# DNA-binding mechanism and evolution of Replication Protein A

Clément Madru[1], Markel Martinez-Carranza[1], Sébastien Laurent[2], Alessandra C. Alberti[1], Maelenn Chevreuil[3], Bertrand Raynal[3], Ahmed Haouz[4], Rémy A. Le Meur[5], Marc Delarue[1], Didier Flament[2], Mart Kuprovic[6], Pierre Legrand[1,7], Ludovic Sauguet[1*]

[1]Architecture and Dynamics of Biological Macromolecules, Institut Pasteur, Université Paris Cité, CNRS UMR 3528, Paris (France), [2]Univ Brest, Ifremer, CNRS, Biologie et Ecologie des Ecoystèmes marins profonds (BEEP), F-29280, Plouzané, France, [3]Molecular Biophysics Platform, C2RT, Institut Pasteur, Université Paris Cité, CNRS UMR 3528, Paris (France), [4]Crystallography Platform, C2RT, Institut Pasteur, Université Paris Cité, CNRS UMR 3528, Paris (France), [5]Biological NMR Platform & HDX, C2RT, Institut Pasteur, Université Paris Cité, CNRS UMR 3528, Paris (France), [6]Archaeal Virology Unit, Institut Pasteur, Université Paris Cité, CNRS UMR 6047, Paris (France) [7]Synchrotron SOLEIL, HelioBio group, L'Orme des Merisiers, 91190, Saint-Aubin (France).

[*]To whom correspondence should be addressed: ludovic.sauguet@pasteur.fr (LS)

## ABSTRACT

Replication Protein A (RPA) is a heterotrimeric single stranded DNA-binding protein with essential roles in DNA replication, recombination and repair, in both eukaryotic and archaeal cells. By using an integrative approach that combines three crystal structures, four cryo-EM structures in complex with single-stranded DNA (ssDNA) of different lengths, we extensively characterized RPA from *Pyrococcus abyssi* in different states. These structures show two essential features conserved in eukaryotes: a trimeric core and a module that promotes cooperative binding to ssDNA, as well as a newly identified archaeal-specific domain. These structures reveal for the first time how ssDNA is handed over from one RPA complex to the other, and uncover an unanticipated mechanism of self-association on ssDNA tracts. This work constitutes a significant step forward in the molecular understanding of the structure and DNA-binding mechanism of RPA, with far-reaching implications for the evolution of this primordial replication factor in Archaea and Eukarya.

## INTRODUCTION

In all forms of life, single-stranded DNA-binding proteins (SSBs) are essential components of the DNA replication machinery (Bain et al., 2018; Oliveira, 2021; Taib et al., 2021). They play vital roles in nearly all aspects of DNA metabolism by protecting exposed single-stranded DNA (ssDNA) and acting as platforms onto which DNA-processing enzymes can assemble (Brosey et al., 2013; Marceau, 2012). In Bacteria, SSB is the major single-stranded DNA binding protein. The archetypal *Escherichia coli* SSB encompasses a single oligonucleotide/oligosaccharide binding (OB) domain that assembles into homotetrameric complexes (Bianco, 2017). In Eukarya, the single-stranded DNA binding function is primarily achieved by the heterotrimeric Replication Protein A (RPA) complex (Wold, 1997). While RPA and SSBs share similar OB-fold DNA-binding domains, the heterotrimeric architecture of RPA is more complex than the homo-oligomeric assemblies of SSBs. Composed of three protein subunits, denoted as Rpa1, Rpa2, and Rpa3 (or RPA70, RPA32, and RPA14 in humans), RPA contains multiple OB-folds with different DNA-binding properties (**Figure 1a**). Over the past two decades, structural and biochemical studies have mostly been focused on the eukaryotic RPA by using X-ray crystallography (Bochkareva, 2002; Deng et al., 2007; Fan and Pavletich, 2012; Feldkamp et al., 2013; Jiang et al., 2006; Seeber et al., 2016), NMR (Brosey et al., 2009; Park, 2005), SAXS (Brosey et al., 2013), and cryo-electron microscopy (Yates et al., 2018).

Archaea, which remains the most unexplored branch of the tree of life, possess an intriguing blend of bacterial and eukaryotic features as well as aspects that are unique to this domain of life (Duggin and Bell, 2006). While archaeal chromosomes resemble those of most Bacteria, their DNA replication machineries are closely related to their eukaryotic counterparts, serving as powerful models for understanding the function and evolution of the complex eukaryotic replication machineries. Thus, it is of considerable interest to understand how the simple bacterial-like chromosomes of Archaea are replicated by a eukaryotic-type replication apparatus (Barry and Bell, 2006). Interestingly, Archaea display a highly differentiated distribution of RPA, with a wide range of domain architectures with varying numbers of OB domains (MacNeill, 2021; Taib et al., 2021). No structure of archaeal RPA has been determined so far, except for structures of isolated OB domains that have been determined in the framework of structural genomics projects (PDBids: 2K50 and 3DM3). Determining the structure of a full-length archaeal RPA is required to fully appreciate its cellular function in the third domain of life.

We report three crystal structures and four cryo-EM structures of *P. abyssi* RPA (hereafter referred to PabRPA), in its apo form as well as bound to ssDNA substrates of different lengths. The structures include a trimeric core that is shared with eukaryotes and additional domains that are specific to Archaea. These structures reveal for the first time how ssDNA is handed over from one RPA molecule to the other. PabRPA is composed of two distinct modules, a trimerization core (Tri-C) with high-affinity for ssDNA and a second module located in the Rpa1 N-terminal region, which promotes cooperative binding to ssDNA. By using an integrative approach that combines X-ray crystallography, cryo-electron microscopy, and extensive biophysical analysis, we investigated the role of each individual

domain of PabRPA and uncovered an unanticipated mechanism of self-association on ssDNA tracts, which has far-reaching implications for the DNA-binding mechanism of RPA in archaea and eukaryotes. We found that PabRPA forms tetrameric supercomplexes in the absence of DNA. In the presence of DNA, these tetramers undergo large conformational changes to form functional nucleoprotein filaments, which coat and protect ssDNA. This work also clarifies the evolutionary history of RPA, suggesting that RPA evolved in Archaea from a bacterial-like SSB ancestor, and that the eukaryotic RPA was inherited from an archaeal RPA ancestor.

## RESULTS

### Architecture of the archaeal RPA heterotrimeric complex

Due to its modular nature, RPA is extremely flexible and can adopt multiple conformations. X-ray crystallography studies have therefore focused on individual domains or truncated RPA trimeric cores (Bochkareva, 2002; Deng et al., 2007; Fan and Pavletich, 2012; Feldkamp et al., 2013; Jiang et al., 2006; Seeber et al., 2016). Here, we have biochemically reconstituted, crystallized and determined the 3.7 Å X-ray crystal structure of a heterotrimeric complex of *P. abyssi* Rpa1, Rpa2, and Rpa3 subunits (**Figure 1b**). An initial, partial model was obtained by using phase information derived from anomalous scattering data collected at the Zn K-edge. Model building was facilitated by determining two additional X-ray crystal structures of the N-terminal region of Rpa1 (1-180) at 1.8 Å resolution (**Figure 1c**) and the trimerization core bound to ssDNA at 3.2 Å resolution (**Supplementary figure S1 & Supplementary table S1**). Most of Rpa1(2-358), Rpa2(1-184) and Rpa3(6-117) subunits have been modelled in the electron density map, the only flexible region being the C-terminus of Rpa2(185-268). Bioinformatic studies have shown that this region contains a Winged-Helix (WH) domain that is conserved from Archaea to Eukarya, and recruits protein factors involved in DNA metabolism (MacNeill, 2021; Makarova and Koonin, 2013).

PabRPA contains four OB domains, named OB-1, OB-2, OB-3 and OB-4, which share the highest structural similarity with OB-A, OB-C, OB-D and OB-E in eukaryotic RPA, respectively (**Figure 1a**). Like in its eukaryotic OB-C counterparts, the amino acid sequence of PabRPA OB-2 contains a Zn-finger motif that is inserted between β-strands β1 and β2, and a short helical domain that is inserted between β-strands β3 and β4 (**Figure 1b & Supplementary figure S2**). The trimerization core (Tri-C) adopts a compact quaternary structure consisting of OB-2 of Rpa1, OB-3 of Rpa2, and OB-4 of Rpa3, which is the smallest subunit. Similar to the eukaryotic RPA, heterotrimerization of PabRPA is primarily mediated through a three-helix bundle formed by a C-terminal α-helix from each subunit (**Supplementary Movie S1**). PabRPA also displays intriguing specificities. Interestingly, the trimerization helix of Rpa2 is followed by a helix-turn-helix motif, acting as a pedestal that supports the trimerization helical bundle and stabilizes it by extending the contacts between the Rpa2 and Rpa3 subunits (**Figure 1b & Supplementary figure S2**). Thus, the total buried surface is 1209 $Å^2$ for the PabRpa3 trimerization helix, but only 774 $Å^2$ for its eukaryotic counterpart (PDBid: 4GOP). The chemical nature of the interactions is varied and includes a network of salt bridges and polar bonds, as well as extensive contacts between hydrophobic residues. Interestingly, the C-

109 terminal extension of Rpa2 seems to be specific to Thermococcales, suggesting that it may play
110 a role in further stabilizing the RPA trimerization core in organisms living at extremely high
111 temperatures.

112 The N-terminal region of Rpa1 adopts an extended conformation that contrasts with the
113 compact structure of the trimerization core and is shown here to play a critical role in the
114 cooperative binding of PabRPA to ssDNA. Therefore, we named this independent DNA-
115 binding module, the "cooperativity module". The OB-1 domain of Rpa1 is bordered by an N-
116 terminal tri-helical bundle, which has never been described so far and is referenced in the Pfam
117 database (Mistry et al., 2021) as a domain of unknown function (DUF2240). This domain is
118 absent in eukaryotes, but is broadly present in Archaea, being located at the N-terminal end of
119 Rpa1 and Rpa1-like subunits, with very rare exceptions (**Figure 1c**). It is present in nearly all
120 Thermoplasmata, Halobacteria and Asgard, but absent from Thermoproteota, which do not
121 encode Rpa or Rpa-like proteins (**Supplementary figure S3**). In the following, we refer to this
122 N-terminal domain as the Archaeal Rpa1 Connecting Domain (ARCD). Interestingly, in the
123 PabRPA crystal structure, ARCD stacks against the OB-4 domain of neighboring Rpa3 chain
124 within a lattice contact (**Figure 1b & Supplementary movie S1**). In the following sections, we
125 describe the biological function of this new domain and show that it contributes to cooperative
126 binding, by pre-connecting together RPA molecules before they bind to DNA.

127 **DNA-binding properties of PabRPA trimerization core and cooperativity module**

128 The DNA-binding properties of PabRPA were investigated by electrophoretic mobility
129 shift assays (EMSA), surface plasmon resonance (SPR) and biolayer interferometry (BLI)
130 (**Figure 2a** & **Supplementary figure S4**). As expected, we found that PabRPA specifically
131 binds to a 24-mer or a 32-mer random ssDNA with a dissociation constant in the low nanomolar
132 range, but hardly binds to dsDNA or ssRNA (**Supplementary figure S4a-c**). As the use of
133 homopolymeric ssDNA substrates is more convenient for structural studies, we verified that
134 PabRPA binds to a poly-dT ssDNA, with an affinity similar to random ssDNA (1.25±0.6 nM
135 vs 0.71±0.01 nM) (**Figure 2a** & **Supplementary figure S4d**). The DNA-binding properties of
136 the PabRPA trimerization core and cooperativity module were further investigated by using
137 BLI. Overall, our data show that binding to ssDNA is primarily ensured by the trimerization
138 core. Yet, while both modules are required for an optimal binding to ssDNA, the trimerization
139 core binds ssDNA with a 35-fold higher affinity (Kd=18.8±6.0nM) than the cooperativity
140 module (Kd=650±150nM). Furthermore, the cooperativity module shows much faster
141 dissociation kinetics than the trimerization core does. Indeed, about 60% of the PabRPA
142 cooperativity module dissociates from ssDNA in 10 seconds, while in the same period, the
143 trimerization core remains almost fully associated to ssDNA.

144 **Crystal structure of the PabRPA trimerization core bound to ssDNA**

145 To gain insights into the structural basis for the high-affinity interaction between
146 PabRPA and ssDNA, we determined the 3.2 Å crystal structure of the trimerization core bound
147 to poly-dT20 ssDNA (**Figure 2b**). The modelled ssDNA lies in a channel that extends from the

148 subunits Rpa1 to Rpa2, showing 14 contiguous nucleotides that traverse the OB-2 and OB-3
149 DNA-binding grooves. The DNA-binding grooves are lined by the β-strands of each OB
150 domain and bordered by their connecting loops. Key DNA contacts are made by aromatic or
151 hydrophobic residues that stack or make van der Waals interactions with the ssDNA bases, and
152 basic residues contacting the phosphodiester backbone (**Figure 2b**). Additionally, several polar
153 and acidic residues contact DNA bases. Strikingly, Rpa3 does not contact ssDNA in the
154 structure of PabRPA. The DNA-binding properties of each subunit of PabRPA were tested
155 individually by using SPR. While both Rpa1 and Rpa2 individual subunits are capable of
156 binding ssDNA, no interaction was observed for the Rpa3 subunit (**Supplementary figure**
157 **S4c**). This property is shared with the eukaryotic RPA where Rpa3 plays a structural role at the
158 heart of the trimerization core but is not primarily involved in ssDNA binding (Bochkareva et
159 al., 1998; Brill and Bastin-Shanower, 1998; Sibenaller et al., 1998).

160 Comparing the structures of ssDNA-bound PabRPA and the eukaryotic RPA from a
161 fungus *Ustilago maydis* (Fan and Pavletich, 2012) also reveal intriguing differences
162 (**Supplementary figure S5**). Indeed, although the chemical nature of the interactions and the
163 overall ssDNA pathway through RPA are conserved, the DNA-binding grooves of their OBs
164 are substantially different in Archaea and Eukarya. In the structure of PabRPA, basic residues
165 of the β3-β4 helical insertion in OB-2 contact the DNA phosphodiester backbone and delineate
166 a crevice that is narrower than in the structure of the eukaryotic RPA (**Figure 2b** &
167 **Supplementary figure S5a**). The helical insertion also exists in eukaryotes but adopts a
168 different conformation and is devoid of conserved basic residues (**Supplementary figure S5a**),
169 thereby suggesting that it may be an evolutionary relic derived from an archaeal ancestor. DNA
170 contacts are also substantially different at the level of OB-3 (OB-D in Eukarya). In PabRPA,
171 an archaeal-specific β-hairpin is inserted within the α1- β1 connecting loop (266-311) and
172 makes extensive contacts with ssDNA (**Figure 2b** & **Supplementary figure S2**). This is
173 particularly interesting as we show later that this loop acts as a molecular switch, which
174 promotes a tetrameric super-complex of PabRPA in absence of ssDNA. On the other hand, the
175 contacts between the OB-2 Zn-finger domain and ssDNA are remarkably conserved between
176 Archaea and Eukarya. Indeed, in the vicinity of their Zn-fingers domains, the interacting
177 residues and the two neighboring bases are perfectly superimposable.

178 **Cryo-EM structure of the full-length PabRPA bound to short ssDNA**

179 We collected a cryo-EM SPA (single-particle analysis) dataset of a purified complex of
180 the full-length PabRPA and a 35 nucleotide-long poly-dT ssDNA (dT35) (**Figure 2c and**
181 **Supplementary figure S6**). Given the small size of this protein-nucleic acid complex (96 kDa)
182 and its intrinsic flexibility due to the modular architecture of RPA, our best 3D reconstruction
183 is limited to a global resolution of 6.3 Å (**Supplementary table S2**). However, the 3D
184 reconstruction shows distinguishable structural features, such as the trimeric helical bundle, and
185 individual OB domains that correspond to the trimerization core. Yet, the PabRPA DNA-bound
186 Tri-C can be fitted into the EM density map by using rigid-body refinement (**Supplementary**
187 **figure S6**). The refined map, as well as the 2D-class averages, show additional density for OB-
188 1 that is positionally flexible. This additional density also shows connectivity to the Tri-C,

189 which can be accounted for by ssDNA traversing the DNA-binding grooves of the OB-2 and
190 OB-1 domains. The single-stranded DNA region across OB-1 and OB-2 was modelled by
191 superimposing PabRPA OB-1 onto the DNA-bound crystal structure of *Homo sapiens* OB-A
192 (Bochkarev et al., 1997) (**Figure 2c** & **Supplementary figure S5b**). Comparisons with our
193 PabRPA apo crystal structure reveal that OB-1 and OB-2 are positioned differently in the DNA-
194 bound cryo-EM reconstruction, adopting a more extended conformation than in the crystal
195 structure. Electrostatic potential surface analysis reveals that repositioning the OB-1 domain
196 does extend the DNA-binding channel from the trimerization core to the cooperativity module,
197 thereby facilitating the passage of ssDNA from one module to the other.

198 **Structural basis for the assembly of multiple PabRPA molecules on ssDNA**

199 *In cellulo*, RPA binds and protects much longer ssDNA tracts than the ones used in the
200 assays described above. The ability of PabRPA to bind to long ssDNA substrates was initially
201 assessed by using negative-staining electron microscopy (**Figure 3a**). By using a concentration
202 of PabRPA that accounts for the binding of one PabRPA to ~30 nucleotides, PabRPA formed
203 nucleoprotein filaments that fully coated a 6.4kb pM13 single-stranded plasmid, showing that
204 purified PabRPA is capable of self-association on long ssDNA. Then, a cryo-EM structural
205 analysis of multiple RPA molecules bound to DNA was undertaken, by using a poly-dT100
206 oligonucleotide to form an RPA/ssDNA complex containing multiple RPA molecules
207 (**Supplementary figure S7** & **Supplementary table S2**). Our best reconstruction was
208 determined at a subnanometer overall resolution of 8.2 Å showing two central PabRPA
209 complexes sandwiched by two additional molecules that are only partially defined in the cryo-
210 EM map (**Figure 3b**). Distinguishable structural features, such as the trimeric helical bundle,
211 and individual OB domains can be observed for the two central molecules in the 2D-class
212 averages, as well as in the refined 3D model. Docking the model of the PabRPA/dT35 complex
213 into the PabRPA/dT100 cryo-EM reconstruction shows that adjacent RPA complexes are
214 connected through an interaction between the Rpa1 OB-1 domain of one RPA complex with
215 the trimerization core of one downstream complex (**Figure 3c & Supplementary movie 2)**.
216 However, the exact positioning and orientation of OB-1 with respect to the trimerization core
217 of the neighboring PabRPA complex could not be determined confidently.

218 To explore the molecular details of this interaction, we ran an AlphaFold prediction of
219 the protein/protein interactions between two individual PabRPA complexes. The prediction was
220 obtained with a very high level of confidence, predicting that OB-1 contacts simultaneously
221 OB-3 and OB-4 from the adjacent trimerization core (**Figure 3c & Supplementary figure**
222 **S8a**). AlphaFold predicted protein complex can be nicely fitted in the cryo-EM volumes. DNA
223 was modelled according to our ssDNA-bound crystal and cryo-EM structures of PabRPA
224 (**Figure 2**). Strikingly, the interaction between OB-1(n) and OB-3(n-1) brings together the 3'-
225 end and 5'-end of the modelled ssDNA from two neighboring RPA complexes. The surface
226 electrostatic potential reveals a continuous positively charged cleft that can guide ssDNA from
227 one RPA molecule to the other. Thus, PabRPA nucleoprotein filaments form a polar and
228 continuous groove that efficiently binds and protect ssDNA from nucleases (**Figure 3e &**
229 **supplementary movie 2**).

230 The role of the PabRPA cooperativity module was further investigated by using
231 microscale-thermophoresis (MST), an approach that was successfully used to study DNA-
232 binding cooperativity of the yeast RPA (**Figure 3d**) (Yates et al., 2018). As expected, plotting
233 as a linear Hill plot the ssDNA-binding data of full-length PabRPA on long ssDNA (dT100)
234 shows a Hill coefficient of 2.0, thereby confirming that it cooperatively binds to ssDNA.
235 However, plotting ssDNA-binding data of a PabRPA construct deleted from its entire
236 cooperativity module shows a Hill coefficient of 1.0, indicating that this truncated PabRPA
237 construct no longer binds ssDNA cooperatively. Thereby, the critical role of this module in
238 DNA binding cooperativity has been established both by the Cryo-EM structure and by MST.
239 The role of the Archaea-specific domain ARCD was also investigated by using MST.
240 Interestingly, deleting ARCD impaired only partially the binding cooperativity of PabRPA (Hill
241 coefficient of 1.8), indicating that OB-1 plays a more prominent role in cooperativity than
242 ARCD does **(Figure 3d)**. Furthermore, ARCD was not visible in the density of the
243 PabRPA/dT100 cryo-EM reconstruction. This suggests that it participates in PabRPA activity
244 by pre-connecting PabRPA complexes together before they bind DNA but becomes dispensable
245 once the complex is bound to DNA (**Supplementary figure S8b**). Consistently, we verified
246 that deleting ARCD does not alter the DNA-binding affinity of PabRPA to a dT35 short ssDNA
247 suggesting that ARCD is involved in protein/protein interactions rather than protein/DNA
248 interactions (**Supplementary figure S8c**).

