UNIVERSITÀ DEGLI STUDI
DI MILANO

-

Ph.D. school in

Molecular and Cellular Biology

XXIX Cycle

Molecular bases of SVP regulatory functions in *Arabidopsis thaliana*

Roberto Ferrari

Ph.D. Thesis

Scientific tutor:

Prof. **Paolo Pesaresi**

Academic year: 2016-2017

TABLE OF CONTENTS:

PART I: SVP PROTEIN PARTNERS

MATERIALS AND METHODS

ABSTRACT

Flowering time regulation has a strong impact on plant life cycle, since it allows plants to flower and to reproduce under environmental permissive conditions. Several genes are involved in the regulatory pathways that determine the floral transition step, i.e. the switch from the plant vegetative phase to the reproductive phase and the consequent flower formation and fruit set. Among those genes, *SHORT VEGETATIVE PHASE* (*SVP*), a MADS box transcription factor, acts as strong repressor of the so called florigen promoting genes, *FLOWERING LOCUS T* (*FT*) and *SUPPRESSOR OF OVEREXPRESSION OF CO 1* (*SOC1*). Moreover, *SVP* has been also reported to act as a repressor of flower homeotic gene expression, thus ensuring the correct maintenance of floral meristem identity. Due to the relevance of SVP in both such important plant developmental stages, during my Ph.D. research program I tried to elucidate the molecular mechanisms at the basis of SVP activities. That has been done through different and complementary strategies that had the dual aim to identify SVP protein partners and to move the first steps towards the comprehension of the role of chloroplasts and chloroplast-nucleus signaling pathways in SVP functions.

Co-immunoprecipitation assays followed by Mass Spectrometry analyses have allowed to draw up a list of Arabidopsis putative robust SVP interactors involved, at different levels, in chromatin organization and histone modification. Interestingly, the detailed characterization of the major Arabidopsis trimethyltransferase enzyme, SET DOMAIN GROUP 2 (SDG2), has revealed the existence of an SVP-SDG2 containing protein complex able to regulate the expression of *SVP* gene at the vegetative and reproductive meristems, by affecting the H3K4 methylation pattern within the first exon of *SVP*.

Furthermore, our interests on the role of chloroplast-nucleus communication and its possible interactions with the flowering time regulation, have been met through the detailed characterization of two chloroplast-located PENTATRICO-PEPTIDE-REPEAT (PPR) containing proteins, which share three main features: i) they are part of the chloroplast gene expression machinery, ii) they are involved in chloroplast-nucleus communication, iii) they have been reported to be target genes of SVP by ChiP-seq assays.

The detailed characterization of the Arabidopsis PPR proteins, GENOME UNCOUPLED 1 (GUN1) and CHLOROPLAST RNA PROCESSING 1 (AtCRP1), has provided the first preliminary insights into how chloroplast-nucleus signaling mechanisms

may enable higher plants to more effectively adapt to the ever-changing internal and external conditions and mitigate detrimental effects to fitness.

INTRODUCTION

Vegetative and reproductive phase in *Arabidopsis thaliana*

Model organisms have been and still are an essential molecular biology tool. Through the study of those organisms it has been possible to decipher the molecular pathways behind all the biological processes so far known.

In plant molecular biology, the model species *Arabidopsis thaliana* is one of the most powerful living tools that can be used, due to its completely sequenced and annotated genome and the availability of a complete set of mutants for almost all the genes of its genome (NASC: *http://arabidopsis.info/* and ABRC *https://abrc.osu.edu/*). *Arabidopsis thaliana* is an annual dicotyledonous plant adapted to live in temperate climates. The life cycle of any annual plant can be divided into two main phases: the vegetative and the reproductive phase.

During the vegetative phase, which starts with seed germination, plants produce leaves and roots, two organs extremely important for providing sugar and other chemicals, respectively, thus to ensure the collection of all the nutrients and energies needed to support the reproductive phase. In *A. thaliana* the vegetative phase consists of a rosette structure formed by the radial disposition of leaves, normally about 13-15 leaves in wild type (WT) plants grown under long-day (LD) conditions (Dorca-Fornell *et al.*, 2011; Valentim *et al.*, 2015). Since all the Arabidopsis organs have a post embryonic formation, it means that all the leaves have to be formed *ex novo* from a pool of undifferentiated cells placed at the center of the rosette structure: the shoot apical meristem (SAM). SAM is formed by a pool of stem cells capable to be maintained undifferentiated by the activity of *WUSHEL* (*WUS*) and *SHOOT MERISTEMLESS* (*STM*) homeobox transcription factors that together with the cytokinins (CKs) and their protein receptors, such as CUP-SHAPED COTYLEDON (CUC), repress the expression of the *CLAVATA* (*CLV*) class genes responsible of the daughter stem cell differentiation into leaf organs, driven by the gibberellin (GAs) and auxin signals (Laux *et al.*, 1996; Leibfried *et al.*, 2005; Aguilar-Martínez *et al.*, 2015).

During the reproductive phase the plant produces sexual organs, the flowers, which will allow the organism to form the new generation of plants through ovule fertilization and seed formation. Flowers in *A. thaliana* are organized on a vertical structure called “stem”, that

arises from the previous SAM and becomes an inflorescence meristem (IM), another indeterminate meristem capable to maintain itself in an undifferentiated state thanks to the *TERMINAL FLOWERS 1* (*TFL1*) (Shannon and Meeks-Wagner, 1991) gene. Successively, the expression of *LEAFY* (*LFY*) (Schultz and Haughn, 1991), *APETALA1* (*AP1*), and *CAULIFLOWER* (*CAL*) (Kempin, Savidge and Yanofsky, 1995; Gregis *et al.*, 2008) leads to the differentiation of cells in the peripheral zone of the IM into a determinate fate meristem called floral meristem (FM), which will form flowers. The development of the floral meristem into flowers is finely regulated by homeotic class genes as nicely described by the "ABC model" (Bowman, Smyth and Meyerowitz, 1991; Jack, 2004), in which the action of A class genes (*AP1, APETALA 2*), B (*APETALA 3, PISTILLATA*), C (*AGAMOUS*) alone or combined are responsible for the correct formation of four concentric flower whorls: sepals, petals, stamens and carpels, respectively.

Like it occurs for any other annual plant, also for Arabidopsis, the switching time from the vegetative to the reproductive phase is a critical step that deeply involves the organization of meristems, also at the molecular level. This irreversible step, so called "floral transition", needs to be perfectly coordinated with all the internal and external stimuli that plants can face during their development to guarantee a correct seed set and the perpetuating of the species.

Flowering: a matter of stimuli

Flowering in Arabidopsis is promoted by a set of few genes, called floral pathway integrators: *FLOWERING LOCUS T* (*FT*), *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*) and *LFY* (Li *et al.*, 2008).

FT is the master of the florigen genes in *Arabidopsis thaliana*, its expression is activated by *CONSTANS* (*CO*) in the phloem companion cells during LD period (Tiwari *et al.*, 2010), and from there it moves to the SAM via endoplasmic reticulum membranes, where it can interact with the bZIP transcription factor FD and form the complex FT-FD, which requires a 14-3-3 protein to be stabilized. This protein complex is needed for *SOC1* activation, thus to commit the floral transition (Li *et al.*, 2008, 2016; Pin and Nilsson, 2012). *FT* shares a similar function with its paralog *TWIN SISTER OF FT* (*TSF*), even if *tsf* mutant shows only

a mild late flowering phenotype compared to *ft* mutant in Arabidopsis (Jang, Torti and Coupland, 2009; Andrés and Coupland, 2012).

SOC1 is expressed both in the young leaves and SAM, and it is actively regulated by *CONSTANS* (*CO*) and *AGAMOUS LIKE 24* (*AGL24*) under LD conditions (Hepworth *et al.*, 2002; Lee *et al.*, 2008; Liu *et al.*, 2008). *SOC1* can regulate its own expression through a positive feedback loop and further activate the expression of the FM identity gene, *LFY* (Lee *et al.*, 2008; Dorca-Fornell *et al.*, 2011; Tao *et al.*, 2012). Afterwards, both FT-FD protein complex and LFY actively regulate the expression of *AP1* in the SAM, another important FM identity gene (Jack, 2004; Li *et al.*, 2008; Tao *et al.*, 2012; Valentim *et al.*, 2015), thus to ensure the irreversible commitment of the shoot meristem to a reproductive meristem.

Flowering is the most critical step in a plant life cycle; therefore, it is extremely important to promote it only in presence of optimal environmental conditions. To do so, plants need to perceive the different signals coming from external and internal cues to best adapt to them. *A. thaliana* has different pathways involved in the introgression of such stimuli and thus to control the floral transition (Dorca-Fornell *et al.*, 2011). Each of those pathways culminates to the activation of the common integrator genes at the core of the florigen pathway, described above (Wang *et al.*, 2014). In this sense, chloroplasts have been recently discovered to play a fundamental role in the introgression of environmental stimuli that influence floral transition (Feng *et al.*, 2016). For instance, in Arabidopsis a chloroplast-derived signal has been reported to be critical for high light-regulated flowering mediated by the *FLOWERING LOCUS C* (*FLC*) (see vernalization pathway section). Feng et al. (2016) have demonstrated, indeed, that PTM, a PHD transcription factor involved in chloroplast retrograde signaling (Sun *et al.*, 2011), perceives such a signal and mediates transcriptional repression of *FLC* through recruitment of FVE, a component of the histone deacetylase complex. Thus, the data indicate that chloroplasts function as a main environmental sensor in the plant cell, capable to perceive light stimuli from outside and, through chloroplast-to-nucleus communication, also known as the retrograde signaling pathway (Chan *et al.*, 2016), those signals are translated into changes of nuclear gene expression, with the aim to modulate the cell metabolism to best face the environmental changes (Colombo *et al.*, 2016). Through very similar mechanisms the nucleus can communicate with chloroplasts, known as anterograde signaling pathway, and adapt their activities to the cell needs (Pogson, Ganguly and Albrecht-Borth, 2015). In the past decade,

substantial progress has been made in elucidating chloroplast-nucleus communication, with the identification of several components involved at different levels in the retrograde and anterograde signaling pathways. Among them, the PENTATRICO PEPTIDE REPEAT (PPR) containing proteins play important roles both in chloroplast-to-nucleus and nucleus-to-chloroplast communication (for further details see PART II of this thesis) (Susek, Ausubel and Chory, 1993; Koussevitzky *et al.*, 2007; Colombo *et al.*, 2016).

i. Exogenous stimuli:

Arabidopsis has four different pathways responsible to perceive the external stimuli. Those are the vernalization pathway, the photoperiod pathway, the thermosensory pathway, and the stress flowering pathway.

1. Vernalization pathway

Vernalization consists of a long period of exposure to low temperatures, thus to avoid flowering during the unfavorable winter season.

FLOWERING LOCUS C (*FLC*) is one of the master regulators, since it directly represses the expression of florigen genes (Hepworth *et al.*, 2002; Yun *et al.*, 2012). *FLC* activity is mainly regulated in a transcriptional manner, through epigenetic changes driven by FRIGIDA (FRI) and VERNALIZATION (VRN) proteins.

FRI binds to *FLC* sequence and, together with chromatin remodeling factors, can activate *FLC* expression through epigenetic positive markers such as trimethylation of lysine 4 histone 3 (H3K4me3) (Choi *et al.*, 2011; Jiang *et al.*, 2011; Yun *et al.*, 2012). *FLC* expression is subordinated to the presence of *FRI* active gene, *Arabidopsis thaliana* lacking *fri*, like Columbia (Col-0) or Landsberg erecta (Ler-0) ecotypes, can flower just few weeks after germination when grown in LD photoperiod. Moreover, during vernalization, *FLC* is transcriptionally regulated by the expression of long non coding RNAs (lncRNAs), COOLAIR and COLDAIR (Swiezewski *et al.*, 2009; Heo and Sung, 2011).

VRN proteins, instead, belong to the family of polycomb-group (PcG) proteins, a class of proteins present also in the animal kingdom, responsible of chromatin remodeling status

and accessibility (Orlando, 2016). Among those, VRN2 (known also as POLYCOM REPRESS COMPLEX 2 (PRC2)) can directly target the first intron of *FLC* and repress it through a trimethylation of H3K27. The expression of *VNR2* is promoted by the long exposure to cold temperature occurring during the vernalization (Gendall *et al.*, 2001; Sheldon *et al.*, 2009). Besides VNR2, also VNR1 and VNR3 have an active role in this process by stabilizing the repression of *FLC* (Bond *et al.*, 2009) after the vernalization, even when temperatures increase again in the spring period. This maintenance repressive role has been also demonstrated for *VERNALIZATION INSENSITIVE 3* (*VIN3*), which codifies for a PHD finger protein induced during the vernalization (Bond *et al.*, 2009).

