Annu. Rev. Biochem. 1989. 58:1051–77


# MOLECULAR MECHANISMS OF TRANSCRIPTIONAL REGULATION IN YEAST

*Kevin Struhl*

Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02115



## CONTENTS



## PERSPECTIVES AND SUMMARY

Regulation of gene expression is a fundamental aspect of many biological phenomena such as cell growth, morphology, the development of multicellular organisms, the response to environmental conditions, and disease. The processes of decoding genes and synthesizing appropriate amounts of gene products are complex, and regulation can occur at one or more of the



0066-4154/89/0701-1051$02.00

various steps along the pathway. Nevertheless, a major point of gene control occurs at the first step, namely the initiation of messenger RNA synthesis.

Over the past several years, many observations have increasingly pointed to common molecular mechanisms of transcriptional regulation in eukaryotic organisms ranging from humans to yeasts (for previous reviews, see 1–4). The subunit structure and catalytic properties of RNA polymerase II as well as the posttranscriptional modifications of the primary transcripts are conserved throughout the eukaryotic kingdom. A sequence resembling TATAAA is found in most eukaryotic promoters at a position near the mRNA initiation site. Genetic analyses of yeast and higher eukaryotic promoters have revealed enhancer or UAS elements that activate transcription when located at long distances from the initiation site. The largest subunit of RNA polymerase II contains an unusual repeating heptapeptide tail at the C-terminus and is otherwise highly conserved (5, 6). More recently, it has been shown that yeast and higher eukaryotic cells contain structurally similar and functionally analogous transcription factors that recognize essentially identical sequences (7–14). Moreover, yeast proteins can activate transcription in a wide variety of eukaryotic organisms (15–18), and vertebrate proteins can stimulate transcription in yeast cells (19–22). Such functional interchangeability makes it possible to study molecular mechanisms in vivo and in vitro by using mixtures of yeast and mammalian components.

Much of our knowledge of the molecular mechanisms involved in transcriptional regulation has come from studies of the baker's yeast *Saccharomyces cerevisiae*. The relative simplicity and rapid growth rate of the organism, the availability of powerful genetic selections and screens to identify important genes and gene products and to obtain mutant strains, the ease of cloning essentially any gene, and the ability to alter the genome at will by performing exact replacements of normal chromosomal sequences with mutated derivatives constructed in vitro have been crucial for the relatively advanced state of knowledge. At present, many yeast promoters have been analyzed in detail, and a number of specific transcription factors have been identified and characterized for their DNA-binding and transcriptional activation functions. In some cases, the regulatory circuits for altering the level of transcription under appropriate conditions have been outlined.

## YEAST PROMOTER ELEMENTS

The yeast genome specifies approximately 5000 protein-coding genes that are densely clustered on 16 linear chromosomes. The average yeast gene is transcribed about 5–10 times during each cell cycle, which results in a steady-state level of 1–2 mRNA molecules/cell. Although many genes are transcribed constitutively at the average level, RNA levels of different genes

vary over 2–3 orders of magnitude. Moreover, some genes are transcribed at variable rates depending on the physiological conditions, and groups of genes are regulated coordinately. Despite the different transcriptional regulatory properties, essentially all yeast promoters contain three basic kinds of DNA sequence elements (Figure 1). The properties of these upstream (UAS), TATA, and initiation (I) elements have been reviewed previously (1–3) and are briefly summarized below.

Upstream elements are short DNA sequences, typically 10–30 bp in length, that are located relatively far from the mRNA start site and are required for transcription. Depending on the gene, upstream elements can be located anywhere from 100 to 1500 bp upstream of the initiation site. For genes whose transcription rates vary according to the physiological conditions, the upstream element is usually the major determinant of the particular regulatory properties of a given promoter. Genes subject to a common control mechanism contain upstream elements that are similar in DNA sequence, whereas noncoordinately regulated genes contain upstream elements with different DNA sequences. Upstream elements are analogous to mammalian enhancer sequences in that they function in both orientations and at long and variable distances with respect to other promoter elements and the mRNA initiation site. In contrast to enhancers, UAS elements do not activate transcription when located downstream of the mRNA initiation site (24, 25).

In addition to upstream elements, TATA elements are necessary for transcriptional initiation of most yeast genes; deletion of these elements greatly reduces the mRNA level (26–29). TATA elements, which have been historically defined by their similarity to the sequence TATAAA, are typically located anywhere between 40 and 120 bp upstream of the mRNA initiation

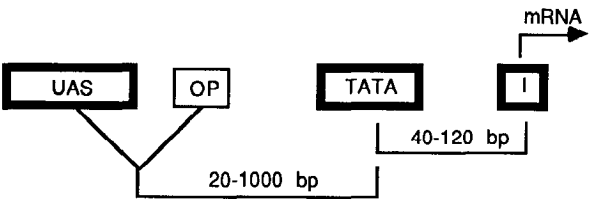


*Figure 1* *cis*-acting elements for a simple yeast promoter.

The initiator element (I) is important for determining where transcription begins (arrow). The TATA element, located 40 to 120 base pairs upstream from the initiation site, is required for transcription. The upstream element (UAS), which is located at variable and sometimes very long distances from the other elements, is important for transcription and usually determines the particular regulatory properties of the promoter. Some promoters contain operator elements (OP), which mediate negative control of transcription; operators are generally located at variable distances upstream of TATA elements, but their position with respect to UAS elements depends on the promoter. See text for details.

sites. This variability in location differs from the situation in most other eukaryotes, in which TATA sequences are almost always located 25–30 bp away from the initiation site. The presence of the conserved TATAAA sequence in many different kinds of promoters has led to the view that TATA elements have a general role in the transcription process much like RNA polymerase. However, more recent observations have argued against this view and instead have suggested the existence of functionally distinct classes of TATA elements.

The initiator element, located near the actual mRNA start site, is the primary determinant of where transcription begins. In yeast, accurate initiation is still observed when the distance to the TATA or upstream element is varied or when transcription depends on "foreign" promoter elements located at different positions from the natural elements (29–32). In contrast, selection of higher eukaryotic mRNA start sites is determined not by specific DNA sequences but rather by distance (25–30 bp) from the TATA element. The spacing between yeast initiation sites and TATA elements is much more flexible and somewhat larger, although there are limits to the distance over which a TATA element can act (roughly 40–120 bp). However, the nonstringent spacing requirements account for why some yeast genes have many more initiation sites than TATA elements. The DNA sequence requirements for yeast initiator elements are poorly understood, although it is clear that many different sequences can carry out the function. Unlike upstream or TATA elements, the initiator element is relatively unimportant for determining the rate of transcriptional initiation.

Some yeast promoters contain a fourth class of element, the operator, that represses the level of transcriptional initiation (reviewed in 33). The properties of yeast operators are similar to those of upstream elements in that they are short sequences that function bidirectionally and at variable distances upstream of TATA elements. Like upstream elements, similar operator sequences in different promoters provide the basic mechanism for coordinate regulation of transcription (34, 35). Operators can repress transcription when located upstream of upstream elements, but repression is generally much more efficient when the operator lies between the upstream and TATA element (28, 34, 36). One major exception to this rule is the mating-type silencer that efficiently represses transcription when located at distances as far as 2 kb upstream or downstream from the mRNA initiation sites (37–39).

## SPECIFIC DNA-BINDING PROTEINS

It is almost an axiom of molecular biology that the function of promoter DNA sequences is to act as targets for the specific DNA-binding proteins that actually activate or repress transcription. As might be expected from the wide

variety of upstream promoter elements, yeast cells contain many specific transcription factors that recognize different DNA sequences. Unlike higher eukaryotic factors that were identified by direct DNA-binding assays, yeast regulatory proteins were first identified by mutations that alter the transcription of a specific gene or set of genes. Using various experimental procedures, it has been shown in a number of cases that these mutations define genes encoding specific DNA-binding proteins. The genetic approach is particularly valuable because the mutant strains provide direct information about the function(s) of the DNA-binding protein in vivo and facilitate the cloning of the genes.

The availability of the cloned genes has allowed for a detailed analysis of the structures and functions of these yeast DNA-binding proteins. For several cases, analyses of deletion proteins indicate that small autonomous domains containing less than 100 amino acid residues are sufficient for specific DNA-binding activity. Although detailed structural information is not yet available, it is clear from the primary sequences and biochemical properties of the DNA-binding domains that different structural motifs are employed for DNA sequence recognition. These motifs include the helix-turn-helix of bacterial repressors and activators (40), the presumptive zinc finger (41), the putative leucine zipper (42), and possibly others that have yet to be described. Studies of yeast DNA-binding proteins are most advanced for the GAL4, GCN4, MAT$\alpha$2, and HAP1 proteins.

