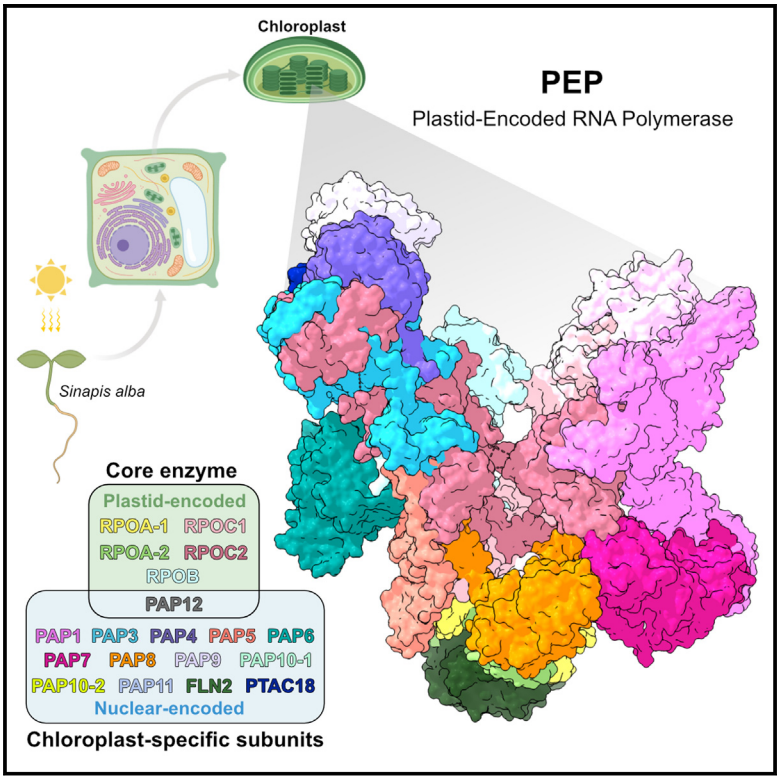
Molecular Cell

Article

Structure of the multi-subunit chloroplast RNA polymerase

Graphical abstract



Highlights

- Cryo-EM structure of the 19-subunit plastid-encoded RNA polymerase complex (PEP)
- The plastid-encoded subunits adopt a prokaryotic-like RNA polymerase core structure
- Chloroplast-specific factors (PAPs) arrange around the core in a unique fashion

Authors

Paula F.V. do Prado, Frederik M. Ahrens, Monique Liebers, Noah Ditz, Hans-Peter Braun, Thomas Pfannschmidt, Hauke S. Hillen

Correspondence

thomas.pfannschmidt@botanik.uni-hannover.de (T.P.), hauke.hillen@med.uni-goettingen.de (H.S.H.)

In brief

do Prado et al. report the cryo-EM structure of the plastid-encoded RNA polymerase complex from white mustard. The structure reveals the architecture of the major chloroplast transcription apparatus and provides insights into its unique associated factors.

do Prado et al., 2024, Molecular Cell *84*, 910–925
March 7, 2024
https://doi.org/10.1016/j.molcel.2024.02.003

CellPress

CellPress
OPEN ACCESS

Molecular Cell

Article

Structure of the multi-subunit chloroplast RNA polymerase

Paula F.V. do Prado,[1,2,3,7] Frederik M. Ahrens,[4,7] Monique Liebers,[4] Noah Ditz,[5] Hans-Peter Braun,[5] Thomas Pfannschmidt,[4,*] and Hauke S. Hillen[1,2,3,6,8,*]

[1]University Medical Center Göttingen, Department of Cellular Biochemistry, Humboldtallee 23, 37073 Göttingen, Germany
[2]Max Planck Institute for Multidisciplinary Sciences, Research Group Structure and Function of Molecular Machines, Am Fassberg 11, 37077 Göttingen, Germany
[3]Cluster of Excellence "Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells" (MBExC), University of Göttingen, 37075 Göttingen, Germany
[4]Institute of Botany, Plant Physiology, Leibniz University Hannover, Herrenhäuser Str. 2, 30419 Hannover, Germany
[5]Institute of Plant Genetics, Plant Molecular Biology and Plant Proteomics, Leibniz University Hannover, Herrenhäuser Str. 2, 30419 Hannover, Germany
[6]Göttingen Center for Molecular Biosciences (GZMB), Research Group Structure and Function of Molecular Machines, University of Göttingen, 37077 Göttingen, Germany
[7]These authors contributed equally
[8]Lead contact
*Correspondence: thomas.pfannschmidt@botanik.uni-hannover.de (T.P.), hauke.hillen@med.uni-goettingen.de (H.S.H.)
https://doi.org/10.1016/j.molcel.2024.02.003

SUMMARY

Chloroplasts contain a dedicated genome that encodes subunits of the photosynthesis machinery. Transcription of photosynthesis genes is predominantly carried out by a plastid-encoded RNA polymerase (PEP), a nearly 1 MDa complex composed of core subunits with homology to eubacterial RNA polymerases (RNAPs) and at least 12 additional chloroplast-specific PEP-associated proteins (PAPs). However, the architecture of this complex and the functions of the PAPs remain unknown. Here, we report the cryo-EM structure of a 19-subunit PEP complex from *Sinapis alba* (white mustard). The structure reveals that the PEP core resembles prokaryotic and nuclear RNAPs but contains chloroplast-specific features that mediate interactions with the PAPs. The PAPs are unrelated to known transcription factors and arrange around the core in a unique fashion. Their structures suggest potential functions during transcription in the chemical environment of chloroplasts. These results reveal structural insights into chloroplast transcription and provide a framework for understanding photosynthesis gene expression.

INTRODUCTION

Plants convert light energy into chemical energy by means of photosynthesis. This process takes place in specialized cell organelles called chloroplasts. These emerged from an endosymbiotic event in which a cyanobacteria-like photosynthetic bacterium was engulfed by a mitochondriate eukaryotic cell. During subsequent evolution, the endosymbiont was integrated into the developmental and metabolic processes of the host cell, leading to a stable mutual dependency.[1] Similar to mitochondria, chloroplasts maintain a dedicated genome, the plastome, which is highly conserved between plant species. It carries ~120 genes mostly encoding components of the photosynthetic machinery and the molecular machinery required for the expression of the plastome. Chloroplast gene expression is therefore required for photosynthesis, which ultimately forms the basis for all terrestrial life on earth.[2,3]

Chloroplast gene expression is carried out by dedicated molecular machineries that differ substantially from those in the nucleus or in bacteria. Transcription of the plastome requires two specialized RNA polymerases (RNAPs). A nuclear-encoded, single-subunit, phage-like RNAP (NEP) that is active primarily during early plastid development drives the expression of organellar housekeeping genes, including the plastid-encoded RNA polymerase (PEP) and some genes for other processes such as protein import or lipid synthesis.[4,5] The PEP in turn transcribes preferentially photosynthesis-related genes and tRNAs.[6] It is comprised of four plastid-encoded core subunits of bacterial origin as well as additional chloroplast-specific subunits, which are encoded in the nucleus and imported to chloroplasts.[7–10] These PEP-associated proteins (PAPs) are essential for plastid gene expression, as loss-of-function mutants exhibit albino or ivory phenotypes similar to those reported for inactivation mutants of core PEP subunits.[11–14] Furthermore, the PEP complex requires transient interaction with nuclear-encoded prokaryotic-like σ factors for promoter-specific transcription initiation.[15]

Sequence homologies suggest that the catalytic core of PEP is related to other multi-subunit RNAPs, such as those found in

bacteria, archaea, or cell nuclei.[16] Bacterial RNAPs consist of four subunits, α, β, β′, and ω. Based on sequence similarities, the PEP subunits RPOA and RPOB correspond to α and β, respectively, and they are hence also referred to as α and β. As in its presumable evolutionary ancestor, cyanobacterial RNAP, the β′ subunit of PEP is split into two proteins, RPOC1 and RPOC2, which are thus referred to as β′ and β″, respectively.[16–19] To date, no ω-like subunit has been identified in PEP. In contrast to the core subunits, the PAPs do not share sequence homology to proteins involved in other transcription systems. Sequence analyses of PAP genes suggest similarities to known protein domains with a wide range of different functions, from nucleic-acid binding over redox biology to metabolic, ambiguous, or even unknown functions.[8,20] Although some PAPs have been structurally and biochemically characterized in isolation,[21,22] their precise functions in chloroplast gene expression remain largely unknown. Intriguingly, half of the PAPs have been suggested to also localize to the nucleus, where they may be involved in the genetic control of photomorphogenesis.[18] Thus, the PAPs may also play a role in coordinating nuclear and chloroplast gene expression.

Taken together, the PEP complex is a unique molecular machinery of dual genetic and evolutionary origin. However, the molecular structure of this large complex and the role of the numerous chloroplast-specific PAPs remain unknown. Recently, a low-resolution envelope of the PEP complex was reported, which revealed its overall shape but did not allow for atomic modeling of its structure.[23] Here, we report the structure of the 19-subunit PEP complex from *Sinapis alba*, determined by single-particle cryoelectron microscopy (cryo-EM) at an overall resolution of 3.5 Å. The structure reveals the architecture of this complex and provides a high-resolution molecular snapshot of the major plastid transcription machinery. Comparison with bacterial and eukaryotic RNAPs shows that while the core polymerase fold is highly conserved, unique structural features in its periphery mediate interactions with the PEP-specific subunits. The PAPs form distinct clusters that bind around the polymerase core in a unique fashion, and their structures suggest possible functions during chloroplast gene expression. Together, these data provide a framework for understanding chloroplast gene transcription in molecular detail.

RESULTS

Structure of the PEP complex

In order to determine the structure of the PEP complex, we employed a protocol that we have previously described and shown to reproducibly yield highly purified native PEP complex preparations from *Sinapis alba* chloroplasts.[8,24] This PEP complex has been extensively characterized *in vitro* with regards to its transcriptional properties, promoter recognition, and inhibitor susceptibility.[25,26] Further, its migration behavior in native gel systems and its subunit composition were determined.[7,8] We isolated the PEP complex using this protocol (Figure S1), and confirmed the presence of all 20 previously identified subunits using mass spectrometric analysis of the purified fractions (Tables S1 and S2). We then analyzed the PEP fraction by single-particle cryo-EM, which resulted in a 3D reconstruction of the complex at an overall resolution of 3.5 Å, with focused refinement maps extending up to 2.8 Å resolution (Figures 1, S2, and S3; Table 1). This allowed us to construct a model of the PEP complex that comprises five core polymerase subunits (RPOA, RPOB, RPOC1, RPOC2, and PAP12), as well as 14 additional chloroplast-specific PEP subunits (PAP1, PAP3-11, FLN2, and pTAC18) (Figure 1; Video S1; Table S3). The model thus contains all previously identified essential PEP subunits with the exception of PAP2.[8]

The polymerase core of the PEP complex is formed by subunits RPOA, RPOB, RPOC1, RPOC2, and PAP12 and adopts a conserved crab claw-like structure typical for multi-subunit RNAPs across all domains of life.[27–33] The chloroplast-specific PAPs are arranged peripherally around the core and protrude in five different directions (Figures 1A and 1B). Based on their arrangement, we classify them into five structural clusters (Figure 1C). The first is formed by PAP1, PAP11, and PAP7, which bind at one side of the polymerase core. The second cluster is formed by PAP5 and PAP8, which wrap around the base of the polymerase. Two further structural clusters are formed by the homologous subunits PAP6 and FLN2, which each associate with one copy of PAP10 (PAP10-1 and PAP10-2, respectively). The FLN2/PAP10-2 cluster is located at the base of the polymerase, while the PAP6/PAP10-1 cluster is located on the side of the polymerase opposite from the PAP1-PAP11-PAP7 cluster. The fifth PAP-cluster protrudes upward from the core and consists of PAP3, PAP4, PAP9, and PTAC18. Taken together, the PAPs form extensive interactions with the core RNAP subunits and with each other and extend peripherally from both lobes of the polymerase. Notably, however, the front and back of the PEP, where the DNA and RNA strands are expected to enter and exit during transcription, are not obscured by PAPs.

Comparison to other multi-subunit RNAPs

The structure of the PEP complex enables us to conduct a structure-based comparison between the chloroplast transcription machinery and other multi-subunit RNAPs, in particular those found in bacteria. As expected from their high degree of sequence similarity,[23] the polymerase core subunits of PEP structurally resemble their bacterial counterparts (Figures 2A and 2B). In particular, RPOA corresponds to α, RPOB corresponds to β, and RPOC1 and RPOC2 together correspond to β′ (Figures 2A–2C and S4A–S4C). RPOA forms a homodimer that acts as a platform on which the subunits RPOB, RPOC1, and RPOC2 assemble. It is structurally and functionally analogous to the α_2 dimer of bacterial RNAP and the Rpb3/Rpb11 dimer in RNA polymerase II (RNA Pol II).[28,29] The two lobes of the crab claw are formed by RPOB and RPOC1/RPOC2, and in between them is a cleft which accommodates the active site and a DNA-RNA hybrid during transcription. The active site is formed by a conserved domain in RPOC1 (residues 378–519), which forms one of the two psi-barrels characteristic of multi-subunit RNAPs. It contains three conserved aspartate residues (D489, D491, and D493) that are expected to coordinate the two metal ions required for catalysis (Figure 2D). The cleft is spanned by the bridge helix of RPOC2, a conserved element in multi-subunit RNAPs involved in strand separation and catalysis[34] (Figures 2B and 2C). In contrast to previous structures of other multi-subunit RNAPs,[28,29] both the active site as well as

