amino acid positions 235, 238, and 239 is aligned along the DNA with direct contacts to ±4, ±3, and ±1, respectively (Fig. 6). Given that hydrophobic interactions between alanine residues and 5-methyl groups of thymines are often found in protein-DNA complexes, it seems likely that Ala-238 and Ala-239 interact respectively with the thymine methyl groups at ±3 and ±1. However, conclusive proof of this hypothesis awaits a high-resolution structure of the protein-DNA complex.

Although highly conserved features of protein families are presumed to be fundamentally important for function, many such features of bZIP proteins are not essential for GCN4 DNA binding. The spacing between the zipper and basic region can be altered by inserting an integral number of helical turns; the invariant asparagine and conserved alanines can be substituted, and in some instances, generate proteins with altered binding specificity; and the canonical leucine residues in the zipper can be varied considerably. Thus, it seems very likely that there are eukaryotic transcriptional regulatory factors that lack some or many of the defining characteristics of bZIP proteins, yet nevertheless are structurally and functionally homologous.

NATURE OF THE GCN4 TRANSCRIPTIONAL ACTIVATION REGION

Extensive deletion analysis has defined the regions of GCN4 required for transcriptional activation in vivo (Fig. 7) (Hope and Struhl 1986; Hope et

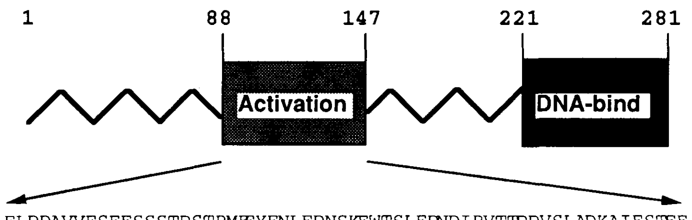

Figure 7 DNA-binding and transcriptional activation functions of GCN4. Shown are locations of the DNA-binding domain (black box), transcriptional activation region (gray box), and nonessential portions (wavy lines) of GCN4 (281 amino acids in length; boundary residues indicated) (Hope and Struhl 1986). Shown below is the sequence of the activation region with acidic residues underlined. Although the acidic region is 60 amino acids in length, a variety of derivatives containing only 40 residues display full activity in vivo (Hope et al. 1988).

al. 1988). Surprisingly, approximately 60% of the GCN4 coding sequence can be deleted without significantly affecting the ability of the protein to activate transcription. However, deleted proteins retaining only the DNA-binding domain cannot activate transcription. In fact, such derivatives bind DNA in vivo and repress transcription from certain promoters, thus inhibiting cell growth in the absence of amino acids (Hope and Struhl 1986). For transcriptional activation in vivo (Hope and Struhl 1986; Hope et al. 1988) or in vitro (Ponticelli and Struhl 1990), a short region of GCN4 located in the center of the protein is required in addition to the DNA-binding domain. This transcriptional activation region is functionally autonomous; it stimulates transcription when fused to a heterologous DNA-binding domain such as the LexA repressor, and its activity is independent of its spacing and orientation with respect to the DNA-binding domain (Hope and Struhl 1986; Hope et al. 1988).

Short Acidic Sequences Are Sufficient for Transcriptional Activation

The transcriptional activation region of GCN4 maps within a central region of the protein that contains 30% acidic residues over a 60-amino-acid stretch (Fig. 7) (Hope and Struhl 1986). Surprisingly, different portions of the GCN4 acidic region are equally capable of activating transcription even though their primary sequences are dissimilar. As few as 35–40 amino acids from this acidic region joined to the GCN4 DNA-binding domain are sufficient for transcriptional activation (Hope et al. 1988). Thus, the yeast GCN4 transcriptional activation function is defined by a short acidic region whose precise sequence is relatively unimportant. In accord with these nonstringent sequence requirements, the GCN4 activation region is much more sensitive to proteolytic cleavage than the DNA-binding domain (Hope et al. 1988). Functional studies on yeast GAL4 (Ma and Ptashne 1987a) and the Jun oncoprotein (Struhl 1988) indicate that transcriptional activation is mediated by acidic regions that are not homologous to each other or to the GCN4 acidic region. Moreover, acidic character is the common feature of activation regions selected from *Escherichia coli* DNA segments (Ma and Ptashne 1987b). Thus, transcriptional activation regions do not have defined tertiary structures such as are found in active sites of enzymes or in conventional structural domains.

Although many different acidic sequences can serve as transcriptional activation regions, and negative charge is clearly important, other structural features influence the level of transcriptional stimulation. Progressive deletion of the GCN4 activation region indicates that transcriptional

activity is directly related to the length of the acidic region but not precisely related to the number of acidic residues (Hope et al. 1988). 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. This repeating unit could be an amphipathic α-helix, since the GCN4 acidic region is compatible with such a structure, and stepwise loss in activation potential is correlated with the removal of two α-helical turns (Hope et al. 1988). Similarly, the level of activation mediated by selected *E. coli* segments is only loosely correlated with net negative charge (Ma and Ptashne 1987b), and acidic peptides of identical amino acid composition but distinct sequence and α-helical potential can have very different transcriptional activation properties (Giniger and Ptashne 1987). However, because many other acidic activation regions are unlikely to form amphipathic helices, a simple relationship between this structure and function appears unlikely.

