Structural studies of RNA polymerase II complexes with human TFIIS and capping enzymes

Dissertation

For the award of the degree

"Doctor rerum naturalium"

Division of Mathematics and Natural Sciences

of the Georg-August-Universität Göttingen

Within the doctoral program

International Max Planck Research School for Molecular Biology

of the Georg-August University School of Science (GAUSS)

submitted by

Gaurika Garg

from Delhi, India

Göttingen 2022

Thesis Committee Members

Thesis Committee

Prof. Dr. Patrick Cramer
Max Planck Institute for Multidisciplinary Sciences, Department of Molecular Biology, Göttingen, Germany

Prof. Dr. Markus Zweckstetter
Max Planck Institute for Multidisciplinary Sciences, Structure determination of proteins using NMR group, Göttingen, Germany

Dr. Alex Faesen
Max Planck Institute for Multidisciplinary Sciences, Biochemistry of signal dynamics group, Göttingen, Germany

Members of the Examination Board

Referee: Prof. Dr. Patrick Cramer
Max Planck Institute for Multidisciplinary Sciences, Department of Molecular Biology, Göttingen, Germany

2nd Referee: Prof. Dr. Markus Zweckstetter
Max Planck Institute for Multidisciplinary Sciences, Structure determination of proteins using NMR group, Göttingen, Germany

Further members of the Examination Board

Dr. Alex Faesen
Max Planck Institute for Multidisciplinary Sciences, Biochemistry of signal dynamics group, Göttingen, Germany

Prof. Dr. Henning Urlaub
Max Planck Institute for Multidisciplinary Sciences, Bioanalytical mass spectrometry group, Göttingen, Germany

Prof. Dr. Gregor Eichele
Max Planck Institute for Multidisciplinary Sciences, Genes and behavior group, Göttingen, Germany

Prof. Dr. Hauke Hillen
Max Planck Institute for Multidisciplinary Sciences, Structure and function of molecular machines group, Göttingen, Germany

Date of the oral examination: November 2, 2022

Table of Contents

Acknowledgments ..... *VI*

Manuscripts and Contributions ..... *VII*

Summary ..... *IX*

1. Introduction *1*

1.1 An overview of Transcription **1**

1.2 Transcription by Pol II **2**

1.2.1 Pol II features 2

1.2.2 Nucleotide addition cycle 2

1.2.3 The stages of a transcription cycle 3

1.2.4 Transcription factor II S 6

1.3 Co-transcriptional processing **7**

1.4 mRNA capping **8**

1.4.1 Importance of the 5' cap 10

1.4.2 Capping pathways 11

1.5 Capping enzymes **12**

1.5.1 RNA guanylytransferase and 5' phosphatase (RNGTT) 13

1.5.2 mRNA guanine N7 methyltransferase (RNMT) 15

1.5.3 mRNA 2' O cap methyltransferase 1 (CMTR1) 16

1.5.4 Co-transcriptional capping 17

1.6 Scope of this dissertation **19**

2 Structure of a backtracked hexasomal intermediate of nucleosome transcription *23*

3 Structural insights into human co-transcriptional capping *57*

4 Discussion and outlook *105*

4.1 Pol II EC-TFIIS: novel findings **105**

4.2 Co-transcriptional capping: novel findings **106**

4.3 Future perspectives **108**

4.3.1 The missing links 108

4.3.2 Towards the complete model of co-transcriptional capping 109

4.3.3 Understanding the uncharacterized caps 112

4.3.4 Concluding remarks 112

References *115*

Abbreviations and acronyms ..... *VIII*

Declaration ..... *X*

Curriculum vitae ..... *XI*

Acknowledgments

Throughout my Ph.D. studies, I have received enormous support and encouragement from the people around me. I would first like to thank my supervisor, Prof. Patrick Cramer for providing me with the opportunity to work in an exciting scientific atmosphere and trusting me to work on an essential aspect of transcription. Patrick's feedback not only helped me learn a lot about presenting my research but also encouraged me to think of the bigger picture.

I would also like to thank my thesis advisory committee members Dr. Alex Faesen and Prof. Markus Zweckstetter or their valuable suggestions throughout my Ph.D. A special thanks to all the committee members for joining my thesis examination board. I would like to thank all my colleagues for a scientifically stimulating, helpful, and friendly working atmosphere. Especially all the people who have been a pleasure to work with on the manuscripts or have provided feedback for the work here. A big thanks to Janine Blümel, Almuth Burgdorf and Kirsten Backs for keeping the lab running.

I am also grateful to all the members of the International Max Planck Research School. A special thanks to the molecular biology program coordinators Steffen and Kerstin who were a source of invaluable support at every point through my journey in Göttingen. They take care of the program wonderfully and have been a huge support for not only the studies, but extracurriculars and advice.

I would like to extend a special thank you to the few colleagues who have become important friends. They have been supportive, given me scientific advice, been helpful whenever needed, and are a vital part of my life here. I would also like to express my gratitude to all the friends who are here, have moved away, or are from back in India. Thank you for all the advice, for cheering me through my thesis writing, for always listening and having my back, for new adventures, all the great laughs and fun times together, and for keeping me sane. I appreciate you all and it would be hard to imagine my Ph.D. journey without these friendships.

And finally, I owe the biggest thank you to my family. Especially my parents and my brother who have always being there for me. Even from miles away, they have been the biggest source of positivity and motivation through thick and thin. I truly appreciate their constant support, presence, and unconditional love. I thank them for pushing me to achieve my dreams and for always believing in me to walk my path.

Manuscripts and Contributions

Manuscript 1: Structure of a backtracked hexasomal intermediate of nucleosome transcription

Lucas Farnung*, Moritz Ochmann*, **Gaurika Garg***, Seychelle M. Vos*, Patrick Cramer.

Molecular Cell., 2022, ISSN 1097-2765

doi: https://doi.org/10.1016/j.molcel.2022.06.027

(* These authors contributed equally)

Contributions:

Gaurika Garg purified the mammalian Pol II. She assembled the complex *in vitro* and prepared the sample for cryo-EM. Further, she screened the samples, and collected data cryo-EM data. Finally, she processed the data and built the model for mammalian Pol II-TFIIS complex. This refers to the work represented in figures 6 and S5 completely, as well as parts of figures S2, S6 and table S1.

Manuscript 2: Structural insights into human co-transcriptional capping

Gaurika Garg*, Christian Dienemann, Lucas Farnung, Andreas Linden, Julianne Schwarz, Henning Urlaub, Patrick Cramer (in preparation)

(* These authors contributed equally)

Contributions:

Gaurika Garg designed the experimental strategy. She purified all the proteins required for the study. She assembled the complexes *in vitro* and prepared the sample for cryo-EM. Further, she screened the samples and collected cryo-EM data. Finally, she processed the data and built models for the six resulting density maps. She wrote the manuscript and prepared the figures and tables for the manuscript with inputs from all the other authors.

Prof. Dr. Patrick Cramer

Summary

Transcription is a conserved fundamental cellular process carried out by RNA polymerases (Pol) that results in the synthesis of an RNA transcript. Transcription of protein coding genes by Pol II is the first step in conversion of the genetic information into cellular proteome. The Pol II transcription cycle is divided into three distinct phases of initiation, elongation, and termination. In this dissertation, we provide insights into two essential aspects of elongation: Pol II backtracking and co-transcriptional capping. Backtracking is a result of Pol II trying to overcome barriers like the nucleosome. It results in the nascent RNA being extruded into a pore, causing Pol II arrest. Transcription factor II S (TFIIS), rescues Pol II under such circumstances. TFIIS travels through the Pol II crevice, facilitates RNA cleavage with its domain III in the pore, generating a new 3' end in Pol II active site. The first manuscript (published) focuses on the details of Pol II passage through the nucleosome with the help of TFIIS, where I resolve the structure of the mammalian Pol II-TFIIS. The structure reveals that the mammalian Pol II has a pre-opened crevice allowing insertion of human TFIIS without structural changes, unlike its yeast counterpart. The TFIIS domain III remains mobile, which might enable regulation of TFIIS entry into Pol II active site. The second manuscript (in-preparation), and the primary focus of this dissertation is the co-transcriptional capping of the 5' end of the emerging RNA. In humans, all mRNAs carry a 5' cap composed of a 7-methylguanosine attached via an inverted phosphate to the first transcribed nucleotide (nt) which is methylated at 2' O. This structure is essential to protect the RNA from exonucleolytic attacks and serves as a marker for 'self-RNA,' maintaining an antiviral state in the cell. DRB sensitivity inducing factor (DSIF), an elongation factor facilitates capping. Through this dissertation, I propose the first comprehensive model of mammalian co-transcriptional capping based on six structural intermediates of the process. The structures show that the human capping enzymes use Pol II stalk as a stable anchor to bind and act on the pre-mRNA at the exit channel. These structures enable us to capture the RNA movement between enzyme active sites, and the major structural rearrangements of DSIF that exemplify the importance of its modular nature. Failure to rescue a backtracked Pol II or to incorporate a 5' cap result in pre-mature transcription termination. In summary, our elucidated structures help unravel the molecular mechanisms governing transcription elongation and provide a robust framework for the fundamental process of co-transcriptional capping.

1. Introduction

1.1 An overview of Transcription

The central dogma first stated in 1957 by Francis Crick reports the sequence of information transfer in living organisms[1]. The framework describes the normal flow of biological information between the three major biopolymers: deoxyribonucleic acid (DNA) ribonucleic acid (RNA) and protein[2]. Transcription, a fundamental biological process of the dogma, involves copying the information from DNA segments into RNA molecules, of which the messenger RNAs (mRNAs) are translated into proteins[3].

Transcription of the cellular genome is carried out by multi-subunit enzymes called DNA-dependent RNA polymerases (RNAP)[4]. Bacteria and archaea possess one RNAP, whereas eukaryotes have three to five RNAPs responsible for transcribing different subset of genes[5]. All of these multi-subunit RNAPs share a common structural core. This structurally conserved core consists of five subunits, mainly comprising the functional elements of the active center and thereby a shared molecular mechanism[5–7]. Eukaryotic RNAPs carry up to ten additional subunits, which is largely linked with RNAPs having to transcribe through the histone compacted, and hence more restrictive eukaryotic DNA[8,9].

The multi-subunit polymerases (Pol) that transcribe the eukaryotic genome are Pol I, Pol II, Pol III, and additionally Pol IV and Pol V in plants[4]. Pol I, II, and III are responsible for synthesizing the 25S ribosomal RNA (rRNA), messenger RNA (mRNA), and small RNAs including transfer RNAs and 5S rRNA, respectively[10]. While Pol is the central player in transcription, the development of promoter-specific in vitro transcription assays[11,12] also revealed the importance and need for general transcription factors (GTFs)[2,13]. Pols together with the GTFs traverse through nucleosomes (the fundamental unit with histone proteins and DNA wrapped around) to generate the corresponding transcripts. Chromatin remodelers and histone chaperones help with this process[14,15].

RNA Polymerase II (Pol II) transcription is a complex yet coordinated event that results in the synthesis of an mRNA[16]. Pol II dependent transcription cycle is divided into three distinct phases of initiation, elongation, and termination (Figure 1)[17]. Initiation begins with Pol II and certain GTFs assembling at the promoter to start the synthesis of the precursor mRNA (pre-mRNA)[18]. Initiation is followed by elongation wherein the Pol II moves through the gene

extending the transcript[16]. Once the Pol II reaches the end of the gene, transcription termination occurs by release of Pol II[19]. After termination, the transcription machinery including Pol II is recycled for the next round of initiation, thereby completing the cycle[17]. During the transcription cycle, a number of essential pre-mRNA processing events occur, resulting in the mature mRNA which is exported to the cytoplasm for translation[20]. Transcription is subject to regulation at various stages which is critical for gene expression and is the key to understanding the basis of cellular differentiation, identity and maintenance. In the next section, we therefore look at the details of Pol II transcription.

1.2 Transcription by Pol II

1.2.1 Pol II features

Eukaryotic Pol II is a twelve-subunit protein complex, with a molecular weight of approximately 500,000 Da[3]. The Pol II core enzyme is formed by two large subunits, namely Rpb1 and Rpb2, that flank its positively charged active center [3]. Additionally, a mobile clamp flanks one side of the cleft, with the smaller subunits are distributed around the periphery[5,21], and the Rpb4-Rpb7 sub-complex forming the so-called stalk[5,22].

A unique feature of Pol II is the Rpb1 C-terminal domain (CTD), which consists of the consensus heptapeptide YSPTSPS repeated 26 (yeast) or 52 (human) times[5]. The CTD undergoes reversible phosphorylation at Y1, S2, T4, S5 and S7 as Pol II transcribes through a gene[23,24]. This tail-like feature of Pol II is therefore a target of multiple kinases that result in specific phosphorylation patterns through the transcription cycle[25]. The phosphorylation pattern acts as a signal for recruitment of several transcription and RNA processing factors, while the CTD itself serves as a binding platform for these factors[26].

1.2.2 Nucleotide addition cycle

Over the years, structural studies of Pol II elongation complexes (ECs) containing nucleic acids have helped shed light on how Pol II catalyzes the DNA-dependent addition of nucleotides[27–29].

In a conserved multi-step process, a nucleotide triphosphate (NTP) binds the Pol II active site that is in a pre-insertion state[28]. This NTP is then selected for based on Watson-Crick pairing with the DNA template, as well as contacts with Pol II[30]. As a result, the NTP is sandwiched between the template DNA, 3' RNA nucleotide, and the Pol II features called the bridge helix and the trigger loop[28,29]. Presence of the correct cognate base leads to nucleotide insertion to the 3' end of the pre-mRNA transcript by phosphodiester bond formation and pyrophosphate release[31,32]. Formation of the pyrophosphate bond involves a nucleophilic attack by the RNA 3'-OH on the NTP gamma-phosphate in a metal ion dependent manner[5,27].

Pol II then translocates along the DNA by one base pair, thereby emptying the active site for the next incoming nucleotide (post-translocated state)[33]. An incoming nucleotide can then bind the active site of Pol II in this state, likely locking the bridge helix and the trigger loop in that position[28]. This cycle is repeated as the Pol II goes through the DNA template, resulting in elongation of the pre-mRNA transcript[29].

1.2.3 The stages of a transcription cycle

Initiation starts with the formation of pre-initiation complex (PIC)[34]. This entails the assembly of Pol II together with its initiation factors at the promoter DNA[35]. Years of structural and biochemical work have led to a conventional PIC assembly model. First, the transcription factor (TF) II D first recognizes promoter elements leading to TFIIA mediated binding of the TATA binding protein (TBP)[36]. These promoter elements upstream of the transcription start site (TSS) serve as nucleation points for transcription initiation[37]. In metazoans, there can be additional elements present downstream the TSS[38]. This is followed by TFIIB binding, that in turn allows for Pol II-TFIIF complex recruitment[34]. Subsequently, the multi-subunit mediator complex associates with the CTD, and makes additional contacts with the rest of the Pol II body[39,40]. Finally, TFIIE, followed by TFIIH recruitment result in the complete PIC[39]. This induces rearrangements within the PIC that in turn lead to DNA unwinding, and formation of an open promoter complex and the so-called initial 'transcription bubble'[41]. This initially transcribing complex (ITC) can then use the unwound DNA and synthesize the pre-mRNA *de novo* [42].

Once the RNA reaches a certain critical length of approximately 13-14 nt [39], the initiation factors dissociate from the ITC as the upstream DNA winds back and Pol II moves downstream through the transcription bubble[42] . The ITC then transitions to an elongation complex (EC)

through an event called promoter escape[17,43]. Promoter escape is triggered as TFIIH kinase module (containing the CDK7 kinase) phosphorylates the CTD and destabilizes the PIC[39]. Mediator stimulates TFIIH activity, thereby allowing for rapid promoter escape[44].

After promoter escape, Pol II enters the early elongation phase[45]. At this stage, the essential elongation factor called 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) sensitivity-inducing factor (DSIF) binds Pol II[46]. It is partially conserved in bacteria and archaea and is a heterodimer composed of two subunits called SPT4 and SPT5[47]. SPT5 consists of an N-terminal NGN (NusG N-terminal) domain, seven Kyrpides, Ouzounis, Woese (KOW) motifs and a mobile C-terminal repeat region (CTR)[48,49]. Structures of the yeast[14] and human[50] Pol II-DSIF ECs revealed that DSIF uses its modular KOW domains to clamp the upstream DNA and the exiting RNA, while the CTR can serve as a recruiting platform for processing factors[51]. Association of DSIF with the transcription machinery is enhanced by interaction with the emerging RNA, and therefore DSIF was shown to preferentially bind ECs with RNA length of 22-24 nt[52]. DSIF likely remains bound to Pol II throughout transcription and undergoes rearrangements due to its highly modular structure to allow association of various factors downstream[53].

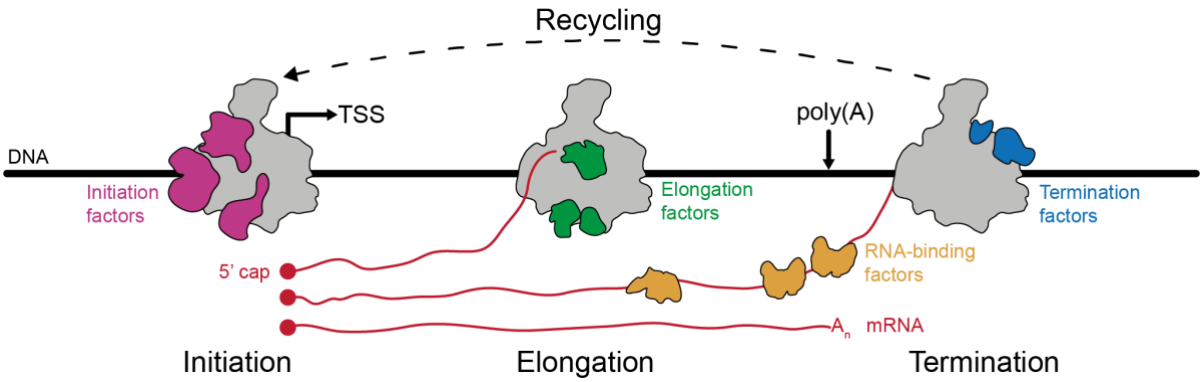


Figure 1: Pol II transcription cycle. Initiation factors (pink) assemble with RNA polymerase II at the gene promoter to initiate pre-mRNA synthesis. After the synthesis of about 10 nucleotides, the cycle enters the elongation phase during which Pol II extends the nascent transcript. Pol II-CTD phosphorylation state plays an important role in recruiting factors such as the capping machinery. The transcription machinery and the mature mRNA are released at termination and recycled for another round of transcription. Pol II in grey; DSIF in green; RNA in red; 5'cap as a red circle; $A_{(n)}$ is 3' polyadenylation; DNA as a black solid line. Adapted from M. Hantsche and Cramer 2016[54].

In metazoans, the early elongation phase is characterized by a common feature called promoter-proximal pausing[55]. It is a state where Pol II is transcriptionally competent but not actively transcribing[56]. Based on Pol II density peaks, promoter-proximal pausing occurs 50 to 80 nt downstream of the TSS and is believed to serve as the early elongation checkpoint (EEC)[57,58]. Pausing functions in regulating the development of an organism, environmental responses, and possibly acts as a kinetic window for various pre-mRNA processing events[59,60]. The two elongation factors, DSIF and the negative elongation factor (NELF), act in concert to stabilize the paused state[45,61]. NELF comprises four subunits namely NELF-A-B-C/D-E[61]. In this state, the DNA-RNA hybrid is in a 'titled' state, likely caused by post-translocation, wherein the DNA template backtracks by one position and the RNA does not, making it incompatible with elongation[62]. Moreover, NELF sterically blocks the access of TFIIS to Pol II active site which is required to rescue the backtracked Pol II and resume RNA synthesis[62,63]. We talk about TFIIS and its functions more in detail in *section 1.2.4.*

The kinase CDK9 of the positive transcription elongation factor b (P-TEFb) plays an instrumental role in overcoming the pause[64].It phosphorylates Pol II CTD, DSIF and NELF, that cause the transition to productive elongation[62]. Structural and biochemical studies suggest that this phosphorylation plays multiple roles that together lead to pause release. Firstly, NELF binding is influenced by SPT5 and NELF phosphorylation[65,66]. Secondly, it facilitates binding of Pol II associated factor (PAF) complex, which sterically competes with NELF for Pol II binding surface[67,68]. And finally, phosphorylation of the linker from Pol II to the CTD recruits elongation factor SPT6, which in turn loosens the DSIF-RNA interaction[67]. Failure to transition from the paused to the productively elongating state leads to premature termination called transcription attenuation, that involves another multi-subunit complex called the integrator complex[69].

The resulting EC bound to DSIF, PAF and SPT6 forms the activated elongation complex (EC*), which can productively elongate the pre-mRNA[67]. PAF is composed of 5 subunits Paf1, Leo1, Ctr9, Cdc73, Rtf1 and an additional subunit WDR61 in humans[70]. It is required for processive elongation by mammalian Pol II, that to a large extent is mediated allosterically by its subunit Rtf1[71,72]. Additionally, PAF is also needed for chromatin transcription, as it is implicated in histone modifications like ubiquitination and methylation, and can therefore have both positive and negative impacts on transcription[73]. Human SPT6 is a 200,000 Da conserved elongation factor that binds around the Pol II stalk[74]. It interacts with Pol II via S2

phosphorylated CTD linker, and also with DSIF KOW domains[75]. It stimulates transcription *in vitro*, *in vivo* and serves as a histone chaperone for chromatin transcription[74,76].

Once Pol II reaches the end of the gene body which is marked by a polyadenylation (polyA) site, transcription terminates with the release of the newly synthesized and partially processed RNA[77]. Termination is critical for avoiding aberrant downstream transcription into the DNA as well as Pol II recycling[78]. Termination signals can be internal, such as DNA or RNA sequences, or external relying on termination factors[79,80]. Amongst the three transcription phases, termination remains the most elusive one with limited structural or biochemical information. However, two models have been suggested in the literature. The allosteric model relies on indirect destabilization of the EC* upon polyA site encounter by 3' processing factors, whereas the torpedo model describes an active displacement of Pol II by RNA-exonuclease post RNA release[81,82]. Following Pol II dissociation upon termination, the CTD is reverted to a hypo-phosphorylated state, making Pol II transcription competent and ready to start a new cycle[80]. In the context of chromatin, recycling is facilitated by gene-looping that brings promoters and terminators in spatial proximity[19].

1.2.4 Transcription factor II S

During elongation, Pol II pausing is often induced by DNA sequence, base misincorporation or nucleosome encounters[83]. While often the Pol II can escape the pause, it can also move backward on the DNA template, in a process termed as backtracking[84]. This dislodges the 3' end of the transcript into a pore beneath the Pol II active site, leading to its arrest[84]. In order to escape this state, Pol II requires TFIIS, an RNA cleavage-stimulatory factor[85]. TFIIS stimulates the intrinsic RNA-cleavage activity of Pol II active center, which cleaves the RNA fragment in the pore resulting in the generation of a new 3' end to resume transcription[28]. The RNA cleavage activity however resides with TFIIS itself in the human system[28]. Backtracking is prevalent across the genome, especially in the 5' and the 3' ends where Pol II stalling is observed for promoter-proximal pausing or 3' end processing, respectively[86].

TFIIS was the first Pol II associated factor to be purified[87]. It is 309 amino acid (aa) in length and 33,970 Da in molecular weight[88]. TFIIS comprises three independently folding domains named I, II and III[88]. Domain I (aa 1-130), is a four-helix bundle that has been observed to interact with the initiation factors[89]. However, it is not necessary for the biochemical

activity of TFIIS, and therefore its function remains unknown[90]. Domain II (aa 131-240) forms a three-helix bundle[88]. It is connected to domain III (aa 265-309) via an unstructured interdomain linker (aa 241-264)[85]. Domain III is composed of three anti-parallel beta-sheets that form a zinc ribbon and is critical for TFIIS activity[84].

Structural studies of yeast[27,84] and human (*publication 1*)[91] Pol II-TFIIS complex have shed light on the mechanism of Pol II rescue by TFIIS and its implications in nucleosome transcription. The backtracked Pol II has approximately 10 nt in its pore[28]. TFIIS binds the Pol II funnel opening with its Domain II[84]. The interdomain linker penetrates the funnel leading up to the Pol II active center[28]. Domain III extends into the Pol II active site and positions its two highly conserved acidic residues D290 and E291 along with an activated water molecule to stimulate nucleophilic cleavage of the backtracked RNA[92]. The cleavage liberates the RNA from the pore, resetting the tilted DNA-RNA hybrid state, generating a new 3' end and a free active site for the incoming NTP[27,84]. In context of chromatin, TFIIS mediated reactivation allows repeated attempts at passing the nucleosome, therefore increasing the likelihood of overcoming the nucleosome barrier[93].

1.3 Co-transcriptional processing

Pol II nascent transcripts undergo various processing events to form the mature RNA that can be translated into proteins[77]. Some of these events are temporally coupled to transcription, and are therefore termed as co-transcriptional processes[94]. This functional coupling likely allows for efficient pre-mRNA maturation. First evidence for co-transcriptional coupling of RNA processing came from early experiments that suggested that intron removal (splicing) and transcription are spatially and temporally coupled[94,95]. The vast majority of RNA undergo three major processing events, namely, 5' capping, splicing and 3' end processing before their release[96]. Dynamic CTD phosphorylation patterns, and an intricate web of interactions between the transcription machinery and the pre-mRNA processing factors is responsible for ensuring efficient mature mRNA biogenesis[97].

Capping is an early co-transcriptional processing event, and a characteristic feature of Pol II transcripts[98]. It involves the addition of a 7-methyl guanosine cap structure to the RNA 5' end resulting in a cap(0) sructure (5' m^7GpppXRNA, where X is the first transcribed nucleotide)[98]. In all human mRNAs, the 2'O of the first transcribed nucleotide is further methylated (5' ^{m7}GpppX$m_{O\text{-}2}$pRNA) resulting in a cap(1) structure[99]. As co-transcriptional

capping is the focus of this dissertation, the current literature and details of capping are extensively covered in *section 1.4* and *section 1.5*.

Pre-mRNA splicing is essential for gene expression and involves the removal of non-coding introns and ligation of the coding regions of the gene called exons[100]. It is catalyzed by an assembly called the spliceosome, that comprises of U1, U2, U4/6 and U5 small nuclear ribonucleopreotein particles (snRNPs)[101]. Splicing occurs in a two-step trans-esterification mechanism[101]. In the first set of steps, U1 snRNP binds and cleaves at the 5' end of the exon, following which the U2 snRNP attaches to a conserved branch point sequence just upstream of the 3' splice site adjacent to the exon[102]. U5 and U4/U6 then bind to form a pre-catalytic complex which allows for the recruitment of U6 to the 5' splice site and the branch point [103]. The 5' splice site is then cleaved exposing a free 3'OH group and forming an intron lariat[103]. In the second step, U5 positions the 5' and the 3' end together by base pairing with both, leading to exon fusion, and excision of the lariat[103]. Spliceosome can then disassemble, and components recycled[104]. Splicing often begins co-transcriptionally and continues post-transcriptionally[105]. Recent structural studies have suggested a growing loop model, which suggests that while the spliceosome is assembled co-transcriptionally, the catalytic steps may occur independent of Pol II[106].

The 3' end processing often entails the synthesis of a 3' polyA signal, that in humans can be up to 1500 bp upstream of the termination site[107]. It confers stability to the mRNA and is required for nuclear export[108]. Trans-acting proteins like cleavage/polyadenylation-specificity factor (CPSF) and cleavage stimulation factor (CstF) cut the pre-mRNA approximately 10-30 nt downstream of the polyA site prior to polyadenylation[109]. Subsequently, additional termination factors that tether to the CTD and include an exonuclease are recruited, which facilitate the release of the mRNA from the EC[110].

1.4 mRNA capping

Occurring during early elongation, capping is the first co-transcriptional event, and a unique feature of Pol II transcripts[107]. Conserved from viruses to humans, capping is a key modification of essential biological importance that serves several functions[111]. It plays a critical role in mRNA stability, facilitates maturation and export, promotes translation initiation, and serves as a hallmark of 'self-RNA'[112]. Details of these individual functions are

described in *section 1.4.1*. Mutating the capping apparatus and thereby disrupting the cap addition has experimentally been shown to be lethal in yeast, confirming its diverse roles and its importance in cell viability[113]. Figure 2 shows a conventional eukaryotic capped pre-mRNA.

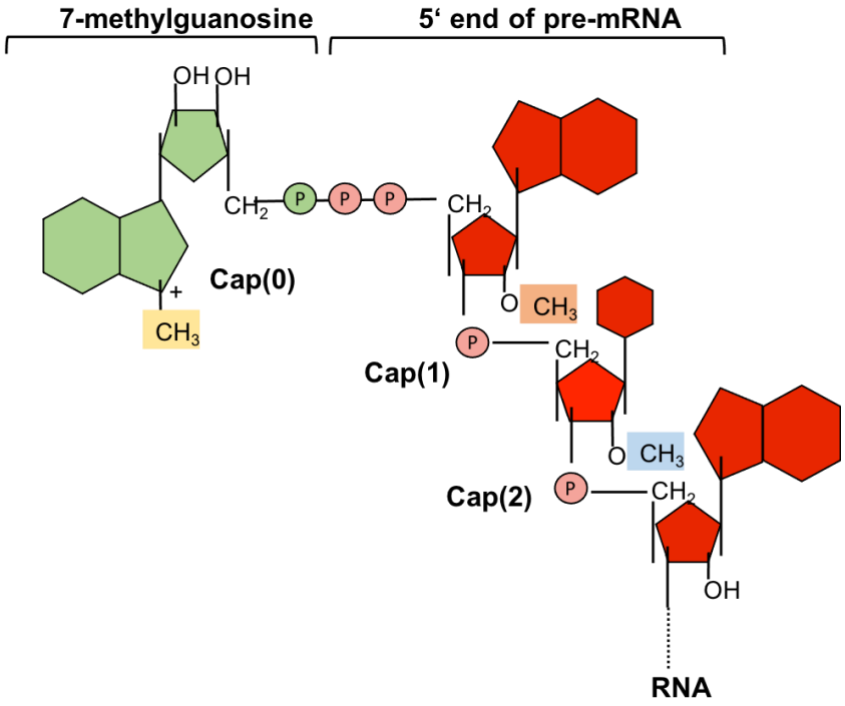


Figure 2: Structure of the RNA cap. Cap(0) is added as a 7-methylguanosine modification shown in green. Cap(1) and cap(2) additionally include a methylation of the 2’ OH of the first (orange) and the second (blue)transcribed nucleotide, respectively. Pre-mRNA elements are shown in red.

Formation of the cap occurs in several catalytic steps and requires the activity of series of enzymes[43]. The catalytic steps and the enzymes involved are described in the sections below. However, the eukaryotic capping process however may not always reach completion[114]. In mammals, protein belonging to the DXO (decapping exoribonuclease)/Dom3Z and XRN families function as quality control systems of the 5’ cap[115]. For instance, the decapping exonuclease DXO recognizes and digests the partially capped but unmethylated transcripts[115].

1.4.1 Importance of the 5' cap

A 5' cap modification is indispensable for the growth and viability of cells[113]. Over the years, biochemical and structural studies of the capping enzymes (CEs) together with protein complexes and/or RNA substrate have revealed a diverse range of functions.

5' end protection. Capping stabilizes the transcript by protecting the emerging 5' pre-mRNA end and later the mature mRNA by shielding the 5' end and thereby blocking the 5'-3'exonuclease attack on the mRNA[116].

Pre-mRNA splicing and nucleocytoplasmic export. Cap(0) has been shown to play an important role in efficient pre-mRNA splicing in mammalian[117] and yeast[118] systems. It facilitates this process by helping in recruitment and binding of the nuclear cap-binding complex (CBC)[119]. CBC is a heterodimer composed of CBC80 and CBC20 that recognizes and binds the cap co-transcriptionally[120]. The resulting cap-CBC complex is critical for efficient interaction and binding of the 5' splice site to the U1snRNP[106]. It also is critical for mRNA assembly into a ribonucleoprotein particle (RNP) which likely contains signals for export to the cytoplasm through the nucleopore complex[101]. Additionally, it is required for export of the spliceosome machinery, that is capping of the U snRNAs[121]. The CBC in turn promotes downstream processing, nucleocytoplasmic export, and other critical cellular functions[122].

mRNA turnover. The level of eukaryotic gene expression is defined by the mRNA synthesis[114]. On the other hand, de-capping determines the mRNA half-life[123]. An mRNA that lacks the cap is rapidly degraded by 5'-3' exonucleases[112]. Cap removal is a critical step in the 5'-3' mRNA degradation and can be triggered for various reasons[124]. This suggests that the interplay between synthesis of capped mRNA and de-capping to degrade the mRNA allows fine-tuning and control of gene expression.

Innate immunity. One of the most important roles of a capped mRNA is in differentiating the cellular transcripts from the foreign ones[125]. In humans, cap(1) methylation is present on all mRNA molecules and allows the cell to discriminate against foreign RNAs and generate an innate immune response[126]. In response to viral infections type I interferon (IFN) response is triggered by cytoplasmic pattern recognition receptors (PRRs) like retinoic acid inducible gene-I (RIG-I)[127] and MDA5[128]. RIG-I was shown to recognize the 5'GpppXRNA and cap(0) RNA with similar affinity, while the presence of cap(1) significantly diminished the recognition[129]. Additionally, when the innate immune response is triggered, IFN induces the expression of IFN-induced proteins with tetratricopeptide repeats (IFIT proteins) which inhibit

the translation of mRNAs lacking cap(1) modification[130]. Finally, IFN induce the transcription of certain IFN stimulated genes (ISGs), including cap methyltransferase 1 (CMTR1 or also called as ISG95), which is the enzyme responsible for formation of cap(1) in humans[131]. CMTR1 in turn was shown to promote the rapid protein expression of other ISGs, without affecting their mRNA levels[132]. Taken together, these results suggest that CMTR1 itself and its 2'O methylation activity help establish an antiviral cellular state.

1.4.2 Capping pathways

Capping involves the addition of an m^7G linked via an inverted 5'-5' triphosphate bridge to the first nucleotide of the mRNA[133]. This minimal m^7GpppX-RNA is often referred to as the cap(0). In higher eukaryotes, cap(0) is methylated at ribose 2'OH of the first nucleotide, giving rise to $m^7GpppXm_{O\text{-}2}$-RNA, or the cap(1) RNA[116]. Cap(1) is the hallmark of cellular RNAs, whereas cap(0) is recognized by the innate immune system as foreign[125]. Approximately half of the capped RNAs undergo another methylation at the 2'OH of the second transcribed nucleotide, resulting in the cap(2) structure[128]. Further modifications of the cap are also possible, such as methylation at N6 position when the first nucleotide is an adenine ($m^7Gppp^{N\text{-}6}Am_{O\text{-}2}$-RNA)[134].

Conventional pathway. Three enzymatic activities are responsible for generation of the cap(0) 5' end: an RNA triphosphatase (TPase), a guanylyltransferase (GTase), and a methyltransferase (MTase)[98]. In higher eukaryotes, cap(0) often serves as a substrate for further modifications, which includes a series of methylations[135]. The most critical modification present on all human mRNA transcripts is the 2'-O methylation to form the cap(1) mRNA[136].

Eukaryotic and almost all viral capping is accomplished in three consecutive catalytic steps to achieve the minimal cap(0)[116]: i) hydrolysis of the 5' γ-phosphate (5'pppX) of pre-mRNA to a diphosphate (5'ppX) by an RNA 5' triphosphatase (TPase)[137]; ii) transfer of a guanosine monophosphate nucleoside (GMP) moiety to the RNA diphosphate terminus (5'GpppX) by an RNA guanylyltransferase (GTase). This is mediated in a two-step reaction that involves the formation of a covalent GMP-enzyme intermediate followed by transfer of the GMP moiety and pyrophosphate (PPi) release[138]; iii) methylation at guanine-N7 position by RNA methyltransferase to produce the 7-methylguanosine cap ($^{m7}GpppX$)[139]. In higher eukaryotes, cap(1) is the most prevalent 5' end cap structure which requires a fourth step, iv)

methylation of the 2'OH ribose of the first nucleotide by cap methyltransferase (2'O MTase)[140]. S-adenosylmethionine (SAM) usually acts as a methyl donor and is converted to S-adenosylhomocysteine (SAH)[98]. The reactions can be summarized by as follows:

i. 5'pppX-RNA $\xrightarrow{\text{TPase}}$ 5'ppX-RNA + Pi

ii. GTase + GTP $\xrightarrow{\text{GTase}}$ GTase-GMP + PPi

iii. GTase-GMP + 5'ppX-RNA $\xrightarrow{\text{GTase}}$ 5'GpppX-RNA + GTase

iv. 5'GpppX-RNA + SAM $\xrightarrow{\text{N7MTase}}$ 5' m^7GpppX-RNA + SAH

v. 5'GpppX-RNA + SAM $\xrightarrow{\text{2'O MTase}}$ 5' $^{m7}GpppXm_{o\text{-}2}RNA$ + SAH

Unconventional pathways. The first evidence of unconventional capping pathways came from viral studies in the 1970s. It has since been shown that some viruses can synthesize RNA caps identical to the cellular cap using completely different mechanisms[141]. There are three well-studied pathways: capping with GDP (*rhabdoviruses*), cap snatching (*influenza virus*) and capping with m7GMP (*alphavirus*)[142]. For instance, E.Decroly et al., in their review describe the positive-sense single stranded (ss +) RNA alphavirus synthesizing a cap(0) using m^7GMP. The first step involves a TPase yielding a diphosphate RNA by cleavage of the γ-phosphate, and a methylation of a GTP molecule at N7 position by a guanine-N7-MTase. GTase then binds the N7-methylated-GTP covalently using a histidine, followed by the transfer of m^7GMP molecule to the diphosphate end. This results in the formation of cap(0), identical in structure to conventionally capped RNA, albeit achieved by a different path[141]. Divergent pathways converging to a consensus structure suggests that there is high selective evolutionary pressure to maintain the cap structure.

1.5 Capping enzymes

Two landmark studies in 1974 and 1976 revealed the function and mechanism of capping[133,143]. It has since been shown that, while the steps required to achieve capped RNA remain evolutionarily conserved from virus to eukaryotes, the enzymes responsible for cap

addition differ drastically in composition and structure[144]. In the following sections I discuss the CEs involved in the human cap(1) formation. I will focus on the structural aspects of these enzymes and compare them to well-studied CEs from other species. Figure 3 provides a summary of the human enzymes responsible for cap(1) formation, focusing on their activity and the isolated structures available in literature.

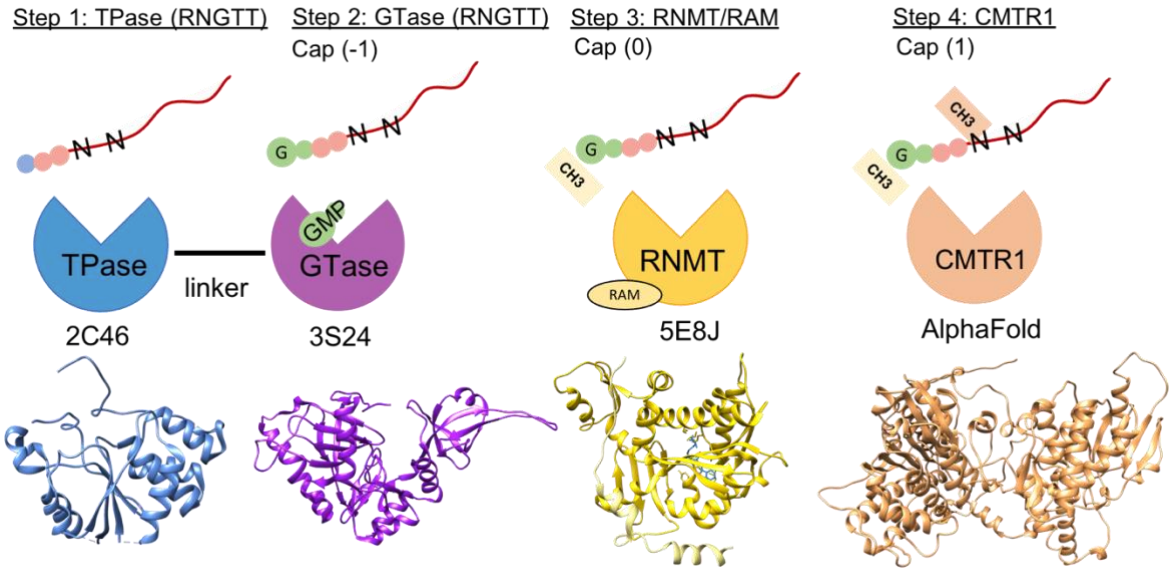


Figure 3: Structures of the human capping enzymes. Addition of the 5' cap is shown as a cartoon representation along with the enzymes responsible for each step. Crystal structures of the RNGTT (shades of blue) and RNMT/RAM (shades of yellow), as well as the predicted AlphaFold model of CMTR1 (orange) is shown.

1.5.1 RNA guanylytransferase and 5' phosphatase (RNGTT)

RNGTT (also referred to as mRNA capping enzyme or Mce1) catalyzes the first two steps of the capping process in metazoans[98]. The N-terminus harbors the TPase activity, while the C-terminus carries the GTase activity, and the two are linked via a flexible linker[145]. In yeast, for instance, TPase and GTase activities are harbored by two separate polypeptides, Cet1 and Ceg1, respectively,[146] which interact together to form a heterodimer[23].

RNA triphosphatases. Two distinct families of RNA TPases exist: metal-dependent (lower eukaryotes like fungi and protozoa) which are sub-classified into three groups, and metal independent (nematodes, metazoan, and plants)[126]. Metazoan and plant TPase domain contain a signature HCxxxxxR(S/T) motif[137]. This motif is a characteristic feature of the cysteine

phosphatase superfamily and catalyzes the removal of 5' phosphate in a ping-pong mechanism of action[142]. First, a nucleophilic attack by active cysteine results in a covalent cysteinyl-S-phosphoester intermediate, and second, the covalent intermediate is hydrolyzed to release inorganic phosphate[142].

X-ray crystallography of the RNGTT TPase domain (1-210 aa) by A. Changela et al., has shed light on the structure and mechanism of this RNA-specific bifunctional enzyme. It consists of five ß-strands, which form a central twisted sheet. These are flanked by two and four α-helices on either side. The signature motif comprises a phosphate binding loop (P-loop), and together with other secondary structure elements forms a deep positively charged pocket. This is required to interact with the negatively charged phosphates of the RNA and guide it to the active site. Additionally, they showed that Arg132 is indispensable for the activity. The nucleophilic cysteine 126 (C126) is situated within the P-loop at the base of this cavity[137].

When compared to its yeast counterpart, Cet1, which is metal-dependent, no structural similarity between the two is observed. Cet1 is biochemically and structurally most similar to the metal-dependent TPases encoded by DNA viruses[147]. These TPases belong to the sub-category called tunnel metalloenzyme family, with a triphosphate group as the preferred substrate; the crystal structure of Cet1 from budding yeast revealed a homodimeric arrangement, and a central eight-stranded anti-parallel ß-barrel (tunnel)[137]. The active site is located at the center of the tunnel which is also responsible for coordinating the metal ion[146]. Mutating conserved tunnel residues in *Saccharomyces cerevisiae* has shown to been lethal[148].

RNA guanylytransferase. GTases usually function in a reversible two-step ping-pong reaction as described in *section 1.4*. Structural studies of the various GTases have highlighted a bipartite domain architecture which comprises an N-terminal nucleotidyltransferase (NTase) domain, and a C-terminal oligonucleotide-binding (OB) domain[149]. The active site lysine lies within a characteristic Kx(D/N)G motif[149]. RNGTT GTase domain crystal structure (PDB code 3S24) revealed that it comprises seven α-helices and fifteen β-strands which are distributed to have an overall GTase/DNA ligase fold. The NTase has an ATP grasp fold. The active site lysine (Lys294) is buried in a positive cleft formed by the interface of the two sub-domains [138]. A general mechanism for the GTases has therefore emerged from these studies. First, magnesium and GTP bind an 'open' conformation, and induce domain 'closure' which facilitates the formation of the enzyme-(lysyl-N)-GMP intermediate. Second, GTP cleavage results in return to an 'open' conformation and release of pyrophosphate. Transfer of GMP to

diphosphate RNA substrate and release of the partially capped RNA require an additional round of closure and opening, respectively[142]. This results in a partial cap structure 5'GpppXRNA, which we annotate here as cap(-1).

RNGTT shares sequence similarity with DNA and RNA ligases that are dependent on adenosine triphosphate or nicotine adenine dinucleotide and belong to the covalent NTase superfamily[149]. It also has high sequence identity (98%) with the other mammalian GTases[138]. On the other hand, it shares low sequence similarity with the *Chlorella* virus GTase PBCV-1[150] and the *Candida albicans* GTase Cgt1[149]. Despite the low sequence similarity, large structural similarity between the PBCV-1 GTase and RNGTT is observed when the structures are superimposed. Comparison revealed high structural conservation for the GTP binding site, and the elements that form the hydrophobic core active cleft, especially with PBCV-1[138,150]. However, large variations are observed in the OB-fold region (462-552 aa)[138]. While high structural similarity exists with the viral GTase, the yeast *C.albican* and *S.cerevesiae* structures showed large deviation in shape and charge distribution[144]. Consequently, while the overall fold of the enzyme is evolutionarily conserved, distinct surface and charge distribution could reflect differences in pre-mRNA capping process between mammalian and unicellular eukaryotic systems[138].

1.5.2 mRNA guanine N7 methyltransferase (RNMT)

Human RNMT is a 476 aa protein that comprises an N-terminal regulatory domain (1-120 aa) and a C-terminal catalytic domain (121-476 aa)[135]. The N-terminus mediates the recruitment to TSS[139]. An activating subunit of RNMT, called RNMT-activating miniprotein (RAM) consists of an N-terminal RNMT activation domain (1-55 aa), an RNA-binding domain (56-90 aa), and a C-terminal nuclear localization domain (91-118 aa)[151]. It is only present in vertebrates, and residues 2-45 comprise the minimal unit that binds the RNMT catalytic domain to enhance its activity[152].

