Annu. Rev. Biochem. 2003. 72:449–79
doi: 10.1146/annurev.biochem.72.121801.161520

First published online as a Review in Advance on March 19, 2003

THE RNA POLYMERASE II CORE PROMOTER

Stephen T. Smale[1] and James T. Kadonaga[2]
[1]*Howard Hughes Medical Institute and Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, California 90095-1662; email: smale@mednet.ucla.edu*
[2]*Section of Molecular Biology, University of California, San Diego, La Jolla, California 92093-0347; email: jkadonaga@ucsd.edu*

Key Words gene regulation, transcription

■ **Abstract** The events leading to transcription of eukaryotic protein-coding genes culminate in the positioning of RNA polymerase II at the correct initiation site. The core promoter, which can extend ~35 bp upstream and/or downstream of this site, plays a central role in regulating initiation. Specific DNA elements within the core promoter bind the factors that nucleate the assembly of a functional preinitiation complex and integrate stimulatory and repressive signals from factors bound at distal sites. Although core promoter structure was originally thought to be invariant, a remarkable degree of diversity has become apparent. This article reviews the structural and functional diversity of the RNA polymerase II core promoter.

CONTENTS

INTRODUCTION

Transcription of a eukaryotic protein-coding gene is preceded by multiple events; these include decondensation of the locus, nucleosome remodeling, histone modifications, binding of transcriptional activators and coactivators to enhancers and promoters, and recruitment of the basal transcription machinery to the core promoter. The core promoter includes DNA elements that can extend ~35 bp upstream and/or downstream of the transcription initiation site. Most core promoter elements appear to interact directly with components of the basal transcription machinery. The basal machinery can be defined as the factors, including RNA polymerase II itself, that are minimally essential for transcription in vitro from an isolated core promoter. The vast majority of studies of the basal machinery have been performed with promoters containing a TATA box as an essential core element. A stable preinitiation complex can form in vitro on TATA-dependent core promoters by association of the basal factors in the following order: TFIID/TFIIA, TFIIB, RNA polymerase II/TFIIF, TFIIE, and then TFIIH. The properties of the basal factors and the mechanisms by which they stimulate transcription initiation from TATA-dependent promoters have been the subject of several recent reviews (1–8). The mechanisms by which sequence-specific transcription factors and coregulators influence the frequency of transcription initiation have also been reviewed (4, 9–13).

Although core promoters for RNA polymerase II were originally thought to be invariant, they have been found to possess considerable structural and functional diversity (14, 15). Furthermore, it appears that core promoter diversity makes an important contribution to the combinatorial regulation of gene expression (15, 16). In this article, we review the basic properties of the most common core elements and our current knowledge of the strategies by which sequence specific motifs in the core promoter participate in combinatorial regulation.

PROPERTIES OF RNA POLYMERASE II CORE PROMOTER ELEMENTS

TATA Box

The TATA box (also named the Goldberg-Hogness box after its discoverers) was the first core promoter element identified in eukaryotic protein-coding genes. The discovery of the TATA box in 1979 emerged from a comparison of the 5′ flanking sequences in a number of *Drosophila*, mammalian, and viral protein-coding genes (17, 18). In virtually every RNA polymerase II-transcribed gene examined, the sequence TATAAA was present 25 to 30 bp upstream of the transcription start site. The development of transfection and in vitro transcription assays made it possible to demonstrate that mutations in the TATA box usually reduced or abolished the activity of cellular and viral promoters (18–22). If

transcription initiation from the mutant promoter remained detectable, the initiation sites were often displaced from the correct location. In *Saccharomyces cerevisiae*, TATA boxes were also found to be critical for transcription initiation; but in this organism, the element was located 40–120 bp from the start site [reviewed in (23)].

PREVALENCE Following the early studies, it was speculated that the TATA box might be strictly conserved and essential for transcription initiation from all protein-coding genes from yeast to man. However, as the promoters for more and more genes were sequenced and characterized, the prevalence of the TATA box diminished. Recent database analyses of *Drosophila* genes revealed that the TATAAA consensus sequence, or a sequence with one mismatch from the consensus, was present in 43% of 205 core promoters (24) or, in another study, in 33% of 1941 potential promoters (25). A database analysis of human genes revealed that TATA boxes were present in 32% of 1031 potential core promoters (26).

TATA RECOGNITION Studies from the Roeder and Parker labs provided the first evidence that a protein binds specifically to the TATA sequence and is responsible for TATA activity (27, 28). Roeder and colleagues identified a biochemical activity, transcription factor IID (TFIID), that elutes from a phosphocellulose column between 0.6 and 1 M NaCl (29). This activity was essential for the activity of TATA-containing core promoters and was capable of binding the core promoter from the adenovirus major late promoter in a DNase I footprinting assay (28, 29). However, further purification and cloning of TFIID proved to be unusually difficult because, in *Drosophila* and man, it appeared to be a heterogeneous, multiprotein complex (30).

The initial cloning of the TATA-binding protein (TBP) gene was facilitated by its discovery in *Saccharomyces cerevisiae* and by the demonstration that the *S. cerevisiae* protein could be purified as a single polypeptide rather than a large multiprotein complex (31). Peptide sequences obtained from the purified yeast protein led to the isolation of TBP cDNA clones from several eukaryotes [reviewed in (32)]. The *Drosophila* and human homologues of yeast TBP were found to be the TATA-binding subunits of the multisubunit TFIID complex (33–35). TBP was also found to be a component of distinct multisubunit complexes that contribute to transcription initiation by RNA polymerases I and III [reviewed in (36)]. TBP-associated factors (TAFs), which are components of the TFIID complex, have been identified and their genes cloned. The biochemical activities contributed by specific TAFs include core promoter recognition (see below), an acetyltransferase activity that uses histones and other proteins as substrates, a kinase activity, ubiquitin activating and conjugating activities, and coactivator functions conferred by protein-protein interactions with gene-specific transcription factors. An in-depth discussion of TAF structure and function is

beyond the scope of this article, but it has been the topic of recent reviews (37–41).

Analyses of TBP-TATA cocrystals revealed a novel mechanism of DNA binding (32, 42–45). The DNA-binding region of TBP folds into a structure that resembles a saddle. This molecular saddle consists of two quasi-symmetrical domains, each containing 89–90 amino acids. The N-terminal domain contacts the 3′ half of a consensus TATA box, and the C-terminal domain contacts the 5′ half. Each half of the large concave surface of the saddle consists of a 5-stranded antiparallel β-sheet. Eight of the 10 β-strands contact the minor groove of the duplex DNA. TBP binding to the minor groove relies on extensive hydrophobic interactions. TBP also induces kinks in the DNA at both the 5′ and 3′ ends of the TATA box and partially unwinds the duplex due to the insertion of phenylalanine residues. The distorted DNA structure is restricted to the region that is directly contacted by TBP, as the flanking DNA duplex is largely unperturbed.

ROLE IN TRANSCRIPTION DIRECTIONALITY In the TBP-DNA cocrystals, TBP is bound in a polar manner to the asymmetrical TATA sequences, TATAAAAG and TATATAAA (42, 43). The polar binding of TBP can, in theory, lead to the assembly of a properly oriented preinitiation complex (containing RNA polymerase II and other general factors) and, therefore, can influence the direction of transcription. Indeed, the orientation of an asymmetrical TATA box has been shown to influence the direction of transcription in vitro from simple synthetic core promoters (46, 47). However, multiple lines of evidence suggest that the contribution of the TATA box to directionality in the context of native promoters may be minimal. First, although TBP binds a consensus TATA box in one orientation in the TBP-DNA cocrystals, it can bind in both orientations in solution with only a modest preference toward the correct orientation (48–50). Furthermore, in the context of more complex synthetic promoters or native promoters, the main determinant of directionality appears to be the relative locations of the activator binding sites, TATA box, and other core promoter elements (46, 47). When the orientation of a consensus, asymmetric TATA box was reversed in a promoter containing distal activator binding sites, the direction of transcription was not reversed; the strength of the promoter was merely reduced, due to the lower affinity of TBP binding in the reverse (nonpreferred) orientation. Recent studies have demonstrated that distal activators can indeed enhance the polarity of TBP binding, which may be a dominant mechanism in determining the direction of transcription (51).

TATA CONSENSUS SEQUENCE Consensus sequences for TATA function and TBP binding have been difficult to define. A binding site selection analysis identified the sequence 5′-TATATAAG-3′ as the optimal TBP recognition sequence (52). However, several other studies revealed that a wide variety of A/T-rich sequences can function as TATA boxes and can interact with TBP (53–57). TBP-DNA cocrystals have been prepared with 10 different TATA sequences to

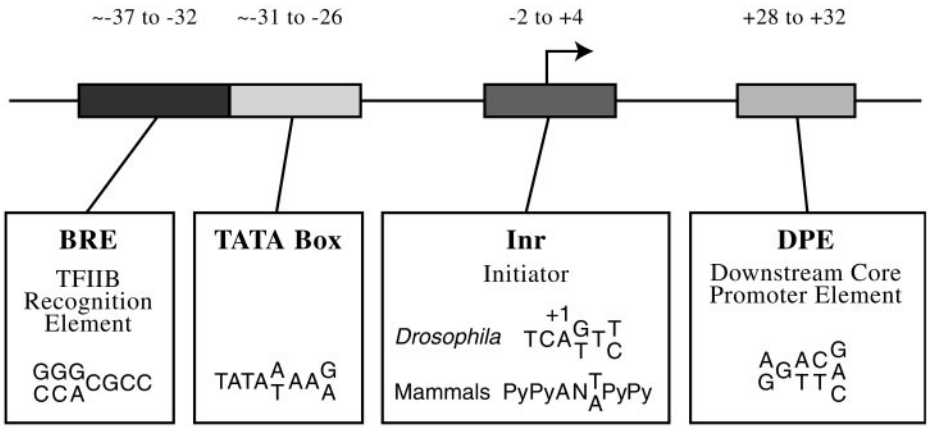

Figure 1 Core promoter motifs. This diagram depicts some of the sequence elements that can contribute to basal transcription from a core promoter. Each of these sequence motifs is found in only a subset of core promoters. A particular core promoter many contain some, all, or none of these elements. The TATA box can function in the absence of BRE, Inr, and DPE motifs. In contrast, the DPE motif requires the presence of an Inr. The BRE is located immediately upstream of a subset of TATA box motifs. The DPE consensus was determined with *Drosophila* core promoters. The Inr consensus is shown for both mammals and *Drosophila*.

examine the structural basis of the promiscuous binding of TBP (57). The results revealed that TBP can induce a similar conformational change in each of the TATA sequences examined. The structural results, combined with statistical data (58), revealed that C:G or G:C base pairs can be accommodated at all but three positions of the TATA box (positions 2, 4, and 5 of the sequence 5′-TATA-AAAG-3′) (57). The TATA definition that resulted from these structural studies was: T≫c>a≈g/A≫t/T≫a≈c/A≫t/T≫a/A≫g>c≈t/A≈T>g>c/G≈A>c≈t (see Figure 1 for a more simplified version of this consensus). It is important to note, however, that this definition only predicts how well each nucleotide can be tolerated at each position, not whether TBP can bind a particular nucleotide sequence. This definition also does not take into account the possibility that TBP may be able to function without forming the stable, kinked structure that is observed in the TBP-TATA cocrystals (59, 60). This possibility may be particularly relevant in promoters that contain other strong, core elements (e.g., an Inr) (56, 61) (see below).

