HHS Public Access

Author manuscript

Mol Cell. Author manuscript; available in PMC 2023 February 03.

Published in final edited form as:

Mol Cell. 2022 February 03; 82(3): 660–676.e9. doi:10.1016/j.molcel.2021.12.020.

Structural visualization of de novo transcription initiation by *Saccharomyces cerevisiae* RNA polymerase II

Chun Yang[1,†], **Rina Fujiwara**[1,2,†], **Hee Jong Kim**[1,2,3], **Pratik Basnet**[4], **Yunye Zhu**[4], **Jose J. Gorbea Colón**[1,2], **Stefan Steimle**[1], **Benjamin A. Garcia**[1,3], **Craig D. Kaplan**[4], **Kenji Murakami**[1,5,*]

[1]Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, U.S.A.

[2]Biochemistry and Molecular Biophysics Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

[3]Epigenetics Institute, Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

[4]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA.

[5]Lead contact

Summary:

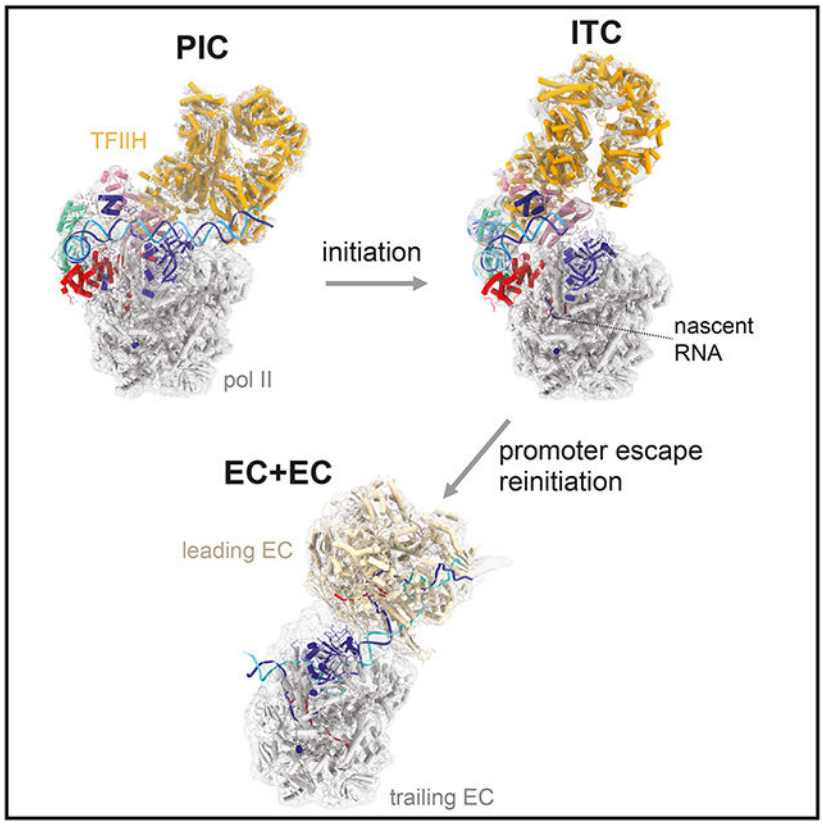
Previous structural studies of the initiation-elongation transition of RNA polymerase II (pol II) transcription have relied on the use of synthetic oligonucleotides, often artificially discontinuous to capture pol II in the initiating state. Here we report multiple structures of initiation complexes converted *de novo* from a 33-subunit yeast pre-initiation complex (PIC) through catalytic activities and subsequently stalled at different template positions. We determine that PICs in the initially-transcribing state (ITC) can synthesize a transcript of ~26 nucleotides before transitioning to an elongation complex (EC) as determined by the loss of general transcription factors (GTFs). Unexpectedly, transition to an EC was greatly accelerated when an ITC encountered a downstream EC stalled at promoter proximal regions, and resulted in a collided head-to-end dimeric EC complex. Our structural analysis reveals a dynamic state of TFIIH, the largest of GTFs, in PIC/ITC with distinct functional consequences at multiple steps on the pathway to elongation.

Graphical Abstract

*Correspondence to: kenjim@pennmedicine.upenn.edu.
†Contributed equally to this work.

Author contributions: C.Y. and R.F. conducted the experiments, C.Y., R.F.,H.J.K. performed cryo-EM image analysis, C.Y., R.F., S.S., J.J.G., H.J.K., and K.M. conducted model building, P.B., Y.Z., and C.D.K. performed genetic analyses. C.Y., R.F., B.A.G., and K.M. prepared figures, designed the study, and wrote the manuscript. C.D.K. edited the manuscript.

Declaration of interests: Authors declare no competing interests.



Keywords

Transcription initiation; promoter escape; TFIIH; RNA polymerase II; pre-initiation complex; initially-transcribing complex; elongation complex; promoter opening; start-site scanning; yeast

Introduction

RNA polymerase II (pol II) and the six general transcription factors (GTFs) assemble in a transcription pre-initiation complex (PIC) on promoter DNA upstream of transcription start sites (TSSs), open the double-stranded DNA, and select a TSS (Conaway and Conaway, 1993; Kornberg, 2007). Following TSS recognition, the PIC transitions to an initially-transcribing complex (ITC), and subsequently transitions to an elongation complex (EC) upon promoter escape, which entails loss of GTFs and, in vivo, recruitment of elongation factors. After transcription initiation is completed on the order of a few seconds (Baek et al., 2021; Nguyen et al., 2021; Rosen et al., 2020), a subset of GTFs may remain at the promoter, forming a platform for assembly of a following round of transcription (Yudkovsky et al., 2000). This set of transitions is universal across all eukaryotes and overlaid by many additional regulatory factors in initiation, such as Mediator, and elongation factors (e.g. Spt4/5, known as DSIF in metazoans, among others) (Adelman and Lis, 2012; Conaway and Conaway, 2012; Wade and Struhl, 2008). Moreover, transcription occurs in discontinuous “bursts” (Bahar Halpern et al., 2015; Chubb et al., 2006; Donovan et al., 2019; Lenstra et al., 2015; Zenklusen et al., 2008), where polymerases from successive rounds of transcription

initiate within short time intervals followed by periods where promoters are not active. In metazoans, the bursting phenomenon is further regulated by so-called promoter-proximal pausing (Bartman et al., 2019; Gressel et al., 2017; Shao and Zeitlinger, 2017) yet the nature and spectrum of functional interactions between successive transcribing polymerases remain unknown (Ehrensberger et al., 2013).

The largest GTF TFIIH, comprising 10 subunits, is an integral component of the PIC; the translocase subunit Ssl2 (XPB in humans) acts as a molecular motor during promoter opening, TSS scanning in yeast, and initial RNA chain elongation (Bradsher et al., 2000; Dvir et al., 1997; Fazal et al., 2015; Fishburn et al., 2015; Qiu et al., 2020; Spangler et al., 2001; Zhao et al., 2021). The other subunits comprise the six-subunit structural core and the three-subunit kinase termed TFIIK, which performs pol II CTD phosphorylation (Feaver et al., 1994; Gibbons et al., 2012; Greber et al., 2017; Greber et al., 2019; Kokic et al., 2019; Luo et al., 2015; Schultz et al., 2000; Svejstrup et al., 1995; van Eeuwen et al., 2021a; van Eeuwen et al., 2021b). Structural studies of open promoter complexes provided information about locations of GTFs and the DNA path (Aibara et al., 2021; He et al., 2013; He et al., 2016; Plaschka et al., 2016; Schilbach et al., 2021; Schilbach et al., 2017). Open complexes in these studies have utilized DNA duplexes opened due to sequence non-complementarity, thus a TFIIH-dependent natural open complex has yet to be described. Additionally, some biochemical data suggest TFIIH is required not only for promoter opening but also for subsequent events including initial RNA chain elongation (Bradsher et al., 2000; Dvir et al., 1997; Fazal et al., 2015; Fujiwara et al., 2019; Luse, 2013, 2019; Spangler et al., 2001), which is not explained by a model, predicted by structural studies, that promoter escape is completed by ejection of TFIIB from the RNA exit tunnel of pol II (Bushnell et al., 2004; Sainsbury et al., 2013).

We have recently developed an in vitro transcription system, in which pol II and six GTFs (TFIIA, TBP, TFIIB, TFIIE, TFIIF, and TFIIH) isolated from the yeast *Saccharomyces cerevisiae*, melt double-stranded promoter DNA, and initiate RNA synthesis de novo with high efficiency (Fujiwara and Murakami, 2019; Murakami et al., 2013a; Murakami et al., 2015a). Resulting post-initiation complexes could be stalled at different template positions through omission of GTP through a series of G-less promoter templates and isolated in abundant homogeneous form by glycerol gradient sedimentation (Fujiwara et al., 2019). This system allowed us to determine the timing of promoter escape and the composition of protein complexes associated with different lengths of RNA. Almost all of the post-initiation complexes retained the GTFs when pol II was stalled at positions before ~+27 relative to the TSS +1, whereas, when stalled after ~+45, most complexes had completed promoter escape as defined by loss of GTFs, followed by assembly of a follow-on pre-initiation complex (PIC) at the promoter for re-initiation.

Here we report cryo-EM structures of such de novo initiation complexes on two different templates (Figure 1A). Structural analysis with a “G-less 26” template, which has a G-free region between the TSS (+1) and a G-stop at +26 of the *SNR20* promoter (Fujiwara et al., 2019), revealed an ITC containing all the GTFs, pol II, and a nascent RNA on the open promoter DNA as well as three distinct forms of the PIC. Importantly, these ITCs retained GTFs even though 26 nt of RNA had been synthesized. By contrast, when polymerases

were allowed to transcribe further by use of a “G-less 49” template, which has a G-free region between the TSS (+1) and a G-stop at +49, revealed successive early elongation complexes (EC+EC), in which two polymerases that completed promoter escape were in close contact with each other. A combination of these structures with previous biochemical and biophysical studies (Fazal et al., 2015; Fujiwara et al., 2019) point to dissociation of TFIIH from pol II (Rpb7) and from the DNA in the ITC as key steps in the transition from initiation to elongation.

Results

Isolation of bona fide PICs and ITCs and Cryo-EM analysis

In order to understand the structural basis of early phases of initiation and promoter clearance, we obtained ITCs by de novo transcription using a G-less 26 *SNR20* promoter fragment template in the absence of elongation factors. This procedure followed our previously established approach generating very high efficiency initiation on a per template basis, enabling structural characterization of complexes (Fujiwara et al., 2019) (Figure 1A). Briefly, our initiation complex assembly entails combining a G-less *SNR20* promoter fragment and excess pol II and GTFs relative to DNA in a solution of high ionic strength, dilution to physiologic conditions allowing PIC assembly, followed by addition of NTPs to initiate promoter opening by TFIIH and RNA synthesis by pol II. We also added Sub1, the yeast homolog of PC4 (Bardwell et al., 1994; Henry et al., 1996), which slightly increased the activity (Fujiwara et al., 2019). Sub1 is recruited to PICs in vivo in yeast (Sikorski et al., 2011) and deletion of *SUB1* confers defects in start site selection and has extensive genetic interactions with initiation factors (Braberg et al., 2013; Zhao et al., 2021). ITCs were stalled at position +26 relative to the TSS (+1) by use of chain-terminating 3’-O-methyl GTP instead of normal 3’-OH GTP. We note that G-less 26 complexes were extremely susceptible to long-range backtracking. Inclusion of 4’-thio UTP instead of normal UTP partially arrested pol II in the ITC, which otherwise would completely revert to a closed complex (PIC) (Fujiwara et al., 2019). Reaction mixtures were sedimented in a 10-40% glycerol gradient to remove free nucleotides and excess free GTFs and pol II. The resulting ITCs contained GTFs and pol II, which were apparently stoichiometric, and transcripts of ~20-26 nucleotides initiating from positions +1 to +7 (Figure 1A, S1A-B). Due to similarity in size, ITCs were not separable on the gradient from residual PICs that did not engage in transcription, those that collapsed back from ITCs, or both.

Aliquots of peak fractions were embedded in vitreous ice, imaged by cryo-EM, and found to contain multiple particle populations as expected for an active rection mixture (Figure S1C). We imaged ~4 million particles using Titan Krios electron microscopes equipped with a K3 direct electron detector. The 2D classes each exhibited two distinct densities, one attributable to well-ordered corePIC and the other attributable to a disordered TFIIH (Figure S1D). In some classes, DNA was readily identifiable on TFIIH. 1.8 million particles were selected through 2D class averaging and subjected to *ab initio* calculation of an initial map (Figure S2A). To sort out variability in positions of TFIIH and DNA, the ~1.8 million particles were subjected to iterative global 3D classifications, which revealed three forms of PICs (hereinafter PIC1, PIC2, PIC3) and one form of the ITC, accounting for 454K, 117K, 70K,

and 322K particles, respectively. In each form of PIC and ITC, TFIIH and DNA were poorly ordered compared to pol II. To reveal TFIIH and DNA in more detail (Figure S2A and S3A), three masks were created, the first containing pol II, the second containing TFIIH, and the third containing DNA-TFIIA-TBP-TFIIE (Tfa1 and Tfa2 WH domains)-TFIIF (Tfg2 WH domains)-TFIIB (cyclin domains). Three segments were subtracted from images with respective masks, subjected to local 3D classifications and refinement, and then composited back to the entire complex (Figure S2A and S3A), resulting in four maps at resolutions of 3.0-4.6 Å (PIC1), 4.0-7.3 Å (PIC2), 4.1-11.8 Å (PIC3), and 3.1-9.9 Å (ITC), respectively.

The three forms of the PIC, all in a closed state, differed from each other in locations and conformations of TFIIH and the path of DNA (Figures 1B-D and S5). PIC1, the most populated class of the PIC, was a good match to previous structures of yeast PIC (PDB ID 5fmf and 6gym) (Dienemann et al., 2019; Murakami et al., 2015b). In PIC1, TFIIH was resolved at near atomic resolution, allowing us to define two different DNA-binding modes for putative DNA translocation, as described in detail below. PIC2 and PIC3 were refined to 4.0-7.3 Å (4.0 Å for pol II, 7.3 Å for TFIIH) and 4.1-11.8 Å (4.1 Å for pol II, 11.8 Å for TFIIH), respectively, and were distinct from PIC1 in the position and conformation of TFIIH and DNA. The ITC revealed an open promoter DNA and a short DNA-RNA hybrid in the pol II active center at 3.1-9.9 Å resolution (3.1 Å for pol II, 9.9 Å for TFIIH) (Figure 3).

Two DNA-binding modes of TFIIH in PIC1

Focused 3D classification of TFIIH in PIC1 revealed two forms of TFIIH at 4.6 Å and 7.6 Å resolution (orange vs steel blue in Figures 1E-F, and S4). One form had a good match to previous structures of the pre-translocation state of TFIIH in the PIC (Dienemann et al., 2019; Murakami et al., 2015b; Schilbach et al., 2017) (orange, upper panel of Figure 1E), while the other form revealed a ~60° rotation of the domain that consists of Tfb5 and the C-terminal region of Tfb2, accompanied by a rotation of the C-terminal ATPase domain of Ssl2 (Ssl2C) relative to the rest of TFIIH (steel blue, lower panel of Figure 1E). In the former (orange in Figures 1E-F), a ~13-bp segment of DNA double helix is bent, deep within the DNA-binding groove between the two ATPase domains, in close contact with the five DNA binding motifs (Ic, IVa, IV, V, Vb, as previously defined (Fairman-Williams et al., 2010)) (referred to as the “strong-binding” state). In the strong-binding state, the DNA preferentially associated with Ssl2C rather than the N-terminal ATPase domain of Ssl2 (Ssl2N) (Movie S1). By contrast, in the latter (steel blue in Figures 1E-F), the DNA is relatively straight, in primary contact with the DNA binding motif Ic of Ssl2N (referred to as the “weak-binding” state). The motifs IV, V, and Vb of Ssl2C are detached from the DNA upon the rotation of Ssl2C along with Tfb5-Tfb2C (Figure 1F), enabling a slight rotation of the DNA along its axis compared to the strong-binding state (Figure S4). This suggests that the weak-binding state may represent the post-translocation state, although nucleotides were not directly resolved. Cycling between alternating preferential DNA binding by Ssl2C and Ssl2N in the strong- and weak-binding states may drive DNA translocation, where Tfb5 serves to play a key role in regulating Ssl2’s catalytic activity, as previously suggested (Coin et al., 2006; Kappenberger et al., 2020; Ranish et al., 2004).

A distinct form of the PIC shows dissociation of Tfb3 from the Rpb4/7 stalk and is implicated in altered PIC properties

PIC2 and PIC3 differ from PIC1 in locations and conformations of TFIIH and DNA path, as readily apparent in initial rounds of 3D classification (Figure S2A), and there are several notable differences between three forms of the PIC (Figure 2). First, PIC2 and PIC3 differ from PIC1 by ~20 Å and ~30 Å shifts in the location of TFIIH (Figure 2A), and by repositioning of Ssl2 on DNA by one turn of dsDNA (~10 bp), accompanied by greater degrees of DNA distortion ~20-30 bp downstream of the TATA box (Figures 2B-C and S5). Second, PIC2 and PIC3 revealed TFIIH in the weak-binding state, while PIC1 primarily revealed the strong-binding state. Third, in PIC2 and PIC3, the RING finger domain of Tfb3 (one of three TFIIK subunits) are dissociated from Rpb4/7, resulting in a shift in their positions by ~8 Å and ~20 Å relative to that in PIC1, such that TFIIH, as well as TFIIE, less closely contacts pol II (lower panels of Figures 1B-D). Irrespective of these significant differences between three forms, promoter DNA is nevertheless associated only with GTFs and not with pol II in all forms.

