The RNA polymerase II preinitiation complex

Through what pathway is the complex assembled?

Donal S Luse

Department of Molecular Genetics; Lerner Research Institute; Cleveland Clinic; Cleveland, OH USA

Keywords: RNA polymerase II, preinitiation complex, general transcription factors, TBP, TFIIB, TFIIF, TFIIE, TFIIH

The general transcription factors required for the assembly of the RNA polymerase II preinitiation complex at TATA-dependent promoters are well known. However, recent studies point to two quite distinct pathways for assembly of these components into functional transcription complexes. In this review, the two pathways are compared and potential implications for gene regulatory mechanisms are discussed.

Introduction

Our understanding of the assembly of the RNA polymerase II (RNAP II) transcription complex is based on studies in test tube systems. The general transcription factors (GTFs) required to form the RNAP II preinitiation complex (PIC) and the pathway through which the PIC assembles are thought to be well-understood.¹⁻⁴ This view is called into question by a number of developments. First, two quite distinct pathways for preinitiation complex formation have recently been described.^{5,6} Second, it is now appreciated that pausing by RNAP II at about 50 bp downstream of transcription start is a major point for metazoan gene regulation.⁷⁻⁹ The mechanistic basis for this process is not fully understood, and in particular it is not known how the loading of factors at the PIC stage affects pausing downstream. Third, when the search for GTFs was begun it was thought that eukaryotic promoters were always anchored by a TATA sequence. The canonical TATA box has been identified in only a small minority of mammalian promoters.¹⁰⁻¹³ GTF requirements for these non-TATA promoters have not been determined and the pathway(s) used for PIC assembly at such promoters have not been explored. However, one recent study suggests a previously unrecognized simplicity to RNAP II promoter structure.¹⁴ The application of in vitro approaches to a broader range of promoters will be crucial in characterizing the generality of the RNAP II GTFs and RNAP II promoter architecture.

What Are the GTFs and What Roles Do They Play at TATA Box Promoters?

In order to assemble a RNAP II PIC competent to initiate transcription on a fully double-stranded template, the following minimal set of factors is required: TBP (TATA box binding protein), TFIIB, TFIIF, TFIIE and TFIIH. This is true for both mammalian⁶ and yeast RNAP II,⁵ but it should be stressed that these factors have all been identified and characterized using promoters that are clearly TATA-dependent. This core GTF set has been extensively reviewed,¹⁻⁴ so the present review will focus on important functional points. The TATA box binding protein, TBP, normally occurs as a subunit of the multisubunit TFIID factor complex, but TBP alone is sufficient for PIC assembly. As its name implies, TBP marks the promoter through its interaction with the TATA box, which in mammalian TATA-containing genes is located about 30 bp upstream of transcription start.¹² When TBP binds the TATA element it bends the DNA at roughly a right angle.¹⁵ At some promoters, DNA recognition is assisted by a second factor, TFIIB, which can bind to DNA elements (BREs) adjacent to the TATA box.^{16,17} TFIIB also interacts with TBP and crucially, one TFIIB domain interacts extensively with RNAP II, reaching into the catalytic center of the polymerase.¹⁸ It has been suggested that TFIIB is partially responsible for selecting the transcription start site¹⁹⁻²² (but see also²³).

The other factor with a clear role at the beginning of transcription is TFIIH, the only GTF with enzymatic activity. TFIIH contains two helicases, XPB (Ssl2p in yeast) and XPD (Rad3p in yeast), as well as a kinase, Cdk7 (Kin28p in yeast). Unlike other DNA-dependent RNA polymerases, RNAP II cannot drive separation of the template and non-template strands. In the absence of superhelical tension in the template, RNAP II must rely on the XPB helicase subunit of TFIIH to melt the promoter DNA.²⁴ For templates in which the DNA will spontaneously unwind, or on artificial DNAs where the template and non-template strands are not complementary near the transcription start site (bubble templates), TFIIH is not required for transcription.^{25,26} TFIIH's location in the PIC is not absolutely clear but there is a consensus that it interacts with the DNA downstream of the rest of the complex.^{6,27,28}

Correspondence to: Donal S Luse; Email: lused@ccf.org

Submitted: 09/26/13; Revised: 11/01/13; Accepted: 11/02/13

Citation: Luse DS. The RNA polymerase II preinitiation complex: Through what pathway is the complex assembled?. *Transcription* 2014; 5; PMID: 24406342; <http://dx.doi.org/10.4161/trns.27050>

Figure 1. Diagrammatic representation of RNAP II PIC assembly pathways for human⁶ and yeast⁵ components as recently reported. In both cases the template is indicated by the black line. The TATA element is located under the TBP symbol and transcription would proceed from left to right. The asterisk denotes the location of the subcomplex of the Rpb4 and Rpb7 subunits of human RNAP II. *For the human pathway:* important associations are indicated but the components are not drawn to scale. In the initial complex (A), DNA downstream of the TATA box does not have a defined location. After addition of TFIIF (B), DNA location is fixed. The addition of TFII E (C) creates an extended interaction interface of TFIIF and TFII E across the face of RNAP II. TFIIF joins the complex in the final step (D). TFIIF interacts primarily with the downstream DNA; it associates with TFII E and RNAP II through a limited interface near RNAP II subunits 4 and 7 (asterisk). *For the yeast system:* an initial complex (E) can be assembled on the template with all components other than RNAP II and TFIIF. This complex can be purified but TFIIB is then lost (indicated by the dashed outline on TFIIB). The complete, transcriptionally competent PIC (F) is assembled by the addition of TFIIB and RNAP II/TFIIF.

