Published in final edited form as:
Nature. 2009 September 10; 461(7261): 186–192. doi:10.1038/nature08449.

Defining mechanisms that regulate RNA polymerase II transcription *in vivo*

Nicholas J. Fuda[1], **M. Behfar Ardehali**[1], and **John T. Lis**[1]
[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA

Abstract

In the eukaryotic genome, the thousands of genes that encode messenger RNA are transcribed by a molecular machine called RNA polymerase II. Analysing the distribution and status of RNA polymerase II across a genome has provided crucial insights into the long-standing mysteries of transcription and its regulation. These studies identify points in the transcription cycle where RNA polymerase II accumulates after encountering a rate-limiting step. When coupled with genome-wide mapping of transcription factors, these approaches identify key regulatory steps and factors and, importantly, provide an understanding of the mechanistic generalities, as well as the rich diversities, of gene regulation.

The genetic information encoded in the DNA of eukaryotic genes is transcribed into RNA by large molecular machines called RNA polymerases. One of these machines, RNA polymerase II (Pol II), transcribes all the protein-coding genes. The control of Pol II activity is highly modulated at individual genes, and this specific regulation is critical for both the homeostasis of cells and the programmed development of multicellular organisms. The execution of this regulation is dictated by combinatorial molecular interactions of transcription factors with each other and with specific DNA sequences at each gene. Modern biochemical and molecular methods coupled with genetics and genomics have identified thousands of factors that participate in regulated transcription[1]. Most of these factors are proteins, but a growing number of them are RNAs. They enable Pol II to gain access to the gene's promoter, to initiate RNA synthesis at the transcription start site (TSS) of the gene and to generate a productively elongating transcription complex that produces a full-length RNA transcript.

The thousands of transcription factors involved in the transcription process may be true regulatory factors or simply critical cogs in the cycle of transcription. True regulatory factors are likely to represent only a fraction of the total number of factors that are important for gene expression. As an analogy, consider a motor vehicle: a car has numerous crucial components and processes that are required to achieve acceleration and proper speed (cylinders, spark plugs, tyres and so on), but components regulated by the driver are limited to the ignition, the steering wheel, the accelerator and brake pedals, and the gear stick. Therefore, it is important to identify the true regulatory factors and the associated biochemical processes that execute gene regulation. The status and local density of the ultimate target of regulation, the transcription machine Pol II, have proved extremely useful in assessing the steps in the transcription cycle that are rate limiting and are altered *in vivo* by particular transcription factors (the driver in the above analogy).

Correspondence should be addressed to J.T.L. (jtl10@cornell.edu).

The authors declare no competing financial interests.

In this Review, we discuss how in-depth mechanistic analysis of individual genes coupled with large-scale analysis of transcription-factor binding over an entire genome can distinguish the key steps at which transcription is regulated, and how these steps can be accelerated in an activator-dependent manner.

Gene promoters and factor interactions

The DNA sequences in and around specific gene promoters provide the code that dictates when, where and at what level specific genes are transcribed. This code comes in three parts: the core promoter, the region proximal to the core promoter, and the more distant enhancer sequences (Fig. 1). In various combinations, the elements of the core promoter sequence target the assembly of distinct preinitiation complexes (PICs) composed of the general transcription factors (GTFs)[2]. Promoter-proximal regions and more distant enhancer sequences direct the binding of specific transcription factors, called activators or repressors (see page 199 for a more detailed discussion of enhancers). Although activators or repressors can interact directly with components associated with the core promoter, they execute their regulation predominantly through co-regulators, which are often multiprotein complexes. Some of the co-regulators can interact directly with Pol II and GTFs and influence expression. Others can reorganize nucleosomes or covalently modify chromatin, and change the chromatin architecture of the gene. This can in turn influence transcription-factor associations and the transcriptional status of Pol II.

Although present evidence suggests that many steps in the transcription process may be rate limiting, the question remains whether these rate-limiting steps are actual points of regulation. To meet this criterion, these steps should be regulated by factors in response to particular physiological, environmental or developmental signals. Although transcription regulatory factors that act as repressors can also modulate specific steps, we focus here on activators, as they seem to predominate as critical modulators of gene expression in eukaryotes.

In-depth analyses of individual genes, or sets of co-regulated genes, have revealed critical mechanistic insights into transcription factors that take part in regulation in response to specific cellular signals. This information, when coupled with more recent large-scale analyses of the associations of such factors over the entire genome (which have been carried out by individual laboratories, as well as by the Encyclopedia of DNA Elements (ENCODE) and modENCODE consortia), allows the generality of particular regulatory mechanisms to be assessed. These genome-wide efforts efficiently appraise the collections of genes that associate with particular transcription factors and thereby define many potential participants in any regulatory mechanism. They also reveal the regulatory circuitry of gene expression networks and how these networks respond to cellular signalling[3]. Interpreting how the transcription factors and gene circuitry respond to signals and lead to transcriptional regulation requires that we identify not only the factors that respond to signals but also the rate-limiting steps in transcription.

Rate-limiting steps in transcription

The transcription cycle consists of at least eight distinct major steps at which transcription could be rate limiting and activators could potentially act to increase the rate of transcription (Fig. 2). The transcription cycle begins with Pol II gaining access to the promoter, which in some cases requires the promoter being cleared of nucleosomes that obscure access to Pol II and the GTFs (step 1). A PIC assembles on the core promoter (step 2). The DNA is then unwound, and Pol II initiates transcription (step 3). Early-elongating Pol II gets a stable grip on both the DNA and the growing RNA chain, escapes/clears the core promoter and proceeds to the promoter-proximal pause region (step 4). The paused Pol II complex is then hyperphosphorylated and escapes from the pause region in an unknown manner, either terminating or entering productive elongation (step 5). If it has not terminated, Pol II must then

productively elongate through the entire body of the gene (step 6). After this, Pol II undergoes termination (step 7), and it can reinitiate to start a new round of transcription (step 8).

