Plastid RNA polymerases: orchestration of enzymes with different evolutionary origins controls chloroplast biogenesis during the plant life cycle

Thomas Pfannschmidt¹,²,³,⁴*, Robert Blanvillain¹,²,³,⁴, Livia Merendino¹,²,³,⁴, Florence Courtois¹,²,³,⁴, Fabien Chevalier¹,²,³,⁴, Monique Liebers¹,²,³,⁴, Björn Grübler¹,²,³,⁴, Elisabeth Hommel¹,²,³,⁴ and Silva Lerbs-Mache¹,²,³,⁴

¹ Université Grenoble-Alpes, F-38000 Grenoble, France
² CNRS, UMR5168, F-38054 Grenoble, France
³ CEA, IRTSV, Laboratoire de Physiologie Cellulaire & Végétale, F-38054 Grenoble, France
⁴ INRA, USC1359, F-38054 Grenoble, France

* To whom correspondence should be addressed. Email: Thomas.Pfannschmidt@ujf-grenoble.fr

Received 9 July 2015; Revised 9 August 2015; Accepted 11 August 2015

Editor: Christine Foyer

Abstract

Chloroplasts are the sunlight-collecting organelles of photosynthetic eukaryotes that energetically drive the biosphere of our planet. They are the base for all major food webs by providing essential photosynthates to all heterotrophic organisms including humans. Recent research has focused largely on an understanding of the function of these organelles, but knowledge about the biogenesis of chloroplasts is rather limited. It is known that chloroplasts develop from undifferentiated precursor plastids, the proplastids, in meristematic cells. This review focuses on the activation and action of plastid RNA polymerases, which play a key role in the development of new chloroplasts from proplastids. Evolutionarily, plastids emerged from the endosymbiosis of a cyanobacterium-like ancestor into a heterotrophic eukaryote. As an evolutionary remnant of this process, they possess their own genome, which is expressed by two types of plastid RNA polymerase, phage-type and prokaryotic-type RNA polymerase. The protein subunits of these polymerases are encoded in both the nuclear and plastid genomes. Their activation and action therefore require a highly sophisticated regulation that controls and coordinates the expression of the components encoded in the plastid and nucleus. Stoichiometric expression and correct assembly of RNA polymerase complexes is achieved by a combination of developmental and environmentally induced programmes. This review highlights the current knowledge about the functional coordination between the different types of plastid RNA polymerases and provides working models of their sequential expression and function for future investigations.

Key words: Chloroplast biogenesis, developmental regulation, gene expression, nucleo–plastid interaction, plants, plastid RNA polymerases.

Introduction

Today’s plastids are a group of morphologically and functionally diverse cellular organelles that are surrounded by a double membrane and are found specifically in plant and algae cells. The most prominent representative of these organelles is the chloroplast, which is found in green tissues of vascular plants, mosses, and green algae. This plastid type is the site of photosynthesis, but also of nitrate and sulphate fixation, and parts of many other important cellular biosynthetic pathways.
including amino acid and lipid biosynthesis. Besides chloroplasts, other non-green plastid forms exist in plants such as amyloplasts and elaioplasts, which are important for starch and oil accumulation, respectively, in roots and storage organs, or chromoplasts, which generate carotenoids responsible for the colours in fruits or flowers. All these different plastid types develop from a common undifferentiated precursor form, the proplastid, which is also the form involved in cytoplasmic inheritance. The respective plastid type that is generated from this proplastid is determined by the tissue context of the cell in which it resides. Furthermore, the different plastid types are convertible into each other, even after differentiation, when respective environmental and/or cellular influences change (e.g. the chloroplast-to-chromoplast conversion upon fruit ripening or in the opposite direction following illumination). Thus, plant plastids display a remarkable morphological and functional diversity and flexibility (Lopez-Juez and Pyke, 2005; Jarvis and Lopez-Juez, 2013).

Plant plastids evolved from a monophyletic evolutionary event called primary endosymbiosis whereby a heterotrophic mitochondriated protist engulfed a photosynthetically active, free-living cyanobacterium. In the course of evolution, this cyanobacterium was ‘tamed’ by the host cell and transformed into a stable and inheritable cell compartment, the plastid (Cavalier-Smith, 1999; Reyes-Prieto et al., 2007). This included a massive horizontal gene transfer from the endosymbiont towards the genome of the host, reshaping its biochemical properties (Martin et al., 2002; Timmis et al., 2004; Reyes-Prieto et al., 2006). Current genome analyses strongly suggest that the process of endosymbiosis was facilitated by an infection of the host cell with Chlamydiae (Ball et al., 2013). The result of this complex evolutionary event was the generation of the first photosynthetic eukaryote. This ancient lineage split into three branches, the ‘green’ lineage (chloroplastida), the ‘red’ lineage (rhodophyta), and the glaucophyta. From the green lineage around 450 million years ago, the green algae and plants developed. The benefit provided by plastids was such an advantage that several secondary and even tertiary endosymbiotic events occurred within the green and red algae resulting in the generation of further ‘plastidiated’ eukaryotes such as the dinoflagellates, the diatoms, the euglenoids, and the apicomplexa (Delwiche, 1999; Stoebe and Maier, 2002). This fascinating complexity is one of the prime examples of the power of evolution within the tree of life.

As a remnant of their cyanobacterial origin, today’s plastids still contain their own genome, the plastome, and a fully functional machinery for its expression. The plastome is comprised of around 120 genes in plants and up to around 200 genes in some red algae; it encodes mainly components of the photosynthetic and gene expression machineries (Sugiuira, 1992; Martin et al., 2002; Green, 2011). This is, however, only a small part of the original coding capacity of the genome of the ancient cyanobacterial precursor. Recent proteome analyses identified up to 1660 non-redundant proteins within Arabidopsis thaliana plastids [available in the online databases AT_CHLORO and the Plant Proteome Data Base (PPDB)], indicating that the vast majority of the chloroplast proteome is encoded in the nucleus, translated in the cytosol, and subsequently imported into the organelle (Sun et al., 2009; Ferro et al., 2010). The present protein catalogues are probably not complete, as bioinformatic predictions of transit peptides identify many more potential plastid-localized proteins in the genome. Estimates range from 2000 to 3500 proteins that may be present in plastids. Low abundance and spatio-temporal expression variations probably represent the major constraints in detection and identification of the lacking proteins (Abdallah et al., 2000; Van Wijk and Baginsky, 2011).

The dependency of plastids on the import of proteins from the cytosol gained the host cell full control over development and differentiation of the organelle (i.e. the cell type can fully determine the plastid type). Therefore, usually only one type of plastid is found in a given cell. Rare exceptions are some cells at an early developmental stage in variegation mutants that possess both normal chloroplasts and aberrant plastids. Segregation of these plastids during further cell division subsequently leads to the formation of white and green patches or stripes in the leaves of the mutants (Wetzel et al., 1994; Sakamoto, 2003).

On the other hand, all major plastid protein complexes involved in photosynthesis or gene expression contain several plastid-encoded subunits that are essential for the assembly of these complexes (Allen et al., 2011; Pfälz and Pfannschmidt, 2013). Thus, the plastids retained some influence on the build-up of their own protein machineries and are, therefore, regarded as genetically semi-autonomous. The plastid protein complexes, in consequence, represent an evolutionary patchwork of subunits encoded in two different genetic compartments. Stoichiometric production and correct assembly of these subunits thus require a faithful and tight communication between the two compartments (Rodermel, 2001; Pogson et al., 2008; Woodson and Chory, 2008). This communication must coordinate the expression of nuclear genes for plastid proteins (typically present in one or few gene copies in the nucleus) with that of plastid-encoded genes (present in up to around 100 plastome copies per plastid and up to ~100 plastids per cell). This huge difference in gene copy number needs to be adjusted by the gene expression machineries in the nucleus and plastids (Lopez-Juez and Pyke, 2005; Dietzel et al., 2008). The RNA polymerases found within the different genetic compartments are among the key regulators of this complex interplay. By selective transcription of genes essential for the corresponding developmental stage of the tissue, they initiate a coordinated adjustment of gene expression events according to the needs of the organelles.