TFII E has been considered primarily as a loading factor for TFIIF,²⁹⁻³² but it can have other roles. TFIIF strongly stimulates transcription in assay systems with supercoiled templates that do not require TFIIF, leading to the suggestion that TFIIF facilitates promoter melting.^{33,34} This idea is reinforced by work with the archaeal TFE factor, an analog of one of the subunits of TFIIF, which is also thought to stimulate template strand separation.³⁵ The archaea lack TFIIF; those organisms require only TBP and TFIIB analogs, as well as TFE in some cases, for PIC formation.^{36,37}

TFIIF was first identified as a RNAP II binding factor³⁸ and it has been suggested that one of its functions is to facilitate loading of RNAP II into the PIC. Several lines of evidence point to an important role for TFIIF in stabilizing TFIIB within the PIC, including structural³⁹⁻⁴¹ and functional studies.^{40,42} Experiments with bubble templates have also suggested a role for TFIIF in the initiation step itself,^{43,44} a function which is probably linked to the interaction of TFIIB and TFIIF (reviewed in⁴⁵). Recent results concerning the effect of phosphorylation on TFIIF function complicate interpretation of earlier findings. TFIIF that is modified by casein kinase 2 (CK2) remains fully functional in supporting PIC formation, but phospho-TFIIF is not stably retained within the PIC.⁴² Thus, it is possible to generate a PIC that is fully initiation competent but contains no TFIIF. This point becomes particularly significant when considering pathways for PIC formation and the role of TFIIF in PIC assembly.

TFIIF is not a GTF in the strictest sense, since it is not required to assemble a transcriptionally active complex, either with human or yeast components. However, recent pathway assembly studies include TFIIF.^{5,6} In yeast, TFIIF appears at promoters along with the core GTFs.⁴⁶ TFIIF has been reported to stabilize TBP-DNA

interactions,⁴⁷ to increase the fraction of TBP-bound templates⁴⁸ and to alter the ability of the TFIID complex to interact with the TATA element.⁴⁹

Two Recent Studies Provide Substantially Different Viewpoints on the Pathway Through Which GTFs and RNAP II Join the Promoter to Assemble a PIC

He et al.⁶ used human RNAP II, human core GTFs, TFIIF and an optimized TATA box promoter⁵⁰ to study PIC formation and structure. Complexes were assembled stepwise from highly purified or recombinant components and examined primarily by cryo-electron microscopy. The results were interpreted using the known crystal structures of the individual components and subcomplexes. A diagrammatic representation of models generated in this study is shown in Figure 1. The first structure examined contained template bound with only TBP, TFIIF, TFIIB and RNAP II (complex A in Fig. 1). Additional complexes were assembled by adding TFIIF (complex B) followed by TFII E (complex C) and finally TFII H (complex D), which is the “conventional” assembly order for a RNAP II PIC. Interestingly, DNA was not visible in the initial complex (A), apparently because the template does not adopt a single position in that case. This may reflect the fact that in this complex, RNAP II only contacts a single domain of TFIIB and does not make any other GTF contacts. The addition of TFIIF resulted in a substantial change to the nascent human PIC, with DNA now locked into a single position (B). This suggests that a major function of TFIIF is to “remodel” the growing PIC in an essential way. Consistent with earlier crosslinking studies with yeast components,²⁸ TFII E was observed to bind across a broad face of RNAP II, spanning

the central catalytic cleft (C). With the addition of TFIIE, the combination of TFIIF and TFIIE provides a continuous band of protein extending across the template and central catalytic cleft to the Rpb4/7 subcomplex (complex C; asterisk). This TFIIF/TFIIE superstructure could be important in confining DNA to the catalytic cleft. TFIIH completes the complex by interacting primarily with the downstream DNA. In the model of the complete human PIC structure, the only contact between TFIIH and RNAP II, and between TFIIE and TFIIH, occurs near the Rpb4/7 subcomplex (D).

Several of the findings in this study are noteworthy, particularly in the context of proposed roles for the individual GTFs. The complexes were assembled using TFIIA. The authors did not comment specifically about any requirement for TFIIA in their system, but its use would be consistent with the reported role of TFIIA in stabilizing TBP on the template.^{47,48} RNAP II joined the human complex without TFIIF, indicating that TFIIF is not absolutely required as a chaperone to load RNAP II into the nascent PIC. The structural results suggest instead that a major function of TFIIF is to reorganize the growing PIC in an essential way, setting the final path of the template. However, as noted above, other recent studies indicate that once TFIIF has acted, it can leave the PIC without significant effect.⁴² This would indicate that any reorientation of components driven by TFIIF should be stable in the absence of that factor. He et al. propose that the C-terminal section of the larger subunit of TFIIF has an important functional role in the fully assembled PIC, but it is not clear how this can be reconciled with the ability of PICs to lose TFIIF without activity loss, or the fact that TFIIF which lacks the C-terminal domain of its larger subunit continues to function in PIC formation.^{51,52} The close association of TFIIH in the human PIC with the subcomplex of polymerase subunits 4 and 7 provides a potential explanation for the apparent requirement for Rpb4/7 specifically in the assembly of an active RNAP II PIC (Rpb4 and Rpb7 are not needed for RNAP II to elongate a nascent RNA⁵³). The proposed locations of TFIIE and TFIIH in the human PIC do not obviously predict that TFIIE should be important in TFIIH loading, as suggested in earlier studies.²⁹⁻³² Indeed, He et al. note that TFIIE is less stable within the complex prior to TFIIH addition.

He et al. did not report either the fraction of their complexes which can melt the template in response to ATP or the fraction of those open complexes which can synthesize RNA.⁶ This leaves open the possibility that only a small fraction of the PICs might be transcriptionally active. The active subset would not necessarily have the same structure as the major species shown in the models. This concern is based on the frequent observation that only a small minority of RNAP II PICs assembled *in vitro* are actually able to initiate transcription (see, for example,^{5,42}). Assessing the transcriptional competence for PICs can be challenging, since complexes may fail prior to run-off through multiple mechanisms. Depending on the promoter, mammalian elongation complexes containing relatively short RNAs are sufficiently stable so that after halting for several minutes, all of the complexes can be extended to run-off. For example, one study reported that human RNAP II complexes with nascent RNAs as

short as 7 nt remained fully active for subsequent elongation.⁵⁴ Determining the proportion of the human PICs that can support synthesis of such minimally-elongated complexes should provide the best estimate of the transcriptional competence of these complexes. Synthesis of very short RNAs (3–4 mers) can proceed by a multi-round abortive mechanism⁵⁵⁻⁵⁷ which would overstate the transcriptional contribution of a subpopulation of complexes.