Any of these major steps could, in principle, be rate limiting, and the distribution of Pol II across a gene can suggest which steps are rate limiting for that gene. The Pol II density across many genes has been determined in a plethora of individual gene studies[4]; moreover, a wealth of data has been obtained in recent genome-wide chromatin immunoprecipitation (ChIP) studies examining Pol II distribution across the genomes of several organisms: *Saccharomyces cerevisiae*[5], *Drosophila melanogaster*[6,7] and *Homo sapiens*[8]. In each organism, these studies have identified different classes of gene on the basis of their Pol II distribution: no Pol II, Pol II evenly distributed and Pol II enrichment at the 5′ end. Genes without Pol II are in an 'off' state, and are limited by the step at which the promoter is cleared of nucleosomes (step 1) or the step at which a PIC assembles (step 2). An even distribution of Pol II suggests that Pol II recruitment (step 2) is the rate-limiting step: none of the downstream steps leads to an accumulation of Pol II in other regions of the gene[9]. An enrichment in Pol II at the 5′ end suggests that steps downstream of Pol II recruitment (steps 3–5) are rate limiting. Because ChIP localization with a single Pol-II-specific antibody cannot distinguish between steps 3, 4 and 5, more experiments pinpointing the exact rate-limiting step need to be performed. The transition between PIC formation (step 2) and promoter escape (step 4) is marked by the unwinding of DNA, formation of a transcription bubble with a stable RNA–DNA duplex and lengthening of the nascent transcripts associated with Pol II. Transcription-bubble formation and RNA length can be distinguished by permanganate mapping of unpaired thymidines in the transcription bubble[10] and run-on assays[11,12], respectively. In addition, the transition between initiation and pausing (step 4) is marked by phosphorylation of the Pol II carboxy-terminal domain (CTD) repeats on Ser 5 by the kinase subunit of the GTF TFIIH (CDK7 in *Drosophila*), and productive elongation (step 6) is generally marked by phosphorylation of Pol II CTD repeats on Ser 2 by the kinase complex positive transcription elongation factor b (P-TEFb; CDK9–cyclin T in *Drosophila*). Therefore, using specific antibodies to examine these phosphorylation marks on genes with 5′-end Pol II peaks can help distinguish the rate-limiting step for those genes[13].

Regulating Pol II recruitment

Many genes regulated by the recruitment of Pol II have promoters covered with nucleosomes. Activators at these genes recruit nucleosome remodellers and nucleosome-modifying enzymes to allow GTFs and Pol II access to the promoter (Fig. 2, step 1) (see page 193 for details on nucleosome remodellers). *PHO5* in *S. cerevisiae* is one of the best studied of the genes that are regulated in this manner (Box 1). In other examples, it has been shown that both human and yeast activators interact with the SWI/SNF remodelling complexes (Swi/Snf complex in yeast) and positively stimulate transcription from nucleosome-containing templates[14]. In addition, recruitment of histone-modifying enzymes (for example recruitment of the histone acetyltransferase Gcn5 to galactose-inducible genes by the yeast activator Gal4 ([ref. 15])) provides another means by which activators influence and modulate the outcome of transcription by modifying promoter chromatin state.

In other genes, the promoter is free from nucleosomes, but Pol II recruitment is still rate limiting (step 2). During activated transcription, recruited Pol II quickly progresses into productive elongation and becomes relatively uniformly distributed across the gene[16]. At these genes, PIC assembly must be upregulated by activators. Extensive *in vitro* studies have shown activators can interact with many GTFs: TATA-binding protein (TBP), TFIID, TFIIA and TFIIB[17]. Activators also recruit the coactivator Mediator, which can interact with GTFs and increase expression[18,19]. These interactions might increase the binding of GTFs to the promoter or stabilize the PIC, allowing more efficient recruitment of Pol II. Additionally, activator-

dependent recruitment of chromatin-modifying enzymes results in distinctive chromatin marks on promoters. Domains associated with GTFs can bind to these marks[20,21], and these interactions can further aid in stabilizing PIC formation.

Regulating post-recruitment steps

In vivo Pol II distributions have also indicated that post-recruitment steps can be rate limiting. Enrichment in Pol II at the 5′ ends of genes suggests that steps between recruitment and productive elongation (steps 3–5) are rate limiting in these genes. Some of them may be regulated at initiation (step 3) or promoter escape (step 4). In the case of regulation at initiation, the Pol II associated with the 5′ end of the gene is contained within a PIC, and activators may regulate open-complex formation by recruiting or stimulating factors important for this step. For example, Mediator can interact with two GTFs crucial for unwinding DNA and forming open complexes: TFIIE and TFIIH[18,22]. Therefore, activators recruiting Mediator may increase the rate of open-complex formation. In the case of regulation at promoter escape/clearance, the Pol II associated with the 5′ end of the gene has initiated transcription but cannot transcribe to the promoter-proximal pause region owing to the instability of the RNA–DNA duplex and the inability of Pol II to break contacts with factors establishing the PIC. This can lead to abortive initiation[4]. Activators may mitigate these problems, but results on the extent of regulation at step 4 or how this happens *in vivo* are limited so far. TFIIH is again important for this step, not only for further unwinding of downstream DNA but also for the TFIIH-dependent Ser 5 CTD phosphorylation that occurs around this step, which may aid in breaking Pol II contacts with some promoter-bound factors[23]. Indeed, an activator can promote this phosphorylation *in vitro*[24], and Mediator enhances the TFIIH-dependent phosphorylation of the CTD[19].

Assays other than ChIP have shown that the Pol II that is enriched on the 5′ ends of many genes is already engaged in transcription but is held paused[12]. Directed studies of specific genes in the 1980s showed that Pol II was at high density on the 5′ ends of some genes, and this Pol II was extensively characterized in focused studies of *Drosophila Hsp70* and other heat-shock genes (Box 2; reviewed in [ref. 25]). Upon activation, the paused Pol II on *Hsp70* is released into productive elongation, and Pol II becomes evenly distributed across the gene. This indicates that the activator is regulating the transition from the paused state to productive elongation (step 5). P-TEFb is a major switch that has a critical role in facilitating the transition of Pol II from promoter-proximal pause sites into productive elongation[26] at most (if not all) genes; inhibition of P-TEFb dramatically decreases global transcription[27]. P-TEFb interacts directly with some activators[28–30], but others rely on different mechanisms to recruit P-TEFb indirectly (reviewed in [ref. 31]). Although P-TEFb is important for pause escape, Pol II still elongates many dozens of base pairs from the canonical *Hsp70* pause sites when P-TEFb is inhibited during heat shock[26]. Therefore, there may be other P-TEFb-independent mechanisms for releasing paused Pol II. In addition, elongation requires nucleosome loss or remodelling to occur, and it has been proposed that nucleosomes block the escape from pausing[32].

