nature communications

Article

https://doi.org/10.1038/s41467-024-54266-2

Architecture of the spinach plastid-encoded RNA polymerase

Received: 30 January 2024

Accepted: 5 November 2024

Published online: 13 November 2024

Check for updates

Tongtong Wang[1,4], Guang-Lei Wang[2,3,4], Ying Fang[1,4], Yi Zhang[1,4], Wenxin Peng[1], Yue Zhou[1], Aihong Zhang[1], Long-Jiang Yu[2,3] ✉ & Congming Lu[1] ✉

The plastid-encoded RNA polymerase serves as the principal transcription machinery within chloroplasts, transcribing over 80% of all primary plastid transcripts. This polymerase consists of a prokaryotic-like core enzyme known as the plastid-encoded RNA polymerase core, and is supplemented by newly evolved associated proteins known as PAPs. However, the architecture of the plastid-encoded RNA polymerase and the possible functions of PAPs remain unknown. Here, we present the cryo-electron microscopy structure of a 19-subunit plastid-encoded RNA polymerase complex derived from spinach (*Spinacia oleracea*). The structure shows that the plastid-encoded RNA polymerase core resembles bacterial RNA polymerase. Twelve PAPs and two additional proteins (FLN2 and pTAC18) bind at the periphery of the plastid-encoded RNA polymerase core, forming extensive interactions that may facilitate complex assembly and stability. PAPs may also protect the complex against oxidative damage and has potential functions in transcriptional regulation. This research offers a structural basis for future investigations into the functions and regulatory mechanisms governing the transcription of plastid genes.

Chloroplasts are the photosynthetic organelles in green eukaryotic organisms and play a crucial role in capturing sunlight and converting it into energy. They possess their own genome, a remnant of an endosymbiotic event between a eukaryotic cell and a photosynthetic cyanobacterium. The residual chloroplast genome is relatively small, encoding only 75–85 of the ~3500–4000 proteins present in a typical chloroplast[1,2]. However, the proper expression of chloroplast genes is vital for chloroplast biogenesis, as well plant growth and development[3,4].

Chloroplast genes are transcribed by two distinct types of RNA polymerases: nucleus-encoded polymerase (NEP), resembling the T3-T7 phage-type RNA polymerase, and plastid-encoded polymerase (PEP), a bacterial-type multisubunit polymerase[5,6]. PEP represents the primary transcription machinery in chloroplasts and predominantly transcribes genes related to photosynthesis[7,8].

In algae and land plants, PEP exhibits a bacterial-like RNA polymerase origin, with a catalytic core inherited from its cyanobacterial endosymbiotic ancestor. This core consists of two α subunits, one β subunit, one β' subunit, and one β'' subunit (2αββ'β'') encoded by the plastid genes *rpoA*, *rpoB*, *rpoC1*, and *rpoC2*, respectively[4,9]. In tobacco (*Nicotiana tabacum*), knockout mutant plants deficient in any of these Rpo proteins exhibit albino or yellowish phenotypes and seedling death due to impaired chloroplast development, underscoring the indispensable role of Rpo subunits in chloroplast biogenesis.

Despite genetic evidence pointing toward a prokaryotic structure of the PEP core[9], the biochemical purification of PEP from plant chloroplasts has revealed many more subunits than initially thought[10]. Indeed, twelve distinct PEP-associated proteins (PAP1–PAP12) and two additional proteins (FRUCTOKINASE-LIKE 2 [FLN2] and PLASTID

[1]State Key Laboratory of Wheat Improvement, College of Life Sciences, Shandong Agricultural University, Taian, Shandong 271018, China. [2]Photosynthesis Research Center, Key Laboratory of Photobiology, Institute of Botany, Chinese Academy of Sciences, 100093 Beijing, China. [3]University of Chinese Academy of Sciences, 100049 Beijing, China. [4]These authors contributed equally: Tongtong Wang, Guang-Lei Wang, Ying Fang, Yi Zhang.
✉e-mail: longer@ibcas.ac.cn; cmlu@sdau.edu.cn

TRANSCRIPTIONALLY ACTIVE 18 [pTAC18]) have been identified as tightly associated with the PEP core[9,11]. Genetic studies have also demonstrated that loss of function of any of these PAP proteins often leads to albino or yellowish phenotypes and severely impairs PEP-mediated transcription, highlighting that these PAPs are essential for PEP activity, revealing the intricate organization of the PEP complex, and suggesting that PAPs have essential structural roles in the complex[12–25].

Although the composition of the PEP complex has been revealed, its precise 3D architecture and the possible functions of PAPs remain unknown. Here, we present the cryo-electron microscopy (cryo-EM) structure of the PEP complex isolated from spinach (*Spinacia oleracea*). Our results reveal structural insights into the chloroplast transcription apparatus and provide a structural basis for further studies on the molecular mechanisms and functions of transcription in chloroplasts.

Results

Overall structure of the PEP complex

We purified the PEP complex from spinach leaves (Supplementary Fig. 1a–d). The purified PEP complex was able to extend RNA in an in vitro transcription elongation reaction, indicating that it is catalytically active (Supplementary Fig. 1e). Mass spectrometry analysis confirmed the presence of four polymerase core subunits (α, β, β', and β'') and 14 proteins associated with the PEP core (PAP1–PAP12, FLN2, and pTAC18)[9,11] (Supplementary Fig. 1d and Supplementary Table 1). Our structural analysis of PEP described below revealed that FLN2 and pTAC18 are stably and uniformly associated with PEP. We thus designated FLN2 and pTAC18 as PAP13 and PAP14, respectively.

We determined the structure of the PEP complex by single-particle cryo-EM, yielding a 3D reconstruction of the complex at an overall resolution of 3.16 Å (Fig. 1, Supplementary Fig. 2a–g and Supplementary Table 2). The maps were of sufficient quality to allow building and refinement of an almost complete model of the PEP complex with the assistance of AlphaFold2[26,27], revealing the features of the complex at molecular detail (Fig. 1 and Supplementary Table 3). The model consisted of a 19-subunit PEP complex, comprising five subunits of the PEP core (two copies of α labeled α1 and α2, β, β', and β'') and 14 PAP subunits (PAP1, PAP3–PAP9, two copies of PAP10 [also reported as thioredoxin Z, Trx Z] labeled as $PAP10_1$ and $PAP10_2$, and PAP11–14) (Fig. 1, Supplementary Fig. 3 and Supplementary Table 3). Thus, the model contains all previously identified essential PEP subunits except PAP2[10,11].

Structure of the PEP core

Chloroplasts originated ~1.5 billion years ago when an ancient cyanobacterium was engulfed by a eukaryotic host cell. A typical bacterial RNA polymerase (RNAP) has five core subunits: 2α subunits, β, β', and ω. The RNAP β' subunit is split into two proteins corresponding to the N-terminal region (γ subunit encoded by *rpoC1*) and the C-terminal region (β' subunit encoded by *rpoC2*) of the equivalent bacterial β' subunit in cyanobacteria. Thus, cyanobacterial RNAP consists of six subunits: 2α subunits, β, γ, β', and ω. The split of RNAP β' subunit also occurs in the PEP core, with the *rpoC* gene that encodes RNAP β' subunit in *E. coli* being split in spinach into two genes, *rpoC1* and *rpoC2*, that encode the β' and β'' subunits of PEP, respectively. This split does not result in significant structural differences between the PEP core and bacterial RNAPs (Supplementary Fig. 4a, b)[28,29], as the PEP core subunits are similar in sequence to their bacterial RNAP counterparts (Supplementary Fig. 5) and the structures of the PEP core and bacterial RNAP are also similar (Fig. 2a–c).

The structure of the PEP core resembles the structures of cyanobacterial RNAP (PDB 8GZG)[29] and *E. coli* RNAP (PDB 6GH5)[30] (Fig. 2a–c)[31–33]. The PEP core has a "crab claw" appearance with two arms: a lower arm (the lobe-protrusion-claw) and an upper arm (the clamp-claw). The PEP core and cyanobacterial RNAP show highly similar folds in their domains, such as the clamp, lobe, protrusion, and flap, which together form the DNA and RNA channels (Fig. 2d–g). The active site is formed by a conserved domain in the β' subunit (residues 378–519) and consists of three conserved aspartate residues (D489, D491, and D493). The PEP core and cyanobacterial RNAP show essentially the same folds for key motifs in the active site, such as the catalytic loop, the trigger loop (TL), the bridge helix (BH), and the β'' rim helices (β''-RH), involved in nucleotide triphosphate (NTP) incorporation (Fig. 2h). In addition, the lid, rudder, and fork-loop that are essential elements for catalytic activity are conserved between the PEP core and cyanobacterial RNAP (Fig. 2e, f). The gene encoding the ω subunit was considered to have been lost from the plastid genome during plant evolution[9]. PAP12 in land plants shows low sequence similarity to the cyanobacterial RNAP subunit ω (Supplementary Fig. 4d). Intriguingly, the location and structure of PAP12 within the PEP core are similar to those of the ω subunit in cyanobacterial RNAP (Fig. 2a, b and Supplementary Fig. 4c), Thus, we define PAP12 as the ω subunit of the PEP core. These structural comparisons indicate that the overall structure of the PEP core is highly structurally similar to cyanobacterial RNAP.

However, the PEP core does show some distinct folds unlike those of bacterial RNAPs, providing the binding sites for PAPs. One distinct feature is that the β'' subunit contains a 790-residue sequence, referred to as insertion 3 (termed SI3; residues 348-1138), in the middle of the TL that forms a key mobile element within the active site of RNAP[34–37]. SI3 is composed of repeats of the conserved sandwich-barrel-hybrid motif (SBHM), with various repeat numbers in different species[29,33,38]. *E. coli* SI3 contains two SBHMs, whereas cyanobacterial SI3 has nine. PEP SI3 in plants also has nine SBHMs but is ~160 residues larger than cyanobacterial SI3. Cyanobacterial SI3 consists of a head, a body, a fin, and a tail, which together form a seahorse-shaped arch. The tail and fin form the base of cyanobacterial SI3, while the head and two adjacent SBHMs of the body comprise the tip[29,39].

We detected densities for four ordered domains within PEP SI3 (Fig. 3a, b). They correspond to the tail, the fin, part of the body, and the head of cyanobacterial SI3, respectively[29,39]. PEP SI3 forms an arch that is similar to that of cyanobacterial SI3 (Fig. 3a, b). The tail associates with β''-RH, near the secondary channel of RNAP, and contacts PAP5 and PAP8. The fin also associates with β''-RH. Unlike the fin of cyanobacterial SI3, the fin of PEP SI3 (designated SI3-fin hereafter) contains an additional 96-residue insertion (residues 962–1058), resulting in a larger SI3 fin that encloses the PAP3 S1 domain and interacts with PAP4 and PAP6. The part of the body adjacent to the head (residues 558–618 and 793–877) adopts orientations distinct from those of cyanobacterial SI3. The remaining parts of the PEP SI3 body are largely flexible, resulting in an unresolved structure. The base and tip of the PEP SI3 are instead bridged by a complex formed by PAP3, PAP14, PAP4, and PAP9, which may be required to stabilize the SI3 arch.

β'' also presents additional specific features unlike those of the cyanobacterial RNAP β'' subunit. The jaw domain that is conserved in cyanobacterial and bacterial RNAP β'' subunits is absent in the PEP β'' subunit, and the corresponding space is instead occupied by a PAP1-unique domain (Fig. 3c, d). In addition, a βαβ motif (residues 962–1,058) adjacent to the secondary channel shows a different conformation from that in cyanobacterial RNAP and provides binding sites for PAP7 and PAP8 (Fig. 3c, e).

β' displays some specific folds not observed in cyanobacterial RNAP, which may also provide binding sites for PAPs (Fig. 4). A sequence insertion in β' (β'-SI1; residues 545–581) forms a specific disordered loop near the C terminus of β'. The β'-blade shows a structural difference from the cyanobacterial RNAP. The β'-SI1 and the β'-blade are connected to each other and embrace the C-terminal helix of PAP8. The β'-SI1 also contacts PAP5 (Fig. 4a, b). β' contains clamp helices (Fig. 4a, c), designated as β'CH in bacterial RNAP[40]. β'CH is