Murakami et al.⁵ have recently described a PIC assembled using yeast RNAP II, yeast GTFs (TBP, TFIIB, TFIIF, TFIIE and TFIIH, as well as TFIIA) and a natural yeast promoter. Their goal was to obtain a nearly homogeneous PIC with recombinant or highly purified factors. They observed initially that PICs assembled with concentrated components tend to precipitate. However, a surprising intermediate containing TBP, TFIIA, TFIIB, TFIIE and TFIIH could be assembled on the template at high salt in the *absence* of polymerase (complex E in Fig. 1). After dialysis and the addition of RNAP II and TFIIF, the resulting complex was transcriptionally active (complex F). Further investigation of complex E showed that it was stable to purification except that TFIIB was not retained in that case. The authors noted that while TFIIA was present in their assembly assays it had no effect on the ability of the PICs to synthesize RNA.

All of the yeast PICs could melt their templates in response to ATP. The set of long (~100–140 nt) run-off transcripts generated in single-round assays with the PICs match the collection of initiation sites used *in vivo* from the promoter in question.⁵ However, only 4 to 17.5% of the purified, apparently homogeneous PICs could support run-off RNA synthesis. This does not necessarily reflect the presence of inactive complexes in the preparations. As just noted, it is necessary to examine all of the RNAs made by PICs, including shorter transcripts, in order to evaluate transcriptional competence. Determining the fraction of yeast PICs which can advance as far as the earliest stable elongation complex could be difficult, since initiation by yeast RNAP II may involve repeated abortive synthesis of short transcripts as part of a scanning mechanism through which the polymerase ultimately locates the transcription start site.^{23,40,44} Murakami et al. noted that reactions in which all components were simply mixed at stoichiometric levels yielded much lower levels of transcript generated per template.⁵ This could be relevant in comparison with the human PICs, which were assembled step-wise on attached templates and then released by restriction digest but were not further purified.⁶

While this review was being prepared for publication, the report on the three-dimensional structure of the yeast PIC became available.⁵⁸ The PICs used in the structural study differed from those described in the initial report⁵ by the addition of one factor, TFIIS. Yeast TFIIS has been reported to be part of yeast PICs assembled from extracts.⁵⁹ The authors did not comment on the rationale for adding TFIIS to the PICs nor did they report what effect, if any, TFIIS has on the transcriptional capabilities of the PICs. The detailed structure of the yeast PIC, determined by cryo-electron microscopy along with a combination of chemical crosslinking and mass spectroscopy,⁵⁸ is consistent with the schematic representation in Figure 1E and F.

The results of the yeast studies raise a number of interesting and perplexing questions, particularly in contrast to the findings with the human RNAP II PIC.⁶ The central difference from the human assembly pathway, and indeed from conventional ideas about assembly, is that almost the entire assembly process with the yeast factors takes place on the template without RNAP II (see complex **E**). Murakami et al.^{5,58} argue that efficient loading of RNAP II into a transcriptionally competent PIC is only possible through the complex **E** intermediate. TFIIF could be serving as a chaperone for RNAP II in the yeast system, although Murakami et al. did not report whether RNAP II could enter the intermediate complex prior to TFIIF. If TFIIF functions as a remodeler of the template within the PIC, as proposed by He et al.,⁶ TFIIF must be able to produce this effect even when it enters the complex last. In addition, if yeast TFIIF is needed only at assembly and not at initiation, as is the case for human PICs,⁴² the yeast complex **F** should continue to support transcription if TFIIF were lost. An earlier study⁶⁰ suggested that effective assembly of TFIIB at the promoter requires the presence of RNAP II, but in the Murakami et al. approach TFIIB can be recruited into the complex **E** intermediate in the absence of RNAP II. It is somewhat surprising that TFIIE can be a stable component of the yeast complex **E**, given that other studies with both yeast²⁸ and human⁶ components indicate that TFIIE has extensive interactions with one face of the RNA polymerase in the complete PIC. The presence of TFIH in the yeast intermediate complex would support the concept that TFIIE is important for TFIH loading, but it is not consistent with the idea that an interaction interface for TFIIE and TFIH on RNAP II is involved in stabilizing TFIH within the complex (see^{6,28}).

Why Would the Assembly Pathway for the RNAP II PIC Be Significant in the Context of Gene Regulation?

Perhaps most importantly, the stable complex **E** characterized by Murakami et al. is similar to the scaffold complex (TFIID, TFIIA, TFIIE, TFIH and Mediator) proposed by Hahn and colleagues⁶¹ as the intermediate in reinitiation of transcription after the first round of RNA synthesis. Efficient reinitiation, which would presumably be important at genes expressed at high levels, would be favored by a PIC pathway in which all of the factors remain stably attached to the promoter except those that associate most strongly with RNAP II (i.e., TFIIB and TFIIF). This requirement could not be met by the complex **A–D** pathway proposed for human PIC formation. It has been suggested that efficient reinitiation of transcription in yeast involves a direct transfer of newly-terminated RNAP II to the promoter through a looping mechanism which depends on TFIIB.^{62,63} The fact that association of TFIIB with complex **E** is only partially stable would be consistent with this model.

