REVIEW

nature structural & molecular biology

Structure and mechanism of the RNA polymerase II transcription machinery

Steven Hahn

Advances in structure determination of the bacterial and eukaryotic transcription machinery have led to a marked increase in the understanding of the mechanism of transcription. Models for the specific assembly of the RNA polymerase II transcription machinery at a promoter, conformational changes that occur during initiation of transcription, and the mechanism of initiation are discussed in light of recent developments.

Regulation of transcription, the synthesis of RNA from a DNA template, is one of the most important steps in control of cell growth and differentiation. Transcription is carried out by the enzyme RNA polymerase (Pol) along with other factors termed general transcription factors. The general factors are involved in recognition of promoter sequences, the response to regulatory factors and conformational changes essential to the activity of Pol during the transcription cycle[1,2]. Advances made over the past 11 years[3–5] have revealed the structures of bacterial and eukaryotic Pols, several of the key general transcription factors, and most recently, structures and models of Pol II interacting with general transcription factors[6–8]. Combined with biochemical and genetic studies, these structures provide emerging views on the mechanism of the transcription machinery, the dynamic nature of protein-protein and protein-DNA interactions involved, and the mechanism of transcriptional regulation.

Although the transcription machinery of eukaryotes is much more complex than that of prokaryotes or archaea, the general principles of transcription and its regulation are conserved. Bacteria and archaea have only one Pol, whereas eukaryotes use three nuclear enzymes, Pol I–III, to synthesize different classes of RNA. The nuclear Pols share five common subunits, with the remainder showing strong similarity among the eukaryotic and archaeal enzymes[2,9]. Although these enzymes have many more subunits than bacterial Pol, subunits that make up most of Pol II are homologous to subunits from all cellular Pols, suggesting that all these enzymes have the same basic structure and mechanism[10]. In bacteria, the σ subunit is the sole general transcription factor–like polypeptide. σ recognizes promoter sequences, promotes conformational changes in the Pol–DNA complex upon initiation and interacts directly with some transcription activators. In eukaryotes, σ factor function has been replaced by a much larger set of polypeptides, with each of the three forms of Pol having their own set of associated general transcription factors[2,11,12]. The Pol II transcription machinery is the most complex, with a total of nearly 60 polypeptides (**Table 1**), only a few of which are required for transcription by the other nuclear Pols. In contrast, archaea use a simplified version of a Pol II-Pol III-like system, relying on only two essential general, factors, TBP (TATA-binding protein) and TFB (related to the Pol II and Pol III general factors TFIIB and Brf1)[9].

Fred Hutchinson Cancer Research Center and Howard Hughes Medical Institute, 1100 Fairview Ave N., A1-162, Seattle, Washington 98109, USA. Correspondence should be addressed to S.H. (shahn@fhcrc.org).

Published online 27 April 2004; doi:10.1038/nsmb763

The RNA Pol II transcription cycle

Pol II transcription typically begins with the binding of gene-specific regulatory factors near the site of transcription initiation. These factors can act indirectly on the transcription machinery by recruiting factors that modify chromatin structure, or directly by interacting with components of the transcription machinery. In the simplest form of gene activation, both the direct and indirect mechanisms result in recruitment of the transcription machinery to a core promoter (the minimal DNA sequence needed to specify nonregulated or basal transcription; **Fig. 1**)[13,14]. The core promoter serves to position Pol II in a state termed the preinitiation complex (PIC), analogous to the bacterial closed complex. In this state, Pol II and the general factors are all bound to the promoter but are not in an active conformation to begin transcription. Next, a marked conformational change occurs in which 11–15 base pairs (bp) of DNA surrounding the transcription start site are melted and the template strand of the promoter is positioned within the active site cleft of Pol to form the open complex[15]. Initiation of transcription begins with synthesis of the first phosphodiester bond of RNA. In many systems, multiple short RNAs (of three to ten bases), termed abortive products, are synthesized before Pol productively initiates synthesis of full-length RNAs[16,17]. After synthesis of ~30 bases of RNA, Pol is thought to release its contacts with the core promoter and the rest of the transcription machinery and enter the stage of transcription elongation. Factors that promote productive RNA chain synthesis, RNA processing, RNA export and chromatin modification can all be recruited to elongating Pol II[18]. After initiation of transcription by Pol II *in vitro*, many of the general transcription factors remain behind at the promoter in the scaffold complex[19]. This complex presumably marks genes that have been transcribed and enables the typically slow step of recruitment to

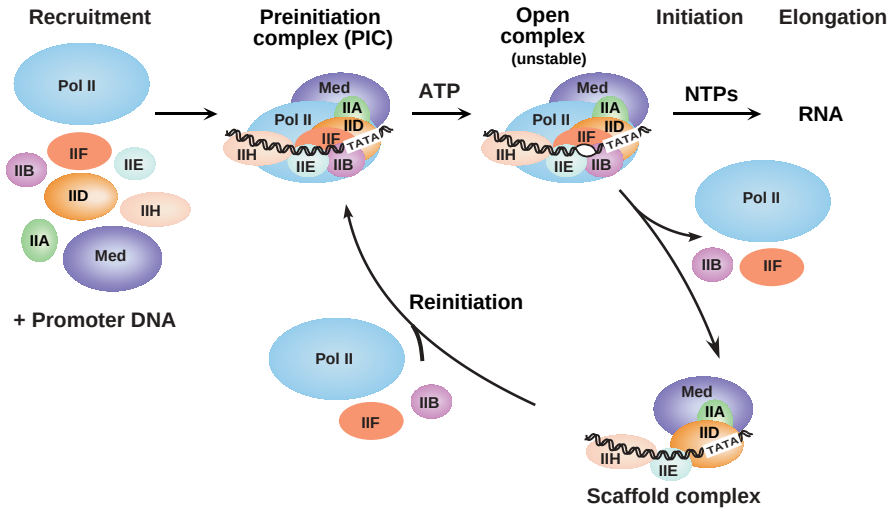

Figure 1 The pathway of transcription initiation and reinitiation for RNA Pol II. See **Table 1** for a description of each transcription factor and Mediator (Med).

be bypassed in subsequent rounds of transcription. Certain transcription activation domains can stabilize this scaffold–promoter complex *in vitro*. The scaffold complex can then rapidly recruit the remaining general factors to promote transcription reinitiation.

Promoter recognition using low-affinity interactions

Recognition of the core promoter by the transcription machinery is essential for correct positioning and assembly of Pol II and the general factors. Sequence elements found in core promoters include the TATA element (TBP-binding site), BRE (TFIIB-recognition element), Inr (initiator element) and DPE (downstream promoter element)[20]. Most promoters contain one or more of these elements, but no one element is absolutely essential for promoter function. The promoter elements are binding sites for subunits of the transcription machinery and serve to orient the transcription machinery at the promoter asymmetrically to direct unidirectional transcription.

The core domain of TBP consists of two imperfect repeats forming a saddle-shaped molecule that binds the widened minor groove of an 8-bp TATA element, unwinding about a third of a helical turn and bending the DNA ~80 Å toward the major groove[21,22] (**Fig. 2a**). At TATA-containing promoters, formation of this protein–DNA complex is the initial step in assembly of the transcription machinery. Although the TBP molecule is symmetrically shaped, the protein surface of the two repeats is very divergent, forming a large asymmetric protein-DNA interface, creating a platform for binding other components of the transcription machinery. Biochemical studies have elegantly shown that TBP does not bind to TATA elements with high orientation specificity[23], leading to the finding that other promoter elements in combination with TATA determine the orientation of transcription machinery assembly at a promoter. The BRE element was first recognized as a sequence contributing to high-affinity binding of TFIIB and TFB to the human and archaeal TBP–DNA complex[24,25]. In archaea, where the DNA-binding surface of the two TBP imperfect repeats is more symmetrical than that of eukaryotic TBPs, the BRE is the primary determinant of transcription orientation[26,27].

The other two core promoter elements with proven function, Inr and DPE, probably serve as binding sites for the TAF (TBP-associated factor) subunits of the general factor TFIID. A combination of two TAFs (TAF1 and TAF2) specifically binds the Inr, and selection for an optimal TAF1-2 binding sequence leads to identification of a sequence closely resembling the Inr element[28]. Additionally, UV crosslinking has shown that TAF1 and TAF2 are normally positioned close to the Inr[29] and that TAF6 and TAF9 lie close to the DPE[30]. Proper function of a DPE-containing promoter requires an Inr element, probably because these elements cooperatively promote the correct binding of TFIID[30]. In summary, specific binding of the transcription machinery at the core promoter derives from cooperative binding of two or more general transcription factor subunits to degenerate, low-specificity promoter elements. The combination of these elements varies between promoters and, in some cases, the core promoter elements determine activator and enhancer specificity[31].