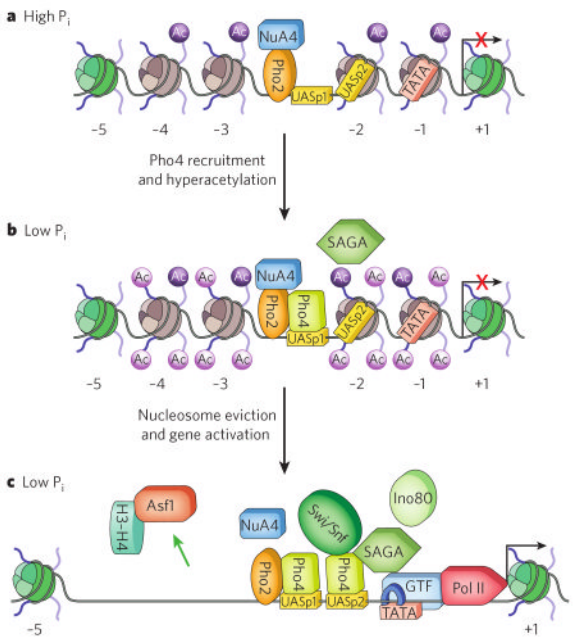
At present, the case for regulation at later stages (steps 6–8) in the transcription cycle is hard to make, but hints of such regulation exist[33]. It seems probable, for some genes, that cells have evolved means of at least modest regulation at these stages in response to cellular signals. Activator-dependent loss of nucleosomes aids in elongation (step 6). Additionally, the activator-dependent GTF-stabilizing interactions discussed earlier are important for recycling and reinitiation of Pol II (step 8). Some GTFs can remain associated with the promoter after the Pol II has escaped, and they form a scaffold that allows Pol II to initiate efficiently in successive rounds of transcription[34].

Box 1

The *Saccharomyces cerevisiae PHO5* gene is regulated at the chromatin-opening step

Transcription from the *Saccharomyces cerevisiae* acid-phosphatase gene *PHO5* (see figure, panel **a**) is regulated at the level of activator recruitment and eviction of four positioned nucleosomes (brown, −1 to −4) from the upstream regulatory and promoter region[51,52]. Pho2, a homeodomain-containing activator, and the histone acetyltransferase complex NuA4, which acetylates histones H4 and H2A (purple Ac) before induction, are both present at the promoter. Phosphate (P_i) starvation (see figure, panel **b**) induces *PHO5* by activating the cyclin-dependent-kinase inhibitor Pho81 (not shown), which inhibits the Pho80–Pho85 kinase complex (also not shown) and allows accumulation of the active unphosphorylated form of the basic helix–loop–helix activator Pho4 in the nucleus[53]. Pho4 binds mainly to the low-affinity UASp1 within the hypersensitive site that is flanked by two positioned nucleosomes on each side, and cooperatively interacts with Pho2. This Pho4–Pho2 complex triggers disruption of the positioned nucleosomes, and this event is concurrent with Pho4 binding to the high-affinity UASp2 and induction of transcriptional activation in a manner that depends on the acidic transactivation domain of Pho4 and on NuA4 ([refs 54–56]).

After Pho4 binding, the positioned nucleosomes become hyperacetylated (light mauve Ac) through the histone-acetyltransferase activity of the SAGA subunit Gcn5 and then undergo remodelling (see figure, panel **b**) before being evicted (see figure, panel **c**, green arrow) from the promoter. Both Swi/Snf and Ino80 complexes have been implicated in chromatin remodelling at *PHO5* ([refs 56, 57]). The H3–H4 histone chaperone Asf1 has also been shown to play a part in the eviction process[58,59]. Although Gcn5, Asf1 and chromatin remodellers are not essential for *PHO5* induction, their deletion results in a kinetic delay in the loss of nucleosomes and gene activation. These observations indicate that multiple mechanisms are in place for remodelling and eviction of the positioned nucleosomes at *PHO5*. The co-regulated *PHO8* gene is dependent on Gcn5 and Swi/Snf[60], indicating that these nucleosome modifications and remodelling events can have a range of effects on the Pho4-mediated activation of this co-regulated gene family.



Other phosphate-responsive genes are also induced during P_i starvation. But the degree of sensitivity to environmental P_i and the extent of expression on induction vary greatly among these genes. A recent study showed that variabilities in the activation threshold and transcription range of phosphate-system genes are governed, respectively, by the accessibility of high-affinity Pho4-binding sites before induction and the affinity and number of these Pho4-binding sites[61], highlighting the role of activator binding-site accessibility and nucleosome positioning on the dynamic range of transcriptional output.

Benefits of regulating at different steps

As suggested from this discussion, activators can act during distinct steps in transcription *in vivo*. Certain activators, such as Sp1 in mammals, target early steps in the cycle, whereas others, such as those with an acidic activation domain, can target early elongation/escape from pausing. Studies suggest that the distinct sets of targets may be independent of one another. The very strong viral acidic activator VP16 seems to act at both early and pausing escape steps[35]. The effect of *Drosophila* activator HSF on nucleosome removal could be separated from its effects in stimulating transcription on the *Hsp70* gene[36]. The ability of activators to stimulate multiple slow steps can lead to a much more rapid and robust activation through a kinetic synergism (reviewed in [ref. 37]).

The different steps in transcription provide multiple targets for the evolution of regulatory mechanisms. A block at early stages of promoter accessibility provides a means of placing a gene under tight control. An activator that stimulates nucleosome removal to unmask the promoter would allow that first step to occur; however, the gene could then require additional activators to stimulate later steps that eventually produce a messenger RNA. Thus, the activation of a gene could be regulated by a combination of signals that each acts on particular activators and their targeted steps, resulting in tight control; an example of such a gene is *PHO5* (Box 1).

The promoter-proximal paused Pol II seems to provide a means of achieving a rapid, and perhaps synchronous, activation of gene expression[38]. The paused Pol II has already progressed through multiple processes that can be slow and stochastic, and a transcriptional activator, acting on a preloaded paused Pol II, allows a rapid transition into productive elongation. Genes with paused Pol II seem not to be in a completely transcriptionally 'off' state[12]. Therefore, regulation of pausing may sacrifice tight control of RNA production in favour of the uniform and rapid response of a gene. The heat-shock genes are a classic example of this regulation: their rapid induction seems critical in responding to a stress that is normally lethal (Box 2). Other stress-response genes, such as those responsible for DNA-damage, unfolded-protein and immune-response pathways, are also enriched in paused Pol II[6,12]. In the early embryo, narrow bands of cells must respond rapidly and uniformly to developmental signals, and genes that respond to these signals are also highly enriched in paused Pol II at the developmental stage at which they must be turned on[7].

