Regulation of gene expression through post-initiation controls

S Vidyalakshmi and V Ramamurthy*
Department of Biotechnology, PSG College of Technology, Coimbatore 641 004, India

Received 7 June 2004; revised 17 February 2005; accepted 27 February 2005

Synthesis of messenger RNA in an eukaryotic cell can be regulated at multiple levels. Although the regulatory processes at the initiation step help in the discrimination between whether or not a gene need to be transcribed in a particular cell at a particular time, the decision to go through with the complete synthesis could still be revised through controls at the succeeding steps of the process. This article reviews the regulatory controls exercised at the transcription elongation step and the factors which participate in this process. Inspite of the plethora of factors contributing to the elongation process, the main catalytic activity is entrusted with the RNA polymerase II, and other factors mainly assist or modify the polymerase to alter its enzymatic efficiency.

Keywords: CTD phosphorylation, dephosphorylation, RNA polymerase II, transcription elongation
IPC Code: Int. Cl.7 C07H21/02; C12P19/32

Introduction

Expression of each gene requires combinatorial interactions among several transcription regulatory proteins. The overall number of proteins involved in regulation of the eukaryotic transcription process is large, of which only a few participate at any given time. The interactions among them to orchestrate gene expression can be expected to be complex due to the number of participants and the diversity of the contexts under which they come together. In the mRNA transcription process, the polymerase responsible for synthesis is the central molecule, around which all the support characters work.

Transcription could be broadly divided into five stages of progressive events. The stages may be distinguished not only based on the outcome of the event, but also on the participating molecules, and perhaps also on the regulatory handles for the cell to choose alternative outcomes. The five stages of transcription cycle are pre-initiation, initiation, promoter clearance, elongation, and termination. The transcription process could be regulated at any of these stages by modifying the region where the process is to occur or modifying the participants involved in the process, their recruitment and clearance from the site of action as well as their synthesis and recycling. The core promoter is the minimal stretch of contiguous DNA sequence that is sufficient to direct accurate initiation of transcription by the RNA polymerase II machinery. The elaborate regulations of the Pol II transcription cycle may, to a large extent, underlie organismal complexity and animal diversity. In eukaryotic cells, the packaging of DNA into chromatin plays an important role in regulating the accessibility of the transcriptional factors. Histone acetylation is an important mechanism by which nucleosomes are modified. The core histones can be acetylated on the lysine side chains of their amino terminal region, decreasing their positive charge and presumably reducing their affinity for negatively charged DNA or other chromatin proteins. The importance of histone acetylation in transcription has been implicated by the identification of histone acetyl transferase (HAT) activities within co-regulatory complexes. HAT complexes, such as the SAGA complex in budding yeast, contribute to control transcription by altering chromatin to serve as a template for transcription. SAGA is a 1.8 MDa transcriptional co-activator and contains several different classes of co-activator proteins, such as SPT (Suppressor of Ty insertions), ADA (transcriptional adapters) and TAFs (TBP Associated Factors). Thus, the SAGA complex unites the ability to modify histone tails and directly interact with transcriptional activators.

At one level, by modifying the DNA packaging proteins, histones and high mobility group proteins,
the availability of DNA to the transcription factors could be regulated\textsuperscript{15}. The diversity in promoter composition affords another level of control by restricting the choice of the promoter-reactive factors and their interaction with distal signals such as enhancers and silencers. Another important aspect of transcription regulation is the turn-off of the initiated process. This may be achieved by inactivation of the transcription initiating factors, once the process has been activated\textsuperscript{16}. This review focuses on the next level of action, viz. the post-initiation events during the transcription process, and its contribution to expression of eukaryotic genes. Further processes, not discussed in this article, including transcription termination and transcript processing are also modulated to control the output.

**Pre-initiation and Initiation Events**

Recent structural studies of Pol II complexes mark the beginning of a detailed structure-function analysis of the transcription cycle\textsuperscript{17}. Accurate and efficient transcription from the core promoter requires RNA polymerase II (RNAP II) along with auxiliary general transcription factors (TF) IIA, IIB, IID, IIE, IIF, and IIH. The formation of the transcription pre-initiation complex (PIC) is a sequential event of progressive assembly of the factors (Fig. 1). In the stepwise PIC assembly, TFIID and TFIIB are the first two factors that interact with the core promoter and hence the recognition of core promoter motifs.