EVOLUTION OF THE TATA BOX Although most studies of TATA boxes have been performed in yeast, *Drosophila*, and man, analogous elements have been found in more ancient eukaryotes as well as in the archaea. In the promoters of several archaeal species, an 8-bp AT-rich sequence is located ~24 bp upstream of the transcription start site [reviewed in (62, 63)]. This sequence, originally called Box

A, is now known to interact with the archaeal homologue of TBP (64, 65). An X-ray crystal structure of TBP from an archaeal hyperthermophile, *Pyrococcus woesei*, revealed a saddle structure similar to that found in eukaryotic TBP (66). Interestingly, the DNA-binding activity of *P. woesei* TBP is optimal at high temperatures, consistent with the fact that this organism grows at 105°C (66). Another notable difference between the archaeal and eukaryotic TBPs is that the archaeal protein exhibits greater symmetry, both in its primary sequence and electrostatic charge distribution [reviewed in (49, 62)]. As discussed below (see BRE section), this increased symmetry decreases the protein's ability to bind TATA boxes in a polar manner.

TATA boxes, or AT-rich sequences located at a fixed distance upstream of the transcription start site, have been identified in essentially all animals, plants, and fungi that have been examined. In addition, TATA-like sequences are found in a number of the more recently evolved protists. For example, promoters in the protozoan parasite, *Entamoeba histolytica*, contain a sequence at −30 that matches the consensus GTATTTAAA(G/C) (67). Like its higher eukaryotic counterpart, this TATA-like element contributes to both promoter strength and start-site selection (68). However, despite the existence of TATA boxes and TBP in the archaea, TATA-like sequences are not apparent in many deep-branching eukaryotes, such as the most ancient parasitic protists [reviewed in (69)]. The absence of TATA boxes in these organisms is likely to reflect the divergence that occurred after these organisms branched from the main line of eukaryotic evolution (69). This divergence is also apparent in some protists that contain TATA boxes, such as *E. histolytica*. In addition to TATA-like and Inr-like elements, core promoters in this organism contain an unusual sequence matching the consensus GAACT (67). This element is found at variable locations within *E. histolytica* core promoters, yet it plays an important role in both promoter strength and start-site placement (68). Thus, although the studies of archaeal transcription originally suggested that core promoter structure would be highly conserved throughout the eukaryotic lineages, tremendous diversity is now apparent.

START-SITE SELECTION IN TATA-CONTAINING PROMOTERS The mechanism that determines the distance from the TATA box to the transcription start site has been the subject of a number of studies [reviewed in (70)]. A key finding emerged from an analysis of the basal factors responsible for the unusually long distance from the TATA box to the transcription start site in *S. cerevisiae* (generally 40–120 bp). By swapping basal factors purified from *S. cerevisiae* and *Schizosaccharomyces pombe*, TFIIB and RNA polymerase II were found to dictate this distance (71). That is, when *S. cerevisiae* TFIIB and RNA polymerase II were combined with the other *S. pombe* basal factors, transcription initiated 40–120 bp downstream of the TATA box. When *S. pombe* TFIIB and RNA polymerase II were combined with the other *S. cerevisiae* basal factors, transcription initiated 25–30 bp downstream of TATA. Studies of archaeal factors

confirmed that its RNA polymerase and TFIIB homolog, TFB, measure the distance from the TATA box to the start site (72). TFIIB was further implicated in this process by the observation that mutations in its N-terminal charged cluster domain can shift the location of the start sites by a few nucleotides in yeast and mammalian promoters (73–75). RNA polymerase II mutations that alter the transcription start site have also been identified (76).

The precise mechanism by which TFIIB and RNA polymerase measure the TATA to start site distance is not known. Most of the TFIIB mutations that alter the start site have no effect on the interaction between TFIIB and the other basal factors with which it interacts, which include TBP, RNA polymerase II, and TFIIF (70, 77, 78). Instead, a recent study found that these mutations alter the conformation of TFIIB (70). One possibility is that this conformational change shifts the location of the RNA polymerase II catalytic center on the promoter. An alternative hypothesis is that the TFIIB mutation induces a conformational change in RNA polymerase II, which alters its position on the promoter (79).

Interestingly, electron crystallography of a yeast RNA polymerase II/TFIIB/TFIIE complex revealed that the distance between the TATA box and the active center of the polymerase is ~30 bp (80). Similarly, the melting of yeast promoter DNA has been found to begin ~20 bp downstream of the TATA box, similar to the distance observed in metazoans (81). These results suggest that, in yeast, the unusually large distance between the TATA box and start site may not be determined by the distance between TBP and the RNA polymerase II catalytic center in a stable transcription preinitiation complex. Rather, a scanning mechanism has been proposed in which the catalytic center is translocated further downstream after initially melting the DNA 20 bp downstream of the TATA box (81).

TBP-RELATED FACTORS Although TBP appears to be the major TATA-binding protein, multicellular animals from nematodes to humans express at least one additional TBP-related factor (TRF) or TBP-like factor (TLF) [reviewed in (82, 83)]. The first TRF, TRF1, was identified in *Drosophila* as a tissue-restricted regulatory protein (84). This protein has the potential to bind consensus TATA boxes, interact with TFIIA and TFIIB (the two basal factors that interact directly with TBP), and substitute for TBP in an in vitro transcription assay (85). In *Drosophila*, TRF1 is also the major component of the RNA polymerase III transcription factor, TFIIIB, whereas TBP is a component of this complex in other eukaryotes (86). Most recently, TRF1 was shown to bind preferentially the sequence TTTTCT, referred to as the TC box, within the core promoter of the *Drosophila tudor* gene (87). Microarray and chromatin immunoprecipitation assays provided further evidence that *tudor* is a direct target of TRF1 (87). TRF1 assembles into a multiprotein complex with associated factors that are distinct from those found in TFIID (85). Together, these findings suggest that TRF1 may be functionally similar to TBP and contribute to the activation of a specific subset of protein-coding genes.

Although TRF1 has been identified only in *Drosophila*, a second TRF, TRF2 (also known as TLF, TLP, and TRP), has been identified in several multicellular animals, but not in plants or fungi [reviewed in (82, 83)]. TRF2 is more distantly related to TBP than is TRF1, but it is likely to fold into a similar saddle structure. TFIIA and TFIIB interactions are retained in TRF2, but the phenylalanines, which (in TBP) are responsible for unwinding and kinking the DNA helix, are missing (82, 83). Consistent with the divergence in the DNA-binding surface, TRF2 cannot bind consensus TATA boxes.

To analyze the function of TRF2 in *Caenorhabditis elegans* early development, RNA interference was used to disrupt its expression (88, 89). The results revealed that TRF2 is essential for development and for the expression of a specific subset of genes. A 300-bp fragment of the *C. elegans* pes-10 promoter has been identified by expression and in situ studies as a direct target of TRF2 (88). In the absence of TRF2, a number of genes were also found to be aberrantly upregulated, which led to the proposal that TRF2 may be a negative regulator of transcription (89). More recently, an analysis of a purified TRF2 complex from *Drosophila* revealed that the transcription factor DREF (DNA replication-related element binding factor) is a TRF2-associated factor (90). A previously described DREF target gene, the PCNA gene, was found to require the DREF-TRF2 complex for the activity of one of its two promoters. Interestingly, an analysis of conserved motifs in 1941 *Drosophila* core promoters revealed that the DREF consensus site, the DRE, was among the most prevalent elements identified (25). However, unlike the TATA, Inr, and downstream promoter element (DPE) motifs, the DREs were not confined to a particular location relative to the transcription start site. These results suggest that DREF may target TRF2 to a subset of core promoters. Because TRF2 can interact with TFIIA and TFIIB, it may then promote the assembly of a productive preinitiation complex.

Initiator Element

The early comparisons of promoter sequences from efficiently transcribed protein-coding genes revealed that most contained an adenosine at the transcription start site (+1), a cytosine at the −1 position, and a few pyrimidines surrounding these nucleotides (91). A large deletion including this initiator region from a sea urchin histone H2A gene revealed that the efficiency of transcription was reduced and that the start-site locations became more heterogeneous (19). Subsequent studies of other metazoan TATA-containing promoters revealed similar results (91–95). In *S. cerevisiae*, where transcription does not initiate at a strictly defined distance from the TATA box, disruption of sequences in the vicinity of the transcription start site resulted in the use of alternative initiation sites (96–99). Collectively, these studies demonstrated that sequences in close proximity to the transcription start site contribute to accurate initiation and the strength of TATA-containing promoters.

The initiator element (Inr) was defined as a discrete core promoter element that is functionally similar to the TATA box and can function independently of

a TATA box in an analysis of the lymphocyte-specific terminal transferase (TdT) promoter (100, 101). Transcription from this promoter initiates at a single start site, yet the region between −25 and −30 is G/C-rich and is unimportant for promoter activity. An extensive mutant analysis revealed that the sequence between −3 and +5 is necessary and sufficient for accurate transcription in vitro and in vivo (100, 102). This region matched the start site consensus sequence observed in the early studies (91). By itself, the TdT Inr supports a very low level of specific initiation by RNA polymerase II. In nuclear extracts, its activity is comparable to that of an isolated TATA box lacking an Inr at the start site (100, 103). The two elements function synergistically with one another when separated by 25 bp. Most importantly, when an Inr is inserted into a synthetic promoter downstream of six binding sites for transcription factor Sp1 (in the absence of a TATA box), the Inr supports high levels of transcription that initiate at a specific start site within the Inr. When the Inr is inserted at a different location relative to Sp1 sites, RNA synthesis consistently begins at the nucleotide dictated by the Inr. In the absence of the Inr, transcription begins from heterogeneous start sites at much lower frequencies.

INR CONSENSUS SEQUENCE AND PREVALENCE Analyses of randomly generated and specifically targeted Inr mutants by in vitro transcription and transient transfection using human cell lines led to the functional consensus sequence Py Py A(+1) N T/A Py Py (102, 104) (Figure 1). The functional consensus defined in *Drosophila* is virtually identical (104). Only a subset of the pyrimidines at the −2, +4, and +5 positions appears to be essential for Inr activity, but the activity increases with increasing numbers of pyrimidines in these positions (102, 104). This functional consensus is similar, but not identical, to the mammalian and *Drosophila* Inr consensus sequences determined by database analysis: Py C A(+1) N T Py Py in mammals (58, 91) and T C A(+1) G/T T Py in *Drosophila* (24, 25, 105, 106) (Figure 1). The *Drosophila* consensus, or a sequence containing one mismatch, was present in 69% of 205 core promoters or, in a separate study, in 69% of 1941 core promoters (24, 25). To our knowledge, the prevalence of the Inr in mammalian promoters has not been determined using rigorous methods similar to those used to calculate its prevalence in *Drosophila*.

Interestingly, transcription does not need to begin at the +1 nucleotide for the Inr to function. RNA polymerase II has been redirected to alternative start sites by reducing ATP concentrations within a nuclear extract, by altering the spacing between the TATA and Inr in a promoter containing both elements, and by dinucleotide initiation strategies (47, 107, 108). In all of these studies, the Inr continued to increase the efficiency of transcription initiation from the alternative sites.

