Running head: PEP subunit composition

Corresponding author:
Name: Thomas Pfannschmidt
Address: Junior Research Group “Plant acclimation to environmental changes: Protein analysis by MS” at the Institute of General Botany and Plant Physiology, Department of Plant Physiology, Friedrich-Schiller-University of Jena, Dornburger Str. 159, 07743 Jena, Germany
Tel.: ++49 (0)3641-949 236
Fax: ++49 (0)3641-949 232
E-mail: Thomas.Pfannschmidt@uni-jena.de

Journal research area: Biochemical processes and macromolecular structures
Titel: Identification of essential subunits in the plastid-encoded RNA polymerase complex reveals building blocks for proper plastid development

Authors: Sebastian Steiner, Yvonne Schröter, Jeannette Pfalz and Thomas Pfannschmidt

Address: Junior Research Group “Plant acclimation to environmental changes: Protein analysis by MS”, Department of Plant Physiology, Institute of General Botany and Plant Physiology, Friedrich-Schiller-University Jena, Dornburger Str. 159, D-07743 Jena, Germany
Footnotes:
Financial source: This work was supported by grants from the Deutsche Forschungsgemeinschaft to T. P. and the research group FOR 804, by the Freistaat Thüringen and the Bundesministerium für Bildung und Forschung (Nachwuchsgruppe “Pflanzliche Anpassung an Umweltveränderungen: Proteinanalyse mittels MS”).

Corresponding author:
Thomas Pfannschmidt, E-mail: Thomas.Pfannschmidt@uni-jena.de
Abstract
The major RNA polymerase activity in mature chloroplasts is a multi-subunit, *E. 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 *rpo* genes a number of additional nuclear 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 one represent rather transient or peripheral components. Here we present a catalogue of true PEP subunits which 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 knock-out mutants reveal an albino or pale-green phenotype. Our study provides a clearly defined subunit catalogue of the basic PEP complex generating the base for a better understanding of chloroplast transcription regulation. In addition, the data support a model which links PEP complex assembly and chloroplast build-up 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 an 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 - 120 genes including genes for photosynthesis proteins, genes for the RNA polymerase (rpo genes), genes for ribosomal subunits and RNAs as well as for tRNAs (Sugiura, 1992). The vast majority of chloroplast proteins, however, is encoded in the nucleus and must be imported from the cytosol (Abdallah et al., 2000; Soll and Schleiff, 2004). As a result all multi-protein complexes in plastids are comprised of a patchwork of plastid and nucleus-coded 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 which 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 re-imported 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 multi-enzyme 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, nuclear-encoded plastid RNA polymerase (NEP) and a prokaryotic-type, multi-subunit, plastid-encoded RNA polymerase (PEP) (Hess and Borner, 1999; Cahoon and Stern, 2001; Lysenko and Kusnetsov, 2005; Shiina et al., 2005; Liere et al., 2011). In Arabidopsis NEP is encoded by two nuclear gene copies (rpoTp and rpoTmp, respectively) 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) which exhibit ~ 26-50% sequence homology to cor-
responding genes from cyanobacteria generating the so-called core enzyme (Igloi and Kossel, 1992). This core enzyme is supplemented by a number of nuclear encoded σ-factors which 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 has been mainly investigated by biochemical means. In plastids the DNA and its associated or interacting proteins (including the RNA polymerase) are organised in so-called nucleoids or plastid nuclei, very large structures which 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 PEND-green fluorescent protein fusion described in detail localisation and distribution of nucleoids in plastids from different plant cell types (Terasawa and Sato, 2005). Purified nucleoids were very useful in 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 unsoluble RNA polymerase preparation called transcriptionally active chromosome (TAC) which can be precipitated by ultra-centrifugation. It represents a high-molecular weight DNA/RNA-protein complex containing approx. 40 - 60 proteins, which is capable of in vitro transcription as it is, 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 which 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 Gartenby, 1985; Lerbs et al., 1985; Rajasekhar et al., 1991; Lakhani et al., 1992; Pfannschmidt and Link, 1997).

