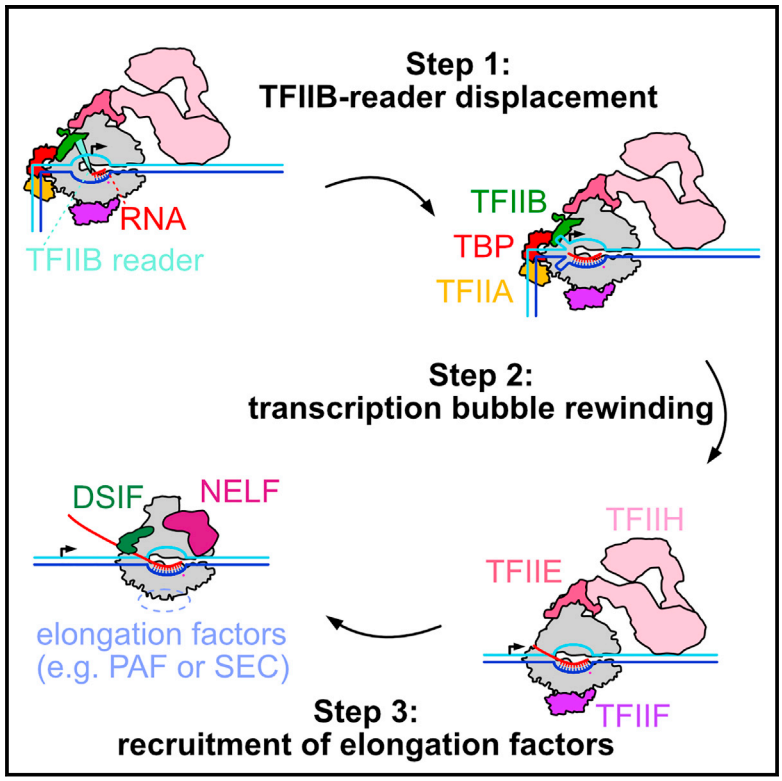
Molecular Cell

Article

Three-step mechanism of promoter escape by RNA polymerase II

Graphical abstract



Highlights

- Visualization of initiation-elongation transition intermediates *de novo* initiated from the PIC
- Initial RNA synthesis repositions TFIIB reader but does not displace TFIIB
- Transcription bubble rewinding coincides with the displacement of TFIIA, TBP, and TFIIB
- Recruitment of DSIF and NELF facilitates the displacement of TFIIE and TFIIH

Authors

Yumeng Zhan, Frauke Grabbe, Elisa Oberbeckmann, Christian Dienemann, Patrick Cramer

Correspondence

christian.dienemann@mpinat.mpg.de (C.D.), patrick.cramer@mpinat.mpg.de (P.C.)

In brief

Zhan et al. reconstitute the initiation-elongation transition of RNA polymerase II (RNA Pol II) and provide high-resolution structures of transcription intermediates during promoter escape. They conclude on the molecular mechanisms of promoter escape by mammalian Pol II.

Zhan et al., 2024, Molecular Cell *84*, 1–12
May 2, 2024
https://doi.org/10.1016/j.molcel.2024.03.016

CellPress

Molecular Cell

CellPress
OPEN ACCESS

Article

Three-step mechanism of promoter escape by RNA polymerase II

Yumeng Zhan,[1] Frauke Grabbe,[1] Elisa Oberbeckmann,[1] Christian Dienemann,[1,2,*] and Patrick Cramer[1,*]
[1]Max Planck Institute for Multidisciplinary Sciences, Department of Molecular Biology, Am Fassberg 11, 37077 Göttingen, Germany
[2]Lead contact
*Correspondence: christian.dienemann@mpinat.mpg.de (C.D.), patrick.cramer@mpinat.mpg.de (P.C.)
https://doi.org/10.1016/j.molcel.2024.03.016

SUMMARY

The transition from transcription initiation to elongation is highly regulated in human cells but remains incompletely understood at the structural level. In particular, it is unclear how interactions between RNA polymerase II (RNA Pol II) and initiation factors are broken to enable promoter escape. Here, we reconstitute RNA Pol II promoter escape *in vitro* and determine high-resolution structures of initially transcribing complexes containing 8-, 10-, and 12-nt ordered RNAs and two elongation complexes containing 14-nt RNAs. We suggest that promoter escape occurs in three major steps. First, the growing RNA displaces the B-reader element of the initiation factor TFIIB without evicting TFIIB. Second, the rewinding of the transcription bubble coincides with the eviction of TFIIA, TFIIB, and TBP. Third, the binding of DSIF and NELF facilitates TFIIE and TFIIH dissociation, establishing the paused elongation complex. This three-step model for promoter escape fills a gap in our understanding of the initiation-elongation transition of RNA Pol II transcription.

INTRODUCTION

Transcription by RNA polymerase II (RNA Pol II) starts with the formation of the pre-initiation complex (PIC), which consists of RNA Pol II and a set of general transcription factors (GTFs): TFIIA, TFIIB, TFIID, TFIIE, TFIIH, and Mediator.[1–4] Upon PIC positioning on the promoter, the translocase activity of TFIIH pumps downstream DNA into the RNA Pol II active center cleft and applies torque to induce melting of the DNA duplex.[5,6] As RNA synthesis commences, the PIC transitions to an initially transcribing complex (ITC), which still contains GTFs. When the RNA reaches a certain length, the ITC is converted to an elongation complex (EC), and the transcribing complex escapes from the promoter region.[7]

During promoter escape, the transcription complex undergoes structural and compositional changes.[7,8] As initial transcription proceeds, the upstream end of the initially melted DNA bubble remains in place, whereas the downstream junction of the bubble is further unwound, enlarging the DNA bubble[9,10] in a process termed DNA scrunching.[11–13] During early transcription, the ITC tends to release the nascent RNA in a process termed abortive transcription.[10,14] When the transcript reaches a length of 7 nucleotides (nt) or longer, the extended transcription bubble rewinds from the upstream end in a process termed DNA unscrunching.[9–13] This leads to the formation of a stable transcription bubble that is consistent with an EC.[7,9] After bubble rewinding, the transcription complex reaches full stability, marked by the end of abortive transcription[10] and independence of TFIIH translocase activity for efficient elongation.[9]

In addition to the transitions in the DNA and RNA structure during initial transcription, the GTFs are released from RNA Pol II.[7,8] Structural studies of RNA Pol II-TFIIB complexes showed that the TFIIB-reader domain projects into the RNA Pol II active center cleft[15] and would clash with an RNA longer than 7 nt.[16] Structural comparison between RNA Pol II-TFIIB complexes and ECs also showed that the TFIIB-linker domain clashes with the upstream DNA after bubble rewinding.[15] These findings led to the early hypothesis that extension of the nascent RNA beyond 6 nt triggers release of TFIIB and promoter escape.[15] However, biochemical studies later showed that TFIIB can stay associated with RNA Pol II until the RNA reaches 12–13 nt.[17] To accommodate TFIIB in the early transcribing complex with a longer RNA, TFIIB needs to rearrange within the transcription complex. However, structural information on such putative conformational changes is lacking.

The fate of TFIIF is unclear as different mechanisms for TFIIF displacement were proposed. It has been reported that TFIIF is not displaced during the initiation-elongation transition but rather travels with the early EC[18,19] and stimulates elongation.[20,21] Suggested mechanisms for displacement of TFIIF involve casein kinase (CK2)-dependent phosphorylation[22] or the competition with multi-subunit elongation factor Paf1 complex (PAF)[23,24] and super elongation complex (SEC),[25] which bind to RNA Pol II at sites that overlap with TFIIF binding. However, the exact molecular mechanism by which TFIIF is removed from RNA Pol II remains elusive.

The mechanism of TFIIE and TFIIH release from RNA Pol II is also unclear. For TFIIE, it was suggested that dissociation from

Please cite this article in press as: Zhan et al., Three-step mechanism of promoter escape by RNA polymerase II, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.016

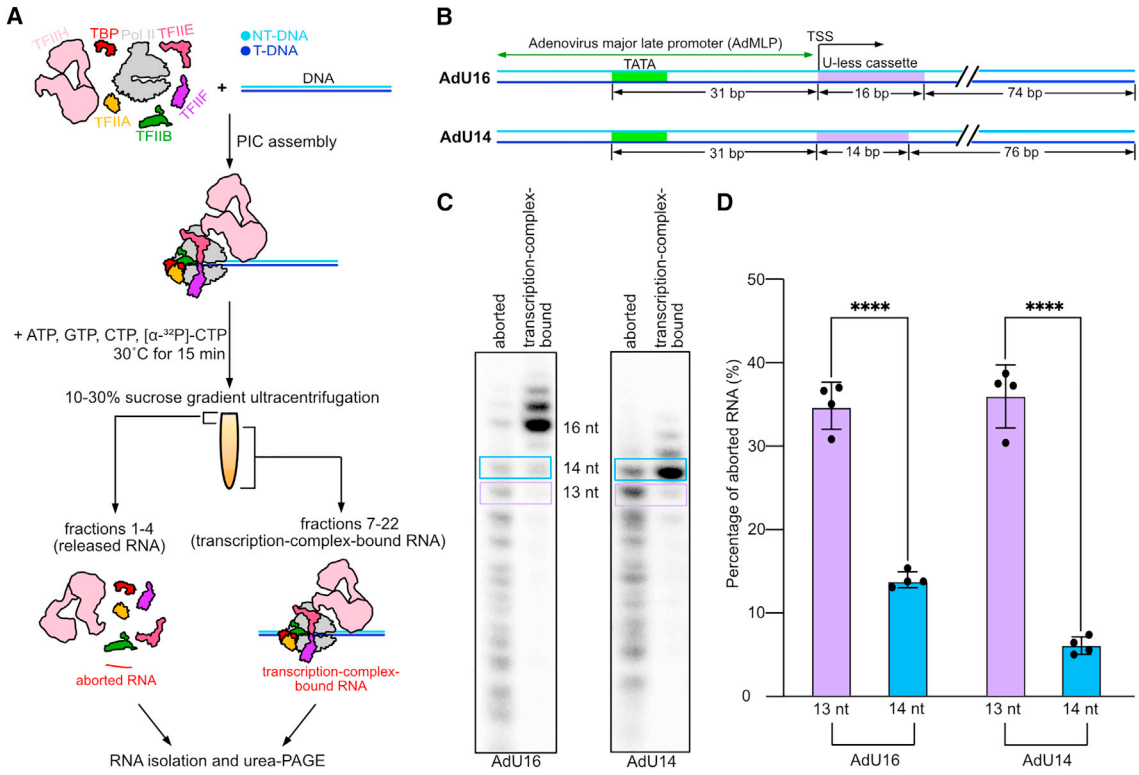


Figure 1. RNA Pol II enters stable RNA chain elongation when nascent RNA is 14 nt long

(A) Scheme of the experimental setup to measure aborted and transcription-complex-bound RNA. NT-DNA, non-template DNA; T-DNA, template DNA.

(B) Scheme of the promoter DNA templates used in the experiments. TATA-box and U-less cassettes are highlighted in green and purple, respectively. The lengths of the U-less cassette, the total length of the DNA template, and the distance between the transcription start site (TSS) and TATA box are indicated.

(C) Urea-PAGE analysis of aborted and transcription-complex-bound RNA products from transcription reactions with AdU16 and AdU14 promoter DNAs. Experiments were repeated four times. Rectangular boxes denote bands of interest that were quantified.

(D) Quantification of the fraction of aborted RNAs with lengths of 13 nt and 14 nt from transcription reactions using AdU16 and AdU14 promoter DNAs. Mean ± standard deviation (SD) values are indicated. *p* values were calculated with unpaired two-tailed t-tests (****$p < 0.0001$). The color code is the same as (C).

RNA Pol II occurs before the RNA transcript reaches 10 nt, whereas TFIIH would leave the early EC after the transcript reaches 30 nt.[18] Another study proposed that TFIIE and TFIIH subunits dissociate in separate steps: TFIIE-α and the TFIIH kinase module dissociate first, whereas TFIIE-β and the core of TFIIH dissociate later.[26]

To study the mechanisms of promoter escape, an efficient *in vitro* system for *de novo*, promoter-dependent transcription is required that allows for reconstitution of the initiation-elongation transition in the test tube and subsequent structural characterization of the transition intermediates. Two recent studies reported such *in vitro* transcription initiation systems for the yeast *Saccharomyces cerevisiae*.[27,28] However, the initiation-elongation transition differs between the yeast and human systems. In particular, yeast RNA Pol II undergoes initial promoter scanning, a process that is absent in human transcription,[29–32] and yeast RNA Pol II does not pause in the promoter-proximal region, a process that occurs in the human system and depends on the negative elongation factor (NELF) that is absent from yeast.[33]

Here, we describe an experimental setup for efficient mammalian *in vitro* promoter-dependent transcription and perform structural investigations of the reaction mixture by cryo-electron microscopy (cryo-EM). We capture five intermediates during promoter escape by mammalian RNA Pol II and determine their structures. We visualize key steps during promoter escape, where the ITC undergoes stepwise structural and compositional changes upon transitioning to an EC. Our results lead to a simple, three-step model for promoter escape by RNA Pol II.

