Investigation of the redox signaling involved in the chloroplast biogenesis.

Soumiya Sankari Muthukumar

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : CSV- Chimie et Sciences du Vivant
Spécialité : Biologie Structurale et Nanobiologie
Unité de recherche : Institut de Biologie Structurale

Etude de la signalisation redox impliquée dans la biogenèse du chloroplaste

Investigation of the redox signaling involved in the chloroplast biogenesis.

Présentée par :

SOUMIYA SANKARI MUTHUKUMAR

Direction de thèse :

David COBESSI Université Grenoble Alpes	Directeur de thèse
Robert BLANVILLAIN Université Grenoble Alpes	Co-directeur de thèse

Rapporteurs :

Julio SAEZ-VASQUEZ
DIRECTEUR DE RECHERCHE, CNRS Occitanie Ouest
Arnaud POTERZMAN
DIRECTEUR DE RECHERCHE, CNRS Alsace

Thèse soutenue publiquement le **7 décembre 2023**, devant le jury composé de :

Catherine CORBIER PROFESSEURE DES UNIVERSITES, Université de Lorraine École Nationale Supérieure d'Agronomie et des Industries Alimentaires (ENSAIA)	Présidente
Annabelle VARROT DIRECTRICE DE RECHERCHE, CNRS Alpes	Examinatrice
Eve de ROSNY MAITRESSE DE CONFERENCES, Université Grenoble Alpes	Examinatrice
Julio SAEZ-VASQUEZ DIRECTEUR DE RECHERCHE, CNRS Occitanie Ouest	Rapporteur
Arnaud POTERZMAN DIRECTEUR DE RECHERCHE, CNRS Alsace	Rapporteur

Table of Contents

Acknowledgements

In the journey of my life, I would not have reached my current position, nor would I be able to move forward, without the unwavering support and encouragement of my father, Muthukumar Thangavel, my mother, Theivathai Vimala Muthaiah, and my sister, Sindhu Shivani Muthukumar. Their belief in my potential, and their unyielding support have been the greatest gifts I've ever received. I am profoundly grateful for everything they have done and for believing in me. I also thank my aunt and uncle for their constant support and motivation.

I thank my supervisor, Dr. David Cobessi for his supervision for three years. I also thank my co-supervisor, Dr. Robert Blanvillain, for their unwavering support and guidance throughout my doctoral journey. Their consistent mentorship, encouragement, and counsel have been invaluable throughout my entire Ph.D. tenure.

I would also like to express my appreciation to the jury of my thesis defence committee for accepting the responsibility of evaluating my work. I thank my CSI members, Dr. Annabelle Varrot, Dr. Catherine Bougault and Dr. Gabrielle Tichtinsky, for their constructive feedback and guidance.

I am grateful to Dr. Christel Carles and Dr. Gilles Vachon, who provided unwavering support and invaluable insights during our lab presentations and meetings. I also like to thank the support I received from the members of LPCV, Grenoble, IBS-GSY and Metallo that enabled me complete my experiments. I am grateful for their support and troubleshooting guidance. I am also grateful for the technical assistance and troubleshooting guidance provided by our collaborators from IBS, namely Dr. Gregory Effantin, Dr. Elisabetta-Erba Boeri, and Dr. Caroline Mas.

Marie, I’ve always been inspired by your work ethics, calmness and patience that you possess and thank you for your cheerful approach and kindness towards me. Anne-Marie thank you for being supportive and for your help whenever I needed. Kimi, your help and assistance when I began my journey in Grenoble was welcoming and warm. Your professional guidance was helpful and motivating. I would like to thank Leon Jenner and Elham for their support and guidance during emergencies and for administration works right from my first year to till date.

Kateryna, I always admired your ethics and thank you for your advices and guidance. Your courage and determination were always inspirational to me.

Lucia, we began and will be ending our PhD together. I had a great time with you sharing our office space together. Thank you for your constant motivation, encouragement and your presence. I thank Jean-Baptiste and Vangeli for their support and motivation to keep moving forward, especially for the past few months during the thesis writing. Borys, I am grateful for your cheerful support, constant presence and laughter that lightens up difficult times. Quentin, you always make me laugh and smile through happiest and challenging times since this year and I will always be grateful to you for that. Your words of belief and encouragement were uplifting when I needed to hear the most. Thank you for being there for me. I also thank everyone who traversed my life for creating a meaningful and adventurous impact, teaching many lessons to learn and unlearn.

Abbreviations

ATP	Adenosine triphosphate
BiFc	Bimolecular fluorescence complementation assay
BS3	Bissulfosuccinimidyl suberate
CC	Correlation co-efficient
Co-IP	Co-immuno precipitation assay
CRY	Cryptochrome
Cryo-EM	Cryogenic electron microscopy
CSP41a	Chloroplastic stem-loop binding protein a
CSP41b	Chloropalstic stem loop binding protein b
CTF	Contrast transfer function
cTP	chloroplast transit peptide
DED	Direct electron detector
DIC	Differential interference contrast
DNA	Deoxyribonucleic acid
DQE	Detective quantum efficiency
DTT	Dithiothreitol
EM	Electron microscopy
EMSA	Electrophoretic mobility shift assay
EXAFS	Extended x-ray absorption fine structure
FEG	Field emission gun
FLN2	Fructokinase-like protein 2
FSC	Fourier shell correlation
FSD3	Fe (iron) superoxide dismutase 3
GFP	Green fluorescent protein
GLKs	Golden-2-likes
HY5	Elongated hypocotyl 5
ICP-MS	Inductively coupled plasma mass spectrometry
ITC	Isothermal titration calorimetry
K_a	Association constant

K_d	Dissociation constant
LHC	Light harvesting complex
mRNA	messenger ribonucleic acid
MS	Mass spectrometry
NAD	Nicotinamide adenosine diphosphate
NADP	Nicotinamide adenosine dinucleotide phosphate
NEP	Nuclear-encoded plastid RNA polymerase
NHS	N-hydroxysulfosuccinimide
NMR	Nuclear magnetic resonance
NY	Yellow fluorescent protein N-terminal
PAPs	PEP-associated proteins
PDB	Protein data bank
PEP	Plastid-encoded RNA polymerase
PET	Photosynthetic electron transport
PhANGs	Photosynthesis associated nuclear genes
PhAPGs	Photosynthesis associated plastid genes
PHYB	Phytochrome B
PIF	Phytochrome interacting factors
PL	Proximity labelling
PRIN2	Plastid redox insensitive 2
PSI	Photosystem I
PSII	Photosystem II
RFP	Red fluorescent protein
RNA	Ribonucleic acid
ROS	Reactive oxygen species
rpo	RNA polymerase genes
SDS-PAGE	Sodium dodecyl sulphate – poly acrylamide gel electrophoresis
SIG	Sigma factors
SOD	Superoxide dismutase
SPA	Single particle analysis

SST	Sodium silico tungstate
TAC	Transcriptionally active chromosome
TADs	Transcriptionally active domain
TAE	Tris-acetic acid EDTA
TbID	Turbo ID
TEM	Transmission electron microscopy
tRNA	transfer ribonucleic acid
Trx	Thioredoxin
UDP	Uridine diphosphate
UV-A	Ultra-violet A
XANES	X-ray absorption near edge spectroscopy
YC	Yellow fluorescent protein C-terminal
YFP	Yellow fluorescent protein

Bases in nucleic acids (single letter abbreviations)

A	Adenine
C	Cytosine
G	Guanine
T	Thymine
U	Uracil

Amino acid residues (single letter and three letter) abbreviations

A, Ala	Alanine
C, Cys	Cysteine
D, Asp	Aspartate
E, Glu	Glutamate
F, Phe	Phenylalanine
G, Gly	Glycine
H, His	Histidine
I, Ile	Isoleucine
K, Lys	Lysine
L, Leu	Leucine

M, Met	Methionine
N, Asn	Aspargine
P, Pro	Proline
Q, Gln	Glutamine
R, Arg	Arginine
S, Ser	Serine
T, Thr	Threonine
V, Val	Valine
W, Trp	Tryptophan
Y, Tyr	Tyrosine

Units

°C	celsius
µL	microlitre
µm	micrometer
µM	micromolar
Å	angstrom
Da	dalton
nm	nanometre
MDa	mega Dalton
ms	milliseconds
mg	milligram
mL	millilitre
mM	millimolar
Kb	kilo bases
kDa	kilo Dalton
kV	kilo volt

List of figures

List of tables

Résumé en Français

Au cours de la biogenèse du chloroplaste, l'assemblage de l'appareil photosynthétique est soumis à des régulations transcriptionnelles se produisant dans le noyau et les plastes. En particulier, l'activité de l'ARN polymérase du plaste (PEP) est couplée à la transcription nucléaire pour l'expression coordonnée des gènes associés à la photosynthèse dans le plaste (*PhAPGs*) et dans le noyau (*PhANGs*). Après induction par la lumière, les *PhAPGs* sont spécifiquement transcrits par la PEP dont le cœur catalytique, composé des quatre sous-unités (α, β, β', β''), est alors associé à 12 protéines (PAPs) essentielles à la biogenèse du chloroplaste, ainsi qu'à d'autres protéines régulant la transcription. Parmi les PAPs, PAP4/FSD2 et PAP9/FSD3 sont des superoxyde dismutases à fer qui protègeraient la PEP des espèces réactives de l'oxygène (ROS) produites dans le chloroplaste lors des premières réactions photosynthétiques. Une stratégie de marquage de proximité de PAP8 a permis l'identification d'une autre protéine charnière du redox, PRIN2 (Plastid Redox Insensitive 2), décrite comme interagissant avec la thiorédoxine PAP10 et la protéine de liaison à l'ARN CSP41b (Chloroplast Stem-loop binding Protein). Toutes ces protéines sont essentielles à l'activité de la PEP comme en témoignent le phénotype albinos des mutants PAPs, déficients pour la photosynthèse. Cette thèse par l'étude de l'activité superoxyde dismutase des protéines purifiées PAP4 et PAP9 pose la question de leur rôle structural ou catalytique au sein de la PEP. L'interaction entre PRIN2 et CSP41b est étudiée par des techniques biophysiques telles que la chromatographie d'exclusion, la calorimétrie isotherme, la spectrométrie de masse et la cryo-EM. La structure de CSP41b a été déterminée par cryo-EM à une résolution de 3,4 Å. L'interaction entre PRIN2 et CSP41b, et autres partenaires de la PEP, ont été testées sur des cellules d'épiderme d'oignon à l'aide de la complémentation de fluorescence bimoléculaire. Une stratégie d'étiquetage de proximité a été conçue pour l'identification *in planta* des interactions de PRIN2. Les constructions génétiques ont ainsi été clonées et testées en expression transitoire. Cette étude s'inscrit dans un projet plus large qui vise à mettre en lumière les innovations fonctionnelles chez les angiospermes autour du contrôle de la transcription chloroplastique.

Abstract in English

During chloroplast biogenesis, the assembly of the photosynthetic apparatus is under transcriptional regulations occurring in the nucleus and plastids. Yet, the activity of the Plastid-encoded RNA Polymerase (PEP) is coupled to nuclear transcription for the coordinated expression of Photosynthesis-associated-plastid genes (*PhAPGs*) and Photosynthesis-associated-nuclear genes (*PhANGs*). The *PhAPGs* are specifically transcribed by the light-activated PEP complex that comprises four catalytic subunits (α, β, β', β''), 12-PEP-associated proteins (PAPs) and fleeting interactors, such as redox-associated proteins. Among the 12 PAPs, PAP4 (*FSD2*) and PAP9 (*FSD3*) are Fe-superoxide dismutases found only in chloroplasts and protect the PEP complex from oxidative stress. These proteins are essential for facing the surge in reactive oxygen species (ROS) that occur during the first photosynthetic reactions. A strategy of PAP8 proximity labelling led to the identification of another pivotal redox protein PRIN2 (Plastid Redox Insensitive 2) previously reported to interact with the thioredoxin PAP10 and the RNA-binding protein CSP41b (Chloroplast Stem-loop binding Protein). All these proteins are essential to the PEP activity as attested by their photosynthetically-deficient mutant phenotypes. The thesis investigates the superoxide dismutase activity of purified PAP4 and PAP9 proteins. The question of PAP4 and PAP9's role to be structural or catalytic in the PEP complex was aimed to be answered. The structure of CSP41b was characterised by cryo-EM at 3.4 Å resolution. The study aimed to identify the interactions between PRIN2 and CSP41b by biophysical techniques such as size-exclusion chromatography, isothermal titration calorimetry, mass spectrometry and cryo-EM. This interaction and others were tested in onion epidermal cells using bimolecular fluorescence complementation assay. For the unbiased fishing of PRIN2 interactors, a proximity labelling strategy was designed. The genetic constructions were cloned and tested in transient experiments proving its feasibility. The study presented here is part of a broader project that aims to highlight the functional innovations around plastid transcription specific to angiosperms.

1. INTRODUCTION

1.1 Biogenesis of chloroplast

The chloroplast originated about 1.5 billion years ago. It evolved from an endosymbiotic relationship between a single-celled eukaryote containing a mitochondrion and an engulfed cyanobacterium that give rise to a novel organelle (Dyall et al., 2004). This evolutionary success of the chloroplast is considered to be the key driver behind the ancient Earth's "Great Oxygenation Event." The modern green lineage (viridiplantae) can be traced back to this initial aquatic photosynthetic eukaryote in the form of a green alga, and it is regarded as monophyletic due to the detection of a single instance of cyanobacterial endosymbiosis. Subsequent endosymbiotic events led to the development of more complex organisms, such as photosynthetic red algae or apicomplexa. Approximately 450 million years ago, during the course of evolution, the green lineage colonized terrestrial environments, giving rise to land plants capable of thriving in oxygen-rich atmospheres and dealing with water scarcity. These chloroplasts, stemming from the engulfed cyanobacteria, continue to play a central role in photosynthetic reactions. The chloroplast underwent myriad changes, losing some original functions and gaining new, previously lacking, features, which are well illustrated in the different plastid types seen in the different organs of higher plants despite a small degree of genetic autonomy (Timmis et al., 2004).

Meristematic cells contain colourless proplastids between 0.2 µm and 1 µm. It has very limited internal membrane vesicles that appear as invaginations of the inner envelope. The embryo and the cell types that are not specialised metabolically also contain ten to twenty non-photosynthetic proplastids per cell (Pyke & Leech, 1992; Waters and Pyke, 2004). Other examples are amyloplasts that are differentiated starch-storing plastids in root cells (Neuhaus & Emes, 2000); leucoplasts are lipid-storing plastids found in bulbs and seeds; and chromoplasts are plastids that accumulate pigments, primarily isoprenoids, carotenoids and xanthophylls (Weston & Pyke, 1999).

The size of the chloroplast varies between 5 and 10 µm in diameter and 3-4 µm in thickness. (Waters et al., 2004). Chloroplasts are composed of six distinct sub-organellar structures that are described as:

1. Inner and outer envelope membranes,
2. Internal thylakoid membrane,

3. Three discrete aqueous compartments (the intermembrane space of the envelope, the stroma and the thylakoid lumen) (Keegstra & Cline, 1999; Jarvis & Robinson, 2004; Gutensohn et al., 2006).

The thylakoids are flat vesicles that extend parallel to the central chloroplast axis. They also appear individually as stromal thylakoids or organised in stacks/grana with a lumen. The stroma is the site of carbon fixation. The thylakoid membranes harbour four main protein-pigment complexes involved in the photosynthetic electron transport chain: PSI, PSII, cytochrome b6/f complex, and ATP synthase (Waters & Langdale, 2009). PSII and its leading light-harvesting complex (LHC) are limited to the granular membrane and are not in contact with the stroma. PSI is present exclusively in the stroma-exposed thylakoids. This allows redistribution of light as per prevailing light conditions (Anderson, 2000). Chloroplasts also contain varying quantities of large starch granules and tiny lipid droplets called plastoglobuli. The chloroplast envelopes and thylakoids have different lipid compositions from the rest of the cell membranes, primarily galactolipids instead of phospholipids.

Chloroplasts possess their own genetic material (plastome) and are genetically semi-autonomous (Börner et al., 2015). The plastome is a circular DNA of about 120 – 160 kb in size that encodes approximately 120 transcripts corresponding to components of the photosynthetic apparatus, protein subunits of the ribosomes and the catalytic subunits of RNA polymerase, ribosomal rRNAs and tRNAs. It is present in several copies associated in compact structures called nucleoids (Bock, 2007; Sugiura, 1992). As chloroplasts mature, these nucleoids are relocated from the inner envelope to the thylakoid membranes (Powikrowska et al., 2014). However, from proteomic analysis, 2,500 – 3,500 different proteins have been identified within the chloroplasts that far exceed their coding capacity (Ferro et al., 2010; Zychlinski et al., 2005; Zybailov et al., 2008). Thus, most of the chloroplast proteins are nuclear encoded proteins that are imported into the chloroplast.

1.1.1. The transition from etioplast to chloroplast in angiosperms

The molecular mechanisms governing chloroplast biogenesis remain elusive due to their rapidity and intricacy (Pogson et al., 2015). Perception of light is essential in angiosperms for differentiating chloroplasts from chlorophyll-free proplastids.

When the seedling is buried in the soil at dark, germination is followed with hypocotyl elongation, leading to **skotomorphogenesis** (Solymosi & Schoefs, 2010).

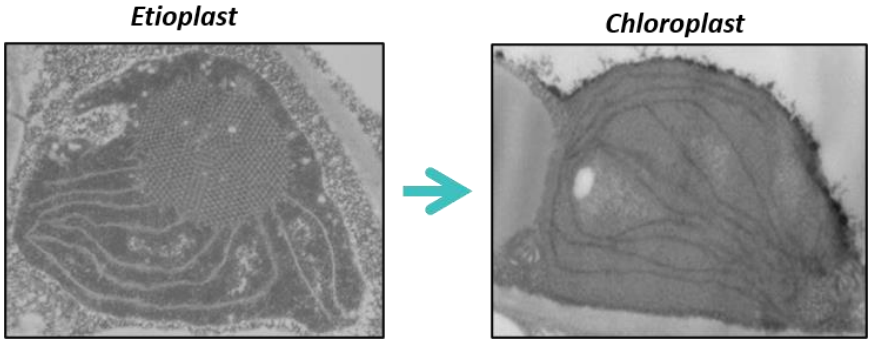


Figure 1.1: *Higher magnification electron micrographs displaying etioplast to chloroplast transition. Source: Martin et al., 2016; Pogson and Albrecht, 2011. A typical etioplast displays a semi-crystalline structure named a prolamellar body on which are inserted pro-thylakoids whereas a young photosynthetic chloroplast rapidly displays a network of thylakoids not yet connected by grana stacks.*

The elongated hypocotyl permits the shoot apex to reach the surface of the soil and receive the sunlight. The apical hook directs the non-developing cotyledons downwards and thus preserves the quiescent shoot apical meristem while the stem is growing and protrudes from the soil (Liebers *et al.*, 2017). During the dark growth, the cotyledons remain yellowish and minuscule with no expansion. At the cellular level in the etiolated seedlings, the proplastids develop into yellow etioplasts, an intermediary stage of the plastid, incapable of performing photosynthesis. In angiosperms, as soon as the germinating seedling receives light, these yellow etioplasts quickly develop into chloroplasts within a few hours (Figure 1.1) (Dubreuil et al., 2018; Armarego-Marriott et al., 2020; Armarego-Marriott et al., 2019; Pogson et al., 2015). Significant morphological changes known as photomorphogenesis occur in the seedling, such as repression of the hypocotyl elongation and opening of the cotyledons after light perception (Pogson *et al.*, 2015). Upon light perception by photoreceptors such as phytochrome B (PHYB) capturing red/far-red light or cryptochrome (CRY) capturing blue/UV-A light, greening occurs. In response to the perception of their activating photons, the photoreceptors translocate from the cytoplasm into the nucleus and regroup in photobodies, which are subnuclear membrane-less compartments coalescing due to liquid-liquid phase separation (Mo et al., 2002; Wang et al., 2021). The presence of photobodies initiates a transition from proplastids to chloroplasts and the assembly of the photosynthetic machinery (PS). Chloroplast biogenesis, however,

results not only from environmental control over nuclear gene expression but also from a significant restructuring of plastid gene expression, leading to an overall increase in transcriptional activity favouring genes associated with photosynthes is in plastids *(PhAPGs)* and genes associated with photosynthesis in the cell nucleus *(PhANGs)*. This shift in gene expression is crucial, as the photosynthetic apparatus (PS) consists of multi-protein complexes synthesized following transcription of both *PhAPGs* and *PhANGs*. The perception of the first rays of light triggers a precise and strong coordination of *PhAPGs* and *PhANGs* transcription involving anterograde signals from nucleus to plastids, retrograde signals from plastids to the nucleus, as well as protein trafficking between the organelles and redox signals mainly related to the build-up of the PS into the thylakoid membrane.

1.1.2. Signalling between nucleus and plastid

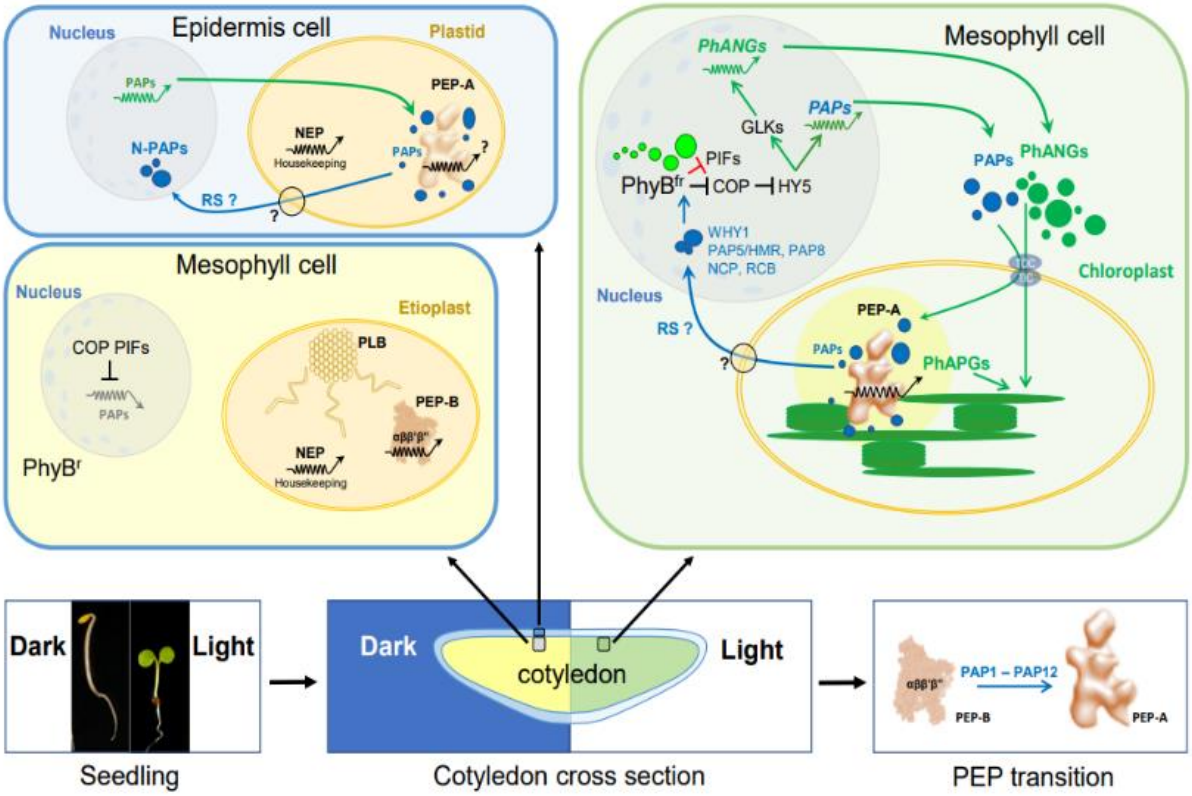


Figure 1.2: *Retrograde signalling by dually localized proteins during de-etiolation. The scheme depicts an overview of the proposed spatial-developmental actions of nucleo-plastidic proteins in retrograde signalling. The left box illustrates the skoto- to photomorphogenesis of* Arabidopsis *seedlings. The middle box illustrates the schematic cross-section through the respective cotyledons that develop either etiolated or green mesophyll cells, respectively. The working model involves the light-induced rearrangement of subunits within the PEP complex. PEP-B represents the* E. coli*-like core enzyme. It is given as the simplified crystallographic structure of the* E. coli *RNA polymerase. The addition of PAPs converts PEP-B into a structurally larger and more complex PEP-A given as a 3D envelope (Ruedas et al., 2022,). Epidermal cells in the dark may contain a fully assembled PEP-A. A potential retrograde signalling (RS) of nuclear-localized PAPs (N-PAPs) in the epidermis is unknown. Mesophyll cells in the dark contain etioplasts with a prolamellar body (PLB) and a PEP-B with basal transcriptional activity. Chloroplast biogenesis is both repressed by COP-mediated protein degradation and PIFs transcriptional activity. After illumination, repression is released, leading to the transcription of PhANGs and PhAPGs. Plastid-imported PhANGs assemble with PhAPGs to build the photosynthetic apparatus. Thylakoids are represented as dark green ovals. Source: Liebers et al. 2022 Darwin Review*

For a long time, it was believed that the nucleus had the sole responsibility of overseeing chloroplast biogenesis by supplying all essential structural elements and protein factors to the organelle. However, it has become increasingly apparent in recent years that the chloroplast also provides vital regulatory signals contributing to this coordination. The two types of regulation are anterograde signalling (describing the nucleus-to-chloroplast signalling) and retrograde signalling (chloroplast-to-nucleus signalling) (Woodson & Chory, 2008). The concept of these signals for chloroplast development was discovered in experiments where the plastid development was chemically or genetically halted. It resulted in a parallel inhibition of expression of the nuclear genes that encoded plastid photosynthetic proteins such as the small RuBisCo subunit or subunits of LHCs of the photosystems (Bradbeer et al., 1979; Oelmüller et al., 1986). During the early steps of chloroplast biogenesis, retrograde signals from plastids have been named biogenic signals (Pogson et al., 2008). The biogenic signals can be distinguished into five classes: pigment precursors or plastid pigments, signals from plastid gene expression, reactive oxygen species that are generated during photosynthesis, photosynthesis-related redox signals and changes in metabolite levels (Liebers et al., 2022).

In response to light exposure, the expression of approximately 30% of all nuclear-encoded genes is changed as compared to dark-grown seedlings. The most prominently upregulated genes constitute the chloroplast-targeted proteins (Ma et al., 2001). Most of the plastid-localized proteins are encoded in the nucleus, translated by 80S ribosomes in the cytosol as preproteins and imported into the organelle via the TOC-TIC machinery in the outer and inner envelope membranes (Schleiff et al., 2002; Bédard & Jarvis, 2005; Smith, 2006). The preproteins contain a chloroplast transit peptide (cTP) (Kessler & Schnell, 2006). The cTPs have copious amounts of positively charged, hydroxylated small amino acids and a low abundance or absence of acidic, large hydrophobic amino acids. They also have no conspicuous sequence conservation among all the translocated proteins. A phosphorylation site at a serine or threonine residue appears to be a standard feature. Proteins and a chaperone bind this site, termed a guidance complex (May & Soll, 2000). The cTP is cleaved during the import of the preproteins into the plastid.

1.2 Plastid-encoded RNA polymerase and PEP-associated proteins

1.2.1. Nucleus-encoded RNA polymerase and Plastid-encoded RNA polymerase

Chloroplast genomes in vascular plants contain genes with significant sequence similarities to bacterial RNA polymerase (rpo) genes (Igloi and Kossel, 1992). These *rpo* genes are typically organized in a large operon consisting of *rpoB*, *rpoC1*, and *rpoC2*, with another operon containing the single gene *rpoA*, along with several genes for ribosomal components, and encoding the α, β, β′, and β′′ subunits respectively. *RpoC1* and *rpoC2* likely emerged from a split of the original *rpoC* gene in the cyanobacterial ancestor (Green, 2011). In dicotyledonous plants, an intron is found within the *rpoC1* gene, which is absent in monocotyledons, suggesting potential evolutionary constraints and differences in transcription machinery between these plant clades (Igloi and Kossel, 1992). The organization of *rpo* genes in the green algae *Chlamydomonas reinhardtii* consists of monocistronic units dispersed across the plastome, indicating a difference in plastome organization within the green lineage (Maul *et al.*, 2002). The subunit α is presumed to stabilize the complex and may exist as a dimer, similar to its bacterial counterpart. The basic RNA polymerase complex, referred to as the core enzyme with a stoichiometry of α2, β, β′, and β″, is capable of transcriptional elongation *in vitro* and likely *in vivo*. Plastid knockout mutants in tobacco or nuclear knockdown mutants in Arabidopsis lacking functional or structural *rpo* genes exhibit albino or yellowish phenotypes with arrested plastid development, highlighting the essential role of α, β, β′, and β″ subunits in chloroplast development. These mutant plants can survive when grown on sucrose-supplemented medium, indicating that the absence of chloroplast function can be partially compensated for by an external carbon source.

Nuclear-encoded plastid RNA polymerase (NEP) and Plastid-encoded RNA polymerase (PEP) are required to transcribe plastid-encoded genes of higher plants. NEP is a T3-T7 bacteriophage-type polymerase that transcribes housekeeping genes in the plastid (Hedtke et al., 1997). Two different NEP enzymes are present in monocotyledons, namely RPOTp and RPOTm. RPOTp is explicitly located in the plastids. It is the primary RNA polymerase that transcribes NEP-controlled genes such as those encoding the α, β, β′, and β″ subunits of the PEP polymerase, YCF2 and ACCD genes during initial development. RPOTm is located only in the mitochondria (Chang et al., 1999; Ikeda & Gray, 1999; Kusumi et al., 2004). It transcribes the rRNA operon at the Pc promoter during seed imbibition (Courtois et al., 2007). In dicots,

a third NEP enzyme RPOTmp is present in both plastids and mitochondria (Hedtke et al., 2000). Contrary to their difference in target genes, the activity of both the NEP enzymes is highest during the initial days of seed germination (Demarsy et al., 2006).

The promoters recognised by the NEPs are classified as class Ia that is characterised by the presence of a YRTA motif (Y = T or C and R = A or G) upstream of the transcription start site. Class Ib has an additional GAA box, approximately 20 nucleotides upstream of the YRTA motif. Class II does not contain the consensus sequence YRTA (Hübschmann & Börner, 1998; Liere & Maliga, 1999; Weihe & Börner, 1999; Pfannschmidt, 2010). Nevertheless, the factors that mediate promoter recognition and transcription initiation by NEP are not yet characterised.

PEP transcribes over 80% of all plastid genes (Zhelyazkova et al., 2012). Many plastid genes have conserved -35 (TTGACA) and -10 (TATAAT) promoter sequences (Sugiura, 1992). The PEP is a bacterial-type multi-subunit enzyme (Allison et al., 1996). The catalytic core of PEP is made up of α, β, β', and β" subunits. Nuclear-encoded sigma factors (SIGs) are also required along with the catalytic core for PEP promoter specificity (Hanaoka et al., 2003). In *A. thaliana*, six sigma factors have been identified and well-characterised. The mutants of SIG2 and SIG6 exhibited chlorophyll deficiency during the early chloroplast developmental stage. This indicated that they are the most essential SIG factors for chloroplast development. In prokaryotic multimeric RNAPs, there is a common catalytic core of two large subunits, a dimer of α subunits and a monomer of ω subunit (Cramer, 2002; Hirata et al., 2008; Murakami, 2015). In the etiolated mustard seedlings, the PEP was present in its prokaryotic composition (α2, β, β', β'' subunits). Nonetheless, in mature chloroplasts, a more extensive PEP complex featuring additional subunits has been isolated and characterized (Pfannschmidt & Link, 1994). Through biochemical purification and proteomic analysis, the catalytic core of PEP has been identified in association with at least 12 distinct proteins, resulting in an overall molecular mass exceeding 900 kDa (Suzuki et al., 2004; Steiner et al., 2011). In PEP-impaired mutants, the albino phenotype could not be rescued by the activity of NEP despite a yet poorly understood mechanism of compensation in gene expression that remains polymerase specific. Moreover, in transcription activity analysis, it was observed that most of the plastid genes were transcribed without PEP. This suggests that the transcript level generated by NEP

alone is insufficient for initiating photosynthetic activity; NEP alone can transcribe most of the chloroplast genome (Zhelyazkova et al., 2012).

1.2.2. Role of PAPs during chloroplast biogenesis

The PEP complex consisting of the four catalytic subunits is collectively referred to as the PEP-B complex. When PEP-associated proteins, known as PAPs, are introduced along with sigma factors (SIG 1 – 6) and other proteins, they transform the PEP-B complex into a larger and an intricate structural entity known as the PEP-A complex of about 1.1 MDa (Figure 1.3).

From the gel filtration and mass spectrometry analyses of *A. thaliana* samples, 35 proteins (named pTAC1-35) were identified in a DNA-protein complex called transcriptionally active chromosome (TAC) (Pfalz et al., 2006). Among the TAC, a number of proteins were also identified as a smaller complex PEP-PAPs purified using Heparin/Sepharose chromatography (Steiner et al., 2011). An albino or pale green phenotype was observed in most mutants of the ten PAPs originally identified. Then 2 additional PAPs (PAP11/MurE and PAP12/pTAC7) were conceptually added to the PEP-PAP complex regarding the mutant phenotype and its suspected association to the PEP despite their absence in gel filtration analysis. Therefore 12 PAPs proteins are considered essential for PEP transcription activity and are named PAPs (PEP-associated proteins). During the etioplast to chloroplast transition, the PEP is activated by restructuration into PEP-A including the nuclear-encoded PAPs (Yagi et al., 2012). Phylogenetic analyses suggest that the appearance of *PAP* genes is connected to the conquest of land as it seems restricted only to terrestrial plants (Pfalz and Pfannschmidt, 2013). Hence, these genes can serve as an evolutionary marker signifying the emergence of multicellular plants with different functional organs (de Vries et al., 2016). However, these genes have been lost in Gymnosperms and other clades whereas they became essential in all flowering plants. Indeed, the gene inactivation of any PAPs in maize, rice and *A. thaliana* halted proper chloroplast development and led to albino phenotypes (Pfalz & Pfannschmidt, 2013). This suggests that the expression of the twelve PAP genes and the restructuring of the twelve PAPs in the PEP core denote the vital steps in early chloroplast biogenesis of angiosperms.

Table 1: List of PAPs and their functions

Subunit	Id	Protein domain	Loss of function
PAP1	pTAC3	PPR	Albino/ivory
PAP2	pTAC2	PPR	Pale-green
PAP3	pTAC10	S1-like (PDE312)	Albino/ivory
PAP4	FSD3	Iron superoxide dismutase 3	Pale-green
PAP5	pTAC12	HEMERA/RAD23	Albino/ivory
PAP6	FLN1	Fructokinase	Albino/ivory
PAP7	pTAC14	SET domain	Albino/ivory
PAP8	pTAC6	UNKNOWN	Albino/ivory
PAP9	FSD2	Iron superoxide dismutase 2	Pale-green
PAP10	TrxZ	Thioredoxin	Albino/ivory
PAP11	MurE-like	UNKNOWN	Albino
PAP12	pTAC7	UNKNOWN	Albino/ivory

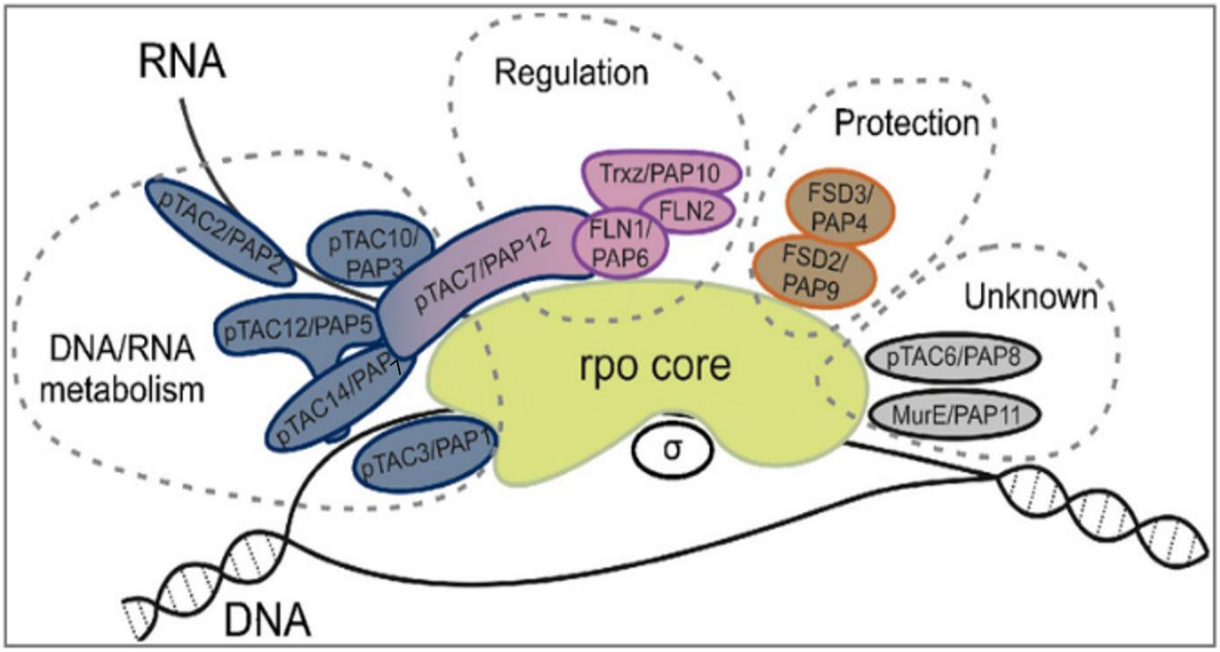


Figure 1.3: *Core of PEP and its necessary proteins for activating transcription of PhAPGs. All the proteins except the rpo core subunits are encoded in nuclear genome, translated in cytosol and are transported to the chloroplast. The three main categories of these proteins, such as DNA/RNA or both-binding, regulatory proteins and proteins that protect the PEP against reactive oxygen species are mentioned in the diagram. pTAC6 and MurE/PAP11 are reported to possess unknown functions. Source: Kindgren & Strand, 2015.*

1.2.3. Interactions of PAPs with PEP

Interactions between some of the PAPs are reported only using non-direct observations, yeast-two hybrid assays (Yu et al., 2013) and fluorescent microscopy (Myouga et al., 2008). It has been reported by (Chang et al., 2017) that PAP3 does not interact directly with the core PEP components. However, it could interact with PAP4, PAP7, PAP9, PAP10 and PAP12 through its carboxyl-terminal region downstream of the S1 domain. PAP5 and PAP8 were observed to be dually located in the nucleus and chloroplast but perform different functions in the respective organelles. Both PAP5 and PAP8 are necessary for the photobodies formation. Upon light exposure, PAP5/HMR is reported to trigger the phytochrome-mediated degradation of PIF1 and PIF3 in the nucleus and trigger transcription of plastid genes in chloroplasts (Qiu et al., 2015; Galvão et al., 2012). PAP5/HMR could directly interact with PAP7 (Gao et al., 2011) and PAP12 (Yu et al., 2013). PAP8 has been reported to be involved in HY5 stabilization, transcription of GLKs and phytochrome-mediated degradation of PIF1 and PIF3 (Liebers et al., 2020) (Figure 1.2). Due to their duality, it is hypothesised that PAP5 and PAP8 are key coordinators between nuclear and chloroplast gene expression during chloroplast biogenesis, which requires further investigation to resolve. Furthermore, interactions between PAP5 and PAP8 were observed by NMR studies (Liebers et al., 2020). PAP4 and PAP9 are reported to be superoxide dismutases that protect the plastid from oxidative stress during the first photosynthetic reactions. It is hypothesised that PAP9 could interact with PAP4 by forming a hetero-complex (Myouga et al., 2008). PAP10 / TrxZ has a thioredoxin domain that could interact with PAP6, FLN2 and PRIN2 (Wimmelbacher & Börnke, 2014; Díaz et al., 2018). Hence, it is suggested that PAP10 is involved in the redox-regulated plastid transcription. The mutants of fructokinase-like proteins FLN1 (PAP6) and FLN2 exhibited albino phenotype and delayed greening (Gilkerson et al., 2012). In *in vitro* studies, they have been reported to form heterodimers or homodimers (Riggs & Callis, 2017). A similar phenotype was observed in the T-DNA insertion lines of PAP12/pTAC7 and PAP11/MurE (MurE-like) despite not being physically linked to PEP in gel filtration and mass spectrometry analysis (Garcia et al., 2008; Pfalz & Pfannschmidt 2013).

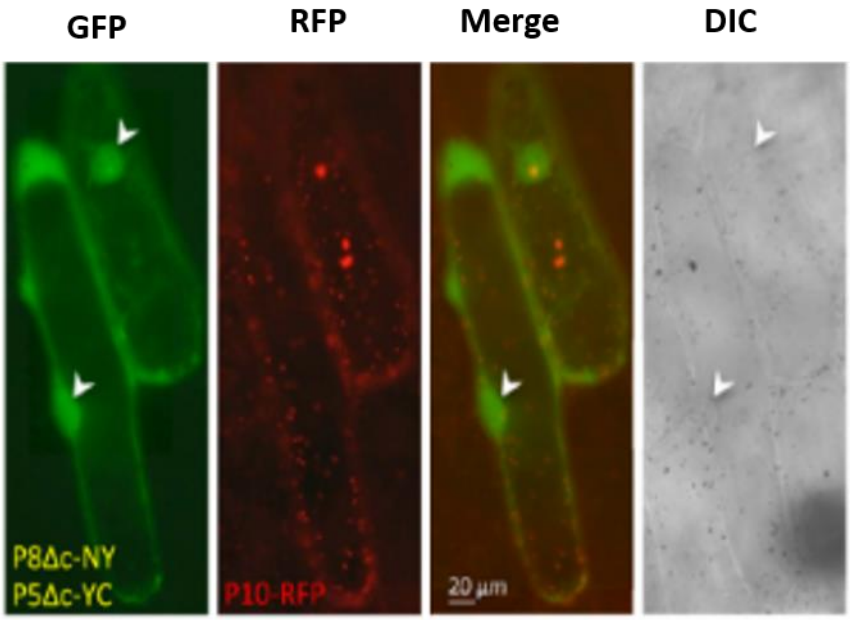


Figure 1.4: *BiFC localisation for PAP8$^{\Delta ctp}$–NY construct with PAP5$^{\Delta ctp}$-YC construct and PAP8$^{\Delta ctp}$ –YC with PAP5$^{\Delta ctp}$ -NY with PAP10-RFP as internal positive control transiently expressed in onion cells. Δc denotes without ctp. Nuclei are indicated by arrowheads as observed with differential interference contrast (DIC). Source: Liebers et al., 2020.*

1.3 Formation of reaction oxygen species during photosynthesis. The role of superoxide dismutases

During the first photosynthetic reactions, a storm of reactive oxygen species such as superoxide (O_2^-), hydrogen peroxide (H_2O_2), hydroxyl radical (OH), and singlet oxygen ($^1O_2$) is produced at the water splitting complex of PSII within the organelle. These reactive oxygen species could disrupt the PEP complex ultimately hindering photosynthesis. Besides its detrimental effect of being a threat to the developmental plastid, research has shown that it has essential physiological roles in plant development. Plants possess well-developed enzymatic and non-enzymatic defence mechanisms against ROS (Alscher et al., 2002). Superoxide radicals can damage sulphur-containing amino acids, metals and Fe-S clusters. Due to the toxic nature of ROS, numerous scavenging enzymes are present in almost every cellular compartment (Kärkönen & Kuchitsu, 2015). The first line of defence against ROS in cells is provided by **superoxide dismutases,** which convert superoxide (O_2^-) and H_2O to hydrogen peroxide (H_2O_2) and molecular oxygen (O_2) (Bowler et al., 1992).

$$O_2^{\cdot-} + O_2^{\cdot-} + 2H^+ \longrightarrow H_2O_2 + O_2$$

SODs protect the plastids from oxidative damage during photosynthesis from oxygen toxicity due to photoreduction of molecular oxygen to O_2^- by electrons from photosystem I (Mehler, 1951). Depending upon the metal cofactor present at its catalytic site, SODs are classified into three types (Figure 1.5), namely:

1. Fe-SOD (Iron Superoxide dismutase)
2. Mn-SOD (Manganese Superoxide dismutase)
3. Copper-Zinc SOD (Cu / ZnSOD).

Fe-SODs are found only in plants and prokaryotes and not in animals. Several conserved regions of Fe-SOD sequences present in cyanobacteria and plants are absent in non-photosynthetic bacteria (Bowler et al., 1994). Proteins that emerged in the later stages of plant evolution have the potential to play a role in regulating various cellular processes (Yagi & Shiina, 2012). Furthermore, they may offer opportunities to enhance the redox regulatory system inherited from cyanobacteria (Balsera et al., 2014). Three genes transcribe Fe-SODs in plants. In *Arabidopsis thaliana* genome,

1. Three Fe-SOD genes - Fe Superoxide dismutase 1 (FSD1, FSD2, FSD3)

2. Three Cu / Zn-SOD are present in plant and localised in cytoplasm (CSD1) and peroxisome (CSD3) and chloroplast (CSD2).
3. An Mn-SOD in mitochondria (Kliebenstein et al., 1998).

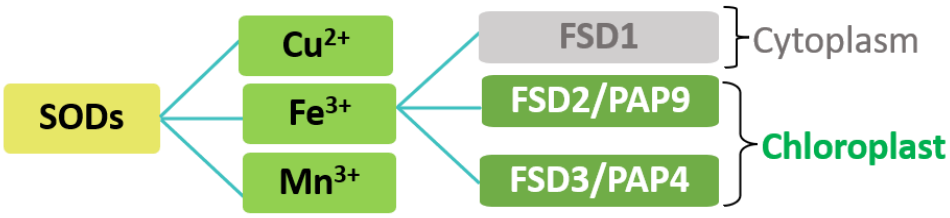


Figure 1.5: *Classification of superoxide dismutases based on their metal cofactor. Adapted from Myouga et al, 2008.*

Among the Fe-SODs, PAP4 (FSD3) and PAP9 (FSD2) are classified with the PAPs. FSD1 protein is reported to be present among chloroplast proteins in peripheral thylakoid (Peltier et al., 2002), stroma (Peltier et al., 2006) and envelope (Ferro et al., 2003) of purified Arabidopsis chloroplasts, as well as in the plasma membrane (Marmagne et al., 2004) in Arabidopsis cell suspension (Brugiere et al., 2004) from proteomic analysis of different plant samples. The study also reported that FSD2 and CSD2 proteins are present specifically in chloroplasts (Kleffmann et al., 2004). The abundance of RuBisCo and other photosynthetic proteins is a limiting factor to this approach, which prevents the detection of less abundant proteins (Baginsky et al., 2005). Therefore, determining the precise subcellular localisation of the three FeSODs is challenging. FeSODs are enzymatic dimers, with each monomer containing one iron ion. The structural configuration of the FeSOD monomer closely resembles that of the MnSOD monomer but differs significantly from Cu/ZnSODs (Pilon et al., 2011). The significant difference between PAP9 and other Fe-SODS is the additional residues in the C-terminal region. Deleting PAP4 and PAP9 genes has been reported to produce albino phenotypes, meaning greening or photosynthesis has not occurred. The double mutant genotype subjected to high light conditions quickly dies from photobleaching. However, in low light conditions, they were able to green again (Myouga et al, 2008). Presence of SOD in PEP led us to investigate whether PAP4 and PAP9 are essential for maintaining the structure of the active PEP complex or whether they have a catalytic role in protecting the PEP complex from reactive oxygen species.

1.4 CSP41b (Chloroplast stem-loop binding protein b) and PRIN2 (Plastid Redox Insensitive 2)

In the transition from dark-to-light growth supporting chloroplast biogenesis, the reshaping of the PEP involves a number of PAPs associated with the redox. Therefore, the activation of PEP entails mechanisms for redox control that have an impact on the expression of plastid genes (Pfannschmidt & Liere, 2005). Yet, the PEP complex is also a significant target of photosynthetic redox signals (Steiner et al., 2009), although the molecular details behind this redox regulation have been elusive. Various theories have been postulated to explain the mechanisms by which redox signals from the photosynthetic electron transport are connected to the expression of plastid genes. Photosynthesis exerts a significant impact on PEP-dependent gene expression within plastids. Various mechanisms have been suggested to connect redox signals originating from photosynthetic electron transport to the regulation of plastid gene expression. Nonetheless, the precise workings of the thiol-mediated pathway responsible for the redox regulation of PEP activity remain undiscovered (Pfannschmidt & Liere, 2005). It is observed that the phosphorylation of SIG1 is affected by the oxidised state of the PQ pool, which leads to the regulation of relative transcription of photosynthetic reaction centre genes psbA (PSII) and psaA/B (PSI). Additionally, the proof that a thio-mediated signal is involved in the redox regulation of PEP components was shown by experiments performed with kinase inhibitors and dithiothreitol (DTT) (Steiner et al., 2009).

Besides PAPs, numerous “fleeting” interactors have been reported to interact with the PEP complex. In the redox-mediated retrograde signalling, a protein called Plastid Redox Insensitive2 (PRIN2) is reportedly involved with PEP interactions. The analysis of plastid transcriptome from Arabidopsis seedlings and rosette plants reported that PRIN2 is vital for the complete expression of *PhANGs*. The *PhANG* expression levels in *prin2* seedlings grown under controlled conditions were lower than in wild-type seedlings. Moreover, the *prin2-1* and *prin2-2* mutants displayed impaired *PhANGs* regulation when photosynthetic electron transport (PET) was inhibited or exposed to excess light. This confirms that PRIN2 is required for nuclear transcription.

When the photosynthetic activity is developed progressively during the greening process and upon activation of photosynthetic electron transport (PET), the reduction of PET components leads to the subsequent reduction of TRX through the FTR system. TRX facilitates the conversion of dimeric PRIN2 into its active monomeric form by reducing a disulfide bond. This

process results in the full initiation of light-activated transcription for *PhANGs*. A retrograde signal tightly links nuclear and chloroplast genomes (Díaz et al., 2018).

The *ys1* mutant, characterised by compromised PEP activity, did not exhibit proper regulation of LHCB1.1 and LHCB2.4 genes under excessive light conditions, resembling the response seen in *prin2* mutants. This indicated that the modifications in PEP activity, rather than PRIN2 solely, are the source of redox-mediated retrograde signals. From experiments performed with PRIN2 and promoter fragments of PEP-dependent *psaA* and NEP-dependent *ycf1*, it is observed that PRIN2 has DNA binding capacity, both alone and as a part of a complex with CSP41b. However, the DNA interaction is reported to be non-specific.

In the photosynthetic eukaryotes, two copies (a and b) of CSP41 (chloroplast stem-loop binding protein of 41 kDa) are reported to exist and are of cyanobacterial origin. CSP41 proteins were also present in PEP-enriched preparations (Pfannschmidt et al., 2000). In a study by (Zybailov et al., 2008), chloroplast stromal proteins were grouped into seven abundance classes and CSP41b was found in the group of highest abundance proteins. The substantial amounts of CSP41 proteins contradict the notion of a specific catalytic role but instead suggest a broader, non-specific function that requires considerably large quantities of the proteins. The reported function of CSP41 proteins is their role in plastid transcription, ribosomal biogenesis, RNase activity and interactions with heteroglycans in the cytosol (Leister, 2014). Also, physical interactions between both CSP41 proteins are reported. The effects observed in leaves of plants deficient in CSP41b suggest that the CSP41a-CSP41b complexes stabilise precursor rRNAs and untranslated mRNAs in a redox-dependent manner. It appears significant during low translational activity prominently in the absence of light. The cellular morphology, photosynthesis and circadian rhythm affect *csp41b* mutants in *A. thaliana* (Hassidim et al., 2007). Nevertheless, the seeds were unviable in mutants lacking *csp41a* and *csp41b* (Beligni & Mayfield, 2008). Notable phenotypic effects were not observed in the *csp41a* mutants compared to *csp41b* mutants (Qi et al., 2012). This could mean that CSP41b is more pivotal than CSP41a. CSP41b is reported to be accumulated predominantly in mature leaves (Fettke et al., 2011). In *A. thaliana* mutants with decreased PEP transcription levels, contrasting behaviour compared to *csp41b* mutant with mature leaves but with young leaves (Chi et al., 2008; Chateigner-Boutin et al., 2011). There is an overlap between the mRNA transcription

levels from sets bound by CSP41 complexes (a and b) or those transcribed by PEP. Hence, it can be inferred that sufficient transcript levels of PEP are present in young leaves and do not require CSP41 complexes. However, in older leaves, the transcript levels depend on CSP41 complexes. It corresponds with the fact that leaf ageing leads to chloroplast transcription stability (Klaff & Gruissem, 1991). It is observed that CSP41 binds with various chloroplast RNAs such *as rbcL, psaA, psaB and PSII* core proteins, 16S and 23S rRNAs (Qi et al., 2012). The stability of two target RNAs was shown to be decreased in mutants lacking CSP41b. The destabilisation of 16S and 23S rRNA precursor forms could lead to lesser functional ribosomes and decreased translation of chloroplast genes. This could lead to decreased synthesis and transcription rate of PEP. However, it's important to note that the direct impact on transcription and translation, resulting from the binding of CSP41 to target transcripts, cannot be definitively excluded at this point.

In a study conducted by Dmitry et al. in 2014, it was revealed that PRIN2 and CSP41b interact directly, forming a complex believed to be vital for PEP-dependent transcription during embryo development. PRIN2 possesses conserved cysteine residues that may participate in its formation as a monomer or dimer. To delve into the details of the study, the researchers employed a co-immunoprecipitation (co-IP) method to identify proteins that interact with PRIN2 within living organisms. During this co-IP experiment, they identified up to 17 protein bands that were absent in the negative control samples. Most of the proteins they identified were exclusive to one experiment and were likely non-specific interactions. However, they found one specific protein, CSP41b, in both experiments. Interestingly, the mutant with impaired CSP41b function displayed a phenotype resembling that of the *prin2* mutant, suggesting a possible functional connection between CSP41b and PRIN2. Mutants with defective CSP41b function have previously exhibited issues related to chloroplast transcription and overall plant development. In an electrophoretic mobility shift assay (EMSA), PRIN2 and CSP41b were found to bind with a labelled DNA probe. PRIN2 formed at least two distinct DNA-protein complexes, while CSP41b formed only one such complex. These proteins demonstrated their ability to interact with DNA *in vitro*, and when combined, they formed a heterodimeric PRIN2/CSP41b complex upon DNA binding (Kremnev et al, 2014). The *prin2.2* and *csp41b-2* mutant plants exhibited distinct phenotypes and demonstrated compromised expression of genes encoded in the chloroplast. Both mutants displayed reduced growth rates,

pale leaves, and abnormalities in chloroplast structures. Additionally, both mutant types displayed defects in embryo development, including altered chloroplast development patterns and paler embryos compared to the wild type. Notably, the *csp41b-2prin2.2* double mutant was found to be lethal to embryos, with impaired ovules and seeds that failed to germinate and were dark in colour (Figure 1.6 A). Transmission Electron Microscopy (TEM) analysis further revealed that chloroplast development was impaired in both single mutants, with fewer thylakoid membranes and changes in plastid structures (Figure 1.6 B). In summary, both PRIN2 and CSP41b are proposed to play roles in regulating the transcription of chloroplast genes dependent on PEP complex.

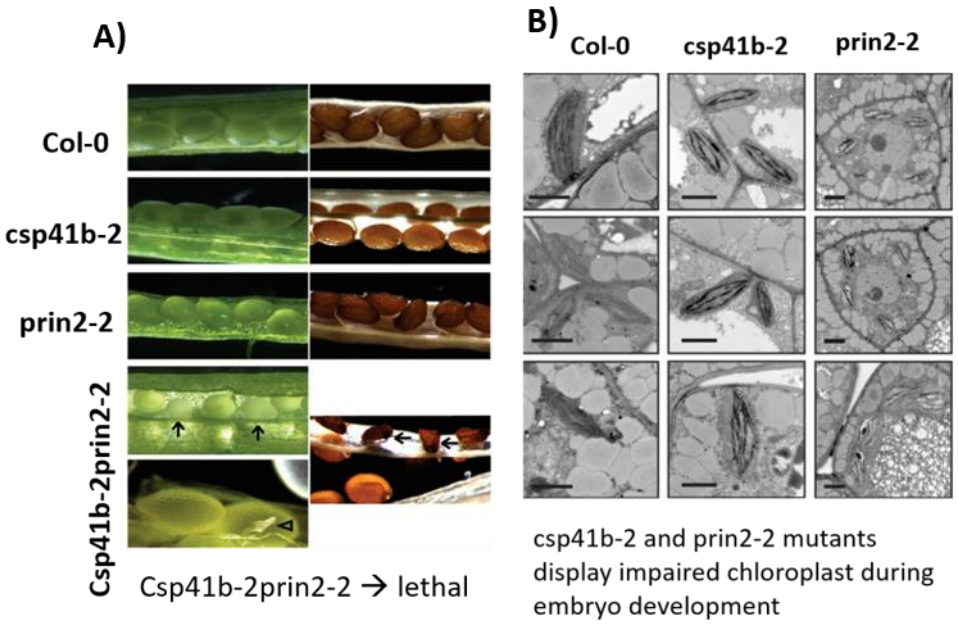


Figure 1.6: *(A) Siliques from self-fertilized CSP41b-2prin2.2/csp41b-2prin2.2 double mutant. (B) the csp41b-2 and 2prin2.2 mutants displaying impaired chloroplast during embryo development. Figures adapted from Kremnev & Strand, 2014.*

Among the proteins of interest found in experiments of proximity labelling with PAP8 performed at the LPCV by Dr. Robert Blanvillain and Dr. François-Xavier Gillet (Figure 1.7 A), CSP41b, PRIN2 and superoxide dismutases were observed. Proximity labelling combined with MS-based quantitative proteomics is used for identifying protein-protein interactions. It could detect weak, transient or hydrophobic protein-protein interactions in their native state. A catalytic enzyme (biotin ligase) such as TurboID is fused to a bait protein (here, PRIN2). TurboID (35 kDa) was developed using yeast-display-based directed evolution of BirA (Branon

et al., 2018). TurboID-based PL has efficient biotinylation at room temperature (25°C). It could identify quickly and dynamically without damaging the living cells (Mair et al., 2019). When external biotin is applied, TurboID converts biotin into diffusible, short-lived activated biotin adenylate intermediates that are transferred to the ε-amino group of surface-exposed lysine residues of proximal proteins within a 10-nm radius. The biotinylated proteins are captured by streptavidin beads and analysed by MS (Samavarchi-Tehrani et al., 2020).

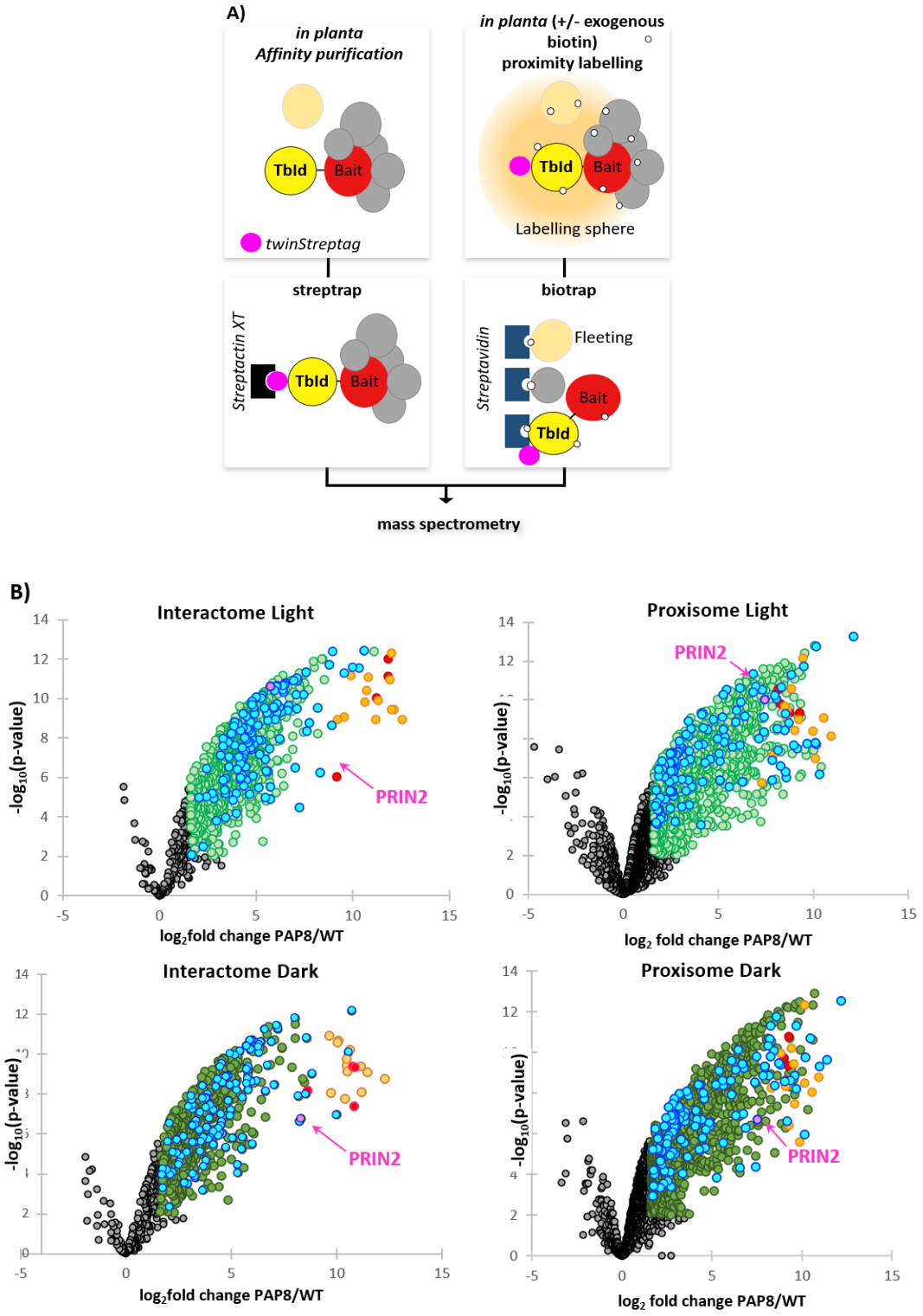


Figure 1.7: *(A) Schematic diagram of PAP8 proximity labelling. (B) Proxisome and Interactome data sets for PAP8 proximity labelling and affinity purification. The commonly found proteins are coloured in green, PAPs are coloured in orange, rpo subunits are coloured in red; nonspecific hits are in grey.*

1.5. OBJECTIVES

The aim of the thesis focuses on elucidating the roles and interactions of proteins engaged in redox signalling within the PEP complex, using an integrated approach comprising both *in vivo* and *in vitro* techniques.

1. The first objective deals with the participation in the purification of chloroplasts from *Sinapis alba* for PEP purification and structurally characterise the PEP complex by cryo-EM.
2. The second objective of this research seeks to address the superoxide dismutases, PAP4 and PAP9. To investigate this, the *in vitro* SOD activity of PAP4 and PAP9 is evaluated using the pyrogallol method.
3. The third objective of this thesis aims to discern the interactions between PRIN2 and CSP41b through the application of *in vitro* biophysical techniques such as isothermal titration calorimetry, native and denaturing mass spectrometry and isolation of CSP41b-PRIN2 complex by size-exclusion chromatography.
4. The fourth objective deals with the structural characterisation of CSP41b by cryo-EM and attempts to isolate the CSP41b-PRIN2 complex by using a crosslinker for subsequent cryo-EM studies.
5. The fifth objective deals with the approaches to identify CSP41b and PRIN2 interactions *in vivo* such as by utilising BiFC assay. This method enables the direct visualization of protein-protein interactions within living cells. To identify interacting redox partners with PRIN2 and CSP41b, a proximity labelling approach is employed. This method could shed light on the redox partners and the positioning of redox proteins within the active PEP complex, a matter that remains unknown. Proximity labelling on transiently expressed constructs in *Nicotiana benthamiana* was performed.

2. MATERIAL AND METHODS

2.1. Chloroplast fractionation for PEP purification from *Sinapis alba*

2.1.1. Chloroplast fractionation from *Sinapis alba*

Two trays of *Sinapis alba* seeds were sowed on day 1 and the trays were placed at 4°C (dark condition) for one night. On day 2, the trays were placed in a dark room at room temperature. The trays were transferred to phytotron on day 3 (16 h light / 8 h dark light regime) at 21°-23°C. On day 7, the cotyledons were harvested quickly by scissors. The material was transferred immediately to the cold aluminium foil in order to reduce internal enzyme activities and wounding responses. 100 g of plant material was placed in a homogenizer, and 200 mL of homogenisation medium (HM, Table 2) was added. The material was homogenised with short, low pulses (3 x 3 s). The suspension was filtered through three layers of nylon mesh and the liquid was collected in a large beaker. The collected homogenate was filtered through the nylon mesh in order to remove small cell wall particles. The nylon mesh was discarded with the remaining plant material. It was repeated until all the plant materials were homogenised. The suspension was transferred to 500 mL polycarbonate tubes and centrifuged in rotor JLA 10.5 (*Beckman Coulter Avanti J-20*) at 6084 *g* for 5 mins at 4°C. The supernatant was discarded and the sediments were suspended carefully and slowly with a soft paintbrush in a small amount of HM. The suspended sediments were combined in one measuring cylinder. Just before performing the density gradient centrifugation, the suspension was poured in a glass potter to remove the chloroplast aggregates carefully by gently up and down movement of the plunger. 60 mL of the suspension was collected. 20%, 40% and 80% Percoll® (Sigma-Aldrich) gradient solutions (12 tubes) were made in 50 mL polycarbonate centrifuge tubes. The homogenised chloroplast suspension was poured carefully into the Percoll gradients and centrifuge in Rotor JS 13.1 (*Beckman Coulter Avanti J-20*) 4696 *g* at for 35 min at 4°C. The collected intact chloroplast was diluted with the dilution buffer (Table 2) in a ratio of 1:1. It was poured into standard centrifuge tubes, and the chloroplasts were centrifuged at 4000 *g* for 12 mins at 4°C. The sediment contains the purified chloroplast. The supernatant was discarded. The chloroplast pellet was suspended in equal volumes of lysis buffer (Table 2) using a paintbrush.

Table 2: Composition of buffers used for chloroplast fractionation

Buffers	Composition
Homogenisation medium (HM)	0.33 M Sorbitol, 50 mM HEPES-KOH pH 8.0, 5 mM $MgCl_2$, 2 mM Na_2-EDTA, 0.3 mM DTT
Dilution buffer	0.05 M Tris-HCl pH 8.0, 0.01 M $MgCl_2$, 0.04 M 2-beta mercaptoethanol
Lysis buffer	0.05 M Tris-HCl pH 7.6, 4 mM Na_2-EDTA, 0.04 M 2-beta mercaptoethanol, 0.01 M NaF, 1 % (v/v) Triton X-100, 25 % (v/v) Glycerol
20% Percoll gradient	20% (v/v) Percoll, 0.33 M Sorbitol, 50 mM HEPES-KOH pH 8.0, 5 mM $MgCl_2$, 2 mM Na_2-EDTA, 0.3 mM DTT
40% Percoll gradient	40% (v/v) Percoll, 0.33 M Sorbitol, 50 mM HEPES-KOH pH 8.0, 5 mM $MgCl_2$, 2 mM Na_2-EDTA, 0.3 mM DTT
80% Percoll gradient	80% (v/v) Percoll, 0.33 M Sorbitol, 50 mM HEPES-KOH pH 8.0, 5 mM $MgCl_2$, 2 mM Na_2-EDTA, 0.3 mM DTT

2.2. Biochemical and biophysical experiments on PAP4, PAP9, CSP41b and PRIN2

2.2.1. Purification of CSP41b, PRIN2, PAP4 and PAP9

The CSP41b and PRIN2 coding regions were cloned into pET28a+ vector backbone that contains a TEV cleavage site at N-terminal, kanamycin resistance gene, lac operator and poly-histidine tag present at C-terminal, were overexpressed in *E. coli* BL21 (DE3) strain in LB medium containing 50 µg/mL kanamycin. PAP4 and PAP9 constructs were cloned in the pET21d vector backbone with the poly-histidine tag at C-terminal and were overexpressed in *E. coli* ROSETTA2 strain in LB medium containing 50 µg/mL kanamycin and 50 µg/mL chloramphenicol. The cells were incubated at 37°C until OD reached between 0.8 to 1.0. 0.5 mM IPTG was added to the culture and the culture was incubated at 16°C. After overnight expression, the culture was centrifugated at 5,500 g for 25 minutes at 16°C. The pellet was re-suspended in lysis buffer (250 mM NaCl, 20 mM imidazole, 50 mM Tris-HCl pH 8.0) containing protease inhibitor (complete protease inhibitor cocktail tablet (Roche)) and sonicated with the following conditions: amplitude = 70, pulse on = 2 sec, pulse off = 6 sec, duration = 3 min). The lysate was centrifuged at 15,000 g for 40 min at 4°C and filtered using a 0.45 µm membrane filter before loading onto a Ni-NTA column. The column was washed with lysis buffer followed by elution with different concentrations of imidazole (50 mM, 100 mM, 150 mM, 200 mM, 250 mM and 500 mM). The fractions were analysed by SDS-PAGE. The fractions containing proteins were pooled and dialyzed overnight at 4°C to remove imidazole and concentrated using Amicon® ultra centrifugal filter (Merck Millipore) (Pore size of the filters used: 30,000 MWCO for CSP41b, PAP4 and PAP9; 10,000 MWCO for PRIN2). The fractions were then loaded on size exclusion column using FPLC and Superdex® s200 16/600 (GE Healthcare) containing 25 mM Tris-HCl pH 8.0, 150 mM NaCl. The collected fractions containing pure proteins were pooled and stored at -20°C with 50% (v/v) glycerol for further experiments. The expected molecular weights for CSP41b = 39.1 kDa, PRIN2 =14.6 kDa, PAP4 = 26.9 kDa, PAP9 = 30.8 kDa.

2.2.2. Estimation of super-oxide dismutase activity by pyrogallol

SOD activity assay was performed to verify the residual superoxide dismutase activity that could be due the presence of a weak amount of Fe since it as reported that both SODs are active; PAP4 being the most active. The SOD activity of purified PAP4 and PAP9 were analysed using pyrogallol. The autooxidation of pyrogallol leads to the production of a yellow-coloured product purpurogallin absorbing at 420 nm. SOD inhibits the autooxidation of pyrogallol at

alkaline pH, by removing the superoxide. 7 mM pyrogallol was dissolved in aTris-succinate-EDTA buffer at pH 8.2 and the pyrogallol auto-oxidation was monitored by recording the absorbance increase at 420 nm. After 3 mins, PAP4 or PAP9 at several concentrations (50 µM, 100 µM, 200 µM, 500 µM and 1 mM) were added to the medium and the absorbance was monitored for 3 min. A control with 5 µM Mn-SOD (Invitrogen™) was used in this experiment.

2.2.3. Isothermal Titration Calorimetry for CSP41b and PRIN2

Isothermal titration calorimetry makes it possible to measure the heat released or absorbed during the interaction between two molecules in solution, generally, a biological macromolecule and a ligand. The measurement of the heat released or absorbed is carried out by using a microcalorimeter. It measures the temperature difference between the reference cell and the cell containing the protein sample. It compensates for the heat lost or gained by injecting energy to maintain a constant temperature. The ligand placed in a syringe is injected in small quantities into the cell containing the protein sample and heat released or absorbed during the interaction is measured at each injection. The interaction curve obtained determines the stoichiometry, which gives the number of binding sites between the two molecules, n), the enthalpy (heat absorbed or released during the interaction, ΔH°), association constant (affinity, K_a) between two molecules (Figure 2.1). Subsequently these data are used to calculate the free energy variation of the system (ΔG°), entropy (ΔS°) as well as the dissociation constant K_d.

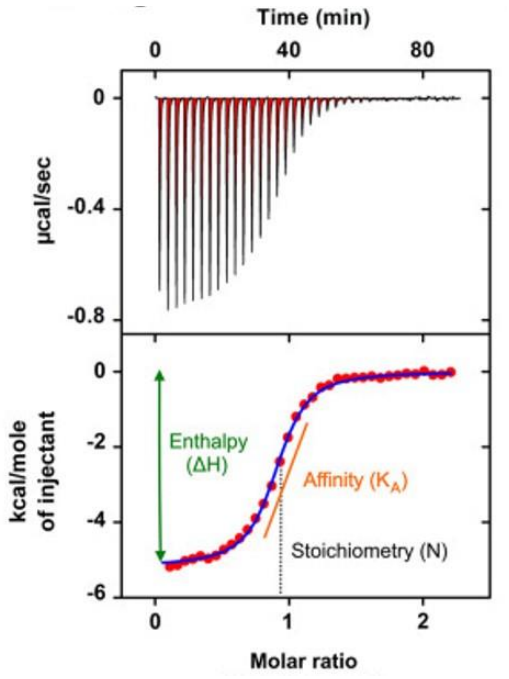


Figure 2.1: *The expected ITC figure is composed of an upper panel (time vs µcal/sec) and lower panel (molar ratio vs kcal/mole of the injectant). The lower panel displays the schematic representation of the calculations performed by the software. It provides the ΔH (enthalpy) which is the difference between the initial H value and the plateau H value (line in green with double arrowheads); K_A (association constant) is the slope value (orange line) that intercepts the exponential phase of the isothermal curve (blue line); N (stoichiometry) provides the molar ratio (dotted black line). Adapted from Saponaro, 2018*

The ITC experiments were carried out at Integrated Structural Biology Grenoble platform (ISBG) with the help of Dr. Caroline Mas. CSP41b and PRIN2 were purified in a buffer containing 50 mM NaCl and 10 mM Tris-HCl pH 8.0. 30 µM CSP41b was placed in the cell and 300 µM PRIN2 was placed in the syringe. Titration was made by 16 injections of 2.5 µL, reference power 5 µcal/sec, stirring speed 750 rpm and were carried out at 20°C on MicroCal ITC 200 device (Malvern). The initial injection was set at volume 1 µL, duration 2 sec, spacing 180 sec and filter period of 5 sec. The rest of the injections (2-16) were set at volume 2.5 µL, duration 5 sec, spacing 180 sec and filter period of 5 sec. The same conditions were duplicated.

2.2.4. Native and Denaturing mass spectrometry (MS) analysis for CSP41b and PRIN2

CSP41b and PRIN2 were purified as described and dialyzed in a buffer containing 250 mM Ammonium acetate pH 7.0 for native mass spectrometry analyses. The MS experiments were carried out at IBS mass spectrometry platform by Dr. Elisabetta Boeri Erba.

2.3. Cryo-EM experiments on CSP41b, CSP41b-PRIN2 complex

2.3.1. Principles of negative staining electron microscopy

Negative staining electron microscopy is employed to assess the quality and uniformity of purified proteins while offering limited-resolution insights into their structural characteristics (Orlova and Saibil, 2011). This method is characterised by its swiftness and the minimal protein quantity requirement, typically less than 0.1 mg/mL. The fundamental principle involves the selective deposition of heavy metal ions, onto a continuous carbon substrate. Protein particles adhere to a previously glow-discharged surface. A droplet of the protein suspension is applied to an electron microscopy support film and subsequently immersed in a heavy metal salt solution, typically uranyl acetate. The sample is then blotted to create a thin film and allowed to desiccate. Although uranyl acetate is the most commonly employed stain, offering superior contrast, specific protein structures exhibit improved preservation when stained with alternative substances like tungsten or molybdenum salts. The heavy metal staining process results in a dense coating outlining the contours of the biological assembly, thereby providing insights into particle size, shape, symmetry, and the overall uniformity of the sample. This technique earns its name 'negative staining' due to the visualization of macromolecular shapes through exclusion rather than binding with the stain. Consequently, the three-dimensional structure may undergo some degree of flattening, and the stain may not uniformly cover the entire molecule, leading to potential distortions or omissions in the image data. For comprehensive three-dimensional structure determination, cryogenic methods are generally preferred.

2.3.2. Principles of single particle cryogenic electron microscopy

Like X-ray crystallography and Nuclear Magnetic Resonance (NMR), Cryo-electron microscopy (cryo-EM) is a method for determining the structure of biological molecules. It is feasible to preserve samples in their native, hydrated state even in a high-vacuum environment using cryo-electron microscopy (cryo-EM).

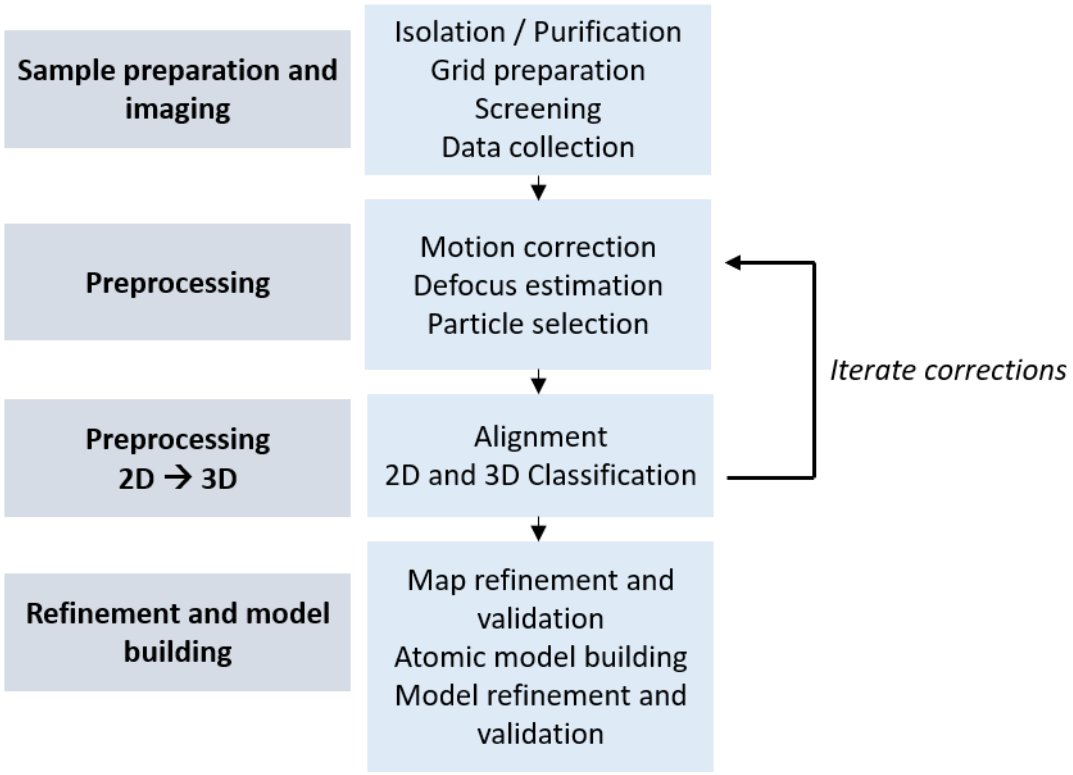


Figure 2.2: *Single particle cryo-EM analysis workflow. Adapted from Saibil, 2022.*

Single-particle cryo-electron microscopy (cryo-EM) utilizes an electron transmission technique that capitalizes on electrons' ability to traverse an exceedingly thin layer of vitreous ice containing embedded proteins. The workflow is depicted in figures 2.2 and 2.3. The sample must be maintained at low temperatures during both transfer and observation within the electron microscope (EM), which is achieved by vitrification. Vitrification refers to the transformation of liquid water into an amorphous state without triggering the nucleation of ice crystals. Rapid vitrification is instrumental in preventing ice crystal formation, as nucleation depends on factors like time, temperature, and pressure. To achieve this, a common method involves rapidly freezing aqueous solutions, applying them to form a thin layer, and promptly immersing them into either liquid ethane or propane, which are cooled to approximately -

182°C using liquid nitrogen for efficient heat transfer from the specimen. The choice of using liquid ethane is preferred over liquid nitrogen for rapid cooling because liquid ethane operates near its freezing point rather than its boiling point. This prevents its evaporation and the subsequent creation of an insulating gas layer. At 90% humidity, a substantial evaporation rate occurs, leading to notable osmotic and conformational changes in the specimen. Conversely, at 100% humidity, evaporation is negligible (Kanno H et al, 1976). This preservation process involves ensuring that the specimen remains cold, devoid of surface contaminants, and mechanically and thermally stable within a warm microscope. Once cooled, the sample is maintained at a temperature around -170°C, close to the temperature of liquid nitrogen (-196°C). Additionally, the low temperature significantly slows down the damaging effects of the electron beam. However, it's important to note that the presence of a continuous carbon film in the imaging process can introduce additional background scattering, which reduces image contrast. To mitigate this, perforated carbon films are often employed. These films allow imaging of the sample in areas of ice that are suspended over holes in the supporting film, enhancing image quality and contrast.