Table 1 ***S. cerevisiae*** **RNA Pol II general transcription factors and coactivators**

Factor		No. of subunits	Function
TFIIA		2	Stabilizes TBP and TFIID-DNA binding. Blocks transcription inhibitors. Positive and negative gene regulation.
TFIIB		1	Binds TBP, Pol II and promoter DNA. Helps fix transcription start site.
TFIID	TBP	1	Binds TATA element and deforms promoter DNA. Platform for assembly of TFIIB, TFIIA and TAFs.
	TAFs	14	Binds INR and DPE promoter elements. Target of regulatory factors.
Mediator		24	Binds cooperatively with Pol II. Kinase and acetyltransferase activity. Stimulates basal and activated transcription. Target of regulatory factors.
TFIIF[a]		3	Binds Pol II and is involved in Pol II recruitment to PIC and in open complex formation.
TFIIE		2	Binds promoter near transcription start. May help open or stabilize the transcription bubble in the open complex.
TFIIH		10	Functions in transcription and DNA repair. Kinase and two helicase activities. Essential for open complex formation. Mutations in IIH can cause human disease.
SAGA[b]	TAFs	5	Function unknown.
	Spts, Adas, Sgfs	9	SAGA structure. Interact with TBP, TFIIA and Gcn5.
	Gcn5	1	Histone acetyltransferase.
	Tra1	1	Essential subunit, activator target. Third largest yeast protein. Component of the NuA4 HAT complex.
	Ubp8	1	Ubiquitin protease.

[a]Yeast has one extra nonessential subunit compared with other organisms studied. [b]Yeast also contain SLIK/SALSA, a closely related complex.

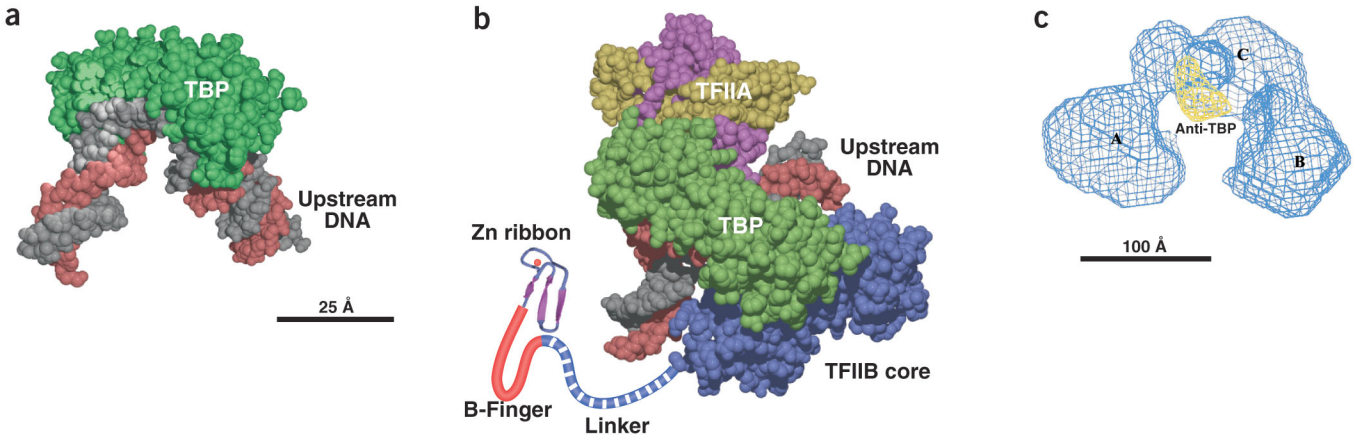

Figure 2 General transcription factor structures. (**a**) Structure of TBP (green) bound to TATA-DNA with B-form DNA (gray and red) modeled upstream and downstream of the TATA box[21,22]. (**b**) Structure model of the TBP–TFIIA–TFIIB–DNA complex. TBP (green) is shown from above binding to the TFIIB core domain (TFIIBc, blue) and TFIIA (large subunit magenta, small subunit yellow). The zinc ribbon domain (β-strands with red Zn atom) that connects to the B-finger domain is normally located in the PIC within the Pol II active site and is connected to the TFIIBc domain through a flexible linker. This model is a composite of the DNA-TBP-TFIIBc, DNA-TBP-TFIIA, and the TFIIB Zn ribbon NMR and crystal structures[8,41–43,50,125]. (**c**) EM structure of human TFIID with density due to anti-TBP in yellow and the three TFIID lobes labeled A–C[57]. Panel **c** was provided by E. Nogales.

The role of TBP at TATA-less promoters. It was initially believed that most Pol II promoters contained a TATA element; however, subsequent sequence analysis has shown that only ~30% of mRNA genes analyzed in *Drosophila melanogaster* contain a recognizable TATA[32]. Although TBP can recognize divergent AT-rich sequences because of its DNA-binding mechanism (see below), promoters that do not have a TATA-like sequence ~30 bp upstream of the transcription start site would be incompatible with specific binding by TBP[33]. In *D. melanogaster* and human promoters, many of these non-TATA-containing promoters have some combination of Inr and DPE elements[20].

Must TBP bind DNA in order to function? At promoters with a functional TATA, mutation of TATA away from the consensus severely decreases transcription[34]. In biochemical studies of the yeast HIS4 promoter, mutation of the TATA to a GC-rich sequence allows recruitment of the transcription machinery to a promoter at a reduced level, but transcription initiation is completely abolished[35]. These results demonstrate that at one class of promoter, assembly of the transcription machinery into a productive complex requires that TBP bind the TATA element as seen in the crystal structure. Although TBP has tremendous flexibility in the ability to bind variants of the TATA sequence, not all sequences are compatible with TBP binding. For example, a C or G in certain positions in the TATA element is incompatible with the DNA-binding surface of TBP[33]. Because many human and *D. melanogaster* promoters have no recognizable TATA element or even AT-rich regions upstream from the transcription start site[32], this suggests that if TBP interacts with DNA at these promoters, it must do so by a different mechanism from that seen at classical TATA elements. In support of this model, a mutation at the TBP-binding surface that abolished detectable binding to a TATA element *in vitro* blocked transcription from a TATA-containing promoter but not from an Inr-containing promoter[36]. At promoters lacking TATA, TBP may nucleate protein-protein interactions among the general transcription factors and interact nonspecifically with DNA, whereas DNA bending is facilitated by interaction of other factors such as TAFs with Inr and DPE elements.

Although only one gene encodes TBP in yeast and most archaea, higher eukaryotes have one or two copies of genes encoding TBP-related factors (TRFs) in addition to TBP[37,38]. It is well established that TRFs promote transcription from a subset of Pol II genes in a cell-type-specific fashion. Trf1, unique to insect cells, binds a TC-rich sequence rather than a TATA element and promotes transcription from a small subset of Pol II promoters as well as all *D. melanogaster* Pol III transcription[39,40]. Trf2, conserved among *D. melanogaster*, mouse and human, also directs transcription from a subset of promoters. Like TBP, Trf1 and Trf2 are both components of multisubunit complexes, although the identity of most of the Trf-associated factors is not yet known.

General factors interacting directly with TBP

TFIIA and TFIIB are the two general factors that interact specifically and independently with TBP. The X-ray structure of these factors bound to the TBP–DNA complex has shown that both TFIIA and TFIIB recognize TBP and the DNA distorted by TBP binding[41–43] (**Fig. 2b**). Both factors recognize the DNA backbone and, as mentioned above, TFIIB can also make base-specific contacts with the BRE. TFIIA is a heterodimer composed of two domains, the C-terminal domain contacting TBP-DNA and the N-terminal domain pointing directly away from TBP. TFIIA stabilizes TBP-DNA binding[44] and strongly promotes binding of TFIID to DNA through an anti-repression mechanism by competing with the TAF1 N-terminal domain (TAND) that occludes the DNA-binding surface of TBP when TFIID is not bound to DNA[45–47]. This effect is particularly marked using human TFIID and certain transcription activators, where a considerable change in the DNA-binding activity of TFIID is observed in the presence of TFIIA and activator[48]. TFIIA can also compete with the negative regulatory factors Mot1 and NC2 to promote TBP binding *in vitro*[2].

