

The RNA polymerase II core promoter: a key component in the regulation of gene expression

Jennifer E.F. Butler¹ and James T. Kadonaga^{2,3,4}

¹Vollum Institute, Oregon Health Sciences University, Portland, Oregon 97201, USA; ²Section of Molecular Biology, University of California, San Diego, La Jolla, California 92093, USA

The development, growth, and survival of eukaryotic organisms require the proper regulation of tens of thousands of genes. By complex formulae that have yet to be solved, the expression of each of these thousands of genes is controlled by a wide variety of mechanisms (e.g., see Lefstin and Yamamoto 1998; Roeder 1998; Struhl 1999; Glass and Rosenfeld 2000; Lee and Young 2000; Lemon and Tjian 2000; Strahl and Allis 2000; Courey and Jia 2001; Smale 2001; White 2001; Zhang and Reinberg 2001; Emerson 2002; McKenna and O'Malley 2002; Narlikar et al. 2002; Orphanides and Reinberg 2002; West et al. 2002). In this review, we will focus on one of these regulatory components—the RNA polymerase II core promoter.

Why study the core promoter?

First, the analysis of core promoters contributes fundamental insights into the mechanisms by which transcription occurs in eukaryotes. This basic knowledge is part of the foundation of our molecular understanding of biology. Second, the cascade of events that precede the activation of transcription must eventually lead to the basal transcriptional machinery at the core promoter. In this manner, the core promoter is the ultimate target of action of all of the factors that are involved in the regulation of transcription by RNA polymerase II.

What is the core promoter?

In this review, we define the core promoter as the minimal stretch of contiguous DNA sequence that is sufficient to direct accurate initiation of transcription by the RNA polymerase II machinery (for review, see Struhl 1987; Weis and Reinberg 1992; Smale 1994, 1997, 2001;

Smale et al. 1998; Burke et al. 1998). Typically, the core promoter encompasses the site of transcription initiation and extends either upstream or downstream for an additional ~35 nt. Thus, in many instances, the core promoter will comprise only about 40 nt. There are several sequence motifs—which include the TATA box, initiator (Inr), TFIIB recognition element (BRE), and downstream core promoter element (DPE)—that are commonly found in core promoters (Fig. 1). These motifs each have specific functions that relate to the transcription process, and will be discussed below in greater detail. It is important to note that each of these core promoter elements is found in some but not all core promoters. (For instance, it is a common misconception that all core promoters contain a TATA box.) It appears that there are no universal core promoter elements.

In addition to the core promoter, other *cis*-acting DNA sequences that regulate RNA polymerase II transcription include the proximal promoter, enhancers, silencers, and boundary/insulator elements (e.g., see Blackwood and Kadonaga 1998; Bulger and Groudine 1999; West et al. 2002). These elements contain recognition sites for a variety of sequence-specific DNA-binding factors that are involved in transcriptional regulation. The proximal promoter is the region in the immediate vicinity of the transcription start site (roughly from -250 to +250 nt). Enhancers and silencers can be located many kbp from the transcription start site and act either to activate or to repress transcription. Boundary/insulator elements appear to prevent the spreading of the activating effects of enhancers or the repressive effects of silencers or heterochromatin. We will later discuss specific functional interactions between transcriptional enhancers and core promoter motifs.

What acts at the core promoter?

The core promoter is the site of action of the RNA polymerase II transcriptional machinery (for review, see Orphanides et al. 1996; Hampsey 1998; Myer and Young 1998; Roeder 1998; Lee and Young 2000; Dvir et al. 2001; White 2001; Woychik and Hampsey 2002). RNA poly-

³Corresponding author.

E-MAIL jkadonaga@ucsd.edu; FAX (858) 534-0555.

⁴Dedicated to Robert G. Roeder on the occasion of his 60th birthday.

Article and publication are at <http://www.genesdev.org/cgi/doi/10.1101/gad.1026202>.

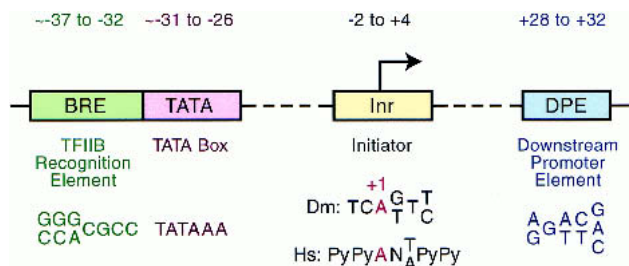


Figure 1. Core promoter elements. Some core promoter motifs that can participate in transcription by RNA polymerase II are depicted. Each of these elements is found in only a subset of core promoters. Any specific core promoter may contain some, all, or none of these motifs. The BRE is an upstream extension of a subset of TATA boxes. The DPE requires an Inr, and is located precisely at +28 to +32 relative to the A₊₁ nucleotide in the Inr. The DPE consensus was determined with *Drosophila* transcription factors and core promoters. The Inr consensus sequence is shown for both *Drosophila* (Dm) and humans (Hs).

merase II is a multisubunit enzyme that catalyzes the synthesis of mRNA from the DNA template. Accurate and efficient transcription from the core promoter requires the polymerase along with auxiliary factors that are commonly termed the “basal” or “general” transcription factors, which include transcription factor (TF) IIA, TFIIB, TFIID, TFIIE, TFIIIF, and TFIIH. With TATA box-dependent core promoters, it has been found that the purified factors can assemble into a transcription preinitiation complex (PIC) in the following order: TFIID, TFIIB, RNA polymerase II-TFIIIF complex, TFIIE, and then TFIIH. For the purposes of this review, it is important to remember TFIID and TFIIB. TFIID is a multisubunit protein that consists of TBP (the TATA box-binding protein) and approximately 13 TBP-associated factors (TAFs; Burley and Roeder 1996; Albright and Tjian 2000; Berk 2000; Verrijzer 2001; Tora 2002). [There are multiple TBP-containing complexes, and hence, the TAFs that are involved in RNA polymerase II transcription are termed TAF_{II}250, TAF_{II}150, etc. The TAF nomenclature has been revised recently (Tora 2002).] TFIIB is a single polypeptide that interacts with TBP as well as with the DNA upstream of the TATA box. In the stepwise PIC assembly mechanism described above, TFIID and TFIIB are the first two factors that interact with the core promoter. Accordingly, it appears that these two factors have a critical role in the recognition of core promoter motifs.

Overview

Core promoters are much more than simple DNA scaffolds for the basal transcription machinery. Rather, core promoter elements are dynamic and vital participants in the regulation of transcriptional activity. In this review, we will describe individual core promoter elements, and then discuss the function of core promoter motifs in the regulation of gene expression.

RNA polymerase II core promoter elements

The TATA box

The TATA box was the first eukaryotic core promoter motif to be identified (Goldberg 1979; Breathnach and Chambon 1981). An early notable feature of the TATA box was its resemblance to the -10 region (Pribnow box) of prokaryotic promoters (Pribnow 1975a,b), but it presently appears that the eukaryotic TATA box and prokaryotic -10 region are not homologous. In metazoans, the TATA box is typically located about 25–30 nt upstream of the transcription start site. In the yeast *Saccharomyces cerevisiae*, the TATA box has a more variable position that ranges from about 40 to 100 nt upstream of the start site. The consensus sequence for the TATA box is TATAAA. It has been observed, however, that a wide range of sequences can function as a TATA box in vivo (Singer et al. 1990). Thus, upstream promoter sequences that resemble TATAAA (perhaps with one or two mismatches from this consensus) may be functional TATA motifs. As noted above, TATA boxes are found in a subset of core promoters. For instance, it was estimated that approximately 43% of 205 core promoters in *Drosophila* contain a TATA box (Kutach and Kadonaga 2000). In humans, it was found that about 32% of 1031 potential promoter regions contain a putative TATA box motif (Suzuki et al. 2001).