The RNA polymerases and their associated regulatory proteins comprise factors of eukaryotic, prokaryotic, and phage ancestry, and reflect the complex evolutionary integration of a prokaryotic gene expression system into a eukaryotic host cell. Most of our current knowledge about this interplay was obtained by studying chloroplast biogenesis, while much less is known about these processes in other plastid types (e.g. in chromoplasts or amyloplasts), although some studies exist (Kahlau and Bock, 2008; Valkov et al., 2009). For reasons of simplicity, this review focuses largely on processes occurring during the light-dependent transition from the proplastid/epplast to the chloroplast, accepting that some aspects of the described events might be different in other plastid types.
The different types of plastid RNA polymerase

Plastids possess two types of RNA polymerase: (i) one multi-subunit, plastid-encoded prokaryotic-type RNA polymerase (PEP), and (ii) two different single-subunit, nuclear-encoded phage-type RNA polymerase(s) (NEP) (Shiina et al., 2005; Liere et al., 2011; Börner et al., 2015). The first requires interaction with sigma-like factors that mediate promoter recognition as well as with additional PEP-associated proteins (PAPs) of various functions, all encoded by the nuclear genome (Fig. 1) (Schweer et al., 2010; Lerbs-Mache, 2011; Pfalz and Pfannschmidt, 2013; Yu et al., 2014; Chi et al., 2015). A proper function of the plastid transcription machinery therefore requires first the action of the eukaryotic nuclear RNA polymerase II, which is responsible for the expression of the nuclear-encoded components. The development of functional plastids, then, depends on the coordinated action of the two different types of plastid RNA polymerases, which initiate a sequential and highly organized expression cascade of essential plastid genes followed by further maturation of transcripts and appropriate translation. The RNA metabolism and transcript maturation as well as the regulation of plastid translation are highly sophisticated processes, especially in mature chloroplasts, and involve a high number of regulatory proteins (Barkan and Goldschmidt-Clermont 2000; Monde et al., 2000; Eberhard et al., 2002), which cannot be covered by this review. It is, however, important to keep in mind that these levels of gene expression are tightly interlinked. While plastid RNA polymerases provide transcripts for proteins (including transcripts for components of the large and small ribosome subunits) and tRNAs as substrate for translation (Williams-Carrier et al., 2014), plastid translation is, in turn, essential for the generation of plastome-encoded Rpo subunits. In addition, both levels are highly affected by the number of plastid DNA copies available (Udy et al., 2012). Transcription, RNA maturation, and translation thus exhibit a mutual interdependency that needs to be highly coordinated. This coordination during differentiation and build-up phases of the organelle is mediated primarily by developmental programmes yielding the typical tissue-specific plastid types such as chromoplasts, amyloplasts, or chloroplasts. In fully matured chloroplasts, a functional regulation that depends on environmental cues such as light variations becomes more dominant. This functional regulation is mediated mainly by the chloroplast itself, for instance by redox signals from photosynthesis (Pfannschmidt et al., 2003). The orchestration of the different transcriptional machineries, thus,

![Fig. 1. Schematic overview summarizing gene location, protein trafficking, and transcriptional activity of all components of the plastid transcriptional machinery and their mutual interdependency. The three genetic compartments of plant cells are depicted. Coloured boxes within them represent genes, with ovals and circles corresponding protein components. The nomenclature of genes and proteins given in these symbols follows that given in the text. Black arrows indicate the traffic of information from a gene via transcripts and translation (indicated by white boxes) towards the place of action of the corresponding protein. Blue symbols: Nuclear genes of NEP enzymes are represented by boxes. Plastid-located NEPs (ovals) are depicted above their target genes in plastids (classes II and III, including rpoA, -B, -C1, and -C2). Rpo subunits are translated within the plastid and generate the PEP core complex. Green symbols: PAP genes in the nucleus are probably expressed as a regulon requiring an as-yet-unknown master regulator (orange oval) for coordinated expression. PAP subunits assemble around the PEP core. Relative positions of PAPs are based on literature data describing protein–protein interactions in a yeast two-hybrid system (see main text), PAPs drawn with a red line display either a true (full line) or potential (broken line) second location in the nucleus. The still-unknown pathways of nuclear trafficking via a cytosolic or a plastid route are indicated by broken red arrows. Yellow symbols: Nuclear-encoded plastid sigma factors are transported into the plastid and mediate specific promoter recognition for the PEP complex (blue Rpo subunits assembled with all green PAPs) for the target genes (yellow boxes: classes I and II as well as sigma factor-specific genes psbN, atpH, and ndhF). For definition of classes I, II and III see text.](https://academic.oup.com/jxb/article-abstract/66/22/6957/2893280)
requires the integration of developmental as well as physiological regulation networks and integrates signals from the nucleus towards the plastid (anterograde signalling) and signals from the plastid towards the nucleus (retrograde or plastidial signalling) (Pogson and Albrecht, 2011).

The bacteria-like DNA-dependent RNA polymerase of plastids (PEP)

Core complex

Transcriptional activity in plastids was already reported in the 1960s (Kirk, 1964), but the nature of this activity remained obscure until molecular biology techniques allowed the sequencing of plastid genes (Hudson et al., 1988; Igloi et al., 1990). Nowadays, it is common knowledge that plastids possess a RNA polymerase complex of prokaryotic origin that is composed of subunits homologous to the well-studied subunits of RNA polymerase from Escherichia coli or other bacteria. So far, all sequenced chloroplast genomes of vascular plants have revealed the presence of genes that exhibit significant sequence homologies to bacterial RNA polymerase (rpo) genes (Fig. 1) (Igloi and Kossel, 1992). A comparison was performed of structural Rpo homology between Arabidopsis and E. coli at the protein level using Phyre2 and found homologies that ranged from 37% (RpoC2) to 85% (RpoA) and 95% (RpoB, RpoC1) (with respect to the lengths of the proteins). The genes for these proteins are typically arranged as a large operon comprising the genes rpoB–rpoC1–rpoC2 and a single gene rpoA being part of another operon that contains several genes for ribosomal components (Green, 2011). These rpo genes encode the catalytic subunit RpoB (or β-subunit), the subunit RpoC1 (or β′-subunit), which has no clearly assigned function yet, and the subunit RpoC2 (or β″-subunit), which is believed to possess a DNA binding function. Most likely, the genes rpoC1 and rpoC2 emerged from a split of the original rpoC gene of the cyanobacterial ancestor. In addition, in dicotyledonous plants, an intron can be found within the rpoC1 gene that is not present in monocotyledons. This observation might suggest different evolutionary constraints for the evolving transciption machinery in these two plant clades (Igloi and Kossel, 1992). Whether this is connected to specific functional differences of the RNA polymerase subunits between the two clades and/or whether this reflects an adaptation to the different developmental programmes underlying chloroplast biogenesis in monocotyledons and dicotyledons still needs to be elucidated. Interestingly, the rpo genes in the green algae Chlamydomonas reinhardtii are monocistronic units dispersed over the plastome, indicating that the plastome organization within the green lineage was not thoroughly maintained (Maul et al., 2002). The subunit RpoA (or α-subunit) probably serves as a stabilizing subunit of the complex and, in analogy to its bacterial counterpart, is assumed to occur as a dimer, although direct supporting evidence does not exist. The basic RNA polymerase complex with the stoichiometry α2, β, β′, β″ is referred to as the core enzyme and is capable of transcriptional elongation in vitro and probably also in vivo. Plastid knockout mutants in tobacco or nuclear knockout mutants in Arabidopsis, both with structural or functional deficiency for any of these rpo genes, display an albino or yellowish phenotype with arrested plastid development, indicating the absolute requirement of Rpo subunits for proper chloroplast development. All these mutant plants have been reported to be viable on sucrose-supplemented medium, indicating that the missing chloroplast function can largely be rescued by an external carbon source (Allison et al., 1996; Hajdukiewicz et al., 1997; De Santis-Maciossek et al., 1999; Chateigner-Boutin et al., 2008, 2011; Zhou et al., 2009).