TFIIB contains two domains conserved in the Pol III and archaea factors Brf1 and TFB: an N-terminal zinc ribbon domain (ribbon) connected by a flexible linker to the C-terminal core domain (TFIIBc) that binds TBP-DNA (**Fig. 2b**). Both the ribbon and core domains bind cooperatively to RNA Pol II; neither isolated TFIIB domain detectably interacts with Pol II[6,49]. The functional surface of the ribbon domain has been conserved in TFIIB, Brf1 and TFB and is essential for recruitment of Pol II to the PIC[50]. For RNA Pol III, the Brf1 ribbon domain is required for normal formation of the open complex, a function probably conserved in the Pol II system (see below). The linker connecting the ribbon and core domains contains a short

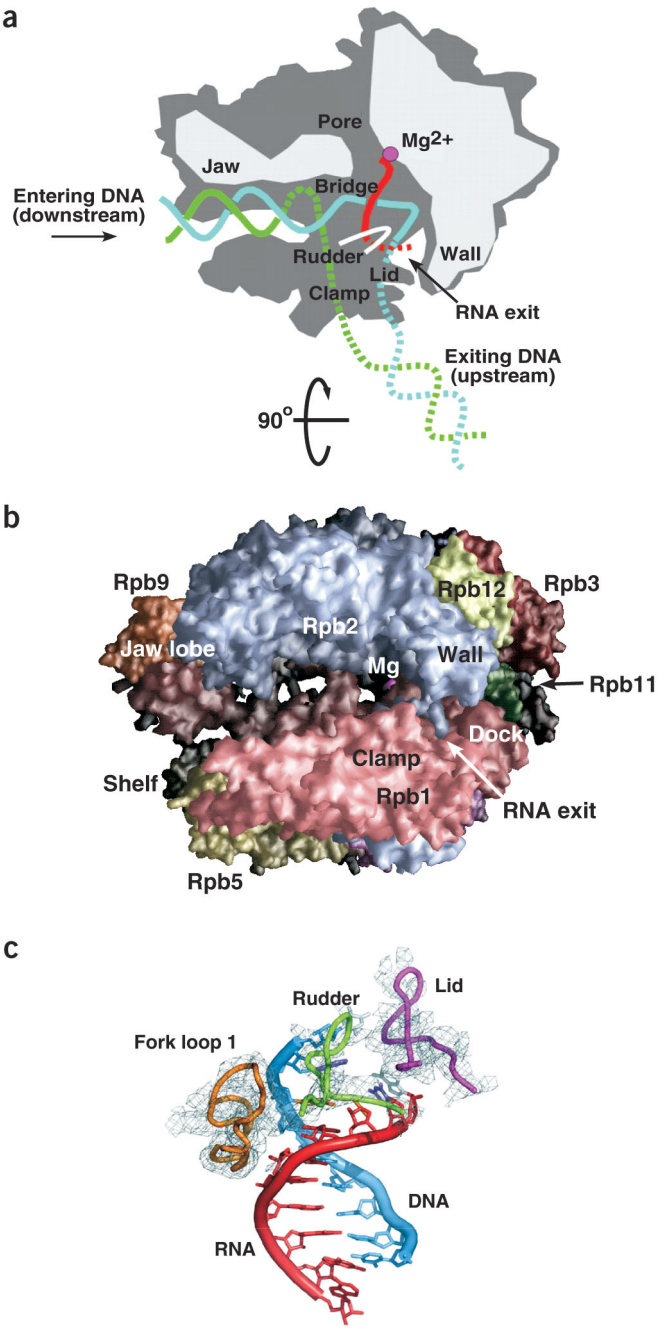

Figure 3 The ten-subunit yeast RNA Pol II structure and structure of the elongation complex. (**a**) Top view of Pol II with the DNA template strand (blue), nontemplate strand (green), RNA (red) and active site Mg (magenta). This image adapted from ref. 126. (**b**) Side view of Pol II looking into the active site cleft[70–72]. Rpb1, light red; Rpb2, light blue; Rpb12, light green; Rpb3, red; Rpb11, dark green; Rpb5, dark yellow; Rpb9, orange; active site Mg, magenta sphere. (**c**) Details of the interactions between Pol II and the DNA-RNA hybrid in elongating Pol II. This image from ref. 81.

conserved block of sequence that forms a loop termed the B-finger, which is positioned in the active site of Pol II[8], where it functions in determining the transcription start site (see below).

TAFS are shared subunits of two large complexes

TFIID is a complex composed of TBP and ~14 TAFs, nearly all of which have been conserved through evolution[51–53]. The TAFs function in promoter recognition and in positive and negative regulation of transcription. Although yeast contains only one form of TFIID, at least six TAFs in mammalian and *D. melanogaster* TFIID have alternative subunits that change the composition of TFIID in a cell type– and development-specific fashion. TAFs have been implicated in gene regulation in both biochemical and genetic studies. In certain *in vitro* systems, TAFs can be functional targets for transcription activators[54,55].

A subset of TAFs have DNA-binding activity and at least one of these TAFs has protein acetylase and ubiquitylation activity, and in addition, can bind acetylated nucleosomes[56]. Some TAFs are also subunits of complexes lacking TBP involved in covalent chromatin modification and transcriptional coactivation such as yeast SAGA and SLIK/SALSA (Spt-Ada-Gcn5-acetylase and SAGA-like complex) and the human complexes pCAF and STAGA (p300/CBP and Spt-TAF-Gcn5-acetylase)[52]. A core of five TAFs is found in both TFIID and the acetylase–coactivator complexes.

Multiple histone-fold-domain interactions shape TFIID structure. The structure of human, *D. melanogaster* and yeast TFIID has been determined at low resolution by electron microscopy[57,58] (**Fig. 2c**). TFIID contains three lobes (termed A–C) arranged in a horseshoe shape, observed in both closed and open configurations. Immune localization of TBP in this complex shows that TBP lies in the center lobe on the inside of the horseshoe, presumably exposing its DNA-binding surface[57]. Comparison of this structure to that of TFIID with either TFIIA or TFIIB localizes these two additional factors to either side of the TBP-binding site in the center lobe. Depending on the experimental conditions used, TFIID protects 40–60 bp of DNA from DNAse I cleavage[47,48]. From the shape of TFIID, the two noncentral lobes potentially provide a large surface for interaction with DNA. Again, further structural studies should clarify whether the known DNA-interacting TAFs are on these surfaces.

The structures of several TAF subunits were solved by X-ray crystallography[59,60]. These studies, combined with sequence analysis, have shown that 9 of 13 conserved TAF subunits have histone fold domains (HFDs)[61]. TAFs form at least five histone-like pairs essential for the function of TFIID (TAF pairs 4-12, 6-9, 3-10, 10-8 and 11-13). Mapping the location of the HFD TAFs to the TFIID structure using electron microscopy has given unexpected results[62], the most surprising of which is that the TAFs shared between TFIID and SAGA are not found in a substructure of TFIID but are distributed among the three lobes. Although it is not known whether the HFDs are involved in DNA binding, their potential DNA interaction must be different from those of the nucleosome, as many of the side chains in the HFDs that interact with DNA in the nucleosome are not conserved in the TAFs[63]. Because biochemical and structural studies have shown that Pol II and other general transcription factors interact extensively with promoter DNA (see below), it seems likely that at least some of the extensive protein-DNA interactions between TAFs and promoter DNA must change during PIC formation to allow access of these other factors to the promoter.

TFIID is not universally required at all promoters. Molecular genetic studies have shown that promoters vary widely in the requirement for TAFs to promote normal gene regulation[8,64–66]. Although most genes are dependent on at least some TAFs for normal regulation, an important class of promoters seems independent of any TAF. In yeast, these completely TAF-independent promoters recruit TBP but not the TAFs upon gene activation[67,68]. Because TFIID is a large complex protecting

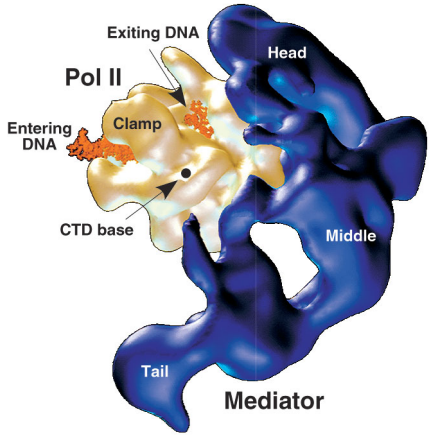

Figure 4 EM structure of the Pol II–mediator complex. Mediator (dark blue) with head, middle and tail domains. Pol II (gold) with modeled upstream and downstream DNA (orange). The dot represents the presumed beginning of the CTD. Figure from ref. 83.

40–60 bp of promoter DNA, it would be expected that substantial structural differences exist between PICs formed with and without TAFs. However, at some TFIID-independent promoters, a TAF-containing complex such as SAGA may functionally replace the TFIID, consistent with results suggesting that TFIID and SAGA function overlap at many yeast genes[69].

Pol II as the center of attention

Pol II lies at the center of the transcription machinery, interacting with the general transcription factors in the PIC, breaking these interactions upon initiation and promoter clearance and associating with another set of factors during elongation and termination. Nearly all Pol II subunits have clear counterparts in the other two nuclear Pols and in archaea. Pol II subunits can be classified into three overlapping categories: subunits of the core domain having homologous counterparts in bacterial Pol (Rpb1, 2, 3 and 11), subunits shared between all three nuclear polymerases (Rpb5, 6, 8, 10 and 12) and subunits specific to Pol II but not essential for transcription elongation (Rpb4, 7 and 9).

