



**HAL**  
open science

## Three-Dimensional Envelope and Subunit Interactions of the Plastid-Encoded RNA Polymerase from *Sinapis alba*

Rémi Ruedas, Soumiya Sankari Muthukumar, Sylvie Kieffer-Jaquinod, François-Xavier Gillet, Daphna Fenel, Grégory Effantin, Thomas Pfannschmidt, Yohann Couté, Robert Blanvillain, David Cobessi

### ► To cite this version:

Rémi Ruedas, Soumiya Sankari Muthukumar, Sylvie Kieffer-Jaquinod, François-Xavier Gillet, Daphna Fenel, et al.. Three-Dimensional Envelope and Subunit Interactions of the Plastid-Encoded RNA Polymerase from *Sinapis alba*. *International Journal of Molecular Sciences*, 2022, 23 (17), pp.9922-10.3390/ijms23179922 . hal-03774879

**HAL Id: hal-03774879**

**<https://hal.science/hal-03774879>**

Submitted on 12 Sep 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Article

# Three-Dimensional Envelope and Subunit Interactions of the Plastid-Encoded RNA Polymerase from *Sinapis alba*

Rémi Ruedas <sup>1,†</sup> , Soumiya Sankari Muthukumar <sup>1,2</sup>, Sylvie Kieffer-Jaquinod <sup>3</sup>, François-Xavier Gillet <sup>2,‡</sup>, Daphna Fenel <sup>1</sup>, Grégory Effantin <sup>1</sup> , Thomas Pfannschmidt <sup>2,§</sup>, Yohann Couté <sup>3</sup> , Robert Blanvillain <sup>2,\*</sup> and David Cobessi <sup>1,\*</sup>

<sup>1</sup> CNRS, CEA, IBS, University Grenoble Alpes, 38000 Grenoble, France

<sup>2</sup> CNRS, CEA, INRAE, IRIG-LPCV, University Grenoble-Alpes, 38000 Grenoble, France

<sup>3</sup> INSERM, CEA, University Grenoble Alpes, UMR BioSanté U1292, CNRS, CEA, FR2048, 38000 Grenoble, France

\* Correspondence: robert.blanvillain@univ-grenoble-alpes.fr (R.B.); david.cobessi@ibs.fr (D.C.)

† Current address: Institut de Biologie Intégrative de la Cellule, UMR9198 I2BC, Avenue de la Terrasse, 91191 Gif-sur-Yvette, France.

‡ Current address: UMR 5240 MAP: CNRS, UCB, INSA, Bayer CropScience, Domaine Lyon Tech-La Doua 10, rue Dubois, Bât Lwoff, RDC, 69622 Villeurbanne, France.

§ Current address: Institut für Botanik, Leibniz-Universität Hannover, Herrenhäuser Str. 2, 30419 Hannover, Germany.



**Citation:** Ruedas, R.;

Muthukumar, S.S.;

Kieffer-Jaquinod, S.; Gillet, F.-X.;

Fenel, D.; Effantin, G.;

Pfannschmidt, T.; Couté, Y.;

Blanvillain, R.; Cobessi, D.

Three-Dimensional Envelope and Subunit Interactions of the Plastid-Encoded RNA Polymerase from *Sinapis alba*. *Int. J. Mol. Sci.* **2022**, *23*, 9922. <https://doi.org/10.3390/ijms23179922>

Academic Editor: Bartolome Sabater

Received: 13 July 2022

Accepted: 26 August 2022

Published: 31 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** RNA polymerases (RNAPs) are found in all living organisms. In the chloroplasts, the plastid-encoded RNA polymerase (PEP) is a prokaryotic-type multimeric RNAP involved in the selective transcription of the plastid genome. One of its active states requires the assembly of nuclear-encoded PEP-Associated Proteins (PAPs) on the catalytic core, producing a complex of more than 900 kDa, regarded as essential for chloroplast biogenesis. In this study, sequence alignments of the catalytic core subunits across various chloroplasts of the green lineage and prokaryotes combined with structural data show that variations are observed at the surface of the core, whereas internal amino acids associated with the catalytic activity are conserved. A purification procedure compatible with a structural analysis was used to enrich the native PEP from *Sinapis alba* chloroplasts. A mass spectrometry (MS)-based proteomic analysis revealed the core components, the PAPs and additional proteins, such as FLN2 and pTAC18. MS coupled with crosslinking (XL-MS) provided the initial structural information in the form of protein clusters, highlighting the relative position of some subunits with the surfaces of their interactions. Using negative stain electron microscopy, the PEP three-dimensional envelope was calculated. Particles classification shows that the protrusions are very well-conserved, offering a framework for the future positioning of all the PAPs. Overall, the results show that PEP-associated proteins are firmly and specifically associated with the catalytic core, giving to the plastid transcriptional complex a singular structure compared to other RNAPs.

**Keywords:** *Sinapis alba*; plastid-encoded RNA polymerase; PEP associated proteins; transcription; photomorphogenesis; photosynthesis; chloroplast biogenesis

## 1. Introduction

DNA-dependent RNA polymerases (RNAPs) are central enzymes of gene expression, which transcribe the genetic information encoded in DNA into single-stranded RNAs, some of which are suitable for translation. RNAPs exist in highly varying degrees of complexity ranging from single subunit enzymes in T3/T7 phages to highly multimeric enzymes in eukaryotes. Eubacterial multimeric RNAPs share a common catalytic core composed of two large subunits called  $\beta$  and  $\beta'$ , a dimer of  $\alpha$  subunits and a monomer of the  $\omega$  subunit [1–3]. For specific transcriptional activity, RNAPs require additional proteins such as  $\sigma$  factors that mediate the recognition of gene promoters and are essential to initiate

transcription. The three-dimensional structures of RNAPs have been solved for eukaryotic and prokaryotic RNAPs in several states [3–5]. Structural comparisons of RNAPs have shown that, even when the sequence identity is low, the overall shape of the five core subunits is largely conserved [3]. Furthermore, homologous regions at the structural level have been identified between the bacterial and eukaryotic RNAPs, suggesting that the fold is better conserved than the amino acid sequences. The essential residues and regions for effective transcription are, however, conserved, indicating that the enzymes share a common transcription mechanism [1]. In eukaryotes, several RNAPs are involved in the transcription of nuclear genes (RNAPs I, II and III), while a specific phage-type RNAP transcribes the mitochondrial DNA. Plant cells, in addition, possess a third genome in plastids with complex transcriptional machinery to express it. Plastids evolved from the engulfment of an ancient cyanobacterium into a mitochondriate proto-eukaryote around 1.5 billion years ago [6]. Thereafter, a massive lateral transfer of cyanobacterial genes into the nucleus reshaped the two genomes [7]. As a result, most plastome (chloroplast DNA, cpDNA) of today's plastids contains only about 120 genes [8], encoding (i) components of the plastid gene expression machinery (the core subunits of the prokaryotic-type RNA polymerase, ribosomal proteins, tRNAs and rRNAs); (ii) subunits of each of the major functional photosynthesis-related complex (e.g., ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), photosystem I and II (PSI and PSII), cytochrome b6f complex, NADH dehydrogenase and the ATP synthase) and (iii) a few proteins involved in other essential processes, such as protein import, fatty acid synthesis or protein homeostasis (e.g., YCF1 and 2, AccD and ClpP1) [9,10]. Despite the limited number of plastid genes, chloroplasts contain 2500–3500 different proteins [11]; thus, the vast majority of chloroplast proteins are encoded by the nuclear genome and must be post-translationally imported. The expression of the cpDNA is, however, essential to chloroplast biogenesis and functions since drug-based or genetic impairments of plastid gene expression result in albinism [12].

Transcription of the plastome involves a single-subunit nuclear-encoded T3/T7 phage-type RNA polymerase (NEP) and the multi-subunit plastid-encoded prokaryotic-type RNA polymerase (PEP). Briefly, the NEP enzyme transcribes the so-called 'house-keeping' genes (including *rpo* genes encoding the core subunits of the PEP), while the PEP preferentially transcribes genes encoding proteins of the photosynthetic complexes, as well as tRNA genes [13,14]. However, some plastid genes possess promoters for NEP and PEP so that they can be transcribed by both RNA polymerases [15]. Furthermore, the division of labor between the two RNA polymerases changes with the developmental stage, and a clear-cut separation between NEP and PEP transcribed genes remains difficult [16]. The catalytic core enzyme of PEP comprises four subunits called  $\alpha$ ,  $\beta$ ,  $\beta'$  and  $\beta''$ , encoded by the genes *rpoA*, *rpoB*, *rpoC1* and *rpoC2*, respectively [17,18]. Biochemical studies performed in dark-grown mustard revealed that the core subunits assemble to form the prokaryotic-like enzyme PEP-B [19–21]. In angiosperm, seedlings illumination initiates a light signaling cascade that triggers photomorphogenesis and chloroplast biogenesis. This involves a structural reorganization of the PEP-B enzyme by association, with additional subunits resulting in a much larger multi-subunit PEP-A complex. Biochemical purifications performed in several plants revealed that the complex comprises at least 16 different proteins with an overall molecular mass of more than 900 kDa [22,23]. MS analyses of the mustard PEP-A complex allowed the identification of 10 PEP-associated proteins (PAPs) that are stably bound to the complex. Two additional proteins (PAP11/MurE and PAP12/pTAC7) were then added to the list of PAPs according to a set of criteria, including biochemistry (presence in the complex) and genetics (albino syndrome of the mutant) [18]. These PAPs are all encoded by the nuclear genome and must be imported in the stroma from the cytosol. The genetic inactivation of any of these 12 PAPs causes a severe block or disturbance of chloroplast biogenesis, indicating that the reorganization of the PEP complex represents a critical step in chloroplast biogenesis [12,23–31]. Therefore, understanding chloroplast biogenesis associated with photosynthesis in angiosperms requires studying the nuclear-

encoded PAPs that, added to PEP, regulate gene expression while protecting the machinery from the threatening reactions of photosynthesis.