GAL4 The GAL4 protein (881 amino acids) is required for transcription of the *GAL* (galactose metabolizing) genes that occurs in galactose medium. GAL4 binds to four sites within the upstream regulatory region required for bidirectional activation of the *GAL1* and *GAL10* genes as well as to sites within several other promoters (43–45). Sequence comparison of the binding sites indicates that a 17-bp sequence of imperfect dyad symmetry is important for recognition by GAL4. In vivo footprinting experiments indicate that in galactose medium the GAL4-binding sites in the genome are occupied most if not all of the time (44, 46), and in certain cases it appears that GAL4 can bind cooperatively to adjacent sites (47). Interestingly, GAL4 is bound to its genomic sites even when *GAL* transcription is not induced (glycerol medium), although it is not bound in glucose medium when *GAL* transcription is catabolite repressed (44, 46).

The GAL4 DNA-binding domain is localized to the N-terminal 73 amino acids (48), and it contains two pairs of cysteine residue sequences that resemble coordination sites for zinc ions (41, 50). An important role for zinc is inferred from gal4 mutants that are functional only in the presence of high concentrations of zinc ions (50, 51). In addition, mutations of the cysteine residues abolish DNA-binding activity (51–53).

The yeast ARGR2 and PPR1 activator proteins contain amino acid sequences that resemble the region of GAL4 that contains the pairs of cysteine residues even though these proteins recognize very different DNA sequences (54, 55). In contrast, LAC9, a protein from the related yeast *Kluveromyces lactis* that can functionally substitute for GAL4 in *S. cerevisiae* cells, has an additional sequence in its DNA-binding domain that is very similar to that found in GAL4. Thus, it is likely that this additional sequence is involved in recognition of the DNA target, whereas the region that includes the cysteine pairs is more involved in maintaining overall structure and perhaps nonspecific interactions to DNA. Consistent with this hypothesis, a synthetic peptide containing the zinc finger of ADR1 can form a discrete structure in the presence of zinc that binds DNA nonspecifically, but it fails to recognize its normal target sequence (56).

GCN4 GCN4 protein (281 amino acids) binds specifically to the promoter regions of many amino acid biosynthetic genes and induces their transcription in response to amino acid starvation (57, 58). Saturation point mutagenesis of the *HIS3* regulatory site indicates that GCN4 recognizes a 9-bp region with optimal binding to the dyad-symmetric sequence ATGA(C/G)TCAT (59). The optimal binding site binds to GCN4 with higher affinity than the native *HIS3* site, and it induces transcription to higher levels, thus suggesting that GCN4 protein levels are limiting in vivo. The optimal GCN4 binding site also represents the consensus of presumptive recognition sequences from 15 genes subject to coordinate induction by GCN4 (59). Interestingly, none of these naturally occurring sites are identical to the consensus sequence but instead have 1–2 deviations. Thus, it appears that yeast cells maintain a fine balance between the level of GCN4 protein and the affinities of the many binding sites in the genome such that transcription of the coregulated genes can be induced even though GCN4 does not fully occupy the sites.

The dyad-symmetric nature of the optimal GCN4 binding site strongly suggests that GCN4 binds as a dimer to adjacent half sites. GCN4 does indeed bind as a dimer because synthetic mixtures of wild-type and deleted GCN4 proteins yield functional heterodimers at the expected frequency (60). The GCN4 recognition site is unusual because it is very short and because mutation of the central C:G base pair, even to the symmetric G:C counterpart, significantly reduces binding to DNA. This suggests that the central base pair is part of the half site recognized by a GCN4 monomer and consequently that the half sites overlap (60). Moreover, it follows that the optimal binding site must contain two nonequivalent half sites, ATGAC and ATGAG. Interestingly, GCN4 binds well to a 10-bp site containing adjacent ATGAC half sites, whereas it fails to bind to a 10-bp site containing adjacent ATGAG half sites (J. W. Sellers, and K. Struhl, unpublished). Thus, GCN4 binds to

overlapping half sites whose optimal sequence is ATGAC, and the protein is surprisingly flexible in that it can bind DNA even when a base pair is inserted at the middle of the recognition sequence.

Deletion analysis of GCN4 reveals that the 60 C-terminal amino acids are fully sufficient for specific DNA-binding and for dimerization (60, 61). Strikingly, the GCN4 DNA-binding domain shows 45% amino acid identity to the C-terminal region of the jun oncoprotein that causes fibrosarcomas in chickens (62). Moreover, the GCN4 DNA-binding domain can be functionally replaced by the homologous jun region to generate a protein that possesses DNA-binding properties indistinguishable from those of GCN4 on a series of target sites (7). The fact that GCN4 and jun recognize identical DNA sequences suggests that the amino acid residues involved in direct contacts to DNA are located within the 30-residue stretch that is most highly conserved.

Recently, it has been proposed that GCN4 belongs to a class of proteins that utilizes a new motif for DNA binding, the leucine zipper (42). These proteins, which include the jun, fos, and myc oncoproteins and the C/EBP enhancer–binding protein, all contain 4–5 leucine residues that could be viewed as being periodically repeated every two turns of an $\alpha$ helix. The hypothetical model is that the leucine residues are important for interdigitating two $\alpha$ helixes, one from each monomer unit, that provide the structural basis for the dimer formation. However, a variety of mutations of the conserved leucines have little if any effect on DNA binding (J. Sellers, K. Struhl, unpublished).

HAP1 The HAP1 protein binds to the upstream regulatory elements of two cytochrome *c* genes, *CYC1* and *CYC7*, and induces their transcription (63, 64). Specific DNA-binding in vitro and transcriptional induction in vivo are stimulated by heme. Although both binding sites compete for HAP1 binding, activation through the *CYC1* site is about 10-fold more efficient. A derivative of the protein, HAP1-18, appears to have an altered specificity because it fails to bind the *CYC1* site while retaining the ability to bind the *CYC7* site. Strains containing HAP1-18 show increased levels of *CYC7* expression and decreased levels of *CYC1* expression.

Surprisingly, the *CYC1* and *CYC7* binding sites have no obvious sequence similarity, thus leading to the suggestion that HAP1 recognizes two qualitatively different DNA sequences, possibly by utilizing a single DNA-binding domain (64). However, in accord with proposed sequence relationships between these two sites (65), a particular base-pair substitution of the *CYC7* site generates a new site with *CYC1*-like properties; high expression in the presence of HAP1 and low expression dependent on HAP1-18 (66). The possibility that the *CYC1* and *CYC7* sites might be related, though divergent, forms of the same sequence has not been eliminated.

MATα2 The MATα2 protein (210 amino acids) regulates yeast cell type by binding specifically to operator sequences of several **a**-specific genes and repressing their transcription (34). The α2 operators are unusual in that they are very large (roughly 30 bp), having highly conserved sequences at both ends with an approximate twofold symmetry (CATGTAA) and lacking sequence similarities in the middle. In accord with this unusual arrangement, α2 protects the two end portions from DNase I, dimethylsulfate, and hydroxyl radical, while leaving the middle accessible to attack by either of these reagents (67). A 7-bp substitution in the center of the operator has little effect on repressor binding. These observations suggest that α2 binds as a dimer to half sites whose centers are two and one-half helical turns apart and does not contact the middle of the operator. Surprisingly, α2 also binds with similar affinity to an operator derivative with the central 13 bp deleted such that the half sites are immediately adjacent. Thus, α2 is remarkably flexible in that it can bind to differently spaced operator half sites even when located on the same or opposite side of the DNA helix (67).

Proteolytic fragmentation of α2 indicates that the protein is composed of two independently folding structural domains (67). The 79-residue C-terminal fragment interacts specifically with operator DNA although with a reduced affinity that is characteristic of half-site binding. The 102-residue N-terminal segment does not bind DNA itself, but instead is necessary for the dimerization that permits the protein to bind simultaneously to both half sites at high affinity. The purified α2 dimers are very unusual in that the monomer subunits are covalently linked through the disulfide bonds between cysteine residues in the N-terminal region. Although α2 dimers might form in vivo by such a mechanism, it now appears more likely that the cysteine residues serve as zinc-binding sites that hold the monomer subunits together (67).

Interestingly, the DNA-binding domain of α2, a protein that specifies cell type in yeast, shows strong primary sequence similarities to the homeobox motif of *Drosophila* proteins that selects the choice of developmental pathway (68). Amino acid substitutions within the α2 homeo-like region abolish function, presumably by virtue of a loss of DNA-binding activity (69). These proteins also show some sequence similarity to the helix-turn-helix motif of bacterial repressors and activators (70), suggesting that this structure may be involved in specific DNA-binding by α2.

