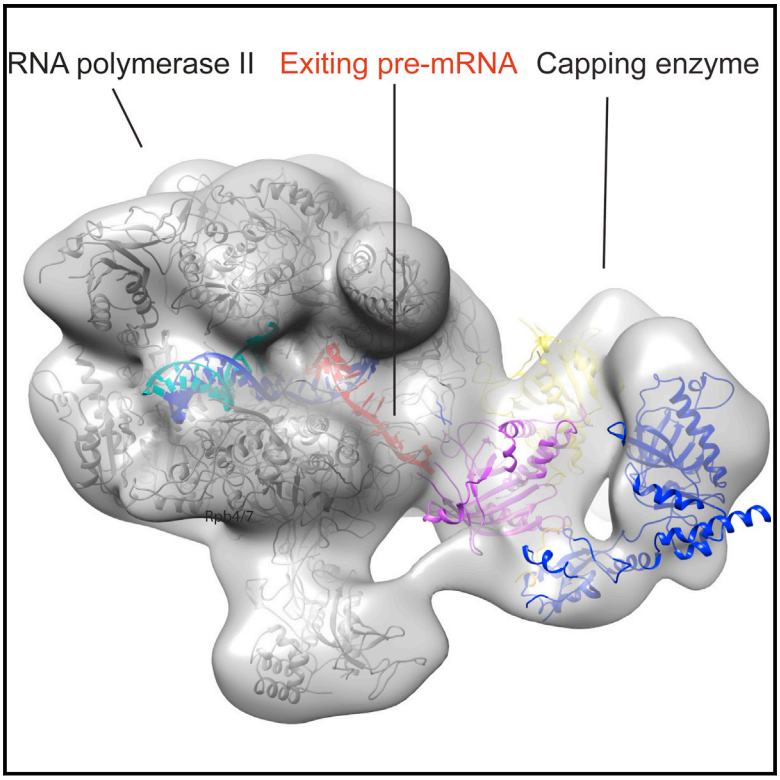
Article

Molecular Cell

Molecular Basis of Transcription-Coupled Pre-mRNA Capping

Graphical Abstract



Authors

Fuensanta W. Martinez-Rucobo, Rebecca Kohler, Michiel van de Waterbeemd, ..., Franz Herzog, Holger Stark, Patrick Cramer

Correspondence

patrick.cramer@mpibpc.mpg.de

In Brief

Co-transcriptional capping is the first step in pre-mRNA processing and required for mRNA stability and translation. Martinez-Rucobo et al. show that the capping enzyme binds around the RNA exit tunnel of RNA polymerase II to ensure seamless RNA protection.

Highlights

- Stable binding of capping enzyme to Pol II requires RNA with a 5′-triphosphate
- Architecture of Pol-II-capping enzyme complex is derived from cryo-EM and crosslinking
- Capping enzyme binds at RNA exit tunnel, explaining seamless RNA protection

Martinez-Rucobo et al., 2015, Molecular Cell *58*, 1079–1089
June 18, 2015
http://dx.doi.org/10.1016/j.molcel.2015.04.004

Molecular Cell
Article

Molecular Basis of Transcription-Coupled Pre-mRNA Capping

Fuensanta W. Martinez-Rucobo,[1] Rebecca Kohler,[1,2] Michiel van de Waterbeemd,[3] Albert J.R. Heck,[3] Matthias Hemann,[1] Franz Herzog,[1] Holger Stark,[4] and Patrick Cramer[2,*]
[1]Gene Center Munich and Department of Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Straße 25, 81377 Munich, Germany
[2]Department of Molecular Biology, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany
[3]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University Padualaan 8, 3584 CH Utrecht, the Netherlands
[4]Max-Planck-Institute for Biophysical Chemistry, Three-dimensional Electron Cryo-Microscopy, Am Fassberg 11, 37077 Göttingen, Germany
*Correspondence: patrick.cramer@mpibpc.mpg.de
http://dx.doi.org/10.1016/j.molcel.2015.04.004

SUMMARY

Capping is the first step in pre-mRNA processing, and the resulting 5′-RNA cap is required for mRNA splicing, export, translation, and stability. Capping is functionally coupled to transcription by RNA polymerase (Pol) II, but the coupling mechanism remains unclear. We show that efficient binding of the capping enzyme (CE) to transcribing, phosphorylated yeast Pol II (Pol IIp) requires nascent RNA with an unprocessed 5′-triphosphate end. The transcribing Pol IIp-CE complex catalyzes the first two steps of capping, and its analysis by mass spectrometry, cryo-electron microscopy, and protein crosslinking revealed the molecular basis for transcription-coupled pre-mRNA capping. CE docks to the Pol II wall and spans the end of the RNA exit tunnel to position the CE active sites for sequential binding of the exiting RNA 5′ end. Thus, the RNA 5′ end triggers its own capping when it emerges from Pol II, to ensure seamless RNA protection from 5′-exonucleases during early transcription.

INTRODUCTION

Capping of the eukaryotic pre-mRNA transcript is the first event in co-transcriptional RNA processing and is essential for cell viability (Ghosh and Lima, 2010; Shuman, 2001). Capping relies on three enzymatic activities, described below for the yeast *Saccharomyces cerevisiae*. First, the RNA triphosphatase Cet1 hydrolyses the gamma phosphate of the 5′-RNA triphosphate end (Rodriguez et al., 1999; Tsukamoto et al., 1997). Then the RNA guanylyltransferase Ceg1 transfers a GMP moiety to the RNA diphosphate end, forming a 5′-5′ link (Shibagaki et al., 1992). Finally, the RNA guanine-N7 methyltransferase Abd1 methylates the terminal guanine base at the N7 position (Mao et al., 1995), resulting in a mature mRNA cap structure.

Cet1 and Ceg1 form a stable complex called the capping enzyme (CE). Yeast Ceg1 is only active when bound to Cet1 (Cho et al., 1998). Metazoan CE consists of a single polypeptide with triphosphatase and guanylyltransferase domains (Ghosh and Lima, 2010; Shuman, 2001). The crystal structure of CE from *S. cerevisiae* is known (Gu et al., 2010), and structures of the triphosphatase and guanylyltransferase components of CE are available for viral (Changela et al., 2005; Håkansson et al., 1997; Håkansson and Wigley, 1998), fungal (Bisaillon and Shuman, 2001; Doamekpor et al., 2014; Fabrega et al., 2003; Gu et al., 2010; Lima et al., 1999), and mammalian enzymes (Changela et al., 2001; Chu et al., 2011; Ghosh et al., 2011).

CE recruitment involves binding of Ceg1 to the carboxy-terminal domain (CTD) of Pol II (Cho et al., 1997; Schroeder et al., 2000). The CTD changes its phosphorylation state during the transcription cycle, and this triggers recruitment of protein factors to Pol II (Buratowski, 2009; Meinhart et al., 2005; Perales and Bentley, 2009; Schuhmacher and Eick, 2013). Ceg1 binds to the CTD when the CTD is phosphorylated at serine-5 (Ser5) residues (Ho and Shuman, 1999), which occurs during transcription initiation. CTD binding to guanylyltransferases was characterized structurally (Doamekpor et al., 2014; Fabrega et al., 2003; Ghosh et al., 2011). Because the CTD is very extended and flexibly linked to Pol II (Cramer et al., 2001), CE could be bound to the CTD but reside very far away from the polymerase surface.

However, there is evidence that capping takes place on the Pol II surface. Capping can occur already upon synthesis of a short RNA of around 20 nt in vitro (Chiu et al., 2002; Coppola et al., 1983; Hagler and Shuman, 1992; Jove and Manley, 1982; Rasmussen and Lis, 1993). In vivo, chromatin immunoprecipitation revealed a sharp occupancy peak for CE subunits just downstream of the transcription start site, in contrast to Abd1, which was observed about 100 nt further downstream (Lidschreiber et al., 2013; Mayer et al., 2010). In the human system, guanylylation is up to 10,000-fold faster for RNA that was part of a transcribing Pol II complex, compared to free RNA, and this functional coupling was only to a minor extent dependent on the CTD (Moteki and Price, 2002). These results indicate the existence of a CE binding site on the Pol II surface. A binding site outside the CTD is also suggested by the observation that

a variant of Cet1 that no longer interacted with Ceg1 was still recruited to the 5′ ends of genes in vivo (Takase et al., 2000).

In order to understand the molecular basis for co-transcriptional capping, the structure of CE in complex with a transcribing Pol II enzyme is required. A pioneering study reported on the crystallization of a complex containing Pol II and CE without nucleic acids (Suh et al., 2010). However, the electron density ascribed to CE in that study could not be unambiguously assigned, because it could not be fitted with known CE structures. In addition, the putative location of CE at the Pol II foot domain (Suh et al., 2010) was far away from the RNA exit site. This location could not explain co-transcriptional capping, because RNA has to be around 15 nt in length to reach the Pol II surface (Andrecka et al., 2009; Kettenberger et al., 2004), and thus short, ~20 nt RNAs could not reach CE at the foot domain. Thus, the site of CE interaction with a transcribing Pol II complex remains unclear.

Determination of the structure of a transcribing Pol II complex with CE has been hampered due to its heterogeneity and flexibility. First, the quaternary structure of CE remains unclear. In particular, *S. cerevisiae* CE contains a Cet1 dimer bound to either one or two Ceg1 molecules, resulting in heterotetrameric or heterotrimeric complexes (Fabrega et al., 2003; Lehman et al., 1999; Lima et al., 1999). It is unclear whether one or two copies of Ceg1 are required for CE function, although the formation of a Cet1 dimer is known to be required in vivo and stabilizes the Cet1 active site in vitro (Hausmann et al., 2003). Second, CE shows very high intrinsic flexibility. Structural studies revealed flexible hinges between Cet1 and Ceg1 (Gu et al., 2010), and also between the two Ceg1 domains NT and OB (Chu et al., 2011; Fabrega et al., 2003; Gu et al., 2010; Håkansson et al., 1997; Håkansson and Wigley, 1998). These flexibilities are apparently required for CE function, because exiting RNA must first bind in the Cet1 active site, must then move to the Ceg1 active site and can then separate from CE.

Here, we show that purified CE exists as a mixture of hetero-tetrameric and hetero-trimeric variants and that it forms a stable complex with transcribing Pol II when the exiting RNA transcript contains a triphosphate at its 5′ end. We then demonstrate that transcribing Pol II-CE complexes are active in catalyzing the first two steps of capping. We derive the three-dimensional architecture of the transcribing Pol II-CE complex by cryo-electron microscopy (cryo-EM) and protein chemical crosslinking coupled to mass spectrometry (CX-MS), which revealed closed and open states of Ceg1. We show that CE spans the end of the RNA exit tunnel of Pol II and that its two active sites face exiting RNA. Our results provide the molecular basis for understanding how capping is functionally coupled to early Pol II transcription and show how nascent pre-mRNA is continuously protected from degradation by 5′-exonucleases.

RESULTS

Exiting RNA Enhances Capping Enzyme Binding to Pol II

We purified a recombinant *S. cerevisiae* CE variant consisting of the catalytic domain of Cet1 (residues 241–549, referred to as Cet1) and full-length Ceg1 (Figure 1A) (Experimental Procedures). Recombinant CE bound to endogenous yeast Pol II at substoichiometric levels when assayed with size exclusion chromatography (Figure 1B). Higher levels of CE remained bound when Pol II was pre-phosphorylated in vitro using P42 MAP kinase (Figure 1B, "Pol IIp") (Suh et al., 2010). The kinase preferentially phosphorylated Ser5 residues of the CTD (Figure S1A). These results were consistent with the known preferred interaction of CE with the Ser5-phosphorylated CTD (Cho et al., 1997; Ho and Shuman, 1999; McCracken et al., 1997; Yue et al., 1997).

To investigate whether CE binding to Pol II was influenced by the presence of nucleic acids, we formed a transcribing Pol IIp complex with a DNA-RNA scaffold containing a 23-nt RNA transcript with a 5′-triphosphate end (Figure 1C, "RNA23-5′ppp"). This nucleic acid scaffold efficiently stabilized CE binding to Pol IIp (Figure 1D), whereas the same scaffold containing a 5′-monophosphate end did not (Figure 1E). This showed that a high level of CE binding to transcribing Pol IIp was dependent on the presence of the RNA 5′-triphosphate end. When we altered the length of the RNA transcript in transcribing complexes, CE binding was enhanced when RNA was at least 17 nt long (Figure S1B). These results were consistent with structural data indicating that RNA must be at least 15 nt in length to reach the Pol II surface (Andrecka et al., 2009; Kettenberger et al., 2004). CE binding depended more on the presence of the RNA 5′-triphosphate than on Pol II phosphorylation (Figures S1C and S1D). Taken together, the presence of an RNA 5′-triphosphate on the Pol II surface strongly enhances CE binding to transcribing Pol II.