RESULTS

Reconstitution of the RNA Pol II initiation-elongation transition

To reconstitute the initiation-elongation transition of transcription *in vitro*, we assembled the RNA Pol II PIC on promoter DNA templates using the human GTFs TBP, TFIIA, -IIB, -IIF, -IIE, and -IIH and *S. scrofa* RNA Pol II, which is 99.9% identical to human RNA Pol II (Figure 1A). As a DNA template, we used a variant of the adenovirus major late promoter (AdMLP) containing a U-less cassette, which allowed us to stall RNA Pol II after synthesis of a 16-nt RNA (Figures 1B, top, and S1A, top). Transcription was initiated by the addition of ATP, cytidine triphosphate (CTP), guanosine triphosphate (GTP), and [α-^{32}P]-CTP, and reactions were incubated at 30°C for 15 min.

Please cite this article in press as: Zhan et al., Three-step mechanism of promoter escape by RNA polymerase II, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.016

To monitor abortive transcription, aborted RNA and transcription-complex-bound RNA were separated by sucrose gradient centrifugation (Figure 1A). The upper factions of the gradient, which are devoid of RNA Pol II (Figure S1B), contained excess GTFs and aborted RNA transcripts that were released from RNA Pol II (fractions 1–4). The lower fractions (7–22) contained RNA bound to the transcription complex (Figures S1B and S1C). The fractions were pooled, and RNAs were resolved by denaturing polyacrylamide-gel electrophoresis (PAGE) (Figure 1C, left). We observed released, short aborted RNAs and the full-length 16-nt transcript that was stably associated with the transcription complex (Figure 1C, left). In summary, we established an efficient, promoter-dependent *in vitro* transcription system yielding ~0.26 transcripts per molecule RNA Pol II, which allows us to study the transition from initiation to elongation.

When we compared aborted and transcription-complex-bound RNAs produced by *in vitro* transcription from the AdU16 promoter, we observed a strong reduction of abortive transcription at around 14 nt of RNA (Figure 1C, left). We therefore designed the AdU14 promoter that stalls RNA Pol II at 14 nt of RNA (Figure 1B, bottom, and S1A, bottom) and resolved aborted and transcription-complex-bound transcripts by denaturing PAGE (Figure 1C, right). Quantification of aborted and transcription-complex-bound RNAs from both the AdU16 and AdU14 promoters showed a strong decrease in the fraction of aborted transcripts when the RNA length grew from 13 to 14 nt (Figures 1C and 1D). This indicates that RNA extension from 13 to 14 nt leads to increased stability of the early transcribing complex. These observations are consistent with previous reports showing that the transcribing complex is unstable and requires TFIIH ATPase activity until the transcript length increases beyond 14 nt.[34,35] Thus, the transition from abortive transcription to stable RNA chain elongation is completed when the RNA reaches 14 nt of length.

Cryo-EM structures of initiation-elongation transition intermediates

We next aimed to structurally characterize early transcription complexes during the transition from abortive transcription to stable elongation by cryo-EM. Guided by our biochemical results, we used the AdU14 promoter that stalls RNA Pol II when the nascent RNA is 14 nt long. We then prepared cryo-EM samples from the transcription reaction (STAR Methods; Figure 2A) and collected single-particle cryo-EM data. After extensive data processing, we could resolve five structures: three ITC and two EC structures (Figures 2B, S2, and S3; Tables 1 and 2). Due to the nature of the single-particle analysis, it has to be noted that these structures likely represent the most populated and stable intermediates in the reaction mixture, whereas other rare or unstable states may not have been resolved. Altogether, our experimental setup allowed us to capture and structurally characterize mammalian transcription intermediates obtained by *de novo* transcription that, in contrast to reconstituted complexes, more likely represent true functional intermediates of the initiation-elongation transition.

The three ITC structures contain all GTFs and an extended DNA bubble similar to a previously reported ITC structure.[37] The DNA-RNA hybrids in all three ITCs are in the pre-translocated state; however, the number of visible RNA nt in the three ITC structures is different (Figure 2B). Because the active site region of the three ITCs is resolved to ~3 Å, we could conclude that the three ITCs contain 8, 10, and 12 nt of ordered RNA, respectively, and we therefore called these structures ITC8, ITC10, and ITC12. RNA sequences were built based on the length of the ordered RNA because the map quality did not allow us to unambiguously identify the exact sequence register at which the ITC structures have stopped transcribing. With their pre-translocated ordered RNAs being only 8 and 10 nt long, ITC8 and ITC10 contain incomplete DNA-RNA hybrids that only reach to register −7 and −9, respectively (Figures 2B and 2C). By contrast, the 12-nt ordered RNA in ITC12 forms a hybrid until register −10, which represents a fully formed DNA-RNA hybrid before strand separation at register −11 (Figures 2B and 2C).

The two EC structures both contain 14-nt RNAs but showed differences in the occupancy and conformation of TFIIE (Figures 2B and S2). These structures were called EC14a and EC14b. Both of these structures contain a DNA-RNA hybrid of 10 base pairs (bp) in the post-translocated state and a rewound transcription bubble with an upstream DNA trajectory that is similar to the paused RNA Pol II EC.[38]

Growing RNA repositions the TFIIB reader but does not displace TFIIB

It was predicted that the 5′-end of the initially transcribed RNA interferes with the reader domain of TFIIB when RNA extension proceeds beyond 5–6 nt.[16] We therefore compared the conformations of the B-reader among our three ITC structures as they contain ordered RNAs of different lengths (Figure 2C). In ITC8, residues 54–55 and 61–64 of the B-reader are further disordered compared with that of the open complex (OC)[36] (PDB: 7nw0), indicating that the DNA-RNA hybrid reaching to register –7 leads to displacement of this part of the B-reader. When the hybrid reaches register –9 in ITC10, B-reader residues 52–53 and 65 appear to be further disordered compared with that of ITC8. Upon the formation of a full 10-bp DNA-RNA hybrid in ITC12, we could not observe density for residues 48–79 of the TFIIB-reader domain. Since the TFIIB reader can stably form in an OC without RNA (PDB: 7nw0, mammalian OC; PDB: 4bbr, yeast RNA Pol II-TFIIB)[16,36] (Figure 2C, left), we conclude that the presence of the RNA in the active center cleft causes the observed displacement of the TFIIBreader. Thus, we show that the RNA strand of the DNA-RNA hybrid and the B-reader directly compete for the same space in the active center cleft of RNA Pol II and that the growing RNA can displace the B-reader during early transcription via clashes between the 5′-end of the RNA and the B-reader.

Consistent with this, it was shown that early transcription complexes are prone to stall when the RNA reaches a length of 7–9 nt[9,39] and that this stalling can be eliminated by the TFIIB point mutation R66L, which destabilizes a conserved salt bridge within the B-reader.[9] Modeling based on our structures suggests that formation of a full 10-bp DNA-RNA hybrid requires disruption of the stabilizing salt bridge formed between E51 and R66 (Figure 2C). This provides a possible mechanism for how the B-reader and RNA compete—salt bridge formation stabilizes the B-reader, whereas the growing RNA destabilizes the salt bridge and thereby also the B-reader.

Please cite this article in press as: Zhan et al., Three-step mechanism of promoter escape by RNA polymerase II, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.016

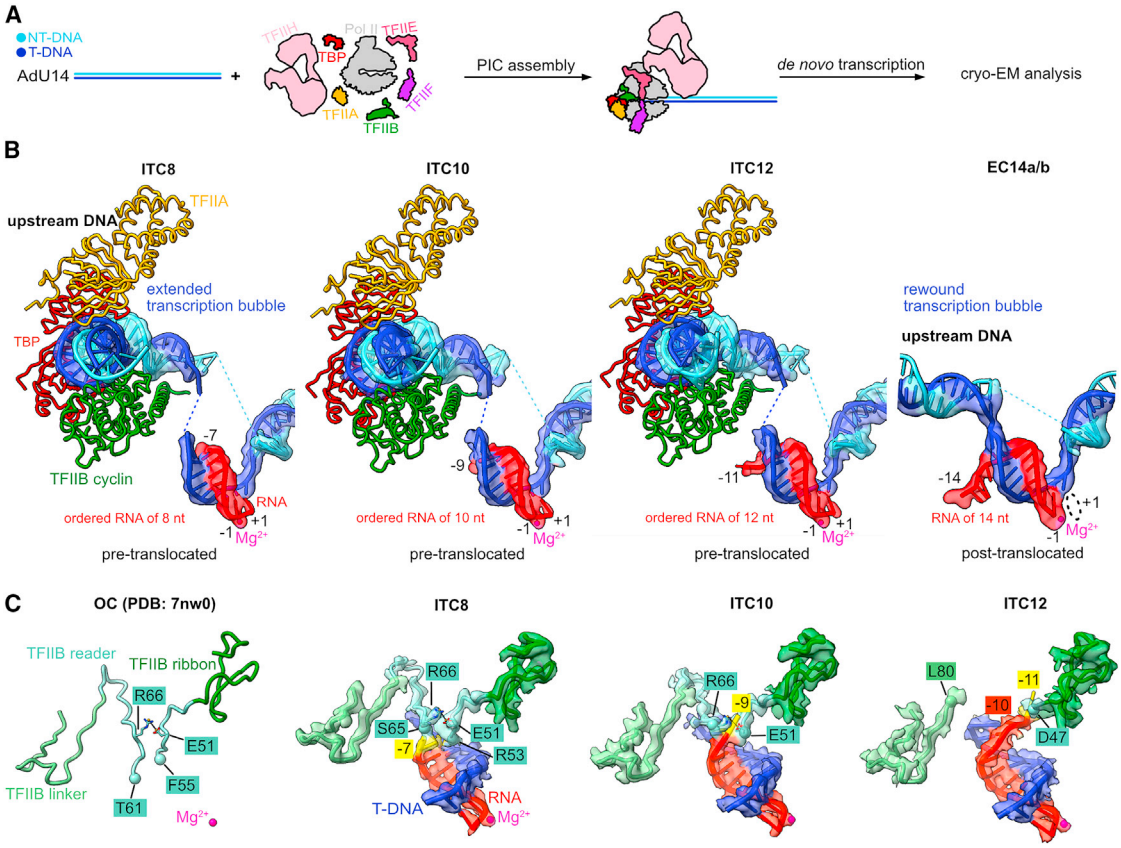


Figure 2. Visualization of early transcription intermediates by cryo-EM

(A) Scheme of experimental setup for cryo-EM sample preparation. For details, see STAR Methods. NT-DNA, non-template DNA; T-DNA, template DNA.

(B) Cryo-EM density and structures of the five transcription intermediates ITC8, ITC10, ITC12, and EC14a/b. The cryo-EM densities around the upstream DNA were low-pass filtered to 5 Å for better visualization. The register of the RNA is indicated with respect to the active site. The active site Mg^{2+} is shown as a magenta sphere. For clarity, RNA Pol II and the rest of the GTFs are not shown except for TFIIA, TBP, and the TFIIB-cyclin.

(C) Cryo-EM density and structures of TFIIB reader, linker, and ribbon and the growing RNA transcript in ITC8, ITC10, and ITC12. For comparison, the structure of TFIIB reader, linker, and ribbon from OC (PDB: 7nw0.[36]) is shown. TFIIB and the DNA-RNA hybrid are shown as cartoons. The salt bridge formed by E51 and R66 is indicated by black dashes. The boundaries of the structured parts of TFIIB are shown as spheres. The first visible base at the 5′-end of the RNA is highlighted in lemon, and its position relative to the active site (−1) is indicated. The active site Mg^{2+} is shown as a magenta sphere. The template DNA at the RNA Pol II active site is shown in blue. For clarity, the rest of the DNA, TFIIB-cyclin, RNA Pol II, and the other GTFs are omitted.