Sigma factors

The typical promoters of PEP-transcribed genes resemble prokaryotic promoters and correspondingly comprise −35 (TGACA)- and −10 (TATAAT)-like-sequence motifs (Shiina et al., 2005; Liere et al., 2011; Chi et al., 2015). For specific promoter recognition and transcription initiation, the PEP core complex is thus dependent on the interaction with prokaryotic-like sigma factors (Allison, 2000), which were first identified by biochemical means (Bülow and Link, 1988; Lerbs et al., 1988). In contrast to the other Rpo subunits, the genes for these factors were found to be encoded in the nucleus (Liu and Troxler, 1996; Tanaka et al., 1996). Like many other genes that were originally located in the genome of the engulfed cyanobacterium, they went through a horizontal gene transfer towards the nucleus of the host cell and are therefore under the control of the nuclear transcription system. In Arabidopsis, six different sigma factors are known and are designated Sig1–Sig6 (Fig. 1) (Schweer et al., 2010; Lerbs-Mache, 2011; Chi et al., 2015). In contrast to mutants with defects in the core components, single sigma-factor T-DNA inactivation mutants in Arabidopsis do not display severe phenotypic defects, pointing to a certain level of functional redundancy between the factors. Exceptions to this are sig2 and sig6 knockout mutants, which exhibit light-green phenotypes indicating a more pronounced role of the encoded Sig2 and Sig6 factors in plastid transcription in the early stages of chloroplast differentiation. Apparently, their loss can only partly be replaced by the other sigma factors during the early steps of seedling development. Studies from several laboratories indeed have revealed partial redundancy for sigma factor functions but also identified gene-specific functions for these regulatory factors (Schweer et al., 2010; Lerbs-Mache, 2011; Malik Ghulam et al., 2012). The specific functions, gene expression profiling, regulation of sigma activity/specificity by post-transcriptional modifications, and some evolutionary aspects have already been described and discussed in detail in a number of previous reviews (Lysenko and Kusnetsov, 2005; Schweer et al., 2010; Lerbs-Mache, 2011; Chi et al., 2015). This review will focus mainly on some evolutionary aspects of sigma diversification and functional integration into the eukaryotic host during land plant evolution.

Evolutionary origin of plant sigma factors

Previous phylogenetic analysis of sigma factor genes indicated that, in plants, they all derived from the principal cyanobacterial sigma factor SigA (Hakimi et al., 2000; Mache et al., 2002;
Lysenko and Kusnetsov, 2005). This suggests that, out of the multiple sigma factors present in the ancient cyanobacterium, only the primary sigma factor SigA has been retained in the chloroplast ancestor. Comparison of intron positions of plant sigma factor genes further suggests that Sig1 and Sig2 originated from this SigA common ancestor by gene duplication and that all other sigma factors were later derived from Sig2 (Lysenko and Kusnetsov, 2005). The early evolutionary origin of Sig1 and Sig2 was recently confirmed by showing the presence of the corresponding sigma factor genes in the nuclear genomes of the liverwort Marchantia polymorpha (Kanazawa et al., 2013; Ueda et al., 2013), of the moss Physcomitrella patens (Hara et al., 2001; Ichikawa et al., 2004), and of the spikemoss Selaginella moellendorffii (Banks et al., 2011). In addition to these two genes, the nuclear genome of M. polymorpha encodes two other different sigma factors. One was named Mp_SIGX because it does not relate to anyone of the six higher plant sigma factors (Kanazawa et al., 2013). The other one found to be orthologous to Sig5 of higher plants. Thus, Sig5 has evolved over the same time scale as Sig1 and Sig2 but probably does not originate from Sig2, nor from the cyanobacterial ancestor, because none of the cyanobacterial sigma factors group together with Sig5 (Mache et al., 2002). Sig5 has the weakest similarity to other plant sigma factors. Noticeably, A. thaliana Sig5 is orthologous to E. coli sigma factor RpoH, which is known to be implicated in heat-shock and stress responses of the bacterium (Mache et al., 2002). It seems that this function has been evolutionary conserved in basal and higher land plants, as suggested by the induction of Sig5 gene expression under several stress conditions in liverwort as well as in Arabidopsis (Nagashima et al., 2004; Kanazawa et al., 2013). However, the evolutionary origin of Sig5 remains to be elucidated.

Diversification of plant sigma factors during plant evolution, and co-evolution of sigma factors and promoters As mentioned above, results obtained so far from non-vascular land plants reveal the presence of sigma orthologues to higher plant factors Sig1, Sig2, and Sig5. The other sigma factors, namely Sig3, Sig4, and Sig6, emerged during the evolution of vascular plants after the separation of lycophytes and euphyllophytes. They probably originated by gene duplication from Sig2. Sig4 represents the most recent acquisition of a new sigma factor. It has been found so far only in Superrosidae (Rosales and Geraniales). Its absence in currently known nuclear genomes of monocotyledonous plants suggests that it evolved after the separation of dicots and monocots.

The recent emergence of Sig4 is of special interest as it can be related to a specific functional diversification of sigma function (Schweer et al., 2010; Lerbs-Mache, 2011; Chi et al., 2015). While Sig2 and Sig6 seem to be more general factors required for transcription initiation of many various genes or operons, Sig3 and Sig4 exclusively recognize the psbN and atpH promoters (Sig3) (Zghidi et al., 2007) and the ndhF promoter (Sig4) in Arabidopsis (Fig. 1) (Favory et al., 2005). The first insights into functional specification of sigma factors in non-vascular land plants came from a recent analysis of the plastid transcriptome of a sig1 mutant of M. polymorpha. These results revealed changes that are similar to those observed in mutants of higher plants, suggesting conservation of principal sigma activity of Sig1. However, the ndhF gene is also transcribed by Sig1 in liverwort. A comparison of ndhF promoter regions from liverwort and higher plants reveals considerable alterations in promoter elements between liverwort, several monocotyledons species, and Arabidopsis, testifying to co-evolution of sigma factors and promoters (Ueda et al., 2013). Co-evolution of Sig4 and ndh gene expression is also testified by the concomitant lack of Sig4 gene expression (Zhang et al., 2015) and the loss of ndh genes from the plastid genome in some Geraniaceae (Blazier et al., 2011; Weng et al., 2014).

Geraniaceae plastid genomes are highly rearranged (Weng et al., 2014) and are characterized by remarkably elevated evolutionary rates, especially in rpoB, rpoC1, and rpoC2 genes (Guisinger et al., 2008; Weng et al., 2012). In addition, numerous gene duplications of Sig5 and Sig6 have been observed in various species (Zhang et al., 2015). Therefore, Geraniaceae represent an attractive system to study plastid–nuclear genome co-evolution with respect to the PEP transcriptional apparatus. Variations in genes for RNA polymerase subunits have already been shown to be useful for taxonomy classification and phylogenetic analyses (Taillardat-Bisch et al., 2003; Iyer et al., 2004; Lane and Darst, 2010). Using Geraniaceae as a model system, coordinated evolution rates between interacting plastid and nuclear genes coding for PEP subunits and sigma factors, respectively, have recently been shown (Zhang et al., 2015). The authors suggested that this strong correlation of evolutionary rates might provide an explanation for the observed plastome–genome incompatibility between species within the Geraniaceae.

Emergence of sigma factor phosphorylation as a mechanism for regulating activity/specificity of transcription Plant sigma factors are composed of evolutionarily conserved C-terminal parts and variable N-terminal parts. The N-terminal parts have probably been acquired during the process of integration of the ancestral genes after duplication into different regions of the host’s nuclear genome. The N-terminal parts have functionally diverged during evolution. They are incompatible with prokaryotic transcription systems as shown for E. coli (Hakimi et al., 2000; Mache et al., 2002), and they are not interchangeable between different plant sigma factors (Schweer et al., 2009). In contrast to their bacterial counterparts, plant sigma factors can be modified post-translationally by phosphorylation, and this modification seems to occur exclusively in the N-terminal variable region. The functional consequences of sigma factor phosphorylation have so far been investigated in vivo for Sig1 (Shimizu et al., 2010) and Sig6 (Schweer et al., 2010) in Arabidopsis; however, using bioinformatics, hypothetical phosphorylation sites were predicted in all plant sigma factors. Phosphorylation has been observed in Sig2 by an in vitro approach, and phosphorylation of Sig1 and Sig6 have been associated with changes in the binding efficiency of RNA polymerase–sigma complexes to selected promoters (Türkeri et al., 2012). Phosphorylation of Sig1 is regulated by
redox signals and serves to adapt photosystem stoichiometry to daily changing light conditions via changes in psbA and psaA transcript levels (Shimizu et al., 2010).

Direct phosphorylation of sigma factors represents a mechanism of transcriptional regulation that has been acquired after the endosymbiotic event. However, it is not clear yet whether this mechanism evolved before or after the separation of Viridiplantae and Rhodophyta. Western blot analyses of the four sigma factors of the red alga Cyanidioschyzon merolae (note that these sigma factors cluster together and are not related to Sig1–Sig4 of higher plants as shown for Cyanidium caldarium; Mache et al., 2002) indicate post-translational modifications for Sig3 and Sig4 (Kanesaki et al., 2012). The nature of these modifications is not known, but phosphorylation might be considered.

