Structure of transcribing RNA polymerase II-nucleosome complex

Lucas Farnung[1*], Seychelle M. Vos[1*], and Patrick Cramer[1]

[1]Max Planck Institute for Biophysical Chemistry, Department of Molecular Biology, Am Fassberg 11, 37077 Göttingen, Germany. *These authors contributed equally.

Transcription of eukaryotic protein-coding genes requires passage of RNA polymerase II (Pol II) through chromatin. Pol II passage is impaired by nucleosomes and requires elongation factors that help Pol II to efficiently overcome the nucleosomal barrier[1-4]. How the Pol II machinery transcribes through a nucleosome remains unclear because structural studies have been limited to Pol II elongation complexes formed on DNA templates lacking nucleosomes[5]. Here we report the cryo-electron microscopy (cryo-EM) structure of transcribing Pol II from the yeast *Saccharomyces cerevisiae* engaged with a downstream nucleosome core particle (NCP) at an overall resolution of 4.4 Å with resolutions ranging from 4-6 Å in Pol II and 6-8 Å in the NCP. Pol II and the NCP adopt a defined orientation that could not be predicted from modelling. Pol II contacts DNA of the incoming NCP on both sides of the nucleosomal dyad with its domains 'clamp head' and 'lobe'. Comparison of the Pol II-NCP structure to known structures of Pol II complexes reveals that the elongation factors TFIIS, DSIF, NELF, PAF1 complex, and SPT6 can be accommodated on the Pol II surface in the presence of the oriented nucleosome. Further structural comparisons show that the chromatin remodelling enzyme Chd1, which is also required for efficient Pol II passage[6,7], could bind the oriented nucleosome with its motor domain. The DNA-binding region of Chd1 must however be released from DNA when Pol II approaches the nucleosome, and based on published data[8,9] this is predicted to stimulate Chd1 activity and to facilitate Pol II passage. Our results provide a starting point for a mechanistic analysis of chromatin transcription.

To investigate how RNA polymerase II (Pol II) passes through chromatin, we determined the structure of Pol II transcribing into a nucleosome core particle (NCP). We initially attempted to obtain structures of elongating Pol II located at different positions within the NCP. To this end we placed transcription bubbles at different NCP locations, but these

31 efforts were unsuccessful. We therefore reconstituted Pol II elongation complexes on pre-
32 assembled, extended NCPs with different lengths of extranucleosomal DNA. The
33 extended NCP that enabled structure determination comprised 145 base pairs (bp) of
34 DNA with the Widom 601 sequence, 15 bp of extranucleosomal DNA, a 9-nucleotide (nt)
35 3'-DNA overhang, and a 10 nt RNA transcript annealed to the DNA overhang (Extended
36 Data Fig. 1a). Pol II can bind the terminal DNA-RNA hybrid duplex of the extended
37 NCP, resulting in a Pol II-NCP complex.

38 To test whether the reconstituted Pol II-NCP complex could elongate RNA, we
39 incubated the complex with nucleoside triphosphates and transcription factor (TF) IIS,
40 which is known to support Pol II elongation on nucleosomal templates[4]. We observed
41 RNA elongation (Extended Data Fig. 1b), demonstrating that the reconstituted complex
42 was functional. RNA elongation paused at several positions and was almost completely
43 blocked after addition of ~35 nt (Extended Data Fig. 1b). The block to RNA elongation
44 likely corresponds to the known major pause site around superhelical location (SHL) -5
45 of the NCP[3]. When we repeated the RNA elongation assay without histones, we observed
46 extension to the end of the template (not shown), demonstrating that the presence of the
47 NCP impairs RNA extension in this system.

48 To prepare a sample for cryo-electron microscopy (cryo-EM), we incubated
49 reconstituted Pol II-NCP complex with nucleoside triphosphates and subjected the
50 resulting complexes to size exclusion chromatography. TFIIS was initially included but
51 lost during complex purification (Extended Data Fig. 1c, d). The purified sample was
52 analysed by cryo-EM (Methods). Analysis of the obtained cryo-EM single particle data
53 revealed a well-defined Pol II-NCP complex that could be refined to an overall resolution
54 of 4.4 Å. Multi-body refinement resulted in a Pol II body at a resolution of 4.3 Å and a
55 NCP body at 6.9 Å (Extended Data Fig. 2). The resulting reconstruction clearly showed
56 nucleic acid backbones and grooves, well-defined protein secondary structure within Pol
57 II, and the histone octamer fold of the NCP (Extended Data Fig. 3; Supplementary Video
58 1).

59 We next placed crystal structures of the Pol II elongation complex[10] (PDB code
60 3HOV) and an NCP containing Widom 601 DNA[11] (PDB code 3LZ0) into the cryo-EM

61 density. The structures were locally adjusted to fit the density, connecting DNA was
62 modelled, and the stereochemistry was refined (Methods). Pol II was observed in the
63 active, post-translocated state (Extended Data Fig. 3f). However, we did not observe
64 additional states of the complex with Pol II transcribing into the NCP, despite extensive
65 sorting of the particle images. It is likely these states are unstable and did not survive
66 preparation of the cryo-EM grids.

67 The structure reveals transcribing Pol II in front of a downstream nucleosome
68 (Fig. 1). The orientation of the NCP with respect to Pol II is defined, and differs from the
69 predicted orientation. To predict the Pol II-NCP orientation, we positioned structures of
70 Pol II and the NCP on the designed nucleic acid scaffold assuming that the connecting
71 DNA adopts the canonical B-form conformation. In this model, the nucleosomal disc is
72 rotated by ~90° compared to the orientation observed in our experimental structure
73 (Extended Data Fig. 4a). Comparison of the model with our structure also reveals that
74 half a turn of nucleosomal DNA has been detached from the histone octamer surface. In
75 particular, DNA base pairs between SHL -7 and SHL -6.5 are lifted from the octamer by
76 up to 5 Å, and this enables feeding of the DNA into the polymerase active site cleft.

77 These observations indicate that the defined orientation of the NCP results from
78 Pol II-NCP contacts. Indeed, the structure reveals two major contacts between Pol II and
79 the nucleosome (Fig. 2). Two Pol II domains that flank the polymerase cleft contact the
80 NCP around its dyad axis. The clamp head domain in Pol II subunit Rpb1 contacts
81 nucleosomal DNA around SHL +0.5, whereas the lobe domain of Rpb2 forms contacts
82 around SHL -0.5. Pol II-NCP contacts involve the loops β4-β5 (Rpb1 residues 187-194)
83 and α8-α9 (Rpb2 residues 333-345) that protrude from the clamp head and lobe,
84 respectively.

85 At the beginning of genes, Pol II often pauses. The sites of Pol II pausing are
86 often found upstream of the first nucleosome in the promoter-proximal region ('+1
87 nucleosome'). Pausing requires the elongation factors DSIF and NELF, raising the
88 question whether these factors can be accommodated in the presence of the oriented
89 nucleosome. Indeed, superposition of the Pol II-NCP structure onto our recent paused
90 elongation complex structure (PEC)[12] reveals that DSIF and NELF can be accommodated

91 on the Pol II surface without clashes with the NCP (Fig. 3a). The resulting model of
92 paused Pol II located in front of the +1 nucleosome reflects the situation *in vivo*, as
93 follows. In the fruit fly *Drosophila melanogaster*, the +1 nucleosome adopts a defined
94 location 135-145 bps downstream of the transcription start site[13-15]. This places the
95 upstream edge of the +1 nucleosome at ~60-70 bp downstream of the TSS. Provided that
96 ~15 bp of DNA are accommodated in the Pol II cleft, the active site of paused Pol II
97 would be located 45-55 bp downstream of the TSS, in good agreement with known sites
98 for promoter-proximal pausing[15-17].

