A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex

Gail Miller[1] and Steven Hahn[1]

[1]Division of Basic Sciences, Fred Hutchinson Cancer Research Center,

1100 Fairview Ave N

Seattle, Washington 98109, USA.

Email: shahn@fhcrc.org, Telephone: 206 667 5261, Fax: 206 667 6497

Running Title: Path of promoter DNA in the PIC

Key Words: RNA Pol II, preinitiation complex, hydroxyl radical, general transcription factor

To directly map the position of promoter DNA within the RNA Polymerase II (Pol II) transcription Preinitiation Complex (PIC), FeBABE was tethered to specific sites within the *HIS4* promoter and used to map exposed surfaces of Pol II and the general transcription factors in proximity to DNA. Our results distinguish between previously proposed models for PIC structure and demonstrate that downstream promoter DNA is positioned over the central cleft of Pol II with DNA upstream of TATA extending toward the Pol II subunit Rpb3. Also mapped were segments of TFIIB, IIE, IIF, and IIH in proximity to promoter DNA. DNA downstream of the transcription bubble maps to a path between the two helicase subdomains of the TFIIH subunit Rad25 (XPB). Together, our results show how the general factors and Pol II converge on promoter DNA within the PIC.

Transcription of protein-coding genes by Pol II requires the recruitment of Pol II and the general transcription factors (GTFs) to promoters to form a PIC[1-3], analogous to the Closed Complex state of bacterial RNAP. Prior to transcription initiation, the PIC transitions into the Open Complex state, where ATP and the TFIIH helicase subunit XPB promote melting of DNA surrounding the transcription start site and positioning of the template strand of the promoter within the active site cleft of Pol II[1,4-7]. In metazoans, ~12 base pairs (bp) of DNA are melted in the Open Complex, and transcription typically initiates at a single start site ~30-bp downstream from the TATA element. However, transcription start site selection in *S. cerevisiae* is more complex since initiation occurs between 40-120-bp from TATA[8]. Despite these differences in start site position, the upstream edge of the transcription bubble maps to ~20-bp downstream of TATA both in metazoan systems and *S. cerevisiae*[7,9-11]. These observations led to the suggestion that yeast Pol II "scans" downstream DNA for a suitable initiation site[9].

Much insight into the mechanism of Pol II transcription has been gained from crystal structures of yeast Pol II[12-16]. Structural studies have shown that for elongating Pol II, the template DNA strand lies inside the deep central cleft located between Rpb1 and Rpb2,

the two largest subunits of Pol II[5,16]. However, the mechanism of how Pol II assembles with the GTFs and promoter DNA in the PIC remains unclear.

Protein-DNA cross-linking studies have provided important insights into PIC structure at the adenovirus major late promoter (AdMLP)[17-22]. These studies showed that the transcription machinery makes extensive interactions with promoter DNA in minimal PICs assembled from purified Pol II and GTFs. Cross-linking performed after treatment with the detergent Sarkosyl[19,20] revealed that Rpb1 and Rpb2, the two largest subunits of Pol II, interact with promoter DNA over a 60-bp region. TFIIB and the small subunit of TFIIF (RAP30) both cross-link to DNA upstream and downstream of TATA whereas the large subunit of TFIIF (RAP74) cross-links to DNA downstream of TATA[19,20]. These studies agree with results obtained using yeast Pol II and a subset of the GTFs at the AdMLP[22]. Both subunits of human TFIIE cross-link to the promoter region overlapping the transcription bubble region, and the human TFIIH subunit XPB cross-links to DNA downstream of the transcription bubble region[19]. However, more extensive cross-links with human TFIIE, TFIIF and TFIIH were detected in the absence of Sarkosyl[17,18,21]. These observations led to different models for the structure of the PIC, and distinct mechanisms were proposed for promoter melting by the TFIIH helicase[17,19].

More recently, separate studies examining TFIIB interaction with Pol II have provided two additional models of PIC structure[23-25]. X-ray structure analysis of a complex of Pol II and TFIIB[23] showed that the TFIIB B-finger domain projects into the active site cleft and suggested a location for the TFIIB core domain outside of the cleft region near the Pol II dock domain. This led to a model for PIC structure where promoter DNA runs along the outer edge of the Pol II clamp domain below the active site cleft. In a separate study, photo-cross-linking and hydroxyl radical probes were placed on TFIIB and assembled into PICs to map the interaction surfaces between TFIIB and Pol II[24,25]. These studies showed that in the PIC, the TFIIB core domain is in a location considerably different from that predicted from the Pol II-TFIIB structure. These results led to a model for PIC structure in which promoter DNA is positioned directly over the central cleft of Pol II rather than the outer edge of the clamp.