Pol II structure. A breakthrough in understanding the mechanism of transcription was achieved with the high-resolution structures of bacterial Pol and the Pol II enzyme[70–72]. Since the initial structural description of bacterial Pol, the structures of Pol in complex with the elongation factor GreB[73], with σ factor (holoenzyme)[74–76] as well as of holoenzyme in complex with a fork junction DNA[77] (analogous to an intermediate in open complex formation) have been determined. The Pol II structure and models for protein interactions have been determined for the 10-subunit and 12-subunit enzymes without DNA[78,79], the 10-subunit enzyme in two different transcribing complexes[80,81] in complex with the general factor TFIIB[6,8] and the elongation factor TFIIS[82]. Lower-resolution EM structures have also been obtained for Pol II binding to the mediator complex[83] as well as to the general factor TFIIF[7]. The structures of these multifactor complexes are beginning to reveal the assembly mechanism for the general transcription machinery and identifying conformational changes in protein and DNA that must occur during transcription initiation.

The highest-resolution Pol II structures are those of the ten-subunit enzyme lacking the Rpb4 and Rpb7 subunits[72] (**Fig. 3a**,**b**). These two subunits are important for transcription initiation but not for elongation. Pol II is composed of four mobile elements termed core, clamp, shelf, and jaw lobe that move relative to each other. The core element (Rpb3, 10, 11, 12 and regions of Rpb1 and Rpb2 forming the active center)[80] accounts for about half the mass of Pol II and is composed mainly of subunits common to all cellular Pols. At the center of the enzyme is a deep cleft where incoming DNA enters from one side and the active site is buried at the base. This cleft is formed by all four mobile elements and has been observed in both closed and open conformations in the ten-subunit enzyme. The shelf and jaw lobe elements move relatively little and can rotate parallel to the active site cleft. The clamp element, connected to the core through a set of flexible switches, moves with a large swinging motion of up to 30 Å to open and close the cleft. Recent work has revealed the structure of the complete 12-subunit enzyme, showing Rpb4-Rpb7 binding to a pocket formed by Rpb1, 2 and 6 at the base of the clamp[78,79]. Rpb7 in this pocket acts as a wedge to lock the clamp in the closed conformation. This striking finding has important implications for the mechanism of initiation and suggests that double-stranded DNA never enters the active site cleft. Rather, it has been proposed that during open complex formation, the single-stranded DNA template strand is inserted deep into the cleft to reach the active site. This mechanism is probably preserved in Pol I, Pol III and archaeal Pol because these enzymes contain subunits homologous to Rpb4-Rpb7. In addition to locking the position of the clamp, Rpb4-Rpb7 also provides a binding surface for other factors and possibly for RNA exiting the elongating Pol.

Structure of the Pol II elongation complex. Much insight into the mechanism of Pol II has been gained from structures of the elongating complex. The first structure determined was of Pol II transcribing a 3′-tailed template that had backtracked by one base with respect to the nucleotide addition site[80]. Several important details of the interaction of the enzyme with the nucleic acids were not visible, probably owing to heterogeneity of the complex. In new work, the elongation complex was instead assembled from a 5′-tailed DNA oligonucleotide and a nine-base RNA complementary to the 5′ tail, generating a complex in the post-translocation state[81]. The structure of this complex was determined to a resolution of 3.5 Å, revealing new important details of the mechanism of elongation. This structure clearly shows an 8-bp RNA-DNA hybrid and interactions of the enzyme with both ends of the transcription bubble as well as with the RNA-DNA hybrid (**Fig. 3c**). The Pol II loop termed lid seems to act as a wedge to drive apart the DNA and RNA strands at the upstream end of the transcription bubble and guide the RNA strand toward the RNA exit groove. The rudder loop interacts with single-stranded DNA after separation from the RNA strand, probably preventing reassociation with the exiting RNA. Finally, the newly revealed fork loop 1 interacts with the RNA-DNA hybrid, possibly stabilizing it. These three protein loops also interact with each other, forming a network of protein-protein and protein-nucleic acid interactions stabilizing the elongation complex.

The Pol II CTD molds itself to a binding partner. Pol II undergoes regulatory phosphorylation and dephosphorylation as part of the transcription cycle with the Rpb1 C-terminal domain (CTD) the target of this modification[18]. The CTD, which is unique to Pol II, contains 25–52 repeats of the tandemly repeated heptad sequence YSPTSPS, with both Ser2 and Ser5 the sites of phosphorylation. The CTD acts as a platform for assembly of factors that regulate transcription initiation, elongation, termination and mRNA processing. Pol II with a hypophosphorylated CTD is initially recruited to promoters during PIC formation and is phosphorylated at Ser5 during transcription initiation. Two cyclin

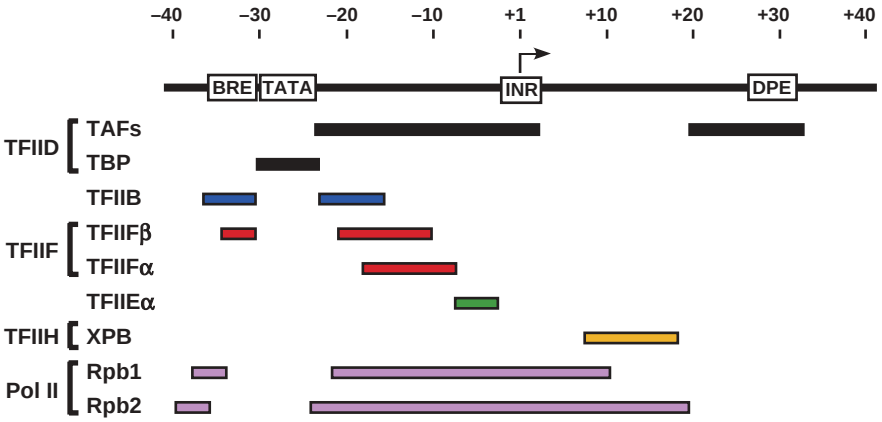

Figure 5 Summary of human general transcription factor protein-DNA crosslinks at a promoter. Top line represents promoter DNA with the position of functional elements indicated. Arrow represents the transcription start site. The bars below the line show the extent of protein-DNA crosslinking with the indicated general transcription factor using a crosslinker incorporated into the DNA backbone and crosslinked after treatment with Sarkosyl[98]. Use of a larger major groove crosslinking reagent without detergent challenge has shown even more extensive crosslinks across the promoter region[99,127].

dependent kinases, Cdk7 and Cdk8, are components of the PIC and target the CTD for phosphorylation[2,84]. Although previous work has suggested that only Cdk7 positively regulates transcription, new work in yeast indicates that both kinases can promote transcription *in vivo* and *in vitro*, as inhibition of both kinases together is required for maximal inhibition of transcription[84]. Phosphorylation of the CTD by these kinases also destabilizes the PIC, leading to formation of the scaffold complex. After initiation, other kinases such as Cdk9/Ctk1 phosphorylate Ser2 (Ser2-P), resulting in recruitment of the RNA processing, polyadenylation and termination factors to elongating Pol II, allowing coupling of transcription and RNA processing[18,85].

Until recently, the structural basis for CTD action was unclear as the many biochemically determined CTD-binding partners had no obvious structural relationship. New structures of CTD interactions with two different binding partners have revealed that the CTD seems to mold itself to its binding partner, adopting different conformations. A complex of a single copy of the CTD with Ser2-P and Ser5-P bound to the Pin1 peptidylproline isomerase has been solved[86]. This structure shows that the CTD binds as an extended coil, projecting every third residue onto one face of the coil. In another study, the structure of a four heptad Ser2-P CTD repeat has been solved in complex with the guanylyltransferase Cgt1 (ref. 87). In contrast to the CTD-Pin1 structure, 17 amino acids of the CTD repeats bound to an extended surface of Ctg1, anchored at both ends by electrostatic interactions with Ser5-P and with extensive hydrophobic CTD-Cgt1 interactions in between. This extensive surface contact between the CTD and Ctg1 suggests that mutation of any single residue would be unlikely to have a major effect on binding, a prediction borne out by mutagenesis studies[87]. The flexibility of the CTD, combined with covalent modification by phosphorylation, provides a way for the CTD to interact with multiple structurally dissimilar partners, a paradigm that may hold true for some transcription activators and their targets. For example, many activation regions are very insensitive to mutagenesis and the strength of the activator often is dependent on the simple length of the activation region[88–90], suggesting that some activators do not interact with their target as a folded globular domain. Interaction of these activators with an extended surface of their binding partners, similar to the interaction of the CTD with Ctg1, could explain these unusual properties of activators.

Structure of mediator and connection to Pol II function. Mediator is a large protein complex that interacts with Pol II, in part through the CTD, binding to the nonphosphorylated form. Mediator was first identified in yeast, where it is composed of 24 subunits and is essential for both basal and activated transcription in unfractionated systems[2]. The yeast mediator binds cooperatively with Pol II and a subset of the general factors at an intermediate step in PIC formation[35]. Related mediator complexes have been found in all eukaryotes examined, although the mediator subunits are the least conserved of all the members of the transcription machinery, consistent with the idea that many mediator subunits serve as regulatory factor targets[91,92]. Biochemical fractionation has shown that ~40% of mediator is in a stable complex with Pol II, consistent with studies showing that mediator can be recruited to promoters independent of the rest of the transcription machinery[93,94]. Mediator consists of three to four domains or modules[95]. In agreement with this, EM structures of the yeast Pol II–Mediator complex show three domains (head, middle and tail) with mediator binding centered on the Rpb3 and Rpb11 subunits, on the side opposite to the active site cleft[83] (**Fig. 4**). It is not yet known how the large mediator complex fits into the context of the rest of the transcription machinery nor how it transmits signals from regulatory factors to Pol II.

