Regulation of gene expression via the core promoter and the basal transcriptional machinery

Tamar Juven-Gershon, James T. Kadonaga

Section of Molecular Biology, 0347, University of California, San Diego, La Jolla, CA 92093-0347, USA

Abstract

The RNA polymerase II core promoter is a structurally and functionally diverse transcriptional regulatory element. There are two main strategies for transcription initiation – focused and dispersed initiation. In focused initiation, transcription starts from a single nucleotide or within a cluster of several nucleotides, whereas in dispersed initiation, there are several weak transcription start sites over a broad region. Focused promoters exhibit the combined qualities of both focused and dispersed promoters.

Introduction

The core promoter lies at the center of the transcription process, yet it is often an overlooked component in the regulation of gene expression. The core promoter serves as a DNA region that directs the accurate initiation of transcription by RNA polymerase II. In the past, the core promoter has been presumed to be a generic entity that functions by a single universal mechanism, but it is now clear that there is widespread diversity in core promoter structure and function. In this review, we will discuss some key features of the RNA polymerase II core promoter and provide some examples of the role of the core promoter and the basal transcription machinery in the regulation of gene expression.

Focused versus dispersed transcription initiation

Examination of the patterns of transcription initiation reveals two different modes of transcription initiation – focused and dispersed.
dispersed promoters because of the biological significance of the regulated genes with which the focused promoters are associated. The analysis of focused core promoters has led to the discovery of sequence motifs such as the TATA box, BRUs (upstream TFIIB recognition element), Inr (initiator), MTE (motif ten element), DPE (downstream promoter element), DCE (downstream core element), and XCPE1 (X core promoter element 1) (Fig. 2). In contrast, dispersed promoters generally lack BRE, TATA, DPE, and MTE motifs (Sandelin et al., 2007; Carninci et al., 2006). It is likely that there are fundamental differences in the mechanisms of transcription from focused versus dispersed promoters. For the remainder of this review, we will mainly describe studies of focused core promoters.

**Basal transcription factors**

The focused core promoter, which typically encompasses −40 to +40 relative to the +1 transcription start site, is the location at which the RNA polymerase II machinery initiates transcription. Purified RNA polymerase II can synthesize RNA from a DNA template but is not able to recognize the core promoter. This process requires additional factors that are commonly known as the “general” or “basal” transcription factors, which include TFIIA (transcription factor for RNA polymerase II A), TFIIB, TFIID, TFIIE, TFIIF, and TFIIH. These factors do not act in a “general” manner at all core promoters, and hence, we will refer to them as the “basal” transcription factors.

With TATA-driven core promoters, transcription can be achieved in vitro with purified RNA polymerase II, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH. However, these same factors are not able to mediate transcription from a DPE-driven promoter (Lewis et al., 2005). In addition, NC2 (negative cofactor 2; also known as Dr1-Drap1), which was identified as repressor of TATA-dependent transcription, was found to be an activator of DPE-dependent transcription (Willy et al., 2000; Hsu et al., 2008). TATA- versus DPE-dependent transcription appears to be controlled, at least in part, by a simple circuit in which TBP (TATA box-binding protein) activates TATA transcription and represses DPE transcription, whereas NC2 and Mot1 (an ATPase that removes TBP from DNA) block TBP function (Hsu et al., 2008; van Werven et al., 2008) and thus promote DPE transcription and repress TATA transcription (Hsu et al., 2008).

TFIID is a key basal transcription factor that is involved in the recognition of focused core promoters (for review, see Thomas and Chiang, 2006). TFIID is a multisubunit complex that comprises TBP and about a dozen TAFs (TBP-associated factors). There are multiple potential points of interaction of TFIID with the core promoter. The TBP subunit binds to the TATA box, the TAF1 and TAF2 subunits recognize the Inr, the TAF1 subunit is in close proximity to the DCE, and the TAF6 and TAF9 subunits appear to interact with the DPE.

There are also mechanisms of core promoter recognition that do not involve the canonical TFIID complex. For instance, as discussed below, there are TRFs (TBP-related factors) with functions that are distinct from those of TBP.

