Eukaryotic Gene Expression: Basics & Benefits

P N RANGARAJAN

Lecture 2

Gene Regulation in Eukaryotes: Diversity in core promoter elements

Recap……………….

In eukaryotes, there are three RNA polymerases (I, II and III).

Eukaryotic RNA Polymerase II is a multi subunit complex .

Unlike prokaryotic RNA polymerase sigma subunit, none of the eukaryotic RNA polymerase II subunits are involved in promoter recognition and DNA unwinding.

RNA polymerase II cannot recognize its target promoters directly and it relies on a host of **general transcription factors** to perform this function

RNA Polymerase IV

In addition to the three eukaryotic RNA polymerases (I, II and III, a fourth RNA polymerase, was recently identified in plants which facilitates the production of small interfering RNA (siRNA) involved in RNA-directed DNA methylation, transcriptional silencing and formation of Heterochromatin.

Although this plant specific RNA polymerase IV appears to be nonessential for viability, it exhibits unique properties not shared by other nuclear RNA polymerases.

Herr, A.J., et al., 2005. RNA polymerase IV directs silencing of endogenous DNA. *Science 308:118–* 120.

Kanno, T., et al., 2005. Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nature Genetics 37:761–765.*

Onodera, Y. et al., 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* 120:613–622.

The RNA Polymerase II is associated with six general transcription factors known as TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, where "TF" stands for "transcription factor" and "II" for the RNA Polymerase II.

Purification and biochemical characterization of the general transcription factors as well as analysis of their function in cell-free transcription assay systems indicated that the RNA polymerase II and the general transcription factors are assembled in a sequential manner:

First, a complex is formed amongst TFIID, TFIIA, and TFIIB (DAB) which recognizes the TATA sequence in the promoter and binds to it .

A stable closed complex containing DAB, hypophoshorylated RNA Pol II and TFIIF is then formed. TFIIF prevents RNA polymerase II from binding non-specifically to DNA.

TFIIE and TFIIH finally join leading to DNA unwinding, promoter clearance and nascent RNA synthesis and In the presence of nucleotides, initiation of transcription begins with the phosphorylation of the
C-terminal domain (CTD) of RNA polymerase II.

Thus the function of sigma factor in prokaryotes is carried
out by the general transcription factors in eukaryotes.

In this class……….

All protein coding genes need not necessarily contain the same core promoter sequences (TATA sequence for example).

Just as variants of -35 and -10 promoter sequences can contribute to
differential gene regulation in prokaryotes by virtue of binding to different sigma factors, variations in core promoter sequences and general transcription factors can contribute to differential gene regulation in eukaryotes as well.

Eukaryotic core promoter sequences and proteins binding to it are not monolithic, invariant, and conserved in all promoters of an organism.

THERE ARE NO UNIVERSAL CORE PROMOTER ELEMENTS IN AN ORGANISM

What is a core promoter?

Core promoter is the minimal stretch of contiguous DNA sequence that is sufficient to direct accurate initiation of transcription by the RNA polymerase II and associated factors.

Usually, the core promoter covers a promoter region upto 35 bp upstream and downstream of the transcription initiation site.

In addition to the core promoter, other *cis*-acting DNA sequences that regulate RNA polymerase II transcription include:

Proximal promoter (encompassing −250 to +250 nt from transcription start site)

