CHAPTER
2

Yeast Promoters

Kevin Struhl

In the early 1960s, Jacob and Monod (Jacob and Monod 1961, Jacob et al. 1964) developed the current conceptual framework for the regulation of gene expression by dividing genes into three parts. The *structural* gene directly encodes a gene product; the *promoter* is responsible for the expression of this structural gene; and *regulatory* sequences ensure that the gene product is synthesized only in the correct environmental circumstances.

Although this framework could in principle support essentially any molecular model, these authors specifically proposed that (1) gene regulation occurs primarily by controlling the frequency of transcription initiation; (2) the promoter serves as a recognition sequence for RNA polymerase; and (3) regulatory sequences are binding sites for specific regulatory proteins that alter the rate of transcription initiation. The experimental basis for these proposals derived from a genetic analysis of mutations that alter the expression and/or the regulation of specific structural genes. This approach has clearly proven to be valid, because inferences drawn from the genetic properties of promoter/regulatory elements were ultimately shown to fit the biochemical facts.

Genetic, biochemical, and physical analyses of prokaryotic genes carried out during the past 20 years have extended the Jacob-Monod model into a detailed molecular description, which is the subject of Chapter 1. These basic concepts of gene regulation, protein-DNA interactions, and DNA sequence recognition have been developed by correlating the DNA sequences of wild-

type and mutant genes with their biochemical properties and with their effects *in vivo*. Most important, many aspects of gene regulation have been reproduced *in vitro* with purified components.

A general scheme for transcription initiation, and its positive and negative regulation, is shown in Figure 2–1. As suggested originally, the promoter and regulatory sites are indeed DNA sequences to which specific proteins bind. A

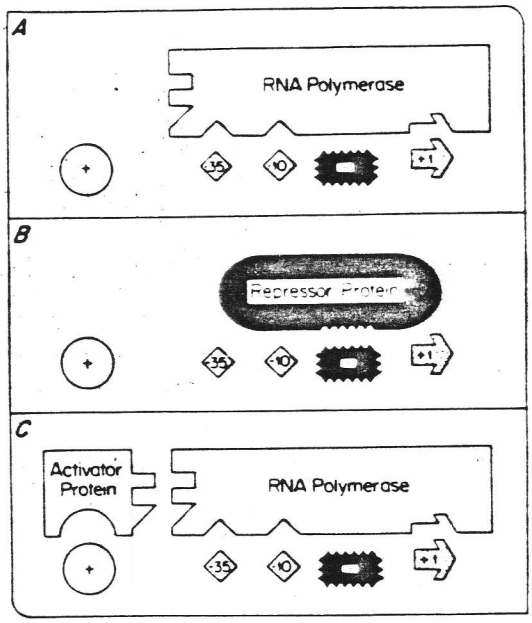

FIGURE 2–1 Transcriptional regulation in *E. coli*. Promoter/regulatory sequences for a hypothetical gene are illustrated in the bottom sections of panels A to C: transcription initiation site (defined as +1 and shown as a rightward arrow); −10 and −35 promoter elements (shaded diamonds); operator site (black box with ruffled edges); positive control site (open circle). The proteins involved in transcription regulation are indicated as follows: RNA polymerase (shaded box); repressor protein (black oval); activator protein (open box). Specific interactions between proteins and DNA sequences are indicated by interlocking shapes and patterns. Panel A describes basal level transcription, panel B describes negative control (repression), and panel C describes positive control (activation). See text for details.

prokaryotic promoter consists of two distinct elements that together are necessary and sufficient both for RNA polymerase holoenzyme binding and for transcription initiation at a discrete site. Thus, transcription initiation can be viewed as an enzyme-substrate interaction between RNA polymerase and promoter DNA. Negative control of transcription is achieved by repressor proteins that bind to operator sites located within the region of DNA that interacts with RNA polymerase. The presence of bound repressor in this

region prevents functional interactions between polymerase and promoter DNA, with the result that transcription initiation is inhibited. With regard to positive control, activator proteins bind to positive regulatory sites usually located upstream of but close to the promoter sequences. Their stimulatory effects on transcription are believed to result from protein-protein interactions. Another general mechanism of transcriptional regulation involves altered forms of RNA polymerase, which are distinguished by their σ subunits. Since promoter recognition depends on the σ subunit, these distinct enzyme forms interact with different DNA sequences. It is presumed, however, that these different RNA polymerases initiate transcription by the same basic mechanism.

Eukaryotic organisms are considerably more complicated than bacteria, and at the molecular level, there is a large collection of differences between them. Nevertheless, it is now clear that a major aspect of gene control in eukaryotic organisms is determined by the level of RNA synthesis. How do eukaryotic organisms initiate transcription, and how do they control the rate in response to environmental and developmental cues? Are the mechanistic principles essentially the same as those found in prokaryotic organisms, or are they qualitatively different? Although our understanding is still limited, the genetic properties of eukaryotic promoter/regulatory elements are now fairly well defined. What is clear is that these properties are distinct from those of prokaryotic promoter elements, and that as a consequence, the underlying molecular mechanisms must be different.

2.1 TRANSCRIPTION IN YEAST

This review is concerned with promoter/regulatory elements in the baker's yeast, *Saccharomyces cerevisiae*. Although yeast is a simple microorganism, it has most of the characteristics of higher eukaryotic cells (for general reviews, see Mortimer and Schild 1980; Petes 1980; Strathern et al. 1982; Struhl 1983a). Yeast genes are distributed among 16 linear chromosomes, each of which contains multiple DNA replication origins, a centromere, and two telomeres. Yeast DNA is confined to the nucleus, and it is complexed with histones in a discrete chromatin structure that is not apparent in bacterial cells. Unlike the situation in *E. coli*, transcription is not carried out by a single RNA polymerase, but rather by three distinct RNA polymerases. RNA polymerase I transcribes only the ribosomal RNA genes, which accounts for about 70% of total RNA in the cell. RNA polymerase III transcribes tRNA genes and the 5S ribosomal RNA gene; these RNAs represent approximately 30% of the cellular total. RNA polymerase II transcribes all the approximately 5 to 10,000 protein coding genes, but like other eukaryotic organisms, this represents only 1% of the total RNA. Since almost all gene regulation involves the protein coding genes, most of the current work has focused on transcription by RNA polymerase II (for a brief summary, see Guarente 1984). This chapter is concerned exclusively with RNA polymerase II.

The primary transcripts of RNA polymerase II become translatable mRNAs upon post-transcriptional addition of 5′ caps and 3′ polyA sequences and upon transport to the cytoplasm. For those genes that have intervening sequences, RNA splicing of the primary transcripts is essential to produce functional mRNA. Yeast mRNAs, like those of other eukaryotic species, are relatively stable. The average half-life of 20 minutes represents 20% of a cell division cycle and, therefore, is equivalent to about 4 hours for mammalian cells. Although regulation of gene expression can take place at many different stages along the pathway between gene and function, transcription initiation remains perhaps the most important point of gene control. Thus, this chapter neither deals with elongation or termination of transcription nor covers other aspects of RNA metabolism such as processing, transport, or degradation.

General surveys of yeast transcription indicate that roughly 50% of the genome is transcribed under normal growth conditions (Hereford and Rosbash 1977; Kaback et al. 1979). This means that the yeast genome contains approximately 5,000 protein coding genes that, by necessity, are packed closely together. However, from the genetic map, it is clear that genes of similar function are not clustered together, but rather are scattered essentially at random around the genome. Two notable exceptions are the genes involved in galactose utilization (*gal7*, *gal10*, and *gal1*) and genes of the mating type locus (*mata1* and *mata2*).

Analysis of thousands of individual yeast DNA segments indicates that most genes are transcribed at similar levels (St. John and Davis 1979), approximately one to two molecules per cell at the steady state (Struhl and Davis 1981). It seems likely that many, if not most, yeast genes are always expressed at this basal level because in any particular regulatory situation, transcription rates are altered for only a small number of genes. However, most studies have focused on genes whose expression is regulated in response to particular environmental or developmental cues. These cues include regulation as a function of exogenously added carbon sources, of amino acid starvation, of time within the cell division cycle, and of cell type.

A detailed undestanding of transcription initiation requires knowledge about the *cis*-acting DNA sequences that are essential for the expression and regulation of specific structural genes, the proteins that interact with these sequences, and the molecules that affect the activity of these proteins such that they influence transcription in the appropriate physiological manner. Most of this chapter is devoted to the DNA sequences and the genetic properties of promoter/regulatory elements for specific yeast genes.

2.2 METHODS FOR STUDYING YEAST PROMOTERS

Our views of yeast promoter/regulatory sequences are inextricably linked with experimental methodology. Rather than interjecting technical comments throughout the chapter, it should be more efficient and more informative to

consider the various approaches in one section. I shall focus primarily on genetic experiments since they form the basis of most of our current knowledge.

The major advances have all depended on gene cloning and yeast DNA transformation techniques (for a review of yeast molecular genetics, see Struhl 1983a). Classical genetic approaches such as those employed in prokaryotic organisms yielded few mutants and little insight. One reason for this is that yeast genes are not organized into operons, thus making it difficult to distinguish undesired mutations in structural genes from those affecting its expression or regulation. A second reason, which will become apparent later, is that yeast promoters often contain redundant information; thus, point mutations (the predominant class *in vivo*) frequently do not cause phenotypic effects that are easily observed.

The standard approach proceeds as follows: A particular gene is isolated by molecular cloning; mutations within this cloned DNA are introduced at will by the appropriate enzymatic or chemical treatments; and finally, mutant DNAs are re-introduced into yeast cells, whereupon their phenotypic consequences are assessed. Several aspects of this approach are worth noting. First, *in vitro* mutations of unlimited variety can be created in a systematic manner without regard to their *in vivo* phenotypes. Thus, by comparing similar derivatives that are or are not functional, the extent and the properties of a given promoter/regulatory element can be clearly defined. Second, since the mutant DNAs are introduced back into the intact organism from which they were derived, they are subject to all normal forms of cellular regulation. Third, the mutant DNAs are introduced by transforming yeast cells using a selectable marker that is unrelated to the phenotype to be tested; the resulting transformed strains can be propagated indefinitely and treated like normal yeast strains. Fourth, the mutant DNAs are re-introduced in defined ways with respect to copy number per cell and chromosomal location. Of particular importance is that newly introduced DNA can be re-integrated exactly at its normal chromosomal location or maintained as part of an artificially constructed minichromosome. Therefore, although the experimental manipulations are quite different from those of classical genetics, the end results are equivalent and interchangeable.

Several schemes for analyzing mutant phenotypes are shown in Figure 2–2 and Table 2–1. By using different kinds of yeast vectors, mutant DNAs can integrate into the genome by homologous recombination, or they can replicate as autonomous molecules. In the case of replicating vectors, mutant DNAs can be introduced either in low (1 to 3) or high (5 to 40) copy numbers per cell, and the molecules can be either linear or circular. Transformation via autonomous replication is convenient because the event occurs in a single step and the frequency is high. A major disadvantage, however, is that the DNA molecules, most of which are small and circular, are structurally different from real yeast chromosomes, which are long and linear. Moreover, these molecules are to varying extents unstable during meiosis and mitosis, and their copy number varies among different cells within a population; these

problems can sometimes lead to quantitative errors. Integrative transformation, although it occurs at much lower frequency, avoids most of these problems. Specifically, the transforming DNA is present at one copy per cell in the normal chromosomal location of the gene of interest. A minor consideration is that the vector sequences are also present.

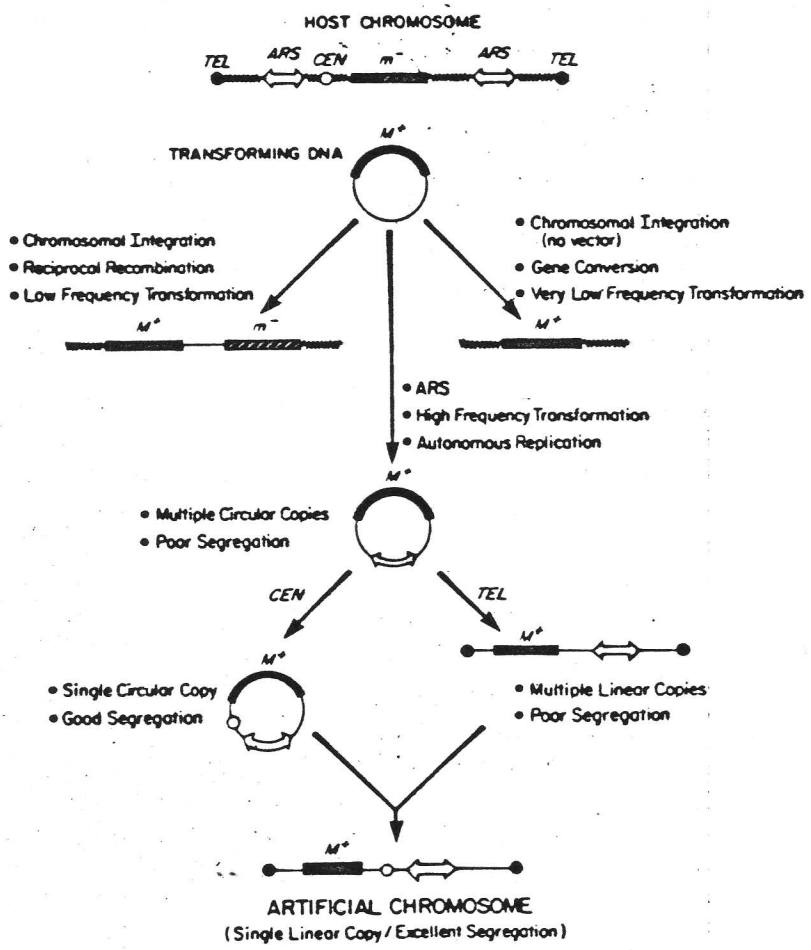

FIGURE 2–2 Yeast vectors. The top line represents a typical host chromosome with the following genetic elements: telomere (*TEL*, closed circle); centromere (*CEN*, open circle); multiple replication origins (*ARS*, bidirectional arrow); mutant DNA sequences corresponding to the vector gene (m^-, striped bar); other chromosomal sequences (wavy line). The circular transforming DNA (diagrammed below the host chromosome) consists of an M^+ marker (solid bar) and vector sequences (wavy line) that include an *E. coli* selectable marker and replication origin. Transforming DNA can integrate into the chromosome by reciprocal recombination or by gene conversion (only the chromosomal region near the *m* locus is shown). Addition of cloned *ARS*, *CEN*, or *TEL* elements produces vectors with properties listed in this figure and described more fully in Table 2–1.