Various biochemical purification procedures yielded in highly purified RNA polymerase preparations which were able to recognise 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 E. coli enzyme (2α, β,
but a much more complex structure comprised of around 20 – 30 subunits. This apparent contradiction was resolved with the identification of the prokaryotic core subunits \( \alpha \), \( \beta \), \( \beta' \) and \( \beta'' \) in various soluble RNA polymerase preparations by using various experimental approaches including western analysis, Edman degradation and mass finger prints (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 reorganisation during light-dependent chloroplast maturation. In etioplasts or young greening chloroplasts PEP displays the expected \textit{E. coli}-like structure but is reorganised into a much more complex “eukaryotic”-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 post-translational modifications of the \textit{rpo} subunits since i) these exhibit differences in the apparent molecular weight between etioplast and chloroplasts and ii) 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 why even highly purified RNA polymerase preparations from chloroplasts of several species exhibit approx. 10 – 15 proteins in addition to the \textit{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 \textit{rpo}A gene was fused to a His tag. The PEP enzyme from tobacco chloroplasts could be then purified \textit{via} nickel-affinity chromatography. Even this affinity-tag purified RNA polymerase preparation revealed a highly complex subunit composition indicating that the additional subunits co-purify due to an interaction with the \textit{rpo} subunits and/or associated non-\textit{rpo} subunits and therefore belong directly to the complex (Suzuki et al., 2004).

To understand chloroplast transcription and its regulation it is mandatory to identify all additional PEP subunits and characterise its potential functions. Using modern mass spectrometry a number of the non-\textit{rpo} subunits have been identified in the last few years, but sev-
eral subunits still remained unknown (Pfannschmidt et al., 2000; Suzuki et al., 2004; Schroter 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 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* L.) chloroplasts after heparin-Sepharose chromatography and blue native two-dimensional gel electrophoresis. We aimed to determine those subunits which 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 ten additional proteins. Combining these biochemical data with phenotypic analyses of corresponding knock-out mutants from *Arabidopsis* we could define the essential subunits of the basic PEP complex and present a comprehensive catalogue of its components. A potential role of PEP subunit assembly as decisive checkpoint in chloroplast development is discussed.

**RESULTS**

**Basic PEP subunit composition as defined by biochemical purification and mass spectrometry**

We used 7 day-old white light grown mustard seedlings as source for PEP preparations as reported earlier (Tiller and Link, 1993). Intact chloroplasts were isolated from cotyledons by homogenisation and sucrose gradient centrifugation, lysed in a buffer containing the non-ionic detergent Triton X-100 and transcriptionally active fractions were subsequently enriched by heparin Sepharose (HS) chromatography. Comparable preparations have been partially characterised earlier and contain RNA polymerases, sigma-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 2D blue native (BN) PAGE as recently described (Schroter et al., 2010). We took advantage from the observation that the PEP complex possesses a size of more than 1000 kDa being by far the largest protein complex in the HS fractions. Due to this large size the protein complex displays very slow migration behaviour in the 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 denatur-
ing second dimension producing a distinct ladder of protein subunits (Fig. 1) which can be clearly distinguished from background bands or staining artefacts 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 analysed three different protein purifications each prepared from an independent biological replicate. Only proteins which reproducibly occurred in all preparations were regarded as candidates for true components of the complex. In addition, this list of subunits was compared to that of 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 which 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 analysed by mass spectrometry. The respective protein spots were cut out, subjected to in-gel tryptic digestion and peptide masses were determined by ESI-MS/MS.

We identified the mustard proteins by the masses of the homologue peptides in the Arabidopsis sequence or other species in the Brassicales database (Table I and 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; Schroter 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 kDa followed by the β-subunit (encoded by rpoB) at 118 kDa. The β´-subunit (encoded by rpoC1) was found at around 85 kDa and, unexpectedly, in a second, smaller variant at about 72 kDa. All RpoC1 peptides detected in our mass spectrometric measurements 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 kDa which matches precisely the predicted size of 38 kDa (Igloi and Kossel, 1992).
A second group of proteins identified here is comprised of PTAC2, 3, 6, 10, 12 and 14 at apparent masses of 107, 110, 37, 76, 70 and 52 kDa, 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 characterised 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 on its potential function is missing (Table I). PTAC12 has been not 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 localisation. In all cases the apparent molecular weights were close to the predicted theoretical ones.

A third group of subunits is comprised of proteins which exhibit functions not directly related to gene expression. At 72, 29 and 26 kDa we identified the two iron superoxide dismutases FSD3 and FSD2, respectively. FSD2 at 29 kDa 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 kDa. The large one differs from its theoretical size of 26 kDa while the small one fits precisely. This suggests that FSD3 generates a stable, probably trimeric complex which can only partly resolved by the change to the second dimension SDS-PAGE.

The protein band at 52 kDa 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 phospho-fructokinase family. An orthologous protein was also found in the tobacco PEP (Suzuki et al., 2004) and a corresponding orthologue called FLN1 was recently characterised in *Nicotiana benthamiana*. In a yeast-two-hybrid screen FLN1 interacted with TrxZ (Arsova et al., 2010), a novel thioredoxin-like protein identified as 13 kDa subunit of the mustard PEP complex here, and recently (Schroter et al., 2010). All these non-rpo subunits were regarded as essential components of the PEP-complex and therefore named PEP-associated proteins (PAP) 1-10.

