

Structural basis of RNA polymerase inhibition by viral and host factors

Finn Werner (✉ f.werner@ucl.ac.uk)

University College London <https://orcid.org/0000-0002-3930-3821>

Simona Pilotto

University College London <https://orcid.org/0000-0002-3251-7376>

Thomas Fouqueau

University College London

Natalya Lukoyanova

Birkbeck College <https://orcid.org/0000-0002-2051-0812>

Carol Sheppard

Imperial College London <https://orcid.org/0000-0002-3285-3804>

Soizick Lucas-Staat

Institut Pasteur

Luis Miguel Díaz-Santín

University College London

Dorota Matelska

University College London <https://orcid.org/0000-0003-0798-3479>

David Prangishvili

Institut Pasteur

Alan Cheung

University of Bristol <https://orcid.org/0000-0001-6430-5127>

Article

Keywords: RNA polymerases, virology, allosteric RNAP inhibition

Posted Date: June 11th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-591334/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 Nature Communications on September 17th, 2021. See the published version at <https://doi.org/10.1038/s41467-021-25666-5>.

1 Structural basis of RNA polymerase inhibition by viral 2 and host factors

3 Simona Pilotto¹, Thomas Fouqueau¹, Natalya Lukoyanova², Carol Sheppard³, Soizick Lucas-
4 Staat⁴, Luis Miguel Díaz-Santín², Dorota Matelska¹, David Prangishvili⁵, Alan C. M. Cheung⁶,
5 and Finn Werner^{1,*}

6 ¹RNAP laboratory, Institute for Structural and Molecular Biology, University College London,
7 Gower Street, WC1E 6BT London, United Kingdom

8 ²Institute for Structural and Molecular Biology, Birkbeck College, Malet Street WC1E 7HX
9 London, United Kingdom

10 ³Section of Virology, Department of Infectious disease, Imperial College London, W2 1PG,
11 London, United Kingdom

12 ⁴Department of Microbiology, Institut Pasteur, 75015 Paris, France

13 ⁵Ivane Javakhishvili Tbilisi State University, 0179 Tbilisi, Georgia

14 ⁶School of Biochemistry, University of Bristol, Bristol, BS8 1TD, United Kingdom

15 * correspondence: f.werner@ucl.ac.uk

16 Abstract

17 The inhibition of RNA polymerases activity plays an important role in the regulation of
18 transcription in response to environmental changes and in the virus-host relationship. Here
19 we present the high-resolution structures of two such RNAP-inhibitor complexes that provide
20 the structural basis underlying RNAP inhibition in archaea. The Acidianus two-tailed virus
21 (ATV) encodes the RIP factor that binds to the inside the DNA-binding channel of RNAP,
22 inhibiting transcription by occlusion of binding sites for nucleic acid and the transcription
23 initiation factor TFB. Infection with the Sulfolobus Turreted Icosahedral Virus (STIV) induces
24 the expression of the host factor TFS4, which binds in the RNAP secondary channel similarly
25 to eukaryotic transcript cleavage factors. In contrast to RIP, TFS4 binding allosterically induces
26 a widening of the DNA binding channel which disrupts trigger loop and bridge helix motifs.
27 Importantly, the conformational changes induced by TFS4 are closely related to inactivated

28 states of RNAP in other domains of life indicating a deep evolutionary conservation of
29 allosteric RNAP inhibition.

30

31 Introduction

32 RNA polymerases (RNAPs) are the engines of transcription and an important target for the
33 regulation of gene expression relevant for health and disease. Gene-specific factors regulate
34 transcription dependent on promoter DNA sequence motifs, while global repression or
35 attenuation of all RNAs is enabled by regulators that directly interact with RNAP
36 independently of the template DNA ^{1,2}. The underlying mechanisms of RNAP inhibition are
37 not only of academic interest but also in the context of the emerging antibiotic resistance
38 crisis, since many frontline antibiotics including Rifampicin are RNAP inhibitors ³. By exploring
39 the structural basis of RNAP inhibition we not only learn about the mechanisms of
40 transcription, but also inform the rational design of novel antibiotics ⁴. In this context, viral
41 transcription factors that inhibit the host RNAP are a rich hunting ground. Viruses are the most
42 abundant pathogens found in nature and important drivers of evolution because of their
43 ability to facilitate horizontal gene transfer ⁵. Once viruses have infected a cell, they are
44 subjected to similar needs: to bypass cellular immunity and to rewire the host gene expression
45 machinery for their own benefit – chiefly to produce virus particles.

46 Well studied bacteriophages like T7 prevent the expression of host genes by inhibiting the
47 host transcription machinery and replace it with virus-encoded components that exclusively
48 transcribe viral genes. The bacteriophage T7 factors gp2 and P7 target the RNAP β' subunit
49 directly and not only interfere with σ^{70} binding but also alter the conformational dynamics of
50 RNAP during 'open complex' formation and DNA melting, an essential step during early
51 transcription initiation ⁶⁻⁸. Inhibition of host transcription is an obvious strategy for viruses that
52 encode their own RNAP, but archaeal viruses are entirely dependent on the host transcription
53 machinery, as their genomes do not encode recognisable RNAP subunit genes. Hence, viral
54 promoters utilise host promoter motifs which are active in *in vitro* transcription experiments
55 using host RNAP and general initiation factors ⁹. We have recently characterised the molecular
56 mechanisms of the highly toxic RNAP inhibitor RIP (RNAP Inhibitory Protein, aka ORF145)
57 encoded by the Acidianus Two-tailed Virus (ATV) virus that infects crenarchaea including

58 *Sulfolobus* species^{2,10}. RIP has undergone a fascinating functional diversification; it is
59 evolutionary related to a capsid protein ORF131, but it binds RNAP with high affinity and
60 efficiently inhibits transcription initiation and elongation¹⁰. But the global inhibition of RNAP
61 is not restricted to virally-encoded factors, as a plethora of cellular factors bind directly to and
62 inhibit RNAP in response to unfavourable environmental conditions including stress, the
63 stationary growth phase and heat shock. The host encoded and constitutively expressed Gfh1
64 (Gre-factor homologue-1)^{11,12} and DksA (DnaK suppressor A)¹³⁻¹⁵ are inhibitors of bacterial
65 RNAP that enable the fine-tuning of transcription, and belong to the group of factors that act
66 through the NTPs entry funnel and pore of RNAP, or secondary channel¹⁵. These also include
67 positive regulators such as TFIS^{16,17}, TFS^{18,19} and GreA/B^{20,21} that enhance elongation and
68 resolve stalled and backtracked transcription elongation complexes (TEC) by transcript
69 cleavage. We have previously described the molecular mechanisms of the archaeal TFS
70 paralogue TFS4, a potent RNAP inhibitor²². TFS4 destabilises RNAP-DNA complexes and
71 inhibits catalysis by decreasing the affinity for NTP substrates²². Intriguingly, TFS4 cannot be
72 detected in the cell during normal growth conditions, but it is strongly upregulated in
73 response to viral infection with STIV (*Sulfolobus* Turreted Icosahedral Virus)²³.
74 We have applied single particle cryo-electron microscopy (cryo-EM) to characterise the
75 structural basis of RNAP inhibition in archaea. Here, we present the high-resolution structures
76 of the inhibitors RIP and TFS4 in complex with the 13-subunit RNAP from *Sulfolobus*
77 *acidocaldarius*. Our results reveal intricate interaction networks between inhibitors and RNAP
78 that rationalise their mode of action. RIP binds to the RNAP clamp inside the DNA binding
79 channel and interferes with transcription by occluding DNA and transcription factors binding
80 sites. Intriguingly the C-terminal tail of RIP interacts with the RNAP similarly to the B-linker of
81 the basal factor TFB (homologous to TFIIIB in eukaryotes). In contrast, TFS4 interacts with
82 RNAP through the secondary channel. Rather than interfering with the binding of transcription
83 factors or nucleic acids, TFS4 inhibits RNAP by inducing large-scale conformational changes
84 that result in the opening of the DNA-binding channel, and the disruption of active site motifs
85 including the bridge helix and the trigger loop. As similar structural perturbations occur in
86 other inhibited states of RNAP, our results demonstrate that the allosteric inhibition of the
87 RNA polymerase is evolutionary conserved across all domains of life.

88 Results

89 The cryo-EM structure of the Saci apo-RNAP

90 We first determined the cryo-EM structure of the apo-RNAP from *Sulfolobus acidocaldarius*
 91 (Saci) at 2.9 Å resolution (Table 1 and figure S1).

92

	apo-RNAP	RNAP/RIP	RNAP/TFS4	RPO4/7 ^a
Data collection and processing				
Microscope	Titan Krios	Titan Krios	Titan Krios	-
Voltage (kv)	300	300	300	-
Detector	K3	K2 Summit	K3	-
Electron exposure (e/Å ² /frame)	1.1115	1.01	1.13	-
Defocus range (µm)	1.3 - 2.5	1.5 - 3.5	1.5 - 3	-
Data collection mode	Super-resolution 30° tilt	Counting 30° tilt	Super-resolution	-
Physical pixel size (Å/pixel)	1.085	1.047	1.085	-
Symmetry imposed	C1	C1	C1	-
Initial particle images	1,286,432	600,640	1,161,535	-
Final particle images	423,157	151,237	350,682	-
Map resolution (Å) ^b	2.88	3.27	2.61	3.75
Map sharpening B-factor (Å ²)	-57.7772	-81.4405	-32.6741	-81.6388
Refinement				
Model composition				
Chains	13	14	14	2
Residues	3234	3373	3305	290
Ligands	1 Mg 6 Zn 1 3Fe-4S	1 Mg 6 Zn 1 3Fe-4S	1 Mg 8 Zn 1 3Fe-4S	none
Map to model cc score	0.85	0.84	0.83	0.81
Molprobit statistics				
Clashscore	5.66	9.58	4.67	9.82
Ramachandran favoured (%)	96.50	97.00	97.12	95.80
Ramachandran outliers (%)	0.00	0.00	0.00	0.00
Bond lengths (%)	0.03 ^c	0.03 ^c	0.03 ^c	0.00
Bond angles (%)	0.04 ^c	0.04 ^c	0.04 ^c	0.00
Rotamer outliers (%)	0.00	0.00	0.00	0.00
Cβ outliers (%)	0.00	0.00	0.00	0.00
Molprobit score	1.54	1.68	1.40	1.80

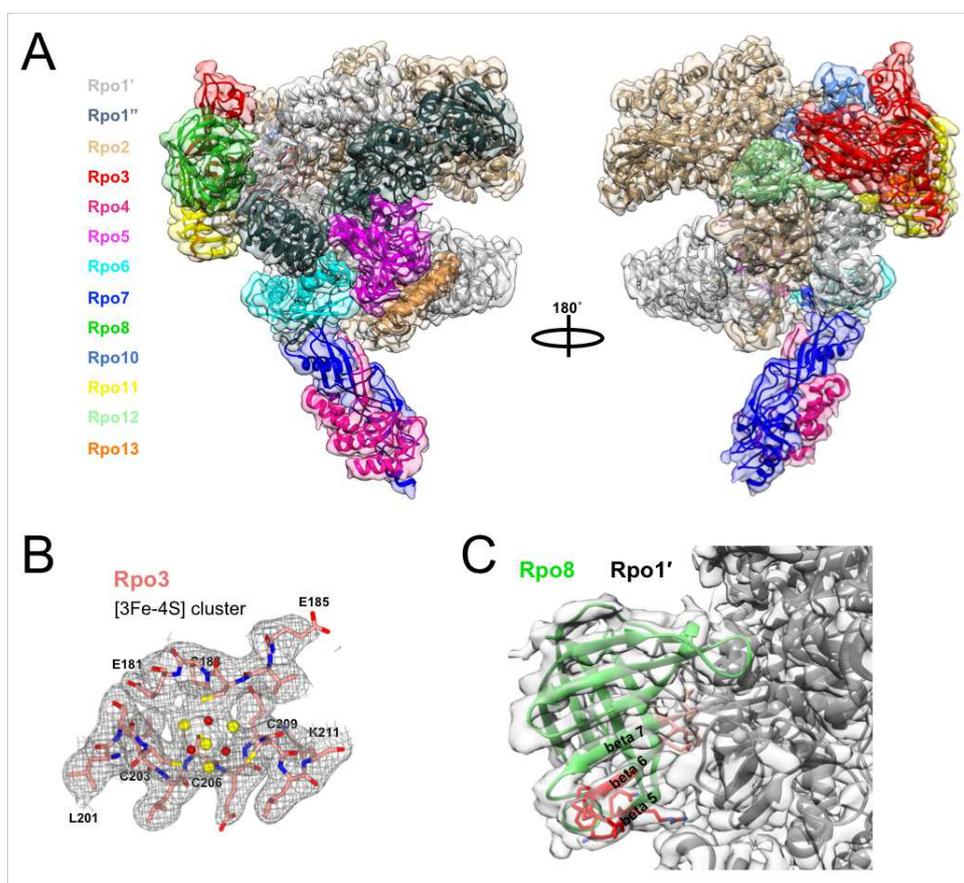
CaBLAM outliers (%)	2.2	1.4	1.9	2.5
CA geometry outliers (%)	0.69	0.39	0.59	1.42
^a From multibody refinement of RNAP/TFS4 dataset				
^b Gold-standard FSC 0.143 cutoff criteria				
^c It refers to the Fe ₃ S ₄ cluster				

93 **Table 1. Cryo-EM data collection and refinement statistics**

94

95 The Saci RNAP consists of 13 subunits (Rpo1-13, figure 1A) with a molecular weight of 405
96 kDa. The active site environment is well resolved, with a clearly defined bridge helix, trigger
97 loop, and the catalytic aspartate triad (Figure 5B and E). The overall structure displays a strong
98 correspondence to crenarchaeal Sso and Ssh RNAPs and eukaryotic RNAPII (figure S2).