A wish list for future approaches

Although many powerful methodologies have been developed for investigating mechanisms of gene regulation *in vivo*, there follows a wish list of key tools and approaches for the future. This list is not meant to be comprehensive, and the approaches described benefit both from the interplay with *in vitro* studies, which provide critical tests of mechanisms and quantification of binding and rate constants for factor interactions, and from structural studies, which provide insight into the precise molecular architectures of proteins and larger macromolecular complexes.

First in the list are protein–DNA crosslinking technologies (for molecular imaging), which produce snapshots of transcription-factor interactions on specific genes *in vivo* and thereby set critical limits in evaluating models of transcriptional regulation. This approach is especially powerful when applied at high temporal and spatial resolution to track the recruitment kinetics and location of specific factors on specific genes during a time course of their activation[13, 39]. These crosslinking methods can be used at individual genes, as well as at the whole genome level. As the resolution of these assays improves, so too does the power of this data in evaluating mechanistic models of transcriptional regulation. Ultimately, the utility of these assays would benefit from the development of this technology to allow mapping of contacts at single-nucleotide resolution and at sufficient kinetic resolution to resolve known major steps in the transcription cycle (Fig. 2), as well as steps yet to be discovered.

Second are highly sensitive microscopy methods, which should provide a strong complement to biochemical methods for examining protein–DNA interactions by allowing observation of the recruitment and dynamics of proteins in real time. The tracking of factors during a time course of the rapid and synchronous activation of a regulated gene will be greatly enhanced when microscopic imaging technology is sufficiently sensitive to examine the recruitment and dynamics of individual proteins on a single chromatid *in vivo*. Tracking proteins at specific loci is now possible on polytene chromosomes or in diploid cells, where genes are amplified in tandem, but single-chromatid tracking of Pol II and particular transcription factors would offer a comprehensive and ordered view of the process and provide the detail that is often masked in measurements that rely on averaging events at many gene copies in a single cell or biochemical measurements of genes in a population of cells.

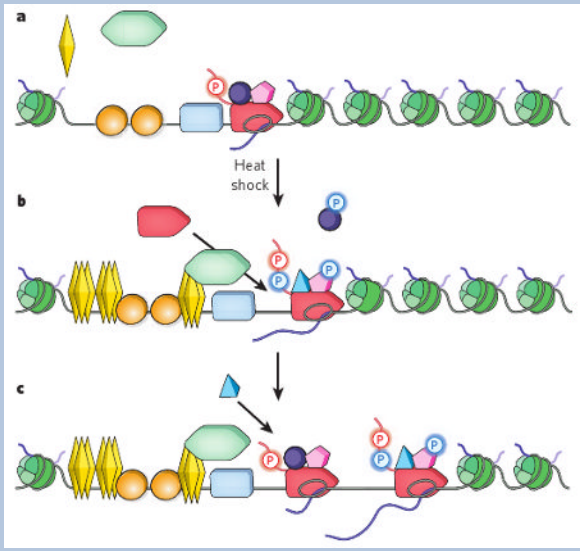
Third are methodologies that evaluate the catalytic and modification state of the key proteins, which, along with the tracking of protein– DNA interactions *in vivo*, are also critical. The antibodies that detect the phosphorylation status of Pol II have been crucial in assessing the activity state of Pol II at various positions along a gene and during the time course of gene activation. Additional antibodies, or other detection reagents, that can evaluate the modification status of transcription factors could certainly provide valuable insights into the way in which different modifications influence each other, and how the final modification code influences the mechanisms of activation. Ultimately, the development of highly effective chromatin purification schemes and highly sensitive mass spectrometry should allow the examination of the complete range of proteins and protein modifications in a particular region and under any condition. There has already been some success in such an examination of a repetitive region of the genome[40], and taking this to the level of specific genes would be extremely powerful.

Box 2

The *Drosophila Hsp70* gene is regulated at the pause-escape step

Drosophila Hsp70 was one of the first genes discovered to have promoter-proximal paused Pol II, and has been extensively studied. As a result, *Hsp70* has served as the model for genes regulated at the step of early elongation. Its promoter resides in a nucleosome-free region extending to about 250 bases downstream of the TSS[36,62] (see figure, panel **a**). This open promoter is bound by GAGA factor (GAF, orange circles) and GTFs (blue rectangle) [63]. Studies have suggested that the GAGA elements are crucial for setting up the paused Pol II (red rocket)[64–66], which is partially phosphorylated (red P). And *in vitro* evidence suggests that GAF bound to the promoter can recruit nucleosome remodellers to maintain this nucleosome-free state[67]. This open promoter allows Pol II to initiate and transcribe 20–40 bases downstream of the TSS, where it is held paused. This pausing is, at least partially, mediated by the SPT4–SPT5 complex (pink pentagon) and the NELF complex (purple circle). *In vivo*, NELF is present on uninduced *Hsp70*, and it is still present, but at lower levels, after heat shock[68] (see figure, panel **c**). Furthermore, NELF depletion *in vivo* reduces

the amount of engaged Pol II on uninduced *Hsp70* ([ref. 69]). Additionally, the downstream sequence may also be important for pausing. When the sequence within 30 bases downstream of the *Hsp70* TSS is switched with the sequence from another gene, the amount of pausing markedly decreases[64]. This may indicate that either the factors binding to downstream elements or the intrinsic pause-inducing characteristics of the transcribed sequence, or both, have a role in pausing. The paused Pol II is phosphorylated by the TFIIH subunit CDK7 on Ser 5 of the CTD repeats. This phosphorylation may be involved in pausing. A temperature-sensitive mutant of CDK7 decreases the amount of paused polymerase on *Hsp70* at non-permissive temperatures[70]; whether this affects pausing directly or at an earlier step remains to be resolved.



Heat shock (see figure, panel **b**) causes the transcriptional activator HSF (yellow diamonds) to trimerize and stably bind upstream of *Hsp70* ([ref. 71]). Such a temperature shift also activates HSF, resulting in the recruitment of coactivators (green hexagon), a rapid general loss of nucleosome protection across the gene[36] and release of the paused Pol II into productive elongation. Upon heat shock, P-TEFb (blue triangle) is recruited to the gene[13, 72] and phosphorylates (blue P) the CTD, SPT5 and NELF subunits; the NELF complex dissociates from the Pol II complex; and Pol II releases from the pause sites, allowing rapid recruitment of new Pol II to the gene (see figure, panel **c**). Although Pol II still resides in the canonical pause sites under these conditions, it is estimated that the pause is of much shorter duration, with Pol II escaping every 4 s rather than once every 10 min before heat-shock induction[63].