Early evidence indicated that there is a difference in the interactions that held components together in the PIC and in the elongation complex. During promoter clearance, RNAP II undergoes a fundamental change in the nature of interactions that hold the complex together. A process that is obligatory to promoter clearance is abortive initiation. This is a cyclic catalytic reaction that generates a nested set of released RNA oligomers up to 12 nucleotides long. Usually just a few percent of the initiated transcripts become complete, useful products during this process\textsuperscript{18}. In a number of genes, the transcription complex is found to be paused at 20 to 40 nucleotides downstream of the transcription start site. This pause may be interpreted as a transitional state from the initiation to the elongation phase of transcription. The elongation competence of the transcription complex is influenced by the RNA-DNA hybrid within the transcription bubble and the interaction of RNAP II with the segment of DNA immediately downstream of the bubble\textsuperscript{19,20}. When the advancing transcription complex reaches a template region where one or both of these interactions are weak, such as locations encoding long stretches of U in the transcript, RNAP II slides upstream along the template to establish a more favorable interaction. The upstream translocation of the transcription bubble displaces the 3′-end of the RNA from the active site of RNAP II, resulting in transcriptional arrest\textsuperscript{21-23}. Arrested polymerases retain their transcripts, but cannot resume RNA synthesis unless the transcription complex moves downstream or the transcript is cleaved at the active site\textsuperscript{24-30}. Either of these events realigns the 3′-end with the catalytic center and allows transcription to continue\textsuperscript{31}. The transcriptional complex may also be stalled in the promoter-proximal region without upstream translocation, but unable to continue transcription\textsuperscript{32}. The relationship of transcript length, transcript sequence and elongation competence for promoter proximal transcriptional complexes is not well understood.

**Post-initiation Fidelity in Transcription Control**

Capping serves as a pre-requisite to elongate initiated transcripts and, thus, contribute as a transcriptional checkpoint, since only capped mRNAs would be transcribed efficiently\textsuperscript{33}. Although the pre-initiation and initiation stages of transcription have received the most attention during the past decade, the past few years have been a watershed for biochemical studies of the transcription elongation complex—a diverse collection of transcription elongation factors and nuclear proteins that regulate the activity of RNAP II during the elongation phase of messenger RNA synthesis have been identified and biochemically characterized\textsuperscript{34-42}. After transcription initiation, the transcription complex is modified to an elongation complex. In addition to the general initiation factors, at least 16 elongation factors have been reported to either increase the overall rate of transcription or prevent premature pausing and arrest of the transcript and thus contribute to regulate the expression of many eukaryotic genes.

The largest subunit of RNAP II is unique in that it contains an unusual domain at its carboxy terminus comprised of tandem repeats with a consensus sequence of Y-S-P-T-S-P\textsuperscript{43-44}. This consensus repeat has been conserved through evolution although the number of the repeat varies in different species. The carboxy terminal domain (CTD) of RNAP II plays an essential role in transcription catalyzed by RNAP II through its reactivity with other transcription factors.
factors and the modifications it undergoes. A similar, functionally diverse domain is absent in RNAP I and RNAP III. The CTD plays a critical role by mediating the interaction of core RNAP II with the factors necessary for pre-initiation complex assembly and respond to other transcriptional regulators. Two of the serine residues (Ser 2 and Ser 5) found in the C-terminal repeat can be conditionally phosphorylated. Complete phosphorylation of CTD will cause a massive increase in the acidity of this domain. RNAP II with an unmodified CTD (RNAP II A) is assembled preferentially in the transcription pre-initiation complex formed at the promoter, whereas RNAP II with a hyperphosphorylated CTD (RNAP II O) is associated with elongation complexes. Upon completion of the transcription, RNAP II O must be dephosphorylated by a CTD phosphatase to regenerate RNAP II A and complete the cycle (Fig. 2). CTD phosphorylation is temporally correlated with the formation of a competent elongation complex. Although the causal relationship between phosphorylation and elongation is still to be established, the available data suggest this to be the case indicating a strong correlation between both the events. Transcription complexes paused near the transcriptional start site on a number of Drosophila genes contain RNAP II A, while the RNAP II O form is tethered to proteins associated with the promoter through its CTD during promoter clearance. This could be interpreted as phosphorylation triggers promoter clearance. An alternate interpretation is that the disruption of RNAP II O interaction with the promoter-
associated factors and phosphorylation is a necessary step to generate an elongation-competent form of RNAP II. At the time of transcript initiation, a CTD kinase stably interacts with the pre-initiation complex, most likely through TFIIH and phosphorylates CTD. Transcription pausing may result when the phosphate incorporated into the CTD is removed during the elongation process and not restored by the CTD kinase. Alternatively, the act of pausing may trigger dephosphorylation of the CTD. It will be of considerable interest to know if phosphate turnover occurs during transcript elongation, whether or not the rate of turnover is gene-specific and what the consequence of this turnover is on pausing and termination.