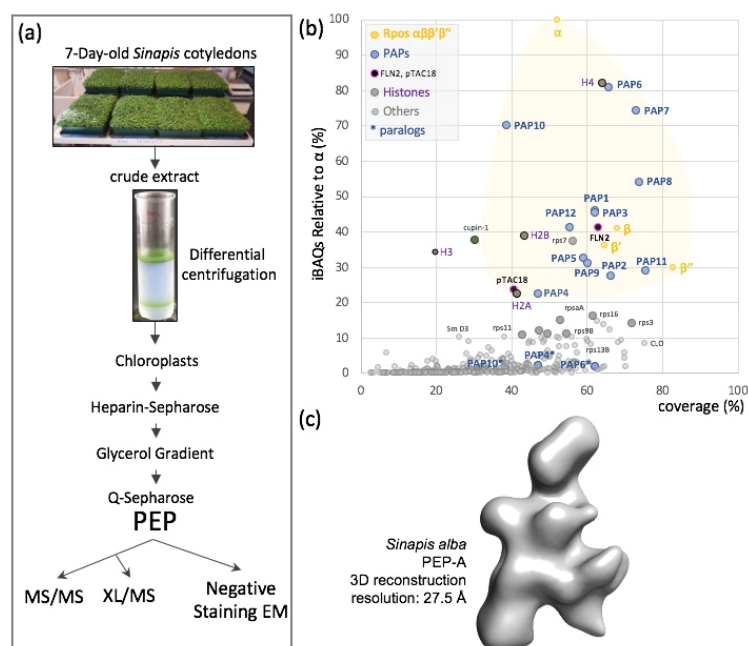
In contrast to RNAPs I, II and III, for which several three-dimensional structures were solved, the PEP-A structure remains unknown. Based on sequence homology, it is assumed that the PEP core enzyme would resemble that of the bacterial RNA polymerase (bRNAP). With the exception of PAP9, whose 3D structure was recently solved [32], only structure predictions of PAPs have been calculated based on their amino acid sequences, searching structural databases for homologous domains [18].

Here, we report the characterization of the PEP complex purified from *S. alba* cotyledons. A MS-based proteomic analysis identified all known PEP subunits and additional members, such as FLN2 and pTAC18. A chemical crosslinking coupled to MS approaches highlighted some interacting peptides in the PEP complex and provided initial structural information in the form of protein clusters, highlighting the relative position of some subunits with their surfaces of interaction. Using negative stain electron microscopy, we calculated the first 3D envelope of the PEP-A complex, showing together with the MS analyses that the PAPs are firmly and reproducibly associated with the catalytic core, each likely at its specific site. Interestingly, some surfaces of the interactions between the core and PAPs correspond to conserved regions of PAP-containing clades that are otherwise variable when bRNAPs are also considered.

## 2. Results

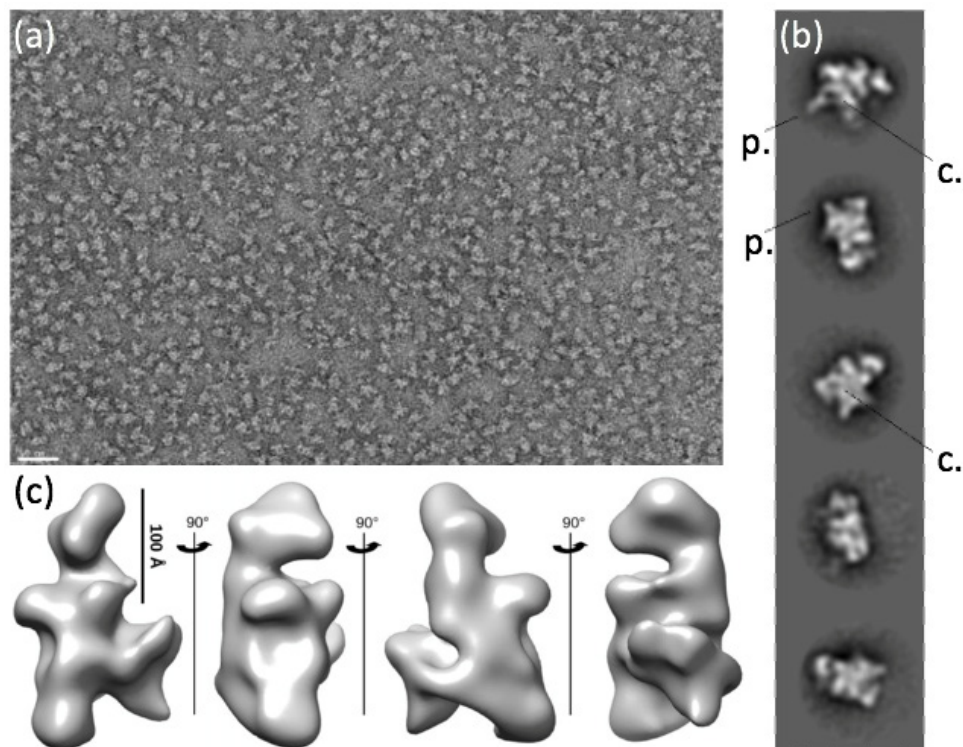
### 2.1. The PEP Complex and Its Associated Proteins

We used a MS-based label-free quantitative proteomic analysis to characterize the *S. alba* PEP-enriched fraction isolated from the chloroplasts of mustard cotyledon. An established purification scheme was used with slight modifications [33] (Figure 1a).



**Figure 1.** PEP composition and three-dimensional envelope. (a) Organelle fractionation, purification scheme and sample processing for mass spectrometry (MS/MS) or crosslinking mass spectrometry (XL/MS) or negative staining electron microscopy (eM). (b) Mass spectrometry data presented as relative iBAQ values to that of  $\alpha$  (iBAQr) as a function of the corresponding protein coverage expressed in percentage. Subunits  $\alpha$ ,  $\beta$ ,  $\beta'$  and  $\beta''$  are in yellow, PAPs in blue, suspected permanent residents in black, histones in magenta and suspected purification contaminants in different shades of grey. In the shaded yellow area fall all the expected components of the PEP-A complex and correspond to the major protein mass contribution to the purified sample. (c) *Sinapis alba* PEP-A envelope calculated from negative staining EM acquisitions (see Figure 2 for details).

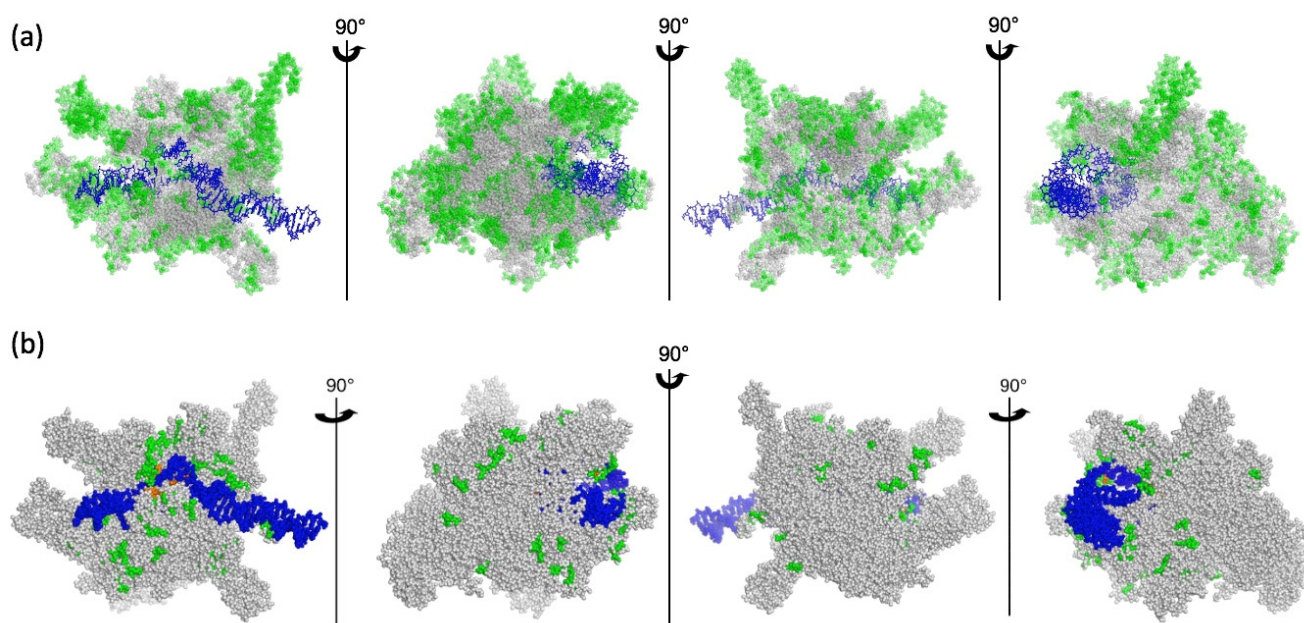
More than 400 different proteins were reproducibly identified and quantified in three independent preparations of PEP (Table S1). Their relative abundances within the PEP fraction were approximated using their extracted iBAQ values [34], showing that these proteins were distributed over four orders of magnitude. Among the 24 most abundant proteins, representing ~60% of the total amounts of proteins within the fraction, we identified the four core subunits ( $\alpha$ ,  $\beta$ ,  $\beta'$  and  $\beta''$ ) and the twelve PAPs (Figure 1b and Figure S1). The  $\alpha$  subunit was found to be approximately twice more abundant than the  $\beta$  subunit, consistent with a stoichiometry of two  $\alpha$  subunits per one  $\beta$  subunit in the catalytic core complex, as described in eubacterial RNAPs. Besides the 16 known PEP subunits, we identified PAP6/FLN1 paralogue fructokinase-like protein 2 (FLN2) and pTAC18 that were identified as a subunit of the plastid transcriptionally active chromosome [35]. The shortlist also contains two unexpected proteins, one homologous of the *A. thaliana* At4g36700 corresponding to a late embryogenesis abundant protein of the RmlC-like cupin superfamily and the chloroplast ribosomal protein Rps7. Whereas cupin may be found due to a spurious interaction related to its high abundance in the young seedling, the presence of Rps7 may be due to the proximity of the PEP to the ribosome. Such a proximity is observed in bacteria and is referred to as transcription–translation coupling. The remaining intruders among the 24 highly abundant proteins belong to the family of histones, suggesting that some nucleosomes copurify with the PEP fraction. This contamination is likely due to nuclei associated with the chloroplast envelopes. All other detected proteins are in the background noise (low stoichiometry of the peptides) corresponding to low-abundant proteins and reflecting the high sensitivity of mass spectrometry. In electron microscopy, though, the contaminant proteins within the sample did not interfere with the structural analysis of the PEP complex, since the individual particles appeared homogenous enough for the calculation of its three-dimensional envelope (Figures 1c and 2 below).