TATA-INR SPACING AND TRANSCRIPTION DIRECTIONALITY Studies of TATA-Inr spacing have shown that the two elements act synergistically when separated by 25–30 bp but act independently when separated by more than 30 bp (47). When

separated by 15 or 20 bp, synergy is retained, but the location of the start site is dictated by the location of the TATA box rather than the location of the Inr (i.e., initiation occurs 25 bp downstream of TATA) (47). Although the Inr element is not symmetrical and therefore has the potential to dictate the direction of transcription from simple synthetic promoters, its contribution to directionality from native promoters and more complex synthetic promoters appears to be minimal (47). Similar to the results obtained with the TATA box, reversal of an Inr in the context of a synthetic promoter containing distal activator binding sites reduced promoter strength, but it did not reverse the direction of transcription.

INR RECOGNITION Most studies of Inr recognition have focused on recognition by TFIID (109). The potential for TFIID recognition of the Inr was suggested by the finding that the TFIID complex is essential for Inr activity (i.e., TBP is insufficient) and that TFIID footprints on some TATA-containing promoters, such as the AdML and *Drosophila hsp70* promoters, extend downstream of the TATA box to approximately +40 (27, 28, 35, 103, 110–113). Although the contacts at the start-site and downstream regions of the AdML promoter were subsequently found to be independent of the Inr (114), DNase I footprinting studies with highly purified human TFIID and simple synthetic promoters containing only TATA and Inr elements revealed weak TFIID interactions at the Inr (115). Analysis of probes containing Inr mutations revealed a close correlation between the Inr contacts and the nucleotides required for Inr function. In the presence of a strong activator bound to distal sites, a much stronger DNase I footprint extending from −35 to +30 was detected on synthetic promoters containing both TATA and Inr elements (114). When the Inr was disrupted by a point mutation, a much weaker footprint was detected that was confined to the TATA box. Disruption of the TATA box eliminated the entire footprint, with the exception of an enhanced DNase I cleavage site at the Inr. Cooperative binding of TFIID was also disrupted by increasing or decreasing the TATA-Inr spacing by 5 bp (114). An independent study used a binding site selection assay to define the nucleotides near the transcription start site of the *Drosophila hsp70* promoter that are required for optimal TFIID binding (116). The selected sequences matched the functional Inr consensus sequence, which provided strong evidence that the Inr consensus sequence is a TFIID recognition site.

Several studies have confirmed that TFIID specifically interacts with the Inr (117–120). Notably, an analysis of TFIID binding to core promoters containing both Inr and DPE elements (see below) revealed that the Inr is essential for stable TFIID binding (120). Although stable binding of the intact TFIID complex to an isolated Inr lacking an upstream TATA box or downstream DPE has not been reported, Verrijzer and colleagues have detected stable Inr binding by a complex consisting of two TAFs, $TAF_{II}250$ and $TAF_{II}150$ (121) [TAFs 1 and 2 in the new TAF nomenclature (122)]. In this study, the Inr consensus was identified in a binding site selection analysis as the DNA sequence preferred by the $TAF_{II}150$-$TAF_{II}250$ complex.

The domains of $TAF_{II}150$ and $TAF_{II}250$ that are responsible for Inr recognition have not been determined. However, their involvement is consistent with functional studies that have shown that a trimeric TBP-$TAF_{II}250$-$TAF_{II}150$ complex is sufficient for Inr activity in reconstituted transcription assays (118). $TAF_{II}150$ was also implicated in Inr activity by the finding that *Drosophila* $TAF_{II}150$ possesses a core promoter-binding activity (123) and that human $TAF_{II}150$ corresponds to a biochemical activity in HeLa cell extracts that is required for Inr function (124, 125).

The synergistic function of TATA and Inr elements was found to correlate with the cooperative binding of TFIID to the two elements (114). The general transcription factor TFIIA was found to be critical for the cooperative binding of TFIID to the Inr element (114). This observation is consistent with earlier evidence that TFIIA can induce a conformational change in the TFIID complex, which alters its contacts with DNA in the vicinity and downstream of the transcription start site (126–128). In a separate study, crosslinking of $TAF_{II}250$ to the Inr of the AdML promoter was greatly enhanced in the presence of TFIIA, presumably due to a TFIID conformational change; this provided further support for the importance of both $TAF_{II}250$ and TFIIA for Inr function (129).

TBP BINDING TO TATA-LESS PROMOTERS Given that the TFIID complex contains subunits that recognize both TATA and Inr elements, a key question is whether the TBP subunit of TFIID must contact the −30 region of TATA-less promoters that contain only an Inr. One study of synthetic core promoters suggested that a TBP contact is necessary because the strength of synthetic Inr-containing promoters was found to be roughly proportional to the A/T content of the −30 region (56). That is, A/T nucleotides at the −30 region had a profound influence on promoter strength even if they had little resemblance to the TATA consensus sequence and were unable to interact stably with TBP. A more definitive analysis of this issue made use of a TFIID complex containing a mutant TBP subunit that cannot bind DNA (61). Two promoters that contain Inr elements but lack TATA boxes (the β-polymerase and TdT promoters) were tested. The mutant TFIID protein was inactive on the β-polymerase promoter, but it remained active on the TdT promoter, which suggested that at least some promoters can function in the absence of a TBP-DNA contact.

The implication of the studies described above, that TBP can contact sequences that have minimal similarity to the TATA consensus sequence, seemed surprising in light of the sequence requirements for the formation of the stable TBP-TATA saddle structure (57). However, a potential explanation was provided by studies showing that stable TBP binding occurs in two steps (59, 60). Possibly, at Inr-containing promoters that require a TBP interaction with a TATA-less −30 sequence, TBP only forms the unstable unbent protein-DNA complex.

OTHER INR-BINDING PROTEINS In addition to TFIID, three other proteins have been reported to recognize Inr elements: RNA polymerase II, TFII-I, and YY-1. Purified RNA polymerase II initiates transcription inefficiently from Inr elements

in the absence of other basal transcription factors (130). Although RNA polymerase II is generally thought to recognize DNA nonspecifically, these results suggest that it possesses a weak, intrinsic preference for Inr-like sequences. In the absence of TAFs, RNA polymerase II, TBP, TFIIB, and TFIIF can form a stable complex on TATA-less promoters that contain Inr elements (130, 131). A subset of the nucleotides within the Inr consensus sequence is required for polymerase recognition and complex formation, but it is not known whether the preferred recognition sequence matches the Inr consensus (131). One hypothesis consistent with the recognition of Inr elements by both TFIID and RNA polymerase II is that TFIID initially recognizes the Inr during preinitiation complex formation. When the polymerase is recruited to the promoter, its intrinsic preference for the Inr may contribute to its proper positioning.

TFII-I was discovered as a factor capable of binding the Inr element within the AdML promoter (132, 133). The structure of TFII-I is complicated, with an unusual DNA-binding domain and six helix-loop-helix motifs that support both homomeric and heteromeric interactions (134, 135). Its ability to stimulate transcription in vitro from Inr-containing promoters and to promote the assembly of a preinitiation complex in an Inr-dependent manner supported the hypothesis that it is a general transcription factor dedicated to the recognition of Inr elements (132, 133). Subsequent evidence revealed that TFII-I is identical to BAP-135, a major phosphorylation target of Bruton's tyrosine kinase (136); SPIN, a protein that binds distal elements in the c-fos promoter and stabilizes the binding of serum response factor and Phox 1 to the c-fos promoter (137, 138); and endoplasmic reticulum stress response element binding factor, which induces the transcription of glucose-regulated protein genes by binding distal promoter elements (139). Furthermore, immunodepletion studies demonstrated that TFII-I is not required for the in vitro function of consensus Inr elements (140). Although several lines of evidence suggest that TFII-I contributes to the function of the Inr element in the T cell receptor Vβ 5.2 promoter, its complicated structure has made it difficult to determine its precise role (141–143).

YY-1 was discovered as a C2H2 zinc finger protein that binds a distal element in the adeno-associated virus (AAV) P5 promoter (144). YY-1 represses transcription through this element in the absence of the adenovirus E1A protein, but it activates transcription in the presence of E1A (144). A subsequent study provided evidence that YY-1 can activate the AAV P5 core promoter by binding its consensus Inr element (145). YY-1 interacts directly with both TFIIB and RNA polymerase II, and the three proteins were found to be sufficient for transcription from the AAV P5 Inr (146, 147). Although these data provide strong evidence that YY-1 can stimulate Inr-dependent transcription in vitro in an assay reconstituted with pure proteins, its contribution to Inr activity in vivo and in nuclear extracts is less certain. Specifically, a mutation at the +2 position of the AAV P5 Inr was found to abolish YY-1 binding, but it had no effect on the activity of the native P5 promoter or the activity of a synthetic promoter containing the P5 Inr (102, 104). In contrast, a mutation at the +3 position

reduced promoter activity but had little effect on YY-1 binding (102, 104). A separate study found that YY-1 does not contribute to the function of the consensus Inr within the β-polymerase promoter, which binds YY-1 with high affinity (131). The precise roles of sequence-specific DNA-binding proteins like TFII-I and YY-1 in Inr function will require further exploration. However, from a more general perspective, it would not be surprising to find that Inr elements in specific genes are recognized by factors other than TFIID and RNA polymerase II.

EVOLUTION OF THE INR Although most studies of Inr elements have been performed in *Drosophila* and man, sequences in the vicinity of the start site contribute to start-site placement and promoter activity in many organisms. In archaeal promoters, the transcription start site usually contains a purine preceded by a pyrimidine [reviewed in (62, 63)]. All archaeal promoters that have been studied require an upstream TATA box for their activity; start-site mutations, however, often reduce promoter strength or shift the transcription start site to other nearby purines. Because archaeal TBP, TFB, and RNA polymerase are sufficient for archaeal transcription, the archaeal Inr is probably recognized by the RNA polymerase (63, 148).

In *S. cerevisiae* promoters, the precise locations of transcription start sites are determined by the DNA sequences in their immediate vicinity (96–99). A positioning sequence is necessary because, as mentioned above, transcription initiation by yeast RNA polymerase II begins 40–120 bp downstream of the TATA box. The sequences observed in the vicinity of yeast transcription start sites often match the consensus PuPuPyPuPu (149). Mutation of this sequence results in repositioning of the start site, but promoter strength is often unaffected or reduced only modestly (149). As in the archaea, it has been suggested that all yeast promoters rely on the presence of a TATA box (150). These properties are quite different from those observed in metazoans, where an Inr often functions in the absence of a TATA box and can greatly enhance promoter strength. Thus, the yeast Inr may be a relatively passive contributor to promoter activity. Rather than being a major contributor to core promoter recognition, it may represent a preferred initiation site for the RNA polymerase after it scans downstream from its initial interaction site, which is 20 bp downstream of the TATA box (81), see above. It is noteworthy, however, that one gene in *S. cerevisiae*, the *GAL80* gene, contains a functional Inr sequence, CACT, that exhibits greater similarity to the metazoan Inr consensus and appears to function in a TATA-independent manner (151, 152).