**PEP-associated proteins (PAPs)**

While genetic data imply a prokaryotic structure of the plastid RNA polymerase complex, biochemical purifications of this enzyme from chloroplasts yielded protein complexes with a much higher subunit number (15–20 depending on the species). These multi-protein complexes thus resemble more the eukaryotic-type RNA polymerases from the nucleus (Lerbs et al., 1985; Rajasekhar et al., 1991). Biochemical studies on mustard cotyledons then revealed that the enzyme in its simple prokaryotic composition (i.e. α, β, β′, β″) could be purified from etioplasts of dark-grown seedlings. This complex can be referred to as the PEP core complex and forms a holo-enzyme together with the sigma factors for transcription initiation. It is able to rapidly acquire additional subunits upon light exposure, which triggers chloroplast formation (Pfannschmidt and Link, 1994). These additional subunits were subsequently identified by mass spectrometry approaches in mustard, tobacco, and Arabidopsis (Pfannschmidt et al., 2000; Suzuki et al., 2004; Pfalz et al., 2006; Steiner et al., 2011). The corresponding genes were identified in the Arabidopsis nuclear genome by homology searches and all displayed a predicted chloroplastic transit peptide at the N terminus. The majority of the subunits of the plastid-encoded RNA polymerase are thus imported from the cytosol (see below).

A major challenge in studying the additional subunits of the plastid RNA polymerase complex by biochemical means is to discriminate between bona fide subunits from contaminating, irrelevant, or non-essential subunits. Plastid DNA and the enzymes responsible for its replication and expression are organized into huge structures resembling bacterial nucleoids. These plastid nucleoids can be visualized in vivo by various optical methods, including DAPI staining, revealing that they are first located at the proplastid envelope and then attach to thylakoids in fully differentiated chloroplasts (Krupinska et al., 2014; Powikrowska et al., 2014). Although nucleoids appear as distinctly shaped bodies in microscopy, they cannot be isolated as such since they do not possess any enclosing membrane or other stabilizing outer structure, and therefore precise subunit catalogues are difficult to obtain. The best approach so far is the purification of the transcriptionally active chromosome (TAC), a huge DNA–protein complex that is capable of in vitro RNA synthesis upon addition of nucleotides. It consists of around 40–60 proteins and was analysed in detail by mass spectrometry in Arabidopsis, mustard, and spinach (Pfalz et al., 2006; Melonek et al., 2012). A total of 35 proteins were identified in Arabidopsis from which 18 were novel and were called pTACs (Pfalz et al., 2006). Many of them are not involved in transcription directly but cover other important functions including structural stabilization of the DNA, membrane anchoring, RNA processing, and ribosome formation. The plastid RNA polymerase complex is an intrinsic part of this TAC, but it can also be purified by alternative protocols as a soluble enzyme activity. This enzyme complex is much smaller than the TAC and requires exogenously added DNA as template for transcription (Pfalz and Pfannschmidt, 2013). The smallest complex reproducibly purified from mustard chloroplasts comprises ten subunits in addition to the Rpo subunits (Steiner et al., 2011) and was formerly assigned as peak A enzyme (Pfannschmidt and Link, 1994). The core complex consisting only of Rpo subunits could be purified as peak B enzyme from etioplasts or young immature chloroplasts (Pfannschmidt and Link, 1994). Complexes with intermediate subunit composition have never been found, strongly suggesting that the large peak A enzyme is only stable when all ten additional subunits are present at the same time. It therefore can be defined as the minimum RNA polymerase complex of chloroplasts, which may be larger by transient addition of further subunits.

The potential functions of these ten subunits were predicted by domain analysis of the corresponding genes in Arabidopsis and verified in some cases by functional characterization of the respective T-DNA inactivation mutants (see below for detailed description). Intriguingly, all these mutants display an albino/ivory or pale-green phenotype with arrested plastid development, no or minor PEP activity, and an enhanced NEP activity. The mutations are seedling lethal on soil, but the albino mutants can be maintained and grown on sucrose-supplemented medium (Schröter et al., 2010). These phenotypic defects are highly reminiscent of those of plants in which PEP activity was inactivated by genetic approaches (plastomic Δrpo mutants), indicating that the presence of each single subunit is as essential for the activity of PEP as the rpo-core subunits themselves. Therefore, it is postulated that proteins that (i) are present in this minimum RNA polymerase complex, and (ii) cause severe defects in chloroplast biogenesis when inactivated (albinos) define a set of true RNA polymerase subunits that have been named PEP-associated proteins (PAPs) (Pfalz and Pfannschmidt, 2013). Not surprisingly these subunits were also found within the group of pTAC proteins (Pfalz et al., 2006). This allows a distinction to be made between true RNA polymerase subunits (PAPs) and more peripheral components among the pTACs that are less directly involved in transcription and for which chloroplast biogenesis phenotypes in inactivation mutants are less pronounced or even wild-type like.

In Arabidopsis, so far 12 different PAPs have been identified that meet the mentioned criteria (Fig. 1). In co-expression analyses, they display a high degree of correlation in their transcript accumulation patterns, suggesting that they probably generate a regulon (Steiner et al., 2011). For PAP1,
it was demonstrated that both RNA and protein are rapidly induced by illumination (Yagi et al., 2012). More detailed in silico expression analyses were performed using available databases at the eFP browser to test this observation for the other pap genes. To this end, the average expression value for the pap gene regulon under all stages of the life cycle (as far as available) was calculated and compared it with the highest expression value found in mature rosette leaves (set to 100%) (Fig. 2, see bar diagram). In general, high pap transcript accumulation was found in green tissues containing chloroplasts (such as carpel, sepals, rosette leaves, and the shoot apical meristem). In contrast, only low expression was observed in non-green tissues such as flower petals and stamen, pollen, roots, and dry seeds (Fig. 2). Interestingly, imbibition of seeds alone led already to significant induction of pap genes. This suggests that pap genes might be induced not only by light but also by internal developmental programmes. In summary, the data indicated that pap genes are mainly expressed in chloroplast-containing tissues and in tissues that are prone to generate chloroplasts (Fig. 2). This coincides with the initial biochemical observations describing a light-induced restructuring of the PEP complex upon etioplast-to-chloroplast transition (Pfannschmidt and Link, 1994). It is thus conceivable that pap gene expression is enhanced by nuclear RNA polymerase II as part of the developmental programmes that serve seedling development and photomorphogenesis. The controlling factors/regulators for this distinct response remain to be determined but most likely are connected to

**Fig. 2.** Expression analysis of the PAP 'regulon' during the life cycle of Arabidopsis. Using the eFP browser (http://bar.utoronto.ca), the developmental series of expression levels was obtained for an average of 11 PAPs (centre plot). Means and standard deviations were calculated and plotted as percentages of the highest score set to 100% (marked by arrow head). The stage of life cycle tested can be identified by the positions numbered on the x-axis. 1, Dry seed (ds); 2, imbibed seed, 24 h (imb); 3, first node (not shown on the life cycle); 4, flower stage 12, stamens (not shown); 5, cauline leaf (cl); 6, cotyledon (c); 7, root (r); 8, entire rosette after transition to flowering (not shown); 9, flower stage 9 (f9); 10, flower stage 10/11 (f10); 11, flower stage 12 (f12); 12, flower stage 15 (f15); 13, flower stage 12, carpels (not shown); 14, flower stage 12, petals (not shown); 15, flower stage 12, sepals (not shown); 16, flower stage 15, carpels (floral diagram); 17, flower stage 15, petals (floral diagram); 18, flower stage 15, sepals (floral diagram); 19, flower stage 15, stamen (floral diagram); 20, flowers stage 15, pedicels (pl); 21, leaf 1 + 2 (not shown); 22, leaf 7, petiole (p); 23, leaf 7, distal half (px); 24, leaf 7, proximal half (px); 25, hypocotyl (h); 26, root (r); 27, rosette leaf 2 (rl2); 28, rosette leaf 4 (rl4); 29, rosette leaf 6 (rl6); 30, rosette leaf 8 (rl8); 31, rosette leaf 10 (rl10); 32, rosette leaf 12 (rl12); 33, senescing leaf (sl); 34, shoot apex, inflorescence (ifm); 35, shoot apex, transition (ft); 36, shoot apex, vegetative (sam); 37, stem, second internode (not shown); 38, mature pollen (mp); 39, seeds stage 3 with siliques (s3); 40, seeds stage 4 with siliques (s4); 41, seeds stage 5 with siliques (s5); 42, seeds stage 6 without siliques (s6); 43, seeds stage 7 without siliques (s7); 44, seeds stage 8 without siliques (s8); 45, seeds stage 9 without siliques (s9); 46, seeds stage 10 without siliques (s10); 47, vegetative rosette (arrowhead, reference set at 100%). Each value was then color coded and attributed to the schematic features on the life cycle. F, fertilization; EG, embryogenesis (inset picture depicts embryos in their photosynthetic phase); G, germination; PMG, photomorphogenesis (green arrow, light-dependent etioplast-to-chloroplast transition); FT, flowering transition; FMG: floral morphogenesis; GG, gametogenesis; FG, female gametophyte. Colours range from yellow (no expression) to red (full expression as described above for the plot); white, not documented. Displayed features were inspired by the eFP browser developmental map.
the processes in the early steps of germination and embryo development.