**Figure 2.** Negative-staining electron microscopy and 3D envelope of the PEP-A complex. (a) Overview image of the grid after negative staining. Note the homogeneity of the sample and the lack of other protein complexes. The white scale bar represents 50 nm. (b) Two-dimensional classes of PEP. (c) Three-dimensional envelope of PEP at 27.5 Å resolution calculated from 17,567 particles.

## 2.2. Patches of Specific Residues Are at the Surface of the PEP Catalytic Core

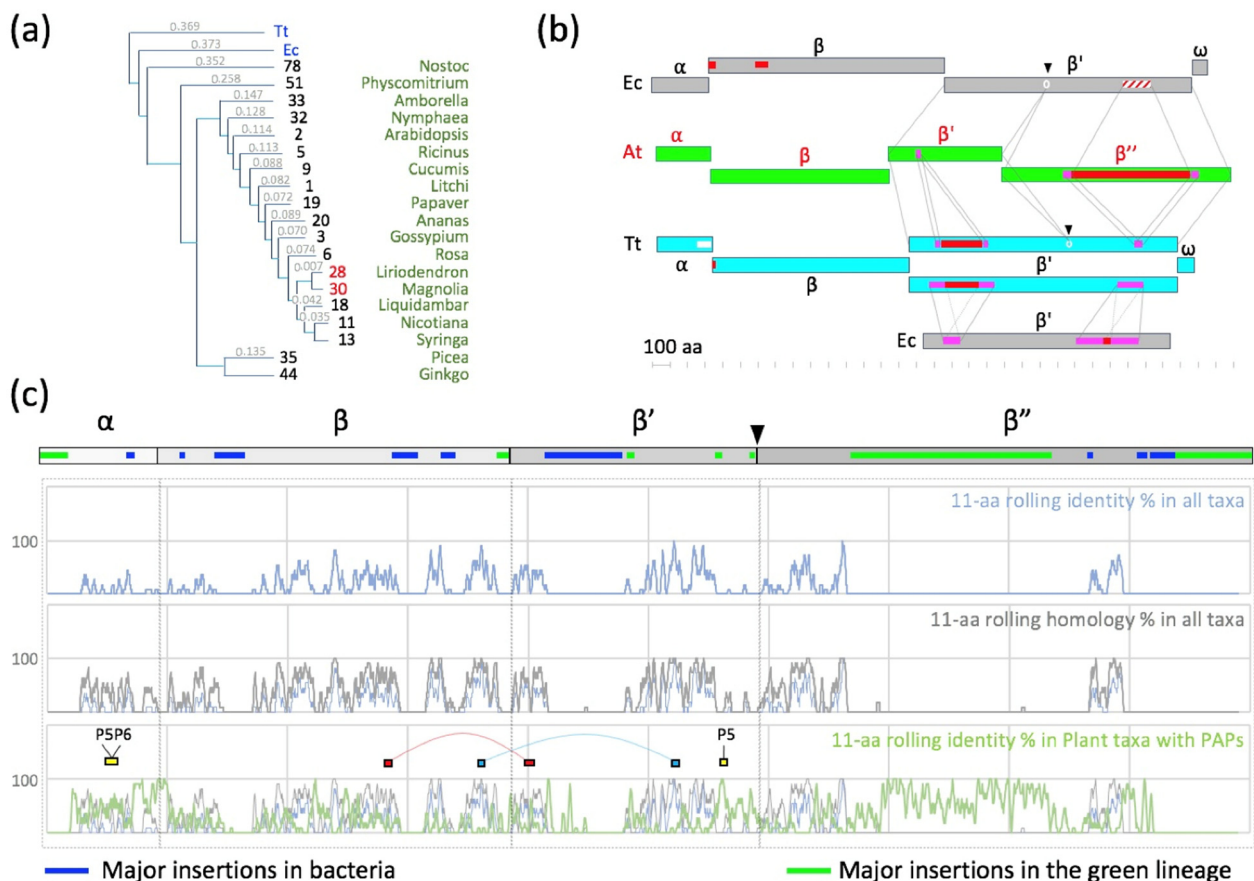
We hypothesized that the emergence of PAPs in the green lineage became essential for chloroplast biogenesis in angiosperms when all the PAPs acquired the capacity to bind to the core enzyme, hereby controlling its transcriptional activity in a “go/no go” switch that remains to be elucidated. It is then implying that surfaces of interactions on the core have evolved, possibly generating innovations (differences with ancestors) that are under selection pressure for conservation (Figure 3). To highlight the differences in the PEP core complex that could be evolutionarily associated with PAP interactions compared to eubacterial RNAPs, we performed a detailed sequence alignment analysis of the  $\alpha$ ,  $\beta$ ,  $\beta'$  and  $\beta''$  core subunits from various species chosen in the tree of the green lineage, as proposed by Finet et al. [36]. These sequences were found to be well-conserved within the green lineage (Figure 4 and Figures S2–S5). The lowest sequence identity is observed when comparing *Physcomitrium* to other species, the sequence of the  $\alpha$  subunit being the most divergent. Sequence conservation appears to be high in the domains of the  $\beta$ ,  $\beta'$  and  $\beta''$  subunits that bear the catalytic activity, while it is lower for the  $\alpha$  subunits that are responsible for the assembly of the core [37]. Sequence comparisons with RNAPs from bacteria and cyanobacteria reveal that the regions that are essential for the transcription activity are conserved, and the bacterial  $\beta'$  subunit can be aligned with the  $\beta'$  and  $\beta''$  subunits of the PEP (Figure 4b, Figures S4 and S5).



**Figure 3.** Mapping variable sites of the core subunits. View of the *E. coli* RNAP (PDB entry: 6GH5) without the  $\omega$  subunit and the  $\sigma_{54}$  factor. The double-stranded DNA is colored in blue. The core subunits are drawn in spheres. (a) Mapping the variable sites as homologous in grey and nonhomologous or gaps in green. (b) Mapping only amino acid functional differences between bRNAP and PEP, as given in the sequence alignments (Figures S2–S5). The residues colored in green and orange are those displaying a strong modification of functional groups for at least 3 consecutive amino acids.

Whereas the catalytic activity is carried by the  $\beta$  and  $\beta'$  subunits in bRNAPs, it is supported in the PEP by the  $\beta$ ,  $\beta'$  and  $\beta''$  subunits. Unlike in *E. coli*, the  $\beta$  subunit of the PEP does not have the additional  $\beta_{i4}$ ,  $\beta_{i9}$  and  $\beta_{i11}$  domains [38]. However, the  $\beta''$  subunit of the PEP contains a long plant-specific insertion of several hundred residues between regions  $\beta'b8$  and  $\beta'b9$  that does not exist either in the  $\beta'$  subunit from *E. coli* RNAP or in the  $\beta'$  subunit from *T. thermophilus* RNAP (Figure 4b,c and Figure S5). The  $\beta''$  subunit of RNAP from angiosperms also lacks a part of the  $\beta'b10$  region observed in the RNAP from

*Nostoc*. Nevertheless, most of the strictly conserved residues described for the catalytic cores of RNAPs [39] are conserved in the PEP. The amino acid homologies were mapped on the *E. coli* 3D structure (Figure 3). Most of the variable sites in the PEP sequences are located at the surface of the catalytic core of the bRNAP, supporting the assumption that some of these innovations may be required for the interaction with PAPs (Figure 3a). The overall difference in amino acid functionalities, however, is rather limited outside of the  $\beta''$  large insertion that remains invisible in these representations (Figure 3b).



