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

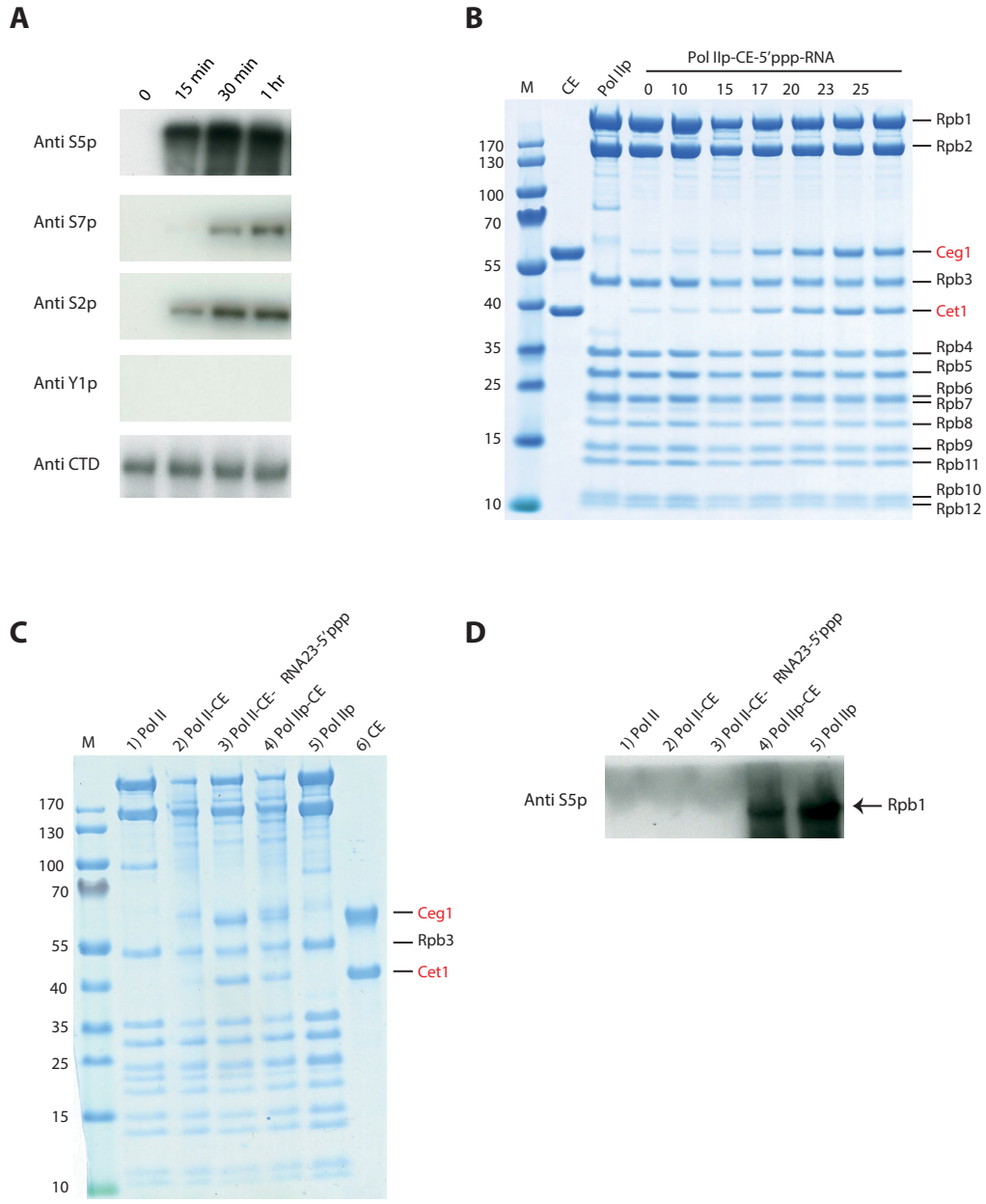

A
0
15 min
30 min
1 hr
Anti S5p
Anti S7p
Anti S2p
Anti Y1p
Anti CTD
B
M
CE
Pol IIp
Pol IIp-CE-5'ppp-RNA
0
10
15
17
20
23
25
170
130
100
70
55
40
35
25
15
10
Rpb1
Rpb2
Ceg1
Rpb3
Cet1
Rpb4
Rpb5
Rpb6
Rpb7
Rpb8
Rpb9
Rpb11
Rpb10
Rpb12
C
M
1) Pol II
2) Pol II-CE
3) Pol II-CE-
RNA23-5'ppp
4) Pol IIp-CE
5) Pol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

