Regulation of the initiation of eukaryotic transcription

Grace Gill

Department of Pathology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, U.S.A.

Introduction

Transcription is the fundamental process whereby the genetic information encoded in the DNA is first expressed. During development, expression of specific genes is turned on or off at precise times and in particular cells, giving rise to the diversity and specificity of cell function. Transcriptional regulation also enables cells to respond to environmental cues such as the availability of nutrients or viral infection. Thus failure to properly regulate transcription can lead to severe developmental abnormalities or disease. Regulation of gene expression is determined in large part by the activity of transcriptional activator proteins that bind specific DNA sequences near the gene. The first step in transcribing a protein-coding gene, i.e. the binding of the RNA polymerase II machinery to the promoter, is subject to extensive regulation. In general, wrapping of the DNA around nucleosomes and packaging into higher-order chromatin inhibits binding by the transcription machinery. Therefore chromatin-modifying complexes, which can either enhance or relieve chromatin-mediated repression, are critical to regulated transcription (see Chapter 4 in this volume by A.P. Wolffe). This chapter will focus on the factors that influence the efficiency of binding by the transcription machinery to DNA that has already been made accessible, namely the DNA sequence of the core promoter and interactions with transcriptional activators and co-activators. Synergistic activation by several transcription factors working

\textsuperscript{1}e-mail: grace_gill@hms.harvard.edu
together provides additional specificity and potency to gene regulation. As described below, studies on the functional interactions of the RNA polymerase II transcription machinery with promoters and activators have provided new insights into the mechanisms that regulate the initiation of eukaryotic transcription.

**A typical eukaryotic promoter**

Several functionally distinct types of DNA elements that can contribute to gene regulation in eukaryotes have been identified. These include the core promoter, regulatory promoter and enhancer sequences. Additional sequences may also contribute to regulation of gene expression in chromatin. The core promoter serves as a binding site for the RNA polymerase II transcription machinery and determines both the start site and the direction of transcription. The core promoter is sufficient to support correctly initiated transcription *in vitro*. Common core promoter elements include the TATA box, which is generally located 25–30 bp upstream of the start site, the initiator (Inr), which flanks the start site, and the downstream promoter element (DPE) found approximately 30 bp downstream of the start site (Figure 1) [1,2]. Core promoters can include a TATA box alone, an Inr element alone or both together. The DPE is found in combination with an Inr at some promoters lacking TATA boxes. The sequence of the core promoter not only determines the basal, or unregulated, level of transcription, but may also contribute to regulation of the gene [3,4].

Although the core promoter is sufficient to support correctly initiated transcription of a gene *in vitro*, expression of most genes *in vivo* is dependent on additional regulatory sequences. Important regulatory sequences located within

![Figure 1. Typical promoter elements that regulate transcription of a gene in higher eukaryotes](image)

The consensus sequences of the common core promoter elements and their positions relative to the start site of transcription are shown. Both the regulatory promoter and the enhancer are composed of binding sites for multiple transcription factors. The actual sequence, size and position of these elements varies greatly among different promoters. Py denotes a pyrimidine nucleotide.
a few hundred base pairs 5′ of the core promoter are often referred to as the regulatory promoter. Enhancers, which are functionally defined as DNA elements that increase transcription of a gene in a position- and orientation-independent manner, can be located at great distances either upstream or downstream of the core promoter. Binding sites for many transcription factors are found in both regulatory promoters and enhancers, and the distinction between these transcriptional control elements is sometimes blurred. Detailed analysis of enhancers such as the interferon-β enhancer has revealed that these DNA sequences serve as a template for the assembly of a precisely configured DNA–protein complex termed the enhanceosome [5]. As described below, cooperative protein–protein interactions between many activators bound to the enhancer contribute to both the specificity and potency of enhancer-mediated transcriptional activation.