**Figure 4.** Phylogeny and sequence alignments of the core subunits. (a) Phylogram obtained with the Clustal Omega multi-alignment algorithm. Branch length presented as a cladogram. Major taxa included from the collection presented in the data source (Excel sheet sorted). A major incongruence from the angiosperm phylogeny tree (version IV: <http://www.mobot.org> accessed on 1 January 2022) is noted for Magnoliales and likely due to the study of chloroplast genes with cytoplasmic inheritance. (b) Schematic representation of the sequence context of *E. coli* (Ec), *T. thermophilus* (Tt) and *A. thaliana* (At) RNAP or PEP subunits as the output of a dot plot analysis performed using dotmatcher (<https://www.bioinformatics.nl/cgi-bin/emboss/dotmatcher> accessed on 1 January 2022). Insertions are represented in red or dashed red, with the duplicated area in pink. The splits of the bacterial  $\beta'$  in the PEP  $\beta'$  and  $\beta''$  are presented with a light-grey circle and a black triangle separating the shared regions. (c) Global alignment represented as the 11-aa rolling identity (blue) or homology (grey) percentages calculated for all taxa. In green is the 11-aa rolling identity percentage calculated in a subset of taxa corresponding to plants with detected PAPs (green). The black triangle is the evolutionary split of the *rpoC* gene in the *rpoC1* and *rpoC2* genes in the cyanobacteria. Red and blue rectangles represent dipeptides between  $\beta$ - $\beta'$ , while yellow rectangles represent interacting peptides with PAPs, as found in the XL-MS analysis (see below).

### 2.3. A Chloroplast Catalytic Core Surrounded with Nuclear-Encoded Proteins

We then investigated the 3D structure of the fully assembled PEP complex by using negative-staining electron microscopy. The overview images of the stained complexes dis-

played well-separated molecules of various shapes but very limited aggregation (Figure 2) and no disturbance by other complexes (such as nucleosomes). The homogeneity of the sample was probed by *ab initio* 2D classification of the individual complex images that revealed several well-defined 2D classes (Figure 2b). The overall shapes of the classes are multiple, but they are all consistent in sizes with dimensions varying between 150 and 280 Å. Some 2D classes of PEP displayed a more compact center, sometimes with a clear stain-filled pocket surrounded by several protrusions of various sizes (Figure 2b). From the particles isolated by 2D classification, a 3D map at 27.5 Å resolution could be determined (Figure 2c), which recapitulates the features seen in the 2D classes, such as the central cavity (depression) and the peripheral protrusions. The resolution was not sufficient, though, to confidently fit the catalytic core of the *E. coli* RNAP in the PEP-A envelope.

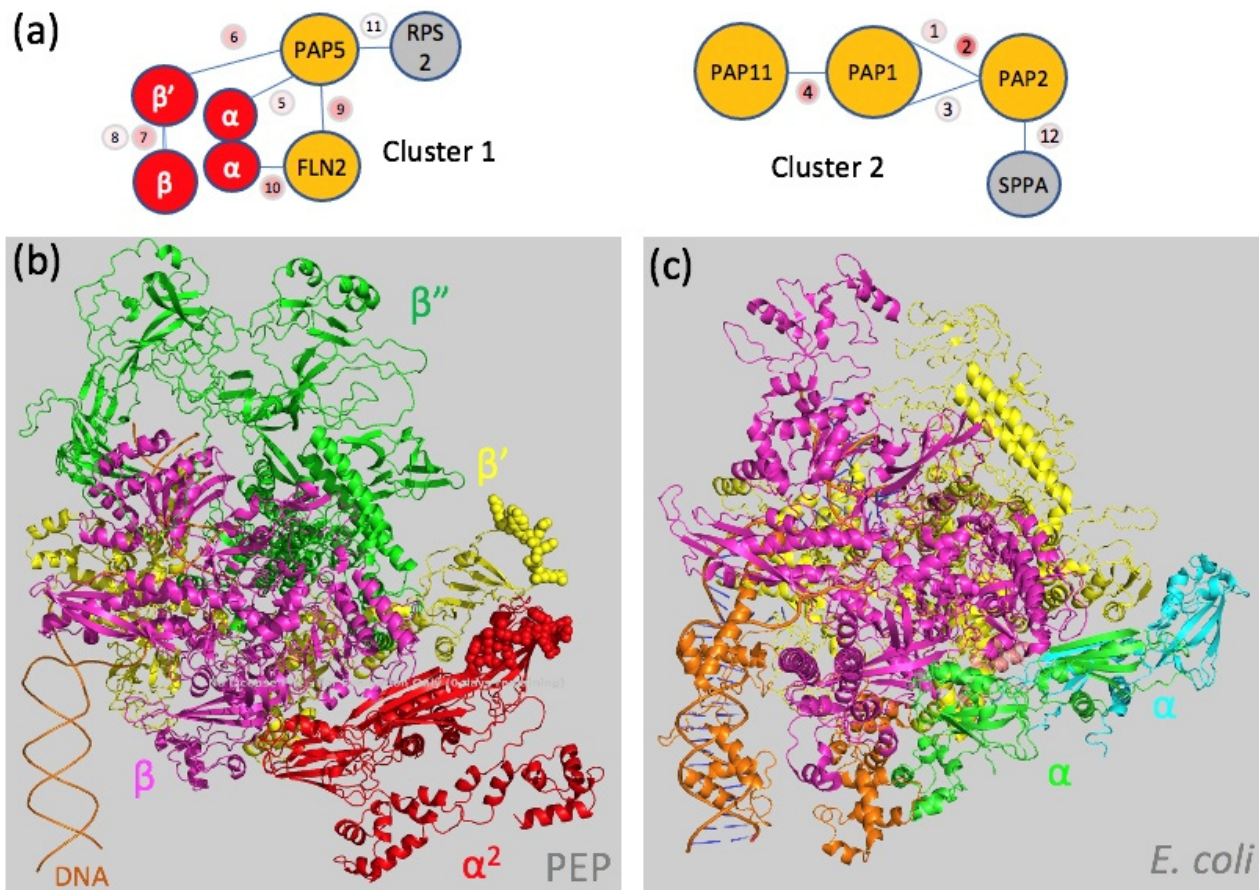
In order to obtain information about the relative position of the PEP subunits within the complex, we used a biochemical crosslinking coupled to MS. To this end, we treated the PEP-enriched fraction from two independent purifications (replicates 2 and 3 of the preparations used for the proteomic discovery) with Disuccinimidyl Dibutyric Urea (DSBU) before tryptic digestion and the MS analyses. This strategy allowed to reliably identify 39 interprotein dipeptides, 12 of which contained PEP core subunits or PAPs, suggesting a spatial proximity between these subunits within the PEP complex (Table 1 and Table S2).

The core subunits were partly accessible to the DSBU treatment, since two dipeptides linking the  $\beta$  and  $\beta'$  subunits were identified, suggesting that the associated PAPs do not cover the core completely but leave some gaps that allow the crosslinker molecules to access the core. Structure analyses of the RNAPs from *E. coli* (PDB entries: 3LU0 [38] and 6GH5 [40]) and *T. thermophilus* (PDB entry: 6ASG [41]) do not allow to model the dipeptides observed, suggesting that these regions in the PEP have different conformations despite their sequence conservation (Figures S2–S5). PAP5 and FLN2 were found to both interact with the same peptide of the  $\alpha$  subunit, indicating that PAP5 interacts with one monomer while FLN2 interacts with the second monomer (Table 1 and Figures 2c and 5). A distinct region of PAP5 was found in close vicinity to the KNYQNER peptide of the  $\beta'$  subunit (Table 1) that belongs to an insertion of conserved residues found only in angiosperms after the  $\beta'a12$  domain (Figure 2c and Figure S4). This result supports the assumption that surface-localized residues that are not conserved between the catalytic cores of bRNAPs and PEP but conserved in plants have evolved towards the interactions with the PAPs (see below). We also found a PAP5-FLN2 dipeptide, suggesting that the  $\alpha$ ,  $\beta$  and  $\beta'$  subunits PAP5 and FLN2 may form a structural cluster within the fully assembled PEP complex (Figure 5a). A second cluster appears to be formed by PAP1, PAP2 and PAP11/MurE-like for which dipeptides were also found (Table 1 and Figure 5a).