Human RNMT structure has a high structural similarity to the *Encephalitozoon cuniculi*, cap methyltransferase, Ecm1 (PDB code 1RI5). These MTases display a typical class I MTase Rossmann fold motif (RFM), a conserved VLxI/LxxGxGxDL motif, and a deep SAM/SAH binding pocket[153]. RNMT differs from Ecm1 in additionally possessing the RNMT lobe, which is also present in its yeast (*S. cerevisiae* Abd1 and *S.pombe* Pcm1) and viral (Vaccinia virus D1)

counterparts. Additionally, it differs also in the length of helix A (170-194 aa), and some other loops[154].

Structural study of RNMT-RAM by D.Varshney and colleagues revealed that the negatively charged residues of RAM (35-45 aa) interact with the positively charged groove of RNMT. This groove comprises RNMT α-helix hinge (319-415 aa) and the lobe. They showed that the binding results in optimally orienting critical substrate binding residues in RNMT active site. They propose that it does so by stabilizing the RNMT lobe and the α-helix hinge, thereby orienting the active site helix A in a more favorable position for substrate binding and methylation[155]. Comparison between the viral and human MTase complexes reveals parallel modes of action between RAM and Vaccinia virus D12, in that they both allosterically activate their cognate RNMT[156]. An evolutionary advantage of such an allosteric system might be that controlling the RAM expression could allow the cells to regulate cap methylation levels.

1.5.3 mRNA 2' O cap methyltransferase 1 (CMTR1)

CMTR1 is a 95,300 Da protein composed of 835 aa[140]. CMTR1 as the human 2'O MTase was only identified recently by F.Belanger and colleagues[140]. It is a multi-domain protein that comprises a disordered N-terminus containing a nuclear-localisation signal (2-19 aa), a G-patch domain (85-133 aa), an MTase domain (170-550 aa), a non-functional GTase-like domain (560-729 aa) and a WW domain (755-790 aa)[157]. Similar to RNMT, the CMTR1 MTase domain has an RFM fold[154]. However, in contrast, CMTR1 relies on a conserved catalytic KDKE triad which is responsible for optimally positioning the 2'OH of the base, allowing a nucleophilic attack on SAM methyl group[158].

The human CMTR1 structure by Smietanski M et al., revealed the characteristic RFM domain core, and the peripheral regions that resembled the other cap modifying MTases like the vaccinia virus VP39[157]. They also showed that the RNA binds in a deep pocket, and that the active site does not interact with RNA bases, suggesting a sequence independent nature for the reaction[157].

Deletion of the C-terminal GTase-like and WW domains has been shown to reduce CMTR1 activity in vitro, which suggests its role in either substrate recruitment or structural stabilization[140]. The WW domain has been implicated in interacting with the S5 phosphorylated CTD of Pol II[140]. Although CMTR1 requires a cap(-1) structure, it methylates cap(-1) and cap(0) equally, and is therefore independent of the N-7 methylation[140].

Additionally, CMTR1 has been observed to be enriched at TSS while RNMT not[159]. This raises the possibility that CMTR1 may act prior to RNMT.

1.5.4 Co-transcriptional capping

Over the years, several studies on cap addition have been conducted for comprehensive structural and biochemical characterization of CEs from different taxonomical orders. Biochemical studies have revealed remarkable features of the capping process such as the selectivity of CEs for Pol II transcripts and the high efficiency of co-transcriptional capping in comparison to that of free RNA[24,160]. Co-transcriptional capping has been suggested to occur between initiation and early elongation, and as soon as the emerging RNA becomes available, around 20-25 nt[161].The stimulation of capping during early elongation is likely two-fold: Pol II dependent (CTD and direct interactions) and Pol II independent (elongation and pausing factors) mechanisms[162]. Additionally, capping has been proposed to serve as a quality check for the mRNA, thereby preventing the extension of uncapped transcripts[163].

Co-transcriptional capping has been proposed to require additional factors that facilitate the recruitment and the stimulation of the CEs. Discovery of selectivity of the CEs for the Pol II transcripts directed the focus towards studying the inherent features of Pol II, such as the CTD[149]. Pol II CTD was found to be essential for pre-mRNA processing and recruitment of the CEs coincided with its Ser2 and/or Ser5 phosphorylation[164]. The CDK7 kinase subunit of TFIIH phosphorylates the CTD at these positions[149]. Several models based on biochemical characterization have been proposed for the recruitment of capping factors. The tethering model is based on the fact that CEs can stably bind to phosphorylated CTD heptad and hypothesizes that it acts as a scaffold for tethering the CE to elongating Pol II[165]. Second, the allosteric model is based on evidence that CTD repeats phosphorylated at Ser5 and not Ser2 increases the formation of covalent Pol II-CE intermediates, and allosterically activate the CE[166]. Both RNGTT and RNMT recruitment has been shown to rely on CTD phosphorylation[139,167]. Crystal structures of individual CEs with the CTD have confirmed the importance of these interactions[113]. However, studies have also shown that inhibiting CTD phosphorylation causes only a modest reduction in capping and that CTD alone is not sufficient for efficient capping[168]. For instance, the yeast CE interacts with both the CTD and the Pol II Rpb1 foot domain[169].

Stimulation of the capping process additionally involves other factors. RNGTT has been shown to bind to transcription factors such as the Spt5 subunit of DSIF[170]. Studies have shown that under limited RNGTT conditions, DSIF stimulates capping[170,171]. In the same study, increasing TPase domain amounts was shown to abolish the stimulation by DSIF[170]. Therefore, while DSIF plays a generally important role during elongation, its precise role in capping is unclear.

DSIF binds Pol II and recruits NELF during early elongation, which induces promoter-proximal pausing during which capping likely occurs[62]. Studies show that the recruitment of RNGTT can relieve the inhibition by NELF, and that the TSS proximal mRNAs are uncapped, while those that are distal are largely capped[172]. This has given rise to a 'capping checkpoint model' which ensures that only the transcripts that are properly capped are subjected to productive elongation[163]. Therefore, implying that pausing ensures recruitment of CEs, which in turn recruits pause release factors upon signal for complete cap[108]. In yeast however, NELF is absent and there is no promoter-proximal pausing which means that the timing of cap addition and its regulation is likely to differ from that of metazoans[160].

Regarding methylation of the cap, chromatin immunoprecipitation sequencing (ChIP-seq) studies have suggested RNMT occupancy throughout the gene length[46]. RNMT along with its activating subunit RNMT-activating mini protein (RAM) has also been observed to stimulate Pol II-dependent transcription independent of capping and was shown interact with transcription elongation components such as SPT6 and PAF[153]. It however remains unknown how the elongation factors stimulate capping, when methylation occurs and if it serves other functions.

Finally, structural evidence for co-transcriptional capping machinery came from a yeast study by F.Martinez-Rucobo et al. They solved the yeast Pol II-CE complex and revealed that the capping machinery docks onto the Pol II wall, positioning the CE active site at the RNA exit tunnel. While the study was limited due to resolution, it gave important topological insights into the mechanism of co-transcriptional capping[169].

Mammalian capping pathway is catalytically fairly well understood owing to being a subject of intense biochemical investigation. However, the timing, the order of recruitment of CEs and their structures in complex with the transcription machinery still remain elusive.

1.6 Scope of this dissertation

Over the last years, advances in single particle cryo-EM, have allowed high-resolution structural studies of the pre-initiation complex[39], the elongation complex[50], the paused complex[62], and the activated elongation complex[67]. However, our structural and mechanistic understanding of fundamental processes like Pol II transcription through a nucleosome and human co-transcriptional capping remain limited.

In ***manuscript 1*** *(published)*, we aim to answer some open questions regarding the passage of Pol II transcription through nucleosomes with emphasis on the role of elongation factors. In particular, the role of TFIIS in rescuing backtracked Pol II has been well studied and described in *section 1.2.4*. We further investigated how TFIIS coordinates Pol II rescue with the EC* and the elongation factors to help transcription through the nucleosome.

We proposed a three-step model that explains Pol II passage through nucleosome with the help of TFIIS. This was based on: structures of the Pol II EC*-nucleosome (Farnung, L.) Pol II EC*-TFIIS (Vos, S.) and the Pol II EC-TFIIS (Garg, G.), and biochemical data revealing the efficiency of Pol II transcription through the nucleosome in the presence of various factors (Ochmann, M.). While the mammalian Pol II EC-TFIIS complex showed similar binding interface as in the previously resolved yeast structures, it did also reveal some differences[28]. For instance, the Pol II crevice is in a pre-opened state in mammalian system, and therefore TFIIS binding does not induce any further opening as in the yeast system. Additionally, the trigger loop and TFIIS domain III remain mobile in the human system as opposed to a locked trigger loop and stabilized domain III in the yeast, potentially having implications in TFIIS entry regulation.

In ***manuscript 2*** *(in- preparation)*, we cover the major focus of the dissertation, which is to understand the molecular mechanisms involved in human co-transcriptional capping. In the recent years, single-particle studies of the human transcription complexes have shed light on how Pol II interacts with a large number of factors in the different stages of transcription. Figure 4 illustrates some of these structures and emphasizes on the minimal structural understanding of co-transcriptional capping.

Biochemical and structural advances in the field have proposed that capping serves as an early elongation check point sincce failure to cap the nascent transcript results in pre-mature transcription termination at the promoter-proximal region[108]. Given its importance in

the cell, viruses have also evolved various capping mechanisms for efficient translation and to escape host immune responses[148]. Recent work by Martinez-Rubico et al., sheds light on the yeast co-transcriptional capping machinery, and highlights the importance of this temporal coupling[169]. However, the study does not include elongation factors and mammalian CEs differ vastly from their yeast homologs. Structural studies with the mammalian co-transcriptional capping complexes are therefore a critical missing link in the understanding of this fundamental process in humans.

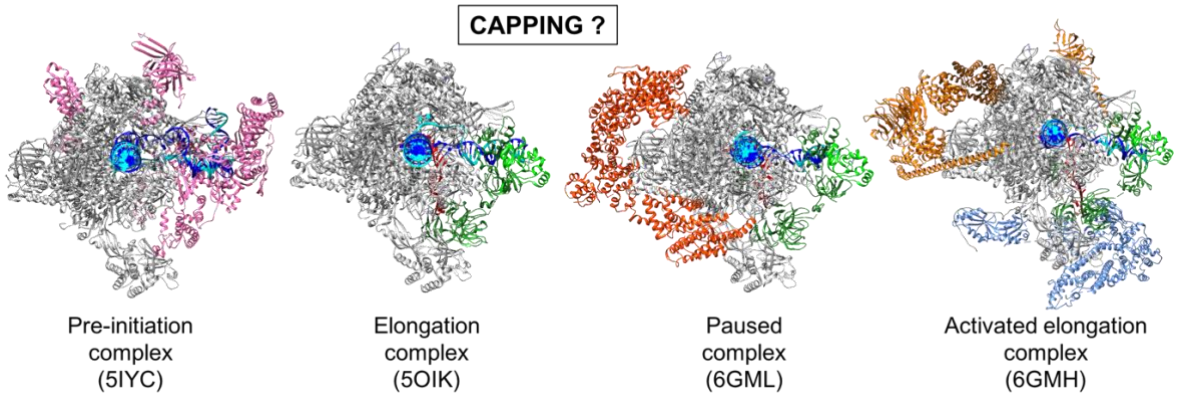


Figure 4: Structural depiction of the transcription stages. Over the years, a concerted effort to capture structural intermediates of the transcription stages has been made. However, capping which occurs between during the early elongation stage is still a prominent missing link in the picture. The PDB code for each structure is indicated in brackets.

Through this dissertation, I structurally answer major long-standing questions in the field. How does the 5' pre-mRNA end move between the multiple active sites? How does the capping machinery interact with the actively transcribing Pol II? What role do elongation factors play? And finally, taken together, how are these steps of capping coordinated with transcription? The answers to these questions are discussed in detail in *section 4.* In summary, I propose a model for human co-transcriptional capping which entails the following steps. First, RNGTT is recruited to the Pol II surface, with the TPase domain binding the RNA and the GTase domain anchoring it to the stalk. Upon γ-phosphate cleavage, TPase loses its affinity, and GTase domain swings-in to displace the TPase domain as it guanylates the diphosphate end RNA, resulting in a cap(-1) structure. The RNA is extended further until it can be methylated. Around 29 nt, CMTR1 can dock onto the stalk with its C-terminal, while its MTase

domain binds and adds a methyl group to the RNA. Large rearrangements in DSIF are observed throughout, exemplifying its modular nature, which are likely required for RNGTT binding and activity.

Chapter 2: Manuscript I

2 Structure of a backtracked hexasomal intermediate of nucleosome transcription[91]

Lucas Farnung[1,2*†], Moritz Ochmann[1*], Gaurika Garg[1*], Seychelle M. Vos[1,3*], Patrick Cramer[1†]

[1] Max Planck Institute for Multidisciplinary Sciences, Department of Molecular Biology, Am Fassberg 11, 37077 Göttingen, Germany.

[2] Current address: Department of Cell Biology, Blavatnik Institute, Harvard Medical School, Boston 02115, MA, USA

[3] Current address: Massachusetts Institute of Technology, Department of Biology, Cambridge 02142, MA, USA

*These authors contributed equally.

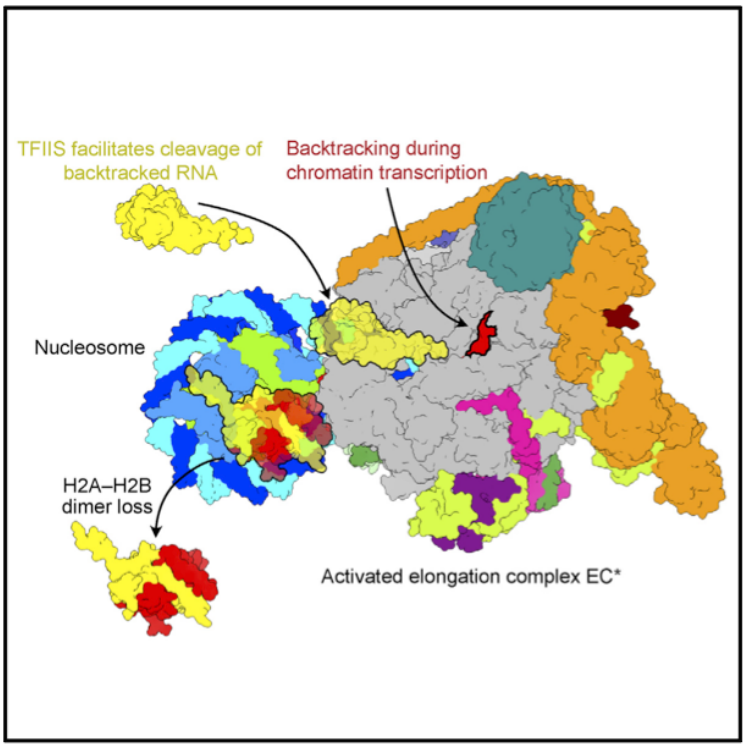
[†]Correspondence: lucas_farnung@hms.harvard.edu, patrick.cramer@mpinat.mpg.de

Molecular Cell

Article

Structure of a backtracked hexasomal intermediate of nucleosome transcription

Graphical abstract



Authors

Lucas Farnung, Moritz Ochmann, Gaurika Garg, Seychelle M. Vos, Patrick Cramer

Correspondence

lucas_farnung@hms.harvard.edu (L.F.), patrick.cramer@mpinat.mpg.de (P.C.)

In brief

Farnung et al. show that the transcription factors DSIF, SPT6, PAF1 complex, and TFIIS enable efficient nucleosome passage. They provide cryo-EM data that show H2A–H2B dimer loss and RNA polymerase II backtracking during nucleosomal traversal. They demonstrate how TFIIS stimulates the cleavage of the backtracked RNA to facilitate nucleosome passage.

Highlights

- SPT6, DSIF, PAF1 complex, and TFIIS stimulate transcription through chromatin
- Nucleosome traversal by RNA Pol II results in H2A–H2B dimer loss
- RNA Pol II can backtrack on nucleosomal substrates
- Cryo-EM reveals mammalian RNA Pol II-TFIIS complexes

Farnung et al., 2022, Molecular Cell *82*, 1–9
September 1, 2022 © 2022 Elsevier Inc.
https://doi.org/10.1016/j.molcel.2022.06.027

CellPress

Please cite this article in press as: Farnung et al., Structure of a backtracked hexasomal intermediate of nucleosome transcription, Molecular Cell (2022), https://doi.org/10.1016/j.molcel.2022.06.027

Molecular Cell

CellPress

Article

Structure of a backtracked hexasomal intermediate of nucleosome transcription

Lucas Farnung,[1,2,4,*] Moritz Ochmann,[1,4] Gaurika Garg,[1,4] Seychelle M. Vos,[1,3,4] and Patrick Cramer[1,5,*]
[1]Max Planck Institute for Multidisciplinary Sciences, Department of Molecular Biology, Am Fassberg 11, 37077 Göttingen, Germany
[2]Present address: Department of Cell Biology, Blavatnik Institute, Harvard Medical School, Boston, MA 02115, USA
[3]Present address: Massachusetts Institute of Technology, Department of Biology, Cambridge, MA 02142, USA
[4]These authors contributed equally
[5]Lead contact
*Correspondence: lucas_farnung@hms.harvard.edu (L.F.), patrick.cramer@mpinat.mpg.de (P.C.)
https://doi.org/10.1016/j.molcel.2022.06.027

SUMMARY

During gene transcription, RNA polymerase II (RNA Pol II) passes nucleosomes with the help of various elongation factors. Here, we show that RNA Pol II achieves efficient nucleosome passage when the human elongation factors DSIF, PAF1 complex (PAF), RTF1, SPT6, and TFIIS are present. The cryo-EM structure of an intermediate of the nucleosome passage shows a partially unraveled hexasome that lacks the proximal H2A–H2B dimer and interacts with the RNA Pol II jaw, DSIF, and the CTR9trestle helix. RNA Pol II adopts a backtracked state with the RNA 3′ end dislodged from the active site and bound in the RNA Pol II pore. Additional structures and biochemical data show that human TFIIS enters the RNA Pol II pore and stimulates the cleavage of the backtracked RNA and nucleosome passage.

INTRODUCTION

Eukaryotic transcription occurs within chromatin, but it remains unclear how RNA Pol II passes through nucleosomes (Chang et al., 2014; Petesch and Lis, 2012; Teves et al., 2014). Studies of nucleosome passage by RNA Pol II have been carried out *in vivo* and *in vitro* and have explained many aspects of this intricate process (Kujirai and Kurumizaka, 2020; Studitsky et al., 2016; Teves et al., 2014). These studies showed that RNA Pol II enters the nucleosome and stalls when it reaches superhelical location (SHL) −1, which is located just in front of the nucleosome dyad, both *in vivo* (Weber et al., 2014) and *in vitro* (Bondarenko et al., 2006; Ehara et al., 2019; Kujirai et al., 2018). To overcome the nucleosomal barrier, RNA Pol II requires the help of elongation factors, including DSIF, PAF (Guermah et al., 2006; Lai and Pugh, 2017; Venkatesh and Workman, 2015), TFIIS (Guermah et al., 2006; Kettenberger et al., 2003), and SPT6 (Žumer et al., 2021).

RNA Pol II passage through a nucleosome can displace one of the two histone H2A–H2B dimers from the histone octamer, generating a transient hexasome that contains one H2A–H2B dimer and one H3-H4 tetramer (Cole et al., 2014; Jamai et al., 2007; Kimura and Cook, 2001; Kireeva et al., 2002; Ramachandran et al., 2017). Recent results in *Drosophila* indicate that the +1 nucleosome lacks either the proximal or the distal H2A–H2B dimer, forming a hexasome (Ramachandran et al., 2017). The proximal H2A–H2B dimer is first encountered by RNA Pol II during transcription. Displacement of the proximal H2A–H2B dimer depends on RNA Pol II elongation (Bortvin and Winston, 1996; Ramachandran et al., 2017) and is associated with RNA Pol II stalling (Kulaeva et al., 2009), whereas loss of the distal H2A–H2B dimer facilitates nucleosome passage, both *in vitro* (Kulaeva et al., 2009) and *in vivo* (Ramachandran et al., 2017). Histones are retained locally upon RNA Pol II passage with the help of the histone chaperones SPT6 and FACT (Jeronimo et al., 2019; Kaplan et al., 2003).

Understanding the mechanisms underlying RNA Pol II passage through nucleosomes requires structural information. Structures of elongating yeast RNA Pol II enzymes in complex with a nucleosome showed that the incoming nucleosome engages with the RNA Pol II clamp head and lobe domains that are located on opposite sides of the polymerase cleft (Farnung et al., 2018; Kujirai et al., 2018). Similar orientations of the nucleosome were observed at different RNA Pol II stalling sites within the nucleosome (Kujirai et al., 2018). Structures of yeast RNA Pol II in the presence of the elongation factors Spt4/Spt5 (corresponding to human DSIF) and Elf1 showed a similar orientation of the nucleosome and suggested that the factors prevented trapping of the nucleosome by RNA Pol II (Ehara et al., 2019). These pioneering studies (Ehara et al., 2019; Farnung et al., 2018; Kujirai et al., 2018) did however not include additional elongation factors that are critical for nucleosome passage.

Here, we provide the structure of a nucleosome-transcribing RNA Pol II intermediate that contains these critical elongation factors. For the first time, we observe directly that nucleosome transcription can coincide with loss of the H2A–H2B dimer and

Please cite this article in press as: Farnung et al., Structure of a backtracked hexasomal intermediate of nucleosome transcription, Molecular Cell (2022), https://doi.org/10.1016/j.molcel.2022.06.027

involves RNA Pol II backtracking. We also obtain mechanistic details into how the human elongation factor TFIIS helps RNA Pol II pass through a nucleosome.

RESULTS

Human elongation factors stimulate nucleosome passage

To investigate nucleosome passage by mammalian RNA Pol II, we formed a RNA Pol II elongation complex on a nucleosomal DNA template (Figure 1A) in the presence of transcription elongation factors DSIF, PAF, the dissociable PAF subunit RTF1, SPT6, positive transcription elongation factor b (P-TEFb), and substoichiometric amounts of TFIIS (Figure S1A). As in previous experiments (Farnung et al., 2018, 2021; Žumer et al., 2021), we started transcription elongation by adding 1 mM of NTPs and performed quantitative time course analyses (Figure S1D). In the presence of DSIF, TFIIS was essential for nucleosome passage (Figures S1B and S1C), and the additional inclusion of PAF and SPT6 resulted in the formation of 4 times more full-length RNA product relative to samples where only DSIF and TFIIS were included (Figures 1B and 1C).

Further introduction of RTF1 revealed an additional increase of full-length product by a factor of three (Figures 1B and 1C), consistent with the previously demonstrated strong RNA extension-stimulatory activity of RTF1 (Vos et al., 2020). RTF1 also changed the frequency and location of RNA Pol II stalling within the nucleosome. In the absence of RTF1, a typical distribution of stalling sites within the nucleosome at bp 1 (SHL −5) and at bp 42 (SHL −1) was observed, whereas RTF1 led to a major reduction of stalling at bp 42 and the emergence of a stalled intermediate at bp 64 (SHL 0) (Figure 1B).

To better quantify the differences in stalling at specific sites within the nucleosome, we defined a "barrier index." The barrier index is calculated as the ratio of intensities of the band at a given stall site (barrier) to the band intensities of the barrier band and all longer RNA products that result from passage of RNA Pol II past the barrier (Figure S1E). Consequently, a barrier index of 1 reflects that the barrier cannot be overcome, whereas a barrier index of 0 means that no barrier exists. Calculation of barrier indices indicates that the bp 42 site represents a major barrier that is reduced 3-fold in the presence of RTF1 (Figure 1D).

Structure of nucleosome transcription intermediate

To understand how human elongation factors assist RNA Pol II in nucleosome passage, we determined the cryo-EM structure of a transcribing RNA Pol II-nucleosome complex containing human elongation factors (EC*) (Figure 2). We obtained the structure by using the same nucleosomal DNA template as employed in our biochemical experiments. This DNA template contained a T-less cassette up to nucleosomal bp 42. EC* was formed on the substrate as described previously (Figures S2A and S2B) (Vos et al., 2020). Transcription elongation was then started by the addition of 3′-dATP, CTP, GTP, and UTP. This set of NTPs is predicted to cause stalling at the end of the T-less cassette. EC* was allowed to transcribe into the nucleosome for 15 min.

The nucleosome transcription intermediate was purified by size exclusion chromatography and lightly cross-linked with glutaraldehyde. The sample was used for cryo-EM analysis and a reconstruction was derived at an overall resolution of 3.6 Å from 32,933 particles (Figures S3 and S6A–S6D). For modeling, we docked the structure of the complete EC* (Vos et al., 2020) into our density and adjusted it locally. We did not observe any major changes compared with the EC* structure and did not observe density for TFIIS in nucleosome-bound particle classes. The density for the nucleosome could be modeled with the canonical nucleosome structure (Vasudevan et al., 2010) after the proximal H2A–H2B dimer was removed. The resulting refined structure shows good stereochemistry (Table S1).

Displacement of proximal H2A–H2B dimer and nucleosome contacts

The structure reveals the RNA Pol II elongation complex, the bound factors DSIF, PAF, RTF1, and SPT6, and a partially unraveled nucleosome (Figure 2). A striking feature of the structure is the absence of the proximal H2A–H2B dimer from the nucleosome, leading to a hexasome with ~90 bp of DNA wrapped around it (Figure 3). Unexpectedly, RNA Pol II adopts a backtracked state, with the RNA 3′ end being dislodged from the RNA Pol II active site and 10 nucleotides of RNA occupying a backtrack site in the pore and funnel of RNA Pol II (Figure 4A).

Our structure also shows several contacts between the transcription elongation complex and the partially unraveled hexasome. First, the RPB1 jaw domain of RNA Pol II is positioned in close proximity to helix α2 of histone H4, the loop 2 region of the exposed histone H4 and H3 loop 2 region between H3 α2 and α3 via a flexible loop (RPB1 residues 1,263–1,280) (Figure S7A). Second, the N-terminal tail of DSIF subunit SPT5 (residues 1–176) is positioned near the DNA at SHL +4.5 (Figure S7B). Third, the trestle helix of the PAF subunit CTR9 is located closely to the nucleosomal DNA at SHL −2 of the hexasome (Figure S7C). Superposition of our structure with the previously reported structures of the human nucleosome-FACT complex (Liu et al., 2020) shows that FACT could bind the transcription machinery without clashes (Figure S8). It is possible that FACT could redeposit a H2A–H2B dimer in the observed conformation.

RNA Pol II backtracking at a nucleosomal barrier

In our structure, RNA Pol II adopts a backtracked state. RNA Pol II is backtracked by 10 bp to nucleosomal bp 32. Backtracked RNA extends into the RNA Pol II pore and emerges from the RNA Pol II funnel (Figure 4A). The DNA-RNA hybrid is not tilted (Figure 4B), in contrast to previous structures of yeast backtracked complexes that showed a tilted hybrid (Cheung and Cramer, 2011). Backtracking is likely a result of RNA Pol II encountering a strong nucleosomal barrier at the end of the T-less cassette. During nucleosome passage, RNA Pol II must break contacts between histones and the DNA. If histone-DNA contacts are too strong, RNA Pol II stalls and becomes prone to backtracking (Bintu et al., 2012). Indeed, nucleosome passage may generally involve backtracking, and this can explain the requirement of TFIIS for successful passage (Kireeva et al., 2005). Complementary to previous biophysical observations, we observe rewrapping of nucleosomal DNA after RNA Pol II backtracking (Bintu et al., 2012; Gaykalova et al., 2015).

Please cite this article in press as: Farnung et al., Structure of a backtracked hexasomal intermediate of nucleosome transcription, Molecular Cell (2022), https://doi.org/10.1016/j.molcel.2022.06.027

Molecular Cell
Article

CellPress

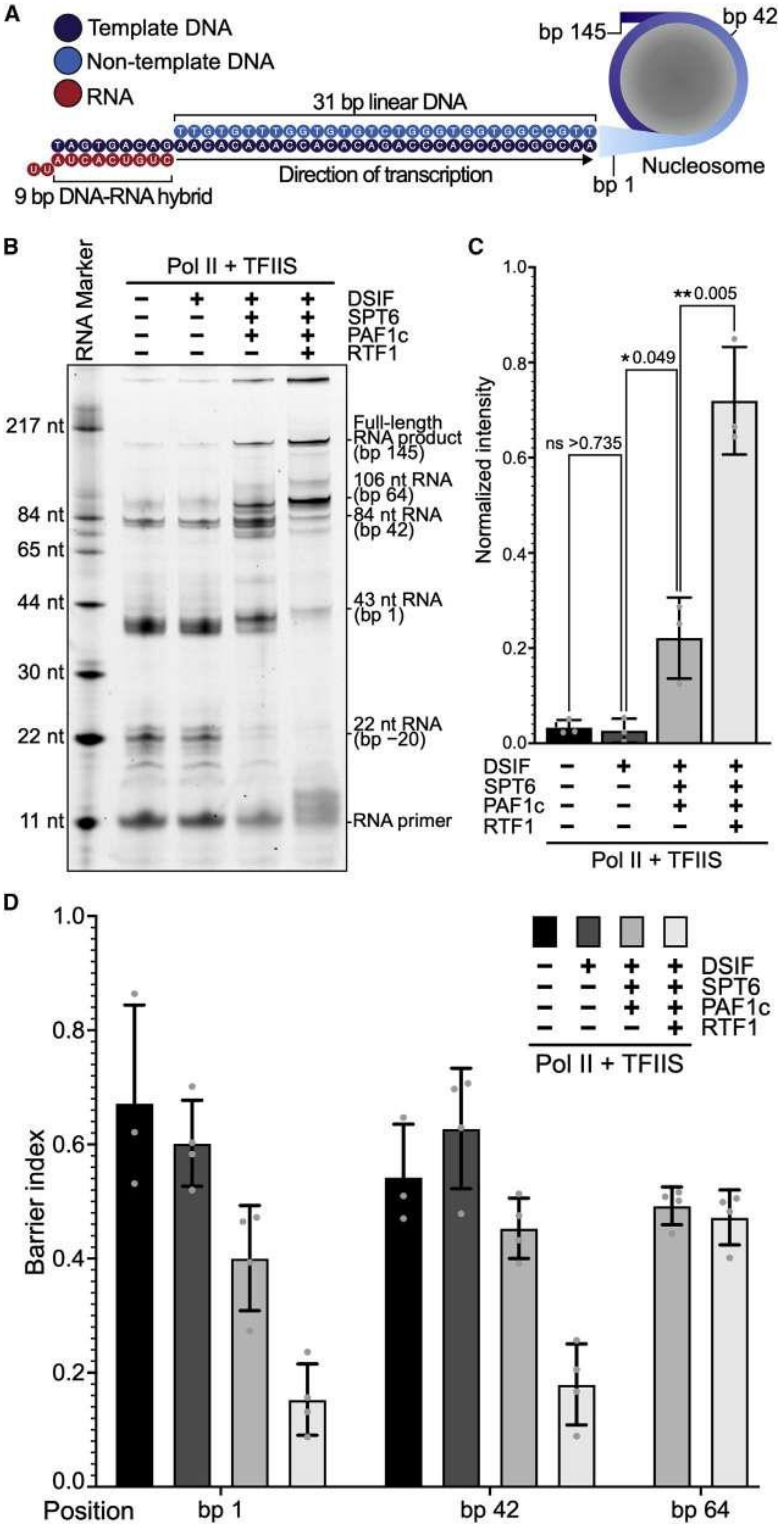


Figure 1. Human elongation factors enable nucleosome passage by RNA Pol II

(A) Design of nucleosomal substrate. Extranucleosomal DNA, nucleosomal DNA, and nucleosomal bp 1 and 42 are indicated.

(B) RNA extension by mammalian RNA Pol II and its dependence on elongation factors TFIIS, DSIF, SPT6, PAF, and RTF1. The RNA primer is labeled with a 5′ FAM label. RNA lengths are indicated on the left side of the gel. The positions of prominent stalling products are indicated on the right side of the gel.

(C) Bar chart of full-length product RNA intensity (n = 3). The full-length RNA signal was quantified by measuring the integrated signal density, subtracting the background, and normalizing it against the signal of the complete lane and the whole gel (Methods). Mean ± standard deviation (SD) values are shown. p values were determined using unpaired two-tailed t tests (*$p < 0.05$ and **$p < 0.01$).

(D) Prominent stalling positions within the nucleosome are quantified using the barrier index (n = 4). The barrier index is the ratio of the barrier band intensity relative to the intensity of the barrier band and all larger RNA extension products (Figure S1E). Inclusion of elongation factors alters the position of specific stall sites. Mean ± SD values are shown.

See also Figure S1.

Please cite this article in press as: Farnung et al., Structure of a backtracked hexasomal intermediate of nucleosome transcription, Molecular Cell (2022), https://doi.org/10.1016/j.molcel.2022.06.027

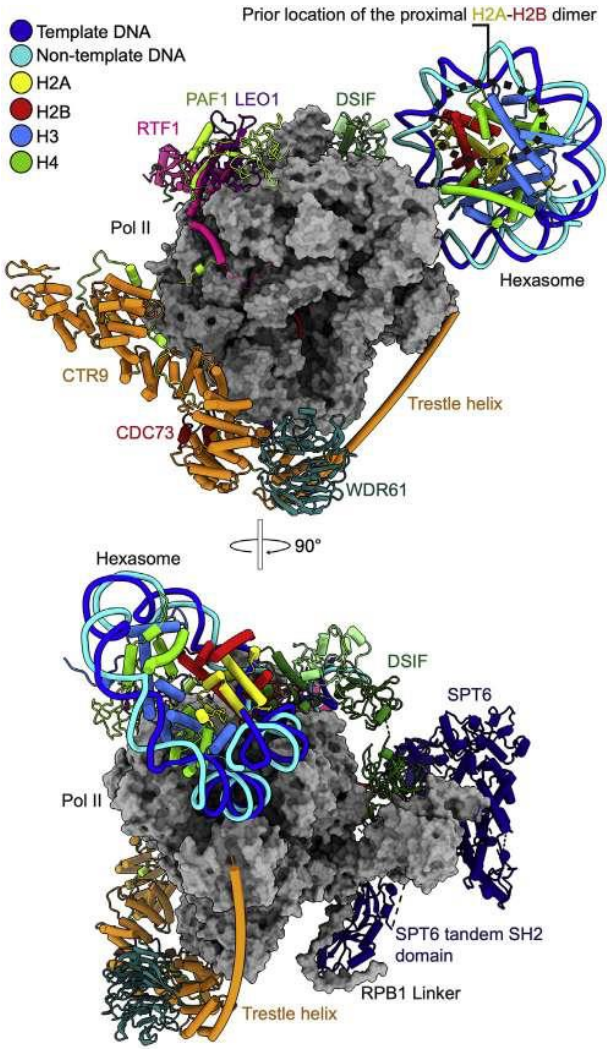


Figure 2. Structure of RNA Pol II EC* intermediate of nucleosome passage

Side and front views of the EC*-hexasome structure. RNA Pol II is colored silver, and the elongation factors DSIF, SPT6, PAF (CTR9, PAF1, LEO1, CDC73, WDR61), and RTF1 are colored green, blue, orange, lime, purple, red, teal, and pink, respectively. Histones H2A, H2B, H3, and H4 are colored yellow, red, blue, and green, respectively. DNA is blue (template strand) and cyan (non-template strand). The location of the displaced H2A–H2B dimer is indicated by a dashed ellipse.

See also Figures S3, S6, and S7.

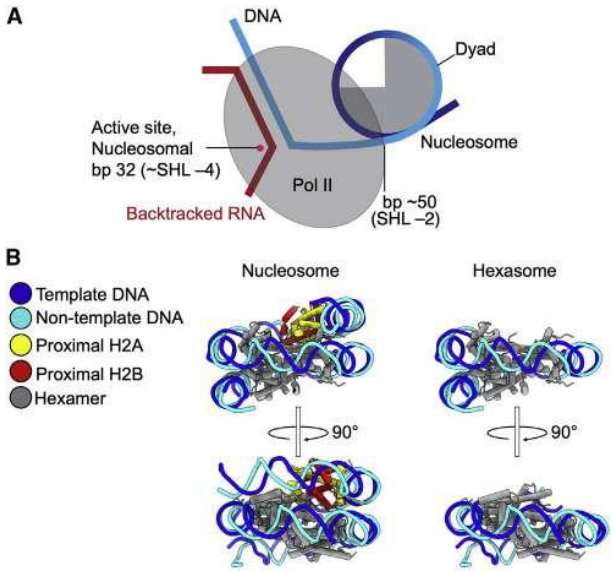


Figure 3. Loss of proximal H2A–H2B dimer

(A) Cartoon diagram of the backtracked state of the intermediate EC*-hexasome structure.

(B) Comparison of the hexasome in our structure with a canonical nucleosome (PDB: 3LZ0). The proximal H2A–H2B dimer is colored yellow and red. Other histone proteins are colored gray.

See also Figure S3.

We confirmed the formation of a backtracked intermediate and determined the precise register of the backtracked complex biochemically (Figures 4C and 4D). To this end, we formed the EC*-nucleosome complex, started transcription elongation by addition of CTP, GTP, UTP, and 3′-dATP, and purified the complex using size exclusion chromatography. We obtained a stable intermediate that showed RNA extension to nucleosomal bp 42 (Figure 4C). After addition of a 2-fold molar excess of TFIIS, we observed cleavage of 10 nucleotides from the RNA 3′ end, consistent with the RNA Pol II backtracking position observed in our structure. A catalytically dead variant of TFIIS (DE282-283AA, DEAA) did not stimulate cleavage, providing a negative control (Figures 4C and 4D). Together, these results directly show that a nucleosome can cause RNA Pol II backtracking and that cleavage of backtracked RNA in the resulting intermediate can be stimulated by TFIIS.

TFIIS-containing structures reveal a mobile domain III

We were able to resolve the structure of the stably backtracked RNA Pol II-nucleosome intermediate likely because TFIIS was lost during sample preparation. To study how TFIIS could reactivate a backtracked complex, we additionally solved the structure of EC* with TFIIS (DEAA) on linear DNA at a resolution of 3.8 Å (Figures 5, S4, and S6E–S6I). The structure of the TFIIS-bound EC* reveals a post-translocated state of the DNA-RNA hybrid. Domain II of human TFIIS binds the exposed jaw domain of RPB1, as observed for yeast TFIIS (Cheung and Cramer, 2011; Kettenberger et al., 2003). Domain III of TFIIS inserts into the RNA Pol II pore similar to what was observed in yeast RNA Pol II-TFIIS structures (Cheung and Cramer, 2011; Kettenberger et al., 2003). However, domain III is not well resolved in our EC*-TFIIS reconstruction, indicating mobility or partial occupancy in the pore.

The interdomain linker of TFIIS runs through a crevice between the RNA Pol II funnel domain and the shelf module, as in the yeast RNA Pol II-TFIIS structures (Cheung and Cramer, 2011; Kettenberger et al., 2003). In contrast to yeast RNA Pol II, however, the crevice is in a pre-opened state in the mammalian RNA Pol II EC (Bernecky et al., 2016). To investigate this further, we also

Please cite this article in press as: Farnung et al., Structure of a backtracked hexasomal intermediate of nucleosome transcription, Molecular Cell (2022), https://doi.org/10.1016/j.molcel.2022.06.027

Molecular Cell
Article

CellPress

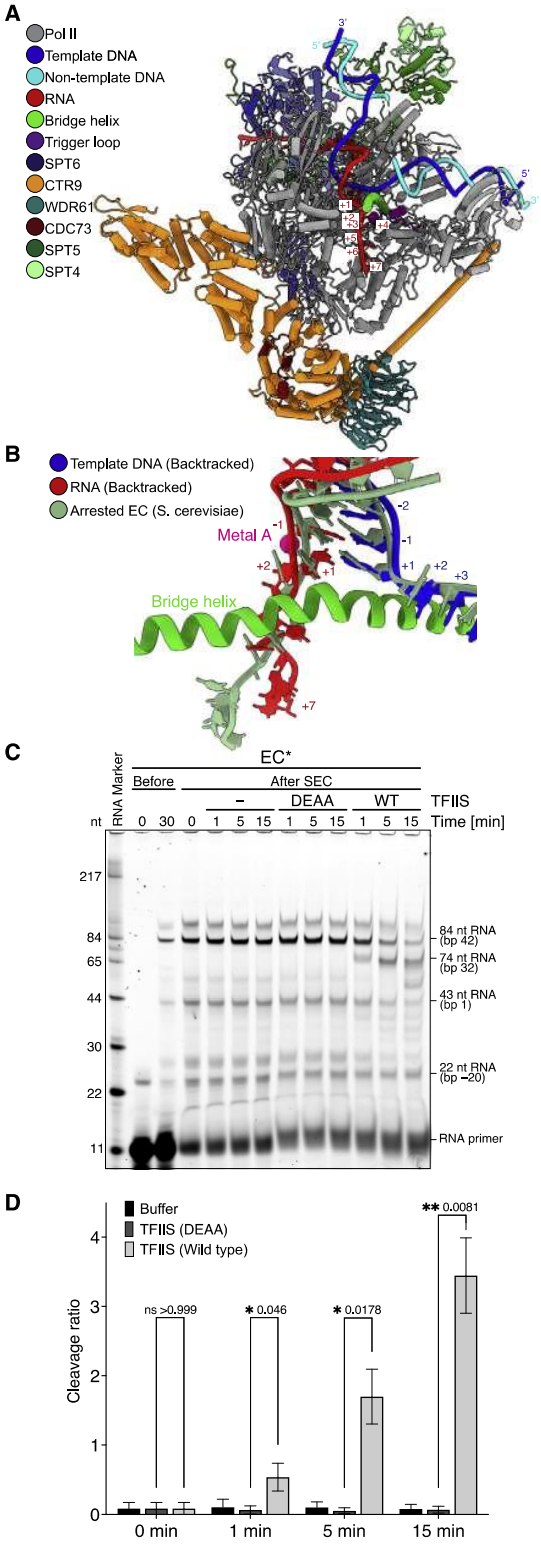


Figure 4. RNA Pol II adopts a backtracked state

(A) Backtracked state of RNA Pol II observed in the intermediate structure. Viewed from the side. Backtracked RNA in the pore and funnel is indicated (register +1 represents the nucleotide addition site, +2 etc. backtracked positions).

(B) Comparison of the location of backtracked RNA with that observed in a yeast backtracked complex (Cheung and Cramer, 2011) (PDB: 3PO2) after the superposition of active site domains.

(C) Biochemical demonstration that the intermediate adopts a backtracked state. After purification by size exclusion chromatography, TFIIS was added to the EC*-nucleosome intermediate complex and incubated for various amounts of time. Quenched reactions were separated by denaturing gel electrophoresis. The prominent stalling product at nucleosome bp 42 is only cleaved upon the inclusion of wild-type (WT) TFIIS.

(D) Quantification of the cleavage ratio after different time points (n = 3). The integrated signal density of the uncleaved and cleaved bands were measured, and after the background was subtracted, the density was used to determine the cleaved-uncleaved ratio in the presence of buffer, a catalytically dead variant, or wild-type TFIIS (STAR Methods). Mean ± SD values are shown. p values were determined using unpaired two-tailed t tests (*p < 0.05 and **p < 0.01).

See also Figure S1.

determined the mammalian EC-TFIIS structure in the absence of elongation factors at 3.1 Å resolution (Figures 6, S5, and S6J–S6O). In this structure, we observed domain II and the interdomain linker at the same position and again weak density for domain III. The RNA Pol II trigger loop also remains partially mobile in all TFIIS-containing structures and does not adopt the locked state that was previously observed for the yeast RNA Pol II-TFIIS structures (Cheung and Cramer, 2011; Kettenberger et al., 2003). Thus, the TFIIS-containing structures show that TFIIS can reach the active site of RNA Pol II but that the catalytic domain III does not stably bind at the RNA Pol II active site as observed previously in the yeast system.

DISCUSSION

Together with published biochemical data, our results converge on a model that can explain nucleosome passage by RNA Pol II with the help of TFIIS (Video S1). In this model, the activated elongation complex EC* first transcribes into the nucleosome, unraveling three turns of nucleosomal DNA, thereby exposing the proximal H2A–H2B dimer. The loss of interactions between the H2A–H2B dimer and the DNA may destabilize the dimer and could lead to its loss and formation of a hexasome. Loss of H2A–H2B may be counteracted by binding of the N-terminal tail of DSIF subunit SPT5 (Evrin et al., 2022; Farnung et al., 2021). In a second step, RNA Pol II progresses by another three turns of DNA before it stalls at a major nucleosomal barrier at SHL −1. The stalled complex may be prone to backtracking, which leads to the reformation of some of the histone-DNA contacts that likely stabilize the backtracked state. Alternatively, the H2A–H2B dimer could be lost because of stalling and subsequent backtracking. Backtracking by more than one turn of DNA may be prevented by binding of backtracked RNA into the backtrack site that lines the RNA Pol II pore and funnel (Cheung and Cramer, 2011). In a third step, TFIIS can rescue the backtracked complex by stimulating RNA cleavage, thereby removing backtracked nucleotides from the RNA to generate a

Please cite this article in press as: Farnung et al., Structure of a backtracked hexasomal intermediate of nucleosome transcription, Molecular Cell (2022), https://doi.org/10.1016/j.molcel.2022.06.027

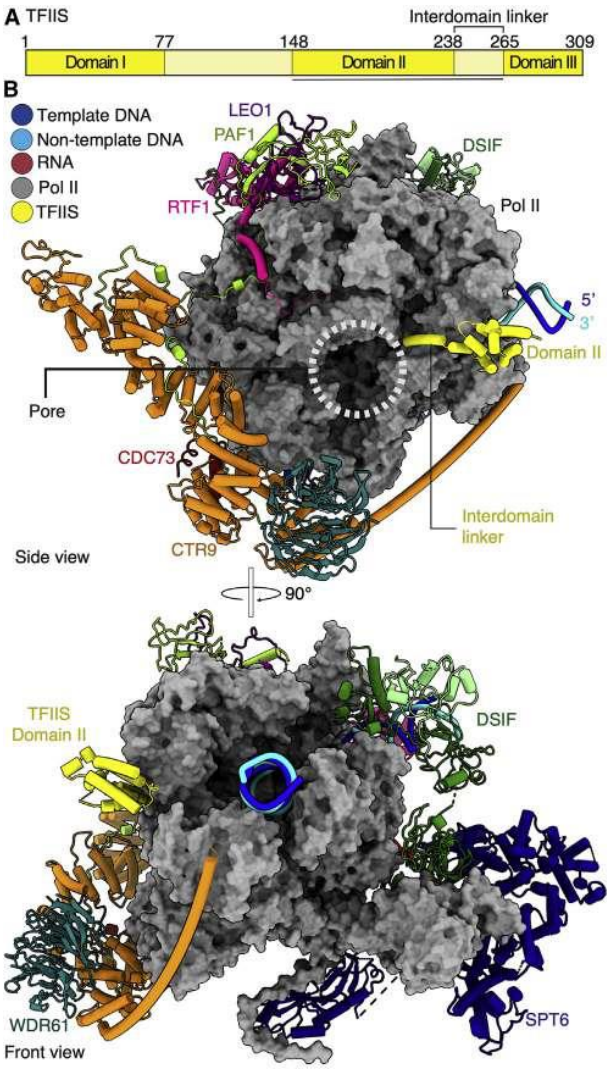


Figure 5. Structure of mammalian EC*-TFIIS elongation complex

(A) Primary structure representation of human TFIIS. Domain boundaries are indicated. Black line below domain II indicates region included in the structure. (B) Structure of EC* with bound human TFIIS (yellow). Side and top views are shown. The RNA Pol II pore is indicated with a dashed circle.
See also Figure S4.