TABLE 2-1 Properties of Yeast Transformation

Property	*Chromosomal Integration*	*Gene Conversion*	*Episomal Replicator*	*Chromosomal Replicator*	*Mini chromosome*	*Linear DNAs*
Vector	I	I	E	R	C	L
Transformation Frequency	10	1	10,000	10,000	10,000	10,000
Autonomous Replication	None	None	Circular	Circular	Circular	Linear
Copies per cell	1	1	5–40	3–30	1	5–30
Vector Sequences	Yes	No	Yes	Yes	Yes	Yes
Integration Frequency	1	1	Variable	10^{-5}	10^{-7}	NT
Required Elements	Yeast DNA	Yeast DNA	2μ ARS	ARS	ARS, CEN	ARS, TEL
Mitotic loss	0.1%	0	30%	30%	1%	30%
Meiotic loss	1–10%	0	90%	90%	30%	90%

Note: Each column represents a particular mode of yeast transformation (Hinnen et al. 1978; Beggs 1978; Struhl et al. 1979; Clarke and Carbon 1980; Szostak and Blackburn 1982; Struhl 1983a). Yeast vectors, categorized as I, E, R, C, and L, all contain sequences that permit replication in *E. coli* cells as well as genetic markers that allow for selection in *E. coli*. The transformation frequency is measured in colonies per μg. In cases involving high rates of mitotic loss, the number of copies per cell represents an average. The integration frequency is listed as events per generation; for I vectors, the frequency of 1 means that all transformation events require integration. Mitotic loss is measured on a per generation basis, and meiotic loss is measured by tetrad analysis. Required genetic elements are abbreviated as follows: ARS, autonomously replicating segment; CEN, centromeric DNA; TEL, telomeric DNA; 2μ, endogenous yeast plasmid sequences (see Figure 2-1).

The best possible *in vivo* assay involves exact replacement of the normal yeast chromosomal gene with the derivative created *in vitro*. Gene replacement is accomplished by two successive homologous recombination events (Figure 2-3). In the first step, the transforming DNA including all vector sequences is integrated into the yeast genome at the locus of interest; in the second step, this DNA is excised. Fifty percent of the time, the crossover points for these recombination events are correctly positioned for gene replacement. Although a variety of tricks are used to select for the integration and excision steps, the end result is that all vector sequences are eliminated, and the mutant gene to be tested is present in one copy per cell precisely at its normal chromosomal location.

In general, the results obtained with replicative vectors agree with those obtained with integrative vectors. However, there are an accumulating number of discrepancies. Since gene replacement represents the true *in vivo* test, experiments involving replicating vectors must be viewed with at least some suspicion. In the long run, these discrepancies may provide clues to molecular mechanisms.

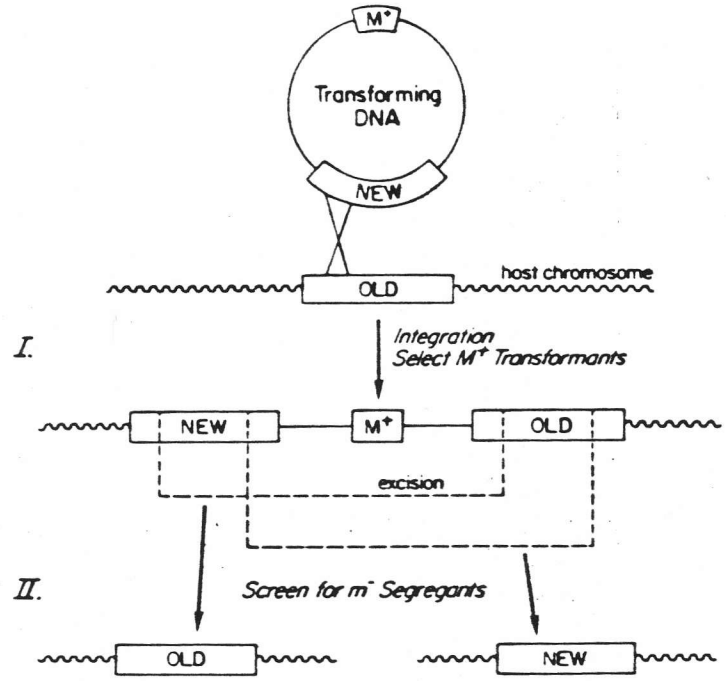

FIGURE 2-3 Gene replacement. The basic method involves integration of the transforming DNA followed by excision (Scherer and Davis 1979). The top part of the figure shows a homologous recombination event between a circular transforming DNA molecule and the normal linear chromosome (wavy line). The transforming DNA contains a selectable marker (M^+) and the "new" allele to be tested; the original chromosome contains the "old" allele. Integration (step I) results in a chromosome containing both alleles separated by the M^+ vector. This step is greatly facilitated by appropriate cleavage of the transforming DNA, which increases transformation frequency and directs the site of integration (Orr-Weaver et al. 1981). The lower part of the figure illustrates the excision step (step II). Two classes of m^- segregants are observed, both of which lack all vector sequences. The desired class represents replacement of the old allele by the new allele present on the transforming DNA; the other class is equivalent to the original host strain. Segregants are identified by replica plating, or more recently by direct selection (Boeke et al. 1984). Two additional methods permit the direct selection for gene replacement events (Rothstein 1983; Struhl 1983b).

Mutant phenotypes are assessed in a number of different ways. The most direct and important assays are to quantitate RNA levels and to determine the transcription initiation site. Many studies, however, use the more indirect method of enzymologically assaying the protein product. In many cases, phenotypes can be qualitatively determined by physiological means. For example, if the gene of interest encodes an amino acid biosynthetic enzyme, cells that fail to express the gene are unable to grow in a medium lacking the amino acid. Thus, the relative growth rate in such a medium is indicative of

the level of expression. Of more importance, such growth properties provide the basis for genetic selections or screens. Finally, mutations within a given promoter/regulatory region are often assessed by fusion to a structural gene that can be assayed easily both genetically and biochemically. The most common fusions involve the *E. coli* β-galactosidase (*lacZ*) structural gene; they can be assayed quantitatively by enzymatic activity and qualitatively by the blue indicator dye, Xgal (Guarante and Ptashne 1981; Rose et al. 1981). Gene fusions are particularly valuable for dissecting complicated regulatory situations because the gene product being assayed does not influence the regulation that is under study.

2.3 UPSTREAM PROMOTER ELEMENTS

In studying the promoter region of a given gene, the initial experiments invariably are designed to determine the minimum contiguous sequence necessary for wild-type levels of transcription. This is accomplished by creating a series of mutations that successively remove DNA sequences adjacent to the 5′ end of the mRNA coding region. Ideally, deletion mutants that remove upstream sequences up to a boundary point will behave indistinguishably from the wild-type gene, while more extensive deletions will impair transcription significantly. If this occurs, it indicates two strong conclusions. First, the most deleted derivative that confers normal levels of transcription contains the entire promoter region; sequences further upstream of the deletion break point are presumably unimportant. Second, the more extensive deletions decrease transcription because essential promoter sequences are removed. Specifically, something between the two defining deletion break points is essential for transcription.

2.3.1 Yeast Promoters Are Relatively Large but Differ in Size

Sequential 5′ deletion analysis has been carried out for a number of yeast genes, and all the results indicate that yeast promoters are large when compared to their prokaryotic counterparts (Faye et al. 1981; Struhl 1981a; Beier and Yound 1982; Guarente et al. 1982a; Struhl 1982a; Donohue et al. 1983; Guarente and Mason 1983; Martinez-Arias and Casadaban 1984; Siliciano and Tatchell 1984; Sarokin and Carlson 1984; Wright and Zitomer 1984). In every case, sequences more than 80 base pairs upstream from the mRNA start are critical for wild-type levels of transcription. Indeed, in some cases sequences as far as 450 base pairs away are implicated as upstream promoter elements. Moreover, the promoter regions of different genes are not equivalent in size. For example, the *pet56* and *mata* promoters are entirely included within a 100 base pair region immediately upstream from the RNA start (Siliciano and Tatchell 1984; Struhl 1985a), whereas the *suc2* promoter occupies a region of at least 450 base pairs (Sarokin and Carlson 1984). In contrast, all the *E. coli* promoter elements are located less than 45 base pairs away from

the RNA start, and the promoter regions are essentially identical in size. The observations that critical upstream sequences of yeast and other eukaryotic promoters are located far away from the transcription initiation site provided the first indications that these upstream elements are not direct sites of interaction for RNA polymerase II.

2.3.2 Defining the Boundaries of Upstream Promoter Elements

Sequential 5′ deletions define the upstream boundary of the promoter element that is furthest upstream of the transcription initiation site. However, in interpreting the results of sequential 5′ deletion analysis, it is important to remember that deletion mutants cannot be viewed simply as truncated DNAs. Instead, the deletion break points actually represent fusion points between the gene of interest and some unrelated sequence. The conclusions listed here are valid only if these unrelated fusion sequences do not contain any functional promoter elements. Moreover, the fusion creates a novel joint, which fortuitously could serve as an upstream element. These considerations are not mere academic exercise. It has been clear from early experiments that yeast genes are subject to position effects: that is, the phenotype of a particular deletion mutant depends on the fused sequences that are adjacent to the deletion break point (Struhl 1981b). Moreover, the sequences responsible for these position effects are located more than 300 base pairs from the RNA coding sequences.

Although in any individual case it is difficult to eliminate such artifactual possibilities, several arguments have been advanced that increase the validity of the conclusions described. First, in most experiments, the same unrelated sequence is fused to all the deleted derivatives of a given gene; thus, its effects should be constant except when a fortuitous novel joint is generated. In such cases, when sequential deletions demarcate a single region that separates wild-type derivatives from damaged ones, the clear implication is that this region is critical for transcription initiation. Second, in the initial experiments on the *his3* gene, a consistent phenotypic pattern was observed even though the fusion sequences for individual deletion mutants usually differed. Subsequent experiments employing a constant fusion sequence defined the same region as being part of the upstream promoter element. Thus, with appropriate controls, sequential 5′ deletion analysis defines the upstream boundary of the upstream promoter element. The precision with which this boundary is determined depends upon the distance between the end points of the two critical deletion mutants.

To define the downstream boundary, sequential 3′ deletion analysis is employed. However, as will be discussed in later sections, upstream promoter elements are necessary but not sufficient to constitute a fully functional yeast promoter. At least two other elements, the TATA box and the initiation region, are also important. Thus, to assay for a functional upstream element, the other promoter elements must be present. These other elements can be

derived from the gene of interest (in which case the sequential 3′ deletions are actually internal deletions within the gene), or they can be derived from a different gene (in which case they are termed *promoter fusions*).

2.3.3 Upstream Elements Are Defined by Short Sequences

From the best defined cases, it is clear that the upstream elements are relatively short stretches of DNA. For example, the *cyc1* (Guarente et al. 1984), *gal1,10* (West et al. 1984; Giniger et al. 1985), *his3* (Struhl 1982b, 1985a), *his4* (Donahue et al. 1983; Hinnebusch et al. 1985), *leu2* (Martinez-Arias and Casadaban 1984), and *mata* (Siliciano and Tatchell 1984) upstream elements are entirely located within 15 to 40 base pair regions. At this stage, it is not clear how much and which parts of these small regions are actually critical determinants. Such information awaits analysis of point mutations within the upstream elements.

The same conclusion was also reached from two separate lines of experiments. In both cases, His$^+$ revertants were selected from strains containing *his3* promoter mutations. In the first example, the *his3* promoter mutation was located on a circular minichromosome. The revertants are caused by DNA rearrangements within the minichromosome molecule, which probably arose by breakage of dicentric molecules (Scherer et al. 1982). Surprisingly, *his3* expression in these revertants was often under novel control mechanisms (Scherer 1985). Thus, the minichromosome contains a number of DNA sequences that can act as upstream elements when juxtaposed to the RNA coding region. In the second example, the *his3* promoter mutations were located in the chromosome, and the revertants were due to unlinked suppressor mutations (Oettinger and Struhl 1985). Again, *his3* expression was subject to novel control mechanisms. It is interesting that the suppressors are allele specific in that they restore expression for only one particular *his3* promoter mutation. From this and a number of other considerations, the interpretation of these suppressor mutations is that they activate transcription from cryptic upstream promoter elements (Oettinger and Struhl 1985). Since, in both sets of experiments, the revertants were relatively easy to obtain, cryptic upstream elements must occur frequently. This means that such elements must be defined by short DNA sequences.

2.3.4 Upstream Promoter Elements Confer Promoter Specificity

Given that upstream elements for a number of different genes have been defined in functional terms and localized to small regions, it is possible to compare their nucleotide sequences meaningfully. The result is that elements for different genes do not, in general, share common sequences. The analogous situation in *E. coli* is quite the opposite in that the upstream promoter element, the −35 sequence, is fairly well conserved among all genes. The implication from this result is that different proteins act at these different

upstream elements. This provides extremely strong support for the suggestion that RNA polymerase II does not interact directly with upstream regions. In more biochemical terms, it suggests that yeast cells have a large number of specific transcription factors that interact with specific upstream elements.

If transcription initiation is mediated by different proteins acting at their cognate sites, one would expect that the upstream elements also serve as regulatory sequences. Indeed, for most of the genes that have been studied, the upstream promoter elements are necessary for transcription, but only under particular physiological conditions. For example, the *gal1, 10* element is responsible for high levels of RNA synthesis in a medium containing galactose but not glucose as a carbon source (Guarente et al. 1982a; Johnston and Davis 1984; West et al. 1984; Struhl 1984). Similarly, the *cyc1* element activates transcription as a function of oxygen and intracellular heme levels (Guarente and Mason 1983; Lowry et al. 1983; Guarente et al. 1984), the *adr1* and *suc2* elements are effective in many circumstances, but not in glucose medium (Beier and Young 1982; Sarokin and Carlson 1984), and there are many other examples. From these observations and others to be discussed later, it now appears that the basis for promoter specificity and transcription control generally resides in the particular nucleotide sequences of different upstream promoter elements.