**Table 1.** Characterization of the proximal proteins in the *S. alba* PEP fraction using crosslinking-MS. Selection of the 12 best hetero-dipeptides is presented with the corresponding protein partners, crosslink score, peptide sequences, position and crosslinked amino acid with a relative position to the peptide. The overall dipeptides are given in Table S2.

| #  | Protein 1 Names    | Protein 2 Names | xLinkScore | Peptide 1                | From | To  | aa 1 | Peptide 2         | From | To  | aa 2 |
|----|--------------------|-----------------|------------|--------------------------|------|-----|------|-------------------|------|-----|------|
| 1  | PAP1/pTac3         | PAP2/pTac2      | 72.00      | [KELGAGQRPLPETMIALVR]    | 131  | 149 | K1   | [GQLEKSSAAR]      | 753  | 762 | K5   |
| 2  | PAP1/pTac3         | PAP2/pTac2      | 194.61     | [KELGAGQRPLPETMIALVR]    | 131  | 149 | K1   | [GQLEKSSAAR]      | 753  | 762 | K5   |
| 3  | PAP1/pTac3         | PAP2/pTac2      | 49.10      | [ENEDSSSFGSSEAVSALER]    | 50   | 68  | S15  | [GQLEKSSAAR]      | 753  | 762 | S6   |
| 4  | MURE               | PAP1/pTac3      | 133.68     | [ELKPR]                  | 608  | 612 | K3   | [VQKAR]           | 564  | 568 | K3   |
| 5  | SaRpoA             | PAP5/PTAC12     | 57.62      | [GYSLKMSNNFEDR]          | 156  | 168 | Y2   | [IKRDPLAMR]       | 365  | 373 | K2   |
| 6  | PAP5/PTAC12        | SaRpoC1         | 99.25      | [KLGRPHPFIDPTK]          | 208  | 220 | K1   | [KNYQNER]         | 683  | 689 | K1   |
| 7  | SaRpoC1            | SaRpoB          | 108.78     | [IFGPIKSGIBABGNYSR]      | 60   | 75  | Y15  | [LTPQVAKESYAPEDR] | 733  | 748 | K7   |
| 8  | SaRpoC1            | SaRpoB          | 52.00      | [FRETLLGKR]              | 489  | 497 | K8   | [SKQGGQR]         | 969  | 975 | S1   |
| 9  | PAP6/FLN1;<br>FLN2 | PAP5/PTAC12     | 89.14      | [KLELVGSMGEDDDSS}        | 602  | 617 | K1   | [NWSVLKSTPELR]    | 481  | 492 | K6   |
| 10 | PAP6/FLN1;<br>FLN2 | SaRpoA          | 121.48     | [MLTVQPDLMNKGYLER]       | 505  | 521 | Y14  | [GYSLKMSNNFEDR]   | 156  | 168 | K5   |
| 11 | PAP5/PTAC12        | RPS2A; RPS2B    | 39.00      | [APQPAGESSSFPYSGKNPGRSR] | 128  | 148 | S20  | [EVATAIR]         | 137  | 143 | T4   |
| 12 | PAP2/pTac2         | SPPA            | 67.62      | [GGLFKESEVILSR]          | 503  | 515 | S7   | [GQISDQLKSR]      | 135  | 144 | K8   |



**Figure 5.** Mapping protein interactions on the core complex. (a) Protein clusters determined from the XL-MS analysis (Table 1) are schematically presented with the link and scores; grey bubbles correspond to the protein not belonging to the PEP-A purified complex: RPS2, Ribosomal Protein S2; SPPA, light-inducible chloroplast protease complex associated with thylakoid membranes. Cluster 1 composed of the PAP5, FLN2,  $\alpha$  and  $\beta'$  subunits. Cluster 2 composed of PAP1, 2 and 11. (b) Model of the PEP core complex from *A. thaliana* built from the  $\alpha$ ,  $\beta$ ,  $\beta'$  and  $\beta''$  subunits modeled using AlphaFold [42] and superimposed onto the *E. coli* RNAP catalytic core and colored as follows:  $\alpha$  subunit in red,  $\beta$  subunit in pink,  $\beta'$  subunit in yellow and  $\beta''$  in green. The van der Waals spheres display the peptides of the  $\alpha$  and  $\beta'$  subunits that are nearby to PAP5 and FLN2 (Table 1). (c) View of the catalytic core from the *E. coli* RNAP (PDB entry: 6GH5 [40]).

### 3. Discussion

The purification protocol used in this study allowed us to retrieve a stable PEP complex with a limited amount of contaminant proteins. The core subunits and previously described PEP-associated proteins are the most abundant proteins. The three MS-based proteomic characterizations of the *Sinapis alba* PEP fraction revealed the presence of FLN1 (PAP6) and FLN2, two fructokinase-like proteins whose gene deletion lead, respectively, to an albino phenotype or a delayed greening [43]. FLN2 is the paralogous protein of FLN1, and despite its fructokinase domain, sugar-phosphorylating activity remains to be detected [26]. They can form homodimers or heterodimers in vitro [44]. Characterization of the proximal proteins in the *S. alba* PEP fraction using XL-MS showed that FLN1 or FLN2 interact with the  $\alpha$  subunit of the catalytic core. Based on the sequence, it is not possible to distinguish which FLN paralog binds to the  $\alpha$  subunit due to the high sequence identity between FLN1 and FLN2 that display the same identified peptide sequence. The part of the  $\alpha$  subunit observed in this interaction (GY(157)SLK(160)MSNNFEDR) is the same that also interacts with PAP5, involving Y157 and K160 in the dipeptide bond with PAP5 and FLN1/FLN2, respectively. Considering that the complex has a homogenous structure with correctly

positioned partners, steric hindrance would not allow for two proteins with predicted different folds (PAP5 and PAP6/FLN1 or FLN2) to interact with the same region of the  $\alpha$  subunit. The MS-based proteomic characterization of the *S. alba* PEP fraction also suggested that the  $\alpha$  subunit is twice more abundant than the  $\beta$  subunit. Together, these observations are consistent with a stoichiometry of two  $\alpha$  subunits per one  $\beta$  subunit [37] in the PEP core complex. Hence, this supports the assumption that the PEP core resembles that of bRNAPs. In the PEP, PAP5 and FLN2 form a cluster with the  $\alpha$ ,  $\beta$  and  $\beta'$  subunits, suggesting that they can associate early during a de novo PEP-B-to-PEP-A transformation. XL-MS also revealed the presence of two other closely related proteins, PAP1 and PAP2. Both PAP1 and PAP2 possess pentatricopeptide repeats involved in RNA binding. Among the PAPs with predicted nucleic acid-binding domains, PAP1 possesses a SAP domain known for DNA or RNA binding, while PAP3 has a S1-like domain predicted to interact with RNA [18]. Since dipeptides between PAP1 and PAP2 are found, both proteins may sit on the PEP as a heterodimer, PAP1 also being involved in interactions with PAP11/MurE-like (Table S2), and the three proteins form a second cluster containing the largest PAPs.

The presence of closely related proteins, such as PAP6 and FLN2 or the two superoxide dismutases PAP4 and PAP9, raises the question of the PEP subunit composition. Even if the 3D classifications did not reveal any significant variability in the 3D envelope, PEP heterogeneous complexes could exist. Furthermore, the PEP complex of our preparations could contain additional subunits such as FLN2 or pTAC18, not detected previously in gel-based MS analyses. It remains open whether these subunits represent loosely or tightly associated PEP subunits. The initial discovery of pTAC18 in the TAC already placed this protein conceptually close to the PEP [35]. Further biochemical analyses associated with a high-resolution cryo-EM map of the PEP and new XL-MS experiments with other crosslinkers will likely resolve the question about the bona fide PEP subunit composition and the potential existence of stage-specific differences.

Indeed, the PEP envelope was calculated at a resolution that does not allow fitting of the map with homologous structures of the catalytic core or PAPs such as PAP9 [32] or high-confidence PAP models. However, the proposed fitting of the catalytic core of the *E. coli* RNAP (PDB entry: 3LU0) [38] revealed the remaining space for the subsequent positioning of the PAPs (Figure S6). It is noteworthy that further 3D classifications did not reveal any significant variability in the 3D envelope of the PEP, suggesting that the protrusions that we attribute to the PAPs are firmly associated with the catalytic core. Despite the recognition of some structural features such as the cleft and stalk, the overall shape of the *S. alba* active PEP envelope is different from that of RNAPs II and III (Figure S7). The use of novel algorithms such as AlphaFold [42] is still limited to predicting larger complexes such as PEP-A in particular to address the spatial organization of the PAPs with the PEP core enzyme.

A sequence comparison (Figures S2–S5) shows that the four insertion regions characterized in *E. coli* RNAP [38] do not exist either in PEP or in the RNAP from *Nostoc*. The high sequence identity between the catalytic core of the bacteria and plastids suggests that the overall shape of the PEP core and the associated catalytic activity are conserved. The bacterial  $\beta'$  subunit has likely been split into two subunits during evolution after the separation of the eubacteria and cyanobacteria branches, the latest uniquely sharing the  $\beta''$  subunit with the chloroplast [45]. The sequence alignment showed that the  $\beta'$  and  $\beta''$  subunits of the PEP can be, respectively, aligned with the N-terminal and C-terminal parts of the  $\beta'$  subunit from bRNAPs. In addition, a very long insertion in the  $\beta''$  subunit of plastids and cyanobacteria (Phe364-Ser1093 in *A. thaliana*) is not observed in the C-terminal part of the  $\beta'$  subunit from bRNAPs. This insertion is located in the trigger loop region at the surface of the bRNAPs (Figure S5). With such a length, this region could be an additional domain in the PEP associated with oxygenic photosynthesis.