A recent study has documented the existence of pre-PICs at many mammalian RNAP II promoters. These nascent PICs contain RNAP II but lack a transcription bubble because they have not yet loaded TFIH.⁶⁴ The existence of such complexes would not have been predicted from the PIC pathway as described

by Murakami et al.⁵ in which RNAP II joins a TFIH-containing pre-PIC, but it is consistent with the human PIC assembly process in which TFIH loads last.⁶ It has been reported that partial PICs can be assembled in yeast; in that case the nascent PICs included TFIIE but lacked both RNAP II and TFIH.⁶⁵

The PIC assembly pathway is also important because it can potentially dictate the content of the transcription complex downstream of initiation. This is particularly relevant when considering TFIIF and the loading of Gdown1 into the transcription complex. Gdown1 is the 13th subunit of mammalian RNAP II, present in one third to one half of RNAP II isolated from tissue sources.⁶⁶ Gdown1 strongly facilitates pausing by RNAP II *in vitro* when it is added to transcription complexes in the earliest stage of elongation.⁶⁷ However, the pathway through which Gdown1 might enter the complex *in vivo* is problematic because Gdown1 displaces TFIIF from free RNAP II and from transcript elongation complexes.^{67–69} This may be explained by the significant overlap of the surfaces on RNAP II which are involved in the interaction of Gdown1 or TFIIF with free polymerase.⁶⁹ Since TFIIF is required to assemble a PIC, Gdown1 should be able to inhibit RNAP II transcription at the initiation stage. Results supporting this prediction have been obtained using supercoiled plasmid templates that do not require the full set of GTFs for transcription.⁶⁸ One solution to this apparent paradox may be found in the observation, noted above, that CK2-phosphorylated TFIIF is not retained in the PIC after complex assembly.⁴² The resulting TFIIF-less complexes should be immediately available to load Gdown1. CK2 does co-localize to at least a subset of promoters *in vivo*.⁷⁰

Modification of TFIIF provides one plausible pathway to accommodate both PIC assembly and the introduction of Gdown1 early in transcript elongation. Alternatively, Mediator might be involved in reconciling TFIIF's presence in the PIC with the apparently higher affinity of Gdown1 for RNAP II. Gdown1 was originally described as a factor which renders transcription Mediator-dependent in certain assays.⁶⁶ In addition, the orientation of RNAP II when in complex with Mediator is determined by TFIIF.⁷¹

Important Cautions: Non-TATA Promoters and Non-Canonical Initiation Mechanisms

The factor requirements and assembly pathways for PIC formation have only been studied for a handful of clearly TATA-dependent RNAP II promoters. Many promoters have consensus elements in addition to, or instead of, TATA (reviewed in^{11,13,72}), including the initiator element (INR) surrounding the start site of transcription and various other elements located at roughly 15 to 35 bp downstream of transcription start. These last sequences are thought to be recognized by the non-TBP subunits of TFIID, the TAFs (see⁴⁹ and references therein). It is not necessarily the case that promoters with defined downstream elements simply require a substitution of TFIID for TBP to optimize transcription. For example, it has been reported that phosphorylation of TFIID is important in controlling recognition of promoters with a certain class of downstream elements.⁷⁰ It is also important to note that

TBP is a member of a small family of related transcription factors; some of these factors also recognize TATA elements while others do not (see for example⁷³; TBP-like factors are reviewed in⁷⁴).

The most abundant promoters in mammalian cells reside within the so-called CG islands.^{11,75} These sequences appear to lack any of the well-characterized promoter elements. The minimal GTF set for these promoters has not been established and thus the PIC assembly pathway(s) have not been determined. One important distinction from the well-studied TATA-dependent promoters is that the TATA-less promoters can direct initiation to a number of well-separated downstream locations,¹¹ in contrast to the narrow window (28–33 bp from TATA) observed for mammalian TATA promoters.¹² TATA-less promoters often appear to lack directionality; transcription at these elements can either proceed into the gene or in an antisense direction.^{76,77} Transcript elongation is only efficient in the direction of the gene but the basis for this difference has not been studied in vitro.

Venters and Pugh have recently provided a novel view of RNAP II promoter architecture, based on the locations of TFIIB binding in the genome of K562 cells as revealed by the ChIP-exo approach.¹⁴ These authors propose that the large majority of the sites occupied by TFIIB occur within a consensus promoter sequence that includes a TBP-bound TATA box flanked by BRE elements, with an INR element downstream at the distance expected from consensus TATA box promoters. A single RNAP II promoter architecture provides a potentially powerful unification of promoter classes that have generally been considered distinct, in particular TATA-containing vs. non-TATA promoters. Interpretation of the ChIP-exo results should be considered with some caution. For example, the assignment of each of the promoter elements allowed for as many as three mismatches with consensus, resulting in an estimated false positive rate of about 20%.¹⁴ Nevertheless, this study seems to establish TBP and TFIIB as universal RNAP II GTFs. It suggests that it may be possible to understand all mammalian RNAP II PIC structure and initiation mechanism using the known GTFs and the constraints of their assembly at TATA box promoters. Anchoring PICs at the position of TFIIB binding provides important insight into bidirectional transcription, particularly for CG islands. Opposing transcription complexes were typically found 70–180 bp apart and occupancy at these sites was not correlated, indicating that divergent transcription probably does not arise from a single promoter element that can support transcription in either direction.¹⁴ This is consistent with the results in an earlier report, in which four CG islands that support bidirectional transcription were studied in detail.⁷⁸

Promoters in budding yeast seem to combine features of metazoan promoters containing canonical TATA boxes and metazoan promoters that support multiple transcription start sites. A recent study in which the structure of yeast PICs was explored by the ChIP-exo method indicated that most *S. cerevisiae* RNAP II promoters do contain a functional TATA element, although this sequence may have significant mismatches with the TATA consensus.⁴⁶ The canonical set of GTFs all appeared to be localized at these TATA and TATA-like elements, as one would expect for a metazoan TATA promoter.⁴⁶ However, the

location of yeast initiation sites relative to the TATA elements can be quite variable, from 40 to 120 bases downstream of TATA.^{40,79} For example, in the well-defined yeast in vitro system of Murakami et al.,⁵ transcription initiates at about 30, 60 and 70 bp downstream of the TATA element of the *HIS4* promoter. It has been suggested that yeast RNAP II scans downstream to select one of several potential transcription start sites (see^{23,40,44,79} for more detailed discussion).