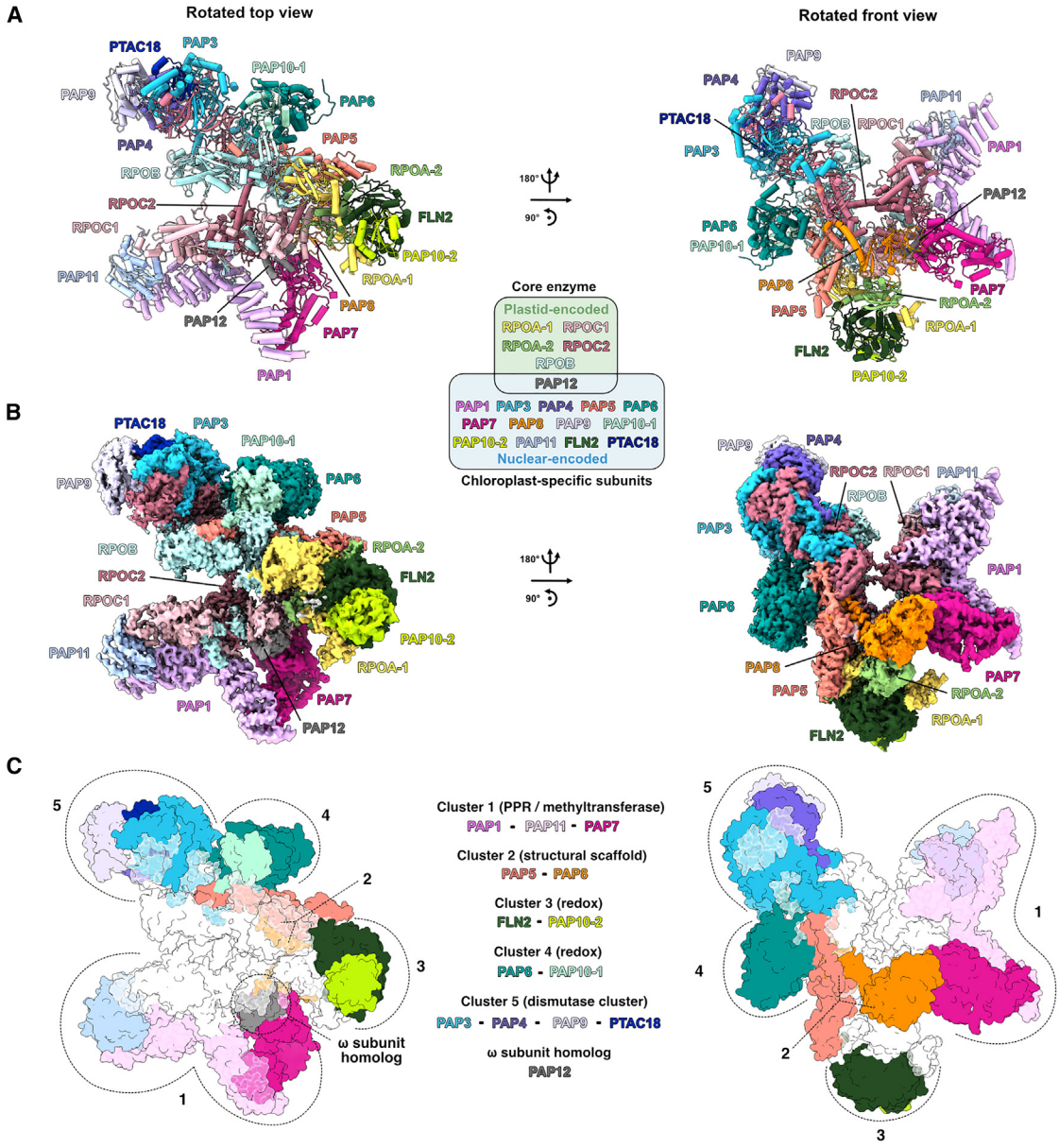


Figure 1. Structure of the *S. alba* PEP complex

(A) Cartoon representation of the PEP complex colored by subunits. Helices are depicted as cylinders.

(B) Cryo-EM map of the PEP complex (map G) colored by subunits.

(C) Schematic representation of the PEP complex with PAP clusters and omega subunit highlighted. The core RNAP subunits are colored white, while the PAPs are colored as in (A).

See also Figures S1–S3 and Table S3.

the bridge helix region are poorly resolved in our PEP reconstruction due to conformational flexibility between the two lobes of the PEP complex, and we do not observe density for the catalytic residues or bound metal ions. Thus, the PEP complex appears to be conformationally dynamic in the absence of nucleic acids. In addition to the RNAP-like subunits previously identified, PAP12 is also part of the polymerase core of PEP. It structurally corresponds to the bacterial RNAP subunit ω and the RNA Pol II subunit Rpb6 (Figures 2A, 2B, and S4D). Like ω, it binds at the base of the cleft near the C terminus of RPOC2 and may prevent its aggregation during RNAP assembly.[29,35,36]

Despite their highly conserved overall structure, the core subunits also show striking differences to other multi-subunits RNAPs (Figures S4C and S5). In particular, RPOC2 contains a large lineage-specific insertion in its trigger loop (TL) (residues 337–1,148), a conserved element within the active site of all multi-subunit RNAPs which is directly involved in catalysis.[37–42] Insertions within the TL, called lineage-specific sequence

Table 1. Cryo-EM data collection, refinement, and validation statistics

Data collection and processing				
Magnification	81,000×			
Voltage (kV)	300			
Electron exposure (e^{-}/Å^2)	40			
Defocus range (μm)	0.3–2.0			
Pixel size (Å)	1.05			
Symmetry imposed	C1			
Initial particle images (no.)	3,425,348			
	Map A; consensus	Map B; RPOB lobe	Map C; RPOC lobe	
EMDB ID	18499	18500	18501	
Final particle images (no.)	46,170	123,874	46,170	
Map resolution (Å)	3.46	2.95	3.52	
FSC threshold	0.143	0.143	0.143	
Map sharpening *B* factor (Å^2)	−58.3	−59.9	−71.0	
	Map D; foot	Map E; central RPOB lobe	Map F; PAP3-cluster	Map G; composite
EMDB ID	18502	18503	18504	18496
Final particle images (no.)	123,874	123,874	123,874	–
Map resolution (Å)	3.34	2.89	2.83	–
FSC threshold	0.143	0.143	0.143	–
Map sharpening *B* factor (Å^2)	−91.9	−74.7	−68.0	–
Refinement	PDB: 8QMA			
Model resolution (Å) (local resolution-filtered map A)	3.5			
FSC threshold	0.143			
Model composition				
Non-hydrogen atoms	52,078			
Protein residues	6,421			
Ligands	ZN: 1, FE: 2, SAM: 1			
Mean *B* factors (Å^2) (from local resolution-filtered map A)				
Protein	111.98			
Ligand	117.74			
RMSDs				
Bond lengths (Å)	0.004			
Bond angles (°)	0.692			
Validation				
MolProbity score	1.61			
Clashscore	5.82			
Poor rotamers (%)	0.74			
Ramachandran plot				
Favored (%)	95.75			
Allowed (%)	4.03			
Disallowed (%)	0.22			

insertion 3 (SI3), can be found in several bacterial lineages, including *E. coli* and cyanobacteria.[16] In *E. coli*, SI3 is composed of two sandwich-barrel hybrid-motif (SBHM) domains and is involved in the regulation of RNAP activity.[43,44] In cyanobacterial RNAP, SI3 is much larger and consists of eight SBHM motifs.[45] Recent structural data show that it adopts an elongated, arch-like architecture, which spans across the polymerase surface and contacts σ factor during transcription initiation.[19,45] In the PEP, SI3 is even larger and adopts a unique structure. We observed density for five ordered domains within SI3 (Figures 2C and 2E). Three of them correspond to the "tail," "fin," and parts of the "body" of cyanobacterial SI3, respectively, and occupy locations similar as these on the polymerase (Figures S4C, S5A, and S5B).[19,45] The tail domain (SHBM1,

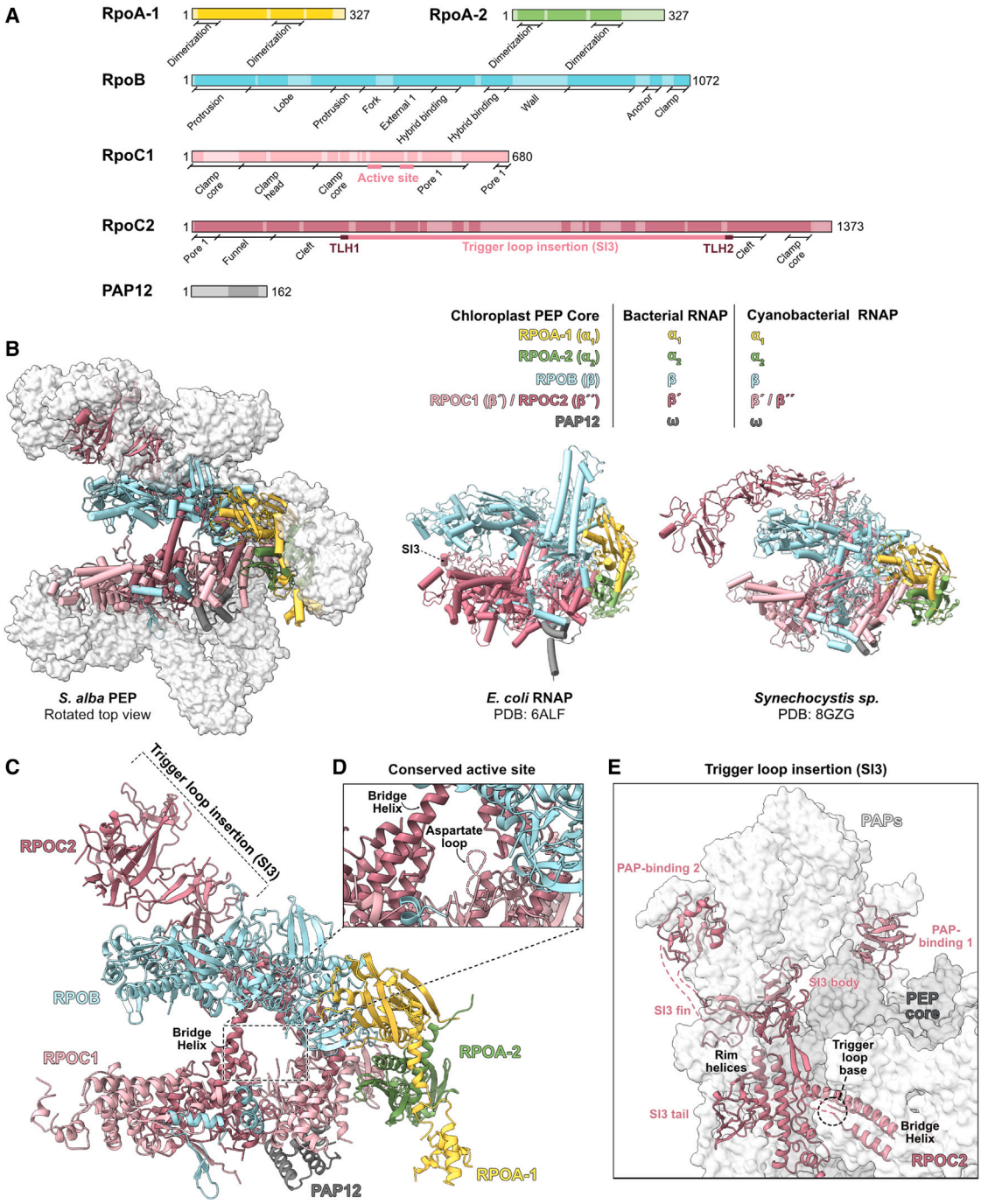


Figure 2. Structure of the polymerase core of PEP

(A) Schematic bar representation of core RNAP subunits with functional domains annotated based on the structure of RNA Pol II.[28] Regions that were not visible in the structure are indicated transparently.

(B) Structure-based comparison of the polymerase core of PEP (left), with prokaryotic RNAPs from *E. coli* (center) (PDB: 6ALF [34]), and *Synechocystis* sp. (cyanobacterium) (right) (PDB: 8GZG [19]). The core subunits of PEP are shown as cartoon while the PAPs are shown as transparent gray surface.

(C) Cartoon representation of the polymerase core of PEP (RPOA-1 and -2, RPOB, RPOC1, RPOC2, and PAP12) with structural details indicated.

(D) Close-up view of the active site of PEP. The loop containing the conserved catalytic aspartate residues D489, D491, and D493 is depicted with a dashed line.

(E) Close-up view of the ordered domains in the trigger loop insertion (SI3) of PEP. The interacting PAPs are shown as transparent gray surfaces.

See also Figures S3–S5 and Table S3.

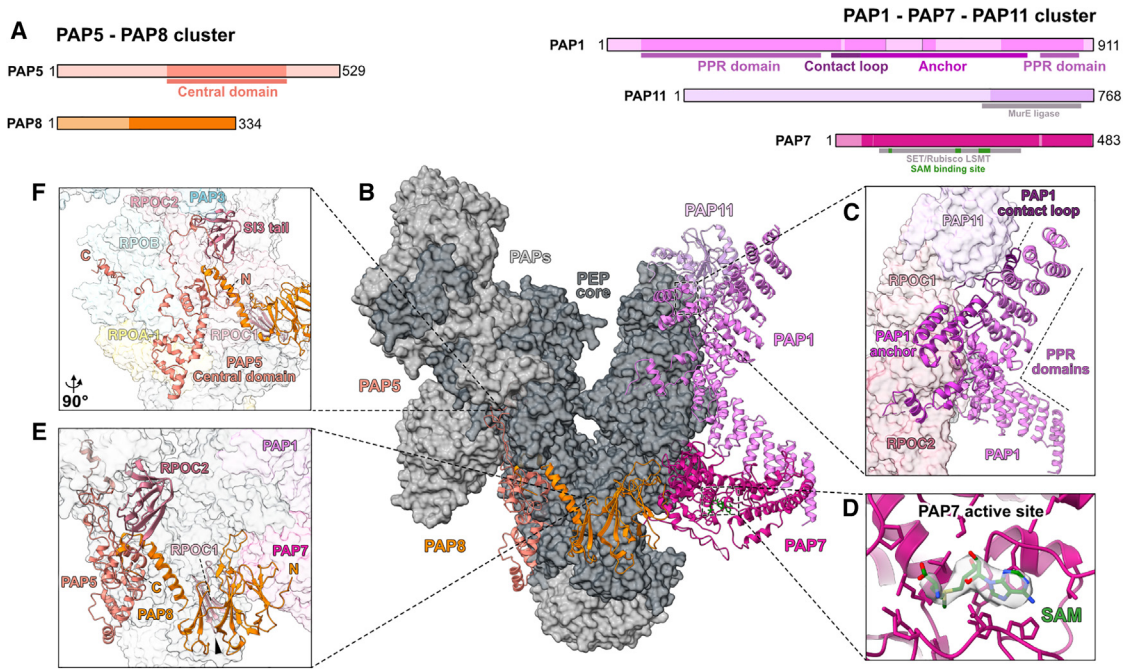


Figure 3. PAP1-PAP7-PAP11 and PAP5-PAP8 clusters

(A) Schematic depiction of PAP subunits as in Figure 2A. Regions that were not visible in the structure are indicated transparently. Structural domains are indicated as solid lines below the primary sequence diagram.