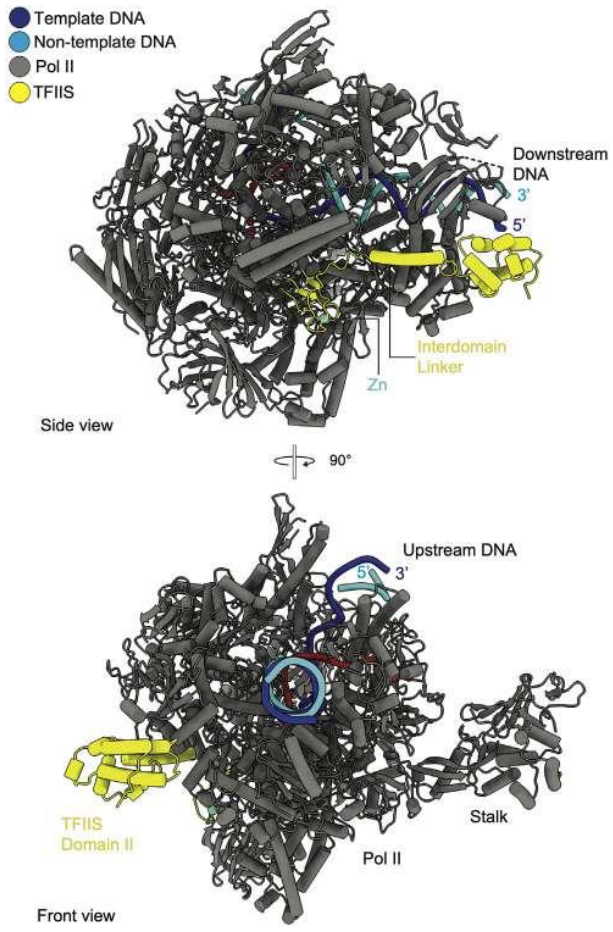


Figure 6. Structure of mammalian RNA Pol II-TFIIS elongation complex

The structure is shown in side and top views. TFIIS is colored yellow, with the zinc ion domain III depicted as a cyan sphere.
See also Figure S5.

new 3′ end from which transcription resumes. Upon further RNA Pol II progression, the hexasome may be released from DNA and RNA Pol II completes nucleosome passage (Hsieh et al., 2013). Although we do not see a direct interaction between PAF and the nucleosome in our structure, our biochemical data as well as that of others suggest that PAF and TFIIS are essential for nucleosomal traversal (Kim et al., 2010).

This three-step model of elongation-factor-assisted nucleosome passage by RNA Pol II provides a framework for understanding how the histone chaperone FACT counteracts the loss of histones during nucleosome passage (Hsieh et al., 2013). Loss of the proximal H2A–H2B leads to displacement of the nucleosome, which is detrimental because it exposes DNA and leads to loss of the information provided by post-translational modifications of histones. FACT could bind the exposed H2A–H2B dimer during backtracking, to prevent its loss and enable full transfer of the histone octamer to upstream DNA. Alternatively, we speculate that FACT could bind the displaced H2A–H2B dimer and deposit it back onto upstream DNA after RNA Pol II passage (Farnung et al., 2021; Liu et al., 2020).

Finally, our work provides new insights into understanding RNA Pol II regulation at the stage of stalling, backtracking, arrest, and reactivation by TFIIS. These processes have been the focus of several mechanistic studies in the yeast system that had shown that an acidic hairpin in domain III of TFIIS reaches the RNA Pol II active site to stimulate hydrolytic cleavage of backtracked RNA (Cheung and Cramer, 2011; Kettenberger et al., 2003, 2004; Wang et al., 2009). However, a similar mechanism has not been demonstrated structurally for the human transcription system yet. Our high-resolution structures of the mammalian

Please cite this article in press as: Farnung et al., Structure of a backtracked hexasomal intermediate of nucleosome transcription, Molecular Cell (2022), https://doi.org/10.1016/j.molcel.2022.06.027

RNA Pol II complexes bound to human TFIIS that uncover some differences between the yeast and human systems. The binding sites of TFIIS domain II and the interdomain linker are the same in both systems, whereas domain III of TFIIS retains some mobility when bound inside the RNA Pol II pore in the human system. We speculate that this mobility observed in the mammalian system enables regulation of TFIIS entry into the RNA Pol II active site and thus more intricate regulation of TFIIS-mediated RNA cleavage and reactivation of arrested RNA Pol II.

Limitations of the study

The functional assays performed in this study were performed in an *in-vitro*-reconstituted system. We cannot distinguish between improper EC* formation, failure of the RNA to anneal to the nucleosome substrate, and dead-end complex formation due to overextended DNA-RNA hybrids. Due to these considerations, we refrain from quantifying the fraction of productive complexes.

We note that RNA Pol II interacts with additional elongation factors that can positively or negatively regulate transcription elongation and traversal of RNA Pol II through chromatin. For example, we were not able to bind the transcription elongation factor ELOF1 to our human complexes. Its homolog Elf1, however, plays an important role in yeast nucleosome traversal. It remains to be seen whether ELOF1 plays a similar role in the human system.

We note that we observe stalling at bp 42 both on linear DNA and nucleosomal DNA. The stalling, however, is significantly increased on a nucleosomal template, strongly supporting the role of DNA rewrapping in stalling and backtracking during nucleosome transcription.

Although all EC*-nucleosome particles that we observe lack the H2A–H2B dimer, additional transcription elongation factors could influence the amount of H2A–H2B dimer loss. It needs further investigation *in vitro* and *in vivo* to clarify the process of nucleosome passage and nucleosome retention during transcription. It remains unknown whether the H2A–H2B dimer is lost prior to or during backtracking.

Our movie showing transcription through chromatin is based on published cryo-EM structures and structures presented here. The movie, however, still requires extrapolation between the different states and is therefore partially speculative.

The heterogeneity of the sample and flexibility of the nucleosome-RNA Pol II interface also only allowed a detailed cryo-EM reconstruction by employing masked refinements and combination of multiple maps to generate a composite map.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
- METHOD DETAILS
  - Preparation of protein components
  - Nucleosome preparation
  - RNA extension assays
  - TFIIS cleavage-stimulatory assay
  - Sample preparation for cryo-EM
  - Cryo-EM analysis and image processing
  - Model building
  - Figure generation
  - Movie generation
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2022.06.027.

ACKNOWLEDGMENTS

We thank past and present members of the Cramer laboratory. We thank C. Oberthür for help with protein purification. We thank C. Dienemann and U. Steuerwald for maintenance of cryo-EM resources. P.C. was supported by the Deutsche Forschungsgemeinschaft (SFB860, EXC 2067/1-390729940) and the ERC Advanced Investigator Grant CHROMATRANS (grant agreement no. 693023).

AUTHOR CONTRIBUTIONS

L.F., S.M.V., and M.O. purified protein components. S.M.V. performed initial cryo-EM experiments of elongation complexes with TFIIS. L.F. and G.G. prepared complexes for cryo-EM and performed data acquisition. M.O. performed biochemical experiments. G.G., S.M.V., and L.F. processed and analyzed the EC-TFIIS, EC*-TFIIS, and EC*-hexasome datasets, respectively. L.F. designed and coordinated experiments. P.C. supervised research. L.F., M.O., G.G., S.M.V., and P.C. prepared the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 30, 2022
Revised: May 17, 2022
Accepted: June 21, 2022
Published: July 19, 2022

REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). Phenix: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221. https://doi.org/10.1107/S0907444909052925.

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.

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.

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.

Bintu, L., Ishibashi, T., Dangkulwanich, M., Wu, Y.-Y., Lubkowska, L., Kashlev, M., and Bustamante, C. (2012). Nucleosomal elements that control the topography of the barrier to transcription. Cell *151*, 738–749. https://doi.org/10.1016/j.cell.2012.10.009.

Please cite this article in press as: Farnung et al., Structure of a backtracked hexasomal intermediate of nucleosome transcription, Molecular Cell (2022), https://doi.org/10.1016/j.molcel.2022.06.027

Bondarenko, V.A., Steele, L.M., Újvári, A., Gaykalova, D.A., Kulaeva, O.I., Polikanov, Y.S., Luse, D.S., and Studitsky, V.M. (2006). Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II. Mol. Cell *24*, 469–479. https://doi.org/10.1016/j.molcel.2006.09.009.

Bortvin, A., and Winston, F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. Science *272*, 1473–1476. https://doi.org/10.1126/science.272.5267.1473.

Chang, H.-W., Kulaeva, O.I., Shaytan, A.K., Kibanov, M., Kuznedelov, K., Severinov, K.V., Kirpichnikov, M.P., Clark, D.J., and Studitsky, V.M. (2014). Analysis of the mechanism of nucleosome survival during transcription. Nucleic Acids Res. *42*, 1619–1627. https://doi.org/10.1093/nar/gkt1120.

Cheung, A.C.M., and Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature *471*, 249–253. https://doi.org/10.1038/nature09785.

Cole, H.A., Ocampo, J., Iben, J.R., Chereji, R.V., and Clark, D.J. (2014). Heavy transcription of yeast genes correlates with differential loss of histone H2B relative to H4 and queued RNA polymerases. Nucleic Acids Res. *42*, 12512–12522. https://doi.org/10.1093/nar/gku1013.

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.

Dyer, P.N., Edayathumangalam, R.S., White, C.L., Bao, Y., Chakravarthy, S., Muthurajan, U.M., and Luger, K. (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. Methods Enzymol. *375*, 23–44. https://doi.org/10.1016/s0076-6879(03)75002-2.

Ehara, H., Kujirai, T., Fujino, Y., Shirouzu, M., Kurumizaka, H., and Sekine, S.I. (2019). Structural insight into nucleosome transcription by RNA polymerase II with elongation factors. Science *363*, 744–747. https://doi.org/10.1126/science.aav8912.

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.

Evrin, C., Serra-Cardona, A., Duan, S., Mukherjee, P.P., Zhang, Z., and Labib, K.P.M. (2022). Spt5 histone binding activity preserves chromatin during transcription by RNA polymerase II. EMBO J. *41*, e109783. https://doi.org/10.15252/embj.2021109783.

Farnung, L., Ochmann, M., Engeholm, M., and Cramer, P. (2021). Structural basis of nucleosome transcription mediated by Chd1 and FACT. Nat. Struct. Mol. Biol. *28*, 382–387. https://doi.org/10.1038/s41594-021-00578-6.

Farnung, L., Vos, S.M., and Cramer, P. (2018). Structure of transcribing RNA polymerase II-nucleosome complex. Nat. Commun. *9*, 5432. https://doi.org/10.1038/s41467-018-07870-y.

Farnung, L., Vos, S.M., Wigge, C., and Cramer, P. (2017). Nucleosome-Chd1 structure and implications for chromatin remodelling. Nature *550*, 539–542. https://doi.org/10.1038/nature24046.

Gaykalova, D.A., Kulaeva, O.I., Volokh, O., Shaytan, A.K., Hsieh, F.-K., Kirpichnikov, M.P., Sokolova, O.S., and Studitsky, V.M. (2015). Structural analysis of nucleosomal barrier to transcription. Proc. Natl. Acad. Sci. USA *112*, E5787–E5795. https://doi.org/10.1073/pnas.1508371112.

Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: meeting modern challenges in visualization and analysis. Protein Sci. *27*, 14–25. https://doi.org/10.1002/pro.3235.

Guermah, M., Palhan, V.B., Tackett, A.J., Chait, B.T., and Roeder, R.G. (2006). Synergistic functions of SII and p300 in productive activator-dependent transcription of chromatin templates. Cell *125*, 275–286. https://doi.org/10.1016/j.cell.2006.01.055.

Hsieh, F.-K., Kulaeva, O.I., Patel, S.S., Dyer, P.N., Luger, K., Reinberg, D., and Studitsky, V.M. (2013). Histone chaperone FACT action during transcription through chromatin by RNA polymerase II. Proc. Natl. Acad. Sci. USA *110*, 7654–7659. https://doi.org/10.1073/pnas.1222198110.

Jamai, A., Imoberdorf, R.M., and Strubin, M. (2007). Continuous histone H2B and transcription-dependent histone H3 exchange in yeast cells outside of replication. Mol. Cell *25*, 345–355. https://doi.org/10.1016/j.molcel.2007.01.019.

Jeronimo, C., Poitras, C., and Robert, F. (2019). Histone recycling by FACT and Spt6 during transcription prevents the scrambling of histone modifications. Cell Rep. *28*, 1206–1218.e8. https://doi.org/10.1016/j.celrep.2019.06.097.

Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. Science *301*, 1096–1099. https://doi.org/10.1126/science.1087374.

Kettenberger, H., Armache, K.-J., and Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. Cell *114*, 347–357. https://doi.org/10.1016/s0092-8674(03)00598-1.

Kettenberger, H., Armache, K.-J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. Mol. Cell *16*, 955–965. https://doi.org/10.1016/j.molcel.2004.11.040.

Kim, J., Guermah, M., and Roeder, R.G. (2010). The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. Cell *140*, 491–503. https://doi.org/10.1016/j.cell.2009.12.050.

Kimura, H., and Cook, P.R. (2001). Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. J. Cell Biol. *153*, 1341–1353. https://doi.org/10.1083/jcb.153.7.1341.

Kireeva, M.L., Hancock, B., Cremona, G.H., Walter, W., Studitsky, V.M., and Kashlev, M. (2005). Nature of the nucleosomal barrier to RNA polymerase II. Mol. Cell *18*, 97–108. https://doi.org/10.1016/j.molcel.2005.02.027.

Kireeva, M.L., Walter, W., Tchernajenko, V., Bondarenko, V., Kashlev, M., and Studitsky, V.M. (2002). Nucleosome remodeling induced by RNA polymerase II loss of the H2A/H2B dimer during transcription. Mol. Cell *9*, 541–552. https://doi.org/10.1016/s1097-2765(02)00472-0.

Kujirai, T., Ehara, H., Fujino, Y., Shirouzu, M., Sekine, S.-I., and Kurumizaka, H. (2018). Structural basis of the nucleosome transition during RNA polymerase II passage. Science *362*, 595–598. https://doi.org/10.1126/science.aau9904.

Kujirai, T., and Kurumizaka, H. (2020). Transcription through the nucleosome. Curr. Opin. Struct. Biol. *61*, 42–49. https://doi.org/10.1016/j.sbi.2019.10.007.

Kulaeva, O.I., Gaykalova, D.A., Pestov, N.A., Golovastov, V.V., Vassylyev, D.G., Artsimovitch, I., and Studitsky, V.M. (2009). Mechanism of chromatin remodeling and recovery during passage of RNA polymerase II. Nat. Struct. Mol. Biol. *16*, 1272–1278. https://doi.org/10.1038/nsmb.1689.

Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. Nat. Rev. Mol. Cell Biol. *18*, 548–562. https://doi.org/10.1038/nrm.2017.47.

Liu, Y., Zhou, K., Zhang, N., Wei, H., Tan, Y.Z., Zhang, Z., Carragher, B., Potter, C.S., D'Arcy, S., and Luger, K. (2020). FACT caught in the act of manipulating the nucleosome. Nature *577*, 426–431. https://doi.org/10.1038/s41586-019-1820-0.

Petesch, S.J., and Lis, J.T. (2012). Overcoming the nucleosome barrier during transcript elongation. Trends Genet. *28*, 285–294. https://doi.org/10.1016/j.tig.2012.02.005.

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.

Ramachandran, S., Ahmad, K., and Henikoff, S. (2017). Transcription and remodeling produce asymmetrically unwrapped nucleosomal intermediates. Mol. Cell *68*, 1038–1053.e4. https://doi.org/10.1016/j.molcel.2017.11.015.

Scheres, S.H.W. (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.

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.

Studitsky, V.M., Nizovtseva, E.V., Shaytan, A.K., and Luse, D.S. (2016). Nucleosomal barrier to transcription: structural determinants and changes in

Please cite this article in press as: Farnung et al., Structure of a backtracked hexasomal intermediate of nucleosome transcription, Molecular Cell (2022), https://doi.org/10.1016/j.molcel.2022.06.027

chromatin structure. Biochem. Mol. Biol. J. *2*, 8. https://doi.org/10.21767/2471-8084.100017.

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.

Teves, S.S., Weber, C.M., and Henikoff, S. (2014). Transcribing through the nucleosome. Trends Biochem. Sci. *39*, 577–586. https://doi.org/10.1016/j.tibs.2014.10.004.

Vasudevan, D., Chua, E.Y.D., and Davey, C.A. (2010). Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. J. Mol. Biol. *403*, 1–10. https://doi.org/10.1016/j.jmb.2010.08.039.

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. Nat. Rev. Mol. Cell Biol. *16*, 178–189. https://doi.org/10.1038/nrm3941.

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.

Vos, S.M., Farnung, L., Linden, A., Urlaub, H., and Cramer, P. (2020). Structure of complete Pol II–DSIF–PAF–SPT6 transcription complex reveals RTF1 allosteric activation. Nat. Struct. Mol. Biol. *27*, 668–677. https://doi.org/10.1038/s41594-020-0437-1.

Wang, D., Bushnell, D.A., Huang, X., Westover, K.D., Levitt, M., and Kornberg, R.D. (2009). Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. Science *324*, 1203–1206. https://doi.org/10.1126/science.1168729.

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.

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

Žumer, K., Maier, K.C., Farnung, L., Jaeger, M.G., Rus, P., Winter, G., and Cramer, P. (2021). Two distinct mechanisms of RNA polymerase II elongation stimulation in vivo. Mol. Cell *81*, 3096–3109.e8. https://doi.org/10.1016/j.molcel.2021.05.028.

Please cite this article in press as: Farnung et al., Structure of a backtracked hexasomal intermediate of nucleosome transcription, Molecular Cell (2022), https://doi.org/10.1016/j.molcel.2022.06.027

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
BL21 (RIL) DE3	Agilent	230245
XL1 Blue	Agilent	200249
Hi5 cells	Expression Systems	94-002F
Sf9 cells	Thermo Fischer	12659017
Sf21 cells	Expression Systems	94-003S
ESF 921	Expression Systems	96-001-01
Biological samples		
Sus scrofa thymus	Locally sourced	N/A
Chemicals, peptides, and recombinant proteins		
Phusion® HF Buffer	New England Biolabs	B0518S
Phusion DNA Polymerase	In house	N/A
TspRI	New England Biolabs	R0582S
3'-dATP	Jena Bioscience	NU-1123L
NTP Set	Thermo Fischer	R0481
Glutaraldehyde	Electron Microscopy Sciences	16200
Deposited data		
EC*–Hexasome	This paper	PDB: 8A3Y, EMD-15127
EC*–TFIIS	This paper	PDB: 8A3Z, EMD-15128
EC-TFIIS	This paper	PDB: 8A40, EMD-15129
EC* structure	Vos et al., 2020	PDB: 6TED
Raw gel data	This paper	https://doi.org/10.17632/6s5pj5x2zj.1
Oligonucleotides		
Forward DNA primer: ACGAAGCGTAGCATCACTGTCTTG	Integrated DNA Technologies	N/A
Reverse DNA primer: ATCAGAATCCCGGTGCCG	Integrated DNA Technologies	N/A
RNA /56-FAM/UUAUCACUGUC	Integrated DNA Technologies	N/A
Recombinant DNA		
pET3a_H2A_fl	Farnung et al., 2017	Internal ID: 3320
pET3a_H2B_fl	Farnung et al., 2017	Internal ID: 3321
pET3a_H3_fl	Farnung et al., 2017	Internal ID: 3322
pET3a_H4_fl	Farnung et al., 2017	Internal ID: 3323
pIDTSmartKan_Farnung-Construct_+42	This paper	Internal ID: 3946
438A_CTR9-6xHis_CDC73_LEO1_SKI8_PAF1	Vos et al., 2018	Internal ID: 3245
438C_MBP-6xHis_RTF1	Vos et al., 2020	Internal ID: 3129
pKSUMO_10xHis-3C-SPT4_SPT5	Bernecky et al. (2017)	Internal ID: 2439
438C_NHis6-MBP-N10-TEV-SUPT6H_H	Vos et al., 2018	Internal ID: 3425
1C_NHis6-MBP-N10-TEV-TFIIS_H	This paper	Internal ID: 3619
1C_NHis6-MBP-N10-TEV-TFIIS_DE282-283AA_H	This paper	Internal ID: 3628
pACEBac1_ His8-TEV-Cdk9(1-372)+GST-TEV-hCycT1(1-272)	Vos et al., 2018	Internal ID: 2842
Software and algorithms		
RELION (version 3.0)	Scheres, 2012; Zivanov et al., 2018	https://github.com/3dem/relion
cryoSPARC	Punjani et al., 2017	https://www.cryosparc.com

(*Continued on next page*)

Please cite this article in press as: Farnung et al., Structure of a backtracked hexasomal intermediate of nucleosome transcription, Molecular Cell (2022), https://doi.org/10.1016/j.molcel.2022.06.027

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Warp (version 1.0.9)	Tegunov and Cramer, 2019	http://www.warpem.com
Phenix (version 1.19.2)	Adams et al., 2009; Afonine et al., 2018	http://www.phenix-online.org
PyMol (version 2.2.2)	Schrodinger, LLC, 2015	http://www.pymol.org
UCSF ChimeraX (version 1.3)	Goddard et al., 2018	https://www.cgl.ucsf.edu/chimerax/
Coot (version 0.9.5)	Emsley et al., 2010	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
ISOLDE (version 1.0)	Croll, 2018	https://isolde.cimr.cam.ac.uk/
ImageJ2 – Fiji (version 2.3)	Schindelin et al., 2012	https://imagej.net/software/fiji/
GraphPad Prism (version 9.3.1)		https://www.graphpad.com/scientific-software/prism/
Other		
Slide-A-Lyzer™MINI Dialysis Units (20,000 MWCO)	Thermo Scientific	69555
Model 491 Prep Cell	Bio-Rad	1702927
R 2/2 UltrAuFoil	Quantifoil	N1-A16nAu20

RESOURCE AVAILABILITY

Lead contact

Requests for materials should be directed and will be fulfilled by the lead contact, Patrick Cramer (patrick.cramer@mpinat.mpg.de).

Materials availability

Materials are available from the lead contact Patrick Cramer upon request under a material transfer agreement with the Max Planck Society.

Data and code availability

- The cryo-EM reconstructions and final models for the Pol II-DSIF-SPT6-PAF1-TFIIS-hexasome elongation complex, the Pol II-DSIF-SPT6-PAF1-TFIIS elongation complex, and the Pol II-TFIIS elongation complex have been deposited in the Electron Microscopy Data Base under ID codes EMD: EMD-15127, EMD-15128 and EMD-15129, respectively, and in the Protein Data Bank under ID codes 8A3Y, 8A3Z and 8A40 respectively. The raw gel data has been deposited at https://doi.org/10.17632/6s5pj5x2zj.1
- 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 MODELS AND SUBJECT DETAILS

Necessary plasmid progagation and bacmid isolation were performed in *Escherichia coli* DH5a (plasmid propagation) or *Escherichia coli* DH10 EMBacY (bacmid isolation). Recombinant protein expression was performed in BL21 (DE3) RIL. Recombinant protein expression in insect cells was performed using Sf9, Sf21, and Hi5 cell lines. Recombinant expressions in insect cells were harvested after 72 hours.

Sus scrofa RNA polymerase II was prepared from pig thymus.

METHOD DETAILS

Preparation of protein components

Sus scrofa RNA polymerase II, *Homo sapiens* DSIF, PAF, RTF1, TFIIS, TFIIS (D282A, E283A, DEAA), P-TEFb (CDK9 1-272, cyclin T1 1-372) and SPT6 were purified as previously described (Vos et al., 2018, 2020). PAF, RTF1, P-TEFb, and SPT6 were recombinantly expressed in insect cells. DSIF and TFIIS were recombinantly expressed in *E. coli*. DSIF contains SPT4 and N-terminally tagged SPT5 with a His10 tag and a 3C protease cleavage site. PAF contains CDC73, LEO1, PAF1, SKI8, and CTR9 with a C-terminal TEV protease cleavage site, MBP tag, and His6 tag. RTF1 carries an N-terminal His6 tag and TEV protease cleavage site. TFIIS and TFISS DEAA are N-terminally tagged with a His6 tag, Mocr tag, and TEV protease cleavage site. P-TEFb with CDK9 (residues 1-272, N-terminal His8 tag and TEV protease cleavage site), and cyclin T1 (residues 1-372, N-terminal GST tag and TEV protease cleavage site). SPT6 was N-terminally tagged with a His6 tag, MBP tag, and TEV protease cleavage site.

Please cite this article in press as: Farnung et al., Structure of a backtracked hexasomal intermediate of nucleosome transcription, Molecular Cell (2022), https://doi.org/10.1016/j.molcel.2022.06.027

Peak fractions for all column purification steps were assessed by SDS–PAGE and Coomassie staining.

The *S. scrofa* RNA polymerase II purification was performed with or without using a Sephacryl S-300 16/60 column (GE Healthcare Life Sciences) as the final gel filtration step. Pol II was concentrated in 100,000 MWCO Amicon Ultra Centrifugal Filters (Merck), aliquoted, snap-frozen, and stored at -80 °C until use.

Cell pellets for all purifications from recombinantly expressed proteins were lysed in lysis buffer with 20 mM Na HEPES pH 7.4, 10% (v/v) glycerol, 30 mM imidazole, 1 mM DTT, 0.284 μg ml^{-1} leupeptin, 1.37 μg ml^{-1} pepstatin A, 0.17 mg ml^{-1} PMSF and 0.33 mg ml^{-1} benzamidine, and 300 mM NaCl (SPT6, RTF1, TFIIS, TFIIS DEAA, PAF, P-TEFb) or 500 mM NaCl (DSIF) by sonication and cleared by centrifugation. The proteins were then applied to a 5 mL HisTrap column, equilibrated in lysis buffer. The column was washed with 10 CV lysis buffer, >3 CV high salt buffer (20 mM Na HEPES pH 7.4, 10% (v/v) glycerol, 30 mM imidazole, 1 mM DTT, 0.284 μg ml^{-1} leupeptin, 1.37 μg ml^{-1} pepstatin A, 0.17 mg ml^{-1} PMSF and 0.33 mg ml^{-1} benzamidine, and 1000 mM NaCl), and >3 CV lysis buffer.

For the DSIF purification, the nickel column was developed with nickel elution buffer (20 mM Na HEPES pH 7.4, 10% (v/v) glycerol, 500 mM imidazole, 1 mM DTT, 0.284 μg ml^{-1} leupeptin, 1.37 μg ml^{-1} pepstatin A, 0.17 mg ml^{-1} PMSF and 0.33 mg ml^{-1} benzamidine) including 500 mM NaCl. Fractions containing DSIF were pooled and mixed with 3C protease. DSIF was subsequently dialyzed in a Slide-A-Lyzer 10-kDa MWCO (ThermoFisher Scientific) and dialyzed overnight against Q buffer (20 mM Na HEPES pH 7.4, 10% (v/v) glycerol, 30 mM imidazole, 1 mM DTT, and 300 mM NaCl). The protein was applied to a tandem HisTrap HP 5 mL/HiTrap Q HP 5 mL, equilibrated in Q buffer. The columes were subsequently washed with Q buffer. The HisTrap column was removed, and DSIF was eluted using a gradient with Q buffer and high salt buffer (no protease inhibitors). Fractions containing DSIF were pooled, concentrated using a 50,000 MWCO Amicon Ultra Centrifugal Filters (Merck), and applied to a HiLoad S200 16/600 pg column equilibrated in SE buffer 500 (300 mM NaCl, 20 mM Na HEPES pH 7.4, 10% (v/v) glycerol and 1 mM DTT/TCEP).

For the P-TEFb purification, the nickel column was washed with lysis buffer with 150 mM NaCl. A HiTrap S HP column was attached and P-TEFB was eluted using nickel elution buffer with 150 mM NaCl. The nickel column was removed. The HiTrap S HP column was washed with lysis buffer with 150 mM NaCl. The protein was eluted in a gradient using high salt buffer.

For RTF1, TFIIS, and TFIIS DEAA, the nickel column was developed with nickel elution buffer with 300 mM NaCl.

For SPT6, the MBP-tagged SPT6 was eluted from the nickel column onto an amylose column using nickel elution buffer. The amylose column was washed with lysis buffer. Subsequently, SPT6 was eluted by amylose elution buffer (lysis buffer + 117 mM maltose).

For PAF, the nickel column was washed with low salt buffer (lysis buffer with 150 mM NaCl). A HiTrapQ column was attached to the nickel column. PAF was eluted onto the HiTrapQ column using nickel elution buffer with 150 mM NaCl. The HisTrap column was removed and washed with low salt buffer. PAF was then eluted using a gradient from low salt buffer to high salt buffer over 9 CV.

For the P-TEFb, RTF1, TFIIS, SPT6, and PAF purifications, fractions containing the respective proteins were pooled and mixed with 1.5 mg TEV protease. All proteins were subsequently dialyzed in a Slide-A-Lyzer 7 kDa MWCO (ThermoFisher Scientific) and dialyzed overnight against lysis buffer. Dialyzed and cleaved protein was then applied to a 5-ml HisTrap column equilibrated in Lysis buffer to remove uncleaved protein, the His or His-MBP tag and TEV protease.

For the P-TEFb purification, the flow-through was subsequently incubated with glutathione beads before size exclusion chromatography. The sample was subsequently concentrated using a 10,000 MWCO Amicon Ultra Centrifugal Filter (Merck), and applied to a HiLoad S200 16/600 pg column equilibrated in SE buffer 300 (300 mM NaCl, 20 mM Na HEPES pH 7.4, 10% (v/v) glycerol and 1 mM DTT).

For the RTF1 purification, the flowthrough was collected and concentrated in a 30,000 MWCO Amicon Ultra Centrifugal Filter. RTF1 was then applied to a HiLoad S200 16/600 pg column equilibrated in SE buffer 300 (300 mM NaCl, 20 mM Na HEPES pH 7.4, 10% (v/v) glycerol and 1 mM DTT).

For the TFIIS and TFIIS DEAA purification, the flowthrough was collected and concentrated in a 10,000 MWCO Amicon Ultra Centrifugal Filter. RTF1 was then applied to a HiLoad S75 16/600 pg column equilibrated in SE buffer 300 (300 mM NaCl, 20 mM Na HEPES pH 7.4, 10% (v/v) glycerol and 1 mM DTT).

For the SPT6 purification, the flowthrough was collected and concentrated in a 50,000 MWCO Amicon Ultra Centrifugal Filter. SPT6 was then applied to a HiLoad S200 16/600 pg column equilibrated in SE buffer 300.

For the PAF purification, the flowthrough was collected and concentrated in a 100,000 MWCO Amicon Ultra Centrifugal Filter. PAF was then applied to a Superose 6 Increase 10/300 column equilibrated in SE buffer 300.

After gel filtration, the fractions containing the protein(s) of interest were pooled, concentrated using Amicon Ultra Centrifugal Filters, aliquoted, snap-frozen, and stored at -80 °C until use.

Nucleosome preparation

Xenopus laevis histones were expressed and purified as described (Dyer et al., 2004; Farnung et al., 2017). DNA fragments for nucleosome reconstitution were generated by PCR. A vector containing the modified Widom 601 sequence was used as a template for PCR. Super-helical locations are assigned based on previous publications (Farnung et al., 2021), assuming direction of transcription from negative to positive SHLs. Large-scale PCR reactions were performed with two PCR primers (forward primer: ACG AAG CGT AGC ATC ACT GTC TTG, reverse primer: ATC AGA ATC CCG GTG CCG) at a scale of 50 mL. This yielded the following PCR product: ACG AAG CGT AGC ATC ACT GTC TTG TGT TTG GTG TGT CTG GGT GGT GGC CGT TTT CGT TGT TTT TTT CTG TCT CGT GCC

Please cite this article in press as: Farnung et al., Structure of a backtracked hexasomal intermediate of nucleosome transcription, Molecular Cell (2022), https://doi.org/10.1016/j.molcel.2022.06.027

Molecular Cell
Article

CellPress

TGG TGT CTT GGG TGT AAT CCC CTT GGC GGT TAA AAC GCG GGG GAC AGC GCG TAC GTG CGT TTA AGC GGT GCT AGA GCT GTC TAC GAC CAA TTG AGC GGC CTC GGC ACC GGG ATT CTG AT. PCR product purification, TspRI digest, nucleosome reconstitution and subsequent purification with a PrepCell were performed as described (Farnung et al., 2021). This substrate was used for the structural and biochemical investigation of the EC*-hexasome complex. Quantification of the reconstituted nucleosome was achieved by measuring absorbance at 280 nm. Molar extinction coefficients were determined for protein and nucleic acid components and were summed to yield a molar extinction coefficient for the reconstituted nucleosome.

RNA extension assays

RNA extension assays were performed on the same nucleosomal template substrate used for the structural studies of the EC*-hexasome complex. As previously described (Farnung et al., 2021), a 6-FAM 5′-labelled 11-nt RNA (5′-/56-FAM/ rUrUrA rUrCrA rCrUrG rUrC-3′) was utilized to track the location of Pol II during transcription. All subsequent concentrations refer to the concentration in the final reaction. The final concentrations of buffer components were 100 mM NaCl, 20 mM Na·HEPES, pH 7.4 (@ 20°C), 3 mM $MgCl_2$, 4% (v/v) glycerol and 1 mM DTT. The final volume for each RNA extension reaction was 50 μL. RNA (120 nM) and nucleosomal template (120 nM) were mixed in equimolar ratios and incubated for 5 min on ice. *S. scrofa* Pol II (150 nM) was added to the mixture and incubated for 5 min on ice. Depending on the assay condition, DSIF, SPT6, PAF, RTF1 (each 225 nM) and P-TEFb (250 nM), 2× compensation buffer, water and ATP (1 mM) were added to the reaction and incubated for 15 min at 30°C. Transcription elongation was started by the addition of CTP, GTP and UTP (1 mM each) and depending on the assay condition, either wild-type TFIIS or the mutant TFIIS (DEAA) (90 nM) were added to the reaction. Five microliters of the reactions were quenched after 10 s, 30 s, 1 min, 5 min, 10 min, 30 min and 60 min in 5 μl 2× stop buffer (6.4 M urea, 50 mM EDTA, pH 8.0, 1× TBE buffer). Samples were treated with 40 μg proteinase K for 30 min at 37°C, denatured at 95°C for 10 min, and separated by denaturing gel electrophoresis. 3 μl of sample were applied to an 8 M urea, 1× TBE buffer, 12% acrylamide:bis-acrylamide 19:1 gel, run in 0.5× TBE buffer at 300 V for 35 min. RNA extension products were visualized using the 6-FAM label and a Typhoon 9500 FLA imager at an excitation wavelength of 473 nm and emission wavelength range of >520 nm.

All RNA extension assays were performed independently at least three times. RNA extension products of interest (74 nt, 84 nt, 106 nt and full-length) were quantified using Fiji's integrated density values (Schindelin et al., 2012). The signal intensity from the products of the RNA extension assays (box size 0.62 x 0.16 cm^2) were normalized by subtracting the background, followed by normalization against the total intensity of the respective reaction lane (box size 0.62 x 6.45 cm^2) and by normalization against the sum from the signal of the respective lane across all reactions. The normalization of the signals was performed to control for errors during gel loading and other external/non-transcription related factors affecting the signal intensity. Graphs were prepared in Prism 9. Bars and points in the graphs represent mean values while error bars show the standard deviation between experiments. The following P values were applied *$P < 0.05$, **$P < 0.01$. A unpaired two-tailed t test was employed to determine statistical significance between conditions.

TFIIS cleavage-stimulatory assay

TFIIS cleavage assays were performed on the same nucleosomal template and with the same RNA used for the RNA extension assays. All subsequent concentrations refer to the concentration in the initial RNA extension reaction prior to size exclusion chromatography. The final concentrations of buffer components were 100 mM NaCl, 20 mM Na·HEPES, pH 7.7 (@ 20°C), 3 mM $MgCl_2$, 4% (v/v) glycerol and 1 mM DTT. The final volume for the initial RNA extension reaction was 50 μl. RNA (20 μM) and nucleosomal template (1.04 μM) were mixed and incubated for 5 min on ice. *S. scrofa* Pol II (1.04 μM) was added to the mixture and incubated for 5 min on ice. DSIF, SPT6, PAF, RTF1 (each 1.56 μM), P-TEFb (250 nM), 5× compensation buffer, water and 3'-dATP (400 μM) were added and incubated for 15 min at 30°C. Transcription elongation was started by the addition of CTP, GTP and UTP (1 mM each), allowing transcription of the T-less cassette to nucleosomal base pair 42. The transcription reaction was incubated for 30 min at 30°C. The samples were subsequently centrifuged and applied onto a Superose 6 3.2/300 Increase size exclusion column (GE Healthcare), equilibrated in complex buffer (100 mM NaCl, 20 mM Na·HEPES, pH 7.4 (@ 20°C), 3 mM $MgCl_2$, 1 mM TCEP, 4% (v/v) glycerol). Fractions eluted from the size exclusion column containing the EC*-nucleosome complex were verified using SDS−PAGE. The concentration of the EC*-nucleosome complex after gel filtration was calculated using the molar extinction coefficient at 280 nm and the measured absorbance at this wavelength. The concentration varied between 88.6 – 90.5 nM.

The complex of EC* transcribed to nucleosomal base pair 42 was used for RNA cleavage assays. Complex buffer, in complex buffer diluted mutant TFIIS (DEAA) or wild-type TFIIS (139.2 nM) was mixed with the EC*-nucleosome complex (69.6 nM) to a final volume of 13 μL. The cleavage reaction was incubated at 20°C. Three microliters of the reactions were quenched after 1 min, 5 min, 15 min in 3 μl 2× stop buffer (6.4 M urea, 50 mM EDTA, pH 8.0, 1× TBE buffer). Samples were treated with 40 μg proteinase K for 30 min at 37°C, denatured at 95°C for 10 min, and separated by denaturing gel electrophoresis. Specifically, 164 fmol were applied to an 8 M urea, 1× TBE buffer, 12% acrylamide:bis-acrylamide 19:1 gel, run in 0.5× TBE buffer at 300 V for 35 min. RNA extension products were visualized as described in the RNA extension assay using the 6-FAM label and a Typhoon 9500 FLA imager at an excitation wavelength of 473 nm and emission wavelength range of >520 nm.

The TFIIS cleavage-stimulatory assay was performed independently three times. The signal intensities from the uncleaved (84 nt; box size 0.45 x 0.23 cm^2) and cleaved products (74 nt; box size 0.45 x 0.80 cm^2) of the RNA cleavage assay were normalized by subtracting the background before calculating the cleavage ratio by dividing the uncleaved RNA signal from the cleaved signal.

Please cite this article in press as: Farnung et al., Structure of a backtracked hexasomal intermediate of nucleosome transcription, Molecular Cell (2022), https://doi.org/10.1016/j.molcel.2022.06.027

Graphs were prepared in Prism 9. Bars and points in the graphs represent mean values while error bars show the standard deviation between experiments. The following P values were applied *P < 0.05, **P < 0.01. A unpaired two-tailed t test was employed to determine statistical significance between conditions.

Sample preparation for cryo-EM

EC*-nucleosome complex for cryo-EM was assembled in a final buffer containing 100 mM NaCl, 20 mM Na·HEPES, pH 7.4 (@ 20°C), 3 mM $MgCl_2$, 1 mM DTT and 4% (v/v) glycerol in a total volume of 100 μL. Complex formation was performed as described for the transcription assay with final concentrations of nucleosome (744 nM), Pol II (744 nM), DSIF (1120 nM), SPT6 (1120 nM), PAF (1120 nM), RTF1 (1120 nM), and P-TEFb (100 nM). Water, compensation buffer, and 0.4 mM 3'-dATP were added to allow for P-TEFb phosphorylation. The reaction was incubated for 20 min at 30°C. The transcription reaction was started by the addition of 1mM of each CTP, GTP, UTP, and TFIIS (450 nM). The reaction was incubated for 15 min at 30°C. 0.5 μL 0.5 M EDTA were added to quench the reaction. The sample was then centrifuged and applied onto a Superose 6 3.2/300 Increase size exclusion column (GE Healthcare), equilibrated in complex buffer (100 mM NaCl, 20 mM Na·HEPES, pH 7.4 (@ 20°C), 3 mM $MgCl_2$, 1 mM DTT, 4% (v/v) glycerol). Eluted fractions were analysed using SDS-PAGE and denaturing RNA gels.

EC*-TFIIS for cryo-EM was assembled in a final buffer containing 100 mM NaCl, 20 mM Na·HEPES, pH 7.4 (@ 20°C), 3 mM $MgCl_2$, 1 mM DTT and 4% (v/v) glycerol. RNA (5′-/56-FAM/ rUrUrA rArGrG rArArU rUrArA rGrUrC rGrUrG rCrGrU rCrUrA rArUrA rArCrCrG rGrArG rArGrG rGrArA rCrCrC rArCrU-3′) and template DNA (5′-GCT CCC AGC TCC CTG CTG GCT CCG AGT GGG TTC TGC CGC TCT CAA TGG-3′) were annealed to form a hybrid (50 μM) using the thermocycler. All following concentrations refer to final concentrations in 100 μL complex formation reaction. Pol II (1000 nM) and hybrid (3000 nM) were incubated for 15 min on ice. Non-template DNA (5′-CCA TTG AGA GCG GCC CTT GTG TTC AGG AGC CAG CAG GGA GCT GGG AGC-3′) (5000 nM) was added, and the sample incubated for 10 min on ice. DSIF (2000 nM), SPT6 (2000 nM), PAF (2000 nM), RTF1 (2000 nM), and P-TEFb (100 nM), TFIIS (DEAA) (3000 nM), water, and a 10-fold compensation buffer was added to reach the final buffer conditions, and the sample was incubated for 20 min at 30°C after addition of 1 mM ATP. The sample was subsequently centrifuged and applied onto a Superose 6 3.2/300 Increase size exclusion column (GE Healthcare), equilibrated in complex buffer (100 mM NaCl, 20 mM Na·HEPES, pH 7.4 (@ 20°C), 3 mM $MgCl_2$, 1 mM DTT, 4% (v/v) glycerol). Eluted fractions were analysed using SDS-PAGE.

EC-TFIIS for cryo-EM was assembled in a final buffer containing 100 mM NaCl, 20 mM Na·HEPES, pH 7.4 (@ 20°C), 3 mM $MgCl_2$, 1 mM DTT and 4% (v/v) glycerol. RNA, template DNA and non-template DNA are identical to the substrate used for the EC*-TFIIS complex. RNA and template DNA were annealed to form a hybrid (25 μM) using the thermocycler. Pol II (1240 nM) and hybrid (1860 nM) were incubated for 15 min on ice. Non-template DNA (2230 nM) was added, and the sample incubated for 10 min on ice. TFIIS (DEAA) (6200 nM), water, and a 10-fold compensation buffer was added to reach the final buffer conditions, and the sample further incubated for 30 min at 30°C. The sample was then centrifuged and applied onto a Superose 6 3.2/300 Increase size exclusion column (GE Healthcare), equilibrated in complex buffer (100 mM NaCl, 20 mM Na·HEPES, pH 7.4 (@ 20°C), 3 mM $MgCl_2$, 1 mM DTT, 4% (v/v) glycerol). Eluted fractions were analysed using SDS-PAGE.

Peak fractions of all complexes were individually cross-linked using 0.1% (v/v) glutaraldehyde for 10 min on ice. The cross-linker was quenched with 8 mM apartate and 2 mM lysine and incubated for 10 min on ice. Fractions were then dialyzed against the dialysis buffer (100 mM NaCl, 20 mM Na·HEPES, pH 7.4 (@ 20°C), 3 mM $MgCl_2$, 1 mM DTT) for 4 hours.

For grid preparation, R2/2 gold grids, Au 200 mesh (Quantifoil) were glow-discharged twice for 100 s. 2 μL of the dialyzed complexes with an approximate concentration of 120-160 nM (EC*-nucleosome and EC*-TFIIS) and 180-200 nM (EC-TFIIS) were applied to each side of the grid. For the EC*-TFIIS preparation, an additional equimolar amount of TFIIS (DEAA) (compared to complex concentration) was added to the sample immediately before freezing. After an incubation of 10 s, grids were blotted and vitrified by plunging into liquid ethane using a Vitrobot Mark IV (Thermo Fisher). The Vitrobot was operated at 4°C and 100% humidity, and the blotting done using blot force of 5 and blot time of 4.5-8 s. The grids were clipped and subsequently stored in liquid nitrogen before data acquisition.