FLC is highly expressed in leaves and SAM tissues where it can play its role by directly repressing *FT* and *SOC1*, respectively, by forming a protein complex together with SHORT VEGETATIVE PHASE (SVP), another strong flowering repressor gene (Searle *et al.*, 2006; Andrés and Coupland, 2012). Nevertheless, the presence of an *FLC*-independent vernalization pathway in Arabidopsis has also been shown. *FLC* paralogs, such as *FLOWERING LOCUS M* (*FLM*) and *MADS AFFECTING FLOWERING* (*MAF*) family genes, act as repressor of floral transition in a *FLC* similar way (Ratcliffe *et al.*, 2001; Scortecci, Michaels and Amasino, 2003).

2. Photoperiod pathway

In Arabidopsis, both light quality and day length can strongly affect the floral transition. In a wild type situation, plants will flower faster in LD than in short-day (SD) photoperiod. The light-dependent activation of *GIGANTEA* (*GI*) and *Flavin binding, Kelch repeat, F-box protein 1* (*FKF1*) is responsible of *CO* expression, a zinc finger transcription factor that acts as the master regulator of photoperiod pathway (Yanovsky and Kay, 2002; Valverde *et al.*, 2004). In particular GI-FKF1 protein complex allows *CO* expression by promoting the degradation of *CO* repressors CYCLING DOF FACTORs (CDFs) (Fornara, de Montaigu and Coupland, 2010). Arabidopsis has different classes of photoreceptor proteins involved in the light sensory detection, each class is specialized in the absorbance of a particular type of light (Briggs and Christie, 2002; Huq and Quail, 2002; Demarsy and Fankhauser, 2009; Rizzini *et al.*, 2011). Since *GI* and *FKF1* are involved in the circadian clock gene regulation,

CO expression is also modulated during the day time, reaching its expressional peak in the afternoon, if plants are grown under LD conditions (Sawa *et al.*, 2007; Baudry *et al.*, 2010).

CO regulation is also affected in a post-translational manner through PHOTOMORPHOGENIC 1 (COP1), an ubiquitin ligase enzyme, which degrades CO protein in the dark and by Phytochrome B (PhB) and HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 1 (HOS1) in the morning. On the contrary, the stability of CO protein is ensured by the binding with FKF1 protein and Phytochrome A (PhA), that can enhance CO stability under daylight (Valverde et al. 2004; Liu et al. 2011; Lazaro et al. 2015).

Once CO has been accumulated in the leaves, it can activate *FT* expression and start then the floral transition. CO is a transcription factor unable to bind to the DNA, it needs to form a protein complex with a nuclear factor Y (NY-F) for the recognition of the specific promoter element present at the 5' end of *FT* promoter region (Kumimoto *et al.*, 2010).

3. Thermosensory pathway

Besides vernalization, Arabidopsis flowering is dramatically influenced by relatively fast temperature changes. Plants need, thus, to perceive short term stimuli using a different pathway than the one involved in the epigenetic modulation of *FLC*. The thermosensory pathway is mainly an *FLC*-independent mechanism that introgresses, via *SVP-FLM* regulation, the temperature stimuli (Lee *et al.*, 2007; Li *et al.*, 2016). *FLOWERING LOCUS M* (*FLM*) and *SVP* are regulated via transcriptional-splicing and post-translationally, respectively. In presence of low temperature (13-17°C), *FLM* undergoes alternative splicing to generate the FLM-β protein isoform. FLM-β interacts with SVP, and together repress flowering through the binding to *FT*, *TSF* and *SOC1* sequences. On the opposite, high temperatures (over 25°C) enable the splicing of *FLM-γ* isoform, an FLM-β competitor for the interaction with SVP, which lacks of the DNA binding domain, therefore the new formed protein complex FLM-γ-SVP is not able anymore to repress the florigen expression. Moreover, high temperature drastically affects the stability of SVP, causing a fast degradation of the protein (Lee *et al.*, 2013; Hwan Lee *et al.*, 2014). SVP-mediated thermosensory regulation can also indirectly repress *FT* expression by activating the MYB transcription factor, EARLY FLOWERING MYB PROTEIN (EFM), which together with

JMJ30 H3K36 Me2 demethylase can modify chromatin status nearby the florigen sequence (Yan *et al.*, 2014).

Beside this *SVP-FML* dependent mechanism, Blázquez et al. 2003 have been able to show the convergence of the autonomous flowering pathway with the thermosensory pathway by affecting the *FT* expression through an alternative *SVP* mechanism.

4. Stress flowering induction

Different stress conditions can drastically affect the floral transition in Arabidopsis, generating a stress escape response. So far, the discussion is still open in literature if it is the case to consider this escape response either a real flowering pathway or not, because any stress stimulus, at the end, converges into the photoperiod and autonomous pathway (Riboni *et al.*, 2014). It is anyway clear that under unfavorable living conditions, plants switch to the reproductive phase, with the aim to set seeds as fast as possible in order to ensure the survival of the species.

Among all the stress escape responses, the drought escape response (DER) is one of the best studied mechanisms in Arabidopsis. It is a *GI*-dependent response, in which *GI* activation under LD condition degrades the *FT* repressor CYCLIN DOF FACTOR (CDF) causing the early floral transition. *FT*, together with *SOC1*, are also activated in LD by the stress hormone abscisic acid (ABA) signal, generated during the DER. Interestingly, ABA shows an opposite effect when DER occurs in SD conditions (Riboni *et al.*, 2013). Beside drought, high salt concentration can also delay flowering.

The salinity escape response (SER) has been shown to involve different flowering regulator genes by employing different molecular mechanisms. One of them converges in the autonomous pathway involving a DELLA protein, a negative regulator of GA and ethylene hormone signaling (see autonomous pathway section) (Achard *et al.*, 2006). SER also delays flowering by repressing *CO* and *FT* expression, and by promoting the expression of BROTHER OF FT AND TFL1 (BFT) protein, which competes with FT for the FT-FD interaction in the SAM, and leads to a late floral transition (Kim, Kim and Park, 2007; Li *et al.*, 2007; Ryu *et al.*, 2014). Although in the photoperiod pathway *GI* and *CO* show a positive relation, in SER they don't. More in details, *gi* mutant has been shown to have an enhanced salt tolerance, whereas *GI* overexpression causes saline sensitivity, pointing to a negative

role of *GI* in the stress tolerance mechanism. *GI*, in fact, is a competitor for the kinase SALT OVERLY SENSITIVE (SOS) protein complex formation, which is responsible for the salt tolerance in Arabidopsis (Kim *et al.*, 2013).

Flowering can also be accelerated through the lack of certain nutrients. For instance, low nitrate levels can trigger flowering, whereas nitric oxide accumulation represses flowering by altering the photoperiod and autonomous pathways (He *et al.*, 2004; Liu *et al.*, 2013). Moreover, Ohto et al. (2001) described the response of Arabidopsis to sucrose concentration, that can modulate the time of flowering through the transcription regulation of *LFY* in the SAM.

ii. Endogenous stimuli

Not only environmental cues can affect floral transition in *Arabidopsis thaliana*, but flowering also requires to perceive internal stimuli. There are other three flowering pathways that are involved in the integration of endogenous stimuli in *A. thaliana*: the autonomous pathway, the GA pathway and the age pathway.

1. Autonomous pathway

The autonomous pathway acts in parallel with the vernalization pathway and culminates with the repression of *FLC*. As a matter of fact, the autonomous pathway mutants, showing a late flowering phenotype, can be fully rescued by the vernalization treatment. The first evidence of the existence of this pathway came from Koornneef et al. (1991), that reported different Arabidopsis mutant loci responsible of a late flowering phenotype regardless of the photoperiod length. *FCA*, *MULTICOPY SUPPRESSOR OF IRA1 4* (*MSI4*), *FY* and *FPA* where thus defined as components of the new autonomous floral pathway. Nowadays, in addition to these four loci, other three more genes have been identified as component of the same pathway: *FLOWERING LATE KH MOTIF* (*FLK*), *FLOWERING LOCUS D* (*FLD*), and *LUMINIDEPENDENS* (*LD*) (Lee *et al.*, 1994; Sanda and Amasino, 1996; Lim *et al.*, 2004). Not many advances have been made in the understanding of this floral pathway, probably due to the fact that any of these autonomous genes seems to act independently from

the others by regulating directly the expression of *FLC*. *FY* encodes a polyadenylation factor, whereas *FCA, FPA* and *FLK* encode RNA-binding proteins, although it is still not clear whether they repress either *FLC* splicing or maturation, by directly interacting with the *FLC* mRNA. Something more is known on the mechanism through which FCA and FY interact with each other and repress *FLC.* In fact, the FCA-FY protein complex is responsible of the polyadenylation of the third intron of *FCA* causing an alternative spliced and truncated form of *FCA* mRNA, which translates an inactive form of the protein. Marquardt et al. (2006) speculate that a similar polyadenylation mechanism is responsible for the *FLC* mRNA abundance decrease. The *LD* gene has been described as a coding sequence for an homeodomain factor with an RNA-binding region of unknown function (Lee *et al.*, 1994). The Epigenetic *FLC* autonomous repression is also regulated by the two deacethylase enzymes FLD and FVE (Simpson, 2004).

2. Gibberellin acid (GA) pathway

Under short-day conditions, *FT* expression and the relative floral transition are strongly dependent from GA pathway, although a direct GA regulation of *FT* has not been demonstrated, yet (Hisamatsu and King, 2008). On the opposite, the mechanism by which GAs can promote the expression of both *SOC1* and *LYF* in the SAM, through the modulation of the two GA-responsive transcription factors, GAMYB and DELLA, is well known (Mutasa-Göttgens and Hedden, 2009). In particular, DELLA proteins, which normally act as transcription factor repressors, are degraded in presence of GA. The GA nuclear receptor GID1, in presence of GA, can bind DELLA proteins causing their conformational changes and allowing their binding to the ubiquitinase proteins for degradation (Murase *et al.*, 2008).

Among all the GA variants present in nature (about 136), only a small number has biological activity, and only the GA_4 has been shown to be active in Arabidopsis. Since GA_4 covers many roles in different development pathways, its abundance has to be finely regulated. To do so, Arabidopsis has a complex system which involves GA biosynthesis, turnover and signal transduction, tightly regulated by environmental and developmental cues (Yamaguchi, 2008). *2-oxoglutarate- dependent dioxygenases*, *GA 20-oxidase* (*GA20ox*) and *GA 3-oxidase* (*GA3ox*) are the three major genes involved in the GA biosynthetic pathway, whereas the turnover is mainly regulated by *GA 2-oxidase* (*GA2ox*), which inactivates GA

molecule by hydroxylation (Mutasa-Göttgens and Hedden, 2009; Fornara, de Montaigu and Coupland, 2010). Interestingly, GA abundance increases in the SAM before floral transition, but not the corresponding GA biosynthetic enzymes. These data indicate that GAs are not directly synthesized in the SAM but somewhere else, and then translocated into the meristem. In line with that, *GA20ox* expression is mainly localized in leaves where it is regulated by far-red light through the PHYTOCROME B (Hisamatsu *et al.*, 2005; Eriksson *et al.*, 2006; Hisamatsu and King, 2008).

Although the GA pathway mainly promotes flowering under SD, it has been found that mutants lacking GA pathway regulatory genes, such as *ga1-3* and *gid1*, can also affect flowering under LD or continuous light growth conditions, suggesting a contribution of gibberellins also in the photoperiod pathway, when *CO* is active (Wilson, Heckman and Somerville, 1992; Griffiths *et al.*, 2006).

3. Age pathway

The Age pathway also promotes flowering under short-day conditions, and beside GA pathway, it is not influenced by external cues. The *SQUAMOSA PROMOTER BINDING LIKEs* (*SPLs*) transcription factor family is the master regulator of this endogenous pathway. In particular the family is divided into two subgroups based on the protein size, termed *SPL3* and *SPL9*, which are coding for transcription factors mostly active in either the SAM or leaves, respectively (Wang, 2014). In the SAM, *SPLs* genes actively promote the expression of *AP1*, *LFY*, *SOC1* and its redundant gene *FRUITFULL* (*FUL*), whereas they promote flowering by repressing the florigen repressor genes *AP2- like* transcription factors, through an miRNA (miR172) mechanism in the leaves (Aukerman and Sakai, 2003; Wang, Czech and Weigel, 2009; Wu *et al.*, 2009).

SPL3 activation in the SAM is strongly regulated through the abundance of miR156, a miRNA which binds to 3' UTR region of SPL3 subgroup mRNA, causing transcript cleavage and translation inhibition (Gandikota *et al.*, 2007). miR156 abundance is temporally regulated in plants, it means that is higher in seedlings and decreases by time, allowing flowering of old plants also in SD conditions (Wang, Czech and Weigel, 2009). miR156 abundance has been demonstrated to be affected also by sugar, the product of

photosynthesis, in particular high levels of sugar downregulate miR156 expression at both transcriptional and post-transcriptional level (Yang *et al.*, 2013).

Remarkably, all the flowering pathways, so far described, are part of a complex network of molecular mechanisms aimed to integrate any environmental and endogenous signal that can ensure plants to pursue flowering under any permissive condition (Figure 1).