Although all *Arabidopsis pap* knockout mutants display highly similar phenotypes, the predicted functions and structures of the respective PAPs are quite diverse and bring a multitude of functional as well as structural features to the complex. For an exhaustive summary, all 12 *pap* gene sequences including their predicted functional domains and their corresponding predicted three-dimensional structure around the *E. coli* RNA polymerase core complex have been compiled in Fig. 3 (for more details see the figure legend). Our current working hypothesis is that all the PAPs are required for the formation and/or stability of the PEP complex and that the lack of any of these subunits is detrimental to the whole complex, hence leading to defects broader than the one corresponding solely to the functional loss of any particular subunit. According to structural prediction of functional domains, the proteins can be subdivided into different groups comprising gene expression- or redox-related functions and a third group with unknown or metabolism-related functions. In addition, our predictions identified numerous PAPs that possess a nuclear localization signal (NLS) in addition to the N-terminal chloroplastic transit peptide (*PAP*1, -5, -7, -8, -9, and -12) (Fig. 3). They therefore probably belong to a special group of proteins recently identified that are dually localized to plastids and the nucleus (Krause and Krupinska, 2009). However, only for PAP5/HMR/pTAC12 has a dual localization in the nucleus and plastids been shown so far, both in *Arabidopsis* and in maize (Fig. 1) (Chen et al., 2010; Pfalz et al., 2015). The functionality of the NLS in the other PAPs awaits experimental proof. For the nucleus-localized PAP5/HMR/pTAC12 of *Arabidopsis*, a function in photomorphogenesis has been attributed (Chen et al., 2010). The mutant allele *hmr-5* shows defects in photomorphogenesis corresponding to hypocotyl elongation under red light and an albino phenotype that is typical of *pap* mutants (Chen and Chory, 2011). It could be demonstrated that HMR physically interacts with phytochrome-interacting-factors (PIFs) and mediates the degradation of PIF1 and PIF3 (Galvao et al., 2012), known to be an important step in the transition from skotomorphogenesis to photomorphogenesis. A 9 aa transcriptional activation domain at the C terminus of PAP5/HMR/pTAC12 mediates the expression of PIF target genes (Qiu et al., 2015), indicating that PAP5/HMR/pTAC12 can directly affect gene expression in the nucleus. This coincides with the action of plastidial PAP5/HMR/pTAC12, as shown recently in maize. It could be demonstrated that the orthologous *ZmpTAC12* possesses a binding activity for

---

**Fig. 3.** Predicted protein structure of the PAP proteins in *Arabidopsis* and a structural model of the PEP complex. PEP core: Structural model borrowed from bacterial-type rpo core complex (Ebright lab at http://rutchem.rutgers.edu); PAP1–PAP12: yellow, chloroplastic transit peptide (cTP); red, bipartite or monopartite NLS (bNLS or mNLS). PAP1: magenta, PNLR domain, stabilization of rbcL RNA; dark green, PPR (pentatricopeptide repeat) motif, RNA binding involved in RNA maturation and translation; cyan, SAP domain, a putative DNA binding motif involved in chromosomal rearrangement. PAP2: dark green, PPR domain composed of nine motives; light green, SMR small MutS related involved in mismatch repair. PAP3: cyan, S1-like domain, involved in RNA binding. PAP4 and PAP9: cyan and sky blue, SOD Fe/Mn superoxide dismutase C-terminal and N-terminal domains, respectively. PAP5: gold, glutamine-rich region; green, PEST signature of short-lived molecules (adapted from Chen et al., 2010). PAP6: green, fructokinase with substrate and ATP binding site. PAP7: magenta, SET domain, involved in methyl transferase activity; cyan, Rubisco binding domain sharing similarity with histone binding motives. PAP8: no homology with known functional domain. PAP10: magenta, thioredoxin domain. PAP11: pink and magenta, MurE ligase domain, involved in the incorporation of amino acids into peptidoglycan. PAP12: no homology with known functional domains. Blue lines indicate protein regions used for structural models predicted by Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2).
single-stranded nucleic acids, but without displaying a gene-specific function (Pfalz et al., 2015). Its essentiality for PEP complex formation could also be confirmed in this monocotyledonous plant. Interestingly, ZmpTAC12 exists in two differentially migrating forms that are present in both the nucleus and plastids. This poses questions for the trafficking of these proteins within the cell (see below) and the impact of the chloroplastidial PAP5/HMR/pTAC12 function on its nuclear action; i.e. would a lack of the PEP complex interfere with photomorphogenesis because the greening has become impossible and how is the retrograde signalling impacted by PAP5/HMR/pTAC12 loss of function?

Transcriptional expression of the PAPs has been observed not only in green tissues but also in tissues preceding their greening, e.g. during imbibition of seeds (Fig. 2), whereas the fully assembled PEP complex has only been observed in light-exposed cotyledons or fully developed leaves. It is possible that all potentially dually localized PAPs may hold different functions in plastid transcription and in photomorphogenesis as described for PAP5/HMR/pTAC12. It has been surmised that a subgroup of the PAPs may assemble into a complex in both the nucleus and the plastid and participate in gene regulation, by interacting with two different RNA polymerases. Extensive research will be necessary to elucidate these events in detail. Urgent questions that remain open are those for the targeting of newly expressed PAPs to nucleus and plastid, and determining which localization signal becomes dominant under which condition. Does nuclear localization require differential translational initiation for removal of the chloroplastic transit peptide or does trafficking towards the nucleus occur through a plastidial route in an unknown process, as proposed for Whirly1 (Fig. 1) (Isemer et al., 2012)?

In the PEP complex, one major functional group is comprised of PAPs involved in DNA/RNA metabolism and gene expression regulation, while the second group is related to redox regulation and reactive oxygen species protection (Steiner et al., 2011). The first group includes PAP1, PAP2, PAP3, and PAP7 (but also PAP5 and PAP12, as described recently; Kindgren and Strand, 2015). PAP1 is a protein with a predicted size of 110 kDa and an apparent size of 99 kDa in SDS-PAGE (Yagi et al., 2012). The gene-derived amino acid sequence displays a predicted SAP domain that is known to be involved in DNA/RNA binding (Fig. 3). It also displays additional domains such as a pentatricopeptide repeat (PPR) motif and a domain associated with the stabilization of rbcL transcripts. PPR proteins are known to be involved in RNA metabolism such as processing, splicing, editing, and translation (Delannoy et al., 2007; Schmitz-Linneweber and Small, 2008; Pfalz et al., 2009). It is reported that pap1 mRNA and the corresponding protein accumulate within 1–3 h, respectively, after illumination of dark-grown Arabidopsis seedlings (Yagi et al., 2012). PAP2 is another PPR-containing protein of 89 kDa that displays many structural similarities with Pap1 (Fig. 3) or a small MutS-related mismatch repair domain protein, MRL1, mainly because of the characteristic fold of the PPR domain (Pfalz et al., 2006; Johnson et al., 2010; Steiner et al., 2011). A more detailed analysis of the specific functions of Pap1 or Pap2 regarding their RNA target(s) is, however, still required. PAP3, a 77 kDa protein with an SL-like domain, was found to interact with other subunits of the PEP complex and PAP12 in particular (see below) (Yu et al., 2013). It is predicted to interact with RNAs through its SL-like domain (Fig. 3) in Nicotiana benthamiana (León et al., 2012). PAP7/pTAC14 is a SET domain-containing protein of 49 kDa that may play a role in the transfer of methyl groups on target proteins (Gao et al., 2011). However, so far there is no evidence for this kind of function for PAP7/pTAC14.

