Review

CellPress

Core promoters in transcription: old problem, new insights

Ananda L. Roy[1] and Dinah S. Singer[2]

[1] Department of Developmental, Molecular, and Chemical Biology, Tufts University School of Medicine, Boston, MA, USA
[2] Experimental Immunology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

Early studies established that transcription initiates within an approximately 50 bp DNA segment capable of nucleating the assembly of RNA polymerase II (Pol II) and associated general transcription factors (GTFs) necessary for transcriptional initiation; this region is called a core promoter. Subsequent analyses identified a series of conserved DNA sequence elements, present in various combinations or not at all, in core promoters. Recent genome-wide analyses have provided further insights into the complexity of core promoter architecture and function. Here we review recent studies that delineate the active role of core promoters in the transcriptional regulation of diverse physiological systems.

The core promoter: a platform for transcription initiation

Cellular differentiation and function depend on the accurate and regulated transcription of the genome, which encompasses both the 2–3% of the genome that encodes proteins and the approximately 90% that is transcribed into noncoding RNA including ribosomal and tRNAs, long noncoding RNAs, miRNAs, and other regulatory RNAs [1,2]. Integral to this regulated expression is the ability to initiate transcription at precise genomic sites.

Core promoters are defined as the DNA segment of 50–100 bp within which transcription initiates [3]. Genome-wide structural analyses have identified a series of conserved DNA sequence elements that are often, but not universally, associated with core promoters (Figure 1). The core promoter functions as a platform on which the transcription machinery assembles. Among the factors recruited to core promoters are the enzyme RNA Pol II, which transcribes protein-coding and many non-protein-coding RNAs (e.g., long noncoding RNAs, miRNAs), and the multiple GTFs and cofactors required for RNA synthesis and biogenesis. It is estimated that this transcriptional complex is well over a megadalton in size and can occupy over a hundred base pairs of DNA around the transcriptional start site (TSS) [3].

The assembly of the transcription machinery at the core promoter is initiated by the interaction of specific transcription factors with their cognate upstream regulatory sequences (e.g., enhancers, silencers). These specific transcription factors mediate both cell-intrinsic, tissue-specific signals and extrinsic signals that modulate the transcription of a given gene. However, all of these signaling events need to be integrated at the core promoter to achieve the appropriate level of transcription initiation. Although considerable effort has been expended to identify the upstream regulatory sequences that contribute to proper regulation of transcription, much less attention has been focused on characterizing the structure and function of core promoters. In this review, we summarize the evidence that the core promoter functions both as a platform to integrate upstream signaling and as an active participant in regulating RNA Pol II-mediated transcription. Moreover, given that genome-wide studies indicate that many genes lack an identifiable core promoter element, we also discuss the implications of these recent findings.

Structure of canonical core promoter elements

Many core promoter sequences are common to a large number of protein-coding genes and have been conserved in evolution. Among these canonical core promoter elements are the TATAA box, the Initiator (Inr), the TFIIB recognition element (BRE), downstream promoter element (DPE), and downstream core element (DCE). Additional elements – the motif ten element (MTE) and XCP1 – have been described but occur at lower frequencies (Box 1). Although not formally considered a canonical core promoter element, the CCAAT box, located between −50 bp and −80/−100 bp upstream of the TSS has been conserved in evolution from bacteria to metazoans. However, the precise role of canonical core promoter elements in regulating the transcription of all genes is unclear as not all core promoter elements are found associated with all genes. In the following sections, we briefly describe the salient features of some of these elements.

Transcription initiation complex nucleators: the TATA box and Inr

The two most common core promoter elements associated with protein-coding genes are the TATAA box and the Inr, which occur either together or separately in most eukaryotic promoters. The TATA box (TATAA), the first core promoter element to be identified, was biochemically found to be located 20–30 bp upstream of the TSS and serves as the binding site for the GTF TFIID [4]. However, recent studies show that the optimal spacing between the first T

Corresponding authors: Roy, A.L. (ananda.roy@tufts.edu); Singer, D.S. (dinah.singer@mail.nih.gov).
Keywords: transcription; core promoter; RNA Pol II; noncoding RNAs.

0968-0004/
© 2015 Elsevier Ltd. All rights reserved. http://dx.doi.org/10.1016/j.tibs.2015.01.007