Similar to yeast and the archaea, start-site sequences that diverge considerably from the metazoan Inr have been identified in several protists, which include *E. histolytica* and *Giardia lamblia* (67, 68, 153, 154). For example, in *E. histolytica*, the Inr consensus sequence is AAAAATTCA (67). Interestingly, the start-site sequences in one of the most ancient eukaryotes, the parasitic protist *Trichomonas vaginalis*, conform closely to the metazoan Inr consensus (155, 156).

Although the *T. vaginalis* genome is AT-rich, functional TATA boxes have not been detected. Rather, promoter activity and the location of the start site are strongly dependent on the *T. vaginalis* Inr (69, 156). It is not yet known whether the apparent absence of TATA boxes in this organism reflects the loss of the TBP gene after branching from the main line of eukaryotic descent, or whether a TBP exists that contributes to transcription initiation from the TATA-less promoters of *T. vaginalis*.

Downstream Promoter Element

The DPE was identified as a downstream core promoter motif that is required for the binding of purified TFIID to a subset of TATA-less promoters [reviewed in (15, 157, 158)]. The DPE is conserved from *Drosophila* to humans and is typically but not exclusively found in TATA-less promoters (24). The DPE acts in conjunction with the Inr, and the core sequence of the DPE is located at precisely +28 to +32 relative to the A_{+1} nucleotide in the Inr motif (24). The DPE consensus sequence is shown in Figure 1.

BINDING OF TFIID TO DPE-DEPENDENT CORE PROMOTERS A typical DPE-dependent promoter contains an Inr and a DPE. Mutation of either the DPE or the Inr results in a loss of TFIID binding and basal transcription activity (120). Hence, TFIID binds cooperatively to the DPE and Inr motifs. In addition, a single nucleotide increase or decrease in the spacing between the DPE and Inr results in a several-fold decrease in TFIID binding and transcriptional activity (24). Consistent with this strict Inr-DPE spacing requirement, all of the ~18 characterized DPE-dependent promoters in *Drosophila* possess the identical spacing between the Inr and DPE sequence motifs (24, 120, 159). Thus, the DPE and Inr function together as a single core promoter unit. In this respect, the DPE differs from the TATA box, which is able to function independently of the presence of an Inr.

DNase I footprinting analysis of the binding of purified TFIID to DPE-dependent promoters revealed an extended region of protection (from about −10 to about +35) that encompasses the Inr and DPE motifs (24, 120). The pattern of DNase I protection and hypersensitivity is consistent with the close association of the DNA in a specific orientation along the surface of TFIID from the Inr to the DPE (24, 120). TBP alone does not bind to TATA-less, DPE-dependent promoters. Photocrosslinking studies with purified TFIID indicated that $TAF_{II}60$ and $TAF_{II}40$ [which are designated TAF6 and TAF9 in the new TAF nomenclature (122)] are in close proximity to the DPE (159). Genetic analyses of $TAF_{II}60$ and $TAF_{II}40$ have been carried out in *Drosophila* (160, 161). These studies revealed that $TAF_{II}60$ and $TAF_{II}40$ are encoded by essential genes. Mutations that alter the amino acid sequences of $TAF_{II}60$ and $TAF_{II}40$ have been found to increase as well as to decrease the expression of DPE- or putative-DPE-containing genes, but it is not yet known whether these mutations affect the function of TFIID at DPE-dependent promoters (160, 161).

THE DPE SEQUENCE MOTIF The DPE was initially identified in the *Drosophila Antennapedia* P2 and *jockey* core promoters, and therefore, the early attempts to determine a DPE consensus sequence were biased toward the DPE sequence motifs in those two promoters. To circumvent this bias, core promoter libraries with randomized sequences in the DPE (or in the vicinity of the DPE) were generated and subjected to in vitro transcription analysis. These studies led to the identification of a range of sequences that can function as a DPE motif (24). This information was then used to identify putative DPE-dependent core promoters in *Drosophila*. In a database of 205 core promoters, it was estimated that about 29% contain a TATA box and no DPE, 26% possess a DPE but no TATA, 14% contain both TATA and DPE motifs, and 31% do not appear to contain either a TATA or a DPE. This analysis further identified nonrandom sequences outside of the DPE core motif (+28 to +32). For instance, a G nucleotide is overrepresented at position +24, and the presence of a G at +24 results in a two- to fourfold higher level of basal transcription (24). Thus, the complete DPE appears to consist of the core motif along with other preferred nucleotides, such as G_{+24}, between the Inr and DPE core. The DPE is also present in human core promoters (159, 162). The analysis of the DPE consensus in humans suggests that it is similar but not identical to that in *Drosophila* (159, Alan K. Kutach, Scott M. Iyama, and J.T. Kadonaga, unpublished data). The frequency of occurrence of the DPE motif in the human genome remains to be determined.

THE DPE VERSUS THE TATA BOX There are similarities and differences between the DPE and TATA box. For example, both the DPE and TATA box are recognition sites for the binding of TFIID. On the other hand, the TATA box, but not the DPE, can function independently of an Inr. If a TATA-dependent promoter is inactivated by mutation of the TATA motif, then core promoter activity can be restored by the addition of a DPE at its downstream position (120).

A key difference between TATA- versus DPE-dependent transcription was revealed by the identification of an activity that stimulates DPE-dependent transcription and represses TATA-dependent transcription (163). This activity was purified and found to be mediated by NC2/Dr1-Drap1, which was initially identified as a repressor of TATA-dependent transcription (164). The observation that NC2/Dr1-Drap1 activates DPE-dependent transcription indicates that it functions differently at DPE- and TATA-dependent promoters. In addition, a mutant form of NC2/Dr1-Drap1 was found to activate DPE transcription but not to repress TATA transcription. Hence, the ability of NC2/Dr1-Drap1 to activate DPE transcription is distinct from its ability to repress TATA transcription. These findings indicate that NC2/Dr1-Drap1 is a multifunctional factor that can discriminate between DPE- and TATA-dependent core promoters.

TFIIB Recognition Element

The TFIIB recognition element (BRE) is the only well-characterized element in the core promoters of protein-coding genes that is recognized by a factor other

than TFIID (or the TRF1 and TRF2 complexes). Initial evidence that a functionally significant element exists immediately upstream of some TATA boxes was provided by mutant analyses of archaeal promoters (165–168). The X-ray crystal structure of a TBP-TFIIB-TATA ternary complex subsequently revealed that TFIIB interacts with the major groove upstream of the TATA box and with the minor groove downstream of the TATA box (169). Protein-DNA crosslinking studies confirmed that TFIIB is in close proximity to the upstream sequences (170, 171).

Compelling evidence that TFIIB interacts with DNA in a sequence-specific manner emerged from two studies, one involving an analysis of the T6 promoter from the archaeal *Sulfolobus shibatae* virus (172). As in other archaeal promoters, a sequence immediately upstream of the TATA box was found to be critical for promoter activity. Archaeal TBP and TFB (the archaeal homolog of TFIIB) bound cooperatively to the promoter when both the TATA box and upstream element were present. A binding site selection analysis revealed no sequence preferences in this upstream region when TBP was analyzed alone. However, in the presence of TFB, strong preferences for purines were observed 3 and 6 bp upstream of the TATA box, with weaker nucleotide preferences at other positions. A parallel study with human TFIIB established the existence of a eukaryotic BRE that prefers a 7-bp sequence: G/C G/C G/A C G C C (173). Recognition of the BRE was found to be mediated by a helix-turn-helix motif at the C-terminus of TFIIB (169, 173, 174). Interestingly, this motif is missing in yeast and plants, which suggests that the BRE may not contribute to gene regulation in these organisms.

As mentioned above, one difference between archaeal TBP and eukaryotic TBP is that the two halves of the archaeal DNA-binding domain exhibit greater symmetry and therefore are incapable of binding most TATA boxes in a polar manner [reviewed in (49, 62)]. Although the relative locations of distal activator proteins and the TATA box appear to be the primary determinants of transcription directionality in eukaryotes, this mechanism may not function in the archaea due to the relative simplicity of the promoter. A series of elegant structural and biochemical studies has shown that archaeal TFB facilitates the polar binding of TBP to the TATA box via its interaction with the BRE (174–176).

Although the interaction between the archaeal TFB and BRE clearly enhances the assembly of a preinitiation complex and transcription initiation, the function of the human TFIIB-BRE interaction appears to be very different. This interaction was originally reported to stimulate RNA polymerase II transcription in an in vitro assay reconstituted with purified basal factors (173). However, it was also observed that the BRE is a repressor of basal transcription in vitro with crude nuclear extracts as well as in vivo in transfection assays (177). This repression and the TFIIB-BRE interaction were relieved when transcriptional activators were bound to distal sites, which resulted in an increased amplitude of transcriptional activation. These results suggest that the function of the BRE may have

expanded during evolution. In the archaea, it stimulates promoter activity, but in eukaryotes, it may also repress transcription.

Proximal Sequence Element

Although this article focuses on core elements found in the promoters of eukaryotic protein-coding genes, a brief description of a well-characterized and highly conserved element within the core promoters of small nuclear RNA (snRNA) genes is pertinent. This element, the proximal sequence element (PSE), is located between −45 and −60 relative to the transcription start site of snRNA genes (178–183). The PSE is essential for basal transcription and dictates the location of the transcription start site. The PSE consensus sequence has been found to vary among organisms (178, 180). In humans, the consensus is T C A C C N T N A C/G T N A A A A G T/G (180).

One of the most intriguing features of the PSE is that, within a single organism, it supports transcription initiation by RNA polymerase II at some snRNA genes and by RNA polymerase III at a distinct subset of snRNA genes (178, 180–183). The core promoter sequences that determine which polymerase transcribes a given snRNA gene vary among organisms. In humans, the presence of a TATA box 15–20 nucleotides downstream of the PSE leads to the recruitment of RNA polymerase III to the promoter, whereas RNA polymerase II transcribes snRNA promoters containing a PSE in the absence of a TATA box (184, 185).

The PSE is recognized by a unique multiprotein complex called SNAPc, PBP, or PTF (181–183, 186–188). The human SNAPc complex is essential for the activity of both RNA polymerase II- and RNA polymerase III-transcribed snRNA promoters (189). The smallest subassembly of SNAPc that can bind the PSE in a sequence-specific manner consists of SNAP190 (residues 84–505), SNAP43 (residues 1–268), and SNAP50 (190). The SNAPc-PSE interaction is mediated, in part, by an atypical Myb domain within SNAP190 (190, 191). SNAPc-directed transcription also requires TBP, which binds DNA in snRNA promoters that contain a TATA box in addition to a PSE (181–183).

Other Core Promoter Elements

There are likely to be a variety of other DNA sequence elements that contribute to core promoter activity. In the analysis of a new putative core promoter motif, it is important to consider its effect upon the basal transcription process, such as the binding of TFIID or TFIIB to the core promoter. It is also useful to examine the frequency of occurrence of the sequence motif—that is, to determine whether the element is used in multiple core promoters. Another related consideration is the specific location of the motif in the core promoter. For instance, the BRE, TATA, Inr, and DPE are located at distinct positions relative to the +1 start site, which is consistent with their roles in the assembly of the transcription initiation

complex. It is possible, however, that other core promoter elements with slightly different functions may be located at variable positions relative to the start site.