https://doi.org/10.1038/s41467-024-54266-2

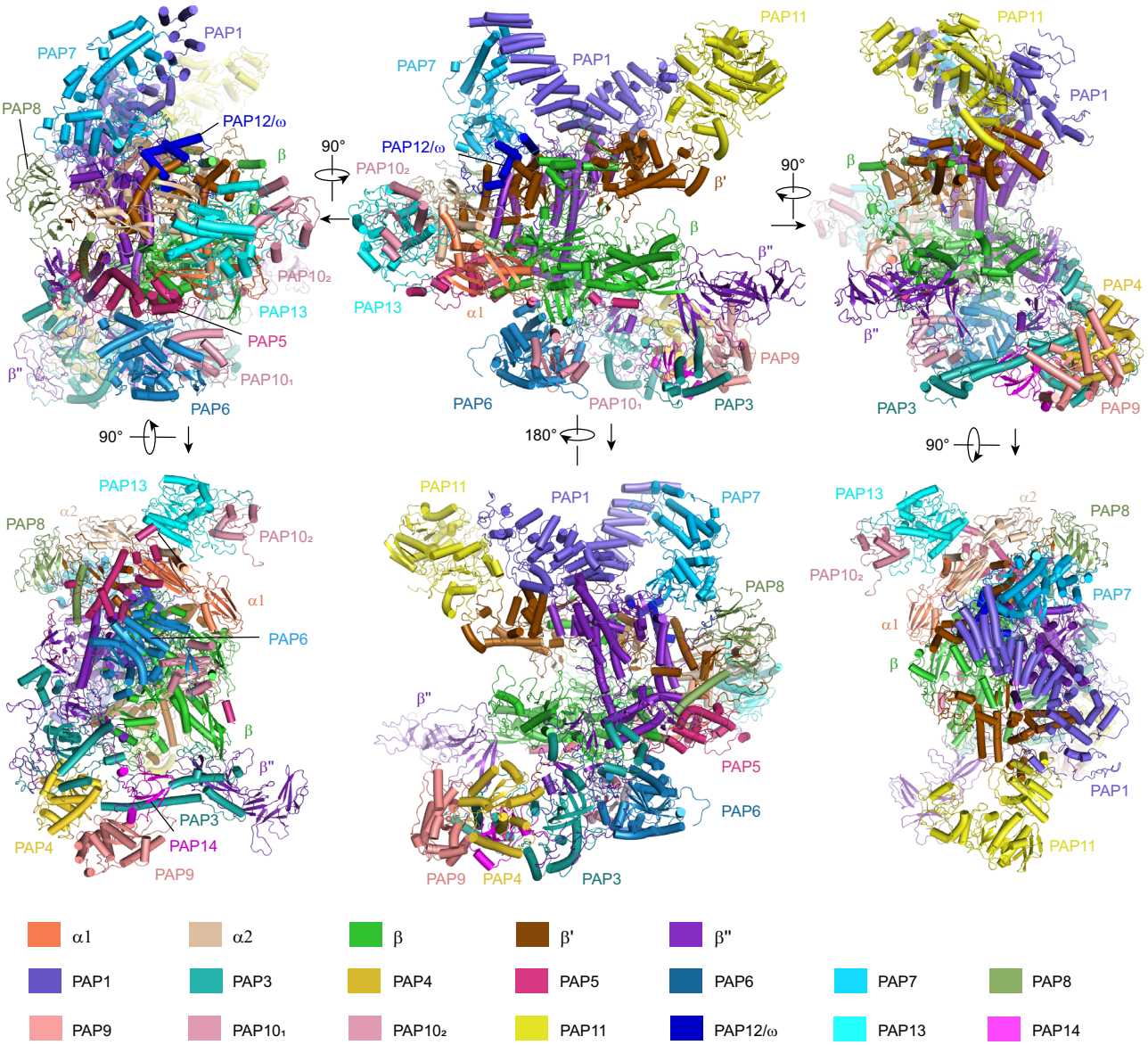


Fig. 1 | Overall architecture of the *Spinacia oleracea* (*Sp.*) PEP complex. Different views show the overall arrangement of the subunits in the *Sp.* PEP complex. The subunits are shown as cartoons and colored individually as indicated. α, RNA polymerase α subunit; β, RNA polymerase β subunit; β', RNA polymerase β' subunit; β", RNA polymerase β" subunit; PAP1, plastid transcriptionally active chromosome protein 3 (pTAC3); PAP3, plastid transcriptionally active chromosome protein 10 (pTAC10); PAP4, iron superoxide dismutase 3 (FSD3); PAP5, plastid transcriptionally active chromosome protein 12 (pTAC12/HEMERA); PAP6, fructokinase-like protein 1 (FLN1); PAP7, plastid transcriptionally active chromosome protein 14 (pTAC14); PAP8, plastid transcriptionally active chromosome protein 6 (pTAC6); PAP9, iron superoxide dismutase 2 (FSD2); PAP10, thioredoxin Z (Trx Z); PAP11, UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-2,6-diaminopimelate ligase MurE homolog; PAP12/ω, plastid transcriptionally active chromosome protein 7 (pTAC7); PAP13, fructokinase-like protein 2 (FLN2); PAP14, plastid transcriptionally active chromosome protein 18 (pTAC18). Among these subunits, the two copies of the α subunit are labeled as α1 and α2 and given different colors, while the two copies of the PAP10 subunit are labeled as $PAP10_1$ and $PAP10_2$ and colored the same due to their different locations.

extended compared to cyanobacterial RNAP, with two more helical turns due to a 14-residue insertion, which provides an interface for PAP1. The N terminus of β' has a distinct structure that extends towards PAP1 (Fig. 4a, c). Notably, the toe of the β' clamp has a flexible structure unlike that in cyanobacterial RNAP (Fig. 4a, d). PAP11 interacts with the clamp-toe, possibly stabilizing this flexible structure.

β contains two disordered loops, loop 1 and loop 2, which form a closed state in cyanobacterial RNAP but adopt an open state to embrace PAP6 in PEP (Fig. 4e, f).

Arrangement of PAPs in the PEP complex

Our structure reveals the detailed arrangement of PAPs peripherally associated with the PEP core. In the overall structure, the PAPs occupy one of three regions: two regions surrounding the two arms of the PEP core, forming a larger clamp-like structure, and one region at the intersection of the two arms, near the α dimer (Fig. 5a, b).

The PAPs interact extensively with the PEP core and with each other (Fig. 5c). We grouped the 13 PAPs into five clusters according to their location and potential functions (Fig. 5a–c). The first cluster consists of PAP1, PAP7, and PAP11, which are all located on the upper arm. The second cluster contains PAP5 and PAP8, which are positioned at the intersection of the two arms, near the α dimer. The third cluster comprises PAP3 and PAP14, which are situated in the lower arm and protrude outward from the PEP core. The fourth cluster is a heterodimer formed by PAP4 and PAP9, which is docked on PAP3. The fifth cluster is formed by two heterodimers, PAP6–$PAP10_1$ and

https://doi.org/10.1038/s41467-024-54266-2

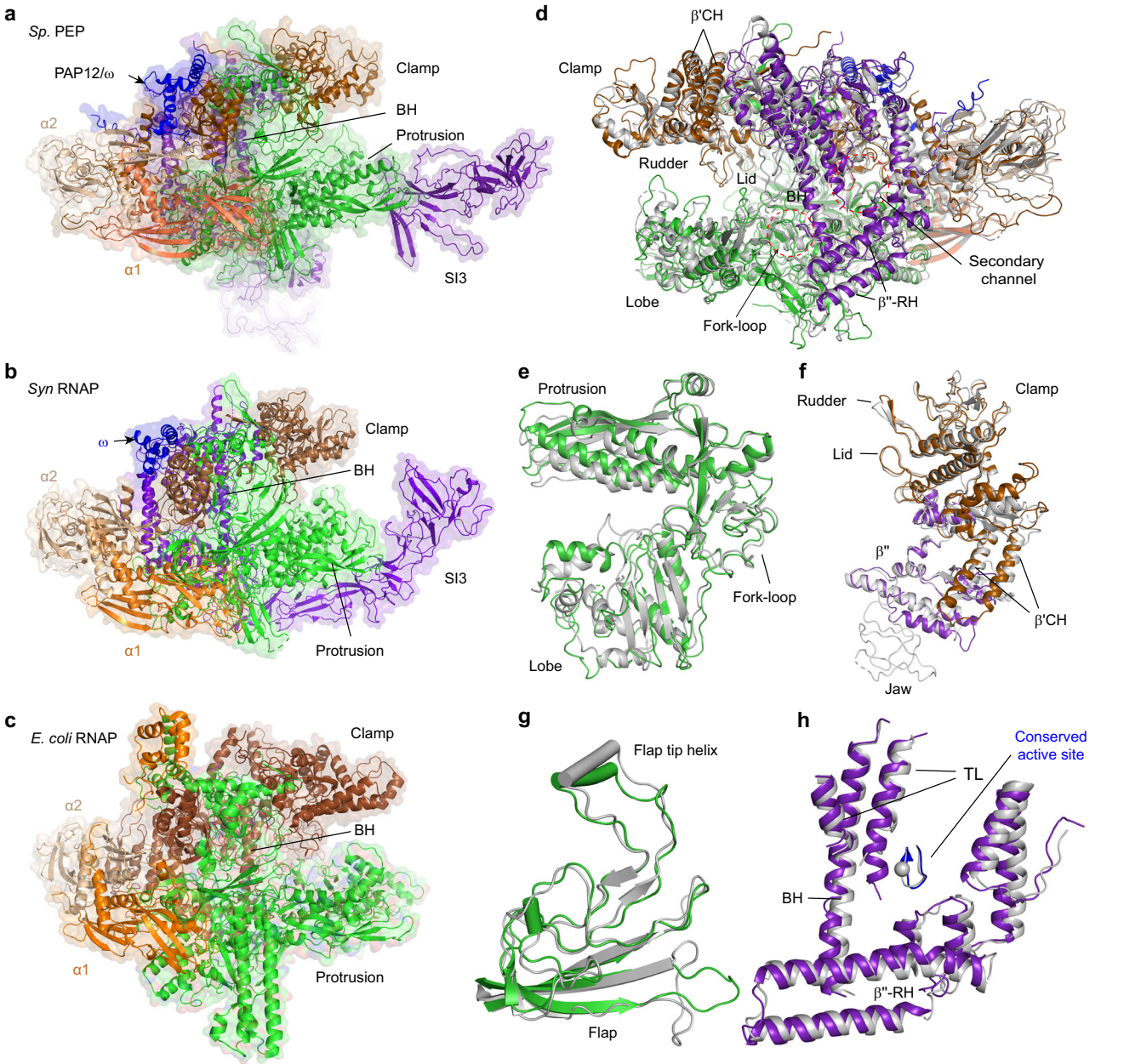


Fig. 2 | Comparison of the *Spinacia oleracea* (*Sp.*) PEP core RNA polymerase and bacterial RNAP. Overall structural similarity between the *Sp.* PEP core polymerase (**a**), *Synechocystis* sp. PCC 6803 (*Syn*) RNAP (PDB 8GZG)[29] (**b**) and *E. coli* RNAP (PDB 6GH5)[30] (**c**). **d** Superimposition of the structures for the *Sp.* PEP core polymerase and *Syn* RNAP (root mean square deviation (r.m.s.d.) of 1.907 Å for Cα atoms by PyMOL) showing that the *Sp.* PEP core polymerase adopts the same folds as *Syn* RNAP. **e**–**g** Structural comparison of the key structural domains (clamp, protrusion, lobe, flap, and jaw) between the *Sp.* PEP core polymerase and *Syn* RNAP. **h** Structural comparison of the key structural motifs in the active site (catalytic loop, trigger loop, bridge helix, and rim helices) between the *Sp.* PEP core polymerase and *Syn* RNAP. α1, α2, RNA polymerase α subunit; β, RNA polymerase β subunit; β', RNA polymerase β' subunit; β'', RNA polymerase β'' subunit; PAP12/ω, plastid transcriptionally active chromosome protein 7 (pTAC7); TL, trigger loop; BH, bridge helix; β''-RH, β'' rim helices; β'CH, clamp helices; SI3, the β'' arch domain of sequence insertion 3.

PAP13–PAP10$_2$, which are positioned on the lower arm and on the α dimer, respectively.

PAP1, PAP7, and PAP11 embrace the clamp-claw

Cluster 1, comprising PAP1, PAP7, and PAP11, is formed by their structural tethering and is situated along the side of the clamp-claw (Fig. 6a). PAP1 contains pentatricopeptide repeats (PPRs), with nine N-terminal PPRs (PAP1-PPR$_N$) and two complete PPRs and one partial PPR at its C terminus (PAP1-PPR$_C$). Although PAP1 is situated on the side of the RNA exit channel, it is not directly adjacent to the RNA exit channel, and it does not display a pronounced positively charged surface, as would be expected for a nucleic acid–binding protein (Supplementary Fig. 6a). Thus, PAP1 may not directly interact with RNA during transcription. PAP1-PPR$_N$ contacts the termini of β, β', and β'' and these termini show changed conformations unlike those seen for the equivalent subunit in cyanobacterial RNAP (Figs. 4c, 6b), suggesting that PAP1-PPR$_N$ may stabilize the PEP core. Located between PAP1-PPR$_N$ and PAP1-PPR$_C$, PAP1 contains a unique domain (residues 394–782) that binds PAP11, the β'-clamp including β'CH, and β'' of the PEP core (Fig. 6a, c). However, large parts of this unique domain are not visible in our reconstruction, suggesting that they are mobile. Notably, the ordered parts of this unique domain wrap