2.3.5 Poly dA:dT Sequences Act as Upstream Promoter Elements for Constitutive Expression

Most yeast genes under study have been chosen because their expression is regulated in some interesting manner. However, as mentioned in the introduction, many genes are expressed at the same level under all conditions; that is, they are not regulated. Thus, these studies may represent the special cases, and consequently, they may have overlooked a typical upstream promoter element.

Indeed, it appears that some upstream promoter elements are not regulatory sites. The best examples of this occur in the *his3* and *his4* genes, both of which are subject to transcription regulation as a function of amino acid starvation. Deletion of either upstream promoter element reduces transcription levels but does not affect regulatory ability (Struhl 1982a; Donohue et al. 1983). Thus, unlike the situation for other genes, it is possible to separate functionally promoter function from regulatory function. Of course, it is possible that these upstream promoter sequences specify a form of regulation that is unknown at the present time.

For at least three different genes, naturally occurring stretches of poly(dA-dT) serve as upstream promoter elements for constitutive expression (Struhl 1985a). The upstream element necessary for the normal basal level of *his3* expression is defined within a 17 base pair region that contains 15 dT residues in the coding strand. As will be described later, this region is distinct from the regulatory site that responds to amino acid starvation. Sequential 5' deletion analysis indicates that this same 17 base pair poly(dA-dT) region

serves as the upstream element for *pet56*, a gene that is located 200 base pairs away from *his3* and is transcribed in the opposite direction. In the case of *ded1*, deletion of a 34 base pair region that contains 28 dT residues in the coding strand significantly reduces transcription below the wild-type level. It is worth noting that in terms of mRNA molecules per cell, constitutive *ded1* levels are five times higher than those of *his3* or *pet56*. This suggests that longer stretches of poly(dA-dT) are more effective upstream promoter elements. Although it is not known how many yeast genes use poly(dA-dT) sequences as upstream elements, such homopolymer regions are often found at appropriate positions with respect to transcription initiation sites.

The influence of long poly(dA-dT) sequences was first observed in *cis*-acting mutants of the *adr2* gene that constitutively overproduce the gene product (Russell et al. 1983). Normally, *adr2* expression is very low in glucose medium but high in ethanol medium, whereas in the mutants, *adr2* expression is high under both conditions. DNA sequence analysis of two of these mutants indicates that the 20 base pair (dA-dT) sequence located around −200 has been expanded to a 54 or 55 base pair homopolymer stretch. Thus, these abnormally large poly(dA-dT) tracts cause high constitutive *adr2* expression, presumably by acting like natural poly(dA-dT) tracts that are actually used as upstream elements for wild-type genes.

There are two mechanisms by which poly(dA-dT) sequences might activate transcription. One possibility is that by analogy to proposed transcription factors that recognize different upstream elements, a specific protein recognizes poly(dA-dT) regions. A more attractive suggestion is that the transcription machinery recognizes the unusual structure of poly(dA-dT) sequences. Such sequences have a helix repeat of 10.0 base pairs instead of the normal 10.5 (Peck and Wang 1981; Rhodes and Klug 1981), and they are associated with kinks in DNA (Marini et al. 1982). Of particular interest is the observation that poly(dA-dT) regions prevent nucleosome formation *in vitro* (Kunkel and Martinson 1981; Prunell 1982). In this view, poly(dA-dT) sequences behave as constitutive upstream elements because they do not require specific transcription factors for their activity.

2.3.6 Upstream Elements Are Functional in Both Orientations

In general, the nucleotide sequences of yeast upstream promoter elements do not show twofold rotational symmetry and, thus, are directional with respect to the transcription initiation site. To examine whether this directionality is an essential aspect of promoter function, several upstream elements have been inverted with respect to the rest of the promoter region (Guarente and Hoar 1984; Struhl 1984; Hinnebusch et al. 1985). In all derivatives where such elements have been inverted, the genes are transcribed with equal efficiency as compared to derivatives with the normal orientation. Moreover, in the cases of the *cyc1* and *gal1,10* elements, transcription is subject to the normal regulation and is initiated from the proper site (Guarente and Hoar 1984;

Struhl 1984). The *cyc1* example is particularly instructive because in the normal yeast chromosome it is the only gene that could be dependent upon this upstream element. This clearly indicates that the element acts bidirectionally even though in the normal yeast cell, there is no particular reason why it should do so.

Yeast can use bidirectional upstream elements to activate coordinately two divergently transcribed genes. The invertible *gal1, 10* element is normally located between two divergently transcribed and coregulated genes. In principle, it is difficult to tell if there is one element that functions bidirectionally or if there are two elements, each functioning in only one direction. However, fine-scale deletion analysis and inversion of small DNA fragments tends to support the one-element model (West et al. 1984; Giniger et al. 1985). The divergently transcribed *mata1, a2* genes represent another example in that a 12 base pair region is critical for the expression of both genes (Siliciano and Tatchell 1984). Although these genes are subject to many coordinate controls, it is not known if this short region serves as a regulatory site. A somewhat different version of this phenomenon is represented by the divergently transcribed *his3* and *pet56* genes (Struhl 1985a). Although these genes have different cellular functions and are not coordinately controlled, deletion analysis of each gene implicates the same 17 base pair poly(dA-dT) region as the upstream promoter element for constitutive transcription.

It has been suggested that upstream promoter elements for different genes are recognized by specific transcription factors. As is discussed in Section 2.3.7, other sequences besides the upstream element are critical for transcription initiation. Clearly, such sequences must be recognized by proteins, and it is highly likely that such proteins are different from those that interact with upstream elements. What, then, is the relationship between the proteins that are associated with different elements of the promoter? Without biochemical data, one can only speculate. Nevertheless, the inference from the observation that asymmetric upstream elements function bidirectionally tends to argue against specific and direct protein-protein interactions. Given that a protein has to recognize an asymmetric sequence, it is hard to imagine how a bound protein could be symmetrically disposed to the rest of the promoter. And if a protein is bound asymmetrically, its contacts to proteins associated with other sequences must be different if one element is inverted with respect to the other. It is possible that a protein dimer could provide the necessary symmetry for bidirectionality, but it seems unlikely that such a dimer could interact symmetrically with an asymmetric sequence. If this speculation is correct, a more likely mechanism by which upstream promoter elements exert their effects is that the bound protein alters the DNA structure in the vicinity of the element, and that this structural change is responsible for transcription activation.

2.3.7 Upstream Elements Act at a Long and Variable Distance

As mentioned in the introduction, transcription initiation in *E. coli* can be viewed as an enzyme-substrate interaction between RNA polymerase and

promoter DNA sequences. Accordingly, the spatial relationships between both promoter elements and the transcription initiation site are determined rather precisely. Mutations that change the spacing between both promoter elements even by a single base pair can have major effects on promoter function.

The striking feature of yeast and other eukaryotic promoters is that the location of upstream elements with respect to the TATA box and the RNA startsite is not a critical determinant of transcription initiation. One indication of this general phenomenon is that upstream elements for different genes are located at various distances from their structural sequences. Anecdotal evidence accumulated over the past few years is that promoter fusions between the upstream element of gene A and the downstream region of gene B are generally functional even though the spacing is not constant. Indeed, the fusion points are usually chosen by fortuitous restriction enzyme sites.

In several cases, the spacing question has been addressed in a more systematic fashion. Starting with a wild-type gene or with a particular promoter fusion, the upstream element was moved relative to the RNA startsite by deleting or inserting DNA. The inherent problem in such experiments is to distinguish between the effects of spacing per se and those caused by specific sequences that are deleted or inserted. In the case of the *his3* element, variation in spacing was achieved by inserting multiple, tandem copies of a short sequence that by itself has no phenotypic effect (Struhl 1982c). Studies of the *gal1,10* and *cyc1* upstream elements accomplished this by taking advantage of the fact that these elements activate transcription only under certain environmental conditions (Struhl 1984; Guarente and Hoar 1984). Since these elements are highly specific DNA sequences, it is extremely unlikely that inserted DNA would contain such a sequence or that a deletion event would be able to create a new one. Thus, by testing any derivative for the proper regulatory response, it is possible to follow the element of interest. The results are that in the appropriate medium, moving either of these upstream elements over a range of hundreds of base pairs neither affects the level of transcription nor alters the transcription initiation site. These upstream elements can exert their effects over a range of at least 600 base pairs, and this may not be the upper limit.

The conclusion from these experiments is that the upstream elements act at a long and variable distance and that the spacing between elements is not critical for function. This conclusion provides further support for the idea that upstream elements are not direct recognition sites for RNA polymerase. This property, along with that of bidirectionality, also suggests that the proteins that bind to the different upstream elements do not mediate their effects by direct contacts to proteins that presumably bind to other parts of the promoter.

2.3.8 Upstream Elements Are not Functional when Placed Downstream from the Initiation Site

Experiments with the *gal* (Struhl 1984) and *cyc1* (Guarente and Hoar 1984) upstream elements indicate that although they can confer their functional

effects at various locations, these locations must be upstream from the RNA initiation point. When these elements are placed in a downstream position, no transcription activation is observed.

Several considerations indicate that the result is due to an inherent property of these elements and not to experimental artifact. To eliminate possible distance effects, the elements were placed close to the transcription initiation point, well within their active ranges. The *gal* element was located within the *his3* structural gene either 100 or 330 base pairs downstream from the initiation site. In the *cyc1* experiment, the element was placed 200 base pairs from the initiation site but within an intervening sequence that would be spliced out after the primary transcription event. Specifically, a fragment containing the *rps51* intervening sequence and some flanking coding sequence was inserted in frame near the 5′ end of a *cyc1-lacZ* fusion. In this way, the expression level could be monitored by *lacZ* enzyme activity. In both experiments, it was important to prove that the observed lack of transcription was not due to unstable RNA species or other effects of the unusual constructs. This was accomplished by placing a second upstream element in an upstream position. The result was that these novel alleles could be transcriptionally activated, but only in the presence of the second upstream element. These controls all demonstrate that transcription activation could have been detected in the relevant molecules. The fact that none was observed indicates that these upstream elements do not function when placed downstream from the site of transcription initiation.

Enhancer elements of mammalian promoters generally activate transcription when placed downstream from the initiation site (see Chapter 3). The results described in this section suggest that yeast upstream promoter elements may differ from enhancers. However, it is possible that yeast elements are intrinsically able to function in a downstream position, but that their effects are blocked by specific sequences in the initiation region. Suggestions that this may be the case are discussed later.

2.3.9 Sequences Block the Effects of Upstream Elements

Although it is clear that upstream elements can function at a long and variable distance, anecdotal evidence from yeast and other eukaryotic genes suggests that they become less effective in activating transcription as they become located further away from the RNA startsite. Such observations can be explained by postulating either that upstream elements are less effective as a function of distance per se, or that the transcription effects of upstream elements are blocked by specific sequences, which would occur more frequently as the distance increases.

The clearest demonstration of the blocking hypothesis comes from experiments in which the *E. coli lexA* operator site was inserted at various locations betwen the *gal1, 10* upstream element and its TATA box (Brent and Ptashne 1984). In these circumstances, the *lexA* operator has only a slight influence on

gal1, 10-dependent transcription. However, when such yeast cells produce the *lexA* protein (by virtue of an appropriate expression vector), transcription is decreased five- to tenfold. Therefore, transcription initiation is inhibited by a heterologous protein that is bound between an upstream promoter element and the RNA startsite. This transcription inhibition is probably not due to steric blocking of the presumptive *gal1, 10* transcription factor because the *lexA* protein has no effect when its operator is located extremely close to, but upstream of, the *gal1, 10* element.

Two somewhat less defined examples provide additional evidence. In one experiment, the *gal1, 10* element was fused to a set of derivatives containing increasing amounts of *his3* upstream sequences (Struhl 1984). If distance alone were important, the expected result would be that the level of *gal1, 10*-dependent transcription would decrease gradually as it moved further away from the *his3* startsite. Instead, the element activates transcription at its maximum level at all distances less than 380 base pairs from the RNA start, but at distances of 430 base pairs or more, the element is five- to tenfold less effective. This suggests that some sequence within a 50 base pair interval inhibits functioning of the *gal1, 10* element; within this interval lies the transcription initiation region for the *pet56* gene. In the other example, the placement of a presumptive transcription terminator downstream from the *gal1, 10* region inhibits transcription five- to tenfold (Brent and Ptashne 1984). A mutated derivative of this terminator does not confer this inhibitory effect. Given that there is no evidence for transcription initiation from the *gal* element itself, it seems unlikely that this result is due to termination per se. A better explanation for this observation and that of the *gal-his3* fusions is that transcription inhibition is caused by proteins that recognize these specific sequences.

The implication from all these results is that something "moves" from an upstream element toward the rest of the promoter. The question, of course, is: What is moving? Is it the transcription factor that recognizes the upstream element, RNA polymerase II, or some other protein? Or is the apparent movement indicative of a structural change, such as local supercoiling, which begins at the upstream element and is propagated bidirectionally?

Besides being of mechanistic interest, the results described above are worth mentioning for other reasons. First, the fact that bound proteins block transcription activation by upstream elements means that mutants that relieve this block should be due to a lack of binding. Thus, for any given sequence of interest, it should be possible by a positive genetic selection to isolate mutants that have a defective site or, more important, a defective protein that recognizes the site. Second, the existence of blocking sequences, particularly those defined by initiation and termination regions, provides a way for yeast to organize its genes into autonomous functional domains. This is important for an organism whose genes are closely packed, and whose upstream can function at long and variable distances.

2.4 THE TATA PROMOTER ELEMENT

Upstream elements have been the major focus of the work on yeast promoters. In part, this is due simply to the fact that they were the first elements to be defined. More important, however, these sites constitute much of the basis for the regulation of transcription initiation in response to appropriate physiological conditions. Nevertheless, as will become clear from the next two sections, the upstream elements are only one of the critical components of a yeast promoter.