Our biochemical approach did neither identify sigma 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 do exist only in sub-stoichiometric 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 knock-outs in Arabidopsis mutants**

The association of proteins into a multi-subunit protein complex is usually reflected in a common functional commitment of these proteins. Here, we studied the composition of the plastid PEP complex and one, therefore, would expect that beside the *rpo* subunits also the non-*rpo* subunits exhibit functions which are somehow related to transcription. However, the functional assignments of only four subunits (PAP1, PAP2, PAP3 and PAP7) are related to gene expression while that 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* knock out mutant collections for presence of PAP-deficient lines. Isolated knock out 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 cruciferes and combination of biochemical and genetic data from both species provides a useful tool for analysing the function of novel proteins as demonstrated recently (Schroter et al. 2010). The screening for potential mutants was further complemented by a survey of literature and data bases for descriptions of potential phenotypic effects in PAP knock-out mutants.

We found studies and data base entries describing detailed phenotypes of knock out lines for most of the non-*rpo* subunits (Table II) with the exception of PAP3, PAP6 and PAP7. For these subunits we isolated homozygous knock-out lines from respective collections, tested the repression of PAP transcript accumulation by RT-PCR and finally checked the phenotypic appearance of confirmed knock-out lines in Petri-dishes on standard MS medium (Fig. 3). This provided a complete survey of the phenotypes for knock-out mutants of all PAPs in *Arabidopsis*. For PAP1/PTAC3 the reported knock-out lines exhibited an albino phenotype while that for PAP2/PTAC2 displayed a slightly greenish phenotype which 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). Knock-out lines
for PAP4 and PAP9 (FSD2 and FSD3) were reported to exhibit pale-green phenotypes in single knock-out 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 knock-out line which also displayed an albino phenotype. Recently, the orthologous gene was analysed in Nicotiana 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 knock-out line was described which exhibited an albino phenotype (Pfalz et al. 2006). Finally, PAP10/TrxZ has been recently reported to be the first thioredoxin whose knock-out results in a visible phenotype, again with an albino appearance (Arsova et al., 2010; Schroter et al., 2010).

Thus, knock-outs 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 sucrose-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 colouring 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 sucrose 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-Maciossek et al., 1999). These data indicate that regardless of the predicted function knock-outs of PAPs result in the same appearance as rpo-gene knock-out lines.

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 co-expression patterns of the genes for PAPs using the Arabidopsis thaliana co-response database (Steinhauser et al., 2004; Lisso et al., 2005; Usadel et al., 2005). Etioplast-chloroplast transition during photomorphogenesis is a major step in seedling development which 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. Co-expression patterns within array data from the AtGenExpress were obtained (Supplemental Figure S1). In the category “developmental series” PAPs exhibited strong co-regulation (average $r_s$ 0.901) which was clearly different from the controls. This indicates that PAP expression appears to be co-regulated 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 $\beta^s$ variant, however, is an unexpected and novel finding. Earlier studies probably missed this subunit because it migrates in the mass range between 70 – 80 kDa 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 $rpoC_1$ gene is the only rpo-gene with an intron which exists, however, only in dicot plants (Igloi et al., 1990). It is conceivable to assume that the two proteins simply represent translation products from spliced and non-spliced 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 $rpoC_1$ splicing. This defect results in PEP-deficiency of the mutant implying that a complete splicing of $rpoC_1$
transcripts is essential for generation of a functional PEP complex (Chateigner-Boutin et al., 2011). An alternative possibility for two RpoC1 variants originates from early characterisations of the rpoBC1C2 transcript maturation in spinach 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 to the apparent molecular weight of 72 kDa of β-′-s.

We detected a peptide from this variant covering the alternative splice site by a few amino-acids but no 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 which requires the action of the PPR protein OTP70 to be matured. Unspliced rpoC1 transcripts appear to be preferentially edited and, 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(21806) coding either for serine or leucine (Chateigner-Boutin and Small, 2007) which eventually could affect translation or post-translational events. Alternatively one could speculate that the binding of the editing factor re-directs the splicing machinery towards 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 post-translational modification of the β-′-l subunit. The type of modification and its effect on functionality require further investigations.

As a side aspect we observed that the α-subunit band did not exhibit an increased staining strength as observed for the 52 kDa band containing PTAC14 and FLN1. This observation is just a hint but suggests that the α-subunit not necessarily exists in two copies per complex as predicted from simple adaptations of the E. coli structure 2α, β, β-′. It is equally likely that the structure of the PEP core enzyme could be an α, β, β-′-l, 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.

