

Structure of the human RNA Polymerase I Elongation Complex

Yanhui Xu (✉ xuyh@fudan.edu.cn)

Fudan University <https://orcid.org/0000-0002-1408-5178>

Dan Zhao

Shanghai Medical College of Fudan University

Weida Liu

Fudan University

Ke Chen

Shanghai Medical College of Fudan University

Zihan Wu

Fudan University

Huirong Yang

Fudan University

Article

Keywords: Eukaryotic RNA polymerase I, RNA Polymerase I Elongation, ribosomal DNA

Posted Date: June 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-576610/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Cell Discovery on October 20th, 2021. See the published version at <https://doi.org/10.1038/s41421-021-00335-5>.

Structure of the human RNA Polymerase I Elongation Complex

Dan Zhao^{1#}, Weida Liu^{1#}, Ke Chen^{1#}, Zihan Wu¹, Huirong Yang^{1*}, and Yanhui Xu^{1,2,3,4*}

¹Fudan University Shanghai Cancer Center, Institutes of Biomedical Sciences, State Key Laboratory of Genetic Engineering, Shanghai Key Laboratory of Radiation Oncology, and Shanghai Key Laboratory of Medical Epigenetics, Shanghai Medical College of Fudan University, Shanghai 200032, China.

²The International Co-laboratory of Medical Epigenetics and Metabolism, Ministry of Science and Technology, China, Department of Systems Biology for Medicine, School of Basic Medical Sciences, Shanghai Medical College of Fudan University, Shanghai 200032, China.

³Human Phenome Institute, Collaborative Innovation Center of Genetics and Development, School of Life Sciences, Fudan University, Shanghai 200433, China

⁴State Key Laboratory of Reproductive Regulation and Breeding of Grassland Livestock School of Life Sciences, Inner Mongolia University, Hohhot, 010070, China

These authors contributed equally to this work.

* To whom correspondence should be addressed. E-mail: xuyh@fudan.edu.cn

19 **Abstract**

20 Eukaryotic RNA polymerase I (Pol I) transcribes ribosomal DNA and generates RNA for ribosome
21 synthesis. Pol I accounts for the majority of cellular transcription activity and dysregulation of Pol I
22 transcription leads to cancers and ribosomopathies. Despite extensive structural studies of yeast Pol
23 I, structure of human Pol I remains unsolved. Here, we determined the structures of the human Pol
24 I in the pre-translocation, post-translocation, and backtracked states at near-atomic resolution. The
25 single-subunit peripheral stalk lacks contacts with the DNA-binding clamp and is more flexible than
26 the two-subunit stalk in yeast Pol I. Compared to yeast Pol I, human Pol I possesses a more closed
27 clamp, which makes more contacts with DNA and may support more efficient transcription in
28 human cells. The Pol I structure in the post-cleavage backtracked state shows that the C-terminal
29 zinc ribbon of RPA12 inserts into an open funnel and facilitates “dinucleotide cleavage” on
30 mismatched DNA-RNA hybrid. Critical disease-associated mutations are mapped on Pol I regions
31 that are involved in catalysis and complex organization. In summary, the structures provide new
32 sights into human Pol I complex organization and efficient proofreading, consistent with
33 requirement of efficient transcription of ribosomal DNA in human cells.

34 **Introduction**

35 Among the three eukaryotic RNA polymerases (Pol I, Pol II and Pol III)¹, Pol I accounts for up to
36 60% of cellular transcriptional activity^{2, 3}. The human Pol I (hPol I) is located in the nucleolus and
37 synthesizes 47S pre-ribosome RNA^{4, 5}, which is further processed into mature 18S, 5.8S and 28S
38 ribosomal RNAs (rRNAs), the essential structural and enzymatic components of ribosomes. The Pol
39 I-mediated pre-rRNA transcription is strictly required for ribosome biogenesis². Dysregulation of
40 Pol I transcription is associated with human diseases and the Pol I transcription machinery is
41 considered to be the drug target for anti-cancer therapy⁶⁻⁹.

42 The hPol I consists of 13 subunits and the yeast Pol I (yPol I) has similar counterparts with an
43 additional yeast-specific subunit, A14. Among the 10 core subunits of the human Pol I, five subunits
44 (RPABC1, RPABC2, RPABC3, RPABC4, and RPABC5) are shared by the three RNA polymerases
45 (Pol I, II, and III) and two subunits (RPAC1 and RPAC2) are shared by Pol I and Pol III, indicating
46 a highly conserved catalytic core¹⁰. The general transcription factors TFIIE and TFIIIF are
47 dissociable and serve as regulatory complexes in Pol II. In contrast, the TFIIIF-/TFIIE-like subunits
48 are evolved as bona fide constitutive subunits, the PAF53/PAF49 heterodimer in Pol I and C37/C53
49 dimer in Pol III¹¹. Additionally, the yPol I stalk is formed by A14/A43 heterodimer whereas the
50 functional human Pol I stalk consists of only one subunit, RPA43, and the A14 counterpart has not
51 been identified. The human ribosomal DNAs (rDNAs) are larger and more complex in organization
52 than that in yeast^{12, 13}, suggesting distinct structure and function of Pol I.

53 The structure of yPol I has been elaborately studied in the past decades. The overall architecture of
54 yPol I structure was reported in early studies¹⁴⁻¹⁶. Recent studies determined yPol I structures in
55 distinct nucleotides binding states¹⁶⁻¹⁹. It remains incompletely understood how hPol I is assembled
56 and what the differences are between hPol I and yPol I. Here, we present cryo-electron microscopy
57 (cryo-EM) structures of hPol I elongation complex (EC) in the pre-translocation, post-translocation,
58 and backtracked states. These structures reveal human Pol I-specific structural features and
59 molecular mechanism of backtracking and RNA cleavage of the mismatched DNA-RNA hybrid for
60 proofreading.

61 **Results**

62 **Complex assembly and structure determination of the human Pol I elongation complexes**

63 The 13-subunit hPol I was overexpressed in Expi293F cells and purified to homogeneity for
64 structural and biochemical analyses (Supplementary information, Fig. S1a). The purified complex
65 exhibited DNA-dependent RNA elongation activity on DNA template and RNA cleavage activity
66 on a mismatched DNA-RNA hybrid, indicative of a functional Pol I (Supplementary information,
67 Fig. S1b). To elucidate the molecular mechanism of Pol I-mediated transcription elongation, we
68 assembled three complexes mimicking the elongating complexes in the pre-translocation, post-
69 translocation, and backtracked states, respectively (Fig. 1a; Supplementary information, Table S1).
70 The Pol I EC in the post-translocation state (EC^{post}) was assembled by incubating the purified Pol I
71 and a DNA-RNA hybrid scaffold consisting of DNA with 11 mismatched base pairs (bp) and RNA
72 with 8 nucleotides (nt)²⁰. The EC in the pre-translocation state (EC^{pre}) was assembled by adding
73 CMPCPP (Cytidine-5'-[(α , β)-methylene]-triphosphate), a non-hydrolysable nucleotide, to the
74 EC^{post} . The DNA-RNA scaffold used in assembly of EC in the backtracked state (EC^{bt}) consists of
75 an additional mismatched base pair at -1 site, relative to the NTP addition site in the elongation
76 complex (Supplementary information, Table S2).

77 The structures of Pol I EC in the three states were determined using cryo-EM single particle
78 reconstruction (Supplementary information, Fig. S2 and Table S1). The cryo-EM maps of EC^{post} ,
79 EC^{pre} , and EC^{bt} were refined to 2.8 Å, 2.9 Å, and 3.0 Å resolution, respectively. The cryo-EM maps
80 of the three structures reveal well-ordered core complex and support unambiguous tracing of
81 residues and nucleotides (Fig. 1c; Supplementary information, Fig. S3). The structural models were
82 manually built aided by the yPol I structures^{18,21} as template and secondary structure prediction. The
83 peripheral regions were relatively flexible and the structural model was built by fitting structural
84 templates into the cryo-EM maps, followed by manual adjustment.

85
86 **Overall structure of hPol I elongation complex in the post-translocation state**

87 The Pol I structures in the three states adopt almost identical overall fold except for distinct
88 organization of nucleic acids within the catalytic center and slight difference in modular organization
89 of EC^{bt} (Supplementary information, Fig. S4a). The structure of EC^{post} will be discussed below in
90 analyzing the shared structure features (Fig. 1; Supplementary information, Movie S1).

91 The cryo-EM map of Pol I EC^{post} reveals a rigid globular core formed by 10 core subunits (Fig. 1b).
92 The two Pol I-specific large subunits, RPA1 and RPA2, create the central DNA-binding cleft and
93 are surrounded by the rest subunits. The five Pol I/II/III-shared components (RPABC1, RPABC2,
94 RPABC3, RPABC4 and RPABC5) are surrounding subunits and integrated into the core module

95 through binding RPA1 and PRA2. The RPAC1 and RPAC2, two shared subunits of Pol I/III, interact
96 with each other and pack against the wall of Pol I. The TFIIS-like subunit RPA12 has an N-terminal
97 zinc ribbon (N-ribbon), which packs against the dimerization domain of TFIIF-like PAF53/PAF49
98 heterodimer and the lobe of RPA2. The C-terminal ribbon (C-ribbon) of RPA12 inserts into the
99 funnel in the backtracked state, consistent with its role in proofreading (detailed below).
100 The cryo-EM map reveals relatively weak density around the three peripheral subunits (Fig. 1b). We
101 performed glutaraldehyde crosslinking to prepare EC^{post} followed by structure determination,
102 generating improved cryo-EM density around RPA43, PAF53 and PAF49 (Supplementary
103 information, Fig. S2). Structural models were built by docking the structural templates of these
104 subunits in yPol I into the cryo-EM map followed by manual adjustment. The flexible stalk consists
105 of one subunit (RPA43) and lacks the yeast counterpart A14 (Fig. 2a; Supplementary information,
106 Fig. S5a). The N-terminal dimerization domains of PAF53 and PAF49 interact with each other and
107 together form a triple beta barrel-like fold, similar to the dimerization domain of TFIIF²². The linker
108 region of PAF49 winds over PRA2, RPAC1 and RPABC5, and facilitates positioning the PAF53-
109 PAF49 heterodimer on the lobe. The TFIIE-like C-terminal tandem winged helix (tWH) domain of
110 PAF53 and the positively charged C-terminal tail of PAF49 were not observed due to their flexibility.
111

112 **The hPol I has a more closed DNA-binding clamp than yPol I**

113 The hPol I EC adopts an overall architecture generally similar to that of yPol I EC¹⁸ (Fig. 2a;
114 Supplementary information, Movie S2), consistent with the high sequence similarity of the core
115 subunits^{15, 16}. Structural comparison shows that the central core and DNA-RNA hybrid are well
116 superimposed and the major difference exists in the conformational arrangement of the clamp (Fig.
117 2a, b). The yPol I have a relatively closer clamp-stalk association with the two modules stably
118 bridged by yPol I-specific insertions, residues 1133-1168 of A135 (RPA2 in human Pol I) and
119 residues 52-68 of Rpb6 (RPABC2 in human Pol I) (Fig. 2b, c, d; Supplementary information, Fig.
120 S6b, c). The intermodular association is further supported by the N-terminal region (residues 12-24)
121 of the yPol I-specific subunit A14 (Supplementary information, Fig. S5a). Due to the lack of
122 stabilizations by equivalent regions, the stalk and clamp are more separated in hPol I, as measured
123 by ~68 Å between residues P125 (RPA43 in hPol I) of stalk and R101 (RPA1 in hPol I) of clamp,
124 compared to the separation of equivalent residues by ~64 Å in yPol I (Fig. 2b). As a result, the clamp
125 in hPol I is positioned closer (by ~4 Å) to the DNA than that in yPol I, generating a more closed
126 DNA entry cleft (Fig. 2b, e, f). Positively charged residues K197, R1663, R1659 (Fig. 2e), R418,
127 K423, K429 and K424 (Fig. 2f) of the clamp are brought into close contact with the phosphate
128 groups of both template and nontemplate strands and may stabilize the EC complex during

129 transcription elongation. In contrast, yPol I has much less clamp-DNA contacts. Other structural
130 differences were observed in the two human-specific insertions (residues 721-732 and 1078-1131)
131 (Supplementary information, Fig. S6a) in RPA1 and slight positional shifts of their adjacent domains
132 (Fig. 2g, h).