The GCN4 Activation Domain Is Monomeric

Because GCN4 binds DNA as a dimer, it brings two acidic activation regions to the promoter. However, a Fos-Jun heterodimeric DNA-binding domain containing only a single GCN4 acidic region activates transcription (Fig. 8) (Oliviero and Struhl 1991). The clearest example is the combination of Fos containing the GCN4 acidic region and the Jun DNA-binding domain; neither protein can activate transcription alone because the Fos derivative is unable to bind DNA and the Jun derivative lacks an activation region. Nevertheless, this heterodimer activates transcription as efficiently as a Fos-Jun combination in which both moieties contain a GCN4 acidic region. Thus, the activation domain is a monomeric structure, and the number of acidic regions on a DNA-bound protein does not significantly affect the level of transcription.

The monomeric nature of acidic activation domains increases the complexity of regulation that can be mediated by protein families that contain common dimerization motifs such as the leucine zipper and the helix-loop-helix (Murre et al. 1989). For example, proteins lacking an activation region can stimulate transcription if they associate into DNA-binding heteromers with partners that contain an activation domain. Conversely, the ability of a gene product to stimulate transcription does not necessarily indicate that the protein itself contains an activation function. In fact, conventional mapping of the transcriptional activation function on such a protein would instead uncover a motif necessary for oligomerization and/or DNA binding. For these reasons, an individual protein may serve as a transcriptional activator or repressor, depending

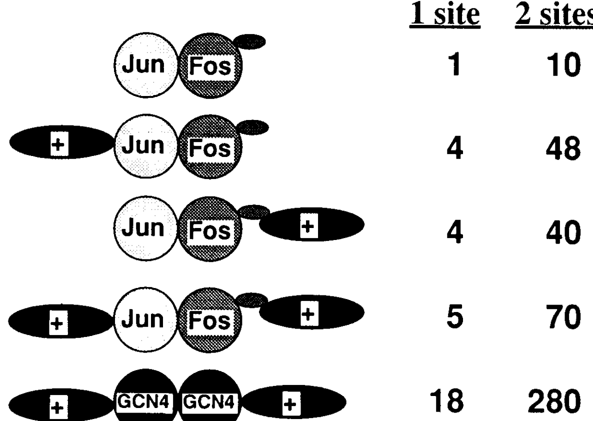

Figure 8 Transcriptional activities of Fos-Jun heterodimers containing GCN4 activation regions. The indicated heterodimers were generated in yeast cells by cointroducing Fos and Jun molecules (bZIP domains indicated by shaded circles; the small oval on the Fos moiety represents nondeleted sequences outside the bZIP region that weakly contribute to transcriptional activation) that either did or did not contain a GCN4 activation region (black oval with plus). For each heterodimer (and for GCN4 homodimers), the relative level of transcription from promoters containing one or two optimal AP-1 target sites upstream of the *his3* TATA region and structural gene is indicated. The indicated levels of transcription are almost exclusively due to Fos-Jun heterodimers; yeast cells containing only the Jun or Fos derivative confer very little, if any, transcriptional activity. For details see Oliviero and Struhl (1991).

on the environmental or developmental situation that affects the presence or activity of other cross-oligomerizing members of the protein family.

TRANSCRIPTIONAL ACTIVATION MECHANISMS

It has been hypothesized that the DNA-binding domains of transcription factors serve merely to bring the protein to the DNA target, whereupon the activation region can interact with a component(s) of the basic transcription machinery (Brent and Ptashne 1985). Given that acidic regions are short and variable in sequence, it is likely that they constitute acidic surfaces necessary for protein-protein interactions (Hope and Struhl 1986). In accord with this idea, the yeast GAL4 activator protein cannot stimulate transcription by a heterologous transcription machinery such as bacteriophage T7 RNA polymerase (Chen et al. 1987). Since acidic regions are necessary for yeast activator proteins to function in mammalian cells (Kakidani and Ptashne 1988; Webster et al. 1988) and for

mammalian activator proteins to function in yeast cells (Schena and Yamamoto 1988; Struhl 1988), it is likely that they contact some part of the basic transcription machinery that is conserved functionally throughout the eukaryotic kingdom.