CrossMark

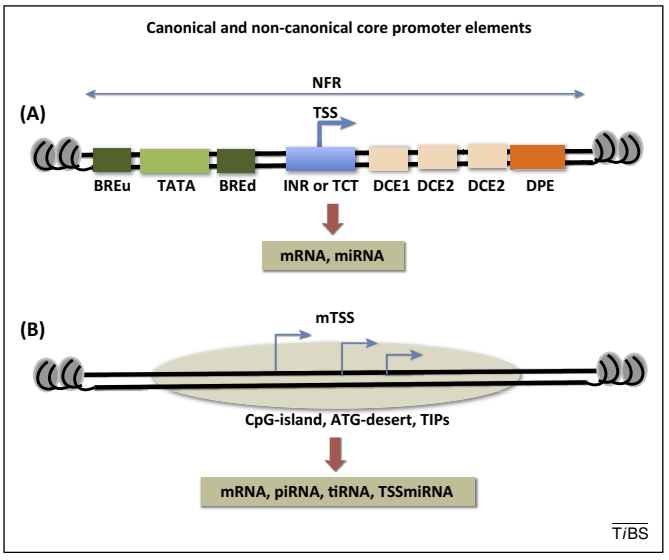

Figure 1. Canonical **(A)** and noncanonical **(B)** core promoter elements. The various established core promoter elements are shown in (A). Although most mammalian genes do not appear to have any of these elements, some protein-coding and miRNA promoters have canonical core promoter elements associated with them. Not all of these elements are present in all promoters and many are present in lineage-specific genes. While the Initiator (Inr) element is utilized by TFIID, the TCT element is utilized by the TATA-binding protein (TBP)-related factor TRF2 in ribosomal protein-coding genes. TSS, transcription start site; NFR, nucleosome-free region. (B) A large majority of mammalian genes appear to lack classical core promoter elements but instead contain broad regions (100–150 bp) associated with transcription initiation. The established noncanonical elements or regions are CpG islands, ATG deserts, and transcription initiation platforms (TIPs). These are present in mRNA-coding as well as various noncoding RNAs, including piwi-interacting RNA (piRNA), transcription initiation-associated RNA (tiRNA), and TSS-associated miRNA (TSSmiRNA). These promoters are frequently characterized by multiple TSSs (mTSSs).

of the TATA box and the +1 of the TSS is 30–31 bp [5]. The TATA box is found in only about 5–7% of eukaryotic promoters. Among the promoters that lack a TATA box, many contain the Inr, although the Inr element is found as often in TATA-containing promoters [6,7].

The Inr (consensus sequence YYA^{+1}NT/AYY) spans the TSS and can independently direct accurate transcription initiation [8–11]. Nearly 70% of *Drosophila* promoters appear to have an Inr element, either alone or in combination with a TATA box; no clear estimates are available for the distribution of the Inr element in mammalian promoters [6].

Although the TATA and Inr can be found together in some core promoters [7], they appear to serve distinct functional families of genes. In general, the core promoters of tissue-specific genes are anchored by a TATA box and generally initiate transcription at a single discrete site or at a tightly clustered locus [12]. By contrast, Inr elements are found in core promoters of ubiquitously expressed or 'housekeeping' genes and initiate transcription over multiple, dispersed sites across a region of up to approximately 150 bp [12–14]. Thus, the promoters of immunoglobulin and globin genes contain only the TATAA box whereas actin and terminal deoxynucleotidyl transferase (TdT) promoters contain only Inr elements. Genes subject to complex regulatory patterns, such as the MHC class I genes, often contain both Inr and TATAA elements [15].

Both the TATA box and the Inr provide a platform for the assembly of transcription pre-initiation complexes (PICs) by the GTF TFIID. The TATA element serves as the binding site for the TFIID component TATA-binding protein (TBP) to nucleate the formation of a transcription initiation complex and the subsequent recruitment of RNA Pol II [3]. Similarly, TFIID nucleates PIC assembly on Inr promoters, although the precise nature of the interaction is unknown.

DPE

Various additional core promoter elements that modulate the activities of the TATAA and Inr elements have been identified. Among these is the DPE found in both *Drosophila* and mammalian Inr promoters [6]. The DPE (consensus sequence A/GGA/TCGTG), which occurs mostly but not exclusively in TATA-less promoters, acts in conjunction with the Inr and is located at +28 to +32 relative to the initiating nucleotide at +1 within the Inr motif [6]. Although the DPE, like the TATA box, is a recognition site for the binding of components of the GTF TFIID, it does not act independently but depends on the presence of an Inr.

BRE

The requirement for core promoter element sequences upstream of the TATA box was first observed in archaeal genes through mutational analysis. Structural analysis of TBP–TFIIB DNA as well as functional studies identified a 7-bp sequence element named the BRE). These motifs can be present on either side of the TATA box, with the upstream BRE (BREu) at −38 to −32 (consensus sequence G/CG/CG/ACGCC) and the downstream BRE (BREd) at −23 to −17 (consensus sequence G/ATT/AT/GT/GT/GT/G). Interestingly, the BRE not only functions in activating transcription but also can repress transcription [6].

DCE

Another core promoter element is the DCE, which has canonical sequences at positions +6 to +11, +16 to +21, and +30 to +34 relative to the TSS (consensus sequence

Box 1. Canonical and noncanonical core prompter elements

Canonical core promoter elements

TATA box: ∼ −25 relative to the TSS; consensus sequence TATAAA or TATATA.

Inr: spans the TSS; consensus sequence YYANT/AYY.

DPE: ∼+30 relative to the TSS; consensus sequence A/GGA/TCGTG.

BREu: ∼−35 relative to the TSS; consensus sequence G/CG/CG/ACGCC.

BREd: ∼−20 relative to the TSS; consensus sequence G/ATT/AT/GT/GT/GT/G.

DCE: ∼+9, +18, +32 relative to the TSS; consensus sequence CTTC, CTGT, AGC.

MTE: ∼+23 relative to the TSS; consensus sequence CG/CAA/GCG/CAACG.

Noncanonical promoter elements

CpG island: spans the TSS; 0.5–2 kb DNA with high density of CpG dinucleotides.

ATG desert: spans ∼1 kb both upstream and downstream of the TSS; a segment of DNA with lower frequency of ATG trinucleotides than the surrounding sequences.

Transcription initiation platform (TIP): overlaps the TSS; highly variable 0.4–10 kb with high CpG content.

Download English Version:

https://daneshyari.com/en/article/2030970

Download Persian Version:

https://daneshyari.com/article/2030970

Daneshyari.com