133

134 **The single-subunit stalk of human Pol I**

135 In yPol I, A43 and A14 subunits form the stalk (Supplementary information, Fig. S5a). It is known
136 that A43 interacts with transcription factor Rrn3²³, an essential transcription initiation factor
137 conserved in human and yeast²⁴⁻²⁷. However, A14 counterpart has not been identified in human Pol
138 I. The purified human Pol I complex was catalytically active in our in vitro transcription assay
139 (Supplementary information, Fig. S1a), suggesting that A14 is indispensable for human Pol I
140 elongation activity.

141 The cryo-EM map of the stalk is relatively weak, consistent with the relatively small contact surface
142 between RPA43 and the Pol I core (Fig. 1b). Similar to Rpb7 of Pol II²⁸ and C25 of Pol III²⁹, the
143 human stalk subunit RPA43 adopts an extended fold and consists of a Tip domain and an OB domain
144 (Fig. 1c)¹⁶. The predicted C-terminal positively charged tail of RPA43 is invisible. The yeast A43-
145 A14 heterodimer has been believed to direct the dimerization of Pol I¹⁶. However, no obvious
146 dimerization of human Pol I was observed (Supplementary information, Fig. S1c).

147 The above structural differences between hPol I and yPol I may reflect distinct functional
148 requirements of Pol I in the two species. For example, the clamp makes more contacts with the entry
149 DNA in hPol I, suggesting a more stabilized hPol I-DNA engagement during transcription elongation,
150 in line with the transcription of more complex and longer rDNA substrate in human cells³⁰. The
151 flexible stalk in hPol I is similar to that of Pol II but differs from the relatively fixed stalk-clamp of
152 yPol I, suggesting that the flexible stalk may accommodate binding of more complicated regulatory
153 proteins in human cells. The functional effect of these human-specific structural features (clamp and
154 stalk) in hPol I requires further investigation.

155

156 **The catalytic center of Pol I EC in the pre- and post-translocation states**

157 The EC^{post} structure reveals characteristic DNA-RNA hybrid and catalytic center in the post-
158 translocation state (Fig. 3a). The “metal A” magnesium cation is coordinated by three highly
159 conserved aspartate residues (D592, D590, D588) of RPA1 and binds 3' end of the growing RNA
160 transcript. The nucleotides are mainly stabilized by subunits RPA1, RPA2, and RPABC1, and most
161 of the nucleotide-binding residues are conserved across species^{15, 16}.

162 The cryo-EM map of the EC^{pre} shows well-resolved density of the CMPCPP at the +1 nucleotide

163 addition site (Fig. 3b), similar to previously reported structures of yeast Pol I and Pol II in the pre-
164 translocation state^{31, 32}. The phosphate group of the CMPCPP is stabilized by R684 and R923 of
165 RPA2, two invariant residues in yeast and human Pol I. The conserved residues N586 and R552 of
166 RPA1 are located near the 3' and 2' hydroxyl group of the CMPCPP, respectively, similar to that in
167 yeast Pol I and Pol II structures^{32, 33}. Residue P554 of RPA1 is located close to the cytosine of the
168 CMPCPP. These interactions serve to recognize all type of NTP in the A site. Residue T978 of
169 RPA1 points towards the CMPCPP, and this conserved residue may participate in detecting the base
170 pairing of +1 site in yeast Pol II³⁴. Residue Y687 of RPA2 around the CMPCPP is involved in
171 proofreading (discussed below). The positioning of the NTP substrate in the active site is similar to
172 that in yPol I structure¹⁹ (Supplementary information, Fig. S5b), indicating a highly conserved
173 catalytic mechanism.

174 The rudder, fork loops 1/2 and loops A/B work together to stabilize the transcription fork and prevent
175 re-association of the template and nontemplate strands (Fig. 3c). Compared to that of yPol I, the
176 rudder (residues 403-416 of RPA1) in hPol I EC^{post}/EC^{pre} is closer to the DNA-RNA scaffold due to
177 the more closed clamp and may better stabilize the transcription fork. Other elements are similarly
178 positioned in yPol I and hPol and stabilize the transcription bubble by preventing re-association of
179 the template and nontemplate strands (Supplementary information, Fig. S6a).

180

181 **Structure of Pol I in the backtracked state reveals the post-cleavage conformation**

182 Eukaryotic RNA polymerases can move backwards on DNA template to remove the RNA 3' end
183 nucleotides from the active site when encountering transcription barriers, such as mismatched NTP
184 addition³⁵ and UV-induced damaged DNA³⁶. It is known that transcription arrest occurs on the
185 backtracked Pol II and reactivation of the arrest Pol II requires an additional transcription factor IIS
186 (TFIIS), which cleaves the mismatched RNA³⁷. The equivalent cleavage factors, RPA12 in Pol I
187 and RPC11 in Pol III, are incorporated into Pol I and Pol III, respectively, and cleave the mismatched
188 RNA for proofreading^{15, 38}. In Pol I, the N-ribbon of RPA12 resembles that of the Pol II subunit
189 Rpb9 and the C-ribbon resembles that of TFIIS²⁰. Although the backtracking of Pol II has been
190 reported^{37, 39, 40}, the post-cleavage state was not observed in previous studies, in which the negatively
191 charged residues D and E of TFIIS have been mutated to deactivate its cleavage activity^{37, 40}. The
192 mechanism of backtracking and RNA cleavage in Pol I-mediated transcription remains incompletely
193 understood.

194 To obtain the structure of Pol I EC^{bt}, we assembled the complex with the nucleotide at the -1 site of
195 the template DNA converted from A to T, generating a T-U mismatch (Fig. 4a). Consistent with the
196 cleavage of dinucleotide in the in vitro transcription assay (Supplementary information, Fig. S1b,

197 lanes 5-9), the remaining 6-nt RNA was evidently observed in the cryo-EM map. The metal A was
198 invisible at the original position near the three aspartic acids (Supplementary information, Fig. S4d).
199 The structure of EC^{bt} represents a Pol I in the backtracked state after cleavage of the mismatched
200 RNA.

201 The C-ribbon of RPA12 was not observed in the EC^{pre} and showed very weak density in the EC^{post}.
202 In contrast, the cryo-EM of EC^{bt} reveals well-ordered C-ribbon of RPA12 within the funnel (Fig.
203 4b). Consistently, the bridge helix is slightly extended (Supplementary information, Fig. S4c) and
204 the funnel and the cleft of EC^{bt} are slightly wider than that of EC^{post} to permit the entry of RPA12
205 C-ribbon (Fig. 4c), which may otherwise clash with the rim of the funnel in the EC^{post}. A conserved
206 tyrosine located in the active site is called “gating tyrosine”, which can block backward movement
207 of RNA. In yeast Pol II EC^{bt} without TFIIS, the “gating tyrosine” Y769 of Rpb2 clearly blocks the
208 backtracked RNA⁴⁰. In hPol I EC^{bt}, Y687 is “opened” by the C-ribbon in EC^{bt} and permits the
209 backward translocation of 3' RNA from the active site (Fig. 4d; Supplementary information, Fig.
210 S4d). The tip residues D106 and E107 of the C-ribbon are in close proximity to the bridge helix and
211 are ~5 Å away from the 3' end of the modeled RNA (Supplementary information, Fig. S4e). These
212 residues are invariant in yeast and human Pol I and may coordinate the nucleophilic water and
213 magnesium cation to cleave the phosphodiester bond. The arrangement of this RPA12 tip is
214 consistent with the previously proposed S_N2 mechanism to cleave the scissile phosphodiester bond
215 in Pol II³⁷. While it has been proposed that metal B is important for the cleavage activity³⁷, metal A
216 was observed in EC^{pre}/EC^{post} but not in EC^{bt} (Supplementary information, Fig. S4d), suggesting that
217 metal A may also be involved in RNA cleavage.

218 Compared to EC^{post}, the template strand in the EC^{bt} has obvious positional shift (Fig. 4e). For
219 example, the T⁻³ of the EC^{post} moves to the -1 site in EC^{bt}, which is located between the positions -
220 1 and +1 in the EC^{post}. This base tilts by ~13° to generate base pair with the first RNA base at the -
221 1 in EC^{bt}. The G⁻² of the EC^{post} moves over the bridge helix and is positioned to the +1 site in the
222 EC^{bt} and this unpaired base points towards the downstream DNA duplex. The catalytic center and
223 the DNA-RNA hybrid position of Pol I EC^{bt} are generally similar to that of the reactivation
224 intermediate Pol II³⁷ (Supplementary information, Fig. S5d), in which the mutated TFIIS resulted in
225 a similar rotation of the gating tyrosine whereas the long backtracked RNA was not cleaved. In
226 contrast, no DNA-RNA hybrid rearrangement was observed in the structure of Pol II EC containing
227 mutated TFIIS and a short-mismatched RNA⁴⁰. Distinct from the above observations in Pol II, Pol
228 I leads to rearrangement of the catalytic center and cleaves the substrate containing a short-
229 mismatched RNA in the presence of active RPA12.

230

231 **Disease-associated mutations of the human Pol I**

232 The human Pol I-mediated transcription is critical for ribosome production, regulation of cell growth
233 and proliferation. Mutations of Pol I subunits result in perturbation of ribosome biogenesis during
234 development and lead to ribosomopathies, such as severe neurodegenerative disease, Acrofacial
235 Dysostosis type Cincinnati (AFDCIN) and Treacher Collins Syndrome (TCS)⁴¹⁻⁴⁵ (Fig. 5a).

236 The mutation E593Q near the active site of PRA1 causes AFDCIN, a cranioskeletal malformation
237 syndrome⁴¹. A recent study shows that E593Q-containing Pol I stably binds rDNA chromatin and
238 inhibits wild-type Pol I condensate formation, leading to repression of rRNA transcription⁴⁶.
239 Notably, residue E593 is positioned near the conserved magnesium-binding aspartate residues
240 (D592, D590 and D588) (Fig. 5b). The replacement of E to Q may affect the geometry of the
241 magnesium coordination and nucleotide addition, thus decrease Pol I transcription and generate a
242 dominant-negative polymerase.

243 Mutations E47K, T50I, L51R, R56C, L82S and G99S in the N-terminal helices of RPAC2 are
244 associated with TCS^{42, 43} (Fig 5c, d). These residues involve intermolecular interactions between
245 RPAC1 and RPAC2 and the above mutations may lead to the destabilization of RPAC1-RPAC2
246 heterodimer and affect Pol I activity. Moreover, residue R56 of RPAC2 binds RPA1 on residues
247 E651 and E655 in Pol I, but does not generate similar charge-charge interaction in Pol III (Fig. 5c;
248 Supplementary information, Fig. S5e). As a shared subunit of Pol I and Pol III, the deletion of
249 RPAC2 in zebra fish impaired Pol I-supported transcription of 47S RNA, but had no effect on Pol
250 III-supported transcription of 5S RNA⁴⁷. Thus, dysfunction of Pol I may play a major role in
251 RPAC2-related TCS. Moreover, TCS can also result from RPA2 subunit mutations. Mutation S682R
252 of RPA2 may destabilize the bridge helix of RPA1 (Fig. 5e); mutation R1003C of RPA2 is located
253 at the DNA-RNA hybrid-binding regions and may affect nucleotides association⁴⁴ (Fig. 5f). These
254 mutations may decrease Pol I activity and thus effect ribosome biogenesis in development.

255 **Discussion**

256 In this study, we reconstituted human Pol I complex and determined the cryo-EM structures of Pol
257 I in the pre-translocation, post-translocation, and backtracked states. The overall human Pol I EC
258 structures reveal a more closed DNA-binding clamp and a flexible one-subunit stalk. The Pol I in
259 the post-cleavage backtracked state has wider cleft and RPA12 in the funnel may favor the RNA
260 cleavage upon backtracking of the mismatched RNA. Structural comparison reveals structural
261 differences between human Pol I and other polymerases, which may reflect their functional
262 differences. Our structures suggests potential implication of disease-associated mutations of hPol I
263 and provides basis for further studies of human Pol I transcription initiation.

264 In human cells, the total number of rDNA repeats was estimated to be in the range of approximately
265 400 copies, and each rDNA repeat (~43 kb) consists of regulatory elements within an intergenic
266 spacer (IGS) of ~30 kb and the 47S pre-rRNA coding region of ~13 kb^{5, 48}. While in *S. cerevisiae*,
267 each rDNA repeat (~9.1 kb) consists of ~6.6 kb 35S coding region and a short IGS^{30, 49}. The *S.*
268 *cerevisiae* and mammalian Pol I elongate at an average speed of 60 nt/s and 91 nt/s, respectively⁵⁰.
269 ⁵¹. The difference in transcription speed may result from the observed structural differences.
270 Compared to yPol I, hPol I has a more closed DNA-binding clamp, which may generate more stable
271 binding of the rDNA during transcription elongation and support higher transcription efficiency
272 and/or processivity.

