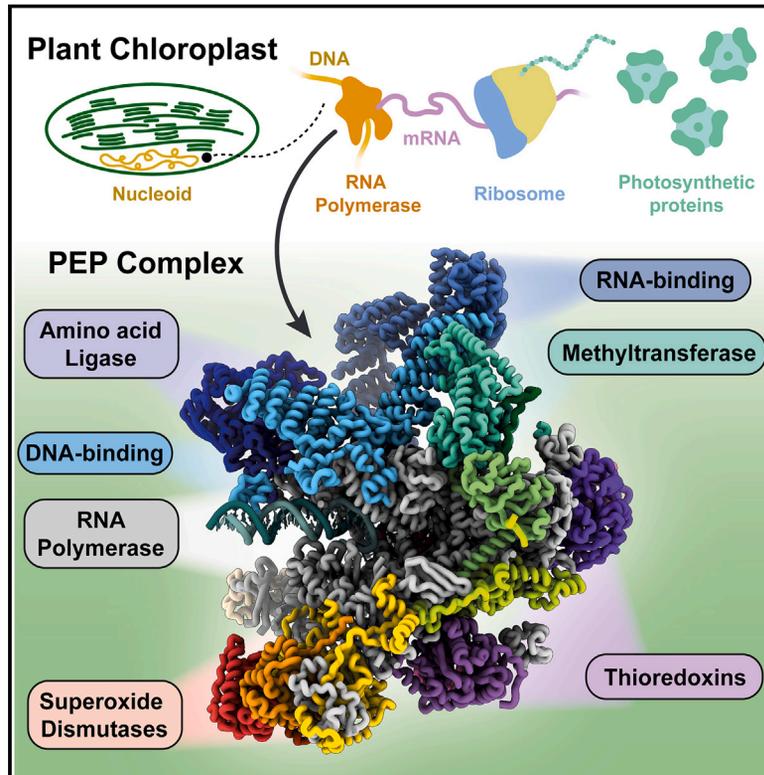


# Structure of the plant plastid-encoded RNA polymerase

## Graphical abstract



## Authors

Ángel Vergara-Cruces, Ishika Pramanick, David Pearce, Vinod K. Vogirala, Matthew J. Byrne, Jason K.K. Low, Michael W. Webster

## Correspondence

michael.webster@jic.ac.uk

## In brief

Structural characterization of the chloroplast RNA polymerase that transcribes photosynthetic genes provides insight into its composition, assembly, and evolution.

## Highlights

- Structure of the chloroplast transcription complex
- Fifteen nuclear-encoded subunits encase the plastid-encoded polymerase
- Subunits PAP1 and PAP2 interact with the DNA and the mRNA, respectively
- Structure-guided insights into enzymatic activities of subunits



## Article

# Structure of the plant plastid-encoded RNA polymerase

Ángel Vergara-Cruces,<sup>1,5</sup> Ishika Pramanick,<sup>1,5</sup> David Pearce,<sup>1,2</sup> Vinod K. Vogirala,<sup>3</sup> Matthew J. Byrne,<sup>3,6</sup> Jason K.K. Low,<sup>4</sup> and Michael W. Webster<sup>1,7,\*</sup>

<sup>1</sup>Department of Biochemistry and Metabolism, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK

<sup>2</sup>School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK

<sup>3</sup>Electron Bio-Imaging Centre (eBIC), Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, UK

<sup>4</sup>School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2050, Australia

<sup>5</sup>These authors contributed equally

<sup>6</sup>Present address: Exscientia, Heatley Road, Oxford Science Park, Oxford OX4 4GE, UK

<sup>7</sup>Lead contact

\*Correspondence: [michael.webster@jic.ac.uk](mailto:michael.webster@jic.ac.uk)

<https://doi.org/10.1016/j.cell.2024.01.036>

## SUMMARY

Chloroplast genes encoding photosynthesis-associated proteins are predominantly transcribed by the plastid-encoded RNA polymerase (PEP). PEP is a multi-subunit complex composed of plastid-encoded subunits similar to bacterial RNA polymerases (RNAPs) stably bound to a set of nuclear-encoded PEP-associated proteins (PAPs). PAPs are essential to PEP activity and chloroplast biogenesis, but their roles are poorly defined. Here, we present cryoelectron microscopy (cryo-EM) structures of native 21-subunit PEP and a PEP transcription elongation complex from white mustard (*Sinapis alba*). We identify that PAPs encase the core polymerase, forming extensive interactions that likely promote complex assembly and stability. During elongation, PAPs interact with DNA downstream of the transcription bubble and with the nascent mRNA. The models reveal details of the superoxide dismutase, lysine methyltransferase, thioredoxin, and amino acid ligase enzymes that are subunits of PEP. Collectively, these data provide a foundation for the mechanistic understanding of chloroplast transcription and its role in plant growth and adaptation.

## INTRODUCTION

Chloroplasts are photosynthetic organelles derived from a cyanobacterial ancestor. Consequently, both the chloroplast genome and the molecular machinery that expresses chloroplast genes have prokaryotic features.<sup>1</sup> Although the chloroplast genome encodes fewer genes than its bacterial ancestor, the molecular machinery that transcribes them has increased in complexity. Numerous chloroplast-specific transcription proteins have evolved to integrate a prokaryotic gene expression system into a eukaryotic host cell and to support the regulation of photosynthesis by developmental and environmental cues.<sup>2–4</sup> Chloroplast transcription is activated by light during plant development.<sup>5</sup> This change underpins production of the photosynthetic proteins that defines the morphology and bioenergetic capacity of the organelle.

The presence of a multi-subunit transcription enzyme in chloroplasts was identified 50 years ago.<sup>6</sup> Subsequent studies revealed that this plastid-encoded RNA polymerase (PEP) contains subunits homologous to bacterial RNA polymerase (RNAP) and relies on some prokaryotic regulatory mechanisms, such as sigma-dependent initiation.<sup>7</sup> Yet PEP is significantly larger than bacterial RNAP due to the stable association of

numerous PEP-associated proteins (PAPs) that are not homologous to bacterial transcription proteins. At least twelve PAPs are required for the activity of PEP to be sufficient for chloroplast biogenesis in *Arabidopsis*.<sup>2,3</sup> Thus, despite the similarity between the polymerase subunits of PEP and bacterial RNAP, PAPs impart essential but largely unknown functions.

PEP isolated from white mustard (*Sinapis alba*) comprises four polymerase subunits that are similar to cyanobacterial RNAP ( $\alpha_2$ ,  $\beta\beta'$ ) and at least fourteen PAPs (PAP1–PAP12, FLN2, and pTAC18).<sup>8,9</sup> Identification of homologous proteins in other flowering plant species indicates this subunit composition is likely to be broadly conserved.<sup>3,10</sup> One set of PAPs potentially regulate transcriptional activity through domains that are associated with nucleic acid binding activity (PAP1/pTAC3, PAP2/pTAC2, and PAP3/pTAC10).<sup>11–13</sup> A second set of PAPs provide PEP with additional enzymatic activities: a lysine methyltransferase (PAP7/pTAC14), two iron superoxide dismutases (FeSODs; PAP4/FSD3 and PAP9/FSD2), a thioredoxin (PAP10/TRXz), and an amino acid ligase (PAP11/MurE).<sup>8,14–16</sup>

It remains unclear, however, whether these activities represent the essential role of each PAP in chloroplast biogenesis. In addition, the possible roles of the remaining PAPs (PAP5/pTAC12/HEMERA, PAP6/FLN1, PAP8/pTAC6, PAP12/pTAC7, FLN2,



and pTAC18) remains unclear from analysis of their amino acid sequences. To better understand the roles of PAPs in chloroplast transcription, as well as the mechanistic similarity between PEP and bacterial RNAP, we sought to determine the structure of PEP purified from chloroplasts.

## RESULTS

### Cryo-EM analysis of PEP

PEP was purified by chromatographic separation of chloroplast lysate from *S. alba* cotyledons (Figures 1A and S1A). Nineteen unique subunits were identified by liquid chromatography-tandem mass spectrometry (LC-MS/MS): the four polymerase core subunits ( $\alpha$ ,  $\beta$ ,  $\beta'$ , and  $\beta''$ ), the twelve named PAPs (PAPs 1–12), two proteins recently identified to co-purify with PEP (FLN2 and pTAC18), and PRIN2, a protein involved in the control of PEP activity by redox signals (Table S1).<sup>9,17</sup> Our structural analysis of PEP described below revealed that FLN2, pTAC18, and PRIN2 are stably and uniformly associated with PEP. We therefore propose to name these subunits PAP13 (FLN2), PAP14 (pTAC18), and PAP15 (PRIN2).

The molecular mass of PEP was measured to be  $\sim 1.1$  MDa by mass photometry (Figure S1B), in agreement with the predicted value (Table S2). We observed extension of RNA in a transcription elongation reaction (Figure S1C), confirming the purified PEP sample contains active polymerase. Analysis by negative-stain electron microscopy revealed PEP to be a star-shaped molecule that is 23 nm in its maximum dimension (Figure S1D).

An initial reconstruction of PEP was obtained by cryoelectron microscopy (cryo-EM) at a resolution of 2.5 Å (Figures S2A–S2G; Table S3). Significant conformational variation was observed within the dataset, limiting the resolution of peripheral regions of the molecule in particular. Improved reconstructions were obtained with resolutions of 2.3–2.9 Å by focused refinement of nine overlapping regions (Figures S2H and S2I; Table S4). The composite map generated from these reconstructions displayed density of improved clarity and uniformity, enabling construction of a structural model that includes all nineteen subunits of PEP that were identified by mass spectrometry (Figure S2J).

Although most sequences could be modeled with confidence within the composite map, the reconstruction contained regions resolved less well due to heterogeneity in the imaged particles. To interpret the cryo-EM data more completely, we constructed an additional PEP model that integrated computationally predicted models for additional domains that could be confidently placed in filtered or enhanced maps (Figure S3A; Table S5). These regions were at the periphery of the complex and included parts of PAP2 and PAP11 and all of PAP15/PRIN2.

Additional support for the accuracy of the structural model was obtained by cross-linking coupled to mass spectrometry (CLMS) (Figure S3B; Table S6). Residues connected by 41 of the 43 unique disuccinimidyl sulfoxide (DSSO) crosslinks were separated by less than the maximum expected cross-linking distance of 30 Å in our structural model (Figure S3B). Similarly, residues connected by 26 of the 36 unique sulfosuccinimidyl 4,4'-azipentanoate (sulfo-SDA) crosslinks were separated by less than the maximum cross-linking distance of 20 Å. Thus, our structural model is consistent with CLMS data.

The model of PEP contains nineteen unique subunits and two copies of the PAP10 and  $\alpha$  subunits, revealing PEP to be a 21-subunit complex (Figures 1B and 1C). PAPs encase the core polymerase and make extensive interactions with the outside surface but do not enter the central cleft that contains the transcription active site. Each PAP contacts at least one of the core polymerase subunits, with the exception of PAP9. The PAPs also closely interconnect with each other, with each PAP contacting at least one other PAP subunit, with the exception of PAP15/PRIN2. The model shows that the 5 arms within the overall star-shape comprise: (1) the  $\beta''$  sequence insertion 3 ( $\beta''$ -SI3) with PAP3, PAP4, PAP9, PAP14, and PAP15, (2) PAP11 and the C-terminal region of PAP1, (3) PAP2, PAP7, and the N-terminal region of PAP1, (4) PAP6 and one PAP10, and (5) PAP13 and the other PAP10.

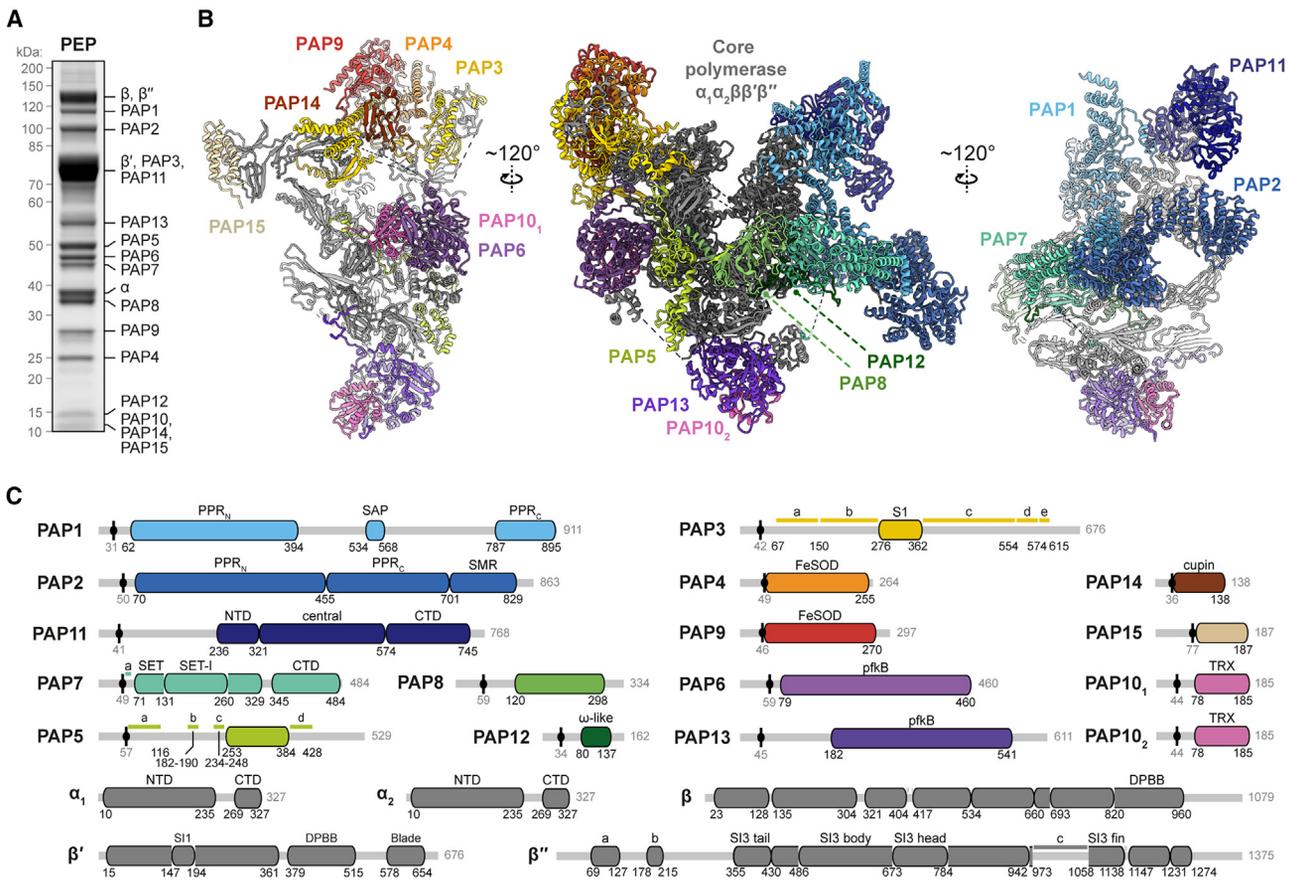
### Structural features of the PEP core polymerase

The amino acid sequences of the core polymerase subunits of PEP are similar to their bacterial RNAP counterparts. Approximately 40% of residues are identical in the aligned concatenated sequences of the core polymerase subunits of *S. alba* PEP and RNAP of the cyanobacteria *Synechocystis* sp. *PCC 6803*. Our model shows this similarity extends to the structural level (Figure 2A). Each pair of homologous subunits show only minor difference in backbone position following superposition ( $<1.6$  Å backbone root-mean-square deviation [RMSD]).

Bacterial RNAPs typically comprise 5 core subunits:  $\alpha_2\beta\beta'\omega$ . The core PEP polymerase differs in the division of  $\beta'$  into an N-terminal region (subunit  $\beta'$  encoded by the RPOC1 gene) and a C-terminal region (subunit  $\beta''$  encoded by the RPOC2 gene). This feature is shared by cyanobacteria and, consistent with structural analysis of cyanobacterial RNAP, the split of the  $\beta'$  subunits does not produce significant structural differences between PEP and other bacterial RNAPs (Figure S4A).<sup>18,19</sup>

A significant exception to the overall similarity of PEP to bacterial RNAPs is the presence of an insertion in the  $\beta''$  subunit, approximately 800 residues in length. This region, termed sequence insertion 3 ( $\beta'$ -SI3 or  $\beta''$ -SI3), contains sandwich-barrel-hybrid motifs (SBHMs) that vary in number between species. The  $\beta'$ -SI3 domain of *Escherichia coli* (*E. coli*) has two SBHMs, whereas  $\beta''$ -SI3 of cyanobacteria has nine SBHMs. In cyanobacteria, the  $\beta''$ -SI3 forms a seahorse-shaped arch, in which three SBHMs comprise the “tail” and “fin,” four SBHMs comprise the “body,” and two SHMBs comprise the “head.”<sup>19</sup> The sequence of PEP  $\beta''$ -SI3 resembles cyanobacteria in the tail and the fin (sequence identity of 30%–40%) but the remaining sequence is longer by approximately 150 amino acids and displays limited similarity. Whether this dissimilarity underlies plastid-specific functions is unknown.

Our model reveals that PEP  $\beta''$ -SI3 forms an arch strikingly like that of cyanobacterial  $\beta''$ -SI3 despite the lack of an equivalent continuous series of SHBM domains (Figure 2B). The base of the arch, comprising the tail and fin, do not contact each other and instead associate with the  $\beta''$  rim helices, similar to cyanobacterial RNAP.<sup>19</sup> This contrasts the *E. coli*  $\beta'$ -SI3 SBHM domains, which are located on the opposing side of the polymerase cleft. The two SBHM domains of the  $\beta''$ -SI3 body that connect the tail to the head in cyanobacteria are absent in PEP, and the



**Figure 1. Structure of PEP**

(A) SDS-PAGE of PEP purified from *S. alba* chloroplasts with identified subunits indicated.

(B) Structural model of PEP (center) and clipped views of the  $\beta$  and  $\beta'$  lobes (left and right, respectively).

(C) Domain organization of *S. alba* PEP subunits determined from the structural model. Positions of chloroplast target peptide cleavage sites are indicated with black circle and line.

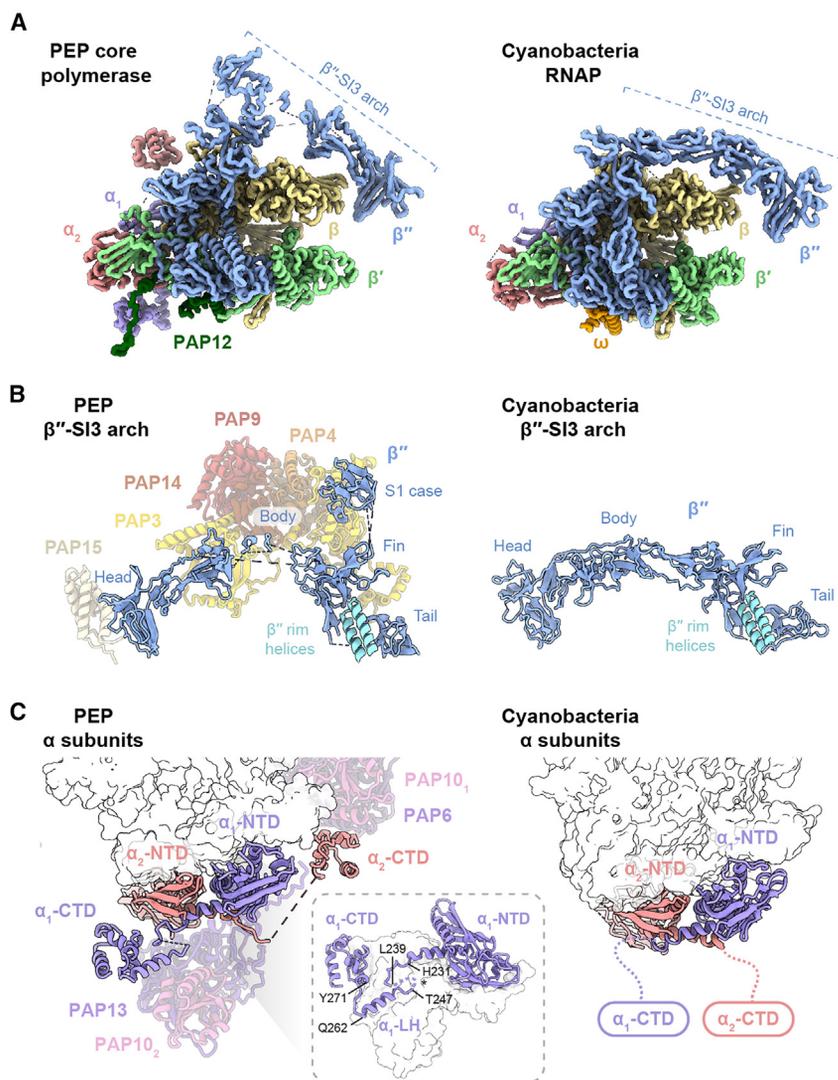
Alternative protein name of PAP13 is FLN2; alternative of PAP14 is pTAC18; alternative of PAP15 is PRIN2. Additional sequence annotations: PAP3: a,  $\beta''$ -S13 fin binding region; b, S1 case; c, arch bridge; d,  $\beta$  binding region; e,  $\beta''$ -S13 body binding region; PAP5: a, PIR1; b, PAP8 binding motif; c, rim helix binding motif; d,  $\beta$  binding region; PAP7: a,  $\alpha$ -CTD binding motif;  $\beta'$ : a, rim helices; b, bridge helix; c, S1 case; PAP4 and PAP9: FeSOD, iron superoxide dismutase; PAP6 and PAP13: pfkB, pfkB family carbohydrate kinase domain; PAP10: TRX, thioredoxin. See also [Figures S1–S3](#) and [Tables S1–S5](#).

corresponding region is largely disordered. The base and the tip of the PEP  $\beta''$ -S13 are instead bridged by a complex comprised of PAP3 and PAP14/pTAC18, which provide a structural scaffold, and the FeSODs PAP4 and PAP9. The PEP  $\beta''$ -S13 fin domain contains a sequence absent in cyanobacteria that encloses the PAP3 S1 domain, which we name the  $\beta''$  S1 case domain (Figure 2B). Despite these differences, the PEP  $\beta''$ -S13 arch is very similar in length to that of cyanobacterial RNAP (13.5 nm in both). The  $\beta''$ -S13 head is consequently also positioned adjacent to the upstream DNA channel of the polymerase. The tip of the arch, comprising the head and two adjacent SBHM domains of the body, were resolved only upon filtering the cryo-EM maps, indicating the arch displays conformational variation.