99 Release of Pol II from pausing and passage of Pol II through nucleosomes
100 requires several protein factors, including TFIIS, the PAF1 complex (PAF), and SPT6[4,18-
101 22]. Superposition with our recent structure of the activated Pol II complex EC*[23] shows
102 that PAF and SPT6 can be accommodated in the presence of the NCP (Fig. 3b). In the
103 resulting model, the C-terminus of the PAF subunit CTR9 is in close proximity to SHL
104 +7 of nucleosomal DNA, explaining the known interaction of CTR9 with DNA[24]. The
105 SPT6 N-terminal region is also located near the nucleosome, and is known to interact
106 with the histone chaperone Spn1/IWS1[25,26], which could bind histones when Pol II
107 progresses and unravels nucleosomal DNA.

108 Nucleosome passage also requires the ATP-dependent chromatin remodelling
109 enzyme Chd1[6,7]. Comparison of the Pol II-NCP structure with our recent structure of the
110 Chd1-NCP complex[27] after superposition of their NCPs revealed that the ATPase motor
111 and the regulatory double chromodomain of Chd1 can be accommodated on the
112 nucleosome surface without clashes with Pol II (Extended Data Fig. 4b). However, the
113 DNA-binding region (DBR) of Chd1 would have to be displaced from its known position
114 on extranucleosomal DNA[27,28] when transcribing Pol II approaches the NCP. Previous
115 studies showed that displacement of the DBR stimulates Chd1 ATPase activity[8,9].
116 Together these observations suggest that when Pol II approaches a Chd1-NCP complex,
117 Pol II displaces the DBR of Chd1, thereby triggering full activation of the remodelling
118 enzyme, which then loosens DNA-histone interactions. It however remains unclear how
119 liberated histones will be transferred from downstream to upstream DNA, which involves
120 the elongation factor FACT[1,29].

121 It was reported that Pol II passage can lead to transient displacement of one of the
122 two H2A-H2B dimers, converting the histone octamer to a hexamer[30]. The observed
123 Pol II-NCP contacts suggest that the orientation of the NCP is maintained during further
124 Pol II transcription. In this case, the proximal H2A-H2B dimer would encounter the
125 clamp head and lobe domains when Pol II progresses by ~4-5 turns of DNA. The H2A-
126 H2B dimer protrudes from the surface of the nucleosomal disc and would therefore clash
127 with the lobe, possibly leading to its displacement. This model of nucleosome passage
128 however raises the question how downstream DNA can be propelled into the active
129 center cleft.

130 In conclusion, our structure of a transcribing Pol II-NCP complex provides a
131 starting point for analysing the mechanisms of chromatin transcription. In the future,
132 structures of Pol II-NCP complexes with elongation factors such as TFIIS, PAF, SPT6,
133 IWS1, CHD1 and FACT should be determined. This should reveal how these factors
134 assist Pol II in nucleosome passage. Ultimately, such studies could also show how the
135 histone octamer is transferred from downstream to upstream DNA, to prevent loss of
136 histones and to maintain histone modifications during transcription. We note that after
137 our work was completed, others reported similar cryo-EM structures of Pol II-NCP
138 complexes using a slightly different experimental strategy[31].

139 **Acknowledgements** We thank past and present members of the Cramer laboratory. We thank C. Burzinski
140 for purification of *S. cerevisiae* Pol II and TFIIS. S.M.V. was supported by an EMBO Long-Term
141 Fellowship (ALTF-725-2014). P.C. was supported by the Deutsche Forschungsgemeinschaft (SFB1064,
142 SFB860), the European Research Council Advanced Investigator Grant TRANSREGULON (grant
143 agreement No. 693023), and the Volkswagen Foundation.

144 **Author Contributions** L.F. designed and carried out all experiments except biochemical experiments that
145 were performed by L.F. and S.M.V.. P.C. designed and supervised research. L.F., S.M.V, and P.C.
146 interpreted the data and wrote the manuscript.

147 **Author Information** Readers are welcome to comment on the online version of the paper.

148 ## REFERENCES

149 1. Belotserkovskaya, R. *et al.* FACT Facilitates Transcription-Dependent
150 Nucleosome Alteration. *Science* **301,** 1090–1093 (2003).
151 2. Teves, S. S., Weber, C. M. & Henikoff, S. Transcribing through the
152 nucleosome. *Trends Biochem Sci* **39,** 577–586 (2014).

153 3. Studitsky, V. M., Clark, D. J. & Felsenfeld, G. Overcoming a nucleosomal
154 barrier to transcription. *Cell* **83,** 19–27 (1995).
155 4. Kim, J., Guermah, M. & Roeder, R. G. The human PAF1 complex acts in
156 chromatin transcription elongation both independently and cooperatively
157 with SII/TFIIS. *Cell* **140,** 491–503 (2010).
158 5. Martinez-Rucobo, F. W. & Cramer, P. Structural basis of transcription
159 elongation. *Biochim. Biophys. Acta* **1829,** 9–19 (2013).
160 6. Kwak, H. & Lis, J. T. Control of transcriptional elongation. *Annu. Rev.*
161 *Genet.* **47,** 483–508 (2013).
162 7. Hauk, G. & Bowman, G. D. Structural insights into regulation and action of
163 SWI2/SNF2 ATPases. *Curr Opin Struct Biol* **21,** 719–727 (2011).
164 8. McKnight, J. N., Jenkins, K. R., Nodelman, I. M., Escobar, T. & Bowman,
165 G. D. Extranucleosomal DNA binding directs nucleosome sliding by Chd1.
166 *Mol Cell Biol* **31,** 4746–4759 (2011).
167 9. Nodelman, I. M. *et al.* Interdomain Communication of the Chd1 Chromatin
168 Remodeler across the DNA Gyres of the Nucleosome. *Mol Cell* **65,** 447–
169 459.e6 (2017).
170 10. Sydow, J. F. *et al.* Structural Basis of Transcription: Mismatch-Specific
171 Fidelity Mechanisms and Paused RNA Polymerase II with Frayed RNA. *Mol*
172 *Cell* **34,** 710–721 (2009).
173 11. Vasudevan, D., Chua, E. Y. D. & Davey, C. A. Crystal structures of
174 nucleosome core particles containing the '601' strong positioning sequence.
175 *J Mol Biol* **403,** 1–10 (2010).
176 12. Vos, S. M., Farnung, L., Urlaub, H. & Cramer, P. Structure of paused
177 transcription complex Pol II-DSIF-NELF. *Nature* **560,** 601–606 (2018).
178 13. Mavrich, T. N. *et al.* Nucleosome organization in the Drosophila genome.
179 *Nature* **453,** 358–362 (2008).
180 14. Gilmour, D. S. & Lis, J. T. RNA polymerase II interacts with the promoter
181 region of the noninduced hsp70 gene in Drosophila melanogaster cells. *Mol*
182 *Cell Biol* **6,** 3984–3989 (1986).
183 15. Zeitlinger, J. *et al.* RNA polymerase stalling at developmental control genes
184 in the Drosophila melanogaster embryo. *Nat Genet* **39,** 1512–1516 (2007).
185 16. Gilchrist, D. A. *et al.* Pausing of RNA Polymerase II Disrupts DNA-
186 Specified Nucleosome Organization to Enable Precise Gene Regulation. *Cell*
187 **143,** 540–551 (2010).
188 17. Nechaev, S. *et al.* Global Analysis of Short RNAs Reveals Widespread
189 Promoter-Proximal Stalling and Arrest of Pol II in Drosophila. *Science* **327,**
190 335–338 (2010).
191 18. Adelman, K. *et al.* Efficient Release from Promoter-Proximal Stall Sites
192 Requires Transcript Cleavage Factor TFIIS. *Mol Cell* **17,** 103–112 (2005).
193 19. Ng, H. H., Robert, F., Young, R. A. & Struhl, K. Targeted Recruitment of
194 Set1 Histone Methylase by Elongating Pol II Provides a Localized Mark and
195 Memory of Recent Transcriptional Activity. *Mol Cell* **11,** 709–719 (2003).
196 20. Yoh, S. M., Cho, H., Pickle, L., Evans, R. M. & Jones, K. A. The Spt6 SH2
197 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and
198 export. *Genes Dev* **21,** 160–174 (2007).