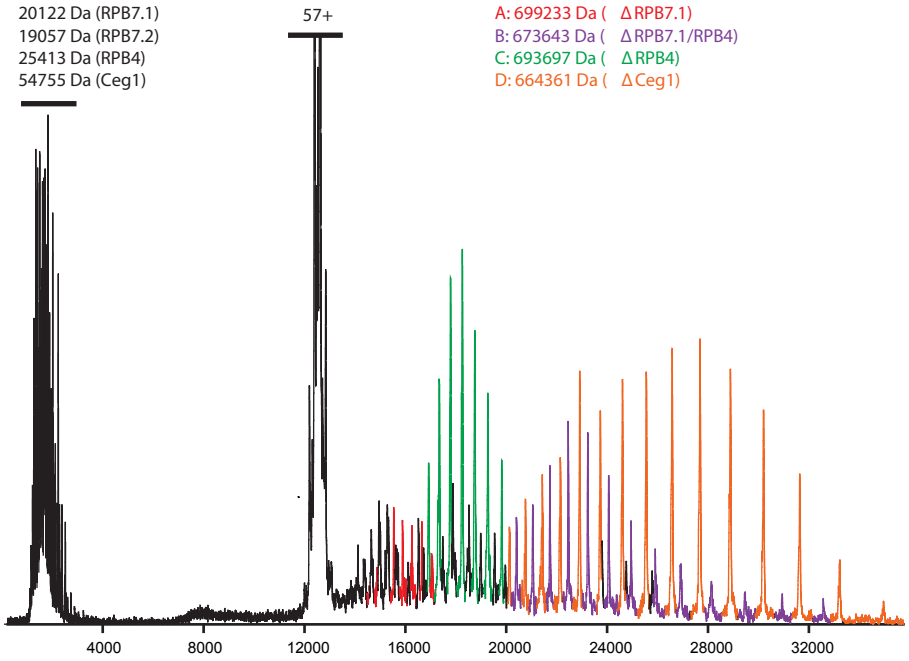

20122 Da (RPB7.1)
19057 Da (RPB7.2)
25413 Da (RPB4)
54755 Da (Ceg1)
57+
A: 699233 Da (ΔRPB7.1)
B: 673643 Da (ΔRPB7.1/RPB4)
C: 693697 Da (ΔRPB4)
D: 664361 Da (ΔCeg1)
4000
8000
12000
16000
20000
24000
28000
32000