In bacterial RNAPs, the linkers connecting the N-terminal and C-terminal domains of the  $\alpha$  subunits are flexible. This flexibility allows the C-terminal domains ( $\alpha$ -CTDs) to regulate a variety of transcriptional processes through interactions with DNA, initiation activators and elongation regulators.<sup>20–22</sup> Based on previous

structural analyses of bacterial RNAPs, we expected the  $\alpha$ -CTDs not to be resolved in our PEP sample due to the lack of equivalent regulatory factors. By contrast, we identify that each of the  $\alpha$ -CTDs stably interacts with either PAP6 or PAP13, which are homologous subunits (Figure 2C). The linker between the  $\alpha$  subunit N-terminal and C-terminal domains is approximately twice as long in PEP than in bacterial RNAPs (Figures 2C and S4B). The additional linker sequence in one  $\alpha$  subunit contains a helical segment that contacts PAP13, which we term the  $\alpha$ -linker helix ( $\alpha$ -LH).

The PEP core polymerase subunits have additional sequences that are absent in cyanobacterial RNAP and contribute to PAP binding sites. The PEP  $\beta'$  sequence insertion 1 ( $\beta'$ -S11) is structurally unlike that of characterized bacterial RNAPs and contacts PAP11 (Figure S4C). A partially disordered loop near the  $\beta'$  C terminus contains a PEP-specific insertion and produces a flap that encloses the PAP8 C-terminal helix (Figure S4D). PAP8 also interacts with a domain of  $\beta'$  adjacent to this loop that we term the  $\beta'$ -blade domain due to its structural



**Figure 2. Structural comparison of PEP core polymerase and bacterial RNAP**

(A) Overall structural similarity between PEP core polymerase (left) and cyanobacteria RNAP (right). Structure used for comparison is PDB: 8GZG.

(B) Comparison of the  $\beta''$ -SI3 arch domains of PEP (left) and cyanobacteria RNAP (right).

(C) Comparison of the  $\alpha$  subunits of PEP (left) and cyanobacteria RNAP (right). The  $\alpha$ -CTDs are not resolved in reconstruction of cyanobacterial RNAP due to flexibility in the linker (dashed lines). By contrast, the  $\alpha$ -CTDs and the linker of  $\alpha_1$  are ordered in PEP. Details of the PEP  $\alpha_1$  subunit (inset) with residues of the linker indicated.

See also Figure S4.

Despite low overall sequence similarity, PAP12 and  $\omega$  have conserved residues at the interfaces with  $\beta$ ,  $\beta'$ , and  $\beta''$  (Figures 3A and S4G). The position of these residues suggest PAP12 acts similar to  $\omega$  as a complex assembly chaperone that mediates binding of  $\beta''$  to the remainder of the complex.<sup>26</sup> Outside the region homologous to  $\omega$ , PAP12 contains N-terminal and C-terminal sequences that traverse the groove between the  $\beta'$  and  $\alpha$  subunits. Thus, PAP12 likely promotes the stability of the PEP core polymerase both through interactions like those of  $\omega$  with bacterial RNAP and additional contacts specific to PEP.

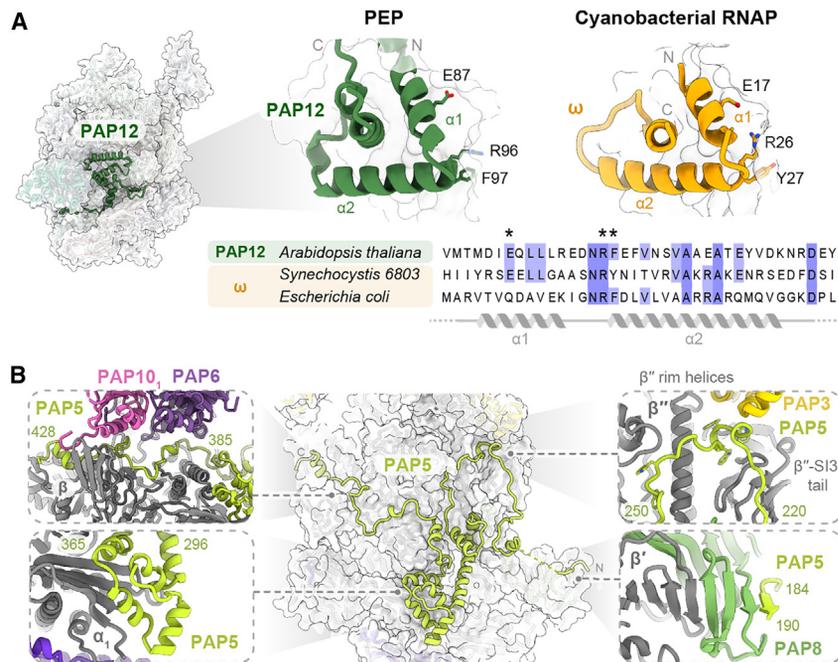
Multivalent contacts between PAP5 and the other PEP subunits suggest it is important to organization of the complex during assembly. PAP5 crosses the  $\beta$  lobe of the core polymerase, interacting with the  $\alpha$ ,  $\beta$  and  $\beta'$  and  $\beta''$  subunits (Figure 3B). An N-terminal motif of PAP5 embedded within a disordered glutamine-rich sequence binds PAP8. A tryptophan-rich motif of PAP5 wraps about the  $\beta''$  rim helices, likely stabilizing their interaction with the core polymerase and mediating their binding to the SI3-tail domain. The C-terminal portion of PAP5 traverses the  $\alpha$  and  $\beta$  subunits and contributes to the interface bound by PAP6. Collectively, the structural features of PAP5 suggest it performs a critical role in complex formation and stability that account for its requirement for accumulation of assembled PEP.<sup>27</sup>

In addition to its role as a subunit of the PEP complex in the chloroplast, PAP5 (also called pTAC12 or HEMERA) is also present in the nucleus, where it contributes to phytochrome signaling.<sup>28</sup> Dual targeting of PEP subunits is not unique to PAP5 and has also been identified for PAP8.<sup>29</sup> PAPs form oligomeric assemblies in the nucleus of unclear composition and structure.<sup>30</sup> In its nuclear role, PAP5 interacts with phytochrome A (phyA) through its N-terminal sequence (phyA-interacting region 1, PIR1), consisting of residues 1–115.<sup>31</sup> Interestingly, this region is not resolved in our model of PEP. This suggests that

resemblance to a single blade of  $\beta$ -propeller proteins (Figure S4E). The  $\beta'$ -blade domain is also present in cyanobacterial RNAP. Lastly, a short sequence within the  $\beta''$  C terminus interacts with the interface between PAP1 and PAP2 via residues not conserved in bacterial RNAPs (Figure S4F). PEP lacks only one domain that is present in cyanobacterial RNAP: the  $\beta$ - $\beta'$  module 2 (BBM2).

### PAPs contribute to PEP assembly and stability

The PAPs most extensively integrated with the core polymerase are PAP12 and PAP5. Each makes interactions with multiple core polymerase subunits and with other PAPs, suggesting they are important to complex assembly and stability. In support of a role in mediating the assembly of PEP, we identified that PAP12 is a structural homolog of the bacterial  $\omega$  subunit (Figure 3A).  $\omega$  is highly conserved, is essential in some bacteria, and is required for full transcription activity in cyanobacteria.<sup>23,24</sup> Given its importance, the apparent absence of a PEP homolog of  $\omega$  was surprising.<sup>25</sup>



**Figure 3. Contribution of PAP12 and PAP5 to PEP complex assembly and stability**

(A) Structural similarity between PAP12 in PEP (left) and  $\omega$  in cyanobacteria RNAP (right). Aligned amino acid sequences of the conserved region with residues predicted to be critical to the role of  $\omega$  in complex assembly indicated (asterisks). Structure used for comparison is PDB: 8GZG.

(B) Structural details of the interactions between the PAP5 subunit with PEP subunits. See also Figure S4.

the nuclear and chloroplastic roles are mediated through distinct regions of PAP5.

### Interactions of PEP with nucleic acids

In our cryo-EM reconstruction of PEP we observe density in the DNA entrance channel, DNA exit channel and mRNA exit channel that likely represents endogenous nucleic acids co-purified with the complex (Figure S5A). To better understand how PEP interacts with DNA and RNA during transcription, we reconstituted a PEP transcription elongation complex (PEP-TEC) using a nucleic acid scaffold and obtained a cryo-EM reconstruction at an overall resolution of 2.7 Å (Figures S5B–S5G; Table S3). A composite map was generated by combining four reconstructions obtained by focused classification and refinement that had improved density for nucleic acids (Figures S5H and S5I; Table S4).

The DNA-RNA hybrid and downstream DNA were clearly resolved, allowing us to build a structural model of PEP-TEC (Figure 4A; Table S5). The transcription active site of PEP closely resembles that of the bacterial TEC (Figure 4B). The identity and approximate positions of residues involved in coordination of magnesium ions and positioning the RNA and template DNA are conserved.

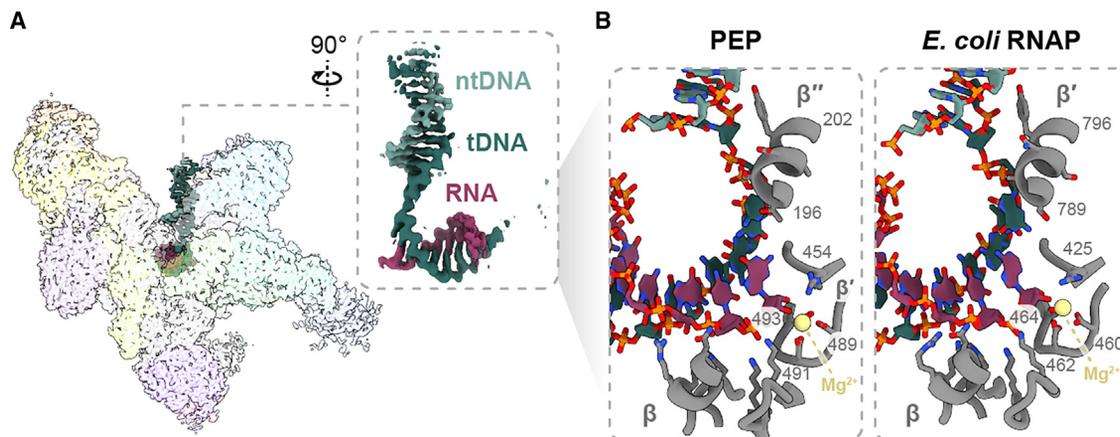
Previous data suggest four PAPs may be involved in direct interactions with nucleic acids: PAP1, PAP2, PAP3, and PAP15/PRIN2. PAP1 and PAP2 contain pentatricopeptide repeats (PPRs), which in other proteins confer sequence-specific RNA-binding activity. PAP1 also contains a SAP motif, which is a domain associated with DNA binding, and PAP2 also contains an SMR, which is a domain that in some proteins imparts nuclease activity. PAP3 contains an S1 domain, which was observed to confer RNA-binding activity to purified PAP3 in isolation from PEP.<sup>13</sup> PAP15/PRIN2 does not contain canonical

nucleic-acid-binding domains but interacts with DNA *in vitro*.<sup>32</sup> The involvement of PAPs in nucleic acid binding in the context of transcription is an important but unresolved question.

Our model of PEP-TEC shows that DNA entering the polymerase interacts with the SAP motif of PAP1 (PAP1-SAP) (Figure 5A). PAP1-SAP is within a region of ~400 amino acids between the nine N-terminal PPRs (PAP1-PPR<sub>N</sub>) and two complete and two partial PPRs at the C terminus (PAP1-PPR<sub>C</sub>) (Figure S6A). PAP1-SAP borders the DNA entrance channel on the  $\beta'$  lobe and contacts the DNA approximately 20–24 nt upstream of the transcription active site. Density for both PAP1-SAP and the downstream DNA is of limited resolution in the PEP-TEC reconstruction, indicating their positions vary with respect to the remainder of PEP (Figures 5A and S5G). An interaction between the PAP1-PPRs and nucleic acids was not observed, consistent with the absence of amino acids associated with sequence-specific RNA binding in PAP1 (Figure S6A). The PAP1-PPR<sub>N</sub> domain contacts numerous other subunits, likely performing an essential role as a molecular scaffold of the complex.

The DNA between PAP1-SAP and the transcription bubble passes a segment of the  $\beta'$  subunit that protrudes into the polymerase cleft (Figure 5B). This segment, which we term the  $\beta'$ -protrusion, is specific to PEP and absent in bacterial RNAP. The  $\beta'$ -protrusion is only partially ordered and contains conserved basic residues that likely contact the phosphate backbone of the downstream DNA.

PAP2 contains eighteen pentatricopeptide repeats (PPRs) and a C-terminal SMR domain (Figure S6B). The C-terminal PPRs 12–18 (PAP2-PPR<sub>C</sub>) are closely bound to the SMR domain and together bind PAP1, thereby connecting PAP2 to the remainder of PEP (Figure S6C). The interaction of the SMR domain with PAP1 conceals residues that potentially impart nuclease activity, supporting the conclusion that PAP2 is non-catalytic in the context of PEP. The N-terminal PPRs 1–11 (PAP2-PPR<sub>N</sub>) form a semi-circular arch that curves inwards to the mRNA exit channel of the polymerase (Figure 5C). PAP2-PPR<sub>N</sub> is resolved only in filtered cryo-EM maps, suggesting it is flexible and suspended in this location rather than in direct contact with the polymerase domains bordering the exit channel. Between the mRNA exit channel and PAP2-PPR<sub>N</sub> we observe continuous cryo-EM density,



**Figure 4. Structure of PEP transcription elongation complex**

(A) Cryo-EM map of PEP-TEC. Density assigned to nucleic acids indicated in inset (opaque) and density assigned to proteins (transparent) colored according to Figure 1.

(B) Structural details of the transcription active sites of PEP (left) and *E. coli* RNAP (right) in corresponding views. Three conserved aspartate residues in the catalytic loop coordinate a magnesium ion that positions the RNA 3' end in both enzymes: PEP  $\beta'$  residues 489, 491, and 493 and *E. coli*  $\beta'$  residues 460, 462, and 464. Conserved residues critical to the positioning of template DNA are indicated: PEP  $\beta''$  residues 196–202 and *E. coli*  $\beta'$  residues 789–796. The trigger loop is not ordered due to the absence of incoming nucleotide triphosphate. Sequence used for comparison is PDB: 6ALH.

See also Figure S5 and Tables S3–S5.

leading us to hypothesize that PAP2 interacts with the nascent mRNA. Three conserved basic residues of PAP2-PPR1 (K80, R87, and K92) project toward the mRNA exit channel and likely contribute to mRNA binding.

The PAP2 residues on the concave surface of PPRs 3–10 could support the modular sequence-specific RNA-binding mode characterized for other PPR proteins. The RNA sequence predicted to have an optimal interaction with PAP2 is AAUUGCGU (Figure 5D). Although we did not observe RNA bound to the concave surface of PAP2-PPR<sub>N</sub> in the expected position, this region is limited in resolution and the mRNA used did not contain a predicted target motif. A search of the *S. alba* chloroplast genome for the predicted target motif revealed 18 sites, of which 11 are within regions predicted to be transcribed (Table S7). Correspondence was not observed between the position of the predicted high-affinity binding sites of PAP2 and genes differentially expressed in plants lacking PAP2.<sup>12</sup>

PAP3 interacts with RNA via its S1 domain when in isolation from PEP.<sup>13</sup> This raised the possibility that PAP3 regulates PEP transcription in this manner. Contrary to this expectation, our PEP-TEC model shows PAP3 does not contact nucleic acids. PAP3 is a structural component of the  $\beta''$ -SI3 arch located far from the DNA and RNA channels of the polymerase. The S1 domain is entirely encased by protein, concealing the loops of the OB fold associated with RNA binding (Figure S6D).<sup>34</sup> We conclude that PAP3 does not interact with nascent transcripts during elongation.

Structural overlay of PEP with a model of a cyanobacterial initiation complex<sup>19</sup> indicates that plant sigma factors ( $\sigma$ ), which are likely structurally homologous to their bacterial counterparts,<sup>35</sup> can be accommodated within the PEP complex without significant steric overlap (Figure 5E). In the bacterial initiation complex, domains of  $\sigma$  that are on the RNAP surface are located adjacent to the DNA exit channel. This surface of PEP is not bound by

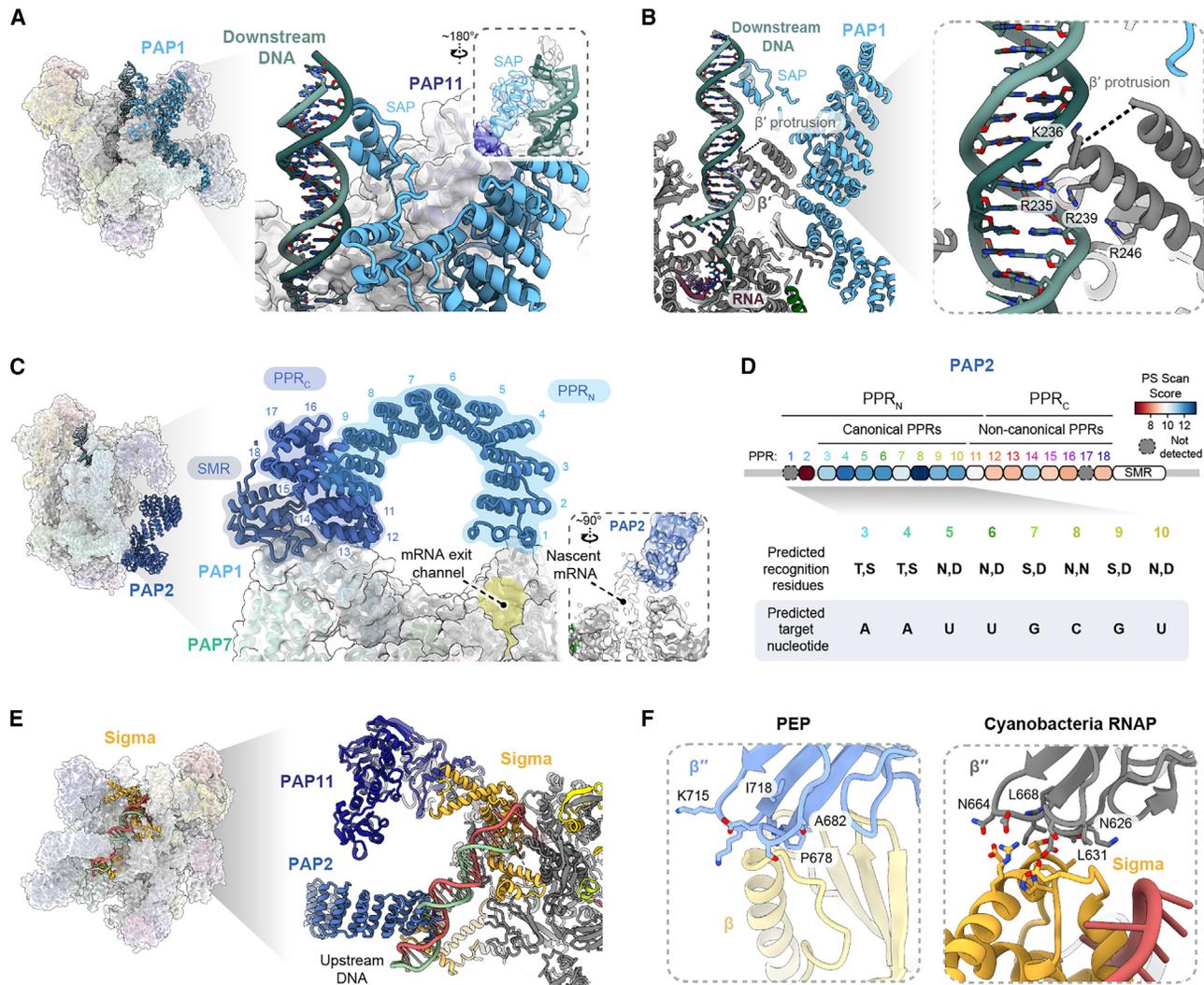
PAPs, and the expected  $\sigma$ -binding surface of PEP is structurally conserved with bacterial RNAP.

Structural comparison with bacterial initiation complexes revealed potential involvement of PAP2 and PAP11 in transcription initiation. The position of PAP2-PPR<sub>N</sub> in PEP-TEC overlaps with the expected location of the  $-35$  promoter element and the  $\sigma_4$  domain. PAP2-PPR<sub>N</sub> was observed to be flexibly associated with PEP, and a minor rotation about the connection point would avoid overlap and place the basic residues of PAP2-PPR1 adjacent to the DNA. PAP11 is adjacent to the modeled position of the  $\sigma_2$  domain, suggesting it may contribute to binding of  $\sigma$ .