Other PAPs may fall into the category of subunits necessary for (redox-dependent or redox-related) gene regulation. PAP6/FLN1 is a protein of 49 kDa and, like its paralogous protein FLN2, displays a phosphofructokinase domain that spans the entire length of the protein (Steiner et al., 2011; Gilkerson et al., 2012). However, a fructokinase activity of the protein has not been confirmed (Arsova et al., 2010) and its true function is still elusive. FLN2, interestingly, was not identified as a PAP and its corresponding knockout mutant displays a less severe phenotype. It thus represents a more transient or peripheral component within the TAC (Gilkerson et al., 2012, Pfalz and Pfannschmidt, 2013). Other protein modifications are possible within the PEP complex by interaction with the thioredoxin domain of the 12 kDa protein PAP10/TrxZ, which has been proven to interact with Pap6 (Arsova et al., 2010; Schröter et al., 2010). It might confer redox regulation of the transcription activity; however, experimental evidence for this is still lacking. Interestingly PAP10 is the first protein for which it was shown that, despite the specific loss of its functional domain, it still rescues the pap10 mutant phenotype with respect to albinism and plastid gene expression (Wimmelbacher and Börnke, 2014). This coincides with the proposed model of PAPs as essential structural components of the PEP complex (see above). Two other paralogous proteins found in the complex are PAP4 and PAP9, both with iron superoxide dismutase domains (Myouga et al., 2008). These 26 and 29 kDa proteins may serve as protection against oxidative stresses generated during the first activities of the photosynthetic apparatus. A partially functional redundancy is observed in the single mutants, as they show pale-green phenotypes rather than the albino phenotype observed with the double mutant. Interestingly, PAP4 displays a highly aberrant apparent molecular weight in SDS-PAGE, suggesting that it might exist as a trimer (Steiner et al., 2011).

Three proteins among the PAPs are less well understood. PAP11 is a MurE domain-containing protein with a theoretical mass of 85 kDa, as concluded from structure prediction and its degree of homology to bacterial proteins able to incorporate amino acids into peptidoglycan (Garcia et al., 2008), but such a function is difficult to reconcile with plastid transcription. PAP8, a 31 kDa protein, is the most enigmatic protein of the complex, as it displays no homologies to any known proteins so far and can only be linked to the formation of the PEP complex by its physical presence within it. No other data about it are available, but sequence analysis of the gene identified a potential NLS, suggesting the possibility of having a second nuclear function. The same is true also for PAP12, a 12.47 kDa protein that was shown to interact with Pap3, Pap5, Pap6, and Pap7 in a yeast-two-hybrid
approach (Yu et al., 2013). Future studies will have to unravel the precise functions for each of these proteins in the PEP complex and whether or not the functions are related to transcription.

**Phage-type DNA-dependent RNA polymerases of plastids (NEPs)**

First indications for the existence of a nuclear-encoded plastid-localized RNA polymerase activity came from studies on plant systems that either do not possess or do not express the plastid rpo genes due to gene loss or ribosome deficiency (Morden et al., 1991; Hess et al., 1993). Since plastids in these plants cannot express the Rpo subunits, they were expected to be defective in plastid transcription. However, numerous plastid transcripts were readily detectable, pointing to a second, probably nuclear-encoded enzyme activity. A first candidate for this activity was purified from spinach as a potential single-subunit RNA polymerase of 110 kDa (Lerbs-Mache, 1993). Further studies using trans-plastomic tobacco lines in which the rpo genes were genetically inactivated confirmed the presence of a nuclear-encoded transcription system in plastids that uses a distinct set of promoters (see below). In order to distinguish between the two activities, the terms nuclear-encoded polymerase (NEP) and plastid-encoded polymerase (PEP) were coined (Hajdukiewicz et al., 1997). Final proof then came from studies in Arabidopsis that identified two nuclear genes for a T3/T7 phage-like single-subunit RNA polymerase being targeted to mitochondria and plastids, respectively (Hedtke et al., 1997). These genes were called rpoT for RNA polymerase of phage T3/T7 type. In Arabidopsis, three different genes exist that encode RNA polymerases targeted to mitochondria (RpoTm) and plastids (RpoTp) and, in addition, according to transit peptide::reporter gene fusion experiments, an enzyme (RpoTmp) that is targeted to both organelles (Fig. 1) (Hedtke et al., 2000; Liere et al., 2011). The gene variants probably emerged from duplication events multiplying the gene for the mitochondrial RNA polymerase, redirecting this RNA polymerase activity also to plastids by acquisition of corresponding transit peptides. However, so far a gene for RpoTmp has not been detected in monocotyledons, suggesting that it is a late evolutionary acquisition in dicotyledons (Liere et al., 2011). All rpoT genes encode phage-type enzymes composed of only a single subunit of around 110 kDa. In vitro experiments with recombinant enzyme activities or overexpression of these proteins indicate faithful promoter recognition, transcription initiation, and elongation, as well as for a specific transcription initiation factor (Liere et al., 2004; Kühn et al., 2007). Specificity factors mediating promoter recognition and transcription initiation as known for yeast or human mitochondria have not been identified yet in plants (Richter et al., 2010). Nevertheless, the NEP enzymes specifically recognize three classes of distinct promoters in vivo designated class Ia, class Ib, and class II (also called type 1a, type 1b, and type 2). Class Ia contains a conserved YRT motif several nucleotides upstream of the transcription start site (Liere and Maliga, 1999). Class Ib contains an additional GAA box around 20 nt upstream of the YRT motif, and class II represents all promoters without the YRT motif (Weihe and Börner, 1999). These distinct types of promoters suggest the existence of protein factors that specifically direct the NEP enzymes to these sites, and at least one protein factor called CDF2 has been reported in spinach that is able to direct RpoTmp to the PC promoter of the rRNA operon (Bligny et al., 2000) (see also below). However, no specificity factor could be unambiguously identified by mass spectrometry or other techniques. A cross-interaction with specificity factors of the PEP enzyme as well as with the PEP complex itself also appears very unlikely, as NEP enzymes have never been identified in any mass spectrometry analysis of PEP complexes. Furthermore, neither NEP nor sigma factors (with the exception of Sig2), or eukaryotic transcription factors with predicted plastidial localization (Wagner and Pfannschmidt, 2006), have ever been identified in nucleoid-enriched fractions or TAC preparations (Pfätz et al., 2006; Majeran et al., 2012; Melonek et al., 2012). This lack of evidence suggests that NEP enzymes as well as transcription regulators are probably present in very low abundance or at very restricted time points outside the harvesting time point of current mass spectrometry studies.

The regulation of NEP activity is still an open question. NEP activity cannot be inferred from the protein level since it has been shown that during plastid maturation the NEP protein level decreases but its activity increases (Cahoon et al., 2004). In addition, plastidial RpoTmp activity differs in different plant species. While in Arabidopsis RpoTmp activity is limited to the seed imbibition stage and declines rapidly after cold release (Courtois et al., 2007), in spinach it can be detected during all developmental stages (Bligny et al., 2000; Azevedo et al., 2008). Some results suggest that the intraplasmidial localization of the enzyme might play a role in its activity in spinach. For instance, immunodetection with specific antisera detected RpoTp and RpoTmp in both stroma and the membrane fraction of spinach chloroplasts (Azevedo et al., 2006). These studies demonstrate an increase in thylakoid membrane association of RpoTmp during leaf development that is mediated by a RING H2-protein named NIP for NEP interacting protein. A regulatory role of this membrane association was suggested because it correlates with a decrease in rDNA transcription (Azevedo et al., 2006, 2008). In contrast, regulation of RpoTmp activity in Arabidopsis should be different. RpoTmp exclusively transcribes the rRNA gene operon at the PC promoter, and this activity is limited to the period of seed imbibition (Courtois et al., 2007).

The plastid-localized NEP polymerases are very active at early steps of seed germination (Demarsy et al., 2006, 2012; see below), and it is at this specific time that they are intensively expressed. In particular, accumulation of RpoTmp transcripts is strongly increased during seed stratification and then declines after root protrusion (Demarsy et al., 2006). Mitochondrial RpoTmp activity was found to be essential for the expression of a subset of mitochondrial genes coding for the subunits of respiratory chain complexes I and IV (Kühn et al., 2009). These expression data were obtained at an early step of plant development (when the first pair of leaves comes out) and at the rosette stage. However, it was not investigated
Plastid RNA polymerases govern chloroplast biogenesis

Whether the mitochondrial RpoTmp enzyme is also active during seed germination, as shown for the plastid equivalent.

Interestingly, these early steps of seed germination are characterized by a strong production of anti-sense RNAs (Demarsy et al., 2012). NEP polymerases are known to exhibit a high readthrough transcription activity (Legen et al., 2002). This might cause the generation of anti-sense RNAs that are complementary to coding regions that are placed on the opposite DNA strand. As suggested for the psbT gene, hybridization of anti-sense RNAs with the corresponding sense RNAs might inhibit translation (Zghidi-Abouzid et al., 2011). In addition, in plants where the expression of RNase J was silenced, anti-sense RNAs were accumulated at high levels (Sharwood et al., 2011). This increased presence of anti-sense RNAs was causing a strong blockage of translation in these silenced plants. Interestingly, in situations where the NEP activity is very high, as for seeds at early steps of germination (Demarsy et al., 2012) and tobacco rpo-knockout plants that are deficient in PEP activity (Legen et al., 2002), transcripts were also not translated. Whether the anti-sense RNAs are just an unspecified by-product of the NEP readthrough transcription activity or whether they possess a distinct role in regulation of translation is not yet understood.