Figure S3

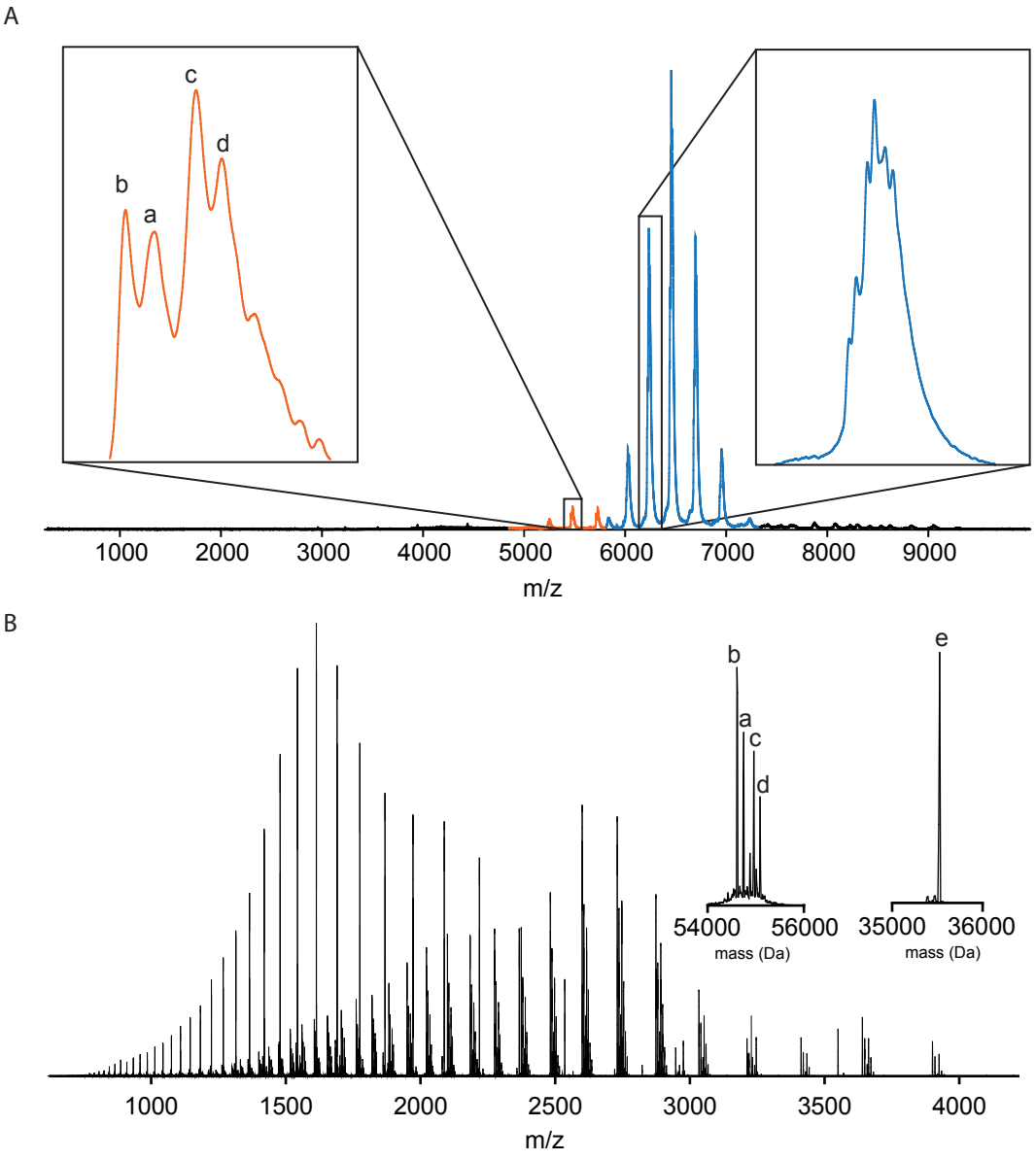
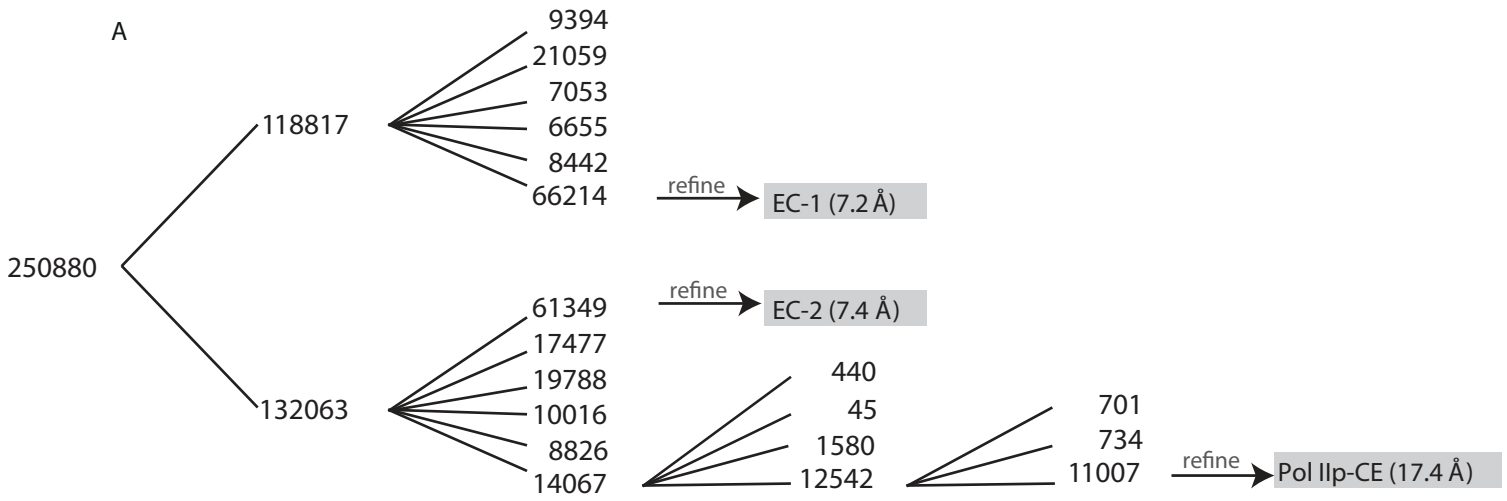
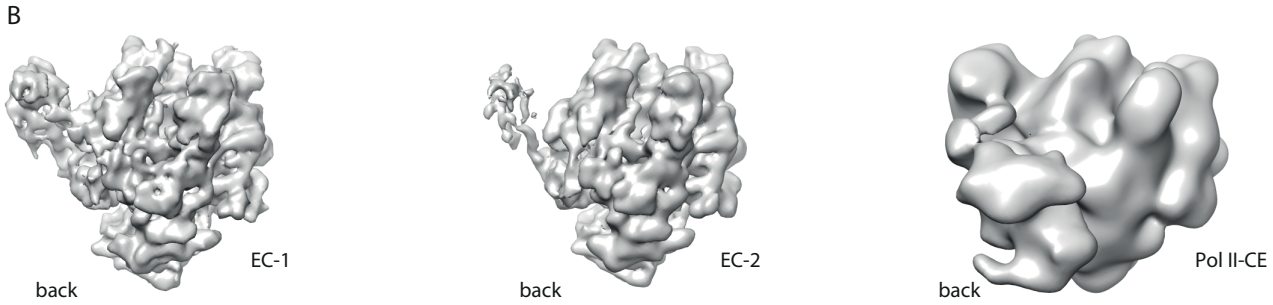
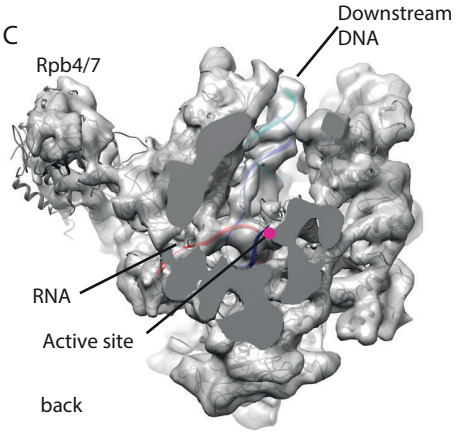
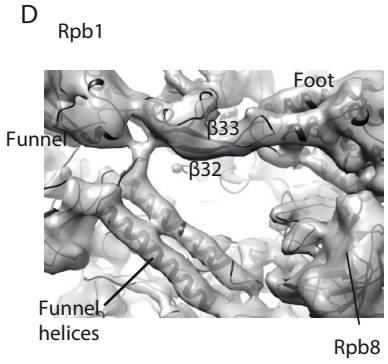