Several studies have demonstrated that P-TEFb is important for releasing the paused polymerase upon induction of *Hsp70*. *In vitro* assays show that P-TEFb relieves the inhibitory effects of SPT4-SPT5 and NELF[73]. Depletion or inhibition of P-TEFb severely reduces *Hsp70* RNA expression[70,74], and P-TEFb inhibition, either before or after heat shock, blocks Pol II escape from the 5′ end of the gene[26]. Additionally, TFIIS is important for Pol II escape from the pause sites through its maintenance of paused Pol II in an elongation-competent state[75]. Depletion of TFIIS impedes the release of Pol II from the pause and reduces the rate of *Hsp70* mRNA production.

The evaluation of Pol II activity state is enhanced by nuclear run-on assays that measure transcriptionally engaged RNA polymerase complexes. RNA polymerases that are in an elongation state or simply associated with DNA can be detected by ChIP assays, whereas only

the former are detected by nuclear run-on assays. An approach called GRO-seq, which uses massively parallel sequencing to measure nascent run-on transcripts, has greatly enhanced the sensitivity of nuclear run-on assays and provides a genome-wide analysis of all transcriptionally engaged polymerases[12]. Fourth in our wish list is further development of the GRO-seq assay and the continued examination of short RNAs[41,42] and RNAs associated with chromatin[43], which may allow various states of elongating Pol II to be distinguished (for example the promoter-proximal paused, arrested, abortively initiating and productively elongating states) and further enhance our understanding of the transcription cycle. The utility of these GRO-seq and derivative assays will be enhanced by the development of strategies that allow single-nucleotide resolution and thereby enable the location of Pol II to be precisely defined relative to sequence elements and particular transcription factors.

Although the mapping of protein–DNA interactions *in vivo* at specific genes is well developed, the determination of protein–protein interactions (which are equally important) is much less so and is the fifth item in our wish list. High-resolution microscopy methods that provide subwavelength resolution, for example fluorescence resonance energy transfer[44] and stochastic optical reconstruction microscopy/photo-activated localization microscopy[45,46], have the resolution to assess whether proteins are separated by tens of nanometres or less and thus evaluate whether these proteins are close enough to be in contact. Other optical techniques such as fluorescence cross-correlation spectroscopy[47] make it possible to assess whether pairs of proteins are part of the same complex. The use of biological amplification provided by polytene chromosomes[47] or tandem polymers of genes[48] has provided a glimpse of the potential of optical methods in viewing transcription at specific genes. With improvements in fluorescent labels and detection methods, these approaches should in principle be applicable to factors associated with specific genes, allowing factor–factor associations to be tracked in real time. In addition, recent studies in *S. cerevisiae* have used photoreactive amino acids to provide detailed protein–protein contacts during initiation *in vivo*[49]. Extending this analysis to other steps and other organisms will allow *in vivo* protein–protein interactions to be examined during the transcription cycle in unprecedented detail.

Last, pronounced augmentation of the optical and molecular imaging described here can be achieved by depleting or inhibiting specific factor interactions. Re-examining (re-imaging) the consequences of such experimental treatments can provide critical tests of proposed mechanisms. Although depleting factors with RNA interference is convenient and can generally be used to disrupt factors, sorting primary effects from secondary effects is difficult. Drugs that target specific transcription-factor kinases have been particularly useful, especially when effects are examined immediately after cells have been treated[23,26]. Ultimately, cell-permeable drugs[26], or RNA-aptamer-based drugs[50] synthesized in cells, that target protein–protein interactions during transcription will be extremely useful for assessing the primary effects of such perturbations.

Outlook

The rapid advancement of techniques in biochemistry and microscopy is providing powerful methods to examine the molecular details of biological processes in living cells. These techniques, when coupled with sophisticated approaches to genetically alter and chemically inhibit transcription factors, will provide a new understanding of the transcription cycle, including more detailed knowledge of the known steps in transcription and perhaps identification of new steps. Analysis of individual genes will continue to reveal important mechanistic information about transcription-factor function. Such studies of single genes will be complemented with genome-wide assays to investigate the generality of discoveries and identify specific mechanisms used by individual and co-regulated genes. The next decade will undoubtedly yield exciting insights into the mechanisms of transcription and its regulation.

Acknowledgments

We thank members of the Lis laboratory for discussions and critical reading of the manuscript. Work in our laboratory is supported by US National Institutes of Health grant GM25232.

References

1. Venter JC, et al. The sequence of the human genome. Science 2001;291:1304–1351. [PubMed: 11181995]
2. Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT. The RNA polymerase II core promoter — the gateway to transcription. Curr Opin Cell Biol 2008;20:253–259. [PubMed: 18436437]
3. Lee TI, et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science 2002;298:799–804. [PubMed: 12399584]
4. Saunders A, Core LJ, Lis JT. Breaking barriers to transcription elongation. Nature Rev Mol Cell Biol 2006;7:557–567. [PubMed: 16936696]
5. Venters BJ, Pugh BF. A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. Genome Res 2009;19:360–371. [PubMed: 19124666] This paper examines the genome-wide distribution of Pol II in yeast and suggests that yeast have Pol II enrichment on the 5′ ends of many genes.
6. Muse GW, et al. RNA polymerase is poised for activation across the genome. Nature Genet 2007;39:1507–1511. [PubMed: 17994021] This paper examines the genome-wide distribution of Pol II in *Drosophila* S2 cells and demonstrates that there is Pol II enrichment on many promoters in a stalled or paused state.
7. Zeitlinger J, et al. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. Nature Genet 2007;39:1512–1516. [PubMed: 17994019] This paper examines the genome-wide distribution of Pol II in *Drosophila* embryos and suggests that promoter-proximal stalling/pausing occurs at many developmental genes.
8. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. Cell 2007;130:77–88. [PubMed: 17632057] This paper examines the promoter architecture in human cells and suggests that there is Pol II enrichment on promoters and that Pol II seems to have initiated transcription at most genes.
9. Ptashne M, Gann A. Transcriptional activation by recruitment. Nature 1997;386:569–577. [PubMed: 9121580] This paper reviews the evidence for regulation by Pol II recruitment in bacteria and yeast.
10. Mirkovitch J, Darnell J Jr. Mapping of RNA polymerase on mammalian genes in cells and nuclei. Mol Biol Cell 1992;3:1085–1094. [PubMed: 1384813]
11. Rougvie AE, Lis JT. The RNA polymerase II molecule at the 5′ end of the uninduced *hsp70* gene of *D. melanogaster* is transcriptionally engaged. Cell 1988;54:795–804. [PubMed: 3136931]
12. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 2008;322:1845–1848. [PubMed: 19056941] This paper examines the transcriptionally engaged polymerase across the genome in human cells, and demonstrates a 5′-end enrichment in engaged Pol II at many genes.
13. Boehm AK, Saunders A, Werner J, Lis JT. Transcription factor and polymerase recruitment, modification, and movement on *dhsp70 in vivo* in the minutes following heat shock. Mol Cell Biol 2003;23:7628–7637. [PubMed: 14560008]
14. Peterson CL, Workman JL. Promoter targeting and chromatin remodeling by the SWI/SNF complex. Curr Opin Genet Dev 2000;10:187–192. [PubMed: 10753786]
15. Larschan E, Winston F. The *S. cerevisiae* SAGA complex functions *in vivo* as a coactivator for transcriptional activation by Gal4. Genes Dev 2001;15:1946–1956. [PubMed: 11485989]
16. Ahn SH, Kim M, Buratowski S. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. Mol Cell 2004;13:67–76. [PubMed: 14731395]
17. Stargell LA, Struhl K. Mechanisms of transcriptional activation *in vivo*: two steps forward. Trends Genet 1996;12:311–315. [PubMed: 8783941]
18. Esnault C, et al. Mediator-dependent recruitment of TFIIH modules in preinitiation complex. Mol Cell 2008;31:337–346. [PubMed: 18691966]