21. Yoh, S. M., Lucas, J. S. & Jones, K. A. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev* **22,** 3422–3434 (2008).
22. Bortvin, A. & Winston, F. Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science* **272,** 1473–1476 (1996).
23. Vos, S. M. *et al.* Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **560,** 607–612 (2018).
24. Musso, M., Bianchi-Scarrà, G. & Van Dyke, M. W. The yeast CDP1 gene encodes a triple-helical DNA-binding protein. *Nucleic Acids Res* **28,** 4090–4096 (2000).
25. 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).
26. Li, S. *et al.* The elongation factor Spn1 is a multi-functional chromatin binding protein. *Nucleic Acids Res* **190,** 351 (2017).
27. Farnung, L., Vos, S. M., Wigge, C. & Cramer, P. Nucleosome-Chd1 structure and implications for chromatin remodelling. *Nature* **550,** 539–542 (2017).
28. Sundaramoorthy, R. *et al.* Structure of the chromatin remodelling enzyme Chd1 bound to a ubiquitinylated nucleosome. *eLife* **7,** (2018).
29. Formosa, T. *et al.* Defects in SPT16 or POB3 (yFACT) in Saccharomyces cerevisiae Cause Dependence on the Hir/Hpc Pathway: Polymerase Passage May Degrade Chromatin Structure. *Genetics* **162,** 1557–1571 (2002).
30. Kireeva, M. L. *et al.* Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. *Mol Cell* **9,** 541–552 (2002).
31. Kujirai, T. *et al.* Structural basis of the nucleosome transition during RNA polymerase II passage. *Science* eaau9904 (2018). doi:10.1126/science.aau9904

FIGURE LEGENDS

Figure 1 | Structure of Pol II-NCP complex. a-c, Cartoon model viewed from the front (**a**), side (**b**), and top (**c**). Pol II subunits Rpb1, Rpb2, Rpb3, Rpb4, Rpb5, Rpb6, Rpb7, Rpb8, Rpb9, Rpb10, Rpb11, Rpb12, template DNA, non-template DNA, RNA, H2A, H2B, H3, and H4 are coloured in silver, sand, ruby, deep purple, slate, cyan, deepblue, forest green, orange, purple, yellow, green, blue, cyan, red, yellow, crimson, light blue, and green, respectively. Colour code used throughout. Histone octamer dyad axis is indicated as a black line with white outline.

237 **Figure 2 | Pol II-NCP contacts.** Rpb1 clamp head and Rpb2 lobe domains contact
238 nucleosomal DNA around the NCP dyad. Clamp head and lobe domains are shown.
239 Loops contacting nucleosomal DNA are indicated.

240 **Figure 3 | Accommodation of elongation factors. a**, Superposition of Pol II-NCP
241 structure with the mammalian paused elongation complex PEC (PDB code 6GML)
242 provides the location of DSIF and NELF. **b**, Superposition of the mammalian activated
243 elongation complex EC* (PDB code 6GMH) provides the location of PAF1 complex
244 (PAF) and SPT6.

245 **METHODS**

246 No statistical methods were used to predetermine sample size. The experiments were
247 not randomized, and the investigators were not blinded to allocation during
248 experiments and outcome assessment.

249 **Protein preparation**

250 *S. cerevisiae* Pol II was essentially purified as described[32]. *Xenopus laevis* histones were
251 cloned, expressed, and purified as described[27,33,34]. *S. cerevisiae* TFIIS with an N-terminal
252 GST tag was cloned, expressed, and purified as described[35].

253 **Preparation of extended NCP**

254 DNA fragments for nucleosome reconstitution were generated by PCR essentially as
255 described[27,36]. A vector containing the Widom 601 sequence was used as a template for
256 PCR. Large-scale PCR reactions were performed with two PCR primers (CGC TGT TTT
257 CGA ATT TAC CCT TTA TGC GCC GGT ATT GAA CCA CGC TTA TGC CCA
258 GCA TCG TTA ATC GAT GTA TAT ATC TGA CAC GTG CCT, reverse: ATC AGA
259 ATC CCG GTG CCG AG) at a typical scale of 25 mL. The reaction was distributed into
260 48-well PCR plates (100 µl per well) and PCR was conducted with the following steps: 1.
261 98 °C for 1 min, 2. 98 °C for 10 s, 3. 72 °C for 45 s, cycle between steps 2 and 3 35
262 times, 4. 72 °C for 10 min, 5. Pause at 5 °C. PCR products were pooled, phenol
263 chlorofom extracted, ethanol precipitated, and resuspended in 1 ml H_2O buffer. The
264 sample was ethanol precipitated, resuspended in 500 µl H_2O buffer, and digested using
265 TspRI (NEB) to generate the DNA 3'-overhang for RNA binding. Again, the digested
266 DNA was phenol chlorofom extracted, ethanol precipitated and resuspended in 1 ml H_2O
267 buffer. The DNA was then applied to a ResourceQ 6 ml (GE Healthcare) and eluted with
268 a gradient from 0–100% TE high salt buffer (10 mM Tris pH 8.0, 1 M NaCl, 1 mM
269 EDTA pH 8.0) over a total volume of 60 mL at a flow rate of 2 mL/min. Peak fractions
270 were analysed on a 1% (v/v) TAE agarose gel and fractions containing the DNA product
271 were pooled, ethanol precipitated, resuspended in 100 µL H_2O and stored at –20 °C.
272 Preparation of extended NCP was performed as described[27]. Quantification of the
273 reconstituted nucleosome was achieved by measuring absorbance at 280 nm. Molar

274 extinction coefficients were determined for protein and nucleic acid components and
275 were summed to yield a molar extinction coefficient for the reconstituted extended NCP.

276 **Reconstitution of Pol II-NCP complex**

277 A RNA with the sequence /56-FAM/UCU CAC UGG A was purchased from Integrated
278 DNA Technologies, resuspended in water, and stored at -80ºC. To reconstitute the Pol II-
279 NCP complex, Pol II, extended NCP and RNA were mixed at a molar ratio of 1:0.5:0.66
280 and incubated for 10 min at 30ºC. Compensation buffer was added after 10 min to a final
281 buffer of 150 mM KCl, 3 mM $MgCl_2$, 20 mM Na·HEPES pH 7.5, 4% (v/v) glycerol,
282 1 mM DTT. For RNA extension, we added TFIIS (1:0.57, relative to Pol II) and NTPs (1
283 mM final concentration) and incubated for 30 min at 30 °C. The reaction was centrifuged
284 (21,000*g*, 4 °C, 10 min), and applied to a Superose 6 Increase 3.2/300 column
285 equilibrated in gel filtration buffer (100 mM NaCl, 3 mM $MgCl_2$, 20 mM Na·HEPES pH
286 7.5, 5% (v/v) glycerol, 1 mM DTT). Peak fractions were analysed by SDS-PAGE and
287 cross-linked with 0.1% (v/v) glutaraldehyde and incubated for 10 min on ice. The cross-
288 linking reaction was quenched for 10 min using a concentration of 2 mM lysine and 8
289 mM aspartate. The sample was transferred to a Slide-A-Lyzer MINI Dialysis Unit 20,000
290 MWCO (Thermo Scientific), and dialysed for 6 h against 600 ml dialysis buffer (100 mM
291 NaCl, 2 mM $MgCl_2$, 20 mM Na·HEPES pH 7.4, 20 mM Tris·HCl pH 7.5, 1 mM DTT).