It appears that TBP is the predominant TATA box-binding protein, but it is also important to consider that there are TBP-related factors (TRFs) that are closely related to TBP (for review, see Berk 2000). In humans, there is at least one TRF (TRF2; also known as TLP, TRF, TRP), whereas in *Drosophila*, there are at least two TRFs (TRF1 and TRF2; *Drosophila* TRF2 is homologous to human TRF2). There are no apparent TRFs in the yeast *S. cerevisiae*. TRF1 (which is termed TRF in earlier papers) has been found in *Drosophila*, but does not appear to exist in humans. TRF1 can bind to TATA box motifs (Crowley et al. 1993; Hansen et al. 1997), but it is also able to bind to a motif termed the TC box, which is not bound efficiently by TBP (Holmes and Tjian 2000). These and other data suggest that TRF1 mediates transcription in a complex that is distinct from TFIID at TC box-containing promoters (Holmes and Tjian 2000). Moreover, in *Drosophila*, transcription by RNA polymerase III involves TRF1, rather than TBP as in other organisms such as yeast and humans (Takada et al. 2000). TRF2 generally does not appear to bind to TATA box motifs (Maldonado 1999; Moore et al. 1999; Ohbayashi et al. 1999; Rabenstein et al. 1999; Teichmann et al. 1999), but has been found to be required for expression of a specific set of genes (Dantonel et al. 2000; Kaltenbach et al. 2000). It remains to be determined whether TRF2 participates in basal transcription at TRF2-binding sites.

The initiator (Inr)

The Inr element encompasses the transcription start site, and was identified in a variety of eukaryotes (e.g.,

see Corden et al. 1980; Breathnach and Chambon 1981; Hultmark et al. 1986; Struhl 1987). The Inr was defined functionally as a discrete element in an extensive analysis of the core promoter of the murine terminal deoxynucleotidyltransferase (TdT) gene (Smale and Baltimore 1989; Smale 1994, 1997; Smale et al. 1998). Inr elements are found in both TATA-containing as well as TATA-less core promoters. The consensus for the Inr in mammalian cells is Py-Py(C)-A₊₁-N-T/A-Py-Py (Corden et al. 1980; Bucher 1990; Javahery et al. 1994; Lo and Smale 1996; Smale et al. 1998). In *Drosophila*, the Inr consensus is T-C-A₊₁-G/T-T-C/T (Hultmark et al. 1986; Purnell et al. 1994; Arkhipova 1995; Kutach and Kadonaga 2000). The A₊₁ position is designated at the +1 start site because transcription commonly initiates at this nucleotide. More generally, however, transcription initiates at a single site or in a cluster of multiple sites in the vicinity of the Inr (and not necessarily at the A₊₁ position).

A variety of factors have been found to interact with the Inr element. First, there is considerable evidence that TFIID binds to the Inr in a sequence-specific manner (e.g., see Kaufmann and Smale 1994; Martinez et al. 1994; Purnell et al. 1994; Burke and Kadonaga 1996; Oelgeschläger et al. 1996). More specifically, it appears that TAF_{II}150 and TAF_{II}250 (i.e., TAF2 and TAF1; Tora 2002) are the key subunits of TFIID that interact with the Inr (Verrijzer et al. 1994, 1995; Kaufmann et al. 1998; Chalkley and Verrijzer 1999). Aside from TFIID binding to the Inr, it has also been observed that purified RNA polymerase II (or RNA polymerase II along with TBP, TFIIB, TFIIF) is able to recognize the Inr and to mediate transcription in an Inr-dependent manner in the absence of TAFs (Carcamo et al. 1991; Weis and Reinberg 1997). These results suggest that TFIID and RNA polymerase II may recognize and interact with the Inr at different steps in the transcription process.

Two different sequence-specific DNA-binding factors, TFII-I and YY1, have been found to interact with the Inr. TFII-I is a basic-helix-loop-helix (bHLH) protein that binds to Inr and E-box (CACGTG) elements and stimulates transcription in vitro (Roy et al. 1991, 1997; Cheriya et al. 1998). TFII-I also interacts with other sequence-specific DNA-binding factors (Roy et al. 1991; Grueneberg et al. 1997), and may thus participate in the communication between promoter-binding factors and the basal transcriptional machinery. YY1 is a zinc finger protein that binds to the Inr motifs in the AAV P5 and human DNA polymerase β core promoters (Usheva and Shenk 1994; Weis and Reinberg 1997). With the AAV P5 promoter (as a supercoiled DNA template), transcription was observed with purified YY1, TFIIB, and RNA polymerase II (Usheva and Shenk 1994). In contrast, the binding of YY1 to the polymerase β core promoter did not correlate with transcriptional activity (Weis and Reinberg 1997; see also Lo and Smale 1996). These findings collectively suggest that sequence-specific DNA-binding proteins such as TFII-I and YY1 may participate in Inr-dependent transcription at a subset of promoters.

The downstream core promoter element (DPE)

The DPE was identified as a downstream core promoter binding site for purified *Drosophila* TFIID (Burke and Kadonaga 1996). TFIID binds cooperatively to the Inr and DPE motifs, as mutation of either the Inr or the DPE results in loss of TFIID binding to the core promoter (Burke and Kadonaga 1996). The DPE is found most commonly in TATA-less promoters. With naturally occurring TATA-less core promoters, mutation of the DPE motif results in a 10- to 50-fold reduction in basal transcription activity, as observed in the analysis of about 18 *Drosophila* core promoters (Burke and Kadonaga 1996, 1997; Kutach and Kadonaga 2000). Although the DPE has been studied mainly in *Drosophila*, it is also present in humans (e.g., see Burke and Kadonaga 1997; Zhou and Chiang 2001).

The DPE is located precisely at +28 to +32 relative to the A₊₁ position in the Inr. All of the known DPE-containing promoters possess identical spacing between the Inr and DPE motifs, and the alteration of the spacing between the Inr and DPE by a single nucleotide causes a several-fold reduction in TFIID binding and basal transcription activity (Burke and Kadonaga 1997; Kutach and Kadonaga 2000). The consensus sequence for the DPE is estimated to be A/G₊₂₈-G-A/T-C/T-G/A/C. There is also a minor preference for G at +24 (Kutach and Kadonaga 2000). Although the DPE consensus sequence is somewhat degenerate, it should be considered that both DPE and Inr motifs are required in DPE-dependent promoters and that the spacing between the DPE and Inr is invariant (which enables the cooperative binding of TFIID to the two motifs). Thus, the functional consensus for DPE-dependent core promoters consists of the Inr and DPE motifs with the DPE located at +28 to +32 relative to A₊₁.