In PIC2 and PIC3, accounting for ~18% and 10% of PICs, the RING finger domain of Tfb3 was found to be dissociated from Rpb4/7 (the pol II “stalk”, Figures 1C-D). This dissociation may relate to a TFIIK-independent subpathway for initiation in the absence of promoter scanning, which results in utilization of upstream TSSs (~31bp downstream of the TATA box, a location characteristic of transcription initiation in metazoans rather than *S. cerevisiae*)(Murakami et al., 2015a)(see lane 1 in Figure 2D, indicated by red arrow). To investigate the requirement for Tfb3 in promoting initiation at downstream TSSs and therefore putative function in promoter scanning, we asked if Tfb3 alone could substitute for TFIIK in in vitro run-off transcription assays. We found that the N-terminus of Tfb3 (Tfb3ΔC), comprising the RING finger domain (residues 1-70) and the ARCH (Rad3) anchor domain (residues 71-148), but deleted the C-terminal region (residues 149-321) including the binding site for the cyclin kinase (van Eeuwen et al., 2021a), supported essentially WT activity at downstream TSSs using an *SNR20* promoter fragment (*SNR20* 91W) as initiation template in vitro (lanes 7-10 in Figure 2D). We then asked if amino acids supporting Tfb3/Rpb7 interaction were required for the activity of Tfb3ΔC. We substituted Tfb3 R64/K65, positively charged amino acids at the interface of the Tfb3 RING domain and Rpb7, for alanines and observed that downstream initiation on this same template was defective (lanes 12-15 in Figure 2D). Similarly, with a promoter variant *SNR20* 31D whose TSS was brought from the wild-type location 91 bp downstream of the TATA box to a location 31 bp downstream (Murakami et al., 2015a), Tfb3ΔC supported transcription both from this TSS (+1, blue) as well as further downstream TSSs (~+30 to ~+50) (lane 10 in Figure 2E). In contrast, the Tfb3ΔC R64A/K65A mutant exclusively initiated transcription from the upstream TSS (lane 15 in Figure 2E). These results suggest that Tfb3 requires specific contacts with Rpb7 for initiation at downstream TSSs, putatively through support of scanning to those positions and that the dissociation of the RING finger domain from Rpb4/7 shifts transcription initiation to sites upstream, as previously observed by the removal of TFIIK.

We determined if the effect of *tfb3* mutations on TSS utilization might be revealed *in vivo* (Figure 2F). One of our labs has previously developed plate phenotypes that are perfectly predictive of in vivo initiation defects for general transcription factors. In this system, we have found that mutants that shift TSSs to upstream positions confirm sensitivity to the drug mycophenolic acid due to inability to use a downstream TSS at the *IMD2* gene (Kaplan et al., 2012; Malik et al., 2017; Zhao et al., 2021). Furthermore, inability to induce *IMD2* can also be detected through use of a sensitive *IMD2Δ::HIS3* transcriptional reporter. We observed sensitivity to mycophenolic acid (MPA) as well as inability to induce *IMD2Δ::HIS3* for *tfb3* alleles, predictive of upstream shifts in TSSs at the *IMD2* promoter. We found that *tfb3* alleles R64D and R64A were much more sensitive to MPA than alleles K65A and K65D. Consistent with this, in PIC1, Arg64 of Tfb3 forms a salt bridge with Asp166 of Rpb7, while Lys65 of Tfb3 may contribute to stability of the RING finger domain itself. We note that we have previously identified an *rpb7* D166G allele as sensitive to MPA and an upstream TSS shifter *in vivo* (Braberg et al., 2013). These results support Tfb3-Rpb7 interactions as required for normal TSS utilization in vivo in *Saccharomyces cerevisiae*.

Bona fide ITC structure

The structures described above indicate potential conformational changes within the PIC occurring during or potentially subsequent to TFIIH-mediated translocation, some of which likely are consequential for initiation during promoter scanning. We now turn to description of a bona fide ITC formed by initiation de novo. Consistent with our biochemical analysis (Fujiwara et al., 2019)(see also Figures S1A-B), we observe a structure that has synthesized ~26 nt of RNA yet retains GTFs. The locations of retained GTFs in the ITC largely correspond to those in PIC3, except for differences in orientations of TFIIH and TFIIE (Figure 3A). In a similar manner to PIC3, the Tfb3 RING finger domain is dissociated from Rpb7 in the ITC (not visualized in the map), resulting in ~30 Å shifts in the location of TFIIH relative to PIC1 (Figure 3B). We could not identify ITC classes in which the Tfb3 RING finger domain is associated with Rpb4/7. The ITC reveals association of factors with upstream promoter DNA spanning ~38 bp, while retaining TFIIH at the downstream end. The ~38 bp upstream segment (positions −116 to −79) extending from about 16 bp upstream of the bend at the TATA box (−100 to −93) to about 22 bp of DNA downstream (from −116 to −79) is suspended above the pol II cleft and positioned by the general transcription factors (TFIIA, -IIB, -IIE, -IIF, TBP) as in the PIC (Figure 3C). Given that RNA synthesis to ~+20–+26 has occurred, observed retention of TFIIH downstream DNA binding has occurred either through dissociation and reassociation or through TFIIH translocation to the downstream position.

The DNA downstream of the 38-bp segment was entirely missing due to flexibility, except densities attributable to ~10-nt template strand in the pol II active center (Figures 3D and S6F) and ~9-bp segment of DNA double helix along the DNA binding groove between two ATPase domains of Ssl2 (Figure S3C) . The downstream end of the 38-bp segment (−79) corresponds to the upstream edge of the bubble (Pal et al., 2005; Plaschka et al., 2016), to which the hairpin of the Tfa1 WH domain (TFIIE) projected (Figure 3C), as previously observed in a pre-melted open complex (Plaschka et al., 2016). Given that transcripts were exclusively initiated from downstream TSSs rather than upstream TSSs in the presence of

3’-O-methyl GTP (Figures 1A and S1A), the single stranded bubble that was initially filled in the active center of pol II, as in the pre-melted open complex, must be displaced by downstream single stranded DNA entering the pol II cleft, through sliding and formation of a larger bubble, while maintaining the upstream edge of the bubble at −79. It may be noted that the path of the DNA at the downstream end of the ITC differs from that in previous pre-melted open complexes (He et al., 2016; Schilbach et al., 2017)(Figure S6A-C): a disordered density was observed extending from the bridge helix to ~10 bp of DNA downstream in the pol II cleft, and the DNA further downstream entering the pol II cleft was not observed. The DNA double helix might be partially displaced from the pol II downstream channel due to dissociation of the Tfb3 RING finger domain from Rpb7 and the ~30 Å shift in the location of TFIIH (Figure S6F). Considering that the Tfb3-Rpb7 interaction is required for downstream TSS utilization (Figure 2), this structural change might have occurred after the initiation of transcription (see DISCUSSION).

The density in the active center of the ITC was not sufficient to assign the nucleotide sequence (i.e., the translocation register of the nascent RNA), but was a good match to the 5-bp DNA-RNA hybrid previously observed by X-ray crystallographic studies (Cheung and Cramer, 2011; Liu et al., 2011): an elongated density corresponding to ~5 ribonucleotides of RNA at positions i–1 to i–5 of RNA relative to the nucleotide addition site (i+1), and ~10 nucleotides of the template strand at positions i+1 to i−9 were observable (Figure 3D). We note that the 5-bp DNA-RNA hybrid observed in the ITC likely represented ensemble of transcripts initiated from different TSSs, given TSSs are equally distributed at +1 to +7 on this promoter template (Figure 1A). The N-terminal region (residues 22-61) of TFIIB, except the B-finger (or B-reader), was discerned in the RNA exit tunnel as in previous X-ray crystallography (Bushnell et al., 2004; Kostrewa et al., 2009; Liu et al., 2010; Sainsbury et al., 2013). This suggests that pol II was subjected to extensive backtracking after transcribing ~26 nt RNA and was arrested in registers compatible with TFIIB (Figures S6E-F). Consistent with this, transcripts of the ITC stalled at +27 were rapidly degraded to ~12 nt or shorter upon addition of TFIIS (Fujiwara et al., 2019). Also, there was a density attributable to the backtracked RNA (at positions i+3 to i+5) in the position compatible with the domain III of TFIIS (Kettenberger et al., 2004) in the pol II funnel. This TFIIS-compatible path of the backtracked RNA observed in this study is roughly consistent with that in previous study by X-ray crystallography (Cheung and Cramer, 2011) (purple density in Figures 3D and S6D). The extensive pol II backtracking may relate to the largely unwound open promoter in the ITC as pol II obviates the need for unwinding DNA for backtracking. Apparently such pol II backtracking was not counteracted by the TFIIH downstream DNA binding, and potentially permitted through dissociation and reassociation of TFIIH.

The structure of ECs colliding head-to-end (EC+EC).

In contrast to the G-less 26 complex, which our studies here reveal is an ITC that retains GTFs, our previous biochemical studies demonstrated that assemblages initiating on our G-less 49 template actually contained two pol II complexes, a pol II at +49 that escaped the promoter, and another pol II that initiated transcription by re-utilizing the promoter to generate the ~25 nt RNAs (thus referred to as a re-initiation complex) (Fujiwara et al.,

2019). Notably, we observed that a significant population of ITCs from the second round of transcription completed promoter escape as defined by the loss of GTFs in contrast to the G-less 26 complex (see Figure 4A). Also, ~25 nt transcripts from the second round of transcription were stably retained in the G-less 49 complex, whereas transcripts of similar lengths of the G-less 26 complex were prone to dissociate by extensive backtracking of pol II upon removal of ATP during gradient sedimentation, thus requiring inclusion of 4'-thio UTP to limit backtracking for our structural study (Fujiwara et al., 2019). To elucidate the underlying mechanism of this difference, the G-less 49 complex was isolated and subjected to cryo-EM analysis in a similar manner to the G-less 26 complex. A transcription reaction with the G-less 49 template was initiated by adding NTPs (ATP, CTP, and UTP) with chain-terminating 3'-O-methyl GTP. Following gradient sedimentation, two major peaks of the re-initiation complex were revealed (fractions 17-18 and 22-24 in Figure 4A). Two initial rounds of 2D classification of the slower sedimenting fractions (fractions 17-18) yielded a set of well-ordered homogeneous classes of two colliding pol II complexes (referred to as EC+EC) (Figures 4C and 4E). After interactive 3D classifications to remove some residual PICs and single ECs (119K and 886K particles, respectively), both of which were likely derived from EC+PIC assemblages (Figure 4A), the structure of EC+EC, accounting for 180K particles, was refined to 3.5 Å resolution (Figures 5 and S7). By contrast, the faster sedimenting fractions (fractions 22-24) yielded similar classes of two colliding pol II molecules, one of which was associated with a set of GTFs (referred to as EC+ITC) (Figures 4D and 4F), consistent with protein analysis by SDS-PAGE (Figure 4B). However, the considerable variability in the distance between two pol II molecules prevented refinement of EC+ITC past about 15-Å resolution.

In the structure of EC+EC, two colliding ECs span over ~74 bp of DNA (from positions −9 to +65 relative to TSS) (Figure 5). There was a well-ordered density corresponding to TFIIF only on the trailing EC, but not the leading EC (Figure 5A). A previous crystallographic model of an EC complex with a 9-bp DNA-RNA hybrid (PDB ID: 5C4J) (Barnes et al., 2015) was fitted into two corresponding densities with some deviations in the non-template strand of the transcription bubble. A structure of pol II (PDB ID: 5U5Q) was fitted also without any deviations except a ~10°-rotation of Rpb4/7 subunits of the leading EC, that enabled a direct contact with TFIIF of the trailing EC (Figures 5B-C).

In the DNA-RNA hybrid of the ECs, purines and pyrimidines were distinguishable based on differences in size, allowing us to assign EC registers (Figures 6A-D, and S9A-B). The leading EC is backtracked one bp from the G-stop +49 and stalled at +48 (for the location of the nucleotide addition site of pol II, see Figure 6G), while the trailing EC is stalled at +13. There is a weak density attributable to phosphodiester of putative 3'-O-methyl-GMP at +49. The locations of the transcription bubbles determined by EM are roughly consistent with our previous KMnO4 assays and TFIIS cleavage assays (Fujiwara et al., 2019). These data suggest that the trailing pol II that had reached ~+25 to transcribe a ~25-nt RNA (based on RNA gels (Figure 4A),) was subjected to extensive (~12bp) backtracking, and arrested at +13. For the trailing EC to have synthesized ~25 nt, two ECs would require substantial structural changes in the protein component or/and the DNA component to avoid a steric clash when both would be in register without backtracking. Consistent with the trailing pol II backtracking, there was density attributable to backtracked RNA in the pore

and the funnel, but not in the leading EC (Figures 6C-E). Two-body refinement revealed a ~6°-rotational motion relative to each other, with a 35-bp spacing between two nucleotide addition sites (Movie S2).

In the observed state with the trailing EC backtracked, specific protein-protein interactions were established at the interface between the two ECs (Figure 5B). There are two major points of contact: the first point of contact involves two loops (residues 148-168 and residues 185-197) protruding from Rpb1 clamp of the trailing EC, and a loop of Rpb2 protrusion (residues 97-113) and the Rpb12 zinc ribbon (residues 35-50) of the leading EC. The second point of contact involves the tip of the dimerization domain of Tfg1, the largest subunit of TFIIF, of the trailing EC, and the tip of Rpb7 of the leading EC. Notably, previous exonuclease footprinting of ECs colliding head-to-end in the absence of TFIIF exhibited greater variability in the distance between two ECs upon head-to-end collision as well as much more extensive backtracking of the trailing EC (~50 bp backtracking upon encountering a leading EC)(Saeki and Svejstrup, 2009). Our structural and biochemical data suggest that some degree of specificity of the EC+EC is conferred by TFIIF.

The template DNA is overall Z-shaped with two kinks at the two active centers of pol II (Figures 6A-B). The 24-bp DNA (from +14 to +37) bridging between two active centers was clearly discerned and modeled with a straight B-form DNA (Figure 6B). The density of the DNA-RNA hybrid in each active center was traceable (Figures 6B-D). In the leading EC, 16 ribonucleotides of the 49-nt transcript were visualized: 9 ribonucleotides form a hybrid with the template DNA (positions from +40 to +48), while a stretch of adjacent seven ribonucleotides extended into the RNA exit tunnel (Figures 6A-B and 6G). In the trailing pol II, 15 ribonucleotides of the ~25-nt transcript were discernible (Figures 6C and 6G): 9 ribonucleotides formed a hybrid with the template DNA (positions from +5 to +13), and an adjacent five ribonucleotides of the backtracked RNA lay in the pol II pore and funnel (Figure 6E); two ribonucleotides of the backtracked RNA at positions i+2 and i+3 were in the position in the pore as previously observed by X-ray crystallography (Cheung and Cramer, 2011; Wang et al., 2009), while three ribonucleotides at positions i+4, i+5, and i+6 lie on a positively charged patch composed of Lys619 and Lys620 of Rpb1 in the funnel in a position not observed in previous structures, suggesting flexibility for backtracked RNA within the pol II funnel except the i+2 and i+3 positions (Figure 6F). This backtracked RNA is incompatible with TFIIS (Figure S9C) and must be displaced from this site for TFIIS-induced transcription resumption from the backtracked state (Cheung and Cramer, 2011). These observations firmly establish the arrested state of a collided head-to-end dimeric EC.

Promoter escape of an ITC is accelerated by a transcribing pol II at promoter proximal regions

The trailing EC stalled at +13 contrasts to the G-less 26 ITC (Figure 3) that failed to escape promoter in a single round of transcription (Figure 3), in that the latter has retained GTFs while the former has not, though both are proposed to have transcribed similar distances. This observation suggests that a promoter proximal pol II might provoke promoter escape as defined by loss of GTFs of the follow-on ITC, allowing its conversion to an EC. Support

for this possibility came from cryo-EM analysis of the EC+ITC, an intermediate on the pathway to promoter escape of the trailing ITC (Figures 4D and 4F). All 2D class averages showed a large (~500 kDa) density attributable to TFIIH in the space between the two polymerases (indicated by orange arrow heads in Figure 4D). The assignment of TFIIH was further validated by a comparison with a 2D projection from a 3D model of EC+coreITC (ITC lacking TFIIH) (Figure S10B). Of the eight class averages we obtained, the top four populated classes maintain a similar spacing between the EC and the ITC as in the EC+EC, presumably through the direct protein-protein interactions described above (upper row in Figure 4D, accounting for 5170 particles). In these class averages, the DNA double helix is accommodated in the pol II downstream cleft of the trailing ITC, while TFIIH is dissociated from the DNA and displaced from the position observed in the G-less 26 ITC (schematically illustrated in Figure 4F, see also Figure S10A). This conformational change of the ITC, as an irreversible critical transition from initiation to elongation, was evidently facilitated by the presence of the leading EC. In the other classes, the EC and the ITC are apparently separated from each other, suggesting that the trailing ITC that had failed to escape promoter on its first attempt was subjected to extensive backtracking (lower row in Figure 4D, accounting for 4537 particles), requiring TFIIS for transcription resumption from the backtracked state. Altogether, we propose that ITCs that reach a paused pol II ahead of them have GTF release facilitated through direct interactions between complexes.

Discussion

Structural and mechanistic studies of transcription initiation involving TFIIH have been hampered by poor efficiency of initiation reaction in vitro (commonly ~0.01-0.1 transcripts per PIC). Previous structural models of transition from initiation to elongation were derived from complexes with artificially open templates, and not obtained by the catalytic activity of TFIIH. Thus how TFIIH directs promoter melting, TSS scanning, and promoter escape (Dvir et al., 1997; Fishburn et al., 2016; Luse, 2013; Qiu et al., 2020; Spangler et al., 2001; Zhao et al., 2021) remains to be resolved. To dispel this long-standing mystery of the transcription initiation process, we have developed a highly efficient in vitro reconstitution from the yeast at quality and quantity amenable to structure determination (Fujiwara and Murakami, 2019) using the strong promoter *SNR20* (Fazal et al., 2015; Kuehner and Brow, 2006; Murakami et al., 2015a). By imaging reaction intermediates in this highly efficient in vitro transcription system, we have generated a multi-step description of pol II transcription from initiation by the 33-subunit PIC to formation of the elongation complex. Combined with our previous biochemical analysis, our structural data have revealed a branched pathway for promoter escape when multiple pol II molecules load on a gene in rapid succession. Promoter escape, viewed in the past as no more than dissociation of pol II from promoter, now appears mechanistically varied in our in vitro system, with potential regulatory consequences.