292 **RNA elongation assays**

293 All concentrations refer to the final concentrations used in 70 µL reactions. Pol II (150
294 nM) was mixed with linear DNA or extended NCP (75 nM) and RNA primer (100 nM) at
295 30ºC. Compensation buffer was then added to achieve final assay conditions of 150 mM
296 KCl, 3 mM $MgCl_2$, 20 mM Na·HEPES pH 7.5, 4% (v/v) glycerol, 1 mM DTT. TFIIS (90
297 nM) was added immediately before addition of NTPs. NTPs were added at a final
298 concentration of 1 mM. 10 µL aliquots were taken at 0, 1, 5, 10, 20, and 30 minutes after
299 NTP addition and were quenched by mixing the sample with 10 µL of Stop Buffer (6.4 M
300 urea, 1x TBE, 50 mM EDTA pH 8.0). Reactions were treated with 4 µg proteinase K for
301 30 minutes at 30ºC. The reactions were ethanol precipitated and resuspended in 5 µL of
302 Stop Buffer and applied to a 7 M urea, 13.5% 19:1 Bis-acrylamide, 1x TBE sequencing

303 gel run at 50W for 35 minutes in 1x TBE. RNA Products were visualized using the 6-
304 FAM label and a Typhoon 9500 FLA Imager (GE Healthcare Life Sciences).

305 **Cryo-EM and image processing**

306 The reconstituted and purified Pol II-NCP complex sample was applied to R2/2 gold
307 grids (Quantifoil). The grids were glow-discharged for 100 s before sample application of
308 2 µl on each side of the grid. The sample was subsequently blotted for 8.5 s and vitrified
309 by plunging into liquid ethane with a Vitrobot Mark IV (FEI Company) operated at 4 °C
310 and 100% humidity. Cryo-EM data were acquired on a Titan Krios transmission electron
311 microscope (FEI/Thermo) operated at 300 keV, equipped with a K2 summit direct
312 detector (Gatan). Automated data acquisition was carried out using FEI EPU software at
313 a nominal magnification of 130,000×. Image stacks of 40 frames were collected in
314 counting mode over 9 s. The dose rate was 5.12 e^- per $Å^2$ per s for a total dose of 46 e^-
315 $Å^{-2}$. A total of 9188 image stacks were collected.

316 Frames were stacked and subsequently processed. Micrographs were CTF and
317 motion corrected using Warp[37]. Image processing was performed with RELION 3.0-beta
318 1[38] , unless noted otherwise. Post-processing of refined models was performed with
319 automatic *B*-factor determination in RELION. Particles were picked using the neural
320 network BoxNet2 of Warp, yielding 1,041,545 particle positions. Particles were extracted
321 with a box size of 320^2 pixel, and normalized. Particles were subdivided into three
322 batches. Using a 30 Å low-pass filtered model from a previously published model
323 (EMDB-3626), we performed iterative rounds of hierarchical 2D and 3D classification
324 with image alignment and activated fast subset option (Extended Data Fig. 2c). 111,895
325 particles in batch 1 and 2 that showed Pol II-bound NCP were subjected to masked
326 classification with the mask encompassing the NCP and global 2D classification. Batch 3
327 also showed Pol II-bound NCP and comprised 44,741 particles that were subjected to a
328 masked 3D classification encompassing the Pol II-NCP complex. Particles from batch 1,
329 2, and 3 with NCP-bound Pol II were merged, 2D classified, and CTF refined[38].

330 The final reconstruction was obtained from a 3D refinement with a mask that
331 encompasses the Pol II-NCP complex. The Pol II-NCP reconstruction was obtained from
332 49,703 particles with an overall resolution of 4.4 Å (gold-standard Fourier shell

333 correlation 0.143 criterion)[39]. The final map was sharpened with a *B*-factor of −182 Å^2.
334 Similarly to previous yeast Pol II cryo-EM structures, we observed an enrichment of
335 certain Pol II orientations, exemplified in the angular distribution plot [35,40], but this did
336 not compromise the resolution in any direction (Extended Data Fig. 2d, f). Local
337 resolution estimates were determined using the built-in RELION tool. The Pol II-NCP
338 complex was subsequently multi-body refined using two bodies encompassing Pol II or
339 NCP with an estimated resolution of 4.3 Å for Pol II and 6.9 Å for NCP[41] (Fourier shell
340 correlation 0.143 criterion).

341 **Model building**

342 Crystal structures of the *S. cerevisiae* Pol II EC (PDB code 3HOV) [10], and *X. laevis*
343 nucleosome with Widom 601 sequence[11] (PDB code 3LZ0), were placed into the density
344 using UCSF Chimera[42]. Extranucleosomal DNA and nucleosomal DNA from SHL −7 to
345 SHL −5 were built using COOT[43]. Secondary structure restraints were applied and the
346 model was real-space refined against the post-processed EM map using PHENIX[44].
347 Figures were generated using PyMol[45] and UCSF Chimera and ChimeraX[42].

348

349 **EXTENDED DATA FIGURE LEGENDS**

350 **Extended Data Figure 1 | Formation of Pol II-NCP complex.**

351 **a**, Schematic of nucleic acid scaffold used for RNA elongation assays and complex
352 reconstitution for cryo-EM analysis. **b**, RNA elongation assays performed with Pol II
353 (150 nM), NCP or linear DNA (75 nM), TFIIS (90 nM) with a FAM labelled 10mer RNA
354 primer. Reactions were quenched at various time points. **c**, Formation of the Pol II-NCP
355 complex on a Superose 6 Increase 3.2/30 size exclusion chromatography column. Blue
356 and red curve shows absorption at 280 nm and 260 nm milli absorption units,
357 respectively. **d**, SDS–PAGE of peak fraction from gel filtration (c) used for cryo-EM grid
358 preparation containing Pol II and histones. The identity of the bands was confirmed by
359 mass spectrometry. All experiments were performed at least three times.

360 **Extended Data Figure 2 | Cryo-EM structure determination**

361 **a**, Representative micrograph, low pass filtered to enhance contrast. **b**, 2D class averages
362 show Pol II in close proximity to a nucleosome. **c**, Angular distribution. Shading from
363 blue to yellow indicates the number of particles at a given orientation. **c**, Sorting and
364 classification tree used to reconstruct the Pol II-NCP complex at 4.4 Å resolution. **e**,
365 Estimation of the resolution. The dark green line indicates the Fourier shell correlation
366 between the half maps of the reconstruction. Resolution is given for the FSC 0.143. **f**,
367 Local resolution estimation. Shading from red to blue indicates the local resolution
368 according to the colour gradient. The multi-body refinement densities are shown.
369 Absolute values are given.

370 **Extended Data Figure 3 | Cryo-EM densities.**

371 **a-c**, Non-sharpened maps after multibody refinement in (a) front, (b) side, and (c) top
372 views. Pol II density is shown in silver, NCP density in light blue. Rpb4-7 is visible at
373 lower contours (not shown). **d**, NCP density from multi-body refinement resolves the
374 histone octamer fold and nucleosomal DNA. **e**, Rpb1 funnel helices on the Pol II surface
375 are well defined. **f**, DNA-RNA hybrid in the Pol II active site adopts a post-translocated
376 conformation with a free site for the nucleoside triphosphate substrate. **g**, Rpb2 hybrid
377 binding region shows Pol II structure is very well resolved.