The fundamental concept underlying electron optical lenses involves the deflection of electrons, small negatively charged particles, through the influence of an electromagnetic field. Analogous to a traditional light microscope, an electron microscope (EM) comprises critical components, including an electron source, an array of lenses, and an image detection system, which can take the form of a viewing screen, photographic film, or digital camera. The grid is then placed in the electron microscope, where it will be exposed to an electron beam. Grids feature predefined hole sizes, shapes, and arrangements. The ideal hole size typically ranges from 1 to 2 µm. These grids are supported by either carbon (e.g., Quantifoils®, C-flat®) or gold (Russo and Passmore, 2016). An additional carbon film is often employed to enhance stability and facilitate CTF fitting. Grid screening involves the evaluation of protein concentration and stability, ice thickness and uniformity across the grid, and the phase of the ice (amorphous or crystalline). This process is essential for achieving high-resolution data collection.

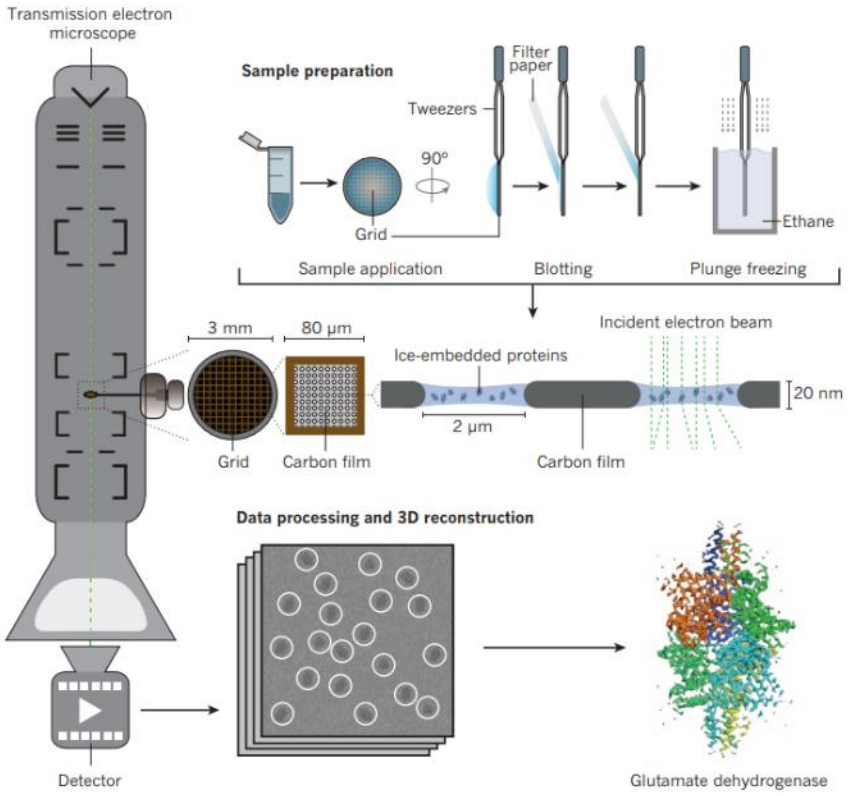


Figure 2.3.: *Schematic representation of the workflow in single-particle cryo-EM. Source: Adapted from Fernandez-Leiro and Scheres, 2016*

Electron Sources: The conventional electron source relies on a tungsten filament heated to temperatures between 2000° and 3000°C. Under these conditions, the electron energy surpasses the work function of tungsten. Electrons thermally emitted from the filament are accelerated by an electric field between the anode and the filament. In modern high-performance microscopes, the preferred electron source is the field emission gun (FEG). The FEG emits a beam of electrons with a smaller diameter, higher coherence, and approximately 500 times greater brightness, along with a narrower energy distribution. This improvement is achieved through the utilization of a single crystal tungsten emitter sharpened to yield a tip radius of approximately 10 to 25 nm. The emitter tip is coated with ZrO_2, which reduces the work function for electrons. Electrons are extracted from the crystal tip through a strong potential gradient at the emitter surface (field emission) and then accelerated through voltages ranging from 100 to 300 kilovolts.

The Interaction of Electrons with Specimens: Utilising electrons for imaging offers the distinct advantage of high resolution owing to their short wavelength. However, the robust interaction between electrons from the primary electron beam and the specimen leads to radiation-induced damage within the sample. The nature of this interaction is contingent upon both electron damage and the composition of the sample. Among the electrons engaging with the specimen, some scatter without losing energy (known as elastic scattering), while others transfer energy to the specimen (referred to as inelastic scattering). This energy transfer from incident electrons can result in various effects, including the ionization of atoms within the specimen, the induction of X-ray emissions, the rearrangement of chemical bonds, the generation of free radicals, or the initiation of secondary electron scattering. All of these interactions bring about alterations in the specimen's structure. Radiation-induced damage represents a noteworthy constraint when aiming for high-resolution imaging of biological molecules. Typical electron exposure levels employed for biological samples fall within the range of 1 to 20 electrons per square Ångström ($e^-/Å^2$). Biological specimens can withstand electron exposures ranging from 100 to 500 ($e^-/Å^2$). However, the finest details of the specimen begin to exhibit alterations at 10 $e^-/Å^2$ or less, depending on variables such as specimen temperature and chemical composition. Consequently, radiation damage dictates the experimental parameters and imposes limitations on the resolution achievable in determining biological structures. To mitigate radiation damage during processes such as area selection, alignment, and focusing specialized low-dose systems are employed. These systems divert the electron beam until the final stage of image recording.

Glow Discharge: Glow discharge is a technique where oxygen is divided into radicals with a high affinity for materials like carbon. This reaction cleans the surface, resulting in increased hydrophilicity. The charge at the bottom becomes negative. The process typically lasts for about 10 to 60 seconds in an air plasma environment.

Phase Plates and Energy Filters: The minimal phase shifts caused by biological specimens in scattered electrons often yield images with insufficient contrast. In phase contrast light microscopy, the visualization of phase objects is facilitated through the deployment of a quarter-wave phase plate. This plate introduces visible contrast by altering the phase of scattered light by 90° relative to the transmitted beam, resulting in constructive interference.

In contact with the atoms of the sample, the electrons sent by the microscope will then be dispersed and detected by a camera to form a two-dimensional image of the objects present (micrograph). During image processing, these 2D projections are first aligned and classified then combined to obtain a 3D representation of the object studied.

Imaging: CCD cameras record electron interactions indirectly, as the interaction of electrons with the scintillator component in the CCD camera generates photons that the device detects. Notably, due to the phenomenon of backscattering, the system produces noise, where a single electron can potentially generate a signal that registers as two independent events. While CCD cameras enable automation in image processing, they do have a limitation in resolution stemming from this effect. Hence, the film-based system continues to provide a higher Detective Quantum Efficiency (DQE), resulting in higher-resolution images. DQE quantifies the detector's ability to record incoming signals (SNR_{in}) and offers insights into efficiency at various resolutions or spatial frequencies (SNR_{out} / SNR_{in}). A significant advancement occurred with the introduction of Direct Electron Detectors (DED or DDD) in Single Particle Analysis (SPA) cryo-EM. This revolution was accompanied by the development of new processing software. DDD systems significantly enhance DQE when compared to previous technologies, leading to the acquisition of higher-resolution images.

Phase Contrast, Defocus, Optical Aberrations, and the CTF: Biological molecules, owing to their composition of elements like hydrogen (H), oxygen (O), nitrogen (N), and carbon (C), do not inherently provide substantial amplitude contrast. Consequently, in electron microscopy (EM), image contrast predominantly relies on phase contrast. During scattering, electrons undergo a 90-degree phase shift. Furthermore, the distinctive paths electrons follow after elastic scattering, compared to the unscattered beam, introduce an additional phase shift contributing to phase contrast. The magnitude of this additional path length due to scattering varies with the specific scattering angle.

CTF Estimation and Correction: To rectify optical distortions, the Contrast Transfer Function (CTF) is calculated. Determining the defocus setting is a crucial step, conducted initially for each micrograph and subsequently for subregions or individual particles within the micrographs. For higher-resolution analyses, more localized CTF determination becomes necessary.

Motion Correction: The generation of enhanced images involves the utilization of motion-corrected frames that have been obtained through dose fractionation and subsequent merging.

Particle picking: During particle selection, the coordinates of particles or regions of interest are identified, a task that can be performed manually or through automated methods, with or without the use of templates. The particle-picking programs are typically trained using manually selected particles as reference points to guide the automated selection process. Parameters are adjusted to optimize the selection of genuine particles while minimizing the inclusion of unwanted features to prevent bias. A notable portion of the initially selected particles may represent irrelevant objects, which are subsequently discarded during the 2D classification stage. The selection of the most appropriate particle-picking program remains contingent on the specific characteristics of the sample.

Single Particle Alignment, 2D Classification, and 3D Reconstruction: Following the selection of particles, the initial steps involve 2D alignment and classification. 2D classes offer insights into particle orientation distribution and homogeneity. With the aid of precise initial models and well-designed software, 3D classification and refinement can proceed efficiently.

Validation of map: To validate the map, it must exhibit consistency with the image data and reveal details in line with the reported resolution. Incorporating concepts from crystallography, the B factor (also known as the temperature factor) is employed to describe local uncertainty in density. This uncertainty may arise from structural disorders or defects in image processing accuracy. B factor correlation involves rescaling Fourier amplitudes to suppress dominant low-frequency terms that can obscure finer structural details. This rescaling is applied locally to account for amplitude falloff. With noisy image data, iterative refinement can produce spurious high-resolution noise. A common approach for assessing the resolution of an EM map involves dividing the dataset into two independent halves and refining them separately. This allows for the comparison of their Fourier transforms. The resulting 3D correlation is averaged within resolution shells, generating a line plot known as the Fourier shell correlation (FSC). Ideally, the FSC curves at0.143 and 0.542 should initiate at 1, reflecting the expected match between the two half datasets at low resolution, before

gradually declining in a sigmoidal fashion to zero correlation beyond the information limit inherent in the data.

Fitting and Validating Atomic Models: The process of fitting and validating atomic models hinges primarily on the resolution of the map, as it dictates the level of detail that can be incorporated. This detail ranges from accommodating entire subunits or domains at lower resolutions, to capturing secondary structures at intermediate resolutions, and eventually resolving finer features like side chain rotamers, water molecules, and ions at higher resolutions. The validation of models encompasses both the model itself and its alignment with the map. The alignment's validation is achieved through the employment of a map-to-model Fourier shell correlation (FSC). Notably, the evaluation criteria for the final model should differ from those employed during its creation. Key criteria encompass:

1. Model-to-map density correlation.
2. Model geometry.
3. Detection of clashes.
4. Assessment of local scores along the sequence to highlight problematic regions.
5. Examination of the consistency in local B factors.

In cases where a related known structure exists, a practical approach is to initiate the process by constructing a homology model using the protein sequence of interest. When working with low-resolution maps, a model can be placed within the map while adhering to distance constraints applied to the constituent atoms, as outlined by Croll and Read (2021). At intermediate resolutions, the fine-tuning of subdomains or secondary structures often necessitates adjustments at hinge points within the model. In contrast, at atomic resolutions, the potential exists for de novo model building. Programs initially developed for atomic modelling in crystallography have been adapted for application with electron microscopy (EM) maps. A key distinction lies in the fact that crystallographic models are employed to refine map phases (such as Coot, Refmac, and Phenix), whereas, in EM, the map itself serves as the final output. Current methods focus on optimizing the alignment between the model and the map to maximize the model-to-map density correlation.

2.3.3 Negative stain electron microscopy sample preparation

CSP41b samples was provided to IBS EM platform. The experiment was performed by Daphna Fenel. The samples were absorbed to the clean side of a carbon film of a mica. It was stained with sodium silico tungstate (SST) at 1% in distilled water (pH 7.0 – 7.5). and transferred to a 400-mesh carbon grid. The images were collected under low dose conditions (<10 e^-/A^2) with defocus values between 1.2 and 2.5 µm on a Tecnai 12 LaB6 electron microscope at 120 kV accelerating voltage using CCD camera Gatan Orius 1000.

2.3.4. Cryo-EM grid preparation

Cryo-electron microscopy. 3.5 µL of sample were applied to 1.2/1.3 Ultrafoil holey carbon grids (Quantifoil Micro Tools GmbH, Germany) and they were plunged and frozen in liquid ethane with a Vitrobot Mark IV (Thermo Fisher Scientific) (6 s blot time, blot force 0). The sample was observed at the beamline CM01 of the ESRF (Grenoble, France) (Kandiah et al., 2019) with a Titan Krios G3 (Thermo Fisher Scientific) at 300 kV equipped with an energy filter (Bioquantum LS/967, Gatan Inc, USA) (slit width of 20 eV). 4185 images were recorded automatically on a K2 Summit direct detector (Gatan Inc., USA) with EPU (Thermo Fisher Scientific). Movies were recorded for a total exposure of 5 s and 125 ms per frame resulting in 40 frame's movies with a total dose of ~40 e−/$Å^2$. The magnification was 165,000x (0.83 Å/pixel at the camera level). The defocus of the images was varied between −0.5 and −1.0 µm. The phase plate position was changed automatically every ~100 images, which corresponds to an accumulated dose of ~50 nC on each phase plate position.

***3D reconstruction**:* The movies were first drift-corrected with motioncor2 (Zheng SQ et al., 2017). The remaining image processing was done in RELION 3.08 (Kimanius et al., 2016) and 3.1 (Zivanov et al, 2018). CTF estimation was done with GCTF (Zhang et al, 2016). An initial set of particles (box size of 128 pixels, sampling of 1.66 Å/pixel) was obtained by auto-picking with a gaussian blob. After 2D classification (run1), the best looking 2D class averages were used to both generate an *ab-initio* initial 3D model (mask diameter: 120 Å, symmetry C1) and do a template-based picking. Following two more 2D classifications (run2 and 3), a first 3D classification (symmetry C1, circular mask of 120 Å diameter, 5 classes) was performed using the particles selected from 2D classification (run3) and using as a reference the 3D model determined from the particles selected in run1. The two best classes were pooled together

and a first 3D refinement (symmetry C2, circular mask of 120 Å diameter) resulted in a first 3D reconstruction at 3.8 Å. After re-extraction (box size 256 pixels, sampling of 0.83 Å/pixel), another 3D refinement (symmetry C2, circular mask of 120 Å diameter) followed by a last 3D classification (C1 symmetry, no alignment, tight mask, 3 classes) resulted in the final set of particles (129 038 particles). The final 3D reconstruction (symmetry C2, tight mask) was calculated and a resolution of 3.4 Å was determined by Fourier Shell Correlation (FSC) at 0.143. Particles polishing, beam tilt correction, magnification anisotropy and CTF refinement per particle were then attempted but did not result in any significant improvement.

2.3.5. CSP41b model building and validation

UCSF ChimeraX (Pettersen EF et al, 2021) was used to analyse and interpret the cryo-EM map. Using CCP4i (Potterton et al, 2003) molecular replacement Chainsaw (Schwarzenbacher et al., 2004) package, a polyalanine model was made from the sequence of CSP41b sequence. Using Map to model in Phenix software package (Torices et al., 2015) the starting model was placed in the map. COOT (Emsley et al., 2010) was used to analyse the model. Using Map symmetry in Phenix, the dimeric model was built. ChimeraX was used to create figures. The built model was finally refined by phenix.real_space_refine in Phenix.

Using Alphafold2 (Jumper et al., 2021) model of CSP41b (Uniprot ID: Q9SA52) without its cTP, the model was first placed in the cryo-EM map using Phenix. Then, using the map symmetry, the dimer was built. The atomic coordinates were then refined by energy minimization and dynamic. The model was manually rebuilt using COOT and further refined, after several cycles of building and refinement, the model was validated using Phenix. Refinement and validation statistics are provided in Table 7.

2.3.6. BS3 crosslinking of CSP41b-PRIN2 complex for cryo-EM

The main chemical bridging agent used to stabilise the CSP41b-PRIN2 complex was bissulfosuccinimidyl suberate, also known as BS3 (Thermo Fisher). It is a reagent with a length of 11.4 Å commonly used for the chemical bridging of proteins. It is a coupling agent which has two reactive N-hydroxysulfosuccinimide (NHS) ester groups and which allows the formation of covalent bonds between the primary amine groups of lysine but also those of histidine or that of residue in the N-terminal region. Various conditions for crosslinking CSP41b-PRIN2 were tested. CSP41b and PRIN2 was crosslinked at room temperature in the

presence of BS3 at concentrations of 0.5 mM, 1 mM and 1.5 mM. The incubation period was varied between 45 minutes and 90 minutes among these samples. The reaction was stopped by adding 50 mM Tris pH 8.0. A fraction of the sample was analysed by 15% SDS-PAGE to check if the subunits were cross-linked. Samples from conditions 1 and 2 (Table 3) were analysed by negative-stain electron microscopy at IBS EM platform.

Table 3: Conditions for crosslinking CSP41b-PRIN2 with BS3

S. N.	Csp41b : PRIN2	BS3 concentration	Incubation time
1.	1 : 10	0.5 mM BS3	45 minutes
2.	1 : 10	0.5 mM BS3	90 minutes
3.	1 : 10	1 mM BS3	45 minutes
4.	1 : 10	1 mM BS3	90 minutes
5.	1 : 10	1.5 mM BS3	45 minutes
6.	1 : 10	1.5 mM BS3	90 minutes
7.	1 : 3	0.5 mM BS3	45 minutes
8.	1 : 3	0.5 mM BS3	90 minutes
9.	1 : 3	1 mM BS3	45 minutes
10.	1 : 3	1 mM BS3	90 minutes
11.	1 : 3	1.5 mM BS3	45 minutes
12.	1 : 3	1.5 mM BS3	90 minutes

2.4 *In vivo* interaction experiments on CSP41b and PRIN2

2.4.1. Cloning for *in vivo* analysis of CSP41b-PRIN2 interactions

gDNA and cDNA of *A. thaliana* Col-0 were extracted using DNeasy Plant Miniprep kit (Qiagen) and QuantiTech reverse transcription kit (Qiagen), respectively. Primers were designed by checking the loci of the genes CSP41b and PRIN2 (Figure 2.4) using ApE software (Davis and Jorgensen, 2022). The promoter region of PRIN2 was cloned from gDNA and the open reading frames CSP41b$^{\Delta ctp}$, ctpPRIN2 and PRIN2$^{\Delta ctp}$ were cloned from cDNA. The list of primers and PCR conditions are tabulated below (Table 4, 5).

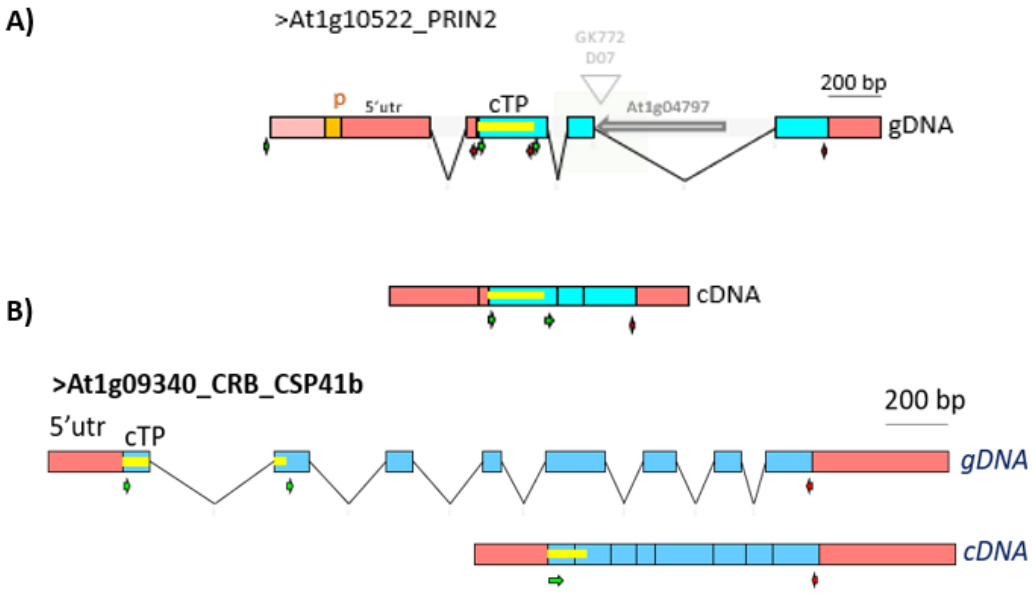


Figure 2.4: *Illustration of gene loci of PRIN2 (A) and CSP41b (B). Exons in blue; untranslated regions in red; cTP in yellow; Introns as broken lines; p, promoter region of* PRIN2; *and oligonucleotides as green and red arrows according to the cloning strategy of each selected fragment.*

50 µL of reactions were run on thermal cycler (BioRad c1000 thermal cycler). The number of cycles was set between 30 - 40. Phusion™ High fidelity DNA polymerase (Thermo scientific™) was used for amplification. After PCR, the reaction mixture was analysed by 1% TAE DNA agarose electrophoresis run in electrophoresis tank (Biorad DNA mini gel system). 1X SYBR™ Safe DNA gel stain (Invitrogen™) was added to the agarose for visualisation. 1X TAE buffer (Euromedex) was used as running buffer. 6X loading buffer (Thermo Scientific™) was mixed with the PCR reaction mixture. 1kb+ DNA ladder (Invitrogen™) was used as a marker. The electrophoresis tank was run at 110 V until the bands were well migrated. Transilluminator unit (Accuris Smart Blue™ Blue light transilluminator) and GelDoc machine (BioRad) were used

for visualisation. The required DNA bands were cut with a sterile scalpel and purified using gel NEB Monarch® DNA extraction kit. The final elution was with sterile milliQ H_2O and DNA concentration (ng/µL) was performed with NanoDrop™ 2000/2000c spectrophotometer.

The respective fragments after PCR reactions, were cloned using Zero Blunt™ TOPO™ PCR cloning kit with PCR™ - Blunt II TOPO™ vector (Invitrogen™) following the provided protocol with the kit. 50 µL of One Shot™ TOP10 chemically competent *E. coli* cells were thawed on ice. 10 µL of ligation mixture were mixed with the competent cells and kept on ice for 5 – 15 mins. Heat shock was given at 42°C for 30s and immediately transferred to ice. 250 µL of SOC media (Invitrogen™) was added and incubated for 1 hour at 37°C at 180 rpm. LB agar (Invitrogen™) plates with respective antibiotics for the plasmids were used for growth. 200 µL of culture were plated and kept at 37°C overnight. 3 – 6 colonies from each plate were screened by plasmid isolation (QIAPrep spin miniprep kit, Qiagen) followed by double restriction enzyme digestion with respective sites that were introduced into the plasmids. The restriction digestion was set at 37°C for 1 hour. The restricted fragments were screened by DNA electrophoresis. The plasmids were stored at -20°C and the plasmids with expected band length were sent for sequencing to Eurofins. The plasmids pMCD03, pMCD07, pMCD08, pMC09 were cloned by Marie-Catherine Ducarre and plasmids pBB654a were cloned by Dr. Robert Blanvillain. pFX024 was cloned by Dr. François-Xavier Gillet.

Table 4: List of primers used for cloning:

	Primer name	**Sequence**
CSP41b (1146 bp)	oCsp41b_fXhoI	5’-ctcgagatggcgaagatgatgatgttgcaacagc-3’
	oCsp41b_rSacII	5’-ccgcggttgaagaacaagtttcttgctcagaatcatgtcg-3’
PRIN2 (549 bp)	octPRIN2_fXhoI	5’-ctcgagatggcttcaatgcacgaagctctg-3’
	oPRIN2_rSacII	5’-ccgcggatcagtgccggtccattccttcc-3’
N-YFP	oNY_fBHI	5’-ggatccatggtgagcaagggcgaggagc-3’
	oNY_rKpnI	5’-ggtacctccggacatgatatagacgttgtgg-3’
C-YFP	oYC_fBHI	5’-ggatccatggccgacaagcagaagaacg-3’
	oYC_rKpnI	5’-ggtacccttgtacagctcgtccatgcc-3’
pPRIN2	opprin2_fSacI	5’-gagctccgagccacacgagaggaatctc-3’

(797 bp)	opprin2_rXhoI	5’-ctcgagtactaagctttgcttcactcttcaacc-3’
ctpPRIN2 (219 bp)	octPRIN2_fXhoI	5’-ctcgagatggcttcaatgcacgaagctctg-3’
	octPRIN2_rSmaI	5’-cccgggtctgcaaacgaaccctctc-3’
PRIN2 (345 bp)	oPRIN2_fNcoI	5’-ccatgggcgctgctgagtacaagtttcc-3’
	oPRIN2_rXbaI	5’-tctagactaatcagtgccggtccattcc-3’

Table 5: List of plasmids

Plasmids	Antibiotic	Inserts
pSSM28	Kanamycin	CSP41b_XhoISacII in TOPO-blunt II
pSSM41	Carbenicillin	N-YFP + CSP41b_XhoISacII
pSSM42	Carbenicillin	C-YFP + CSP41b_XhoISacII
pSSM43	Carbenicillin	GFP + CSP41b_XhoISacII
pMCD3	Kanamycin	PRIN2_XhoISacII in TOPO-bluntII
pMCD07	Carbenicillin	N-YFP + PRIN2_XhoISacII
pMCD08	Carbenicillin	C-YFP + PRIN2_XhoISacII
pMCD09	Carbenicillin	GFP + PRIN2_XhoISacII
pRB1001	Carbenicillin	Coilin
pBB301	Carbenicillin	PAP10 + dsred
pBB330	Carbenicillin	PAP4 + dsred
pSSM03	Kanamycin	pPRIN2_SacIXhoI in TOPO-bluntII
pBB654a	Kanamycin	ctpPRIN2_XhoISmaI in TOPO-blunt II
pSSM15i	Kanamycin	PRIN2_NcoIXbaI in TOPO-bluntII
pSSM38	Carbenicillin	pPAP8+ctpPRIN2+Twin-Strep+HA+TurboID+PRIN2
pSSM39	Carbenicillin	pPRIN2+ctpPRIN2+Twin-Strep+HA+TurboID+PRIN2
pFX024	Carbenicillin	pPAP8+ctpPAP8+Twin-Strep+HA+TurboID+PAP8

2.4.2. Bimolecular fluorescence complementation assay for CSP41b-PRIN2 interactions in onion epidermal cells

For *in vivo* techniques, a BiFC assay was used. BiFC allows for direct visualisation of protein-protein interactions in living cells. N- and C-terminal fragments of yellow fluorescent protein (YFP) were truncated and ligated to N- and C- -terminal regions of the proteins of interest, PRIN2 and CSP41b. Different constructs with N- and C-terminal fragments of YFP with PAP8, PAP4 and PAP10 were also designed to test their interactions with PRIN2 and CSP41b, respectively. The principle of BiFC is based on *in vivo* reconstitution of yellow fluorescent protein in onion cells. If the two proteins interact, the fluorescent protein will reconstitute and form a fluorescent complex. The signal can be visualised by epifluorescence microscopy without any special treatment to the cells. The technique provides insight into the sub-cellular localisation of the interacting complex. It does not require capturing signals at two different wavelengths, such as in FRET, which is a significant advantage for using BiFC. The analysis is sensitive with less background as fluorescence recovery requires the interaction of the two proteins when the non-fluorescent portions are brought nearby (Miller et al., 2015). It provides insight into the initial appearance of the interaction. The dynamics of the protein-protein interactions cannot be investigated as the fluorescent fragments re-assembly is irreversible.

2.4.3. Preparation of DNA and gold mixture and bombardment in onion cells

2.5 µg/mL of plasmid was diluted in sterile water to a final volume of 20 µL. (1 µg of plasmid for 1 shoot). 20 µL of gold particles (Seashell technology; 0.6 µm; 30 mg/mL) was taken and vortexed until it was completely suspended. 30 µL of binding buffer (Seashell technology) was added to the gold suspension. It was placed on ice for 5 minutes. 70 µL of precipitation buffer (Seashell technology) was added to binding buffer + Gold suspension mixture and incubated for 10 mins on ice until the gold particles had settled down. The supernatant was discarded and rinsed with 500 µL of 100% ethanol. It is kept on ice. 500 µL of ethanol (without gold) was discarded and 30 µL of 100% ethanol was added. It was resuspended well. 10 µL of gold suspension was placed in the centre of the macro carrier and allowed to dry. The rupture disk, macro-carrier and micro carrier were assembled in the biolistic canon chamber (Bio-rad PDS-1000/He). The onion piece was placed. The chamber was vacuumed until the gauge reached

27 mmHg and the switch was moved to “Hold” position. It was switched to “Fire” until the pop sound due to breakage of rupture disk (1100 psi) was heard. The onion slice was kept in dark for 24h-48h and the peel was observed under the fluorescence microscope. The imaging was performed in a Nikon Axioscope microscope equipped with an AxioCam MRc camera. Nikon’s Zen software was used for picture acquisition.

2.4.4. Sterilisation of *Nicotiana benthamiana* seeds

Nicotiana benthamiana seeds were taken in a microcentrifuge tube and 1 mL of sterilisation solution (2.6% bleach and 0.1% Triton X-100) and kept aside for 5 minutes. It was centrifuged for 30 seconds at the highest speed. The supernatant was discarded and the seeds were rinsed with sterile H_2O thrice. ½ MS + 1% Sucrose pH 5.4 media plates were used for sowing the seeds. The plates were stored at 4°C for two days for vernalisation. The plate was transferred to phytotron (21°-24°C 16 h light / 8 h dark light regime) and for 12 days. The seedlings were transferred to pots and kept at phytotron until they were four-weeks old.

2.4.5. *Agrobacterium tumefaciens* infiltration

The plasmid was desalted in sterile H_2O using MF-Millipore™ 0.025 µm membrane filter paper (Merck) for 5 mins. To 50 µL of electro-competent *Agrobacterium tumefaciens* strain C58C1 pMP90 cells, 1 µL of desalted plasmid with approximate concentration of 250 ng/µL was added. The electroporation unit was set up at 2500 V, 5 millisecs. After electroporation, 1 mL of LB media was added to the cells and incubated at 28°C for 3 hours. 300 µL of the culture was plated in LB agar plates with 50 µg/mL rifamycin (Carl Roth®), 50 µg/mL gentamycin (Carl Roth®) and 50 µg/mL spectinomycin (TOKU-E) and incubated at 28°C for 2 days. Two days before infiltration, the transformed binary vector colonies were resuspended in 6 mL of LB media containing 50 µg/mL rifamycin, 50 µg/mL gentamycin and 50 µg/mL spectinomycin and incubated at 28°C at 200 rpm until OD_{600nm} was 0.5-1.0. The cell culture was mixed with infiltration media (10 mM MES pH 5.5, 10 mM $MgCl_2$ and 200 µM acetosyringone (3’,5’-dimethoxy-4’-hydroxyacetophenone, Merck)) in the ratio of 1:1. Infiltration was performed used a needleless syringe on the bottom surface of four-weeks old *N. benthamiana* leaves in patches. The pots were placed in a cool place and transferred to phytotron maintain temperature of 21°-24°C with 60% humidity and 16 h light / 8 h dark light regime.

2.4.6. Transient expression Gus assay in leaves of *Nicotiana benthamina*

Proximity labelling can also be attempted in transiently expressed Tb-ID constructs with proteins of interest in *Nicotiana benthamiana*. The tobacco leaves were incised and incubated in 80% cold acetone for 20 mins. The acetone solution was discarded. The leaves were rinsed with rinse buffer and the solution was discarded. The leaves were incubated with Gus solution without X-Gluc for 10 mins. The solution was discarded. Gus solution with X-Gluc (Thermo Scientific™) was added and vacuum infiltrated for 10 mins. The leaves were incubated at 37°C until blue colour emerged. The Gus solution with X-Gluc was discarded and 80% ethanol was added to remove the chlorophylls. Once the leaves were bleached, the leaves were photographed and recorded (Figure 3.22).

2.4.7. Transient expression proximity labelling in *Nicotiana benthamiana*

The *Agrobacterium* inoculum containing the plasmids pSSM38, pSSM39 and pFX024 was infiltrated in leaves of four weeks old *N. benthamiana* plants using a needless syringe. The pots were kept in shade overnight and transferred to phytotron. After three days of incubation, 200 µM biotin (Thermo Fisher Scientific™) were infiltrated in the leaves as patches using a needless syringe and incubated for 30 minutes. The leaves were incised using scissors and flash frozen in batches of 700 mg with liquid nitrogen and stored at -70°C until use for sample preparation. The protocol for TurboID sample preparation was adapted from Zhang et al, 2019. The frozen leaf material (700 mg of leaves for each construct) was ground to a fine powder using liquid nitrogen and 1 mL of RIPA lysis buffer (Table 6) in a mortar. The samples were mixed in a vortex machine and were immediately centrifuged at 16,500 g for 10 min and the lysates were collected. 200 µL of streptavidin-coated magnetic beads (Dynabeads™ MyOne™ Streptavidin C1, Invitrogen) were washed twice with RIPA lysis buffer and the lysates were incubated with the equilibrated beads on a rotator overnight at 4°C. The following day, the beads were sequentially washed using the following buffers mentioned in Table 6, once with 1 mL buffer I, followed by once with buffer II and once with buffer III. The beads were washed twice in 50 mM Tris-HCl, pH 7.5 and six more times in 50 mM ammonium bicarbonate, pH 8.0 to remove detergents. Finally, the beads were resuspended in 1 mL of 50 mM ammonium bicarbonate. The suspension was used for Western blot analysis and the rest of the beads was flash-frozen in liquid nitrogen and stored at -80°C.

Table 6: Composition of buffers used for TurboID sample preparation

Buffer	Components
RIPA lysis buffer	50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 1 mM EDTA, 1% NP40 (v/v), 0.1% SDS (w/v), 0.5% sodium deoxycholate (w/v), 1 mM DTT, 1 tablet of complete™ Protease Inhibitor Cocktail
Buffer I	2% SDS in H_2O.
Buffer II	50 mM HEPES (pH 7.5), 500 mM NaCl, 1 mM EDTA, 0.1% deoxycholic acid (w/v), 1% Triton X-100
Buffer III	10 mM Tris-HCl, pH 7.4, 250 mM LiCl, 1 mM EDTA, 0.1% deoxycholic acid (w/v), 1% NP40 (v/v)

2.4.8. Western blot analysis

12% acrylamide:bis-acrylamide gels were prepared. The samples were loaded after mixing with loading dye. It was set to run in Biorad® electrophoretic system at 100 V until the bands were well separated. The gel was stained with InstantBlue® Coomassie stain (Abcam) and the other gel was used for western blot analysis. Two antibodies were used to detect the presence of transformed constructs as they possess 2X strep tag and HA tag. Streptactin-HRP (Thermo Fisher Scientific™) was used to detect 2X strep tag in the biotinylated proteins and rabbit polyclonal anti-HA antibody (Sigma Aldrich) was used to detect the HA tag.

Non-specific sites were blocked by soaking the membranes in 5% non-fat dry milk in TBS (plus Tween20 0.1%) and incubated overnight at 4°C or 1 hour at room temperature under constant agitation. The following day, Streptactin-HRP was dissolved in 5% non-fat dry milk with TBS-T for a final concentration of 1:1000. The membranes were incubated for 1 hour at room temperature at constant agitation. The membranes were washed thrice with TBS-T (3 x 5 mins) and rinsed with TBS-T for (5 mins). When using Rabbit polyclonal anti-HA, it was added at 1:4000 dissolved in 5% non-fat dry milk with TBS-T and incubated overnight at 4°C with constant agitation. The membrane is washed 6 x 5 mins with TBS-T and a secondary rabbit anti-HRP at 1:10,000 for 1 hour at room temperature in agitation.

The immuno-detected protein is revealed using a chemiluminescence reagent (Biorad ®, clarity #170560) by incubating the membrane for 1 min with a mix of solution A and solution B (500 µL each).

3.RESULTS

3.1. *Sinapis alba* PEP ENVELOPE

ARTICLE 1: Three-Dimensional Envelope and Subunit Interactions of the Plastid-Encoded RNA Polymerase from *Sinapis alba*

A main objective in our laboratory is to solve the 3D structure of the PEP complex and understand its dynamics during the dark-to-light transition. In that line of research, the study in Ruedas et al, 2022 aimed to understand the PEP complex and its associated proteins in *Sinapis alba* by examining their structural organization and interactions. This was achieved through a combination of techniques, including proteomic analysis, electron microscopy, and biochemical crosslinking. The purification method effectively isolated a stable PEP complex with minimal impurities. Negative staining electron microscopy showed well-separated PEP molecules with distinct shapes and minimal interference from other complexes. The analysis revealed distinct 2D classes and a 3D map with certain features, although the exact placement of the *E. coli* RNAP's catalytic core within the PEP-A envelope remains uncertain. By utilising cross-linking mass spectrometry (XL-MS) treatment with a crosslinker (DSBU), 39 interprotein dipeptides, with 12 involving PEP core subunits or PAPs were revealed, indicating their close proximity within the PEP complex. The accessibility of the crosslinker to core subunits suggested that associated PAPs do not completely cover the core. Conformational differences were observed in PEP regions with dipeptides compared to *E. coli* RNAP. More than 400 proteins were identified in the PEP-enriched fraction isolated from chloroplasts, with the most abundant proteins being the core subunits and twelve PAPs. The presence of histones suggested nuclear contamination associated with chloroplast envelopes. The study also identified two fructokinase-like proteins, FLN1 and FLN2, and pTAC18 in all the preparations used. FLN2 interacts with the α subunit of the PEP core enzyme, similar to another protein, PAP5. The α subunit was found to be twice as abundant as the β subunit in the PEP complex, supporting the idea that the PEP core shares similarities with bacterial RNA polymerases. PAP5 and FLN2 seem to associate early in the PEP transformation process, indicating their cooperative role in the complex. Additionally, two related proteins, PAP1 and PAP2, were identified. PAP1 and PAP2 could form a heterodimer within the PEP complex and interact with PAP11/MurE-like proteins. The presence of closely related proteins like PAP6 and FLN2 as well as pTAC18 raises questions about the PEP's subunit composition, indicating that the catalytic core may be associated to 14 PAPs instead of 12. Efforts to fit the PEP structure with known structures of the catalytic core were limited by resolution. The PAPs are closely associated

with the catalytic core, and the overall shape of the PEP differs from other RNA polymerases. Sequence comparisons between the PEP and bacterial RNA polymerases showed conservation in the core but revealed unique features in the PEP's catalytic subunits, suggesting evolutionary divergence. In conclusion, the proteomic analysis provided valuable insights into the composition and interactions of the PEP complex and shed light on its unique characteristics and potential evolutionary adaptations. However, further research is needed to precisely determine the subunit composition and for obtaining a cryo-EM structure. I participated in the chloroplast fractionation from *Sinapis alba.*

International Journal of
Molecular Sciences

MDPI

Article

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

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

1 CNRS, CEA, IBS, University Grenoble Alpes, 38000 Grenoble, France
2 CNRS, CEA, INRAE, IRIG-LPCV, University Grenoble-Alpes, 38000 Grenoble, France
3 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.

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

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

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 β and β′, a dimer of α subunits and a monomer of the ω subunit [1–3]. For specific transcriptional activity, RNAPs require additional proteins such as σ 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 α, β, β′ and β″, 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).

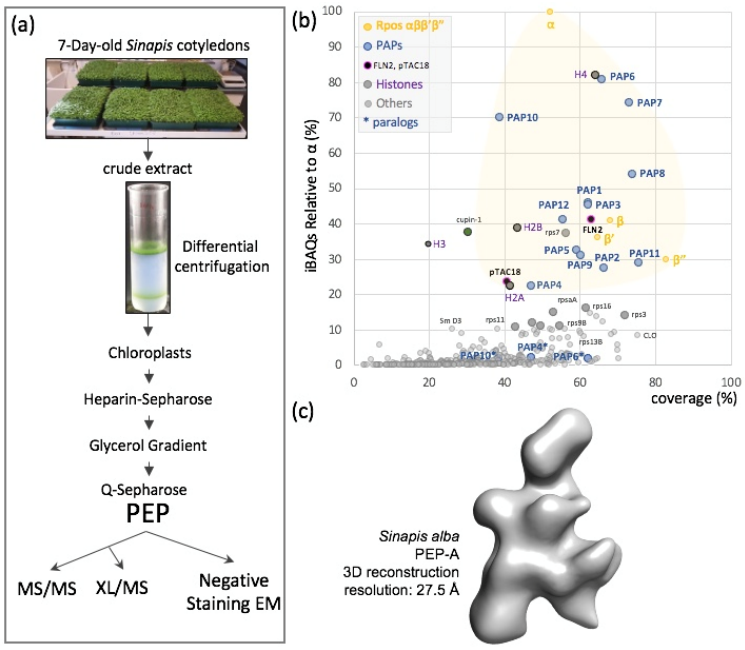


Figure 1. PEP composition and three-dimensional envelope. (**a**) Organelle fractionation, purification scheme and sample processing for mass spectrometry (MS/MS) or crosslinking mass spectrometry (XL/MS) or negative staining electron microscopy (eM). (**b**) Mass spectrometry data presented as relative iBAQ values to that of α (iBAQr) as a function of the corresponding protein coverage expressed in percentage. Subunits α, β, β′ and β″ 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 (α, β, β′ and β″) and the twelve PAPs (Figure 1b and Figure S1). The α subunit was found to be approximately twice more abundant than the β subunit, consistent with a stoichiometry of two α subunits per one β subunit in the catalytic core complex, as described in eubacterial RNAPs. Besides the 16 known PEP subunits, we identified PAP6/FLN1 paralogous 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).

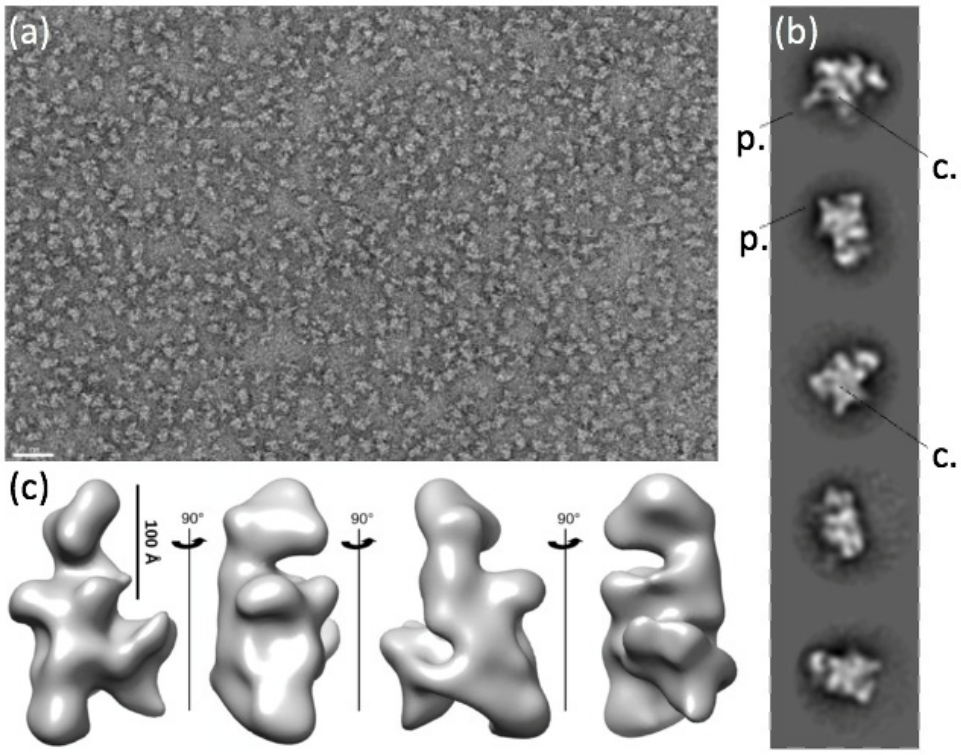


Figure 2. Negative-staining electron microscopy and 3D envelope of the PEP-A complex. (**a**) Overview image of the grid after negative staining. Note the homogeneity of the sample and the lack of other protein complexes. The white scale bar represents 50 nm. (**b**) Two-dimensional classes of PEP. (**c**) 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 α, β, β′ and β″ 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 α subunit being the most divergent. Sequence conservation appears to be high in the domains of the β, β′ and β″ subunits that bear the catalytic activity, while it is lower for the α 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 β′ subunit can be aligned with the β′ and β″ subunits of the PEP (Figure 4b, Figures S4 and S5).

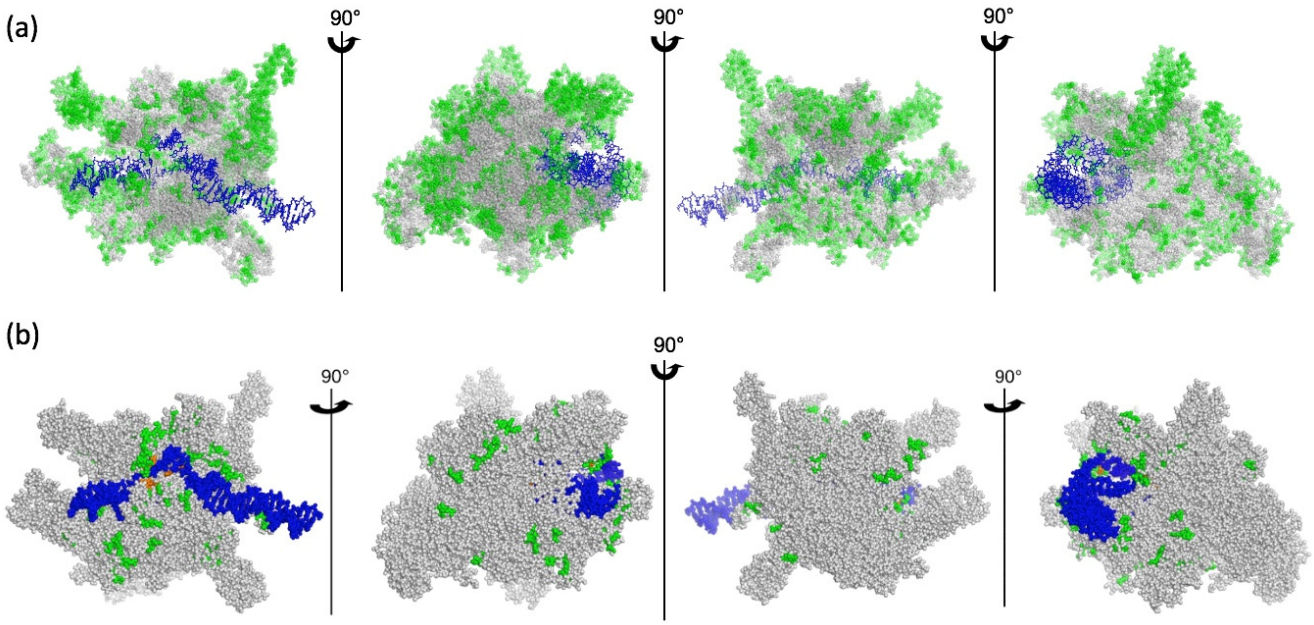


Figure 3. Mapping variable sites of the core subunits. View of the *E. coli* RNAP (PDB entry: 6GH5) without the ω subunit and the σ54 factor. The double-stranded DNA is colored in blue. The core subunits are drawn in spheres. (**a**) Mapping the variable sites as homologous in grey and nonhomologous 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 β and β′ subunits in bRNAPs, it is supported in the PEP by the β, β′ and β″ subunits. Unlike in *E. coli*, the β subunit of the PEP does not have the additional βi4, βi9 and βi11 domains [38]. However, the β″ subunit of the PEP contains a long plant-specific insertion of several hundred residues between regions β′b8 and β′b9 that does not exist either in the β′ subunit from *E. coli* RNAP or in the β′ subunit from *T. thermophilus* RNAP (Figure 4b,c and Figure S5). The β″ subunit of RNAP from angiosperms also lacks a part of the β′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 β″ large insertion that remains invisible in these representations (Figure 3b).

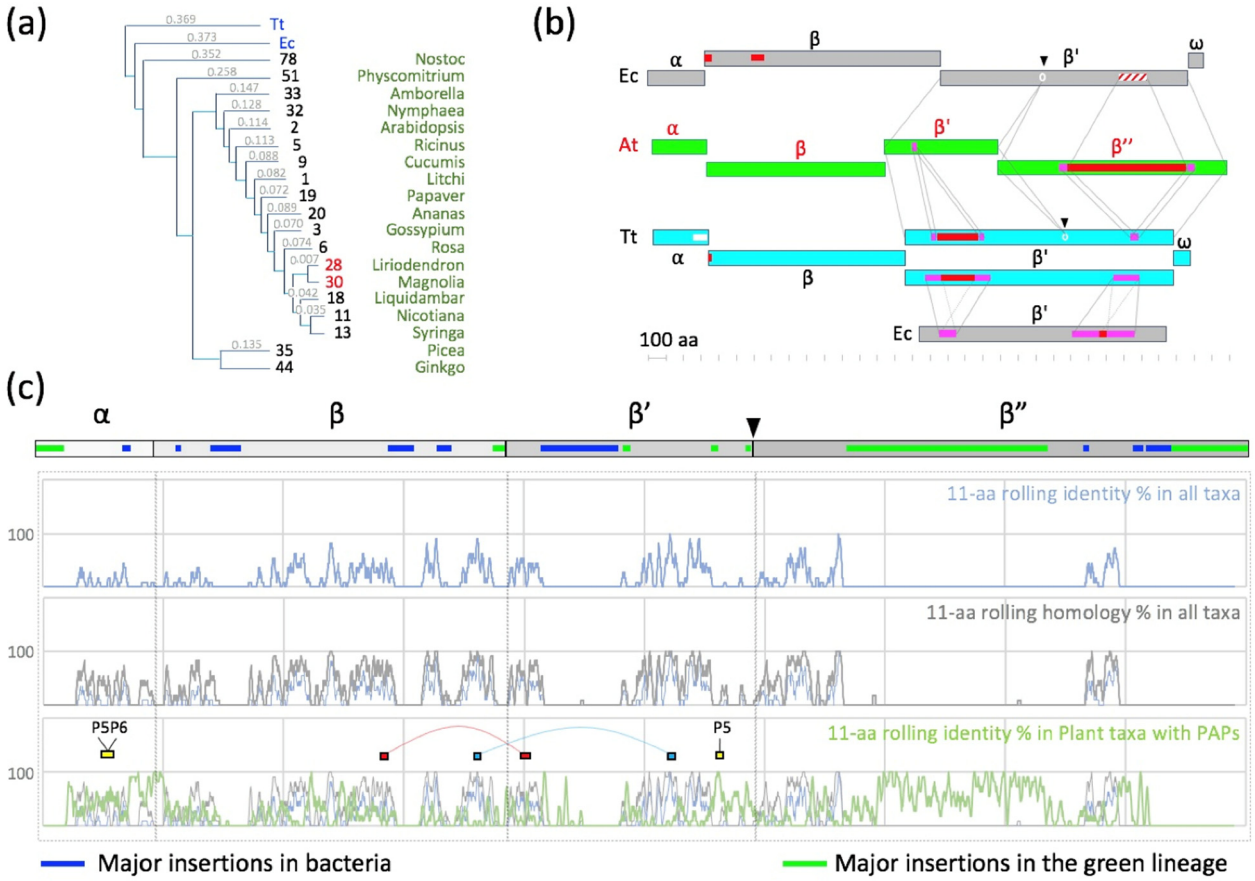


Figure 4. Phylogeny and sequence alignments of the core subunits. (**a**) Phylogram obtained with the Clustal Omega multi-alignment algorithm. Branch length presented as a cladogram. Major taxa included from the collection presented in the data source (Excel sheet sorted). A major incongruence from the 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 β′ in the PEP β′ and β″ 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 β-β′, 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 β and β′ 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 α 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 β′ subunit (Table 1) that belongs to an insertion of conserved residues found only in angiosperms after the β′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 α, β and β′ 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	[IFGPIKSGIBABGNYR]	60	75	Y15	[LTPQVAKESSYAPEDR]	733	748	K7
8	SaRpoC1	SaRpoB	52.00	[FRETLLGKR]	489	497	K8	[SKQGGQR]	969	975	S1
9	PAP6/FLN1; FLN2	PAP5/PTAC12	89.14	[KLELVGSMGEDDDSS}	602	617	K1	[NWSVLKSTPELR]	481	492	K6
10	PAP6/FLN1; FLN2	SaRpoA	121.48	[MLTVQPDLMNDKGYLER]	505	521	Y14	[GYSLKMSNNFEDR]	156	168	K5
11	PAP5/PTAC12	RPS2A; RPS2B	39.00	[APQPAGESSSFPSYGKNPGSR]	128	148	S20	[EVATAIR]	137	143	T4
12	PAP2/pTac2	SPPA	67.62	[GGLFKESEVILSR]	503	515	S7	[GQISDQLKSR]	135	144	K8

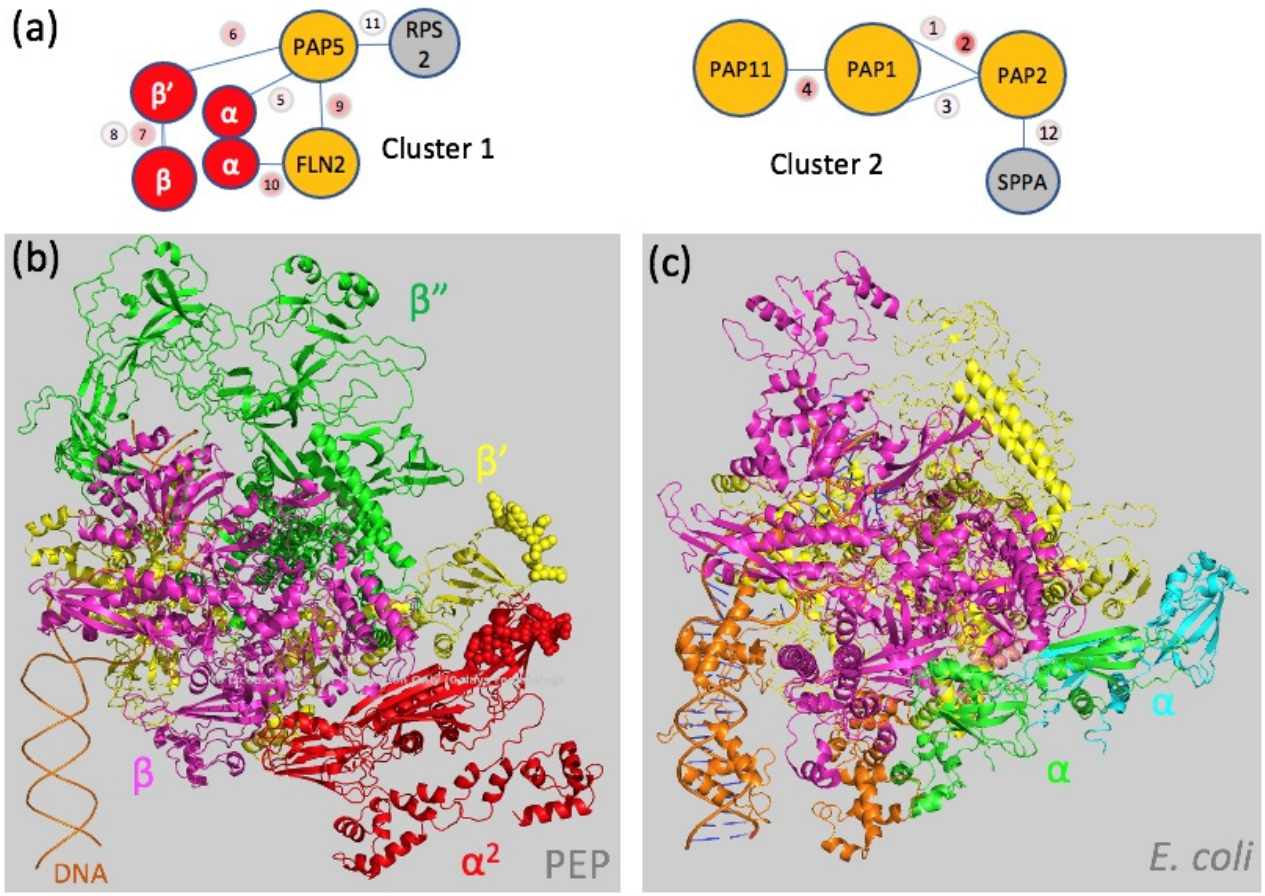


Figure 5. Mapping protein interactions on the core complex. (**a**) Protein clusters determined from the XL-MS analysis (Table 1) are schematically presented with the link and scores; grey bubbles correspond to the protein not belonging to the PEP-A purified complex: RPS2, Ribosomal Protein S2; SPPA, light-inducible chloroplast protease complex associated with thylakoid membranes. Cluster 1 composed of the PAP5, FLN2, α and β′ subunits. Cluster 2 composed of PAP1, 2 and 11. (**b**) Model of the PEP core complex from *A. thaliana* built from the α, β, β′ and β″ subunits modelized using AlphaFold [42] and superimposed onto the *E. coli* RNAP catalytic core and colored as follows: α subunit in red, β subunit in pink, β′ subunit in yellow and β″ in green. The van der Waals spheres display the peptides of the α and β′ 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 α subunit of the catalytic core. Based on the sequence, it is not possible to distinguish which FLN paralog binds to the α subunit due to the high sequence identity between FLN1 and FLN2 that display the same identified peptide sequence. The part of the α 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 α subunit. The MS-based proteomic characterization of the *S. alba* PEP fraction also suggested that the α subunit is twice more abundant than the β subunit. Together, these observations are consistent with a stoichiometry of two α subunits per one β 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 α, β and β′ 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 heterogenous 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 β′ subunit has likely been split into two subunits during evolution after the separation of the eubacteria and cyanobacteria branches, the latest uniquely sharing the β″ subunit with the chloroplast [45]. The sequence alignment showed that the β′ and β″ subunits of the PEP can be, respectively, aligned with the N-terminal and C-terminal parts of the β′ subunit from bRNAPs. In addition, a very long insertion in the β″ subunit of plastids and cyanobacteria (Phe364-Ser1093 in *A. thaliana*) is not observed in the C-terminal part of the β′ 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 × 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-µm nylon mesh, then centrifuged 3 min at 6084.1× *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× *g*, 4 °C. The fractions containing the chloroplasts were then pooled, diluted in homogenization buffer and centrifuged 10 min at 4000× *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× *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× *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× *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 α, β, β′ and β″ 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 α, β, β′ and β″ 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 µm × 5 mm PepMap C18, Thermo Scientific) and separated in a 75 µm × 250 mm C18 column (Reprosil-Pur 120 C18-AQ, 1.9 µm, 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 ≥ 6 amino acids, peptide score ≥ 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 µM 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 °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 ~1:50 was added, and the digestion was performed overnight at 37 °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 µm × 5 mm PepMap C18, Thermo Scientific) and separated using a Pharmafluidics µPAC™ column of 200 cm in length (with a pillar array backbone at

an interpillar distance of 2.5 µm) 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 Å/pixel) under low-dose conditions (<10 e-/Å^2) with defocus values between −1.2 and −2.5 µm 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 Å 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 Å mask diameter) that was then used to calculate a 3D map (C1 symmetry and 320 Å mask diameter) at 27.5 Å 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 archaebacteria. *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 MChevalier, 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]

Enhancing the resolution of the PEP complex for cryo-EM was the subsequent objective, for which a greater quantity of chloroplasts was required for PEP purification. In order to increase the yield of chloroplast fractionated from *Sinapis alba*, certain modifications were implemented to purify the chloroplasts. Specifically, the number of trays containing *Sinapis alba* seedlings was reduced from eight to two trays for convenient handling and controlling wastage. Moreover, the density gradient ultracentrifugation step involved utilising different concentrations of percoll such as 20%, 40% and 80% percoll as opposed to the previous approach, which employed only 40% and 80% density layers. The 20% percoll gradient served as a layer facilitating uniform accumulation of chloroplasts based on their structure and size, before entering into the 40% and 80% percoll gradient layers. Subsequently, the intact chloroplasts were collected (Figure 3.1 F), flash frozen in liquid nitrogen and stored at -80°C. This resulted in isolation of approximately 30 mL of unbroken chloroplasts each time while following these modifications and were ready to be used for PEP purification.

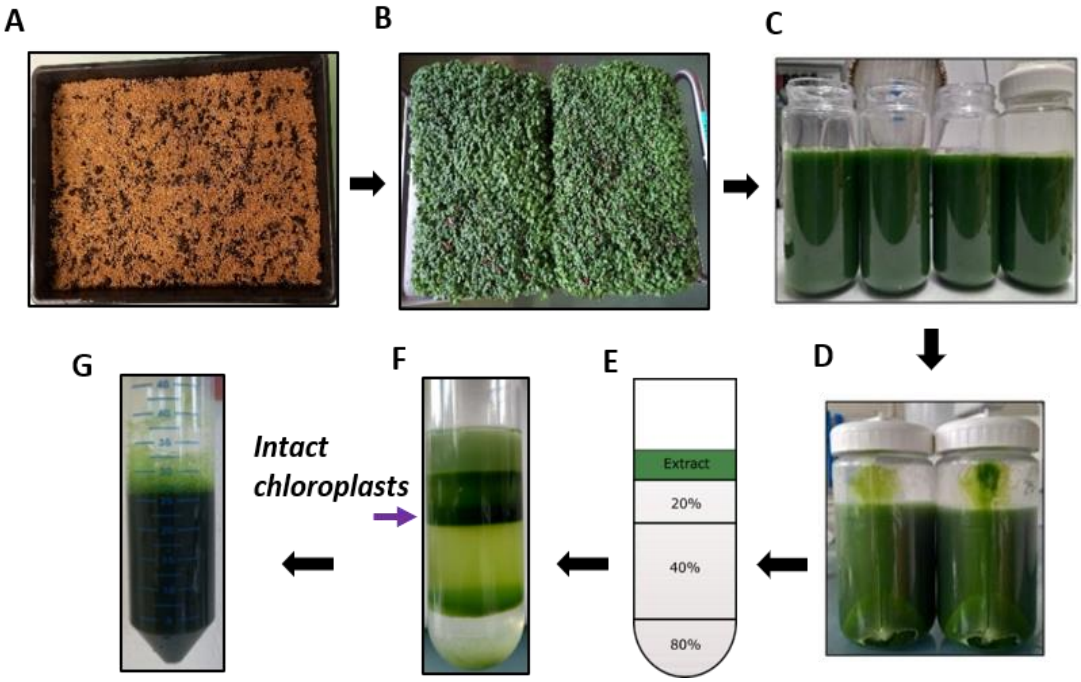


Figure 3.1: *Chloroplast fractionation from Sinapis alba**.** (A) Sinapis alba seeds sowed on day 1. (B) Sinapis alba on day 6 before harvesting. (C) The harvested cotyledons were homogenized in HM. (D) The suspension after centrifugation. The supernatant was discarded and the pellet was resuspended in HM buffer. (E) Schematic diagram of Percoll density gradient that was used to fractionate chloroplasts. (F) Percoll density gradient after centrifugation. The layer containing intact chloroplasts is indicated by arrowhead. (G) Chloroplast suspension suspended in lysis buffer was flash frozen in liquid nitrogen and stored at -80°C.*

3.2. SUPEROXIDE DISMUTASE ACTIVITY OF PAP4 AND PAP9

According to the findings of Steiner et al, 2011, PAP4 was identified to be in a trimeric state based on its electromobility in SDS-PAGE. In addition to a monomer of PAP9, four superoxide dismutases were therefore suspected to be present in the PEP complex. This discovery raises the question of whether their role within the PEP is primarily structural or if they serve to neutralise the reactive oxygen species that are generated during initial photosynthetic reactions. Consequently, experiments to measure their superoxide dismutase activity in *in vitro* purified PAP4 and PAP9 proteins and structural characterisation of PAP9 was conducted (Favier et al, 2021).

3.2.1. Purification of PAP4 and PAP9

PAP4 and PAP9 were purified with a yield of approximately 4500 mg/L and 1540 mg/L. The expected molecular weight for PAP4 is 26.9 kDa and PAP9 is 30.8 kDa. In gel filtration analyses, a single peak was observed at an elution volume of 78 - 80 mL for PAP4 corresponding to its expected molecular weight, with no aggregates and a peak at 80 mL for PAP9 corresponding to its expected size, along with peaks denoting some aggregates at 45 mL. Purity accessed by SDS-PAGE analyses showed that expected molecular weights were obtained and the proteins were pure for further studies (Figure 3.2A, C).

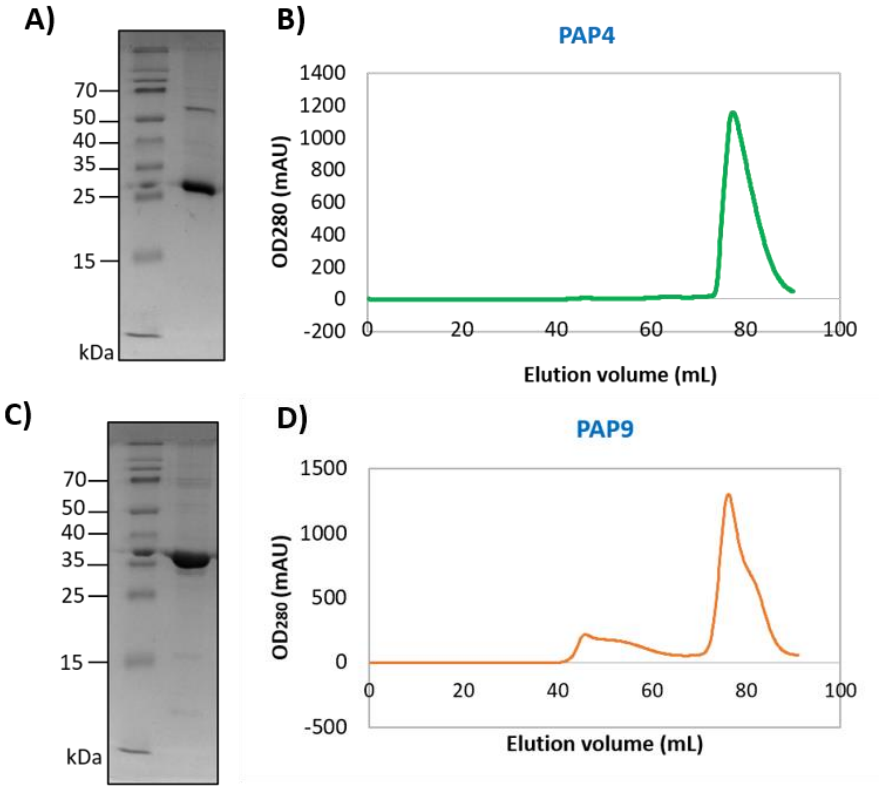


***Figure 3.2**: SDS PAGE gels showing purified PAP4 (A) and PAP9 (B) after size-exclusion chromatography. Profiles of PAP4 (B) and PAP9 (D) after size exclusion chromatography.*

3.2.2. Estimation of PAP4 and PAP9 superoxide dismutase activity

The superoxide dismutase activity for PAP4 and PAP9 was tested using pyrogallol. Pyrogallol undergoes auto-oxidation and leads to the formation of a yellow-coloured product, purpurogallin that absorbs at 420 nm. Upon the presence of the purified SOD enzyme, the auto-oxidation of pyrogallol is inhibited and a decrease in absorbance should be observed. In the experiment, the auto-oxidation of pyrogallol was allowed to occur for 3 minutes and the respective protein, either PAP4 or PAP9 are added. The absorbances were plotted against time and a control, standard Mn-SOD of 5 µm was used as it is a commercially available purified superoxide dismutase (Figure 3.3). The results showed that PAP9 has no SOD activity and PAP4 has little SOD activity corresponding to standard control. PAP9 shows little SOD activity at a very high concentration of 1 mM.

The crystal structure of PAP9 was solved by Dr. David Cobessi, GSY group at IBS. The structure of PAP9 was symmetrically dimeric (Favier et al., 2021). However, the buried catalytic site revealed zinc ion instead of an iron (Figure 3.4B). The superoxide dismutase activity tests performed on purified PAP9 showed no prominent activity, as zinc exists only as Zn^{2+} in its

redox state. That could explain why the PAP9 superoxide dismutase activity reported is weak in comparison with PAP4 which displays activity in the experiment. At high PAP9 concentration (1 mM), a weak SOD activity was observed suggesting that some PAP9 molecules have iron.

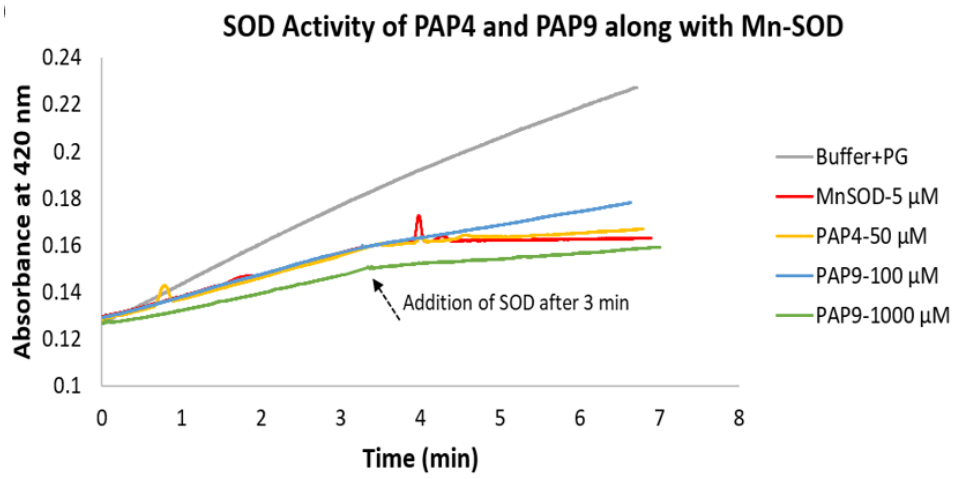


***Figure 3.3**: Graph showing SOD activity of PAP4 and PAP9. The SOD enzyme was added after 3 minutes to the reaction mixture containing pyrogallol as indicated by the arrow. 5 µM Mn-SOD (red) was used as a positive control and displays SOD activity. The buffer and pyrogallol mixture (grey) were used as a negative control. PAP9 at 1000 µM (green) shows very less SOD activity compared to PAP4 at 50 µM (yellow) and 100 µM (blue).*

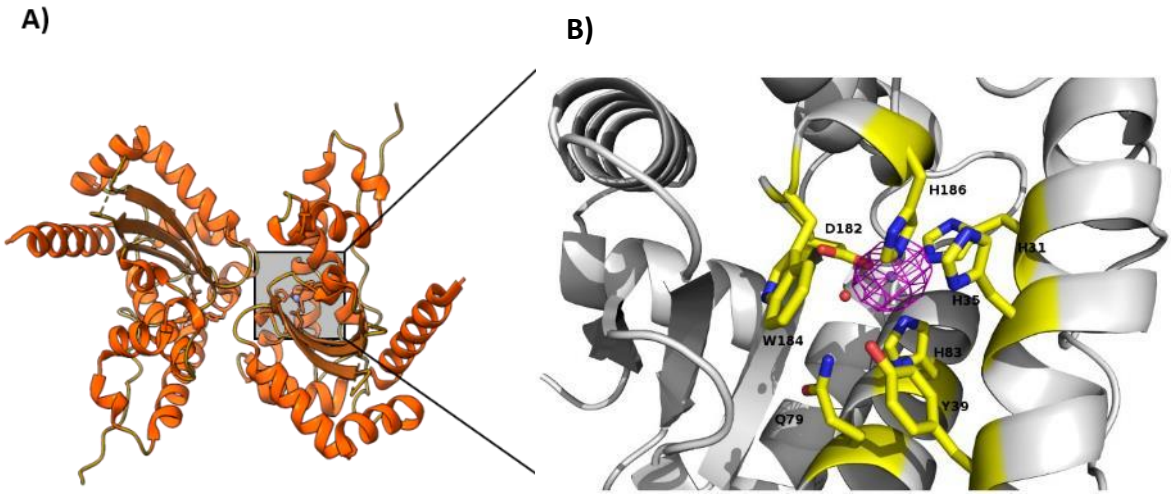


***Figure 3.4**: (A) 3D model of PAP9 in its dimeric state from X-ray diffraction. (B) View of the catalytic active site of PAP9. The catalytic active residues are shown as sticks in yellow. Anomalous electron density corresponding to zinc ion is shown in magenta and the water molecule in red. Source: Favier et al, 2021*

The catalytically active residues are displayed in Figure 3.4. The following experiments to be conducted would be to test the functionality *in planta* and mutate the SOD catalytic site and check if the phenotype of greening is restored or not, since deletion of PAPs have been reported to exhibit albino phenotype.

Article 2: The plastid-encoded RNA polymerase-associated protein PAP9 is a superoxide dismutase with unusual structural features

The study Favier et al, 2021 involves the analysis of PAP9 in the context of its evolution, subcellular localisation, structure, and dynamics. From the phylogeny analysis of PAP9, sequence similarities with At-PAP9 were identified in various clades, including salt-water algae (chlorophytes), suggesting that plastid-localized superoxide dismutases (SODs) were acquired early in evolution. Notably, the C-terminal domain of PAP9 has undergone significant changes during evolution, with some species having substantial insertions and alterations in this region. The presence and characteristics of the C-terminal domain of PAP9 vary across different plant clades, indicating that it may have different roles in various species. PAP9 was found to be mostly localized within plastids, but some of it were also detected in stromules, which are dynamic tubular structures connecting plastids. There were occasional signals in the cytosol and nucleus, suggesting a degree of cytoplasmic and nuclear localisation. The localisation of PAP9 may be influenced by the fusion of GFP to its C-terminus, which could affect its functionality. Mass spectrometry was used to assess the mass of PAP9-6His and $^{15}N,^{13}C$-6His-PAP9 under denaturing conditions. The experimental mass of PAP9-6His matched its amino acid sequence, while the $^{15}N,^{13}C$-6His-PAP9 showed an incomplete labelling. Native mass spectrometry detected monomers and dimers of PAP9, suggesting potential oligomeric states. The X-ray crystallography revealed the structure of PAP9, indicating two domains similar to those observed in FeSODs or MnSODs. A zinc ion was found in the catalytic centre instead of the expected iron ion, suggesting potential differences in catalytic activity. The C-terminal part of PAP9, a flexible region, was not observed in the crystal structure and was suggested to behave dynamically. Solution-State NMR analyses were conducted to investigate the structural and dynamic properties of PAP9, particularly the regions not observed in the crystal structure. Some segments of the protein exhibited fast dynamics, and their behaviour was similar to that of free peptides or small proteins. The C-terminal tail of PAP9 was found to

be dynamic, lacking any defined secondary structure. Comparisons with other SODs from the Protein Data Bank (PDB) revealed that the fold of PAP9, the ligands involved in metal coordination, and residues closing the catalytic site are conserved. While PAP9 shares structural similarities with other SODs, the presence of a zinc ion in its catalytic centre sets it apart. I performed the superoxide dismutase activity test using pyrogallol method on PAP9, which displayed very little activity only at 1mM.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 30 June 2021
doi: 10.3389/fpls.2021.668897

The Plastid-Encoded RNA Polymerase-Associated Protein PAP9 Is a Superoxide Dismutase With Unusual Structural Features

Adrien Favier[1], Pierre Gans[1], Elisabetta Boeri Erba[1], Luca Signor[1], Soumiya Sankari Muthukumar[1], Thomas Pfannschmidt[2†], Robert Blanvillain[2] and David Cobessi[1*]*

[1] Université Grenoble Alpes, CEA, CNRS, IBS, Grenoble, France, [2] Université Grenoble-Alpes, CNRS, CEA, INRA, IRIG-LPCV, Grenoble, France

In Angiosperms, the plastid-encoded RNA polymerase (PEP) is a multimeric enzyme, essential for the proper expression of the plastid genome during chloroplast biogenesis. It is especially required for the light initiated expression of photosynthesis genes and the subsequent build-up of the photosynthetic apparatus. The PEP complex is composed of a prokaryotic-type core of four plastid-encoded subunits and 12 nuclear-encoded PEP-associated proteins (PAPs). Among them, there are two iron superoxide dismutases, FSD2/PAP9 and FSD3/PAP4. Superoxide dismutases usually are soluble enzymes not bound into larger protein complexes. To investigate this unusual feature, we characterized PAP9 using molecular genetics, fluorescence microscopy, mass spectrometry, X-ray diffraction, and solution-state NMR. Despite the presence of a predicted nuclear localization signal within the sequence of the predicted chloroplast transit peptide, PAP9 was mainly observed within plastids. Mass spectrometry experiments with the recombinant *Arabidopsis* PAP9 suggested that monomers and dimers of PAP9 could be associated to the PEP complex. In crystals, PAP9 occurred as a dimeric enzyme that displayed a similar fold to that of the FeSODs or manganese SOD (MnSODs). A zinc ion, instead of the expected iron, was found to be penta-coordinated with a trigonal-*bipyramidal* geometry in the catalytic center of the recombinant protein. The metal coordination involves a water molecule and highly conserved residues in FeSODs. Solution-state NMR and DOSY experiments revealed an unfolded C-terminal 34 amino-acid stretch in the stand-alone protein and few internal residues interacting with the rest of the protein. We hypothesize that this C-terminal extension had appeared during evolution as a distinct feature of the FSD2/PAP9 targeting it to the PEP complex. Close vicinity to the transcriptional apparatus may allow for the protection against the strongly oxidizing aerial environment during plant conquering of terrestrial habitats.

Keywords: plastid-encoded RNA polymerase, iron superoxide dismutase, chloroplast biogenesis, NMR, X-ray crystallography

OPEN ACCESS

Edited by:
Ning Li,
Hong Kong University of Science and Technology, China

Reviewed by:
Guang Zhu,
Hong Kong University of Science and Technology, China
Takashi Shiina,
Kyoto Prefectural University, Japan

****Correspondence:***
Robert Blanvillain
robert.blanvillain@cea.fr
David Cobessi
david.cobessi@ibs.fr

†Present address:
Thomas Pfannschmidt, Pflanzenphysiologie, Institut für Botanik, Leibniz-Universität Hannover, Hanover, Germany

Specialty section:
This article was submitted to Plant Proteomics and Protein Structural Biology, a section of the journal Frontiers in Plant Science

Received: *17 February 2021*
Accepted: *28 May 2021*
Published: *30 June 2021*

Citation:
Favier A, Gans P, Boeri Erba E, Signor L, Muthukumar SS, 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. doi: 10.3389/fpls.2021.668897

INTRODUCTION

In the green lineage, the photosynthetic reactions in the chloroplast convert light energy into chemical energy with the release of di-oxygen. Other metabolic pathways take place in chloroplasts such as the biosynthesis of amino acids, fatty acids, vitamins, and hormones. Hence, the chloroplast functions sustain most life forms on Earth (Jarvis and López-Juez, 2013). According to the endosymbiosis theory of the origin of organelles, chloroplasts have evolved from a single ancient cyanobacterium engulfed around 1.5 billion years ago into a mitochondriate proto-eukaryote (Bobik and Burch-Smith, 2015). During evolution, a massive gene transfer occurred from the cyanobacterium into the nucleus of the host cell (Martin et al., 2002). Thus, the nuclear genome could encode from 1500 to 4500 chloroplast proteins whereas the plastid genome (plastome) encodes for about hundred proteins (Zybailov et al., 2008). The plastome (cpDNA) mainly encodes: (1) components of the plastid gene expression machinery (RNA polymerase, ribosomal proteins, tRNAs, and rRNAs), (2) subunits of each major functional photosynthesis-related complex (e.g., RuBisCO, Photosystem I and II, the cytochrome b_6f complex, NADPH dehydrogenase, and ATP synthase), and (3) a few proteins involved in other processes (e.g., ClpP1 and YCF3) (Sugiura, 1992; Majeran et al., 2012; Yu et al., 2014). Hence, the vast majority of chloroplast proteins are encoded by the nuclear genome. The pre-proteins are imported into the chloroplast from the cytosol mainly by the TOC-TIC machinery of the chloroplast envelope that recognizes and cleaves specific transit peptides (cTPs) at their N-terminal extremity (Jarvis, 2008). Once in the stroma, the proteins are properly folded. Since most of the protein complexes in the chloroplast contain nuclear and chloroplast-encoded proteins, coordination in expression of both genomes is essential (Liebers et al., 2017).