Three distinct forms of the PIC were identified in this study: relative to PIC1 in a form similar to previous structures (Dienemann et al., 2019; Murakami et al., 2015b), PIC2 and PIC3 exhibited ~20 Å and ~30 Å shifts in the location of TFIIH, and repositioning of Ssl2 on DNA by one turn of dsDNA, along with greater degrees of DNA distortion. In PIC1, the location of TFIIH is constrained by the contact with Rpb4/7, whereas, in PIC2/PIC3, upon the dissociation of the RING finger domain of Tfb3 from Rpb4/7, the positional

constraint of TFIIH is relieved, and TFIIH is anchored to the core PIC presumably by protein-protein contacts between the C-terminal flexible region of Tfa1 (TFIIE) and BSD domains of Tfb1 (TFIIH), supported by many cross-links formed between them (Murakami et al., 2013b; Robinson et al., 2016; Schilbach et al., 2017). Distinct forms of the PIC relate to the differing properties in TSS utilization; we have demonstrated that a mutation R64A/K65A of Tfb3 at the interface with Rpb4/7 exclusively initiated transcription from upstream TSSs in vitro, locations characteristic of transcription initiation in metazoans rather than *S. cerevisiae,* as previously observed by the removal of TFIIK(Murakami et al., 2015a). This explains why, in our transcription reactions in the presence of TFIIK, we generally observe transcripts initiated from downstream TSSs canonical in the yeast system but also to a lesser extent from upstream TSSs, where most other eukaryotes initiate (see lane 5 in Figure 2D). It is likely that additional factors redundantly enforce downstream TSS usage in yeast *in vivo*. Upon the dissociation of the RING finger domain from pol II and the resulting repositioning of TFIIH, the DNA entering the pol II cleft may disengage from the downstream pol II channel (Figures S6E-F) and its unwinding by the TFIIH translocase may not be coupled to the delivery of a TSS (especially downstream TSS) to the active center of pol II. Notably, many residues of the Tfb3, other than R64 and K65, at the interface with Rpb4/7 are not conserved between yeast and humans (Gervais et al., 2001), which may contribute to differences in TSS utilization between yeast and metazoans. It has also been proposed that fundamental characteristics of TFIIH processivity are responsible for differences in initiation mechanisms (Tomko et al., 2021).

In a similar manner to PIC2/PIC3, the structure of the ITC also revealed the dissociation of the Tfb3 RING finger domain from Rpb4/7 and the repositioning of TFIIH. Moreover, the repositioning of TFIIH likely induced partial DNA disengagement from the pol II channel, in contrast to previous structures of human ITCs, in which DNA was accommodated in the pol II channel and with DNA further downstream simultaneously bound by TFIIH (He et al., 2016)(Figures S6A-C). DNA disengagement in our ITC might occur after the synthesis of a relatively long transcript of ~26 nt RNA or subsequent pol II backtracking. Notably a recent study of mammalian PICs demonstrated that the dissociation of the RING finger domain from Rpb4/7 is accompanied by the conversion from closed to open complexes (Aibara et al., 2021), pointing to the dissociation of the RING finger domain from Rpb4/7 as a conserved mechanism, with the timing of the dissociation being different among species.

Previous real-time observations of single PICs demonstrated that yeast TFIIH translocase reels dozens of base pairs of downstream DNA irrespective of the presence or absence of TFIIK (Fazal et al., 2015). These results indicated that TFIIH must be sufficiently stable in the ITC to promote DNA scrunching in the absence of Tfb3-PIC contacts with either Rpb7 or TFIIE. However, this DNA translocation no longer results in PIC scanning to downstream TSSs. Biochemical analysis of the long-persisting ITC indicated that promoter escape is not completed by the steric clash of growing RNA with TFIIB and additional distinct mechanisms are required (Fujiwara et al., 2019; Luse, 2019). We have shown that addition of the capping enzyme and Spt4/5 increases the frequency of promoter escape of about ~30% of ITCs, but not all ITCs, *in vitro*. (Fujiwara et al., 2019). Unexpectedly, we find promoter escape is greatly accelerated in ITCs encountering another pol II at promoter proximal regions, where TFIIH is dissociated from DNA due to steric occlusion by two

successive transcribing polymerases (Figure 7). For highly transcribed genes that initiate frequently, e.g., *HSP70* promoter at 0.25/sec (O'Brien and Lis, 1991), polymerases would be located every ~120 bp assuming a constant elongation rate, *e.g.,* ~30 bp/sec *in vivo* (Larson et al., 2011; Lenstra et al., 2016; Mason and Struhl, 2005). For genes that initiate less frequently, abundant evidence suggests that transcription occurs in discontinuous bursts (or pulses) in which multiple polymerases within each pulse may run adjacent to one another (Bahar Halpern et al., 2015; Chubb et al., 2006; Donovan et al., 2019; Lenstra et al., 2015; Zenklusen et al., 2008). Under these conditions, it makes sense that a fraction of productive yeast PICs may encounter a pol II from a preceding round of transcription *in vivo*.

Lastly, it is noteworthy that the formation of EC+EC is sterically compatible with other transcription factors that function in early transcription. For example, pol II-Mediator complex (Robinson et al., 2016; Schilbach et al., 2017; Tsai et al., 2017) is compatible with EC+EC (Figure S11A). Additionally, the structure of pol II (EC)-DSIF-NELF complex (Vos et al., 2018) in a canonical form of promoter-proximal paused pol II in mammalian systems, is compatible with EC+EC except a partial steric clash with DSIF (Figure S11B). It is likely that promoter escape may be synergistically regulated by distinct mechanisms involving contributions of additional factors. We now have established the capability of visualizing bona fide intermediates on the pathway to elongation and determining the basis for differences between them.

Limitations of the study

Although particle numbers used for each reconstruction were relatively small, our gradient sedimentation followed by RNA and protein analyses strongly support that our structures are representative of true reaction intermediates on the main reaction pathway. However, our study does not establish that our structures are only representative of true reaction intermediates. For example, as pol II in the ITC was subjected to extensive backtracking when stalled either by a G-stop or by a downstream EC at promoter proximal regions, ITCs identified in this study represent a backtracked state and may differ from those during the nucleotide addition steps. In addition to such structural bias introduced by isolation of complexes, PICs and the ITC were considerably variable in structure. During extensive computational classifications, some other structural states may have escaped detection. Thus mere existence of these structures does not establish that they are only major on-pathway structural states. Additionally, our reaction intermediates in this study have been obtained in the absence of elongation factors, such as the capping enzyme and Spt4/5, which can significantly affect the reaction pathway (Fujiwara et al., 2019). Finally it may be noted that our structures of PICs and ITCs had some preferred orientations and resulting resolution anisotropy, likely due to the use of promoter DNA fragments of over 200 bp including the core promoter and downstream TSSs. Such issues may potentially affect map interpretation and model building, and can be improved in the future.

STAR METHODS

RESOURCE AVAILABILITY

Lead Contact—Further information and requests for reagents should be directed to and will be fulfilled by the Lead Contact, Kenji Murakami (kenjim@pennmedicine.upenn.edu).

Materials Availability—Plasmids and strains generated in this study are available upon request from Lead Contact with a completed Materials Transfer Agreement.

Data and Code Availability

- The Cryo-EM maps have been deposited at the Electron Microscope Data Base (EMDB) under accession codes EMD-23904 (PIC1), EMD-23905 (PIC2), EMD-23906 (PIC3), EMD-23907 (TFIIH weak binding state), EMD-23908 (ITC), and EMD-23789 (EC+EC), and are publicly available as of the data of publication. And the atomic coordinates have been deposited at the Protein Data Bank (PDB) under accession codes 7ML0 (PIC1), 7ML1 (PIC2), 7ML2 (PIC3), 7ML3 (TFIIH weak binding state), 7ML4 (ITC), and 7MEI (EC+EC), and are publicly available as of the data of publication. Raw cryo-EM images are deposited at EMPIAR under accession code EMPIAR-10865 and are publicly available as of the data of publication. These accession numbers are also listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Microbe strains—*E.coli* Rosetta 2 (DE3) strains were used to express TFIIA, TFIIB, TBP, and Sub1. *E.coli* BL21 (DE3) strains were used to express Tfb3ΔC, the Tfb3 mutant R64A/K65A, and TFIIA. *E.coli* cells were grown in lysogeny broth (LB) medium or terrific broth (TB) medium supplemented with antibiotics at 37°C for pre-culture and at 18°C or 16°C for protein expression as described in the method details below. *S. cerevisiae* CB010 were used as parent strains to construct the specific mutants to express TFIIH, TFIIE, TFIIK, and pol II. *S. cerevisiae* cells were grown in yeast extract (1.5% w/v), peptone (2% w/v), dextrose (3% w/v) medium supplemented with adenine (0.016% w/v) at 30°C. All information of *S. cerevisiae* stains used for transcription-related growth phenotypes of tfb3 alleles is listed in the key resources table and the *S. cerevisiae* cells were grown under conditions described in the method details below.

METHOD DETAILS

Protein purification—Recombinant TFIIA, TFIIB, TBP, and Sub1 were overexpressed and purified from *E.coli*. TFIIE, TFIIF, TFIIH, and pol II were isolated from yeast as previously described (Fujiwara and Murakami, 2019).

TFIIB was expressed as previously published (Bratkowski et al., 2018). TFIIB fused with a SUMO-tag was expressed in *E.coli* Rosetta 2(DE3) in lysogeny broth (LB) medium supplemented with 50 μg/ml kanamycin. The lysate in buffer containing 50 mM Tris (pH 7.5), 10 μM $Zn(OAc)_2$, 500 mM NaCl was subjected to HisTrap™ HP (Cytiva), followed by cleavage using Ulp1 protease. The protein mixture was loaded onto HisTrap™ HP (Cytiva) to remove the SUMO-tag. Finally, the flowthrough containing TFIIB was loaded onto HiTrap Heparin (Cytiva) to run gradient elution from 300 mM NaCl to 2 M NaCl in buffer containing 50 mM Tris (pH 7.5), 10 μM $Zn(OAc)_2$.

Toa1 and Toa2 of TFIIA were expressed separately in the insoluble fraction in *E.coli* Rosetta 2(DE3) cells in lysogeny broth (LB) medium supplemented with 50 μg/ml kanamycin as previously published (Murakami et al., 2012). TFIIA was reconstituted with refolded Toa1 and Toa2 in buffer containing 20 mM Tris (pH 7.5), 100 mM NaCl, 10% glycerol, and 3 mM DTT, and further purified using a Superdex G200 column (Cytiva). Alternatively, TFIIA was expressed in the soluble fraction as previously described (Adachi et al., 2017). Briefly, TFIIA was expressed in *E.coli* BL21(DE3) cells for 20 hours at 16°C in terrific broth (TB) medium. Cells were lysed in buffer containing 30 mM Tris (pH 7.5), 500 mM NaCl, 10% glycerol, and 10 mM 2-mercaptoethanol, and protease inhibitors (0.17 mg/mL phenylmethylsulfonyl fluoride, 0.284 μg/mL leupeptine, 1.37 μg/mL pepstatin A, 0.33 mg/mL benzamidine) and the debris was removed by centrifugation. The supernant was loaded onto a Ni column and TFIIA was eluted using gradient from buffer A (30 mM Tris (pH 7.5), 150 mM NaCl, 10% glycerol, and 5 mM 2-mercaptoethanol, and 50 mM imidazole) to buffer B (buffer A + 500 mM imizazole). Fractions containing TFIIA were pooled, supplemented with 2 mM DTT, 3C precession, and dialyzed in buffer containing 30 mM Tris (pH 7.5), 150 mM NaCl, 10% glycerol, and 2 mM DTT. After the linker between Toa1 and Toa2 was cleaved by 3C precession, TFIIA was further purified using Superdex 200 (Cytiva). Stable TFIIA was then loaded onto a UNO Q column, and eluted with gradient from buffer A (20 mM HEPES (pH 7.6), 10% glycerol, 3 mM DTT, 150 mM potassium acetate) to buffer B (buffer A + 1M potassium acetate).

TBP was expressed in *E.coli* Rosetta 2(DE3) in TB medium supplemented with 50 μg/ml kanamycin as previously published (Murakami et al., 2012). The lysate in buffer containing 20 mM Tris (pH 7.5), 100 mM NaCl, and 3 mM DTT was subjected to HiTrap Heparin HP (Cytiva) followed by gradient elution from 100 mM NaCl to 1000 mM NaCl in buffer containing 20 mM Tris (pH 7.5), and 3 mM DTT. After dialysis against buffer containing 20 mM Tris (pH 7.5), 100 mM NaCl, and 3 mM DTT, the protein mixtures were loaded onto HiTrap SP (Cytiva) to run gradient elution from 100 mM NaCl to 1000 mM NaCl in buffer containing 20 mM Tris (pH 7.5), and 3 mM DTT.

Sub1 fused with a His6-tag was expressed in *E.coli* Rosetta 2(DE3) in TB medium supplemented with 100 μg/ml ampicillin as previously published (Fazal et al., 2015). The lysate in buffer containing 30 mM Tris (pH 7.5), 500 mM NaCl, 0.1% Triton X-100, 50 mM imidazole, 1 mM DTT, and protease inhibitors was subjected to HisTrap HP column (Cytiva) followed by gradient elution from 50 mM imidazole to 500 mM imidazole in buffer containing 30 mM Tris (pH 7.5), 200 mM NaCl, and 5% glycerol. The elution was further

purified using HiTrap Heparin (Cytiva) and HiTrap SP (Cytiva) in buffers containing 20 mM HEPES (pH 7.6), 5% glycerol, 4 mM DTT, and 150-1000 mM potassium acetate.

TFIIF was isolated from a yeast strain carrying a TAP-tag on *TFG2* as previously published (Murakami et al., 2012). Cells were resuspended in lysis buffer 200 (50 mM Tris (pH 7.5), 1 mM EDTA, 1 μM $Zn(OAc)_2$, 200 mM ammonium sulfate, 0.1% 3-(decyldimethyl-ammonio) propanesulfonate (Sigma), 2 mM DTT, and protease inhibitors) and lysed by bead beating. After addition of 0.25% PEI and 55% ammonium sulfate, lysed cells were stirred for 1 h and centrifuged for 90 min. The pellet was re-suspended with 1 L buffer 25 (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 μM Zinc Acetate, 25 mM ammonium sulfate, 0.05% 3-(decyldimethyl-ammonio) propanesulfonate (Sigma), 2 mM DTT, and protease inhibitors). After centrifugation for 1 h, the supernatant was loaded onto 50 mL IgG column, then washed with 1 L buffer (250) (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 μM Zinc Acetate, 0.05% 3-(decyldimethyl-ammonio) propanesulfonate (Sigma), 2 mM DTT, and protease inhibitors) with the mM concentration of ammonium sulfate in parentheses, 1 L buffer (500), 1 L buffer (25), and 1 L buffer (200). TFIIF bound to IgG column was incubated with 1 mg TEV for 15 h, eluted with 300 mL buffer 200, and then loaded onto HiTrap Heparin (Cytiva). TFIIF was eluted by salt gradient of concentration from 200 mM to 800 mM ammonium sulfate. Fractions containing three-subunit TFIIF were pooled, concentrated, and stored at −80°C.

Pol II was isolated from a yeast strain carrying a TAP-tag on *RPB3* was purified in the essentially same manner as that for TFIIF except that detergent 3-(decyldimethyl-ammonio) propanesulfonate (Sigma) was omitted in all buffers.

TFIIE was isolated from a yeast carrying a TAP-tag on *TFA2* as previously described (Fazal et al., 2015) with minor modifications. Cells were resuspended in buffer 250 (50 mM HEPES (pH 7.6), 1 mM EDTA, 5% glycerol, 250 mM ammonium sulfate, 5 mM 2-mercaptoethanol, and protease inhibitors) and lysed by bead beating. Following the addition of 0.1% PEI, lysed cells were stirred for 1 h and centrifuged. The cleared lysate was loaded onto 50 ml IgG column, then washed with 1 L buffer 250, and then 1 L buffer 400 (50 mM HEPES (pH 7.6), 1mM EDTA, 5% glycerol, 400 mM ammonium sulfate, 2 mM DTT, and protease inhibitors) plus 300 mM potassium acetate, followed by 1 L buffer 0 (50 mM HEPES (pH 7.6), 1 mM EDTA, 5% glycerol, 2 mM DTT, and protease inhibitors) plus 300 mM potassium acetate. TFIIE bound to IgG column was incubated with 1 mg TEV for 15 h, and was eluted with 300 mL buffer A (50 mM HEPES (pH 7.6), 150 mM potassium acetate, 5 mM DTT, 2 mM magnesium acetate, and 5% glycerol), followed by concentration, and then loaded onto HiTrap Heparin (Cytiva). TFIIE was eluted by salt gradient of concentration from 150 mM to 1 M potassium acetate. Fractions containing TFIIE were pooled and loaded onto Superdex G200 column (Cytiva) equilibrated with buffer G (20 mM HEPES (pH 7.6), 300 mM potassium acetate, 5% glycerol, 2 mM magnesium acetate, and 5 mM DTT). Fractions containing TFIIE were pooled and concentrated. Aliquoted TFIIE was flash frozen and stored at −80°C.

TFIIH was isolated from yeast *tfb6Δ* as previously described (Murakami et al., 2012) with minor modifications. In short, TFIIH carrying a TAP-tag on *TFB4* was grown in

100 L of YPAD medium to OD 10.0. Cells were lysate by bead beating in Buffer 400 (50 mM HEPES pH 7.6, 400 mM potassium acetate,1 mM EDTA, 5% glycerol, 5 mM 2-mercaptoethanol, and protease inhibitors). Following the addition of 100 mM ammonium sulfate and 0.2% PEI, lysed cells were stirred for 1 h and centrifuged. The cleared lysate was loaded onto 50 ml IgG column. The column was washed with 2 L Buffer 300 (50 mM HEPES pH 7.6, 300 mM potassium acetate, 1 mM EDTA, 5% glycerol, 2 mM DTT, and protease inhibitors), then resuspended in Buffer 300 and allowed to settle overnight. IgG column was washed with 1 L Buffer 300 and incubated with 1 mg TEV for 15 hours. After elution with 50 mL Buffer 300, sample was concentrated, and then loaded onto HiTrap Q (Cytiva). TFIIH was eluted by salt gradient of concentration from 300 mM to 1.2 M potassium acetate. Fractions containing 10-subunit and 7-subunit TFIIH were pooled separately and concentrated. Aliquoted TFIIH was flash frozen and stored at −80°C.