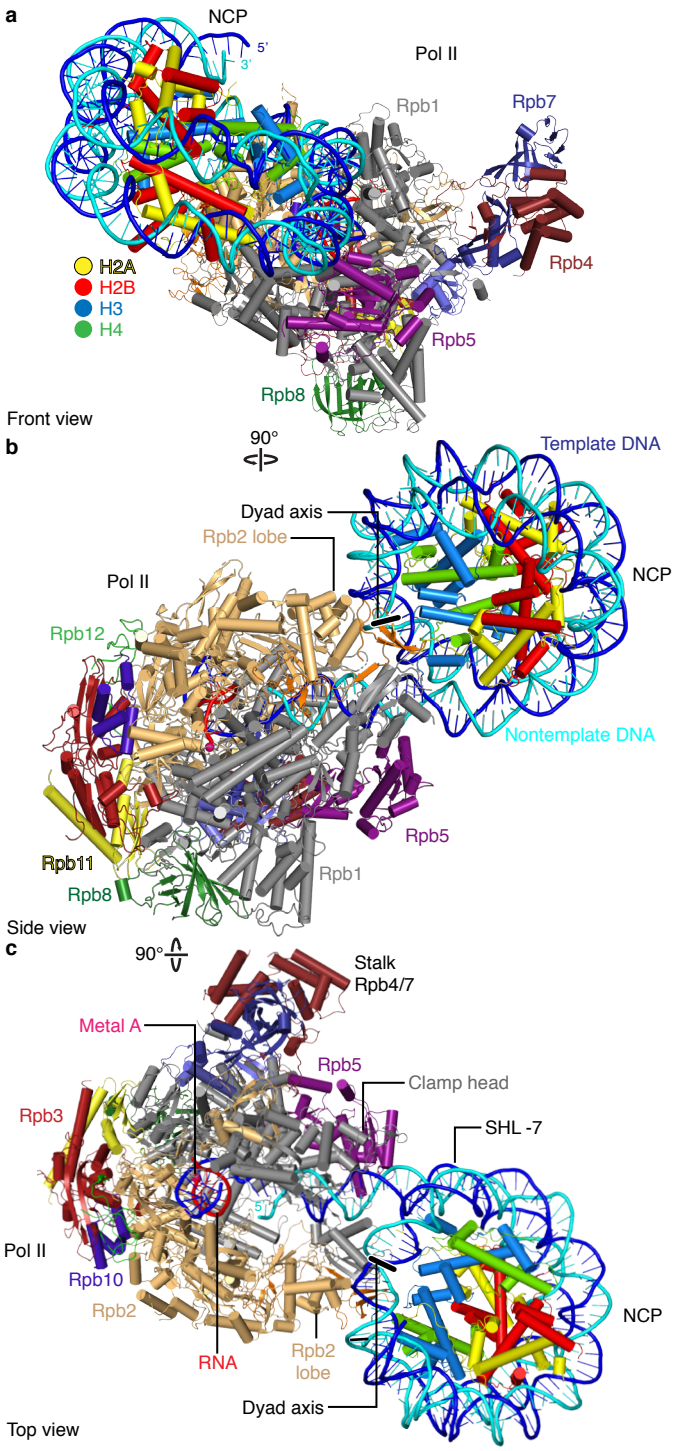
378 **Extended Data Figure 4 | Additional models and structural comparisons**.

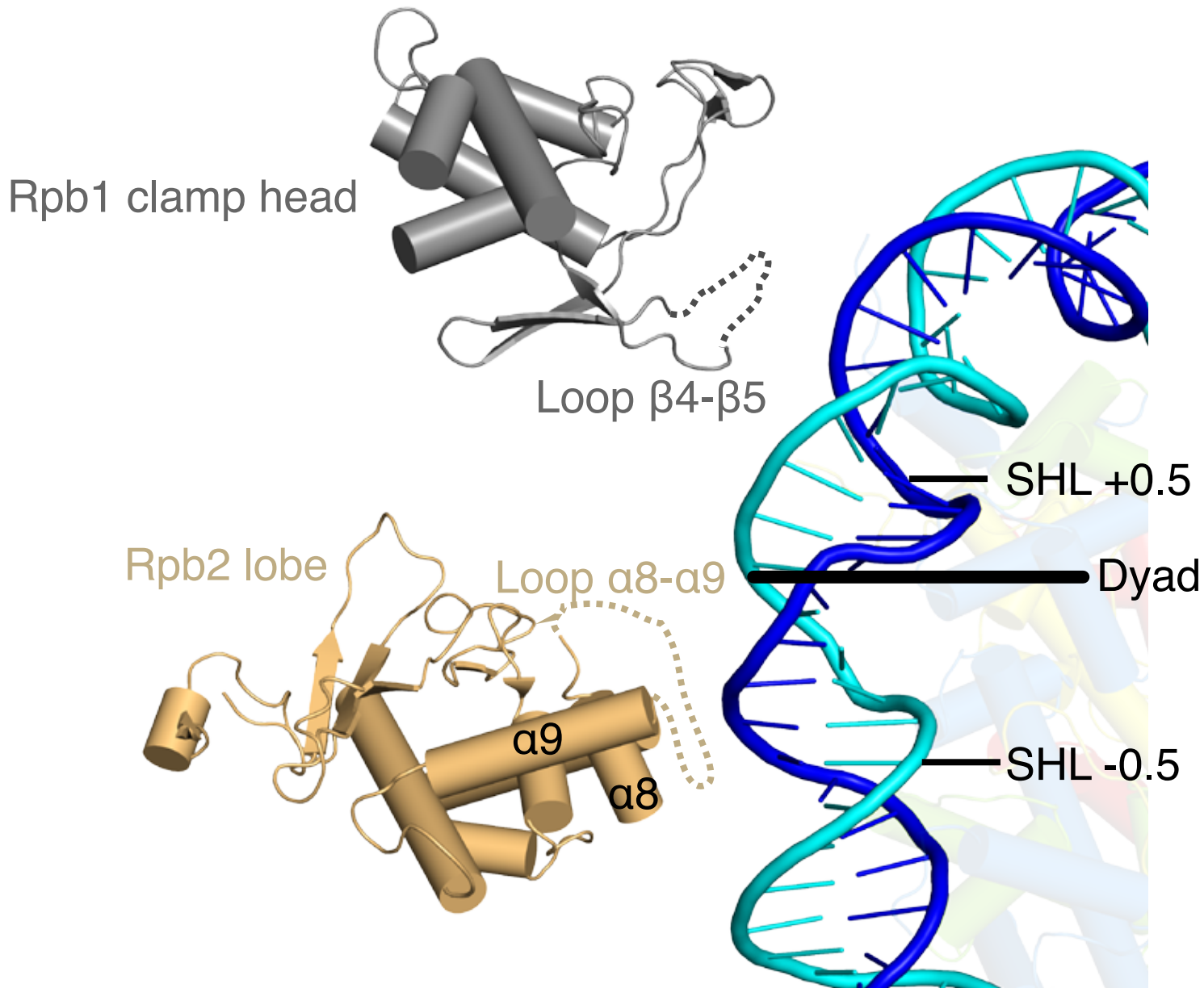
379 **a,** Downstream DNA of a Pol II elongation complex was modelled with a total length of
380 15 base pairs, followed by a mononucleosome (PDB code 3LZ0, shown in grey) based on
381 our nucleic acid scaffold design. The experimental structure is coloured with the same
382 colouring scheme used throughout the publication. Difference in rotational position of the
383 NCP is indicated. **b**, Superposition with the Chd1-NCP complex (PDB code 5O9G) by
384 aligning the NCPs reveals that the DNA-binding region (DBR) of Chd1 must be
385 displaced from downstream DNA when Pol II approaches the NCP. Chd1 double
386 chromodomain, ATPase lobe 1, ATPase lobe 2, and DNA-binding region are coloured in
387 purple, orange, sea green, and pink, respectively.