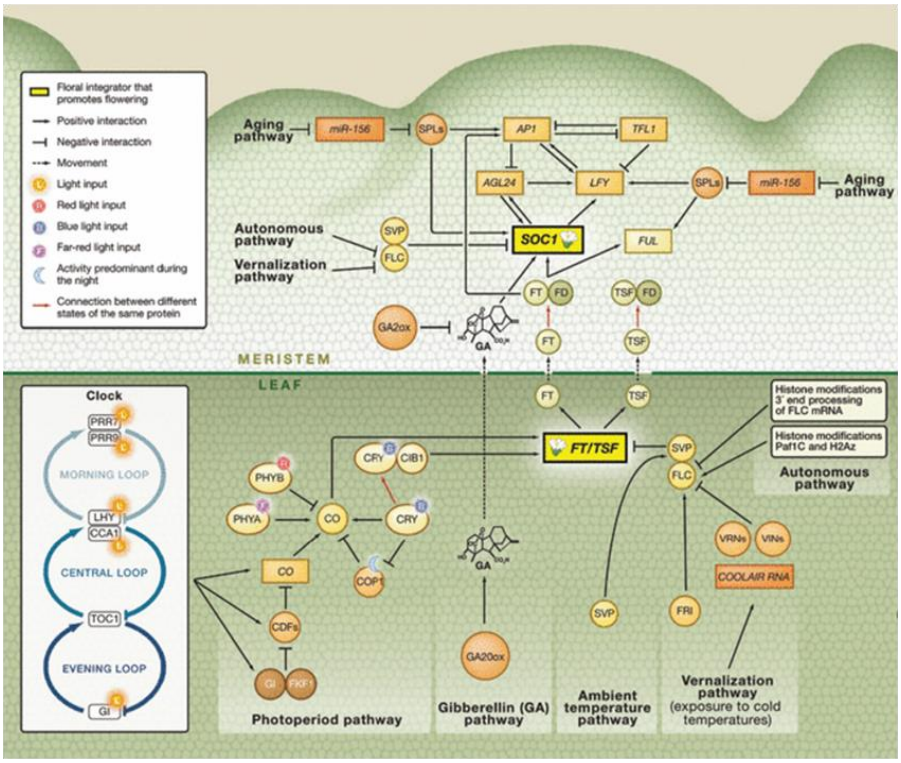

Figure 1: A scheme of the flowering pathway regulatory networks interacting together in leaf and meristem. Adapted from Fornara et al. 2010.

MADS box transcription factors, at the basis of *Arabidopsis thaliana* developmental control

The MADS box transcription factor family counts 107 gene members in the *Arabidopsis thaliana* genome, interestingly only few of them have been studied so far, revealing several important roles in any Arabidopsis developmental step (Parenicová *et al.*, 2003). The acronym MADS derives from the initial of four important transcription factors *MINICHROMOSOME MAINTENANCE 1* (*MCM1*), *AGAMOUS, DEFICIENS* and *SERUM RESPONSE FACTOR* (*SRF*), from *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Antirrhinum majus* and *Homo sapiens*, respectively (Norman *et al.*, 1988; Passmore *et al.*, 1988; Schwarz-Sommer *et al.*, 1990; Sommer *et al.*, 1990; Yanofsky *et al.*, 1990). Plant MADS box family is divided into two main groups: MADS box type I (SRF-like) and type II (MEF2-like), based on sequence conservation (Masiero *et al.*, 2011).

Arabidopsis MADS box type I group is composed of 60 members further divided into Mα, Mβ and Mγ clades. All of them are short single exon coding genes located on the I and V chromosome. Not much is known about their functions, which have been nicely reviewed by Masiero et al. (2011), that assigned to them a role during the reproductive development, in particular during female gametophyte and seed development (Köhler *et al.*, 2003; Colombo *et al.*, 2008; Steffen *et al.*, 2008).

MADS box type II group, instead, is formed by Mδ and MICK-like clades. Those are located all over the five Arabidopsis chromosomes and appear longer in size than type I, with several introns and exons. MICK-like transcription factors are named on the four highly conserved protein domains: MADS domain (M), Intervening domain (I), coiled-coil keratin-like domain (K) and the C-terminal domain (C). The MADS domain is the highly conserved N-terminus domain, which recognizes specific DNA sequences called CArG boxes [$CC(A/T)_6GG$], allowing the protein to bind to its target genes (Riechmann, Wang and Meyerowitz, 1996). I and K domains are involved in protein dimerization, whereas the C domain is the less conserved domain and it is involved in specific MADS box protein multimeric complex formation and transcriptional activation (Davies *et al.*, 1996; Fan *et al.*, 1997; Honma and Goto, 2001; Parenicová *et al.*, 2003; van Dijk *et al.*, 2010). Notably, type

I and II do not share any sequence similarity with each other, except for the MADS domain (De Bodt *et al.*, 2003).

Unlike type I factors, type II MADS box genes are much more studied also at functional level. They are involved in all the Arabidopsis developmental processes, from the embryo to the fruit development, and sometimes they display activities in more than just one developmental step (Figure 2). *AGL15*, for instance, has been demonstrated to be involved in embryogenesis, as well as a repressor of floral transition (Smaczniak et al. 2012a; Adamczyk et al. 2007; Harding et al. 2003). In addition, *SVP* has been shown to have an important role both in floral transition, where it acts as a strong repressor, and in floral development, as an inflorescence meristem identity gene (Hartmann *et al.*, 2000; Gregis *et al.*, 2006, 2008, 2009). In Smaczniak et al. (2012a), all the MADS box type II members whose functions are known, 31 in total, are listed. Some of them have been described previously, in this thesis, as flowering pathway regulators (*FLC*, *SOC1*, *AP1*,*CAL*, *FUL*, *AGL24* and *SVP*), as well as the homeotic genes involved in the ABC model of flower formation (*AP3*, *PI* and *AG* with exception for *AP2*), together with the homeotic class D gene, *SEEDSTICK* (*STK*) and *SHATTERPROOF* (*SHP*) *1-2*, involved in specific carpel and ovule formation (Yanofsky *et al.*, 2000; Favaro *et al.*, 2003). Furthermore, like type I MADS box factors, the type II are involved in gametophyte development, but specifically for the male gametophyte formation (pollen), as it was reported for *AGL65-66-104* (Verelst, Saedler and Münster, 2007). Moreover, others MADS factors have been also reported to be involved in root formation (*AGL17-21*), leaf formation (*AGL16*) and fruit development and senescence (*FUL* and *SHP 1-2*, respectively).

In summary, all the information available on MADS box genes, lead to a straight and clear conclusion: MADS box transcription factors are a class of genes at the basis of *Arabidopsis thaliana* developmental control.

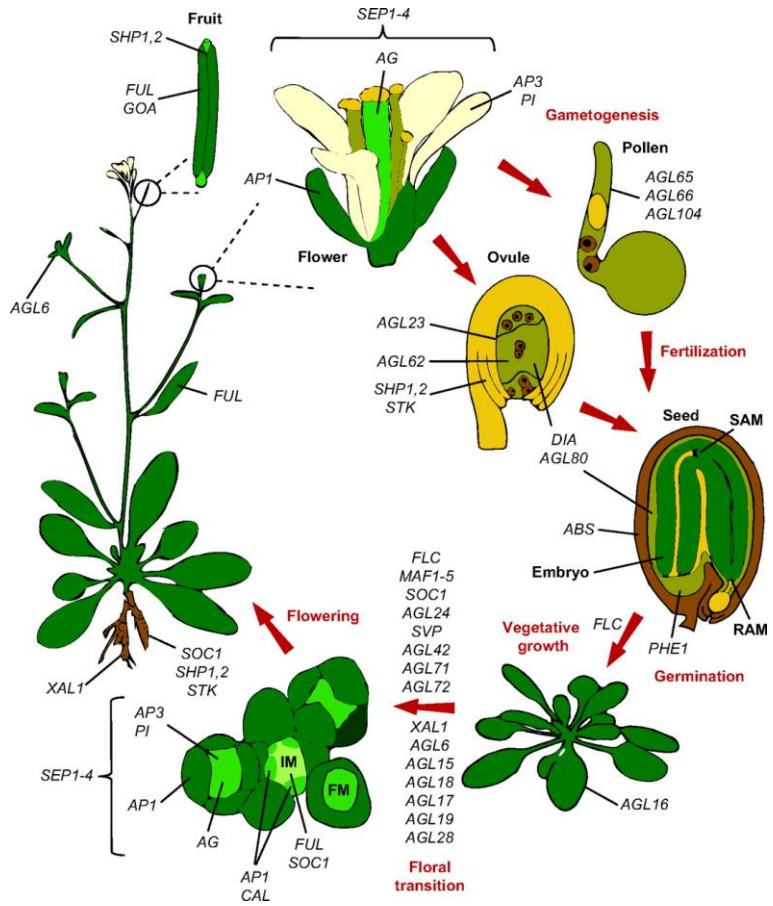

Figure 2: Functions of MADS box genes throughout the life cycle of *Arabidopsis thaliana*. For each developmental step, the involved MADS box transcription factors are indicated, starting from the seedling until the seed formation. (Smaczniak et al. 2012a).

MADS box protein complexes

Transcription factors (TF) form usually homodimers or heterodimers with other TFs, thus increasing the selectivity of the DNA binding sites, and creating a large number of DNA binding complexes based on a small number of TFs (de Folter *et al.*, 2005). Also MADS box TFs need to form multimeric protein complexes to exert their role (Schwarz-Sommer *et al.*, 1992). The mechanism of protein complex formation among the MICK MADS box clade has been largely studied, whereas very little is known about the other MADS box clades. MICK TFs preferably interact with MICK proteins, even if there are rare exceptions, such

as the interaction of MICK TFs with Mα type I MADS box factors (Espinosa-Soto et al. 2014).

In particular, the MICK's K-domain was demonstrated being essential for the multimeric protein complex formation. The K-domain is supposed to fold into three alpha-helixes, where the first two (K1 and K2) are responsible of the dimer specification, and the third one (K3) together with the C-domain, is responsible for the higher order protein complex formation (Figure 3) (Yang, Fanning and Jack, 2003; Yang and Jack, 2004). In fact, MICK TFs have been reported to form mainly tetramer structures, by combining two different protein homodimers or heterodimers in the same complex, enhancing the specificity and the stability of the DNA binding. Based on this model, four MADS box TFs forming a tetramer can bind to two different CArG boxes in close proximity on the same regulatory region of a target gene, forming a DNA loop necessary for its further stabilization (Figure 3) (West and Sharrocks, 1999).

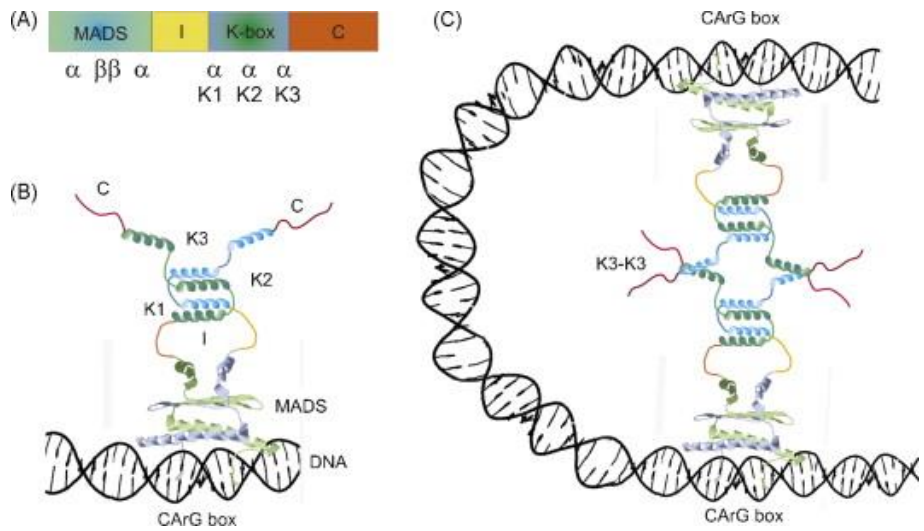

Figure 3: MICK MADS box protein structure. (A) MICK protein domain composition, with regards to M, I, K and C domains. The presence of α-helixes or β-sheet in the secondary structure is indicated as α and β, respectively. (B) MADS box TF dimer binds a target DNA sequence. (C) A tetramer MADS box TF complex binds to two CArG box sequences present on the same regulatory region of a target gene, causing a DNA looping to further stabilize the DNA-protein interaction. (Immink, Kaufmann and Angenent, 2010).