Sequence divergence with the *T. thermophilus* and *E. coli* RNAPs is mainly observed between residues located at the surface of the core complex. Since the nuclear-encoded PAPs seems to have appeared with the terrestrialization of the green lineage (first appearance in

fresh water algae and mosses), it is likely that the evolution of novel cell types requested some control of the PEP catalytic core activity, providing the capacity to generate novel plastid types. The PAPs, acting as signaling components expressed after phytochrome activation in the nucleus of angiosperms, may have been required to control PEP activity by the nucleus in order to synchronize the transcription of the photosynthesis-associated nuclear genes (*PhANGs*) and photosynthesis-associated plastid genes (*PhAPGs*) for the proper building of the photosynthetic apparatus upon first illumination. Due to their dual localization, some of the PAPs such as PAP5/HEMERA [46] and PAP8 [47–49] provided a potential regulatory link between the nucleus and plastids in the expression of photosynthesis genes. It remains to be solved whether their nuclear or their plastid function evolved first.

In conclusion, this study opens the road for an in-depth structural description of the PEP complex responsible for the expression of photosynthesis-associated plastid genes. This complex possesses a well-defined structure with subunits that are specifically associated with the catalytic core, providing essential functions related to efficient transcription, post-transcriptional modifications and protections against the threats of photosynthesis reactions.

#### 4. Materials and Methods

**Chloroplast isolation:** Six to seven-day-old *Sinapis alba* cotyledons were collected and homogenized using a blender with short pulses ( $3 \times 3$  s): 100 g approximately of fresh material in 200 mL homogenization buffer containing 50 mM HEPES-KOH, pH 8.0, 0.3 M sorbitol, 5 mM  $MgCl_2$ , 2 mM EDTA and 0.3 mM DTT. The suspension obtained was then filtered through a 56- $\mu$ m nylon mesh, then centrifuged 3 min at  $6084.1 \times g$  at 4 °C. The pellet was resuspended in homogenization buffer and poured in a potter to remove all the chloroplast aggregates. The suspension was then loaded on a linear percoll gradient (35% percoll, 50 mM HEPES-KOH, pH 8.0, 0.3 M sorbitol, 5 mM  $MgCl_2$ , 2 mM EDTA and 0.3 mM DTT) and centrifuged 50 min at  $4696 \times g$ , 4 °C. The fractions containing the chloroplasts were then pooled, diluted in homogenization buffer and centrifuged 10 min at  $4000 \times g$ , 4 °C to remove percoll. The pellet containing the chloroplasts was solubilized in the lysis buffer containing 50 mM Tris HCl, pH 7.6, 25% glycerol (*w/v*), 10 mM NaF, 4 mM EDTA, 1 mM DTT and 1% Triton X-100 (*w/v*) and poured in a potter for homogenization. The suspension was then centrifuged 1 h at  $15,000 \times g$ , 4 °C and the supernatant frozen in liquid nitrogen and stored at  $-80$  °C before using it to purify the PEP.

**PEP purification:** After thawing, the chloroplast lysate was mixed overnight at 4 °C with heparin resin equilibrated with 50 mM HEPES, pH 7.6, 10% (*w/v*) glycerol, 10 mM  $MgCl_2$ , 80 mM  $(NH_4)_2SO_4$ , 1 mM DTT and 0.1% (*w/v*) Triton X-100. The resin was extensively washed with 50 mM HEPES, pH 7.6, 10% (*w/v*) glycerol, 10 mM  $MgCl_2$ , 80 mM  $(NH_4)_2SO_4$ , 1 mM DTT and 0.1% Triton X-100 (*w/v*) before elution over 10 fractions of 1 mL with 50 mM HEPES, pH 7.6, 10% (*w/v*) glycerol, 10 mM  $MgCl_2$ , 1.2 M  $(NH_4)_2SO_4$ , 1 mM DTT and 0.1% Triton X-100 (*w/v*). The fractions were then subjected to SDS-PAGE and Western blot analyses with anti-PAP8 antibodies [47]. The fractions containing PAP8 and, therefore, the PEP were pooled; loaded on a 35–15% glycerol gradient (50 mM HEPES, pH 7.6, 35–15% (*w/v*) glycerol, 10 mM  $MgCl_2$  and 0.01% (*w/v*) Triton X-100) and centrifuged at  $97,083 \times g$  on a SW55-Ti rotor (Beckmann Coulter) for 16 h at 4 °C.

The gradient was then analyzed using SDS-PAGE and Western blot. The fractions containing the PEP were pooled before the last step of purification or frozen in liquid nitrogen and stored at  $-80$  °C. The pool containing the PEP was mixed overnight with Q-Sepharose resin (Amersham) pre-equilibrated in 50 mM HEPES, pH 7.6, 10% glycerol (*w/v*), 10 mM  $MgCl_2$  and 0.01% (*w/v*) Triton X-100. The complex was eluted using a 0–1 M NaCl gradient. The fractions containing the PEP were pooled and concentrated at  $2000 \times g$  on a 100-kDa cutoff membrane. The purified PEP was then frozen in liquid nitrogen and kept at  $-80$  °C before analyses.

Sequence alignments: Full-length coding sequences of the  $\alpha$ ,  $\beta$ ,  $\beta'$  and  $\beta''$  subunits were retrieved from Blastp. The protein sequences were aligned using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/> accessed on 1 January 2022) and then colored using the BOXSHADE server using default parameters. The domains of the  $\alpha$ ,  $\beta$ ,  $\beta'$  and  $\beta''$  subunits of the PEP were assigned based on those described [37,39].

MS-based proteomic analyses: Three PEP preparations from independently grown plant batches were analyzed. For this, purified PEP from chloroplasts was solubilized in Laemmli buffer and stacked in the top of a 4–12% NuPAGE gel (Invitrogen). After staining with R-250 Coomassie Blue (Bio-Rad), the proteins were digested in gel using trypsin (modified sequencing purity, Promega), as previously described [49]. The resulting peptides were analyzed by online nano-liquid chromatography coupled with MS/MS (Ultimate 3000 RSLCnano and Q-Exactive Plus, Thermo Fisher Scientific) using a 140-min gradient. For this purpose, the peptides were sampled on a precolumn (300  $\mu\text{m} \times 5$  mm PepMap C18, Thermo Scientific) and separated in a 75  $\mu\text{m} \times 250$  mm C18 column (Reprosil-Pur 120 C18-AQ, 1.9  $\mu\text{m}$ , Dr. Maisch). The MS and MS/MS data were acquired by Xcalibur (Thermo Fisher Scientific). Peptides and proteins were identified by Mascot (version 2.7, Matrix Science) through concomitant searches against the NCBI database (*Sinapis alba* strain: S2 GC0560-79 (white mustard) taxonomy, BioProject PRJNA214277, July 2020 download), the UniProt database (*Sinapis alba* taxonomy, February 2021 download), a homemade database containing the sequences of classical contaminant proteins found in proteomic analyses (human keratins, trypsin, etc.) and the corresponding reversed databases. Trypsin/P was chosen as the enzyme, and two missed cleavages were allowed. Precursor and fragment mass error tolerances were set, respectively, at 10 and 20 ppm. Peptide modifications allowed during the search were: Carbamidomethyl (C, fixed), Acetyl (Protein N-term, variable) and Oxidation (M, variable). The Proline software [50] was used for the compilation, grouping and filtering of the results (conservation of rank 1 peptides, peptide length  $\geq 6$  amino acids, peptide score  $\geq 25$ , allowing to reach a false discovery rate of the peptide spectrum match identifications  $< 1\%$ , as calculated on the peptide spectrum match scores by employing the reverse database strategy and the minimum of one specific peptide per identified protein group). Proline was then used to perform a MS1 label-free quantification of the identified protein groups based on razor and specific peptides. Intensity-based absolute quantification (iBAQ) [34] values were calculated from MS1 intensities of razor and specific peptides. The iBAQ values of each protein were normalized by the sum of the iBAQ values of all the quantified proteins in each sample before summing the values of the three replicates to generate the final iBAQ value. The gene names for the identified proteins were annotated after the Blastp search for the *A. thaliana* proteome.

Crosslinking coupled to MS analyses: A few micrograms of two PEP preparations used for mass spectrometry-based proteomic analyses (replicates 2 and 3) were crosslinked during 1 h at room temperature using 100  $\mu\text{M}$  of DSBU in HEPES buffer, pH 7.8. To quench the crosslinking reaction, one microliter of 1 M ammonium bicarbonate was added and the sample incubated for 15 min at room temperature. To reduce disulfide bonds, 100 mM DTT solution was added to obtain a final concentration of 3.5 mM, and the mixture was incubated at 56  $^{\circ}\text{C}$  for 30 min in a ThermoMixer. For the alkylation of cysteines, 50 mM IAA solution was added to a final concentration of 8 mM, and the mixture was incubated at room temperature in the dark for 20 min. Freshly prepared trypsin solution to an enzyme/protein ratio of  $\sim 1:50$  was added, and the digestion was performed overnight at 37  $^{\circ}\text{C}$ . To quench the enzymatic digestion, a final TFA concentration of 1% (*v/v*) was added. Micro-spin columns (Harvard Apparatus) were then used to desalt the samples using 5% ACN, 0.1% TFA as the washing solution and 75% ACN, 0.1% TFA as the elution buffer.

The resulting peptides were analyzed by online nano-liquid chromatography coupled with MS/MS (Ultimate 3000 RSLCnano and Orbitrap Exploris 480 for replicate 2 and Q-Exactive HF for replicate 3, Thermo Fisher Scientific). Peptides were sampled on a precolumn (300  $\mu\text{m} \times 5$  mm PepMap C18, Thermo Scientific) and separated using a Pharmafluidics  $\mu\text{PAC}^{\text{TM}}$  column of 200 cm in length (with a pillar array backbone at

an interpillar distance of 2.5  $\mu\text{m}$ ) using a 240-min method. Data were acquired in the data-dependent MS/MS mode with stepped higher-energy collision-induced dissociation (HCD) and normalized collision energies (20%, 25% and 35% for Orbitrap Exploris 480 and 22%, 27% and 30% for Q-Exactive HF).