To distinguish between all these models proposed for PIC structure, we used a strategy to directly establish the path of promoter DNA in the *S. cerevisiae* PIC. The Fe-EDTA probe FeBABE was tethered to specific locations on promoter DNA and hydroxyl radical cleavage was used to map proximity and position of both Pol II and the GTFs relative to the promoter. In contrast to previous DNA-protein cross-linking studies, this approach maps the position of amino acids on proteins that are exposed and in proximity to specific positions of promoter DNA. We find that DNA downstream of TATA is positioned over the central cleft of Pol II and that DNA upstream of TATA bends around the back face of Pol II toward the Pol II subunit Rpb3. This approach also located the surfaces of the general factors TFIIB, TFIIE, TFIIF and TFIIH near to DNA allowing us to distinguish between previously proposed models for PIC structure and helicase function. Finally, our data shows that despite differences in transcription start site between yeast and human Pol II, the yeast PIC assembles at a position equivalent to human Pol II with respect to the TATA element.

Results

FeBABE-tethered *HIS4* templates support *in vitro* transcription

To directly probe the path of promoter DNA in the yeast PIC, the hydroxyl radical generating reagent FeBABE[26] was tethered to specific positions on the yeast *HIS4* promoter. FeBABE is ~12 Å in length and has previously been linked to nucleic acids via a phosphorothioate on the nucleic acid backbone[27-29].

We conjugated FeBABE to the *HIS4* promoter between nucleotides –87 and +1, where +1 indicates the most upstream transcription start site. From alignment of TATA elements, this corresponds to promoter positions –55 to +33 of the AdMLP (**Fig. 1a**), and is equivalent to the region of the AdMLP used in previous cross-linking studies, allowing a direct comparison between the two systems. FeBABE was conjugated to both the non-template and template DNA strands in clusters of 5 with an estimated conjugation efficiency of 70-80% (**Fig. 1a**, **Supplementary Fig. 1a**, and data not shown). In preliminary studies, we found that insertion of a single FeBABE at *HIS4* gave very weak

protein cleavage, leading us to construct promoters with multiple FeBABE positions to circumvent this problem. 17 different probes were constructed with FeBABE conjugated to the non-template strand spanning a total of 10 bases (i.e., 1 FeBABE every 2 bases) (**Fig. 1a**). The 17 probes were numbered from –5 to –1 upstream of TATA and from 1 to 11 downstream of TATA. All 17 probes are essentially identical except for the position of the 5 FeBABEs; each probe is 590-bp, contains the *HIS4* promoter (encompassing regions upstream and downstream of the TATA element including the proposed transcription bubble region and the transcription start site), a cluster of 5 FeBABEs, a single upstream Gal4 binding site and is biotinylated at the 5' end to allow coupling to Dynabeads for *in vitro* transcription and for PIC formation/hydroxyl radical cleavage experiments.

To evaluate the ability of Pol II to transcribe RNA from phosphorothioate-modified and FeBABE-conjugated DNA, *in vitro* transcription assays were performed using immobilized promoter templates (**Fig. 1b**). Phosphorothioate-modification of promoters showed equivalent activity when compared to a control promoter containing no phosphorothioate residues. However, a ~50% decrease in transcription was observed when FeBABE was conjugated to probe 11, which was not surprising since this covers the transcription start site. Otherwise, conjugating FeBABE to the promoter had little or no effect on the ability of Pol II to transcribe *HIS4*.

Mapping the domains of Pol II close to promoter DNA in the PIC

Hydroxyl radical cleavage was first used to locate the surfaces of Rpb1 and Rpb2 in close proximity to *HIS4* promoter DNA within the PIC. Nuclear extracts containing a triple Flag-tag on either the C terminus of Rpb1 or Rpb2 were used to form PICs on immobilized FeBABE-conjugated *HIS4* templates. These PICs contain not only Pol II and GTFs but also a single dimer of the Gal4-VP16 activator and co-activators such as SAGA, Mediator and NuA4[30,31]. After PIC assembly, templates were washed to purify the complexes. Cleavage of peptide bonds near the iron center of the tethered FeBABE was activated by the addition of sodium ascorbate and hydrogen peroxide to generate hydroxyl radicals. Because the hydroxyl radicals diffuse to a limited extent under the

conditions used, FeBABE probes for proteins located within ~15-20 Å of the Fe-EDTA center[24-26,32]. Anti-Flag Western blots were used to visualize the Rpb1 and Rpb2 cleavage products (**Figs. 2a, 2b** and **3**). Compared to a control experiment using *HIS4* promoter DNA with no phosphorothioate residues, specific cleavage fragments were observed using FeBABE-conjugated probes. In **Figures 2a** and **b**, additional experiments were carried out with probe 6 in which either FeBABE or hydrogen peroxide was individually omitted.

The sizes of the peptide fragments and their corresponding cleavage sites were calculated using *in vitro*-translated Flag-tagged Rpb1 and Rpb2 peptides as molecular standards[24], and the results are summarized in **Supplementary Figures 2a-f** and **Supplementary Table 1**. Based on the sizes of the peptide fragments, the Rpb1 cleavage sites are located in the Clamp core, Clamp head and Jaw domains[14] (**Figs. 2a** and **2b**). Rpb1 was cleaved by probes 1 to 8, covering an extensive region of promoter DNA downstream of TATA. Cleavage sites of Rpb2 were located in the protrusion, lobe and fork domains[14] (**Fig. 3**). Rpb2 was cleaved by probes -3 and -2 and 3 to 6, spanning DNA both upstream and downstream of TATA.