The first indications of a second promoter element came from the same sequential 5′ deletion analyses used to define the upstream element (Faye et al. 1981; Struhl 1981a). In deletion mutants lacking the upstream element, the transcription level was reduced ten- to thirtyfold when compared to those derivatives that contained the element. More extensive deletions did not further reduce this level of expression until a new boundary was reached. Deletions beyond this boundary point lowered transcription by at least another factor of 10—that is, usually below the limit of detectability.

The proof that yeast promoters contain at least two distinct elements is that the upstream element is not sufficient to direct wild-type levels of RNA synthesis. Specifically, mutations for several genes were obtained that reduced expression approximately tenfold but that retained the intact upstream element (Struhl 1982c; Guarente and Mason 1983; Siliciano and Tatchell 1984). Thus, two separate regions of DNA are necessary to constitute a promoter, and neither region is sufficient by itself.

2.4.1 Localizing the TATA Promoter Element

In comparing the DNA sequences of eukaryotic promoter regions from a wide variety of organisms, the only common feature is a short (dA-dT) region whose canonical sequence is TATAAA. In higher eukaryotes, this sequence is invariably located 25 to 30 base pairs from the transcription initiation site. In yeast, the promoter regions are extremely AT rich, making it difficult to decide which among several candidates is the "best" TATA box. However, the distance from the RNA startsite usually averages around 60 base pairs, but it is much more variable and can apparently be located 100 base pairs away. It was, of course, expected from the beginning that the TATA sequence would indeed be a required promoter element. As is discussed below, it has been demonstrated in several cases that yeast TATA boxes define the downstream promoter element. For many genes, however, it is assumed without any evidence that particular TATA sequences are actually functionally important.

The upstream boundaries of TATA elements have been mapped by sequential deletion analysis. The DNA molecules used in these experiments all contain an intact upstream element, which may or may not derive from the same gene as the TATA box under study. The most common approach is to begin the deletion series at a point just downstream from the upstream

element, and to delete sequences gradually toward the initiation site (Guarente and Mason 1983; Siliciano and Tatchell 1984; Struhl 1984). As expected, deletions beyond a certain point greatly reduce transcription, thus defining the upstream boundary. In a related approach, a promoter region is disrupted at various locations by insertion of an oligonucleotide linker containing a site for restriction enzyme cleavage (Struhl 1982c). In this way, the promoter can be divided into upstream and downstream halves and then recreated in all possible pairwise combinations. The end result is a matrix of deletion mutants whose phenotypes can be compared to determine which sequences are critical and which are not.

The conclusion from such mapping experiments is that the upstream boundary of the TATA promoter element corresponds closely to the position of TATA sequences. The deletions that remove these TATA sequences are those which are functionally defective. This information strongly suggests that the TATA box is a critical component for transcription initiation *in vivo*, although sequences further downstream could also be essential.

Mapping the downstream boundary of the TATA box by sequential 3′ deletions is somewhat more complicated because the phenotypic assays all require an intact structural gene. Thus, the deletion series must begin at a position just upstream from the transcription initiation site and proceed upstream from there. Such an experiment has been performed only for the *his3* promoter, and the result is that the entire TATA element is located somewhere between nucleotides −32 and −52 (Struhl 1982c).

2.4.2 The TATA Element is Directional and Sequence Specific

The localization of the TATA promoter element assumes that (1) mutations that cause reduced expression are due to the deletion of important sequences and (2) sequences deleted in derivatives with wild-type phenotypes are not essential. However, deletion mutants not only remove specific sequences but they also change the spacing relationships between other regions. Thus, the transcription defects could be caused by such altered spacing relationships.

The fact that the distance between upstream and TATA elements is not a critical factor of promoter activity strongly suggests that phenotypes of TATA deletions are due to the removal of specific sequences. This was shown directly by experiments in which the various parts of the *his3* TATA region were replaced by a 31 base pair fragment of coliphage M13 DNA (Struhl 1982b, c). This foreign sequence was inserted in both orientations and in varying numbers of tandem copies. The result was that the M13 sequence could functionally replace the *his3* TATA box but only in one orientation. In the other orientation, the M13 sequence behaved as neutral DNA because it did not influence transcription when placed between the intact *his3* upstream and TATA elements. Therefore, since derivatives differing only by the orientation of the M13 fragment have identical spacing relationships but opposite phenotypes, the TATA element is directional and sequence specific.

2.4.3 The TATA Element is a General Feature of Yeast Promoters

Unlike upstream elements, which have different nucleotide sequences depending on the gene, essentially all yeast genes have TATA-like sequences in the promoter region. In all cases tested, these TATA sequences are contained within a region necessary for transcription initiation. Moreover, although the TATA element is necessary for transcription, upstream elements can be functionally interchanged (Guarente et al. 1982a; Beier and Young 1982; Struhl 1984). In such promoter fusions, the source of the TATA element does not in general affect the particular regulatory properties conferred by the upstream element. Finally, deletion mapping of the *his3* TATA box indicates that the same 20 base pair region is critical for transcription activation mediated by either the *his3* or the *gal* upstream element (Struhl 1984). Thus, the TATA element must interact with a protein that is a basic component of the process of transcription initiation. One obvious candidate, especially in view of the sequence homology with the *E. coli* – 10 promoter element, is that this protein is RNA polymerase II. However, in higher eukaryotic cells, a TATA binding protein that is easily separable from RNA polymerase II has been characterized (Davison et al. 1983; Parker and Topol 1984). Considerations to be discussed in Section 2.5 also suggest that a similar situation may be true in yeast.

2.5 ELEMENTS THAT SELECT INITIATION SITES

In the preceding sections, the focus has been on the level of transcription. However, analysis of a large number of yeast genes indicates that the 5′ ends of mRNA species are highly selected. Indeed, for many genes, transcription begins at a unique site, although there are several cases of extreme 5′ heterogeneity—for example, *cyc1* (Faye et al. 1979), *his1* (Hinnebusch and Fink 1983a), and *ura3* (Rose and Botstein 1983). In formal terms, it should be noted that the 5′ ends of mRNA species may not correspond to initiation sites. The possibility that such ends are produced by rapid RNA processing of randomly initiated transcripts has never been rigorously excluded. In this chapter, however, we shall assume that RNA end points correspond to sites of initiation. This assumption is supported by the fact that randomly initiated transcripts have not been observed *in vivo* under conditions where they should have been detectable, by *in vitro* transcription results obtained with higher eukaryotic genes, and by the fact that all genetic results presented earlier are internally consistent with this assumption.

What genetic elements are responsible for selecting where transcription is initiated? In higher eukaryotes, the TATA box is responsible for selecting the site (see Chapter 3). The evidence for this is that the TATA sequence is invariably located 25 to 30 base pairs from the start of transcription even when the normal initiation region is deleted and that deletion of the TATA sequence alters the initiation sites. In yeast, the situation is not so apparent. As discussed

in the previous section, TATA sequences tend to be found at various distances away from the initiation sites. However, in most of these cases, there is no evidence that any particular TATA-like sequence is functionally important, and generally there are several regions that at least resemble the canonical sequence.

5′ mapping of transcripts produced from derivatives lacking TATA elements have generally not been performed, largely because the transcription level is very low. However, in deletion mutants that remove the *mata*1a TATA box, the following results were obtained (Siliciano and Tatchell 1984). The wild-type gene initiates transcription from about seven different sites. In certain TATA deletion strains, transcription in general is decreased, and initiation from the upstream-most sites is relatively less than initiation from more downstream sites. This result is rather hard to interpret. The fact that there are multiple initiation sites may mean that there are several functional TATA-like sequences, only some of which are deleted. In addition, it is hard to assess the contributions of the new sequences that are in the same position as the deleted TATA box. In more general terms, in a situation where a critical sequence has been removed with serious functional consequences, it is impossible to know whether other aspects of the promoter are behaving properly.

Various derivatives of the *his3* promoter indicate that the TATA element is not sufficient to direct the initiation site (Chen and Struhl 1985). Specifically, the distance between the TATA region and the initiation site was altered by an 8 base pair insertion and by 4 and 16 base pair deletions. The sequences that are necessary and sufficient for TATA element function were present in all derivatives. The finding is that in all cases, transcription initiation is indistinguishable from the wild-type gene. Both the two normal initiation sites are observed at equivalent levels to the wild type. This result is in striking contrast to results on higher eukaryotic genes both *in vivo* and *in vitro*. In those cases, altering the spacing in a similar manner would cause initiation at new sites located the usual distance from the TATA element.

If the TATA element does not specify the initiation site, something else must do so. This question was addressed by constructing hybrid promoters consisting of the upstream and TATA elements from the *ded1* gene fused to the *his3* initiation region and RNA coding sequences (Chen and Struhl 1985). In such cases, transcription is initiated at the normal *his3* sites. This result demonstrates that sequences downstream from the *his3* TATA element are sufficient to direct accurate initiation. The upstream boundary of this initiator element is at nucleotide −10; the downstream boundary has not been determined. In the converse experiment, the *his3* upstream and TATA elements were fused to the *ded1* initiation region and structural gene. As expected, the correct *ded1* RNA startsites are used. Therefore, the initiation region itself is responsible for the selection of the proper transcription initiation sites.

These experiments also directly show that a precise distance between the TATA element and the initiation site is not critical for promoter function. The maximum distance appears to be about 90 to 100 base pairs. The best evi-

dence for this is that when the *ded1* TATA element is about 90 and 100 base pairs from the two normal *his3* initiation sites, the proximal site is preferentially used. When this TATA element is located closer, transcription is initiated equally from both sites. The minimal distance between the TATA element and the initiation site is less defined, although it is probably about 30 to 40 base pairs. Finally, it is worth noting that the initiation region can contribute to the overall level of transcription. Fusions between the *his3* upstream and TATA elements and the *ded1* initiation region are transcribed at higher rates than the normal *his3* gene.

The existence of an initiator element constitutes a major distinction between yeast promoters and higher eukaryotic promoters. This result is also of interest for other reasons. First, it suggests that a specific protein interacts with the initiation region. Supporting evidence for this idea comes from the *gal-his3* fusion experiment discussed earlier. The relevant result is that something within a 50 base pair region inhibited *gal* activation of *his3* expression (Struhl 1984). This region includes the initiation region for the *pet56* gene, which is transcribed in the opposite direction from *his3*, but it does not include the presumptive TATA box for this gene. Second, the existence of such an element may explain why yeast upstream promoter elements, unlike their counterparts in higher organisms, are unable to activate transcription when they are placed downstream from the RNA start. It may be that the initiation region, which of course is present, may block activation. In fact, in both experiments that demonstrated the inability of upstream elements to work in a downstream location, this possibility was tested (Guarente and Hoar 1984; Struhl 1984). Specifically, the experiment was to insert the initiation region between an upstream element and a test gene containing a TATA element. In both cases, these downstream consequences, which included the initiation region, blocked transcription activation. And third, despite many attempts, accurate yeast RNA polymerase II transcription has never been achieved *in vitro*. One speculation for this might be that the initiation process requires a TATA binding protein as well as an initiator binding protein. It might be harder to simulate a two-protein interaction than the single TATA binding protein interaction that is presumed to operate in higher eukaryotic cells.

2.6 TRANSCRIPTION REGULATION

By definition, transcription regulation means that RNA levels of a particular gene are higher in one circumstance than in another. There are two basic ways in which genes are transcriptionally regulated. In positive control, the critical regulatory protein(s) stimulates transcription initiation, whereas in negative control, the regulatory protein represses transcription that would otherwise occur. The distinction is important because it is expected from prokaryotic examples that the molecular mechanisms of positive and negative control will be different.

Since the critical step in transcription regulation involves an interaction between a regulatory protein and a target site in the gene of interest, the question of positive or negative control can be resolved by obtaining information about either component. However, it is generally easier to examine the regulatory site, because it must be located somewhere near the gene that has already been cloned. Two complementary approaches have been employed. The first method involves the isolation of mutations that functionally destroy the regulatory site. The phenotype of such regulatory mutations is that the gene of interest will be transcribed constitutively—that is, at the same level under the relevant physiological conditions. For positive control, the expression level will be low, whereas for negative control, the constitutive level will be high. In the second method, a small segment of DNA containing the regulatory site is fused to a test gene that is not subject to the relevant form of regulation. An increase in transcription of the test gene under the appropriate physiological conditions indicates a positive control site, whereas a decrease indicates a negative control site. In addition to determining the nature of regulatory sites, both methods are useful for localizing the critical DNA sequences.

2.6.1 Positive Control Sites

By definitions and experiments described, most of the regulatory sites that have been examined are positive control sites. Deletion of the sequences responsible for proper regulation of the *gal* (Guarente et al. 1982a; Johnston and Davis 1984; West et al. 1984), *cyc1* (Faye et al. 1981; Guarente et al. 1984), *adr1* (Beier and Young 1982), *suc2* (Sarokin and Carlson 1984), *his3* (Struhl 1982a), and *his4* (Donohue et al. 1983) genes reduces transcription under the appropriate physiological conditions. Furthermore, promoter fusions in which the *gal*, *his4*, and *adr1* elements are placed in front of unrelated genes result in increased transcription and proper regulation. The inference from these results is that the presumptive regulatory proteins that bind to these sequences stimulate the rate of transcription initiation.

In the cases of the *gal* and *cyc1* genes, the regulatory sites are equivalent to the upstream promoter sequences discussed earlier; that is, these sites are essential for transcription but only under the relevant environmental circumstances. Thus, in many cases, positive control reflects the fact that upstream elements are necessary for transcription, and that many different DNA sequences can suffice. The properties of these genetic elements and the molecular inferences that can be drawn from them have been discussed in detail in earlier sections.

Since the *his3* and *his4* genes are coregulated, it is not surprising that the regulatory sites have almost identical DNA sequences (Struhl 1982a; Donohue et al. 1983). However, in both genes, the regulatory sites do not correspond with the upstream promoter element. There are a number of mutations that greatly reduce the basal level and do not affect the regulation, and there are also mutations that eliminate any regulatory response but do not change the

basal transcription rate. Fine structure mapping indicates that the upstream promoter element and the regulatory site are distinct and separable DNA sequences. The properties of these regulatory sites have not been investigated in the same detail as for the *gal* or *cyc1* sites. Nevertheless, it seems that these sites are similar in that they function in both orientations and do not have a precise distance requirement (Hinnebusch et al. 1985).