**PEP-associated proteins (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; compare Table I). The other group consists of PAP4-6 and PAP8-10 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 characterised. PAP2, PAP5 and PAP8 are also known as PTAC2, PTAC12 and PTAC6 and the corresponding knock-out mutants all display a specific PEP-deficient plastid gene expression phenotype (Pfalz et al., 2006). Intriguingly, the same has been observed in knock-out 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 uncharacterised PAP1, PAP3 and PAP7 mutants since they exhibit comparable phenotypes. PAP5 or PTAC12/HEMERA is special among all these proteins since it has been recently 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 signalling probably in ubiquitin-mediated proteolysis since it exhibits some similarities to the yeast RAD23 protein. Dual localisation in nucleus and plastids within the same plant cell has been 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 SODs from which the first one has been identified here for the first time as a PAP while the second was described already 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 could be also demonstrated for PAP6 and PAP10. PAP6 is also called FLN1 and its sequence suggest that it belongs to the class of the phosphofructokinases, however, it could be shown that this enzyme lost its ability to recognise 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 which apparently cannot be replaced by another one since the knock-out results in an albino phenotype (Arsova et al., 2010; Schroter et al., 2010). Nevertheless, despite these investigations only little is known about the true PAP functions and further characterisations will be necessary to unravel the specific roles of the distinct PAPs in the RNA polymerase complex.

**Impact of PAP gene knock-outs on plastid development**

The major common feature of all PAPs, regardless of their predicted/detected function, is that a knock-out of the corresponding gene always results in a severe defect in chloroplast develop-
development. In knock-out mutants of PAP2, 4-6 and 8-10 this is accompanied by high NEP-dependent and a low PEP-dependent transcript accumulation (Table II and references therein). This suggests that PAP knock-outs cause a block of PEP activity which prevents the transition of plastid transcription from a NEP-dependent into PEP-dependent mode in the same manner as it could be observed for rpo-gene knock-out mutants of tobacco (Allison et al., 1996; Hajdukiewicz et al., 1997; De Santis-Maciossek et al., 1999) and rpo-knock down mutants of Arabidopsis (Chateigner-Boutin et al., 2008; Zhou et al., 2009; Chateigner-Boutin et al., 2011).

For Sinapis alba extensive biochemical data exist which 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). Combination of these data with the biochemical results reported here suggests an explanation for the observed phenotypes in the Arabidopsis PAP knock-out mutants (Figure 4). In early seedling development the rpo-subunits of PEP are expressed by the nuclear-encoded NEP enzyme representing a first essential check-point 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 prokaryotic-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 eukaryotic-like enzyme complex, the PEP-A enzyme. This, most likely, involves first a post-translational 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 nuclear-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 – 48 hours 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 check-point 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 undisputable the reason for it remains obscure. In principle there are three possibilities why lack of any PAP results in block or severe disturbance of plastid development (Figure 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 or third, the proteins are required for the integrity of the complex itself. One could imagine that failure in each one of these processes lead 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 able of faithful transcription rendering it unlikely that addition of PAPs is essential for 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 of their predicted function is potentially involved in gene expression (Table II). The NEP expression profile observed in some mutants, however, point more to a complete inactivation of the PEP-A complex which is unlikely if just one regulatory event is affected. Knock-outs 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 knock-out mutant lines. The ivory phenotype of many of the mutants indicates that the plastids although PEP-deficient are still active, producing carotenoids and are also able to divide since older leaves of plants grown on sucrose-supplemented medium appear yellow. This suggests that the plastids are arrested in a 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 knock-out 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 the lacking 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 programme of PAP5/PTAC12/HEMERA could serve both genetic compartments.
CONCLUSIONS
We generated a comprehensive and complete catalogue of the subunits of the chloroplast RNA polymerase from mustard comprising five subunits encoded by plastid rpo genes and ten 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 co-expression pattern which is mainly determined by developmental programmes pointing to the reconfiguration of the PEP complex as an essential step in plastid and plant development. Our check-point model explains the white/ivory/pale-green phenotypes of PAP knock-out 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 knock-out 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, L., var. Albatros) were grown for 7 days 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.

Heparin-Sepharose chromatography of chloroplast proteins
Two kg of cotyledons were homogenised 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 sucrose gradient centrifugation, lysed and subjected to a heparin-Sepharose CL-6B (HS) 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 (NH₄)₂SO₄. 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 dialysed 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 first dimension a blue native PAGE with 4 – 12 % acrylamide gradient gels followed by a denaturing SDS PAGE in 7 - 17 % acrylamide gradient gels as a second dimension as recently described (Schroter et al., 2010).

