EXPERIMENTAL and MOLECULAR MEDICINE, Vol. 34, No. 4, 259-264, September 2002

The DPE, a core promoter element for transcription by RNA polymerase II

James T. Kadonaga[1,2]

[1] Section of Molecular Biology, University of California, San Diego, La Jolla, CA 92093-0347, USA.
[2] Correspondence: Tel, +1-858-534-4608; Fax, +1-858-534-0555; E-mail, jkadonaga@ucsd.edu

Accepted 29 Augest 2002

Abbreviations: BRE, TFIIB recognition element; DPE, downstream core promoter element; Inr, initiator element; LINE, long interspersed nuclear element; NC2, negative cofactor 2 (NC2 is also known as Dr1-Drap1); nt, nucleotides; TAF, TBP-associated factor; TBP, TATA box-binding protein; TFIIB, RNA polymerase II basal transcription factor B; TFIID, RNA polymerase II basal transcription factor D.

Overview

The core promoter is an important yet often overlooked component in the regulation of transcription by RNA polymerase II. In fact, the core promoter is the ultimate target of action of all of the factors and coregulators that control the transcriptional activity of every gene. In this review, I describe our current knowledge of a downstream core promoter element termed the DPE, which is a TFIID recognition site that is conserved from *Drosophila* to humans. The DPE is located from +28 to +32 relative to the +1 transcription start site, and is mainly present in core promoters that lack a TATA box motif. Moreover, in *Drosophila*, the DPE appears to be about as common as the TATA box. There are distinct mechanisms of basal transcription from DPE- versus TATA-dependent core promoters. For instance, NC2/Dr1-Drap1 is a repressor of TATA-dependent transcription and an activator of DPE-dependent transcription. In addition, DPE-specific and TATA-specific transcriptional enhancers have been identified. These findings further indicate that the core promoter is an active participant in the regulation of eukaryotic gene expression.

Keywords: DPE, TATA, Inr, core promoter, RNA polymerase II

Regulation of Transcription by RNA Polymerase II

The eukaryotic cell is confronted with the challenge of properly regulating each of its tens of thousands of genes. When it is considered that each gene has its own unique expression program, it becomes evident that the control of gene activity requires an enormous amount of resources in terms of information (*i.e.*, instructions for the regulation of each gene) and effectors (*i.e.*, factors that mediate the gene expression programs).

Transcription is a key step at which gene activity is controlled. Much of the information that specifies the transcriptional program of a gene is encoded in its DNA sequence. These cis-acting sequences include enhancers, silencers, proximal promoter regions, core promoters, and boundary/insulator elements (see, for example: Struhl, 1987; Weis and Reinberg, 1992; Smale, 1994, 1997, 2001; Blackwood and Kadonaga, 1998; Bulger and Groudine, 1999; Butler and Kadonaga, 2002; West *et al.*, 2002). Enhancer and silencer elements contain recognition sites for a variety of sequence-specific DNA-binding factors, and can act from long distances (such as tens of kbp) from the transcription start site. The proximal promoter region also contains multiple recognition sites for sequence-specific DNA-binding factors, and is typically located from about -40 to about -250 relative to the +1 start site. The core promoter is generally located within -40 to +40 of the start site, and is recognized by the basal RNA polymerase II transcriptional machinery. Boundary/insulator elements act to block the long-range influence of enhancers and silencers.

Trans-acting factors are the effectors of the transcriptional programs of genes. These factors include RNA polymerase II and the basal/general transcription factors (*i.e.*, the basal transcriptional machinery), sequence-specific DNA-binding factors that interact with promoters and enhancers, ATP-utilizing chromatin remodeling factors that mobilize nucleosomes, transcriptional mediators that promote interactions between enhancer-binding factors and the basal transcriptional machinery, and an assortment of enzymes that catalyze acetylation, deacetylation, phosphorylation, ubiquitinylation, and methylation of histones and other proteins (see, for example: Burley and Roeder, 1996; Orphanides *et al.*, 1996; Hampsey 1998; Lefstin and Yamamoto, 1998; Myer and Young, 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; Dvir *et al.*, 2001; White, 2001; Zhang and Reinberg, 2001; Emerson, 2002; McKenna and O'Malley, 2002; Narlikar *et al.*, 2002; Orphanides and Reinberg, 2002).