Since RNAP IIA and RNAP IIO have distinct roles in the transcription cycle, CTD kinases and CTD phosphatase can act as positive or negative regulators of transcription depending on the point in the transcription cycle at which they function. For example, phosphorylation of the CTD concomitant with transcript initiation might stimulate transcription, whereas phosphorylation of free RNAP II would reduce the amount of RNAP IIA available for recruitment to the promoter and, hence, inhibit transcription. Conversely, CTD phosphatase that dephosphorylates RNAP IIO in the initiated or elongation complex may well inhibit transcription, whereas dephosphorylation of RNAP IIO upon completion of the transcript would stimulate transcription. Elucidation of the various CTD kinases and phosphatases that modulate the level of RNAP II phosphorylation in vivo, how these activities are coordinated, and the consequences they have on the activity of RNAP II at discrete steps in the transcription process at different situations will go a long way in understanding the transcriptional process.

Transcription elongation by RNAP II is a dynamic process that does not occur at a constant rate\textsuperscript{52}. There is a wide variation in the time the polymerase spends in inserting nucleotides into a growing RNA chain. The dwell time at a specific site can vary as a function of the template and transcript sequence and structure\textsuperscript{32,53,54}. Throughout its elongation phase, RNAP II can encounter constraints leading to pause, arrest and termination. Although pausing presents a temporary impediment, transcriptional arrest has more potent consequences (Fig. 3). Transcriptional arrest within the coding region of the gene would effectively repress mRNA synthesis from the affected gene.
Regulation of gene expression at the elongation stage of transcription has been described for many genes\(^{55,56}\). Two distinct classes of early elongation complexes have been observed\(^ {32,34,36,38,57-60}\). One class undergoes abortive elongation and gives rise to short transcripts; whereas, the second class surmounts early blocks and carries out productive elongation. Several negative transcription elongation factors have been identified to cause abortive elongation. Factors such as DSIF and NELF associate with RNAP II and prevent transcription elongation. This inhibition can be overcome by other positive factors\(^ {61}\).

Six general positive elongation factors (SII, P-TEF-b, TFIIF, Elongin, 11-19 lysine-rich leukemia protein and the Cockayne syndrome complementation group B protein) have been defined biochemically using \textit{in vitro} transcription assays and found to increase the efficiency of transcription elongation by RNAP II. SII and P-TEFb prevent the transcription complex from engaging into a prematurely arrested complex. The other elongation factors, viz. TFIIF, Elongin, 11-19 lysine-rich leukemia protein and Cockayne syndrome complementation group B protein, act to increase the overall rate of RNA chain elongation by RNAP II through suppressing the transient pausing of polymerase at many sites along the DNA template\(^ {62-64}\).

Arrest sites are often but not always associated with tracts of A-T base pairs, although no consensus sequences have emerged. T-rich sites are scattered...
throughout the eukaryotic genome that can participate in blocking of elongating RNAP II. In addition to the T-rich arrest sites, pausing by RNAP II may depend on DNA sequences downstream of the arrest site\cite{65}. Recent studies indicated the characterization of a novel RNA Polymerase II arrest site, which lacks a weak 3' RNA-DNA hybrid in the tyrosine hydroxylase gene\cite{66}. During the process of DNA dependent transcriptional arrest, RNAP II undergoes a dramatic conformational change that results in a loss of contact between the 3' end of the nascent transcript and the polymerase catalytic site\cite{64,67,68}. This loss of contact between 3' end of nascent transcript and RNAP II is due, in part, to the change from the monotonic mode of elongation to a discontinuous one. Escape from such transcriptional impediments requires the presence of transcription factor SII, a 38 kDa protein originally purified from Ehrlich ascites tumor cells by Sekimizu \textit{et al}\cite{69}. SII is a unique cellular factor that exerts arrest-relief activity on RNAP II by preventing it from stopping on the template DNA. SII allows the passage of arrested RNAP II from DNA-dependent pause sites by promoting reiterative cleavage by RNAP II and re-extension of the nascent transcript held in the active site. Cleavage of the nascent transcript requires a physical interaction between the RNA and RNAP II and is inhibited by low concentrations of α-amanitin. It is likely that SII-induced transcript cleavage reactivates arrested RNAP II by realigning its catalytic site with the newly formed 3' end of the nascent transcript.