Quaternary Structure of the Transcribing Pol II-CE Complex

We used native mass spectrometry (MS) to investigate the quaternary structure of CE and the stoichiometry of CE bound to the transcribing Pol IIp complex (Experimental Procedures; Table S1). Native MS showed that a major fraction of purified CE has a molecular mass of 180.7 kDa, consistent with a hetero-tetramer comprising two copies of Cet1 and two copies of Ceg1 (theoretical MW = 180.5 kDa, Figure 2A). In agreement with our earlier data (Lorenzen et al., 2007), free Pol IIp had a mass of 518.8 kDa. The transcribing Pol IIp complex showed a mass of 539 kDa, consistent with an additional mass of 20.2 kDa for the DNA-RNA scaffold (Figure 2B). Analysis of the transcribing Pol IIp-CE complex revealed a molecular mass of 719.8 kDa (Figure 2C). This corresponded to a hetero-tetrameric CE bound to the transcribing Pol IIp (theoretical MW = 719.1 kDa) and could be confirmed by a tandem MS spectrum (Figure S2). A fraction of the transcribing Pol IIp-CE complexes contained a hetero-trimeric CE missing one Ceg1 molecule (Figure 2C, green peaks). Thus, the transcribing Pol IIp complex can associate with hetero-tetrameric (Ceg1-Cet1-Cet1-Ceg1) and hetero-trimeric (Cet1-Cet1-Ceg1) forms of CE. Two Cet1 copies are present in both these complexes, consistent with a requirement for Cet1 dimerization in vivo (Lehman et al., 2001).

Both CE Enzymes Are Active in Reconstituted Complexes

We then used negative mode MS to investigate whether CE was catalytically active within a transcribing Pol IIp-CE complex (Experimental Procedures). Accurate mass analysis of the free

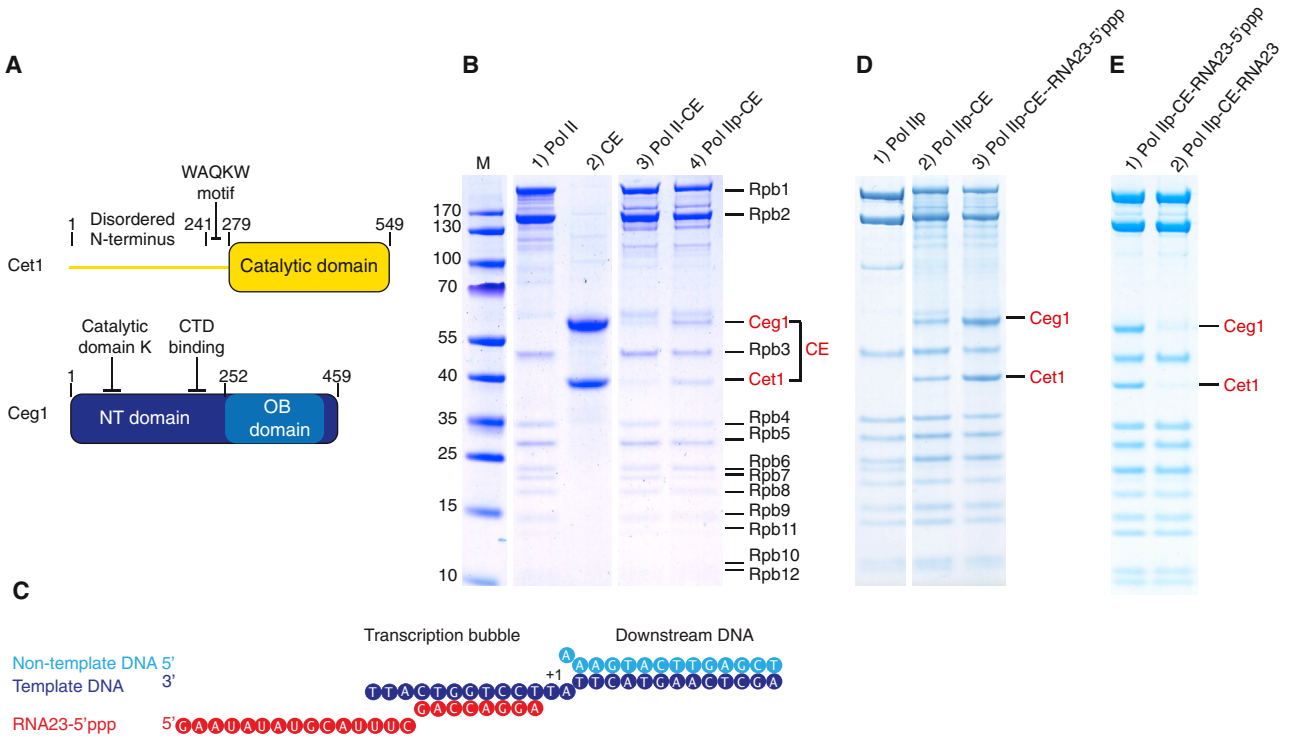


Figure 1. Preparation of the Transcribing Pol IIp-CE Complex

(A) Schematic of domain architecture of the CE triphosphatase (Cet1) and guanylyltransferase (Ceg1) subunits.

(B) SDS-PAGE analysis of purified Pol II, recombinant CE, and complexes of Pol II-CE, and Pol IIp-CE after size exclusion chromatography.

(C) DNA-RNA scaffold used to form a transcribing Pol IIp complex. The DNA template, DNA non-template, and RNA strand are in blue, cyan, and red, respectively. The RNA contains a 5′-triphosphate (5′ppp).

(D) CE binding to Pol IIp is enhanced in the presence of the DNA-RNA scaffold shown in (C). SDS-PAGE analysis of the purified complex after size exclusion chromatography. Pol II (4.7 mg/ml) was incubated with a 2-fold excess of nucleic acid scaffold for 20 min at 18°C before addition of a 3-fold molar excess of CE (6.9 mg/ml) and further incubation for 60 min at 20°C. The formed complex was subjected to size exclusion chromatography on a Superose 6 column.

(E) CE binding is strongly dependent on the presence of the RNA 5′-ppp triphosphate end in the transcribing Pol IIp-CE complex (left lane). The same nucleic acid scaffold lacking the 5′-triphosphate only enables weak CE retention on Pol IIp after size exclusion chromatography (right lane). Assay conducted as in (D).

See also Figure S1.

nucleic acid scaffold was consistent with the presence of a 5′-triphosphate end (“5′ppp”) on the RNA (Figure 3A). The RNA mass remained unchanged following incubation with Pol IIp (Figure 3B). However, when we added CE to the transcribing Pol IIp complex, the initial RNA transcript peak disappeared, and two additional peaks were detected (Figure 3C). The first additional peak was the product of Cet1 activity because it corresponded in mass to 5′-diphosphorylated RNA (Figure 3C, “RNA23-5′pp”). The second additional peak showed the mass of RNA23-5′pp with an added GMP moiety (Figure 3C, “RNA23-5′pppG”) and corresponded to the 5′-guanylated RNA and was the product of both Cet1 and Ceg1 activities.

These data revealed that both enzymatic activities of recombinant CE were intact in reconstituted transcribing Pol IIp-CE complexes. The MS analysis also showed that Cet1 action was complete, whereas Ceg1 action was partial. Ceg1 action was apparently incomplete because of partial co-purification of the cofactor GTP (Figure S3). To test this, we pre-incubated CE with GTP, removed the excess of GTP (Experimental Procedures), and added transcribing Pol IIp. This indeed led to a strong increase in the peak corresponding to the Ceg1 product (Figure 3D, RNA23-5′pppG). However, GTP addition led to partial dissociation of CE from Pol II, as observed by native MS (Figure 2D versus Figure 2C). CE dissociation could not be prevented when GTP was replaced by the non-reactive GTP analog guanosine-5′-[(α,β)-methyleno] triphosphate (GpCpp, also referred to as GMPCPP) (Figure 2E). These results indicated that Cet1 is active in the transcribing Pol IIp-CE complex, and that the newly formed RNA 5′-diphosphate end was moved to the Ceg1 active site, where guanylation of RNA23-5′pp took place in a fraction of complexes and destabilized CE binding to Pol II.

Capping Enzyme Spans the Pol II RNA Exit Tunnel

These results enabled structural studies of a transiently stable and catalytically active transcribing Pol IIp-CE complex by cryo-electron microscopy (EM) (Experimental Procedures). Investigation of the sample by negative stain EM revealed single particles of the expected size and showed extra density on the Pol II surface (Figure S4G). A sample of the same composition was stabilized by protein crosslinking and used for preparing

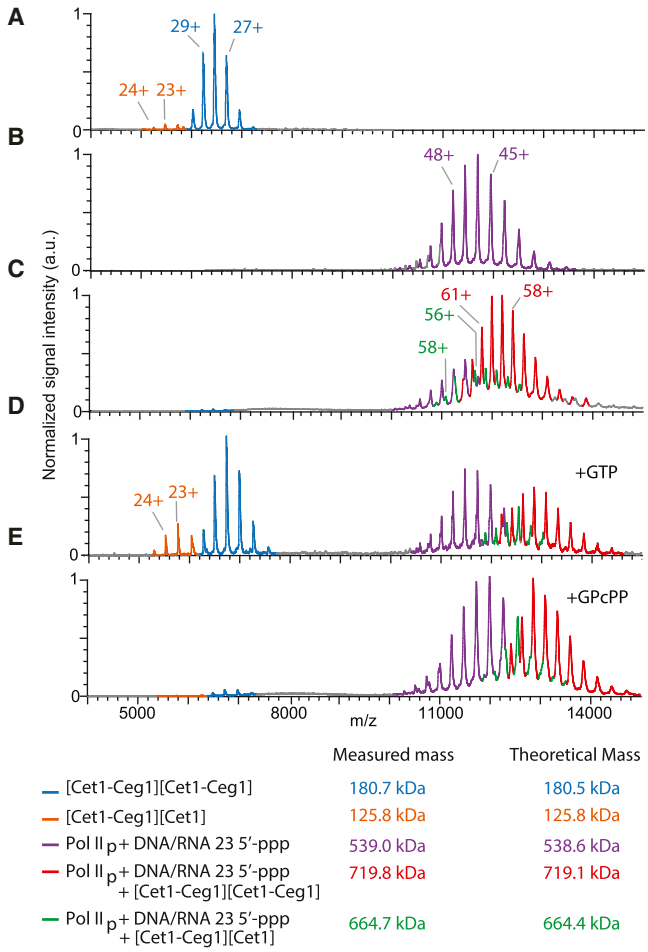


Figure 2. Quartenary Structure of the Transcribing Pol IIp-CE Complex

Native MS analysis of the transcribing Pol IIp-CE complex and its components. For a complete list of the measured masses, see Table S1.

(A) Purified CE (blue) is a hetero-tetramer of 180.7 kDa.

(B) The transcribing Pol IIp complex (violet) has a mass of 539 kDa.

(C) The transcribing Pol IIp complex predominantly binds the CE hetero-tetramer Cet1-Cet1-Ceg1-Ceg1 (red) but also a Cet1-Cet1-Ceg1 hetero-trimer (green).

(D) GTP addition induces CE dissociation from the transcribing Pol IIp complex after RNA guanylation (violet). Excess of CE hetero-trimer (orange) and hetero-tetramer (blue) in the complex mixture could not prevent CE dissociation.

(E) Addition of the non-reactive GTP analog GpCpp cannot prevent CE dissociation from the transcribing Pol IIp complex.

See also Figure S2 and Table S1.

cryo-EM grids (Figure S4H). Micrographs were collected on a FEI Titan Krios microscope with a direct detection device and yielded 250,880 single particles. Using a hierarchical computational image classification scheme (Figure S4A), we obtained two populations of images that were used to reconstruct CE-free Pol II elongation complexes at 7.2-Å (EC-1; 66,214 particles) and 7.4 Å-resolution (EC-2; 61,349 particles) (Figure S4B). In these reconstructions, many details were observed, including the RNA path from the polymerase active site to the end of the RNA exit tunnel (Figure S4C).

We also obtained a particle class that contained strong CE density and led to a reconstruction of the transcribing Pol IIp-CE complex at a resolution of 17.4 Å from 11,007 particle images (Figure 4). CE spans the outer end of the RNA exit tunnel on the Pol II surface (Figure 4A). The CE density is partitioned in a major part, which contacts extensively the area around the RNA exit tunnel on the Pol II core (contact area A, Figure 4A), and a minor part forming a small contact area with the Rpb4/7 subcomplex of Pol II (contact area B, Figure 4A). The high degree of CE flexibility and conformational heterogeneity of the complexes prevented structure determination of the Pol IIp-CE complex at higher resolution (Figure 4D).