Eviction of TBP, TFIIA, and TFIIB coincides with bubble rewinding

As RNA synthesis proceeds, the extended, early DNA bubble rewinds from upstream, and this rewinding coincides with the transition from abortive to stable RNA synthesis.[9,11–13] Consistent with this, we observe a strong difference in upstream DNA when we compare our ITC and EC structures. Whereas the ITC structures contain extended upstream bubbles, our EC structures show that upstream DNA is rewound and positioned where it is observed in a reconstituted paused EC.[38]

To investigate the structural changes within the transcription complex upon bubble rewinding, we compared the structures of ITC12 and EC14a (Figure 3A). In EC14a, we do not observe density for the upstream promoter complex comprising TBP, TFIIA, and TFIIB (Figure 3A, right). Therefore, displacement of the upstream promoter complex from RNA Pol II coincides with the rewinding of the transcription bubble. Superposition of the ITC with the EC structures shows that the trajectory of the rewound upstream DNA duplex clashes with the TFIIB-cyclin and -linker domains (Figure 3B). TBP and TFIIA are mainly binding to the upstream DNA and the TFIIB-cyclin domain within the ITC (Figure 3B). Therefore, our structural data suggest that rewinding of the overextended DNA bubble within the ITC may lead to eviction of the upstream promoter complex from RNA Pol II. However, it cannot be excluded that the upstream promoter complex may dissociate before bubble rewinding.

Although the TFIIB-cyclin readily dissociates at the time of bubble rewinding, release of the B-ribbon in EC14a is caused by the 14-nt-long transcript, which starts clashing with the B-ribbon at a length of 12 nt (Figure 3C). This is consistent with a previous study showing that TFIIB is only destabilized when the RNA is 12–13 nt long.[17] In summary, eviction of TBP, TFIIA, and TFIIB coincides with DNA bubble rewinding and RNA extension beyond 12 nt, which converts the ITC to an early EC.

Please cite this article in press as: Zhan et al., Three-step mechanism of promoter escape by RNA polymerase II, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.016

Table 1. Cryo-EM data acquisition and processing

	Dataset 1	Dataset 2	Dataset 3
Magnification	81,000×	81,000×	81,000×
Voltage (kV)	300	300	300
Electron exposure (e^{-}/Å^2)	40.06	40.00	41.47
Defocus range (μm)	0.7–1.7	0.7–1.7	0.7–1.7
Pixel size (Å)	1.05	1.05	1.05
Initial particle images (no.)	6,386,813	4,204,757	2,593,001

TFIIF, TFIIE, and TFIIH remain bound to the early EC

It was proposed that bubble rewinding during the conversion from the ITC to the early EC also triggers the release of TFIIE and TFIIH.[7,33] In contrast to the proposed model, we observe density for TFIIF, TFIIE, and TFIIH in our cryo-EM reconstructions of both ECs (Figures 4A, S2, and S3). Although the TFIIF dimerization domain is bound to RNA Pol II in both ECs, we could not observe density for the TFIIF-β winged-helix (WH) domain (Figures 3A and 4A). The TFIIF-β WH domain serves as a binding site for TFIIE-β in the PIC and ITC, and dissociation of the WH domain could influence the conformation and occupancy of TFIIE within the early EC.

Indeed, when we classified the particle populations of the ECs after masking the region around TFIIE, we observed two distinct conformations of TFIIE (Figures 4A, S2, and S3). In EC14a, the density for both subunits of TFIIE is visible (Figure 4A, top). However, in EC14b, clear density is only present for the TFIIE-α subunit, which additionally has rotated around the RNA Pol II stalk (Figure 4A, middle, and bottom). TFIIE-α serves as the main interaction partner for TFIIH within the ITC[37,36,41,42] and TFIIE is crucial for the binding of TFIIH to the PIC.[42] In line with this, we find that TFIIE enhances the binding of TFIIH to the early EC (Figures S4A and S4B) and observe density for TFIIH in the cryo-EM reconstructions of both EC14a and EC14b (Figures 4A, S2, and S3). In summary, TFIIF, TFIIE, and TFIIH remain bound to RNA Pol II in the early EC, demonstrating that neither RNA extension nor upstream DNA rewinding can trigger their release from the early EC.

DSIF and NELF facilitate the displacement of TFIIE and TFIIH

From these results, it remained unclear how TFIIF, TFIIE, and TFIIH are removed from the early EC. We therefore asked whether the general elongation factor DSIF, the capping

Table 2. Refinement and validation statistics

	ITC8 model	ITC10 model	ITC12 model	EC14a model	EC14b model
PDB code	PDB: 8S51	PDB: 8S52	PDB: 8S5N	PDB: 8S55	PDB: 8S54
Map code	EMDB: EMD-19718	EMDB: EMD-19719	EMDB: EMD-19743	EMDB: EMD-19726	EMDB: EMD-19720
Symmetry imposed	C1	C1	C1	C1	C1
Map resolution (Å) at FSC = 0.143	3.1	2.9	3.4	3.4	3.3
Map resolution range (Å)	3.0–11.4	2.6–7.9	2.8–8.0	3.1–10.4	3.0–10.6
Refinement					
Initial models used (PDB code)	7NW0, 5IYD, 5FLM	7NW0, 5IYD, 5FLM	7NW0, 5IYD, 5FLM	7NW0, 5FLM	7NW0, 5FLM
Model composition					
Non-hydrogen atoms	44,796	44,876	44,728	38,140	36,737
Protein residues	5,272	5,277	5,249	4,514	4,343
Nucleotides	124	127	129	95	95
Ligands	Zn: 10, Mg: 1	Zn: 10, Mg: 1	Zn: 10, Mg: 1	Zn: 9, Mg: 1	Zn: 9, Mg: 1
Mean B factors (Å^2)					
Protein	113.48	117.32	125.33	151.00	144.58
Nucleotides	165.15	201.77	198.91	201.38	199.84
Ligand	132.44	139.67	172.74	204.85	195.06
RMS deviations					
Bond lengths (Å)	0.004	0.004	0.004	0.005	0.005
Bond angles (°)	0.721	0.715	0.698	1.176	1.168
Validation					
MolProbity score	1.50	1.52	1.67	1.88	1.87
Clashscore	4.89	5.14	5.17	9.43	8.84
Poor rotamers (%)	0.00	0.00	0.00	0.00	0.00
Ramachandran plot					
Disallowed (%)	0.00	0.00	0.00	0.00	0.00
Allowed (%)	3.71	3.66	5.79	5.63	5.90
Favored (%)	96.29	96.34	94.21	94.37	94.10

Please cite this article in press as: Zhan et al., Three-step mechanism of promoter escape by RNA polymerase II, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.016

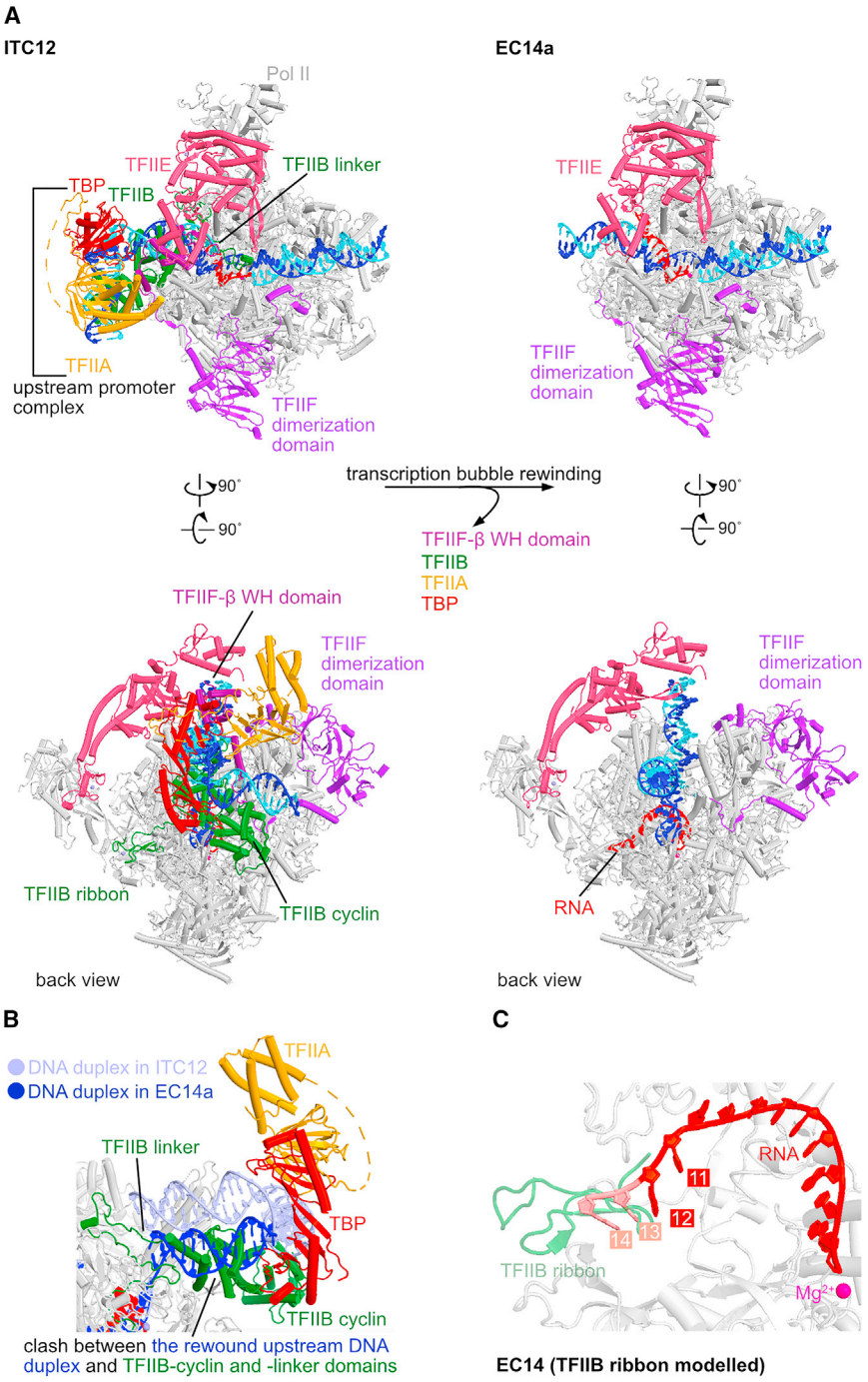


Figure 3. Eviction of TBP, TFIIA, and TFIIB coincides with transcription bubble rewinding

(A) Structures of ITC12 (left) and EC14a (right) are shown as cartoons for comparison.

(B) Superposition of ITC12 and EC14a reveals clashes between the rewound upstream DNA duplex with TFIIB-linker and -cyclin domains. For clarity, TFIIE, TFIIF, and TFIIH are omitted.

(C) Eviction of TFIIB ribbon by RNA. The TFIIB ribbon clashes with the RNA at the length of 13–14 nt and is displaced in EC14. The first two RNA bases that would clash with the TFIIB ribbon in EC14 are colored in salmon.

enzyme RNGTT, or the pausing factor NELF, which binds when RNA reaches lengths of 24–50 nt,[33,43] may be responsible for the eviction of the remaining GTFs.

To test for eviction of TFIIE from the early EC, we first assembled an EC24 complex with TFIIF, TFIIE, and a 24-nt RNA (STAR Methods; Figures S4C and S4D, left). Although TFIIF is likely displaced by elongation factors,[23–25,40] we still included TFIIF in the experiments as it is present in the early EC (Figures 3A and 4A) and may interact with TFIIE and stabilize it. We first added DSIF together with the capping enzyme RNGTT but could not observe the eviction of TFIIF or TFIIE (Figure S4D, middle). However, when adding DSIF in combination with NELF, we observed displacement of TFIIE from the EC24, whereas TFIIF remains associated (Figure S4D, right). However, DSIF and NELF fail to displace TFIIE from an EC14 with 14 nt of RNA (Figure S4E, right), which is in line with previous studies showing that the binding of DSIF and NELF requires an RNA of at least 22 nt length.[44]

TFIIH binding to RNA Pol II is mainly dependent on the presence of TFIIE[42] (Figures S4A and S4B), and displacement of TFIIE would therefore likely result in dissociation of TFIIH as well. To test whether DSIF and NELF can dissociate both TFIIE and TFIIH from the early EC, we assembled an EC24 containing TFIIF, TFIIE, and core TFIIH and monitored the binding of TFIIE-α and the TFIIH translocase XPB with RNA Pol II in gradient ultracentrifugation (Figures 4B and S4F). In the absence of DSIF and NELF, a large proportion of TFIIE-α and XPB are associated and therefore co-migrated with RNA Pol II (Figures S4F, left, and S4G, top). However, in the presence of DSIF and NELF, the fractions of RNA Pol II-bound TFIIE-α and XPB were reduced (Figure 4B) as TFIIE-α and XPB shifted from RNA Pol II-bound fractions to lower molecular weight fractions that are devoid of RNA Pol II (Figures S4F, right, and S4G, bottom). We speculate that the decrease of TFIIH in Figure 4B is not more drastic, as TFIIH might be dissociated from RNA Pol II but still bound to the DNA. Unfortunately, we cannot distinguish this in our experimental setup. Nonetheless, our results show that recruitment of DSIF and NELF facilitates dissociation of both TFIIE and TFIIH from the early EC and that this relies on RNA extension beyond 22 nt.