The RNA polymerase II machinery

RNA polymerase II requires a number of additional factors to specifically initiate transcription of a gene, dependent on the promoter sequences. One of the great advances in transcription research in recent years has been the identification and cloning of these general transcription factors (GTFs). The GTFs include TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH [6]. TFIID is the only GTF that binds specifically and independently to the core promoter. Both the TATA-box-binding protein (TBP) and TBP-associated factors (TAF$_{IHS}$) in the multiprotein TFIID complex contribute to binding the core promoter. TBP binds the TATA box, TAF$_{II250}$ and TAF$_{II150}$ together bind to the Inr element, and TAF$_{II60}$ contributes to binding to the DPE [2,6,7]. Interaction of TFIIB with DNA just 5′ of the TATA box can also influence binding of the transcription machinery to the core promoter [8]. Using highly purified components, researchers have defined an order of assembly of the GTFs on a promoter. TFIID, together with TFIIA, binds first to the promoter, followed by TFIIB, an RNA polymerase II–TFIIF complex, then TFIIE and lastly TFIIH [6,9]. Recent studies have suggested, however, that many of these factors bind to each other off the DNA, leading to a simplified two-step reaction in which TFIID binds together with TFIIA, followed by binding of a large pre-assembled complex including RNA polymerase II and the remaining GTFs (Figure 2) [6]. This form of the polymerase, termed the holoenzyme, is also associated with accessory factors, such as SRB (suppressor of RNA polymerase B) and mediator proteins in yeast and p300/CBP [cAMP-response-element-binding protein (CREB)-binding protein] in humans. Although these factors are not required for basal transcription in vitro, they have been found to contribute to regulated gene expression [10]. These studies have revealed a stunning complexity in the eukaryotic transcription machinery, with over 50 proteins assembled at the core promoter.
Biochemical studies using model templates such as the adenovirus major late promoter led to the identification of the GTFs required to support specific initiation of transcription by RNA polymerase II [6]. As more promoters are analysed, however, it has become clear that the same set of GTFs may not function at all promoters. The TAFIIs in the TFIID complex, for example, are required for transcription from a core promoter lacking a TATA box in vitro, although TBP alone is sufficient to support transcription from a promoter with a consensus TATA box [11]. Studies in yeast confirm that the requirement of TAFIIs for transcription in vivo is dependent upon the sequence of the core promoter [12]. Similarly, TFIIE has been found to be dispensable for transcription of certain promoters in vitro and in vivo [13,14]. In fact, recent studies have revealed a surprising diversity of GTFs, including two TBP-related factors, at least one tissue-specific TAFII and a testis-specific TFIIA isoform, further supporting the notion that distinct sets of factors may assemble at specific promoters [15–18]. Although the mechanisms that determine which basal transcription factors are required at a given promoter are not well under-
stood, the exact composition of the transcription machinery that assembles at
the promoter is likely to contribute to regulation.

**Activators and co-activators**

Regulated expression of a gene depends in large part on the activity of
transcription factors that bind to specific sequences in the regulatory promoter
or enhancer. Many of the dynamic changes in gene expression in response to a
developmental programme or external signal are due to changes in the levels or
activity of specific transcriptional activators and repressors. Although the activ-
ity of promoter-specific repressors is important for eukaryotic gene regulation,
repression mechanisms will not be discussed here. Transcriptional activators
have at least two functional domains: a DNA-binding domain and an activation
domain [19]. The DNA-binding domain binds a specific sequence present in
the regulatory promoter or enhancer, and thereby serves to localize the activa-
tor near the gene to be regulated. The activation domain mediates protein–
protein interactions with other activators, co-activators and GTFs to increase
transcription. Although the first activation domains to be identified were
noticeably rich in acidic amino acids, many different amino acid sequences have
been found to function as activation domains [20]. Studies of several
transcriptional activators, in particular the VP16 protein from herpes simplex
virus, which does not bind DNA directly but is localized to the promoter by
binding to a DNA-binding protein, have identified interactions between
activation domains and the GTFs TFIIA, TBP, TFIIB, TFIIF and TFIHH [20].
Although it has not been established that all of the observed interactions are
physiologically relevant, it is likely that activator interactions with many
different components of the RNA polymerase II transcriptional machinery can
enhance transcription. Thus, there are many targets for activation domains, and,
in fact, a single activator can interact with multiple general transcription factors
or co-activators to stimulate transcription [21].