Cet1 Contacts Pol II around the Wall

To aid placement of known CE crystal structures into the EM density, we subjected the transcribing Pol IIp-CE complex to CX-MS analysis (Figure 5; Experimental Procedures). From two independent measurements, 527 high-confidence lysine-lysine crosslinks were obtained that resulted in 337 distance restrains (Figure 5A; Table 1; Table S2). A total of 252 crosslinks were observed within Pol II and explained with the Pol II crystal structure, except for one crosslink that exceed the maximum allowed Cα distance of 27 ± 3 Å and involved a flexible protein region (Figure 5C). For Cet1, 51 out of 52 intra-crosslinks were explained with the capping enzyme structure (PDB code 3KYH) (Gu et al., 2010), and one unexplained crosslink again involved a region of conformational flexibility (Figure 5C). These intra-protein crosslinks within Cet1 and within Pol II thus provided positive controls for the CX-MS measurements.

The analysis also yielded 27 crosslinks between Cet1 and Pol II that identified the major CE density as the Cet1 dimer (Figure 5B). Analysis of these inter-protein crosslinks guided placement of the Cet1 dimer crystal structure into the EM density (Figures 4A–4C). One Cet1 subunit resided near Pol II subunits Rpb3 and Rpb12 (Figure 5B, yellow Cet1 copy), and another Cet1 subunit resides at the end of the RNA exit tunnel (Figure 5B, purple Cet1 copy). The latter Cet1 subunit approached the Pol II wall and a flexible extension from the wall called "flap loop" (β30-β31, Rpb2 residues 919–934) (Cramer et al., 2001). The flap loop is not conserved in Pol I and Pol III, and this may explain why CE specifically interacts with Pol II. CE interaction with Pol II is dominated by Cet1 contacts with the Pol II core (contact area A, Figure 4A), explaining why Cet1 remains active in vivo when the Cet1-Ceg1 interaction is perturbed (Takase et al., 2000).

Ceg1 Remains Mobile in the Transcribing Pol II-CE Complex

In contrast to the defined position of the Cet1 homodimer, EM and MS indicated a high degree of variability for Ceg1. The minor part of the CE density corresponded to one flanking Ceg1 molecule ("proximal Ceg1"), whereas density for a second Ceg1 molecule ("distal Ceg1") was only partial (Figures 4B and 4C). The distal Ceg1 molecule is likely absent in a subset of particles containing a heterotrimeric Cet1-Cet1-Ceg1 complex and may be flexible in other particles. We therefore placed only the proximal Ceg1 molecule into the EM density, using Ceg1 from the *S. cerevisiae* CE structure (PDB code 3KYH) (Figures 4A–4C). In

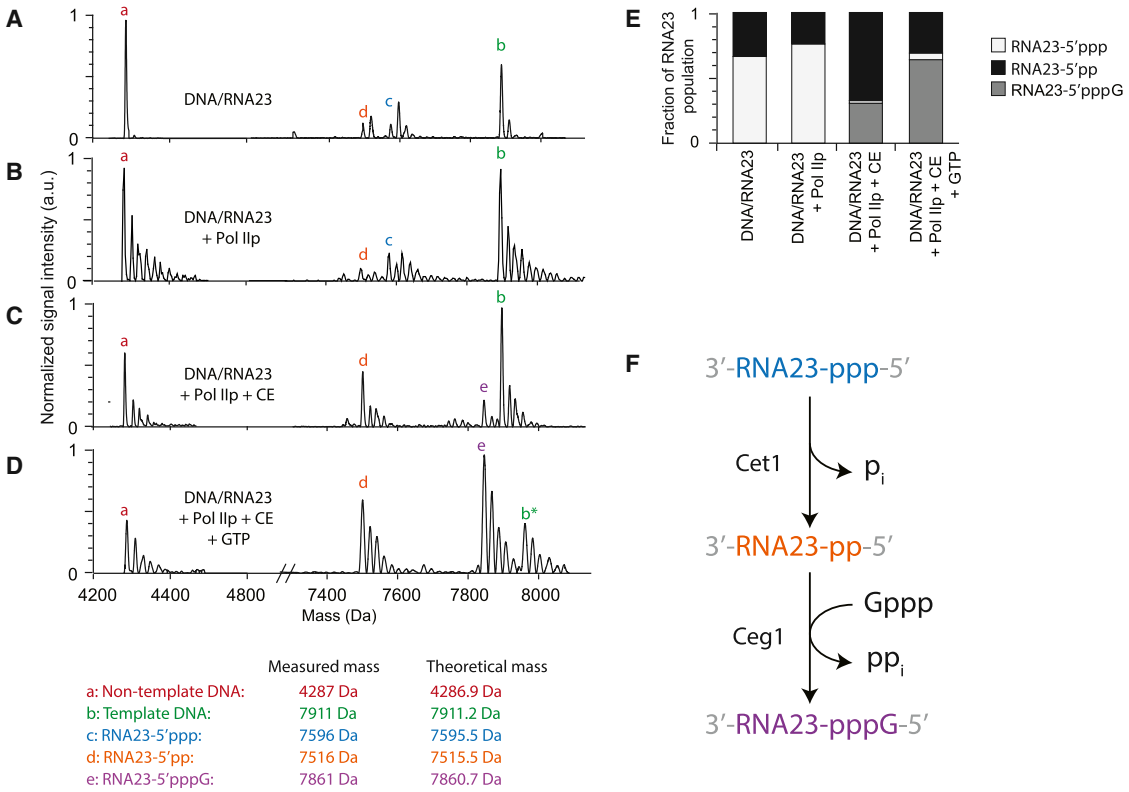


Figure 3. Cet1 and Ceg1 Are Catalytically Active in the Transcribing Pol IIp-CE Complex

Analysis of the nucleic acids by negative mode MS. Peaks corresponding to the different RNA and DNA species are labeled. The additional peaks represent sodium and/or potassium ion adducts.

(A) Analysis of the free nucleic acids scaffold by negative mode MS. The in-vitro-transcribed 23 nt RNA in transcribing Pol IIp-CE complexes contains a 5′-triphosphate (5′ppp, peak "c"). Peaks from the DNA non-template and template strands are denoted "a" and "b."

(B) The RNA peaks remain essentially unchanged upon addition of Pol IIp.

(C) The initial 5′ppp RNA is hydrolyzed to 5′-diphosphorylated RNA (5′pp, peak "d") by Cet1, and a fraction of it is also guanylated by Ceg1-catalyzed addition of a GMP moiety (5′pppG, peak "e").

(D) The RNA23-5′pppG peak ("e") is increased after addition of Ceg1 incubated with GTP. Peak "b*" is shifted with respect to other panels only because the DNA template strand was differentially brominated in the used sample (Experimental Procedures).

(E) Bar plot depicting the fraction of each RNA population in the spectra in (A)–(D).

(F) Schematic of CE catalytic turnover of RNA23-5′ppp to the product RNA23-5′pppG.

See also Figure S3.

the resulting model, most of Ceg1 is covered with EM density, and the Ceg1 OB domain remained near the Cet1 motif WAQKW that it binds in the CE crystal structure (Gu et al., 2010). The model indicated that contact area B is formed between the Ceg1 OB domain and the Pol II subunit Rpb7 (Figure 4A). The model further showed that exiting RNA has to be about 17 and 23 nt long to reach the active sites of Cet1 and Ceg1, respectively (Figures S5C and S5D), readily explaining our binding data (Figure S1B).

The CX-MS analysis further showed that Ceg1 remains intrinsically flexible. The open Ceg1 conformation observed in the *S. cerevisiae* CE crystal structure (PDB code 3KYH) (Gu et al., 2010) could explain only about 60% of the obtained crosslink distances for Ceg1 (114 out of a total of 196) (Figure S5A), whereas 171 out of 196 protein crosslinks could be explained when the closed Ceg1 conformation observed in the structure of a homologous viral guanylyltransferase (PDB code 1CKM) (Håkansson et al., 1997) was considered (Figures 5C–5E; Figures S5A and S5B; Experimental Procedures). The remaining 25 crosslinks that span distances larger than 40 Å indicate that Ceg1 can adopt at least one additional conformation that was thus far not trapped in crystals. A model of CE with Ceg1 in the closed conformation (Experimental Procedures) explained 96 out of 114 Cet1-Ceg1 inter-protein crosslinks (Figure 5E). Thus, Ceg1 adopted several different conformations in the transcribing Pol IIp-CE complex of which we modeled only the one previously observed in the CE crystal structure. These results led to the three-dimensional architecture of a transcribing, active Pol IIp-CE complex, and showed that CE remained flexible in this complex, reflecting its functional nature.

DISCUSSION

Here, we describe how CE interacts with an early transcribing yeast Pol II complex. Biochemical data reveal that efficient

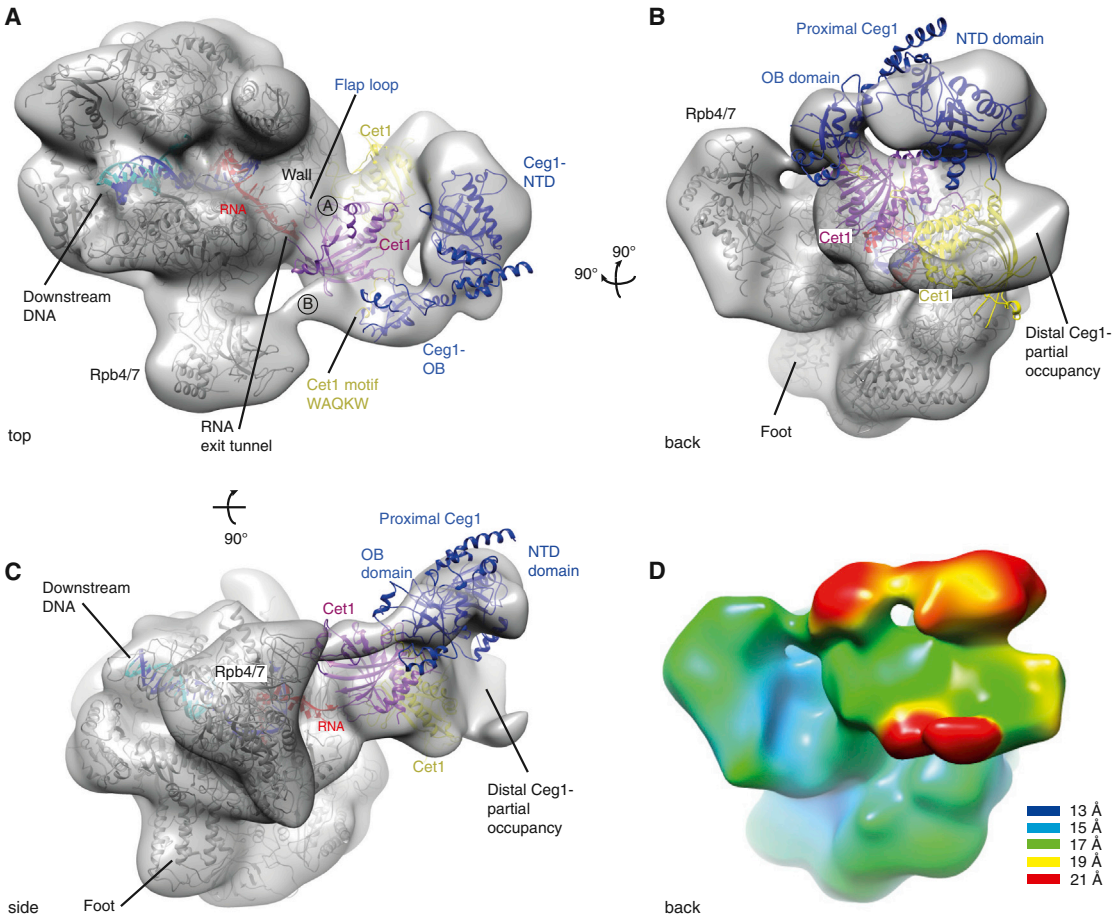


Figure 4. Cryo-EM Reveals that CE Spans the End of Pol II RNA Exit Tunnel

(A) Crystal structures of transcribing Pol II and CE were manually docked into the EM density of the transcribing Pol IIp-CE complex (gray surface) and are depicted as ribbon models (Pol II, gray; DNA, blue/cyan, RNA, red; Cet1 dimer, purple and yellow, and Ceg1 domains, blue). Cet1 placement was facilitated by crosslinking (Figure 5). CE forms two contact areas (encircled letters A and B) with Pol II around the wall and RNA exit tunnel.

(B) Back view of transcribing Pol IIp-CE complex with the proximal Ceg1 monomer (blue) placed into the EM density. Partial density for the distal Ceg1 monomer is indicated.

(C) Side view of the transcribing Pol IIp-CE complex.

(D) Low-pass-filtered local resolution map for the transcribing Pol IIp-CE complex. The resolution for this reconstruction ranges from ~13 Å near the Pol II active center to ~17 Å for the rigid parts of CE (Cet1) and ~21 Å for the outer, flexible Ceg1 molecules on both sides of the Cet1 dimer. View as in (B).