https://doi.org/10.1038/s41467-024-54266-2

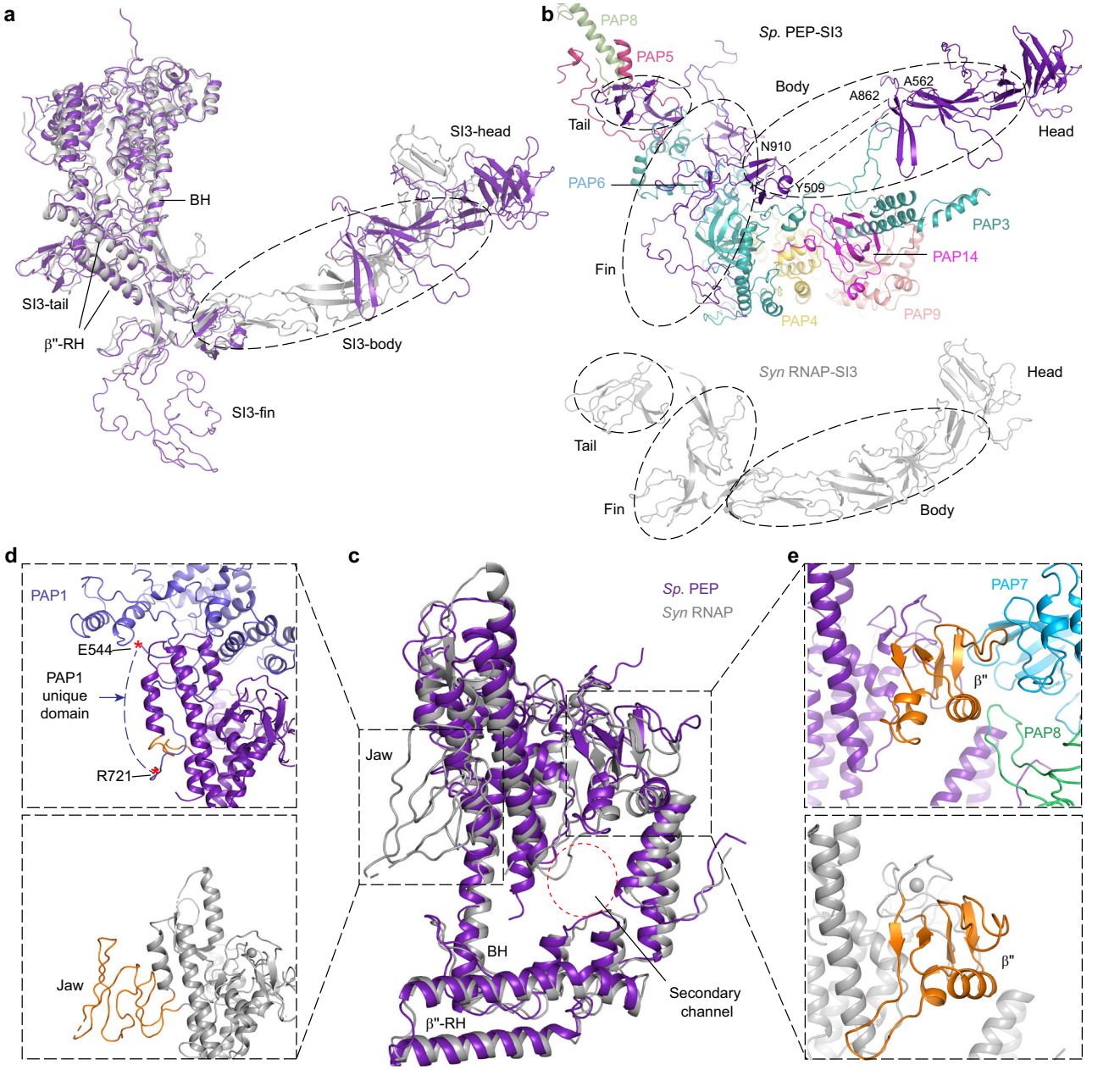


Fig. 3 | Comparison of the β" subunit of the *Spinacia oleracea* (*Sp.*) PEP core polymerase and cyanobacterial RNAP. a Comparison of the β" subunit of the *Sp.* PEP core polymerase and *Synechocystis* sp. PCC 6803 (*Syn*) RNAP (PDB 8GZG)[29] (r.m.s.d. of 5.702 Å for Cα atoms by PyMOL). **b** Comparison of the β" arch domain of sequence insertion 3 (SI3) in the *Sp.* PEP core polymerase and *Syn* RNAP (r.m.s.d. of 2.965 Å for Cα atoms by PyMOL, amino acids [aa] 340–1131). The PEP β" SI3 arch shows a larger fin that provides a binding site for PAPs. **c** Comparison of the region of β" other than SI3 between the *Sp.* PEP core polymerase and *Syn* RNAP (r.m.s.d. of 1.137 Å for Cα atoms by PyMOL). **d** Close-up view of the jaw domain in *Syn* RNAP and the corresponding position in the PEP core polymerase. The dashed line represents the unique domain of PAP1 connecting residues E544 and R721, which occupies the position corresponding to the jaw in *Syn* RNAP. **e** Close-up view of the βαβ motif adjacent to the secondary channel in the *Sp.* PEP core polymerase and *Syn* RNAP. The βαβ motif in the *Sp.* PEP core polymerase contributes to the binding site for PAP7 and PAP8. β", RNA polymerase β" subunit; PAP1, plastid transcriptionally active chromosome protein 3 (pTAC3); PAP3, plastid transcriptionally active chromosome protein 10 (pTAC10); PAP4, iron superoxide dismutase 3 (FSD3); PAP5, plastid transcriptionally active chromosome protein 12 (pTAC12/HEMERA); PAP6, fructokinase-like protein 1 (FLN1); PAP7, plastid transcriptionally active chromosome protein 14 (pTAC14); PAP8, plastid transcriptionally active chromosome protein 6 (pTAC6); PAP9, iron superoxide dismutase 2 (FSD2); PAP14, plastid transcriptionally active chromosome protein 18 (pTAC18). SI3, the β" arch domain of sequence insertion 3; BH, bridge helix; β"-RH, β" rim helices.

around the clamp of PEP, occupying a position in the jaw of bacterial RNAP that is absent in the PEP core (Figs. 3d, 6c). PAP2 was first identified as part of the transcriptionally active PEP in 2011, when PAP notation was originally defined[10]. An interaction was previously detected between PAP1 and PAP2 from the PEP complex purified from white mustard (*Sinapis alba*)[11]. However, we did not observe PAP2 in our structure. We did detect an additional density close to the N-terminal part of PAP1 in low-pass-filtered maps, which may correspond to PAP2 (Supplementary Fig. 2h). We refrained from modeling into this density due to its limited map quality. The absence of PAP2 in our structure may be due to its weak association in the PEP and/or its high flexibility.

PAP7 adopts a SET domain methyltransferase fold commonly found in histone lysine methyltransferases[41]. Similar to the Rubisco

https://doi.org/10.1038/s41467-024-54266-2

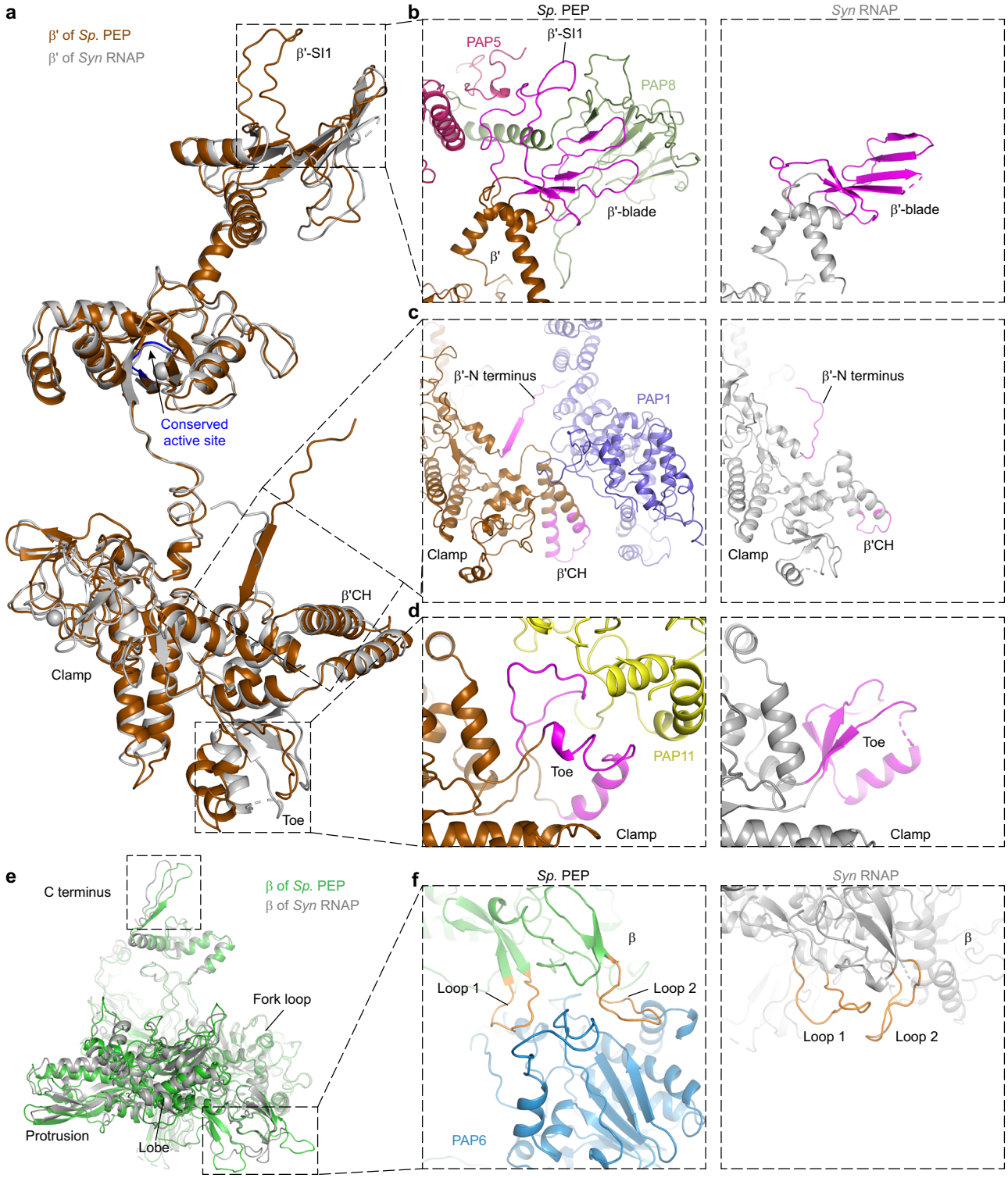
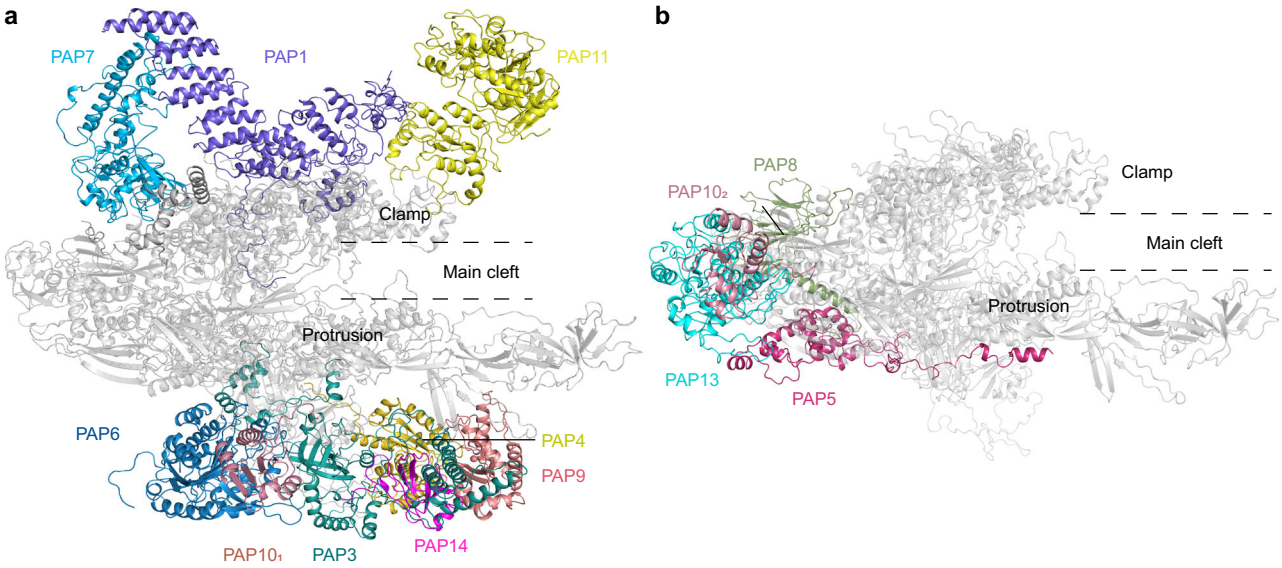


Fig. 4 | Comparison of the β and β' subunits of the *Spinacia oleracea* (*Sp.*) PEP core polymerase and cyanobacterial RNAP. a Structural superimposition of the β' subunit between the *Sp.* PEP core polymerase and *Synechocystis* sp. PCC 6803 (*Syn*) RNAP (PDB 8GZG)[29] (r.m.s.d. of 1.257 Å for Cα atoms by PyMOL). The conserved active site is highlighted in dark blue. Close-up view of the β' SI1 and the β' blade domain (**b**), the β' N terminus and β'CH (**c**), and the toe of the clamp (**d**). Structural domains showing a change in conformation are highlighted in magenta. **e** Structural superimposition of the β subunit between the *Sp.* PEP core polymerase and *Syn* RNAP (r.m.s.d. of 1.280 Å for Cα atoms by PyMOL). **f** Close-up view of two loops close to the β lobe domain of PEP, labeled as loop 1 and loop 2. The two loops adopt different conformation compared to those of *Syn* RNAP (right). The two loops are highlighted in orange. β, RNA polymerase β subunit; β', RNA polymerase β' subunit; PAP1, plastid transcriptionally active chromosome protein 3 (pTAC3); PAP5, plastid transcriptionally active chromosome protein 12 (pTAC12/HEMERA); PAP8, plastid transcriptionally active chromosome protein 6 (pTAC6); PAP11, UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-2,6-diaminopimelate ligase MurE homolog; β'-SI1, a sequence insertion in β' (residues 545–581); β'CH, clamp helices.