The other known basal transcription factors participate in the early steps in transcription as follows. TFIIB interacts with TBP and assists in the recruitment of polymerase to the core promoter. TFIIF can bind to core promoter sequences at the BRUs and BREd motifs in a manner that is dependent upon the binding of TBP to the TATA box (for review, see Deng and Roberts, 2007). TFIHIA appears to promote the binding of TBP to the TATA box. TFIIIE, TFIIF, and TFIIH act subsequent to the binding of TFIID and TFIIB to the core promoter and mediate the unwinding of DNA and the early steps in the transcription process.

**Core promoter motifs**

The focused core promoter is diverse in terms of its structure and function. There are several known sequence motifs that can contribute to core promoter activity (Fig. 2), and it is likely that many other core promoter elements remain to be discovered. There are no universal core promoter elements. A brief overview of several core promoter motifs is as follows.

**The initiator (Inr)**

The Inr encompasses the transcription start site and is probably the most commonly occurring core promoter motif (Ohler et al., 2002; FitzGerald et al., 2006; Gershenzon et al., 2006). The function of the Inr as a distinct core promoter element was articulated by Smale and Baltimore (1989). Although several factors have been found to interact with the Inr, the binding of TFIID correlates best with Inr activity (discussed in Smale and Kadonaga, 2003). Functional analyses have determined that the Inr consensus is YYANWYY (IUPAC nucleotide code) in humans and TCAKTY in Drosophila. In rice and Arabidopsis, a YR Inr motif (with R + 1) has been identified (Yamamoto et al., 2007b). Inr-like sequences have also been described in Saccharomyces cerevisiae (Yang et al., 2007).

Computational analyses of Drosophila promoters have suggested an Inr consensus of TCAKTY (Ohler et al., 2002; FitzGerald et al., 2006), which is nearly identical to the Drosophila Inr consensus of TCAKTY determined via functional studies, such as the binding of TFIID.
The TATA box and BRE motifs

The TATA box is the first core promoter motif that was discovered \cite{Goldberg1979} as well as the best known core promoter element. The metazoan TATA box consensus is TATATAAAR, where the upstream T is usually located at −31 or −30 relative to the A+1 (or G+1) position in the Inr \cite{Carninci2006, Fonjavin2006}. As noted above, the TATA box is recognized and bound by the TBP subunit of the TFIIID complex. Both the TATA box and TBP are conserved from archaeabacteria to humans \cite{Reeve2003}. The TATA box is also present in plants \cite{Molina2005, Yamamoto2007}. Although the TATA box is a well known core promoter motif, it is present in only about 10%–15% of mammalian core promoters \cite{Carninci2006, Kim2005, Cooper2006}.

The BRE (TFIIH recognition element) was initially identified as a TFIIH binding sequence that is immediately upstream of a subset (~10%–30%) of TATA box elements \cite{Lagrange1998}. In addition, a second TFIIH recognition site, the BREd (downstream TFIIH recognition element), was found immediately downstream of the TATA box \cite{Deng2005}. The discovery of the BREd led to the renaming of the original BRE as BREu for upstream BRE \cite{Deng2005}. Both the BREu and BREd function in conjunction with a TATA box and have been found to increase as well as to decrease the levels of basal transcription \cite{Lagrange1998, Evans2001, Deng2005}. More recent studies suggest a distinct role for the BREu in transcriptional regulation \cite{Juven-Gershon2008}; discussed below.

DPE and MTE motifs

The DPE (downstream core promoter element) was identified as a TFIIH recognition site that is downstream of the Inr \cite{Burke1997, Kutach2000, Lim2004}. The DPE is located precisely from +28 to +33 relative to the A+1 and is conserved from Drosophila to humans \cite{Burke1997}. The DPE does not appear to be present in S. cerevisiae. The DPE is a recognition site for TFIIH, which binds cooperatively to the Inr and DPE motifs. The spacing between the Inr and DPE is critical for transcriptional activity of DPE-dependent promoters \cite{Kutach2000}. DPE-dependent promoters typically contain only DPE and Inr motifs. In some cases, however, TATA, Inr, and DPE motifs can be found in the same core promoter.