2.6.2 Negative Control Sites

Although positive control is likely to be the major form of transcription regulation, several yeast genes are under negative control. In the case of the mating-type (*mat*) genes, yeast cells have three identical copies in the genome, but only the copy at the mating-type locus is expressed. Deletion analysis indicates that both the nonexpressed (silent) copies can be transcribed if particular DNA sequences are removed (Abraham et al. 1982, 1984). Thus, the simplest hypothesis is that silent copies are not expressed because they each contain a negative regulatory site. Initial experiments, which were performed with multicopy, circular plasmids, suggested that there were two sites (E and I) that were important in negative control of each silent copy. However, later analysis by the method of gene replacement indicates that only the E site is involved (Brand et al. 1985). The E regulatory elements are located approximately 1,000 base pairs from the genes under their control. Therefore, it appears that this negative regulatory site confers its effects over long distances. Repression by the E site is not confined only to the mating-type genes. When the mating-type genes are replaced by the *trp1* gene, *trp1* expression is now subject to this control mechanism (Brand et al. 1985). However, the *trp1* DNA fragment used in this experiment is identical to the one that is strongly influenced by position effects (Struhl 1981b); as discussed earlier, this fragment probably lacks an upstream promoter element.

A number of yeast genes involved in the metabolism of poor carbon sources are transcribed at reduced levels when glucose is present in the medium: this effect is termed *catabolite repression*. Analysis of promoter fusions indicates that the DNA sequences mediating catabolite repression are located upstream of the TATA element (Guarente et al. 1982; Beier and Young 1982; Sarokin and Carlson 1984). However, in most of these experiments, the catabolite repression site was fused to DNA segments that contain only the TATA element. Since such DNA segments are transcriptionally inactive, it is difficult to determine whether the glucose effect represents a true repression mechanism or simply the lack of a functional activator protein. Indeed, catabolite repression in *E. coli* represents a positive control mechanism because the CAP regulatory protein, together with its required cofactor cAMP, activates transcription (reviewed by deCrombrugghe et al. 1984). Since glucose-grown cells contain low levels of intracellular cAMP, catabolite repression is actually a consequence of the lack of a functional activator protein rather than the presence of a repressor.

Evidence that catabolite repression in yeast indeed occurs by a negative control mechanism comes from *gal-his3* fusions in which the catabolite repression sequence is placed upstream from the entire *his3* promoter region (Struhl 1985b). In glucose medium, the regulatory sequence overrides the *his3* promoter elements and reduces transcription below the normal basal level. Such a result cannot be explained by the lack of a functional activator protein because if this were the case, the *his3* promoters should have functioned normally. This conclusion is supported by deletion mutations of the *gal* segment that are defective in glucose repression yet fully functional in terms of galactose control (West et al. 1984).

There are two kinds of haploid yeast cells, **a** and α, which are distinguished by the genes that are expressed at the mating-type locus. From a series of genetic studies, it was proposed that the mating-type genes encode regulatory proteins that control the expression of target genes with **a**-specific or α-specific functions (MacKay and Manney 1974; Strathern et al. 1981). One specific suggestion was that the *a2* gene product is a repressor of **a**-specific gene functions. Analysis of the **a**-specific gene *ste6* indicates that its transcription is controlled by the *a2* gene (Wilson and Herskowitz 1984). Moreover, a sequence located about 200 base pairs from the RNA startsite is implicated in this regulation because deletion of this region abolishes *a2* control of transcription. In the presence of the *a2* product, deletion of the *ste6* regulatory region results in an increase of transcription, thus indicating that *ste6* expression is subject to negative control. The definitive evidence for negative control comes from the DNA binding properties of the *a2* protein (Johnson and Herskowitz 1985); these are discussed in the next section.

Prokaryotic repressor proteins inhibit transcription initiation by binding to sites that overlap the promoter region and thus interfere with functional RNA polymerase binding. In a similar vein, the SV40 T antigen binds to sites where transcription factors are known to interact (Rio and Tjian 1983). In a related but somewhat different situation described earlier, the *E. coli lexA* protein inhibits transcription activation in yeast when the *lexA* operator is placed between the *gal* and TATA promoter elements (Brent and Ptashne 1984).

In contrast, it appears that negative regulatory sites in yeast are located relatively far from the genes they control and, more important, that they confer their repressing effects even when located upstream of an intact promoter region. This was directly demonstrated for the catabolite repression site by the appropriate *gal-his3* fusions (Struhl 1985b). In the cae of the *ste6* gene, the negative control site appears to be located between the upstream promoter element and the TATA element. When the relative locations of the control site and the upstream element are reversed, it seems that the gene is still under negative control although the repression is less effective (Johnson and Herskowitz 1985). Thus, the *a2* operator site also exerts its effects when located upstream of an inact promoter region. The most dramatic example is seen in the negative control of the silent mating-type genes mediated through

the E site. In normal yeast cells, the E sites are located at least 1 kb from the divergently transcribed genes at the silent mating-type loci (Abraham et al. 1982). At the mating-type locus, the identical genes are transcribed by virtue of a bidirectional upstream promoter element that lies between them (Siliciano and Tatchell 1984).

The implication from these observations on negative control sites is that unlike the situation in *E. coli*, repression does not involve steric competition between the presumptive regulatory protein and the transcription apparatus. Although the nature of these repression mechanisms is unknown, there are clear similarities between the properties of negative and positive control sites. In both cases, the relevant elements act at long distances from the genes they control. Furthermore, it appears that they function in both orientations and that their precise location with respect to the mRNA coding region is not critical. The comparison is most easily seen in the experiments involving the *gal* DNA segment that specifies both galactose induction and catabolite repression. In a series of promoter fusions in which the spacing and the orientation of the *gal* segment are varied with respect to the target gene, these different control mechanisms are inseparable (West et al. 1984; Struhl 1984). Thus, even though the identical sequences are unlikely to be important for positive and negative control, the properties of the regulatory sites are similar. This might mean that positive and negative control may represent opposite sides of the same mechanism. If, for example, activator proteins alter chromatin structure to allow access of the basic transcription factors, then repressor proteins might cause the promoter region to become inaccessible.

2.7 REGULATORY PROTEINS

In all the previous sections, I considered the properties of promoter and regulatory elements. It is, of course, a basic assumption of molecular biology that functional specificity in DNA sequences must be associated with a corresponding specificity defined by a protein. A number of inferences about such proteins can be drawn from the genetic properties of the DNA sequence elements since these are the sites that the proteins presumably recognize. Nevertheless, a detailed molecular picture of transcription initiation and its regulation is impossible without purified proteins and *in vitro* transcription systems. These matters are the weak links in our understanding of yeast promoters.

Despite a number of attempts by competent investigators, no one has been able to obtain cell-free extracts that are capable of accurate transcription initiation. In contrast, *in vitro* transcription systems are available for many higher eukaryotic organisms. As discussed previously, the reason for this difference may involve the more complex requirements for accurate initiation.

In terms of proteins that recognize specific DNA sequence elements, genetic studies have suggested a number of candidates. These presumptive

regulatory proteins are defined by *trans*-acting mutations that abolish the ability of a particular gene(s) to be regulated properly. By using additional genetic and physiological criteria, some additional clues about the properties of these presumptive regulatory proteins can be inferred. However, the major problem with such analyses is that a regulatory mutation can cause its phenotype in many different ways, some of which may be very indirect. Without biochemical evidence, it is essentially impossible to determine if a mutation directly affects the transcription process by altering a specific protein-DNA interaction. Very recently, however, it has been demonstrated that three proteins do indeed bind to specific DNA sequences, the *α2* repressor protein and the *gal4* and *gcn4* positive regulatory proteins.

2.7.1 Binding by the *α2* Repressor Protein

From genetic studies, it was proposed that *α2* represses the transcription of genes such as *ste6*, which encode functions necessary for haploid **a** cells (Strathern et al. 1981; Wilson and Herskowitz 1984). The demonstration of this prediction, however, was performed in a novel manner (Johnson and Herskowitz 1985). As expected, the template for these binding studies was a recombinant DNA molecule that contained the *ste6* gene. However, the *α2* protein was not purified in its native form from a wild-type yeast strain. Instead, the analysis was performed with an *α2-lacZ* fusion protein that contained most of the *α2* structural gene at the N terminus and an enzymatically active β-galactosidase domain at the C terminus. This fusion protein confers *α2* function *in vivo* because it complements the defects of *α2* mutants. It was purified to near homogeneity on the basis of its β-galactosidase activity and then tested for its DNA binding properties.

In the first experiment, a plasmid containing the *ste6* gene was cleaved by restriction endonucleases into a number of DNA fragments and incubated with the *α2-lacZ* protein. Using a nitrocellulose filter binding assay, it was shown that the protein bound specifically to one DNA fragment. Moreover, this DNA fragment included sequences immediately upstream of the *ste6* mRNA coding region. By using a number of different restriction enzymes, the binding site was localized to a region about 200 base pairs from the transcription initiation site. Confirmation of these results and further details about the binding site were obtained from DNase I footprinting experiments. In subsequent experiments, it was demonstrated that extracts from wild-type *α* strains contain a protein that binds to the same DNA sequences. This protein has been purified through several columns, and it presumably represents the native *α2* protein. All these results indicate that *α2* is a specific DNA binding protein. Most important, the binding site determined by these *in vitro* experiments corresponds to the position of *ste6* sequences that are critical for *α2*-mediated regulation *in vivo*.

2.7.2 Binding by the *gal4* Activator Protein

It has been long suggested that transcription activation of the *gal7*, *gal10*, and *gal1* genes depends on specific DNA binding by the *gal4* gene product. As mentioned earlier, the target regulatory sites for the *gal1, 10* genes have been defined by deletion analysis. Two independent lines of experimentation now demonstrate that regulation of the galactose genes indeed depends on *gal4* binding to the appropriate regulatory sites (Bram and Kornberg 1985; Giniger et al. 1985).

In one set of experiments, *gal4* binding was demonstrated by *in vitro* methods similar to those described for the *α2* repressor (Bram and Kornberg 1985). Filter binding assays were used to identify and partially purify a protein that bound specifically to the cognate regulatory sites. Moreover, this binding activity was observed only in strains that were engineered to overproduce the *gal4* protein.

In the other set of experiments, *gal4* binding *in vivo* was demonstrated by methylation protection (Giniger et al. 1985). This is a standard method for analyzing DNA-protein contacts *in vitro*, and it is similar to DNase I footprinting except that the structural probe is a chemical agent rather than an enzyme. To adapt this method for *in vivo* studies, the experimental protocol was modified in the following manner. Yeast cells growing exponentially in galactose medium were treated with dimethylsulfate. Due to its small size, this methylating agent gets into living cells and modifies DNA, primarily at guanine residues. However, since the entire yeast genome is modified, the *gal1, 10* methylation sites had to be identified by indirect end labeling (also known as genomic sequencing) (Church and Gilbert 1984). The results of this experiment were that when compared to purified genomic DNA, specific guanine residues within the regulatory region showed altered reactivity under conditions of transcription activation. More important, these alterations were observed only in *gal4*^{1} strains.

To prove that these altered methylation patterns were due to *gal4* binding, the analogous experiment was performed in *E. coli*, using two specially designed plasmids. One plasmid contained the *gal1, 10* target sequences, while the other plasmid contained the *gal4* gene cloned into an expression vector. When both plasmids coexisted in *E. coli* cells, the *in vivo* methylation pattern showed the characteristic alterations observed during transcription activation in yeast. In contrast, these alterations were not seen when the *gal4* gene was deleted from the expression vector plasmid. Because *E. coli* represents a heterologous host organism, the alterations in methylation patterns almost certainly reflect the direct effects of the *gal4* protein. This finding is supported by the fact that methylation patterns constitute individual signatures of specific proteins. For example, the bacteriophage λ repressor and *cro* proteins recognize the same DNA sequences but nevertheless produce different methylation patterns (Johnson et al. 1978).

Thus, it is clear from these experiments that yeast has proteins that recognize specific DNA sequences. Furthermore, the comparison of *in vitro*

binding experiments with genetic identification of the regulatory sites indicates that transcription control is mediated by the direct action of these binding proteins to their cognate regulatory sites.

2.7.3 DNA Binding and Transcription Activation Are Separable Functions of the *gal4* Protein

In *E. coli*, positive regulatory proteins have separate domains for DNA binding and transcription activation. There exist mutants of the λ*cI* protein (Guarente et al. 1982b) and P22*cII* protein (Hochschild et al. 1983) that are defective in positive control although they bind to their recognition sites equally as well as their wild-type counterparts. The ingenious experiment described next demonstrates that DNA binding and transcription activation are separable functions for a yeast regulatory protein (Brent and Ptashne 1985).

The target gene is a derivative of a *gal-lacZ* fusion in which the *gal* upstream regulatory site is replaced by the *E. coli lexA* operator site. As expected, *lacZ* activity is not subject to galactose control even in the presence of the *lexA* protein (produced with an appropriate expression vector). In contrast, however, a *lexA-gal4* hybrid protein does confer galactose inducibility. Thus, the hybrid protein binds via the *lexA* protein-operator interaction, but it stimulates transcription by virtue of the *gal4* domain. In addition, the hybrid protein has lost its ability to bind to the *gal* upstream regulatory site. Therefore, specific binding to the *gal* regulatory site is not necessary for transcription activation by the *gal4* protein. This result also indicates that galactose regulation by the *gal4* protein occurs at the activation step rather than at the DNA binding step.

2.7.4 Binding by the *gcn4* Positive Regulatory Protein

Under conditions of amino acid starvation, transcription of many amino acid biosynthetic genes is coordinately induced above the basal level. Extensive deletion analysis of the *his3* (Struhl 1982 a, b; Struhl et al. 1985) and *his4* (Donahue et al. 1983; Hinnebusch et al. 1985) promoter regions, combined with DNA sequence comparisons of other coregulated genes (Hinnebusch and Fink 1983a), implicates the conserved TGACTC sequence as a regulatory site. Epistatic relationships among *trans*-acting mutations suggest that the *gcn4* gene product has the most direct role in the transcription regulation process (Hinnebusch and Fink 1983b; Penn et al. 1983).