TATA-BINDING PROTEIN Unlike the above activator and repressor proteins that affect relatively few genes of related function, a protein binding to TATA elements should be important for the transcription of many genes. As the genetic approach is restricted to transcription factors that are nonessential for growth and can be mutated to produce simple phenotypes, it is not useful for identifying a TATA-binding protein. Initial evidence for a TATA-binding

activity was obtained by photofootprinting in vivo, which revealed physical changes in the *GAL1,10* TATA elements only under conditions when the genes were transcriptionally active (71).

More recently, such a protein has been identified and purified by its ability to substitute for the mammalian TATA-binding transcription factor (TFIID) in a reconstituted in vitro transcription system (13, 14). The yeast factor binds to two of the three genetically identified TATA elements in the *CYC1* promoter, is surprisingly small (25 kd by sedimentation and gel filtration), and appears to be a basic protein (13). Interestingly, transcription stimulated in vitro by the yeast TATA-binding protein is initiated 30 bp downstream from a *CYC1* or a mammalian TATA element, a distance typical of mammalian promoters (13). This observation suggests that the difference between yeast and mammalian promoters regarding the spacing between the TATA element and mRNA initiation site is due not to the TATA-binding proteins, but rather to differences in the basic transcription machinery that interacts with the TATA factor.

## TRANSCRIPTIONAL ACTIVATION

The binding of activator proteins to upstream promoter elements is necessary but not sufficient for stimulating transcription in vivo. GAL4 and GCN4 deletion proteins containing only the intact DNA-binding domain do not activate transcription, but instead can actually repress transcription when their recognition sites are located between a heterologous upstream element and TATA sequence (48, 61). Conversely, regions of GAL4 or GCN4 lacking their own DNA-binding domains can activate transcription when fused to a heterologous DNA-binding domain such as that of the *Escherichia coli* LexA repressor (61, 72). These LexA hybrid proteins activate a promoter having a LexA operator as the upstream element by binding via the LexA repressor-operator interaction and stimulate transcription utilizing the GAL4 and GCN4 activation function. Thus, DNA binding and transcriptional activation are distinct functions that are located in separate parts of the protein.

Deletion analyses have localized the GCN4 and GAL4 transcriptional activation functions to short regions that contain a relatively high proportion of acidic residues (61, 73). In both cases, surprisingly large portions of the protein-coding sequence can be removed without significantly affecting the transcriptional activation function. The GCN4 transcriptional activation function is localized to a short acidic region in the center of the protein, whereas GAL4 contains two separate activation regions. In the case of GCN4, derivatives retaining as few as 35–40 amino acids from the acidic region are sufficient for wild-type levels of transcriptional activation when fused directly to the DNA-binding domain (74). The distance and orientation of the GCN4

and GAL4 activation regions with respect to their DNA-binding domains is functionally unimportant. This indicates that the activation region encodes an autonomous function and that there is no requirement for a spacer between the activation region and DNA-binding domain.

Several lines of evidence indicate that yeast transcriptional activation functions are defined by short acidic regions with little sequence homology. First, different portions of the GCN4 acidic region are equally capable of activating transcription even though their primary sequences are dissimilar (61, 74). Second, the GCN4 and GAL4 transcriptional activation regions are all acidic but have no other noticeable sequence similarities (61, 73, 74). In this regard, other yeast activator proteins also have acidic regions that may be required for transcriptional stimulation in vivo. Third, acidic character is the common feature of functional transcriptional activation regions selected from short *E. coli* DNA segments (75). Fourth, single amino acid changes in a GAL4 derivative that increase or decrease the level of activation are almost always associated with an increased or decreased negative charge, respectively (76). These results strongly suggest that transcriptional activation regions are not defined by a specific primary sequence but rather a more general property such as net negative charge.

Despite the clear importance of acidic character, it appears that functional transcriptional activation regions must have additional structural features. An acidic 15-residue peptide whose sequence could form two turns of an $\alpha$ helix confers some transcriptional activity when fused to the GAL4 DNA-binding domain (77). However, a related peptide with the exact length and composition but a different sequence fails to stimulate transcription. Progressive fine-scale deletions of the GCN4 transcriptional activation region yields a series of small, stepwise reductions in activity (74). GCN4 activity appears directly related to the size of the transcription activation region remaining, whereas there is no such precise relationship of transcriptional activity to the number of acidic residues. These experiments strongly suggest that transcriptional activation regions do not have a defined tertiary structure such as found in active sites or domains in a protein, a view that is confirmed by proteolysis experiments involving GCN4 (74).

Nevertheless, 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. Independent evidence for some kind of structure comes from an unusual chymotrypsin digestion pattern observed for GCN4 and deleted derivatives that depends on the presence of a functional transcriptional activation region (74). A clue to the structure of the GCN4 activation region is that the boundaries defining the stepwise levels of activation may occur every seven amino acid residues, a repeat unit consistent with two turns of an $\alpha$ helix (74). The GCN4 activation

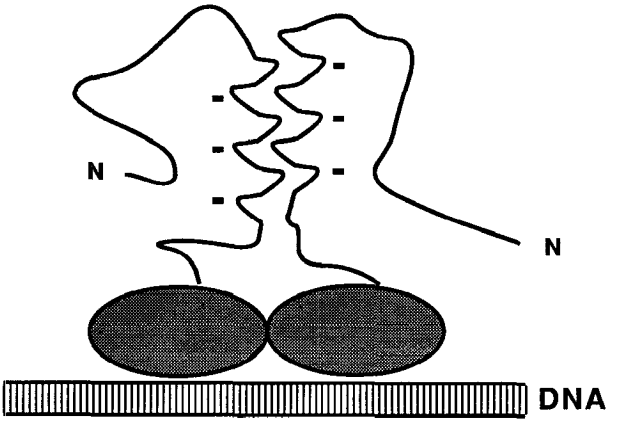


*Figure 2* Structural model for a typical transcriptional activator protein.
The regions of an activator protein include a dimeric DNA-binding domain (shaded ovals) interacting with a dyad symmetric sequence (arrows), a transcriptional activation region that is hypothesized to be a dimer of interacting $\alpha$ helixes (wavy lines) with acidic residues exposed (−), and a nonessential region shown as unstructured. See text.

region could form three amphipathic $\alpha$ helixes with acidic and hydrophobic residues tending to be clustered along separate surfaces. Similarly, the sequence of the synthetic peptide described above that functions as an activation region is consistent with forming two turns of an amphipathic helix (77). However, as the major GAL4 and other activation regions are unlikely to form amphipathic helixes, a simple relationship between this structure and function appears unlikely.

A structural model has been proposed in which the activation region is a dimer of intertwined $\alpha$ helixes, one helix from each monomer (74) (Figure 2). The formation/stability of this structure should be facilitated by the stability of dimeric DNA-binding domains. Moreover, as the LexA domain binds very poorly to its operator because of weak dimerization (78), it is likely that transcriptional stimulation through the LexA domain also requires that the activation region facilitate dimerization (7, 61, 75). By this model, amphipathic helixes could form functional activation regions as they would easily permit a structure involving interacting hydrophobic residues that are protected from solvent and exposed acidic residues.

The short acidic regions that are sufficient for activation are likely to be surfaces that are used for interactions with other proteins of the transcription machinery. It seems extremely unlikely that these short acidic regions of limited homology could encode catalytic activities such as topoisomerases, nucleases, methylases, etc that might be involved in transcription. Acidic regions of DNA-binding proteins are likely to be general requirements for

transcriptional activation in all eukaryotic organisms. GAL4 activates transcription from appropriate target promoters in mammalian cells, and in flies and plants; the acidic activation region is required in all cases (15–18). Conversely, transcriptional activation in yeast cells by the jun oncoprotein and the glucocorticoid receptor requires acidic sequences in addition to the DNA-binding domains (20, 22). The implication of these results is that the acidic activation regions contact some part of the transcription machinery that is conserved functionally throughout the eukaryotic kingdom. The obvious candidates for such an interaction are TATA-binding proteins, RNA polymerase II, or histones.

## REGULATION

In all forms of transcriptional regulation, there must be at least two distinct physiological conditions that are defined operationally by different RNA levels of a particular gene or set of genes. Furthermore, there must be a mechanism by which one of these physiological states can be converted to another state by some molecular signal. Such a signal can be caused either by a change in the external environment (the presence or the absence of a particular compound), or by an internal change governed by a particular developmental program. Although transcriptional control is executed by DNA-binding proteins, there must be additional regulatory molecules that govern when these regulatory proteins execute their roles in transcription. Thus, the DNA-binding proteins interact not only with specific nucleotide sequences, but also (directly or indirectly) with signal molecules that distinguish between the two physiological states.