See also Figure S4.

binding of CE to Pol II requires the presence of an exiting RNA transcript with an intact 5′-triphosphate end, in addition to the known interaction with the Ser5-phosphorylated CTD. This finding enabled us to reconstitute and investigate a transcribing Pol IIp-CE complex by native mass spectrometry, cryo-EM, and CX-MS, resulting in the three-dimensional architecture of the active complex. Mass spectrometry showed that both enzymatic activities of CE were fully functional inside this complex and suggested that the RNA 5′ end resided mainly in the active site of Ceg1. Consistent with this, Ceg1 adopted both an open and a closed conformation, and guanylation was observed to facilitate CE release from Pol II.

Our data revealed that CE docks on the Pol II wall and spans the end of the RNA exit tunnel of Pol II. This shows that the first two steps in pre-mRNA capping can take place when the nascent RNA reaches the polymerase surface. CE is ideally positioned for instant recognition of the exiting RNA 5′ end. This mechanism ensures a seamless protection of RNA from degradation by 5′-exonucleases right from the beginning of transcription. Our data also indicate that Ceg1 remains mobile in the transcribing Pol IIp-CE complex. Ceg1 mobility is likely required for CE function but unfortunately sets a limit to the resolution of the EM analysis.

Our results converge with published work on the following model for transcription-coupled pre-mRNA capping (Figure 6, stages 1–6). During transcription initiation, unphosphorylated Pol II assembles with general transcription factors on promoter DNA, and RNA synthesis commences (stage 1). When the RNA transcript reaches a length of around 13 nt, general factors are apparently released (Cabart et al., 2011), and a transcribing

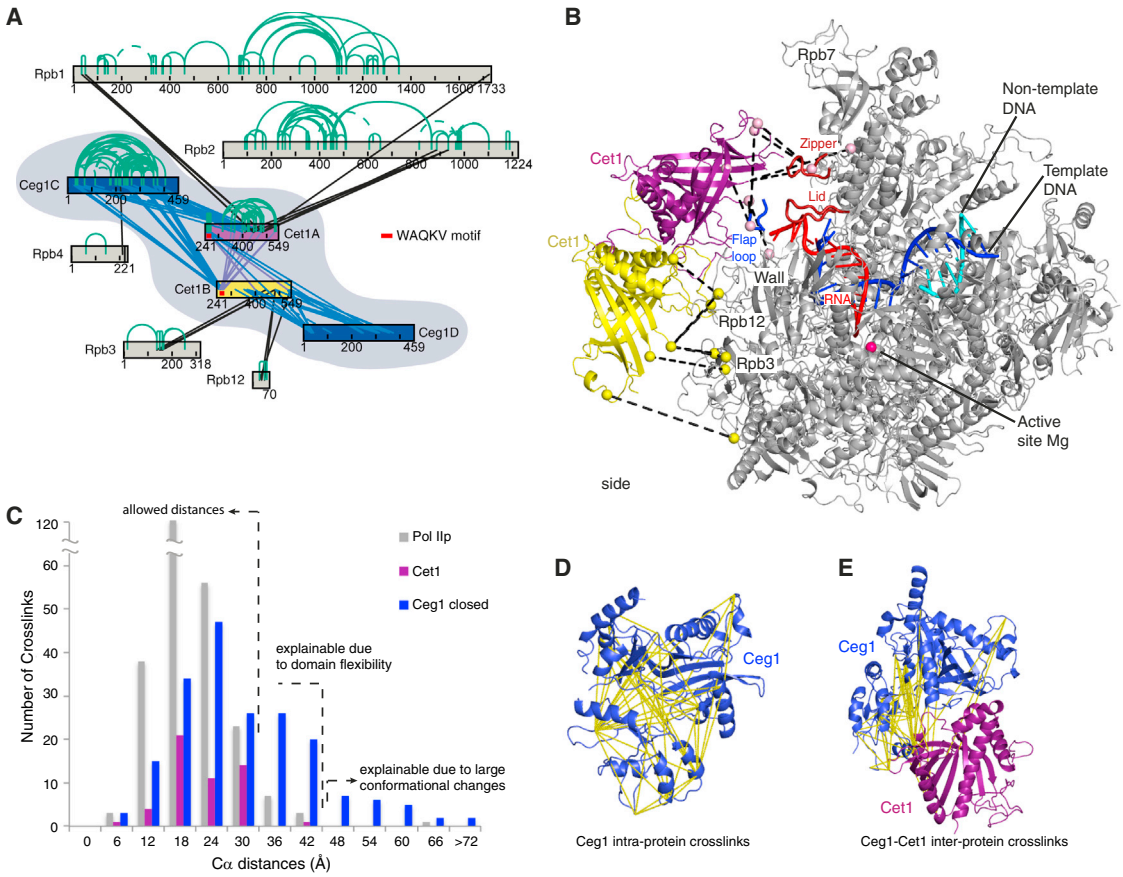


Figure 5. Crosslinking Defines the Architecture of the Transcribing Pol IIp-CE Complex

(A) Crosslink map of the transcribing Pol IIp-CE complex. Crosslinks between Pol II subunits were excluded for clarity. Two copies of Cet1 and Ceg1 are shown. Intra-protein crosslinks are in green, inter-protein crosslinks between Pol II and CE in black, inter-protein crosslinks between Cet1 and Ceg1 in blue, and crosslinks between Cet1 molecules in lilac.

(B) Inter-protein crosslinks between Pol II and CE (black dotted lines) guided placement of one Cet1 subunit (yellow) near Pol II subunits Rpb3 and Rpb12, and the other Cet1 subunit (purple) at the end of the RNA exit tunnel. This placement also fits the EM density (Figure 4). The view is from the side but rotated by 180 degrees around a vertical axis compared to the side view in Figure 4C.

(C) Cα distance distribution for observed lysine-lysine crosslinks. Crosslinks spanning distances of 31–40 Å are explained by the conformational flexibility of mobile protein loops. Distance restraints indicate that Ceg1 resides predominantly in the closed conformation, but alternative states apparently coexist.

(D) Intra-protein crosslinks on the *S. cerevisiae* Ceg1 structure (PDB code 3KYH) in the closed state after domain alignment onto the viral enzyme structure (PDB code 1CKM).

(E) Inter-protein Cet1-Ceg1 crosslinks between the *S. cerevisiae* Ceg1 structure in the closed state as in (D).

See also Figure S5 and Table S2.

complex is formed that contains a Ser5-phophorylated CTD and can bind CE (stage 2). When the RNA reaches a length of around 17 nt, CE docks to the Pol II surface around the end of the RNA exit tunnel (stage 3). Cet1 then hydrolyzes the 5′-triphosphate end of the nascent RNA, resulting in a diphosphate end that is transferred to the Ceg1 active site and coupled to a GMP moiety (stage 4). Guanylation stimulates dissociation of CE from the polymerase surface, and a subsequent decrease in Ser5 phosphorylation of the CTD leads to complete release of CE from Pol II. Abd1 binds the Ser2-phosphorylated CTD and catalyzes addition of the methyl group to the cap (stage 5). Failure to methylate the cap results in decapping and RNA degradation in a surveillance mechanism (Jiao et al., 2010). The complete cap binds the cap-binding complex (CBC), which stabilizes the RNA further and stimulates processive elongation (stage 6).

EXPERIMENTAL PROCEDURES

Sample Production and Purification

S. cerevisiae genes encoding Cet1 (241–549) and full-length Ceg1 were cloned into a bicistronic pOPINE vector. In order to obtain a ribosomal binding site (RBS) before each gene, a PCR product containing Cet1-RBS-Ceg1-Strep2 tag was cloned in the vector. Only Ceg1 contained a Strep2 tag. The plasmid was transformed into *E. coli* Rosetta (DE3) pLysS (Novagen). Cells were grown in TB medium at 37°C to an OD_{600} of 0.6. Expression was induced with 0.5 mM IPTG for 16 hr at 21°C. Cells were lysed by sonication in buffer L (20 mM Tris [pH 8.0], 100 mM NaCl, 3 mM DTT, 2.5 mM $MgCl_2$, 10 μg/ml DNase, 14.2 μg leupeptin, 68.5 μg pepstatin A, 8 mg PMSF, 16 mg benzamidine). After

Table 1. Summary of Crosslinks and Distance Restraints Identified in the Transcribing Pol IIp-CE Complex by CX-MS

Protein 1	Protein 2	Restraints	Crosslinks
Inter-crosslinks			
Cet1	Ceg1	73	114
Cet1	Pol II	17	27
	Rpb9	1	1
	Rpb1	5	6
	Rpb2	4	6
	Rpb3	4	10
	Rpb12	3	4
Ceg1	Pol II	1	1
	Rpb4	1	1
Pol II	Pol II	50	91
	Total	141	233
Intra-crosslinks			
	Ceg1	58	81
	Cet1	35	52
	Pol II	103	161
	Total	196	294

The numbers of inter-protein crosslinks between CE and Pol II subunits and of intra-protein crosslinks for the individual proteins are given. A distance restraint is defined by a unique peptide-peptide combination that is often identified by more than one crosslink (lysine-lysine combination).

centrifugation at 15,000 × *g* for 20 min the supernatant was loaded onto a Strep-Tactin macro prep column (IBA) equilibrated in buffer L. Strep-tagged Ceg1 was eluted with buffer L containing 2.5 mM desthiobiotin and Cet1 co-purified.

The eluted CE complex was further purified using a HiTrap Heparin HP column (GE Healthcare) with a gradient from buffer H1 (20 mM HEPES, 100 mM NaCl, 3 mM DTT, 5 mM $MgCl_2$) to buffer H2 (buffer H1 plus 1M NaCl). Pure CE complex was obtained after gel filtration on a Superdex 200 column (GE Healthcare) in buffer GF (5 mM HEPES [pH 7.25], 100 mM KCl, 3 mM DTT). The eluted complex was concentrated to 7 mg/ml, aliquoted, and stored at −80°C. *S. cerevisiae* Pol II was prepared as described (Sydow et al., 2009), and the final gel-filtration step was carried out in buffer GF. Synthetic DNA and RNA oligonucleotides were purchased from Metabion. RNA containing a 5′ triphosphate end was transcribed in vitro from a modified pSP64 plasmid (Promega) containing an autocleavable 3′-terminal hepatitis delta virus ribozyme. The product RNA was purified as previously described (Müller et al., 2006).

To obtain Ser5-phosphorylated Pol II (Pol IIp) 1 mg of purified Pol II was incubated with 17 μl P42 MAP kinase (NEB, 100 U/μl) at 30°C for 15 min. The sample was subjected to gel filtration using a Superose 6 column (GE Healthcare) equilibrated in buffer GF. Pol IIp was concentrated to 3–4 mg/ml, aliquoted, and stored at −80°C. The phosphorylation stage of the Pol II-CTD was analyzed by western blotting with anti-phospho CTD antibodies (Figure S1A). To obtain the CE-Pol II complex a 3-fold molar excess of CE was incubated with Pol II for 60 min at 20°C followed by gel filtration in buffer GF. To obtain the transcribing Pol IIp-CE complex, Pol II was incubated with a 2-fold excess of nucleic acid scaffold prepared as described (Kettenberger et al., 2004) and incubated for 20 min at 20°C before addition of CE.

Mass Spectrometric Analyses

Complex samples were analyzed in 200 mM ammonium acetate (pH 7.3) at a protein concentration of 0.5–2 μM obtained by six sequential concentrations and dilution steps at 4°C using a 10-kDa molecular weight cutoff centrifugal filter. Proteins were introduced into the mass spectrometer using electrospray ionization with gold-coated borosilicate capillaries. Native mass spectrometry was performed using either an Orbitrap Exactive EMR (ThermoFisher Scientific) with extended mass range or a QTOF II (Waters) modified for high mass detection (Tahallah et al., 2001; van den Heuvel et al., 2006). Typical operating parameters were: capillary voltage 1,300–1,500 V, sample cone voltage 75–125 V, extraction cone voltage 0–35 V, collision cell voltage 75–160 V, source region pressure 8 mbar, collision gas pressure 0.015 mbar. All mass spectra were calibrated using CsI clusters.

Mass spectrometry of proteins under denaturing conditions was achieved by diluting samples in 1%–10% (V/V) formic acid/water solutions followed by analysis using a Micromass LCT mass spectrometer. Oligonucleotide spectra were acquired on the same instrument in negative ion mode after dilution in 25%–50% (V/V) methanol/water solutions. Data analysis was performed with the Water Masslynx 4.1 software. All measurements were done with template DNA with sequence 5′-TTA CTG GTC CTT ATT CAT GAA CTC GA-3′, except for the data shown in Figure 3D that were obtained with bromine-labeled template DNA at the position indicated by B in the sequence 5′-TTA CTG GBC CTT ATT CAT GAA CTC GA-3′.