The frequency of occurrence of the DPE was also investigated. The analysis of a set of 205 *Drosophila* core promoters with thoroughly mapped start sites revealed that the DPE appears to be as common as the TATA box (Kutach and Kadonaga 2000). The presence of putative TATA and DPE motifs in *Drosophila* core promoters was estimated to be as follows: ~29% have a TATA box but no DPE; 26% contain a DPE but no TATA box; 14% possess both TATA and DPE motifs; and 31% do not appear to have either a TATA box or a DPE. The frequency of occurrence of the DPE in human core promoters has yet to be determined.

What binds to the DPE? The DPE is bound by TFIID but not by TBP, and hence, it seems likely that TAFs are involved in the interaction of TFIID with DPE-dependent core promoters. Consistent with this notion, photocrosslinking studies with purified TFIID revealed that TAF_{II}60 and TAF_{II}40 (i.e., TAF6 and TAF9; Tora 2002) are in close proximity to the DPE (Burke and Kadonaga 1997). Thus, it appears that these two TAFs, which possess histone fold motifs that are similar to those in histones H4 and H3, interact with the DPE. The functions of *Drosophila* TAF_{II}40 and TAF_{II}60 have also been investigated in vivo (Soldatov et al. 1999; Aoyagi and Wassarman 2001). *e(y)1/TAF_{II}40* and *TAF_{II}60* are essen-

tial genes in *Drosophila*. A mutation in *e(y)1/TAF_{II}40* (which alters the C-terminal 25 amino acid residues and reduces the level of TAF_{II}40 protein) was observed to decrease expression of the *white* gene, which contains a TATA-less promoter with a weak DPE (Kutach and Kadonaga 2000), and the *yellow* gene, which has a TATA-containing promoter with a 4 out of 5 match to the current DPE consensus (Soldatov et al. 1999). On the other hand, a mutation in *TAF_{II}60* (which results in a Tyr-Tyr insertion in the TAF_{II}60 protein) causes a threefold increase in the transcript levels of *Stellate* (Aoyagi and Wassarman 2001), which is a DPE-dependent gene (Kutach and Kadonaga 2000). This finding suggests that TAF_{II}60 might have a minor role in DPE-dependent transcription, or alternatively, that the Tyr-Tyr insertion in the mutant TAF_{II}60 protein might not impair its function at the DPE. It will be important to investigate further, both in vitro and in vivo, the relation between specific TAFs and DPE-dependent transcription.

Are there different mechanisms of basal transcription from TATA- versus DPE-dependent core promoters? Biochemical studies led to the identification of an activity that stimulates transcription from DPE-dependent (and TATA-less) core promoters and represses transcription from TATA-dependent (and DPE-less) core promoters. This factor was purified and then found to be NC2/Dr1-Drap1 (Willy et al. 2000). NC2/Dr1-Drap1 has been extensively studied as a repressor of TATA-dependent transcription (for review, see Maldonado et al. 1999). The ability of NC2/Dr1-Drap1 to discriminate between DPE- and TATA-dependent promoters indicates that there are fundamental differences in the mechanisms of transcription from DPE- versus TATA-dependent promoters. In addition, a mutant version of NC2/Dr1-Drap1 was found to be fully active for stimulation of DPE-dependent transcription but unable to repress TATA-dependent transcription (Willy et al. 2000). Hence, the ability of NC2/Dr1-Drap1 to activate DPE-dependent transcription is distinct from its ability to repress TATA-dependent transcription.

The TFIIB recognition element (BRE)

The BRE is a TFIIB binding site that is located immediately upstream of some TATA boxes (Lagrange et al. 1998). TFIIB is able to bind directly to the BRE in a sequence-specific manner (Lagrange et al. 1998). The interaction of TFIIB with the BRE was further illuminated by x-ray crystallography of a TFIIB-TBP-DNA complex (Tsai and Sigler 2000). The BRE consensus is G/C-G/C-G/A-C-G-C-C (where the 3' C of the BRE is followed by the 5' T of the TATA box), and at least a 5 out of 7 match with the BRE consensus was found in 12% of a collection of 315 TATA-containing promoters (Lagrange et al. 1998). In vitro transcription experiments with purified basal transcription factors revealed that the BRE facilitates the incorporation of TFIIB into productive transcription initiation complexes (Lagrange et al. 1998). In other studies, the BRE was observed to have a negative effect on basal transcription by in vitro transcription

with a crude extract or by transient transfection analysis (Evans et al. 2001). Additional experiments suggested that the Gal4-VP16 activator can increase transcription, at least in part, by disruption of the repressive TFIIB-BRE interaction (Evans et al. 2001). These results are not necessarily contradictory, and it would be interesting and important to understand the basis for the function of the BRE as a positive or negative core promoter element.

In parallel to the studies of the BRE in humans, an apparently analogous BRE motif was identified in Archaea (Qureshi and Jackson 1998). Archaea contain homologs of TBP and TFIIB that are termed TBP and TFB, and archaeal promoters contain a TATA box sequence (also known as the A box) as well as an upstream TFB-binding site that is called the TFB recognition element (BRE; Qureshi and Jackson 1998). The archaeal BRE increases the efficiency of transcription in vitro (Qureshi and Jackson 1998). Both the eukaryotic and archaeal BREs are located immediately upstream of the TATA box, but the eukaryotic BRE is a GC-rich sequence (Lagrange et al. 1998) whereas the archaeal BRE is an unrelated sequence that is somewhat AT-rich (Qureshi and Jackson 1998).

CpG islands

The CpG (i.e., CG) dinucleotide is underrepresented in vertebrate genomes due to methylation at the 5 position of the cytosine ring and subsequent deamination of the 5-methylcytosine to give a TpG dinucleotide, which is not repaired by the DNA repair machinery. However, there are stretches of DNA, termed CpG islands, that are relatively GC-rich and overrepresented in CpG dinucleotides that are mostly unmethylated (e.g., see Bird 1986, 1993; Gardiner-Garden and Frommer 1987; Antequera and Bird 1993; Adachi and Lieber 2002). CpG islands, which generally range in size from 0.5 to 2 kbp, contain promoters for a wide variety of genes. CpG islands typically lack TATA or DPE core promoter elements, but contain multiple GC box motifs that are bound by Sp1 and related transcription factors (e.g., see Brandeis et al. 1994; Macleod et al. 1994). In addition, transcription from CpG islands initiates from multiple weak start sites that are often distributed over a region of about 100 nt, which is in contrast to transcription from TATA or DPE-dependent core promoters that occurs from a single site or localized cluster (of less than 10 nt) of sites. The analysis of 1031 human genes revealed that about half of the potential promoter regions are located in CpG islands (Suzuki et al. 2001).

From the core promoter perspective, CpG islands may contain multiple weak core promoters rather than a single strong core promoter. The presence of Sp1 binding sites in CpG islands is particularly notable. Not only does Sp1 contribute to the maintenance of the hypomethylated state of CpG islands (Brandeis et al. 1994; Macleod et al. 1994), but it may also function in concert with the basal transcription factors to mediate transcription initiation. It has been found, for example, that Sp1 binding sites in conjunction with an Inr motif can acti-

vate transcription in the absence of a TATA box (e.g., see Smale and Baltimore 1989; Smale et al. 1990; Emami et al. 1995). Hence, it is possible that CpG island promoters consist of multiple Sp1+Inr pairs that collectively generate the array of start sites that are observed.