Assembly of Pol II with the transcription machinery

A major question that remains is how Pol assembles with the rest of the transcription machinery at a promoter. Important information on the architecture of the PIC comes from a study using photoreactive probes placed in promoter DNA and assembled into minimal PICs formed with TBP[96–99] (**Fig. 5**). These studies have shown that the transcription machinery makes extensive interactions with promoter DNA between positions –43 and +24 with respect to the transcription start site. Two RNA Pol II subunits (Rpb1 and 2) make extensive DNA interactions over 60 bp. TFIIB and the small subunit of TFIIF (TFIIFβ) both interact with DNA on either side of the TATA and the large TFIIF subunit interacts with DNA downstream of TATA. TFIIE interacts with promoter DNA just upstream of the transcription start site whereas the TFIIH helicase subunit interacts downstream and possibly upstream of the transcription start site. Any structural model for the PIC must account for these extensive protein-DNA interactions.

The Pol II-TFIIB interface. Recently, two lines of evidence have shown how TFIIB interacts with Pol II, which in one case led to a model for a complex of Pol II with TBP, TFIIB and DNA. In the first set of experiments, photocrosslinking and hydroxyl radical generating probes were placed on TFIIB near the functional surface of the ribbon domain and assembled into PICs[6]. Mapping the interaction of the ribbon domain with respect to the two largest Pol II subunits has shown that the ribbon domain fits into a pocket formed by the wall, dock and clamp domains near the RNA exit point. In this model, the functional surface of the ribbon interacts with the dock domain, a region of Rpb1 best conserved in Pol II, Pol III and archaea Pol, all of which use a TFIIB-like factor for initiation.

In new work from the Kornberg laboratory, the structure of a complex of TFIIB and the ten-subunit Pol II has been determined[79] (**Fig. 6a**). In this structure, the position of the TFIIB ribbon domain agrees with the binding seen in the complete PIC as described above. Additionally, the conserved portion of the linker between the TFIIB ribbon and core domains enters the RNA exit channel and into the active site cleft analogous to the path of σ region 3.2. This linker sequence forms a hairpin-like structure termed the B-finger that is predicted to be located very near the upstream end of the RNA-DNA hybrid in elongating Pol II. In this location, the B-finger would block productive elongation much like σ region 3.2, and may help stabilize or position the transcription bubble during open complex formation. The remainder of the TFIIB linker C-terminal to the B-finger was observed to exit back out though the RNA exit channel. The proposed location of the B-finger within the active site cleft is also consistent with recent results showing that in the open complex, archaeal TFB crosslinks to the template strand close to the transcription start site[100,101].

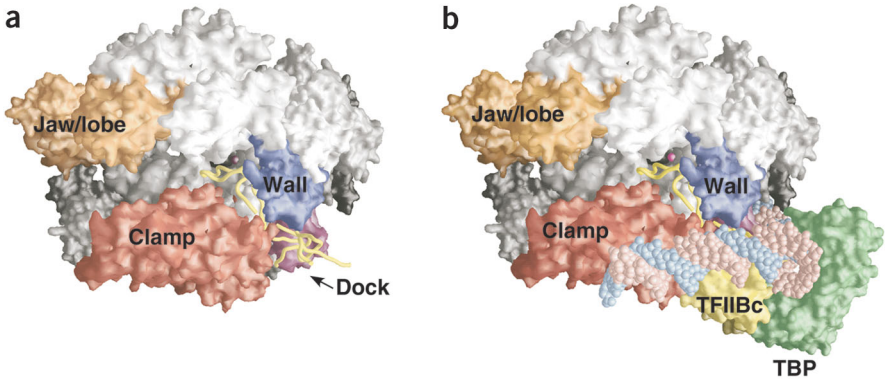

Figure 6 The Pol II–TFIIB complex and a model for the structure of the PIC. (**a**) X-ray structure of the Pol II–TFIIB complex with the TFIIB N-terminal domain backbone (yellow), and the clamp, wall, dock and jaw/lobe domains in red, dark blue, purple and orange[8], respectively. (**b**). Model for PIC structure based on the Pol II–TFIIB structure[8]. Colors are the same as in **a** with TBP (green), the N-terminal TFIIB core domain (TFIIBc; yellow) and DNA (red and blue).

In the Pol II-TFIIB structure, electron density from the TFIIB core domain is located adjacent to the Pol dock domain, on the opposite side of the ribbon domain interaction surface. Modeling the TFIIB-TBP-DNA structure to this location led to a model for a complex of Pol II with these factors and a predicted path of DNA in the preinitiation complex (**Fig. 6b**). In this model, TBP mediates promoter DNA bending around Pol II and the DNA downstream from TBP runs along the outer edge of the clamp element. This model will guide biochemical tests for interaction of the TFIIB core domain and promoter DNA with Pol II and will also serve as a template for modeling assembly of other general transcription factors. An unanswered question is whether the interaction of TFIIB differs between the Pol II–TFIIB complex and the PIC. Further structural and biochemical studies on higher-order assemblies of Pol II and the general factors are needed to answer this question.

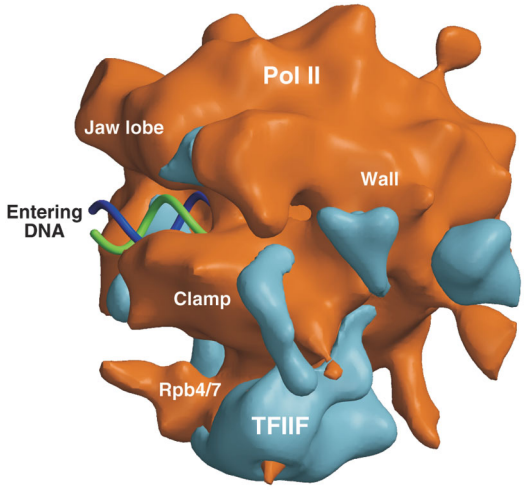

Figure 7 EM structure of the Pol II–TFIIF complex. Pol II is a orange surface with TFIIF density in blue. The orientation is similar to that of **Figures 3b** and **6**. Also shown is a model for downstream double-stranded DNA entering the active site cleft (blue and green helix). Figure was provided by F. Asturias.

Pol II and TFIIF interact over an extended surface. In other new work, cryo-EM has been used to determine the structure of yeast Pol II in complex with the general factor TFIIF[7] (**Fig. 7**). TFIIF binds Pol II as a heteromer and contains two subunits that are conserved among human, insects and yeast, termed Tfg1 and Tfg2 in yeast (Rap74 and Rap30 in humans). The N termini of both conserved subunits form a dimerization domain and the C termini of both subunits are winged helix domains[3]. Biochemical and structural analysis has implicated regions of TFIIF involved in protein-protein interactions with Pol II, TFIIB and the FCP1 phosphatase, as well as nonspecific protein-DNA interactions[102–104]. In the EM structure, TFIIF has been observed to interact with a highly extended surface of Pol II along the edge of the clamp element as well as with the Rpb4/7 subunits. The structure of Pol II with the Tfg2 subunit alone has implicated Tfg2 binding to the extended clamp region and Tfg1 binding to Rpb4/7, although this experiment should be interpreted cautiously as the Tfg1/2 dimerization domain is unlikely to fold normally with only one subunit[102].

Because the extended TFIIF domains along the clamp element are in a similar general location to that of σ in the bacterial holoenzyme, the authors speculate that TFIIF is the structural homolog of σ factor. Previous sequence comparison has suggested weak similarity between two regions of Tfg2/Rap30 and σ[105,106]. However, a more extensive comparison with many σ and Tfg2 family members does not reveal any marked similarity between these polypeptides (H.-T. Chen, Fred Hutchinson Cancer Research Center, and S.H., unpublished data). Nevertheless, both factors may play some of the same roles in the initiation process, such as helping to promote or stabilize opening of the DNA strands upon open complex formation.

Combining the TFIIB-Pol II structure model with the Pol II-TFIIF model, Asturias and co-workers propose a model for the structure of a minimal PIC in which TFIIF interacts with DNA downstream from the TATA and helps position DNA along the active site cleft of Pol II[7]. One complication in this proposal is that it does not seem to agree with protein-DNA crosslinking results that show a close overlap between promoter sequences upstream and downstream of the TATA contacted by both TFIIB and TFIIF[96,97,99]. This also suggests the possibility that the arrangement of the general factors on Pol II could change in higher-order assemblies with DNA, a possibility that can be addressed by further structural and biochemical studies.