Data analysis was conducted using MeroX 2.0 [51]. The following settings were applied: proteolytic cleavage: C-ter at Lys and Arg with 3 missed cleavages allowed, peptide length 4–30 amino acids, fixed modification: alkylation of Cys by IAA and variable modification: oxidation of Met, crosslinker: DSBU with specificity towards Lys, Ser, Thr, Tyr and N-ter for site 1 and 2, analysis mode: RISEUP mode, maximum missing ions: 2, precursor mass accuracy: 10 ppm, product ion mass accuracy: 30 ppm, signal-to-noise ratio: 2, precursor mass correction activated, pre-score cutoff at 10% intensity, FDR cut-off: 1% and minimum score cut-off: 30. Crosslinks identified in the two replicates were then combined using Merox.

Negative staining electron microscopy: Ten microliters of PEP were added to a glow discharge grid coated with a carbon-supporting film for 3 min, and the grid was stained with fifty microliters of Sodium Silico Tungstate (SST) (1% (*w/v*) in distilled water (pH 7–7.5)) for 2 min. The excess solution was soaked by a filter paper, and the grid was air-dried. The images were taken at 30,000 magnification (2.2  $\text{\AA}$ /pixel) under low-dose conditions ( $<10 \text{ e}^-/\text{\AA}^2$ ) with defocus values between  $-1.2$  and  $-2.5 \mu\text{m}$  on a Tecnai 12 (Thermo Fischer Scientific) LaB6 electron microscope operating at 120 kV using a Gatan Orius 1000 CCD camera.

Determination of the PEP envelope: The image processing was entirely done in RELION [52]. The CTF parameters of each micrograph were determined with CTFFIND4 [53], and the particles were auto-picked in RELION with the Laplacian of the Gaussian option. Two-dimensional classification was then performed in 50 classes using a 350  $\text{\AA}$  mask diameter that resulted in the selection of 17,567 particles. The latter were then used to create an ab initio model (C1 symmetry and 300  $\text{\AA}$  mask diameter) that was then used to calculate a 3D map (C1 symmetry and 320  $\text{\AA}$  mask diameter) at 27.5  $\text{\AA}$  resolution (at FSC = 0.143) (Figure S8).

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/ijms23179922/s1>.

**Author Contributions:** Conceptualization, R.B. and D.C.; R.R., F.-X.G., Y.C., S.K.-J., S.S.M., D.F., G.E., R.B. and D.C. performed the research. Y.C. and S.K.-J. contributed the MS data. D.F. and G.E. contributed the EM data. D.C. and R.B. wrote the manuscript with contributions from R.R., F.-X.G., Y.C., S.K.-J., D.F., G.E. and T.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Agence National de la Recherche (grant ANR-17-CE11-0031-02). This work used the platforms of the Grenoble Instruct-ERIC Centre (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-0005-02) and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). The proteomic experiments were partly supported by ProFI (ANR-10-INBS-08-01 grant).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The MS and crosslinking coupled to the MS data and were deposited to the ProteomeXchange Consortium via the PRIDE [54] partner repository with the dataset identifiers PXD032738 and PXD032739, respectively. Each spectrum corresponding to an interprotein link with the best scores were manually checked. The eM data were deposited into the EMDB with accession code EMD-14571.

**Acknowledgments:** We thank Maha Chieb for the technical help with the initial setup of the chloroplast preparation. IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG, CEA).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Cramer, P. Multisubunit RNA polymerases. *Curr. Opin. Struct. Biol.* **2002**, *12*, 89–97. [[CrossRef](#)]
2. Hirata, A.; Klein, B.J.; Murakami, K.S. The X-ray crystal structure of RNA polymerase from Archaea. *Nature* **2008**, *451*, 851–854. [[CrossRef](#)] [[PubMed](#)]
3. Murakami, K.S. Structural biology of bacterial RNA polymerase. *Biomolecules* **2015**, *5*, 848–864. [[CrossRef](#)]
4. Lee, J.; Borukhov, S. Bacterial RNA polymerase-DNfA interaction—the driving force of gene expression and the target for drug action. *Front. Mol. Biosci.* **2016**, *3*, 73. [[CrossRef](#)] [[PubMed](#)]
5. Hanske, J.; Sadian, Y.; Müller, C.W. The cryo-EM resolution revolution and transcription complexes. *Curr. Opin. Struct. Biol.* **2018**, *52*, 8–15. [[CrossRef](#)] [[PubMed](#)]
6. Bobik, K.; Burch-Smith, T.M. Chloroplast signaling within, between and beyond cells. *Front. Plant Sci.* **2015**, *6*, 781. [[CrossRef](#)]
7. Martin, W.; Rujan, T.; Richly, E.; Hansen, A.; Cornelsen, S.; Lins, T.; Leister, D.; Stoebe, B.; Hasegawa, M.; Penny, D. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12246–12251. [[CrossRef](#)]
8. Sugiura, M. The chloroplast genome. *Plant Mol. Biol.* **1992**, *19*, 149–168. [[CrossRef](#)]
9. Majeran, W.; Friso, G.; Asakura, Y.; Qu, X.; Huang, M.; Ponnala, L.; Watkins, K.P.; Barkan, A.; van Wijk, K.J. Nucleoid-enriched proteomes in developing plastids and chloroplasts from maize leaves: A new conceptual framework for nucleoid functions. *Plant Physiol.* **2012**, *158*, 156–189. [[CrossRef](#)]
10. Yu, Q.B.; Huang, C.; Yang, Z.N. Nuclear-encoded factors associated with the chloroplast transcription machinery of higher plants. *Front. Plant Sci.* **2014**, *5*, 316. [[CrossRef](#)]
11. Zybailov, B.; Rutschow, H.; Friso, G.; Rudella, A.; Emanuelsson, O.; Sun, Q.; van Wijk, K.J. Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE* **2008**, *3*, e1994. [[CrossRef](#)] [[PubMed](#)]
12. Pfalz, J.; Pfannschmidt, T. Essential nucleoid proteins in early chloroplast development. *Trends Plant Sci.* **2013**, *18*, 186–194. [[CrossRef](#)] [[PubMed](#)]
13. Hajdukiewicz, P.T.; Allison, L.A.; Maliga, P. The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J.* **1997**, *16*, 4041–4048. [[CrossRef](#)]
14. Williams-Carrier, R.; Zoschke, R.; Belcher, S.; Pfalz, J.; Barkan, A. A major role for the plastid-encoded RNA polymerase complex in the expression of plastid transfer RNAs. *Plant Physiol.* **2014**, *164*, 239–248. [[CrossRef](#)]
15. Weihe, A.; Börner, T. Transcription and the architecture of promoters in chloroplasts. *Trends Plant Sci.* **1999**, *4*, 169–170. [[CrossRef](#)]
16. Zhelyazkova, P.; Sharma, C.M.; Förstner, K.U.; Liere, K.; Vogel, J.; Börner, T. The primary transcriptome of barley chloroplasts: Numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *Plant Cell* **2012**, *24*, 123–136. [[CrossRef](#)] [[PubMed](#)]
17. Börner, T.; Aleynikova, A.Y.; Zubo, Y.O.; Kusnetsov, V.V. Chloroplast RNA polymerases: Role in chloroplast biogenesis. *Biochim. Biophys. Acta* **2015**, *1847*, 761–769. [[CrossRef](#)]
18. Pfannschmidt, T.; Blanvillain, R.; Merendino, L.; Courtois, F.; Chevalier, F.; Liebers, M.; Grübler, B.; Hommel, E.; Lerbs-Mache, S. Plastid RNA polymerases: Orchestration of enzymes with different evolutionary origins controls chloroplast biogenesis during the plant life cycle. *J. Exp. Bot.* **2015**, *66*, 6957–6973. [[CrossRef](#)]
19. Pfannschmidt, T.; Link, G. Separation of two classes of plastid DNA-dependent RNA polymerases that are differentially expressed in mustard (*Sinapis alba* L.) seedlings. *Plant Mol. Biol.* **1994**, *25*, 69–81. [[CrossRef](#)]
20. Pfannschmidt, T.; Ogrzewalla, K.; Baginsky, S.; Sickmann, A.; Meyer, H.E.; Link, G. The multisubunit chloroplast RNA polymerase A from mustard (*Sinapis alba* L.). Integration of a prokaryotic core into a larger complex with organelle-specific functions. *Eur. J. Biochem.* **2000**, *267*, 253–261. [[CrossRef](#)]
21. Yagi, Y.; Shiina, T. Recent advances in the study of chloroplast gene expression and its evolution. *Front. Plant Sci.* **2014**, *5*, 61. [[CrossRef](#)] [[PubMed](#)]
22. Suzuki, J.Y.; Ytterberg, A.J.; Beardslee, T.A.; Allison, L.A.; Wijk, K.J.; Maliga, P. Affinity purification of the tobacco plastid RNA polymerase and in vitro reconstitution of the holoenzyme. *Plant J.* **2004**, *40*, 164–172. [[CrossRef](#)] [[PubMed](#)]
23. Steiner, S.; Schröter, Y.; Pfalz, J.; Pfannschmidt, T. Identification of essential subunits in the plastid-encoded RNA polymerase complex reveals building blocks for proper plastid development. *Plant Physiol.* **2011**, *157*, 1043–1055. [[CrossRef](#)] [[PubMed](#)]
24. Myouga, F.; Hosoda, C.; Umezawa, T.; Iizumi, H.; Kuromori, T.; Motohashi, R.; Shono, Y.; Nagata, N.; Ikeuchi, M.; Shinozaki, K. A heterocomplex of iron superoxide dismutases defends chloroplast nucleoids against oxidative stress and is essential for chloroplast development in Arabidopsis. *Plant Cell* **2008**, *20*, 3148–3162. [[CrossRef](#)]
25. Garcia, M.; Myouga, F.; Takechi, K.; Sato, H.; Nabeshima, K.; Nagata, N.; Takio, S.; Shinozaki, K.; Takano, H. An Arabidopsis homolog of the bacterial peptidoglycan synthesis enzyme MurE has an essential role in chloroplast development. *Plant J.* **2008**, *53*, 924–934. [[CrossRef](#)]
26. Arsova, B.; Hoja, U.; Wimmelbacher, M.; Greiner, E.; Ustün, S.; Melzer, M.; Petersen, K.; Lein, W.; Börnke, F. 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* **2010**, *22*, 1498–1515. [[CrossRef](#)]