Surprisingly, when researchers analysed transcription *in vitro* using puri-
fied activators, GTFs and RNA polymerase II, it was found that these compo-
nents were not sufficient to generate high levels of activated transcription [22].
These *in vitro* studies revealed that additional co-activators were required in
addition to the basal transcription machinery and sequence-specific activators.
Subsequent biochemical and genetic studies have identified numerous co-activ-
ators that contribute to activated transcription [21]. Many co-activators have
been shown to bind directly to activators, providing a mechanism for targeting
co-activator function to specific promoters. For example, the interaction of
TAFII110 with the glutamine-rich activation domains of the transcriptional
activator Sp1 is important for Sp1-mediated activation [23]. Similarly, the thy-
roid-hormone-receptor-associated protein 220 component of the human medi-
ator complex binds directly to the vitamin D and thyroid hormone receptors
in a ligand-dependent manner [24]. Some co-activators, such as the TAFII8 in
Mechanisms of activation

Multiple steps are required to produce a correctly initiated full-length transcript. These steps include binding of the GTFs and RNA polymerase II to the promoter, melting of the template strands, formation of the first phosphodiester bond (initiation), release of the RNA polymerase II complex from the promoter (promoter clearance), movement of the polymerase through the gene (elongation) and termination. In addition, the components of the transcription machinery are recycled to allow multiple rounds of
transcription from a single template. Each of these steps could, in principle, be subject to negative regulation by repressors or positive regulation by activators.

It is perhaps not surprising that the first step in this process, binding of the RNA polymerase II transcription complex to the promoter, is subject to extensive regulation. As shown in Figure 2, binding of the RNA polymerase II machinery to the promoter is thought to occur in at least two steps — binding of TFIID together with TFIIA to the core promoter, and subsequent binding of an RNA polymerase II holoenzyme complex to the resulting DNA–protein assembly. Although there are clear examples of activators that stimulate elongation, many transcriptional activators act to stimulate recruitment of the RNA polymerase II machinery to the promoter (Figure 3) [19]. Many activators, when bound to nearby sites, function to increase TFIID binding to the core promoter in vitro and in vivo [9,25]. Studies using immobilized template assays to analyse proteins bound at the promoter have also shown that, under conditions where TBP/TFIID is bound at the promoter, activators can increase recruitment of holoenzyme components, including TFIIB and RNA polymerase II [26,27]. Activators function to increase binding of the transcription machinery to the promoter in at least two ways: (i) the additional affinity provided by a simple protein contact and (ii) stabilization of a conformation that increases assembly of the GTFs on the promoter (Figure 4).

The additional affinity provided by binding to an activator localized at the promoter is, in many cases, sufficient to increase binding of the transcription machinery to the core promoter [19]. Experiments using fusion proteins dra-
matically illustrate that increasing the local concentration of the transcription machinery through many diverse types of interactions results in increased transcription. Direct fusion of TBP to a heterologous DNA-binding domain produces high levels of transcription from promoters with the appropriately positioned binding sites [28,29]. Similarly, a heterodimerization domain, when localized to a promoter by fusion to a DNA-binding domain, was shown to activate transcription when the corresponding dimerization domain was fused to TFIIB [30]. Artificial recruitment of several other holoenzyme components also leads to activated levels of transcription [19]. The results of such ‘activator bypass’ experiments argue that the activator does not have to interact in any special way with the transcription machinery, since an activation domain can easily be substituted by a DNA-binding domain or a dimerization domain. Consistent with the idea that many activators increase the affinity of the transcription machinery for the promoter, but do not otherwise modify its activity, increasing the concentration of holoenzyme in an in vitro reaction is sufficient to obtain activated levels of transcription [31].