Evidence for a Functional Interaction between GCN4 and TFIID

The yeast *his3* promoter contains two functional TATA-like elements, T_R and T_C, but GCN4 can only activate transcription dependent on the T_R element (Fig. 9) (Struhl 1986). The T_R element is defined primarily by the canonical TATA sequence, TATAAA, that interacts with the general transcription factor IID. Almost all single mutations of this sequence significantly reduce transcription in vitro (Chen and Struhl 1988), with the resulting levels being strongly correlated with the level of TFIID-dependent transcription in vitro (Wobbe and Struhl 1990). In contrast, the T_C element is surprisingly tolerant of single base substitution muta-

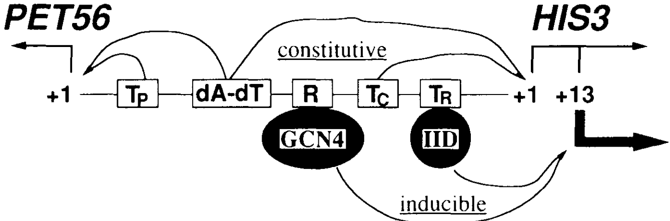

Figure 9 Independent regulation of divergently transcribed genes by selective interaction between GCN4 and TATA elements. *his3* (a histidine biosynthetic gene) and *pet56* (a gene required for mitochondrial function) are divergently transcribed from initiation sites only 191 bp apart. Constitutive transcription utilizes the shaded promoter elements (Struhl 1986); both genes require a poly(dA)•poly(dT) upstream element that can function in combination with the *his3* T_C (which is responsible for essentially all transcription initiated at +1) and the *pet156* T_P TATA elements. The poly(dA)•poly(dT) sequence also can function with T_R (Harbury and Struhl 1989; not shown in figure). GCN4 activation occurs only in combination with the *his3* T_R element, which by genetic and biochemical criteria is a classic TFIID-dependent TATA element (Chen and Struhl 1988; Ponticelli and Struhl 1990; Wobbe and Struhl 1990). This selectivity explains why GCN4 activates *his3* but not *pet56* transcription. In addition, *his3* transcription is activated only from the +13 site because GCN4 cannot function in combination with T_C, and T_R is too close to the +1 site for efficient initiation. The mechanism of T_C-mediated transcription is clearly distinct from the TFIID-dependent transcription from a canonical TATA element (Mahadevan and Struhl 1990; Ponticelli and Struhl 1990).

tions (Mahadevan and Struhl 1990), and it does not support transcription in yeast nuclear extracts under conditions where all known T_R-dependent phenomena can be accurately simulated (Ponticelli and Struhl 1990). The mechanism of T_C-dependent transcription is unknown, yet is clearly distinct from classic TATA- and TFIID-dependent transcription. It may involve a novel TATA factor, an effect of chromatin, or an alternative mechanism involving TFIID. Another distinction between the two *his3* TATA elements is that overproduction of GAL4 squelches transcription dependent on T_R but not T_C (Gill and Ptashne 1988). Thus, both genetic and biochemical evidence indicates that GCN4 activation is specific to TFIID-dependent transcription from the classic TATA element T_R. This specificity provides strong evidence for an interaction between GCN4 and TFIID, presumably mediated by the acidic activation domain (Struhl 1987b).

In a related set of experiments, wild-type GCN4 protein is required specifically for *his4* transcription that depends on a canonical TATA element (Pellman et al. 1990). Unlike BAS1 and BAS2, proteins that bind to the *his4* promoter and are necessary for basal transcription in the absence of this TATA element (Arndt et al. 1987), GCN4 cannot support TATA-independent transcription. Interestingly, however, GCN4 derivatives deleted for various parts of the acidic activation region stimulate low levels of TATA-independent transcription that is initiated from the correct site. Analysis of a series of GCN4 deletion mutants indicates that the discrimination between TATA-dependent and TATA-independent *his4* transcription is correlated with the strength of activation (Pellman et al. 1990). Thus, as is the case for *his3* transcription, GCN4 functions efficiently only in combination with conventional TATA elements that are TFIID interaction sites.

Functional interactions between activator proteins and TFIID have been inferred from several independent lines of evidence. First, functional distinctions between TATA elements similar to those described for T_R and T_C have been observed for a variety of different activator proteins; i.e., only certain combinations of enhancer and TATA elements are compatible for activation (Homa et al. 1988; Simon et al. 1988; Harbury and Struhl 1989). Second, functional TATA elements that support equivalent levels of basal TFIID-dependent transcription in vitro (Wobbe and Struhl 1990) respond extremely differently to GAL4-mediated activation in vivo (Harbury and Struhl 1989). Third, upstream activators and TFIID can cooperatively interact with DNA (Sawadogo and Roeder 1985; Horikoshi et al. 1988b) in a manner that may involve the acidic activation region (Horikoshi et al. 1988a). Fourth, TFIID directly and specifically interacts with the acidic activation domain encoded by the her-

pesvirus VP16 protein (Stringer et al. 1990). Analysis of mutant proteins indicates that the quality of the interaction in vitro is moderately correlated with the level of transcriptional stimulation in vivo (Ingles et al. 1991). This last observation suggests that the combinatorial nature of upstream activator and TATA-binding proteins may reflect direct, allosteric interactions that influence the activity of the transcription machinery. However, other lines of evidence suggest the possibility that TFIIB is the target of acidic activator proteins (Lin and Green 1991). Thus, an alternative model for TATA element specificity may involve interactions between TFIIB and TFIID that are differentially affected by the acidic activation region and the sequence of the TATA element.