273 In all our structures, the stalk subunit RPA43 is flexible, as evidenced by the relatively weak cryo-
274 EM density. It is known that RPA43 involves transcription initiation³³. Recently reported structures
275 of yeast Pol I initiation complexes shows that yeast Pol I initiates transcription in the presence of
276 transcription core factor and Rrn3 without TATA box-binding protein (TBP)³³. In contrast, the
277 human Pol I initiation complex involves upstream binding factor (UBF), RRN3 and TBP-containing
278 selectivity factor 1 (SL1)^{52, 53}. Expect for RRN3, other transcription factors and promoter elements
279 are not conserved between human and yeast⁵⁴. The compositional and conformational differences in
280 the stalk of hPol I and yPol I may accommodate distinct transcription initiation mechanism.

281 The mechanistic studies of polymerase proofreading have been mainly focused on Pol II. When Pol
282 II encounters mismatched DNA-RNA, RNA fraying occurs and Pol II pauses and backtracks by one
283 base pair. However, further backtracking is hindered by the gating tyrosine. Transcription elongation
284 continues if the mismatched RNA is cleaved by intrinsic cleavage activity of the polymerase^{37, 40}.
285 However, if the mismatched RNA goes backtrack beyond the gating tyrosine at some occasions,
286 transcription arrest occurs and TFIIS is required to bind to reactivate the arrested Pol II. Our study
287 suggests a mechanism of coordinated backtracking and RNA cleavage in Pol I-mediated
288 transcription. The mismatched DNA-RNA possibly facilitates the Pol I to open the cleft and funnel

289 to permit the entry of RPA12, which pushes the gating tyrosine Y687 aside to open the gate and
290 allows the mismatched RNA to pass through. The catalytic center rearranges the nucleotides and
291 exposes the scissile phosphodiester bond between nucleotide -1 and +1 to the negatively charged
292 RPA12 tip, which may facilitate the cleavage of the phosphodiester bond. Reactivation and continue
293 transcription may occur upon the addition of next NTP. These evidences are consistent with previous
294 report that yeast Pol I transcribes faster than Pol II and pauses less often, and Pol I is more efficient
295 in backtrack recovery than Pol II¹⁷.

296 **References:**

- 297 1 Roeder RG, Rutter WJ. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms.
298 *Nature* 1969; **224**:234-237.
- 299 2 Moss T, Stefanovsky VY. At the center of eukaryotic life. *Cell* 2002; **109**:545-548.
- 300 3 Warner JR. The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* 1999; **24**:437-440.
- 301 4 Brian, McStay, Ingrid, Grummt. The epigenetics of rRNA genes: from molecular to chromosome biology.
302 *Annual Review of Cell & Developmental Biology* 2008.
- 303 5 McStay B, Grummt I. The epigenetics of rRNA genes: from molecular to chromosome biology. *Annu Rev*
304 *Cell Dev Biol* 2008; **24**:131-157.
- 305 6 Sharifi S, Bierhoff H. Regulation of RNA Polymerase I Transcription in Development, Disease, and Aging.
306 *Annu Rev Biochem* 2018; **87**:51-73.
- 307 7 Drygin D, Rice WG, Grummt I. The RNA polymerase I transcription machinery: an emerging target for the
308 treatment of cancer. *Annu Rev Pharmacol Toxicol* 2010; **50**:131-156.
- 309 8 Ferreira R, Schneekloth JS, Jr., Panov KI, Hannan KM, Hannan RD. Targeting the RNA Polymerase I
310 Transcription for Cancer Therapy Comes of Age. *Cells* 2020; **9**.
- 311 9 Wei T, Najmi SM, Liu H *et al.* Small-Molecule Targeting of RNA Polymerase I Activates a Conserved
312 Transcription Elongation Checkpoint. *Cell Rep* 2018; **23**:404-414.
- 313 10 Russell J, Zomerdijk JC. The RNA polymerase I transcription machinery. *Biochem Soc Symp* 2006:203-
314 216.
- 315 11 Vannini A, Cramer P. Conservation between the RNA polymerase I, II, and III transcription initiation
316 machineries. *Mol Cell* 2012; **45**:439-446.
- 317 12 Grummt, Roth, R M, Paule. Ribosomal RNA transcription in vitro is species specific. *Nature* 1982.
- 318 13 Ingrid, Grummt. Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus.
319 *Genes & development* 2003.
- 320 14 Schultz P, Celia H, Riva M, Sentenac A, Oudet P. Three-dimensional model of yeast RNA polymerase I
321 determined by electron microscopy of two-dimensional crystals. *EMBO J* 1993; **12**:2601-2607.
- 322 15 Kuhn CD, Geiger SR, Baumli S *et al.* Functional architecture of RNA polymerase I. *Cell* 2007; **131**:1260-
323 1272.
- 324 16 Fernandez-Tornero C, Moreno-Morcillo M, Rashid UJ *et al.* Crystal structure of the 14-subunit RNA
325 polymerase I. *Nature* 2013; **502**:644-649.
- 326 17 Lisica A, Engel C, Jahnel M *et al.* Mechanisms of backtrack recovery by RNA polymerases I and II. *Proc*
327 *Natl Acad Sci U S A* 2016; **113**:2946-2951.
- 328 18 Neyer S, Kunz M, Geiss C *et al.* Structure of RNA polymerase I transcribing ribosomal DNA genes.
329 *Nature* 2016; **540**:607-610.
- 330 19 Tafur L, Sadian Y, Hanske J, Wetzel R, Weis F, Muller CW. The cryo-EM structure of a 12-subunit variant
331 of RNA polymerase I reveals dissociation of the A49-A34.5 heterodimer and rearrangement of subunit A12.2.
332 *Elife* 2019; **8**.
- 333 20 Lin CW, Moorefield B, Payne J, Aprikian P, Mitomo K, Reeder RH. A novel 66-kilodalton protein
334 complexes with Rrn6, Rrn7, and TATA-binding protein to promote polymerase I transcription initiation in
335 *Saccharomyces cerevisiae*. *Mol Cell Biol* 1996; **16**:6436-6443.
- 336 21 Tafur L, Sadian Y, Hoffmann NA *et al.* Molecular Structures of Transcribing RNA Polymerase I. *Mol Cell*
337 2016; **64**:1135-1143.

338 22 Geiger SR, Lorenzen K, Schrieck A *et al.* RNA polymerase I contains a TFIIF-related DNA-binding
339 subcomplex. *Mol Cell* 2010; **39**:583-594.

340 23 Peyroche G, Milkereit P, Bischler N *et al.* The recruitment of RNA polymerase I on rDNA is mediated by
341 the interaction of the A43 subunit with Rrn3. *EMBO J* 2000; **19**:5473-5482.

342 24 Pilsl M, Crucifix C, Papai G *et al.* Structure of the initiation-competent RNA polymerase I and its
343 implication for transcription. *Nat Commun* 2016; **7**:12126.

344 25 Moorefield B, Greene EA, Reeder RH. RNA polymerase I transcription factor Rrn3 is functionally
345 conserved between yeast and human. *Proceedings of the National Academy of Sciences* 2000; **97**:4724-4729.

346 26 Stepanchick A, Zhi H, Cavanaugh AH, Rothblum K, Schneider DA, Rothblum LI. DNA binding by the
347 ribosomal DNA transcription factor *rrn3* is essential for ribosomal DNA transcription. *J Biol Chem* 2013;
348 **288**:9135-9144.

349 27 Miller G, Panov KI, Friedrich JK, Trinkle-Mulcahy L, Lamond AI, Zomerdijk JCBM. hRRN3 is essential
350 in the SL1-mediated recruitment of RNA Polymerase I to rRNA gene promoters. *The EMBO Journal* 2001;
351 **20**:1373-1382.

352 28 Armache K. Crystal structures of the complete 12-subunit RNA polymerase II and its subcomplex Rpb4/7,
353 and modeling of RNA polymerases I and III. *Ludwig-Maximilians-Universität München* 2005.

354 29 Jasiak AJ, Armache KJ, Martens B, Jansen RP, Cramer P. Structural Biology of RNA Polymerase III:
355 Subcomplex C17/25 X-Ray Structure and 11 Subunit Enzyme Model: *Molecular Cell*. *Molecular Cell* 2006;
356 **23**:71-81.

357 30 Woolford JL, Jr., Baserga SJ. Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics* 2013;
358 **195**:643-681.

359 31 Cheung AC, Sainsbury S, Cramer P. Structural basis of initial RNA polymerase II transcription. *EMBO J*
360 2011; **30**:4755-4763.

361 32 Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD. Structural basis of transcription: role of
362 the trigger loop in substrate specificity and catalysis. *Cell* 2006; **127**:941-954.

363 33 Sadian Y, Baudin F, Tafur L *et al.* Molecular insight into RNA polymerase I promoter recognition and
364 promoter melting. *Nat Commun* 2019; **10**:5543.

365 34 Da LT, Pardo-Avila F, Xu L *et al.* Bridge helix bending promotes RNA polymerase II backtracking through
366 a critical and conserved threonine residue. *Nat Commun* 2016; **7**:11244.

367 35 Komissarova N, Kashlev M. RNA Polymerase Switches between Inactivated and Activated States By
368 Translocating Back and Forth along the DNA and the RNA. *Journal of Biological Chemistry* 1997;
369 **272**:15329-15338.

370 36 Sanz-Murillo M, Xu J, Belogurov GA *et al.* Structural basis of RNA polymerase I stalling at UV light-
371 induced DNA damage. *Proc Natl Acad Sci U S A* 2018; **115**:8972-8977.

372 37 Cheung AC, Cramer P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*
373 2011; **471**:249-253.

374 38 Alic N, Ayoub N, Landrieux E *et al.* Selectivity and proofreading both contribute significantly to the
375 fidelity of RNA polymerase III transcription. *Proceedings of the National Academy of Sciences of the United*
376 *States of America* 2007; **104**:10400-10405.

377 39 Sydow JF, Brueckner F, Cheung AC *et al.* Structural basis of transcription: mismatch-specific fidelity
378 mechanisms and paused RNA polymerase II with frayed RNA. *Mol Cell* 2009; **34**:710-721.

379 40 Wang D, Bushnell DA, Huang X, Westover KD, Levitt M, Kornberg RD. Structural basis of transcription:

380 backtracked RNA polymerase II at 3.4 angstrom resolution. *Science* 2009; **324**:1203-1206.

381 41 Weaver KN, Watt KE, Hufnagel RB *et al.* Acrofacial Dysostosis, Cincinnati Type, a Mandibulofacial
382 Dysostosis Syndrome with Limb Anomalies, Is Caused by POLR1A Dysfunction. *Am J Hum Genet* 2015;
383 **96**:765-774.

384 42 Dauwerse JG, Dixon J, Seland S *et al.* Mutations in genes encoding subunits of RNA polymerases I and
385 III cause Treacher Collins syndrome. *Nat Genet* 2011; **43**:20-22.

386 43 Watt KEN, Neben CL, Hall S, Merrill AE, Trainor PA. tp53-dependent and independent signaling
387 underlies the pathogenesis and possible prevention of Acrofacial Dysostosis-Cincinnati type. *Hum Mol Genet*
388 2018; **27**:2628-2643.

389 44 Sanchez E, Laplace-Builhe B, Mau-Them FT *et al.* POLR1B and neural crest cell anomalies in Treacher
390 Collins syndrome type 4. *Genet Med* 2020; **22**:547-556.

391 45 Kara B, K roglu  , Peltonen K *et al.* Severe neurodegenerative disease in brothers with homozygous
392 mutation in POLR1A. *European Journal of Human Genetics* 2017; **25**:315-323.

393 46 Ide S, Imai R, Ochi H, Maeshima K. Transcriptional suppression of ribosomal DNA with phase separation.
394 *Sci Adv* 2020; **6**.

395 47 Noack Watt KE, Achilleos A, Neben CL, Merrill AE, Trainor PA. The Roles of RNA Polymerase I and III
396 Subunits Polr1c and Polr1d in Craniofacial Development and in Zebrafish Models of Treacher Collins
397 Syndrome. *PLoS Genet* 2016; **12**:e1006187.

398 48 Gibbons JG, Branco AT, Yu S, Lemos B. Ribosomal DNA copy number is coupled with gene expression
399 variation and mitochondrial abundance in humans. *Nat Commun* 2014; **5**:4850.