Cryo-EM analysis and image processing

Cryo-EM data was acquired on a Titan Krios G2 transmission electron microscope (FEI) operated at 300 keV, equipped with a K3 summit direct detector and a GIF BioQuantum energy filter (Gatan).

Data acquisition for the EC*-nucleosome data set was performed using Serial EM at a nominal magnification of 81,000×, corresponding to a pixel size of 1.05 Å/pixel in nanoprobe EF-TEM mode. A total of 4 datasets were collected. A total of 36,000 movie stacks of 40 frames were collected in counting mode over 2-2.2 s, with a defocus range of 0.5 to 2 μm. The dose rate was ~17-19 e^- per $Å^2$ per s for a total dose of ~36-40 e^- per $Å^2$.

Data acquisition for the EC*-TFIIS data set was performed using Serial EM at a nominal magnification of 81,000×, corresponding to a pixel size of 1.05 Å/pixel in nanoprobe EF-TEM mode. A total of 12,787 movie stacks of 40 frames were collected in counting mode over 2.2 s, with a defocus range of 0.5 to 2 μm. The dose rate was ~17.6 e^- per $Å^2$ per s for a total dose of 38.9 e^- per $Å^2$.

Data acquisition for the Pol II-TFIIS EC was performed using SerialEM at a nominal magnification of 105,000×, corresponding to a pixel size of 0.834 Å/pixel in nanoprobe EF-TEM mode. A total of 14,601 movie stacks of 40 frames were collected in counting mode over 2.2 s, with a defocus range of 0.5 μm to 4.0 μm. The dose rate was ~18.12 e^- per $Å^2$ per s for a total dose of 39.87 e^- per $Å^2$.

Please cite this article in press as: Farnung et al., Structure of a backtracked hexasomal intermediate of nucleosome transcription, Molecular Cell (2022), https://doi.org/10.1016/j.molcel.2022.06.027

Micrographs for all data sets were stacked, the contrast transfer function (CTF) was estimated, and the micrographs were dose-weighted using Warp (Tegunov and Cramer, 2019). Particles were picked using in-house trained instances of the neural network BoxNet2 as implemented in Warp. Subsequent image processing was done in RELION 3.0.7 (Scheres, 2012; Zivanov et al., 2018) unless otherwise mentioned.

Global refinements/classifications refer to refinement/classifications that encompass a spherical mask or a user-generated mask that encompasses the whole complex. Masked classifications/refinements refer to refinements/classifications that only encompass a certain area of the complex using a user-generated mask. The area of masking is indicated.

For the EC*-nucleosome dataset a total of 9,819,841 particles were picked from the four datasets. Particles were extracted with a box size of 360 pixels. The extracted particles were initially imported into cryoSPARC for 2D classification to remove non-aligning particles using 2D classification. Subsequently multiple rounds of heterogeneous refinements were performed. The remaining 1,992,107 particles were exported to RELION 3.1.0. To further select particles containing both the PAF1 complex and nucleosome, masked classifications were performed with masks encompassing CTR9 and WDR61, the nucleosome or both. The remaining 32,933 particles were subsequently refined and a reconstruction was obtained at 3.6 Å resolution. EC* densities except for the nucleosome were subtracted from the particles, and a masked refinement of the nucleosome was performed to obtain better resolution for the hexasome.

For the EC*-TFIIS EC dataset a total of 1,713,362 particles were picked. Particles were extracted with a box size of 360 pixels. The extracted particles were initially imported into cryoSPARC for two rounds of 3D classification and subsequent removal of non-aligning particles using 2D classification. The remaining 532,036 particles were exported to RELION 3.1.0. To further select particles containing both the PAF1 complex and TFIIS, masked classifications were performed with masks encompassing CTR9 and WDR61. The remaining 51,218 particles were subsequently refined, and a reconstruction was obtained at 3.8 Å resolution.

For the Pol II-TFIIS EC dataset a total of 3,617,066 particles were picked. Particles were extracted with a box size of 300 pixels (twice binned to 150 pixels). Multiple rounds of 3D classification were used for initial data cleaning and resulted in 1,204,375 particles. 3D refinement of these particles against an initial model generated in cryoSPARC (Punjani et al., 2017) resulted in an overall map of 3.01 Å. After un-binning the data, the particles were CTF refined. Two subsequent rounds of classification were performed to select for Pol II-TFIIS, which resulted in 339,432 particles. 3D refinement of these particles yielded a map with a resolution of 2.9 Å. The map showed excellent density for Pol II and TFIIS. To improve TFIIS domain III density, a masked classification on TFIIS was performed. The resulting 82,497 particles were used for a final refinement with an overall resolution of 3.1 Å.

Local resolutions of relevant maps were estimated using the RELION built-in tool.

Model building

For the EC*-hexasome structure, structures of EC* (PDB code 6TED), nucleosome (PDB code 3LZ0) were rigid body docked into the final map and refined using COOT (Emsley et al., 2010). The proximal H2A–H2B dimer was removed. The atomic model was real-space refined using PHENIX (Adams et al., 2010; Afonine et al., 2018) with secondary structure restraints against the final local-resolution filtered map. Density in the active site of Pol II allowed unambiguous assignment of DNA and DNA-RNA register. The atomic model was real-space refined using PHENIX with secondary structure restraints.

For the EC*-TFIIS structure, the activated elongation complex (PDB code 6TED) and *H. sapiens* TFIIS (PDB code 5IYB) were rigid body docked and real space refined using ISOLDE (Croll, 2018) and COOT. SPT6 was modeled using PDB code 7OPC after removing residues 515-530. TFIIS domain II and the interdomain linker were modelled after docking. TFIIS domain III was removed due to lack of high-resolution features. The model was refined in PHENIX with secondary structure restraints applied.

For the Pol II-TFIIS structure, structures of *B. taurus* activated Pol II (PDB code 5OIK), *H. sapiens* TFIIS (PDB code 5IYB), and nucleic acid (PDB code 6GML) were rigid body docked into the global refinement and refined using COOT. Pol II core and shelf modules were rigid body docked and modelled separately to identify potential conformational changes upon TFIIS binding. TFIIS domain II and the interdomain linker were modelled into the density after docking, while domain III was not due to lack of resolution. The atomic model was real-space refined using PHENIX with secondary structure restraints against the global EM refinement.

Figure generation

Figures and movie were generated using The PyMOL Molecular Graphics System, Version 2.2.2 Schrödinger, LLC., and UCSF ChimeraX (version 1.3) (Goddard et al., 2018).

Movie generation

The movie was generated with USCF ChimeraX (version 1.3) using the structure of mammalian EC* (PDB 6TED) and *X. laevis* nucleosome (PDB 3LZ0).

The movie was created by morphing between a large number of interpolated models. The last model presented in the movie shows the resolved EC*-hexasome structure.

The movie also features an animation of the H2A–H2B dimer loss observed in the EC*-hexasome structure. However, it is unknown at what exact location the dimer dissociates from the nucleosome. The movie shows this event once the dimer loses its last interaction with DNA. Nevertheless, the dissociation could also occur at a different position prior or during backtracking. Possible positions for H2A–H2B dimer loss are indicated in the movie.

Please cite this article in press as: Farnung et al., Structure of a backtracked hexasomal intermediate of nucleosome transcription, Molecular Cell (2022), https://doi.org/10.1016/j.molcel.2022.06.027

The effect of movement for the transcribing linear DNA was created by morphing three DNA models with an identical sequence by an offset of 1 bp. The rotation and DNA unwrapping of the nucleosome in front of the EC* was generated by aligning the entry DNA of the nucleosome with the linear DNA downstream of the polymerase for each transcribed base pair. During the nucleosome passage part of the movie, where the nucleosome is slowly unwrapped, the nucleosome was aligned to the linear DNA using two separate parameters; a) the "distance parameter" by defining the reference segment on the linear DNA for alignment of the nucleosome, and b) the "degree of unwrapping" of the nucleosome by defining the segment of the nucleosomal DNA used for alignment. The optimal set of parameters were determined by screening different combinations and choosing the model with the least unwrapped nucleosome without clashes with the EC*.

With the nucleosome approaching the polymerase, the linear DNA was gradually shortened to create the continuous appearance of DNA moving into the polymerase for transcription.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of signal from denaturing PAGE gels were performed using the integrated density function of ImageJ2 2.3.0. Statistical analysis of these quantified intensities were performed in Excel 16.62 and Prism 9.3.1.

The exact quantifications and statistical analysis are described in the STAR Methods section.

All graphs and diagrams derived from quantification of denaturing PAGE were generated in Prism and display mean values while the error bars represent the standard deviation SD. The P-values for statistical significance were determined by unpaired two-tailed t tests with Welch's correction in Prism. The number of experimental replicates are described in the figure legends.

Molecular Cell, Volume *82*

Supplemental information

Structure of a backtracked hexasomal intermediate of nucleosome transcription

Lucas Farnung, Moritz Ochmann, Gaurika Garg, Seychelle M. Vos, and Patrick Cramer

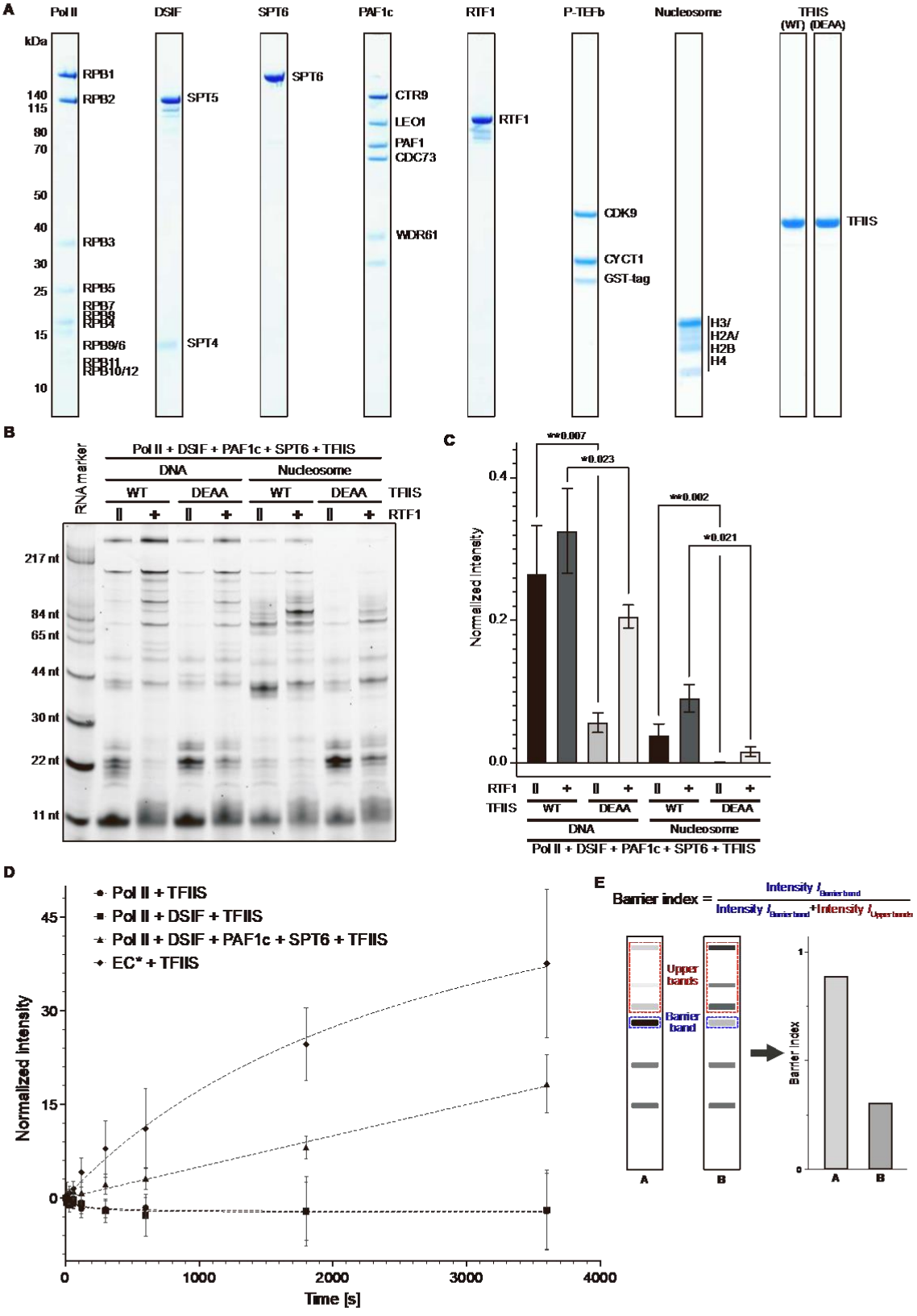

A
Pol II
DSIF
SPT6
PAF1c
RTF1
P-TEFb
Nucleosome
TFIIS (WT) (DEAA)
kDa
RPB1
RPB2
RPB3
RPB5
RPB7
RPB8
RPB4
RPB9/6
RPB11
RPB10/12
SPT5
SPT4
SPT6
CTR9
LEO1
PAF1
CDC73
WDR61
RTF1
CDK9
CYCT1
GST-tag
H3/
H2A/
H2B
H4
TFIIS
B
RNA marker
Pol II + DSIF + PAF1c + SPT6 + TFIIS
DNA
Nucleosome
WT
DEAA
TFIIS
RTF1
217 nt
84 nt
65 nt
44 nt
30 nt
22 nt
11 nt
C
Normalized Intensity
RTF1
TFIIS
Pol II + DSIF + PAF1c + SPT6 + TFIIS
D
Pol II + TFIIS
Pol II + DSIF + TFIIS
Pol II + DSIF + PAF1c + SPT6 + TFIIS
EC* + TFIIS
Normalized intensity
Time [s]
E
Barrier index =
Upper bands
Barrier band
Barrier index

Figure S1. Protein components and transcription assays. Related to Figure 1 and 4.

(A) SDS-PAGE of all purified protein components (0.75 µg proteins per lane).

(B) Transcription assay of the EC* with and without RTF1 on either linear or a nucleosomal template DNA in the presence of wild-type TFIIS or a catalytically dead TFIIS variant (DEAA).

(C) Quantification of the full-length RNA product of the transcription assay depicted in (B) (n = 4). The full-length RNA signal was determined by measuring the integrated signal density, subtracting the background and normalizing the value through the signal of the lane and the gel (Methods). Mean ± SD values are shown. P-values were determined using unpaired two-tailed t tests (*P < 0.05, **P < 0.01).

(D) Quantification of the full-length RNA product from transcription assay time courses of Pol II elongation complexes on a nucleosomal template in the presence of different combinations of elongation factors DSIF, TFIIS, SPT6, PAF1, RTF1 (n = 3). Mean ± SD values are shown. P-values were determined using unpaired two-tailed t tests (*P < 0.05, **P < 0.01).

(E) Schematic of the principle and calculation of the barrier index.

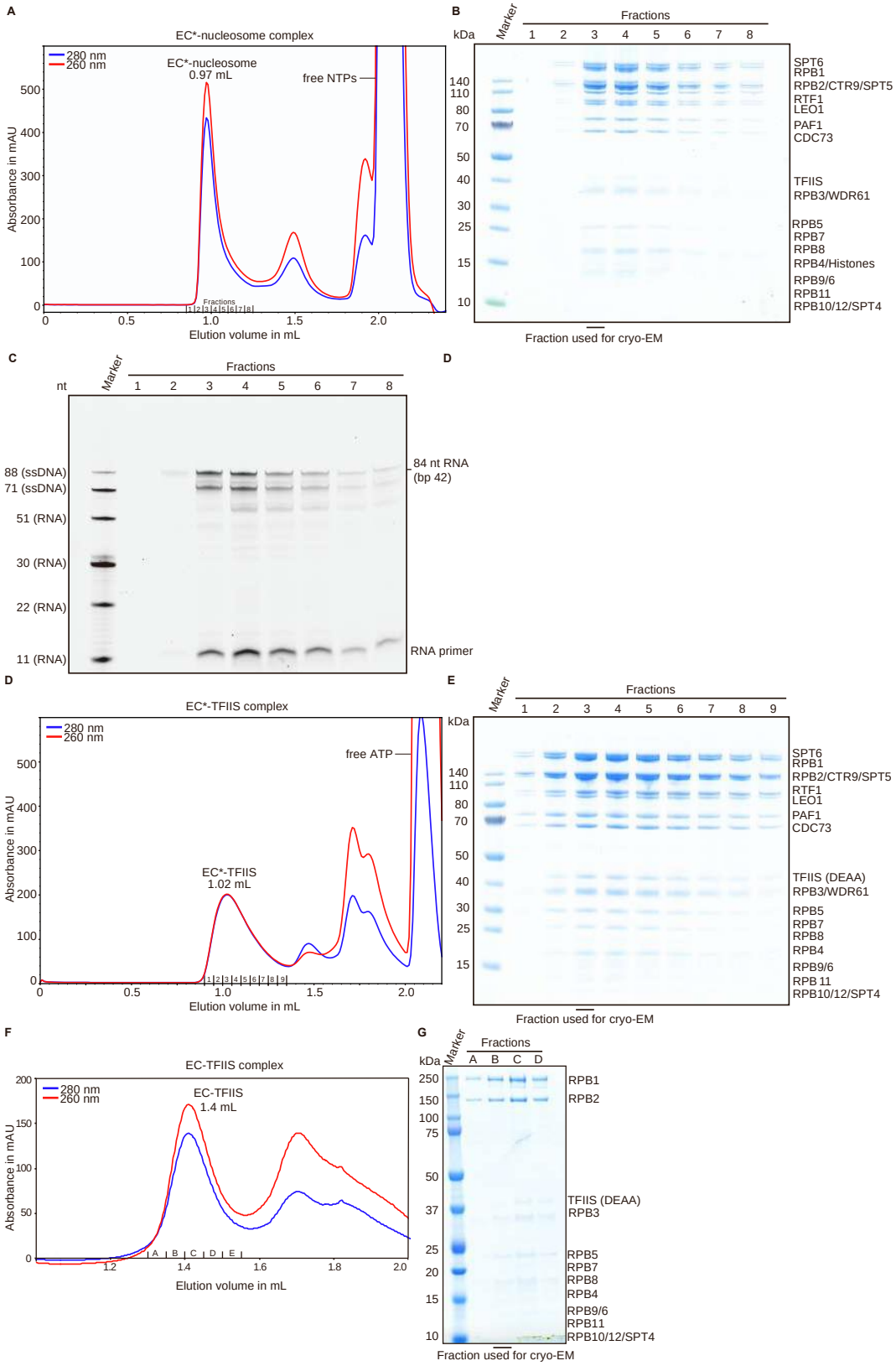

A
EC*-nucleosome complex
280 nm
260 nm
EC*-nucleosome
0.97 mL
free NTPs
Absorbance in mAU
Elution volume in mL
Fractions
B
kDa
Marker
Fractions
1 2 3 4 5 6 7 8
140
110
80
70
50
40
30
25
15
10
SPT6
RPB1
RPB2/CTR9/SPT5
RTF1
LEO1
PAF1
CDC73
TFIIS
RPB3/WDR61
RPB5
RPB7
RPB8
RPB4/Histones
RPB9/6
RPB11
RPB10/12/SPT4
Fraction used for cryo-EM
C
nt
Marker
Fractions
1 2 3 4 5 6 7 8
88 (ssDNA)
71 (ssDNA)
51 (RNA)
30 (RNA)
22 (RNA)
11 (RNA)
84 nt RNA
(bp 42)
RNA primer
D
EC*-TFIIS complex
280 nm
260 nm
free ATP
EC*-TFIIS
1.02 mL
Absorbance in mAU
Elution volume in mL
E
Marker
Fractions
1 2 3 4 5 6 7 8 9
kDa
140
110
80
70
50
40
30
25
15
SPT6
RPB1
RPB2/CTR9/SPT5
RTF1
LEO1
PAF1
CDC73
TFIIS (DEAA)
RPB3/WDR61
RPB5
RPB7
RPB8
RPB4
RPB9/6
RPB 11
RPB10/12/SPT4
Fraction used for cryo-EM
F
EC-TFIIS complex
280 nm
260 nm
EC-TFIIS
1.4 mL
Absorbance in mAU
Elution volume in mL
G
kDa
Marker
Fractions
A B C D
250
150
100
75
50
37
25
20
15
10
RPB1
RPB2
TFIIS (DEAA)
RPB3
RPB5
RPB7
RPB8
RPB4
RPB9/6
RPB11
RPB10/12/SPT4
Fraction used for cryo-EM

Figure S2. Formation of complexes. Related to Figure 2, 5 and 6.

(A) Chromatogram of EC*-nucleosome complex formation (Superose 6 Increase 3.2/300).

(B) SDS-PAGE of EC*-nucleosome complex formation.

(C) Denaturing gel of EC*-nucleosome complex formation.

(D) Chromatogram of EC*-TFIIS complex formation (Superose 6 Increase 3.2/300).

(E) SDS-PAGE of EC*-TFIIS complex formation.

(F) Chromatogram of Pol II-TFIIS elongation complex formation (Superose 6 Increase 3.2/300).

(G) SDS-PAGE of Pol II-TFIIS elongation complex formation.

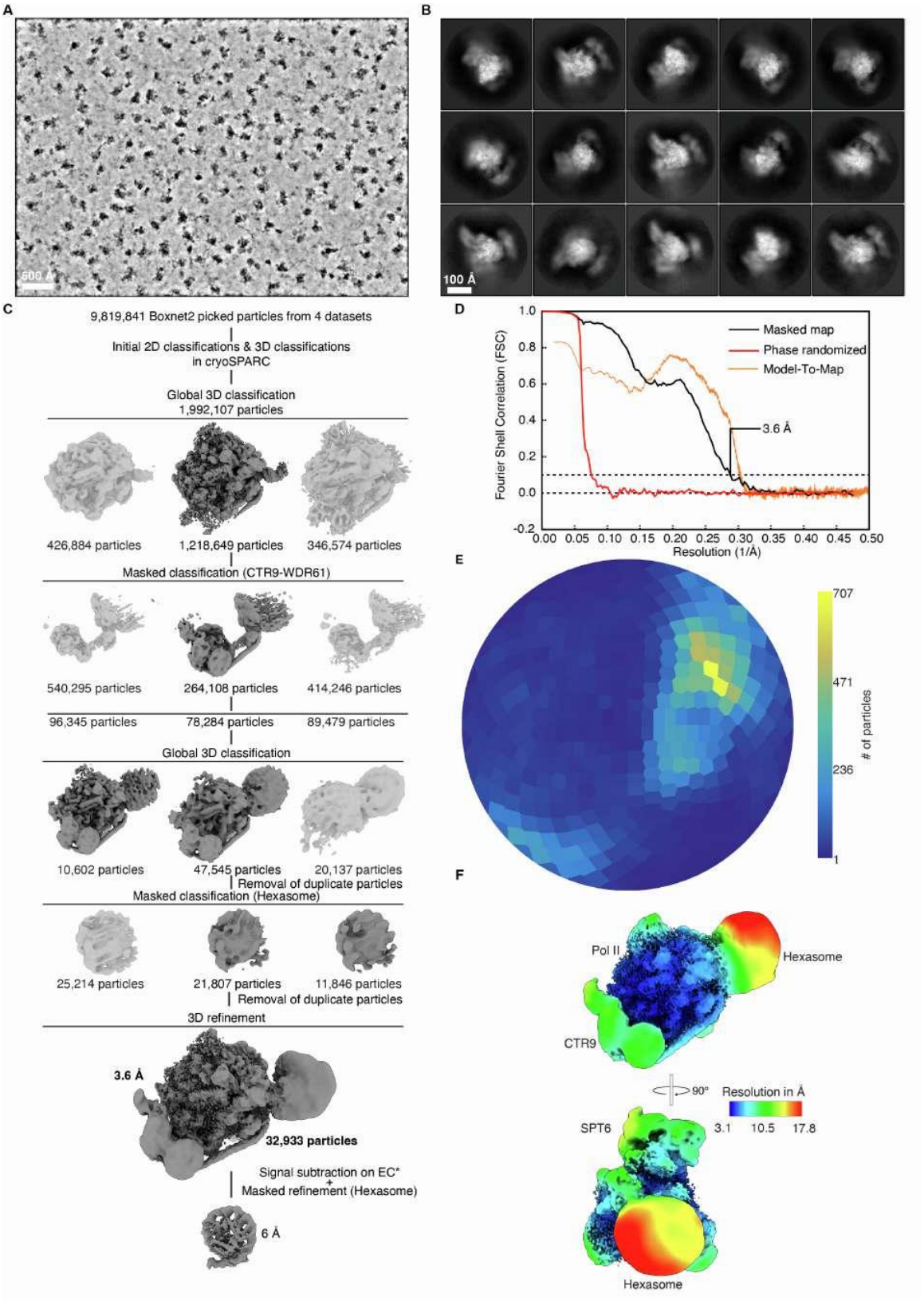

A
500 Å
B
100 Å
C
9,819,841 Boxnet2 picked particles from 4 datasets
Initial 2D classifications & 3D classifications in cryoSPARC
Global 3D classification
1,992,107 particles
426,884 particles
1,218,649 particles
346,574 particles
Masked classification (CTR9-WDR61)
540,295 particles
264,108 particles
414,246 particles
96,345 particles
78,284 particles
89,479 particles
Global 3D classification
10,602 particles
47,545 particles
20,137 particles
Removal of duplicate particles
Masked classification (Hexasome)
25,214 particles
21,807 particles
11,846 particles
Removal of duplicate particles
3D refinement
3.6 Å
32,933 particles
Signal subtraction on EC*
+
Masked refinement (Hexasome)
6 Å
D
Masked map
Phase randomized
Model-To-Map
3.6 Å
Fourier Shell Correlation (FSC)
Resolution (1/Å)
E
707
471
236
1
# of particles
F
Pol II
Hexasome
CTR9
90°
Resolution in Å
3.1
10.5
17.8
SPT6
Hexasome

Figure S3. Data acquisition, processing, and data quality metrics for the EC*-hexasome complex. Related to Figure 2.

(A) Representative denoised micrograph of data collection with scale bar (50 nm).

(B) 2D classes of final refinement show RNA polymerase II, transcription elongation factors, and nucleosome-like shape with scale bar of 10 nm.

(C) Sorting and classification tree of EC*-hexasome complex dataset.

(D) FSC curve of overall map. Resolution at FSC threshold criterion 0.143 is indicated.

(E) Angular distribution of particles employed to reconstruct EC*-hexasome complex.

(F) Local resolution of EC*-hexasome complex.

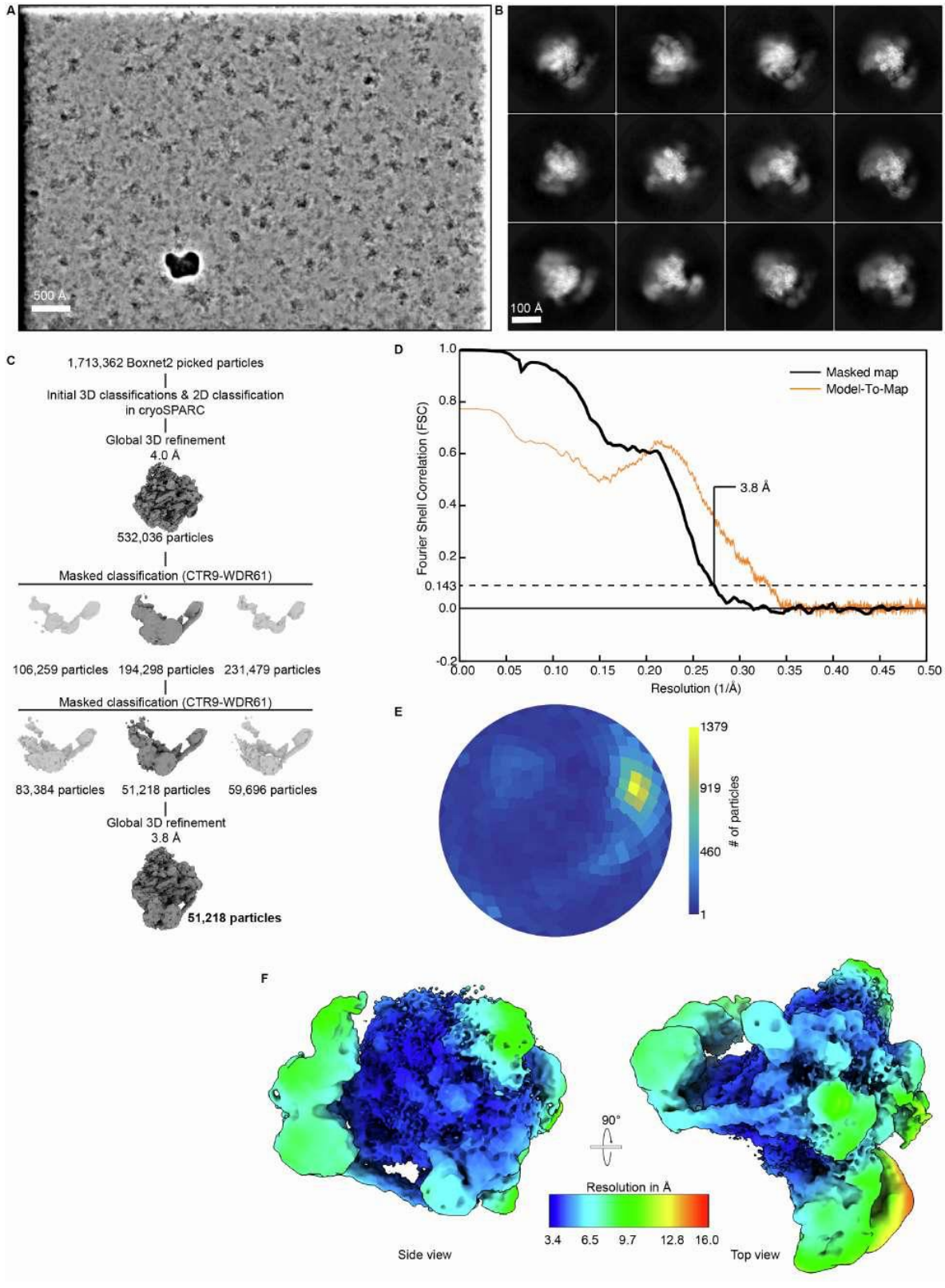

A
500 Å
B
100 Å
C
1,713,362 Boxnet2 picked particles
Initial 3D classifications & 2D classification in cryoSPARC
Global 3D refinement
4.0 Å
532,036 particles
Masked classification (CTR9-WDR61)
106,259 particles
194,298 particles
231,479 particles
Masked classification (CTR9-WDR61)
83,384 particles
51,218 particles
59,696 particles
Global 3D refinement
3.8 Å
51,218 particles
D
Masked map
Model-To-Map
3.8 Å
Fourier Shell Correlation (FSC)
Resolution (1/Å)
E
1379
919
460
1
# of particles
F
90°
Resolution in Å
3.4 6.5 9.7 12.8 16.0
Side view
Top view

Figure S4. Data acquisition, processing, and data quality metrics for the EC*-TFIIS complex. Related to Figure 5.

(A) Representative denoised micrograph of data collection with scale bar (50 nm).

(B) 2D classes of final refinement show RNA polymerase II, transcription elongation factors, and TFIIS with scale bar of 10 nm.

(C) Sorting and classification tree of EC*-TFIIS complex dataset.

(D) FSC curves of EC*-TFIIS complex. Resolution at FSC threshold criterion 0.143 is indicated.

(E) Angular distribution of particles employed to reconstruct EC*-TFIIS complex.

(F) Local resolution of map EC*-TFIIS complex.

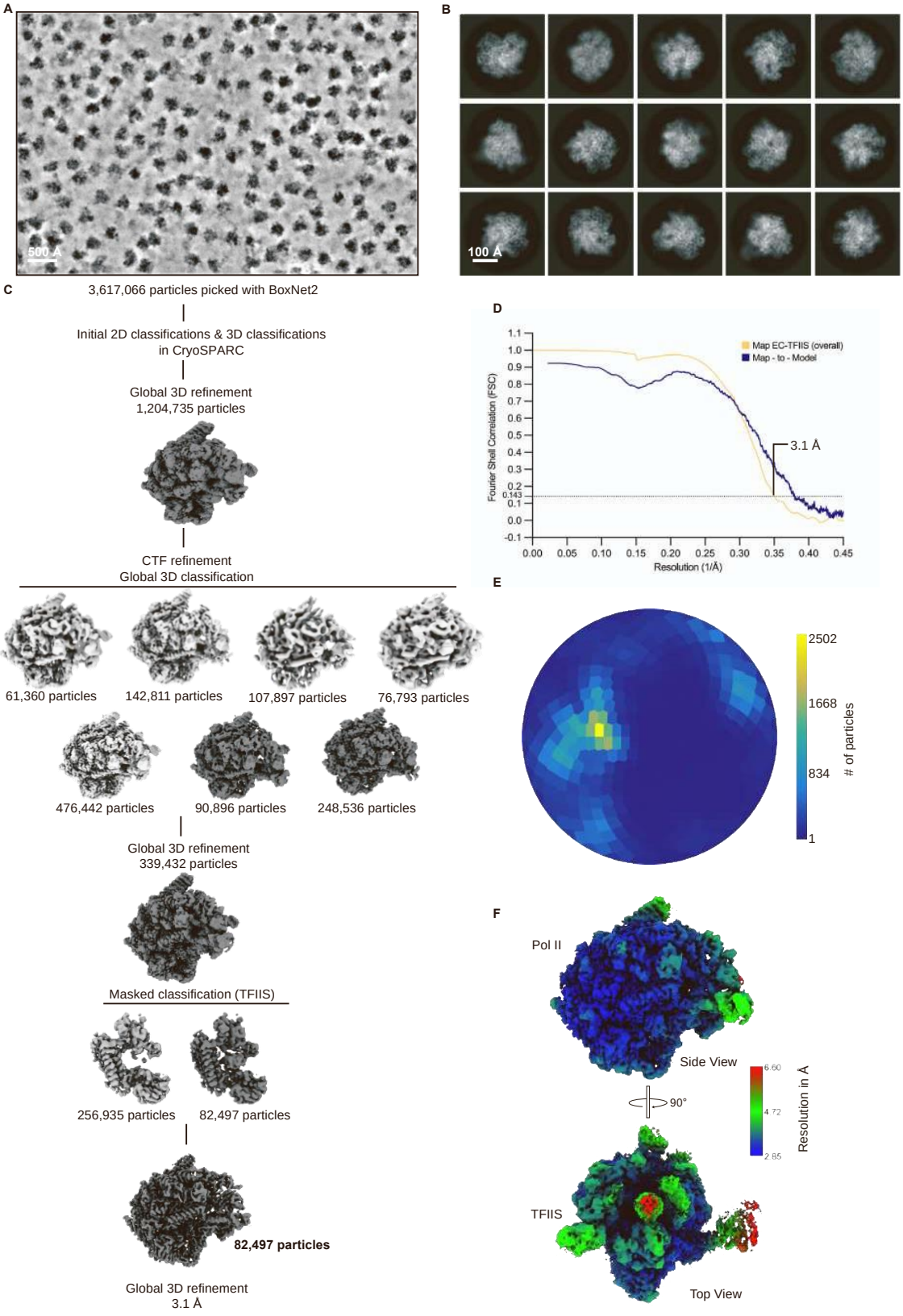

A
500 Å
B
100 Å
C
3,617,066 particles picked with BoxNet2
Initial 2D classifications & 3D classifications
in CryoSPARC
Global 3D refinement
1,204,735 particles
CTF refinement
Global 3D classification
61,360 particles
142,811 particles
107,897 particles
76,793 particles
476,442 particles
90,896 particles
248,536 particles
Global 3D refinement
339,432 particles
Masked classification (TFIIS)
256,935 particles
82,497 particles
82,497 particles
Global 3D refinement
3.1 Å
D
Map EC-TFIIS (overall)
Map - to - Model
Fourier Shell Correlation (FSC)
3.1 Å
Resolution (1/Å)
E
2502
1668
834
1
# of particles
F
Pol II
Side View
90°
TFIIS
Top View
6.60
4.72
2.85
Resolution in Å

Figure S5. Data acquisition, processing, and data quality metrics for the Pol II-TFIIS EC complex. Related to Figure 6.

(A) Representative denoised micrograph of data collection with scale bar (50 nm).

(B) 2D classes of final refinement show RNA polymerase II and TFIIS with scale bar of 10 nm.

(C) Sorting and classification tree of Pol II-TFIIS EC complex dataset.

(D) FSC curve of Pol II-TFIIS elongation complex and map-to-model. Resolution at FSC threshold criterion 0.143 is indicated.

(E) Angular distribution of particles employed to reconstruct Pol II-TFIIS elongation complex.

(F) Local resolution of Pol II-TFIIS elongation complex.

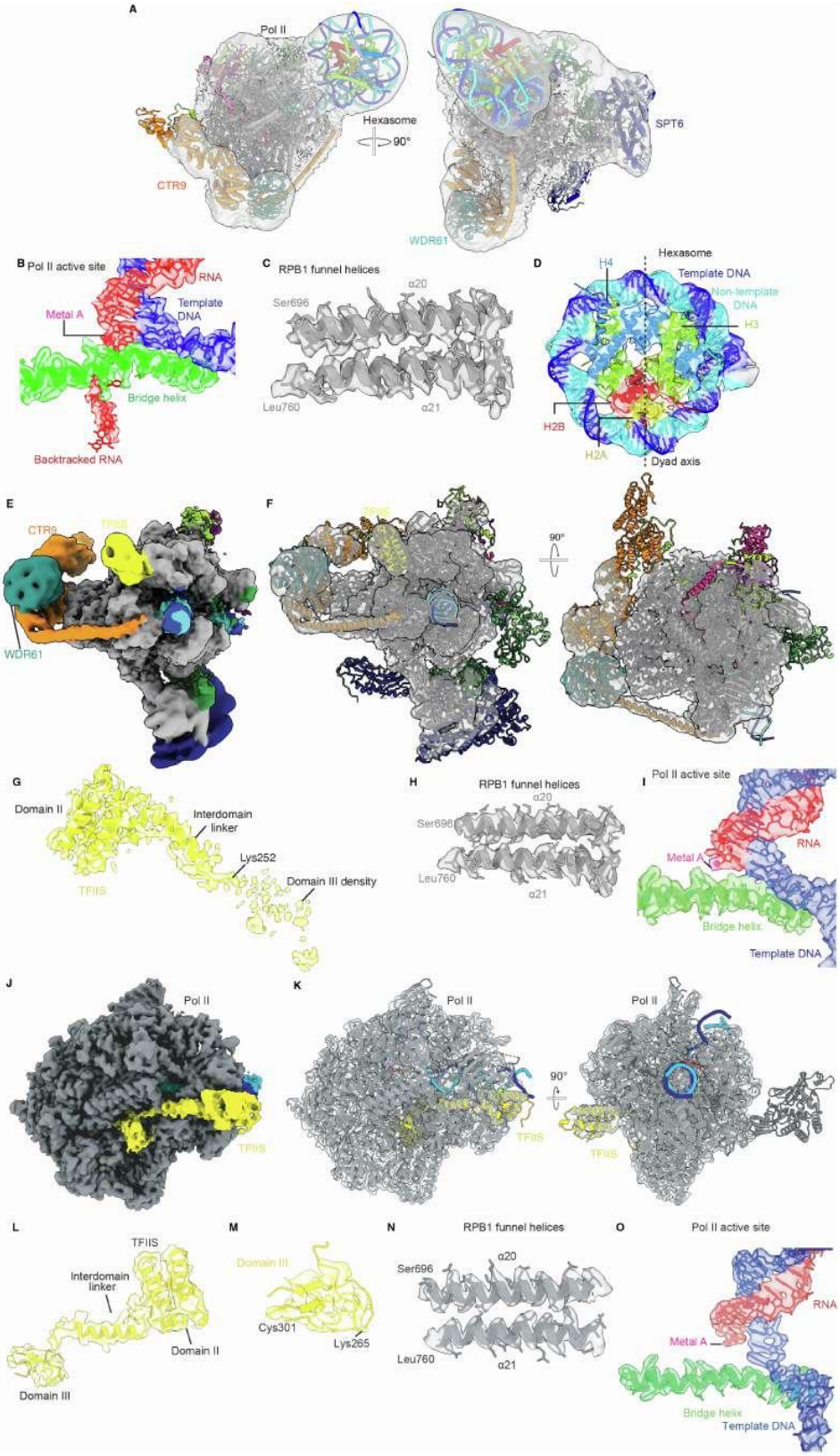

A
Pol II
Hexasome
90°
CTR9
SPT6
WDR61
B Pol II active site
RNA
Template DNA
Metal A
Bridge helix
Backtracked RNA
C RPB1 funnel helices
α20
Ser696
Leu760
α21
D
Hexasome
H4
Template DNA
Non-template DNA
H3
H2B
H2A
Dyad axis
E
CTR9
TFIIS
WDR61
F
90°
G
Domain II
Interdomain linker
Lys252
Domain III density
TFIIS
H
RPB1 funnel helices
α20
Ser696
Leu760
α21
I
Pol II active site
RNA
Metal A
Bridge helix
Template DNA
J
Pol II
TFIIS
K
Pol II
TFIIS
90°
Pol II
TFIIS
L
TFIIS
Interdomain linker
Domain II
Domain III
M
Domain III
Cys301
Lys265
N
RPB1 funnel helices
α20
Ser696
Leu760
α21
O
Pol II active site
RNA
Metal A
Bridge helix
Template DNA

Figure S6. Cryo-EM densities of complexes. Related to Figure 2, 5, and 6.

(A) EC*-hexasome structure with corresponding cryo-EM map. Cryo-EM map is shown in grey.

(B) Pol II active site of EC*-hexasome structure structure with corresponding density. Metal A is shown as a pink sphere.

(C) RPB1 funnel helices with corresponding density from EC*-hexasome complex.

(D) Hexasome with corresponding density (map from masked refinement) from EC*-hexasome complex.

(E) Cryo-EM map of EC*-TFIIS complex.

(F) EC*-TFIIS structure with corresponding cryo-EM map. Cryo-EM map is shown in grey.

(G) TFIIS with corresponding density from EC*-TFIIS complex. Domain III is indicated.

(H) RPB1 funnel helices with corresponding density from EC*-TFIIS complex.

(I) Pol II active site of EC*-TFIIS structure with corresponding density. Metal A is shown as a pink sphere.

(J) Cryo-EM map of Pol II-TFIIS EC complex.

(K) Pol II-TFIIS elongation complex structure with corresponding cryo-EM map. Cryo-EM map is shown in grey.

(L) TFIIS with corresponding density from Pol II-TFIIS elongation complex.

(M) TFIIS domain density from Pol II-TFIIS elongation complex.

(N) RPB1 funnel helices with corresponding density from Pol II-TFIIS elongation complex.

(O) Active site of Pol II-TFIIS EC structure with corresponding density. Metal A is shown as a pink sphere.

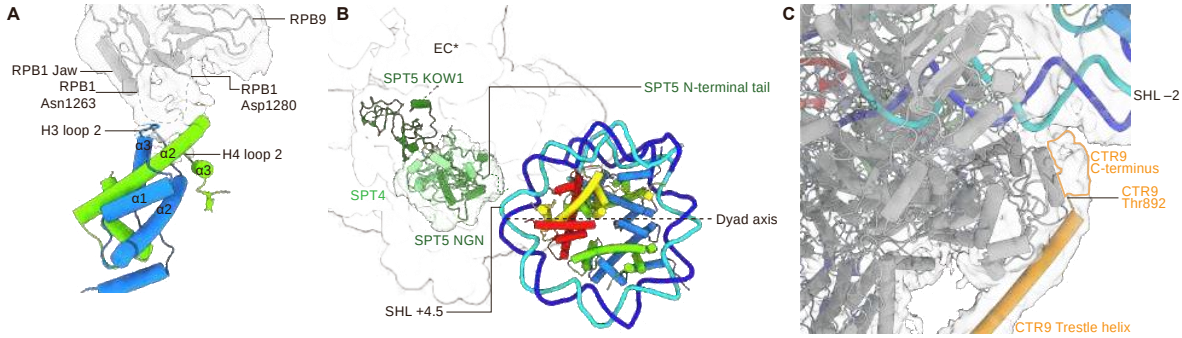


Figure S7. Interactions of Pol II and transcription elongation factors with the hexasome. Related to Figure 2.

(A) Interaction of the RPB1 jaw domain with H3–H4. Density in grey.

(B) Interaction of DSIF with nucleosomal DNA at SHL +4.5. Density in green. EC* indicated as outline. SPT5 N-terminal tail indicated as dashed line.

(C) Interaction of the CTR9 C-terminus with nucleosomal DNA at SHL –2. Density shown in grey and putative CTR9 density highlighted in orange.

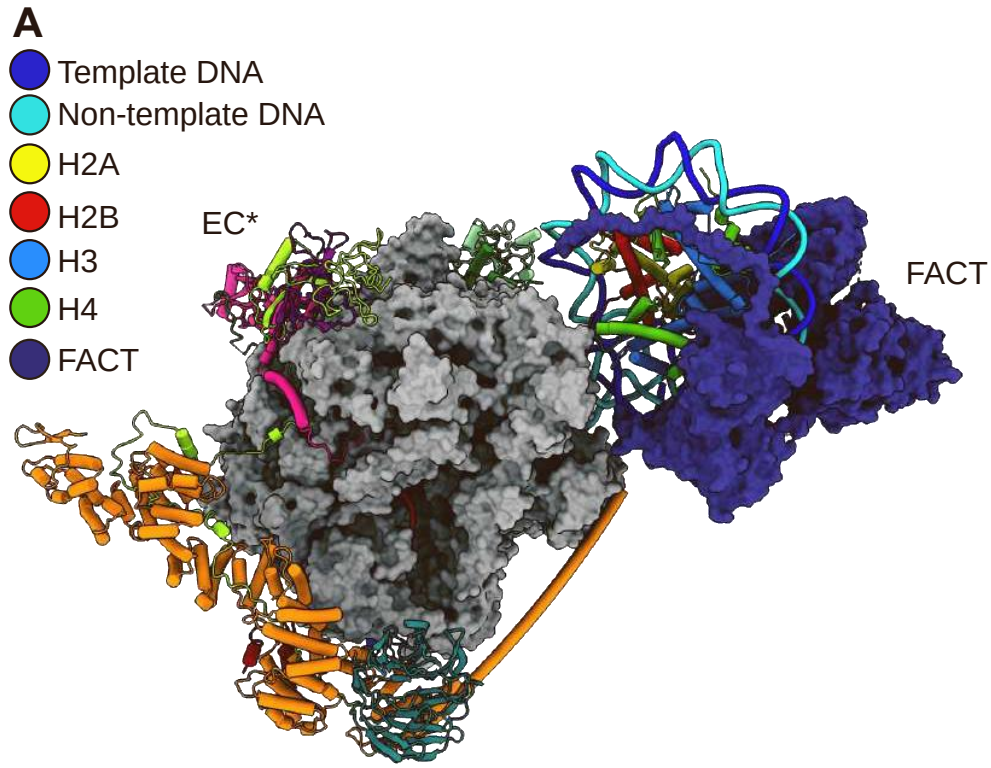


Figure S8. Binding of FACT is compatible with the observed nucleosome position in front of the transcribing EC* complex. Related to Figure 2.