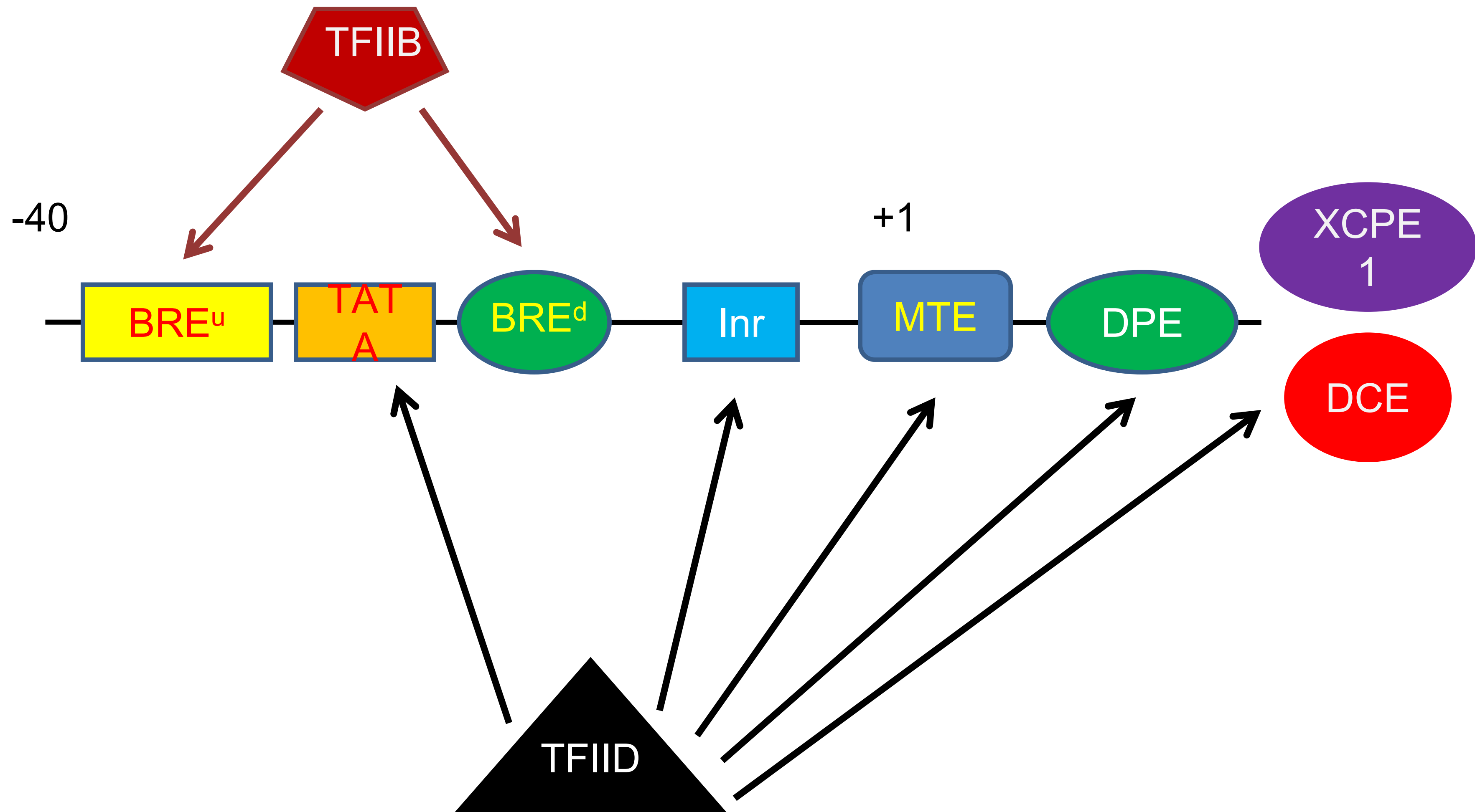
Enhancers

Silencers

Boundary/insulator elements

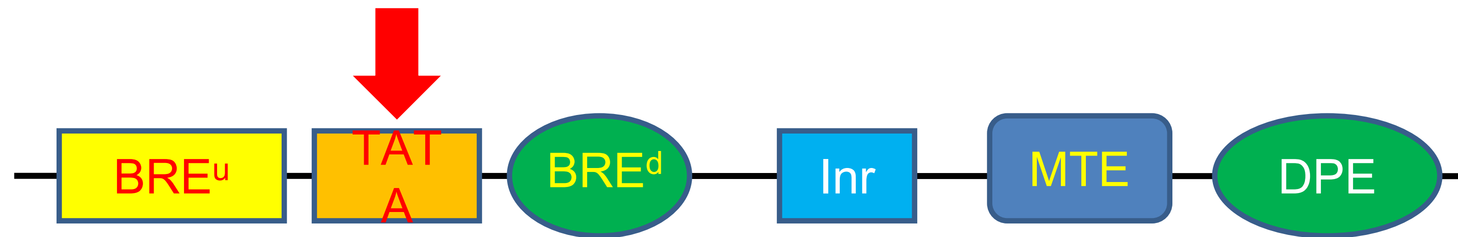
These elements contain binding sites for several transcription factors which are located several kilobases upstream of transcription start site

CORE PROMOTER SEQUENCES OF A PROTEIN CODING GENE



Each of these core promoter elements is present in only a subset of core promoters while some promoters lack all known core promoter elements.