19. Kim YJ, Bjorklund S, Li Y, Sayre MH, Kornberg RD. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell 1994;77:599–608. [PubMed: 8187178]
20. Vermeulen M, et al. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell 2007;131:58–69. [PubMed: 17884155]
21. Jacobson RH, Ladurner AG, King DS, Tjian R. Structure and function of a human TAFII250 double bromodomain module. Science 2000;288:1422–1425. [PubMed: 10827952]
22. Sakurai H, Fukasawa T. Functional connections between mediator components and general transcription factors of *Saccharomyces cerevisiae*. J Biol Chem 2000;275:37251–37256. [PubMed: 10973956]
23. Liu Y, et al. Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. Mol Cell Biol 2004;24:1721–1735. [PubMed: 14749387]
24. Spilianakis C, et al. CIITA regulates transcription onset via Ser5-phosphorylation of RNA Pol II. EMBO J 2003;22:5125–5136. [PubMed: 14517250]
25. Core LJ, Lis JT. Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science 2008;319:1791–1792. [PubMed: 18369138]
26. Ni Z, et al. P-TEFb is critical for the maturation of RNA polymerase II into productive elongation *in vivo*. Mol Cell Biol 2008;28:1161–1170. [PubMed: 18070927]
27. Chao SH, Price DH. Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription *in vivo*. J Biol Chem 2001;276:31793–31799. [PubMed: 11431468]
28. Eberhardy SR, Farnham PJ. Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the *cad* promoter. J Biol Chem 2002;277:40156–40162. [PubMed: 12177005]
29. Wittmann BM, Fujinaga K, Deng H, Ogba N, Montano MM. The breast cell growth inhibitor, estrogen down regulated gene 1, modulates a novel functional interaction between estrogen receptor α and transcriptional elongation factor cyclin T1. Oncogene 2005;24:5576–5588. [PubMed: 15940264]
30. Nowak DE, et al. RelA Ser276 phosphorylation is required for activation of a subset of NF-κB-dependent genes by recruiting cyclin-dependent kinase 9/cyclin T1 complexes. Mol Cell Biol 2008;28:3623–3638. [PubMed: 18362169]
31. Peterlin BM, Price DH. Controlling the elongation phase of transcription with P-TEFb. Mol Cell 2006;23:297–305. [PubMed: 16885020]
32. Mavrich TN, et al. Nucleosome organization in the *Drosophila* genome. Nature 2008;453:358–362. [PubMed: 18408708]
33. Moore MJ, Proudfoot NJ. Pre-mRNA processing reaches back to transcription and ahead to translation. Cell 2009;136:688–700. [PubMed: 19239889]
34. Yudkovsky N, Ranish JA, Hahn S. A transcription reinitiation intermediate that is stabilized by activator. Nature 2000;408:225–229. [PubMed: 11089979]
35. Blau J, et al. Three functional classes of transcriptional activation domain. Mol Cell Biol 1996;16:2044–2055. [PubMed: 8628270]
36. Petesch SJ, Lis JT. Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at *Hsp70* loci. Cell 2008;134:74–84. [PubMed: 18614012]
37. Herschlag D, Johnson FB. Synergism in transcriptional activation: a kinetic view. Genes Dev 1993;7:173–179. [PubMed: 8436289] The paper proposes a role for kinetic synergism in the mechanism of transcriptional activation.
38. Boettiger AN, Levine M. Synchronous and stochastic patterns of gene activation in the *Drosophila* embryo. Science 2009;325:471–473. [PubMed: 19628867]
39. Mason PB, Struhl K. Distinction and relationship between elongation rate and processivity of RNA polymerase II *in vivo*. Mol Cell 2005;17:831–840. [PubMed: 15780939]
40. Dejardin J, Kingston RE. Purification of proteins associated with specific genomic loci. Cell 2009;136:175–186. [PubMed: 19135898]
41. Seila AC, et al. Divergent transcription from active promoters. Science 2008;322:1849–1851. [PubMed: 19056940]
42. Preker P, et al. RNA exosome depletion reveals transcription upstream of active human promoters. Science 2008;322:1851–1854. [PubMed: 19056938]