https://doi.org/10.1038/s41467-024-54266-2



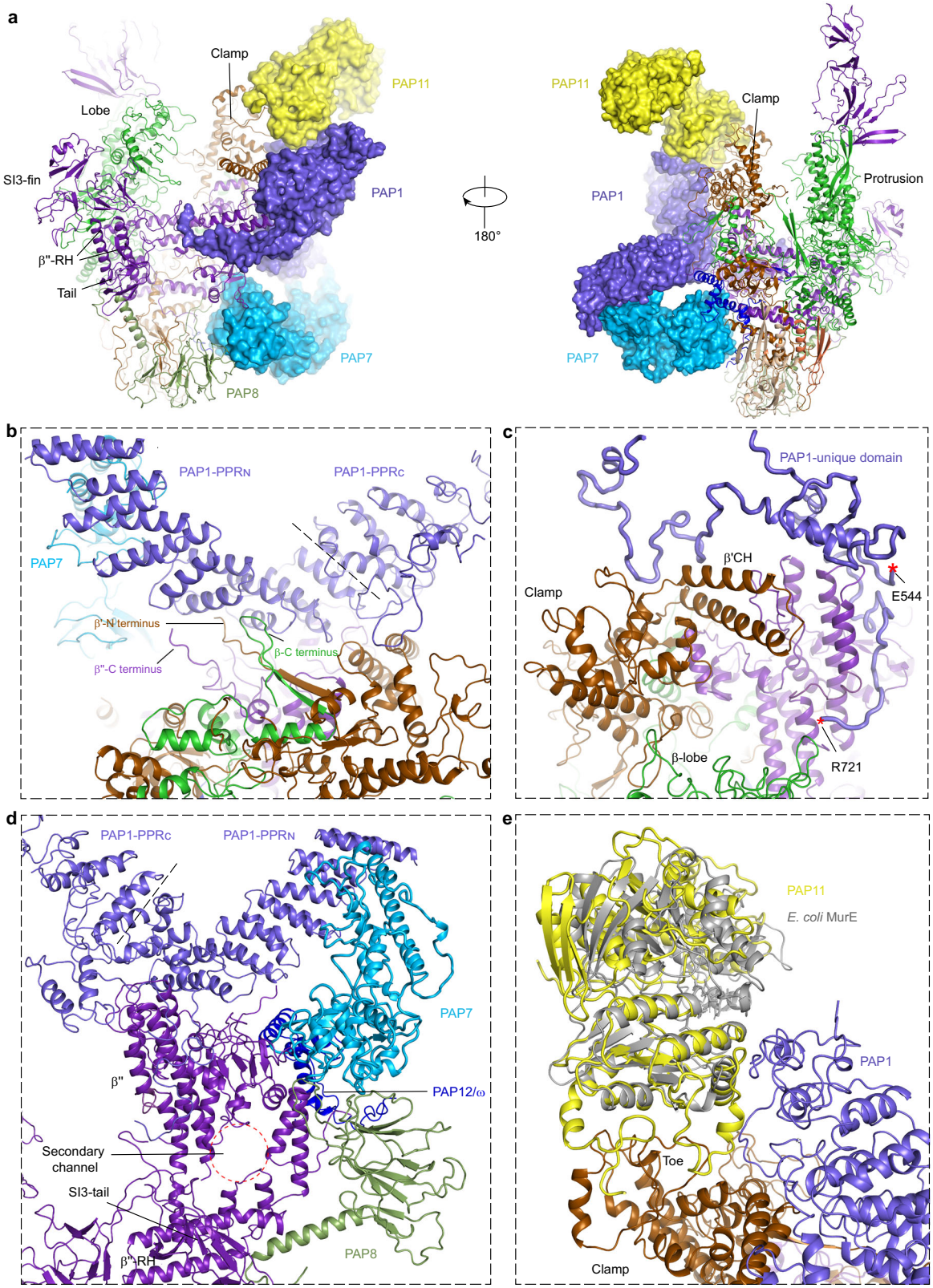
c

Cluster	PAP subunit	Interacting core subunits	Interacting PAP subunits
1	PAP1	β, β', β"	PAP7, PAP11
	PAP7	β', PAP12/ω	PAP8, PAP1
	PAP11	β'	PAP1
2	PAP5	α1, β, β', β"	PAP8, PAP6, PAP3, $PAP10_2$
	PAP8	β', β"	PAP7, PAP5
3	PAP3	β"	PAP4, PAP9, PAP14, PAP6, PAP5
	PAP14	possibly β"	PAP3, PAP4, PAP9
4	PAP4	β"	PAP9, PAP3, PAP14
	PAP9	none	PAP4, PAP3, PAP14
5	PAP6	β, β"	$PAP10_1$, PAP3, PAP5
	PAP13	α1, α2	$PAP10_2$
	$PAP10_1$	β	PAP6, PAP5
	$PAP10_2$	none	PAP13

Fig. 5 | Arrangement of PAPs in the *Spinacia oleracea* (*Sp.*) PEP complex. Position of PAPs located on the upper arm and the lower arm (**a**) and the intersection of the two arms (**b**). All PAPs are presented as cartoons. The PAPs are colored individually as indicated in Fig. 1; the *Sp.* PEP core subunits are colored in gray. **c** Summary table showing that the PAPs interact extensively with the PEP core subunits and with each other. α1, α2, RNA polymerase α subunit; β, RNA polymerase β subunit; β', RNA polymerase β' subunit; β'', RNA polymerase β'' subunit; PAP1, plastid transcriptionally active chromosome protein 3 (pTAC3); PAP3, plastid transcriptionally active chromosome protein 10 (pTAC10); PAP4, iron superoxide dismutase 3 (FSD3); PAP6, fructokinase-like protein 1 (FLN1); PAP7, plastid transcriptionally active chromosome protein 14 (pTAC14); PAP8, plastid transcriptionally active chromosome protein 6 (pTAC6); PAP9, iron superoxide dismutase 2 (FSD2); $PAP10_1$, $PAP10_2$, thioredoxin Z (Trx Z); PAP11, UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-2,6-diaminopimelate ligase MurE homolog; PAP12/ω, plastid transcriptionally active chromosome protein 7 (pTAC7); PAP13, fructokinase-like protein 2 (FLN2); PAP14, plastid transcriptionally active chromosome protein 18 (pTAC18).

large subunit methyltransferase (RBLSMT) in *Pisum sativum*, PAP7 contains an N-terminal SET domain (Supplementary Fig. 6b). The residues in the active site of RBLSMT are conserved in the equivalent positions in PAP7[42]. The structure of PAP7 indicates that it may function as a lysine methyltransferase within the PEP. The SET domain of PAP7 is bound to the PEP core through β'' and the ω subunit and contacts PAP8, while its C-terminal domain interacts with the N terminus of PAP1 (Fig. 6d). Notably, the β'' region binding to both PAP7 and PAP8 has a distinct conformation, unlike that in cyanobacterial RNAP, and is adjacent to the secondary channel serving as an entry point for substrate NTPs and an exit route for RNA 3' ends during backtracking in bacterial RNAPs[31,43,44] (Figs. 3e, 6d). Thus, PAP7, together with PAP8, may stabilize the PEP core and regulate NTP binding or PEP backtracking.

https://doi.org/10.1038/s41467-024-54266-2



PAP11 contains a MurE-like domain, similar to bacterial MurE, an amino acid ligase involved in peptidoglycan biosynthesis. However, homologs of the enzymes upstream of MurE in the peptidoglycan biosynthetic pathway are absent in Arabidopsis (*Arabidopsis thaliana*)[23]. The presence of this MurE-like domain in PAP11 thus raises a question about the role of PAP11, which interacts with the clamp-toe of β' and PAP1 (Fig. 6a, e). Notably, these interactions are mediated by two folds in the C-terminal domain of PAP11 that are absent in *E. coli* MurE (PDB 1E8C)[45] (Fig. 6e). In addition, the clamp-toe of β' interacting with PAP11 adopts a conformation distinctly unlike that of bacterial RNAPs (Fig. 4d). These observations indicate that PAP11 may stabilize the PEP core.

PAP1, PAP7, and PAP11 are essential for PEP activity and chloroplast development in Arabidopsis[12,21,23], indicating that they are

Fig. 6 | PAP1, PAP7, and PAP11 stabilize the *Spinacia oleracea* (*Sp.*) PEP core polymerase through embracing the clamp-claw. a PAP1, PAP7, and PAP11 (shown as surface presentation) embrace the clamp-claw of the *Sp*. PEP core polymerase. **b** Detailed interactions of the N-terminal pentatricopeptide repeats (PPRs) of PAP1 (PAP1-PPR_N) and the C-terminal PPRs of PAP1 (PAP1-PPR_C) with the β, β', and β'' subunits of the *Sp*. PEP core polymerase. **c** Detailed interactions of the PAP1-unique domain (aa 394–782) located between PAP1-PPR_N and PAP1-PPR_C with the β'-clamp and β'' of the *Sp*. PEP core polymerase. **d** Detailed interactions of PAP7 with the *Sp*. PEP core polymerase subunits (β'', PAP12/ω), PAP1, and PAP8. **e** Detailed interactions of PAP11 with the β‘ clamp and PAP1, and structural comparison of PAP11 and *E. coli* MurE (PDB 1E8C)[45] (r.m.s.d. of 3.226 Å for Cα atoms by PyMOL). β, RNA polymerase β subunit; β‘, RNA polymerase β' subunit; β'', RNA polymerase β'' subunit; PAP1, plastid transcriptionally active chromosome protein 3 (pTAC3); PAP7, plastid transcriptionally active chromosome protein 14 (pTAC14); PAP11, UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-2,6-diaminopimelate ligase MurE homolog; PAP12/ω, plastid transcriptionally active chromosome protein 7 (pTAC7); β''-RH, rim helices; SI3, the β'' arch domain of sequence insertion 3; β'CH, clamp helices.

essential components of the PEP. Taken together, these results indicate that the PAP1–PAP7–PAP11 cluster may stabilize the PEP core through embracing its clamp-claw.

PAP5 and PAP8 contribute to the assembly and stability of the PEP core

PAP5 and PAP8 wrap around the intersection of the two arms of the PEP core and engage in multiple interactions with the PEP core and other PAPs (Fig. 5b, c). PAP5 adopts a unique fold, which looks like a long thread with several random loops (Fig. 7a, b). PAP5 has multivalent contacts with the PEP core (α, β, β', and β'') and other PAPs (PAP3, PAP8, PAP6, and PAP10) (Fig. 7a–f). Starting from interacting with the β'- SI1, PAP5 wraps around the SI3-tail and β'' rim helices, passes through the PEP α subunits, and extends to the protrusion domain of the lobe-protrusion-claw. PAP5 forms a major platform for the docking of α-homodimer and the PAP6–$PAP10_1$ heterodimer.

PAP8 also adopts a unique fold (Fig. 7g, h). Only a portion of PAP8 (residues 111–324) was resolved in our PEP reconstruction; this showed that PAP8 contains three layers of β-sheets with an additional helix at its C terminus and sits near the secondary channel of the PEP. PAP8 interacts with the PEP core by forming an antiparallel β-fold through its two β-sheets and the β' blade domain and inserting a hairpin loop (residues 199–226) that protrudes from the middle layer β-sheet to contact β'' and PAP7. PAP8 also interacts with the PEP core through the binding of its C-terminal helix into the β' blade domain, the β'-SI1, the β'' rim helix, and the β''-SI3 tail. In addition, the C-terminal helix of PAP8 contacts PAP5, and one layer of β-sheets contacts PAP7.

Genetic and biochemical studies showed that PAP5 and PAP8 are important for PEP activity[13]. The structural features of PAP5 and PAP8 in the PEP indicate that they perform a structural role in the assembly and stability of PEP.

PAP3 and PAP14 scaffold the lobe-protrusion-claw

PAP3 and PAP14 form a scaffold for the lobe-protrusion-claw, providing the docking site for the PAP4–PAP9 heterodimer (Fig. 8a, b). To illustrate how PAP3 interacts with the PEP core and other PAPs, we divided PAP3 into three domains: a short N-terminal domain (NTD; aa 72–150), a middle domain (MD; aa 160–380), and a C-terminal domain (CTD; aa 381–618) (Fig. 8b). PAP3-NTD contacts the body, fin, and tail of SI3, as well as β''-RH. PAP3-MD interwinds with the SI3-fin, forming a concave surface for the major docking site of PAP4. PAP3-CTD interacts with the SI3 body (Fig. 8c–e). PAP14 is a cupin-like protein, belonging to the most functionally diverse protein family in plants[11,46]. PAP14 interlinks the PAP3-MD and PAP3-CTD and contacts PAP4 and PAP9 (Fig. 8a, b).

A knockout of *PAP3* results in an albino phenotype due to defects in PEP activity and chloroplast development in plants such as Arabidopsis and rice (*Oryza sativa*)[14–16]. Moreover, a *pap3* mutant in maize (*Zea mays*) does not accumulate a fully assembled PEP complex[47]. These characteristics of PAP3 reflect its role as an essential component of the PEP. The broad interactions of PAP3 with β'' and other PAPs indicate that PAP3 and PAP14 stabilize the PEP core by scaffolding the lobe-protrusion-claw.

Some PAPs have protective roles against oxidative stress

PAP4 (also reported as FE SUPEROXIDE DISMUTASE 3 [FSD3]) and PAP9 (also reported as FSD2) belong to the family of manganese/iron superoxide dismutases (SODs). The PAP4–PAP9 heterodimer is situated on the lobe-protrusion-claw mainly through PAP3 (Fig. 8a, b). PAP4 is embedded in a concave surface formed by PAP3-MD and SI3-fin, while PAP9 contacts PAP3-CTD and PAP14 (Fig. 8d, e). PAP4 interacts with PAP9 using the same interface as in the PAP9 homodimer[48] (Supplementary Fig. 7a). Although PAP4 and PAP9 are structurally highly similar, a specific C-terminal loop in PAP4 interacts with PAP3-MD, while a specific loop in PAP9 interacts with PAP3-CTD (Supplementary Fig. 7b, c). The loss of PAP4 and PAP9 function in a *pap4 pap9* double mutant was previously shown to lead to impaired PEP activity and chloroplast development[17]. *PAP4* overexpression only partially rescued the phenotypes of *pap9* mutants, while *PAP9* overexpression did not rescue the seedling death observed in *pap4* mutants[49]. In addition, *pap4 pap9* double mutant plants display a more severe phenotype than either single mutant[17]. These earlier results suggest that PAP4 and PAP9 have distinct roles in the PEP, which is consistent with our finding that PAP4 and PAP9 bind at distinct locations in the PEP. Therefore, the PAP4–PAP9 heterodimer may stabilize the PEP SI3 arch by interacting extensively with the PEP core, PAP3, and PAP14.

The structures of PAP4 and PAP9 show conserved histidine and aspartate residues at the active sites for SODs[50]. We detected iron ions within the active site of PAP9 but not PAP4 (Supplementary Fig. 7c, d). PAP4 and PAP9 exhibit SOD activity when tested in vitro, and *pap4* and *pap9* single mutants and *pap4 pap9* double mutants produced more reactive oxygen species and are more sensitive to oxidative stress than the wild type[17]. Therefore, we propose that PAP4 and PAP9 protect the PEP against oxidative damage by neutralizing superoxide radicals produced by photosynthesis.

PAPs with potential redox activities have structural roles

The last cluster is composed of four PAP subunits: two identical PAP10s (also reported as Trx Z), PAP6 (also reported as FLN1), and PAP13 (also reported as FLN2). The two PAP10s interact with PAP6 and PAP13 to form two heterodimers: PAP6–$PAP10_1$ and PAP13–$PAP10_2$ (Fig. 9a).

PAP10s are thioredoxins. PAP10 adopts the canonical αβα thioredoxin fold with two redox-active cysteine residues in the signature motif (CGPC) and has a specific extended loop that is absent in spinach Trx M (Supplementary Fig. 8a). PAP10 was previously proposed to reduce PAP6, PAP13, and PLASTID REDOX INSENSITIVE 2 (PRIN2)[22,51]. However, a catalytic-deficient PAP10 variant complemented the phenotypes observed in *pap10* knockout mutants[18]. The PAP10 catalytic cysteine residues are buried in PAP6 and not close to any cysteine residues of PAP6 (Supplementary Fig. 8b). Thus, the thioredoxin activity of PAP10 may not be crucial in the PEP. Loss of PAP10 function leads to severely decreased PEP activity and impaired chloroplast development[22]. $PAP10_1$ interacts with PAP5 and β through a specific extended loop, whereas $PAP10_2$ does not interact with the PEP core (Fig. 9b, c). Thus, PAP10 may have a structural rather than catalytic role in the PEP.