TATA BOX



The TATA box was the first eukaryotic core promoter element
to be identified in the year 1979 by Michael Goldberg at Stanford University.

TATA box consensus sequence: TATAWAAR

The TATA box is typically located about 25–30 nt upstream of
the transcription start site (+1) and the first nucleotide of
the TATA sequence above (T) is usually present at −31 or −30
position relative to +1.

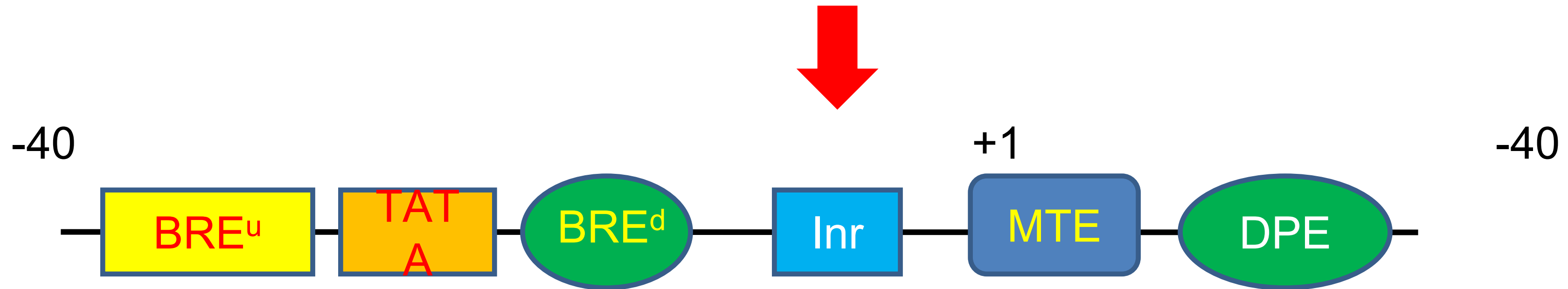
TATA box bears resemblance to the −10 region (Pribnow box)
of prokaryotic promoters but the eukaryotic TATA box and
prokaryotic −10 region are not homologous.

The consensus sequence for the TATA box is TATAAA. However, several variants of this sequence can function as a TATA box *in vivo* and those with one or two mismatches from this consensus sequence can still function as TATA box.

As noted above, TATA boxes are not present in all the promoters transcribed by RNA polymerase II. It was estimated that approximately 43% of 205 core promoters in *Drosophila* contain a TATA box while in humans, 32% of 1031 potential promoter regions contain a putative TATA box motif .

During transcription initiation, TATA box serves as the binding site for the general transcription factor TATA binding protein (TBP) which is a component of the multi protein complex, TFIID.

INITIATOR (Inr)



Inr was first identified by Smaleb and Baltimore in 1989 in the promoter of lymphocyte-specific terminal deoxynucleotidyltransferase gene which lacks TATA box.

Inr is a 17 bp sequence that is sufficient for accurate basal transcription of this gene both *in vitro* and *in vivo* and encompasses the transcription start site.

The discovery of Inr explained how transcription is initiated from the promoters of protein coding genes that lack TATA box.

Inr Consensus sequence:

YY**A**NWYY Humans
TC**A**KTY *Drosophila*

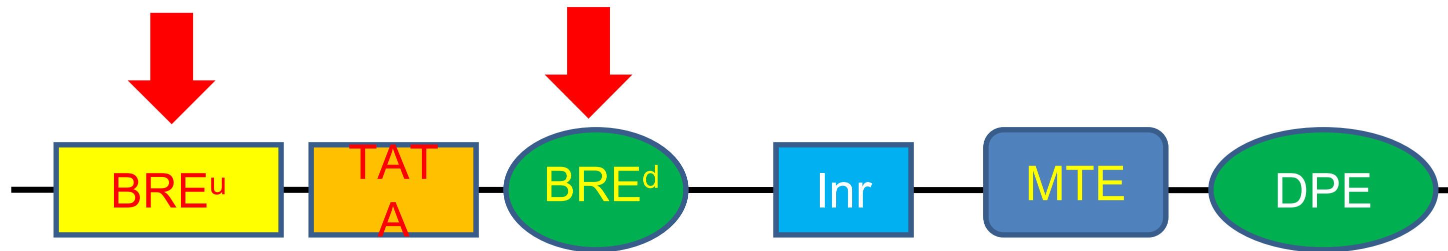
A IS THE TRANSCRIPTION START SITE

Computational analysis of several mammalian transcription start sites revealed that the mammalian Inr consensus is YR in which R corresponds to the +1 start site.

Inr binds to a number of proteins of which TFIIID binding is the most important

IUPAC nucleotide code	Base
A	**Adenine**
C	Cytosine
G	Guanine
T (or U)	Thymine (or Uracil)
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base

BRE (TFIIB RECOGNITION ELEMENT)



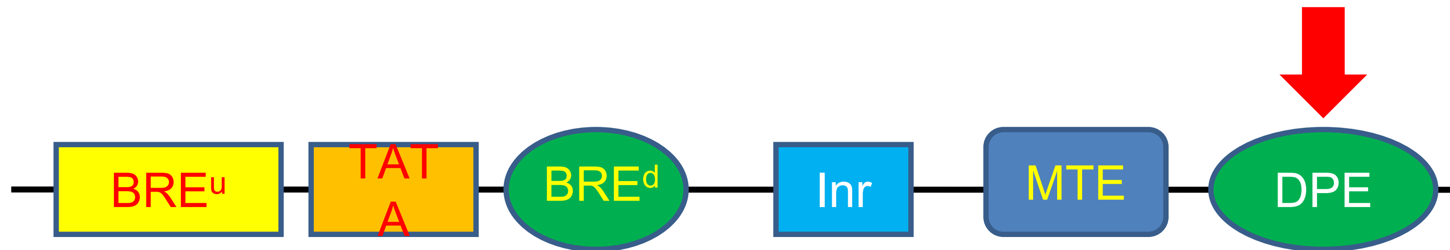
The BRE was originally identified by Lagrange et al., in the year 1998 as a TFIIB binding element . BRE may be present either upstream (BRE^u) or downstream (BRE^d) of certain TATA sequences.

BRE^u consensus sequence: 5' G/C-G/C-G/A-C-G-C-C 3'
BRE^d consensus sequence: 5' G/A-T-T/G/A-T/G-G/T-T/G-T/G 3'

Depending on the promoter context, the BREs can have a positive or negative effect on transcription.

Interestingly, BRE is present more frequently in promoters which do not contain a TATA box than TATA-containing promoters (28.1% of TATA-less promoters contain BRE^u compared to 11.8% of TATA containing promoters have BRE^u).

DPE (DOWNSTREAM CORE PROMOTER ELEMENT)

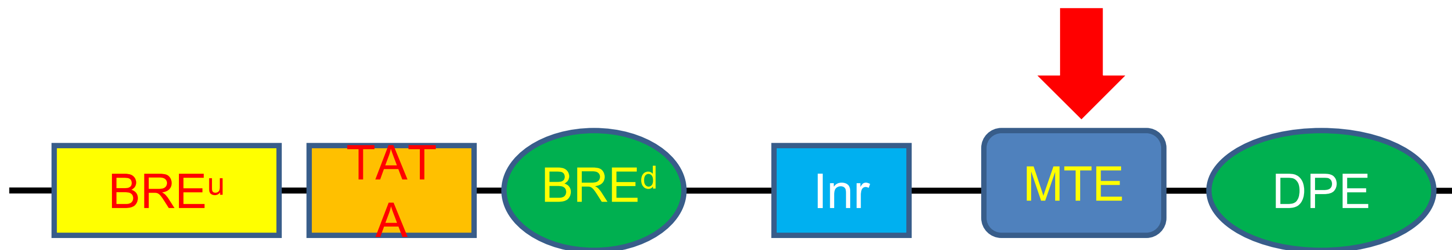


DPE was first identified as a downstream TFIID recognition sequence that has a role in basal transcription activity by Burke and Kadonaga in 1996.