EM Sample Preparation and Data Collection

The purified Pol IIp-CE-nucleic acids complex was crosslinked with 1.6 mM of an equimolar mixture of heavy- and light-labeled disuccinimidyl suberate (DSS-d0/d12, d is deuterium, Creative Molecules) for 30 min at 30°C as indicated in the CXMS methods section and subjected to gel filtration on a Superose 6 column (GE Healthcare) equilibrated in buffer GF. The crosslinked sample was concentrated to 0.25 mg/ml. For negative stain EM, samples of the Pol IIp-CE complex were applied to Quantifoil grids coated with a second layer of continuous carbon. The sample was allowed to bind for 1 min, washed with several drops of distilled water, stained for 1 min in a 50 μl drop of 2% (w/v) uranyl formate solution, and blotted with filter paper. Cryo EM samples of the transcribing Pol IIp-CE complex were prepared on Quantifoil grids coated with a second layer of continuous carbon. The sample was incubated for 45 s on the carbon foil, blotted, and vitrified in liquid ethane with a FEI Vitrobot Mark IV. Data were acquired on a C_s corrected FEI Titan Krios at 300 kV acceleration voltage. Micrographs were collected using a FEI Falcon II direct detector with defocus values between −1 and −4 μm at a nominal magnification of 74,000× (1.0 Å/pixel). The total dose was 25 $e^-/Å^2$. Micrographs were binned by a factor of 2 in Fourier space (2 Å/pixel) and used for image processing.

EM Image Processing and 3D Reconstruction

Particles (250,880) were selected semi-automatically using EMAN Boxer (Ludtke et al., 1999). CTF parameters were estimated using CTFFIND (Mindell and Grigorieff, 2003). CTF correction and 3D reconstruction were performed in RELION (Scheres, 2012a, 2012b). The selected particles were extracted with a 148^2 pixel box and normalized during the pre-processing procedure in RELION. 3D classification and refinement were also done in RELION. Initial unsupervised 3D classification in RELION led to a subset of particles that contained density for CE. This subset contained 1.5% of the particles. Directly using unsupervised classification was not very effective. We thus added one step of supervised classification to the procedure. The map obtained from the initial unsupervised RELION classification with density for CE was used as a first reference for supervised classification to divide the large data set into two smaller subsets. To obtain a second reference map for supervised classification, the initial CE density was erased from the first map with the volume eraser tool in UCSF Chimera (Pettersen et al., 2004). All particles were aligned to both references in RELION, and the particles were sorted into two populations (118,817 particles and 132,063 particles) according to the value _rlnLogLikeliContribution. Both populations were then subjected to one round of 3D classification in RELION with six classes each.

The 118,817 particles in the first population yielded one class (66,214 particles) representing a Pol II elongation complex (EC-1) that was refined to a resolution of 7.2 Å, plus five classes representing less well-resolved ECs or bad maps likely containing ill-aligned particles. The 132,063 particles in a second population yielded one EC (61,349 particles) that was refined with RELION to a resolution of 7.4 Å, one complex with strong CE features, and four classes representing less well-resolved ECs or poor maps. The CE-containing particles were subjected to two further rounds of 3D classification into four classes

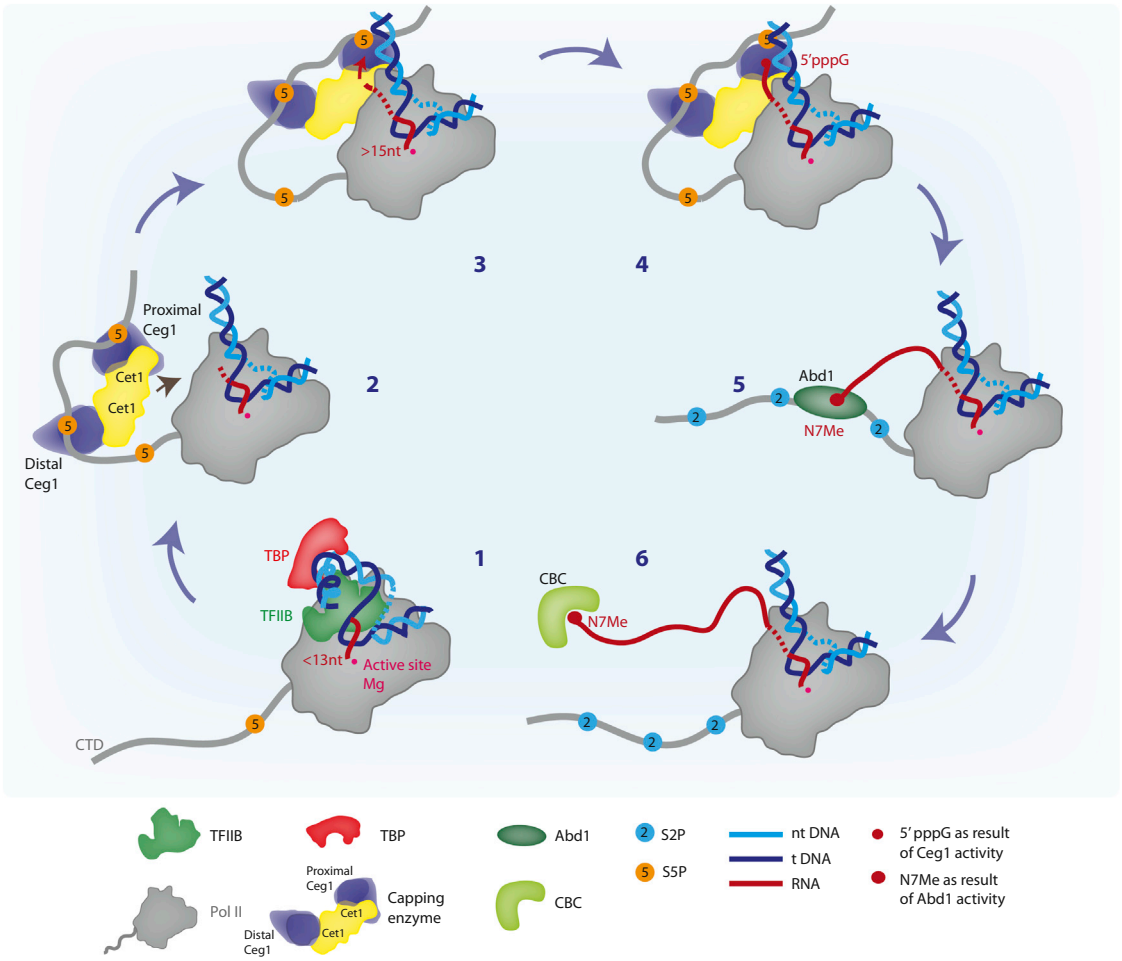


Figure 6. Model for Transcription-Coupled Pre-mRNA Capping

The model depicts six stages of transcription-coupled pre-mRNA capping. For details compare text in the discussion paragraph.

and three classes with the aim to get a more homogeneous subset of particles. However, both rounds resulted in a large main population and only small additional populations, meaning that further separation into more homogeneous subsets was not possible. Likely, the flexibility of CE resulted in gradual heterogeneity of the particles that could not be further resolved. The final 11,007 particles for the transcribing Pol IIp-CE complex were refined to 17.4-Å resolution with RELION. Sampling was 7.5 degrees with five pixels offset search range and one pixel offset search step for 3D classification, and reference maps for 3D classification and refinement were filtered to 45–50 Å. A regularization parameter of four and wide, soft spherical masks of 290 Å diameter were used. Classifications were iterated until convergence.

Post-processing was done in RELION, using soft, wide particle shaped masks produced in SPIDER (Shaikh et al., 2008). B-factors were estimated automatically, and FSC weighting was applied. The B-factor was then used for local resolution estimates using the program RESMAP (Kucukelbir et al., 2014). Local resolution maps (LRMs) were obtained using method83 implemented to run on a Graphics Processing Unit. A sliding window of 20 (transcribing Pol IIp-CE complex) or 44 (transcribing Pol IIp complex) was centered around each voxel, extending the original half-maps through mirroring at the borders. The FSC at 0.5 was then calculated within the window and assigned to the central voxel. The conservative value of 0.5 was chosen to compensate for the higher FSC variance introduced by smaller window sizes. All structural figures were generated using UCSF Chimera (Pettersen et al., 2004) and PyMOL.

CXMS Analysis and Modeling

The transcribing Pol IIp-CE complex was crosslinked with 1.6 mM of an equimolar mixture of heavy- and light-labeled disuccinimidyl suberate (DSS-d0/d12, d is deuterium, Creative Molecules) as described (Herzog et al., 2012). Crosslinked proteins were digested and the peptides were enriched and analyzed by liquid chromatography coupled to an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific). Crosslink spectra were searched by the xQuest software and crosslink identifications were manually validated at false discovery rates <5% (Walzthoeni et al., 2012).

To fit the Ceg1 crosslink distances in the allowed distance range, the OB and NT domains of Ceg1 were independently superposed to the corresponding domains of available guanylyltransferase structures (Chu et al., 2011; Fabrega et al., 2003; Ghosh et al., 2011; Gu et al., 2010; Håkansson et al., 1997; Håkansson and Wigley, 1998) using Chimera. The *C. virus* structure was used as reference for the final superposition (Figure S5B). To avoid clashes with Cet1 in the CE structure, the resulting *S. cerevisiae* Ceg1 model in the closed conformation was rotated together with the Cet1 WAQKW motif. Crosslink distances were generated with PyMOL and analyzed manually.

ACCESSION NUMBERS

The EM maps in this manuscript were deposited in the EMDB with accession codes EMD-2922 (transcribing Pol IIp-CE complex) and EMD-2923 (transcribing Pol IIp complex).

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2015.04.004.

AUTHOR CONTRIBUTIONS

F.W.M.-R. and P.C. designed the study. F.W.M.-R. carried out biochemical experiments. R.K. and F.W.M.-R. carried out electron microscopy. M.H. and F.H. carried out CXMS analysis. M.v.d.W. and A.J.R.H. carried out native MS and RNA processing analysis. F.W.M.-R. and P.C. wrote the manuscript with input from all authors. P.C. supervised the project, except for CXMS and native MS, which were supervised by F.H. and A.J.R.H., respectively, and EM sample preparation and data acquisition, which was supervised by H.S.

ACKNOWLEDGMENTS

We would like to thank Dimitry Tegunov (Max Plank Institute of Biochemistry in Munich), Clemens Plaschka, Otto Berninghausen, and Charlotte Ungewickell for advice and help (Gene Center Munich). F.W.M.-R. was supported by the BGF 2011. R.K. was supported by the SNSF and the European Commission (FP7-PEOPLE-IEF). F.H. was funded by the Bavarian Research Network for Molecular Biosystems and by an LMU*excellent* junior grant. M.v.d.W. and A.J.R.H. were supported by the Stichting voor Fundamenteel Onderzoek der Materie (FOM; project 12PR3033-2) and the Netherlands Science Organization (NWO) funded proteomics facility Proteins @ Work (no. 184.032.201). P.C. was supported by the DFG (SFB860), the Advanced Investigator Grant TRANSIT of the European Research Council, and the Volkswagen Foundation.

Received: January 20, 2015
Revised: February 27, 2015
Accepted: March 31, 2015
Published: May 7, 2015

REFERENCES

Andrecka, J., Treutlein, B., Arcusa, M.A., Muschielok, A., Lewis, R., Cheung, A.C., Cramer, P., and Michaelis, J. (2009). Nano positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex. Nucleic Acids Res. *37*, 5803–5809.

Bisaillon, M., and Shuman, S. (2001). Structure-function analysis of the active site tunnel of yeast RNA triphosphatase. J. Biol. Chem. *276*, 17261–17266.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol. Cell *36*, 541–546.

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

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.

Changela, A., Martins, A., Shuman, S., and Mondragón, A. (2005). Crystal structure of baculovirus RNA triphosphatase complexed with phosphate. J. Biol. Chem. *280*, 17848–17856.

Chiu, Y.L., Ho, C.K., Saha, N., Schwer, B., Shuman, S., and Rana, T.M. (2002). Tat stimulates cotranscriptional capping of HIV mRNA. Mol. Cell *10*, 585–597.

Cho, E.J., Takagi, T., Moore, C.R., and Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev. *11*, 3319–3326.

Cho, E.J., Rodriguez, C.R., Takagi, T., and Buratowski, S. (1998). Allosteric interactions between capping enzyme subunits and the RNA polymerase II carboxy-terminal domain. Genes Dev. *12*, 3482–3487.

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. USA *108*, 10104–10108.

Coppola, J.A., Field, A.S., and Luse, D.S. (1983). Promoter-proximal pausing by RNA polymerase II in vitro: transcripts shorter than 20 nucleotides are not capped. Proc. Natl. Acad. Sci. USA *80*, 1251–1255.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science *292*, 1863–1876.

Doamekpor, S.K., Sanchez, A.M., Schwer, B., Shuman, S., and Lima, C.D. (2014). How an mRNA capping enzyme reads distinct RNA polymerase II and Spt5 CTD phosphorylation codes. Genes Dev. *28*, 1323–1336.

Fabrega, C., Shen, V., Shuman, S., and Lima, C.D. (2003). Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol. Cell *11*, 1549–1561.