Tryptic in-gel digestion and LC/ESI-MS/MS analysis
Peptide generation of proteins from silver stained gels was performed by tryptic in-gel digestion of cut-out spots using a described protocol (Mortz et al., 1994; Stauber et al., 2003) with minor modifications. Peptides from digested proteins were analyzed by liquid chromatography/electrospray ionization-tandem mass spectrometry (LC/ESI-MS/MS) using a LCQ™ DecaXP ion trap mass spectrometer (Thermo Finnigan, San Jose, CA). Nano-LC was performed using an UltiMate™ Nano LC and Famos™ Autosampler HPLC unit and a reverse phase C18 PepMap™100, 3 µm, 100Å Nano-column (75 µm i.d. x 15 cm) (Dionex Corporation, Sunnyvale, CA). 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 µl/min 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 95% B for 18 min, switched back to 5% B in 0.5 min and held 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 comprised of a full MS scan of the mass range m/z 450 to 1200, followed by 3 MS/MS scans for the three most abundant ions. Sample run and data acquisition was performed using the Xcalibur™ software (Version1.3 © Thermo Finnigan 1998-2001).

MS data analysis
For peak list generation the Ceate DTA tool of TurboSEQUEST® (v. 27 (rev. 12) © 1999-2002 Molecular Biotechnology Univ. of Washington J.Eng/S.Morgan/J.Yates Licensed to ThermoFinnigan Corp.) was used with default settings. Database search was conducted with TurboSEQUEST® (v. 27 (rev. 12) © 1999-2002 Molecular Biotechnology Univ. of Washington J.Eng/S.Morgan/J.Yates Licensed to ThermoFinnigan Corp.) against a Brassicales protein database of NCBI (NCBI Brassicales 2008.09.09.; 154464 sequences). The enzyme specific-
itivity was set to trypsin strict and no missed cleavages were permitted. As variable modifications the carboxy-amido-methylation of cysteine (57.0293), oxidation of methionine (15.9949) and phosphorylation of serine, threonine and tyrosine (79.9663) were included. The mass tolerance for precursor ions was set to 1.5 Da and 0 Da for fragment ions. Calculated cross-correlation ($X_{corr}$) values for significantly matching sequences had to be equal or above 1.5, 2.0, or 3.5 for singly, doubly, or triply charged precursor ions, respectively, and the $\Delta$Corr 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 and consequently peptides match to several equivalent proteins of *A. thaliana* and other brassicales. Therefore the protein entry of the first complete sequence of *A. thaliana* within the list of matching entries is given in the results. Alternatively a representative species is given in the case that the *A. thaliana* sequence is not matching.

**RpoC$_1$ transcript analysis**

*RpoC$_1$*-transcript splice sites were analysed by RT-PCR using cDNA from *A. thaliana* Col-0 and *Sinapis alba* generated from total RNA with random hexamer oligonucleotides. Primers were *rpoC1*-fwd: 5´-AATTGGCTTAGTTTCTCCTCAG-3´ and *rpoC1*-rev: 5´-CCCTTCTTCTTCATTGTTTCC-3´. Preparation of total RNA for cDNA synthesis, PCR and RT-PCR programmes followed standard procedures described at The Arabidopsis Information Resource (http://www.arabidopsis.org).

**In silico and genetic analyses of Arabidopsis PAP knock out 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 NCBI (http://www.ncbi.nlm.nih.gov/pubmed/) for publications concerning either the encoded proteins or corresponding knock-out or knock-down mutant lines in various plant species. In addition, *Arabidopsis* T-DNA knock-out mutant collections at The Arabidopsis Information Resource (http://www.arabidopsis.org) were screened for respective tagged lines. Identified knock-out lines were compared to published data and checked for phenotypic descriptions in research reports, the “Chloroplast Function Database” (Myouga et al., 2010) (compare 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 detection of T-DNA location were: LB3: 5´-
TAGCATCTGAATTTACAAATCTCGATACAC-3’; CS16115-(pap3/ptac) forward: 5’-TCAGGGAGCGTTTGTTGACAT-3’; CS16155-(pap3/ptac10) reverse: 5’-GGTGATCAGAGAGCAGCCCTT-3’; GK-443A08-forward: 5’-CAAAATAGCGAGTCCTACGGTG-3’; GK-443A08-reverse: 5’-GATCAATTCCCAAAGGAAGC-3’; SAIL_566_F06-forward: 5’-AGAAGGTTCAGAGAGCAGCCCT-3’; SAIL_566_F06-reverse: 5’-TGCAGAGAATGATCAATCGTG-3’. Primers for RT-PCR were: CS16155-f: 5’-TCAGGGAGCGTTTGTTGACAT-3’; At3g48500-r: 5’-TCAGTCTGTCAAGACTTGAG-3’; pap6-fwd: 5’-CAAGATCAAGTTTAAGGACGG-3’; pap6-rev: 5’-GTTCCATCAGTCAACCAG-3’; pap7-fwd: 5’-CAACAACAGAAACGAATCCT-3’ and pap7-rev: 5’-CAGAGAACTTTATAGCATCCA-3’. PCR and RT-PCR programmes followed standard procedures described at The Arabidopsis Information Resource (http://www.arabidopsis.org).