Investigation of chloroplast development in plants

Light-dependent chloroplast development in vascular plants has been investigated extensively as part of photomorphogenesis and represents a prime example for the influence of light on developmental programmes in biological systems. The most common biological test system for chloroplast development used over the years is the light-induced de-etiolation of dark-grown seedlings when shifted from dark to the light (Pogson and Albrecht, 2011). Dark-grown angiosperms such as Arabidopsis and Sinapis typically perform a special developmental programme called skotomorphogenesis exhibiting hypocotyl elongation, formation of an apical hook, and yellow cotyledons that contain a dark-specific plastid type, the etioplast. These plastids lack a thylakoid membrane system but instead possess a so-called prolamellar body, which is comprised of paracrystalline structures of protochlorophyllide, NADPH, and the enzyme protochlorophyll-oxidoreductase. Upon illumination, chlorophyll biosynthesis and thylakoid membrane formation are initiated, and within a few hours a fully developed chloroplast is generated while in parallel the apical hook is opening and hypocotyl elongation is stopped (Solymosi and Schoefs, 2010). In Arabidopsis, numerous mutants with defects at different steps of these developmental programmes have been developed, and essential regulators such as constitutive photomorphogenesis in the dark (cop) (a repressor of photomorphogenesis) or hy5 (the major activator of the light-dependent response) were identified (Casal and Yanovsky, 2005). Extensive research led to an understanding and build-up of the complex regulatory networks behind these light-dependent developmental programmes (Jiao et al., 2007), and chloroplast biogenesis is mainly understood as a consequence of the action of these networks.

The etioplast–chloroplast transition is, however, only one case of chloroplast biogenesis among others. Under natural conditions, chloroplast biogenesis in seedlings occurs more often as a concomitant process of germination without a prolonged dark phase, since the seeds often are not fully covered from the light. Furthermore, while studies on seedling development are well accepted as developmental models, these approaches neglect the chloroplast formation in growing or mature plants. Here, chloroplasts are permanently generated in all meristems of the aerial and illuminated parts of a plant starting from proplastids in the meristematic stem cells. Recent studies of chloroplast formation in the shoot apical meristem of Arabidopsis have revealed very dynamic and tissue-dependent revertible chloroplast development (Charvu et al., 2012). These observations suggest that, in plants beyond the seedling stage, modified or different regulation pathways of chloroplast biogenesis might be active that cannot be fully represented by studies on seedlings. It must be noted that the primary steps of chloroplast biogenesis occur only in meristematic cells, while in mature green cells chloroplasts multiply by division without running through the early developmental programme.

Finally, Arabidopsis (but also many other plants) displays a transient greening stage of the embryo during seed formation within their siliques (Allorent et al., 2013). This includes chloroplast formation within the early embryo, which is then reversed by de-differentiation into so-called eoplasts in the late embryo during the desiccation of the seed. It should be noted that, while in Arabidopsis the maturing seeds turn to a brown colour, other plants maintain their seeds in a green stage until full maturation of seeds, provoking a question about the function of chloroplasts during this stage (Borisjuk et al., 2013). The primary chloroplast biogenesis in Arabidopsis embryos is largely not understood and requires further detailed investigation (see below).

Additional complexity in this context is introduced by the differences in plastid development between monocotyledonous and dicotyledonous plants. Leaf development in monocotyledonous plants depends on the activity of a basal meristem leading to an age-dependent gradient of cells and plastids from the bottom towards to the tip of the leaf (for a recent review, see Pogson et al., 2015). In conclusion, we need to learn much more about the decisive and limiting steps in chloroplast biogenesis of terrestrial plants. One major aspect (and the focus of this review) is the interaction of the various RNA polymerases within the genetic compartments of the plant cell, which are required for the faithful formation of chloroplasts.

Regulation and interplay of plastid RNA polymerases during embryogenesis, germination, and seedling development

The initial model for the interaction of NEP and PEP enzymes has been developed based on research data obtained in
transplastomic tobacco lines in which plastid rpo genes were inactivated by the insertion of an aadA gene cassette (Allison et al., 1996; Hajdukiewicz et al., 1997). Homoplastomic tissues and lines of these Δ rpo-plants are deficient in plastid PEP activity but retain the NEP enzyme activity (see above), allowing a discrimination between PEP- and NEP-dependent transcription units. It was proposed that the NEP enzyme is responsible for the early expression of so-called ‘housekeeping’ genes (including rpo genes and other gene expression-related genes), while the PEP enzyme transcribes genes required for later chloroplast biogenesis such as the genes for the photosynthetic complexes (Hajdukiewicz et al., 1997). In simple terms, the model proposes that the NEP enzyme transcribes the rpo genes on the plastome of proplastids or non-green plastids in meristematic cells. The resulting PEP enzyme then becomes important for the expression of photosynthesis genes and the build-up of the plastid photosynthetic apparatus. This proposed sequence of action of the two types of RNA polymerases is mainly based on their differential use of different types and/or combinations of promoter elements (see above). According to this usage, plastome genes were subdivided into three classes that possess promoters only for PEP (class I) (photosystem genes), for PEP and NEP (class II) (most other genes), or only for NEP (class III) (the rpo operon, accD and ycf2) (see Fig. 1). This model was a major breakthrough as it could explain the complex process of chloroplast biogenesis as a simple sequential action of different RNA polymerases activating different gene groups depending on the developmental stage.

Data and observations raised in the last 15 years expanded our knowledge about the action of PEP and NEP transcription systems. There is growing evidence that both polymerases are active at the same time in all green or non-green tissues but with different degrees of activity. Systematic comparison of plastid transcription in green and white tissues of Δ rpoA tobacco plants and in primary leaves in the barley mutant albostrians revealed that PEP is the predominant transcriptional activity in green tissues, generating most of the mRNAs, tRNAs, and rRNAs (Legen et al., 2002; Zhelyazkova et al., 2012). The NEP enzymes display a minor activity in green tissues but are enhanced in white tissues of the mutants where they transcribe even photosynthesis genes from remote promoters, although to a low degree (Courtois et al., 2007; Zhelyazkova et al., 2012). An enhanced NEP activity in white plastids of PEP-deficient mutants appears to be characteristic for this type of mutants, as it has also been reported in many other albino mutants and especially in mutants with pap gene expression defects (Pfalz and Pfannschmidt, 2013). Apparently, PEP deficiency causes a condition in plastids that leads to either a compensatory or accidental enhancement of NEP activity, although this cannot rescue the chloroplast biogenesis programme. A recent study performed in maize proposes a predominant role of PEP in the generation of tRNAs supporting active translation in chloroplasts (Williams-Carrier et al., 2014). This coincides with the low levels of photosynthesis proteins found in white plastids. At the same time, this negatively affects the chlorophyll biosynthesis pathway, since the glutamyl-tRNA trnE represents the precursor of aminolevulinic acid, the entrance substrate of the tetrapyrrole biosynthesis pathway (Rüdiger and Grimm, 2006; Tanaka and Tanaka, 2007). These and other new data (e.g. about the requirement of PAP subunits) allow refining of the initial sequential model of NEP and PEP activity and adapting it to the timely and spatial dynamics that occur during tissue development in the life cycle of a plant.