TFIIK was isolated from yeast carrying a TAP-tag on *TFB3* in the same manner as described above for TFIIH purification except that the IgG eluate was puried by a Superdex G200 column (Cytiva) instead of HiTrap Q (Cytiva).

Cryo-EM sample preparation with the G-less 26 template—The G-less 26 DNA fragment (−122/+97) (Fujiwara et al., 2019) was amplified by PCR and purified using a Superose 6 10/300 (Cytiva) in buffer 300 (20 mM Hepes (pH 7.6), 300 mM potassium acetate, 5 mM DTT, and 2 mM magnesium acetate). To assemble PICs on the G-less 26 DNA template, the following were mixed in 515 μL of buffer 300 containing additional 5% glycerol: 0.26 μM DNA template, 0.5 μM TFIIA, 0.7 μM TFIIB, 1.2 μM TBP, 0.6 μM TFIIE, 1.04 μM TFIIF, 0.44 μM TFIIH, 0.44 μM TFIIK, 1.04 μM pol II, and 0.6 μM Sub1. The mixture was then diluted by adding an equal volume of buffer 10 (20 mM Hepes (pH 7.6), 10 mM potassium acetate, 5 mM magnesium sulfate, 5 mM DTT) and incubated on ice for 24 hours. After pre-incubation for 20 min at 30°C, 3/4th of the PIC mixture received 2x NTP solution consisting of 1.6 mM ATP, 1.34 mM CTP, 2 mM 4'-thio UTP, 0.5 μM 3'-O-methyl GTP, 10 mM magnesium acetate, and 0.5 U/μL RNaseOUT in buffer 10, and 1/4th of the mixture received 2x NTP solution containing 44 nM [α-^{32}P] CTP (33 μCi). Transcription initiation was carried out for 20 min at 30°C and a total of 1.5 mL of the cold sample was immediately loaded onto three pre-cooled glycerol gradients (500 μL per gradient) prepared with buffer A (20 mM Hepes (pH7.6), 50 mM potassium acetate, 5 mM DTT, and 2 mM magnesium acetate, and 10% glycerol (v/v)) and buffer B (20 mM Hepes (pH7.6), 50 mM potassium acetate, 5 mM DTT, and 2 mM magnesium acetate, 0.125% glutaraldehyde, and 40% glycerol (v/v)). The 1/4th of the mixture (500 μL) that was incubated with hot NTP solution was loaded onto a glycerol gradient without glutaraldehyde. After centrifugation for 14 h at 30,000 rpm in a Beckman SW60 Ti rotor, the gradients were fractionated using a PGF Piston Gradient Fractionator (BioComp Instruments, Inc.) into ~100 μL per fraction and crosslinking reaction was quenched by addition of 50 mM glycine (pH 7.6). To perform RNA analysis of the fractions, 70 μL from non-crosslinked sample was incubated for 15 min at 42°C with 160 μL of stop buffer containing 390 mM sodium acetate (pH 5.5), 8 mM EDTA, 0.6% SDS, 0.06 mg/mL glycogen, 0.03 mg/mL proteinase K, and 0.03 mg/mL salmon sperm DNA, and subjected to

ethanol precipitation. RNA was then analyzed by urea denaturing gel. For protein analysis, 20 μL per fraction was analyzed by SDS-PAGE (Figure S1A).

Cryo-EM grid preparation and data collection of the G-less 26 sample—The appropriate fractions from crosslinked sample were pooled and concentrated ~8 fold with a 100k MWCO spin concentrator and dialyzed against buffer 50 (20 mM Hepes (pH 7.6), 50 mM potassium acetate, 5 mM DTT, and 4 mM magnesium acetate) for 45 min. For cryo-EM grid preparation, 2.7 μL of the G-less 26 sample were applied onto glow-discharged Quantifoil R0.6/1 200-mesh holey carbon grids (Electron Microscopy Sciences), blotted for ~1.7 second, and plunge frozen in liquid ethane with a Leica EM CPC manual plunger (Leica Microsystems). The grids were loaded onto a Titan Krios electron microscope operating at 300kV equipped with Gatan K3 Summit direct electron detector with Gatan quantum energy filter (slit width of 20 eV) at CryoEM core facility at University of Massachusetts. The data were collected automatically using SerialEM at a nominal magnification of 81,000x, with a defocus range of −0.5 μm to −2.5 μm, and with a 30 frame exposure taken over ~2.4 sec with a total electron dose of ~45 $e^{-}/Å^2$. A total of 15895 images was collected.

Cryo-EM data processing of the G-less 26 sample—All image processing of the G-less 26 sample was performed using RELION 3.0 and 3.1 (Zivanov et al., 2018). A total of 15,895 images was processed in two sets (8,000 and 7,895 images) in the same manner. The movie frames were aligned using RELION's own implementation with a binning factor of 2 and the CTF was determined using CTFFIND-4.1 (Mindell and Grigorieff, 2003). At this point, a total of 56 images were excluded for the further analysis. Initially, particles were picked automatically from 700 images and subjected to a few rounds of reference-free 2D classification. Some of the resulting 2D classes were low-pass filtered to 20Å and then used to pick particles from the two sets of the data, resulting in 1,940,218 and 1,872,221 particles from the first and second sets of images, respectively. These particles are separately subjected to three rounds of reference-free 2D classification. The 2D classes containing detailed features were selected from the first set and used to generate an initial model. We then performed 3D classification, using the initial model with a low-pass filter of 60Å, from 1,637,582 particles and 1,705,180 particles separately. After three rounds of global 3D classification, the resulting EM maps were aligned on pol II, and maps that had TFIIH at similar location relative to pol II were combined, resulting into three groups. The first group (454,296 particles) that was similar to the canonical PIC (Murakami et al., 2015b) was subjected to per-particle CTF refinement by first estimating beam shift, trefoil, and 4^{th} order aberrations, then magnification anisotropy, and finally per-particle defocus, and per-micrograph astigmatism. To improve the map quality of TFIIH, which was poorly ordered in the entire map, a soft mask was created around TFIIH of the 3D refined map, subtracted, and the subtracted images were used to generate an initial model. Then 3D classification was performed with image alignment. The resulting classes revealed two interpretable maps: strong (137,466 particles) and weak (101,497 particles) binding states of TFIIH, which were 3D-refined and post-processed to a resolution of 4.6Å and 7.6Å, respectively. All the reported resolutions are based on the gold-standard Fourier shell correlation (FSC) using 0.143 criterion. To further improve map quality of TFIIH, the maps were segmented into 3

bodies for multibody refinement: Tfb3, Rad3, Tfb1, Ssl1, Tfb4, and Tfb2N in body 1, Ssl2N in body 2, and Ssl2C, Tfb2C, and Tfb5 in body 3. The core PIC (cPIC) maps for both strong and weak binding states were generated by reverting the TFIIH maps to obtain entire maps, then 3D refinement using global search, and finally postprocessing pol II and TFIIB, and the rest of cPIC separately by applying appropriate masks. Resolution for the cPIC parts ranged between 3.0Å and 3.9Å.

The maps in the second group (384,412 particles) were combined and subjected to one more round of global 3D classification. For structural determination of PIC2, the resulting maps that showed the density for downstream dsDNA (117,450 particles) were combined and 3D refined. As in PIC1, focused refinement of TFIIH was performed, resulting in 33,159 particles in the best class. The per-particle CTF was determined as described above, and then TFIIH was 3D refined. This resulted TFIIH map with a resolution of 7.3Å. Multibody refinement (Nakane et al., 2018) was performed as described above to improve the map quality of TFIIH. To improve the map quality corresponding to downstream dsDNA, a soft mask around the DNA was created from the entire map obtained by reverting TFIIH and 3D refining using only local search, subtracted, and 3D classified. The map containing the best DNA density (11,028 particles) was reverted and 3D refined using only local search. This yielded the DNA density at 12.1Å resolution. To obtain cPIC2, the particles in the best TFIIH class was reverted and 3D refined using global search. Post-processing of the cPIC was performed as in PIC1. The resolution ranged between 4.0Å and 6.4Å.

To reconstruct PIC3, maps in the third group (307,173 particles) were combined and subjected to another round of 3D classification. Similar to PIC2 analysis above, classes containing downstream dsDNA (69,513 particles) were merged and 3D refined, and then TFIIH was subjected to a focused refinement. The best class was 3D refined and post-processed to a resolution of 11.8Å. The cPIC for PIC3 was obtained in the same manner as PIC1 and the resolution ranged between 4.1 and 7.6Å.

To reconstruct ITC, three maps that revealed the upstream edge of the bubble after one round of 3D classification of the third group (Figure S2A) were subjected to another round of 3D classification (Figure S3A). The resulting maps (227,346 particles) containing no downstream dsDNA were combined and then per-particle CTF was determined in the same manner as PIC particles above. The 3D auto-refined map did not show clear density for the DNA-RNA hybrid although some density was apparent in the active site of pol II, indicating variability of the DNA/RNA hybrid in position and length. Thus, to improve the quality of the hybrid density, a soft mask around the active site was generated, subtracted, and the resulting images were subjected to focused 3D classification without image alignment with higher regularization parameter (T=30). Classes that contained strong density for the hybrid (120,006 particles) were combined, reverted and 3D refined to obtain the entire map. cITC was postprocessed with a mask to a resolution of 3.1Å. The quality of the density for upstream DNA, TFIIE, TBP, Tfg2WH was also improved by a similar manner as the hybrid but with alignment during focused 3D classification, which yielded a map (86,069 particles) with a resolution of 6.8Å. To improve the map quality of TFIIH in the ITC, focused 3D classification with alignment was performed after per-particle CTF refinement. The best class (45,780 particles) was combined with focused refined map resulting from a subset of

particles in the second group that did not show density for downstream dsDNA, then the particles were subjected to one round of 3D classification. This yielded a 9.9Å TFIIH map for the ITC.

Model building of PIC1-3 and ITC—Maps with and without B-factor sharpening were used to build models of PIC1-3. For cPIC, the previous model from the yeast PIC (PDB ID: 5OQJ) was used as an initial template. For TFIIH, the previous model from the 3.9 Å resolution cryo-EM structure in a form of DNA repair (PDB ID: 7K01) (van Eeuwen et al., 2021a) was used as an initial model. Promoter DNA was manually built by combining short (~10bp) B-form DNA segments. A combined model containing cPIC, TFIIH, and promoter DNA was iteratively subjected to manual refinement (a combination of real-space refinement, regularization, and rigid body fit of domains) with Coot (Emsley et al., 2010) and rigid body refinement with Phenix1.16 (Liebschner et al., 2019). Through the refinement, secondary structures and base pairs of DNA double helix were maintained. The ITC was modeled essentially as in PIC1-3. The model of the 5-nt DNA-RNA hybrid with TFIIB was built using the previous X-ray crystallographic model (PDB ID: 4BBS) as a template, and refined using Coot and Phenix. Figures were prepared using UCSF Chimera (Pettersen et al., 2004).

Cryo-EM sample preparation with the G-less 49 template—The G-less 49 DNA fragment (−122/+97) (Fujiwara et al., 2019) was amplified by PCR and purified using Superose 6 10/300 (Cytiva) in buffer 300. To assemble PIC on the G-less 49 DNA template, the following were mixed in 240 μL of buffer 300: 0.26 μM DNA template, 0.4 μM TFIIA, 1.2 μM TFIIB, 2.4 μM TBP, 0.6 μM TFIIE, 1 μM TFIIF, 0.6 μM holoTFIIH, 0.36 μM TFIIK, 1.04 μM pol II, and 0.4 μM Sub1. The mixture was then diluted by adding an equal volume of buffer 10 (20 mM Hepes (pH 7.6), 10 mM potassium acetate, 5 mM magnesium sulfate, 5 mM DTT) and incubated on ice for 24 hours. After pre-incubation for 20 min at 30°C, 3/4th of the PIC mixture received 2x NTP solution consisting of 1.6 mM ATP, 1.6 mM CTP, 1 mM UTP, 0.5 μM 3'-O-methyl GTP, 10 mM magnesium acetate, and 0.5 U/μL RNaseOUT in buffer 10, for cryo-EM analysis, while 1/4th of the mixture received 2x NTP solution containing 44 nM [α-^{32}P] UTP (33 μCi) for characterization of proteins and RNA (Figures 4A-B). Transcription initiation was carried out for 20 min at 30°C and the sample was immediately loaded onto a gradient prepared with buffer A and buffer B as for the G-less 26 complex. 240 μl (with [α-^{32}P] UTP) and 720 μl (with cold UTP) were sedimented without and with glutaraldehyde, respectively. After centrifugation for 13 h at 30,000 rpm in a Beckman SW60 Ti rotor, the gradients were fractionated using a PGF Piston Gradient Fractionator (BioComp Instruments, Inc.) into ~130 μL per fraction and crosslinking reaction was quenched by addition of 40 mM glycine (pH 7.6). To perform RNA analysis of the fractions, 100 μL from non-crosslinked sample was incubated for 15 min at 42°C with 110 μL of stop buffer containing 390 mM sodium acetate (pH 5.5), 8 mM EDTA, 0.6% SDS, 0.06 mg/mL glycogen, 0.03 mg/mL proteinase K, and 0.03 mg/mL salmon sperm DNA, and subjected to ethanol precipitation, followed by RNA analysis by urea denaturing gel (Figure 4A). For protein analysis, 20 μL per fraction was analyzed by SDS-PAGE (Figure 4B).

To prepare cryo-EM grids, samples were dialyzed into EM buffer (20 mM HEPES (pH 7.6), 50 mM potassium acetate, 5 mM DTT, 2 mM magnesium acetate) for 30 minutes prior to making grids. EC+EC samples were applied to R1.2/1.3 400 mesh quantifoil holey carbon grids (Electron Microscopy Sciences), and EC+ITC samples were applied to R2/2 300 mesh quantifoil holey carbon grids (Electron Microscopy Sciences). All grids were glow-discharged (easiGlow, Pelco) for 2 min before deposition of 2uL of dialyzed sample, and subsequently blotted for 1.5 (EC+EC samples) or 2 seconds (EC +ITC samples) using Whatman Grade 41 filter paper (Sigma-Aldrich) and flash-frozen in liquid ethane with a Leica EM CPC manual plunger (Leica Microsystems). EM grids were prepared in batches and the freezing conditions were optimized by screening on a FEI TF20 microscope operating at 200 kV and equipped with a FEI Falcon III direct electron detection camera at the Electron Microscopy Research Lab (the University of Pennsylvania).

For EC+EC, two datasets (8872 and 7742 micrographs) were collected at Frederick National Laboratory (sponsored by the National Cancer Institute) using a Titan Krios transmission electron microscope operating at 300 kV, equipped with a K3 direct electron detector and a Bioquantum energy quantum filter. Images were collected by image shift and at a nominal magnification of 81,000x in super-resolution mode (pixel size of 0.54 Å) at a defocus range between −1 and −2.5 μm. The exposure time was 3.2 s at a nominal dose of 50 e^{-}/Å2, movies were divided into 40 frames.

Cryo-EM images of EC+ITC were collected at the Pacific Northwest Cryo-EM center using a Titan Krios transmission electron microscope operating at 300 kV, equipped with a K3 direct detection camera (Gatan) and a Bioquantum energy quantum filter. Data was collected by image shift and at a nominal magnification of 105,000x in super-resolution mode (pixel size of 0.415 Å) at a defocus range between – 0.9 and −2.2 μm. A total of 29,626 images were collected over 5 days. The exposure time was 2.1 s at a nominal dose of 45 e^{-}/Å2, movies were divided into 66 frames.

Image processing and 3D reconstruction of the EC+EC—Cryo-EM images of EC+EC were processed by a combination of cryoSPARC v3.1 (Punjani et al., 2017), Relion 3.1 (Zivanov et al., 2018), and Topaz (Bepler et al., 2019). The two datasets were motion-corrected with MotionCorr2 (Zheng et al., 2017) , and then CTF corrected with CTFFIND4 (Mindell and Grigorieff, 2003). A total of 1,630,930 particles were extracted with 340 pixel box after particle-picking using Topaz from the first dataset, and then the resultant particles were screened by two rounds of reference-free 2D classification, from which classes containing two ECs, accounting for 101,566 particles, were selected to calculate initial model. Subsequently one round of 3D classification was carried out, yielding four reasonable 3D classes. From the second dataset, 934,816 particles were extracted with 340 pixel box with Topaz, and then was subjected to three rounds of reference-free 2D classification followed by one round of 3D classification using the map obtained from the first dataset as a reference, yielding two reasonable 3D classes. The four 3D classes from the first dataset and the two classes from the second dataset were combined to perform further iterative rounds of 2D and 3D classifications, resulting in two maps showing the leading EC and the trailing EC. The resulting two 3D classes, accounting for a total of 107,093 particles, were subjected to 3D auto-refinement with a soft-edged mask, CTF-refinement,

Bayesian polishing, and Post-processing, leading to a reconstruction of the entire structure at 4.22 angstrom resolution. To push resolution, each elongation complex was subtracted using soft-edged masks encompassing the leading EC and the trailing EC respectively with 220 pixel box for each, and subjected to focused 3D classification followed by 3D auto-refinement. Lastly, a 3.5 angstrom map of the leading EC containing 57,690 particles and a 3.5 angstrom map of the trailing EC containing 66,261 particles were combined to generate a composite map using the vop maximum command in UCSF chimera (Pettersen et al., 2004).

All 2D classification, 3D classification, 3D refinement, Bayesian polishing, CTF-refinement, mask creation and post-process procedures described above were employed with Relion 3.1.0. Local resolution estimation for each elongation complex was performed with cryoSPARC v3.1. Resolution was reported on the basis of the gold-standard Fourier shell correlation (FSC) (0.143 criterion).

