

Identification of Essential Subunits in the Plastid-Encoded RNA Polymerase Complex Reveals Building Blocks for Proper Plastid Development¹[C][W][OA]

Sebastian Steiner, Yvonne Schröter, Jeannette Pfalz, and Thomas Pfannschmidt*

Junior Research Group, Department of Plant Physiology, Institute of General Botany and Plant Physiology, Friedrich-Schiller-University Jena, D-07743 Jena, Germany

The major RNA polymerase activity in mature chloroplasts is a multisubunit, *Escherichia coli*-like protein complex called PEP (for plastid-encoded RNA polymerase). Its subunit structure has been extensively investigated by biochemical means. Beside the “prokaryotic” subunits encoded by the plastome-located RNA polymerase genes, a number of additional nucleus-encoded subunits of eukaryotic origin have been identified in the PEP complex. These subunits appear to provide additional functions and regulation modes necessary to adapt transcription to the varying functional situations in chloroplasts. However, despite the enormous progress in genomic data and mass spectrometry techniques, it is still under debate which of these subunits belong to the core complex of PEP and which ones represent rather transient or peripheral components. Here, we present a catalog of true PEP subunits that is based on comparative analyses from biochemical purifications, protein mass spectrometry, and phenotypic analyses. We regard reproducibly identified protein subunits of the basic PEP complex as essential when the corresponding knockout mutants reveal an albino or pale-green phenotype. Our study provides a clearly defined subunit catalog of the basic PEP complex, generating the basis for a better understanding of chloroplast transcription regulation. In addition, the data support a model that links PEP complex assembly and chloroplast buildup during early seedling development in vascular plants.

Chloroplasts are the typical organelles of green plant cells, which originated from a cyanobacteria-like ancestor during endosymbiosis (Blankenship, 2002; Buchanan et al., 2002). They still possess many remnants of this prokaryotic origin, including its own genetic system. This consists of a plastid chromosome, the so-called plastome, and a fully functional transcriptional and translational apparatus for the expression of the genetic information on it. In vascular plants, the plastome contains a largely conserved set of 100 to 120 genes, including genes for photosynthesis proteins, genes for the RNA polymerase (*rpo* genes), and genes for ribosomal subunits and RNAs as well as for tRNAs (Sugiura, 1992). The vast majority of chloroplast proteins, however, are encoded in the nucleus

and must be imported from the cytosol (Abdallah et al., 2000; Soll and Schleiff, 2004). As a result, all multiprotein complexes in plastids are composed of a patchwork of plastid- and nucleus-encoded subunits. The core proteins of large complexes (for instance, of the photosystems) are usually encoded in the plastome, while peripheral subunits typically appear to be encoded in the nucleus. This distribution reflects two evolutionary tendencies that occurred during the establishment of endosymbiosis. First, most genes from the cyanobacteria-like ancestor were lost to the nucleus of the host cell, and essential proteins had to be reimported and assembled into the complexes. During evolution, this was easier achieved for peripheral than for core proteins, which usually represent the pace-makers for complex assembly. Second, the organelle also gained novel proteins from the eukaryotic host cell, which conferred new properties to the prokaryotic multienzyme complexes of the endosymbiont. Both strategies led to the transfer of a large proportion of developmental and functional control from the symbiont to the nucleus of the host cell and, by this means, led to a complete integration of the organelle into the cell (Martin et al., 2002; Stoebe and Maier, 2002; Herrmann et al., 2003; Greiner et al., 2011).

The evolutionary patchwork of chloroplast protein complexes becomes especially obvious in the plastid transcription machinery. Multiple lines of evidence indicate that the transcription of plastomic genes depends on the activity of a phage-type, single-subunit, nucleus-encoded plastid RNA polymerase (NEP) and

¹ This work was supported by grants from the Deutsche Forschungsgemeinschaft (to T.P. and the research group FOR 804), by the Freistaat Thüringen, and by the Bundesministerium für Bildung und Forschung (Nachwuchsgruppe “Pflanzliche Anpassung an Umweltveränderungen: Proteinanalyse Mittels MS”).

* Corresponding author; e-mail thomas.pfannschmidt@uni-jena.de.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Thomas Pfannschmidt (thomas.pfannschmidt@uni-jena.de).

[C] Some figures in this article are displayed in color online but in black and white in the print edition.

[W] The online version of this article contains Web-only data.

[OA] Open Access articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.111.184515

a prokaryote-type, multisubunit, plastid-encoded RNA polymerase (PEP; Hess and Börner, 1999; Cahoon and Stern, 2001; Lysenko and Kuznetsov 2005; Shiina et al., 2005; Liere et al., 2011). In *Arabidopsis thaliana*, NEP is encoded by two nuclear gene copies (*rpoTp* and *rpoTmp*), each with a different target sequence directing the encoded protein either to plastids or, via dual targeting, to plastids and mitochondria. A third gene product encoded by *rpoTm* is directed exclusively to mitochondria (Hedtke et al., 2000). The PEP subunits, in contrast, are encoded by a set of plastome-located genes (*rpoA* and the *rpoB/C₁/C₂* operon) that exhibit approximately 26% to 50% sequence homology to corresponding genes from cyanobacteria, generating the so-called core enzyme (Iglói and Kossel, 1992). This core enzyme is supplemented by a number of nucleus-encoded σ -factors that provide the necessary promoter specificity to the complex (Link, 1996; Allison, 2000; Schweer et al., 2010). PEP is the major RNA polymerase activity in mature chloroplasts and represents the predominant target for environmental regulation, such as light-induced redox control of chloroplast transcription (Link, 2003; Pfannschmidt and Liere, 2005).

Initially, the structure, identity, and subunit composition of the chloroplast transcription machinery were mainly investigated by biochemical means. In plastids, the DNA and its associated or interacting proteins (including the RNA polymerase) are organized in so-called nucleoids or plastid nuclei, very large structures that represent bacteria-like assemblies of several plastome copies and numerous proteins with various functions in nucleoid structure and gene expression. Recently, a microscopic study using a plastid envelope DNA-binding protein-GFP fusion described in detail the localization and distribution of nucleoids in plastids from different plant cell types (Terasawa and Sato, 2005). Purified nucleoids were very useful in the determination of gene-specific transcription activities, but due to the high number of proteins within the complex, a detailed subunit analysis of the RNA polymerases was not feasible (Sakai et al., 2004). Therefore, a number of different biochemical purification procedures were developed aiming to enrich more distinct RNA polymerase complexes from chloroplasts. Basically, two types of plastid RNA polymerase preparations can be distinguished. The first represents an insoluble RNA polymerase preparation called transcriptionally active chromosome (TAC), which can be precipitated by ultracentrifugation. It represents a high- M_r DNA/RNA-protein complex containing approximately 40 to 60 proteins that is capable of *in vitro* transcription, resembling the nucleoids in this respect (Hallick et al., 1976; Reiss and Link, 1985; Little and Hallick, 1988; Krause and Krupinska, 2000; Pfalz et al., 2006). The second type of preparation usually includes a detergent treatment, resulting in a soluble RNA-polymerase activity that requires externally added DNA for transcriptional activity. Many studies concentrated on these soluble preparations, since these allowed a precise molecular

analysis of the promoter specificity and cis-element usage of the purified transcription complex (Bradley and Gatenby, 1985; Lerbs et al., 1985; Rajasekhar et al., 1991; Lakhani et al., 1992; Pfannschmidt and Link, 1997).

Various biochemical purification procedures yielded highly purified RNA polymerase preparations that were able to recognize specifically the typical prokaryotic 10 and 35 promoter boxes of many plastid genes. However, these RNA polymerases did not exhibit the expected subunit structure 2α , β , β' , and β'' , resembling that of the *Escherichia coli* enzyme (2α , β , β'), but a much more complex structure composed of around 20 to 30 subunits. This apparent contradiction was resolved with the identification of the prokaryotic core subunits α , β , β' , and β'' in various soluble RNA polymerase preparations by using various experimental approaches, including western analysis, Edman degradation, and mass fingerprints (Hu and Bogorad, 1990; Hu et al., 1991; Pfannschmidt et al., 2000). Interestingly, these subunits were also found in the TAC, indicating that TAC and soluble RNA polymerases represent two different biochemical preparations of the same complex rather than two separate RNA polymerase classes, as originally assumed (Little and Hallick, 1988; Suck et al., 1996). In addition, it turned out that the PEP enzyme undergoes a structural reorganization during light-dependent chloroplast maturation. In etioplasts or young greening chloroplasts, PEP displays the expected *E. coli*-like structure but is reorganized into a much more complex "eukaryote"-like RNA polymerase in mature chloroplasts (Hu and Bogorad, 1990; Hu et al., 1991; Pfannschmidt and Link, 1994; Pfannschmidt et al., 2000). This probably involves a number of still unknown posttranslational modifications of the *rpo* subunits, since (1) these exhibit differences in the apparent M_r between etioplast and chloroplasts and (2) the PEP enzyme appears to change its promoter recognition properties during the etioplast-chloroplast transition (Pfannschmidt and Link, 1997). The recruitment of further subunits with additional enzymatic activities has been interpreted as an evolutionary adaptation of the RNA polymerase complex and its functions to the specific conditions in the chloroplast (Link, 1996; Pfannschmidt and Liere, 2005). This, so far, is the best explanation for why even highly purified RNA polymerase preparations from chloroplasts of several species exhibit approximately 10 to 15 proteins in addition to the *rpo* subunits (Rajasekhar et al., 1991; Khanna et al., 1992; Lakhani et al., 1992; Pfannschmidt and Link, 1994; Rajasekhar and Tewari, 1995; Boyer and Hallick, 1998; Suzuki et al., 2004). The general criticism that these additional proteins may simply represent contaminants of the biochemical purification procedures has been recently invalidated by an elegant transplastomic approach in which the *rpoA* gene was fused to a His tag. The PEP enzyme from tobacco (*Nicotiana tabacum*) chloroplasts could be then purified via nickel-affinity

chromatography. Even this affinity tag-purified RNA polymerase preparation revealed a highly complex subunit composition, indicating that the additional subunits copurify due to an interaction with the *rpo* subunits and/or associated non-*rpo* subunits and therefore belong directly to the complex (Suzuki et al., 2004).