Yeast cells use a variety of molecular mechanisms to regulate the activity of specific transcription factors. In one mechanism, the activity of a DNA-binding protein is altered by the binding of a small molecule. For example, HAP1 protein requires heme for efficient DNA binding in vitro and transcriptional activation in vivo (63). In another mechanism, the activity of a DNA-binding protein is affected by the binding of another protein. The GAL80 protein inhibits transcriptional activation by interacting directly with GAL4 such that it masks the primary acidic transcriptional activation region; it does not affect the ability of GAL4 to bind DNA (79–81). The GAL4-GAL80 interaction is eliminated when cells are grown in galactose medium, presumably by an interaction between GAL80 and galactose (or related metabolite). A third mechanism involves alteration by covalent modification of the DNA-binding protein. The heat-shock transcription factor is present under all growth conditions, but during heat shock it appears to be phosphorylated (82, 83). As the target genes are expressed only during heat-shock conditions, presumably the phosphorylated form represents the active transcription factor.

By analogy with the acidic transcriptional activation regions of GAL4 and GCN4, phosphorylation of the heat-shock factor might be required for activation simply by increasing the negative charge of the protein. Another potential example of phosphorylation involves the SNF1 protein kinase that is required for transcriptional activation by GAL4 and ADR2 (84); however it is not clear if the effects of SNF1 are direct.

In some cases, yeast cells regulate transcription by controlling the amount of a specific DNA-binding protein rather than by modulating the activity of the protein. GCN4 activates transcription only during amino acid starvation because it is not synthesized under normal growth conditions. During starvation, GCN4 protein levels increase 30–50-fold, while GCN4 mRNA levels remain relatively unchanged (85, 86). This translational regulation involves four small open reading frames present in the extremely large 5' untranslated leader of GCN4 mRNA (87, 88). As expected, the open reading frames prevent translation in normal circumstances, whereas it appears that the basic rules of translational initiation are circumvented under starvation. As GCN4 effectively regulates protein synthesis by controlling the amount of amino acid precursors, the unusual translational control mechanism is sensible because the effector protein is sensitive to the process it controls.

A different method for regulating transcription by controlling the amount of a DNA-binding protein is exemplified by the proteins that determine yeast mating type. Unlike GCN4, whose level changes in response to environmental conditions, the MAT$\alpha$2, MAT$\alpha$1, and MAT**a**1 proteins regulate transcription of their target genes by being present only in the relevant cell types (reviewed in 89). The segregation of transcriptional regulatory proteins into different cell types probably represents the major mechanism for developmental regulation of genes in multicellular organisms.

## MORE COMPLEX PHENOMENA

In previous sections, I discussed the basic components required for transcriptional regulation. For the simplest promoters, an upstream activator protein binds to its target site and stimulates transcription through the acidic activation region in a process that also requires the binding of a common TATA factor to its target site. Other promoters may contain multiple upstream, TATA, initiator, or operator elements that allow for distinct regulatory properties. Although some of these may be viewed as a set of independently acting simple promoters, it is becoming increasingly clear that there are additional complexities that represent new principles.

### *Multiple Proteins Recognizing the Same Target Sequence*

The existence of a protein distinct from GCN4 but recognizing similar DNA sequences was initially inferred from the observation that the GCN4 binding

site in the *HIS3* promoter activates low levels of transcription in the absence of GCN4 protein (90). Such a protein, yAP-1, was identified (10, 11) by virtue of its ability to bind a site recognized by AP-1, the mammalian transcription factor when DNA-binding domain is structurally and functionally related to GCN4 (8, 9). yAP-1 and GCN4 show about 35% sequence identity in their DNA-binding domains and they bind very similar sequences, although with an altered specificity (91). Genetic experiments indicate that GCN4 and yAP-1 activate transcription from promoters whose upstream elements contain related sequences. However, their in vivo roles must differ because yAP-1 cannot compensate for the defects caused by *gcn4* mutations. The role of yAP-1 is unclear because mutant strains lacking the protein have nearly normal growth properties (91). Perhaps yAP-1 is important for low-level, constitutive expression of some of the amino acid biosynthetic genes that are induced by GCN4 under starvation conditions.

Several lines of evidence suggest that there are additional yeast proteins with DNA-binding properties similar to those of GCN4 and yAP-1. First, proteins interacting with the AP-1 binding site are present in extracts lacking both GCN4 and yAP-1 (91). Second, a yeast protein(s) has been isolated by virtue of its ability to bind a sequence recognized by the mammalian ATF/CREB factor (92). This sequence is similar to the functional 10-bp GCN4-binding site that contains an additional G:C base pair in the middle of the site. Third, this 10-bp GCN4-binding site represses *HIS3* transcription in the absence of GCN4; in the presence of GCN4, *HIS3* transcription is partially induced probably due to competition between GCN4 and the putative repressor protein for the binding site (J. Sellers, K. Struhl, unpublished).

Another example involves an upstream element in the *CYC1* promoter that interacts both with HAP1, which induces transcription in response to heme, and with RC2, a distinct protein found in crude extracts (63, 93). The protein-DNA contacts for these two interactions are indistinguishable, suggesting that binding by these proteins is mutually exclusive (63). The role of RC2 in vivo is unclear.

The existence of multiple proteins that recognize related sequences increases the sensitivity and flexibility for coordinately and independently regulating subclasses of genes. Given the increase in complexity, it seems likely such families of proteins will be important for regulatory processes of basic importance for the organism. For example, the bacteriophage λ repressor and cro proteins recognize similar but not identical sequences that control the developmental decision between lysis and lysogeny (94). In the case of the GCN4-related proteins, one might imagine that it is crucial for yeast cells to precisely regulate amino acid biosynthesis due to its crucial role in protein synthesis and hence cell growth.

### *Multiple Proteins Necessary for a Single DNA-Binding Event*

It is generally assumed that a given DNA-binding protein is fully capable of interacting with its target site. However, there are instances in which two proteins together can bind DNA whereas neither protein can bind alone. Both the HAP2 and HAP3 proteins are required for binding to an upstream element in the *CYC1* promoter and are present in the final protein-DNA complex (95). The HAP2-HAP3 complex forms in the absence of DNA and can be purified to near homogeneity through a number of chromatographic steps (96). The HAP2-HAP3 complex has a remarkable functional relationship to the mammalian CP1 binding factor that itself is a complex between the CP-1A and CP-1B proteins (97). Both complexes recognize the same DNA sequences, and functional yeast-mammalian hybrid complexes can be created; HAP2 is functionally equivalent to CP-1B and HAP3 is equivalent to CP-1A (12). It is unknown whether binding by the heteromeric complex involves specific DNA contacts being mediated through HAP2, HAP3, or both together.

Two other examples can be found among the proteins that regulate mating type. In one case, regulatory sites in three $\alpha$-specific genes are bound by a combination of the MAT$\alpha$1 transcriptional activator and a second protein PTRF; neither protein binds alone (98). PRTF itself appears to bind some promoter elements whose consensus sequence is dyad symmetric. The $\alpha$-specific genes contain mutated and functionally defective versions of the PTRF element as well as an additional element termed Q (99). It has been suggested that $\alpha$1 and PTRF interact respectively with the Q and defective PTRF elements in the $\alpha$-specific genes, with both proteins being necessary for high-affinity binding (98, 99).

The second example involves the MAT$\alpha$2 and MAT**a**1 proteins that are both required to bind promoter sites in diploid-specific genes (100). The target sequence for this combination of proteins is large and includes a sequence resembling an $\alpha$2-binding site (35, 101). However, the **a**1-$\alpha$2 combination does not bind to a simple $\alpha$2 operator sequence, thus suggesting that **a**1 protein alters the binding specificity of $\alpha$2 repressor (100). Different portions of $\alpha$2 are required for the different binding activities because removal of the 62 N-terminal residues abolishes **a**1-$\alpha$2 binding, but does not effect simple $\alpha$2 binding (100, 102). It is likely the N-terminal region of $\alpha$2 is not required for specific contacts to DNA, but rather for protein-protein contacts to **a**1. Although many amino acid substitutions in the $\alpha$2 DNA-binding domain abolish both binding functions, some mutations specifically affect $\alpha$2 activities, whereas others specifically affect **a**1-$\alpha$2 activities (69, 103). It has been suggested that **a**1-$\alpha$2 might be a heterodimer that recognizes a different sequence from the presumptive homodimeric $\alpha$2 repressor (100).

The principle that the combined action of two proteins can be necessary for

DNA-binding increases the precision of transcriptional regulation. Such combinatorial regulation makes it possible to influence gene expression only when two specific physiological conditions occur; e.g. regulation of diploid-specific genes only in **a**/$\alpha$ cells or in complex cell-type regulation in higher organisms. Combinatorial DNA-binding might also increase the flexibility of regulation by having a given protein associate with a variety of different proteins to yield heteromeric species with distinct sequence recognition properties.