Two RNA polymerases are involved in plastid transcription: a nuclear-encoded RNA polymerase (NEP) and the plastid-encoded RNA polymerase (PEP). The NEP, a T3–T7 bacteriophage type RNA polymerase, transcribes the *rpo* genes (rpoA, B, C1, and C2), encoding the four subunits of the catalytic core of the PEP, and other housekeeping genes (Kremnev and Strand, 2014; Börner et al., 2015). During chloroplast biogenesis, the PEP core is reshaped in a multi-subunit RNA-polymerase of at least 16 different proteins (MW: ∼1 MDa), which mainly transcribes photosynthesis related genes. The active PEP complex is composed of four rpo core subunits, and 12 nuclear-encoded PEP-associated proteins (PAPs) (Pfannschmidt et al., 2015). Mutations in most of the *pap* genes yield albino/ivory plants incapable of photosynthesis with a defect in the expression of PEP-dependent genes indicating that the PEP is not fully functional (Pfannschmidt et al., 2015). This shared phenotype triggered the idea of a PAPs-related developmental block corresponding to an epistasis effect. This effect occurs when all components are required for the stability of the entire complex ensuring that photosynthesis could be launched only if all the functions are present (Liebers et al., 2018).

The PAPs can be divided into four groups according to their hypothetical functions (Yu et al., 2014). PAP sequence analyses and biochemical studies allowed to characterize four PAPs with potential known catalytic activities: PAP4, PAP7, PAP9, and PAP10. PAP7 belongs to methyltransferases (Gao et al., 2011), PAP10 is a thioredoxin (TrxZ) (Steiner et al., 2011) while PAP4 (FSD3) and PAP9 (FSD2) are both iron superoxide dismutases (FeSOD) (Myouga et al., 2008). Formation of superoxide radicals mainly occurs in electron transport chains of photosynthesis and respiration. Therefore, PAP4 and PAP9 may serve as protection against oxidative stresses generated during the first activities of the photosynthetic apparatus (Pfannschmidt et al., 2015). Indeed, superoxide radicals can damage sulfur containing amino acids, metals, and Fe-S clusters. SODs are cellular defenses against superoxide by catalyzing the dismutation of superoxide into hydrogen peroxide according to the overall reaction: $2O_2^- + 2H^+ \rightarrow H_2O_2 + O_2$ (Pfannschmidt, 2003; Abreu and Cabelli, 2010).

Besides the MnSODs and the copper-zinc SODs (Cu/ZnSODs, where Cu is the redox center), three iron superoxide dismutases (FeSODs) were characterized in plants. Dimeric MnSODs are found in the matrix of the mitochondria, with one Mn ion per monomer. Cu/ZnSODs are dimeric SODs found in the cytosol, peroxisomes, and plastids. Each monomer contains one Cu and one Zn ion. FeSODs are dimeric enzymes with one iron ion bound to each monomer. The fold of the FeSOD monomer is roughly similar to that of the MnSOD monomer and is completely different from the Cu/ZnSODs (Pilon et al., 2011). In plants, FSD1 is a cytoplasmic FeSOD, while PAP4 and PAP9 are FeSODs only observed in the chloroplast, both associated to the PEP (Myouga et al., 2008; Steiner et al., 2011). Surprisingly, the oligomeric assembly of PAP4 and PAP9 differ from that observed for FeSODs. PAP9 was reported as being a monomer in the PEP and PAP4 as a trimer (Steiner et al., 2011). In *Arabidopsis thaliana*, PAP4 and PAP9 could form a heterodimeric complex in the chloroplast nucleoids (Myouga et al., 2008). The *pap4–pap9* double mutant displayed an albino phenotype with no chloroplast development while the *pap4* or *pap9* single inactivation mutants showed pale green phenotypes and sensitivity to oxidative stress indicating some compensation effect but no full redundancy between the two proteins (Myouga et al., 2008). These observations strongly suggested that a heterodimeric complex PAP4/PAP9 could protect the transcriptionally active chromosome (TAC) during the early stages of chloroplast development from the superoxide radical produced during photosynthesis in the thylakoid membranes (Myouga et al., 2008). To better characterize PAP9 and understand how plastid-localized FeSODs were embedded in the PEP, we studied PAP9 using phylogenetic approaches, *in planta* experiments, mass spectrometry, X-ray diffraction, and solution-state NMR.

MATERIALS AND METHODS

Accessions

PAP9 At5g51100; accessions from the green lineage are given in **Supplementary Table 1**. Full-length coding sequences were retrieved from Blastp (**Supplementary Table 2**). The protein

sequences were aligned using Clustal Omega[1]. The prediction of chloroplast pre-sequences (**Supplementary Table 3**) were established using ChloroP[2] (Emanuelsson et al., 1999). The predictions of the nuclear localization signals (NLS) were performed using NLS_Mapper[3] and are given in **Supplementary Table 4** (Kosugi et al., 2009). Clustal Omega color-code as followed: [red (AVFPMILW): small + hydrophobic (includes aromatic Y); blue: (DE), acidic; magenta: (RHK), basic; green: (HSTYHCNGQ), hydroxyl + sulfhydryl + amine + G].

Peptide Synthesis

The peptide ^{226}QREQEGTETEDEENPDDEVPEVYLDSDIDVSE VD^{259} corresponding to the last 34 residues of PAP9 was synthesized by Proteomic Solution with a purity (HPLC) of 98.29%. Its molecular mass (MW: 3925.85 Da) was checked using mass spectrometry.

Transient Transformation of Onion Cells

Gold Carrier Particles (Seashell technology) were coated with 1 μg of the expression vector and 1 μg of an internal control such as PAP10-RFP (Liebers et al., 2020). Gold particles were delivered into onion cells using a particle gun (BioRad). The transformed cells were allowed to express the construct for 16–24 h before fluorescence observation using proper filters. Signal profiles of the two fluorescence channels were acquired on pictures using ImageJ.

Cloning and Vector Construction

PAP9$^{\Delta cTP}$ (271 aa/31 kDa) in pBB408 corresponds to PAP9$^{\Delta cTP}$-6His in the pEt21d backbone: RT-PCR fragment was obtained from seedling cDNA amplified with oP9ΔcTP_FNco (5′-CCATGGGTGTTATCACAGCTGG)/oP9_RNot (5′-GCGGCC GCGTCAACCTCAGATACATCGATG), A-tailed and cloned in pGem-Teasy (pBB399a) then digested with *Nco*I, *Not*I and cloned in pET21d. PAP9-GFP in pAF04 (pEZS-NL backbone, Stanford): RT-PCR fragment was obtained from seedling cDNA amplified with oPAP9_FXho (5′-CTC GAGATGATGAATGTTGCAGTGACAGCC) and oPAP9_ RBH (5′-GGATCCCCGTCAACCTCAGATACATCGATGTCAC) cloned as above then digested with *Xho*I *Bam*HI and ligated in pEZS-NL. pBB301 (PA10-RFP) was used as internal control (Liebers et al., 2020).

Protein Expression and Purification

PAP9-6His (for ΔcTP-PAP9-6His) was overexpressed in *E. coli* Rosetta2 strain in LB with 100 μg/mL ampicillin and 34 μg/mL chloramphenicol. 6His-PAP9 (for ΔcTP-6His-PAP9) was overexpressed in *E. coli* Rosetta2 strain in LB with 100 μg/mL ampicillin and 50 μg/mL kanamycin. Cells were grown overnight in 50 mL of LB with antibiotics at 37°C. One liter of LB (with antibiotics) was then inoculated with the first culture to reach an initial OD_{600} of 0.1. Growth was continued at 37°C. When the OD_{600} reached 0.6, the temperature was decreased to 16°C and isopropyl β-D-1-thiogalactopyranoside was added to give a final concentration of 0.5 mM. After an overnight induction, bacteria were harvested at 6619 g, for 25 min, at 4°C. The cell pellet was resuspended in 30 mL of lysis buffer (50 mM Tris–HCl pH 8.0, 0.5 M NaCl, 20 mM imidazole) containing a Complete Protease inhibitor Cocktail tablet (Roche). The lysate was centrifuged at 15,000 *g*, for 40 min, at 4°C. The purification was performed at room temperature. The supernatant was applied onto a NiNTA column in 50 mM Tris–HCl pH 8.0, 0.5 M NaCl, 20 mM imidazole. Proteins were eluted in one step in a buffer containing 50 mM Tris–HCl pH 8.0, 0.1 M NaCl, 300 mM imidazole. Then the eluate was diluted 2 times in 50 mM Tris–HCl pH 8.0 and loaded on a MonoQ column. Elution was performed using a linear NaCl gradient from 0 to 1 M in 50 mM Tris–HCl pH 8.0. The fractions containing PAP9-6His or 6His-PAP9 were pooled and concentrated with an Amicon Ultra 4 mL centrifugal filter and a 10 kDa membrane cut-off before loading on a HiLoad 16/60 Superdex 200 and then eluted with 10 mM Tris–HCl pH 8.0, 50 mM NaCl. The fractions containing the pure protein were pooled and concentrated for further experiments or stored at −20°C with 50% (v/v) glycerol.

^{15}N,^{13}C-6His-PAP9 was expressed in minimum media M9 supplemented with $^{15}NH_4Cl$, ^{13}C-glucose and antibiotics. Briefly, 5 mL of LB were inoculated with *E. coli* Rosetta2 stock glycerol overexpressing 6His-PAP9. After 10 h of growing, 1 mL was added to 100 mL of minimum media supplemented as described above. After 1 night growing, when OD_{600} was close to 2, the overnight culture was centrifuged to inoculate 1 L of minimum media M9 supplemented with $^{15}NH_4Cl$ and ^{13}C-glucose and antibiotics. Cell growth, overexpression and purification followed the procedure described above for 6His-PAP9 and PAP9-6His.

Enzymatic Assays

The superoxide dismutase activity of PAP9 was tested using pyrogallol. The pyrogallol auto-oxidation is characterized by increase of absorbance at 420 nm and superoxide dismutase inhibits the pyrogallol auto-oxidation. Briefly, 7 mM pyrogallol was dissolved in a *Tris*-succinate-EDTA buffer pH 8.2 and the pyrogallol auto-oxidation was followed by monitoring the absorbance increase at 420 nm. After 180 s, PAP9 at several concentrations (50, 100, 200, 500 μM, and 1 mM) or 5 μM Mn-SOD were added into the medium and the absorbance was monitored for further 3 min. Experiments were repeated three times for each concentration and the curves were plotted. Each curve correspond to the average of three enzymatic assays (**Supplementary Figure 1**).

LC/ESI and Native Mass Spectrometry

Liquid chromatography electrospray ionization mass spectrometry (LC/ESI-MS) was used to assess the masses of the intact PAP9-6His, and ^{15}N,^{13}C-6His-PAP9. All solvents were HPLC grade (Chromasolv, Sigma-Aldrich) and trifluoroacetic acid (TFA) was from Acros Organics (puriss, p.a.). Solvent A was 0.03% TFA in water, solvent B contained 95% acetonitrile, 5% water, and 0.03% TFA. A 6210 LC/ESI-TOF mass spectrometer interfaced with an HPLC binary pump system (Agilent Technologies) was used. The mass spectrometer was calibrated

[1]https://www.ebi.ac.uk/Tools/msa/clustalo/

[2]http://www.cbs.dtu.dk/services/ChloroP/

[3]http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi

in the mass-to-charge (*m/z*) range 300–3000 using a standard calibrant (ESI-L, low concentration tuning mix, Agilent Technologies) before the measurements of protein samples. MS acquisition was carried out in positive ion mode and mass spectra were recorded in the 300–3200 *m/z* range. ESI source temperature was set at 573 K, nitrogen was used as drying gas (7 L/min) and as nebulizer gas (10 psi). The capillary needle voltage was set at 4000 V. Spectra acquisition rate was of 1.03 spectra/s. The MS spectra were acquired and the data processed with MassHunter workstation software (v. B.02.00, Agilent Technologies) and with GPMAW software (v. 7.00b2, Lighthouse Data, Denmark). Immediately before the MS analysis, the protein samples were diluted to a final concentration of 8 μM using solvent A. Samples were kept at 10°C in the autosampler and 8 μL of each sample were injected into the system. They were first trapped and desalted on a reverse phase-C8 cartridge (Zorbax 300SB-C8, 5 μm, 300 μm ID × 5 mm, Agilent Technologies) for 3 min at a flow rate of 50 μL/min with 100% solvent A and then eluted and separated on a RP-HPLC column (Jupiter Proteo, 4 μm, 90 Å, 1 mm ID × 50 mm, Phenomenex) using a linear gradient from 5 to 95% solvent B in 15 min.

PAP9-6His was also analyzed by native MS (Boeri Erba and Petosa, 2015; Boeri Erba et al., 2020). Protein ions were generated using a nanoflow ESI (nano-ESI) source. Nanoflow platinum-coated borosilicate ESI capillaries were bought from Thermo Electron SAS (Courtaboeuf, France). MS analyses were carried out on a quadrupole time-of-flight mass spectrometer (Q-TOF Ultima, Waters Corporation, Manchester, United Kingdom). The instrument was modified for the detection of high masses (Sobott et al., 2002; van den Heuvel et al., 2006). The following instrumental parameters were used: capillary voltage = 1.2–1.3 kV, cone potential = 40 V, RF lens-1 potential = 40 V, RF lens-2 potential = 1 V, aperture-1 potential = 0 V, collision energy = 30–140 V, and microchannel plate (MCP) = 1900 V. All mass spectra were calibrated externally using a solution of cesium iodide (6 mg/mL in 50% isopropanol) and were processed with the Masslynx 4.0 software (Waters Corporation, Manchester, United Kingdom) and with Massign software package (Morgner and Robinson, 2012).

Solution-State NMR

One milligram of the 34 amino-acids C-terminal peptide of PAP9 was dissolved in 25 mM Na phosphate, pH 6.5 to a final concentration of 1 mM. For assignment of the peptide, homonuclear TOCSY, NOESY, and sensitivity-enhanced ^{13}C-HSQC experiments were recorded at 25°C on a Bruker ADVANCE III spectrometer operating at ^{1}H frequency of 600 MHz and equipped with a triple resonance pulsed field gradient cryoprobe.

For assignment of 6His-PAP9, 100 μM of ^{15}N,^{13}C-6His-PAP9 in a 90:10 H_2O:D_2O, 10 mM Tris pH 8.0, 50 mM NaCl were used. Heteronuclear 3D Best-TROSY-HNCA, Best-TROSY-HNCACB, Best-TROSY-HNCOCANH (Favier and Brutscher, 2011; Solyom et al., 2013), sensitivity-enhanced ^{13}C-HSQC and ^{15}N-SOFAST experiments were recorded at 298 K on Bruker ADVANCE III HD spectrometers operating either at ^{1}H frequency of 600 or 700 MHz and equipped with a triple resonance pulsed field gradient cryoprobe. [^{15}N,^{1}H]-TRACT (to estimate the global correlation time) (Lee et al., 2006) and DOSY experiments (for measuring the translational diffusion) (Morris and Johnson, 1992) were recorded at 298 K on a Bruker ADVANCE III HD spectrometer operating at ^{1}H frequency of 700 MHz.

Crystallization, Data Collection, and Structure Resolution

6His-PAP9 and PAP9-6His at 5 mg/mL in 10 mM Tris–HCl, pH 8.0, 50 mM NaCl (+10% glycerol for 6His-PAP9) were subjected to crystallization using the sitting-drop vapor-diffusion technique and the high throughput crystallization facility at the EMBL, Grenoble, at 4°C. Crystallization hits were optimized using Limbro plates, at 293 K. Crystals of PAP9-6His were grown in PEG3350 from 15 to 19%, 0.1 M Bis–Tris pH 6.5, 0.2 M $NaNO_3$, for data collection. Crystals of 6His-PAP9 were grown in Bis–Tris pH 7.5, PEG3350 18%, 0.2 M $NaNO_3$.

Diffraction data for PAP9-6His were collected on ID23-1 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, at 100 K, using a PILATUS detector and two crystals. Anomalous data at the peak and after the peak of the zinc K-edge for PAP9-6His and native data for 6His-PAP9 were collected on FIP-BM30A (Roth et al., 2002) at the ESRF, at 100 K, using an ADSC 315r detector. Diffraction data (**Table 1**) were processed and scaled using XDS (Kabsch, 2010).

Phasing was performed by molecular replacement using Phaser (McCoy et al., 2007) from CCP4 (Collaborative Computational Project, Number 4 (CCP4), 1994). To calculate the phases, the crystal structure of the eukaryotic FeSOD from *Vigna unguiculata* (PDB entry: 1UNF) (Muñoz et al., 2005) was used as a model after modifications based on sequence alignment with PAP9 from *A. thaliana* using CHAINSAW (Stein, 2008) from CCP4. The refinements and rebuilding were done using PHENIX (Adams et al., 2010) and COOT (Emsley et al., 2010), respectively. The model refinements were performed with the non-crystallographic symmetry and the water molecules were added using PHENIX in the last stages of the refinement. Refinement statistics are summarized in **Table 2**. Atomic coordinates and X-ray data for PAP9-6His were deposited in the PDB with the accession number 7BJK. Since 6His-PAP9 is similar to PAP9-6His, the diffraction data and the 3D-structure were not reported in the PDB.

RESULTS

Phylogeny of PAP9 in the Green Lineage

Significant sequence similarities with At-PAP9 were found as early as in clades representing the chlorophytes, indicating that salt-water algae acquired plastid-localized SODs early in evolution. However, sequence alignments (**Figure 1**) identified a critical domain, outside of the SOD catalytic domain (**Figure 2A**), at the C-terminal (C-ter) of the protein, which had strongly changed during evolution. Whereas absent in early separated clades (as represented by Chlamydomonas), a significant insertion after the last well-conserved arginine

TABLE 1 | Statistics of data collection.

	PAP9-6His	PAP9-6His	PAP9-6His
Wavelength (Å) and beamline	0.976250 (ID23-1)	1.280867 (FIP-BM30A)	1.284809 (FIP-BM30A)
Resolution range (Å)	48.20–2.25 (2.31–2.25)	107.0–2.59 (2.75–2.59)	48.69–3.14 (3.33–3.14)
Space group	C2	C2	C2
Unit cell parameters (Å, °)	a = 214.09, b = 83.01, c = 118.24, β = 115.759	a = 215.36, b = 83.39, c = 118.65, β = 115.57	a = 217.63, b = 83.86, c = 120.33, β = 116.13
Molecules in au	5	5	5
Number of total reflections	321,204 (13,098)	436,955 (66,107)	251,369 (38,638)
Unique reflections	83,998 (5642)	115,026 (17,963)	65,945 (10,382)
Average multiplicity	3.82 (2.32)	3.80 (3.68)	3.81 (3.72)
Data completeness (%)	94.5 (86.0)	99.0 (95.9)	99.2 (96.9)
R_{sym} (%)	10.8 (77.9)	13.5 (80.4)	15.1 (69.5)
$<I/\sigma_{(I)}>$	7.87 (1.03)	8.65 (1.82)	8.87 (2.00)
CC (1/2) (%)	99.5 (60.5)	99.2 (73.0)	99.1 (71.0)

$R_{sym} = \Sigma\Sigma| I_i - I_m| / \Sigma\Sigma I_i$, where I_i is the intensity of the measured reflection and I_m is the mean intensity of this reflection.
Values indicated in parentheses correspond to the statistics in the highest resolution shell.

(Arg262) is found in *Selaginella* with a large proportion of acidic residues representing one third of the amino acids (**Figure 2B**). The C-terminal of PAP9 in its long form (i.e., 40 residues) is not essential in higher Angiosperms since different clades have a shorter domain of approximately 20 residues in Physcomitrella, basal clades of the ANA grade, Apiales from Eudicots, Alismatales, and Asparagales from Monocots. Interestingly, the PAP9 C-terminus is either totally absent in Gyngko and Pinus or present as the short sequence in Picea, suggesting that there is no *bona fide* PAP9 referring to the involvement of the protein to the PEP function. These observations corroborate the hypothesis according to which Gymnosperms had favored a different use of PEP complex canceling the use of some PAPs that are not found anymore in the clade. In most Eudicots, a largely acidic tail with a well-conserved tyrosine (**Figure 2C**) may be involved in the PEP function as it could also play the role of electron donor with manganese clusters or as a signaling residue.

TABLE 2 | Refinement statistics.

	PAP9-6His
Resolution (Å)	48.20–2.25 (2.28–2.25)
R_{cryst} ($\sigma_F = 0$) (%)	17.94 (33.96)
R_{free} ($\sigma_F = 0$) (%)	22.10 (38.11)
Number of atoms	8997
Water molecules	399
B average (Å^2)	51.82
RMSD bonds (Å)	0.007
RMSD angle (°)	0.884
Ramachandran favored (%)	91.5
Ramachandran allowed (%)	7.4
Ramachandran disallowed (%)	0.5

Values indicated in parentheses correspond to the statistics in the highest resolution shell.
$R_{cryst} = \Sigma|| F_{obs}| - |F_{calc}|| / \Sigma| F_{obs}|$. R_{free} (Brünger, 1992) is the same as R_{cryst} but calculated for 5% data omitted from the refinement.

Subcellular Localization of PAP9-GFP Proteins

Some of the proteins associated to the PEP, like PAP9, possess a predicted NLS (Pfannschmidt et al., 2015). However, the putative NLS of PAP9 (**Figure 2A** and **Supplementary Table 4**) is nested within the cTP (**Figure 2A** and **Supplementary Tables 3,4**), which is conceptually cleaved off during plastid import through the TOC/TIC machinery. Hence the question arises whether the predicted sequence is actually a *bona fide* NLS. Since the NLS sequence at this position is not conserved in other species, it does not likely play an important role in PAP9 localization. This is experimentally supported by the transient localization of PAP9-GFP (**Figures 3A,B**), which appears to be mostly plastidial. However, the clear labeling of the stromules (**Figure 3B**), indicates that a part of the pool of fluorescent molecules is found in the stroma, released from the PEP/PAP complex. In some images, we could also detect some signals in the cytosol and nucleus (**Supplementary Figure 2**). The GFP fluorescent profile across plastids is more spread than that of the RFP, indicating that the PAP9-GFP signal is not as restricted as that of PAP10-RFP used here as specific marker of the PEP complex (Liebers et al., 2020). The translational fusion of GFP at the C-terminus may alter the function of the corresponding domain so that the localization may not reflect precisely that of PAP9. Such a perturbation has been observed for HMR/PAP5 (Chen et al., 2010) and pTAC6/PAP8 (Liebers et al., 2020) for which C-terminal GFP fusions alter the localization and/or the functionality of the protein.

Mass Spectrometry Analyzes

We utilized MS to assess the mass of PAP9-6His and ^{15}N,^{13}C-6His-PAP9 under denaturing conditions. The experimental mass of PAP9-6His was 30,848 Da, matching the amino acidic

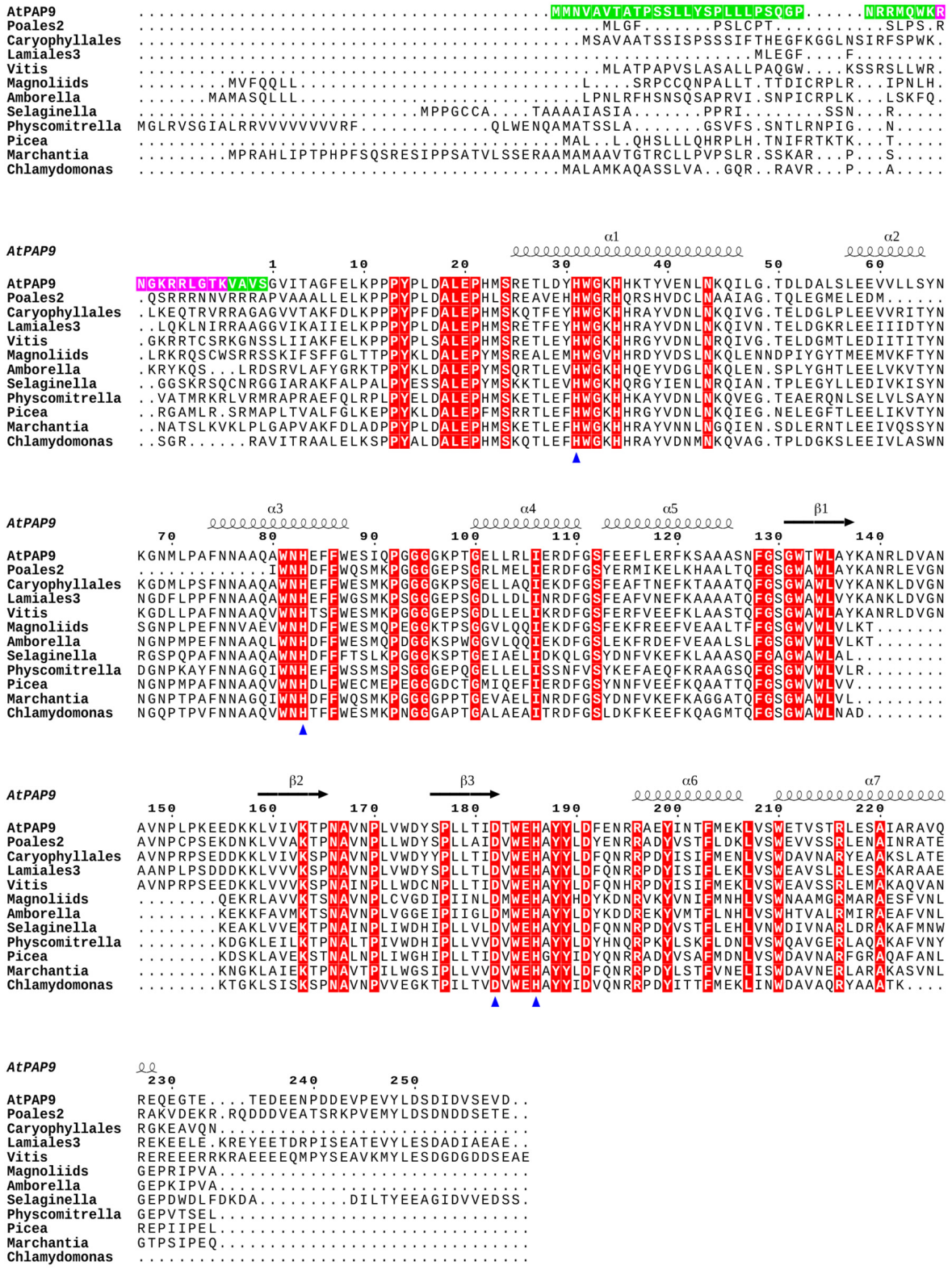


FIGURE 1 | PAP9 secondary structures mapping on a sequence alignment including orthologous proteins from different clades of the green lineage. The PAP9 secondary structure from *Arabidopsis thaliana* is drawn as followed: the α-helices are displayed as squiggles and β-strands as arrows. The conserved residues are highlighted in red. The residues involved in the metal binding, Zn^{2+} in the crystal structure of *A. thaliana* PAP9, are indicated with a blue triangle. The cTP and NLS of the *A. thaliana* PAP9 are highlighted in green and magenta, respectively. The drawing was prepared using ESPript (Robert and Gouët, 2014).

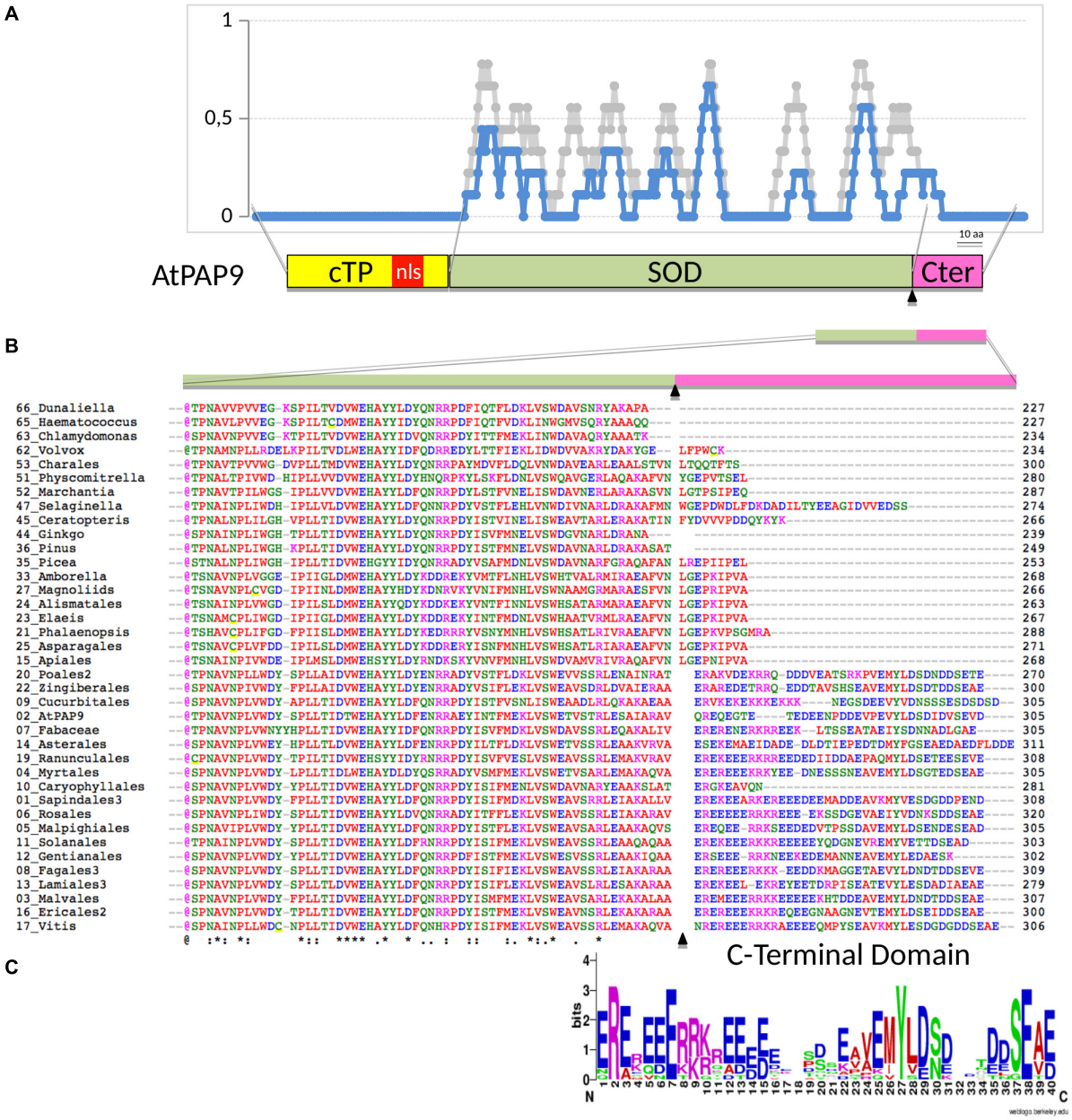


FIGURE 2 | Sequence alignment of predicted orthologous PAP9 protein found in representatives of phylogenetic clades as described in **Supplementary Table 1**. **(A)** Amino acid identity (in blue) and similarity (in light gray) given in a 10-aa moving ratio. The alignment corresponds to the full dataset given in **Supplementary Table 2** marked with a cross (column named in this figure). Schematic illustration of the PAP9 domains: cTP, chloroplast transit peptide in yellow; NLS, predicted nuclear localization signal in red; SOD, superoxide dismutase domain in khaki; C-terminal domain in magenta. **(B)** Amino acid partial alignment of the accessions as above, the 2-digit number corresponds to the clade given in **Supplementary Table 1**. cTP, transit peptide as predicted with ChloroP1.1 (www.cbs.dtu.dk) underlined in yellow and described in **Supplementary Table 3**. (*), (:), or (.), conserved, strongly similar or weakly similar amino acid properties (standards from www.uniprot.org). Amino acids colors as in Clustal Omega [red (AVFPMILW): small + hydrophobic (includes aromatic Y); blue (DE): acidic; magenta (RHK): basic; green (STYHCNGQ): hydroxyl + sulfhydryl + amine + G). bNLS, bipartite NLS as predicted with NLS mapper (http://nls-mapper.iab.keio.ac.jp). **(C)** AA sequence logo (Crooks et al., 2004) generated on the C-terminal domain alignment of PAP9 with the 40-aa-long stretch from clade 22 (Zingiberales) to clade 17 (Vitis) excluding clade 10 (Caryophyllales).

sequence 1-270 (**Figure 4A**) and ^{15}N,^{13}C-6His-PAP9 displayed a mass of 34,670 Da. The calculated mass of the fully labeled protein is 34,801 Da, taking into account Met at N-terminal that has not been cleaved because the second residue before the 6His-Tag is Lys (Hirel et al., 1989); the difference between both mass resulting from an incomplete labeling (**Figure 4B**). To investigate the oligomeric state of PAP9-6His, we used native MS and we detected monomers and dimers (**Figure 4C**).

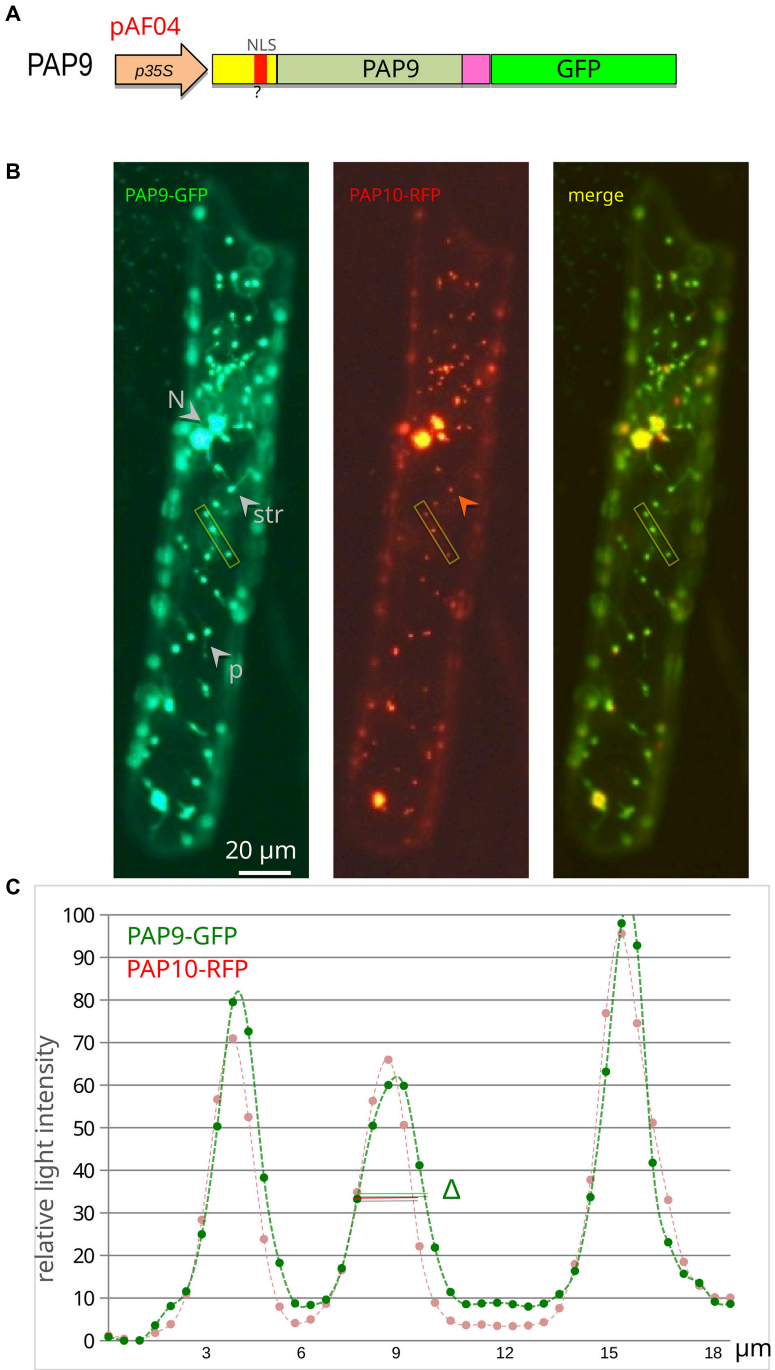


FIGURE 3 | PAP9 is localized in plastids. **(A)** Schematic illustration of the PAP9-GFP construction in pAF04. p35S, CaMV35S promoter region; cTP, chloroplast transit peptide in yellow; ?, predicted NLS (nuclear localization signal) in red; C-terminal domain in magenta; GFP in green. **(B)** Transiently expressed PAP9-GFP in onion epidermal cells. N, nucleus; str, stromule; p, plastid. The red arrowhead points to the absence of red fluorescence in stromules. The yellow rectangle represents the analyzed segment in panel **(C)**. **(C)** Fluorescent signal quantitative profile on an 18-µm-long segment of the image across three plastids. Δ represents the difference in width of the GFP signal compared to the red signal of PAP10-RFP.

X-Ray Structure Analyzes

Five molecules of PAP9 are in the asymmetric unit. Four of them form two dimers. The fifth interacts with a molecule from another asymmetric unit to form also a dimer. Both monomers in the dimer are related by a non-crystallographic twofold axis. The monomers are very similar with a value of root mean square deviation (RMSD) ranging from 0.14 to 0.21 Å between monomers when calculated between the Cα atoms. The buried

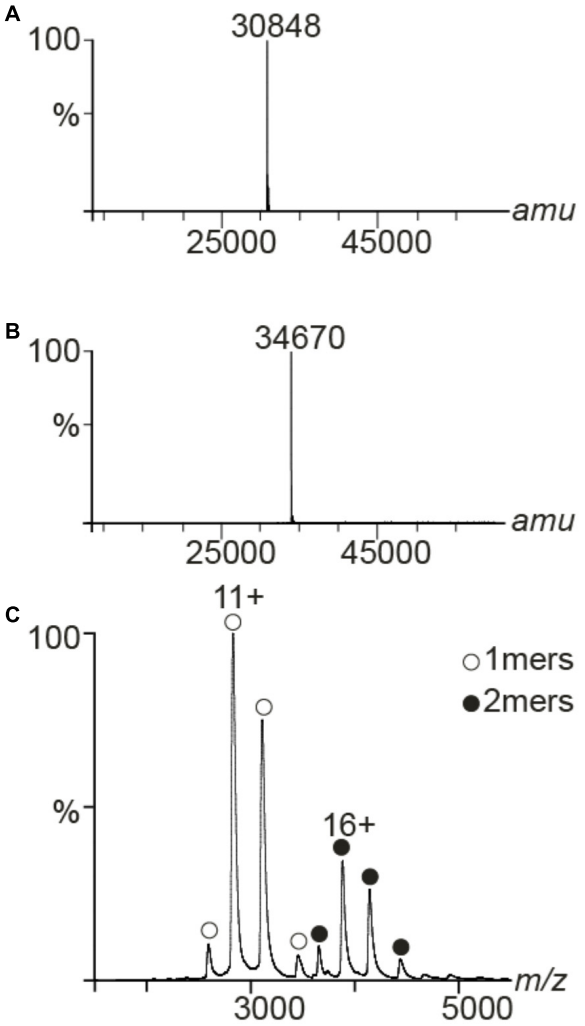


FIGURE 4 | MS spectra of PAP9. Deconvoluted spectra of PAP9-6His **(A)** and $^{15}N,^{13}C$-6His-PAP9 **(B)**. Under denaturing conditions the accurate mass of PAP9-6His was 30,848 and 34,670 Da for $^{15}N,^{13}C$-6His-PAP9. **(C)** Native MS spectrum of the PAP9-6His. It formed two distinct oligomers, such as monomers (1mer, 30,848 ± 1 Da) and dimers (2mers, 61,697 ± 2 Da).

area calculated using PISA (Krissinel and Henrick, 2007) in the dimer interface is 1785 Å^2. PAP9 is folded in two domains similar to those observed in FeSODs or MnSODs. The N-terminal domain extends from Gly1 to Gly93 and contains three α-helices. The C-terminal domain (Gly94–Gln229) displays an α/β fold with a three anti-parallel β-strands sandwiched by four α-helices and the N-terminal domain (**Figure 5**). No electron density is observed for residues from Arg141 to Glu155 and for the last 29 residues from Gly231 to Asp259 suggesting flexibility. Crystallographic analysis of the 6His-PAP9, produced to decrease the C-terminal flexibility, did not allow to better observe the electron density of the C-terminal part and the structures of 6His-PAP9 and PAP9-6His were similar. The catalytic center is at the interface of the N- and C-terminal domains. Surprisingly a zinc ion, instead of the expected iron ion, is penta-coordinated in the catalytic center. Anomalous difference electron density map calculated at the zinc K-edge showed a strong peak of anomalous density (**Figure 6**) while the map computed with diffraction data collected just after the zinc K-edge does not show any strong peak. The zinc ion is penta-coordinated by the His31, His83 side chains of the N-terminal domain, the Asp182, and His186 side chains from the α/β fold domain, and a water molecule supposed to mimic the position of the hydroxide ion (**Figure 6**). The arrangement of the five coordinating ligands forms a trigonal bipyramid with His31 and the water molecule as the axial ligands. The side chains of His35, Tyr39, Gln79, and Trp184 close the catalytic site (**Figure 6**). Since PAP9 mainly binds Zn^{2+} in our expression/purification steps no catalytic activity could be observed excepted at very high PAP9 concentrations (**Supplementary Figure 1**).

Structure Comparisons and the PAP9 Family

Rms deviations calculated using PDBefold (Krissinel and Henrick, 2004) between the monomer of PAP9 and more than 200 monomers of SODs from the PDB range from 0.71 Å (PAP9 vs. FeSOD from *V. unguiculata*, PDB entry: 1UNF) to 1.6 Å with the FeSOD from *Aquifex pyrophilus* (PDB entry: 1COJ) (Lim et al., 1997). The structure comparisons revealed that the fold of PAP9, the ligands involved in the metal coordination and residues closing the catalytic site are conserved. Dimer interface comparison with FeSOD from *V. unguiculata* revealed also a conservation of residues involved in interactions by hydrogen bonds between the subunits. The Glu185 carboxylate group from one monomer interacts with the Ser130 hydroxyl group involving a water molecule and also with the His186 imidazole ring of the catalytic center from the other monomer. Additionally, the hydroxyl group of Ser130 interacts with the hydroxyl group of Ser130 from the other monomer (**Figure 7**). The main difference originates from the metal center occupied by a zinc ion in AtPAP9 instead of an iron ion. The conserved interaction described in FeSOD from *V. unguiculata* between His35 of one monomer and Tyr188 of the other monomer is not observed in PAP9. The residues Gly156 to Ser164 of the cytosolic FeSOD from *V. unguiculata* corresponding to Val144 to Pro152 of the flexible loop Arg141–Glu155 in PAP9 are not observed in the electron density.

Sequence comparisons between PAP9 and SODs of the PDB showed that the flexible C-terminal part (Gly231 to Asp259) of PAP9 is not observed in the sequences of SODs of the PDB. The longest C-terminal extension is observed in FeSOD of *Helicobacter pylori* (PDB entry: 3CEI) (Esposito et al., 2008). However, it is 19 residues shorter than in PAP9 and is folded as a kinked α-helix that interacts with the N-terminal domain. The 29 last residues unobserved in the electron density map of PAP9 are found in several sequences reported as plastid SODs. Indeed, the PAP9 C-terminal part alone, used in alignment searches of the UniProtKB database restricted to plants, matches FeSODs; some of which being not annotated as plastid-localized, despite individual detection of a chloroplast transit peptide using the ChloroP prediction tool. Most of hits are *bona fide*

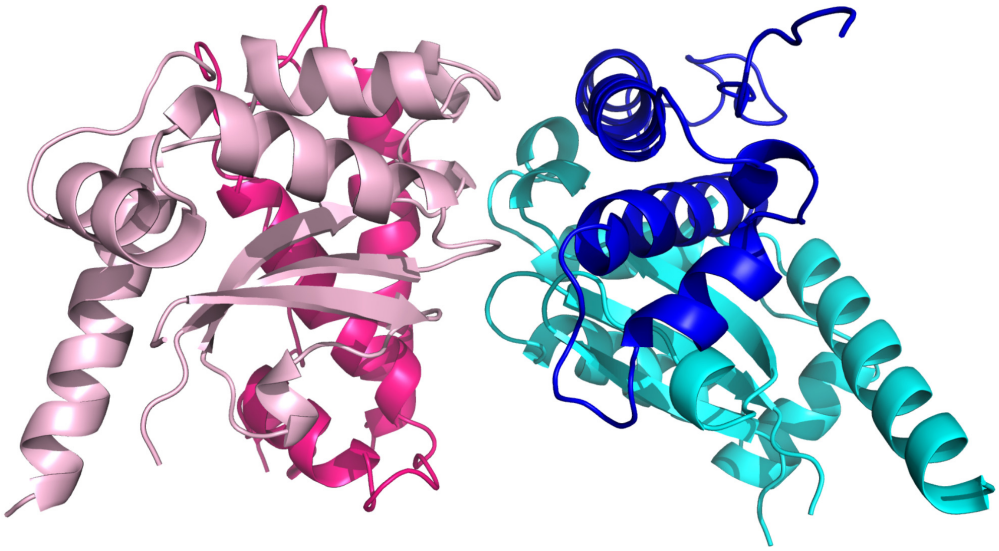


FIGURE 5 | View of the PAP9 dimer. The β-strands are drawn in arrows and the α-helices are represented in ribbons. The N-terminal domains are colored in dark pink and dark blue. The C-terminal domains are in cyan and light pink.

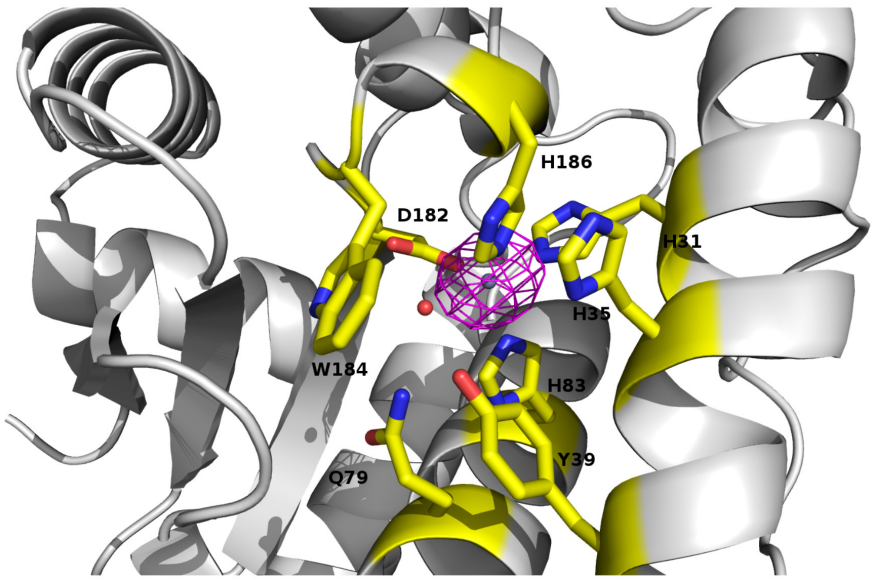


FIGURE 6 | View of the catalytic site of PAP9 superimposed with the anomalous electron density map calculated at the zinc K-edge. Residues of the catalytic site and closing the catalytic site are drawn in sticks. The zinc ion is drawn as gray sphere. The water molecule corresponding to the hydroxide ion is represented as a red sphere.

PAP9 orthologous SODs, and the C-terminal sequence represents a signature of this protein family. In addition, the sequence homology between PAP9 and PAP4/FSD3 (MW: 25657.94 Da) from *A. thaliana* is very high, suggesting that both FeSODs have a similar fold. However, PAP4 does not have the C-terminal extension found in PAP9. PAP9 and PAP4 should be functionally distinct and partially redundant as suggested by comparison of individual light-green phenotypes to the more severe albino phenotype of the double mutant (Myouga et al., 2008).

Solution-State NMR Analyses

Two segments, suggesting a dynamic structure, are not observed in the crystal structure of PAP9, i.e., the loop Arg141–Glu155 and the C-terminal part Gly231–Asp259 and are supposed to behave a fast dynamic. In order to further investigate the structural and dynamic properties of these unseen parts in the PAP9 crystal structure, we produced ^{15}N,^{13}C-6His-PAP9. In our conditions (see section "Materials and Methods"), only about forty peaks can be observed above the background in the ^{15}N-SOFAST spectrum

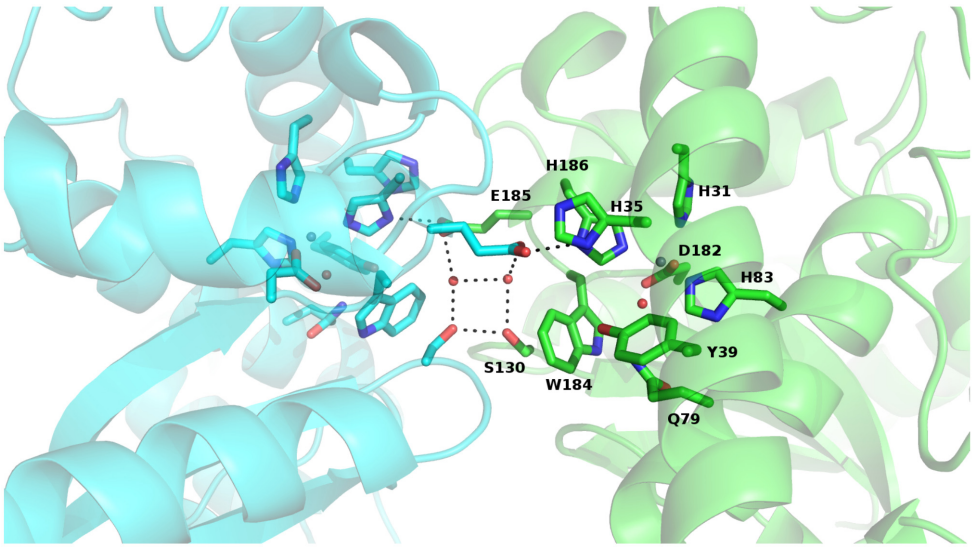


FIGURE 7 | View of the conserved interactions between both monomers of PAP9 *and* observed in FeSODs, with *each monomer* of PAP9 in a different color. The residues involved are drawn in sticks, the hydrogen bonds are represented in dark dashed lines and the water molecules are shown as spheres. The zinc ion is drawn as gray sphere. The β-strands are drawn in arrows and the α-helices are represented in ribbons.

in agreement with the presence of some dynamic residues. The most intense residues have an apparent rotational correlation time of 3 ns measured using [^{15}N,^{1}H]-TRACT technique (Lee et al., 2006), a value near those expected for free peptides or small proteins such as ubiquitin. In the other hand, the translational diffusion coefficient measured using DOSY experiment at 293 K is of 7×10^{-7} cm^2/s, indicating that PAP9, from the point of view of translational diffusion, behaves like an object of 80 kDa. For such molecular weight, the residues located in the structured regions of the protein are expected to be line broadened supporting the fact that only the flexible residues can be observed in the NMR spectra. These results indicate that the observed residues have a fast movement while being included in a much larger species. We performed a set of 3D-experiments to assign these residues: HNCA, HNCACB, and HNCOCANH. Of these residues, only fifteen present detectable correlations in HNCACB experiments. A first analysis allows characterizing unambiguously a GTxTx sequence that corresponds only to the GTETE sequence located in the C-terminal tail of PAP9. In order to help to identify other residues within this part and characterize secondary structures, we studied a peptide composed of the 34 last residues of PAP9. We have entirely assigned the protons and carbons of the peptide using homonuclear TOCSY, NOESY, and ^{13}C-HSQC experiments at natural abundance. SSP program (Marsh et al., 2006) using Cα, Cβ, Hα chemical shift data sets show that the peptide does not present any secondary structure propensity at all (**Supplementary Figure 3**). In the same way, the ^{13}C-HSQC experiment of the integer ^{15}N,^{13}C-6His-PAP9 presents the very similar correlations than those observed for the peptide (**Supplementary Figure 4**), strongly suggesting that the C-terminal tail in 6His-PAP9 is also dynamic. Analysis of the observable Cα and Cβ chemical shift values in the protein together with comparison of those of the peptide allowed us to assign the Gly231–Glu238 and the Ser251–Asp259 stretches. Assignments of Asn239, Val247–Leu249 can be proposed on basis of the HNCA experiment. The assigned ^{1}H-^{15}N correlation spectrum of 6His-PAP9 is shown in **Figure 8**. No residue of the Gln226-Glu230 stretches were identified in agreement with their position in the last helix of the protein. Interestingly, the correlations of the residues, when observable, located in the middle of the tail: Asn239-Asp250 showed weaker intensities than those in the Gly231–Glu238 and Ser251–Asp259 stretches.

DISCUSSION

In Angiosperms, the developmental program following germination in the dark is skotomorphogenesis. Inside the cell, chloroplast biogenesis is blocked, allowing for the formation of yellow etioplasts without the chlorophylls. After light perception etiolated seedlings start the photomorphogenesis program leading to chloroplast biogenesis (Liebers et al., 2018). This essential step toward photo-autotrophy involves the rapid assembly of the photosynthetic apparatus within the thylakoid membranes. Jointly, chlorophylls are quickly synthesized from the stored precursors, protochlorophyllides, by the light-activated protochlorophyllide oxidoreductase (POR). Chlorophylls are then inserted in the light harvesting antenna proteins. Transcription of photosynthesis associated plastid genes is ensured by PEP and is rapidly promoted after light perception owing to the PAP assembly into the active PEP complex. Two of the PAPs are FeSODs (Myouga et al., 2008; Steiner et al., 2011). FeSODs catalyze the dismutation of superoxide radicals into peroxides and may protect the transcriptional machinery from the newly acquired photosynthetic capacity (Pfannschmidt et al., 2015). Once

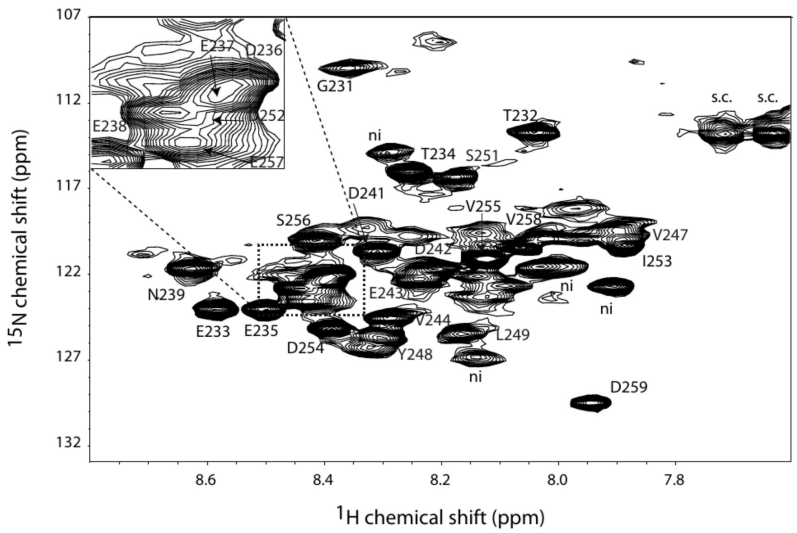


FIGURE 8 | ^{1}H-^{15}N correlation spectrum of PAP9 with the assigned amino acid residue labels annotated "ni" standing for not identified.

the chloroplast is formed and fully photosynthetically active, the PEP activity substantially decreases.

Transmembrane translocation of PAP9 into the chloroplast results from the recognition of its N-terminal plastid transit peptide by the transmembrane TOC/TIC machinery. Fluorescence microscopy experiments showed that PAP9 is mainly located in the chloroplast stroma (**Figure 3**); the stroma localization may result from the lack of developed thylakoids in onion epidermal cells. Therefore, the predicted nuclear localization sequence observed within the cTP (**Figure 2A** and **Supplementary Table 4**) may not serve a localization purpose. It is cleaved off instead during the chloroplast import leading to a mature protein of 30,848 Da as observed using mass spectrometry analysis in denaturing conditions (**Figure 4A**). The native MS data indicated that PAP9 assembles as dimers. Monomers were also detected, suggesting protein dynamics during assembly. The ionization efficiency of the different oligomeric states affects the relative abundance of the different species in the MS spectra. Therefore, it is not possible to judge whether the monomers are more abundant that the dimers. Moreover, the native MS experiments were performed at 5 μM concentration and in ammonium acetate, which is a different buffer used for purification, NMR, and crystallographic experiments. The buffer conditions may affect the relative abundance of the species.

In the crystals, PAP9 is a symmetric dimer (**Figure 5**) as revealed by the low RMSD values between both monomers. The buried surface of the dimer interface suggests that the dimer is the biological form of PAP9. The FeSODs and MnSODs are active as dimeric or tetrameric (dimer of dimers) enzymes (Perry et al., 2010). In the PEP, PAP9 has been observed as a monomer (Steiner et al., 2011); a form of the protein also observed in our mass spectrometry analyses. The main difference between PAP9 analyzed here, and the FeSODs or MnSODs, is the metal ion bound to the catalytic site. In our crystal structure a zinc ion, instead of an iron ion, is penta-coordinated by a water molecule, supposed to mimic the position of the hydroxide ion (Muñoz et al., 2005), the His31, His83, Asp182, and His186 side chains (**Figure 6**) as observed in the cytoplasmic FeSODs and MnSODs. The zinc ion cannot be the catalytic ion to perform the dismutation of superoxide since it has only the redox state II, in opposition to Fe and Mn that both have several redox states from II to VI and II to VIII, respectively. Since PAP9 is an active FeSOD even when overexpressed in *E. coli* (Myouga et al., 2008), the replacement of Fe by Zn occurred during either overexpression or purification of the protein although the metal center is hidden from the solvent by the conserved side chains of His35, Tyr39, Gln79, and Trp184 (**Figure 6**). As observed in FeSODs and MnSODs, His186 from the catalytic site of one monomer interacts also with the Glu185 carboxylate group from the other monomer. All these observations suggest a conserved catalytic mechanism in cytoplasmic and plastid FeSODs.

The main difference between PAP9 and the other FeSODs, and even MnSODs, is the additional residues of the C-terminal part. In the crystal structures of PAP9-6His and 6His-PAP9, no electron density was observed for the 29 last residues of the C-terminal part resulting from flexibility. Proteolysis can be excluded since the correct molecular weight of the 6His-tagged PAP9 was observed using mass spectrometry (**Figure 4A**). The flexibility does also not result from the construction of the over-expressed recombinant protein since the electron density of the C-terminal part is not observed for 6His-PAP9. The only observable residues of ^{13}C,^{15}N-6His-PAP9 using NMR correspond essentially to the C-terminal residues whose dynamic is identical to that of the free peptide (**Supplementary Figure 4**). This result clearly shows that the C-terminal part is flexible with its central part (weaker intensities of correlations) not as free as the two other parts, probably due to some interactions of this part with residues at the protein surface. As in FeSOD from *V. unguiculata* (Muñoz et al., 2005), no electron density is observed for residues Val144 to Pro152 of the loop

Arg141–Glu155 suggesting flexibility in cytosolic FeSODs (FSD1) from plants and PAP9. The C-terminal extension observed in PAP9 could then allow distinguishing between PAP9, as a component of the PEP, and other plant FeSODs. We hypothesize that the C-terminal tail anchors PAP9 to the PEP complex and its observed flexibility arises from the isolation of a subunit that normally belongs to a larger multisubunit complex.

The C-terminal part of the protein had strongly changed during evolution (**Figures 1**, **2**). It is absent in early clades of the green lineage. A first significant C-terminal modification is found in Charales and Physcomitrella while a second longer fragment appears in *Selaginella*. Such events are dating back to the conquest of fresh waters and terrestrial life. It is then possible that the C-terminal part could have appeared along with a complete set of new features for controlling chloroplast transcription; namely the assembly of PEP-PAP complex. The acquisition of these features, including SOD activities in a stoichiometry of four units per complex (three PAP4 and one PAP9), may provide sufficient protection of the organelle while the photosynthetic cells are exposed to a more oxidizing environment. This C-terminal part is totally absent in Gymnosperms, which seem to have evolved a completely different strategy of photo-autotrophy acquisition with, for example, no light regulation of chloroplast biogenesis since seedlings can green in darkness.

The PEP is composed of at least 16 subunits of unknown structures. Interactions between some of them were only reported by using non-direct observations, using yeast-two-hybrid assays (Yu et al., 2013) and fluorescent microscopy (Myouga et al., 2008). We have recently shown by NMR that PAP5 interacts with PAP8 (Liebers et al., 2020). PAP9 was proposed to interact with PAP4 therefore forming a hetero-complex of FeSODs (Myouga et al., 2008), and we show here that PAP9 can have several oligomeric states. Surprisingly, neither this heterocomplex nor the PAP9 dimer have been described (Steiner et al., 2011) suggesting that the PEP is probably a dynamic complex, still poorly characterized at the level of its structure and composition.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: http://www.wwpdb.org/, 7BJK.

AUTHOR CONTRIBUTIONS

RB and DC designed the research. AF, PG, EBE, LS, SSM, RB, and DC performed the research. EBE and LS contributed mass spectrometry data. AF and PG contributed NMR data. RB and DC wrote the manuscript with contributions from AF, PG, EBE, LS, and TP. All authors approved the manuscript.

FUNDING

This work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UAR 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). This work was supported by the Agence National de la Recherche (ANR-17-CE11-0031).

ACKNOWLEDGMENTS

The diffraction experiments were conducted on beamline FIP-BM30A and ID23-1 at the ESRF (Grenoble, France). We thank the beamline staff for technical help, Auriane Bron and Florence Prunier-Bossion for their technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.668897/full#supplementary-material

Supplementary Figure 1 | Enzymatic assay of PAP9. The superoxide dismutase activity of PAP9 was tested using pyrogallol. The pyrogallol auto-oxidation was followed by monitoring the absorbance increase at 420 nm. After 180 s, PAP9 at several concentrations [50 (orange), 100 (gray), 200 (yellow), 500 μM (light blue), and 1 mM (green)] or 5 μM Mn-SOD (dark blue) were added into the medium and the absorbance was monitored for further 3 min.

Supplementary Figure 2 | Transiently expressed PAP9-GFP in onion epidermal cells. N, nucleus; str, stromule; p, plastid. The red arrowhead points to the absence of red fluorescence in stromules.

Supplementary Figure 3 | Secondary structure propensity (SSP) scores for the Cter-PAP9 peptide (circles) and the C-terminal tail of integer PAP9 (squares) using $^{13}C\alpha$, $^{13}C\beta$, and Hα chemical shifts. Between residues 245 and 252, the SSP score (star) was obtained from the $^{13}C\alpha$ and Hα chemical shifts only. Positive values represent α-structure propensity and negative values represent β-structure propensity. The SSP is near zero along the sequence indicating the absence of any secondary structure in the peptide. The numbering of the residues corresponds to the whole protein.

Supplementary Figure 4 | Overlay of ^{1}H-^{13}C correlation spectra (sensitivity-enhanced HSQC) of PAP9 and the Cter-PAP9 peptide. The peptide signals and the PAP9 peaks are shown in red and in black, respectively. All resonances of the peptide are observable in the PAP9 spectrum indicating the presence of the same mobility in the C-terminal tail of PAP9.

REFERENCES

Abreu, I. A., and Cabelli, D. E. (2010). Superoxide dismutases – a review of the metal-associated mechanistic variations. *Biochim. Biophys. Acta.* 1804, 263–274. doi: 10.1016/j.bbapap.2009.11.005

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst. D* 66, 213–221. doi: 10.1107/S0907444909052925

Bobik, K., and Burch-Smith, T. M. (2015). Chloroplast signaling within, between and beyond cells. *Front. Plant Sci.* 6:781. doi: 10.3389/fpls.2015.00781

Boeri Erba, E., and Petosa, C. (2015). The emerging role of native mass spectrometry in characterizing the structure and dynamics of macromolecular complexes. *Protein Sci.* 24, 1176–1192. doi: 10.1002/pro.2661

Boeri Erba, E., Signor, L., and Petosa, C. (2020). Exploring the structure and dynamics of macromolecular complexes by native mass spectrometry. *J. Proteomics* 222:103799. doi: 10.1016/j.jprot.2020.103799

Börner, T., Aleynikova, A. Y., Zubo, Y. O., and Kusnetsov, V. V. (2015). Chloroplast RNA polymerases: role in chloroplast biogenesis. *Biochim. Biophys. Acta* 1847, 761–769. doi: 10.1016/j.bbabio.2015.02.004

Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472–475. doi: 10.1038/355472a0

Chen, M., Galvao, R. M., Li, M., Burger, B., Bugea, J., Bolado, J., et al. (2010). *Arabidopsis* HEMERA/pTAC12 initiates photomorphogenesis by phytochromes. *Cell* 141, 1230–1240. doi: 10.1016/j.cell.2010.05.007

Collaborative Computational Project, Number 4 (CCP4) (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* 50, 760–763. doi: 10.1107/S0907444994003112

Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004

Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8, 978–984. doi: 10.1110/ps.8.5.978

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Cryst. D* 66, 486–501. doi: 10.1107/S0907444910007493

Esposito, L., Seydel, A., Aiello, R., Sorrentino, G., Cendron, L., Zanotti, G., et al. (2008). The crystal structure of the superoxide dismutase from *Helicobacter pylori* reveals a structured C-terminal extension. *Biochim. Biophys. Acta* 1784, 1601–1606. doi: 10.1016/j.bbapap.2008.04.024

Favier, A., and Brutscher, B. (2011). Recovering lost magnetization: polarization enhancement in biomolecular NMR. *J. Biomol. NMR* 49, 9–15. doi: 10.1007/s10858-010-9461-5

Gao, Z.-P., Yu, Q.-B., Zhao, T.-T., Ma, Q., Chen, G.-X., and Yang, Z.-N. (2011). A functional component of the transcriptionally active chromosome complex, Arabidopsis pTAC14, interacts with pTAC12/HEMERA and regulates plastid gene expression. *Plant Physiol.* 157, 1733–1745. doi: 10.1104/pp.111.184762

Hirel, P. H., Schmitter, M. J., Dessen, P., Fayat, G., and Blanquet, S. (1989). Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid. *Proc. Natl. Acad. Sci. U.S.A.* 86, 8247–8251. doi: 10.1073/pnas.86.21.8247

Jarvis, P. (2008). Targeting of nucleus–encoded proteins to chloroplasts in plants. *New Phytol.* 179, 257–285. doi: 10.1111/j.1469-8137.2008.02452.x

Jarvis, P., and López-Juez, E. (2013). Biogenesis and homeostasis of chloroplasts and other plastids (2013). *Nat. Rev. Mol. Cell. Biol.* 14, 787–802. doi: 10.1038/nrm3702

Kabsch, W. (2010). XDS. *Acta Cryst.* D 66, 125–132. doi: 10.1107/S0907444909047337

Kosugi, S., Hasebe, M., Tomita, M., and Yanagawa, H. (2009). Systematic identification of yeast cell cycle-dependent nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10171–10176. doi: 10.1073/pnas.0900604106

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. doi: 10.3389/fpls.2014.00385

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Cryst. D* 60, 2256–2268. doi: 10.1107/S0907444904026460

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797. doi: 10.1016/j.jmb.2007.05.022

Lee, D., Hilty, C., Wider, G., and Wüthrich, K. (2006). Effective rotational correlation times of proteins from NMR relaxation interference. *J. Magn. Reson.* 178, 72–76. doi: 10.1016/j.jmr.2005.08.014

Liebers, M., Chevalier, F., Blanvillain, R., and Pfannschmidt, T. (2018). PAP genes are tissue- and cell-specific markers of chloroplast development. *Planta* 248, 629–646. doi: 10.1007/s00425-018-2924-8

Liebers, M., Gillet, F. X., Israel, A., Pounot, K., Chambon, L., Chieb, M., et al. (2020). Nucleo-plastidic PAP8/pTAC6 couples chloroplast formation with photomorphogenesis. *EMBO J.* 39:e104941. doi: 10.15252/embj.2020104941

Liebers, M., Grübler, B., Chevalier, F., Lerbs-Mache, S., Merendino, L., Blanvillain, R., et al. (2017). Regulatory shifts in plastid transcription play a key role in morphological conversions of plastids during plant development. *Front. Plant Sci.* 19:23. doi: 10.3389/fpls.2017.00023

Lim, J. H., Yu, Y. G., Han, Y. S., Cho, S., Ahn, B. Y., Kim, S. H., et al. (1997). The crystal structure of an Fe-superoxide dismutase from the hyperthermophile *Aquifex pyrophilus* at 1.9 Å resolution: structural basis for thermostability. *J. Mol. Biol.* 270, 259–274. doi: 10.1006/jmbi.1997.1105

Majeran, W., Friso, G., Asakura, Y., Qu, X., Huang, M., Ponnala, L., et al. (2012). Nucleoid-enriched proteomes in developing plastids and chloroplasts from maize leaves: a new conceptual framework for nucleoid functions. *Plant Physiol.* 158, 156–189. doi: 10.1104/pp.111.188474

Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006). Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci.* 15, 2795–2804. doi: 10.1110/ps.062465306

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., et al. (2002). Evolutionary analysis of ***Arabidopsis***, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U.S.A.* 99, 12246–12251. doi: 10.1073/pnas.182432999

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Cryst.* 40, 658–674. doi: 10.1107/S0021889807021206

Morgner, N., and Robinson, C. V. (2012). Massign: an assignment strategy for maximizing information from the mass spectra of heterogeneous protein assemblies. *Anal. Chem.* 84, 2939–2948. doi: 10.1021/ac300056a

Morris, K. F., and Johnson, C. S. (1992). Diffusion-ordered two-dimensional nuclear magnetic resonance spectroscopy. *J. Am. Chem. Soc.* 114, 3139–3141. doi: 10.1021/ja00034a071

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. doi: 10.1110/ps.04979505

Myouga, F., Hosoda, C., Umezawa, T., Iizumi, H., Kuromori, T., Motohashi, R., et al. (2008). A heterocomplex of iron superoxide dismutases defends chloroplast nucleoids against oxidative stress and is essential for chloroplast development in *Arabidopsis*. *Plant Cell* 20, 3148–3162. doi: 10.1105/tpc.108.061341

Perry, J. J., Shin, D. S., Getzoff, E. D., and Tainer, J. A. (2010). The structural biochemistry of the superoxide dismutases. *Biochim. Biophys. Acta* 1804, 245–262. doi: 10.1016/j.bbapap.2009.11.004

Pfannschmidt, T. (2003). Chloroplast redox signals: how photosynthesis controls its own genes. *Trends Plant Sci.* 8, 33–41. doi: 10.1016/s1360-1385(02)00005-5

Pfannschmidt, T., Blanvillain, R., Merendino, L., Courtois, F., Chevalier, F., Liebers, M., et al. (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. doi: 10.1093/jxb/erv415

Pilon, M., Ravet, K., and Tapken, W. (2011). The biogenesis and physiological function of chloroplast superoxide dismutases. *Biochim. Biophys. Acta* 1807, 989–998. doi: 10.1016/j.bbabio.2010.11.002

Robert, X., and Gouët, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucl. Acids Res.* 42, W320–W324. doi: 10.1093/nar/gku316

Roth, M., Carpentier, P., Kaïkati, O., Joly, J., Charrault, P., Pirocchi, M., et al. (2002). FIP: a highly automated beamline for multiwavelength anomalous diffraction experiments. *Acta Cryst.* D 58, 805–814. doi: 10.1107/s0907444902003943

Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A., and Robinson, C. V. (2002). A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* 74, 1402–1407. doi: 10.1021/ac0110552

Solyom, Z., Schwarten, M., Geist, L., Konrat, R., Willbold, D., and Brutscher, B. (2013). BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J. Biomol. NMR* 55, 311–321. doi: 10.1007/s10858-013-9715-0

Stein, N. (2008). CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Cryst.* 41, 641–643. doi: 10.1107/S0021889808006985

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. doi: 10.1104/pp.111.184515

Sugiura, M. (1992). The chloroplast genome. *Plant Mol. Biol.* 19, 149–168. doi: 10.1007/BF00015612

van den Heuvel, R. H., van Duijn, E., Mazon, H., Synowsky, S. A., Lorenzen, K., Versluis, C., et al. (2006). Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal. Chem.* 78, 7473–7483. doi: 10.1021/ac061039a

Yu, Q. B., Huang, C., and Yang, Z. N. (2014). Nuclear-encoded factors associated with the chloroplast transcription machinery of higher plants. *Front Plant Sci.* 5:316. doi: 10.3389/fpls.2014.00316

Yu, Q. B., Lu, Y., Ma, Q., Zhao, T. T., Huang, C., Zhao, H. F., et al. (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. doi: 10.1111/j.1399-3054.2012.01718.x

Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q., et al. (2008). Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS One* 3:e1994. doi: 10.1371/journal.pone.0001994

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

3.3. BIOPHYSICAL AND BIOCHEMICAL CHARACTERIZATION OF CSP41b AND PRIN2

The transition from dark-to-light growth triggers significant changes in the PEP complex, involving multiple PAPs associated with redox mechanisms, which affect the expression of plastid genes. CSP41b and PRIN2 were found to interact directly, forming a complex crucial for PEP-dependent transcription during embryo development. PRIN2 and CSP41b play critical roles in redox-mediated retrograde signalling and PEP activation as reported by Diaz et al, 2018. CSP41b is also reported to be crucial for the complete expression of *PhANGs*, and its interaction with PRIN2 is necessary for light-activated transcription. This research aims to highlight the interplay of CSP41b and PRIN2 in the regulation of plastid gene transcription and their roles in plant development and redox-mediated retrograde signalling. In pursuit of this objective, *in vitro* biophysical techniques such as isothermal titration calorimetry, native and denaturing mass spectrometry, and the isolation of the CSP41b-PRIN2 complex *via* size-exclusion chromatography have been employed. By using an integrated approach, it is aimed to understand the interaction of recombinant CSP41b and PRIN2 proteins *in vitro*, with no previous research done on the structural characterisation of CSP41b and PRIN2.

3.3.1. Purification of CSP41b and PRIN2

CSP41b and PRIN2 were purified with a yield of approximately 5800 mg/L and 7000 mg/L respectively. In gel filtration, only one peak was observed at an elution volume of 74 mL and 98 mL for CSP41b and PRIN2 suggesting that CSP41b was largely purified as dimer and PRIN2 as monomer, respectively. The purity was accessed by SDS-PAGE analyses and showed that the proteins were pure for further studies (Figure 3.5A, C). CSP41b was provided to the EM platform for quality control by negative stain electron microscopy.

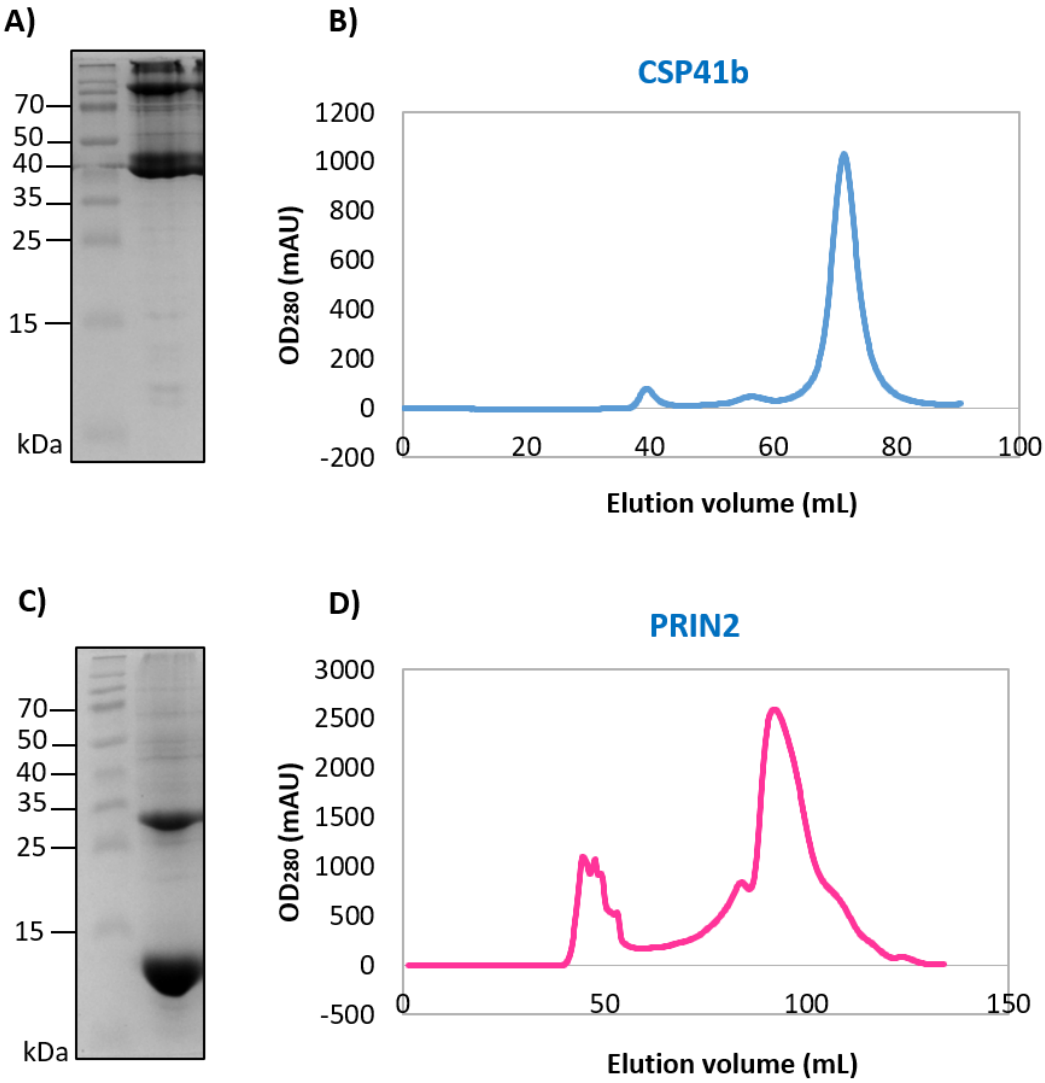


***Figure 3.5**: Purification of CSP41b and PRIN2. SDS-PAGE of the eluted fractions of CSP41b (A) and PRIN2 (C) after size exclusion chromatography. Profiles of CSP41b (B) and PRIN2 (D) after size exclusion chromatography.*

3.3.6. Isothermal titration calorimetry analysis of CSP41b and PRIN2

ITC was performed at the ITC platform at Partnership for Structural Biology to check any signs of kinetic interactions between PRIN2 and CSP41b. The experiments were performed in different buffer conditions using various protein concentrations. After performing the experiment in various concentrations and buffer conditions, weak exothermic signals were observed, with a stoichiometry of 1, which is also consistent with the existing literature (Diaz et al, 2018). The K_a value was *= 3.67E5 ± 1.73E5 M^{-1}.* The K_d value was calculated from the formula, $K_d = 1/K_a$. *K_d = 2.72 ± 5.780 µM.* We can conclude that there might be weak

interactions between PRIN2 (300 µM) and CSP41b (30 µM) in 50 mM NaCl and 10 mM Tris-HCl pH 8.0 (Figure 3.6).

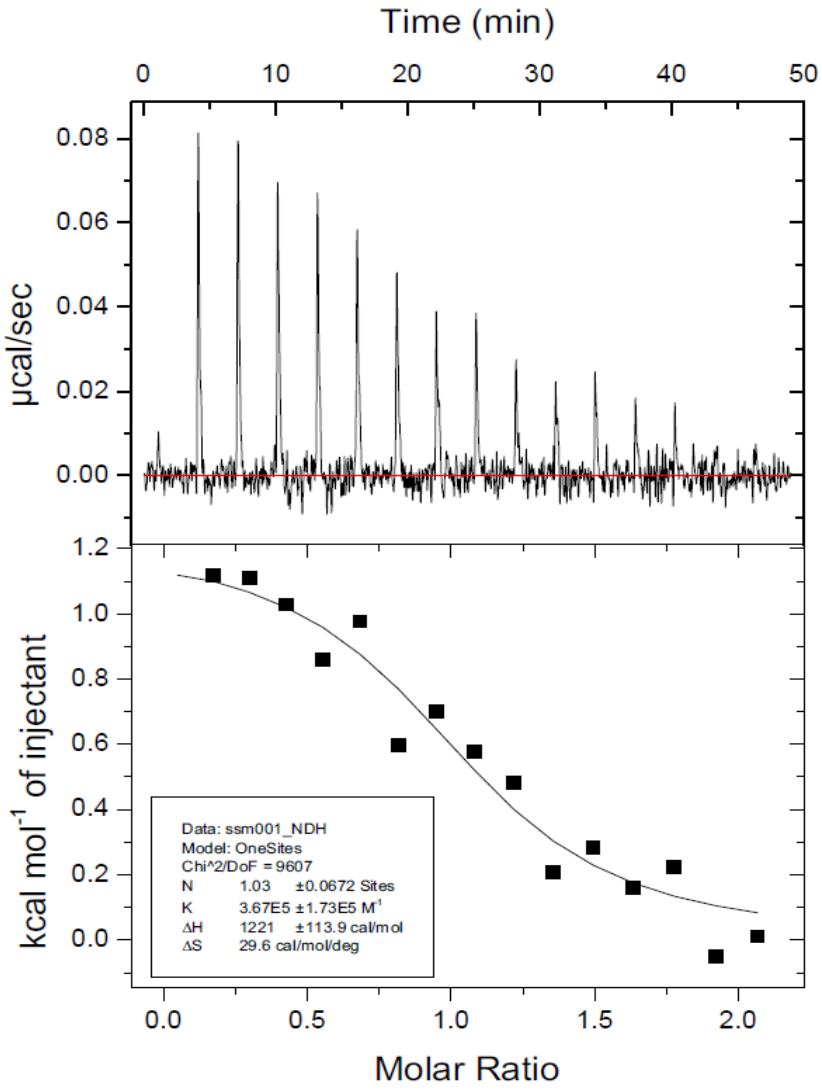


Figure 3.6: *ITC spectrum for interaction between PRIN2 and CSP41b in 50 mM NaCl and 10 mM Tris-HCl with K_d = 2.72 ± 5.780 µM, K_a = 3.67E5 ± 1.73E5 M^{-1} and N (stoichiometry) = 1*

3.3.5. Mass spectrometry analysis of CSP41b and PRIN2

The CSP41b-PRIN2, PRIN2 and CSP41b samples in 150 mM NaCl and 25 mM Tris-HCl pH 8.0 were provided to MS platform at IBS. The experiment and analysis were performed by Dr. Elisabetta-Erba Boeri. Since the objective is to investigate whether PRIN2 and CSP41b form a complex *in vitro* or not, native mass spectrometry was performed as it is helpful in identifying protein-protein complexes.