249 **PabRPA forms a tetrameric supercomplex that dissociates upon binding to ssDNA**

250 Unexpectedly, aside from assessing the coating of ssDNA by PabRPA, we also
251 discovered that PabRPA specifically assembles as a tetrameric supercomplex in the absence of
252 DNA by using SEC-SLS (**Figure 4a**). SAXS measurement confirmed that PabRPA
253 oligomerizes spontaneously in solution, forming a compact structure with a radius of gyration
254 (Rg) of 57 Å and a Dmax of 207 Å (**Supplementary figure S9a**). Upon addition of a poly-
255 dT25 ssDNA substrate, the PabRPA tetrameric form dissociates into a smaller complex. To
256 decipher the molecular basis of PabRPA oligomerization in the absence of DNA, the structure
257 of the PabRPA tetrameric supercomplex was determined by cryo-EM at 3.35 Å resolution
258 (**Figure 4b & Supplementary figure S10**). The PabRPA tetrameric supercomplex forms from
259 two dimers (AB and CD) arranged into a C2-symmetrical assembly, which interact through
260 contacts involving their trimerization cores. Interactions within the AB and CD dimers involve
261 their OB-3 domains, while the two dimers are held together by interactions involving their OB-
262 4 domains (**Figure 4c**). PabRPA monomers interact specifically to form the tetramer, burying
263 a great deal of exposed protein surface (2076 $Å^2$ per monomer). These observations suggest
264 that assembly of the PabRPA tetrameric complex is thermodynamically favorable in the
265 absence of ssDNA. To evaluate the influence of temperature, the light scattering experiment
266 was repeated at 65°C to get closer to the natural growth temperature of *P. abyssi*. At this
267 temperature, the tetrameric state of PaBRPA was maintained and the intrinsic viscosity
268 increased from 5.0ml/g at 20°C to 5.7ml/g at 65°C showing no change in the overall compact
269 structure of the tetramer but a small gain in flexibility (**Supplementary figure S9b**).

This oligomerization mechanism is strikingly different from the one observed in the DNA-bound PabRPA filaments, which form upon contacts between the cooperativity module and the trimerization core of adjacent RPA molecules. Yet, oligomerization of PabRPA molecules in the absence of ssDNA seems to be independent of the cooperativity module, which was not visible in the cryo-EM reconstruction and appeared to be flexible. SEC-SLS experiments were reproduced in presence or absence of ssDNA by using a truncated version of PabRPA that encompasses only its trimerization core, and lacks the cooperativity module (**Figure 4a**). Consistently, the trimerization core was shown to form a tetramer in solution, and a clear shift into a smaller complex was induced by adding dT25 ssDNA, thereby demonstrating that assembly/dissociation of the PabRPA tetrameric supercomplex is modulated through binding of ssDNA to the Tri-C. It is striking that the OB-3 DNA binding groove is completely occluded in the PabRPA tetrameric supercomplex (**Figure 4c & Supplementary movie S3**). Indeed, many interfacial residues that stabilize the PabRPA tetrameric assembly are also found to bind ssDNA, rationalizing why addition of DNA competes with these intermolecular interactions and promotes dissociation of the PabRPA tetrameric supercomplex. Interestingly, most contacts within the AB and CD dimers involve the α1-β1 β-hairpin of OB-3, which also make extensive contacts with ssDNA (**Figure 2b & Supplementary figure S5**). Thus, the β-hairpin plays a dual role: it drives oligomerization of PabRPA in the absence of DNA, but contacts extensively the substrate in the presence of DNA. Therefore, we hypothesize that this archaeal-specific β-hairpin acts as a molecular switch that promotes a tetrameric super-complex of PabRPA in the absence of ssDNA, contributes to binding the DNA substrate and dissociate the tetramer in the presence of DNA.

Finally, by using a subset of the dataset (**Supplementary figure S10**), we determined an asymetric 3.8 Å cryo-EM structure of the PabRPA tetrameric supercomplex showing the interaction between the cooperativity module of one RPA complex and the OB-4 domain of another RPA (**Figure 4d**). Remarkably, the ARCD/OB-4 interface observed in this new cryo-EM structure matches the one observed in the PabRPA apo crystal structure (**Figure 1b**). This new structure shows how the archaeal-specific ARCD connects two RPA molecules in a conformation that is compatible with their adjacent assembly onto ssDNA.

## DISCUSSION

In all forms of life, coating and protecting exposed ssDNA from nucleases is essential (Oliveira, 2021). In Eukarya, RPA has been shown to be an abundant multi-domain heterotrimeric protein complex, which is essential to all DNA processing events in DNA replication, recombination and repair (Bain et al., 2018; Brosey et al., 2013; Dueva and Iliakis, 2020). Here, we explored the quaternary structure of PabRPA and the conformational changes that occur upon DNA-binding.

### Self-association ssDNA-binding mechanism of PabRPA

We have extensively characterized the RPA from *P. abyssi*, a hyperthermophilic archaeon, and determined three crystal structures and four cryo-EM structures of PabRPA in

309 different states. By using an integrative approach, we uncovered an unanticipated mechanism
310 of self-association of PabRPA on ssDNA (**Figure 4e**). These structures reveal for the first time
311 how ssDNA is handed over from one RPA molecule to the other. In its apo form, PabRPA
312 assembles into tetrameric surpercomplexes of four RPA protomers, whose association is
313 mediated by interactions within OB domains of their trimerization cores (**Figure 4e**).
314 Additional interactions are observed within the cooperativity module and the trimerization
315 cores of two neighboring RPA molecules. Furthermore, the Archaea-specific ARCD domain of
316 one PabRPA molecule has been shown to bind specifically to the OB-4 domain of an adjacent
317 molecule, in the crystal and cryo-EM structures of PabRPA, which were determined in their
318 apo forms. This interaction allows pre-assembly of PabRPA molecules in a conformation that
319 is pre-determined to efficiently coat ssDNA, once the tetramer is dissociated. PabRPA primarily
320 binds ssDNA through its trimerization core, which triggers dissociation of the tetrameric
321 supercomplex. Binding to ssDNA also induces a conformational change within the
322 cooperativity module, untightening the interaction between ARCD(n+1) and OB-4(n) and
323 promoting the binding of the OB-1(n+1) domain to the OB-3(n) and OB-4(n) domains of the
324 neighboring PabRPA complex. This interaction is essential for handing over ssDNA between
325 two adjacent RPA protomers, resulting in coating and protection of ssDNA regions. This
326 cooperative binding mechanism enables the rapid and efficient coating of ~100 nucleotides per
327 tetramer of PabRPA. The Rpa2 WH domain does not contribute to stabilization of the
328 PabRPA/ssDNA filaments, but instead is ideally positioned to recruit other protein factors
329 involved in DNA metabolism (**Figure 4e**).

330 **Comparison with eukaryotic RPA**

331 The structures of PabRPA show three essential features conserved in eukaryotes: a
332 trimeric core, a module that promotes cooperative binding to ssDNA and a flexible WH domain,
333 as well as Archaea-specific additional domains. The proposed DNA-binding mechanism of
334 PabRPA displays profound similarities with the dynamic model proposed for the yeast RPA
335 (Yates et al., 2018). Indeed, both studies on PabRPA and yeast RPA support a linear
336 arrangement of ssDNA within RPA (Yates et al., 2018), which differs with the compact horse-
337 shoe conformation that was observed in the crystal structure from *U. maydis* (Fan and Pavletich,
338 2012). PabRPA possesses a trimerization core and a cooperativity module, which show distinct
339 DNA-binding properties. While the trimerization core associates with ssDNA in a high-affinity
340 stable complex, the cooperativity module binds ssDNA with much lower affinity and higher
341 dissociation kinetics. These findings echo earlier structural, biochemical and biophysical
342 studies showing that the eukaryotic RPA can associate with ssDNA in different modalities, a
343 low-affinity mode involving the Rpa1 N-terminal OB domains and a high-affinity compact
344 mode involving all four major OBs (Arunkumar et al., 2003; Brosey et al., 2015,
345 2013)(Blackwell and Borowiec, 1994; Fan and Pavletich, 2012). Moreover, multiple RPA
346 molecules assemble on ssDNA by using similar protein-protein interactions in *P. abyssi* and *S.*
347 *cerevisiae*. Yet, the OB-4(n)/OB-1(n+1) protein-protein interaction in PabRPA (OB-E(n)/OB-
348 A(n+1) in eukaryotes, respectively) plays a critical role when multiple RPAs are associated
349 with ssDNA. In PabRPA, OB-1 (OB-A in eukaryotes) has been shown to be part of a
350 cooperativity module that interacts with the trimerization core of a downstream RPA, thereby

351 enabling the passage of ssDNA across adjacent RPA molecules. Similarly, structural and
352 biophysical studies support that a similar OB-A/OB-E interaction also exists in the yeast RPA
353 and is required for promoting RPA-RPA complexes (Yates et al., 2018). In addition, the OB-4
354 domain of PabRpa3 does not bind DNA but is a dedicated platform for protein-protein
355 interactions, a property that is shared with the eukaryotic RPA (Bochkareva et al., 1998; Brill
356 and Bastin-Shanower, 1998; Sibenaller et al., 1998). Similarly to previous study on eukaryotic
357 RPA(Yates et al., 2018), the PAbRPA2-WH domain remains unseen in both crystallography
358 and cryo-EM structures, most likely due to its highly flexible attachment to the RPA trimeric
359 core. This conserved flexibility is probably a key element that guaranties its accessibility.

360 A recent structure of the RPA-like human CST complex revealed a decameric assembly
361 bound to telomeric DNA (Lim et al., 2020). In Archaea, the PabRPA tetrameric supercomplex
362 may play a regulatory role by clustering multiple PabRPA complexes, which can rapidly and
363 efficiently bind to exposed ssDNA tracts. Our work strongly suggests that
364 association/dissociation of the PabRPA tetrameric supercomplex is controlled by a β-hairpin
365 molecular switch, which is located in the N-terminal region of Rpa2. However, multimerization
366 of RPA, such as the PabRPA tetrameric supercomplex, has not been reported in eukaryotes.
367 Yet, the equivalent N-terminal region of Rpa32 in eukaryotes hosts important regulatory
368 phosphorylation sites that control the DNA damage signaling and repair activities of RPA
369 (Dutta and Stillman, 1992; Kõivomägi et al., 2011, p. 1; Maréchal and Zou, 2015; Pan et al.,
370 1994; Stephan et al., 2009). This observation highlights the key regulatory roles that the N-
371 terminal region of Rpa2 (Rpa 32 in eukaryotes) play in DNA-binding activities of RPA.
372 PabRPA also shows intriguing specificities, such as the Archaea-specific ARCD domain, which
373 allows pre-assembly of RPA molecules in a conformation that is pre-determined to efficiently
374 coat ssDNA. Strikingly, while this domain is widely distributed among Archaea, it is absent in
375 eukaryotes. Instead of ARCD, the eukaryotic Rpa1 N-terminal extension includes an extra OB
376 domain that is dedicated to protein-protein interactions, and is linked to OB-A by a 50 amino-
377 acid long linker with no predicted secondary structures. Aside from its role in connecting RPA
378 protomers together, one may hypothesize that ARCD also recruits genome maintenance factors
379 once PabRPA is bound to DNA.

380 While the newly characterized ARCD domain is widely distributed across Archaea,
381 being present in nearly all Rpa1 and Rpa1-like subunits, the Rpa2 C-terminal helicoidal
382 extension and the β-hairpin molecular switch of PabRPA may be specific to particular groups
383 of Archaea. Interestingly, Archaea possess a highly differentiated distribution of RPA
384 (MacNeill, 2021; Taib et al., 2021). *Pyrococcus furiosis* and *Thermococcus kodakarensis*
385 encode three RPA subunits, which form a heterotrimer with high affinity for ssDNA (Komori
386 and Ishino, 2001; Nagata et al., 2019). *Methanosarcina acetivorans* also encode three RPA-like
387 SSBs, but they do not interact, and all form homodimers (Robbins et al., 2004). Non-canonical
388 RPA molecules, which display intriguing DNA-binding properties, have also been
389 characterized in *Haloferax volcanii* but the exact stoichiometry of their complexes is unknown
390 (Stroud et al., 2012). In Archaea, RPA displays a wide range of domain architectures with
391 varying numbers of OB domains. For example, one or three additional OB domains are
392 respectively inserted between the OB-1 and OB-2 domains in *M. acetivorans* and *M. jannaschii*

393 Rpa1 subunits (MacNeill, 2021), and this may affect the DNA-binding properties of these
394 RPAs. Therefore, extending this study to RPA from other archaeal species would be of
395 particular interest.

396 **Evolution of RPA in Archaea and Eukarya**

397 This work constitutes a significant step towards the molecular understanding of the
398 structure and DNA-binding mechanism of RPA, with important implications for the evolution
399 of this vital replication factor in Archaea and Eukarya. In Bacteria, the archetypal SSB is the
400 major single-stranded DNA binding protein (Theobald et al., 2003). It encompasses a single
401 OB domain and assembles into homotetrameric complexes (Meyer and Laine, 1990).
402 Eukaryotes also encode single OB-fold SSBs, but their function is restricted to DNA damage
403 repair, whereas the major ssDNA-binding component of the replisome consists of a
404 heterotrimeric RPA complex (Brosey et al., 2013). In addition to RPA, several RPA-like
405 complexes have evolved to perform various specialized roles, as in the case of the CST (Cdc13-
406 Stn1-Ten1) complex, which is essential for telomer maintenance (Lim et al., 2020; Miyake et
407 al., 2009; Surovtseva et al., 2009). By contrast, Archaea display a patchier distribution of SSB
408 and RPA proteins, with some archaeal lineages encoding either one or both of the two systems
409 (Raymann et al., 2014; Wadsworth, 2001). The genomic neighborhoods of the archaeal RPA
410 genes are also highly variable (**Supplementary figure S3**). Such SSB/RPA distribution
411 patterns prompt questions regarding the emergence of the trimeric RPAs and the provenance of
412 the different eukaryotic ssDNA-binding proteins. These questions are challenging to address
413 using sequence-based comparisons due to small size and high divergence of the OB domains.
414 Instead, we performed here an all-against-all structural comparison of the OB domains from
415 representative SSB and RPA originating from Bacteria, Eukarya and Archaea as well as
416 eukaryotic RPA-like CST complex (**Figure 5a**). The available structures were supplemented
417 with AlphaFold models (Jumper et al., 2021) of additional crenarchaeal and thaumarchaeal
418 SSBs and trimeric RPAs from Asgard archaea, the postulated ancestors of eukaryotes (Liu et
419 al., 2021; Spang et al., 2015). The structures formed three major branches in the structural
420 distance-based tree. The first branch included all bacterial SSBs. In the second clade,
421 thaumarchaeal, crenarchaeal and eukaryotic SSBs grouped with the archaeal and eukaryotic
422 OB-1 domain of Rpa1 (OB-A in eukaryotes), with OB-2 of Rpa1 (OB-C in eukaryotes) forming
423 a sister group to this assemblage. Finally, the third clade included two sister clades
424 corresponding to OB-3 and OB-4 domains of archaeal and eukaryotic Rpa2 and Rpa3,
425 respectively. It was previously noted that it is difficult to distinguish single-OB domain SSBs
426 and Rpa3 (Taib et al., 2021). Structural comparisons clearly distinguish the two sets of proteins
427 and strongly suggest an evolutionary relationship between the archaeal SSBs and the N-
428 terminal OB domain of archaeal and eukaryotic Rpa1 subunits, rather than with Rpa3. These
429 observations are consistent with the evolution of heterotrimeric RPA from a single-OB SSB
430 ancestor, likely within the archaeal branch (Gamsjaeger et al., 2015; Kerr, 2003; Morten et al.,
431 2015). Given the shared non-structured C-terminal extension in bacterial and archaeal SSBs
432 and the ability to oligomerize, it is likely that they have evolved from a common ancestor, which
433 might have functioned already in the last universal cellular ancestor (LUCA). Within archaea,
434 this SSB gave rise to the multidomain Rpa1 through tandem duplication followed by insertion

435 of a Zn-finger subdomain into OB-2 (**Figure 5b**). By contrast, Rpa2 and Rpa3 appear to be
436 structural paralogs, with Rpa2 accruing a characteristic C-terminal WH domain. Interestingly,
437 the PabRPA OB-3 and OB-4 domains respectively share the highest structural similarity with
438 the OB-domains from the Ctc1 and Ten1 subunits of the human telomeric maintenance CST
439 complex. More generally, the OB domains from eukaryotic RPA and RPA-like complexes,
440 such as the CST, grouped with the corresponding domains of archaeal PabRPA-like homologs,
441 including those from Asgard archaea, suggesting that the eukaryotic RPA and its derivatives
442 have been inherited from the archaeal ancestors.

443 RPA, which evolved from its SSB ancestor by accretion of complexity, has acquired the
444 ability to accomplish a broad range of biological functions in DNA replication, recombination
445 and repair (Oliveira, 2021). In addition, this increased level of complexity probably also
446 enabled more sophisticated mechanisms of modulation and regulation of RPA. Indeed, RPA is
447 a protein platform that recruits many different key actors involved in genome maintenance
448 (Brosey et al., 2013). The strong structural similarity observed between PabRPA and the human
449 telomeric maintenance CST complex is also intriguing. Indeed, telomer maintenance in linear
450 genomes are defining features of eukaryotes compared to prokaryotes, which possess circular
451 genomes. Therefore, the inherited ancestral RPA may have provided the primitive eukaryotes
452 with key genome maintenance activities that were required for the emergence of the eukaryotic
453 domain of life.

454 **MATERIEL & METHODS**

455 **Cloning, protein expression, and purification**

456 The open reading frames (ORFs) of the Rpa1, Rpa2 and Rpa3 genes from *P. abyssi* were
457 optimized and synthesized by GeneArt (Thermo Fisher). For individual subunit expressions,
458 ORFs were inserted into the pRSFduet(+) multiple cloning site 1 with a TEV-cleavable N-
459 terminal 14-His tag. For PabRPA complexes co-expressions, ORFs were inserted into the
460 pRSFduet(+) multiple cloning site 1 as a polycistronic Rpa3-Rpa2-Rpa1 construct separated by
461 ribosome binding sites (RBS), with a TEV-cleavable N-terminal His14-tagged Rpa3 fusion
462 protein. The folowing RPA isoforms were cloned from the pRSFduet(+) constructs using the
463 Q5 site directed mutagenesis kit (New England Biolabs): PabRpa1 cooperativity module
464 ($Rpa1_{1-193}$), PabRPA trimerization core Tri-C ($Rpa1_{193-358}$/$Rpa2_{1-179}$/Rpa3), PabRPA ΔARCD
465 ($Rpa1_{64-358}$/Rpa2/Rpa3), PabRPA ΔARCD-OB-1 ($Rpa1_{193-358}$/Rpa2/Rpa3), PabRPA ΔWH
466 (Rpa1/$Rpa2_{1-179}$/Rpa3).

467 Proteins were expressed in BL21 Star (DE3) strain from *E. coli* (Invitrogen) at 37°C in
468 LB medium supplemented with 100 µg/mL of kanamycin. Recombinant protein expression was
469 induced by adding 0.25 mM IPTG. Cells were then incubated overnight at 20°C, collected by
470 centrifugation, resuspended in buffer A (0.02 M Na-HEPES at pH 8, 0.5 M NaCl, 0.02 M
471 imidazole) supplemented with complete EDTA-free protease inhibitors (Roche), and lysed with
472 a Cell-Disruptor. Lysates were then heated for 10 min at 60°C and centrifuged 30 min at 20000
473 g. PabRPA purifications were performed using a three-step protocol including nickel affinity,

474 anion exchange and size exclusion chromatography. The clear cell lysate was loaded onto 5 mL
475 HisTrap columns (Cytiva) connected to an ÄKTA purifier (Cytiva). Elutions were performed
476 using a linear gradient of imidazole (buffer B, 0.02 M Na-HEPES at pH 8, 0.5 M NaCl, 0.5 M
477 imidazole). Protein fractions were combined, dialyzed in buffer C (0.02 M Na-HEPES pH 8,
478 0.1 M NaCl), loaded onto 5 ml HiTrap Q FF columns (Cytiva) and eluted with a linear gradient,
479 by mixing buffer C with buffer D (0.02 M Na-HEPES pH 8, 2 M NaCl). Depending on the
480 applications, the 14-His tag was removed following an overnight TEV-protease cleavage.
481 Purifications were finally polished using exclusion-size chromatography in buffer E (0.02 M
482 Na-Hepes pH 8, 0,15 M NaCl) on a Superdex 200 10/300 (Cytiva).

483 **Crystallization**

484 Crystallization trials were performed using the hanging drop vapor diffusion technique
485 in 2 µL drops (1:1 reservoir to protein ratio) equilibrated against 500 µL of reservoir solution.
486 PabRPA crystals were obtained at 18°C in 0.1 M Tris pH 8.5, 1.2 M ammonium sulfate with a
487 protein solution at 10 mg/mL. After 72 hours, crystals were transferred in a dehydration solution
488 containing the original mother liquor supplemented with 25% glycerol and incubated for 1 hour
489 at 4 °C before flash-freezing in liquid nitrogen. PabRpa1 cooperativity module crystals were
490 obtained at 18°C using a crystallization solution made of 0.01 M NiCl2, 20% w/v PEG MME
491 2K and 0.1 M Tris pH 8.5 with a protein solution at 40 mg/mL. Crystals were then directly
492 flash-frozen in liquid nitrogen without additional cryoprotection. The PabRPA Tri-C/dT20
493 complex was reconstituted by mixing the protein solution at 0.5 mg/mL with a 1.2X excess of
494 poly-dT20 ssDNA (Eurogentec) in buffer E. The mixture was then concentrated ~30 times to
495 obtain a final $OD_{280nm}$ value of 40. Crystals were grown at 4°C in 10% w/v PEG 8K, 0.1 M
496 imidazole pH 8 and 0.2 M calcium acetate, and cryoprotected with 25% ethylene glycol. X-ray
497 data were collected at the SOLEIL synchrotron on beamlines PX1 and PX2.