A
c
d
b
a
1000
2000
3000
4000
5000
6000
7000
8000
9000
m/z
B
b
a
c
d
e
54000
56000
mass (Da)
35000
36000
mass (Da)
1000
1500
2000
2500
3000
3500
4000
m/z

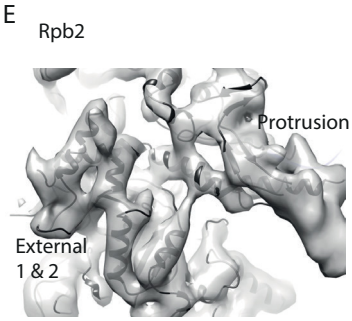
Figure S4

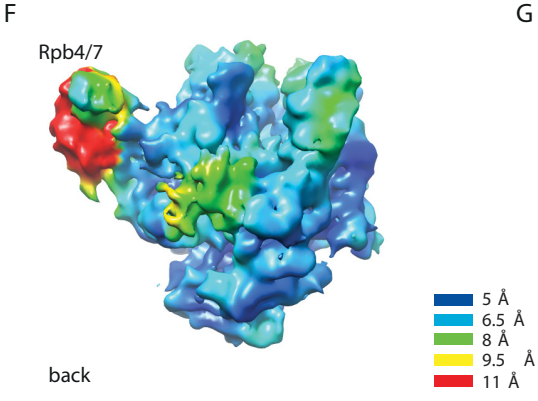

A
250880
118817
9394
21059
7053
6655
8442
66214
refine
EC-1 (7.2 Å)
132063
61349
17477
19788
10016
8826
14067
refine
EC-2 (7.4 Å)
440
45
1580
12542
701
734
11007
refine
Pol IIp-CE (17.4 Å)