Among all the MICK protein interactions, protein complexes involving homeotic floral genes are the best studied ones, thus they represent a good model to explain the mechanism

of MADS box TF tetramer structure formation. As already discussed above, floral organs develop from floral meristem thanks to the action, alone or combined, of specific genes grouped into four different classes: A, B, C and D. AP1, a gene of class A, is responsible for sepals formation when it acts alone, however it is responsible of petal formation when it acts together with B class genes (*AP3* and *PI*). *AP3* and *PI* can also specify pistil formation when also the C class gene, *AG*, is actively transcribed, whereas single *AG* activity regulates the formation of carpels in flowers. Ovule formation is then due to the activity of the D class genes, *STK* and *SHP*. Interestingly, studies of this "quartet" model revealed the importance of dimer formation as a compulsory mechanism for homeotic gene functionality. In fact, it was also demonstrated *in vitro*, that MADS box homeotic TFs cannot bind to CArG box sequences as monomers, but they need the presence of other homeotic TFs to play their role (Schwarz-Sommer *et al.*, 1992). Moreover, in 2009 Immink et al. revealed the existence of a E class gene *SEPALLATA* (*SEP*) family, that has a "glue" protein role, necessary for homeotic ABCD protein dimerization. Immink's data, combined with all the high-throughput yeast two- and three-hybrid data collected during the last few years, definitively helped to define the homeotic protein-protein interaction network responsible of floral organ formation in Arabidopsis. Even if MADS box TFs belonging to the same homeotic class can share a certain grade of functional redundancy, it was possible to define the protein interactors involved in each tetramer complex (Figure 4): one AP1 and one SEP4 homodimer for sepals, AP1-SEP and PI-AP3 tetramer for petals, the combination PI-AP3 and AG-SEP for stamens, one AG and one SEP homodimer for carpels, and finally an AG-SHP and STK-SEP quartet for ovule identity specification (Immink, Kaufmann and Angenent, 2010; Theißen, Melzer and Rümpler, 2016).

Beside this homeotic regulatory mechanism, MADS box protein complex formation has been demonstrated to be also involved in MADS box TF nuclear transportation. Indeed, it was shown that AP3 and PI heterodimerization is a necessary prerequisite for their import into the nucleus (McGonigle, Bouhidel and Irish, 1996).

MADS protein-protein dimerization shown in the ABCD(E) model, it is only an example of the MADS interaction mechanisms. Further examples of protein complex formation can be also observed in other MICK MADS box dependent processes, such as floral transition and FM maintenance identity. For instance, FUL-SVP and FUL-SOC1 heterodimers have regulatory roles during floral transition, as well as the FLC-SVP

heterodimer (Li *et al.*, 2008; Balanzà, Martínez-Fernández and Ferrándiz, 2014). Furthermore, other MADS box heterodimers were found to have key roles in FM identity, such as AP1-AGL24 and AP1-SVP complexes (Gregis *et al.*, 2006).

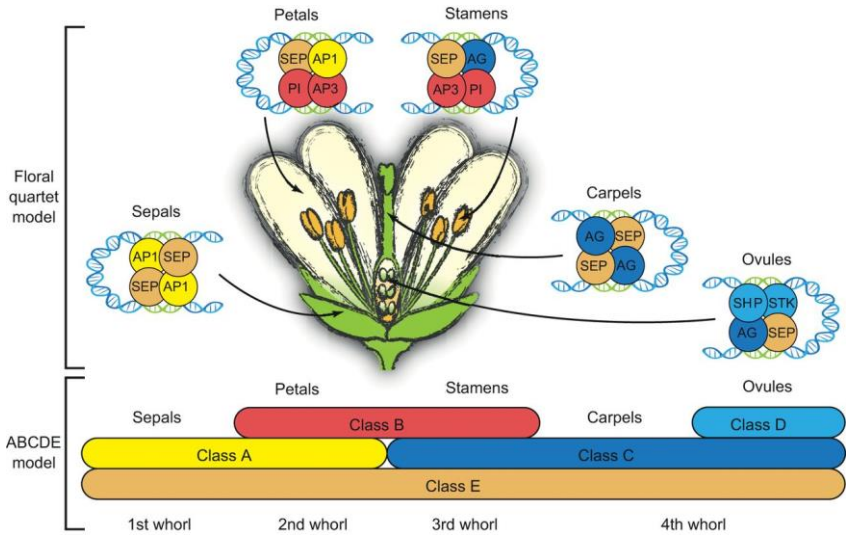

Figure 4: Schematic representation of ABCD(E) model of floral organ specification, with particular regard to homeotic gene tetramer formation. (Theißen, Melzer and Rümpler, 2016).

Interestingly, in some cases MADS box factors were co-immunoprecipitated with other proteins rather than MADS box TFs, indicating that MADS box dependent processes require interactions with a plethora of proteins with different functions (Acevedo *et al.*, 2004; Gregis *et al.*, 2006; Helliwell *et al.*, 2006; Cseke *et al.*, 2007).

i. Chromatin remodeling enzyme interactors

Beside MADS protein-protein interaction, MADS box factors were also found to form protein complexes with other classes of proteins, such as other TFs belonging to homeobox clade, as well as chromatin and nucleosome remodeling factors (Smaczniak et al. 2012b; Helliwell et al. 2006; Messenguy & Dubois 2003).

It is known from literature that homeotic MADS box heterodimers are able to cause changes in chromatin accessibility around their target genes, by modifying the chromatin

landscape in a flexible mode according to the plant developmental needs. However, it was not clear if this mechanism was due to either a direct activity of MADS protein complex on the histone conformation, or to the presence of other chromatin remodeling factors in the high order MADS box protein complexes (Pajoro *et al.*, 2014). The answer to this question came from different immunoprecipitation assays and mass spectrometry analyses performed during the last decade on the MADS box protein complexes, involving the homeotic proteins AP1, AP3, AG and SEP3 (Immink et al. 2009; Smaczniak et al. 2012b; Hill et al. 2008). These analyses revealed the presence of different classes of chromatin remodeling factors interacting with MADS box proteins. Among those epigenetic factors, some of them were highly represented in most of the IPs, such as the RELATIVE OF EARLY FLOWERING 6 (REF6) protein (Lu *et al.*, 2011), an histone H3K27 demethylase, as well as the CHROMATIN REMODELING (CHR) class, in particular CHR4-7 and 11. From the same IPs, nucleosome remodeling factors, such as PIKLE (PKL) and ATINO 80 (INO80), specific for just one or few homeotic MADS box protein complexes, were co-immunoprecipitated with SEP3 and AG/AP1, respectively (Smaczniak et al. 2012b). Moreover, PKL has been found to be a common protein partner also capable of interacting with another transcription factor belonging to the bZIP family, such as HY5, involved in hypocotyl cell elongation (Jing *et al.*, 2013), thus pointing out a common mechanism of interaction between TFs and chromatin remodeling factors, rather than being a MADS box specific mechanism.

Interestingly, not only the MADS box proteins involved in the floral ABCD(E) model were found to interact directly with chromatin remodeling factors. AGL15, a MADS box factor active in embryos, was immunoprecipitated together with two histone deacetylase, SIN3 and HDA19 (Hill, Wang and Sharyn E Perry, 2008), pointing to the existence of an epigenetic role of MADS box protein complexes not only with regards to flower development, but probably as a general mechanism of MADS box protein activities.

SVP **an important MADS box gene**

Arabidopsis thaliana At2g22540 gene locus encodes the SHORT VEGETATIVE PHASE (SVP), a MADS box MICK TF known to be involved in two developmental critical

steps. *SVP* gene is organized in 9 short exons and 8 introns (Figure 5), and encodes a nuclear 240 aminoacid protein of about 26,9 kDa. Hartmann et al. (2000) firstly described the early flowering phenotype of *svp* homozygous plants, defining *SVP* as a potent flower repressor gene. Its expression was detected either in the vegetative tissues (leaves, roots and SAM), where it displayed two alternative spliced mRNA forms, as well as in the floral meristem, before flower organs arise, where Hartmann found a single mRNA spliced form corresponding to the whole *SVP* cDNA length. The alternative *SVP* splicing form (1,3 kb) has been demonstrated, several years later, having a different dimerization behavior compared to the 1,7 kb longer version. In fact, only the longer mRNA form is able to interact with all the *SVP*'s interactor partners so far characterized (Severing *et al.*, 2012).

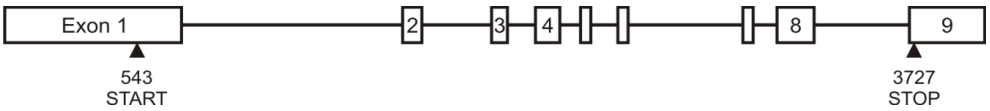

Figure 5: Scheme of At2g22540 (*SVP*) genomic locus. Exons and introns are represented as open blocks and black lines, respectively. Positions of the start and stop codons are indicated by arrow heads.

Furthermore, the role of *SVP* in flowering control has been well studied, showing the involvement of this MADS box factor in several floral transition pathways present in Arabidopsis. In particular, SVP was defined as a strong interactor of FLC involved in the repressive control of *FT* and *SOC1* floral integrators under inductive condition (LD), making then *SVP* an essential part of the vernalization pathway (Figure 6) (Scortecci, Michaels and Amasino, 2003; Jack, 2004; Li *et al.*, 2008; Fornara, de Montaigu and Coupland, 2010). Interestingly, a reciprocal positive auto-regulation loop of *SVP* and *FLC* was also reported, thanks to a Chromatin ImmunoPrecipitation-sequencing (ChIP-seq) experiment, through which it was possible to find putative CArG box target sequences on *SVP* and *FLC* genomic regions (Gregis *et al.*, 2013). *SVP* was also found to be a pivotal gene in the integration of the thermosensory pathway. As already discussed above (see *Exogenous stimuli* section), environmental temperatures can affect SVP protein stability, thereby promoting or inhibiting flowering (Lee *et al.*, 2007, 2013; Hwan Lee *et al.*, 2014; Galvão *et al.*, 2015). Finally, in 2014 Andrés et al. discovered a further *SVP* repressing role in the GA pathway, as repressor of gibberellin synthesis. These data are also consistent with the SVP-CArG box sequence

found, through ChIP-seq analysis, on the promoter region of *STIMPY* (*STIP*), a GA pathway component (Gregis *et al.*, 2013).

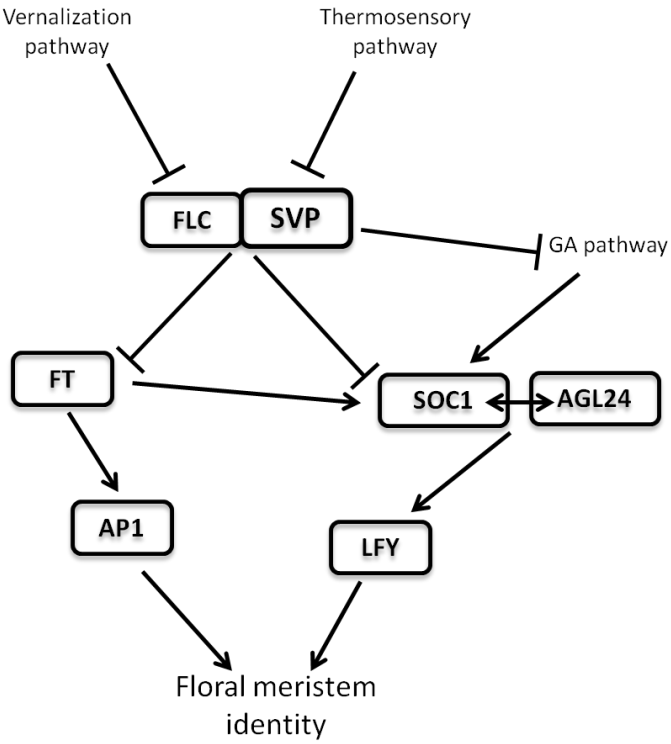

Figure 6: *SVP* floral pathway regulation scheme. *SVP* is involved as a repressor both in the vernalization and thermosensory pathways upstream the *FT* and *SOC1* floral integrator genes, as well as a repressor of GA pathway by affecting the GA synthesis.

SVP was firstly named as *AGL22* due to the high aminoacid sequence homology with another component of the *AGAMOUS LIKE* family, *AGL24*. Beside this homology, *SVP* and *AGL24* have an opposite role during the Arabidopsis vegetative phase, the first one acts as a strong floral repressor, whereas *AGL24* was reported being a floral meristem identity gene, that promotes flowering together with *SOC1*, downstream *FT* gene (Tao *et al.*, 2012; Li *et al.*, 2016). Notably, during the vegetative phase, *SVP* exhibits an epistatic role on *AGL24*. Despite this, *SVP* and *AGL24* share the same floral meristem identity role. In particular, the role of *SVP* as repressor regulator, together with *AGL24*, *AP1* and *SOC1*, of class B and C homeotic genes during the early stage of flower development, aimed to maintain the

undifferentiated identity of floral meristem, was nicely reported by Gregis et al. in 2006, 2008 and 2009. Although, the single homozygous *svp* mutant is not able to affect flower development, due to the redundant role of AGL24 in the repressive complex, constitutive *SVP* overexpression, in addition to delaying flowering in both LD and SD growth conditions, is sufficient to display a drastic phenotypic effect on flower development, giving rise to aberrant sepals and petals structures, resembling a leaf structure organ, as well as, occasionally, causing the formation of new floral meristems in the external whorls, which lead to the formation of new aberrant flowers instead of petals organs. Interestingly, the same flower phenotype has been also reported in plants overexpressing *AGL24* (Gregis *et al.*, 2009).