The MTE (motif ten element) was found to be a functionally active core promoter element that corresponds to an overrepresented sequence (termed motif 10) that was identified in Drosophila core promoter regions \cite{Ohler2004, Lim2004}. The MTE is located immediately upstream of the DPE at precisely +18 to +27 relative to the A+1 in the Inr and is conserved from Drosophila to humans. DNase I footprinting analyses suggest that the MTE, like the DPE, is a recognition site for TFIIH. The MTE functions cooperatively with the Inr but can act independently of the DPE as well as the TATA box. There is, however, synergy between the MTE and DPE as well as between the MTE and TATA box.

These studies led to the design of a super core promoter (SCP) that contains a TATA, Inr, MTE, and DPE in a single promoter \cite{Juven-Gershon2006}. The SCP is the strongest core promoter observed in vitro and in cultured cells and yields high levels of transcription in conjunction with transcriptional enhancers. These findings indicate that gene expression levels can be modulated via the core promoter.

Role of the core promoter in the regulation of gene expression

Transcriptional regulation is achieved not only by diversity in enhancers but also by diversity in core promoter structure \cite{Smale2001, Butler2002}. This effect is seen, in particular, in the area of enhancer–promoter communication. For instance, the Drosophila AE1 and IAB enhancers preferentially activate the TATA-containing even-skipped core promoter relative to the TATA-less and DPE-containing white core promoter \cite{Ohatsuki1998}. In addition, DPE- as well as TATA-specific enhancers were identified in an enhancer-trapping screen in Drosophila \cite{Butler2001}. Thus, some activators prefer TATA-dependent promoters, whereas others prefer DPE-dependent promoters.

More recently, the analysis of the Drosophila homeotic (Hox) genes has revealed new insights into the role of the core promoter in a regulatory network \cite{Juven-Gershon2008}. In this study, it was found that nearly all of the Hox genes, which were previously known to have TATA-less promoters, contain DPE-dependent core promoters \cite{Juven-Gershon2008}. This observation suggested that at least some of the transcription factors that regulate the Hox gene network might be DPE-specific activators. Following this hypothesis, it was found that Caudal, a sequence-specific DNA-binding transcription factor and key regulator of the Hox genes, is a DPE-specific activator. In addition, Caudal-mediated activation of the Antennapedia P2 enhancer–promoter region as well as the Sex combs reduced enhancer–promoter region was observed to be dependent upon the DPE motifs in their respective core promoters. These findings collectively indicate an important role of the DPE in the regulation of the Hox genes.

Further investigation showed that the function of Caudal is more complex than a simple matter of specificity for the DPE relative to the TATA box. Specifically, the BREu motif suppresses the ability of Caudal to activate transcription via the TATA box but not the DPE. Hence, as depicted in Fig. 4, there are three levels of Caudal activation – strong
other core promoter motifs. Transcription levels can be further regulated by the presence of the BREu as well as both TATA and DPE motifs can be activated by either DPE- or TATA-specific promoters with the appropriate core promoter elements. The core promoter containing factors. Transcription factors bind to enhancers but only activate transcription from core promoters. The use of DPE- and TATA-specific core promoters is essential to consider and to incorporate these factors in the analysis of differentiation of other cell types.

Conclusions and perspectives

The core promoter and the basal transcriptional machinery are two important yet relatively unexplored dimensions in the regulation of gene expression. It is now apparent that diversity in the structure and function of core promoters and basal transcription factors contributes to developmental processes that lead to organismal complexity (Levine and Tjian, 2003). Thus, in the future, it will be important to consider why Caudal might act as a DPE-specific activator. In a simple sense, it could be imagined that DPE specificity would be useful in the construction of regulatory networks. As in the wiring of a printed circuit board, there could be connections between transcriptional enhancers and their cognate core promoters. The use of DPE- and TATA-specific activators would enable the construction of more sophisticated and effective connections between enhancers and promoters (Fig. 5).