To understand chloroplast transcription and its regulation, it is necessary to identify all additional PEP subunits and characterize its potential functions. Using modern mass spectrometry, a number of the non-*rpo* subunits have been identified in the last few years, but several subunits still remained unknown (Pfannschmidt et al., 2000; Suzuki et al., 2004; Schröter et al., 2010). Furthermore, the highly varying subunit composition of the different transcription complexes described above suggests that the RNA polymerase represents a dynamic protein complex with many subunits only transiently attached. This raises the question of which subunits represent true and essential components of the basic RNA polymerase complex. In order to answer it, we have performed mass spectrometry with all subunits of PEP preparations from mustard (*Sinapis alba*) chloroplasts after heparin-Sepharose (HS) chromatography and blue native two-dimensional gel electrophoresis. We aimed to determine those subunits that can be reproducibly purified in order to distinguish between permanent and transient protein components. We identified all *rpo* subunits, including one novel variant of RpoC1 and 10 additional proteins. Combining these biochemical data with phenotypic analyses of corresponding knockout mutants from Arabidopsis, we could define the essential subunits of the basic PEP complex and present a comprehensive catalog of its components. A potential role of PEP subunit assembly as a decisive checkpoint in chloroplast development is discussed.

RESULTS

Basic PEP Subunit Composition as Defined by Biochemical Purification and Mass Spectrometry

We used 7-d-old white light-grown mustard seedlings as a source for PEP preparations, as reported earlier (Tiller and Link, 1993). Intact chloroplasts were isolated from cotyledons by homogenization and Suc gradient centrifugation, lysed in a buffer containing the nonionic detergent Triton X-100, and transcriptionally active fractions were subsequently enriched by HS chromatography. Comparable preparations have been partially characterized earlier and contain RNA polymerases, σ -factors, several DNA- and RNA-binding proteins, DNA polymerase, and kinase activities (Tiller and Link, 1993; Pfannschmidt and Link, 1994; Baginsky et al., 1997). The PEP enzyme was then further purified from such fractions by two-dimensional (2D) blue native (BN)-PAGE, as recently described (Schröter et al., 2010). We took advantage of the observation that the

PEP complex possesses a size of more than 1,000 kD, being by far the largest protein complex in the HS fractions. Due to this large size, the protein complex displays very slow migration behavior in BN-PAGE. No other proteins or protein complexes from the HS fractions were observed to migrate in this area of the gel. The subunit composition of the PEP complex was revealed by subsequent separation on a denaturing second dimension, producing a distinct ladder of protein subunits (Fig. 1) that can be clearly distinguished from background bands or staining artifacts due to its perpendicular arrangement and characteristic spot shape. Theoretically, some single proteins could be accidentally retained in this gel area because of technical inconsistencies, such as unspecific retardation within the PEP complex during the separation or because of biological variations in the plant material. In order to exclude these possibilities, we analyzed three different protein purifications, each prepared from an independent biological replicate. Only proteins that reproducibly occurred in all preparations were regarded as candidates for true components of the complex. In addition, this list of subunits was compared with that of

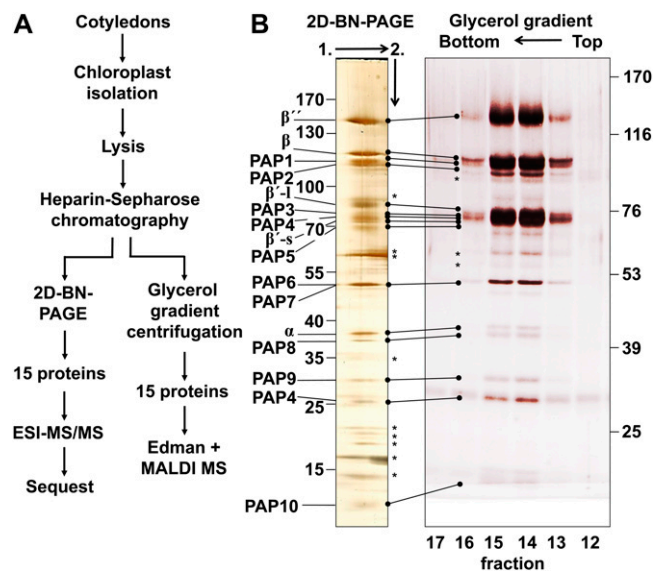


Figure 1. Comparison of subunit composition of the plastid RNA polymerase from mustard after 2D BN-PAGE and glycerol gradient centrifugation. A, Purification schemes and resulting proteins. B, PEP subunit composition obtained by 2D BN-PAGE (left, large gel, 7%–17%) and SDS-PAGE after glycerol gradient centrifugation (right, mini gel, 5%–15%). Two representative gels are shown. Total protein (150 μ g) was separated and fixed, and proteins were stained with silver. Running directions of the first and second dimensions are indicated by arrows. Sizes of marker proteins separated in parallel on the same gel are given in the margins. Single subunits within the PEP complexes that gave significant hits in the databases are indicated by consecutive numbering. Corresponding proteins within the two preparations are connected by lines. Asterisks mark proteins not reproducibly found in the complexes. For identity and detailed data of mass spectrometry, see Table I and Supplemental Table S1. MALDI, Matrix-assisted laser-desorption ionization. [See online article for color version of this figure.]

highly purified PEP preparations after glycerol gradient centrifugation (Pfannschmidt and Link, 1994). In glycerol gradient centrifugation, the large PEP complex exhibits the fastest sedimentation of all protein complexes in the HS fractions and, therefore, can be easily separated from smaller complexes or single proteins. Only protein bands that appeared in both preparations were regarded as permanent PEP subunits. By this means, 15 different protein bands were reproducibly identified in the PEP complex (Fig. 1), which were then analyzed by mass spectrometry. The respective protein spots were cut out, subjected to in-gel tryptic digestion, and peptide masses were determined by electrospray ionization-tandem mass spectrometry (ESI-MS/MS).

We identified the mustard proteins by the masses of the homolog peptides in the Arabidopsis sequence or other species in the Brassicales database (Table I; Supplemental Table S1). In total, we measured 15 spots and identified 16 distinct protein sequences. We could confirm the identification of all subunits recently found by mass spectrometry in the mustard PEP (Loschelder et al., 2004; Schröter et al., 2010) but also found three novel components not described yet as PEP subunits. In particular, we found all *rpo* gene products (α , β , β' , β'') representing the “classical” core of the PEP complex. The β'' -subunit (encoded by *rpoC₂*) was the largest subunit at 141 kD, followed by the β -subunit (encoded by *rpoB*) at 118 kD. The β' -subunit (encoded by *rpoC₁*) was found at around 85 kD and, unexpectedly, in a second, smaller variant at about 72 kD. All RpoC₁ peptides detected in our mass spectrometric measure-

ments were found for both proteins, with only one exception. This special peptide occurred only among those detected from the larger β' -variant and is located approximately in the middle of the RpoC₁ protein sequence (Fig. 2). This and the wide distribution of the identified peptides in the sequence suggest that the smaller β' -variant is a genuine gene product rather than a result of degradation. For a defined assignment, we named these two variants β' -l and β' -s (for large and small, respectively). The α -subunit (encoded by *rpoA*) was identified at 38 kD, which matches precisely the predicted size of 38 kD (Igloi and Kossel, 1992).

A second group of proteins identified here is composed of PTAC2, -3, -6, -10, -12, and -14, at apparent masses of 107, 110, 37, 76, 70, and 52 kD, respectively. All were described to be part of the transcriptionally active chromosome (Pfalz et al., 2006). PTAC2, -3, -10, and -14 contain a number of diverse functional domains related to DNA/RNA binding or interaction. These domains, however, are mainly characterized by domain prediction, and true functional assignments based on experimental evidence are lacking. PTAC6 is the most enigmatic PEP subunit, since it contains no known protein motif and any experimental clue to its potential function is missing (Table I). PTAC12 has not been described yet as a subunit of the soluble PEP complex. It has been reported to be potentially involved in protein degradation (Table I; Chen et al., 2010); however, this function was mainly attributed to its nuclear localization. In all cases, the apparent M_r values were close to the predicted theoretical ones.