The SVP floral regulator mechanism proposed by Gregis et al. (2006), consists of a repressor tetramer complex made up of LEUNING (LEU) and SEUSS (SEU), AP1 and redundantly SVP or AGL24, that regulates *AG* and *AP3* expression during the early stages of flower formation (stage 1-2) (see Figure 7). Later in the development, the repressor complex is unfolded, thus allowing homeotic gene expression and correct flower development. AP1-SEP3 tetramer complex represses *SVP* and *AGL24* expression during this later stage. More recently, Liu et al. (2009) implemented the homeotic gene repressor complex described above, with the identification of two chromatin remodeling factor interactors, TERMINAL FLOWER 2-LIKE/LIKE HETEROCHORMATIN PROTEIN 1 (TFL2/LHP1) and SAP18, member of the SIN3 histone deacethylase complex, that were reported to interact specifically with SVP and AGL24, respectively (Figure 7).

Because of the *SVP* dual role both in vegetative and reproductive phase, during the last 4 years, efforts have been undertaken to identify new *SVP*'s target genes using a ChIP-seq approach (Tao *et al.*, 2012; Gregis *et al.*, 2013). *SVP* was found to be the direct regulator of many genes involved in widely different pathways. SVP was found, for instance, to be the regulator of genes involved in the vernalization pathway (*VRN2*), in the photoperiod pathway (*GI*, *PRR7* and *PHYA*) and the autonomous pathway (*FLK* and *FLD*), in addition to the already known target genes *FT*, *TSF* and *SOC1*, as well in the hormonal signaling pathways such as auxin (*BIG*), jasmonate (*COI1*) and gibberellins (*STIP*) receptor genes. SVP was also reported as a regulator of meristem identity gene both in the SAM, regulating PHB, CLV, KAN1 and KAN2 expression, and also in the FM, with respect to *PIN1*, *WUS* and *ARF3*. Interestingly, SVP was also found to bind its own genomic locus, either during

vegetative phase, binding within its first exon and the first intron, and also in the floral meristem, where it was found to bind additionally to 2 kb upstream its 5'UTR. Furthermore, SVP has been reported to bind the regulatory regions of nuclear genes encoding chloroplast located PPR proteins, pointing to a major role of chloroplast-nuclear communication in the regulation of flowering time and inflorescence meristem identity.

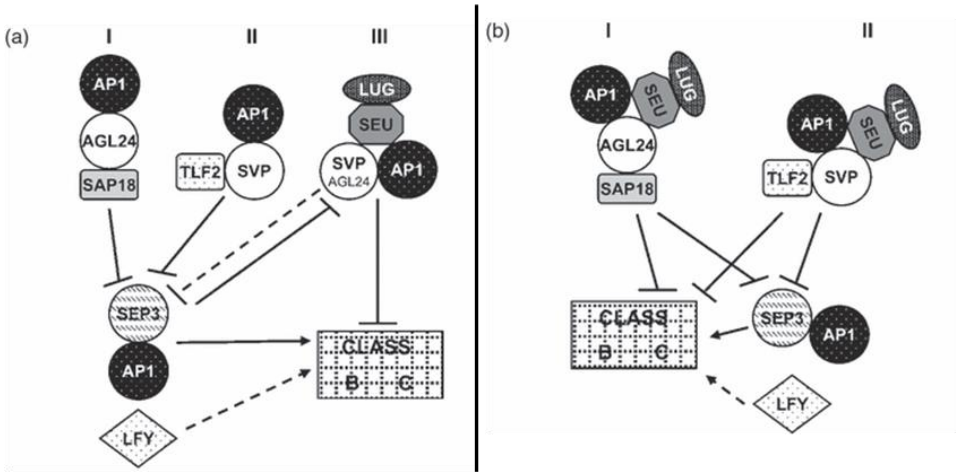

Figure 7: Model of SVP repressor complex acting in flowers. (a) During the stage 1 and 2 of flower development, AP1, SEU, LEU together with AGL24 or SVP (III) form a protein complex responsible of floral meristem maintenance by repressing homeotic genes of class B and C. In parallel I and II AGL24-SAP18 and SVP-TFL2 heterodimers indirectly repress homeotic class B and C genes through *SEP3*. Later on, starting from stage 3, SEP3-AP1 heterodimer represses the expression of *SVP* and *AGL24*, whereas on the other side, it activates the expression of *AG*, *PI* and *AP3*, together with the protein LFY. (b) Alternatively, homeotic genes expression is repressed through the formation of a higher order protein complex recruiting all the repressor proteins in the same complex (Gregis *et al.*, 2009).

SVP and the chloroplast-nucleus communication

An important role for chloroplasts in the introgression of external light stimuli and in the determination of flowering time via the *FLC*-dependent pathway has been discovered, recently (for more details refer to PART II of this thesis). In particular, Feng et al. (2016) have proposed an active role of chloroplasts in light stimuli perception during long-day (LD) conditions, capable to repress *FLC* expression by removing active histone acetylation and trimethylation markers nearby its genomic locus. Those recent findings reveal the

importance of the chloroplast-nucleus communication for the transition from the vegetative to the reproductive stage, beside signaling the physiological and developmental needs of plastids (Koussevitzky *et al.*, 2007).

In this context, Pentatrico-Peptide-Repeat (PPR) containing proteins are a class of nuclear-encoded and chloroplast-imported proteins, which are mainly involved in chloroplast-nucleus communication (Barkan and Small, 2014; Tadini *et al.*, 2016). Interestingly, SVP ChIP-seq analysis performed by Gregis *et al.* (2013) discovered the presence of putative SVP-CArG box sequences within the genomic loci of different nuclear *PPR* genes, revealing a possible involvement of SVP in the expression regulation of these genes. From another perspective, it is possible to speculate a putative involvement of SVP in the chloroplast-nucleus communication by tuning the accumulation of chloroplast-located PPR proteins.

Among the *PPRs* genes found in the SVP ChIP-seq analysis, we focused our attention on GENOME UNCOUPLED 1 (GUN1), mainly involved in the chloroplast-nucleus signaling pathway, as well as a key component of chloroplast stromal-membrane-associated gene expressional machinery (Colombo *et al.*, 2016), and on CHLOROPLAST RNA PROCESSING1 (CRP1), firstly described in *Zea mays*, and reported to be part, together with GUN1, of megadalton protein complexes associated to the chloroplast nucleoids (Olinares, Ponnala and van Wijk, 2010).

Ph.D. PROJECT'S GOALS

The aim of this Ph.D. thesis was to decipher the molecular mechanisms through which SVP plays its dual role, both at the level of vegetative tissues in the SAM, as one of the major flowering repressor genes, and at the level of IM, where SVP acts as a meristem identity gene to ensure a correct flower development. At the beginning of this Ph.D. project, it was largely known that MADS box transcription factors play their roles by forming large protein complexes (i.e. floral homeotic gene quaternary complexes), however biochemical studies aimed to identified the SVP protein interactors were not reported, yet.

Within this frame, one of the main objective of the Ph.D. project was the identification of SVP protein partners both in the SAM and in the inflorescence meristem of *Arabidopsis thaliana* plants.

Part I of this thesis is dedicated to the description of the different strategies employed to identify the SVP interactors. Furthermore, the detailed functional characterization of one of the SVP interactors, SET DOMAIN GROUP 2 (SDG2) that acts as a histone trimethyltransferase enzymes, is reported.

As stated in the Introduction section, chloroplasts also appear to work as main sensors of environmental conditions that determine the flowering time in higher plants. This notion is supported further by the finding that several SVP target genes, identified by ChIP-seq analyses (Gregis et al. 2013), encode chloroplast-located proteins.

A second objective of this project was the detailed functional characterization of two SVP target genes, encoding chloroplast located Pentatrico Peptide Repeat proteins, known to play a major role in chloroplast-nucleus communication.

Part II of this thesis describes, indeed, the functional characterization of the nuclear gene *CHLOROPLAST RNA PROCESSING 1* (*AtCRP1*), encoding a PPR protein whose expression is probably regulated by SVP and that has been identified, inside the chloroplast, as part of a large protein complex containing, among the other polypeptides, the GENOME UNCOUPLED 1 (GUN1) protein, also a member of the PPR protein family found to be a putative SVP gene target. The GUN1 protein, has been studied into details in the lab where this Ph.D. project has been carried out and it is considered a master regulator within the plastid-to-nucleus retrograde signaling pathway.

PART I:

SVP PROTEIN PARTNERS

MATERIALS AND METHODS

Plants material and growth conditions

Arabidopsis thaliana wild type and mutant plants used in this work were grown under controlled growth chamber conditions with 16 h light / 8 h dark cycle at 22°C on soil, except for those employed for the ChIP analysis, which were grown under 8 h light / 16 h dark cycle at the same temperature condition on soil. Plants used for experiments on the vegetative phase, were sampled within fourteen days after sowing, whereas inflorescence material, was collected five days after the bolting stage.

All the mutant plants were in Columbia background. Mutant plants are listed in RESULTS section, except for the *svp* mutant, which corresponds to the deletion mutant *svp-41* (Hartmann *et al.*, 2000). Mutants genotyping was analyzed through PCR using the primers listed in the Table 1.

Target gene		Sense primer (5´ to 3´)	Antisense primer (5´ to 3´)
SVP	WT	GACCCACTAGTTATCAGCTCAG	AAGTTATGCCTCTCTAGGAC
	mutant	GACCCACTAGTTATCAGCTCAG	AAGTTATGCCTCTCTAGGTT
SDG2	WT	GAGAGACTGTACAGAGATTC	CTTCCCCAAGGATGAATCAG
	mutant	ATAATAACGCTGCGGACATCTACATTTT	TGGTTCACGTAGTGGGCCATCG
DEK domain	WT	CCAAAAAGGGGGAAATCTGG	ATCGAACGAGGCTAAGCCTC
	mutant	CCAAAAAGGGGGAAATCTGG	TGGTTCACGTAGTGGGCCATCG
ATHD2	WT	AAATAGCCCCAAACCCACTG	GATATTGCTCGATGTTGGGG
	mutant	TAGCATCTGAATTTCATAACCAATCTCGATACAC	GATATTGCTCGATGTTGGGG
HTA5	WT	AGCAAAACCCTAAAGCCCAC	AACTCCTGAGAAGCAGATCC
	mutant	GGGCTACACTGAATTGGTAGCTC	AACTCCTGAGAAGCAGATCC
ATHD2A	WT	AAGCCAGTTACAGTGACTCC	GCTCTTCTTCCTCAGAATCC
	mutant	GGGCTACACTGAATTGGTAGCTC	GCTCTTCTTCCTCAGAATCC
GCN5	WT	AAGATTGTGCTAGTCGCTCC	TTAGCACCAGATTGGAGACC
	mutant	TGGTTCACGTAGTGGGCCATCG	TTAGCACCAGATTGGAGACC
HSP	WT	TTCGGGTTTTGTCCCATTCC	TTCACCTCTGATACGACGAG
	mutant	TGGTTCACGTAGTGGGCCATCG	TTCACCTCTGATACGACGAG
NAP1	WT	TAACTGAGCGAGATGAAGGG	CATCCTCATCATCTTCCTCG
	mutant	GGGCTACACTGAATTGGTAGCTC	CATCCTCATCATCTTCCTCG
35S::SVP-GFP		GGATGACGCACAATCCCACTATCC	AGAGAACGGAGAGTTCTTCAGC
pSVP::SVP-GFP		GCAACTAACGGAAGAGAACGAG	AGTCGTGCTGCTTCATGTGG

Table 1: List of primers used for genotyping analysis.

Generation of Arabidopsis transgenic lines

For the overexpression of *SVP-GFP* chimera, the *SVP* genomic locus (from start codon to the last nucleotide before the stop codon) was cloned into pB7FWG2 destination vector, under the control of the *35S-CaMV* constitutive promoter, and introduced into the *svp* mutant background by *Agrobacterium tumefaciens*-mediated transformation. *35S-CaMV::SVP-GFP svp* complemented plants were then selected by virtue of their resistance to the Basta herbicide. The *SVP-GUS* transgenic line was generated by cloning the whole *SVP* genomic locus, starting from the promoter region (-3000 bp from the ATG start codon) until the last nucleotide before the stop codon, into pBGWFS7 destination vector. The GUS construct was then introduced in WT plants by Agrobacterium-mediated transformation, as described above. Transformant plants were then crossed with the *SDG2/sdg2 svp* mutant plants. From the segregating F2 population, different *SVP-GUS* transgenic lines in *svp* and *sdg2* backgrounds were obtained. Primers used for gene cloning are listed in Table 2.

Name of the construct	Sense primer (5´ to 3´)	Antisense primer (5´ to 3´)
35S::SVP-GFP	$ATCCGGCGAGAAACTGAACC	#CCACCACCATACGGTAAGCTGC
pSVP::SVP-GUS	$ATCCGGCGAGAAACTGAACC	#CCACCACCATACGGTAAGCTGC
GGGGACAAGTTTGTACAAAAAAGCAGGCT$; GGGGACCACTTTGTACAAGAAAGCTGGGT#		

Table 2: List of primers used for the generation of SVP-GFP and SVP-GUS protein chimeras.