400 49 Albert B, Leger-Silvestre I, Normand C *et al.* RNA polymerase I-specific subunits promote polymerase
401 clustering to enhance the rRNA gene transcription cycle. *J Cell Biol* 2011; **192**:277-293.

402 50 Dundr M, Hoffmann-Rohrer U, Hu Q *et al.* A kinetic framework for a mammalian RNA polymerase in
403 vivo. *Science* 2002; **298**:1623-1626.

404 51 French SL, Osheim YN, Cioci F, Nomura M, Beyer AL. In exponentially growing *Saccharomyces*
405 *cerevisiae* cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by
406 the number of active genes. *Mol Cell Biol* 2003; **23**:1558-1568.

407 52 Friedrich JK, Panov KI, Cabart P, Russell J, Zomerdijk JC. TBP-TAF complex SL1 directs RNA
408 polymerase I pre-initiation complex formation and stabilizes upstream binding factor at the rDNA promoter.
409 *J Biol Chem* 2005; **280**:29551-29558.

410 53 Panov KI, Friedrich JK, Russell J, Zomerdijk JC. UBF activates RNA polymerase I transcription by
411 stimulating promoter escape. *EMBO J* 2006; **25**:3310-3322.

412 54 Schneider DA. RNA polymerase I activity is regulated at multiple steps in the transcription cycle: recent
413 insights into factors that influence transcription elongation. *Gene* 2012; **493**:176-184.

414 55 Mastronarde DN. Automated electron microscope tomography using robust prediction of specimen
415 movements. *Journal of Structural Biology* 2005; **152**:36-51.

416 56 Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA. MotionCor2: anisotropic correction
417 of beam-induced motion for improved cryo-electron microscopy. *Nature methods* 2017; **14**:331-332.

418 57 Zhang K. Gctf: Real-time CTF determination and correction. *Journal of structural biology* 2016; **193**:1-
419 12.

420 58 Scheres SH. RELION: implementation of a Bayesian approach to cryo-EM structure determination.
421 *Journal of structural biology* 2012; **180**:519-530.

422 59 Kimanius D, Forsberg BO, Scheres SH, Lindahl E. Accelerated cryo-EM structure determination with
423 parallelisation using GPUs in RELION-2. *Elife* 2016; **5**.

424 60 Pettersen EF, Goddard TD, Huang CC *et al.* UCSF Chimera--a visualization system for exploratory
425 research and analysis. *J Comput Chem* 2004; **25**:1605-1612.

426 61 Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta crystallographica* 2004;
427 **60**:2126-2132.

428 62 Adams PD, Grosse-Kunstleve RW, Hung LW *et al.* PHENIX: building new software for automated
429 crystallographic structure determination. *Acta crystallographica* 2002; **58**:1948-1954.

430 63 Amunts A, Brown A, Bai XC *et al.* Structure of the yeast mitochondrial large ribosomal subunit. *Science*
431 2014; **343**:1485-1489.

432 64 Chen VB, Arendall WB, 3rd, Headd JJ *et al.* MolProbity: all-atom structure validation for macromolecular
433 crystallography. *Acta crystallographica* 2010; **66**:12-21.

434 65 DeLano WL. The PyMOL Molecular Graphics System. *on World Wide Web* <http://www.pymol.org> 2002.

435 66 Goddard TD, Huang CC, Meng EC *et al.* UCSF ChimeraX: Meeting modern challenges in visualization
436 and analysis. *Protein Sci* 2018; **27**:14-25.

437 67 Bernecky C, Herzog F, Baumeister W, Plitzko JM, Cramer P. Structure of transcribing mammalian RNA
438 polymerase II. *Nature* 2016; **529**:551-554.

439 68 Wang Q, Li S, Wan F *et al.* Structural insights into transcriptional regulation of human RNA polymerase
440 III. *Nature Structural & Molecular Biology* 2021; **28**:220-227.

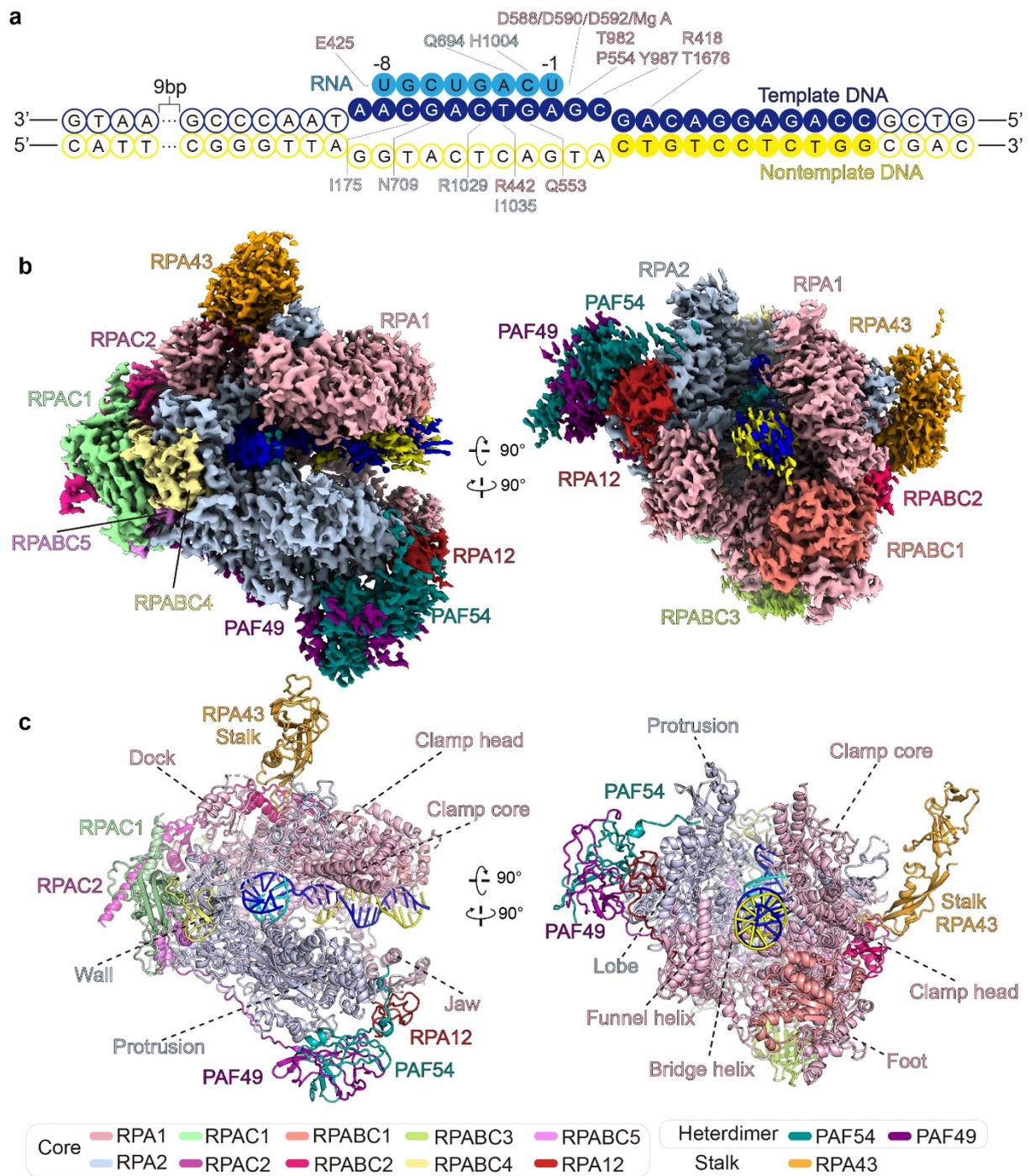
441

442 **Acknowledgments:** We thank the Center of Cryo-Electron Microscopy of Fudan University for the
443 supports on cryo-EM data collection and data analyses. This work was supported by grants from the
444 National key R&D program of China (2016YFA0500700), the National Natural Science Foundation
445 of China (32030055, 31830107, 31821002), the Shanghai Municipal Science and Technology Major
446 Project (2017SHZDZX01), Shanghai Municipal Science and Technology Commission
447 (19JC1411500), the National Ten-Thousand Talent Program (Y. X.), the National Program for
448 support of Top-Notch Young Professionals (Y. X.), and the Strategic Priority Research Program of
449 the Chinese Academy of Sciences (grant no. XDB08000000).

450 **Author contributions:** D. Z. and K. C. prepared the samples for structural analyses and performed
451 biochemical analyses; W. L. and K. C. checked negative staining samples and prepared the cryo-EM
452 samples, and W. L. performed cryo-EM data collection, data processing, and model building; D. Z.,
453 H. Y. and Y. X. analyzed the data, prepared the figures and wrote the manuscript; Y. X. supervised
454 the project.

455 **Competing interests:** Authors declare no competing interests.

456 **Data and materials availability:** The cryo-EM maps and model coordinates will be deposited to
457 the Electron Microscopy Data Band (EMDB) and Protein Data Bank (PDB) upon acceptance of the
458 manuscript.



459

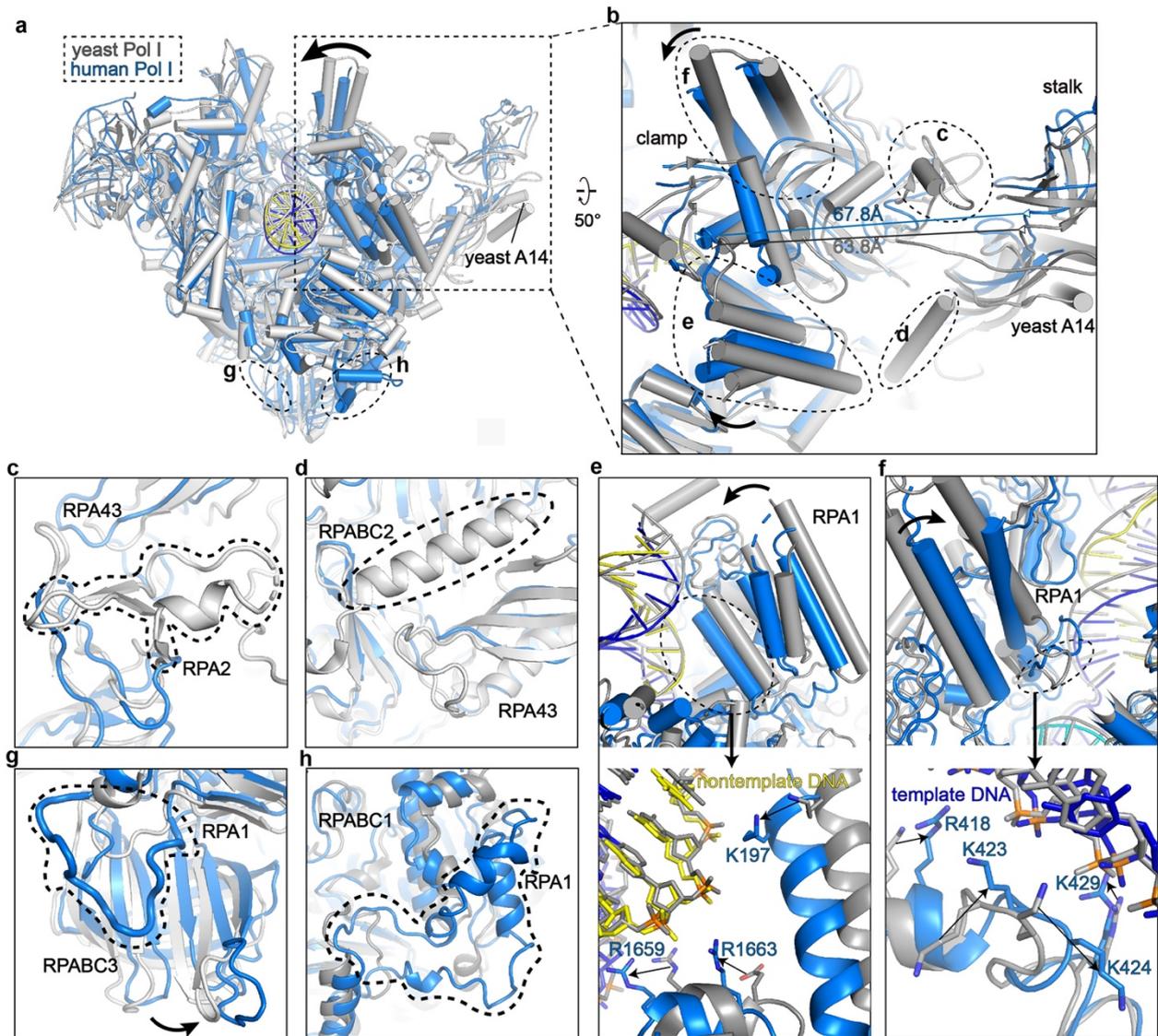
460

Fig. 1. Cryo-EM structure of the human Pol I EC^{post}.