In Drosophila, the DPE consensus sequence for DPE is RGWYVT (where R is A/G; W is A or T; Y is C/T; V is A/C/G)

The DPE functions cooperatively with the Inr, and the spacing between the Inr and DPE is very important for optimal transcription.

MTE (MOTIF TEN ELEMENT)



In Drosophila, the MTE is present from +18 to +27 relative to transcription start site (+1) .

The consensus sequence for MTE is: CSARCSSAAC (where S is G/C; R is A/G).

As observed in case of DPE, MTE also functions cooperatively with the Inr and the spacing between Inr-MTE is crucial for its function.

MTE synergistically activates transcription together with the TATA and DPE motifs. Based on this observation, a Super Core Promoter (SCP) containing optimized versions of the TATA box, Inr, MTE, and DPE was designed by Juven-Gershon et al., in the year 2006. The SCP is one of the strongest known core promoters which exhibits high affinity for the binding of TFIID.

DCE (DOWNSTREAM CORE ELEMENT)

The DCE (downstream core element) was first identified in the human beta-globin promoter by Lewis et al., in the year 2000 and subsequently in the adenovirus major late promoter.

The DCE consists of three sub elements:

S_I, CTTC from +6 to +11
S_{II}, CTGT from +16 to +21
S_{III}, AGC from +30 to +34.

XCPE1 (X CORE PROMOTER ELEMENT)

XCPE1, which is present from −8 to +2 relative to the +1 transcription start site functions as a positive cis-acting element in presence of certain sequence-specific transcription factors that bind to far upstream promoter sequences in the promoter.

About 1% of human core promoters contain an XCPE1 motif and is often present in TATA-less promoters.

XCPE1 consensus sequence is DSGYGGRASM (where D is A/G/T; S is G/C; Y is T/C; R is G/A; M is A/C)

Frequency of core promoter elements in eukaryotic promoters

Core element	Frequency in promoters	
	vertebrates	flies
TATA	10-16%	33-43%
Inr	55%	69%
BRE	12-62%	----
DPE	48%	40%
MTE	----	8.5%

Heintzman and Ren 2007, Cell. Mol. Life Sci. 386-400

Transcriptional regulation via core promoter sequences

The core promoter is the ultimate target of the action of sequence-specific transcription factors that bind to specific upstream promoter elements and many such transcription factors function only in conjunction with specific core promoter sequences.

Thus, core promoters are not only involved in proper positioning of the RNA polymerase II transcriptional machinery, but also function as important *cis-acting regulatory elements* thereby providing another level of transcriptional regulation.

Certain enhancer sequences are functionally coupled to specific core promoter sequences. Certain transcription factors which bind to enhancer sequences can activate transcription only when the promoter contains specific core promoter sequences. For example, the transcription factor Sp1 in conjunction with the general transcription factors can enhance transcription from a TATA-containing core promoter but not from a DPE-containing core promoter.

Examples of transcriptional regulation via core promoter sequences

Differential usage of two TATA sequences in the *his3* gene promoter of *Saccharomyces cerevisiae*

The S. cerevisiae *his3* gene promoter contain two TATA boxes known as T_C and $T_{R.}$ The downstream T_R (regulatory TATA) has a canonical TATAAA sequence, whereas the upstream T_C (constitutive TATA) is an AT-rich region of ~30 nt that lacks a canonical TATA sequence.

T_C is used when transcription levels are low while T_R is used when *his3* gene is to be transcribed at high levels*.

*** For mechanistic insights, refer: Mahadevan, S. and Struhl, K. 1990 &**
Iyer, V. and Struhl, K. 1995.

Certain upstream activators function in conjunction with specific core promoter elements

Simon et al., (1988) observed that in case of the human hsp70 promoter, it was observed that the transcription factor E1A can activate transcription only when hsp7- TATA box (TATAA) is present but not when it is substituted by the SV40 early promoter TATA box (TATTTAT).

Thus, the specific TATA box sequence of the *hsp70 promoter* is important for its transcription activation by E1A transcription factor.