In cyanobacteria, residues of the  $\beta''$ -SI3 head interact with  $\sigma$  near the promoter  $-10$  element. These residues are not conserved in PEP, indicating that an equivalent interaction between the  $\beta''$ -SI3 head and  $\sigma$ , if it occurs during PEP initiation, is mediated by a distinct interface (Figure 5F). The PEP  $\beta''$ -SI3 head domain does, however, contain a positively charged surface located close to the path of the DNA exiting the polymerase (Figure S6E). Although this potentially contacts negatively charged upstream DNA, an interaction was not observed in our reconstruction of PEP-TEC. Thus, a stable contact does not appear to be made during elongation but may contribute to other processes in the transcription cycle. PAP15/PRIN2 is bound to the  $\beta''$ -SI3 head. The position of PAP15/PRIN2 indicates it has the potential to contribute to DNA-binding activity of PEP, consistent with its observed ability to bind DNA in isolation (Figure S6E).<sup>32</sup> The surface of PAP15/PRIN2 adjacent to the DNA exit channel is negatively charged, indicating that if such an interaction occurs it is not solely electrostatic.

### Redox roles of PAPs

The transcriptional activity of PEP is regulated by redox signaling cascades arising from photosynthetic processes.<sup>36</sup> PAP10, also



**Figure 5. Interactions of PEP with nucleic acids**

(A) Structural details of the interaction between the PAP1-SAP domain and the DNA entering PEP approximately 20 nt downstream of the transcription site. Inset: overlay of PEP-TEC structural model and cryo-EM map filtered to 4 Å resolution in the region of the contact between PAP1-SAP and DNA.

(B) PEP-TEC structural model clipped view showing the β'-protrusion, a PEP-specific feature that interacts with DNA approximately 10 nt downstream of the transcription bubble. Four basic residues that likely contact the DNA phosphate backbone are indicated. Within the protrusion, β' residues 226–233 were not resolved (dashed line).

(C) The N-terminal PPR repeats of PAP2 are positioned adjacent to the mRNA exit channel. Inset: continuous density in the cryo-EM reconstruction focused on the mRNA exit channel shows a path for the emerging mRNA to the N-terminal PPR repeats of PAP2.

(D) Analysis of the PAP2 by PPRCODE<sup>33</sup> shows the PPR array can be divided into an N-terminal portion (PAP2-PPR<sub>N</sub>) that contains canonical PPRs (repeats 3–11) and divergent PPRs (repeats 1 and 2), and a C-terminal portion (PAP2-PPR<sub>C</sub>) that contain non-canonical PPRs predicted not to interact specifically with RNA (repeats 12–18). RNA-binding specificity predicted based on the PPR motif sequence for repeats 3–10 is shown.

(E) Superposition of PEP-TEC with sigma factor and nucleic acids from cyanobacterial initiation complex. Structure used for comparison is PDB: 8GZG.

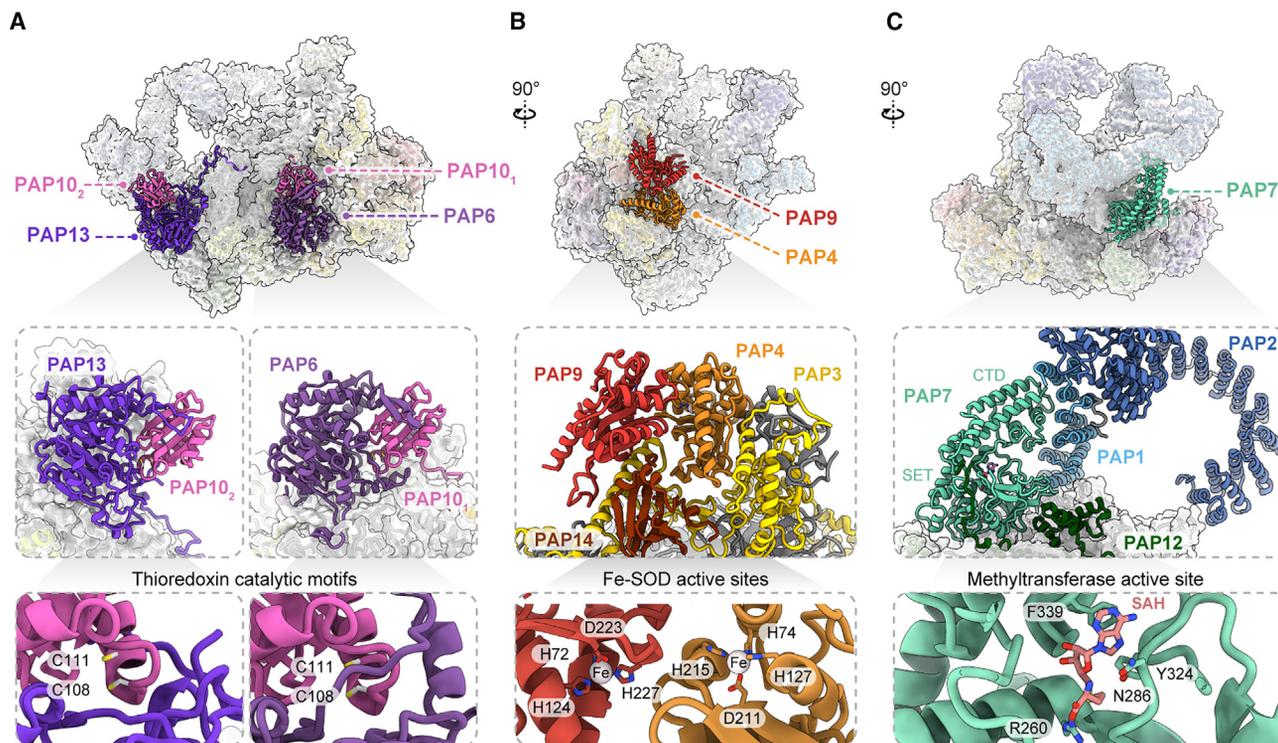
(F) Structural comparison of the β''-SI3 head domain of PEP and cyanobacteria RNAP in complex with sigma. Cyanobacteria residues that interact with sigma are two loops: β'' residues 626–631 (sequence NKDISL) and 664–668 (sequence NDIL). The corresponding regions of *S. alba* PEP are different in sequence: β'' residues 678–682 (sequence PESSA) and 715–718 (sequence KKRI) respectively. Structure used for comparison is PDB: 8GZG.

See also Figure S6 and Table S7.

called TRXz, is a thioredoxin that can be reduced by NADPH-Trx reductase C (NTRC).<sup>37</sup> PAP10 has been proposed to in turn reduce the PEP subunits PAP6/FLN1, PAP13/FLN2, and PAP15/PRIN2.<sup>15,37,38</sup> PAP10 is essential to PEP activity and chloroplast biogenesis, but, surprisingly, its thioredoxin activity is not.<sup>39</sup> How redox signaling regulates PEP transcription activ-

ity, and the contribution of PAP10 to this pathway, therefore remains unclear.

Our model reveals that PEP contains two copies of the PAP10 thioredoxin subunit. Each PAP10 is bound to the structurally homologous pseudokinase subunits PAP6 or PAP13 through corresponding surfaces (Figure 6A). The heterodimers bind distinct



**Figure 6. Structural details of enzymatic PAPs**

(A) Two PAP10 thioredoxin subunits are associated with the homologous pseudokinases PAP6/FLN1 or PAP13/FLN2 on distinct surfaces of PEP. The thioredoxin CGPC motifs are buried by these interactions and the catalytic cysteine residues are in a reduced state (bottom panel).

(B) A heterodimer of FeSOD subunits PAP4 and PAP9 is a structural component of the  $\beta''$ -SI3 arch. The active site iron ions are separated by 18 Å (bottom panel).

(C) Lysine methyltransferase subunit PAP7 interacts with the PAP1 N terminus through its C-terminal domain (CTD), likely stabilizing the position of PAP2. The PAP7 binding pocket for cofactor SAH contains residues conserved in other methyltransferase enzymes (bottom panel).

See also [Figure S7](#).

surfaces of the core polymerase: PAP6 interacts with  $\beta$ ,  $\beta'$ , and PAP5, whereas PAP13 binds the  $\alpha$  subunits. Thus, each has a unique position within PEP despite the sequence similarity between PAP6 and PAP13 (~40% identity with the chloroplast targeting peptide excluded).

Within each heterodimer, the catalytic CGPC thioredoxin motif of PAP10 is at the interface with PAP6 or PAP13. It is unlikely, however, that PAP6 and PAP13 are targets of PAP10 thioredoxin activity during PEP complex assembly: the PAP10 catalytic cysteine residues are all reduced and are not close to cysteine residues of PAP6 or PAP13. The model sheds light on experiments that suggested an inter-protein disulfide bond supports PAP10 binding to PAP6 and PAP13.<sup>15</sup> Mutation of a PAP10 catalytic cysteine was observed to reduce PAP6 binding, and our model shows that this residue is located at the interface with PAP6 and alteration could therefore destabilize the contact. Similarly, the PAP6 and PAP13 cysteine residues identified as important for binding PAP10 are within the core pfkB domains ([Figure S7A](#)), and their alteration may destabilize the fold and thereby impair binding.

#### Additional enzymatic activities of PAPs

Two PEP subunits are FeSOD enzymes: PAP4/FSD3 and PAP9/FSD2. Loss of both PAP4 and PAP9 prevents chloroplast biogenesis in *Arabidopsis* and produces sensitivity to oxidative

stress.<sup>14</sup> The essential role of PAP4 and PAP9 may therefore be to neutralize superoxide radicals produced by photosynthesis to protect the transcription machinery or the plastid genome. Yet FeSODs are not generally associated with larger complexes, and the presence of two in PEP raises questions about the molecular relationship between PAP4 and PAP9.

Our model shows that a heterodimer of PAP4 and PAP9 is bound to the  $\beta''$ -SI3 arch ([Figure 6B](#)). PAP4 is embedded in a deep cavity of the arch that is bordered on one side by a domain comprising the  $\beta''$ -SI3 case domain, and on the other by PAP14/pTAC18 and a  $\alpha$ -helix of PAP3. PAP9 interacts with PAP4 through an interface that was previously observed in a homodimeric PAP9 complex<sup>40</sup> and also contacts PAP3.

The amino acid sequences of PAP4 and PAP9 with the chloroplast targeting peptide excluded are ~55% identical, and their folds are correspondingly similar (backbone RMSD 0.36 Å). Specificity for the arrangement of the heterodimer in PEP is conferred by numerous residues conserved in, and specific to, either PAP4 or PAP9 that are mostly on the surfaces ([Figure S7B](#)). The N and C termini of PAP4 are unlike PAP9 and extensively interact with PAP3. Conversely, an extended loop present in PAP9 but not PAP4 interacts with PAP3. PAP9 is more surface-exposed than PAP4. Thus, in addition to greater SOD activity than PAP9,<sup>14</sup> PAP4 likely has the predominant structural role in PEP.

PAP11 is essential to chloroplast biogenesis in *Arabidopsis*.<sup>16</sup> The amino acid sequence of PAP11 closely resembles that of bacterial MurE, an amino acid ligase involved in peptidoglycan biosynthesis. Yet homologs of the enzymes upstream of MurE in the peptidoglycan biosynthetic pathway are absent in *Arabidopsis*, raising a significant question as to the role of PAP11. We identify that PAP11 binds the  $\beta'$ -SI1 domain and PAP1 at the periphery of PEP. These interactions are mediated by PAP11 residues in the C-terminal domain that are absent in bacterial MurE (Figure S7C). The position of the substrate-binding pocket of PAP11, predicted by structural alignment with *E. coli* MurE, is within a cleft formed by the N-terminal and central of the three globular domains. In PAP11, the N-terminal domain and loops of the central domain predicted to be involved in catalysis are poorly resolved in the cryo-EM reconstruction (Figure S7D).

Our model does not fully define whether PAP11 is likely to contribute to the transcription activity of PEP. We observe, however, that the PAP11 C-terminal domain contacts the PAP1-SAP motif that is bound to downstream DNA (Figure S7E). The cryo-EM map indicates the PAP1 sequences on either side of the SAP motif are not well ordered. It is therefore likely that the interaction of the SAP motif with PAP11 is important for positioning it such that it can interact more stably with DNA.

The structure of PAP7 supports its predicted role as a lysine methyltransferase. PAP7 is structurally similar to rubisco large subunit methyltransferase (RBLsMT) and SETD6, which comprise an N-terminal SET domain interrupted by a helical insertion (SET-I) and followed by a C-terminal domain.<sup>41,42</sup> Both the SET-I and the C-terminal domain of PAP7 display structural similarity to RBLsMT and SETD6 (Figure S7F). The cryo-EM reconstruction revealed density within the active site of PAP7 consistent with the cofactor product of methyltransferase activity, S-adenosyl homocysteine (SAH) (Figure S7G). Residues that bind SAH in RBLsMT are conserved in PAP7.<sup>43</sup>

Although our data support a role of PAP7 in methylation, the model of PEP indicates this is likely not its only function. The PAP7 SET domain is bound to the core polymerase via  $\beta''$ , PAP12, and PAP8. The PAP7 C-terminal domain interacts with the PAP1 N terminus that is the binding site for PAP2. We therefore hypothesize that PAP7 aids in positioning PAP2 at the mRNA exit channel (Figure 6C). PAP7 is essential to chloroplast biogenesis,<sup>44</sup> and further analysis will be needed to assess whether its methyltransferase activity or its role in complex stabilization underlies this.

## DISCUSSION

### PAPs encase a prokaryotic core polymerase

The structural models presented here reveal that the chloroplast transcription machinery consists of a plastid-encoded core enzyme encased by nuclear-encoded subunits (Figure 7A). The precise composition of PEP and the position of the PAP subunits with respect to the polymerase were not known prior to this work. We identify that PEP contains fifteen unique PAPs and two copies of the thioredoxin subunit PAP10 (Figure 1). Notably, the PAP12 subunit is homologous to the conserved  $\omega$  subunit of bacterial RNAP, suggesting it originates from a plastid gene transferred

to the nuclear genome (Figure 3A). The overall arrangement of subunits in PEP shows clear similarity with that of the photosynthetic complexes, in which the plastid-encoded reaction center is generally surrounded by nuclear-encoded subunits.

The PAP subunits augment the prokaryotic polymerase with DNA-binding and RNA-binding capability, imparted by PAP1 and PAP2, respectively (Figure 5), and enzymatic activities, imparted by PAP4, PAP7, PAP9, PAP10, and PAP11 (Figure 6). These additional activities are distributed across the surface of the polymerase (Figure 7B). Each PAP interacts directly with at least one core polymerase subunit with the exception of PAP9. The PAPs also interact extensively with each other to form a continuous belt that spans the  $\beta$  and  $\beta'$  lobes of the polymerase (Figure 1B). These connections do not, however, prevent significant movement of the lobes with respect to each other, as indicated by the effect of focused refinement of our cryo-EM data (Figure S2).

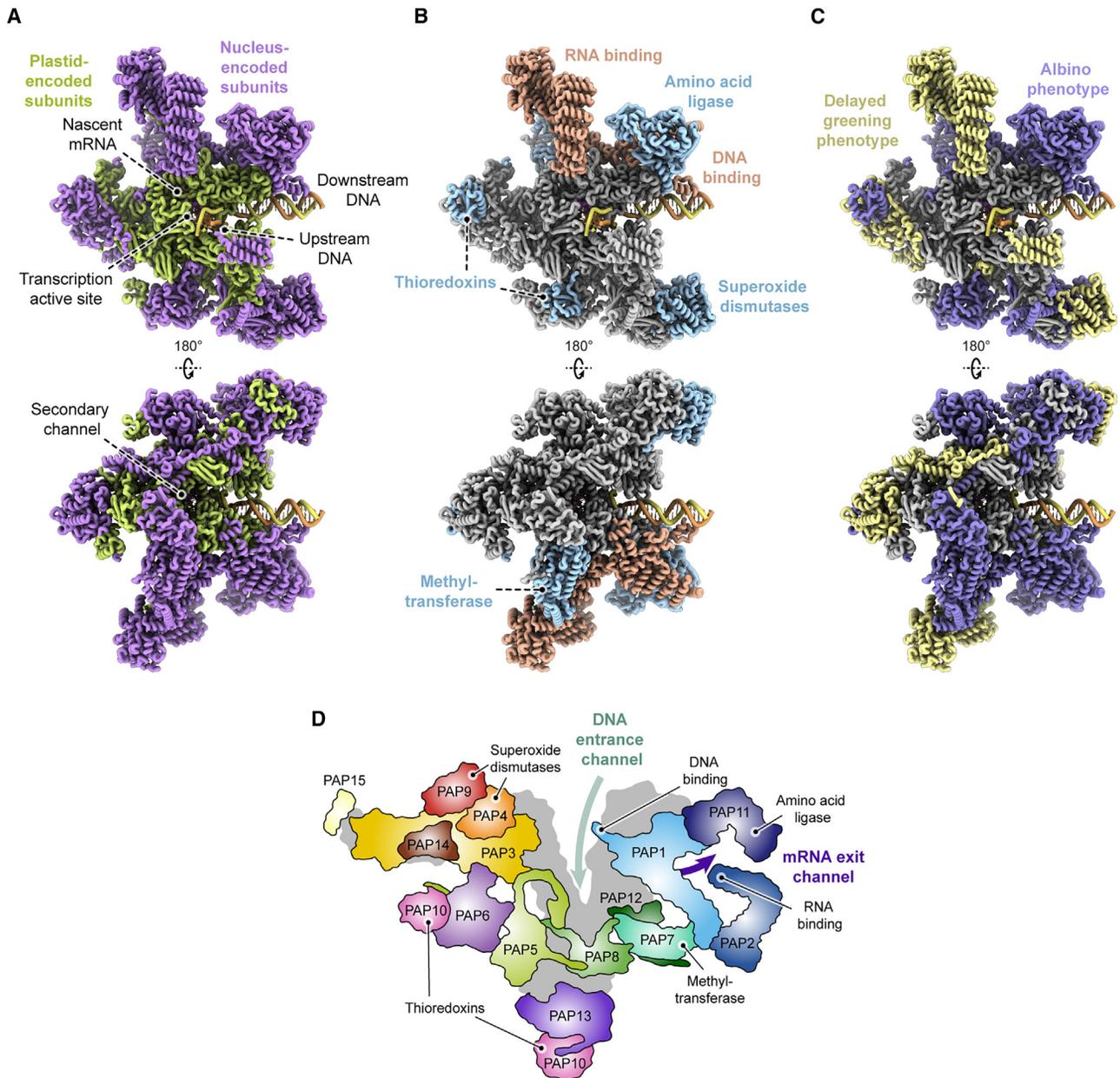
It is notable that all fifteen PAP subunits are resolved within our cryo-EM reconstructions, indicating their positions do not substantially vary with respect to the core polymerase. Furthermore, all regions of PAPs that are predicted to form independent structural domains were resolved (Figure 1C). Thus, PEP adopts a structural state that is relatively well defined in the absence of additional regulatory proteins or nucleic acid sequence elements.

### Structural basis of chloroplast transcription

The high level of conservation between PEP and bacterial RNAP in terms of both polymerase architecture (Figure 2A) and individual residues involved in nucleotide addition (Figure 4B) indicate the chloroplast transcription reaction likely occurs by a mechanism closely resembling that of bacteria. As PAPs are associated with the outer surface of the polymerase and do not enter the central cleft, we conclude they do not contribute directly to the nucleotide addition cycle. In addition, PAPs do not associate with the polymerase secondary channel that in bacterial RNAP is the binding site of transcript cleavage factors involved in reactivating backtracked complexes (Figure 7A).<sup>45,46</sup>

A significant reduction in the levels of PEP-dependent transcripts is a hallmark of mutant plants lacking any individual PAP.<sup>3</sup> Whether this represents a direct positive effect of each PAP on transcriptional output is unclear, however, as the integrity of the polymerase complex may be concurrently lost. Our data demonstrate direct involvement of two PAPs in chloroplast transcription elongation. PAP1 contacts the DNA entering the polymerase (Figure 5A). This may modulate transcription processivity and backtracking and aid the clearance of proteins bound to the downstream DNA. PAP2 contacts nascent mRNA that exits the polymerase (Figure 5D). This suggests PAP2 functions as a co-transcriptional RNA chaperone that likely regulates pausing and intrinsic termination.<sup>47</sup> In bacteria, contact with the nascent mRNA supports the transcription elongation regulation activity of NusA, a protein without a known homolog in chloroplasts.<sup>48</sup>

Involvement of the other PAPs in transcription regulation was not observed in the canonical TEC we studied. However, additional structural states occur during the transcription cycle and their characterization is expected to shed light on additional aspects of regulatory control that may involve other PAPs. For



**Figure 7. Activities and properties of PEP subunits**

(A–C) Structural model of PEP-TEC colored by (A) genome origin of each subunit: plastid (green) and nucleus (purple), (B) identified PAP activity: nucleic acid binding (orange) and additional enzymatic activity (blue), and (C) phenotypic consequence of null mutation in each PAP subunit: albino (purple) and delayed greening (yellow).

(D) A topological representation of the PAP subunits shows their relative positions within the PEP complex, inter-subunit connectivity, nucleic acid binding activities identified in this study, and likely additional enzymatic activities.

example, PAPs form a belt that spans the lobes of the polymerase and likely effects clamp opening movements associated with initiation and elongation and swiveling movements associated with pausing.<sup>49,50</sup> A further PAP-dependent transcription regulatory mechanism likely involves the controlled release of the  $\alpha$ -CTDs. The  $\alpha$ -CTDs are essential to transcription initiation in bacteria and depend on flexible association with the polymerase to bind promoter DNA and activator proteins.<sup>51,52</sup> A mecha-

nistically similar process in chloroplasts depends on the release of the  $\alpha$ -CTDs from their binding sites on the pseudokinase subunits, PAP6 and PAP13.