In yeast, the restriction of GCN4 activation to classic TATA-dependent promoters is utilized as an important regulatory mechanism. For example, GCN4 binds about 50 bp upstream of the TATA-like elements of both the divergently transcribed *his3* and *pet56* genes, yet despite its ability to function bidirectionally, it only induces *his3* transcription (Fig. 9). The basis for this discordant regulation is that *his3* contains the T_R element that responds to GCN4 activation, whereas the *pet56* TATA-like sequence behaves like T_C and hence is not responsive to upstream activator proteins (Struhl 1986). Thus, functionally distinct TATA elements provide a mechanism for closely packed and divergently transcribed genes to be regulated independently. This may be particularly important for eukaryotic organisms, because they rely on bidirectional upstream elements that act over long distances.

Biochemical Evidence for a GCN4-RNA Polymerase II Interaction

Affinity chromatography indicates a direct and selective interaction between GCN4 and RNA polymerase II (Brandl and Struhl 1989). RNA polymerase II was selectively retained on a GCN4-Sepharose column under conditions where proteins that copurify over three or four ion-exchange columns flow through. Conversely, GCN4 binds to an RNA polymerase II column but not to control columns. Surprisingly, the GCN4 DNA-binding domain is necessary and sufficient for this interaction with RNA polymerase II. It is unlikely that this interaction reflects a trivial ionic effect because deletion of 11 carboxy-terminal residues of GCN4 has a minor effect on overall charge but eliminates the interaction.

It is tempting to believe that this selective interaction between the GCN4 DNA-binding domain and RNA polymerase II is relevant for transcription, but direct evidence is lacking. Of course, any functional role for this interaction would be mechanistically distinct from the role of the

acidic activation region. In addition, it would be contrary to the commonly held view that the DNA-binding domain does not have a direct role in transcription other than targeting the protein to the promoter. However, the potential importance of DNA-binding domains in the transcription process has been implicated by studies on *E. coli* activator proteins (Hochschild et al. 1983; Irwin and Ptashne 1987; Bushman and Ptashne 1988), by the existence of glucocorticoid receptor or HAP1 derivatives that bind DNA normally but fail to activate transcription (Kim and Guarente 1989; Schena et al. 1989), and by experiments on synergistic activation to be discussed below (Oliviero and Struhl 1991).

GCN4 Can Activate Transcription When Its Binding Site Replaces the TATA Element

Although GCN4 normally activates transcription when bound upstream of a TATA element, it can stimulate transcription when bound at the position of the TATA element in a *gal-his3* promoter (Fig. 10) (Chen and Struhl 1989). This TATA-independent transcription requires the GCN4 acidic activation region, and it occurs from normal initiation sites. In contrast to normal TATA-dependent activation, the initiation pattern resembles that mediated by the T_C element, rather than the T_R element, and it does not respond to GAL4 protein. Instead, GCN4 activation from the promoter-proximal element requires at least two additional elements in the *gal* enhancer region that are distinct from the GAL4 sites and from TATA elements (Brandl and Struhl 1990). The most important of these elements, termed Q, interacts with a yeast protein (Brandl and Struhl 1990; Chasman et al. 1990) and corresponds with a nucleosome positioning sequence (Fedor et al. 1988). Consistent with (but hardly conclusive for) a potential role for nucleosome positioning, the level of transcription is significantly reduced upon small increases in the distance between Q and the GCN4 site.

These observations suggest the possibility of an alternative mechanism for transcriptional activation in which TFIID either is not required or is not directly bound to DNA. Furthermore, if TFIID is not functionally involved in this case, the requirement for the GCN4 activation region disfavors the view that TFIID is the functionally important target of acidic activation domains. Finally, these results provide additional evidence for combinatorial specificity between distal and proximal promoter elements. In particular, GAL4 can function with T_R, and Q can function with GCN4 (Fig. 10), but other combinations such as GAL4 and GCN4 or Q and T_R are nonfunctional.

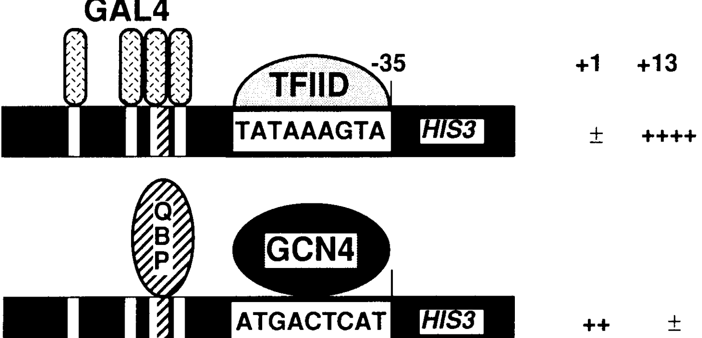

Figure 10 TATA-independent activation by GCN4. (*Top*) The *his3*-G17 promoter consists of the 365-bp *gal* enhancer fragment with four GAL4-binding sites fused to the *his3* T_R TATA element (which interacts with TFIID) and structural gene (Chen and Struhl 1988). Because transcription depends on GAL4 and the TATA element, it occurs only in galactose medium and is initiated with a very strong preference for the +13 site. (*Bottom*) The *his3*-GG1 promoter is identical to the *his3*-G17 promoter, except that the TATA element has been precisely replaced by a GCN4-binding site. Transcription from *his3*-GG1 does not involve GAL4, occurs in both glucose and galactose medium, and is initiated preferentially from the +1 site (Chen and Struhl 1989). Instead, transcription requires GCN4 (including the acidic activation region) and a distinct protein (QBP) that interacts with a site that partially overlaps one of the GAL4 sites (Brandl and Struhl 1990).