Denaturing mass spectrometry analyses of CSP41b and PRIN2 were also performed to estimate their accurate molecular weights. The observed mass for PRIN2 was 14682 Da and 14860 Da. The expected mass was 14682.58 (Figure 3.7 B). The additional 178 Da (+14860 Da) may be due to the alpha-N-gluconylation of the His-tagged proteins. The observed masses for CSP41b were 39126.11 Da, 39304.41 Da and 39486.59 Da. The expected mass was 39126.11 Da resulting from additional residues from cloning. The additional mass +178 Da could be due to alpha-N-gluconylation. The 39486 Da peak could be due to two gluconoylation. (Figure 3.7 A).

The native mass spectrometry spectrum of CSP41b and PRIN2 in 250 mM ammonium acetate pH 7.0 displayed a spectrum that corresponded to CSP41b and PRIN2 and not to CSP41b-PRIN2 (Figure 3.7 C, D). From analysing PRIN2 alone (Native MS), PRIN2 in its monomeric state at 14 kDa and low abundance signal ranging between 80 – 90 kDa was observed. From analysing CSP41b alone (Native MS), it's dimeric state at 78 kDa and low abundance signals that correspond to monomers at 39 kDa and high oligomeric states were observed. However, in the spectra of the complex, the main signals corresponded to the CSP41b monomer and dimer at 78 kDa. Low abundance species that represented at 117 kDa and 156 kDa were also observed that may be CSP41b trimers (3x39=117 kDa) and tetramers (4x39=156 kDa). Oligomeric states of CSP41b were observed *in planta* (Kremnev & Strand, 2014) but no new signals that belonged to PRIN2-CSP41b were observed.

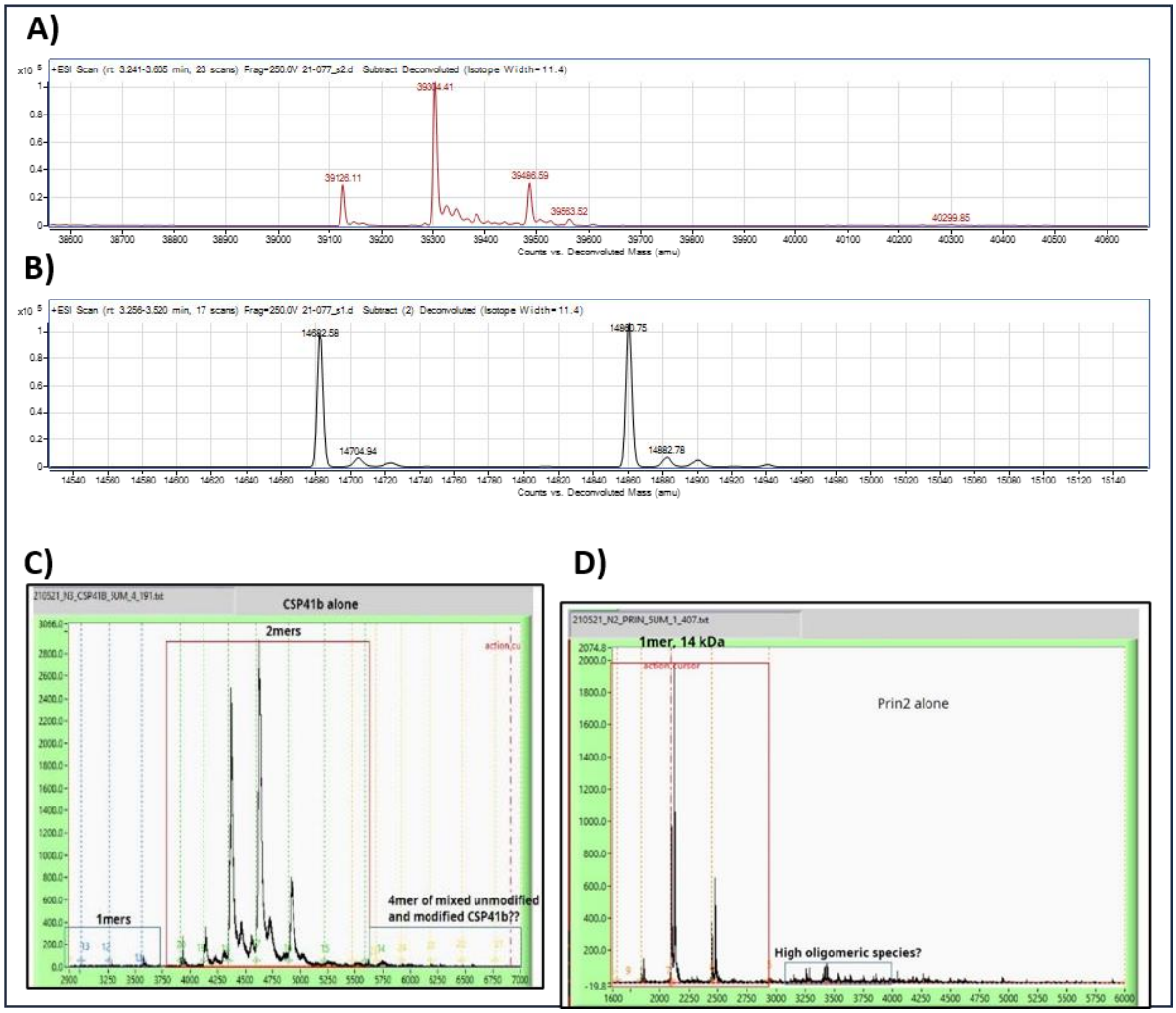


***Figure 3.7**: Denaturing MS spectra of CSP41b (A) and PRIN2 (B). Native MS spectra for CSP41b (C) and PRIN2(D)*

3.3.7. Isolation of CSP41b-PRIN2 complex by size exclusion chromatography

The isolation of CSP41b-PRIN2 complex (100 µM each) was attempted by size exclusion chromatography using and Superdex® S200 10/300 and Superdex® S75 10/300 in buffer compositions such as 150 mM NaCl 25 mM Tris-HCl pH 8.0 and 75 mM NaCl 25 mM Tris-HCl pH 8.0 to reproduce results already obtained. Different incubation timings were tested. The size-exclusion chromatogram did not show any peak corresponding to CSP41b-PRIN2 complex.

3.4. STRUCTURAL CHARACTERIZATION OF CSP41b and PRIN2

3.4.1. Negative staining electron microscopy of CSP41b

Negative staining electron microscopy is important for checking the quality of the sample and to assess its homogeneity. Homogenous solution is important for single particle cryo-EM structure resolution. The CSP41b samples for negative staining were provided to the IBS EM platform. Two conditions of the samples, *i.e.* with and without NaCl, were stained with 1% MFT Sodium silico-tungstate (SST). At 150 mM NaCl and 25 mM Tris-HCl pH 8.0, homogenous solutions of CSP41b were observed (Figure 3.8 A). This sample was provided to the IBS EM platform for cryo-EM experiments.

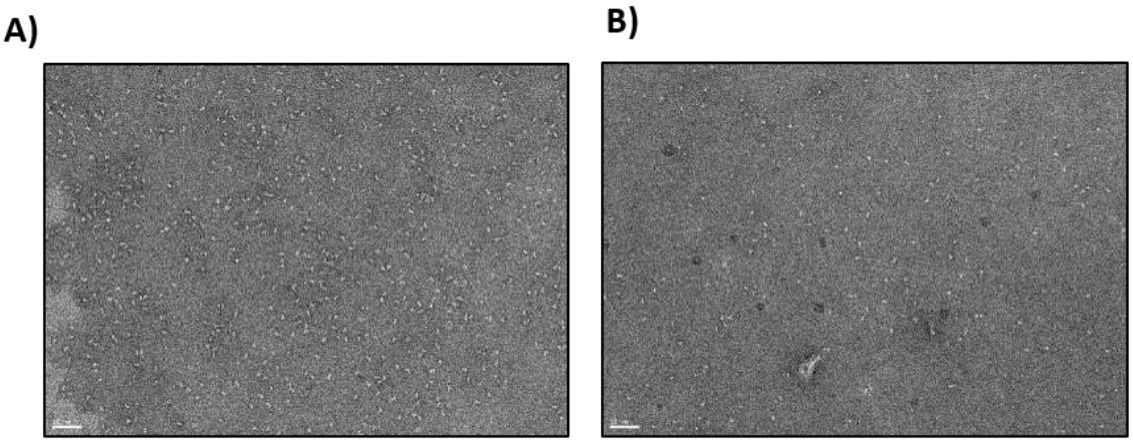


Figure 3.8: *(A) Homogenous solution of CSP41b in the presence of 150mM NaCl at MFT SST 1 %. (B) Non-homogenous solution in the absence of NaCl at MFT SST 1%.*

3.4.2. Cryo-EM structure of CSP41b at 3.4 Å resolution

Search for CSP41b sequence within the protein sequences in the PDB revealed that no 3D structure of CSP41b was solved and that the closest homolog to CSP41b is the UDP-glucose 4-epimerase from *Thermotoga maritima* with 25% sequence identity. This sequence identity suggested that both proteins share a common fold and maybe also a common dimeric state. Since no 3D structure of CSP41b was solved, attempts to solve its 3D structure by X-ray crystallography were performed but all failed probably resulting from flexibility. Attempts to crystallize the protein after cleavage of the 6His-Tag also failed. The 3D structure was then solved using cryo-EM while the size of the dimer classified it as a very small protein for structural studies by cryo-EM as the technique is better adapted for macromolecules ranging above 100 kDa.

The data collection was performed by Dr. Gregory Effantin on the Krios microscope CM01 at the European Synchrotron Radiation Facilities (ESRF). An electron density map at 3.4 Å global resolution was calculated which provided a dimer of CSP41b. The refinement and validation statistics are provided in table 7.

Both monomers can be superposed with a rmsd value of 0.728 Å when calculated with all the atoms. This low value shows that both monomers are roughly identical. The structure comparison with PDB using PDBeFOLD (Krissinel & K. Henrick, 2004), displayed that the monomer of CSP41b can be superimposed on the monomer of the UDP-N-acetylglucosamine 4-epimerase from *Bacillus cereus* (PDB entry 3M2P) (Berman et al, 2004) with a rmsd value of 1.79 Å, while the dimers of both proteins cannot be superposed.

The 3D structure analyses revealed that CSP41b has 2 domains, with a Rossmann fold domain in N-terminal domain. Function of the C-terminal domain is unknown whereas the Rossmann fold domain is well known to bind cofactors such as NAD^+, $NADP^+$ in dehydrogenesases (Rao & Rossmann, 1973) or nucleotide for example such as ATP in tRNA-synthetases (Moras, 1992). CSP41b has a few flexible regions: only residues 86 – 89 located in a loop and last residues in C-terminal part, were not assigned due to very low local resolution in these regions (Figure 3.9 C, D).

The electrostatic potential surface of CSP41b (Figure 3.10 A, B) showed a large crevice positively charged containing several conserved basic residues belonging to both N-terminal and C-terminal domains. The conserved basic residues are highlighted in grey in figure 3.10 B, C. This observation strongly suggests that both domains participate in the binding of ligands that have negative charges such as RNA or DNA. Since CSP41b is supposed to interact with RNA stem loops during the translation, this crevice may bind stem loop from mRNA during translation or during the transcription-translation coupling for stabilization.

CSP41b sequence alignments revealed not only conserved basic residues, but also some negative residues that form a negative patch at the protein surface. Moreover, sequences analyses in several taxonomies using PlantEnsemble and blast program show that CSP41b is present in most of the clads from photosynthetic bacteria such as Nostoc to angiosperms. This observation suggests that the gene encoding CSP41b was translocated from the archae genome to the nucleus after endosymbiosis.

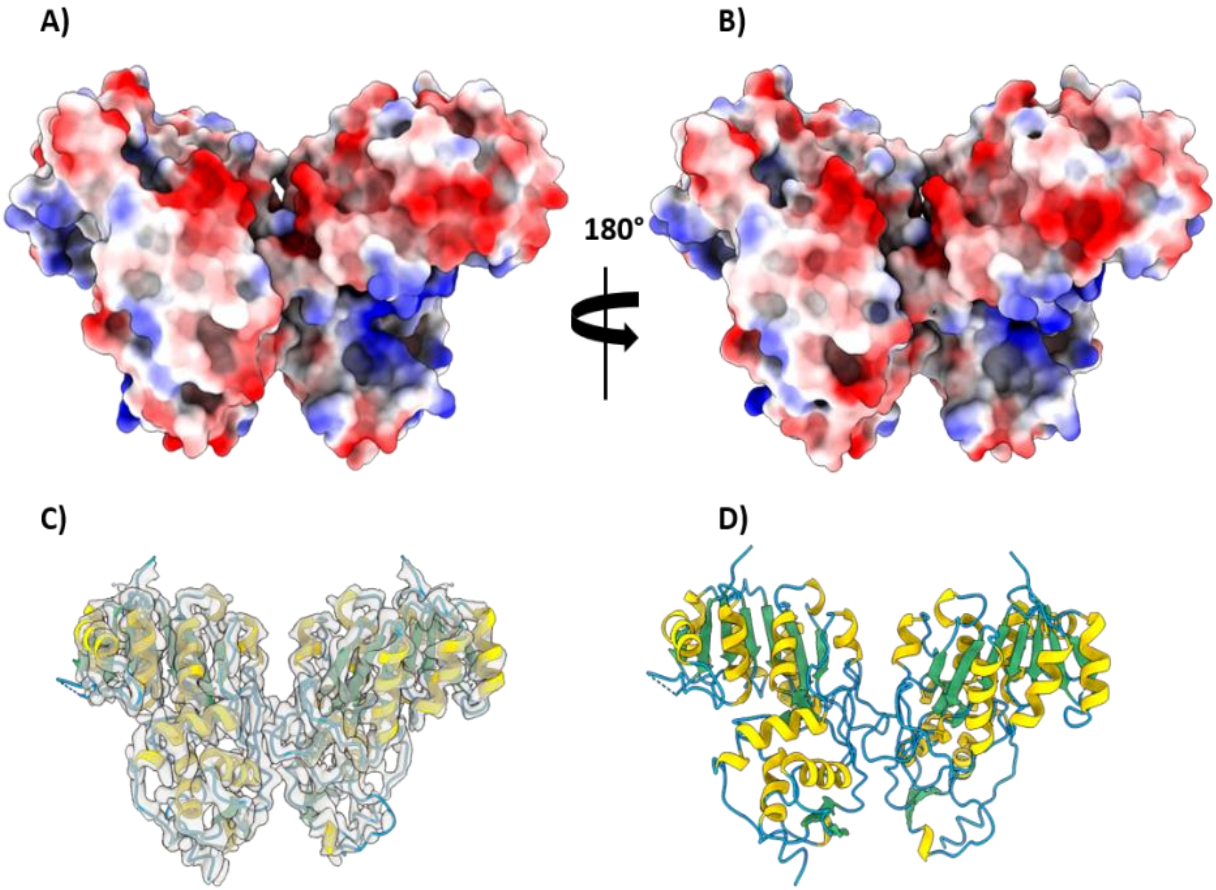
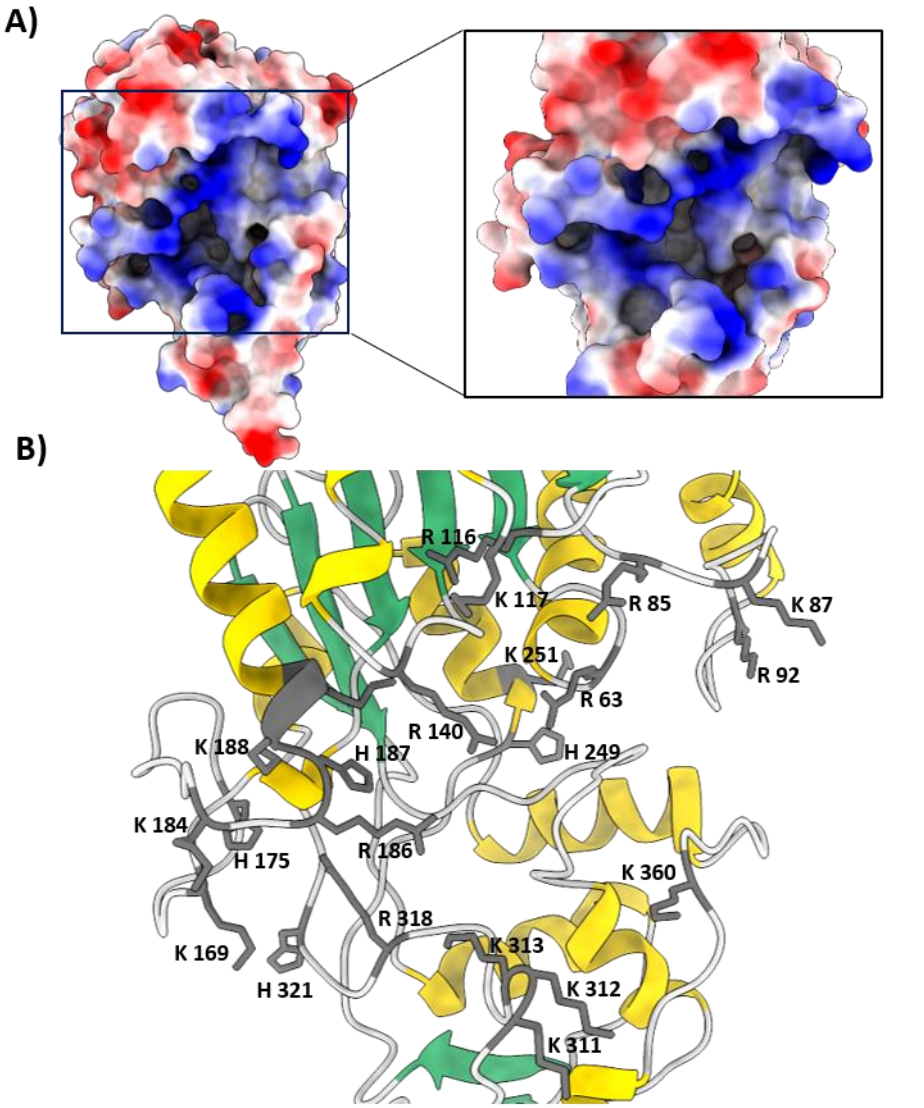


Figure 3.9: *(A), (B) – Electrostatic surface potentials of CSP41b. The regions represented in blue correspond to positively charged regions, while the regions represented in red correspond to negatively charged patches. (C) 3D model of CSP41b fitted inside the map (light grey) at 3.4 Å. (D) 3D model of CSP41b. Helices are represented in yellow, β-strands are represented in green and loops are represented in blue.*



C)

SSEKKILIMGGTRFIGLFLSRILVKEGHQVTLFTRGKSPIAKQLPGESDQ
DFADFSSKILHLKGDRKDYDFVKSSLSAEGFDVVYDINGREAEEVEPIL
EALPKLEQYIYCSSAGVYLKSDILPHCEEDAVDPKSRHKGKLETESLLQS
KGVNWTSIRPVYIYGPLNYNPVEEWFFHRLKAGRPIPVPNSGIQISQL
GHVKDLATAFLNVLGNEKASREIFNISGEKYVTFDGLAKACAKAGGFP
EPEIVHYNPKEFDFGKKKAFPFRDQHFFASVEKAKHVLGWKPEFDLV
EGLTDSYNLDFGRGTFRKEADFTTDDMILSKKLVLQ

Figure 3.10: *(A) – Electrostatic surface potentials of a monomer of CSP41b and zoomed region to show the conserved basic residues. The regions in blue correspond to positively charged regions, while the regions in red correspond to negatively charged patches. (B) The conserved residues at the crevice shown in grey are highlighted. The helices are in yellow and the strands are in green. (C) Sequence of CSP41b from residue 51 without cTP. The residues highlighted in blue corresponds to coils, helices in yellow and residues highlighted in bold correspond to conserved basic residues.*

Table 7: Cryo-EM data collection, refinement and validation statistics

Data collection and processing		
Nominal magnification		165,000 x
Voltage (kV)		300 kV
Electron resource ($e^-/Å^2$)		40
Defocus range (μm)		-0.5 to -1.0
Calibrated pixel size (Å)		0.83
Symmetry imposed		C2
Final particles images (no.)		129,038
Map resolution (Å) at FSC = 0.143		3.4
Refinement		
Model resolution (Å) at FSC = 0.143		3.4
RMS deviations	Bonds (Å)	0.003
	Angles (°)	0.468
Validation		
MolProbity score		1.84
Clash score		7.63
Ramachandran plot (%)	Favoured	95.07
	Allowed	4.93
	Outliers	0.00
C-β deviations		0.00
Rotamer outliers (%)		1.28
CαBLAM outliers %		3.22

3.4.3. Structure of PRIN2

The crystal structure of PRIN2 was solved by Alicia Vallet, Remi Ruedas and Dr. David Cobessi from NMR and GSY at IBS. After finding that secondary structure of PRIN2 was only 4 α-helices by NMR, the crystal structure of PRIN2 was solved by molecular replacement by searching 4 α-helices. The structure is made up of four α-helices (Figure 3.11). The construct used for crystallisation was only its coding region without cTP, with a 6His tag at the N-terminal in the pET28a+ plasmid. The first N-terminal cysteine was not in the coding sequence as it was

predicted to be a part of the cTP. It was introduced after the crystallographic studies as it was described to be essential by Diaz et al, 2018. But the introduction of the Cys in N-terminal led to the formation of PRIN2 oligomers that prevented any structural studies of PRIN2. The other cysteine residue Cys51 is buried.

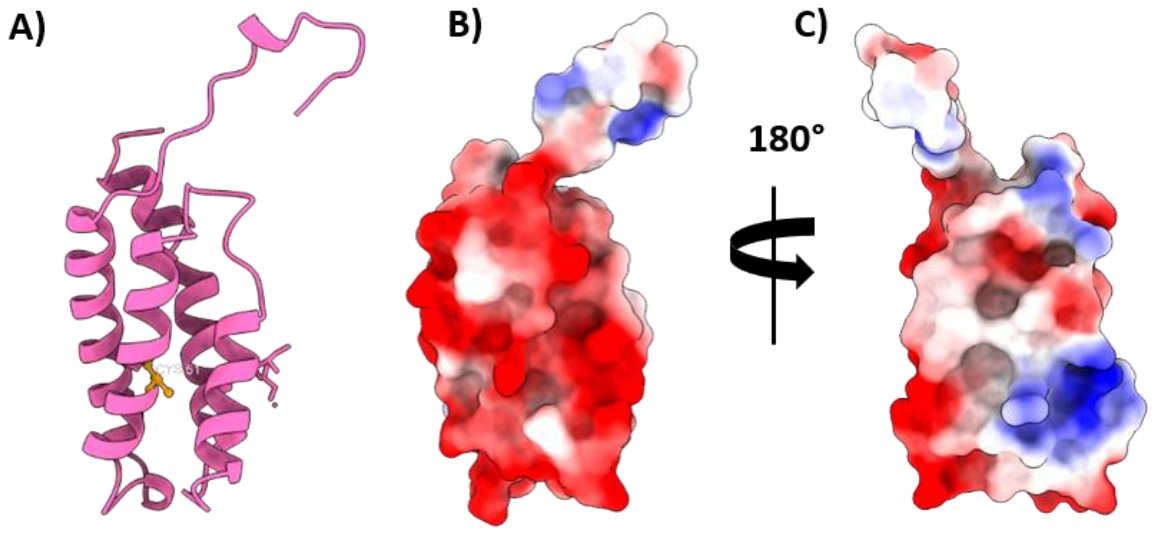


***Figure 3.11**: 3D model of PRIN2 in pink at 1.6 Å resolution from X-ray diffraction (A). Cys 51 residue is shown in orange as sticks. (B) (C) Electrostatic surface potential of PRIN2. The regions in blue correspond to positively charged regions, while the regions in red correspond to negatively charged patches.*

3.4.4. CSP41b-PRIN2 complex stabilization by cross-linking experiments for cryo-EM

BS3 crosslinker was used to crosslink CSP41b and PRIN2 for complex formation, since ITC revealed a weak interaction, for cryo-EM experiment. Concentrations of BS3 at 0.5 mM and 1 mM were used and the incubation time was tested at 45 minutes and 90 minutes with CSP41b and PRIN2 at 1:10 and 1:3 ratios. The samples after incubation were analysed by SDS-PAGE. The bands observed were similar with multiple bands to samples 1 and 2, that corresponded to 0.5 mM BS3 at 1:10 ratio of CSP41b : PRIN2 with incubation periods of 45 minutes and 90 minutes The cross-linked samples of CSP41b and PRIN2 (1 and 2) were provided to Daphna Fenel at IBS EM platform to observe the homogeneity of the sample. Both the samples (Figure 3.12 B, C) were observed to be highly concentrated and non-homogenous This could not be used to further analysis by cryo-EM. The samples could be further purified by size-exclusion

chromatography and only the bands corresponding to the complex formation could be provided to the EM platform to get the complex structure.

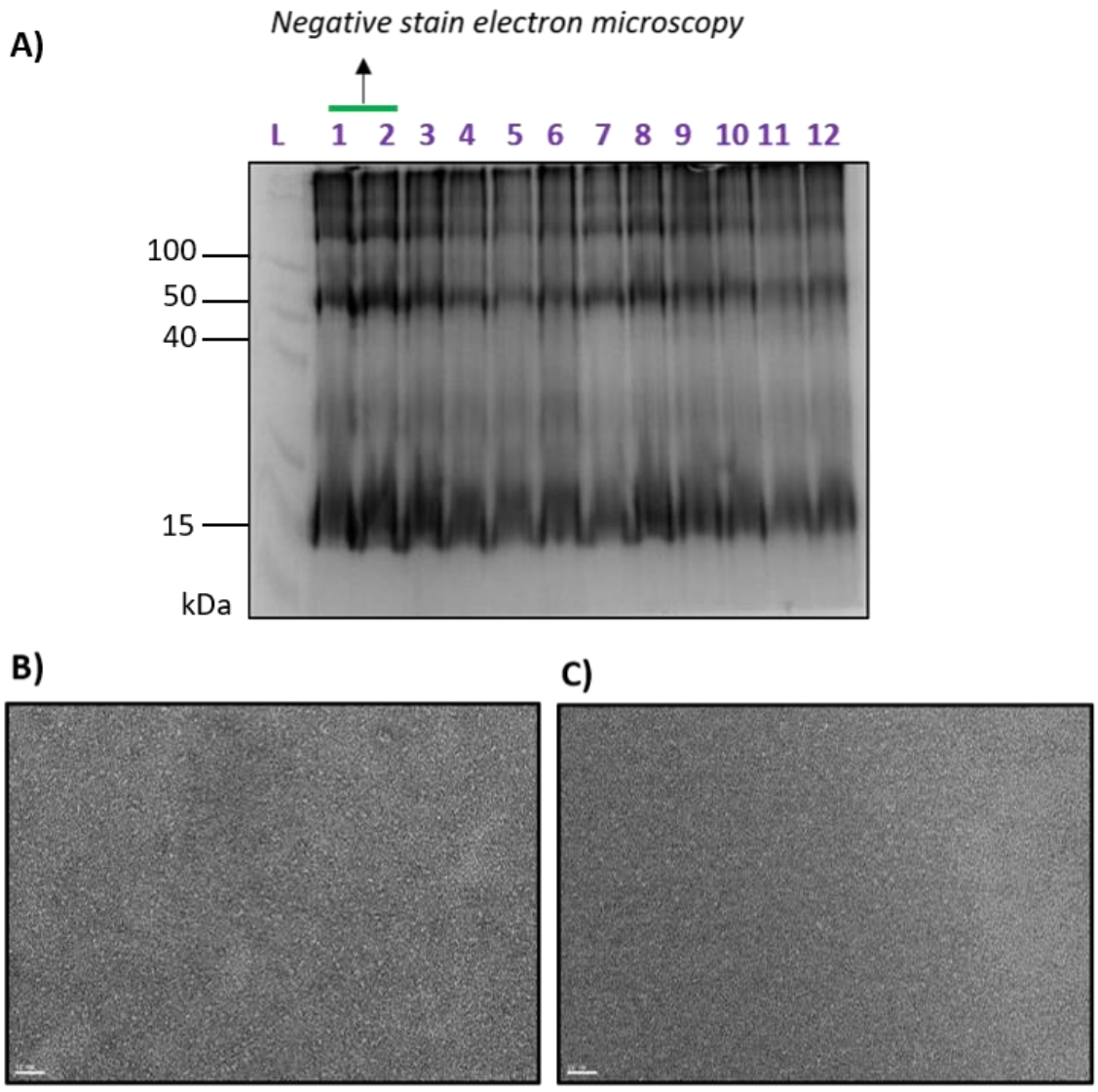


Figure 3.12: *(A) SDS PAGE gel of cross-linked CSP41b-PRIN2 where lanes 1 to 12 correspond to different concentrations and incubation period of the samples. L- Ladder. Samples 1 and 2 at were provided to IBS EM platform for negative stain electron microscopy. (B) (C) Samples 1 and 2 after negative staining with MFT SST 1% showing non-homogenous solution.*

The complex between PRIN2 and CSP41b was not isolated or well characterized using these approaches. This could be due to the existence of a complex but with a weak affinity between both proteins or that maybe a result of experimental conditions. However, the low affinity complex may be the strongest hypothesis since the complex was observed on gel filtration but without reproducibility. Since the complex could not be structurally characterized, it was modeled using AlphaFold2 by Dr. David Cobessi. The complex calculated by AlphaFold2 allows

the binding of RNA to CSP41b since the positively charged crevice is fully accessible. (Figure 3.13)

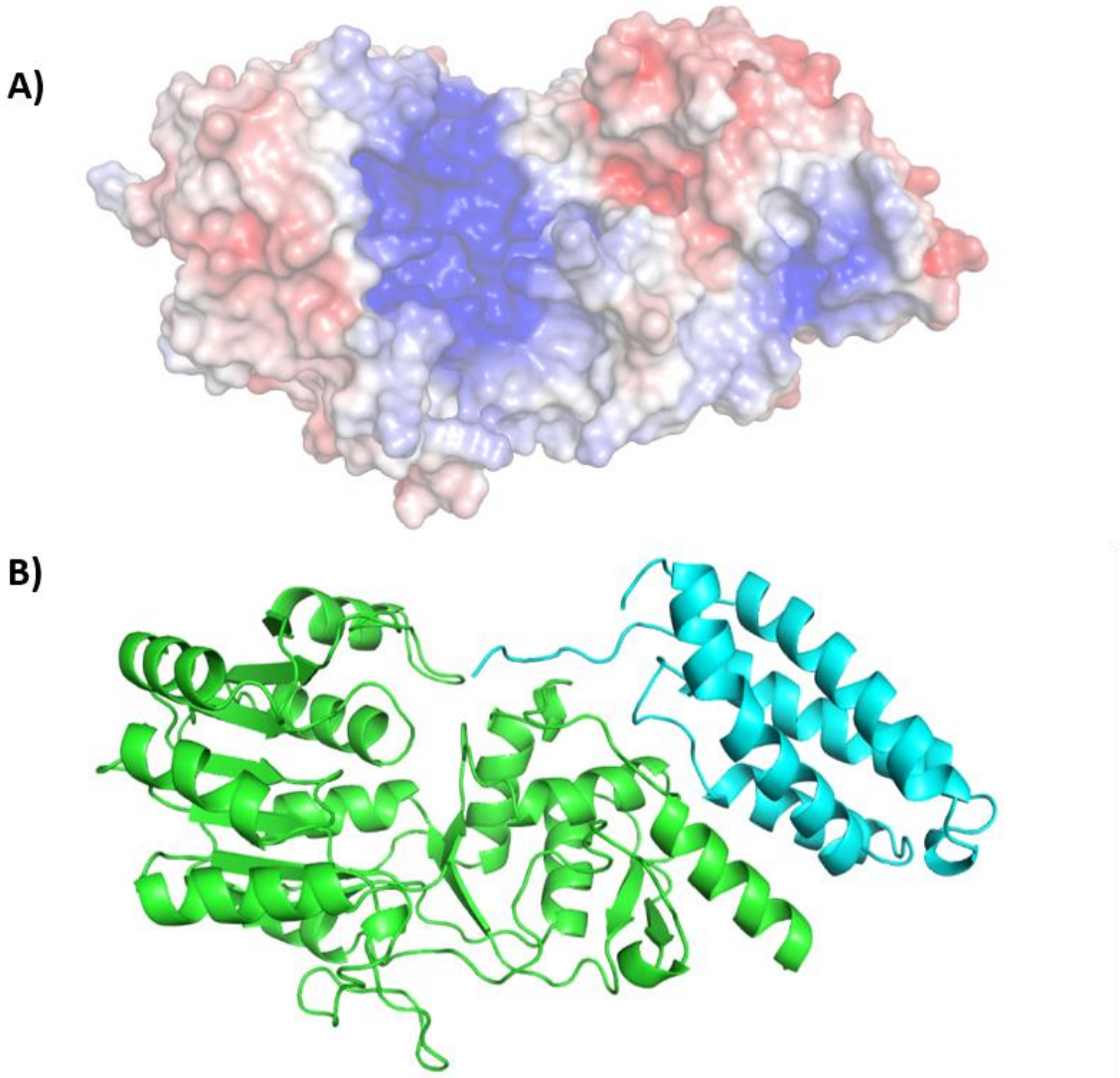


Figure 3.13: *Electrostatic surface potential for CSP41b-PRIN2 complex calculated by AlphaFold2 and the complex in cartoon. (A) The regions represented in blue correspond to positively charged regions, while the regions represented in red correspond to negatively charged patches. (B). CSP41b is highlighted in green and PRIN2 is highlighted in cyan. The helices are drawn in ribbons and β-sheets in arrows.*

3.5. TESTING CSP41b and PRIN2 *in vivo* INTERACTIONS

3.5.1. Bimolecular fluorescence complementation assay for CSP41b and PRIN2

In order to test protein interactions in vivo, the coding sequences of *PRIN2* and CSP41b were cloned, using RT-PCR, in translational fusions with fluorescent markers (GFP, NY and YC) under the strong ubiquitous and constitutive CaMV35S promoter (refer methods 2.4.1). The localisation pattern of CSP41b-PRIN2 is expected to be in the chloroplast like the PAPs due to the presence of predicted chloroplast transit peptide. The localisation pattern of the construct 35S::CSP41b-GFP (pSSM43) was established in onion epidermal cells using transient expression (Figure 3.14 A).The signal is observed in many tiny spots of about 1 micrometer corresponding to a similar pattern observed with the strictly chloroplastic PAP10-GFP or PAP10-DsRed known to accumulate in the onion epidermoplasts. In addition, the fluorescent signal is also present in strings connected to the plastid known as stromules and generated by evagination of the plastid envelope containing some stroma. This result therefore indicates that the protein can freely move within the stroma or is attached to the inner membrane. This could be due to an over expression effect conducting to the saturation of the natural sites of CSP41b, or be specific to CSP41b that is loose in the stroma (Figure 3.14 B). It was not possible to detect a convincing signal in the nucleus even when plastids accumulate around it (Figure 3.14 B, cell on the right panel) marking a clear difference with dually-localized PAPs such as PAP5, 7, 8, and 12.

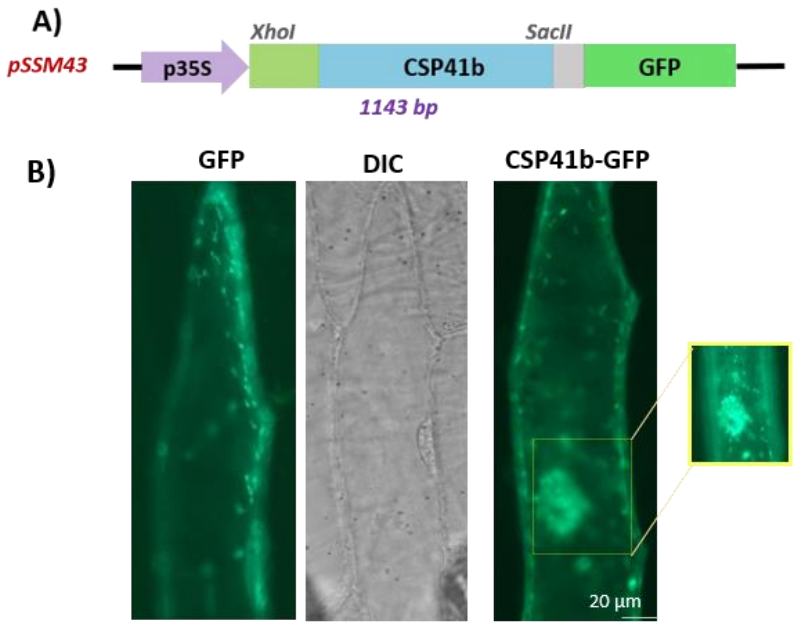


Figure 3.14: *Localisation of CSP41b-GFP in onion cells. (A) pSSM43 construct with CSP41b cTP and coding region with GFP. (B) Images showing CSP41b-GFP signal in plastids.*

The localisation pattern of PRIN2-GFP corresponding to the genetic construction pMCD09 (35S::PRIN2-GFP) was established along with pBB301 (35S::PAP10-DsRed) in onion epidermal cells using transient expression (Figure 3.15 A, B). PAP10 is reported to be a thioredoxin specific to the PEP for which the localisation pattern in GFP fusion, as observed in confocal microscopy, is strictly specific to nucleoids with 1 or 2 speckles detected in plastids and no leakage in the stroma as well as no signal in stromules (Liebers et al., 2020). The GFP signal of PRIN2-GFP is concomitant to the DsRed signal indicating that PRIN2 is localized in plastid as it was already described. The merged image confirms that the protein accumulation pattern of PRIN2 and PAP10 is superimposable (Figure 3.15 C) although the resolution of the epifluorescence microscopy does not allow concluding whether PRIN2 is strictly present in the nucleoid. However, this experiment is comforting the strategy to use fluorescent tags to perform protein-protein interaction assays in BiFC.

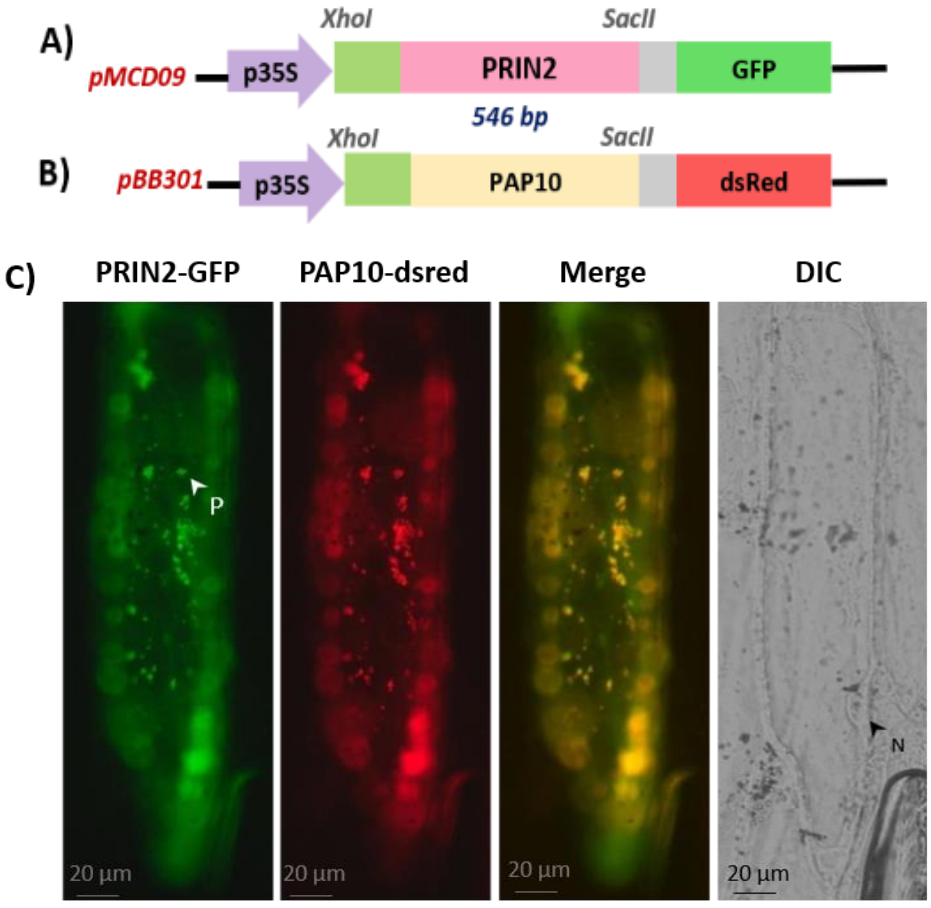


Figure 3.15: *Localisation of PRIN2-GFP in onion cells. (A) pMCD09 construct with PRIN2 cTP and coding region with GFP. (B) pBB301 construct with PAP10 cTP and coding region with dsRed. (C) Images showing PRIN2-GFP signal and PAP10-dsred signal at plastids. Nuclei (N) and plastid (P) are pointed by arrowheads as observed under DIC.*

After testing the localisation pattern of individual proteins in fusion with GFP, the BiFC assay was attempted with the split-YFP genetic constructions pSSM42a and pSSM41a having CSP41b with N-YFP and C-YFP at the respective terminals (Figure 3.16 A, B). Signal was observed and localised in plastids. The signal could be due to either the formation of CSP41b homodimers or to the close proximity of CSP41b at their localisation sites reuniting the fragmented YFP, that provides the fluorescent signal. The second hypothesis being particularly valid when the localisation is very restricted such as in the nucleoid or the PEP. The Coilin construction was used as an internal control of transfection particularly useful when the BiFC signal is weak. In this case, the BiFC signal was strong and present in every transfected cell as observed with the marking of the Cajal bodies (Figure 3.16 C). Both constructions using CSP41b and the split YFP fragments are therefore useful for other BiFC assays

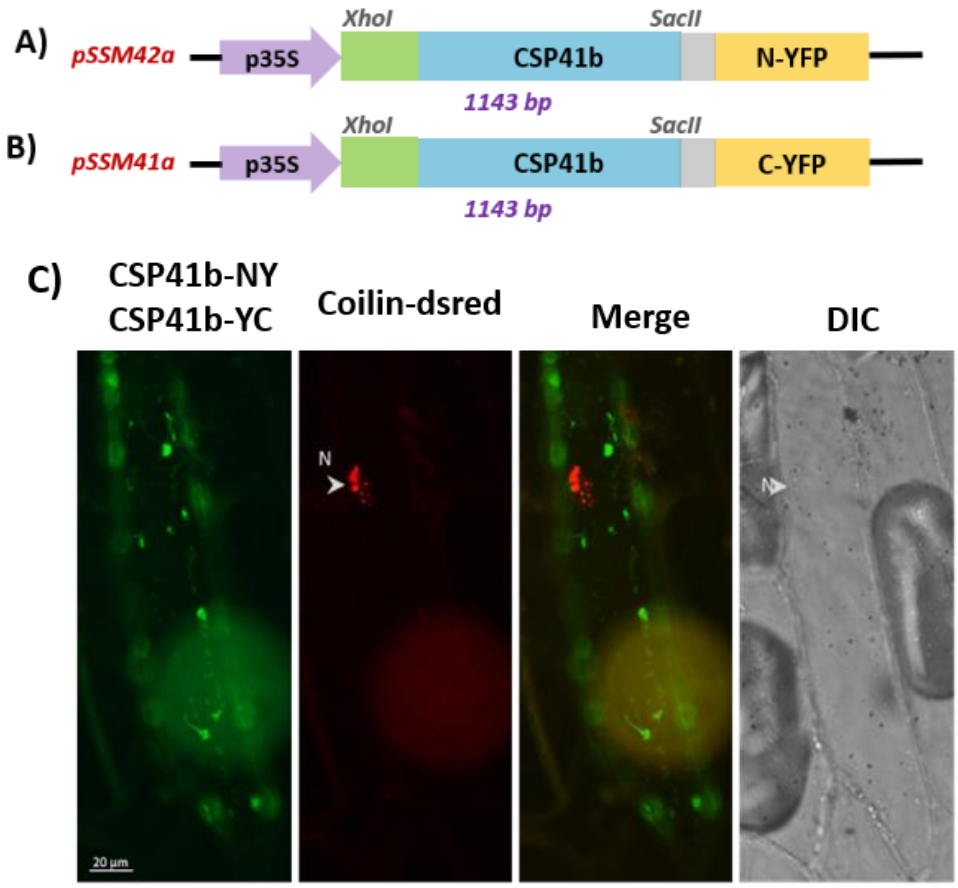


Figure 3.16: *BiFC localisation of CSP41b-NY and CSP41b-YC constructs. (A) (B) – pSSM42a and pSSM41a constructs with CSP41b cTP and coding region with truncated N- and C-terminus of YFP respectively. (C) Images showing CSP41b BiFC signal in plastids. The use of Coilin-dsred, as internal control of transfection, shows transfected cells with a typical Cajal bodies localised signal. Nuclei (N) is pointed by arrowheads as observed under DIC.*

Similarly, PRIN2 genetic constructions, 35S::PRIN2-NY (pMCD07) and 35S::PRIN2-YC (pMCD08) (Figure 3.17 A, B) were also tested. The observed signals were similar to that of PRIN2-GFP alone. As previously described the Coilin-DsRed was used as an internal transfection control with similar outputs. The BiFC signal using PRIN2 was unexpectedly as strong as the PRIN2-GFP signal alone (Figure 3.17 C).

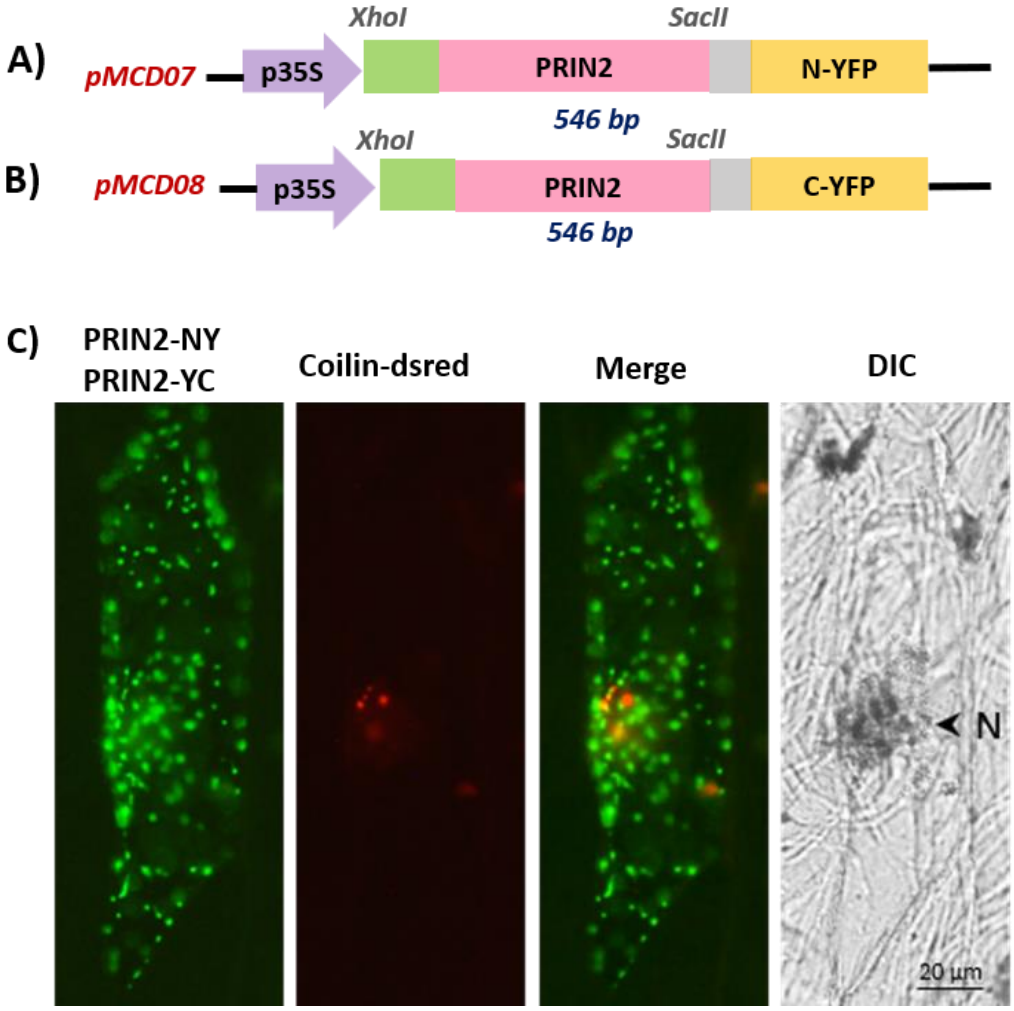


Figure 3.17: *BiFC localisation of PRIN2-NY and PRIN2-YC constructs. (A) (B) pMCD07 and pMCD08 constructs with PRIN2 cTP and coding region with truncated N- and C-terminus of YFP respectively. (C) Images showing PRIN2 BiFC signal and Coilin-DsRed signal in plastids. Nuclei (N) is pointed by arrowheads as observed under DIC.*

The next strategy was to test whether CSP41b and PRIN2 could interact *in vivo* using transient expression in onion cells. The genetic constructions pSSM42a (35S::cTP-CSP41b-NY) and pMCD08 (35S::cTP-PRIN2-YC) were used in association with the internal control (35S::Coilin-DsRed) (Figure 3.18 A, B). A clear signal was observed in plastids (Figure 3.18 C). The merged

imaged shows the distinction between the Cajal body within the nucleus and the plastids. These experiments provide a proof of concept that the two putative partners CSP41b and PRIN2 can be fused to fluorescent tags and localise in the same territory of the chloroplast. A question remain concerning the functionality of these recombinant proteins. This question should only be answered using functional complementation of the mutant phenotype with the respective recombinant gene.

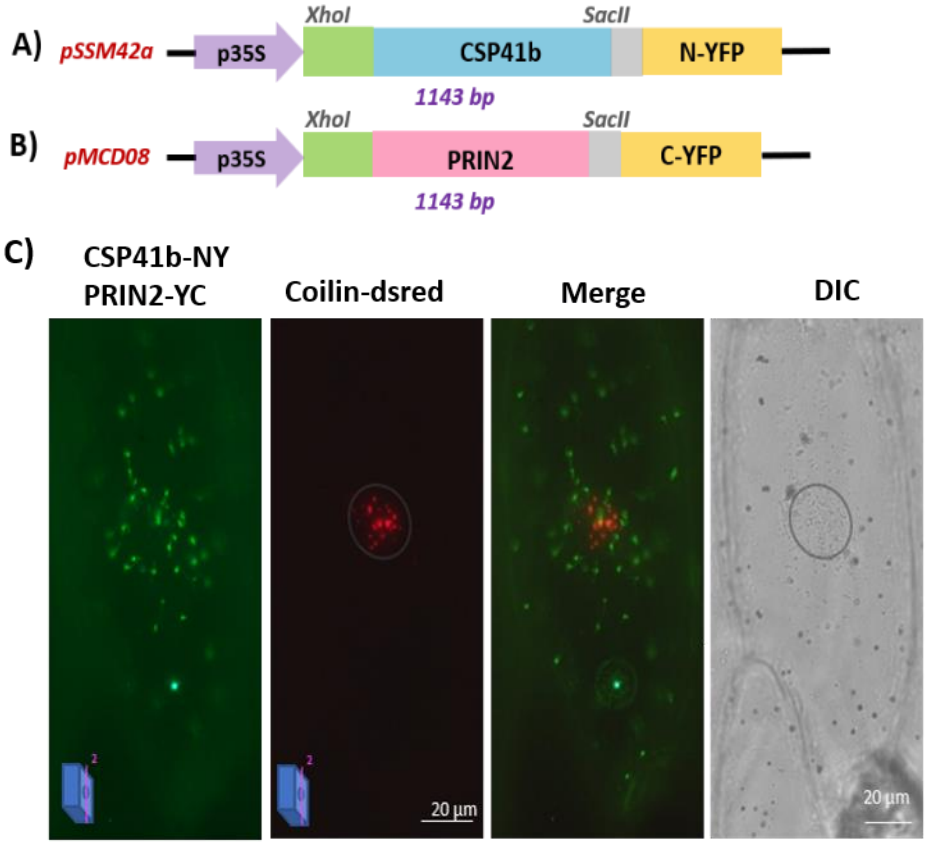


Figure 3.18: *BiFC localisation of CSP41b-NY and PRIN2-YC constructs. (A) (B) pSSM42a and pMCD08 constructs with CSP41b and PRIN2 cTP and coding region with truncated N- and C-terminus of YFP respectively. (C) Images showing CSP41b-PRIN2 BiFC signal in plastids. pRB1001 (Coilin-DsRed) signal corresponds to localisation in Cajal bodies. The pictures corresponding to the GFP channel were taken in 2 positions in the z-axis as indicated by the blue cell scheme and the red or magenta focal plan.*

PRIN2 BiFC interaction with PAP4 was tested with respective constructs with the truncated regions of N- and C-terminus of YFP in different combinations. (Figure 3.19 A, B, C, D). The signal was localised in plastids in both experiments and was similar to the signals that are observed with PAP specific interactions (Figure 3.19 E, F).

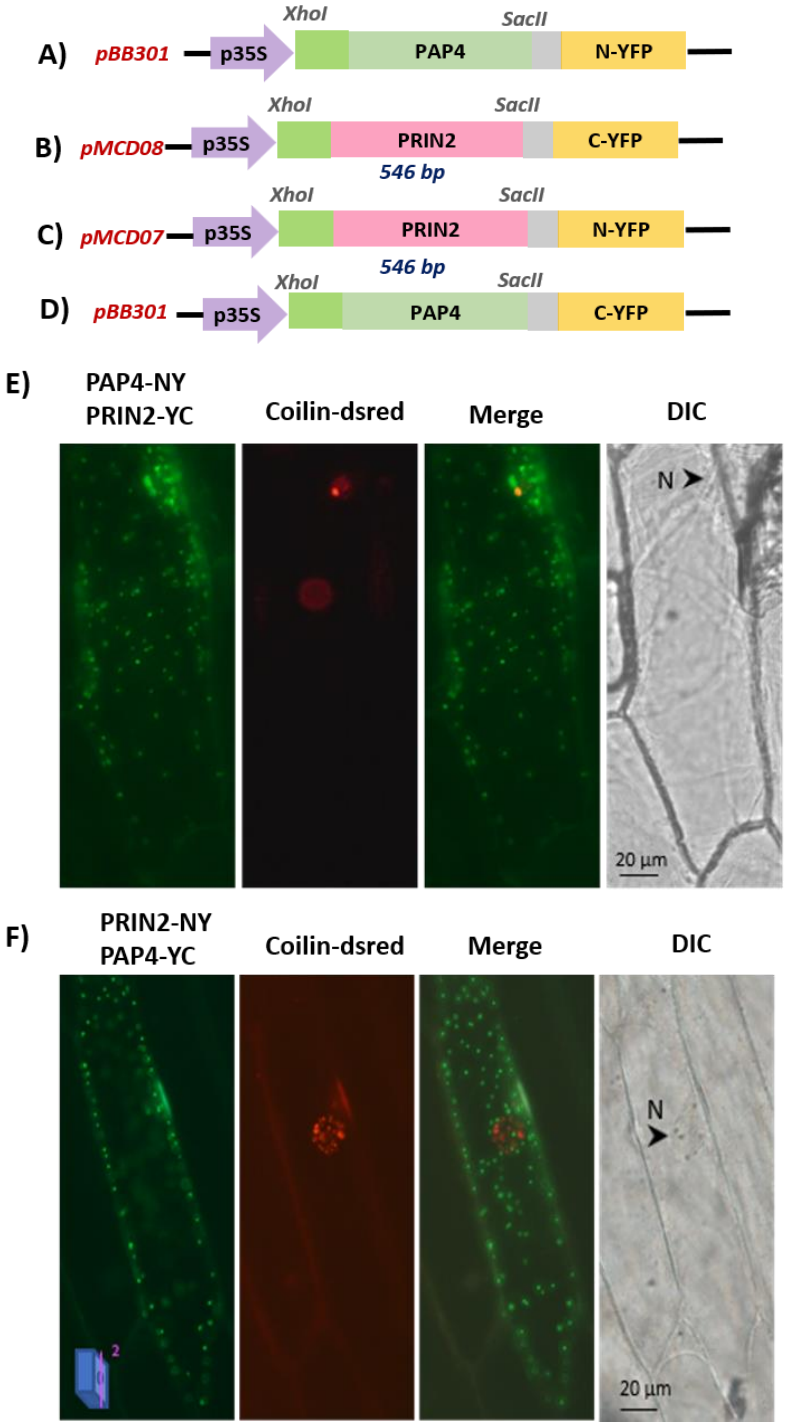


Figure 3.19: *BiFC localisation of PAP4-NY and PRIN2-YC constructs and PRIN2-NY and PAP4-YC constructs. (A) (B) – pMCD09 and pBB301 constructs with PRIN2 and PAP4 cTP and coding regions with truncated C- and N-terminus of YFP respectively. (C) (D) pMCD07 and pBB301 constructs with PRIN2 and PAP4 cTP and coding regions with truncated N- and C-terminus of YFP respectively. (E) (F) Images showing PAP4-PRIN2 BiFC signal in plastids. pRB1001 (Coilin-DsRed) signal corresponds to localisation in Cajal bodies of the nucleus (N). Nuclei (N) is pointed by arrowheads as observed under DIC.*

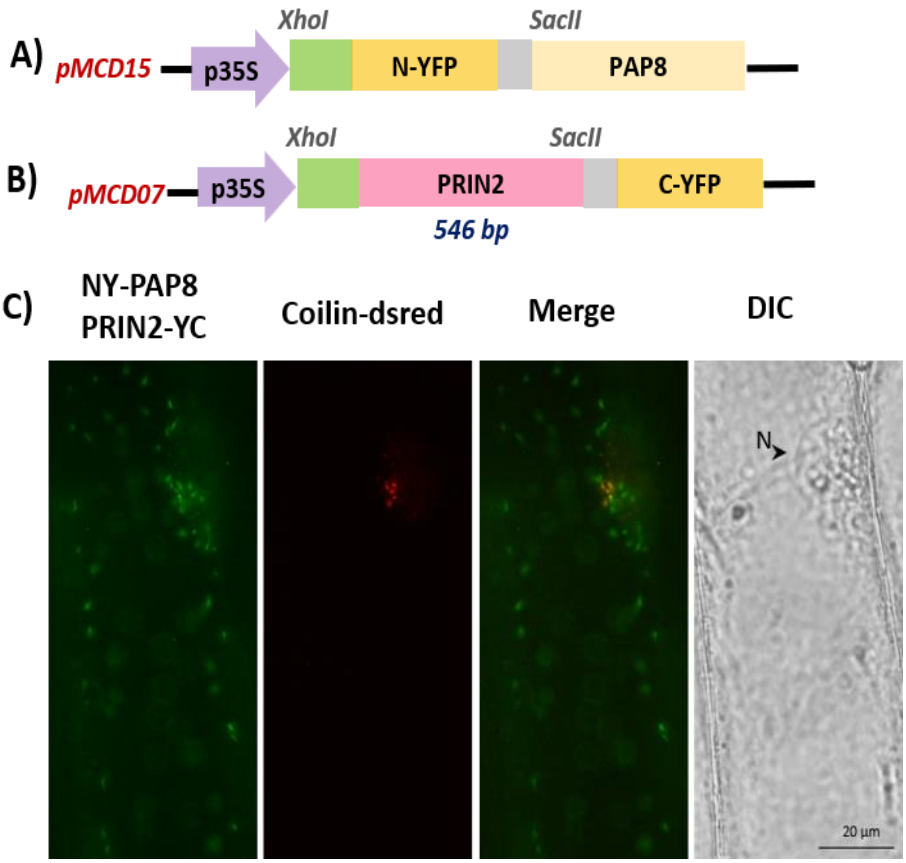


Figure 3.20: *BiFC localisation of NY-PAP8 and PRIN2-YC constructs. (A) (B) pSSM42a and pMCD08 constructs with PAP8 and PRIN2 cTP and coding region with truncated N- and C-terminus of YFP respectively. (C) Images showing PRIN2-PAP8 BiFC signal in plastids. pRB1001 (Coilin-DsRed) signal corresponds to localisation in Cajal bodies. Nucleus (N) is pointed by arrowheads as observed under DIC.*

This shows that PRIN2 is among the proteins that interact closely with PAPs in the active PEP complex. However, PRIN2 interactions with other PAPs such as PAP8 (Figure 3.20 A, B, C) and PAP10, which is a thioredoxin should also be tested to observe their interaction signals. The data from BiFC experiments also confirm that the translated proteins in the living cell context provide interesting patterns that could be challenged in other experiments, such as in planta experiment giving the opportunity to observe the proteins in more relevant plastid types and developmental stages. This also provides a technical confirmation that the regions considered from the gene loci could be further used for complementation experiments, proximity labelling or affinity purification. Other experiments could make use of the structural data to predict some alterations of the surfaces of protein-protein interaction and design mutant clones. All of the above aim at getting comprehensive picture of the functional interactions surrounding the PEP during the delicate transition from dark growth to photomorphogenesis.

3.5.2. Transient assay of PRIN2 proximity labelling in *Nicotiana benthamiana*

The promoter (797 bp), cTP (345 bp) and coding region of PRIN2 (213 bp) was cloned to a proximity labelling vector that has the biotin ligase TurboID (Mair et al., 2019) and Twin streptactin tags. The construct pSSM39 (Figure 3.21) was transformed using *Agrobacterium tumefaciens* to *Arabidopsis thaliana* Col-0 using floral dip method. The seeds from T1 generation were collected and Hygromycin selection should be performed to select the transformed seeds. Proximity labelling is generally done on T3 or T4 to obtain homozygous seeds.

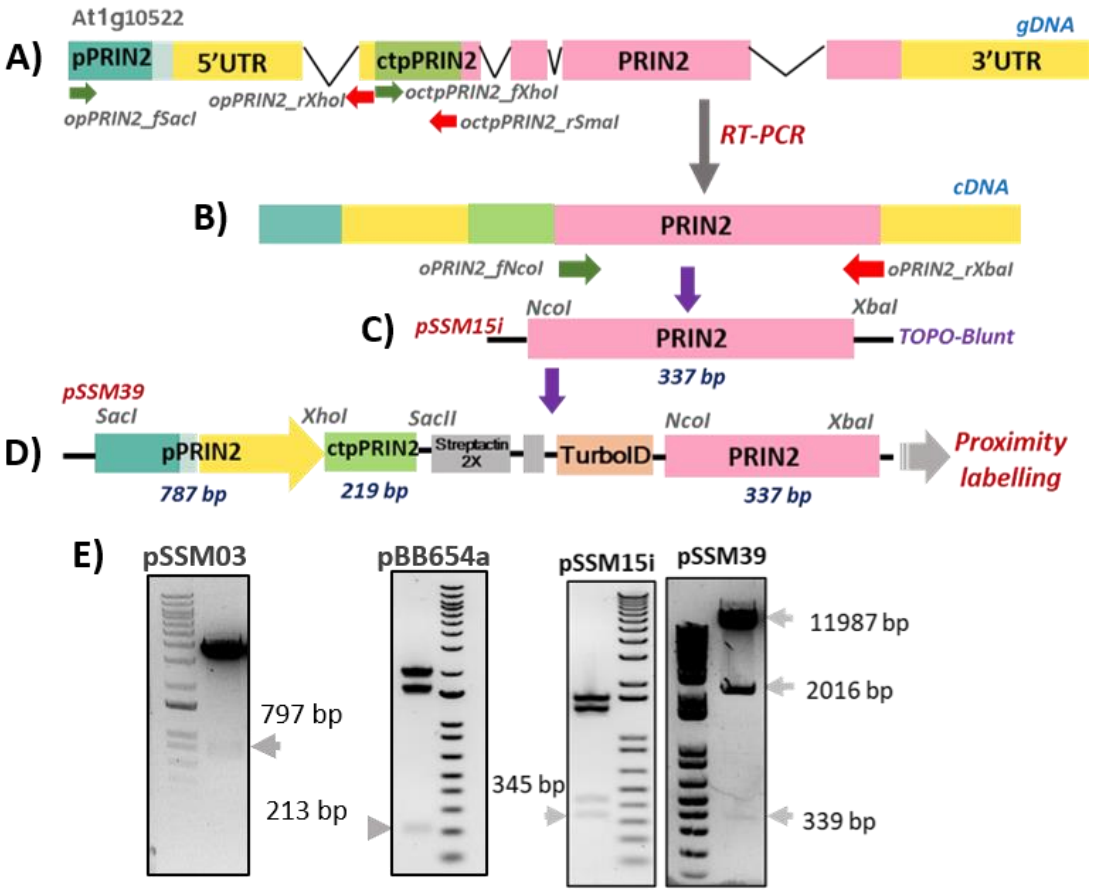


Figure 3.21: *Cloning strategy for PRIN2 regions from gDNA and cDNA of A. thaliana (A) (B) (C). (D) - Proximity labelling gene construct pSSM39. (E) DNA agarose gel pictures showing PRIN2 promoter (pSSM03), cTP (pBB654a), coding region (pSSM15i) and the complete construct cloned in their respective vectors.*

The promoter region of PRIN2 was cloned into a GUS-containing vector (pSSM11) (Figure 3.22), for testing the functionality of the promoter region that was chosen from the PRIN2 gene locus. Transient expression of the GUS clone in *Nicotiana benthamiana* leaves followed by the GUS assay was performed. pBJB16 was used as a positive control. After 24 hours, the

positive clone displayed blue coloration. However, the PRIN2 promoter::GUS clone did not show any coloration after 24 hours. The leaves in which the assay was performed was mature leaves and not young leaves. The activity of PRIN2 promoter region could be the highest during the late to early stages between skoto- to photomorphogenesis. The experiment has to be repeated in young leaves followed by GUS assay in *Arabidopsis thaliana*.

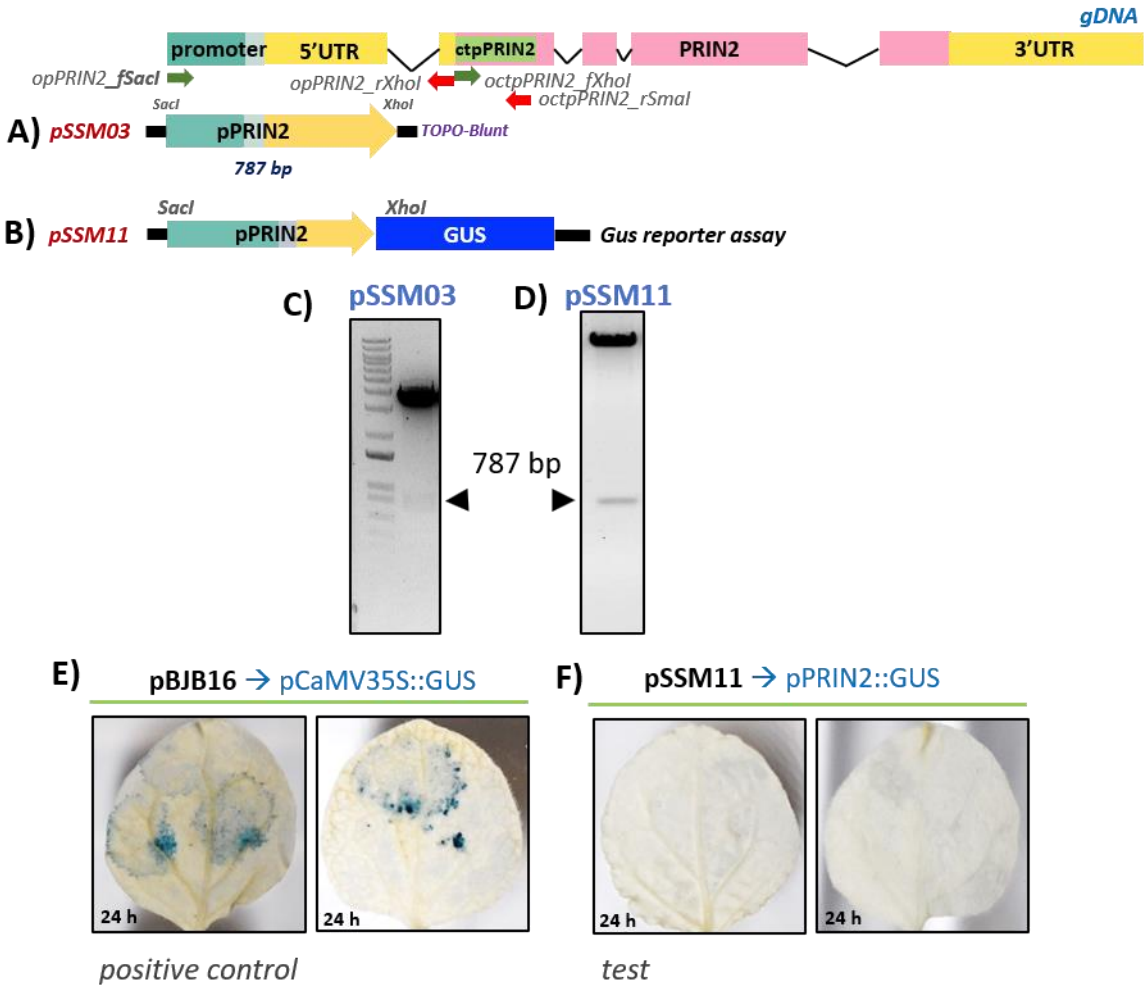
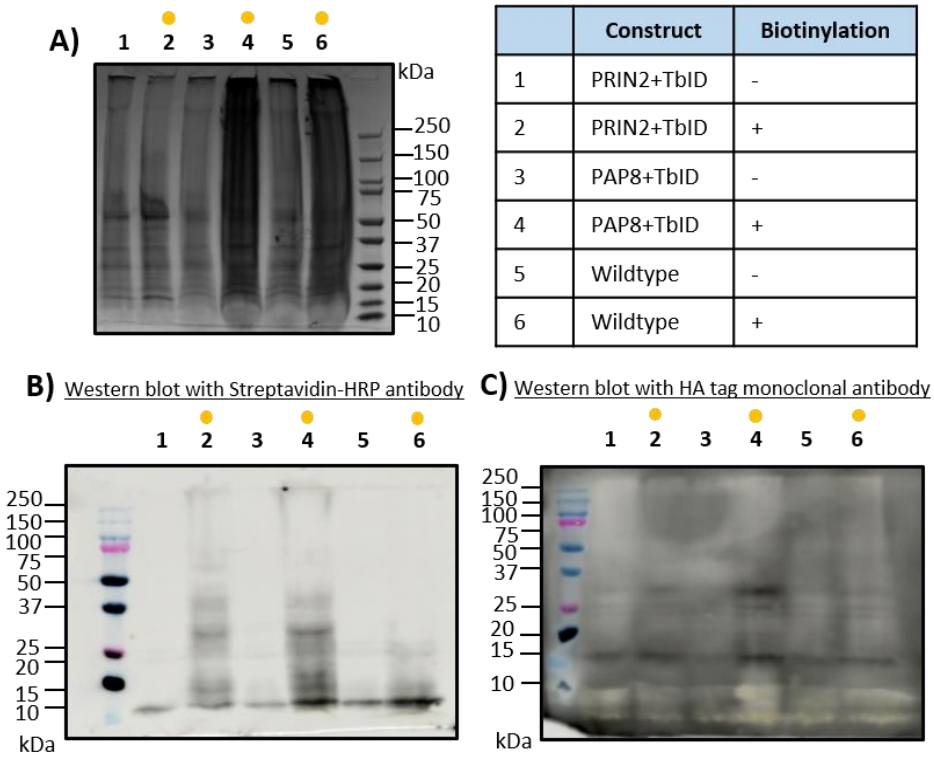


Figure 3.22: *Cloning strategy of PRIN2 promoter for gus assay. (A) strategy for pPRIN2 region pSSM03 from gDNA of A. thaliana. (B) pPRIN2::GUS gene construction pSSM11. (C) (D) DNA agarose gel pictures showing PRIN2 promoter (pSSM03) and pPRIN2::GUS plasmid (pSSM11)*

Proximity labelling was attempted in *Nicotiana benthamina* plants using transiently expressed constructions. Specifically, two test samples, pSSM38 and pSSM39, were utilized for this experiment. A positive control, pFX024, was chosen because previous proximity labelling experiments on PAP8 had been successfully conducted by Dr. Francois-Xavier Gillet and Dr. Robert Blanvillain in the lab. To keep the experiment consistent with previous findings, the

plants were not incubated for more than 30 minutes with exogenous biotin, as previous experiments had demonstrated that 30 minutes were sufficient for effective biotinylation. Consequently, no specific kinetics analysis of the labelling process was conducted before this experiment. For western blot analysis, two separate antibodies were employed to detect the presence of the protein encoded in the pSSM39 construct, which features both a twin strep tag and a HA tag along with TurboID and PRIN2. The results, as depicted in Figure 3.23 B and C, indicated a notably higher level of detection in biotinylated samples (lanes 2 and 4) in both blots



	Construct	Biotinylation
1	PRIN2+TbID	-
2	PRIN2+TbID	+
3	PAP8+TbID	-
4	PAP8+TbID	+
5	Wildtype	-
6	Wildtype	+

Figure 3.23: *(A) Biotinylated proteins from Nicotiana benthamiana. The lanes marked with yellow circles are samples treated with exogenous biotin. The construct pSSM39 has PRIN2 ctp and coding region tagged with the biotin ligase TurboID, twin-strep tag and HA tag. The construct pFX024 has the full coding region of PAP8 fused with TurobID, twin strep tag and HA tag. The biotinylated proteins are revealed using Streptavidin-HRP blot (B) and monoclonal HA antibody-anti-HRP blot (C).*

The blot using Streptavidin-HRP clearly indicates that upon a biotin treatment a number of proteins are marked in rather specific patterns (Fig. 3.23 B). In this first approach it seems that PRIN2-TbID biotinylation pattern is slightly weaker than that of PAP8 but much stronger than that of the wild type. Although fuzzy, it appears that some proteins are shared between the

profiles of PAP8 and PRIN2 and some not. Therefore, this transient assay provides preliminary data for PRIN2 proximity labelling demonstrating that the PRIN2 genetic construction is working and the PRIN2 “proxisomics” could be obtained in *Arabidopsis thaliana*. Modifications can be made in the protocol for isolating the biotinylated proteins from the leaves and the expression pattern has to be confirmed.

4.DISCUSSIONS

4.1. Improving protocol for structural characterisation of PEP complex

The conditions for chloroplast fractionation were optimised for a better PEP purification. Different percoll gradients such as 20%, 40% and 80% / 40% and 80% were tested for find which density gradient would give higher concentrations of intact chloroplasts. The gradient densities with 20%, 40% and 80% percoll gradients consistently provided intact chloroplasts in maximum yield. The chloroplasts were provided for PEP purification. The PEP purification from chloroplasts of *Sinapis alba* has to be improved to achieve better resolution of its 3D structure by cryo-EM. Another experimental approach, similar to studies performed by Floris and Kühlbrandt in 2021, involves utilizing cryo-electron tomography on chloroplast extracts or *Arabidopsis thaliana* cell extracts for identifying the PEP complex. Cryo-ET allows visualization of high-molecular-weight complexes in native cellular extracts. Alongside its technical challenges, the structure of ribosomes and PEP complex in the cellular layer should be distinct, as the shape of the PEP envelope is known. It provides a valuable means to study the PEP complex in TAC and the expressome within the chloroplast. Additionally, an alternative approach is to subject the PEP sample to DNA crosslinking or introduce DNA and sigma factors after RNase digestion. This modification should allow capturing the PEP complex in its transcriptionally active state, offering insights into its functional activity during transcription.

4.2. PAP4 and PAP9 superoxide dismutases

PAP4 and PAP9 are SODs with varying catalytic activities. Notably, PAP9, when purified from *E. coli*, exhibited no SOD activity, a characteristic attributed to the presence of zinc instead of iron in its catalytic site. This alteration may have been during over-expression or purification processes, ultimately abolishing its catalytic function. In plant studies, it was observed that PAP9 displayed lower activity than PAP4 (Myouga et al, 2008). However, the transcripts of the FSD2 gene were found to accumulate to a lesser extent than those of FSD3 in young seedlings. The presence of SODs in the PEP complex has been associated with their potential protective role against ROS generated during photosynthesis. This hypothesis, although untested, raises questions about the precise functions of these SODs in the PEP. Despite the albino phenotype observed in the double mutant pap4/pap9, it remains uncertain whether the SODs serve only a structural role in the PEP, a catalytic and structural role, or other functions. To elucidate the

roles of PAP4 and PAP9 in the PEP, it is essential to mutate their catalytic sites and investigate their structural and catalytic functions *in planta*. This can be achieved by complementing the *pap9/pap9* or *pap4/pap4* mutants with the catalytically inactive PAP4 or PAP9. To ensure proper protein folding after catalytic site mutation, the modified PAP4 and PAP9 should be produced in *E. coli*, purified, and characterized. Techniques such as X-ray crystallography can be employed for structural characterization. Considering the possibility of zinc contamination in PAP4 and PAP9, it may be valuable to explore their presence within the PEP. Starting with pure PEP preparations from young plants, the presence of metals, including Fe^{3+} and Zn^{2+}, can be investigated using techniques like X-ray absorption near edge spectroscopy (XANES), Extended X-ray absorption fine structure (EXAFS), or inductively coupled plasma mass spectrometry (ICP-MS). These methods are dependent on the protein quantity required for the analysis. It is worth noting that analysing metal coordination in 3D protein structures may be limited, especially for metals like Fe^{3+} and Zn^{2+}, which exhibit similarities as seen in the PAP9 structure.

4.3 CSP41b and PRIN2 interaction *in vitro* experiments

The structure of CSP41b was characterised by cryo-EM, achieving a resolution of 3.4 Å. Through this analysis, it was found that the CSP41b monomer shares a structural fold common to that of the UDP-N-acetylglucosamine 4-epimerase found in *Bacillus cereus*. Intriguingly, several basic residues that are conserved in CSP41b are situated within a spacious crevice formed by both the N-terminal region, which contains a Rossmann fold and the C-terminal domain. These conserved positively charged residues could likely play a role in stabilizing stem-loop RNAs. Notably, CSP41b exists as a dimer, which raises questions about its capacity to stabilize two stem-loop RNA molecules simultaneously or whether it operates as a monomer following RNA binding. To further understand this complex, it is proposed that investigations into the 3D structures of CSP41b in association with RNA be conducted using methods such as X-ray crystallography or 2D-NMR, where CSP41b is labelled. These approaches offer insights into the dynamics of the complex. In the event that the complex remains a dimer in solution (as observable in NMR if there are no signal changes), cryo-EM can be employed to reveal its 3D structure.

Meanwhile, docking simulations can be used to postulate potential RNA binding mechanisms. This suggests that CSP41b may bind mRNA stem-loops within the crevice (Figure 4.1 A, B). Consequently, it is essential to perform structural studies to characterize this complex. The docking experiments were performed by Dr. David Cobessi.