Table I. RNA polymerase subunits identified by ESI-MS/MS after 2D BN-PAGE

Subunit, PAPs as given in Figure 1; AGI Accession No., Arabidopsis Genome Initiative gene accession numbers; Mass (kD; theor. cTP/app.), theoretical molecular mass without chloroplast transit peptide and apparent molecular mass observed on the gel; Identity/Protein Domain, identity of PAP and its predicted protein domain(s) as obtained by the Conserved Domain Database (Marchler-Bauer et al., 2011); Function, subunit functions predicted from subdomains or proposed/shown by experiment (ex.); Reference, source for functional classification.

Subunit	AGI Accession No.	Mass (kD; theor. – cTP/app.)	Identity/Protein Domain	Function	Reference
RpoC ₂	AtCg00170	156/141	β'' subunit	DNA binding	Igloi and Kossel (1992)
RpoB	AtCg00190	121/118	β subunit	RNA synthesis	Igloi and Kossel (1992)
PAP1	At3g04260	110/110	pTAC3/SAP domain	DNA/RNA binding	Prediction
PAP2	At1g74850	89/107	pTAC2/PPR repeat, SMR domain	RNA metabolism	Prediction
RpoC ₁	AtCg00180	80/85	β' -l subunit	Unknown	Igloi and Kossel (1992)
PAP3	At3g48500	77/76	pTAC10/S1-like domain	RNA binding	Prediction
PAP4	At5g23310	26/74	Iron superoxide dismutase 3	Superoxide detoxification (ex.)	Myouga et al. (2008)
RpoC ₁	AtCg00180	80/72	β' -s subunit	Unknown	Igloi and Kossel (1992)
PAP5	At2g34640	57/70	pTAC12-HEMERA/RAD23 domain	Probably involved in ubiquitin-mediated proteolysis (ex.)	Chen et al. (2010)
PAP6	At3g54090	49/52	FLN1/pfkB-2 fructokinase	Fru phosphorylation; substrate specificity not detectable (ex.)	Prediction; Arsova et al. (2010)
PAP7	At4g20130	49/52	pTAC14/SET domain, Rubisco substrate binding	Lys methyltransferase, interaction with histones or Rubisco	Prediction
RpoA	AtCg00740	38/38	α -Subunit	Complex stabilization	Igloi and Kossel (1992)
PAP8	At1g21600	31/37	pTAC6	Unknown	Unknown
PAP9	At5g5110	29/29	Iron superoxide dismutase 2	Superoxide detoxification (ex.)	Myouga et al. (2008)
PAP4	At5g23310	26/26	Iron superoxide dismutase 3	Superoxide detoxification (ex.)	Myouga et al. (2008)
PAP10	At3g06730	12/13	TRX-z/thioredoxin WCGPC motif	Redox regulation (ex.)	Arsova et al. (2010)

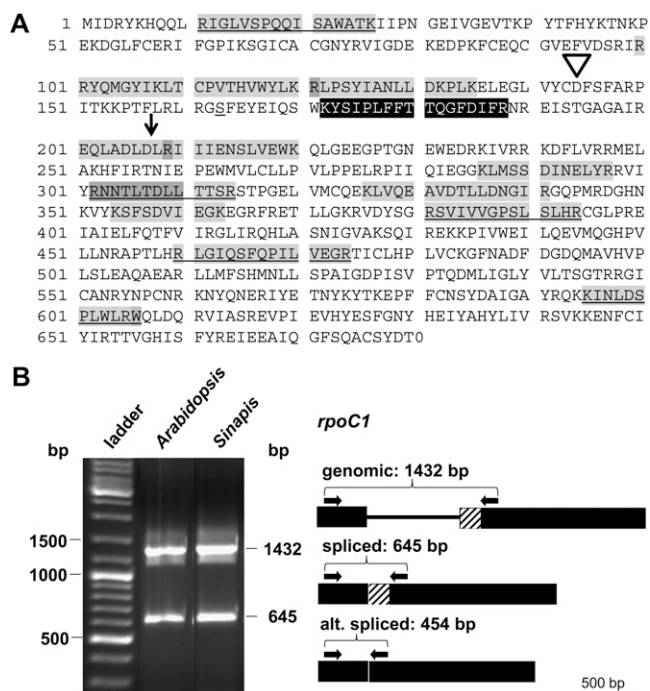


Figure 2. Analysis of the novel β' -s subunit. A, Detected peptides of the two β' -subunits. Peptides identified by mass spectrometry are given within the *RpoC₁* amino acid sequence. Gray background, peptides identified in both subunits. Overlaps between neighboring peptides are indicated in dark gray. Gray background underlined, peptides of β' -s. Black background with white letters, peptide solely identified in the β' -l subunit. Positions of the spliced intron (triangle), the editing site (underlined S), and the proposed alternative splice acceptor site (arrow) are indicated. B, Results of RT-PCR analysis and corresponding gene models indicating potential splice variants of *rpoC₁*. Left panel, ethidium bromide-stained RT-PCR products using *Arabidopsis* and *Sinapis* cDNA and *rpoC₁*-specific primers. The top band represents a product generated from the genomic DNA, as depicted in the right panel. The bottom band represents a product generated from the spliced RNA. Right panel, gene model and splice variants. The coding sequence is given as the thick black bar and the intron as the thin black bar. A potential alternatively spliced area is given as the hatched bar.

A third group of subunits is composed of proteins that exhibit functions not directly related to gene expression. At 72, 29, and 26 kD, we identified the two iron superoxide dismutases FSD3 and FSD2. FSD2 at 29 kD has been found in earlier studies by Edman degradation of mustard PEP subunits, and both enzymes were detected by antibody reactions in nucleoids (Pfannschmidt et al., 2000; Myouga et al., 2008). However, so far, FSD3 has never been described as a PEP subunit. It appears in two bands at 72 and 26 kD. The large one differs from its theoretical size of 26 kD, while the small one fits precisely. This suggests that FSD3 generates a stable, probably trimeric complex that can only be partly resolved by the change to the second dimension SDS-PAGE.

The protein band at 52 kD always displayed a characteristic stronger staining intensity than other proteins. Our mass spectrometry data indicated that it

contains two proteins of identical size, PTAC14 and a protein corresponding to a potential kinase with a domain typical for the phosphofructokinase family. An orthologous protein was also found in the tobacco PEP (Suzuki et al., 2004), and a corresponding ortholog called FLN1 was recently characterized in *Nicotiana benthamiana*. In a yeast-two-hybrid screen, FLN1 interacted with TrxZ (Arsova et al., 2010), a novel thioredoxin-like protein identified as a 13-kD subunit of the mustard PEP complex here and recently (Schröter et al., 2010). All these non-*rpo* subunits were regarded as essential components of the PEP complex and therefore were named PEP-associated proteins (PAP) 1 to 10.

Our biochemical approach identified neither σ -factors nor cpCK2, CSP41, and an annexin-like protein identified earlier in the mustard PEP complex by Edman degradation and mass spectrometry (Pfannschmidt et al., 2000; Ogrzewalla et al., 2002). Sigma factors likely interact very shortly with the RNA polymerase during promoter recognition and probably exist only in substoichiometric amounts, exacerbating their biochemical identification (Schweer et al., 2010). In addition, biochemical observations demonstrated that the plastid transcription kinase (representing cpCK2) can dissociate from the RNA polymerase complex (Baginsky et al., 1999). This suggests that all these proteins likely represent transient or loosely attached components of the PEP complex, which are excluded under our stringent search conditions.

Phenotypic Effects of PAP Gene Knockouts in *Arabidopsis* Mutants

The association of proteins into a multisubunit protein complex is usually reflected in a common functional commitment of these proteins. Here, we studied the composition of the plastid PEP complex; therefore, one would expect that, besides the *rpo* subunits, the non-*rpo* subunits also exhibit functions that are somehow related to transcription. However, the functional assignments of only four subunits (PAP1, PAP2, PAP3, and PAP7) are related to gene expression, while those of PAP4, PAP5, PAP6, PAP8, PAP9, and PAP10 are difficult to reconcile with this function and appear unnecessary and/or dispensable for transcription. In order to understand the structural involvement of PAPs into the PEP complex, we screened *Arabidopsis* knockout mutant collections for the presence of PAP-deficient lines. Isolated knockout lines could potentially indicate the importance of the respective PAP if phenotypic effects are caused by the respective protein deficiency. *Sinapis* and *Arabidopsis* are related crucifers, and the combination of biochemical and genetic data from both species provides a useful tool for analyzing the functions of novel proteins, as demonstrated recently (Schröter et al., 2010). The screening for potential mutants was further complemented by a survey of literature and databases for descriptions of potential phenotypic effects in PAP knockout mutants.