498 **X-ray data collection, processing, model building and refinement**

499 Diffraction data collection and refinement statistics are given in **Supplementary table**
500 **S1**. Crystallographic data were collected on the PROXIMA-1 (Chavas et al., 2021) beamlines
501 at Synchrotron SOLEIL (Saint-Aubin, France) and processed with XDS (Kabsch, 2010)
502 through XDSME (https://github.com/legrandp/xdsme) (Legrand, 2017). The strong diffraction
503 anisotropy was corrected using the STARANISO program (Tickle et al., 2018). Initial phases
504 for the apo-PabRPA crystal structure were obtained by Zn-SAD. The diffraction data from 3
505 crystals collected at the peak absorption wavelength of zinc K-edge (1.2827 Å) were merged to
506 obtain a highly redundant dataset (>60). A unique zinc site was found using the SHELXC/D
507 (Sheldrick, 2008) from which phases were calculated with PHASER (McCoy et al., 2007) and
508 then improved by density modification with the PARROT program (Cowtan, 2010). Thus, a
509 first experimental electron density map at 4.5 Å resolution could be obtained in which an initial
510 model could be assembled and completed using COOT (Emsley et al., 2010) combining the N-
511 terminal region of Rpa1 previously determined and other partial models like the zinc finger.
512 Later, this initial model could be improved with the help of AlphaFold (Jumper et al., 2021)
513 predictions, and refined using a dataset reaching 3.7 Å in the best diffracting direction. The

514 procedure proposed by Terwilliger (Terwilliger et al., 2022) was exploited to iteratively
515 improved AlphaFold models using experimental information. Crystal structures of the PabRpa1
516 cooperativity module and the PabRPA Tri-C/d20T complex were determined by molecular
517 replacement with MOLREP (Vagin and Teplyakov, 1997) using the previously determined apo
518 structure of PabRPA. All refinements were conducted with the BUSTER program (Bricogne,
519 G. et al., 2017) using TLS motion groups and COOT was used for model reconstruction. In the
520 case of PabRPA-apo and PabRPA Tri-C/d20T, local structure similarity restraints (LSSR) were
521 used using hybrid target models constructed from a mix of higher resolution structures from the
522 present work and AlphaFold predicted models (Jumper et al., 2021; Terwilliger et al., 2022).
523 Details for all datasets are summarized in **Supplementary table S1**.

524 **AlphaFold model predictions**

525 AlphaFold model predictions were calculated either using the Google Colab platform
526 and AlphaFold2_advanced form developed by the ColabFold team (Mirdita et al., 2021)
527 (https://colab.research.google.com/github/sokrypton/ColabFold) or using a local installation of
528 ColabFold obtained from the LocalColabFold Github repository
529 (https://github.com/YoshitakaMo/localcolabfold).
530

531 **Cryo-EM sample preparation**

532 The PabRPA/dT35 complex was reconstituted by mixing a poly-dT35 ssDNA
533 (Eurogentec) with a 3-fold excess of PabRPA. The complex was then injected on a superdex
534 200 10/300 in buffer F (0.02 M HEPES pH 8, 0.1 M NaCl and 0.002 M magnesium acetate). 3
535 µL of the eluted fraction ($OD_{280nm}$=1.5) were applied to glow-discharged quantifoil R2/2 300
536 gold mesh grids (Electron Microscopy Sciences). After 20 seconds of pre-incubation, the grids
537 were blotted for 6 seconds (force -2) using a Vitrobot Mark IV (ThermoFischer) at 100%
538 humidity and 25 °C. The PabRPA/d100T complex was reconstituted using the same strategy
539 but using a 5-fold excess of PabRPA. Specimens were prepared similarly with quantifoil R2/2
540 200 copper mesh grids (Electron Microscopy Sciences). For the PabRPA tetrameric
541 supercomplex, 3 µL of purified PabRPA at 0.5 mg/mL were applied to glow discharged
542 Quantifoil R2/2 300 gold mesh grids. The grids were then blotted for 6 seconds (force 0).

543 **Cryo-EM data acquisition and image processing**

544 Movie acquisition was carried out at the Nanoimaging Core Facility (Institut Pasteur,
545 Paris). Details for all datasets are summarized in table 2. The PabRPA/dT35 dataset was
546 collected on a 200 kV Glacios electron microscope (Thermo Fisher) equipped with a Falcon 3
547 detector operating in electron counting mode. The PabRPA/dT100 dataset was collected in the
548 same microscope, operating in integration mode. The PabRPA tetrameric supercomplex dataset
549 was collected on a 300 kV Titan Krios electron microscope (Thermo Fisher) equipped with a
550 K3 detector and a bio-quantum energy filter (Gatan).

551 Motion correction and CTF-estimation of the acquired movies were carried out in
552 Cryosparc v3.3.2 (Punjani et al., 2017). Image processing pipelines for each map are shown in

553 **Supplementary figures S4, S5 & S7**. Blob picking was used to generate initial picks of
554 PabRPA/dT35, which were then filtered by 2D classification and used to train a Topaz picking
555 model (Bepler et al., 2019). Topaz particle picks were used to generate three initial models,
556 which were refined with heterogeneous refinement jobs, including a volume encompassing only
557 the trimeric core. The highest resolution map was refined using non-uniform refinement
558 yielding a 6.24 Å resolution map, and then locally filtered with a lanczos filter. Initial templates
559 of PabRPA/dT100 were generated with blob picking, and then used for template picking.
560 Multiple initial models were generated, and the one that showed the interaction between two
561 RPA complexes was chosen and refined through heterogeneous and non-uniform refinement,
562 achieving 8.24 Å resolution. Initial 2D templates of the PabRPA tetrameric supercomplexes
563 datasets were generated via blob-picking and used to train a Topaz model. Topaz picks were
564 then used to generate three initial models, which were refined using heterogeneous refinement.
565 3D classification yielded a map containing the cooperativity module, which was refined using
566 non-uniform refinement and locally filtered with a lanczos filter, achieving 3.8 Å resolution. A
567 high-resolution map that did not include the cooperativity module was refined with non-
568 uniform refinement and C2 symmetry to achieve 3.35 Å resolution. Details for all datasets are
569 summarized in **Supplementary table S2**.

570 **Building and refinement of cryo-EM models**

571 Procedure for the PabRPA/d35T cryo-EM structure: individual crystal structures of
572 PabRPA Tri-C/d20T (PDBid: 8AAS) and PabRpa1 OB-1 (PDBid: 8AA9) were manually
573 placed in the cryo-EM map and subsequently rigid-body fitted in the density with coot (Emsley
574 et al., 2010; Emsley and Cowtan, 2004). Six different rigid-body groups were used:
575 the trimerization bundle, the four individual OB domains and the Rpa1 zinc-finger domain
576 (**Supplementary figure S6f**). ssDNA modeling was guided by superimposing the PabRRPA
577 Tri-C/d20T (PDBid:8AAS) and the human Rpa1 OB-A/ssDNA (PDBid:1JMC) crystal
578 structures. This initial model was further real-space refined by using Phenix (Adams et al.,
579 2011), with a high weight on ideal geometry and restraints on secondary structures. The final
580 map correlation coefficient is 0.63. Procedure for the PabRPA/d100T cryo-EM structure: two
581 identical models of Tri-C(n)/OB-1(n+1) complex predicted by AlphaFold (Evans et al., 2021)
582 were manually placed and rigid-body fitted in the cryo-EM map using coot (**Supplementary**
583 **figure S7f**). Modelling of the ssDNA substrate was guided by superimposing the PabRPA/d35T
584 cryo-EM structure (PDBid: 8ACE). Finally, the complete model was real-space refined using
585 Phenix, following the same procedure as for the PabRPA/d35T structure (See above). The final
586 map correlation coefficient is 0.64. Procedure for the PabRPA tetrameric supercomplex cryo-
587 EM structure: Four identical models of Pab-RPA Tri-C complex (PDBid: 8AAS) were
588 manually placed and rigid-body fitted in the cryo-EM map using coot. The initial model was
589 then subjected to global real-space refinement program from Phenix using secondary structure
590 restraints. The refined model was further manually inspected and adjusted in coot. Procedure
591 for the PabRPA tetrameric supercomplex cryo-EM structure with the cooperativity module
592 connecting two PabRPA molecules: The model was built by using the same procedure than for
593 the PabRPA tetrameric complex (see above). In addition, Rpa1 ARCD and OB-1 domains
594 crystal structures (PDBid: 8AA9) were manually placed and rigid-body fitted using coot and
595 Phenix. Refinement details for all datasets are summarized in **Supplementary table S2**.

596 **Structure analysis**

597 Electrostatic calculations were performed by using APBS (Jurrus et al., 2018).
598 Interaction surfaces were calculated by using the PISA webserver (Krissinel and Henrick,
599 2007). Surface conservations were calculated using CONSURF (Ashkenazy et al., 2016).
600 Structure comparisons were performed by using DALI (Holm, 2020). All figures were prepared
601 using UCSF Chimera X (Pettersen et al., 2021).

602 **Negative staining microscopy**

603 The PabRPA/pM13 complex was reconstituted by mixing 1 µL of M13mp18 ssDNA
604 (New England Biolabs) at 250 µg/mL with 1 µL of PabRPA at 30 µM (200X excess) in 98 µL
605 of buffer F. 3 µL of the mixture were deposited on glow-discharged carbon-coated copper grids
606 CF400-CU (Electron Microscopy Sciences) and contrasted 3x1 minute in 2% uranyl acetate.
607 Data collection was performed using a Tecnai biotwin T12 (Thermo Fisher) equipped with a
608 LaB6 filament, operating at 120 keV. Images were recorded using an Eagle camera (Thermo
609 Fisher) at a nominal magnification of 49000, using a 3 µm defocus.

610 **Electromobility shift assay**

611 2.5 pmol of 3'-FAM ssDNA-24mer (5'-GCCTGCAGGTCGACTCTAGAGGAT-3') or
612 dsDNA-24mer were mixed with increasing amounts of *Pab*RPA in Tris-HCl 30 mM pH 7.5,
613 NaCl 300 mM, 0.5 g/L bovine serum albumin (BSA), 5% Ficoll (20 µL final volume).
614 Reactions were carried out to reach equilibrium at 60°C for 1 h before migration in 1% agarose
615 gel at 4°C under native conditions (90 V, 30 mA). Images were acquired with Typhoon9400
616 (GE Healthcare).

617 **Surface plasmon resonance (SPR)**

618 Surface plasmon resonance data have been acquired with Reichert SR7000DC
619 spectrometer equipped with a self-sampling injection system Reichert SR7100. Sensor chips
620 consisted of a glass slide coated by a gold film, the latter being already functionalized by a
621 mixed self-assembled monolayer (SAM) of 90% dithiol aromatic PEG6-COOH and 10% of
622 dithiol aromatic PEG3-COOH. Neutravidin was immobilized onto SAM, in both flow cells,
623 using classical EDC:NHS coupling chemistry. Appropriate amounts of DNA were immobilized
624 onto surfaces by successive injections at 25 µL/min of 1 µM of 5'-biotin-TEG labelled 32-mer
625 ssDNA (5'-TGCCAAGCTTGCATGCCTGCAGGTCGACTCTA-3'), ssRNA (5'-
626 UGCCAAGCUUGCAUGCCUGCAGGUCGACUCUA-3') or dsDNA (Eurogentec) diluted in
627 buffer E supplemented with 0.05% tween. Binding assays were performed at 25°C by injecting
628 increasing concentrations of protein solutions for 300 s at 25 µL/min. The raw sensograms were
629 processed by double referencing i.e. subtracting both the signals measured on the reference
630 flow cell and the signals measured for blank injections. Resulting sensorgrams were analyzed
631 by Scrubber2.0 (BiolLogic) following several Monte Carlo fits without mass transport
632 limitation of single equilibrium system for PabRPA, Rpa2, Rpa3 and Rpa1/Rpa2 and two-steps
633 equilibrium system for Rpa1.

634 **Bio-layer interferometry (BLI)**

635 ssDNA-protein binding assays were performed on an Octet RED384 BLI instrument
636 (ForteBio). 3'-biotin-TEG labelled poly-dT35 ssDNA were captured at a concentration of 5
637 µg/mL for 100 s on streptavidin biosensors. Binding experiments were monitored at 25°C in
638 buffer E supplemented with 0.2 mg/mL BSA (Sigma). Data were analyzed with Prism 9.4. A
639 sample reference with buffer-only was subtracted from all curves. Affinities were determined
640 by fitting the concentration dependence of the experimental steady-state signals, using the
641 following equation Req = Rmax*[RPA]/(Kd + [RPA]).

642 **Microscale thermophoresis (MST)**

643 The MST measurements were performed at room temperature on Monolith NT.115
644 Blue/Red (NanoTemper Technologies, GmbH) instrument using standard capillaries. For all
645 experiments, the LED power and IR laser configuration were set to 20% and default,
646 respectively. The MST data were either analyzed with NanoTemper Analysis v2.3, Prism 9.4
647 or Matlab 2016b software. In all cases, the data were fitted using the Hill equation without any
648 constrains. Each experiment was performed at least three times in buffer E. 5'-Cy5 fluorophore
649 labelled poly-dT100 ssDNA (Integrated DNA technologies) was diluted to a concentration of
650 200nM and mixed with an equal volume of a serial dilution series of WT or mutants PabRPA
651 and incubated over night at 20°C before loading into MST capillaries.

652 **Small-Angle X-Ray Scattering (SAXS)**

653 SAXS data of the PabRPA ΔWH complex (50 µl at 1 mg/ml) in buffer E was collected
654 in batch Mode at beamline SWING (synchrotron SOLEIL). The curves were background-
655 subtracted using FOXTROT and analyzed using the ATSAS 3.0.2 software suite. The
656 normalized Kratky was calculated according to (Durand et al., 2010; Pérez et al., 2001) using
657 the value determined by Guinier analysis. P(r) functions were computed from the scattering
658 curves by an indirect transform method in GNOM (Svergun, 1992).

659 **Molecular mass measurements by size exclusion chromatography coupled to static light**
660 **scattering detection (SEC-SLS) and viscometry**

661 The oligomerization states of PabRPA ΔWH and PabRPA Tri-C at 20 µM in presence
662 or absence of 24 µM poly-dT25 ssDNA (Eurogentec) were determined by size exclusion
663 chromatography coupled to a triple detection (concentration detector: UV detector,
664 refractometer; static light scattering 7°, 90°; viscometer) on a Omnisec resolve/reveal
665 instrument (Malvern Panalytical). The column and detectors were equilibrated with the filtered
666 and degazed running buffer E prior measurement. All proteins were injected (100 µl) and eluted
667 at 0.4 ml/min on a Superdex 200 increase 10/300 GL column (Cytiva). Detections were
668 performed at 20°C or 65°C. External calibration was done with BSA using an injection of 100
669 µl at 1.4 mg/ml. The refractive index, static light scattering, and the viscosity measurements
670 were processed to determine the mass average molecular mass and the intrinsic viscosity using
671 the OMNISEC V11.10 software (Malvern Panalytical).

672 **Data avaibility**

673 The PabRpa1 cooperativity module, PabRPA and PabRPA Tri-C/d20T crystal
674 structures were deposited in the Protein Data Bank under accession codes, 8AA9, 8AAJ and
675 8AAS respectively. The all-atom cryo-EM structure of PabRPA tetramer, as well as the 3D-
676 class showing the cooperativity module are deposited under accession code (PDBid: 8AAH;
677 EMDB-15300) and (PDBid: 8AAL; EMDB-15301) respectively. Finally, C-alpha backbone,
678 nucleic acid phosphate trace and the zinc metal ions atomic coordinates of the PabRPA/d35T
679 and PabRPA/d100T complexes are deposited under the accession codes (PDBid: 8ACE;
680 EMDB-15347) and (PDBid: 8ACJ; EMDB-15350), respectively.

681 **Acknowledgments**

682

683 We thank the Platforms for X-ray crystallography (PF6), biophysics, and nanoimaging
684 in the Institut Pasteur. We acknowledge SOLEIL for provision of synchrotron radiation
685 facilities and the staff of beamline PROXIMA-1 (Saint-Aubin, France). We gratefully
686 acknowledge the financial support of ANR (Grants ARCHAPOL and ARCHAPRIM). C.M.
687 was funded by a postdoctoral Pasteur-Roux-Cantarini fellowship. M.M.C was funded by a
688 postdoctoral FRM fellowship (Fondation pour la Recherche Médicale). We would like to thank
689 Sebastien Brûlé (Molecular Biophysics Platform, C2RT, Institut Pasteur) for helping us with
690 SEC-SLS, and Olympe Vaillant for preparing the RPA topology diagram. We also would like
691 to thank Dr. Christophe Crézé for performing the EMSA experiments.

692 **FIGURE LEGENDS**

693 **Figure 1: Structure of the archaeal RPA. a.** Domain diagrams of PabRPA and human RPA.
694 **b.** Two orthogonal views of the full-length PabRPA crystal structure at 3.7 Å with a focused
695 view on the trimeric helical bundle (top right). Crystal lattice contacts within neighboring RPA
696 heterotrimers are shown in grey. **c**. High resolution crystal structure of the Rpa1 cooperativity
697 module at 1.8 Å (top) with a focused view on the newly identified archaeal-specific ARCD
698 domain, which is colored according to sequence conservation. Side chains of residues
699 connecting ARCD to OB-4 are shown as ball-and-sticks. Phylogenetic analysis (bottom)
700 reveals a broad distribution of Rpa1-ARCD domain homologs across Archaea. Genomes
701 lacking ARCD domain are presented in grey.

702 **Figure 2: Structural basis for the DNA-binding activity by PabRPA. a.** DNA-binding
703 properties of PabRPA trimerization core and cooperativity module. Specific binding of
704 PabRPA constructs to immobilized poly-dT35 ssDNA was measured by biolayer
705 interferometry. Steady-state analysis were performed using the average signal measured at the
706 end of the association steps. **b.** 3.2 Å crystal structure of the poly-dT20-bound Tri-C, with two
707 focused views on Rpa1 OB-2 and Rpa2 OB-3 ssDNA-interacting residues. **c.** Cryo-EM
708 structure of full-length PabRPA bound to a poly-dT35 ssDNA. Backbone traces of the poly-
709 dT20-bound Tri-C and the Rpa1 OB-1 crystal structures were rigid-body fitted into the cryo-
710 EM map contoured at a level of 1rmsd. Conformational changes of OB-1, observed between

711 the apo-PabRPA crystal structure (in grey), and the DNA-bound cryo-EM structure (in orange).
712 Electrostatic potential surface analysis showing that the repositioning of Rpa1 OB-1 domain
713 extends the ssDNA-binding channel.

714 **Figure 3: Structural basis for the assembly of multiple PabRPA molecules on ssDNA. a.**
715 Negative staining microscopy of PabRPA bound to M13 ssDNA plasmid. **b.** Cryo-EM 2D
716 classes of PabRPA bound to a poly-dT100. **c.** Cryo-EM structure of PabRPA bound to a poly-
717 dT100 displayed as backbone traces with a focused view on the Tri-C(n)/OB-1(n+1) AlphaFold
718 model used for model building. The cryo-EM map is contoured at a level of 6rmsd **d.**
719 Biophysical assessment of RPA full-length and cooperativity-module truncated constructs
720 binding onto long ssDNA. Microscale thermophoresis binding curves of RPA heterotrimeric
721 complexes interacting with ssDNA, using wild-type PabRPA (blue), ΔARCD mutant (red) or
722 ΔARCD-OB-1 mutant and a Cy5-labeled poly-dT100 ssDNA substrate. $EC_{50}$ values were
723 calculated using the Hill-equation. Hill coefficients were calculated from the Hill-plot
724 transform of the binding data. **e.** Electrostatic potential surface analysis showing that PabRPA
725 forms a continuous positive groove along the DNA path (orange).

726 **Figure 4: PabRPA assembles as a tetrameric super-complex that dissociates upon binding**
727 **to ssDNA. a.** SEC-SLS characterization of PabRPA and PabRPA Tri-C in presence or absence
728 of poly-d25T ssDNA. The theoretical ($MW_{th}$) and calculated molecular weights ($MW_{calc}$) for
729 each complex are given in kDa. **b.** 3.35 Å cryo-EM structure of the tetrameric PabRPA super-
730 structure. **c.** Focused view on critical contacts within the PabRPA tetrameric assembly. **d.** 3D
731 class showing contacts between the cooperativity module and Rpa3 of two neighboring
732 PabRPA molecules within the tetrameric super-structure. **e.** Proposed mechanism for the self-
733 association of PabRPA on ssDNA.

734 **Figure 5: Evolutionary implications for the RPA evolution in Archaea and Eukarya. a.**
735 Structure-based dendrogram produced by DALI, based on average linkage clustering of the Z-
736 scores of the structural similarity matrix. **b**. Current proposal for the evolutionary history of
737 SSB and RPA in the three domains of life.

738 **SUPPLEMENTARY FIGURE AND TABLE LEGENDS**

739 **Figure S1: Electron density in representative regions of the PabRPA crystal structures.**
740 PabRPA (**a**), PabRPA Rpa1 cooperativity module (**b**) and PabRPA Tri-C/20dT (**c**) crystal
741 structures are represented as Cα traces in the left panels. Right panels show close views of
742 functionally important regions of PabRPA. Main chain and side chains atoms are represented
743 as Cα traces and sticks, respectively. The gray mesh is the maximum likelihood 2mFo-DFc
744 electron density map contoured at a level of 1 σ.

745 **Figure S2: Topological diagram of PabRPA.** Two-dimensional diagram of protein topology
746 for three subunits of PabRPA.

747 **Figure S3: Comparison of genomic loci encoding Rpa1 genes in Archaea.** Genomic
748 organizations of the loci encoding Rpa1 genes are shown for representative species from each

749 major archaeal phyla. Each locus is indicated with the UniProt accession number of the Rpa1
750 gene, followed by the species name. All loci are shown as -7 and +7 genes around Rpa1 gene.
751 Rpa1 genes are embedded within highly variable genomic neighborhoods, which typically vary
752 even for relatively closely related organisms. Presence of the Archaea-specific ARCD is
753 indicated with white asterisks. Strikingly, ARCD is present in all Rpa1 sequences, except for
754 the 5 *Thermoplasmatales* members. It is noteworthy that *Methanomassiliicoccales* within
755 *Thermoplasmatota* do contain ARCD, suggesting that ARCD was lost within
756 *Thermoplasmatales* following their divergence from a common ancestor with
757 *Methanomassiliicoccales*. Interestingly, *Aciduliprofundum boonei* which is even closer to the
758 *Thermoplasmatales* also contain ARCD. DNA metabolism proteins encoding genes are
759 represented in color.