It will be important to extend the analysis of core promoter motifs. To understand the full range of mechanisms that are used in the basal transcription process, it is ultimately necessary to identify and to characterize all of the sequence elements that contribute to core promoter activity. In fact, in addition to the motifs described above, a number of core promoter sequences have been found to contribute to transcriptional activity. Some examples of these sequences are as follows (note that the downstream sequences in these promoters appear to be distinct from the DPE). First, the downstream core element (DCE) was identified in the human β-globin promoter (192). The DCE is located from +10 to +45, and it was observed to contribute to transcriptional activity and binding of TFIID. Second, in the human glial fibrillary acidic protein (*gfa*) gene, a downstream promoter element (from +11 to +50) was found to be required for TFIID binding and transcriptional activity (193, 194). Third, the multiple start site downstream element (MED-1) was identified in TATA-less promoters that have unclustered, multiple start sites (195). The MED-1 element was observed to contribute to transcriptional activity in two of three promoters tested (195–197). In the future, it will useful to study these elements further as well as to identify additional core promoter motifs.

CpG Islands

The CpG dinucleotide, a DNA methyltransferase substrate, is underrepresented in the genomes of many vertebrates because 5-methylcytosine can undergo deamination to form thymine, which is not repaired by DNA repair enzymes (198). However, 0.5–2 kbp stretches of DNA exist that possess a relatively high density of CpG dinucleotides. The human genome contains ~29,000 of these CpG islands. Most importantly, it has been estimated that, in mammals, CpG islands are associated with approximately half of the promoters for protein-coding genes (26, 199). During early mammalian development, DNA methylation decreases substantially throughout the genome, followed by de novo methylation to normal levels prior to implantation (198). CpG islands are largely excluded from this phase of de novo methylation, and most remain unmethylated in all tissues and at all stages of development.

Despite the prevalence of promoters associated with CpG islands, the elements that are responsible for their core promoter function remain poorly defined. CpG islands usually lack consensus or near-consensus TATA boxes, DPE elements, or Inr elements (14, 200). In addition, they are often characterized by the presence of multiple transcription start sites that span a region of 100 bp or more. The transcription start sites can coincide with sequences exhibiting weak homology to the Inr consensus or can be unrelated to this sequence. Mutations in the vicinity of the start site can lead to the use of alternative start sites, but promoter strength is often unaffected. In general, it has been difficult to identify core promoter elements within CpG islands that are essential for promoter function. One common feature of CpG islands is the presence of multiple binding sites for transcription factor Sp1 (200–

202). Transcription start sites are often located 40–80 bp downstream of the Sp1 sites; this suggests that Sp1 may direct the basal machinery to form a preinitiation complex within a loosely defined window (14, 103, 200). One possibility is that TFIID subunits that are capable of core promoter recognition (i.e., TBP, $TAF_{II}150$/$TAF_{II}250$, and $TAF_{II}40/60$) then interact with the sequences within that window that are most compatible with their DNA recognition motifs. According to this hypothesis, core promoter recognition within CpG islands relies on the same factors and elements as were discussed above. The key difference is that the binding of basal factors is more strongly dependent on recruitment by activator proteins bound to distal promoter elements.

CONTRIBUTIONS TO COMBINATORIAL GENE REGULATION

The precise reasons for the diversity of eukaryotic core promoters remain largely unknown. Different core promoter classes (e.g., TATA, TATA-Inr, Inr, Inr-DPE) may have evolved initially as functionally equivalent recognition sites for TFIID subunits and their evolutionary precursors. At some eukaryotic promoters, the core elements may continue to serve as functionally equivalent and interchangeable recognition sites for TFIID. However, at other promoters, dramatic differences have been identified, which suggest that core elements can make significant contributions to combinatorial gene regulation strategies. The principle underlying combinatorial regulation is that the limited number of transcription factors within an organism can support a much larger number of gene expression patterns if activation of each gene requires the concerted action of multiple factors. If transcription factors bound to promoter or enhancer elements were capable of activating transcription only when the core promoter contains a specific element or combination of elements, a larger number of gene expression patterns could be obtained. Specific examples of selective communication between core promoter and transcription factors or transcriptional control regions are discussed below, along with hypothetical benefits of selective communication.

Functionally Distinct TATA Sequences

An analysis of the *S. cerevisiae his3* promoter provided the first evidence that core elements can be functionally distinct (203, 204). This promoter contains two TATA boxes, a downstream consensus TATA box (T_R) and an upstream AT-rich sequence (T_C), which functions as a nonconsensus TATA box (205). T_C supports weak transcription of the *his3* gene under all growth conditions, whereas T_R supports strong transcription after the inducible activators GAL4 or GCN4 bind to upstream activating sequences (206). Thus, proper regulation of *his3* transcription appears to rely on the presence of functionally distinct TATA boxes. One benefit of restricting activation by GCN4 and GAL4 to the consensus TATA

box (T_R) is that expression of a nearby, divergently transcribed gene, *pet56*, is unaltered when these factors are induced, because the *pet56* core promoter contains a weak, nonconsensus TATA box (204).

TATA sequences that respond differently to transcriptional activators have also been observed in mammalian cells (207–209). For example, the adenovirus E1A protein stimulates transcription from an *hsp70* promoter regulated by its own consensus TATA box, but E1A cannot stimulate transcription from an *hsp70* promoter containing a nonconsensus TATA box from the SV40 early promoter (207). Similarly, an enhancer associated with the myoglobin gene can activate transcription from the myoglobin promoter, which contains a consensus TATA box, or from an SV40 early promoter after insertion of a consensus TATA box (208). However, the same enhancer cannot activate transcription from an SV40 promoter containing its own nonconsensus TATA box. Although these mammalian studies did not address the benefits of the selective communication, potential benefits were suggested by the following examples.

Restricting the Stimulatory Capacity of Enhancers

Transcriptional enhancers are often located at a considerable distance from their relevant target promoters. Because several enhancers may be in the vicinity of a given promoter and because several promoters may be in the vicinity of a given enhancer, strategies that limit activation by an enhancer to a given promoter could be of great benefit. Indeed, selective communication between enhancers and promoters has been well documented (210, 211).

The myoglobin enhancer analysis cited above and two studies performed in *Drosophila* suggest that core promoter diversity may be critical for selective communication between enhancers and promoters. One study began with the observation that the AE1 enhancer of the *Drosophila* Hox gene cluster is located between the *fushi tarazu* (*ftz*) and *Sex combs reduced* (*Scr*) promoters, but it stimulates transcription only from the *ftz* promoter (212). The *ftz* promoter contains a TATA box, and the *Scr* promoter contains Inr and DPE elements. This observation led to the hypothesis that AE1 preferentially stimulates TATA-containing promoters. Analysis of a series of transgenic *Drosophila* lines demonstrated that AE1 prefers to stimulate transcription from promoters containing TATA boxes. The TATA preference was dependent on competition between the two promoter classes, as AE1 activated the Inr-DPE promoter when the nearby TATA promoter was compromised.

In the second study, P-element-mediated transformation and an enhancer trap strategy were used to introduce a promoter-reporter cassette into various locations of the *Drosophila* genome (213). The promoter-reporter cassette contained both TATA-Inr and Inr-DPE core promoters, with each core promoter linked to a green fluorescent protein (GFP) reporter gene. After integration into the genome, the TATA-Inr promoter-reporter or the Inr-DPE promoter-reporter was deleted by recombination with the FLP or Cre recombinase, respectively, resulting in matched sublines with the different core promoters in identical

genomic locations. Out of 18 enhancers tested, 3 were specific for the DPE-dependent core promoter, whereas 1 was specific for the TATA-dependent core promoter. Moreover, primer extension analysis revealed that there was no detectable activation of the TATA-dependent promoters at the DPE-specific integration sites. The differential activities of TATA-Inr and Inr-DPE promoters when analyzed at identical genomic locations provide evidence that the two types of core promoters exist, at least in part, for the purpose of mediating selective enhancer function.

Restricting Activation by Members of Large Protein Families

In metazoans, most sequence-specific DNA binding proteins are members of large protein families. Multiple family members often recognize similar DNA sequences. Strategies are therefore needed to restrict the activity of a given family member to its relevant target promoters. Selective communication with the core promoter is one such strategy that can be envisioned. Initial support for this hypothesis has been provided by an analysis of the murine terminal transferase (TdT) promoter, which contains a consensus Inr at the transcription start site. The −25 to −30 region of this promoter is G/C-rich and unimportant for promoter function (100, 214). Nevertheless, because the TFIID complex is required for the in vitro activity of this promoter (61), it was anticipated that an engineered TATA box at −30 would effectively substitute for the Inr. Surprisingly, promoter activity was severely compromised when a consensus TATA box was inserted and the Inr disrupted (214). These results suggested that the function of the native promoter depends on the Inr, presumably due to an Inr preference of transcription factors bound to the distal promoter. A subsequent study revealed that Elf-1, which binds 60 bp upstream of the TdT start site and is a member of the large Ets family of DNA-binding proteins, possesses an intrinsic Inr preference (215). Perhaps, as implied above, only a small subset of Ets proteins will exhibit an Inr preference. Family members that can bind the Ets recognition sequence in the TdT promoter, but lack an Inr preference, would be incapable of activating TdT transcription. This hypothesis remains to be tested.

A strong preference for an Inr element has also been observed with a fusion protein between the GAL4 DNA-binding domain and the glutamine-rich activation domains of Sp1 (216). The biological significance of this preference is uncertain, however, because full-length Sp1 is a potent activator of promoters containing either a TATA box or Inr (216). A specific activation domain of c-Fos was found to exhibit a strong preference for activation of TATA-containing promoters (217); deletion of this domain resulted in comparable activation of TATA- and Inr-containing promoters. Transcriptional repression by p53 was also found to depend on a core promoter containing a TATA box (218). Although promoters containing TATA boxes were repressed by p53, comparable promoters containing an Inr instead of a TATA box were resistant to repression.

Finally, GAL4-VP16 activates transcription much more strongly when both TATA and Inr elements are present in the core promoter, whereas other proteins, such as full-length Sp1, are strong activators when the core promoter contains either a TATA box, an Inr, or both elements. (216). The bovine papillomavirus E2 transactivator exhibited a similar preference for core promoters containing both elements (219). The benefit of strengthening a promoter by combining two core elements is apparent in the *Drosophila Adh* gene, which is transcribed from two different promoters. The distal promoter is preferentially used at early stages of development because it contains both TATA and Inr elements (220).

Classes of Genes that Rely on Specific Core Elements

The existence of core promoter diversity has led to considerable interest in the possibility that specific classes of genes will contain specific core elements. The most widely discussed correlation of this type involves CpG-rich promoters, which are frequently associated with ubiquitously expressed housekeeping genes. The retrotransposons termed long interspersed nuclear elements (LINEs) can be used as a second example. In these retrotransposons, the entire promoter region is located downstream of the transcription start site, because these elements are propagated via an RNA intermediate in the absence of a long terminal repeat (LTR). The retrotransposons in *Drosophila* include the *jockey*, *Doc*, *G*, *I*, and *F* elements, all of which contain a DPE in their promoter regions. Thus, the *Drosophila* LINE promoters provide an example of a class of genes that could not function with an upstream TATA box and are entirely dependent on downstream promoter elements. A third class of genes that may be associated with a specific core promoter structure are genes expressed during the earliest stages of mammalian embryogenesis (221). TATA boxes may be nonfunctional during early development, which suggests that the expressed genes contain core promoters that can function in the absence of a TATA box.