Please cite this article in press as: Zhan et al., Three-step mechanism of promoter escape by RNA polymerase II, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.016

CellPress
OPEN ACCESS

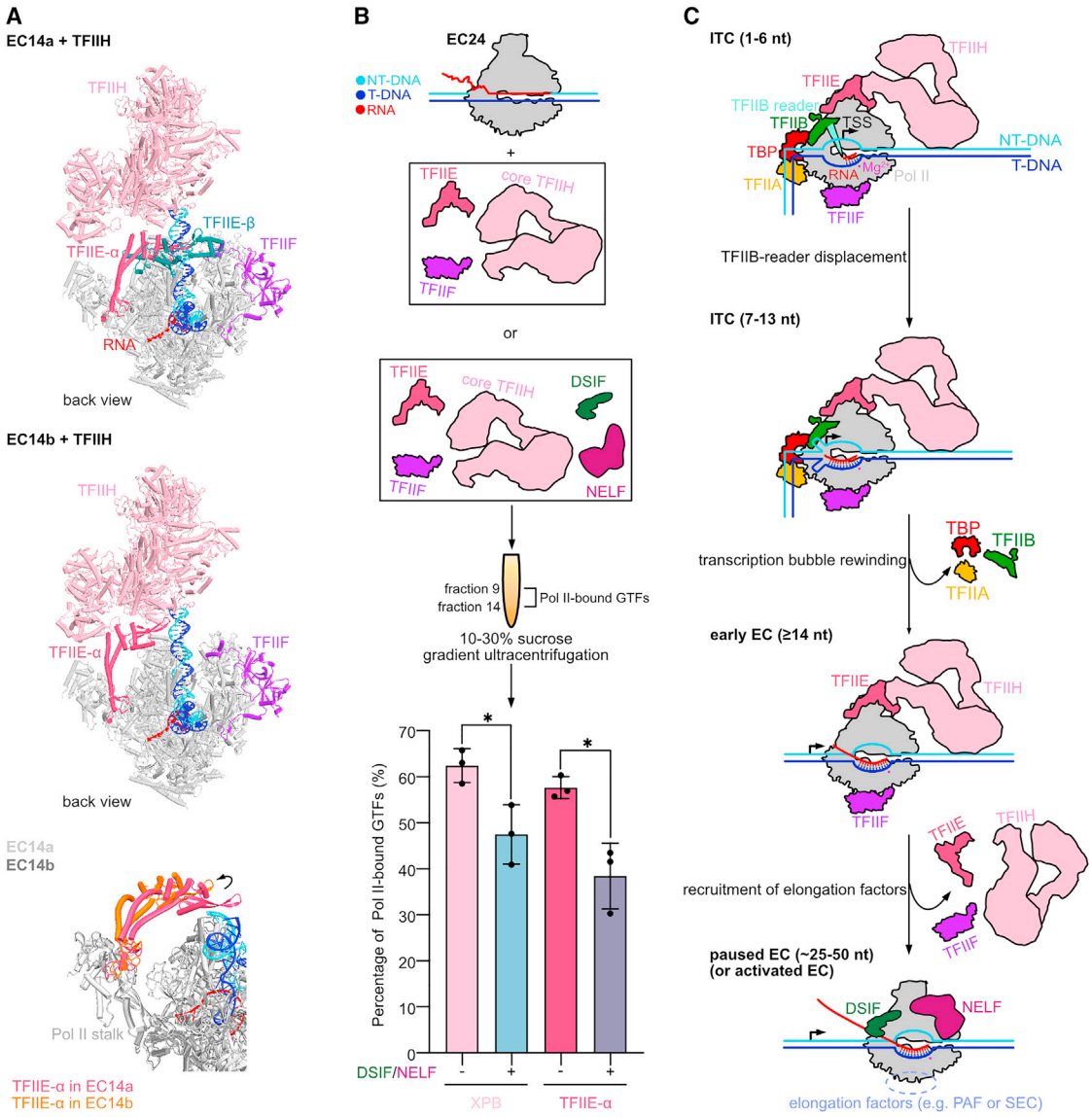


Figure 4. TFIIE, TFIIF, and TFIIH remain bound to the early elongation complex

(A) Structures of EC14a (top) and EC14b (middle) containing TFIIH, TFIIE, and TFIIF. The structure of TFIIH (PDB: 7nvx.[36]) was rigid-body docked into the global refined maps of EC14a+TFIIH and EC14b+TFIIH (STAR Methods). Bottom: zoom-in on TFIIE-α and RNA Pol II of EC14a and b. RNA Pol II of EC14a is colored in light gray. RNA Pol II of EC14b is colored in dark gray. TFIIE-α of EC14a is colored in pink. TFIIE-α of EC14b is colored in orange. The structures of EC14a and EC14b are superimposed using Rpb1. EC14a/b are captured from transcription reaction with AdU14 template after GraFix.

(B) Scheme of the experimental setup to monitor the displacement of TFIIE and TFIIH by DSIF and NELF and quantification of the percentage of RNA Pol II-bound GTFs after gradient ultracentrifugation. Mean ± SD values are indicated. *p* values were calculated with unpaired two-tailed t-tests (*$p < 0.05$). *p* values for XPB without DSIF/NELF-XPB with DSIF/NELF and TFIIE-α without DSIF/NELF-TFIIE-α with DSIF/NELF are 0.0252 and 0.0115, respectively. The experiments were repeated three times and performed in the absence of crosslinker. NT-DNA, non-template DNA; T-DNA, template DNA.

(C) Three-step model of promoter escape by mammalian RNA Pol II. First, the RNA transcript competes with and eventually displaces the TFIIB reader from the RNA Pol II active center when the RNA is around 7–13 nt long. Second, rewinding of the upstream transcription bubble coincides with the eviction of the upstream promoter complex (TBP, TFIIA, and TFIIB) from the RNA Pol II surface. Finally, when the RNA is around 25–50 nt,[33] DSIF and NELF are recruited to the early EC and facilitate the displacement of TFIIE and TFIIH. Association of elongation factors such as PAF or SEC likely displaces TFIIF.[23–25,40]

DISCUSSION

Here, we establish an *in vitro* promoter-dependent mammalian transcription system that enables us to carry out structural studies of early RNA Pol II transcription complex intermediates and investigate the initiation-elongation transition of transcription. Our results converge with published data[7,9,10,14–17,34,35,44] on a three-step model of promoter escape by RNA Pol II (Figure 4C; Video S1). First, when RNA Pol II synthesizes an RNA longer than 6 nt, the transcript begins to compete with the

Please cite this article in press as: Zhan et al., Three-step mechanism of promoter escape by RNA polymerase II, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.016

TFIIB-reader domain for the same binding site in the RNA Pol II active center, which may cause early transcription abort. As the RNA extends, however, it can displace the B-reader without evicting TFIIB. Second, when the RNA reaches a length of ~13–14 nt, the upstream DNA bubble collapses. Rewinding of upstream DNA coincides with the eviction of the upstream promoter complex comprising TBP, TFIIA, and TFIIB, likely by steric clashes between the upstream promoter complex and the rewound DNA. Transcription bubble rewinding also coincides with the repositioning of parts of TFIIE and TFIIF. Third, when the RNA extends to at least 22 nt, which is the minimum RNA length required for the recruitment of the elongation factors DSIF and NELF,[44] binding of these factors to the RNA Pol II surface facilitates displacement of TFIIE and TFIIH. Finally, recruitment of other elongation factors that bind the RNA Pol II lobe region can displace TFIIF.[23–25,40] Displacement of all GTFs from the transcribing RNA Pol II marks the end of promoter escape and the transition to the EC.

Our three-step model generally agrees with previous models but also differs in one important way. Namely, it was previously proposed that the removal of all GTFs from the ITC is triggered by rewinding of the initial extended transcription bubble.[7,33] However, the comparison of our ITC and EC structures, which were formed by *de novo* transcription, strongly suggests that bubble rewinding only coincides with the dissociation of the upstream promoter complex comprising TBP, TFIIA, and TFIIB, whereas the three other GTFs (TFIIE, TFIIH, and TFIIF) remain bound to RNA Pol II at this stage. We speculate that TFIIE and TFIIH may remain associated with the early EC to enable complete Ser5 phosphorylation on the RNA Pol II C-terminal repeat domain (CTD) by the TFIIH kinase CDK7, possibly until high levels of CTD phosphorylation release Mediator,[41,45] which allows recruitment of DSIF and NELF. The recruitment of DSIF and NELF will then facilitate the dissociation of TFIIE and TFIIH. Additionally, as RNA Pol II transcribes toward the +1 nucleosome, which is positioned at ~40–60 bp downstream of the transcription start site (TSS) on actively transcribed mammalian genes,[46–49] the nucleosome may clash with TFIIE and TFIIH and displace them from the RNA Pol II surface.[50]

Our observation that TFIIF is part of the early EC is consistent with reports that TFIIF can remain bound to the early EC after the initiation-elongation transition[18,19] where it likely stimulates early transcription elongation.[20,21] However, TFIIF is largely absent from gene bodies[51–53] and therefore must be displaced soon after the initiation-elongation transition. It was suggested that the activated EC (EC*) is incompatible with TFIIF binding because a region of the PAF complex subunit Leo1 occupies a similar site on the RNA Pol II protrusion as a part of TFIIF[23,24] (Figure S5A). Together with our observation that TFIIF remains bound to the early EC and that it is not displaced in the context of a paused RNA Pol II elongation complex (PEC) containing DSIF and NELF (Figures S4D and S5B), this suggests that TFIIF might get evicted upon the conversion of the PEC to the EC*. Alternatively, Gdown1,[40] components of the SEC[25] (Figure S5C), or phosphorylation by CK2[22] could release TFIIF. Further studies about the interplay of these factors and the PEC will help to elucidate TFIIF displacement further.

Comparison with the bacterial RNA polymerase (RNAP) system shows that promoter escape involves conceptually similar steps, although the involved proteins strongly differ. First, a short RNA is synthesized that starts clashing with the σ3.2 linker in the initiation factor σ upon extension beyond 4 nt.[54–56] The steric presence of the σ3.2 linker in the path of the newly synthesized RNA was shown to be the cause for abortive transcription by RNAP.[54] The magnitude of abortive transcription is highly dependent on the sequence of the initially transcribed region.[57] Although the σ3.2 linker and σ4 are displaced as the RNAP transitions into an EC, the σ factor is not displaced upon transcription bubble rewinding as σ2 and σ3 are compatible with the rewound DNA.[58] Hence, the transcription factor σ typically remains associated with, and translocates with, RNAP upon promoter escape and formation of an early EC, but the affinity of the σ factor for RNAP decreases, enabling dissociation of σ during elongation.[58–61] Displacement of the σ factor can be facilitated by binding of the elongation regulatory factor NusA, which stabilizes promoter-proximally paused RNAP.[58]

With respect to the archaeal transcription system, the RNAP resembles RNA Pol II, and archaea contain the initiation factors TBP, TFB, and TFE that are related to human TBP, TFIIB, and TFIIE, respectively.[62] Promoter escape has not been studied in detail in this system. However, it has been shown that the TFB reader stimulates abortive transcription[63] and is displaced when the RNA extends to 10 nt.[64] Complete release of TFB, however, occurs after transcription bubble rewinding[65] when the RNA is longer than 15 nt.[64] Additionally, it is known that binding of the DSIF-related archaeal elongation factor Spt4-Spt5 occurs at the polymerase clamp[66] and can displace the archaeal TFIIE counterpart TFE.[67]

In summary, promoter escapes from bacteria to eukaryotes shares conceptually similar steps: displacement of TFIIB/TFB-reader[64] or σ3.2-linker[54–56] by the RNA transcript, rewinding of the extended transcription bubble,[9–13,65,68] and regulated displacement of initiation factors by elongation factors.[58,67] We suggest that these conserved features of promoter escape are likely a consequence of the nature of nucleic acid conformation, interaction, and energetics, whereas different proteins evolved around DNA and RNA to catalyze transcription initiation and the transition to elongation. However, it remains to be shown if the underlying mechanisms of promoter escape in the three kingdoms are universal.