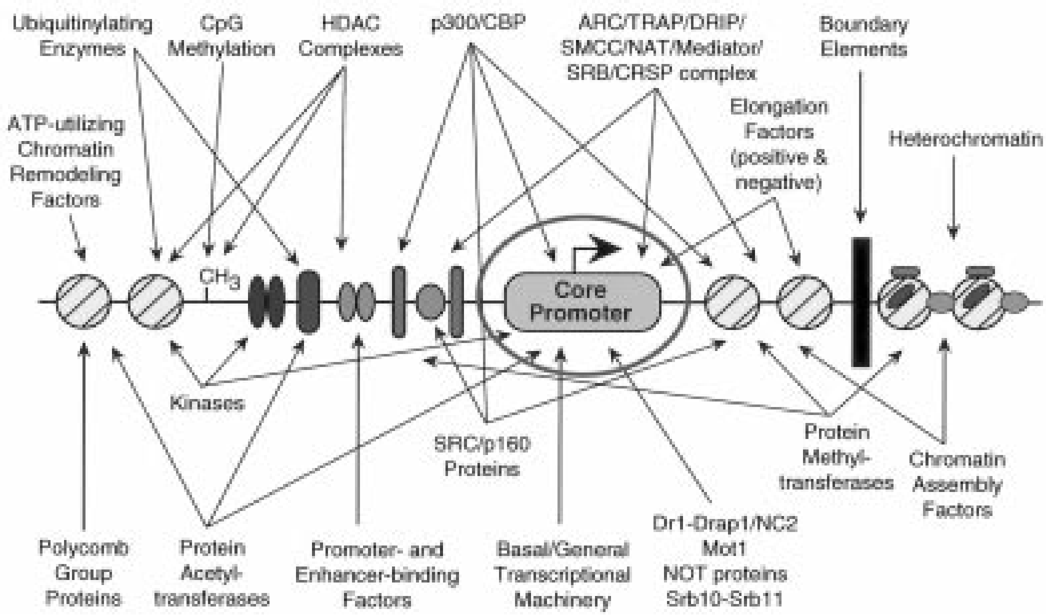

Figure 1. The core promoter is the ultimate target of factors that regulate transcription by RNA polymerase II.

Yet, in the midst of the complexity of transcriptional regulation, it is important to note that the ultimate target of action of all of the transcription factors and coregulators is the core promoter (Figure 1). The core promoter is an important but often overlooked component in the regulation of transcription by RNA polymerase II. Hence, in this review, I will focus on the core promoter. Then, more specifically, I will discuss the DPE, which is a conserved downstream core promoter element.

The RNA Polymerase II Core Promoter

The core promoter encompasses the transcription start site and typically extends ~35 nt either upstream or downstream from the +1 start site. A key function of the core promoter is to direct the initiation of transcription by the basal RNA polymerase II machinery. It is often incorrectly assumed that all core promoters are essentially the same. There is, in fact, considerable variability in the DNA elements that constitute core promoters. These elements include the TATA box, the TFIIB recognition element (BRE), the initiator (Inr), and the downstream core promoter element (DPE). It is important to note that there are no universal core promoter elements. Each of these motifs is found in only a subset of core promoters. In this section, I will briefly describe the TATA, BRE, and Inr core promoter elements.

The TATA box was the first eukaryotic core promoter motif to be identified (Goldberg 1979; Breathnach and Chambon, 1981). It is typically located about 25 to 30 nt upstream of the transcription start site, and has a consensus sequence of TATAAA. The TATA box is bound by the TATA box-binding protein (TBP) subunit of the TFIID complex. It is sometimes incorrectly thought that the TATA box is a component of all core promoters. In fact, the TATA box is found in less than half of all core promoters. For instance, a putative TATA box motif was identified in ~43% of 205 *Drosophila* core promoters (Kutach and Kadonaga, 2000) and ~32% of 1031 human core promoters (Suzuki *et al*., 2001).

A subset of TATA boxes possess an upstream sequence termed the BRE, which is a recognition site for the binding of TFIIB (Lagrange *et al*., 1998). The consensus sequence for the BRE is G/C-G/C-G/A-C-G-C-C, where the 3' C of the BRE is immediately followed by the 5' T of the TATA box. In the analysis of 315 TATA-containing promoters, a motif with at least a 5 out of 7 match with the BRE consensus was found in 12% of the promoters (Lagrange *et al*., 1998). Under different experimental conditions, the BRE has been found to exhibit both positive and negative effects upon transcriptional activity (Lagrange *et al*., 1998; Evans *et al*., 2001).