#### PAP mutant phenotypes reflect PEP structural organization

The loss of any individual PAP has a significant effect on chloroplast biogenesis. *Arabidopsis* null mutants lacking any of

nine PAP subunits display an albino phenotype in which chloroplast biogenesis is arrested at an early stage: PAP1, PAP3, PAP4, PAP6, PAP7, PAP8, PAP10, PAP11, and PAP12/ $\omega$ .<sup>11,12,14–16,44,53,54</sup> Disruption of any of a further 5 subunits causes a pale green phenotype in which chloroplast maturation is delayed and plant growth is inhibited: PAP2, PAP5, PAP9, PAP13/FLN2, and PAP15/PRIN2.<sup>12,14,17</sup> Disruption of the remaining PAP, PAP14/pTAC18 remains unexamined.

The dependence of PEP activity on each of its constituent subunits raises the important question of whether each PAP performs an independent essential role, or whether PAPs depend on each other such that the loss of one PAP impairs the overall assembly or stability of PEP. Mapping the phenotypic categories onto the structural model of PEP reveals the more severe albino phenotypes are associated with PAPs more closely associated with the core polymerase and the less severe delayed greening phenotype are associated with more peripheral PAPs (Figure 7C). The structural data thereby support a model in which an albino phenotype generally arises from compromised PEP complex integrity whereas the less severe phenotypes could be produced by compromised transcription activity in an otherwise intact complex.

### Role of PAPs in PEP assembly and stability

The architecture of PEP indicates that the plastid-encoded core polymerase subunits could assemble prior to the nuclear-encoded subunits. The plastid-encoded polymerase subunits interact extensively through interfaces analogous to bacterial RNAP (Figure 2A), and no interfaces between plastid-encoded polymerase subunits were identified that would depend on the preceding binding of nuclear-encoded PAPs.

This observation supports the viability of a model of chloroplast biogenesis in which a core polymerase complex may assemble prior to chloroplast biogenesis, followed by the incorporation of PAPs upon exposure to light.<sup>55</sup> The transition from the smaller core polymerase (termed “PEP-B”) to the complete complex (“PEP-A”) may underlie the activation of photosynthetic transcript production in the chloroplast that leads to greening. However, our data do not reveal clearly why a complex consisting of only the core polymerase would be inactive in non-green plastids given its similarity to active bacterial RNAP. Interestingly, the extensive interaction network between PAPs (Figure 7D) indicates that almost all the nuclear-encoded subunits, with the exception of PAP13/FLN2 and PAP15/PRIN2, could likewise associate prior to their assembly with the plastid-encoded core polymerase.

Despite the potential for the core polymerase to assemble in the absence of PAPs, our data suggest that PAPs contribute to the folding of the polymerase into a state that is transcriptionally competent. PAP12 is a structural homolog of the bacterial  $\omega$  subunit, and therefore likely similarly aids the assembly of  $\beta''$  with the remainder of the complex (Figure 3A). The  $\beta''$ -SI3 domain is expected not to form a stable arch in the absence of PAP3, PAP4, PAP9, and PAP14/pTAC18 (Figure 2A). Although the role of the  $\beta''$ -SI3 arch of PEP in transcription has not been defined, if it is found to be needed for the transcription of at least one essential chloroplast gene, we predict that the four PAPs within the  $\beta''$ -SI3 arch provide a required structural role.

We hypothesize that other PAPs also perform significant structural roles within the complex related to the correct posi-

tioning of other PAPs. It is notable that PAP2 can contact the nascent mRNA even though it does not interact with proteins surrounding the mRNA exit channel (Figure 5B). The region of PAP2 that contacts RNA is instead positioned by an interaction between PAP2 and PAP1 that is  $\sim 70$  Å away from the PAP2-mRNA contact point. The ability of PAP2 to contact RNA therefore depends on PAP1. The position of the PAP1 domain that contacts PAP2 appears to be stabilized by interactions with PAP7 (Figure 6C). Thus, if contact between PAP2 and RNA is required for the transcription of an essential chloroplast gene, PAP7 likewise facilitates this interaction. PAP8 may, in turn, position PAP7, as it interacts with the side of PAP7 distal to the PAP7-PAP1 contact point. Finally, we hypothesize that PAP11 is involved in positioning the PAP1-SAP domain. If the interactions we observe between PAP1-SAP and downstream DNA (Figure 5A) are essential, a critical role of PAP11 is likely to facilitate this interaction.

We therefore hypothesize the structural roles of each PAP could be classified approximately in the following way. Firstly, two PAPs (PAP12 and PAP5) support assembly of the core polymerase (Figure 3). Secondly, four PAPs (PAP3, PAP4, PAP9, and PAP14/pTAC18) support the formation of the  $\beta''$ -SI3 arch (Figure 2A). Thirdly, three PAPs (PAP1, PAP7, and PAP8) support the position of PAP2 at the mRNA exit channel (Figures 5C and 6C). Finally, one PAP (PAP11) is required to position the PAP1-SAP domain at the DNA entrance channel (Figure S7E).

### Role of PAPs in oxidative stress protection

The PAP4 and PAP9 subunits interact in isolation from the PEP complex and together display SOD activity *in vitro*.<sup>14</sup> PAP4 and PAP9 also independently display SOD activity. It is likely that this depends on the formation of homodimeric complexes, as inter-subunit cooperation has been observed for FeSODs during catalysis.<sup>56</sup> Homodimerization has been observed in the case of PAP9.<sup>40</sup> The activities of PAP4 and of PAP9 are lower, however, than that of the PAP4-PAP9 heterodimer.<sup>14</sup> Our model shows that interactions between the FeSOD subunits and other components of PEP will select for the heterodimeric PAP4-PAP9 complex, rather than homodimeric states (Figure S7B). Thus, the incorporation of PAP4 and PAP9 in PEP is expected to promote their SOD activity by stabilizing their interaction with each other.

The role of PEP-associated SOD activity is not well understood. *Arabidopsis* lacking PAP4 or PAP9 are more sensitive to oxidative stress, suggesting they contribute to the detoxification of reactive oxygen species (ROS).<sup>14</sup> A further hypothesis is that the hydrogen peroxide produced by the SOD reaction could regulate transcription: possibly providing a positive feedback loop in which ROS produced by photosynthetic activity stimulate production of new photosynthetic proteins.<sup>57</sup> Our PEP-TEC model (Figure 4) shows that the FeSOD subunits of PEP are positioned approximately 80 Å from the DNA entering the polymerase, potentially providing support for a model in which transcribed regions of the chloroplast genome are protected from ROS.

### Limitations of the study

This study is limited to a structural analysis of PEP and we propose several hypotheses that require investigation through biochemical experiments and plant genetic manipulation. In

addition, the structural information presented is limited to a model of PEP in an active elongation state. Additional study is needed to understand mechanistic details of chloroplast transcription initiation, regulation of transcription elongation, pausing, and termination. These processes involve protein factors and regulatory DNA sequences beyond the core complex characterized here.<sup>58,59</sup> Finally, the composition and structure of PEP may change during plant development and vary between plant species. Our data are limited to the state of the PEP complex in the cotyledons of *Sinapis alba*.

## STAR★METHODS

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

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead Contact
  - Material Availability
  - Data and Code Availability
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
- **METHOD DETAILS**
  - Plant growth and chloroplast isolation
  - Chloroplast lysate preparation
  - Purification of PEP
  - Mass photometry
  - Negative stain electron microscopy
  - Cryo-EM sample preparation and data acquisition
  - Cryo-EM data processing
  - Model building
  - RNA extension assay
  - Cross-linking mass spectrometry
  - Liquid chromatography mass spectrometry
  - Protein conservation and sequence features

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2024.01.036>.

## ACKNOWLEDGMENTS

We thank Gerhard Saalbach and Carlo Martins for assistance with proteomic analysis, Abbas Maqbool for assistance with mass photometry, David Lawson, Jake Richardson, and Dmitry Ghilarov for assistance with cryo-EM analysis, and members of the JIC horticultural service provision for assistance with obtaining plant material. We acknowledge the Sydney Mass Spectrometry Core Research Facility at the University of Sydney for providing access to mass spectrometers and thank the technical staff for the maintenance of the instruments. We thank all members of the Webster group for critical discussions and Lori Passmore for feedback on the manuscript. This work was funded by the BBSRC Institute Strategic Programme GEN (BB/P013511/1) and the BBSRC Institute Strategic Programme BRIC (BB/X01102X/1), a Royal Society Research Grant (RGS/R2/222157), a John Innes Foundation studentship to A.V.-C., and John Innes Centre strategic funding. We acknowledge the Diamond Light Source for access and the support of the cryo-EM facilities at the UK national electron Bio-Imaging Centre (eBIC) and proposal NT33824, funded by the Wellcome Trust, MRC, and BBSRC.

## AUTHOR CONTRIBUTIONS

A.V.-C., I.P., D.P., and M.W.W. performed sample preparation; V.K.V., M.J.B., and M.W.W. performed cryo-EM data acquisition; I.P. and M.W.W. performed cryo-EM data analysis and atomic model building; A.V.-C. performed biochemical experiments; A.V.-C. and I.P. performed mass photometry; J.K.K.L. performed cross-linking mass spectrometry; A.V.-C. and M.W.W. performed bioinformatic analyses; M.W.W. designed and supervised the research; A.V.-C., I.P., and M.W.W. prepared the manuscript with input from all authors.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 10, 2023

Revised: December 18, 2023

Accepted: January 24, 2024

Published: February 29, 2024

## REFERENCES

1. Barkan, A. (2011). Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol.* 155, 1520–1532.
2. Tadini, L., Jeran, N., Peracchio, C., Masiero, S., Colombo, M., and Pesaresi, P. (2020). The plastid transcription machinery and its coordination with the expression of nuclear genome: plastid-Encoded polymerase, Nuclear-Encoded Polymerase and the Genomes Uncoupled 1-mediated retrograde communication. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 375, 20190399.
3. Pfannschmidt, T., Blanvillain, R., Merendino, L., Courtois, F., Chevalier, F., Liebers, M., Grüber, 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.
4. Yagi, Y., and Shiina, T. (2012). Evolutionary aspects of plastid proteins involved in transcription: the transcription of a tiny genome is mediated by a complicated machinery. *Transcription* 3, 290–294.
5. Apel, K., and Bogorad, L. (1976). Light-induced increase in the activity of maize plastid DNA-dependent RNA polymerase. *Eur. J. Biochem.* 67, 615–620.
6. Smith, H.J., and Bogorad, L. (1974). The polypeptide subunit structure of the DNA-dependent RNA polymerase of *Zea mays* chloroplasts. *Proc. Natl. Acad. Sci. USA* 71, 4839–4842.
7. Chi, W., He, B., Mao, J., Jiang, J., and Zhang, L. (2015). Plastid sigma factors: their individual functions and regulation in transcription. *Biochim. Biophys. Acta* 1847, 770–778.
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. 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.
10. 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.
11. 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.

12. 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.
13. Jeon, Y., Jung, H.J., Kang, H., Park, Y.I., Lee, S.H., and Pai, H.S. (2012). S1 domain-containing STF modulates plastid transcription and chloroplast biogenesis in *Nicotiana benthamiana*. *New Phytol.* *193*, 349–363.
14. Myouga, F., Hosoda, C., Umezawa, T., Iizumi, H., Kuromori, T., Motohashi, R., Shono, Y., Nagata, N., Ikeuchi, M., and Shinozaki, K. (2008). A hetero-complex of iron superoxide dismutases defends chloroplast nucleoids against oxidative stress and is essential for chloroplast development in *Arabidopsis*. *Plant Cell* *20*, 3148–3162.
15. Arsova, B., Hoja, U., Wimmelbacher, M., Greiner, E., Ustün, 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.
16. 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.
17. Kindgren, P., Kremnev, D., Blanco, N.E., de Dios Barajas López, J., Fernández, A.P., Tellgren-Roth, C., Kleine, T., Small, I., and Strand, A. (2012). The plastid redox insensitive 2 mutant of *Arabidopsis* is impaired in PEP activity and high light-dependent plastid redox signalling to the nucleus. *Plant J.* *70*, 279–291.
18. Bergslund, K.J., and Haselkorn, R. (1991). Evolutionary relationships among eubacteria, cyanobacteria, and chloroplasts: evidence from the rpoC1 gene of *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* *173*, 3446–3455.
19. Shen, L., Lai, G., You, L., Shi, J., Wu, X., Pui, M., Gu, Z., Feng, Y., Yuzenkova, Y., and Zhang, Y. (2023). An Sl3- $\sigma$  arch stabilizes cyanobacteria transcription initiation complex. *Proc. Natl. Acad. Sci. USA* *120*, e2219290120.
20. Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R.L. (1993). A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* *262*, 1407–1413.
21. Benoff, B., Yang, H., Lawson, C.L., Parkinson, G., Liu, J., Blatter, E., Ebright, Y.W., Berman, H.M., and Ebright, R.H. (2002). Structural basis of transcription activation: the CAP-alpha CTD-DNA complex. *Science* *297*, 1562–1566.
22. Mah, T.F., Kuznedelov, K., Mushegian, A., Severinov, K., and Greenblatt, J. (2000). The alpha subunit of *E. coli* RNA polymerase activates RNA binding by NusA. *Genes Dev.* *14*, 2664–2675.
23. Mao, C., Zhu, Y., Lu, P., Feng, L., Chen, S., and Hu, Y. (2018). Association of  $\omega$  with the C-terminal Region of the  $\beta'$  Subunit Is Essential for Assembly of RNA polymerase in *Mycobacterium tuberculosis*. *J. Bacteriol.* *200*, e00159–18.
24. Gunnellus, L., Hakkila, K., Kurkela, J., Wada, H., Tyystjärvi, E., and Tyystjärvi, T. (2014). The omega subunit of the RNA polymerase core directs transcription efficiency in cyanobacteria. *Nucleic Acids Res.* *42*, 4606–4614.
25. Kurkela, J., Fredman, J., Salminen, T.A., and Tyystjärvi, T. (2021). Revealing secrets of the enigmatic omega subunit of bacterial RNA polymerase. *Mol. Microbiol.* *115*, 1–11.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.
31. Galvão, R.M., Li, M., Kothadia, S.M., Haskel, J.D., Decker, P.V., Van Buskirk, E.K., and Chen, M. (2012). Photoactivated phytochromes interact with HEMERA and promote its accumulation to establish photomorphogenesis in *Arabidopsis*. *Genes Dev.* *26*, 1851–1863.
32. Kremnev, D., and Strand, A. (2014). Plastid encoded RNA polymerase activity and expression of photosynthesis genes required for embryo and seed development in *Arabidopsis*. *Front. Plant Sci.* *5*, 385.
33. Yan, J., Yao, Y., Hong, S., Yang, Y., Shen, C., Zhang, Q., Zhang, D., Zou, T., and Yin, P. (2019). Delineation of pentatricopeptide repeat codes for target RNA prediction. *Nucleic Acids Res.* *47*, 3728–3738.
34. Bycroft, M., Hubbard, T.J., Proctor, M., Freund, S.M., and Murzin, A.G. (1997). The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell* *88*, 235–242.
35. Puthiyaveetil, S., McKenzie, S.D., Kayanja, G.E., and Ibrahim, I.M. (2021). Transcription initiation as a control point in plastid gene expression. *Biochim. Biophys. Acta Gene Regul. Mech.* *1864*, 194689.
36. Pfannschmidt, T.N., Nilsson, A., and Allen, J.F. (1999). Photosynthetic control of chloroplast gene expression. *Nature* *397*, 625–628.
37. Yoshida, K., and Hisabori, T. (2016). Two distinct redox cascades cooperatively regulate chloroplast functions and sustain plant viability. *Proc. Natl. Acad. Sci. USA* *113*, E3967–E3976.
38. 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.
39. 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.
40. Favier, A., Gans, P., Boeri Erba, E., 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.
41. Chang, Y., Levy, D., Horton, J.R., Peng, J., Zhang, X., Gozani, O., and Cheng, X. (2011). Structural basis of SETD6-mediated regulation of the NF-kB network via methyl-lysine signaling. *Nucleic Acids Res.* *39*, 6380–6389.
42. Trievel, R.C., Beach, B.M., Dirk, L.M., Houtz, R.L., and Hurley, J.H. (2002). Structure and catalytic mechanism of a SET domain protein methyltransferase. *Cell* *111*, 91–103.
43. Trievel, R.C., Flynn, E.M., Houtz, R.L., and Hurley, J.H. (2003). Mechanism of multiple lysine methylation by the SET domain enzyme RuBisCO LSMT. *Nat. Struct. Biol.* *10*, 545–552.
44. Grübler, B., Merendino, L., Twardziok, S.O., Mininno, M., Allorent, G., Chevalier, F., Liebers, M., Blanvillain, R., Mayer, K.F.X., Lerbs-Mache, S., et al. (2017). Light and plastid signals regulate different sets of genes in the albino mutant Pap7-1. *Plant Physiol.* *175*, 1203–1219.
45. Opalka, N., Chlenov, M., Chacon, P., Rice, W.J., Wriggers, W., and Darst, S.A. (2003). Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell* *114*, 335–345.
46. Abdelkareem, M., Saint-André, C., Takacs, M., Papai, G., Crucifix, C., Guo, X., Ortiz, J., and Weixlbaumer, A. (2019). Structural basis of

- transcription: RNA polymerase backtracking and its reactivation. *Mol. Cell* 75, 298–309.e4.
47. Said, N., and Wahl, M.C. (2021). Transcription complexes as RNA chaperones. *Transcription* 12, 126–155.
  48. Guo, X., Myasnikov, A.G., Chen, J., Crucifix, C., Papai, G., Takacs, M., Schultz, P., and Weixlbaumer, A. (2018). Structural basis for NusA stabilized transcriptional pausing. *Mol. Cell* 69, 816–827.e4.
  49. Chakraborty, A., Wang, D., Ebricht, Y.W., Korlann, Y., Kortkhonja, E., Kim, T., Chowdhury, S., Wigneshweraraj, S., Irschik, H., Jansen, R., et al. (2012). Opening and closing of the bacterial RNA polymerase clamp. *Science* 337, 591–595.
  50. 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.
  51. Feng, Y., Zhang, Y., and Ebricht, R.H. (2016). Structural basis of transcription activation. *Science* 352, 1330–1333.
  52. Hubin, E.A., Lilic, M., Darst, S.A., and Campbell, E.A. (2017). Structural insights into the mycobacteria transcription initiation complex from analysis of X-ray crystal structures. *Nat. Commun.* 8, 16072.
  53. Chang, S.H., Lee, S., Um, T.Y., Kim, J.K., Do Choi, Y., and Jang, G. (2017). pTAC10, a key subunit of plastid-encoded RNA polymerase, promotes chloroplast development. *Plant Physiol.* 174, 435–449.
  54. Yu, Q.B., Lu, Y., Ma, Q., Zhao, T.T., Huang, C., Zhao, H.F., Zhang, X.L., Lv, R.H., and Yang, Z.N. (2013). TAC7, an essential component of the plastid transcriptionally active chromosome complex, interacts with FLN1, TAC10, TAC12 and TAC14 to regulate chloroplast gene expression in *Arabidopsis thaliana*. *Physiol. Plant.* 148, 408–421.
  55. 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.
  56. Muñoz, I.G., Moran, J.F., Becana, M., and Montoya, G. (2005). The crystal structure of an eukaryotic iron superoxide dismutase suggests intersubunit cooperation during catalysis. *Protein Sci.* 14, 387–394.
  57. Pilon, M., Ravet, K., and Tapken, W. (2011). The biogenesis and physiological function of chloroplast superoxide dismutases. *Biochim. Biophys. Acta* 1807, 989–998.
  58. Ding, S., Zhang, Y., Hu, Z., Huang, X., Zhang, B., Lu, Q., Wen, X., Wang, Y., and Lu, C. (2019). mTERF5 acts as a transcriptional pausing factor to positively regulate transcription of chloroplast psbEFLJ. *Mol. Plant* 12, 1259–1277.
  59. Xiong, H.B., Pan, H.M., Long, Q.Y., Wang, Z.Y., Qu, W.T., Mei, T., Zhang, N., Xu, X.F., Yang, Z.N., and Yu, Q.B. (2022). AtNusG, a chloroplast nucleoid protein of bacterial origin linking chloroplast transcriptional and translational machineries, is required for proper chloroplast gene expression in *Arabidopsis thaliana*. *Nucleic Acids Res.* 50, 6715–6734.
  60. 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, e42166.
  61. Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* 14, 331–332.
  62. Rohou, A., and Grigorieff, N. (2015). CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* 192, 216–221.
  63. Meng, E.C., Goddard, T.D., Pettersen, E.F., Couch, G.S., Pearson, Z.J., Morris, J.H., and Ferrin, T.E. (2023). UCSF ChimeraX: tools for structure building and analysis. *Protein Sci.* 32, e4792.
  64. 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.
  65. Tan, Y.Z., Baldwin, P.R., Davis, J.H., Williamson, J.R., Potter, C.S., Carragher, B., and Lyumkis, D. (2017). Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* 14, 793–796.
  66. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* 75, 861–877.
  67. Holm, L., Laiho, A., Törönen, P., and Salgado, M. (2023). DALI shines a light on remote homologs: one hundred discoveries. *Protein Sci.* 32, e4519.
  68. Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
  69. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589.
  70. Jamali, K., Käll, L., Zhang, R., Brown, A., Kimanius, D., and Scheres, S.H.W. (2023). Automated model building and protein identification in cryo-EM maps. Preprint at bioRxiv. <https://doi.org/10.1101/2023.05.16.541002>.
  71. Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Carazo, J.M., Sorzano, C.O.S., and Vargas, J. (2021). DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* 4, 874.
  72. Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* 11, 2301–2319.
  73. Crowder, D.A., Sarpe, V., Amaral, B.C., Brodie, N.I., Michael, A.R.M., and Schriemer, D.C. (2023). High-sensitivity proteome-scale searches for crosslinked peptides using CRIMP 2.0. *Anal. Chem.* 95, 6425–6432.
  74. Chen, Z.L., Meng, J.M., Cao, Y., Yin, J.L., Fang, R.Q., Fan, S.B., Liu, C., Zeng, W.F., Ding, Y.H., Tan, D., et al. (2019). A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat. Commun.* 10, 3404.
  75. Katoh, K., Rozewicki, J., and Yamada, K.D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20, 1160–1166.
  76. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191.
  77. Madeira, F., Pearce, M., Tivey, A.R.N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., and Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* 50, W276–W279.
  78. Emms, D.M., and Kelly, S. (2022). SHOOT: phylogenetic gene search and ortholog inference. *Genome Biol.* 23, 85.
  79. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018.
  80. Webster, M.W., Takacs, M., Zhu, C., Vidmar, V., Eduljee, A., Abdelkareem, M., and Weixlbaumer, A. (2020). Structural basis of transcription-translation coupling and collision in bacteria. *Science* 369, 1355–1359.
  81. Li, L., Fang, C., Zhuang, N., Wang, T., and Zhang, Y. (2019). Structural basis for transcription initiation by bacterial ECF  $\sigma$  factors. *Nat. Commun.* 10, 1153.
  82. Kang, J.Y., Olinares, P.D., 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.
  83. Du, X., Zeng, T., Feng, Q., Hu, L., Luo, X., Weng, Q., He, J., and Zhu, B. (2020). The complete chloroplast genome sequence of yellow mustard