TFIIE and TFIIH function in steps after PIC formation. The general factors TFIIE and TFIIH function primarily in steps after PIC formation and can be at least partially dispensable on promoters with a preformed transcription bubble[2]. TFIIE binds independently to Pol II[107] and is thought to stimulate both the kinase and helicase activities of TFIIH[2,108]. Biochemical analysis has suggested that TFIIE interacts with several other general transcription factors and may functionally interact with double-stranded and single-stranded promoter DNA. TFIIE is probably a heterodimer[109] and electron crystallography studies suggest that TFIIE binds near the Pol II cleft, consistent with the observed crosslinking of TFIIE to DNA immediately surrounding the transcription start site in the PIC[110]. The structure of the central core domain of the TFIIE β has been determined to be a winged helix domain by NMR[111]. It has been proposed that this domain interacts with single-stranded promoter DNA, but conclusive evidence for this has not been obtained.

TFIIH, which has a dual function in transcription and transcription-coupled DNA mismatch repair, is composed of two domains, a core domain containing two DNA helicase activities and a kinase domain termed CAK-containing Cdk7 (refs. 112,113). Mutations in the human XPD helicase cause the diseases xeroderma pigmentosum and trichothiodystrophy. These mutations affect nucleotide excision repair and can also affect basal transcription and transcription activated by certain nuclear receptors[114]. The low-resolution structure of the TFIIH core domain has been determined by EM and is ring-shaped with the two helicase subunits located on either side of a prominent protrusion[115,116]. The center of the ring seems to have dimensions sufficient to accommodate double-stranded DNA, although it is not clear if DNA normally enters the ring. Models for how TFIIH fits into the PIC are speculative as there is no information available on the docking of TFIIH with any other factor. The XPB helicase is essential for open complex formation and is the only TFIIH subunit seen to crosslink to promoter DNA.

Open complex formation and the transcription start site

A major unanswered question about the mechanism of Pol II initiation is how melting of the DNA strands is initiated during open complex formation. Pol II is unique among cellular Pols in requiring the action of an ATP-dependent DNA helicase (XPB) for open complex formation. This requirement is puzzling as all cellular Pols have the same overall structure and catalytic mechanism. Aside from Pol II, archaeal and other eukaryotic Pols may use a mechanism similar to that seen in bacteria in which aromatic side chains from one of the general transcription factors act as a wedge to stabilize separation of the two DNA strands[77].

Helicase requirement for opening the transcription bubble. Although helicase activity is typically defined *in vitro* as the ability to processively remove a paired oligonucleotide from single-stranded DNA, not all helicases act processively. Helicases act by destabilizing double-stranded nucleic acids through the ATP hydrolysis–dependent motion of two separate domains that interact with single- and double-stranded nucleic acids[117]. Using this mechanism, the XPB helicase that binds to promoter DNA as a subunit of TFIIH probably initiates unwinding by introducing torsional strain in the DNA near the transcription start site. Because of uncertainty in the location of the XPB-DNA interaction[3], the mechanism of helicase action is also uncertain. If the XPB helicase motifs bind at the site of single-stranded bubble formation, XPB would directly initiate strand unwinding. If the helicase domain interacts only with downstream DNA, then this torsional strain would lead to initial DNA opening upstream from the point of destabilization. In either case, one or more of the general transcription factors likely acts like σ region 3.1 to trap the single-stranded bubble and promote the insertion of this single-stranded region into the active site of the enzyme. From the EM structure and protein-DNA crosslinking studies, TFIIE or the TFIIF subunit Tfg2 may be in a position to promote this reaction. Understanding this reaction will require more precise localization of the XPB helicase domain at the promoter and the identification of amino acids in general transcription factors located near the initial site of DNA melting. Also unexplained is how the related archaeal factors TBP and Tfb promote open complex formation without the requirement for any other general factor.

Transcription start site selection. Another major unsolved problem is understanding how the start site of transcription is selected. At TATA-containing promoters in vertebrates and *D. melanogaster*, the transcription start site is located ~30 bp downstream from the beginning of the TATA sequence. However, at *Saccharomyces cerevisiae* TATA-containing promoters, the TATA seems to define a window of ~40–120 bp in which transcription starts at preferred DNA sequences[1]. Transcription start sites in *Schizosaccharomyces pombe* are less heterogeneous, with initiation beginning 25-40 bp downstream from the TATA element[118]. A model to explain the transcription initiation site in higher eukaryotes would be that TFIIB binding to both Pol II and promoter DNA sets the distance needed for the DNA to travel from the TFIIB-binding site on Pol II to the active site of the enzyme[8]. It is not obvious why this would be different in yeasts in which the transcription machinery is largely conserved. Extensive genetic and biochemical studies in *S. cerevisiae* have identified mutations in Pol II subunits, TFIIB and TFIIF that alter the transcription start site[119–123].

Based on the Pol II-TFIIB structure, it has been proposed that the tip of the B-finger might play a role in recognition of the transcription start site[8]. This residue is not conserved between yeasts and human and it has been postulated that, lacking a stable protein-DNA interaction, the promoter DNA would slip through the enzyme active site until a sequence that stably bound to the active site was located. This model is consistent with *in vivo* mapping experiments of *S. cerevisiae* promoters that have found evidence of single-stranded DNA over a wide region between the TATA and transcription start site[124]. Ultimately, determining the mechanism of start site selection will involve mapping the location of promoter DNA in the PIC both before and after ATP addition.

Concluding remarks

As exemplified by the similarities in cellular RNA polymerases, the general mechanism of transcription is similar in all cells. Despite this overall conservation, the transcription machinery is much more complex in eukaryotes, with the function of bacterial σ factor distributed among several general transcription factors. In eukaryotes, the transcription start site is determined in part by the precise binding of the transcription machinery to a promoter and this seems driven not by any one high-affinity protein-DNA interaction, but rather by multiple low-affinity and low-specificity protein-DNA interactions. Because all transcription requires TBP but many promoters do not have a recognizable TATA element, the precise role of TBP in nucleating the assembly of the PIC at these non-TATA promoters is an open question. Although much progress has been made in structural analysis, an important challenge is to determine the structure of the PIC to test whether structures and models of single general transcription factor–Pol II complexes reflect the structure of the much larger PIC. Also important will be to determine the mechanism of the marked conformational changes that accompany transition to the open com-

plex state. This mechanism, whereby single-stranded DNA is positioned in the active site of Pol, is one of the most mysterious aspects of both bacterial and eukaryotic transcription. In future studies, integration of the structures discussed above with biochemical and structural work on activators and their targets will get at the heart of the mechanism of gene regulation and will examine in atomic detail how regulatory signals are transmitted to the transcription machinery.

ACKNOWLEDGMENTS

I thank D. Bushnell, R. Kornberg, and F. Asturias for figures and for communication of results before publication, E. Nogales for figures, B. Moorefield and H.-T. Chen for their comments on the manuscript and H.-T. Chen for discussions and help with figures. This work was supported by a grant from the US National Institutes of Health. S.H. is an associate investigator of the Howard Hughes Medical Institute.

COMPETING INTERESTS STATEMENT

The author declares that he has no competing financial interests.

Received 12 November 2003; accepted 22 March 2004
Published online at http://www.nature.com/natstructmolbiol/