While this paper was under revision, another study[69] reported structures of early transcription complexes containing 2 to 17 nt of RNA. Similar to our results, they report on conformational changes of the TFIIB-reader domain and transcription bubble rewinding. Based on their structures, they suggest that the TFIIB reader stabilizes the initial RNA transcript when it is shorter than 6 nt and that the B-reader is fully displaced at 7 nt RNA length. However, we observe density for the TFIIB reader as it competes with the growing RNA once transcription proceeds beyond 6 nt. This is likely due to the higher cryo-EM map quality of our structures that allowed us to assign the density of TFIIB. Additionally, the other study[69] suggests that during early transcription, the scrunched single-stranded template DNA is accumulated in a narrow "channel T" of RNA Pol II and that the release of the scrunched template DNA triggers dissociation of all GTFs from RNA Pol II except TFIIF and TFIIE. The scrunching mechanism for release of the upstream promoter complex as

well as the observation that TFIIF and TFIIE remain bound to RNA Pol II are compatible with our conclusions.

Limitations of the study

Due to the highly dynamic nature of promoter escape and the flexibility of transcription complexes, extensive classification of the cryo-EM images was performed to sort out the conformational and compositional heterogeneity of the particles. In turn, certain states might be filtered out during data processing due to their low occurrence. Additionally, the transcription reaction was subjected to GraFix and dialysis before plunge freezing, so the intermediates we captured may be putative direct intermediates during promoter escape due to the extensive handling time. Besides, the use of a 14-nt U-less cassette in our experimental design prevents observation of intermediates that transcribe further than 14 nt (outside of misincorporation events). Moreover, our competition assay could not distinguish whether DSIF and NELF prevent the rebinding of TFIIE and TFIIH after they dissociate stochastically or whether DSIF and NELF actively displace TFIIE and TFIIH by forming a transient intermediate with all four factors bound to RNA Pol II. The molecular details of such displacement remain to be understood. Finally, we were not able to resolve TFIIH at high resolution, likely due to the presence of ATP during sample preparation. This limits our understanding on the mechanism of TFIIH during promoter escape.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Material availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Bacteria strains
  - Insect cell lines
  - Pig thymus tissue
- METHOD DETAILS
  - Cryo-EM sample preparation with AdU14 promoter DNA
  - Visualization of abortive transcription
  - Biochemical analysis of the binding of TFIIH to EC14
  - Competition assays
  - Biochemical analysis of the retention of GTFs on EC14
  - Cryo-EM data collection and processing
  - Model building
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2024.03.016.

ACKNOWLEDGMENTS

We thank current and past members of the Cramer laboratory, in particular U. Steuerwald for maintaining the electron microscopy facility; C. Oberthür for purifying DSIF and NELF; G. Garg for providing RNGTT; G. Kokic, J. Abril-Garrido, J. Walshe, and S. Schilbach for discussions; and O. Dybkov for providing the glass plates for RNA sequencing gel and advice on radioactivity experiments.

AUTHOR CONTRIBUTIONS

Y.Z. designed and carried out all experiments and data analysis unless stated otherwise. Y.Z. and C.D. collected cryo-EM data. F.G. purified general transcription factors. E.O., C.D., and P.C. supervised research. Y.Z., E.O., and C.D. interpreted the data. Y.Z., C.D., and P.C. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 2, 2023
Revised: January 4, 2024
Accepted: March 16, 2024
Published: April 10, 2024

REFERENCES

1. Boeger, H., Bushnell, D.A., Davis, R., Griesenbeck, J., Lorch, Y., Strattan, J.S., Westover, K.D., and Kornberg, R.D. (2005). Structural basis of eukaryotic gene transcription. FEBS Lett. *579*, 899–903. https://doi.org/10.1016/j.febslet.2004.11.027.
2. Nogales, E., Louder, R.K., and He, Y. (2017). Structural Insights into the Eukaryotic Transcription Initiation Machinery. Annu. Rev. Biophys. *46*, 59–83. https://doi.org/10.1146/annurev-biophys-070816-033751.
3. Roeder, R.G. (2019). 50+ years of eukaryotic transcription: an expanding universe of factors and mechanisms. Nat. Struct. Mol. Biol. *26*, 783–791. https://doi.org/10.1038/s41594-019-0287-x.
4. Cramer, P. (2019). Organization and regulation of gene transcription. Nature *573*, 45–54. https://doi.org/10.1038/s41586-019-1517-4.
5. Kim, T.K., Ebright, R.H., and Reinberg, D. (2000). Mechanism of ATP-dependent promoter melting by transcription factor IIH. Science *288*, 1418–1422. https://doi.org/10.1126/science.288.5470.1418.
6. 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. USA *112*, 3961–3966. https://doi.org/10.1073/pnas.1417709112.
7. Luse, D.S. (2013). Promoter clearance by RNA polymerase II. Biochim. Biophys. Acta *1829*, 63–68. https://doi.org/10.1016/j.bbagrm.2012.08.010.
8. Dvir, A. (2002). Promoter escape by RNA polymerase II. Biochim. Biophys. Acta *1577*, 208–223. https://doi.org/10.1016/s0167-4781(02)00453-0.
9. Pal, M., Ponticelli, A.S., and Luse, D.S. (2005). The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. Mol. Cell *19*, 101–110. https://doi.org/10.1016/j.molcel.2005.05.024.
10. Holstege, F.C.P., Fiedler, U., and Timmers, H.T.M. (1997). Three transitions in the RNA polymerase II transcription complex during initiation. EMBO J. *16*, 7468–7480. https://doi.org/10.1093/emboj/16.24.7468.
11. Kapanidis, A.N., Margeat, E., Ho, S.O., Kortkhonjia, E., Weiss, S., and Ebright, R.H. (2006). Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. Science *314*, 1144–1147. https://doi.org/10.1126/science.1131399.
12. Fazal, F.M., Meng, C.A., Murakami, K., Kornberg, R.D., and Block, S.M. (2015). Real-time observation of the initiation of RNA polymerase II transcription. Nature *525*, 274–277. https://doi.org/10.1038/nature14882.

13. Revyakin, A., Liu, C.Y., Ebright, R.H., and Strick, T.R. (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. Science *314*, 1139–1143. https://doi.org/10.1126/science.1131398.

14. Luse, D.S., and Jacob, G.A. (1987). Abortive initiation by RNA polymerase II in vitro at the adenovirus 2 major late promoter. J. Biol. Chem. *262*, 14990–14997. https://doi.org/10.1016/S0021-9258(18)48127-6.

15. Kostrewa, D., Zeller, M.E., Armache, K.J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. (2009). RNA polymerase II-TFIIB structure and mechanism of transcription initiation. Nature *462*, 323–330. https://doi.org/10.1038/nature08548.

16. Sainsbury, S., Niesser, J., and Cramer, P. (2013). Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. Nature *493*, 437–440. https://doi.org/10.1038/nature11715.

17. Čabart, P., Újvári, A., Pal, M., and Luse, D.S. (2011). Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes. Proc. Natl. Acad. Sci. USA *108*, 15786–15791. https://doi.org/10.1073/pnas.1104591108.

18. Zawel, L., Kumar, K.P., and Reinberg, D. (1995). Recycling of the general transcription factors during RNA polymerase II transcription. Genes Dev. *9*, 1479–1490. https://doi.org/10.1101/gad.9.12.1479.

19. Joo, Y.J., Ficarro, S.B., Chun, Y., Marto, J.A., and Buratowski, S. (2019). In vitro analysis of RNA polymerase II elongation complex dynamics. Genes Dev. *33*, 578–589. https://doi.org/10.1101/gad.324202.119.

20. Price, D.H., Sluder, A.E., and Greenleaf, A.L. (1989). Dynamic Interaction between a *Drosophila* Transcription Factor and RNA Polymerase II. Mol. Cell. Biol. *9*, 1465–1475. https://doi.org/10.1128/mcb.9.4.1465-1475.1989.

21. Bengal, E., Flores, O., Krauskopf, A., Reinberg, D., and Aloni, Y. (1991). Role of the Mammalian Transcription Factor-IIF, Factor-IIS, and Factor-IIX during Elongation by RNA Polymerase II. Mol. Cell. Biol. *11*, 1195–1206. https://doi.org/10.1128/mcb.11.3.1195-1206.1991.

22. Újvári, A., Pal, M., and Luse, D.S. (2011). The functions of TFIIF during initiation and transcript elongation are differentially affected by phosphorylation by casein kinase 2. J. Biol. Chem. *286*, 23160–23167. https://doi.org/10.1074/jbc.M110.205658.

23. Xu, Y., Bernecky, C., Lee, C.T., Maier, K.C., Schwalb, B., Tegunov, D., Plitzko, J.M., Urlaub, H., and Cramer, P. (2017). Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex. Nat. Commun. *8*, 15741. https://doi.org/10.1038/ncomms15741.

24. Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018). Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. Nature *560*, 607–612. https://doi.org/10.1038/s41586-018-0440-4.

25. Chen, Y., Vos, S.M., Dienemann, C., Ninov, M., Urlaub, H., and Cramer, P. (2021). Allosteric transcription stimulation by RNA polymerase II super elongation complex. Mol. Cell *81*, 3386–3399.e10. https://doi.org/10.1016/j.molcel.2021.06.019.

26. Compe, E., Genes, C.M., Braun, C., Coin, F., and Egly, J.M. (2019). TFIIE orchestrates the recruitment of the TFIIH kinase module at promoter before release during transcription. Nat. Commun. *10*, 2084. https://doi.org/10.1038/s41467-019-10131-1.

27. Fujiwara, R., Damodaren, N., Wilusz, J.E., and Murakami, K. (2019). The capping enzyme facilitates promoter escape and assembly of a follow-on preinitiation complex for reinitiation. Proc. Natl. Acad. Sci. USA *116*, 22573–22582. https://doi.org/10.1073/pnas.1905449116.

28. Yang, C., Fujiwara, R., Kim, H.J., Basnet, P., Zhu, Y., Gorbea Colón, J.J., Steimle, S., Garcia, B.A., Kaplan, C.D., and Murakami, K. (2022). Structural visualization of *de novo* transcription initiation by *Saccharomyces cerevisiae* RNA polymerase II. Mol. Cell *82*, 660–676.e9. https://doi.org/10.1016/j.molcel.2021.12.020.

29. Giardina, C., and Lis, J.T. (1993). DNA melting on yeast RNA polymerase II promoters. Science *261*, 759–762. https://doi.org/10.1126/science.8342041.

30. 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. https://doi.org/10.1074/jbc.M116.724583.

31. Qiu, C., Jin, H., Vvedenskaya, I., Llenas, J.A., Zhao, T., Malik, I., Visbisky, A.M., Schwartz, S.L., Cui, P., Čabart, P., et al. (2020). Universal promoter scanning by Pol II during transcription initiation in *Saccharomyces cerevisiae*. Genome Biol. *21*, 132. https://doi.org/10.1186/s13059-020-02040-0.

32. Murakami, K., Mattei, P.J., Davis, R.E., Jin, H., Kaplan, C.D., and Kornberg, R.D. (2015). Uncoupling Promoter Opening from Start-Site Scanning. Mol. Cell *59*, 133–138. https://doi.org/10.1016/j.molcel.2015.05.021.

33. Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev. *33*, 960–982. https://doi.org/10.1101/gad.325142.119.

34. Dvir, A., Conaway, R.C., and Conaway, J.W. (1997). A role for TFIIH in controlling the activity of early RNA polymerase II elongation complexes. Proc. Natl. Acad. Sci. USA *94*, 9006–9010. https://doi.org/10.1073/pnas.94.17.9006.

35. Dvir, A., Conaway, R.C., and Conaway, J.W. (1996). Promoter escape by RNA polymerase II. A role for an ATP cofactor in suppression of arrest by polymerase at promoter-proximal sites. J. Biol. Chem. *271*, 23352–23356. https://doi.org/10.1074/jbc.271.38.23352.

36. Aibara, S., Schilbach, S., and Cramer, P. (2021). Structures of mammalian RNA polymerase II pre-initiation complexes. Nature *594*, 124–128. https://doi.org/10.1038/s41586-021-03554-8.

37. 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. https://doi.org/10.1038/nature17970.

38. Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018). Structure of paused transcription complex Pol II-DSIF-NELF. Nature *560*, 601–606. https://doi.org/10.1038/s41586-018-0442-2.

39. Hieb, A.R., Baran, S., Goodrich, J.A., and Kugel, J.F. (2006). An 8 nt RNA triggers a rate-limiting shift of RNA polymerase II complexes into elongation. EMBO J. *25*, 3100–3109. https://doi.org/10.1038/sj.emboj.7601197.

40. Jishage, M., Malik, S., Wagner, U., Uberheide, B., Ishihama, Y., Hu, X.P., Chait, B.T., Gnatt, A., Ren, B., and Roeder, R.G. (2012). Transcriptional Regulation by Pol II(G) Involving Mediator and Competitive Interactions of Gdown1 and TFIIF with Pol II. Mol. Cell *47*, 491–492. https://doi.org/10.1016/j.molcel.2012.07.023.