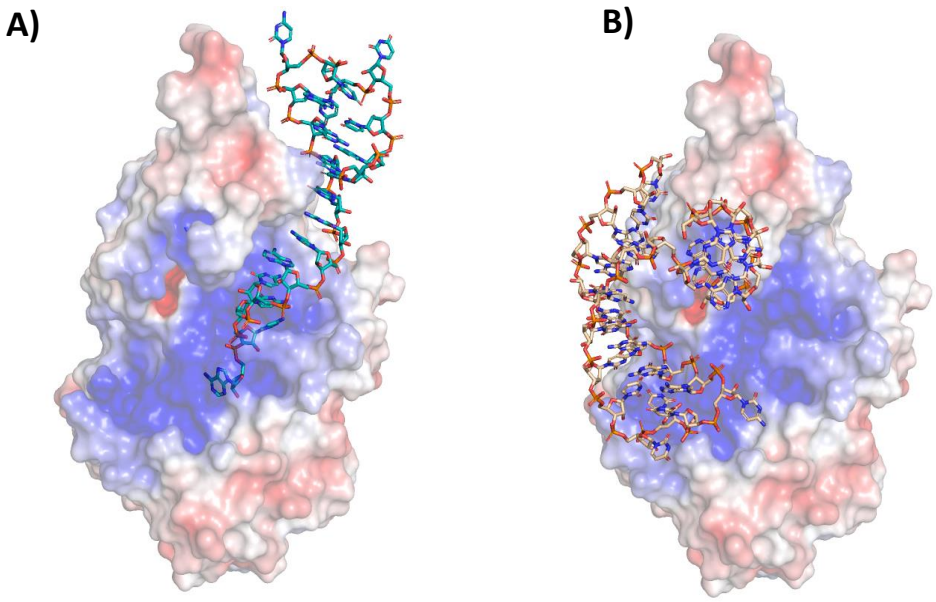


Figure 4.1: *(A) Docking of CSP41b with the stem loop UCUUUUCAGAGCCACCCA (PDB entry: 4L8R; Tan et al., 2013). (B) Docking of CSP41b with the stem loop GGCCAAAGGCCCUUUUCAGGGCCACC (PDB entry: 4TUX; Zhang et al., 2013). The regions represented in blue correspond to positively charged regions, while the regions represented in red correspond to negatively charged patches. The DNA is represented as sticks.*

Sequence comparisons have indicated the presence of CSP41b in *Archaea* and in a majority of photosynthetic organisms. Therefore, it might not be essential for PEP transcription, but its significance may lie in transcription by prokaryotic RNA polymerases or in the connection between transduction and transcription in photosynthetic organisms, including chloroplasts. This involvement may occur as a monomer that interacts with mRNA within a protein complex bridging the ribosome and RNA polymerase.

Intriguingly, mass spectrometry analyses of PEP proteins from Ruedas et al, 2022, have revealed the presence of PRIN2 and CSP41b, alongside ribosomal proteins. The potential involvement of these proteins in a related mechanism can be investigated through studies of the transcription/translation complex, known as the expressome, within the chloroplast, similar to studies conducted in bacteria (Webster et al. 2020) (Webster and Weixlbaumer,

2021). As CSP41b has origins in cyanobacteria, it may interact with one or more catalytic subunits of RNA polymerase in these organisms, suggesting its potential interaction with catalytic PEP subunits within chloroplasts. Furthermore, it may be in proximity to PAP8, as CSP41b was identified in the pool of proteins biotinylated in proximity labelling experiments with PAP8 performed by Dr. François-Xavier Gillet.

CSP41b could also play a role in protecting RNA or facilitating RNA modification. CSP41b has been observed in several oligomeric states in plants and as a dimer in these experiments, with its active form remaining uncertain. There is speculation that its involvement may lie in the monomeric form, given the presence of a patch of negative residues close to the dimer interface. mRNA binding in the positively charged crevice could lead to conformational changes involving crevice closure and the movement of these patches, possibly resulting in monomerization. Consequently, further structural studies of CSP41b bound to oligonucleotides are essential. In addition to CSP41b's binding to mRNA, it interacts with the monomeric protein PRIN2 (Kremnev et al., 2014; Diaz et al., 2018). PRIN2 operates as a monomer within the PEP and combining this information with the hypothesis of CSP41b monomerization upon mRNA binding, it is likely that CSP41b exists as a monomer when associated with PRIN2 and mRNA. While EMSA and co-IP experiments have suggested a weak interaction between PRIN2 and CSP41b, it is necessary to confirm this interaction through other biophysical methods such as ITC and SEC-MALLS.

The modelling of the CSP41b and PRIN2 complex has revealed that the binding of PRIN2 does not obstruct the positively charged crevice essential for mRNA binding (Figure 4.2). Also, specific residues from the first TAD boxes of PRIN2 interact with CSP41b, while the other two TADs remain readily accessible (Figure 4.2). To gain a better understanding of this complex, further experiments are required to successfully isolate it. Alternatively, both PRIN2 and CSP41b can be introduced along with a stem-loop mRNA into the PEP to fully characterize the complex. In addition to its three TAD boxes, PRIN2 also possesses two conserved cysteine residues. The first cysteine, located in the N-terminal region, was omitted from our genetic construct designed for PRIN2 overexpression. This cysteine plays a role in the formation of a disulfide bridge, enabling PRIN2 to dimerize. In its monomeric form, PRIN2 may potentially bind to Trxz within the PEP to activate PEP. In this monomeric state, it could also interact with CSP41b, facilitating the coupling of transcription and translation. Diaz et al. have proposed

that the electron flow involved in the redox reaction of PRIN2 is associated with the FTR/TRX system, a system closely linked to photosynthesis. However, this system's involvement with photosystem occurs only after the initial assembly of the photosystem. As a result, the electrons used in the redox reaction must originate from another chloroplastic redox pathway before the assembly of photosystem.

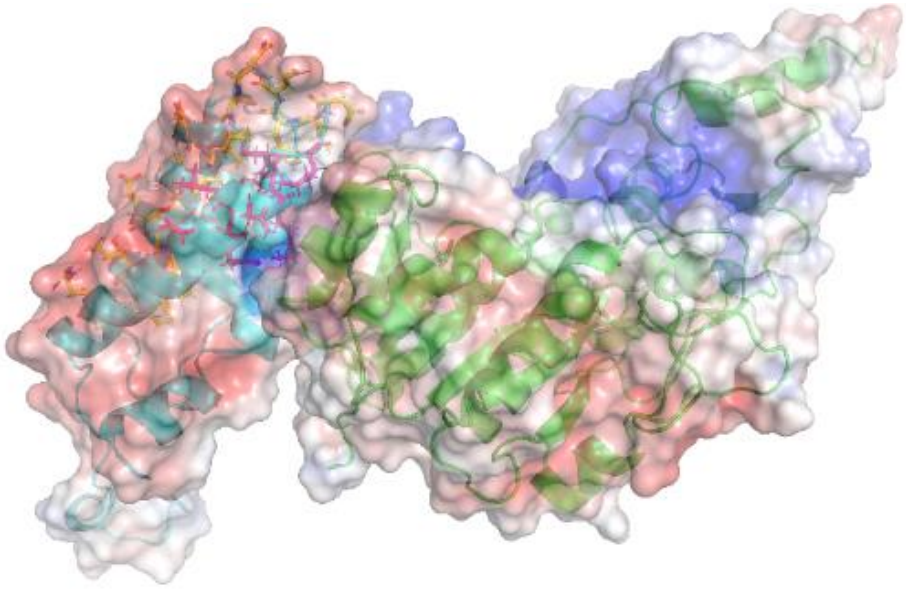


Figure 4.2: *PRIN2/CSP41b complex with the residues of the 3TADs of PRIN2 depicted as sticks. The first TAD is coloured in pink, PRIN2 is coloured in cyan and CSP41b in green. The regions represented in blue correspond to positively charged regions, while the regions represented in red correspond to negatively charged patches.*

Given the observed weak interactions between CSP41b and PRIN2, the BS3 crosslinker was introduced to stabilize these complexes for cryo-EM analyses. But, the negative-staining electron microscopy experiments revealed a non-uniform solution. Further modifications of the purification conditions for the crosslinked complexes are necessary to facilitate structural analyses through cryo-EM. The most likely alternative experiment involves determining the 3D structure of PEP bound to PRIN2 and CSP41b through cryo-EM, as purifying all three partners have been successful. It should be noted that only a weak interaction between PRIN2 and CSP41b was observed by *in vitro* experiments. To confirm this interaction, various biophysical experiments must be performed, followed by thorough characterization using structural approaches under varying physio-chemical conditions, such as buffer compositions, concentration of proteins and isolation temperatures.

4.4. *In vivo* interaction experiments on PRIN2 with CSP41b and PAPs

The BiFC constructs of CSP41b and PRIN2 designed for testing *in vivo* interactions within onion epidermal cells were successfully cloned into their respective vectors. These genetic constructions contained truncated N-terminal and C-terminal YFP regions, as well as GFP and RFP coding regions. BiFC experiments were also conducted with PRIN2 and PAP4, and fluorescence was observed in a location similar to that observed for PAP8 and PAP5 (Figure 1.4) (Liebers et al., 2020). This pattern is also similar to the signal observed in the BiFC studies of CSP41b/PRIN2. This suggests that PRIN2/CSP41b and PAP4 might be in close proximity within the PEP complex. It is worth considering the hypothesis that these proteins could potentially relocate to the PEP surface to repair damage caused by reactive oxygen species. In such a scenario, fluorescence might not be observed. This observation suggests that the CSP41b/PRIN2 complex, if it exists, is located within a region of the chloroplast where protein and RNA concentrations are high, similar to the expressome, for example. BiFC experiments could also be conducted using PAP9 and PAP10/TrxZ, as they also participate in scavenging ROS, to determine if they also possess similar fluorescence patterns. Nevertheless, it is essential to consider the possibility that the close proximity of both proteins allows YFP to reconstitute, resulting in fluorescence. In such a scenario, the concentration of the two proteins in close proximity might be sufficient for fluorescence to be detected. To address this potential spatial proximity, other constructs can be used by cloning N-terminal-YFP and C-terminal-YFP at the N-termini and C-termini of both PRIN2 and CSP41. The length of the linker must also be carefully considered to ensure that YFP reconstitution occurs within a PRIN2/CSP41b complex and not solely due to proximity. The precise localisation of these proteins within the PEP can be determined by characterising the 3D structure of the PEP complex. To further identify proteins in their vicinity, cross-linking experiments coupled with mass spectrometry can be performed using pure PEP preparations. This technique, combined with trypsin digestion, can describe subunit interactions, as previously demonstrated by Ruedas et al., 2022. It is essential to complement these BiFC experiments with additional approaches, as close proximity of proteins can lead to fluorescence even in the absence of direct interaction.

One complementary method is proximity labelling combined with affinity purification, making use of the cloning with a twin-Strep Tag because it enhances the precision of the experiments.

Affinity purification provides insights into protein interactions under native conditions, while proximity labelling uncovers both strong and transient interactions as it operates *in vivo* and does not require native conditions for further preparation of the samples. It could also confirm the interaction between PRIN2 and CSP41b, as well as to explore the interactome and proxisome of PRIN2. The components of the PRIN2 gene, including the promoter, chloroplast transit peptide, and open reading frame, have been successfully cloned into the proximity labelling vector and are ready for transformation in mutant *A. thaliana* using *Agrobacterium tumefaciens*. Based on the initial transiently expressed GUS assay conducted in *Nicotiana benthamiana* leaves, there was no detectable expression of the GUS reporter gene driven by the PRIN2 promoter construct in mature leaves. It is possible that the selected region for the PRIN2 promoter may exhibit high activity in very young leaves rather than in mature leaves, which were the focus of this initial assessment. To explore this possibility further, it may be advisable to replicate the experiment in young leaves to confirm whether the PRIN2 promoter demonstrates increased activity in that specific leaf developmental stage. Working on the preliminary experiments conducted with the PRIN2 proximity labelling construct that were transiently expressed in *Nicotiana benthamiana* plants, the construct pSSM39 has expression of biotinylated proteins as observed in the western blot (Figure 3.23 B, C). It is also worthwhile to investigate the expression patterns using the plasmid pSSM38. Notably, pSSM38 features the PAP8 promoter instead of the PRIN2 promoter. Since the expression patterns of PAP8 and PRIN2 are similar, and the chosen region of the PAP8 promoter has previously demonstrated success in expressing the TurboID construct within the pFX024 construct, it presents a logical choice for further investigation. Moreover, in the context of western blot analysis, an antibody specific to the PRIN2 coding region can be employed alongside the twin-Strep Tag and HA antibodies. This multi-antibody approach provides a confirmation of the expression patterns and interactions occurring within the system. These experiments have the potential to uncover the redox partners that activate PRIN2 before photosystem assembly in the chloroplast. The experiments can be conducted under both light and dark conditions, enabling the observation of the expressome through proximity labelling coupled with affinity purification.

5. CONCLUSIONS

In the first section, my contribution focussed on enhancing the protocol of chloroplast fractionation from *Sinapis alba* with the aim of enhancing the efficacy of purification. It discusses the importance of obtaining chloroplasts of sufficient yield and outlines the potential application of cryo-electron tomography to study the PEP complex in its native environment. The utilization of cryo-electron tomography and DNA crosslinking methods provides promising avenues to investigate the PEP complex's transcriptional activity. In the second section, the study delves into the catalytic activities of PAP4 and PAP9, raising questions about their precise roles within the PEP. My participation was involved in estimating the superoxide dismutase activity of the *in vitro* PAP9 and PAP4 proteins. Future experiments involving mutagenesis and structural characterization would shed light on their role and necessity as a superoxide dismutase. Additionally, the investigation of metal coordination within PAP4 and PAP9 could uncover further insights into their roles. The third section delves into the interaction of CSP41b and PRIN2, with implications for their involvement in RNA binding and potential redox mechanisms. The need for structural studies and biophysical experiments is evident to confirm their interactions and unravel their roles during chloroplast biogenesis. In the fourth section, *in vivo* interaction experiments, such as BiFC and proximity labelling, provide valuable insights into the spatial relationships between various proteins within the PEP complex. Further experimentation in BiFC, where proteins could be mutated and ligating the N-terminal and C-terminal truncated YFP at the either ends of the protein is needed to confirm and characterize these interactions under different *in vivo* physiological conditions. Complementary approaches such as proximity labelling would help in exploring potential associations in native conditions, prompting further exploration of these protein interactions and their roles in mitigating the impact of reactive oxygen species. Complementary approaches, such as proximity labelling combined with affinity purification, provide promising aspects for investigating these interactions under native conditions. These experiments hold promise in elucidating the regulatory mechanisms involving PRIN2 and its redox partners.

6. Publications

1. Ruedas Rémi, Soumiya Sankari Muthukumar, Sylvie Kieffer-Jaquinod, François-Xavier Gillet, Daphna Fenel, Grégory Effantin, Thomas Pfannschmidt, Yohann Couté, Robert Blanvillain, and David Cobessi. (2022). 'Three-Dimensional Envelope and Subunit Interactions of the Plastid-Encoded RNA Polymerase from *Sinapis alba*'. *International Journal of Molecular Sciences*. 23, (17): 9922. doi:10.3390/ijms23179922

2. Favier, A., Gans, P., Boeri Erba, E., Signor, L., Muthukumar, S. S., Pfannschmidt, T., Blanvillain, R., & Cobessi, D. (2021). 'The Plastid-Encoded RNA Polymerase-Associated Protein PAP9 Is a Superoxide Dismutase with unusual structural features'. *Frontiers in Plant Science.* 12:668897. https://www.frontiersin.org/articles/10.3389/fpls.2021.668897

7. REFERENCES

Allison, L. A., Simon, L. D., & Maliga, P. (1996). 'Deletion of rpoB reveals a second distinct transcription system in plastids of higher plants'. *The EMBO Journal*, 15(11), 2802–2809. https://doi.org/10.1002/j.1460-2075.1996.tb00640.x

Alscher, R. G., Erturk, N., & Heath, L. S. (2002). 'Role of superoxide dismutases (SODs) in controlling oxidative stress in plants'. *Journal of Experimental Botany*, 53(372), 1331–1341.

Angell, C. A., & Kanno, H. (1976). 'Density Maxima in High-Pressure Supercooled Water and Liquid Silicon Dioxide'. *Science*, 193(4258), 1121–1122. https://doi.org/10.1126/science.193.4258.1121

Armarego-Marriott, T., Kowalewska, Ł., Burgos, A., Fischer, A., Thiele, W., Erban, A., Strand, D., Kahlau, S., Hertle, A., Kopka, J., Walther, D., Reich, Z., Schöttler, M. A., & Bock, R. (2019). 'Highly Resolved Systems Biology to Dissect the Etioplast-to-Chloroplast Transition in Tobacco Leaves'. *Plant Physiology*, 180(1), 654–681. https://doi.org/10.1104/pp.18.01432

Armarego-Marriott, T., Sandoval-Ibañez, O., & Kowalewska, Ł. (2020). 'Beyond the darkness: Recent lessons from etiolation and de-etiolation studies'. *Journal of Experimental Botany,* 71(4), 1215–1225. https://doi.org/10.1093/jxb/erz496

Baginsky, S., Kleffmann, T., von Zychlinski, A., & Gruissem, W. (2005). 'Analysis of shotgun proteomics and RNA profiling data from *Arabidopsis thaliana* chloroplasts'. *Journal of Proteome Research*, 4(2), 637–640. https://doi.org/10.1021/pr049764u

Baginsky, S., Tiller, K., Pfannschmidt, T., & Link, G. (1999). 'PTK, the chloroplast RNA polymerase-associated protein kinase from mustard (*Sinapis alba*), mediates redox control of plastid in vitro transcription'. *Plant Molecular Biology,* 39(5), 1013–1023. https://doi.org/10.1023/A:1006177807844

Balsera, M., Uberegui, E., Schürmann, P., & Buchanan, B. B. (2014). 'Evolutionary Development of Redox Regulation in Chloroplasts'. *Antioxidants & Redox Signaling*, 21(9), 1327–1355. https://doi.org/10.1089/ars.2013.5817

Bédard, J., & Jarvis, P. (2005). 'Recognition and envelope translocation of chloroplast preproteins'. *Journal of Experimental Botany*, 56(419), 2287–2320.

Beligni, M. V., & Mayfield, S. P. (2008). '*Arabidopsis thaliana* mutants reveal a role for CSP41a and CSP41b, two ribosome-associated endonucleases, in chloroplast ribosomal RNA metabolism'. *Plant Molecular Biology,* 67(4), 389–401. https://doi.org/10.1007/s11103-008-9328-2

Bock, R. (2007). 'Structure, function, and inheritance of plastid genomes'. In R. Bock (Ed.), C*ell and Molecular Biology of Plastids* (pp. 29–63). https://doi.org/10.1007/4735_2007_0223

Börner, T., Aleynikova, A. Yu., Zubo, Y. O., & Kusnetsov, V. V. (2015). 'Chloroplast RNA polymerases: Role in chloroplast biogenesis'. *Biochimica et Biophysica Acta (BBA) - Bioenergetics,* 1847(9), 761–769. https://doi.org/10.1016/j.bbabio.2015.02.004

Bowler, C., Montagu, M. V., & Inze, D. (1992). 'Superoxide Dismutase and Stress Tolerance'. *Annual Review of Plant Physiology and Plant Molecular Biology,* 43(1), 83–116. https://doi.org/10.1146/annurev.pp.43.060192.000503

Bowler, C., Van Camp, W., Van Montagu, M., Inzé, D., & Asada, K. (1994). 'Superoxide dismutase in plants'. *Critical Reviews in Plant Sciences,* 13:3, 199-218, DOI: 10.1080/07352689409701914

Bradbeer, J. W., Atkinson, Y. E., Börner, T., & Hagemann, R. (1979). 'Plastid-synthesised RNA may control the cytoplasmic synthesis of plastid polypeptides'. *Nature,* 279(5716), Article 5716. https://doi.org/10.1038/279816a0

Branon, T. C., Bosch, J. A., Sanchez, A. D., Udeshi, N. D., Svinkina, T., Carr, S. A., Feldman, J. L., Perrimon, N., & Ting, A. Y. (2018). 'Efficient proximity labelling in living cells and organisms with TurboID'. *Nature Biotechnology,* 36(9), Article 9. https://doi.org/10.1038/nbt.4201

Brugière, S., Kowalski, S., Ferro, M., Seigneurin-Berny, D., Miras, S., Salvi, D., Ravanel, S., d'Hérin, P., Garin, J., Bourguignon, J., Joyard, J., & Rolland, N. (2004). 'The hydrophobic proteome of mitochondrial membranes from Arabidopsis cell suspensions'. *Phytochemistry,* 65(12), 1693–1707. https://doi.org/10.1016/j.phytochem.2004.03.028

Chambon, L., Gillet, F.-X., Chieb, M., Cobessi, D., Pfannschmidt, T., & Blanvillain, R. (2022). 'PAP8/pTAC6 is part of a nuclear protein complex and displays RNA recognition motifs of viral origin'. *International Journal of Molecular Sciences*, 3059. https://doi.org/10.3390/ijms23063059

Chang, S. H., Lee, S., Um, T. Y., Kim, J.-K., Do Choi, Y., & Jang, G. (2017). 'PTAC10, a Key Subunit of Plastid-Encoded RNA Polymerase, Promotes Chloroplast Development'. *Plant Physiology,* 174(1), 435–449. https://doi.org/10.1104/pp.17.00248

Chieb, M., Liebers, M., Chevalier, F., Lerbs-Mache, S., Blanvillain, R., & Pfannschmidt, T. (2018). 'Determination of the DNA/RNA-Associated Subproteome from Chloroplasts and Other Plastid Types'. In E. Maréchal (Ed.), Plastids (Vol. 1829, pp. 253–271). *Springer US.* https://doi.org/10.1007/978-1-4939-8654-5_17

Cramer, P. (2002). 'Multisubunit RNA polymerases'. *Current Opinion in Structural Biology*, 12(1), 89–97. https://doi.org/10.1016/S0959-440X(02)00294-4

Croll, T. I., & Read, R. J. (2021). 'Adaptive Cartesian and torsional restraints for interactive model rebuilding'. *Acta Crystallographica Section D: Structural Biology,* 77(4), 438–446. https://doi.org/10.1107/S2059798321001145

D. Liebschner, P.V. Afonine, M.L. Baker, G. Bunkóczi, V.B. Chen, T.I. Croll, B. Hintze, L.W. Hung, S. Jain, A.J. McCoy, N.W. Moriarty, R.D. Oeffner, B.K. Poon, M.G. Prisant, R.J. Read, J.S. Richardson, D.C. Richardson, M.D. Sammito, O.V. Sobolev, D.H. Stockwell, T.C. Terwilliger, A.G. Urzhumtsev, L.L. Videau, C.J. Williams, and P.D. Adams. (2019). 'Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix'. Acta Crystallographica Section D: Structural Biology, 75(10), 861-877. https://doi.org/10.1107/s2059798319011471

Davis, M. W., & Jorgensen, E. M. (2022). 'ApE, A Plasmid Editor: A Freely Available DNA Manipulation and Visualization Program'. *Frontiers in Bioinformatics,* 2. https://doi.org/10.3389/fbinf.2022.818619

de Vries, S., de Vries, J., von Dahlen, J. K., Gould, S. B., Archibald, J. M., Rose, L. E., & Slamovits, C. H. (2018). 'On plant defense signaling networks and early land plant evolution'. *Communicative & Integrative Biology,* 11(3), 1–14. https://doi.org/10.1080/19420889.2018.1486168

Díaz, M. G., Hernández-Verdeja, T., Kremnev, D., Crawford, T., Dubreuil, C., & Strand, Å. (2018). 'Redox regulation of PEP activity during seedling establishment in *Arabidopsis thaliana'*. *Nature Communications,* 9(1), Article 1. https://doi.org/10.1038/s41467-017-02468-2

Dubreuil, C., Jin, X., Barajas-López, J. de D., Hewitt, T. C., Tanz, S. K., Dobrenel, T., Schröder, W. P., Hanson, J., Pesquet, E., Grönlund, A., Small, I., & Strand, Å. (2018). 'Establishment of Photosynthesis through Chloroplast Development Is Controlled by Two Distinct Regulatory Phases'. *Plant Physiology,* 176(2), 1199–1214. https://doi.org/10.1104/pp.17.00435

Elizabeth Potterton, Peter Briggs, Maria Turkenburg and Eleanor Dodson. (2003). 'A graphical user interface to the CCP4 program suite'. *Acta Crystallographica Section D Biological Crystallography*, 59(7), 1131-1137. https://doi.org/10.1107/s0907444903008126

Favier, A., Gans, P., Boeri Erba, E., Signor, L., Muthukumar, S. S., Pfannschmidt, T., Blanvillain, R., & Cobessi, D. (2021). 'The Plastid-Encoded RNA Polymerase-Associated Protein PAP9 Is a Superoxide Dismutase with unusual structural features'. *Frontiers in Plant Science.* 12:668897. https://www.frontiersin.org/articles/10.3389/fpls.2021.668897

Fernandez-Leiro, R., & Scheres, S. H. W. (2017). 'A pipeline approach to single-particle processing in RELION'. *Acta Crystallographica. Section D, Structural Biology,* 73(Pt 6), 496–502. https://doi.org/10.1107/S2059798316019276

Ferro, M., Brugière, S., Salvi, D., Seigneurin-Berny, D., Court, M., Moyet, L., Ramus, C., Miras, S., Mellal, M., Gall, S. L., Kieffer-Jaquinod, S., Bruley, C., Garin, J., Joyard, J., Masselon, C., & Rolland, N. (2010). 'AT_CHLORO, a Comprehensive Chloroplast Proteome Database with Subplastidial Localization and Curated Information on Envelope Proteins'. *Molecular & Cellular Proteomics,* 9(6), 1063–1084. https://doi.org/10.1074/mcp.M900325-MCP200

Fettke, J., Nunes-Nesi, A., Fernie, A. R., & Steup, M. (2011). 'Identification of a novel heteroglycan-interacting protein, HIP 1.3, from *Arabidopsis thaliana'*. *Journal of Plant Physiology*, 168(12), 1415–1425. https://doi.org/10.1016/j.jplph.2010.09.008

Floris, D., & Kühlbrandt, W. (2021). 'Molecular landscape of etioplast inner membranes in higher plants'. *Nature Plants,* 7(4), Article 4. https://doi.org/10.1038/s41477-021-00896-z

Galvão, R. M., Li, M., Kothadia, S. M., Haskel, J. D., Decker, P. V., Buskirk, E. K. V., & Chen, M. (2012). 'Photoactivated phytochromes interact with HEMERA and promote its accumulation to establish photomorphogenesis in Arabidopsis'. *Genes & Development,* 26(16), 1851–1863. https://doi.org/10.1101/gad.193219.112

Gao, Z.-P., Yu, Q.-B., Zhao, T.-T., Ma, Q., Chen, G.-X., & Yang, Z.-N. (2011). 'A functional component of the Transcriptionally Active Chromosome Complex, Arabidopsis pTAC14, interacts with pTAC12/HEMERA and regulates plastid gene expression'. *Plant Physiology,* 157(4), 1733–1745. https://doi.org/10.1104/pp.111.184762

Gerschman, R., Gilbert, D. L., Nye, S. W., Dwyer, P., & Fenn, W. O. (1954). 'Oxygen poisoning and x-irradiation: A mechanism in common'. *Science*, 119(3097), 623–626.

Gilkerson, J., Perez-Ruiz, J. M., Chory, J., & Callis, J. (2012). 'The plastid-localized pfkB-type carbohydrate kinases FRUCTOKINASE-LIKE 1 and 2 are essential for growth and development of *Arabidopsis thaliana'. BMC Plant Biology,* 12(1), 102. https://doi.org/10.1186/1471-2229-12-102

Gingras, A.-C., Abe, K. T., & Raught, B. (2019). 'Getting to know the neighborhood: Using proximity-dependent biotinylation to characterize protein complexes and map organelles'. *Current Opinion in Chemical Biology,* 48, 44–54. https://doi.org/10.1016/j.cbpa.2018.10.017

Gray, B. H., Lee, L. H., & Wyman, J. F. (1985). 'An automated analysis for superoxide dismutase enzyme activity'. *Journal of Analytical Toxicology*, 9(1), 36-39. https://doi.org/10.1093/jat/9.1.36

Gutensohn, M., Fan, E., Frielingsdorf, S., Hanner, P., Hou, B., Hust, B., & Klösgen, R. B. (2006). 'Toc, Tic, Tat et al.: Structure and function of protein transport machineries in chloroplasts'. *Journal of Plant Physiology*, 163(3), 333–347. https://doi.org/10.1016/j.jplph.2005.11.009

Hanaoka, M., Kanamaru, K., Takahashi, H., & Tanaka, K. (2003). 'Molecular genetic analysis of chloroplast gene promoters dependent on SIG2, a nucleus-encoded sigma factor for the plastid-encoded RNA polymerase, in *Arabidopsis thaliana*'. *Nucleic Acids Research,* 31(24), 7090–7098. https://doi.org/10.1093/nar/gkg935

Hassidim, M., Yakir, E., Fradkin, D., Hilman, D., Kron, I., Keren, N., Harir, Y., Yerushalmi, S., & Green, R. M. (2007). 'Mutations in CHLOROPLAST RNA BINDING provide evidence for the involvement of the chloroplast in the regulation of the circadian clock in Arabidopsis'. *The Plant Journal*, 51(4), 551–562. https://doi.org/10.1111/j.1365-313X.2007.03160.x

Heck, A. J. R. (2008). 'Native mass spectrometry: A bridge between interactomics and structural biology'. *Nature Methods*, 5(11), Article 11. https://doi.org/10.1038/nmeth.1265

Hedtke, B., Börner, T., & Weihe, A. (1997). 'Mitochondrial and Chloroplast Phage-Type RNA Polymerases in Arabidopsis'. *Science,* 277(5327), 809–811. https://doi.org/10.1126/science.277.5327.809

Hirata, A., Klein, B. J., & Murakami, K. S. (2008). 'The X-ray crystal structure of RNA polymerase from Archaea'. *Nature,* 451(7180). https://doi.org/10.1038/nature06530

Jarvis, P., & Robinson, C. (2004). 'Mechanisms of protein import and routing in chloroplasts'. *Current Biology:* CB, 14(24), R1064-1077. https://doi.org/10.1016/j.cub.2004.11.049

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes

B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. (2021). 'Highly accurate protein structure prediction with AlphaFold'. *Nature*. 596(7873):583-589. doi: 10.1038/s41586-021-03819-2.

Kandiah, E.; Giraud, T.; de Maria Antolinos, A.; Dobias, F.; Effantin, G.; Flot, D.; Hons, M.; Schoehn, G.; Susini, J.; Svensson, O.; et al. (2019). 'CM01: A Facility for Cryo-Electron Microscopy at the European Synchrotron'. *Acta Crystallographica Section D: Structural Biology,* 75, 528–535, doi:10.1107/S2059798319006880.

Kärkönen, A., & Kuchitsu, K. (2015). 'Reactive oxygen species in cell wall metabolism and development in plants'. *Phytochemistry,* 112, 22–32. https://doi.org/10.1016/j.phytochem.2014.09.016

Keegstra, K., & Cline, K. (1999). 'Protein import and routing systems of chloroplasts'. *The Plant Cell,* 11(4), 557–570. doi: 10.1105/tpc.11.4.557. PMID: 10213778; PMCID: PMC144212.

Kessler, F., & Schnell, D. J. (2006). 'The function and diversity of plastid protein import pathways: A multilane GTPase highway into plastids'. *Traffic,* 7(3), 248–257. https://doi.org/10.1111/j.1600-0854.2005.00382.x

Kimanius, D.; Forsberg, B.O.; Scheres, S.H.; Lindahl, E. (2016). 'Accelerated Cryo-EM Structure Determination with Parallelisation Using GPUs in RELION-2'. *eLife*, 5, doi:10.7554/eLife.18722.

Kindgren P & Strand A. (2015). 'Chloroplast transcription, untangling the Gordian Knot'. *New Phytologist*. 206(3), 889-891. https://doi.org/10.1111/nph.13388

Kindgren, P., Kremnev, D., Blanco, N. E., de Dios Barajas López, J., Fernández, A. P., Tellgren-Roth, C., Small, I., & Strand, Å. (2012). 'The plastid redox insensitive 2 mutant of Arabidopsis is impaired in PEP activity and high light-dependent plastid redox signalling to the nucleus'. *The Plant Journal,* 70(2), 279–291. https://doi.org/10.1111/j.1365-313X.2011.04865.x

Kleffmann, T., Russenberger, D., von Zychlinski, A., Christopher, W., Sjölander, K., Gruissem, W., & Baginsky, S. (2004). 'The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions'. *Current Biology:* CB, 14(5), 354–362. https://doi.org/10.1016/j.cub.2004.02.039

Kliebenstein, D. J., Monde, R.-A., & Last, R. L. (1998). 'Superoxide dismutase in Arabidopsis: An eclectic enzyme family with disparate regulation and protein localization'. *Plant Physiology,* 118(2), 637–650.

Kremnev, D., & Strand, Å. (2014). 'Plastid encoded RNA polymerase activity and expression of photosynthesis genes required for embryo and seed development in Arabidopsis'. *Frontiers in Plant Science, 5*. https://www.frontiersin.org/articles/10.3389/fpls.2014.00385

Lambert, J.-P., Tucholska, M., Go, C., Knight, J. D. R., & Gingras, A.-C. (2015). 'Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes'. *Journal of Proteomics*, 118, 81–94. https://doi.org/10.1016/j.jprot.2014.09.011

Leavitt, S., & Freire, E. (2001). 'Direct measurement of protein binding energetics by isothermal titration calorimetry'. *Current Opinion in Structural Biology*, 11(5), 560–566. https://doi.org/10.1016/S0959-440X(00)00248-7

Leister, D. (2014). 'Complex(iti)es of the ubiquitous RNA-binding CSP41 proteins'. *Frontiers in Plant Science,* 5. https://www.frontiersin.org/articles/10.3389/fpls.2014.00255

Liebers, M., Cozzi, C., Uecker, F., Chambon, L., Blanvillain, R., & Pfannschmidt, T. (2022). 'Biogenic signals from plastids and their role in chloroplast development'. *Journal of Experimental Botany,* 73, 7105–7125. https://doi.org/10.1093/jxb/erac344

Liebers, M., Gillet, F.-X., Israel, A., Pounot, K., Chambon, L., Chieb, M., Chevalier, F., Ruedas, R., Favier, A., & Gans, P. (2020). 'Nucleo-plastidic PAP 8/pTAC 6 couples chloroplast formation with photomorphogenesis'. *The EMBO Journal*, 39(22), e104941.

Liebers, M., Grübler, B., Chevalier, F., Lerbs-Mache, S., Merendino, L., Blanvillain, R., & Pfannschmidt, T. (2017). 'Regulatory Shifts in Plastid Transcription Play a Key Role in Morphological Conversions of Plastids during Plant Development'. *Frontiers in Plant Science,* 8. https://www.frontiersin.org/articles/10.3389/fpls.2017.00023

Ma, L., Li, J., Qu, L., Hager, J., Chen, Z., Zhao, H., & Deng, X. W. (2001). 'Light Control of Arabidopsis Development Entails Coordinated Regulation of Genome Expression and Cellular Pathways'. *The Plant Cell*, 13(12), 2589–2607. https://doi.org/10.1105/tpc.010229

Mair, A., Xu, S.-L., Branon, T. C., Ting, A. Y., & Bergmann, D. C. (2019). 'Proximity labelling of protein complexes and cell-type-specific organellar proteomes in Arabidopsis enabled by TurboID'. *eLife*, 8, e47864. https://doi.org/10.7554/eLife.47864

Marklund, S., & Marklund, G. (1974). 'Involvement of the superoxide anion radical in the autoxidation of pyrogallol and a convenient assay for superoxide dismutase'. *European Journal of Biochemistry,* 47(3), 469–474.

Martín, G., Leivar, P., Ludevid, D., Tepperman, J. M., Quail, P. H., & Monte, E. (2016). 'Phytochrome and retrograde signalling pathways converge to antagonistically regulate a light-induced transcriptional network'. *Nature Communications,* 7(1). https://doi.org/10.1038/ncomms11431

May, T., & Soll, J. (2000). '14-3-3 Proteins Form a Guidance Complex with Chloroplast Precursor Proteins in Plants'. *The Plant Cell,* 12(1), 53–63. https://doi.org/10.1105/tpc.12.1.53

McCord, J. M., & Fridovich, I. (1968). 'The reduction of cytochrome c by milk xanthine oxidase'. *Journal of Biological Chemistry*, 243(21), 5753–5760. PMID: 4972775.

Mehler, A. H. (1951). 'Studies on reactions of illuminated chloroplasts. II. Stimulation and inhibition of the reaction with molecular oxygen'. *Archives of Biochemistry and Biophysics,* 34(2), 339-351. https://doi.org/10.1016/0003-9861(51)90012-4

Miller, K. E., Kim, Y., Huh, W.-K., & Park, H.-O. (2015). 'Bimolecular Fluorescence Complementation (BiFC) Analysis: Advances and Recent Applications for Genome-Wide

Interaction Studies'. *Journal of Molecular Biology*, 427(11), 2039–2055. https://doi.org/10.1016/j.jmb.2015.03.005

Murakami, K. S. (2015). 'Structural Biology of Bacterial RNA Polymerase'. *Biomolecules,* 5(2), Article 2. https://doi.org/10.3390/biom5020848

Myouga, F., Hosoda, C., Umezawa, T., Iizumi, H., Kuromori, T., Motohashi, R., Shono, Y., Nagata, N., Ikeuchi, M., & Shinozaki, K. (2008). 'A Heterocomplex of Iron Superoxide Dismutases Defends Chloroplast Nucleoids against Oxidative Stress and Is Essential for Chloroplast Development in Arabidopsis'. *The Plant Cell,* 20(11), 3148–3162. https://doi.org/10.1105/tpc.108.061341

Neuhaus, H. E., & Emes, M. J. (2000). 'Nonphotosynthetic metabolism in plastids'. *Annual Review of Plant Biology,* 51(1), 111–140.

Oelmüller, R., Levitan, I., Bergfeld, R., Rajasekhar, V. K., & Mohr, H. (1986). 'Expression of nuclear genes as affected by treatments acting on the plastids'. *Planta,* 168(4), 482–492. https://doi.org/10.1007/BF00392267

Paul Emsley and Bernhard Lohkamp and William G. Scott and Kevin Cowtan. (2010). 'Features and Development of Coot'. *Acta Crystallographica Section D - Biological Crystallography,* 66, 486-501

Peltier, G., & Cournac, L. (2002). 'Chlororespiration'. *Annual Review of Plant Biology,* 53, 523–550. https://doi.org/10.1146/annurev.arplant.53.100301.135242

Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. (2021). 'UCSF ChimeraX: Structure visualization for researchers, educators, and developers'. *Protein Science,* 30(1), 70-82. https://doi.org/10.1002/pro.3943

Pfalz, J., & Pfannschmidt, T. (2013). 'Essential nucleoid proteins in early chloroplast development'. *Trends in Plant Science,* 18(4), 186-194. https://doi.org/10.1016/j.tplants.2012.11.003

Pfannschmidt, T., & Liere, K. (2005). 'Redox Regulation and Modification of Proteins Controlling Chloroplast Gene Expression'. *Antioxidants & Redox Signaling,* 7(5–6), 607–618. https://doi.org/10.1089/ars.2005.7.607

Pfannschmidt, T., & Link, G. (1994). 'Separation of two classes of plastid DNA-dependent RNA polymerases that are differentially expressed in mustard (*Sinapis alba L.)* seedlings'. *Plant Molecular Biology,* 25(1), 69–81. https://doi.org/10.1007/BF00024199

Pfannschmidt, T., Blanvillain, R., Merendino, L., Courtois, F., Chevalier, F., Liebers, M., Grübler, B., Hommel, E., & Lerbs-Mache, S. (2015). 'Plastid RNA polymerases: Orchestration of enzymes with different evolutionary origins controls chloroplast biogenesis during the plant life cycle'. *Journal of Experimental Botany*, 66(22), 6957–6973. https://doi.org/10.1093/jxb/erv415

Pilon, M., Ravet, K., & Tapken, W. (2011). 'The biogenesis and physiological function of chloroplast superoxide dismutases'. *Biochimica Et Biophysica Acta*, 1807(8), 989–998. https://doi.org/10.1016/j.bbabio.2010.11.002

Pogson, B. J., & Albrecht, V. (2011). 'Genetic dissection of chloroplast biogenesis and development: An overview'. *Plant Physiology,* 155(4), 1545–1551. https://doi.org/10.1104/pp.110.170365

Pogson, B. J., Ganguly, D., & Albrecht-Borth, V. (2015). 'Insights into chloroplast biogenesis and development'. *Biochimica et Biophysica Acta (BBA) - Bioenergetics,* 1847(9), 1017–1024. https://doi.org/10.1016/j.bbabio.2015.02.003

Pogson, B. J., Woo, N. S., Förster, B., & Small, I. D. (2008). 'Plastid signalling to the nucleus and beyond'. *Trends in Plant Science*, 13(11), 602–609. https://doi.org/10.1016/j.tplants.2008.08.008

Powikrowska, M., Oetke, S., Jensen, P. E., & Krupinska, K. (2014). 'Dynamic composition, shaping and organization of plastid nucleoids'. *Frontiers in Plant Science,* 5. https://www.frontiersin.org/articles/10.3389/fpls.2014.00424

Pyke, K. A., & Leech, R. M. (1992). 'Chloroplast Division and Expansion Is Radically Altered by Nuclear Mutations in *Arabidopsis thaliana'*. *Plant Physiology,* 99(3), 1005–1008. https://doi.org/10.1104/pp.99.3.1005

Qi, Y., Armbruster, U., Schmitz-Linneweber, C., Delannoy, E., de Longevialle, A. F., Rühle, T., Small, I., Jahns, P., & Leister, D. (2012). 'Arabidopsis CSP41 proteins form multimeric complexes that bind and stabilize distinct plastid transcripts'. *Journal of Experimental Botany,* 63(3), 1251–1270. https://doi.org/10.1093/jxb/err347

Qiu, Y., Li, M., Pasoreck, E. K., Long, L., Shi, Y., Galvão, R. M., Chou, C. L., Wang, H., Sun, A. Y., Zhang, Y. C., Jiang, A., & Chen, M. (2015). 'HEMERA Couples the Proteolysis and Transcriptional Activity of PHYTOCHROME INTERACTING FACTORs in Arabidopsis Photomorphogenesis'. *The Plant Cell,* 27(5), 1409–1427. https://doi.org/10.1105/tpc.114.136093

Schwarzenbacher, R., Godzik, A., Grzechnik, S. K., & Jaroszewski, L. (2004). The importance of alignment accuracy for molecular replacement. Acta Crystallographica Section D Biological Crystallography, 60(7), 1229-1236. https://doi.org/10.1107/s0907444904010145

Riggs, J. W., & Callis, J. (2017). 'Arabidopsis fructokinase-like protein associations are regulated by ATP'. *Biochemical Journal*, 474(11), 1789–1801. https://doi.org/10.1042/BCJ20161077

Ruedas Rémi, Soumiya Sankari Muthukumar, Sylvie Kieffer-Jaquinod, François-Xavier Gillet, Daphna Fenel, Grégory Effantin, Thomas Pfannschmidt, Yohann Couté, Robert Blanvillain, and David Cobessi. (2022). 'Three-Dimensional Envelope and Subunit Interactions of the Plastid-Encoded RNA Polymerase from *Sinapis alba'*. *International Journal of Molecular Sciences*. 23, (17): 9922. doi:10.3390/ijms23179922

Russo, C. J., & Passmore, L. A. (2016). 'Ultrastable gold substrates: Properties of a support for high-resolution electron cryomicroscopy of biological specimens'. *Journal of Structural Biology*, 193(1), 33–44. https://doi.org/10.1016/j.jsb.2015.11.006

Samavarchi-Tehrani, P., Samson, R., & Gingras, A.-C. (2020). 'Proximity dependent biotinylation: Key enzymes and adaptation to proteomics approaches'. *Molecular & Cellular Proteomics,* 19(5), 757-773. https://doi.org/10.1074/mcp.r120.001941

Saponaro A. (2018). 'Isothermal Titration Calorimetry: A Biophysical Method to Characterize the Interaction between Label-free Biomolecules in Solution'. *Bio Protocol*, 8(15), e2957. https://doi: 10.21769/BioProtoc.2957

Schleiff, E., Soll, J., Sveshnikova, N., Tien, R., Wright, S., Dabney-Smith, C., Subramanian, C., & Bruce, B. D. (2002). 'Structural and guanosine triphosphate/diphosphate requirements for transit peptide recognition by the cytosolic domain of the chloroplast outer envelope receptor, Toc34.' *Biochemistry*, 41(6), 1934-1946. https://doi.org/10.1021/bi011361+

Smith, M. D. (2006). 'Protein import into chloroplasts: An ever-evolving story'. *Canadian Journal of Botany*, 84(4), 531-542. https://doi.org/10.1139/b06-050

Solymosi, K., & Schoefs, B. (2010). 'Etioplast and etio-chloroplast formation under natural conditions: The dark side of chlorophyll biosynthesis in angiosperms'. *Photosynthesis Research*, 105(2), 143–166. https://doi.org/10.1007/s11120-010-9568-2

Steiner, S., Dietzel, L., Schröter, Y., Fey, V., Wagner, R., & Pfannschmidt, T. (2009). 'The role of phosphorylation in redox regulation of photosynthesis genes psaA and psbA during photosynthetic acclimation of mustard'. *Molecular Plant,* 2(3), 416-429. https://doi.org/10.1093/mp/ssp007

Steiner, S., Schröter, Y., Pfalz, J., & Pfannschmidt, T. (2011). 'Identification of Essential Subunits in the Plastid-Encoded RNA Polymerase Complex Reveals Building Blocks for Proper Plastid Development.' *Plant Physiology,* 157(3), 1043–1055. https://doi.org/10.1104/pp.111.184515

Sugiura, M. (1992). 'The chloroplast genome. In R. A. Schilperoort & L. Dure (Eds.), 10 Years Plant Molecular Biology (pp. 149–168)'. *Springer Netherlands*. https://doi.org/10.1007/978-94-011-2656-4_10

Suzuki, J. Y., Jimmy Ytterberg, A., Beardslee, T. A., Allison, L. A., van Wijk, K. J., & Maliga, P. (2004). 'Affinity purification of the tobacco plastid RNA polymerase and in vitro reconstitution of the holoenzyme'. *The Plant Journal,* 40(1), 164–172. https://doi.org/10.1111/j.1365-313X.2004.02195.x

Varadi, M et al. (2022). 'AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy model's. *Nucleic Acids Research*, 50(D1), D439-D444. https://doi.org/10.1093/nar/gkab1061

Wang, X., Jiang, B., Gu, L., Chen, Y., Mora, M., Zhu, M., Noory, E., Wang, Q., & Lin, C. (2021). 'A photoregulatory mechanism of the circadian clock in Arabidopsis'. *Nature Plants,* 7(10), Article 10. https://doi.org/10.1038/s41477-021-01002-z

Waters, M. T., & Langdale, J. A. (2009). 'The making of a chloroplast'. *The EMBO Journal,* 28(19), 2861–2873. https://doi.org/10.1038/emboj.2009.264

Waters, M. T., Fray, R. G., & Pyke, K. A. (2004). 'Stromule formation is dependent upon plastid size, plastid differentiation status and the density of plastids within the cell'. *The Plant Journal,* 39(4), 655–667. https://doi.org/10.1111/j.1365-313X.2004.02164.x

Webster, M. W., & Weixlbaumer, A. (2021). 'Macromolecular assemblies supporting transcription-translation coupling'. *Transcription*, 12(4), 103–125. https://doi.org/10.1080/21541264.2021.1981713

Webster, M. W., Takacs, M., Zhu, C., Vidmar, V., Eduljee, A., Abdelkareem, M., & Weixlbaumer, A. (2020). 'Structural basis of transcription-translation coupling and collision in bacteria'. *Science* (New York, N.Y.), 369(6509), 1355–1359. https://doi.org/10.1126/science.abb5036

Weston, E. L., & Pyke, K. A. (1999). 'Developmental Ultrastructure of Cells and Plastids in the Petals of Wallflower (*Erysimum cheiri*)'. *Annals of Botany,* 84(6), 763–769. https://doi.org/10.1006/anbo.1999.0981

Whatley, J. M. (1978). 'A suggested cycle of plastid developmental interrelationships'. *New Phytologist,* 80(3), 489-502. https://doi.org/10.1111/j.1469-8137.1978.tb01581.x

Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F., & Quandt, D. (2011). 'The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function'. *Plant Molecular Biology,* 76(3), 273–297. https://doi.org/10.1007/s11103-011-9762-4

Wimmelbacher, M., & Börnke, F. (2014). 'Redox activity of thioredoxin z and fructokinase-like protein 1 is dispensable for autotrophic growth of *Arabidopsis thaliana'*. *Journal of Experimental Botany,* 65(9), 2405–2413. https://doi.org/10.1093/jxb/eru122

Woodson, J. D., & Chory, J. (2008). 'Coordination of gene expression between organellar and nuclear genomes'. *Nature Reviews Genetics,* 9(5), Article 5. https://doi.org/10.1038/nrg2348

Xu, S.-L., Shrestha, R., Karunadasa, S. S., & Xie, P.-Q. (2023). 'Proximity labelling in plants'. *Annual Review of Plant Biology,* 74(1), 285–312. https://doi.org/10.1146/annurev-arplant-070522-052132

Yagi, Y., & 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(6), 290–294. https://doi.org/10.4161/trns.21810

Yagi, Y., & Shiina, T. (2014). 'Recent advances in the study of chloroplast gene expression and its evolution'. *Frontiers in Plant Science,* 5. https://www.frontiersin.org/articles/10.3389/fpls.2014.00061

Yu, Q.-B., Lu, Y., Ma, Q., Zhao, T.-T., Huang, C., Zhao, H.-F., Zhang, X.-L., Lv, R.-H., & 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'*. *Physiologia Plantarum,* 148(3), 408–421. https://doi.org/10.1111/j.1399-3054.2012.01718.x

Zhang, K. (2016). ‘Gctf: Real-Time CTF Determination and Correction’. *Journal of Structural Biology,* 193, 1–12, doi:10.1016/j.jsb.2015.11.003.

Zhang, Y., Song, G., Lal, N. K., Nagalakshmi, U., Li, Y., Zheng, W., Huang, P., Branon, T. C., Ting, A. Y., Walley, J. W., & Dinesh-Kumar, S. P. (2019). ‘TurboID-based proximity labelling reveals that UBR7 is a regulator of N NLR immune receptor-mediated immunity’. *Nature Communications,* 10, 3252. https://doi.org/10.1038/s41467-019-11202-z

Zhelyazkova, P., Sharma, C. M., Förstner, K. U., Liere, K., Vogel, J., & Börner, T. (2012). ‘The Primary Transcriptome of Barley Chloroplasts: Numerous Noncoding RNAs and the Dominating Role of the Plastid-Encoded RNA Polymerase’. *The Plant Cell,* 24(1), 123–136. https://doi.org/10.1105/tpc.111.089441

Zheng, S.Q.; Palovcak, E.; Armache, J.-P.; Verba, K.A.; Cheng, Y.; Agard, D.A. (2017). ‘MotionCor2: Anisotropic Correction of Beam-Induced Motion for Improved Cryo-Electron Microscopy’. *Nature Methods,* 14, 331–332, doi:10.1038/nmeth.4193.

Zivanov, J.; Nakane, T.; Forsberg, B.O.; Kimanius, D.; Hagen, W.J.; Lindahl, E.; Scheres, S.H. (2018). ‘New Tools for Automated High-Resolution Cryo-EM Structure Determination in RELION-3’. *eLife,* 7, doi:10.7554/eLife.42166.

Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q., & Wijk, K. J. van. (2008). ‘Sorting Signals, N-Terminal Modifications and Abundance of the Chloroplast Proteome’. *PLOS ONE,* 3(4), e1994. https://doi.org/10.1371/journal.pone.0001994

Zychlinski, A. von, Kleffmann, T., Krishnamurthy, N., Sjölander, K., Baginsky, S., & Gruissem, W. (2005). ‘Proteome Analysis of the Rice Etioplast: Metabolic and Regulatory Networks and Novel Protein Functions’. *Springer Molecular & Cellular Proteomics,* 4(8), 1072–1084. https://doi.org/10.1074/mcp.M500018-MCP200

8. SUPPLEMENTARY MATERIAL

Supplementary Material

Three-dimensional envelope and subunit interactions of the plastid-encoded RNA polymerase from *Sinapis alba*

Authors: Rémi Ruedas[1,†], Soumiya Sankari Muthukumar[1,2], Sylvie Kieffer-Jaquinod[3], François-Xavier Gillet[2,‡], Daphna Fenel[1], Grégory Effantin[1], Thomas Pfannschmidt[2,&], Yohann Couté[3], Robert Blanvillain[2*] and David Cobessi[1*]

**corresponding authors: robert.blanvillain@cea.fr, david.cobessi@ibs.fr*

Supplemental tables

Table S1: MS-based proteomic characterization of *S. alba* PEP fraction.

Table S2: characterization of proximal proteins in *S. alba* PEP fraction using crosslinking-MS.

Supplemental figures

Figure S1: abundance-based ranking of proteins quantified by MS in PEP enriched samples. Distribution of abundances represented as Log2 of normalized and summed iBAQ values of individual proteins detected in three independent PEP samples (Table S1). Identified subunits of individual complexes are color-coded (blue: PAPs; orange: α, β, β´ and β´´ subunits; green: histones). The annotated zoom-in shows the 24 most abundant proteins in the ranking.

Figure S2-5: sequence alignment of the α, β, β´ and β´´ subunits from PEP of angiosperms with those of the RNAPs from *E. coli*, *T. thermophilus* and Nostoc. S3) Sequence alignment of the α subunits, S2) sequence alignment of the β subunits S3) Sequence alignment of the N-terminal part from β´ subunit with β ´ subunit from PEP, S4) Sequence alignment of the C-terminal part from β´ subunit with the β´´ subunit from PEP. The residues conserved more than 50 % are in red, those mutated in similar residues are in blue. The strictly conserved residues described by Lane & Darst (Lane & Darst, 2010) are highlighted in gray. The blue triangles show mutations observed among the strictly conserved residues described (Lane & Darst, 2010). The non-conservative mutations, at least three in a row in the β or β´ domain in *E. coli* and *T. thermophilus*, are high-lighted in green and displayed on the *E. coli* structure (PDB entry: 6GH5). Those colored in orange are nearby to the DNA, those in green are located at the surface of the subunits. The domains described for all-RNA polymerase (a) and the bRNAPs (b) are also given and highlighted in yellow and cyan respectively. The name of the RNAP domains are also given and highlighted in purple and green (Lane & Darst, 2010; Sutherland & Murakami, 2018).

Figure S6: view of the catalytic core from the *E. coli* RNAP (PDB entry: 3LU0 (Opalka *et al*., 2010)) manually fitted into the envelope of PEP using Chimera (Pettersen *et al*., 2004).

Figure S7a and S7b: overall shape of the a) human RNA polymerase II (EMDB entry: EMD-2194; Kassube et al., 2013) and b) yeast RNA polymerase III (EMDB entry: EMD-1753; Vanini et al., 2010) solved at 25 and 21 Å respectively.

Figure S8: FSC curve for the PEP 3D reconstruction calculated between two independent half maps (gold standard FSC). The dotted line represents the FSC=0.143 cutoff used to determine the resolution.

Data source: rpos collection from the green lineage

Figure S1 : abundance-based ranking of proteins quantified by MS in PEP enriched samples. Distribution of abundances represented as Log2 of normalized and summed iBAQ values of individual proteins detected in three independent PEP samples (Table S1). Identified subunits of individual complexes are color-coded (blue: PAPs; orange: α, β, β´ and β´´ subunits; magenta: histones). The annotated zoom-in shows the 24 most abundant proteins in the ranking.

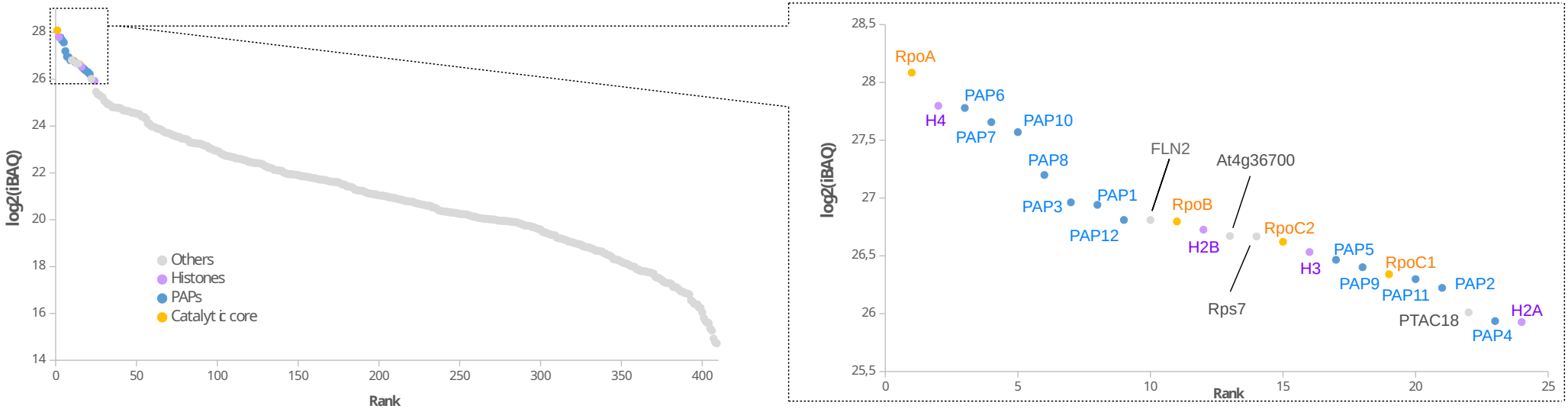
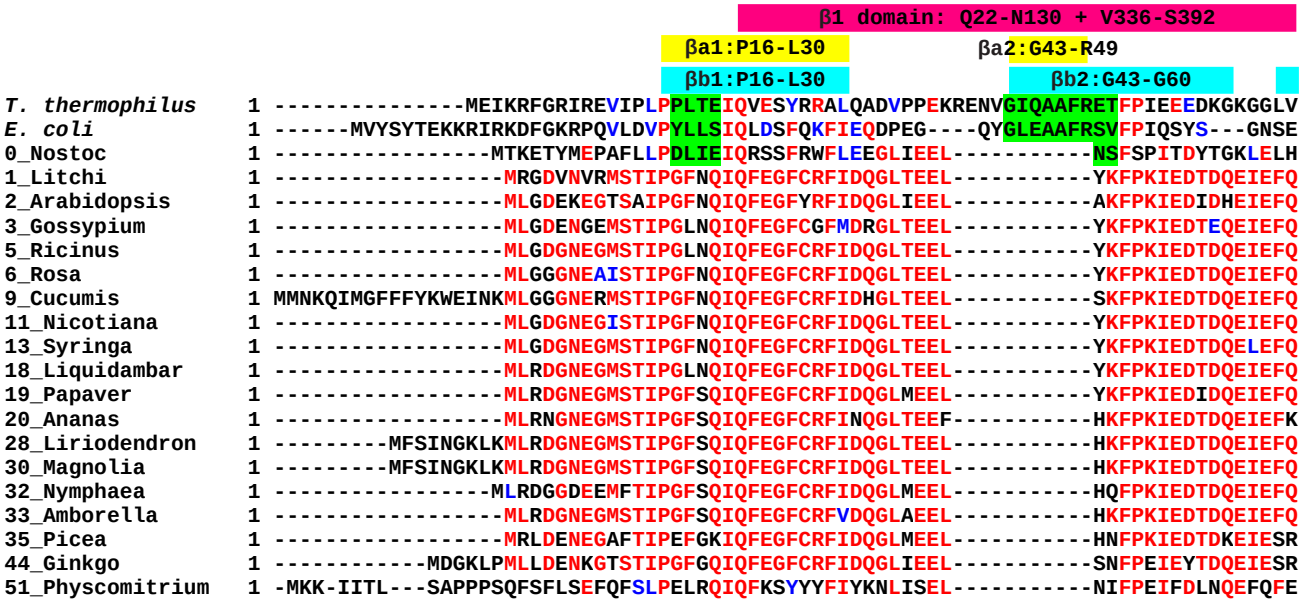


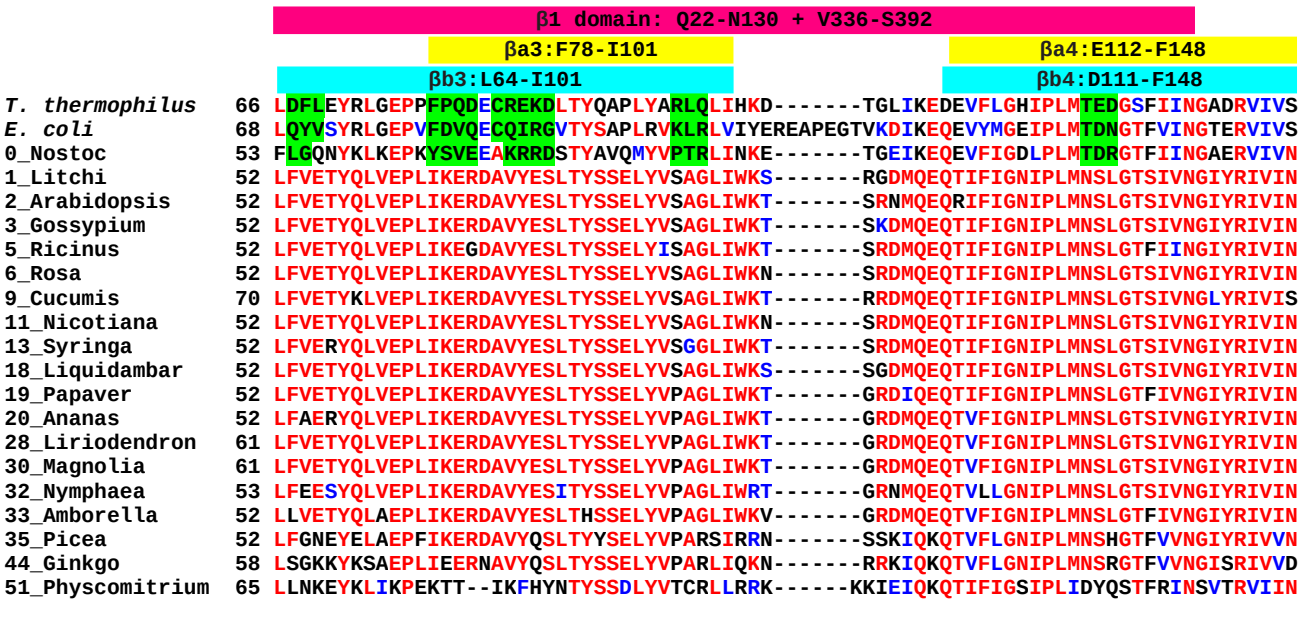
Figure S2: sequence alignment of the α subunits from PEP of angiosperms with those of the RNAPs from *E. coli*, *T. thermophilus* and Nostoc. The residues conserved more than 50 % are in red, those mutated in similar residues are in blue. The strictly conserved residues described by Lane & Darst (Lane & Darst, 2010) are highlighted in gray. The blue triangles show mutations observed among the strictly conserved residues described (Lane & Darst, 2010). The non-conservative mutations, at least three in a row in the β or β' domain in *E. coli* and *T. thermophilus*, are highlighted in green and displayed on the *E. coli* structure (PDB entry: 6GH5). Those colored in orange are nearby to the DNA, those in green are located at the surface of the subunits. The domains described for all-RNA polymerase (a) and the bRNAPs (b) are also given and highlighted in yellow and cyan respectively. The name of the RNAP domains are also given and highlighted in purple and green (Lane & Darst, 2010; Sutherland & Murakami, 2018).

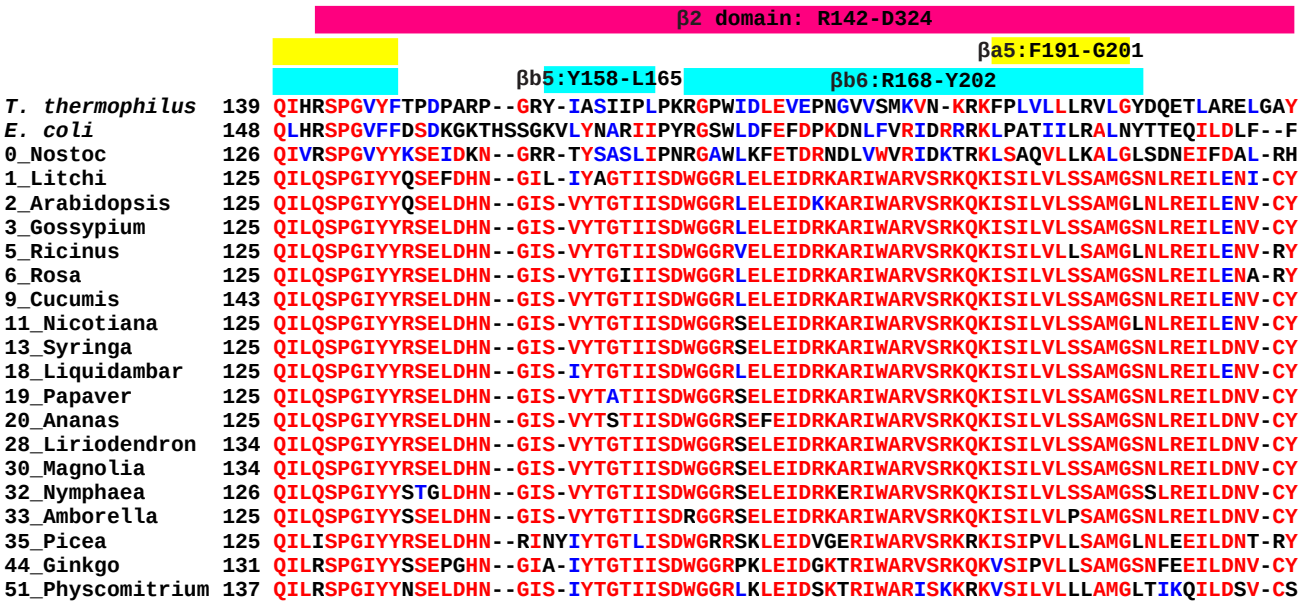
```
T. thermophilus     1 ------------------------------------------------------------------------------
E. coli             1 ------------------------------------------------------------------------------
0_Nostoc            1 ------------------------------------------------------------------------------
1_Litchi            1 ------------------------------------------------------------------------------
2_Arabidopsis       1 ------------------------------------------------------------------------------
3_Gossypium         1 ------------------------------------------------------------------------------
5_Ricinus           1 ------------------------------------------------------------------------------
6_Rosa              1 ------------------------------------------------------------------------------
9_Cucumis           1 ------------------------------------------------------------------------------
11_Nicotiana        1 ------------------------------------------------------------------------------
13_Syringa          1 ------------------------------------------------------------------------------
18_Liquidambar      1 ------------------------------------------------------------------------------
19_Papaver          1 ------------------------------------------------------------------------------
20_Ananas           1 ------------------------------------------------------------------------------
28_Liriodendron     1 ------------------------------------------------------------------------------
30_Magnolia         1 ------------------------------------------------------------------------------
32_Nymphaea         1 ------------------------------------------------------------------------------
33_Amborella        1 ------------------------------------------------------------------------------
35_Picea            1 ------------------------------------------------------------------------------
44_Ginkgo           1 ------------------------------------------------------------------------------
51_Physcomitrium    1 MAAVMSAQAMADAVRVVRASISELGKSAKKSDAFSARDLKSRSGFTESVDRLGLVNLQVQARKKRVRGLRLVPTRVVSAI

                                                                                  α-NTD
T. thermophilus     1 ----------------------------MLDSK----------LKAPVFTVRTQGREYGEFVLEPLERGFGVTLGNP
E. coli             1 -------------------------MQGSVTEF----------LKPRLVDIEQVSSTHAKVTLEPLERGFGHTLGNA
0_Nostoc            1 ------------------------------------------MAQFQIECVESNTEESRNHYSKFVLEPLERGQGTTVGNA
1_Litchi            1 ----------------------------MVREKVK--VFTRTLQWKCVESRTESKRLYYGRFILSPLMKGQADTIGIA
2_Arabidopsis       1 ----------------------------MVREKVK--VSTRTLQWKCVESKRDSKRLYYGRFILSPLMKGQADTIGIA
3_Gossypium         1 ----------------------------MVREKVKVSTSTRTRQWKCVESRTDSKRLYYGRFILSPLMKGQADTIGIA
5_Ricinus           1 ----------------------------MIREKVT--ISTRTLQWKCVESRTDNKRLFYGRFILSPLMKGQADTIGIA
6_Rosa              1 ----------------------------MVREKVT--VSTRTLQWKCVESRADSKRLYYGRFILAPLMKGQADTIGIA
9_Cucumis           1 ----------------------MILSKNITMVREKIR--VSTRTLKWKCVESRADSKRLYYGRFILSPLMKGQGDTIGIA
11_Nicotiana        1 ----------------------------MVREKVT--VSTRTLQWKCVESRTDSKRLYYGRFILSPLMKGQADTIGIA
13_Syringa          1 ----------------------------MVREKVT--VSTRTLQWKCVESREDSKRLYYGRFILSPLMKGQADTIGIA
18_Liquidambar      1 ----------------------------MVREKVT--VSTRTLQWKCVESRADSKRLYYGRFILSPLMKGQADTIGIA
19_Papaver          1 ----------------------------MLREEVA--VSTRTLQWKCVESRADSKRLYYGRFILSPLLKGQADTIGIA
20_Ananas           1 ----------------------------MVREEVA--GYTRTLQWKCVESRVDSKRLYYGRFILSPLMKGQADTIGIA
28_Liriodendron     1 ----------------------------MVREEVA--VSTRTLQWKCVESRTDSKRLYYGRFILAPLMKGQADTIGIA
30_Magnolia         1 ----------------------------MVREEVA--VSTRTLQWKCVESRTDSKRLYYGRFILAPLMKGQADTIGIA
32_Nymphaea         1 ----------------------------MVREEVP--VSTRTLQWKCVESRADSKRLYYGRFVLSPLMKGQADTIGIA
33_Amborella        1 ----------------------------MVRKEVP--VSNRTLQWKCVESKADSKRLYYGRFVLSPLMKGQADTIGTA
35_Picea            1 ----------------------------MIRDEIS--VSIQTLRWKCIESRAYSERLHYGRFALSPLRKGRADTIGIA
44_Ginkgo           1 ----------------------------MIRDEIS--VSIQTLRWKCIESRVNGKRLHYGRFALSPLQKGQANTIGIA
51_Physcomitrium   81 EGSNSTTADAPVDEDVLAWTKAYRAENSTAITRDETLKSNAQSALQWKCVETQVEGERLHYGRFAVSPFRSGQANTVGVS

                                  α-NTD
T. thermophilus    40 LRRILLSSIPGTAVTSVYIEDVLHEFSTIPGVKEDVVEIILNLKELVVRFLNPSLQTVTLLLKAEGPKEVKARDFLPVAD
E. coli            43 LRRILLSSMPGCAVTEVEIDGVLHEYSTKEGVQEDILEILLNLKGLAVRVQGKDEV--ILTLNKSGIGPVTAADITHDGD
0_Nostoc           40 LRRVLLSNLEGTAVTAVRIAGVSHEFATVPGVREDVLELILMRMKEVILKNYSSQPQ--IGRLLVNGPATITAAHFDLPSE
1_Litchi           49 MRRALLGEIEGTCITRAKFEKIPHECSTILGIQESVHEILMNLKEIVLRSNLYGTR--DALLCVKGPGYVTAQDILLPPS
2_Arabidopsis      49 MRRALLGEIEGTCITRAKSENIPHDYSNIAGIQESVHEILMNLNEIVLRSNLYGTR--NALICVQGPGYITARDIILPPA
3_Gossypium        51 MRRALLGELEGTCITRAKSEKIPHEYSTIVGIQESVHEILMNLKEIVLRGNLYGTR--NAFICAKGPGYVTAQDIILPPS
5_Ricinus          49 TRRALLGEIEGTCITRAKSEKIPHEFSTIAGIQESIHEILMNLKEIVLRSNLYGTC--DASICVKGPGYITAQDIILPPF
6_Rosa             49 MRRALLGEIEGTCITRARSEKIPHEYSTIVGIQESVHEILMNLKEIVLRSNLYGTR--NASICVKGPGYVTAQDIILPPS
9_Cucumis          57 MRKALLGEIEGTCITRAKSEKIPHEYSTIVGIQESVHEILMNLKEIVLRSNLYGTR--DASICVKGPGCVTAQDIILPPS
11_Nicotiana       49 MRRALLGEIEGTCITRVKSEKVPHEYSTITGIQESVHEILMNLKEIILRSNLYGTS--DASICVKGPGSVTAQDIILPPY
13_Syringa         49 MRRTLLGEIEGTCITRVKSENVPHEYSTIAGIQESVHEILMNLKEIVLRSNLYGTW--DASICVRGPGYVTAQDIILPPY
18_Liquidambar     49 MRRALLGEIEGTCITRAKSEKIPHEYSTIVGIQESVHEILMNLKEVVLRSNLYGTR--DASICVRGPGYVTAQDIISPPS
19_Papaver         49 MRRALLGEIEGTCITRAKSDKIPHEYSTIVGIEESVHEILMNLKEIVLRSNLYGTR--DASICVRGPGYVTAQDIISPPS
20_Ananas          49 MRRALLGEIEGTCITRAKSEKVPHEYSTIVGIEESVHEILLNLKEIVLRSNLYGVR--DALICVRGPRYVTAQDIISPPS
28_Liriodendron    49 MRRALLGEIEGTCITRVKSEKVPNEYYTIVGIEESVHEILMNLKEIVLRSHLYGTR--DASICVRGPRYVTAQDIIPPPS
30_Magnolia        49 MRRALLGEIEGTCITRAKSEKVPHEYSTIVGIEESVHAILMNLKEIVLGSHLYGTR--DASICVRGPRYVTAQDIIPPPS
32_Nymphaea        49 MRRALLGELEGTCITRAKSDKVPHEYSTIVGIEESVHEILMNLKKIVLRSDLYGTL--NASICVRGPRHVTAQDIISPPS
33_Amborella       49 MRRSLLGEIEGTCITCAKSERVPHEYSTIVGIEESVHEILMNLKEIIVRSDLYGTR--DAFICVRGPKYVTAQDIISPPS
35_Picea           49 MRRVLLGEVEGTCITHVKLENIKHEYSAIIGIEESVHDILMNLKEIVLRSDSYGIR--GASICIVGPRNVTAQDIILPPS
44_Ginkgo          49 MRRALLGEVEGTCITHAKFENMTHEYSAIMGIEESVHDISINLRGIVLQSDPYGIR--EASIYSVGPRDVTAQDIILPPS
51_Physcomitrium  161 MQKALLGEVEGAAVSCATFKNVKSEYAAMKGVEETPMDILVNLKELVIRSDSDEPQ--KAIISAIGPGPVTAGDIVLPPS
```

```
                                 α-NTD
T. thermophilus   120 VEIMNPDLHIATL-EEGGRLNMEVRVDRGVGYVPAEKHGI----KDRINAIPVDAVFSPVRRVAFQVEDTRLGQRTDLDK
E. coli           121 VEIVKPQHVICHLTDENASISMRIKVQRGRGYVPASTRIHSEEDERPIGRLLVDACYSPVERIAYNVEAARVEQRTDLDK
0_Nostoc          118 VEVIDPTQYVATI-AEGGKLEMEFRIERGKGYRTVERGRE---EATSLDFLQIDSVFMPVRKVNYSVEEARGDGSITKDR
1_Litchi          127 VEIIDNTQHIASL-AEPIDFCIGLQIERNRGYNIKAPNNF---QD---RSYSIDAVFMPVRNANYSIHSYGSGN-EKQEI
2_Arabidopsis     127 VEIIDNTQHIATL-TEPIDLCIELKIERNRGYSLKMSNNF---ED---RSYPIDAVFMPVENANHSIHSYGNGN-EKQEI
3_Gossypium       129 VEIVDNTQHVASL-TEPIDLCIGLQIERNRGYGIKTPKNF---HD---GSYPIDAVFMPVRNANHSIHCYGNDN-EKQEI
5_Ricinus         127 VEIIDNTQHIASL-TEPIDLCIGLQIERNRGYRIKPTNNF---HE---GSYPIDAVFMPVRNANHSVHSYGNGN-EKQEI
6_Rosa            127 VEIVDNTQHIANL-TEPINLCIQLQIERNRGYRIKTPNNF---QD---GSYPIDAVFMPVRNANHSIHSYVNGN-EKQEI
9_Cucumis         135 VEIVDNTQHIANL-TEPINFCIELKIERNRGYHIQTPNNF---QD---ASYPMDAIFMPVRNVNHSIHSYVNGN-EKQEI
11_Nicotiana      127 VEIVDNTQHIASL-TEPIDFCIGLQIERNRGYLIKTPHNF---QD---GSYPIDAVFMPVRNANHSIHSYGNGN-EKQEI
13_Syringa        127 VEIVDNTQHIASL-TEPIDLCIGLQIERNRGYLIKMPDNF---QD---GSYPIDAVFMPVRNANHSIHSYGNGN-EKQEI
18_Liquidambar    127 VEIVDNTQHIASL-TEPIDLCIGLKMERNRGYRIKTPNNF---QD---RSYPIDAVFMPVRNANHSIHSYGNEN-EKQEI
19_Papaver        127 VEIVDTTQHIASL-TEPMDLCIGLQIERNRGYRMKTPNNA---QD---GSFTIDAVFMPVRNANHSIHSYGNGN-EKQEI
20_Ananas         127 VEIVDTTQHIANL-TEPIDLCIGLQIKRDRGYRMEPINDS---QD---VSYPIDAVSMPVRNANHSIHSYGNGN-EKQEI
28_Liriodendron   127 VEIVDTTQHIASL-TEPIDLCIELQIERDRGYRMKTPNNY---QD---GSYPIDAVSMPVRNANHSIHSYGNRN-EKQEI
30_Magnolia       127 VEIVDTTQHIASL-TEPIDLCIELQIERDRGYRMKTPNNY---QD---GSYPIDAVSMPVRNANHSIHSYGNGN-EKQEI
32_Nymphaea       127 VEIVDTTQHIAVL-TEPVDLCIELQIERGRGYCTRTPNNY---QD---GSYPIDAVSMPVRNANHSIHSYGNGN-EKQEI
33_Amborella      127 VKIVDTTQHIASI-TEPINFCIGLEIERNRGYRTRTTN-----QD---ASYPIDAVSMPVRNANYSIQCYGNEN-EKQEI
35_Picea          127 VKIIDTTQHIARL-TKSITSDIRLQIEKNRGYIIHSPNNY---QD---GIFPIDAVFMPVRDANYSIHSYGSGN-EIREV
44_Ginkgo         127 VRIIDTTQHIASL-KESITSDIRSQIGKDRGYRIQSPNNS---RD---GIFSTDAVSMPVRNANYSIHSYGNGN-EMQEI
51_Physcomitrium  239 LEVTDPTQHIAYL-TKEVSLDIELDVEKGCGYRMGDHTKS---GD---GRFYIDSVFMPVRNANYSVHSYESEPDVTQEI

                                 α-NTD                                                  α-CTD
T. thermophilus   195 LTLRIWTDGSVTPLEALNQAVEILREHLTYFSNPQAAAVAAPEEAKE-----------------PEAPPEQEEELDLPL
E. coli           201 LVIEMETNGTIDPEEAIRRAATILAEQLEAFVDLRDVRQ--PE---V-----------------KEEKPEFDPILLRPV
0_Nostoc          194 LLLEVWTNGSISPQEALSSAAGILVDLFNPLKDISLEPTD-------T-----------------GSEISEDPTAQIPI
1_Litchi          199 LFLEIWTNGGLTPKEALHEASRNLIDLFIPFLQAAE--ENLHLENNQYKVSLPYLTFRDRVAKLKKKK--KEIAGKSIFI
2_Arabidopsis     199 LFLEIWTNGSLTPKEALHQASRNLINLFIPFLHVEE--ETFYLENNQHQVTLPFFPFHNRLVNLRKKKKTKELAFQYIFI
3_Gossypium       201 LFLEIWTNGSLTPKEALHEASRNLIDLFIPFLHTEE--ENLHLENNQHDVTLPFFPFHDRLVKLTKKK--KEIALKYIFI
5_Ricinus         199 LFLEIWTNGSLTPKEALHEASRNLIDLFIPFLHAEE--ENFHLEKNQHKVTLPLFTFYDRLTKLRKNQ--NEITLKYVFI
6_Rosa            199 LFIEIWTNGSLTPKEALHEASRNLIDLFIPFLHAEE--ENFHFENNQHKVTLPLFTFHDKFANPRKSK--TEITLKYIFI
9_Cucumis         207 LFLEIWTNGSLTPKEALHEASRNLIDLFIPFLHAEEEKENFHFKNNKQKVTLPLFTFHEKLAKLRKKK--KERALKYIFI
11_Nicotiana      199 LFLEIWTNGSLTPKEALHEASRNLIDLFIPFLHMEE--DNLYLQDNQHTVPLSPFTFHDKLAKLIKNK--KKIALKSIFI
13_Syringa        199 LFLEIWTNGSLTPKEALHEASRNLIDLFIPFLHKEE--ENLPLEDNQHTVPLSPLTFHEKLDKLRKNK--KKIALKSIFI
18_Liquidambar    199 LFLEIWTNGSLTPKEALHEASRNLIDLFIPFLHAEE--ENLHLEDNQHKTILPLFTFHDRLAKLKKNK--KKKALKSIFI
19_Papaver        199 LFLEIWTNGSLTPKEALHEASRNLIDLFIPFLHAEERDTN--LEDNQNIVTIPFFTFQNKLDKL--SKSKKDIALKCIFI
20_Ananas         199 LFLEIWTNGSLTPKEALYEASRNLIDLFIPFLHAEEEDIN--FEENKNRFTLPFLTFQDRLTNL--KKNKKGIPLNCIFI
28_Liriodendron   199 LFLEIWTNGSLTPKEALHEASRNLIDLFIPFLHAEEQDIV--LEDNPNRFTVPLFTFHDRLANI--RKNKKGIALKCIFI
30_Magnolia       199 LFLEIWTNGSLTPKEALHEASRNLIDLFIPFLHAEEQDIN--LEDNPNRFTVPFFTFHDRLANI--RKNKKGIALKCIFI
32_Nymphaea       199 LFLEIWTNGSLTPKEALYEASRNLIDLFIPFLHAEEQDINRNMEDNLKRASVPFFTFDDGLDNM--K---REIILKRIFI
33_Amborella      197 LFLEIWTNGSLTPKEALYEASRNLIDLFIPFLHAEERDLH--LDDDQNIFTAPFFPFHEGLAYIGIGESKKRIALKRIFI
35_Picea          199 LFLEIWTNGGLTPREALSEASRILIELFIPFLHGEEQNI--DGMNNRKGSNMPPFPLSHVLTDTGET-KEKKIAFQHIFI
44_Ginkgo         199 LFSEIWTNGSLAPGEALYEASRNLIDLFIPFMRAEEQNI--DGIDNWNGYNMSLFPFSHISTDTEK--MEEEIAFKHIFI
51_Physcomitrium  312 LFLEIWTNGSITPEEALHEAARCLIDLFLPFLHPKKKEVTNSATKMHKSFTMSQFNSS-------AEMSAKEVDLRHVYV

                                 α-CTD
T. thermophilus   257 EELGLSTRVLHSLKEEGIESVRALLALNLKDLKNIPGIGERSLEEIKEALEKKGFTLKE----------------
E. coli           258 DDLELTVRSANCLKAEAIHYIGDLVQRTEVELLKTPNLGKKSLTEIKDVLASRGLSLGMRLENW---PPASIADE
0_Nostoc          249 EELQLSVRAYNCLKRAQVNSVADLLDYTQEDLLEIKNFGQKSAEEVVEALQRRL-GITLPQERGSKHP-------
1_Litchi          275 DQPELSPRIYNCLKKSNIHTLFDLLNKSQDDLMKIEHFRIEDVKQILGILEKK----------------------
2_Arabidopsis     277 DQLELPPRIYNCLKKSNIHTLLDLLNNSQEDLIKIEHFHVEDVKKILDILEKK----------------------
3_Gossypium       277 DQSELPPRIYNCLKKSNIHTLLDLLNNSREDLMKIEHFRIEDVKQILGILEKK----------------------
5_Ricinus         275 DQSELTPKIYNCLKRSNIHTLSDLLNKSQEDLMKIEHFRIDDVKHILGILEIEK-NFAI----------------
6_Rosa            275 DQSELPPRVYNCLKRSNIHTLLDLLNNSQEDLIKMKHLRIEDVKHILNILEKKK-HFE-----------------
9_Cucumis         285 DQSELPPRIYNCLKRCNIHTLFDLLNNSPDELMKIKHFRIEDVKHILDILEMEK-NFA-----------------
11_Nicotiana      275 DQSELPSRIYNCLKMSNIYTLLDLLNNSQEDLMKIEHFRSEDVKRILGILEKYF-VIDLAKNKF-----------
13_Syringa        275 DQSEFSPRVYNCLKRSNIYTLLDLLNNSQEDLMKMEHLRLEDIKQILVILEKHF-AIDLPKKNF-----------
18_Liquidambar    275 DQSELPPRIYTCLKRSNIHTLLDLLNNNQEDLMKIEHFRIEDVKKILGILEKHF-AIDLPKNKF-----------
19_Papaver        275 DQSELSPRIYNCLKRSNIYTLLDLLNKSQEDLMKIDHFRIEDVKQILDILQKRF-AIDLPKNKFSFLNPLE----
20_Ananas         275 DQLELPSRTYNCLKRSNIHTLLDLLSNSQEDLMRIEYFRMEDVKQILDTLQKHF-AIDLPKNKFSF---------
28_Liriodendron   275 DQSELPPRTYNCLKRSNIHTLLDLLSNSQEDLMRIEHFRIEDVKQILDILQKHF-TIDLPKNKF-----------
30_Magnolia       275 DQSELPPRTYNCLKRSNIHTLLDLLSNSQEDLMRIEHFRIEDVKQILDILQKHF-TIDLPKNKF-----------
32_Nymphaea       274 DQLELPPRTYNCLKRSNIHTLLDLLSKSQEDLMRIEHFRVEDVKQIFDILQKGF-TIDLLKNFK-NSNQFESR--
33_Amborella      275 DQLELSPRTYNCLKRSNIHTLLDLLSKSQEDLRKIQHFRVEDVKRVLDILKKRF-AMNLFT--------------
35_Picea          276 DQLEFPPRVYNCLRRANIHTLSDLLNYSREDLMRIERFGKESVEQVFEILRKRF-AIDPPRN-------------
44_Ginkgo         275 DQSELPPRVYNCLKRVNIHTLSDLLNYSRDDPTRIENFGNQSVEQILEVLQKRF-AIDLPKKKF-----------
51_Physcomitrium  385 DQLRIPSKAYNSLKRANINTVSDLLDYTQDDLLSIPNFGRKSVDDILEALQAQF-SIDLPENNPLCN--------
```

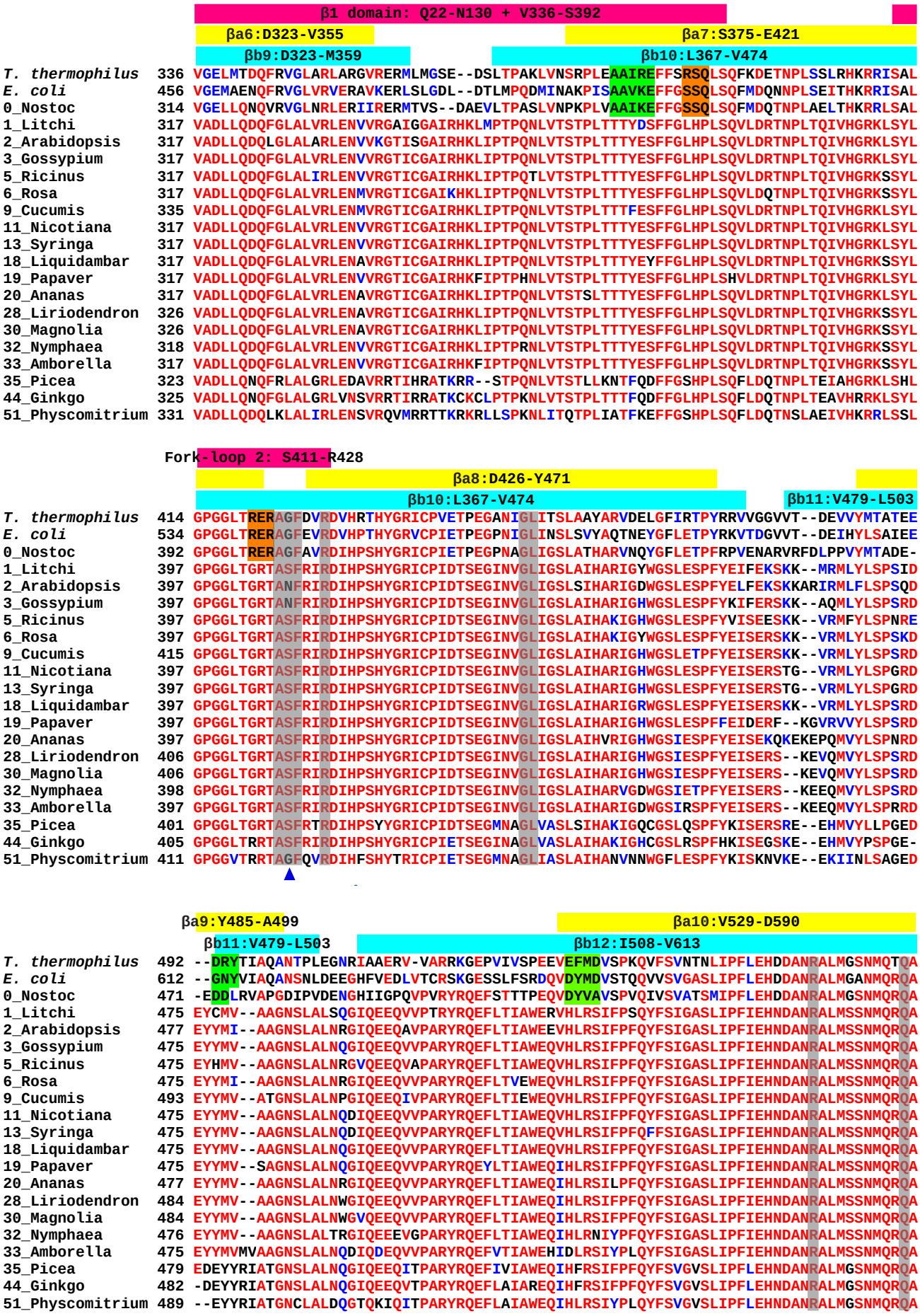
Figure S3: sequence alignment of the β subunits from PEP of angiosperms with those of the RNAPs from *E. coli*, *T. thermophilus* and Nostoc. The residues conserved more than 50 % are in red, those mutated in similar residues are in blue. The strictly conserved residues described by Lane & Darst (Lane & Darst, 2010) are highlighted in gray. The blue triangles show mutations observed among the strictly conserved residues described (Lane & Darst, 2010). The non-conservative mutations, at least three in a row in the β or β' domain in *E. coli* and *T. thermophilus*, are highlighted in green and displayed on the *E. coli* structure (PDB entry: 6GH5). Those colored in orange are nearby to the DNA, those in green are located at the surface of the subunits. The domains described for all-RNA polymerase (a) and the bRNAPs (b) are also given and highlighted in yellow and cyan respectively. The name of the RNAP domains are also given and highlighted in purple and green (Lane & Darst, 2010; Sutherland & Murakami, 2018).