760 **Figure S4: Nucleic-acids binding properties of PabRPA. a.** Electrophoretic mobility shift
761 assay of ssDNA-24mer and dsDNA-24mer (2.5 pmol) in presence of decreasing concentrations
762 of PabRPA (312, 62, 12, 2.5, 0.5, 0.1 pmol). **b.** PabRPA binding specificity on nucleic acids.
763 Specific binding of PabRPA heterotrimer on immobilized ssDNA-32mer (left), dsDNA-32mer
764 (middle) and ssRNA-32mer measured by surface plasmon resonance (RU: resonance units). **c.**
765 Role of PabRPA subunits on *ss*DNA binding. Specific binding of Rpa1, Rpa2, Rpa3, or
766 Rpa1/Rpa2 complex on immobilized ssDNA-32mer measured by surface plasmon resonance.

767 **Figure S5: Comparison of archaeal and eukaryotic RPA ssDNA binding determinants. a.**
768 Comparison of DNA-bound *P. abyssi* and *U. maydis* (4GOP) Tri-C structures with focused
769 views on Rpa1 (purple and blue) and Rpa2 (green). ssDNA-contacting residues are shown as
770 spheres. **b.** Superimposition of the *P. abyssi* Rpa1 OB-1 (in purple) with the *H. sapiens* Rpa1
771 OB-A (in blue) (1JMC) crystal structures reveals conserved ssDNA-binding amino-acids.

772 **Figure S6: Cryo-EM structure determination of PabRPA bound to a short ssDNA. a.** Size-
773 exclusion chromatography (SEC) profile (left) and SDS-PAGE analysis (right) of PabRPA
774 bound to a poly-dT35 ssDNA **b.** Data processing workflow (detailed in methods). **c.** Resolution
775 of final reconstructions determined by gold standard FSC at the 0.143 criterion. **d.** Local
776 resolution evaluations of the final reconstructions (top) and Euler angle distributions of particles
777 used for the final reconstructions (bottom). **e.** Cryo-EM maps at varying contour levels. **f.**
778 Comparison of the DNA-bound RPA cryo-EM and crystal structure (left) and cryo-EM map
779 colored by RPA subunits (right).

780 **Figure S7: Cryo-EM structure determination of PabRPA bound to a long ssDNA. a.** Size-
781 exclusion chromatography (SEC) profile (left) and SDS-PAGE analysis (right) of PabRPA
782 bound to a poly-dT100 ssDNA **b.** Data processing workflow (detailed in methods). **c.**
783 Resolution of final reconstructions determined by gold standard FSC at the 0.143 criterion. **d.**
784 Local resolution evaluations of the final reconstructions (left) and Euler angle distributions of
785 particles used for the final reconstructions (right). **e.** Cryo-EM maps at varying contour levels
786 **f.** Model building workflow.

**Figure S8: ARCD is required for early-steps of the ssDNA-RPA complex assembly and becomes flexible after binding to ssDNA. a.** Left: superposition of the five AlphaFold-Multimer predicted models (Evans et al., 2021) for the interaction of one PabRPA-ΔWH heterotrimer with two first domains of PabRPA1 (ARCD-OB1, 1-180). Right: AlphaFold generated plots of MSA sequence coverage (top), of predicted LDDT vs position along the protein sequence (middle) and of predicted aligned error (bottom, where each protein chain of the complexe is plotted against itself and against the other chain). **b.** On left, fit of the RPA(n)/ARCD-OB-1(n+1) AlphaFold model containing the ARCD domain in the cryo-EM map. The ARCD domain is not observed either in the map or in the 2D classes. On right, superimposition with the apo-PabRPA crystal structure. The lattice contact between ARCD and OB-4 fits with the AlphaFold prediction. **c.** Specific binding of PabRPA (in purple) and ΔARCD mutant (in pink) to immobilized poly-dT35 ssDNA measured by biolayer interferometry. Steady-state analysis were performed using the average signal measured at the end of the association steps.

**Figure S9: PabRPA oligomerization state in solution. a.** In solution study of PabRPA structure by SAXS. On top left, SAXS profile of PabRPA (in purple) with superimposition of fitted scattering pattern that allowed the generation of the distance distribution function P(r) (red dotted line). On right, distance distribution functions P(r) derived from the scattering patterns of PabRPA obtained using the program GNOM. On bottom left, dimensionless Kratky plot scattering. Scattering derived parameters are provided in the table. **b.** PabRPA molecular mass (Green) and intrinsic viscosity (orange) measurements by SEC-SLS coupled with an inline viscometer. Measurements were performed at 20°C (higher panel) and 65°C (lower panel).

**Figure S10: Cryo-EM structure determination of the tetrameric RPA super-structure. a.** Size-exclusion chromatography (SEC) profile (left) and SDS-PAGE analysis (right) of PabRPA **b.** Data processing workflow (detailed in methods). **c.** Resolution of final reconstructions determined by gold standard FSC at the 0.143 criterion. **d.** Local resolution evaluations of the final reconstructions (top) and Euler angle distributions of particles used for the final reconstructions (bottom). **e.** Representative views of the cryo-EM map showing secondary structure elements and key regions of the tetrameric RPA assembly.

**Table S1: X-ray diffraction data collection and refinement statistics**. Values in parentheses are for highest-resolution shell. Dataset from single crystal used per structure. *Values calculated after truncation by STARANISO. Estimated resolution limits along the three crystallographic directions *a**, *b**, *c**.

**Table S2: Cryo-EM data collection, refinement and validation statistics**

**Movie S1: Architecture of PabRPA**

**Movie S2: Structural and functional characterization of PabRPA in complex with ssDNA of different lengths**

825 **Movie S3: PabRPA forms tetrameric complexes that dissociates upon binding to ssDNA**

826 **REFERENCES**


827 Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Echols, N., Headd, J.J., Hung, L.-W.,
828 Jain, S., Kapral, G.J., Grosse Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner,
829 R.D., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H.,
830 2011. The Phenix software for automated determination of macromolecular structures.
831 Methods 55, 94–106. https://doi.org/10.1016/j.ymeth.2011.07.005
832 Arunkumar, A.I., Stauffer, M.E., Bochkareva, E., Bochkarev, A., Chazin, W.J., 2003.
833 Independent and Coordinated Functions of Replication Protein A Tandem High
834 Affinity Single-stranded DNA Binding Domains. Journal of Biological Chemistry
835 278, 41077–41082. https://doi.org/10.1074/jbc.M305871200
836 Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016.
837 ConSurf 2016: an improved methodology to estimate and visualize evolutionary
838 conservation in macromolecules. Nucleic Acids Res 44, W344–W350.
839 https://doi.org/10.1093/nar/gkw408
840 Bain, F.E., Fischer, L.A., Chen, R., Wold, M.S., 2018. Single-Molecule Analysis of
841 Replication Protein A–DNA Interactions, in: Methods in Enzymology. Elsevier, pp.
842 439–461. https://doi.org/10.1016/bs.mie.2017.11.016
843 Barry, E.R., Bell, S.D., 2006. DNA Replication in the Archaea. Microbiology and Molecular
844 Biology Reviews 70, 876–887. https://doi.org/10.1128/MMBR.00029-06
845 Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A.J., Berger, B., 2019.
846 Positive-unlabeled convolutional neural networks for particle picking in cryo-electron
847 micrographs. Nature Methods 16, 1153–1160. https://doi.org/10.1038/s41592-019-
848 0575-8
849 Bianco, P.R., 2017. The tale of SSB. Progress in Biophysics and Molecular Biology 127,
850 111–118. https://doi.org/10.1016/j.pbiomolbio.2016.11.001
851 Blackwell, L.J., Borowiec, J.A., 1994. Human replication protein A binds single-stranded
852 DNA in two distinct complexes. Mol Cell Biol 14, 3993–4001.
853 https://doi.org/10.1128/mcb.14.6.3993-4001.1994
854 Bochkarev, A., Pfuetzner, R.A., Edwards, A.M., Frappier, L., 1997. Structure of the single-
855 stranded-DNA-binding domain of replication protein A bound to DNA. Nature 385,
856 176–181. https://doi.org/10.1038/385176a0
857 Bochkareva, E., 2002. Structure of the RPA trimerization core and its role in the multistep
858 DNA-binding mechanism of RPA. The EMBO Journal 21, 1855–1863.
859 https://doi.org/10.1093/emboj/21.7.1855
860 Bochkareva, E., Frappier, L., Edwards, A.M., Bochkarev, A., 1998. The RPA32 Subunit of
861 Human Replication Protein A Contains a Single-stranded DNA-binding Domain.
862 Journal of Biological Chemistry 273, 3932–3936.
863 https://doi.org/10.1074/jbc.273.7.3932
864 Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P.,
865 Sharff, A., Smart, O.S., Vonrhein, C., Womack, T.O., 2017. BUSTER version 2.10.3.
866 Cambridge, United Kingdom: Global Phasing Ltd.
867 Brill, S.J., Bastin-Shanower, S., 1998. Identification and Characterization of the Fourth
868 Single-Stranded-DNA Binding Domain of Replication Protein A. Mol Cell Biol 18,
869 7225–7234. https://doi.org/10.1128/MCB.18.12.7225
870 Brosey, C.A., Chagot, M.-E., Ehrhardt, M., Pretto, D.I., Weiner, B.E., Chazin, W.J., 2009.
871 NMR Analysis of the Architecture and Functional Remodeling of a Modular

872 Multidomain Protein, RPA. J. Am. Chem. Soc. 131, 6346–6347.
873 https://doi.org/10.1021/ja9013634
874 Brosey, C.A., Soss, S.E., Brooks, S., Yan, C., Ivanov, I., Dorai, K., Chazin, W.J., 2015.
875 Functional Dynamics in Replication Protein A DNA Binding and Protein Recruitment
876 Domains. Structure 23, 1028–1038. https://doi.org/10.1016/j.str.2015.04.008
877 Brosey, C.A., Yan, C., Tsutakawa, S.E., Heller, W.T., Rambo, R.P., Tainer, J.A., Ivanov, I.,
878 Chazin, W.J., 2013. A new structural framework for integrating replication protein A
879 into DNA processing machinery. Nucleic Acids Research 41, 2313–2327.
880 https://doi.org/10.1093/nar/gks1332
881 Chavas, L.M.G., Gourhant, P., Guimaraes, B.G., Isabet, T., Legrand, P., Lener, R.,
882 Montaville, P., Sirigu, S., Thompson, A., 2021. PROXIMA-1 beamline for
883 macromolecular crystallography measurements at Synchrotron SOLEIL. J
884 Synchrotron Rad 28, 970–976. https://doi.org/10.1107/S1600577521002605
885 Cowtan, K., 2010. Recent developments in classical density modification. Acta Crystallogr D
886 Biol Crystallogr 66, 470–478. https://doi.org/10.1107/S090744490903947X
887 Deng, X., Habel, J.E., Kabaleeswaran, V., Snell, E.H., Wold, M.S., Borgstahl, G.E.O., 2007.
888 Structure of the Full-length Human RPA14/32 Complex Gives Insights into the
889 Mechanism of DNA Binding and Complex Formation. Journal of Molecular Biology
890 374, 865–876. https://doi.org/10.1016/j.jmb.2007.09.074
891 Dueva, R., Iliakis, G., 2020. Replication protein A: a multifunctional protein with roles in
892 DNA replication, repair and beyond. NAR Cancer 2, zcaa022.
893 https://doi.org/10.1093/narcan/zcaa022
894 Duggin, I.G., Bell, S.D., 2006. The Chromosome Replication Machinery of the Archaeon
895 Sulfolobus solfataricus. Journal of Biological Chemistry 281, 15029–15032.
896 https://doi.org/10.1074/jbc.R500029200
897 Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P., Fieschi, F.,
898 2010. NADPH oxidase activator p67phox behaves in solution as a multidomain
899 protein with semi-flexible linkers. Journal of Structural Biology 169, 45–53.
900 https://doi.org/10.1016/j.jsb.2009.08.009
901 Dutta, A., Stillman, B., 1992. cdc2 family kinases phosphorylate a human cell DNA
902 replication factor, RPA, and activate DNA replication. The EMBO Journal 11, 2189–
903 2199. https://doi.org/10.1002/j.1460-2075.1992.tb05278.x
904 Emsley, P., Cowtan, K., 2004. *Coot* : model-building tools for molecular graphics. Acta
905 Crystallogr D Biol Crystallogr 60, 2126–2132.
906 https://doi.org/10.1107/S0907444904019158
907 Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of *Coot*.
908 Acta Crystallogr D Biol Crystallogr 66, 486–501.
909 https://doi.org/10.1107/S0907444910007493
910 Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R.,
911 Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A.,
912 Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P.,
913 Jumper, J., Hassabis, D., 2021. Protein complex prediction with AlphaFold-Multimer
914 (preprint). Bioinformatics. https://doi.org/10.1101/2021.10.04.463034
915 Fan, J., Pavletich, N.P., 2012. Structure and conformational change of a replication protein A
916 heterotrimer bound to ssDNA. Genes & Development 26, 2337–2347.
917 https://doi.org/10.1101/gad.194787.112
918 Feldkamp, M.D., Frank, A.O., Kennedy, J.P., Patrone, J.D., Vangamudi, B., Waterson, A.G.,
919 Fesik, S.W., Chazin, W.J., 2013. Surface Reengineering of RPA70N Enables
920 Cocrystallization with an Inhibitor of the Replication Protein A Interaction Motif of

921 ATR Interacting Protein. Biochemistry 52, 6515–6524.
922 https://doi.org/10.1021/bi400542z
923 Gamsjaeger, R., Kariawasam, R., Gimenez, A.X., Touma, C., McIlwain, E., Bernardo, R.E.,
924 Shepherd, N.E., Ataide, S.F., Dong, Q., Richard, D.J., White, M.F., Cubeddu, L.,
925 2015. The structural basis of DNA binding by the single-stranded DNA-binding
926 protein from *Sulfolobus solfataricus*. Biochemical Journal 465, 337–346.
927 https://doi.org/10.1042/BJ20141140
928 Holm, L., 2020. DALI and the persistence of protein shape. Protein Science 29, 128–140.
929 https://doi.org/10.1002/pro.3749
930 Jiang, X., Klimovich, V., Arunkumar, A.I., Hysinger, E.B., Wang, Y., Ott, R.D., Guler, G.D.,
931 Weiner, B., Chazin, W.J., Fanning, E., 2006. Structural mechanism of RPA loading on
932 DNA during activation of a simple pre-replication complex. EMBO J 25, 5516–5526.
933 https://doi.org/10.1038/sj.emboj.7601432
934 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O.,
935 Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C.,
936 Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R.,
937 Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M.,
938 Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W.,
939 Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure
940 prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-
941 03819-2
942 Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L.E., Brookes, D.H., Wilson,
943 L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D.W., Dolinsky, T., Konecny, R.,
944 Koes, D.R., Nielsen, J.E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G., Holst,
945 M.J., McCammon, J.A., Baker, N.A., 2018. Improvements to the APBS biomolecular
946 solvation software suite. Protein Science 27, 112–128.
947 https://doi.org/10.1002/pro.3280
948 Kabsch, W., 2010. *XDS*. Acta Crystallogr D Biol Crystallogr 66, 125–132.
949 https://doi.org/10.1107/S0907444909047337
950 Kerr, I.D., 2003. Insights into ssDNA recognition by the OB fold from a structural and
951 thermodynamic study of Sulfolobus SSB protein. The EMBO Journal 22, 2561–2570.
952 https://doi.org/10.1093/emboj/cdg272
953 Kõivomägi, M., Valk, E., Venta, R., Iofik, A., Lepiku, M., Balog, E.R.M., Rubin, S.M.,
954 Morgan, D.O., Loog, M., 2011. Cascades of multisite phosphorylation control Sic1
955 destruction at the onset of S phase. Nature 480, 128–131.
956 https://doi.org/10.1038/nature10560
957 Komori, K., Ishino, Y., 2001. Replication Protein A in Pyrococcus furiosus Is Involved in
958 Homologous DNA Recombination. Journal of Biological Chemistry 276, 25654–
959 25660. https://doi.org/10.1074/jbc.M102423200
960 Krissinel, E., Henrick, K., 2007. Inference of Macromolecular Assemblies from Crystalline
961 State. Journal of Molecular Biology 372, 774–797.
962 https://doi.org/10.1016/j.jmb.2007.05.022
963 Legrand, P., 2017. XDSME: XDS Made Easier. GitHub (https://github.com/legrandp/xdsme).
964 Lim, C.J., Barbour, A.T., Zaug, A.J., Goodrich, K.J., McKay, A.E., Wuttke, D.S., Cech, T.R.,
965 2020. The structure of human CST reveals a decameric assembly bound to telomeric
966 DNA. Science 368, 1081–1085. https://doi.org/10.1126/science.aaz9649
967 Liu, Y., Makarova, K.S., Huang, W.-C., Wolf, Y.I., Nikolskaya, A.N., Zhang, X., Cai, M.,
968 Zhang, C.-J., Xu, W., Luo, Z., Cheng, L., Koonin, E.V., Li, M., 2021. Expanded
969 diversity of Asgard archaea and their relationships with eukaryotes. Nature 593, 553–
970 557. https://doi.org/10.1038/s41586-021-03494-3

971 MacNeill, S.A., 2021. Remote Homology Detection Identifies a Eukaryotic RPA DBD-C-like
972 DNA Binding Domain as a Conserved Feature of Archaeal Rpa1-Like Proteins.
973 Frontiers in Molecular Biosciences 8. https://doi.org/10.3389/fmolb.2021.675229
974 Makarova, K.S., Koonin, E.V., 2013. Archaeology of Eukaryotic DNA Replication. Cold
975 Spring Harbor Perspectives in Biology 5, a012963–a012963.
976 https://doi.org/10.1101/cshperspect.a012963
977 Marceau, A.H., 2012. Functions of Single-Strand DNA-Binding Proteins in DNA Replication,
978 Recombination, and Repair, in: Keck, J.L. (Ed.), Single-Stranded DNA Binding
979 Proteins. Humana Press, Totowa, NJ, pp. 1–21. https://doi.org/10.1007/978-1-62703-
980 032-8_1
981 Maréchal, A., Zou, L., 2015. RPA-coated single-stranded DNA as a platform for post-
982 translational modifications in the DNA damage response. Cell Res 25, 9–23.
983 https://doi.org/10.1038/cr.2014.147
984 McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J.,
985 2007. *Phaser* crystallographic software. J Appl Crystallogr 40, 658–674.
986 https://doi.org/10.1107/S0021889807021206
987 Meyer, R.R., Laine, P.S., 1990. The single-stranded DNA-binding protein of Escherichia coli.
988 Microbiol Rev 54, 342–380. https://doi.org/10.1128/mr.54.4.342-380.1990
989 Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M., 2021.
990 ColabFold - Making protein folding accessible to all (preprint). Bioinformatics.
991 https://doi.org/10.1101/2021.08.15.456425
992 Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L.,
993 Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., Bateman, A., 2021.
994 Pfam: The protein families database in 2021. Nucleic Acids Research 49, D412–D419.
995 https://doi.org/10.1093/nar/gkaa913
996 Miyake, Y., Nakamura, M., Nabetani, A., Shimamura, S., Tamura, M., Yonehara, S., Saito,
997 M., Ishikawa, F., 2009. RPA-like Mammalian Ctc1-Stn1-Ten1 Complex Binds to
998 Single-Stranded DNA and Protects Telomeres Independently of the Pot1 Pathway.
999 Molecular Cell 36, 193–206. https://doi.org/10.1016/j.molcel.2009.08.009
1000 Morten, M.J., Peregrina, J.R., Figueira-Gonzalez, M., Ackermann, K., Bode, B.E., White,
1001 M.F., Penedo, J.C., 2015. Binding dynamics of a monomeric SSB protein to DNA: a
1002 single-molecule multi-process approach. Nucleic Acids Res 43, 10907–10924.
1003 https://doi.org/10.1093/nar/gkv1225
1004 Nagata, M., Ishino, S., Yamagami, T., Ishino, Y., 2019. Replication protein A complex in
1005 *Thermococcus kodakarensis* interacts with DNA polymerases and helps their effective
1006 strand synthesis. Bioscience, Biotechnology, and Biochemistry 83, 695–704.
1007 https://doi.org/10.1080/09168451.2018.1559722
1008 Oliveira, M.T. (Ed.), 2021. Single Stranded DNA Binding Proteins, Methods in Molecular
1009 Biology. Springer US, New York, NY. https://doi.org/10.1007/978-1-0716-1290-3
1010 Pan, Z.Q., Amin, A.A., Gibbs, E., Niu, H., Hurwitz, J., 1994. Phosphorylation of the p34
1011 subunit of human single-stranded-DNA-binding protein in cyclin A-activated G1
1012 extracts is catalyzed by cdk-cyclin A complex and DNA-dependent protein kinase.
1013 Proc. Natl. Acad. Sci. U.S.A. 91, 8343–8347. https://doi.org/10.1073/pnas.91.18.8343
1014 Park, C.-J., 2005. Solution structure of the DNA-binding domain of RPA from
1015 Saccharomyces cerevisiae and its interaction with single-stranded DNA and SV40 T
1016 antigen. Nucleic Acids Research 33, 4172–4181. https://doi.org/10.1093/nar/gki736
1017 Pérez, J., Vachette, P., Russo, D., Desmadril, M., Durand, D., 2001. Heat-induced unfolding
1018 of neocarzinostatin, a small all-β protein investigated by small-angle X-ray scattering
1019 1 1Edited by M. F. Moody. Journal of Molecular Biology 308, 721–743.
1020 https://doi.org/10.1006/jmbi.2001.4611