1. Hampsey, M. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol. Mol. Biol. Rev.* **62**, 465–503 (1998).
2. Lee, T.I. & Young, R.A. Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* **34**, 77–137 (2000).
3. Woychik, N.A. & Hampsey, M. The RNA polymerase II machinery: structure illuminates function. *Cell* **108**, 453–463 (2002).
4. Borukhov, S. & Nudler, E. RNA polymerase holoenzyme: structure, function and biological implications. *Curr. Opin. Microbiol.* **6**, 93–100 (2003).
5. Murakami, K.S. & Darst, S.A. Bacterial RNA polymerases: the wholo story. *Curr. Opin. Struct. Biol.* **13**, 31–39 (2003).
6. 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**, 437–447 (2003).
7. Chung, W.H. *et al.* RNA Polymerase II/TFIIF Structure and Conserved Organization of the Initiation Complex. *Mol. Cell* **12**, 1003–1013 (2003).
8. 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 Å. *Science* **303**, 983–988 (2004).
9. Bell, S.D. & Jackson, S.P. Transcription in *Archaea*. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 41–51 (1998).
10. Ebright, R.H. RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J. Mol. Biol.* **304**, 687–698 (2000).
11. Schramm, L. & Hernandez, N. Recruitment of RNA polymerase III to its target promoters. *Genes Dev.* **16**, 2593–2620 (2002).
12. Grummt, I. Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev.* **17**, 1691–1702 (2003).
13. Ptashne, M. & Gann, A. *Genes and Signals* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2002).
14. Cosma, M.P. Ordered recruitment: gene-specific mechanism of transcription activation. *Mol. Cell* **10**, 227–236 (2002).
15. Wang, W., Carey, M. & Gralla, J.D. Polymerase II promoter activation: closed complex formation and ATP-driven start site opening. *Science* **255**, 450–453 (1992).
16. Luse, D.S. & Jacob, G.A. Abortive initiation by RNA polymerase II *in vitro* at the adenovirus 2 major late promoter. *J. Biol. Chem.* **262**, 14990–14997 (1987).
17. Holstege, F.C.P., Fiedler, U. & Timmers, H.T.M. Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J.* **16**, 7468–7480 (1997).
18. Bentley, D. The mRNA assembly line: transcription and processing machines in the same factory. *Curr. Opin. Cell Biol.* **14**, 336–342 (2002).
19. Yudkovsky, N., Ranish, J.A. & Hahn, S. A transcription reinitiation intermediate that is stabilized by activator. *Nature* **408**, 225–229 (2000).
20. Smale, S.T. & Kadonaga, J.T. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* **72**, 449–479 (2003).
21. Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520 (1993).
22. Kim, J.L., Nikolov, D.B. & Burley, S.K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527 (1993).
23. Cox, J.M. *et al.* Bidirectional binding of the TATA box binding protein to the TATA box. *Proc. Natl. Acad. Sci. USA* **94**, 13475–13480 (1997).
24. Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D. & Ebright, R.H. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev.* **12**, 34–44 (1998).
25. Qureshi, S.A. & Jackson, S.P. Sequence-specific DNA binding by the *S. shibatae* TFIIB homolog, TFB, and its effect on promoter strength. *Mol. Cell* **1**, 389–400 (1998).
26. Bell, S.D., Kosa, P.L., Sigler, P.B. & Jackson, S.P. Orientation of the transcription preinitiation complex in archaea. *Proc. Natl. Acad. Sci. USA* **96**, 13662–13667 (1999).
27. Littlefield, O., Korkhin, Y. & Sigler, P.B. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc. Natl. Acad. Sci. USA* **96**, 13668–13673 (1999).
28. Chalkley, G.E. & Verrijzer, C.P. DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250–TAF(II)150 complex recognizes the initiator. *EMBO J.* **18**, 4835–4845 (1999).
29. Oelgeschlager, T., Chiang, C.-M. & Roeder, R.G. Topology and reorganization of a human TFIID-promoter complex. *Nature* **382**, 735–738 (1996).
30. Burke, T.W. & Kadonaga, J.T. The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes Dev.* **11**, 3020–3031 (1997).
31. Butler, J.E. & Kadonaga, J.T. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev.* **15**, 2515–2519 (2001).
32. Ohler, U., Liao, G.C., Niemann, H. & Rubin, G.M. Computational analysis of core promoters in the *Drosophila* genome. *Genome Biol.* **3**, 0087.1–0087.12 (2002).
33. Patikoglou, G.A. *et al.* TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* **13**, 3217–3230 (1999).
34. Wobbe, C.R. & Struhl, K. Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription *in vitro*. *Mol. Cell. Biol.* **10**, 3859–3867 (1990).
35. 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).
36. Martinez, E. *et al.* Core promoter-specific function of a mutant transcription factor TFIID defective in TATA-box binding. *Proc. Natl. Acad. Sci. USA* **92**, 11864–11868 (1995).
37. Hochheimer, A. & Tjian, R. Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. *Genes Dev.* **17**, 1309–1320 (2003).
38. Davidson, I. The genetics of TBP and TBP-related factors. *Trends Biochem. Sci.* **28**, 391–398 (2003).
39. Holmes, M.C. & Tjian, R. Promoter-selective properties of the TBP-related factor TRF1. *Science* **288**, 867–870 (2000).
40. Takada, S., Lis, J.T., Zhou, S. & Tjian, R. A TRF1:BRF complex directs *Drosophila* RNA polymerase III transcription. *Cell* **101**, 459–469 (2000).
41. Nikolov, D.B. *et al.* Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377**, 119–128 (1995).
42. Geiger, J.H., Hahn, S., Lee, S. & Sigler, P.B. Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science* **272**, 830–836 (1996).
43. Tan, S., Hunziker, Y., Sargent, D.F. & Richmond, T.J. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* **381**, 127–134 (1996).
44. Weideman, C.A. *et al.* Dynamic interplay of TFIIA, TBP, and TATA DNA. *J. Mol. Biol.* **271**, 61–75 (1997).
45. Kokubo, T., Swanson, M.J., Nishikawa, J.I., Hinnebusch, A.G. & Nakatani, Y. The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA-binding protein. *Mol. Cell. Biol.* **18**, 1003–1012 (1998).
46. Liu, D. *et al.* Solution structure of a TBP-TAF(II)230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP. *Cell* **94**, 573–583 (1998).
47. Sanders, S.L., Garbett, K.A. & Weil, P.A. Molecular characterization of *Saccharomyces cerevisiae* TFIID. *Mol. Cell. Biol.* **22**, 6000–6013 (2002).
48. Chi, T., Lieberman, P., Ellwood, K. & Carey, M. A general mechanism for transcriptional synergy by eukaryotic activators. *Nature* **377**, 254–257 (1995).
49. Pardee, T.S., Bangur, C.S. & Ponticelli, A.S. 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.* **273**, 17859–17864 (1998).
50. Hahn, S. & Roberts, S. The zinc ribbon domains of the general transcription factors TFIIB and Brf: conserved functional surfaces but different roles in transcription initiation. *Genes Dev.* **14**, 719–730 (2000).
51. Albright, S.R. & Tjian, R. TAFs revisited: more data reveal new twists and confirm old ideas. *Gene* **242**, 1–13 (2000).
52. Green, M.R. TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes. *Trends Biochem. Sci.* **25**, 59–63 (2000).
53. Tora, L. A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev.* **16**, 673–675 (2002).
54. Chen, J.-L., Attardi, L.D., Verrijzer, C.P., Yokomori, K. & Tjian, R. Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell* **79**, 93–105 (1994).
55. Thut, C.J., Chen, J.L., Klemm, R. & Tjian, R. p53 transcriptional activation mediated by coactivators $TAF_{II}40$ and $TAF_{II}60$. *Science* **267**, 100–104 (1995).
56. Wassarman, D.A. & Sauer, F. TAF(II)250: a transcription toolbox. *J. Cell Sci.* **114**, 2895–2902 (2001).
57. Andel, F. 3rd, Ladurner, A.G., Inouye, C., Tjian, R. & Nogales, E. Three-dimensional structure of the human TFIID-IIA-IIB complex. *Science* **286**, 2153–2156 (1999).
58. Brand, M., Leurent, C., Mallouh, V., Tora, L. & Schultz, P. Three-dimensional structures of the TAFII-containing complexes TFIID and TFTC. *Science* **286**, 2151–2153 (1999).
59. Xie, X. *et al.* Structural similarity between TAFs and the heterotetrameric core of the histone octamer. *Nature* **380**, 316–322 (1996).
60. Werten, S. *et al.* Crystal structure of a subcomplex of human transcription factor TFIID formed by TATA binding protein-associated factors hTAF4 (hTAF(II)135) and hTAF12 (hTAF(II)20). *J. Biol. Chem.* **277**, 45502–45509 (2002).
61. Gangloff, Y.G., Romier, C., Thuault, S., Werten, S. & Davidson, I. The histone fold is a key structural motif of transcription factor TFIID. *Trends Biochem. Sci.* **26**,