388 **Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics**
389 **Supplementary Video 1 | Cryo-EM density and structure of the Pol II-NCP complex**
390

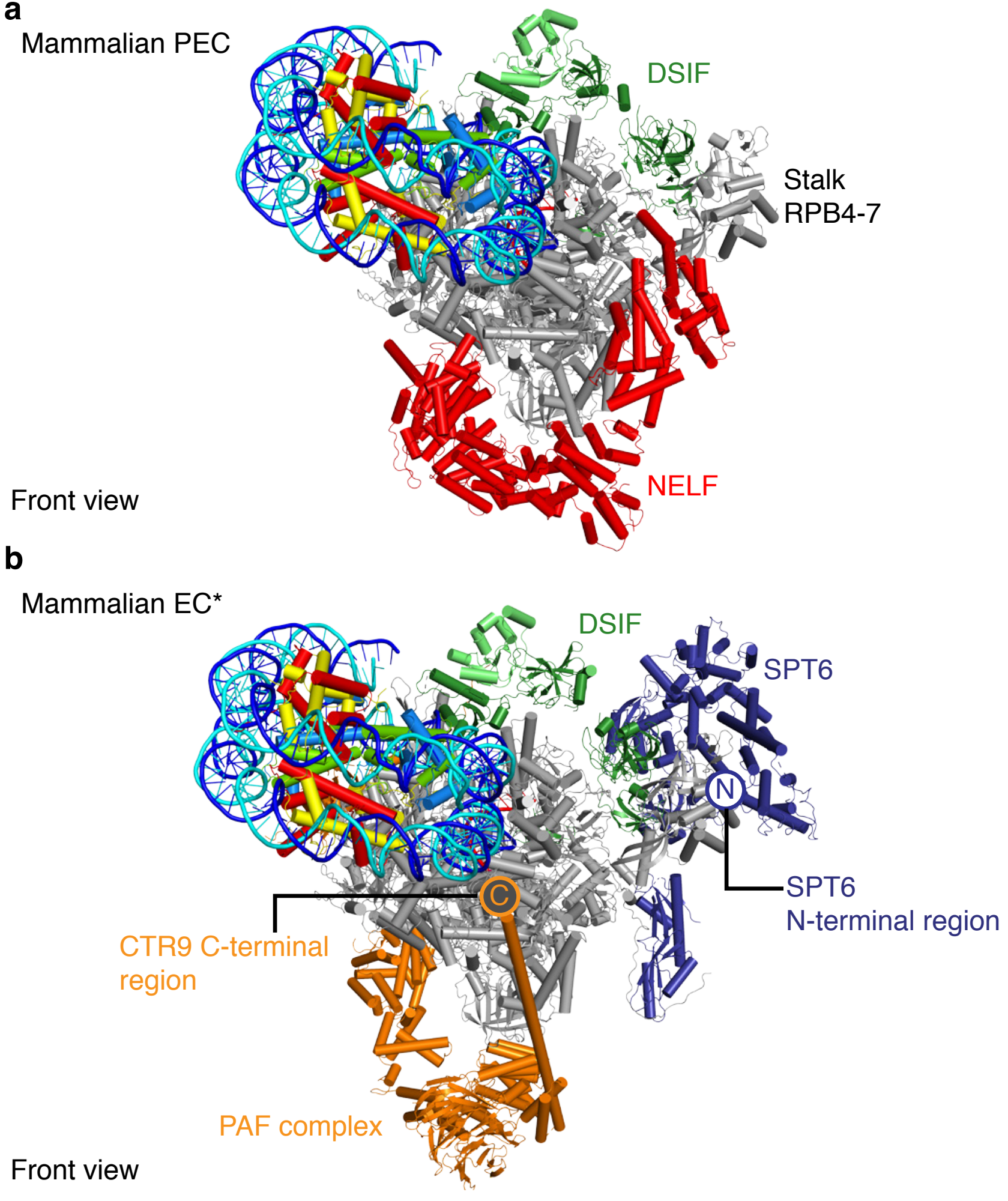
METHODS AND EXTENDED DATA REFERENCES

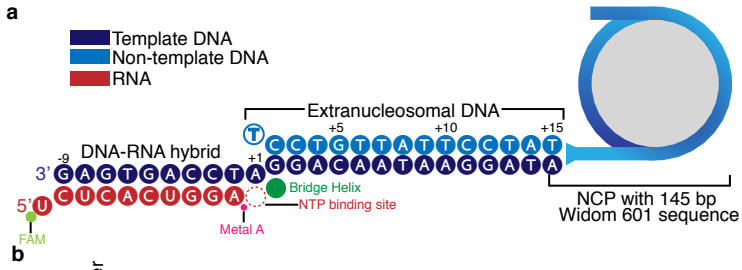
32. Lehmann, E., Brueckner, F. & Cramer, P. Molecular basis of RNA-dependent RNA polymerase II activity. *Nature* **450,** 445–449 (2007).
33. Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol* **375,** 23–44 (2004).
34. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol Biol* **119,** 1–16 (1999).
35. Xu, Y. *et al.* Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex. *Nat Comms* **8,** 15741 (2017).
36. Maskell, D. P. *et al.* Structural basis for retroviral integration into nucleosomes. *Nature* **523,** 366–369 (2015).
37. Tegunov, D., bioRxiv, P. C.2018. Real-time cryo-EM data pre-processing with Warp. *biorxiv.org* doi:10.1101/338558
38. Zivanov, J. *et al.* RELION-3: new tools for automated high-resolution cryo-EM structure determination. *biorxiv.org* (2018). doi:10.1101/421123
39. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* **333,** 721–745 (2003).
40. Ehara, H. *et al.* Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* **357,** 921–924 (2017).
41. Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7,** 1485 (2018).
42. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25,** 1605–1612 (2004).
43. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 (2010).
44. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221 (2010).
45. PyMOL. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

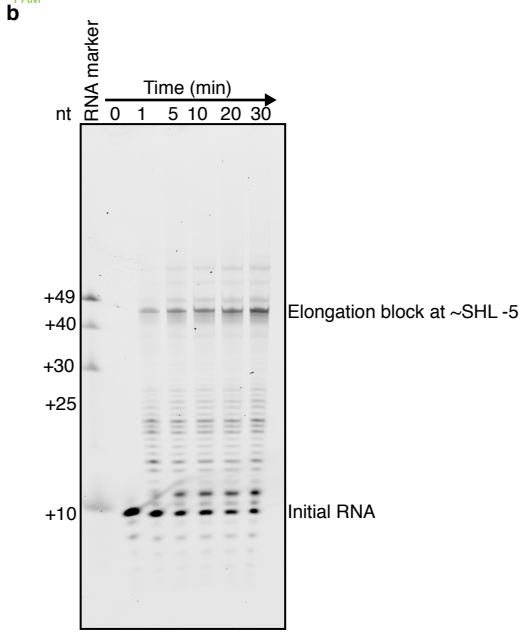

a
NCP
5'
3'
Pol II
Rpb1
Rpb7
Rpb4
H2A
H2B
H3
H4
Rpb5
Rpb8
Front view
90°
b
Template DNA
Dyad axis
Rpb2 lobe
NCP
Pol II
Rpb12
Nontemplate DNA
Rpb5
Rpb11
Rpb1
Rpb8
Side view
c
90°
Stalk
Rpb4/7
Metal A
Rpb5
Clamp head
Rpb3
SHL -7
5'
Pol II
Rpb10
Rpb2
NCP
RNA
Rpb2
lobe
Dyad axis
Top view