## *Multiple Proteins Necessary for Activation or Repression*

In the above examples where two proteins are necessary to bind a target sequence, it follows that both proteins are necessary for the transcriptional induction or repression that is mediated through the site. However, in the situations described below, it appears that two proteins are necessary for activation or repression even though each protein can bind independently and simultaneously to the target. As discussed for combinatorial DNA-binding, synergy between activators or repressors increases both the precision and flexibility of transcriptional regulation. In addition, synergy is economical because it minimizes the number of distinct transcription factors that are necessary to achieve the wide variety of regulatory responses.

Transcriptional activation mediated by the UAS1 element in the *CYC1* promoter occurs in the presence of heme and requires the HAP1 activator protein. However, more detailed analysis of UAS1 indicates that it is actually composed of two parts, only one of which is bound by HAP1 (104). Thus, HAP1 is necessary but not sufficient for activation of the *CYC1* promoter. The RAF1 protein identified in crude extracts that binds to the other part of UAS1 may also be required for this activation (63).

Binding of the MAT$\alpha$2 protein to operators of **a**-specific genes is necessary but not sufficient for transcriptional repression (105). As mentioned earlier, $\alpha$2 binds to the ends of the operator, leaving the middle free. Although the middle of the operator is not important for $\alpha$2 binding, it is crucial for repression in vivo. The GRM protein, which is present in all cell types, binds to the center of the operator even at the same time as when $\alpha$2 is bound at the ends. Moreover, GRM and $\alpha$2 bind cooperatively to the operator, and the N-terminal domain of $\alpha$2 is necessary for cooperative binding in vitro (105) and repression in vivo (102). These observations have suggested that simultaneous, and possibly cooperative, binding of GRM and $\alpha$2 is required for repression.

Another example of synergy is represented by the mating-type silencer that efficiently represses transcription when located at distances as far as 2 kb upstream or downstream from the mRNA initiation sites (37–39). Three distinct elements, two of which interact with known DNA-binding proteins,

are involved in silencer function. Surprisingly, no individual element is either necessary or sufficient for silencer function, but transcriptional repression occurs with any two of the elements (106). Presumably, any two of the proteins that bind these elements can act together to repress transcription, whereas no individual protein is capable.

There are three basic models to explain synergy. First, one protein that carries out the activation or repression might bind efficiently to DNA only in the presence of the other protein. Such cooperative binding would imply a very specific interaction between the two proteins, and hence seems unlikely to be a general explanation. Second, the proteins may form a relatively nonspecific interaction once both are bound to DNA. By the model invoking dimeric acidic activation regions, one might imagine that acidic regions of two distinct proteins might associate to yield a heterodimer whose structure results in a higher level of function. Third, the two proteins might not interact directly, but may both contact a common target. In this view, the target would require a threshold of contacts, which neither protein alone would be able to achieve.

## *Activation and Repression by the Same Protein*

Several instances have been described that are consistent with the idea that a single protein can serve either as a transcriptional activator or repressor depending on the promoter. Based on the consensus sequences for binding and the properties of *mcm1* mutant strains, it has been suggested that GRM, the protein that binds cooperatively with the $\alpha 2$ repressor to **a**-specific promoters (105), might be identical (or related) to PTRF, the protein that binds in combination with the $\alpha 1$ activator to $\alpha$-specific promoters (98). If true, a single non-cell-type-specific protein could act as a co-activator or a co-repressor depending on the cell-type-specific protein with which it acts in combination. In addition, sequences that bind these proteins appear to act on their own as weak upstream promoter elements in **a**-cells (107), suggesting that PTRF/GRM can also act as an independent activator.

Proteins initially identified by their ability to bind to elements in the mating type silencer many also represent other examples of the same phenomenon (108, 109). Two of the silencer elements act as transcriptional activating elements when fused upstream of TATA elements (106, 109). The RAP1 protein, also known as GRF, binds to the upstream promoter elements of some ribosomal protein genes and to the telomers of chromosomes in addition to binding a silencer element (108, 109). In the appropriate context, the ribosomal promoter element functions as a silencer element (108). It will be interesting to determine if distinct regions of RAP1 are involved in transcriptional activation or repression.

## *Functional Distinctions Between TATA Elements*

It is commonly assumed that TATA sequences are general promoter elements that are recognized by a common transcription factor that is part of the basic transcriptional machinery. However, the *HIS3* promoter region contains two distinct classes of TATA elements, constitutive ($T_C$) and regulatory ($T_R$), that are defined by their interactions with upstream promoter elements, selectivity of initiation sites, and chromatin structure (110). Transcription dependent on $T_C$ is initiated equally from two sites, +1 and +12, whereas transcription dependent on $T_R$ initiates preferentially from the +12 site; this selectivity is determined primarily by the distance between $T_R$ and the initiation sites. Transcriptional activation by GCN4 and GAL4 occurs only in combination with the $T_R$ element, not $T_C$. As assayed by its ability to activate transcription in combination with GAL4, the sequence TATAAA in the *HIS3* promoter is fully sufficient for $T_R$ function. Saturation mutagenesis of this sequence revealed that 17 out of the 18 possible single mutations abolish $T_R$ function (TATATA being the only exception) (111).

The high sequence specificity of the *his3* $T_R$ element provides strong genetic evidence for a $T_R$-binding protein. It is very likely that this $T_R$ protein is the yeast TATA-binding protein identified by in vitro transcription using reconstituted mammalian factors (13, 14). However, the $T_C$ element necessary for constitutive *HIS3* expression does not have a sequence that fits the $T_R$ rules. Thus, it is almost certain that the $T_R$ and $T_C$ elements interact with different proteins, an explanation that easily accounts for the function distinctions between $T_R$ and $T_C$ elements with regard to their interactions with upstream activator proteins (110, 111). This idea also accounts for why overproduction of GAL4 "squelches" *HIS3* transcription from the +12 but not the +1 initiation site; presumably GAL4 is titrating out the $T_R$ protein but not the $T_C$ protein (112).

By analogy with bacterial $\sigma$ factors that interact with the core RNA polymerase to generate distinct holoenzymes that recognize different promoter sequences, yeast cells may contain multiple proteins that carry out a related "downstream element function" but have different specificities for DNA-binding. Two additional lines of evidence support this view. First, two TATA point mutations have the novel property of activating transcription in combination with GCN4 but not with GAL4 (113). The simplest interpretation of these results is that there are two "TATA-binding" proteins; $T_R$, which recognizes TATAAA and interacts functionally with both GCN4 and GAL4, and a distinct protein that recognizes a related sequence but interacts only with GCN4. Second, the $T_R$ element was replaced by random-sequence oligonucleotides, and functional promoter elements were selected by virtue of their ability to activate transcription in combination with GAL4. In addition to the

expected $T_R$-like sequences, other functional elements having no sequence resemblance to $T_R$ were obtained (V. Singer, K. Struhl, unpublished).

## *Poly(dA-dT) Sequences and the Effect of Chromatin*

Many yeast promoters contain poly(dA-dT) homopolymer sequences greater than 10 bp in length located upstream of the mRNA initiation site. For several constitutively expressed genes these poly(dA-dT) sequences act as the upstream promoter elements required for transcription (114). In addition, *adr2* promoter mutations that lengthen the natural poly(dA-dT) tract cause high constitutive levels of transcription (115). Interestingly, the *DED1* poly(dA-dT) element stimulates transcription by bacteriophage T7 RNA polymerase in yeast cells by a factor similar to its ability to stimulate the natural *DED1* promoter (116). These similar enhancement effects on two very different transcription machineries suggests that this poly(dA-dT) element mediates its effects through the chromatin template. As poly(dA-dT) sequences inhibit nucleosome formation in vitro (117, 118), it has been proposed that this particular class of upstream elements might act by excluding nucleosomes, not by binding specific proteins (114).

Consistent with this idea, several observations indicate that transcription from some promoters can be stimulated by direct alterations of chromatin structure. Inhibition of nucleosome formation by preventing the synthesis of histone H4 leads to an increase of *PHO5* transcription (119). Altering the balance of histones by gene dosage can influence transcription from certain promoters (120). Small N-terminal deletions of histone H4 do not affect cell growth, but they increase transcription from the silent mating type genes, presumably by interfering with silencer function (121). These experiments provide direct evidence that normal chromatin structure represents a transcriptionally repressed state.

Such chromatin structural models have been weakened recently by the identification of proteins binding to yeast poly(dA-dT) elements. The *DED1* poly(dA-dT) element stimulates transcription in vitro under conditions where nucleosome formation does not occur (122). Moreover, this effect is blocked by an excess of competing oligonucleotide, suggesting that transcriptional stimulation depends on a protein(s) binding to the element. In a separate line of experiments, a protein that specifically recognizes homopolymer (dA-dT) sequences at least 9 bp in length has been purified, and the gene encoding this protein has been cloned (E. Winter, A. Varshavsky, unpublished). Surprisingly, this poly(dA-dT)–binding protein appears to act as a transcriptional repressor because a mutant strain lacking the protein shows increased levels of transcription from an artificial promoter containing a poly(dA-dT) upstream element (E. Winter, A. Varshavsky, unpublished). However, the existence of specific binding proteins does not exclude the possibility that poly(dA-dT)

sequences might also influence transcription by virtue of their unusual DNA structures.