The Inr element encompasses the transcription start site. The Inr was identified in mammals, *Drosophila*, and yeast (see, for instance: Corden *et al*., 1980; Breathnach and Chambon, 1981; Hultmark *et al*., 1986; Struhl, 1987), and was later defined as a discrete functional element (Smale and Baltimore, 1989; Smale 1994, 1997). The consensus for the Inr is Py-Py(C)-A_{+1}-N-T/A-Py-Py in mammals (Corden *et al*., 1980; Bucher, 1990; Javahery *et al*., 1994; Lo and Smale, 1996; Smale *et al.,* 1998) and T-C-A_{+1}-G/T-T-C/T in *Drosophila* (Hultmark *et al*., 1986; Arkhipova, 1995; Purnell *et al*., 1994; Kutach and

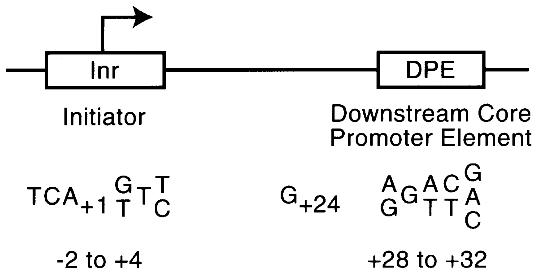

Figure 2. Features of DPE-dependent core promoters. DPE-dependent core promoters require both DPE and Inr motifs for the cooperative binding of TFIID. The DPE is located precisely at +28 to +32 relative to the A_{+1} nucleotide in the Inr. In addition, at position +24, a G nucleotide is preferred. The Inr and DPE consensus sequences are for *Drosophila* core promoters.

Kadonaga, 2000). Transcription will commonly (but not always) initiate at the A_{+1} nucleotide.

The DPE

The DPE is a downstream core promoter element

The DPE was identified in a study of the binding of purified *Drosophila* TFIID to TATA-less core promoters (Burke and Kadonaga, 1996). The DPE is most commonly found in TATA-less promoters, although some promoters contain both DPE and TATA motifs (Kutach and Kadonaga, 2000). The core of the DPE motif is located precisely at +28 to +32 relative to the A_{+1} nucleotide of the Inr (Kutach and Kadonaga, 2000; Butler and Kadonaga, 2002). In addition, there are small (about two-fold) yet distinct nucleotide preferences at specific positions between the Inr and DPE motifs. For example, a G nucleotide is preferred at position +24 (Kutach and Kadonaga, 2000). The DPE consensus is depicted in Figure 2. Although the DPE was originally identified in *Drosophila*, it is also present in humans (see, for example: Burke and Kadonaga, 1997; Zhou and Chiang, 2001). Thus, the DPE is conserved from *Drosophila* to humans. It remains to be determined whether or not a homologue of the DPE is present in the yeast *Saccharomyces cerevisiae*.

The DPE functions in coordination with the Inr

By DNase I footprinting analysis of wild-type and mutant core promoters, it was found that TFIID binds cooperatively to the Inr and DPE motifs (Burke and Kadonaga, 1996). In addition, core promoter activity is strongly reduced (typically by about 20- to 50-fold) upon mutation of either the Inr or DPE motifs in DPE-containing promoters (Burke and Kadonaga, 1996, 1997; Kutach and Kadonaga, 2000). In fact, all of the ~18 confirmed DPE-dependent *Drosophila* core promoters have been found to possess the identical spacing between the Inr and DPE motifs (Burke and Kadonaga 1996, 1997; Kutach and Kadonaga 2000). Moreover, the insertion or deletion of a single nucleotide between the Inr and DPE motifs causes a several-fold reduction in the binding of TFIID and basal transcription activity (Kutach and Kadonaga, 2000). These observations indicate that DPE-dependent basal transcription involves the precise cooperative binding of TFIID to the DPE and Inr motifs.