1021 Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris,
1022 J.H., Ferrin, T.E., 2021. UCSF CHIMERAX : Structure visualization for researchers,
1023 educators, and developers. Protein Science 30, 70–82.
1024 https://doi.org/10.1002/pro.3943
1025 Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A., 2017. cryoSPARC: algorithms for
1026 rapid unsupervised cryo-EM structure determination. Nat Methods 14, 290–296.
1027 https://doi.org/10.1038/nmeth.4169
1028 Raymann, K., Forterre, P., Brochier-Armanet, C., Gribaldo, S., 2014. Global Phylogenomic
1029 Analysis Disentangles the Complex Evolutionary History of DNA Replication in
1030 Archaea. Genome Biology and Evolution 6, 192–212.
1031 https://doi.org/10.1093/gbe/evu004
1032 Robbins, J.B., Murphy, M.C., White, B.A., Mackie, R.I., Ha, T., Cann, I.K.O., 2004.
1033 Functional Analysis of Multiple Single-stranded DNA-binding Proteins from
1034 Methanosarcina acetivorans and Their Effects on DNA Synthesis by DNA Polymerase
1035 BI. Journal of Biological Chemistry 279, 6315–6326.
1036 https://doi.org/10.1074/jbc.M304491200
1037 Seeber, A., Hegnauer, A.M., Hustedt, N., Deshpande, I., Poli, J., Eglinger, J., Pasero, P., Gut,
1038 H., Shinohara, M., Hopfner, K.-P., Shimada, K., Gasser, S.M., 2016. RPA Mediates
1039 Recruitment of MRX to Forks and Double-Strand Breaks to Hold Sister Chromatids
1040 Together. Molecular Cell 64, 951–966. https://doi.org/10.1016/j.molcel.2016.10.032
1041 Sheldrick, G.M., 2008. A short history of *SHELX*. Acta Crystallogr A Found Crystallogr 64,
1042 112–122. https://doi.org/10.1107/S0108767307043930
1043 Sibenaller, Z.A., Sorensen, B.R., Wold, M.S., 1998. The 32- and 14-Kilodalton Subunits of
1044 Replication Protein A Are Responsible for Species-Specific Interactions with Single-
1045 Stranded DNA. Biochemistry 37, 12496–12506. https://doi.org/10.1021/bi981110+
1046 Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E.,
1047 van Eijk, R., Schleper, C., Guy, L., Ettema, T.J.G., 2015. Complex archaea that bridge
1048 the gap between prokaryotes and eukaryotes. Nature 521, 173–179.
1049 https://doi.org/10.1038/nature14447
1050 Stephan, H., Concannon, C., Kremmer, E., Carty, M.P., Nasheuer, H.-P., 2009. Ionizing
1051 radiation-dependent and independent phosphorylation of the 32-kDa subunit of
1052 replication protein A during mitosis. Nucleic Acids Research 37, 6028–6041.
1053 https://doi.org/10.1093/nar/gkp605
1054 Stroud, A., Liddell, S., Allers, T., 2012. Genetic and Biochemical Identification of a Novel
1055 Single-Stranded DNA-Binding Complex in Haloferax volcanii. Frontiers in
1056 Microbiology 3. https://doi.org/10.3389/fmicb.2012.00224
1057 Surovtseva, Y.V., Churikov, D., Boltz, K.A., Song, X., Lamb, J.C., Warrington, R., Leehy,
1058 K., Heacock, M., Price, C.M., Shippen, D.E., 2009. Conserved Telomere Maintenance
1059 Component 1 Interacts with STN1 and Maintains Chromosome Ends in Higher
1060 Eukaryotes. Molecular Cell 36, 207–218. https://doi.org/10.1016/j.molcel.2009.09.017
1061 Svergun, D.I., 1992. Determination of the regularization parameter in indirect-transform
1062 methods using perceptual criteria. J Appl Crystallogr 25, 495–503.
1063 https://doi.org/10.1107/S0021889892001663
1064 Taib, N., Gribaldo, S., MacNeill, S.A., 2021. Single-Stranded DNA-Binding Proteins in the
1065 Archaea, in: Oliveira, M.T. (Ed.), Single Stranded DNA Binding Proteins. Springer
1066 US, New York, NY, pp. 23–47. https://doi.org/10.1007/978-1-0716-1290-3_2
1067 Terwilliger, T.C., Poon, B.K., Afonine, P.V., Schlicksup, C.J., Croll, T.I., Millán, C.,
1068 Richardson, J.S., Read, R.J., Adams, P.D., 2022. Improved AlphaFold modeling with
1069 implicit experimental information (preprint). Biochemistry.
1070 https://doi.org/10.1101/2022.01.07.475350

1071 Theobald, D.L., Mitton-Fry, R.M., Wuttke, D.S., 2003. Nucleic Acid Recognition by OB-
1072 Fold Proteins. Annu. Rev. Biophys. Biomol. Struct. 32, 115–133.
1073 https://doi.org/10.1146/annurev.biophys.32.110601.142506
1074 Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., Bricogne, G.,
1075 2018. STARANISO (http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi).
1076 Cambridge, United Kingdom: Global Phasing Ltd.
1077 Vagin, A., Teplyakov, A., 1997. *MOLREP* : an Automated Program for Molecular
1078 Replacement. Journal of Applied Crystallography 30, 1022–1025.
1079 https://doi.org/10.1107/S0021889897006766
1080 Wadsworth, R.I.M., 2001. Identification and properties of the crenarchaeal single-stranded
1081 DNA binding protein from Sulfolobus solfataricus. Nucleic Acids Research 29, 914–
1082 920. https://doi.org/10.1093/nar/29.4.914
1083 Wold, M.S., 1997. REPLICATION PROTEIN A: A Heterotrimeric, Single-Stranded DNA-
1084 Binding Protein Required for Eukaryotic DNA Metabolism. Annual Review of
1085 Biochemistry 66, 61–92. https://doi.org/10.1146/annurev.biochem.66.1.61
1086 Yates, L.A., Aramayo, R.J., Pokhrel, N., Caldwell, C.C., Kaplan, J.A., Perera, R.L., Spies,
1087 M., Antony, E., Zhang, X., 2018. A structural and dynamic model for the assembly of
1088 Replication Protein A on single-stranded DNA. Nat Commun 9, 5447.
1089 https://doi.org/10.1038/s41467-018-07883-7
1090

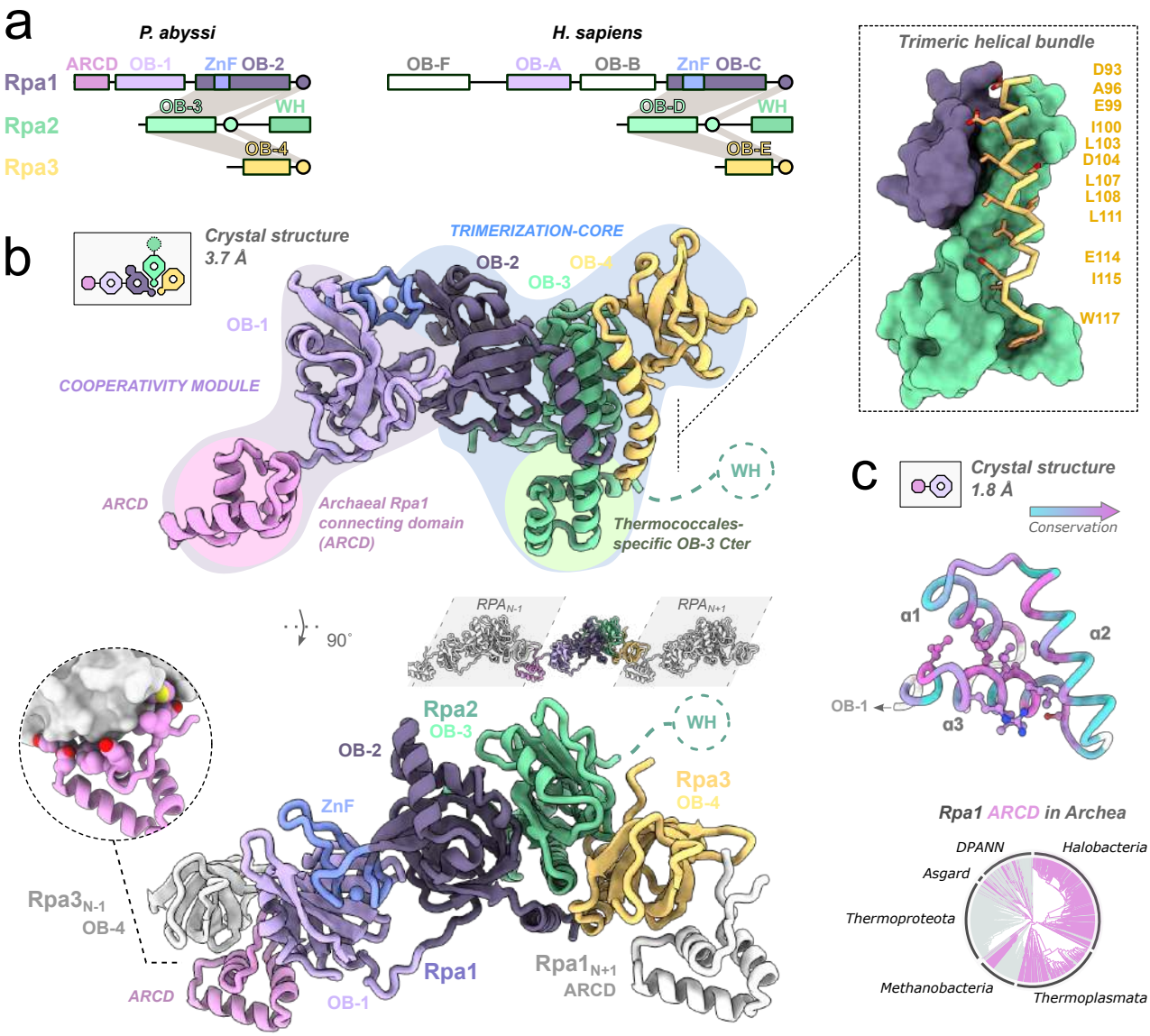



Figure 1: Structure of the archaeal RPA. a. Domain diagrams of PabRPA and human RPA. b. Two orthogonal views of the full-length PabRPA crystal structure at 3.7 Å with a focused view on the trimeric helical bundle (top right). Crystal lattice contacts within neighboring RPA heterotrimers are shown in grey. c. High resolution crystal structure of the Rpa1 cooperativity module at 1.8 Å (top) with a focused view on the newly identified archaeal-specific ARCD domain, which is colored according to sequence conservation. Side chains of residues connecting ARCD to OB-4 are shown as ball-and-sticks. Phylogenetic analysis (bottom) reveals a broad distribution of Rpa1-ARCD domain homologs across Archaea. Genomes lacking ARCD domain are presented in grey.

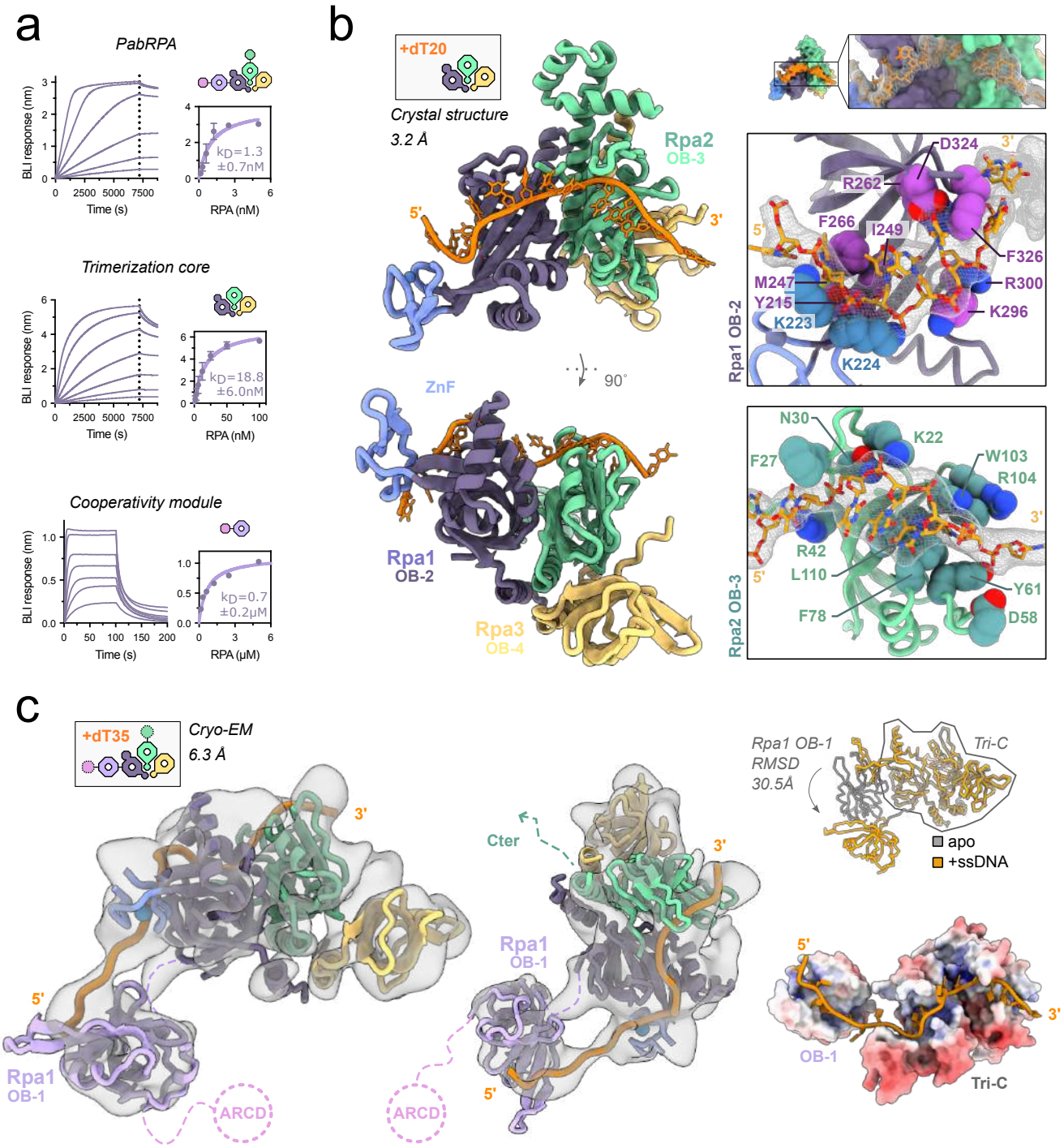



Figure 2: Structural basis for the DNA-binding activity by PabRPA. a. DNA-binding properties of PabRPA trimerization core and cooperativity module. Specific binding of PabRPA constructs to immobilized poly-dT35 ssDNA was measured by biolayer interferometry. Steady-state analysis were performed using the average signal measured at the end of the association steps. b. 3.2 Å crystal structure of the poly-dT20-bound Tri-C, with two focused views on Rpa1 OB-2 and Rpa2 OB-3 ssDNA-interacting residues. c. Cryo-EM structure of full-length PabRPA bound to a poly-dT35 ssDNA. Backbone traces of the poly-dT20-bound Tri-C and the Rpa1 OB-1 crystal structures were rigid-body fitted into the cryo-EM map contoured at a level of 1rmsd. Conformational changes of OB-1, observed between the apo-PabRPA crystal structure (in grey), and the DNA-bound cryo-EM structure (in orange). Electrostatic potential surface analysis showing that the repositioning of Rpa1 OB-1 domain extends the ssDNA-binding channel.

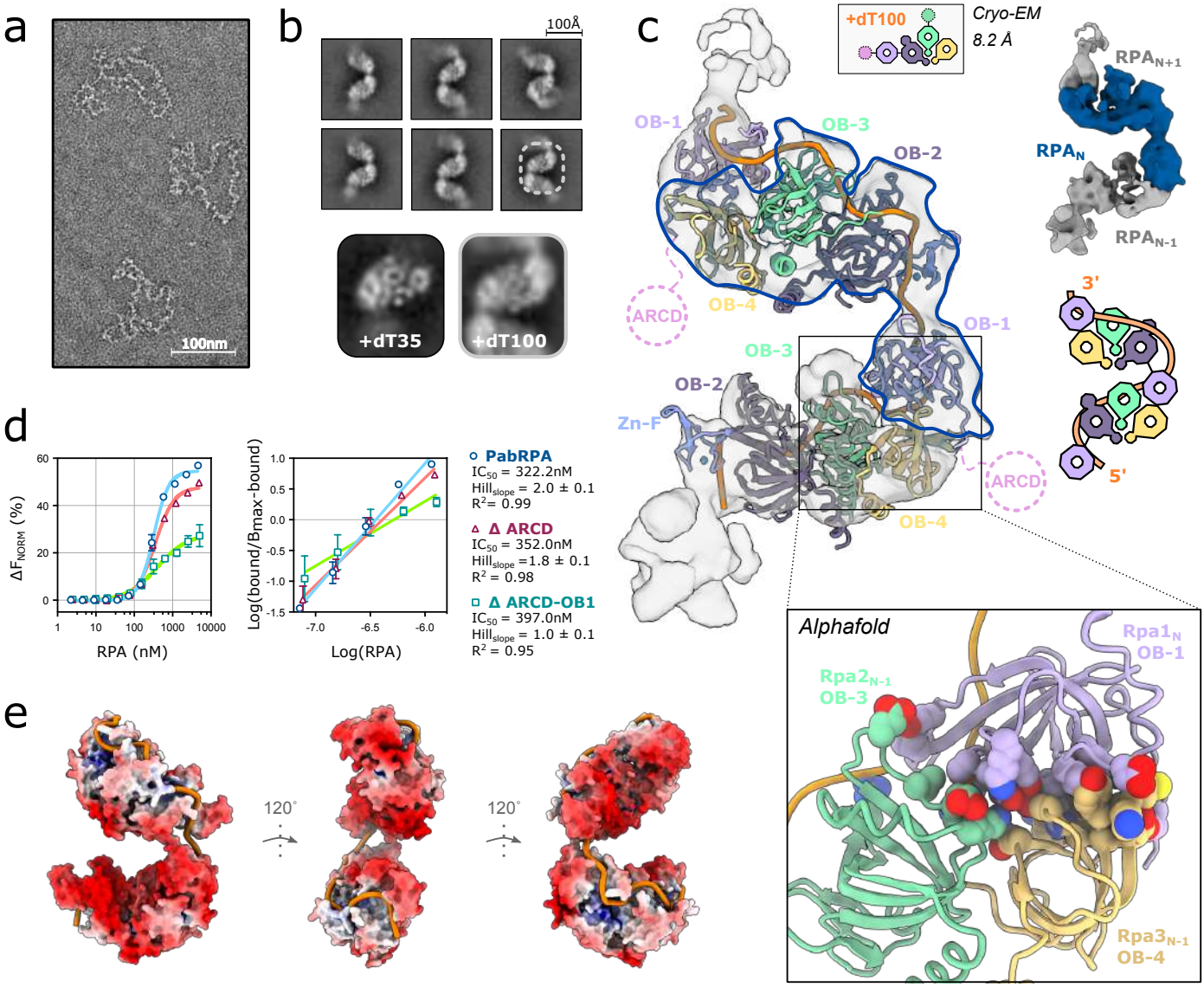



Figure 3: Structural basis for the assembly of multiple PabRPA molecules on ssDNA. a. Negative staining microscopy of PabRPA bound to M13 ssDNA plasmid. b. Cryo-EM 2D classes of PabRPA bound to a poly-dT100. c. Cryo-EM structure of PabRPA bound to a poly-dT100 displayed as backbone traces with a focused view on the Tri-C(n)/OB-1(n+1) AlphaFold model used for model building. The cryo-EM map is contoured at a level of 6rmsd d. Biophysical assessment of RPA full-length and cooperativity-module truncated constructs binding onto long ssDNA. Microscale thermophoresis binding curves of RPA heterotrimeric complexes interacting with ssDNA, using wild-type PabRPA (blue), ΔARCD mutant (red) or ΔARCD-OB-1 mutant and a Cy5-labeled poly-dT100 ssDNA substrate. EC50 values were calculated using the Hill-equation. Hill coefficients were calculated from the Hill-plot transform of the binding data. e. Electrostatic potential surface analysis showing that PabRPA forms a continuous positive groove along the DNA path (orange).

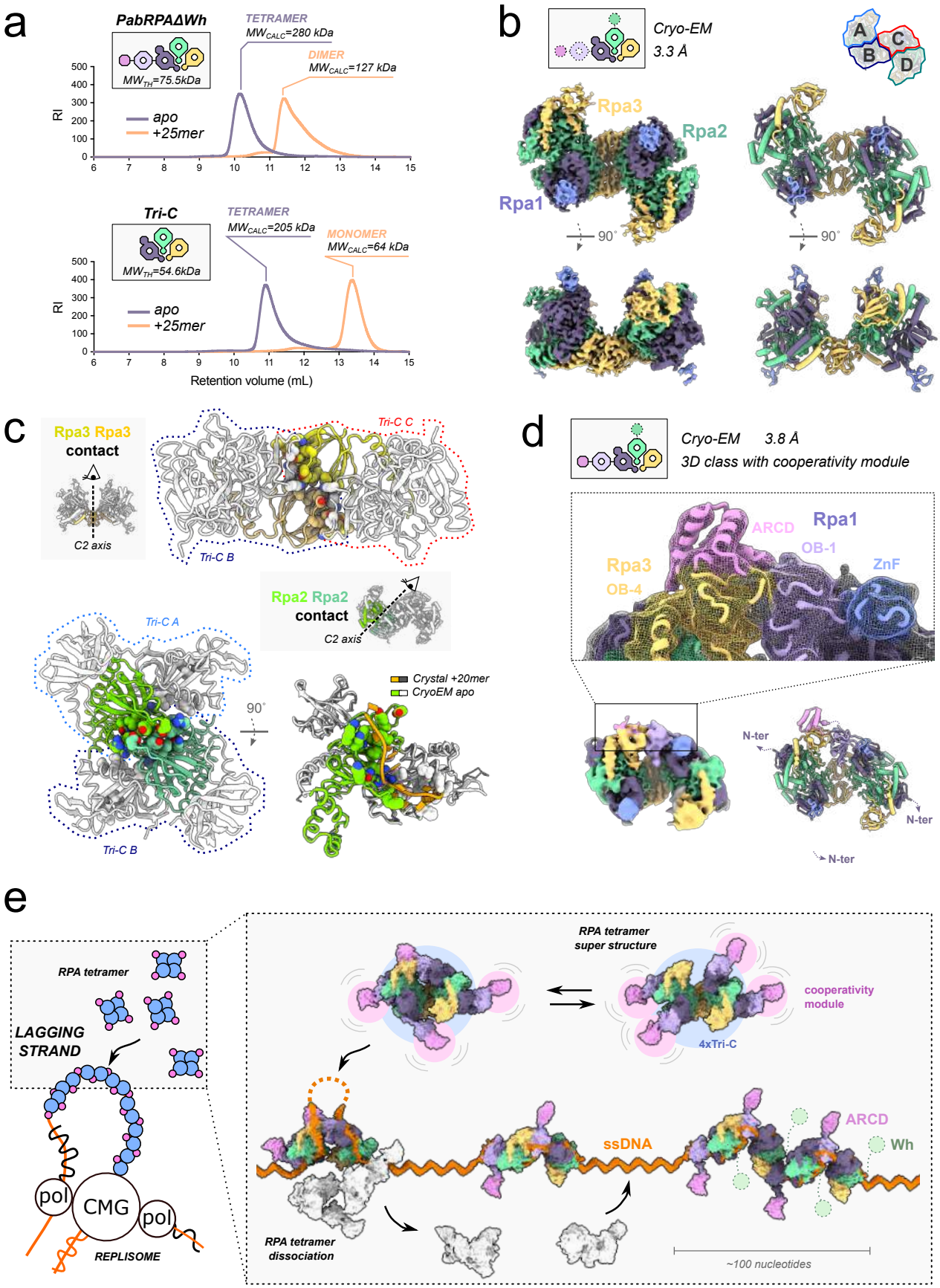



Figure 4: PabRPA assembles as a tetrameric super-complex that dissociates upon binding to ssDNA. a. SEC-SLS characterization of PabRPA and PabRPA Tri-C in presence or absence of poly-d25T ssDNA. The theoretical (MWth) and calculated molecular weights (MWcalc) for each complex are given in kDa. b. 3.35 Å cryo-EM structure of the tetrameric PabRPA super-structure. c. Focused view on critical contacts within the PabRPA tetrameric assembly. d. 3D class showing contacts between the cooperativity module and Rpa3 of two neighboring PabRPA molecules within the tetrameric super-structure. e. Proposed mechanism for the self-association of PabRPA on ssDNA.