Challenges for the Future

The recent results on the structure and the genomic distribution of PICs hold out the promise of significant new insights into fundamental transcription mechanisms and the control of gene expression at the transcriptional level. Transcript initiation at yeast promoters does not follow the same rules regarding the spacing of transcription start and the TATA element as seen with mammalian TATA box promoters. This could indicate a basic mechanistic difference in the yeast and mammalian RNAP II transcription machinery. This difference could be reflected in the differences in the yeast and human PIC assembly pathways and PIC structures.^{5,6,58} Does the proposed simplification of mammalian promoter structure from the ChIP-exo study¹⁴ allow at least a partial unification of these apparently different mechanisms?

As noted above, the purified yeast PICs can unwind their templates and at least a subset can begin transcription at the same three start sites used at *HIS4* in vivo.⁵ These initiation sites are distributed over 40 bases, with the promoter-proximal site at the typical metazoan distance of ~30 bases downstream from TATA box. The yeast PIC provides the opportunity to test different models for multiple initiation sites. One scanning mechanism for initiation involves upstream start sites which are used abortively before productive transcription begins from downstream sites. Are short RNAs corresponding to upstream abortive starts actually observed with the yeast PICs? As an alternative approach, it may be possible to follow the dimensions of the transcription bubble after ATP addition to the yeast PICs. Does the melted region appear over the entire range of potential start sites simultaneously or does it first appear only over the upstream initiation site? Can γ -thio-ATP (which inhibits the XPB helicase of TFIIF⁸⁰) be used to trap initiations at only the most upstream location?

The results of the recent ChIP-exo study suggest that the large majority of mammalian RNAP II promoters conform to a basic consensus with a TATA box and an INR-anchored start site about 30 bases downstream of TATA.¹⁴ However, the authors note that about a quarter of mRNA 5' ends map much farther downstream of the TATA sequence.¹⁴ This raises the possibility that for at least a subset of mammalian promoters, PIC assembly would be based on a single TATA element but multiple start sites would be possible, some much more than 30 bases downstream of TATA. Promoters of this type could be functional analogs of yeast TATA promoters. As individual mammalian promoters are examined from the ChIP-exo study set, it will be particularly important to identify and characterize any single TATA/multi-start promoters. The questions posed above for yeast TATA promoters concerning

dimensions of the transcription bubble and the possibility of abortive scanning as a precursor to downstream initiation could all be posed for these mammalian promoters as well. It will also be important to further characterize promoters from the ChIP-exo study whose TATA elements deviate significantly (i.e., three mismatches) from consensus. Can these promoters support transcription *in vitro* with only TBP and the remaining GTFs, or must TBP be part of TFIID for these promoters to function? If this proves to be the case, are such promoters enriched in downstream promoter elements expected to be recognized by the TAF subunits of TFIID⁴⁹?

Single-start TATA promoters that confine initiation to a narrow downstream range (as seen with mammalian consensus TATA promoters) and multi-start TATA promoters that are not constrained in TATA to +1 distance (as seen with yeast and possibly with some mammalian promoters) provide a striking mechanistic contrast. The He et al. model of the mammalian PIC proposes a pathway for template melting in which DNA is driven into the confined interior of the PIC by the action of the XPB helicase.⁶ A limited ability to accommodate DNA within the PIC would be consistent with a limited range of functional distances between the TATA box (which fixes the upstream edge of the bubble⁵⁴) and the start site of transcription. A limited capacity to accommodate the growing bubble as transcript elongation begins

could also dictate the location of the “bubble collapse” transition that is linked to promoter clearance for mammalian RNAP II.⁵⁴

It is not possible for the earliest stages of transcription in the yeast system to conform exactly to the mammalian model, if constraints within the mammalian PIC structure prevent initiation more than about 30 bases from the TATA element. In this context it is interesting that the yeast PIC structure features a clear open space between the template-bound GTFs on the one side and RNAP II on the other,⁵⁸ in contrast to the mammalian PIC.⁶ One can envision that in the yeast structure, room is available to extrude long stretches of single-stranded DNA to accommodate far-downstream initiation sites. If collapse of the initial transcription bubble drives PIC disassembly and promoter clearance in the yeast system, then presumably bubble collapse could only be linked to productive initiation events and not to any abortive upstream initiations during scanning. As a corollary, if the more “open” PIC structure seen with yeast components is a requirement for supporting far-downstream initiations, then any mammalian TATA-based PICs that support initiation much further than 30 bases downstream from a single TATA element may also adopt a structure similar to the yeast PIC.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