The DPE versus the TATA box

In some respects, the DPE is a downstream counterpart to the TATA box. For instance, when a TATA-dependent promoter is inactivated by mutation of the TATA motif, transcriptional activity can be restored to the mutant TATA promoter by the addition of a DPE sequence at its normal (+28 to +32) downstream position (Burke and Kadonaga, 1996). How common is the DPE relative to the TATA box? In an analysis of 205 core promoters in *Drosophila*, it was estimated that 29% of the promoters contain a TATA box only (no DPE), 26% contain a DPE only (no TATA), 14% contain both TATA and DPE motifs, and 31% do not appear to contain either a TATA or a DPE motif (Kutach and Kadonaga, 2000). Thus, it appears that the DPE is about as common as the TATA box in *Drosophila*. It will also be interesting to determine the frequency of occurrence of the DPE in human core promoters.

What interacts with the DPE?

Because the DPE was originally identified as an element that is required for the binding of TFIID, some subunit (or subunits) of TFIID is likely to interact with the DPE motif. First, it was found that the TBP subunit of TFIID, which binds to TATA boxes, does not appear to bind to the DPE (Burke and Kadonaga, 1996). Then, photocrosslinking studies with purified TFIID revealed that the $dTAF_{II}60$ and $dTAF_{II}40$ subunits of TFIID are in close proximity to the DPE motif (Burke and Kadonaga, 1997). [Note that the TAF nomenclature has been recently revised (Tora, 2002). In the new nomenclature, $dTAF_{II}60$ and $dTAF_{II}40$ are TAF6 and TAF9, respectively.] Thus, TAF6 and TAF9, which possess histone fold motifs that resemble those in histones H4 and H3, may interact with the DPE, possibly as an $\alpha_2\beta_2$ heterotetramer that is analogous to the $(H3\text{-}H4)_2$ tetramer. Genetic studies of TAF6 and TAF9 in *Drosophila* have revealed that these polypeptides are encoded by essential genes (Soldatov *et al.*, 1999; Aoyagi and Wassarman, 2001), but the effect of the loss of either of these TAFs upon DPE-dependent transcription in vivo remains to be determined. Lastly, it should be noted that factors other than TFIID may also interact with the DPE.

NC2/Dr1-Drap1 discriminates between DPE- and TATA-dependent promoters

It is often assumed that all core promoters function by the same mechanism. In fact, there are a variety of mechanisms by which core promoters function. Thus, the core promoter not only directs the initiation of transcription by RNA polymerase II, but is also a cis-acting regulatory element (for recent reviews, see: Smale, 2001; Butler and Kadonaga, 2002).

In the analysis of DPE-dependent transcription, an activity that stimulates DPE-dependent transcription and represses TATA-dependent transcription was identified (Willy *et al*., 2000). Upon purification, this activity was found to be mediated by a factor termed NC2 (negative cofactor 2; also known as Dr1-Drap1). NC2 was originally identified as a repressor of TATA-dependent transcription (for review, see: Maldonado *et al*., 1999), and its ability to activate DPE-dependent transcription was unexpected. In addition, a mutant version of NC2 was found to be able to activate DPE-dependent transcription but unable to repress TATA-dependent transcription. Hence, the ability of NC2 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 TATA- and DPE-dependent core promoters, and reflect the fundamental differences in the mechanisms of TATA- versus DPE-driven basal transcription.

DPE motifs are used as downstream core promoter elements in LINEs (long interspersed nuclear elements)

Why might a gene contain a DPE or TATA motif in its core promoter? One possibility is that the gene may require a downstream promoter region. That is, the transcription unit, including the regulatory sequences, may need to be located downstream of the +1 start site. In fact, this situation is indeed the case with non-LTR retrotransposons termed LINEs. These retrotransposons, which include the jockey, Doc, G, I, and F elements in *Drosophila*, possess DPE motifs in their core promoters. They propagate via the use of internal promoters that are entirely downstream of the transcription start site. Thus, the *Drosophila* LINE promoters provide examples in which DPE motifs are used as downstream core promoter elements in vivo.