99



100

101 **Figure 1. The complete structure of the crenarchaeal RNA polymerase.** (A) Cryo-EM map and cartoon
102 representation of *Sulfolobus acidocaldarius* RNA polymerase. This composite map is illustrated as
103 semi-transparent surface and obtained by merging the cryo-EM structure of apo-RNA polymerase with
104 the stalk resulting from the multi-body refinement of the TFS4-bound RNA polymerase. Each subunit
105 is highlighted according to the colour code shown on the left, the cryo-EM map is illustrated as semi-
106 transparent surface. (B) The Rpo3 [3Fe-4S] ligand is shown as stick and spheres, sulphur atoms are

107 shown in yellow and the iron in dark red. (C) Cryo-EM map and cartoon representation of Rpo8. The
108 β 5–6 loop is highlighted in red, and the conserved GGLLM motif in orange. Corresponding sequence
109 alignment reported in figure S3b)

110

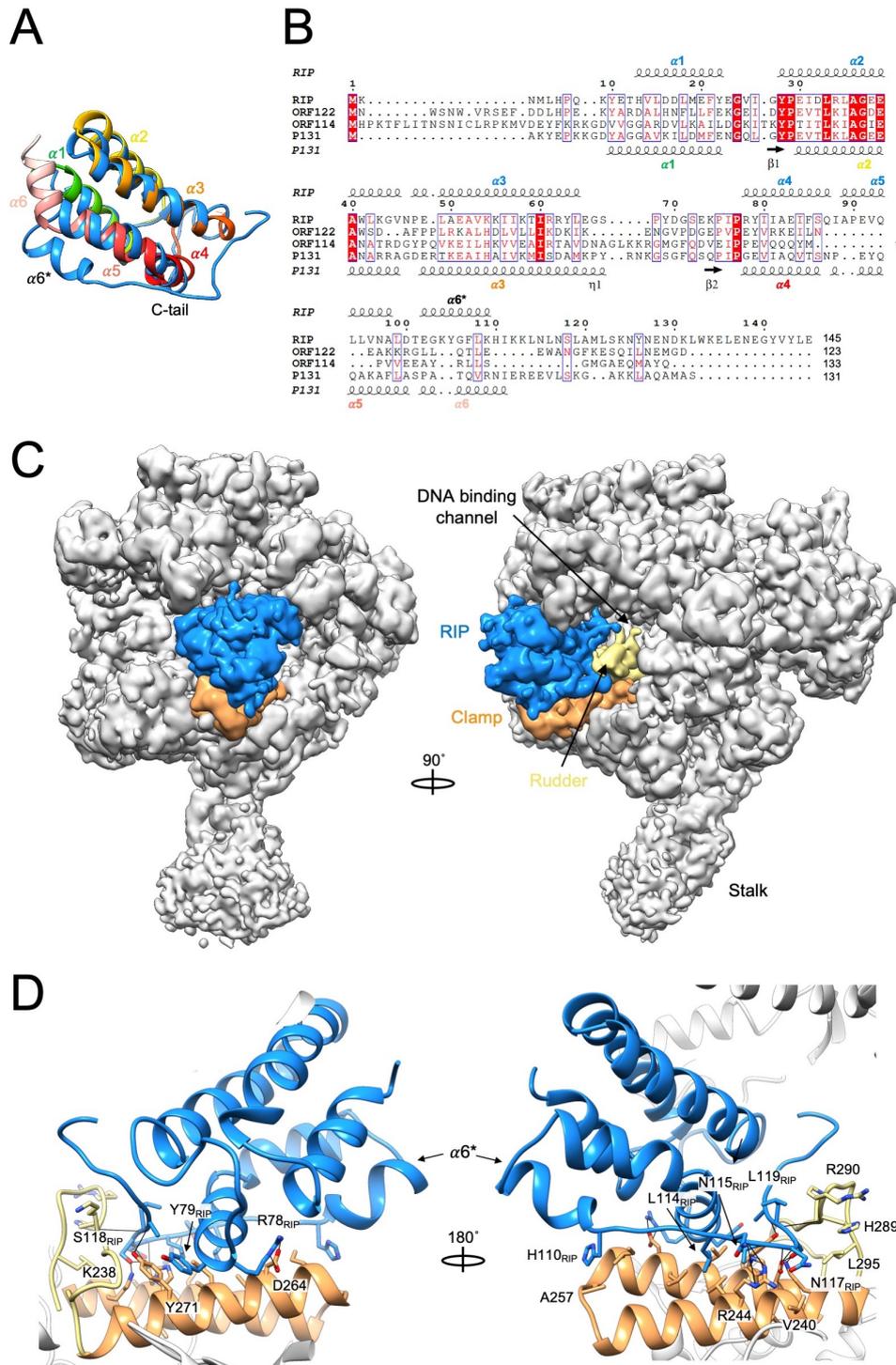
111 The Saci RNAP structure includes eight conserved metal centres, six zinc coordinating motifs,
112 a redox-inactive²⁴ cubane [Fe3-S4] iron-sulphur cluster, as well as the catalytic magnesium ion
113 (Mg^A) that is characteristic for multisubunit, double-psi beta barrel (DPBB) RNAPs (Figures 1B,
114 S2, and S3a). The high resolution enabled us to identify five prolines in the *cis* configuration
115 and an additional zinc finger domain not previously reported (Figure S3a). Based on our map,
116 we rebuilt the structure of Rpo8 starting from loop β 5-6 which led to the repositioning of the
117 conserved GGLLM motif²⁵ at the interface with Rpo1', as seen in all eukaryotic RPB8 (Figure
118 1C; supplementary material and figure S3b).

119

120 [ATV RIP adopts a five-helical bundle structure](#)

121 In order to gain insights into the structural basis of RIP function, and to identify the accurate
122 binding mode of RIP, we solved the cryo-EM structure of the RNAP-RIP complex (Figure 2). A
123 10-fold excess of bacterially expressed recombinant RIP was incubated with Saci RNA
124 polymerase and crosslinked with BS³ (see methods). Data were collected using the same
125 approach used for the apo-RNAP and the processing provided a map with a final map
126 resolution of 3.3 Å (Figure 2 and S4, statistics reported in Table 1).

127



128

129 Figure 2. RIP forms a plug on top the DNA binding channel of RNAP. (A) Structure superimposition of
 130 RIP and its viral relative ORF131/P131 (pdb code 3faj). RIP is shown in blue and ORF131/P131 in a
 131 gradient of colours from the N-terminus in green to C-terminus in light pink with the helix numbering
 132 indicated. The orientation of the RIP $\alpha 6$ helix (indicated with an asterisk) diverges from P131. (B) The
 133 corresponding multiple sequence alignment of viral proteins closely related to RIP has been performed
 134 in PROMALS3D. They include SMV-1 ORF122 and ORF114, as well as the ATV P131 (ORF131). The

135 secondary structure annotation generated by Esprpt3 for RIP and P131 are shown, respectively, above
136 and below the sequence alignment. (C) The cryo-EM map of the Saci RNAP-ATV RIP complex reveals
137 that RIP forms a plug in the DNA binding channel. RIP is coloured in blue, the RNAP clamp and rudder
138 motifs in orange and yellow, respectively. (D) Enlarged view highlighting selected residues of RIP, RNAP
139 clamp head and rudder motifs that are involved in the interaction shown in stick representation. The
140 detailed RIP-RNAP interfaces analysis is illustrated in figure S6a.

141

142 The amino acid sequence of RIP suggests that it is a paralog of the viral P131/ORF131 capsid
143 protein ¹⁰. The ancestral gene of ORF131 underwent gene duplication and speciation and the
144 RIP paralogue acquired the novel function of tightly binding to RNAP and inhibiting its activity
145 ¹⁰. The structure of RIP confirms this close relationship, as both RIP and ORF131 adopt a highly
146 superimposable coil bundle structure with six α -helices and a rmsd of 1.7 Å (Figure 2A). The
147 alpha-helical bundle of RIP is connected to a RIP-specific long C-terminal tail that is not
148 conserved in any of the ORF131 paralogues (Figure 2B) ¹⁰. We could resolve RIP encompassing
149 residues 10-127 (of 145 aa) within the RNAP-RIP complex, and the secondary structure content
150 of RIP is in good agreement with CD-spectra recorded with 'free' RIP which predicted a
151 predominantly alpha-helical structure ¹⁰. The first five α -helices are highly conserved between
152 RIP and ORF131. The insertion of three additional residues, L⁹⁹-D¹⁰⁰-T¹⁰¹, allows the α 6* helix
153 to fold partially back on α 5, and project the C-terminal tail inside the DNA-binding channel of
154 RNAP. The difference in the position and orientation of the sixth α -helix is a crucial feature
155 that allows the tail of RIP to adopt the correct position required for its interaction with RNAP
156 (Figure 2A, and D). We prepared an improved sequence alignment of RIP/ORF131
157 paralogues, informed by our cryo-EM structure, which allowed us to identify the sequence
158 determinants and the structural features responsible for the functional specialisation of RIP.

159

160 RIP forms a plug in the DNA binding channel of RNAP

161 Our recent cross-linking/mass spectrometry studies suggested that RIP interacts with the
162 RNAP clamp inside the DNA-binding channel of the archaeal RNAP ¹⁰. The cryo-EM structure
163 of the RNAP-RIP complex provides the detailed structural basis for this interaction. The
164 compact structure of RIP allows it to fit snugly in the DNA-binding channel of RNAP between
165 the Rpo1' clamp, and the Rpo2 protrusion and lobe motifs, respectively, on each side of the

166 channel (Figure 2C). The C-terminal tail of RIP forms an intricate network of interactions with
167 the RNAP clamp and rudder motifs with an interface area of 1,127 Å² that includes both
168 hydrogen bonds and hydrophobic interactions (Figure 2D, and S6a and c). The RIP tail makes
169 a 90° bend at residue N117 and the region L114-M119 adopts an unusual L-shaped
170 conformation that fits into a pocket between rudder and clamp (Figure 2D). On the opposite
171 side of the DNA binding channel, RIP interacts with the RNAP Rpo2 protrusion and lobe motifs
172 mainly via hydrophobic interactions and with a small interface area of 302 Å². Taken together,
173 the tight interaction network between RIP and RNAP provides a structural rationale for the
174 extreme salt-resistant binding that persists at up to 2M NaCl ¹⁰.

175 The extremity of the RIP tail (128-145) that was not resolved in the structure is enriched in
176 negatively charged residues (6/18) that are not conserved in ORF131; these possibly mimic
177 the negative charge of the DNA template phosphodiester backbone. The importance of the
178 C-terminal tail for RNAP binding is in good agreement with a mutagenesis analysis of RIP¹⁸,
179 which showed that a C-terminal truncation (RIP Δ114-145) abrogated the RNAP binding and
180 inhibitory activities of RIP, while not compromising the extreme heat-stability of the protein ¹⁰.
181 Some of the critical residues of Rpo1' that are involved in RIP binding including K238, R244,
182 D241, H272, and R290 are strictly conserved among archaeal RNAPs but not with bacterial
183 RNAPs (Figure S6d). This is in good agreement with the observations that (i) ATV has a broad
184 archaeal host range, which includes inhibition of the euryarchaeal *M. jannaschii* RNAP *in vitro*
185 but not of the *E. coli* RNAP (CMS and FW, unpublished data), and (ii) the fact that recombinant
186 RIP can be overexpressed in large quantities in *E. coli* while being extremely toxic to
187 *Sulfolobus acidocaldarius* ¹⁰. The flexible RNAP clamp can adopt open and closed states which
188 result in changes in the width of the DNA-binding channel in response to the engagement of
189 RNAP with the DNA template, and binding of initiation and elongation factors, respectively
190 ^{26,27}. Similarly to other DPBB RNAPs, the archaeal enzyme cycles between multiple distinct
191 conformational states as RNAP progresses through the transcription cycle. We have previously
192 applied smFRET to monitor the conformational changes of the *M. jannaschii* RNAP in solution
193 and found that RIP strongly favoured one fixed closed conformation of the clamp ^{9,27}. The RIP-
194 RNAP interaction network that involves both sides of the RNAP DNA binding channel provides
195 a persuasive structural rationale for the nanomolar binding affinity and RIP's ability to lock the
196 clamp in a fixed conformation ⁹.

197

198 Structural determinants of RIP inhibition during transcription initiation and 199 elongation

200 The structure and function of the archaeal preinitiation complex (PIC) is conserved with the
201 eukaryotic RNAPII system; the combination of TBP and TFB is necessary and sufficient to
202 enable RNAP recruitment and start site-specific transcription initiation at basal levels²⁸. TFE
203 activates transcription by inducing conformational changes in RNAP and enhancing DNA
204 strand separation during the closed to open PIC transition²⁹⁻³¹. Electrophoretic mobility shift
205 assay (EMSA) experiments using the *Saccharolobus solfataricus* RNAP and initiation factors
206 demonstrated that RIP interferes with PICs by counteracting their formation and by
207 destabilising preformed PICs¹⁰. A superposition of the RIP-RNAP structure with the closed
208 PIC (from yeast, pdb 6gyk) reveals that the C-terminal tail of RIP (residues 117 - 123) overlaps
209 with the TF(II)B residues 90-120 encompassing the B-linker and B-linker helix by adopting a
210 similar binding mode to the RNAP rudder (Figure 3A and B). Consequently, RIP inhibits the
211 PIC by competing with TFB binding to RNAP. In addition, RIP occludes the binding site to the
212 nucleic acid scaffold in both closed and open PIC (pdb 5iyd), and thus would prevent the
213 loading of the DNA in the RNAP active site (Figure 3C).

214 In contrast to the strong inhibition of initiation factor-dependent and promoter-directed
215 transcription initiation, RIP has a smaller effect on elongation, and EMSAs showed that RIP,
216 albeit in a limited fashion, can bind to TECs consisting of RNAP and a DNA/RNA scaffold¹⁰.
217 A superposition of the RNAP-RIP structure with a eukaryotic TEC (pdb 5oik) reveals that the
218 downstream duplex DNA and the unwound template DNA strand passes underneath RIP
219 while the non-template strand clashes with RIP (Figure 3F and G). To probe this mechanism
220 biochemically, we carried out EMSA experiments with dual labelled components; RIP was
221 radiolabelled using ³²P while the DNA/RNA elongation scaffold was labelled with the
222 fluorescent dye Cy3 on the DNA template strand. The low mobility bands detected by the
223 fluorescence signal in the upper panel of figure 3D correspond to bona fide TECs. The
224 addition of ³²P-RIP counteracts the Cy3-TEC signal in a concentration-dependent manner.
225 Vice versa, the addition of increasing amounts of Cy3-labelled nucleic acid scaffold leads to
226 an increase in the signal corresponding to free ³²P-RIP. In summary, the EMSAs support a

232 surface representation with the RNA polymerase highlighted in grey, RIP in blue, TBP in light purple,
233 TFIIB in red, TFE in light green, and promoter DNA in pink. (B) Enlarged view of the the C-terminal tail
234 of RIP mimicking the TFIIB linker in front of the RNAP rudder motif. (C) Superimposition of the RNAP-
235 RIP complex with the open transcription initiation complex from *Homo sapiens* (5iyb). (D) RIP and
236 DNA/RNA scaffold compete for RNAP binding. Electrophoretic mobility shift assay (EMSA) using Cy3-
237 labelled DNA scaffold (top panel) and ³²P-labelled RIP protein (bottom panel). The (+) signs indicate
238 0.2 μM Saci RNA polymerase, and 0.4 μM RIP. Dose-response titration includes increasing
239 concentrations of RIP (0.2; 1; 2 μM) or DNA/RNA scaffold (0.1; 0.5; 1 μM). (E) Western blot showing the
240 expression of RIP during late infection using polyclonal antibodies raised against RIP and Alba serving
241 as control. (F) Superimposition of RNAP-RIP complex with the *Bos taurus* RNAP-DNA-RNA transcription
242 elongation complex, TEC (pdb 5oik). (G) Enlarged view highlighting the clash of RIP with nucleic acids,
243 in particular the non-template DNA strand (NTS).