43. Wuarin J, Schibler U. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. Mol Cell Biol 1994;14:7219–7225. [PubMed: 7523861]
44. Bhaumik SR, Raha T, Aiello DP, Green MR. *In vivo* target of a transcriptional activator revealed by fluorescence resonance energy transfer. Genes Dev 2004;18:333–343. [PubMed: 14871930]
45. Patterson GH, Lippincott-Schwartz J. Selective photolabeling of proteins using photoactivatable GFP. Methods 2004;32:445–450. [PubMed: 15003607]
46. Huang B, Wang W, Bates M, Zhuang X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. Science 2008;319:810–813. [PubMed: 18174397]
47. Yao J, Munson KM, Webb WW, Lis JT. Dynamics of heat shock factor association with native gene loci in living cells. Nature 2006;442:1050–1053. [PubMed: 16929308]
48. Janicki SM, et al. From silencing to gene expression: real-time analysis in single cells. Cell 2004;116:683–698. [PubMed: 15006351]
49. Chen HT, Warfield L, Hahn S. The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. Nature Struct Mol Biol 2007;14:696–703. [PubMed: 17632521]
50. Shi H, Hoffman BE, Lis JT. RNA aptamers as effective protein antagonists in a multicellular organism. Proc Natl Acad Sci USA 1999;96:10033–10038. [PubMed: 10468557]
51. Boeger H, Griesenbeck J, Strattan JS, Kornberg RD. Nucleosomes unfold completely at a transcriptionally active promoter. Mol Cell 2003;11:1587–1598. [PubMed: 12820971]
52. Svaren J, Hörz W. Transcription factors vs nucleosomes: regulation of the *PH05* promoter in yeast. Trends Biochem Sci 1997;22:93–97. [PubMed: 9066259]
53. O'Neill EM, Kaffman A, Jolly ER, O'Shea EK. Regulation of PHO4 nuclear localization by the PHO80–PHO85 cyclin-CDK complex. Science 1996;271:209–212. [PubMed: 8539622]
54. Barbaric S, Munsterkotter M, Goding C, Horz W. Cooperative Pho2-Pho4 interactions at the *PHO5* promoter are critical for binding of Pho4 to UASp1 and for efficient transactivation by Pho4 at UASp2. Mol Cell Biol 1998;18:2629–2639. [PubMed: 9566882]
55. Svaren J, Schmitz J, Horz W. The transactivation domain of Pho4 is required for nucleosome disruption at the *PHO5* promoter. EMBO J 1994;13:4856–4862. [PubMed: 7957055]
56. Reinke H, Hörz W. Histones are first hyperacetylated and then lose contact with the activated *PHO5* promoter. Mol Cell 2003;11:1599–1607. [PubMed: 12820972]
57. Steger DJ, Haswell ES, Miller AL, Wente SR, O'Shea EK. Regulation of chromatin remodeling by inositol polyphosphates. Science 2003;299:114–116. [PubMed: 12434012]
58. Korber P, et al. The histone chaperone Asf1 increases the rate of histone eviction at the yeast *PHO5* and *PHO8* promoters. J Biol Chem 2006;281:5539–5545. [PubMed: 16407267]
59. Adkins MW, Howar SR, Tyler JK. Chromatin disassembly mediated by the histone chaperone Asf1 is essential for transcriptional activation of the yeast *PHO5* and *PHO8* genes. Mol Cell 2004;14:657–666. [PubMed: 15175160]
60. Barbaric S, et al. Redundancy of chromatin remodeling pathways for the induction of the yeast *PHO5* promoter *in vivo*. J Biol Chem 2007;282:27610–27621. [PubMed: 17631505]
61. Lam FH, Steger DJ, O'Shea EK. Chromatin decouples promoter threshold from dynamic range. Nature 2008;453:246–250. [PubMed: 18418379]
62. Wu C. The 5′ ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. Nature 1980;286:854–860. [PubMed: 6774262]
63. Lis J. Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation. Cold Spring Harb Symp Quant Biol 1998;63:347–356. [PubMed: 10384299]
64. Lee H, Kraus KW, Wolfner MF, Lis JT. DNA sequence requirements for generating paused polymerase at the start of *hsp70*. Genes Dev 1992;6:284–295. [PubMed: 1737619]
65. Wang YV, Tang H, Gilmour DS. Identification *in vivo* of different rate-limiting steps associated with transcriptional activators in the presence and absence of a GAGA element. Mol Cell Biol 2005;25:3543–3552. [PubMed: 15831460]
66. Lee C, et al. NELF and GAGA factor are linked to promoter-proximal pausing at many genes in *Drosophila*. Mol Cell Biol 2008;28:3290–3300. [PubMed: 18332113]
67. Tsukiyama T, Becker PB, Wu C. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. Nature 1994;367:525–532. [PubMed: 8107823]

68. Wu CH, et al. Molecular characterization of *Drosophila* NELF. Nucleic Acids Res 2005;33:1269–1279. [PubMed: 15741180]
69. Wu CH, et al. NELF and DSIF cause promoter proximal pausing on the *hsp70* promoter in *Drosophila*. Genes Dev 2003;17:1402–1414. [PubMed: 12782658]
70. Schwartz BE, Larochelle S, Suter B, Lis JT. Cdk7 is required for full activation of *Drosophila* heat shock genes and RNA polymerase II phosphorylation *in vivo*. Mol Cell Biol 2003;23:6876–6886. [PubMed: 12972606]
71. Ardehali MB, et al. Spt6 enhances the elongation rate of RNA polymerase II *in vivo*. EMBO J 2009;28:1067–1077. [PubMed: 19279664]
72. Lis JT, Mason P, Peng J, Price DH, Werner J. P-TEFb kinase recruitment and function at heat shock loci. Genes Dev 2000;14:792–803. [PubMed: 10766736]
73. Renner DB, Yamaguchi Y, Wada T, Handa H, Price DH. A highly purified RNA polymerase II elongation control system. J Biol Chem 2001;276:42601–42609. [PubMed: 11553615]
74. Ni Z, Schwartz BE, Werner J, Suarez JR, Lis JT. Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. Mol Cell 2004;13:55–65. [PubMed: 14731394]
75. Adelman K, et al. Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. Mol Cell 2005;17:103–112. [PubMed: 15629721]