Identification of DPE-specific transcriptional enhancers

Transcriptional enhancers are able to activate transcription over tens of kbp, and therefore must be able to interact specifically with their cognate promoters (for recent review on enhancer-promoter specificity, see Butler and Kadonaga, 2002). It therefore seemed possible that DPE motifs could be involved in enhancer-promoter specificity. To test this hypothesis, the relative ability of enhancers to activate transcription from TATA- versus DPE-dependent promoters was tested in vivo in *Drosophila* (Butler and Kadonaga, 2001). These studies led to the identification of DPE-specific as well as TATA-specific transcriptional enhancers. Out of 18 enhancers tested, three were specific for a DPE-dependent core promoter, and one was specific for a TATA-dependent core promoter. Moreover, primer extension analysis revealed no detectable TATA-dependent transcription with the DPE-specific enhancers. The remaining 14 enhancers activated transcription from both DPE- and TATA-dependent promoters. In a separate study involving promoter competition, it was found that the *Drosophila* AE1 and IAB5 enhancers preferentially activate transcription from the TATA-containing *even-skipped* core promoter relative to the TATA-less (and weak DPE-containing) *white* core promoter (Ohtsuki *et al*., 1998). These studies collectively indicate that some transcriptional enhancers function specifically with DPE- or TATA-dependent core promoters. In this manner, the presence or absence of a DPE or TATA motif in the core promoter might be an important component in the regulation of a gene. In addition, it is possible that core promoters with both DPE and TATA motifs could be bifunctional promoters that are able to interact with both DPE- and TATA-specific enhancers.

Summary and Perspectives

In conclusion, it is important to consider the core promoter as an active participant in the regulation of gene expression. There are a variety of core promoter elements as well as multiple distinct mechanisms of basal transcription. It is also likely that there are additional core promoter motifs that have yet to be discovered and characterized. In the future, there will be many exciting and important experiments to be performed on the role of core promoter motifs in transcriptional regulation.

Acknowledgments

I thank Sharon Wampler, Curtis Tyree, Catherine George Takizawa, Tom Burke, Alan Kutach, Patricia Willy, Jenny Butler, Scott Iyama, Tammy Juven-Gershon, Tom Boulay, Chin Yan Lim, M. Cristina Ramones, and Emily Dong for their analysis of basal RNA polymerase II transcription in my laboratory. The study of basal RNA polymerase II transcription in the laboratory of J.T.K. is supported by a grant from the NIH (GM41249).

References

Aoyagi N, Wassarman DA. Developmental and transcriptional consequences of mutations in *Drosophila* $TAF_{II}60$. Mol Cell

Biol 2001;21:6808-19

Arkhipova IR. Promoter elements in *Drosophila* melanogaster revealed by sequence analysis. Genetics 1995;139:1359-69

Blackwood EM, Kadonaga JT. Going the distance: a current view of enhancer action. Science 1998;281:60-3

Breathnach R, Chambon P. Organization and expression of eucaryotic split genes coding for proteins. Annu Rev Biochem 1981;50:349-83

Bucher P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. J Mol Biol 1990;212:563-78

Bulger M, Groudine M. Looping versus linking: toward a model for long-distance gene activation. Genes Dev 1999;13: 2465-77

Burley SK, Roeder RG. Biochemistry and structural biology of transcription factor IID (TFIID). Annu Rev Biochem 1996;65: 769-99

Burke TW, Kadonaga JT. *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes Dev 1996;10:711-24

Burke TW, Kadonaga JT. The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAF_{II}60 of *Drosophila*. Genes Dev 1997;11: 3020-31

Butler JEF, Kadonaga JT. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. Genes Dev 2001;15:2515-9

Butler JEF, Kadonaga JT. The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes Dev 2002; in press

Corden J, Wasylyk B, Buchwalder A, Sassone-Corsi P, Kedinger C, Chambon P. Promoter sequences of eukaryotic protein-coding genes. Science 1980; 209:1406-14

Courey AJ, Jia S. Transcriptional repression: the long and the short of it. Genes Dev 2001;15:2786-96

Dvir A, Conaway JW, Conaway RC. Mechanism of transcription initiation and promoter escape by RNA polymerase II. Curr Opin Genet Dev 2001;11:209-14

Emerson B. Specificity of gene regulation. Cell 2002;109:267-70

Evans R, Fairley JA, Roberts SGE. Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. Genes Dev 2001;15:2945-9

Glass CK, Rosenfeld MG. The coregulator exchange in transcriptional functions of nuclear receptors. Genes Dev 2000;14:121-41

Goldberg ML. 1979. Sequence analysis of *Drosophila* histone genes. Ph.D. Dissertation. Stanford University.