250–257 (2001).
62. Leurent, C. *et al.* Mapping histone fold TAFs within yeast TFIID. *EMBO J.* **21**, 3424–3433 (2002).
63. Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. & Richmond, T.J. Crystal structure of the nucleosome core particle at 2.9Å resolution. *Nature* **389**, 251–260 (1997).
64. Freiman, R.N. *et al.* Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development. *Science* **293**, 2084–2087 (2001).
65. Hiller, M.A., Lin, T.Y., Wood, C. & Fuller, M.T. Developmental regulation of transcription by a tissue-specific TAF homolog. *Genes Dev.* **15**, 1021–1030 (2001).
66. Shen, W.C. *et al.* Systematic analysis of essential yeast TAFs in genome-wide transcription and preinitiation complex assembly. *EMBO J.* **22**, 3395–3402 (2003).
67. Kuras, L., Kosa, P., Mencia, M. & Struhl, K. TAF-containing and TAF-independent forms of transcriptionally active TBP *in vivo*. *Science* **288**, 1244–1248 (2000).
68. Li, X.Y., Bhaumik, S.R. & Green, M.R. Distinct classes of yeast promoters revealed by differential TAF recruitment. *Science* **288**, 1242–1244 (2000).
69. Lee, T.I. *et al.* Redundant roles for the TFIID and SAGA complexes in global transcription. *Nature* **405**, 701–704 (2000).
70. Zhang, G. *et al.* Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**, 811–824 (1999).
71. Cramer, P. *et al.* Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288**, 640–649 (2000).
72. Cramer, P., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* **292**, 1863–1876 (2001).
73. Opalka, N. *et al.* Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell* **114**, 335–345 (2003).
74. Mekler, V. *et al.* Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **108**, 599–614 (2002).
75. Murakami, K.S., Masuda, S. & Darst, S.A. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* **296**, 1280–1284 (2002).
76. Vassylyev, D.G. *et al.* Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417**, 712–719 (2002).
77. 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**, 1285–1290 (2002).
78. Armache, K.J., Kettenberger, H. & Cramer, P. Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **100**, 6964–6968 (2003).
79. Bushnell, D.A. & Kornberg, R.D. Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: implications for the initiation of transcription. *Proc. Natl. Acad. Sci. USA* **100**, 6969–6973 (2003).
80. Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001).
81. Westover, K.D., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* **303**, 1014–1016 (2004).
82. Kettenberger, H., Armache, K.J. & Cramer, P. Architecture of the RNA polymerase II–TFIIS complex and implications for mRNA cleavage. *Cell* **114**, 347–357 (2003).
83. Davis, J.A., Takagi, Y., Kornberg, R.D. & Asturias, F.A. Structure of the yeast RNA polymerase II holoenzyme: mediator conformation and polymerase interaction. *Mol. Cell* **10**, 409–415 (2002).
84. Liu, Y. *et al.* Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex. *Mol. Cell. Biol.* **24**, 1721–1735 (2004).
85. Ahn, S.H., Kim, M. & Buratowski, S. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. *Mol. Cell* **13**, 67–76 (2004).
86. Verdecia, M.A., Bowman, M.E., Lu, K.P., Hunter, T. & Noel, J.P. Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat. Struct. Biol.* **7**, 639–643 (2000).
87. Fabrega, C., Shen, V., Shuman, S. & Lima, C.D. Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Mol. Cell* **11**, 1549–1561 (2003).
88. Hope, I.A., Mahadevan, S. & Struhl, K. Structural and functional characterization of the short acidic transcriptional activation region of yeast GCN4 protein. *Nature* **333**, 635–640 (1988).
89. Cress, W.D. & Triezenberg, S.J. Critical structural elements of the VP16 transcriptional activation domain. *Science* **251**, 87–90 (1991).
90. Jackson, B.M., Drysdale, C.M., Natarajan, K. & Hinnebusch, A.G. Identification of seven hydrophobic clusters in GCN4 making redundant contributions to transcriptional activation. *Mol. Cell. Biol.* **16**, 5557–5571 (1996).
91. Malik, S. & Roeder, R.G. Transcriptional regulation through mediator-like coactivators in yeast and metazoan cells. *Trends Biochem. Sci.* **25**, 277–283 (2000).
92. Boube, M., Joulia, L., Cribbs, D.L. & Bourbon, H.M. Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. *Cell* **110**, 143–151 (2002).
93. Cosma, M.P., Tanaka, T. & Nasmyth, K. Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell* **97**, 299–311 (1999).
94. Rani, P.G., Ranish, J.A. & Hahn, S. RNA polymerase II (Pol II)-TFIIF and Pol II-mediator complexes: the major stable Pol II complexes and their activity in transcription initiation and reinitiation. *Mol. Cell. Biol.* **24**, 1709–1720 (2004).
95. Kang, J.S. *et al.* The structural and functional organization of the yeast mediator complex. *J. Biol. Chem.* **276**, 42003–42010 (2001).
96. Forget, D. *et al.* RAP74 induces promoter contacts by RNA polymerase II upstream and downstream of a DNA bend centered on the TATA box. *Proc. Natl. Acad. Sci. USA* **94**, 7150–7155 (1997).
97. Kim, T.-K. *et al.* Trajectory of DNA in the RNA polymerase II transcription preinitiation complex. *Proc. Natl. Acad. Sci. USA* **94**, 12268–12273 (1997).
98. Kim, T.K., Ebright, R.H. & Reinberg, D. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science* **288**, 1418–1422 (2000).
99. 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**, 1122–1131 (2004).
100. Bartlett, M.S., Thomm, M. & Geiduschek, E.P. Topography of the euryarchaeal transcription initiation complex. *J. Biol. Chem.* **279**, 5894–5903 (2004).
101. Renfrow, M.B. *et al.* Transcription factor B contacts promoter DNA near the transcription start site of the archaeal transcription initiation complex. *J. Biol. Chem.* **279**, 2825–2831 (2004).
102. Gaiser, F., Tan, S. & Richmond, T.J. Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 Å resolution. *J. Mol. Biol.* **302**, 1119–1127 (2000).
103. Kamada, K., Roeder, R.G. & Burley, S.K. Molecular mechanism of recruitment of TFIIF-associating RNA polymerase C-terminal domain phosphatase (FCP1) by transcription factor IIF. *Proc. Natl. Acad. Sci. USA* **100**, 2296–2299 (2003).
104. Nguyen, B.D. *et al.* NMR structure of a complex containing the TFIIF subunit RAP74 and the RNA polymerase II carboxyl-terminal domain phosphatase FCP1. *Proc. Natl. Acad. Sci. USA* **100**, 5688–5693 (2003).
105. Sopta, M., Burton, Z.F. & Greenblatt, J. Structure and associated DNA-helicase activity of a general transcription initiation factor that binds to RNA polymerase II. *Nature* **341**, 410–414 (1989).
106. Henry, N.L. *et al.* TFIIF-TAF-RNA polymerase II connection. *Genes Dev.* **8**, 2868–2878 (1994).
107. Bushnell, D.A., Bamdad, C. & Kornberg, R.D. A minimal set of RNA Pol II transcription protein interactions. *J. Biol. Chem.* **271**, 20170–20174 (1996).
108. Ohkuma, Y. Multiple functions of general transcription factors TFIIE and TFIIH in transcription: possible points of regulation by trans-acting factors. *J. Biochem.* **122**, 481–489 (1997).
109. Sayre, M.H., Tschochner, H. & Kornberg, R.D. Purification and properties of *S. cerevisiae* RNA polymerase II general initiation factor a. *J. Biol. Chem.* **267**, 23383–23387 (1992).
110. 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**, 773–779 (1996).
111. Okuda, M. *et al.* Structure of the central core domain of TFIIEβ with a novel double-stranded DNA-binding surface. *EMBO J.* **19**, 1346–1356 (2000).
112. Coin, F. & Egly, J.M. Ten years of TFIIH. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 105–110 (1998).
113. Takagi, Y. *et al.* Revised subunit structure of yeast TFIIH and reconciliation with human TFIIH. *J. Biol. Chem.* **278**, 43897–43900 (2003).
114. Dubaele, S. *et al.* Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients. *Mol. Cell* **11**, 1635–1646 (2003).
115. Schultz, P. *et al.* Molecular structure of human TFIIH. *Cell* **102**, 599–607 (2000).
116. Chang, W.H. & Kornberg, R.D. Electron crystal structure of the transcription factor and DNA repair complex, core TFIIH. *Cell* **102**, 609–613 (2000).
117. Caruthers, J.M. & McKay, D.B. Helicase structure and mechanism. *Curr. Opin. Struct. Biol.* **12**, 123–133 (2002).
118. Choi, W.S., Yan, M., Nusinow, D. & Gralla, J.D. *In vitro* transcription and start site selection in *Schizosaccharomyces pombe*. *J. Mol. Biol.* **319**, 1005–1013 (2002).
119. Hekmatpanah, D.S. & Young, R.A. Mutations in a conserved region of RNA polymerase II influence the accuracy of mRNA start site selection. *Mol. Cell. Biol.* **11**, 5781–5791 (1991).
120. Berroteran, R.W., Ware, D.E. & Hampsey, M. The sua8 suppressors of *S. cerevisiae* encode replacements of conserved residues within the largest subunit of RNA polymerase II and affect transcription start site selection similarly to sua7 (TFIIB) mutations. *Mol. Cell. Biol.* **14**, 226–237 (1994).
121. Pinto, I., Ware, D.E. & Hampsey, M. The yeast SUA7 gene encodes a homolog of human transcription factor TFIIB and is required for normal start site selection *in vivo*. *Cell* **68**, 977–988 (1992).
122. 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**, 4427–4440 (2001).
123. Sun, Z.W., Tessmer, A. & Hampsey, M. Functional interaction between TFIIB and the Rpb9 (Ssu73) subunit of RNA polymerase II in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **24**, 2560–2566 (1996).
124. Giardina, C. & Lis, J.T. DNA melting on yeast RNA polymerase II promoters. *Science* **261**, 759–762 (1993).
125. Chen, H.T., Legault, P., Glushka, J., Omichinski, J.G. & Scott, R.A. Structure of a (Cys3His) zinc ribbon, a ubiquitous motif in archaeal and eucaryal transcription. *Protein Sci.* **9**, 1743–1752 (2000).
126. Bushnell, D.A., Cramer, P. & Kornberg, R.D. Structural basis of transcription: α-amanitin-RNA polymerase II cocrystal at 2.8 Å resolution. *Proc. Natl. Acad. Sci. USA* **99**, 1218–1222 (2002).
127. 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**, 8168–8177 (2000).