### *Upstream Activator Proteins Function When Bound at the TATA Position*

It is generally assumed that upstream activator proteins such as GAL4 and GCN4 stimulate transcription only when bound upstream of a TATA element. However, GCN4 can efficiently activate transcription from a *GAL-HIS3* promoter in which the conventional TATA element is replaced by a GCN4-binding site (123). In other words, GCN4 can activate transcription in the absence of a TATA element when bound close to the mRNA initiation site. Transcription occurs from wild-type initiation sites and requires both the GCN4 DNA-binding domain and the acidic activation region, but it is not affected by changing the spacing between the GCN4-binding site and the mRNA start sites. GCN4 is not sufficient for this TATA-independent activation; a sequence in the GAL region distinct from the GAL4 binding sites is also required. Thus, GCN4 functions both upstream of a TATA element and in place of a TATA element, suggesting that there might not be an intrinsic difference between an upstream and TATA activator protein (123).

The idea that the downstream element function can be carried out by multiple proteins, including conventional upstream activator proteins, can easily account for transcriptional activity from promoters lacking the conserved TATAAA sequence. For example, the *TRP3* promoter lacks the conserved TATAAA sequence and instead contains a GCN4-binding site 28 bp upstream of the mRNA start site that presumably mediates induction in response to amino acid starvation. It is important to note that even when GCN4 replaces the TATA function, at least two distinct promoter elements are necessary for transcription. However, not all combinations of upstream and downstream elements result in functional promoters.

## MOLECULAR MECHANISM OF ACTIVATION AND REPRESSION

Although RNA polymerase II can catalyze RNA synthesis on a variety of artificial templates in vitro, it does not bind to specific DNA sequences and does not initiate transcription, even randomly, on normal double-stranded templates. A view of the transcriptional initiation process is that RNA polymerase II initiates mRNA synthesis at discrete sites upon recognizing a transcription complex composed of upstream activator proteins, TATA factors, and the DNA (Figure 3). The complex is formed/stabilized by specific interactions between the proteins and cognate DNA sequences, and by pro-

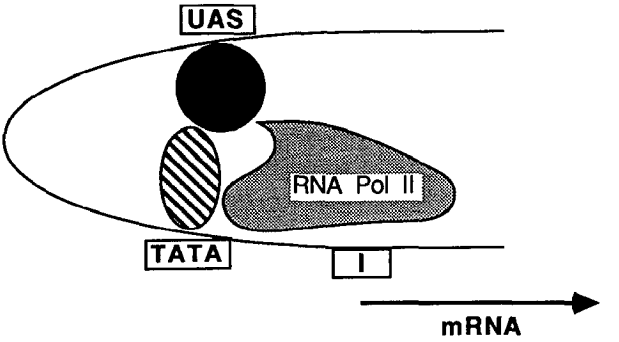


*Figure 3* Molecular model for transcriptional activation.

An upstream activator protein (black circle), TATA-binding protein (striped oval), and RNA polymerase II are shown as interacting with their target promoter elements (boxes) and with each other. The protein-protein interactions and the RNA polymerase II association with the initiator element are hypothetical. The DNA (line) is illustrated as bending to allow for the protein-protein interactions. See text.

tein-protein interactions between the various components. In this sense, the DNA serves as a specific scaffold for the assembly of an active transcription complex. The crucial mechanistic questions are which proteins are in direct contact, and how is RNA polymerase II activated.

Assuming that the acidic transcriptional activation regions are surfaces that contact other proteins, several observations suggest that the TATA-binding protein is a likely target. From the functional distinctions between the *HIS3* transcription $T_R$ and $T_C$ elements (110), it has been suggested that the GCN4 and GAL4 activation regions stimulate transcription by associating with a protein that binds $T_R$, whereas they are unable to interact with the protein that recognizes $T_C$ (1, 113). The proposed interaction between GAL4 and the $T_R$ proteins also explains why overproduction of GAL4 selectively inhibits *HIS3* transcription that depends on the $T_R$ element (112). Finally, the interaction of the mammalian TATA factor and its target DNA site appears to be altered by GAL4 derivatives that contain a functional activation region (124).

The proposed interaction between upstream activators and TATA factors does not explain how RNA polymerase II is activated for transcription. The fact that transcription can be initiated accurately, though inefficiently, in vitro from promoters containing only a "typical" TATA element has led to the belief that the major TATA factor is part of the basic transcription machinery and hence associated with RNA polymerase II. In principle, this presumptive interaction could either increase the binding of the TATA factor to its target site, or allosterically affect the TATA protein such that it would be able to promote transcription more efficiently.

An important issue is whether upstream activator proteins directly contact

RNA polymerase II or whether they affect transcription indirectly through the TATA factor. The fact that LexA hybrid proteins activate transcription when bound upstream of a TATA element argues against a strict requirement for a direct contact to RNA polymerase II. However, GCN4 can activate transcription when its binding site replaces the TATA element (123), and affinity chromatography indicates that GCN4 interacts directly with RNA polymerase II in vitro (125). Surprisingly, the GCN4 DNA-binding domain is necessary and sufficient for this interaction in vitro; the acidic activation region does not appear to be involved. Thus, GCN4 might facilitate the formation of an active initiation complex by utilizing different regions of the protein for contacting RNA polymerase II and a TATA factor (Figure 3). However, the significance of the GCN4-RNA polymerase II interaction for transcriptional activation in vivo remains to be determined.

By any model, RNA polymerase II must be the ultimate target for the activities of TATA and upstream activator proteins. In this regard, the largest subunit of RNA polymerase II from yeast to human contains a conserved seven-amino-acid sequence that is repeated many times at the C-terminus (126, 127). This "tail" is required for transcription in vitro (128) and in vivo (129, 130), although it is not important for basic RNA polymerizing activity. Although the yeast tail contains 26 repeats, derivatives containing only 13 repeats have minimal effects on cell growth, and those containing 10–12 repeats are viable at normal (but not high or low) temperatures (129, 130). Interestingly, the yeast tail can be functionally replaced by the analogous tail from hamster, but not by the more divergent tail from flies (130). It has been proposed that the tail might be an interaction site for TATA and or upstream activator proteins. The idea is attractive because of the repeating nature of both the tail and the acidic activation region (74), and because it provides an explanation for the synergism between upstream activator proteins that is often observed.

A basic property of upstream activator proteins is their ability to function when bound at long and variable distances from TATA elements and mRNA initiation sites. By analogy with well-documented examples of prokaryotic regulatory proteins, it is now believed that action at a distance is explained by looping out of the intervening DNA such that relevant proteins can directly interact (reviewed in 131). However, unlike the prokaryotic examples, which involve highly specific contacts between identical protein molecules, looping involved in eukaryotic transcriptional activation would presumably involve relatively nonspecific contacts between different proteins. Looping does not easily explain how activators can function at relatively short and variable distances because of the high energy involved in generating short loops. In this case, such flexibility might reflect variable conformations of the large

nonessential parts of the activator protein such that its critical acidic surface can interact with the other protein (1).

Given the above ideas about activation, repressor proteins could inhibit transcription by a variety of molecular mechanisms. The simplest model is steric hindrance, in which the binding of a repressor protein to its operator prevents binding of an upstream activator or TATA protein to its cognate promoter element. Steric hindrance might account for regulation of the *BAR1* gene (107) and be involved in the examples where multiple proteins bind very relaxed sequences. However, a bacterial repressor protein (132) or derivatives of yeast activator proteins lacking the acidic activation region (48, 61) can inhibit transcription when bound at various positions between an upstream and TATA element. In these cases, it is likely that the repressors interfere with the putative interaction between the transcriptional activator proteins rather than preventing DNA binding. Such interference might simply reflect steric constraints between the relevant proteins, or an increased difficulty of DNA looping due to the presence of a bound protein within the loop.

Neither of these steric hindrance models can explain how some repressors can function when bound upstream of promoter elements. By analogy with transcriptional activators, it seems that repressor proteins that function at a distance must require a transcriptional repression function that is distinct from DNA binding. Consistent with this analogy, N-terminal deletions of $\alpha 2$ behave as negative control mutants in that they do not repress transcription even though they can bind DNA (102). Perhaps, the putative negative regulatory regions of repressors interact with the same target as that recognized by the acidic activation regions, thus forming a distinct and stable complex that prevents the formation of an active initiation complex. The idea that activators and repressors may interact with a common target also suggests a mechanism for how a single protein could act as a positive or negative regulatory protein depending on the promoter context.