41. 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. https://doi.org/10.1038/nature24282.

42. Maxon, M.E., Goodrich, J.A., and Tjian, R. (1994). Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. Genes Dev. *8*, 515–524. https://doi.org/10.1101/gad.8.5.515.

43. Garg, G., Dienemann, C., Farnung, L., Schwarz, J., Linden, A., Urlaub, H., and Cramer, P. (2023). Structural insights into human co-transcriptional capping. Mol. Cell *83*, 2464–2477.e5. https://doi.org/10.1016/j.molcel.2023.06.002.

44. Missra, A., and Gilmour, D.S. (2010). Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the *Drosophila* RNA polymerase II transcription elongation complex. Proc. Natl. Acad. Sci. USA *107*, 11301–11306. https://doi.org/10.1073/pnas.1000681107.

45. Søgaard, T.M.M., and Svejstrup, J.Q. (2007). Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. J. Biol. Chem. *282*, 14113–14120. https://doi.org/10.1074/jbc.M701345200.

46. Schones, D.E., Cui, K.R., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z.B., Wei, G., and Zhao, K.J. (2008). Dynamic regulation of nucleosome positioning in the human genome. Cell *132*, 887–898. https://doi.org/10.1016/j.cell.2008.02.022.

47. Mavrich, T.N., Jiang, C.Z., Ioshikhes, I.P., Li, X.Y., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., et al. (2008). Nucleosome organization in the *Drosophila* genome. Nature *453*, 358–362. https://doi.org/10.1038/nature06929.

48. Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. (2007). Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. Nature *446*, 572–576. https://doi.org/10.1038/nature05632.

49. Weber, C.M., Ramachandran, S., and Henikoff, S. (2014). Nucleosomes Are Context-Specific, H2A.Z-modulated barriers to RNA polymerase. Mol. Cell *53*, 819–830. https://doi.org/10.1016/j.molcel.2014.02.014.

50. Abril-Garrido, J., Dienemann, C., Grabbe, F., Velychko, T., Lidschreiber, M., Wang, H., and Cramer, P. (2023). Structural basis of transcription reduction by a promoter-proximal +1 nucleosome. Mol. Cell *83*, 1798–1809.e7. https://doi.org/10.1016/j.molcel.2023.04.011.

51. Krogan, N.J., Kim, M., Ahn, S.H., Zhong, G.Q., Kobor, M.S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S., and Greenblatt, J.F. (2002). RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. Mol. Cell. Biol. *22*, 6979–6992. https://doi.org/10.1128/MCB.22.20.6979-6992.2002.

52. Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nat. Struct. Mol. Biol. *17*, 1272–1278. https://doi.org/10.1038/nsmb.1903.

53. Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature *487*, 128. https://doi.org/10.1038/nature11266.

54. Murakami, K.S., Masuda, S., and Darst, S.A. (2002). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. Science *296*, 1280–1284. https://doi.org/10.1126/science.1069594.

55. Li, L.T., Molodtsov, V., Lin, W., Ebright, R.H., and Zhang, Y. (2020). RNA extension drives a stepwise displacement of an initiation-factor structural module in initial transcription. Proc. Natl. Acad. Sci. USA *117*, 5801–5809. https://doi.org/10.1073/pnas.1920747117.

56. Basu, R.S., Warner, B.A., Molodtsov, V., Pupov, D., Esyunina, D., Fernández-Tornero, C., Kulbachinskiy, A., and Murakami, K.S. (2014). Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. J. Biol. Chem. *289*, 24549–24559. https://doi.org/10.1074/jbc.M114.584037.

57. Hsu, L.M., Cobb, I.M., Ozmore, J.R., Khoo, M., Nahm, G., Xia, L., Bao, Y., and Ahn, C. (2006). Initial transcribed sequence mutations specifically affect promoter escape properties. Biochemistry *45*, 8841–8854. https://doi.org/10.1021/bi060247u.

58. Mooney, R.A., Darst, S.A., and Landick, R. (2005). Sigma and RNA polymerase: an on-again, off-again relationship? Mol. Cell *20*, 335–345. https://doi.org/10.1016/j.molcel.2005.10.015.

59. Mukhopadhyay, J., Kapanidis, A.N., Mekler, V., Kortkhonjia, E., Ebright, Y.W., and Ebright, R.H. (2001). Translocation of σ^{70} with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. Cell *106*, 453–463. https://doi.org/10.1016/s0092-8674(01)00464-0.

60. Kapanidis, A.N., Margeat, E., Laurence, T.A., Doose, S., Ho, S.O., Mukhopadhyay, J., Kortkhonjia, E., Mekler, V., Ebright, R.H., and Weiss, S. (2005). Retention of transcription initiation factor σ^{70} in transcription elongation: single-molecule analysis. Mol. Cell *20*, 347–356. https://doi.org/10.1016/j.molcel.2005.10.012.

61. Pukhrambam, C., Molodtsov, V., Kooshkbaghi, M., Tareen, A., Vu, H., Skalenko, K.S., Su, M., Yin, Z., Winkelman, J.T., Kinney, J.B., et al. (2022). Structural and mechanistic basis of σ-dependent transcriptional pausing. Proc. Natl. Acad. Sci. USA *119*, e2201301119. https://doi.org/10.1073/pnas.2201301119.

62. Blombach, F., Matelska, D., Fouqueau, T., Cackett, G., and Werner, F. (2019). Key Concepts and Challenges in Archaeal Transcription. J. Mol. Biol. *431*, 4184–4201. https://doi.org/10.1016/j.jmb.2019.06.020.

63. Werner, F., and Weinzierl, R.O.J. (2005). Direct modulation of RNA polymerase core functions by basal transcription factors. Mol. Cell. Biol. *25*, 8344–8355. https://doi.org/10.1128/MCB.25.18.8344-8355.2005.

64. Dexl, S., Reichelt, R., Kraatz, K., Schulz, S., Grohmann, D., Bartlett, M., and Thomm, M. (2018). Displacement of the transcription factor B reader domain during transcription initiation. Nucleic Acids Res. *46*, 10066–10081. https://doi.org/10.1093/nar/gky699.

65. Spitalny, P., and Thomm, M. (2003). Analysis of the open region and of DNA-protein contacts of archaeal RNA polymerase transcription complexes during transition from initiation to elongation. J. Biol. Chem. *278*, 30497–30505. https://doi.org/10.1074/jbc.M303633200.

66. Martinez-Rucobo, F.W., Sainsbury, S., Cheung, A.C.M., and Cramer, P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. EMBO J. *30*, 1302–1310. https://doi.org/10.1038/emboj.2011.64.

67. Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R.H., Michaelis, J., and Werner, F. (2011). The Initiation Factor TFE and the Elongation Factor Spt4/5 Compete for the RNAP Clamp during Transcription Initiation and Elongation. Mol. Cell *43*, 263–274. https://doi.org/10.1016/j.molcel.2011.05.030.

68. Hsu, L.M. (2002). Promoter clearance and escape in prokaryotes. Biochim. Biophys. Acta *1577*, 191–207. https://doi.org/10.1016/S0167-4781(02)00452-9.

69. Chen, X., Liu, W., Wang, Q., Wang, X., Ren, Y., Qu, X., Li, W., and Xu, Y. (2023). Structural visualization of transcription initiation in action. Science *382*, eadi5120. https://doi.org/10.1126/science.adi5120.

70. 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. https://doi.org/10.1038/s41467-019-10745-5.

71. Bernecky, C., Plitzko, J.M., and Cramer, P. (2017). Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. Nat. Struct. Mol. Biol. *24*, 809–815. https://doi.org/10.1038/nsmb.3465.

72. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296. https://doi.org/10.1038/nmeth.4169.

73. Scheres, S.H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. *180*, 519–530. https://doi.org/10.1016/j.jsb.2012.09.006.

74. Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J.H., Lindahl, E., and Scheres, S.H.W. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. eLife *7*, e42166. https://doi.org/10.7554/eLife.42166.

75. Tegunov, D., and Cramer, P. (2019). Real-time cryo-electron microscopy data preprocessing with Warp. Nat. Methods *16*, 1146–1152. https://doi.org/10.1038/s41592-019-0580-y.

76. Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., and Adams, P.D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. Acta Crystallogr. D Struct. Biol. *74*, 531–544. https://doi.org/10.1107/S2059798318006551.

77. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. Acta Crystallogr. D Struct. Biol. *75*, 861–877. https://doi.org/10.1107/S2059798319011471.

78. Schrödinger, L., and Delano, W. (2020). PyMOL v2.5.0 (Schrödinger, Inc.).

79. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization

system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612. https://doi.org/10.1002/jcc.20084.

80. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501. https://doi.org/10.1107/S0907444910007493.

81. Croll, T.I. (2018). ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. Acta Crystallogr. D Struct. Biol. *74*, 519–530. https://doi.org/10.1107/S2059798318002425.

82. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/Nmeth.2019.

83. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J.M., and Cramer, P. (2016). Structure of transcribing mammalian RNA polymerase II. Nature *529*, 551–554. https://doi.org/10.1038/nature16482.

84. Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., et al. (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. Nat. Methods *5*, 53–55. https://doi.org/10.1038/Nmeth1139.

85. Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. J. Struct. Biol. *152*, 36–51. https://doi.org/10.1016/j.jsb.2005.07.007.

86. Davis, I.W., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res. *32*, W615–W619. https://doi.org/10.1093/nar/gkh398.

Please cite this article in press as: Zhan et al., Three-step mechanism of promoter escape by RNA polymerase II, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.016

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-SUPT5H	Proteintech	Cat# 16511-1-AP; RRID: AB_2878268
Rabbit polyclonal anti-ERCC3	Proteintech	Cat# 10580-1-AP; RRID: AB_2262174
Rabbit polyclonal anti-TFIIE-α/GTF2E1	Abcam	Cat# ab28177; RRID: AB_778322
Rabbit polyclonal anti-Rpb3	Bethyl Laboratories	Cat# A303-771A; RRID: AB_11218388
Bacterial and virus strains		
E.coli BL21 CodonPlus (DE3) RIL	Agilent	Cat# 230245
E. coli LOBSTR-BL21(DE3)-RIL	Kerafast	Cat# EC1002
E.coli DH10EMBacY	Geneva Biotech	N/A
E.coli XL-1 Blue	Agilent	Cat# 200249
Biological samples		
Sus scrofa thymus	Locally sources	N/A
Chemicals, peptides, and recombinant proteins		
Sus scrofa RNA polymerase II	Vos et al.[38]	N/A
Homo sapiens TBP	Aibara et al.[36]	N/A
Homo sapiens TFIIA	Aibara et al.[36]	N/A
Homo sapiens TFIIB	Aibara et al.[36]	N/A
Homo sapiens TFIIF	Aibara et al.[36]	N/A
Homo sapiens TFIIE	Aibara et al.[36]	N/A
Homo sapiens TFIIH	Kokic et al.[70]	N/A
Glutaraldehyde 25%	EMS	Cat# 16200
Proteinase K	New England Biolabs	Cat# P8107S
Invitrogen UltraPure 0.5M EDTA, pH 8.0	Thermo Fisher Scientific	Cat# 15575020
Invitrogen Ambion Sodium Acetate (3M), pH 5.5	Thermo Fisher Scientific	Cat# AM9740
NTP Set, 100 mM Solution	Thermo Fisher Scientific	Cat# R0481
Urea (RNase-free)	Panreac AppliChem	Cat# A1049
2x RNA Loading Dye	New England Biolabs	Cat# B0363S
DL-Dithiothreitol solution, 1M	Sigma-Aldrich	Cat# 43816
40% Acrylamide/bis-acrylamide 19:1	Sigma-Aldrich	Cat# A9926
TRIS borate-EDTA buffer solution (10x)	Sigma-Aldrich	Cat# 93290
Phusion DNA Polymerase	House sourced	N/A
Phusion® HF Buffer Pack	New England Biolabs	Cat# B0518S
dNTP Set, 100 mM Solutions	Thermo Fisher Scientific	Cat# R0186
DMSO	Sigma-Aldrich	Cat# D8418
GlycoBlue™ Coprecipitant	Thermo Fisher Scientific	Cat# AM9516
UTP, [α-^{32}P]- 3000Ci/mmol, 10mCi/ml, EasyTide, 250 μCi	Perkin Elmer	Cat# BLU507H250UC
Deposited data		
ITC8 cryo-EM reconstruction	This study	EMD-19718
ITC10 cryo-EM reconstruction	This study	EMD-19719
ITC12 cryo-EM reconstruction	This study	EMD-19743
EC14a cryo-EM reconstruction (composite map)	This study	EMD-19726
EC14b cryo-EM reconstruction (composite map)	This study	EMD-19720
ITC8 model	This study	PDB-8S51