Other core promoter sequences

In addition to the core promoter motifs described above, other DNA sequences in the core promoter region have been found to contribute to transcriptional activity in a variety of genes. Some examples are as follows (note that the downstream sequences in these promoters appear to be distinct from the DPE). The human β -globin promoter has a downstream promoter sequence from +10 to +45 that is termed the downstream core element (DCE; Lewis et al. 2000). Mutations in the DCE were observed to reduce the efficiency of transcription and TFIID binding. The promoter of the human glial fibrillary acidic protein (*gfa*) gene has a TATA box as well as a downstream sequence from +11 to +50 that is required for TFIID binding and transcriptional activity (Nakatani et al. 1990a,b). The analysis of TATA-less promoters with unclustered, multiple start sites led to the identification of a downstream motif termed MED-1 (multiple start site element downstream; Ince and Scotto 1995). The MED-1 motif was observed to contribute to transcriptional activity in two promoters but not in a third (Ince and Scotto 1995; Benson et al. 1999; Rudge and Johnson 1999). It is likely that there are more core promoter elements to be discovered. It will be of particular interest to know how these elements function in relation to the transcriptional machinery.

Modes of transcriptional regulation via core promoter elements

The core promoter is the ultimate target of the action of sequence-specific transcription factors and coregulators. Current evidence indicates that many transcriptional enhancers and factors exhibit core promoter specificity. Hence, core promoters are not passive elements that serve only to direct the proper placement of the RNA polymerase II transcriptional machinery. Rather, core promoter motifs are *cis*-acting regulatory elements. In this manner, the core promoter provides another level of transcriptional regulation. Some instances of enhancer-core promoter specificity are as follows.

TATA₁ versus TATA₂

In this section, we will describe some examples of differential regulation between two different TATA box motifs. In most of these cases, one of the two TATA box elements has a canonical TATA sequence (TATAAA), whereas the other TATA box has a noncanonical TATA sequence.

The *his3* promoter in *S. cerevisiae* contains two TATA

boxes, termed T_C and T_R, that direct transcription from distinct start sites (Struhl 1986, 1987). The downstream T_R (regulatory TATA) has a canonical TATAAA sequence, whereas the upstream T_C (constitutive TATA) is an AT-rich region of about 30 nt that lacks a canonical TATA sequence (Mahadevan and Struhl 1990). T_C is used preferentially at low levels of transcription. However, upon induction of the *his3* gene, there is an increase in transcription from T_R but not from T_C. Thus, these experiments demonstrate differential usage of two TATA motifs (Fig. 2A). With the *his3* gene, the arrangement of the weak TATA T_C upstream of the strong TATA T_R is an important component of the preferential usage of T_C relative to T_R at low transcription levels. Upon *his3* induction, the transcriptional activators function specifically with the stronger downstream T_R (Iyer and Struhl 1995). Factors that preferentially repress T_C-directed transcription have also been identified (e.g., see Collart and Struhl 1994). In addition, some TAFs were found to be required for T_C-dependent transcription but not T_R-dependent transcription (Moqtaderi et al. 1996, 1998). These findings, along with the observation that TAF dependence correlates with TAF occupancy at promoters (Kuras et al. 2000; Li et al. 2000), suggest that TFIID interacts with T_C but not with T_R. As noted above, there appears to be only a single TBP gene in *S. cerevisiae*, and it is therefore unlikely that TATA-related factors (TRFs) are involved in T_R or T_C transcription.

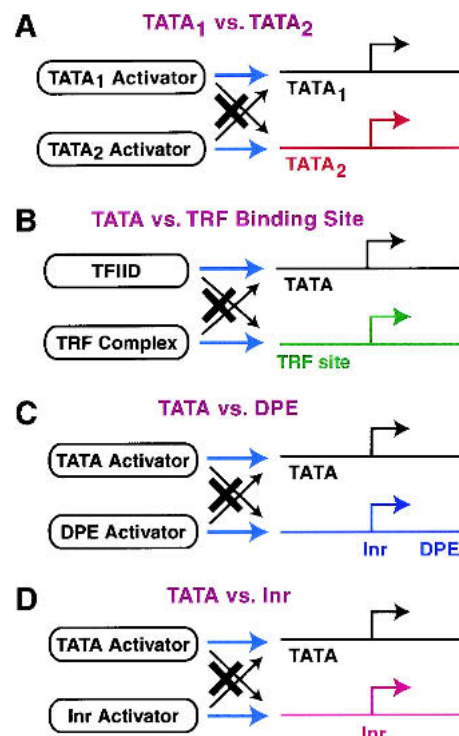


Figure 2. Some modes of regulation via the core promoter. These models are discussed in the text. TATA₁ vs. TATA₂ regulation usually involves a canonical TATA sequence (TATAAA) relative to a weak, noncanonical TATA sequence.

TATA versus TATA specificity has also been observed with mammalian enhancers and activators. For instance, in the analysis of the transactivation of the human *hsp70* promoter by E1A, it was found that substitution of the *hsp70* TATA (TATAA) with the SV40 early promoter TATA (TATTTAT) resulted in a loss of E1A inducibility but not in heat inducibility (Simon et al. 1988). Hence, the specific TATA box sequence of the *hsp70* promoter is important for its induction by E1A.

The human myoglobin enhancer provides another example of TATA versus TATA specificity. The myoglobin enhancer is able to activate the myoglobin promoter but not the SV40 promoter. However, when the SV40 early promoter TATA sequence (TATTTAT) was changed to the myoglobin TATA sequence (TATAAAA), the resulting promoter was activated by the myoglobin enhancer (Wefald et al. 1990). Thus, the myoglobin enhancer functions specifically with the TATAAAA sequence relative to the TATTTAT sequence.

The analysis of TAF recruitment by upstream activating sequences (UASs; enhancer elements in *S. cerevisiae*) has revealed a mechanism for enhancer-promoter specificity. In yeast, there are TAF-dependent (TAF_{dep}) and TAF-independent (TAF_{ind}) promoters (Kuras et al. 2000; Li et al. 2000). By chromatin immunoprecipitation analysis, it was found that TBP associates with both types of promoters, whereas TAFs associate with TAF_{dep} promoters but not with TAF_{ind} promoters. It was observed that the UAS of a TAF_{ind} gene is unable to recruit TAFs to a TAF_{dep} promoter and to activate its transcription (Li et al. 2002). These experiments indicate that a mechanism for enhancer-promoter specificity is the differential recruitment of TAFs by the enhancers. In related studies, it was found that a Rap1-containing activator recruits TFIID to ribosomal protein (RP) promoters, which generally lack canonical TATA box motifs (Mencía et al. 2002). This Rap1-containing activator also stimulates transcription that is directed by the *his3* T_C element (which, as noted above, does not have a canonical TATA sequence) but not by the canonical T_R TATA box.

The analysis of B-cell-specific transcription of immunoglobulin (Ig) promoters led to the identification of another potential mechanism of TATA versus TATA specificity. A key factor that is required for activation of Ig promoters is the B-cell-specific coactivator termed OCA-B (Luo et al. 1992; Luo and Roeder 1995; Roeder 1998). OCA-B exhibits specificity for transcription from the IgH promoter, which contains a noncanonical TATA sequence (TAAATATA) relative to the H2B promoter, which contains a canonical TATA sequence (TATAT AAA). It has been suggested that the basis for this specificity is the TATA sequence (Roeder 1998). Hence, TATA versus TATA specificity could be mediated by transcriptional coactivators such as OCA-B. Moreover, coactivators also appear to be involved in TATA-containing versus TATA-less transcription. For instance, NC2/Dr1-Drap1 represses TATA-dependent transcription and activates DPE-dependent transcription (Willy et al. 2000). In addition, cofactors that specifically activate

transcription from TATA-less core promoters have been partially purified (Martinez et al. 1998).