MECHANISTIC EVENTS THAT DIFFER AMONG CORE PROMOTER CLASSES

A complete understanding of the mechanistic events that differ among core promoter classes can be realized only after the basal factors required for transcription initiation from each class have been identified. At this time, the reconstitution of basal transcription with a complete set of purified proteins has been achieved only with promoters containing consensus TATA boxes. Identification of the basal factors used at other core promoter classes will also be required for a mechanistic understanding of the selective communication between core promoters and regulatory factors bound to distal sites. Although this goal has not yet been achieved, a number of studies have provided insight into unique mechanistic events that occur at different types of core promoters.

The mechanisms resulting in selective communication between activators and core promoters can be envisioned most easily when the core promoter is recognized by factors other than the intact TFIID complex. For example, some core promoters are recognized by the TRF1 or TRF2 complexes. These complexes are likely to be the targets of defined sets of transcriptional activators and coactivators that cannot interact with TFIID.

A similar scenario can be envisioned at the *S. cerevisiae his3* promoter, which was discussed above. Although metazoans contain very little free TBP, yeast TBP is not tightly associated with the TAF complex; this allows TBP and the TAF complex to be recruited independently to target promoters (222, 223). Genetic analyses have shown that some promoters in yeast are dependent on TAFs, whereas others are TAF-independent (41). At the *his3* locus, Tc-directed transcription is TAF-dependent, whereas T_R-directed transcription is TAF-independent (224, 225). Chromatin immunoprecipitation studies have shown that TAFs associate only with the TAF-dependent T_C promoter and with other TAF-dependent promoters (222, 223). Interestingly, upstream activating sequences associated with a TAF-independent promoter are unable to recruit TAFs to a TAF-dependent promoter (226). GCN4 and GAL4 cannot recruit TAFs to either T_C or T_R, which explains why these activators cannot stimulate transcription from the TAF-dependent T_C promoter. If core promoter function relies on a TAF recognition event, transcription will occur only if the activators and coactivators are capable of recruiting the TAF complex. Although sequence-specific interactions between yeast TAFs and core promoters have not been observed, the promoters that require TAF recruitment, such as the promoters in the ribosomal protein genes, usually contain weak, nonconsensus TATA boxes (227, 228). Furthermore, core promoter sequences downstream of the TATA box have been shown to be responsible for the TAF dependence of some yeast promoters (229). Thus, the yeast core promoter sequence along with the activator bound to the upstream activating sequences appear to communicate by dictating the selective recruitment of the TAF complex, even though discrete TAF-binding elements, comparable to the Inr and DPE in metazoans, may not exist.

In metazoans, tissue-specific TAFs have been identified that are likely to contribute to combinatorial regulation by supporting selective communication between activators and the basal machinery (40). However, because the tissue-specific TAFs do not appear to recognize core promoter elements, their contribution to selective gene activation may be distinct from the contribution of core promoter diversity.

Although the mechanisms of selective communication are relatively easy to envision when the core promoters are recognized by factors other than the intact TFIID complex, it is important to note that most core promoter diversity that has been documented in metazoans involves the variable occurrence of TATA, Inr, and DPE elements. All of these elements are recognized by subunits of the same TFIID complex (103, 114, 120). If the same complex recognizes all core elements, how do these elements support selective communication with enhanc-

ers and distal activators? The mechanisms responsible are likely to involve differences in the mechanisms of basal transcription initiation from the various core promoter classes.

A few factors that are uniquely required for basal transcription from Inr and Inr-DPE core promoters have been identified. The most intriguing factors are NC2/Dr-1-DRAP1, TIC-2, and TIC–3. NC2/Dr1-DRAP1 was described above as a factor that stimulates transcription from Inr-DPE promoters in *Drosophila* but represses transcription from TATA promoters in both *Drosophila* and mammals (163, 164). Perhaps this factor is recruited to core promoters by a subset of distal activators or enhancers, which could provide a mechanism for selective communication between the activator and core promoter. TIC-2 and TIC-3 were identified as activities in HeLa cell nuclear extracts that are required for transcription from an Inr-containing core promoter (140). These factors had no effect on the activity of TATA or TATA-Inr core promoters and therefore could contribute to selective communication. Biochemical studies identified two other factors that are required for Inr activity: $TAF_{II}150$ and TIC-1 (118, 124, 125, 140). Unlike NC2/Dr1-DRAP1, TIC-2, and TIC-3, however, these factors also stimulate transcription from TATA-containing promoters, even though they are not required for TATA-directed transcription.

Although the above-mentioned selectivity factors are likely to contribute to core promoter preferences of transcriptional activators, the mechanisms responsible for the preferences do not necessarily involve selectivity factors. Activators may instead influence parameters of the transcription initiation reaction that are important for only one core promoter class. For example, the rate-limiting steps at TATA promoters may differ from those at Inr or Inr-DPE promoters (108). If an activator influences an event that is rate-limiting at only one core promoter class, it would preferentially stimulate transcription from that class. One biochemical difference between the initiation reaction at different core promoters is that transcription reinitiation appears to be more efficient at TATA-containing promoters than at TATA-less promoters (230). The conformation of the TFIID complex also appears to differ when it is bound to different core promoters; this might make it competent for activation by different subsets of transcriptional activators (231).

Finally, the affinity of the TFIID-DNA interaction may be important for activation by transcription factors such as VP16 and E2, which prefer promoters containing two core elements. As described above, TFIID binds with much higher affinity to core promoters containing two elements (i.e., TATA-Inr or Inr-DPE) (114, 120). VP16 and E2 may belong to a class of activators that stimulates transcription by a mechanism that benefits from the higher affinity of TFIID binding. Activators such as Sp1, which stimulate with equal efficiency from core promoters containing only one core element, may recruit TFIID to the core promoter, either directly or indirectly, via a mechanism that is less dependent on the affinity of TFIID for the core promoter.

CONCLUDING REMARKS

The studies cited above demonstrate that core promoters for RNA polymerase II exhibit considerable diversity, which has the potential to contribute to differential gene regulation. The recent computational studies by Ohler et al. (25) suggest that the TATA box, Inr, and DPE are the predominant core promoter elements whose locations relative to the transcription start site are restricted, at least in *Drosophila*. However, there is a high probability that additional core elements will exist whose locations relative to the start site are not conserved. Less prevalent core elements with restricted locations may also be found. Furthermore, elements within the core promoter that are more analogous to gene-specific elements will almost certainly be identified in many promoters. The detailed dissection of core promoters for new genes will be required to identify these elements.

In the study of core promoters that depend on the common TATA, Inr, and DPE elements, the most important goal for the immediate future is to identify the complete set of basal factors required for the activity of the DPE and Inr. The identification of these factors will be necessary for the long-term goal of determining precisely how different combinations of core elements contribute to the differential regulation of transcription.

ACKNOWLEDGMENTS

We thank Gerald Rubin and Robert Tjian for communicating results prior to their publication. We apologize to colleagues whose work was inadvertently omitted. Studies of RNA polymerase II core promoters in the laboratories of S.T.S. and J.T.K. were supported by the Howard Hughes Medical Institute and the NIH (GM41249), respectively.

The *Annual Review of Biochemistry* is online at http://biochem.annualreviews.org

LITERATURE CITED

1. Orphanides G, Lagrange T, Reinberg D. 1996. *Genes Dev.* 10:2657–83
2. Hampsey M. 1998. *Microbiol. Mol. Biol. Rev.* 62:465–503
3. Myer VE, Young RA. 1998. *J. Biol. Chem.* 273:27757–60
4. Roeder RG. 1998. *Cold Spring Harbor Symp. Quant. Biol.* 58:201–18
5. Lee TI, Young RA. 2000. *Annu. Rev. Genet.* 34:77–137
6. Dvir A, Conaway JW, Conaway RC. 2001. *Curr. Opin. Genet. Dev.* 11:209–14
7. Kornberg RD. 2001. *Biol. Chem.* 382: 1103–7
8. Woychik NA, Hampsey M. 2002. *Cell* 108:453–63
9. Carey M. 1998. *Cell* 92:5–8
10. Kornberg RD. 1999. *Trends Cell. Biol.* 9:M46–49
11. Myers LC, Kornberg RD. 2000. *Annu. Rev. Biochem.* 69:729–49
12. Lemon B, Tjian R. 2000. *Genes Dev.* 14:2551–69
13. Näär AM, Lemon BD, Tjian R. 2001. *Annu. Rev. Biochem.* 70:475–501
14. Smale ST. 1994. *Transcription: Mechanisms and Regulation*, ed. RC Conaway, JW Conaway, pp. 63–81. New York: Raven

15. Butler JEF, Kadonaga JT. 2002. *Genes Dev.* 16:2583–92
16. Smale ST. 2001. *Genes Dev.*15:2503–8
17. Goldberg ML. 1979. PhD Diss. Stanford Univ.
18. Breathnach R, Chambon P. 1981. *Annu. Rev. Biochem.* 50:349–83
19. Grosschedl R, Birnstiel ML. 1980. *Proc. Natl. Acad. Sci. USA* 77:1432–36
20. Wasylyk B, Derbyshire R, Guy A, Molko D, Roget A, et al. 1980. *Proc. Natl. Acad. Sci. USA* 77:7024–28
21. Grosveld GC, Shewmaker CK, Jat P, Flavell RA. 1981. *Cell* 25:215–26
22. Hu SL, Manley JL. 1981. *Proc. Natl. Acad. Sci. USA* 78:820–24
23. Struhl K. 1989. *Annu. Rev. Biochem.* 58:1051–77
24. Kutach AK, Kadonaga JT. 2000. *Mol. Cell. Biol.* 20:4754–64
25. Ohler U, Liao G, Niemann H, Rubin GM. 2002. *Genome Biol.* 3:0087.1–87.12
26. Suzuki Y, Tsunoda T, Sese J, Taira H, Mizushima-Sugano J, et al. 2001. *Genome Res.* 11:677–84
27. Parker CS, Topol J. 1984. *Cell* 36:357–69
28. Sawadogo M, Roeder RG. 1985. *Cell* 43:165–75
29. Matsui T, Segall J, Weil PA, Roeder RG. 1980. *J. Biol. Chem.* 255:11992–96
30. Nakajima N, Horikoshi M, Roeder RG. 1988. *Mol. Cell. Biol.* 8:4028–40
31. Buratowski S, Hahn S, Sharp PA, Guarente L. 1988. *Nature* 334:37–42
32. Burley SK, Roeder RG. 1996. *Annu. Rev. Biochem.* 65:769–99
33. Dynlacht BD, Hoey T, Tjian R. 1991. *Cell* 66:563–76
34. Tanese N, Pugh BF, Tjian R. 1991. *Genes Dev.* 5:2212–24
35. Zhou Q, Lieberman PM, Boyer TG, Berk AJ. 1992. *Genes Dev.* 6:1964–74
36. Goodrich JA, Tjian R. 1994. *Curr. Opin. Cell Biol.* 6:403–9
37. Davidson I, Romier C, Lavigne AC, Birck C, Mengus G, et al. 1998. *Cold Spring Harbor Symp. Quant. Biol.* 63:233–41
38. Shen WC, Apone LM, Virbasius CM, Li XY, Monsalve M, Green MR. 1998. *Cold Spring Harbor Symp. Quant. Biol.* 63:219–27
39. Bell B, Tora L. 1999. *Exp. Cell Res.* 246:11–19
40. Albright SR, Tjian R. 2000. *Gene* 242: 1–13
41. Green MR. 2000. *Trends Biochem. Sci.* 25:59–63
42. Kim JL, Nikolov DB, Burley SK. 1993. *Nature* 365:520–27
43. Kim YC, Geiger JH, Hahn S, Sigler PB. 1993. *Nature* 365:512–20
44. Juo ZS, Chiu TK, Leiberman PM, Baikalov I, Berk AJ, Dickerson RE. 1996. *J. Mol. Biol.* 261:239–54
45. Nikolov DB, Chen H, Halay ED, Hoffman A, Roeder RG, Burley SK. 1996. *Proc. Natl. Acad. Sci. USA* 93:4862–67
46. Xu L, Thali M, Schaffner W. 1991. *Nucleic Acids Res.* 19:6699–704
47. O'Shea-Greenfield A, Smale ST. 1992. *J. Biol. Chem.* 267:1391–402
48. Cox JM, Hayward MM, Sanchez JF, Gegnas LD, van der Zee S, et al. 1997. *Proc. Natl. Acad. Sci. USA* 94:13475–80
49. Tsai FT, Littlefield O, Kosa PF, Cox JM, Schepartz A, Sigler PB. 1998. *Cold Spring Harbor Symp. Quant. Biol.* 63:53–61
50. Liu Y, Schepartz A. 2001. *Biochemistry* 40:6257–66
51. Kays AR, Schepartz A. 2002. *Biochemistry* 41:3147–55
52. Wong J, Bateman E. 1994. *Nucleic Acids Res.* 22:1890–96
53. Hahn S, Buratowski S, Sharp PA, Guarente L. 1989. *Proc. Natl. Acad. Sci. USA* 86:5718–22
54. Singer VL, Wobbe CR, Struhl K. 1990. *Genes Dev.* 4:636–45
55. Wobbe CR, Struhl K. 1990. *Mol. Cell. Biol.* 10:3859–67
56. Zenzie-Gregory B, Khachi A, Garraway