We found studies and database entries describing detailed phenotypes of knockout lines for most of the non-*rpo* subunits (Table II), with the exception of PAP3, PAP6, and PAP7. For these subunits, we isolated homozygous knockout lines from respective collections, tested the repression of PAP transcript accumulation by reverse transcription (RT)-PCR, and finally checked the phenotypic appearance of confirmed knockout lines in petri dishes on standard Murashige and Skoog medium (Fig. 3). This provided a complete survey of the phenotypes for knockout mutants of all PAPs in Arabidopsis. For PAP1/PTAC3, the reported knockout lines exhibited an albino phenotype, while those for PAP2/PTAC2 displayed a slightly greenish phenotype that was also reflected in the plastid ultrastructure (Pfalz et al., 2006; Myouga et al., 2010). For PAP3/PTAC10, the isolated T-DNA insertion line exhibited an albino-like phenotype in the seedling stages and turned into an ivory phenotype in later stages (Fig. 3C). Knockout lines for PAP4 and PAP9 (FSD2 and FSD3) were reported to exhibit pale-green phenotypes in single knockout lines (with leaves being paler in FSD3 than in FSD2) and a full albino phenotype in the double mutant (Myouga et al., 2008). PAP5/PTAC12 seedlings were found to be white (Chen et al., 2010), while older plants turned into an ivory phenotype (Pfalz et al., 2006). For PAP6/FLN1, we isolated an Arabidopsis knockout line that also displayed an albino phenotype. Recently, the orthologous gene was analyzed in *N. benthamiana* by virus-induced gene silencing. Intriguingly, down-regulation of FLN1 expression resulted in white sectors in the affected leaves, while the same experiment with the paralogous protein FLN2 (which we did not identify as PAP) produced no apparent phenotypic variations (Arsova et al., 2010). For PAP7/PTAC14, we isolated an Arabidopsis T-DNA insertion line and again observed an albino phenotype. For PAP8/PTAC6, a knockout line was described that exhibited an albino phenotype (Pfalz et al., 2006). Finally, PAP10/TrxZ has recently been reported to be the first thioredoxin

whose knockout results in a visible phenotype, again with an albino appearance (Arsova et al., 2010; Schröter et al., 2010).

Thus, knockouts of all subunits defined as PAPs by our biochemical approach result either in a complete block or a severe retardation of chloroplast development. In all cases, the developmental deficiencies were so strong that the mutants were only viable on Suc-supplemented medium. This makes the results highly comparable even if they were generated in different laboratories. It should be noted that we used the phenotypic description “albino” as it was found in the literature. A more detailed phenotypic analysis indicated in most cases that the albino turned into an ivory phenotype, usually clearly visible as yellowish coloring in the older stages (indicated in Table II). An ivory phenotype indicates carotenoid biosynthesis and, therefore, active and dividing plastids, which, however, cannot perform the transition into fully developed chloroplasts. This is consistent with the electron micrographs available for many PAP mutants, displaying plastids without thylakoid membrane systems and high accumulation of plastoglobuli (Table II). It also coincides with the analyses describing the respective plastid gene expression profiles, which in all cases investigated revealed a NEP-dependent transcript accumulation pattern. All these observations correspond to observations in transplastomic tobacco lines in which the *rpo* subunits had been knocked out. These exhibited an albino-like phenotype, indicating the necessity of the PEP enzyme for early chloroplast development. Typically, such plants were viable when grown on medium supplemented with Suc and displayed increased expression of the NEP-transcribed genes of the plastome, while PEP-transcribed genes were largely reduced (Allison et al., 1996; Hajdukiewicz et al., 1997; De Santis-Maclossek et al., 1999). These data indicate that regardless of the predicted function, knockout of PAPs results in the same appearance as *rpo* gene knockout lines.

Table II. Phenotyping of Arabidopsis PAP knockout mutants

Growth phenotype, Developmental appearance of knockout mutants (Suc, viable only on Suc-supplemented medium); Plastid Structure, plastid morphology in knockout/silenced mutants (t., thylakoids; p., plastoglobuli enrichment); Molecular Phenotype, NEP expression profile of plastid transcript accumulation; Reference, source of phenotypic descriptions; n.d., not described. Classification as “ivory” is based on our own observations.

Subunit	Growth Phenotype	Plastid Structure	Molecular Phenotype	Reference
PAP1	Albino/ivory; Suc	n.d.	n.d.	Myouga et al. (2010)
PAP2	Pale-yellow/green; Suc	Impaired; few t.; p.	NEP	Pfalz et al. (2006)
PAP3	Albino/ivory; Suc	n.d.	n.d.	This work
PAP4	Pale-green; Suc	Impaired; no t.; p.	NEP	Myouga et al. (2008)
PAP5	Albino/ivory; Suc	Impaired; no t.; p.	NEP	Pfalz et al. (2006); Chen et al. (2010)
PAP6	Albino/ivory; Suc	n.d.	n.d.	This work
PAP7	Albino/ivory; Suc	n.d.	n.d.	This work
PAP8	Albino/ivory; Suc	Impaired; no t.; p.	NEP	Pfalz et al. (2006)
PAP9	Pale-green; Suc	Impaired; no t.; p.	NEP	Myouga et al. (2008)
PAP10	Albino/ivory; Suc	Impaired; no t.; p.	NEP	Arsova et al. (2010); Schröter et al., 2010)

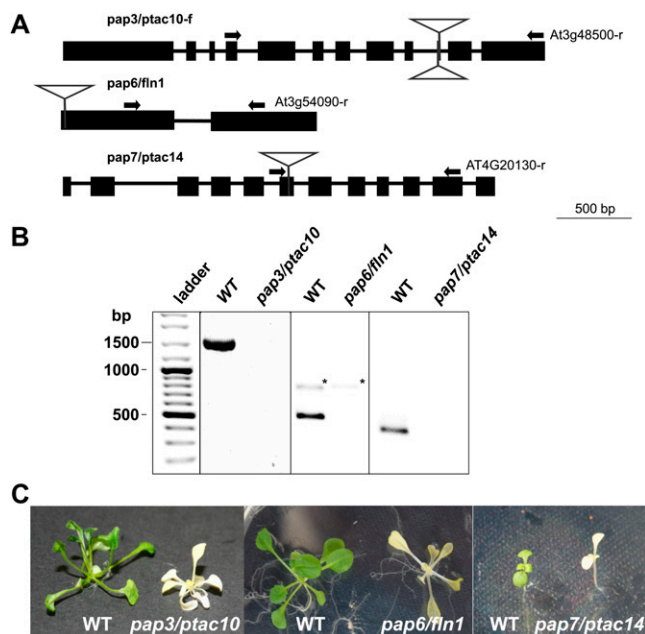


Figure 3. Identification and analysis of homozygous *pap3/ptac10*, *pap6/fln1*, and *pap7/ptac14* T-DNA insertion mutants. A, Schematic presentation of the corresponding genes showing the positions of the T-DNA insertions as confirmed by PCR and subsequent sequencing. B, RT-PCR amplification of the *pap3/ptac10*, *pap6/fln1*, and *pap7/ptac14* genes using gene-specific primers given in A. Lines homozygous for *pap3/ptac10*, *pap6/fln1*, and *pap7/ptac14* T-DNA insertion fail to express the wild-type (WT) allele. Asterisks indicate bands derived by genomic DNA. C, Wild-type and homozygous *pap3/ptac10*, *pap6/fln1*, and *pap7/ptac14* plants germinated on petri dishes with medium. Seeds of the T-DNA insertion line were surface sterilized and placed on sterile agar plates containing Murashige and Skoog medium supplemented with 2% Suc.

The phenotypic commonalities suggest that the PAPs are related to each other in a structural and/or developmental context. In order to obtain further support for such a potential relation, we determined coexpression patterns of the genes for PAPs using the Arabidopsis coresponse database (Steinhauser et al., 2004; Lisso et al., 2005; Usadel et al., 2005). The etioplast-chloroplast transition during photomorphogenesis is a major step in seedling development that involves parallel changes in thousands of nuclear genes (Ma et al., 2001). In order to distinguish PAP expression patterns from these light-induced developmental changes, we used nuclear genes encoding plastid protein components not involved in photosynthesis (e.g. RNA metabolism, metabolic pathways) as controls. Coexpression patterns within array data from AtGenExpress were obtained (Supplemental Fig. S1). In the category “developmental series,” PAPs exhibited strong coregulation (average r_s of 0.901), which was clearly different from the controls. This indicates that PAP expression appears to be coregulated, supporting the notion that PAPs are related in a structural/developmental context.

DISCUSSION

The PEP Core Enzyme

Our mass spectrometry data identified all *rpo* gene products in the PEP complex and are in good accordance with earlier reports (Hu and Bogorad, 1990; Hu et al., 1991; Pfannschmidt et al., 2000; Suzuki et al., 2004). The identification of the β' -s variant, however, is an unexpected and novel finding. Earlier studies probably missed this subunit because it migrates in the mass range between 70 and 80 kD, where at least five different PEP subunits of similar size are located, which may mask each other if not separated in a high-resolution gel system as used here. The *rpoC₁* gene is the only *rpo* gene with an intron that exists, however, only in dicot plants (Igloi et al., 1990). It is conceivable that the two proteins simply represent translation products from spliced and nonspliced variants. However, the intron sequence encodes several stop codons distributed over the complete intron (Supplemental Fig. S2), making it unlikely that unspliced transcripts are translated. This conclusion is confirmed by observations in the *otp70* mutant of Arabidopsis, which displays a defect in *rpoC₁* splicing. This defect results in PEP deficiency of the mutant, implying that a complete splicing of *rpoC₁* transcripts is essential for the generation of a functional PEP complex (Chateigner-Boutin et al., 2011). An alternative possibility for two RpoC₁ variants originates from early characterizations of the *rpoBC₁C₂* transcript maturation in spinach (*Spinacia oleracea*) via S1 mapping analyses. These suggested the existence of a second splice acceptor site within exon 2 of the *rpoC₁* gene, giving rise to a second, smaller version of the transcript and its resulting protein product (Hudson et al., 1988). This smaller product would fit the apparent molecular mass of 72 kD of β' -s. We detected a peptide from this variant covering the alternative splice site by a few amino acids but not one within the area between the two splice acceptor sites. Instead, we detected a single β' -l peptide. The analyses in the Arabidopsis mutant *otp70* indicated that exactly in that area, an editing site exists that requires the action of the PPR protein OTP70 to be matured. Unspliced *rpoC₁* transcripts appear to be preferentially edited; therefore, rapid splicing in the wild type eventually prevents *rpoC₁* transcripts from being fully edited (Chateigner-Boutin et al., 2011). This could generate two pools of transcripts with differing sequence at *rpoC₁* residue 21,806, coding either for Ser or Leu (Chateigner-Boutin and Small, 2007), which eventually could affect translation or posttranslational events. Alternatively, one could speculate that the binding of the editing factor redirects the splicing machinery toward the second splice acceptor site, resulting in a smaller transcript and hence a smaller translation product. However, our RT-PCR approach did not detect alternative *rpoC₁* splice variants, suggesting that the β' -s subunit is likely generated by posttranslational modification of the β' -l

subunit. The type of modification and its effect on functionality require further investigation.