(B) Structures and locations of the PAP1-PAP7-PAP11 and PAP5-PAP8 clusters in the PEP complex. Both clusters are shown as cartoons, while the polymerase core subunits are shown as dark gray surface and the remaining PAPs as light gray surface.

(C) Close-up view of PAP1 with structural features annotated. Neighboring subunits are shown as transparent surface colored as in Figure 1.

(D) Close-up view of the PAP7 active site. The co-purified SAM cofactor and residues within 5 Å distance are shown as sticks. The cryo-EM density for the ligand is shown as transparent surface (map G).

(E) Detailed view of PAP8 interactions. Neighboring subunits are shown as transparent surface and colored as in Figure 1. PAP8 forms an inter-subunit β sheet with RPOC1 (indicated with the black arrowhead) and interactions with the SI3 tail domain in RPOC2.

(F) Detailed view of PAP5 interactions. Neighboring subunits are shown as transparent surface colored as in Figure 1. PAP6 was omitted for clarity. The PAP5 central domain contacts multiple other subunits.

See also Figure S6 and Table S3.

residues 351–425) is located next to the rim helices of β′, near the secondary channel of the RNAP, which serves as an entry site for substrate nucleoside triphosphates (NTPs).[29,46,47] The fin (SHBM2, residues 435–484; SHBM8, residues 944–952 and 1,064–1,136) and the first part of the body (SHBM3, residues 485–503 and 921–943) sit on top of the rim helices and occupy a similar location as Rpb9 in RNA Pol II.[28] The other two resolved domains of SI3 adopt unique folds that differ from those observed in *E. coli* or cyanobacteria. One of them is part of the body (PAP-binding domain 1, residues 558–618 and 793–877), while the other is an insertion within the fin (PAP-binding domain 2, residues 973–1,058). The remaining segments of SI3 are not clearly visible in our reconstruction, suggesting conformational flexibility. However, low-pass-filtered reconstructions suggest that they may extend toward the protrusion of RPOB, as observed in cyanobacterial SI3 (Figure S6A).[19,45] Strikingly, all five ordered domains of SI3 interact extensively with PAPs. The fin, tail, and body segment interact with PAP3 and form a platform that tethers the PAP3-PAP4-PAP9-pTAC18 cluster to the polymerase. The tail additionally contacts PAP5 and PAP8, and the fin and body interact with PAP6. The two unique domains of SI3 form extensive interactions with PAP3, and PAP-binding domain 2 also interacts with PAP4. Thus, the structural comparison shows that the polymerase core is conserved between PEP and other multi-subunit RNAPs, but contains additional features that mediate interactions with the chloroplast-specific PAP clusters.

PAP1, PAP7, and PAP11 form a multi-functional hub on the side of PEP

The first and largest PAP cluster is formed by PAP1, PAP7, and PAP11, and tethers structurally diverse factors to the side of the polymerase. PAP1 is located on the RPOC side of the polymerase, next to the RNA exit channel, where it forms a horn-shaped concave structure (Figures 3A and 3B). Its N-terminal region is composed of a pentatricopeptide repeat (PPR) domain with 9 PPRs (residues 64–400), and its C-terminal region contains two additional PPRs (residues 811–883). In between the PPR domains, PAP1 contains a unique domain that we term "anchor." This domain contains a "contact loop," which binds PAP11 and RPOC1 (residues 420–474), and a region with which PAP1 is anchored to RPOC2 (residues 475–787) (Figure 3C). Large parts of the anchor domain are not visible in our reconstruction, suggesting that they are mobile. This includes a region previously

suggested to adopt a SAP domain fold (residues 535–567)[48] as well as a large segment which may reach toward the polymerase cleft (residues 617–738). The ordered parts of this domain wrap around the clamp of the polymerase, and occupy a position close to the jaw in bacterial RNAP and Rpb5 in RNA Pol II, respectively (Figure S6D). PPR proteins comprise a large family of RNA-binding proteins involved in different steps of organellar gene expression, ranging from RNA editing and processing to translation.[49,50] Although PAP1 is situated on the side of the polymerase where RNA is expected to exit during transcription, it is not in direct vicinity of the RNA exit channel. In addition, it does not display a pronounced positively charged surface, as could be expected for a nucleic-acid-binding protein (Figure S6B). Thus, it is unclear whether PAP1 directly interacts with RNA during chloroplast transcription. Structural comparison shows that PAP1 occupies a comparable location as the stalk in RNA Pol II, which interacts with co-transcriptionally acting RNA processing enzymes[51] (Figure S6D). Thus, PAP1 may similarly act as a hub for binding additional factors that could mediate co-transcriptional processes. In particular, cross-linking data indicate that PAP1 interacts with PAP2.[23] Consistent with this, we observe additional density close to the N-terminal part of PAP1 in low-pass-filtered maps, which may correspond to PAP2 (Figure S6C). However, we refrained from modeling into this density due to its limited quality. Figure S6

PAP7 acts as a bridge between PAP1 and the polymerase core (Figure 3B). It forms multiple contacts with the first three PPRs of PAP1 on one side, and with RPOC2 and PAP12 on the other. PAP7 adopts a SET domain methyltransferase fold commonly found in histone lysine methyltransferases.[52] Structural similarity searches suggest that it is most similar to SETD3,[53] which methylates histidine residues in actin, and the Rubisco large subunit methyltransferase (LSMT) (Figure S6E). In its putative active site, we observe density that is consistent with a co-purified S-adenosylmethionine (SAM) molecule, suggesting that PAP7 may indeed act as a methyltransferase in the PEP complex (Figure 3D).

PAP11 binds on top of PAP1, cradled in between the contact loop of PAP1 and the clamp head of RPOC1 (residues 143–195) (Figures 3B and 3C). We observe clear density only for the C-terminal domain of PAP11 (residues 575–766). However, superposition of the full-length AlphaFold model of PAP11 onto the PEP complex shows that the entire protein could be accommodated without major clashes and positions the unmodeled parts of PAP11 within a large extra density visible in low-pass filtered maps of the complex (Figure S6F). This indicates that the N-terminal part of PAP11 is present but likely flexible. Based on sequence comparisons, PAP11 has been suggested to resemble bacterial Mur ligases, in particular MurE, which catalyze the addition of amino-acid linkers to the sugar backbone of peptidoglycans during bacterial cell-wall synthesis.[54,55] Consistent with this, the resolved domain of PAP11 adopts a similar fold as domain 3 of MurE, which is involved in substrate binding (Figure S6G). However, residues important for substrate binding and catalysis in Mur ligases are only partially conserved in PAP11.[55] Thus, it is unclear whether PAP11 fulfills a similar enzymatic function in the context of chloroplast transcription.

Taken together, the PAP1-PAP7-PAP11 cluster may integrate multiple enzymatic activities and could additionally serve as a platform for the binding of additional factors that mediate co-transcriptional processes.

PAP5 and PAP8 are unique structural subunits of the PEP complex

The second cluster, comprising PAP5 and PAP8, wraps around the base of the polymerase core and forms multiple interactions with both the core and other PAPs (Figure 3B). Only the C-terminal domain of PAP8 is resolved in the PEP reconstruction, which sits in front of the secondary channel anchored to the polymerase core through an inter-subunit beta sheet formed with RPOC1 (Figure 3E). It adopts a structurally unique fold, but occupies a similar position as Rpb8 in RNA Pol II.[28] PAP8 additionally interacts with PAP7 through an extended loop (residues 199–226) and with PAP5 via an extended helical element (residues 302–334). The latter is located C-terminally of the ordered domain and runs along RPOC1 to the base of the rim helices in RPOC2. PAP5 binds next to the rim helices, at the opposite side of the polymerase cleft from PAP7. Large parts of it appear to be mobile, as only its central domain shows clear density (residues 206–429). Similar to PAP8, it adopts a unique fold with no detectable similarity to known proteins. It is anchored to the polymerase core through interactions with RPOA-1, RPOC2, and RPOB, and is flanked by extended regions that meander along the polymerase core (Figure 3F). The N-terminal segment (residues 206–254) runs upward along the tail domain of SI3 in RPOC2 and contacts PAP3 and PAP6, while the C-terminal segment runs along the side of the polymerase between RPOB and PAP6.

In summary, these observations confirm recent interaction data[56] and suggest that PAP5 and PAP8 fulfill structural roles in the PEP complex by facilitating interactions between the core polymerase subunits and more peripheral PAPs.

Redox enzymes are coupled to the chloroplast transcription apparatus

The third and fourth clusters, formed by PAP6/PAP10-1 and FLN2/PAP10-2, are structurally highly similar but are located at different locations on the core enzyme (Figures 1A, 4A, and 4B). The FLN2/PAP10-2 dimer is anchored to the bottom of the complex through interactions with the C-terminal domain of RPOA-1 (residues 236–304), while the PAP6/PAP10-1 dimer is located at the RPOB-side of the PEP complex and interacts with RPOB, PAP5, and PAP3 (Figure 4B). PAP6 and FLN2 are paralogs and hence closely resemble each other (Figure 4C). They are structurally related to the pfkB carbohydrate kinase family, in particular fructokinases (FRKs).[22,57] However, PAP6 and FLN2 in the PEP complex adopt a more closed conformation than FRKs, which would likely preclude substrate binding (Figure S7A). In addition, the lid domain, which closes upon substrate binding by *Vibrio cholerae* FRK,[58] adopts a different conformation in PAP6 and FLN2 and mediates interactions with PAP10. Compared with FRKs, both PAP6 and FLN2 contain insertions that are involved in interactions with core polymerase subunits of PEP and with other PAPs (Figure S7B). These insertions differ between the two proteins, explaining why they can bind at distinct locations within the PEP complex and cannot fully complement each other.[57] These data suggest that the FRK fold has been repurposed to serve a structural role in the PEP complex, consistent with previous biochemical findings.[22,57]

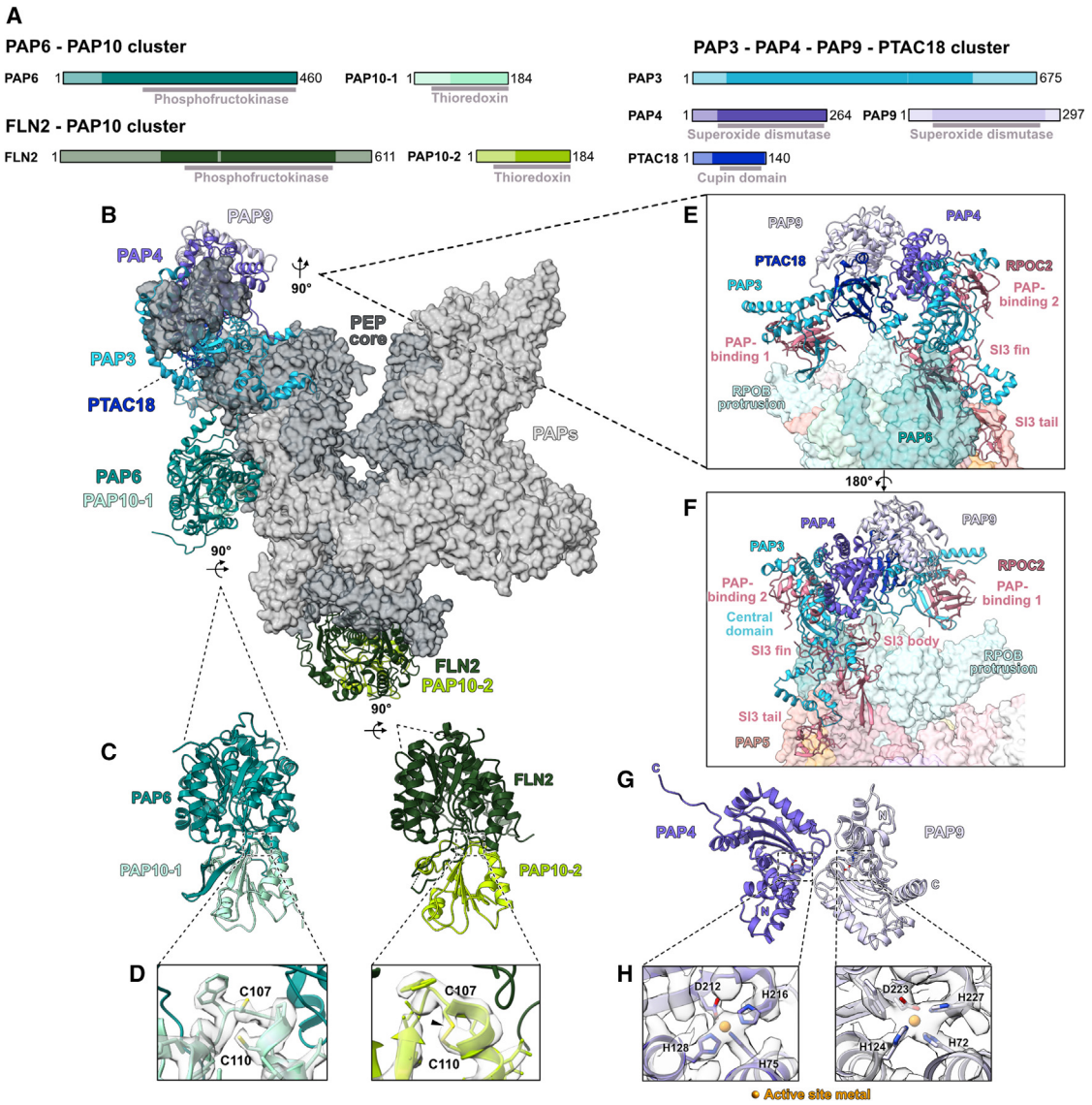


Figure 4. Redox and superoxide dismutase clusters

(A) Schematic depiction of PAP subunits as in Figure 2A. Regions that were not visible in the structure are indicated transparently. Structural domains are indicated as solid lines below the primary sequence diagram.

(B) Structures and locations of the PAP6-PAP10-1, FLN2-PAP10-2, and PAP3-PAP4-PAP9-PTAC18 clusters in the PEP complex. The clusters are shown as cartoons, while the polymerase core subunits are shown as dark gray surface and the remaining PAPs as light gray surface.

(C) Detailed view of the PAP6-PAP10-1 and FLN2-PAP10-2 clusters.

(D) Close-up view of the boxed region in (C) highlighting the conserved cysteine residues (C107 and C110) in PAP10. The cryo-EM density (map G) is shown as transparent surface. In PAP10-1, the cryo-EM density indicates reduced cysteine residues while in PAP10-2 it is consistent with a disulfide-bridge (black arrowhead).