Hampsey M. Molecular genetics of the RNA polymerase II general transcription machinery. Microbiology and Molecular Biology Reviews 1998;62:465-503

Hultmark D, Klemenz R, Gehring WJ. Translational and transcriptional control elements in the untranslated leader of the heat-shock gene hsp22. Cell 1986;44:429-38

Javahery R, Khachi A, Lo K, Zenzie-Gregory B, Smale ST. DNA sequence requirements for transcriptional initiator activity in mammalian cells. Mol Cell Biol 1994;14:116-27

Kutach AK, Kadonaga JT. The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. Mol Cell Biol 2000;20:4754-64

Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev 1998;12:34-44

Lee TI, Young RA. Transcription of eukaryotic protein-coding genes. Annu Rev Genet 2000;34:77-137

Lefstin JA, Yamamoto KR. Allosteric effects of DNA on transcriptional regulators. Nature 1998;392:885-8

Lemon B, Tjian R. Orchestrated response: a symphony of transcription factors for gene control. Genes Dev 2000;14: 2551-69

Lo K, Smale ST. Generality of a functional initiator consensus sequence. Gene 1996;182:13-22

Maldonado E, Hampsey M, Reinberg D. Repression: targeting the heart of the matter. Cell 1999;99:455-8

McKenna NJ, O'Malley BW. Combinatorial control of gene expression by nuclear receptors and coregulators. Cell 2002; 108:465-74

Myer VE, Young RA. RNA polymerase II holoenzymes and subcomplexes. J Biol Chem 1998;273:27757-60

Narlikar GJ, Fan HY, Kingston RE. Cooperation between complexes that regulate chromatin structure and transcription. Cell 2002;108:475-87

Ohtsuki S, Levine M, Cai HN. Different core promoters possess distinct regulatory activities in the *Drosophila* embryo. Genes Dev 1998;12:547-56

Orphanides G, Lagrange T, Reinberg D. The general transcription factors of RNA polymerase II. Genes Dev 1996; 10:2657-83

Orphanides G, Reinberg D. A unified theory of gene expression. Cell 2002;108:439-51

Purnell BA, Emanuel PA, Gilmour DS. TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes. Genes Dev 1994;8:830-42

Roeder RG. Role of general and gene-specific cofactors in the regulation of eukaryotic transcription. Cold Spring Harbor Symp Quant Biol 1998;58:201-18

Smale ST, Baltimore D. The “initiator” as a transcription control element. Cell 1989;57:103-13

Smale ST. Core promoter architecture for eukaryotic protein-coding genes. In *Transcription: Mechanisms and Regulation.* (Conaway RC, Conaway JW, eds), 1994, 63-81, Raven Press, New York, NY

Smale ST. Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. Biochim Biophys Acta 1997;1351:73-88

Smale ST, Jain A, Kaufmann J, Emami KH, Lo K, Garraway IP. The initiator element: a paradigm for core promoter heterogeneity within metazoan protein-coding genes. Cold Spring Harbor Symp Quant Biol 1998;58:21-31

Smale ST. Core promoters: active contributors to combinatorial gene regulation. Genes Dev 2001;15:2503-8

Soldatov A, Nabirochkina E, Georgieva S, Belenkaja T, Georgiev P. TAF_{II}40 protein is encoded by the e(y)1 gene: biological consequences of mutations. Mol Cell Biol 1999;19: 3769-78

Strahl BD, Allis CD. The language of covalent histone modifications. Nature 2000;403:41-5

Struhl K. Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast. Cell 1987;49:295-7

Struhl, K. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. Cell 1999;98:1-4

Suzuki Y, Tsunoda T, Sese J, Taira H, Mizushima-Sugano J, Hata H, Ota T, Isogai T, Tanaka T, Nakamura Y, Suyama A, Sakaki Y, Morishita S, Okubo K, Sugano S. Identification and characterization of the potential promoter regions of 1031 kinds of human genes. Genome Res 2001;11:677-84

Tora L. A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. Genes Dev 2002;16:673-5

Weis L, Reinberg D. Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. FASEB J 1992;6:3300-9

West AG, Gaszner M, Felsenfeld G. Insulators: many functions, many mechanisms. Genes Dev 2002;16:271-88

White RJ. 2001. Gene Transcription: Mechanisms and Control, Blackwell Science., Oxford, UK

Willy PJ, Kobayashi R, Kadonaga JT. A basal transcription factor that activates or represses transcription. Science 2000; 290:982-4

Zhang Y, Reinberg D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev 2001;15:2343-60

Zhou T, Chiang CM. The intronless and TATA-less human TAF_{II}55 gene contains a functional initiator and a downstream promoter element. J Biol Chem 2001;276:25503-11