Hydroxyl radical cleavage experiments of Rpb1 and Rpb2 were also carried out with FeBABE conjugated to the promoter template strand (**Supplementary Figs. 1a** and **1b**). Results were nearly identical to those above except that cleavage was not detected in the protrusion domain of Rpb2 from probes 3 and 4 (**Supplementary Fig. 1b**). The Pol II cleavage sites are PIC-specific. When purified Pol II was subjected to hydroxyl radical cleavage generated by Fe/H_2O_2 in solution, different cleavage sites were observed (data not shown).

Site-specific hydroxyl radical cleavage was also used to probe the proximity of surfaces of the Pol II subunits Rpb3, Rpb4, Rpb5, Rpb7 and Rpb9 to promoter DNA. Of particular interest were Rpb5 and Rpb7 as these subunits had previously been shown to cross-link to the AdMLP[19,22]. However, no cleavage was detected within any of these Pol II subunits suggesting that these subunits are not near DNA in the yeast PIC. An

alternative explanation is that other proteins in the PIC block the access of the hydroxyl radical to these subunits.

Model for the path of promoter DNA in the PIC

The calculated FeBABE cleavage sites for Rpb1 and Rpb2 were mapped on the surface of the X-ray crystal structure of 12 subunit Pol II[12,13] where strong/medium and weak cleavages are highlighted in blue and cyan, respectively (**Fig. 4a**). Four residues on either side of the calculated cleavage sites are highlighted to reflect the precision in measurement of cleavage fragments[24]. Using the cleavage sites of Rpb1 and Rpb2 as a guide, we modeled the path of *HIS4* promoter DNA within the PIC. From the location of the cut sites on the Pol II surface, which are distributed on either side of the central cleft, the promoter DNA downstream of TATA most likely lies outside and runs along the middle of the central cleft as straight B-form DNA (**Fig. 4b**). Rpb2 was also cleaved by probes placed on DNA upstream of TATA. From the location of the cut sites on the Rpb2 protrusion domain, upstream promoter DNA most likely bends around the back face of Pol II near Rpb2 and extends toward Rpb3 (**Fig. 4b**). From this predicted path of the promoter DNA, all cleavage sites of Rpb1 and Rpb2 are within the ~30 Å range of FeBABE conjugated to the phosphorothioate group. The observed FeBABE cleavage pattern is not consistent with other proposed models for PIC structure where DNA is proposed to either wrap around Pol II for almost a complete turn[18,21] or run along the outer edge of the Pol II clamp domain[23]. In mammalian cells, melting of the DNA template occurs at -12-bp (-44, yeast numbering), located 19-bp downstream from TATA. In this model position -12 is located directly over the active site cleft (highlighted red/purple in **Fig. 4**).

The predicted position of promoter DNA is entirely consistent with the model for PIC structure proposed by Chen and Hahn[25], based on mapping the position of the TFIIB core domain proximal to the Pol II wall domain using TFIIB-linked FeBABE. The observed path of promoter DNA requires no deviation from B-form DNA other than the bend imposed by TBP. Also shown on these models are TBP (green ribbon) and TFIIB (yellow ribbon).

It is apparent from mapping the sites of cleavage to the Pol II surface that large regions of Rpb1 and Rpb2 are protected from cleavage. As demonstrated below, promoter probes giving rise to Rpb1 and Rpb2 cleavage are also in proximity to the other general transcription factors. A likely explanation for protection of Pol II from more extensive cleavage is that much of the Pol II surface in this region is not exposed and is protected by interaction with other general transcription factors.

Mapping the Rad25 helicase domains close to promoter DNA

DNA-protein cross-linking studies at the AdMLP have reported the positions of promoter DNA near the GTFs. However, these reported locations of TFIIE, TFIIF and TFIIH remain uncertain since two different cross-linking studies obtained distinct cross-linking patterns. For these reasons, we used FeBABE cleavage to locate the surfaces of TFIIE, TFIIF and TFIIH near promoter DNA. Given that the interactions of TFIIB with promoter DNA, TBP and Pol II have been extensively studied, we first examined the cleavage of TFIIB as a control. FeBABE positioned on either side of TATA (probes –1 and 2 to 4) cleaves TFIIB as expected, in the B-finger, linker and core domains (**Supplementary Figs. 3a, 3b, 4a, 4b** and **4c**; **Supplementary Table 1**). Cut sites within the core domain correspond to the specific contact points between promoter DNA and the core domain observed in the DNA-TBP-TFIIB crystal structure[33,34].

Cleavage experiments were then carried out for the Rad25 subunit of TFIIH. Yeast TFIIH is composed of 10 subunits including two helicases and a kinase[1,3], and helicase activity of the Rad25 subunit (XPB) is essential for open complex formation[6]. The human XPB helicase is the only TFIIH subunit observed to cross-link to promoter DNA and separate cross-linking studies place XPB at different locations on promoter DNA[17,19]. Moreover, these studies could not unambiguously identify the helicase domain as the DNA proximal domain.