**Action of RNA polymerases during embryogenesis**

The plant life cycle of a diploid dicotyledonous angiosperm such as *Arabidopsis* starts with the fertilization of the egg cell, resulting in a zygote. This zygote develops via distinct gene expression programmes and developmental steps (octant, globular stage, heart stage, and torpedo stage) into a fully developed plant embryo comprising two cotyledons, a shoot axis and a primary root embedded into a seed shell (Le et al., 2010). In *Arabidopsis* but also in many other plants, this involves a transient photosynthetic stage in which the complete embryo turns green and generates fully developed chloroplasts. The function of these chloroplasts is not fully understood yet, but repression of this photosynthetic phase by electron transport inhibitors (e.g. DCMU) negatively affects the germination ability of the resulting mature seeds (Allorent et al., 2015). The development of the embryo, still connected to the mother plant, might be influenced by hormones from the mother plant at this stage, but may also include independent processes that are controlled by the embryo itself. A clear cut separation of these influences is, however, difficult at this stage. Based on the observations given above, it must be assumed that all steps of the RNA polymerase-driven expression cascade for the full development of a chloroplast become active in the embryo. This includes the activation of NEP and PEP enzymes, as well as the expression of interacting sigma factors and PAPs. Indeed, expression and action of both NEP and PEP activities were identified in this particular stage (Allorent et al., 2013). Likewise activation of pap gene expression could be identified in publicly available expression data sets during the embryonic stage of *Arabidopsis* (Kremnev and Strand, 2014) (Fig. 2). As already mentioned above, *Arabidopsis* seeds lose their chloroplasts during maturation and turn into brown seeds with chlorophyll-less embryos. This requires an active block of chloroplast function followed by degradation of chloroplast membrane structures, protein components, and pigments within the embryo. Seed maturation, desiccation, and dormancy are largely determined by hormonal regulation via abscisic acid from the mother plant. A recent study reports that abscisic acid induces the expression of RSH2 and RSH3, two plastid-localized enzymes capable of synthesizing guanosine tetraphosphate (ppGpp), an alarmone that induces the stringent response in bacteria (Yamburenko et al., 2015). This ancient signalling pathway of probable endosymbiotic origin has been retained in plants; however, no real physiological function has been attributed to plant ppGpp yet (van der Biezen et al., 2000; Takahashi et al., 2004). As for bacterial RNA polymerases, the molecule was shown to inhibit transcriptional activities of purified PEP preparations *in vitro* (Sato, 2009). It is thus conceivable...
that abscisic acid shuts down PEP activity in chloroplasts of embryos during seed maturation by enhancing ppGpp production. This would lead to a halt in gene expression and could initiate subsequent disassembly of the photosynthetic apparatus. It has been shown recently that the plastids in the desiccated seeds do not fully return to the proplastid stage but retain the transcriptional apparatus at both the protein and transcript level in order to allow an immediate regain of transcription during germination and subsequent seed development (Allorent et al., 2013).

**Action of RNA polymerases during germination and plant development**

Upon imbibition and germination of the seed, a second phase of chloroplast development can be observed within the now-growing embryo. This process is, however, restricted to the tissue of the cotyledons, which is in contrast to the embryo stage during seed formation where all tissues become green. Apparently during seedling development, chloroplast formation in the hypocotyl and root is either activated or not promoted. This would require either expression of an inhibiting factor or the repression of an activating factor. A recent study suggests that chloroplast formation in the roots is actively repressed by hormone signals from the shoot (Kobayashi et al., 2012). This observation also demonstrates that chloroplast formation is a developmental subprogramme that can be switched on and off and does not simply represent a concomitant process of illumination. In the cotyledons, it has been shown that the same RNA polymerase expression cascade runs as before in the seed embryo, leading to a quick generation of chloroplasts in the leaf blade tissues (Demarsy et al., 2006). NEP transcripts both for RpoTp and RpoTmp are already present in the dry seed (Demarsy et al., 2006; Allorent et al., 2013). Upon imbibition and germination, NEP transcripts and activities increase and peak at 2–3 d after germination. They trigger a rapid expression of plastid rpo genes, increasing subsequently PEP activity. This coincides with a high pap gene expression detected after illumination of the seed (Fig. 2). Whether NEP and PAP expression is controlled by the same regulators in the nucleus is not known yet. PEP activity stored in the eoplast after de-differentiation of embryo chloroplasts appears to be important for efficient germination, as inhibition of this activity by tagetin treatment delays significantly germination (Demarsy et al., 2006). However, these starter molecules are apparently not sufficient to induce full chloroplast development and require supplementation of activity by generation of novel PEP complexes. This process of complementation probably becomes interrupted when germination occurs in the dark. In mustard, this leads to accumulation of the PEP core enzyme, which is rapidly reconstructed by the addition of PAPs upon illumination (Pfannschmidt and Link, 1994). The precise sequence of events that occur during this light-triggered reconstruction of PEP is largely not understood and will require careful dissection in future studies. Existing data suggest that it may include regulation at the levels of transcription, translation, and even post-translational mechanisms (Steiner et al., 2011; Yagi et al., 2012). The strong albino phenotype of pap knockout mutants indicates that the process of PEP reconstruction is an essential bottleneck in the generation of chloroplasts. Interruption of any single step in it should lead to similar phenotypes. A detailed characterization of all pap mutants as well as additional genetic screens for regulatory factors will be necessary to fully understand this process. The high speed at which chloroplasts are generated in dark-grown seedlings upon illumination suggests that the time window for the course of events is probably in the hour to minute range. In the meristems of aerial parts of adult plants, such as the shoot apical meristem, it should be expected that the processes run even faster, since no artificial pausing due to darkening does occur. A study analysing the presence of thylakoid membranes in plastids of the different cell layers in the shoot apex of Arabidopsis could identify fairly developed thylakoids in the L1 and L3 layers, while in the L2 layer true proplastids were identified (Charuvi et al., 2012). This suggests that the differential action of NEP and PEP enzyme must be under a tight cellular control. The described secondary loss of chloroplasts in epidermal cells of leaves (Charuvi et al., 2012) could be explained by a targeted repressive effect on PEP activity e.g. by ppGpp, followed by a specific degradation of plastids similar to the processes described for the late embryonic phase. How these processes are performed and regulated is largely unexplored.

The events described above are based mainly on observations in dicotyledonous plants. In principle, it should be expected that similar processes also occur in monocotyledonous plants, but adapted to the specific developmental gradient of the plastids caused by the action of the basal meristem (Mullet, 1993). Indeed, many similarities have been reported. In maize, lack of ZmpTAC2, ZmpTAC10, and ZmpTAC12 (the orthologues of Arabidopsis PAP2, PAP3, and PAP5) each leads to PEP deficiency exactly as described for Arabidopsis (Pfalz et al., 2015). Analysis of the barley mutant albostrians indicates enhanced NEP activities in plastids of tissues with PEP deficiency (Zhelyazkova et al., 2012). Nevertheless, there also exist differences from dicotyledonous plants, for instance the existence of two variants of ZmpTAC12 (Pfalz et al., 2015). Whether this represents a species-specific difference or a difference between the plant clades needs to be elucidated in the future.

**Concluding remarks**

The transcription of plastidial genes is essential for all plastid types regardless of whether they perform photosynthesis or not. Proper plastid development therefore involves in all cases the action of the plastid RNA polymerases. It should be noted that, in green tissues with mature chloroplasts, the organelles do multiply by fission rather than by biogenesis from proplastids. The role of plastid RNA polymerases here is different from in meristems, embryos, or cotyledons, and is mainly restricted to the maintenance of transcript levels to allow sufficient protein production in the mature chloroplast. Future studies focusing on plastid development will therefore...
need to address a complete understanding of functioning and regulation of plastid RNA polymerases in both a spatial and time-resolved manner at the cellular and tissue levels. This will also include the search for essential (master?) regulators controlling the expression of the nuclear-encoded components of the plastid transcription machineries (NEPs, Sigl–6, and PAPs) and putative signalling molecules for retrograde signals from the plastids. Identification of such regulators and signalling molecules may help us to understand the coupling (and its limitations) between plastid development and developmental programmes that control plant morphology and provides interesting targets for potential biotechnological applications. A detailed understanding of the action of plastid RNA polymerases will be essential for future ‘green’ biotechnology and bioenergy applications. This includes functional and structural (crystallographic) analyses of the complex. Necessary experimental tools are readily available, and corresponding working models await experimental proof promising fundamental new discoveries in this research field.

Acknowledgements

Work in Team 7 ‘Nucleo-plastidic interaction: Chloroplast biogenesis and redox control’ at the Laboratoire de Physiologie Cellulaire & Végétale (LPCV) was funded by the DFG to TP (PI323-4 and PI323-5), by Université Grenoble-Alpes (AGIR2015) to RB and the Labex GRAL (Grenoble Alliance for Integrated & Structural Biology) to LPCV. Copyright for colored scanning electron microscopy images of the Arabidopsis flower used in Fig. 2: J. Berger and D. Weigel, Max-Planck-Institute for Developmental Biology, Tübingen, Germany.

References


Legen J, Kemp S, Krause K, Profanter B, Herrmann RG, Maier RM. 2002. Comparative analysis of plastid transcription profiles of entire plastid...
chromosomes from tobacco attributed to wild-type and PEP-deficient transcription machineries. The Plant Journal 31, 171–188.


Yamurenko MV, Zubo YO, Börner T. 2015. Abscisic acid affects transcription of chloroplast genes via protein phosphatase 2C-dependent activation of nuclear genes: repression by guanosine-3’-5’-bisphosphate and activation by sigma factor 5. The Plant Journal 82, 1030–1041.