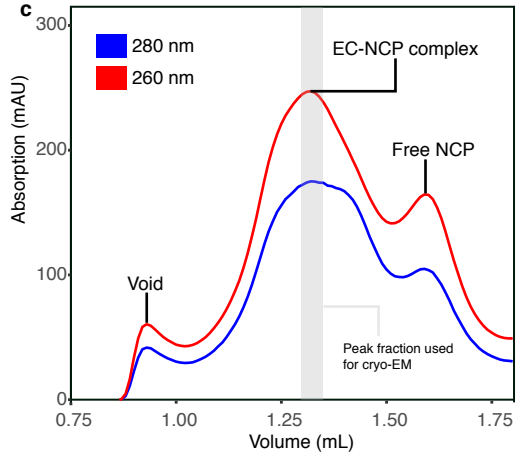

Rpb1 clamp head
Loop β4-β5
SHL +0.5
Rpb2 lobe
Loop α8-α9
Dyad
α9
α8
SHL -0.5

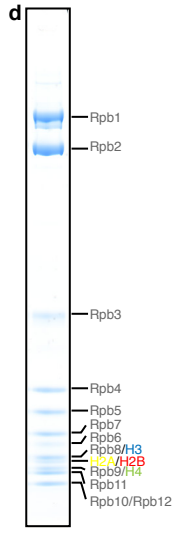
Top view

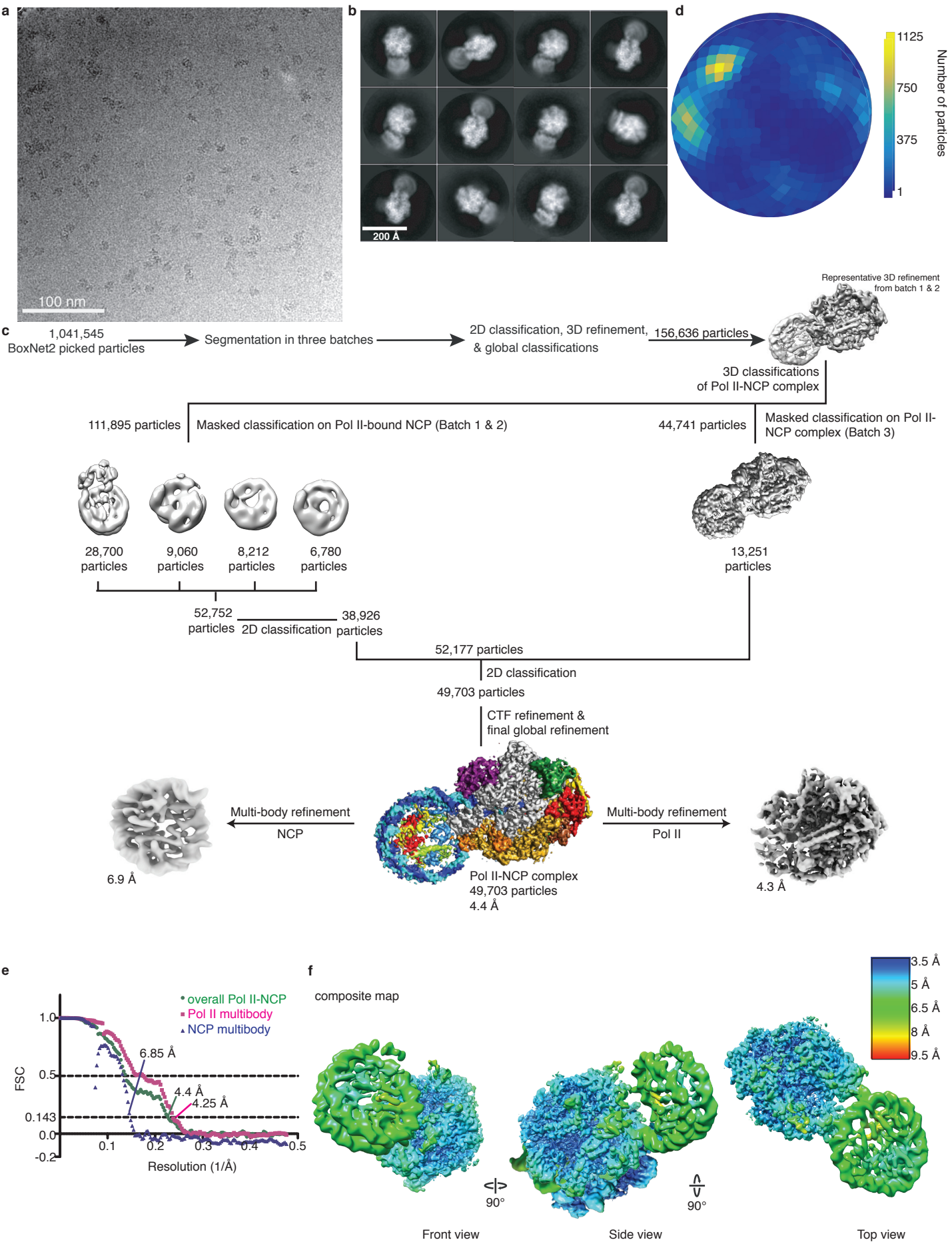

a
Mammalian PEC
DSIF
Stalk
RPB4-7
NELF
Front view
b
Mammalian EC*
DSIF
SPT6
N
SPT6
N-terminal region
C
CTR9 C-terminal
region
PAF complex
Front view