(E) Rotated detailed view of the PAP3-PAP4-PAP9-PTAC18 cluster. The RPOC2 SI3 is shown as cartoons with PAP3-interacting domains annotated. Surrounding PAPs are shown as transparent surface and colored as in Figure 1. PAP3 anchors on top of SI3 tail and fin and is clamped by the PAP-binding domains 1 and 2, with which it forms inter-subunit secondary structures.

(F) Rotated view of (E).

(G) Detailed view of the PAP4-PAP9 SOD heterodimer.

(H) Close-up view of PAP4 and PAP9 conserved active site residues. The cryo-EM density (map G) is consistent with a coordinated metal ion, depicted as orange sphere. See also Figure S7 and Table S3.

PAP6 and FLN2 each interact with one molecule of PAP10, the only member of the conserved plastid-specific thioredoxin family TrxZ.[22] PAP10 adopts a thioredoxin fold with two redox-active cysteine residues (C107 and C110) (Figure 4D). While the oxidation state of these residues cannot be unambiguously inferred from the structural data, focused refinements around these clusters

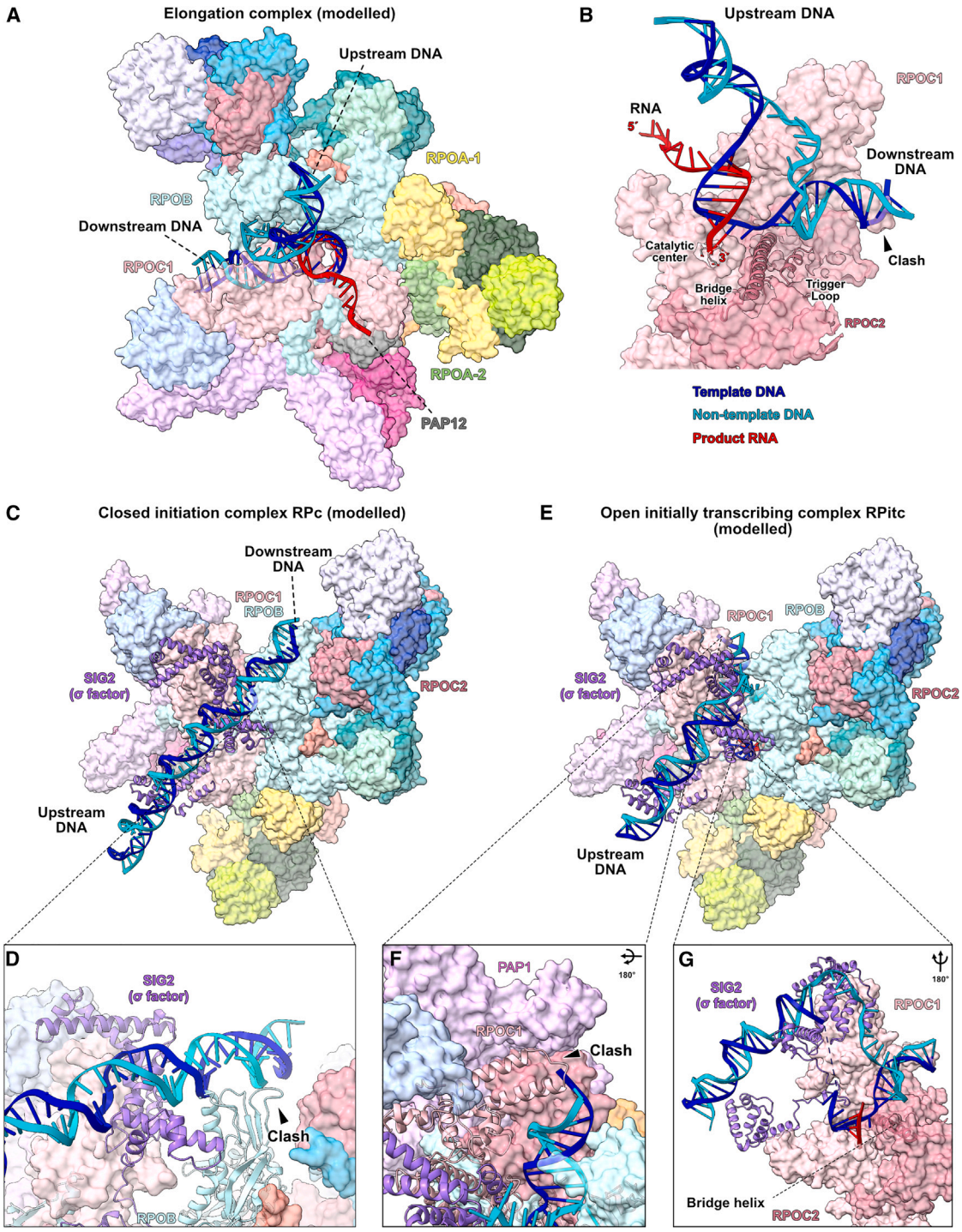


Figure 5. Models of active chloroplast transcription complexes

(A) Model of an actively elongating PEP complex (EC). The nucleic acid was positioned based on structural superposition with the cyanobacterial EC (PDB: 8SYI [45]). The PEP is shown as transparent surface colored as in Figure 1, and the nucleic acids are shown as ribbons. The template strand is shown in blue, the nontemplate strand in cyan, and the RNA in red. No large structural rearrangements or loss of factors are necessary to form an elongation complex.

(B) Details of the active site in the modeled EC. The bridge helix is shown as a cartoon and the active site aspartates as sticks. The nucleic acid can bind to the PEP complex without major clashes in the active site. The downstream DNA clashes with an extended helical segment in RPOC1, which may undergo conformational movements upon DNA binding.

(C) Model of a closed initiation complex (RPc). The model was constructed by superimposing an AlphaFold model of the *S. alba* SIG2 and the PEP complex with an *E. coli* RPc (PDB: 6PSQ [62]). The PEP is shown as a transparent surface colored as in Figure 1. SIG2 is shown as a cartoon in purple. The nucleic acids are

(legend continued on next page)

indicate that the two cysteines may form a disulfide linkage in PAP10-2, while for PAP10-1 the density is more consistent with two reduced cysteine residues (Figure 4D). Structural comparisons show that PAP6 and FLN2 occupy the expected substrate binding surface of PAP10, resulting in the redox-active cysteines being buried at the protein-protein interface (Figure S7C). Previous biochemical studies have suggested several cysteine residues in PAP6 and FLN2 to be involved in inter-subunit disulfide formation with PAP10.[22] However, these are not in close vicinity to the redox-active cysteines of PAP10 in the PEP complex. This suggests that FLN2 and PAP6 are likely not direct redox substrates of PAP10 in the context of the PEP complex, and instead act to tether PAP10 to the PEP complex.

Overall, the structure reveals that two potentially redox-active enzymes are part of the chloroplast transcription machinery and are associated with the PEP complex in a unique fashion through interactions with a repurposed FRK fold.

SODs may protect the PEP from oxidative damage

The fifth PAP cluster, formed by PAP3, PAP4, PAP9, and PTAC18, is located above the PAP6-PAP10-1 cluster and is anchored to the PEP complex primarily through interactions between PAP3 and SI3 in RPOC2 (Figures 4A, 4B, 4E, and 4F). The N-terminal region of PAP3 interacts with the SI3 tail, fin, and body domains, while its central domain forms intertwined structures with the two PAP-binding domains (Figures 4E and 4F). In addition, it also directly contacts one of the PAP6 extensions. Thus, PAP3 forms a structural scaffold that anchors this PAP cluster to the polymerase core and on which the other PAPs assemble.

PAP4 and PAP9 bind on top of PAP3 (Figures 4B and 4E). Both belong to the family of manganese/iron superoxide dismutases (SODs) and are structurally highly similar. The structure of PAP9 in the PEP complex is virtually identical to a recently reported crystal structure of this subunit in isolation[21] (root-mean-square distance [RMSD] 0.798 Å) (Figure S7D). As suggested previously,[59] PAP9 and PAP4 form a heterodimer in the PEP complex. They adopt an identical relative orientation to each other as two monomers of PAP9 in the crystal structure (Figure S7E), explaining why they can partially substitute for each other in single-knockout plants while double knockouts lead to albino phenotypes.[59] It has previously been suggested that PAP9 possesses a C-terminal disordered tail, which is supposedly absent in PAP4.[21] In the PEP complex, the C-terminal tail of PAP9 is not visible; instead the C terminus of PAP4 forms an elongated peptide that makes additional contacts with PAP3 (Figure 4G). SODs typically coordinate manganese or iron ions via conserved histidine and aspartate residues.[60] In both PAP4 and PAP9, these active site residues are present and the density is consistent with a coordinated metal ion (Figure 4H), supporting previous studies demonstrating their SOD activity *in vitro*.[59] The structure thus suggests that PAP4 and PAP9 confer SOD activity to the PEP complex.

The final constituent of this cluster is PTAC18, which interacts primarily with the C-terminal part of PAP3 (Figure 4E). It adopts a cupin-like fold, a functionally diverse group of proteins that are widespread in plants.[61] This subunit was only recently discovered as part of the PEP,[23] and no functional data is currently available. However, its peripheral location in the complex suggests that it may facilitate transcription-coupled processes rather than influencing the catalytic activity of the PEP directly.

In summary, this PAP cluster may confer superoxide-detoxification activity to the PEP complex and could serve as a hub for interactions with further proteins.

Model for chloroplast transcription

To investigate whether the PAPs could modulate transcription by directly interacting with nucleic acids, we performed structural comparisons to model active states of the complex. Superimposition of PEP with the recent structure of a cyanobacterial transcription elongation complex (EC)[45] reveals that the DNA and RNA can be accommodated without major clashes (Figure 5A). As expected, the DNA-RNA hybrid is positioned in the cleft with the 3′ end of the RNA close to the active center (Figure 5B). The binding of DNA may be accompanied by conformational movements in the polymerase core, as the downstream duplex would otherwise clash with an extended helical segment in the clamp (RPOC1 residues 223–235) (Figure 5B). The downstream DNA duplex may also contact the central domain of PAP1, which occupies a similar position as Rpb5 in RNA Pol II (Figure S6D). The RNA emerges next to PAP1 and PAP7, which could thus facilitate co-transcriptional processes. Taken together, the comparison indicates that nucleic acid interactions during transcription elongation are mainly mediated by the core polymerase subunits of PEP and that no large conformational changes in the PEP are required to form the active complex.

Multi-subunit RNAPs require the action of auxiliary factors to initiate transcription from promoter DNA. In bacteria, this is mediated by σ factors, which interact with conserved promoter *cis* elements and facilitate formation of a closed initiation complex (RPc), which can subsequently transition into an open, initially transcribing complex (RPitc).[63] Chloroplasts contain several different σ factors, which are responsible for correct promoter recognition and transcription initiation.[18] To model the structure of PEP initiation complexes, we superimposed PEP and an AlphaFold model of the *S. alba* sigma factor SIG2 with structures of an *E. coli* RPc and a cyanobacterial RPitc[19,62] (Figures 5C–5G). The models show that in both cases, SIG2 as well as the promoter DNA could be accommodated without major clashes and that none of the PAPs are located in direct vicinity to the promoter DNA. Thus, initiation complexes could likely be

depicted as in (A). Disordered regions of SIG2 with low predicted local distance difference test (pLDDT) scores were omitted. Both SIG2 and DNA can be accommodated without large conformational rearrangements or loss of factors.

(D) Close-up view of the minor clash between the RPOB protrusion and the DNA duplex in the RPc.

(E) Model of an open initially transcribing complex (RPitc). The model was constructed by superimposing an AlphaFold model of the *S. alba* SIG2 and the PEP complex with a cyanobacterial RPitc (PDB: 8GSG [19]). The PEP is shown as transparent surface colored as in Figure 1. SIG2 is depicted as in (C). Both SIG2 and DNA can be accommodated without large conformational rearrangements or loss of factors.

(F) Close-up view of the minor clash between RPOC1 and the DNA duplex in the RPitc. The clash is identical to that seen in the EC in (B).

(G) Close-up view of the active site in the modeled RPitc. SIG2 and nucleic acids can be accommodated without major clashes.

See also Figure S6.

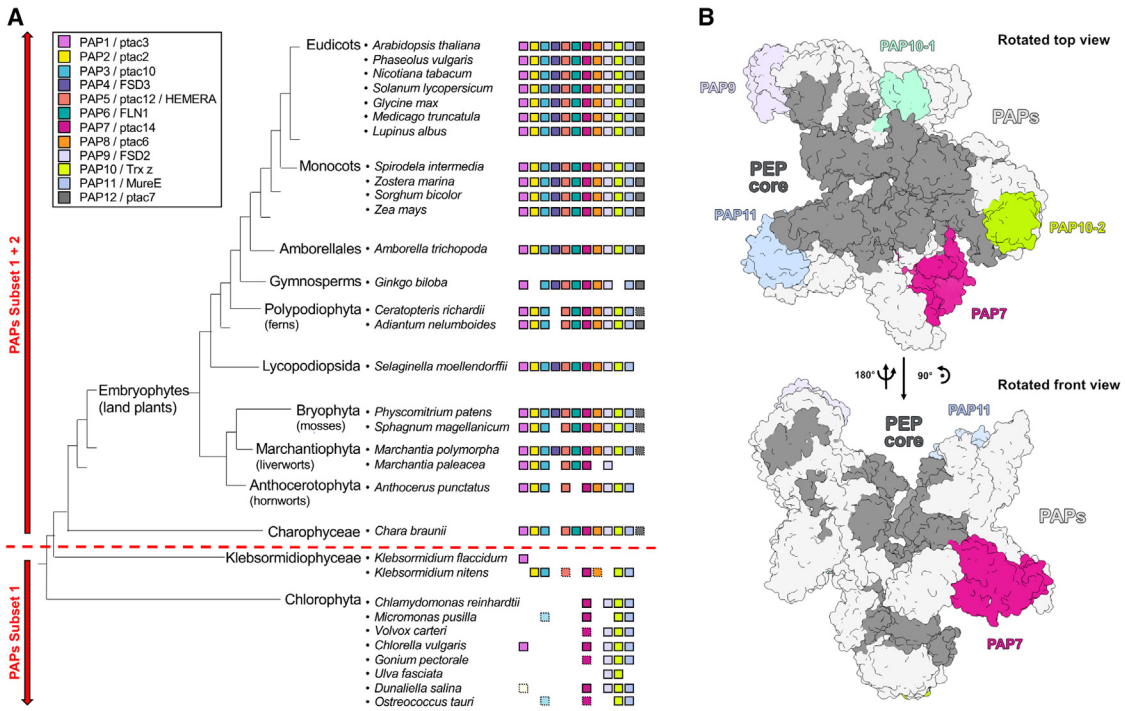


Figure 6. Distribution of PAP orthologs within the Chloroplastida

(A) Cladogram representing the evolutionary tree of Chloroplastida and the appearance of *Arabidopsis pap* gene orthologs among major clades. Orthologs were identified by reciprocal best hit search. Each PAP is represented by a square with a unique color code as in Figure 1, (see legend in top left corner). Symbols in fading colors with a dashed line depict orthologs with e values higher then 1e−10. Annotations of gymnosperm sequences such as *Picea* and *Pinus* were found to be of low quality, thus only data for *Ginkgo biloba* are given here (for details see Table S4). The topology of the tree follows earlier reports.[66]

(B) Schematic depiction of the *S. alba* PEP complex. Core RNAP subunits are colored in dark gray and PAPs in light gray. Subset 1 PAPs are highlighted and colored as in (A).