Phenotypic appearance of the homozygous lines was tested on MS medium containing 2% sucrose as described recently (Schroter et al., 2010).

Co-expression analysis of PAPs

For analysis of co-expression we used the Arabidopsis co-response database (http://csbdb.mpimp-golm.mpg.de, (Steinhauser 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 non-parametric Spearman’s Rho rank correlation $r_s$ (ranging from +1 to -1) was obtained for each pair, linked to a value-dependent colour code for visualisation and given as a matrix (compare Supplemental Figure S1).

Supplemental Data

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

**Supplemental Figure S1.** Co-expression analysis of PAPs.

**Supplemental Figure S2.** Computer translation of the unspliced $rpoC_1$ 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 co-expression analysis.

LITERATURE CITED


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: -


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


10.1016/j.plphysci.2011.01.005


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


FIGURE LEGENDS

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. 150 µg of total protein were separated, fixed and proteins were stained with silver. Running directions of first and second dimensions are indicated by arrows. Sizes of marker proteins separated in parallel on the same gels are given in the margins. Single subunits within the PEP complexes which gave significant hits in the data bases 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.

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. Grey background: Peptides identified in both subunits. Overlaps between neighbouring peptides are indicated in dark grey. Grey 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 (arrowhead) are indicated. B) Results of RT-PCR analysis and corresponding gene models indicating potential splice variants of rpoC1. Left panel: EtBr-stained RT-PCR products using Arabidopsis and Sinapis cDNA and rpoC1-specific primers. The upper band represents a product generated from the genomic DNA as depicted in the right panel. The lower band represents a product generated from the spliced RNA. Right panel: Gene model and splice variants. Coding sequence is given as thick black bar, the intron as thin black bar. A potential alternatively spliced area is given as hatched bar.

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 position 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 panel A. Lines homozygous for the pap3/ptac10, pap6/fln1 and pap7/ptac14 T-DNA insertion fail to express the wild-type allele. Asterisks indicate bands derived by ge-
nomic DNA. C) WT and homozygous pap3/ptac10, pap6/fln1 and pap7/ptac14 plants germinated on Petri-dishes with media. Seeds of the T-DNA insertion line were surface-sterilised and placed on sterile agar plates containing MS medium supplemented with 2% sucrose.

**Figure 4:** Check-point model describing the reconfiguration of the plastid RNA polymerase complex as 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 sigma factors is assumed. Upon initiation of photomorphogenesis it is modified first by post-translational changes of rpo-subunits (via unknown modifying enzymes) and second by addition of PAPs generating the structurally more complex PEP-A. White arrows indicate the flow of events required for PEP-A build-up. Thin black arrows indicate the action or involvement of nuclear encoded proteins delivered in a fixed sequence which follows a distinct developmental programme in the nucleus. Thick black arrows indicate transcription activity. Dotted lines indicate the possible impact of a PAP gene knock-out 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 knock-out mutants. For further details see text.
Tables