## EVOLUTION

It is now clear that the mechanism of transcription is remarkably conserved throughout the eukaryotic kingdom. Yeast upstream activator proteins function in a variety of eukaryotic organisms, vertebrate transcription factors function in yeast cells, and the yeast and mammalian TATA factors are functionally interchangeable. Such functional conservation undoubtedly indicates that the basic mechanism of transcriptional initiation has existed since the first eukaryotic organisms.

Surprisingly, yeast and mammalian cells have structurally related transcrip-

tion factors that recognize essentially identical sequences even though the homologues perform different functions in their respective organisms. For example, GCN4 and HAP2,3 activate the amino acid biosynthetic and oxygen-regulated genes, respectively, in yeast, whereas their evolutionary counterparts, jun/AP-1 and CP-1, activate a variety of genes whose functions appear unrelated. Assuming that the original eukaryotic organisms contained multiple genes that utilized these conserved binding sites, it would be difficult to alter the sequence recognition properties of the regulatory protein without affecting the transcription of many genes. Similar arguments have been used to explain why the genetic code is essentially universal and why eukaryotic TATA elements and prokaryotic −10 sequences are extremely similar even though transcriptional and translational mechanism are quite different (133).

Despite all the similarities, there are a few significant differences between yeast and mammalian transcriptional mechanisms. First, unlike mammalian activators, yeast proteins cannot activate transcription when bound downstream of the mRNA initiation site. However, this property does not reflect an intrinsic difference between activator proteins because yeast activators can function at a downstream position in mammalian cells, and mammalian factors cannot function downstream in yeast cells. Second, the distance between TATA element and mRNA initiation sites is rather large and variable in yeast promoters when compared to the short fixed distance in mammalian promoters. Again this does not reflect basic differences between the yeast and mammalian TATA factors; in mammalian in vitro transcription systems, the yeast TATA factor stimulates initiation at the distance expected for a mammalian promoter. In both these cases, it is clear that there is some difference in the basic transcription machineries of yeast and higher eukaryotes such that they respond differently to upstream activator proteins and TATA factors. It is tempting to speculate that this difference involves the initiator element. Perhaps, a factor(s) involved in correct initiation in yeast also blocks the ability of upstream activator proteins to function in a downstream position. In any event, it is clear that transcriptional regulatory mechanisms in yeast and higher eukaryotes, though not identical, are extremely similar, and that continuing analysis of yeast genes will continue to uncover new and general principles that are relevant to all eukaryotic organisms.

## ACKNOWLEDGMENTS

I thank colleagues in the yeast gene regulation field for many valuable discussions over the years. This work was supported by grant GM 30186 from the National Institutes of Health.

## Literature Cited


1. Struhl, K. 1987. *Cell* 49:295–97
2. Guarente, L. 1987. *Annu. Rev. Genet.* 21:425–52
3. Guarente, L. 1988. *Cell* 52:303–5
4. McKnight, S. L., Tjian, R. 1986. *Cell* 46:795–805
5. Allison, L. A., Moyle, M., Shales, M., Ingles, C. J. 1985. *Cell* 42:599–610
6. Corden, J. L., Cadena, D. L., Ahearn, J. M. Jr., Dahmus, M. E. 1985. *Proc. Natl. Acad. Sci. USA* 82:7934–38
7. Struhl, K. 1987. *Cell* 50:841–46
8. Bohmann, D., Bos, T. J., Admon, A., Nishimura, T., Vogt, P. K., Tjian, R. 1987. *Science* 238:1386–92
9. Angel, P., Allegretto, E. A., Okino, S. T., Hattori, K., Boyle, W. J., et al. 1988. *Nature* 332:166–71
10. Harshman, K. D., Moye-Rowley, W. S., Parker, C. S. 1988. *Cell* 53:321–30
11. Jones, R. H., Moreno, S., Nurse, P., Jones, N. C. 1988. *Cell* 53:659–67
12. Chodosh, L. A., Olesen, J., Hahn, S., Baldwin, A. S., Guarente, L., Sharp, P. A. 1988. *Cell* 53:25–35
13. Buratowski, S., Hahn, S., Sharp, P. A., Guarente, L. 1988. *Nature* 334:37–42
14. Cavallini, B., Huet, J., Plassat, J.-L., Sentenac, A., Egly, J.-M., Chambon, P. 1988. *Nature* 334:77–80
15. Kakidani, H., Ptashne, M. 1988. *Cell* 52:161–67
16. Webster, N., Jin, J. R., Green, S., Hollis, M., Chambon, P. 1988. *Cell* 52: 169–78
17. Fischer, J. A., Giniger, E., Maniatis, T., Ptashne, M. 1988. *Nature* 332:853–56
18. Ma, J., Przibilla, E., Hu, J., Bogorad, L., Ptashne, M. 1988. *Nature* 334:631–33
19. Lech, K., Anderson, K., Brent, R. 1988. *Cell* 52:179–84
20. Struhl, K. 1988. *Nature* 332:649–50
21. Metzger, D., White, J. H., Chambon, P. 1988. *Nature* 334:31–36
22. Schena, M., Yamamoto, K. R. 1988. *Science* 241:965–67
23. Struhl, K. 1983. *Nature* 305:391–96
24. Guarente, L., Hoar, E. 1984. *Proc. Natl. Acad. Sci. USA* 81:7860–64
25. Struhl, K. 1984. *Proc. Natl. Acad. Sci. USA* 81:7865–69
26. Struhl, K. 1982. *Proc. Natl. Acad. Sci. USA* 79:7385–89
27. Guarente, L., Mason, T. 1983. *Cell* 32:1279–86
28. Siliciano, P. G., Tatchell, K. 1984. *Cell* 37:969–78
29. Nagawa, F., Fink, G. R. 1985. *Proc. Natl. Acad. Sci. USA* 82:8557–61
30. Chen, W., Struhl, K. 1985. *EMBO J.* 4:3273–80
31. Hahn, S., Hoar, E. T., Guarente, L. 1985. *Proc. Natl. Acad. Sci. USA* 82:8562–66
32. McNeil, J. B., Smith, M. 1986. *J. Mol. Biol.* 187:363–78
33. Brent, R. 1985. *Cell* 42:3–4
34. Johnson, A. D., Herskowitz, I. 1985. *Cell* 42:237–47
35. Miller, A. M., MacKay, V. L., Nasmyth, K. A. 1985. *Nature* 314:598–603
36. Struhl, K. 1985. *Nature* 317:822–24
37. Abraham, J., Nasmyth, K. A., Strathern, J. N., Klar, A. J. S., Hicks, J. B. 1984. *J. Mol. Biol.* 176:307–31
38. Feldman, J. B., Hicks, J. B., Broach, J. R. 1984. *J. Mol. Biol.* 178:815–34
39. Brand, A. H., Breeden, L., Abraham, J., Sternglanz, R., Nasmyth, K. 1985. *Cell* 41:41–48
40. Pabo, C. O., Sauer, R. T. 1984. *Annu. Rev. Biochem.* 53:293–321
41. Miller, J., McLachlan, A. D., Klug, A. 1985. *EMBO J.* 4:1609–14
42. Landschulz, W. H., Johnson, P. F., McKnight, S. L. 1988. *Science* 240:1759–64
43. Bram, R., Kornberg, R. 1985. *Proc. Natl. Acad. Sci. USA* 82:43–47
44. Giniger, E., Varnum, S. M., Ptashne, M. 1985. *Cell* 40:767–74
45. Bram, R. J., Lue, N. F., Kornberg, R. D. 1986. *EMBO J.* 5:603–8
46. Selleck, S. B., Majors, J. 1987. *Mol. Cell. Biol.* 7:3260–67
47. Giniger, E., Ptashne, M. 1988. *Proc. Natl. Acad. Sci. USA* 85:382–86
48. Keegan, L., Gill, G., Ptashne, M. 1986. *Science* 231:699–704
49. Berg, J. 1986. *Science* 232:485–87
50. Johnston, M. 1987. *Nature* 328:353–55
51. Johnston, M., Dover, J. 1988. *Genetics* 120:63–74
52. Johnston, M., Dover, J. 1987. *Proc. Natl. Acad. Sci. USA* 84:2401–5
53. Witte, M. M., Dickson, R. C. 1988. *Mol. Cell. Biol.* 8:3726–33
54. Salmeron, J. M., Johnston, S. A. 1986. *Nucleic Acids Res.* 14:7767–81
55. Wray, L. V., Witte, M. M., Dickson, R. C., Riley, M. I. 1987. *Mol. Cell. Biol.* 7:1111–21
56. Parraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., et al. 1988. *Science* 241:1489–92