The nucleosome–FACT structure (PDB ID 6UPK) was superimposed onto the EC*-hexasome structure**.** No clashes are observed between FACT and the EC*. It is possible that FACT can redeposit a H2A–H2B dimer in the observed conformation. FACT with subunits SPT16 and SSRP1 is shown in purple.

Table S1. Cryo-EM data collection, refinement, and validation statistics. Related to Figures 2, 5, and 6

	EC*-hexasome (EMD-15127) (PDB 8A3Y)	EC*-TFIIS (EMDB-15128) (PDB 8A3Z)	EC-TFIIS (EMDB-15129) (PDB 8A40)
Data collection and processing			
Magnification	81,000	81,000	105,000
Voltage (kV)	300	300	300
Electron exposure (e^{-}/Å^2)	43-45	43-45	39.87
Defocus range (μm)	0.25-4	0.25-4	0.25-4
Pixel size (Å)	1.05	1.05	0.834
Symmetry imposed	C1	C1	C1
Initial particle images (no.)	9,819,841	1,713,362	3,617,066
Final particle images (no.)	32,933	51,218	82,497
Map resolution (Å) FSC threshold	3.6 0.143	3.8 0.143	3.1
Map resolution range (Å)	3.1-17.8	3.4-16	2.8-4.7
Refinement			
Initial models used (PDB code)	6TED, 3LZ0	6TED, 5IYB	5IYB, 5OIK, 6GMH
Map sharpening *B* factor (Å^2)	-36	-86	-10
Model composition Non-hydrogen atoms Protein residues Nucleotides Ligands	 65,421 7894 293 10	 61,283 7437 106 10	 34,003 4054 81 7
B factors (Å^2) Protein Nucleotide Ligand	 201 486 170	 49.80 111.84 68.20	 47.41 70.06 67.48
R.m.s. deviations Bond lengths (Å) Bond angles (°)	 0.005 0.867	 0.003 0.552	 0.004 0.698
Validation MolProbity score Clashscore Poor rotamers (%)	 1.91 7.91 0.76	 2.05 10.34 1.67	 1.45 4.43 0.00
Ramachandran plot Favored (%) Allowed (%) Disallowed (%)	 92.16 7.28 0.55	 95.02 4.94 0.04	 96.53 3.44 0.02

3 Structural insights into human co-transcriptional capping

Gaurika Garg[1*], Christian Dienemann[1], Lucas Farnung[2], Andreas Linden[3], Julianne Schwarz[3], Henning Urlaub[3, 4], Patrick Cramer[1†]

[1] Max Planck Institute for Multidisciplinary Sciences, Department of Molecular Biology, Am Fassberg 11, 37077 Göttingen, Germany.

[2] Department of Cell Biology, Blavatnik Institute, Harvard Medical School, Boston 02115, MA, USA

[3] Max Planck Institute for Multidisciplinary Sciences, Bioanalytical Mass Spectrometry Group, Am Fassberg 11, 37077 Göttingen

[4] University Medical Center Göttingen, Georg-August-Universität, Robert-Koch-Straße 40, 37075 Göttingen

[†]Correspondence: patrick.cramer@mpinat.mpg.de

Structural insights into human co-transcriptional capping

Gaurika Garg[1], Christian Dienemann[1], Lucas Farnung[2], Andreas Linden[3], Julianne Schwarz[3], Henning Urlaub[3, 4], Patrick Cramer[1†]

[1] Max Planck Institute for Multidisciplinary Sciences, Department of Molecular Biology, Am Fassberg 11, 37077 Göttingen, Germany.

[2] Department of Cell Biology, Blavatnik Institute, Harvard Medical School, Boston 02115, MA, USA

[3] Max Planck Institute for Multidisciplinary Sciences, Bioanalytical Mass Spectrometry Group, Am Fassberg 11, 37077 Göttingen

[4] University Medical Center Göttingen, Georg-August-University, Robert-Koch-Straße 40, 37075 Göttingen

[†]Correspondence: patrick.cramer@mpinat.mpg.de

SUMMARY

Co-transcriptional capping of the nascent pre-mRNA 5' end prevents degradation of RNA polymerase (Pol) II transcripts and suppresses the innate immune response. Here we provide mechanistic insights into the three major steps of human co-transcriptional pre-mRNA capping based on six different cryo-EM structures. The human mRNA capping enzyme first docks to the Pol II stalk to position its triphosphatase domain near the RNA exit site. The capping enzyme then moves and the pre-mRNA 5'-diphosphate end is transferred to the guanylyl transferase. Addition of a GMP moiety can occur when the RNA reaches ~22 nucleotides in length. For subsequent RNA methylation, the methyltransferase CMTR1 binds between the Pol II stalk and the guanylyl transferase and can receive RNA once it is grown to ~27 nucleotides. The observed rearrangements of capping factors on Pol II appear to be triggered by the completion of catalytic reaction steps and are accommodated on the Pol II surface by domain movements in the modular elongation factor DSIF.

INTRODUCTION

Capping of the nascent pre-mRNA is an essential event that occurs early during RNA polymerase II (Pol II) transcription (McCracken et al., 1997; Shatkin, 1976; Wen et al., 1998). Capping facilitates further pre-mRNA processing, nuclear export, and efficient mRNA translation (Changela et al., 2001; Furuichi et al., 1977; Konarska et al., 1984). In higher eukaryotes and many viruses, the predominant cap structure is a 7-methylguanosine linked via a triphosphate bridge to the first transcribed nucleotide, which is methylated on the ribose O-2 position (Furuichi et al., 1977;

Shatkin, 1976). This structure is often referred to as cap1 and is annotated as m7G(5')ppp(5')XmRNA, where Xm is the methylated first transcribed RNA nucleotide (Cowling, 2010; Furuichi et al., 1977). In humans, cap1 is present on all Pol II transcripts (Darnellt, 1975). The cap1 structure serves as a crucial player of cellular immunity in two ways. First, it is recognized as self-RNA and prevents the triggering of an innate immune response (Devarkar et al., 2016; Stepinski et al., 2010). Second, it helps establish an antiviral state by promoting the expression of genes required during the type I interferon response (Williams et al., 2020). An RNA with a cap0 (m7G(5')ppp(5')XRNA) structure that lacks the methyl group on the first transcribed RNA nucleotide is recognized as non-self and activates a RIG-I or MDA5 mediated innate immune response in the host (Stepinski et al., 2010; Züst et al., 2011).

The formation of a 5' cap entails three major consecutive catalytic steps. In mammals, the first two steps are catalyzed by a bifunctional RNA guanylytransferase and 5'-phosphatase (RNGTT). In the first step, RNGTT catalyzes the removal of the RNA gamma-phosphate via its N-terminal triphosphatase (TPase) domain resulting in 5'ppXRNA and inorganic phosphate (Pi). This is followed by a second step, the addition of a GMP moiety by the RNGTT C-terminal guanylytransferase (GTase) domain that uses GTP as a substrate (Changela et al., 2001; Rodriguez et al., 1999). The resulting partial cap, 5'GpppX, that we annotate as cap(-1) serves as a substrate for several methylation events that usually use S-adenosylhomocysteine (SAH) as the methyl donor (Tomasz and Shatkin, 1976). Conventionally, in the third step, RNA methyltransferase (RNMT) together with the RNMT activating miniprotein (RAM) adds a methyl group to N-7 of the cap guanosine to generate the cap(0) structure (5'm7GpppX) (Varshney et al., 2018). Further, the cap-specific methyltransferase 1 (CMTR1) methylates the 2'O of the first transcribed nucleotide, resulting in the cap(1) structure (5'GpppX$m_{O\text{-}2}$RNA). However, as CMTR1 has a similar affinity for cap(-1) and for cap(0) structures, the order of methylation is unclear and it is possible that CMTR1 acts prior to RNMT (Smietanski et al., 2014; Stepinski et al., 2010).

There is strong evidence for the co-transcriptional nature of pre-mRNA capping, in particular from *in vivo* data (Bentley, 2014; Gonzalez et al., 2018; Neugebauer, 2002). However, there is also biochemical and structural evidence for co-transcriptional capping. First, the phosphorylated C-terminal repeat domain (CTD) of the largest Pol II subunit RPB1 can serve as a recruitment platform for enzymes involved in capping (Rasmussen and Lis, 1993; Rodriguez et al., 1999). Second, mammalian capping has been shown to occur during early elongation, with the general elongation factor DRB sensitivity inducing factor (DSIF) stimulating the process (Mandal et al., 2004). Interestingly, DSIF was shown to stimulate guanylyation only for the full-length RNGTT, while it had no effect on the GTase domain alone. This DSIF induced enhancement was abolished in the presence of increasing amounts of the TPase domain (Mandal et al., 2004; Wen and Shatkin, 1999). Additionally, DSIF was shown to bind to elongation complexes containing

atleast 22 nt RNA, and preferentially to complexes with 25 nt RNA (Cheng and Price, 2008; Palangat et al., 2005). These observations therefore leave the precise role and extent of DSIF involvement in capping unclear. Third, recruitment of RNGTT to Pol II relieves the inhibition by negative elongation factor (NELF), according to the capping checkpoint model (Kachaev et al., 2020; Mandal et al., 2004). Finally, although yeast uses capping enzymes that are entirely distinct from the mammalian counterparts, early low-resolution structural studies have shown that a yeast capping enzyme can bind to the Pol II surface near the end of the RNA exit tunnel of Pol II, suggesting how the nascent RNA 5' end can be capped as soon as it reaches the Pol II surface (Martinez-Rucobo et al., 2015). Such direct interaction between a polymerase and a capping enzyme was also observed for the poxviral machinery (Grimm et al., 2019).

There are currently no structural insights into mammalian co-transcriptional capping complexes, and structural information is only available for isolated mammalian capping factors (Decroly et al., 2012). Therefore, the mechanisms that underlie the functional coupling of transcription to the various steps of capping are unknown. Here we report structures of six transcribing Pol II complexes with various factors including RNGTT, DSIF and CMTR1. From these results emerges a model for the highly dynamic multistep process of co-transcriptional pre-mRNA capping.

RESULTS

Reconstitution of human capping machinery on the Pol II surface

In order to prepare functional complexes of transcribing Pol II with human capping factors, we first purified endogenous porcine (*S. scrofa*) Pol II, which is almost identical to the human Pol II, and recombinant human factors RNGTT, DSIF, NELF, RNMT/RAM, and CMTR1 (Methods). We also prepared RNAs of various lengths carrying different cap structures, i.e. uncapped, cap(-1) and cap(0) (Methods). We were able to reconstitute RNA guanylyation *in vitro* (Fig S1) and observed that efficient capping of RNAs that are associated with their 3' end with the Pol II active site in transcribing complexes requires a minimum length of 22-24 nucleotides (nt) (Fig S1B) (Shatkin and Manley, 2000). We then checked the binding of RNGTT and CMTR1 variants to the transcribing Pol II complex in the presence and absence of DSIF, and assembled the different complexes *in vitro* for their imaging by single-particle cryo-EM (Methods, Fig S1E,F). The complexes were further subjected to crosslinking mass-spectrometry (Methods, Fig S7).

Structure of transcribing Pol II-TPase complex

To visualize the first step in co-transcriptional pre-mRNA capping and to understand the structural dynamics in the absence of DSIF, we prepared a minimal transcribing Pol II-RNGTT complex. We used a 22 nt uncapped RNA, which corresponds to the minimal RNA length required for activity

(Methods). In order to trap the 5' triphosphate end of the RNA in the TPase domain active site, we prevented its dissociation after catalysis by using a RNGTT variant with a mutation in the TPase active site (Cys126 to Ser126, and called it C126S) (Fig 1A) (Changela et al., 2001). We imaged the complex by cryo-EM and resolved its structure to a nominal resolution of 3.6 Å with the Pol II core resolved at 3.1 Å and the peripheral regions at 3.5 Å to 7.5 Å (Fig S2).

In the resulting structure 1, we observed both the domains of the RNGTT located docked onto the Pol II surface (Fig 1 B and D). RNGTT domain organisation reveals how the TPase and the GTase remain mobile relative to each other. Apart from TPase domain N-terminal helices in proximity to the GTase OB-domain, no extensive interface is observed (Fig 1 C).

The TPase domain is positioned at the end of the Pol II RNA exit tunnel and receives the RNA. The positively charged active site pocket, formed largely by the characteristic phosphate-binding loop (P-loop) of the TPase domain faces the RNA 5' end (Fig 1F). This allows the emerging negatively charged phosphate backbone of the nascent RNA to interact with the TPase, guiding it to the active site Cys126 residue for catalysis. RNA can be modelled up to 19 nt, and additional density, likely from the RNA backbone is observed in the TPase active site (dashed line in Fig 1F).

The Pol II-TPase interface reveals two primary contacts (Fig 1E,F). One of these contacts involves a surface loop (residues 114-118) (Changela et al., 2001) that gets ordered in the complex structure by forming a hydrophobic interface with the RPB7 tip domain (Armache et al., 2005). The other contacts are formed between the TPase domain N-terminal region and the RPB2 flap loop (Bernecky et al., 2016) (Fig 1F).

The GTase domain binds the stalk and adopts an open conformation, as the GTase active site cleft is not bound by a substrate (Hâkansson and Wigley, 1998). The details of the GTase interaction and activity are discussed below.

Structure of the transcribing Pol II-GTase complex

In order to gain insights into the second step of co-transcriptional capping, we assembled transcribing Pol II with a 22 nt uncapped RNA and wild-type (WT) RNGTT containing an intact TPase active site in the absence of GTP (Methods). These conditions enabled TPase action but prevented GTase action, which requires the presence of GTP (Methods). We imaged the complex by cryo-EM and resolved the structure (Fig. 2) at a nominal resolution of 3.2 Å with the Pol II core resolved at 2.9 Å and the periphery of the complex between 3.0 Å to 6.8 Å (Fig S3).

The resulting structure 2 reveals the GTase domain of RNGTT, whereas the TPase that was observed in structure 1 is mobile (Fig. 2, A and C). The exiting RNA now extends into the GTase active site (Fig 2D), instead of towards the TPase active site as observed in structure 1. The RNA is bound in the positively charged cleft of the GTase and extends to the active site residue lysine 294.

The Pol II-GTase interface remains unchanged in both structures 1 and 2. We observe a loop from the oligonucleotide binding fold (OB-fold) of the GTase domain (residues 476-484; OB-loop) inserting into the crevice formed between the RPB1 dock and the RPB7 tip domain (Fig 2B) (Armache et al., 2005).. The anchoring OB-loop is lacking in the OB-fold of the structurally similar Chlorella virus PBCV-1 viral capping enzyme (Chu et al., 2011), suggesting that the OB-domain might have evolved in higher eukaryotes to confer GTase domain binding to the polymerase.

The TPase domain is not well resolved (Fig 2A), indicating that it becomes mobile upon completion of the first reaction. The transient stabilization of the TPase domain on Pol II is supported by the observation that this domain alone is unable to bind to a transcribing complex, in contrast to the GTase domain which readily binds a transcribing Pol II complex (Fig S1D). These observations suggest that the GTase domain serves as the primary RNGTT anchor on Pol II surface, likely due to the stable interaction of the GTase OB-loop with the Pol II stalk.

RNA transfer from TPase to GTase

Comparison of structures 1 and 2 (Fig. 3A) elucidates the transfer of the RNA substrate 5’ end from the TPase to the GTase domain of RNGTT. In structure 1, the TPase domain is positioned at the RNA exit tunnel, ready to accept the (5’)pppXRNA end. Superposition with structure 2 reveals that the GTase domain has moved closer to the RNA exit site in structure 2. Rigid-body docking of the TPase into a lowpass-filtered map of structure 2 shows that the TPase is displaced from the Pol II stalk (dashed outline in Fig 2A). Furthermore, the conformation of the GTase has changed dramatically from an open state in structure 1 to a partiually closed state in structure 2. Closure of the GTase domain is achieved by a swivel movement of the nucleotidyltransferase domain (NT-domain) by ~25 Å. This is in contrast to the idea proposed based on the isolated crystal structure (Fig 3B) (Håkansson et al., 1997), which involved the movement of the OB-fold instead.

In summary, after completion of TPase activity, the RNA 5’ end is released from the TPase domain active site, binds the GTase domain active site, leading to a closure of the GTase domain and its movement nearer to the RNA exit channel, which in turn results in displacement of the transiently stabilized TPase domain.

Structure of the transcribing Pol II-DSIF-RNGTT complex

While RNGTT alone can efficiently cap RNA, DSIF has been shown to have a stimulatory effect on cap guanylyation under limited capping enzyme conditions (Ref). Additionally, promoter proximal pausing has been correlated to capping in serving as capping check point (Ref). Interestingly, superposition of our minimal structures 1 and 2 onto the transcribing Pol II-DSIF complex (Bernecky et al., 2016) showed that the TPase domain clashes with the KOWx-KOW4 domain of the DSIF subunit SPT5 and the GTase OB-fold shows a minor clash with KOW4 (Fig

S4). Therefore, to investigate how the general elongation factor DSIF affects the binding and activity of RNGTT, we assembled a complex of transcribing Pol II with a 24 nt uncapped RNA (the most efficiently capped substrate, Fig S1), DSIF, RNGTT and NELF, assuming it may stabilize the complex (Nilson et al., 2015). We imaged the complex by cryo-EM and determined two structures from the data (structures 3 and 4) (Fig. 4). Our structures reveal both DSIF and RNGTT bound to Pol II. However, NELF was only observed in a particle class that lacked RNGTT.

Structure 3 (Fig 4A) was resolved at a nominal resolution of 3.8 Å with the Pol II core at 2.6 Å and periphery between 4.0 Å to 8.6 Å. It captures the transcribing complex in the presence of both DSIF and RNGTT with RNA directed towards the TPase domain. RNGTT is anchored to the Pol II stalk via the GTase OB-loop as described above, and the NT-domain adopts an open conformation as in structure 1. The resolution for SPT5 KOWx-KOW4 domains and the TPase domains is limited, likely because KOWx-KOW4 and the TPase domain compete for the same binding site, which leads to heterogeneity due to a mixture of states. The TPase was rigid-body docked based on the active site orientation and its conformation in structure 1. SPT5 domains were modelled into the density, apart from KOWx-KOW4, which was rigid body docked. This revealed large deviations of SPT5 from the canonical conformation (Bernecky et al., 2016). The TPase is sandwiched between KOW2-KOW3 and the GTase, and likely displaces KOWx-KOW4 from the end of the RNA exit channel, moving it nearer to the RPB7 OB-fold. Additionally, the KOW1-L1 domain is moved towards KOW2-KOW3. Therefore, accommodating the TPase domain requires large domain movements in SPT5, especially of its KOWx-KOW4 domain.

Structure 4 (Fig. 4B) was resolved at a nominal resolution of 3.6 Å with the Pol II core at 2.6 Å and the periphery resolved between 4.0 and 8.5 Å. It captures the RNA extending towards the GTase domain. The GTase OB-loop consistently anchors RNGTT to Pol II, and the NT-domain is in a partially closed state, similar to structure 2. Interestingly, KOWx-KOW4 could be docked as a rigid body and shows that KOW4 is now positioned near the GTase OB-domain, where the TPase was previously bound. No density for the TPase domain was observed. We therefore conclude that after TPase action, the TPase domain is displaced, and the GTase NT-domain can move to the RNA exit channel.

Comparison of our structures 1-4 suggests that capping enzymes can be accommodated in different states due to the modular nature of SPT5, which contains several mobile domains that are bound dynamically around the Pol II stalk and RNA exit tunnel. To accommodate RNGTT, SPT5 adopts a different conformation compared to its canonical state (Bernecky et al., 2016), especially an altered position of its KOWx-KOW4 domain. These rearrangements of DSIF domains around RNGTT could sterically limit RNGTT mobility, thereby limiting reaction dynamics. This may in turn explain the known increase in guanylylation efficiency under limiting capping enzyme concentrations (Mandal et al., 2004).

Structures of the transcribing Pol II with CMTR1

To investigate the last step of capping, RNA methylation, we first tried to bind the methyltransferase complex RNMT/RAM to a transcribing Pol II. Despite many trials, we were not able to obtain stoichiometric binding or to observe density for RNMT on the Pol II surface (data not shown). We therefore turned to the methyltransferase CMTR1, which is responsible for co-transcriptional 2'OH methylation of the first ribonucleotide (Stepinski et al., 2010).

We assembled a transcribing Pol II-DSIF-RNGTT-CMTR1 complex with uncapped RNA of various lengths. Whereas we did not observe density for CMTR1 when using a 26 nt uncapped RNA, use of a 29 nt uncapped RNA revealed CMTR1 density on Pol II. We imaged the complex by cryo-EM and resolved two structures (structures 5 and 6). Structure 5 captured a transcribing Pol II-DSIF-RNGTT-CMTR1 complex at a nominal resolution of 3.9 Å, with the Pol II core resolved at 3.0 Å and the periphery between 4.0 and 9.5 Å (Fig. S6). Structure 6 showed a transcribing Pol II-DSIF-CMTR1 complex at a nominal resolution of 4.0 Å with the Pol II core at 3.0 Å and the periphery resolved between 4.0 to 9.8 Å (Fig S6). We used the crystal structure of the CMTR1 MTase domain (PDB code 4N48) and the C-terminal region of a model of CMTR1 predicted by AlphaFold2 (Ronneberger et al., 2021; Stepinski et al., 2010) to build an atomic model and to trace the RNA from the Pol II active site to the CMTR1 active site.

Structure 5 showed that the SPT5 KOW5 domain, the RNGTT GTase domain and CMTR1 can co-occupy the Pol II surface (Fig. 5). The observation of KOW5 indicates that DSIF is bound but that the other DSIF domains are mobile. The GTase domain is bound with its OB-fold as in the other structures, and is in an open conformation with its active site empty. In contrast to structures 1 and 2, however, the NT-domain of the GTase has moved away from the RNA exit channel by up to 27 Å and 40 Å, respectively. This movement apparently accommodates CMTR1, that interacts directly with Pol II, which is described below.

Structure 6 shows that DSIF adopts its canonical conformation (Bernecky et al., 2016) and that CMTR1 is bound to the Pol II stalk and to exiting RNA (Fig. 6). CMTR1 uses its C-terminal GTase-like domain to bind the Pol II RPB7 OB-fold, constituting the only Pol II-CMTR1 contact in our structures, suggesting its role in recruitement of CMTR1 to the Pol II surface (Fig 6C). Additionally, CMTR1 C-terminus deletion *in vitro* has been shown to decrease its methylation activity (Smietanski et al., 2014). Therefore, the GTase-like domain may also help stabilize a conformation of CMTR1 that supports efficient RNA methylation. RNA can be unambiguously traced from Pol II to the CMTR1 methyltransferase domain (Fig 6B), and binds in the methyltransferase (MTase) active site essentially as observed in a CMTR1-RNA crystal structure (Stepinski et al., 2010). Considering that CMTR1 does not recognize 5'pppXRNA (Fig S 1D), the RNA apparently went through the first two capping steps with RNGTT, was released from the

GTase domain and is in the guanylylated cap(-1) state, enabling its binding to the CMTR1 active site.

Superposition of structures 5 and 6 (Fig 6D) reveal a minor clash between GTase OB-fold and SPT5 domain KOW4. It also shows a major clash between the GTase NT-domain and the CMTR1 MTase domain. The CMTR1 MTase domain is moved towards the RNA exit channel by 22 Å in structure 6, whereas the C-terminal region interacting with the Pol II stalk remains relatively unaltered. Therefore, the CMTR1 swinging-in movement and the rearrangement of the modular DSIF subunit SPT5 may be coupled to a displacement of RNGTT after it has completed its function.

DISCUSSION

In this study we determined six cryo-EM structures of Pol II elongation complex intermediates involved in mammalian co-transcriptional capping. Together with known functional results, these structures lead to a model for the process of mammalian co-transcriptional capping (Fig. 7). In this model, RNGTT is first recruited to a transcribing Pol II-DSIF complex during early elongation through interactions with the Pol II stalk, which are enabled by conformational changes in DSIF, especially movements of its KOWx-KOW4 domain. The TPase of RNGTT is transiently stabilized at the end of the RNA exit tunnel of Pol II and removes the 5' gamma-phosphate of the emerging nascent RNA. In a second step, the RNA 5' diphosphate end is released from the TPase, which is displaced, allowing for movement of the GTase NT-domain nearer the RNA exit site. When the nascent RNA grows to 22-24 nt, it can reach the GTase active site, which uses GTP to add GMP to the 5' diphosphate end of the RNA, generating the cap(-1) structure. In a third step, CMTR1 can dock on to the Pol II stalk, bind and methylate RNA once it reached a length of ~29 nt. CMTR1 binding triggers displacement of RNGTT and resets DSIF to its original conformation.

Additionally, our minimal structures 1 and 2 shed light on the individual capping steps, illustrating the mechanism of RNA handover between active sites. This is crucial for understanding how RNGTT functions in isolation, as capping can occur efficiently without DSIF (Moteki and Price, 2002). It also leads us to suggests that the observed transient DSIF rearrangements are apparently a result of RNGTT binding, as is observed with structures 3, 4 and 5. Consistent with this observation, in structure 6, which lacks RNGTT, DSIF displays its canonical arrangement.

Comparison of our mammalian structures with structural information on yeast capping complexes reveals that the human and yeast co-transcriptional capping systems are entirely distinct. Whereas in both systems the capping enzyme is recruited near the RNA exit site of Pol II, the capping enzyme-Pol II interactions are different. In our structures the capping enzyme is anchored via the human GTase that binds to the Pol II stalk, whereas the yeast capping enzyme uses its TPase Cet1 to bind near the Pol II wall (Martinez-Rucobo et al., 2015). Also, human mRNAs possess the

cap(1) modification and a corresponding enzyme, CMTR1, that introduces this modification and is resolved here at a specific position on Pol II, whereas yeast mRNAs lack a cap(1) modification and the corresponding methyltransferase (Sripati et al., 1976). This difference relates to the innate immune system in human, which is absent in yeast, emphasizing the importance of cap(1) in self-RNA identification and in acting as a gatekeeper for the innate immune response, where sensors and effectors recognize cap(-1) and cap(0) with similar affinity (Devarkar et al., 2016).

Finally, it remains unclear how the methyltransferase complex RNMT/RAM fits into our model of co-transcirptional capping. We could not observe binding of RNMT/RAM to Pol II complexes and thus additional factors may be required to recruit this methyltransferase. It may also be that cap(0) methylation occurs further downstream, after cap(1) methylation is completed. This idea is supported by the fact that CMTR1 can efficiently act on both cap(-1)/cap(0) transcripts (Stepinski et al., 2010), that in some viral machineries cap(1) modification is introduced before cap(0) (Decroly et al., 2012), and that CMTR1 occupies the beginning of genes *in vivo*, whereas RNMT does not (Liang et al., 2022). Nevertheless, our structures and our molecular model of co-transcriptional capping provide a basis for understanding the essential and fundamental process of mammalian co-transcriptional pre-mRNA capping in mechanistic detail.

AUTHOR CONTRIBUTIONS

G.G. carried out all experiments and data analysis, unless stated otherwise. C.D. performed data acquisition for structure 2 and advised on data acquisition and processing. A.L. and J.S. performed crosslinking-MS analysis, supervised by H.U. L.F. cloned the RNGTT plasmid (insect cells), advised on the experimental approach and on data processing for structure 2. P.C. supervised research. G.G., and P.C. prepared the manuscript, with input from all other authors.

ACKNOWLEDGEMENTS

We thank past and present members of the Cramer laboratory. We thank C. Oberthür for help with protein purification. We thank C. Dienemann and U. Steuerwald for maintenance of cryo-EM resources. We thank A-H.Rosch for providing the 29 nt RNA. We thank I.F. for sharing NELF. P.C. was supported by the ERC Advanced Investigator Grant CHROMATRANS (grant agreement no. 693023).

DECLARATION OF INTERESTS

The authors declare no competing interests.

FIGURES

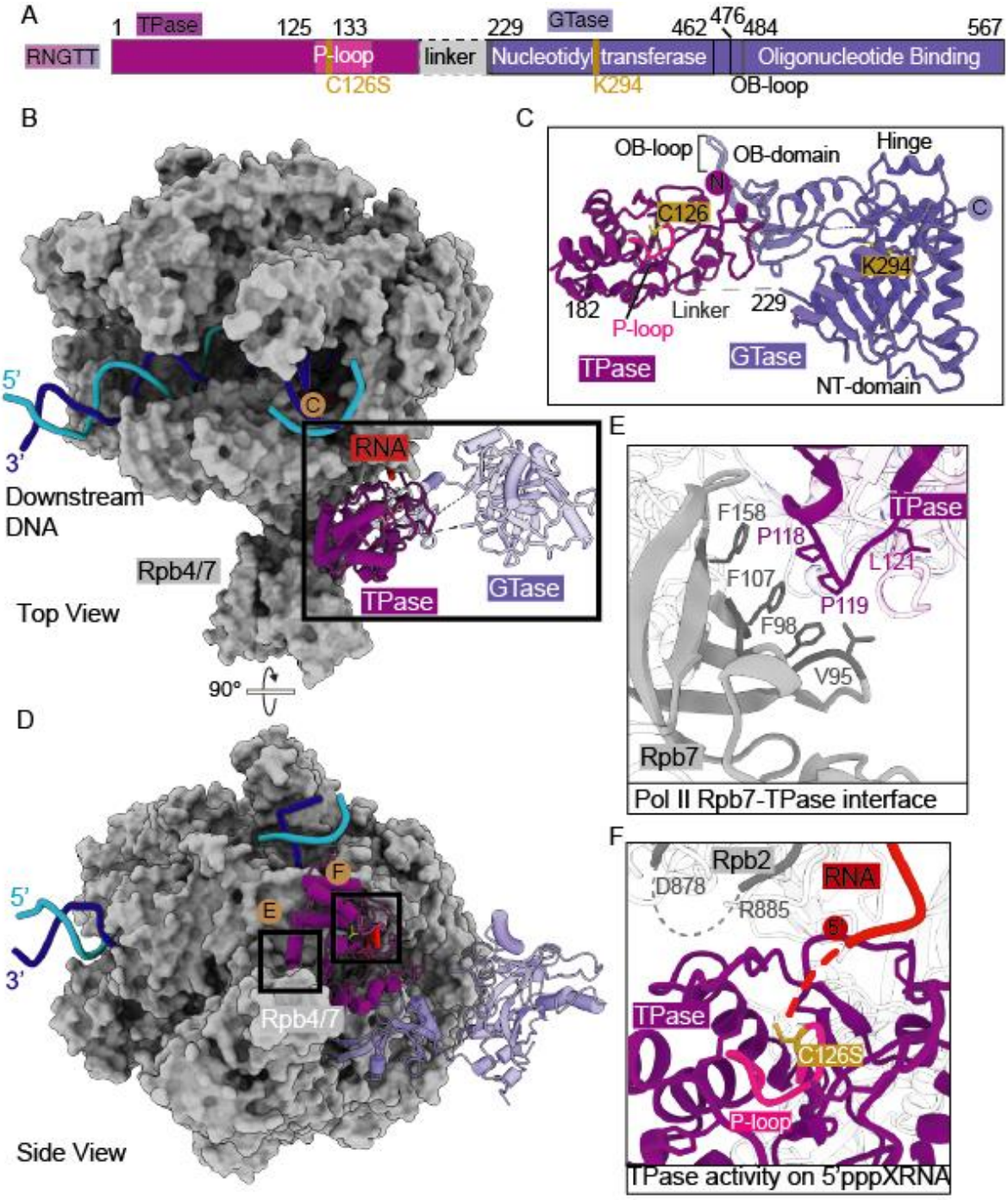

A
1
TPase
125
133
229
GTase
462
476
484
567
RNGTT
P-loop
linker
Nucleotidyl transferase
Oligonucleotide Binding
C126S
K294
OB-loop
B
Downstream DNA
Rpb4/7
Top View
RNA
TPase
GTase
90°
C
OB-loop
OB-domain
Hinge
C126
K294
182
P-loop
Linker
229
TPase
GTase
NT-domain
D
Side View
Rpb4/7
E
F158
P118
TPase
L121
F107
P119
F98
V95
Rpb7
Pol II Rpb7-TPase interface
F
Rpb2
D878
R885
RNA
TPase
C126S
P-loop
TPase activity on 5'pppXRNA

Figure 1: Structure of the transcribing Pol II-TPase complex (structure 1). Related to Figures S1 and S2.

(A) RNGTT domain architecture, elaborating the TPase and GTase domains in purple and slate blue, respectively. The active sites of the bifunctional RNGTT are in yellow, and the TPase domain P-loop is in magenta.

(B) Structure 1 top view. Overview of the transcribing Pol II-RNGTT complex in the TPase phase of the reaction. Shown are Pol II (grey), RNA (red) and RNGTT (purple and slate blue).

(C) Structural overview of the complete RNGTT. The relative organisation reveals how the domains remain mobile relative to each other. Apart from TPase domain N-terminal helices in proximity to the GTase OB-domain, no extensive interface is observed.

(D) Structure 1 side view. Overview of transcribing Pol II-RNGTT in the TPase step rotated 90° around the y-axis.

(E) Pol II-TPase interface. Hydrophobic residues from Rbp7 (grey) and the surface loop residues (magenta) of the TPase are depicted, that likely form a transiently stabilized interface.

(F) RNA-TPase and second Pol II-TPase interface. Important residues (magenta) guiding and interacting with the RNA, probable RNA trajectory (dashed red), and active site cysteine are shown (PDB code 3S24). RPB2 flap loop residues also likely stabilize the TPase domain by interacting with its C-terminus.

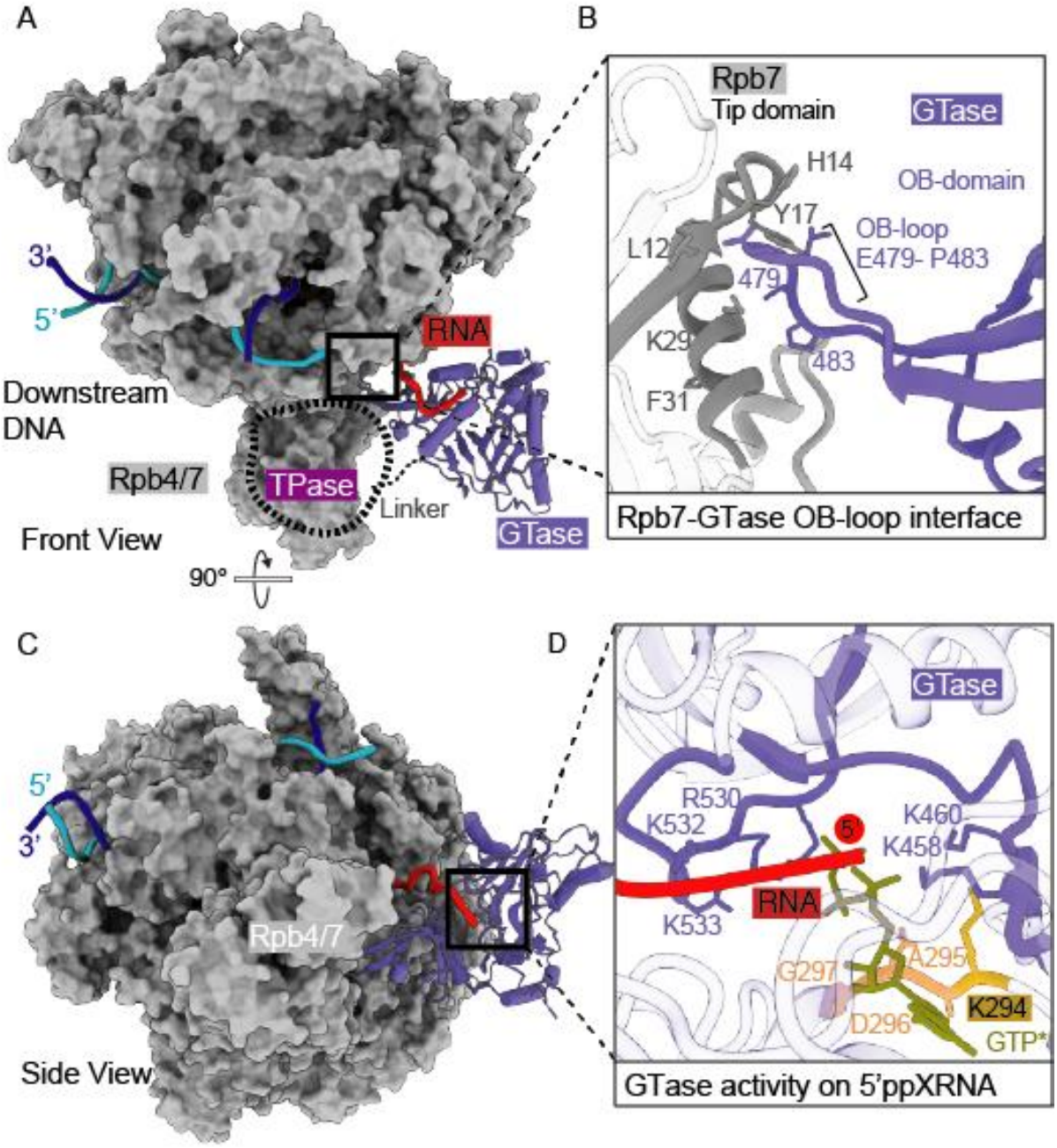

A
B
3'
5'
RNA
Downstream
DNA
Rpb4/7
TPase
Linker
GTase
Front View
90°
Rpb7
Tip domain
GTase
H14
OB-domain
Y17
OB-loop
E479- P483
L12
479
K29
483
F31
Rpb7-GTase OB-loop interface
C
D
5'
3'
Rpb4/7
Side View
GTase
R530
K532
K460
K458
5'
RNA
K533
G297
A295
K294
D296
GTP*
GTase activity on 5'ppXRNA

Figure 2: Structure of the transcribing Pol II-GTase complex (structure 2). Related to Figures S1 and S3.

(A) Structure 2 top view. Overview of the transcribing Pol II-RNGTT complex in the at the GTase step of the reaction. Shown are Pol II (grey), RNA (red), and GTase (slate blue) connected via a flexible linker (dashed line) to the TPase domain (dashed outline).

(B) Pol II-GTase interface. RPB7 tip domain (grey) and OB-loop residues (slate blue) are shown. Important residues that likely contribute to a tight anchor between Pol II and the GTase domain via ionic, hydrogen or hydrophic interactions have been shown as side chains.

(C) Structure 2 side view. Overview of transcribing Pol II-RNGTT at the GTase step rotated 90° around the y-axis.

(D) RNA-GTase interface. Important residues guiding and interacting with the RNA (slate blue), RNA trajectory (red), and active site lysine 294 (yellow) are shown. GTP (olive) is superimposed from the viral crystal structure (PDB code 1CKN), and was not present in our sample preparation.

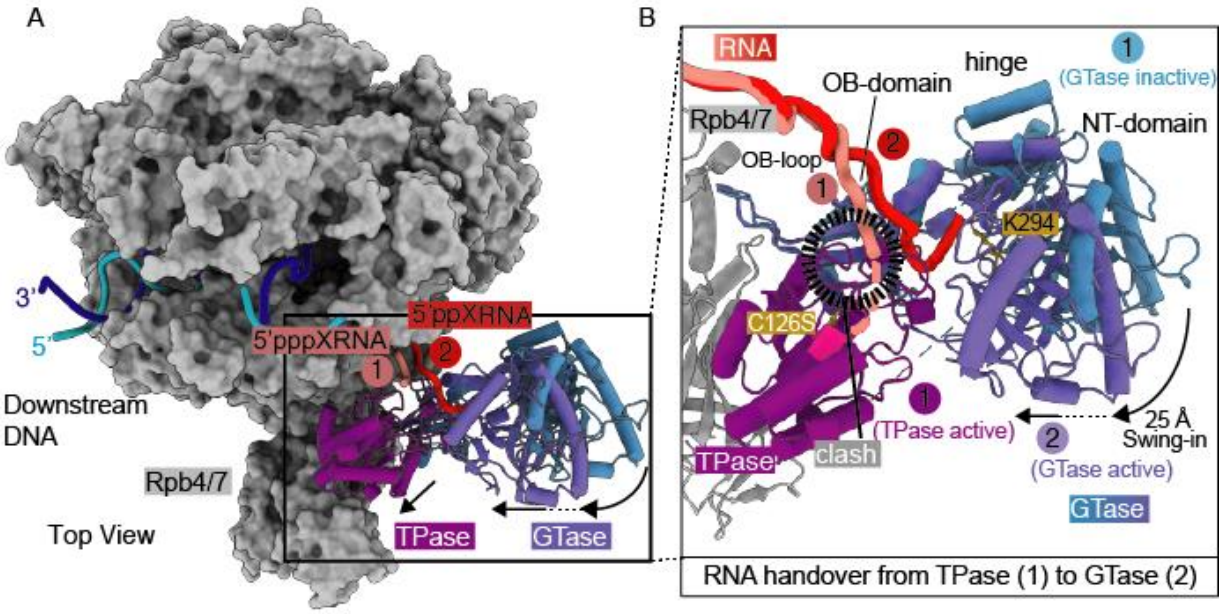


Figure 3: RNA handover between the RNGTT active sites. Related to Figure S4.

(A) Superposition of structure 1 and 2. RNA (structure 1 in salmon and structure 2 in red), GTase domain (structure 1 in steel blue and structure 2 in slate blue), TPase domain (structure 1 in magenta) are depicted.

(B) RNA handover between active sites. The TPase domain cleaves the gamma phosphate from the nascent triphosphate RNA. GTase NT-domain recognizes the RNA diphosphate end, swings-in by approximately 25 Å using the flexible hinge region, and displaces the TPase to capture the RNA. The clash (black dashed outline) illustrates the requirement for TPase domain displacement.

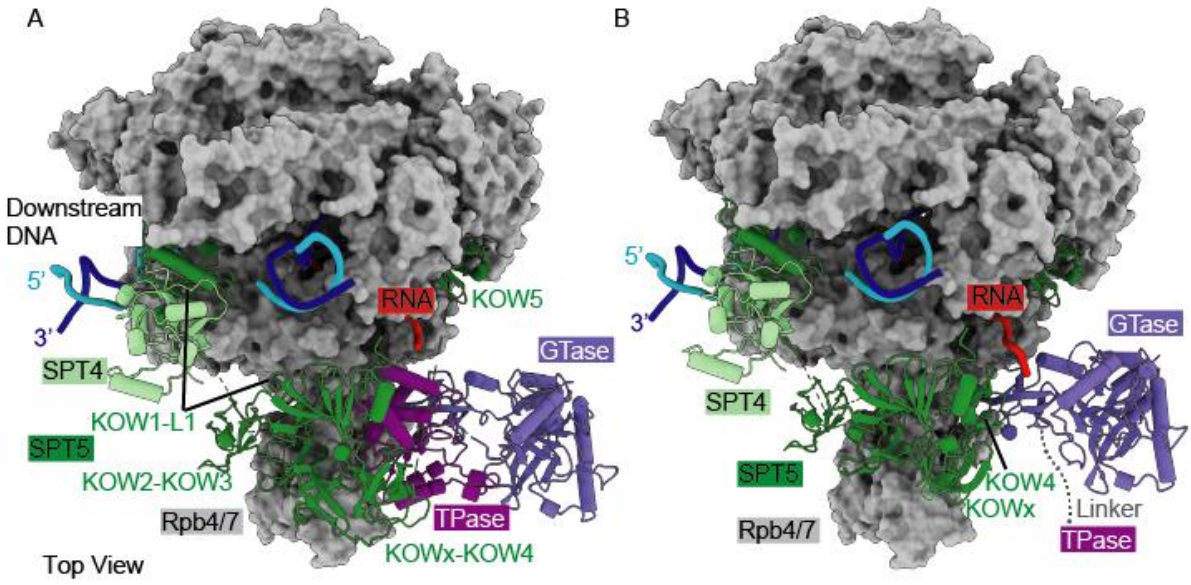


Figure 4: Structures of the transcribing Pol II-DSIF-RNGTT complexes highlighting the modular nature of DSIF (structures 3 and 4). Related to Figure S5.

(A) Structure 3 top view. Structure of the transcribing Pol II-DSIF-RNGTT complex in. Pol II (grey), DSIF (shades of green), RNGTT (purple and slate blue), and RNA (red) are shown. KOW1-L1 is moved away from the upstream DNA. KOW2-KOW3 occupies a similar position as in the transcribing Pol II elongation complex (PDB code 5OIK). KOWx-KOW4 is displaced by TPase domain away from the RNA exit channel and towards RPB7 OB-fold.

(B) Structure 4 top view. Structure of the transcribing Pol II-DSIF-RNGTT complex in the GTase step (structure 4). After TPase domain is displaced (no density observed), and the GTase NT-domain moves in. KOW4 moves in to position itself again the RNA exit channel where TPase was bound in structure 3.