PAP6 is structurally related to the pfkB carbohydrate kinase family, in particular fructokinases (FRKs)[20,22] (Supplementary Fig. 8c).

https://doi.org/10.1038/s41467-024-54266-2

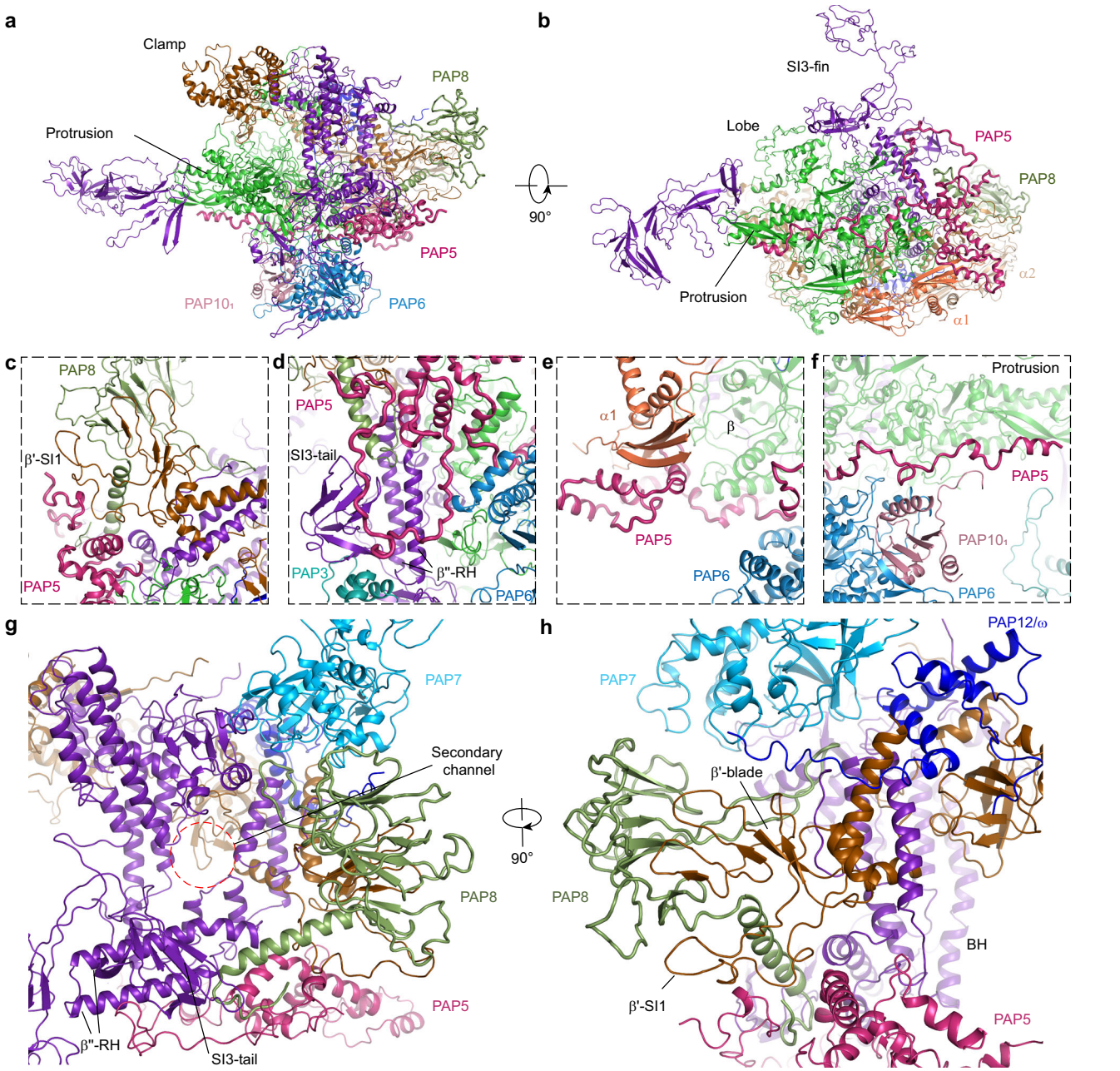


Fig. 7 | PAP5 and PAP8 contribute to assembly and stability of the *Spinacia oleracea* (*Sp.*) PEP core polymerase. a, **b** Structures and locations of PAP5 and PAP8 shown from different views. **c**–**f** Detailed views of the interactions between PAP5 and other subunits. **g**, **h** Detailed views of the interactions between PAP8 and other subunits shown from different views. α1, α2, RNA polymerase α subunit; β, RNA polymerase β subunit; β', RNA polymerase β' subunit; β'', RNA polymerase β'' subunit; PAP3, plastid transcriptionally active chromosome protein 10 (pTAC10); PAP5, plastid transcriptionally active chromosome protein 12 (pTAC12/HEMERA); PAP6, fructokinase-like protein 1 (FLN1); PAP7, plastid transcriptionally active chromosome protein 14 (pTAC14); PAP8, plastid transcriptionally active chromosome protein 6 (pTAC6); $PAP10_1$, thioredoxin Z (Trx Z); PAP12/ω, plastid transcriptionally active chromosome protein 7 (pTAC7); BH, bridge helix; β''-RH, rim helices; SI3, the β'' arch domain of sequence insertion 3; β'-SI1, a sequence insertion in β' (residues 545–581).

PAP6 adopts a closed conformation, unlike typical FRKs, which would most likely prevent substrate binding. PAP6 contains specific folds that mediate its interactions with the SI3 fin, loop 1 and loop 2 of β, PAP3, and PAP5 (Fig. 9b and Supplementary Fig. 8c). Loss of PAP6 function also results in an albino, seedling-lethal phenotype with severe defects in PEP activity[18–20]. Therefore, PAP6 performs a structural role in the PEP.

PAP13 and PAP6 have highly similar structures and sequences, and are conserved across plant species (Supplementary Fig. 8d and Supplementary Fig. 9). The PAP13–$PAP10_2$ heterodimer interacts with the N-terminal domains of the α-homodimer through PAP13 (Fig. 9c). Thus, the heterodimer might function in the early steps of PEP core assembly[52], which is in line with the crucial role of PAP13 in PEP transcription[20,25].

Taken together, the structural features of the PAP6–$PAP10_1$ and PAP13–$PAP10_2$ heterodimers indicate that they stabilize the PEP complex by interacting extensively with the PEP core and other PAPs.

PAPs may not regulate transcription directly

To investigate whether PAPs regulate transcription by directly interacting with nucleic acids and the σ factor, we obtained a model of the

https://doi.org/10.1038/s41467-024-54266-2

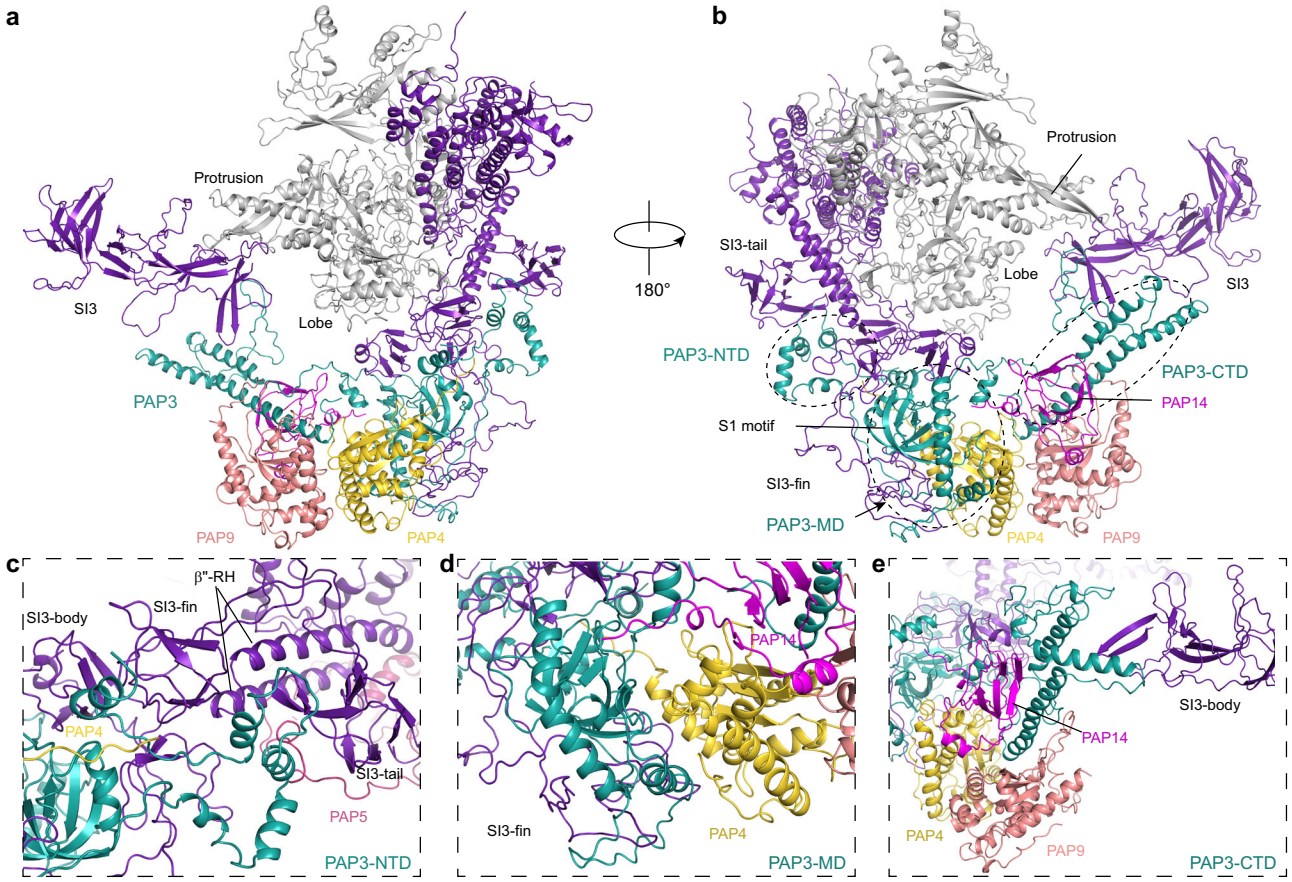


Fig. 8 | PAP3, PAP14, PAP4, and PAP9 scaffold the lobe-protrusion-claw. **a**, **b** Structure and locations of PAP3, PAP14, PAP4, and PAP9 shown from different views. PAP3, PAP14, PAP4, and PAP9 are shown and colored as indicated in Fig. 1. The β subunit is shown as a gray cartoon. To show its detailed interactions with the PEP core subunits and other PAPs, PAP3 was divided into three domains: the short N-terminal domain (PAP3-NTD; aa 72–150), the middle domain (PAP3-MD; aa 160–380) and the C-terminal domain (PAP3-CTD; aa 381–618). **c**–**e** Detailed view of the interactions between PAP3-NTD, PAP3-MD, and PAP3-CTD with PEP SI3, β''-RH, PAP14, PAP4, and PAP9. PAP3, plastid transcriptionally active chromosome protein 10 (pTAC10); PAP4, iron superoxide dismutase 3 (FSD3); PAP5, plastid transcriptionally active chromosome protein 12 (pTAC12/HEMERA); PAP9, iron superoxide dismutase 2 (FSD2); PAP14, plastid transcriptionally active chromosome protein 18 (pTAC18); SI3, the β'' arch domain of sequence insertion 3; β''-RH, rim helices.

PEP-promoter DNA open complex (PEPitc) by superposing an AlphaFold2 model of the spinach σ factor (SigF) and PEP with a cyanobacterial open, initially transcribing complex (RPitc; PDB 8GZG)[29] (Fig. 10a, b). The σ factor and promoter DNA could be accommodated well without major clashes. PAPs do not locate in direct vicinity to the promoter DNA. The σ-binding surface of the PEP is structurally conserved with respect to that of bacterial RNAP. This surface is not bound by PAPs and thus there is no steric clash between the PAPs and the σ factor. PAP11 contacts the σ factor and may thus contribute to its binding. The position of PAP1 at the direct vicinity of the entrance for the promoter DNA indicates that it might contribute to the DNA-binding activity of the PEP indirectly. Notably, there is a minor clash between the DNA and the β'CH in the clamp, which may be associated with the longer β'CH of the spinach PEP compared to cyanobacterial RNAP (Figs. 4c, 10c), suggesting small conformational changes in the PEPitc upon DNA binding.

We further obtained a structural model of the PEP elongation complex (PEP EC) by superposition with the structure of a cyanobacterial transcription elongation complex (EC; PDB 8SYI)[39] (Fig. 10d). The DNA and RNA could be accommodated well without major clashes. There are no obvious interactions between PAPs and nucleic acids. The DNA–RNA hybrid is positioned in the main cleft, and the 3' end of the RNA is close to the catalytic active center. We observed a minor clash between the downstream double-stranded DNA and β'CH in the PEP EC, as in the cyanobacterial PEPitc (Fig. 10e).

Discussion

We report here a unique architecture for the chloroplast transcription machinery that consists of the PEP core surrounded by many PAPs. Our results provide structural insights into chloroplast transcription and PAP functions.

Our model reveals that the PEP core resembles bacterial RNAPs (Fig. 2). Thus, the chloroplast transcription reaction likely involves a mechanism resembling that of bacterial RNAPs. Our model also indicates that PAPs do not enter the main cleft and are not associated directly with the second channel of the PEP core, as they bind through outer surfaces of the PEP core. We therefore propose that PAPs do not contribute directly to the NTP addition cycle. In addition, the results of modeling for PEP transcription initiation and elongation complexes indicate that the PEP accommodates the DNA–RNA hybrid in its active site in the same manner as cyanobacterial RNAP and likely uses the same mechanism as bacterial RNAPs to catalyze the RNA extension reaction (Fig. 10). Thus, our results suggest that the PEP core of the chloroplast transcription machinery is likely to use the same mechanism as cyanobacterial RNAP.