See also Table S4.

formed without large structural rearrangements or loss of PAPs. The only exception is a minor clash between the DNA and the protrusion of RPOB in the RPc and between the DNA and the same helical element of RPOC1 as in the EC also in the RPitc, suggesting small conformational changes upon DNA binding (Figures 5D and 5F).

In sum, these comparisons enable us to propose plausible models for active chloroplast transcription complexes. Surprisingly, the models do not show obvious interactions between the PAPs and nucleic acids, suggesting that they do not modulate transcription directly. Consistent with this, surface charge analysis of the PEP complex shows that while the DNA/RNA-binding cleft is predominantly positively charged, the remaining surface shows a relatively even charge distribution with only few clusters of positive charge that are not in the vicinity of the modeled nucleic acids (Figure S6H).

The presence of PEP complex components is conserved across land plants

The unique structure of the PEP complex and the composition and arrangement of PAPs raises the question of how conserved the observed architecture of the chloroplast transcription apparatus is. To address this, we performed homology searches in publicly available databases in order to investigate the presence of PAP orthologs in other species within the green lineage of plastid evolution (Figure 6A; Table S4). This analysis revealed the presence of full sets of the 12 PAPs within all land plants, including vascular species and bryophytes. In contrast, green algae possess only a subset of PAPs that occur in some species-specific combinations. This comparison shows that PAP7, PAP9, PAP10, and PAP11 form a relatively conserved subset that we term "subset 1," which is present in both lineages and may thus have evolved before the separation of Chlorophyta and Streptophyta. The evolution of the PEP complex in these two groups then apparently diverged, consistent with a previously reported unusual organization of the plastid *rpo* genes in *Chlamydomonas reinhardtii*.[64] The complete set of PAPs, comprising in addition to subset 1 also PAP2, PAP3, PAP5, PAP6, PAP8, PAP11, and PAP12 (subset 2), occurs first in the streptophyte alga *Chara braunii*, which is part of the Zygnematophyceae/Charophyceae/Coleochaetophyceae (ZCC) grade in which polyplastidy evolved.[65] Intriguingly, these evolutionary PAP subsets do not correlate to the PAP clusters identified in the structure of the PEP complex (Figure 6B). This suggests that the evolutionary subsets do not represent different stages of structural evolution of the complex. Instead, it is possible that subset 1 PAPs evolved first as a functional adaptation to the specific redox conditions within chloroplasts and were integrated into PEP later on.

Taken together, our evolutionary analysis suggests that the genes for PAPs evolved before the appearance of flowering

plants, during the terrestrialization event when streptophyte algae conquered land, and that the PEP complex with its PAPs was established very early in plant evolution. The architecture of the PEP complex reported here is therefore likely conserved across most (if not all) land plants, but may differ in green algae.

DISCUSSION

Here we present the high-resolution structure of the multi-subunit transcription machinery in chloroplasts. It reveals the architecture of this complex and provides detailed insights into the structure and arrangement of both the conserved catalytic core as well as the chloroplast-specific PAPs.

The structure shows that the polymerase core of the PEP complex is structurally similar to that of other multi-subunit RNAPs, consistent with its evolutionary origin from a cyanobacteria-like ancestor.[1] It provides evidence that the RPOA subunit indeed forms a dimer in the complex, an assumption inferred from the bacterial RNAP structure but not demonstrated so far. The structure also reveals that PAP12 is a homolog of the bacterial ω subunit. Such a homolog was long-sought in plants but never identified and it was widely believed to be lost during evolution. Compared with bacterial RNAP, the PEP core subunits contain several insertions, in particular a large insertion within the TL (SI3). The structure shows that the SI3 of PEP adopts a unique structure that differs from that observed in other RNAPs and that it has been adopted to mediate interactions with the chloroplast-specific PAPs. In addition to this tethering function, this may also provide a mechanism for relaying structural changes from the active site of RNAP to the PAPs at the periphery and vice versa. The TL undergoes refolding during transcript elongation,[67] which could in turn also cause rearrangements of SI3 and associated PAPs. In support of this, recent structures of cyanobacterial RNAP, which is likely the evolutionary ancestor of PEP and also contains a large TL insertion, show that the binding of an incoming NTP leads to substantial movements of SI3.[45] Thus, structural changes that occur during transcription in the active site of PEP could potentially lead to large-scale rearrangements of the complex. In addition, SI3 may also facilitate interactions with additional auxiliary transcription factors. In the recently reported structure of the cyanobacterial RNAP holoenzyme, SI3 forms an arch across the polymerase body and contacts σ factor via its head domain.[19] In our PEP structure, the head domain and several other parts of SI3 are mobile, but may become more ordered during distinct stages of the transcription cycle.

The structure of the PEP complex also provides detailed structural and molecular insights in the chloroplast-specific PAPs and rationalizes a large body of previous genetic and biochemical data.[13,17,18,20,68] We observe that the PAPs form distinct clusters, which bind at different locations around the polymerase core and integrate PAPs with distinct functions. The PEP complex appears to be unique among all known multi-subunit RNAPs, as for most PAPs, neither their structure nor their location on the polymerase core resemble known polymerase-associated factors from other transcription systems, making it difficult to draw functional analogies. For many PAPs, potential functions within the PEP complex have been predicted based on sequence analysis and *in vitro* data.[8,14,22,48,56,59,69,70] The structure of the PEP complex provides insights regarding which of the PAPs may indeed possess these proposed activities. For example, PAP7 adopts a methyltransferase fold and co-purifies with a cofactor molecule bound in the active site, suggesting that it is a *bona fide* methyltransferase. However, its methylation substrate remains to be determined. PAP4 and PAP9 appear to be active SODs, which could potentially serve to protect the PEP complex against reactive oxygen species arising from the photosynthetic light reactions in the nearby thylakoid membrane. The situation is less clear for the thioredoxin PAP10 and its interacting partners, PAP6 and FLN2. While previous *in vitro* studies demonstrated that isolated recombinant PAP10 indeed possesses redox activity,[22] the structure of the PEP complex shows that both copies of PAP10 are tightly associated with PAP6 and FLN2, respectively, and the redox-active cysteines are buried within the interaction surface with no obvious substrate residues in close vicinity. This suggests that neither PAP6 nor FLN2 are direct redox substrates of PAP10. Consistent with this, genetic studies demonstrate that mutagenesis of the redox-active cysteines does not prevent chloroplast transcription and proper organelle biogenesis.[71] Nevertheless, it is possible that structural rearrangements in PEP that occur during the transcription cycle may lead to the dissociation of PAP10, thus liberating it for interactions with other substrates. However, transiently expressed PAP10 has been shown to lead to dense plastid foci, suggesting that PAP10 likely does not exist as a monomeric soluble protein the plastid stroma.[72] This may differ in *Chlamydomonas reinhardtii*, where no direct orthologs of FLN2 and PAP6 can be detected, and TrxZ/PAP10 has been proposed to have regulatory functions within the Calvin-Benson-Bassham cycle (Figure 6).[73] PAP10 has been proposed to play a role in the light-dependent activation of PEP transcription, which may involve redox-mediated phosphorylation events.[22,74] This is particularly intriguing, as both FLN2 and PAP6 are structurally related to FRKs, which phosphorylate carbohydrates. While structural comparisons suggest that the conformation of both FLN2 and PAP6 would not allow substrate binding as observed in FRKs, this raises the possibility that redox-triggered rearrangements in the PEP complex may render them active. Taken together, it remains unclear whether PAP10 acts as a redox enzyme or sensor in the context of chloroplast transcription, or whether the thioredoxin and FRK folds have been repurposed for a different function in the PEP complex. Similarly, it remains to be determined whether PAP11 harbors a MurE-like catalytic activity in the context of the PEP. Cross-species complementation experiments suggest that it is not involved in peptidoglycan synthesis in *A. thaliana*, pointing toward a repurposing of this fold in the chloroplast transcription apparatus.[55] Notably, such repurposing of functionally unrelated folds has also been observed in the human mitochondrial transcription system,[75,76] and may thus be a common feature of organellar transcription systems. Very few hypotheses on the function of PAP3, PAP5, and PAP8 have been proposed, as they show very little sequence homology to other proteins.[18] The structure of the PEP complex reveals that these PAPs adopt unique structures and suggests that they fulfill architectural roles in the complex. In summary, the structure of the PEP complex provides detailed insights into

the structures and potential functions of the PAPs, which will enable rational design of genetic and functional studies to test these hypotheses.

The structure also allows us to propose models for actively transcribing and initiating PEP complexes. This analysis reveals that most PAPs are located far away from where nucleic acids are expected to bind during transcription, suggesting that they are likely not directly involved in the process of transcription. This is consistent with early models proposing a structural role in complex stability rather than a functional involvement of PAPs in transcription.[13] An exception to this hypothesis is PAP1, which binds next to the RNA exit channel and may interact with additional factors and/or with downstream DNA. For the remaining PAPs, their peripheral arrangement in distinct functional modules is consistent with the hypothesis that PAPs have been acquired to adapt the transcription apparatus to the chemically unique environment of plastids.[7] A limitation to this interpretation is that considerable parts of some PAPs were not visible in our cryo-EM reconstruction, indicating conformational flexibility. It is thus possible that these regions become ordered and interact with nucleic acids during some stages of the transcription cycle. In addition, rearrangements of the core RNAP during transcription may lead to movement of PAPs, as described above.[45]

Finally, our evolutionary analysis suggests that the architecture of the PEP complex determined here is conserved across land plants. Our survey of PAP orthologs points to the terrestrialization as a decisive evolutionary event in the formation of the PEP complex. The conservation of genes for PEP complex subunits correlates well with a highly conserved plastome organization and PEP promoter specificity in plant chloroplast gene expression.[17,18] Interestingly, the PEP subunits PAP5/Hemera and PAP8/pTAC6 are dually localized proteins present in both chloroplasts and the nucleus. In the latter, both proteins have been shown to be essential components of phytochrome-mediated plant photomorphogenesis.[56,77] They apparently evolved in land plants, when light-controlled development of a multicellular body plan with multiple plastids per cell became important,[78] and may provide a retrograde signaling link connecting PEP and nuclear RNAPs to coordinate nuclear and plastid photosynthesis gene expression.[79] Early studies indicate that the PAPs are rapidly recruited to the core RNAP during the light-induced etioplast-chloroplast transition—the developmental phase with the highest demand for photosynthesis proteins.[7,25] This assembly of the PEP complex occurs without any detectable intermediate assembly states, suggesting that the complex in chloroplasts is functional and stable only in its entirety, which explains the strong albinoic phenotypes of *pap* inactivation mutants. In addition, it is noteworthy that in some plants within the embryophyta, PAPs may have further evolved to adopt new functions, since *pap* mutants of *Physcomitrium patens* do not develop the typical albinoic phenotypes as in vascular plants.[55,80] Elucidating the evolution and functional diversity of PAPs will be an exciting field for future research, which will be facilitated by the availability of structural information.

In summary, the structure of the PEP complex provides molecular insight into the architecture of the chloroplast transcription apparatus and will serve as a rational framework for future structure-guided studies aimed at obtaining a detailed molecular understanding of chloroplast gene expression and its regulation.

Limitations of the study

The cryo-EM reconstruction showed several residual unassigned densities, which were not of sufficient quality for us to confidently interpret and model them (see STAR Methods). It is possible that with a larger dataset, additional conformational and compositional states could be identified, which would allow for modeling further parts. Furthermore, the modeling of nucleic-acid-bound PEP complexes (Figure 5) is based on the assumption that the architecture and composition of the PEP complex observed in this study corresponds to that of the active enzyme and does not change during different stages of chloroplast transcription. However, it is possible that structural rearrangements occur during nucleic-acid binding and transcription, or that the composition of the PEP varies during different stages of the transcription cycle. Whether these complexes indeed adopt the proposed architecture will require structural data on elongation and initiation complexes of the PEP.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Isolation and lysis of *S. alba* chloroplasts
  - Affinity chromatography and glycerol gradient centrifugation
  - Electrophoresis and Immuno-detection
  - Mass spectrometry of PEP fractions
  - Cryo-EM sample preparation and data collection
  - Cryo-EM data processing and analysis
  - Model building and refinement
  - Homology searches

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2024.02.003.