Thomas \textit{et al}\cite{70}, using an \textit{in vitro} transcription assay, have shown that, in the presence of SII, RNAP II can proofread using 3'-5' nuclease activity. When SII was added to the transcription reaction, RNAP II could quantitatively remove the misincorporated nucleotides from the nascent transcript during rapid chain elongation. Addition of SII to a transcription reaction \textit{in vitro} strikingly altered the RNA base composition, suggesting that SII governs stable incorporation of more correct and fewer incorrect nucleotides into the growing nascent RNA chain synthesized by RNAP II. The basis of discrimination between the incorporation of correct and incorrect base was the slow addition of the next nucleotide to the mismatched terminus. Thus, in addition to its role in transcriptional elongation, SII can regulate the transcriptional fidelity and proofreading by inducing the nuclease activity of RNAP II. Since SII binds directly to RNAP II and does not recognize specific DNA sequences by itself, it could be considered as a general transcription factor\cite{71}. However, a tissue specific SII binding factor, FESTA has been identified\cite{72}. Using two-hybrid assay, it was found that FESTA directly interacts through its C-terminal tail and activates SII under \textit{in vivo} conditions. Tissue specific expression of FESTA suggests that it interacts with SII and thereby its effect on elongation regulation could be limited to a few tissues like kidney and spleen, and suggests other SII activity modifiers in other tissues. The direct \textit{in vivo} role of SII needs further exploration.

It has been observed that TFIIF, ELL, and Elongin can all dramatically increase the ability of RNAP II to bind to and extend DNA primers in a DNA template-directed transcription reaction. TFIIF mutations that reduce transcription elongation activity also reduce this DNA primer binding activity, suggesting that the mechanism by which TFIIF facilitates RNAP II to extend DNA primers with ribonucleotides in a DNA template-directed manner is a consequence of its participation in the formation of ternary elongation complex more effectively\cite{73}. With the RNAP II binding to the 3'-hydroxyl terminus of DNA, similar to its binding to the 3'-end of pre-formed RNA molecule, the RNAP II catalytic site would then add ribonucleotides to the DNA 3'-hydroxyl terminus as if it were the 3'-end of a nascent transcript. The observation that TFIIF, ELL and Elongin can promote extension of DNA primers by RNAP II is consistent with the model that these transcription factors facilitate proper positioning of the 3'-ends of the nucleic acid primers in the polymerase catalytic site. Similarly, they could suppress transient pausing of the transcription complex by preventing slippage of RNAP II upstream to the 3'-end of the nascent transcript. Gù and Reines observed that TFIIF can decrease the rate at which RNAP II gets arrested, suggesting that TFIIF can inhibit backsliding of RNAP II and the consequent displacement of the 3'-end of the nascent transcript from the polymerase catalytic site and the template\cite{74}.

TFIIF, ELL and Elongin may expedite elongation not only by increasing the forward rate of nucleotide addition by RNAP II, but also by inhibiting SII-induced nascent transcript cleavage. Inhibition of SII-induced nascent transcript cleavage by TFIIF, ELL or Elongin could be a manifestation of their intrinsic elongation promoting activity. Whether TFIIF, ELL...
and Elongin inhibit SII-induced transcript cleavage by controlling the orientation of the 3'-ends of nascent transcripts or by simply interacting with the polymerase elongation complex and sterically blocking interaction with SII is still unclear.

Elongin has been shown to suppress transient pausing of RNAP II at multiple sites along the DNA. It was initially purified from rat liver nuclei by its ability to increase the catalytic rate of transcription elongation from promoter-specific transcription. It was a heterotrimer of three subunits (A, B and C) with apparent molecular masses of approximately 110, 18 and 15 kDa, respectively. Biochemical and mechanistic studies have shown that Elongin A is the transcriptionally active subunit of the complex and that Elongin B and C can regulate the transcription elongation activity of Elongin A. In vitro experimental data have demonstrated that Elongin C can interact directly with Elongin A in the absence of Elongin B and can increase the specific activity of Elongin A. On the other hand, Elongin B cannot alter the specific activity of Elongin A, nor is capable of direct physical interaction with Elongin A in the absence of Elongin B and C. This suggests that the RNAP II binding and elongation activating domain of Elongin A are different. Elongins A also contains the Elongin BC binding motif, T(S)LXXXCXXX(V/L/I) at their carboxy terminal region. Consistent with the conservation of this sequence motif, Elongins A2 and A3 were shown to form stable complex with Elongin BC. However, unlike with Elongin A, the elongation activation of both Elongin A2 and Elongin A3 were not influenced by binding with Elongin BC. Although the elongation activating domain of Elongin A, between residues 520-690, is conserved among other Elongins, there are subtle differences in the response of the Elongins to activation by Elongin B and C. Also, it is evident that homologues of Elongin A are expressed in a tissue-specific manner and this may provide functional diversity either through their variation in transcription elongation activity, or controlling the availability of the regulatory subunits of Elongin BC.