244

245 Temporal RIP expression pattern supports its role for viroid maturation

246 The role of RIP for virus fitness and function is still a matter of debate. As RIP directly interferes
247 with nucleic acid binding and affects host and virus promoters alike, the early expression of
248 RIP at high levels appears problematic. However, the global attenuation of transcription could
249 benefit the virus by preventing transcription-dependent host defence mechanisms including
250 the CRISPR IIIb system ³². Alternatively, RIP could disengage RNAP from the actively
251 transcribed viral genome aiding the DNA packaging into virus particles. The former would be
252 associated with an expression pattern during early stage, and the latter during a late infection
253 stage. To address this question we analysed the temporal gene expression of RIP over a time
254 course of 72 hours. Exponentially growing *Acidianus* cells were infected with ATV and samples
255 taken at regular intervals; RIP protein levels were detected by immunoblotting using a
256 polyclonal antibody raised against recombinant RIP and compared to the expression levels of
257 Alba, a chromatin protein that serves as control. In good agreement with its high toxicity, RIP
258 could not be detected in *Acidianus* cell extracts during the early and middle stages of
259 infection but was strongly upregulated at the end of the time course (72 hrs p.i.) just prior to
260 cell lysis that occurs 96 hours post infection (Figure 3E). This suggests that RIP function is
261 important during late infection, e. g. by dissociating RNAPs from the actively transcribed viral
262 genome and thereby assisting maturation of the virion particle.

263

264 TFS4 is a cleavage factor paralog evolved into a repressor

265 Unlike RIP, the TFS4 inhibitor is encoded by the host genome but only expressed in response
266 to viral infection. Like RIP, TFS4 binds tightly to RNAP and efficiently represses transcription
267 ²². In order to investigate the structural basis for TFS4 inhibition and compare it to RIP's, we
268 solved the cryo-EM structure of the RNAP-TFS4 complex at 2.6 Å resolution (Figure 4, S5 and
269 Table 1 for statistics). The map has been further refined by multibody refinement which
270 allowed us to obtain a medium resolution map of the stalk at 3.8 Å where all the known
271 structural features were identified and correctly modelled inside the map (Figure S5f-g and
272 table 1). Archaeal TFS paralogues have a domain configuration akin to A12, RPB9 and C11 of
273 RNAPI, II, and III, respectively, composed of two zinc-ribbon domains, ZR^N and ZR^C, which are
274 connected by a long linker ²² (Figure 4D). As proven by the cryo-EM map, the TFS4 ZR^N
275 interacts with RNAP between the upper jaw and the lobe (Figure 4A and C) ^{16,33}. However,
276 while the linker of A12 binds to a solvent-exposed surface of the RNAPI jaw, TFS4 induces a
277 conformational change that allows its linker region to bind to a surface that is otherwise
278 occluded in apo-RNAP. The insertion of the TFS4 linker likely forces a swinging movement of
279 the jaw and clamp motifs, causing the opening of the DNA-binding channel described below
280 (Movie S1). The N-terminal segment of the TFS4 linker forms two β-addition motifs (β3 and
281 β4) by providing one antiparallel and one parallel strand to two β-sheets in the upper jaw of
282 RNAP (Figure 4C and D). The C-terminal segment (β5) of the linker, unexpectedly, folds on
283 itself and packs on the TFS ZR^C domain that interacts with the rim helices of the funnel (Figure
284 4A-C).

285 The TFS4 ZR^C binds in the NTP-entry funnel of RNAP ^{16,33} in a manner that is related to the
286 eukaryotic RNAPI and III subunits RPA12 and C11 (RPC10 in human) and the RNAPII
287 elongation factor TFIIIS^{34,35,36}, but without reaching through the pore into the active site (Figure
288 4B and C) ²². Importantly, the TFS4 ZR^C domain clashes with and displaces the trigger loop
289 without occluding the pore. The chemical nature of the interactions between the ZR^N and ZR^C
290 domains with RNAP is different (Figure S6b), the ZR^N-jaw/lobe interactions are dominated by
291 a network of hydrophobic interactions, whereas the ZR^C-funnel interactions are facilitated by
292 hydrogen bonds and salt bridges. Intriguingly, two aspartates (D702 and 705) form a
293 negatively charged patch on the rim helices of the RNAP facing the highly positively charged
294 surface which are specific of the TFS4 ZR^C domain and not conserved in the other factors or

295 RNAP subunits (Figure 4B and D). The TFS4-specific positively charged surface may allow
296 alternative binding orientations inside the secondary channel. Site directed mutagenesis has
297 emphasised the importance of three consecutive lysine residues (K76/77/78) in the TFS4 ZR^C
298 domain as determinants for RNAP binding and inhibition. The positively charged surface
299 formed by the three lysines together with arginine 57, 65 and 82 is likely responsible for the
300 initial binding inside the negatively charged channel (Figure 4B-D), which is followed by a
301 remodelling of RNAP and TFS4 itself to enable the insertion of the TFS4 linker and ZR^N
302 between the lobe and jaw elements. The ability of the ZR^C domain of the cleavage
303 transcription factors/subunits to switch binding mode on the RNA polymerase in and out of
304 the NTP entry channel is a well characterised mechanism of regulation during transcription
305 elongation ³⁶. However, TFS4 is the first example of a ZR^C domain that (i) physically clashes
306 with the trigger loop, and (ii) induces conformational changes that allosterically inhibit RNAP
307 (Figure 5).

308

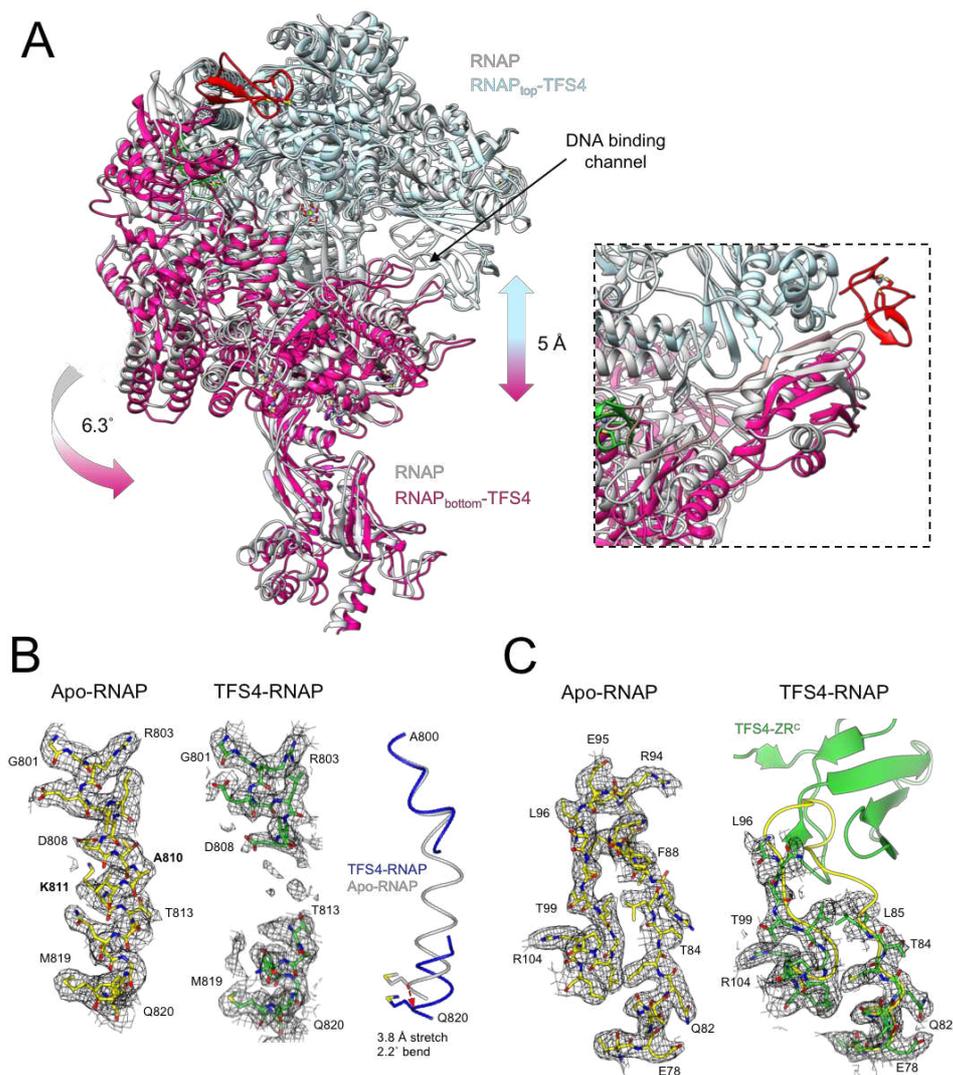
318 implemented with the secondary structure annotation for TFS4 and domains organisation reported
319 below the sequence alignment.

320

321 *Allosteric modus operandi of TFS4*

322 TFS4 overexpression induces retardation of cell growth *in vivo*, and TFS4 is a potent inhibitor
323 of RNAP *in vitro*²², lowering the K_M for substrate NTP binding of ~50-fold and, like RIP,
324 destabilises PICs and TECs²². Unlike RIP however, the binding sites of TFS4 with RNAP do not
325 overlap with any of the RNAP interaction sites of the DNA or transcription initiation and
326 elongation factors and does not occlude the secondary channel. TFS4 allosterically inhibits
327 the RNA polymerase by inducing conformational changes driven by the displacement of the
328 upper jaw, encompassing Rpo1'', Rpo5, 6 and 13, and the Rpo1' clamp head (Figure 5A). The
329 jaw displacement is stabilized within the RNAP structure by the replacement of the two lost
330 beta strands with two newly formed strands provided by TFS4 (enlargement in figure 5A). The
331 superposition between the apo- and the TFS4-bound RNAP reveals a swinging movement of
332 the jaw in unison with the clamp head by 6.3° concomitant with the splaying of the DNA-
333 binding channel by 5 Å (Movie S1). Considering that the width of the DNA double helix is 23.7
334 Å, the observed widening of the DNA-binding channel is likely to impair the close interactions
335 of the DNA template with the RNAP, which provides the structural rationale for TFS4
336 destabilisation of RNAP-nucleic acid interactions²². The Rpo1' bridge helix spans across the
337 DNA binding channel and is 'anchored' on either side of it. The TFS4-induced opening
338 stretches and bends the bridge helix by 3.8 Å and 2.2° leading to the loss of density between
339 residues T808 to K811, likely due to the unwinding of the helix in the TFS4-bound RNAP
340 (Figure 5B). The ensemble of bridge helix and trigger loop plays a key role in the substrate
341 nucleotide binding in the active site and its translocation cycle, i. e. the molecular mechanism
342 underlying transcription elongation³⁷. The TFS4 ZR^C in the NTP entry pore clashes with the tip
343 of the trigger loop resulting in its displacement which is confirmed by the increased disorder
344 visible in the EM map between residues R86 to E95 of Rpo1'' in the RNAP-TFS4 complex
345 (Figure 5C and S7e and f).

346 In summary, the changes induced by TFS4 binding include the (i) widening of the DNA binding
347 channel, (ii) melting of the bridge helix, and (iii) displacement of the trigger loop. This three-
348 pronged attack makes for a formidable intervention with the catalytic mechanisms of RNAP.



349

350 **Figure 5. TFS4-induced allosteric inhibition of RNAP.** (A) Superimposition of the apo-RNAP and TFS4-
 351 bound RNAP illustrating largescale conformational changes resulting in a widening of the DNA-binding
 352 channel of 5 Å. Rpo2 subunits of apo- and TFS4-bound RNAP were superimposed in Chimera and the
 353 moving structural elements are highlighted in hot pink. In the enlargement a focus on the TFS4 linker
 354 position compared to the apo-RNAP showing how it replaces perfectly the beta sheets of the jaw. (B)
 355 Effect of the TFS4 binding on the bridge helix. From left the helix modelled inside the cryo-EM map for
 356 the apo-RNAP and the TFS4-bound one. The latter is shown with a high contour level to highlight the
 357 loss of density in the central section. On the right the superimposition of the bridge helix in the two
 358 structures is presented in worm style to appreciate the extend of the stretch and bending of the helix.
 359 (C) The TFS4 ZR^C domain sterically clashes with the tip of the trigger loop, leading to loss of density,
 360 while the trigger loop is structurally well-resolved in the apo-RNAP.

361 Discussion

362 Common themes of RNAP inhibition in all domains of life

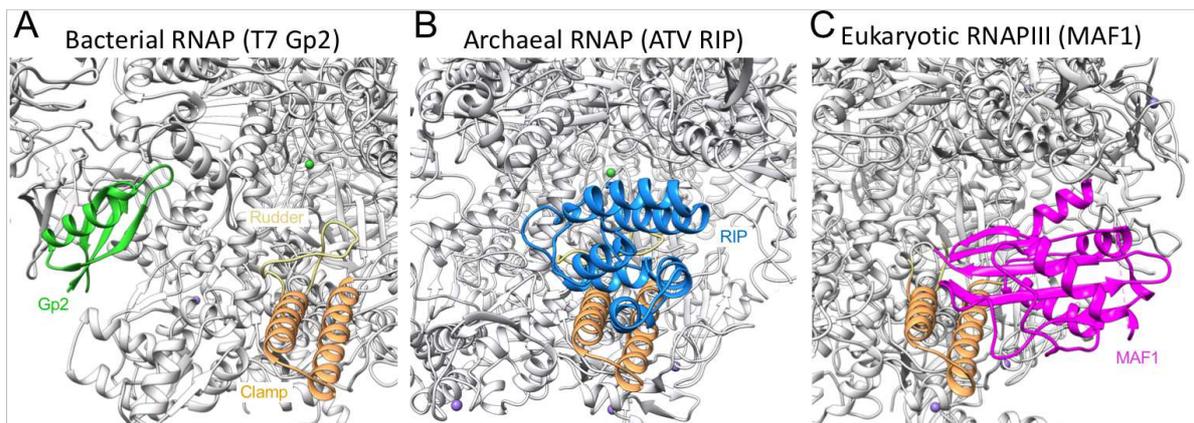
363 As RNAPs are important therapeutic targets, the structural basis and mechanisms of their
364 inhibition have been studied in great detail in bacteria and eukaryotes – while nothing is
365 known about archaeal RNAP. Arguing from first principles, inhibition can be achieved by (i)
366 preventing direct interactions between RNAP and basal factors, nucleic acid template, or
367 substrate NTP, or (ii) via allosteric mechanisms that alter the structure or conformation of
368 RNAP in a way that abrogates catalysis. Our two novel structures of RNAP-inhibitor complexes
369 both add new features and highlight common themes of intervening with RNAP function.

370

371 RIP is a classic competitive inhibitor of RNAP

372 RIP is a small single-domain protein which binds to the inside of the RNAP clamp and rudder
373 motifs which is incompatible with the binding and function of the essential initiation factor
374 TFB. The C-terminal tail of RIP and the B-linker strand and α -helix motifs of TFB (and TFIIB)
375 interact with RNAP in an identical fashion (Figure 3B)³⁸. RIP forms a ‘plug’ in the DNA-binding
376 channel of RNAP that interferes with the formation of the closed PIC, and it would prevent the
377 loading of the template DNA strand into the RNAP active site during the transition from the
378 closed to the open PIC (Figure 3A-C). The binding of RIP is furthermore incompatible with the
379 binding of the DNA scaffold in the transcription elongation complex (Figure 3C). All of the
380 observations above are in perfect agreement with published biochemical interaction analyses,
381 which suggested a competitive inhibition mechanism of RIP. The initiation factor TFE induces
382 an opening of the RNAP clamp²⁷ which stabilises the PIC and activates transcription³⁰.
383 Somewhat counterintuitively, complete TFE-containing PICs are more sensitive to RIP as
384 compared to minimal PIC lacking TFE¹⁸. The structure of the RNAP-RIP complex rationalises
385 this observation, as the TFE-induced opening the clamp²⁷ makes the RIP binding site more
386 accessible. The occlusion of the DNA-binding channel is a reliable and direct mechanism of
387 inhibition exploited by all domains of life. Indeed, the bacteriophage T7 Gp2 binds RNAP
388 within the DNA channel and effectively inhibits initiation of the *E. coli* RNAP by interfering
389 with the σ^{70} initiation factor⁷. Among the eukaryotic transcription systems, the cellular negative
390 regulator MAF1 specifically inhibits RNAPIII in response to stress and nutrient deficiency.