- Orphanides G, Lagrange T, Reinberg D. The general transcription factors of RNA polymerase II. *Genes Dev* 1996; 10:2657-83; PMID:8946909; <http://dx.doi.org/10.1101/gad.10.21.2657>
- Roeder RG. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* 1996; 21:327-35; PMID:8870495
- Hahn S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* 2004; 11:394-403; PMID:15114340; <http://dx.doi.org/10.1038/nsmb763>
- Vannini A, Cramer P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell* 2012; 45:439-46; PMID:22365827; <http://dx.doi.org/10.1016/j.molcel.2012.01.023>
- Murakami K, Calero G, Brown CR, Liu X, Davis RE, Boeger H, Kornberg RD. Formation and fate of a complete 31-protein RNA polymerase II transcription preinitiation complex. *J Biol Chem* 2013; 288:6325-32; PMID:23303183; <http://dx.doi.org/10.1074/jbc.M112.433623>
- He Y, Fang J, Taatjes DJ, Nogales E. Structural visualization of key steps in human transcription initiation. *Nature* 2013; 495:481-6; PMID:23446344; <http://dx.doi.org/10.1038/nature11991>
- Nechaev S, Adelman K. Promoter-proximal Pol II: when stalling speeds things up. *Cell Cycle* 2008; 7:1539-44; PMID:18469524; <http://dx.doi.org/10.4161/cc.7.11.6006>
- Gilmour DS. Promoter proximal pausing on genes in metazoans. *Chromosoma* 2009; 118:1-10; PMID:18830703; <http://dx.doi.org/10.1007/s00412-008-0182-4>
- Zhou Q, Li T, Price DH. RNA polymerase II elongation control. *Annu Rev Biochem* 2012; 81:119-43; PMID:22404626; <http://dx.doi.org/10.1146/annurev-biochem-052610-095910>
- Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y, Hume DA. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet* 2007; 8:424-36; PMID:17486122; <http://dx.doi.org/10.1038/nrg2026>
- Juven-Gershon T, Hsu J-Y, Theisen JWM, Kadonaga JT. The RNA polymerase II core promoter - the gateway to transcription. *Curr Opin Cell Biol* 2008; 20:253-9; PMID:18436437; <http://dx.doi.org/10.1016/j.ccb.2008.03.003>
- Ponjavic J, Lenhard B, Kai C, Kawai J, Carninci P, Hayashizaki Y, Sandelin A. Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters. *Genome Biol* 2006; 7:R78; PMID:16916456; <http://dx.doi.org/10.1186/gb-2006-7-8-r78>
- Dikstein R. The unexpected traits associated with core promoter elements. *Transcription* 2011; 2:201-6; PMID:22231114; <http://dx.doi.org/10.4161/trns.2.5.17271>
- Venters BJ, Pugh BF. Genomic organization of human transcription initiation complexes. *Nature* 2013; 502:53-8; PMID:24048476
- Kim JL, Nikolov DB, Burley SK. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* 1993; 365:520-7; PMID:8413605; <http://dx.doi.org/10.1038/365520a0>
- Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev* 1998; 12:34-44; PMID:9420329; <http://dx.doi.org/10.1101/gad.12.1.34>
- Deng W, Roberts SGE. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes Dev* 2005; 19:2418-23; PMID:16230532; <http://dx.doi.org/10.1101/gad.342405>
- Sainsbury S, Niesser J, Cramer P. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* 2013; 493:437-40; PMID:23151482; <http://dx.doi.org/10.1038/nature11715>
- Pinto I, Wu W-H, Na JG, Hampsey M. Characterization of *suaf* mutations defines a domain of TFIIB involved in transcription start site selection in yeast. *J Biol Chem* 1994; 269:30569-73; PMID:7982976
- Pardee TS, Bangur CS, Ponticelli AS. The N-terminal region of yeast TFIIB contains two adjacent functional domains involved in stable RNA polymerase II binding and transcription start site selection. *J Biol Chem* 1998; 273:17859-64; PMID:9651390; <http://dx.doi.org/10.1074/jbc.273.28.17859>
- Hawkes NA, Roberts SG. The role of human TFIIB in transcription start site selection *in vitro* and *in vivo*. *J Biol Chem* 1999; 274:14337-43; PMID:10318856; <http://dx.doi.org/10.1074/jbc.274.20.14337>
- Cho EJ, Buratowski S. Evidence that transcription factor IIB is required for a post-assembly step in transcription initiation. *J Biol Chem* 1999; 274:25807-13; PMID:10464320; <http://dx.doi.org/10.1074/jbc.274.36.25807>
- Yang C, Ponticelli AS. Evidence that RNA polymerase II and not TFIIB is responsible for the difference in transcription initiation patterns between *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Nucleic Acids Res* 2012; 40:6495-507; PMID:22510268; <http://dx.doi.org/10.1093/nar/gks323>
- Tirole F, Busso D, Coin F, Egly JM. Reconstitution of the transcription factor TFIIB: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol Cell* 1999; 3:87-95; PMID:10024882; [http://dx.doi.org/10.1016/S1097-2765\(00\)80177-X](http://dx.doi.org/10.1016/S1097-2765(00)80177-X)
- Parvin JD, Sharp PA. DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell* 1993; 73:533-40; PMID:8490964; [http://dx.doi.org/10.1016/0092-8674\(93\)90140-L](http://dx.doi.org/10.1016/0092-8674(93)90140-L)
- Pan G, Greenblatt J. Initiation of transcription by RNA polymerase II is limited by melting of the promoter DNA in the region immediately upstream of the initiation site. *J Biol Chem* 1994; 269:30101-4; PMID:7982911