ACKNOWLEDGMENTS

We thank Christian Dienemann and Ulrich Steuerwald (MPI-NAT cryo-EM facility) for technical assistance. This work was funded by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy EXC 2067/1-390729940 (H.S.H.), Hertha Sponer College (P.F.V.d.P.), FOR2848 (P10, H.S.H.), SFB1565 (project number 469281184, P13, H.S.H.), PF323-7 (T.P.), the framework of the Priority Programme "MAdLand - Molecular Adaption to Land: Plant Evolution to Change (SPP2237, T.P.), PF323-9 (T.P.), and by the European Union (ERC Starting Grant MitoRNA, grant agreement no. 101116869, to H.S.H.). Views and opinions expressed are those of the

author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

AUTHOR CONTRIBUTIONS

F.M.A. prepared *S. alba* chloroplasts and purified the PEP complex. P.F.V.d.P. prepared cryo-EM grids and collected cryo-EM data. P.F.V.d.P. and H.S.H. processed cryo-EM data and built the structural model. M.L. and F.M.A. performed homology searches for *pap* genes. N.D. performed mass-spectrometric measurements. H.-P.B. analyzed mass spectrometric data. H.S.H. and T.P. conceived the study, supervised research and wrote the manuscript, with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 25, 2023
Revised: January 26, 2024
Accepted: February 6, 2024
Published: February 29, 2024

REFERENCES

1. Archibald, J.M. (2015). Endosymbiosis and eukaryotic cell evolution. Curr. Biol. *25*, R911–R921.
2. Ort, D.R., Merchant, S.S., Alric, J., Barkan, A., Blankenship, R.E., Bock, R., Croce, R., Hanson, M.R., Hibberd, J.M., Long, S.P., et al. (2015). Redesigning photosynthesis to sustainably meet global food and bioenergy demand. Proc. Natl. Acad. Sci. USA *112*, 8529–8536.
3. Jarvis, P., and López-Juez, E. (2013). Biogenesis and homeostasis of chloroplasts and other plastids. Nat. Rev. Mol. Cell Biol. *14*, 787–802.
4. Hedtke, B., Börner, T., and Weihe, A. (1997). Mitochondrial and chloroplast phage-type RNA polymerases in Arabidopsis. Science *277*, 809–811.
5. Zhelyazkova, P., Sharma, C.M., Förstner, K.U., Liere, K., Vogel, J., and Börner, T. (2012). The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. Plant Cell *24*, 123–136.
6. Williams-Carrier, R., Zoschke, R., Belcher, S., Pfalz, J., and Barkan, A. (2014). A major role for the plastid-encoded RNA polymerase complex in the expression of plastid transfer RNAs. Plant Physiol. *164*, 239–248.
7. Pfannschmidt, T., Ogrzewalla, K., Baginsky, S., Sickmann, A., Meyer, H.E., and Link, G. (2000). The multisubunit chloroplast RNA polymerase A from mustard (Sinapis alba L.): integration of a prokaryotic core into a larger complex with organelle-specific functions. Eur. J. Biochem. *267*, 253–261.
8. Steiner, S., Schröter, Y., Pfalz, J., and Pfannschmidt, T. (2011). Identification of essential subunits in the plastid-encoded RNA polymerase complex reveals building blocks for proper plastid development. Plant Physiol. *157*, 1043–1055.
9. Suzuki, J.Y., Ytterberg, A.J., Beardslee, T.A., Allison, L.A., Wijk, K.J., and Maliga, P. (2004). Affinity purification of the tobacco plastid RNA polymerase and in vitro reconstitution of the holoenzyme. Plant J. *40*, 164–172.
10. Finster, S., Eggert, E., Zoschke, R., Weihe, A., and Schmitz-Linneweber, C. (2013). Light-dependent, plastome-wide association of the plastid-encoded RNA polymerase with chloroplast DNA. Plant J. *76*, 849–860.
11. Allison, L.A., Simon, L.D., and Maliga, P. (1996). Deletion of rpoB reveals a second distinct transcription system in plastids of higher plants. EMBO J. *15*, 2802–2809.
12. Legen, J., Kemp, S., Krause, K., Profanter, B., Herrmann, R.G., and Maier, R.M. (2002). 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.
13. Pfalz, J., and Pfannschmidt, T. (2013). Essential nucleoid proteins in early chloroplast development. Trends Plant Sci. *18*, 186–194.
14. Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K.J., and Oelmüller, R. (2006). pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. Plant Cell *18*, 176–197.
15. Lerbs-Mache, S. (2011). Function of plastid sigma factors in higher plants: regulation of gene expression or just preservation of constitutive transcription? Plant Mol. Biol. *76*, 235–249.
16. Lane, W.J., and Darst, S.A. (2010). Molecular evolution of multisubunit RNA polymerases: sequence analysis. J. Mol. Biol. *395*, 671–685.
17. Börner, T., Aleynikova, A.Yu., Zubo, Y.O., and Kusnetsov, V.V. (2015). Chloroplast RNA polymerases: role in chloroplast biogenesis. Biochim. Biophys. Acta *1847*, 761–769.
18. Pfannschmidt, T., Blanvillain, R., Merendino, L., Courtois, F., Chevalier, F., Liebers, M., Grübler, B., Hommel, E., and Lerbs-Mache, S. (2015). Plastid RNA polymerases: orchestration of enzymes with different evolutionary origins controls chloroplast biogenesis during the plant life cycle. J. Exp. Bot. *66*, 6957–6973.
19. Shen, L., Lai, G., You, L., Shi, J., Wu, X., Puiu, M., Gu, Z., Feng, Y., Yuzenkova, Y., and Zhang, Y. (2023). An SI3-σ arch stabilizes cyanobacteria transcription initiation complex. Proc. Natl. Acad. Sci. USA *120*, e2219290120.
20. Kindgren, P., and Strand, Å. (2015). Chloroplast transcription, untangling the Gordian knot. New Phytol. *206*, 889–891.
21. Favier, A., Gans, P., Boeri Erba, E.B., Signor, L., Muthukumar, S.S., Pfannschmidt, T., Blanvillain, R., and Cobessi, D. (2021). The plastid-encoded RNA polymerase-associated protein PAP9 is a superoxide dismutase with unusual structural features. Front. Plant Sci. *12*, 668897.
22. Arsova, B., Hoja, U., Wimmelbacher, M., Greiner, E., Üstün, S., Melzer, M., Petersen, K., Lein, W., and Börnke, F. (2010). 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.
23. Ruedas, R., Muthukumar, S.S., Kieffer-Jaquinod, S., Gillet, F.-X., Fenel, D., Effantin, G., Pfannschmidt, T., Couté, Y., Blanvillain, R., and Cobessi, D. (2022). Three-dimensional envelope and subunit interactions of the plastid-encoded RNA polymerase from Sinapis alba. Int. J. Mol. Sci. *23*, 9922.
24. Chieb, M., Liebers, M., Chevalier, F., Lerbs-Mache, S., Blanvillain, R., and Pfannschmidt, T. (2018). Determination of the DNA/RNA-associated subproteome from chloroplasts and other plastid types. Methods Mol. Biol. *1829*, 253–271.
25. Pfannschmidt, T., and Link, G. (1994). Separation of two classes of plastid DNA-dependent RNA polymerases that are differentially expressed in mustard (Sinapis alba L.) seedlings. Plant Mol. Biol. *25*, 69–81.
26. Pfannschmidt, T., and Link, G. (1997). The A and B forms of plastid DNA-dependent RNA polymerase from mustard (Sinapis alba L.) transcribe the same genes in a different developmental context. Mol. Gen. Genet. *257*, 35–44.
27. Werner, F., and Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. Nat. Rev. Microbiol. *9*, 85–98.
28. Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science *292*, 1863–1876.
29. Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999). Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. Cell *98*, 811–824.
30. Hirata, A., Klein, B.J., and Murakami, K.S. (2008). The X-ray crystal structure of RNA polymerase from Archaea. Nature *451*, 851–854.
31. Engel, C., Sainsbury, S., Cheung, A.C., Kostrewa, D., and Cramer, P. (2013). RNA polymerase I structure and transcription regulation. Nature *502*, 650–655.

32. Fernández-Tornero, C., Moreno-Morcillo, M., Rashid, U.J., Taylor, N.M.I., Ruiz, F.M., Gruene, T., Legrand, P., Steuerwald, U., and Müller, C.W. (2013). Crystal structure of the 14-subunit RNA polymerase I. Nature *502*, 644–649.

33. Hoffmann, N.A., Jakobi, A.J., Moreno-Morcillo, M., Glatt, S., Kosinski, J., Hagen, W.J.H., Sachse, C., and Müller, C.W. (2015). Molecular structures of unbound and transcribing RNA polymerase III. Nature *528*, 231–236.

34. Kang, J.Y., Olinares, P.D.B., Chen, J., Campbell, E.A., Mustaev, A., Chait, B.T., Gottesman, M.E., and Darst, S.A. (2017). Structural basis of transcription arrest by coliphage HK022 Nun in an Escherichia coli RNA polymerase elongation complex. eLife *6*, e25478.

35. Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H., and Severinov, K. (2001). Bacterial RNA polymerase subunit ω and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. Proc. Natl. Acad. Sci. USA *98*, 892–897.

36. Ghosh, P., Ishihama, A., and Chatterji, D. (2001). Escherichia coli RNA polymerase subunit omega and its N-terminal domain bind full-length beta' to facilitate incorporation into the alpha2beta subassembly. Eur. J. Biochem. *268*, 4621–4627.

37. Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., and Kornberg, R.D. (2006). Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. Cell *127*, 941–954.

38. Zhang, J., Palangat, M., and Landick, R. (2010). Role of the RNA polymerase trigger loop in catalysis and pausing. Nat. Struct. Mol. Biol. *17*, 99–104.

39. Vassylyev, D.G., Vassylyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007). Structural basis for substrate loading in bacterial RNA polymerase. Nature *448*, 163–168.

40. Toulokhonov, I., Zhang, J., Palangat, M., and Landick, R. (2007). A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. Mol. Cell *27*, 406–419.

41. Mishanina, T.V., Palo, M.Z., Nayak, D., Mooney, R.A., and Landick, R. (2017). 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.

42. Mazumder, A., Lin, M., Kapanidis, A.N., and Ebright, R.H. (2020). Closing and opening of the RNA polymerase trigger loop. Proc. Natl. Acad. Sci. USA *117*, 15642–15649.

43. Windgassen, T.A., Mooney, R.A., Nayak, D., Palangat, M., Zhang, J., and Landick, R. (2014). Trigger-helix folding pathway and SI3 mediate catalysis and hairpin-stabilized pausing by Escherichia coli RNA polymerase. Nucleic Acids Res. *42*, 12707–12721.

44. Kang, J.Y., Mishanina, T.V., Bellecourt, M.J., Mooney, R.A., Darst, S.A., and Landick, R. (2018). RNA polymerase accommodates a pause RNA hairpin by global conformational rearrangements that prolong pausing. Mol. Cell *69*, 802–815.e5.

45. Qayyum, M.Z., Imashimizu, M., Leanca, M., Vishwakarma, R.K., Riaz-Bradley, A., Yuzenkova, Y., and Murakami, K.S. (2024). Structure and function of the Si3 insertion integrated into the trigger loop/helix of cyanobacterial RNA polymerase. https://doi.org/10.1101/2024.01.11.575193.

46. Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science *288*, 640–649.

47. Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S.A. (2000). A structural model of transcription elongation. Science *289*, 619–625.

48. Yagi, Y., Ishizaki, Y., Nakahira, Y., Tozawa, Y., and Shiina, T. (2012). 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.

49. Schmitz-Linneweber, C., and Small, I. (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci. *13*, 663–670.

50. Barkan, A., and Small, I. (2014). Pentatricopeptide repeat proteins in plants. Annu. Rev. Plant Biol. *65*, 415–442.

51. Garg, G., Dienemann, C., Farnung, L., Schwarz, J., Linden, A., Urlaub, H., and Cramer, P. (2023). Structural insights into human co-transcriptional capping. Mol. Cell *83*, 2464–2477.e5.

52. Dillon, S.C., Zhang, X., Trievel, R.C., and Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biol. *6*, 227.

53. Guo, Q., Liao, S., Kwiatkowski, S., Tomaka, W., Yu, H., Wu, G., Tu, X., Min, J., Drozak, J., and Xu, C. (2019). Structural insights into SETD3-mediated histidine methylation on β-actin. eLife *8*, e43676.

54. Smith, C.A. (2006). Structure, Function and Dynamics in the mur Family of Bacterial Cell Wall Ligases. J. Mol. Biol. *362*, 640–655.

55. Garcia, M., Myouga, F., Takechi, K., Sato, H., Nabeshima, K., Nagata, N., Takio, S., Shinozaki, K., and Takano, H. (2008). An Arabidopsis homolog of the bacterial peptidoglycan synthesis enzyme MurE has an essential role in chloroplast development. Plant J. *53*, 924–934.

56. Liebers, M., Gillet, F.X., Israel, A., Pounot, K., Chambon, L., Chieb, M., Chevalier, F., Ruedas, R., Favier, A., Gans, P., et al. (2020). Nucleo-plastidic PAP8/pTAC6 couples chloroplast formation with photomorphogenesis. EMBO J. *39*, e104941.

57. Gilkerson, J., Perez-Ruiz, J.M., Chory, J., and Callis, J. (2012). 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.

58. Paul, R., Chatterjee, S., Nath, S., and Sen, U. (2018). Large-scale conformational changes and redistribution of surface negative charge upon sugar binding dictate the fidelity of phosphorylation in Vibrio cholerae fructokinase. Sci. Rep. *8*, 16925.

59. Myouga, F., Hosoda, C., Umezawa, T., Iizumi, H., Kuromori, T., Motohashi, R., Shono, Y., Nagata, N., Ikeuchi, M., and Shinozaki, K. (2008). 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.

60. Perry, J.J.P., Shin, D.S., Getzoff, E.D., and Tainer, J.A. (2010). The structural biochemistry of the superoxide dismutases. Biochim. Biophys. Acta *1804*, 245–262.

61. Dunwell, J.M., Purvis, A., and Khuri, S. (2004). Cupins: the most functionally diverse protein superfamily? Phytochemistry *65*, 7–17.

62. Chen, J., Chiu, C.E., Campbell, E.A., and Darst, S.A. (2020). *Escherichia coli* RNA polymerase closed complex (TRPc) with TraR and rpsT P2 promoter. https://pdbj.org/emnavi/quick.php?id=EMD-20460.

63. Feklístov, A., Sharon, B.D., Darst, S.A., and Gross, C.A. (2014). Bacterial sigma factors: A historical, structural, and genomic perspective. Annu. Rev. Microbiol. *68*, 357–376.

64. Maul, J.E., Lilly, J.W., Cui, L., dePamphilis, C.W., Miller, W., Harris, E.H., and Stern, D.B. (2002). The Chlamydomonas reinhardtii Plastid Chromosomelslands of genes in a sea of repeats. Plant Cell *14*, 2659–2679.

65. de Vries, J., Stanton, A., Archibald, J.M., and Gould, S.B. (2016). Streptophyte terrestrialization in light of plastid evolution. Trends Plant Sci. *21*, 467–476.

66. Liu, G.-Q., Lian, L., and Wang, W. (2022). The molecular phylogeny of land plants: progress and future prospects. Diversity *14*, 782.

67. Martinez-Rucobo, F.W., and Cramer, P. (2013). Structural basis of transcription elongation. Biochim. Biophys. Acta *1829*, 9–19.

68. Yagi, Y., and Shiina, T. (2014). Recent advances in the study of chloroplast gene expression and its evolution. Front. Plant Sci. *5*, 61.