Interaction with an activator has also been found to promote a conformational change that enhances binding of the transcription machinery to the promoter. Studies of the mechanisms of transcriptional activation by the Zebra activator from Epstein–Barr virus have revealed that Zebra interacts with TFIID. In this case, however, increasing the concentration of TFIIA–TFIID to ensure binding to the promoter was not sufficient to obtain activated levels of transcription. Interaction with the Zebra activator stabilizes a conformation of the TFIIA–TFIID–TATA complex with increased affinity for the RNA polymerase II holoenzyme [32]. Although the exact nature of this conformational change is not known, other activators have been observed to alter the interaction of TFIID with the core promoter, as judged by DNaseI footprinting, suggesting that the conformation of TFIID may be commonly subject to regulation [9]. Interestingly, the pattern of the TFIID footprint varies on different core promoters, raising the possibility that the sequence of the core promoter may influence the conformation of TFIID and thereby the response to certain activators.

**Transcriptional synergy**

In eukaryotes, different combinations of a relatively small number of transcription factors are used to generate a large array of specific transcription patterns in a process often referred to as combinatorial control. The effect of two or more activators working together at a promoter is generally much more than additive [5]. Synergistic activation has important consequences for the specificity of gene regulation, since it ensures that a gene will only be expressed when each of the activators that regulate it are functional. For example, induction of the gene encoding interferon-β in response to virus infection requires the transcription factor nuclear factor κB (NF-κB). Not all
signals that activate NF-κB, however, result in stimulation of interferon-β expression. Interferon-β transcription is induced only when all of the transcription factors, including NF-κB, activating transcription factor 2 (ATF2)/c-Jun and interferon regulatory factor-3 (IRF-3), function together at the enhancer [33,34]. Thus the interferon-β enhancer serves to integrate different external signals. Several mechanisms are likely to contribute to synergistic activation [5]. First, interactions between activators may allow for co-operative binding and increased stability of the activator–DNA complex. Secondly, non-competitive interaction of multiple activators with different surfaces of the transcriptional machinery should increase the effective interaction energy. Thirdly, if multiple activators stimulate different steps in the transcription reaction, such as recruitment of TFIID and the holoenzyme, the effects may be greater than additive. It is likely that all of these mechanisms of synergy contribute to the high levels of activation mediated by enhancers.

**Perspectives**

Although studies of several model promoters and activators has revealed much about regulated transcription initiation in eukaryotes, several questions remain to be answered. In particular, regulation of very few cellular genes has been studied in sufficient detail to understand the complex interplay between the sequence of the core promoter, the requirement for individual GTFs and the activity of specific activators. In addition, many co-activators have been identified, including the multi-subunit human mediator complex, and the mechanisms by which these co-activators function with activators to regulate transcription at specific promoters continues to be a subject of active investigation. It is interesting to note that several co-activators have associated kinase or acetyltransferase activities, which may contribute to their function. Since gene regulation in vivo occurs in the context of chromatin, it is of significant interest to understand how the functions of activators, co-activators and GTFs described here are co-ordinated with the activity of chromatin-modifying complexes. Further studies of transcriptional activation mechanisms should advance our knowledge of how transcription of each of the more than 60000 human genes is properly regulated.

**Summary**

- DNA sequences that determine transcriptional regulation of a typical eukaryotic gene consist of a core promoter, which serves as a binding site for the GTF TFIID, and regulatory promoter or enhancer sequences, which bind transcriptional activators.
The RNA polymerase II transcription machinery consists of over 50 proteins which are thought to bind to the core promoter in as few as two steps: binding of TFIIA–TFIID, followed by binding of a large pre-assembled holoenzyme complex consisting of the remaining GTFs, RNA polymerase II and associated regulatory proteins.

Activators function to increase binding of the transcription machinery to the promoter in at least two ways: (i) simple protein–protein interactions with activators increases the affinity of the transcription machinery for the promoter, and (ii) some activators stabilize a conformation of the TFIIA–TFIID–DNA complex that enhances binding of the holoenzyme.

Recent studies have identified many co-activators that function with activators to increase transcription by the RNA polymerase II transcription machinery. Although some co-activators may serve as bridges to connect activators with the transcription machinery, the mechanism of action of many co-activators has not yet been determined.