Synergistic Activation Does Not Depend on the Number of Acidic Regions

Transcriptional enhancement by activator proteins is synergistic in that promoters containing multiple protein-binding sites upstream of a TATA element are often 10–100 times more active than analogous promoters containing single binding sites. Such synergy is frequently observed when the multiple binding sites are recognized by distinct, and even evolutionarily distant, proteins. Such promiscuous synergy is a fundamental aspect of eukaryotic transcription and constitutes an important basis for the extraordinarily diverse patterns of gene expression mediated by enhancers (Struhl 1991).

Although cooperative DNA binding of transcription factors to adjacent promoter sites is likely to account for some cases of synergy (Driever et al. 1989; Schmid et al. 1989; Struhl et al. 1989; Tsai et al. 1989), the promiscuity of the phenomenon strongly suggests that there must be alternative mechanisms. In accord with this view, synergistic activation has been observed in vitro under conditions where the binding

sites for a given activator protein are fully occupied (Carey et al. 1990; Lin et al. 1990). One such alternative mechanism is that acidic activation regions associated with DNA-binding proteins bound to adjacent promoter sites interact synergistically with a common target of the basic transcription machinery (for review, see Ptashne 1988). In such a model, the common target would respond in a nonlinear fashion to the number and/or quality of acidic activation regions.

To determine whether synergy depends on the number of acidic activation regions, the transcriptional activity of Fos-Jun heterodimers containing one or two GCN4 acidic regions was assayed on promoters containing one or two target sites (Fig. 8) (Oliviero and Struhl 1991). As mentioned previously, the number of GCN4 acidic regions on such heterodimers does not affect the level of transcription from promoters containing one target site. More importantly, all the heterodimer combinations and wild-type GCN4 protein stimulate transcription 10–15% more efficiently on analogous promoters containing two adjacent sites. In other words, there is a dramatic difference in transcriptional activation when two GCN4 acidic domains are located on two DNA-bound proteins, as opposed to the situation when the same two acidic regions are located on a single DNA-bound molecule. Thus, transcriptional synergy does not depend on the number of aciaic activation regions, but instead on the number of proteins bound to the promoter. Although the length, quality, and probably number of acidic domains contribute to the level of transcription, they do not appear to be responsible for the synergistic effects.

The above observation argues against the prevailing view that synergy reflects a nonlinear response of a "common target" to the number and/or quality of acidic activation regions. In addition, it seems that cooperative binding may not account for synergistic activation by GCN4. Such cooperativity has not been observed in DNA-binding experiments carried out in vitro (Oliviero and Struhl 1991), and it is very likely that single binding sites are frequently occupied by GCN4 in vivo (Hope and Struhl 1986; Brandl and Struhl 1990), thus making it unlikely that cooperative binding increases promoter occupancy by a factor of 10.

The alternative explanation is that the DNA-binding domain plays a more direct role in transcription than simply targeting the protein to the promoter. This idea is supported by the existence of glucocorticoid receptor or HAP1 protein derivatives that bind DNA normally but fail to activate transcription (Kim and Guarente 1989; Schena et al. 1989) and is consistent with the interaction in vitro between the GCN4 DNA-binding domain and RNA polymerase II (Brandl and Struhl 1989). In specific models of this type, the DNA-binding domain might alter DNA structure,

affect nucleosome distribution on the chromatin template (possibly to increase access of the general transcription factors and/or RNA polymerase), or interact with the basic RNA polymerase II transcriptional machinery either directly or indirectly through an adapter protein(s). By any of these models, the acidic activation region presumably would carry out a different function from the DNA-binding domain in the overall process of transcriptional enhancement.

REFERENCES

Agre, P., P.F. Johnson, and S.L. McKnight. 1989. Cognate DNA binding specificity retained after leucine zipper exchange between GCN4 and C/EBP. *Science* **246:** 922–926.

Arndt, K. and G. Fink. 1986. GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5′TGACTC 3′sequences. *Proc. Natl. Acad. Sci.* **83:** 8516–8520.

Arndt, K.T., C. Styles, and G.R. Fink. 1987. Multiple global regulators control *HIS4* transcription in yeast. *Science* **237:** 874–880.

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:** 1386–1392.

Brandl, C.J. and K. Struhl. 1989. Yeast GCN4 transcriptional activator protein interacts with RNA polymerase II *in vitro*. *Proc. Natl. Acad. Sci.* **86:** 2652–2656.

———. 1990. A nucleosome-positioning sequence is required for GCN4 to activate transcription in the absence of a TATA element. *Mol. Cell. Biol.* **10:** 4256–4265.