Analysis of SVP containing protein complexes

i. Nuclear protein isolation

Intact native nuclear protein complexes, were obtained from WT, *35S-CaMV::SVP-GFP* and *SVPp::SVP-GFP* seedling and inflorescence materials (0,5 g) as following:

1. Materials were broken in liquid nitrogen using a 15 cm mortar and pestle.

2. 30 mL of Buffer 1 were added, under shaking, to the powder in presence of liquid nitrogen, which freezes the buffer forming blocks. Those blocks will be then broken and mixed with the powder.
3. Powdered material was let thaw in the mortar and subsequently was poured through a 50 µm mesh into a centrifuge tube.
4. The intact nuclei were than pelleted at 4000 xg for 20 min at 4°C.
5. 1 mL of the supernatant was then collected and treated with 2 µL of protease inhibitor cocktail. After a maximum speed centrifugation in a bench-top centrifuge for 30 min at 4°C, the supernatant was collected and considered as extranuclei fraction.
6. Nuclei pellet obtained from step 4, was then gently resuspended in 10 mL of Buffer 1.
7. Nuclei were then again pelleted at 3000 xg for 10 min at 4 °C.
8. 10 mL of Buffer 1 were used to resuspend once again the nuclei pellet.
9. Nuclei were then again pelleted at 2000 xg for 10 min at 4 °C.
10. Intact nuclei pellet was then resuspended with 1 mL of Buffer 2 and moved to a 1,5 mL tube.
11. Intact nuclei in their native state were then obtained through a centrifugation in a bench-top centrifuge at 2000 xg for 8 min at 4°C.
12. Nuclei were then disrupted by adding 2/3 (v/v) of Buffer 3 to the pellet. Nuclei disruption was obtained under shaking for 45 min at 4°C.
13. To ensure the complete nuclei membrane disruption, the material was also sonicated twice for 20 seconds at amplitude of 15%. Between the two sonication rounds, the samples were chilled in ice.
14. Nuclear protein complexes in their native state, were then obtained as clear supernatant after a centrifugation at 2500 xg for 15 min at 4°C in a bench-top centrifuge.

Buffer 1 (Nuclei extraction buffer):

50 mM MES pH 8,5, 25 mM KCl, 5 mM $MgCl_2$, 30% (v/v) glycerol, 5% (w/v) sucrose. To be added fresh: 10 mM βmercaptoethanol, 1 mM DTT, 1 mM PMSF, 5 µg/mL Chymostatin, 5 µg/mL Leupeptin, 5 µg/mL Antiparin and 0,3% Triton.

Buffer 2 (Nuclei extraction buffer without triton):

Buffer 1 without triton.

Buffer 3 (Native extraction buffer):

0,45 M KCl, 25% (v/v) glycerol, 20 mM Hepes pH 8,0, 3 mM $MgCl_2$, 0,2 mM EDTA, 2 mM DTT and 1:200 protease inhibitor cocktail (Serva protease inhibitor mix #39103, 64,2 mg in DMSO).

ii. Protein co-immunoprecipitation and trypsin digestion

Native nuclear SVP protein complexes from WT and chimeric *SVP-GFP* plants, either under native or constitutive promoter, were further co-immunoprecipitated using the µMACS Epitope Tag Protein Isolation Kits (MACS molecular n°130-091-125), according to Smaczniak et al. (2012c), as following:

1. 50 µL of α-GFP microbeads were added to each native nuclear protein sample obtained previously, and then incubated for 1 h on a rotating device at 4°C.
2. The µ-Columns were placed in the µMACS separator and calibrated with 200 µL of lysis buffer.
3. The immunoprecipitated sample from step 1, was then applied onto the column and let run through it by gravity flow.
4. The immobilized beads were washed six times with 200 µL of lysis buffer.
5. Two washes with 200 µL of buffer 2 were applied to the immobilized beads.

6. 20 µL of 8 M urea (freshly prepared) were then applied onto the column and incubated for 5-10 min at room temperature.
7. Nuclear denaturated proteins were finally eluted into 1,5 mL low-protein-binding tube, with 50 µL of 8 M urea.

At this point, the proteins so eluted, had to be prepared for the following mass spectrometry (MS) analysis. To do so, the proteins were digested with trypsin enzyme and small polypeptides were obtained. According to (Smaczniak et al. 2012c), the trypsin digestion was performed as following:

1. Protein eluate (50 µL starting volume), was diluted four times with 50 mM ammonium bicarbonate (freshly prepared), to decrease the urea concentration to 2 M.
2. 10 µL of 45 mM DTT were added to the sample, and incubated for 30 min at 37°C under shaking (550 rpm) to reduce protein disulfide bonds.
3. The mixture was chilled slowly at room temperature for 5 min.
4. The thiol groups eventually present on the aminoacid residues, were then alkylated by adding 10 µL of 100 mM iodoacetamide and incubating the samples in the dark for 30 min.
5. 15 µL of 0,1 µg μL^{-1} trypsin was added to the sample and digestion was performed overnight (16-24 h) at 37°C.
6. Digestion was then stopped by adding 2 µL of 10% (v/v) TFA thus to bring the pH to a value of 3.
7. The sample was then centrifuged at maximum speed in a bench-top centrifuge for 5 min and the clear supernatant was then moved into a new 1,5 mL low-protein-binding tube, to make sure no solid contaminant was present in the sample.
8. The material so obtained, was ready for the subsequent MS analysis.

iii. Mass Spectrometry and bioinformatics analysis

Digested polypeptides obtained from three biological replicates from WT, *SVPp::SVP-GFP* and *35S-CaMV::SVP-GFP* were then analyzed with an Orbitrap Mass Spectrometry instrument. Raw MS data obtained were further analyzed employing a MaxQuant v.1.2.2.5 software (Smaczniak et al. 2012c), using a LFQ quantification strategy, which allows to compare protein abundance between IP samples and IP control. Basically, data obtained from WT samples were consider as unspecific protein interactors, and then subtracted from the chimeric lines samples data as background noise. Furthermore, also the protein known to be not localized in the nucleus, were considered as IP contaminants and then treated as background noise, as well as all the protein found to belong to other organisms rather than *Arabidopsis thaliana*. Short list of significative peptides was then made combining the data obtained from both the *SVPp::SVP-GFP* and *35S-CaMV::SVP-GFP* samples MS analysis, which display a good p-value score, an high enrichment fold and a false discovery rate (FDR) of 0,01. Each peptide present in this short list, was then associated to a known protein present in the database, based on the aminoacidic sequence. Strong putative SVP's interactors list, was then generated considering also pertinence of biological function, role in floral transition or flower development, as well as suitable cellular compartmentalization of each protein with regard to the SVP functions.

Nucleic acid analysis

Plant DNA was extracted according to Varotto et al. 2000. Total RNA isolation, from either vegetative or inflorescence tissues, was performed using the NucleoSpin - RNA isolation from plant kit (Machery-Nagel n° 740949.50). Subsequently, 2 μg of total RNA were employed for first-strand cDNA synthesis using GoScript Reverse Transcription System (Promega) according to the supplier's instructions. qRT-PCR analysis was carried out on a CFX96 Real-Time system (Bio-Rad), the primers used are listed in Table 3. *SAND* and *UBIQUITIN* genes were used as internal standards. Data from three biological and three technical replicates were analyzed with Bio-Rad CFX Manager software (version 3.1).

GUS staining analysis

Inflorescence samples, employed for the GUS histochemical assay, were previously fixed in 90% (v/v) acetone at -20°C for 20 min. Samples were then washed 3 times with 50 mM NaPi pH 7,2 before adding the GUS staining solution (50 mM NaPi pH7,2, 5 mM EDTA, 0,1% (v/v) triton X-100, 5 mM ferrocyanide, 5 mM ferricyanide and 1 mg/mL X-gluc). GUS staining solution was let infiltrate the sample for 5 min under vacuum (2,5 inHg), and then incubate for 14 h at 37°C in the dark. Staining reaction was stopped washing the sample 3 times with 70% (v/v) Ethanol, each washing step last 1 h, to remove all the GUS solution residues. Samples were prepared for optical microscopy observation with a cellulose digestion process employing a clearing solution (160 g of chloral hydrate [Sigma-Aldrich] dissolved in a solution consisting of 100 ml of water and 50 ml of glycerol) for 3 h. A Zeiss AxioImager AZ microscope equipped with differential interference contrast optics was used to observe the assay; images were captured with a Zeiss Axiocam MRc3 camera using Zen Imaging Software (Zen 2011 SP1).

Chromatin immunoprecipitation analysis

i. Material crosslinking

Twelve days old seedling were collected (0,5 g) in 25 mL MC buffer in a 50 mL tube kept in ice. Chromatin crosslinking was performed adding 1% (v/v) formaldehyde to the MC buffer under vacuum (2,5 inHg) for 20 min. Crosslinking reaction was then stopped with glycine to a final concentration of 125 mM under vacuum for other 5 min. Sample was then washed 3 times in MC buffer, to remove all the formaldehyde residues, and dried on a paper tissues. The sample material was grinded to a very fine powder in liquid nitrogen using a mortar. Crosslinked chromatin samples were then stored at -80°C waiting to be immunoprecipitated.

MC buffer

0,1 M sucrose, 50 mM NaCl and 10 mM NaPi pH7

ii. Chromatin immunoprecipitation

α-H3K4me3 antibody from Millipore (07-473) was employed for the ChIP assay, according to Gregis et al. 2013 with some small differences, as hereinafter described:

1. 30 µL and 10 µL of A-dynabeads (Invitrogen n°10003D) for each IP sample and No-Ab control, respectively, were washed twice in a clean 1,5 mL tube with 1 mL of PBS buffer added with 0,1% (v/v) triton X-100. To remove the washing buffer, the beads were let attached to a magnet.
2. A second round of washing was applied to the beads, using 1 mL of ChIP Dilution Buffer (CDB).
3. The beads were then resuspended in 100 µL final volume of CDB and incubated with 1,5 µL of α-H3K4me3 antibody for each sample for 4 h at 4°C in rotation. No-Ab tube beads, was incubated without adding the antibody, as a control for the eventual unspecific cross-reaction between dynabeads and chromatin.
4. Crosslinked sample frozen powder was resuspended in 30 mL of Extraction Buffer 1 (EB1) and left in ice for 5 min to let thaw slowly the material.
5. The suspension was then filter through Miracloth into a new 50 mL tube.
6. Intact nuclei were then pelleted at 3000 xg for 20 min at 4°C.
7. The supernatant was then discarded and pellet was gently resuspended in 1 mL of Extraction Buffer 2 (EB2) and then transferred in a new 1,5 mL tube.
8. Solution was then spinned at 12000 xg for 10 min at 4 °C to pellet the nuclei.
9. Nuclei pellet was then resuspended in 300 µL of Extraction Buffer 3 (EB3).
10. 300 µL of EB3 were laid into a clean 1,5 mL tube, and the 300 µL resuspended pellet from step 6 were then carefully layer on the top.

11. Nuclei were then separated from the rest of cytosolic and plastid material through a 1 h centrifugation at 14000 xg at 4°C.

12. The supernatant was discarded and the nuclei pellet gently resuspended in 300 µL of Nuclei Lysis Buffer (NLB), which caused the disruption of nuclei integrity and the releasing of the chromatin. A small aliquot (1-2 µL) from this step was taken for each sample, as a pre-sonication chromatin control.

13. Chromatin was then fragmented by sonication (amplitude 15%). Five round of sonication were applied to the sample, each round consisted in five sonication pulses of 30 sec each, followed by 30 sec of rest. The chromatin fragmentation step was performed at 4°C.

14. Fragmented chromatin was spinned at 10000 xg for 5 min at 4°C and supernatant was recovered in a clean 1,5 mL tube.

15. A small aliquot (1-2 µL) of clean fragmented chromatin was taken apart as a post-sonication chromatin control.

16. Quality and size of the fragmented chromatin so obtained, was checked on 2% (w/v) agarose electrophoresis gel. Chromatin fragment enrichment around 500 bp size was considered of good quality.

17. From the fragmented chromatin sample, two aliquots of 10 µL were transferred in a clean 1,5 mL tube as Input (stored at -20°C) and No-Ab sample, respectively. The remaining chromatin material was employed as IP sample.

18. IP and No-Ab samples were then diluted 1 to 10 with CDB, to dilute the SDS present in the NLB from 10% to 1% (v/v)

19. Antibody conjugated and No-Ab Dynabeads from step 3, were then washed twice with CDB and resuspended in 30 µL and 10 µL in the same buffer, respectively.

20. IP and No-Ab samples were then incubated overnight at 4°C in rotation with the conjugated and No-Ab dynabeads, respectively.

21. Each incubated sample was then washed twice, for 10 min under rotation at 4°C with Low Salt Wash Buffer, High Salt Wash Buffer, LiCl Wash Buffer and TE Buffer.