(*Continued on next page*)

Please cite this article in press as: Zhan et al., Three-step mechanism of promoter escape by RNA polymerase II, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.016

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ITC10 model	This study	PDB-8S52
ITC12 model	This study	PDB-8S5N
EC14a model	This study	PDB-8S55
EC14b model	This study	PDB-8S54
ITC8+TFIIH cryo-EM reconstruction	This study	EMD-19721
ITC10+TFIIH cryo-EM reconstruction	This study	EMD-19722
ITC12+TFIIH cryo-EM reconstruction	This study	EMD-19723
EC14a+TFIIH cryo-EM reconstruction	This study	EMD-19724
EC14b+TFIIH cryo-EM reconstruction	This study	EMD-19725
EC14a cryo-EM reconstruction (global map)	This study	EMD-19795
EC14b cryo-EM reconstruction (global map)	This study	EMD-19796
EC14 core Pol II cryo-EM reconstruction (focused map)	This study	EMD-19797
Experimental models: Cell lines		
Sf9 cells	Oxford expression technology	Cat# 600100
Sf21 cells	Expression Systems	Cat# 94-003F
High Five cells	Expression Systems	Cat# 94-002F
Oligonucleotides		
EC14 template 5′ – TTC TGC TGG CCC TCG CTG TCA AAA TTG CGG CTG GGT GTG AGG ACG AAC GCG CCC CCA CCC CCT TTT ATA GCC CCC CTT CAG GAA CAC – 3′	This work	Integrated DNA Technologies
EC14 non-template 5′ – GTG TTC CTG AAG GGG GGC TAT AAA AGG GGG TGG GGG CGC GTT CGT CCT CAC AGG GTC GGC GTT TTT TGA CAG CGA GGG CCA GCA GAA – 3′	This work	Integrated DNA Technologies
EC14 RNA 5′ – rUrUrU rCrCrC rArGrC rCrGrC rArA – 3′	This work	Integrated DNA Technologies
EC24 template 5′ – GAC AAT CTT AGC GCA GAA GTC ATG CCC GCT TTT GAG AAA AAG TTC TCT CCC CTT CTG CTG GCC – 3′	This work	Integrated DNA Technologies
EC24 non-template 5′ – GGC CAG CAG AAC CCC TCT CTT GTT TTT CTC AAA AGC GGG CAT GAC TTC TGC GCT AAG ATT GTC – 3′	This work	Integrated DNA Technologies
EC24 RNA 5′ – rUrCrC rCrGrG rUrCrG rUrCrU rUrGrG rGrGrA rGrArG rArArC – 3′	This work	Integrated DNA Technologies
EC24 template_short 5′ – GTC ATG CCC GCT TTT GAG AAA AAG TTC TCT CCC CTT CTG CTG GCC – 3′	This work	Integrated DNA Technologies
EC24 non-template_short 5′ – GGC CAG CAG AAC CCC TCT CTT GTT TTT CTC AAA AGC GGG CAT GAC – 3′	This work	Integrated DNA Technologies
EC14 RNA_short 5′ – rCrUrU rGrGrG rGrArG rArGrA rArC – 3′	This work	Integrated DNA Technologies
AdU14 & AdU16 PCR forward primer 5′ – GTG TTC CTG AAG GGG GGC TAT AAA AGG GGG TGG GG – 3′	This work	Sigma-Aldrich
AdU14 & AdU16 PCR reverse primer 5′ – GAC AAT CTT AGC GCA GAA GTC ATG CCC GCT TTT GAG AAA AAG TTC TCT CC – 3′	This work	Sigma-Aldrich
Recombinant DNA		
438A-hTBP	Aibara et al.[36]	N/A
pOPINF-hTFIIB	Aibara et al.[36]	N/A
438A-hTFIIA	Aibara et al.[36]	N/A
pETDuet-1-hTFIIE	Aibara et al.[36]	N/A
pAHS3C-hTFIIF	Aibara et al.[36]	N/A
438C-XPD-p52-p34-p8-p62-p44-XPD	Kokic et al.[70]	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
438B-CCNH-CDK7-MAT1	Kokic et al.[70]	N/A
pKSUMO_10xHis-3C-SPT4_SPT5	Bernecky et al.[71]	N/A
438-A_NELFA_B_E_H6-TEV-D	Vos et al.[38]	N/A
1B_Nhis6_TEV-RNGTT_H	Garg et al.[43]	N/A
pUC119-AdMLU14	This study	N/A
pUC119-AdMLU16	This study	N/A
Software and algorithms		
cryoSPARC	Punjani et al.[72]	https://cryosparc.com/
RELION 3.1	Scheres[73]; Zivanov et al.[74]	https://github.com/3dem/relion
Warp 1.0.9	Tegunov and Cramer[75]	http://www.warpem.com
PHENIX 1.20.1	Afonine et al.[76]; Liebschner et al.[77]	http://www.phenix-online.org
PyMol 2.5.4	Schrödinger and Delano[78]	http://www.pymol.org
UCSF Chimera	Pettersen et al.[79]	https://www.cgl.ucsf.edu/chimera/
UCSF Chimera X 1.6.1	Pettersen et al.[79]	https://www.cgl.ucsf.edu/chimerax/
Coot 0.9.8.7	Emsley et al.[80]	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
ISOLDE 1.6	Croll[81]	https://isolde.cimr.cam.ac.uk/
ImageJ 1.53t	Schindelin et al.[82]	https://imagej.nih.gov/ij/index.html
Prism 10.0.0	GraphPad Software Inc (California, USA)	https://www.graphpad.com/
Other		
Titan Krios G2	FEI/Thermo Fisher Scientific	N/A
QuantumLS energy Filter	Gatan	N/A
K3 Summit Direct Electron Detector	Gatan	N/A
Quantifoil™ R3.5/1, copper, mesh 200	Quantifoil	N/A
Slide-A-Lyzer™ MINI Dialysis Devices (20 kDa MWCO)	Thermo Fisher Scientific	Cat# 69590
Typhoon™ 9500 FLA imager	Typhoon™ 9500 FLA imager	N/A

RESOURCE AVAILABILITY

Lead contact

Correspondence and request for materials should be addressed to Christian Dienemann (christian.dienemann@mpinat.mpg.de).

Material availability

Materials are available from Christian Dienemann upon request under a material transfer agreement with the Max Planck Society.

Data and code availability

- The cryo-EM reconstructions and final models were deposited to the Electron Microscopy Database (EMDB) and to the Protein Data Bank (PDB) under the following accession codes: EMDB-19718 and PDB 8S51 for ITC8, EMDB-19719 and PDB 8S52 for ITC10, EMDB-19743 and PDB 8S5N for ITC12, EMD-19797 (focused map) for EC14 core Pol II, EMDB-19795 (global map), EMDB-19726 (composite map) and PDB 8S55 for EC14a, EMDB-19796 (global map), EMDB-19720 (composite map) and PDB 8S54 for EC14b, EMDB-19721 for ITC8+TFIIH, EMDB-19722 for ITC10+TFIIH, EMDB-19723 for ITC12+TFIIH, EMDB-19724 for EC14a+TFIIH and EMDB-19725 for EC14b+TFIIH.
- 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 STUDY PARTICIPANT DETAILS

Bacteria strains

Escherichia coli strains XL1-Blue, DH10 EMBacY, BL21 CodonPlus (DE3) RIL and LOBSTR-BL21(DE3)-RIL were cultured in standard LB media with the respective antibiotics at 37°C while shaking.

Insect cell lines

Sf9, Sf21, and Hi5 cell lines were used for recombinant protein expression in insect cells. All cell lines were grown in ESF921 (Expression Technologies) at 27°C.

Pig thymus tissue

Pig thymus tissues were harvested from a local butcher in Göttingen, flash-frozen in liquid nitrogen and stored at -80°C before use.

METHOD DETAILS

Cryo-EM sample preparation with AdU14 promoter DNA

AdML promoter scaffolds containing U-less cassettes of different lengths were inserted into pUC119 vectors as previously described.[36] DNA templates were amplified by large-scale PCR, purified by ion-exchange chromatography with a ResourceQ column (5 mL, Cytiva) followed by phenol-chloroform extraction. The sequences of the DNA templates used for cryo-EM sample preparation are listed below. TATA-box and TSS are highlighted in bold. U-less cassettes are underlined. AdU14 template (139 bp): 5′-GTG TTC CTG AAG GGG GGC **TAT AAA A**GG GGG TGG GGG CGC GTT CGT CCT C**A**C ACC CAG CCG CAA TTT TGA CAG CGA GGG CCA GCA GAA GGG GAG AGA ACT TTT TCT CAA AAG CGG GCA TGA CTT CTG CGC TAA GAT TGTC-3′.

S. scrofa Pol II and human initiation factors (TBP, TFIIA, TFIIB, TFIIE, TFIIF and TFIIH) were purified as previously described.[36,38,50,70,83] Sample preparation was performed based on published protocol[36] with modifications. Before PIC formation, three complexes, DNA-TFIIA-TFIIB-TBP, Pol II-TFIIF and TFIIH-TFIIE were incubated as three separate samples at room temperature for 5-10 min. Then, the pre-assembled samples were mixed and further incubated at 30°C for 1 h. Transcription was initiated by the addition of 0.5 mM NTP mix consisting of ATP, CTP and GTP. Transcription was allowed to proceed for 2∼10 min at 30°C in reaction buffer containing 20 mM K-HEPES pH 7.5, 110 mM KCl, 3% glycerol, 16 mM $MgCl_2$, 0.5 mM DTT. Afterwards, 6/7th of the resulting transcription reaction was loaded onto a 15%-40% (w/v) sucrose gradient while mildly cross-linked by GraFix[84] at 32,000 rpm (SW60 rotor) for 16h at 4 °C. After centrifugation, the gradient was fractionated manually from the top and quenched with 10 mM lysine and 40 mM aspartate for 10 min on ice. In parallel to GraFix, the 1/7th of the same transcription reaction was loaded onto a 15%-40% (w/v) sucrose gradient but without cross-linker to monitor the distribution of the transcription complexes. After ultracentrifugation, fractions from non-crosslinked sample were isopropanol precipitated and analyzed by urea-PAGE. Corresponding fractions that contained the transcribed complexes from the cross-linked sample were then pooled and dialyzed against the Buffer D (20 mM K-HEPES pH 7.5, 90 mM KCl, 1% glycerol (v/v), 1mM DTT) at 4°C for 7-8 h to remove sucrose and glycerol. Subsequently, a thin piece of home-made continuous carbon film (∼2.7 nm) was floated onto the dialyzed sample and the particles were allowed to adsorb onto the film at 4°C for 15 min. The film was then picked up by a holey-carbon grid (Quantifoil R3.5/1, copper, mesh 200) and 3.5 μL of Buffer D was immediately added onto the carbon film prior to vitrification with a Vitrobot Mark IV (FEI) operated at 4 °C and 100% humidity. The grids were blotted for 1-1.5 s with blot force of 5 before plunge-freezing in liquid ethane.

Visualization of abortive transcription

Abortive transcription was visualized by gradient ultracentrifugation. For that, promoter dependent *in vitro* transcription reactions were performed with AdU14 and AdU16 promoter DNA. The sequence of AdU14 template was same as cryo-EM sample preparation. The sequence of AdU16 template is listed below. TATA box and TSS are highlighted as bold. U16-less cassette is underlined. AdU16 template (139 bp): 5′-GTG TTC CTG AAG GGG GGC **TAT AAA A**GG GGG TGG GGG CGC GTT CGT CCT C**A**C ACC CAG CCG CAA CGT TTT CAG CGA GGG CCA GCA GAA GGG GAG AGA ACT TTT TCT CAA AAG CGG GCA TGA CTT CTG CGC TAA GAT TGTC-3′.

PIC assembly was carried out in the same way as sample preparation. For each 20 μL reaction, 14 pmol Pol II, 71 pmol TFIIF, 15 pmol DNA template, 93 pmol TFIIA, 47 pmol TBP, 47 pmol TFIIB, 21 pmol TFIIE and 21 pmol TFIIH were used. Transcription was initiated with 0.5 mM ATP, 0.5 mM CTP, 0.5 mM GTP and 2 μCi [α-^{32}P]-CTP and allowed to take place for 15 min at 30°C in reaction buffer containing 20 mM K-HEPES pH 7.5, 110 mM KCl, 3% glycerol, 16 mM $MgCl_2$, 0.5 mM DTT. Transcription reactions were then loaded onto a 10%-30% (w/v) sucrose gradient in buffer containing 20 mM K-HEPES pH 7.5, 110 mM KCl, 3% glycerol (v/v), 1 mM DTT, 0.5 mM $MgCl_2$. Ultracentrifugation was performed at 55,000 rpm for 3 h at 4°C with a swinging bucket rotor S55S (Thermo Scientific). The gradients were fractioned manually from top. Fractions 1-4 and 7-22 were pooled separately and digested with proteinase K in the presence of GlycoBlue (Thermo Fisher) for 30 min at 37°C followed by isopropanol precipitation. The precipitated RNA pellets were resuspended in 2x RNA loading dye (NEB) and loaded onto an RNA sequencing gel (7 M urea, 1x TBE, 20% acrylamide:bis-acrylamide 19:1). To achieve similar exposure of the phosphorus screen, 2/15th of the RNA suspension from fraction 7-22 were loaded whereas the total RNA suspension from fraction 1-4 was loaded. For quantification see below.