IP, Smale ST. 1993. *Mol. Cell. Biol.* 13:3841–49
57. Patikoglou GA, Kim JL, Sun L, Yang SH, Kodadek T, Burley SK. 1999. *Genes Dev.* 13:3217–30
58. Bucher P. 1990. *J. Mol. Biol.* 212: 563–68
59. Hoopes BC, LeBlanc JF, Hawley DK. 1992. *J. Biol. Chem.* 267:11539–46
60. Zhao X, Herr W. 2002. *Cell* 108:615–27
61. Martinez E, Zhou Q, L'Etoile ND, Oelgeschläger T, Berk AJ, Roeder RG. 1995. *Proc. Natl. Acad. Sci. USA* 92:11864–68
62. Bell SD, Jackson SP. 1998. *Cold Spring Harbor Symp. Quant. Biol.* 63:41–51
63. Bell SD, Jackson SP. 2001. *Curr. Opin. Microbiol.* 4:208–13
64. Marsh TL, Reich CI, Whitelock RB, Olsen GJ. 1994. *Proc. Natl. Acad. Sci. USA* 91:4180–84
65. Rowlands T, Baumann P, Jackson SP. 1994. *Science* 264:1326–29
66. DeDecker BS, O'Brien R, Fleming PJ, Geiger JH, Jackson SP, Sigler PB. 1996. *J. Mol. Biol.* 264:1072–84
67. Purdy JE, Mann BJ, Pho LT, Petri WA Jr. 1994. *Proc. Natl. Acad. Sci. USA* 91:7099–103
68. Singh U, Rogers JB, Mann BJ, Petri WA Jr. 1997. *Proc. Natl. Acad. Sci. USA* 94:8812–17
69. Liston DR, Lau AO, Ortiz D, Smale ST, Johnson PJ. 2001. *Mol. Cell. Biol.* 21:7872–82
70. Fairley JA, Evans R, Hawkes NA, Roberts SGE. 2002. *Mol. Cell. Biol.* 22:6697–705
71. Li Y, Flanagan PM, Tschochner H, Kornberg RD. 1994. *Science* 263:805–7
72. Bell SD, Jackson SP. 2000. *J. Biol. Chem.* 275:12934–40
73. Pinto I, Ware DE, Hampsey M. 1992. *Cell* 68:977–88
74. Bangur CS, Pardee TS, Ponticelli AS. 1997. *Mol. Cell. Biol.* 17:6784–93
75. Hawkes NA, Roberts SGE. 1999. *J. Biol. Chem.* 274:14337–43
76. Berroteran RW, Ware DE, Hampsey M. 1994. *Mol. Cell. Biol.* 14:226–37
77. Cho E-J, Buratowski S. 1999. *J. Biol. Chem.* 274:25807–13
78. Ranish JA, Yudkovsky N, Hahn S. 1999. *Genes Dev.* 13:49–63
79. Zhang D-Y, Carson DJ, Ma J. 2002. *Nucleic Acids Res.* 30:3078–85
80. Leuther KK, Bushnell DA, Kornberg RD. 1996. *Cell* 85:773–79
81. Giardina C, Lis JT. 1993. *Science* 261: 759–62
82. Dantonel J-C, Wurtz J-M, Poch O, Moras D, Tora L. 1999. *Trends Biochem. Sci.* 285:335–39
83. Berk AJ. 2000. *Cell* 103:5–8
84. Crowley TE, Hoey T, Lui JK, Jan LN, Jan LY, Tjian R. 1993. *Nature* 361:557–61
85. Hansen SK, Takada S, Jacobson RH, Lis JT, Tjian R. 1997. *Cell* 91:71–83
86. Takada S, Lis JT, Zhou S, Tjian R. 2000. *Cell* 101:459–69
87. Holmes MC, Tjian R. 2000. *Science* 288: 867–70
88. Kaltenbach L, Horner MA, Rothman JH, Mango SE. 2000. *Mol. Cell* 6:705–13
89. Dantonel J-C, Quintin S, Lakatos L, Labouesse M, Tora L. 1999. *Mol. Cell* 6:715–22
90. Hochheimer A, Zhou S, Zheng S, Holmes MC, Tjian R. 2002. *Nature* 420: 439–45
91. Corden J, Wasylyk B, Buchwalder A, Sassone-Corsi P, Kedinger C, Chambon P. 1980. *Science* 209:1405–14
92. Talkington CA, Leder P. 1982. *Nature* 298:192–95
93. Dierks P, van Ooyen A, Cochran MD, Dobkin C, Reiser J, Weissmann C. 1983. *Cell* 32:695–706
94. Concino MF, Lee RF, Merryweather JP, Weinmann R. 1984. *Nucleic Acids Res.* 12:7423–33
95. Tokunaga K, Hirose S, Suzuki Y. 1984. *Nucleic Acids Res.* 12:1543–58
96. Chen W, Struhl K. 1985. *EMBO J.* 4:3273–80

97. Hahn S, Hoar ET, Guarente L. 1985. *Proc. Natl. Acad. Sci. USA* 82:8562–66
98. McNeil JB, Smith M. 1985. *Mol. Cell. Biol.* 5:3545–51
99. Nagawa F, Fink GR. 1985. *Proc. Natl. Acad. Sci. USA* 82:8557–61
100. Smale ST, Baltimore D. 1989. *Cell* 57:103–13
101. Smale ST, Jain A, Kaufmann J, Emami KH, Lo K, Garraway IP. 1998. *Cold Spring Harbor Symp. Quant. Biol.* 63:21–31
102. Javahery R, Khachi A, Lo K, Zenzie-Gregory B, Smale ST. 1994. *Mol. Cell. Biol.* 14:116–27
103. Smale ST, Schmidt MC, Berk AJ, Baltimore D. 1990. *Proc. Natl. Acad. Sci. USA* 87:4509–13
104. Lo K, Smale ST. 1996. *Gene* 182:13–22
105. Hultmark D, Klemenz R, Gehring WJ. 1986. *Cell* 44:429–38
106. Arkhipova IR. 1995. *Genetics* 139: 1359–69
107. Kadonaga JT. 1990. *J. Biol. Chem.* 265: 2624–31
108. Zenzie-Gregory B, O'Shea-Greenfield A, Smale ST. 1992. *J. Biol. Chem.* 267: 2823–30
109. Smale ST. 1997. *Biochim. Biophys. Acta* 1351:73–88
110. Conaway JW, Hanley JP, Garrett KP, Conaway RC. 1991. *J. Biol. Chem.* 266: 7804–11
111. Pugh BF, Tjian R. 1991. *Genes Dev.* 5:1935–45
112. Purnell BA, Gilmour DS. 1993. *Mol. Cell. Biol.* 13:2593–603
113. Martinez E, Chiang CM, Ge H, Roeder RG. 1994. *EMBO J.* 13:3115–26
114. Emami KH, Jain A, Smale ST. 1997. *Genes Dev.* 11:3007–19
115. Kaufmann J, Smale ST. 1994. *Genes Dev.* 8:821–29
116. Purnell BA, Emanuel PA, Gilmour DS. 1994. *Genes Dev.* 8:830–42
117. Wang JC, Van Dyke MW. 1993. *Biochim. Biophys. Acta* 1216:73–80
118. Verrijzer CP, Chen J-L, Yokomori K, Tjian R. 1995. *Cell* 81:1115–25
119. Bellorini M, Dantonel JC, Yoon JB, Roeder RG, Tora L, Mantovani R. 1996. *Mol. Cell. Biol.* 16:503–12
120. Burke TW, Kadonaga JT. 1996. *Genes Dev.* 10:711–24
121. Chalkley GE, Verrijzer CP. 1999. *EMBO J.* 18:4835–45
122. Tora L. 2002. *Genes Dev.* 16:673–75
123. Verrijzer CP, Yokomori K, Chen J-L, Tjian R. 1994. *Science* 264:933–41
124. Kaufmann J, Verrijzer CP, Shao J, Smale ST. 1996. *Genes Dev.* 10:873–86
125. Kaufmann J, Ahrens K, Koop R, Smale ST, Müller R. 1998. *Mol. Cell. Biol.* 18:233–39
126. Horikoshi M, Carey MF, Kakidani H, Roeder RG. 1988. *Cell* 54:665–69
127. Horikoshi M, Hai T, Lin YS, Green MR, Roeder RG. 1988. *Cell* 54:1033–42
128. Chi TH, Carey M. 1996. *Genes Dev.* 10:2540–50
129. Oelgeschläger T, Chiang C-M, Roeder RG. 1996. *Nature* 382:735–38
130. Carcamo J, Buckbinder L, Reinberg D. 1991. *Proc. Natl. Acad. Sci. USA* 88:8052–56
131. Weis L, Reinberg D. 1997. *Mol. Cell. Biol.* 17:2973–84
132. Roy AL, Meisterernst M, Pognonec P, Roeder RG. 1991. *Nature* 354:245–48
133. Roy AL, Malik S, Meisterernst M, Roeder RG. 1993. *Nature* 365:355–59
134. Roy AL, Du H, Gregor PD, Novina CD, Martinez E, Roeder RG. 1997. *EMBO J.* 16:7091–104
135. Cheriyath V, Roy AL. 2001. *J. Biol. Chem.* 276:8377–83
136. Yang W, Desiderio S. 1997. *Proc. Natl. Acad. Sci. USA* 94:604–9
137. Grueneberg DA, Henry RW, Brauer A, Novina CD, Cheriyath V, et al. 1997. *Genes Dev.*11:2482–93
138. Kim DW, Cheriyath V, Roy AL, Cochran BH. 1998. *Mol. Cell. Biol.* 18:3310–20