Table I: RNA polymerase subunits identified by ESI-MS/MS after 2D-BN-PAGE. Subunit: PAPs as given in Figure 1. Accession: At gene accession numbers; MW kDa (theor. –cTP/app.): Theoretical molecular weight without chloroplast transit peptide and apparent molecular weight observed in gel. Identity/protein domain: Identity of PAP and its predicted protein domain(s) as obtained by CD data base (Marchler-Bauer et al., 2011). Function: Subunit functions predicted from sub-domains or proposed/shown by experiment (ex.). Reference: Source for functional classification.
<table>
<thead>
<tr>
<th>Subunit</th>
<th>AGI Accession</th>
<th>MW in kDa (theor. –cTP/ app.)</th>
<th>Identity/protein domain</th>
<th>Function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>RpoC₂</td>
<td>AtCg00170</td>
<td>156/ 141</td>
<td>Beta’- subunit</td>
<td>DNA-binding</td>
<td>(Igloi and Kos-ssel, 1992)</td>
</tr>
<tr>
<td>RpoB</td>
<td>AtCg00190</td>
<td>121/ 118</td>
<td>Beta subunit</td>
<td>RNA synthesis</td>
<td>(Igloi and Kos-ssel, 1992)</td>
</tr>
<tr>
<td>PAP1</td>
<td>At3g04260</td>
<td>110/ 110</td>
<td>pTAC3/SAP domain</td>
<td>DNA/RNA-binding</td>
<td>Prediction</td>
</tr>
<tr>
<td>PAP2</td>
<td>At1g74850</td>
<td>89/ 107</td>
<td>pTAC2/PPR repeat, SMR domain</td>
<td>RNA metabolism</td>
<td>Prediction</td>
</tr>
<tr>
<td>RpoC₁</td>
<td>AtCg00180</td>
<td>80/ 85</td>
<td>Beta’-l subunit</td>
<td>unknown</td>
<td>(Igloi and Kos-ssel, 1992)</td>
</tr>
<tr>
<td>PAP3</td>
<td>At3g48500</td>
<td>77/ 76</td>
<td>pTAC10/S1-like domain</td>
<td>RNA-binding</td>
<td>Prediction</td>
</tr>
<tr>
<td>PAP4</td>
<td>At5g23310</td>
<td>26/ 74</td>
<td>Iron superoxide dismutase 3</td>
<td>Superoxide detoxification (ex.)</td>
<td>(Myouga et al., 2008)</td>
</tr>
<tr>
<td>RpoC₁</td>
<td>AtCg00180</td>
<td>80/ 72</td>
<td>Beta’-s subunit</td>
<td>unknown</td>
<td>(Igloi and Kos-ssel, 1992)</td>
</tr>
<tr>
<td>PAP5</td>
<td>At2g34640</td>
<td>57/ 70</td>
<td>pTAC12-HEMER/A RAD23 do- main</td>
<td>Probably involved in ubiquitin-mediated proteolysis (ex.)</td>
<td>(Chen et al., 2010)</td>
</tr>
<tr>
<td>PAP6</td>
<td>At3g54090</td>
<td>49/ 52</td>
<td>FNL1/pfK-B-2 fructo-kinase</td>
<td>Fructose phosphorylation; substrate specificity not detectable (ex.)</td>
<td>Prediction; (Arsova et al., 2010)</td>
</tr>
<tr>
<td>PAP7</td>
<td>At4g20130</td>
<td>49/ 52</td>
<td>pTAC14/SET-domain, Rubis-subs- bind</td>
<td>Lysine methyltransferase, interaction with histones or Rubisco</td>
<td>Prediction</td>
</tr>
<tr>
<td>RpoA</td>
<td>AtCg00740</td>
<td>38/ 38</td>
<td>Alpha subunit</td>
<td>Complex stabilization</td>
<td>(Igloi and Kos-ssel, 1992)</td>
</tr>
<tr>
<td>PAP8</td>
<td>At1g21600</td>
<td>31/ 37</td>
<td>pTAC6</td>
<td>unknown</td>
<td>unknown</td>
</tr>
<tr>
<td>PAP9</td>
<td>At3g51100</td>
<td>29/ 29</td>
<td>Iron superoxide dismutase 2</td>
<td>Superoxide detoxification (ex.)</td>
<td>(Myouga et al., 2008)</td>
</tr>
<tr>
<td>PAP4</td>
<td>At5g23310</td>
<td>26/ 26</td>
<td>Iron superoxide dismutase 3</td>
<td>Superoxide detoxification (ex.)</td>
<td>(Myouga et al., 2008)</td>
</tr>
<tr>
<td>PAP10</td>
<td>At3g06730</td>
<td>12/ 13</td>
<td>TRX-z/ thioredoxin WCGPC motif</td>
<td>Redox regulation (ex.)</td>
<td>(Arsova et al., 2010)</td>
</tr>
</tbody>
</table>
Table II: Phenotyping of *Arabidopsis* PAP knock-out mutants. Growth phenotype: Developmental appearance of knock-out mutants, Suc: Viable only on sucrose-supplemented medium; Plastid structure: Plastid morphology in knock-out/silenced mutants (t.: thylakoids; p.: plastoglobuli enrichment); Molecular phenotype: NEP, NEP expression profile of plastid transcript accumulation; Reference: Source of phenotypic descriptions. N. d.: not described. Classification as “ivory” is based on own observations.