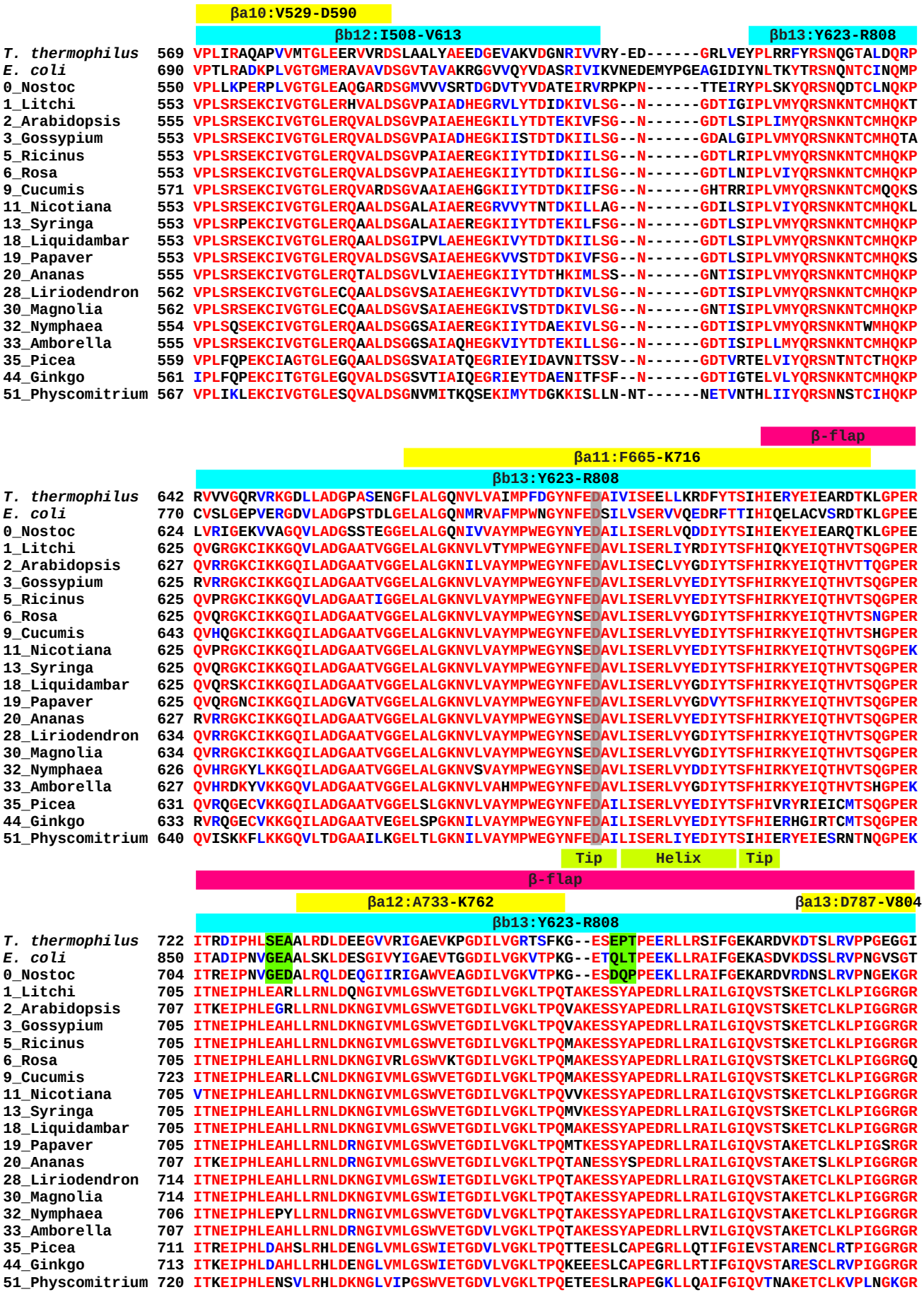

β1 domain: Q22-N130 + V336-S392
βa1:P16-L30
βa2:G43-R49
βb1:P16-L30
βb2:G43-G60
T. thermophilus 1 --------------MEIKRFGRIREVIPLPPLTEIQVESYRRALQADVPPEKRENVGIQAAFRETFPIEEEDKGKGGLV
E. coli 1 ------MVYSYTEKKRIRKDFGKRPQVLDVPYLLSIQLDSFQKFIEQDPEG----QYGLEAAFRSVFPIQSYS---GNSE
0_Nostoc 1 ----------------MTKETYMEPAFLLPDLIEIQRSSFRWFLEEGLIEEL-----------NSFSPITDYTGKLELH
1_Litchi 1 ----------------MRGDVNVRMSTIPGFNQIQFEGFCRFIDQGLTEEL-----------YKFPKIEDTDQEIEFQ
2_Arabidopsis 1 ----------------MLGDEKEGTSAIPGFNQIQFEGFYRFIDQGLIEEL-----------AKFPKIEDIDHEIEFQ
3_Gossypium 1 ----------------MLGDENGEMSTIPGLNQIQFEGFCGFMDRGLTEEL-----------YKFPKIEDTEQEIEFQ
5_Ricinus 1 ----------------MLGDGNEGMSTIPGLNQIQFEGFCRFIDQGLTEEL-----------YKFPKIEDTDQEIEFQ
6_Rosa 1 ----------------MLGGGNEAISTIPGFNQIQFEGFCRFIDQGLTEEL-----------YKFPKIEDTDQEIEFQ
9_Cucumis 1 MMNKQIMGFFFYKWEINKMLGGGNERMSTIPGFNQIQFEGFCRFIDHGLTEEL-----------SKFPKIEDTDQEIEFQ
11_Nicotiana 1 ----------------MLGDGNEGISTIPGFNQIQFEGFCRFIDQGLTEEL-----------YKFPKIEDTDQEIEFQ
13_Syringa 1 ----------------MLGDGNEGMSTIPGFNQIQFEGFCRFIDQGLTEEL-----------YKFPKIEDTDQELEFQ
18_Liquidambar 1 ----------------MLRDGNEGMSTIPGLNQIQFEGFCRFIDQGLTEEL-----------YKFPKIEDTDQEIEFQ
19_Papaver 1 ----------------MLRDGNEGMSTIPGFSQIQFEGFCRFIDQGLMEEL-----------YKFPKIEDIDQEIEFQ
20_Ananas 1 ----------------MLRNGNEGMSTIPGFSQIQFEGFCRFINQGLTEEF-----------HKFPKIEDTDQEIEFK
28_Liriodendron 1 ---------MFSINGKLKMLRDGNEGMSTIPGFSQIQFEGFCRFIDQGLTEEL-----------HKFPKIEDTDQEIEFQ
30_Magnolia 1 ---------MFSINGKLKMLRDGNEGMSTIPGFSQIQFEGFCRFIDQGLTEEL-----------HKFPKIEDTDQEIEFQ
32_Nymphaea 1 ----------------MLRDGGDEEMFTIPGFSQIQFEGFCRFIDQGLMEEL-----------HQFPKIEDTDQEIEFQ
33_Amborella 1 ----------------MLRDGNEGMSTIPGFSQIQFEGFCRFVDQGLAEEL-----------HKFPKIEDTDQEIEFQ
35_Picea 1 ----------------MRLDENEGAFTIPEFGKIQFEGFCRFIDQGLMEEL-----------HNFPKIEDTDKEIESR
44_Ginkgo 1 -----------MDGKLPMLLDENKGTSTIPGFGQIQFEGFCRFIDQGLIEEL-----------SNFPEIEYTDQEIESR
51_Physcomitrium 1 -MKK-IITL---SAPPPSQFSFLSEFQFSLPELRQIQFKSYYYFIYKNLISEL-----------NIFPEIFDLNQEFQFE

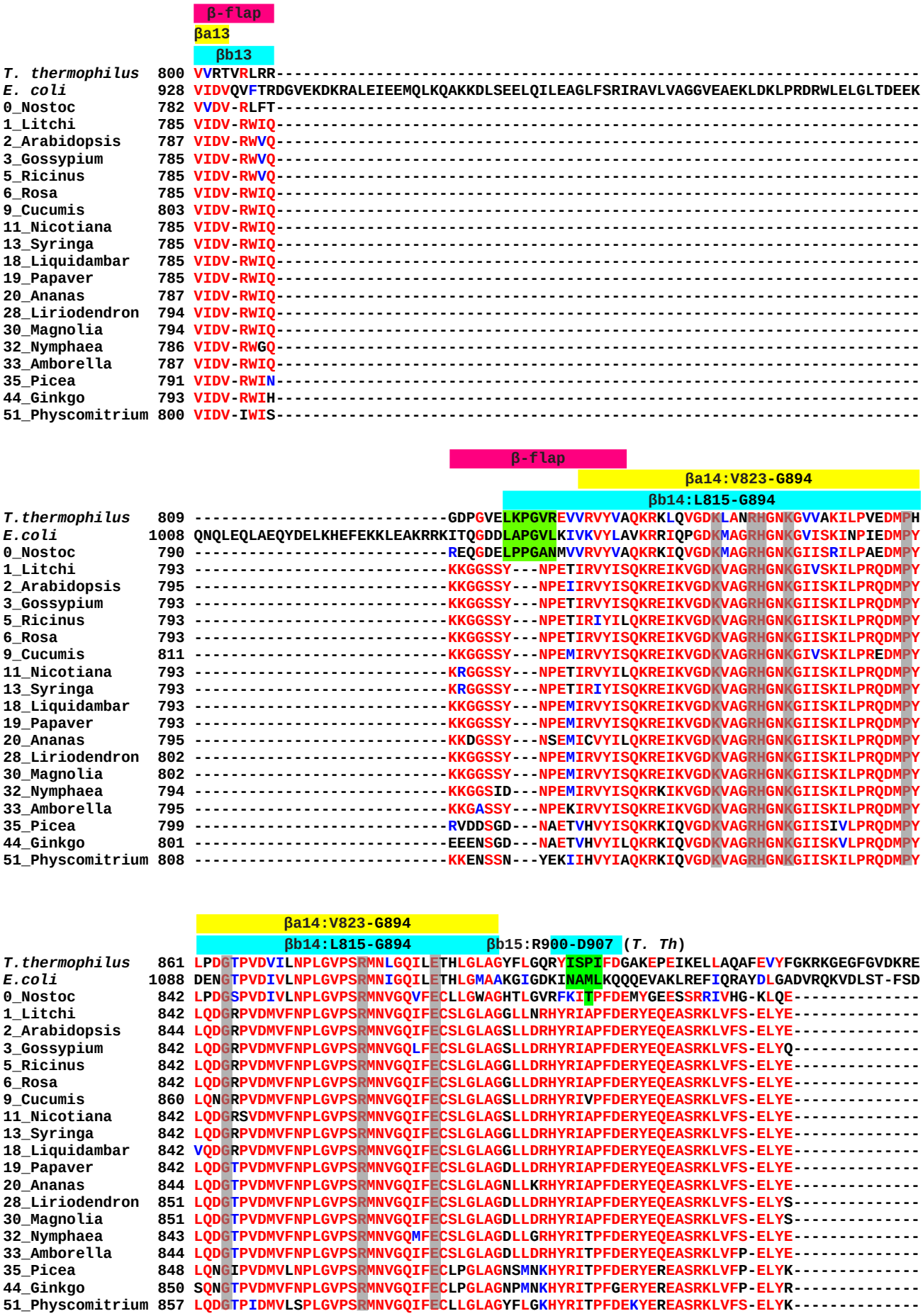

β1 domain: Q22-N130 + V336-S392
βa3:F78-I101
βa4:E112-F148
βb3:L64-I101
βb4:D111-F148
T. thermophilus 66 LDFLEYRLGEPPFPQDECREKDLTYQAPLYARLQLIHKD-------TGLIKEDEVFLGHIPLMTEDGSFIINGADRVIVS
E. coli 68 LQYVSYRLGEPVFDVQECQIRGVTYSAPLRVKLRLVIYEREAPEGTVKDIKEQEVYMGEIPLMTDNGTFVINGTERVIVS
0_Nostoc 53 FLGQNYKLKEPKYSVEEAKRRDSTYAVQMYVPTRLINKE-------TGEIKEQEVFIGDLPLMTDRGTFIINGAERVIVN
1_Litchi 52 LFVETYQLVEPLIKERDAVYESLTYSSELYVSAGLIWKS-------RGDMQEQTIFIGNIPLMNSLGTSIVNGIYRIVIN
2_Arabidopsis 52 LFVETYQLVEPLIKERDAVYESLTYSSELYVSAGLIWKT-------SRNMQEQRIFIGNIPLMNSLGTSIVNGIYRIVIN
3_Gossypium 52 LFVETYQLVEPLIKERDAVYESLTYSSELYVSAGLIWKT-------SKDMQEQTIFIGNIPLMNSLGTSIVNGIYRIVIN
5_Ricinus 52 LFVETYQLVEPLIKEGDAVYESLTYSSELYISAGLIWKT-------SRDMQEQTIFIGNIPLMNSLGTFIINGIYRIVIN
6_Rosa 52 LFVETYQLVEPLIKERDAVYESLTYSSELYVSAGLIWKN-------SRDMQEQTIFIGNIPLMNSLGTSIVNGIYRIVIN
9_Cucumis 70 LFVETYKLVEPLIKERDAVYESLTYSSELYVSAGLIWKT-------RRDMQEQTIFIGNIPLMNSLGTSIVNGLYRIVIS
11_Nicotiana 52 LFVETYQLVEPLIKERDAVYESLTYSSELYVSAGLIWKN-------SRDMQEQTIFIGNIPLMNSLGTSIVNGIYRIVIN
13_Syringa 52 LFVERYQLVEPLIKERDAVYESLTYSSELYVSGGLIWKT-------SRDMQEQTIFIGNIPLMNSLGTSIVNGIYRIVIN
18_Liquidambar 52 LFVETYQLVEPLIKERDAVYESLTYSSELYVSAGLIWKS-------SGDMQEQTIFIGNIPLMNSLGTSIVNGIYRIVIN
19_Papaver 52 LFVETYQLVEPLIKERDAVYESLTYSSELYVPAGLIWKT-------GRDIQEQTIFIGNIPLMNSLGTFIVNGIYRIVIN
20_Ananas 52 LFAERYQLVEPLIKERDAVYESLTYSSELYVPAGLIWKT-------GRDMQEQTVFIGNIPLMNSLGTSIVNGIYRIVIN
28_Liriodendron 61 LFVETYQLVEPLIKERDAVYESLTYSSELYVPAGLIWKT-------GRDMQEQTVFIGNIPLMNSLGTSIVNGIYRIVIN
30_Magnolia 61 LFVETYQLVEPLIKERDAVYESLTYSSELYVPAGLIWKT-------GRDMQEQTVFIGNIPLMNSLGTSIVNGIYRIVIN
32_Nymphaea 53 LFEESYQLVEPLIKERDAVYESITYSSELYVPAGLIWRT-------GRNMQEQTVLLGNIPLMNSLGTSIVNGIYRIVIN
33_Amborella 52 LLVETYQLAEPLIKERDAVYESLTHSSELYVPAGLIWKV-------GRDMQEQTVFIGNIPLMNSLGTFIVNGIYRIVIN
35_Picea 52 LFGNEYELAEPFIKERDAVYQSLTYYSELYVPARSIRRN-------SSKIQKQTVFLGNIPLMNSHGTFVVNGIYRIVVN
44_Ginkgo 58 LSGKKYKSAEPLIEERNAVYQSLTYSSELYVPARLIQKN-------RRKIQKQTVFLGNIPLMNSRGTFVVNGISRIVVD
51_Physcomitrium 65 LLNKEYKLIKPEKTT--IKFHYNTYSSDLYVTCRLLRRK------KKIEIQKQTIFIGSIPLIDYQSTFRINSVTRVIIN


β2 domain: R142-D324
βa5:F191-G201
βb5:Y158-L165
βb6:R168-Y202
T. thermophilus 139 QIHRSPGVYFTPDPARP--GRY-IASIIPLPKRGPWIDLEVEPNGVVSMKVN-KRKFPLVLLLRVLGYDQETLARELGAY
E. coli 148 QLHRSPGVFFDSDKGKTHSSGKVLYNARIIPYRGSWLDFEFDPKDNLFVRIDRRRKLPATIILRALNYTTEQILDLF--F
0_Nostoc 126 QIVRSPGVYYKSEIDKN--GRR-TYSASLIPNRGAWLKFETDRNDLVWVRIDKTRKLSAQVLLKALGLSDNEIFDAL-RH
1_Litchi 125 QILQSPGIYYQSEFDHN--GIL-IYAGTIISDWGGRLELEIDRKARIWARVSRKQKISILVLSSAMGSNLREILENI-CY
2_Arabidopsis 125 QILQSPGIYYQSELDHN--GIS-VYTGTIISDWGGRLELEIDKKARIWARVSRKQKISILVLSSAMGLNLREILENV-CY
3_Gossypium 125 QILQSPGIYYRSELDHN--GIS-VYTGTIISDWGGRLELEIDRKARIWARVSRKQKISILVLSSAMGSNLREILENV-CY
5_Ricinus 125 QILQSPGIYYRSELDHN--GIS-VYTGTIISDWGGRVELEIDRKARIWARVSRKQKISILVLLSAMGLNLREILENV-RY
6_Rosa 125 QILQSPGIYYRSELDHN--GIS-VYTGIIISDWGGRLELEIDRKARIWARVSRKQKISILVLSSAMGSNLREILENA-RY
9_Cucumis 143 QILQSPGIYYRSELDHN--GIS-VYTGTIISDWGGRLELEIDRKARIWARVSRKQKISILVLSSAMGSNLREILENV-CY
11_Nicotiana 125 QILQSPGIYYRSELDHN--GIS-VYTGTIISDWGGRSELEIDRKARIWARVSRKQKISILVLSSAMGLNLREILENV-CY
13_Syringa 125 QILQSPGIYYRSELDHN--GIS-VYTGTIISDWGGRSELEIDRKARIWARVSRKQKISILVLSSAMGSNLREILDNV-CY
18_Liquidambar 125 QILQSPGIYYRSELDHN--GIS-IYTGTIISDWGGRLELEIDRKARIWARVSRKQKISILVLSSAMGSNLREILENV-CY
19_Papaver 125 QILQSPGIYYRSELDHN--GIS-VYTATIISDWGGRSELEIDRKARIWARVSRKQKISILVLSSAMGSNLREILDNV-CY
20_Ananas 125 QILQSPGIYYRSELDHN--GIS-VYTSTIISDWGGRSEFEIDRKARIWARVSRKQKISILVLSSAMGSNLREILDNV-CY
28_Liriodendron 134 QILQSPGIYYRSELDHN--GIS-VYTGTIISDWGGRSELEIDRKARIWARVSRKQKISILVLSSAMGSNLREILDNV-CY
30_Magnolia 134 QILQSPGIYYRSELDHN--GIS-VYTGTIISDWGGRSELEIDRKARIWARVSRKQKISILVLSSAMGSNLREILDNV-CY
32_Nymphaea 126 QILQSPGIYYSTGLDHN--GIS-VYTGTIISDWGGRSELEIDRKERIWARVSRKQKISILVLSSAMGSSLREILDNV-CY
33_Amborella 125 QILQSPGIYYSSELDHN--GIS-VYTGTIISDRGGRSELEIDRKARIWARVSRKQKISILVLPSAMGSNLREILDNV-CY
35_Picea 125 QILISPGIYYRSELDHN--RINYIYTGTLISDWGRRSKLEIDVGERIWARVSRKRKISIPVLLSAMGLNLEEILDNT-RY
44_Ginkgo 131 QILRSPGIYYSSEPGHN--GIA-IYTGTIISDWGGRPKLEIDGKTRIWARVSRKQKVSIPVLLSAMGSNFEEILDNV-CY
51_Physcomitrium 137 QILRSPGIYYNSELDHN--GIS-IYTGTIISDWGGRLKLEIDSKTRIWARISKKRKVSILVLLLAMGLTIKQILDSV-CS

```
T. thermophilus   215 GELVQGLM----------------------------------------------------------------------
E. coli           226 EKVIFEIRDNKLQMELVPERLRGETASFDIEANGKVYVEKGRRITARHIRQLEKDDVKLIEVPVEYIAGKVVAKDYIDES
0_Nostoc          202 PEYFQKTIE----------------------------------------------------------------------
1_Litchi          201 PEIFLSFLTD-------------------------KE--------------------------------------
2_Arabidopsis     201 PEIFLSFLTD-------------------------KE--------------------------------------
3_Gossypium       201 PEIFLSFLTD-------------------------KE--------------------------------------
5_Ricinus         201 PEIFLSFLND-------------------------KE--------------------------------------
6_Rosa            201 PEIFLSFLND-------------------------KE--------------------------------------
9_Cucumis         219 PEIFLSFLND-------------------------KE--------------------------------------
11_Nicotiana      201 PEIFLSFLSD-------------------------KE--------------------------------------
13_Syringa        201 PEIFLSFLND-------------------------KE--------------------------------------
18_Liquidambar    201 PEIFLSFLND-------------------------KE--------------------------------------
19_Papaver        201 PEIFLSFPND-------------------------KE--------------------------------------
20_Ananas         201 PEIFLSFPND-------------------------KE--------------------------------------
28_Liriodendron   210 PEIFLSFPND-------------------------KE--------------------------------------
30_Magnolia       210 PEIFLSFPND-------------------------KE--------------------------------------
32_Nymphaea       202 PEIFLSFPNE-------------------------KE--------------------------------------
33_Amborella      201 PEILLYFPNE-------------------------KE--------------------------------------
35_Picea          202 PEKIFFLLKK-------------------------KKGRW-----------------------------------
44_Ginkgo         207 PEIFLSFL---------------------------NGRQ-----------------------------------
51_Physcomitrium  213 SKIFLDFLKE-------------------------KK-------------------------------------

                                                                    β2 domain: R142-D324

                                                                                  βb7:A234-K280
T. thermophilus   223 ------------------------DE------------------SVFAMRPEEALIRLFTLLRPGDPPKR--DKAVA
E. coli           306 TGELICAANMELSLDLLAKLSQSGHKRIETLFTNDLDHGPYISETLRVDPTNDRLSALVEIYRMMRPGEPPTR--EAAES
0_Nostoc          211 ----------------------------------------------KEGQFSEEEALMELYRKLRPGEPPTVLG--GQQ
1_Litchi          213 ----------------------------------------------KKKIGSKENAILEFYQQFACVGGDPIFSESLCK
2_Arabidopsis     213 ----------------------------------------------KKKIGSKENAILEFYQQFSCVGGDPIFSESLCK
3_Gossypium       213 ----------------------------------------------KKKIGSKENAILEFYQQFSCVGGDPVFSESLCK
5_Ricinus         213 ----------------------------------------------KKKIGSKENAILEFYQQFTCVGGDPVFSESLCK
6_Rosa            213 ----------------------------------------------KKKIGSKENAILEFYQQFACVGGDPVVSESLCK
9_Cucumis         231 ----------------------------------------------KKKIGSKENAILEFYQQFSCVGGDPVFSESLCK
11_Nicotiana      213 ----------------------------------------------RKKIGSKENAILEFYQQFACVGGDPVFSESLCK
13_Syringa        213 ----------------------------------------------RKKIGSKENAILEFYQQFACVGGDPVFSESLCK
18_Liquidambar    213 ----------------------------------------------KKKIGSKENAILEFYQQFACVGGDPVFSESLCK
19_Papaver        213 ----------------------------------------------KKKIGSKENAILEFYQQFACVGGDPVFSESLCK
20_Ananas         213 ----------------------------------------------KKKIGSKENAILEFYQQFACVGGDPVFSESLCK
28_Liriodendron   222 ----------------------------------------------KKKIGSRENAILEFYQQFACVGGDPVFSESLCK
30_Magnolia       222 ----------------------------------------------KKKIGSRENAILEFYQQFACVGGDPVFSESLCK
32_Nymphaea       214 ----------------------------------------------KKKISSKENAILEFYQKFACVGGDPVFSESLCK
33_Amborella      213 ----------------------------------------------KKKIGSKENAILEFYQQFSCVGGDPVFSESLCK
35_Picea          217 ------------------------ER------------------EEYIWSKEKAILEFYKKLYCVSGDLVFSESLCK
44_Ginkgo         219 ------------------------KR------------------KKYLRSEENAILEFHKKLYCVGGDLVFSESLCK
51_Physcomitrium  225 ------------------------KK------------------KEHLQSTEDAMVELYKQLYYIGGDLLFSESIRK

                                  β2 domain: R142-D324
                                                                                  βa6:D323-V355
                                  βb7:A234-K280                  βb8:V302-G316    βb9:D323-M359
T. thermophilus   256 YVYGLIADPRRYDLGEAGRYKAEEKLGIRLSGRTLARFEDGEFKDEVFLPTLRYLFALTAGVPGHEVDDIDHLGNRRIRT
E. coli           384 LFENLFFSEDRYDLSAVGRMKFNRSLLREEIEGS------GILSKDDIIDVMKKLIDIRNGK--GEVDDIDHLGNRRIRS
0_Nostoc          242 LLDSRFFDPKRYDLGRVGRYKLNKKLRLSVPDTMRVLTSS------DILAAVDYLINLEYDI--GNIDDIDHLGNRRVRS
1_Litchi          246 ELQKKF-FHQRCELGRIGRRNMNRRLNLNIPQNNTFLLPR------DVLAAADHLIELKFGM--GTLDDMNHLKNKRIRS
2_Arabidopsis     246 ELQKKF-FHQRCELGRIGRRNINWRLNLNIPQNNIFLLPR------DVLAAADHLIGMKFGM--GTLDDMNHLKNKRIRS
3_Gossypium       246 ELQKKF-FQQRCELGRIGRRNMNQRLNLNIPQNNTFLLPR------DILAAADRLIGMKFGM--GPLDDMNHLKNKRIRS
5_Ricinus         246 ELQKKF-FQQRCELGRIGRLNMNRRLNLDIPHNNTFLLPR------DILAAADHLIGMKFGM--GTLDDMNHLKNKRIRS
6_Rosa            246 ELQKKF-FQQRCELGRIGRRNMNRRLNLDIPQNNTFLLPR------DILAAADHLIGMKFGM--GTLDDMNHLKNKRIRS
9_Cucumis         264 ELQKKF-FQQRCELGRIGRRNLNQRLNLDIPENNTFLLPR------DILAAADHLIGLKFGM--GTLDDMNHLKNKRIRS
11_Nicotiana      246 ELQKKF-FQQRCELGRIGRRNMNRRLNLDIPQNNTFLLPR------DILAAADHLIGLKFGM--GALDDMNHLKNKRIRS
13_Syringa        246 ELQKKF-FQQRCELGRIGRRNMNRRLNLDIPQNNTFLLPR------DILAAADHLIELKFGM--GTLDDMNHLKNKRIRS
18_Liquidambar    246 ELQKKF-FQQRCELGRIGRRNMNRRLDLNIPQNNTFLLPR------DILAAADHLIGMKFGM--GTLDDMNHLKNKRIRS
19_Papaver        246 ELQKKF-FQQRCELGRIGRRNMNRRLNLDIPQNNTFLLPR------DVLAAADHLIGMKFGM--GTLDDMNHLKNKRIRS
20_Ananas         246 ELQKKF-FQQRCELGRIGRRNMNRRLNLDIPQNNTFLLPR------DVLAAADHLIGMKFGM--GTFDDMNHLKNKRIRS
28_Liriodendron   255 ELQKKF-FQQRCELGRIGRRNMNRRLNLDIPQNNTFLLPR------DVLAAADHLIGMKFGM--GTLDDMNHLKNKRIRS
30_Magnolia       255 ELQKKF-FQQRCELGRIGRRNMNRRLNLDIPQNNTFLLPR------DVLSAADHLIRMKFGM--GTLDDMNHLKNKRIRS
32_Nymphaea       247 ELQKKF-FQQRCELGRIGRRNMNQRLNLDIPQNNTFLLPR------DVLAAADHLIGMKFGM--GTLDDMNHLKNKRIRS
33_Amborella      246 ELQKRF-FQQRCELGRIGRQNMNQRLNIDIPQNNTFLLPR------DVLAATDHLIGMKFGM--GTLDDMNHLKNKRIRS
35_Picea          252 ELQEKF-FRQRCELGKIGRRNPNQKLNLDIPENEIFSLPQ------DVLAAVDYLIGVKFGM--GTLDDIDHLRNRRIRS
44_Ginkgo         254 ELQKKS-LQQRCELGRIGRRNPNQKLNLDIPENEIFSLPQ------DVLAAADYSIRVKFGM--GTLDDMDHLKNKRIRS
51_Physcomitrium  260 ELQKKF-FQQRCELGKIGRLNVNKKLSLDIPENEFFLLPQ------DILAAIDYLIKIKFGI--GTLDDIDHLKNRRIRS
```


β1 domain: Q22-N130 + V336-S392
βa6:D323-V355
βa7:S375-E421
βb9:D323-M359
βb10:L367-V474
T. thermophilus 336 VGELMTDQFRVGLARLARGVRERMLMGSE--DSLTPAKLVNSRPLEAAIREFFSRSQLSQFKDETNPLSSLRHKRRISAL
E. coli 456 VGEMAENQFRVGLVRVERAVKERLSLGDL--DTLMPQDMINAKPISAAVKEFFGSSQLSQFMDQNNPLSEITHKRRISAL
0_Nostoc 314 VGELLQNQVRVGLNRLERIIRERMTVS--DAEVLTPASLVNPKPLVAAIKEFFGSSQLSQFMDQTNPLAELTHKRRLSAL
1_Litchi 317 VADLLQDQFGLALVRLENVVRGAIGGAIRHKLMPTPQNLVTSTPLTTTYDSFFGLHPLSQVLDRTNPLTQIVHGRKLSYL
2_Arabidopsis 317 VADLLQDQLGLALARLENVVKGTISGAIRHKLIPTPQNLVTSTPLTTTYESFFGLHPLSQVLDRTNPLTQIVHGRKLSYL
3_Gossypium 317 VADLLQDQFGLALVRLENVVRGTICGAIRHKLIPTPQNLVTSTPLTTTYESFFGLHPLSQVLDRTNPLTQIVHGRKLSYL
5_Ricinus 317 VADLLQDQFGLALIRLENVVRGTICGAIRHKLIPTPQTLVTSTPLTTTYESFFGLHPLSQVLDRTNPLTQIVHGRKSSYL
6_Rosa 317 VADLLQDQFGLALVRLENMVRGTICGAIKHKLIPTPQNLVTSTPLTTTYESFFGLHPLSQVLDQTNPLTQIVHGRKSSYL
9_Cucumis 335 VADLLQDQFGLALVRLENMVRGTICGAIRHKLIPTPQNLVTSTPLTTTFESFFGLHPLSQVLDRTNPLTQIVHGRKLSYL
11_Nicotiana 317 VADLLQDQFGLALVRLENVVRGTICGAIRHKLIPTPQNLVTSTPLTTTYESFFGLHPLSQVLDRTNPLTQIVHGRKLSYL
13_Syringa 317 VADLLQDQFGLALVRLENVVRGTICGAIRHKLIPTPQNLVTSTPLTTTYESFFGLHPLSQVLDRTNPLTQIVHGRKLSYL
18_Liquidambar 317 VADLLQDQFGLALVRLENAVRGTICGAIRHKLIPTPQNLVTSTPLTTTYEYFFGLHPLSQVLDRTNPLTQIVHGRKSSYL
19_Papaver 317 VADLLQDQFGLALVRLENVVRGTICGAIRHKFIPTPHNLVTSTPLTTTYESFFGLHPLSHVLDRTNPLTQIVHGRKLSYL
20_Ananas 317 VADLLQDQFGLALVRLENAVRGTICGAIRHKLIPTPQNLVTSTSLTTTYESFFGLHPLSQVLDRTNPLTQIVHGRKLSYL
28_Liriodendron 326 VADLLQDQFGLALVRLENAVRGTICGAIRHKLIPTPQNLVTSTPLTTTYESFFGLHPLSQVLDRTNPLTQIVHGRKSSYL
30_Magnolia 326 VADLLQDQFGLALVRLENAVRGTICGAIRHKLIPTPQNLVTSTPLTTTYESFFGLHPLSQVLDRTNPLTQIVHGRKSSYL
32_Nymphaea 318 VADLLQDQFGLALVRLENVVRGTICGAIRHKLIPTPRNLVTSTPLTTTYESFFGLHPLSQVLDRTNPLTQIVHGRKSSYL
33_Amborella 317 VADLLQDQFGLALVRLENVVRGTICGAIRHKFIPTPQNLVTSTPLTTTYESFFGLHPLSQVLDRTNPLTQIVHGRKSSYL
35_Picea 323 VADLLQNQFRLALGRLEDAVRRTIHRATKRR--STPQNLVTSTLLKNTFQDFFGSHPLSQFLDQTNPLTEIAHGRKLSHL
44_Ginkgo 325 VADLLQNQFGLALGRLVNSVRRTIRRATKCKCLPTPKNLVTSTPLTTTFQDFFGLHPLSQFLDQTNPLTEAVHRRKLSYL
51_Physcomitrium 331 VADLLQDQLKLALIRLENSVRQVMRRTTKRKRLLSPKNLITQTPLIATFKEFFGSHPLSQFLDQTNSLAEIVHKRRLSSL
Fork-loop 2: S411-R428
βa8:D426-Y471
βb10:L367-V474
βb11:V479-L503
T. thermophilus 414 GPGGLTRERAGFDVRDVHRTHYGRICPVETPEGANIGLITSLAAYARVDELGFIRTPYRRVVGGVVT--DEVVYMTATEE
E. coli 534 GPGGLTRERAGFEVRDVHPTHYGRVCPIETPEGPNIGLINSLSVYAQTNEYGFLETPYRKVTDGVVT--DEIHYLSAIEE
0_Nostoc 392 GPGGLTRERAGFAVRDIHPSHYGRICPIETPEGPNAGLIGSLATHARVNQYGFLETPFRPVENARVRFDLPPVYMTADE-
1_Litchi 397 GPGGLTGRTASFRIRDIHPSHYGRICPIDTSEGINVGLIGSLAIHARIGYWGSLESPFYEIFEKSKK--MRMLYLSPSID
2_Arabidopsis 397 GPGGLTGRTANFRIRDIHPSHYGRICPIDTSEGINVGLIGSLSIHARIGDWGSLESPFYELFEKSKKARIRMLFLSPSQD
3_Gossypium 397 GPGGLTGRTANFRIRDIHPSHYGRICPIDTSEGINVGLIGSLAIHARIGHWGSLESPFYKIFERSKK--AQMLYLSPSRD
5_Ricinus 397 GPGGLTGRTASFRIRDIHPSHYGRICPIDTSEGINVGLIGSLAIHAKIGHWGSLESPFYVISEESKK--VRMFYLSPNRE
6_Rosa 397 GPGGLTGRTASFRIRDIHPSHYGRICPIDTSEGINVGLIGSLAIHAKIGYWGSLESPFYEISERSKK--VRMLYLSPSKD
9_Cucumis 415 GPGGLTGRTASFRIRDIHPSHYGRICPIDTSEGINVGLIGSLAIHARIGHWGSLETPFYEISERSKK--VRMLYLSPSRD
11_Nicotiana 397 GPGGLTGRTASFRIRDIHPSHYGRICPIDTSEGINVGLIGSLAIHARIGHWGSLESPFYEISERSTG--VRMLYLSPGRD
13_Syringa 397 GPGGLTGRTASFRIRDIHPSHYGRICPIDTSEGINVGLIGSLAIHARIGHWGSLESPFYEISERSTG--VRMLYLSPGRD
18_Liquidambar 397 GPGGLTGRTASFRIRDIHPSHYGRICPIDTSEGINVGLIGSLAIHARIGRWGSLESPFYEISERSKK--VRMLYLSPSRD
19_Papaver 397 GPGGLTGRTASFRIRDIHPSHYGRICPIDTSEGINVGLIGSLAIHARIGHWGSLESPFFEIDERF--KGVRVVYLSPSRD
20_Ananas 397 GPGGLTGRTASFRIRDIHPSHYGRICPIDTSEGINVGLIGSLAIHVRIGHWGSIESPFYEISEKQKEKEPQMVYLSPNRD
28_Liriodendron 406 GPGGLTGRTASFRIRDIHPSHYGRICPIDTSEGINVGLIGSLAIHARIGHWGSIESPFYEISERS--KEVQMVYLSPSRD
30_Magnolia 406 GPGGLTGRTASFRIRDIHPSHYGRICPIDTSEGINVGLIGSLAIHARIGHWGSIESPFYEISERS--KEVQMVYLSPSRD
32_Nymphaea 398 GPGGLTGRTASFRIRDIHPSHYGRICPIDTSEGINVGLIGSLAIHARVGDWGSIETPFYEISERS--KEEQMVYLSPSRD
33_Amborella 397 GPGGLTGRTASFRIRDIHPSHYGRICPIDTSEGINVGLIGSLAIHARIGDWGSIRSPFYEISERS--KEEQMVYLSPRRD
35_Picea 401 GPGGLTGRTASFRTRDIHPSYYGRICPIDTSEGMNAGLVASLSIHAKIGQCGSLQSPFYKISERSRE--EHMVYLLPGED
44_Ginkgo 405 GPGGLTRRTASFRIRDIHPSHYGRICPIETSEGINAGLVASLAIHAKIGHCGSLRSPFHKISEGSKE--EHMVYPSPGE-
51_Physcomitrium 411 GPGGVTRRTAGFQVRDIHFSHYTRICPIETSEGMNAGLIASLAIHANVNNWGFLESPFYKISKNVKE--EKIINLSAGED
βa9:Y485-A499
βa10:V529-D590
βb11:V479-L503
βb12:I508-V613
T. thermophilus 492 --DRYTIAQANTPLEGNRIAAERV-VARRKGEPVIVSPEEVEFMDVSPKQVFSVNTNLIPFLEHDDANRALMGSNMQTQA
E. coli 612 --GNYVIAQANSNLDEEGHFVEDLVTCRSKGESSLFSRDQVDYMDVSTQQVVSVGASLIPFLEHDDANRALMGANMQRQA
0_Nostoc 471 -EDDLRVAPGDIPVDENGHIIGPQVPVRYRQEFSTTTPEQVDYVAVSPVQIVSVATSMIPFLEHDDANRALMGSNMQRQA
1_Litchi 475 EYCMV--AAGNSLALSQGIQEEQVVPTRYRQEFLTIAWERVHLRSIFPSQYFSIGASLIPFIEHNDANRALMSSNMQRQA
2_Arabidopsis 477 EYYMI--AAGNSLALNRGIQEEQAVPARYRQEFLTIAWEEVHLRSIFPFQYFSIGASLIPFIEHNDANRALMSSNMQRQA
3_Gossypium 475 EYYMV--AAGNSLALNQGIQEEQVVPARYRQEFLTIAWEQVHLRSIFPFQYFSIGASLIPFIEHNDANRALMSSNMQRQA
5_Ricinus 475 EYHMV--AAGNSLALNRGVQEEQVAPARYRQEFLTIAWEQVHLRSIFPFQYFSIGASLIPFIEHNDANRALMSSNMQRQA
6_Rosa 475 EYYMI--AAGNSLALNRGIQEEQVVPARYRQEFLTVEWEQVHLRSIFPFQYFSIGASLIPFIEHNDANRALMSSNMQRQA
9_Cucumis 493 EYYMV--ATGNSLALNPGIQEEQIVPARYRQEFLTIEWEQVHLRSIFPFQYFSIGASLIPFIEHNDANRALMSSNMQRQA
11_Nicotiana 475 EYYMV--AAGNSLALNQDIQEEQVVPARYRQEFLTIAWEQVHLRSIFPFQYFSIGASLIPFIEHNDANRALMSSNMQRQA
13_Syringa 475 EYYMV--AAGNSLALNQDIQEEQVVPARYRQEFLTIAWEQVHLRSIFPFQFFSIGASLIPFIEHNDANRALMSSNMQRQA
18_Liquidambar 475 EYYMV--AAGNSLALNQGIQEEQVVPARYRQEFLTIAWEQVHLRSIFPFQYFSIGASLIPFIEHNDANRALMSSNMQRQA
19_Papaver 475 EYYMV--SAGNSLALNQGIQEEQVVPARYRQEYLTIAWEQIHLRSIFPFQYFSIGASLIPFIEHNDANRALMSSNMQRQA
20_Ananas 477 EYYMV--AAGNSLALNRGIQEEQVVPARYRQEFLTIAWEQIHLRSILPFQYFSIGASLIPFIEHNDANRALMSSNMQRQA
28_Liriodendron 484 EYYMV--AAGNSLALNWGIQEEQVVPARYRQEFLTIAWEQIHLRSIFPFQYFSIGASLIPFIEHNDANRALMSSNMQRQA
30_Magnolia 484 EYYMV--AAGNSLALNWGVQEEQVVPARYRQEFLTIAWEQIHLRSIFPFQYFSIGASLIPFIEHNDANRALMSSNMQRQA
32_Nymphaea 476 EYYMV--AAGNSLALTRGIQEEEVGPARYRQEFLTIAWEQIHLRNIYPFQYFSIGASLIPFIEHNDANRALMSSNMQRQA
33_Amborella 475 EYYMVMVAAGNSLALNQDIQDEQVVPARYRQEFVTIAWEHIDLRSIYPLQYFSIGASLIPFIEHNDANRALMSSNMQRQA
35_Picea 479 EDEYYRIATGNSLALNQGIQEEQITPARYRQEFIVIAWEQIHFRSIFPFQYFSVGVSLIPFLEHNDANRALMGSNMQRQA
44_Ginkgo 482 -DEYYRIATGNSLALNQGIQEEQVTPARYRQEFLAIAREQIHFRSIFPFQYFSVGVSLIPFLEHNDANRALMGSNMQRQA
51_Physcomitrium 489 --EYYRIATGNCLALDQGTQKIQITPARYRQEFLAIAWEQIHLRSIYPLQYFSVGVSLIPFLEHNDANRALMGSNMQRQA


βa10:V529-D590
βb12:I508-V613
βb13:Y623-R808
T. thermophilus 569 VPLIRAQAPVVMTGLEERVVRDSLAALYAEEDGEVAKVDGNRIVVRY-ED------GRLVEYPLRRFYRSNQGTALDQRP
E. coli 690 VPTLRADKPLVGTGMERAVAVDSGVTAVAKRGGVVQYVDASRIVIKVNEDEMYPGEAGIDIYNLTKYTRSNQNTCINQMP
0_Nostoc 550 VPLLKPERPLVGTGLEAQGARDSGMVVVSRTDGDVTYVDATEIRVRPKPN------TTEIRYPLSKYQRSNQDTCLNQKP
1_Litchi 553 VPLSRSEKCIVGTGLERHVALDSGVPAIADHEGRVLYTDIDKIVLSG--N------GDTIGIPLVMYQRSNKNTCMHQKT
2_Arabidopsis 555 VPLSRSEKCIVGTGLERQVALDSGVPAIAEHEGKILYTDTEKIVFSG--N------GDTLSIPLIMYQRSNKNTCMHQKP
3_Gossypium 553 VPLSRSEKCIVGTGLERQVALDSGVPAIADHEGKIISTDTDKIILSG--N------GDALGIPLVMYQRSNKNTCMHQTA
5_Ricinus 553 VPLSRSEKCIVGTGLERQVALDSGVPAIAEREGKIIYTDIDKIILSG--N------GDTLRIPLVMYQRSNKNTCMHQKP
6_Rosa 553 VPLSRSEKCIVGTGLERQVALDSGVPAIAEHEGKIIYTDTDKIILSG--N------GDTLNIPLVIYQRSNKNTCMHQKP
9_Cucumis 571 VPLSRSEKCIVGTGLERQVARDSGVAAIAEHGGKIIYTDTDKIIFSG--N------GHTRRIPLVMYQRSNKNTCMQQKS
11_Nicotiana 553 VPLSRSEKCIVGTGLERQAALDSGALAIAERGERVVYTNTDKILLAG--N------GDILSIPLVIYQRSNKNTCMHQKL
13_Syringa 553 VPLSRPEKCIVGTGLERQAALDSGALAIAERGKIIYTDTEKILFSG--N------GDTLSIPLVMYQRSNKNTCMHQKP
18_Liquidambar 553 VPLSRSEKCIVGTGLERQAALDSGIPVLAEHEGKIVYTDTDKIILSG--N------GDTLSIPLVMYQRSNKNTCMHQKP
19_Papaver 553 VPLSRSEKCIVGTGLERQVALDSGVSAIAEHEGKVVSTDTDKIVFSG--N------GDTLSIPLVMYQRSNKNTCMHQKS
20_Ananas 555 VPLSRSEKCIVGTGLERQTALDSGVLVIAEHEGKIIYTDTHKIMLSS--N------GNTISIPLVMYQRSNKNTCMHQKP
28_Liriodendron 562 VPLSRSEKCIVGTGLECQAALDSGVSAIAEHEGKIVYTDTDKIVLSG--N------GDTISIPLVMYQRSNKNTCMHQKP
30_Magnolia 562 VPLSRSEKCIVGTGLECQAALDSGVSAIAEHEGKIVSTDTDKIVLSG--N------GNTISIPLVMYQRSNKNTCMHQKP
32_Nymphaea 554 VPLSQSEKCIVGTGLERQAALDSGGSAIAERGKIIYTDAEKIVLSG--N------GDTISIPLVMYQRSNKNTWMHQKP
33_Amborella 555 VPLSRSEKCIVGTGLERQAALDSGGSAIAQHEGKVIYTDTEKILLSG--N------GDTISIPLLMYQRSNKNTCMHQKP
35_Picea 559 VPLFQPEKCIAGTGLEGQAALDSGSVAIATQEGRIEYIDAVNITSSV--N------GDTVRTELVIYQRSNTNTCTHQKP
44_Ginkgo 561 IPLFQPEKCITGTGLEGQVALDSGSVTIAIQEGRIEYTDAENITFSF--N------GDTIGTELVLYQRSNKNTCMHQKP
51_Physcomitrium 567 VPLIKLEKCIVGTGLESQVALDSGNVMITKQSEKIMYTDGKKISLLN-NT------NETVNTHLIIYQRSNNSTCIHQKP
β-flap
βa11:F665-K716
βb13:Y623-R808
T. thermophilus 642 RVVVGQRVRKGDLLADGPASENGFLALGQNVLVAIMPFDGYNFEDAIVISEELLKRDFYTSIHIERYEIEARDTKLGPER
E. coli 770 CVSLGEPVERGDVLADGPSTDLGELALGQNMRVAFMPWNGYNFEDSILVSERVVQEDRFTTIHIQELACVSRDTKLGPEE
0_Nostoc 624 LVRIGEKVVAGQVLADGSSTEGGELALGQNIVVAYMPWEGYNYEDAILISERLVQDDIYTSIHIEKYEIEARQTKLGPEE
1_Litchi 625 QVGRGKCIKKGQVLADGAATVGGELALGKNVLVTYMPWEGYNFEDAVLISERLIYRDIYTSFHIQKYEIQTHVTSQGPER
2_Arabidopsis 627 QVRRGKCIKKGQILADGAATVGGELALGKNILVAYMPWEGYNFEDAVLISECLVYGDIYTSFHIRKYEIQTHVTTQGPER
3_Gossypium 625 RVRRGKCIKKGQILADGAATVGGELALGKNVLVAYMPWEGYNFEDAVLISERLVYEDIYTSFHIRKYEIQTHVTSQGPER
5_Ricinus 625 QVPRGKCIKKGQVLADGAATIGGELALGKNVLVAYMPWEGYNFEDAVLISERLVYEDIYTSFHIRKYEIQTHVTSQGPER
6_Rosa 625 QVQRGKCIKKGQILADGAATVGGELALGKNVLVAYMPWEGYNSEDAVLISERLVYGDIYTSFHIRKYEIQTHVTSNGPER
9_Cucumis 643 QVHQGKCIKKGQILADGAATVGGELALGKNVLVAYMPWEGYNFEDAVLISERLVYEDIYTSFHIRKYEIQTHVTSHGPER
11_Nicotiana 625 QVPRGKCIKKGQILADGAATVGGELALGKNVLVAYMPWEGYNSEDAVLISERLVYEDIYTSFHIRKYEIQTHVTSQGPEK
13_Syringa 625 QVQRGKCIKKGQILADGAATVGGELALGKNVLVAYMPWEGYNFEDAVLISERLVYEDIYTSFHIRKYEIQTHVTSQGPER
18_Liquidambar 625 QVQRSKCIKKGQILADGAATVGGELALGKNVLVAYMPWEGYNFEDAVLISERLVYGDIYTSFHIRKYEIQTHVTSQGPER
19_Papaver 625 QVQRGNCIKKGQILADGVATVGGELALGKNVLVAYMPWEGYNFEDAVLISERLVYGDVYTSFHIRKYEIQTHVTSQGPER
20_Ananas 627 RVRRGKCIKKGQILADGAATVGGELALGKNVLVAYMPWEGYNSEDAVLISERLVYEDIYTSFHIRKYEIQTHVTSQGPER
28_Liriodendron 634 QVRRGKCIKKGQILADGAATVGGELALGKNVLVAYMPWEGYNSEDAVLISERLVYGDIYTSFHIRKYEIQTHVTSQGPER
30_Magnolia 634 QVRRGKCIKKGQILADGAATVGGELALGKNVLVAYMPWEGYNSEDAVLISERLVYGDIYTSFHIRKYEIQTHVTSQGPER
32_Nymphaea 626 QVHRGKYLKKGQILADGAATVGGELALGKNVSVAYMPWEGYNSEDAVLISERLVYDDIYTSFHIRKYEIQTHVTSQGPER
33_Amborella 627 QVHRDKYVKKGQVLADGAATVGGELALGKNVLVAHMPWEGYNFEDAVLISERLVYGDIYTSFHIRKYEIQTHVTSHGPEK
35_Picea 631 QVRQGECVKKGQILADGAATVGGELSLGKNVLVAYMPWEGYNFEDAILISERLVYEDIYTSFHIVRYRIEICMTSQGPER
44_Ginkgo 633 RVRQGECVKKGQILADGAATVEGELSPGKNILVAYMPWEGYNFEDAILISERLVYEDIYTSFHIERHGIRTCMTSQGPER
51_Physcomitrium 640 QVISKKFLKKGQVLTDGAAILKGELTLGKNILVAYMPWEGYNFEDAILISERLIYEDIYTSIHIERYEIESRNTNQGPEK
Tip
Helix
Tip
β-flap
βa12:A733-K762
βa13:D787-V804
βb13:Y623-R808
T. thermophilus 722 ITRDIPHLSEAALRDLDEEGVVRIGAEVKPGDILVGRTSFKG--ESEPTPEERLLRSIFGEKARDVKDTSLRVPPGEGGI
E. coli 850 ITADIPNVGEAALSKLDESGIVYIGAEVTGGDILVGKVTPKG--ETQLTPEEKLLRAIFGEKASDVKDSSLRVPNGVSGT
0_Nostoc 704 ITREIPNVGEDALRQLDEQGIIRIGAWVEAGDILVGKVTPKG--ESDQPPEEKLLRAIFGEKARDVRDNSLRVPNGEKGR
1_Litchi 705 ITNEIPHLEARLLRNLDQNGIVMLGSWVETGDILVGKLTPQTAKESSYAPEDRLLRAILGIQVSTSKETCLKLPIGGRGR
2_Arabidopsis 707 ITKEIPHLEGRLLRNLDKNGIVMLGSWVETGDILVGKLTPQVAKESSYAPEDRLLRAILGIQVSTSKETCLKLPIGGRGR
3_Gossypium 705 ITNEIPHLEAHLLRNLDKNGIVMLGSWVETGDILVGKLTPQVAKESSYAPEDRLLRAILGIQVSTSKETCLKLPIGGRGR
5_Ricinus 705 ITNEIPHLEAHLLRNLDKNGIVMLGSWVETGDILVGKLTPQMAKESSYAPEDRLLRAILGIQVSTSKETCLKLPIGGRGR
6_Rosa 705 ITNEIPHLEAHLLRNLDKNGIVRLGSWVKTGDILVGKLTPQMAKESSYAPEDRLLRAILGIQVSTSKETCLKLPIGGRGQ
9_Cucumis 723 ITNEIPHLEARLLCNLDKNGIVMLGSWVETGDILVGKLTPQMAKESSYAPEDRLLRAILGIQVSTSKETCLKLPIGGRGR
11_Nicotiana 705 VTNEIPHLEAHLLRNLDKNGIVMLGSWVETGDILVGKLTPQVVKESSYAPEDRLLRAILGIQVSTSKETCLKLPIGGRGR
13_Syringa 705 ITNEIPHLEAHLLRNLDKNGIVMLGSWVETGDILVGKLTPQMVKESSYAPEDRLLRAILGIQVSTSKETCLKLPIGGRGR
18_Liquidambar 705 ITNEIPHLEAHLLRNLDKNGIVMLGSWVETGDILVGKLTPQMAKESSYAPEDRLLRAILGIQVSTSKETCLKLPIGGRGR
19_Papaver 705 ITNEIPHLEAHLLRNLDRNGIVMLGSWVETGDILVGKLTPQMTKESSYAPEDRLLRAILGIQVSTAKETCLKLPIGSRGR
20_Ananas 707 ITKEIPHLEAHLLRNLDRNGIVMLGSWVETGDILVGKLTPQTANESSYSPEDRLLRAILGIQVSTAKETSLKLPIGGRGR
28_Liriodendron 714 ITNEIPHLEAHLLRNLDKNGIVMLGSWIETGDILVGKLTPQTAKESSYAPEDRLLRAILGIQVSTAKETCLKLPIGGRGR
30_Magnolia 714 ITNEIPHLEAHLLRNLDKNGIVMLGSWIETGDILVGKLTPQTAKESSYAPEDRLLRAILGIQVSTAKETCLKLPIGGRGR
32_Nymphaea 706 ITNEIPHLEPYLLRNLDRNGIVMLGSWVETGDVLVGKLTPQTAKESSYAPEDRLLRAILGIQVSTAKETCLKLPIGGRGR
33_Amborella 707 ITNEIPHLEAHLLRNLDRNGIVMLGSWVETGDVLVGKLTPQTAKESSYAPEDRLLRVILGIQVSTAKETCLKLPIGGRGR
35_Picea 711 ITREIPHLDAHSLRHLDENGLVMLGSWIETGDVLVGKLTPQTTEESLCAPEGRLLQTIFGIEVSTARENCLRTPIGGRGR
44_Ginkgo 713 ITKEIPHLDAHLLRHLDENGLVMLGSWIETGDVLVGKLTPQKEEESLCAPEGRLLRTIFGIQVSTARESCLRVPIGGRGR
51_Physcomitrium 720 ITKEIPHLENSVLRHLDKNGLVIPGSWVETGDVLVGKLTPQETEESLRAPEGKLLQAIFGIQVTNAKETCLKVPLNGKGR


β-flap
βa13
βb13
T. thermophilus 800 VVRTVRLRR
E. coli 928 VIDVQVFTRDGVEKDKRALEIEEMQLKQAKKDLSEELQILEAGLFSRIRAVLVAGGVEAEKLDKLPRDRWLELGLTDEEK
0_Nostoc 782 VVDV-RLFT
1_Litchi 785 VIDV-RWIQ
2_Arabidopsis 787 VIDV-RWVQ
3_Gossypium 785 VIDV-RWVQ
5_Ricinus 785 VIDV-RWVQ
6_Rosa 785 VIDV-RWIQ
9_Cucumis 803 VIDV-RWIQ
11_Nicotiana 785 VIDV-RWIQ
13_Syringa 785 VIDV-RWIQ
18_Liquidambar 785 VIDV-RWIQ
19_Papaver 785 VIDV-RWIQ
20_Ananas 787 VIDV-RWIQ
28_Liriodendron 794 VIDV-RWIQ
30_Magnolia 794 VIDV-RWIQ
32_Nymphaea 786 VIDV-RWGQ
33_Amborella 787 VIDV-RWIQ
35_Picea 791 VIDV-RWIN
44_Ginkgo 793 VIDV-RWIH
51_Physcomitrium 800 VIDV-IWIS
β-flap
βa14:V823-G894
βb14:L815-G894
T.thermophilus 809 GDPGVELKPGVREVVRVYVAQKRKLQVGDKLANRHGNKGVVAKILPVEDMPH
E.coli 1008 QNQLEQLAEQYDELKHEFEKKLEAKRRKITQGDDLAPGVLKIVKVYLAVKRRIQPGDKMAGRHGNKGVISKINPIEDMPY
0_Nostoc 790 REQGDELPPGANMVVRVYVAQKRKIQVGDKMAGRHGNKGIISRILPAEDMPY
1_Litchi 793 KKGGSSY---NPETIRVYISQKREIKVGDKVAGRHGNKGIVSKILPRQDMPY
2_Arabidopsis 795 KKGGSSY---NPEIIRVYISQKREIKVGDKVAGRHGNKGIISKILPRQDMPY
3_Gossypium 793 KKGGSSY---NPETIRVYISQKREIKVGDKVAGRHGNKGIISKILPRQDMPY
5_Ricinus 793 KKGGSSY---NPETIRIYILQKREIKVGDKVAGRHGNKGIISKILPRQDMPY
6_Rosa 793 KKGGSSY---NPETIRVYISQKREIKVGDKVAGRHGNKGIISKILPRQDMPY
9_Cucumis 811 KKGGSSY---NPEMIRVYISQKREIKVGDKVAGRHGNKGIVSKILPREDMPY
11_Nicotiana 793 KRGGSSY---NPETIRVYILQKREIKVGDKVAGRHGNKGIISKILPRQDMPY
13_Syringa 793 KRGGSSY---NPETIRIYISQKREIKVGDKVAGRHGNKGIISKILPRQDMPY
18_Liquidambar 793 KKGGSSY---NPEMIRVYISQKREIKVGDKVAGRHGNKGIISKILPRQDMPY
19_Papaver 793 KKGGSSY---NPEMIRVYISQKREIKVGDKVAGRHGNKGIISKILPRQDMPY
20_Ananas 795 KKDGSSY---NSEMICVYILQKREIKVGDKVAGRHGNKGIISKILPRQDMPY
28_Liriodendron 802 KKGGSSY---NPEMIRVYISQKREIKVGDKVAGRHGNKGIISKILPRQDMPY
30_Magnolia 802 KKGGSSY---NPEMIRVYISQKREIKVGDKVAGRHGNKGIISKILPRQDMPY
32_Nymphaea 794 KKGGSID---NPEMIRVYISQKRKIKVGDKVAGRHGNKGIISKILPRQDMPY
33_Amborella 795 KKGASSY---NPEKIRVYISQKREIKVGDKVAGRHGNKGIISKILPRQDMPY
35_Picea 799 RVDDSGD---NAETVHVYISQKRKIQVGDKVAGRHGNKGIISIVLPRQDMPY
44_Ginkgo 801 EEENSGD---NAETVHVYILQKRKIQVGDKVAGRHGNKGIISKVLPRQDMPY
51_Physcomitrium 808 KKENSSN---YEKIIHVYIAQKRKIQVGDKVAGRHGNKGIISKILPRQDMPY
βa14:V823-G894
βb14:L815-G894
βb15:R900-D907 (T. Th)
T.thermophilus 861 LPDGTPVDVILNPLGVPSRMNLGQILETHLGLAGYFLGQRYISPIFDGAKEPEIKELLAQAFEVYFGKRKGEGFGVDKRE
E.coli 1088 DENGTPVDIVLNPLGVPSRMNIGQILETHLGMAAKGIGDKINAMLKQQQEVAKLREFIQRAYDLGADVRQKVDLST-FSD
0_Nostoc 842 LPDGSPVDIVLNPLGVPSRMNVGQVFECLLGWAGHTLGVRFKITPFDEMYGEESSRRIVHG-KLQE
1_Litchi 842 LQDGRPVDMVFNPLGVPSRMNVGQIFECSLGLAGGLLNRHYRIAPFDERYEQEASRKLVFS-ELYE
2_Arabidopsis 844 LQDGRPVDMVFNPLGVPSRMNVGQIFECSLGLAGSLLDRHYRIAPFDERYEQEASRKLVFS-ELYE
3_Gossypium 842 LQDGRPVDMVFNPLGVPSRMNVGQLFECSLGLAGSLLDRHYRIAPFDERYEQEASRKLVFS-ELYQ
5_Ricinus 842 LQDGRPVDMVFNPLGVPSRMNVGQIFECSLGLAGGLLDRHYRIAPFDERYEQEASRKLVFS-ELYE
6_Rosa 842 LQDGRPVDMVFNPLGVPSRMNVGQIFECSLGLAGGLLDRHYRIAPFDERYEQEASRKLVFS-ELYE
9_Cucumis 860 LQNGRPVDMVFNPLGVPSRMNVGQIFECSLGLAGSLLDRHYRIVPFDERYEQEASRKLVFS-ELYE
11_Nicotiana 842 LQDGRSVDMVFNPLGVPSRMNVGQIFECSLGLAGSLLDRHYRIAPFDERYEQEASRKLVFS-ELYE
13_Syringa 842 LQDGRPVDMVFNPLGVPSRMNVGQIFECSLGLAGGLLDRHYRIAPFDERYEQEASRKLVFS-ELYE
18_Liquidambar 842 VQDGRPVDMVFNPLGVPSRMNVGQIFECSLGLAGGLLDRHYRIAPFDERYEQEASRKLVFS-ELYE
19_Papaver 842 LQDGTPVDMVFNPLGVPSRMNVGQIFECSLGLAGDLLDRHYRIAPFDERYEQEASRKLVFS-ELYE
20_Ananas 844 LQDGTPVDMVFNPLGVPSRMNVGQIFECSLGLAGNLLKRHYRIAPFDERYEQEASRKLVFS-ELYE
28_Liriodendron 851 LQDGTPVDMVFNPLGVPSRMNVGQIFECSLGLAGDLLDRHYRIAPFDERYEQEASRKLVFS-ELYS
30_Magnolia 851 LQDGTPVDMVFNPLGVPSRMNVGQIFECSLGLAGDLLDRHYRIAPFDERYEQEASRKLVFS-ELYS
32_Nymphaea 843 LQDGTPVDMVFNPLGVPSRMNVGQMFECSLGLAGDLLGRHYRITPFDERYEQEASRKLVFS-ELYE
33_Amborella 844 LQDGTPVDMVFNPLGVPSRMNVGQIFECSLGLAGDLLDRHYRITPFDERYEQEASRKLVFP-ELYE
35_Picea 848 LQNGIPVDMVLNPLGVPSRMNVGQIFECLPGLAGNSMNKHYRITPFDERYERAASRKLVFP-ELYK
44_Ginkgo 850 SQNGTPVDMVFNPLGVPSRMNVGQIFECLPGLAGNPMNKHYRITPFGERYERAASRKLVFP-ELYR
51_Physcomitrium 857 LQDGTPIDMVLSPLGVPSRMNVGQIFECLLGLAGYFLGKHYRITPFDEKYERAASRKLVFS-ELYK

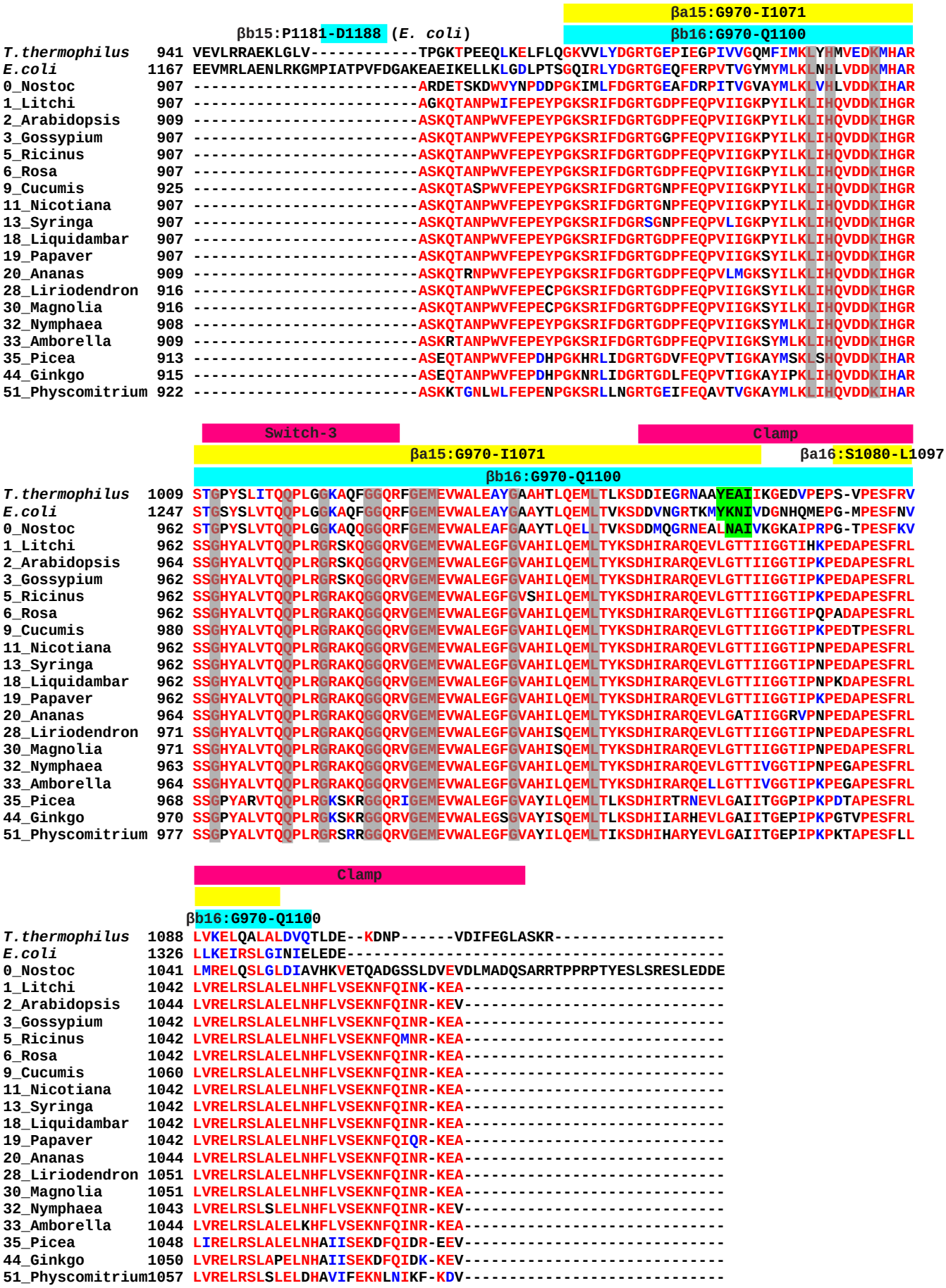
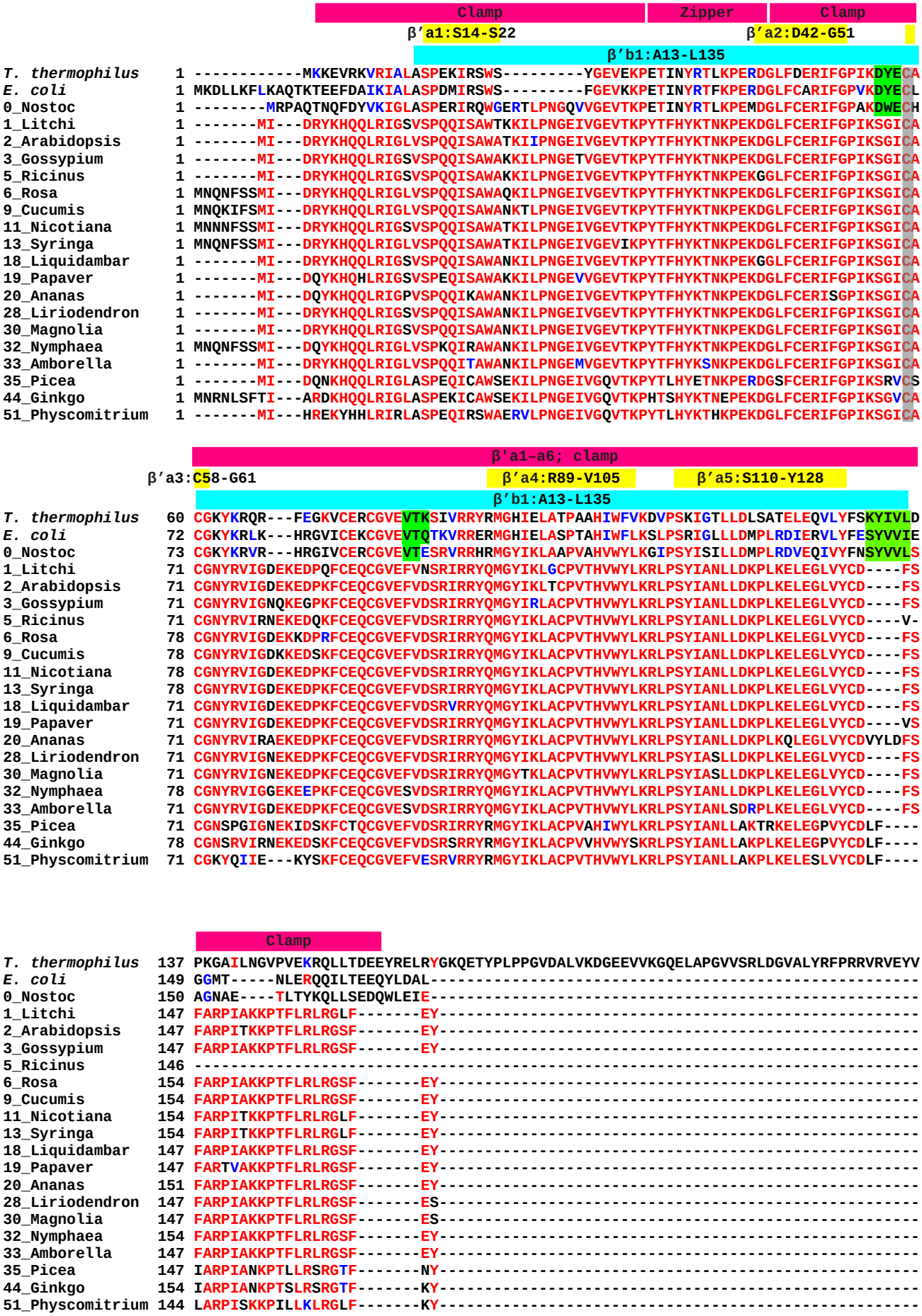
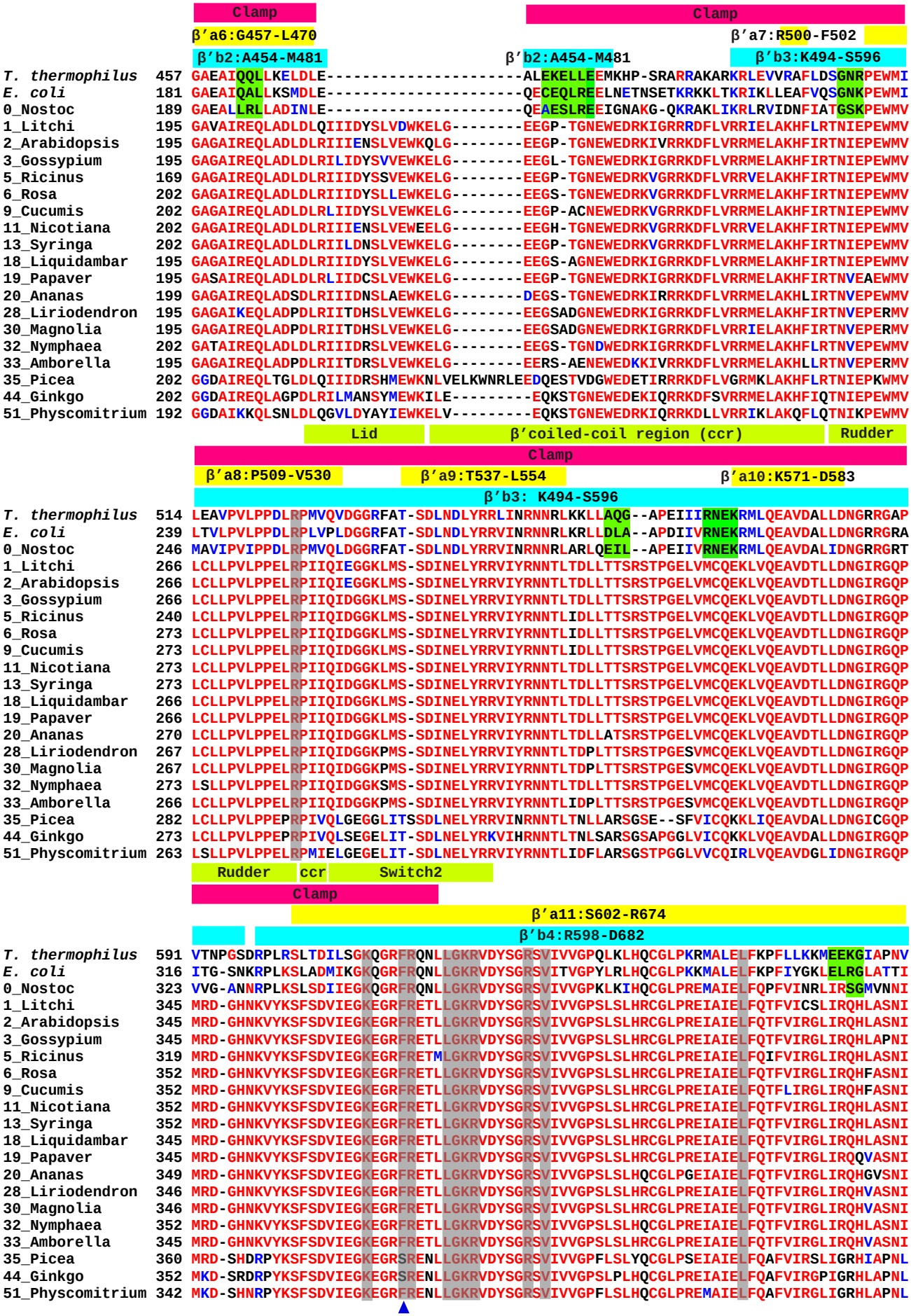
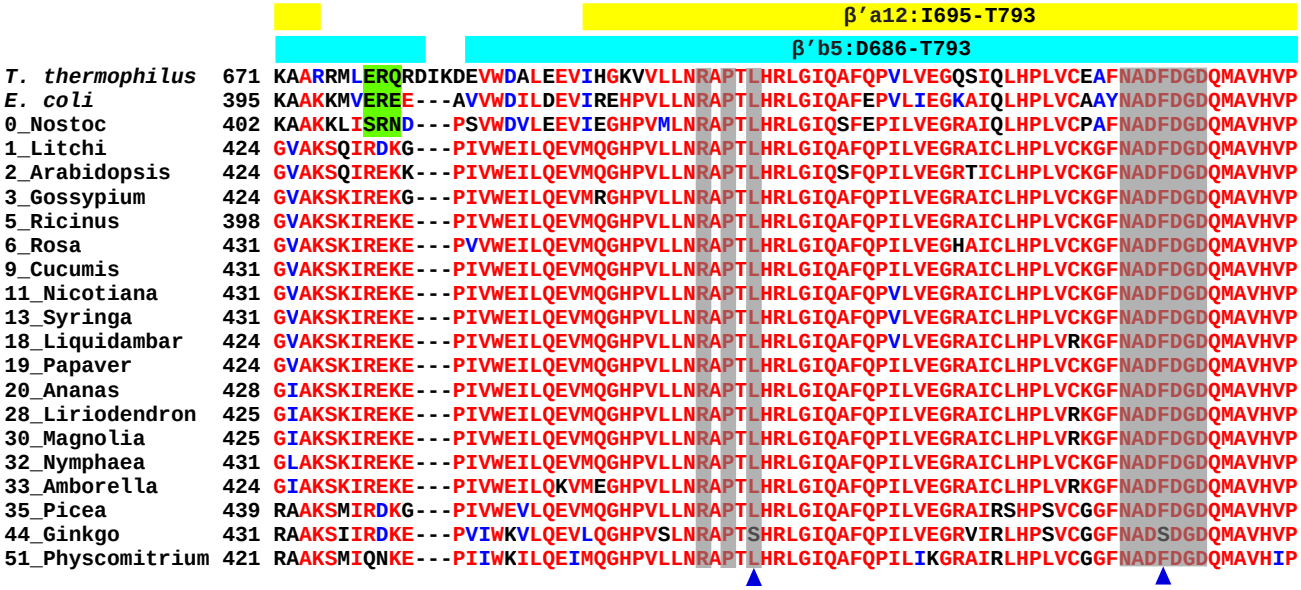