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

Bushman, F.D. and M. Ptashne. 1988. Turning λ Cro into a transcriptional activator. *Cell* **54:** 191–197.

Carey, M., Y.-S. Lin, M.R. Green, and M. Ptashne. 1990. A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives. *Nature* **345:** 361–364.

Chasman, D.I., N.F. Lue, A.R. Buchman, J.W. LaPointe, Y. Lorch, and R.D. Kornberg. 1990. A yeast protein that influences the chromatin structure of UAS_G and functions as a powerful auxiliary gene activator. *Genes Dev.* **4:** 503–514.

Chen, W. and K. Struhl. 1988. Saturation mutagenesis of a yeast *his3* TATA element: Genetic evidence for a specific TATA-binding protein. *Proc. Natl. Acad. Sci.* **85:** 2691–2695.

———. 1989. Yeast upstream activator protein GCN4 can stimulate transcription when its binding site replaces the TATA element. *EMBO J.* **8:** 261–268.

Chen, W., S. Tabor, and K. Struhl. 1987. Distinguishing between mechanisms of eukaryotic transcriptional activation with bacteriophage T7 RNA polymerase. *Cell* **50:** 1047–1055.

Curran, T. and B.J. Franza. 1988. Fos and Jun: The AP-1 connection. *Cell* **55:** 395–397.

Driever, W., G. Thoma, and C. Nüsslein-Volhard. 1989. Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding sites for the bicoid morphogen. *Nature* **340:** 363–367.

Falvey, E. and N.D.F. Grindley. 1987. Contacts between γδ resolvase and the γδ *res* site. *EMBO J.* **6:** 815–821.

Fedor, M.J., N.F. Lue, and R.D. Kornberg. 1988. Statistical positioning of nucleosomes by specific protein binding to an upstream activating sequence in yeast. *J. Mol. Biol.* **204:** 109–127.

Gartenberg, M.R., C. Ampe, T.A. Steitz, and D.M. Crothers. 1990. Molecular characterization of the GCN4-DNA complex. *Proc. Natl. Acad. Sci.* **87:** 6034–6038.

Gill, G. and M. Ptashne. 1988. Negative effect of the transcriptional activator GAL4. *Nature* **334:** 721–723.

Giniger, E. and M. Ptashne. 1987. Transcription in yeast activated by a putative amphipathic α helix linked to a DNA binding unit. *Nature* **330:** 670–672.

Glass, C.K., J.M. Holloway, O.V. Devary, and M.G. Rosenfeld. 1988. The thyroid hormone receptor binds with opposite transcriptional effects to a common sequence motif in thyroid hormone and estrogen response elements. *Cell* **54:** 313–323.

Gonzalez, G.A., K.K. Yamamoto, W.H. Fischer, D. Karr, P. Menzel, W. Biggs, W.W. Vale, and M.R. Montminy. 1989. A cluster of phosphorylation sites on the cyclic-AMP regulated nuclear factor CREB predicted by its sequence. *Nature* **337:** 749–752.

Hai, T., F. Liu, W.J. Coukos, and M.R. Green. 1989. Transcription factor ATF cDNA clones: An extensive family of leucine zipper proteins able to selectively form DNA-binding heterodimers. *Genes Dev.* **3:** 2083–2090.

Hai, T., F. Liu, E.A. Allegretto, M. Karin, and M.R. Green. 1988. A family of immunologically related transcription factors that includes multiple forms of ATF and AP-1. *Genes Dev.* **2:** 1216–1226.

Harbury, P.A.B. and K. Struhl. 1989. Functional distinctions between yeast TATA elements. *Mol. Cell. Biol.* **9:** 5298–5304.

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:** 451–457.

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

———. 1988. Mechanisms of gene regulation in the general control of amino acid biosynthesis in *Saccharomyces cerevisiae. Microbiol. Rev.* **52:** 248–273.

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

Hoeffler, J.P., T.E. Meyer, Y. Yun, J.L. Jameson, and J.F. Haebner. 1988. Cyclic AMP-responsive DNA-binding protein: Structure based on a cloned placental cDNA. *Science* **242:** 1430–1433.

Homa, F.L., J.C. Glorioso, and M. Levine. 1988. A specific 15-bp TATA box promoter element is required for expression of a herpes simplex virus type I late gene. *Genes Dev.* **2:** 40–53.

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:** 177–188.

———. 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* **46:** 885–894.

———. 1987. GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. *EMBO J.* **6:** 2781–2784.

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:** 635–640.

Horikoshi, M., M.F. Carey, H. Kakidani, and R.G. Roeder. 1988a. Mechanism of action of a yeast activator: Direct effect of GAL4 derivatives on mammalian TFIID-promoter interactions. *Cell* **54:** 665–669.

Horikoshi, M., T. Hai, Y.-S. Lin, M. Green, and R.G. Roeder. 1988b. Transcription factor ATF interacts with the TATA factor to facilitate establishment of a preinitiation complex. *Cell* **54:** 1033–1042.

Hu, J.C., E.K. O'Shea, P.S. Kim, and R.T. Sauer. 1990. Sequence requirements for coiled-coils: Analysis of repressor-GCN4 leucine zipper fusions. *Science* **250:** 1400–1403.