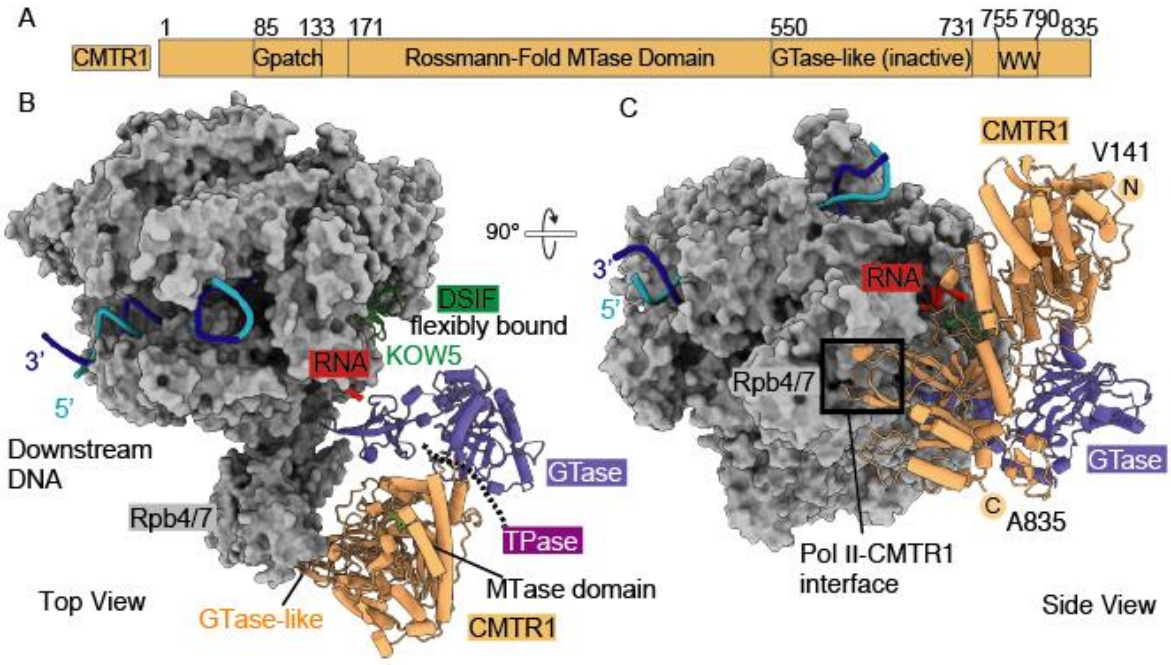


Figure 5: Structure of the transcribing Pol II-DSIF-RNGTT-CMTR1 intermediate (structure 5). Related to Figure S6.

(A) Domain architecture of the human CMTR1 (sandy brown). Important elements and domain boundaries are indicated.

(B) Structure 5 top view. Overview of the transcribing Pol II-DSIF-RNGTT-CMTR1 intermediate. RNA (red) is directed towards CMTR1 (sandy brown), and GTase domain (slate blue) is in an open conformation and moved away from the RNA exit tunnel. Pol II (grey), KOW5 (forest green) are also shown.

(C) Structure 5 side view. Overview of the transcribing Pol II-DSIF-RNGTT-CMTR1 intermediate. Pol II-CMTR1 interface is highlighted (black square).

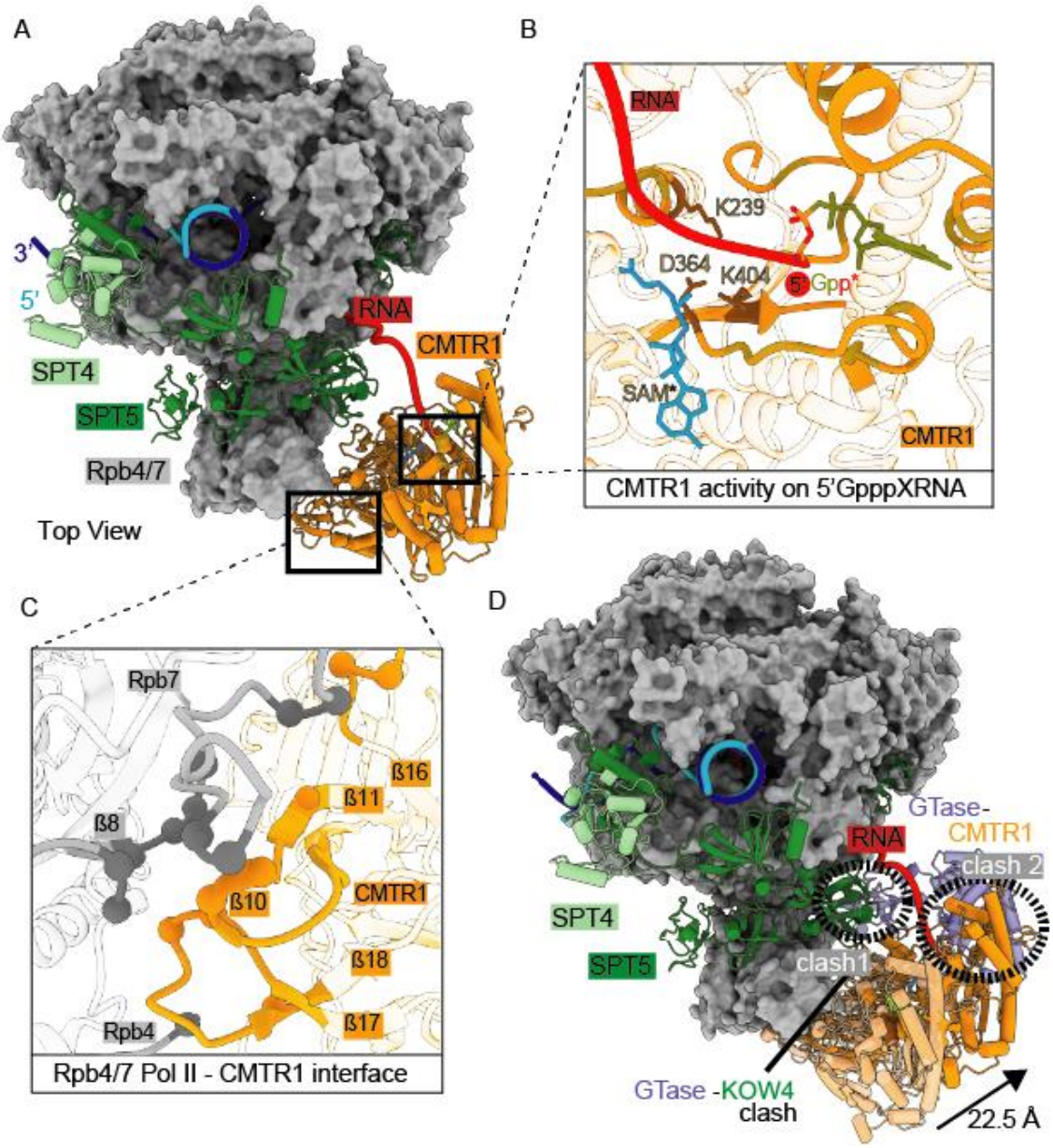

A
3'
5'
RNA
CMTR1
SPT4
SPT5
Rpb4/7
Top View
B
RNA
K239
D364
K404
5'Gpp*
SAM*
CMTR1
CMTR1 activity on 5'GpppXRNA
C
Rpb7
ß16
ß11
ß8
ß10
CMTR1
ß18
ß17
Rpb4
Rpb4/7 Pol II - CMTR1 interface
D
GTase-
CMTR1
RNA
clash 2
SPT4
SPT5
clash1
GTase -KOW4
clash
22.5 Å

Figure 6: Structure of the transcribing Pol II-DSIF-CMTR1 bound to RNA (structure 6). Related to Figure S6.

(A) Structure 6 top view. Overview of transcribing Pol II-DSIF-CMTR1 complex. Shown are Pol II (grey), RNA (red), DSIF (shades of green), and CMTR1 (orange).

(B) RNA-CMTR1 interface. Residues of the active KDK-triad (saddle brown) and additional residues (goldenrod) involved in RNA (red) path into the active site are shown (PDB code 4N48). RNA is shown with a cap-1 structure and SAM is shown in dodger blue (superimposed from PBD 4N48, not used in sample preparation).

(C) RPB7 OB-fold (grey) loop residues (124-142) between the ß8 and ß9 form an extensive interface with the residues between ß10- ß11 and ß17- ß18 of the inactive GTase-like domain of CMTR1 (orange). C-alpha atoms of the mentioned, and some additional residues within the allowed distance to form hydrophobic, ionic, or H-bond interactions are shown.

(D) Superposition of structure 5 and 6 show the 22.5 Å movement of the CMTR1 MTase domain towards the RNA exit channel. DSIF also moves back into the canonical conformation. Both DSIF and CMTR1 in the new position clash with GTase. Hence, these movements likely cause GTase ejection, evident in loss of any GTase density in structure 6.

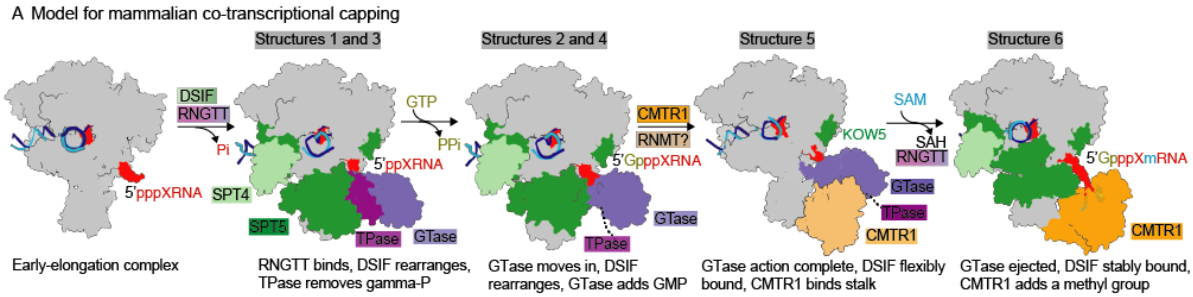


Figure 7: Model for mammalian co-transcriptional capping.

As the RNA starts to appear from the RNA exit tunnel during early elongation, elongation factor DSIF binds. RNGTT is recruited to Pol II via direct interactions and recognizes the RNA around 22-24 nt. This induces DSIF rearrangement. TPase binds 5'pppXRNA and removes the gamma-phosphate generating 5'ppXRNA. GTase domain displaces the TPase domain upon cleavage completion, and moves in to add a GMP moiety resulting in 5'GpppXRNA. As the RNA grows to around 29 nt, cap-1 end of the RNA is efficiently transferred to CMTR1 as an intermediate with RNGTT and CMTR1 is observed. Loss of GTase affinity for the RNA, binding of CMTR1, and return of DSIF into its canonical conformation, likely contribute to RNGTT ejection from the complex. Upon cap1 formation, CMTR1 should also be released from the complex, and the RNA poised for further downstream processing.

Lead Contact

Correspondence should be addressed to P.C. (patrick.cramer@mpinat.mpg.de). Requests for materials should be addressed to P.C. (patrick.cramer@mpinat.mpg.de).

Materials Availability

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

Method Details

Cloning and protein purification

H. sapiens cDNA was used as a polymerase chain reaction (PCR) template for cloning RNGTT into a modified pFastBac vector (a gift from S. Gradia, UC Berkeley, vector 438-C, Addgene: 55220) via ligation independent cloning. The construct contains an N-terminal 6×His tag followed by a maltose binding protein (MBP) tag and a tobacco etch virus (TEV) protease cleavage site. It was then cloned into series 1B via ligation independent cloning (Gradia et al., 2017). RNGTT (C126S inactive variant) was cloned using site directed mutagenesis and the TPase and GTase domains were cloned using around-the-horn PCR (Moore and Prevelige, 2002). A cDNA clone encoding full-length human CMTR1 (pcDNA3.1-C-(k)DYK vector) was obtained from GenScript (clone ID NM_015050.3).

Sus scrofa RNA polymerase II, *Homo sapiens* DSIF, NELF and RNA were purified as previously described (Bernecky et al., 2016; Fianu et al.; Qin and Pyle, 1999; Rinaldi et al., 2015; Vos et al., 2018). RNGTT and CMTR1 variants were expressed and purified as described below unless otherwise stated.

Plasmids were transformed into *Escherichia coli (E.coli)* BL21(DE3)RIL cells. They were grown at 37 °C until OD600nm reached 0.5-0.6. Induction was carried out using 0.75 mM IPTG overnight at 18 °C. Cells were harvested by centrifugation, resuspended in lysis buffer (300 mM NaCl, 20 mM Na-HEPES pH 7.4, 10% (v/v) glycerol, 30 mM imidazole pH 8.0, 1 mM DTT, 0.284 µg ml^{-1} leupeptin, 1.37 µg ml^{-1} pepstatin A, 0.17 mg ml^{-1} PMSF, and 0.33 mg ml^{-1} benzamidine), flash frozen and stored at -80 °C.

All steps were done at 4°C unless otherwise stated. All buffers contained 1 mM DTT and protease inhibitors unless otherwise stated. *E.coli* pellets were sonicated in the lysis buffer. The lysate was centrifuged at 27,000 rpm, clarified with 0.45 µm filter, and then loaded onto a 5 mL HisTrap HP column (GE Healthcare Life Sciences). The elution was carried out with a linear gradient in the elution buffer (lysis buffer containing 500 mM Imidazole). The eluted protein was cleaved with TEV protease overnight and dialysed into the lysis buffer containing 150 mM Imidazole. The cleaved protein was loaded again onto the HisTrap HP column to remove cleaved tag and uncleaved

protein. The flow-through (containing the protein) was then loaded onto a 5 mL HiTrap Q column (GE Healthcare Life Sciences) for anion exchange. It was eluted with a linear gradient using buffer (300 mM NaCl, 20 mM Na-HEPES pH 7.4, 10% (v/v) glycerol) and elution buffer (1 M NaCl, 20 mM Na-HEPES pH 7.4, 10% (v/v) glycerol). Finally, the residual degradation products and aggregates were removed using gel filtration chromatography with a HiLoad 16/600 Superdex 200 pg column (GE Healthcare Life Sciences) equilibrated in a gel filtration buffer (300 mM NaCl, 20 mM Na-HEPES pH 7.4, 10% (v/v) glycerol). Peak fractions were concentrated with Amicon spin concentrator of an appropriate molecular weight cut-off to a concentration of 3-5 mg/mL.

RNA guanylytransferase assay

RNA guanlylyation assays were performed with a 20 nt-24 nt RNA substrate. Transcribing Pol II elongation complexes (EC) were assembled with 75 pmol Pol II as previously described (Bernecky et al., 2016; Vos et al., 2018). 50 pmol CDK7 was added, and the complex was incubated 5 min at 37 °C. 150 pmol DSIF, 150 pmol RNGTT, and 0.5 mM GTP were added either to free RNA (control) or an assembled EC. Finally, water, and a 2-fold compensation buffer were added to reach the final buffer conditions of 150 mM NaCl, 20 mM Na·HEPES, pH 7.4 (@ 20°C), 5 mM $MgCl_2$, 1 mM DTT, 10 % (v/v) glycerol and a reaction volume of 20 µL. The reaction was further incubated for 5 min at 37 °C. DNase I (RNase free, 2 units, NEB) was added to remove the template and the non-template, followed by addition of Proteinase K (0.8 units, NEB). Finally, 10% glycerol was added to be able to load the sample on a denaturing Urea-PAGE gel (12 µL applied to an 8 M urea, 1 x TBE, 20% acrylamide:bis-acrylamide 19:1 gel). The gel was run in 0.5 x TBE at 300 V for 150 min. Products were visualized using Syber Gold staining and a Typhoon 9500 FLA Imager (GE Healthcare Life Sciences).

Sample preparation for cryo-EM

Pol II-RNGTT for cryo-EM was assembled in a final buffer containing 100 mM NaCl, 20 mM Na·HEPES, pH 7.4 (@ 20°C, 3 mM $MgCl_2$, 1 mM DTT and 4% (v/v) glycerol. RNA(5′-rGrUrA rArCrC rGrGrA rGrArG rGrGrA rArCrC rCrArC rUrGrC-3′) and template DNA (5′-GCT CCC AGC TCC CTG CTG GCT CCG AGT GGG TTC TGC CGC TCT CAA TGG-3′) were annealed to form a hybrid (50 µM) using the thermocycler. All following concentrations refer to final concentrations in 100 µL complex formation reaction. Pol II (1 µM) and hybrid (1 µM) were incubated for 15 min on ice. Non-template DNA (5′-CCA TTG AGA GCG GCC CTT GTG TTC AGG AGC CAG CAG GGA GCT GGG AGC-3′) (1.5 µM) was added, and the sample incubated for 10 min on ice. CDK7 (0.5 µM) was added, and the sample further incubated for 5 min at 30 °C.

To assemble the Pol II-RNGTT complexes, RNGTT (C126S/WT) (3 µM) was added. To

assemble the Pol II-DSIF-RNGTT complex, DSIF (1.5 µM), RNGTT (3 µM) and NELF (4 µM) were added. To assemble the Pol II-DSIF-RNGTT-CMTR1, DSIF (1.5 µM), RNGTT (3 µM), CMTR1 (3 µM), and NELF (4 µM) were added. Finally, water, and a 10-fold compensation buffer were added to reach the final buffer conditions, and the sample further incubated for 30 min at 30°C. The sample was then centrifuged and applied onto a Superose 6 3.2/300 Increase size exclusion column (GE Healthcare), equilibrated in complex buffer (100 mM NaCl, 20 mM Na·HEPES, pH 7.4 (@ 20°C), 3 mM $MgCl_2$, 1 mM DTT, 4% (v/v) glycerol). Eluted fractions were analysed using SDS-PAGE.

Peak fractions of all complexes were individually cross-linked using 0.1% (v/v) glutaraldehyde for 8 min on ice. The cross-linker was quenched with 8 mM apartate and 2 mM lysine and incubated for 10 min on ice. Fractions were then dialyzed against the dialysis buffer (100 mM NaCl, 20 mM Na·HEPES, pH 7.4 (@ 20°C), 3 mM $MgCl_2$, 1 mM DTT) for 4 hours.

For grid preparation, R2/2 gold grids, Au 200 mesh (Quantifoil) were glow-discharged twice for 100 s. 2 µL of the dialyzed complexes with an approximate concentration of 200 nM complex were applied to each side of the grid. After an incubation of 10 s, grids were blotted and vitrified by plunging into liquid ethane using a Vitrobot Mark IV (Thermo Fisher). The Vitrobot was operated at 4°C and 100% humidity, and the blotting done using blot force of 5 and blot time of 3.5-6 s. The grids were clipped and subsequently stored in liquid nitrogen for data acquisition.

Crosslinking mass spectrometry

Sample 1 and 2 refers to Pol II-RNGTT, sample 3 to Pol II-DSIF-RNGTT-NELF, and sample 4 Pol II-DSIF-RNGTT-CMTR1. 70-110 µg aliqutos of the protein complex were crosslinked with 2 mM (sample 1, 2) or 0.5 mM BS3 (sample 3, 4). The cross-linking reaction was carried out for 45 min at room temperature (RT) and was quenched with 50 mM TRIS for 15 min at RT. Cross-linked proteins were either precipitated with 4 volumes ice-cold acetone o/n and washed with 80% ethanol (sample 3, 5) or processed directly. The crosslinked proteins were resuspended in 8 M Urea/50 mM ammonium bicarbonate (ABC, pH 8.0) (samples 3) or in 2 M Urea/50 mM ammonium bicarbonate (ABC, pH 8.0) (sample 1,2,4). Proteins were reduced for 30 min at RT with 10 mM dithiotreitol (DTT) and were alkylated for 30 min at RT in the dark by adding iodacetamide (IAA) to a final concentration of 55 mM. Samples were diluted to 1 M Urea and digested for 30 min at 37 °C with 4 µl Pierce Universal Nuclease (250 U/µl) in the presence of 2 mM $MgCl_2$. Trypsin (Promega) digest was performed o/n at 37 °C in a 1:50 enzyme/protein ratio, the reaction was terminated with 0.2 % (v/v) FA. Tryptic peptides were desalted on MicroSpin Columns (Harvard Apparatus) following manufacturer´s instruction and vacuum-dried.

Dried peptides were either resuspended in 50 µl 30 % acetonitrile/0.1 % TFA and enriched by peptide size exclusion chromatography (Superdex Peptide PC3.2/300 column, GE Healthcare, flow rate 50 µl/min; sample 1, 2) or resuspended in 50 µl 10 mM ammonium hydroxide and

separated via ph Reverse Phase (bRP) on a RP-Vanquish Flex UHPLC-System (Waters XBridge C18 column (3.5 µm particles, 1.0 mm x 150 mm) at a flow rate of 50 µl/min and collected with fraction concatenation (sample 3, 4)(Yang et al., 2012).

Fractions containing the cross-linked peptides were vacuum dried and dissolved in 4% ACN/0.1% TFA (v/v) for subsequent LC-MS/MS analysis.LC-MS/MS analysis: Cross-linked peptides derived from pSEC or bRP were analyzed as technical duplicates on a Thermo QExactive HF-X (sample 1, 2) or a Thermo Obitrap Exploris mass spectrometer (sample 3 and 4). Peptides were loaded onto a Dionex Ultimate 3000 RSLCnano equipped with a custom column (ReproSil-Pur 120 C18-AQ, 1.9 µm pore size, 75 µm inner diameter, 30 cm length, Dr. Maisch GmbH).

Peptides were separated applying the following gradient: mobile phase A consisted of 0.1 % formic acid (FA, v/v), mobile phase B of 80 % ACN/0.08 % FA (v/v). The gradient started at 5 % B, increasing to 10 (for bRP fractions), 15 or 20 % B within 3 min (for pSEC fractions), followed by a continuous increase to 48 % B within 45 min, then keeping B constant at 90 % for 8 min. After each gradient, the column was again equilibrated to 5 % B for 2 min. The flow rate was set to 300 nL/min.

MS1 spectra were acquired with a resolution of 120,000 in the orbitrap (OT) covering a mass range of 380–1600 m/z. Dynamic exclusion was set to 30 s. The top 30 abundant precursor ions (charge state 3-8) were triggered for HCD fragmentation (30 % NCE). MS2 spectra were recorded with a resolution of 30,000 in OT and maximal injection time was set to 54 ms and the isolation window was set to 1.4 or 1.6 m/z.

For data analysis, raw files were searched against a database containing the sequences of the proteins of the complex and analyzed via pLink 2.3.9 at a false discovery rate (FDR) of 1 %(Chen et al., 2019). Carbamidomethylation of cysteines was set as fixed modification, oxidation of methionines as variable modification. The database contained all proteins within the complex. For further analysis only interaction sites with 3 cross-linked peptide spectrum matches were taken into account. Cross-links were displayed with xiNET and Xlink Analyzer combined with UCSF Chimera. Due to the highly dynamic nature of the complex, and the presence of Pol II dimers in the sample, some long-distance crosslinks were observed. In order to filter the data reasonably, high threshold values between 14-18 were used in UCSF chimera (Combe et al., 2015; Kosinski et al., 2015; Pettersen et al., 2004).

Cryo-EM analysis and image processing

Cryo-EM data was collected on a Titan Krios G2 transmission electron microscope (FEI), equipped with a K3 direct detector and a GIF Quantum Filter (Gatan), operated at 300 keV. All data sets were acquired using SerialEM (Mastronarde, 2005) at a nominal magnification of 81,000×,

corresponding to a pixel size of 1.05 Å/pixel in nanoprobe EF-TEM mode with a slith width of 20 eV.

For the transcribing Pol II-RNGTT (C126S) structure, a total of 4698 movie stacks over 40 frames was collected in counting mode over 1.2 s, with a defocus range of −0.5 μm to −2.0 μm. The dose rate was 33.89 e^- per $Å^2$ per s for a total dose of 42.43 e^- per $Å^2$.

For the transcribing Pol II-RNGTT (WT), a total of 3236 movie stacks over 40 frames was collected in counting mode over 2.07 s, with a defocus range of −0.5 μm to −2.0 μm. The dose rate was 19.86 e^- per $Å^2$ per s for a total dose of 41.26 e^- per $Å^2$.

For the transcribing Pol II-DSIF-RNGTT, a total of 2900 movie stacks over 40 frames was collected in counting mode over 2.13 s, with a defocus range of −0.5 μm to −2.0 μm. The dose rate was 18.93 e^- per $Å^2$ per s for a total dose of 40.32 e^- per $Å^2$.

For the transcribing Pol II-DSIF-RNGTT-CMTR1, a total of 8620 movie stacks over 40 frames was collected in counting mode over 1.9 s, with a defocus range of −0.5 μm to −2.0 μm. The dose rate was 15.06 e^- per $Å^2$ per s for a total dose of 40.65 e^- per $Å^2$.

Micrographs for all the data sets were subjected to motion correction, dose weighting, contrast transfer function (CTF) estimation and particle picking on the fly using Warp (Tegunov and Cramer, 2019). Subsequent image processing was done in RELION 3.0.7 or 3.1.0 (Zivanov et al., 2018). Particles were extracted binned with a box of 150 pixels and a pixel size of 2.1 Å/pixel. These particles were then subjected to an initial cleanup by 2D classification in cryoSPARC (Punjani et al., 2017). This was followed by an *ab initio* model generation, and a heterogeneous refinement to remove the non-aligning particles.

For the Pol II-RNGTT (C126S) dataset, a total of 3,049,188 particles were picked. After the initial cyoSPARC steps, the remaining 1,364,586 particles were exported to RELION 3.1.0 and unbinned. To select particles containing Pol II and RNGTT, a global classification was performed with masks encompassing the region of extra density. Subsuquet masked classifications were performed to get the TPase. 34,647 particles were finally refined and a reconstruction was obtained at 3.6 Å resolution. Local resolutions of the map were estimated using the RELION built-in tool (Fig S2).

For the Pol II-RNGTT (wild type; WT) dataset, a total of 3,261,338 particles were picked. After the initial cyoSPARC pipeline, the remaining 1,069,668 particles were exported to RELION 3.1.0 and unbinned. To select particles containing RNGTT, classifications were performed with masks encompassing RNGTT or the stalk. 132,949 particles were selected. Finally, masked classification for GTase NT-domain was performed, 21,150 particles were selected refined and a reconstruction was obtained at 3.2 Å resolution. Local resolutions of the map were estimated using the RELION built-in tool (Fig S3).

For the Pol II-DSIF-RNGTT dataset, a total 3,299,682 particles were picked. After the initial cyoSPARC pipeline, the remaining 1548,296 particles were exported to RELION 3.1.0 and unbinned. To select particles containing DSIF and RNGTT, masked classification was performed with mask encompassing the RNGTT and SPT5 domains, followed by a mask for TPase and KOW2-KOW3. 64,321 particles were selected and refined resulting in structure 3. To select particles with RNA bound to the GTase, masked classification on GTase followed by GTase together with SPT5 domains was performed. 73,948 particles were selected and refined resulting in structure 4. Both the structures 3 and 4 were globally refined and an overall reconstruction was obtained at 3.8 Å and 3.6 Å resolution. Local resolutions of the map were estimated using the RELION built-in tool (Fig S5).

For the Pol II-DSIF-RNGTT-CMTR1 dataset, a total of 3,299,064 particles were picked. After the initial cyoSPARC pipeline, the remaining 1,240,321 particles were exported to RELION 3.1.0 and unbinned. To select particles containing RNGTT, or CMTR1 classifications were performed with masks encompassing both. Two classes of interest were obtained. 48,148 particles were selected and refined resulting in structure 5. 252,156 particles were selected and further classified with a mask encompassing DSIF, CMTR1 and SPT5. 31,427 particles were then selected and refined resulting in structure 6. Both the structures 5 and 6 were globally refined and an overall reconstruction was obtained at 3.9 Å and 4.0 Å resolution, respectively. Local resolutions of the map were estimated using the RELION built-in tool (Fig S6).

Model building

For all the transcribing capping complex structures (Pol II-RNGTT, Pol II-DSIF-RNGTT, Pol II-DSIF-RNGTT-CMTR1), structures of Pol II and DSIF from the paused elongation complex (PDB code 6GML) were separately ridig body docked. Structures of the TPase domain (PDB code 2C46), GTase domain (3S24), and CMTR1 (PDB code 4N48, predicted AF-Q8N1G2) modelled into the final map and refined using COOT (Emsley and Cowtan, 2004). Pol II stalk was modelled in separately for each density. The model was real-space refined using PHENIX (Afonine et al., 2018) with secondary structure restraints against the final local-resolution filtered map.

For the PolII-RNGTT(C126S) structure, GTase chain F was modelled into the density for the open conformation. TPase domain chain A was used, and surface loop residues were modelled to fit the density. For the PolII-RNGTT(WT) structure, GTase domain chain D was used. The OB-domain was modelled into the density, and the NT-domain was separately docked for the closed conformation.

For Pol II-DSIF-RNGTT structure, SPT5 domains were separately modelled. KOWx-KOW4 and the TPase domains were rigid body docked into the locally filtered maps generated using Relion 3.1. GTase from structure 1 and 2 were rigid body docked into densities for structures

3 and 4, respectively.

For the Pol II-DSIF-RNGTT-CMTR1 structure, MTase domain crystal structure and AlphFold2 predicted C-terminal region were ridid body docked separately. For all the structures, density in the active site of Pol II allowed unambiguous assignment of DNA and DNA-RNA register until the RNA exit channel.

The RNA was then built into the density to trace the path up to the active sites of the enzymes based on the sequence and the length used. The RNA in the active site of CMTR1 was docked together with the crystal structure. GTP in GTase active site, and SAM in CMTR1 active site were also superimposed from the crystal structures 1CKN and 4N48, respectivel, and were not present in sample preparation.. The models were real-space refined using PHENIX with secondary structure restraints against the global refinement.

Figure generation

Figures were generated using UCSF ChimeraX (version 1.3) (Goddard et al., 2018) and Adobe Illustrator.

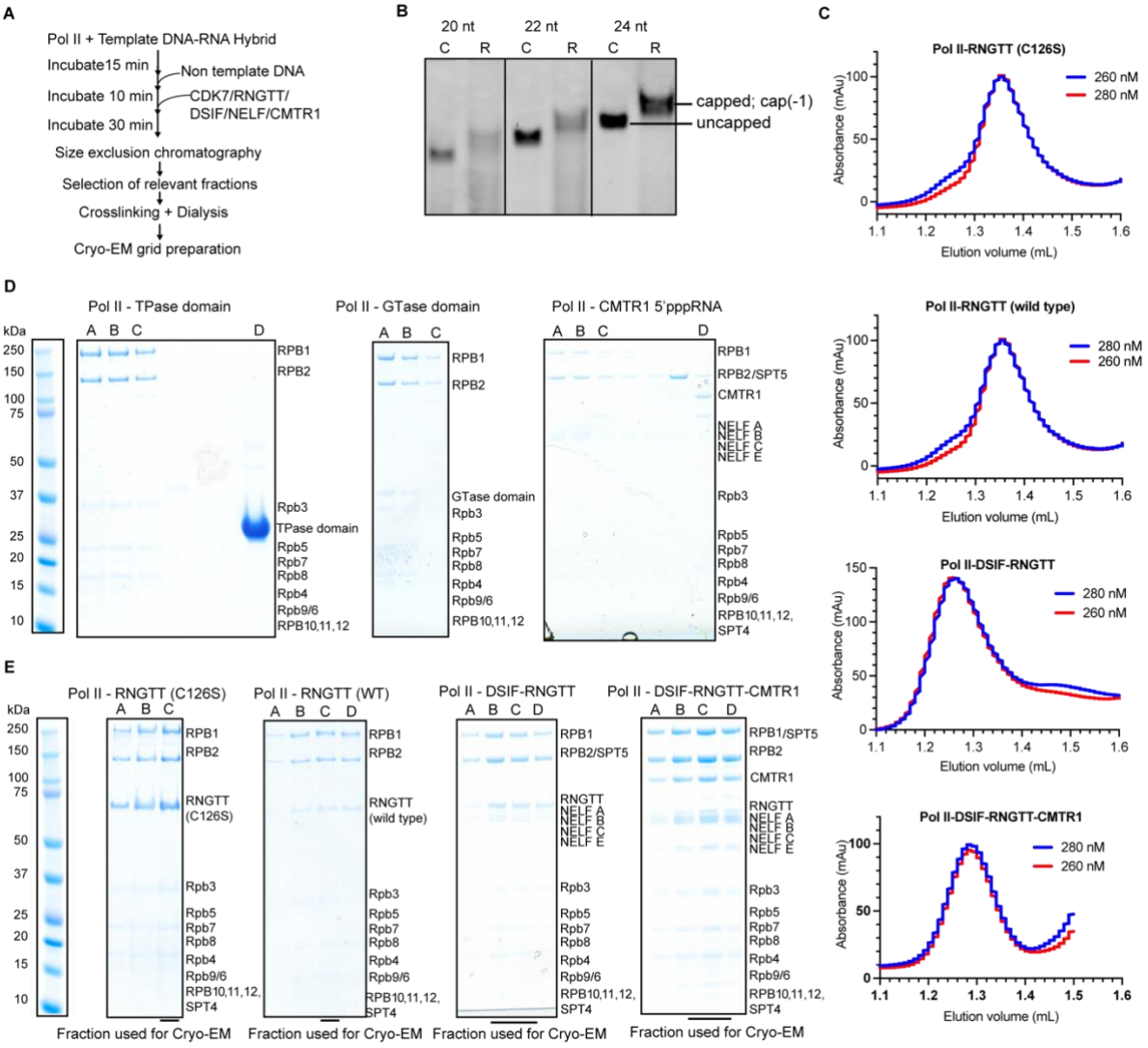


Figure S1: Protein components, GTase assay and complex formation. Related to Figure 1 to 6.

(A) SDS-PAGE of all purified protein components (0.75 µg per lane).

(B) GTase assay in the presence of CDK7, RNGTT, DSIF with a 20 nt, 22 nt and 24 nt uncapped substrate. Each reaction has a corresponding free RNA control.

(C) Complex assembly protocol.

(D) SDS-PAGE of binding trials by gel filtration for transcribing Pol II-TPase (C126S), Pol II-GTase, and Pol II-CMTR1 (with 5'pppXRNA).

(D) SDS-PAGE for complex assembly of transcribing Pol II-RNGTT (C126S), Pol II-RNGTT (WT), Pol II-DSIF-RNGTT and Pol II-DSIF-RNGTT-CMTR1 complexes.

(E) Gel filtration profiles for transcribing Pol II-RNGTT (C126S), Pol II-RNGTT (WT), Pol II-DSIF-RNGTT and Pol II-DSIF-RNGTT-CMTR1 complexes.

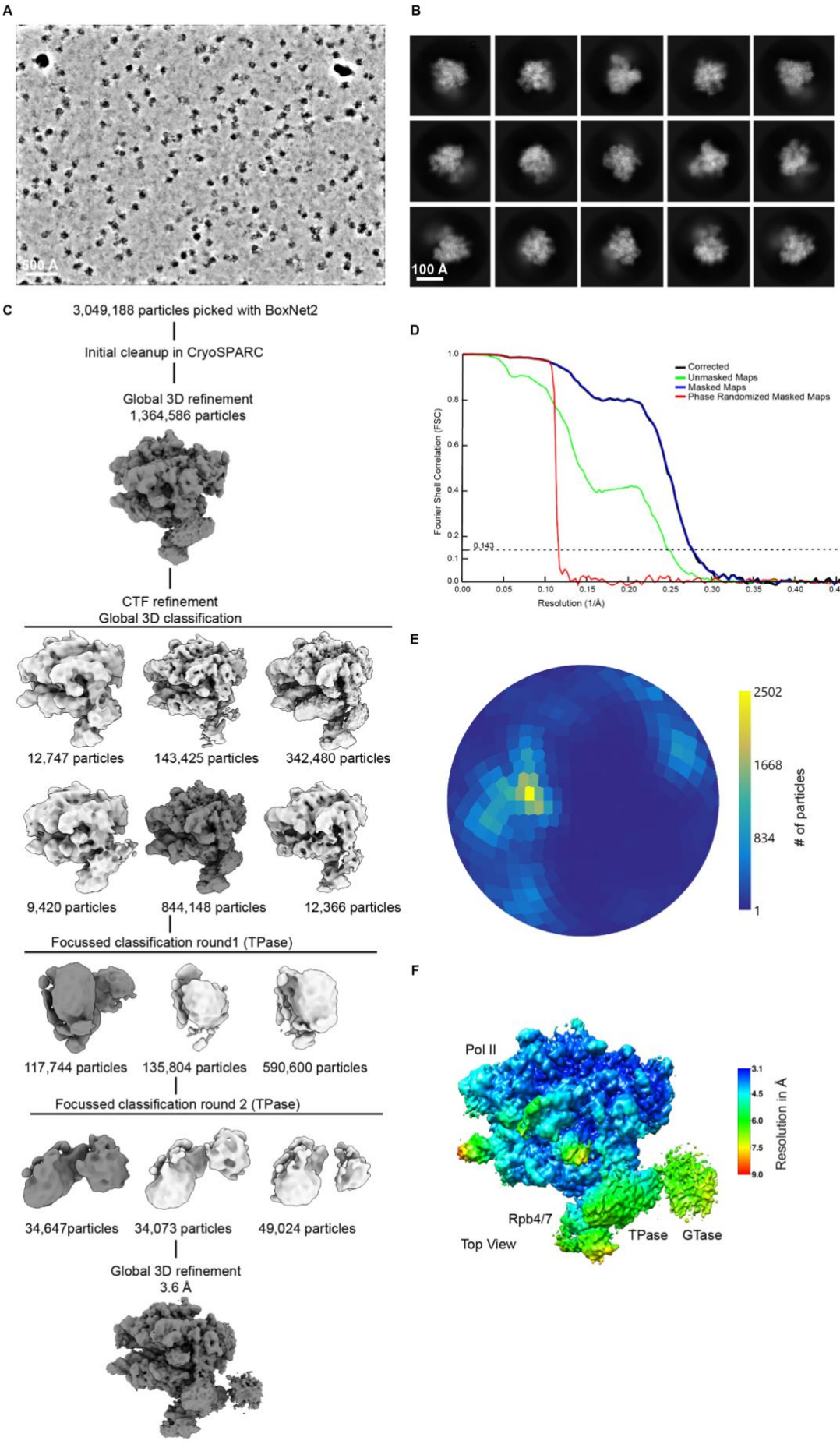

A
100 Å
B
100 Å
C
3,049,188 particles picked with BoxNet2
Initial cleanup in CryoSPARC
Global 3D refinement
1,364,586 particles
CTF refinement
Global 3D classification
12,747 particles
143,425 particles
342,480 particles
9,420 particles
844,148 particles
12,366 particles
Focussed classification round1 (TPase)
117,744 particles
135,804 particles
590,600 particles
Focussed classification round 2 (TPase)
34,647particles
34,073 particles
49,024 particles
Global 3D refinement
3.6 Å
D
Corrected
Unmasked Maps
Masked Maps
Phase Randomized Masked Maps
Fourier Shell Correlation (FSC)
0.143
Resolution (1/Å)
E
2502
1668
834
1
# of particles
F
Pol II
Rpb4/7
TPase
GTase
Top View
3.1
4.5
6.0
7.5
9.0
Resolution in Å

Figure S2: Data acquisition, processing, and data quality metrics for the transcribing Pol II-RNGTT (C126S) complex. Related to Figure 1.

(A) Representative denoised micrograph of data collection with scale bar (500 Å).

(B) 2D classes show RNA polymerase II and fuzzy density for RNGTT with scale bar of 100 Å.

(C) Sorting and classification tree of transcribing Pol II-RNGTT(C126S) complex dataset.

(D) FSC curve of overall map.

(E) Angular distribution of particles used for final recobstrontion of structure 1.

(F) Local resolution of Pol II-RNGTT (C126S) complex.

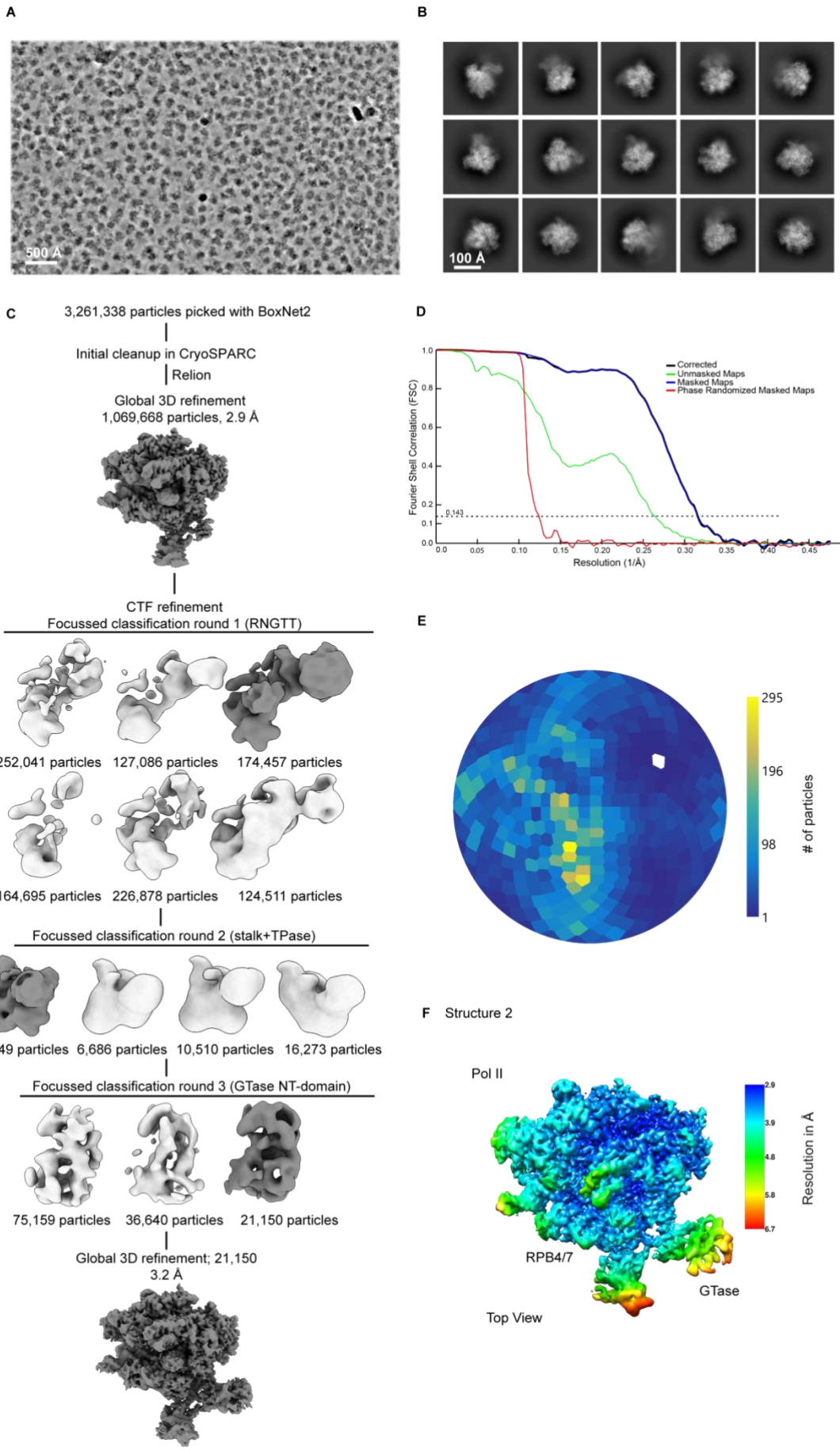

A
500 Å
B
100 Å
C
3,261,338 particles picked with BoxNet2
Initial cleanup in CryoSPARC
Relion
Global 3D refinement
1,069,668 particles, 2.9 Å
CTF refinement
Focussed classification round 1 (RNGTT)
252,041 particles
127,086 particles
174,457 particles
164,695 particles
226,878 particles
124,511 particles
Focussed classification round 2 (stalk+TPase)
132,949 particles
6,686 particles
10,510 particles
16,273 particles
Focussed classification round 3 (GTase NT-domain)
75,159 particles
36,640 particles
21,150 particles
Global 3D refinement; 21,150
3.2 Å
D
Corrected
Unmasked Maps
Masked Maps
Phase Randomized Masked Maps
Fourier Shell Correlation (FSC)
0.143
Resolution (1/Å)
E
295
196
98
1
# of particles
F
Structure 2
Pol II
RPB4/7
GTase
Top View
2.9
3.9
4.8
5.8
6.7
Resolution in Å

Figure S3: Data acquisition, processing, and data quality metrics for transcribing Pol II-RNGTT (WT) complex. Related to Figure 2.

(A) Representative denoised micrograph of data collection with scale bar (500 Å).

(B) 2D classes show RNA polymerase II and fuzzy density for RNGTT with scale bar of 100 Å.

(C) Sorting and classification tree of transcribing Pol II-RNGTT(C126S) complex dataset.

(D) FSC curve of overall map.

(E) Angular distribution of particles used for final recobstrontion of structure1.

(F) Local resolution of Pol II-RNGTT (WT) complex.