391 Similarly to RIP and T7 Gp2, MAF-1 binds inside the DNA-binding channel of RNAPIII and
 392 occludes the binding site of TFIIIB to the RNAP clamp and rudder motifs (Figure 6)³⁹. What is
 393 singular about RIP and novel in the field, is the unique ability of a viral protein to mimic the
 394 binding mode and thereby prevent the binding of the universally conserved host basal
 395 initiation factor TFB.
 396
 397



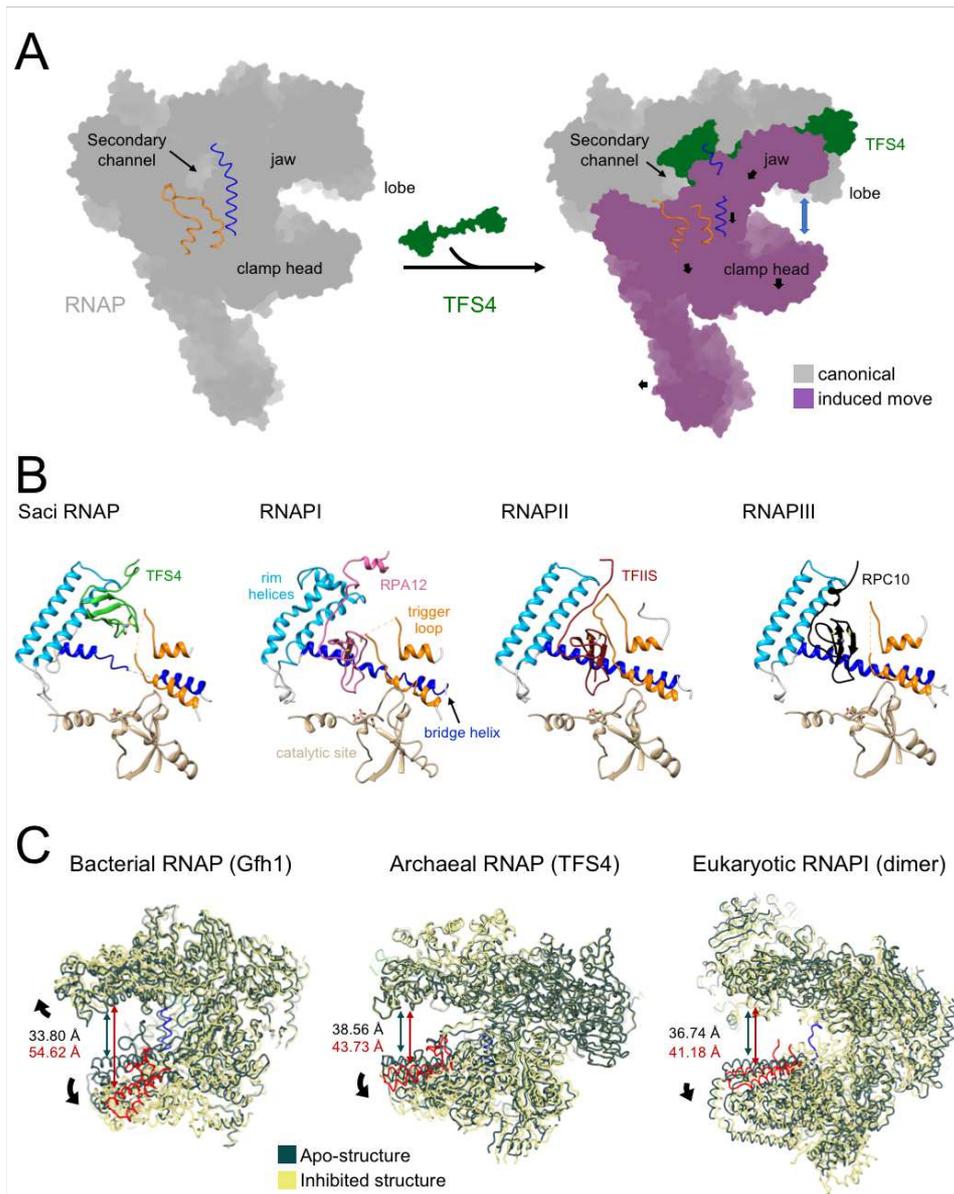
398
 399 **Figure 6. The DNA-binding channel is a common target for transcription inhibition in all domains of**
 400 **life.** The structures of the RNAP from (A) *E. coli* (pdb code 4lk0), (B) *Saci* and (C) the *S. cerevisiae* RNAPIII
 401 (pdb code 6tut) are all shown in grey ribbons with the clamp and rudder highlighted in light orange
 402 and yellow, respectively. The viral inhibitor T7 Gp2 is shown in green, RIP from ATV in blue, and the
 403 cellular negative regulator MAF1 in magenta. The Mg ion in the active site of the channel is shown as
 404 a green sphere.

405

406 TFS4 exploits the intrinsic conformational flexibility of cellular RNAPs

407 Archaea, including Sulfolobales, encode functionally diversified paralogues of the transcript
 408 cleavage elongation factor TFS^{22,40}. *Sso* TFS1 and TFS4 share the same domain organisation
 409 and a 28% sequence identity and both bind competitively to RNAP, yet while TFS1 stimulates
 410 elongation, TFS4 has evolved into a potent inhibitor of RNAP²². Although the structure of the
 411 RNAP-TFS1 complex has not been determined yet, it is very likely that the acidic loop residing
 412 in the TFS1 ZR^C has to penetrate deep into the active site through the NTP entry funnel like
 413 TFIIIS^{16,22} because the molecular mechanisms of TFS1 and TFIIIS are strictly conserved^{18,19}. Our
 414 structure shows that the TFS4 ZR^C binds in the RNAP funnel, but not as deep and in a different

415 orientation compared to TFIIIS (Figure 7B) ^{16,41}. A superposition of the apo-RNAP and RNAP-
416 TFS4 complex reveals global as well as local changes of RNAP structure that account for the
417 efficient inhibition, with an overall opening of the RNAP through the DNA-binding channel,
418 and alteration of the bridge helix and trigger loop motifs in the active site (Figure 5 and 7A).
419 The NTP entry funnel provides a crucial orifice to the RNAP active site that not only allows
420 NTPs to enter, and the RNA 3' terminus to exit the RNAP in backtracked TECs, but also serves
421 as binding site for a plethora of transcription regulators ¹⁵. These include negative regulators
422 such as the bacterial Gfh1 ^{42,43} and DksA ^{44,45} that have pleiotropic effects on transcription.
423 DksA in conjunction with ppGpp alters the promoter-specificity of RNAP by repressing
424 stringent promoters during amino acid starvation, and the binding of ppGpp, even in the
425 absence of DksA, restricts the conformational flexibility and thereby inhibits RNAP ⁴⁶. Gfh1
426 inhibits all catalytic activities of RNAP and in particular enhances pausing ^{42,47,48}. Like TFS1 and
427 TFS4 in archaea, the bacterial GreA, GreB, Gfh1 and DksA factors compete with each other
428 for RNAP binding, and changes in their relative expression level is likely to be an integral part
429 of their regulation. In the case of TFS1 and 4, the former is constitutively expressed and
430 engaged with RNAP, while the latter is exclusively expressed in response to viral infection ²³.
431 TFS4 and Gfh1 are not homologous but share intriguing functional properties that likely result
432 from convergent evolution. Both bind to the RNAP funnel and allosterically inhibit RNAP which
433 results in the widening of the DNA binding channel and – albeit to a different degree - a
434 distortion of the bridge helix and trigger loop in the active site (Figure 7C).
435



436

437 **Figure 7. Commonalities into allosterically exploiting the RNAP structural flexibility.** (A) Schematic
 438 representation of the conformational changes induced by TFS4 binding. The RNAP is shown in grey,
 439 and the part of the structure subjected to conformational change is highlighted in purple with the
 440 movement direction indicated by arrows. Main features of RNAP including the secondary channel, jaw,
 441 lobe and clamp head, as well as bridge helix and trigger loop are indicated. (B) Comparison of the
 442 archaeal Saci RNAP with eukaryotic RNAPI, II and III and the locations of the ZR^C domains of subunits
 443 TFS4, RPA12, TFIIIS, and RPC11 (pdb codes 6rqh, 5xon, and 7ae3, respectively for the eukaryotic
 444 factors), relative to the rim helices and the active site motifs (aspartate triad, bridge helix and trigger
 445 loop), while all other ZR^C domains reach towards the catalytic magnesium ion (green sphere). (C)
 446 The structural changes associated with the inhibition of the archaeal, bacterial and eukaryotic RNAPI
 447

448 by TFS4, Gfh1 and Pol I dimerization, respectively, involve a widening of the DNA binding-channel. The
449 structures of the free, or apo forms of RNAP are shown in dark teal, and the inhibited conformations of
450 the RNAP-TFS4, RNAP-Gfh1 and RNAPI dimers are shown in yellow, respectively. The RNAP clamp
451 helices are highlighted in red and the bridge helix distortions in blue.

452

453 What is the origin of the conformational dynamics of DPBB RNAP in all domains of life? The
454 NTP translocation cycle of RNAP involves tightly coordinated conformational changes as it
455 cycles through post- and pre-translocated, and of NTP pre-insertion and insertion states that
456 have been captured by high-resolution X-ray structures of the bacterial RNAP^{49,50}. As the
457 archaeal RNAP progresses through the transcription cycle, the flexible RNAP clamp adopts
458 distinct conformational states which alter the width of the DNA-binding channel²⁷. Elongation
459 is a discontinuous process interrupted by pausing and termination which involves inhibited
460 states. Recently it has been shown that the bacterial RNAP in the context of an RNA hairpin
461 and elemental pause - which is likely preceding termination - has a widened DNA-binding
462 channel, which is reminiscent of the archaeal TFS4-RNAP complex⁵¹. This is different from a
463 closed RNAP clamp that is characteristic for processive elongation complexes^{49,50}. Eukaryotic
464 RNAPI, II and III include RPA12, RPR9 and RPC10 subunits, respectively, that are paralogs of
465 TFIIIS but stably incorporated into the RNAP rather than reversibly associated with it (Figure
466 S7a-d). Importantly, the binding of ZR^C domains of RPA12, RPC10 and TFIIIS inside the RNAP
467 funnel is nearly identical, while only TFS4 clashes with the trigger loop (Figure 7B and S7e-f).
468 The ZR^C domains of RPA12 and RPC10 are thought to move in and out of the NTP entry funnel,
469 with the 'inside funnel' state of A12 and RPC10 which might play a role in transcription
470 termination of RNAPI and RNAPIII, respectively^{36,52,53,54}. Finally, in the inactive dimeric form of
471 RNAPI, the two molecules face each other symmetrically in a 'handshake' orientation where
472 the DNA-binding channel of one molecule interacts with the stalk of the other³⁴. In this
473 conformation, the RNAPI jaw and lobe modules have moved outwards and the DNA-binding
474 channel is widened by 5 Å⁵⁵ similarly to the inhibited state of the archaeal TFS4-RNAP
475 complex (Figure 7C).

476 We posit that the allosteric changes induced by TFS4 are related to several inhibited states of
477 DPBB RNAP, including transcription effectors (like Gfh1), oligomerisation equilibria (RNAPI)
478 and more general inhibited states of RNAPs associated with pausing and termination. In the

479 structurally well characterised cases discussed above, the DNA-binding channel is widened
480 and the interactions with the template DNA weakened. The common denominator of
481 inhibition is the ability to lock them in a specific conformation. In a likely scenario, RNAP
482 subunits and factors have evolved vertically by gene duplication and speciation (e. g. TFIIIS,
483 TFS1 and -4, RPA12, RPB9 and RPC10) or by convergent evolution (e. g. GreA/B and Gfh1
484 relative to TFS1) to exploit the inherent conformational flexibility of RNAP to modulate and
485 fine tune transcription.

486

487 [The role of RNAP inhibition in the virus-host relationship](#)

488 The biological function of RIP for the ATV virus is still a matter of debate, but its tight
489 regulation and strong induction during the very late stages of lytic infection supports a role in
490 virus particle assembly, possibly by dissociating transcribing RNAP from viral genomes. The
491 strong and stable interactions between RIP and the RNAP clamp and rudder make RIP a lethal
492 protein, and this absolute mechanism to shut down RNAP is apparently the preferred option
493 for viruses and bacteriophages like ATV and T7 ⁶.

494 In comparison, host-encoded factors such as TFS4 and Gfh1 seem to opt for a more versatile
495 approach by allosterically interfering with RNAP function. This provides an opportunity for the
496 cell to temporarily pause transcription, mount additional defence mechanisms, and eventually
497 reactivate the gene expression programme once the favourable conditions have returned.
498 This is the case for the well-studied Gfh1 which adopts an inactive conformation during normal
499 growth conditions and undergoes a reversible conformational change upon acidification of
500 the medium ²². The plasmid-driven expression of TFS4 stops cell growth similar to virus
501 infection, but it is not known yet whether TFS4 is the sole agent to trigger the host growth
502 retardation in response to infection ²². The induction of a quiescent state is an important
503 preamble for persistence of many pathogenic bacteriophages ⁵⁶, hence, we hypothesise that
504 TFS4 may play a similar role in archaea.

505 The detailed structural basis and the mechanisms underlying that the inhibition of archaeal
506 DPBB RNAP described in this manuscript has the potential to aid the design of novel drugs
507 targeting the RNAPs of bacterial pathogens and RNAPI in cancer therapy ⁵⁷. These include
508 small effector proteins that bind tightly to RNAP with high specificity, like RIP, and either deny
509 access to the DNA binding channel or essential transcription factors. Such effectors could be

510 eventually delivered to their eukaryotic or bacterial targets by recombinant viruses or
511 bacteriophages, respectively^{58,59}. Moreover, agents that bind to RNAP in a polydentate
512 fashion like TFS4 and prevent conformational changes that are critical for catalysis are possibly
513 more refractive to elicit the fast emergence of resistance mutations.

514 Material and Methods

515 Protein expression and purification

516 Saci RNA polymerase, 6xHis tagged at the C-terminus of Rpo8, was expressed in Saci strain
517 MW001 and purified according to established protocols²⁹.

518 ATV RIP (ORF145) tagged at the C-terminus with 6xHis sequence and Sso TFS4 were
519 expressed in *E. coli* and purified as previously described^{10,22}.