The DNA binding properties of *gcn4* protein have been investigated by using a new and general method for analyzing protein-DNA interactions (Hope and Struhl 1985). Specifically, the *gcn4* protein coding sequences were cloned into a vector containing a promoter for SP6 RNA polymerase, RNA was synthesized by transcribing the template with this enzyme, and *gcn4* protein was synthesized as a pure ^{35}S-labeled species by *in vitro* translation of this mRNA. DNA binding activity was detected by incubating the labeled

protein with specific DNA fragments and separating protein-DNA complexes from free protein by electrophoresis in native acrylamide gels.

Four lines of evidence indicate that *gcn4* encodes a specific DNA binding protein that is involved in transcriptional regulation of amino acid biosynthetic genes (Hope and Struhl 1985). First, *gcn4* protein binds specifically to the promoter regions of four genes subject to general control (*his3, his4, trp5, arg4*), whereas it does not bind to analogous regions of four unregulated genes (*ded1, gal1,10, ura3, trp1*). Second, analysis of deletion mutants of the *his3* promoter region indicates that a 20 base pair region is necessary and sufficient for *gcn4* binding. This region corresponds precisely to the *his3* sequences that are critical for regulation *in vivo*; indeed, the same deletion mutants were tested both *in vitro* and *in vivo*. Third, DNase I protection experiments indicate that the TGACTC sequence within the *his3* regulatory region interacts directly with *gcn4* protein. Fourth, a synthetic *gcn4* mutant protein lacking the 40 C-terminal amino acids has no specific or nonspecific DNA binding activity; this correlates with a *gcn4* mutant gene (Hinnebusch 1984) that fails to induce transcription *in vivo*.

Thus, the *gcn4* gene, whose gene product is necessary for coordinate induction *in vivo*, encodes a protein that binds specifically to *his3* regulatory sequences and to promoter regions of other coregulated genes. Moreover, at the level of DNA sequences, there is a direct correlation between DNA binding *in vitro* and transcription activation *in vivo*. However a *gcn4* mutant protein consisting of the 92 C-terminal acids is fully capable of binding the *his3* regulatory sites *in vitro*; this protein is unable to activate transcription *in vivo* (I. Hope and K. Struhl, unpublished observations). Thus, it appears that like the *gal4* protein, DNA binding and transcriptional activation are separable properties of the *gen4* protein.

2.8 OTHER ASPECTS OF REGULATION

Most studies of eukaryotic gene regulation are focused on specific regulatory sequences and the proteins that recognize them. However, this approach overlooks a critical aspect of regulation, which from a biological perspective, may be the most interesting. In all forms of transcription 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 at least one of these physiological states can be converted to the other state; otherwise, the gene of interest will be transcribed at a constitutive level. If we consider the early stages of this change in physiological state, there must be a signal that initiates all the molecular mechanisms that result in altered transcription initiation. 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. The key point here is that the regulatory signal cannot be simply the regulatory protein that interacts

with the promoter region of interest. Another regulatory factor must be responsible either for the synthesis/degradation of the DNA binding protein or for the activation/inhibition of the DNA binding protein. Without this additional factor, there is no possible mechanism by which the change of physiological state can be interpreted in such a manner as to affect transcription; that is, transcription control is executed by the regulatory proteins that bind to specific DNA sequences. However, there must be additional regulatory molecules that govern when these regulatory proteins execute their roles in transcription initiation. 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. The classic example of this aspect of gene regulation is the *E. coli lac* repressor that binds both to operator DNA and to lactose analogues.

The usual approach to these issues involves the isolation of mutations that eliminate the regulatory behavior of a given gene or set of genes. The next step is to determine the genetic properties of these regulatory mutations (*cis-trans* and dominant-recessive tests) as well as epistatic relationships between them. By these means, it is usually possible to obtain a formal description of a regulatory pathway. The problem with this approach is that the complexity of the pathway depends on the number of mutations isolated. Unfortunately, there are usually many indirect ways to alter the cell's physiological state. This makes it difficult to determine whether a regulatory mutation defines an interesting mechanistic feature or a side issue. To get beyond this point, other information is needed, which usually comes from an inspired guess or from a serendipitous observation.

A description of all yeast regulatory mechanisms under study could fill several review articles, especially since this would involve the vast subject of cellular metabolism and physiology. What I shall do here is to describe briefly three specific cases that exemplify different solutions to the general problem.

The situation that appears most analogous to the *E. coli lac* operon is the induction of *gal7, 10, 1* expression upon addition of galactose to the medium. As described in earlier sections, induction is mediated by binding of the *gal4* protein to target sites located upstream of the *gal* structural genes. From the observations that the *gal4* protein is expressed constitutively (Perlman and Hopper 1979) and that *gal* induction occurs extremely rapidly upon the addition of galactose (St. John and Davis 1981), it is likely that the signal molecule is galactose or some direct metabolite of it. The simplest view of galactose induction is that the *gal4* protein activates transcription only when it is physically associated with the signal molecule.

cyc1 regulation represents a related but somewhat different circumstance. This gene is expressed at high levels when cells are grown in nonfermentable carbon sources such as lactate or glycerol, but it is expressed much more poorly when cells are grown in fermentable carbon sources such as glucose or galactose. Unlike *gal* induction, this regulatory phenomenon can be achieved with a wide variety of compounds; thus, it is unlikely that the exogenously added agents (or their direct metabolites) are all signal molecules. Instead,

several lines of evidence suggest that intracellular levels of heme might be the signal (Guarente and Mason 1983; Guarente et al. 1984). First, *hem1* mutants, which are unable to synthesize heme, have extremely low *cyc1* levels. Second, exogenous addition of heme precursors to these mutants restores high *cyc1* levels. Third and most important, these heme biosynthetic mutants have high *cyc1* levels when nonmetabolizable heme analogues are added. This last fact strongly suggests that the low levels seen in the *hem1* mutants are a direct consequence of heme deficiency.

Physiologically speaking, heme is a sensible signal molecule. For cells to grow with nonfermentable carbon sources, they must carry out oxidative phosphorylation in the mitochondria. This process requires cytochrome c (encoded by *cyc1*) and its heme cofactor. Since this is not necessary for cells to grow with fermentable carbon sources, the level of intracellular heme is directly related to the cell's competence for oxidative phosphorylation. In terms of the molecular mechanisms of transcription control, the *cyc1* site specifying heme regulation has been identified by deletion analysis. Although there is no direct evidence that implicates a protein that binds to this regulatory site, the best candidate is the *hap1* gene product. *hap1* mutants are insensitive to normal *cyc1* control even when exogenous heme or heme analogue is added. Thus, the current view is similar to the *gal* situation in that the *hap1*-heme complex is the active DNA binding protein, whereas the apoprotein is inactive (Guarente et al. 1984).

It should be noted here that all these experiments critically depended on *cyc1-lacZ* fusions. As mentioned in the methodology section, fusions are valuable because the gene product being assayed does not influence the regulation of interest. Here, heme is essential for cytochrome c activity in addition to its role in *cyc1* transcription regulation. By measuring *lacZ* activity in strains that also contain the wild-type *cyc1* gene, complications in interpretation are avoided.

The final example, general control of amino acid biosynthetic genes, appears to occur by a novel regulatory mechanism. Unlike the situation in *E. coli*, individual biosynthetic pathways are not usually subject to specific transcription regulatory mechanisms. Instead, transcription of at least 30 (and probably 50 to 100) genes involved in amino acid biosynthesis is coordinately increased in response to general amino acid starvation (Schurch et al. 1974; Wolfner et al. 1975). This starvation response can be elicited by using a variety of metabolic poisons that inhibit the synthesis of any one of a number of different amino acids. The analogous physiological phenomenon in *E. coli* is stringent-relaxed control. Here, the probable signal molecule for amino acid starvation is guanosine tetraphosphate (ppGpp), which is an aberrant side-product of GTP hydrolysis produced by translationally stalled ribosomes (reviewed by Cashel 1975; Gallant 1979). It has been hypothesized that stringent-relaxed control is mediated by a direct interaction between ppGpp and RNA polymerase (Travers 1980). Information concerning the sites and

proteins involved in general amino acid control has come from genetic analysis of *cis*- and *trans*-acting mutations that abolish the regulatory phenomenon and from DNA binding experiments. As described earlier, the *gcn4* protein binds to the promoter regions of these coordinately controlled genes by recognizing the TGACTC regulatory sequences (Hope and Struhl 1985).

The insight into the mechanism of general control comes from the unusual structure of the *gcn4* gene (Thireos et al. 1984; Hinnebusch 1984). Unlike typical eukaryotic genes, which contain short RNA leaders prior to the AUG initiation codon, the *gcn4* leader is 600 bases in length. Furthermore, this leader contains several AUG codons that specify an incorrect reading frame for the *gcn4* protein. Studies of eukaryotic translation initiation, particularly in yeast, indicate that protein synthesis begins at the 5′ proximal AUG codon and that it cannot be reinitiated at more downstream AUG codons (Sherman et al. 1980). In other words, by all the normal rules of translation initiation, the *gcn4* protein should not be made. Indeed, analysis of *gcn4-lacZ* protein fusions demonstrates that this expectation is true under normal growth conditions. The surprising and revealing result is that the protein levels of this *gcn4-lacZ* fusion are 50 times higher under conditions of amino acid starvation, whereas the RNA levels are affected by no more than a factor of 5 (Thireos et al. 1984; Hinnebusch 1984). Thus, it appears that the basic rules of translation initiation are circumvented under conditions of amino acid starvation.

What appears to be novel in general amino acid biosynthesis control is that regulation does not occur by changing the activity of the DNA binding protein. Instead, transcription control is apparently achieved by altering the amount of the DNA binding protein. This translation control mechanism neatly explains how a metabolic signal that occurs under a variety of experimental conditions can be transmitted to the effector molecules that directly cause transcription regulation. Although the actual signal is not known, it is very likely to be related directly to the process of translation; in this way starvation for any amino acid would produce the signal. Thus, the presumptive translation effects that produce the signal may also be involved in the mechanism and control of translation initiation.

2.9 COMPLEX PROMOTER ORGANIZATION

For simplicity's sake, one usually discusses different aspects of promoter function on an individual basis. However, it is equally useful to consider the overall organization of particular promoters. In this section, I briefly describe the current views on four well-studied examples. Two matters of interest are the following. First, although genes in yeast are closely packed together, adjacent genes are usually unrelated and, hence, subject to different forms of regulation. Second, many individual genes are subject to more than one control mechanism.

2.9.1 Distinct Upstream Regulatory Sites of the *cyc1* Gene

Transcription of the *cyc1* gene is subject to catabolite repression in that RNA levels are approximately 20 times higher in lactate medium as compared to glucose medium. This regulation is mediated by either of two separate regions of DNA, which have partial sequence homology and which are located approximately 275 and 210 base pairs from the transcription initiation site (Guarente et al. 1984). In lactate medium, these upstream elements are equally efficient in promoting *cyc1* transcription. However, in glucose medium, almost all *cyc1* transcription is due to the upstream site (UAS1); therefore, the downstream site (UAS2) is a more efficient repression site. A single base pair mutation in UAS2 weakens the repression such that transcription activity in glucose medium is increased ten- to twentyfold. *cyc1* transcription requires heme as a cofactor because mutants unable to synthesize heme are completely defective in *cyc1* transcription. As described earlier in the section on promoter fusions, this heme effect is not due to its interaction with cytochrome c protein.

Although these two sites are fairly homologous in DNA sequence, their regulatory properties are distinctly different (Guarente et al. 1984). The repression effect mediated by UAS1 is abolished by adding nonmetabolizable heme analogues to the growth medium. Repression mediated by UAS2 is refractory to such environmental intervention. Since these analogues act as gratuitous inducers, it is likely that UAS1 but not UAS2 mediates catabolite repression by responding to intracellular heme levels. A further distinction between these two regulatory sites is apparent from *trans*-acting mutations. *hap1* mutations abolish transcription dependent upon UAS1, while *hap2* mutations greatly reduce transcription dependent upon UAS2. It is possible that the *hap* genes encode proteins that interact with the individual control sites. *hap1*, in particular, is an excellent candidate because *hap1* mutants are uninducible even with heme analogues.

2.9.2 Constitutive and Regulatory Promoters for *his3* and *pet56*

In the normal yeast genome, *his3* and *pet56* are adjacent genes that perform unrelated functions. They are transcribed in opposite directions from initiation sites that are separated by only 200 base pairs. Under normal growth conditions, both genes are transcribed at a similar basal level. Although each gene has its own TATA element, a 20 base pair region of poly dA:dT located between the genes serves as the upstream promoter element for both (Struhl 1985a). Thus, this constitutive element acts bidirectionally to activate transcription of two unrelated genes.

Along with many other genes involved in amino acid biosynthesis, *his3* transcription is induced under conditions of amino acid starvation. This regulation depends on an upstream site located between the *his3* and *pet56* TATA elements. Thus, constitutive and regulated expression of the *his3* gene depends on two separate upstream elements. However, although the *his3*

regulatory sequence functions in both orientations and is repeated several times in the divergent promoter region, *pet56* expression is not subject to amino acid regulation.

Several lines of evidence indicate that the constitutive and regulatory promoters for *his3* expression are qualitatively different (Struhl et al. 1985). First, under normal conditions, *his3* transcription is initiated equally from two different sites (*1* and *12*). However, *his3* induction is due entirely to increased transcription from the *12* site. The same selectivity of initiation sites is observed during galactose induction of *gal-his3* fusions (Struhl 1984) and in revertants of *his3* promoter mutations that lack the normal upstream promoter element (Oettinger and Struhl 1985). These revertants, which are due to *ope* suppressor mutations, express *his3* in a regulated manner but only from the *12* site. Second, attempts to place the *pet56* gene under galactose control by standard promoter fusions were unsuccessful. Thus, the *pet56* transcript, like the *his3* transcript initiating at *1*, is not activated by upstream regulatory elements. Third, small deletions in the *his3* TATA region abolish regulation but not constitutive expression (Struhl 1982a). Fourth, in nuclear chromatin, the *his3* TATA region is hypersensitive to micrococcal nuclease (Struhl 1982b). This structural feature depends specifically on the presence of the *his3* constitutive upstream element. It is not observed when *his3* transcription is mediated by the *gal* element or by the *ope* suppressor mutations. Thus, this feature of chromatin structure is not associated with transcription but with transcription that initiates at *1* (Oettinger and Struhl 1985; Struhl et al. 1985). The explanation for these observations is that constitutive and regulatory promoters are distinguished not only by their upstream elements but also by the TATA elements and by initiation sites.