22. Chromatin was then eluted from the beads through an incubation with 250 μL of Elution Buffer at 65°C under gently shaking. Elution step was repeated twice, and the supernatant was collected into a new 1,5 mL tube.
23. Input sample from step 17 was let slowly thaw and then was resuspended in 500 μL final volume of TE buffer.
24. IP, No-Ab and Input samples, were then de-crosslinked overnight at 65°C under gently shaking, in presence of 200 mM NaCl. The De-crosslinking step is extremely important to recover chromatin from protein bonds and make it available for the further qRT-PCR analysis.
25. De-crosslinking step was then stopped incubating the sample for 1 h at 45°C in presence of 10 mM EDTA, 35 mM Tris-HCl pH 6,5 and 20 μg of proteinase K.
26. DNA was then recovered through phenol/chloroform/isoamyl alcohol washing and precipitated at -20°C for 3 h in presence of 1/10 (v/v) sodium acetate and 3 volumes of ethanol.
27. DNA pellet was then washed twice with 70% (v/v) ethanol and resuspended in clean water, ready to be analyzed by qRT-PCR.

Extraction Buffer 1

0,4 M sucrose, 10 mM Tris-HCl pH 8, 10 mM $MgCl_2$, 1 mM EDTA, 5 mM β-mercaptoethanol and 1:200 protease inhibitor cocktail (Serva protease inhibitor mix #39103, 64,2 mg in DMSO).

Extraction Buffer 2

0,25 M sucrose, 10 mM Tris-HCl pH 8, 10 mM $MgCl_2$, 1 mM EDTA, 1% (v/v) triton X-100, 5 mM β-mercaptoethanol and 1:200 protease inhibitor cocktail.

Extraction Buffer 3

1,7 M sucrose, 10 mM Tris-HCl pH 8, 2 mM $MgCl_2$, 1 mM EDTA, 0,15% (v/v) triton X-100, 5 mM β-mercaptoethanol and 1:200 protease inhibitor cocktail.

Nuclei Lysis Buffer

50 mM Tris-HCl pH 8, 10 mM EDTA, 1% (v/v) Sodium Dodecyl Sulfate (SDS) and 1:200 protease inhibitor cocktail.

ChIP Dilution Buffer

1,1% (v/v) Triton X-100, 1,2 mM EDTA, 16,7 mM Tris-HCl pH 8, 167 mM NaCl and 1:200 protease inhibitor cocktail.

Elution Buffer

1% (v/v) SDS and 0,1 M $NaHCO_3$.

Low Salt Wash Buffer

150 mM NaCl, 0,1% (v/v) SDS, 1% (v/v) Triton X-100, 2 mM EDTA and 20 mM Tris-HCl pH 8.

High Salt Wash Buffer

500 mM NaCl, 0,1% (v/v) SDS, 1% (v/v) Triton X-100, 2 mM EDTA and 20 mM Tris-HCl pH 8.

LiCl Wash Buffer

0,25 M LiCl, 1% (v/v) NP-40, 1% (v/v) sodium deoxycholate, 1 mM EDTA and 10 mM Tris-HCl pH 8.

TE Buffer

10 mM Tris-HCl pH 8 and 1 mM EDTA.

iii. Real-Time analysis

Input, IP and no Ab fractions were then analyzed by qRT-PCR assay. Primers used for the qRT-PCR analysis are listed in the Table 3. Relative % enrichment of the target genes in the IP versus input fraction of WT, *sdg2* and *svp* mutants was calculated according to

Mizzotti et al. 2014; no-Ab qRT-PCR values were subtracted to the % of input calculation in each sample as unspecific background noise, probably due to the cross-reaction with the chromatin-dynabeads. *Ta3* transposon gene was used as negative ChIP internal control, according to Guo et al. (2010). *ACTIN 7* housekeeping gene, was used for the internal normalization of the relative % enrichment calculation.

Target gene	Sense primer (5´ to 3´)	Antisense primer (5´ to 3´)
	Expressional analysis	
SVP	GGCGAGAGAAAAGATTCAGA	GTCCGCAATTTCTTTACTCATTC
UBIQUITIN	CTGTTCACGGAACCCAATTC	GGAAAAAGGTCTGACCGACA
SAND	CAGACAAGGCGATGGCGATA	GCTTTCTCTCAAGGGTTTCTGGGT
	ChIP analysis	
Promoter.1	GAATGGGTTTGTAGTAGTTGC	CGAAGAAACAGTGAATAGAGG
Promoter.2	CCAACCAATTAGTTACAAGGAG	TGGTTATTGGATATGGTGAGG
Exon	ATCCCGTTCTCGAAAGATCC	TTCCGGTGGAAGAGAAGATG
Intron 1.1	CTGATACATAGGAGTTTACTGTATC	TGAATATTACCGTAGTTAGATACC
Intron 1.2	GACCCACTAGTTATCAGCTCAG	CACCGACAATATACCAAAATCT
Ta3	GTGAATCTTGTTTGGCTCTAAATAACATAAAAAG	GTTACTTTGGTCTGTTTGATGCTCTTTCCCC
ACTIN 7	CGTTTCGCTTTCCTTAGTGTTAGCT	AGCGAACGGATCTAGAGACTCACCTTG

Table 3: List of primers used for qRT-PCR expression analyses and ChIP qRT-PCR analyses.

RESULTS

The SVP-GFP chimeras are functional

MADS box TFs share high homologous aminoacid sequences, therefore highly specific antibodies cannot be generated by using full-length recombinant proteins, as antigens. Furthermore, previous attempts to generate SVP-specific antibodies using short peptides did not give rise to optimal antibodies for biochemical studies, such as Co-ImmunoPrecipitation analysis (CoIP). Therefore, it was necessary to use the SVP chimeric proteins, which could be detected and immunoprecipitated by using commercially available antibodies. In our case, we decided to generate Arabidopsis transgenic plants carrying the SVP MADS box TF fused to the green fluorescent protein (GFP), since the c-terminal GFP fusion was reported to be functional for other MADS box TFs (de Folter *et al.*, 2007). In particular, the SVP genomic sequence, devoid of the promoter region, was fused 3' with GFP coding sequence (see Figure 8a) and placed under the control of the constitutive 35S Cauliflower Mosaic Virus (35S-CaMV) promoter. The construct was introduced into *svp* knockout line via *Agrobacterium tumefaciens*-mediated transformation.

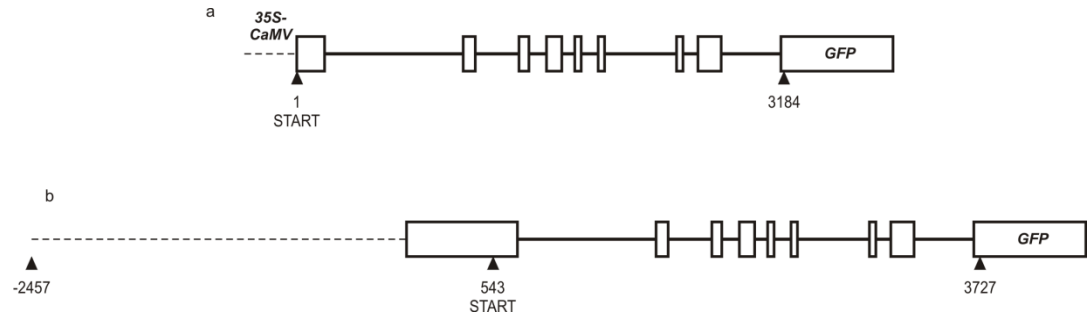

Figure 8: Schematic representation of the *SVP-GFP* chimeric constructs used in thesis. (a) Scheme of the *35S-CaMV::SVP-GFP* construct, which consists of the whole *SVP* genomic locus, starting from the ATG codon, fused to the *GFP* sequence at the 3' end after removing the stop codon, and controlled by the constitutive *35S-CaMV* promoter. (b) Scheme of the *SVPp::SVP-GFP* construct, which consist of the whole *SVP* genomic locus from -2457 bp upstream the 5' UTR region, fused at the 3' end with the *GFP* sequence. Exons and introns are represented as open blocks and black lines, respectively. Promoter regions are represented in dashed lines.

The functionally of the SVP-GFP chimera was initially verified by monitoring its accumulation into the nuclei of transgenic plants using confocal microscopy. As shown in Figure 9a, the GFP fluorescence could be clearly observed inside the nuclei of root cells,

that have been chosen for the observation in order to avoid any background signal given by chlorophyll fluorescence. Furthermore, the selected transgenic lines, overexpressing the SVP-GFP chimera displayed a late flowering phenotype compared to WT plants in LD conditions, calculated on the basis of the number of rosette leaves at the time of flowering (Figure 9c and Table 4), similar to the late flowering phenotype observed in plants overexpressing the SVP protein alone (Lee *et al.*, 2007), as well as the aberrant leafy-like flowers (Figure 9b), thus supporting the complete functionality of the SVP-GFP chimera, and making it a reliable tool for proteomic studies.

Finally, Arabidopsis transgenic lines carrying the SVP-GFP chimera under the control of the *SVP* native promoter, *SVPp::SVP-GFP* (see Figure 8b), and able to fully rescue the early flowering phenotype of *svp* mutant plants (Figure 9c and Table 4) were already available in the lab (see Gregis et al. in 2013) and could also be used for proteomic studies.

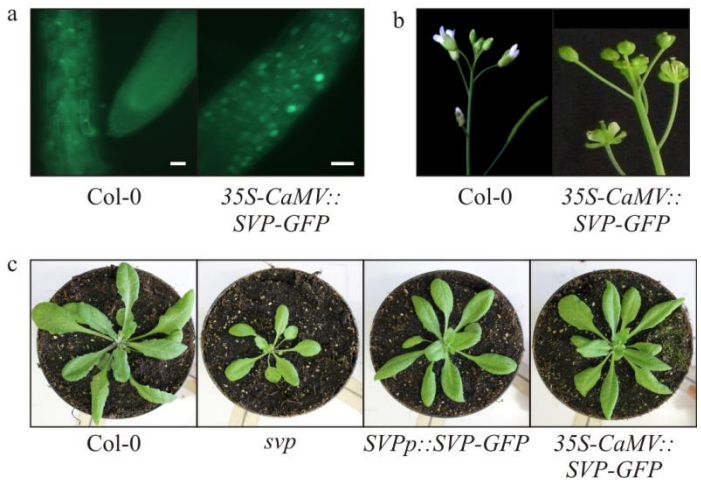

Figure 9: *SVP-GFP* complemented plants. (a) GFP fluorescence detection at confocal microscope in Col-0 and 35S-*CaMV*::*SVP-GFP* root cells. Fluorescence of the SVP-GFP chimera was clearly visible only in the nuclei of 35S-*CaMV*::*SVP-GFP* complemented plants. Bar = 25 µm. (b) Col-0 and *35S-CaMV::SVP-GFP* inflorescence. Plants carrying the 35S-*CaMV*::SVP-GFP construct showed an aberrant leafy-like phenotype of sepals whorl, as well as the total absence of petals whorl, as expected from the overexpression of *SVP*. (c) Vegetative rosette pictures of different *SVP* background plants at the time of flowering in LD condition. *svp* plants showed an early flowering phenotype, whereas the complemented *SVPp::SVP-GFP* and *35S-CaMV::SVP-GFP* plants had a WT-like flowering phenotype.

	Long day
WT	10,9 ± 1,5
svp	5,7 ± 0,5
SVPp::SVP-GFP	9,1 ± 0,9
35S-CaMV::SVP-GFP	12,3 ± 1,3

Table 4: Flowering time data calculated as number of rosette leaves, together with the corresponding standard deviations, for each plant line shown in Figure 9c, and grown under LD conditions.

SVP co-immunoprecipitation assay

The *SVPp::SVP-GFP* and *35S-CaMV::SVP-GFP* transgenic lines, able to fully rescue the *svp* mutant phenotype and with the constructs in homozygosity, were employed for CoIP assays and Mass Spectrometry analyses (CoIP-MS). The line with SVP ectopic expression would have guaranteed high SVP protein accumulation and made the proteomic analysis more simple, whereas the *SVPp::SVP-GFP* construct would have allowed the correct *SVP* temporal and spatial expression, thus the native SVP-GFP protein complex formation in the right tissues and at the right developmental stages. Ideally, a merged list of SVP protein interactors isolated from leaves and inflorescences of both transgenic lines through CoIP under native conditions would have provided robust SVP interactors.

Initially, protocols for the isolation of purified nuclei and extraction of nuclear proteins under native conditions from both leaf and inflorescence were established (see Materials and Methods for a detailed description of the protocols). The quality of nuclear native extracts from frozen materials was evaluated both by Coomassie Brilliant Blue staining, looking for plastid contaminants such as RUBISCO (Figure 10a), and immunoblot analysis where the histone enrichment in nuclear extracts was evaluated (Figure 10a). As shown in Figure 10, RUBISCO was only barely detectable in the nuclear fraction, whereas HISTONE 3 (H3) signal was clearly observed in the same fraction. Moreover, no H3 band was present in the extra-nuclear sample, indicating that no nuclear disruption occurred during the nuclei isolation protocol, hence no nuclear protein complex was lost during the realization of this procedure.