Ingles, C.J., M. Shales, W.D. Cress, S.J. Triezenberg, and J. Greenblatt. 1991. Reduced binding of TFIID to transcriptionally compromised mutants of VP16. *Nature* **351:** 588–590.

Irwin, N. and M. Ptashne. 1987. Mutants of the catabolite activator protein of *Escherichia coli* that are specifically deficient in the gene activation function. *Proc. Natl. Acad. Sci.* **84:** 8315–8319.

Kakidani, H. and M. Ptashne. 1988. GAL4 activates gene expression in mammalian cells. *Cell* **52:** 161–167.

Kim, K.S. and L. Guarente. 1989. Mutations that alter transcriptional activation but not DNA binding in the zinc finger of yeast activator HAP1. *Nature* **342:** 200–203.

Kouzarides, T. and E. Ziff. 1989. Leucine zippers of *fos*, *jun*, and GCN4 dictate dimerization specificity and thereby control DNA binding. *Nature* **340:** 568–571.

Landschulz, W.H., P.F. Johnson, and S.L. McKnight. 1988. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science* **240:** 1759–1764.

Lin, Y.-S. and M.R. Green. 1991. Mechanism of action of an acidic transcriptional activator *in vitro*. *Cell* **64:** 971–981.

Lin, Y.-S., M. Carey, M. Ptashne, and M.R. Green. 1990. How different eukaryotic transcriptional activators can cooperate promiscuously. *Nature* **345:** 359–360.

Ma, J. and M. Ptashne. 1987a. Deletion analysis of GAL4 defines two transcriptional activating segments. *Cell* **48:** 847–853.

———. 1987b. A new class of yeast transcriptional activators. *Cell* **51:** 113–119.

Mahadevan, S. and K. Struhl. 1990. T_C, an unusual promoter element required for constitutive transcription of the yeast *his3* gene. *Mol. Cell. Biol.* **10:** 4447–4455.

Murre, C., P.S. McCaw, H. Vaessin, M. Caudy, L.Y. Jan, Y.N. Jan, C.V. Cabrera, J.N. Buskin, S.D. Hauschka, A.B. Lassar, H. Weintraub, and D. Baltimore. 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58:** 537–544.

Oakley, M.G. and P.B. Dervan. 1990. Structural motif of the GCN4 DNA binding domain characterized by affinity cleaving. *Science* **248:** 847–850.

Oliphant, A.R., C.J. Brandl, and K. Struhl. 1989. Defining sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: Analysis of the yeast GCN4 protein. *Mol. Cell. Biol.* **9:** 2944–2949.

Oliviero, S. and K. Struhl. 1991. Synergistic transcriptional enhancement does not depend on the number of acidic activation domains bound to the promoter. *Proc. Natl. Acad. Sci.* **88:** 224–228.

O'Neil, K.T., R.H. Hoess, and W.F. DeGrado. 1990. Design of DNA-binding peptides based on the leucine zipper motif. *Science* **249:** 774–778.

O'Shea, E.K., R. Rutkowski, and P.S. Kim. 1989a. Evidence that the leucine zipper is a coiled coil. *Science* **243:** 538–542.

O'Shea, E.K., R. Rutkowski, W.F.I. Stafford, and P.S. Kim. 1989b. Preferential

heterodimer formation by isolated leucine zippers from Fos and Jun. *Science* **245:** 646–648.

Pellman, D., M.E. McLaughlin, and G.R. Fink. 1990. Function of the TATA element at *HIS4*. *Nature* **348:** 82–86.

Ponticelli, A.S. and K. Struhl. 1990. Analysis of yeast *his3* transcription *in vitro*: Biochemical support for multiple mechanisms of transcription. *Mol. Cell. Biol.* **10:** 2832–2839.

Ptashne, M. 1988. How eukaryotic transcriptional activators work. *Nature* **335:** 683–689.

Pu, W.T. and K. Struhl. 1991a. The leucine zipper symmetrically positions the adjacent basic regions for specific binding to DNA. *Proc. Natl. Acad. Sci.* **88:** 6901–6905.

———. 1991b. Highly conserved residues in the bZIP domain of yeast GCN4 are not essential for DNA-binding. *Mol. Cell. Biol.* **11:** 4918–4926.

———. 1992. Uracil interference, a rapid and general method for defining protein-DNA interactions involving the 5-methyl group of thymines: The GCN4-DNA complex. *Nucleic Acids Res.* **20:** 771–775.

Rasmussen, R., D. Be[illegible]gnu, E.K. O'Shea, P.S. Kim, and T. Alber. 1991. X-ray scattering indicates that the leucine zipper is a coiled coil. *Proc. Natl. Acad. Sci.* **88:** 561–564.

Richardson, J.S. and D.C. Richardson. 1988. Amino acid preferences for specific locations at ends of α helices. *Science* **240:** 1648–1652.

Roesler, W.J., G.R. Vandenbark, and R.W. Hanson. 1988. Cyclic AMP and the induction of eukaryotic gene transcription. *J. Biol. Chem.* **263:** 9063–9066.