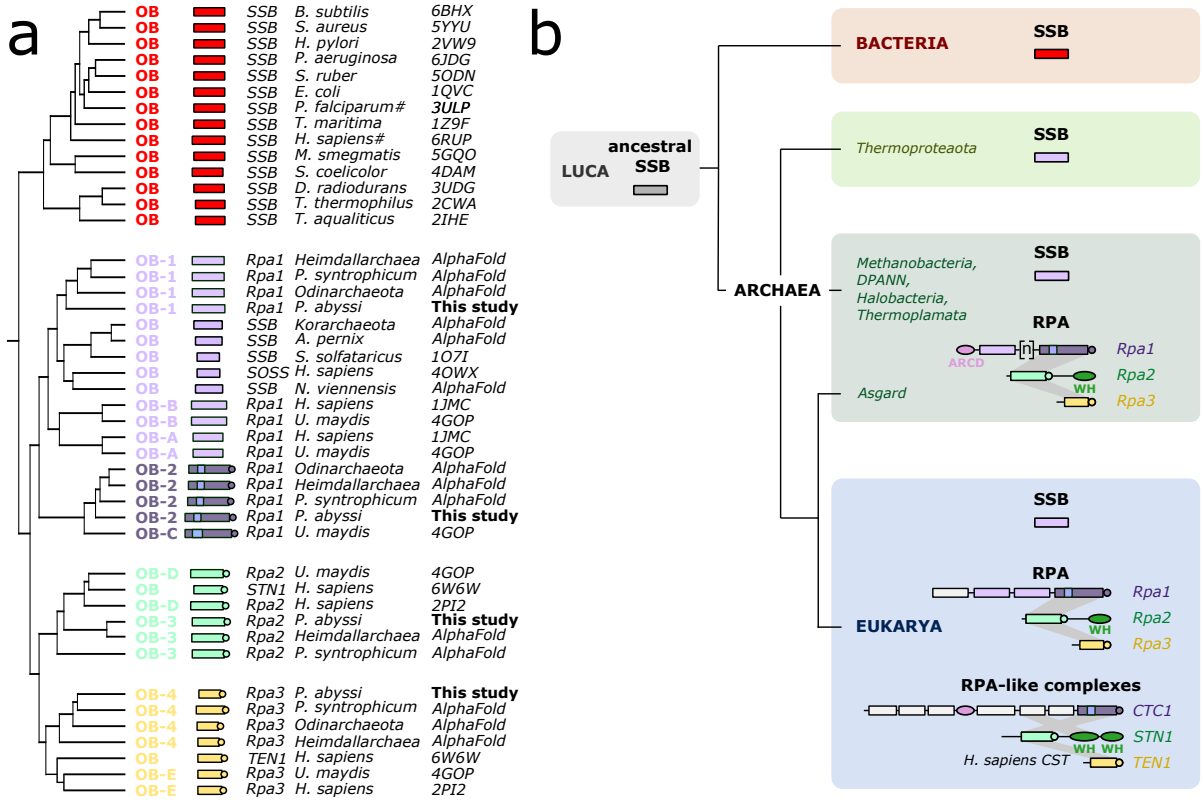



Figure 5: Evolutionary implications for the RPA evolution in Archaea and Eukarya. a. Structure-based dendrogram produced by DALI, based on average linkage clustering of the Z-scores of the structural similarity matrix. b. Current proposal for the evolutionary history of SSB and RPA in the three domains of life.

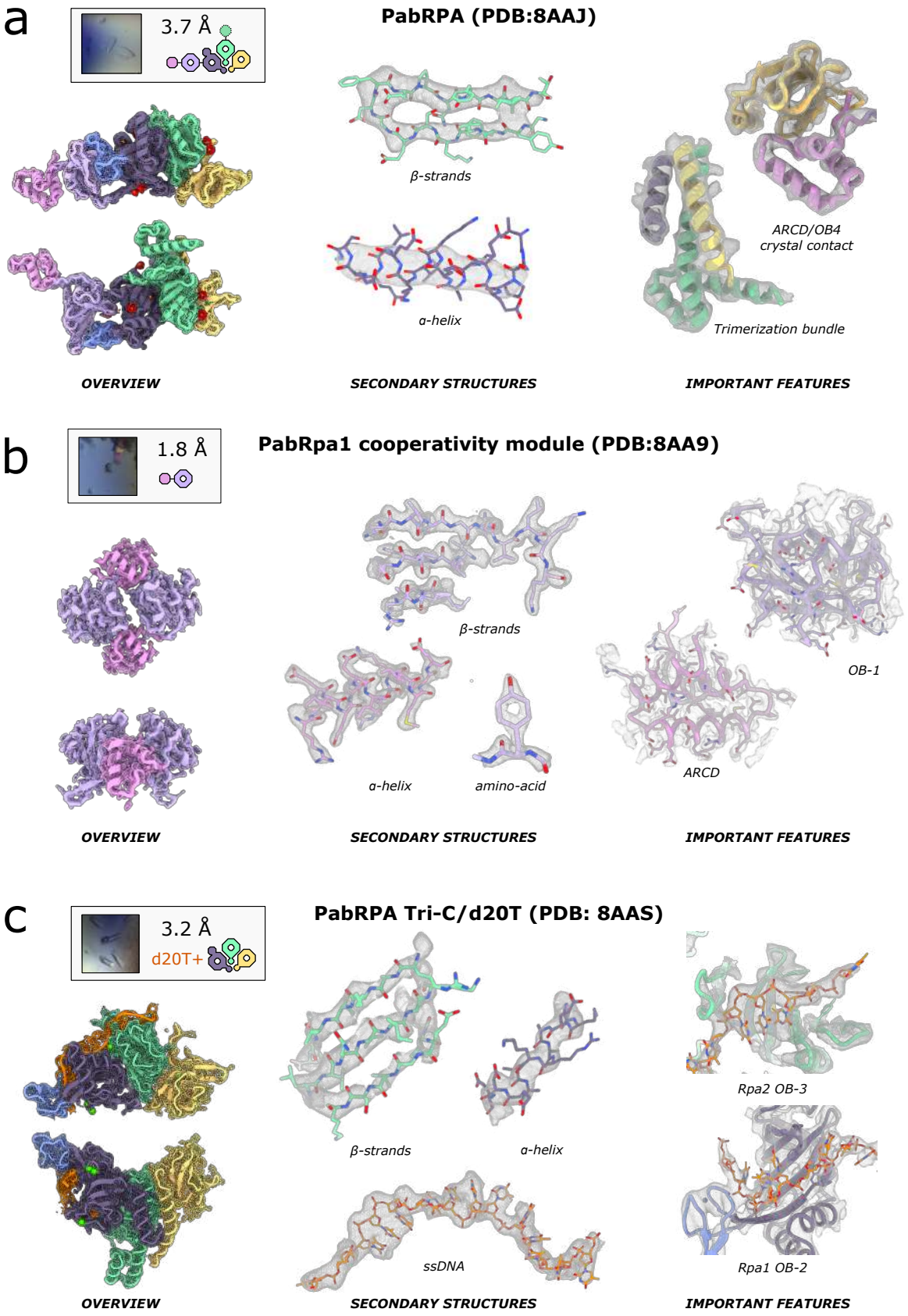



Figure S1: Electron density in representative regions of the PabRPA crystal structures. PabRPA (a), PabRPA Rpa1 cooperativity module (b) and PabRPA Tri-C/20dT (c) crystal structures are represented as Cα traces in the left panels. Right panels show close views of functionally important regions of PabRPA. Main chain and side chains atoms are represented as Cα traces and sticks, respectively. The gray mesh is the maximum likelihood 2mFo-DFc electron density map contoured at a level of 1 σ.

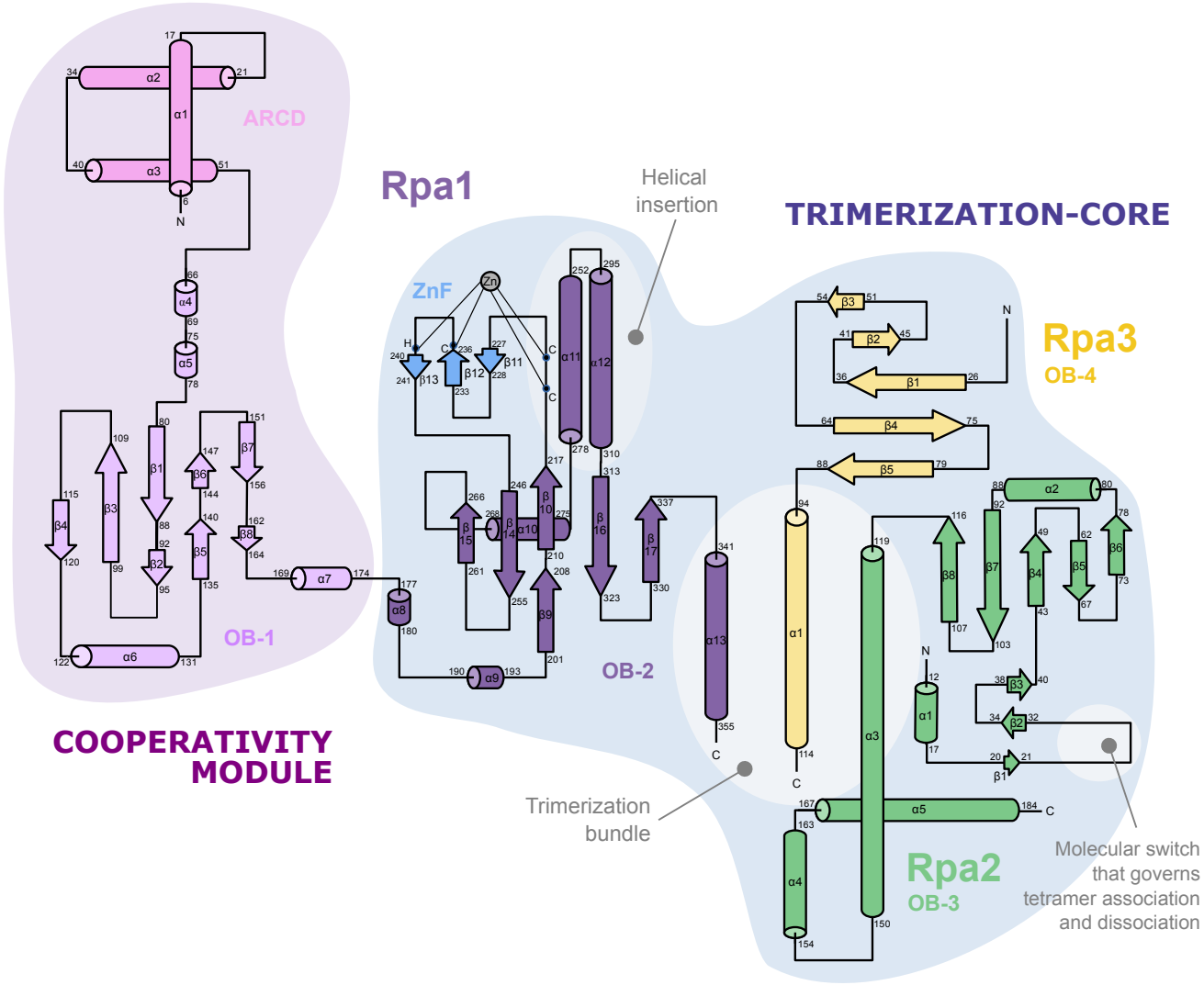
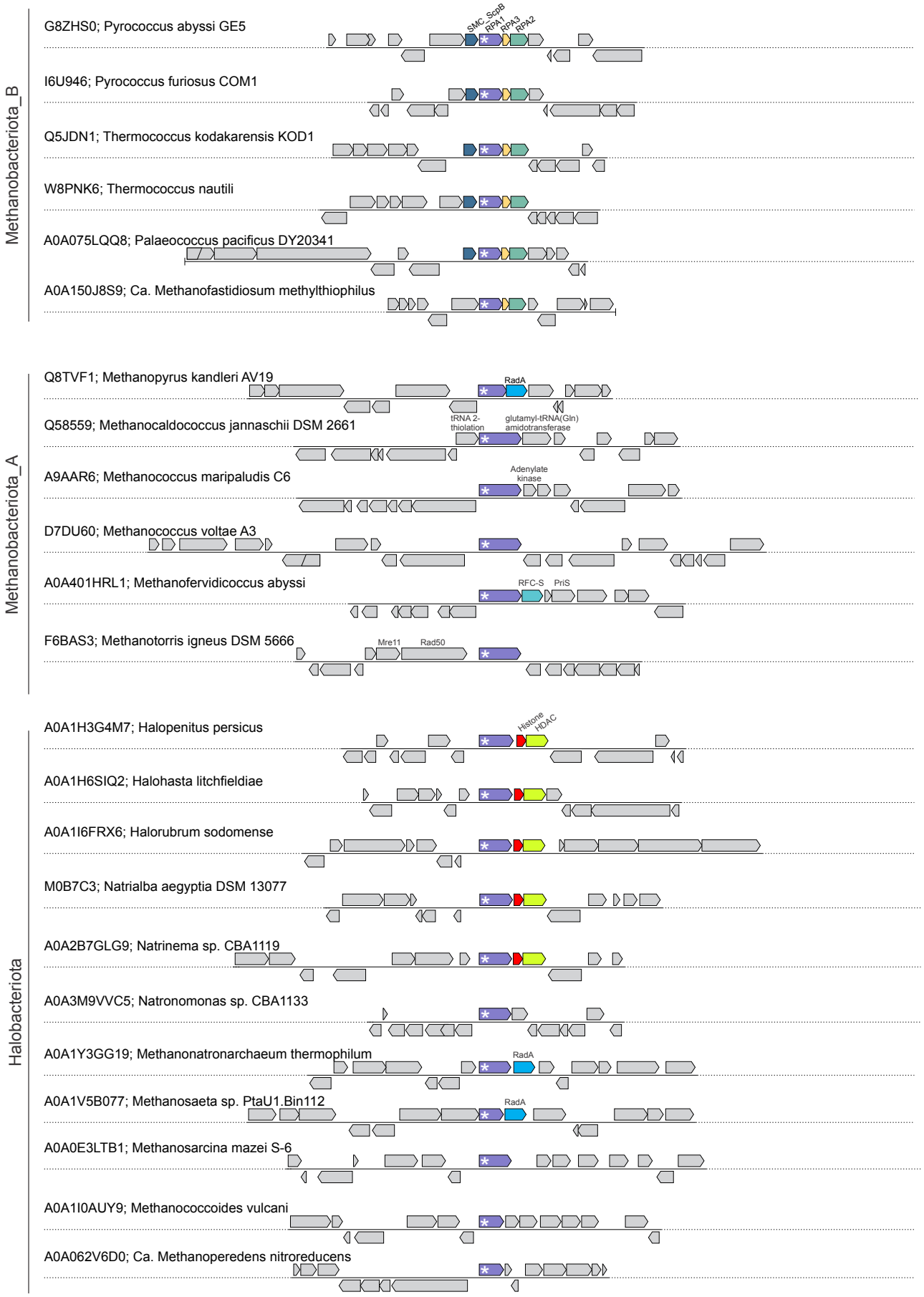
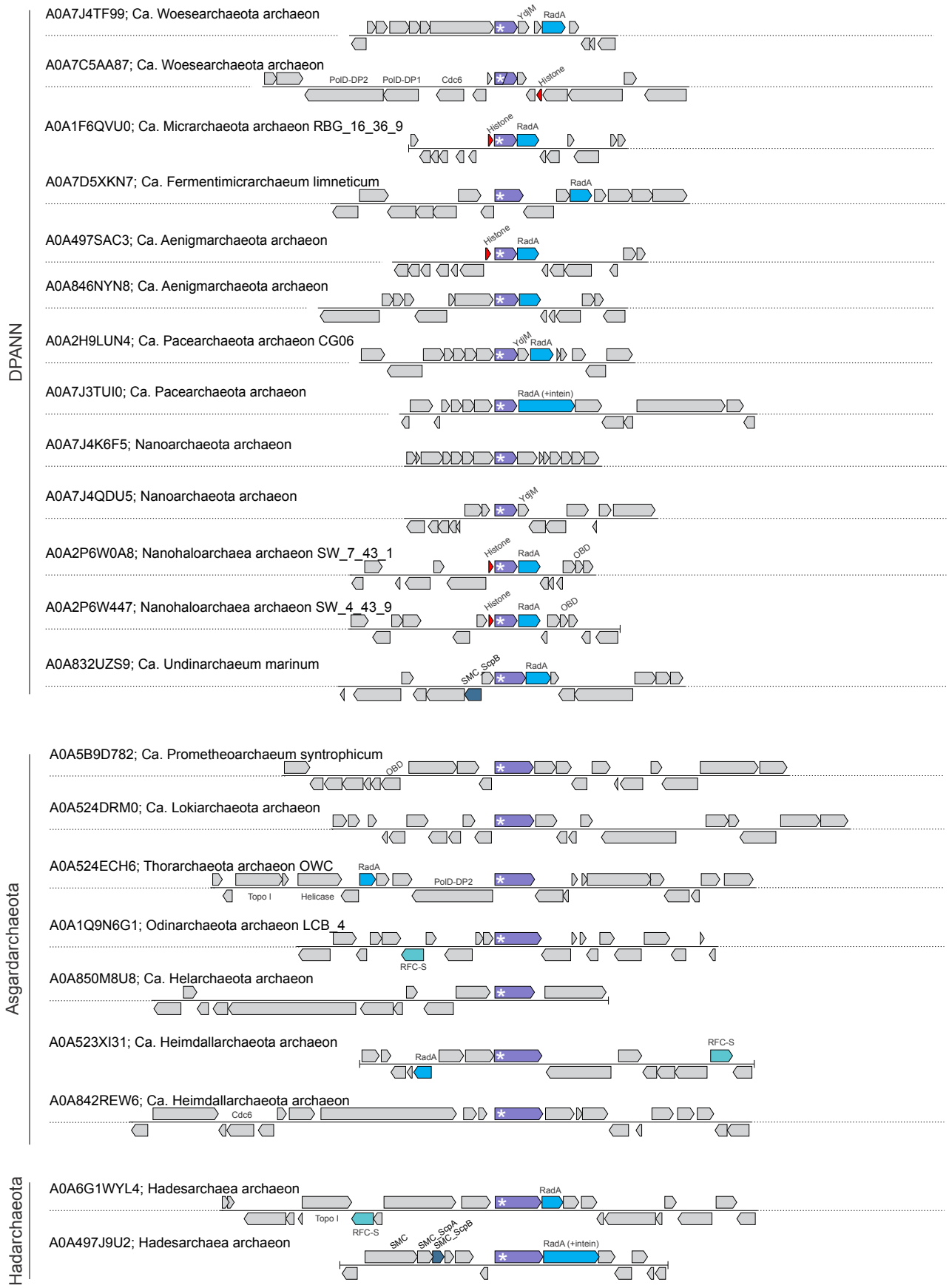



Figure S2: Topological diagram of PabRPA. Two-dimensional diagram of protein topology for three subunits of PabRPA.