Cryo-EM images of EC +ITC were processed using a combination of Relion 3.1.1 (Zivanov et al., 2018), and SPHIRE-crYOLO (Wagner et al., 2019). Datasets were motion-corrected with MotionCorr2 (Zheng et al., 2017) then CTF corrected with CTFFIND4 (Mindell and Grigorieff, 2003). Particles were picked with SPHIRE-crYOLO and then extracted with 530 pixel box. A total of 988,153 particles were subjected to three rounds of 2D classification with Relion, yielding eight 2D classes accounting for ~9,707 particles (Figure 4D).

Model building and refinement of the EC+EC—Structural models were built using COOT (Emsley et al., 2010) and Phenix (Liebschner et al., 2019), the process was described as follows.

For the leading EC, structural models of pol II, the upstream DNA, the downstream DNA, the transcription bubble, and the DNA-RNA hybrid (PDB:5C4J) were placed into the map and subjected to rigid-body refinement with Phenix. The DNA-RNA hybrid was then manually refined using COOT against auto-sharpened map generated by Phenix. The 7 nt extended ssRNA (PDB : 6gml) in the exit tunnel was fitted into density, and then subjected to refinement with Phenix and COOT. The trailing EC was modeled as for the leading EC. The 5 nt backtracked ssRNA (15-19 nt, PDB:3PO2) was rigid body fitted into the density and then subjected to iterative refinement with Phenix and COOT. The TFIIF except Tfg2 WH domain (PDB:5FYW) was fitted into density using UCSF Chimera. The dsDNA bridging two elongation complexes was built with a B form DNA and then manually adjusted using COOT. Lastly, the entire model was interactively subjected to rigid-body refinement with Phenix and manual refinement with COOT. The final refinement was done with Phenix, with validation report (Table 1). All figures were generated using UCSF Chimera (Pettersen et al., 2004).

Tfb3ΔC protein expression and purification—The DNA sequence of WT Tfb3ΔC (residues 1-148) was obtained by PCR using *Saccharomyces cerevisiae* genome as a template and cloned into a pGV1594 vector with BamHI and HindIII restriction enzyme. The Tfb3 mutant R64A/K65A was constructed by QuickChange Site-Directed Mutagenesis using WT Tfb3ΔC as a template. Sequences of plasmids were confirmed by DNA

sequencing. Tfb3ΔC and the Tfb3 mutant R64A/K65A proteins were expressed in BL21 (DE3) cells with lysogeny broth (LB) medium supplemented with 50 μg/ml kanamycin. When OD600 reached 0.6, expression was induced by adding IPTG at a final concentration of 0.3 mM for 18 h at 18°C. The cells were harvested by centrifugation and the pellets were stored at −80°C. Two proteins were purified using the same protocol. The pellets were first resuspended in Ni binding buffer A (20 mM Hepes pH 7.6, 500 mM NaCl, 10 μM $ZnCl_2$, 2 mM DTT and 5% Glycerol) and lysed by sonication. The cell lysate was clarified by centrifugation and the supernatant was loaded onto a HisTrapTM HP column (Cytiva) equilibrated with Buffer A. The Mat-Sumo-Tfb3ΔC fusion proteins were eluted from the column using elution buffer B (20 mM Hepes pH 7.6, 500 mM NaCl, 10 μM $ZnCl_2$, 2 mM DTT, 5% Glycerol and 500 mM imidazole (pH 7.6)) with a gradient. Fractions containing the desired fusion proteins were pooled and dialyzed against Buffer A at 4°C overnight. Ulp1 protease was also added to the sample during the dialysis process to cleave the Mat-Sumo tag. The sample was again loaded onto the HisTrapTM HP column and the flow-through containing the target proteins was collected and mixed with buffer A without NaCl to dilute the NaCl to a final concentration of 100 mM. The diluted sample was applied to a Hi Trap Q column (Cytiva) equilibrated with binding buffer C (20 mM Hepes pH 7.6, 100 mM NaCl, 10 μM $ZnCl_2$, 5 mM DTT and 5% Glycerol), and then eluted from the column using elution buffer D (20 mM Hepes pH 7.6, 1000 mM NaCl, 10 μM $ZnCl_2$, 2 mM DTT, and 5% Glycerol) with a gradient. The target proteins were concentrated and loaded onto a Hi 16/60 Superdex G200 column (Cytiva) equilibrated with a gel filtration buffer (20 mM Hepes pH 7.6, 300 mM KOAc, 10 uM $Zn(OAc)_2$, 2 mM DTT and 5% Glycerol). Protein purity was analyzed using an SDS-PAGE gel and the samples were concentrated and snap-frozen using liquid nitrogen and stored at −80°C until use.

In vitro run-off transcription assay—The *SNR20* 91W promoter and its variant *SNR20* 31D promoter that lacks the region from site −68 to site −7 were previously described (Murakami et al., 2015a). Both promoter fragments (−122/+97) were amplified by PCR and purified using Superose 6 10/300 (Cytiva) in buffer 300. To assemble PIC on templates, the following were mixed in 5 μL of buffer 300: 0.26 μM DNA template, 0.4 μM TFIIA, 1.2 μM TFIIB, 2.4 μM TBP, 0.6 μM TFIIE, 1 μM TFIIF, 0.4 μM coreTFIIH, 1.04 μM pol II, 0.4 μM Sub1 and TFIIK (or Tfb3ΔC mutants) with given amount described above. The mixture was then diluted by adding an equal volume of buffer 10 (20 mM Hepes (pH 7.6), 10 mM potassium acetate, 5 mM magnesium sulfate, 5 mM DTT) and incubated on ice for 24 hours. Run-off transcription reaction was initiated by adding 10 ul of 2x NTP solution consisting of 1.6 mM ATP, 1.6 mM CTP, 1 mM UTP, 1.6 mM GTP, 44 nM [α-^{32}P] UTP (33 μCi), 10 mM magnesium acetate, and 0.5 U/μL RNaseOUT in buffer 10. Transcription was carried out for 30 min at 30°C then stopped by adding 190 ul of stop buffer containing 300 mM sodium acetate (pH 5.5), 5 mM EDTA, 0.7% SDS, 0.1 mg/ml glycogen, 0.013 mg/ml of proteinase K (Sigma). Transcripts were precipitated by adding 700 ml of ethanol, dried, and analyzed by a denaturing 6% acrylamide gel.

Transcription-related growth phenotypes of *tfb3* alleles.—Transcription-related growth phenotypes were analyzed as previously published (Braberg et al., 2013; Zhao et al., 2021). Yeast extract (1% w/v; BD), peptone (2% w/v; BD) and bacto-agar (2% w/v;

BD), supplemented with adenine (0.15 mM) and tryptophan (0.4 mM) (Sigma-Aldrich) comprised YP solid medium. YPD plates contained dextrose (2% w/v, VWR). Minimal media plates were prepared with synthetic complete (SC) or 'Hopkins mix' with appropriate amino acid(s) dropped out, with slight modifications as described in (Kaplan et al., 2012). For studies with MPA (Sigma-Aldrich) and 3-aminotriazole (3-AT) (Sigma-Aldrich), stock solutions (10 mg/ml, in 100% ethanol) of MPA and 3-AT were added to solid or liquid media to achieve desired concentration. *tfb3* alleles were generated by site-directed mutagenesis and verified by Sanger sequencing. Manipulation of *TFB3/tfb3* alleles in *IMD2* or *imd2Δ::HIS3* backgrounds were exactly as described for *ssl2* alleles in Zhao et al. (Zhao et al., 2021).

QUANTIFICATION AND STATISTICAL ANALYSIS

Local resolution estimation of the cryo-EM maps (Figures S7B-C) was performed with cryoSPARC v3.11 (Punjani et al., 2017). Resolution of the cryo-EM maps (Figure S2B-C and S3B, Figures S7D-E, and Table 1) was calculated on the basis of the gold-standard Fourier shell correlation (FSC) (0.143 criterion) using Relion 3.1 (Zivanov et al., 2018). Refinement statistics (Table 1) was generated with PHENIX (Liebschner et al., 2019).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We thank Dr. Sudheer Molugu for assistance with cryo-EM sample screening. We thank UMass CryoEM Facility, Drs. Chen Xu, KangKang Song and Kyounghwan Lee for their assistance in data collection of G-less 26 complexes. We thank NCI National Cryo-EM Facility, Drs. Thomas Edwards, Adam Wier under contract HSSN261200800001 for data collection of the G-less 49 complex (EC+EC). We thank Drs. Janette Myers, Nancy Meyer, and Harry Scott at PNCC for cryo-EM screening and data collection of G-less 49 complexes (EC+ITC). We thank members in the lab and Dr. Peter Geiduschek for helpful discussions.

Funding:

A portion of this research was supported by NIH grant U24GM129547 and performed at the PNCC at OHSU and accessed through EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research. This research was supported by NIH R01-GM123233 to K.M, NIH grants CA196539 and AG031862 to B.A.G, NIH R01-GM120450 to C.D.K. Computational resources were supported by NIH Project Grant S10OD023592.

References

Adachi N, Aizawa K, Kratzer Y, Saijo S, Shimizu N, and Senda T (2017). Improved method for soluble expression and rapid purification of yeast TFIIA. Protein Expr Purif 133, 50–56. [PubMed: 28259734]

Adelman K, and Lis JT (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet 13, 720–731. [PubMed: 22986266]

Aibara S, Schilbach S, and Cramer P (2021). Structures of mammalian RNA polymerase II pre-initiation complexes. Nature 594, 124–128. [PubMed: 33902107]

Baek I, Friedman LJ, Gelles J, and Buratowski S (2021). Single-molecule studies reveal branched pathways for activator-dependent assembly of RNA polymerase II pre-initiation complexes. Mol Cell.

Bahar Halpern K, Tanami S, Landen S, Chapal M, Szlak L, Hutzler A, Nizhberg A, and Itzkovitz S (2015). Bursty gene expression in the intact mammalian liver. Mol Cell 58, 147–156. [PubMed: 25728770]

Bardwell AJ, Bardwell L, Iyer N, Svejstrup JQ, Feaver WJ, Kornberg RD, and Friedberg EC (1994). Yeast nucleotide excision repair proteins Rad2 and Rad4 interact with RNA polymerase II basal transcription factor b (TFIIH). Mol Cell Biol 14, 3569–3576. [PubMed: 8196602]

Barnes CO, Calero M, Malik I, Graham BW, Spahr H, Lin G, Cohen AE, Brown IS, Zhang Q, Pullara F, et al. (2015). Crystal Structure of a Transcribing RNA Polymerase II Complex Reveals a Complete Transcription Bubble. Mol Cell 59, 258–269. [PubMed: 26186291]

Bartman CR, Hamagami N, Keller CA, Giardine B, Hardison RC, Blobel GA, and Raj A (2019). Transcriptional Burst Initiation and Polymerase Pause Release Are Key Control Points of Transcriptional Regulation. Mol Cell 73, 519–532 e514. [PubMed: 30554946]

Bepler T, Morin A, Rapp M, Brasch J, Shapiro L, Noble AJ, and Berger B (2019). Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. Nat Methods 16, 1153–1160. [PubMed: 31591578]

Braberg H, Jin H, Moehle EA, Chan YA, Wang S, Shales M, Benschop JJ, Morris JH, Qiu C, Hu F, et al. (2013). From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. Cell 154, 775–788. [PubMed: 23932120]

Bradsher J, Coin F, and Egly JM (2000). Distinct roles for the helicases of TFIIH in transcript initiation and promoter escape. J Biol Chem 275, 2532–2538. [PubMed: 10644710]

Bratkowski M, Unarta IC, Zhu L, Shubbar M, Huang X, and Liu X (2018). Structural dissection of an interaction between transcription initiation and termination factors implicated in promoter-terminator cross-talk. J Biol Chem 293, 1651–1665. [PubMed: 29158257]

Bushnell DA, Westover KD, Davis RE, and Kornberg RD (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. Science 303, 983–988. [PubMed: 14963322]

Cheung AC, and Cramer P (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature 471, 249–253. [PubMed: 21346759]

Chubb JR, Trcek T, Shenoy SM, and Singer RH (2006). Transcriptional pulsing of a developmental gene. Curr Biol 16, 1018–1025. [PubMed: 16713960]

Coin F, Proietti De Santis L, Nardo T, Zlobinskaya O, Stefanini M, and Egly JM (2006). p8/TTD-A as a repair-specific TFIIH subunit. Mol Cell 21, 215–226. [PubMed: 16427011]

Conaway RC, and Conaway JW (1993). General initiation factors for RNA polymerase II. Annu Rev Biochem 62, 161–190. [PubMed: 8352588]

Conaway RC, and Conaway JW (2012). The Mediator complex and transcription elongation. Biochim Biophys Acta.

Dienemann C, Schwalb B, Schilbach S, and Cramer P (2019). Promoter Distortion and Opening in the RNA Polymerase II Cleft. Mol Cell 73, 97–106 e104. [PubMed: 30472190]

Donovan BT, Huynh A, Ball DA, Patel HP, Poirier MG, Larson DR, Ferguson ML, and Lenstra TL (2019). Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting. EMBO J 38.

Dvir A, Conaway RC, and Conaway JW (1997). A role for TFIIH in controlling the activity of early RNA polymerase II elongation complexes. Proc Natl Acad Sci U S A 94, 9006–9010. [PubMed: 9256425]

Ehrensberger AH, Kelly GP, and Svejstrup JQ (2013). Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. Cell 154, 713–715. [PubMed: 23953103]

Emsley P, Lohkamp B, Scott WG, and Cowtan K (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66, 486–501. [PubMed: 20383002]

Fairman-Williams ME, Guenther UP, and Jankowsky E (2010). SF1 and SF2 helicases: family matters. Curr Opin Struct Biol 20, 313–324. [PubMed: 20456941]

Fazal FM, Meng CA, Murakami K, Kornberg RD, and Block SM (2015). Real-time observation of the initiation of RNA polymerase II transcription. Nature.

Feaver WJ, Svejstrup JQ, Henry NL, and Kornberg RD (1994). Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. Cell 79, 1103–1109. [PubMed: 8001136]

Fishburn J, Galburt E, and Hahn S (2016). Transcription Start Site Scanning and the Requirement for ATP during Transcription Initiation by RNA Polymerase II. J Biol Chem 291, 13040–13047. [PubMed: 27129284]

Fishburn J, Tomko E, Galburt E, and Hahn S (2015). Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation. Proc Natl Acad Sci U S A 112, 3961–3966. [PubMed: 25775526]

Fujiwara R, Damodaren N, Wilusz JE, and Murakami K (2019). The capping enzyme facilitates promoter escape and assembly of a follow-on preinitiation complex for reinitiation. Proc Natl Acad Sci U S A 116, 22573–22582. [PubMed: 31591205]

Fujiwara R, and Murakami K (2019). In vitro reconstitution of yeast RNA polymerase II transcription initiation with high efficiency. Methods 159-160, 82–89. [PubMed: 30905750]

Gervais V, Busso D, Wasielewski E, Poterszman A, Egly JM, Thierry JC, and Kieffer B (2001). Solution structure of the N-terminal domain of the human TFIIH MAT1 subunit: new insights into the RING finger family. J Biol Chem 276, 7457–7464. [PubMed: 11056162]

Gibbons BJ, Brignole EJ, Azubel M, Murakami K, Voss NR, Bushnell DA, Asturias FJ, and Kornberg RD (2012). Subunit architecture of general transcription factor TFIIH. Proc Natl Acad Sci U S A 109, 1949–1954. [PubMed: 22308316]

Greber BJ, Nguyen THD, Fang J, Afonine PV, Adams PD, and Nogales E (2017). The cryo-electron microscopy structure of human transcription factor IIH. Nature 549, 414–417. [PubMed: 28902838]

Greber BJ, Toso DB, Fang J, and Nogales E (2019). The complete structure of the human TFIIH core complex. Elife 8.

Gressel S, Schwalb B, Decker TM, Qin W, Leonhardt H, Eick D, and Cramer P (2017). CDK9-dependent RNA polymerase II pausing controls transcription initiation. Elife 6.