461

(a) Schematic model of the DNA-RNA scaffold used in EC^{post}. RNA is numbered relative to the NTP addition site in the elongation complex. The template DNA is shown in blue, the nontemplate DNA in yellow, and RNA in cyan. The bases that built in our model are shown with color-filled cycles. Residues involved in contacting nucleotides are indicated. (b-c) Cryo-EM map (b) and structural model (c) of the human Pol I EC^{post} in two different views. Color scheme is indicated and used in all the figures below.

466



467

468

Fig. 2. Structural differences between hPol I and yPol I.

469

(a) Structural comparison of hPol I (marine) and yPol I (gray) (PDB: 5M3F)¹⁸ in the EC^{post} state.

470

(b-f) Close-up views of structural differences around the clamp between hPol I and yPol I and the

471

differences are indicated with arrows. Two dashed circles in (c) and (d) indicate the regions in yPol

472

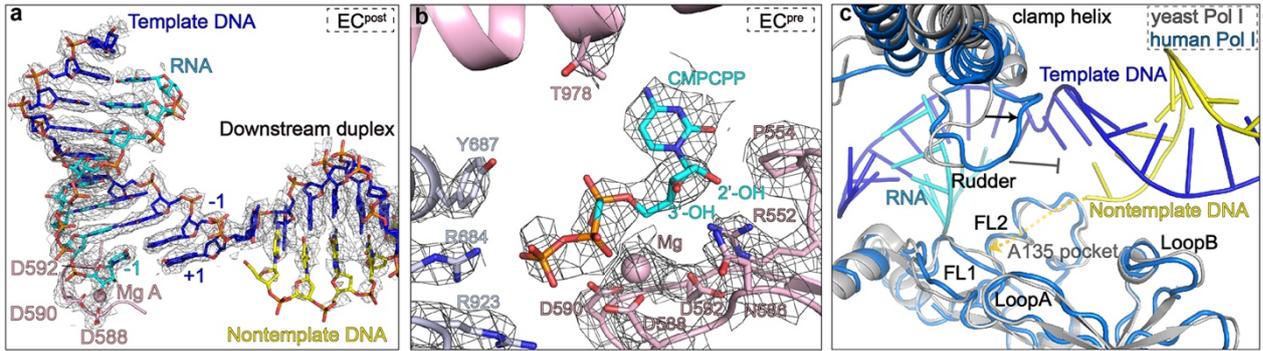
I but not in hPol I. Two different views of the structural difference in the clamp are shown in the top

473

panels of (e) and (f) and more details are shown at the bottom panels. Two dashed circles in (g) and

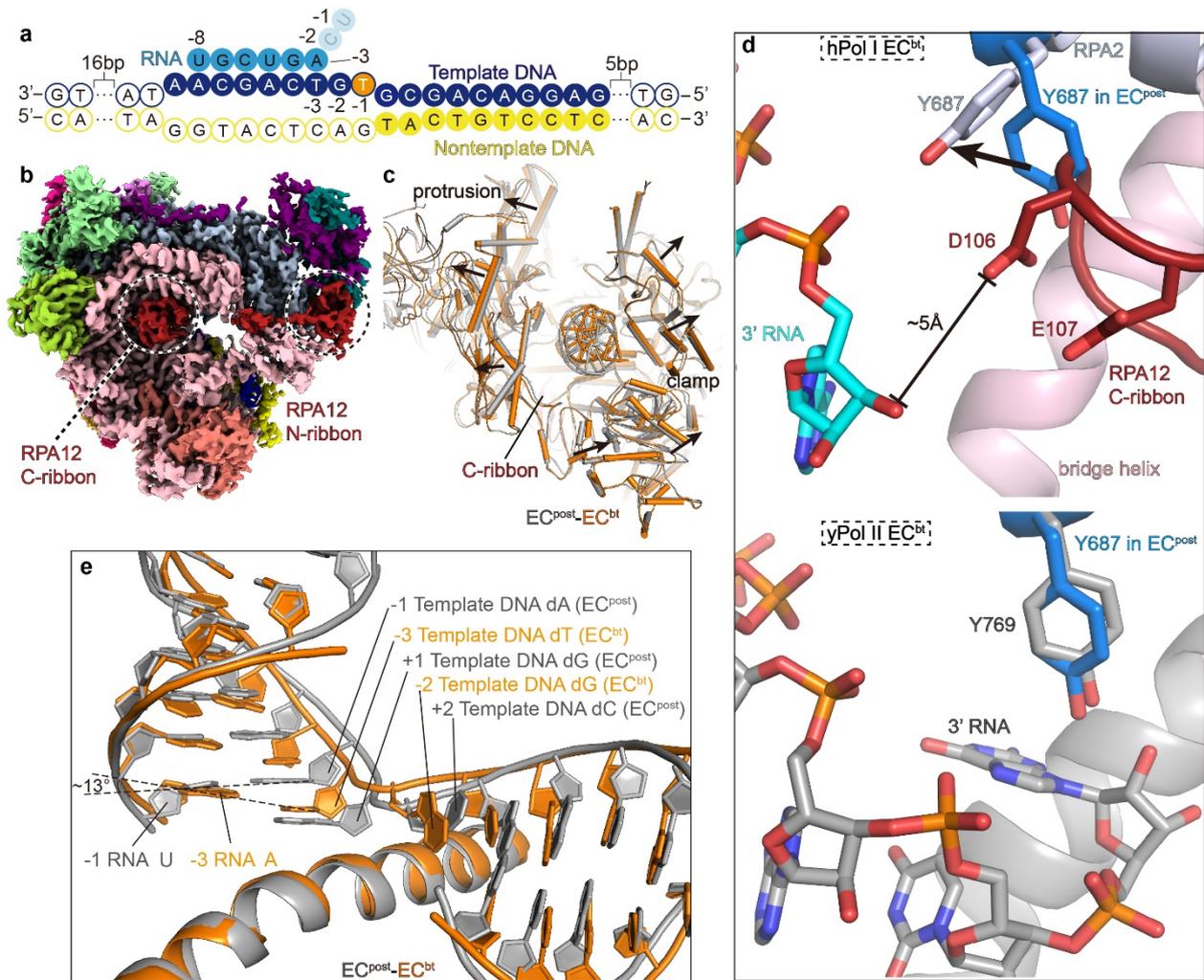
474

(h) indicate the regions of RPA1 that exist in hPol I but not in yPol I.



475
476 **Fig. 3. The catalytic center of Pol I in the post- and pre-translocation states.**

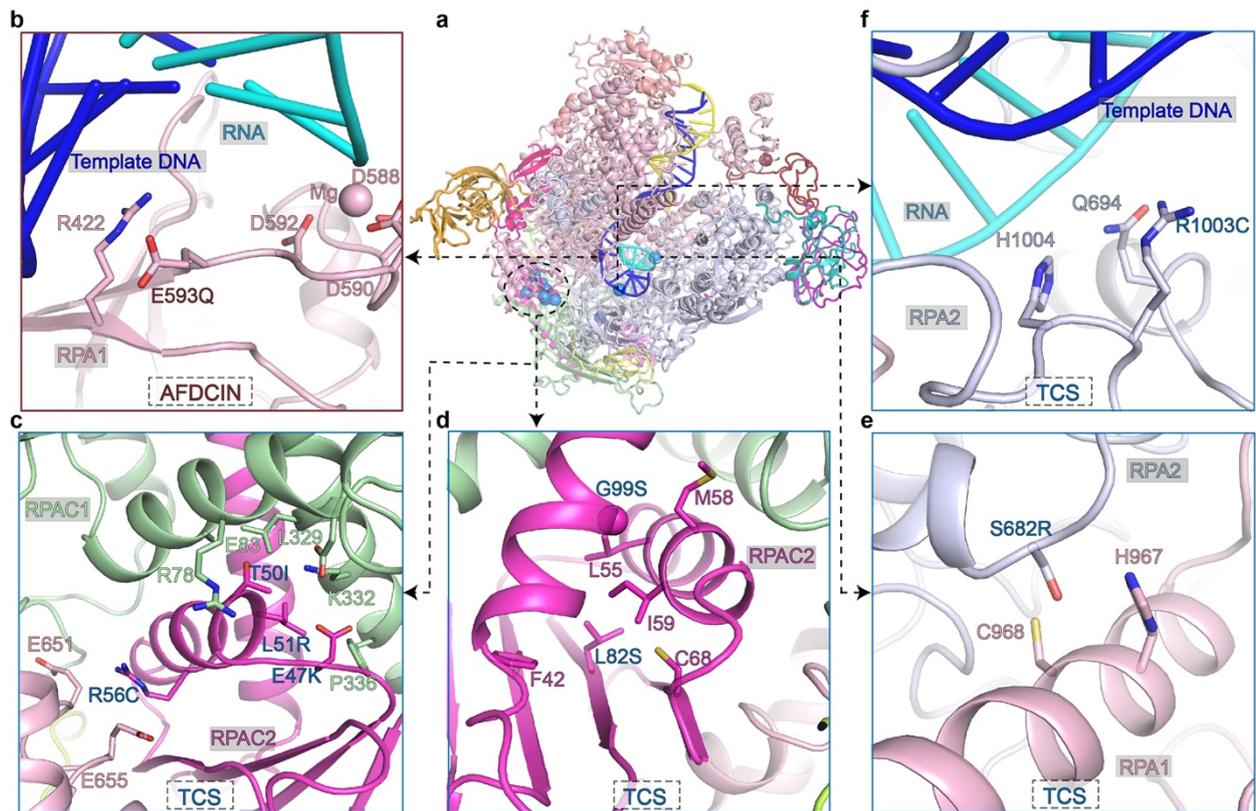
477 **(a)** Cryo-EM map and structural model of the DNA-RNA hybrid in the Pol I EC^{post}. Critical residues
478 involved in catalysis are shown in sticks. The magnesium cation in the metal A position is shown in
479 pink ball. **(b)** Close-up view of the catalytic center of Pol I EC^{pre}. The cryo-EM map is shown in
480 mesh and critical residues are shown in sticks. **(c)** Conformational differences of hPol I (marine) and
481 yPol I (gray) around the transcription fork. The yellow dash line indicates the putative path of the
482 nontemplate strand.



483

484 **Fig. 4. Structure of Pol I EC in the backtracked state.**

485 (a) Schematic diagram of the transcription scaffold used in assembly of Pol I EC^{bt}. The A⁻¹ in
 486 template strand was changed to T⁻¹, generating mismatched base pair of DNA-RNA hybrid. The U⁻¹
 487 and C⁻² of the RNA were cleaved in the assembled EC^{bt} complex. (b) Cryo-EM map of EC^{bt} shows
 488 that the C-ribbon of RPA12 is inserted into the active site. (c) Structure comparison shows that the
 489 cleft of EC^{bt} (orange) is wider than that of EC^{post} (gray). (d) Top panel shows the close-up view of
 490 the tip of RPA12 C-ribbon in the active site and its interaction with the gating tyrosine and 3' end
 491 RNA. Residue Y687 in EC^{post} is shown in sticks and colored in marine. Conformational difference
 492 in residue Y687 in the two states are indicated with black arrow. The bottom panel shows the same
 493 view of yPol II EC^{bt} without TFIIS (gray, PDB: 3GTJ)⁴⁰. Residue Y687 in hPol I EC^{post} is shown
 494 (marine) for comparison. (e) Comparison of the DNA-RNA hybrid in hPol I EC^{bt} (orange) and EC^{post}
 495 (gray).



496

497

Fig. 5. Diseases-associated mutations in human Pol I.

498

(a) Overall structure of the human Pol I EC^{post} with spheres indicating the positions of mutations

499

associated with AFDCIN (red), neurodegenerative disease (black), and TCS (blue). (b-f) Close-up

500

views of the positions of disease-associated mutations. Mutant residues and potential contacting

501

residues are shown in sticks.