69. Gao, Z.-P., Yu, Q.-B., Zhao, T.-T., Ma, Q., Chen, G.-X., and Yang, Z.-N. (2011). 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.

70. Pfalz, J., Holtzegel, U., Barkan, A., Weisheit, W., Mittag, M., and Pfannschmidt, T. (2015). ZmpTAC12 binds single-stranded nucleic acids and is essential for accumulation of the plastid-encoded polymerase complex in maize. New Phytol. *206*, 1024–1037.

71. Wimmelbacher, M., and Börnke, F. (2014). Redox activity of thioredoxin z and fructokinase-like protein 1 is dispensable for autotrophic growth of Arabidopsis thaliana. J. Exp. Bot. *65*, 2405–2413.

72. Chambon, L., Gillet, F.-X., Chieb, M., Cobessi, D., Pfannschmidt, T., and Blanvillain, R. (2022). PAP8/pTAC6 Is Part of a Nuclear Protein Complex and Displays RNA Recognition Motifs of Viral Origin. Int. J. Mol. Sci. *23*, 3059. https://doi.org/10.3390/ijms23063059.

73. Le Moigne, T.L., Gurrieri, L., Crozet, P., Marchand, C.H., Zaffagnini, M., Sparla, F., Lemaire, S.D., and Henri, J. (2021). Crystal structure of chloroplastic thioredoxin z defines a type-specific target recognition. Plant J. *107*, 434–447.

74. Díaz, M.G., Hernández-Verdeja, T., Kremnev, D., Crawford, T., Dubreuil, C., and Strand, Å. (2018). Redox regulation of PEP activity during seedling establishment in Arabidopsis thaliana. Nat. Commun. *9*, 50.

75. Hillen, H.S., Parshin, A.V., Agaronyan, K., Morozov, Y.I., Graber, J.J., Chernev, A., Schwinghammer, K., Urlaub, H., Anikin, M., Cramer, P., et al. (2017). Mechanism of transcription anti-termination in human mitochondria. Cell *171*, 1082–1093.e13.

76. Hillen, H.S., Morozov, Y.I., Sarfallah, A., Temiakov, D., and Cramer, P. (2017). Structural basis of mitochondrial transcription initiation. Cell *171*, 1072–1081.e10.

77. Chen, M., Galvão, R.M., Li, M., Burger, B., Bugea, J., Bolado, J., and Chory, J. (2010). Arabidopsis HEMERA/pTAC12 initiates photomorphogenesis by phytochromes. Cell *141*, 1230–1240.

78. Liebers, M., Grübler, B., Chevalier, F., Lerbs-Mache, S., Merendino, L., Blanvillain, R., and Pfannschmidt, T. (2017). Regulatory shifts in plastid transcription play a key role in morphological conversions of plastids during plant development. Front. Plant Sci. *8*, 23.

79. Liebers, M., Cozzi, C., Uecker, F., Chambon, L., Blanvillain, R., and Pfannschmidt, T. (2022). Biogenic signals from plastids and their role in chloroplast development. J. Exp. Bot. *73*, 7105–7125.

80. Takahashi, A., Sugita, C., Ichinose, M., and Sugita, M. (2021). Moss PPR-SMR protein PpPPR_64 influences the expression of a psaA-psaB-rps14 gene cluster and processing of the 23S–4.5S rRNA precursor in chloroplasts. Plant Mol. Biol. *107*, 417–429.

81. Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc. *11*, 2301–2319.

82. Mastronarde, D.N. (2018). Advanced data acquisition from electron microscopes with SerialEM. Microsc. Microanal. *24*, 864–865.

83. Tegunov, D., and Cramer, P. (2019). Real-time cryo-electron microscopy data preprocessing with Warp. Nat. Methods *16*, 1146–1152.

84. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296.

85. Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres, S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. eLife *7*, 163.

86. Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., and Adams, P.D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. Acta Crystallogr. D Struct. Biol. *74*, 531–544.

87. Casañal, A., Lohkamp, B., and Emsley, P. (2020). Current developments in coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. Protein Sci. *29*, 1069–1078.

88. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82.

89. Croll, T.I. (2018). ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. Acta Crystallogr. D Struct. Biol. *74*, 519–530.

90. Wittig, I., Braun, H.P., and Schägger, H. (2006). Blue native PAGE. Nat. Protoc. *1*, 418–428.

91. Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680–685.

92. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

93. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589.

94. Sreedasyam, A., Plott, C., Hossain, M.S., Lovell, J.T., Grimwood, J., Jenkins, J.W., Daum, C., Barry, K., Carlson, J., Shu, S., et al. (2023). JGI Plant Gene Atlas: an updateable transcriptome resource to improve functional gene descriptions across the plant kingdom. Nucleic Acids Res. *51*, 8383–8401.

95. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682.

96. Holm, L. (2022). Dali server: structural unification of protein families. Nucleic Acids Res. *50*, W210–W215.

97. vanKempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M., Söding, J., and Steinegger, M. (2023). Fast and accurate protein structure search with Foldseek. Nat. Biotechnol. https://doi.org/10.1038/s41587-023-01773-0.

98. Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K.K., Haas, F.B., Vanderstraeten, L., Becker, D., Lang, D., et al. (2018). The chara genome: secondary complexity and implications for plant terrestrialization. Cell *174*, 448–464.e24.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-RpoB	Unpublished, produced against full-length recombinantly expressed *A. thaliana* RpoB	N/A
Rabbit polyclonal anti-PAP8	Liebers et al.[56]	N/A
Goat polyclonal anti-Rabbit IgG (H&L), HRP conjugated	Agrisera	Cat #AS09 602; Lot #2007; RRID:AB_1966902
Chemicals, peptides, and recombinant proteins		
Heparin Sepharose 6 Fast Flow affinity resin	Cytiva	Cat #17099801
Sequencing Grade Modified Trypsin	Promega	Cat #V5111
Deposited data		
Label free proteomic mass spectrometry data	This study; PRIDE	Tables S1 and S2; PRIDE (https://www.ebi.ac.uk/pride/archive/) Project accession: PXD048976
S.alba PEP cryo-EM map A (consensus refinement)	This study	EMD-18499
S.alba PEP cryo-EM map B (RPOB lobe focus refinement)	This study	EMD-18500
S.alba PEP cryo-EM map C (RPOC lobe focus refinement)	This study	EMD-18501
S.alba PEP cryo-EM map D (foot focus refinement)	This study	EMD-18502
S.alba PEP cryo-EM map E (central RPOB lobe focus refinement)	This study	EMD-18503
S.alba PEP cryo-EM map F (PAP3 cluster focus refinement)	This study	EMD-18504
S.alba PEP cryo-EM map G (composite map)	This study	EMD-18496
S.alba PEP model	This study	PDB: 8QMA
Experimental models: Organisms/strains		
Sinapis alba Var. Litember	Raiffeisen Handelsgesellschaft mbH	Formerly: Cat #43193; Currently: Cat #10407348-1; RRID: NCBITaxon_3728
Software and algorithms		
MaxQuant version 2.4.2.0	Tyanova et al.[81]	https://www.maxquant.org/
SerialEM 4.0	Mastronarde[82]	https://bio3d.colorado.edu/SerialEM/#Source
Warp 1.0.9	Tegunov and Cramer[83]	http://www.warpem.com
cryoSPARC 4.2.1	Punjabi et al.[84]	https://cryosparc.com/
Relion 3.1.0	Zivanov et al.[85]	https://github.com/3dem/relion
PHENIX 1.21-5207	Afonine et al.[86]	http://www.phenix-online.org
Coot 0.9.8.92	Casañal et al.[87]	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
ChimeraX 1.61 & 1.7	Pettersen et al.[88]	https://www.cgl.ucsf.edu/chimerax/
ISOLDE 1.6	Croll[89]	https://isolde.cimr.cam.ac.uk/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
Amicon Ultra Centrifugal Filter, 10 kDa MWCO	Merck	Cat #UFC5010
Vivaspin 20 Centrifugal Concentrator Polyethersulfone, 100 kDa MWCO	Sartorius	Cat #VS2002
Transfermembrane ROTI PVDF 0.45 μm	Carl Roth	Cat #T830.1
Titan Krios G2	FEI / Thermo Fisher Scientific	N/A
Quantum LS Energy Filter	Gatan	N/A
K3 Summit Direct Electron Detector	Gatan	N/A
Quantifoil R2/1 holey carbon grids (copper)	Quantifoil	N/A

RESOURCE AVAILABILITY

Lead contact

Correspondence and requests for materials should be addressed to H.S.H. (hauke.hillen@med.uni-goettingen.de).

Materials availability

Materials are available from Hauke S. Hillen or Thomas Pfannschmidt upon request under a material transfer agreement.

Data and code availability

- The cryo-EM density reconstructions were deposited with the Electron Microscopy Database (EMDB) under accession codes EMD-18499 (Map A), EMD-18500 (Map B), EMD-18501 (Map C), EMD-18502 (Map D), EMD-18503 (Map E), EMD-18504 (Map F) and EMD-18496 (Map G). The structure coordinates were deposited with the Protein Data Bank (PDB) under accession code 8QMA. Mass-spectrometry data were deposited with the PRIDE database under accession number PXD048976.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

White mustard (*Sinapis alba* var. Litember) was grown in 40x60 cm trays with draining holes that were filled with 3-4 cm of soil (SP Vermehrung, Cat #11-01500 Einheitserde Werkverband e.V., Sinntal-Altengronau, Germany). Seeds were densely sown to an extent that the soil was uniformly covered with a single layer of seeds. The seeds were then kept moist for the first 2-3 days. After cotyledons emerged, watering was reduced to prevent growth of mold around the seedlings and roots. Growth conditions were approximately 21 °C and 30 % relative humidity with a 16 h day and 8 h night cycle inside a climate chamber (GroBanks BB-XL3 Model, CLF Plant Climatics GmbH, Wertingen, Germany). The light intensity was ~120 μE of white light. The total growth time was 6 days from sowing to harvesting. Cotyledons were then harvested using a pair of scissors, while avoiding as much of the hypocotyls as possible. Harvesting was performed shortly after onset of light to prevent starch accumulation from photosynthesis.

METHOD DETAILS

Isolation and lysis of *S. alba* chloroplasts

PEP was purified from 6-day old white mustard (*Sinapis alba* var. Litember) chloroplasts as described earlier[24] with minor modifications. Therefore, cotyledons were harvested shortly after onset of light and homogenized in ice-cold homogenization buffer (50 mM Tris pH 8.0, 10 mM $MgCl_2$, 330 mM Sorbitol, 1 mM β-mercaptoethanol) using 3 short pulses in a blender. The homogenate was filtered through 4 layers of nylon mesh and pelleted at 4000 xg, 4 °C for 10 minutes. Following resuspension of the pellet in a small amount of homogenization buffer, chloroplasts were isolated from this crude lysate by sucrose gradient centrifugation. To this end, 10 ml of lysate were layered on top of a 30 ml 20-55% (w/w) linear sucrose gradient with a 10 ml 55% sucrose cushion in 50 ml falcon tubes and centrifuged at 3220 xg, 4 °C for 90 minutes. Sucrose solutions were prepared in dilution buffer (50 mM Tris pH 8.0, 10 mM $MgCl_2$, 1 mM β-mercaptoethanol). After centrifugation, the lower band containing intact chloroplasts was collected, 2-fold diluted in dilution buffer and pelleted at 4500 xg, 4 °C for 20 minutes. Lysis of chloroplasts was performed in ice-cold lysis buffer (50 mM Tris pH 7.6, 25% (v/v) glycerol, 1 % (v/v) Triton X-100, 4 mM EDTA, 40 mM β-mercaptoethanol, 0.1 mM PMSF, 10 mM NaF) by thorough resuspension using a Potter-Elvehjem-homogenizer and subsequent stirring for 1 h at 350 rpm on ice. Insoluble cell debris was removed by centrifugation at 3000 xg, 4 °C for 10 minutes. If the resulting pellet remained green, lysis was repeated until the pellet became brownish-white. The supernatant was flash frozen in liquid nitrogen and stored at -80 °C until affinity chromatography.

Affinity chromatography and glycerol gradient centrifugation

Affinity chromatography was performed using heparin-sepharose (HS) FastFlow6 (Cytiva, Marlborough, Massachusetts, USA) with a bed volume of 20 ml that was equilibrated in column buffer (50 mM Tris pH 7.6, 10% (v/v) glycerol, 1 mM β-mercaptoethanol, 0.1 mM EDTA, 10 mM $MgCl_2$, 0.1% (v/v) Triton X-100, 10 mM NaF, 80 mM $(NH_4)_2SO_4$). The solubilized chloroplasts suspension was thawed on ice, adjusted to 10 mM $MgCl_2$, 80 mM $(NH_4)_2SO_4$ and applied to the HS by gravity flow. The HS resin was washed extensively in column buffer and the column was eluted in column buffer containing 1.2 M $(NH_4)_2SO_4$. After dialysis in 50 mM Tris pH 7.6, 10% (v/v) glycerol, 0.1 mM EDTA, 10 mM $MgCl_2$, 10 mM NaF, 0.1 mM PMSF, 1 mM β-mercaptoethanol for 12 h (Type 27 dialysis tube, 12-16 kDa MWCO, pore size 25 A, Biomol GmbH, Hamburg, Germany), the dialysate was again subjected to 5 ml of HS resin, washed and eluted as before. The resulting elution fraction was further concentrated using an ultrafiltration spin column (Vivaspin 20, 100 kDa MWCO, Sartorius, Göttingen, Germany). 1200 µl of this concentrated HS fraction were loaded on a 17 ml 15-35 % (w/v) glycerol gradient (prepared in 50 mM Tris pH 7.6, 10 mM $MgCl_2$, 0.01% (v/v) TX-100, 1 mM β-mercaptoethanol) and centrifuged at 100,000 xg at 4 °C for 16.5 h. The gradient was fractionated into 660 µl fractions and analyzed by SDS-PAGE following silver staining or immunodetection to identify fractions containing the PEP. Fractions 8-10, 11-13 and 14-16 were pooled individually and concentrated in ultrafiltration units (Amicon Ultra 0.5 ml, 10 kDa MWCO, Merck KGaA, Darmstadt, Germany) while slowly exchanging the buffer to 50 mM Tris pH 7.6, 0.1 mM EDTA, 10 mM $MgCl_2$, 10 mM NaF, 0.1 mM PMSF, 1 mM β-mercaptoethanol.