27. Kim TK, Ebright RH, Reinberg D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* 2000; 288:1418-22; PMID:10827951; <http://dx.doi.org/10.1126/science.288.5470.1418>
28. Grünberg S, Warfield L, Hahn S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat Struct Mol Biol* 2012; 19:788-96; PMID:22751016; <http://dx.doi.org/10.1038/nsmb.2334>
29. Li Y, Flanagan PM, Tschochner H, Kornberg RD. RNA polymerase II initiation factor interactions and transcription start site selection. *Science* 1994; 263:805-7; PMID:8303296; <http://dx.doi.org/10.1126/science.8303296>
30. Maxon ME, Goodrich JA, Tjian R. Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIF: a model for promoter clearance. *Genes Dev* 1994; 8:515-24; PMID:7926747; <http://dx.doi.org/10.1101/gad.8.5.515>
31. Ohkuma Y, Hashimoto S, Wang CK, Horikoshi M, Roeder RG. Analysis of the role of TFIIE in basal transcription and TFIIF-mediated carboxy-terminal domain phosphorylation through structure-function studies of TFIIE- α . *Mol Cell Biol* 1995; 15:4856-66; PMID:7651404
32. Bushnell DA, Bamdad C, Kornberg RD. A minimal set of RNA polymerase II transcription protein interactions. *J Biol Chem* 1996; 271:20170-4; PMID:8702741; <http://dx.doi.org/10.1074/jbc.271.33.20170>
33. Timmers HTM. Transcription initiation by RNA polymerase II does not require hydrolysis of the beta-gamma phosphoanhydride bond of ATP. *EMBO J* 1994; 13:391-9; PMID:8313884
34. Holstege FCP, Tantin D, Carey M, van der Vliet PC, Timmers HTM. The requirement for the basal transcription factor IIE is determined by the helical stability of promoter DNA. *EMBO J* 1995; 14:810-9; PMID:7882984
35. Grohmann D, Nagy J, Chakraborty A, Klose D, Fielden D, Ebright RH, Michaelis J, Werner F. The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol Cell* 2011; 43:263-74; PMID:21777815; <http://dx.doi.org/10.1016/j.molcel.2011.05.030>
36. De Carlo S, Lin SC, Taatjes DJ, Hoenger A. Molecular basis of transcription initiation in Archaea. *Transcription* 2010; 1:103-11; PMID:21326901; <http://dx.doi.org/10.4161/trns.1.2.13189>
37. Werner F, Weinzierl ROJ. Direct modulation of RNA polymerase core functions by basal transcription factors. *Mol Cell Biol* 2005; 25:8344-55; PMID:16135821; <http://dx.doi.org/10.1128/MCB.25.18.8344-8355.2005>
38. Sotma M, Carthew RW, Greenblatt J. Isolation of three proteins that bind to mammalian RNA polymerase II. *J Biol Chem* 1985; 260:10353-60; PMID:3860504
39. Eichner J, Chen HT, Warfield L, Hahn S. Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *EMBO J* 2010; 29:706-16; PMID:20033062; <http://dx.doi.org/10.1038/emboj.2009.386>
40. Fishburn J, Hahn S. Architecture of the yeast RNA polymerase II open complex and regulation of activity by TFIIF. *Mol Cell Biol* 2012; 32:12-25; PMID:22025674; <http://dx.doi.org/10.1128/MCB.06242-11>
41. Chen ZA, Jawhari A, Fischer L, Buchen C, Tahir S, Kaminski T, Rasmussen M, Larivière L, Bukowski-Wills JC, Nilges M, et al. Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J* 2010; 29:717-26; PMID:20094031; <http://dx.doi.org/10.1038/emboj.2009.401>
42. Čabart P, Újvári A, Pal M, Luse DS. TFIIF is not required for initiation by RNA polymerase II but it is essential to stabilize TFIIB in early transcription complexes. *Proc Natl Acad Sci USA* 2011; 108:15786-91; PMID:21896726; <http://dx.doi.org/10.1073/pnas.1104591108>
43. Yan Q, Moreland RJ, Conaway JW, Conaway RC. Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. *J Biol Chem* 1999; 274:35668-75; PMID:10585446; <http://dx.doi.org/10.1074/jbc.274.50.35668>
44. Khapersky DA, Ammerman ML, Majovski RC, Ponticelli AS. Functions of *Saccharomyces cerevisiae* TFIIF during transcription start site utilization. *Mol Cell Biol* 2008; 28:3757-66; PMID:18362165; <http://dx.doi.org/10.1128/MCB.02272-07>
45. Luse DS. Rethinking the role of TFIIF in transcript initiation by RNA polymerase II. *Transcription* 2012; 3:156-9; PMID:22771986; <http://dx.doi.org/10.4161/trns.20725>
46. Rhee HS, Pugh BF. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* 2012; 483:295-301; PMID:22258509; <http://dx.doi.org/10.1038/nature10799>
47. Hieb AR, Halsey WA, Betterton MD, Perkins TT, Kugel JF, Goodrich JA. TFIIF changes the conformation of the DNA in TBP/TATA complexes and increases their kinetic stability. *J Mol Biol* 2007; 372:619-32; PMID:17681538; <http://dx.doi.org/10.1016/j.jmb.2007.06.061>
48. Blair RH, Goodrich JA, Kugel JF. Single-molecule fluorescence resonance energy transfer shows uniformity in TATA binding protein-induced DNA bending and heterogeneity in bending kinetics. *Biochemistry* 2012; 51:7444-55; PMID:22934924; <http://dx.doi.org/10.1021/bi300491j>
49. Cianfrocco MA, Kassavetis GA, Grob P, Fang J, Juven-Gershon T, Kadonaga JT, Nogales E. Human TFIID binds to core promoter DNA in a reorganized structural state. *Cell* 2013; 152:120-31; PMID:23332750; <http://dx.doi.org/10.1016/j.cell.2012.12.005>
50. Juven-Gershon T, Cheng S, Kadonaga JT. Rational design of a super core promoter that enhances gene expression. *Nat Methods* 2006; 3:917-22; PMID:17124735; <http://dx.doi.org/10.1038/nmeth937>
51. Lei L, Ren D, Finkelstein A, Burton ZF. Functions of the N- and C-terminal domains of human RAP74 in transcriptional initiation, elongation, and recycling of RNA polymerase II. *Mol Cell Biol* 1998; 18:2130-42; PMID:9528785
52. Újvári A, Pal M, Luse DS. The functions of TFIIF during initiation and transcript elongation are differentially affected by phosphorylation by casein kinase 2. *J Biol Chem* 2011; 286:23160-7; PMID:21566144; <http://dx.doi.org/10.1074/jbc.M110.205658>
53. Edwards AM, Kane CM, Young RA, Kornberg RD. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter *in vitro*. *J Biol Chem* 1991; 266:71-5; PMID:1985924
54. Pal M, Ponticelli AS, Luse DS. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol Cell* 2005; 19:101-10; PMID:15989968; <http://dx.doi.org/10.1016/j.molcel.2005.05.024>
55. Luse DS, Jacob GA. Abortive initiation by RNA polymerase II *in vitro* at the adenovirus 2 major late promoter. *J Biol Chem* 1987; 262:14990-7; PMID:3667620
56. Holstege FCP, Fiedler U, Timmers HTM. Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J* 1997; 16:7468-80; PMID:9405375; <http://dx.doi.org/10.1093/emboj/16.24.7468>
57. Pal M, Luse DS. Strong natural pausing by RNA polymerase II within 10 bases of transcription start may result in repeated slippage and reextension of the nascent RNA. *Mol Cell Biol* 2002; 22:30-40; PMID:11739720; <http://dx.doi.org/10.1128/MCB.22.1.30-40.2002>
58. Murakami K, Elmlund H, Kalisman N, Bushnell DA, Adams CM, Azubel M, Elmlund D, Levi-Kalishman Y, Liu X, Gibbons BJ, et al. Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* 2013; 342:1238724; PMID:24072820; <http://dx.doi.org/10.1126/science.1238724>
59. Kim B, Nesvizhskii AI, Rani PG, Hahn S, Aebersold R, Ranish JA. The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proc Natl Acad Sci U S A* 2007; 104:16068-73; PMID:17913884; <http://dx.doi.org/10.1073/pnas.0704573104>
60. Elsbey LM, O'Donnell AJ, Green LM, Sharrocks AD, Roberts SG. Assembly of transcription factor IIB at a promoter *in vivo* requires contact with RNA polymerase II. *EMBO Rep* 2006; 7:898-903; PMID:16878124; <http://dx.doi.org/10.1038/sj.embor.7400767>
61. Yudkovsky N, Ranish JA, Hahn S. A transcription reinitiation intermediate that is stabilized by activator. *Nature* 2000; 408:225-9; PMID:11089979; <http://dx.doi.org/10.1038/35041603>
62. Singh BN, Hampsey M. A transcription-independent role for TFIIB in gene looping. *Mol Cell* 2007; 27:806-16; PMID:17803944; <http://dx.doi.org/10.1016/j.molcel.2007.07.013>
63. El Kaderi B, Medler S, Raghunayakula S, Ansari A. Gene looping is conferred by activator-dependent interaction of transcription initiation and termination machineries. *J Biol Chem* 2009; 284:25015-25; PMID:19602510; <http://dx.doi.org/10.1074/jbc.M109.007948>
64. Kouzine F, Wojtowicz D, Yamane A, Resch W, Kieffer-Kwon KR, Bandle R, Nelson S, Nakahashi H, Awasthi P, Feigenbaum L, et al. Global regulation of promoter melting in naive lymphocytes. *Cell* 2013; 153:988-99; PMID:23706737; <http://dx.doi.org/10.1016/j.cell.2013.04.033>
65. Zanton SJ, Pugh BF. Full and partial genome-wide assembly and disassembly of the yeast transcription machinery in response to heat shock. *Genes Dev* 2006; 20:2250-65; PMID:16912275; <http://dx.doi.org/10.1101/gad.1437506>
66. Hu X, Malik S, Negroiu CC, Hubbard K, Velalar CN, Hampton B, Grosu D, Catalano J, Roeder RG, Gnatt A. A Mediator-responsive form of metazoan RNA polymerase II. *Proc Natl Acad Sci U S A* 2006; 103:9506-11; PMID:16769904; <http://dx.doi.org/10.1073/pnas.0603702103>
67. Cheng B, Li T, Rahl PB, Adamson TE, Loudas NB, Guo J, Varzavand K, Cooper JJ, Hu X, Gnatt A, et al. Functional association of Gdown1 with RNA polymerase II poised on human genes. *Mol Cell* 2012; 45:38-50; PMID:22244331; <http://dx.doi.org/10.1016/j.molcel.2011.10.022>
68. Jishage M, Malik S, Wagner U, Uberheide B, Ishihama Y, Hu X, Chait BT, Gnatt A, Ren B, Roeder RG. Transcriptional regulation by Pol II(G) involving mediator and competitive interactions of Gdown1 and TFIIF with Pol II. *Mol Cell* 2012; 45:51-63; PMID:22244332; <http://dx.doi.org/10.1016/j.molcel.2011.12.014>
69. Wu YM, Chang JW, Wang CH, Lin YC, Wu PL, Huang SH, Chang CC, Hu X, Gnatt A, Chang WH. Regulation of mammalian transcription by Gdown1 through a novel steric crosstalk revealed by cryo-EM. *EMBO J* 2012; 31:3575-87; PMID:22850672; <http://dx.doi.org/10.1038/emboj.2012.205>