Methanobacteriota_B
G8ZHS0; Pyrococcus abyssi GE5
SMC
ScpB
RPA1
RPA3
RPA2
I6U946; Pyrococcus furiosus COM1
Q5JDN1; Thermococcus kodakarensis KOD1
W8PNK6; Thermococcus nautili
A0A075LQQ8; Palaeococcus pacificus DY20341
A0A150J8S9; Ca. Methanofastidiosum methylthiophilus
Methanobacteriota_A
Q8TVF1; Methanopyrus kandleri AV19
RadA
Q58559; Methanocaldococcus jannaschii DSM 2661
tRNA 2-thiolation
glutamyl-tRNA(Gln) amidotransferase
A9AAR6; Methanococcus maripaludis C6
Adenylate kinase
D7DU60; Methanococcus voltae A3
A0A401HRL1; Methanofervidicoccus abyssi
RFC-S
PriS
F6BAS3; Methanotorris igneus DSM 5666
Mre11
Rad50
Halobacteriota
A0A1H3G4M7; Halopenitus persicus
Histone HDAC
A0A1H6SIQ2; Halohasta litchfieldiae
A0A1I6FRX6; Halorubrum sodomense
M0B7C3; Natrialba aegyptia DSM 13077
A0A2B7GLG9; Natrinema sp. CBA1119
A0A3M9VVC5; Natronomonas sp. CBA1133
A0A1Y3GG19; Methanonatronarchaeum thermophilum
RadA
A0A1V5B077; Methanosaeta sp. PtaU1.Bin112
RadA
A0A0E3LTB1; Methanosarcina mazei S-6
A0A1I0AUY9; Methanococcoides vulcani
A0A062V6D0; Ca. Methanoperedens nitroreducens


DPANN
A0A7J4TF99; Ca. Woesearchaeota archaeon
YdjM
RadA
A0A7C5AA87; Ca. Woesearchaeota archaeon
PolD-DP2
PolD-DP1
Cdc6
Histone
A0A1F6QVU0; Ca. Micrarchaeota archaeon RBG_16_36_9
Histone
RadA
A0A7D5XKN7; Ca. Fermentimicrarchaeum limneticum
RadA
A0A497SAC3; Ca. Aenigmarchaeota archaeon
Histone
RadA
A0A846NYN8; Ca. Aenigmarchaeota archaeon
A0A2H9LUN4; Ca. Pacearchaeota archaeon CG06
YdjM
RadA
A0A7J3TUI0; Ca. Pacearchaeota archaeon
RadA (+intein)
A0A7J4K6F5; Nanoarchaeota archaeon
A0A7J4QDU5; Nanoarchaeota archaeon
YdjM
A0A2P6W0A8; Nanohaloarchaea archaeon SW_7_43_1
Histone
RadA
OBD
A0A2P6W447; Nanohaloarchaea archaeon SW_4_43_9
Histone
RadA
OBD
A0A832UZS9; Ca. Undinarchaeum marinum
SMC
ScpB
RadA
Asgardarchaeota
A0A5B9D782; Ca. Prometheoarchaeum syntrophicum
OBD
A0A524DRM0; Ca. Lokiarchaeota archaeon
A0A524ECH6; Thorarchaeota archaeon OWC
RadA
PolD-DP2
Topo I
Helicase
A0A1Q9N6G1; Odinarchaeota archaeon LCB_4
RFC-S
A0A850M8U8; Ca. Helarchaeota archaeon
A0A523XI31; Ca. Heimdallarchaeota archaeon
RadA
RFC-S
A0A842REW6; Ca. Heimdallarchaeota archaeon
Cdc6
Hadarchaeota
A0A6G1WYL4; Hadesarchaea archaeon
RadA
Topo I
RFC-S
A0A497J9U2; Hadesarchaea archaeon
SMC
SMC_ScpA
SMC_ScpB
RadA (+intein)

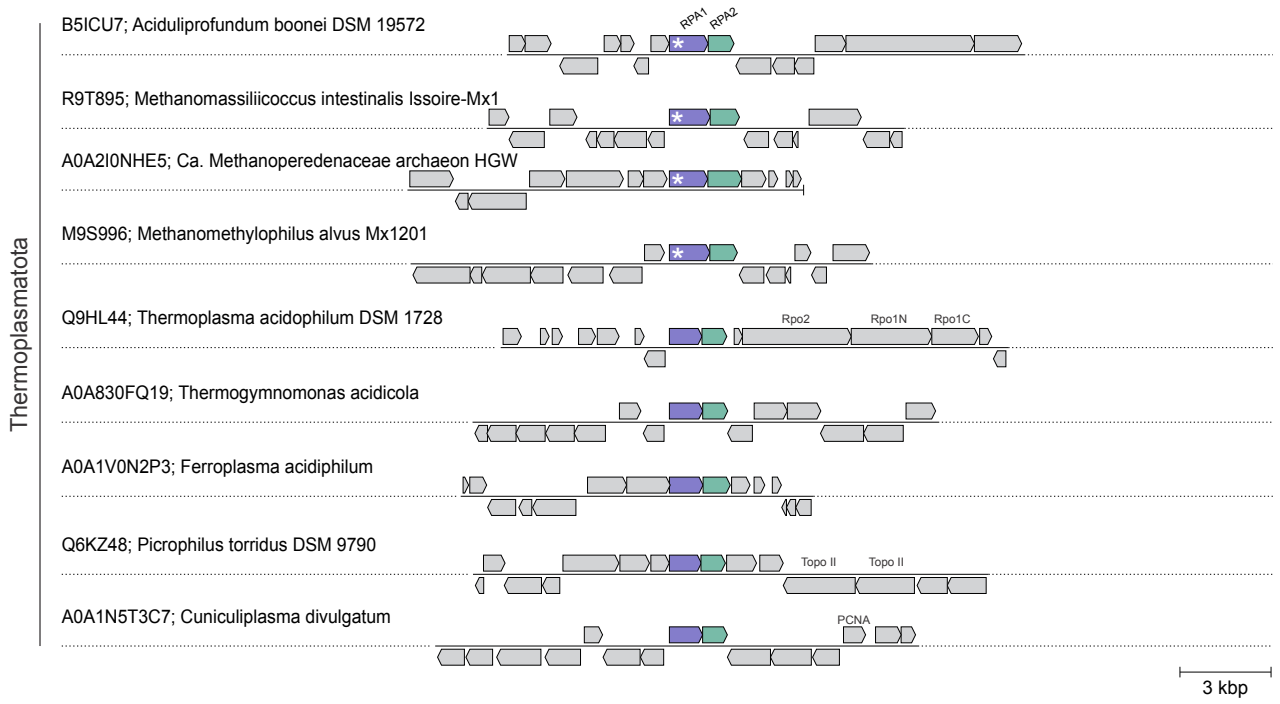
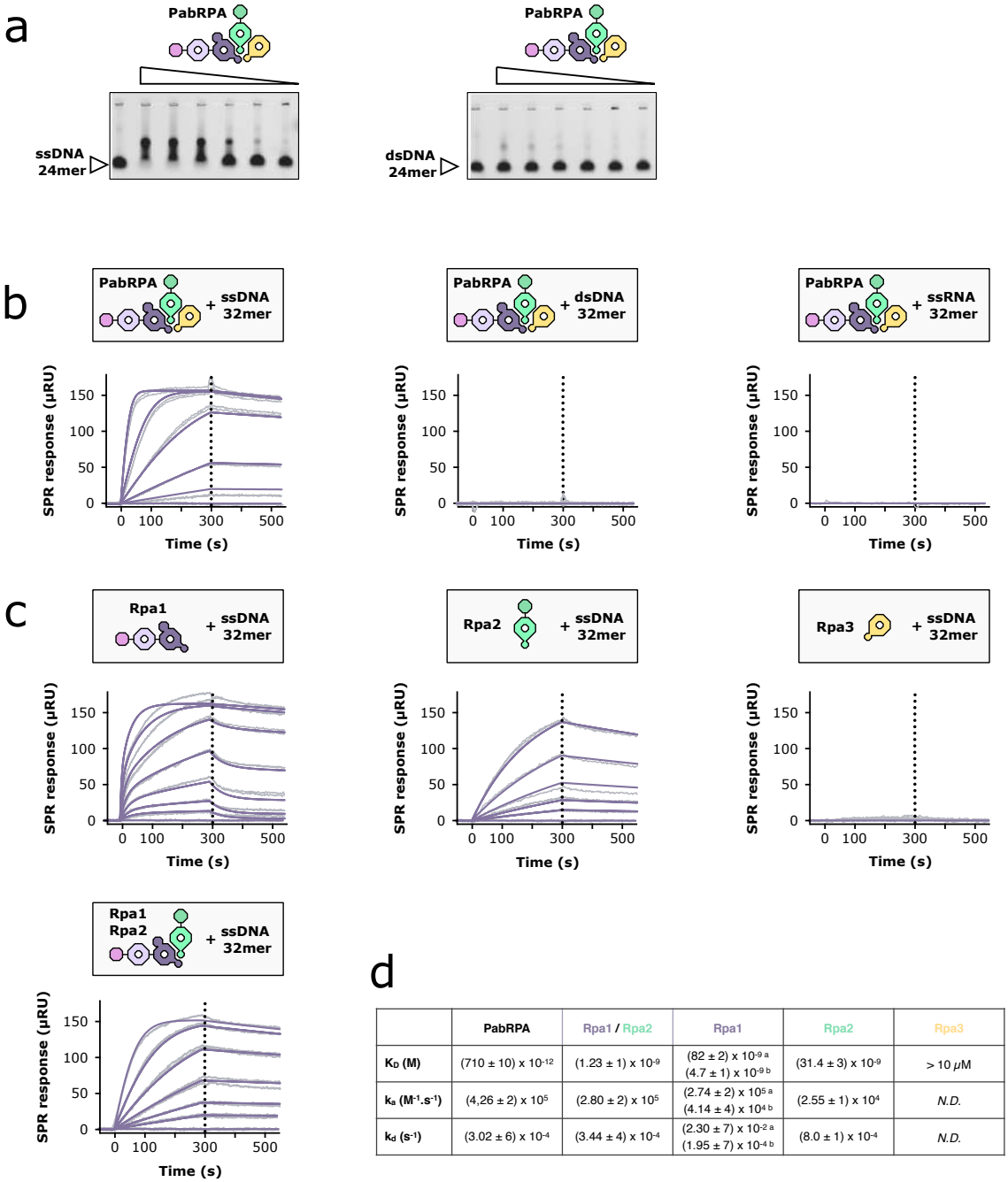



Figure S3: Comparison of genomic loci encoding Rpa1 genes in Archaea. Genomic organizations of the loci encoding Rpa1 genes are shown for representative species from each major archaeal phyla. Each locus is indicated with the UniProt accession number of the Rpa1 gene, followed by the species name. All loci are shown as -7 and +7 genes around Rpa1 gene. Rpa1 genes are embedded within highly variable genomic neighborhoods, which typically vary even for relatively closely related organisms. Presence of the Archaea-specific ARCD is indicated with white asterisks. Strikingly, ARCD is present in all Rpa1 sequences, except for the 5 Thermoplasmatales members. It is noteworthy that Methanomassiliicoccales within Thermoplasmatota do contain ARCD, suggesting that ARCD was lost within Thermoplasmatales following their divergence from a common ancestor with Methanomassiliicoccales. Interestingly, Aciduliprofundum boonei which is even closer to the Thermoplasmatales also contain ARCD. DNA metabolism proteins encoding genes are represented in color.




| | PabRPA | Rpa1 / Rpa2 | Rpa1 | Rpa2 | Rpa3 |
|---|---|---|---|---|---|
| $K_D$ (M) | (710 ± 10) x $10^{-12}$ | (1.23 ± 1) x $10^{-9}$ | (82 ± 2) x $10^{-9}$ [a]<br>(4.7 ± 1) x $10^{-9}$ [b] | (31.4 ± 3) x $10^{-9}$ | > 10 µM |
| $k_a$ ($M^{-1}.s^{-1}$) | (4,26 ± 2) x $10^{5}$ | (2.80 ± 2) x $10^{5}$ | (2.74 ± 2) x $10^{5}$ [a]<br>(4.14 ± 4) x $10^{4}$ [b] | (2.55 ± 1) x $10^{4}$ | *N.D.* |
| $k_d$ ($s^{-1}$) | (3.02 ± 6) x $10^{-4}$ | (3.44 ± 4) x $10^{-4}$ | (2.30 ± 7) x $10^{-2}$ [a]<br>(1.95 ± 7) x $10^{-4}$ [b] | (8.0 ± 1) x $10^{-4}$ | *N.D.* |

Figure S4: Nucleic-acids binding properties of PabRPA. a. Electrophoretic mobility shift assay of ssDNA-24mer and dsDNA-24mer (2.5 pmol) in presence of decreasing concentrations of PabRPA (312, 62, 12, 2.5, 0.5, 0.1 pmol). b. PabRPA binding specificity on nucleic acids. Specific binding of PabRPA heterotrimer on immobilized ssDNA-32mer (left), dsDNA-32mer (middle) and ssRNA-32mer measured by surface plasmon resonance (RU: resonance units). c. Role of PabRPA subunits on ssDNA binding. Specific binding of Rpa1, Rpa2, Rpa3, or Rpa1/Rpa2 complex on immobilized ssDNA-32mer measured by surface plasmon resonance.

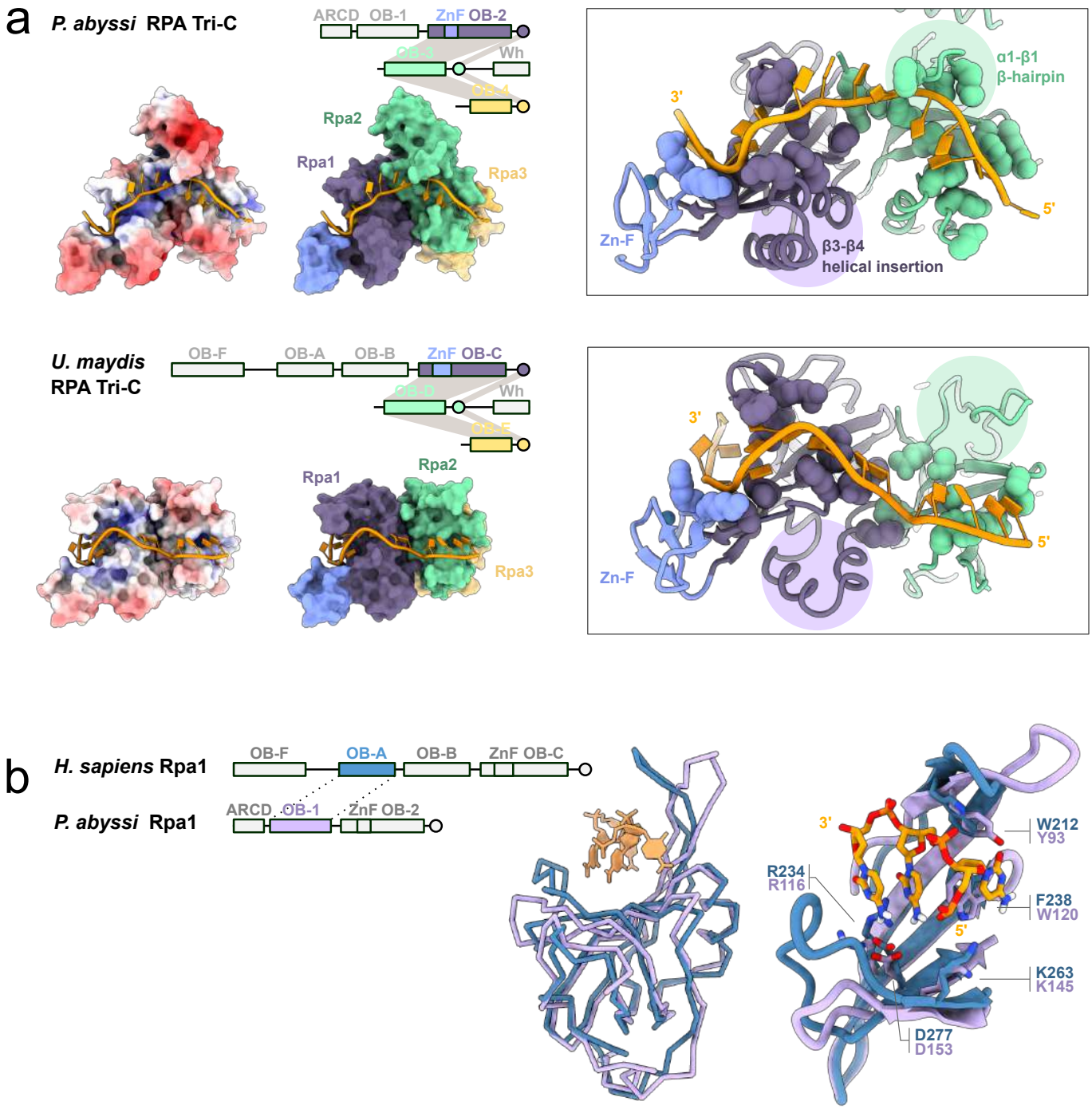



Figure S5: Comparison of archaeal and eukaryotic RPA ssDNA binding determinants. a. Comparison of DNA-bound P. abyssi and U. maydis (4GOP) Tri-C structures with focused views on Rpa1 (purple and blue) and Rpa2 (green). ssDNA-contacting residues are shown as spheres. b. Superimposition of the P. abyssi Rpa1 OB-1 (in purple) with the H. sapiens Rpa1 OB-A (in blue) (1JMC) crystal structures reveals conserved ssDNA-binding amino-acids.

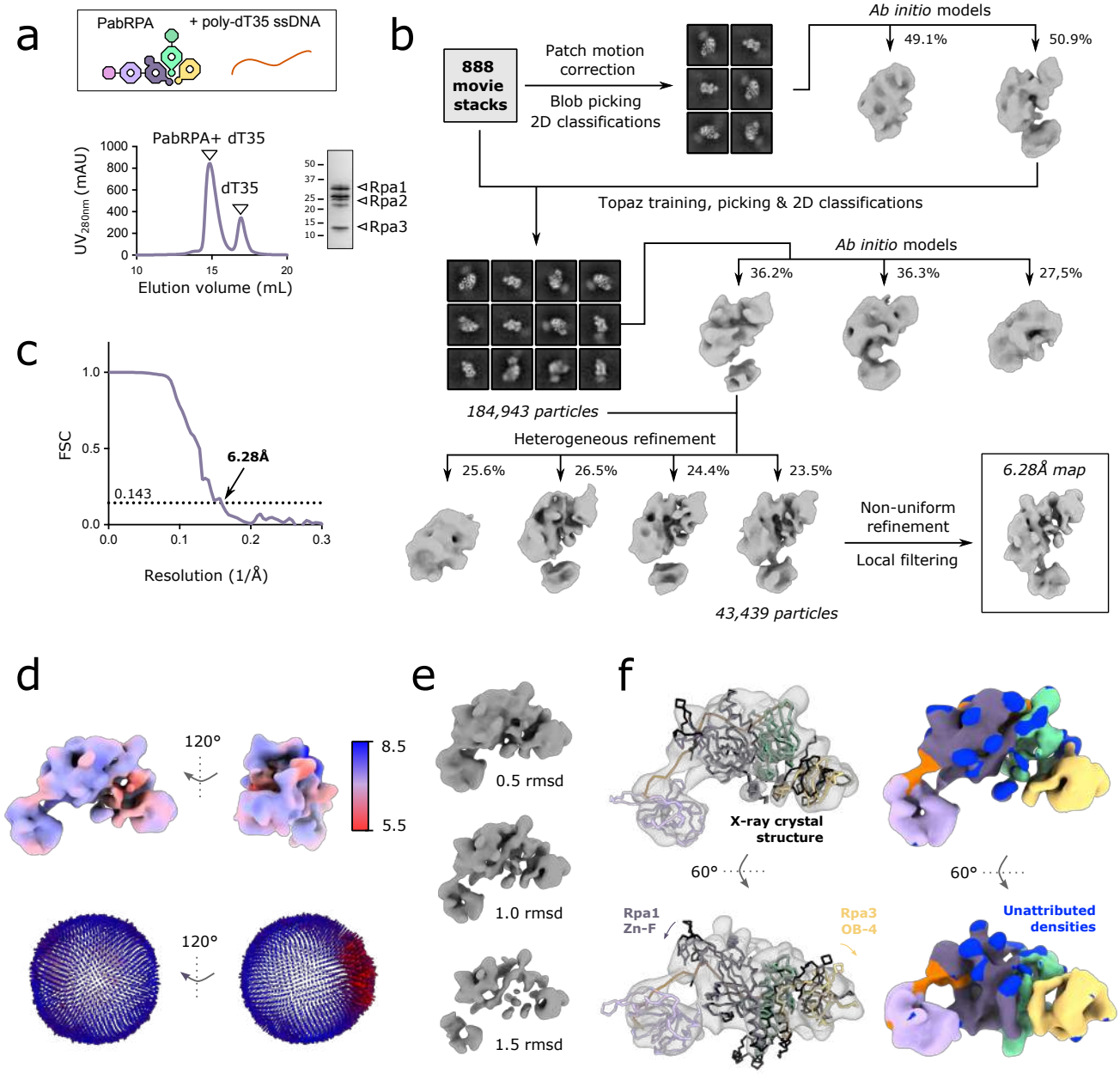



Figure S6: Cryo-EM structure determination of PabRPA bound to a short ssDNA. a. Size-exclusion chromatography (SEC) profile (left) and SDS-PAGE analysis (right) of PabRPA bound to a poly-dT35 ssDNA b. Data processing workflow (detailed in methods). c. Resolution of final reconstructions determined by gold standard FSC at the 0.143 criterion. d. Local resolution evaluations of the final reconstructions (top) and Euler angle distributions of particles used for the final reconstructions (bottom). e. Cryo-EM maps at varying contour levels. f. Comparison of the DNA-bound RPA cryo-EM and crystal structure (left) and cryo-EM map colored by RPA subunits (right).