Sadler, J.R., H. Sasmor, and J.L. Betz. 1983. A perfectly symmetric *lac* operator binds the *lac* repressor very tightly. *Proc. Natl. Acad. Sci.* **80:** 6785–6789.

Sauer, R.T., D.L. Smith, and A.D. Johnson. 1988. Flexibility of the yeast α2 repressor enables it to occupy the ends of its operator, leaving the center free. *Genes Dev.* **2:** 807–816.

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:** 165–175.

Schena, M. and K.R. Yamamoto. 1988. Mammalian glucocorticoid receptor derivatives enhance transcription in yeast. *Science* **241:** 965–967.

Schena, M., L.P. Freedman, and K.R. Yamamoto. 1989. Mutations in the glucocorticoid receptor zinc finger region that distinguish interdigitated DNA binding and transcriptional enhancement activities. *Genes Dev.* **3:** 1590–1601.

Schmid, W., U. Strahle, G. Schutz, J. Schmitt, and H. Stunnenberg. 1989. Glucocorticoid receptor binds cooperatively to adjacent recognition sites. *EMBO J.* **8:** 2257–2263.

Sellers, J.W. and K. Struhl. 1989. Changing fos oncoprotein to a DNA-binding protein with GCN4 dimerization specificity by swapping "leucine zippers." *Nature* **341:** 74–76.

Sellers, J.W., A.C. Vincent, and K. Struhl. 1990. Mutations that define the optimal half-site for binding yeast GCN4 activator protein and identify an ATF/CREB-like repressor that recognizes similar DNA sites. *Mol. Cell. Biol.* **10:** 5077–5086.

Shuman, J.D., C.R. Vinson, and S.L. McKnight. 1990. Evidence of changes in protease sensitivity and subunit exchange rate on DNA binding by C/EBP. *Science* **249:** 771–774.

Simon, M.C., T.M. Fisch, B.J. Benecke, J.R. Nevins, and N. Heintz. 1988. Definition of multiple, functionally distinct TATA elements, one of which is a target in the *hsp70* promoter for E1A regulation. *Cell* **52:** 723–729.

Stringer, K.F., C.J. Ingles, and J. Greenblatt. 1990. Direct and selective binding of an

acidic transcriptional activation domain to the TATA-box factor TFIID. *Nature* **345:** 783–786.

Struhl, K. 1986. Constitutive and inducible *Saccharomyces cerevisiae* promoters: Evidence for two distinct molecular mechanisms. *Mol. Cell. Biol.* **6:** 3847–3853.

———. 1987a. The DNA-binding domains of the *jun* oncoprotein and the yeast GCN4 transcriptional activator are functionally homologous. *Cell* **50:** 841–846.

———. 1987b. Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. *Cell* **49:** 295–297.

———. 1988. The JUN oncoprotein, a vertebrate transcription factor, activates transcription in yeast. *Nature* **332:** 649–650.

———. 1989. Helix-turn-helix, zinc finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. *Trends Biochem. Sci.* **14:** 137–140.

———. 1991. Mechanisms for diversity in gene expression patterns. *Neuron* **7:** 177–181.

Struhl, G., K. Struhl, and P.M. Macdonald. 1989. The gradient morphogen *bicoid* is concentration-dependent transcriptional activator. *Cell* **57:** 1259–1273.

Talanian, R.V., C.J. McKnight, and P.S. Kim. 1990. Sequence-specific DNA binding by a short peptide dimer. *Science* **249:** 769–771.

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

Tsai, S.Y., M.-J. Tsai, and B.W. O'Malley. 1989. Cooperative binding of steroid hormone receptors contributes to transcriptional synergism at target enhancer elements. *Cell* **57:** 443–448.

Tzamarias, D., W.T. Pu, and K. Struhl. 1992. Mutations in the bZIP domain of yeast GCN4 that alter DNA-binding specificity. *Proc. Natl. Acad. Sci.* **89:** 2007–2011.

Umesono, K. and R.M. Evans. 1989. Determinants of target gene specificity for steroid/thyroid hormone receptors. *Cell* **57:** 1139–1146.

vanHeeckeren, W.J., J.W. Sellers, and K. Struhl. 1992. Role of the conserved leucines in the leucine zipper dimerization motif of yeast GCN4. *Nucleic Acids Res.* **20:** 3721–3724.

Vinson, C.R., P.B. Sigler, and S.L. McKnight. 1989. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* **246:** 911–916.

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:** 3316–3319.

Webster, N., J.R. Jin, S. Green, M. Hollis, and M. Chambon. 1988. The yeast UAS_G is a transcriptional enhancer in human HeLa cells in the presence of the GAL4 *trans*-activator. *Cell* **52:** 169–178.

Weiss, M.A., T. Ellenberger, C.R. Wobbe, J.P. Lee, S.C. Harrison, and K. Struhl. 1990. Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature* **347:** 575–578.

Wobbe, C.R. and K. Struhl. 1990. Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription *in vitro*. *Mol. Cell. Biol.* **10:** 3859–3867.