Although the overall structure of the PEP core is highly conserved, the subunits of the PEP core have additional sequence insertions, in particular SI3, compared to other bacterial RNAPs. PEP SI3 adopts a unique structure, not observed in other RNAPs, that facilitates interactions with PAPs. *E. coli* SI3 is involved in stabilizing the open complex, RNA transcription pausing, and intrinsic termination[53–56]. *E. coli*

https://doi.org/10.1038/s41467-024-54266-2

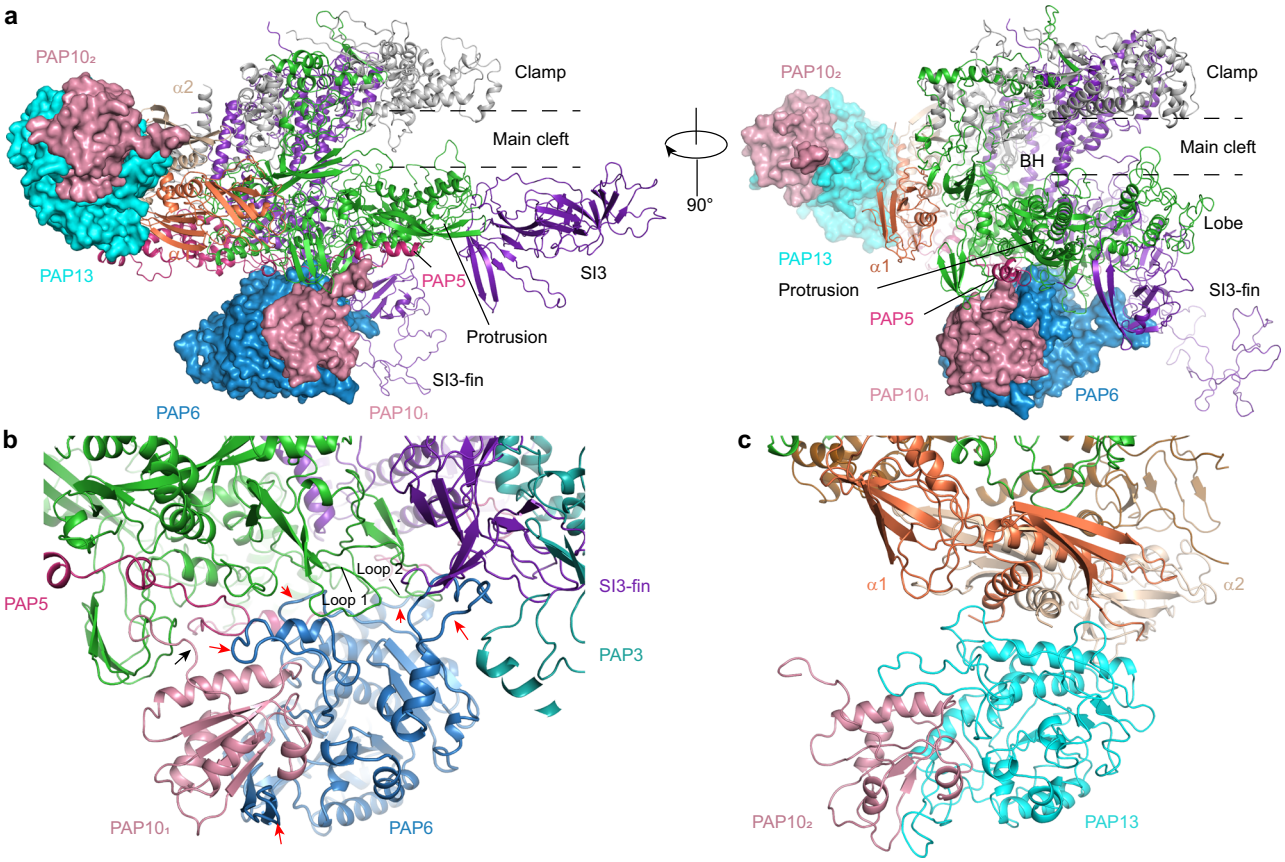


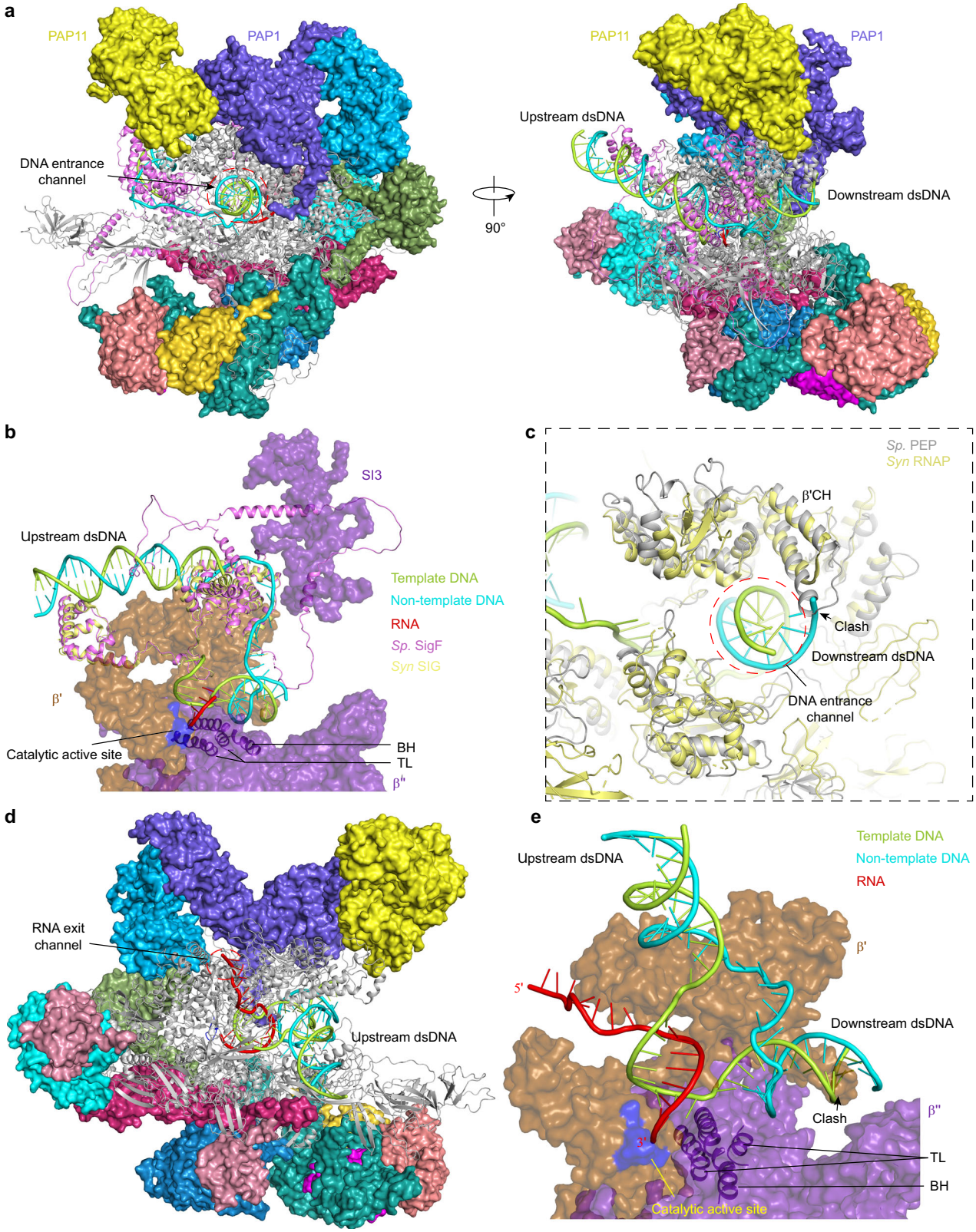
Fig. 9 | Redox clusters of the PAP6–PAP10$_1$ and PAP13–PAP10$_2$ heterodimers. **a** Structure and locations of the PAP6–PAP10$_1$ and PAP13–PAP10$_2$ heterodimers shown from different views. **b** Detailed view of the interactions between the PAP6–PAP10$_1$ heterodimer and the *Spinacia oleracea* (*Sp.*) PEP core polymerase and other PAPs. The black arrow highlights the PAP10$_1$-specific extended loop absent in Trx M, which interacts with PAP5 and the PEP β subunit. The red arrows highlight the folds of PAP6 that are distinctly unlike those in *V. cholerae* fructokinase (FRK) and mediate the interactions with PEP core subunits and other PAPs. **c** Structural details of the interactions between the PAP13–PAP10$_2$ heterodimer and the α-homodimer of the *Sp.* PEP core polymerase. α1, α2, RNA polymerase α subunit; PAP3, plastid transcriptionally active chromosome protein 10 (pTAC10); PAP5, plastid transcriptionally active chromosome protein 12 (pTAC12/HEMERA); PAP6, fructokinase-like protein 1 (FLN1); PAP10$_1$, PAP10$_2$, thioredoxin Z (Trx Z); PAP13, fructokinase-like protein 2 (FLN2); SI3, the β'' arch domain of sequence insertion 3; BH, bridge helix.

SI3 is highly mobile, and its conformation is dependent on the folded/unfolded state of the trigger loop/helix and the binding of transcription factors at the secondary channel of RNAP[57,58]. Recent advances in determining the structure and function of cyanobacterial RNAPs could clarify the potential function of PEP SI3[29,39]. Cyanobacterial SI3 contacts σ via its head to form an SI3-σ arch[29]. This arch stabilizes the transcription initiation complex and maintains cell growth under nutrient-limiting conditions. Thus, PEP SI3 may facilitate interaction with σ factors to stabilize the promoter complex. A study on the function of SI3 during transcription elongation reveals that binding of the incoming NTP results in a large swing motion of SI3[39]. Thus, the structural changes in the active site of the PEP during transcription could potentially lead to large-scale movements of PEP SI3, which in turn may lead to movements of the PEP core subunits and PAPs. The high flexibility of SI3 could explain why our model of the PEP SI3 head has a large shift toward the protrusion compared to cyanobacterial SI3. The complete deletion of cyanobacterial SI3 did not affect the assembly, maturation, or catalytic activity of RNAP[39]. Further investigations of the function of PEP SI3 at different stages of transcription will be required to understand the full array of its biological functions in plants.

Genetic disruption of individual *PAP* genes leads to albino or pale green phenotypes, severely impaired chloroplast development, and reduced PEP-dependent transcription, suggesting that PEP transcription activity is essentially dependent on each PAP. Moreover, the loss of PAP2, PAP3, or PAP5 function resulted in loss of the entire PEP, suggesting that PAPs are essential for PEP structural integrity[47,59]. Our results reveal that PAPs interact extensively with the PEP core and with each other. We thus propose that PAPs play an essential structural role in the PEP complex through maintaining its overall assembly and/or stability; moreover, PAPs depend on each other, such that loss of any one PAP results in loss of PEP integrity, which is essential for PEP activity.

Several PAPs contain additional enzymatic structural features. PAP4 and PAP9 display SOD activity in vitro, and the activity of the PAP4–PAP9 heterodimer is higher than that of PAP4 and PAP9 individually[17], suggesting that the PAP4–PAP9 heterodimer might have greater SOD activity. Thus, the integration of the PAP4–PAP9 heterodimer in the PEP could help protect the transcription apparatus from reactive oxygen species produced during photosynthesis. PAP7 adopts a methyltransferase fold, suggesting that it is a genuine methyltransferase, although its methylation substrate remains to be determined. PAP11 adopts a MurE-like fold. However, it has been suggested that PAP11 is not involved in peptidoglycan biosynthesis in Arabidopsis and that its substrate may have changed during the evolution from bryophytes to seed plants[23]. Whether PAP11 has MurE-like catalytic activity in the PEP remains to be determined. The structural analyses of PAP6, PAP10, and PAP13 indicated that they are unlikely to have additional enzymatic activities in the context of the PEP, but further study will be needed to ascertain this fact.

https://doi.org/10.1038/s41467-024-54266-2



PAPs may be involved in transcriptional regulation. PEP lacks the jaw domain that is conserved in bacterial RNAPs; instead, a PAP1-unique domain occupies the corresponding space (Fig. 3c, d). The jaw of bacterial RNAP plays essential roles in transcriptional initiation and pausing[60,61], suggesting that PAP1 might participate in these aspects of transcription. In addition, PAP1 interacts with β'CH (Fig. 6e), which protrudes into the downstream double-stranded DNA (dsDNA) channel (Fig. 10c). β'CH is longer than its counterpart in cyanobacterial RNAP (Fig. 4a, c), leading to a steric clash with the downstream dsDNA that prevents loading of the downstream dsDNA into the channel (Fig. 10c). Thus, PAP1 may be involved in changing the conformation of β'CH to enable the downstream dsDNA to be properly loaded into the channel.

Our model is limited by several residual unassigned densities in the cryo-EM reconstruction. Our structural information is also limited

Fig. 10 | Models of active chloroplast transcription complexes. a Model of an open, initially transcribing complex of PEP (PEPitc) shown from different views. The model was constructed by superimposing an AlphaFold2 model of the *Spinacia oleracea* (*Sp.*) σ factor (SigF) and the *Sp*. PEP complex with *Synechocystis* sp. PCC 6803 (*Syn*) RPitc (PDB 8GZG)[29]. The nucleic acid molecule was positioned based on structural superposition with *Syn* RPitc (PDB 8GZG)[29]. The PEP core is shown as a cartoon in gray and PAPs are shown as surfaces in the same color as in Fig. 1. The nucleic acids are shown as cartoons. Both SigF and DNA can be accommodated without large conformational rearrangements or loss of factors in PEPitc. **b** Close-up view of the active site in the modeled PEPitc. The active site is shown as a blue surface. The bridge helix (BH) and trigger loop (TL) are shown as cartoons. SigF and nucleic acids can be accommodated without major clashes. **c** Close-up view of the minor clash between β'CH and the downstream DNA in PEPitc. **d** Model of an actively elongating complex of PEP (PEP EC). The model was constructed by superimposing the *Sp*. PEP complex with the *Syn* EC (PDB 8SYI)[39]. The nucleic acid molecule was positioned based on structural superimposition with the *Syn* EC (PDB 8SYI)[39]. No large conformational rearrangements or loss of factors are required to form the PEP EC. **e** Close-up view of the active site in the modeled PEP EC. The clash is identical to that seen with the PEPitc in **c**. β', RNA polymerase β' subunit; β'', RNA polymerase β'' subunit; PAP1, plastid transcriptionally active chromosome protein 3 (pTAC3); PAP11, UDP-N-acetylmuramoyl-L-alanyl-D-glutamate-2,6-diaminopimelate ligase MurE homolog; SI3, the β'' arch domain of sequence insertion 3; BH, bridge helix; TL, trigger loop; β'CH, clamp helices.

to a model of PEP in an open state. In addition, the limited observed electron density in the N-terminal part of PAP11 suggests that PAP11 has a high degree of flexibility. Thus, PAP11 may clash with the σ factor during transcription initiation.