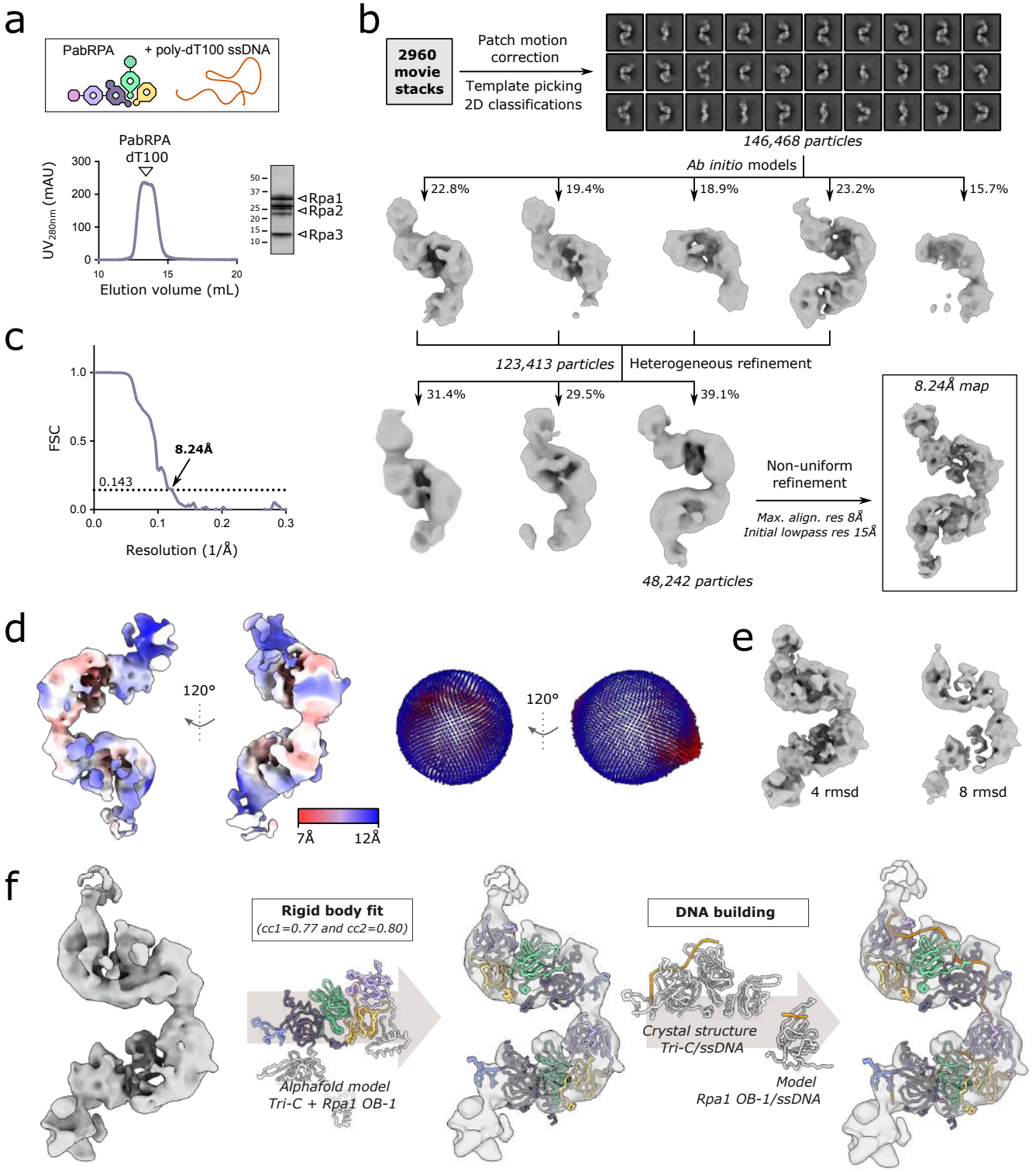



Figure S7: Cryo-EM structure determination of PabRPA bound to a long ssDNA. a. Size-exclusion chromatography (SEC) profile (left) and SDS-PAGE analysis (right) of PabRPA bound to a poly-dT100 ssDNA b. Data processing workflow (detailed in methods). c. Resolution of final reconstructions determined by gold standard FSC at the 0.143 criterion. d. Local resolution evaluations of the final reconstructions (left) and Euler angle distributions of particles used for the final reconstructions (right). e. Cryo-EM maps at varying contour levels f. Model building workflow.

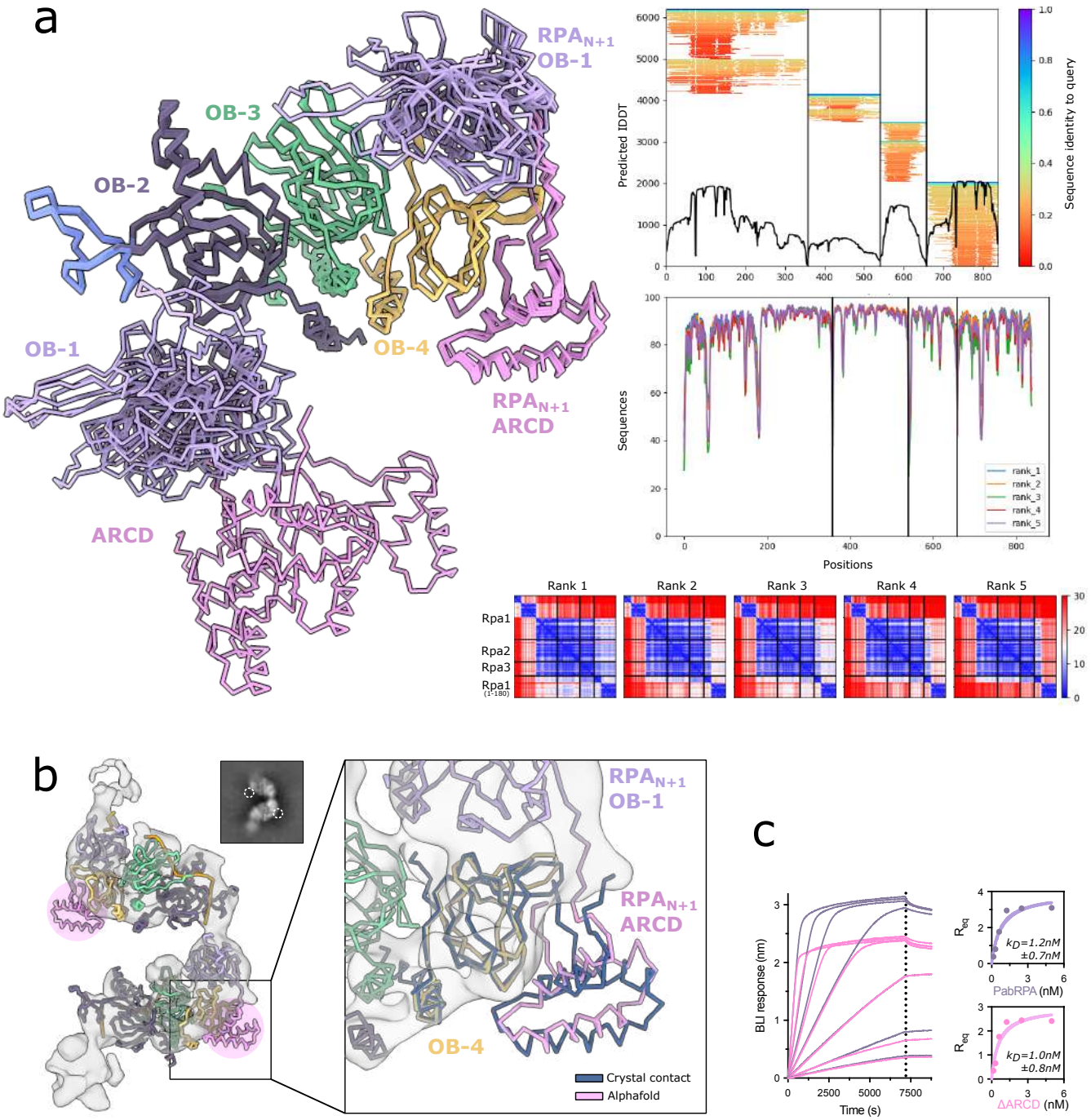
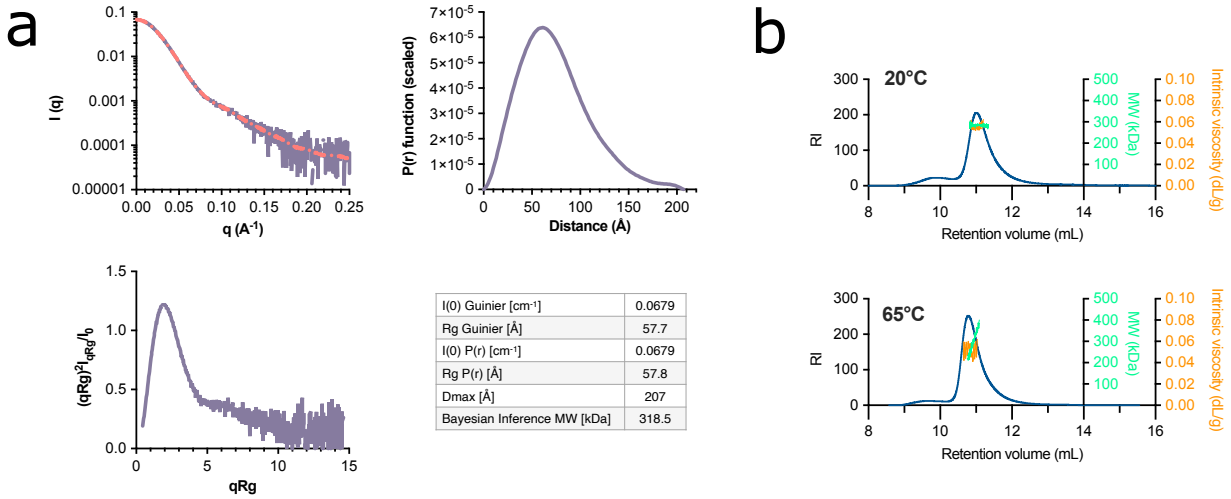



Figure S8: ARCD is required for early-steps of the ssDNA-RPA complex assembly and becomes flexible after binding to ssDNA. a. Left: superposition of the five AlphaFold-Multimer predicted models (Evans et al., 2021) for the interaction of one PabRPA-ΔWH heterotrimer with two first domains of PabRPA1 (ARCD-OB1, 1-180). Right: AlphaFold generated plots of MSA sequence coverage (top), of predicted LDDT vs position along the protein sequence (middle) and of predicted aligned error (bottom, where each protein chain of the complexe is plotted against itself and against the other chain). b. On left, fit of the RPA(n)/ARCD-OB-1(n+1) AlphaFold model containing the ARCD domain in the cryo-EM map. The ARCD domain is not observed either in the map or in the 2D classes. On right, superimposition with the apo-PabRPA crystal structure. The lattice contact between ARCD and OB-4 fits with the AlphaFold prediction. c. Specific binding of PabRPA (in purple) and ΔARCD mutant (in pink) to immobilized poly-dT35 ssDNA measured by biolayer interferometry. Steady-state analysis were performed using the average signal measured at the end of the association steps.




| | |
|---|---|
| I(0) Guinier [cm-1] | 0.0679 |
| Rg Guinier [Å] | 57.7 |
| I(0) P(r) [cm-1] | 0.0679 |
| Rg P(r) [Å] | 57.8 |
| Dmax [Å] | 207 |
| Bayesian Inference MW [kDa] | 318.5 |

Figure S9: PabRPA oligomerization state in solution. a. In solution study of PabRPA structure by SAXS. On top left, SAXS profile of PabRPA (in purple) with superimposition of fitted scattering pattern that allowed the generation of the distance distribution function P(r) (red dotted line). On right, distance distribution functions P(r) derived from the scattering patterns of PabRPA obtained using the program GNOM. On bottom left, dimensionless Kratky plot scattering. Scattering derived parameters are provided in the table. b. PabRPA molecular mass (Green) and intrinsic viscosity (orange) measurements by SEC-SLS coupled with an inline viscometer. Measurements were performed at 20°C (higher panel) and 65°C (lower panel).

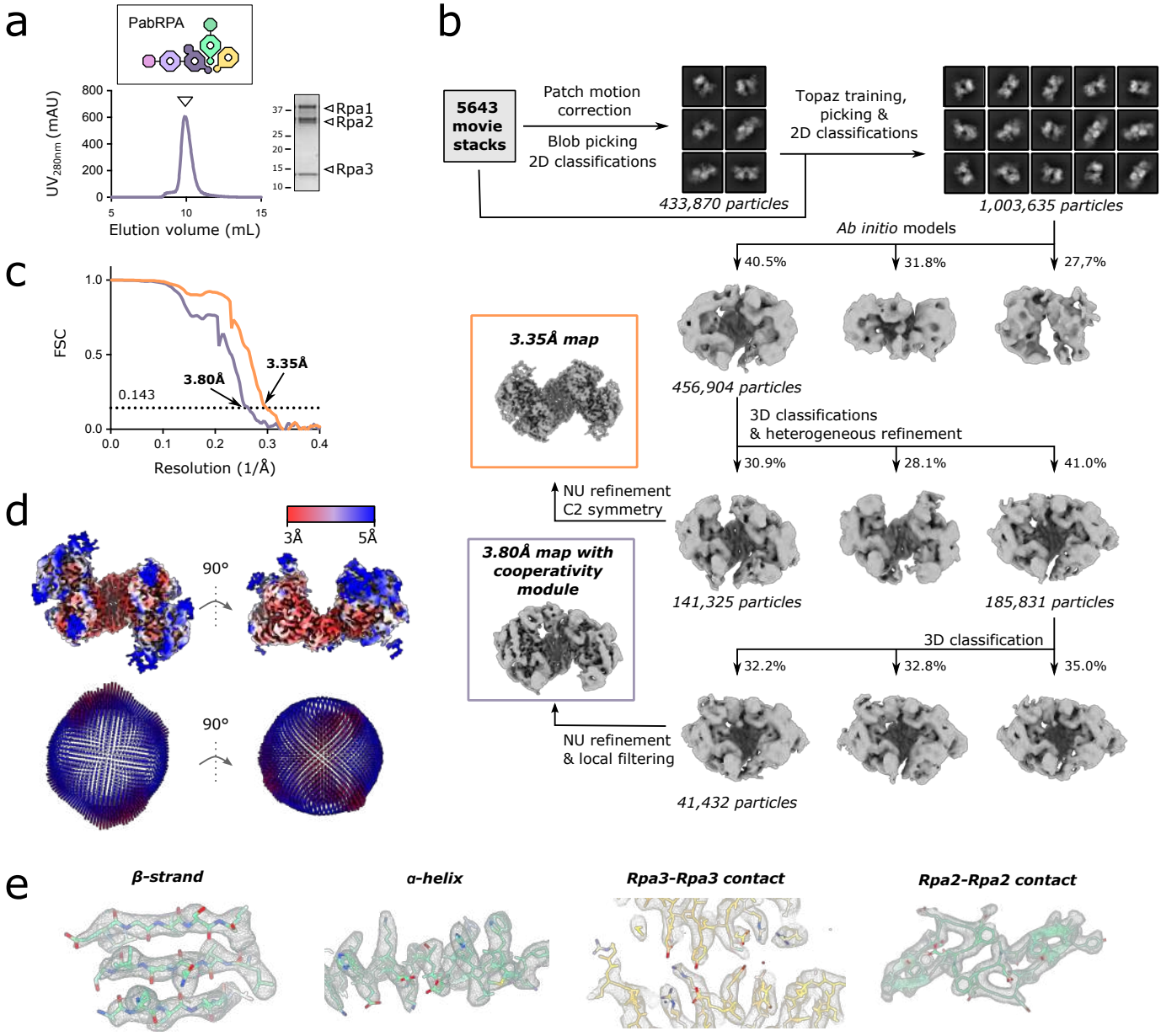



Figure S10: Cryo-EM structure determination of the tetrameric RPA super-structure. a. Size-exclusion chromatography (SEC) profile (left) and SDS-PAGE analysis (right) of PabRPA b. Data processing workflow (detailed in methods). c. Resolution of final reconstructions determined by gold standard FSC at the 0.143 criterion. d. Local resolution evaluations of the final reconstructions (top) and Euler angle distributions of particles used for the final reconstructions (bottom). e. Representative views of the cryo-EM map showing secondary structure elements and key regions of the tetrameric RPA assembly.

| | PabRpa1 cooperativity module (8AA9) | PabRPA (8AAJ) | PabRPA Tri-C/d20T (8AAS) |
|---|---|---|---|
| **Data collection** | | | |
| Space group | *P* $2_1$ | *C* 2 2 $2_1$ | *P* $3_2$ 2 1 |
| Cell dimensions | | | |
| *a*, *b*, *c* (Å) | 52.75, 83.47, 56.19 | 116.53, 169.74, 201.86 | 75.23, 75.23, 215.49 |
| α, β, γ (°) | 90, 115.3, 90 | 90, 90, 90 | 90, 90, 120 |
| Wavelength (Å) | 0.97857 | 0.97857 | 0.97856 |
| Resolution (Å) | 50.0 - 1.8 (1.85 - 1.8) | 49.4 - 3.70 (3.80 - 3.70) | 48.26 - 3.2 (3.28 - 3.2) |
| Estimated resolution limit (Å)* | 1.76, 1.76, 1.91 | 5.46, 4.21, 3.26 | 3.52, 3.52, 3.02 |
| $R_{pim}$ | 0.025 (0.986) | 0.024 (20.52) | 0.016 (2.752) |
| $R_{merge}$ | 0.061 (2.43) | 0.123 (89.34) | 0.080 (14.31) |
| $I / \sigma I$ | 14.6 (0.8) | 13.0 (0.1) | 22.1 (0.3) |
| $CC_{1/2}$ | 0.999 (0.288) | 1.000 (0.130) | 0.999 (0.223) |
| Completeness (%) | 99.9 (99.5) | 99.8 (99.8) | 99.9 (99.9) |
| Redundancy | 7.0 (6.9) | 28.2 (28.4) | 27.1 (27.5) |
| $R_{merge}$* | 0.054 (1) | 0.082 (4.31) | 0.067 (4.19) |
| $I / \sigma I$* | 14.6 (1.2) | 21.4 (1.0) | 27.1 (1.0) |
| $CC_{1/2}$* | 0.999 (0.589) | 1.000 (0.542) | 1.000 (0.454) |
| Completeness (%)* | 86.9 (50.5) | 60.1 (13.3) | 80.7 (21.2) |
| **Refinement** | | | |
| Resolution (Å) | 32.25 - 1.80 (1.85 - 1.80) | 49.4 - 3.70 (3.80 - 3.70) | 44.84 - 3.2 (3.28 - 3.2) |
| No. reflections | 37716 (1857) | 13191 (1036) | 37624 (753) |
| $R_{work}$ / $R_{free}$ (%) | 20.86/22.84 | 23.41/25.09 | 26.28/28.90 |
| No. atoms | | | |
| Protein | 2860 | 5304 | 6372 |
| DNA | - | - | |
| Ligand/ion | 2 | 56 | 258 |
| Water | 290 | - | 335 |
| *B*-factors | | | |
| Protein | 38.6 | 61.9 | 51.04 |
| Ligand/ion | 76.1 | 44.13 | 40.68 |
| Water | 50.8 | 39.49 | 51.69 |
| R.m.s. deviations | | | |
| Bond lengths (Å) | 0.008 | 0.010 | 0.008 |
| Bond angles (□) | 0.95 | 1.45 | 1.18 |

Table S1: X-ray diffraction data collection and refinement statistics. Values in parentheses are for highest-resolution shell. Dataset from single crystal used per structure. *Values calculated after truncation by STARANISO. Estimated resolution limits along the three crystallographic directions a*, b*, c*.

| | PabRPA/dT35 (EMDB-15347) (PDB 8ACE) | PabRPA/dT100 (EMDB-15350) (PDB 8ACJ) | PabRPA tetramer (EMDB-15300) (PDB 8AAH) | PabRPA tetramer with ARCD-OB-1 (EMDB-15301) (PDB 8AAL) |
|---|---|---|---|---|
| **Data collection and processing** | | | | |
| Magnification | 150000 | 150000 | 150000 | 150000 |
| Voltage (kV) | 200 | 200 | 300 | 300 |
| Electron exposure ($e^-/Å^2$) | 40 | 40 | 40 | 40 |
| Defocus range (μm) | -0.8 to –2.8 | -0.8 to –2.8 | -0.8 to –2.8 | -0.8 to –2.8 |
| Pixel size (Å) | 0.96 | 0.96 | 0.86 | 0.86 |
| Symmetry imposed | - | - | C2 | - |
| Initial particle images (no.) | 208094 | 847605 | 1761244 | 1761244 |
| Final particle images (no.) | 43439 | 146468 | 1003635 | 41432 |
| Map resolution (Å) | 6.28 | 8.24 | 3.35 | 3.8 |
| FSC threshold | 0.143 | 0.143 | 0.143 | 0.143 |
| Map resolution range (Å) | 5.5 - 8.5 | 7 - 12 | 3 - 5 | 3 - 5 |
| | | | | |
| **Refinement** | | | | |
| Initial model used (PDB) | 8AA9, 8AAS | 8AA9, 8AAS | 8AAJ | 8AA9, 8AAJ |
| Model resolution (Å)<br>FSC threshold (0.143) | 8.3 | 10.4 | 3.3 | 3.7 |
| Map sharpening *B* factor ($Å^2$) | 413.2 | 742.8 | 130.0 | 82.4 |
| Model composition | | | | |
| Non-hydrogen atoms | 4632 | 9428 | 15198 | 16616 |
| Protein residues | 521 | 1048 | 1868 | 2044 |
| Nucleotides | 21 | 46 | | |
| Ligands | 1 | 2 | 4 | 4 |
| *Mean B* factors ($Å^2$) | | | | |
| Protein | 777.3 | 620.3 | 45.7 | 278.2 |
| Nucleotides | 810.0 | 586.4 | | |
| Ligand | 828.3 | 386.2 | 107.3 | 291.1 |
| R.m.s. deviations | | | | |
| Bond lengths (Å) | 0.025 | 0.030 | 0.004 | 0.004 |
| Bond angles (°) | 1.945 | 2.196 | 0.685 | 0.697 |
| Validation | | | | |
| MolProbity score | 2.41 | 2.18 | 1.66 | 1.75 |
| Clashscore | 31.27 | 50.46 | 7.98 | 11.34 |
| Poor rotamers (%) | 1.74 | 0.65 | 0 | 0 |
| Ramachandran plot | | | | |
| Favored (%) | 96.28 | 98.26 | 96.52 | 96.92 |
| Allowed (%) | 3.33 | 1.36 | 3.26 | 2.8 |
| Disallowed (%) | 0.39 | 0.39 | 0.22 | 0.20 |

Table S2: Cryo-EM data collection, refinement and validation statistics.