TATA versus TRF binding site

TBP and TBP-related factors (TRFs) were discussed above in the section on the TATA box. Differential gene activation by TBP versus TRFs has been observed (Dantonel et al. 2000; Kaltenbach et al. 2000; Holmes and Tjian 2000). In one instance, it was found that a specific motif in the *Drosophila* *tudor* core promoter, termed the TC box, is a binding site for TRF1 but not TBP (Holmes and Tjian 2000). In addition, TRF1 was found to activate the *tudor* promoter both in vitro and in vivo. Hence, it is important to consider that enhancers and activators might function specifically with either TBP- or TRF-containing transcription complexes (Fig. 2B).

TATA versus DPE

The phenomenon of enhancer-promoter specificity has been observed in the analysis of gene expression in *Drosophila* (Li and Noll 1994; Merli et al. 1996). In these studies, however, it was not known whether the effects were due to the proximal promoters (which are bound by sequence-specific factors) or to the core promoters.

The effect of TATA-containing versus TATA-less (and weak DPE-containing) core promoters was examined in promoter competition experiments in *Drosophila* (Ohtsuki et al. 1998). In these studies, a transcriptional enhancer was situated between (or in the vicinity of) two divergently transcribed core promoters: the TATA-containing *even-skipped* core promoter and the TATA-less (and weak DPE-containing; see Kutach and Kadonaga 2000) *white* core promoter. These experiments revealed that the *Drosophila* AE1 and IAB5 enhancers preferentially activate transcription from the TATA-containing *even-skipped* core promoter relative to the TATA-less *white* core promoter (Ohtsuki et al. 1998). In the absence of a competing TATA-containing promoter, however, both enhancers were found to be able to activate transcription from the TATA-less *white* core promoter. Thus, the AE1 and IAB5 enhancers can function with the TATA-less *white* core promoter, but exhibit a strong preference for TATA-containing relative to TATA-less promoters in a competitive situation.

The comparative analysis of DPE and TATA core promoter elements in *Drosophila* led to the identification of DPE- and TATA-specific transcriptional enhancers (Butler and Kadonaga 2001). By using a combination of enhancer-trapping and FLP/Cre excision techniques, those authors created pairs of sister *Drosophila* lines that contain either a DPE- or a TATA-dependent reporter gene at precisely the same genomic position relative to each trapped enhancer. Then, the ability of each enhancer to activate DPE-dependent transcription (in the DPE reporter flies) was compared to its ability to activate TATA-dependent transcription (in the sister TATA reporter flies). The characterization of 18 trapped enhanc-

ers led to the discovery of three DPE-specific enhancers and one TATA-specific enhancer. By primer extension analysis, it was found that the DPE-specific enhancers activate transcription from the +1 start site of DPE-dependent core promoters, but do not exhibit any detectable activity with the TATA-dependent core promoters. These findings indicate that the presence of a DPE or TATA motif in the core promoter can be a key determinant of enhancer function (Fig. 2C).

Why might transcriptional enhancers function specifically with DPE- or TATA-dependent core promoters? Enhancer-core promoter specificity would facilitate the desired interaction of an enhancer with its promoter. For example, this function would be useful for an enhancer that must activate a single promoter that is located in a cluster of genes. In addition, if an enhancer were located many kbp from its core promoter, then this specificity would facilitate the proper association of the enhancer with its promoter. On the other hand, enhancer-core promoter specificity might not be necessary if an enhancer were located immediately upstream of its promoter. Consistent with this notion, only four of 18 enhancers tested were found to be specific for TATA or DPE motifs (Butler and Kadonaga 2001). Core-promoter specificity appears to be a property of a subset of transcriptional enhancers.

TATA versus *Inr*

The presence or absence of an *Inr* element is also an important factor in transcriptional regulation (Fig. 2D). For instance, Gal4-VP16 and Sp1 exhibit differences in their ability to activate TATA-containing promoters relative to *Inr*-containing promoters (Emami et al. 1995). The Gal4-VP16 activator has a strong preference for TATA-containing promoters, whereas Gal4-Sp1 fusion proteins that contain only the glutamine-rich activation region of Sp1 have a strong preference for *Inr*-containing promoters. In another study, it was found that the upstream D' element of the TdT gene (which has a TATA-less, *Inr*-containing core promoter) requires an *Inr* element in the core promoter to activate transcription, and that the addition of a TATA box does not replace the requirement for an *Inr* (Garraway et al. 1996). These results reveal the key role of the *Inr* in transcriptional regulation and further suggest that specific activation domains, such as the glutamine-rich regions of Sp1, are involved in activation via the *Inr* element.

Summary and perspectives

A few final thoughts on core promoters are as follows. Transcriptional enhancers are confronted with a wide variety of core promoters (Fig. 3). Core promoters not only direct the initiation of transcription, but also participate in the specificity of enhancer function.

There are many distinct core promoter elements. Thus, it is probably best not to think of core promoters as simply "TATA-containing" or "TATA-less". Rather,

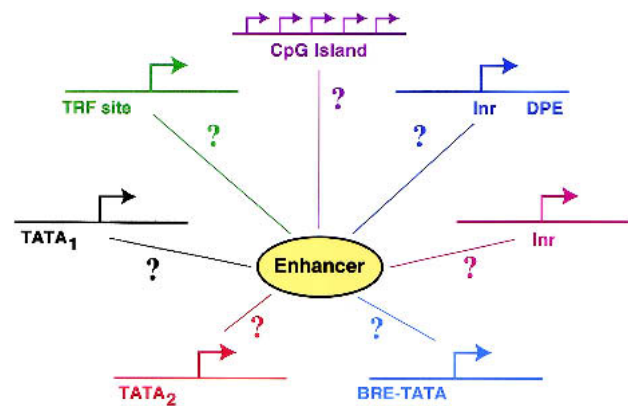


Figure 3. Transcriptional enhancers are confronted with a variety of core promoters.

for each core promoter, the presence or absence of each individual motif should be examined. In addition, it would probably be useful to consider whether any TATA-containing promoter has a canonical TATA box (~TATAAA) or a noncanonical TATA element (i.e., an AT-rich region lacking a TATAAA-like motif).

The identification of core-promoter-specific enhancers indicates the importance of studying transcriptional enhancers in conjunction with their cognate core promoters. For instance, if a DPE-specific enhancer were fused to a TATA-dependent core promoter, then transcriptional activation would not be observed.

The emerging picture of transcriptional regulation is one of unanticipated variety and complexity. While our understanding of eukaryotic gene regulation has advanced considerably, the incredible intricacy of the system has become apparent. In the future, it will be important to build upon the breadth and depth of our current knowledge to illuminate the molecular mechanisms of known phenomena as well as to uncover novel modes of gene regulation.

Acknowledgments

We thank Steve Smale, Robert Tjian, Kevin Struhl, Alexandra Lusser, Tammy Juven-Gershon, Chin Yan Lim, Vassili Alexiadis, and Tom Boulay for critical reading of the manuscript. We thank our many colleagues that have contributed their invaluable advice and insight to this review. We also apologize to colleagues whose work has been inadvertently omitted, and accept responsibility for any inaccuracies. The study of RNA polymerase II core promoters in the laboratory of J.T.K. is supported by a grant from the NIH (GM41249).