βa15:G970-I1071
βb15:P1181-D1188 (E. coli)
βb16:G970-Q1100
T.thermophilus 941 VEVLRRAEKLGLV----------TPGKTPEEQLKELFLQGKVVLYDGRTGEPIEGPIVVGQMFIMKLYHMVEDKMHAR
E.coli 1167 EEVMRLAENLRKGMPIATPVFDGAKEAEIKELLKLGDLPTSGQIRLYDGRTGEQFERPVTVGYMYMLKLNHLVDDKMHAR
0_Nostoc 907 ----------ARDETSKDWVYNPDDPGKIMLFDGRTGEAFDRPITVGVAYMLKLVHLVDDKIHAR
1_Litchi 907 ----------AGKQTANPWIFEPEYPGKSRIFDGRTGDPFEQPVIIGKPYILKLIHQVDDKIHGR
2_Arabidopsis 909 ----------ASKQTANPWVFEPEYPGKSRIFDGRTGDPFEQPVIIGKPYILKLIHQVDDKIHGR
3_Gossypium 907 ----------ASKQTANPWVFEPEYPGKSRIFDGRTGGPFEQPVIIGKPYILKLIHQVDDKIHGR
5_Ricinus 907 ----------ASKQTANPWVFEPEYPGKSRIFDGRTGDPFEQPVIIGKPYILKLIHQVDDKIHGR
6_Rosa 907 ----------ASKQTANPWVFEPEYPGKSRIFDGRTGDPFEQPVIIGKPYILKLIHQVDDKIHGR
9_Cucumis 925 ----------ASKQTASPWVFEPEYPGKSRIFDGRTGNPFEQPVIIGKPYILKLIHQVDDKIHGR
11_Nicotiana 907 ----------ASKQTANPWVFEPEYPGKSRIFDGRTGNPFEQPVIIGKPYILKLIHQVDDKIHGR
13_Syringa 907 ----------ASKQTANPWVFEPEYPGKSRIFDGRSGNPFEQPVLIGKPYILKLIHQVDDKIHGR
18_Liquidambar 907 ----------ASKQTANPWVFEPEYPGKSRIFDGRTGDPFEQPVIIGKPYILKLIHQVDDKIHGR
19_Papaver 907 ----------ASKQTANPWVFEPEYPGKSRIFDGRTGDPFEQPVIIGKSYILKLIHQVDDKIHGR
20_Ananas 909 ----------ASKQTRNPWVFEPEYPGKSRIFDGRTGDPFEQPVLMGKSYILKLIHQVDDKIHGR
28_Liriodendron 916 ----------ASKQTANPWVFEPECPGKSRIFDGRTGDPFEQPVIIGKSYILKLIHQVDDKIHGR
30_Magnolia 916 ----------ASKQTANPWVFEPECPGKSRIFDGRTGDPFEQPVIIGKSYILKLIHQVDDKIHGR
32_Nymphaea 908 ----------ASKQTANPWVFEPEYPGKSRIFDGRTGDPFEQPVIIGKSYMLKLIHQVDDKIHGR
33_Amborella 909 ----------ASKRTANPWVFEPEYPGKSRIFDGRTGDPFEQPVIIGKSYMLKLIHQVDDKIHGR
35_Picea 913 ----------ASEQTANPWVFEPDHPGKHRLIDGRTGDVFEQPVTIGKAYMSKLSHQVDDKIHAR
44_Ginkgo 915 ----------ASEQTANPWVFEPDHPGKNRLIDGRTGDLFEQPVTIGKAYIPKLIHQVDDKIHAR
51_Physcomitrium 922 ----------ASKKTGNLWLFEPENPGKSRLLNGRTGEIFEQAVTVGKAYMLKLIHQVDDKIHAR
Switch-3
Clamp
βa15:G970-I1071
βa16:S1080-L1097
βb16:G970-Q1100
T.thermophilus 1009 STGPYSLITQQPLGGKAQFGGQRFGEMEVWALEAYGAAHTLQEMLTLKSDDIEGRNAAYEAIIKGEDVPEPS-VPESFRV
E.coli 1247 STGSYSLVTQQPLGGKAQFGGQRFGEMEVWALEAYGAAYTLQEMLTVKSDDVNGRTKMYKNIVDGNHQMEPG-MPESFNV
0_Nostoc 962 STGPYSLVTQQPLGGKAQQGGQRFGEMEVWALEAFGAAYTLQELLTVKSDDMQGRNEALNAIVKGKAIPRPG-TPESFKV
1_Litchi 962 SSGHYALVTQQPLRGRSKQGGQRVGEMEVWALEGFGVAHILQEMLTYKSDHIRARQEVLGTTIIGGTIHKPEDAPESFRL
2_Arabidopsis 964 SSGHYALVTQQPLRGRSKQGGQRVGEMEVWALEGFGVAHILQEMLTYKSDHIRARQEVLGTTIIGGTIPKPEDAPESFRL
3_Gossypium 962 SSGHYALVTQQPLRGRSKQGGQRVGEMEVWALEGFGVAHILQEMLTYKSDHIRARQEVLGTTIIGGTIPKPEDAPESFRL
5_Ricinus 962 SSGHYALVTQQPLRGRAKQGGQRVGEMEVWALEGFGVSHILQEMLTYKSDHIRARQEVLGTTIIGGTIPKPEDAPESFRL
6_Rosa 962 SSGHYALVTQQPLRGRAKQGGQRVGEMEVWALEGFGVAHILQEMLTYKSDHIRARQEVLGTTIIGGTIPQPADAPESFRL
9_Cucumis 980 SSGHYALVTQQPLRGRAKQGGQRVGEMEVWALEGFGVAHILQEMLTYKSDHIRARQEVLGTTIIGGTIPKPEDTPESFRL
11_Nicotiana 962 SSGHYALVTQQPLRGRAKQGGQRVGEMEVWALEGFGVAHILQEMLTYKSDHIRARQEVLGTTIIGGTIPNPEDAPESFRL
13_Syringa 962 SSGHYALVTQQPLRGRAKQGGQRVGEMEVWALEGFGVAHILQEMLTYKSDHIRARQEVLGTTIIGGTIPNPEDAPESFRL
18_Liquidambar 962 SSGHYALVTQQPLRGRAKQGGQRVGEMEVWALEGFGVAHILQEMLTYKSDHIRARQEVLGTTIIGGTIPNPKDAPESFRL
19_Papaver 962 SSGHYALVTQQPLRGRAKQGGQRVGEMEVWALEGFGVAHILQEMLTYKSDHIRARQEVLGTTIIGGTIPKPEDAPESFRL
20_Ananas 964 SSGHYALVTQQPLRGRAKQGGQRVGEMEVWALEGFGVAHILQEMLTYKSDHIRARQEVLGATIIGGRVPNPEDAPESFRL
28_Liriodendron 971 SSGHYALVTQQPLRGRAKQGGQRVGEMEVWALEGFGVAHISQEMLTYKSDHIRARQEVLGTTIIGGTIPNPEDAPESFRL
30_Magnolia 971 SSGHYALVTQQPLRGRAKQGGQRVGEMEVWALEGFGVAHISQEMLTYKSDHIRARQEVLGTTIIGGTIPNPEDAPESFRL
32_Nymphaea 963 SSGHYALVTQQPLRGRAKQGGQRVGEMEVWALEGFGVAHILQEMLTYKSDHIRARQEVLGTTIVGGTIPNPEGAPESFRL
33_Amborella 964 SSGHYALVTQQPLRGRAKQGGQRVGEMEVWALEGFGVAHILQEMLTYKSDHIRARQELLGTTIVGGTIPKPEGAPESFRL
35_Picea 968 SSGPYARVTQQPLRGKSKRGGQRIGEMEVWALEGFGVAYILQEMLTLKSDHIRTRNEVLGAIITGGPIPKPDTAPESFRL
44_Ginkgo 970 SSGPYALVTQQPLRGKSKRGGQRVGEMEVWALEGSGVAYISQEMLTLKSDHIIARHEVLGAIITGEPIPKPGTVPESFRL
51_Physcomitrium 977 SSGPYALVTQQPLRGRSRRGGQRVGEMEVWALEGFGVAYILQEMLTIKSDHIHARYEVLGAIITGEPIPKPKTAPESFLL
Clamp
βb16:G970-Q1100
T.thermophilus 1088 LVKELQALALDVQTLDE--KDNP------VDIFEGLASKR----------
E.coli 1326 LLKEIRSLGINIELEDE----------
0_Nostoc 1041 LMRELQSLGLDIAVHKVETQADGSSLDVEVDLMADQSARRTPPRPTYESLSRESLEDDE
1_Litchi 1042 LVRELRSLALELNHFLVSEKNFQINK-KEA----------
2_Arabidopsis 1044 LVRELRSLALELNHFLVSEKNFQINR-KEV----------
3_Gossypium 1042 LVRELRSLALELNHFLVSEKNFQINR-KEA----------
5_Ricinus 1042 LVRELRSLALELNHFLVSEKNFQMNR-KEA----------
6_Rosa 1042 LVRELRSLALELNHFLVSEKNFQINR-KEA----------
9_Cucumis 1060 LVRELRSLALELNHFLVSEKNFQINR-KEA----------
11_Nicotiana 1042 LVRELRSLALELNHFLVSEKNFQINR-KEA----------
13_Syringa 1042 LVRELRSLALELNHFLVSEKNFQINR-KEA----------
18_Liquidambar 1042 LVRELRSLALELNHFLVSEKNFQINR-KEA----------
19_Papaver 1042 LVRELRSLALELNHFLVSEKNFQIQR-KEA----------
20_Ananas 1044 LVRELRSLALELNHFLVSEKNFQINR-KEA----------
28_Liriodendron 1051 LVRELRSLALELNHFLVSEKNFQINR-KEA----------
30_Magnolia 1051 LVRELRSLALELNHFLVSEKNFQINR-KEA----------
32_Nymphaea 1043 LVRELRSLSLELNHFLVSEKNFQINR-KEV----------
33_Amborella 1044 LVRELRSLALELKHFLVSEKNFQINR-KEA----------
35_Picea 1048 LIRELRSLALELNHAIISEKDFQIDR-EEV----------
44_Ginkgo 1050 LVRELRSLAPELNHAIISEKDFQIDK-KEV----------
51_Physcomitrium 1057 LVRELRSLSLELDHAVIFEKNLNIKF-KDV----------

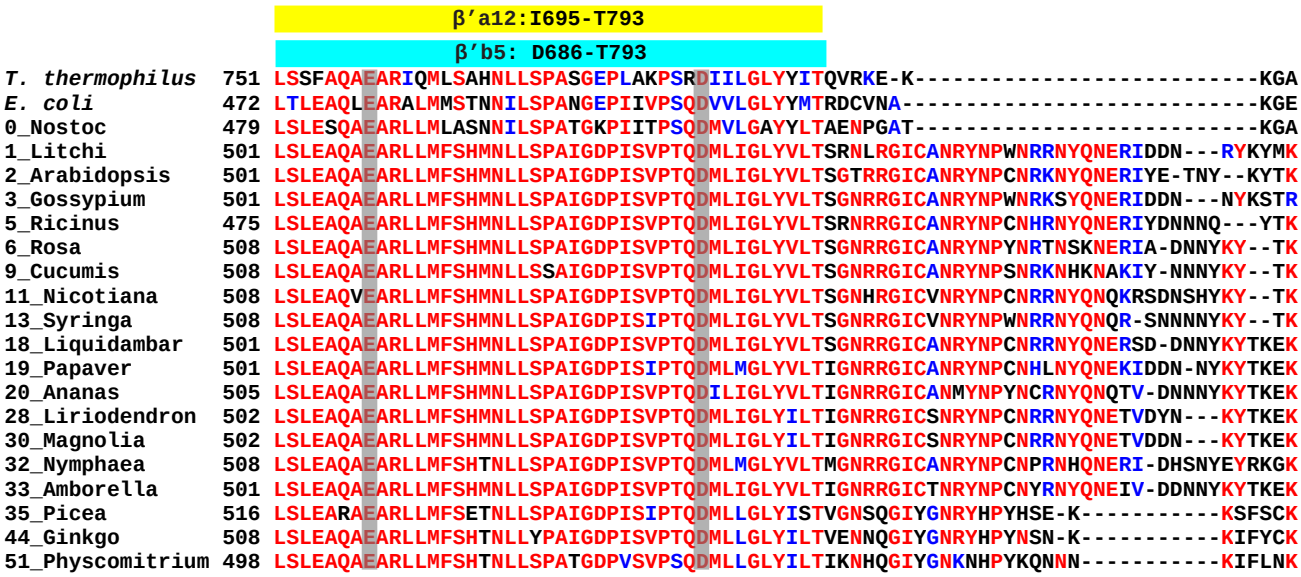
Figure S4: sequence alignment of the β' subunits from PEP of angiosperms with those of the RNAPs from *E. coli*, *T. thermophilus* and Nostoc. The residues conserved more than 50 % are in red, those mutated in similar residues are in blue. The strictly conserved residues described by Lane & Darst (Lane & Darst, 2010) are highlighted in gray. The blue triangles show mutations observed among the strictly conserved residues described (Lane & Darst, 2010). The non-conservative mutations, at least three in a row in the β or β' domain in *E. coli* and *T. thermophilus*, are highlighted in green and displayed on the *E. coli* structure (PDB entry: 6GH5). Those colored in orange are nearby to the DNA, those in green are located at the surface of the subunits. The domains described for all-RNA polymerase (a) and the bRNAPs (b) are also given and highlighted in yellow and cyan respectively. The name of the RNAP domains are also given and highlighted in purple and green (Lane & Darst, 2010; Sutherland & Murakami, 2018).

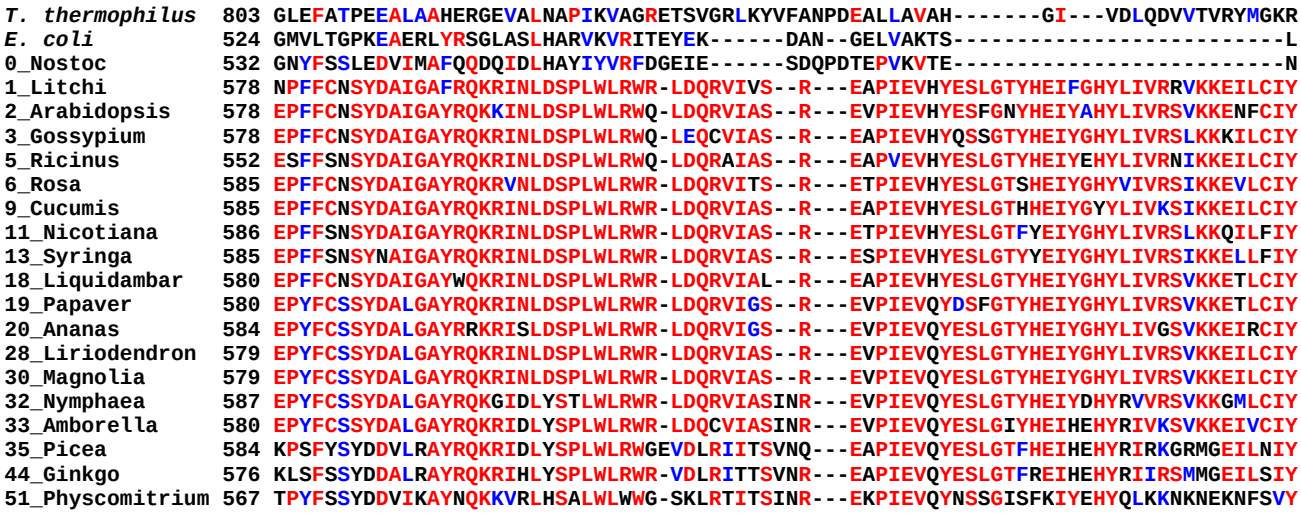

Clamp
Zipper
Clamp
β'a1:S14-S22
β'a2:D42-G51
β'b1:A13-L135
T. thermophilus 1 -----------MKKEVRKVRIALASPEKIRSWS---------YGEVEKPETINYRTLKPERDGLFDERIFGPIKDYECA
E. coli 1 MKDLLKFLKAQTKTEEFDAIKIALASPDMIRSWS---------FGEVKKPETINYRTFKPERDGLFCARIFGPVKDYECL
0_Nostoc 1 --------MRPAQTNQFDYVKIGLASPERIRQWGERTLPNGQVVGEVTKPETINYRTLKPEMDGLFCERIFGPAKDWECH
1_Litchi 1 -------MI---DRYKHQQLRIGSVSPQQISAWTKKILPNGEIVGEVTKPYTFHYKTNKPEKDGLFCERIFGPIKSGICA
2_Arabidopsis 1 -------MI---DRYKHQQLRIGLVSPQQISAWATKIIPNGEIVGEVTKPYTFHYKTNKPEKDGLFCERIFGPIKSGICA
3_Gossypium 1 -------MI---DRYKHQQLRIGSVSPQQISAWAKKILPNGETVGEVTKPYTFHYKTNKPEKDGLFCERIFGPIKSGICA
5_Ricinus 1 -------MI---DRYKHQQLRIGSVSPQQISAWAKKILPNGEIVGEVTKPYTFHYKTNKPEKGGLFCERIFGPIKSGICA
6_Rosa 1 MNQNFSSMI---DRYKHQQLRIGLVSPQQISAWAQKILPNGEIVGEVTKPYTFHYKTNKPEKDGLFCERIFGPIKSGICA
9_Cucumis 1 MNQKIFSMI---DRYKHQQLRIGLVSPQQISAWANKTLPNGEIVGEVTKPYTFHYKTNKPEKDGLFCERIFGPIKSGICA
11_Nicotiana 1 MNNNFSSMI---DRYKHQQLRIGSVSPQQISAWATKILPNGEIVGEVTKPYTFHYKTNKPEKDGLFCERIFGPIKSGICA
13_Syringa 1 MNQNFSSMI---DRYKHQQLRIGLVSPQQISAWATKILPNGEIVGEVIKPYTFHYKTNKPEKDGLFCERIFGPIKSGICA
18_Liquidambar 1 -------MI---DRYKHQQLRIGSVSPQQISAWANKILPNGEIVGEVTKPYTFHYKTNKPEKGGLFCERIFGPIKSGICA
19_Papaver 1 -------MI---DQYKHQHLRIGSVSPEQISAWAKKILPNGEVVGEVTKPYTFHYKTNKPEKDGLFCERIFGPIKSGICA
20_Ananas 1 -------MI---DQYKHQQLRIGPVSPQQIKAWANKILPNGEIVGEVTKPYTFHYKTNKPEKDGLFCERISGPIKSGICA
28_Liriodendron 1 -------MI---DRYKHQQLRIGSVSPQQISAWANKILPNGEIVGEVTKPYTFHYKTNKPEKDGLFCERIFGPIKSGICA
30_Magnolia 1 -------MI---DRYKHQQLRIGSVSPQQISAWANKILPNGEIVGEVTKPYTFHYKTNKPEKDGLFCERIFGPIKSGICA
32_Nymphaea 1 MNQNFSSMI---DQYKHQQLRIGLVSPKQIRAWANKILPNGEIVGEVTKPYTFHYKTNKPEKDGLFCERIFGPIKSGICA
33_Amborella 1 -------MI---DRYKHQQLRIGLVSPQQITAWANKILPNGEMVGEVTKPYTFHYKSNKPEKDGLFCERIFGPIKSGICA
35_Picea 1 -------MI---DQNKHQQLRIGLASPEQICAWSEKILPNGEIVGQVTKPYTLHYETNKPERDGSFCERIFGPIKSRVCS
44_Ginkgo 1 MNRNLSFTI---ARDKHQQLRIGLASPEKICAWSEKILPNGEIVGQVTKPHTSHYKTNEPEKDGLFCERIFGPIKSGVCA
51_Physcomitrium 1 -------MI---HREKYHHLRIRLASPEQIRSWAERVLPNGEIVGQVTKPYTLHYKTHKPEKDGLFCERIFGPIKSGICA
β'a1–a6; clamp
β'a3:C58-G61
β'a4:R89-V105
β'a5:S110-Y128
β'b1:A13-L135
T. thermophilus 60 CGKYKRQR---FEGKVCERCGVEVTKSIVRRYRMGHIELATPAAHIWFVKDVPSKIGTLLDLSATELEQVLYFSKYIVLD
E. coli 72 CGKYKRLK---HRGVICEKCGVEVTQTKVRRERMGHIELASPTAHIWFLKSLPSRIGLLLDMPLRDIERVLYFESYVVIE
0_Nostoc 73 CGKYKRVR---HRGIVCERCGVEVTESRVRRHRMGYIKLAAPVAHVWYLKGIPSYISILLDMPLRDVEQIVYFNSYVVLS
1_Litchi 71 CGNYRVIGDEKEDPQFCEQCGVEFVNSRIRRYQMGYIKLGCPVTHVWYLKRLPSYIANLLDKPLKELEGLVYCD----FS
2_Arabidopsis 71 CGNYRVIGDEKEDPKFCEQCGVEFVDSRIRRYQMGYIKLTCPVTHVWYLKRLPSYIANLLDKPLKELEGLVYCD----FS
3_Gossypium 71 CGNYRVIGNQKEGPKFCEQCGVEFVDSRIRRYQMGYIRLACPVTHVWYLKRLPSYIANLLDKPLKELEGLVYCD----FS
5_Ricinus 71 CGNYRVIRNEKEDQKFCEQCGVEFVDSRIRRYQMGYIKLACPVTHVWYLKRLPSYIANLLDKPLKELEGLVYCD----V-
6_Rosa 78 CGNYRVIGDEKKDPRFCEQCGVEFVDSRIRRYQMGYIKLACPVTHVWYLKRLPSYIANLLDKPLKELEGLVYCD----FS
9_Cucumis 78 CGNYRVIGDKKEDSKFCEQCGVEFVDSRIRRYQMGYIKLACPVTHVWYLKRLPSYIANLLDKPLKELEGLVYCD----FS
11_Nicotiana 78 CGNYRVIGDEKEDPKFCEQCGVEFVDSRIRRYQMGYIKLACPVTHVWYLKRLPSYIANLLDKPLKELEGLVYCD----FS
13_Syringa 78 CGNYRVIGDEKEDPKFCEQCGVEFVDSRIRRYQMGYIKLACPVTHVWYLKRLPSYIANLLDKPLKELEGLVYCD----FS
18_Liquidambar 71 CGNYRVIGDEKEDPKFCEQCGVEFVDSRVRRYQMGYIKLACPVTHVWYLKRLPSYIANLLDKPLKELEGLVYCD----FS
19_Papaver 71 CGNYRVIGDEKEDPKFCEQCGVEFVDSRIRRYQMGYIKLACPVTHVWYLKRLPSYIANLLDKPLKELEGLVYCD----VS
20_Ananas 71 CGNYRVIRAEKEDPKFCEQCGVEFVDSRIRRYQMGYIKLACPVTHVWYLKRLPSYIANLLDKPLKQLEGLVYCDVYLDFS
28_Liriodendron 71 CGNYRVIGNEKEDPKFCEQCGVEFVDSRIRRYQMGYIKLACPVTHVWYLKRLPSYIASLLDKPLKELEGLVYCD----FS
30_Magnolia 71 CGNYRVIGNEKEDPKFCEQCGVEFVDSRIRRYQMGYTKLACPVTHVWYLKRLPSYIASLLDKPLKELEGLVYCD----FS
32_Nymphaea 78 CGNYRVIGGEKEEPKFCEQCGVESVDSRIRRYQMGYIKLACPVTHVWYLKRLPSYIANLLDKPLKELEGLVYCD----FS
33_Amborella 71 CGNYRVIGDEKEDPKFCEQCGVESVDSRIRRYQMGYIKLACPVTHVWYLKRLPSYIANLSDRPLKELEGLVYCD----FS
35_Picea 71 CGNSPGIGNEKIDSKFCTQCGVEFVDSRIRRYRMGYIKLACPVAHIWYLKRLPSYIANLLAKTRKELEGPVYCDLF----
44_Ginkgo 78 CGNSRVIRNEKEDSKFCEQCGVEFVDSRSRRYRMGYIKLACPVVHVWYSKRLPSYIANLLAKPLKELEGPVYCDLF----
51_Physcomitrium 71 CGKYQIIE---KYSKFCEQCGVEFVESRVRRYRMGYIKLACPVTHVWYLKRLPSYIANLLAKPLKELESLVYCDLF----
Clamp
T. thermophilus 137 PKGAILNGVPVEKRQLLTDEEYRELRYGKQETYPLPPGVDALVKDGEEVVKGQELAPGVVSRLDGVALYRFPRRVRVEYV
E. coli 149 GGMT-----NLERQQILTEEQYLDAL----------
0_Nostoc 150 AGNAE----TLTYKQLLSEDQWLEIE----------
1_Litchi 147 FARPIAKKPTFLRLRGLF-------EY----------
2_Arabidopsis 147 FARPITKKPTFLRLRGSF-------EY----------
3_Gossypium 147 FARPIAKKPTFLRLRGSF-------EY----------
5_Ricinus 146 ----------
6_Rosa 154 FARPIAKKPTFLRLRGSF-------EY----------
9_Cucumis 154 FARPIAKKPTFLRLRGSF-------EY----------
11_Nicotiana 154 FARPITKKPTFLRLRGLF-------EY----------
13_Syringa 154 FARPITKKPTFLRLRGLF-------EY----------
18_Liquidambar 147 FARPIAKKPTFLRLRGSF-------EY----------
19_Papaver 147 FARTVAKKPTFLRLRGSF-------EY----------
20_Ananas 151 FARPIAKKPTFLRLRGSF-------EY----------
28_Liriodendron 147 FARPIAKKPTFLRLRGSF-------ES----------
30_Magnolia 147 FARPIAKKPTFLRLRGSF-------ES----------
32_Nymphaea 154 FARPIAKKPTFLRLRGSF-------EY----------
33_Amborella 147 FARPIAKKPTFLRLRGSF-------EY----------
35_Picea 147 IARPIANKPTLLRSRGTF-------NY----------
44_Ginkgo 154 IARPIANKPTSLRSRGTF-------KY----------
51_Physcomitrium 144 LARPISKKPILLKLRGLF-------KY----------

```
T. thermophilus   217 KKERAGLRLPLAAWVEKEAYKPGEILAELPEPYLFRAEEEGVVELKELEEGAFLVLRREDEPVATYFLPVGMTPLVVHGE
E. coli           170 --------------------------------------------------------------------------------
0_Nostoc          172 --------------------------------------------------------------------------------
1_Litchi          167 ---------EIQSWKYSI-------------PLFFTT--------------------------------QGFD-------
2_Arabidopsis     167 ---------EIQSWKYSI-------------PLFFTT--------------------------------QGFD-------
3_Gossypium       167 ---------EIQSWKYSI-------------PLFFTT--------------------------------QGFD-------
5_Ricinus         146 -------------KYSI-------------PLFFTA--------------------------------QGFD-------
6_Rosa            174 ---------EIQSWKYSI-------------PLFFTT--------------------------------PGFD-------
9_Cucumis         174 ---------EIQSWKYSI-------------PLFFTT--------------------------------QGFD-------
11_Nicotiana      174 ---------EIQSWKYSI-------------PLFFTT--------------------------------QGFD-------
13_Syringa        174 ---------EIQSWKYSI-------------PLFFTT--------------------------------QGFD-------
18_Liquidambar    167 ---------EIQSWKYSI-------------PLFFTT--------------------------------QGFD-------
19_Papaver        167 ---------EIQSWKYSI-------------PLFFTT--------------------------------QGFD-------
20_Ananas         171 ---------EIQSRNYSI-------------PLFFTT--------------------------------SGFE-------
28_Liriodendron   167 ---------EIQSRKYSI-------------PLFFTT--------------------------------QDFD-------
30_Magnolia       167 ---------EIQSRKYSI-------------PLFFTT--------------------------------QGFD-------
32_Nymphaea       174 ---------EIQSRKYSI-------------PLFFTT--------------------------------QGFD-------
33_Amborella      167 ---------EIQSRKYSI-------------PLFFTT--------------------------------QCFN-------
35_Picea          167 ---------EIQSWRDII-------------PHYLSAR------------------------SHYLFARGSG-------
44_Ginkgo         174 ---------DIQSWGDIL-------------PHYLSA--------------------------------QGFG-------
51_Physcomitrium  164 ---------EDQSWREIF-------------PRYFSS--------------------------------RGFE-------

T. thermophilus   297 IVEKGQPLAEAKGLLRMPRQVRAAQVEAEEEGETVYLTLFLEWTEPKDYRVQPHMNVVVPEGARVEAGDKIVAAIDPEEE
E. coli           170 ----------------------------------------------------------------------------EE
0_Nostoc          172 ----------------------------------------------------------------------------DQ
1_Litchi          186 --------------------------------------------------------------------------------
2_Arabidopsis     186 --------------------------------------------------------------------------------
3_Gossypium       186 --------------------------------------------------------------------------------
5_Ricinus         160 --------------------------------------------------------------------------------
6_Rosa            193 --------------------------------------------------------------------------------
9_Cucumis         193 --------------------------------------------------------------------------------
11_Nicotiana      193 --------------------------------------------------------------------------------
13_Syringa        193 --------------------------------------------------------------------------------
18_Liquidambar    186 --------------------------------------------------------------------------------
19_Papaver        186 --------------------------------------------------------------------------------
20_Ananas         190 --------------------------------------------------------------------------------
28_Liriodendron   186 --------------------------------------------------------------------------------
30_Magnolia       186 --------------------------------------------------------------------------------
32_Nymphaea       193 --------------------------------------------------------------------------------
33_Amborella      186 --------------------------------------------------------------------------------
35_Picea          193 --------------------------------------------------------------------------------
44_Ginkgo         193 --------------------------------------------------------------------------------
51_Physcomitrium  183 --------------------------------------------------------------------------------

T. thermophilus   377 VIAEAEGVVHLHEPASILVVKARVYPFEDDVEVSTGDRVAPGDVLADGGKVKSDVYGRVEVDLVRNVVRVVESYDIDARM
E. coli           172 ----------------------F-------------------------------------------------GDEFDAKM
0_Nostoc          174 ----------------------IYS-E-----------------------------------------DSVLQGVEVGI
1_Litchi          186 ------------------------TFR-------------------------------------------------NREIST
2_Arabidopsis     186 ------------------------IFR-------------------------------------------------NREIST
3_Gossypium       186 ------------------------TFR-------------------------------------------------SREIST
5_Ricinus         160 ------------------------TFR-------------------------------------------------NREIST
6_Rosa            193 ------------------------TFR-------------------------------------------------NREIST
9_Cucumis         193 ------------------------TFR-------------------------------------------------NREIST
11_Nicotiana      193 ------------------------TFR-------------------------------------------------NREIST
13_Syringa        193 ------------------------TFR-------------------------------------------------NREIST
18_Liquidambar    186 ------------------------TFR-------------------------------------------------NREIST
19_Papaver        186 ------------------------TFR-------------------------------------------------NREIST
20_Ananas         190 ------------------------TFR-------------------------------------------------NREIST
28_Liriodendron   186 ------------------------TFR-------------------------------------------------NREIST
30_Magnolia       186 ------------------------TFR-------------------------------------------------NREIST
32_Nymphaea       193 ------------------------TFR-------------------------------------------------NREIST
33_Amborella      186 ------------------------LFR-------------------------------------------------NREIST
35_Picea          193 ------------------------TFQ-------------------------------------------------EREIAT
44_Ginkgo         193 ------------------------AFQ-------------------------------------------------NREIAT
51_Physcomitrium  183 ------------------------AFQ-------------------------------------------------NKEIAT
```


Clamp
Clamp
β'a6:G457-L470
β'a7:R500-F502
β'b2:A454-M481
β'b2:A454-M481
β'b3:K494-S596
T. thermophilus 457 GAEAIQQLLKELDLE----------ALEKELLEEMKHP-SRARRAKARKRLEVVRAFLDSGNRPEWMI
E. coli 181 GAEAIQALLKSMDLE----------QECEQLREELNETNSETKRKKLTKRIKLLEAFVQSGNKPEWMI
0_Nostoc 189 GAEALLRLLADINLE----------QEAESLREEIGNAKG-QKRAKLIKRLRVIDNFIATGSKPEWMV
1_Litchi 195 GAVAIREQLADLDLQIIIDYSLVDWKELG--------EEGP-TGNEWEDRKIGRRRDFLVRRIELAKHFLRTNIEPEWMV
2_Arabidopsis 195 GAGAIREQLADLDLRIIIENSLVEWKQLG--------EEGP-TGNEWEDRKIVRRKDFLVRRMELAKHFIRTNIEPEWMV
3_Gossypium 195 GAGAIREQLADLDLRILIDYSVVEWKELG--------EEGL-TGNEWEDRKIGRRKDFLVRRMELAKHFIRTNIEPEWMV
5_Ricinus 169 GAGAIREQLADLDLRIIIDYSSVEWKELG--------EEGP-TGNEWEDRKVGRRKDFLVRRVELAKHFIRTNIEPEWMV
6_Rosa 202 GAGAIREQLADLDLRIIIDYSLLEWKELG--------EEGS-TGNEWEDRKVGRRKDFLVRRMELAKHFIRTNIEPEWMV
9_Cucumis 202 GAGAIREQLADLDLRLIIDYSLVEWKELG--------EEGP-ACNEWEDRKVGRRKDFLVRRMELAKHFIRTNIEPEWMV
11_Nicotiana 202 GAGAIREQLADLDLRIIIENSLVEWEELG--------EEGH-TGNEWEDRKVGRRKDFLVRRVELAKHFIRTNIEPEWMV
13_Syringa 202 GAGAIREQLADLDLRIILDNSLVEWKELG--------EEGP-TGNEWEDRKVGRRKDFLVRRMELAKHFIRTNIEPEWMV
18_Liquidambar 195 GAGAIREQLADLDLRIIIDYSLVEWKELG--------EEGS-AGNEWEDRKIGRRKDFLVRRMELAKHFIRTNIEPEWMV
19_Papaver 195 GASAIREQLADLDLRLIIDCSLVEWKELG--------EEGP-TGNEWEDRKIGRRKDFLVRRMELAKHFIRTNVEAEWMV
20_Ananas 199 GAGAIREQLADSDLRIIIDNSLAEWKELG--------DEGS-TGNEWEDRKIRRRKDFLVRRMELAKHLIRTNVEPEWMV
28_Liriodendron 195 GAGAIKEQLADPDLRIITDHSLVEWKELG--------EEGSADGNEWEDRKIGRRKDFLVRRMELAKHFIRTNVEPERMV
30_Magnolia 195 GAGAIREQLADPDLRIITDHSLVEWKELG--------EEGSADGNEWEDRKIGRRKDFLVRRIELAKHFIRTNVEPERMV
32_Nymphaea 202 GATAIREQLADLDLRIIIDRSLVEWKELG--------EEGS-TGNDWEDRKIGRRKDFLVRRMELAKHFLRTNVEPEWMV
33_Amborella 195 GAGAIREQLADPDLRIITDRSLVEWKELG--------EERS-AENEWEDKKIVRRKDFLVRRMELAKHLLRTNVEPERMV
35_Picea 202 GGDAIREQLTGLDLQIIIDRSHMEWKNLVELKWNRLEEDQESTVDGWEDETIRRRKDFLVGRMKLAKHFLRTNIEPKWMV
44_Ginkgo 202 GGDAIREQLAGPDLRILMANSYMEWKILE---------EQKSTGNEWEDEKIQRRKDFSVRRMELAKHFIQTNIEPEWMV
51_Physcomitrium 192 GGDAIKKQLSNLDLQGVLDYAYIEWKELV---------EQKSTGNEWEDRKIQRRKDLLVRRIKLAKQFLQTNIKPEWMV
Lid
β'coiled-coil region (ccr)
Rudder
Clamp
β'a8:P509-V530
β'a9:T537-L554
β'a10:K571-D583
β'b3: K494-S596
T. thermophilus 514 LEAVPVLPPDLRPMVQVDGGRFAT-SDLNDLYRRLINRNNRLKKLLAQG--APEIIIRNEKRMLQEAVDALLDNGRRGAP
E. coli 239 LTVLPVLPPDLRPLVPLDGGRFAT-SDLNDLYRRVINRNNRLKRLLDLA--APDIIVRNEKRMLQEAVDALLDNGRRGRA
0_Nostoc 246 MAVIPVIPPDLRPMVQLDGGRFAT-SDLNDLYRRVINRNNRLARLQEIL--APEIIVRNEKRMLQEAVDALIDNGRRGRT
1_Litchi 266 LCLLPVLPPELRPIIQIEGGKLMS-SDINELYRRVIYRNNTLTDLLTTSRSTPGELVMCQEKLVQEAVDTLLDNGIRGQP
2_Arabidopsis 266 LCLLPVLPPELRPIIQIEGGKLMS-SDINELYRRVIYRNNTLTDLLTTSRSTPGELVMCQEKLVQEAVDTLLDNGIRGQP
3_Gossypium 266 LCLLPVLPPELRPIIQIDGGKLMS-SDINELYRRVIYRNNTLTDLLTTSRSTPGELVMCQEKLVQEAVDTLLDNGIRGQP
5_Ricinus 240 LCLLPVLPPELRPIIQIDGGKLMS-SDINELYRRVIYRNNTLIDLLTTSRSTPGELVMCQEKLVQEAVDTLLDNGIRGQP
6_Rosa 273 LCLLPVLPPELRPIIQIDGGKLMS-SDINELYRRVIYRNNTLIDLLTTSRSTPGELVMCQEKLVQEAVDTLLDNGIRGQP
9_Cucumis 273 LCLLPVLPPELRPIIQIDGGKLMS-SDINELYRRVIYRNNTLIDLLTTSRSTPGELVMCQEKLVQEAVDTLLDNGIRGQP
11_Nicotiana 273 LCLLPVLPPELRPIIQIDGGKLMS-SDINELYRRVIYRNNTLTDLLTTSRSTPGELVMCQEKLVQEAVDTLLDNGIRGQP
13_Syringa 273 LCLLPVLPPELRPIIQIDGGKLMS-SDINELYRRVIYRNNTLTDLLTTSRSTPGELVMCQEKLVQEAVDTLLDNGIRGQP
18_Liquidambar 266 LCLLPVLPPELRPIIQIDGGKLMS-SDINELYRRVIYRNNTLTDLLTTSRSTPGELVMCQEKLVQEAVDTLLDNGIRGQP
19_Papaver 266 LCLLPVLPPELRPIIQIDGGKLMS-SDINELYRRVIYRNNTLTDLLTTSRSTPGELVMCQEKLVQEAVDTLLDNGIRGQP
20_Ananas 270 LCLLPVLPPELRPIIQIDGGKLMS-SDINELYRRVIYRNNTLTDLLATSRSTPGELVMCQEKLVQEAVDTLLDNGIRGQP
28_Liriodendron 267 LCLLPVLPPELRPIIQIDGGKPMS-SDINELYRRVIYRNNTLTDPLTTSRSTPGESVMCQEKLVQEAVDTLLDNGIRGQP
30_Magnolia 267 LCLLPVLPPELRPIIQIDGGKPMS-SDINELYRRVIYRNNTLTDPLTTSRSTPGESVMCQEKLVQEAVDTLLDNGIRGQP
32_Nymphaea 273 LSLLPVLPPELRPIIQIDGGKSMS-SDINELYRRVIYRNNTLTDLLTTSRSTPGELVMCQEKLVQEAVDTLLDNGIRGQP
33_Amborella 266 LCLLPVLPPELRPIIQIDGGKPMS-SDINELYRRVIYRNNTLIDPLTTSRSTPGESVMCQEKLVQEAVDTLLDNGIRGQP
35_Picea 282 LCLLPVLPPEPRPIVQLGEGGLITSSDLNELYRRVINRNNTLTNLLARSGSE--SFVICQKKLIQEAVDALLDNGICGQP
44_Ginkgo 273 LCLLPVLPPEPRPIVQLSEGELIT-SDLNELYRKVIHRNNTLTNLSARSGSAPGGLVICQKKLVQEAVDALLDNGIRGQP
51_Physcomitrium 263 LSLLPVLPPELRPMIELGEGELIT-SDLNELYRRVIYRNNTLIDFLARSGSTPGGLVVCQIRLVQEAVDGLIDNGIRGQP
Rudder
ccr
Switch2
Clamp
β'a11:S602-R674
β'b4:R598-D682
T. thermophilus 591 VTNPGSDRPLRSLTDILSGKQGRFRQNLLGKRVDYSGRSVIVVGPQLKLHQCGLPKRMALELFKPFLLKKMEEKGIAPNV
E. coli 316 ITG-SNKRPLKSLADMIKGKQGRFRQNLLGKRVDYSGRSVITVGPYLRLHQCGLPKKMALELFKPFIYGKLELRGLATTI
0_Nostoc 323 VVG-ANNRPLKSLSDIIEGKQGRFRQNLLGKRVDYSGRSVIVVGPKLKIHQCGLPREMAIELFQPFVINRLIRSGMVNNI
1_Litchi 345 MRD-GHNKVYKSFSDVIEGKEGRFRETLLGKRVDYSGRSVIVVGPSLSLHRCGLPREIAIELFQTFVICSLIRQHLASNI
2_Arabidopsis 345 MRD-GHNKVYKSFSDVIEGKEGRFRETLLGKRVDYSGRSVIVVGPSLSLHRCGLPREIAIELFQTFVIRGLIRQHLASNI
3_Gossypium 345 MRD-GHNKVYKSFSDVIEGKEGRFRETLLGKRVDYSGRSVIVVGPSLSLHRCGLPREIAIELFQTFVIRGLIRQHLAPNI
5_Ricinus 319 MRD-GHNKVYKSFSDVIEGKEGRFRETMLGKRVDYSGRSVIVVGPSLSLHRCGLPREIAIELFQIFVIRGLIRQHLASNI
6_Rosa 352 MRD-GHNKVYKSFSDVIEGKEGRFRETLLGKRVDYSGRSVIVVGPSLSLHRCGLPREIAIELFQTFVIRGLIRQHFASNI
9_Cucumis 352 MRD-GHNKVYKSFSDVIEGKEGRFRETLLGKRVDYSGRSVIVVGPSLSLHRCGLPREIAIELFQTFLIRGLIRQHFASNI
11_Nicotiana 352 MRD-GHNKVYKSFSDVIEGKEGRFRETLLGKRVDYSGRSVIVVGPSLSLHRCGLPREIAIELFQTFVIRGLIRQHLASNI
13_Syringa 352 MRD-GHNKVYKSFSDVIEGKEGRFRETLLGKRVDYSGRSVIVVGPSLSLHRCGLPREIAIELFQTFVIRGLIRQHLASNI
18_Liquidambar 345 MRD-GHNKVYKSFSDVIEGKEGRFRETLLGKRVDYSGRSVIVVGPSLSLHRCGLPREIAIELFQTFVIRGLIRQHLASNI
19_Papaver 345 MRD-GHNKVYKSFSDVIEGKEGRFRETLLGKRVDYSGRSVIVVGPSLSLHRCGLPREIAIELFQTFVIRGLIRQQVASNI
20_Ananas 349 MRD-GHNKVYKSFSDVIEGKEGRFRETLLGKRVDYSGRSVIVVGPSLSLHQCGLPGEIAIELFQTFVIRGLIRQHGVSNI
28_Liriodendron 346 MRD-GHNKVYKSFSDVIEGKEGRFRETLLGKRVDYSGRSVIVVGPSLSLHRCGLPREIAIELFQTFVIRGLIRQHVASNI
30_Magnolia 346 MRD-GHNKVYKSFSDVIEGKEGRFRETLLGKRVDYSGRSVIVVGPSLSLHRCGLPREIAIELFQTFVIRGLIRQHVASNI
32_Nymphaea 352 MRD-GHNKVYKSFSDVIEGKEGRFRETLLGKRVDYSGRSVIVVGPSLSLHQCGLPREIAIELFQTFVIRGLIRQHLASNI
33_Amborella 345 MRD-GHNKVYKSFSDVIEGKEGRFRETLLGKRVDYSGRSVIVVGPSLSLHRCGLPREIAIELFQTFVIRGLIRQHVASNI
35_Picea 360 MRD-SHDRPYKSFSDVIEGKEGRSRENLLGKRVDYSGRSVIVVGPFLSLYQCGLPSEIAIELFQAFVIRSLIGRHIAPNL
44_Ginkgo 352 MKD-SRDRPYKSFSDVIEGKEGRSRENLLGKRVDYSGRSVIVVGPSLPLHQCGLPREIAIELFQAFVIRGPIGRHLAPNL
51_Physcomitrium 342 MKD-SHNRPYKSFSDVIEGKEGRFRENLLGKRVDYSGRSVIVVGPFLSLHQCGLPREMAIELFQAFVIRGLIGRHLAPNL


β'a12:I695-T793
β'b5:D686-T793
T. thermophilus 671 KAARRMLERQRDIKDEVWDALEEVIHGKVVLLNRAPTLHRLGIQAFQPVLVEGQSIQLHPLVCEAFNADFDGDQMAVHVP
E. coli 395 KAAKKMVEREE---AVVWDILDEVIREHPVLLNRAPTLHRLGIQAFEPVLIEGKAIQLHPLVCAAYNADFDGDQMAVHVP
0_Nostoc 402 KAAKKLISRND---PSVWDVLEEVIEGHPVMLNRAPTLHRLGIQSFEPILVEGRAIQLHPLVCPAFNADFDGDQMAVHVP
1_Litchi 424 GVAKSQIRDKG---PIVWEILQEVMQGHPVLLNRAPTLHRLGIQAFQPILVEGRAICLHPLVCKGFNADFDGDQMAVHVP
2_Arabidopsis 424 GVAKSQIREKK---PIVWEILQEVMQGHPVLLNRAPTLHRLGIQSFQPILVEGRTICLHPLVCKGFNADFDGDQMAVHVP
3_Gossypium 424 GVAKSKIREKG---PIVWEILQEVMRGHPVLLNRAPTLHRLGIQAFQPILVEGRAICLHPLVCKGFNADFDGDQMAVHVP
5_Ricinus 398 GVAKSKIREKE---PIVWEILQEVMQGHPVLLNRAPTLHRLGIQAFQPILVEGRAICLHPLVCKGFNADFDGDQMAVHVP
6_Rosa 431 GVAKSKIREKE---PVVWEILQEVMQGHPVLLNRAPTLHRLGIQAFQPILVEGHAICLHPLVCKGFNADFDGDQMAVHVP
9_Cucumis 431 GVAKSKIREKE---PIVWEILQEVMQGHPVLLNRAPTLHRLGIQAFQPILVEGRAICLHPLVCKGFNADFDGDQMAVHVP
11_Nicotiana 431 GVAKSKIREKE---PIVWEILQEVMQGHPVLLNRAPTLHRLGIQAFQPVLVEGRAICLHPLVCKGFNADFDGDQMAVHVP
13_Syringa 431 GVAKSKIREKE---PIVWEILQEVMQGHPVLLNRAPTLHRLGIQAFQPVLVEGRAICLHPLVCKGFNADFDGDQMAVHVP
18_Liquidambar 424 GVAKSKIREKE---PIVWEILQEVMQGHPVLLNRAPTLHRLGIQAFQPVLVEGRAICLHPLVRKGFNADFDGDQMAVHVP
19_Papaver 424 GVAKSKIREKE---PIVWEILQEVMQGHPVLLNRAPTLHRLGIQAFQPILVEGRAICLHPLVCKGFNADFDGDQMAVHVP
20_Ananas 428 GIAKSKIREKE---PIVWEILQEVMQGHPVLLNRAPTLHRLGIQAFQPILVEGRAICLHPLVCKGFNADFDGDQMAVHVP
28_Liriodendron 425 GIAKSKIREKE---PIVWEILQEVMQGHPVLLNRAPTLHRLGIQAFQPILVEGRAICLHPLVRKGFNADFDGDQMAVHVP
30_Magnolia 425 GIAKSKIREKE---PIVWEILQEVMQGHPVLLNRAPTLHRLGIQAFQPILVEGRAICLHPLVRKGFNADFDGDQMAVHVP
32_Nymphaea 431 GLAKSKIREKE---PIVWEILQEVMQGHPVLLNRAPTLHRLGIQAFQPILVEGRAICLHPLVCKGFNADFDGDQMAVHVP
33_Amborella 424 GIAKSKIREKE---PIVWEILQKVMEGHPVLLNRAPTLHRLGIQAFQPILVEGRAICLHPLVRKGFNADFDGDQMAVHVP
35_Picea 439 RAAKSMIRDKG---PIVWEVLQEVMQGHPVLLNRAPTLHRLGIQAFQPILVEGRAIRSHPSVCGGFNADFDGDQMAVHVP
44_Ginkgo 431 RAAKSIIRDKE---PVIWKVLQEVLQGHPVSLNRAPTSHRLGIQAFQPILVEGRVIRLHPSVCGGFNADSDGDQMAVHVP
51_Physcomitrium 421 RAAKSMIQNKE---PIIWKILQEIMQGHPVLLNRAPTLHRLGIQAFQPILIKGRAIRLHPLVCGGFNADFDGDQMAVHIP

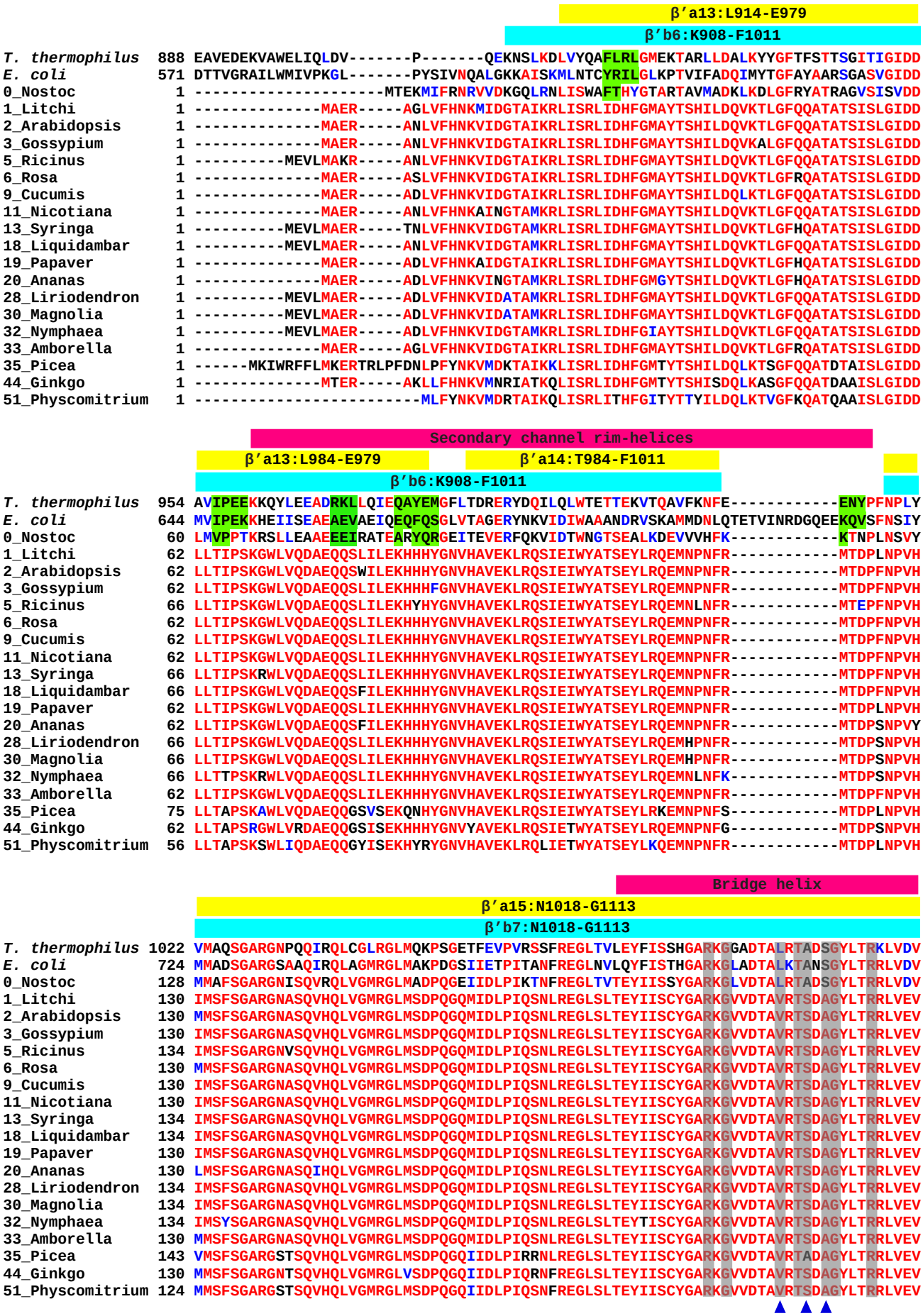

β'a12:I695-T793
β'b5: D686-T793
T. thermophilus 751 LSSFAQAEARIQMLSAHNLLSPASGEPLAKPSRDIILGLYYITQVRKE-K----------KGA
E. coli 472 LTLEAQLEARALMMSTNNILSPANGEPIIVPSQDVVLGLYYMTRDCVNA----------KGE
0_Nostoc 479 LSLESQAEARLLMLASNNILSPATGKPIITPSQDMVLGAYYLTAENPGAT----------KGA
1_Litchi 501 LSLEAQAEARLLMFSHMNLLSPAIGDPISVPTQDMLIGLYVLTSRNLRGICANRYNPWNRRNYQNERIDDN---RYKYMK
2_Arabidopsis 501 LSLEAQAEARLLMFSHMNLLSPAIGDPISVPTQDMLIGLYVLTSGTRRGICANRYNPCNRKNYQNERIYE-TNY--KYTK
3_Gossypium 501 LSLEAQAEARLLMFSHMNLLSPAIGDPISVPTQDMLIGLYVLTSGNRRGICANRYNPWNRKSYQNERIDDN---NYKSTR
5_Ricinus 475 LSLEAQAEARLLMFSHMNLLSPAIGDPISVPTQDMLIGLYVLTSRNRRGICANRYNPCNHRNYQNERIYDNNNQ---YTK
6_Rosa 508 LSLEAQAEARLLMFSHMNLLSPAIGDPISVPTQDMLIGLYVLTSGNRRGICANRYNPYNRTNSKNERIA-DNNYKY--TK
9_Cucumis 508 LSLEAQAEARLLMFSHMNLLSSAIGDPISVPTQDMLIGLYVLTSGNRRGICANRYNPSNRKNHKNAKIY-NNNYKY--TK
11_Nicotiana 508 LSLEAQVEARLLMFSHMNLLSPAIGDPISVPTQDMLIGLYVLTSGNHRGICVNRYNPCNRRNYQNQKRSDNSHYKY--TK
13_Syringa 508 LSLEAQAEARLLMFSHMNLLSPAIGDPISIPTQDMLIGLYVLTSGNRRGICVNRYNPWNRRNYQNQR-SNNNNYKY--TK
18_Liquidambar 501 LSLEAQAEARLLMFSHMNLLSPAIGDPISVPTQDMLIGLYVLTSGNRRGICANRYNPCNRRNYQNERSD-DNNYKYTKEK
19_Papaver 501 LSLEAQAEARLLMFSHMNLLSPAIGDPISIPTQDMLMGLYVLTIGNRRGICANRYNPCNHLNYQNEKIDDN-NYKYTKEK
20_Ananas 505 LSLEAQAEARLLMFSHMNLLSPAIGDPISVPTQDILIGLYVLTIGNRRGICANMYNPYNCRNYQNQTV-DNNNYKYTKEK
28_Liriodendron 502 LSLEAQAEARLLMFSHMNLLSPAIGDPISVPTQDMLIGLYILTIGNRRGICSNRYNPCNRRNYQNETVDYN---KYTKEK
30_Magnolia 502 LSLEAQAEARLLMFSHMNLLSPAIGDPISVPTQDMLIGLYILTIGNRRGICSNRYNPCNRRNYQNETVDDN---KYTKEK
32_Nymphaea 508 LSLEAQAEARLLMFSHTNLLSPAIGDPISVPTQDMLMGLYVLTMGNRRGICANRYNPCNPRNHQNERI-DHSNYEYRKGK
33_Amborella 501 LSLEAQAEARLLMFSHMNLLSPAIGDPISVPTQDMLIGLYVLTIGNRRGICTNRYNPCNYRNYQNEIV-DDNNYKYTKEK
35_Picea 516 LSLEARAEARLLMFSETNLLSPAIGDPISIPTQDMLLGLYISTVGNSQGIYGNRYHPYHSE-K----------KSFSCK
44_Ginkgo 508 LSLEAQAEARLLMFSHTNLLYPAIGDPISVPTQDMLLGLYILTVENNQGIYGNRYHPYNSN-K----------KIFYCK
51_Physcomitrium 498 LSLEAQAEARLLMFSHTNLLSPATGDPVSVPSQDMLLGLYILTIKNHQGIYGNKNHPYKQNNN----------KIFLNK

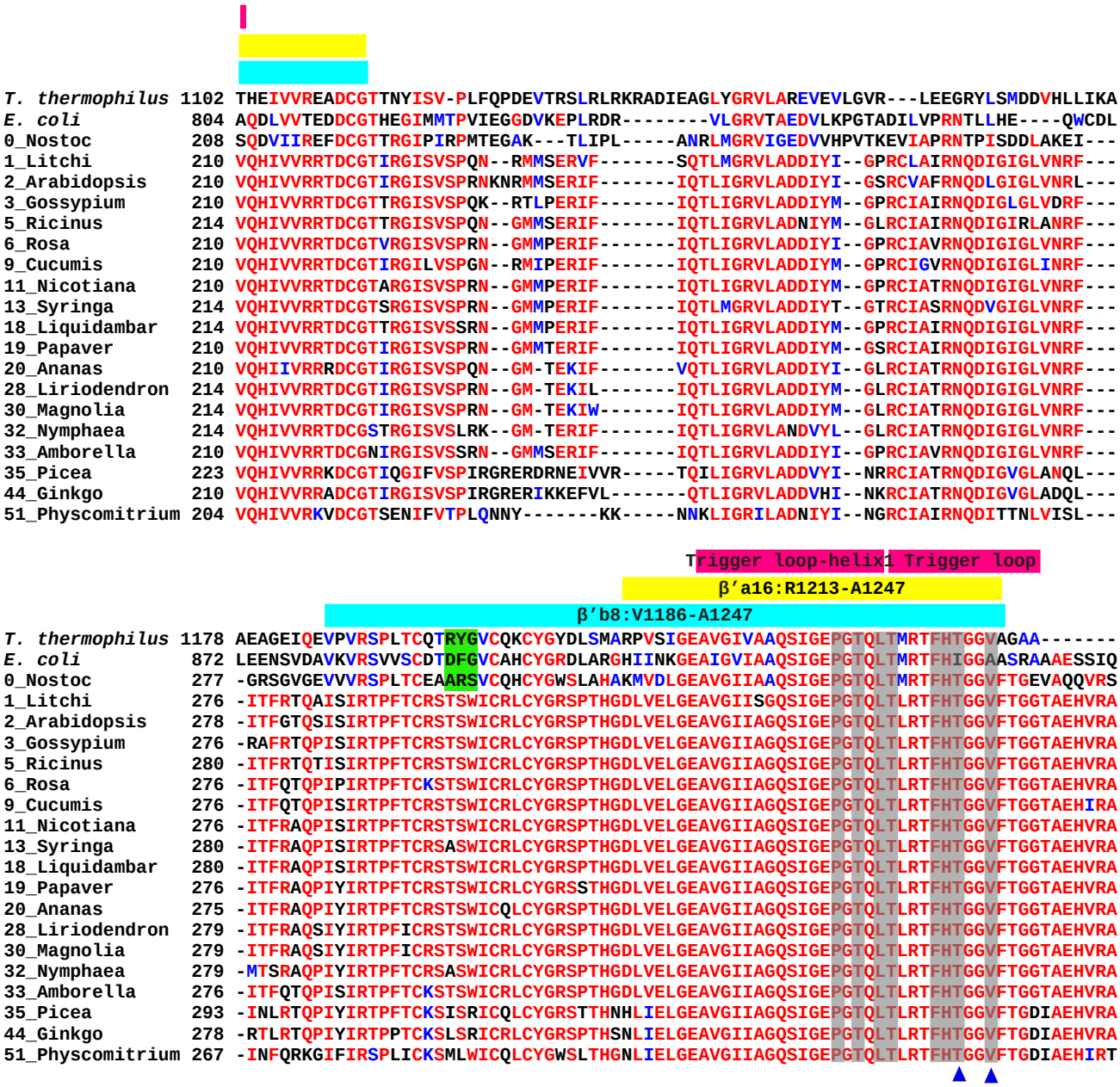

T. thermophilus 803 GLEFATPEEALAAHERGEVALNAPIKVAGRETSVGRLKYVFANPDEALLAVAH-------GI---VDLQDVVTVRYMGKR
E. coli 524 GMVLTGPKEAERLYRSGLASLHARVKVRITEYEK------DAN--GELVAKTS----------L
0_Nostoc 532 GNYFSSLEDVIMAFQQDQIDLHAYIYVRFDGEIE------SDQPDTEPVKVTE----------N
1_Litchi 578 NPFFCNSYDAIGAFRQKRINLDSPLWLRWR-LDQRVIVS--R---EAPIEVHYESLGTYHEIFGHYLIVRRVKKEILCIY
2_Arabidopsis 578 EPFFCNSYDAIGAYRQKKINLDSPLWLRWQ-LDQRVIAS--R---EVPIEVHYESFGNYHEIYAHYLIVRSVKKENFCIY
3_Gossypium 578 EPFFCNSYDAIGAYRQKRINLDSPLWLRWQ-LEQCVIAS--R---EAPIEVHYQSSGTYHEIYGHYLIVRSLKKKILCIY
5_Ricinus 552 ESFFSNSYDAIGAYRQKRINLDSPLWLRWQ-LDQRAIAS--R---EAPVEVHYESLGTYHEIYEHYLIVRNIKKEILCIY
6_Rosa 585 EPFFCNSYDAIGAYRQKRVNLDSPLWLRWR-LDQRVITS--R---ETPIEVHYESLGTSHEIYGHYVIVRSIKKEVLCIY
9_Cucumis 585 EPFFCNSYDAIGAYRQKRINLDSPLWLRWR-LDQRVIAS--R---EAPIEVHYESLGTHHEIYGYYLIVKSIKKEILCIY
11_Nicotiana 586 EPFFSNSYDAIGAYRQKRINLDSPLWLRWR-LDQRVIAS--R---ETPIEVHYESLGTFYEIYGHYLIVRSLKKQILFIY
13_Syringa 585 EPFFSNSYNAIGAYRQKRINLDSPLWLRWR-LDQRVIAS--R---ESPIEVHYESLGTYYEIYGHYLIVRSIKKELLFIY
18_Liquidambar 580 EPFFCNSYDAIGAYWQKRINLDSPLWLRWR-LDQRVIAL--R---EAPIEVHYESLGTYHEIYGHYLIVRSVKKETLCIY
19_Papaver 580 EPYFCSSYDALGAYRQKRINLDSPLWLRWR-LDQRVIGS--R---EVPIEVQYDSFGTYHEIYGHYLIVRSVKKETLCIY
20_Ananas 584 EPYFCSSYDALGAYRRKRISLDSPLWLRWR-LDQRVIGS--R---EVPIEVQYESLGTYHEIYGHYLIVGSVKKEIRCIY
28_Liriodendron 579 EPYFCSSYDALGAYRQKRINLDSPLWLRWR-LDQRVIAS--R---EVPIEVQYESLGTYHEIYGHYLIVRSVKKEILCIY
30_Magnolia 579 EPYFCSSYDALGAYRQKRINLDSPLWLRWR-LDQRVIAS--R---EVPIEVQYESLGTYHEIYGHYLIVRSVKKEILCIY
32_Nymphaea 587 EPYFCSSYDALGAYRQKGIDLYSTLWLRWR-LDQRVIASINR---EVPIEVQYESLGTYHEIYDHYRVVRSVKKGMLCIY
33_Amborella 580 EPYFCSSYDALGAYRQKRIDLYSPLWLRWR-LDQCVIASINR---EVPIEVQYESLGIYHEIHEHYRIVKSVKKEIVCIY
35_Picea 584 KPSFYSYDDVLRAYRQKRIDLYSPLWLRWGEVDLRIITSVNQ---EAPIEVQYESLGTFHEIHEHYRIRKGRMGEILNIY
44_Ginkgo 576 KLSFSSYDDALRAYRQKRIHLYSPLWLRWR-VDLRITTSVNR---EAPIEVQYESLGTFREIHEHYRIIRSMMGEILSIY
51_Physcomitrium 567 TPYFSSYDDVIKAYNQKKVRLHSALWLWWG-SKLRTITSINR---EKPIEVQYNSSGISFKIYEHYQLKKNKNEKNFSVY

```
T. thermophilus   873 LETSPGRILFARIVAEAVEDEKVAWELIQL-------DVPQEKNSLKDLVYQAFLRLGMEKTARLLDALKYYGFTFSTTS
E. coli           570 KDTTVGR-------------AILWMIVPKGLPYSIVNQALGKKAISKMLNTCYRILGLKPTVIFADQIMYTGFAYAARS
0_Nostoc          580 --EDGTR-------------TLLYKFR-------RVRQDAKGNVLSQYI---YT---TPGRVIYNNAIQ-EALAS----
1_Litchi          652 IRTTVGHISLYREIEEAIQGFCRACSYGT-------------------------------------------------
2_Arabidopsis     652 IRTTVGHISFYREIEEAIQGFSQACSYDT-------------------------------------------------
3_Gossypium       652 IRTTVGHISLYREIEEAIQGFFRAYSYDTQSYGI-------------------------------------------
5_Ricinus         626 IRTTVGHISLYREIEEAIQGFCQAGSDGI-----------------------------------------------
6_Rosa            659 VRTTVGHISLYREIEEAIQGFCRAYSYGT-------------------------------------------------
9_Cucumis         659 IRTTVGHISLYREIEEAIQGFCRACSYGT-------------------------------------------------
11_Nicotiana      660 IRTTVGHIALYREIEEAIQGFSRAYSSGT-------------------------------------------------
13_Syringa        659 IRTTVGHISLYREIEEAVQGFSQACSYGTELS----------------------------------------------
18_Liquidambar    654 IRTTVGHISLYREIEEALQGFYRACSYRT-------------------------------------------------
19_Papaver        654 IRTTVGHISFYREIEEAIQGFSQACSYDT-------------------------------------------------
20_Ananas         658 IRTTLGHISFYREIEEAIQGFCRAYSYTI-------------------------------------------------
28_Liriodendron   653 IRTTVGHISFYREIEEAIQGFCRAYLYDT-------------------------------------------------
30_Magnolia       653 IRTTVGHISFYREIEEAIQGFCRAYSYDT-------------------------------------------------
32_Nymphaea       663 IRTTVGHISFYREIEEAVQGFCRSYSYGT-------------------------------------------------
33_Amborella      656 IRTTVGHISFYREIEEAIQGFCRTY-----------------------------------------------------
35_Picea          661 IRTTVGRTRFNREMEEAIQGFAR-SEHPKKSLPALRI-----------------------------------------
44_Ginkgo         652 IRTTVGRIRFNREIEEAIQGFSRASEHPNKSLKAIRI-----------------------------------------
51_Physcomitrium  643 ICTTVGRIIFNQQIEEAIQGTLKASLFRNQSLPAITI-----------------------------------------
```

```
T. thermophilus   946 GITIGIDDAVIPEEKKQYLEEADRKLLQIEQAYEMGFLTD
E. coli           636 GASVGIDDMVIPEKKHEIISEAEAEVAEIQEQFQSGLVTA
0_Nostoc              ----------------------------------------
1_Litchi              ----------------------------------------
2_Arabidopsis         ----------------------------------------
3_Gossypium           ----------------------------------------
5_Ricinus             ----------------------------------------
6_Rosa                ----------------------------------------
9_Cucumis             ----------------------------------------
11_Nicotiana          ----------------------------------------
13_Syringa            ----------------------------------------
18_Liquidambar        ----------------------------------------
19_Papaver            ----------------------------------------
20_Ananas             ----------------------------------------
28_Liriodendron       ----------------------------------------
30_Magnolia           ----------------------------------------
32_Nymphaea           ----------------------------------------
33_Amborella          ----------------------------------------
35_Picea              ----------------------------------------
44_Ginkgo             ----------------------------------------
51_Physcomitrium      ----------------------------------------
```

Figure S5 : sequence alignment of the β" subunits from PEP of angiosperms with those of the RNAPs from *E. coli*, *T. thermophilus* and Nostoc. The residues conserved more than 50 % are in red, those mutated in similar residues are in blue. The strictly conserved residues described by Lane & Darst (Lane & Darst, 2010) are highlighted in gray. The blue triangles show mutations observed among the strictly conserved residues described (Lane & Darst, 2010). The non-conservative mutations, at least three in a row in the β or β' domain in *E. coli* and *T. thermophilus*, are highlighted in green and displayed on the *E. coli* structure (PDB entry: 6GH5). Those colored in orange are nearby to the DNA, those in green are located at the surface of the subunits. The domains described for all-RNA polymerase (a) and the bRNAPs (b) are also given and highlighted in yellow and cyan respectively. The name of the RNAP domains are also given and highlighted in purple and green (Lane & Darst, 2010; Sutherland & Murakami, 2018).