While we were revising this manuscript, three structural studies were published online of PEP complexes purified from tobacco and *Sinapis alba*[62–64]. Our conclusions are consistent with those reached in these three studies.

In summary, we present the architecture of the chloroplast transcription machinery and provide a strong structural basis for further studies into the molecular mechanisms and functions of chloroplast transcription.

Methods

Data reporting

No statistical method was used to predetermine the sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment.

Purification of the spinach PEP complex

About 5 kg of leaves from mature spinach plants were cut into small pieces, which were gently homogenized with a blender in 15 liters of pre-cooled Chloroplast Isolation Buffer (CIB, 20 mM HEPES, pH 8.0, 0.33 M sorbitol, 5 mM $MgCl_2$, 5 mM EGTA, 5 mM EDTA, 10 mM $NaHCO_3$, 60 mM sodium ascorbate). The resulting homogenate was filtered through three layers of Miracloth (Millipore). The crude chloroplasts were collected by centrifugation at 3000 × *g* for 5 min at 4 °C and lysed with 2 liters of Chloroplast Breaking Buffer (50 mM HEPES, pH 7.6, 4 mM EDTA, 20% [v/v] glycerol, 20 mM β-mercaptoethanol, 50 μg ml^{-1} phenylmethylsulfonyl fluoride [PMSF], 0.1 M $(NH_4)_2SO_4$). The insoluble materials were removed by centrifugation at 40,000 × *g* for 30 min at 4 °C. The supernatant was loaded onto a XK16/20 column (1.6-cm diameter, 20-cm length, column volume 30 ml) filled with Heparin Sepharose 6 Fast Flow resin (17099801, Lot: 10314514, Cytiva). The column was washed with 100 ml of a buffer containing 50 mM HEPES, pH 7.6, 4 mM EDTA, 10% (v/v) glycerol, 20 mM β-mercaptoethanol, 50 μg ml^{-1} PMSF. Proteins bound to the resin in the column were eluted with 60 ml of 0.28 M $(NH4)_2SO_4$ and separated on a 0.44–0.88 M continuous sucrose density gradient, in a buffer containing 50 mM HEPES, pH 7.6, 0.1 mM EDTA, 1 mM DTT, 0.5× proteinase inhibitor cocktail, and 5% (v/v) glycerol, by centrifugation using a Beckman Coulter Optima XPN-100 centrifuge with a SW40Ti rotor at 16,0000 × *g* for 23 h at 4 °C. After centrifugation, equal 24 fractions were collected from the top to the bottom of the gradient. Each fraction was resolved by electrophoresis and subjected to immunoblot analysis with an anti-RpoB antibody (polyclonal, PhytoAB, PHY1239) at a 1:5000 dilution. The peak fractions of the PEP complex were collected and concentrated to 5 mg ml^{-1} using a membrane concentrator with a 100-kDa cut-off. Then, the samples were subjected to gel filtration chromatography (Cytiva, Superose 6 Increase 10/300GL) in buffer containing 50 mM HEPES, pH 7.6, 2 mM EDTA, 2% (v/v) glycerol, 10 mM β-mercaptoethanol, and 50 μg ml^{-1} PMSF. The collected PEP complex was used for cryo-EM analysis.

Characterization of the spinach PEP complex

The purified PEP complex was analyzed on a 5% (w/v) BN-PAGE to assess its integrity. The BN-PAGE gel was stained in 25% (v/v) isopropanol, 10% (v/v) acetic acid, and 0.1% (w/v) Coomassie brilliant blue R-250 (CBB) for 30 min. Then, the stained BN-PAGE gel was destained in 5% (v/v) ethanol and 10% (v/v) acetic acid. The ~1-MDa band of the PEP complex was excised from the BN-PAGE gel for mass spectrometry (MS) analysis. The protein composition of the purified PEP complex was analyzed by SDS-PAGE using a 10–16% (w/v) gradient polyacrylamide gel. The SDS-PAGE gel was stained with CBB.

Protein identification of the spinach PEP complex by MS

For MS analysis, the ~1-MDa protein band of the PEP complex excised from the BN-PAGE gel was destained in 25 mM ammonium bicarbonate in 50% (v/v) acetonitrile. The proteins in the gel were reduced with 10 mM dithiothreitol at 37 °C for 1 h, alkylated with 55 mM iodoacetamide in 50 mM ammonium bicarbonate in the dark for 45 min, and digested with trypsin at 37 °C. The digested peptides were extracted from gel with buffers containing 5% (v/v) trifluoroacetic acid and 50% (v/v) acetonitirile by ultrasonic twice. The liquid containing the digested peptides was desalted using StageTips59 and completely freeze dried using a SpeedVac concentrator.

The liquid chromatography coupled tandem mass spectrometry (LC-MS/MS) analysis was performed using a LTQ Orbitrap Elite mass spectrometer (Thermo Scientific, Rockford, IL Waltham, MA) coupled online to an Easy-nLC 1000 in the data-dependent mode. The dried peptides were resuspended in 0.1% (v/v) formic acid with a final concentration 8 μg μL^{-1}, and 4 μL of each sample was injected into a capillary analytic column (inner diameter: 150 μm, length: 25 cm) packed with C18 particles (diameter: 1.9 μm) to separate by reverse phase LC. The mobile phases for the reverse phase LC contain buffer A [0.1% (v/v) formic acid] and buffer B [100% (v/v) acetonitirile, 0.1% (v/v) formic acid], and a 90-min non-linear gradient of buffer B from 3 to 30% (3–8%, 10 min; 8–20%, 60 min; 20–30%, 8 min; 30–100%, 2 min; 100%, 10 min) with a flow rate of 600 nL min^{-1} was used for peptide separation. Precursor ions were measured in the Orbitrap analyzer at 240,000 resolution (at 400 m/z) and a target value of 10^6 ions. The twenty most intense ions from each MS scan were isolated, fragmented, and measured in the linear ion trap. The Collisional dissociation (CID) normalized collision energy was set to 35.

The database search was performed for raw MS file using the software MaxQuant (v2.3.1.0). The *Spinacia oleracea* proteome sequence database was downloaded from Uniprot [https://www.uniprot.org/proteomes/UP001155700]. The parameters used for the database search were set up as follows: the type of search: Standard; the protease used for protein digestion: trypsin; protease cleavage sites: the C-terminus of residues K or R; the maximum number of allowed missed cleavages: 2; the mass tolerance for precursor ions: 20 PPM for the first search and 4.5 PPM for the main search; the mass

tolerance for fragments ions: 0.5 Da; the minimum peptide length: 7; the minimum number of unique peptides for protein identification: 0. The variable modifications include N-terminal acetylation and methionine oxidation. The fix modification includes cysteine carbamidomethylation. The false discovery rates (FDR) for both peptide and protein identifications were set to 0.01. Default values were used for all other parameters.

In vitro transcription assays

The transcription activity of PEP was assessed by visualizing the extension of CY5-labeled RNA. A template DNA (tDNA, 5′-GTAGAGACTTGTGTGTTCCTGCGCGCGGCTACCTGCATCAGAGGT-3′), a non-template DNA (ntDNA, 5′-ACCTCTGATGCAGGTAGCCGCGCGCAGGAACACACAAGTCTCTAC-3′), and a CY5-labeled RNA (5′-CY5-ACAUACGCCGCGCGC-3′) were used for the in vitro transcription assays. The nucleic acid scaffold was reconstituted in vitro by mixing the tDNA (final concentration 5 μM) and CY5-labeled RNA (final concentration 2.5 μM) in 20 μl DEPC-treated water at 98 °C for 2 min, and then cooling the mixture to 10 °C at a rate of 1 °C min^{-1}. The scaffold (final RNA concentration RNA 500 nM) was mixed with the purified PEP complex (1 mM) or *E. coli* RNAP (0.5 U ml^{-1}) and the ntDNA in reconstitution buffer (10 mM Tris-HCl, pH 8.0, 40 mM KCl, 5 mM $MgCl_2$, 0.02 mg ml^{-1} acetylated BSA). Then, the mixture was incubated at room temperature for 1 h. RNA extension was started by the addition of a mixture of ATP, CTP, GTP, and UTP (0.1 mM each) and the mixture was incubated at room temperature for 1 h. The reaction was quenched by the addition of an equal volume of 2× stop buffer (8 M urea, 20 mM EDTA, 5 mM Tris-HCl, pH 7.5). The samples were loaded onto 16% (w/v) Tris-borate-EDTA (TBE)-polyacrylamide (20% [w/v] 19:1 acrylamide:bisacrylamide) gels containing 7 M urea, and electrophoresis was performed at 300 V for 75 min in an ice bath. The gels were scanned with a chemiluminescence imaging system (Fusion FX7, VILBER) to collect CY5 fluorescence.

Cryo-EM data collection

The concentration of the purified PEP complex was adjusted to about 2 mg ml^{-1}. Three microliters of protein solution were applied to a glow-discharged (Solarus plasma cleaner, Gatan) holey carbon grid (Quantifoil grid R2/1, 200 mesh, Cu) that had been treated with H_2 and O_2 mixtures for 30 s. The grid was plunged into liquid ethane cooled by liquid nitrogen using a Vitrobot Mark IV (Thermo Fisher Scientific). The parameters were set as follows: blot force 0, blotting time 3 s, humidity 100%, sample chamber temperature 8 °C. Cryo-EM images were collected on a Titan Krios (Thermo Fisher Scientific, Hillsboro, USA) electron microscope at 300 kV equipped with a K3 camera (Gatan) and at nominal magnification of 62,000 in super-resolution mode. Movies were recorded in super-resolution mode and Fourier-cropped to give a resulting calibrated pixel size of 1.10 Å at the specimen level. An exposure rate of 22.5 e^- per pixel per s was set and a fresh super-resolution gain reference was performed at this dose rate before data acquisition. Each movie consists of 40 frames with a total dose of 50 e− per $Å^2$. In total, 7175 movies were collected for the PEP complex. Data acquisition was carried out using the EPU software (2.12.0.2771, Thermo Fisher Scientific) with defocus values from −0.8 to −2.0 μm.

Image processing of the spinach PEP complex

All of the stacked frames were subjected to motion correction and defocus was estimated using cryoSPARC (v4.2.1)[65]. The F-crop factor was set to 1/2 to generate aligned micrographs in Motion Correction. Particles were picked by crYOLO[66] using a pretrained model. A total of 2,338,580 particles were picked from 7175 micrographs at a threshold of 0.01 with a box size of 200 pixels. These particles were then imported to CryoSPARC (v4.2.1) at 512-pixel box size, and down-sampled to 128 pixels for the first 2D classification with 100 classes. The resulting 1,266,268 particles were selected for a second 2D classification with 50 classes. The resulting 384,187 particles from 21 good 2D classes were selected for 3D reconstruction. These particles were re-extracted at a resolution of 512-pixel box size and subjected to 3D ab initio reconstruction with CryoSPARC (v4.2.1) to build four initial models, followed by heterogeneous refinement. Among the four initial models, the class showing the clearest features of the PEP complex containing 180,924 particles was selected for homogeneous refinement, resulting in a map at a nominal global resolution of 3.16 Å. In this map, two regions had weak electron densities, making them unsuitable for model building. Thus, we attempted to improve the quality of the map via local refinement, which resulted in improved resolutions with a nominal resolution of 3.78 Å (Map 1) and 3.24 Å (Map 2). The resolution of the remaining region was improved very slightly from 3.16 Å to 3.13 Å (Map 3) by particle subtraction and local refinement. The global map, improved Map 1, and improved Map 2 were used for model building.

Model building and refinement of the spinach PEP complex

To build an atomic model of PEP, the model of *Synechocystis* sp. PCC 6803 RNA polymerase complex (PDB 8GZG)[29] and the structures of PAP subunits from *Arabidopsis thaliana* predicted by AlphaFold2[26,27] were fitted into density maps as an initial model in UCSF Chimera (v1.15-42258) (https://www.rbvi.ucsf.edu/chimera). The initial atomic model was manually checked and mutated to fit to the map in COOT (v0.95)[67]. The regions subjected to local refinement above were designated as PAP11 and the part of RpoC2 (aa 562–862) according to the overall structural characteristics; the models of PAP11 and the part of RpoC2 predicted by AlphaFold2 were directly employed and could not be further modified due to their poor electron density map. Realspace refinement was performed in PHENIX (v1.19.2-4158)[68]; COOT (v0.95) was used for model building and PHENIX (v1.19.2-4158) for refinement until convergence. Finally, all parameters were generated by PHENIX (v1.19.2-4158) and figures were drawn with the PyMOL (v3.0.0) Molecular Graphic System (Schrödinger)[69] and UCSF ChimeraX (v1.5)[70].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The cryo-EM density map and model for the spinach PEP complex generated in this study have been deposited in the Electron Microscopy Data Bank and the Protein Data Bank with accession codes EMD-38799 and 8XZV, respectively. Two local refinement maps of PAP11 and part of RpoC2 (residues 562–862) generated in this study have been deposited in the Electron Microscopy Data Bank with accession codes EMD-61151 and EMD-61149, respectively. The following previously published PDB codes were used for model building and structural comparison: 8GZG, 6GH5, 1E8C, 8SYI, 2H21, 7BJK, 2PUK, and 5YGG. The all protein mass spectrometry raw data have been deposited in a ProteomeXchange partner repository with accession code PXD056642. The data that support the findings of this study are presented in the paper and/or the Supplementary Information. Source data are provided with this paper.