As a side aspect, we observed that the α -subunit band did not exhibit an increased staining strength, as observed for the 52-kD band containing PTAC14 and FLN1. This observation is just a hint but suggests that the α -subunit does not necessarily exist in two copies per complex, as predicted from simple adaptations of the *E. coli* structure $2\alpha, \beta, \beta'$. It is equally likely that the structure of the PEP core enzyme could be an α, β, β' -1, β' -s, β'' assembly. Probably, only structural work, including crystallography, will be of sufficient resolution to fully understand the composition and structure of the PEP core complex.

PAPs

We could reproducibly identify 10 essential non-*rpo* protein subunits of the PEP complex, which can be roughly divided into two functional groups. One group consists of PAP1, PAP2, PAP3, and PAP7, with domains or motifs likely involved in gene expression/regulation (SAP, PPR, S1, SET, respectively; Table I). The other group consists of PAP4 to -6 and PAP8 to -10, which all are related to or involved in redox-dependent processes or regulation. The specific functions of PAP1, PAP3, and PAP7 are based only on protein domain predictions, while all other PAPs have been, at least in part, functionally characterized. PAP2, PAP5, and PAP8 are also known as PTAC2, PTAC12, and PTAC6, and the corresponding knockout mutants all display a specific PEP-deficient plastid gene expression phenotype (Pfalz et al., 2006). Intriguingly, the same has been observed in knockout mutants for PAP4 and PAP9 as well as for PAP6 and PAP10, and it is reasonable to expect a similar expression pattern also in the uncharacterized PAP1, PAP3, and PAP7 mutants, since they exhibit comparable phenotypes. PAP5 or PTAC12/HEMERA is special among all these proteins, since it has recently been demonstrated to be dual targeted to nucleus and plastids (Chen et al., 2010). In the nucleus, it appears to be located in so-called nuclear bodies and seems to act in phytochrome signaling, probably in ubiquitin-mediated proteolysis, since it exhibits some similarities to the yeast RAD23 protein. Dual localization in nucleus and plastids within the same plant cell was first demonstrated for the RNA-binding protein Whirly1/PTAC1 (Grabowski et al., 2008), and further analyses suggested that this dual subcellular distribution occurs also for other plant cell proteins (Krause and Krupinska, 2009). Whether PAP5 is involved in plastid protein degradation, however, is not known yet. PAP4 and PAP9 are two superoxide dismutases, the first of which, to our knowledge, has been identified here for the first time as a PAP, while the second was described earlier (Pfannschmidt et al., 2000). A recent independent study could show that these two proteins interact in a yeast two-hybrid assay and that both are located within plastid nucleoids (Myouga et al., 2008). These observations are consistent with our data. Interaction in a yeast two-hybrid assay

also could be demonstrated for PAP6 and PAP10. PAP6 is also called FLN1, and its sequence suggests that it belongs to the class of the phosphofructokinases; however, it could be shown that this enzyme lost its ability to recognize this type of substrate (Arsova et al., 2010). The interacting PAP10 is also called TrxZ and represents a novel type of thioredoxin. It still functions as a "true" thioredoxin in the insulin activation assay (Arsova et al., 2010), but it is the only thioredoxin that apparently cannot be replaced by another one, since the knockout results in an albino phenotype (Arsova et al., 2010; Schröter et al., 2010). Nevertheless, despite these investigations, little is known about the true PAP functions, and further characterizations will be necessary to unravel the specific roles of the distinct PAPs in the RNA polymerase complex.

Impact of PAP Gene Knockouts on Plastid Development

The major common feature of all PAPs, regardless of their predicted/detected functions, is that a knockout of the corresponding gene always results in a severe defect in chloroplast development. In knockout mutants of PAP2, PAP4 to -6, and PAP8 to -10, this is accompanied by high NEP-dependent and low PEP-dependent transcript accumulation (Table II and refs. therein). This suggests that PAP knockouts cause a block of PEP activity that prevents the transition of plastid transcription from a NEP-dependent to a PEP-dependent mode in the same manner as could be observed for *rpo* gene knockout mutants of tobacco (Allison et al., 1996; Hajdukiewicz et al., 1997; De Santis-Maclossek et al., 1999) and *rpo* knockdown mutants of Arabidopsis (Chateigner-Boutin et al., 2008, 2011; Zhou et al., 2008).

For *S. alba*, extensive biochemical data exist that describe the subunit composition of the soluble PEP enzyme in etioplasts, greening chloroplasts, and mature chloroplasts (Pfannschmidt and Link, 1994; Pfannschmidt et al., 2000; Ogrzewalla et al., 2002; Loschelder et al., 2004). The combination of these data with the biochemical results reported here suggest an explanation for the observed phenotypes in the Arabidopsis PAP knockout mutants (Fig. 4). In early seedling development, the *rpo* subunits of PEP are expressed by the nucleus-encoded NEP enzyme, representing a first essential checkpoint in the establishment of the plastid gene expression machinery. In combination with some PEP starter molecules inherited from the parent plant (Demarsy et al., 2006), these first PEP complexes provide effective transcription of PEP-dependent plastid genes in the very early stages of plastid development. Data from etioplasts and greening chloroplasts suggest that these complexes exhibit the basic prokaryote-like PEP structure, representing a core complex consisting of only the *rpo* gene products (PEP-B; Hu and Bogorad, 1990; Hu et al., 1991; Pfannschmidt et al., 2000). With the onset of photomorphogenesis, this PEP-B enzyme is reconfigured into a much more complex eukaryote-like enzyme complex, the PEP-A enzyme. This most likely

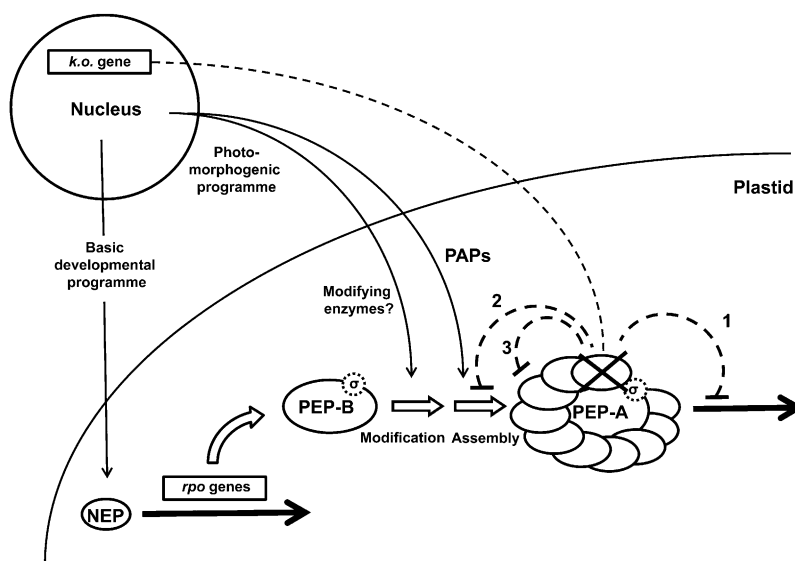


Figure 4. Checkpoint model describing the reconfiguration of the plastid RNA polymerase complex as an essential step in plastid development. The scheme depicts the structural assembly of the plastid-encoded RNA polymerase, starting with the expression of the *rpo* genes by the NEP enzyme, resulting in the formation of the basic PEP-B enzyme. Interaction with σ -factors is assumed. Upon initiation of photomorphogenesis, it is modified first by posttranslational changes of *rpo* subunits (via unknown modifying enzymes) and second by the addition of PAPs, generating the structurally more complex PEP-A. White arrows indicate the flow of events required for PEP-A buildup. Thin black arrows indicate the action or involvement of nucleus-encoded proteins delivered in a fixed sequence that follows a distinct developmental program in the nucleus. Thick black arrows indicate transcription activity. Dotted lines indicate the possible impact of a PAP gene knockout in the nucleus on PEP-A. The lacking subunit is indicated by a cross, its inhibitory feedback by dotted lines. Numbers refer to discussed possibilities causing the observed phenotypes of PAP knockout mutants. For further details, see text.

involves first a posttranslational modification of the PEP-B *rpo* subunits, since the complex changes subunit sizes and promoter recognition properties (Pfannschmidt and Link, 1997), which allow in a second step the assembly of the nucleus-encoded PAPs. The time range for this reconfiguration of the PEP complex parallels the etioplast-chloroplast transition and requires only a small time window of about 16 to 48 h, depending on the growth conditions of the seedlings (Hu and Bogorad, 1990; Hu et al., 1991; Pfannschmidt and Link, 1994). During normal plastid development without an intermediate etioplast stage, this complex conversion likely takes place faster, making it very difficult to resolve this process in a temporal manner. The addition of PAPs to the PEP core complex is the second essential checkpoint in the establishment of the plastid gene expression machinery, and the strong phenotypes of the *Arabidopsis* mutants indicate that it represents an irreplaceable step in chloroplast development.