He Y, Fang J, Taatjes DJ, and Nogales E (2013). Structural visualization of key steps in human transcription initiation. Nature 495, 481–486. [PubMed: 23446344]

He Y, Yan C, Fang J, Inouye C, Tjian R, Ivanov I, and Nogales E (2016). Near-atomic resolution visualization of human transcription promoter opening. Nature 533, 359–365. [PubMed: 27193682]

Henry NL, Bushnell DA, and Kornberg RD (1996). A yeast transcriptional stimulatory protein similar to human PC4. J Biol Chem 271, 21842–21847. [PubMed: 8702984]

Kaplan CD, Jin H, Zhang IL, and Belyanin A (2012). Dissection of Pol II trigger loop function and Pol II activity-dependent control of start site selection in vivo. PLoS Genet 8, e1002627. [PubMed: 22511879]

Kappenberger J, Koelmel W, Schoenwetter E, Scheuer T, Woerner J, Kuper J, and Kisker C (2020). How to limit the speed of a motor: the intricate regulation of the XPB ATPase and translocase in TFIIH. Nucleic Acids Res 48, 12282–12296. [PubMed: 33196848]

Kettenberger H, Armache KJ, and Cramer P (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. Mol Cell 16, 955–965. [PubMed: 15610738]

Kokic G, Chernev A, Tegunov D, Dienemann C, Urlaub H, and Cramer P (2019). Structural basis of TFIIH activation for nucleotide excision repair. Nat Commun 10, 2885. [PubMed: 31253769]

Kornberg RD (2007). The molecular basis of eukaryotic transcription. Proc Natl Acad Sci U S A 104, 12955–12961. [PubMed: 17670940]

Kostrewa D, Zeller ME, Armache KJ, Seizl M, Leike K, Thomm M, and Cramer P (2009). RNA polymerase II-TFIIB structure and mechanism of transcription initiation. Nature 462, 323–330. [PubMed: 19820686]

Kuehner JN, and Brow DA (2006). Quantitative analysis of in vivo initiator selection by yeast RNA polymerase II supports a scanning model. J Biol Chem 281, 14119–14128. [PubMed: 16571719]

Larson DR, Zenklusen D, Wu B, Chao JA, and Singer RH (2011). Real-time observation of transcription initiation and elongation on an endogenous yeast gene. Science 332, 475–478. [PubMed: 21512033]

Lenstra TL, Coulon A, Chow CC, and Larson DR (2015). Single-Molecule Imaging Reveals a Switch between Spurious and Functional ncRNA Transcription. Mol Cell 60, 597–610. [PubMed: 26549684]

Lenstra TL, Rodriguez J, Chen H, and Larson DR (2016). Transcription Dynamics in Living Cells. Annu Rev Biophys 45, 25–47. [PubMed: 27145880]

Liebschner D, Afonine PV, Baker ML, Bunkoczi G, Chen VB, Croll TI, Hintze B, Hung LW, Jain S, McCoy AJ, et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol 75, 861–877. [PubMed: 31588918]

Liu X, Bushnell DA, Silva DA, Huang X, and Kornberg RD (2011). Initiation complex structure and promoter proofreading. Science 333, 633–637. [PubMed: 21798951]

Liu X, Bushnell DA, Wang D, Calero G, and Kornberg RD (2010). Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. Science 327, 206–209. [PubMed: 19965383]

Luo J, Cimermancic P, Viswanath S, Ebmeier CC, Kim B, Dehecq M, Raman V, Greenberg CH, Pellarin R, Sali A, et al. (2015). Architecture of the Human and Yeast General Transcription and DNA Repair Factor TFIIH. Mol Cell 59, 794–806. [PubMed: 26340423]

Luse DS (2013). Promoter clearance by RNA polymerase II. Biochim Biophys Acta 1829, 63–68. [PubMed: 22982364]

Luse DS (2019). Insight into promoter clearance by RNA polymerase II. Proc Natl Acad Sci U S A 116, 22426–22428. [PubMed: 31628251]

Malik I, Qiu C, Snavely T, and Kaplan CD (2017). Wide-ranging and unexpected consequences of altered Pol II catalytic activity in vivo. Nucleic Acids Res 45, 4431–4451. [PubMed: 28119420]

Mason PB, and Struhl K (2005). Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. Mol Cell 17, 831–840. [PubMed: 15780939]

Mindell JA, and Grigorieff N (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. J Struct Biol 142, 334–347. [PubMed: 12781660]

Murakami K, Calero G, Brown CR, Liu X, Davis RE, Boeger H, and Kornberg RD (2013a). Formation and fate of a complete 31-protein RNA polymerase II transcription preinitiation complex. J Biol Chem 288, 6325–6332. [PubMed: 23303183]

Murakami K, Elmlund H, Kalisman N, Bushnell DA, Adams CM, Azubel M, Elmlund D, Levi-Kalisman Y, Liu X, Gibbons BJ, et al. (2013b). Architecture of an RNA polymerase II transcription pre-initiation complex. Science 342, 1238724. [PubMed: 24072820]

Murakami K, Gibbons BJ, Davis RE, Nagai S, Liu X, Robinson PJ, Wu T, Kaplan CD, and Kornberg RD (2012). Tfb6, a previously unidentified subunit of the general transcription factor TFIIH, facilitates dissociation of Ssl2 helicase after transcription initiation. Proc Natl Acad Sci U S A 109, 4816–4821. [PubMed: 22411836]

Murakami K, Mattei PJ, Davis RE, Jin H, Kaplan CD, and Kornberg RD (2015a). Uncoupling Promoter Opening from Start-Site Scanning. Mol Cell 59, 133–138. [PubMed: 26073544]

Murakami K, Tsai KL, Kalisman N, Bushnell DA, Asturias FJ, and Kornberg RD (2015b). Structure of an RNA polymerase II preinitiation complex. Proc Natl Acad Sci U S A 112, 13543–13548. [PubMed: 26483468]

Nakane T, Kimanius D, Lindahl E, and Scheres SH (2018). Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. Elife 7.

Nguyen VQ, Ranjan A, Liu S, Tang X, Ling YH, Wisniewski J, Mizuguchi G, Li KY, Jou V, Zheng Q, et al. (2021). Spatiotemporal coordination of transcription preinitiation complex assembly in live cells. Mol Cell.

O'Brien T, and Lis JT (1991). RNA polymerase II pauses at the 5' end of the transcriptionally induced Drosophila hsp70 gene. Mol Cell Biol 11, 5285–5290. [PubMed: 1922045]

Pal M, Ponticelli AS, and Luse DS (2005). The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. Mol Cell 19, 101–110. [PubMed: 15989968]

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25, 1605–1612. [PubMed: 15264254]

Plaschka C, Hantsche M, Dienemann C, Burzinski C, Plitzko J, and Cramer P (2016). Transcription initiation complex structures elucidate DNA opening. Nature 533, 353–358. [PubMed: 27193681]

Punjani A, Rubinstein JL, Fleet DJ, and Brubaker MA (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14, 290–296. [PubMed: 28165473]

Qiu C, Jin H, Vvedenskaya I, Llenas JA, Zhao T, Malik I, Visbisky AM, Schwartz SL, Cui P, Cabart P, et al. (2020). Universal promoter scanning by Pol II during transcription initiation in Saccharomyces cerevisiae. Genome Biol 21, 132. [PubMed: 32487207]

Ranish JA, Hahn S, Lu Y, Yi EC, Li XJ, Eng J, and Aebersold R (2004). Identification of TFB5, a new component of general transcription and DNA repair factor IIH. Nat Genet 36, 707–713. [PubMed: 15220919]

Robinson PJ, Trnka MJ, Bushnell DA, Davis RE, Mattei PJ, Burlingame AL, and Kornberg RD (2016). Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. Cell 166, 1411–1422 e1416. [PubMed: 27610567]

Rosen GA, Baek I, Friedman LJ, Joo YJ, Buratowski S, and Gelles J (2020). Dynamics of RNA polymerase II and elongation factor Spt4/5 recruitment during activator-dependent transcription. Proc Natl Acad Sci U S A 117, 32348–32357. [PubMed: 33293419]

Saeki H, and Svejstrup JQ (2009). Stability, flexibility, and dynamic interactions of colliding RNA polymerase II elongation complexes. Mol Cell 35, 191–205. [PubMed: 19647516]

Sainsbury S, Niesser J, and Cramer P (2013). Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. Nature 493, 437–440. [PubMed: 23151482]

Schilbach S, Aibara S, Dienemann C, Grabbe F, and Cramer P (2021). Structure of RNA polymerase II pre-initiation complex at 2.9 A defines initial DNA opening. Cell 184, 4064–4072 e4028. [PubMed: 34133942]

Schilbach S, Hantsche M, Tegunov D, Dienemann C, Wigge C, Urlaub H, and Cramer P (2017). Structures of transcription pre-initiation complex with TFIIH and Mediator. Nature 551, 204–209. [PubMed: 29088706]

Schultz P, Fribourg S, Poterszman A, Mallouh V, Moras D, and Egly JM (2000). Molecular structure of human TFIIH. Cell 102, 599–607. [PubMed: 11007478]

Shao W, and Zeitlinger J (2017). Paused RNA polymerase II inhibits new transcriptional initiation. Nat Genet 49, 1045–1051. [PubMed: 28504701]

Sikorski TW, Ficarro SB, Holik J, Kim T, Rando OJ, Marto JA, and Buratowski S (2011). Sub1 and RPA associate with RNA polymerase II at different stages of transcription. Mol Cell 44, 397–409. [PubMed: 22055186]

Spangler L, Wang X, Conaway JW, Conaway RC, and Dvir A (2001). TFIIH action in transcription initiation and promoter escape requires distinct regions of downstream promoter DNA. Proc Natl Acad Sci U S A 98, 5544–5549. [PubMed: 11331764]

Svejstrup JQ, Wang Z, Feaver WJ, Wu X, Bushnell DA, Donahue TF, Friedberg EC, and Kornberg RD (1995). Different forms of TFIIH for transcription and DNA repair: holo-TFIIH and a nucleotide excision repairosome. Cell 80, 21–28. [PubMed: 7813015]

Tomko EJ, Luyties O, Rimel JK, Tsai CL, Fuss JO, Fishburn J, Hahn S, Tsutakawa SE, Taatjes DJ, and Galburt EA (2021). The Role of XPB/Ssl2 dsDNA Translocase Processivity in Transcription Start-site Scanning. J Mol Biol 433, 166813. [PubMed: 33453189]

Tsai KL, Yu X, Gopalan S, Chao TC, Zhang Y, Florens L, Washburn MP, Murakami K, Conaway RC, Conaway JW, et al. (2017). Mediator structure and rearrangements required for holoenzyme formation. Nature 544, 196–201. [PubMed: 28241144]

van Eeuwen T, Li T, Kim HJ, Gorbea Colon JJ, Parker MI, Dunbrack RL, Garcia BA, Tsai KL, and Murakami K (2021a). Structure of TFIIK for phosphorylation of CTD of RNA polymerase II. Sci Adv 7.

van Eeuwen T, Shim Y, Kim HJ, Zhao T, Basu S, Garcia BA, Kaplan CD, Min JH, and Murakami K (2021b). Cryo-EM structure of TFIIH/Rad4-Rad23-Rad33 in damaged DNA opening in nucleotide excision repair. Nat Commun 12, 3338. [PubMed: 34099686]

Vos SM, Farnung L, Urlaub H, and Cramer P (2018). Structure of paused transcription complex Pol II-DSIF-NELF. Nature 560, 601–606. [PubMed: 30135580]

Wade JT, and Struhl K (2008). The transition from transcriptional initiation to elongation. Curr Opin Genet Dev 18, 130–136. [PubMed: 18282700]

Wagner T, Merino F, Stabrin M, Moriya T, Antoni C, Apelbaum A, Hagel P, Sitsel O, Raisch T, Prumbaum D, et al. (2019). SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. Commun Biol 2, 218. [PubMed: 31240256]

Wang D, Bushnell DA, Huang X, Westover KD, Levitt M, and Kornberg RD (2009). Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. Science 324, 1203–1206. [PubMed: 19478184]

Yudkovsky N, Ranish JA, and Hahn S (2000). A transcription reinitiation intermediate that is stabilized by activator. Nature 408, 225–229. [PubMed: 11089979]

Zenklusen D, Larson DR, and Singer RH (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol 15, 1263–1271. [PubMed: 19011635]

Zhao T, Vvedenskaya IO, Lai WK, Basu S, Pugh BF, Nickels BE, and Kaplan CD (2021). Ssl2/TFIIH function in transcription start site scanning by RNA Polymerase II in Saccharomyces cerevisiae. Elife 10.

Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, and Agard DA (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods 14, 331–332. [PubMed: 28250466]

Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, and Scheres SH (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife 7.

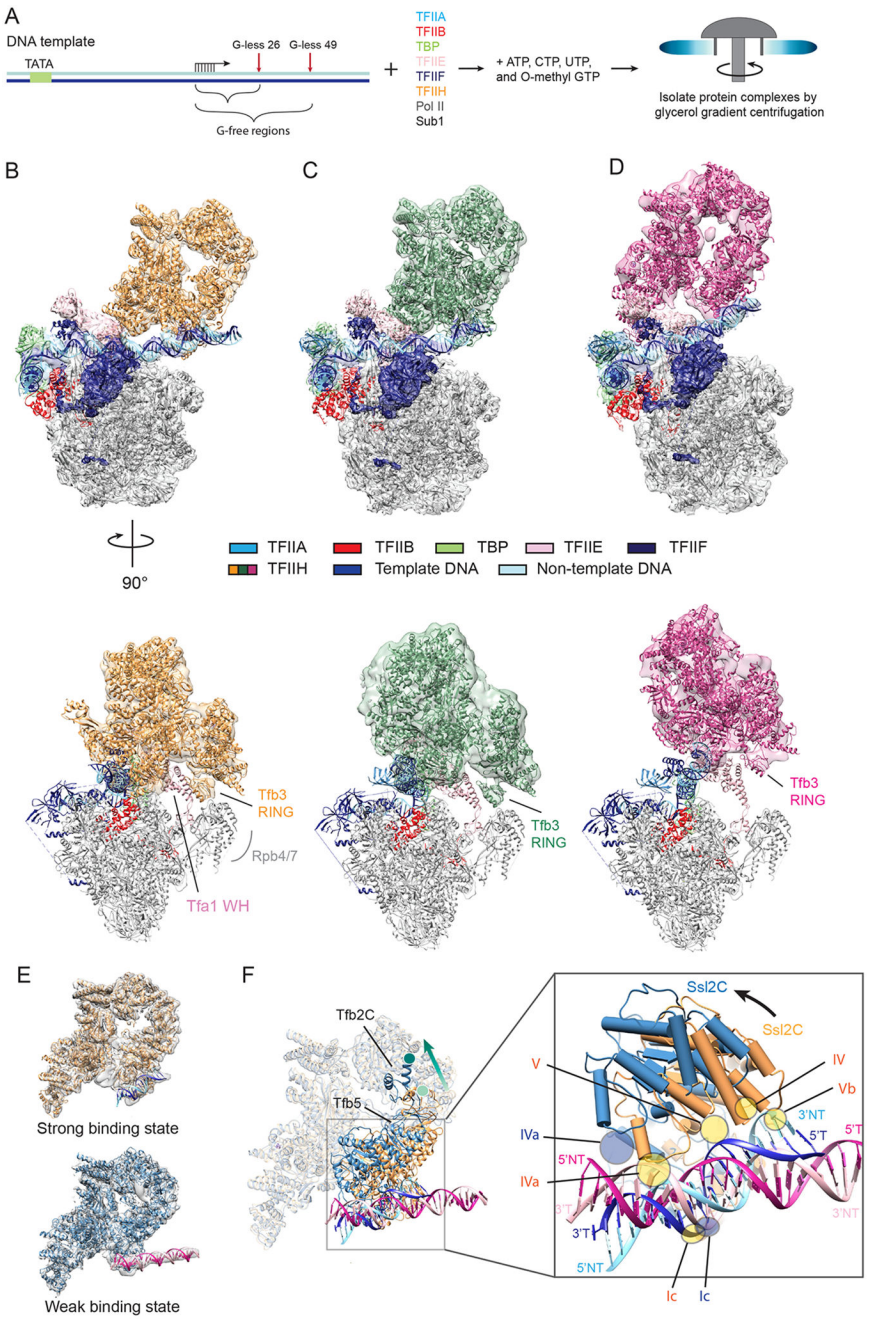


Figure 1. Structures of three forms of pre-initiation complexes.

(A) Schematic representation for isolation of the transcription complexes analyzed in this study. **(B-D)** Models of three forms of PIC obtained on the G-less 26 DNA template; PIC1 (B), PIC2 (C), and PIC3 (D) with composite density maps (top for side view, bottom for front view). For the front view (bottom), only maps for TFIIH are shown to highlight densities attributable to the Tfb3 RING finger domain. Same colors are used throughout the manuscript unless otherwise noted: TFIIA (steel blue), TFIIB (red), TBP (light green), TFIIE (pink), TFIIF (dark blue), TFIIH (orange, green, or magenta), template DNA (blue), non-template DNA (sky blue). The Tfb3 RING finger domain interacts with Rpb4/7 in PIC1

but dissociates in PIC2 and PIC3. **(E)** Composite density maps and models of strong (top) and weak (bottom) binding states of TFIIH, with corresponding models in orange and steel blue, respectively. **(F)** Comparison of TFIIH and downstream DNA in the strong (blue) and weak (purple) binding states. The C-terminal residue of Tfb2C in the two states is marked by green dots and the direction of movement is indicated by an arrow. Inset, Ssl2–DNA interactions suggested by the model. The five DNA binding motifs (Ic, IVa, IV, V, Vb) are shown by yellow and blue circles in the strong and weak binding states, respectively (see also Figure S4). See also Figures S1, S2, S3, S4, S6 and Movie S1.

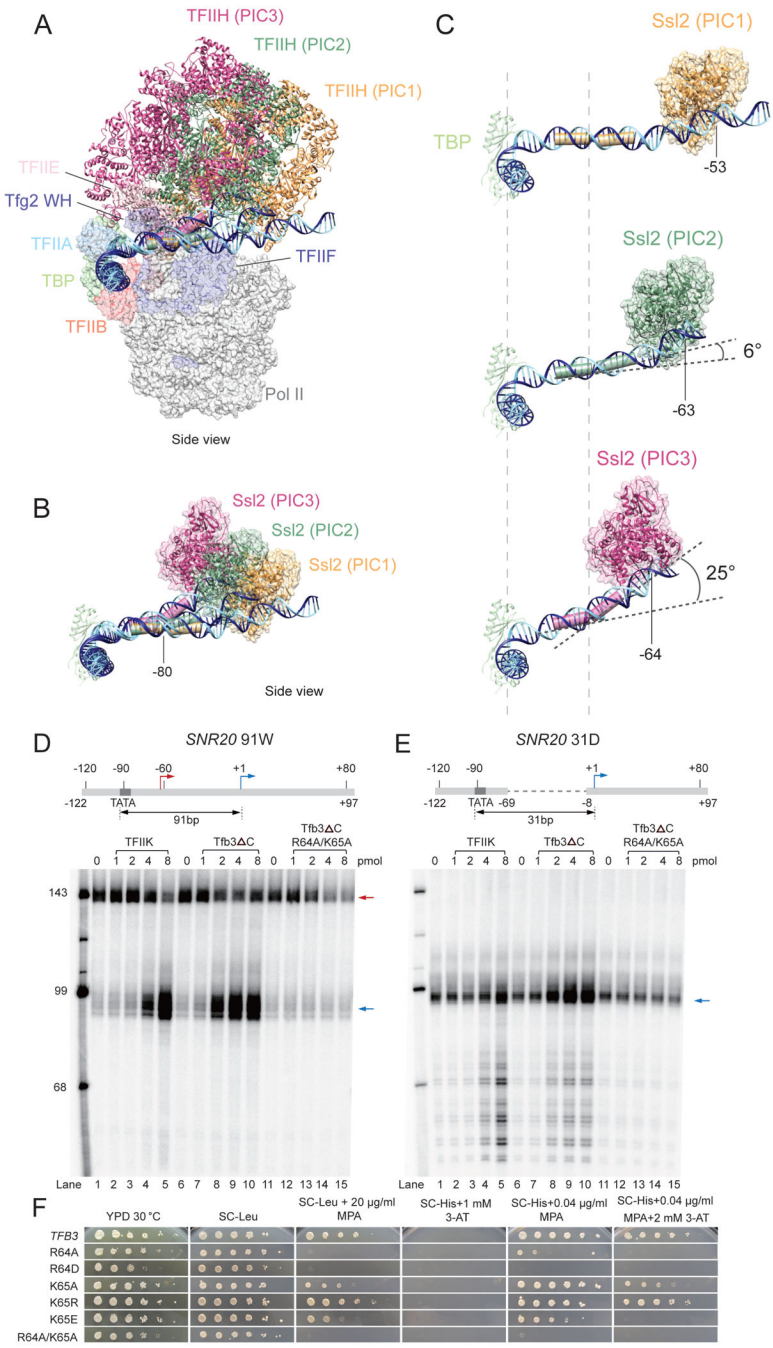


Figure 2. Distortion of promoter DNA in PIC1-3.