502 **Materials and Methods**

503 **Protein expression and purification**

504 The ORFs of 13 subunits of human Pol I were individually subcloned into a modified pCAG vector.
505 PAF53 is the only subunit that is N-terminally Protein A tagged. Except that RPA1 and RPA2 are
506 individually cloned, the expression cassettes of the other 11 subunits were merged into 5 plasmids
507 and all plasmids are co-transfected into Expi293F suspension cells using PEI. After being cultured
508 at 37°C for 72 hours, cells were harvested and lysed in lysis buffer (50 mM HEPES 7.4, 300 mM
509 NaCl, 0.25% CHAPS, 10 μ M ZnCl₂, 5 mM ATP, 5 mM MgCl₂, 10% glycerol, 2 mM DTT, 1 mM
510 PMSF, 1 μ g/mL Benzamidine, 1 μ g/mL Pepstatin and 1 μ g/mL Leupeptin) at 4°C. The supernatant
511 was incubated with IgG beads for 3 hours, and beads was extensively washed with wash buffer
512 (50 mM HEPES pH 7.4, 300 mM NaCl, 0.1% CHAPS, 10% glycerol, 5 mM ATP, 5 mM MgCl₂, 10
513 μ M ZnCl₂ and 2 mM DTT). Protein was digested using Ulp1 protease overnight to remove tags, and
514 the complex was eluted with elution buffer (50 mM HEPES pH 7.4, 300 mM NaCl, 0.1% CHAPS,
515 10% glycerol, 2 mM MgCl₂, 10 μ M ZnCl₂ and 2 mM DTT). The eluted protein was further purified
516 by anion exchange (mono Q) in HEPES 7.4 buffer with increasing concentration of NaCl from 0.1
517 to 1 M. The Pol I complex was eluted at 360 mM NaCl. The complex was concentrated to
518 approximately 2 mg/ml followed by dialysis against low salt buffer (similar to mono Q buffer but
519 containing 150 mM NaCl) overnight. The complex was then flash-frozen in liquid nitrogen and
520 stored at -80°C.

521

522 **Complex assembly**

523 For the preparation of EC^{post}, Pol I was incubated with a 46 base pair transcription scaffold
524 containing an 11-nucleotide mismatched bubble and an 8-nucleotide RNA (Table. S2). The
525 oligonucleotides were dissolved in DEPC H₂O to a final concentration of 100 μ M, mixed in
526 equimolar concentration, heated to 95°C for 5 min and cooled to 20°C at a rate of 1°C per min. To
527 obtain DNA-RNA hybrid, DNA duplex was incubated with a 1.2-fold molar of RNA for 5 min at
528 45°C and then gradually cooled to 4°C. 10 μ L Pol I was incubated with a 1.5-fold molar excess of
529 DNA-RNA scaffold for 10 min at 25°C. Sample was dialyzed against dialysis buffer (25 mM HEPES
530 7.4, 150 mM NaCl, 2 mM MgCl₂, 2 mM DTT) at 4°C using Slide-a-lyzer mini dialysis pins (10,000
531 MW cut-off, Thermo Fisher). The EC^{pre} and EC^{bt} were assembled similarly but with a few
532 modifications. The EC^{pre} was assembled by incubating the EC^{post} with additional 1mM CMPCPP.
533 The EC^{bt} was assembled uniformly as EC^{post}, except with -1 site mismatched template DNA (Table.
534 S2).

535

536 **Transcription assay**

537 Four picomoles of polymerase was incubated for 30 min at 20°C with 4 pmol pre-annealed minimal
538 nucleic-acid scaffold. For RNA elongation, complexes were incubated in the presence of 1 mM
539 NTPs at 28°C for 20 min in transcription buffer (30 mM HEPES 7.5, 100 mM NaCl, 5 mM MgCl₂
540 10 μM ZnCl₂, 10% glycerol, and 2 mM DTT). Reactions were stopped by addition of an equal
541 volume 2×loading buffer (8 M urea, 2×TBE) at different time and incubation for 5 min at 95°C. The
542 FAM-labeled RNA extension products were separated by denaturing gel electrophoresis (0.5 pmol
543 RNA per lane) and visualized with Tanon 4600SF. For RNA cleavage assays, Pol I was incubated
544 with pre-annealed backtrack-scaffold at 16°C, and reactions were stopped at different time and
545 analyzed by gel electrophoresis as above.

546

547 **Cryo-EM sample preparation**

548 For negative staining EM grids preparation, samples (5 μL at a concentration of ~0.035 mg/mL)
549 were applied onto glow-discharged copper grids supported by a continuous thin layer of carbon film
550 for 60 s before negatively stained by 2% (w/v) uranyl formate solution at room temperature. The
551 grids were prepared in the Ar/O₂ mixture for 15 s using a Gatan 950 Solarus plasma cleaning system
552 with a power of 35 W. The negatively stained grids were loaded onto a Thermo Fisher Scientific
553 Talos L120C microscope equipped with a Ceta CCD camera and operating at 120 kV at a nominal
554 magnification of 92,000×, corresponding to a pixel size of 1.58 Å on the specimen.

555 For cryo-EM grids preparation, samples (4 μL at a concentration of ~1.5 mg/mL) were applied to
556 freshly glow-discharged Quantifoil R1.2/1.3 holey gold grids. After incubation 5 s at 4°C and 100%
557 humidity, the grids were blotted for 8.5 s with force 13 in a Thermo Fisher Scientific Vitrobot Mark
558 IV and plunge-frozen in liquid ethane at liquid nitrogen temperature. The grids were prepared in the
559 H₂/O₂ mixture for 20 s using a Gatan 950 Solarus plasma cleaning system with a power of 5 W. The
560 ø 55/20 mm blotting paper (TED PELLA) was used for plunge freezing.

561

562 **Data collection**

563 The cryo-EM grids of three kinds of Pol I EC-state were loaded onto a Thermo Fisher Scientific
564 Titan Krios transmission electron microscope and operated at 300 kV for data collection. All the
565 cryo-EM images were automatically recorded by a Gatan K2 Summit direct electron detector in the
566 super-resolution counting mode using Serial-EM⁵⁵ with a nominal magnification of 130,000× in the
567 NPTEM mode, which yielded a super-resolution pixel size of 0.527 Å on the image plane, and with
568 a defocus ranged from 1.5 to 2.5 μm. Each micrograph stack was dose-fractionated to 32 frames
569 with a total electron dose of ~50 e⁻/Å²⁵⁶ and a total exposure time of 6.94 s. For the dataset of Pol I

570 EC^{pre}, EC^{post}, and EC^{bt} samples, 3,283, 2,074, and 2,854 micrographs were collected for further
571 processing, respectively.

572 The cryo-EM grids of EC^{post}-crosslinking were loaded onto a Thermo Fisher Scientific Arctica
573 transmission electron microscope and operated at 200 kV for data collection. All the cryo-EM
574 images were automatically recorded by a Gatan K3 Summit direct electron detector in the super-
575 resolution counting mode using Serial-EM with a nominal magnification of 36,000× in the TEM
576 mode, which yielded a super-resolution pixel size of 0.55 Å on the image plane, and with a defocus
577 ranged from 1.5 to 2.5 μm. Each micrograph stack was dose-fractionated to 40 frames with a total
578 electron dose of ~50 e⁻/Å² and a total exposure time of 3.009 s. For the dataset of EC^{post}-crosslinking
579 sample, 505 micrographs were collected for further processing.

580

581 **Image processing**

582 For cryo-EM data, drift and beam-induced motion correction were applied on the super-resolution
583 movie stacks using MotionCor2⁵⁶ and binned twofold to a calibrated pixel size of 1.054 Å/pix. The
584 defocus values were estimated by Gctf⁵⁷ from summed images without dose weighting. Other
585 procedures of cryo-EM data processing were performed within RELION v3.0^{58, 59} and cryoSPARC
586 v2^{58, 59} using the dose-weighted micrographs.

587 For the datasets of the Pol I EC^{pre} complex, 1,141,229 particles were picked by automatic particle
588 picking in RELION without reference and subjected to reference-free 2D classification. 946,031
589 particles were selected from good 2D classes for 3D classification in RELION. 382,890 particles
590 were selected from good 3D classes, which were used for the heterogeneous refinement in
591 cryoSPARC and CTF refinement, yielding a reconstruction of Pol I EC^{pre} complex at 2.89 Å
592 resolution.

593 For the datasets of the Pol I EC^{post} complex, 618,806 particles were picked by automatic particle
594 picking in RELION without reference and subjected to reference-free 2D classification. 389,117
595 particles were selected from good 2D classes for 3D classification in RELION. 282,280 particles
596 were selected from good 3D classes, which were used for the heterogeneous refinement in
597 cryoSPARC and CTF refinement, yielding a reconstruction of Pol I EC^{post} complex at 2.81 Å
598 resolution.

599 For the datasets of the Pol I EC^{bt} complex, 676,465 particles were picked by automatic particle
600 picking in RELION without reference and subjected to reference-free 2D classification. 581,698
601 particles were selected from good 2D classes for 3D classification in RELION. 152,653 particles
602 were selected from good 3D classes, which were used for the heterogeneous refinement in
603 cryoSPARC and CTF refinement, yielding a reconstruction of Pol I EC^{bt} complex at 3.01 Å

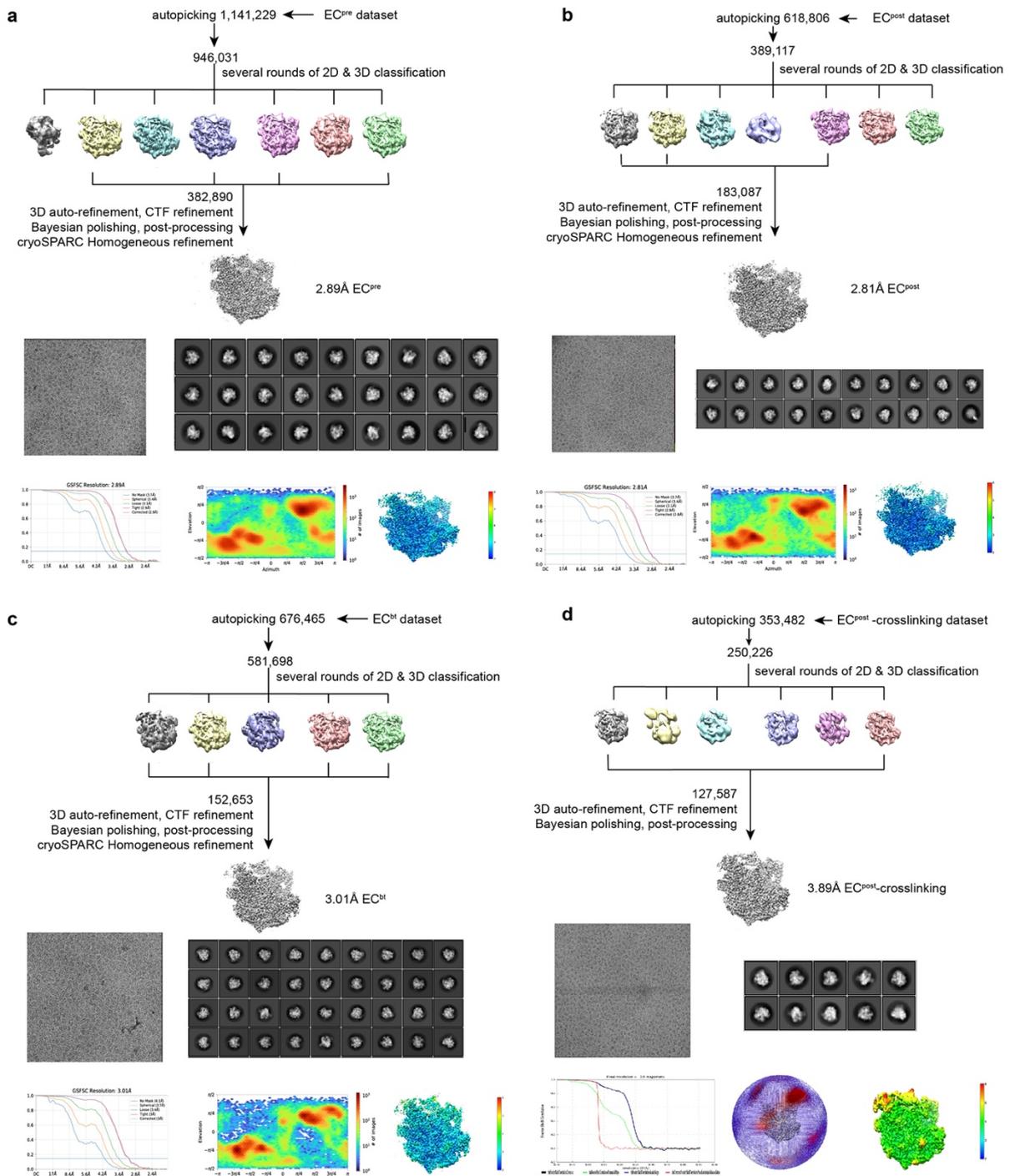
604 resolution.