- (*Sinapis alba* L.) and its phylogenetic relationship to other Brassicaceae species. *Gene* 731, 144340.
84. Vizcaino, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., et al. (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 44, D447–D456.
85. Sousa, F., Civián, P., Foster, P.G., and Cox, C.J. (2020). The chloroplast land plant phylogeny: analyses employing better-fitting tree- and site-heterogeneous composition models. *Front. Plant Sci.* 11, 1062.
86. Karpenahalli, M.R., Lupas, A.N., and Söding, J. (2007). TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. *BMC Bioinformatics* 8, 2.
87. Fukui, K., and Kuramitsu, S. (2011). Structure and function of the Small MutS-related domain. *Mol. Biol. Int.* 2011, 691735.
88. Zhou, W., Lu, Q., Li, Q., Wang, L., Ding, S., Zhang, A., Wen, X., Zhang, L., and Lu, C. (2017). PPR-SMR protein SOT1 has RNA endonuclease activity. *Proc. Natl. Acad. Sci. USA* 114, E1554–E1563.
89. Liu, S., Melonek, J., Boykin, L.M., Small, I., and Howell, K.A. (2013). PPR-SMRs: ancient proteins with enigmatic functions. *RNA Biol.* 10, 1501–1510.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Rabbit anti-(chloroplast RPOB)	Agrisera	Cat#AS15 2867
Chicken anti-Rabbit IgG HRP Conjugate	Agrisera	Cat#AS10 839; RRID:AB_10748986
<b>Chemicals, peptides, and recombinant proteins</b>		
Tris base	Melford	Cat#T60040
Potassium chloride (KCl)	Sigma-Aldrich	Cat#P9541
Potassium acetate (KOAc)	Sigma-Aldrich	Cat#P1190
Sodium chloride (NaCl)	Sigma-Aldrich	Cat#31434
Magnesium chloride (MgCl <sub>2</sub> )	Sigma-Aldrich	Cat#M2393
Calcium chloride (CaCl <sub>2</sub> )	Sigma-Aldrich	Cat#C3306
Zinc chloride (ZnCl <sub>2</sub> )	Sigma-Aldrich	Cat#793523
Glycerol	Thermo Fisher Scientific	Cat#10795711
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich	Cat#E9884
DL-Dithiothreitol (DTT)	Melford	Cat#D11000
Triton X-100	Sigma-Aldrich	Cat#T8787
Deoxyribonuclease I (DNase I)	Sigma-Aldrich	Cat#DN25
Pefabloc	Sigma-Aldrich	Cat#76307
Bovine serum albumin (BSA)	Sigma-Aldrich	Cat#A7030
Sorbitol	Melford	Cat#S23080
Uranyl acetate	TAAB Laboratories	Cat#U001
DEPC treated water	Severn	Cat#20-9000-01
<i>E. coli</i> RNAP	NEB	Cat#M0550S
SYPRO Orange	Thermo Fisher Scientific	Cat#S6650
ATP, CTP, GTP	Thermo Fisher Scientific	Cat#R0441, R0451, R0461
Urea	Melford	Cat#U20200
Disuccinimidyl sulfoxide (DSSO)	Sigma-Aldrich	Cat#909602
DMSO	Sigma-Aldrich	Cat#D8418
Sulfosuccinimidyl 4,4'-azipentanoate (sulfo-SDA)	Sigma-Aldrich	Cat#803340
Dihydrazide sulfoxide (DHSO)	Sigma-Aldrich	Cat#909599
4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM)	Sigma-Aldrich	Cat#74104
Iodoacetamide	Sigma-Aldrich	Cat#I1149
Trypsin/Lys-C mix	Promega	Cat#VA9000
Acetonitrile	Sigma-Aldrich	Cat#271004
Formic acid	Sigma-Aldrich	Cat#695076
Sodium deoxycholate	Sigma-Aldrich	Cat#D6750
EPPS	Sigma-Aldrich	Cat#E9502
Trifluoroacetic acid	Sigma-Aldrich	Cat#302031
<b>Deposited data</b>		
Atomic model of PEP	This paper	PDB: 8R5O
Atomic model of PEP (Integrated)	This paper	PDB: 8R6S
Atomic model of PEP-TEC	This paper	PDB: 8RAS
Atomic model of PEP-TEC (Integrated)	This paper	PDB: 8RDJ
Cryo-EM map of PEP Composite	This paper	EMDB: EMD-18935
Cryo-EM map of PEP Consensus	This paper	EMDB: EMD-18920

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cryo-EM map of PEP Region 1	This paper	EMDB: EMD-18964
Cryo-EM map of PEP Region 2	This paper	EMDB: EMD-18965
Cryo-EM map of PEP Region 3	This paper	EMDB: EMD-18974
Cryo-EM map of PEP Region 4	This paper	EMDB: EMD-18975
Cryo-EM map of PEP Region 5	This paper	EMDB: EMD-18976
Cryo-EM map of PEP Region 6	This paper	EMDB: EMD-18982
Cryo-EM map of PEP Region 7	This paper	EMDB: EMD-18998
Cryo-EM map of PEP Region 8	This paper	EMDB: EMD-18983
Cryo-EM map of PEP Region 9	This paper	EMDB: EMD-18985
Cryo-EM map of PEP-TEC Composite	This paper	EMDB: EMD-19023
Cryo-EM map of PEP-TEC Consensus	This paper	EMDB: EMD-18952
Cryo-EM map of PEP-TEC Region 1	This paper	EMDB: EMD-18986
Cryo-EM map of PEP-TEC Region 2	This paper	EMDB: EMD-18995
Cryo-EM map of PEP-TEC Region 3	This paper	EMDB: EMD-18996
Cryo-EM map of PEP-TEC Region 4	This paper	EMDB: EMD-19007
Cryo-EM map of PEP-TEC (PAP2-mRNA)	This paper	EMDB: EMD-19010
Mass spectrometry data	This paper	PRIDE: PXD045575

**Experimental models: Organisms/strains**

<i>Sinapis alba</i> variety Green Manure Mustard White	Moles Seeds	Cat#VGR28
--	-------------	-----------

**Oligonucleotides**

Sequence of tDNA (for PEP-TEC preparation): 5'-GGCTTTCGCTTTCGCGTCTCTCTAAAATT GCAGTCCC GCGCGCGTAGGACGACTGA CCTCCATTTTAGGAACCAATAA -3'	This paper	N/A
Sequence of ntDNA (for PEP-TEC preparation): 5'- /6-FAM/TTATTTGGTTCCTAAAATGGAGG TCAGTACGTCCTATCGATCTTCGGACTGCAA TTTTAGAGAGACGCGAAAGCGAAAGCC-3'	This paper	N/A
Sequence of mRNA (for PEP-TEC preparation): 5'-CCUGAUGAUUAAAUAACCAAGGAUU UUACCCGGCGCGCG-3'	This paper	N/A
Sequence of RNA (for RNA extension assay) 5'-Cy5/AUACCCGGCGCGCG-3'	This paper	N/A
Sequence of ntDNA (for RNA extension assay) 5'-AGGTCAGTACGTCCTACGCGCGCGA CCAAGACACTACTCACTTC-3'	This paper	N/A
Sequence of tDNA (for RNA extension assay) 5'-GAAGTGAGTAGTGTCTTGGTCGCGCGC CGTAGGACGACTGACCT-3'	This paper	N/A

**Software and algorithms**

EPU	Thermo Fisher Scientific	N/A
RELION v3.1.4	Zivanov et al. <sup>60</sup>	<a href="http://www3.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page">www3.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page</a>
MotionCor2	Zheng et al. <sup>61</sup>	<a href="http://emcore.ucsf.edu/ucsf-software">emcore.ucsf.edu/ucsf-software</a> ; RRID:SCR_016499
CTFFind v4.1	Rhou and Grigorieff <sup>62</sup>	<a href="http://grigoriefflab.umassmed.edu/ctffind4">grigoriefflab.umassmed.edu/ctffind4</a> ;
UCSF Chimera X v1.6.1	Meng et al. <sup>63</sup>	<a href="http://www.rbvi.ucsf.edu/chimerax/">www.rbvi.ucsf.edu/chimerax/</a> ; RRID:SCR_015872
CryoSPARC v4.4.0	Punjani et al. <sup>64</sup>	<a href="http://www.cryosparc.com">www.cryosparc.com</a> ; RRID:SCR_016501
3DFSC	Tan et al. <sup>65</sup>	<a href="https://github.com/LyumkisLab/3DFSC">github.com/LyumkisLab/3DFSC</a>
Phenix v1.20.1	Liebschner et al. <sup>66</sup>	<a href="http://www.phenix-online.org/">www.phenix-online.org/</a> ; RRID:SCR_014224
DALI	Holm et al. <sup>67</sup>	<a href="http://ekhidna2.biocenter.helsinki.fi/dali/">ekhidna2.biocenter.helsinki.fi/dali/</a> ; RRID:SCR_013433

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
COOT v0.9.8.1	Emsley and Cowtan <sup>68</sup>	<a href="http://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/">www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/</a> ; RRID:SCR_014222
AlphaFold2	Jumper et al. <sup>69</sup>	<a href="http://alphafold.ebi.ac.uk/">alphafold.ebi.ac.uk/</a> ; RRID:SCR_023662
ModelAngelo	Jamali et al. <sup>70</sup>	<a href="https://github.com/3dem/model-angelo">github.com/3dem/model-angelo</a>
DeepEMhancer	Sanchez-Garcia et al. <sup>71</sup>	<a href="https://github.com/rsanchezgarc/deepEMhancer">github.com/rsanchezgarc/deepEMhancer</a>
MaxQuant v2.4.2	Tyanova et al. <sup>72</sup>	<a href="http://www.maxquant.org/">www.maxquant.org/</a> ; RRID:SCR_014485
Mass Spec Studio v2.4.0.3545	Crowder et al. <sup>73</sup>	<a href="http://www.msstudio.ca/">www.msstudio.ca/</a>
pLINK v2.3.11	Chen et al. <sup>74</sup>	<a href="http://pfind.org/software/pLink/">pfind.org/software/pLink/</a> ; RRID:SCR_000084
Proteome Discoverer 3.0	Thermo Fisher Scientific	RRID:SCR_014477
CHIMERYS	MSAID	N/A
PPRCODE	Yan et al. <sup>33</sup>	N/A
MAAFT	Katoh et al. <sup>75</sup>	<a href="http://mafft.cbrc.jp">mafft.cbrc.jp</a>
JalView v2.11.3.2	Waterhouse et al. <sup>76</sup>	<a href="http://www.jalview.org/">www.jalview.org/</a> ; RRID:SCR_006459
Clustal Omega	Madeira et al. <sup>77</sup>	<a href="http://www.ebi.ac.uk/Tools/msa/clustalo/">www.ebi.ac.uk/Tools/msa/clustalo/</a> ; RRID:SCR_001591
SHOOT	Emms and Kelly <sup>78</sup>	<a href="http://shoot.bio/">shoot.bio/</a>
FIMO	Grant et al. <sup>79</sup>	<a href="http://meme-suite.org/meme/tools/fimo">meme-suite.org/meme/tools/fimo</a> ; RRID:SCR_001783
AquireMP and DiscoverMP software (version R1.2)	Refeyn	<a href="http://www.refeyn.com">www.refeyn.com</a>

**Other**

CB15V Waring blender	Nisbets	Cat#GF422
BioDesign cheesecloth	Thermo Fisher Scientific	Cat#12967487
Mira cloth	Sigma-Aldrich	Cat#475855
Fiberlite F20-12x50 LEX rotor	Thermo Fisher Scientific	Cat#096-124375
Fiberlite F9-6x1000 LEX rotor	Thermo Fisher Scientific	Cat#096-061075
PES syringe filter 0.45 μm	StarLab	Cat#E4780-1456
HiPrep Heparin FF 16/10 column	Cytiva	Cat#28936549
HiTrap Heparin HP 5 ml column	Cytiva	Cat#17040701
HiPrep Sephacryl S-300 (26/60) column	Cytiva	Cat#17119601
Capto HiRes Q 5/50 column	Cytiva	Cat#29275878
Superose 6 Increase 3.2/300 column	Cytiva	Cat#29091598
Centrifugal concentrator (Vivaspin 500 100,000 MWCO)	Sartorius	Cat#VS0141
NativeMark Unstained Protein Standard (Mass photometry)	Thermo Fisher Scientific	Cat#LC0725
EM grid (Carbon film, 400 mesh, Cu)	Electron Microscopy Sciences	Cat#CF400-CU-50
Quantifoil EM grid (UltrAufoil, R2/2, 200 mesh)	Electron Microscopy Sciences	Cat#Q250AR2A
Zeba spin desalting columns	Thermo Fisher Scientific	Cat#89882
Sep-Pak tC18 cartridge (50 mg)	Waters	Cat#WAT054960
C18-AQ HPLC resin	Dr Maisch	Cat#r119.aq

**RESOURCE AVAILABILITY**

**Lead Contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Michael W. Webster ([michael.webster@jic.ac.uk](mailto:michael.webster@jic.ac.uk)).

**Material Availability**

This study did not generate new unique reagents.

### Data and Code Availability

- Cryo-EM reconstructions and atomic coordinates of PEP and PEP-TEC complexes have been deposited in the Electron Microscopy Data Bank (<https://www.ebi.ac.uk/pdbe/emdb/>) and RCSB Protein Data Bank (<http://www.rcsb.org>) and are publicly available as of the date of publication. PDB accession numbers are 8R5O, 8R6S, 8RAS and 8RDJ for the PEP composite model, PEP integrated model, PEP-TEC composite model and PEP-TEC integrated models respectively. EMDB accession numbers are: EMD-18935 (PEP composite map), EMD-18920 (PEP consensus map), EMD-18964 (PEP focused region 1), EMD-18965 (PEP focused region 2), EMD-18974 (PEP focused region 3), EMD-18975 (PEP focused region 4), EMD-18976 (PEP focused region 5), EMD-18982 (PEP focused region 6), EMD-18998 (PEP focused region 7), EMD-18983 (PEP focused region 8), EMD-18985 (PEP focused region 9), EMD-19023 (PEP-TEC composite map), EMD-18952 (PEP-TEC consensus map), EMD-18986 (PEP-TEC focused region 1), EMD-18995 (PEP-TEC focused region 2), EMD-18996 (PEP-TEC focused region 3), EMD-19007 (PEP-TEC focused region 4), and EMD-19010 (PEP-TEC PAP2-mRNA focused region). Mass spectrometry data and CLMS search results have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD045575.
- 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

Proteins were purified from isolated chloroplasts of *Sinapis alba* cotyledons grown from seeds of Green Manure Mustard White (VGR28) purchased from Moles Seeds.

### METHOD DETAILS

#### Plant growth and chloroplast isolation

White mustard (*Sinapis alba*) was sown in John Innes F2 Starter media and aerial parts of 1-week old seedlings were harvested. Harvested material was placed on ice and all steps were performed at 4°C. Batches of ~400 g plant material were homogenised in 1.6 L of homogenisation buffer comprised of Tris-HCl (10 mM, pH 7.6 at 4°C), KCl (50 mM), MgCl<sub>2</sub> (5 mM), EDTA (2 mM), sorbitol (0.7 M), DTT (1 mM) using a Waring blender (CB15V model, power 4, 20 sec). Homogenate was filtered through BioDesign cheesecloth and three layers of miracloth. Chloroplasts were pelleted by centrifugation (4000 × g, 10 min, 4°C, rotor Fiberlite F9-6x1000 LEX) and washed in homogenization buffer supplemented with BSA (1 g/L). The resulting suspension was centrifuged (3000 × g, 10 min, 4°C, rotor S-4-104) and the resulting pellet was flash-frozen in liquid nitrogen and stored at -80°C.

#### Chloroplast lysate preparation

Chloroplast pellet from ~2 kg of leaf material was resuspended in 360 mL of lysis buffer comprised of Tris-HCl (50 mM, pH 7.6 at 4°C), NaCl (150 mM), MgCl<sub>2</sub> (2 mM), CaCl<sub>2</sub> (1 mM), ZnCl<sub>2</sub> (10 μM), glycerol (10%), DTT (2 mM), Triton X-100 (2%), DNase I (10 μg/mL) and Pefabloc (1 mg/mL). The crude extract was incubated for 30 min at 4°C with stirring to allow digestion of plastid DNA. This lysate was clarified by centrifugation (40,000 × g, 30 min, 4°C, rotor F20-12x50 LEX) and the supernatant filtered through a 0.45 μm PES syringe filter to remove cell debris.

#### Purification of PEP

Clarified chloroplast lysate was applied to a HiPrep Heparin FF 16/10 column that had been equilibrated in PEP Buffer, comprised of Tris (50 mM, pH 7.6 at 4°C), NaCl (150 mM), MgCl<sub>2</sub> (2 mM), ZnCl<sub>2</sub> (10 μM), glycerol (10%) and DTT (2 mM). After washing with 5 column volumes of PEP Buffer, the sample was eluted with PEP Buffer supplemented to 600 mM NaCl. Eluted fractions were collected and dialyzed into PEP buffer containing 150 mM NaCl for 2 hours at 4°C and loaded onto a HiTrap Heparin HP 5 mL column that had been equilibrated in PEP Buffer. Following washing with 3 column volumes of PEP Buffer, step elution was performed with PEP buffer supplemented to 600 mM NaCl. The resulting fractions with a total volume of ~5 mL were pooled and applied to a HiPrep Sephacryl S-300 (26/60) column that had been equilibrated in SEC buffer, comprised of PEP buffer supplemented to 200 mM and with glycerol concentration of 5% rather than 10%.

SEC fractions were resolved by SDS-PAGE and the presence of RPOB was assessed by immunodetection. The primary antibody (anti-RPOB, Agrisera #AS15 2867) was used at 1:1000 dilution, and the secondary antibody (Chicken anti-Rabbit IgG HRP Conjugate, Agrisera: AS10 839) was used at 1:20,000 dilution. The fractions that contained the most RPOB were pooled and applied to a Capto HiRes Q 5/50 column that had been equilibrated in SEC Buffer. The column was washed with 5 mL of SEC buffer and a gradient elution was performed with SEC buffer containing 200 to 1000 mM NaCl over 30 column volumes. Two prominent peaks in absorbance at 280 nm were observed, and immunodetection of RPOB indicated the first peak, which eluted at 300 to 400 mM NaCl, contained PEP. Fractions containing PEP were combined and diluted four-fold in SEC buffer containing 50 mM NaCl. The sample was re-applied to the Capto HiRes Q 5/50 that had been equilibrated in SEC buffer, and step elution with SEC buffer supplemented to 650 mM NaCl was performed to increase the sample concentration. PEP was further concentrated and buffer-exchanged

using a centrifugal concentrator (Vivaspin 500 100,000 MWCO) to ~2-3 mg/mL. Aliquots were flash-frozen in liquid nitrogen and stored at -80°C.

### Mass photometry

Mass photometry measurements were recorded using the Refeyn OneMP mass photometer at 25 °C. Instrument calibration in the molecular mass range 20-1200 kDa was performed using selected masses of NativeMark Unstained Protein Standard (146, 480 and 1048 kDa). 2 µL of purified PEP was applied to 8 µL PEP Buffer on a coverslip to a final concentration of 25 nM. Movies were recorded at 60 frames/sec for 60 sec using the large field of view with AcquireMP software version R1.2. The data were processed using DiscoverMP software version R1.2. Mass of PEP was estimated as the mode of the histogram distribution following fit of a Gaussian distribution.

### Negative stain electron microscopy

Negative stain electron microscopy was used to assess homogeneity and structural features of purified PEP. Grids coated with thin carbon were glow-discharged for 30 sec at 8 mA with a Leica ACE200 vacuum coater. 3.5 µL of purified PEP was diluted and applied to grids and incubated for 2 mins. Following blotting of excess solution, grids were stained with uranyl acetate solution (2% w/v) for 30 sec, before blotting again. Images were collected at room temperature on a Talos F200C transmission electron microscope operated at 200 keV using a Gatan OneView camera with settings: calibrated pixel size of 3.6 Å/pixel, defocus of -0.7 µm and dose of 25 electrons/Å<sup>2</sup>.