Diseases and Elongation Factors

The involvement of several proteins illustrates the complexity of the transcription process and, hence, could be disrupted due to mutational changes in any of the involved participants. In the case of cell-type or tissue-specific transcription modulators, like Myc, Fos, Jun, Rel, Myb, etc., the mutant forms associated with diseased conditions led to the isolation of the wild-type gene. Although one might surmise that alterations to general transcription factors would have catastrophic consequences, recent findings show linkages between mutated general transcription associated factors and human disease. The product of Xeroderma Pigmentosum (XP) complementation group B was found to be the largest subunit of TFIH. Similarly, the product of XP complementation group D was identified to be another subunit of TFIH. Defects in either of these proteins are also associated with a fraction of the patients suffering from Cockayne syndrome and trichothiodystrophy. All these three syndromes are autosomal recessive and are associated with nucleotide excision repair defect. Clinical manifestations of XP include extreme sensitivity to sunlight-induced skin damage, pigmentation defects, pre-disposition to skin cancers, and neurological and ocular abnormalities. People affected with Cockayne syndrome show sunlight sensitivity but are not cancer-prone, short in stature, display mental retardation, microencephaly, and skeletal and retinal abnormalities. Trichothiodystrophy patients show sulphur deficient brittle hair, mental retardation, abnormal facial features and sunlight sensitivity, but are not skin cancer prone. Both the subunits of TFIH exhibit DNA helicase activity and are associated with CTD kinase. Due to helicase activity, both subunits are implicated in nucleotide excision repair of damaged DNA, following lesions due to UV light or chemical agents. Thus, TFIH appears to be involved both in DNA repair as well as transcription.
initiation, and its defect is associated with a few disease conditions showing generalized dysfunction in several systemic processes.

The protein responsible for von Hippel Lindau (VHL) tumor suppressor gene competes with Elongin A for binding with its regulatory subunits Elongin BC. Thus, the transcription elongation rate of Elongin A is indirectly influenced by the presence of the tumor suppressor protein and perhaps the other homologues of Elongin A. Mutations in the VHL gene predispose the individuals to different types of cancer, particularly renal carcinomas and phaeochromocytomas92-95.

Besides mutations, genes could be corrupted through chromosomal translocations. Of the many translocations associated with leukemia, involvement of MLL is frequently observed. At the translocation point the amino terminal half of the protein is fused with any one of six genes to result in transformation of the cells. The only functionally characterized gene of this group is the elongation factor ELL96. ELL translocated fusions with MLL in chromosome 11q23 has been found in several cases of acute myeloid leukemia97,98.

Post-initiation transcription regulation offers the higher eukaryotes a mechanism for subtle changes in the requirement for expression of genes. Transcription elongation process is controlled by different factors acting on RNAP II. The polymerase itself appears to be inefficient in the sense of its poor processivity and acting on RNAP II. The elongation process is controlled by different factors that are required for expression of genes. Transcription through chromosomal translocations. Of the many translocations associated with leukemia, involvement of MLL is frequently observed. At the translocation point the amino terminal half of the protein is fused with any one of six genes to result in transformation of the cells. The only functionally characterized gene of this group is the elongation factor ELL96. ELL translocated fusions with MLL in chromosome 11q23 has been found in several cases of acute myeloid leukemia97,98.

Post-initiation transcription regulation offers the higher eukaryotes a mechanism for subtle changes in the requirement for expression of genes. Transcription elongation process is controlled by different factors acting on RNAP II. The polymerase itself appears to be inefficient in the sense of its poor processivity and proclivity to modifications affecting its function. Competition among the regulators of the elongation factors is another level of control that the cells exercise in controlling the fidelity of transcript formation.

Acknowledgement

The authors thank the Principal and the management of PSG College of Technology for encouragement and support, and the Department of Biotechnology for research grant.

References

20. Nudler E, Gusrav I, Avetissova E, Kozlor M & Goldfarb A, Spatial organization of transcription elongation complex in...


32 Ramamurthy V, Maa M C, Harless M L, Wright D A & Kellems R E, Sequence requirements for transcriptional arrest in exon one of the murine adenosine deaminase gene, Mol Cell Biol, 10 (1990) 1484-1491.


an atypical TATA box which binds TFIIID and transcriptional activity is stimulated by multiple upstream Sp1 binding sites, *J Biol Chem*, 266 (1991) 21765-21772.