2.9.3 Short- and Long-Range Regulation of the Mating-type *α* Locus

Yeast comes in three cell types—**a** haploids, *α* haploids, and **a**/*α* diploids—that are distinguished by a number of biological characteristics. The control of cell type is determined by genes at the mating-type locus. The *α* locus contains two genes, *α1* and *α2*, which are proposed to encode regulatory proteins that control the transcription of many target genes that execute cell-type specific functions (MacKay and Manney 1974; Strathern et al. 1981). As described previously, the *α2* protein indeed binds specifically to target promoters of **a**-specific genes and represses their transcription. From genetic experiments, *α1* is presumed to be a positive regulatory protein that activates transcription of *α*-specific genes. In **a**/*α* diploids, *α2* is necessary for sporulation, whereas *α1* is not.

The *α1* and *α2* genes are transcribed divergently from initiation sites that are separated by approximately 200 base pairs (Siliciano and Tatchell 1984). As in the *his3-pet56* example, a single upstream promoter element, defined within the sequence ATGATGTCTG, is essential for the transcription of both

genes. Since this sequence is short, asymmetric, and unique to the *mata* genes, it follows that the element is recognized by a specific DNA binding protein that functions bidirectionally. In **a**/α diploids, transcription of both genes is reduced. This diploid regulation is mediated by a separate regulatory element that acts bidirectionally to repress transcription (Siliciano and Tatchell 1984). As expected from the roles of *mata* gene products in sporulation, $\alpha 1$ RNA levels are repressed more strongly than $\alpha 2$ levels. This observation is probably accounted for by the fact that the diploid regulation site lies between the upstream promoter element and the $\alpha 1$ TATA box, whereas it lies upstream of the intact $\alpha 2$ promoter. Both the upstream promoter element and the diploid regulatory site represent short-range control of *mata* transcription.

In terms of long-range control, the key fact is that although haploid yeast cells have three copies of mating-type information, only the copy at the mating-type locus is transcribed. With the exception of a 700 base pair region that distinguishes *mata* from *mata*, the copies have identical nucleotide sequences that extend for 2,400 base pairs (Nasmyth et al. 1981; Klar et al. 1981). This means that the sequences that determine whether the mating-type genes are active or silent must lie outside the common region. As described in Section 2.6.2, the silent copies are not expressed because they contain a negative regulatory site (Abraham et al. 1982). These E sites are located more than 1,000 base pairs from the divergent control regions of the silent copies; they are not found at the mating-type locus. Moreover, since there is only one E site for each silent copy, the E site must confer its repressive effect both when it is far upstream and when it is far downstream from the genes under its control. This regulatory site is distinguished from most others by the very long distances involved, and by the ability to function even when downstream from the promoter region.

2.9.4 Cell-cycle and Mother-daughter Control of *HO* Transcription

In addition to the complex regulatory behavior described in the preceding section, haploid yeast cells can interconvert between **a** and α mating types. This is accomplished by specific genomic rearrangements in which the *mat* cassette at the mating-type locus is replaced by a cassette that previously resided at a silent locus (Hicks et al. 1977). Since only the copy at the *mat* locus is transcriptionally active, cell-type switching occurs when the incoming copy carries a different **a** or α allele from the evicted copy. In normal yeast cells, mating-type interconversion is a rare event because it occurs by gene conversion. However, in yeast strains containing the *HO* (homothallism) gene, this process occurs as frequently as every cell division cycle. The *HO* gene encodes a site-specific endonuclease that initiates the interconversion process by producing a double-stranded break at the mating-type locus (Kostriken et al. 1983).

Detailed pedigree analysis of homothallic strains indicates that mating-type interconversion occurs early in the cell cycle and that mother but not

daughter cells are capable (Strathern and Herskowitz 1979). In addition, this process is restricted to haploid cells in that **a**/α diploids do not switch to homozygous **a**/**a** or α/α diploids. As might have been expected, this cell lineage pattern is determined largely by the expression of the *HO* gene (Jensen et al. 1983; Nasmyth 1983). Specifically, the *HO* gene is transcribed (1) early but not late in the cell cycle, (2) in mother but not daughter cells, and (3) in haploid but not diploid cells. Although analysis of promoter/regulatory elements for *HO* transcription is not advanced as for several other genes, several features are already clear.

First, at least 1,400 base pairs upstream from the mRNA coding sequences are required for proper regulation (Nasmyth 1985a). Transcription depends primarily on a TATA element as well as an upstream region located 1,000 to 1,400 base pairs from the initiation site. A deletion mutant that removes the DNA between these two regions still shows mother/daughter and haploid/diploid control of *HO* transcription. It is likely that regulatory sites for these phenomena are located in the upstream region, although their relationship to each other and to a possible promoter element is unclear.

Second, although DNA sequences between the upstream region and TATA element do not affect overall transcription levels, this region is critical for some aspects of cell-cycle control (Nasmyth 1985b). In mutants lacking this region, *HO* is transcribed at times when it should not be early G1 prior to start and also late in the cell cycle. However, these mutants do retain other features of cell-cycle control. A small oligonucleotide sequence (consensus PyCACGAAAA), which is repeated 10 times in the deleted region, plays some role in cell-cycle regulation. When synthetic copies of this sequence replace the entire region between the upstream and TATA elements, *HO* transcription is properly regulated in early G1 but not late in the cycle. The repeated nature of this sequence is important because the regulation becomes more authentic as the number of inserted copies is increased. Moreover, deletions of the native gene that remove most but not all copies have little phenotypic effect.

Third, deletion analysis also indicates that the region between the upstream and TATA elements also includes regulatory sites for haploid/diploid control. At least two separate subregions are sufficient for this regulation. Within these subregions is a nucleotide sequence that, with some variation, is found nine times upstream of the *HO* gene and in front of coregulated genes *mata1* and *ste5* (Siliciano and Tatchell 1984). Thus, the presumptive haploid/diploid and cell-cycle regulatory sites are interpreted between the upstream promoter region and the TATA box.

2.10 MOLECULAR MECHANISMS: INFERENCES AND SPECULATIONS

From the experiments described in this chapter, the most general conclusion is that yeast promoters are very different from their prokaryotic counterparts. It follows directly that the molecular mechanism of transcription initiation

and its regulation must be qualitatively different. The properties of yeast promoter/regulatory elements are summarized schematically in Figure 2–4, and the conclusions derived from them are as follows:

1. It is clear that transciption initiation in yeast cannot be viewed as a simple enzyme-substrate interaction between RNA polymerase II and promoter DNA. Unlike the situation in *E. coli*, there is no precise spacing arrangement of the individual promoter elements and the RNA initiation site.
2. It is likely that a protein distinct from RNA polymerase II specifically recognizes the TATA element. This idea is supported by the variable distance between the TATA sequence and the initiation site, by analogies to mammalian proteins, and by the observations that RNA polymerase II does not recognize specific sequences in *in vitro*.
3. Yeast cells must have many specific transcription factors because upstream promoter elements are required for transcription, yet can consist of different DNA sequences.
4. It is unlikely that transcription activation and positive control are mediated by specific protein-protein interactions between the activator protein and the transcription machinery. Yeast enhancer-like elements act at long and variable distances and also when inverted with respect to the TATA element and the initiation site.
5. Negative control of transcription cannot occur simply by competition for the promoter region between the repressor protein and the transcription machinery. In contrast to *E. coli* repression occurs even when the protein recognition site is located upstream of an intact promoter region.

Therefore, even though yeast is a simple microorganism, the properties of its promoters are qualitatively different from prokaryotic promoters. Instead,

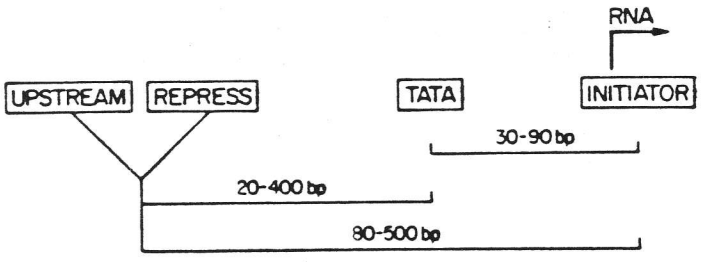

FIGURE 2–4 Promoter/regulatory elements in yeast. *cis*-acting elements of a hypothetical yeast gene are indicated as boxes. The initiator element, which is located near the RNA start (arrow), is important for determining where transcription begins. The TATA element, located 30 to 90 base pairs away from the RNA start, is required for transcription. The upstream promoter element, which can be located at variable distances away from the other elements, is important for transcription and also for regulation. Repressor sites, which are important for negative control, are also located at variable positions upstream of the TATA element. See text.

they are extremely similar to those found in higher eukaryotic organisms (see Chapter 3). Given this information, what is the molecular mechanism for transcription initiation and regulation in yeast cells? In this last section, I summarize the various aspects of yeast promoters in terms of a molecular model (Figure 2–5). At this time, any specific model is highly speculative and is more aptly described as a personal viewpoint. Nevertheless, I hope that this

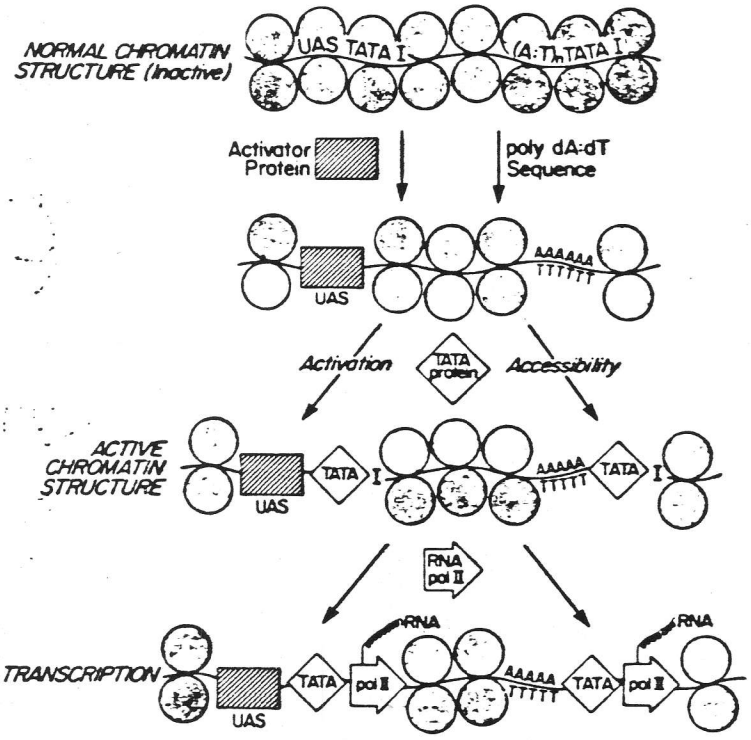

FIGURE 2–5 Molecular models for transcription. As described in the text, these models are highly speculative and are presented mainly to summarize the data. The top part of the figure shows a region of the yeast genome coated wsith nucleosomes (pairs of shaded circles). The promoter sequences of two genes are indicated. The gene on the left contains an upstream promoter element (UAS) typical for a regulated gene, whereas the gene on the right contains poly dA:dT tracts typical of a constitutively expressed gene. Both genes contain TATA and initiator (I) elements. The first step of transcription activation is diagrammed as a disruption in chromatin structure mediated by an activator protein (striped box) or by the unusual properties of the poly dA:dT region. The second step involves interaction of the TATA protein (open diamond) with its cognate promoter element. This is pictured as either activation mediated by the particular protein that binds to the upstream element (for the regulated gene) or as accessibility due to nucleosome exclusion (for the constitutively expressed gene). By either of the proposed mechanisms, the result is an active chromatin structure. The final step is shown as the recognition of this active structure by RNA polymerase II followed by transcription initiation. The precise start point is mediated in some manner by the initiator element.

attempt to organize a large set of observations into a coherent picture will be a useful exercise.

The most basic fact about transcription in eukaryotic organisms is that *in vivo*, the DNA template does not exist free in solution, but rather is associated with histones in a repeating series of structurally discrete nucleosome units. This chromatin structure is best viewed as an inert form of DNA. For example, the DNA in chromatin is far more resistant than purified DNA to the action of nucleases. Thus, for transcription to occur, the basic chromatin structure must be disrupted in a specific way.

There are two ways in which such a structural alteration can be achieved, both of which involve the upstream promoter element. In one of these, poly dA:dT sequences such as those found in constitutive elements can prevent nucleosome formation. In the other, the binding of specific activator proteins to their cognate upstream sequences would also exclude nucleosomes. One reason for suspecting that activator proteins such as *gal4* can disrupt the normal chromatin structure is that they can induce transcription extremely rapidly under the appropriate circumstances. A third possibility is that protein(s) interacting with the TATA element could be responsible for early structural changes. This possibility seems less likely because micrococcal nuclease sensitivity of the *his3* TATA region, which presumably is a measure of a specific protein-DNA interaction, depends on the presence of the upstream element.

Such changes in the chromatin structure, however, are not sufficient for transcription initiation. First, the TATA element is also necessary for transcription to occur. Second, binding of the *lexA* protein to its operator is not sufficient for transcription, whereas binding of the *lexA-gal4* fusion protein is. Thus, there must be an activation step that is distinct from DNA binding.

The nature of this activation is perhaps the most mysterious step in the transcriptional process. The observations that different upstream elements can be functionally associated with a given TATA and initiation region indicate that the activation mechanism must be a general one. In some way, a signal initiated at the upstream element must be transmitted downstream to the TATA box. This suggestion is supported by the fact that activation is inhibited either by bound proteins or by specific sequences between the upstream element and TATA elements.