Rad25 in PICs was cleaved by probes 6 to 9, the region of DNA directly downstream of the predicted location of the metazoan transcription bubble[7,10,11] and 12-bp downstream

from the proposed upstream site of the yeast transcription bubble[9] (**Fig. 5a** and **Supplementary Fig. 5a**). The sizes of the peptide fragments and their corresponding cleavage sites were calculated as described above, and the results are summarized in **Supplementary Figures 3c** and **3d** and **Supplementary Table 1**. Yeast Rad25 (842 amino acids) is a member of the SF-2 family of DNA helicases and cleavage sites map to the DEVH box helicase and the C terminal helicase subdomains of Rad25 (**Fig. 5b**) showing that in the PIC, the helicase domain is exposed and in proximity to promoter DNA.

Although no structure of Rad25 has yet been obtained, the structure of several other SF-2 family members has been determined[35] including the recent structure of *Archaeoglobus fulgidus* XPB[36]. DNA helicases contain two conserved subdomains, one encompassing the DEX(D/H) box subdomain and the other encompassing a C terminal helicase subdomain, with these two subdomains connected by a linker of variable length. ATP binds between these two subdomains and nucleic acid is observed to lie across the top groove between them. Conformational changes mediated by ATP hydrolysis are thought to induce strain into the bound nucleic acid to promote single strand DNA formation and/or remove secondary structure in single stranded nucleic acids. As described in **Supplementary Methods**, a structure model for the two Rad25 helicase subdomains (which lie between positions 349 and 714) was generated based on sequence alignment of eukaryotic XPB family members with Af XPB (**Supplementary Fig. 6**), the structure coordinates of Af XPB, and positioning of these models in the closed conformation by orientation of the subdomains relative to the HCV RNA helicase containing bound RNA[35,36].

The FeBABE cleavage sites within Rad25 were mapped on the surface of this structure model (**Fig. 5c**) where strong/medium and weak cleavages are highlighted in dark blue and cyan, respectively. From the location of the cleavage sites on either side of the central groove between the two helicase subdomains, the order of the probes generating cleavage, and the location of nucleic acid in other helicases of known structure, we predict that DNA downstream of the transcription bubble is likely to lie along the top of

Rad25 between the two subdomains in the approximate direction shown by the arrow, where the arrowhead indicates more downstream promoter DNA. In this view shown, ATP would bind at the bottom of the structure model between the two subdomains.

Mapping the domains of the TFIIF subunits near promoter DNA

Yeast TFIIF likely binds to Pol II as a heteromer composed of three subunits, two of which, Tfg1 (RAP74) and Tfg2 (RAP30) are conserved among human, insects and yeast[3]. The third yeast subunit Tfg3 is less tightly associated, has no mammalian counterpart, and is also a component of TFIID and the chromatin remodeling complexes RSC and INO80[37]. The N termini of human Tfg1 and Tfg2 subunits form a dimerization domain, and the C termini of both subunits are winged helix domains[38]. Cryo-EM has been used to determine the structure of yeast Pol II in complex with TFIIF[39], where Tfg2 has been observed to interact extensively with the surface of Pol II along the central cleft.

Both Tfg1 and Tfg2 were cleaved from an extensive region of promoter DNA both upstream and downstream of TATA (probes -4 to 6), in agreement with the cross-linking results of Coulombe and coworkers[21] (**Fig. 6**; **Supplementary Figs. 3e-h, 5b** and **c**; **Supplementary Table 1**). The cleavage sites of Tfg1 map to the border of the Tfg2 dimerization domain/unstructured domain and to the middle of the unstructured domain as shown in **Figure 6b**. Cleavage of Tfg2 maps to the Tfg1 dimerization and the winged helix domains, as shown in **Figure 6d**.

Mapping the domains of the TFIIE subunits near promoter DNA

TFIIE contains two subunits, Tfa1 (TFIIEα) and Tfa2 (TFIIEβ) and is thought to be a heterodimer. TFIIE is required for recruitment of TFIIH and for regulation of TFIIH kinase and helicase activities[40]. Electron crystallography studies place TFIIE near the Pol II active center[41], consistent with the observed cross-linking of IIE to DNA overlapping the transcription bubble region in the PIC[19].

The exposed amino acid surfaces of Tfa1 and Tfa2 (**Fig. 7**) near promoter DNA within the PIC were located and results are summarized in **Supplementary Figs. 3i-l** and

Supplementary Table 1. Both Tfa1 and Tfa2 were cleaved from a relatively short region of the promoter (probes 2 to 4) just downstream of TATA and overlapping the transcription bubble region, in agreement with cross-linking data from Ebright and coworkers[19] and with EM studies[41]. Other cross-linking studies indicating that TFIIE also binds upstream and downstream of the bubble region[18] is not consistent with our findings. Cleavage of Tfa1 maps to the winged helix domain as shown in **Figure 7b**, while cleavage of Tfa2 maps to the C-terminal region, as shown in **Figure 7d**. The position of the cut sites in Tfa1 and Tfa2 are consistent with the findings of recent EM studies[42] that locate the N-terminus of TFIIEα adjacent to the C-terminal third of TFIIEβ in the human TFIIE dimer. Since only partial crystal structures are available for both TFIIE and TFIIF, these cleavage sites could not be modeled onto a structure.