In yet another instance, it was observed that the human myoglobin enhancer can activate transcription from myoglobin promoter but not the SV40 promoter.

However, when the SV40 early promoter TATA sequence (TATTTAT) was replaced by 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 but not the TATTTAT sequence.

In case of Drosophila, enhancers known as AE1 and IAB5 preferentially activate transcription from TATA containing promoters rather than TATA-less promoters. Thus, the enhancer function is dependent on the presence of a TATA sequence in the core promoter

A transcription factor known as Gal4-VP16 has a strong preference for TATA-containing promoters, whereas another transactivator, Gal4-Sp1 preferentially activates transcription from Inr-containing promoters.

Thus, specific transcription factors function efficiently in conjunction with specific core promoter elements

SUMMARY

THERE ARE MANY DIFFERENT CORE PROMOTER ELEMENTS IN EUKARYOTES

THERE IS NOTHING LIKE UNIVERSAL CORE PROMOTER ELEMENTS IN EUKARYOTES

CORE PROMOTERS ARE ESSENTIAL NOT ONLY FOR THE ASSEMBLY OF RNA POLYMERASE MACHINERY AND TRANSCRIPTION INITIATION BUT ALSO FOR THE MODULATION OF THE FUNCTION OF OTHER UPSTREAM CIS-ACTING ELEMENTS (ENHANCERS) AND TRANSCRIPTION FACTORS (WHICH BIND TO ENHANCER SEQUENCES).

ORIGINAL RESEARCH ARTICLES

Goldberg M.L. *(1979) "Sequence analysis of Drosophila histone genes." Ph.D. dissertation (Stanford University, California).*

Stephen T. Smaleb, and David Baltimore. The "initiator" as a transcription control element (1989) Cell 57:103-113

T. Lagrange, A.N. Kapanidis, H. Tang, D. Reinberg and R.H. Ebright, New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev* **12** (1998), pp. 34–44.

W. Deng and S.G. Roberts, A core promoter element downstream of the TATA box that is recognized by TFIIB, *Genes Dev* **19** (2005), pp. 2418–2423.

W. Deng and S.G. Roberts, A core promoter element downstream of the TATA box that is recognized by TFIIB, *Genes Dev* **19** (2005), pp. 2418–2423.

W. Deng and S.G. Roberts, TFIIB and the regulation of transcription by RNA polymerase II, *Chromosoma* **116** (2007), pp. 417–429.

W. Burke and J.T. Kadonaga, Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters, *Genes Dev* **10** (1996), pp. 711–724.

C.Y. Lim, B. Santoso, T. Boulay, E. Dong, U. Ohler and J.T. Kadonaga, The MTE, a new core promoter element for transcription by RNA polymerase II, *Genes Dev* **18** (2004), pp. 1606–1617.

T. Juven-Gershon, S. Cheng and J.T. Kadonaga, Rational design of a super core promoter that enhances gene expression, *Nat Methods* **3** (2006), pp. 917–922.

ORIGINAL RESEARCH ARTICLES

B.A. Lewis, T.K. Kim and S.H. Orkin, A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts, *Proc Natl Acad Sci U S A* **97** (2000), pp. 7172–7177.

D.H. Lee, N. Gershenzon, M. Gupta, I.P. Ioshikhes, D. Reinberg and B.A. Lewis, Functional characterization of core promoter elements: the downstream core element is recognized by TAF1, *Mol Cell Biol* **25** (2005), pp. 9674–9686.

Y. Tokusumi, Y. Ma, X. Song, R.H. Jacobson and S. Takada, The new core promoter element XCPE1 (X Core Promoter Element 1) directs activator-, mediator-, and TATA-binding protein-dependent but TFIID-independent RNA polymerase II transcription from TATA-less promoters, *Mol Cell Biol* **27** (2007), pp. 1844–1858.

Mahadevan, S. and Struhl, K. 1990. TC, an unusual promoter element required for constitutive transcription of the yeast *HIS3 gene. Mol. Cell. Biol.* ***10: 4447–4455.***

Iyer, V. and Struhl, K. 1995. Mechanism of differential utilization of the *his3 TR and TC TATA elements. Mol. Cell. Biol.* **15: 7059–7066.**

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.***

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.

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.***