Coloring as in Figure 1. **(A)** Comparison of locations of TFIIH and DNA in PIC1-3 relative to Pol II. TFIIH shifts ~ 20 Å and ~30 Å in PIC2 and PIC3, respectively, relative to PIC1. **(B)** Paths of promoter DNA. Ssl2 contacting DNA is shown. **(C)** Ssl2 binds ~47 bp and ~37 bp downstream of the TATA box in PIC1 and PIC2/PIC3, respectively. DNA bends by ~6° and ~25° in PIC2 and PIC3, respectively. Dashed lines indicate positions of the bend at ~−80 and the TATA box at −100. The numbering is relative to the TSS. **(D)** Run-off transcription with increasing amounts of TFIIK, Tfb3ΔC, and Tfb3ΔC R64A/K65A. *SNR20*

91W promoter DNA (−122/+97) was combined with TFIIA, TFIIB, TBP, Sub1, TFIIE, TFIIF, TFIIH-ΔTFIIK, pol II, and the amounts of TFIIK and Tfb3ΔC variants indicated. 2 pmol TFIIH-ΔTFIIK was used in *in vitro* transcription assays. Transcripts initiated from upstream and downstream TSSs are indicated by red and blue arrows. **(E)** Same as (D) with *SNR20* 31D promoter DNA (−122/+97). **(F)** 10-fold serial dilutions of saturated cultures of Tfb3 WT and mutant strains plated on different media. YPD is rich medium with dextrose as a carbon source. SC-Leu is defined, complete medium lacking leucine. MPA was added to this medium (SC-Leu+MPA) to 20 μg/ml final concentration, showing that R64A and R64D are sensitive to this drug, indicative of defects in TSS scanning at the *IMD2* gene, which is required for resistance. The three rightmost panels show mutant phenotypes in an *imd2*Δ*::HIS3* reporter strain, where *HIS3* expression is under control of the *IMD2* promoter. Control defined medium lacking histidine but containing the His3p inhibitor 3-aminotriazole (3-AT) at 1 mM indicates the *IMD2* promoter is not abnormally active in the control condition. Addition of 0.04 μg/ml MPA activates the *IMD2* promoter and defects in activation are consistent with TSS scanning defects and are reflected by no or poor growth on SC-His+0.04μg/ml MPA. Addition of 2 mM 3-AT provides a more stringent test for the His+ phenotype. See also Figure S5.

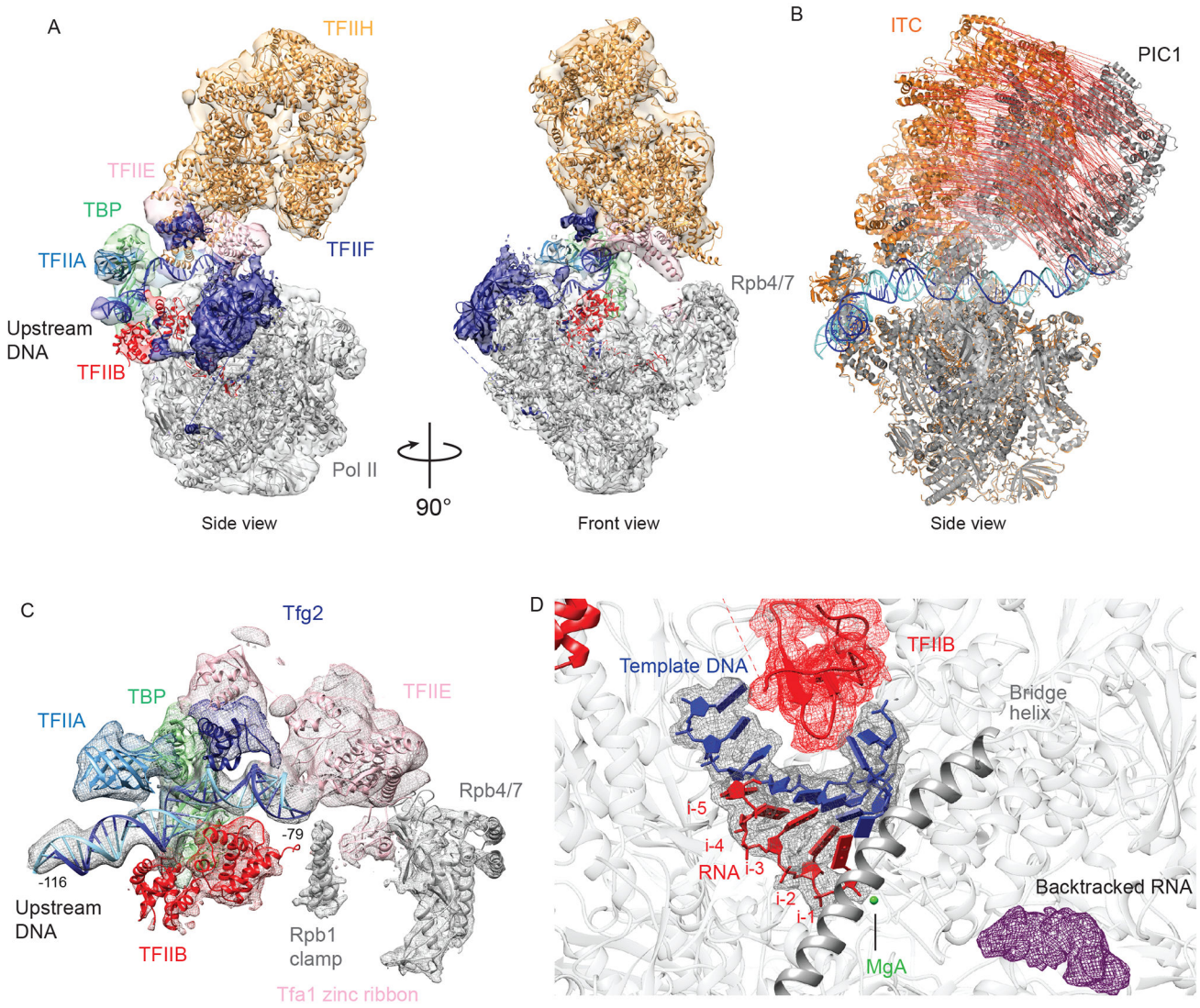


Figure 3. Structure of ITC with the G-less 26 template.

(A) Cryo-EM map and a corresponding model of the ITC. **(B)** Structural changes of ITC relative to PIC1. Two structures are aligned by superimposing pol II. The red lines represent the movement of TFIIH. ITC is shown in orange, while PIC1 is in gray, DNA in both structures is blue and cyan. **(C)** EM density shows the upstream DNA bound to TFIIA, TBP, TFIIB (cyclin domains), TFIIE (Tfa1 and Tfa2 WH1 and WH2 domains), and TFIIF (Tfg2 WH domains). **(D)** EM density map shows the DNA-RNA hybrid in the pol II active center in contact with TFIIB. Density attributable to the backtracked RNA is shown in purple. See also Figure S6.

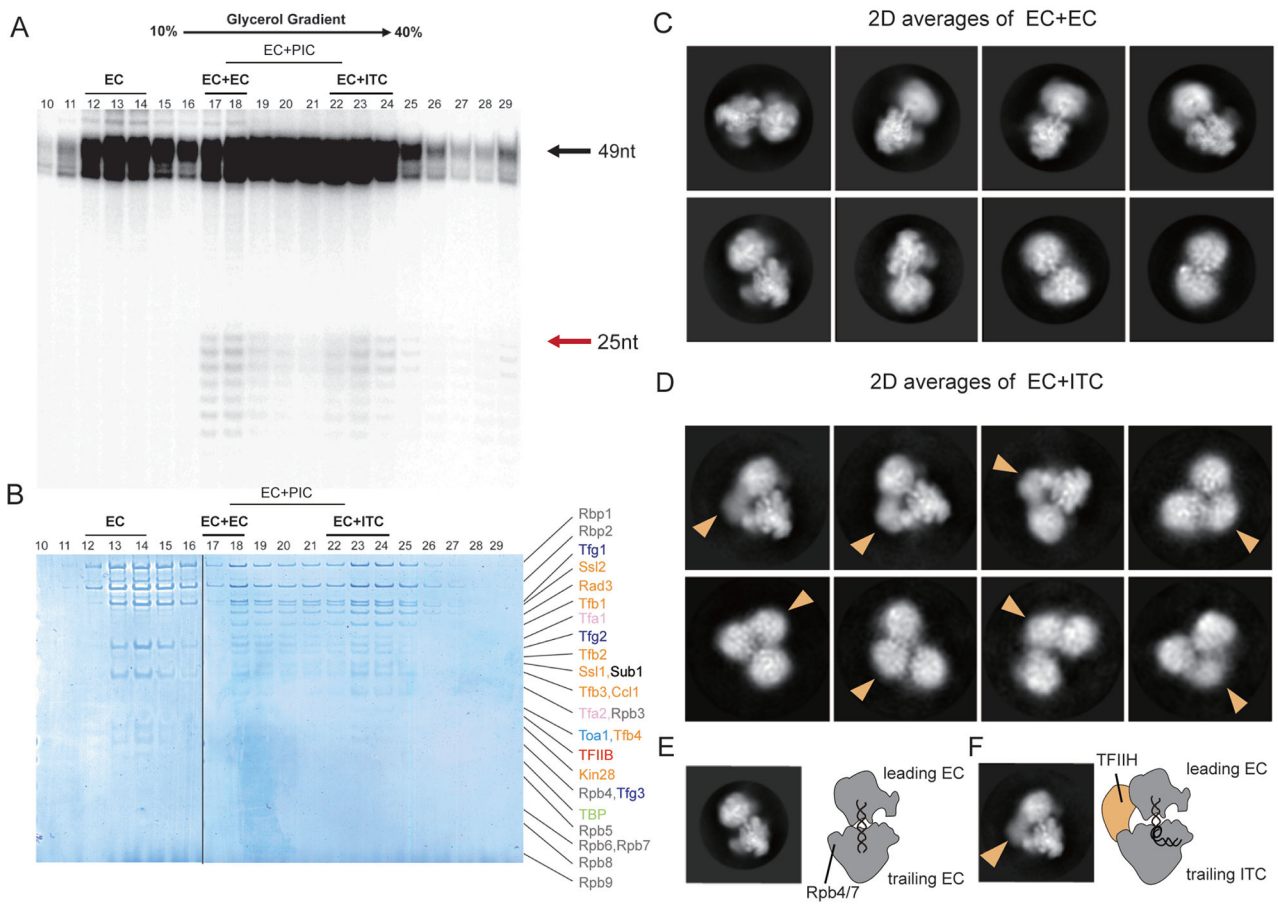


Figure 4. G-less 49 complexes reveal EC+EC and EC+ITC.

(A-B) Transcription complexes with the G-less 49 template were subjected to glycerol gradient sedimentation. RNA analysis of the fractions by denaturing Urea-PAGE gel (A) and protein analysis of the fractions by SDS-PAGE gel (B) revealed EC+EC (fractions 17-18) and EC+ITC (fractions 22-24). 49-nt and 25-nt transcripts from the first and the second rounds of transcription are indicated by black and red arrows. Note that both complexes have some contamination of EC+PIC (fractions 18-21). **(C)** Eight representative reference-free 2D class averages of EC+EC. **(D)** Same as (C) but for EC+ITC. A large density attributable to TFIIH (indicated by orange arrow heads) is located between EC and ITC. **(E)** A representative 2D class average of EC+EC, with a schematic model, showing the two well-featured densities corresponding to the leading EC and trailing EC, respectively. The density of DNA bridging two polymerases is clearly discernable. **(F)** Same as (E) but for EC+ITC, showing a large density attributable to TFIIH (orange) compared with 2D class averages of EC + EC. The density of DNA is clearly discernable as in (E). See also Figures S7 and S8.

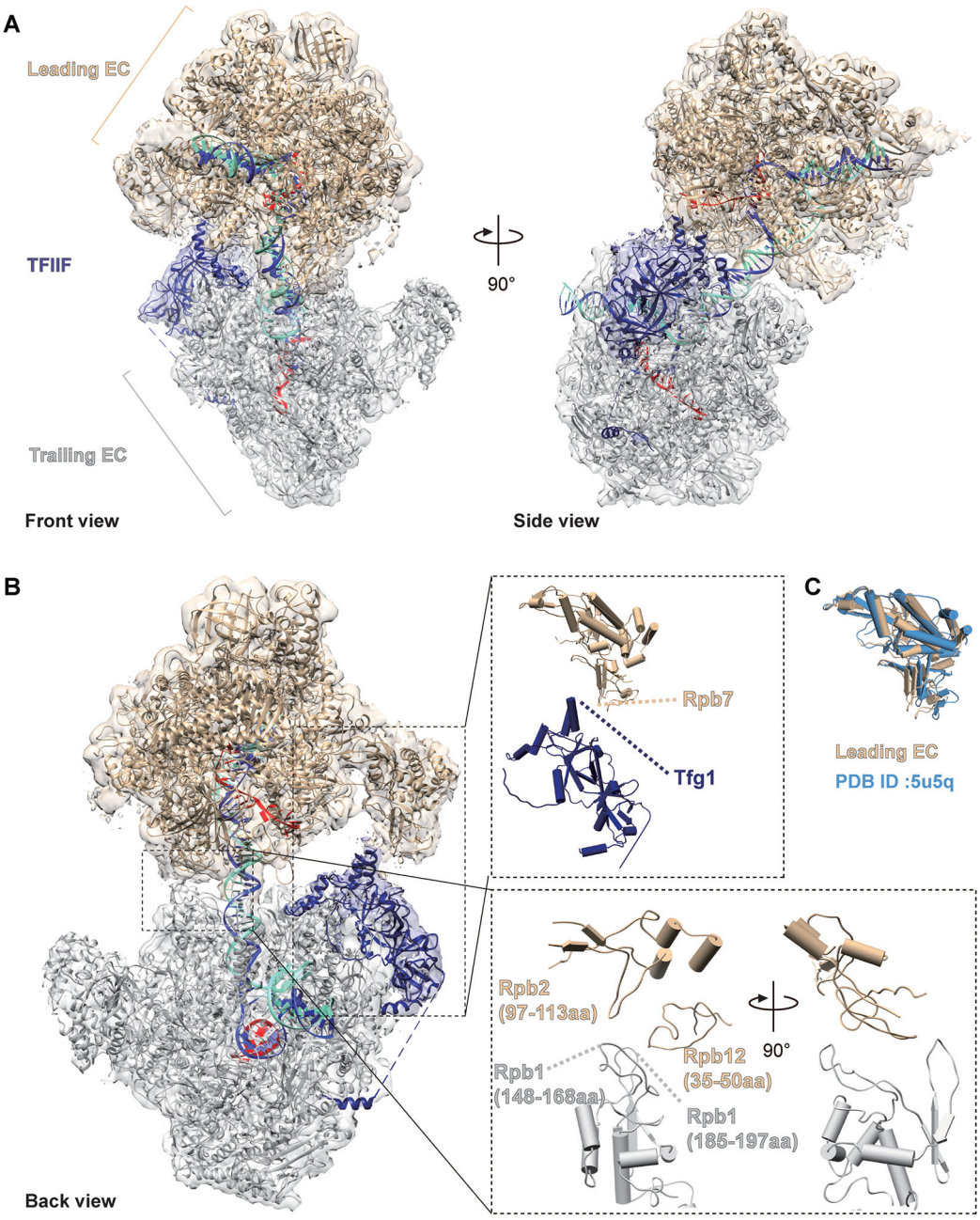


Figure 5. The structure of ECs colliding head-to-end (EC+EC).

(A) Front (left) and side views (right) of the cryo-EM reconstruction with the model. **(B)** Interactions between two ECs viewed from the back. The Rpb4/7 of the leading EC contacts TFIIF associated with the trailing EC, while Rpb2 and Rpb12 of the leading EC contacts the Rpb1 of the trailing EC. **(C)** Superposition of the Rpb4/7 of the leading EC with that of polII previously determined by X-ray crystallography (PDB ID 5U5Q) ,which shows ~10°-rotation of Rpb4/7. The leading EC, the trailing EC, TFIIF, template DNA, non-template DNA and RNA are colored in tan, gray, navy blue, blue, aquamarine and red, respectively. See also Movie S2.