Data was processed using RELION 3.1.1.<sup>64</sup> Particles (239,165) were picked from 1992 micrographs and extracted with downsampling to 7.2 Å/pixel with a box size of 64 pixels. Three sequentially rounds of two-dimensional classification were performed with a mask diameter of 320 Å. A further round of two-dimensional classification on particles re-extracted without downsampling produced a set of 103,775 particles. *Ab initio* model generation and a single round of three-dimensional classification were performed. The final reconstruction was produced from the 40,545 particles assigned to the class with the best features.

### Cryo-EM sample preparation and data acquisition

Purified PEP was applied to a Superose 6 Increase 3.2/300 equilibrated in PEP buffer containing 50 mM NaCl and without glycerol. Sample was eluted at a flowrate of 0.01 mL/min, and peak fractions were pooled and measured to have a concentration of 4 mg/mL.

To produce PEP-TEC, oligonucleotides for the template DNA (tDNA), non-template DNA (ntDNA) and RNA were chemically synthesised with HPLC purification (Integrated DNA Technologies) and resuspended in RNase-free water. Sequences of oligonucleotides are shown in [key resources table](#). To prepare the nucleic acid scaffold, tDNA and RNA were mixed in a 1:1 molar ratio at a final concentration of 24 µM in buffer comprising Tris (10 mM, pH 7.6 at 4°C), KOAc (40 mM), Mg(OAc)<sub>2</sub> (5 mM). Annealing was performed by heating to 95 °C followed by stepwise cooling to 10°C in increments of 5 °C for 5 mins each. Purified PEP was buffer-exchanged using a centrifugal concentrator (Vivaspin 500 100 MWCO) into PEP buffer containing 50 mM NaCl and no glycerol. Purified PEP (1.2 µM final concentration) and tDNA-RNA scaffold (6 µM final concentration) were mixed in PEP Buffer containing 50 mM NaCl and no glycerol and incubated for 30 min at 37 °C. 5' fluorescein-labelled ntDNA (6 µM final concentration) was added and incubated for 30 min at 37 °C. Complex formation was assessed by electrophoretic mobility shift assay with a 6% acrylamide gel, visualised by scanning for fluorescein signal on a Typhoon FLA-9000 imager.

UltrAufoil R2/2 200 mesh grids were glow-discharged for 45 sec at high power settings using a Harrick Plasma Cleaner before sample was applied to the grids. Once glow discharged, cryo-EM grids were prepared using a Vitrobot Mark IV plunger at 4°C and 100% humidity. 3.5 µL of purified protein sample (approximately 1 µM for PEP and 1.2 µM for PEP-TEC) was applied to the glow-discharged grid, followed by 2 sec of blotting using a blot force of 2.

Cryo-EM data collection was carried out on a Titan Krios transmission electron microscope (FEI) operated at 300 keV with a BioQuantum energy filter (slit width 20 eV) and K3 summit direct electron detector (Gatan). Movies were collected using EPU software (see [Table S3](#)). For PEP, 21,347 movies were collected with a defocus range of -0.7 to -2.5 µm, pixel size 0.831 Å and total dose of 50 e<sup>-</sup>/Å<sup>2</sup>. For PEP-TEC, 33,330 movies were collected with a defocus range of -0.8 to -2.0 µm, pixel size 0.831 Å and total dose of 40 e<sup>-</sup>/Å<sup>2</sup>.

### Cryo-EM data processing

#### Consensus reconstruction of PEP

Cryo-EM data processing was performed using RELION 3.1.4.<sup>60</sup> Image frames were aligned and averaged with MotionCor2<sup>61</sup> implemented in RELION and contrast transfer function (CTF) parameters were calculated with CTFFIND4.<sup>162</sup> Micrographs with CTF maximum resolution estimate ≤ 8 Å were retained, resulting in 17,249 micrographs. Templates for particle selection were generated by 2D classification of approximately 6000 particles that were manually selected. Auto-picked particle images (1,888,376) were extracted with four-fold downsampled pixel size of 3.324 Å and a box size of 150 pixels. Iterative rounds of 2D classification with mask diameter of 380 Å were performed to remove low-quality particle images, resulting in a set of 1,678,594 particle images. An initial model was generated using RELION *ab initio* reconstruction. 3D refinement was performed with a soft mask, followed by three rounds of 3D classification with a mask diameter of 428 Å, a single round of 2D classification, and a further round of 3D classification without alignment, producing a set of 613,537 particle images. Particle images were re-extracted with a two-fold downsampled pixel

size of 1.662 Å and a box size of 300 pixels. Following 3D refinement, refinement of CTF parameters was performed to correct higher-order aberrations, anisotropic magnification, per-particle defocus values, and beam tilt correction. Particles were re-extracted with full pixel size of 0.831 Å and a box size of 600 pixels. Following 3D refinement, refinement of CTF parameters was repeated and Bayesian polishing performed. Polished particles were refined and postprocessed to produce the consensus reconstruction of PEP at 2.49 Å resolution.

### Composite reconstruction of PEP

The consensus reconstruction revealed substantial structural variation across the dataset that limited the interpretability of peripheral regions of PEP. The reconstruction was divided into nine overlapping regions and focused 3D classification and focused refinement was performed with a mask around each region. The masked regions were defined by inspection of a preliminary structural model of PEP to include segments predicted to be internally rigid but may move with a degree of independence from the remainder of the complex. The protein regions in the final model corresponding to each masked area are indicated in Table S4. Masks were created by erasing density from a consensus reconstruction map low pass filtered to 6 Å resolution. Following 3D classification without particle alignment using each mask, particles assigned to well defined classes were combined and 3D refinement performed with the same mask. The nine focused maps obtained displayed improved features within the masked area and reduced density outside, consistent with improved alignment on the selected area. Each focused reconstruction was resampled to a pixel size of 0.5 Å, aligned to the consensus reconstruction and combined to create a composite map using UCSF ChimeraX<sup>63</sup> 'volume maximum' command.

### Reconstruction of PEP-TEC

Micrograph pre-processing was performed as for PEP using RELION 3.1.4, resulting in 29,783 micrographs from which 7,035,671 particles were auto-picked. Particles extracted with a box size of 150 pixels and four-fold downsampled pixel size of 3.324 Å were subjected to iterative rounds of 2D classification, from which 2,121,579 particles were retained. 3D refinement was performed, followed by two rounds of 3D classification without alignment with mask diameters of 270 to 320 Å. The well resolved classes contained 1,941,179 particles, from which a reconstruction was obtained.

Inspection of the reconstruction indicated nucleic acids were resolved less clearly than the protein subunits, indicating heterogeneity in the dataset that limited interpretation. To obtain an improved reconstruction, refinement was continued from an intermediate iteration with a soft mask around density arising from the nucleic acid scaffold to selectively align particles based on this density. A single round of 3D classification without alignment was then performed with the mask including the whole nucleic acid scaffold, and particles in the well resolved class (1,920,744 particles) were then subjected to three further rounds of 3D classification without alignment using masks focused on the downstream DNA density in the reconstruction. This yielded 465,084 particles assigned to a class with well resolved nucleic acids.

The resulting particle set was re-extracted in cryoSPARC v4.2.1<sup>64</sup> with a box size of 600 pixels and full pixel size of 0.831 Å. A single round of 2D classification was performed, and 417,374 particles were selected. Non-uniform 3D refinement (including fitting spherical aberration, tetrafoil and anisotropic magnification) was performed, followed by local CTF refinement. A final non-uniform 3D refinement produced the consensus PEP-TEC reconstruction at 2.62 Å resolution. The composite PEP-TEC map was obtained as for PEP by combining reconstructions focused on three regions (upstream DNA, active site, and downstream DNA) and the consensus reconstruction (Table S4).

Local resolution and 3D FSC information was estimated for each reconstruction using Relion, CryoSPARC and 3DFSC.<sup>65</sup>

### Model building

The structural model of PEP was built in a series of steps: (1) construction of an initial model refined within the consensus cryo-EM reconstruction, (2) construction of a 'composite model' refined within the composite map of higher resolution, and (3) construction of an 'integrated model' in which models for regions less resolved in the cryo-EM density were added.

#### Initial model construction

Density at the centre of the reconstruction showed resemblance to the bacterial RNAP 'crab claw' architecture. We therefore first fitted a structure of the *E. coli* RNA polymerase (PDB: 6ZTM)<sup>80</sup> to this region and performed refined in Phenix (v1.20.1)<sup>66</sup> with rigid-body restraints for the two lobes ( $\beta$  and  $\beta'$ ). A structural homology search with the resulting model was performed with Dali,<sup>67</sup> revealing that among deposited structures the closest structural matches for the overall conformation of the PEP polymerase in the absence of nucleic acids were bacterial transcription initiation complexes, such as the *Mycobacterium tuberculosis* RNAP holoenzyme (PDB: 5ZX3).<sup>81</sup> This model was fitted in the cryo-EM density using Coot (v0.9.8.1).<sup>68</sup> Structural predictions were generated for each *S. alba* PEP core polymerase subunit ( $\alpha$ ,  $\beta$ ,  $\beta'$ ,  $\beta''$ ) using AlphaFold.<sup>69</sup> These models were separated into their modular domains and individually aligned to the reference structure (PDB: 5ZX3). Further adjustments were made with Coot and real-space refinement was performed with Phenix. Initial models of each *S. alba* PAP subunit were generated with AlphaFold and fitted approximately within the density by structural alignment to a model of the entire PEP complex built using ModelAngelo.<sup>70</sup> Iterations of manual adjustment in Coot, *de novo* model building, and real-space refinement in Phenix were performed with the consensus EM map.

#### Composite model construction

The initial PEP model built using the consensus reconstruction was separated into nine models corresponding to the nine focused maps, where each region was assigned to the map that had the best signal in that area (Table S4). Each model was trimmed so that only residues with clear signal at the resolution of the focused maps were retained. These models were independently improved by

iterations of manual adjustment in Coot, and *de novo* model building and real-space refinement in Phenix. Water molecules were placed in density using Phenix Douse followed by selection in Coot for density stronger than 12 rmsd. The nine models were combined into a single model of the well-resolved subset of PEP by rigid-body fitting each into the composite cryo-EM map. Connections between chains that were split between modules were manually adjusted in Coot, and real-space refinement in Phenix was then performed using the rigid-body definitions of each of the nine models with minimization.

### **Integrated model construction**

Parts of the structure that were not sufficiently ordered to be modelled at high resolution were excluded from the composite model. To aid interpretation of regions resolved to lower resolution in the cryo-EM reconstructions, we generated models based on structural predictions from AlphaFold and maps that were low-pass filtered or processed with DeepEMhancer.<sup>71</sup> The models for these regions, which we term the ‘model extensions’, should be considered as representative of an ensemble of positions present in the conformationally heterogeneous dataset. To generate the ‘Integrated model’ of PEP, the model extensions were combined with the model built from the high-resolution composite map. Connections between chains were manually adjusted in Coot, followed by real-space refinement in Phenix with the consensus map.

### **PEP-TEC model construction**

To the PEP model, additional sequences of the  $\beta$  and  $\beta'$  subunits that were resolved in the PEP-TEC reconstruction but not PEP were modelled *de novo* (Table S5). A model of the PAP1 SAP domain was generated from the AlphaFold structural prediction of PAP1. Initial models for the nucleic acid chains were generated by modification of those of an *E. coli* TEC (PDB: 6ALH).<sup>82</sup> Iterations of manual adjustment in Coot, *de novo* model building and real-space refinement in Phenix were performed with the PEP-TEC composite EM map.

### **Amino acid sequences for structural models**

Sequences of plastid-encoded subunits were obtained from translations of the chloroplast reference genome.<sup>83</sup> Sequences of nuclear-encoded subunits were obtained from Brassicales Map Alignment Project (BMAP, DOE-JGI) (Table S1). Four corrections in amino acid sequence were made based on unambiguous density in the cryo-EM reconstructions:  $\alpha$  subunit, substitution at residue 67 to F;  $\beta$  subunit, substitution at residue 113 to F and 657 to V; PAP11 substitution at residue 544 to Y.

### **Structure analysis and visualization**

Analysis of structural models was performed with Coot<sup>68</sup> and ChimeraX,<sup>63</sup> and figures were generated with ChimeraX.

### **RNA extension assay**

The transcription activity of purified PEP was assessed by visualising extension of Cy5-labelled RNA. A nucleic acid scaffold was reconstituted *in vitro* using synthetic oligonucleotides (sequences in key resources table) by mixing tDNA (100  $\mu$ M) and RNA (50  $\mu$ M) in DEPC-treated water, incubating for 2 min at 98 °C and then cooling to 10 °C at a rate of 1 °C/min. Scaffold (final concentration of RNA 500 nM) was mixed with either purified PEP (1  $\mu$ M) or *E. coli* RNAP (0.46 U/ $\mu$ L) in reconstitution buffer (10 mM Tris-HCl, pH 8.0, 40 mM KCl, 5 mM MgCl<sub>2</sub>) supplemented with 0.02 mg/mL acetylated BSA. The amount of *E. coli* RNAP added was chosen to approximate an equimolar amount as PEP when assessed by SDS-PAGE stained with SYPRO Orange. Following incubation of PEP with tDNA-RNA scaffold at 37 °C for 20 min, ntDNA was added to a final concentration of 1000 nM and the sample incubated at 37 °C for 20 min. Transcription elongation was started by addition of a mixture of ATP, CTP and GTP (0.1 mM each). Reactions were incubated for 5 min at 30 °C and quenched by addition of an equal volume of 2 $\times$  stop buffer (8 M urea, 20 mM EDTA pH 8.0, 5 mM Tris-HCl pH 7.5). Samples were applied to TBE (Tris-borate-EDTA)-polyacrylamide gels containing 7 M urea (20% 19:1 acrylamide:bisacrylamide) and electrophoresis performed at 300 V for 75 min. Gels were scanned with a Typhoon FLA-9000 imager for Cy5 fluorescence.

### **Cross-linking mass spectrometry**

#### **Protein cross-linking, digestion and MS sample preparation**

Purified PEP was buffer-exchanged into an amine-free buffer (50 mM HEPES, 100 mM NaCl, 2 mM MgCl<sub>2</sub>, 10  $\mu$ M ZnCl<sub>2</sub>, 2 mM DTT, pH 7.5) using Zeba Spin desalting columns (7K MWCO). Crosslinking agents disuccinimidyl sulfoxide (DSSO) was prepared in anhydrous DMSO at 10 mM, and sulfosuccinimidyl 4,4'-azipentanoate (sulfo-SDA), dihydrazide sulfoxide (DHSO) and 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) were prepared in water at 10 mM, 25 mM and 25 mM, respectively. For the cross-linking reaction, sample (~10  $\mu$ g at ~0.5 mg/mL) was incubated at 30 °C with either DSSO (1 mM) for 45 min, sulfo-SDA (1 mM) for 45 min, or DHSO and DMTMM together (2.5 mM each) for 2 hours. The DSSO and sulfo-SDA reactions were quenched by addition of ammonium bicarbonate to a final concentration of 0.1 M, with incubation at 30 °C for 20 min. For the sulfo-SDA reaction, excess cross-linking agent was removed using a Zeba Spin desalting columns before exposure to 365 nm UV for 10 min using a Penn OC Photoreactor M1 (Penn Optical Coatings). For the DHSO/DMTMM reactions, which cannot be quenched, cross-linking agent was removed using Zeba Spin desalting columns. All samples were then snap-frozen in liquid nitrogen and freeze-dried.

Dried, crosslinked samples were resuspended in 20  $\mu$ L of 8 M urea, reduced by incubation with 5 mM DTT at 37 °C for 30 min, and alkylated by incubation with 10 mM iodoacetamide at room temperature for 20 min in darkness. Samples were diluted to 6 M urea by addition of Tris (25 mM, pH 8.0), and were treated with Trypsin/Lys-C mix at an enzyme to substrate ratio of 1:25 (w/w) at 37 °C for 4 hours. Samples were then diluted to 0.75 M urea by addition of Tris (25 mM, pH 8.0) and incubated at 37 °C overnight (~16 hours). Samples were then acidified by addition of 2% (v/v) formic acid and centrifuged at 16,000  $\times$  g for 10 min. The supernatant was

desalted using Sep-Pak tC18 cartridges, eluted in 60:40:0.1 acetonitrile:water:formic acid (v/v/v), flash frozen in liquid nitrogen and freeze-dried.

### Mass spectrometry

Dried peptides were resuspended in a solution of 4% (v/v) acetonitrile and 0.1% (v/v) formic acid and loaded onto a HPLC column containing 1.9  $\mu\text{m}$  C18-AQ (30 cm  $\times$  75  $\mu\text{m}$  inner diameter column packed in-house) using a Dionex UltiMate 3000 UHPLC. Peptides were separated in buffer containing 0.1% (v/v) formic acid and a linear gradient of 8–40% (v/v) acetonitrile over 107 min at 300 nL/min at 55 °C. Mass analyses were performed using a Q-Exactive HF-X mass spectrometer. Following each full-scan MS1 at 120,000 resolution at 200  $m/z$  (350 – 1600  $m/z$ , AGC =  $3 \times 10^6$ , 100 ms max injection time), up to 12 most abundant precursor ions were selected MS2 in a data-dependent manner. Settings used were: HCD,  $R = 30,000$ , AGC =  $2 \times 10^5$ , stepped NCE = (sulfo-SDA: 24, 28, 32) or (DSSO/DHSO/DMTMM: 21, 27, 33), 54 ms max injection time, 1.4  $m/z$  isolation window, minimum charge state of +3; dynamic exclusion of 20 s.

### Identification of cross-linked peptides

All data was processed using MaxQuant (v2.4.2)<sup>72</sup> and the sulfo-SDA and DMTMM datasets were additionally processed with Mass Spec Studio (v2.4.0.3545)<sup>73</sup> or pLINK (v2.3.11),<sup>74</sup> respectively. The following key parameters were generally used: peptide mass between 300–10,000 Da, minimum peptide length of 6 residues, precursor mass tolerance  $\pm 10$  ppm, product-ion mass tolerance of  $\pm 20$  ppm. Allowable variable modification = oxidation (M), allowable static modification = carbamidomethyl (C), enzyme specificity of Trypsin with up to four missed cleavages (excluding the site of cross-linking), and FDR control at 1%.

The search database contained the PEP sequences identified by mass spectrometry (Table S1). Cross-linked spectral matches from MS Studio were also manually visually assessed. All mass spectrometry data and search results have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository<sup>84</sup> with the dataset identifier PXD045575. This analysis identified 13 unique crosslinks for DMTMM (9/12 crosslinks involving modelled residues within the distance upper bound of 25 Å), no crosslinks for DHSO, 43 unique crosslinks for DSSO (39/41 crosslinks involving modelled residues within the distance upper bound of 30 Å) and 36 unique crosslinks for SDA (26/36 crosslinks involving modelled residues within the distance upper bound of 20 Å). As DMTMM and DHSO yielded few or no crosslinks, we used only data derived from crosslinking using DSSO and SDA for assessment of the structural model. Cross-linking peptides identified are listed in Table S6.

### Liquid chromatography mass spectrometry

Protein sample (2  $\mu\text{L}$ , ~40  $\mu\text{g}$  protein) was added to 50  $\mu\text{L}$  of 1.5% sodium deoxycholate (SDC) in EPPS buffer (0.2 M, pH 8.5) and vortexed under heating. Cysteine residues were reduced with DTT, alkylated with iodoacetamide, and proteins digested with trypsin in SDC buffer according to standard procedures. After the digest, the SDC was precipitated by adjusting to 0.2% trifluoroacetic acid (TFA), and the clear supernatant subjected to C18 HPLC. Aliquots were analysed by nanoLC-MS/MS on an Orbitrap Eclipse Tribrid mass spectrometer with a FAIMS Pro Duo source, coupled to an UltiMate 3000 RSLCnano LC system (Thermo Fisher Scientific). The samples were loaded onto a trap cartridge (PepMap Neo Trap Cartridge, C18, 5 $\mu\text{m}$ , 0.3 $\times$ 5mm, Thermo Fisher Scientific) with 0.1% TFA at 15  $\mu\text{L}/\text{min}$  for 3 min. The trap column was then switched in-line with the analytical column (nanoEase M/Z column, HSS C18 T3, 1.8  $\mu\text{m}$ , 100 Å, 250 mm  $\times$  0.75  $\mu\text{m}$ , Waters) for separation using the following gradient of solvents A (water, 0.1% formic acid) and B (80% acetonitrile, 0.1% formic acid) at a flow rate of 0.2  $\mu\text{L}/\text{min}$ : 0–3 min 3% B (parallel to trapping); 3–10 min linear increase B to 7%; 10–100 min increase B to 32%; 100–148 min increase B to 50%; followed by a ramp to 99% B and re-equilibration to 3% B. Mass spectrometry data were acquired with the FAIMS device set to three compensation voltages (-35V, -50V, -65V) at standard resolution for 1 sec each with the following MS settings in positive ion mode: resolution 120K, profile mode, mass range  $m/z$  300–1800, AGC target  $4e^5$ , max inject time 50 ms; MS2: quadrupole isolation window 1 Da, charge states 2–5, threshold  $1e^4$ , HCD CE = 30, AGC target standard, max. injection time dynamic, dynamic exclusion 1 count for 15 s with mass tolerance of  $\pm 10$  ppm.