520

521 Complex assembly and cryo-electron microscopy data collection

522 Apo-RNAP. The RNA polymerase was crosslinked at 0.15 mg/ml in 200 μ l of a buffer
523 containing 20 mM Hepes pH 7.0, 200 mM NaCl, 5 mM MgCl₂, 100 μ M ZnSO₄, 10% glycerol, 5
524 mM DTT, with 2 mM bis(sulfosuccinimidyl)suberate (BS3) for 5 min at 65° C. The reaction was
525 quenched adding 150 mM NH₄HCO₃ at rt for 20 min. The sample was then diluted ten times
526 in the same buffer without glycerol, filtered with a 0.22 μ m filter and concentrated up to 0.4
527 mg/ml in a concentrator with a cutoff of 100 kDa. Sample quality was firstly assessed by
528 negative staining, then 3 μ L of sample at 0.06 mg/ml was spotted on a UltrAuFoil holey grid
529 300 mesh R1.2/1.3 (Quantifoil, Germany) covered with graphene oxide according to a protocol
530 described by Cheng K. and co-workers⁶⁰, and vitrified by plunging in liquid ethane using
531 Vitrobot Mark IV (Thermo Fisher Scientific, USA) at 4°C and 94 % humidity. Data were collected
532 at eBIC National facility (Diamond Light Source, UK) using a Titan Krios microscope (Thermo
533 Fisher Scientific, USA) operated at 300 keV and equipped with a BioQuantum energy filter
534 (Gatan, USA). The images were collected with a post-GIF K3 direct electron detector (Gatan,
535 USA) operated in super resolution mode, at a nominal magnification of 81,000, corresponding
536 to a pixel size of 1.085 Å. The dose rate was set to 21 e⁻ per pixel per second, and a total dose
537 of 44.46 e⁻/Å² was fractionated over 40 frames. An energy slit with a 20 eV width was used
538 during data collection. Data were collected using EPU software (Thermo Fisher Scientific, USA)

539 with a nominal defocus range $-1.0\ \mu\text{m}$ - $-2.5\ \mu\text{m}$. During the data collection microscope stage
540 was tilted to -30° to overcome the preferred orientations observed in previous datasets under
541 these experimental conditions.

542 RNAP/RIP complex. Prior the crosslinking procedure already described, the RNA polymerase
543 was incubated with a 10 molar fold excess of RIP for 5 min at $65^\circ\ \text{C}$. Grids covered with
544 graphene oxide were prepared according to the protocol described above and data were
545 collected in ISMB Birkbeck EM facility using a Titan Krios microscope operated at 300 keV and
546 equipped with a BioQuantum energy filter. The images were collected with a post-GIF K2
547 Summit direct electron detector (Gatan, USA) operated in counting mode, at a nominal
548 magnification of 130,000 corresponding to a pixel size of $1.01\ \text{\AA}$. The dose rate was set to 5.84
549 e^- per pixel per second, and a total dose of $45.5\ \text{e}^-/\text{\AA}^2$ was fractionated over 45 frames. An
550 energy slit with a 20 eV width was used during data collection. Data were collected using EPU
551 software with a nominal defocus range $-1\ \mu\text{m}$ to $-2.5\ \mu\text{m}$. During the data collection
552 microscope stage was tilted to -30° .

553 RNAP/TFS4 complex. The RNA polymerase was incubated and crosslinked as previously
554 described in presence of a 20 molar fold excess of TFS4. The sample at $0.4\ \text{mg/ml}$ of
555 concentration was applied twice on a UltrAuFoil holey grid 300 mesh R1.2/1.3 and vitrified by
556 plunging in liquid ethane using Vitrobot Mark IV at 4°C and 94 % humidity. Data were collected
557 at eBIC National facility using a Titan Krios operated at 300 keV and equipped with a
558 BioQuantum energy. The images were collected with a post-GIF K3 direct electron detector
559 (Gatan, USA) operated in super resolution mode, at a nominal magnification of 81,000,
560 corresponding to a pixel size of $1.085\ \text{\AA}$. The dose rate was set to $21.221\ \text{e}^-$ per pixel per
561 second, and a total dose of $45.2\ \text{e}^-/\text{\AA}^2$ was fractionated over 40 frames. An energy slit with a
562 20 eV width was used during data collection. Data were collected using EPU software with a
563 nominal defocus range $-1\ \mu\text{m}$ to $-2.5\ \mu\text{m}$.

564

565 [Cryo-electron microscopy data processing](#)

566 Apo-RNAP. The dataset of 1,676 movie stacks was aligned, summed and 2X binned using
567 MotionCor2⁶¹, followed by CTF estimation using GCTF⁶². Relion 3.0 software⁶³ was used for
568 template-free particle picking (Laplacian method) and all consequent image processing for
569 this sample. Initially 1,286,432 particles were extracted and downscaled to a/pix of $4.5\ \text{\AA}$. After

570 multiple cycles of 2D and 3D classifications 423,157 best particles were selected and rescaled
571 to the pixel size of 1.085 Å. The selected subset of particles was then refined and, after post-
572 processing, subjected to four cycles of CTF refinement to correct for the effect of stage tilt
573 used during data collection on the initial CTF estimation. Three cycles of CTF refinement were
574 sufficient for the correction, and the fourth cycle didn't provide any further improvement. The
575 final cycle of 3D refinement and post-processing resulted a map with resolution of 2.88 Å as
576 estimated using gold standard Fourier Shell Correlation (FSC) with a 0.143 threshold (Figure
577 S1 and table 1).

578 RNAP/RIP complex. The 2,130 movie stacks were aligned and summed using MotionCor2
579 followed by CTF estimation in GCTF. Particles were picked using Relion 3.0 reference-based
580 method. These 600,640 particles extracted from the micrographs were subjected to multiple
581 rounds of 2D and 3D classifications using cryoSPARC⁶⁴. Multiple approaches were applied to
582 identify the apo-RNAP but no additional species were found likely due to the relatively low
583 number of particles available for the search. The best 151,237 particles were rescaled to the
584 original pixel size, and image processing continued using Relion 3.0. Particles were refined
585 and subjected to three cycles of CTF refinement. After a final 3D refinement and post
586 processing step the resolution of 3.27 Å was estimated using gold standard FSC with 0.143
587 threshold (Figure S3 and table 1).

588 RNAP/TFS4 complex. The dataset of 1760 movie stacks was motion corrected and analysed
589 with Relion 3.0 following the same routine described above for the apo-RNA polymerase
590 sample. After multiple rounds of 2D and 3D classifications, best 505,758 particles were
591 selected out of 1,161,535, and subjected to two cycles of CTF refinement followed by particle
592 polishing. Although the large number of particles, we did not identify the apo-RNAP or any
593 other intermediate conformational species, using either Relion 3.0 and Cryosparc. The last 3D
594 refinement and post-processing steps provided a map with an averaged resolution of 2.59 Å
595 as estimated using gold standard FSC with 0.143 threshold. Analysis of the Euler angles
596 distribution for this map revealed a clear cluster of preferred orientations (data not shown).
597 Although it didn't seem to affect map quality, we decided to prune the refined particles and
598 repeat refinement and post-processing. This step was carried out using an in-house script
599 considering the tilt and rotation angles of the last refinement and pruning the particles based
600 on the CtfFigureOfMerit metadata up to 350,000 particles. The pruning provided a batch of

601 350,682 particles which gave a map with resolution of 2.61 Å at the FSC of 0.143 (Figure S4
602 and table 1). The local resolution of TFS4 map shown in figure S6 resulted to be perfectly
603 compatible with the local resolution of the RNA polymerase in the same area suggesting that
604 the apo- or intermediate species, if present, represent all together less than the 5% of the
605 dataset, which might explain the resulting single class from the 3D classification step. To
606 improve map quality of the stalk which is intrinsically a flexible arm protruding from the main
607 body, we performed a multi-body refinement using the RNAP main domain and the stalk as
608 moving bodies. To do that, we generated two masks with soft edges for the large globular
609 main domain and the stalk, respectively, to which we applied 20° of width for the rotation
610 priors and 5 pixels for the translation between the two bodies, a setup suggested for highly
611 flexible bodies according to the protocol released by Nakane T and Sheres SHW⁶⁵. The multi-
612 body refinement was the only successful approach found improving the stalk resolution up to
613 3.75 Å (figure S5 and table 1).

614

615 Model building and refinement

616 Local resolution was assessed using Relion 3.0 after post-processing, whereas map sharpening
617 and model refinement were carried out in Phenix v1.15.2 and 1.19.2⁶⁶, including rounds of
618 manual editing and refinement in Coot v0.8.9.1⁶⁷ (Table 1). Models of all RNAP subunits, RIP
619 and TFS4 were prepared using Modeller³⁴, then the RNAP complex was assembled using the
620 homologous RNA polymerase from *Sulfolobus shibatae* at 3.2 Å (pdb 4ayb) and refined
621 against the cryo-EM map of RNAP/TFS4 at 2.61 Å. The structure obtained was used to refine
622 the model against the maps of the apo- and RIP-bound RNA polymerase. Following the RNAP
623 refinement, the initial models of RIP and TFS4, obtained using Modeller, were placed inside
624 the extra-densities and refined accordingly. The C-terminal tail of RIP and TFS4 linker, not
625 predicted in the initial model, were built manually in Coot following map density.

626 PDB accession codes: RNA polymerase 7ok0, RNAP/RIP complex 7oq4, RNAP/TFS4 YYY.

627 EMDB accession codes: RNA polymerase EMD-12960, RNAP/RIP complex EMD-13026,

628 RNAP/TFS4 yyyy, RNAP/TFS4 stalk zzzz.

629

630 Interfaces data analyses and sequence alignments

631 The binding interfaces between the RNA polymerase and RIP or TFS4 were analysed using
632 two different programs for consistency, PISA and LigPlot+^{68,69}. As setup in LigPlot+ we used
633 for the H-bonds and salt bridges a maximum of 3.5 Å for donors and 4 Å for acceptors, for the
634 hydrophobic interactions we used the maximum values of 3.5 and 4.3 Å.

635 Structure-based alignments of all Saci RNAP subunits have been carried out with the Match-
636 Align tool in Chimera⁷⁰ after superimposition of the Saci apo-RNAP with *S. shibatae* (4ayb), *S.*
637 *solfatarius* (3hgz), *T. kodakarensis* (6kf9) and *S. cerevisiae* (5vvs) followed by manual editing
638 of the flexible loops. ATV RIP, and P131/ORF131, SMV-1 ORF122 and ORF114 have been aligned
639 using PROMALS3D⁷¹. The alignment of Rpo8/RPB8 was carried out by manual editing after
640 superposition of Saci RNAP structure with 3hgz, 4ayb, and 5vvs (*S. solfataricus*, *S. shibatae*,
641 and *S. cerevisiae*, respectively). Alignment images throughout the text were prepared in
642 Esript3⁷².

643

644 ATV infection and biomass sampling

645 *Acidianus convivator* strain AA9 was grown at 76°C in 1X medium including sulfur⁷³. The cell
646 density was monitored with a Shimadzu spectrophotometer (OD600) and the cell number
647 determined with a Thoma counting chamber at different times points during the growth. Cells
648 from 500 ml of exponentially growing *A. convivator* were collected at OD600 = 0.08 by
649 centrifugation at 3500 rpm for 30 minutes at 4°C, and the pellet was suspended in 500 µl of
650 1X medium. 150 µl of ATV virus preparation⁷⁴ (titer of 1010 virions/ml) was added to the cell
651 suspension, the mixture was incubated at 80°C for 1 hour and diluted with 500 ml of 1X
652 medium including sulfur. Infected cells were grown at 76°C and samples taken at t= 0, 6, 12,
653 18, 48 and 72 hrs post infection.

654

655 Immunodetection

656 Cell pellets were sonicated (20% amplitude, 10 sec pulse mode for 1 hour) and protein content
657 was measured using the Qubit system (Invitrogen). For each sample, the equivalent of 9 µg of
658 total protein was loaded on a 14% Tris-Tricine SDS-PAGE and blotted onto a 0.2 µm PVDF
659 membrane. As primary antibodies, we used a polyclonal antibody from rabbit for RIP, and
660 from sheep for *Sso Alba*, the second target used as additional internal control to assess

661 protein content of the samples. To visualise the two protein targets in different colours we
662 used two different secondary antibodies, the anti-rabbit antibody from donkey was
663 conjugated with Dylight680 dye, the anti-sheep antibody from donkey was conjugated with
664 Dylight488. Fluorescent detection was carried out on Typhoon FLA 9500 scanner (GE
665 Healthcare) for Alexa-488 and Alexa-680.

666

667 EMSA experiments

668 The DNA:RNA scaffold was prepared incubating the DNA template strand labelled with Cy3
669 (Cy3-TS83) with RNA (RNA14) at 76°C for 5 min, then we added the non-template strand
670 (NTS83) and we incubated for further 5 min at 76°C. RIP was cloned into the pKA-vector
671 carrying the phosphorylation site at the N-terminus. It was expressed, purified and labelled
672 with ³²P according to published protocol¹⁰. In order to be able to use detectable signals of
673 both fluorescent scaffold and radioactive RIP, ³²P-RIP was mixed with cold RIP in to 3:1 ratio.
674 We split the competition assay in two halves. In the first half assay we pre-incubated the
675 scaffold (100 nM) with Saci RNAP (200 nM) in 15 µl of buffer 10 mM MOPS pH 6.5, 10 mM
676 MgCl₂, 200 mM NaCl, 7 mM DTT, 10% glycerol, 0.067 mg/ml BSA, and 0.1 mg/ml heparin at
677 65°C for 5 min, then we added increasing amount of ³²P-RIP (0.2-2 µM) and we incubated again
678 at 65°C for 5 min. The competition reactions of the second half assay were prepared by pre-
679 incubating ³²P-RIP (400 nM) with 200 nM Saci RNAP in 15 µl of the same buffer at 65°C for 5
680 min, then adding 0.1-1 µM scaffold and incubating for further 5 min at 65°C. Samples were
681 resolved on a 6% native PAGE and detected using Typhoon (GE).

682 Author contributions

683 SP, NL, LMD and AC worked on cryo-EM data, TF and CS prepared proteins and carried out
684 cross-linking and Western blotting, SLS and DP carried out virus infection experiments, DM
685 contributed to the identification of Rpo8 mismatch, SP and FW wrote the manuscript, and FW
686 conceived and planned the project.

687 Acknowledgements

688 We would like to thank Diamond Light Source for access to the cryo-EM facilities at the UK
689 National electron Bio-Imaging Centre (eBIC, proposal EM20287) funded by the Wellcome

690 Trust, the Medical Research Council UK and the Biotechnology and Biological Sciences
691 Research Council. Cryo-EM data for this investigation were also collected at the ISMB EM
692 facility at Birkbeck College, University of London with financial support from Wellcome
693 (202679/Z/16/Z and 206166/Z/17/Z). We would like to thank Dr. David Houldershaw for IT
694 support and Giulia Zanetti for the script writing. We are very grateful to Jerome Gouge,
695 Anthony Roberts and Christoph Müller for discussions, comments and suggestions. Research
696 in the RNAP laboratory at UCL is funded by a Wellcome Investigator Award in Science
697 'Mechanisms and Regulation of RNAP transcription' to FW (WT 207446/Z/17/Z).