57. Hope, I. A., Struhl, K. 1985. *Cell* 43:177–88
58. Arndt, K., Fink, G. 1986. *Proc. Natl. Acad. Sci. USA* 83:8516–20
59. Hill, D. E., Hope, I. A., Macke, J. P., Struhl, K. 1986. *Science* 234:451–57
60. Hope, I. A., Struhl, K. 1987. *EMBO J.* 6:2781–84
61. Hope, I. A., Struhl, K. 1986. *Cell* 46:885–94
62. Vogt, P. K., Bos, T. J., Doolittle, R. F. 1987. *Proc. Natl. Acad. Sci. USA* 84:3316–19
63. Pfeifer, K., Arcangioli, B., Guarente, L. 1987. *Cell* 49:9–18
64. Pfeifer, K., Prezant, T., Guarente, L. 1987. *Cell* 49:19–27
65. Zitomer, R. S., Sellers, J. W., McCarter, D. W., Hastings, G. A., Wick, P., Lowry, C. V. 1987. *Mol. Cell. Biol.* 7:2212–20
66. Cerdan, M. E., Zitomer, R. S. 1988. *Mol. Cell. Biol.* 8:2275–79
67. Sauer, R. T., Smith, D. L., Johnson, A. D. 1988. *Genes Dev.* 2:807–16
68. Shepherd, J. C. W., McGinnis, W., Carrasco, A. E., DeRobertis, E. M., Gehring, W. J. 1984. *Nature* 310:70–71
69. Porter, S. D., Smith, M. 1986. *Nature* 320:766–68
70. Laughon, A., Scott, M. P. 1984. *Nature* 310:25–30
71. Selleck, S. B., Majors, J. 1987. *Nature* 325:173–77
72. Brent, R., Ptashne, M. 1985. *Cell* 43:729–36
73. Ma, J., Ptashne, M. 1987. *Cell* 48:847–53
74. Hope, I. A., Mahadevan, S., Struhl, K. 1988. *Nature* 333:635–40
75. Ma, J., Ptashne, M. 1987. *Cell* 51:113–19
76. Gill, G., Ptashne, G. 1987. *Cell* 51: 121–26
77. Giniger, E., Ptashne, M. 1987. *Nature* 330:670–72
78. Hurstel, S., Granger-Schnarr, M., Schnarr, M. 1988. *EMBO J.* 7:269–75
79. Luc, N. F., Chasman, D. I., Buchman, A. R., Kornberg, R. D. 1987. *Mol. Cell. Biol.* 7:3446–51
80. Ma, J., Ptashne, M. 1987. *Cell* 50:137–42
81. Johnston, S. A., Salmeron, J. M. Jr., Dincher, S. S. 1987. *Cell* 50:143–46
82. Sorger, P. K., Lewis, M. J., Pelham, H. R. B. 1987. *Nature* 329:81–84
83. Sorger, P. K., Pelham, H. R. B. 1988. *Cell* 54:855–64
84. Celenza, J. L., Carlson, M. 1986. *Science* 233:1175–80
85. Thireos, G., Penn, M. D., Greer, H. 1984. *Proc. Natl. Acad. Sci. USA* 81:5096–100
86. Hinnebusch, A. G. 1984. *Proc. Natl. Acad. Sci. USA* 81:6442–46
87. Mueller, P. P., Hinnebusch, A. G. 1986. *Cell* 42:201–7
88. Tzamarias, D., Alexandraki, D., Thireos, G. 1986. *Proc. Natl. Acad. Sci. USA* 83:4849–53
89. Nasmyth, K. A., Shore, D. 1987. *Science* 237:1162–70
90. Struhl, K., Hill, D. E. 1987. *Mol. Cell. Biol.* 7:104–10
91. Moye-Rowley, W. S., Harshman, K. D., Parker, C. S. 1989. *Genes Dev.* 3:283–92
92. Lin, Y.-S., Green, M. R. 1989. *Proc. Natl. Acad. Sci. USA*. In press
93. Arcangioli, B., Lescure, B. 1985. *EMBO J.* 4:2627–33
94. Johnson, A. D., Poteete, A. R., Lauer, G., Sauer, R. T., Ackers, G. K., Ptashne, M. 1981. *Nature* 294:217–23
95. Olesen, J., Hahn, S., Guarente, L. 1987. *Cell* 51:953–61
96. Hahn, S., Guarente, L. 1988. *Science* 240:317–21
97. Chodosh, L. A., Baldwin, A. S., Carthew, R. W., Sharp, P. A. 1988. *Cell* 53:11–24
98. Bender, A., Sprague, G. F. Jr. 1987. *Cell* 50:681–91
99. Jarvis, E. E., Hagen, D. C., Sprague, G. F. Jr. 1988. *Mol. Cell. Biol.* 8:309–20
100. Goutte, C., Johnson, A. D. 1988. *Cell* 52:875–82
101. Siliciano, P. G., Tatchell, K. 1986. *Proc. Natl. Acad. Sci. USA* 83:2330–34
102. Hall, M. N., Johnson, A. D. 1987. *Science* 237:1007–12
103. Strathern, J., Shafer, B., Hicks, J., McGill, C. 1988. *Genetics* 120:75–81
104. Lalonde, B., Arcangioli, B., Guarente, L. 1986. *Mol. Cell. Biol.* 6:4690–96
105. Keleher, C. A., Goutte, C., Johnson, A. D. 1988. *Cell* 53:927–36
106. Brand, A. H., Micklem, G., Nasmyth, K. 1987. *Cell* 51:709–19
107. Kronstad, J. W., Holly, J. A., MacKay, V. L. 1987. *Cell* 50:369–77
108. Shore, D., Nasmyth, K. 1987. *Cell* 51:721–32
109. Buchman, A. R., Kimmerly, W. J., Rine, J., Kornberg, R. D. 1988. *Mol. Cell. Biol.* 8:210–25
110. Struhl, K. 1986. *Mol. Cell. Biol.* 6:3847–53
111. Chen, W., Struhl, K. 1988. *Proc. Natl. Acad. Sci. USA* 85:2691–95
112. Gill, G., Ptashne, M. 1988. *Nature* 334:721–24

113. Struhl, K., Brandl, C. J., Chen, W., Harbury, P. A. B., Hope, I. A., Mahadevan, S. 1988. *Cold Spring Harbor Symp. Quant. Biol.* 53:701–9
114. Struhl, K. 1985. *Proc. Natl. Acad. Sci. USA* 82:8419–23
115. Russell, D. W., Smith, M., Cox, D., Williamson, V. M., Young, E. T. 1983. *Nature* 304:652–54
116. Chen, W., Tabor, S., Struhl, K. 1987. *Cell* 50:1047–55
117. Kunkel, G. R., Martinson, H. G. 1981. *Nucleic Acids Res.* 9:6869–88
118. Prunell, A. 1982. *EMBO J.* 1:173–79
119. Han, M., Kim, U-J., Kayne, P., Grunstein, M. 1988. *EMBO J.* 7:2221–28
120. Clark-Adams, C. D., Norris, D., Osley, M. A., Fassler, J., Winston, F. 1988. *Genes Dev.* 2:150–59
121. Han, M., Grunstein, M. 1988. *Cell* 55:1137–45
122. Lue, N. F., Buchman, A. R., Kornberg, R. D. 1989. *Proc. Natl. Acad. Sci. USA* 86:486–90
123. Chen, W., Struhl, K. 1989. *EMBO J.* 8:261–68
124. Horikoshi, M., Carey, M. F., Kakidani, H., Roeder, R. G. 1988. *Cell* 54:665–69
125. Brandl, C. J., Struhl, K. 1989. *Proc. Natl. Acad. Sci. USA.* In press
126. Allison, L. A., Moyle, M., Shales, M., Ingles, C. J. 1985. *Cell* 42:599–610
127. Corden, J. L., Cadena, D. L., Ahearn, J. M. Jr., Dahmus, M. E. 1985. *Proc. Natl. Acad. Sci. USA* 82:7934–38
128. Dahmus, M. E., Kedinger, C. 1983. *J. Biol. Chem.* 258:2303–7
129. Nonet, M., Sweetser, D., Young, R. A. 1987. *Cell* 50:909–15
130. Allison, L. A., Wong, J. K., Fitzpatrick, V. D., Moyle, M., Ingles, C. J. 1988. *Mol. Cell. Biol.* 8:321–29
131. Ptashne, M. 1986. *Nature* 322:697–701
132. Brent, R., Ptashne, M. 1984. *Nature* 312:612–15
133. Struhl, K. 1986. *J. Mol. Biol.* 191:221–29

*Annual Review of Biochemistry*
*Volume 58, 1989*



# CONTENTS



*(continued)* 