Electrophoresis and Immuno-detection

The resulting PEP fractions were analyzed on a 4-12% BN-PAGE for integrity of the 1 MDa complex. BN-PAGE was performed according to Wittig et al.,[90] stained in 50% (v/v) ethanol, 10% (v/v) acetic acid, 0.1% (w/v) Coomassie R-250 for 1h, destained in 10% (v/v) ethanol and 7% (v/v) acetic acid and the 1 MDa band of the PEP excised for a mass spectrometry analysis.

SDS-PAGE was performed as described by Laemmli.[91] After electrophoresis, proteins were transferred in towbin buffer (25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS, 15% (v/v) ethanol, pH 8.3) to a PVDF membrane (Roti-PVDF, 0.45 µm pore size, Carl Roth GmbH & Co. KG, Karlsruhe, Germany) using a TE77X semi-dry transfer unit (SERVA Electrophoresis GmbH, Heidelberg, Germany) at 0.8 mA/cm^2 for 75 min. The membrane was blocked in TTBS (50 mM Tris pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween 20) containing 5% (w/v) skimmed milk powder for 1 h and probed with PAP8[56] and RpoB (polyclonal, produced in rabbit against full-length recombinant RpoB) antibodies (both diluted 1:10,000 in 10 ml TTBS) over night. After washing 4 times in TTBS, membranes were incubated with secondary goat anti-rabbit HRP conjugated antibody (AS09 602, Agrisera AB, Vännäs, Sweden, diluted 1:10,000 in 10 ml TTBS) for 2 hours before washing again 4 times in TTBS. Chemiluminescence detection was performed using Lumi-LightPlus western blot substrate (Roche, Basel, Switzerland) in a Lumi-Imager F1 (Roche, Basel, Switzerland).

Mass spectrometry of PEP fractions

Proteins within the PEP complex were identified by tryptic in-gel digestion and liquid chromatography coupled tandem mass spectrometry (LC-MS/MS). The excised protein band was dehydrated using acetonitrile (ACN) and a vacuum centrifuge. Disulfide bridges were reduced in 20 mM dithiothreitol and 100 mM ammonium bicarbonate (ABC) at 56 °C for 30 min before alkylation in 55 mM IAA and 100 mM ABC for 30 min in the dark. Alkylated proteins were washed in 100 mM ABC before tryptic (sequence grade modified trypsin, Promega, Mannheim, Germany) in-gel digestion at 37 °C overnight. Peptides extraction was performed in 5% [v/v] formic acid (FA) and 50% [v/v] ACN for 20 min at 37 °C and 800 rpm. Peptide extraction was repeated with 1% [v/v] FA and 50% [v/v] ACN, followed by a final step with 100% [v/v] ACN. Supernatants of each extraction were pooled and dried in a vacuum centrifuge before storage at -20 °C. For LC-MS analysis, dried peptides were resuspended in 20 µL 0.1% [v/v] FA and 5% [v/v] ACN by ultrasonication for 15 min. The suspension was centrifuged at 13,000 xg for 10 min and transferred into glass vials that were stored in the auto-sampler of an Ultimate 3000 UHPLC (Thermo Fischer Scientific, Dreieich, Germany) at 8 °C. Five microliter peptide suspension were injected into a 20 µL sample loop and loaded onto a 2 cm C18 reverse phase pre-column (Acclaim PepMap 100, 75 µm diameter, 3 µm particle size, 100 Å pore size, Thermo Fischer Scientific, Dreieich, Germany) at a flow rate of 3 µL min^{-1}. Peptides were separated on a 50 cm C18 reversed phase analytical column (Acclaim PepMap 100, 75 µm diameter, 3 µm particle size, 100 Å pore size, Thermo Fischer Scientific, Dreieich, Germany) by applying a non-linear 5% [v/v] ACN to 30% [v/v] ACN gradient in 0.1% [v/v] FA over a time of 63 min at a flow rate of 250 nL min^{-1} and a column temperature of 45° C. The ACN concentration was subsequently increased to 76% [v/v] over a period of 10 min and kept at this value for another 10 min. Equilibration of the column was achieved by decreasing the ACN concentration to 5% [v/v] over a time period of 5 min, followed by additional 10 min. Eluting peptides were transferred into a Q-Exactive MS using a NSI source (both Thermo Fischer Scientific, Dreieich, Germany) equipped with a stainless-steel nano-bore emitter. Ionization of peptides was done using a spray voltage of 2.2 kV and a capillary temperature of 275 °C. The radio-frequency (RF)-level of the S-lens was set to 50% and the MS was run in positive ion mode. A data dependent acquisition (DDA) strategy was used with the top 10 MS/MS scans recorded from 15 to 95 min. Full MS scans were performed at a resolution of 70,000 and a scan range from 400 to 1600 m/z. The automated gain control (AGC) target was set to 1e6 at a maximum injection time of 400 ms. For MS/MS scans, top 10 peptides were fragmented at a normalized collision energy (NCE) of 27% and recorded at a resolution of 17500. The AGC target was set to 1e5 at a maximum injection time of 120 ms. Only ions with a charge of 2-4 above an intensity threshold of 4.2e3 were considered for MS/MS scans at an isolation window of 3 m/z and a dynamic exclusion of 45 s. MS raw files obtained by the Xcalibur software were loaded into the Maxquant version 2.4.2.0[81] and queried against a *Sinapis alba* protein sequence database downloaded from uniprot (24.11.2022). Maxquant default settings were applied with the

following exceptions: the 'refine peaks' function and the 'match between runs' function were enabled. Additionally, iBAQ values[92] were computed to calculate abundance of PEP subunits and contaminants.

Cryo-EM sample preparation and data collection

4μl of isolated PEP complex (pooled fractions 8-10 of the glycerol gradient) were applied to freshly glow-discharged R 2/1 holey carbon grids (Quantifoil), at 4 °C and 95% humidity in a Vitrobot (FEI). Grids were blotted for 7 s with a blot force of 5 immediately before plunge-freezing in liquid ethane.

Cryo-EM data collection was performed with SerialEM[82] using a Titan Krios transmission electron microscope (Thermo Fisher Scientific) operated at 300 keV. Images were acquired in EFTEM mode using a GIF quantum energy filter set to a slit width of 20 eV and a K3 direct electron detector (Gatan) at a nominal magnification of 81,000x corresponding to a calibrated pixel size of 1.05 Å/pixel. Exposures were recorded in counting mode for 2.7 seconds with a dose rate of 16.3 e^-/px/s resulting in a total dose of 40 e^-/Å^2 that was fractionated into 40 movie frames. Images were acquired in 3 by 3 holes per stage movement with active beamtilt compensation as implemented in SerialEM.

Cryo-EM data processing and analysis

Motion correction, CTF-estimation, particle picking and extraction were performed using Warp.[83] Further processing was carried out using cryoSPARC 4.2.1[84] and Relion 3.1.0.[85] A representative micrograph and the cryo-EM processing workflow are depicted in Figures S2A and S2B. For initial processing steps, the dataset was split into three batches containing 1,177,099, 1,476,936 and 771,313 particles, respectively. Initial 2D classification in Cryosparc revealed that only a relatively small subset (~ 4.5%) of particles was sorted into well-defined 2D classes due to the heterogeneous nature of the sample. These particles were selected and used for ab initio model generation followed by non-uniform refinement in Cryosparc, which led to an initial reconstruction of the PEP complex at a resolution of ~ 3.5 Å. However, this reconstruction showed anisotropy due to preferred orientation bias. In order to recover more particles with rare views, the initial PEP structure was used as input model in a supervised classification run (heterogeneous refinement in Cryosparc) together with five "junk" models generated ab initio from the previously excluded particles. Classification of the entire dataset with this strategy led to a subset of 592,666 particles, which was then subjected to one round of unsupervised classification (heterogeneous refinement in Cryosparc). Particles belonging to the best class (210,309 particles) were exported to Relion and subjected to 3D refinement followed by 3D classification with local alignments. Particles belonging to the class with clear density for both lobes of the PEP complex (46,160 particles) were selected and subjected to CTF refinement in Relion followed by non-uniform refinement in Cryosparc, which led to an isotropic reconstruction of the complete PEP complex at 3.46 Å resolution (Map A; Figures S2B–S2D). In this reconstruction the RPOC-lobe of the PEP complex remained less-well resolved than the opposing RPOB-lobe, suggesting conformational flexibility between the two halves of the PEP complex. Local refinement in Cryosparc using a mask encompassing the RPOC-lobe led to an improved map (Map C). To obtain the highest resolution possible for the RPOB-lobe, particles were combined with those from another class in the final 3D classification job which showed clear density for the RPOB-lobe only, resulting in a subset of 123,874 particles. Non-uniform refinement of these particles with a mask encompassing the entire RPOB-lobe led to a reconstruction at 2.95 Å resolution (Map B). To improve local resolution further, local refinements using masks encompassing the foot of the polymerase (Map D), the center of the RPOB-lobe (Map E) and the PAP3-cluster (Map F) were carried out (Figure S2). For model building and visualization purposes, a composite map of all focused maps (Map G) was generated using the vop max command in Chimera.

Model building and refinement

In order to generate an initial model of the PEP complex, we used available AlphaFold[93] models and sequences of the core polymerase subunits from Uniprot (Uniprot IDs: RPOA: A0A6C0M610; RPOB: A0A6C0M5W1, RPOC1: A0A6C0M5W0; RPOC2: A0A6C0M829). For the PAPs, we used sequence data from the Phytozome project[94] to predict AlphaFold models using the ColabFold notebook.[95] The AlphaFold models of the PEP subunits were fit into the cryo-EM density manually using ChimeraX,[88] guided by previously published cross-linking coupled to mass-spectrometry data.[23]After initial fitting, we observed several regions of well-resolved density around the PAP3-PAP4-PAP9-PTAC18 cluster with poor connectivity to the rest of the complex, which we hypothesized to represent parts of the RPOC2 SI3. To model these, we predicted PAP3:SI3 complex models using AlphaFold multimer. Placement of these models confirmed our initial modeling of SI3 and enabled us to confidently interpret the densities with poor connectivity as the PAP-binding domain 1 of SI3 and previously unmodeled parts of PAP3. The initial model was then manually re-built in Coot[87] using maps B,C,D,E,F and G. Some regions of the reconstruction showed less-well resolved density than the remainder of the complex, in particular the cleft with the active site and peripheral region. These regions were modeled by rigid-body fitting AlphaFold models of individual domains (e.g. the catalytic domain in RPOC1 (res. 378 – 519)). In order to obtain a stereochemically sound model, the model was first real-space refined against the composite Map G using phenix.real_space_refine (rigid-body and XYZ) with hydrogens present[86] followed by interactive re-building and refinement using molecular dynamics force fields in ISOLDE[89] within ChimeraX and further real-space refinement in Phenix. The final model was then real-space refined against the global map (Map A, ADP refinement only) locally filtered to local resolution in Cryosparc (Figures S2B and S2C). This resulted in a model with excellent stereochemistry and B-factors that adequately represent the conformational flexibility of the complex (Table 1). A list of all subunits and modeled regions is given in Table S3. We note that the reconstruction contained residual unassigned densities which were not of sufficient quality to confidently interpret and model

them. This includes discontinuous densities around the foot of the PEP complex as well as densities consistent with coordinated ions in FLN2 and PAP6, for example. In addition, we observe several additional densities in lowpass-filtered maps which may represent missing parts of the structure or additional subunits. Figures were prepared with ChimeraX 1.6.1 and structural similarity searches were carried out using the DALI server[96] and FoldSeek.[97]

Homology searches

NCBI blastp searches were performed, using the according *Arabidopsis* PAP protein sequence from TAIR as query. The standard database with non-redundant protein sequences was used for the protein-protein BLAST with standard settings (expected threshold 0.05, word size 5, max matches in a query 0, matrix BLOSUM62, gap cost existence 11, gap cost extension 1 and conditional compositional score matrix adjustment) and a maximum of 5000 target sequences. All blast results are given in the tabs "general blastp results" in the Table S3. The blast results were sorted by organism, e-value and percent identity. To roughly cover all Chloroplastida clades, at least one species from Chlorophyta, Klebsormidiophyceae, Charophyceae, Bryophyta, Lycopodiopsida, Polypodiophyta, Gymnosperms, Amborellales, Monocots and Eudicots were chosen (see "1st hits & rev. blast" in Table S3). "Best hits" were selected for each species, by searching for the hit with the lowest e-value and the highest percent identity. In the case that two or more hits showed identical e-value and percent identity, they were all listed. For some PAPs, blasting with 5000 target sequences was not sufficient to identify species from Lycopodiopsida, Marchantiophyta, Klebsormidiophyceae and/or Chlorophyta. In those cases, an additional blastp search was applied, searching only for hits in the missing taxons ("general blastp results" Table S4). In addition, blastp searches with the *Arabidopsis* PAP protein sequence as query was conducted with the database from Ginkgo DB genome (https://ginkgo.zju.edu.cn/genome/blast/), using standard settings (drop off 0, Match/Mismatch scores 1,-3, Penalty for opening a gap 5 and Penalty for extending a gap 2) and Hornworts from University of Zurich (https://www.hornworts.uzh.ch/en/Blast.html), using standard settings (Max e value 1e-3, Matrix BLOSUM62). For PAP12, *Arabidopsis*, *Amborella* and *Adiantum* proteins were used as queries in NCBI. Also e!EnsemblPlants (https://plants.ensembl.org/Multi/Tools/Blast) was used to search for hits of *Physcomitrium*, *Marchantia* and *Chara*, using *Arabidopsis* as query in the standard settings (E-value threshold 1e-1, dropoff 0, Scoring matrix BLOSUM62, Penalty for opening a gap 11, Penalty for extending a gap 1). In case of gaps in the blast hits for the selected organisms (mainly Lycopodiopsida, Marchantiophyta, Klebsormidiophyceae and/or Chlorophyta), the *Chara* protein sequence, identified during the search with the *Arabidopsis* protein as query, was used as query for the blastp in NCBI. To confirm that the identified hits are likely PAP orthologues, a reciprocal blast search was conducted[98] ("1st hist & rev. blast" in Table S4). Best matches from the initial blast search (see above) were used as query for the reverse blast, looking for *Arabidopsis* or *Chara* proteins as targets. Only if the highest scoring protein of either *Arabidopsis* or *Chara* yielded the PAP that was used for the initial search, the proteins were viewed as potential orthologues. If the reverse blast yielded other proteins as best matches, then the PAP protein and the species was marked in red in the Excel-file and was not included in Figure 6A. If another sequence of the same protein was found as hit, the reciprocal search was considered a best match and only the *Arabidopsis* hit from the initial search was given as a representative.