698 References

- 699 1. Stepanova, E.V., Shevelev, A.B., Borukhov, S.I. & Severinov, K.V. [Mechanisms of
700 action of RNA polymerase-binding transcription factors that do not bind to DNA].
701 *Biofizika* **54**, 773-90 (2009).
- 702 2. Sheppard, C. & Werner, F. Structure and mechanisms of viral transcription factors in
703 archaea. in *Extremophiles* Vol. 21 829-838 (Springer Japan, 2017).
- 704 3. Mosaei, H. & Harbottle, J. Mechanisms of antibiotics inhibiting bacterial RNA
705 polymerase. *Biochem Soc Trans* **47**, 339-350 (2019).
- 706 4. Artsimovitch, I. & Vassylyev, D.G. Is it easy to stop RNA polymerase? *Cell Cycle* **5**, 399-
707 404 (2006).
- 708 5. Forterre, P. The origin of viruses and their possible roles in major evolutionary
709 transitions. *Virus Res* **117**, 5-16 (2006).
- 710 6. Bae, B. et al. Phage T7 Gp2 inhibition of Escherichia coli RNA polymerase involves
711 misappropriation of $\sigma 70$ domain 1.1. in *Proceedings of the National Academy of*
712 *Sciences of the United States of America* Vol. 110 19772-19777 (2013).
- 713 7. Sheppard, C. et al. A non-bacterial transcription factor inhibits bacterial transcription
714 by a multipronged mechanism. *RNA Biol* **10**, 495-501 (2013).
- 715 8. Tabib-Salazar, A. et al. Full shut-off of Escherichia coli RNA-polymerase by T7 phage
716 requires a small phage-encoded DNA-binding protein. *Nucleic Acids Res* **45**, 7697-7707
717 (2017).
- 718 9. Sheppard, C. et al. Repression of RNA polymerase by the archaeo-viral regulator
719 ORF145/RIP. in *Nature Communications* Vol. 7 1-13 (Nature Publishing Group, 2016).
- 720 10. Sheppard, C. et al. Repression of RNA polymerase by the archaeo-viral regulator
721 ORF145/RIP. *Nat Commun* **7**, 13595 (2016).
- 722 11. Agapov, A., Esyunina, D., Pupov, D. & Kulbachinskiy, A. Regulation of transcription
723 initiation by Gfh factors from *Deinococcus radiodurans*. *Biochem J* **473**, 4493-4505
724 (2016).
- 725 12. Laptenko, O. et al. pH-dependent conformational switch activates the inhibitor of
726 transcription elongation. in *EMBO Journal* Vol. 25 2131-2141 (2006).
- 727 13. Furman, R., Sevostyanova, A. & Artsimovitch, I. Transcription initiation factor DksA has
728 diverse effects on RNA chain elongation. *Nucleic Acids Res* **40**, 3392-402 (2012).

- 729 14. Parshin, A. et al. DksA regulates RNA polymerase in Escherichia coli through a network
730 of interactions in the secondary channel that includes Sequence Insertion 1. *Proc Natl*
731 *Acad Sci U S A* **112**, E6862-71 (2015).
- 732 15. Zenkin, N. & Yuzenkova, Y. New Insights into the Functions of Transcription Factors
733 that Bind the RNA Polymerase Secondary Channel. *Biomolecules* **5**, 1195-209 (2015).
- 734 16. Cheung, A.C. & Cramer, P. Structural basis of RNA polymerase II backtracking, arrest
735 and reactivation. *Nature* **471**, 249-53 (2011).
- 736 17. Noe Gonzalez, M., Blears, D. & Svejstrup, J.Q. Causes and consequences of RNA
737 polymerase II stalling during transcript elongation. *Nat Rev Mol Cell Biol* **22**, 3-21
738 (2021).
- 739 18. Hausner, W., Lange, U. & Musfeldt, M. Transcription factor S, a cleavage induction
740 factor of the archaeal RNA polymerase. *J Biol Chem* **275**, 12393-9 (2000).
- 741 19. Lange, U. & Hausner, W. Transcriptional fidelity and proofreading in Archaea and
742 implications for the mechanism of TFS-induced RNA cleavage. *Mol Microbiol* **52**, 1133-
743 43 (2004).
- 744 20. Laptenko, O., Lee, J., Lomakin, I. & Borukhov, S. Transcript cleavage factors GreA and
745 GreB act as transient catalytic components of RNA polymerase. *EMBO J* **22**, 6322-34
746 (2003).
- 747 21. Yuzenkova, Y. et al. Control of transcription elongation by GreA determines rate of
748 gene expression in Streptococcus pneumoniae. *Nucleic Acids Res* **42**, 10987-99 (2014).
- 749 22. Fouqueau, T. et al. The transcript cleavage factor paralogue TFS4 is a potent RNA
750 polymerase inhibitor. in *Nature communications* (2017).
- 751 23. Ortmann, A.C. et al. Transcriptome analysis of infection of the archaeon Sulfolobus
752 solfataricus with Sulfolobus turreted icosahedral virus. *J Virol* **82**, 4874-83 (2008).
- 753 24. Hirata, A., Klein, B.J. & Murakami, K.S. The X-ray crystal structure of RNA polymerase
754 from Archaea. *Nature* **451**, 851-4 (2008).
- 755 25. Briand, J.-F. et al. Partners of Rpb8p, a Small Subunit Shared by Yeast RNA Polymerases
756 I, II, and III. in *Molecular and Cellular Biology* Vol. 21 6056-6065 (2001).
- 757 26. Grohmann, D. et al. The initiation factor TFE and the elongation factor Spt4/5 compete
758 for the RNAP clamp during transcription initiation and elongation. *Mol Cell* **43**, 263-74
759 (2011).
- 760 27. Schulz, S. et al. TFE and Spt4/5 open and close the RNA polymerase clamp during the
761 transcription cycle. *Proc Natl Acad Sci U S A* **113**, E1816-25 (2016).
- 762 28. Werner, F. & Weinzierl, R.O. A recombinant RNA polymerase II-like enzyme capable of
763 promoter-specific transcription. *Mol Cell* **10**, 635-46 (2002).
- 764 29. Blombach, F. et al. Archaeal TFE α/β is a hybrid of TFIIE and the RNA polymerase III
765 subcomplex hRPC62/39. in *eLife* Vol. 4 e08378 (eLife Sciences Publications Ltd, 2015).
- 766 30. Blombach, F., Smollett, K.L., Grohmann, D. & Werner, F. Molecular Mechanisms of
767 Transcription Initiation-Structure, Function, and Evolution of TFE/TFIIE-Like Factors
768 and Open Complex Formation. *J Mol Biol* **428**, 2592-2606 (2016).
- 769 31. Werner, F. & Weinzierl, R.O. Direct modulation of RNA polymerase core functions by
770 basal transcription factors. *Mol Cell Biol* **25**, 8344-55 (2005).
- 771 32. Deng, L., Garrett, R.A., Shah, S.A., Peng, X. & She, Q. A novel interference mechanism
772 by a type IIIB CRISPR-Cmr module in Sulfolobus. *Mol Microbiol* **87**, 1088-99 (2013).
- 773 33. Cramer, P., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: RNA
774 polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863-76 (2001).

- 775 34. Engel, C., Sainsbury, S., Cheung, A.C., Kostrewa, D. & Cramer, P. RNA polymerase I
776 structure and transcription regulation. *Nature* **502**, 650-5 (2013).
- 777 35. Fernandez-Tornero, C., Bottcher, B., Rashid, U.J. & Muller, C.W. Analyzing RNA
778 polymerase III by electron cryomicroscopy. *RNA Biol* **8**, 760-5 (2011).
- 779 36. Girbig, M. et al. Cryo-EM structures of human RNA polymerase III in its unbound and
780 transcribing states. *Nat Struct Mol Biol* **28**, 210-219 (2021).
- 781 37. Fouqueau, T., Zeller, M.E., Cheung, A.C., Cramer, P. & Thomm, M. The RNA polymerase
782 trigger loop functions in all three phases of the transcription cycle. *Nucleic Acids Res*
783 **41**, 7048-59 (2013).
- 784 38. Kostrewa, D. et al. RNA polymerase II-TFIIB structure and mechanism of transcription
785 initiation. *Nature* **462**, 323-30 (2009).
- 786 39. Vorlander, M.K. et al. Structural basis for RNA polymerase III transcription repression
787 by Maf1. *Nat Struct Mol Biol* **27**, 229-232 (2020).
- 788 40. Blombach, F., Matelska, D., Fouqueau, T., Cackett, G. & Werner, F. Key Concepts and
789 Challenges in Archaeal Transcription. *J Mol Biol* **431**, 4184-4201 (2019).
- 790 41. Kettenberger, H., Armache, K.J. & Cramer, P. Architecture of the RNA polymerase II-
791 TFIIS complex and implications for mRNA cleavage. *Cell* **114**, 347-57 (2003).
- 792 42. Deighan, P. & Hochschild, A. Conformational toggle triggers a modulator of RNA
793 polymerase activity. *Trends Biochem Sci* **31**, 424-6 (2006).
- 794 43. Tagami, S. et al. Crystal structure of bacterial RNA polymerase bound with a
795 transcription inhibitor protein. in *Nature* Vol. 468 978-982 (Nature Publishing Group,
796 2010).
- 797 44. Gourse, R.L. et al. Transcriptional Responses to ppGpp and DksA. *Annu Rev Microbiol*
798 **72**, 163-184 (2018).
- 799 45. Perederina, A. et al. Regulation through the secondary channel--structural framework
800 for ppGpp-DksA synergism during transcription. *Cell* **118**, 297-309 (2004).
- 801 46. Zuo, Y., Wang, Y. & Steitz, T.A. The mechanism of E. coli RNA polymerase regulation
802 by ppGpp is suggested by the structure of their complex. *Mol Cell* **50**, 430-6 (2013).
- 803 47. Agapov, A., Esyunina, D. & Kulbachinskiy, A. Gre-family factors modulate DNA damage
804 sensing by *Deinococcus radiodurans* RNA polymerase. *RNA Biol* **16**, 1711-1720 (2019).
- 805 48. Esyunina, D., Agapov, A. & Kulbachinskiy, A. Regulation of transcriptional pausing
806 through the secondary channel of RNA polymerase. *Proc Natl Acad Sci U S A* **113**, 8699-
807 704 (2016).
- 808 49. Vassylyev, D.G., Vassylyeva, M.N., Perederina, A., Tahirov, T.H. & Artsimovitch, I.
809 Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448**,
810 157-62 (2007).
- 811 50. Vassylyev, D.G. et al. Structural basis for substrate loading in bacterial RNA
812 polymerase. *Nature* **448**, 163-8 (2007).
- 813 51. Weixlbaumer, A., Leon, K., Landick, R. & Darst, S.A. Structural basis of transcriptional
814 pausing in bacteria. *Cell* **152**, 431-41 (2013).
- 815 52. Neyer, S. et al. Structure of RNA polymerase I transcribing ribosomal DNA genes.
816 *Nature* **540**, 607-610 (2016).
- 817 53. Sharifi, S. & Bierhoff, H. Regulation of RNA Polymerase I Transcription in Development,
818 Disease, and Aging. *Annu Rev Biochem* **87**, 51-73 (2018).
- 819 54. Fernandez-Tornero, C. RNA polymerase I activation and hibernation: unique
820 mechanisms for unique genes. *Transcription* **9**, 248-254 (2018).

- 821 55. Fernandez-Tornero, C. et al. Crystal structure of the 14-subunit RNA polymerase I.
822 *Nature* **502**, 644-9 (2013).
- 823 56. Rittershaus, E.S., Baek, S.H. & Sasseti, C.M. The normalcy of dormancy: common
824 themes in microbial quiescence. *Cell Host Microbe* **13**, 643-51 (2013).
- 825 57. Ferreira, R., Schneekloth, J.S., Jr., Panov, K.I., Hannan, K.M. & Hannan, R.D. Targeting
826 the RNA Polymerase I Transcription for Cancer Therapy Comes of Age. *Cells* **9**(2020).
- 827 58. Cisek, A.A., Dabrowska, I., Gregorczyk, K.P. & Wyzewski, Z. Phage Therapy in Bacterial
828 Infections Treatment: One Hundred Years After the Discovery of Bacteriophages. *Curr*
829 *Microbiol* **74**, 277-283 (2017).
- 830 59. Kortright, K.E., Chan, B.K., Koff, J.L. & Turner, P.E. Phage Therapy: A Renewed
831 Approach to Combat Antibiotic-Resistant Bacteria. *Cell Host Microbe* **25**, 219-232
832 (2019).
- 833 60. Cheng, K., Wilkinson, M., Chaban, Y. & Wigley, D.B. A conformational switch in
834 response to Chi converts RecBCD from phage destruction to DNA repair. *Nat Struct*
835 *Mol Biol* **27**, 71-77 (2020).
- 836 61. Zheng, S.Q. et al. MotionCor2: Anisotropic correction of beam-induced motion for
837 improved cryo-electron microscopy. in *Nature Methods* Vol. 14 331-332 (2017).
- 838 62. Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12
839 (2016).
- 840 63. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure
841 determination in RELION-3. *Elife* **7**(2018).
- 842 64. Punjani, A., Rubinstein, J.L., Fleet, D.J. & Brubaker, M.A. cryoSPARC: algorithms for
843 rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296
844 (2017).
- 845 65. Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S.H. Characterisation of molecular
846 motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife*
847 **7**(2018).
- 848 66. Afonine, P.V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography.
849 *Acta Crystallogr D Struct Biol* **74**, 531-544 (2018).
- 850 67. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta*
851 *Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
- 852 68. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline
853 state. *J Mol Biol* **372**, 774-97 (2007).
- 854 69. Laskowski, R.A. & Swindells, M.B. LigPlot+: multiple ligand-protein interaction
855 diagrams for drug discovery. *J Chem Inf Model* **51**, 2778-86 (2011).
- 856 70. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research
857 and analysis. *J Comput Chem* **25**, 1605-12 (2004).
- 858 71. Pei, J., Tang, M. & Grishin, N.V. PROMALS3D web server for accurate multiple protein
859 sequence and structure alignments. *Nucleic Acids Res* **36**, W30-4 (2008).
- 860 72. Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: Extracting and rendering
861 sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res*
862 **31**, 3320-3 (2003).
- 863 73. Zillig, W. et al. Screening for Sulfolobales, Their Plasmids and Their Viruses in Icelandic
864 Solfataras. *Systematic and Applied Microbiology* **16**, 609-628 (1994).
- 865 74. Prangishvili, D. et al. Structural and genomic properties of the hyperthermophilic
866 archaeal virus ATV with an extracellular stage of the reproductive cycle. *J Mol Biol* **359**,
867 1203-16 (2006).