27. Chen, M.; Galvão, R.M.; Li, M.; Burger, B.; Bugea, J.; Bolado, J.; Chory, J. *Arabidopsis* HEMERA/pTAC12 initiates photomorphogenesis by phytochromes. *Cell* **2010**, *141*, 1230–1240. [[CrossRef](#)]
28. Gao, Z.P.; Yu, Q.B.; Zhao, T.T.; Ma, Q.; Chen, G.X.; Yang, Z.N. A functional component of the transcriptionally active chromosome complex, *Arabidopsis* pTAC14, interacts with pTAC12/HEMERA and regulates plastid gene expression. *Plant Physiol.* **2011**, *157*, 1733–1745. [[CrossRef](#)]
29. Yagi, Y.; Ishizaki, Y.; Nakahira, Y.; Tozawa, Y.; Shiina, T. Eukaryotic-type plastid nucleoid protein pTAC3 is essential for transcription by the bacterial-type plastid RNA polymerase. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 7541–7546. [[CrossRef](#)]
30. Yu, Q.B.; Lu, Y.; Ma, Q.; Zhao, T.T.; Huang, C.; Zhao, H.F.; Zhang, X.L.; Lv, R.H.; Yang, Z.N. 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.* **2013**, *148*, 408–421. [[CrossRef](#)]
31. Yua, Q.B.; Ma, Q.; Kong, M.M.; Zhao, T.T.; Zhang, X.L.; Zhou, Q.; Huang, C.; Chong, K.; Yang, Z.N. AtECB1/MRL7, a thioredoxin-like fold protein with disulfide reductase activity, regulates chloroplast gene expression and chloroplast biogenesis in *Arabidopsis thaliana*. *Mol. Plant* **2014**, *7*, 206–217. [[CrossRef](#)] [[PubMed](#)]
32. Favier, A.; Gans, P.; Boeri Erba, E.; Signor, L.; Muthukumar, S.S.; Pfannschmidt, T.; Blanvillain, R.; Cobessi, D. The plastid-encoded RNA polymerase-associated protein PAP9 is a superoxide dismutase with unusual structural features. *Front. Plant Sci.* **2021**, *12*, 668897. [[CrossRef](#)]
33. Chieb, M.; Liebers, M.; Chevalier, F.; Lerbs-Mache, S.; Blanvillain, R.; Pfannschmidt, T. Determination of the DNA/RNA-associated Subproteome from Chloroplasts and other plastid types. *Methods Mol. Biol.* **2018**, *1829*, 253–271. [[CrossRef](#)]
34. Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global quantification of mammalian gene expression control. *Nature* **2011**, *473*, 337–342. [[CrossRef](#)] [[PubMed](#)]
35. Pfalz, J.; Liere, K.; Kandlbinder, A.; Dietz, K.J.; Oelmüller, R. pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell* **2006**, *18*, 176–197. [[CrossRef](#)] [[PubMed](#)]
36. Finet, C.; Timme, R.E.; Delwiche, C.F.; Marlétaz, F. Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. *Curr. Biol.* **2010**, *20*, 2217–2222. [[CrossRef](#)] [[PubMed](#)]
37. Sutherland, C.; Murakami, K.S. An introduction to the structure and function of the catalytic core enzyme of *Escherichia coli* RNA polymerase. *EcoSal Plus* **2018**, *8*. [[CrossRef](#)]
38. Opalka, N.; Brown, J.; Lane, W.J.; Twist, K.A.; Landick, R.; Asturias, F.J.; Darst, S.A. Complete structural model of *Escherichia coli* RNA polymerase from a hybrid approach. *PLoS Biol.* **2010**, *8*, e1000483. [[CrossRef](#)]
39. Lane, W.J.; Darst, S.A. Molecular evolution of multisubunit RNA polymerases: Structural analysis. *J. Mol. Biol.* **2010**, *395*, 686–704. [[CrossRef](#)]
40. Glyde, R.; Ye, F.; Jovanovic, M.; Kotta-Loizou, I.; Buck, M.; Zhang, X. Structures of bacterial RNA polymerase complexes reveal the mechanism of DNA loading and transcription initiation. *Mol. Cell* **2018**, *70*, 1111–1120. [[CrossRef](#)]
41. Lin, W.; Das, K.; Degen, D.; Mazumder, A.; Duchi, D.; Wang, D.; Ebright, Y.W.; Ebright, R.Y.; Sineva, E.; Gigliotti, M.; et al. Structural basis of transcription inhibition by fidaxomicin (Lipiarmycin A3). *Mol. Cell* **2018**, *70*, 60–71. [[CrossRef](#)] [[PubMed](#)]
42. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)] [[PubMed](#)]
43. Gilkerson, J.; Perez-Ruiz, J.M.; Chory, J.; Callis, J. The plastid-localized pfkB-type carbohydrate kinases FRUCTOKINASE-LIKE 1 and 2 are essential for growth and development of *Arabidopsis thaliana*. *BMC Plant Biol.* **2012**, *12*, 102. [[CrossRef](#)]
44. Riggs, J.W.; Callis, J. *Arabidopsis* fructokinase-like protein associations are regulated by ATP. *Biochem. J.* **2017**, *474*, 1789–1801. [[CrossRef](#)] [[PubMed](#)]
45. Schneider, G.J.; Hasekorn, R. RNA polymerase subunit homology among cyanobacteria, other eubacteria and archaeobacteria. *J. Bacteriol.* **1988**, *170*, 4136–4140. [[CrossRef](#)]
46. Nevarez, P.A.; Qiu, Y.; Inoue, H.; Yoo, C.Y.; Benfey, P.N.; Schnell, D.J.; Chen, M. Mechanism of dual targeting of the phytochrome signaling component HEMERA/pTAC12 to plastids and the nucleus. *Plant Physiol.* **2017**, *173*, 1953–1966. [[CrossRef](#)]
47. Liebers, M.; Gillet, F.X.; Israel, A.; Pounot, K.; Chambon, L.; Chieb M;Chevalier, F.; Ruedas, R.; Favier, A.; Gans, P.; Boeri Erba, E.; et al. Nucleo-plastidic PAP8/pTAC6 couples chloroplast formation with photomorphogenesis. *EMBO J.* **2020**, *39*, e104941. [[CrossRef](#)]
48. Chambon, L.; Gillet, F.X.; Chieb, M.; Cobessi, D.; Pfannschmidt, T.; Blanvillain, R. PAP8/pTAC6 is part of a nuclear protein complex and displays RNA recognition motifs of viral origin. *Int. J. Mol. Sci.* **2022**, *23*, 3059. [[CrossRef](#)]
49. Casabona, M.G.; Vandenbrouck, Y.; Attree, I.; Couté, Y. Proteomic characterization of *Pseudomonas aeruginosa* PAO1 inner membrane. *Proteomics* **2013**, *13*, 2419–2423. [[CrossRef](#)]
50. Bouyssié, D.; Hesse, A.M.; Mouton-Barbosa, E.; Rompais, M.; Macron, C.; Carapito, C.; Gonzalez de Peredo, A.; Couté, Y.; Dupierris, V.; Burel, A.; et al. Proline: An efficient and user-friendly software suite for large-scale proteomics. *Bioinformatics* **2020**, *36*, 3148–3155. [[CrossRef](#)]
51. Iacobucci, C.; Götze, M.; Ihling, C.; Piotrowski, C.; Arlt, C.; Schäfer MHage, C.; Schmidt, R.; Sinz, A. A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions. *Nat. Protoc.* **2018**, *13*, 2864–2889. [[CrossRef](#)] [[PubMed](#)]
52. Scheres, S.H.W. RELION: Implementation of a bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **2012**, *180*, 519–530. [[CrossRef](#)] [[PubMed](#)]



53. Rohou, A.; Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **2015**, *192*, 216–221. [[CrossRef](#)] [[PubMed](#)]
54. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D.J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **2019**, *47*, D442–D450. [[CrossRef](#)] [[PubMed](#)]