The activation signal could be the movement of a protein. The obvious candidates are the activator protein, the TATA protein, or RNA polymerase II. Several considerations suggest that the TATA protein is perhaps the best choice. First, the fact that methylation footprints of *gal4* binding can be obtained *in vivo* means that the protein must always be bound at the upstream regulatory site. If such a protein moved, another would have to take its place. Second, if RNA polymerase moves, it would have to do so in a transcriptionally inactive form because readthrough transcripts starting near the upstream element have never been observed. Moreover, it is hard to imagine what would happen when a moving polymerase encounters a protein bound to the TATA element. Since the location of the TATA element does not directly determine the initiation site, the polymerase would have to switch to a directional form of movement upon reaching the TATA element.

An alternative mechanism is that the activation signal represents a structural change induced by the activator protein that is propagated in both directions. One possibility is that activation represents the exclusion of nucleosomes from the promoter region. In this way, the critical promoter sequences would be more accessible to the TATA protein and RNA polymerase II. Another possibility is that the upstream element induces a change in local supercoiling. The TATA protein could bind and/or be activated by recognizing such a structural change.

By either specific model, the normal chromatin structure has been disrupted first by the upstream element and then activated in some manner that results in the binding of the TATA protein. The end result of these two step is the creation of an active chromatin structure. The final stage is the binding by RNA polymerase II and the initiation of RNA synthesis.

The basic proposal is that RNA polymerase II does not interact with DNA in the inert, nucleosomal form of chromatin, but rather recognizes the active structure created by the proteins bound to the upstream and TATA elements. Presumably, the enzyme binds in a region near this complex, just downstream from the TATA element. The size of this active region, about 70 base pairs, corresponds to the variability in spacing between the TATA element and initiation site. Finally, the initiator element is proposed to be the particular sequences within the active region that are preferred by RNA polymerase II. This specificity could be due to the polymerase itself or to an initiation factor that positions the enzyme.

This basic view of transcription initiation provides a simple way to understand the basis of regulation. In essence, regulation is defined by the first step. Positive control is achieved by transcription factors, which are functional only in association with cofactors that exist under specific environmental or developmental circumstances. In their active form, these proteins disrupt the inert chromatin structure; in their inactive form, the chromatin remains inert. Negative control is achieved by repressor proteins that also require cofactors for their action. In their functional form, these repressors either alter the chromatin such that the transcription process cannot begin, or they block the activation process that was started by a positive transcription factor. Thus, complex regulation can be viewed as a competition between activator and repressor proteins, each recognizing a specific DNA sequence and each subject to particular physiological controls, to determine the activity state of chromatin.

REFERENCES

Abraham, J., Feldman, J., Nasymth, K., Strathern, J.N., Klar, A.J.S., Broach, J.R., and Hicks, J.B. (1982) *Cold Spring Harbor Symp. quant. Biol.* 47:989–998.

Abraham, J., Nasymth, K.A., Strathern, J.N., Klar, A.J.S., and Hicks, J.B. (1984) *J. Mol. Biol.* 176:307–331.

Beggs, J. (1978) *Nature* 275:104–109.

Beier, D., and Young, E.T. (1982) *Nature* 300:724–727.

Boeke, J.D., LaCroute, F., and Fink, G.R. (1984) *Mol. Gen. Genet.* 197:345–346.
Bram, R.J., and Kornberg, R.D. (1985) *Proc. Natl. Acad. Sci. USA* 82:43–47.
Brand, A.H., Breeden, L., Abraham, J., Sternglanz, R., and Nasmyth, K. (1985) *Cell* 41:41–48.
Brent, R., and Ptashne, M. (1984) *Nature* 312:612–615.
Brent, R., and Ptashne, M. (1985) *Cell* 43:729–736.
Cashel, M. (1975) *Ann. Rev. Microbiol.* 29:301–318.
Chen, W., and Struhl, K. (1985) *EMBO. J.* 4:3273–3280.
Church, G., and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* 81:1991–1995.
Clarke, L., and Carbon, J. (1980) *Nature* 287:504–509.
Davison, B.L., Egly, J-M., Mulvihill, E.R., and Chambon, P. (1983) *Nature* 301:680–686.
deCrombrugghe, B., Busby, S., and Buc, H. (1984) in *Biological Regulation and Development*, vol. III-B, Yamamoto, K., ed. Plenum Press, New York, pp. 129–167.
Donohue, T.F., Daves, R.S., Lucchini, G., and Fink, G.R. (1983) *Cell* 32:89–98.
Faye, G., Leang, D., Tatchell, K., Hall, B.D., and Smith, M. (1981) *Proc. Natl. Acad. Sci. USA* 78:2258–2262.
Gallant, J.A. (1979) *Ann. Rev. Genet.* 13:393–415.
Giniger, E., Varnum, S., and Ptashne, M. (1985) *Cell* 40:767–774.
Guarente, L. (1984) *Cell* 36:799–800.
Guarente, L., and Hoar, E. (1984) *Proc. Natl. Acad. Sci. USA* 81:7860–7864.
Guarente, L., Lalonde, B., Gifford, P., and Alani, E. (1984) *Cell* 36:503–511.
Guarente, L., and Mason, T. (1983) *Cell* 32:1279–1286.
Guarente, L., Nye, J.S., Hochschild, A., and Ptashne, M. (1982b) *Proc. Natl. Acad. Sci. USA* 79:2236–2240.
Guarente, L., and Ptashne, M. (1981) *Proc. Natl. Acad. Sci. USA* 78:2199–2203.
Guarente, L., Yocum, R.R., and Gifford, P. (1982a) *Proc. Natl. Acad. Sci. USA* 79:7410–7414.
Herefore, L.M., and Rosbash, M. (1977) *Cell* 10:453–462.
Hicks, J.B., Strathern, J.N., and Herskowitz, I. (1977) in *DNA Insertions, Plasmids, and Episomes*, Bukhari, A., Shapiro, J., and Adhya, S. eds. Cold Spring Harbor Laboratory, New York, pp. 457–462.
Hinnebusch, A.G. (1984) *Proc. Natl. Acad. Sci. USA* 81:6442–6446.
Hinnebusch, A.G., and Fink, G.R. (1983a) *J. Biol. Chem.* 258:5238–5247.
Hinnebusch, A.G., and Fink, G.R. (1983b) *Proc. Natl. Acad. Sci. USA* 80:5374–5378.
Hinnebusch, A.G., Lucchini, G., and Fink, G.R. (1985) *Proc. Natl. Acad. Sci. USA* 82:498–502.
Hinnen, A., Hicks, J.B., and Fink, G.R. (1978) *Proc. Natl. Acad. Sci. USA* 75:1929–1933.
Hochschild, A., Irwin, N., and Ptashne, M. (1983) *Cell* 32:319–325.
Hope, I.A., and Struhl, K. (1985) *Cell* 43:177–188.
Jacob, F., and Monod, J. (1961) *J. Mol. Biol.* 3:318–356.
Jacob, F., Ullman, A., and Monod, J. (1964) *C.R. Hebd. Seances Acad. Sci.*, Ser. X 258:3125–3138.
Jensen, R., Sprague, G.F., and Herskowitz, I. (1983) *Proc. Natl. Acad. Sci. USA* 80:3035–3039.
Johnson, A.D., and Herskowitz, I. (1985) *Cell* 42:237–247.
Johnson, A.D., Meyer, B.J., and Ptashne, M. (1978) *Proc. Natl. Acad. Sci. USA* 75:1783–1787.
Johnston, M., and Davis, R.W. (1984) *Mol. Cell. Biol.* 4:1440–1448.
Kaback, D.B., Angerer, L.M., and Davidson, N. (1979) *Nucleic Acids Res.* 6:2499–2517.
Klar, A.J.S., Strathern, J.N., Broach, J.B., and Hicks, J.B. (1981) *Nature* 289:239–244.

Kostriken, R., Strathern, J.N., Klar, A.J.S., Hicks, J.B., and Heffron, F. (1983) *Cell* 35:167–174.
Kunkel, G.R., and Martinson, H.G. (1981) *Nucleic Acids Res.* 9:6869–6888.
Lowry, C.V., Weiss, J.L., Walthall, D.A., and Zitomer, R.S. (1983) *Proc. Natl. Acad. Sci. USA* 80:151–155.
MacKay, V., and Manney, T.R. (1974) *Genetics* 76:273–290.
Marini, J.C., Levene, S.D., Crothers, D.M., and Englund, P.T. (1982) *Cold Spring Harbor Symp. Quant. Biol.* 47:279–283.
Martinez-Arias, A., and Casadaban, M. (1984) *Nature* 307:740–742.
Mortimer, R.K., and Schild, D. (1980) *Microbiol. Rev.* 44:519–571.
Nasmyth, K.A. (1983) *Nature* 302:670–676.
Nasymth, K.A. (1985a) *Cell* 42:213–223.
Nasmyth, K.A. (1985b) *Cell* 42:225–235.
Nasmyth, K.A., Tatchell, K., Hall, B.D., Astell, C., and Smith, M. (1981) *Nature* 289:244–250.
Oettinger, M.A., and Struhl, K. (1985) *Mol. Cell. Biol.* 5:1901–1909.
Orr-Weaver, T., Rothstein, R.J., and Szostak, J.W. (1981) *Proc. Natl. Acad. Sci. USA* 78:6354–6358.
Parker, C.S., and Topol, J. (1984) *Cell* 36:357–369.
Peck, L.J., and Wang, J.C. (1981) *Nature* 292:375–378.
Penn, M.D., Galgoci, B., and Greer, H. (1983) *Proc. Natl. Acad. Sci. USA* 80:2704–2708.
Perlman, D., and Hopper, J.E. (1979) *Cell* 16:89–95.
Petes, T.D. (1980) *Ann. Rev. Biochem.* 49:519–571.
Prunell, A. (1982) *EMBO J.* 1:173–179.
Rhodes, D., and Klug, A. (1981) *Nature* 292:378–380.
Rio, D.C., and Tjian, R. (1983) *Cell* 32:1227–1240.
Rose, M., and Botstein, D. (1983) *J. Mol. Biol.* 170:883–904.
Rose, M., Casadaban, M., and Botstein, D. (1981) *Proc. Natl. Acad. Sci. USA* 78:2460–2464.
Rothstein, R.J. (1983) *Methods in Enzymology* 101:202–211.
Russell, D.W., Smith, M., Cox, D., Williamson, V.M., and Young, E.T. (1983) *Nature* 304:652–654.
Sarokin, L., and Carlson, M. (1984) *Mol. Cell. Biol.* 4:2750–2757.
Scherer, S. (1985) *J. Mol. Biol.*, in press.
Scherer, S., and Davis, R.W. (1979) *Proc. Natl. Acad. Sci. USA* 76:4951–4955.
Scherer, S., Mann, C., and Davis, R.W. (1982) *Nature* 298:815–819.
Schurch, A., Miozzari, G., and Hutter, R. (1974) *J. Bacteriol.* 117:1131–1140.
Sherman, F., Stewart, J.W., and Schweingruber, A.W. (1980) *Cell* 20:215–222.
Siliciano, P.G., and Tatchell, K. (1984) *Cell* 37:969–978.
St. John, T.P., and Davis, R.W. (1979) *Cell* 16:443–452.
St. John, T.P., and Davis, R.W. (1981) *J. Mol. Biol.* 152:285–315.
Strathern, J.N., and Herskowitz, I. (1979) *Cell* 17:371–381.
Strathern, J.N., Hicks, J.B., and Herskowitz, I. (1981) *J. Mol. Biol.* 147:357–372.
Strathern, J.N., Jones, E.W., Broach, J.R., ed. (1982) *The Molecular Biology of the Yeast Saccharomyces*, vols. 1 and 2. Cold Spring Harbor Laboratory, New York.
Struhl, K. (1981a) *Proc. Natl. Acad. Sci. USA* 78:4461–4465.
Struhl, K. (1981b) *J. Mol. Biol.* 152:535–552.
Struhl, K. (1982a) *Nature* 300:284–287.

Struhl, K. (1982b) *Cold Spring Harbor Symp. Quant. Biol.* 47:901–910.
Struhl, K. (1982c) *Proc. Natl. Acad. Sci. USA* 79:7385–7389.
Struhl, K. (1983a) *Nature* 305:391–397.
Struhl, K. (1983b) *Gene* 26:231–241.
Struhl, K. (1984) *Proc. Natl. Acad. Sci. USA* 81:7865–7869.
Struhl, K. (1985a) *Proc. Natl. Acad. Sci. USA* 82, in press.
Struhl, K. (1985b) *Nature* 317:822–824.
Struhl, K., Chen, W., Hill, D.E., Hope, I.A., and Oettinger, M.A. (1985) *Cold Spring Harbor Symp. Quant. Biol.* 50:489–503.
Struhl, K., and Davis, R.W. (1981) *J. Mol. Biol.* 152:569–575.
Struhl, K., Stinchcomb, D., Scherer, S., and Davis, R.W. (1979) *Proc. Natl. Acad. Sci. USA* 76:1035–1039.
Szostak, J., and Blackburn, E. (1982) *Cell* 29:245–255.
Thireos, G., Penn, M.D., and Greer, H. (1984) *Proc. Natl. Acad. Sci. USA* 81:5096–5100.
Travers, A.A. (1980) *J. Mol. Biol.* 141:91–97.
West, R.W., Yocum, R.R., and Ptashne, M. (1984) *Mol. Cell. Biol.* 4:2467–2478.
Wilson, K., and Herskowitz, I. (1984) *Mol. Cell. Biol.* 4:2420–2427.
Wolfner, M., Yep, D., Messenguy, F., and Fink, G.R. (1975) *J. Mol. Biol.* 96:273–290.
Wright, C.F., and Zitomer, R.S. (1984) *Mol. Cell. Biol.* 4:2023–2030.
Zitomer, R.S., Montgomery, D.L., Nichols, D.L., and Hall, B.D. (1979) *Proc. Natl. Acad. Sci. USA* 76:3627–3631.