- 868 75. Korkhin, Y. et al. Evolution of complex RNA polymerases: the complete archaeal RNA
869 polymerase structure. *PLoS Biol* **7**, e1000102 (2009).
- 870 76. Wojtas, M.N., Mogni, M., Millet, O., Bell, S.D. & Abrescia, N.G. Structural and
871 functional analyses of the interaction of archaeal RNA polymerase with DNA. *Nucleic*
872 *Acids Res* **40**, 9941-52 (2012).
- 873
- 874

875 Supplementary materials

876

877 Saci Rpo8 modelling and reannotation

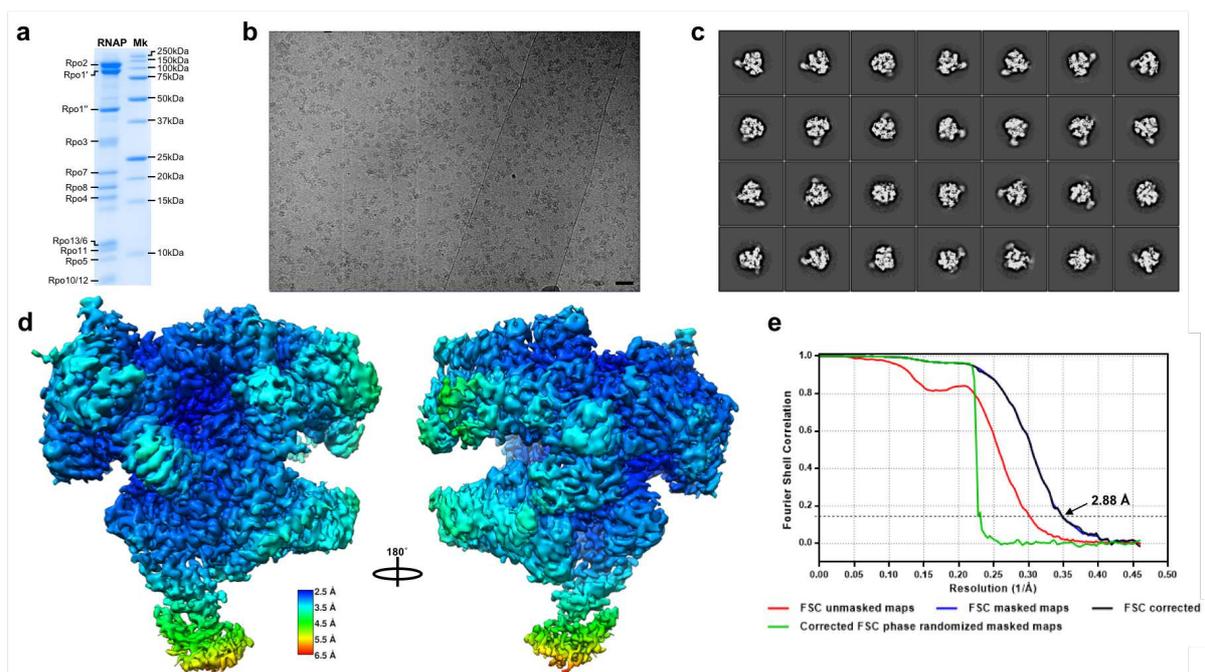
878

879 The structural model of Rpo8 has been generated in Modeller webserver based on the
880 homology with *S. shibatae* Rpo8 and initially fitted inside the cryo-EM map of the apo-RNAP.
881 The model was then manually edited in Coot starting from loop β 5-6 to fit correctly the map
882 followed by refinement in Phenix. Comparisons of Rpo8 between our EM structure and other
883 archaeal RNAP structures revealed a discrepancy. In the deposited Sso and Ssh RNAP X-ray
884 structures^{75,76} the GLLM motif (Saci sequence YGLIV, Sso YGLLV), reported to be essential
885 for the binding to RPB1 in yeast²⁵, was erroneously assigned and rendered the motif solvent
886 exposed. In our structure, the conserved GLLM motif is in the same position seen in all
887 eukaryotic RPB8 at the interface with Rpo1' (Figure 1C). The improved sequence alignment
888 shown in figure 1D predicts also that the β 5-6 loop in *S. shibatae* and *S. solfataricus* should
889 be longer than what initially published. Indeed, the electron density maps of both RNAP
890 structures (codes 3hgz and 4ayb) support our findings showing extra-density proximal to the
891 loop which suggests that both loops are longer and partially disordered (figure S5d).

892

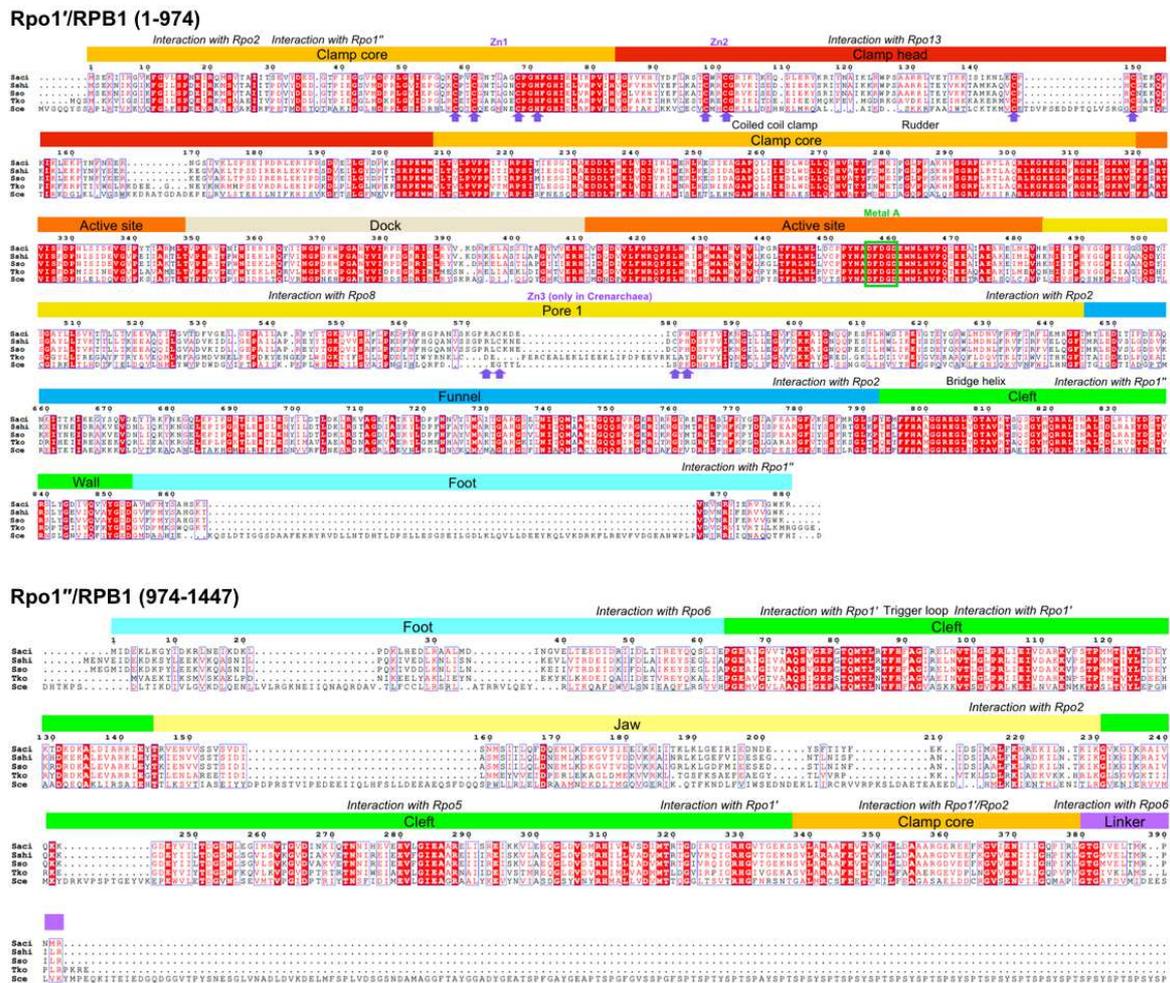
893 Supplementary figures

894

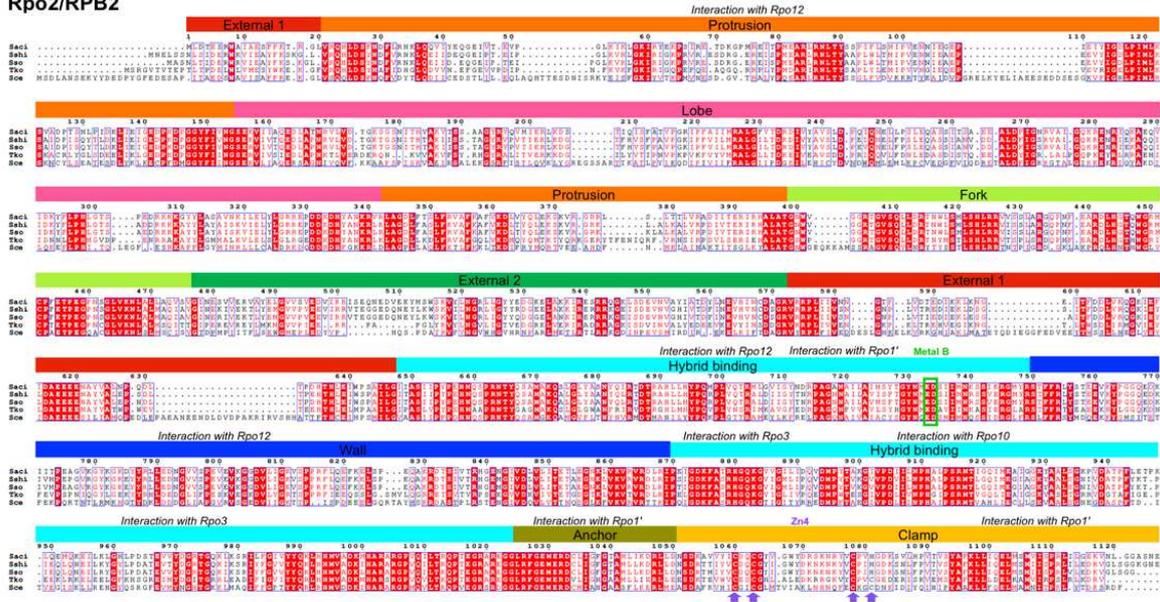


895

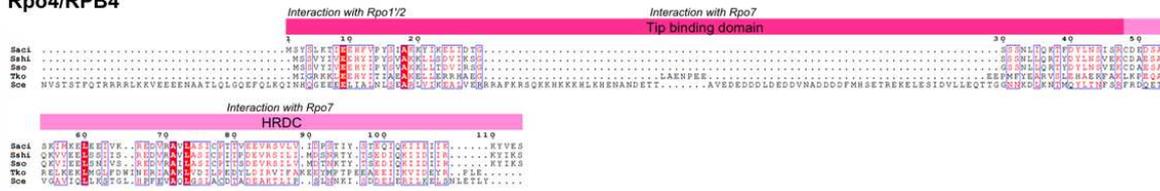
896 Supplementary figure S1. Cryo-EM results and map quality evaluation for the apo-RNA
 897 polymerase. a) SDS-PAGE showing the RNA polymerase before crosslinking. b)
 898 Representative motion-corrected cryo-EM micrograph; scale bar corresponds to 300 Å. c)
 899 Selected averages of the RNA polymerase common views; box size 360.22 Å. d) Local
 900 resolution variation of the RNA polymerase 3D reconstruction; map is presented in two
 901 orientations and coloured as indicated in the scale below. e) Gold standard Fourier Shell
 902 Correlation (FSC) plot obtained from post processing in Relion 3.0. The dashed line represents
 903 0.143 cutoff which indicated a resolution of 2.88 Å. For the curve labelled 'masked', the FSC
 904 was calculated using a mask with soft edges.
 905
 906