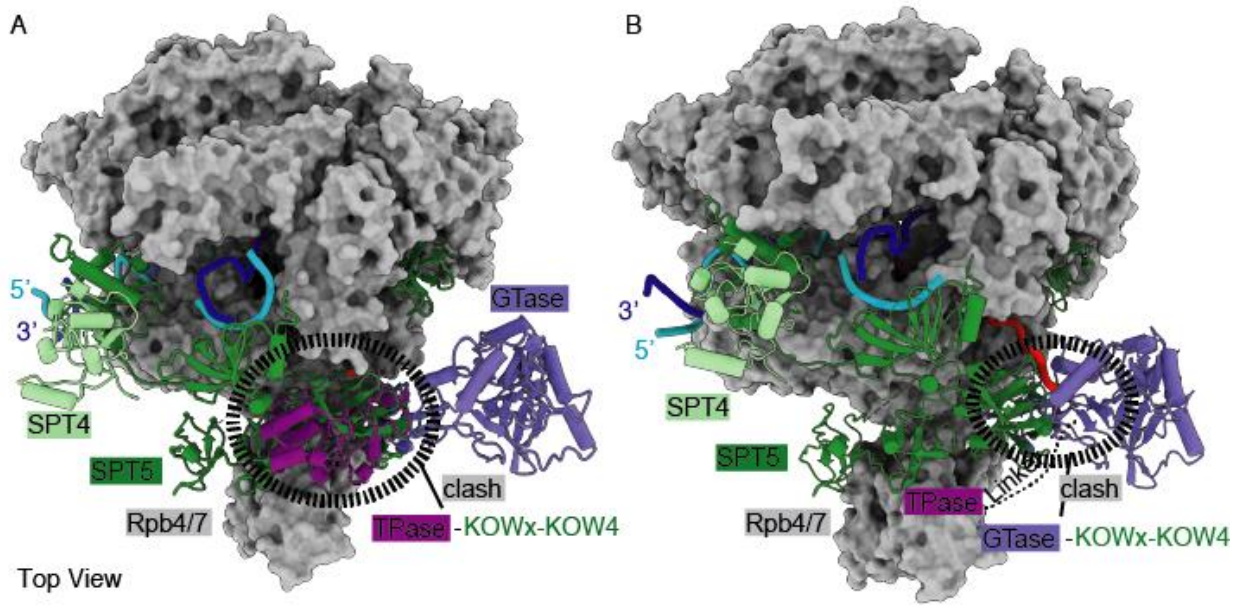


Figure S4: Superposition of structures 1 and 2 with transcribing Pol II elongation complex. Related to figure 1, 2 and 3.

(A) Superposition of Pol II EC-DSIF and structure 1. TPase and SPT5 (forest green) subunit KOWx-KOW4 bind the same site on Pol II stalk, and therefore show heavy clash (black dashed circle).

(B) Superposition of Pol II EC-DSIF and structure 2. GTase and SPT5 (forest green) subunit KOW4 show minor clashes (black dashed circle) around the RNA exit tunnel.

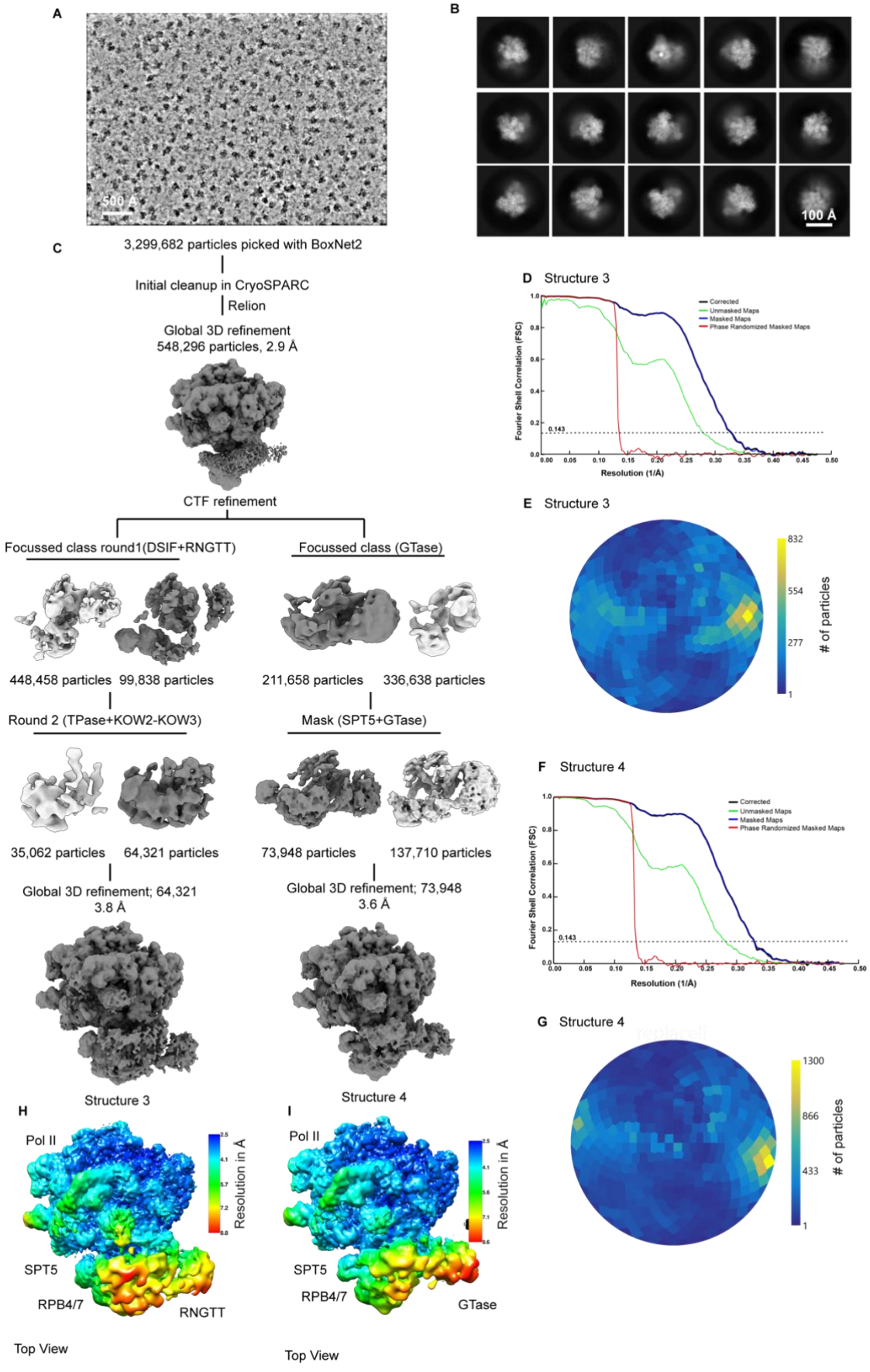

A
500 Å
B
100 Å
C
3,299,682 particles picked with BoxNet2
Initial cleanup in CryoSPARC
Relion
Global 3D refinement
548,296 particles, 2.9 Å
CTF refinement
Focussed class round1(DSIF+RNGTT)
Focussed class (GTase)
448,458 particles
99,838 particles
211,658 particles
336,638 particles
Round 2 (TPase+KOW2-KOW3)
Mask (SPT5+GTase)
35,062 particles
64,321 particles
73,948 particles
137,710 particles
Global 3D refinement; 64,321
3.8 Å
Global 3D refinement; 73,948
3.6 Å
Structure 3
Structure 4
D Structure 3
Corrected
Unmasked Maps
Masked Maps
Phase Randomized Masked Maps
Fourier Shell Correlation (FSC)
0.143
Resolution (1/Å)
E Structure 3
832
554
277
1
# of particles
F Structure 4
G Structure 4
1300
866
433
1
H
Pol II
SPT5
RPB4/7
RNGTT
Resolution in Å
Top View
I
Pol II
SPT5
RPB4/7
GTase
Resolution in Å
Top View

Figure S5: Data acquisition, processing, and data quality metrics for transcribing Pol II-DSIF-RNGTT complex. Related to Figure 4.

(A) Representative denoised micrograph of data collection with scale bar (500 Å).

(B) 2D classes show RNA polymerase II and a fuzzy density for DSIF and RNGTT with scale bar of 100 Å.

(C) Sorting and classification tree of transcribing Pol II-DSIF-RNGTT complex dataset.

(D) FSC curve of overall map for structure 3.

(E) Angular distribution of particles used for final reconstruction of structure3.

(F) FSC curve of overall map for structure 4.

(G) Angular distribution of particles used for final reconstruction of structure3.

(H) Local resolution for structure 3.

(I) Local resolution for structure 4.

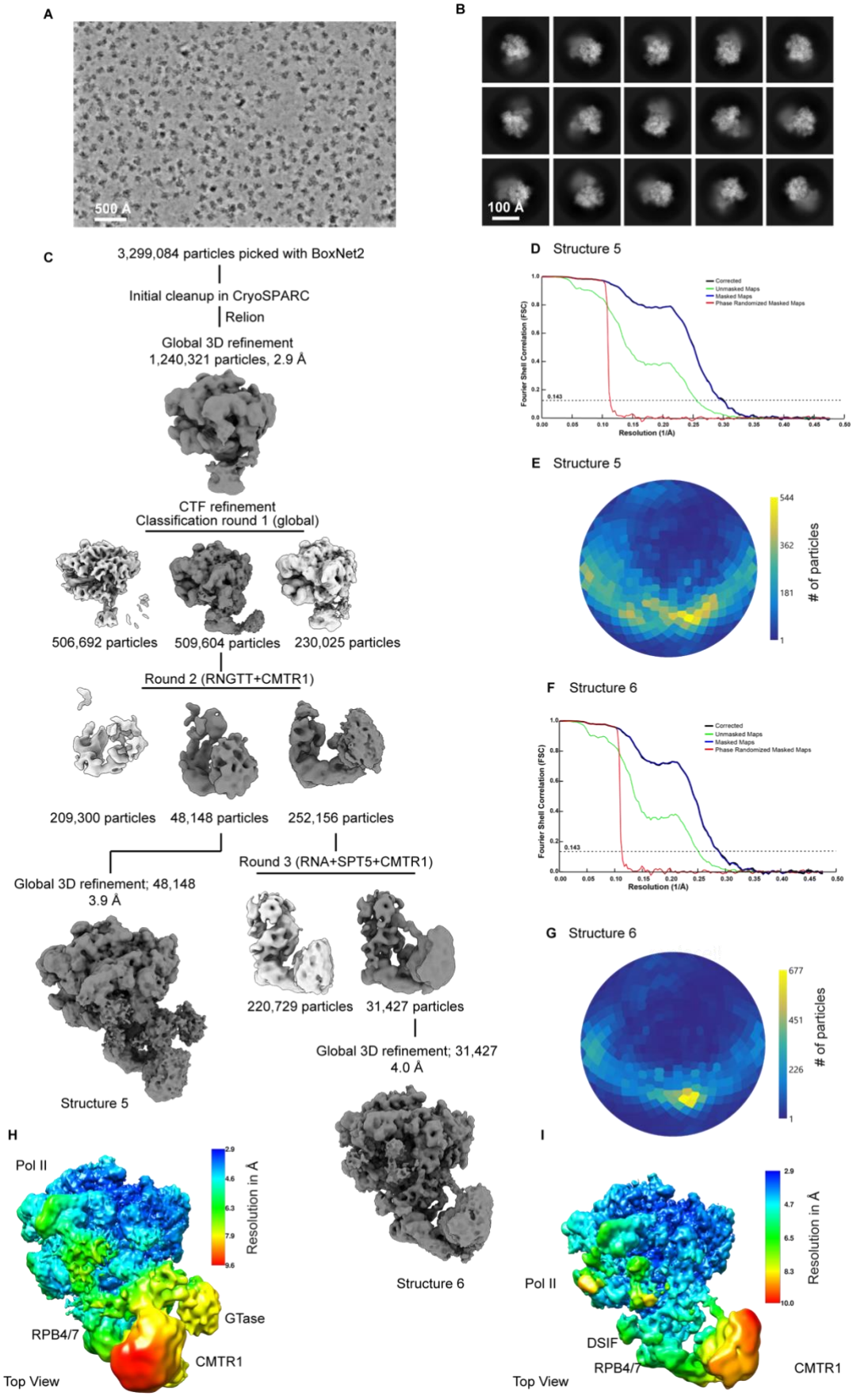

A
500 Å
B
100 Å
C
3,299,084 particles picked with BoxNet2
Initial cleanup in CryoSPARC
Relion
Global 3D refinement
1,240,321 particles, 2.9 Å
CTF refinement
Classification round 1 (global)
506,692 particles
509,604 particles
230,025 particles
Round 2 (RNGTT+CMTR1)
209,300 particles
48,148 particles
252,156 particles
Global 3D refinement; 48,148
3.9 Å
Structure 5
Round 3 (RNA+SPT5+CMTR1)
220,729 particles
31,427 particles
Global 3D refinement; 31,427
4.0 Å
Structure 6
D Structure 5
Corrected
Unmasked Maps
Masked Maps
Phase Randomized Masked Maps
Fourier Shell Correlation (FSC)
Resolution (1/Å)
0.143
E Structure 5
# of particles
F Structure 6
G Structure 6
H
Pol II
Resolution in Å
GTase
RPB4/7
CMTR1
Top View
I
Pol II
DSIF
RPB4/7
CMTR1
Top View

Figure S6: Data acquisition, processing, and data quality metrics for transcribing Pol II-DSIF-RNGTT-CMTR1 complexes. Related to Figure 5 and 6.

(A) Representative denoised micrograph of data collection with scale bar (500 Å).

(B) 2D classes show RNA polymerase II and a fuzzy density for DSIF, RNGTT and CMTR1 with scale bar of 100 Å.

(C) Sorting and classification tree of transcribing Pol II-DSIF-RNGTT-CMTR1 complex dataset.

(D) FSC curve of overall map for structure 5.

(E) Angular distribution of particles used for final reconstruction of structure 5.

(F) FSC curve of overall map for structure 6.

(G) Angular distribution of particles used for final reconstruction of structure 6.

(H) Local resolution for structure 5.

(I) Local resolution for structure 6.

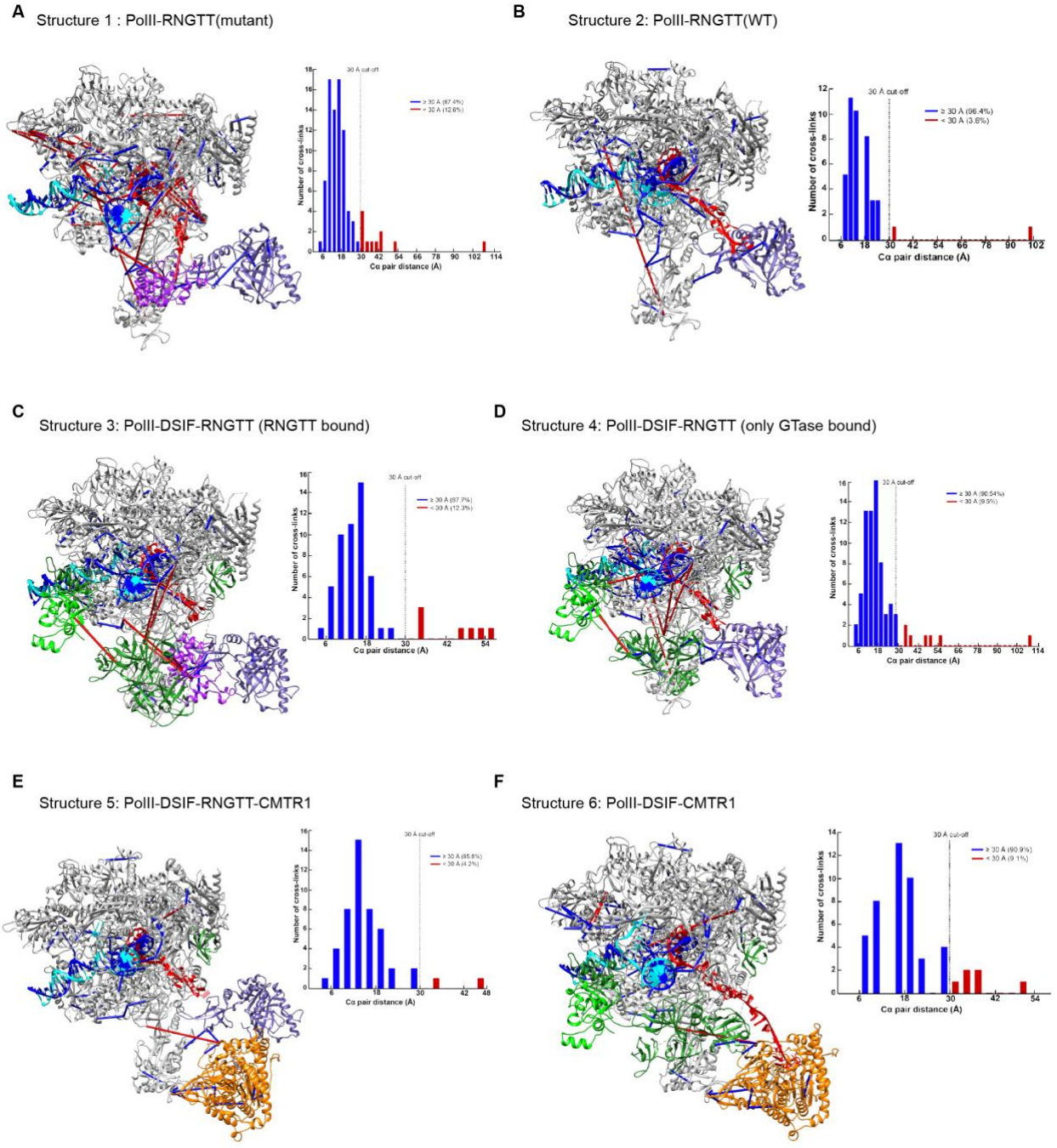

A
Structure 1 : PolII-RNGTT(mutant)
B
Structure 2: PolII-RNGTT(WT)
C
Structure 3: PolII-DSIF-RNGTT (RNGTT bound)
D
Structure 4: PolII-DSIF-RNGTT (only GTase bound)
E
Structure 5: PolII-DSIF-RNGTT-CMTR1
F
Structure 6: PolII-DSIF-CMTR1
Number of cross-links
Cα pair distance (Å)
30 Å cut-off

Figure S7: Cross-linking mass spectrometry representation of the various complexes. Related to figure 1 to 6.

(A) Overview of crosslinks and the corresponding statistics obtained for transcribing Pol II-RNGTT (C126S) complex.

(B) Overview of crosslinks and the corresponding statistics obtained for transcribing Pol II-RNGTT (WT) complex.

(C) Overview of crosslinks and the corresponding statistics obtained for transcribing Pol II-DSIF-RNGTT complex mapped on structure 3.

(D) Overview of crosslinks and the corresponding statistics obtained for transcribing Pol II-DSIF-RNGTT complex mapped on structure 4.

(E) Overview of crosslinks and the corresponding statistics obtained for transcribing Pol II-DSIF-RNGTT-CMTR1 complex mapped on structure 5.

(E) Overview of crosslinks and the corresponding statistics obtained for transcribing Pol II-DSIF-RNGTT-CMTR1 complex mapped on structure 6.

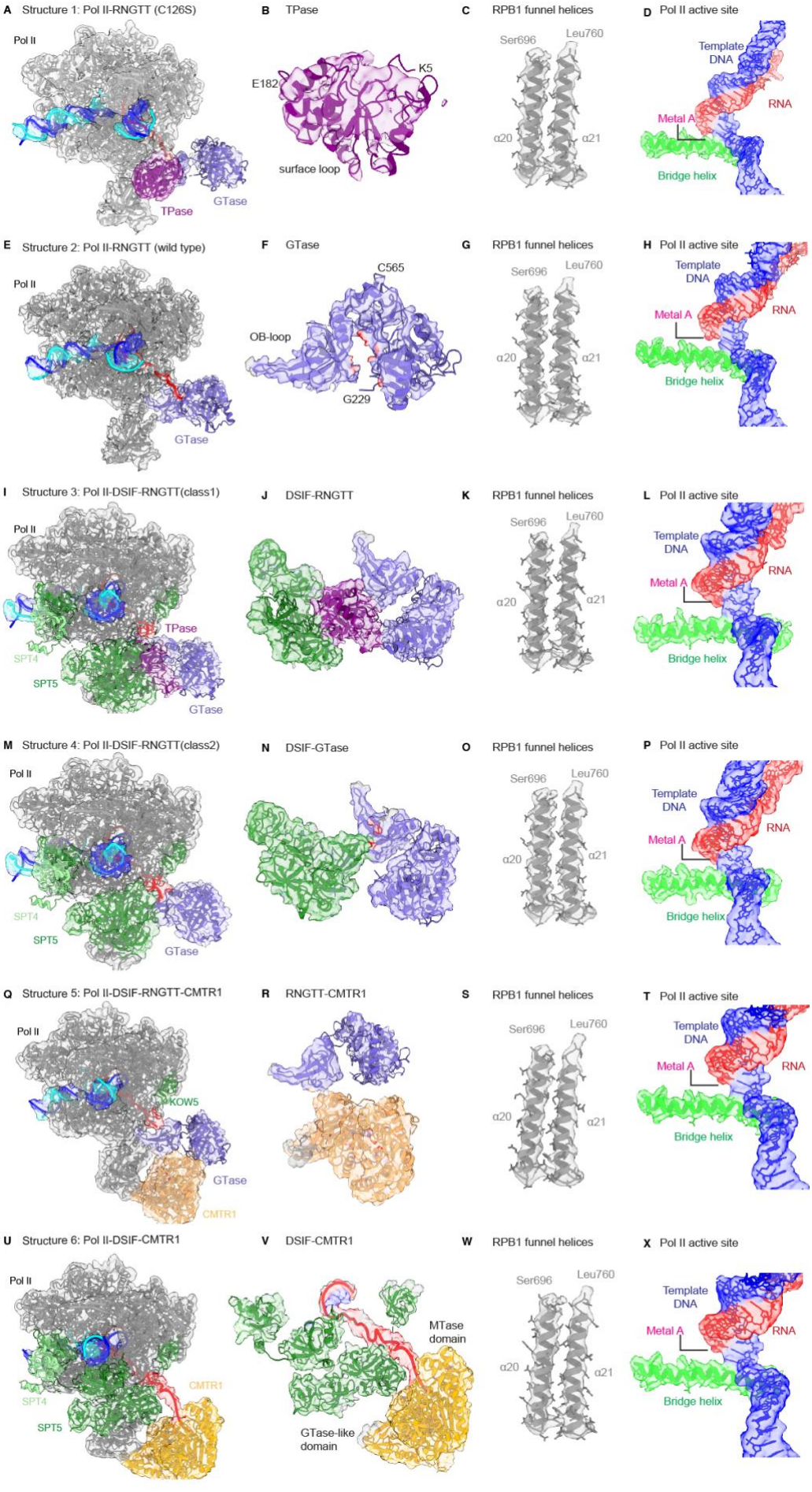

A Structure 1: Pol II-RNGTT (C126S)
Pol II
TPase
GTase
B TPase
E182
K5
surface loop
C RPB1 funnel helices
Ser696
Leu760
α20
α21
D Pol II active site
Template DNA
RNA
Metal A
Bridge helix
E Structure 2: Pol II-RNGTT (wild type)
Pol II
GTase
F GTase
C565
OB-loop
G229
G RPB1 funnel helices
Ser696
Leu760
α20
α21
H Pol II active site
Template DNA
RNA
Metal A
Bridge helix
I Structure 3: Pol II-DSIF-RNGTT(class1)
Pol II
TPase
SPT4
SPT5
GTase
J DSIF-RNGTT
K RPB1 funnel helices
Ser696
Leu760
α20
α21
L Pol II active site
Template DNA
RNA
Metal A
Bridge helix
M Structure 4: Pol II-DSIF-RNGTT(class2)
Pol II
SPT4
SPT5
GTase
N DSIF-GTase
O RPB1 funnel helices
Ser696
Leu760
α20
α21
P Pol II active site
Template DNA
RNA
Metal A
Bridge helix
Q Structure 5: Pol II-DSIF-RNGTT-CMTR1
Pol II
KOW5
GTase
CMTR1
R RNGTT-CMTR1
S RPB1 funnel helices
Ser696
Leu760
α20
α21
T Pol II active site
Template DNA
RNA
Metal A
Bridge helix
U Structure 6: Pol II-DSIF-CMTR1
Pol II
CMTR1
SPT4
SPT5
V DSIF-CMTR1
MTase domain
GTase-like domain
W RPB1 funnel helices
Ser696
Leu760
α20
α21
X Pol II active site
Template DNA
RNA
Metal A
Bridge helix

Figure S8: Cryo-EM densities of complexes. Related to figure 1 to 6.

(A) Pol II-RNGTT (C126S) structure1 with corresponding cryo-EM map. Cryo-EM map is shown in grey.

(B) TPase domain structure with corresponding cryo-EM map. Cryo-EM map is shown in purple.

(C) RPB1 funnel helices with corresponding density from Pol II-RNGTT (C126S) structure.

(D) Active site of Pol II-RNGTT (C126S) structure with corresponding density. Metal A is shown as a pink sphere.

(E) Pol II-RNGTT (WT) structure2 with corresponding cryo-EM map. Cryo-EM map is shown in grey.

(F) GTase domain structure with corresponding cryo-EM map. Cryo-EM map is shown in slate blue.

(G) RPB1 funnel helices with corresponding density from Pol II-RNGTT (WT) structure.

(H) Active site of Pol II-RNGTT (WT) structure with corresponding density. Metal A is shown as a pink sphere.

(I) Pol II-DSIF-RNGTT structure3 with corresponding cryo-EM map. Cryo-EM map is shown in grey.

(J) SPT5-RNGTT structure with corresponding cryo-EM map. Cryo-EM map is shown in green (SPT5), puple (TPase domain) and slate blue (GTase domain).

(K) RPB1 funnel helices with corresponding density from Pol II-DSIF-RNGTT structure.

(L) Active site of Pol II-DSIF-RNGTT structure with corresponding density. Metal A is shown as a pink sphere.

(M) Pol II-DSIF-RNGTT structure4 with corresponding cryo-EM map. Cryo-EM map is shown in grey.

(N) SPT5-RNGTT structure with corresponding cryo-EM map. Cryo-EM map is shown in green (SPT5), and slate blue (GTase domain).

(O) RPB1 funnel helices with corresponding density from Pol II-DSIF-RNGTT structure4.

(P) Active site of Pol II-DSIF-RNGTT structure4 with corresponding density. Metal A is shown as a pink sphere.

(Q) Pol II-DSIF-RNGTT-CMTR1 structure5 with corresponding cryo-EM map. Cryo-EM map is shown in grey.

(R) RNGTT-CMTR1 structure with corresponding cryo-EM map. Cryo-EM map is shown in slate blue (GTase domain) and sandy brown (CMTR1).

(S) RPB1 funnel helices with corresponding density from Pol II-DSIF-RNGTT-CMTR1 structure5.

(T) Active site of Pol II-DSIF-RNGTT-CMTR1 structure5 with corresponding density. Metal A is shown as a pink sphere.

(U) Pol II-DSIF-RNGTT-CMTR1 structure6 with corresponding cryo-EM map. Cryo-EM map is shown in grey.

(V) DSIF-CMTR1-RNA structure with corresponding cryo-EM map. Cryo-EM map is shown in green (SPT5), red (RNA) and orange (CMTR1).

(W) RPB1 funnel helices with corresponding density from Pol II-DSIF-RNGTT-CMTR1 structure6.

(X) Active site of Pol II-DSIF-RNGTT-CMTR1 structure6 with corresponding density. Metal A is shown as a pink sphere.

Table S1: Cryo-EM data collection, refinement and validation statistics.

	Pol II-RNGTT (C126S)	Pol II-RNGTT (WT)
Data collection and processing		
Magnification	81,000	81,000
Voltage (kV)	300	300
Electron exposure (e^{-}/Å^2)	40-43	40-43
Defocus range (μm)	0.25-2.0	0.25-2.0
Pixel size (Å)	1.05	1.05
Symmetry imposed	C1	C1
Initial particle images (no.)	3,049,188	3,261,338
Final particle images (no.)	34,647	21,150
Map resolution (Å)	3.6	3.2
FSC threshold	0.143	0.143
Map resolution range (Å)	3.1-9.0	2.9-6.7
Refinement		
Initial models used (PDB code)	6GML, 2C46, 3S24	6GML, 3S24
Map sharpening *B* factor (Å^2)	-20	-20
Model composition		
Non-hydrogen atoms	36,831	35,383
Protein residues	4376	4219
Nucleotides	85	81
Ligands	9	9
B factors (Å^2)		
Protein	111	49.75
Nucleotide	147.85	73.64
Ligand	149.71	65.66
R.m.s. deviations		
Bond lengths (Å)	0.003	0.004
Bond angles (°)	0.656	0.911
Validation		
MolProbity score	1.94	2.04
Clashscore	12.79	14.82
Poor rotamers (%)	0.00	0.03
Ramachandran plot		
Favored (%)	95.41	94.77
Allowed (%)	4.54	5.18
Disallowed (%)	0.05	0.05

Table S2: Cryo-EM data collection, refinement and validation statistics.

	Pol II-DSIF-RNGTT (TPase bound)	Pol II-DSIF-RNGTT
Data collection and processing		
Magnification	81,000	81,000
Voltage (kV)	300	300
Electron exposure (e^{-}/Å^2)	40-43	40-43
Defocus range (μm)	0.25-2.0	0.25-2.0
Pixel size (Å)	1.05	1.05
Symmetry imposed	C1	C1
Initial particle images (no.)	3,299,682	3,299,682
Final particle images (no.)	64,321	73,948
Map resolution (Å)	3.8	3.6
FSC threshold	0.143	0.143
Map resolution range (Å)	2.5-8.5	2.5-8.5
Refinement		
Initial models used (PDB code)	6GML, 2C46, 3S24	6GML, 2C46, 3S24
Map sharpening *B* factor (Å^2)	-20	-20
Model composition		
Non-hydrogen atoms	41,436	40,090
Protein residues	4953	4782
Nucleotides	90	92
Ligands	9	9
B factors (Å^2)		
Protein	105.27	94.32
Nucleotide	146.18	144.53
Ligand	149	68.20
R.m.s. deviations		
Bond lengths (Å)	0.004	0.003
Bond angles (°)	0.646	0.640
Validation		
MolProbity score	2.01	2.01
Clashscore	13.41	13.95
Poor rotamers (%)	0.00	0.00
Ramachandran plot		
Favored (%)	94.51	94.82
Allowed (%)	5.45	5.13
Disallowed (%)	0.04	0.04

Table S3: Cryo-EM data collection, refinement and validation statistics.

	Pol II-DSIF-CMTR1
Data collection and processing	
Magnification	81,000
Voltage (kV)	300
Electron exposure (e^{-}/Å^2)	40-43
Defocus range (μm)	0.25-4
Pixel size (Å)	1.05
Symmetry imposed	C1
Initial particle images (no.)	3,299,084
Final particle images (no.)	31,427
Map resolution (Å)	4.0
FSC threshold	0.143
Map resolution range (Å)	2.9-10.0
Refinement	
Initial models used (PDB code)	6GML, AlphaFold
Map sharpening *B* factor (Å^2)	-20
Model composition	
Non-hydrogen atoms	50,879
Protein residues	6304
Nucleotides	114
Ligands	10
B factors (Å^2)	
Protein	82.83
Nucleotide	147.71
Ligand	131.85
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.652
Validation	
MolProbity score	1.91
Clashscore	13.76
Poor rotamers (%)	0.00
Ramachandran plot	
Favored (%)	96.15
Allowed (%)	3.79
Disallowed (%)	0.06

REFERENCES

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. Sect. D, Struct. Biol. *74*, 531–544.

Armache, K., Mitterweger, S., Meinhart, A., and Cramer, P. (2005). Structures of Complete RNA Polymerase II and Its Subcomplex , Rpb4 / 7 *. J. Biol. Chem. *280*, 7131–7134.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat. Publ. Gr.

Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J.M., and Cramer, P. (2016). Structure of transcribing mammalian RNA polymerase II. Nature.

Changela, A., Ho, C.K., Martins, A., Shuman, S., and Mondragón, A. (2001). Structure and mechanism of the RNA triphosphatase component of mammalian mRNA capping enzyme. EMBO J. *20*, 2575–2586.

Chen, Z.-L., Meng, J.-M., Cao, Y., Yin, J.-L., Fang, R.-Q., Fan, S.-B., Liu, C., Zeng, W.-F., Ding, Y.-H., Tan, D., et al. (2019). A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. Nat. Commun. *10*, 3404.

Cheng, B., and Price, D.H. (2008). Analysis of factor interactions with RNA polymerase II elongation complexes using a new electrophoretic mobility shift assay. Nucleic Acids Res. *36*, e135.

Chu, C., Das, K., Tyminski, J.R., Bauman, J.D., Guan, R., Qiu, W., Montelione, G.T., Arnold, E., and Shatkin, A.J. (2011). Structure of the guanylyltransferase domain of human mRNA capping enzyme. Proc. Natl. Acad. Sci. U. S. A. *108*, 10104–10108.

Combe, C.W., Fischer, L., and Rappsilber, J. (2015). xiNET: cross-link network maps with residue resolution. Mol. Cell. Proteomics *14*, 1137–1147.

Cowling, V.H. (2010). Regulation of mRNA cap methylation. *302*, 295–302.

Darnellt, J.E. (1975). Blocked 5' Termini. *72*, 1904–1908.

Decroly, E., Ferron, F., Lescar, J., and Canard, B. (2012). Conventional and unconventional mechanisms for capping viral mRNA. Nat. Publ. Gr. *10*.

Devarkar, S.C., Wang, C., Miller, M.T., Ramanathan, A., Jiang, F., and Khan, A.G. (2016). discrimination in capped RNAs by the innate immune receptor RIG-I.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D. Biol. Crystallogr. *60*, 2126–2132.

Fianu, I., Dienemann, C., Aibara, S., and Cramer, P. Cryo-EM structure of mammalian RNA polymerase. Commun. Biol. 1–7.

Furuichi, Y., LaFiandra, A., and Shatkin, A.J. (1977). 5′-Terminal structure and mRNA stability. Nature.

Gonzalez, M.N., Sato, S., Tomomori-Sato, C., Conaway, J.W., and Conaway, R.C. (2018). CTD-dependent and -independent mechanisms govern co-transcriptional capping of Pol II transcripts. BioRxiv 269977.

Gradia, S.D., Ishida, J.P., Tsai, M.-S., Jeans, C., Tainer, J.A., and Fuss, J.O. (2017). Chapter One - MacroBac: New Technologies for Robust and Efficient Large-Scale Production of Recombinant Multiprotein Complexes. In DNA Repair Enzymes: Structure, Biophysics, and Mechanism, B.F.B.T.-M. in E. Eichman, ed. (Academic Press), pp. 1–26.

Grimm, C., Hillen, H.S., Bedenk, K., Szalay, A.A., Grimm, C., Hillen, H.S., Bedenk, K., Bartuli,

J., Neyer, S., Zhang, Q., et al. (2019). Structural Basis of Poxvirus Transcription : Vaccinia Article Structural Basis of Poxvirus Transcription : Vaccinia RNA Polymerase Complexes. 1537–1550.

Hâkansson, K., and Wigley, D.B. (1998). Structure of a complex between a cap analogue and mRNA guanylyl transferase demonstrates the structural chemistry of RNA capping. Proc. Natl. Acad. Sci. U. S. A. *95*, 1505–1510.

Håkansson, K., Doherty, A.J., Shuman, S., and Wigley, D.B. (1997). X-Ray Crystallography Reveals a Large Conformational Change during Guanyl Transfer by mRNA Capping Enzymes. Cell *89*, 545–553.

Kachaev, Z.M., Lebedeva, L.A., and Kozlov, E.N. (2020). Interplay of mRNA capping and transcription machineries. *0*, 1–10.

Konarska, M.M., Padgett, R.A., and Sharp, P.A. (1984). Recognition of Cap Structure In Vitro of mRNA Precursors. *38*, 731–736.

Kosinski, J., von Appen, A., Ori, A., Karius, K., Müller, C.W., and Beck, M. (2015). Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. J. Struct. Biol. *189*, 177–183.

Mandal, S.S., Chu, C., Wada, T., Handa, H., Shatkin, A.J., and Reinberg, D. (2004). Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *101*.

Martinez-Rucobo, F.W., Kohler, R., van de Waterbeemd, M., Heck, A.J.R., Hemann, M., Herzog, F., Stark, H., and Cramer, P. (2015). Molecular Basis of Transcription-Coupled Pre-mRNA Capping. Mol. Cell.

Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. J. Struct. Biol. *152*, 36–51.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D.L. (1997). 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev. *11*, 3306–3318.

Moore, S.D., and Prevelige, P.E.J. (2002). A P22 scaffold protein mutation increases the robustness of head assembly in the presence of excess portal protein. J. Virol. *76*, 10245–10255.

Moteki, S., and Price, D. (2002). Functional coupling of capping and transcription of mRNA. Mol. Cell.

Neugebauer, K.M. (2002). On the importance of being co-transcriptional.

Nilson, K.A., Guo, J., Turek, M.E., Brogie, J.E., Delaney, E., Luse, D.S., and Price, D.H. (2015). THZ1 Reveals Roles for Cdk7 in Co-transcriptional Capping and Pausing. Mol. Cell.

Palangat, M., Renner, D.B., Price, D.H., and Landick, R. (2005). A negative elongation factor for human RNA polymerase II inhibits the anti-arrest transcript-cleavage factor TFIIS. Proc. Natl. Acad. Sci. U. S. A. *102*, 15036–15041.

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.

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.

Qin, P.Z., and Pyle, A.M. (1999). Site-Specific Labeling of RNA with Fluorophores and Other Structural Probes. *70*, 60–70.

Rasmussen, E.B., and Lis, J.T. (1993). In vivo transcriptional pausing and cap formation on three

Drosophila heat shock genes. *90*, 7923–7927.
Rinaldi, A.J., Suddala, K.C., and Walter, N.G. (2015). Chapter 6.
Rodriguez, C.R., Takagi, T., Cho, E.J., and Buratowski, S. (1999). A saccharomyces cerevisiae RNA 5'-triphosphatase related to mRNA capping enzyme. Nucleic Acids Res.
Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Ballard, A.J., Cowie, A., Romera-paredes, B., Nikolov, S., Jain, R., Adler, J., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*.
Shatkin, A.J. (1976). Capping of Eucaryotic mRNAs Review. *1976*.
Shatkin, A.J., and Manley, J.L. (2000). The ends of the affair : Capping and polyadenylation. *7*.
Smietanski, M., Werner, M., Purta, E., Kaminska, K.H., Stepinski, J., Darzynkiewicz, E., Nowotny, M., and Bujnicki, J.M. (2014). structure formation. Nat. Commun. *5*, 1–10.
Sripati, C.E., Groner, Y., and Warner, J.R. (1976). Blocked 5 ' Termini of Yeast mRNA *. J. Biol. Chem. *251*, 2898–2904.
Stepinski, J., Darzynkiewicz, E., and Pelletier, J. (2010). Characterization of hMTr1 , a Human Cap1 2 ˜- O -Ribose. *285*, 33037–33044.
Tegunov, D., and Cramer, P. (2019). Real-time cryo-electron microscopy data preprocessing with Warp. Nat. Methods *16*, 1146–1152.
Tomasz, J.E.N., and Shatkin, A.J. (1976). Mechanism of Formation Blocked and Methylated of Reovirus mRNA V-terminal Sequence , m7GpppGmpC. J. Biol. Chem. *251*, 5043–5053.
Varshney, D., Lombardi, O., Schweikert, G., Dunn, S., Suska, O., and Cowling, V.H. (2018). mRNA Cap Methyltransferase, RNMT-RAM, Promotes RNA Pol II-Dependent Transcription. Cell Rep. *23*, 1530–1542.
Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018). Structure of paused transcription complex Pol II–DSIF–NELF. Nature 1.
Wen, Y., and Shatkin, A.J. (1999). Transcription elongation factor hSPT5 stimulates mRNA capping. Genes Dev.
Wen, Y., Yue, Z., Shatkin, A.J., Handa, H., Shatkin, A.J., and Reinberg, D. (1998). Mammalian capping enzyme binds RNA and uses protein tyrosine phosphatase mechanism. Proc. Natl. Acad. Sci. U. S. A. *95*, 12226–12231.
Williams, G.D., Gokhale, N.S., and Snider, D.L. (2020). crossm.
Yang, F., Shen, Y., Camp 2nd, D.G., and Smith, R.D. (2012). High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. Expert Rev. Proteomics *9*, 129–134.
Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres, S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife *7*.
Züst, R., Cervantes-barragan, L., Habjan, M., Maier, R., Neuman, B.W., Ziebuhr, J., Szretter, K.J., Baker, S.C., Barchet, W., Diamond, M.S., et al. (2011). Ribose 2 ′ - O -methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. Nat. Publ. Gr. *12*, 137–144.

4 Discussion and outlook

While transcription has been well-studied both structurally and biochemically, we lack the understanding of certain aspects which are fundamental to the process. This dissertation focuses on two essential aspects of transcription elongation, backtracking and capping. In the absence of either of these processes, Pol II would cease to elongate the transcript and terminate prematurely[173].

Section 4.1 answers the key findings from my colleagues and I (manuscript 1) regarding Pol II passage through the nucleosome. It then focuses on discussing Pol II-TFIIS structure (my contribution) and future perspectives.

Section 4.2 covers the primary focus of this dissertation, which is structural dissection of the mammalian co-transcriptional capping process. Based on our findings, we first briefly describe the answers to few key open questions in the field. This is followed by elaborating on the proposed model which is based on the individual structures resolved. And finally, I emphasize on the relevance of our results comparing it to the existing literature.

4.1 Pol II EC-TFIIS: novel findings

Through our manuscript and existing literature, we propose a three-step model of Pol II passage through the nucleosome with the help of TFIIS. First, the activated elongation complex (EC*) transcribes into the nucleosome, and the DNA unravelling potentially leads to H2A-H2B dimer loss. Second, as the EC* transcribes further, it stalls and is prone to backtracking. Third, TFIIS rescues the backtracked Pol II by stimulating RNA cleavage and generating a new 3’ end in the active site.

Therefore, solving the structure of Pol II EC-TFIIS was important to understand not only the molecular basis of TFIIS function in humans, but also to differentiate the changes introduced by EC* from those by TFIIS association. Our structure of the Pol II EC-TFIIS revealed three key things. First, the binding interface between Pol II and TFIIS remains the same as that observed in yeast systems. Second, the mammalian Pol II crevice does not need further opening as does the yeast Pol II[84]. Therefore, TFIIS domain II can span the crevice with domain III reaching the Pol II active site with the Pol II subunits showing minimal movement. And third, we note the mobile domain III and Pol II trigger loop, in comparison to a stable binding in the

yeast system[28]. In our manuscript, we therefore speculate that this mobility is associated to regulation of TFIIS entry into the active site.

4.2 Co-transcriptional capping: novel findings

Through this dissertation, we provide a comprehensive structural understanding of the mammalian co-transcriptional machinery. By reconstituting the human capping machinery on the Pol II surface and resolving the cryo-EM structures of six intermediates, we answer the following key questions.

How does the capping machinery interact with the actively transcribing Pol II? We observed direct interactions of the CEs both with Pol II and the RNA. In contrast to the yeast system, mammalian Pol II stalk seems to serve as the primary interface for CE interactions. For RNGTT, TPase domain is transiently stabilized at the stalk likely based on three key factors: TPase domain interaction with RPB2 flab loop; TPase domains active site affinity to the RNA triphosphate end; and GTase domain anchoring to the stalk.

GTase domain is essential for RNGTT to anchor itself to the stalk and position the TPase domain at the Pol II RNA exit channel. In particular, GTase OB-domain loop (which we called the OB-loop) inserts itself in a crevice formed between RPB1 and RPB4/7 interface.

For CMTR1, two primary interfaces are observed. CMTR1 GTase-like domain docks onto RPB7 tip-domain (terminology derived from the yeast system[22]) constituting the only Pol II-CMTR1 interface observed. Additionally, the RNA binds the MTase domain which likely helps in recruitment and stabilization at the RNA exit channel.

How does the 5’ pre-mRNA end move between the multiple active sites? Based on our minimal structures of transcribing Pol II and RNGTT with RNA bound to either the TPase domain or GTase domain active site, we could propose a mechanism for RNA handover. The TPase and the GTase domains of RNGTT are connected by a flexible linker. After the recruitment of RNGTT to the Pol II surface, and its anchoring to the stalk as described above, the TPase domain accepts the triphosphate end of the RNA. It cleaves the γ-phosphate, resulting in a diphosphate end RNA, which serves as a substrate for the GTase domain. The loss of TPase affinity for the diphosphate end coupled to its high affinity of the GTase domain, likely signals the GTase domain for RNA binding. This triggers the NTase domain of the GTase to swing-in, a movement likely allowed due to the flexible linker between the two domains of RNGTT. As a result, GTase domain displaces the transiently stabilized TPase domain and

performs the cap guanylylation. The mRNA handover therefore is a function of affinities and the domain architecture of RNGTT. Upon the completion of RNGTT activity, CMTR1 MTase domain can displace RNGTT and bind the growing nascent cap(-1) or cap(0) RNA to methylate ribose 2'O of the first transcribed nucleotide.

What role do elongation factors play? The structures of actively transcribing Pol II in the presence of DSIF and bound to RNGTT or CMTR1 help shed light on the interplay between early elongation and capping.

RNGTT displaces DSIF from the RNA exit channel causing its drastic rearrangement in comparison to the "canonical" EC conformation[50]. This displacement involves flipping out of the KOWx-KOW4 from the RNA exit channel to the stalk. The TPase domain is therefore now bound in the position where KOWx-KOW4 were bound. Upon the complete of the TPase activity, the TPase domain is displaced by the GTase domain. This displacement of TPase domain could be facilitated by the KOW4 domain, which is observed to be bound next to the GTase. Such a mechanism of action for KOW4 would result in pushing the reaction towards guanylylation and therefore also explain the stimulation of guanylylation in presence of DSIF.

In the presence of CMTR1, DSIF is bound similar to its canonical form. As the RNA is already 29 nt in length, and CMTR1 binds to Pol II stalk, DSIF is bound without any clashes. DSIF stabilizes the emerging pre-mRNA[174] and shows some cross links with CMTR1 C-terminus, which may indicate that CMTR1 activity is facilitated and the conformation stabilized by DSIF.

Finally, taken together, how are these steps of capping coupled to transcription? Co-transcriptional capping of pre-mRNA is more efficient than that of the free RNA[167]. The results described above lead us to propose the first model for the structural basis of mammalian co-transcriptional capping.

We find that as soon as the emerging pre-mRNA end is around 22 nt in length, it can bind the RNGTT. First, RNGTT binds the Pol II stalk and anchors it around the RNA exit channel. Rearrangement of SPT5 domains of DSIF is required to allow this binding. The TPase domain of RNGTT first binds the 5'pppXRNA and cleaves the gamma-phosphate. Then, the TPase domain is displaced by the GTase domain and binds the 5'ppXRNA, which is likely facilitated by SPT5. This is followed by the addition of GMP moiety to the 5' end, resulting in a 5'GpppXRNA. As the pre-mRNA extends to around 29 nt, CMTR1 is able to bind the pre-mRNA. CMTR1 then displaces RNGTT and DSIF moves back into the canonical EC conformation.