The mass spectrometry raw data were processed and quantified in Proteome Discoverer 3.0 (PD) using the search engine CHIMERYS; all tools of the following workflow are nodes of the PD software. Protein database (52,754 entries) was imported into PD adding a reversed sequence database for decoy searches; a database for common contaminants ([maxquant.org](http://maxquant.org), 245 entries) was also included. The database search was performed using the incorporated CHIMERYS with the inferys\_2.1\_fragmentation prediction model, a fragment tolerance of 0.5 Da, enzyme trypsin with 2 missed cleavages, variable modification oxidation (M), fixed modification carbamidomethyl (C). The workflow included the Minora Feature Detector with min. trace length 7, S/N 3, PSM confidence high; the Top N Peak Filter with 10 peaks per 100 Da; Percolator with FDR targets 0.01 (strict) and 0.05 (relaxed). The CHIMERYS search used the inferys\_2.1\_fragmentation prediction model, a fragment tolerance of 0.5 Da, enzyme trypsin with 1 missed cleavage, variable modification oxidation (M), fixed modification carbamidomethyl (C).

The consensus workflow in the PD software was used to evaluate the peptide identifications and to measure the abundances of the peptides based on the LC-peak intensities. For identification, an FDR of 0.01 was used as strict threshold, and 0.05 as relaxed threshold. For protein abundances the average of the top 3 most abundant unique peptides was used. The results were exported into a Microsoft Excel table including data for protein abundances, number of peptides, protein coverage, q-values and PEP-values from Percolator, and the CHIMERYS identification score.

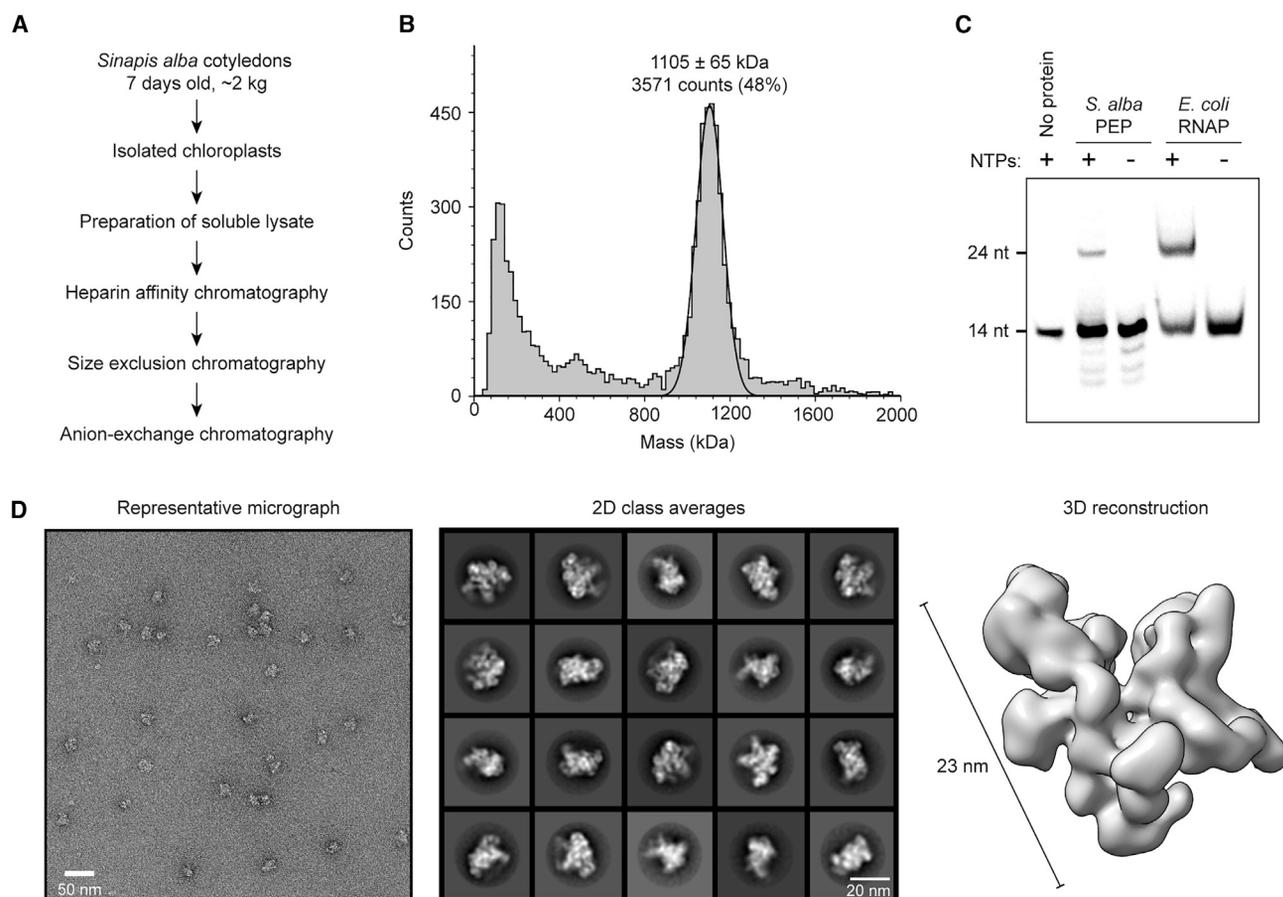
### Protein conservation and sequence features

Sequence alignments were performed with MAFFT<sup>75</sup> and visualized with JalView (v2.11.3.2).<sup>76</sup> Sequence percentage identity was measured using Clustal Omega.<sup>77</sup> Sequences of PAP orthologs in plants and  $\omega$  orthologs in bacteria were obtained using SHOOT.<sup>78</sup>

For analysis of PAP12- $\omega$  homology by multiple sequence alignment, PAP12 and  $\omega$  sequences were aligned as a single set. The positions and predicted RNA nucleotide binding specificity of the PPR proteins PAP1 and PAP2 were analysed by PPRCODE using the PS\_Scan detection algorithm.<sup>33</sup> Locations of PAP2 target sites within the chloroplast genome were identified using FIMO.<sup>79</sup>

Sequences of PEP polymerase subunit  $\alpha$ , orthologs were obtained by BLAST searches of *Arabidopsis thaliana* sequences with a defined set of species. The selected species maximise diversity and are a subset of that used previously to assess chloroplast phylogeny<sup>85</sup>: *Psilotum nudum*, *Angiopteris evecta*, *Equisetum hyemale*, *Adiantum capillus-veneris*, *Pinus thunbergii*, *Cycas revoluta*, *Arabidopsis thaliana*, *Nymphaea alba*, *Huperzia lucidula*, *Isoetes flaccida*, *Selaginella moellendorffii*, *Anthoceros angustus*, *Apopellia endiviifolia*, *Ptilidium pulcherrimum* and *Physcomitrium patens*. Sequences of bacterial RNAP subunit orthologs ( $\alpha$ ,  $\beta$ ,  $\beta'$ ,  $\beta''$ ) were obtained by BLAST searches of *Escherichia coli* sequences with a defined set of species to obtain, a single representative from each of phyla: *Acidobacteria*, *Actinobacteria*, *Aquificae*, *Armatimonadetes*, *Bacteroidetes*, *Caldiserica*, *Chlamydiae*, *Chlorobi*, *Chloroflexi*, *Chrysiogenetes*, *Coprothermobacterota*, *Cyanobacteria*, *Deferribacteres*, *Deinococcus-Thermus*, *Dictyoglomi*, *Elusimicrobia*, *Fibrobacteres*, *Firmicutes*, *Fusobacteria*, *Gemmatimonadetes*, *Lentisphaerae*, *Nitrospirae*, *Planctomycetes*, *Proteobacteria*, *Spirochaetes*, *Synergistetes*, *Thermodesulfobacteria*, *Thermotogae* and *Verrucomicrobia*.

## Supplemental figures



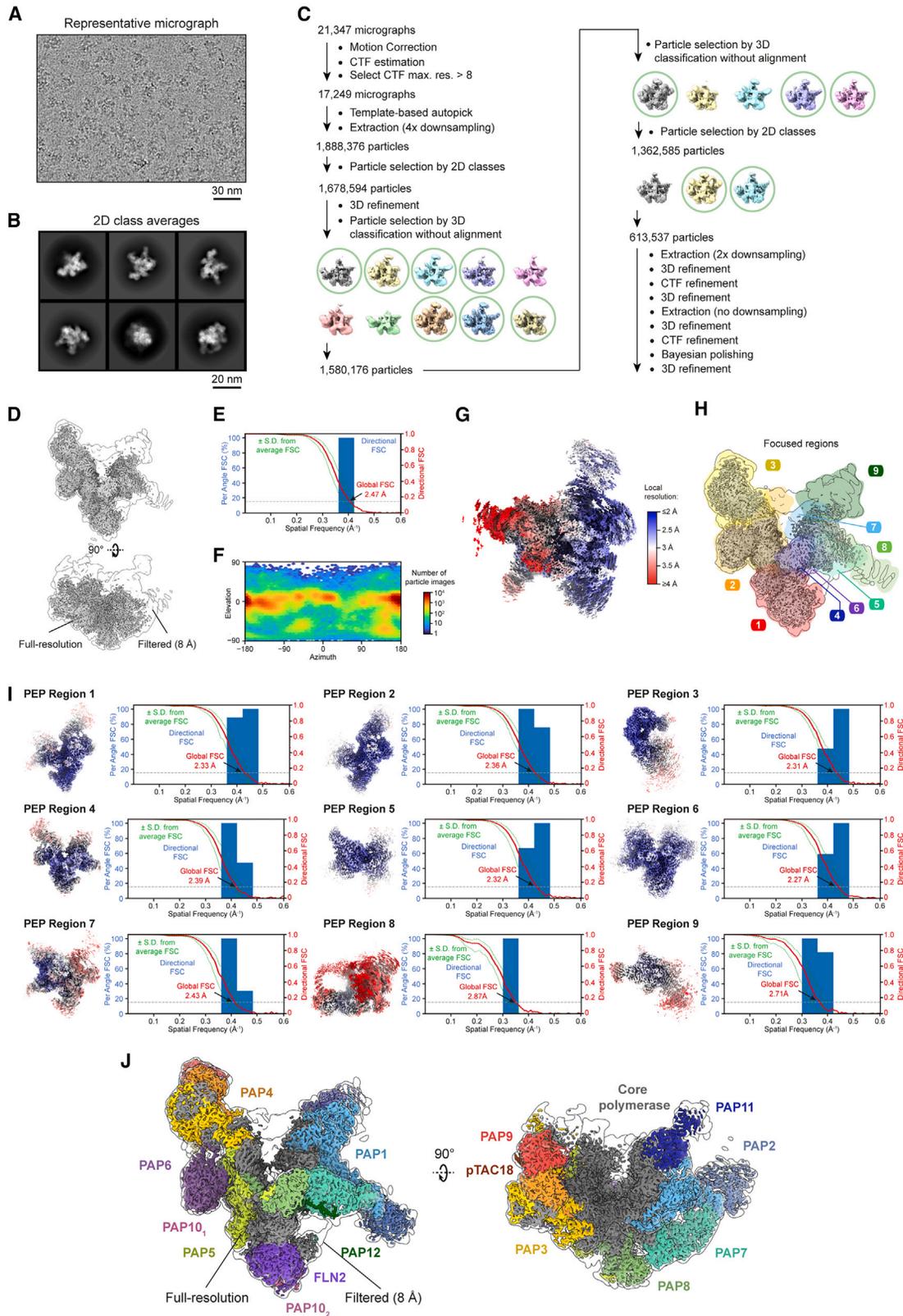
**Figure S1. Characterization of PEP composition and mass, related to Figure 1**

(A) Purification scheme for native PEP from chloroplast lysate of *S. alba* cotyledons.

(B) Analysis of PEP by mass photometry shows a molecular mass estimate of 1105 kDa ( $\sigma = 65$ ), similar to the theoretic molecular mass of 1152 kDa (Table S2).

(C) RNA extension assays with purified PEP and *E. coli* RNAP. An RNA oligomer 14 nt in length in a nucleic acid scaffold loaded with PEP was extended by the addition of ATP, CTP and GTP. The detected RNA product, 24 nt in length, was produced due to the templated requirement for UTP.

(D) Analysis of purified PEP by single-particle negative-stain electron microscopy. Scale bars in representative micrograph is 50 nm, scale bars in 2D class averages is 20 nm, and the longest dimension in the 3D reconstruction is 23 nm as indicated.

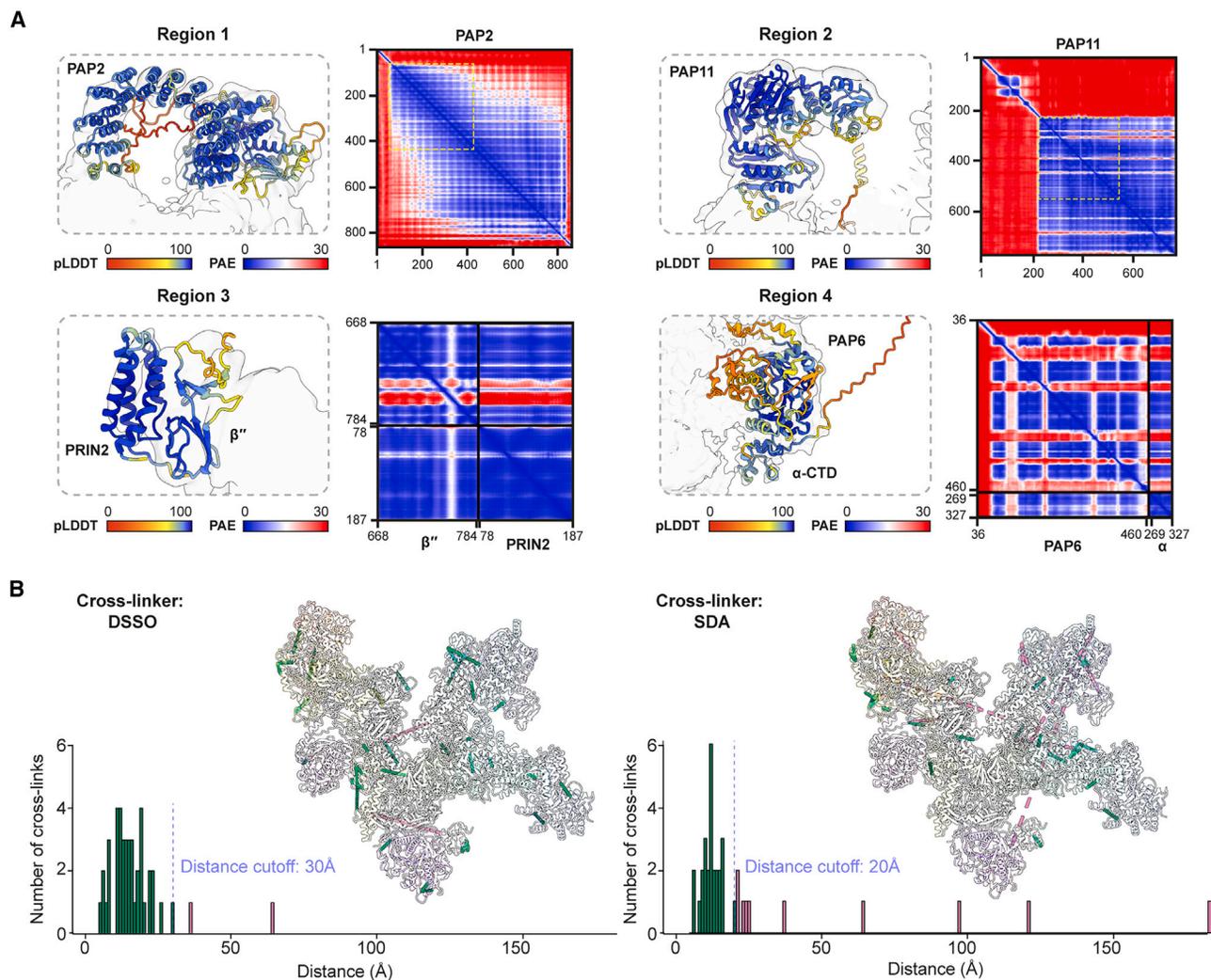


(legend on next page)

---

**Figure S2. Analysis of PEP by cryo-EM, related to Figure 1**

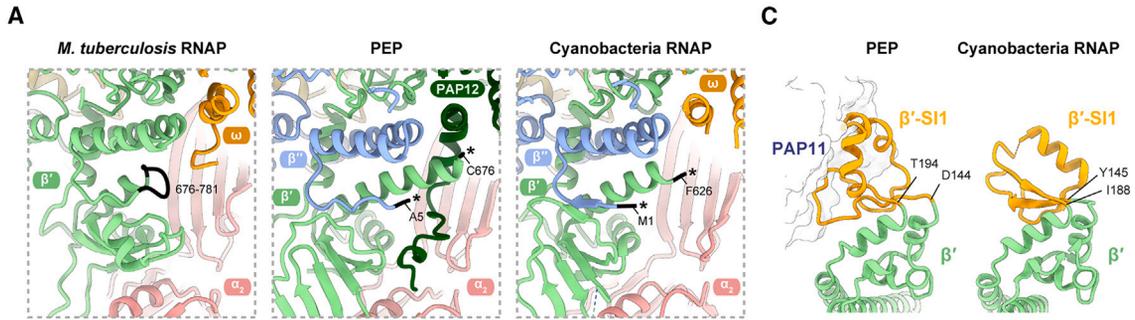
- (A) Representative cryo-EM micrograph of PEP. Scale bars represent 30 nm.
- (B) Selected cryo-EM two-dimensional class averages from the final particle set. Scale bars represent 20 nm.
- (C) Cryo-EM data processing workflow for analysis of PEP. Classes produced by 3D classification are shown, with particles within the indicated classes (green circles) selected for further processing.
- (D) Consensus cryo-EM reconstruction of PEP shown at full resolution (opaque) and low-pass filtered to 8 Å (transparent).
- (E) Fourier shell correlation plot for consensus PEP reconstruction. Dotted lines indicate the 0.143 threshold for gold-standard half map correlation.
- (F) Angular distribution plot for consensus PEP reconstruction.
- (G) Local resolution estimate for consensus PEP reconstruction.
- (H) Regions masked during focused classification and refinement procedures indicated 1 to 9 on consensus cryo-EM reconstruction of PEP.
- (I) Fourier shell correlation plots and local resolution estimates for focused reconstructions of PEP. Dotted lines indicate the 0.143 threshold for gold-standard half map correlation. Local resolution color range as in panel F.
- (J) Composite cryo-EM reconstruction generated from focused reconstructions, shown at full resolution (opaque and colored by assigned subunit), and low-pass filtered to 8 Å showing additional regions not resolved to high resolution (transparent)



**Figure S3. Construction of PEP structural model, related to Figure 1**

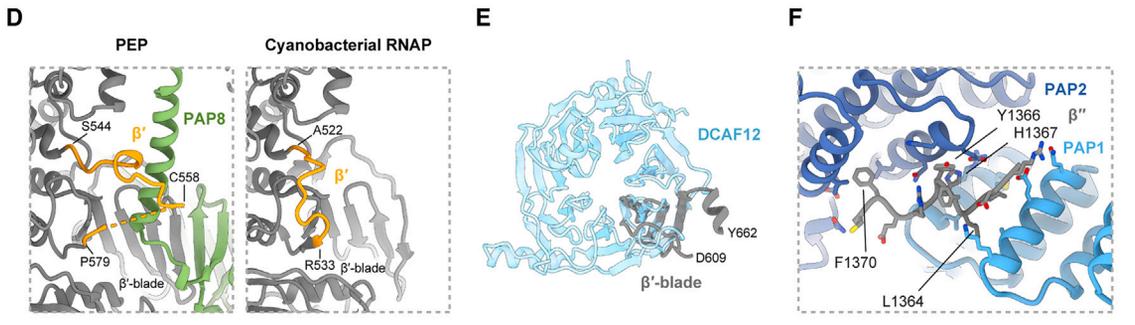
(A) Structural predictions generated with AlphaFold used to interpret regions of PEP that were incompletely resolved in cryo-EM reconstructions. Structural predictions are shown for four representative regions: region 1, from which the PAP2-PPR<sub>N</sub> region (residues 63–419) was used; region 2, from which the PAP11 N-terminal and central domains (residues 230–572) were used; region 3, complex of β''-SI3 head domain and PAP15/PRIN2, which was used in entirety; region 4, complex of PAP6 and α-CTD, from which the α-CTD (residues 269–327) was used. Models for regions 1–3 are shown in PEP cryo-EM reconstruction filtered to 8 Å, and model for region 4 shown within PEP-TEC cryo-EM reconstruction filtered to 5 Å. AlphaFold models are colored by per-residue confidence score (pLDDT). Predicted aligned error (PAE) plots for each region are shown with sections incorporated into the integrated PEP models highlighted (yellow dashed lines). For all predicted models, unstructured loops were removed from the models.

(B) Cross-linking mass spectrometry analysis of purified PEP. Crosslinks mapped onto PEP model following treatment with either DSSO (left) or SDA (right) are colored by those within the distance threshold (green), or not within (pink). Histograms of C $\alpha$ -C $\alpha$  distances between cross-linked residues were determined using structural model of PEP. The cross-linking distance thresholds indicated are 30 Å for DSSO and 20 Å for sulfo-SDA.