References

- Adachi, N. and Lieber, M.R. 2002. Bidirectional gene organization: A common architectural feature of the human genome. *Cell* **109**: 807–809.
- Albright, S.R. and Tjian, R. 2000. TAFs revisited: More data reveal new twists and confirm old ideas. *Gene* **242**: 1–13.
- Antequera, F. and Bird, A. 1993. Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci.* **90**:

- 11995–11999.
- Aoyagi, N. and Wassarman, D.A. 2001. Developmental and transcriptional consequences of mutations in *Drosophila* TAF_{II}60. *Mol. Cell. Biol.* **21**: 6808–6819.
- Arkhipova, I.R. 1995. Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. *Genetics* **139**: 1359–1369.
- Benson, L.Q., Coon, M.R., Krueger, L.M., Han, G.C., Sarnaik, A.A., and Wechsler, D.S. 1999. Expression of *MXI1*, a Myc antagonist, is regulated by Sp1 and AP2. *J. Biol. Chem.* **274**: 28794–28802.
- Berk, A.J. 2000. TBP-like factors come into focus. *Cell* **103**: 5–8.
- Bird, A.P. 1986. CpG-rich islands and the function of DNA methylation. *Nature* **321**: 209–213.
- Bird, A.P. 1993. Functions for DNA methylation in vertebrates. *Cold Spring Harb. Symp. Quant. Biol.* **58**: 281–285.
- Blackwood, E.M. and Kadonaga, J.T. 1998. Going the distance: A current view of enhancer action. *Science* **281**: 60–63.
- Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., and Cedar, H. 1994. Sp1 elements protect a CpG island from de novo methylation. *Nature* **371**: 435–438.
- Breathnach, R. and Chambon, P. 1981. Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50**: 349–383.
- Bucher, P. 1990. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212**: 563–578.
- Bulger, M. and Groudine, M. 1999. Looping versus linking: Toward a model for long-distance gene activation. *Genes & Dev.* **13**: 2465–2477.
- Burke, T.W. and Kadonaga, J.T. 1996. *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes & Dev.* **10**: 711–724.
- Burke, T.W. and Kadonaga, J.T. 1997. The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAF_{II}60 of *Drosophila*. *Genes & Dev.* **11**: 3020–3031.
- Burke, T.W., Willy, P.J., Kutach, A.K., Butler, J.E.F., and Kadonaga, J.T. 1998. The DPE, a conserved downstream core promoter element that is functionally analogous to the TATA box. *Cold Spring Harb. Symp. Quant. Biol.* **63**: 75–82.
- Burley, S.K. and Roeder, R.G. 1996. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.* **65**: 769–799.
- Butler, J.E.F. and Kadonaga, J.T. 2001. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes & Dev.* **15**: 2515–2519.
- Carcamo, J., Buckbinder, L., and Reinberg, D. 1991. The initiator directs the assembly of a transcription factor IID-dependent transcription complex. *Proc. Natl. Acad. Sci.* **88**: 8052–8056.
- Chalkley, G.E. and Verrijzer, C.P. 1999. DNA binding site selection by RNA polymerase II TAFs: A TAF_{II}250-TAF_{II}150 complex recognizes the initiator. *EMBO J.* **18**: 4835–4845.
- Cheriyath, V., Novina, C.D., and Roy, A.L. 1998. TFII-I regulates V β promoter activity through an initiator element. *Mol. Cell. Biol.* **18**: 4444–4454.
- Collart, M.A. and Struhl, K. 1994. *NOT1* (CDC39), *NOT2* (CDC36), *NOT3*, and *NOT4* encode a global-negative regulator of transcription that differentially affects TATA-element utilization. *Genes & Dev.* **8**: 525–537.
- Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Keding, C., and Chambon, P. 1980. Promoter sequences of eukaryotic protein-coding genes. *Science* **209**: 1406–1414.
- Courey, A.J. and Jia, S. 2001. Transcriptional repression: The long and the short of it. *Genes & Dev.* **15**: 2786–2796.
- Crowley, T.E., Hoey, T., Liu, J.-K., Jan, Y.N., Jan, L.Y., and Tjian, R. 1993. A new factor related to TATA-binding protein has highly restricted expression patterns in *Drosophila*. *Nature* **361**: 557–561.
- Dantoni, J.-C., Quintin, S., Lakatos, L., Labouesse, M., and Tora, L. 2000. TBP-like factor is required for embryonic RNA polymerase II transcription in *C. elegans*. *Mol. Cell* **6**: 715–722.
- Dvir, A., Conaway, J.W., and Conaway, R.C. 2001. Mechanism of transcription initiation and promoter escape by RNA polymerase II. *Curr. Opin. Genet. Dev.* **11**: 209–214.
- Emami, K.H., Navarre, W.W., and Smale, S.T. 1995. Core promoter specificities of the Sp1 and VP16 transcriptional activation domains. *Mol. Cell. Biol.* **15**: 5906–5916.
- Emerson, B. 2002. Specificity of gene regulation. *Cell* **109**: 267–270.
- Evans, R., Fairley, J.A., and Roberts, S.G.E. 2001. Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. *Genes & Dev.* **15**: 2945–2949.
- Gardiner-Garden, M. and Frommer, M. 1987. CpG islands in vertebrate genomes. *J. Mol. Biol.* **196**: 261–282.
- Garaway, I.P., Semple, K., and Smale, S.T. 1996. Transcription of the lymphocyte-specific terminal deoxynucleotidyltransferase gene requires a specific core promoter structure. *Proc. Natl. Acad. Sci.* **93**: 4336–4341.
- Glass, C.K., and Rosenfeld, M.G. 2000. The coregulator exchange in transcriptional functions of nuclear receptors. *Genes & Dev.* **14**: 121–141.
- Goldberg, M.L. 1979. "Sequence analysis of *Drosophila* histone genes." Ph.D. dissertation, Stanford University, California.
- Gruenberg, D.A., Henry, R.W., Brauer, A., Novina, C.D., Cheriyath, V., Roy, A.L., and Gilman, M. 1997. A multifunctional DNA-binding protein that promotes the formation of serum response factor/homeodomain complexes: Identity to TFII-I. *Genes & Dev.* **11**: 2482–2493.
- Hampsey, M. 1998. Molecular genetics of the RNA polymerase II general transcription machinery. *Microbiol. Mol. Biol. Rev.* **62**: 465–503.
- Hansen, S.K., Takada, S., Jacobson, R.H., Lis, J.T., and Tjian, R. 1997. Transcription properties of a cell type-specific TATA-binding protein, TRF. *Cell* **91**: 71–83.
- Holmes, M.C. and Tjian, R. 2000. Promoter-selective properties of the TBP-related factor TRF1. *Science* **288**: 867–870.
- Hultmark, D., Klemenz, R., and Gehring, W.J. 1986. Translational and transcriptional control elements in the untranslated leader of the heat-shock gene *hsp22*. *Cell* **44**: 429–438.
- Ince, T.A. and Scotto, K.W. 1995. A conserved downstream element defines a new class of RNA polymerase II promoters. *J. Biol. Chem.* **270**: 30249–30252.
- Iyer, V. and Struhl, K. 1995. Mechanism of differential utilization of the *his3* T_R and T_C TATA elements. *Mol. Cell. Biol.* **15**: 7059–7066.
- Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., and Smale, S.T. 1994. DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell. Biol.* **14**: 116–127.
- Kaltenbach, L., Horner, M.A., Rothman, J.H. and Mango, S.E. 2000. The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during *C. elegans* embryogenesis. *Mol. Cell* **6**: 705–713.
- Kaufmann, J. and Smale, S.T. 1994. Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes & Dev.* **8**: 821–829.