CMTR1 MTase domain then methylates the pre-mRNA first nucleotide at 2'O position resulting in 5'GpppXmRNA.

Therefore, the binding affinity of CEs to the pre-mRNA, their direct interactions with the Pol II, and the modular nature of the DSIF together facilitate the efficient transition between steps and coordinate the complex process of co-transcriptional capping on Pol II surface.

4.3 Future perspectives

In this section I first briefly address the missing aspect in our understanding of TFIIS structure. Then for the rest of the section I primarily focus on co-transcriptional capping starting with discussing the limitations of our study and ways to address them. I then move on to discussing the broader implications of the findings for the field, next questions to be answered, and their future applications.

4.3.1 The missing links

TFIIS. The limitations pertaining to Pol II EC-TFIIS are covered in detail in the study limitation section of *manuscript 1*. A missing piece additionally in our understanding of the Pol II EC-TFIIS is the domain III activity. The structure reveals that TFIIS interdomain linker insertion into the crevice does not require additional opening or changes in Pol II subunits as was observed for the yeast system[27]. However, we did not observe a well-resolved domain III due to mobility. We thus refrain from commenting on the activity, and mechanistic differences that might be present in our reconstituted complex in comparison to yeast. Therefore, improving the resolution of domain III in the active site would be the next step in understanding how exactly does the human TFIIS function, and differs from its counterparts in different species. While we tried processing for domain III, one could still collect more data, and focus the processing on domain III using masks and multi-body refinement which may help with its mobile nature. Additionally, while TFIIS is one of the best characterized elongation factors, relatively little is known about its regulation. Exploring its requirement in the cell under various conditions may further the unravel its yet to be understood mechanisms of regulation and roles.

Pre-mRNA capping. A dynamic process such as co-transcriptional capping of the pre-mRNA transcript is particularly challenging to study with cryo-EM. While our structures give

first insights and relative arrangement of different domains, the lack the resolution and experiments limits the comments on the following aspects.

Reconstitution of the minimal structures allows the observation of cap(-1) formation on a urea gel. However, the experimental design relies on addition of an m^7G to observe a band shift on the gel. This limits us from commenting on the TPase activity. Since the TPase binds directly at the RNA exit channel as the pre-mRNA emerges, the TPase could function earlier than the GTase step and may require a shorter pre-mRNA than the 22 nt required for GTase.

Additionally, our minimal structures suggest that CEs can bind to and function in the absence of elongation factors. The TPase and SPT5 domains compete for binding, which is evident with the structural rearrangement of DSIF. This could lead to three possibilities: one, DSIF binds only after the completion of the TPase activity; two,that the TPase domain displaces already bound DSIF; and three, the process is more stochastic than believed and can efficiently occur even in the absence of DSIF. Therefore, while the structures show how these possibilities might work, the precise point in the cycle where DSIF binding is necessary needs further investigation.

Finally, with our in vitro system, I show that although CMTR1 can biochemically bind Pol II EC with a 22 nt RNA, that for a stable complex, a pre-mRNA of around 29 nt is needed. As no density for CMTR1 is observed at least until 26 nt, we determine that the window for CMTR1 binding should be between 27-29 nt. However, the precise pre-mRNA length at which the 2'O methylation can first occur needs further systematic testing. This can be monitored using fluorescent SAM based assays that detect that are routinely used to report MTase activity.

4.3.2 Towards the complete model of co-transcriptional capping

In the recent years, crystal structures of the various enzymes in isolation or in-complex with their substrates have elaborated on the mechanistic details of the CEs. More recently, its critical function in innate immune response, as well its structural basis of this function has come to light[129]. Our structures answer long standing questions and lead us to propose a comprehensive model of co-transcriptional capping. However, it also open new aspects to study.

What is the precise role of DSIF, and when does it bind to Pol II? As elaborated in the previous section, DSIF shows a large rearrangement in the presence of RNGTT. In order to

understand the cause and the consequence, one needs to look carefully at some aspects. Firstly, in order to check at which pre-mRNA length capping can first occur, the smallest pre-mRNA length at which the TPase activity can be achieved needs to be investigated. Secondly, to understand if DSIF is bound at this point two things should be checked for: at which pre-mRNA length does DSIF bind the transcribing Pol II and if it affects TPase activity. Finally, to understand the extent of DSIF effect, quantification of increase in cap guanylylation upon DSIF additions should be done. Taken together, these experiments will help understand the order of events better. However, a possibility to keep in mind is that as long as RNGTT is recruited to Pol II, these events are stochastic and not inter-dependent/sequential.

How important is the CTD in the light of such CTD-independent mechanisms? This leads us to the next aspect which is the effect of Pol II CTD on co-transcriptional capping. Phosphorylation of CTD by TFIIH kinase CDK7 has been shown to be instrumental for recruitment and tethering of CEs to Pol II. However, activation of co-transcriptional capping is only modestly affected in the absence of CTD or TFIIH[162]. Crystal structure of the mouse GTase domain with a CTD heptad (PDB code: 3RTX) revealed interaction sites on the NTase domain of GTase. Our structures show that this potential CTD-interacting surface of GTase remains exposed upon binding to Pol II body. This could mean that the CTD if bound to the exposed surface keeps RNGTT tethered to Pol II. Additionally, CMTR1 has also been shown to interact with phosphorylated CTD with its WW domain[175]. Interestingly, our structures show that the WW domain remains solvent exposed while the GTase-like domain binds the Pol II stalk. Therefore, in consensus with the literature, CMTR1 WW domain likely serves as its tethering point to the Pol II CTD. Altogether, while CTD seems to be important for efficient co-transcriptional capping, there are CTD independent mechanisms that exist and allow for efficient capping. Therefore, the extent of CTD involvement, and phosphorylation requirement for the recruitment and tethering of the CEs to Pol II during co-transcriptional capping need further evaluation. Pol II is still largely endogenously purified in the field and no CTD deleted mutants are available for thorough biochemical testing. Therefore, *in vivo* approaches correlating the global effect of CDK7 and/or CTD loss on the efficiency of the individual steps of capping may help understand the extent of influence of these factors.

When does RNMT act in humans? Are the cap(0) and cap(1) methylations sequential? Can CMTR1 act prior to RNMT in humans? Finally, despite multiple repeated efforts, we could not bind or visualize RNMT/RAM together with Pol II. On gel filtration, it seemed to travel

throughout with Pol II and gave no clear peak. As expected, upon cryo-EM data collection, no density for RNMT was observed. While R.C. Pillutla et al., showed a weak direct interaction between RNMT and an elongating Pol II, we could not reproduce these results[135]. Similar to our observations, two studies by the V.Cowling group have emerged recently. First shows that RNMT/RAM by itself does not directly interact with Pol II or with RNGTT[139]. Second reveals that RNMT shows significant enrichment at a very small number of TSS' in comparison to CMTR1[159]. Considering that most mature mRNAs carry an m^7G cap[133], and a redundant MTase has not been identified, there must be underlying mechanisms that recruit RNMT *in vivo*. The observations suggest that RNMT binding may require yet unidentified additional factors, or that its reaction with the transcription complex is transient and RNA affinity based. In order to decipher the mechanism behind RNMT recruitment and activity, a systemic approach could involve first diving deep into the biochemical aspects. Defining the minimal pre-mRNA length needed in an actively transcribing complex to recapitulate activity and identifying the factors that might additionally be involved in this step of capping will provide a basis for *in vitro* complex assembly. Finally, we still need to answer if there a sequence to the MTase reactions *in vivo*? In vitro, RNMT can act on both cap(-1) and cap(1) RNAs similarly[136]. Additionally, CMTR1 can also methylate both cap(-1) and cap(0) with similar affinities[136]. If yes, what determines the sequence of these reactions? Considering that we observe CMTR1 readily bind to the EC with both cap(-1) and cap(0) RNA, it might be that CMTR1 functions before RNMT. Deriving on the various capping mechanisms reviewed by E.Decroly and colleagues in their review[141]: on one hand, in negative-sense RNA viruses like the *Rhabdoviridae,* 2'O MTase acts prior to the N7 MTase; on the other hand, in ss(+) RNA viruses like the flavivirus or the coronavirus, the order of the two methylations depends on a combination of kinetics, affinity and RNA secondary structure. As a diverse range of systems exist in nature, identifying the order of events in humans, if any may not be trivial but important to understand. Upon the identification of factors and RNA length required, the complexes can then be assembled *in vitro* for structural studies as described in our manuscripts to capture this missing step of capping.

In summary, the improvement in resolution of our structures might help understanding the precise protein-protein and protein-RNA interactions involved. Additionally, one needs to further investigate the above-mentioned aspects of capping. Altogether, the answers to these

will fill the gaps in the field and lead us to have a complete and robust model for the fundamental process of co-transcriptional capping.

4.3.3 Understanding the uncharacterized caps

The RNA 5' end modifications are diverse, species-specific, and still being discovered. In addition to the fundamental cap(1) modification observed in metazoans, there are other uncharacterized caps that have recently come to light.

In mammals, if the first transcribed nucleotide is an adenosine, then it can be methylated at the N6 position by phosphorylated CTD interacting factor 1 (PCIF1)[176]. PCIF1 was revealed as the cap-specific N6-adenosine methyltransferase (CAPAM) only recently[134]. It was first identified as a 704 aa protein that interacts with the Ser5 phosphorylated Pol II CTD with its WW domain[176]. PCIF1 also interacts with Pol II suggesting that the modification is added co-transcriptionally[177]. While the biological significance of m^6Am is a subject of intense study, it has thus far been implicated in modulating mRNA translation, regulating gene expression[134], and in oncogenic functions such as gastric cancer progression[178]. However, no structural information on the PCIF1 is yet available.

Another modification is the O-2 methylation of the second transcribed nucleotide, which is present on half of the mRNA transcripts[122]. The resulting modification is called cap(2) and annotated as $m^7GpppXm_{o\text{-}2}Xm_{o\text{-}2}RNA$[107]. The U snRNAs all carry the cap(2) modification which required for efficient pre-mRNA splicing[179]. CMTR2, the 770 aa protein responsible for this modification was also recently discovered[180]. A homology model of CMTR2 proposes two domains, an RFM domain and an MTase-like domain[157]. While both the domains are required for cap methylation, only the RFM domain carries the activity[180].

In addition to these modifications, there exists the 5'-gamma-methyl phosphate cap[181], the dimethyl monophosphate cap[182], and a recent study that identified various other novel caps[183]. Structural and biochemical studies of such novel modifications will help unravel the cap landscape and further the understanding of its various biological roles, as well as potential targets for disease treatment.

4.3.4 Concluding remarks

Through this dissertation I provide structural basis of two essential aspects of human transcription elongation: i) structure of the human TFIIS with Pol II and ii) the first structural

model for co-transcriptional pre-mRNA capping. In the absence of either of these processes, Pol II would cease to elongate the transcript and terminate transcription prematurely[184].

Understanding the passage of Pol II through the nucleosome has remained as one of the most fundamental questions in the field. TFIIS is instrumental in Pol II traversing through the gene smoothly. Through the work here, we first portray the critical role of TFIIS during elongation when Pol II backtracks. The Pol II EC-TFIIS structure sheds light on important structural differences between the human and the yeast system. The next step is to understand in-depth its function and regulation. Recently, TFIIS was shown to be essential for viability during heat stress adaptation in plants[185]. Unraveling biological functions such as these highlight that there is much to discover still.

Through our work on pre-mRNA capping, we look at another essential cellular process. RNA modifications represent the most fundamental layer in the biology of a cell, and therefore their loss often has drastic effects on cell viability or is lethal[126]. Although this may limit identification of disease correlation, there are increasing number of associations between RNA modifications and disease onset/progression being uncovered. For instance, CMTR1 overexpression has been implicated in a number of diseases, and especially in asthma and cancer[124]. Understanding RNA modifications have also been instrumental in developing mRNA vaccines that do not generate an immune response in human cells. For instance, the mRNA vaccine against SARS-CoV-2 carries a 2'O methylation of the first nucleotide[186]. In addition, the viral 2'O methylation step has been suggested as a potential promising target for inhibition or inactivation to tackle viruses like the coronavirus[187]. While, the underlying molecular mechanisms are not well understood, owing to the large mechanistic diversity, RNA capping might be an attractive antiviral target. Structural studies such as this therefore provide a strong basis to understand the possible mechanisms for disease ramifications, and design specific therapeutic strategies.

Nowadays, a wide selection of structural biology tools is available to study dynamic assemblies such as the co-transcriptional capping complexes. For instance, cryoDRGN, a recently introduced software that reconstructs 3D density maps from heterogenous datasets, and therefore helps map the continuous conformational changes could be particularly[188]. Such tools can serve as a great asset in understanding the continuous movements required to accomplish a process like co-transcriptional capping. Structural biology remains at the forefront of understanding the molecular basis of large machineries involved in the

transcription cycle. However, complementary findings from other techniques are becoming increasingly important for research progress and an in-depth understanding of the field. For instance, a combinatorial approach that uses advances in predictive biology tools like AlphaFold[189], techniques like cryo-electron tomography, and the data from *in vivo* genome-wide studies is required. Such collaborative studies will not only further the single-molecule level understanding, but also shed light on the on the native cellular context of the fundamental process of transcription to achieve the entire picture.

References

1. CRICK, F. Central Dogma of Molecular Biology. *Nature* **227**, 561–563 (1970).
2. Orphanides, G. & Reinberg, D. A Unified Theory of Gene Expression. *Cell* **108**, 439–451 (2002).
3. Cramer, P. *et al.* Architecture of RNA Polymerase II and Implications for the Transcription Mechanism. *Science (80-.).* **288**, 640–649 (2000).
4. BESSMAN, M. J., KORNBERG, A., LEHMAN, I. R. & SIMMS, E. S. Enzymic synthesis of deoxyribonucleic acid. *Biochim. Biophys. Acta* **21**, 197–198 (1956).
5. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. *Science (80-.).* **292**, 1863 LP – 1876 (2001).
6. Mukherjee, K. & Chatterji, D. Studies on the ω Subunit of Escherichia Coli RNA Polymerase. *Eur. J. Biochem.* **247**, 884–889 (1997).
7. Minakhin, L. *et al.* Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 892–897 (2001).
8. Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179 (2015).
9. Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* **9**, 85–98 (2011).
10. Vannini, A. & Cramer, P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol. Cell* **45**, 439–446 (2012).
11. Lindell, T. J., Weinberg, F., Morris, P. W., Roeder, R. G. & Rutter, W. J. Specific Inhibition of Nuclear RNA Polymerase II by α-Amanitin. *Science (80-.).* **170**, 447–449 (1970).
12. Schedl, P., Roberts, J. & Primakoff, P. In vitro processing of E. coli tRNA precursors. *Cell* **8**, 581–594 (1976).
13. Reinberg, D. & Roeder, R. G. Factors Involved in Specific Transcription by Mammalian RNA Polymerase I1 TRANSCRIPTION FACTOR IIS STIMULATES ELONGATION OF. **262**, 3331–3337 (1987).
14. Kujirai, T. *et al.* Structural basis of the nucleosome transition during RNA polymerase II passage. *Science (80-.).* **362**, 595–598 (2018).
15. Farnung, L., Ochmann, M., Engeholm, M. & Cramer, P. Structural basis of nucleosome transcription mediated by Chd1 and FACT. *Nat. Struct. Mol. Biol.* **28**, 382–387 (2021).
16. Woychik, N. A. Fractions to functions: RNA polymerase II thirty years later. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 311–317 (1998).
17. Svejstrup, J. Q. The RNA polymerase II transcription cycle: cycling through chromatin. *Biochim. Biophys. Acta - Gene Struct. Expr.* **1677**, 64–73 (2004).
18. Conaway, R. C. & Conaway, J. W. General initiation factors for RNA polymerase II. *Annu. Rev. Biochem.* **62**, 161–190 (1993).
19. Shandilya, J. & Roberts, S. G. E. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim. Biophys. Acta* **1819**, 391–400 (2012).
20. Bentley, D. The mRNA assembly line: transcription and processing machines in the same factory. *Curr. Opin. Cell Biol.* **14**, 336–342 (2002).
21. Fu, J. *et al.* Yeast RNA Polymerase II at 5 Å Resolution. *Cell* **98**, 799–810 (1999).
22. Armache, K., Mitterweger, S., Meinhart, A. & Cramer, P. Structures of Complete RNA Polymerase II and Its Subcomplex , Rpb4 / 7 *. *J. Biol. Chem.* **280**, 7131–7134 (2005).
23. Ho, C. K. & Shuman, S. Distinct Roles for CTD Ser-2 and Ser-5 Phosphorylation in the Recruitment and Allosteric Activation of Mammalian mRNA Capping Enzyme. *Mol.*

Cell **3**, 405–411 (1999).
24. McCracken, S. *et al.* 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev.* **11**, 3306–3318 (1997).
25. Eick, D. & Geyer, M. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem. Rev.* **113**, 8456–8490 (2013).
26. Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Mol. Cell* **36**, 541–546 (2009).
27. Cheung, A. C. M. & Cramer, P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471**, 249–253 (2011).
28. Kettenberger, H., Armache, K. & Cramer, P. Structure and Its Interactions with NTP and TFIIS. **16**, 955–965 (2004).
29. Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H. & Artsimovitch, I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448**, 157–162 (2007).
30. Kireeva, M. L. *et al.* Transient Reversal of RNA Polymerase II Active Site Closing Controls Fidelity of Transcription Elongation. *Mol. Cell* **30**, 557–566 (2008).
31. Sosunov, V. *et al.* Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *EMBO J.* **22**, 2234–2244 (2003).
32. Steitz, T. A. Function and Structure. *Class. Cases Neuropsychol.* (2010) doi:10.4324/9780203304112_chapter_i.
33. Kireeva, M., Kashlev, M. & Burton, Z. F. Translocation by multi-subunit RNA polymerases. *Biochim. Biophys. Acta* **1799**, 389–401 (2010).
34. Muhlbacher, W. *et al.* Conserved architecture of the core RNA polymerase II initiation complex. *Nat. Commun.* **5**, 4310 (2014).
35. Schier, A. C. & Taatjes, D. J. Structure and mechanism of the RNA polymerase II transcription machinery. *Genes Dev.* **34**, 465–488 (2020).
36. Finkelstein, A. *et al.* A cDNA encoding RAP74, a general initiation factor for transcription by RNA polymerase II. *Nature* **355**, 464–467 (1992).
37. Seizl, M. *et al.* A Conserved GA Element in TATA-Less RNA Polymerase II Promoters. *PLoS One* **6**, e27595 (2011).
38. Kadonaga, J. T. Perspectives on the RNA polymerase II core promoter. *WIREs Dev. Biol.* **1**, 40–51 (2012).
39. Schilbach, S. *et al.* Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* (2017) doi:10.1038/nature24282.
40. Abdella, R. *et al.* Structure of the human Mediator-bound transcription preinitiation complex. *Science* **372**, 52–56 (2021).
41. Dienemann, C., Schwalb, B., Schilbach, S. & Cramer, P. Promoter Distortion and Opening in the RNA Polymerase II Cleft. *Mol. Cell* **73**, 97-106.e4 (2019).
42. Louder, R. K. *et al.* Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* **531**, 604–609 (2016).
43. Wang, S. P. *et al.* Phylogeny of mRNA capping enzymes. *Proc. Natl. Acad. Sci.* **94**, 9573–9578 (1997).
44. Jeronimo, C. & Robert, F. Kin28 regulates the transient association of Mediator with core promoters. *Nat. Struct. Mol. Biol.* **21**, 449–455 (2014).
45. Wada, T. *et al.* DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**, 343–356 (1998).
46. Lidschreiber, M., Leike, K. & Cramer, P. Cap completion and C-terminal repeat domain kinase recruitment underlie the initiation-elongation transition of RNA polymerase II. *Mol. Cell. Biol.* **33**, 3805–3816 (2013).

47. NandyMazumdar, M. & Artsimovitch, I. Ubiquitous transcription factors display structural plasticity and diverse functions: NusG proteins - Shifting shapes and paradigms. *Bioessays* **37**, 324–334 (2015).
48. Ponting, C. P. Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res.* **30**, 3643–3652 (2002).
49. Kyrpides, N. C., Woese, C. R. & Ouzounis, C. A. KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem. Sci.* **21**, 425–426 (1996).
50. Tegunov, D., Xu, Y., Bernecky, C., Lee, C. & Maier, K. C. transcription elongation complex. (2017) doi:10.1038/ncomms15741.
51. Hartzog, G. A. & Fu, J. The Spt4-Spt5 complex: a multi-faceted regulator of transcription elongation. *Biochim. Biophys. Acta* **1829**, 105–115 (2013).
52. Qiu, Y. & Gilmour, D. S. Identification of regions in the Spt5 subunit of DRB sensitivity-inducing factor (DSIF) that are involved in promoter-proximal pausing. *J. Biol. Chem.* **292**, 5555–5570 (2017).
53. Missra, A. & Gilmour, D. S. Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 11301–11306 (2010).
54. Hantsche, M. & Cramer, P. The Structural Basis of Transcription: 10 Years After the Nobel Prize in Chemistry. *Angew. Chem. Int. Ed. Engl.* **55**, 15972–15981 (2016).
55. Gilmour, D. S. & Lis, J. T. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. *Mol. Cell. Biol.* **6**, 3984–3989 (1986).
56. Strobl, L. J. & Eick, D. Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-myc in vivo. *EMBO J.* **11**, 3307–3314 (1992).
57. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).
58. Pei, Y. & Shuman, S. Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J. Biol. Chem.* **277**, 19639–19648 (2002).
59. Rahl, P. B. *et al.* c-Myc Regulates Transcriptional Pause Release. *Cell* **141**, 432–445 (2010).
60. Williams, L. H. *et al.* Pausing of RNA Polymerase II Regulates Mammalian Developmental Potential through Control of Signaling Networks. *Mol. Cell* **58**, 311–322 (2015).
61. Yamaguchi, Y. *et al.* NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41–51 (1999).
62. Vos, S. M., Farnung, L., Urlaub, H. & Cramer, P. Structure of paused transcription complex Pol II–DSIF–NELF. *Nature* 1 (2018) doi:10.1038/s41586-018-0442-2.
63. Palangat, M., Renner, D. B., Price, D. H. & Landick, R. A negative elongation factor for human RNA polymerase II inhibits the anti-arrest transcript-cleavage factor TFIIS. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15036–15041 (2005).
64. Meisheng, Z. *et al.* Tat Modifies the Activity of CDK9 To Phosphorylate Serine 5 of the RNA Polymerase II Carboxyl-Terminal Domain during Human Immunodeficiency Virus Type 1 Transcription. *Mol. Cell. Biol.* **20**, 5077–5086 (2000).
65. Fujinaga, K. *et al.* Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol. Cell. Biol.* **24**, 787–795 (2004).
66. Yamada, T. *et al.* P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol. Cell* **21**, 227–237 (2006).
67. Vos, S. M. *et al.* Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature* (2018) doi:10.1038/s41586-018-0440-4.

68. Yu, M. *et al.* RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science* **350**, 1383–1386 (2015).
69. Fianu, I. *et al.* Structural basis of Integrator-mediated transcription regulation. *Science* **374**, 883–887 (2021).
70. Van Oss, S. B., Cucinotta, C. E. & Arndt, K. M. Emerging Insights into the Roles of the Paf1 Complex in Gene Regulation. *Trends Biochem. Sci.* **42**, 788–798 (2017).
71. Vos, S. M., Farnung, L., Linden, A., Urlaub, H. & Cramer, P. Structure of complete Pol II-DSIF-PAF-SPT6 transcription complex reveals RTF1 allosteric activation. *Nat. Struct. Mol. Biol.* **27**, 668–677 (2020).
72. Cao, Q.-F. *et al.* Correction for Cao et al., Characterization of the Human Transcription Elongation Factor Rtf1: Evidence for Nonoverlapping Functions of Rtf1 and the Paf1 Complex. *Mol. Cell. Biol.* **35**, 4093 (2015).
73. Hou, L. *et al.* Paf1C regulates RNA polymerase II progression by modulating elongation rate. *Proc. Natl. Acad. Sci.* **116**, 14583–14592 (2019).
74. Endoh, M. *et al.* Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. *Mol. Cell. Biol.* **24**, 3324–3336 (2004).
75. Yoh, S. M., Cho, H., Pickle, L., Evans, R. M. & Jones, K. A. The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev.* **21**, 160–174 (2007).
76. Ardehali, M. B. *et al.* Spt6 enhances the elongation rate of RNA polymerase II in vivo. *EMBO J.* **28**, 1067–1077 (2009).
77. Proudfoot, N. J., Furger, A. & Dye, M. J. Integrating mRNA Processing with Transcription. *Cell* **108**, 501–512 (2002).
78. Cho, E. J., Kobor, M. S., Kim, M., Greenblatt, J. & Buratowski, S. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev.* **15**, 3319–3329 (2001).
79. Baejen, C. *et al.* Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. *Mol. Cell* **66**, 38-49.e6 (2017).
80. Proudfoot, N. J. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* **352**, aad9926 (2016).
81. Kuehner, J. N., Pearson, E. L. & Moore, C. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat. Rev. Mol. Cell Biol.* **12**, 283–294 (2011).
82. Logan, J., Falck-Pedersen, E., Darnell, J. E. & Shenk, T. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc. Natl. Acad. Sci.* **84**, 8306–8310 (1987).
83. Abdelkareem, M. *et al.* Structural Basis of Transcription: RNA Polymerase Backtracking and Its Reactivation. *Mol. Cell* **75**, 298-309.e4 (2019).
84. Kettenberger, H., Armache, K. & Cramer, P. Architecture of the RNA Polymerase II-TFIIS Complex and Implications for mRNA Cleavage. **114**, 347–357 (2003).
85. Olmsted, V. K. *et al.* Yeast transcript elongation factor (TFIIS), structure and function. I: NMR structural analysis of the minimal transcriptionally active region. *J. Biol. Chem.* **273**, 22589–22594 (1998).
86. Sheridan, R. M., Fong, N., D'Alessandro, A. & Bentley, D. L. Widespread Backtracking by RNA Pol II Is a Major Effector of Gene Activation, 5' Pause Release, Termination, and Transcription Elongation Rate. *Mol. Cell* **73**, 107-118.e4 (2019).
87. Sekimizu, K., Kobayashi, N., Mizuno, D. & Natori, S. Purification of a Factor from Ehrlich Ascites Tumor Cells Specifically Stimulating RNA Polymerase II. *Biochemistry* **15**, 5064–5070 (1976).
88. Morin, P. E., Awrey, D. E., Edwards, A. M. & Arrowsmith, C. H. Elongation Factor TFIIS Contains Three Structural Domains: Solution Structure of Domain II. *Proc. Natl.*

Acad. Sci. U. S. A. **93**, 10604–10608 (1996).
89. Pan, G., Aso, T. & Greenblatt, J. Interaction of Elongation Factors TFIIS and Elongin A with a Human RNA Polymerase II Holoenzyme Capable of Promoter-specific Initiation and Responsive to Transcriptional Activators*. *J. Biol. Chem.* **272**, 24563–24571 (1997).
90. Wind, M. & Reines, D. Transcription elongation factor SII. *Bioessays* **22**, 327–336 (2000).
91. Farnung, L., Ochmann, M., Garg, G., Vos, S. M. & Cramer, P. Structure of a backtracked hexasomal intermediate of nucleosome transcription. *Mol. Cell* 1–9 (2022) doi:10.1016/j.molcel.2022.06.027.
92. Wang, D. *et al.* Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science* **324**, 1203–1206 (2009).
93. Kireeva, M. L. *et al.* Nature of the Nucleosomal Barrier to RNA Polymerase II. *Mol. Cell* **18**, 97–108 (2005).
94. Miller, O. L. J. & Beatty, B. R. Visualization of nucleolar genes. *Science* **164**, 955–957 (1969).
95. LeMaire, M. F. & Thummel, C. S. Splicing precedes polyadenylation during Drosophila E74A transcription. *Mol. Cell. Biol.* **10**, 6059–6063 (1990).
96. Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**, 688–700 (2009).
97. Hsin, J.-P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* **26**, 2119–2137 (2012).
98. Shatkin, A. J. Capping of Eucaryotic mRNAs Review. **1976**, (1976).
99. Cowling, V. H. Regulation of mRNA cap methylation. **302**, 295–302 (2010).
100. Staley, J. P. & Guthrie, C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315–326 (1998).
101. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
102. Matlin, A. J. & Moore, M. J. Spliceosome assembly and composition. *Adv. Exp. Med. Biol.* **623**, 14–35 (2007).
103. Smith, D. J., Query, C. C. & Konarska, M. M. 'Nought may endure but mutability': spliceosome dynamics and the regulation of splicing. *Mol. Cell* **30**, 657–666 (2008).
104. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).
105. Baurén, G. & Wieslander, L. Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell* **76**, 183–192 (1994).
106. Zhang, S. *et al.* Structure of a transcribing RNA polymerase II-U1 snRNP complex. *Science* **371**, 305–309 (2021).
107. Shatkin, A. J. & Manley, J. L. The ends of the affair : Capping and polyadenylation. **7**, (2000).
108. de Almeida, S. F. & Carmo-Fonseca, M. Cotranscriptional RNA checkpoints. *Epigenomics* **2**, 449–455 (2010).
109. Millevoi, S. & Vagner, S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res.* **38**, 2757–2774 (2010).
110. Kim, M. *et al.* The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**, 517–522 (2004).
111. Wang, S. P. *et al.* Phylogeny of mRNA capping enzymes. *Proc. Natl. Acad. Sci.* **94**, 9573–9578 (1997).
112. Varani, G. A cap for all occasions. *Structure* **5**, 855–858 (1997).
113. Suh, M.-H. *et al.* A dual interface determines the recognition of RNA polymerase II by RNA capping enzyme. *J. Biol. Chem.* **285**, 34027–34038 (2010).
114. Beelman, C. A. & Parker, R. Degradation of mRNA in eukaryotes. *Cell* **81**, 179–183

(1995).
115. Jiao, X., Chang, J. H., Kilic, T., Tong, L. & Kiledjian, M. A Mammalian Pre-mRNA 5′ End Capping Quality Control Mechanism and an Unexpected Link of Capping to Pre-mRNA Processing. *Mol. Cell* **50**, 104–115 (2013).
116. Furuichi, Y., LaFiandra, A. & Shatkin, A. J. 5′-Terminal structure and mRNA stability. *Nature* at https://doi.org/10.1038/266235a0 (1977).
117. Inoue, K., Ohno, M., Sakamoto, H. & Shimura, Y. Effect of the cap structure on pre-mRNA splicing in Xenopus oocyte nuclei. *Genes Dev.* **3**, 1472–1479 (1989).
118. Fresco, L. D. & Buratowski, S. Conditional mutants of the yeast mRNA capping enzyme show that the cap enhances, but is not required for, mRNA splicing. *RNA* **2**, 584–596 (1996).
119. Gonatopoulos-Pournatzis, T. & Cowling, V. H. Cap-binding complex (CBC). *Biochem. J.* **457**, 231–242 (2014).
120. Topisirovic, I., Svitkin, Y. V, Sonenberg, N. & Shatkin, A. J. Cap and cap-binding proteins in the control of gene expression. *Wiley Interdiscip. Rev. RNA* **2**, 277–298 (2011).
121. Izaurralde, E. & Mattaj, I. W. RNA export. *Cell* **81**, 153–159 (1995).
122. Furuichi, Y. Discovery of m(7)G-cap in eukaryotic mRNAs. *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.* **91**, 394–409 (2015).
123. Caponigro, G. & Parker, R. Multiple functions for the poly(A)-binding protein in mRNA decapping and deadenylation in yeast. *Genes Dev.* **9**, 2421–2432 (1995).
124. Galloway, A. & Cowling, V. H. mRNA cap regulation in mammalian cell function and fate. *Biochim. Biophys. acta. Gene Regul. Mech.* **1862**, 270–279 (2019).
125. Stepinski, J., Darzynkiewicz, E. & Pelletier, J. Characterization of hMTr1 , a Human Cap1 2 ¯ - O -Ribose. **285**, 33037–33044 (2010).
126. Ramanathan, A., Robb, G. B. & Chan, S.-H. mRNA capping: biological functions and applications. *Nucleic Acids Res.* **44**, 7511–7526 (2016).
127. Devarkar, S. C. *et al.* discrimination in capped RNAs by the innate immune receptor RIG-I. (2016) doi:10.1073/pnas.1515152113.
128. Züst, R. *et al.* Ribose 2 ′ - O -methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. *Nat. Publ. Gr.* **12**, 137–144 (2011).
129. Devarkar, S. C. *et al.* Structural basis for m7G recognition and 2'-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 596–601 (2016).
130. Young, D. F. *et al.* Human IFIT1 Inhibits mRNA Translation of Rubulaviruses but Not Other Members of the Paramyxoviridae Family. *J. Virol.* **90**, 9446–9456 (2016).
131. Haline-Vaz, T., Silva, T. C. L. & Zanchin, N. I. T. The human interferon-regulated ISG95 protein interacts with RNA polymerase II and shows methyltransferase activity. *Biochem. Biophys. Res. Commun.* **372**, 719–724 (2008).
132. Williams, G. D., Gokhale, N. S. & Snider, D. L. crossm. (2020).
133. Rottman, F., Shatkin, A. J. & Perry, R. P. Sequences containing methylated nucleotides at the 5' termini of messenger RNAs: possible implications for processing. *Cell* **3**, 197–199 (1974).
134. Sugita, A. *et al.* The cap-specific m6A methyltransferase, PCIF1/CAPAM, is dynamically recruited to the gene promoter in a transcription-dependent manner. *J. Biochem.* **170**, 203–213 (2021).
135. Pillutla, R. C., Yue, Z., Maldonado, E. & Shatkin, A. J. Recombinant human mRNA cap methyltransferase binds capping enzyme/RNA polymerase IIo complexes. *J. Biol. Chem.* **273**, 21443–21446 (1998).
136. Inesta-vaquera, F. & Cowling, V. H. Regulation and function of CMTR1-dependent

mRNA cap methylation. **8**, 1–6 (2017).
137. Changela, A., Ho, C. K., Martins, A., Shuman, S. & Mondragón, A. Structure and mechanism of the RNA triphosphatase component of mammalian mRNA capping enzyme. *EMBO J.* **20**, 2575–86 (2001).
138. Chu, C. *et al.* Structure of the guanylyltransferase domain of human mRNA capping enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10104–8 (2011).
139. Aregger, M. & Cowling, V. H. Human cap methyltransferase (RNMT) N-terminal non-catalytic domain mediates recruitment to transcription initiation sites. *Biochem. J.* **455**, 67–73 (2013).
140. Bélanger, F., Stepinski, J., Darzynkiewicz, E. & Pelletier, J. Characterization of hMTr1, a human Cap1 2'-O-ribose methyltransferase. *J. Biol. Chem.* **285**, 33037–33044 (2010).
141. Decroly, E., Ferron, F., Lescar, J. & Canard, B. Conventional and unconventional mechanisms for capping viral mRNA. *Nat. Publ. Gr.* **10**, (2012).
142. Ghosh, A. & Lima, C. D. Enzymology of RNA cap synthesis. *Wiley Interdiscip. Rev. RNA* **1**, 152–172 (2010).
143. Moss, B., Gershowitz, A., Wei, C. M. & Boone, R. Formation of the guanylylated and methylated 5'-terminus of vaccinia virus mRNA. *Virology* **72**, 341–351 (1976).
144. Shuman, S. Structure, mechanism, and evolution of the mRNA capping apparatus. *Prog. Nucleic Acid Res. Mol. Biol.* (2001) doi:10.1016/S0079-6603(00)66025-7.
145. Hâkansson, K. & Wigley, D. B. Structure of a complex between a cap analogue and mRNA guanylyl transferase demonstrates the structural chemistry of RNA capping. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 1505–10 (1998).
146. Lima, C. D., Wang, L. K. & Shuman, S. Structure and Mechanism of Yeast RNA Triphosphatase: An Essential Component of the mRNA Capping Apparatus. *Cell* **99**, 533–543 (1999).
147. Benarroch, D., Qiu, Z. R., Schwer, B. & Shuman, S. Characterization of a mimivirus RNA cap guanine-N2 methyltransferase. *RNA* **15**, 666–674 (2009).
148. Wen, Y. *et al.* Mammalian capping enzyme binds RNA and uses protein tyrosine phosphatase mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 12226–31 (1998).
149. Fabrega, C., Shen, V., Shuman, S. & Lima, C. D. Structure of an mRNA Capping Enzyme Bound to the Phosphorylated Carboxy-Terminal Domain of RNA Polymerase II. *Mol. Cell* **11**, 1549–1561 (2003).
150. Håkansson, K., Doherty, A. J., Shuman, S. & Wigley, D. B. X-Ray Crystallography Reveals a Large Conformational Change during Guanyl Transfer by mRNA Capping Enzymes. *Cell* **89**, 545–553 (1997).
151. Gonatopoulos-Pournatzis, T., Dunn, S., Bounds, R. & Cowling, V. H. RAM/Fam103a1 is required for mRNA cap methylation. *Mol. Cell* **44**, 585–596 (2011).
152. Protein, C., Unit, U., Sciences, L., Street, D. & Dd, D. Human cap methyltransferase (RNMT) N-terminal non-catalytic domain mediates recruitment to transcription initiation sites. **73**, 67–73 (2013).
153. Varshney, D. *et al.* mRNA Cap Methyltransferase, RNMT-RAM, Promotes RNA Pol II-Dependent Transcription. *Cell Rep.* **23**, 1530–1542 (2018).
154. Fabrega, C., Hausmann, S., Shen, V., Shuman, S. & Lima, C. D. Structure and Mechanism of mRNA Cap (Guanine-N7) Methyltransferase. *Mol. Cell* **13**, 77–89 (2004).
155. Varshney, D. *et al.* Molecular basis of RNA guanine-7 methyltransferase (RNMT) activation by RAM. *Nucleic Acids Res.* **44**, 10423–10436 (2016).
156. Kyrieleis, O. J. P., Chang, J., de la Peña, M., Shuman, S. & Cusack, S. Crystal structure of vaccinia virus mRNA capping enzyme provides insights into the mechanism and evolution of the capping apparatus. *Structure* **22**, 452–465 (2014).
157. Smietanski, M. *et al.* structure formation. *Nat. Commun.* **5**, 1–10 (2014).

158. Li, C., Xia, Y., Gao, X. & Gershon, P. D. Mechanism of RNA 2'-O-methylation: evidence that the catalytic lysine acts to steer rather than deprotonate the target nucleophile. *Biochemistry* **43**, 5680–5687 (2004).
159. Liang, S. *et al.* CMTR1 is recruited to transcription start sites and promotes ribosomal protein and histone gene expression in embryonic stem cells. 1–18 (2022).
160. Ghosh, A., Shuman, S. & Lima, C. D. Structural Insights to How Mammalian Capping Enzyme Reads the CTD Code. *Mol. Cell* **43**, 299–310 (2011).
161. Fujiwara, R., Damodaren, N., Wilusz, J. E. & Murakami, K. The capping enzyme facilitates promoter escape and assembly of a follow-on preinitiation complex for reinitiation. *Proc. Natl. Acad. Sci.* **116**, 22573–22582 (2019).
162. Gonzalez, M. N. *et al.* CTD-dependent and -independent mechanisms govern co-transcriptional capping of Pol II transcripts. *bioRxiv* (2018) doi:10.1101/269977.
163. Kim, H.-J. *et al.* mRNA capping enzyme activity is coupled to an early transcription elongation. *Mol. Cell. Biol.* **24**, 6184–6193 (2004).
164. Glover-Cutter, K., Kim, S., Espinosa, J. & Bentley, D. L. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat. Struct. Mol. Biol.* **15**, 71–78 (2008).
165. Narita, T. *et al.* NELF Interacts with CBC and Participates in 3′ End Processing of Replication-Dependent Histone mRNAs. *Mol. Cell* (2007) doi:10.1016/j.molcel.2007.04.011.
166. Nilson, K. A. *et al.* THZ1 Reveals Roles for Cdk7 in Co-transcriptional Capping and Pausing. *Mol. Cell* (2015) doi:10.1016/j.molcel.2015.06.032.
167. Moteki, S. & Price, D. Functional coupling of capping and transcription of mRNA. *Mol. Cell* (2002) doi:10.1016/S1097-2765(02)00660-3.
168. Perales, R. & Bentley, D. "Cotranscriptionality": The Transcription Elongation Complex as a Nexus for Nuclear Transactions. *Mol. Cell* **36**, 178–191 (2009).
169. Martinez-Rucobo, F. W. *et al.* Molecular Basis of Transcription-Coupled Pre-mRNA Capping. *Mol. Cell* (2015) doi:10.1016/j.molcel.2015.04.004.
170. Wen, Y. & Shatkin, A. J. Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev.* (1999) doi:10.1101/gad.13.14.1774.
171. Mandal, S. S. *et al.* Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. **101**, (2004).
172. Rasmussen, E. B. & Lis, J. T. In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. **90**, 7923–7927 (1993).
173. Neugebauer, K. M. On the importance of being co-transcriptional. (2002) doi:10.1242/jcs.00073.
174. Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M. & Cramer, P. Structure of transcribing mammalian RNA polymerase II. *Nature* (2016) doi:10.1038/nature16482.
175. Inesta-vaquera, F. *et al.* DHX15 regulates CMTR1-dependent gene expression and cell proliferation. **1**, 1–18 (2018).
176. Fan, H. *et al.* PCIF1, a novel human WW domain-containing protein, interacts with the phosphorylated RNA polymerase II. *Biochem. Biophys. Res. Commun.* **301**, 378–385 (2003).
177. Akichika, S. *et al.* Cap-specific terminal N (6)-methylation of RNA by an RNA polymerase II-associated methyltransferase. *Science* **363**, (2019).
178. Zhuo, W. *et al.* m6Am methyltransferase PCIF1 is essential for aggressiveness of gastric cancer cells by inhibiting TM9SF1 mRNA translation. *Cell Discov.* **8**, 48 (2022).
179. Dönmez, G., Hartmuth, K. & Lührmann, R. Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation. *RNA* **10**, 1925–1933 (2004).
180. Werner, M. *et al.* 2 0 - O -ribose methylation of cap2 in human : function and evolution

in a horizontally mobile family. **39**, 4756–4768 (2011).
181. Cosgrove, M. S., Ding, Y., Rennie, W. A., Lane, M. J. & Hanes, S. D. The Bin3 RNA methyltransferase targets 7SK RNA to control transcription and translation. *WIREs RNA* **3**, 633–647 (2012).
182. Xhemalce, B., Robson, S. C. & Kouzarides, T. Human RNA methyltransferase BCDIN3D regulates microRNA processing. *Cell* **151**, 278–288 (2012).
183. Abdelhamid, R. F. *et al.* Multiplicity of 5' cap structures present on short RNAs. *PLoS One* **9**, e102895 (2014).
184. Noe Gonzalez, M., Blears, D. & Svejstrup, J. Q. Causes and consequences of RNA polymerase II stalling during transcript elongation. *Nat. Rev. Mol. Cell Biol.* **22**, 3–21 (2021).
185. Szádeczky-Kardoss, I. *et al.* Elongation factor TFIIS is essential for heat stress adaptation in plants. *Nucleic Acids Res.* **50**, 1927–1950 (2022).
186. Nance, K. D. & Meier, J. L. Modi fi cations in an Emergency: The Role of N1-Methylpseudouridine in COVID-19 Vaccines. (2021) doi:10.1021/acscentsci.1c00197.
187. Paramasivam, A. RNA 2 ′ -O-methylation modi fi cation and its implication in COVID-19 immunity. 20–22 (2020) doi:10.1038/s41420-020-00358-z.
188. Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **18**, 176–185 (2021).
189. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

Abbreviations and acronyms

EC*	Activated elongation complex
CTD	C-terminal domain
CTR	C-terminal repeat region
CBC	Cap binding complex
CMTR1	Cap methyltransferase 1
CAPAM	Cap-specific N6-adenosine methyltransferase
CEs	Capping enzymes
CstF	Cleavage stimulation factor
CPSF	Cleavage/polyadenylation-specificity factor
PRRs	Cytoplasmic pattern recognition receptors
DXO	Decapping exoribonuclease
DNA	Deoxyribonucleic acid
RNAP	DNA-dependent RNA polymerases
DSIF	DRB sensitivity inducing factor
EEC	Early elongation checkpoint
EC	Elongation complex
GTF's	General transcription factors
GTase	Guanylyltransferase
ISG	IFN stimulated genes
IFIT	IFN-induced proteins with tetratricopeptide repeats
ITC	Initially transcribing complex
IFN	Interferon
KOW	Kyrpides, Ouzounis, Woese
MDA5	Melanoma differentiation-associated protein 5
mRNA	Messenger RNA
MTase	Methyltransferase
NusG N-terminal	N-terminal NGN
NTase	N-terminal nucleotidyltransferase
NELF	Negative elongation factor
NTP	Nucleotide triphosphate
nt	Nucleotides
OB-domain	Oligonucleotide-binding domain
PCIF1	Phosphorylated CTD interacting factor 1
PAF	Pol II associated factor
PolyA	Polyadenylation
Pol	Polymerase
P-TEFb	Positive transcription elongation factor b
PIC	Pre-initiation complex
Pre-mRNA	Precursor-mRNA

PPi	Pyrophosphate
RIG-I	Retinoic acid inducible gene-I
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RNGTT	RNA guanylytransferase and 5’ phosphatase
Pol II	RNA polymerase II
RAM	RNMT-activating mini protein
SAH	S-adenosylhomocysteine
SAM	S-adenosylmethionine
snRNPs	Small nuclear ribonucleoprotein particles
TBP	Tata binding protein
TF	Transcription factor
TSS	Transcription start site
TPase	Triphosphatase

Declaration

Herewith I declare, that I prepared the doctoral Thesis "Structural studies of RNA polymerase II complexes with human TFIIS and capping enzymes" on my own and with no other sources and aids than quoted.

Göttingen, 31st August 2022