Discussion

To distinguish between previously proposed models for PIC structure and to map the regions of the GTFs near promoter DNA in the PIC, we tethered FeBABE at specific locations of the *HIS4* promoter. FeBABE was positioned both upstream and downstream of the TATA box from positions -87 to +1 (-55 to +33, human numbering) and used along with yeast nuclear extracts to form PICs. Activation of FeBABE generated hydroxyl radicals, cleaving exposed regions of polypeptides in close proximity to these promoter positions. Our data provides a model for the path of promoter DNA in the PIC across the surface of Pol II and is clearly compatible with only one previously proposed model for PIC structure. From the locations of the cleavage sites on the Pol II surface, which are distributed among Rpb1 and Rpb2 on either side of the central cleft, the promoter DNA downstream of TATA most likely runs outside of and along the center of the Pol II cleft. DNA upstream of TATA bends around the back face of Pol II near Rpb2, extending towards Rpb3. Our model is entirely consistent with the position of TFIIB in PICs mapped using FeBABE tethered to cysteine residues in the TFIIB core, finger, and ribbon domains[24,25] but is very different from other proposed models[18,21,23]. The observed DNA bending around Pol II is in agreement with the DNA bend imposed by the binding of TBP to DNA visualized by X-ray structural analysis[33,34].

Positioning of downstream promoter DNA over the Pol II central cleft is in agreement with most of the DNA-protein cross-linking results in the human and yeast systems for Rpb1 and Rpb2. In these studies, extensive cross-linking between promoter DNA downstream of TATA and the Rpb1 and Rpb2 subunits was observed[18-22]. In contrast, our results also demonstrate that DNA upstream of TATA follows a path near Rpb2 that lies distant from Rpb1. This corresponds to the cross-linking between Rpb2 and upstream DNA observed in previous studies, but not with the weak cross-linking observed to Rpb1[18,20-22] or to Rpb7[22]. Previous cross-linking studies utilized either partially assembled PICs or PICs formed from the minimal set of GTFs, in contrast to the studies reported here where the PICs contain Mediator, activator, and other coactivators. Since EM studies of Pol II-Mediator show that Mediator binds to the face of Pol II

predicted to interact with upstream promoter DNA, it is possible that Mediator and/or another of these coactivators influences the DNA path in the PIC.

A summary of the hydroxyl radical cleavage results are shown in **Figure 8** with specific probes cleaving the GTFs indicated in panel **a**, and results are mapped on the model of the PIC in panels **b-d**. From this summary, it is clear that promoter DNA downstream of TATA is very crowded, being positioned near Pol II, TFIIB, TFIIF, TFIIE, and TFIIH. Of all the GTFs, TFIIF is in proximity to the most extensive region of promoter DNA, extending both upstream and downstream of TATA. The EM structure of the Pol II-TFIIF complex showed that Tfg2, the small TFIIF subunit, makes extensive contacts with a large surface of Pol II, consistent with our cleavage results. In contrast, the EM studies did not reveal appreciable density of Tfg1 in proximity to DNA. Our results demonstrate that in the PIC, Tfg1 is in proximity to nearly 60-bp of promoter DNA. Perhaps the position of Tfg1 is fixed only in higher order PIC complexes.

The mapped position of TFIIF partially overlaps with that for both subunits of TFIIE, which lies downstream of TATA and over the site for initiation of the transcription bubble. As TFIIF, TFIIE, and TFIIB[23,25] lie in this critical region, they are positioned to play important roles in the mechanism of Open Complex formation, where the double stranded DNA is disrupted and the template strand of DNA is inserted within the Pol II active site. Our results are consistent with the previous proposal that TFIIE stabilizes single stranded DNA in the Open Complex[43] as well as with the findings that Tfg1 and the B-finger region of TFIIB lie in the active site of Pol II in the PIC[23,25]. Notably, probes 4 and 5, which lie over the transcription bubble, cleave in the unstructured region of Tfg1, a region previously found to play an important role in transcription start site selection[44,45]. Also, from the position of cleavage of Tfa1 at the N-terminus and of Tfa2 in the C-terminal third, we propose that the dimerization region of TFIIE[42] lies over the site for initiation of single stranded DNA formation. Moreover, probes 2 and 3, which cover this DNA region, cleave Tfa2, the small subunit of TFIIE, in a segment adjacent to the Tfa2 Basic region. This Basic region encompasses the proposed site of single stranded DNA binding[43]. From these positions, TFIIE and/or TFIIF may play a role

analogous to region 2.3 of bacterial sigma factor, where a conserved hydrophobic residue has been proposed to stack on the exposed downstream face of the base pair forming the upstream edge of the transcription bubble[46].