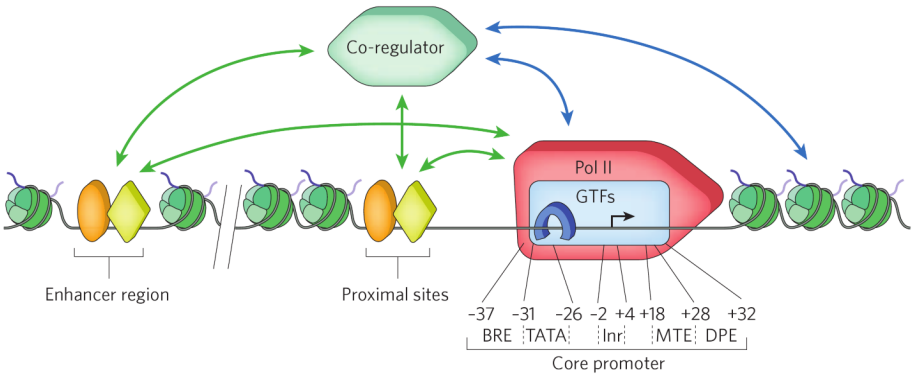


Figure 1. Transcription regulatory interactions

General transcription factors (GTFs) bind to specific sequence elements in the promoter. These elements (the B recognition element (BRE), the TATA box (TATA), the initiator (Inr), the motif ten element (MTE) and the downstream promoter element (DPE)) and their approximate locations relative to the transcription start site (TSS, black arrow) are shown[2]. Transcriptional regulators (orange oval and yellow diamond), which are either activators or repressors, bind to specific DNA sequences located near the core promoter of the gene or various distant regions, called enhancers. The regulators can interact (green arrows) with GTFs, such as TFIID (blue rectangle) and TATA-binding protein (TBP, blue horseshoe), and the Pol II complex (red 'rocket') to enhance or repress transcription. They also interact (green arrows) with co-regulators (green hexagon) that can interact (blue arrows) with the general transcription machinery or chromatin-modifying factors, such as histone modifiers or nucleosome remodellers. The co-regulators can also bind to nucleosomes (green) with various histone modifications, stabilizing the co-regulator binding to the gene. Activators can recruit, stabilize or stimulate these factors, and repressors can disrupt or inhibit these factors.

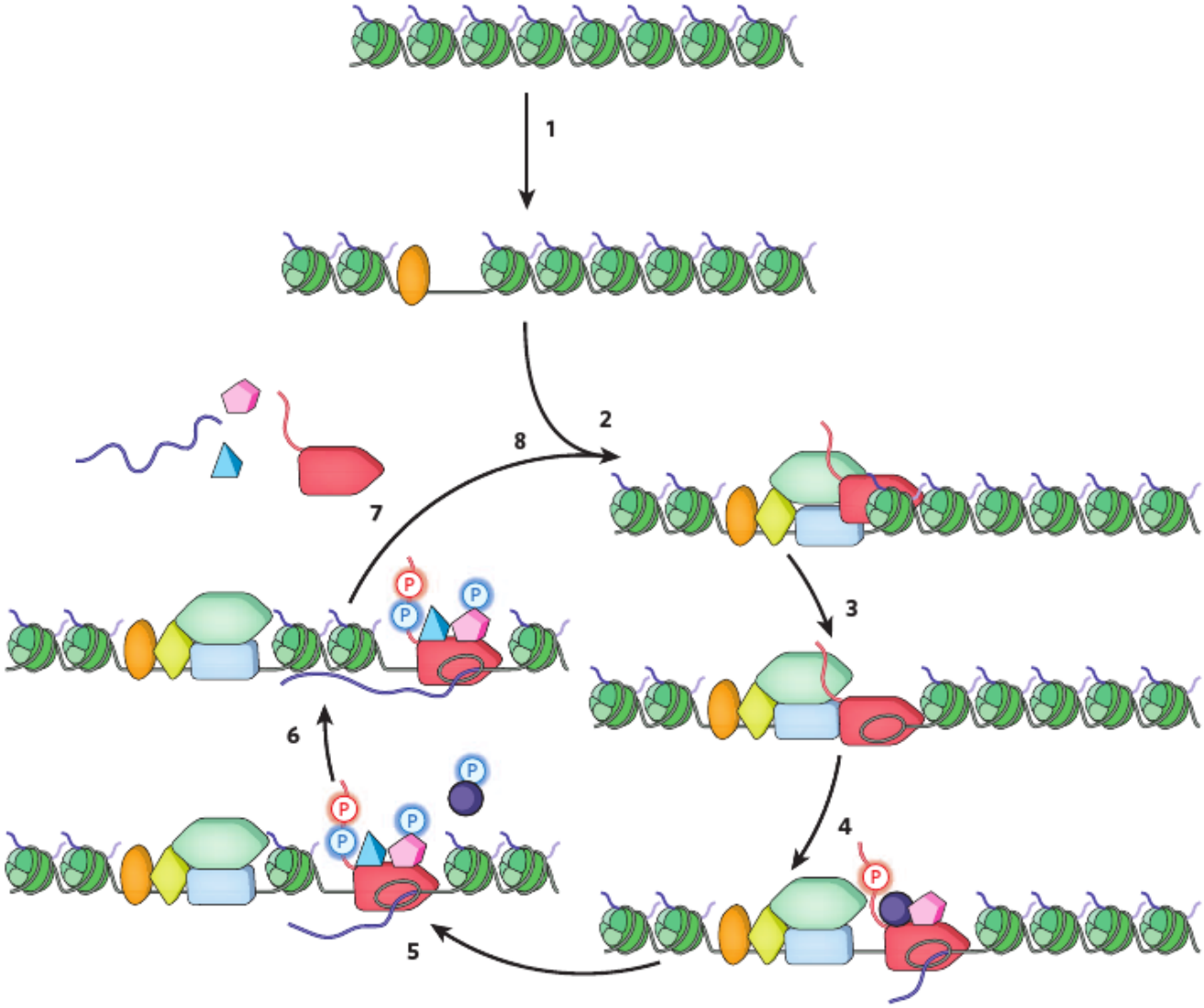


Figure 2. The transcription cycle is a multistep process

Step 1: chromatin opening. The repressed gene and regulatory region are entirely packaged as nucleosomes (green). An activator (orange oval) binds and recruits nucleosome remodellers to clear the promoter. Step 2: PIC formation. A second activator (yellow diamond) binds, promotes the binding of GTFs (blue rectangle) and recruits coactivators (green hexagon), facilitating Pol II (red rocket) entry to the PIC. Step 3: initiation. DNA is unwound (oval inside Pol II) at the TSS, and an open complex is formed. Step 4: promoter escape/clearance. Pol II breaks contacts with promoter-bound factors, transcribes 20–50 bases downstream of the TSS, produces an RNA (purple line) and pauses, partially mediated by SPT4–SPT5 in *Drosophila* (pink pentagon) and negative elongation factor (NELF) complex (purple circle). The Ser residues at position 5 (Ser 5) of the Pol II carboxy-terminal domain (CTD) repeats are phosphorylated (red P) during this step. Step 5: escape from pausing. P-TEFb (blue triangle) is recruited directly or indirectly by the activator and phosphorylates Ser 2 of the Pol II CTD repeats, SPT5 and the NELF subunits (blue Ps). NELF dissociates from the rest of the complex. Pol II escapes from the pause, either terminating or entering productive elongation. Step 6: productive elongation. Nucleosomes are disassembled and reassembled as the Pol II elongation

complex transcribes through the gene. Step 7: termination. After the Pol II complex transcribes the gene, it is removed from the DNA, and the RNA is released. Step 8: recycling. The freed Pol II can reinitiate.