References

1. Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**, 123–135 (2004).
2. Soll, J. & Schleiff, E. Protein import into chloroplasts. *Nat. Rev. Mol. Cell Biol.* **5**, 198–208 (2004).
3. Pogson, B. J. & Albrecht, V. Genetic dissection of chloroplast biogenesis and development: an overview. *Plant Physiol.* **155**, 1545–1551 (2011).

4. Zhang, Y., Tian, L. & Lu, C. Chloroplast gene expression: Recent advances and perspectives. *Plant Commun.* **4**, 100611 (2023).
5. Yagi, Y. & Shiina, T. Recent advances in the study of chloroplast gene expression and its evolution. *Front. Plant Sci.* **5**, 61 (2014).
6. Börner, T., Aleynikova, A. Y., Zubo, Y. O. & Kusnetsov, V. V. Chloroplast RNA polymerases: Role in chloroplast biogenesis. *Biochim. Biophys. Acta* **1847**, 761–769 (2015).
7. Legen, J. et al. Comparative analysis of plastid transcription profiles of entire plastid chromosomes from tobacco attributed to wild-type and PEP-deficient transcription machineries. *Plant J.* **31**, 171–188 (2002).
8. Zhelyazkova, P. et al. The Primary transcriptome of barley chloroplasts: Numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *Plant Cell* **24**, 123–136 (2012).
9. Pfannschmidt, T. et al. Plastid RNA polymerases: orchestration of enzymes with different evolutionary origins controls chloroplast biogenesis during the plant life cycle. *J. Exp. Bot.* **66**, 6957–6973 (2015).
10. Steiner, S., Schröter, Y., Pfalz, J. & Pfannschmidt, T. Identification of essential subunits in the plastid-encoded RNA polymerase complex reveals building blocks for proper plastid development. *Plant Physiol.* **157**, 1043–1055 (2011).
11. Ruedas, R. et al. Three-dimensional envelope and subunit interactions of the plastid-encoded RNA polymerase from *Sinapis alba*. *Int. J. Mol. Sci.* **23**, 9922 (2022).
12. Yagi, Y., Ishizaki, Y., Nakahira, Y., Tozawa, Y. & Shiina, T. Eukaryotic-type plastid nucleoid protein pTAC3 is essential for transcription by the bacterial-type plastid RNA polymerase. *Proc. Natl Acad. Sci. USA* **109**, 7541–7546 (2012).
13. Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K. J. & Oelmüller, R. pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell* **18**, 176–197 (2006).
14. Williams-Carrier, R., Zoschke, R., Belcher, S., Pfalz, J. & Barkan, A. A major role for the plastid-encoded RNA polymerase complex in the expression of plastid transfer RNAs. *Plant Physiol.* **164**, 239–248 (2014).
15. Chang, S. H. et al. pTAC10, a key subunit of plastid-encoded RNA polymerase, promotes chloroplast development. *Plant Physiol.* **174**, 435–449 (2017).
16. Seo, D. H. et al. PEP-ASSOCIATED PROTEIN 3 regulates rice tiller formation and grain yield by controlling chloroplast biogenesis. *Plant Physiol.* **194**, 805–818 (2024).
17. Myouga, F. et al. A heterocomplex of iron superoxide dismutases defends chloroplast nucleoids against oxidative stress and is essential for chloroplast development in *Arabidopsis*. *Plant Cell* **20**, 3148–3162 (2008).
18. Wimmelbacher, M. & Börnke, F. Redox activity of thioredoxin *z* and fructokinase-like protein 1 is dispensable for autotrophic growth of *Arabidopsis thaliana*. *J. Exp. Bot.* **65**, 2405–2413 (2014).
19. Shi, C. et al. Conserved role of fructokinase-like protein 1 in chloroplast development revealed by a seedling-lethal albino mutant of pepper. *Hortic. Res.* **9**, uhab084 (2022).
20. Gilkerson, J., Perez-Ruiz, J. M., Chory, J. & Callis, J. The plastid-localized pfkB-type carbohydrate kinases FRUCTOKINASE-LIKE 1 and 2 are essential for growth and development of *Arabidopsis thaliana*. *BMC Plant Biol.* **12**, 102 (2012).
21. Gao, Z. P. et al. A functional component of the transcriptionally active chromosome complex, Arabidopsis pTAC14, interacts with pTAC12/HEMERA and regulates plastid gene expression. *Plant Physiol.* **157**, 1733–1745 (2011).
22. Arsova, B. et al. Plastidial thioredoxin *z* interacts with two fructokinase-like proteins in a thiol-dependent manner: evidence for an essential role in chloroplast development in *Arabidopsis* and *Nicotiana benthamiana*. *Plant Cell* **22**, 1498–1515 (2010).
23. Garcia, M. et al. An *Arabidopsis* homolog of the bacterial peptidoglycan synthesis enzyme MurE has an essential role in chloroplast development. *Plant J.* **53**, 924–934 (2008).
24. Yu, Q. B. et al. TAC7, an essential component of the plastid transcriptionally active chromosome complex, interacts with FLN1, TAC10, TAC12 and TAC14 to regulate chloroplast gene expression in *Arabidopsis thaliana*. *Physiol. Plant.* **148**, 408–421 (2013).
25. Huang, C. et al. The reduced plastid-encoded polymerase-dependent plastid gene expression leads to the delayed greening of the *Arabidopsis fln2* mutant. *PLoS ONE* **8**, e73092 (2013).
26. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
27. Varadi, M. et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, D439–D444 (2022).
28. Bergsland, K. J. & Haselkorn, R. Evolutionary relationships among eubacteria, cyanobacteria, and chloroplasts: evidence from the *rpoC1* gene of *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **173**, 3446–3455 (1991).
29. Shen, L. et al. An SI3-σ arch stabilizes cyanobacteria transcription initiation complex. *Proc. Natl. Acad. Sci. USA* **120**, e2219290120 (2023).
30. Glyde, R. et al. Structures of bacterial RNA polymerase complexes reveal the mechanism of DNA loading and transcription initiation. *Mol. Cell* **70**, 1111–1120 (2018).
31. Zhang, G. et al. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**, 811–824 (1999).
32. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 Ångstrom resolution. *Science* **292**, 1863–1876 (2001).
33. Lane, W. J. & Darst, S. A. Molecular evolution of multisubunit RNA polymerases: sequence analysis. *J. Mol. Biol.* **395**, 671–685 (2010).
34. Zhang, J., Palangat, M. & Landick, R. Role of the RNA polymerase trigger loop in catalysis and pausing. *Nat. Struct. Mol. Biol.* **17**, 99–104 (2010).
35. Toulokhonov, I., Zhang, J., Palangat, M. & Landick, R. A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Mol. Cell* **27**, 406–419 (2007).
36. Mishanina, T. V., Palo, M. Z., Nayak, D., Mooney, R. A. & Landick, R. Trigger loop of RNA polymerase is a positional, not acid-base, catalyst for both transcription and proofreading. *Proc. Natl Acad. Sci. USA* **114**, E5103–E5112 (2017).
37. Mazumder, A., Lin, M., Kapanidis, A. N. & Ebright, R. H. Closing and opening of the RNA polymerase trigger loop. *Proc. Natl Acad. Sci. USA* **117**, 15642–15649 (2020).
38. Riaz-Bradley, A. Transcription in cyanobacteria: A distinctive machinery and putative mechanisms. *Biochem. Soc. Trans.* **47**, 679–689 (2019).
39. Qayyum, M. Z. et al. Structure and function of the Si3 insertion integrated into the trigger loop/helix of cyanobacterial RNA polymerase. *Proc. Natl Acad. Sci. USA* **121**, e2311480121 (2024).
40. Sevostyanova, A., Belogurov, G. A., Mooney, R. A., Landick, R. & Artsimovitch, I. The β subunit gate loop is required for RNA polymerase modification by RfaH and NusG. *Mol. Cell* **43**, 253–262 (2011).
41. Dillon, S. C., Zhang, X., Trievel, R. C. & Cheng, X. The SET-domain protein superfamily: Protein lysine methyltransferases. *Genome Biol.* **6**, 227 (2005).
42. Trievel, R. C., Flynn, E. M., Houtz, R. L. & Hurley, J. H. Mechanism of multiple lysine methylation by the SET domain enzyme Rubisco LSMT. *Nat. Struct. Mol. Biol.* **10**, 545–552 (2003).
43. Cramer, P. et al. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288**, 640–649 (2000).
44. Korzheva, N. et al. A structural model of transcription elongation. *Science* **289**, 619–625 (2000).

45. Gordon, E. J. et al. Crystal structure of UDP-N-acetylmuramoyl-L-alanyl-D-glutamate: meso-diaminopimelate ligase from *Escherichia coli*. *J. Biol. Chem.* **276**, 10999–11006 (2001).
46. Dunwell, J. M., Purvis, A. & Khuri, S. Cupins: the most functionally diverse protein superfamily? *Phytochemistry* **65**, 7–17 (2004).
47. Pfalz, J. et al. ZmpTAC12 binds single-stranded nucleic acids and is essential for accumulation of the plastid-encoded polymerase complex in maize. *N. Phytol.* **206**, 1024–1037 (2015).
48. Favier, A. et al. The plastid-encoded RNA polymerase-associated protein PAP9 is a superoxide dismutase with unusual structural features. *Front. Plant Sci.* **12**, 668897 (2021).
49. Gallie, D. R. & Chen, Z. Chloroplast-localized iron superoxide dismutases FSD2 and FSD3 are functionally distinct in *Arabidopsis*. *PLoS ONE* **14**, e0220078 (2019).
50. Perry, J. J. P., Shin, D. S., Getzoff, E. D. & Tainer, J. A. The structural biochemistry of the superoxide dismutases. *Biochim. Biophys. Acta* **1804**, 245–262 (2010).
51. Díaz, M. G. et al. Redox regulation of PEP activity during seedling establishment in *Arabidopsis thaliana*. *Nat. Commun.* **9**, 50 (2018).
52. Igarashi, K., Fujita, N. & Ishihama, A. Identification of a subunit assembly domain in the alpha subunit of *Escherichia coli* RNA polymerase. *J. Mol. Biol.* **218**, 1–6 (1991).
53. Windgassen, T. A. et al. Trigger-helix folding pathway and SI3 mediate catalysis and hairpin-stabilized pausing by *Escherichia coli* RNA polymerase. *Nucleic Acids Res* **42**, 12707–12721 (2014).
54. Guo, X. et al. Structural basis for NusA stabilized transcriptional pausing. *Mol. Cell* **69**, 816–827.e4 (2018).
55. Kang, J. Y. et al. RNA polymerase accommodates a pause RNA hairpin by global conformational rearrangements that prolong pausing. *Mol. Cell* **69**, 802–815.e5 (2018).
56. Ray-Soni, A., Mooney, R. A. & Landick, R. Trigger loop dynamics can explain stimulation of intrinsic termination by bacterial RNA polymerase without terminator hairpin contact. *Proc. Natl Acad. Sci. USA* **114**, E9233–E9242 (2017).
57. Abdelkareem, M. et al. Structural basis of transcription: RNA polymerase backtracking and its reactivation. *Mol. Cell* **75**, 298–309.e294 (2019).
58. Shin, Y. et al. Structural basis of ribosomal RNA transcription regulation. *Nat. Commun.* **12**, 528 (2021).
59. Kindgren, P. & Strand, Å. Chloroplast transcription, untangling the Gordian Knot. *N. Phytol.* **206**, 889–891 (2015).
60. Ederth, J., Artsimovitch, I., Isaksson, L. A. & Landick, R. The downstream DNA jaw of bacterial RNA polymerase facilitates both transcriptional initiation and pausing. *J. Biol. Chem.* **277**, 37456–37463 (2002).
61. Drennan, A. et al. Key roles of the downstream mobile jaw of *Escherichia coli* RNA polymerase in transcription initiation. *Biochemistry* **51**, 9447–9459 (2012).
62. do Prado, P. F. V. et al. Structure of the multi-subunit chloroplast RNA polymerase. *Mol. Cell* **84**, 1–16 (2024).
63. Wu, X.-X. et al. Cryo-EM structures of the plant plastid-encoded RNA polymerase. *Cell* **187**, 1127–1144 (2024).
64. Vergara-Cruces, Á. et al. Structure of the plant plastid-encoded RNA polymerase. *Cell* **187**, 1145–1159 (2024).
65. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
66. Wagner, T. & Raunser, S. The evolution of SPHIRE-crYOLO particle picking and its application in automated cryo-EM processing workflows. *Commun. Biol.* **3**, 61 (2020).
67. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Cryst. D.* **66**, 486–501 (2010).
68. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D.* **66**, 213–221 (2010).
69. Delano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific, San Carlos, CA, USA, 2002).
70. Pettersen, E. F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

Acknowledgements

We are grateful to the staff at the National Facility for Protein Science (NFPS) for instrument support and technical assistance during cryo-EM data collection. This work was supported in part by the National Key Research and Development Program of China (2022YFF1001700, C.L.), the Strategic Priority Research Program of CAS (XDA26050402, L.-J.Y.) and the Science & Technology Specific Project in Agricultural High-tech Industrial Demonstration Area of the Yellow River Delta (2022SZX12, L.-J.Y.).

Author contributions

C.L. conceived the project. T.W, Y.F., W.P., and Y.Z. performed sample preparation and characterization. G.-L.W. collected the cryo-EM data. G.-L.W. and L.-J.Y. processed the cryo-EM data and reconstructed the cryo-EM density map. G.-L.W., L.-J.Y., and Y.F. built the structure model and refined the structure. Y.F. and G.-L.W. prepared the figures. G.-L.W., L.-J.Y., Y.F., and C.L. analyzed the structure. Y.Z. and A.Z. contributed to the discussion and comments on results. Y.F., G.-L.W., L.-J.Y., and C.L. jointly wrote the manuscript. All authors contributed to the discussion and commented on the results and the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-54266-2.

Correspondence and requests for materials should be addressed to Long-Jiang Yu or Congming Lu.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.