Ghosh, A., and Lima, C.D. (2010). Enzymology of RNA cap synthesis. Wiley Interdiscip. Rev. RNA *1*, 152–172.

Ghosh, A., Shuman, S., and Lima, C.D. (2011). Structural insights to how mammalian capping enzyme reads the CTD code. Mol. Cell *43*, 299–310.

Gu, M., Rajashankar, K.R., and Lima, C.D. (2010). Structure of the Saccharomyces cerevisiae Cet1-Ceg1 mRNA capping apparatus. Structure *18*, 216–227.

Hagler, J., and Shuman, S. (1992). A freeze-frame view of eukaryotic transcription during elongation and capping of nascent mRNA. Science *255*, 983–986.

Håkansson, 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. USA *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.

Hausmann, S., Pei, Y., and Shuman, S. (2003). Homodimeric quaternary structure is required for the in vivo function and thermal stability of Saccharomyces cerevisiae and Schizosaccharomyces pombe RNA triphosphatases. J. Biol. Chem. *278*, 30487–30496.

Herzog, F., Kahraman, A., Boehringer, D., Mak, R., Bracher, A., Walzthoeni, T., Leitner, A., Beck, M., Hartl, F.U., Ban, N., et al. (2012). Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. Science *337*, 1348–1352.

Ho, C.K., and Shuman, S. (1999). 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.

Jiao, X., Xiang, S., Oh, C., Martin, C.E., Tong, L., and Kiledjian, M. (2010). Identification of a quality-control mechanism for mRNA 5′-end capping. Nature *467*, 608–611.

Jove, R., and Manley, J.L. (1982). Transcription initiation by RNA polymerase II is inhibited by S-adenosylhomocysteine. Proc. Natl. Acad. Sci. USA *79*, 5842–5846.

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.

Kucukelbir, A., Sigworth, F.J., and Tagare, H.D. (2014). Quantifying the local resolution of cryo-EM density maps. Nat. Methods *11*, 63–65.

Lehman, K., Schwer, B., Ho, C.K., Rouzankina, I., and Shuman, S. (1999). A conserved domain of yeast RNA triphosphatase flanking the catalytic core regulates self-association and interaction with the guanylyltransferase component of the mRNA capping apparatus. J. Biol. Chem. *274*, 22668–22678.

Lehman, K., Ho, C.K., and Shuman, S. (2001). Importance of homodimerization for the in vivo function of yeast RNA triphosphatase. J. Biol. Chem. *276*, 14996–15002.

Lidschreiber, M., Leike, K., and Cramer, P. (2013). Cap completion and C-terminal repeat domain kinase recruitment underlie the initiation-elongation transition of RNA polymerase II. Mol. Cell. Biol. *33*, 3805–3816.

Lima, C.D., Wang, L.K., and Shuman, S. (1999). Structure and mechanism of yeast RNA triphosphatase: an essential component of the mRNA capping apparatus. Cell *99*, 533–543.

Lorenzen, K., Vannini, A., Cramer, P., and Heck, A.J. (2007). Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. Structure *15*, 1237–1245.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. *128*, 82–97.

Mao, X., Schwer, B., and Shuman, S. (1995). Yeast mRNA cap methyltransferase is a 50-kilodalton protein encoded by an essential gene. Mol. Cell. Biol. *15*, 4167–4174.

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

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.

Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S., and Cramer, P. (2005). A structural perspective of CTD function. Genes Dev. *19*, 1401–1415.

Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. *142*, 334–347.

Moteki, S., and Price, D. (2002). Functional coupling of capping and transcription of mRNA. Mol. Cell *10*, 599–609.

Müller, M., Weigand, J.E., Weichenrieder, O., and Suess, B. (2006). Thermodynamic characterization of an engineered tetracycline-binding riboswitch. Nucleic Acids Res. *34*, 2607–2617.

Perales, R., and Bentley, D. (2009). "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. Mol. Cell *36*, 178–191.

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.

Rasmussen, E.B., and Lis, J.T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc. Natl. Acad. Sci. USA *90*, 7923–7927.

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. *27*, 2181–2188.

Scheres, S.H. (2012a). A Bayesian view on cryo-EM structure determination. J. Mol. Biol. *415*, 406–418.

Scheres, S.H. (2012b). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. *180*, 519–530.

Schroeder, S.C., Schwer, B., Shuman, S., and Bentley, D. (2000). Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes Dev. *14*, 2435–2440.

Schuhmacher, M., and Eick, D. (2013). Dose-dependent regulation of target gene expression and cell proliferation by c-Myc levels. Transcription *4*, 192–197.

Shaikh, T.R., Gao, H., Baxter, W.T., Asturias, F.J., Boisset, N., Leith, A., and Frank, J. (2008). SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. Nat. Protoc. *3*, 1941–1974.

Shibagaki, Y., Itoh, N., Yamada, H., Nagata, S., and Mizumoto, K. (1992). mRNA capping enzyme. Isolation and characterization of the gene encoding mRNA guanylytransferase subunit from Saccharomyces cerevisiae. J. Biol. Chem. *267*, 9521–9528.

Shuman, S. (2001). Structure, mechanism, and evolution of the mRNA capping apparatus. Prog. Nucleic Acid Res. Mol. Biol. *66*, 1–40.

Suh, M.H., Meyer, P.A., Gu, M., Ye, P., Zhang, M., Kaplan, C.D., Lima, C.D., and Fu, J. (2010). A dual interface determines the recognition of RNA polymerase II by RNA capping enzyme. J. Biol. Chem. *285*, 34027–34038.

Sydow, J.F., Brueckner, F., Cheung, A.C., Damsma, G.E., Dengl, S., Lehmann, E., Vassylyev, D., and Cramer, P. (2009). Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. Mol. Cell *34*, 710–721.

Tahallah, N., Pinkse, M., Maier, C.S., and Heck, A.J. (2001). The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. Rapid Commun. Mass Spectrom. *15*, 596–601.

Takase, Y., Takagi, T., Komarnitsky, P.B., and Buratowski, S. (2000). The essential interaction between yeast mRNA capping enzyme subunits is not required for triphosphatase function in vivo. Mol. Cell. Biol. *20*, 9307–9316.

Tsukamoto, T., Shibagaki, Y., Imajoh-Ohmi, S., Murakoshi, T., Suzuki, M., Nakamura, A., Gotoh, H., and Mizumoto, K. (1997). Isolation and characterization of the yeast mRNA capping enzyme beta subunit gene encoding RNA 5′-triphosphatase, which is essential for cell viability. Biochem. Biophys. Res. Commun. *239*, 116–122.

van den Heuvel, R.H., van Duijn, E., Mazon, H., Synowsky, S.A., Lorenzen, K., Versluis, C., Brouns, S.J., Langridge, D., van der Oost, J., Hoyes, J., and Heck, A.J. (2006). Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. Anal. Chem. *78*, 7473–7483.

Walzthoeni, T., Claassen, M., Leitner, A., Herzog, F., Bohn, S., Förster, F., Beck, M., and Aebersold, R. (2012). False discovery rate estimation for cross-linked peptides identified by mass spectrometry. Nat. Methods *9*, 901–903.

Yue, Z., Maldonado, E., Pillutla, R., Cho, H., Reinberg, D., and Shatkin, A.J. (1997). Mammalian capping enzyme complements mutant Saccharomyces cerevisiae lacking mRNA guanylyltransferase and selectively binds the elongating form of RNA polymerase II. Proc. Natl. Acad. Sci. USA *94*, 12898–12903.

Molecular Cell, Volume *58*

Supplemental Information

Molecular Basis of Transcription-Coupled pre-mRNA Capping

Fuensanta W. Martinez-Rucobo, Rebecca Kohler, Michiel van de Waterbeemd, Albert J.R. Heck, Matthias Hemann, Franz Herzog, Holger Stark, and Patrick Cramer

Figure S1

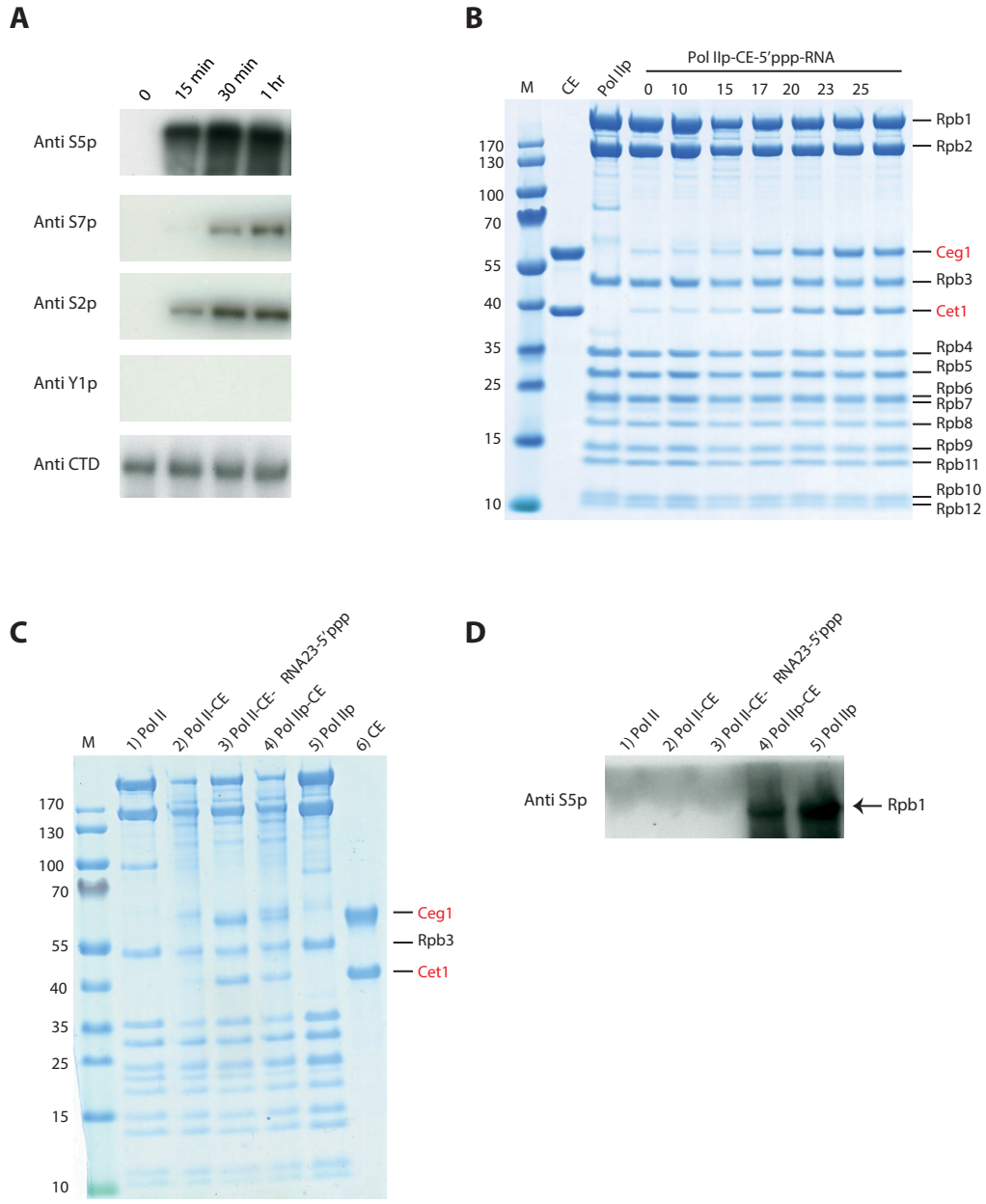

A
0
15 min
30 min
1 hr
Anti S5p
Anti S7p
Anti S2p
Anti Y1p
Anti CTD
B
M
CE
Pol IIp
Pol IIp-CE-5'ppp-RNA
0
10
15
17
20
23
25
170
130
100
70
55
40
35
25
15
10
Rpb1
Rpb2
Ceg1
Rpb3
Cet1
Rpb4
Rpb5
Rpb6
Rpb7
Rpb8
Rpb9
Rpb11
Rpb10
Rpb12
C
M
1) Pol II
2) Pol II-CE
3) Pol II-CE-
RNA23-5'ppp
4) Pol IIp-CE
5) Pol IIp
6) CE
170
130
100
70
55
40
35
25
15
10
Ceg1
Rpb3
Cet1
D
1) Pol II
2) Pol II-CE
3) Pol II-CE-
RNA23-5'ppp
4) Pol IIp-CE
5) Pol IIp
Anti S5p
Rpb1