While the latter observation is indisputable, the reason for it remains obscure. In principle, there are three possibilities why lack of any PAP results in a block or severe disturbance of plastid development (Fig. 4). First, PAP functions are essential for the transcriptional activity or regulation of the complex; second, the proteins are required for the assembly or attachment of further proteins; third, the proteins are required for the integrity of the complex itself. One could imagine that failure in each one of these pro-

cesses leads to inactivity of the PEP-A complex and, hence, to a disturbance in plastid development. However, not all possibilities are equally likely. The basic PEP-B complex is already capable of faithful transcription, rendering it unlikely that the addition of PAPs is essential for the transcriptional activity of PEP-A. Thus, it appears more reasonable that PAPs modulate or regulate transcriptional activity, and at least four subunits seem to confirm this assumption, because their predicted functions are potentially involved in gene expression (Table II). The NEP expression profile observed in some mutants, however, points more to a complete inactivation of the PEP-A complex, which is unlikely if just one regulatory event is affected. Knockouts for most known regulators of chloroplast transcription do not result in such strong phenotypes and usually display just a few gene-specific changes or minor general effects of limited impact (Bollenbach et al., 2009; Schweer et al., 2010; Barkan, 2011; Lerbs-Mache, 2011). This suggests that structural effects are potentially the reason for the observed phenotypes. If one or more of the PAPs are lacking, the whole complex could become unstable and either breaks apart or its further assembly is retarded or blocked. This could result in an inactive PEP complex, causing the observed NEP expression phenotype of the chlorotic knockout mutant lines. The ivory phenotype of many of the mutants indicates that the plastids, although PEP deficient, are still active, producing

carotenoids and also being able to divide, since older leaves of plants grown on Suc-supplemented medium appear yellow. This suggests that the plastids are arrested in an early developmental stage, being unable to reach the next step of development despite the presence of light. This model would be consistent with the phenotypic effects in the knockout lines of PAP5/PTAC12/HEMERA. The lack of PAP5 leads to a block in PEP complex assembly in developing plastids and causes chlorosis. In parallel, it prevents photomorphogenic responses of the young seedlings due to its lack in the nucleus, as apparent from its lacking a response to red/far-red shift experiments (Chen et al., 2010). Since chloroplast development is an intrinsic part of photomorphogenesis, one and the same expression program of PAP5/PTAC12/HEMERA could serve both genetic compartments.

CONCLUSION

We generated a comprehensive and complete catalog of the subunits of the chloroplast RNA polymerase from mustard, comprising five subunits encoded by plastid *rpo* genes and 10 subunits called PAPs encoded by nuclear genes. We identified three novel protein subunits, β' -s, FSD3, and PTAC12, not described yet in the soluble PEP complex. Combining these biochemical data with observations from reverse genetics, we could establish that PAPs represent essential components of the PEP complex. These components display a coexpression pattern that is mainly determined by developmental programs, pointing to the reconfiguration of the PEP complex as an essential step in plastid and plant development. Our checkpoint model explains the white/ivory/pale-green phenotypes of PAP knockout mutants as the likely result of an interruption in this PEP complex reconfiguration. Although we now have a clear picture of the subunit composition of the PEP complex, our knowledge of their precise functions is still rudimentary, and further studies are required to fully understand the processes involved in plastid transcription and its regulation. Complementation of PAP knockout lines with corresponding full-length genes carrying modified functional domains provides a useful tool for this future goal.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Mustard seedlings (*Sinapis alba* var Albatros) were grown for 7 d on soil under continuous white light illumination at 20°C and 60% humidity. Cotyledons were harvested under the growth light, placed on ice, and immediately used for preparation of chloroplasts.

HS Chromatography of Chloroplast Proteins

Two kilograms of cotyledons was homogenized in ice-cold isolation buffer using a Waring blender and filtered through three layers of muslin and one layer of nylon. Chloroplasts were isolated by differential centrifugation

followed by Suc gradient centrifugation, lysed, and subjected to HS CL-6B chromatography as described earlier (Tiller and Link, 1993; Steiner et al., 2009). Bound proteins were eluted with a single high-salt step of 1.2 M $(\text{NH}_4)_2\text{SO}_4$. The elution peak was identified by a protein quantification assay (RC-DC; Bio-Rad). RNA polymerase activity was determined in an in vitro transcription activity assay (Pfannschmidt and Link, 1994). Identified peak fractions were pooled and dialyzed against storage buffer (50 mM Tris-HCl, pH 7.6, 0.1 mM EDTA, 5 mM 2-mercaptoethanol, 0.1% Triton X-100, and 50% glycerol) and stored at -20°C until further use.

2D BN-PAGE

HS peak fractions were subjected to 2D gel electrophoresis using as a first dimension BN-PAGE with 4% to 12% acrylamide gradient gels followed by denaturing SDS-PAGE on 7% to 17% acrylamide gradient gels as a second dimension, as described recently (Schröter et al., 2010).

Tryptic In-Gel Digestion and Liquid Chromatography/ESI-MS/MS Analysis

Peptide generation of proteins from silver-stained gels was performed by tryptic in-gel digestion of cutout spots using a described protocol (Mørtz et al., 1994; Stauber et al., 2003) with minor modifications. Peptides from digested proteins were analyzed by liquid chromatography/ESI-MS/MS using a LCQ-DecaXP ion trap mass spectrometer (Thermo Finnigan). Nano-liquid chromatography was performed using an UltiMate Nano liquid chromatograph with a Famos Autosampler HPLC unit and a reverse-phase C18 PepMap100, 3- μm , 100- \AA Nano column (75 μm i.d. \times 15 cm; Dionex). Peptides were eluted using a three-step gradient with mobile phases A (0.1% HCOOH and 5% acetonitrile in water) and B (0.1% HCOOH and 80% acetonitrile in water). The mobile phase flow was 27 $\mu\text{L min}^{-1}$ with 5% B for the first 8 min, followed by 5% to 50% B in the next 17 min, 50% to 95% B for 0.5 min, held at 95% B for 18 min, switched back to 5% B in 0.5 min, and held at 5% B for 16.5 min. The ion signals from the eluted peptides were collected using a data-dependent scan procedure with four cyclic scan events. The first cycle was composed of a full mass spectrometry scan of the mass-to-charge ratio range 450 to 1,200, followed by three MS/MS scans for the three most abundant ions. Sample run and data acquisition were performed using Xcalibur software (version 1.3; Thermo Finnigan).

MS Data Analysis

For peak list generation, the Ceate DTA tool of TurboSEQUENT (version 27 [revision 12]; Molecular Biotechnology, University of Washington, licensed to Thermo Finnigan) was used with default settings. Database search was conducted with TurboSEQUENT (version 27 [revision 12]) against a Brassicales protein database of the National Center for Biotechnology Information (NCBI Brassicales 2008.09.09; 154,464 sequences). The enzyme specificity was set to trypsin strict, and no missed cleavages were permitted. As variable modifications, the carboxyamidomethylation of Cys (57.0293), oxidation of Met (15.9949), and phosphorylation of Ser, Thr, and Tyr (79.9663) were included. The mass tolerance for precursor ions was set to 1.5 D and 0 D for fragment ions. Calculated cross-correlation values for significantly matching sequences had to be equal to or above 1.5, 2.0, or 3.5 for singly, doubly, or triply charged precursor ions, respectively, and the delta correlation (ΔCorr) values had to exceed 0.1. Proteins were accepted as identified with two or more different significant matching peptides. The database used is highly redundant; consequently, peptides match to several equivalent proteins of Arabidopsis and other Brassicales species. Therefore, the protein entry of the first complete sequence of Arabidopsis within the list of matching entries is given in "Results." Alternatively, a representative species is given in the case in which the Arabidopsis sequence is not matching.

*rpoC*₁ Transcript Analysis

*rpoC*₁ transcript splice sites were analyzed by RT-PCR using cDNA from Arabidopsis (*Arabidopsis thaliana* ecotype Columbia) and mustard generated from total RNA with random hexamer oligonucleotides. Primers were *rpoC*₁-fwd (5'-AATTGGCTTAGTTTCTCCTCAG-3') and *rpoC*₁-rev (5'-CCCTTCT-TCTCCTAATGTGTTCC-3'). Preparation of total RNA for cDNA synthesis,

PCR, and RT-PCR programs followed standard procedures described at The Arabidopsis Information Resource (<http://www.arabidopsis.org>).