139. Parker R, Phan T, Baumeister P, Roy B, Cheriyath V, et al. 2001. *Mol. Cell. Biol.* 21:3220–33
140. Martinez E, Ge H, Tao Y, Yuan CX, Palhan V, Roeder RG. 1998. *Mol. Cell. Biol.* 18:6571–83
141. Manzano-Winkler B, Novina CD, Roy AL. 1996. *J. Biol. Chem.* 271:12076–81
142. Cheriyath V, Novina CD, Roy AL. 1998. *Mol. Cell. Biol.* 18:4444–54
143. Roy AL. 2001. *Gene* 274:1–13
144. Shi Y, Seto E, Chang LS, Shenk T. 1991. *Cell* 67:377–88
145. Seto E, Shi Y, Shenk T. 1991. *Nature* 354:241–45
146. Usheva A, Shenk T. 1994. *Cell* 76:1115–21
147. Usheva A, Shenk T. 1996. *Proc. Natl. Acad. Sci. USA* 93:13571–76
148. Qureshi SA, Bell SD, Jackson SP. 1997. *EMBO J.* 16:2927–36
149. Mosch HU, Graf R, Braus GH. 1992. *EMBO J.* 11:4583–90
150. Brandl CJ, Martens JA, Liaw PC, Furlanetto AM, Wobbe CR. 1992. *J. Biol. Chem.* 267:20943–52
151. Sakurai H, Ohishi T, Fukasawa T. 1994. *Mol. Cell. Biol.* 14:6819–28
152. Ohishi-Shofuda T, Suzuki Y, Yano K, Sakurai H, Fukasawa T. 1999. *Biochem. Biophys. Res. Commun.* 255:157–63
153. Sun CH, Tai JH. 1999. *J. Biol. Chem.* 274:19699–706
154. Yee J, Mowatt MR, Dennis PP, Nash TE. 2000. *J. Biol. Chem.* 275:11432–39
155. Quon DV, Delgadillo MG, Khachi A, Smale ST, Johnson PJ. 1994. *Proc. Natl. Acad. Sci. USA* 91:4579–83
156. Liston DR, Johnson PJ. 1999. *Mol. Cell. Biol.* 19:2380–88
157. Burke TW, Willy PJ, Kutach AK, Butler JEF, Kadonaga JT. 1998. *Cold Spring Harbor Symp. Quant. Biol.* 63:75–82
158. Kadonaga JT. 2002. *Exp. Mol. Med.* 34:259–64
159. Burke TW, Kadonaga JT. 1997. *Genes Dev.*11:3020–31
160. Soldatov A, Nabirochkina E, Georgieva S, Belenkaja T, Georgiev P. 1999. *Mol. Cell. Biol.*19:3769–78
161. Aoyagi N, Wassarman DA. 2001. *Mol. Cell. Biol.* 21:6808–19
162. Zhou T, Chiang C-M. 2001. *J. Biol. Chem.* 276:25503–11
163. Willy PJ, Kobayashi R, Kadonaga JT. 2000. *Science* 290:982–84
164. Maldonado E, Hampsey M, Reinberg D. 1999. *Cell* 99:455–58
165. Reiter WD, Hudepohl U, Zillig W. 1990. *Proc. Natl. Acad. Sci. USA* 87:9509–13
166. Hain J, Reiter WD, Hudepohl U, Zillig W. 1992. *Nucleic Acids Res.* 20:5423–28
167. Hausner W, Frey G, Thomm M. 1991. *J. Mol. Biol.* 222:495–508
168. Palmer JR, Daniels CJ. 1995. *J. Bacteriol.* 177:1844–49
169. Nikolov DB, Chen H, Halay ED, Usheva AA, Hisatake K, et al. 1995. *Nature* 377:119–28
170. Lee S, Hahn S. 1995. *Nature* 376:609–12
171. Lagrange T, Kim T-K, Orphanides G, Ebright Y, Ebright R, Reinberg D. 1996. *Proc. Natl. Acad. Sci. USA* 93:10620–25
172. Qureshi SA, Jackson SP. 1998. *Mol. Cell* 1:389–400
173. Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH. 1998. *Genes Dev.* 12:34–44
174. Tsai FTF, Sigler PB. 2000. *EMBO J.* 19:25–36
175. Bell SD, Kosa PL, Sigler PB, Jackson SP. 1999. *Proc. Natl. Acad. Sci. USA* 96:13662–67
176. Littlefield O, Korkhin Y, Sigler PB. 1999. *Proc. Natl. Acad. Sci. USA* 96:13668–73
177. Evans R, Fairley JA, Roberts SGE. 2001. *Genes Dev.* 15:2945–49
178. Dahlberg JE, Lund E. 1988. In *Structure and Function of the Major and Minor Small Nuclear Ribonucleoprotein Particles*, ed. ML Birnstiel, pp. 38–70. Berlin: Springer-Verlag

179. Skuzeski JM, Lund E, Murphy JT, Steinberg TH, Burgess RR, Dahlberg JE. 1984. *J. Biol. Chem.* 259:8345–52
180. Lobo SM, Hernandez NT. 1994. See Ref. 14, pp. 127–59
181. Henry RW, Ford E, Mital R, Mittal V, Hernandez N. 1998. *Cold Spring Harbor Symp. Quant. Biol.* 63:111–20
182. Hernandez N. 2001. *J. Biol. Chem.* 276: 26733–36
183. Schramm L, Hernandez N. 2002. *Genes Dev.* 16:2593–620
184. Mattaj IW, Dathan NA, Parry HW, Carbon P, Krol A. 1988. *Cell* 55:435–42
185. Lobo SM, Hernandez N. 1989. *Cell* 58:55–67
186. Waldschmidt R, Wanandi I, Seifart KH. 1991. *EMBO J.* 10:2595–603
187. Murphy S, Yoon J-B, Gerster T, Roeder RG. 1992. *Mol. Cell. Biol.* 12:3247–61
188. Sadowski CL, Henry RW, Lobo SM, Hernandez N. 1993. *Genes Dev.* 7:1535–48
189. Henry RW, Mittal V, Ma B, Kobayashi R, Hernandez N. 1998. *Genes Dev.* 12:2664–72
190. Ma B, Hernandez N. 2001. *J. Biol. Chem.* 276:5027–35
191. Mittal V, Ma B, Hernandez N. 1999. *Genes Dev.* 13:1807–21
192. Lewis BA, Kim TK, Orkin SH. 2000. *Proc. Natl. Acad. Sci. USA* 97:7172–77
193. Nakatani Y, Brenner M, Freese E. 1990. *Proc. Natl. Acad. Sci. USA* 87:4289–93
194. Nakatani Y, Horikoshi M, Brenner M, Yamamoto T, Besnard F, et al. 1990. *Nature* 348:86–88
195. Ince TA, Scotto KW. 1995. *J. Biol. Chem.* 270:30249–52
196. Benson LQ, Coon MR, Krueger LM, Han GC, Sarnaik AA, Wechsler DS. 1999. *J. Biol. Chem.* 274:28794–802
197. Rudge TL, Johnson LF. 1999. *J. Cell. Biochem.* 73:90–96
198. Bird A. 2002. *Genes Dev.* 16:6–21
199. Antequera F, Bird A. 1995. *Proc. Natl. Acad. Sci. USA* 90:11995–99
200. Blake MC, Jambou RC, Swick AG, Kahn JW, Azizkhan JC. 1990. *Mol. Cell. Biol.* 10:6632–41
201. Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, et al. 1994. *Nature* 371:435–38
202. Macleod D, Charlton J, Mullins J, Bird AP. 1994. *Genes Dev.* 8:2282–92
203. Struhl K. 1986. *Mol. Cell. Biol.* 6:3847–53
204. Struhl K. 1987. *Cell* 49:295–97
205. Mahadevan S, Struhl K. 1990. *Mol. Cell. Biol.* 10:4447–55
206. Iyer V, Struhl K. 1995. *Mol. Cell. Biol.* 15:7059–66
207. Simon MC, Fisch TM, Benecke BJ, Nevins JR, Heintz N. 1988. *Cell* 52:723–29
208. Wefald FC, Devlin BH, Williams RS. 1990. *Nature* 344:260–62
209. Taylor ICA, Kingston RE. 1990. *Mol. Cell. Biol.* 10:165–75
210. Li X, Noll M. 1994. *EMBO J.* 13:400–6
211. Merli C, Bergstrom DE, Cygan JA, Blackman RK. 1996. *Genes Dev.* 10:1260–70
212. Ohtsuki S, Levine M, Cai HN. 1998. *Genes Dev.*12:547–56
213. Butler JEF, Kadonaga JT. 2001. *Genes Dev.* 15:2515–19
214. Garraway IP, Semple K, Smale ST. 1996. *Proc. Natl. Acad. Sci. USA* 93:4336–41
215. Ernst P, Hahm K, Trinh L, Davis JN, Roussel MF, et al. 1996. *Mol. Cell. Biol.* 16:6121–31
216. Emami KH, Navarre WW, Smale ST. 1995. *Mol. Cell. Biol.* 15:5906–16
217. Metz R, Bannister AJ, Sutherland JA, Hagemeier C, O'Rourke EC, et al. 1994. *Mol. Cell. Biol.* 14:6021–29
218. Mack DH, Vartikar J, Pipas JM, Laimins LA. 1993. *Nature* 363:281–83
219. Ham J, Steger G, Yaniv M. 1994. *EMBO J.* 13:147–57
220. Hansen SK, Tjian R. 1995. *Cell* 82:565–75
221. Majumder S, DePamphilis ML. 1994. *Mol. Cell. Biol.* 14:4258–68

222. Li X-Y, Bhaumik SR, Green MR. 2000. *Science* 288:1242–44
223. Kuras L, Kosa P, Mencia M, Struhl K. 2000. *Science* 288:1244–48
224. Moqtaderi Z, Bai Y, Poon D, Weil PA, Struhl K. 1996. *Nature* 383:188–91
225. Moqtaderi Z, Keaveney M, Struhl K. 1988. *Mol. Cell* 2:675–82
226. Li X-Y, Bhaumik SR, Zhu X, Li L, Shen W-C, et al. 2002. *Curr. Biol.* 12:1240–44
227. Tsukihashi Y, Miyake T, Kawaichi M, Kokubo T. 2000. *Mol. Cell. Biol.* 20:2385–99
228. Mencía M, Moqtaderi Z, Geisberg JV, Kuras L, Struhl K. 2002. *Mol. Cell* 9:823–33
229. Shen WC, Green MR. 1997. *Cell* 90:615–24
230. Yean D, Gralla J. 1997. *Mol. Cell. Biol.* 17:3809–16
231. Knutson A, Castaño E, Oelgeschläger T, Roeder RG, Westin G. 2000. *J. Biol. Chem.* 275:14190–97