Figure S2

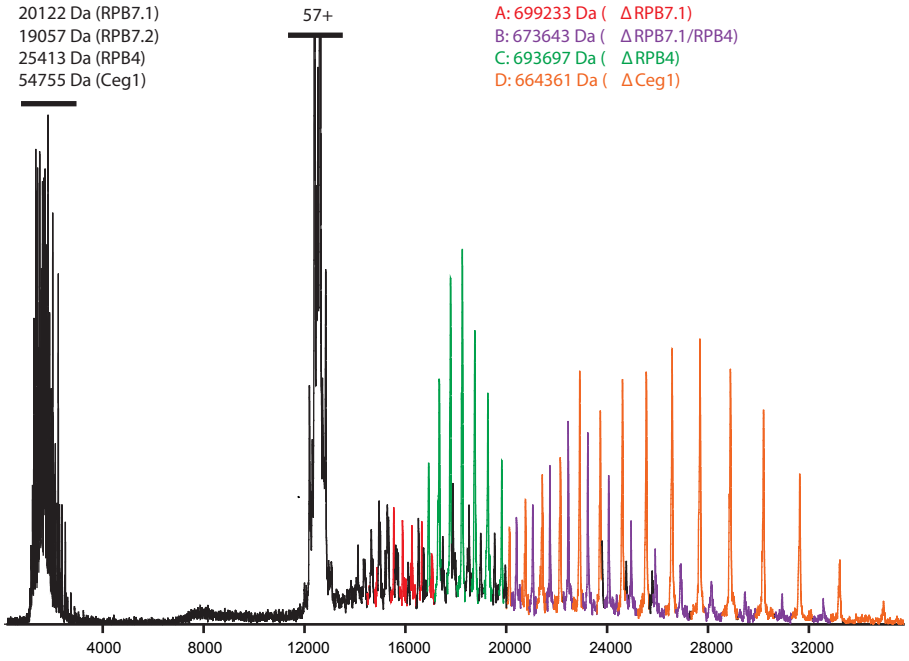

20122 Da (RPB7.1)
19057 Da (RPB7.2)
25413 Da (RPB4)
54755 Da (Ceg1)
57+
A: 699233 Da (ΔRPB7.1)
B: 673643 Da (ΔRPB7.1/RPB4)
C: 693697 Da (ΔRPB4)
D: 664361 Da (ΔCeg1)
4000
8000
12000
16000
20000
24000
28000
32000

Figure S3

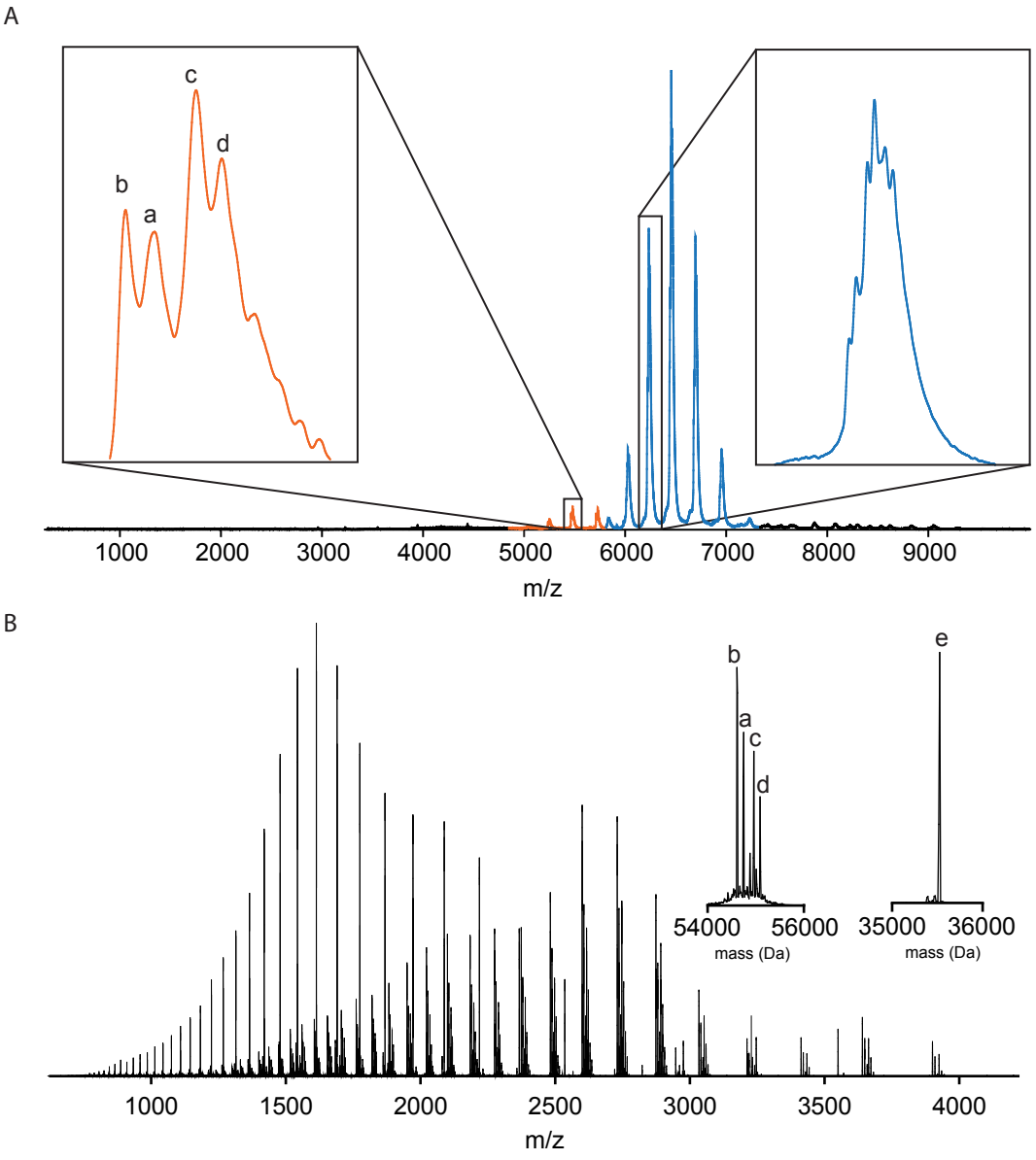
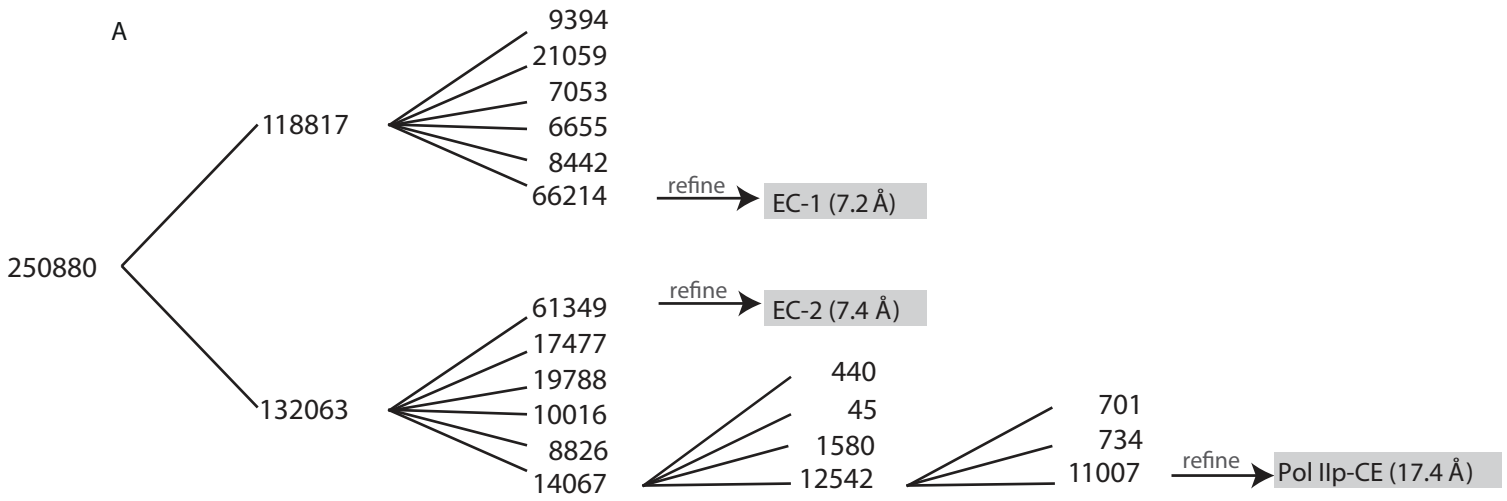
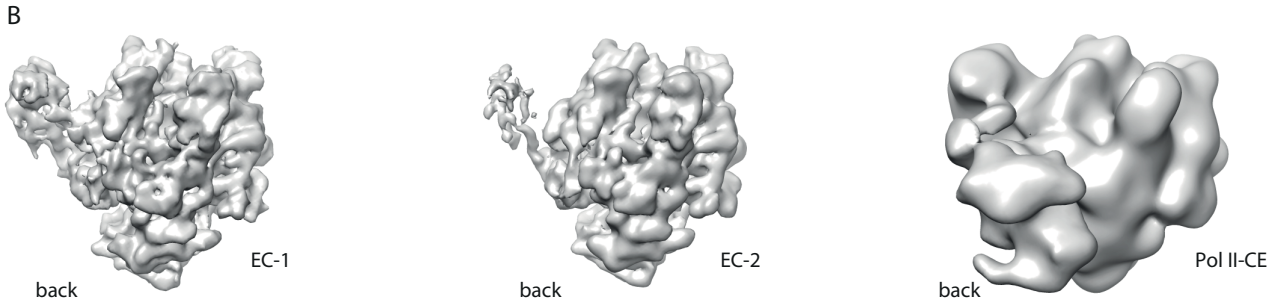
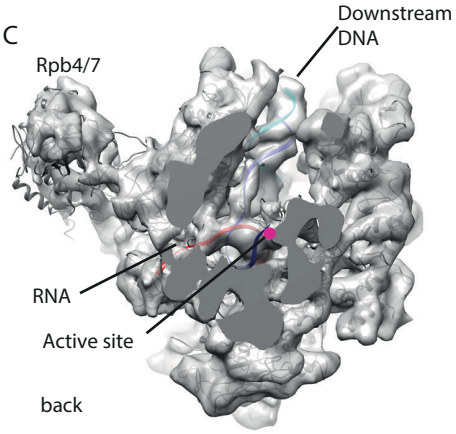
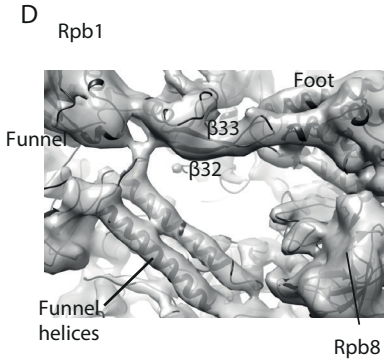

A
c
d
b
a
1000
2000
3000
4000
5000
6000
7000
8000
9000
m/z
B
b
a
c
d
e
54000
56000
mass (Da)
35000
36000
mass (Da)
1000
1500
2000
2500
3000
3500
4000
m/z

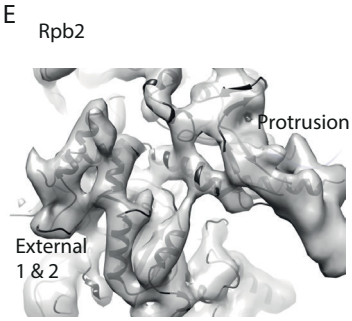
Figure S4

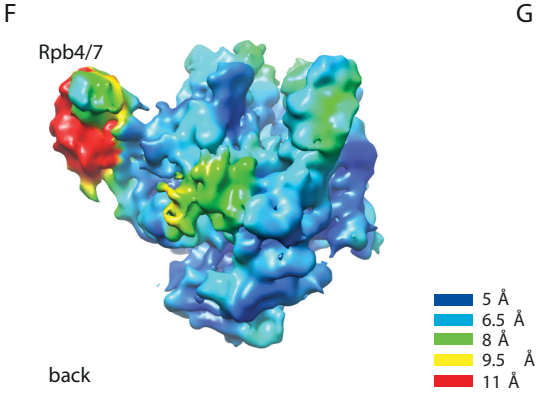

A
250880
118817
9394
21059
7053
6655
8442
66214
refine
EC-1 (7.2 Å)
132063
61349
17477
19788
10016
8826
14067
refine
EC-2 (7.4 Å)
440
45
1580
12542
701
734
11007
refine
Pol IIp-CE (17.4 Å)