In Silico and Genetic Analyses of Arabidopsis PAP Knockout Mutant Plants

Gene identification numbers of the determined proteins obtained from SEQUEST searches as well as trivial names of identified proteins/enzymes were used to screen the PubMed literature database at the NCBI (<http://www.ncbi.nlm.nih.gov/pubmed/>) for publications concerning either the encoded proteins or corresponding knockout or knockdown mutant lines in various plant species. In addition, Arabidopsis T-DNA knockout mutant collections at The Arabidopsis Information Resource (<http://www.arabidopsis.org>) were screened for respective tagged lines. Identified knockout lines were compared with published data and checked for phenotypic descriptions in research reports, the Chloroplast Function Database (Myouga et al., 2010; Table II), and the Chloroplast 2010 database (<http://www.plastid.msu.edu>; Ajjawi et al., 2010). Seeds of unpublished lines CS16115 for PAP3/PTAC10, GK-443A08 for PAP6/FLN1, and SAIL_566_F06 for PAP7/PTAC14 were ordered, and homozygous plants were isolated. Primers for the detection of T-DNA locations were as follows: LB3 (5'-TAGCATCTGAATTCATAACCAATCTCGATACAC-3'), CS16115-(pap3/ptac) forward (5'-TCAGGGAGC-GTTTGTGACAT-3'), CS16115-(pap3/ptac) reverse (5'-GGTGATCAGAGCAGCCCTT-3'), GK-443A08-forward (5'-CAAATAGCGAGTCTCAGGTG-3'), GK-443A08-reverse (5'-GATCAATCCCACAAGGAAGC-3'), SAIL_566_F06-forward (5'-AGAAGGTCCAGATCGTTTGG-3'), and SAIL_566_F06-reverse (5'-TGCAGAGAATGATCAATCGTG-3'). Primers for RT-PCR were as follows: CS16155-f (5'-TCAGGGAGCGTTTGTGACAT-3'), At3g48500-r (5'-TCAGTCTGTCAAGACTTGGAG-3'), pap6-fwd (5'-CAAGATCAAGTTTAA-GGACGG-3'), pap6-rev (5'-GTTCCATCAGTCAACAACCAG-3'), pap7-fwd (5'-CAACAACAGAAACGAATCCT-3'), and pap7-rev (5'-CAGAGAACITTA-TAGCATCCCA-3'). PCR and RT-PCR programs followed standard procedures described at The Arabidopsis Information Resource (<http://www.arabidopsis.org>). The phenotypic appearance of the homozygous lines was tested on Murashige and Skoog medium containing 2% Suc, as described recently (Schröter et al., 2010).

Coexpression Analysis of PAPs

For the analysis of coexpression, we used the Arabidopsis coreponse database (<http://csbdb.mpimp-golm.mpg.de>; Steinhäuser et al., 2004) and searched for expression correlations of all PAPs to each other in the transcript profiles of the AtGenExpress stress series, developmental series, and miscellaneous. The nonparametric Spearman's ρ rank correlation r_s (ranging from +1 to -1) was obtained for each pair, linked to a value-dependent color code for visualization, and given as a matrix (Supplemental Fig. S1).

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers AtCg00170, AtCg00190, At3g04260, At1g74850, AtCg00180, At3g48500, At5g23310, AtCg00180, At2g34640, At3g54090, At4g20130, AtCg00740, At1g21600, At5g51100, and At3g06730.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Coexpression analysis of PAPs.

Supplemental Figure S2. Computer translation of the unspliced *rpoC₁* gene of Arabidopsis.

Supplemental Table S1. Primary data of mass spectrometry for *rpo* and non-*rpo* subunits.

ACKNOWLEDGMENTS

We thank the SALK Institute for making T-DNA insertion lines publicly available. Monique Liebers is acknowledged for her help in the coexpression analysis.

Received August 2, 2011; accepted September 14, 2011; published September 23, 2011.

LITERATURE CITED

- Abdallah F, Salamini F, Leister D** (2000) A prediction of the size and evolutionary origin of the proteome of chloroplasts of Arabidopsis. *Trends Plant Sci* 5: 141–142
- Ajjawi I, Lu Y, Savage LJ, Bell SM, Last RL** (2010) Large-scale reverse genetics in Arabidopsis: case studies from the Chloroplast 2010 Project. *Plant Physiol* 152: 529–540
- Allison LA** (2000) The role of sigma factors in plastid transcription. *Biochimie* 82: 537–548
- Allison LA, Simon LD, Maliga P** (1996) Deletion of *rpoB* reveals a second distinct transcription system in plastids of higher plants. *EMBO J* 15: 2802–2809
- Arsova B, Hoja U, Wimmelbacher M, Greiner E, Ustün S, Melzer M, Petersen K, Lein W, Börnke F** (2010) Plastidial thioredoxin z interacts with two fructokinase-like proteins in a thiol-dependent manner: evidence for an essential role in chloroplast development in *Arabidopsis* and *Nicotiana benthamiana*. *Plant Cell* 22: 1498–1515
- Baginsky S, Tiller K, Link G** (1997) Transcription factor phosphorylation by a protein kinase associated with chloroplast RNA polymerase from mustard (*Sinapis alba*). *Plant Mol Biol* 34: 181–189
- 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 Mol Biol* 39: 1013–1023
- Barkan A** (2011) Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol* 155: 1520–1532
- Blankenship RE** (2002) *Molecular Regulation of Photosynthesis*. Blackwell Science, Oxford
- Bollenbach TJ, Sharwood RE, Gutierrez R, Lerbs-Mache S, Stern DB** (2009) The RNA-binding proteins CSP41a and CSP41b may regulate transcription and translation of chloroplast-encoded RNAs in Arabidopsis. *Plant Mol Biol* 69: 541–552
- Boyer AS, Hallick RB** (1998) Purification and characterization of a soluble DNA-dependent chloroplast RNA polymerase from *Pisum sativum*. *Plant Sci* 137: 13–32
- Bradley D, Gatenby AA** (1985) Mutational analysis of the maize chloroplast ATPase-beta subunit gene promoter: the isolation of promoter mutants in *E. coli* and their characterization in a chloroplast in vitro transcription system. *EMBO J* 4: 3641–3648
- Buchanan BB, Gruissem W, Jones RL** (2002) *Biochemistry and Molecular Biology of Plants*. John Wiley & Sons, Somerset, NJ
- Cahoon AB, Stern DB** (2001) Plastid transcription: a menage à trois? *Trends Plant Sci* 6: 45–46
- Chateigner-Boutin AL, des Francs-Small CC, Delannoy E, Kahlau S, Tanz SK, de Longevialle AF, Fujii S, Small I** (2011) OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript *rpoC1*. *Plant J* 65: 532–542
- Chateigner-Boutin AL, Ramos-Vega M, Guevara-García A, Andrés C, de la Luz Gutiérrez-Nava M, Cantero A, Delannoy E, Jiménez LF, Lurin C, Small I, et al** (2008) CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. *Plant J* 56: 590–602
- Chateigner-Boutin AL, Small I** (2007) A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. *Nucleic Acids Res* 35: e114
- Chen M, Galvão RM, Li MN, Burger B, Bugea J, Bolado J, Chory J** (2010) Arabidopsis HEMERA/pTAC12 initiates photomorphogenesis by phytochromes. *Cell* 141: 1230–1240
- Demarsy E, Courtois F, Azevedo J, Buhot L, Lerbs-Mache S** (2006) Building up of the plastid transcriptional machinery during germination and early plant development. *Plant Physiol* 142: 993–1003
- De Santis-Maclossek G, Kofer W, Bock A, Schoch S, Maier RM, Wanner G, Rüdiger W, Koop HU, Herrmann RG** (1999) Targeted disruption of the plastid RNA polymerase genes *rpoA*, *B* and *C1*: molecular biology, biochemistry and ultrastructure. *Plant J* 18: 477–489
- Grabowski E, Miao Y, Mulisch M, Krupinska K** (2008) Single-stranded DNA-binding protein Whirly1 in barley leaves is located in plastids and the nucleus of the same cell. *Plant Physiol* 147: 1800–1804
- Greiner S, Rauwolf U, Meurer J, Herrmann RG** (2011) The role of plastids in plant speciation. *Mol Ecol* 20: 671–691
- Hajdukiewicz PTJ, Allison LA, Maliga P** (1997) The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J* 16: 4041–4048