HIS4 promoter probes 6-9 all generate cleavage in Rad25, the helicase subunit of TFIIH that is critical for Open Complex formation. The position of Rad25 at this site is consistent with previous cross-linking data placing human XPB downstream of the transcription bubble region at the AdMLP[19]. Importantly, hydroxyl radical cleavage of Rad25 demonstrated that the two helicase subdomains were exposed and in proximity to this segment of the promoter. Mapping of these cleavage patterns to a structure model for the two helicase subdomains suggests that in the PIC, the promoter DNA lies in a groove between the subdomains in a manner similar to nucleic acid interacting with other diverse helicases[35]. This suggests that in the PIC, the Rad25/XPB helicase is poised to induce torsional strain on promoter DNA. Consistent with the model proposed by Ebright and coworkers[19] the helicase would act downstream from the segment that initially becomes single stranded. An alternative model for XPB was recently suggested[47] in which the XPB ATPase, but not the DNA helicase activity, is required for Open Complex formation. In this model, the XPB ATPase serves to induce a conformational change in the PIC during Open Complex formation rather than the helicase acting directly to induce strain on DNA. Our findings suggest that the first model is more likely since Rad25 within the PIC seems in position to act on promoter DNA.

Finally, comparison of our results and the results of protein-DNA cross-linking in the human[17-21] and yeast[22] systems demonstrates that PIC formation occurs in a similar location with respect to the TATA element. This similar organization of the PIC is in contrast to the great disparity in transcriptional start sites between *S. cerevisiae* and human Pol II. A model for start site location in yeast, consistent with available data, is that upon ATP addition, Rad25 generates torsional strain in DNA upstream of its binding site. TFIIF and/or TFIIE would next insert residue(s) between double strand promoter DNA to initiate and stabilize single stranded bubble formation. This would be followed by insertion of the single stranded template strand into the Pol II cleft. Perhaps promoted

by the helicase activity of Rad25, single stranded template DNA is fed through the Pol II cleft until a suitable transcription start site is encountered. This proposed movement would be somewhat analogous to the scrunching mechanism proposed for T7[48] and bacterial Pol at initiation. For bacterial Pol, single-molecule studies have shown that long range (>>8-bp) scrunching can occur, suggesting that the initiation complex can accommodate long-range bubble expansion (A. Revyakin, C.Y. Liu, T. Strick, and R. Ebright, Howard Hughes Medical Institute, Rutgers University, Piscataway, NJ, personal communication). In *S. cerevisiae*, this would involve long single-stranded DNA loops protruding from the central cleft while Pol II retained contacts with the GTFs at the promoter. Based on previous genetics and biochemical findings, TFIIB and TFIIF would play a critical role in the transcription start site selection[44,49-52]. Future studies combining structural analysis, biochemistry and genetics should elucidate the mechanism for transcription start site selection and reveal the basis for the great differences in start site selection shown by yeast and mammalian systems.

Methods

Preparation of phosphorothioate-containing *HIS4* DNA templates Phosphorothioate-containing templates were prepared by ligating the products of two PCR reactions, where one of the PCR products contains phosphorothioates incorporated at specific locations. This is described in detail in **Supplementary Methods**.

FeBABE conjugation to phosphorothioate-containing *HIS4* promoters Conjugation reactions containing 6 mM FeBABE (iron (S)-1-(p-bromoacetamido-benyl) EDTA; Dojindo) and 860 ng phosphorothioate DNA were performed in 20 mM MOPS, pH 7.9. After 16 hours of incubation at 50 ^{0}C, the biotinylated phosphorothioate-containing *HIS4* promoters were immobilized to M280 streptavidin-coated dynabeads (Dynal) and excess FeBABE was removed by washing 3 times with 200 μl of transcription buffer containing 0.05% (w/v) NP-40. FeBABE-tethered probes were directly used in the hydroxyl radical cleavage assay.

***In vitro* transcription with immobilized templates** Immobilized template assays for *in vitro* transcription were performed as described (www.fhcrc.org/labs/hahn). 180 μg nuclear extract and 48 ng activator Gal4-VP16 were used to assemble PICs on immobilized phosphorothioate-containing DNA templates or FeBABE-conjugated DNA templates (215 ng). After 40 min incubation, transcription was initiated by addition of 100 mM of each rNTP. Products were mapped by primer extension as previously described[31].

Hydroxyl Radical Cleavage Assay Hydroxyl radical probing experiments were performed as previously described[24] except that 720 μg yeast nuclear extract, 192 ng Gal4-VP16 activator and 860 ng FeBABE-conjugated *HIS4* promoter immobilized template were used. After PIC formation on immobilized templates, PICs were washed 3 times with 400 μl of transcription buffer containing 0.05% (w/v) NP-40, then Fe-EDTA was activated to generate hydroxyl radicals as previously described[24]. Cleavage

fragments were visualized by anti-Flag Western blot using the LI-COR Bioscience Odyssey Infrared Imaging System.