**TBP-related factors (TRFs) and transciptonal regulation**

There is diversity not only in core promoter elements but also in the basal transcription machinery. This concept is nicely exemplified in studies of the TBP-related factors (TRFs) (for reviews, see: Jones, 2007; Müller et al., 2007; Reina and Hernandez, 2007; Torres-Padilla and Tora, 2007). There are three TRFs, which are generally termed TRF1, TRF2, and TRF3.

TRF1 does not exist in yeast and humans but is present in *Drosophila*. In many eukaryotes, including yeast and humans, TBP participates in transcription by RNA polymerases I, II, and III. However, in *Drosophila*, TRF1 is used instead of TBP for RNA polymerase III transcription (Takada et al., 2000).

TRF2 (also known as TLF, TLP, TRF, and TRP) is present in most eukaryotes and is involved in transcription by RNA polymerase II. TRF2 does not bind to TATA box sequences and cannot replace TBP in vitro. It appears that many genes are regulated by TRF2 instead of TBP—such as the *Drosophila* histone H1 gene (Isogai et al., 2007). The TATA-less H1 linker histone gene is in a cluster of genes that also includes the four TATA-containing core histone genes, which are transcribed with TBP. These findings suggest the use of different transcriptional mechanisms within a cluster of genes.

TRF3 (also known as TPB2 and TBP2) appears to be present only in vertebrates and is the TFIID complex that is most closely related to TBP. TRF3 can bind to TATA boxes and support TATA-dependent transcription (Bártfai et al., 2004; Jallow et al., 2004). TRF3 was found to be important for embryonic development (Bártfai et al., 2004; Jallow et al., 2004). In addition, zebrafish embryos that are depleted of TRF3 exhibit multiple developmental defects and fail to undergo hematopoiesis (Hart et al., 2007).

A particularly striking function of TRF3 was discovered during the analysis of the differentiation of myoblasts into myotubes (Deato and Tjian, 2007; Deato et al., 2008). Myoblasts were found to contain the canonical TFIID complex; however, upon terminal differentiation into myotubes, the TFIID complex was replaced by a TRF3–TAF3-containing complex (Fig. 6). These findings suggest that externally differentiated cells may employ specialized transcription systems that are dedicated to the particular functions of the cells. It will be interesting to see if an analogous effect is observed in the differentiation of other cell types.

**Fig. 4.** Caudal is a DPE-specific activator. Caudal preferentially activates transcription from DPE-dependent core promoters relative to TATA-dependent core promoters. In addition, the presence of a BREu motif upstream of the TATA box further suppresses the ability of Caudal to activate transcription. The BREu motif does not affect the ability of Caudal to activate a DPE-dependent core promoter. The TATA box also does not alter the ability of Caudal to activate transcription via the DPE motif (not shown). Results taken from Juven-Gershon et al. (2008a).

**Fig. 5.** A simplified, hypothetical diagram of activation by DPE- and TATA-specific activators. Transcription factors bind to enhancers but only activate transcription from promoters with the appropriate core promoter elements. The core promoter containing both TATA and DPE motifs can be activated by either DPE- or TATA-specific activators. Transcription levels can be further regulated by the presence of the BREu as well as other core promoter motifs.

**Fig. 6.** Replacement of the canonical TFIID complex by a TRF3–TAF3-containing complex upon terminal differentiation of myoblasts into myotubes. Both complexes bind to TATA box motifs via the TBP or TRF3 subunits. These findings exemplify the establishment of a new basal transcription system upon cell differentiation and suggest that analogous processes may occur in other cell types. Results taken from Deato and Tjian (2007).
enhancers (see, for example, Pfeiffer et al., 2008). The increased appreciation and understanding of core promoter motifs and basal transcription factors will lead to new and exciting discoveries and, ultimately, provide a more complete and accurate view of biological regulation.

Acknowledgments

We are grateful to Joshua Theisen, Timur Yusufzai, Jer-Yuan Hsu, Barbara Rattner, Moriah Eustice, Sharon Torigoe, and Trevor Parry for critical reading of the manuscript. Our research on the RNA polymerase II core promoter is supported by a grant from the National Institutes of Health (GM041249).

References


Deng, W., Roberts, S.G., 2005. A core promoter element downstream of the TATA box is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila melanogaster. Genes Dev. 11, 3020–3031.