- Kaufmann, J., Ahrens, K., Koop, R., Smale, S.T., and Müller, R. 1998. CIF150, a human cofactor for transcription factor IID-dependent initiator function. *Mol. Cell. Biol.* **18**: 233–239.
- Kuras, L., Kosa, P., Mencia, M., and Struhl, K. 2000. TAF-containing and TAF-independent forms of transcriptionally active TBP in vivo. *Science* **288**: 1244–1248.
- Kutach, A.K. and Kadonaga, J.T. 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. *Mol. Cell. Biol.* **20**: 4754–4764.
- Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H. 1998. New core promoter element in RNA polymerase II-dependent transcription: Sequence-specific DNA binding by transcription factor IIB. *Genes & Dev.* **12**: 34–44.
- Lee, T.I., and Young, R.A. 2000. Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* **34**: 77–137.
- Lefstin, J.A. and Yamamoto, K.R. 1998. Allosteric effects of DNA on transcriptional regulators. *Nature* **392**: 885–888.
- Lemon, B. and Tjian, R. 2000. Orchestrated response: A symphony of transcription factors for gene control. *Genes & Dev.* **14**: 2551–2569.
- Lewis, B.A., Kim, T.-K. and Orkin, S.H. 2000. A downstream element in the human β -globin promoter: Evidence of extended sequence-specific transcription factor IID contacts. *Proc. Natl. Acad. Sci.* **97**: 7172–7177.
- Li, X. and Noll, M. 1994. Compatibility between enhancers and promoters determines the transcriptional specificity of *gooseberry* and *gooseberry neuro* in the *Drosophila* embryo. *EMBO J.* **13**: 400–406.
- Li, X.-Y., Bhaumik, S.R., and Green, M.R. 2000. Distinct classes of yeast promoters revealed by differential TAF recruitment. *Science* **288**: 1242–1244.
- Li, X.-Y., Bhaumik, S.R., Zhu, X., Li, L., Shen, W.-C., Dixit, B.L. and Green, M.R. 2002. Selective recruitment of TAFs by yeast upstream activating sequences: Implications for eukaryotic promoter structure. *Curr. Biol.* **12**: 1240–1244.
- Lo, K. and Smale, S.T. 1996. Generality of a functional initiator consensus sequence. *Gene* **182**: 13–22.
- Luo, Y. and Roeder, R.G. 1995. Cloning, functional characterization, and mechanism of action of the B-cell-specific transcriptional coactivator OCA-B. *Mol. Cell. Biol.* **15**: 4115–4124.
- Luo, Y., Fujii, H., Gerster, T., and Roeder, R.G. 1992. A novel B cell-derived coactivator potentiates the activation of immunoglobulin promoters by octamer-binding transcription factors. *Cell* **71**: 231–241.
- Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. 1994. Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island. *Genes & Dev.* **8**: 2282–2292.
- Mahadevan, S. and Struhl, K. 1990. T_C, an unusual promoter element required for constitutive transcription of the yeast *HIS3* gene. *Mol. Cell. Biol.* **10**: 4447–4455.
- Maldonado, E. 1999. Transcriptional functions of a new mammalian TATA-binding protein-related factor. *J. Biol. Chem.* **274**: 12963–12966.
- Maldonado, E., Hampsey, M., and Reinberg, D. 1999. Repression: Targeting the heart of the matter. *Cell* **99**: 455–458.
- Martinez, E., Chiang, C.-M., Ge, H., and Roeder, R.G. 1994. TATA-binding protein-associated factor(s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. *EMBO J.* **13**: 3115–3126.
- Martinez, E., Ge, H., Tao, Y., Yuan, C.-X., Palhan, V., and Roeder, R.G. 1998. Novel cofactors and TFIIA mediate functional core promoter selectivity by the human TAF_{II}150-containing TFIID complex. *Mol. Cell. Biol.* **18**: 6571–6583.
- McKenna, N.J. and O'Malley, B.W. 2002. Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell* **108**: 465–474.
- Mencia, M., Moqtaderi, Z., Geisberg, J.V., Kuras, L., and Struhl, K. 2002. Activator-specific recruitment of TFIID and regulation of ribosomal protein genes in yeast. *Mol. Cell* **9**: 823–833.
- Merli, C., Bergstrom, D.E., Cygan, J.A., and Blackman, R.K. 1996. Promoter specificity mediates the independent regulation of neighboring genes. *Genes & Dev.* **10**: 1260–1270.
- Moore, P.A., Ozer, J., Salunek, M., Jan, G., Zerby, D., Campbell, S., and Lieberman, P.M. 1999. A human TATA binding protein-related protein with altered DNA binding specificity inhibits transcription from multiple promoters and activators. *Mol. Cell. Biol.* **19**: 7610–7620.
- Moqtaderi, Z., Bai, Y., Poon, D., Weil, P.A., and Struhl, K. 1996. TBP-associated factors are not generally required for transcriptional activation in yeast. *Nature* **383**: 188–191.
- Moqtaderi, Z., Keaveney, M., and Struhl, K. 1998. The histone H3-like TAF is broadly required for transcription in yeast. *Mol. Cell* **2**: 675–682.
- Myer, V.E. and Young, R.A. 1998. RNA polymerase II holoenzymes and subcomplexes. *J. Biol. Chem.* **273**: 27757–27760.
- Nakatani, Y., Brenner, M., and Freese, E. 1990a. An RNA polymerase II promoter containing sequences upstream and downstream from the RNA startpoint that direct initiation of transcription from the same site. *Proc. Natl. Acad. Sci.* **87**: 4289–4293.
- Nakatani, Y., Horikoshi, M., Brenner, M., Yamamoto, T., Benard, F., Roeder, R.G., and Freese, E. 1990b. A downstream initiation element required for efficient TATA box binding and *in vitro* function of TFIID. *Nature* **348**: 86–88.
- Narlikar, G.J., Fan, H.-Y., and Kingston, R.E. 2002. Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108**: 475–487.
- Oelgeschläger, T., Chiang, C.-M., and Roeder, R.G. 1996. Topology and reorganization of a human TFIID-promoter complex. *Nature* **382**: 735–738.
- Ohbayashi, T., Makino, Y., and Tamura, T. 1999. Identification of a mouse TBP-like protein (TLP) distantly related to the *Drosophila* TBP-related factor. *Nucleic Acids Res.* **27**: 750–755.
- Ohtsuki, S., Levine, M., and Cai, H.N. 1998. Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. *Genes & Dev.* **12**: 547–556.
- Orphanides, G., Lagrange, T. and Reinberg, D. 1996. The general transcription factors of RNA polymerase II. *Genes & Dev.* **10**: 2657–2683.
- Orphanides, G. and Reinberg, D. 2002. A unified theory of gene expression. *Cell* **108**: 439–451.
- Pribnow, D. 1975a. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc. Natl. Acad. Sci.* **72**: 784–788.
- Pribnow, D. 1975b. Bacteriophage T7 early promoters: Nucleotide sequences of two RNA polymerase binding sites. *J. Mol. Biol.* **99**: 419–443.
- Purnell, B.A., Emanuel, P.A., and Gilmour, D.S. 1994. TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes. *Genes & Dev.* **8**: 830–842.
- Qureshi, S.A. and Jackson, S.P. 1998. Sequence-specific DNA binding by the *S. shibatae* TFIIB homolog, TFB, and its effect on promoter strength. *Mol. Cell* **1**: 389–400.
- Rabenstein, M.D., Zhou, S., Lis, J.T., and Tjian, R. 1999. TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family. *Proc. Natl. Acad. Sci.* **96**: 4791–