Determination of the FeBABE Cleavage Sites The molecular sizes of Rpb1-, Rpb2-, Rad25-, Tfg1-, Tfg2-, Tfa1-, Tfa2- and Sua7 -Flag tagged fragments generated by FeBABE cleavage were determined as previously described[24] using a set of *in vitro* translated Flag-tagged constructs for each protein. Based on multiple experiments we estimate that the cleavage sites are mapped within 8-10 residues.

Yeast Strains, Nuclear Extracts and Antisera A complete list is provided in **Supplementary Methods**.

Co-ordinates of the Rad25 model and of the Pol II-TFIIB-TBP-DNA model are available upon request.

Acknowledgements

We thank R. Ebright for the suggestion of the scrunching mechanism for yeast Pol II, H.T. Chen and L. Warfield for yeast strains and nuclear extracts, and L. Fan and J. Tainer for the coordinates of Af XPB before public release. We thank B. Schmidt and J. Roe for technical advice on efficient conjugation of FeBABE to DNA, and H.T. Chen, N. Mohibullah, B. Moorefield and I. Whitehouse for discussions and for comments on the manuscript. This work was supported by Grant 5RO1GM053451 from the National Institutes of Health. G. Miller was supported in part as a postdoctoral associate of the Howard Hughes Medical Institute.

References

1. Hahn, S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* **11**, 394-403 (2004).
2. Orphanides, G., Lagrange, T. & Reinberg, D. The general transcription factors of RNA polymerase II. *Genes Dev* **10**, 2657-83 (1996).
3. Woychik, N.A. & Hampsey, M. The RNA polymerase II machinery: structure illuminates function. *Cell* **108**, 453-63 (2002).
4. Holstege, F.C., Fiedler, U. & Timmers, H.T. Three transitions in the RNA polymerase II transcription complex during initiation. *Embo J* **16**, 7468-80 (1997).
5. Kettenberger, H., Armache, K.J. & Cramer, P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell* **16**, 955-65 (2004).
6. Tirode, F., Busso, D., Coin, F. & Egly, J.M. Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol Cell* **3**, 87-95 (1999).
7. Wang, W., Carey, M. & Gralla, J.D. Polymerase II promoter activation: closed complex formation and ATP-driven start site opening. *Science* **255**, 450-3 (1992).
8. Hampsey, M. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev* **62**, 465-503 (1998).
9. Giardina, C. & Lis, J.T. DNA melting on yeast RNA polymerase II promoters. *Science* **261**, 759-62 (1993).
10. Holstege, F.C., van der Vliet, P.C. & Timmers, H.T. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *Embo J* **15**, 1666-77 (1996).
11. Pal, M., Ponticelli, A.S. & Luse, D.S. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol Cell* **19**, 101-10 (2005).
12. Armache, K.J., Kettenberger, H. & Cramer, P. Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci U S A* **100**, 6964-8 (2003).
13. Bushnell, D.A. & Kornberg, R.D. Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proc Natl Acad Sci U S A* **100**, 6969-73 (2003).
14. Cramer, P., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863-76 (2001).
15. Kettenberger, H., Armache, K.J. & Cramer, P. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* **114**, 347-57 (2003).
16. Westover, K.D., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* **303**, 1014-6 (2004).
17. Douziech, M. et al. Mechanism of promoter melting by the xeroderma pigmentosum complementation group B helicase of transcription factor IIH revealed by protein-DNA photo-cross-linking. *Mol Cell Biol* **20**, 8168-77 (2000).
18. Forget, D., Langelier, M.F., Therien, C., Trinh, V. & Coulombe, B. Photo-cross-linking of a purified preinitiation complex reveals central roles for the RNA