B
EC-1
back
EC-2
back
Pol II-CE
back


C
Downstream DNA
Rpb4/7
RNA
Active site
back


D
Rpb1
Foot
β33
Funnel
β32
Funnel helices
Rpb8
side


E
Rpb2
Protrusion
External 1 & 2


F
Rpb4/7
back
5 Å
6.5 Å
8 Å
9.5 Å
11 Å

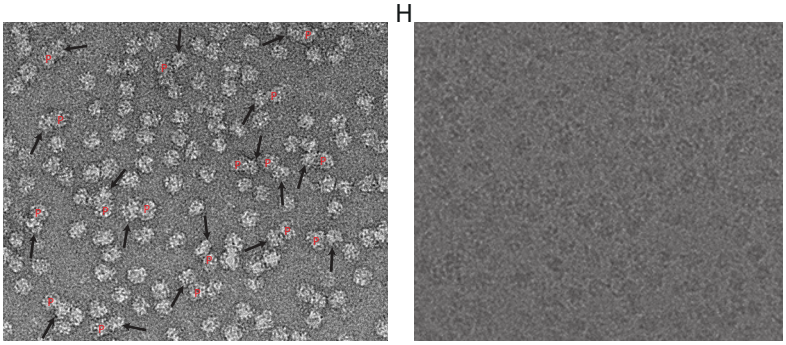

G
H

Figure S5

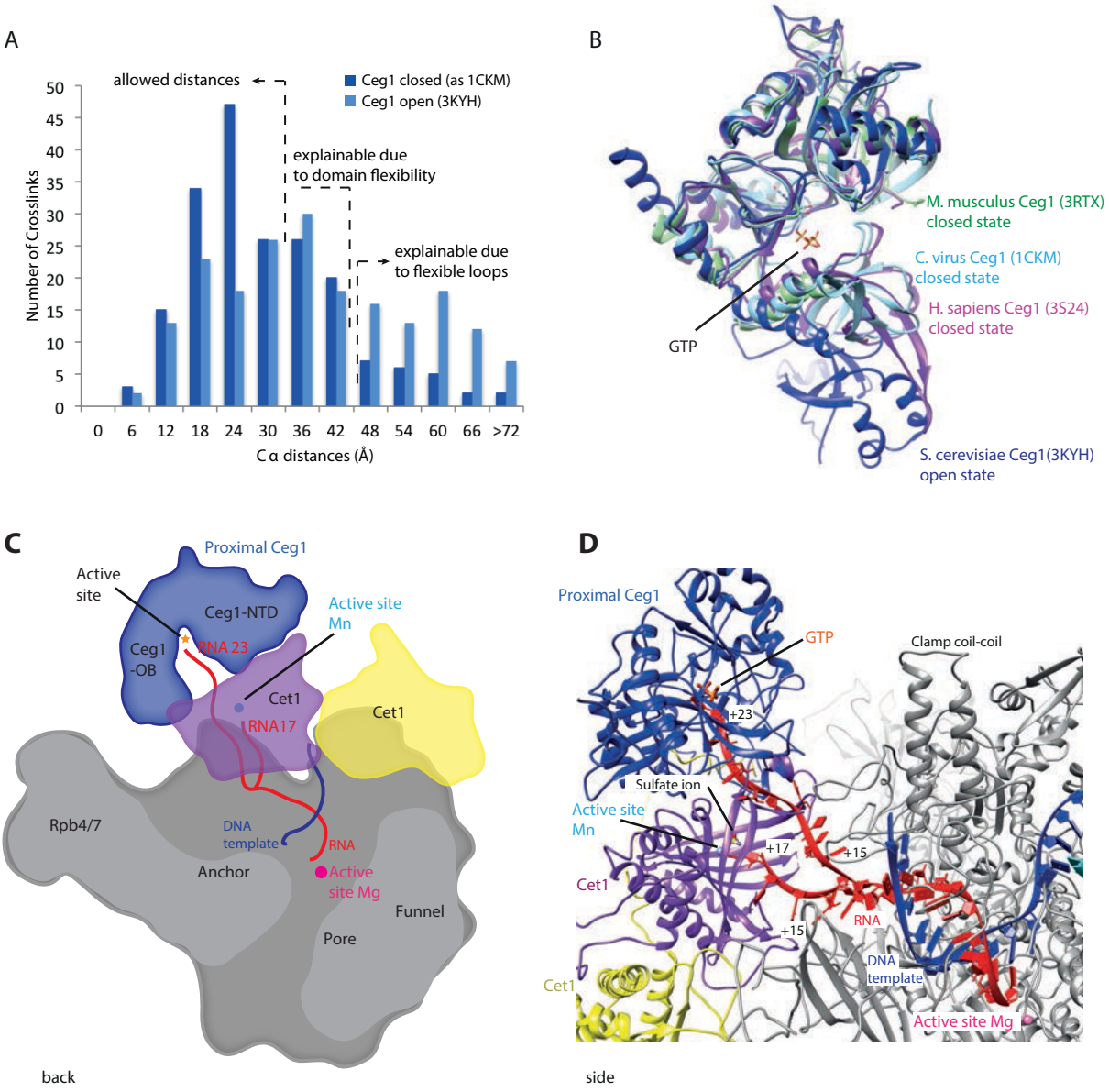

A
allowed distances
Ceg1 closed (as 1CKM)
Ceg1 open (3KYH)
explainable due to domain flexibility
explainable due to flexible loops
Number of Crosslinks
Cα distances (Å)
B
M. musculus Ceg1 (3RTX) closed state
C. virus Ceg1 (1CKM) closed state
H. sapiens Ceg1 (3S24) closed state
GTP
S. cerevisiae Ceg1(3KYH) open state
C
Proximal Ceg1
Active site
Ceg1-NTD
Active site Mn
Ceg1 -OB
RNA 23
Cet1
RNA17
Cet1
Rpb4/7
DNA template
RNA
Anchor
Active site Mg
Funnel
Pore
back
D
Proximal Ceg1
GTP
Clamp coil-coil
+23
Sulfate ion
Active site Mn
+17
+15
Cet1
RNA
+15
DNA template
Cet1
Active site Mg
side

Table S1: Masses observed in native MS analysis of the transcribing Pol IIp-CE complex and its components. Related to figure 2.

Structure	Intact	MSMS(1)	Expected
Pol IIp	520932	518783	513521.56 (2)
Pol IIp ΔRPB4/RPB7	474823		474310.68
[Cet1-Ceg1][Cet1-Ceg1]	180749		180503.64
Pol IIp + Scaffold	538978	538667	538641.5
Pol IIp + Scaffold + [Cet1-Ceg1][Cet1-Ceg1]	719812	719158	719145.14
Pol IIp + Scaffold + [Cet1-Ceg1][Cet1]	664698		664399.74
Scaffold (DNA1+ DNA2+RNA triphosphate)			19858.5

Component	Denatured	MSMS(1)	Expected
Cet1		35505	35506.42
Ceg1		54755	54745.4
DNA 1	4287		4286.9
DNA 2	7911		7911.2
RNA triphosphate	7596		7595.5
RNA diphosphate	7516		7515.5
RNA capped	7861		7860.71
RPB1			191612.11
RPB2			138751.86
RPB3			35297.67
RPB4	25446.8	25413	25414.21
RPB5	25120		25079.11
RPB6	17909		17909.89
RPB7.1	20115.7	20122	
RPB7.2	19051.7		19058.11
RPB8			16501.11
RPB9	14197		14288.2
RPB10	8275		8277.75
RPB11	13656		13615.54
RPB12	7624		7715.97

(1) Average of the summed fragments
(2) Pol II expected mass is without PTMs which explains the large deviation. For the other expected masses the Pol II MSMS mass is used.

Supplemental Figures

Figure S1. Pol IIp-CE complex formation depends on CTD phosphorylation and the length of 5' triphosphate-containing RNA. Related to figure 1.

A. MAP kinase preferentially phosphorylates Ser5 residues of the CTD. For each Western blot the indicated anti-phospho CTD primary antibody was used after phosphorylation of Pol II with P42 MAP Kinase. A control Western blot with anti-CTD antibody confirmed that equal amounts of Pol IIp were loaded per lane.

B. RNA substrates of 17 nt or more in length enhanced CE binding to Pol IIp.

C. CE binding strongly depends on the presence of the RNA 5'-triphosphate end in the non-phosphorylated Pol II elongation complex (lane 3) as well as in the phosphorylated Pol IIp elongation complex (main Fig. 1D).

D. Western blot analysis of Pol II samples used in panel C confirms the phosphorylated and non-phosphorylated state of the Pol II-CTD at Ser5 residues.

Figure S2. Tandem MS of the transcribing Pol IIp-CE complex. Related to figure 2.

Selection and collision induced dissociation of the 57+ charge state of the 719 kDa species in the spectrum of Pol IIp + DNA/RNA23 + CE hetero-tetramer Cet1-Cet1-Ceg1-Ceg1. The low m/z region shows unfolded subunits belonging to Pol IIp or CE. The high m/z region shows their corresponding charge reduced species

Figure S3. MS analysis of CE with partial GTP occupancy in Ceg1. Related to figure 3.

A. Full range native mass spectrum of CE containing hetero-tetrameric (blue) and hetero-trimeric (orange) complexes of Ceg1 and Cet1. The insets show enlarged peaks of both forms of the enzyme revealing heterogeneity that is a result of initiator methionine loss and covalent GMP binding of Ceg1. For the left inset: a) full Ceg1 chain; b) Ceg1 – initiator methionine; c) Ceg1 – initiator methionine + GMP; d) Ceg1 + GMP. For the right inset, some peaks can be assigned, but the presence of two Ceg1 copies causes a larger amount of proteoform combinations.

B. Mass spectrum of denatured CE. The inset shows two regions of the spectrum after deconvolution. On the left, the Ceg1 monomer shows the same distribution of masses as

the native form of the heterotrimeric CE (orange inset). On the right, the Cet1 monomer (e) is mainly present as the full chain form including the initiator methionine.

Figure S4. EM classification scheme and reconstructions. Related to figure 4.

A. Hierarchical classification led to cryo-EM 3D reconstructions of two Pol IIp-DNA/RNA elongation complexes (EC-1, EC-2) and the transcribing Pol IIp-CE complex. The number of particles in the initial data and in each class is given. The first classification step is a supervised classification (Experimental Procedures), all other steps are conventional RELION 3D classifications. Final populations for EC-1, EC-2 and Pol IIp-CE were refined in RELION.

B. Maps of the refined populations EC-1 (7.2 Å), EC-2 (7.4 Å) and Pol IIp-CE (17.4 Å) in back view.

C. EM reconstruction of the free Pol IIp elongation complex (EC-1) at 7.2 Å resolution. A cut-away view reveals RNA density that reaches from the active site to the surface of the polymerase. Downstream DNA density reveals major and minor grooves. The RNA is shown in red, template DNA in blue, and non-template DNA in cyan. The active site is indicated by a magenta dot.

D. Detailed view of Pol II domains of Rpb1 from the cryo-EM density map of the Pol IIp elongation complex (EC-1). Domain names are indicated.

E. Detailed view of Pol II domains of Rpb2 from the cryo-EM density map of the Pol IIp elongation complex (EC-1). Domain names are indicated.

F. Local resolution map for the Pol IIp elongation complex (EC-1). The resolution ranges from <6 Å around the active center to ~8Å on the surface of Pol II. The Rpb4/7 stalk is more flexible, with resolutions between 8 and <11 Å.

G. Negative stain micrograph of transcribing Pol IIp-CE complexes. Polymerases are indicated with a ´P´. Additional density on a subset of particles corresponds to CE and is indicated by arrows.

H. Cryo-EM micrograph of transcribing Pol IIp-CE complexes.

Figure S5. Ceg1 flexibility and RNA transfer between CE active sites. Related to figure 5.

A. The closed Ceg1 conformation explains more of the observed protein crosslinks. Shown are Cα distance distributions for crosslinks of *S. cerevisiae* Ceg1 in the open conformation (as in the 3KYH structure, (Gu et al., 2010)) and in the closed conformation, after superposition of the 3KYH structure with the viral homolog (PDB code 1CKM (Hakansson et al., 1997)).

B. Superposition of structures from *S. cerevisiae* Ceg1 (3KYH) with human (3S24, (Chu et al., 2011)), murine (3RTX, (Ghosh et al., 2011)) and viral (1CKM, (Hakansson et al., 1997)) guanylyltransferases in the closed conformation. GTP from the *C. virus* structure is shown.

C. Schematic cut-away view of the transcribing Pol IIp-CE complex with RNA modeled to reach either the active site of Cet1 or the active site of Ceg1.

D. Side view of the transcribing Pol IIp-CE complex model in cartoon representation. A modeled RNA of 23 nt can reach both the Cet1 and the Ceg1 active centers. A 17 nt RNA is sufficient to reach the Cet1 active center. The RNA register (+15, +17 and +23) is indicated. The active site magnesium ion (Mg) is indicated. A sulfate ion mimicking a phosphate in the Cet1 active center is from the Cet1 structure with code 1D8H (Lima et al., 1999). In the Ceg1 active center GTP from the viral guanylyltransferase structure with code 1CKM (Hakansson et al., 1997) is shown.

Supplemental Tables

Table S1: Masses observed in native MS analysis of the transcribing Pol IIp-CE complex and its components. Related to Figure 2.

Table S2: High-confidence lysine-lysine crosslinks of the transcribing Pol IIp-CE complex and its components. Related to Figure 5.