<table>
<thead>
<tr>
<th>Subunit</th>
<th>Growth phenotype</th>
<th>Plastid structure</th>
<th>Molecular phenotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAP1</td>
<td>Albino/ivory; Suc</td>
<td>n. d.</td>
<td>n. d.</td>
<td>(Myouga et al., 2010)</td>
</tr>
<tr>
<td>PAP2</td>
<td>Pale-yellow green; Suc</td>
<td>Impaired; few t.; p.</td>
<td>NEP</td>
<td>(Pfalz et al., 2006)</td>
</tr>
<tr>
<td>PAP3</td>
<td>Albino/ivory; Suc</td>
<td>n. d.</td>
<td>n. d.</td>
<td>This work</td>
</tr>
<tr>
<td>PAP4</td>
<td>Pale-green; Suc</td>
<td>Impaired; no t.; p.</td>
<td>NEP</td>
<td>(Myouga et al., 2008)</td>
</tr>
<tr>
<td>PAP5</td>
<td>Albino/ivory; Suc</td>
<td>Impaired; no t.; p.</td>
<td>NEP</td>
<td>(Pfalz et al., 2006; Chen et al., 2010)</td>
</tr>
<tr>
<td>PAP6</td>
<td>Albino/ivory, Suc</td>
<td>n. d.</td>
<td>n. d.</td>
<td>This work,</td>
</tr>
<tr>
<td>PAP7</td>
<td>Albino/ivory; Suc</td>
<td>n. d.</td>
<td>n. d.</td>
<td>This work</td>
</tr>
<tr>
<td>PAP8</td>
<td>Albino/ivory; Suc</td>
<td>Impaired; no t.; p.</td>
<td>NEP</td>
<td>(Pfalz et al., 2006)</td>
</tr>
<tr>
<td>PAP9</td>
<td>Pale-green; Suc</td>
<td>Impaired; no t.; p.</td>
<td>NEP</td>
<td>(Myouga et al., 2008)</td>
</tr>
<tr>
<td>PAP10</td>
<td>Albino/ivory; Suc</td>
<td>Impaired; no t.; p.</td>
<td>NEP</td>
<td>(Arsova et al., 2010; Schroter et al., 2010)</td>
</tr>
</tbody>
</table>
Chloroplast isolation

Lysis

Heparin-Sepharose chromatography

2D-BN-PAGE

Glycerol gradient centrifugation

15 proteins

ESI-MS/MS

Sequest

15 proteins

Edman + MALDI MS

PAP1

PAP2

PAP3

PAP4

PAP5

PAP6

PAP7

PAP8

PAP9

PAP10

β'-l

β'-s

α

β

β''
1 MIDRYKHQQL RIGLVSPQQI SAWATKIIPN GEIVGEVTKP YTFHYKTNKI
31 EKDGLFCERI FGPIKSGICA CGNYRVIQDE KEPKFCEQC GVEFVDSRIR
51 RYMQMYIKLT CPVTHVWY1K RLPSYIANLL DKPLKEEGEL YVCDMSFARP
51 ITKPTFRLR RGSFEYEIOQS WKYSIPLFFR TQGFDFRFRN EISTGAGAIR
51 EQLADLDRLI IIENSLVEWKG QLGEEGTGQN EWEDRKLIVRR KDFLVRRMEL
51 AKHFIRTNE PEVMVLCLLP VLPPELRPII QIEGGKLMS DINELYRRTVI
51 YRNNLTDLIL TTSRSTPGEL VMCQEKTVQE AVDTLLDNGI RGQPMRGHN
51 KVKSFSDVI EGKEGRFRET LLGKRVDGSG RSVIVVGPSEL SLHRCGLPREF
51 IALILEFQTFV IRGLIRQHLA SNIGVAKSQI REKKPIVWEI LQEVMPGHPV
51 LNNAPTLPHE LGIQSFQPII VEGRTICLHP LVCKGFAEND DGDQMAVHVP
51 LSLEAQAEAR LLMFShMNll SPAIGDPISV PTQDMLGLY VLTSGTTRGI
51 CANRYNCNR KNYQNERIYE TNYKYTEKPF FCNSYDAIGA YRQKPINLDS
51 PLWLRWQLDQ RVIASREVPI EVHYESFGNY HEIYAHILIV RSVKKENFCI
51 YRTTVGHIS FYREIEBAILQ GFSQACSYDT0

rpoC1

- genomic: 1432 bp
- spliced: 645 bp
- alt. spliced: 454 bp

500 bp
k.o. gene
Nucleus

Basic developmental programme

Photo-morphogenetic programme

Modifying enzymes?

PAPs

PEP-B

Modification Assembly

PEP-A

NEP

rpo genes

Plastid

1

2

3