β'a13:L914-E979
β'b6:K908-F1011
T. thermophilus 888 EAVEDEKVAWELIQLDV-------P-------QEKNSLKDLVYQAFLRLGMEKTARLLDALKYYGFTFSTTSGITIGIDD
E. coli 571 DTTVGRAILWMIVPKGL-------PYSIVNQALGKKAISKMLNTCYRILGLKPTVIFADQIMYTGFAYAARSGASVGIDD
0_Nostoc 1 ----------MTEKMIFRNRVVDKGQLRNLISWAFTHYGTARTAVMADKLKDLGFRYATRAGVSISVDD
1_Litchi 1 ----------MAER-----AGLVFHNKMIDGTAIKRLISRLIDHFGMAYTSHILDQVKTLGFQQATATSISLGIDD
2_Arabidopsis 1 ----------MAER-----ANLVFHNKVIDGTAIKRLISRLIDHFGMAYTSHILDQVKTLGFQQATATSISLGIDD
3_Gossypium 1 ----------MAER-----ANLVFHNKVIDGTAIKRLISRLIDHFGMAYTSHILDQVKALGFQQATATSISLGIDD
5_Ricinus 1 ----------MEVLMAKR-----ANLVFHNKVIDGTAIKRLISRLIDHFGMAYTSHILDQVKTLGFQQATATSISLGIDD
6_Rosa 1 ----------MAER-----ASLVFHNKVIDGTAIKRLISRLIDHFGMAYTSHILDQVKTLGFRQATATSISLGIDD
9_Cucumis 1 ----------MAER-----ADLVFHNKVIDGTAIKRLISRLIDHFGMAYTSHILDQLKTLGFQQATATSISLGIDD
11_Nicotiana 1 ----------MAER-----ANLVFHNKAINGTAMKRLISRLIDHFGMAYTSHILDQVKTLGFQQATATSISLGIDD
13_Syringa 1 ----------MEVLMAER-----TNLVFHNKVIDGTAMKRLISRLIDHFGMAYTSHILDQVKTLGFHQATATSISLGIDD
18_Liquidambar 1 ----------MEVLMAER-----ANLVFHNKVIDGTAMKRLISRLIDHFGMAYTSHILDQVKTLGFQQATATSISLGIDD
19_Papaver 1 ----------MAER-----ADLVFHNKAIDGTAIKRLISRLIDHFGMAYTSHILDQVKTLGFHQATATSISLGIDD
20_Ananas 1 ----------MAER-----ADLVFHNKVINGTAMKRLISRLIDHFGMGYTSHILDQVKTLGFHQATATSISLGIDD
28_Liriodendron 1 ----------MEVLMAER-----ADLVFHNKVIDATAMKRLISRLIDHFGMAYTSHILDQVKTLGFQQATATSISLGIDD
30_Magnolia 1 ----------MEVLMAER-----ADLVFHNKVIDATAMKRLISRLIDHFGMAYTSHILDQVKTLGFQQATATSISLGIDD
32_Nymphaea 1 ----------MEVLMAER-----ADLVFHNKVIDGTAMKRLISRLIDHFGIAYTSHILDQVKTLGFQQATATSISLGIDD
33_Amborella 1 ----------MAER-----AGLVFHNKVIDGTAIKRLISRLIDHFGMAYTSHILDQVKTLGFRQATATSISLGIDD
35_Picea 1 ------MKIWRFFLMKERTRLPFDNLPFYNKVMDKTAIKKLISRLIDHFGMTYTSHILDQLKTSGFQQATDTAISLGIDD
44_Ginkgo 1 ----------MTER-----AKLLFHNKVMNRIATKQLISRLIDHFGMTYTSHISDQLKASGFQQATDAAISLGIDD
51_Physcomitrium 1 ----------MLFYNKVMDRTAIKQLISRLITHFGITYTTYILDQLKTVGFKQATQAAISLGIDD
Secondary channel rim-helices
β'a13:L984-E979
β'a14:T984-F1011
β'b6:K908-F1011
T. thermophilus 954 AVIPEEKKQYLEEADRKLLQIEQAYEMGFLTDRERYDQILQLWTETTEKVTQAVFKNFE----------ENYPFNPLY
E. coli 644 MVIPEKKHEIISEAEAEVAEIQEQFQSGLVTAGERYNKVIDIWAAANDRVSKAMMDNLQTETVINRDGQEEKQVSFNSIY
0_Nostoc 60 LMVPPTKRSLLEAAEEEIRATEARYQRGEITEVERFQKVIDTWNGTSEALKDEVVVHFK----------KTNPLNSVY
1_Litchi 62 LLTIPSKGWLVQDAEQQSLILEKHHHYGNVHAVEKLRQSIEIWYATSEYLRQEMNPNFR----------MTDPLNPVH
2_Arabidopsis 62 LLTIPSKGWLVQDAEQQSWILEKHHHYGNVHAVEKLRQSIEIWYATSEYLRQEMNPNFR----------MTDPFNPVH
3_Gossypium 62 LLTIPSKGWLVQDAEQQSLILEKHHHFGNVHAVEKLRQSIEIWYATSEYLRQEMNPNFR----------MTDPFNPVH
5_Ricinus 66 LLTIPSKGWLVQDAEQQSLILEKHYHYGNVHAVEKLRQSIEIWYATSEYLRQEMNLNFR----------MTEPFNPVH
6_Rosa 62 LLTIPSKGWLVQDAEQQSLILEKHHHYGNVHAVEKLRQSIEIWYATSEYLRQEMNPNFR----------MTDPFNPVH
9_Cucumis 62 LLTIPSKGWLVQDAEQQSLILEKHHHYGNVHAVEKLRQSIEIWYATSEYLRQEMNPNFR----------MTDPFNPVH
11_Nicotiana 62 LLTIPSKGWLVQDAEQQSLILEKHHHYGNVHAVEKLRQSIEIWYATSEYLRQEMNPNFR----------MTDPFNPVH
13_Syringa 66 LLTIPSKRWLVQDAEQQSLILEKHHHYGNVHAVEKLRQSIEIWYATSEYLRQEMNPNFR----------MTDPFNPVH
18_Liquidambar 66 LLTIPSKGWLVQDAEQQSFILEKHHHYGNVHAVEKLRQSIEIWYATSEYLRQEMNPNFR----------MTDPFNPVH
19_Papaver 62 LLTIPSKGWLVQDAEQQSLILEKHHHYGNVHAVEKLRQSIEIWYATSEYLRQEMNPNFR----------MTDPLNPVH
20_Ananas 62 LLTIPSKGWLVQDAEQQSFILEKHHHYGNVHAVEKLRQSIEIWYATSEYLRQEMNPNFR----------MTDPSNPVY
28_Liriodendron 66 LLTIPSKGWLVQDAEQQSLILEKHHHYGNVHAVEKLRQSIEIWYATSEYLRQEMHPNFR----------MTDPSNPVH
30_Magnolia 66 LLTIPSKGWLVQDAEQQSLILEKHHHYGNVHAVEKLRQSIEIWYATSEYLRQEMHPNFR----------MTDPSNPVH
32_Nymphaea 66 LLTTPSKRWLVQDAEQQSLILEKHHHYGNVHAVEKLRQSIEIWYATSEYLRQEMNLNFK----------MTDPSNPVH
33_Amborella 62 LLTIPSKGWLVQDAEQQSLILEKHHHYGNVHAVEKLRQSIEIWYATSEYLRQEMNPNFR----------MTDPFNPVH
35_Picea 75 LLTAPSKAWLVQDAEQQGSVSEKQNHYGNVHAVEKLRQSIEIWYATSEYLRKEMNPNFS----------MTDPLNPVH
44_Ginkgo 62 LLTAPSRGWLVRDAEQQGSISEKHHHYGNVYAVEKLRQSIETWYATSEYLRQEMNPNFG----------MTDPSNPVH
51_Physcomitrium 56 LLTAPSKSWLIQDAEQQGYISEKHYRYGNVHAVEKLRQLIETWYATSEYLKQEMNPNFR----------MTDPLNPVH
Bridge helix
β'a15:N1018-G1113
β'b7:N1018-G1113
T. thermophilus 1022 VMAQSGARGNPQQIRQLCGLRGLMQKPSGETFEVPVRSSFREGLTVLEYFISSHGARKGGADTALRTADSGYLTRKLVDV
E. coli 724 MMADSGARGSAAQIRQLAGMRGLMAKPDGSIIETPITANFREGLNVLQYFISTHGARKGLADTALKTANSGYLTRRLVDV
0_Nostoc 128 MMAFSGARGNISQVRQLVGMRGLMADPQGEIIDLPIKTNFREGLTVTEYIISSYGARKGLVDTALRTADSGYLTRRLVDV
1_Litchi 130 IMSFSGARGNASQVHQLVGMRGLMSDPQGQMIDLPIQSNLREGLSLTEYIISCYGARKGVVDTAVRTSDAGYLTRRLVEV
2_Arabidopsis 130 MMSFSGARGNASQVHQLVGMRGLMSDPQGQMIDLPIQSNLREGLSLTEYIISCYGARKGVVDTAVRTSDAGYLTRRLVEV
3_Gossypium 130 IMSFSGARGNASQVHQLVGMRGLMSDPQGQMIDLPIQSNLREGLSLTEYIISCYGARKGVVDTAVRTSDAGYLTRRLVEV
5_Ricinus 134 IMSFSGARGNVSQVHQLVGMRGLMSDPQGQMIDLPIQSNLREGLSLTEYIISCYGARKGVVDTAVRTSDAGYLTRRLVEV
6_Rosa 130 MMSFSGARGNASQVHQLVGMRGLMSDPQGQMIDLPIQSNLREGLSLTEYIISCYGARKGVVDTAVRTSDAGYLTRRLVEV
9_Cucumis 130 IMSFSGARGNASQVHQLVGMRGLMSDPQGQMIDLPIQSNLREGLSLTEYIISCYGARKGVVDTAVRTSDAGYLTRRLVEV
11_Nicotiana 130 IMSFSGARGNASQVHQLVGMRGLMSDPQGQMIDLPIQSNLREGLSLTEYIISCYGARKGVVDTAVRTSDAGYLTRRLVEV
13_Syringa 134 IMSFSGARGNASQVHQLVGMRGLMSDPQGQMIDLPIQSNLREGLSLTEYIISCYGARKGVVDTAVRTSDAGYLTRRLVEV
18_Liquidambar 134 IMSFSGARGNASQVHQLVGMRGLMSDPQGQMIDLPIQSNLREGLSLTEYIISCYGARKGVVDTAVRTSDAGYLTRRLVEV
19_Papaver 130 IMSFSGARGNASQVHQLVGMRGLMSDPQGQMIDLPIQSNLREGLSLTEYIISCYGARKGVVDTAVRTSDAGYLTRRLVEV
20_Ananas 130 LMSFSGARGNASQIHQLVGMRGLMSDPQGQMIDLPIQSNLREGLSLTEYIISCYGARKGVVDTAVRTSDAGYLTRRLVEV
28_Liriodendron 134 IMSFSGARGNASQVHQLVGMRGLMSDPQGQMIDLPIQSNLREGLSLTEYIISCYGARKGVVDTAVRTSDAGYLTRRLVEV
30_Magnolia 134 IMSFSGARGNASQVHQLVGMRGLMSDPQGQMIDLPIQSNLREGLSLTEYIISCYGARKGVVDTAVRTSDAGYLTRRLVEV
32_Nymphaea 134 IMSYSGARGNASQVHQLVGMRGLMSDPQGQMIDLPIQSNLREGLSLTEYTISCYGARKGVVDTAVRTSDAGYLTRRLVEV
33_Amborella 130 MMSFSGARGNASQVHQLVGMRGLMSDPQGQMIDLPIQSNLREGLSLTEYIISCYGARKGVVDTAVRTSDAGYLTRRLVEV
35_Picea 143 VMSFSGARGSTSQVHQLVGMRGLMSDPQGQIIDLPIRRNLREGLSLTEYIISCYGARKGVVDTAVRTADAGYLTRRLVEV
44_Ginkgo 130 MMSFSGARGNTSQVHQLVGMRGLVSDPQGQIIDLPIQRNFREGLSLTEYIISCYGARKGVVDTAVRTSDAGYLTRRLVEV
51_Physcomitrium 124 MMSFSGARGSTSQVHQLVGMRGLMSDPQGQIIDLPIQSNFREGLSLTEYIISCYGARKGVVDTAVRTSDAGYLTRRLVEV


Trigger loop-helix
Trigger loop
β'a16:R1213-A1247
β'b8:V1186-A1247

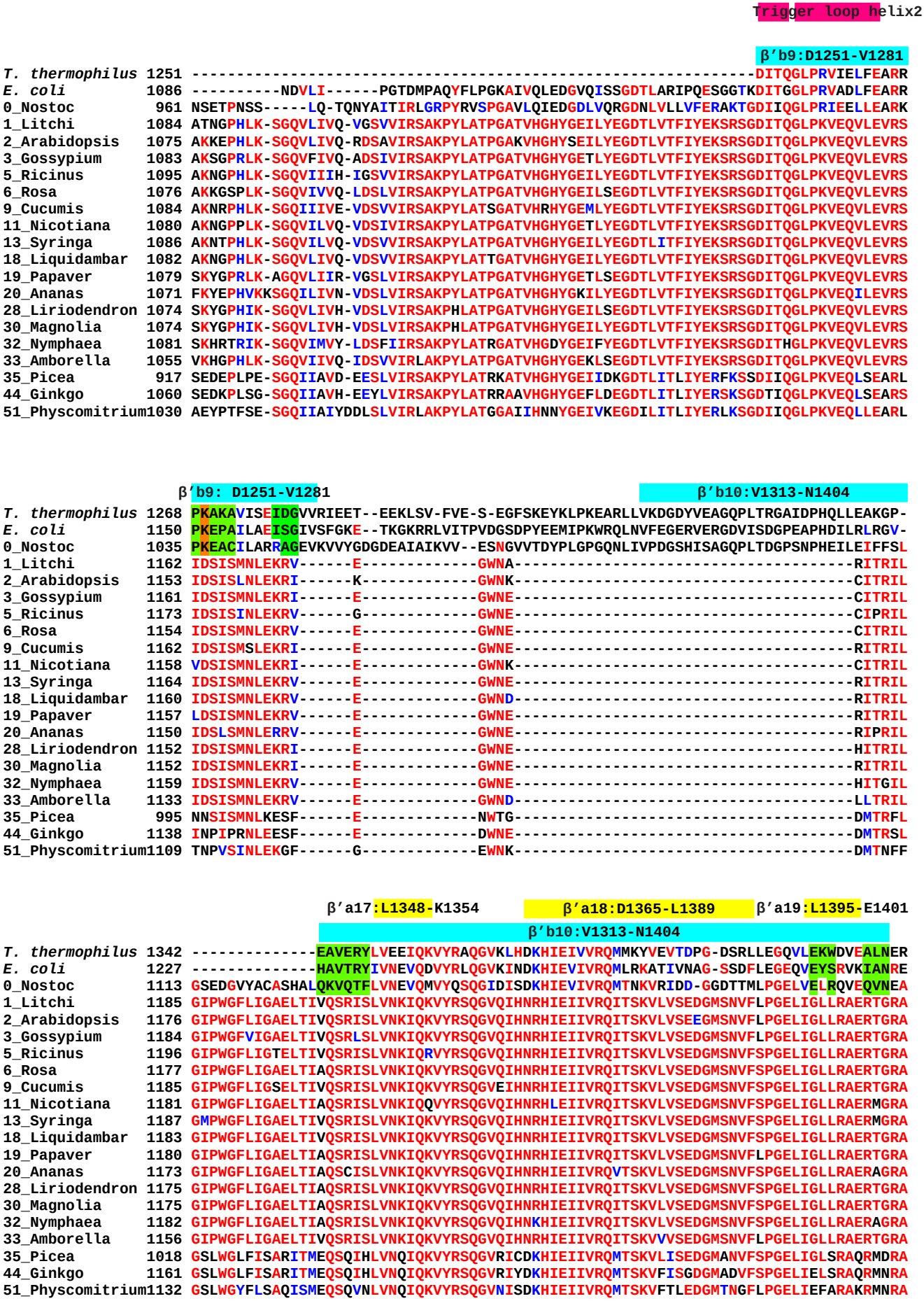
```
T. thermophilus 1251 ------------------------------------------------------------------------------
E. coli          961 ------------------------------------SNVKSVVNSSGKLVITS-------------------------
0_Nostoc         435 GRTTRTNTEKAVKDVASDLAGEVQFAEVVPEQKTDRQGNTTTTAARGGLIWILSGEVYNLPPGAELVVKNGDAIASNGVL
1_Litchi         427 TY-TLNFKERVRKHIYSDSEGEMHWSTDVYHAPEFTYSNVH-LLPKTSHLWILSGGSCGSGVVSFSLYKDQDQLSIHYRS
2_Arabidopsis    429 TY-TFHFKERVRKYIYSDSEGEMHWSTDVSHAPEFTYSNVH-LLPKTSHLWILSGGSCGSSLIRFSIHKDQDQMNIPFLS
3_Gossypium      427 TY-TLNLKERVRKHIYSDSEGEMHWSTDVYHSPEFTYSNVH-LLPKTSHLWILSGGSYKFSVVPFSLHKDQDQINIHYLS
5_Ricinus        431 TY-TLNFKEKVRKHIYSDSEGEMHWSTDVYHAPEFTYSNVH-LLPKTSHLWILSGNSCRSSIVPFSLHKDQDQMNVHSLS
6_Rosa           427 AY-TFNFKERVRKHIYSDSEGEMHWSTDVYHAPEFTYSNVH-LLPKTSHLWILSGGSCRFSAVPPSLHKDQDQTNVHSLS
9_Cucumis        427 TY-TLNLKERVRKHIYSDSEGEMHWSTDVYHAPEFTYSNVH-LLPKTSHLWILSGGSCGCSVVPFSLYKDQDQINVHSLC
11_Nicotiana     427 IS-TLNFKEKVRKHIYSDSDGEMHWSTDVYHAPEFTYGNVH-LLPKTSHLWILLGRPCRSSLVYLSIHKDQDQMNAHFLS
13_Syringa       431 TS-TLNFKEKVRKHIYSDSDGEMHWSTDVYHAPEFTYGNVH-LLPKTSHLWILLGGPCRSSLVSLSLHKDQDQINAHSRS
18_Liquidambar   431 TY-TFNFKERVRKHIYSDSEGEMHWSTDVYHAPEFTYGNVH-LLPKTSHLWILSGGSCRSSVVPFSLHKDQDQMNVHSLS
19_Papaver       427 TS-TFNFKERVRKHIYSDLEGEMHWSTDVYHAPEYTYGNVH-LLPKTSHLWILSGGLCRSSIVPFSLGKDQDQTNVHSLF
20_Ananas        426 TS-TFHFKERVRKHIYSESEGEMHWSTDVYHAPEYTYGNVH-LLPKTSHLWILAGGPCRSSIVSFSLHKDQDQMNVYSLS
28_Liriodendron  430 TS-TFNFKERVRKHIYSDSEGEMHWSTDVYHAPEYRYGNVH-LLPKTSHLWILSGGPCRSSIVPFSLHKDQDQMNVHSLS
30_Magnolia      430 TS-TFNFKERARKHIYSDSEGEMHWSTDVYHAPEYRYGNVH-LLPKTSHLWILSGGPCRSSIVTFSLHKDQDQMNVHSLS
32_Nymphaea      430 TS-TFHFKERVRKHIYSDSEGEMHWSTGVYHAPEYTHGNVH-FLPKTSHLWILSGGPCKSSLVPFSLHKDQDQMNVQSLS
33_Amborella     427 TS-TFNFKEKVRKHIYSDSEGEMHWSTDVYHAPDFTYSNVH-LLPKTSHLWILSGSSYRSSVVPFSIHKDQDQTNVYFLS
35_Picea         444 TS---SFKEKVRKNIYSDLEGEMHWSTNVCHAPEYVHGNVH-PILRTGYLWILSGGIYGSGVVPFPFHKHQDQVDVQPFV
44_Ginkgo        429 TS---PSKEKVMKHIYSDLEGEMHWSTNVCHAPENVHGNVH-LILRTSYLWVLSGGLYESGVVPFPLYKDQDQVNIQFPL
51_Physcomitrium 418 TS---PFKEKVQKYIYSNLSGEMHWSSKVQHSSEYIHSNVH-LLRKTGHIWILAGNFDKDNKFSFIFYQNQDKLDNKLPI


T. thermophilus 1251 ------------------------------------------------------------------------------
E. coli          978 ------------------------------------------------------------RNTELKLIDEFGRTKES
0_Nostoc         515 AETKLTT------LHGGVVRL-------------PEATPGK--------STREIEIITASVVLDQATVTVQSS--QGRNN
1_Litchi         505 VERRYISSLS---VNNDQVRHQLVSSDFSDNK--EDGISDY-SGFNRIIGIGHCNLIHAAILHENSD--LLAK--RRRNR
2_Arabidopsis    507 AERKSISSLS---VNNDQVSQKFFSSDFADPK--KLGIYDY-SELNGNLGTSHYNLIYSAIFHENSD--LLAK--RRRNR
3_Gossypium      505 AERRYISRFS---VNNDQVRHNLFSSDFSDKK--EERIYDY-SELNRIIGTGHCDFIYSAILHENAD--LLAK--RRRNR
5_Ricinus        509 IKRRYISSPSVNSVNNDQVKPKFFSSDFSGKK--PSRIPYY-SELNRIVCTGHCNLIYPAILYENSD--LLAK--RRRNK
6_Rosa           505 VEGRYFSSLS---VNNDQVKHKFFGLNLSGKK--ESCIPDY-SELNRIIYTSHCNLIFPPIRHDN-F--LLTK--RRRNR
9_Cucumis        505 VERRYISSLS---VNNDKVGQKFYGPDLSGKN--ESGIPDY-SELNPILCTGQSNLTYPAIFHGNSD--LLAK--RRRNG
11_Nicotiana     505 GKRRYTSNLS---VTNDQARQKLFSSDFSGKK--EDRIPDY-SDLNRIICAGQYNLVYSPILHENSD--LLSK--RRRNK
13_Syringa       509 VKRRYTSNLS---GTNDPERQKLFSSYFYGKKKXEDRISDY-SDLNRIICNGRCNLIYPTILHQNSD--LFSK--RRRNR
18_Liquidambar   509 VERRYISNLS---VTNDQVRHKLFSSDISGKK--EGRIPDY-SELNRIICSGHCNLIYPAILRENSD--LLAK--RRRNR
19_Papaver       505 AKQRYTPSLS---VTNDQVKQKFCSSESSGTG--GRGVLDY-SGPDRIICNGHCNFIYPPILHESSD--LLAK--RRRNR
20_Ananas        504 VEGRYISNPS---MTNDQVRHKLLDT--SGKK--DRKILDY-SRLDRIISNGHWNFIYPSILQENPD--FLAK--KRRNR
28_Liriodendron  508 VERRYISDLS---VTNDRVRHKLFSSDPSGKK--KERILDY-SGPDRIVSNGHWNFLYPAILHENSD--LLAK--RRRNR
30_Magnolia      508 VERRYISDLS---VTNDRVRHKLFSSDPSSKK--GKGILDY-SGPDRIISNGHWNFLYPATLHENSD--LLAK--RRRNR
32_Nymphaea      508 VQERSISDFS---VNNNRVKHKLFGSDPLARK--GRRISDYAAGLERVISNGDGDFIYPAILRENSY--LLAK--RRRNR
33_Amborella     505 AEGKNISSRSVNTVNNDQGKGKFFSSDFSGKK--ESTIPDY-SEFNRIIDRDHWNLIFPSILHKNYDLFLLAK--RRRNR
35_Picea         520 AKHQSLFDSY---V--DQVEHRSGDSNCYGKE---EQIFSY-SETDRTISNEHRDSIYVTLSPKNYN--MKGK--RQMNR
44_Ginkgo        505 AKHKSLSDSS---VNQDRVKHKSVDSNFSGKE---EKISGY-SGIDRIMSNEHWDSIYSTIPFDNCK--ILGK--KQRNR
51_Physcomitrium 494 AKQ----------------------------------TLNY-FQLKEHFLNNFWNSIYSSIILYNYR--FLEK--K-NNK


T. thermophilus 1251 ------------------------------------------------------------------------------
E. coli          995 YKVPYG------------------------AV----------------------------------------------
0_Nostoc         566 YLVSTGNNQVFN-----------LRATPGTKVQNGQVVAELIDDRYRTTTGGFLKFAGVEVQKKGKA-----------
1_Litchi         575 FLIPFQSIQEQEKELMPH--SGISIEIPIKGVFRKNSIFAYFDDPRYRRKNSGITKYGTIGAHSIVKKEDLIEYRGRGGK
2_Arabidopsis    577 FLIPFQSIQEQEKEFIPQ--SGISVEIPINGIFRRNSIFAFFDDPRYRRKSSGILKYGTLKADSIIQKEDMIEYRG--VQ
3_Gossypium      575 FIIPFQLIQDQEKELMLHSHSGISMEIPINGIFRRKSILAFFDDPRYRRKSSGITKYGTLGAHSIVKREDVIEYRG--VK
5_Ricinus        582 FIIPFQSIQEQEKKLMTRS-SAISIEIPLNGIFRRNSVFAYFDDPQYRRKSSGITKYGAIGVHSIVKKEDLIEYRG--VK
6_Rosa           574 FIIPFQSIQEQEKERMPR--PDISIEIPINGIFRRNSILAYFDDPQYRRKSSGITKYGTVGLHSILKKEDLIEYRG--VK
9_Cucumis        575 FIIQFESLQERKELRPP--SGISIEIPINGLFRRNSILAFFDDPQYRRNSSGITKYGTIGVHSILKKEDLIEYRG--VK
11_Nicotiana     575 FIIPLHSIQELENELMPC--SGISIEIPVNGIFRRNSILAYFDDPRYRRKSSGIIKYGTVETHSVIKKEDLLEYRG--VK
13_Syringa       581 FIIPLQSIQERENELMPR--SGISIEIPPNGIFRRNSILAYFDDPRYRRKSSGITKYGTIEMHSIVKKEDLIEYRG--VK
18_Liquidambar   579 FIIPFQSIQERKEQMPHSNSGISIEIPINGIFRRNSILAYFDDPRYRRKSSGITKYGTIEVHSIVKKEDLIECRG--VK
19_Papaver       575 LIIPFQSNQERDKERIPR--SGISIEIPINGIFRRNSILAYFDDPRYRRNSSGITKYETLEMHSIVKKEDFIEYRR--AK
20_Ananas        572 FIIPLQYDQEREKELIPC--FGISIEIPINGILRRNSILAYFDDPRYRRSSSGITKYGTIEVDSIIKKEDLIEYRG--AK
28_Liriodendron  578 FIIPFQYDQEREKELMPR--SGISIEIPINGILRRDTILAYFDDPRYRRSSSGITKYGTIEVDSIVKKEDLIEYRG--AK
30_Magnolia      578 FIIPFQYDQEREKELMPR--SGISIEIPINGILRRNSILAYFDDPRYRRSSSGITKYGTIEVDSIVKKEDLIEYRG--AK
32_Nymphaea      579 FIIPFQYDPEREKELTPHSSTSITVEIPANGILRRNSILAYFDDPRYRRSSSGITKYGIIEVDSIVKKEGLIEYRR--PK
33_Amborella     580 FIIPFQWIQERENELMLR--SSISIEIPINGVFRKNSILAYFDDPQYRRKSSGITKYGAIGLHSIFKKEDLIEYVG--IK
35_Picea         587 FIVPLQCDKEWGKRIISF--PDAILRIPKSGVLQRNSIFGY-----------------------------------------
44_Ginkgo        574 LIVPLRYDKERKKRRIPC--PNSILRIPRNFLFQRNHILAVLDDPQYRVDSSGILKYGNIRIDSIEKKDDFLEDQG--SR
51_Physcomitrium 534 ----------YEKKL----LFQFMLKLPKNGILKQNDIFAIFNDPKYRIKNSGIIKYGNIKVDLINKKNDIFEDQK--TK
```

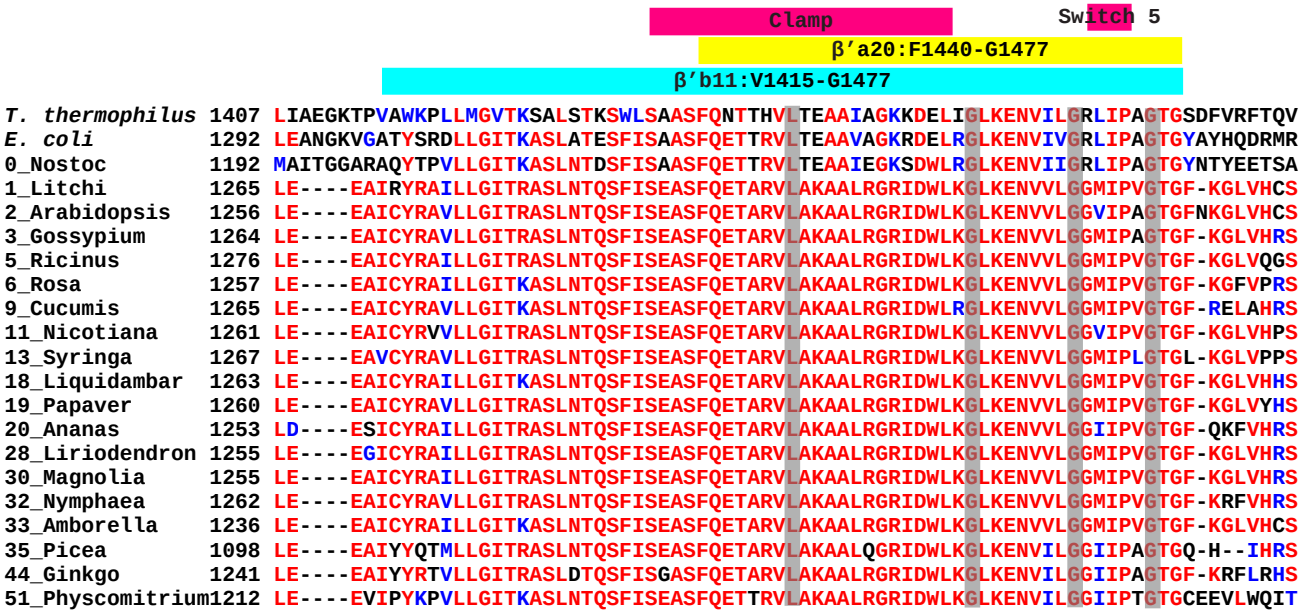
```
T. thermophilus 1251 ------------------------------------------------------------------------------
E. coli         1003 ------------------------------------------------------------------------------
0_Nostoc         622 --KLGYEV-VQGGTLLWIPEETHEVNKD-ISLLLVEDGQFVEAGTEVVK--DIFCQNSGVIEVTQKNDILREVVVKPGEL
1_Litchi         653 KIKPKYQ--MKFDRFFFIPEEVHTLPES--SYVMVRNNSLIGVDTRITL--NRRSQVGGLVRVERKKKR-IELKIFSGDI
2_Arabidopsis    653 KIKTKYE--MKVDRFFFIPEEVHILPES--SAIMVQNYSIIGVDTRLTL--NIRSQVGGLIRVEKKKKR-IELKIFSGDI
3_Gossypium      653 KVKPKYQ--MKVDRFFFIPEEVHILSES--SSIMVRNNSIIGVDTPITL--NTRSQVGGLVRVERKKKR-IELKIFSGNI
5_Ricinus        659 EFKPKYQ--MKVDRFFFIPEEVYILPES--SSLMVRNNSIIGVDTPITL--NTRSRVGGLVRVERKKKK-IELKIFSGDI
6_Rosa           650 EFKPKYQ--TKVDRFFFIPEEVHILPES--SSIMVRNNSIIGIDTRITL--NTRSRVGGLVRIERKKKR-IELKIFSGDI
9_Cucumis        651 DFKPKYQMQMKVDRFFFIPEEVHILPES--SSIMVRNNSIIGVATRLTL--SIRSRVGGLVRVEKKKKR-IELKIFSGDI
11_Nicotiana     651 EFRPKYQ--MKVDRFFFIPEEVHILPGS--SSIMVRNNSIVGVDTQITL--NLRSRVGGLVRVERKKKR-IELKIFSGDI
13_Syringa       657 AFRPKYQ--MKVDRFFFIPEEVHILPGS--SSIMVRNNSLIGVDTQITL--NIRSRVGGFVRVERKKKR-IELQIFSGDI
18_Liquidambar   657 EFKPRYQ--MKVDRFFFIPEEAHILPGS--SSIMVRNNSIIGVDTQITL--NTRSRVGGLVRVERKKKR-IELKIFSGDI
19_Papaver       651 EFRQKYQ--KKVDRFFFIPEEVHILSGS--SSIMVRNNSIIGIDTRITL--NIRSRVGGLVRVERKKKR-IELKIFSGDI
20_Ananas        648 EFSPKYQ--TEVDQFFFILEEVHILPGS--SLIMVRNNSIIGVDTRLALNINTRSRVRGLVRVERKKKY-IELKIFSGDI
28_Liriodendron  654 EFRPKYQ--MKVDRFFFIPEEVHILPGS--SPIMVRNNSIIGVDTRIAL--NTRSRVGGLVRVERKKKK-IELKIFSGDI
30_Magnolia      654 EFRPKYQ--MKVDRFFFIPEEVHILPGS--SSIMVRNNSIIGVDTRIAL--NTRSRVGGLVRVERKKKR-IELKIFSGDI
32_Nymphaea      657 ESRPKYQ--MKVDRFFVIPEEVHILPGS--SSIMVRNNSIIGVDTRITF--NTRSQIGGLVRIEKKKKK-IELKIFSGGI
33_Amborella     656 ELKPKYQTKYYWNKYT-----NHF-------------------------KYKKPSRRIGPSGEKKKR-IELKIFSGEI
35_Picea         626 ------------------------------------------------------------------------------
44_Ginkgo        650 GSRPKYE--IEGGRFLFIPEEVHILHES--SSIMVRNGSIIRTGTQITF--NIESQVGGLVRIERMRKK-IEVRILPGDI
51_Physcomitrium 598 TVRPRYKI-LKEGNFFLLPEEVYILDQSSFSSILVKNNSFIKAGTKITF--NISSKITGFVKIKKKFNN-FKIKILPGSI

T. thermophilus 1251 ------------------------------------------------------------------------------
E. coli         1003 ------------------------------------------------------------------------------
0_Nostoc         696 LMVDDPEAVMGRDNTFVQPGEEFQGT--------VATELRYIQYVE-TPEGPALLSRPVVEFAVPNNPDVPSTTS----V
1_Litchi         726 HFPGEADKISRHSGILIPPETGKKKLKESTGESKKLKKWIYVQRITLTKKKYFVLVRPVVTYEIAD---GINLATLFPQD
2_Arabidopsis    726 HFPDKTDKISRHSGILIPPGRGKKNSK----ESKKFKNWIYVQRITPTKKKFFVLVRPVATYEIAD---SINLATLFPQD
3_Gossypium      726 YFPGERDKISRHSGILIPPGTGKTNSK----ESKKLKNWIYVQRITPTKKKYFVLVRPVTPYEIPD---GLNLATLFPQD
5_Ricinus        732 HFPGETDKISRHSGILIPPGMVKTNSK----ESKKQKNWIYIQRIAPTRKKYFVLVRLVIIYEIAN---GINLETLFPRD
6_Rosa           723 HFPGEMDKIFRHNGILIPPGT---NSK----ESKKRNNWIYVQWITPTKKKYFVLVRPVIIYEIAD---GINLATLFPQD
9_Cucumis        726 HFPGEMDKISRHNGILIPPERVKKNSK----KSKKSKNWIYVQWITPTKKKYFVFVRPVIIYELAD---GINLVKLFPQD
11_Nicotiana     724 HFPGETDKISRHTGVLIPPGTGKRNSK----ESKKVKNWIYVQRITPSKKKFFVLVRPVVTYEITD---GINLATLFPPD
13_Syringa       730 HFPGETDKISRHSGVLIPPGTGNSNSK----ESKKLKNWIYVQRITPSKKKYFVLVRPVVTYEITD---GINLVTLFPPD
18_Liquidambar   730 HFPGETNKIARHSGILIPPGTGKTNSK----ESKKLKNWIYVQRITPTKKKHFVLVRPVVTYEIAD---GINLATLFPQD
19_Papaver       724 HFPGETDKISWHSGILIPPGTGKKNAG----DSKKLKNWIYVQRITPIKKKFFVLVRPVVTYEIAD---GINLATLFPHD
20_Ananas        723 HFPGETDKISRHSGIFIPPETEKKNSK----ESKKWKNWIYVQRITPTKKKYFVSVRPVVTYEISD---GINLATLFPRD
28_Liriodendron  727 HFPGETDKISRHSGILIPPGTGKKNSK----ESKKWKNWIYVQRITPTKKKYFVSVRPVVTYEIAD---GINLGTLFPQD
30_Magnolia      727 HFTGETDKISRHSGILIPPGTGKRNSK----ESKKWKNWIYVQRITPTKKKYFVSVRPVVTYEIAD---GINLGTLFPQD
32_Nymphaea      730 HFPGETDKISRHIGILIPPGARKKMDKGSQGKNWEGKNWVYVQRITPIKKKYFVSVRPVVTYEIAD---GINLVTLFPGD
33_Amborella     703 QFPVEMDKIFRHSGILIPPGRVKKKIK----ESKKLKNWIYVQWITPTKKKYFVLVRPVIIYEIAD---GINLETLFPQD
35_Picea         626 ------------------------------------------------------SNVEYGIPD---GPIMATSFSLD
44_Ginkgo        723 YFPGEIHEISRHNGTLIPPGKIIFD-------EFQSVNWIYFQWITPHKEKPFVPVRSAAEYGIHD---GSNRTAPFYLD
51_Physcomitrium 674 YYPKEKQKNFKQNGILIPPGEKIF-------EQFRAKNWIYLEWIVLSKDNSFFLIRPAIEYKIIFNDNPLTLPIPFYLD

T. thermophilus 1251 ------------------------------------------------------------------------------
E. coli         1003 --------------LAKGDGEQVA----------GGETVANWDPHTMPVIT---------------EVSGFVRFTDMID
0_Nostoc         763 SQQTGRSIQLRAVQRLPYKDSERVKSVE--GVELLRTQLVLEIEQEGEQQDHNASPLAADIELVQDTEDPEVQRLQLVIL
1_Litchi         803 PLREKDNMQLRVVNYILYGKGKPTRGISDTSIQLVRTCLVLNWGQDKKS-SSAEEVRTSFVEVSTNGMIRDFLRIDLVQS
2_Arabidopsis    799 LFREKDNIQLRVFNYILYGNGKPTRGISDTSIQLVRTCLVLNWDKN----SSLEEVRAFFVEVSTKGLIQDFIRIGLVKS
3_Gossypium      799 PFQEKDNMQLRAVNYILYGNGKPTRRISDTSIQLVRTCLVLSWDQDNKS-SFAEEVCASFVEVRTNGLIRDFLRIDLVKS
5_Ricinus        805 LLQEKDNLKLRVVNYILSGNGKPIRGISDTSIQLVRTCLVLNWDQEKKS-SSIEEARASFVEVNTNGLIRDFLRINLVKS
6_Rosa           793 PLRERDNLELRVVNYILYGNGKPIRGISGTSIQLVRTCLILNWDKNKKS-SSIEEAHASFVEVSANGLIQDFLRINLVKS
9_Cucumis        799 LLQERDNLELRVVNYILYGNGKPIRGISGTSIQLVRTCLLLNWDRDKKS-SSIEDARASFVEVSTNGLVRNFLRIDLGKS
11_Nicotiana     797 PLQERDNVQLRIVNYILYGNGKPIRGISDTSIQLVRTCLVLNWNQDKKS-SSCEEARASFVEIRTNGLIRHFLRINLVKS
13_Syringa       803 LLQERDNVQLRVVNYILYGNGKPIRGISDTDIQLVRTCLVLNWDQDKKSSSSEEARASFVEIRTNGLIRHFLRIDLVKS
18_Liquidambar   803 LLQERDNMKLRVVNYILYGNGKPIRGISDTSIQLVRTCLVLNWDQDKKSA-SSGEAHASFVEVRTNGLIRNFLRINLVES
19_Papaver       797 LLQERDNVQLRVVNYILYGNGKPIRGIYHTSIQLVRTCLVLNWNQEKKG-SSIEEVQASFVEVRVNNLIRYFIRMDLVKS
20_Ananas        796 ILQEKDNVQLRVVNYILYGNGKSIRGIYHTSIQLVRTCLVLNWDQEQNG--FIEEVHASFVEVRANDLIRDFIRIELVKS
28_Liriodendron  800 LLQERDNVQLRVVNYILYGNGKPIRGIYHTSIQLVRTCLVLNWDQDRNG--SIEEVHASFVEVGTNDLIRDFIRIDLVKS
30_Magnolia      800 LLQERDNVQLRVVNYILYGNGKPIRGIYHTSIQLVRTCLVLNWDQDRNG--SIEEVHASFVEVGANDLIRDFIRIDLVKS
32_Nymphaea      807 MLQEKDNLRLQVVNYILYGDGKPIRGISHTSIQLVRTCLVLNWDQDKKG--SIEKVQASSAEVRANDLIRYFIRIDLVKS
33_Amborella     776 LLQEKDNLELRVVNYILYGNGKPILGISGTSIQLVRTCLMLNWDQDNKS-SSSEEAHVYFVEVSTTGLIRDFLRINLAKS
35_Picea         646 LSREGDNLQIQVSNSSSYEDGERIQVMSDTSIPLVQTCLGFDWEQIDS---IESEAYASLTSVRTNKIVSNMIQISLIKY
44_Ginkgo        793 LLGEEDNLQVQVGNYILYGDGEQIQVISDTSIQLVRTCSVLNWEQKDS---ME-EAYAFLTEVRINEVVRNFLQISLMKY
51_Physcomitrium 747 LLKEQKKIKIQTVKYILYEDSEEVEINPDTDIQLIQTCLILNWETK----VFIKEAHISFIKIRINKIIKNFFQINLIEN
```

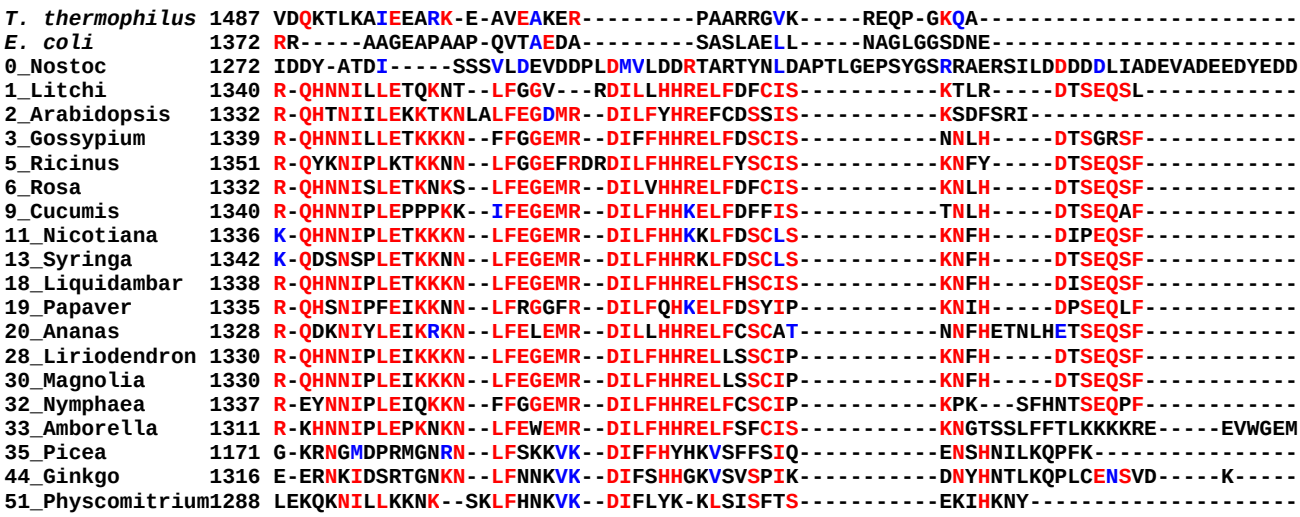
```
T. thermophilus 1251 ------------------------------------------------------------------------------
E. coli         1043 GQTITR-QTDE--LTG-----------------------------------------------------LSSLVVLDSAERT-
0_Nostoc         841 ESLVIRRDITADATQG--------STQTTLEVQDGLTIAPGSVVARTQILSKEGGIVRGVQKGTENVRRCLVLRET----
1_Litchi         882 HISYMR-KRNDPSSSG--LISDNGSDRTNINP--FYSLYF--KARVQQSLSQNQRTLHTLLNRNKKCQSLIILSSSNCFR
2_Arabidopsis    875 HISYIR-KRNNSPDSGLI-------SADHMNP--FYSISPK-SGILQQSLRQNHGTIRMFLNRNKESQSLLILSSSNCFR
3_Gossypium      878 HIFYIR-KRNDPSGSE--LISDNRSDRTNKNP--FYSIYS--NARIQQSFSQNHGTIHTLLNRNKESQSLIILSASNCFR
5_Ricinus        884 HISYISRKRNDPSGSG--PISNNGANHTNINP--FYPIYF--KTRIQQSLKQNQGTISTLLNRNKECQSLIILSSSNCFR
6_Rosa           872 HTSYIR-KRNDPLGSG--LISDNRSDRTNINP--FYSIYS--KERIQQSLRQNQGTFRTLLNRNKESQSLIILSSYNCFQ
9_Cucumis        878 DTAYMR-KRKDPSGSG--LIFNNESDRTNINP--FFSIYS--KTRVPQSPSQNQGTIRTLFNRNKERQSLIILSASNCLQ
11_Nicotiana     876 PISYIG-KRNDPSGSG--LLSDNGSDCTNINP--FSSIYSYSKAKIQQSINQPQGTIHTLLNRNKECQSLIILSAANCSR
13_Syringa       883 PISYIG-KRNNPSGSG--LLSDNGSDCTNINP--FSSIYS--KARIQHSLNQNQGTIHTLLNRNTGFQSLIILSSSNCFR
18_Liquidambar   882 PISYTG-KRNDPSGSG--WISDNGSDRTNINP--FYSTYS--KERIQQSLSQNQGTIRTLLNRNKECQSLRILSSSNCSR
19_Papaver       876 PILYTR-KRNDRGGAGLIWIPDNGSDRTNLNP--FSF---SSKARIQQTFTQHQGTIRALVNRNKESQPLIILSSSNCFQ
20_Ananas        874 TISYTG-KRYDRASSG--LIPDNGLDRTNINP--FYSKAK------IQSLSQHQGTIGTLLNRNKECQSLIILSSSNCSR
28_Liriodendron  878 PISYIG-KRDDTTGSG--LIPDNESDRTNINT--FYSKT-----R-IQSLTQHQGTIRTFLNRNKECQSFLILSSSDCSR
30_Magnolia      878 PISYIG-KRNDTAGSG--LIPDNESDRTNINT--FYSKT-----R-IQSLTQHQGTICTFLNRNKECQSFLILSSSDCSR
32_Nymphaea      885 PILYTG-KRNDGSGS---VIPDTGSYCANTNL--FSSKVK------MKSLSQHQGTVRTFLNRNKEGQSLIVFSSSNCSR
33_Amborella     855 NICYIR-KRNDPLGSG--LISDNRSDCT--NP--FYSIYS--KEKIQQSLRQNQGTIHTLLNRSKESQSLIILSSSNCFQ
35_Picea         723 PLFFMGR-RDNKASSN--LMFHNKLDHT--NL--FYSN------GERQLISKHQGTICSLYNGEEDSGSFMVLSPSDCFR
44_Ginkgo        869 PG---GK-RKNVTGSK--FLFHNRSDQT--NT--FSSN------RGSQFFSKHQGTIRTLPNEEKEGGSFAVLSPSDCSR
51_Physcomitrium 823 INLMNKK-KNNNIILN--YLFKKKR----------YII-NQKDCEKILLLSKTWGIIRTPSNKNQEKSFFLILSPFNLFQ
```

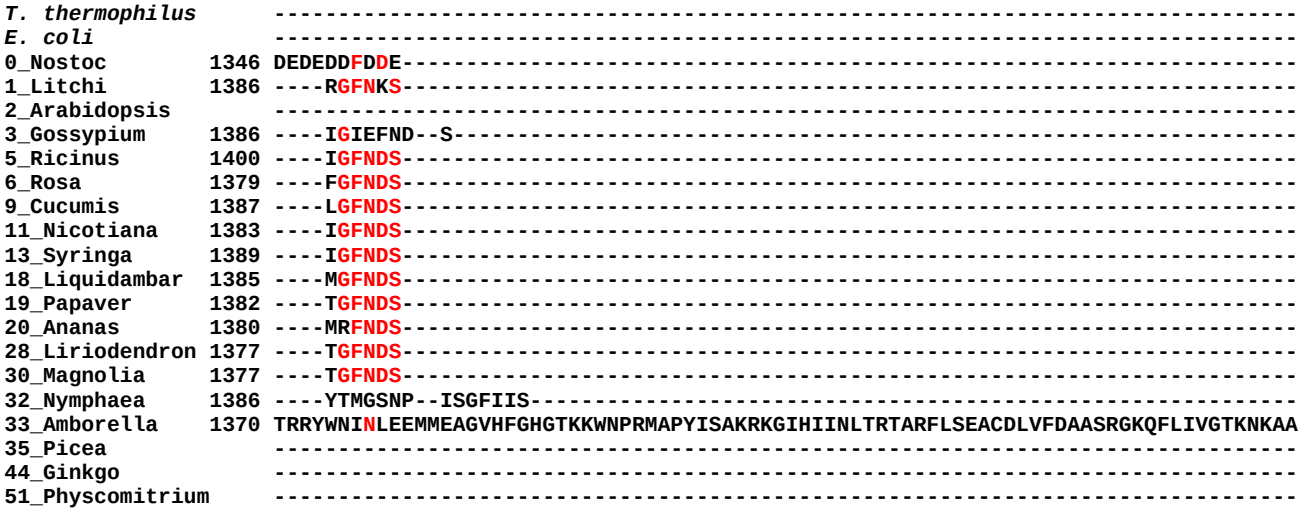
```
T. thermophilus 1251 ------------------------------------------------------------------------------
E. coli         1069 ------------------------------------------------------------------------------AG
0_Nostoc         909 ----------------------------------------------------------DLITVNTSTQPKVKM--
1_Litchi         955 MGPFNDI-KYHNVIKQSIHI------------QKGSLIPIRNSLGPLGT-VLQIANFYSFYY--LITYNQISVT--KYWK
2_Arabidopsis    944 MGPFNHV-KHHNVINQSI--------------KKNTLITIKNSSGPLGT-ATPISNFYSFLP--LLTYNQISLI--KYFQ
3_Gossypium      951 MGPFNDV-KYHNVIKQSI--------------KKDPLIPIKNLLGPLGT-APKIANFYSSFYP-LITHNQTSVA--KYFE
5_Ricinus        958 MDPFNDV-KHHNVIKESI-------------KRDPIIPIRNSLGPLGT-ALQIANLYLFYHLNLITHNRISVT--KYLK
6_Rosa           945 MSPFNDV-KYYDGIKESIKR-----------DRDSLIQITNLLGPLGT-ASQIDLFYSFYHL--LTHNHISVTKYFYLQ
9_Cucumis        951 MDLFNDVKDY-NVIKESS--------------KKDPLISIRNSLGPLGA-APQIVNFYSFYD--LITHNPISLT--KYLQ
11_Nicotiana     951 MGPFKDV-KYHSVIKKSI--------------KKDPLIPIRNSLGPLGT-SLPIENFYSSYH--LITHNQILVT--NYLQ
13_Syringa       956 MGPFNDVIKYHNVIKESIKI------------TKDPLIPLKNSLGPFGT-AFTIANFFSFYH--LITHNQILVN--NYLQ
18_Liquidambar   955 MGPFNDV-KYHNVIKESI--------------KRDPLIPIRNSLGPLGT-ALQIANFYSFYH--LITHNKILVT--KYLQ
19_Papaver       950 MGPFNSVKYNDGVTKEST--------------KRDLRISILNSLGPLGI-VPKFVNFS---SYHLITHNQILVK--KYLL
20_Ananas        943 IGPFNSS-KYNNLTK----------------ESDPLIPIRDSLGLLGAIVPKIANFYSSYH--LITHNQIVLK--KYLL
28_Liriodendron  947 IGPFNGS-KSHKVTKESI--------------KEDPMIPIRNSLGPLGT-VPKIANFYSSYY--LITHNQILLN--KYLL
30_Magnolia      947 IGPFNGS-KSHKVTKESI--------------KEDPMIPIRNSLGPLGT-VSKIANFYSSYY--LITHNQILLN--KYLL
32_Nymphaea      953 IN----VSKYHNVTKESIKE------------KEDTPIPILNLLGPLGT-VPKIHNFSPSYH--SITHNEILLNKYLILD
33_Amborella     926 MSPFKDV-QYSNGIKESI--------------KVEPLIPIRNSLGPLGT-SSQIENFYY----LLKTHNQISVT--KYLQ
35_Picea         790 IVLFNDSKCYDTV-NKSN--------------REDPMRKIIEFSGLLGHLH-SITNRFPSS--HFLTYKKVLSKKHS--I
44_Ginkgo        933 TVLFSGSKYYDTV-KRSI--------------QEDPMMQIIELSGLLGNLH-SIANRFPSP--HLITYNKVLSNKHS--I
51_Physcomitrium 889 TILFDKTKQNLKIENNVEKLFTYEPKKIIKTFNIEKRKNFVEFLGLLGYLQN-ITKSFQLFSCKKFSDKSI---PINFSI
```

```
T. thermophilus 1251 ------------------------------------------------------------------------------
E. coli         1071 GKDLRPALKIV--------DAQG----------------------------------------------------------
0_Nostoc         924 GDLLVAGT---EVATGIFTEESGQVTNVK-----------------------------------------KLG-VKSEELGV
1_Litchi        1017 LDNLKQTFQIC---KFYLMDENGRIYNPDPDSKIVLNPFNLNWYFLHH--------NYCEE--MSTIISLGQFICENVCI
2_Arabidopsis   1004 LDNLKYIFQKI---NSYLIDENGIILNLDPYSNVVLNPFKLNWYFLHQNYHHNYCE----E--TSTIISLGQFFCENVCI
3_Gossypium     1012 LDNLKQAFQVL---NYYLIAENGRIYNFDPCRNIFLNAVNLNWYFPHHHYHH----NYCEE--TSTIISLGQFICENVCI
5_Ricinus       1020 LDNLKQTFRVL---KYYLMDENGRVVNPDPCSNSVLNPFNLNWYFLHHNYHHNYCHNYCEE--SFTIISLGQFICENVCM
6_Rosa          1009 LDNLKQTFQVF---KYYLMDENGRISNSDPCSSILLNPFNLNWHFLDH--------NYCEE--TSTIISLGQFICENLCI
9_Cucumis       1011 LDNLKQTFQVL---KYYLMDENGGIFNSDPCSNIVFNTFNLNWHFLHHNYHNNYCEET--P--TRTRISLGHFFFENVCI
11_Nicotiana    1011 LDNLKQTFQVIK-FKYYLMDENGKIFNPDPCRNIILNPFNLNWYFLHH--------NYCEE--TSKIISLGQFICENVCI
13_Syringa      1019 LDNLKQTFQVI---KYYLMDENEKIYNPEPGSNIILNPFNLNWYFLHH--------NYCQE--TSTIISLGQFICENVCI
18_Liquidambar  1015 LDNLKQTFQVL---NYYLMDENGRIYNPDPCSNIILNPFNLNWYFLHH--------NYCEE--TSTIISLGQFICENVCI
19_Papaver      1010 LDNLKQTFTF-QGLKYYLLDETGRIYNPNLGSHIILNPLNLNWFFLQH--------DYCEE--RATIINLGQFICENVCI
20_Ananas       1001 LDNLKQIFQVLQVLKYCLIDENRRIYNPDPCSNIILNPFHLNWCFFHH--------DYCEE--TSTKISLGQFICENVCL
28_Liriodendron 1007 LDNLKQTFQVL---KYYLMDENGRIYNPNLHSNIIFNPFDLNWCFLRH--------DYCEE--TSTIISLGQFICENVCI
30_Magnolia     1007 LDNLKQTSQVL---KYYLMDENGRIYNPDPRSNIIFNPFDLNWCFLRH--------DYCEE--TSTIISLGQFICENVCI
32_Nymphaea     1014 NNNPKQTFQLL---KYFLVDENGRISNANPCSDIIFNLFGS--CFLPH--------DYCKGTSTTRIISLGQFICENVCL
33_Amborella     984 LDNFTQPFQVL---QYYLMDENGIVYNSDPCSNTRLNPFNLNWHFFHNNYDNNYYQK------KSPIISLGRFFCENVCI
35_Picea         850 FHN---SFNTFQVPKYYFMDENTRISHFDPCRNIISNLLGPNWCSSSS--------EFCKK--IFPVVSPGQLIPESVCI
44_Ginkgo        993 SDN---SGKVSQVSKCYFMGGNTGILNFDSCRNIIFNLFNSNWCSPLS--------NFCKK--KLPAVSLGQLIRESVCI
51_Physcomitrium 965 IDNLKKKIK---ISKWFFLNENKKVQKFFLTQNTILSL--LNWSFPIF--------DLAKK--KTQLFNLGHFFCDGLSI
```


Trigger loop helix2
β'b9:D1251-V1281
T. thermophilus 1251 ----------DITQGLPRVIELFEARR
E. coli 1086 ----------NDVLI------PGTDMPAQYFLPGKAIVQLEDGVQISSGDTLARIPQESGGTKDITGGLPRVADLFEARR
0_Nostoc 961 NSETPNSS-----LQ-TQNYAITIRLGRPYRVSPGAVLQIEDGDLVQRGDNLVLLVFERAKTGDIIQGLPRIEELLEARK
1_Litchi 1084 ATNGPHLK-SGQVLIVQ-VGSVVIRSAKPYLATPGATVHGHYGEILYEGDTLVTFIYEKSRSGDITQGLPKVEQVLEVRS
2_Arabidopsis 1075 AKKEPHLK-SGQVLIVQ-RDSAVIRSAKPYLATPGAKVHGHYSEILYEGDTLVTFIYEKSRSGDITQGLPKVEQVLEVRS
3_Gossypium 1083 AKSGPRLK-SGQVFIVQ-ADSIVIRSAKPYLATPGATVHGHYGETLYEGDTLVTFIYEKSRSGDITQGLPKVEQVLEVRS
5_Ricinus 1095 AKNGPHLK-SGQVIIIH-IGSVVIRSAKPYLATPGATVHGHYGEILYEGDTLVTFIYEKSRSGDITQGLPKVEQVLEVRS
6_Rosa 1076 AKKGSPLK-SGQVIVVQ-LDSLVIRSAKPYLATPGATVHGHYGEILSEGDTLVTFIYEKSRSGDITQGLPKVEQVLEVRS
9_Cucumis 1084 AKNRPHLK-SGQIIIVE-VDSVVIRSAKPYLATSGATVHRHYGEMLYEGDTLVTFIYEKSRSGDITQGLPKVEQVLEVRS
11_Nicotiana 1080 AKNGPPLK-SGQVILVQ-VDSIVIRSAKPYLATPGATVHGHYGETLYEGDTLVTFIYEKSRSGDITQGLPKVEQVLEVRS
13_Syringa 1086 AKNTPHLK-SGQVILVQ-VDSVVIRSAKPYLATPGATVHGHYGEILYEGDTLITFIYEKSRSGDITQGLPKVEQVLEVRS
18_Liquidambar 1082 AKNGPHLK-SGQVLIVQ-VDSVVIRSAKPYLATTGATVHGHYGEILYEGDTLVTFIYEKSRSGDITQGLPKVEQVLEVRS
19_Papaver 1079 SKYGPRLK-AGQVLIIR-VGSLVIRSAKPYLATPGATVHGHYGETLSEGDTLVTFIYEKSRSGDITQGLPKVEQVLEVRS
20_Ananas 1071 FKYEPHVKKSGQILIVN-VDSLVIRSAKPYLATPGATVHGHYGKILYEGDTLVTFIYEKSRSGDITQGLPKVEQILEVRS
28_Liriodendron 1074 SKYGPHIK-SGQVLIVH-VDSLVIRSAKPHLATPGATVHGHYGEILSEGDTLVTFIYEKSRSGDITQGLPKVEQVLEVRS
30_Magnolia 1074 SKYGPHIK-SGQVLIVH-VDSLVIRSAKPHLATPGATVHGHYGEILYEGDTLVTFIYEKSRSGDITQGLPKVEQVLEVRS
32_Nymphaea 1081 SKHRTRIK-SGQVIMVY-LDSFIIRSAKPYLATRGATVHGDYGEIFYEGDTLVTFIYEKSRSGDITHGLPKVEQVLEVRS
33_Amborella 1055 VKHGPHLK-SGQVIIVQ-IDSVVIRLAKPYLATPGATVHGHYGEKLSEGDTLVTFIYEKSRSGDITQGLPKVEQVLEVRS
35_Picea 917 SEDEPLPE-SGQIIAVD-EESLVIRSAKPYLATRKATVHGHYGEIIDKGDTLITLIYERFKSSDIIQGLPKVEQLSEARL
44_Ginkgo 1060 SEDKPLSG-SGQIIAVH-EEYLVIRSAKPYLATRRAAVHGHYGEFLDEGDTLITLIYERSKSGDTIQGLPKVEQLSEARS
51_Physcomitrium 1030 AEYPTFSE-SGQIIAIYDDLSLVIRLAKPYLATGGAIIHNNYGEIVKEGDILITLIYERLKSGDIIQGLPKVEQLLEARL
β'b9: D1251-V1281
β'b10:V1313-N1404
T. thermophilus 1268 PKAKAVISEIDGVVRIEET--EEKLSV-FVE-S-EGFSKEYKLPKEARLLVKDGDYVEAGQPLTRGAIDPHQLLEAKGP-
E. coli 1150 PKEPAILAEISGIVSFGKE--TKGKRRLVITPVDGSDPYEEMIPKWRQLNVFEGERVERGDVISDGPEAPHDILRLRGV-
0_Nostoc 1035 PKEACILARRAGEVKVVYGDGDEAIAIKVV--ESNGVVTDYPLGPGQNLIVPDGSHISAGQPLTDGPSNPHEILEIFFSL
1_Litchi 1162 IDSISMNLEKRV------E------------GWNA----------RITRIL
2_Arabidopsis 1153 IDSISLNLEKRI------K------------GWNK----------CITRIL
3_Gossypium 1161 IDSISMNLEKRI------E------------GWNE----------CITRIL
5_Ricinus 1173 IDSISINLEKRV------G------------GWNE----------CIPRIL
6_Rosa 1154 IDSISMNLEKRV------E------------GWNE----------CITRIL
9_Cucumis 1162 IDSISMSLEKRI------E------------GWNE----------RITRIL
11_Nicotiana 1158 VDSISMNLEKRI------E------------GWNK----------CITRIL
13_Syringa 1164 IDSISMNLEKRV------E------------GWNE----------RITRIL
18_Liquidambar 1160 IDSISMNLEKRV------E------------GWND----------RITRIL
19_Papaver 1157 LDSISMNLEKRV------E------------GWNE----------RITRIL
20_Ananas 1150 IDSLSMNLERRV------E------------GWNE----------RIPRIL
28_Liriodendron 1152 IDSISMNLEKRI------E------------GWNE----------HITRIL
30_Magnolia 1152 IDSISMNLEKRI------E------------GWNE----------RITRIL
32_Nymphaea 1159 IDSISMNLEKRV------E------------GWNE----------HITGIL
33_Amborella 1133 IDSISMNLEKRV------E------------GWND----------LLTRIL
35_Picea 995 NNSISMNLKESF------E------------NWTG----------DMTRFL
44_Ginkgo 1138 INPIPRNLEESF------E------------DWNE----------DMTRSL
51_Physcomitrium 1109 TNPVSINLEKGF------G------------EWNK----------DMTNFF
β'a17:L1348-K1354
β'a18:D1365-L1389
β'a19:L1395-E1401
β'b10:V1313-N1404
T. thermophilus 1342 ----------EAVERYLVEEIQKVYRAQGVKLHDKHIEIVVRQMMKYVEVTDPG-DSRLLEGQVLEKWDVEALNER
E. coli 1227 ----------HAVTRYIVNEVQDVYRLQGVKINDKHIEVIVRQMLRKATIVNAG-SSDFLEGEQVEYSRVKIANRE
0_Nostoc 1113 GSEDGVYACASHALQKVQTFLVNEVQMVYQSQGIDISDKHIEVIVRQMTNKVRIDD-GGDTTMLPGELVELRQVEQVNEA
1_Litchi 1185 GIPWGFLIGAELTIVQSRISLVNKIQKVYRSQGVQIHNRHIEIIVRQITSKVLVSEDGMSNVFLPGELIGLLRAERTGRA
2_Arabidopsis 1176 GIPWGFLIGAELTIVQSRISLVNKIQKVYRSQGVQIHNRHIEIIVRQITSKVLVSEEGMSNVFLPGELIGLLRAERTGRA
3_Gossypium 1184 GIPWGFVIGAELTIVQSRLSLVNKIQKVYRSQGVQIHNRHIEIIVRQITSKVLVSEDGMSNVFLPGELIGLLRAERTGRA
5_Ricinus 1196 GIPWGFLIGTELTIVQSRISLVNKIQRVYRSQGVQIHNRHIEIIVRQITSKVLVSEDGMSNVFSPGELIGLLRAERTGRA
6_Rosa 1177 GIPWGFLIGAELTIAQSRISLVNKIQKVYRSQGVQIHNRHIEIIVRQITSKVLVSEDGMSNVFSPGELIGLLRAERTGRA
9_Cucumis 1185 GIPWGFLIGSELTIVQSRISLVNKIQKVYRSQGVEIHNRHIEIIVRQITSKVLVSEDGMSNVFSPGELIGLLRAERTGRA
11_Nicotiana 1181 GIPWGFLIGAELTIAQSRISLVNKIQQVYRSQGVQIHNRHLEIIVRQITSKVLVSEDGMSNVFSPGELIGLLRAERMGRA
13_Syringa 1187 GMPWGFLIGAELTIVQSRISLVNKIQKVYRSQGVQIHNRHIEIIVRQITSKVLVSEDGMSNVFSPGELIGLLRAERMGRA
18_Liquidambar 1183 GIPWGFLIGAELTIVQSRISLVNKIQKVYRSQGVQIHNRHIEIIVRQITSKVLVSEDGMSNVFSPGELIGLLRAERTGRA
19_Papaver 1180 GIPWGFLIGAELTIAQSRISLVNKIQKVYRSQGVQIHNRHIEIIVRQITSKVLVSEDGMSNVFLPGELIGLLRAERTGRA
20_Ananas 1173 GIPWGFLIGAELTIAQSCISLVNKIQKVYRSQGVQIHNRHIEIIVRQVTSKVLVSEDGMSNVFSPGELIGLLRAERAGRA
28_Liriodendron 1175 GIPWGFLIGAELTIAQSRISLVNKIQKVYRSQGVQIHNRHIEIIVRQITSKVLVSEDGMSNVFSPGELIGLLRAERTGRA
30_Magnolia 1175 GIPWGFLIGAELTIAQSRISLVNKIQKVYRSQGVQIHNRHIEIIVRQITSKVLVSEDGMSNVFSPGELIGLLRAERTGRA
32_Nymphaea 1182 GIPWGFLIGAELTIAQSRISLVNKIQKVYRSQGVQIHNKHIEIIVRQITSKVLVSEDGMSNVFSPGELIGLLRAERAGRA
33_Amborella 1156 GIPWGFLIGAELTIVQSRISLVNKIQKVYRSQGVQIHNRHIEIIVRQITSKVVVSEDGMSNVFLPGELIGLLRAERTGRA
35_Picea 1018 GSLWGLFISARITMEQSQIHLVNQIQKVYRSQGVRICDKHIEIIVRQMTSKVLISEDGMANVFSPGELIGLSRAQRMDRA
44_Ginkgo 1161 GSLWGLFISARITMEQSQIHLVNQIQKVYRSQGVRIYDKHIEIIVRQMTSKVFISGDGMADVFSPGELIELSRAQRMNRA
51_Physcomitrium 1132 GSLWGYFLSAQISMEQSQVNLVNQIQKVYRSQGVNISDKHIEIIVRQMTSKVFTLEDGMTNGFLPGELIEFARAKRMNRA


Clamp
Switch 5
β'a20:F1440-G1477
β'b11:V1415-G1477
T. thermophilus 1407 LIAEGKTPVAWKPLLMGVTKSALSTKSWLSAASFQNTTHVLTEAAIAGKKDELIGLKENVILGRLIPAGTGSDFVRFTQV
E. coli 1292 LEANGKVGATYSRDLLGITKASLATESFISAASFQETTRVLTEAAVAGKRDELRGLKENVIVGRLIPAGTGYAYHQDRMR
0_Nostoc 1192 MAITGGARAQYTPVLLGITKASLNTDSFISAASFQETTRVLTEAAIEGKSDWLRGLKENVIIGRLIPAGTGYNTYEETSA
1_Litchi 1265 LE----EAIRYRAILLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGMIPVGTGF-KGLVHCS
2_Arabidopsis 1256 LE----EAICYRAVLLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGVIPAGTGFNKGLVHCS
3_Gossypium 1264 LE----EAICYRAVLLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGMIPAGTGF-KGLVHRS
5_Ricinus 1276 LE----EAICYRAILLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGMIPVGTGF-KGLVQGS
6_Rosa 1257 LE----EAICYRAILLGITKASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGMIPVGTGF-KGFVPRS
9_Cucumis 1265 LE----EAICYRAVLLGITKASLNTQSFISEASFQETARVLAKAALRGRIDWLRGLKENVVLGGMIPVGTGF-RELAHRS
11_Nicotiana 1261 LE----EAICYRVVLLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGVIPVGTGF-KGLVHPS
13_Syringa 1267 LE----EAVCYRAVLLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGMIPLGTGL-KGLVPPS
18_Liquidambar 1263 LE----EAICYRAILLGITKASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGMIPVGTGF-KGLVHHS
19_Papaver 1260 LE----EAICYRAVLLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGMIPVGTGF-KGLVYHS
20_Ananas 1253 LD----ESICYRAILLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGIIPVGTGF-QKFVHRS
28_Liriodendron 1255 LE----EGICYRAILLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGMIPVGTGF-KGLVHRS
30_Magnolia 1255 LE----EAICYRAILLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGMIPVGTGF-KGLVHRS
32_Nymphaea 1262 LE----EAICYRAVLLGITRASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGMIPVGTGF-KRFVHRS
33_Amborella 1236 LE----EAICYRAILLGITKASLNTQSFISEASFQETARVLAKAALRGRIDWLKGLKENVVLGGMIPVGTGF-KGLVHCS
35_Picea 1098 LE----EAIYYQTMLLGITRASLNTQSFISEASFQETARVLAKAALQGRIDWLKGLKENVILGGIIPAGTGQ-H--IHRS
44_Ginkgo 1241 LE----EAIYYRTVLLGITRASLDTQSFISGASFQETARVLAKAALRGRIDWLKGLKENVILGGIIPAGTGF-KRFLRHS
51_Physcomitrium 1212 LE----EVIPYKPVLLGITKASLNTQSFISEASFQETTRVLAKAALRGRIDWLKGLKENVILGGIIPTGTGCEEVLWQIT

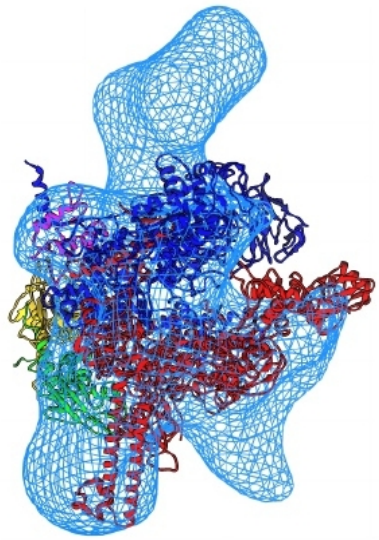

T. thermophilus 1487 VDQKTLKAIEEARK-E-AVEAKER---------PAARRGVK-----REQP-GKQA------------------------
E. coli 1372 RR-----AAGEAPAAP-QVTAEDA---------SASLAELL-----NAGLGGSDNE------------------------
0_Nostoc 1272 IDDY-ATDI-----SSSVLDEVDDPLDMVLDDRTARTYNLDAPTLGEPSYGSRRAERSILDDDDDLIADEVADEEDYEDD
1_Litchi 1340 R-QHNNILLETQKNT--LFGGV---RDILLHHRELFDFCIS----------KTLR-----DTSEQSL------------
2_Arabidopsis 1332 R-QHTNIILEKKTKNLALFEGDMR--DILFYHREFCDSSIS----------KSDFSRI--------------------
3_Gossypium 1339 R-QHNNILLETKKKN--FFGGEMR--DIFFHHRELFDSCIS----------NNLH-----DTSGRSF------------
5_Ricinus 1351 R-QYKNIPLKTKKNN--LFGGEFRDRDILFHHRELFYSCIS----------KNFY-----DTSEQSF------------
6_Rosa 1332 R-QHNNISLETKNKS--LFEGEMR--DILVHHRELFDFCIS----------KNLH-----DTSEQSF------------
9_Cucumis 1340 R-QHNNIPLEPPPKK--IFEGEMR--DILFHHKELFDFFIS----------TNLH-----DTSEQAF------------
11_Nicotiana 1336 K-QHNNIPLETKKKN--LFEGEMR--DILFHHKKLFDSCLS----------KNFH-----DIPEQSF------------
13_Syringa 1342 K-QDSNSPLETKKNN--LFEGEMR--DILFHHRKLFDSCLS----------KNFH-----DTSEQSF------------
18_Liquidambar 1338 R-QHNNIPLETKKKN--LFEGEMR--DILFHHRELFHSCIS----------KNFH-----DISEQSF------------
19_Papaver 1335 R-QHSNIPFEIKKNN--LFRGGFR--DILFQHKELFDSYIP----------KNIH-----DPSEQLF------------
20_Ananas 1328 R-QDKNIYLEIKRKN--LFELEMR--DILLHHRELFCSCAT----------NNFHETNLHETSEQSF------------
28_Liriodendron 1330 R-QHNNIPLEIKKKN--LFEGEMR--DILFHHRELLSSCIP----------KNFH-----DTSEQSF------------
30_Magnolia 1330 R-QHNNIPLEIKKKN--LFEGEMR--DILFHHRELLSSCIP----------KNFH-----DTSEQSF------------
32_Nymphaea 1337 R-EYNNIPLEIQKKN--FFGGEMR--DILFHHRELFCSCIP----------KPK---SFHNTSEQPF------------
33_Amborella 1311 R-KHNNIPLEPKNKN--LFEWEMR--DILFHHRELFSFCIS----------KNGTSSLFFTLKKKKRE-----EVWGEM
35_Picea 1171 G-KRNGMDPRMGNRN--LFSKKVK--DIFFHYHKVSFFSIQ----------ENSHNILKQPFK---------------
44_Ginkgo 1316 E-ERNKIDSRTGNKN--LFNNKVK--DIFSHHGKVSVSPIK----------DNYHNTLKQPLCENSVD-----K-----
51_Physcomitrium 1288 LEKQKNILLKKNK--SKLFHNKVK--DIFLYK-KLSISFTS----------EKIHKNY--------------------


T. thermophilus ----------
E. coli ----------
0_Nostoc 1346 DEDEDDFDDE----------
1_Litchi 1386 ----RGFNKS----------
2_Arabidopsis ----------
3_Gossypium 1386 ----IGIEFND--S----------
5_Ricinus 1400 ----IGFNDS----------
6_Rosa 1379 ----FGFNDS----------
9_Cucumis 1387 ----LGFNDS----------
11_Nicotiana 1383 ----IGFNDS----------
13_Syringa 1389 ----IGFNDS----------
18_Liquidambar 1385 ----MGFNDS----------
19_Papaver 1382 ----TGFNDS----------
20_Ananas 1380 ----MRFNDS----------
28_Liriodendron 1377 ----TGFNDS----------
30_Magnolia 1377 ----TGFNDS----------
32_Nymphaea 1386 ----YTMGSNP--ISGFIIS----------
33_Amborella 1370 TRRYWNINLEEMMEAGVHFGHGTKKWNPRMAPYISAKRKGIHIINLTRTARFLSEACDLVFDAASRGKQFLIVGTKNKAA
35_Picea ----------
44_Ginkgo ----------
51_Physcomitrium ----------

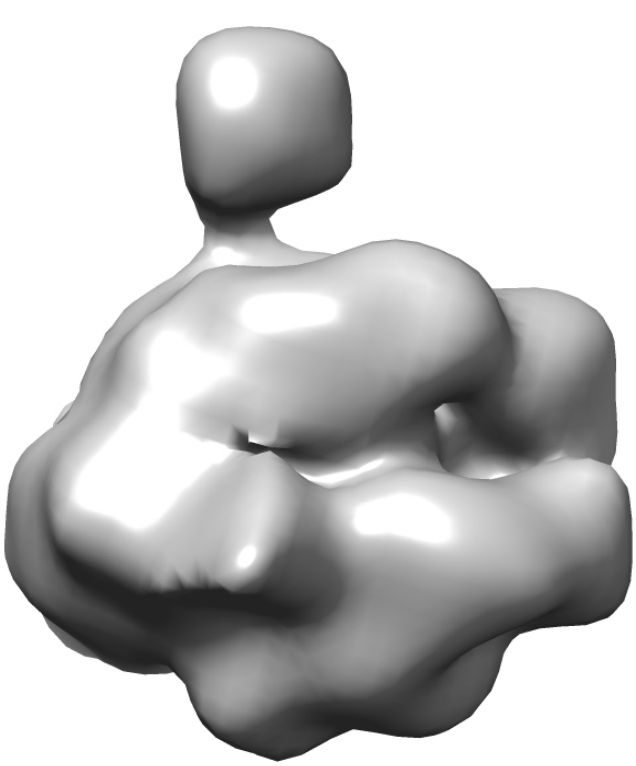
```
T. thermophilus      ------------------------------------------------------------------------
E. coli              ------------------------------------------------------------------------
0_Nostoc             ------------------------------------------------------------------------
1_Litchi             ------------------------------------------------------------------------
2_Arabidopsis        ------------------------------------------------------------------------
3_Gossypium          ------------------------------------------------------------------------
5_Ricinus            ------------------------------------------------------------------------
6_Rosa               ------------------------------------------------------------------------
9_Cucumis            ------------------------------------------------------------------------
11_Nicotiana         ------------------------------------------------------------------------
13_Syringa           ------------------------------------------------------------------------
18_Liquidambar       ------------------------------------------------------------------------
19_Papaver           ------------------------------------------------------------------------
20_Ananas            ------------------------------------------------------------------------
28_Liriodendron      ------------------------------------------------------------------------
30_Magnolia          ------------------------------------------------------------------------
32_Nymphaea          ------------------------------------------------------------------------
33_Amborella    1450 DSVAGAAIKARCHYVNKKWLGGMLTNWYTTETRLHKFRDLRTEQKTGRLNRLPKRDAAVLKRQLSHLQTYLGGIKYMTGL
35_Picea             ------------------------------------------------------------------------
44_Ginkgo            ------------------------------------------------------------------------
51_Physcomitrium     ------------------------------------------------------------------------

T. thermophilus      ------------------------------------------------------------------------
E. coli              ------------------------------------------------------------------------
0_Nostoc             ------------------------------------------------------------------------
1_Litchi             ------------------------------------------------------------------------
2_Arabidopsis        ------------------------------------------------------------------------
3_Gossypium          ------------------------------------------------------------------------
5_Ricinus            ------------------------------------------------------------------------
6_Rosa               ------------------------------------------------------------------------
9_Cucumis            ------------------------------------------------------------------------
11_Nicotiana         ------------------------------------------------------------------------
13_Syringa           ------------------------------------------------------------------------
18_Liquidambar       ------------------------------------------------------------------------
19_Papaver           ------------------------------------------------------------------------
20_Ananas            ------------------------------------------------------------------------
28_Liriodendron      ------------------------------------------------------------------------
30_Magnolia          ------------------------------------------------------------------------
32_Nymphaea          ------------------------------------------------------------------------
33_Amborella    1530 PDIVIIVDQQEEYTALRECITLGIPTICLIDTNCDPDLADISIPANDDAIASIRLILNKLVFAILLYDISGVEVGQHFYW
35_Picea             ------------------------------------------------------------------------
44_Ginkgo            ------------------------------------------------------------------------
51_Physcomitrium     ------------------------------------------------------------------------

T. thermophilus      ------------------------------------------------------------------------
E. coli              ------------------------------------------------------------------------
0_Nostoc             ------------------------------------------------------------------------
1_Litchi             ------------------------------------------------------------------------
2_Arabidopsis        ------------------------------------------------------------------------
3_Gossypium          ------------------------------------------------------------------------
5_Ricinus            ------------------------------------------------------------------------
6_Rosa               ------------------------------------------------------------------------
9_Cucumis            ------------------------------------------------------------------------
11_Nicotiana         ------------------------------------------------------------------------
13_Syringa           ------------------------------------------------------------------------
18_Liquidambar       ------------------------------------------------------------------------
19_Papaver           ------------------------------------------------------------------------
20_Ananas            ------------------------------------------------------------------------
28_Liriodendron      ------------------------------------------------------------------------
30_Magnolia          ------------------------------------------------------------------------
32_Nymphaea          ------------------------------------------------------------------------
33_Amborella    1610 QIGGFQVHAQVLITSWVVIAILLGSAILAVRNPQTIPTDGQNFFEYVLEFIRDVSKTQIGEEYGPWVPFIGTLFLFIFVS
35_Picea             ------------------------------------------------------------------------
44_Ginkgo            ------------------------------------------------------------------------
51_Physcomitrium     ------------------------------------------------------------------------
```

```
T. thermophilus      ----------------------------------------
E. coli              ----------------------------------------
0_Nostoc             ----------------------------------------
1_Litchi             ----------------------------------------
2_Arabidopsis        ----------------------------------------
3_Gossypium          ----------------------------------------
5_Ricinus            ----------------------------------------
6_Rosa               ----------------------------------------
9_Cucumis            ----------------------------------------
11_Nicotiana         ----------------------------------------
13_Syringa           ----------------------------------------
18_Liquidambar       ----------------------------------------
19_Papaver           ----------------------------------------
20_Ananas            ----------------------------------------
28_Liriodendron      ----------------------------------------
30_Magnolia          ----------------------------------------
32_Nymphaea          ----------------------------------------
33_Amborella    1690 NWSGALLPWKIIELPHGELAAPTNDINTTVALALLTRFHKTLIT
35_Picea             ----------------------------------------
44_Ginkgo            ----------------------------------------
51_Physcomitrium     ----------------------------------------
```

Figure S6: view of the catalytic core from the *E. coli* RNAP (PDB entry: 3LU0 (Opalka *et al*., 2010)) manually fitted into the envelope of PEP using Chimera (Pettersen *et al*., 2004).



Figures S7a and S7b: overall shape of the a) human RNA polymerase II (EMDB entry: EMD-2194; Kassube *et al*., 2013) and b) yeast RNA polymerase III (EMDB entry: EMD-1753; Vanini *et al*., 2010) solved at 25 and 21 Å respectively.:



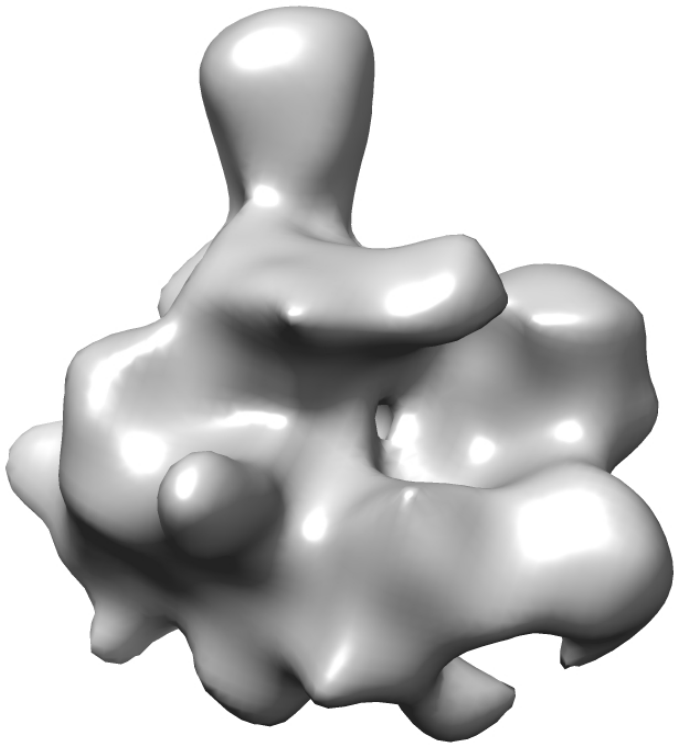


Figure S8: FSC curve for the PEP 3D reconstruction calculated between two independent half maps (gold standard FSC). The dotted line represents the FSC=0.143 cutoff used to determine the resolution.