- 4796.
- Roeder, R.G. 1998. Role of general and gene-specific cofactors in the regulation of eukaryotic transcription. *Cold Spring Harb. Symp. Quant. Biol.* **58**: 201–218.
- Roy, A.L., Du, H., Gregor, P.D., Novina, C.D., Martinez, E., and Roeder, R.G. 1997. Cloning of an Inr- and E-box-binding protein, TFIID-I, that interacts physically and functionally with USF1. *EMBO J.* **16**: 7091–7104.
- Roy, A.L., Meisterernst, M., Pognonec, P., and Roeder, R.G. 1991. Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. *Nature* **354**: 245–248.
- Rudge, T.L. and Johnson, L.F. 1999. Inactivation of MED-1 elements in the TATA-less, initiator-less mouse thymidylate synthase promoter has no effect on promoter strength or the complex pattern of transcriptional start sites. *J. Cell. Biochem.* **73**: 90–96.
- Simon, M.C., Fisch, T.M., Benecke, B.J., Nevins, J.R., and Heintz, N. 1988. Definition of multiple, functionally distinct TATA elements, one of which is a target in the *hsp70* promoter for E1A recognition. *Cell* **52**: 723–729.
- Singer, V.L., Wobbe, C.R., and Struhl, K. 1990. A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. *Genes & Dev.* **4**: 636–645.
- Smale, S.T. 1994. Core promoter architecture for eukaryotic protein-coding genes. In *Transcription: Mechanisms and regulation* (eds R.C. Conaway and J.W. Conaway), pp. 63–81. Raven Press, Ltd., New York.
- Smale, S.T. 1997. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim. Biophys. Acta* **1351**: 73–88.
- Smale, S.T. 2001. Core promoters: Active contributors to combinatorial gene regulation. *Genes & Dev.* **15**: 2503–2508.
- Smale, S.T. and Baltimore, D. 1989. The “initiator” as a transcription control element. *Cell* **57**: 103–113.
- Smale, S.T., Jain, A., Kaufmann, J., Emami, K.H., Lo, K., and Garraway, I.P. 1998. The initiator element: A paradigm for core promoter heterogeneity within metazoan protein-coding genes. *Cold Spring Harb. Symp. Quant. Biol.* **58**: 21–31.
- Smale, S.T., Schmidt, M.C., Berk, A.J., and Baltimore, D. 1990. Transcriptional activation by Sp1 as directed through TATA or initiator: Specific requirement for mammalian transcription factor IID. *Proc. Natl. Acad. Sci.* **87**: 4509–4513.
- Soldatov, A., Nabirochkina, E., Georgieva, S., Belenkaja, T., and Georgiev, P. 1999. TAF_{II}40 protein is encoded by the *e(y)1* gene: Biological consequences of mutations. *Mol. Cell. Biol.* **19**: 3769–3778.
- Strahl, B.D. and Allis, C.D. 2000. The language of covalent histone modifications. *Nature* **403**: 41–45.
- Struhl, K. 1986. Constitutive and inducible *Saccharomyces cerevisiae* promoters: Evidence for two distinct molecular mechanisms. *Mol. Cell. Biol.* **6**: 3847–3853.
- Struhl, K. 1987. Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. *Cell* **49**: 295–297.
- Struhl, K. 1999. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* **98**: 1–4.
- Suzuki, Y., Tsunoda, T., Sese, J., Taira, H., Mizushima-Sugano, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y., et al. 2001. Identification and characterization of the potential promoter regions of 1031 kinds of human genes. *Genome Res.* **11**: 677–684.
- Takada, S., Lis, J.T., Zhou, S., and Tjian, R. 2000. A TRF1:BRF complex directs *Drosophila* RNA polymerase III transcription. *Cell* **101**: 459–469.
- Teichmann, M., Wang, Z., Martinez, E., Tjernberg, A., Zhang, D., Vollmer, F., Chait, B.T., and Roeder, R.G. 1999. Human TATA-binding protein-related factor-2 (hTRF2) stably associates with hTFIIA in HeLa cells. *Proc. Natl. Acad. Sci.* **96**: 13720–13725.
- Tora, L. 2002. A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes & Dev.* **16**: 673–675.
- Tsai, F.T.F. and Sigler, P.B. 2000. Structural basis of preinitiation complex assembly on human Pol II promoters. *EMBO J.* **19**: 25–36.
- Usheva, A. and Shenk, T. 1994. TATA-binding protein-independent initiation: YY1, TFIIB, and RNA polymerase II direct basal transcription on supercoiled template DNA. *Cell* **76**: 1115–1121.
- Verrijzer, C.P. 2001. Transcription factor IID—Not so basal after all. *Science* **293**: 2010–2011.
- Verrijzer, C.P., Yokomori, K., Chen, J.-L. and Tjian, R. 1994. *Drosophila* TAF_{II}150: Similarity to yeast gene TSM-1 and specific binding to core promoter DNA. *Science* **264**: 933–941.
- Verrijzer, C.P., Chen, J.-L., Yokomori, K., and Tjian, R. 1995. Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* **81**: 1115–1125.
- Wefald, F.C., Devlin, B.H., and Williams, R.S. 1990. Functional heterogeneity of mammalian TATA-box sequences revealed by interaction with a cell-specific enhancer. *Nature* **344**: 260–262.
- Weis, L. and Reinberg, D. 1992. Transcription by RNA polymerase II: Initiator-directed formation of transcription-competent complexes. *FASEB J.* **6**: 3300–3309.
- Weis, L. and Reinberg, D. 1997. Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. *Mol. Cell. Biol.* **17**: 2973–2984.
- West, A.G., Gaszner, M., and Felsenfeld, G. 2002. Insulators: Many functions, many mechanisms. *Genes & Dev.* **16**: 271–288.
- White, R.J. 2001. *Gene transcription: Mechanisms and control*, Blackwell Science, Oxford, UK.
- Willy, P.J., Kobayashi, R., and Kadonaga, J.T. 2000. A basal transcription factor that activates or represses transcription. *Science* **290**: 982–984.
- Woychik, N.A. and Hampsey, M. 2002. The RNA polymerase II machinery: Structure illuminates function. *Cell* **108**: 453–463.
- Zhang, Y. and Reinberg, D. 2001. Transcription regulation by histone methylation: Interplay between different covalent modifications of the core histone tails. *Genes & Dev.* **15**: 2343–2360.
- Zhou, T. and Chiang, C.-M. 2001. The intronless and TATA-less human TAF_{II}55 gene contains a functional initiator and a downstream promoter element. *J. Biol. Chem.* **276**: 25503–25511.