polymerase II mobile clamp and TFIIE in initiation mechanisms. *Mol Cell Biol* **24**, 1122-31 (2004).
19. Kim, T.K., Ebright, R.H. & Reinberg, D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* **288**, 1418-22 (2000).
20. Kim, T.K. et al. Trajectory of DNA in the RNA polymerase II transcription preinitiation complex. *Proc Natl Acad Sci U S A* **94**, 12268-73 (1997).
21. Robert, F. et al. Wrapping of promoter DNA around the RNA polymerase II initiation complex induced by TFIIF. *Mol Cell* **2**, 341-51 (1998).
22. Chen, B.S., Mandal, S.S. & Hampsey, M. High-resolution protein-DNA contacts for the yeast RNA polymerase II general transcription machinery. *Biochemistry* **43**, 12741-9 (2004).
23. Bushnell, D.A., Westover, K.D., Davis, R.E. & Kornberg, R.D. Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. *Science* **303**, 983-8 (2004).
24. Chen, H.T. & Hahn, S. Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex. *Mol Cell* **12**, 437-47 (2003).
25. Chen, H.T. & Hahn, S. Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. *Cell* **119**, 169-80 (2004).
26. Datwyler, S.A. & Meares, C.F. Protein-protein interactions mapped by artificial proteases: where sigma factors bind to RNA polymerase. *Trends Biochem Sci* **25**, 408-14 (2000).
27. Joseph, S. & Noller, H.F. Mapping the rRNA neighborhood of the acceptor end of tRNA in the ribosome. *Embo J* **15**, 910-6 (1996).
28. Joseph, S., Weiser, B. & Noller, H.F. Mapping the inside of the ribosome with an RNA helical ruler. *Science* **278**, 1093-8 (1997).
29. Schmidt, B.D. & Meares, C.F. Proteolytic DNA for mapping protein-DNA interactions. *Biochemistry* **41**, 4186-92 (2002).
30. Fishburn, J., Mohibullah, N. & Hahn, S. Function of a eukaryotic transcription activator during the transcription cycle. *Mol Cell* **18**, 369-78 (2005).
31. Ranish, J.A., Yudkovsky, N. & Hahn, S. Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a postrecruitment role for the TATA box and TFIIB. *Genes Dev* **13**, 49-63 (1999).
32. Warfield, L., Ranish, J.A. & Hahn, S. Positive and negative functions of the SAGA complex mediated through interaction of Spt8 with TBP and the N-terminal domain of TFIIA. *Genes Dev* **18**, 1022-34 (2004).
33. Tsai, F.T. & Sigler, P.B. Structural basis of preinitiation complex assembly on human pol II promoters. *Embo J* **19**, 25-36 (2000).
34. Nikolov, D.B. et al. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377**, 119-28 (1995).
35. Caruthers, J.M. & McKay, D.B. Helicase structure and mechanism. *Curr Opin Struct Biol* **12**, 123-33 (2002).
36. Fan, L. et al. Conserved XPB Core Structure and Motifs for DNA Unwinding: Implications for Pathway Selection of Transcription or Excision Repair. *Mol Cell* **22**, 27-37 (2006).

37. Kabani, M., Michot, K., Boschiero, C. & Werner, M. Anc1 interacts with the catalytic subunits of the general transcription factors TFIID and TFIIF, the chromatin remodeling complexes RSC and INO80, and the histone acetyltransferase complex NuA3. *Biochem Biophys Res Commun* **332**, 398-403 (2005).
38. Gaiser, F., Tan, S. & Richmond, T.J. Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 A resolution. *J Mol Biol* **302**, 1119-27 (2000).
39. Chung, W.H. et al. RNA polymerase II/TFIIF structure and conserved organization of the initiation complex. *Mol Cell* **12**, 1003-13 (2003).
40. Ohkuma, Y. & Roeder, R.G. Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation. *Nature* **368**, 160-3 (1994).
41. Leuther, K.K., Bushnell, D.A. & Kornberg, R.D. Two-dimensional crystallography of TFIIB- and IIE-RNA polymerase II complexes: implications for start site selection and initiation complex formation. *Cell* **85**, 773-9 (1996).
42. Jawhari, A. et al. Structure and oligomeric state of human transcription factor TFIIE. *EMBO Rep* (2006).
43. Watanabe, T. et al. The carboxy terminus of the small subunit of TFIIE regulates the transition from transcription initiation to elongation by RNA polymerase II. *Mol Cell Biol* **23**, 2914-26 (2003).
44. Freire-Picos, M.A., Krishnamurthy, S., Sun, Z.W. & Hampsey, M. Evidence that the Tfg1/Tfg2 dimer interface of TFIIF lies near the active center of the RNA polymerase II initiation complex. *Nucleic Acids Res* **33**, 5045-52 (2005).
45. Ghazy, M.A., Brodie, S.A., Ammerman, M.L., Ziegler, L.M. & Ponticelli, A.S. Amino acid substitutions in yeast TFIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II. *Mol Cell Biol* **24**, 10975-85 (2004).
46. Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O. & Darst, S.A. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**, 1285-90 (2002).
47. Lin, Y.C., Choi, W.S. & Gralla, J.D. TFIIH XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape. *Nat Struct Mol Biol* **12**, 603-7 (2005).
48. Cheetham, G.M. & Steitz, T.A. Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286**, 2305-9 (1999).
49. Chen, B.S. & Hampsey, M. Functional interaction between TFIIB and the Rpb2 subunit of RNA polymerase II: implications for the mechanism of transcription initiation. *Mol Cell Biol* **24**, 3983-91 (2004).
50. Faitar, S.L., Brodie, S.A. & Ponticelli, A.S. Promoter-specific shifts in transcription initiation conferred by yeast TFIIB mutations are determined by the sequence in the immediate vicinity of the start sites. *Mol Cell Biol* **21**, 4427-40 (2001).
51. Pinto, I., Wu, W.H., Na, J.G. & Hampsey, M. Characterization of sua7 mutations defines a domain of TFIIB involved in transcription start site selection in yeast. *J Biol Chem* **269**, 30569-73 (1994).

52. Sun, Z.W. & Hampsey, M. Identification of the gene (SSU71/TFG1) encoding the largest subunit of transcription factor TFIIF as a suppressor of a TFIIB mutation in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **92**, 3127-31 (1995).
53. Meinhart, A., Blobel, J. & Cramer, P. An extended winged helix domain in general transcription factor E/IIE alpha. *J Biol Chem* **278**, 48267-74 (2003).