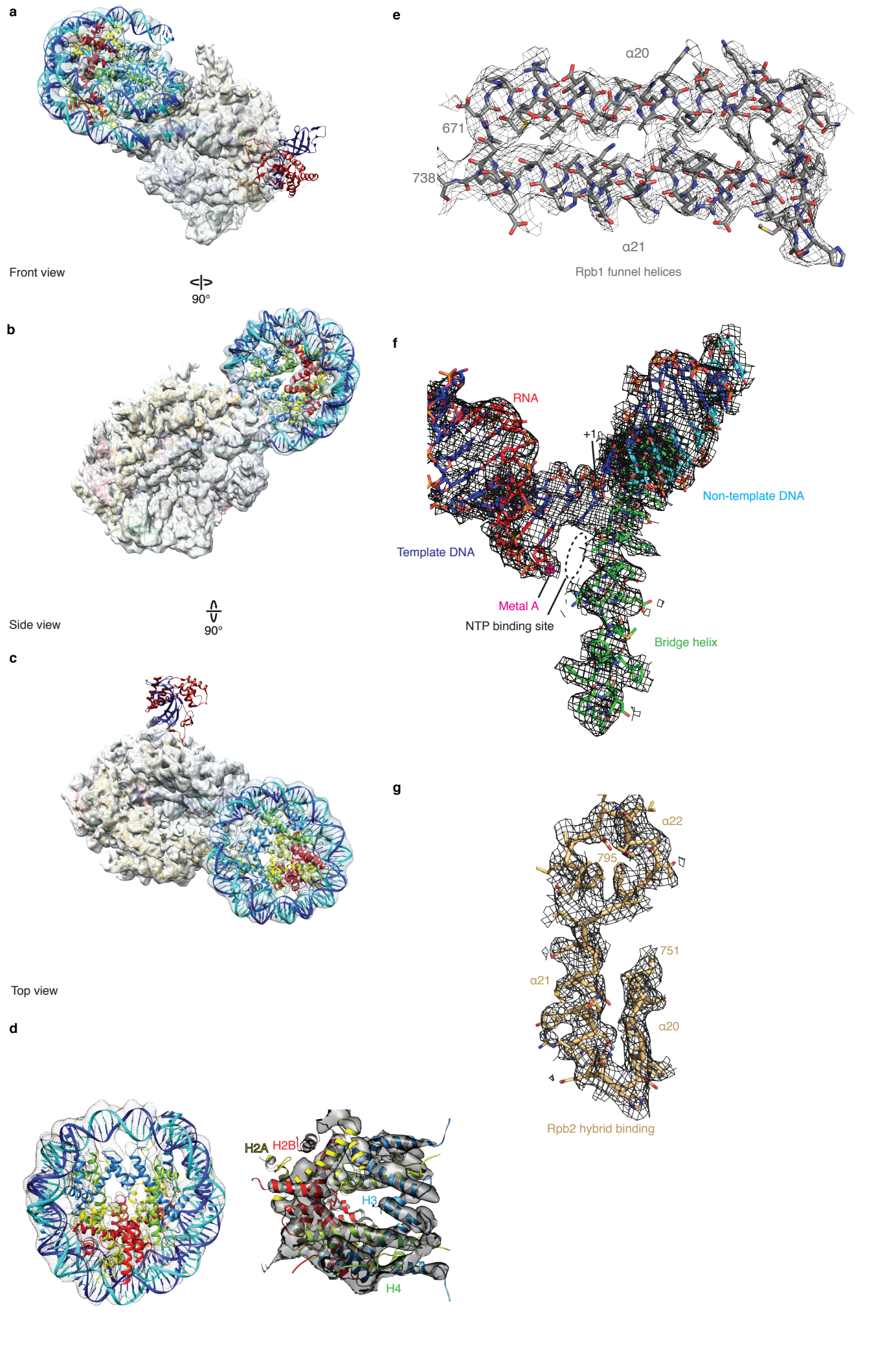

a
Template DNA
Non-template DNA
RNA
Extranucleosomal DNA
DNA-RNA hybrid
3'
5'
FAM
Bridge Helix
NTP binding site
Metal A
NCP with 145 bp
Widom 601 sequence

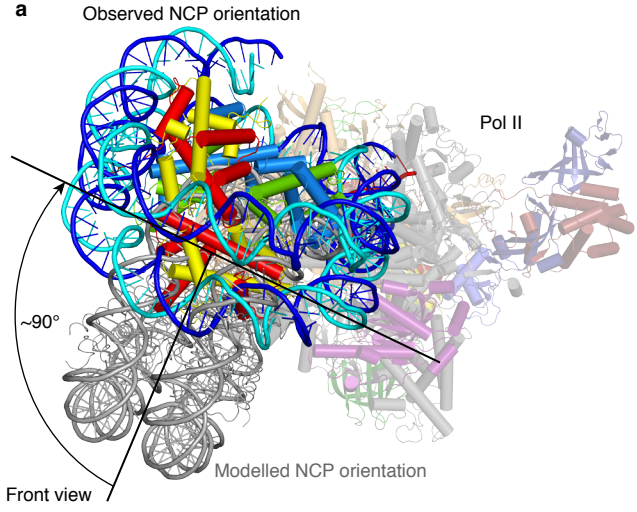

b
RNA marker
nt
Time (min)
0 1 5 10 20 30
+49
+40
+30
+25
+10
Elongation block at ~SHL -5
Initial RNA

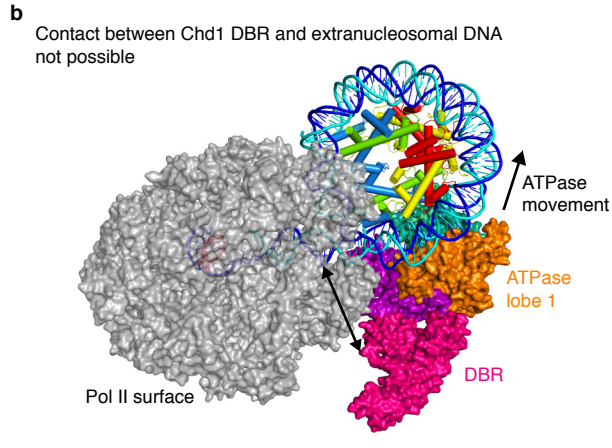

c
280 nm
260 nm
EC-NCP complex
Free NCP
Void
Peak fraction used
for cryo-EM
Absorption (mAU)
Volume (mL)
0
100
200
300
0.75
1.00
1.25
1.50
1.75


d
Rpb1
Rpb2
Rpb3
Rpb4
Rpb5
Rpb7
Rpb6
Rpb8/H3
Rpb7/H2B
Rpb9/H4
Rpb11
Rpb10/Rpb12


a
100 nm
b
200 Å
d
1125
750
375
1
Number of particles
c
1,041,545
BoxNet2 picked particles
Segmentation in three batches
2D classification, 3D refinement,
& global classifications
156,636 particles
Representative 3D refinement
from batch 1 & 2
3D classifications
of Pol II-NCP complex
111,895 particles
Masked classification on Pol II-bound NCP (Batch 1 & 2)
44,741 particles
Masked classification on Pol II-
NCP complex (Batch 3)
28,700
particles
9,060
particles
8,212
particles
6,780
particles
13,251
particles
52,752
particles
2D classification
38,926
particles
52,177 particles
2D classification
49,703 particles
CTF refinement &
final global refinement
Multi-body refinement
NCP
Multi-body refinement
Pol II
6.9 Å
Pol II-NCP complex
49,703 particles
4.4 Å
4.3 Å
e
overall Pol II-NCP
Pol II multibody
NCP multibody
1.0
0.5
0.143
0.0
-0.2
FSC
6.85 Å
4.4 Å
4.25 Å
0.1
0.2
0.3
0.4
0.5
Resolution (1/Å)
f
composite map
3.5 Å
5 Å
6.5 Å
8 Å
9.5 Å
90°
90°
Front view
Side view
Top view


a
Front view
90°
b
Side view
90°
c
Top view
d
H2A
H2B
H3
H4
e
α20
671
738
α21
Rpb1 funnel helices
f
RNA
+1
Non-template DNA
Template DNA
Metal A
NTP binding site
Bridge helix
g
α22
795
751
α21
α20
Rpb2 hybrid binding


a
Observed NCP orientation
Pol II
~90°
Modelled NCP orientation
Front view


b
Contact between Chd1 DBR and extranucleosomal DNA
not possible
ATPase
movement
ATPase
lobe 1
DBR
Pol II surface

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

	Pol II-NCP structure (EMDB-XXXX) (PDB XXXX)
Data collection and processing	
Magnification	130,000
Voltage (kV)	300
Electron exposure (e–/Å^2)	46
Defocus range (μm)	0.25-4
Pixel size (Å)	1.05
Symmetry imposed	C1
Initial particle images (no.)	1,041,545
Final particle images (no.)	49,703
Map resolution (Å)	4.4
FSC threshold	0.143
Map resolution range (Å)	4-12
Refinement	
Initial model used (PDB code)	3HOV, 3LZ0
Map sharpening *B* factor (Å^2)	-181.9
Model composition	
Non-hydrogen atoms	44,012
Protein residues	44,003
Ligands	9
B factors (Å^2)	
Protein	138.10
Ligand	100.64
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.20
Validation	
MolProbity score	1.95
Clashscore	7.37
Poor rotamers (%)	0.37
Ramachandran plot	
Favored (%)	90.01
Allowed (%)	9.9
Disallowed (%)	0.09