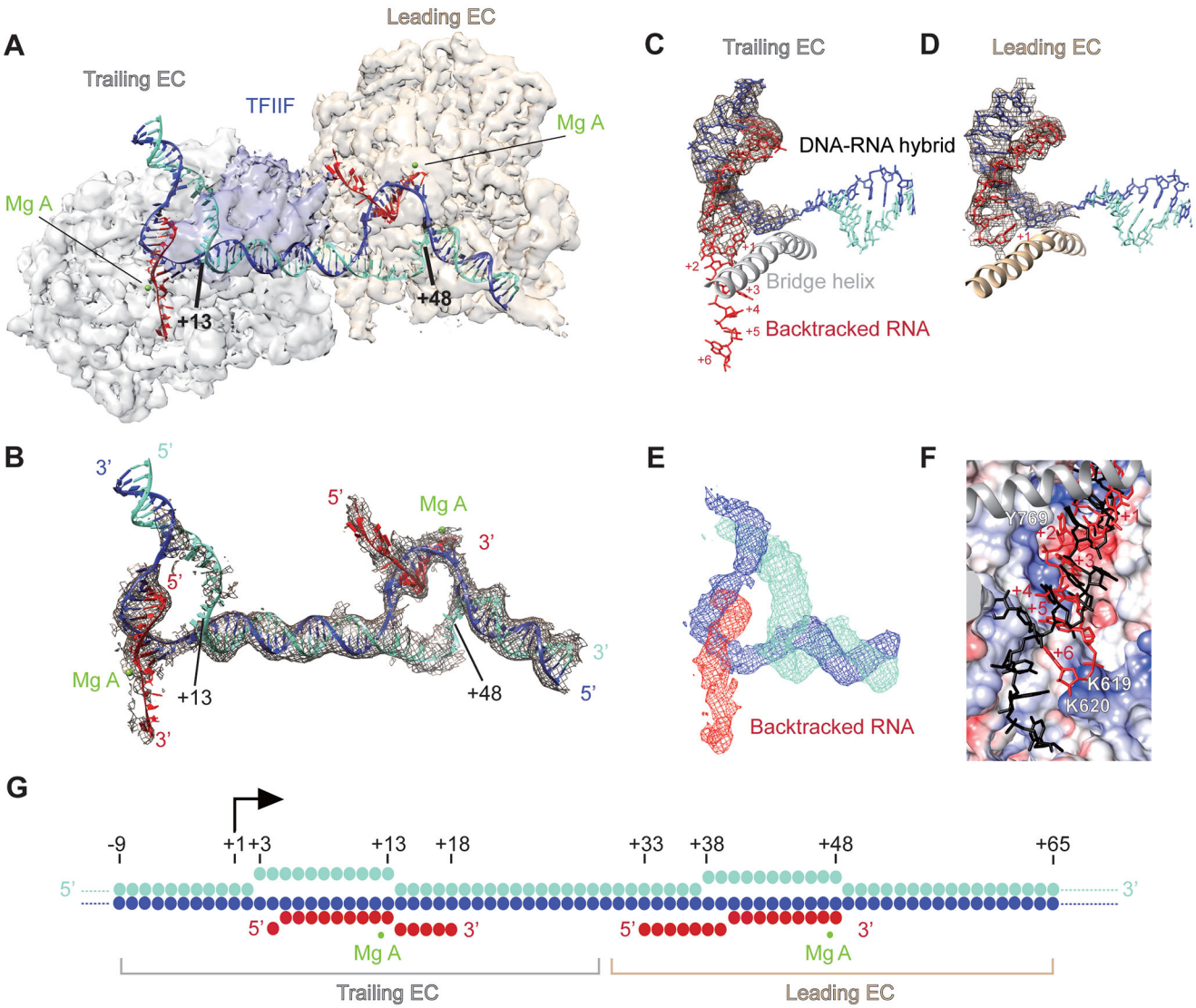


Figure 6. Nucleic acids of ECs colliding head-to-end (EC+EC).

(A) Unsharpened cryo-EM densities of the trailing EC (gray), the leading EC (tan) and TFIIF (navy blue) are shown as surface, contoured at 2.07 sigma. Template DNA, non-template DNA and RNA are colored in blue, aquamarine and red throughout. Mg A in active center is shown as sphere and colored in green. **(B)** Composite cryo-EM density of nucleic acids. Unsharpened cryo-EM densities are shown as mesh (space gray), contoured at level 2.07 sigma. **(C)** The DNA-RNA hybrid of the trailing EC. Sharpened cryo-EM densities of the DNA-RNA hybrid of the trailing EC are shown as mesh (space gray), contoured at level 4.1 sigma. The bridge helix is shown in gray. **(D)** Same as (C) but for the leading EC, contoured at 3.86 sigma. **(E)** Cryo-EM density of nucleic acids of the trailing EC showing the backtracked RNA (red). Unsharpened cryo-EM densities of nucleic acids are shown as mesh (RNA, red; template DNA, blue; Non-template DNA, aquamarine), contoured at level 1.6 sigma. **(F)** Superposition of the backtracked RNA of the trailing EC (red) with the backtrack site of the arrested pol II previously determined by crystallography (PDB:3PO2, black). Backtracked RNA in the trailing EC shifts towards a positive charged patch that

consists of K619 and K620 of Rpb1 in the funnel. **(G)** Schematic of nucleic acids. Modelled nucleotides of promoter template are shown with filled circles. TSS (+1) is indicated by black arrow. See also Figure S9.

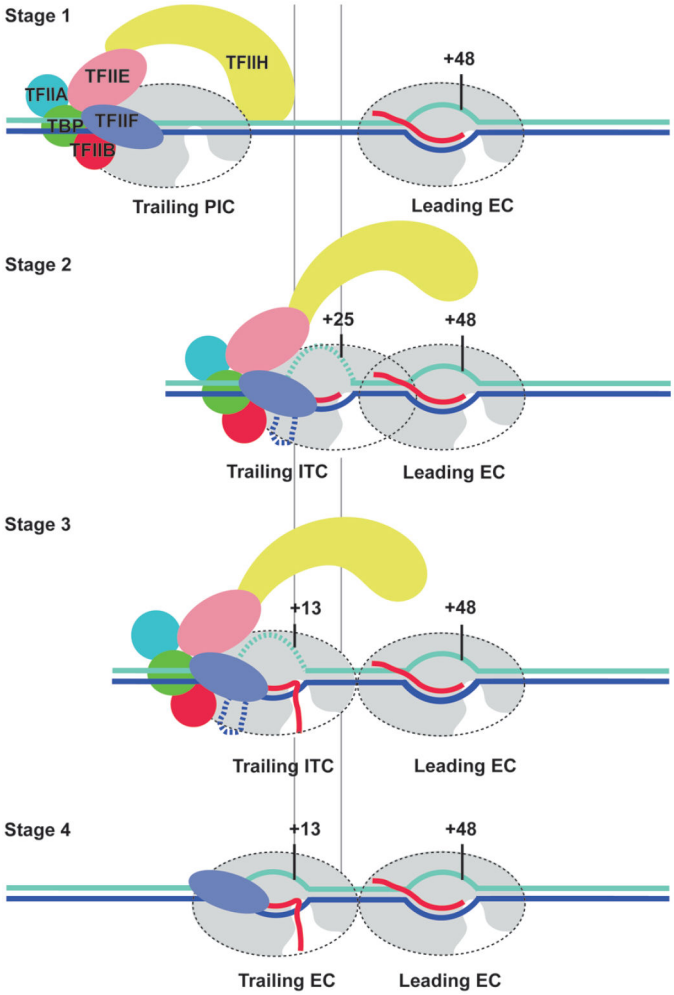


Figure 7. Schematic of promoter escape facilitated by the leading EC.

A trailing PIC is assembled by re-utilizing the promoter, while a leading EC is stalled at promoter proximal regions (Stage 1). The trailing PIC in the initially-transcribing state (ITC) extends a transcript and encounters the leading EC, resulting in dissociation of TFIIH from DNA due to steric occlusion by two transcribing pol II molecules (Stage 2). The trailing ITC is backtracked from +25 to +13 (Stage 3), followed by dissociation of GTFs and bubble collapse (Stage 4). See also Figures S10 and S11.

Table 1.

Cryo-EM Data Collection and Refinement Statistics

	PIC-1	TFIIH-2	PIC-2	PIC-3	ITC	EC+EC
Data collection and processing						
Magnification	81,000	81,000	81,000	81,000	81,000	81000
Voltage (kV)	300	300	300	300	300	300
Electron exposure (e^{-}/Å^2)	45	45	45	45	45	50
Defocus range (μm)	−0.5 to −2.5	−0.5 to −2.5	−0.5 to −2.5	−0.5 to −2.5	−0.5 to −2.5	−1 to −2.5
Pixel size (Å)	0.53	0.53	0.53	0.53	0.53	0.54
Map sharpening B factor (Å)	−10 to −100 *	0 to −233 *	0 to −300 *	0 to −40 *	0 *	−34.8 [e] / −19.9 [f]
Map resolution (Å)	3.0 [a] /3.8 [b] /4.6 [c]	7.6	4.0 [a] /6.4 [b] /7.3 [c] /12.1 [d]	4.1 [a] /7.6 [b] /11.8 [c]	3.1 [a] /6.8 [b] /9.9 [c]	3.5 [e] / 3.5 [f]
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143
EMDB entry	23904	23907	23905	23906	23908	23789
Initial particle #			1848899			1418609
Final particle #	454296	101497	117450	69513	322399	107093
Model Refinement						
Model resolution (Å)	3.40 (2.61)	8.36 (4.16)	4.21 (3.55)	6.71 (3.60)	4.10 (3.32)	4.9 (4.0)
FSC threshold	0.5 (0.143)	0.5 (0.143)	0.5 (0.143)	0.5 (0.143)	0.5 (0.143)	0.5 (0.143)
PDB entry	7ML0	7ML3	7ML1	7ML2	7ML4	7MEI
Model composition						
Non-hydrogen atoms	64,262	22,609	64,586	64,552	62,812	69,294
Protein residues	7,991	2,847	8,081	8,082	7,916	8284
Nucleotides	132	58	114	112	91	179
Ligands	139	129	142	140	138	20
R.m.s deviations						
Bond lengths (Å)	0.018	0.024	0.005	0.008	0.006	0.005
Bond angles (°)	1.114	1.608	0.646	1.131	0.726	0.816
Validation						
MolProbity score	3.71	3.09	3.23	2.70	2.34	2.41
Clashscore	36.14	89.37	25.10	39.55	20.78	21.48
Poor rotamer (%)	20.12	0.00	12.16	0.00	0.38	0.07
Ramachandran plot						
Favored (%)	84.08	84.55	91.04	86.73	90.82	88.62
Allowed (%)	15.61	15.31	8.88	13.13	9.04	11.30
Disallowed (%)	0.31	0.14	0.08	0.14	0.14	0.07
Model vs. Data						
CC (mask)	0.60	0.61	0.60	0.39	0.47	0.69
CC (box)	0.70	0.75	0.77	0.73	0.57	0.87
CC (volume)	0.65	0.59	0.61	0.37	0.56	0.70
CC (peaks)	0.56	0.49	0.53	0.26	0.45	0.67

	PIC-1	TFIIH-2	PIC-2	PIC-3	ITC	EC+EC
CC (main chain)	0.64	0.70	0.70	0.61	0.49	0.77
CC (side chain)	0.64	0.70	0.70	0.63	0.49	0.74

[a] Pol II, TFIIB

[b] DNA, TBP, TFIIE, TFIIF

[c] TFIIH

[d] downstream DNA

[e] Leading EC

[f] Trailing EC

[*] see FigS2, S3 for the B factor values applied.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
BL21(DE3) Competent Cells	Thermo Fisher Scientific	Cat#EC0114
Chemicals, peptides, and recombinant proteins		
25% Glutaraldehyde Solution in H_2O, EM grade	Sigma Aldrich	Cat#111-30-8
4-thiouridine-5'-triphosphate	TriLink	Cat#N-1025-5
3'-O-methylguanosine-5'triphosphate	TriLink	Cat#N-1058-5
100 mM ATP, CTP, UTP	Thermo Fisher Scientific	Cat#R0481
Salmon sperm DNA	Thermo Fisher Scientific	Cat#15632011
Glycogen	Roche	Cat#10901393001
Proteinase K	Sigma Aldrich	Cat#P4850
RNaseOUT Recombinant Ribonuclease Inhibitor	Thermo Fisher Scientific	Cat#10777019
UTP [α-^{32}P]	PerkinElmer	Cat#NEG507H250UC
CTP [α-^{32}P]	PerkinElmer	Cat#BLU008H250UC
SDS, 10% solution	Thermo Fisher Scientific	Cat#AM9822
EDTA	Thermo Fisher Scientific	Cat#15575020
Deposited data		
PIC1 structure	This paper	PDB: 7ML0
PIC2 structure	This paper	PDB: 7ML1
PIC3 structure	This paper	PDB: 7ML2
ITC structure	This paper	PDB: 7ML4
TFIIH weak binding state structure	This paper	PDB: 7ML3
Composite structure of EC+EC	This paper	PDB: 7MEI
Leading EC structure (focused refinement)	This paper	PDB: 7MKA
Trailing EC structure (focused refinement)	This paper	PDB: 7MK9
PIC1 cryo-EM composite map	This paper	EMD: 23904
PIC2 cryo-EM composite map	This paper	EMD: 23905
PIC3 cryo-EM composite map	This paper	EMD: 23906
ITC cryo-EM composite map	This paper	EMD: 23908
TFIIH weak binding state cryo-EM map (Focused refinement)	This paper	EMD: 23907
Composite cryo-EM Map of EC+EC	This paper	EMD: 23789
Leading EC cryo-EM Map	This paper	EMD: 23888
Trailing EC cryo-EM Map	This paper	EMD: 23887
Raw cryo-EM images	This paper	EMPIAR:10865
Experimental models: organisms/strains		
S. cerevisiae: C-TAP *TFB3* and *tfb6Δ* CB010 (*MATα pep4::HIS3 prb1::LEU2 prc1::HISG can1 ade2 trp1 ura3 his3 leu2–3,112 cir-o GAL+ RAF+ SUC+ tfb6::kanMX6 TFB3::TAP::Kl.TRP1*)	Murakami et al. 2012	N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
S. cerevisiae: C-TAP *TFG2* CB010 (*MATα pep4::HIS3 prb1::LEU2 prc1::HISG can1 ade2 trp1 ura3 his3 leu2–3,112 cir-o GAL+ RAF+ SUC+ TFG2::TAP::Kl.TRP1*)	Murakami et al. 2012	N/A
S. cerevisiae: C-TAP *TFA2* CB010 (*MATα pep4::HIS3 prb1::LEU2 prc1::HISG can1 ade2 trp1 ura3 his3 leu2–3,112 cir-o GAL+ RAF+ SUC+ TFA2::TAP::Kl.TRP1*)	Fazal et al. 2015	N/A
S. cerevisiae: C-TAP *RPB3* CB010 (*MATα pep4::HIS3 prb1::LEU2 prc1::HISG can1 ade2 trp1 ura3 his3 leu2–3,112 cir-o GAL+ RAF+ SUC+ RPB3::TAP::Kl.TRP1*)	Murakami et al. 2012	N/A
S. cerevisiae: C-TAP *TFB4* and *tfb6Δ CB010 (MATα pep4::HIS3 prb1::LEU2 prc1::HISG can1 ade2 trp1 ura3 his3 leu2–3,112 cir-o GAL+ RAF+ SUC+ tfb6::kanMX6 TFB4::TAP::Kl.TRP1*)	Murakami et al. 2012	N/A
S. cerevisiae: CKY2194 = *MAT***a** *ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX RPB3::TAP::KlacTRP1 imd2Δ::HIS3* [pCK1632 = pRSII316 *TFB3 URA3 CEN ARS amp*^{r}]	This paper	N/A
S. cerevisiae: CKY2196 = *MAT***a** *ura3-52 his3Δ200 leu2Δ0 or Δ1 trp1Δ63 met15Δ0 lys2-128∂ tfb3Δ::hphMX gal10Δ56 RPB3::TAP::KlacTRP1* [pCK1632 = pRSII316 *TFB3 URA3 CEN ARS amp*^{r}]	This paper	N/A
Recombinant DNA		
Plasmid: Full-length *S.c* TBP (pRSFDuet)	Murakami et al. 2012	N/A
Plasmid: Full-length *S.c* Toa1 (pRSFDuet)	Murakami et al. 2012	N/A
Plasmid: Full-length *S.c* Toa2 (pRSFDuet)	Murakami et al. 2012	N/A
Plasmid: Full-length *S.c* Toa1-Toa2 (pET47b)	Adachi et al. 2017	N/A
Plasmid: Full-length *S.c* TFIIB (pET28)	Bratkowski et al. 2018	N/A
Plasmid: Full-length *S.c* Sub1 (pCold II)	Fazal et al., 2015	N/A
Plasmid: SNR20 G-less 26 promoter fragment [−122/+97]	Fujiwara et al. 2019	N/A
Plasmid: SNR20 G-less 49 promoter fragment [−122/+97]	Fujiwara et al. 2019	N/A
Plasmid: *S.c TFB3* expression vector pCK1664 = pRSII315 *CEN ARS amp*^{r} *LEU2 TFB3*	This paper	N/A
Plasmid: *S.c tfb3* expression vector pCK2836 = pRSII315 *CEN ARS amp*^{r} *LEU2 tfb3* R64A	This paper	N/A
Plasmid: *S.c tfb3* expression vector pCK2826 = pRSII315 *CEN ARS amp*^{r} *LEU2 tfb3* R64D	This paper	N/A
Plasmid: *S.c tfb3* expression vector pCK2837 = pRSII315 *CEN ARS amp*^{r} LEU2 *tfb3* K65A	This paper	N/A
Plasmid: *S.c tfb3* expression vector pCK2827 = pRSII315 *CEN ARS amp*^{r} *LEU2 tfb3* K65R	This paper	N/A
Plasmid: *S.c tfb3* expression vector pCK2828 = pRSII315 *CEN ARS amp*^{r} *LEU2 tfb3* K65E	This paper	N/A
Plasmid: *S.c tfb3* expression vector pCK2838 = pRSII315 *CEN ARS amp*^{r} *LEU2 tfb3* R64A/K65A	This paper	N/A
Oligonucleotides		
Primer: SNR20 G-less 49 and G-less 26 Forward:5'-GCCGTTTCCGATGGG CCACTCGGTGAAAA-3'	This paper	N/A
Primer: SNR20 G-less 49 and G-less 26 Reverse:5'-GGTAATGAGCCTCAT TGAGGTCATTTCAGTTGTTACA-3'	This paper	N/A
Software and algorithms		

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Relion version 3.0/3.1	Zivanov et al. 2018	https://www3.mrc-lmb.cam.ac.uk/relion//index.php/Main_Page/; RRID:SCR_016274
cryoSPARC version 3.1	Punjani et al. 2017	https://cryosparc.com/; RRID:SCR_016501
UCSF Chimera	Pettersen et al. 2004	https://www.cgl.ucsf.edu/chimera/; RRID:SCR_004097
Phenix version 1.18.2	Adams et al. 2010	https://www.phenix-online.org/; RRID:SCR_014224
Coot version 0.8.9.2	Emsley et al., 2010	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/; RRID:SCR_014222
Topaz	Bepler et al., 2019	https://cb.csail.mit.edu/cb/topaz/