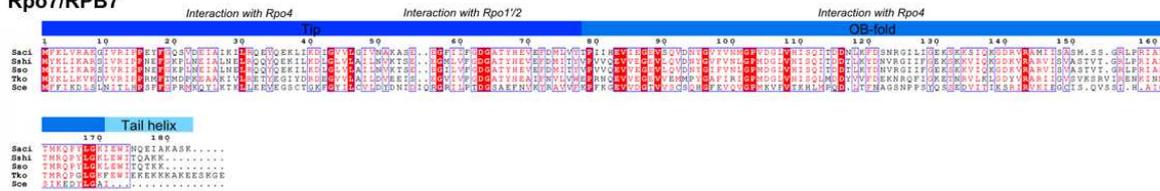
Rpo2/RPB2

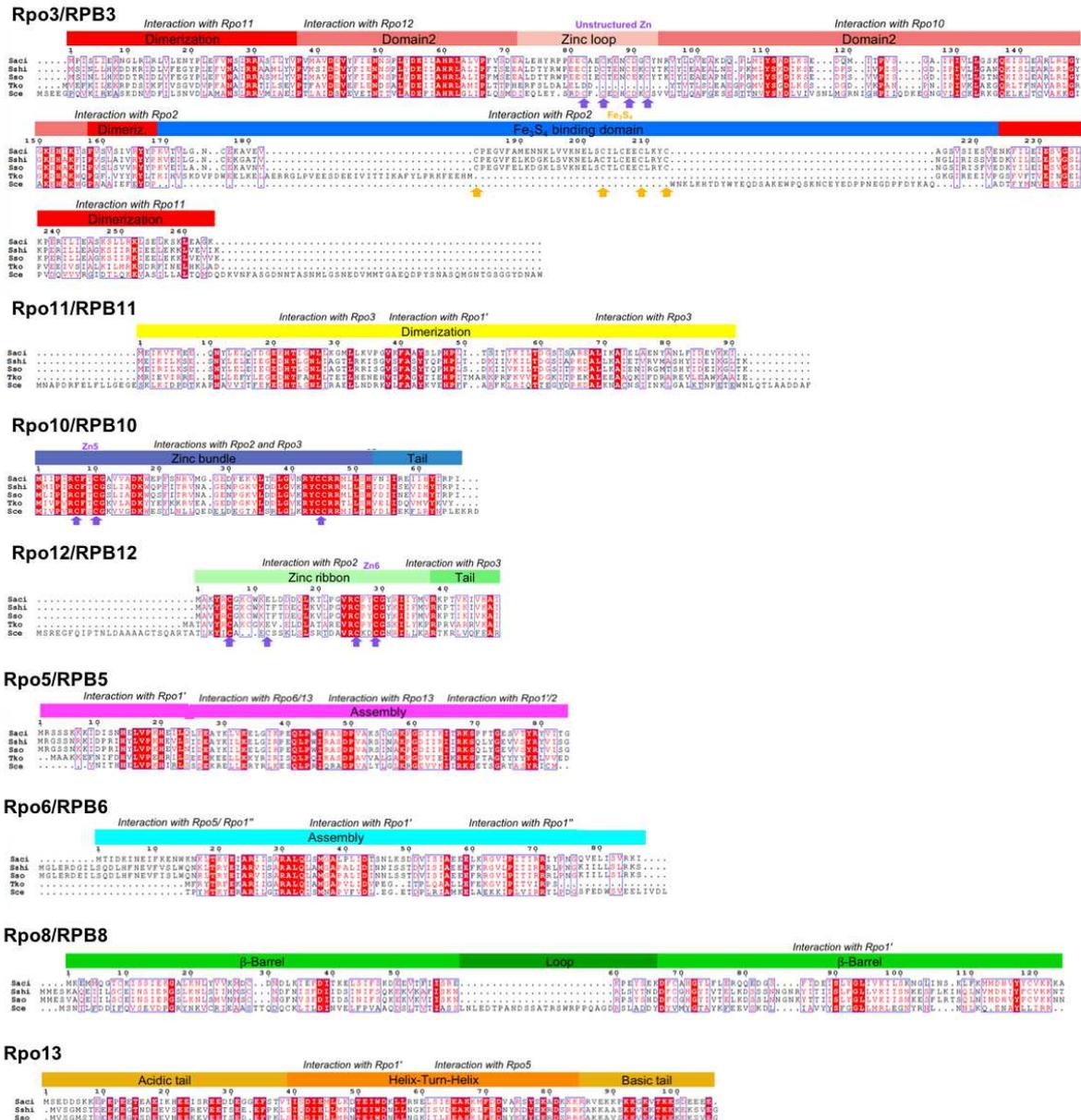


Rpo4/RPB4



Rpo7/RPB7

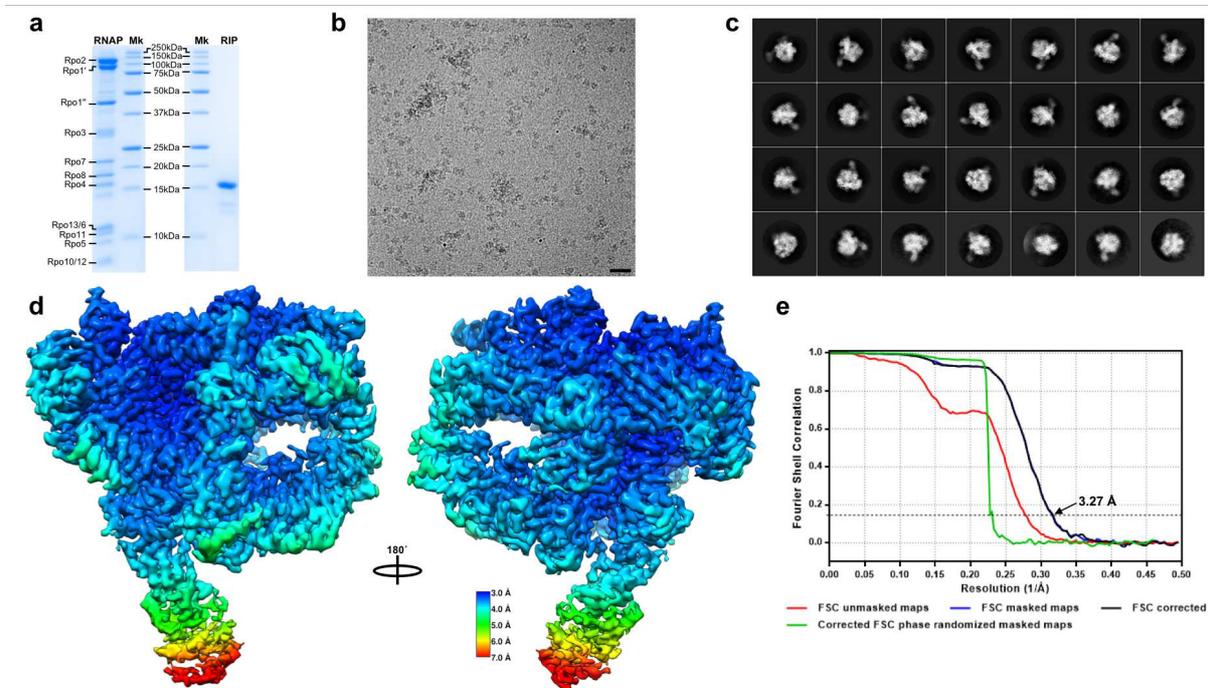




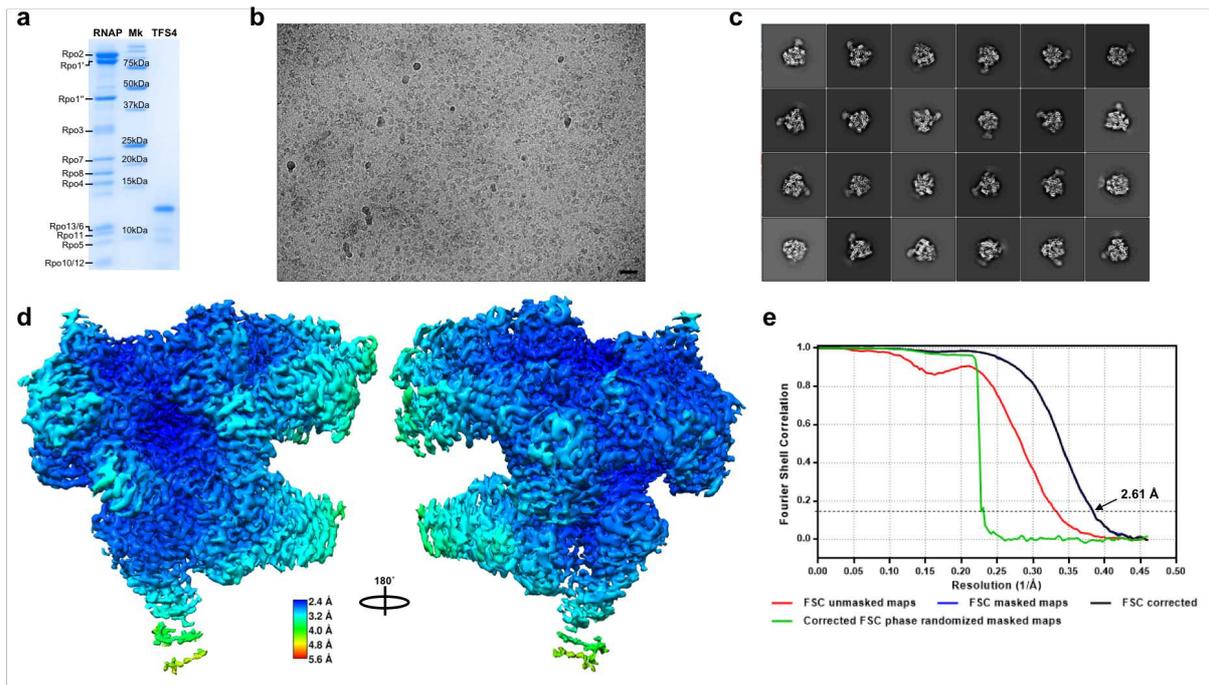
909

910 **Supplementary figure S2.** Structure-based sequence alignment of each single RNAP subunit
 911 from *Sulfolobus acidocaldarius*, *Saccharolobus shibatae* (pdb 4ayb), *Saccharolobus*
 912 *solfataricus* (pdb 3hkh), *Thermococcus kodakarensis* (pdb 6kf3) and the corresponding ones
 913 from *Saccharomyces cerevisiae* (pdb 6gyk). The MSEs have been manually edited to account
 914 for the unstructured loops and the domains assigned according to Cramer et al., 2001 and
 915 Fernandez-Tornero et al., 2013^{33,55}. Magnesium binding sites are highlighted within a green
 916 box, the zinc ligands with purple arrows and the iron-sulfur cluster (Fe₃S₄) ligands in gold
 917 arrows. In archaeal Rpo1' there is a third zinc finger domain which is not conserved in yeast.
 918 The zinc ion in *Saci* Rpo3 is not visible because of the higher flexibility of the loop whereas it
 919 has not been detected in *S. shibatae* or *S. solfataricus* although the cysteine residues are

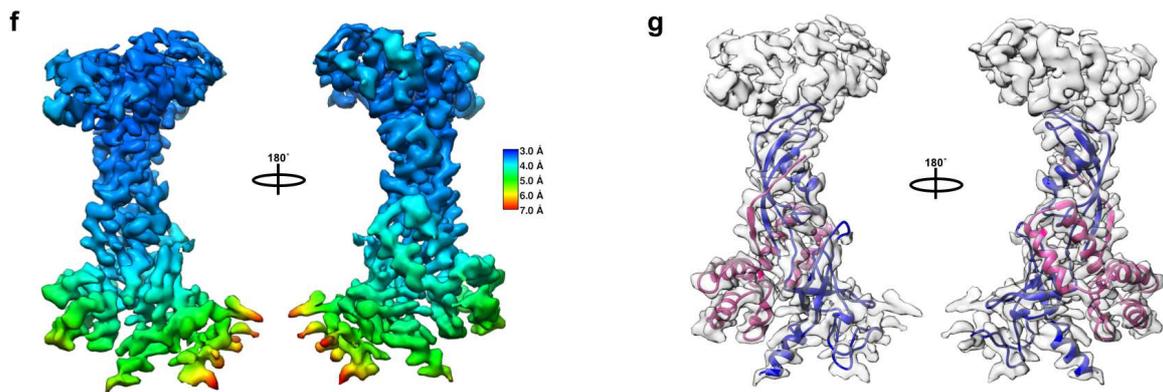
941 the positive signal) confirming that the β 5-6 loop in these two species is longer than what
 942 initially established.
 943
 944



945
 946 **Supplementary figure S4. Cryo-EM results and map quality evaluation for the RNAP-RIP**
 947 **complex.** **a)** SDS-PAGE showing the RNA polymerase and RIP before crosslinking. **b)**
 948 Representative motion-corrected cryo-EM micrograph; scale bar corresponds to 300 Å. **c)**
 949 Selected averages of the RNA polymerase common views; box size 303 Å. **d)** Local resolution
 950 variation of the RNA polymerase 3D reconstruction; map is presented in two orientations and
 951 coloured as indicated in the scale below. **e)** Gold standard Fourier Shell Correlation (FSC)
 952 plot obtained from post processing in Relion 3.0. The dashed line represents 0.143 cutoff
 953 which indicated a resolution of 3.27 Å. For the curve labelled 'masked', the FSC was calculated
 954 using a mask with soft edges.
 955
 956

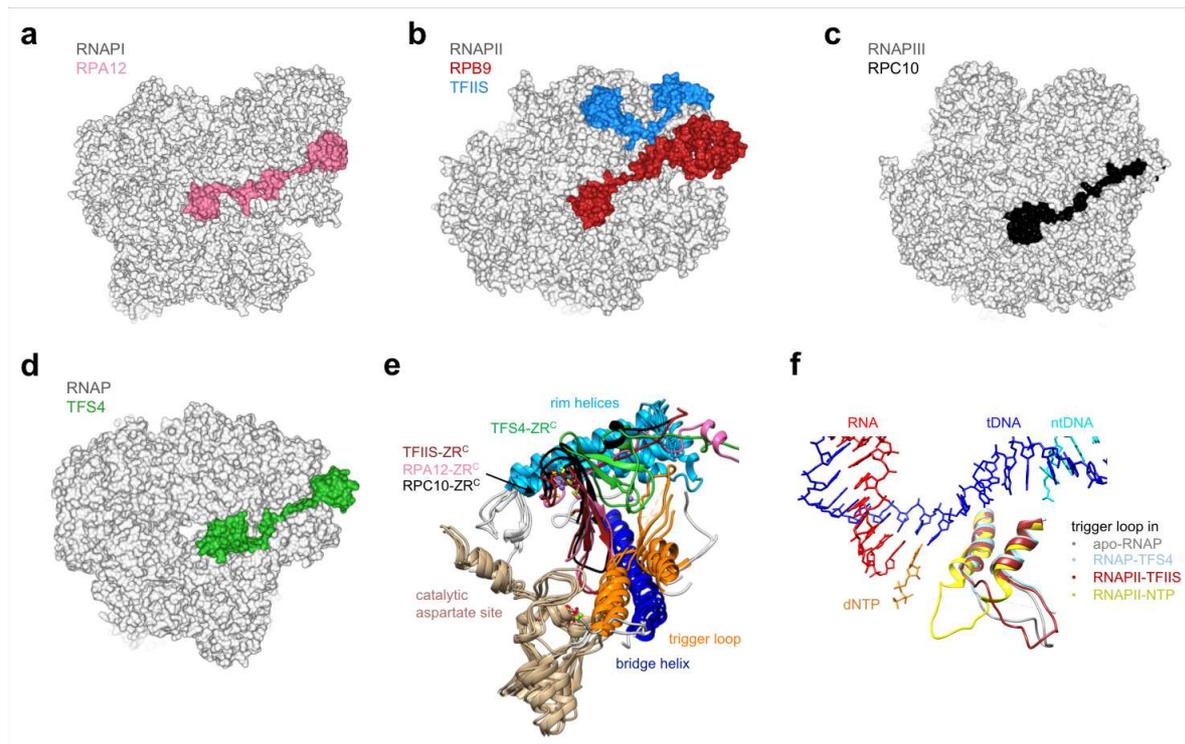


957



958

959 Supplementary figure S5. Cryo-EM results and map quality evaluation for the RNAP-TFS4
 960 complex. **a)** SDS-PAGE showing the RNA polymerase before crosslinking. **b)** Representative
 961 motion-corrected cryo-EM micrograph; scale bar corresponds to 300 Å. **c)** Selected averages
 962 of the RNA polymerase common views; box size 360.22 Å. **d)** Local resolution variation of the
 963 RNA polymerase 3D reconstruction; map is presented in two orientations and coloured as
 964 indicated in the scale below. **e)** Gold standard Fourier Shell Correlation (FSC) plot obtained
 965 from post processing in Relion 3.0. The dashed line represents 0.143 cutoff which indicated a
 966 resolution of 2.61 Å. For the curve labelled 'masked', the FSC was calculated using a mask
 967 with soft edges. **f)** Local resolution variation map of the stalk map region after multi-body
 968 refinement and sharpening. **g)** Rpo4 (in deep pink) and Rpo7 (in blue) subunits modelled
 969 inside the same map shown here as a semi-transparent surface.



984

985 **Supplementary figure S7. TFS-related proteins in RNAPI, II and III and the archaeal RNAP. (a)**
 986 **RNAPI with RPA12 highlighted in pink, (b) RNAPII-TFIIS complex with RPB9 highlighted in blue**
 987 **and TFIIS in dark red, (c) RNAPIII with RPC10 in the 'inside-funnel' state highlighted in black,**
 988 **and (d) RNAP-TFS4 complex with TFS4 highlighted in green. (e) Superimposition of the ZR^C**
 989 **domains of TFS4 and its eukaryotic paralogues TFIIS, A12, and RPC10. All ZR^C domains bar**
 990 **TFS4 adopts an equivalent binding mode inside the secondary channel reaching for the active**
 991 **site, interacting with both the rim helices and the bridge helix whilst not interfering with the**
 992 **trigger loop. In contrast, TFS4 binds at the entrance of the channel and clashes with the trigger**
 993 **loop. (f) Comparison of Trigger loop conformations between apo-RNAP (grey), TFS4-bound**
 994 **(light blue), TFIIS-locked (brown) and NTP-bound (yellow) RNAPs. Locked and closed TLs**
 995 **from yeast RNAPII structures (pdb code 3po3 and 2e2h respectively). RNA, template DNA,**
 996 **non-template DNA and substrate NTP are shown red, blue, cyan and orange**
 997 **respectively. Nucleic acids and substrate NTP were taken from 2e2h to highlight the active**
 998 **site. Bridge helices are omitted for clarity.**

999

Supplementary Files

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

- [Movie1.mp4](#)