70. Lewis BA, Sims RJ 3rd, Lane WS, Reinberg D. Functional characterization of core promoter elements: DPE-specific transcription requires the protein kinase CK2 and the PC4 coactivator. *Mol Cell* 2005; 18:471-81; PMID:15893730; <http://dx.doi.org/10.1016/j.molcel.2005.04.005>
71. Bernecky C, Grob P, Ebmeier CC, Nogales E, Taatjes DJ. Molecular architecture of the human Mediator-RNA polymerase II-TFIIF assembly. *PLoS Biol* 2011; 9:e1000603; PMID:21468301; <http://dx.doi.org/10.1371/journal.pbio.1000603>
72. Valen E, Sandelin A. Genomic and chromatin signals underlying transcription start-site selection. *Trends Genet* 2011; 27:475-85; PMID:21924514; <http://dx.doi.org/10.1016/j.tig.2011.08.001>
73. Isogai Y, Keles S, Prestel M, Hochheimer A, Tjian R. Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev* 2007; 21:2936-49; PMID:17978101; <http://dx.doi.org/10.1101/gad.1608807>
74. Müller F, Demény MA, Tora L. New problems in RNA polymerase II transcription initiation: matching the diversity of core promoters with a variety of promoter recognition factors. *J Biol Chem* 2007; 282:14685-9; PMID:17395580; <http://dx.doi.org/10.1074/jbc.R700012200>
75. Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes Dev* 2011; 25:1010-22; PMID:21576262; <http://dx.doi.org/10.1101/gad.2037511>
76. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 2008; 322:1845-8; PMID:19056941; <http://dx.doi.org/10.1126/science.1162228>
77. Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA. Divergent transcription from active promoters. *Science* 2008; 322:1849-51; <http://dx.doi.org/10.1126/science.1162253>; PMID:19056940
78. Flynn RA, Almada AE, Zamudio JR, Sharp PA. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc Natl Acad Sci U S A* 2011; 108:10460-5; PMID:21670248; <http://dx.doi.org/10.1073/pnas.1106630108>
79. Kuehner JN, Brow DA. Quantitative analysis of in vivo initiator selection by yeast RNA polymerase II supports a scanning model. *J Biol Chem* 2006; 281:14119-28; PMID:16571719; <http://dx.doi.org/10.1074/jbc.M601937200>
80. Dvir A, Conaway RC, Conaway JW. Promoter escape by RNA polymerase II. A role for an ATP cofactor in suppression of arrest by polymerase at promoter-proximal sites. *J Biol Chem* 1996; 271:23352-6; PMID:8798537; <http://dx.doi.org/10.1074/jbc.271.38.23352>