**B**

	NTD	Linker	α-LH	CTD
	231	239	247	271
<b>Plants</b>	<i>Sinapis alba</i> T F I I W N S G L T K E A L C H E R S R N I L F I P L H . . . . . V E E E F Y . . . . . L E N D H Q V T L P L P F F I N K L V N L R Q K K . . . . . K E L A F D Y I F D O L P P I Y N C K K S N E H T L L D L L			
<b>Bacteria</b>	<i>Acetobacter</i> V F I E T D S V P E D T V A L A R I Q D Q L S V F V N . . . . . F E E P K E . . . . . A T F E R Q P E L L . . . . . A F N A A L L K V D E L L E V S A N C K N D N V V I O B I L			



**G**

	Sequence
<b>PAP12</b>	<i>Sinapis alba</i> I K E V G K . V M T M D I E R L L L S E D N R F E P N S V A A E T E V Y E K N R D E Y G G T K . . . . . A I F H L S N R V N D . L E D F R P E A Y A E S D P Y K P G P G Y L K E Y Y T . . . . .
<b>ω</b>	<i>M. tuberculosis</i> M A R V T V Q D A V E K I G N R D L V A R R R R Q I Q S G K D A L V P E E N . . . . . D K V T V I A L R E I E . E E L I N Q I L D V R E R E Q E Q E . . . . . Q A A E I Q A V T A I E G R R . . . . .

(legend on next page)

**Figure S4. Comparison of core PEP polymerase and bacterial RNAP, related to Figures 2 and 3**

(A) Structural consequences of RPOC1-RPOC2 gene split to the structure of PEP. Comparison of structures of *Mycobacterium tuberculosis* RNAP (left), PEP (middle), and *Synechocystis* RNAP (right) at the location of the  $\beta'$  split position. Residues at the position of the split (PEP and *Synechocystis*) and the equivalent position of *Mycobacterium*, which lacks a split, are shown in black with asterisks.

(B) Sequence comparison of the  $\alpha$  subunits of PEP and bacterial RNAP. Multiple sequence alignment of the linker between the  $\alpha$  subunit N-terminal domain ( $\alpha$ -NTD) and C-terminal domain ( $\alpha$ -CTD) from representative species from diverse plants (top) and bacterial phyla (bottom) illustrate a conserved increase in sequence length of the  $\alpha$  linker in PEP relative to bacterial RNAP. The average linker length in the analyzed plant sequences is 41.1 ( $\sigma = 1.4$ ) amino acids and that of analyzed bacterial sequences is 25.5 ( $\sigma = 4.0$ ) amino acids. The insertion in the PEP  $\alpha$  subunit contains a helical segment ( $\alpha$ -LH). The position of the  $\alpha$ -LH in PEP is indicated along with an adjacent segment not ordered in the *S. alba* PEP cryo-EM reconstruction (asterisk).

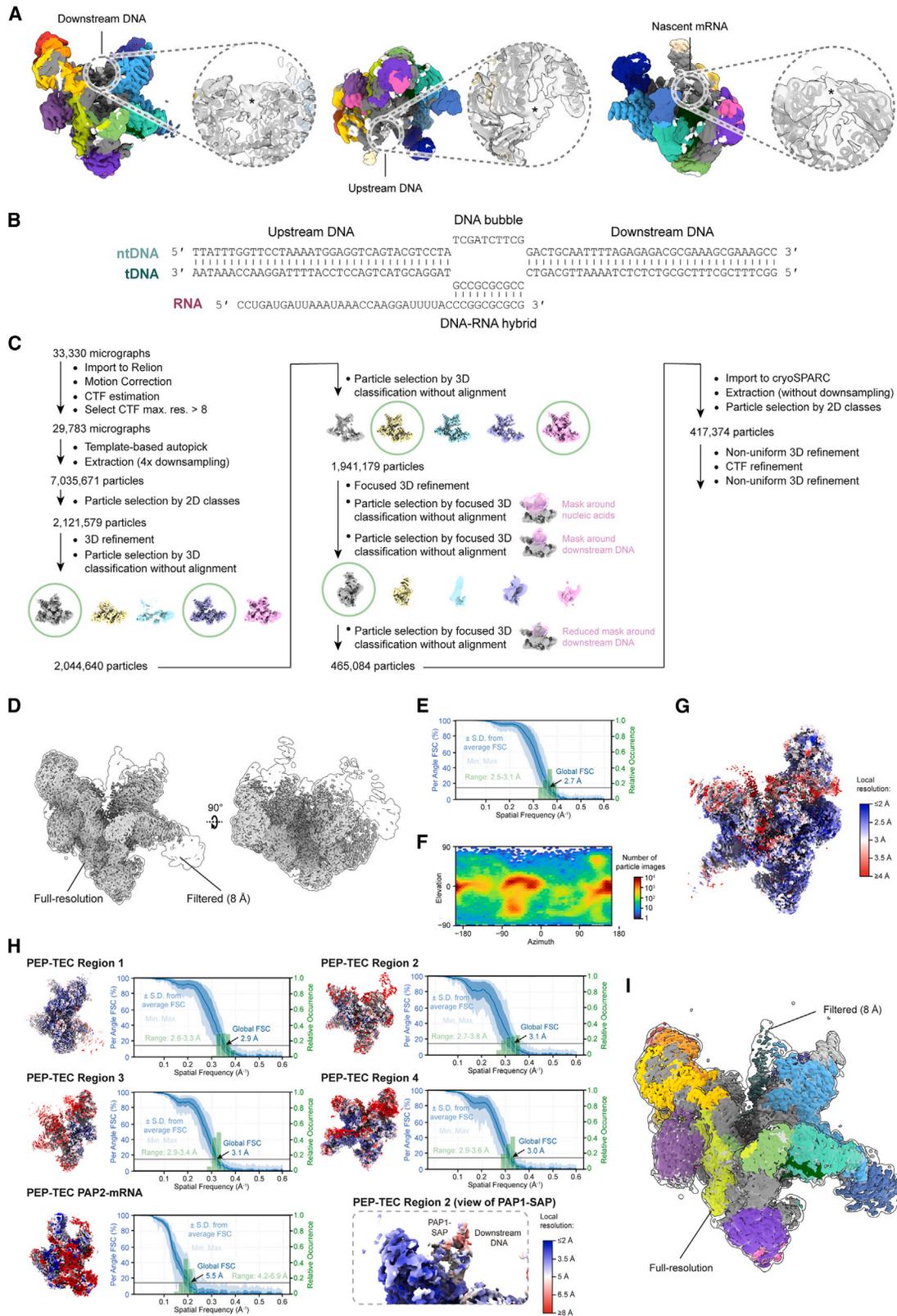
(C) Comparison of  $\beta'$ -SI1 domains of PEP (left) and *Synechocystis* RNAP (right). A structural search of the PDB for the PEP  $\beta'$ -SI1 domain using Dali did not identify any structural match in deposited structures.

(D) A region of  $\beta'$  that is longer in PEP (orange, left) than cyanobacteria RNAP (orange, right) forms a partially ordered loop that encloses part of PAP8.

(E) A search of the PDB for structural similarity to a  $\beta'$  domain specific to PEP and cyanobacteria using Dali revealed close resemblance to a single blade of  $\beta'$  propeller proteins, such as the DCAF12 protein shown.

(F) A short motif at the C terminus of PEP  $\beta''$  interacts with the interface of PAP1 and PAP2.

(G) Multiple sequence alignment shows similarity between the PEP subunit PAP12 from diverse plant species (top) and  $\omega$  from diverse bacterial species (bottom). Structural models shown: cyanobacteria RNAP (PDB: 8GZG), *Mycobacterium tuberculosis* RNAP (PDB: 5ZX3), and DCAF12 (PDB: 8AJM).



(legend on next page)

---

**Figure S5. Analysis of PEP transcription elongation complex by cryo-EM, related to Figure 4**

(A) Cryo-EM reconstruction of PEP filtered to 6 Å resolution shows density for downstream DNA (left, asterisk), upstream DNA (middle, asterisk) and nascent mRNA (right, asterisk). Reconstructions are shown colored by assigned protein according to scheme in Figure 1.

(B) Nucleic acid scaffold used for PEP-TEC reconstitution.

(C) Cryo-EM data processing workflow for analysis of PEP-TEC. Classes produced by 3D classification are shown, with particles within the indicated classes (green circles) selected for further processing.

(D) Consensus cryo-EM reconstruction of PEP-TEC shown at full resolution (opaque) and low-pass filtered to 8 Å (transparent).

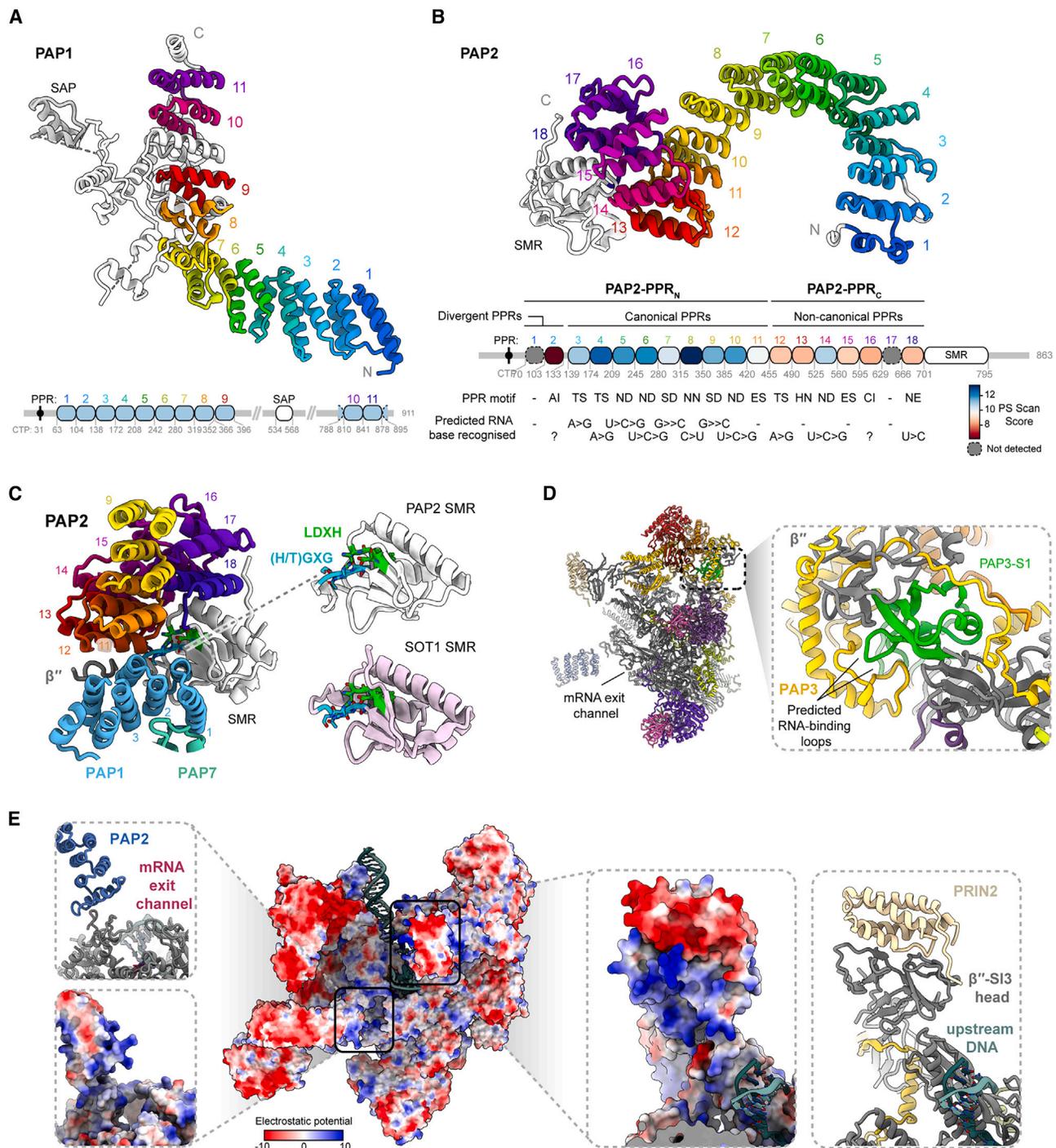
(E) Fourier shell correlation plot for consensus PEP-TEC reconstruction. Dotted lines indicate the 0.143 threshold for gold-standard half map correlation.

(F) Angular distribution plot for consensus PEP-TEC reconstruction.

(G) Local resolution estimate for consensus PEP-TEC reconstruction.

(H) Fourier shell correlation plots and local resolution estimates for focused reconstructions of PEP-TEC. Gray lines indicate the 0.143 threshold for gold-standard half map correlation. Local resolution color range as in (F), except PEP-TEC Region 2 inset of the PAP1-SAP domain, which is indicated.

(I) Composite cryo-EM reconstruction generated from focused reconstructions at full-resolution (opaque and colored by assigned subunit), and low-pass filtered to 8 Å showing additional regions not resolved to high resolution (transparent).



**Figure S6. Interactions between PEP and nucleic acids, related to Figure 5**

(A) Model of PAP1 colored by assigned PPR repeat. PAP1 contains a tandem array of nine PPRs at the N terminus (repeats 1–9), and two PPRs at the C terminus (repeats 10 and 11) that are bordered by two partial PPR repeats (not colored). The PPRs of PAP1 do not sufficiently resemble canonical PPRs to be detected by PPRCODE<sup>33</sup> for prediction of RNA-binding specificity. PPRs were detected using TPRpred<sup>86</sup> and the positions were refined by inspection of the structure.

(B) Model of PAP2 colored by assigned PPR repeat. PAP2 contains a continuous array of 18 PPR-like repeats. The model shows that PAP2-PPR<sub>N</sub> adopts a semi-circular arch, whereas the PAP2-PPR<sub>C</sub> forms a tight bundle closely bound to the PAP2 SMR domain. PPRs 1 and 17 were not sufficiently similar to a canonical PPR sequence to be assigned by PPRCODE and were identified by structural features. RNA-binding specificity predicted based on the PPR motif sequence is shown.

(C) PAP2 is connected to PEP through its interaction with the PAP1 N terminus. The PAP2 SMR domain and PAP2 PPRs 11 and 12 contact PAP1 PPRs 1 and 2. SMR domains of some proteins have DNA and RNA nuclease activities. The (H/T)GXG and LDXH motifs of SMR proteins, such as the SMR-PPR protein SOT1,

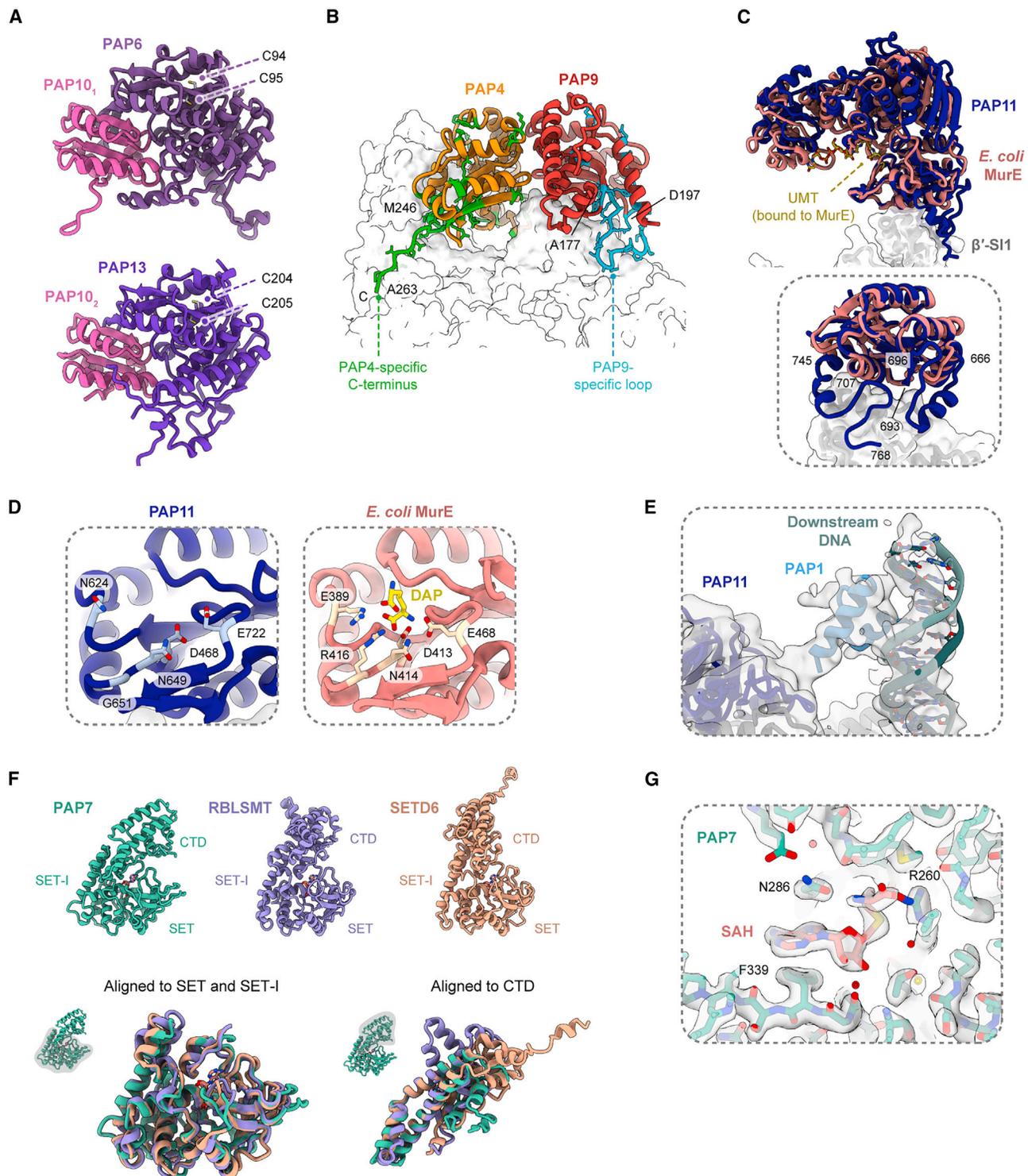
(legend continued on next page)

---

were found to be critical for RNA endonuclease and DNA nicking activity respectively.<sup>87,88</sup> Structural comparison of PAP2 SMR (top) and the SOT1 SMR predicted by AlphaFold (bottom) is shown with catalytic motifs indicated. PAP2 lacks the (H/T)GXXG motif,<sup>89</sup> and this loop instead supports binding of PAP2 SMR to PAP2 PPRs 13 and 14. The LDXH motif of PAP2 has the sequence VDVH, a variation that does not produce a significant structural difference to the predicted structure of the catalytically-active LDVR motif of SOT1. However, the PAP2 LDXH motif is concealed from the solvent by PAP2 PPRs 12 and 13 on one side and PAP1 on the other. Thus, we hypothesize that neither the (H/T)GXXG or LDXH motifs confer nuclease activity to PAP2.

(D) The S1 domain of PAP3 (green) is encased by protein within the  $\beta''$ -SI3 arch and the predicted RNA-binding loops (indicated) are concealed.

(E) Structural model of PEP-TEC in surface representation colored by electrostatic potential (center). Enlargement of positive charged surface of the PAP2-PPR<sub>N</sub> domain (left) and positive charged surface of  $\beta''$ -SI3 head bound to PAP15/PRIN2 (right).



**Figure S7. Structural details of PAP enzymatic activities, related to Figure 6**

(A) Structures of PEP PAP6-PAP10<sub>1</sub> heterodimer (top) and PAP13-PAP10<sub>2</sub> heterodimer (bottom). Cysteine residues of PAP6 and PAP13 that were identified to be required for the interaction with PAP10 in *Arabidopsis*<sup>15</sup> are indicated. As they are located within the center of the pfkB domain, we predict the interaction with PAP10 was abrogated by destabilization of the fold, rather than loss of the ability to form a disulfide bond.

(B) Structure of PEP PAP4-PAP9 heterodimer highlighting residues present in PAP4 but not PAP9 (green), and residues present in PAP9 but not PAP4 (cyan). The most significant differences between the homologous subunits are a C-terminal sequence unique to PAP4 and an extended loop unique to PAP9.

(legend continued on next page)

---

(C) PAP11 is linked to PEP through binding of its C-terminal domain to the  $\beta'$ -Sl1 domain. Structural alignment of PAP11 and *E. coli* MurE shows that the interaction interface of PAP11 involves a helix also present in MurE (*S. alba* PAP11 residues 696–707), and two regions unique to PAP11: an extended loop (*S. alba* PAP11 residues 666–693) and a sequence at the C terminus (*S. alba* PAP11 residues 745–768). Structural model shown for comparison is PDB: 1E8C.

(D) Structural comparison of the C-terminal domains of PAP11 (left) and *E. coli* MurE (right). The C-terminal domain of MurE contains the substrate-binding pocket for the amino acid diaminopimelic acid (DAP), which the enzyme ligates to an amino acid-nucleotide compound, UDP-MurNAc-Ala-Glu (UMAG), to produce UDP-MurNAc-Ala-Glu- $A_2$ pm (UMT). Structural model shown for comparison is PDB:1E8C, and contains the product, UMT, and the region of this that corresponds to the DAP substrate is displayed (yellow). The substrate of PAP11, if it has similar ligase activity, is unknown, and the region corresponding to the DAP-binding pocket of MurE is shown for comparison. The identity of three of the five residues of MurE that contact DAP are conserved in PAP11, suggesting the possible substrate of PAP11 has similarity to DAP, but is likely not identical.

(E) The PAP1-SAP domain interacts with the PAP11 C-terminal domain. Structural model is shown overlaid with PEP-TEC cryo-EM reconstruction filtered to 5 Å resolution.

(F) Structural comparison of PAP7 to homologous lysine methyltransferases RBLSMT (PDB: 2H21) and SETD6 (PDB: 3RC0) (top). Structural alignment of models to the SET domain (left) or CTD (right) shown.

(G) Structural model of PAP7 overlaid with cryo-EM reconstruction showing the presence of bound S-adenosyl homocysteine (SAH), a cofactor product of lysine methyltransferase reaction.