Biochemical analysis of the binding of TFIIH to EC14

Analytical gradient ultracentrifugation was used to monitor the association of TFIIH with EC14. To assemble the Pol II ECs, the template DNA strand was mixed with RNA in equimolar ratio and annealed by first incubating at 75°C for 1 min then slowly cooling down to 4°C at the speed of 1°C per min. Next, the DNA-RNA scaffold was added to Pol II in 1.5x molar excess and incubated for 10 min at 30°C. Finally, the non-template DNA strand was added in 2x molar excess and incubated for another 10 min at 30°C. DNA and RNA oligonucleotides were purchased from Integrated DNA technologies. Sequences used for EC14 assembly were: template: 5′-TTC TGC TGG CCC TCG

CTG TCA AAA TTG CGG CTG GGT GTG AGG ACG AAC GCG CCC CCA CCC CCT TTT ATA GCC CCC CTT CAG GAA CAC-3′; non-template: 5′-GTG TTC CTG AAG GGG GGC TAT AAA AGG GGG TGG GGG CGC GTT CGT CCT CAC AGG GTC GGC GTT TTT TGA CAG CGA GGG CCA GCA GAA-3′; RNA: 5′-rUrUrU rCrCrC rArGrC rCrGrC rArA-3′. TFIIE and TFIIH were added to the preformed EC14 in 2x molar excess to Pol II and incubated at 30°C for 30 min in final buffer containing 20 mM K-HEPES pH 7.5, 100 mM KCl, 4% glycerol, 4 mM $MgCl_2$ and 1 mM DTT. Reactions with and without TFIIE were loaded onto a 10%-30% (w/v) sucrose gradient in buffer containing 110 mM KCl, 20 mM K-HEPES pH 7.5, 3% glycerol (v/v), 1 mM DTT, 0.5 mM $MgCl_2$. Ultracentrifugation was performed at 32,000 rpm for 16 h at 4°C with an SW60 rotor. Fractions were analyzed by SDS-PAGE followed by silver staining.

Competition assays

Analytical gradient ultracentrifugation was used to monitor the competition between TFIIE and elongation factors. DSIF, RNGTT and NELF were purified as previously reported.[38,43,71] The Pol II ECs were assembled as above. Sequences used for EC14 assembly were the same as above. Sequences used for EC24 assembly were: template: 5′-GAC AAT CTT AGC GCA GAA GTC ATG CCC GCT TTT GAG AAA AAG TTC TCT CCC CTT CTG CTG GCC-3′; non-template: 5′- GGC CAG CAG AAC CCC TCT CTT GTT TTT CTC AAA AGC GGG CAT GAC TTC TGC GCT AAG ATT GTC -3′; RNA: 5′-rUrCrC rCrGrG rUrCrG rUrCrU rUrGrG rGrGrA rGrArG rArArC-3′. The preformed EC was then incubated with CAK (1 μM) and ATP (1 mM) for 30 min at 30 °C to phosphorylate the Pol II CTD in final buffer containing 20 mM K-HEPES pH 7.5, 100 mM KCl, 4% glycerol, 4 mM $MgCl_2$ and 1 mM DTT. The phosphorylation reaction was stopped with 8 mM EDTA before adding initiation factors (TFIIE and TFIIF) in 2x molar excess to Pol II and incubated at 30°C for 15 min. Afterwards, elongation factors (DSIF and NELF or DSIF and RNGTT) were added in 2x molar excess to Pol II and incubated at 30°C for 15 min in final buffer containing 20 mM K-HEPES pH 7.5, 100 mM KCl, 4% glycerol, 2.5 mM EDTA and 1 mM DTT. Reactions were loaded onto a 10%-30% (w/v) sucrose gradient in buffer containing 110 mM KCl, 20 mM HEPES (pH 7.5), 3% glycerol (v/v), 1 mM DTT, 0.5 mM $MgCl_2$. Ultracentrifugation was performed at 32,000 rpm for 16 h at 4°C using an SW60 rotor. Gradient fractions were analyzed by SDS-PAGE followed by silver staining. For competition assays in the presence of DSIF and NELF, gradient fractions were also subjected to Western blotting with an antibody against Spt5.

To monitor the competition between TFIIE and core TFIIH with DSIF and NELF, EC24 was assembled as described above albeit with a different DNA scaffold with a shorter downstream DNA (23 bp, comparable to the length of the downstream DNA observed in our EC14 structures, which is ~22 bp) to decrease the tailing effect of TFIIH during gradient ultracentrifugation. Sequences used for EC24 assembly were: template: 5′-GTC ATG CCC GCT TTT GAG AAA AAG TTC TCT CCC CTT CTG CTG GCC-3′; non-template: 5′-GGC CAG CAG AAC CCC TCT CTT GTT TTT CTC AAA AGC GGG CAT GAC-3′; RNA: 5′-rUrCrC rCrGrG rUrCrG rUrCrU rUrGrG rGrGrA rGrArG rArArC-3′. After EC assembly, TFIIE, TFIIF and TFIIH core were added in 1.5x molar excess to Pol II and incubated at 30°C for 30 min. Afterwards, DSIF and NELF were added in 5x molar excess to Pol II and incubated at 30°C for 30 min in final buffer containing 20 mM K-HEPES pH 7.5, 110 mM KCl, 4% glycerol, 3 mM $MgCl_2$ and 1 mM DTT. Reactions were loaded onto a 10%-30% (w/v) sucrose gradient in buffer containing 110 mM KCl, 20 mM HEPES (pH 7.5), 3% glycerol (v/v), 1 mM DTT, 0.5 mM $MgCl_2$. Ultracentrifugation was performed at 55,000 rpm for 3 h at 4°C using a TLS55 rotor. Gradient fractions were analyzed by SDS-PAGE followed by both Western blot analysis and Coomassie blue staining.

Biochemical analysis of the retention of GTFs on EC14

To study the retention of TFIIE, TFIIF and core TFIIH on EC14 in the absence of crosslinker, an EC14 was assembled as described above with the DNA scaffold that contains a short downstream DNA of 23 bp and an RNA of 14 nt. The RNA sequence used for EC14 assembly was: 5′-rCrUrU rGrGrG rGrArG rArGrA rArC-3′. After EC assembly, TFIIE, TFIIF and TFIIH core were added in 1.5x molar excess to Pol II and incubated at 30°C for 30 min in final buffer containing 20 mM K-HEPES pH 7.5, 110 mM KCl, 4% glycerol, 2 mM $MgCl_2$ and 1 mM DTT. Samples were then loaded onto a 10%-30% (w/v) sucrose gradient in buffer containing 110 mM KCl, 20 mM HEPES (pH 7.5), 3% glycerol (v/v), 1 mM DTT, 0.5 mM $MgCl_2$. Control runs of free TFIIE, TFIIF and core TFIIH were performed in parallel. Ultracentrifugation was performed at 55,000 rpm for 3 h at 4°C using a TLS55 rotor. Gradient fractions were analyzed by SDS-PAGE followed by Coomassie blue staining.

Cryo-EM data collection and processing

Cryo-EM data were collected on a FEI Titan Krios transmission electron microscope with a K3 summit direct electron detector (Gatan) and a GIF quantum energy filter (Gatan), operated at 300 keV and with a slit width of 20 eV. Data collection was performed automatically with SerialEM[85] at a nominal magnification of 81,000x (1.05 Å per pixel) with a total dose of around 40 e/$Å^2$ fractionated over 40 frames. The defocus range applied was 0.7 μm to 1.7 μm.

On-the-fly motion correction, contrast-transfer function estimation and particle picking were performed with Warp.[75] Initial cleaning of the datasets was carried out by consecutive rounds of 2D and 3D classification in cryoSPARC[72] to remove ice contamination, falsely picked and aggregated particles. Further processing steps were performed in Relion3.1,[73,74] as described in detail in Figure S2. Focused 3D classification with a mask around core Pol II was performed to separate complexes with open and closed DNA. Particles that contained high resolution features for the DNA-RNA hybrid in the Pol II active site were CTF-refined and polished. These particles were further signal subtracted and classified with a mask around the upstream DNA to separate complexes before and after bubble rewinding. Particles with features of extended bubble were further classified with a mask around core Pol II to separate ITCs with different DNA-RNA

Please cite this article in press as: Zhan et al., Three-step mechanism of promoter escape by RNA polymerase II, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.03.016

hybrids. Particles with features of a rewound bubble were further signal subtracted and classified with a mask around TFIIE to separate different TFIIE conformations. Focused refinements were performed to improve the resolution of Pol II in each class.

Particles with high-resolution features for DNA-RNA hybrid were also signal subtracted and classified with a mask around TFIIH. Particles with concrete features of TFIIH were combined. To study the association of TFIIH before and after bubble rewinding, particle populations with concrete TFIIH features were intersected with particle populations from EC classes and ITC classes, respectively. The identified particles were then subjected to global 3D refinement.

Additionally, particles with features of a rewound bubble were also classified with a mask around the RNA exit tunnel. This resulted in two main species. First, particles with the TFIIB-ribbon occupying the RNA exit tunnel and second, particles with the full-length 14-nt RNA occupying the RNA exit tunnel. Intersection of particles with full-length 14-nt RNA with particles from EC14a resulted in 16,312 mutual particles (25% of EC14a). Intersection of particles with full-length 14-nt RNA with particles from EC14b resulted in 25,046 mutual particles (31% of EC14b). This suggests that the competition between RNA and TFIIB-ribbon in the RNA exit tunnel is independent of TFIIE as particles with full-length 14-nt RNA were similarly distributed in the two classes of different TFIIE states. Particles with the full-length 14-nt RNA were then subjected to focused refinement with a mask around core Pol II. The final composite maps were created with the focused refined map of core Pol II and EC14a/b.

Model building

To facilitate model building, cryo-EM maps were filtered according to local resolution in Relion3.1[73,74] and auto-sharpened in Phenix.[76,77] For modelling of the ITCs, previous published models (PDB: 7nw0,[36] 5iyd,[37] 5flm[83]) were rigid-body docked into focused refinement maps in Chimera[79] and adjusted manually in Coot.[80] The resulting models were then real-space refined in Phenix[76,77] followed by consecutive rounds of rebuilding with Coot[80] and ISOLDE.[81] Final models showed good stereochemistry as validated by Molprobity.[86] For modelling of EC14a/b, core Pol II (without stalk) from previous published model (PDB: 5flm.[83]) was rigid-body docked into focused refinement maps of core Pol II in Chimera[79] and adjusted manually in Coot.[80] The resulting models were then real-space refined in Phenix[76,77] followed by consecutive rounds of rebuilding with Coot[80] and ISOLDE.[81] Afterwards, TFIIE, Pol II stalk from previous published model (PDB: 7nw0.[36]) and core Pol II were rigid-body docked into focused refinement maps in Chimera[79] and real-space refined in Phenix.[76,77] The real-space refined models and TFIIF from a previously published model (PDB: 7nw0.[36]) were rigid-body docked into focused refinement maps in Chimera[79] to generate the final models of ECs. The final EC models were subjected to comprehensive validation (cryo-EM) in Phenix[76,77] and showed good stereochemistry.

QUANTIFICATION AND STATISTICAL ANALYSIS

Signals of RNA transcripts from abortive transcription assays were quantified with ImageJ 1.53.[82] Abortive transcription experiments were performed four times independently. The percentage of aborted RNA was defined as: (transcription signals from fractions 1-4)/(transcription signals from fractions 1-4+7.5*transcription signals from fractions 7-22). Statistical significance of the results was calculated with unpaired two-tailed t-tests using GraphPad Prism 9.5.1.

The western blot signals for XPB (or TFIIE-α) in each individual fraction were quantified with ImageJ 1.53[82] and normalized to the total signal of fraction 1-14. The normalized signals of every fraction were then plotted. The experiments were repeated three times independently. The percentage of Pol II-bound GTFs was defined as the sum of the normalized signals in fractions 9-14. Statistical significance of the results was calculated with unpaired two-tailed t-tests using GraphPad Prism 9.5.1.