605 For the datasets of the EC^{post} (crosslinking) complex, 353482 particles were picked by automatic
606 particle picking in RELION without reference and subjected to reference-free 2D classification.
607 250226 particles were selected from good 2D classes for 3D classification in RELION. 127587
608 particles were selected from good 3D classes, which were used for the heterogeneous refinement in
609 cryoSPARC and CTF refinement, yielding a reconstruction of EC^{post} at 3.89 Å resolution
610 All reported resolutions are based on the gold-standard Fourier shell correlation (FSC) = 0.143
611 criterion. The GSFSC curves were corrected for the effects of a soft mask with high-resolution noise
612 substitution. All cryo-EM maps were sharpened by applying a negative B-factor estimation in
613 cryoSPARC Sharpening Tools. All the visualization and evaluation of the 3D volume map were
614 performed within UCSF Chimera or UCSF ChimeraX⁶⁰, and the local resolution variations were
615 calculated using cryoSPARC .

616

617 **Model building and structure refinement**

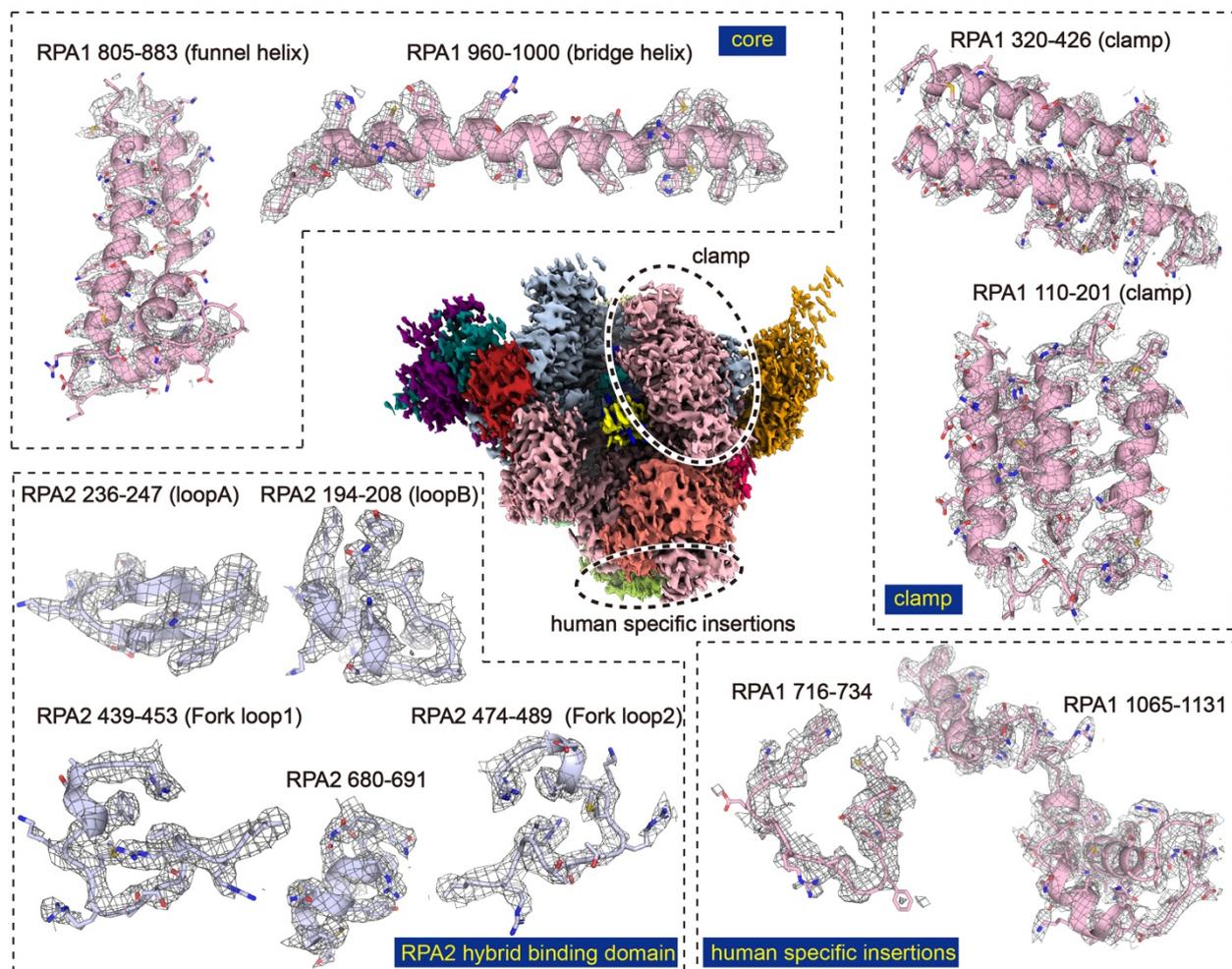
618 The cryo-EM maps of the Pol I EC complexes were used for model fitting. The structures of yeast
619 Pol I EC (PDB: 5M3F)¹⁸ was used as initial structural template, which was docked into the cryo-
620 EM maps by rigid-body fitting using UCSF Chimera⁶⁰. The structural models were built in COOT⁶¹
621 and refined in real space using Phenix⁶² with secondary structure and geometry restraints using the
622 cryo-EM map of the Pol I EC complex. Overfitting of the model was monitored by refining the
623 model in one of the two half maps from the gold-standard refinement approach and testing the
624 refined model against the other map⁶³. Statistics of the map reconstruction and model refinement
625 can be found in Table S1. The final models were evaluated using MolProbity⁶⁴. Maps and model
626 representations in the figures were prepared by PyMOL(<https://pymol.org/>)⁶⁵, UCSF Chimera or
627 UCSF ChimeraX⁶⁶.



639

640 **Supplementary information, Figure S2. Data collection and image processing.**

641 Cryo-EM reconstructions of Pol I EC^{pre} (a), EC^{post} (b), EC^{bt} (c) and EC^{post-crosslinking} (d)
 642 accordingly. In each figure, the top panel is the flow-chart of the cryo-EM image processing; the
 643 middle panels show the representative cryo-EM raw micrograph and 2D classification; the bottom
 644 left panel is the FSC curve of the corresponding map; the bottom middle and right panels are the
 645 local resolution estimation and orientation of the cryo-EM reconstructions of the corresponding map.



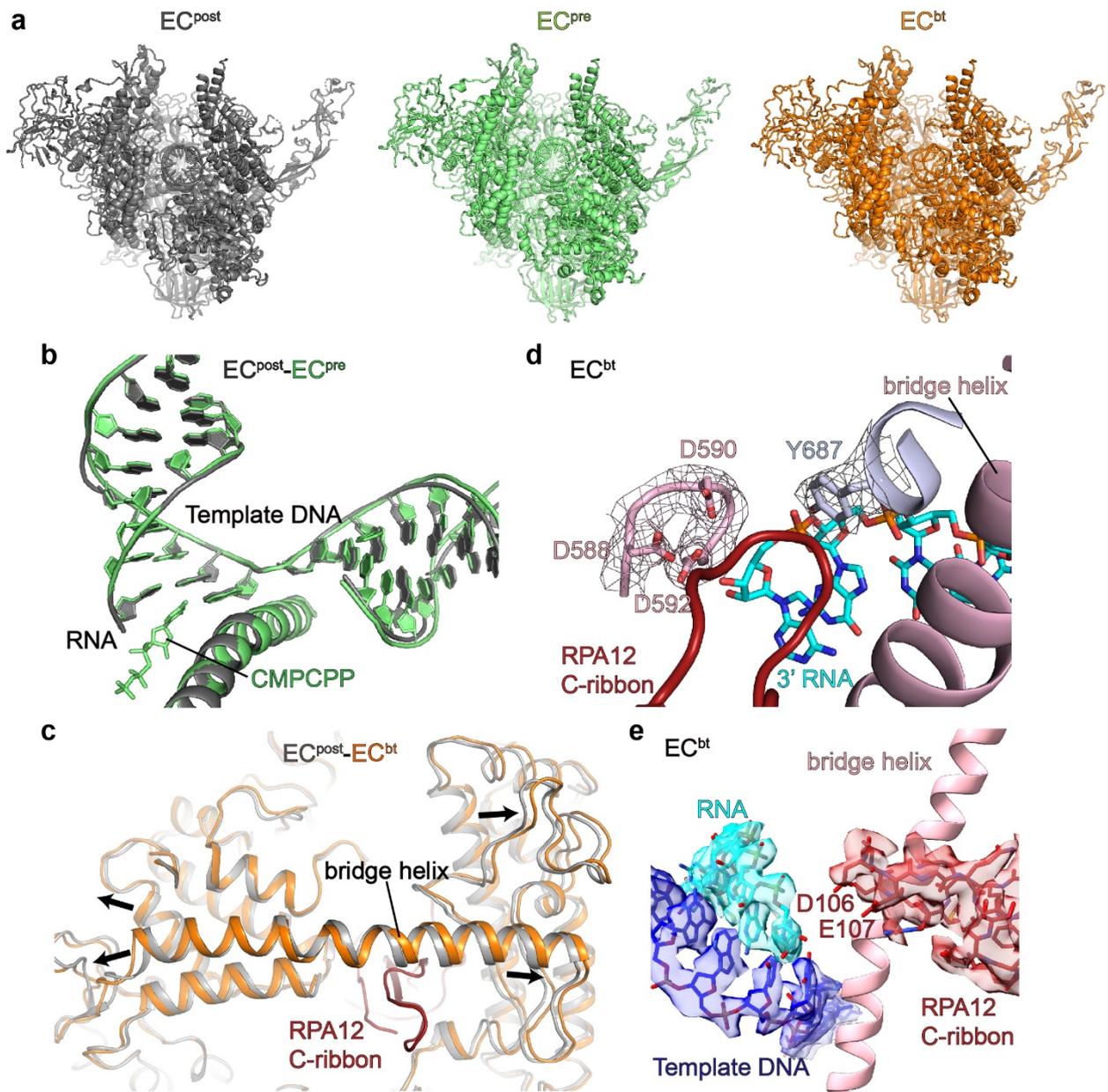
646

647 **Supplementary information, Figure S3. Cryo-EM map and structural model of hPol I EC^{post}.**

648 Overall cryo-EM map of hPol I EC^{post} is shown in the center. The cryo-EM maps of representative

649 regions are shown in mesh and structural models shown in cartoon. Most of the side chains fit in

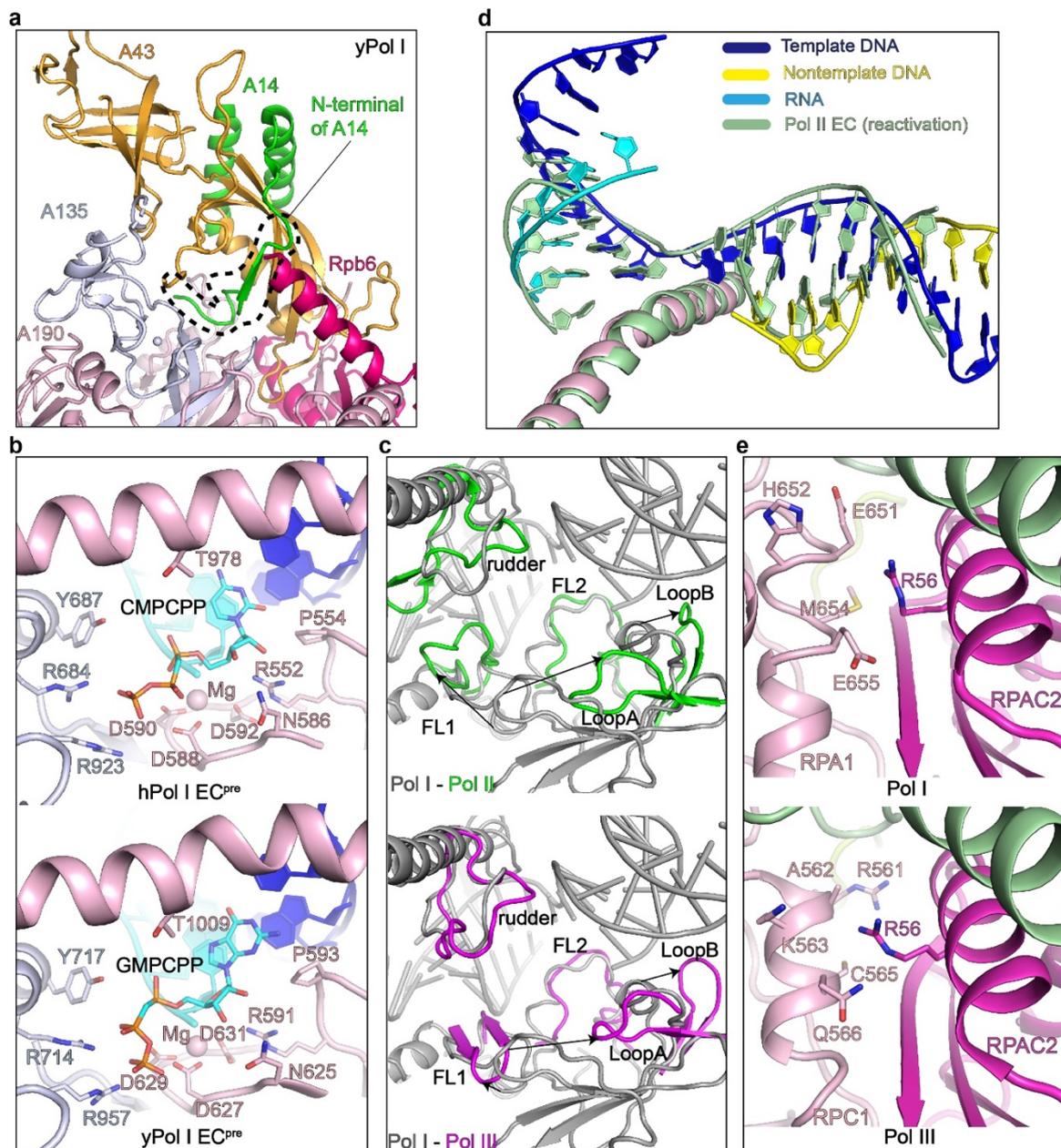
650 corresponding density, indicating the structure was correctly built.