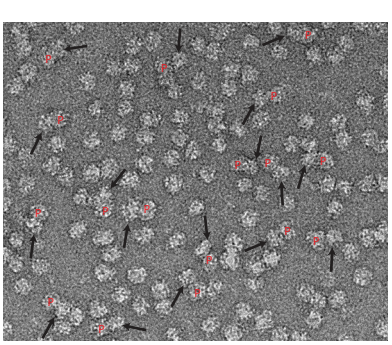

B
EC-1
back
EC-2
back
Pol II-CE
back


C
Downstream DNA
Rpb4/7
RNA
Active site
back


D
Rpb1
Foot
β33
β32
Funnel
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

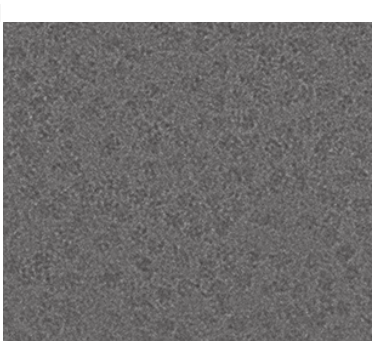

H

Figure S5

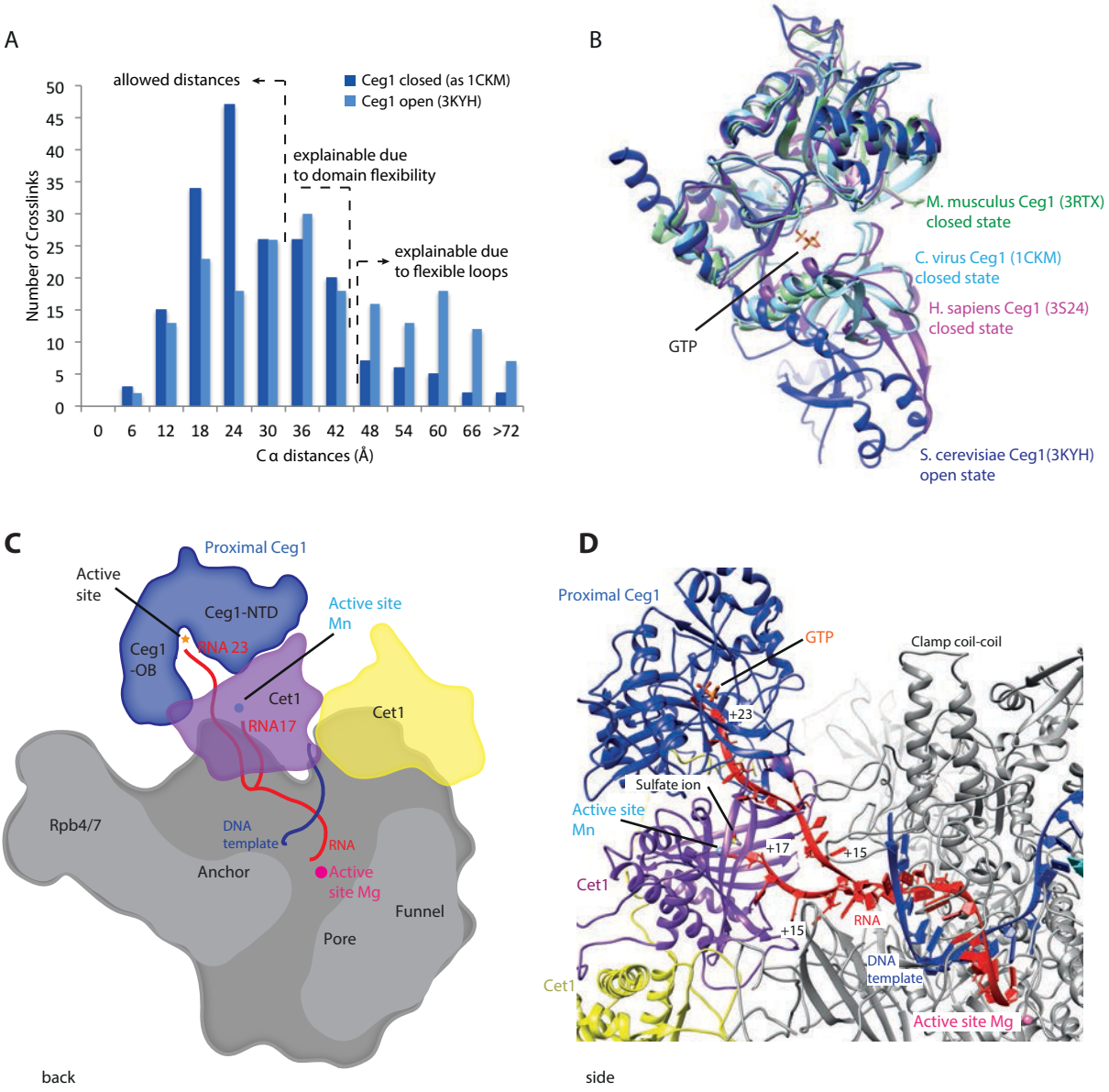

A
allowed distances
Ceg1 closed (as 1CKM)
Ceg1 open (3KYH)
explainable due to domain flexibility
explainable due to flexible loops
Number of Crosslinks
Cα distances (Å)
B
M. musculus Ceg1 (3RTX) closed state
C. virus Ceg1 (1CKM) closed state
H. sapiens Ceg1 (3S24) closed state
GTP
S. cerevisiae Ceg1(3KYH) 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.