- Hallick RB, Lipper C, Richards OC, Rutter WJ** (1976) Isolation of a transcriptionally active chromosome from chloroplasts of *Euglena gracilis*. *Biochemistry* **15**: 3039–3045
- Hedtke B, Börner T, Weihe A** (2000) One RNA polymerase serving two genomes. *EMBO Rep* **1**: 435–440
- Herrmann RG, Maier RM, Schmitz-Linneweber C** (2003) Eukaryotic genome evolution: rearrangement and coevolution of compartmentalized genetic information. *Philos Trans R Soc Lond B Biol Sci* **358**: 87–97, discussion 97
- Hess WR, Börner T** (1999) Organellar RNA polymerases of higher plants. *Int Rev Cytol* **190**: 1–59
- Hu J, Bogorad L** (1990) Maize chloroplast RNA-polymerase: the 180-, 120-, and 38-kilodalton polypeptides are encoded in chloroplast genes. *Proc Natl Acad Sci USA* **87**: 1531–1535
- Hu J, Troxler RF, Bogorad L** (1991) Maize chloroplast RNA polymerase: the 78-kilodalton polypeptide is encoded by the plastid *rpoC1* gene. *Nucleic Acids Res* **19**: 3431–3434
- Hudson GS, Holton TA, Whitfield PR, Bottomley W** (1988) Spinach chloroplast *rpoBC* genes encode three subunits of the chloroplast RNA polymerase. *J Mol Biol* **200**: 639–654
- Igloi GL, Kossel H** (1992) The transcriptional apparatus of chloroplasts. *Crit Rev Plant Sci* **10**: 525–558
- Igloi GL, Meinke A, Döry I, Kössel H** (1990) Nucleotide sequence of the maize chloroplast *rpo B/C1/C2* operon: comparison between the derived protein primary structures from various organisms with respect to functional domains. *Mol Gen Genet* **221**: 379–394
- Khanna NC, Lakhani S, Tewari KK** (1992) Identification of the template binding polypeptide in the pea chloroplast transcriptional complex. *Nucleic Acids Res* **20**: 69–74
- Krause K, Krupinska K** (2000) Molecular and functional properties of highly purified transcriptionally active chromosomes from spinach chloroplasts. *Physiol Plant* **109**: 188–195
- Krause K, Krupinska K** (2009) Nuclear regulators with a second home in organelles. *Trends Plant Sci* **14**: 194–199
- Lakhani S, Khanna NC, Tewari KK** (1992) Two distinct transcriptional activities of pea (*Pisum sativum*) chloroplasts share immunologically related functional polypeptides. *Biochem J* **286**: 833–841
- Lerbs S, Bräutigam E, Parthier B** (1985) Polypeptides of DNA-dependent RNA polymerase of spinach chloroplasts: characterization by antibody-linked polymerase assay and determination of sites of synthesis. *EMBO J* **4**: 1661–1666
- Lerbs-Mache S** (2011) Function of plastid sigma factors in higher plants: regulation of gene expression or just preservation of constitutive transcription? *Plant Mol Biol* **76**: 235–249
- Liere K, Weihe A, Börner T** (2011) The transcription machineries of plant mitochondria and chloroplasts: composition, function, and regulation. *J Plant Physiol* **168**: 1345–1360
- Link G** (1996) Green life: control of chloroplast gene transcription. *Bioessays* **18**: 465–471
- Link G** (2003) Redox regulation of chloroplast transcription. *Antioxid Redox Signal* **5**: 79–87
- Lisso J, Steinhäuser D, Altmann T, Kopka J, Müssig C** (2005) Identification of brassinosteroid-related genes by means of transcript co-response analyses. *Nucleic Acids Res* **33**: 2685–2696
- Little MC, Hallick RB** (1988) Chloroplast *rpoA*, *rpoB*, and *rpoC* genes specify at least three components of a chloroplast DNA-dependent RNA polymerase active in tRNA and mRNA transcription. *J Biol Chem* **263**: 14302–14307
- Loschelder H, Homann A, Ogrzewalla K, Link G** (2004) Proteomics-based sequence analysis of plant gene expression: the chloroplast transcription apparatus. *Phytochemistry* **65**: 1785–1793
- Lysenko EA, Kuznetsov VV** (2005) [Plastid RNA polymerases]. *Mol Biol (Mosk)* **39**: 762–775
- Ma L, Li J, Qu L, Hager J, Chen Z, Zhao H, Deng XW** (2001) Light control of *Arabidopsis* development entails coordinated regulation of genome expression and cellular pathways. *Plant Cell* **13**: 2589–2607
- Marchler-Bauer A, Lu SN, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, et al** (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* **39**: D225–D229
- Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D** (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci USA* **99**: 12246–12251
- Mörtz E, Vorm O, Mann M, Roepstorff P** (1994) Identification of proteins in polyacrylamide gels by mass spectrometric peptide mapping combined with database search. *Biol Mass Spectrom* **23**: 249–261
- Myoung F, Akiyama K, Motohashi R, Kuromori T, Ito T, Iizumi H, Ryusui R, Sakurai T, Shinozaki K** (2010) The Chloroplast Function Database: a large-scale collection of *Arabidopsis* Ds/Spm- or T-DNA-tagged homozygous lines for nuclear-encoded chloroplast proteins, and their systematic phenotype analysis. *Plant J* **61**: 529–542
- Myoung 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*. *Plant Cell* **20**: 3148–3162
- Ogrzewalla K, Piotrowski M, Reinbothe S, Link G** (2002) The plastid transcription kinase from mustard (*Sinapis alba* L.): a nuclear-encoded CK2-type chloroplast enzyme with redox-sensitive function. *Eur J Biochem* **269**: 3329–3337
- Pfalz J, Liere K, Kandlbinder A, Dietz KJ, Oelmüller R** (2006) pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell* **18**: 176–197
- Pfannschmidt T, Liere K** (2005) Redox regulation and modification of proteins controlling chloroplast gene expression. *Antioxid Redox Signal* **7**: 607–618
- 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 Mol Biol* **25**: 69–81
- Pfannschmidt T, Link G** (1997) The A and B forms of plastid DNA-dependent RNA polymerase from mustard (*Sinapis alba* L.) transcribe the same genes in a different developmental context. *Mol Gen Genet* **257**: 35–44
- Pfannschmidt T, Ogrzewalla K, Baginsky S, Sickmann A, Meyer HE, Link G** (2000) The multisubunit chloroplast RNA polymerase A from mustard (*Sinapis alba* L.): integration of a prokaryotic core into a larger complex with organelle-specific functions. *Eur J Biochem* **267**: 253–261
- Rajasekhar VK, Sun E, Meeker R, Wu BW, Tewari KK** (1991) Highly purified pea chloroplast RNA-polymerase transcribes both rRNA and mRNA genes. *Eur J Biochem* **195**: 215–228
- Rajasekhar VK, Tewari KK** (1995) Analyses of the extent of immunological relatedness between a highly purified pea chloroplast functional RNA-polymerase and *Escherichia coli* RNA-polymerase. *J Plant Physiol* **145**: 427–436
- Reiss T, Link G** (1985) Characterization of transcriptionally active DNA-protein complexes from chloroplasts and etioplasts of mustard (*Sinapis alba* L.). *Eur J Biochem* **148**: 207–212
- Sakai A, Takano H, Kuroiwa T** (2004) Organelle nuclei in higher plants: structure, composition, function, and evolution. *Int Rev Cytol* **238**: 59–118
- Schröter Y, Steiner S, Matthäi K, Pfannschmidt T** (2010) Analysis of oligomeric protein complexes in the chloroplast sub-proteome of nucleic acid-binding proteins from mustard reveals potential redox regulators of plastid gene expression. *Proteomics* **10**: 2191–2204
- Schweier J, Türkeri H, Kolpack A, Link G** (2010) Role and regulation of plastid sigma factors and their functional interactors during chloroplast transcription: recent lessons from *Arabidopsis thaliana*. *Eur J Cell Biol* **89**: 940–946
- Shiina T, Tsunoyama Y, Nakahira Y, Khan MS** (2005) Plastid RNA polymerases, promoters, and transcription regulators in higher plants. *Int Rev Cytol* **244**: 1–68
- Soll J, Schleiff E** (2004) Protein import into chloroplasts. *Nat Rev Mol Cell Biol* **5**: 198–208
- Staubert EJ, Fink A, Markert C, Kruse O, Johanningmeier U, Hippler M** (2003) Proteomics of *Chlamydomonas reinhardtii* light-harvesting proteins. *Eukaryot Cell* **2**: 978–994
- 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. *Mol Plant* **2**: 416–429
- Steinhäuser D, Usadel B, Luedemann A, Thimm O, Kopka J** (2004) CSB. DB: a comprehensive systems-biology database. *Bioinformatics* **20**: 3647–3651
- Stoebe B, Maier UG** (2002) One, two, three: nature's tool box for building plastids. *Protoplasma* **219**: 123–130

- Suck R, Zeltz P, Falk J, Acker A, Kössel H, Krupinska K** (1996) Transcriptionally active chromosomes (TACs) of barley chloroplasts contain the alpha-subunit of plastome-encoded RNA polymerase. *Curr Genet* **30**: 515–521
- Sugiura M** (1992) The chloroplast genome. *Plant Mol Biol* **19**: 149–168
- Suzuki JY, Ytterberg AJ, Beardslee TA, Allison LA, Wijk KJ, Maliga P** (2004) Affinity purification of the tobacco plastid RNA polymerase and in vitro reconstitution of the holoenzyme. *Plant J* **40**: 164–172
- Terasawa K, Sato N** (2005) Visualization of plastid nucleoids in situ using the PEND-GFP fusion protein. *Plant Cell Physiol* **46**: 649–660
- Tiller K, Link G** (1993) Phosphorylation and dephosphorylation affect functional characteristics of chloroplast and etioplast transcription systems from mustard (*Sinapis alba* L.). *EMBO J* **12**: 1745–1753
- Usadel B, Kuschinsky AM, Steinhauser D, Pauly M** (2005) Transcriptional co-response analysis as a tool to identify new components of the wall biosynthetic machinery. *Plant Biosyst* **139**: 69–73
- Zhou WB, Cheng YX, Yap A, Chateigner-Boutin AL, Delannoy E, Hammani K, Small I, Huang JR** (2008) The Arabidopsis gene YS1 encoding a DYW protein is required for editing of rpoB transcripts and the rapid development of chloroplasts during early growth. *Plant J* **58**: 82–96