651

652 **Supplementary information, Figure S4. Structural comparison of human Pol I EC in the three**
 653 **states.**

654 **(a)** The structural models of human EC^{post} (gray), EC^{pre} (green) and EC^{bt} (orange). **(b)** Comparison
 655 of the DNA-RNA hybrid in the EC^{post} (gray) and EC^{pre} (green). **(c)** Comparison of the bridge helix
 656 and surround regions in the EC^{post} and EC^{bt}. The EC^{post} is colored in gray and EC^{bt} is colored in
 657 orange except for the RPA12 C-ribbon (red). Structural differences are indicated with arrows. **(d-e)**
 658 Close-up views of the C-ribbon of RPA12 in the active site. The cryo-EM map is shown in mesh (d)
 659 and transparent surface (e), respectively. Most of residues are well-fit into the cryo-EM map.



660

661 **Supplementary information, Figure S5. Structural comparison of hPol I with other RNA**
 662 **polymerases.**

663 (a) Structural model of the stalk of yPol I (PDB: 5M3F)¹⁸. A14 is colored in green and other subunits
 664 are colored as in hPol I in Fig. 1. (b) The catalytic centers are similar in hPol I EC^{pre} and yPol I EC^{pre}
 665 (PDB: 6HKO)¹⁹. Subunits are colored as in Fig. 1 (c) The FL1, FL2, loopA, and loopB in the active
 666 site of hPol I form a narrow tunnel that directs the exit of the nontemplate DNA strand. Pol II (green;
 667 PDB: 5FLM)⁶⁷ and Pol III (purple; PDB: 7D58)⁶⁸ show distinct conformations of these elements.
 668 (d) Comparison of DNA-RNA hybrid in hPol I EC^{bt} in the post-cleavage state and yPol II in the
 669 reactivation state (PDB: 3PO3)³⁷ with color diagram listed. (e) The conserved residue R56 of RPAC2
 670 in Pol I and Pol III (PDB: 7D58)⁶⁸ makes distinct contact with RPA1/PC1 in two enzymes.

a

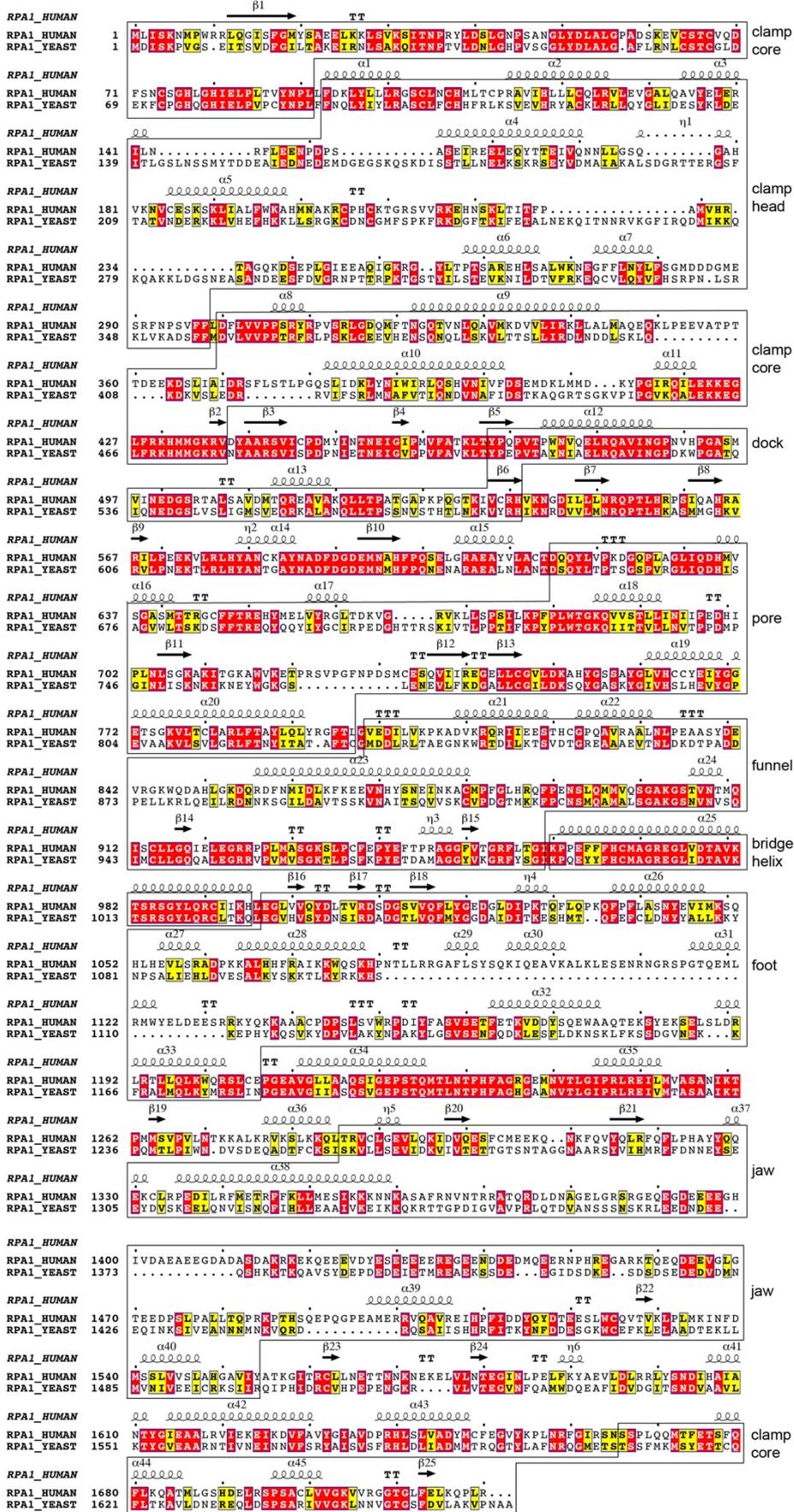


Table S1. Statistics of cryo-EM data collection, refinement and validation statistics.

	EC ^{pre}	EC ^{post}	EC ^{bt}	EC ^{post-crosslinking}
EMDB/PDB				
Data collection and processing				
Magnification	130,000	130,000	130,000	36,000
Voltage (kV)	300	300	300	200
Total electron exposure (e ⁻ /Å ²)	~50	~50	~50	~50
Exposure rate (e ⁻ /pix/s)	~8	~8	~8	~21
Number of frames per movie	32	32	32	40
Automation software	SerialEM	SerialEM	SerialEM	SerialEM
Defocus range (μm)	-1.5 to -2.5	-1.5 to -2.5	-1.5 to -2.5	-1.5 to -2.5
Pixel size (Å)	1.054	1.054	1.054	1.1
Symmetry imposed	C1	C1	C1	C1
Micrographs (no.)	3,283	2,074	2,854	505
Total of extracted particles (no.)	1,141,229	618,806	676,465	353,482
Total of refined particles (no.)	382,890	282,280	152,653	127,587
Local resolution range (Å)	6.0-2.0	6.0-2.0	6.0-2.0	6.0-2.0
Resolution Masked 0.143 FSC (Å)	2.98	2.91	3.16	3.89
Refinement				
Map sharpening B-factor (Å ²)	100.9	83.6	79.4	-170
Initial model used (PDB code)	5M3F	5M3F	5M3F	
Refinement package	Phenix (real space)	Phenix (real space)	Phenix (real space)	
r.m.s. deviations				
Bond lengths (Å)	0.011	0.013	0.013	
Bond angles (°)	1.181	1.035	1.04	
Validation				
MolProbity score	2.46	2.5	2.51	
All-atom clashscore	17.66	20.11	19.89	
Rotamers outliers (%)	0.62	1.05	0.19	
Cβ outliers (%)	0	0	0	
CaBLAM outliers (%)	10.1	10.47	10.08	
B-factors (min/max/mean)				
Protein	13.95/133.07/53.2	17.91/199.99/72.39	14.30/186.28/79.98	
Ligand	28.47/149.49/84.81	28.78/242.64/147.72	60.15/186.26/136.00	
Overall correlation coefficients				
CC (mask)	0.79	0.75	0.74	
CC (peaks)	0.63	0.57	0.56	
CC (volume)	0.76	0.73	0.72	
Ramachandran plot statistics				
Favored (%)	81.5	82.35	81.4	
Allowed (%)	17.23	16.61	17.72	
Disallowed (%)	1.27	1.05	0.89	

678 **Table S2. RNA and DNA oligonucleotides**

EC ^{post/pre} assembly	
nontemplate DNA	5'-CATTTTGGGCCCGCCGGGTTAGGTACTCAGTACTGTCCTCTGGCGAC-3'
template DNA	3'-GTAAAACCCGGCGGCCCAATAACGACTGAGCGACAGGAGACCGCTG-5'
RNA	5'-UGCUGACU-3'
EC ^{bt} assembly	
nontemplate DNA	5'-CATTTTGGGCCCGCCGGGTTAGGTACTCAGTACTGTCCTCTGGCGAC-3'
template DNA	3'-GTAAAACCCGGCGGCCCAATAACGACTGTGCGACAGGAGACCGCTG-5'
RNA	5'-UGCUGACU-3'
Extension assay	
nontemplate DNA	5'-GTACTGTCCTCTGGAC-3'
template DNA	3'-ATAACGACTGAGCGACAGGAGACCTG-5'
RNA	5'-FAM-GUGCUGACU-3'
Cleavage assay	
nontemplate DNA	5'-CATTTTGGGCCCGCCGGGTTAGGTACTCAGTACTGTCCTCTGGCGAC-3'
template DNA	3'-GTAAAACCCGGCGGCCCAATAACGACTGTGCGACAGGAGACCGCTG-5'
RNA	5'-FAM-GUGCUGACU-3'

679

680 **Table S3. Disordered regions that were not modeled in the structure.**

RPA1	1-5, 282-289, 315-317, 525-532, 1227-1238, 1302-1312, 1363-1495
RPA2	1-4, 1085-1092
RPAC1	1-7, 344-346
RPAC2	1-20, 129-133
RPA43	1-45, 205-338
PAF53	1-28, 136-419
PAF49	1-9, 28-34, 59-63, 167-510
RPA12	1-5, 67-79
RPABC1	1-5, 50-55, 211-215
RPABC2	1-50
RPABC3	1-2, 149-150
RPABC4	1-13
RPABC5	65-67

681

682

683 **Supplementary information, Movie S1** Cryo-EM map and structural model of the Pol I EC^{post}.

684 Color scheme is same as Fig. 1.

685 **Supplementary information, Movie S2** Structure comparison of hPol I EC and yPol I EC as in Fig.

686 2.

687

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [MovieS2com2.mov](#)
- [MovieS1com2.mov](#)