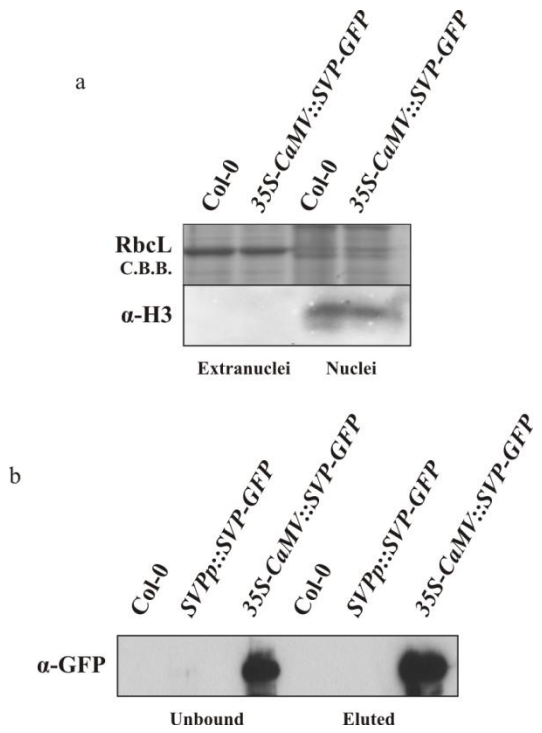

Figure 10: Nuclei isolation and SVP Co-ImmunoPrecipitation (CoIP) controls. (a) The quality of the Nuclei isolation protocol from inflorescences was checked through Coomassie Brilliant Blue (C.B.B.) staining to monitor the accumulation of the large subunit of RUBISCO (RbcL) and immunoblot assay, using an α-Histone 3 (H3) antibody, to assess the enrichment of histones in the nuclei fraction . The RbcL subunit was almost exclusively present in the extranuclei fraction, whereas was only barely detectable in the nuclei fraction. On the contrary, immunoblot analysis revealed the presence of the H3 protein only in the nuclei fraction, thus confirming the good quality of the nuclei isolation protocol. (b) The quality of the SVP-GFP CoIP was checked by immunoblot analysis, using a GFP specific antibody to detect the presence of SVP-GFP chimera. The chimeric protein was highly abundant in the *35S-CaMV::SVP-GFP* samples, both in the unbound and eluted fractions, whereas the GFP signal was detectable nor in WT samples, used as negative control, neither in the eluted fraction of *SVPp::SVP-GFP* samples. Note that identical data were obtained for the isolation of nuclei from seedlings.

Once the nuclei isolation and native protein extraction protocols were correctly established, seedling and inflorescence nuclear native protein extracts from WT, *SVPp::SVP-GFP* and *35S-CaMV::SVP-GFP* plants were prepared and employed, as starting material, in CoIP-MS assays. In particular, the SVP-GFP chimeras, together with their putative interactors, were trapped by an α-GFP antibody and, after few washing steps, aimed to eliminate unbound protein complexes, were eluted and collected (for a detailed description of the protocol, see Materials and Methods). The presence of GFP protein in IP eluates and in the unbound fractions was further checked by immunoblot, using a GFP-specific antibody (Figure 10). GFP-specific signals were only detectable in the *35S-*

CaMV::SVP-GFP samples, both in the unbound and eluted fractions, as consequence of the high protein accumulation which made easier to detect the protein. Clearly, the large amount of SVP-GFP protein was even able to saturate the binding capacity of α-GFP antibody used for the CoIP. As expected, it was not possible to detect any GFP signal in the eluted fraction, obtained from plants expressing the SVP-GFP chimera under the control of the native *SVP* promoter, due to the very low abundance of this protein in the nuclear fraction. As a matter of fact, no SVP-specific immunoblot has yet been published. Nevertheless, all the CoIP samples were further analyzed by Mass spectrometry.

Mass spectrometry and putative SVP interactors

CoIP samples, consisting of SVP-GFP chimeras and all the other proteins which were trapped and eluted together with them, were digested with trypsin allowing then the generation of short peptides, which have been further analyzed through the Q ExactivePlus Orbitrap LC-MS/MS, in collaboration with the proteomic platform at the University of Wageningen (https://www.wur.nl/en/product/Q-ExactivePlus-Orbitrap-LC-MSMS.htm). Based on the UniProt Knowledgebase for Arabidopsis proteins, peptide identification, protein assembly, and protein quantification were performed with MaxQuant software (v1.2.2.5, Max Planck Institute of Biochemistry, Martinsried, Germany), using default parameters.

The high sensitivity of the mass spectrometer allowed the identification of GFP and SVP peptides, as top hits, in both transgenic lines, together with several other peptides belonging to proteins with different functions (Table 5). Mean $\log_2$ ratios of abundances of proteins precipitated from *SVPp::SVP-GFP* and *35S-CaMV::SVP-GFP* versus wild-type lines and corresponding *P* values of significance, derived by Student's *t* test statistics and subsequent adjustment to control the false discovery rate according to Benjamini and Hochberg (1995), were determined. Proteins were considered as robust SVP interactors (see Table 5) when i) they were identified both in *SVPp::SVP-GFP* and *35S-CaMV::SVP-GFP* CoIP samples, ii) they showed more than 1.5-fold difference in abundance between *SVPp::SVP-GFP*, *35S-CaMV::SVP-GFP* and wild-type CoIPs, iii) they fulfilled the

statistical significance criterion of $P \leq 0.05$, iv) the biological pertinence of the results, based on predicted gene function, protein localization and literature data, existed.

Interestingly, among the proteins robustly identified in the co-immunoprecipitates of SVP-GFP, several factors involved in different chromatin processes were present. Indeed, nucleic acid binding proteins, RNA and DNA helicases, histone modification enzymes, protein involved in the maintenance and assembly of nucleosome and histone proteins were among the best SVP putative interactors (Table 5). Interestingly, these data well fit with the literature information, which largely reports about the existence of interactions between MADS box TFs and chromatin remodeling and histone modifying enzymes (Smaczniak et al. 2012b). In particular, the presence of many histone modifying enzymes among the SVP's putative interactors, together with several helicase enzymes, suggest that SVP and, more in general MADS box TFs, can regulate the expression of their target genes by the employment of epigenetic factors able to remodel the histone density and chromatin accessibility nearby the target CArG boxes.

Gene ID	Description
-	Green Fluorescent Protein
At2g22540	Short Vegetative Phase: MADS-box MICK protein; FUNCTION IN: transcriptional activity
At1g51060	P-loop containing nucleoside triphosphate hydrolases superfamily protein; FUNCTIONS IN: helicase activity
At2g45810	DEA(D/H)-box RNA helicase family protein; FUNCTIONS IN: helicase activity
At2g33730	P-loop containing nucleoside triphosphate hydrolases superfamily protein; FUNCTIONS IN: helicase activity
At3g06480	DEAD box RNA helicase family protein; FUNCTIONS IN: helicase activity
At2g42520	P-loop containing nucleoside triphosphate hydrolases superfamily protein; FUNCTIONS IN: helicase activity
At5g05450	P-loop containing nucleoside triphosphate hydrolases superfamily protein; FUNCTIONS IN: helicase activity
At5g51280	DEAD-box protein abstrakt, putative; FUNCTIONS IN: helicase activity
At3g01540	RNA HELICASE DRH1 protein_coding DEAD BOX RNA HELICASE 1 (DRH1)
At5g62190	DEAD/DEAH box RNA helicase PRH75
At3g19760	RNA helicase that may be a component of the Exon Junction Complex
At3g58510	DEA(D/H)-box RNA helicase family protein; FUNCTIONS IN: helicase activity
At5g11170	DEAD/DEAH box RNA helicase family protein ; FUNCTIONS IN: helicase activity
At1g08880	HTA5, a histone H2A protein.
At3g20670	HTA13, a histone H2A protein.
At3g45980	Histone 2B (H2B) protein.
At3g44750	Histone deacetylase. Controls the development of adaxial/abaxial leaf polarity.
At3g53650	Histone superfamily protein;INVOLVED IN: nucleosome assembly;
At5g02570	Histone superfamily protein; INVOLVED IN: nucleosome assembly.
At3g54610	Histone acetyltransferase required to regulate the floral meristem activity
At5g58470	TBP-associated factor 15B (TAF15b); nucleic acid binding.
At5g42520	BASIC PENTACYSTEINE 6 DNA binding
At5g63550	DEK domain-containing chromatin associated protein
At5g22650	ARABIDOPSIS HISTONE DEACETYLASE 2
At4g15180	SET domain protein 2 (SDG2)
At4g40030	Histone superfamily protein
At2g19480	NUCLEOSOME ASSEMBLY PROTEIN 1

Table 5: List of strong putative SVP interactors obtained from CoIP-MS analyses performed on inflorescence tissues. Putative SVP interactors were selected based on the criteria listed above in the main text. Only proteins found in both *SVPp::SVP-GFP* and *35S-CaMV::SVP-GFP* CoIP-MS data are reported in the list. Unfortunately, the attempt to produce a similar list of SVP interactors from seedling tissues failed.

Phenotypes of Arabidopsis mutants lacking some selected SVP interactors

After the identification of putative SVP protein partners, functional genomics studies were initiated in order to validate their interactions with SVP and their involvement in the determination of flowering time and flower meristem identity. Eight genes, among the ones reported in Table 5, were selected (see Table 6) and T-DNA insertional mutants were then

ordered at the Nottingham Arabidopsis Stock Centre (NASC, http://arabidopsis.info/). Those genes belong to different classes of DNA-binding proteins and histone modification enzymes. Among them, we selected proteins with DNA-binding features, like DEK domain and HISTONE SUPERFAMILY PROTEIN (HSP), which are reported to be involved in the correct chromatin assembly (Pendle, 2004; Okada, Singh and Bhalla, 2006), as well as HTA5, a particular histone structural protein involved also in the *FLC* activation in seedling, before vernalization occurred (Deal *et al.*, 2007). The short list of putative SVP interactors (see Table 6) also includes the NUCLEOSOME ASSEMBLY PROTEIN 1 (NAP1), a chromatin binding protein involved in DNA maintenance and repair (Z. Liu *et al.*, 2009), which is also known to interact with others SVP putative protein interactors present in the list, the ARABIDOPSIS THALIANA HISTONE DEACETHYLASE 2 (ATHD2) and ATHD2A, two histone deacetylase proteins reported to be involved in plant organ polarity sensing growth (Colville *et al.*, 2011; Son *et al.*, 2015). Other two histone modifying enzymes were also included in the short list, the GENERAL CONTROL NONREPRESSED-PROTEIN 5 (GCN5), an histone deacetylase, and the SET DOMAIN GROUP 2 (SDG2), an histone trimethyltransferase enzyme. GCN5 is reported to be part of two large deacethylase protein complexes, SAGA and ADA2, involved in seed fatty acid accumulation, iron homeostasis and heat stress response (Hu *et al.*, 2015; Xing *et al.*, 2015; Wang *et al.*, 2016). On the other hand, SDG2 has been reported to be the major trimethyltransferase enzyme in *Arabidopsis thaliana*, by promoting the expression of several genes through the methylation of the fourth lysine residue on the histone 3 (H3K4me3) (Guo *et al.*, 2010). Moreover, SDG2 appears to play a major role in flowering time regulation, either by directly regulating *FLC* expression via a FRI-dependent pathway, or through an FRI-independent flowering pathway active under short-day growth conditions, which causes a flowering delay in the *sdg2* mutant plants (Yun *et al.*, 2012).

Gene ID	Gene name	Description	NASC code
At5g63550	DEK domain	DEK domain-containing chromatin associated protein	SALK_137152.43.05.x
At5g22650	ATHD2	Arabidopsis Histone Deacetylase 2	SAIL_1247 A02
At1g08880	HTA5	Histone H2A Protein	GABI_097F11
At4g15180	SDG2	SET Domain Protein 2	SALK_021008.5600.x
At3g44750	ATHD2A	Histone Deacetylase	GABI_355H03
At3g54610	GCN5	Histone Acetyltransferase	SALK_106557.46.80.x
At4g40030	HSP	Histone Superfamily Protein	SALK_082765.35.95.n
At2g19480	NAP1	Nucleosome Assembly Protein 1	GABI_273H07

Table 6: Eight most prominent SVP's putative interactors selected based on the criteria listed above. In the table are reported the gene IDs, the gene name abbreviations, a short description of their biological function and T-DNA insertional line codes, according to the NASC database.

Plants were then grown on soil under greenhouse conditions, homozygous mutant plants identified by PCR-based genotyping and then checked for their flowering time and flower development phenotypes, in order to support further their interaction with SVP at phenotypic level. Already few days after sowing (DAS), it was possible to recognize differences in growth rate in two of the mutants, *sdg2* and *gcn5*. The delay in growth rate became even larger after 21 DAS, with *sdg2* and *gcn5* plants about three and two times smaller, than WT plant, respectively (Figure 11a). Moreover at 35 DAS, the rosette of *sdg2* plants displayed a characteristic pleiotropic phenotype with a bushy-like shape, curly leaves, and the total absence of secondary inflorescence meristems (Figure 11b).

Concerning the flowering time phenotype of those mutants grown under long-day greenhouse conditions, none of them showed any difference in terms of number of rosette leaves at the time of flowering. However, *sdg2* was able to flower only five days later than the other mutants and WT plants, probably as a consequence of the slow growth rate of *sdg2* plants, which can affect the floral transition step at the level of the SAM.

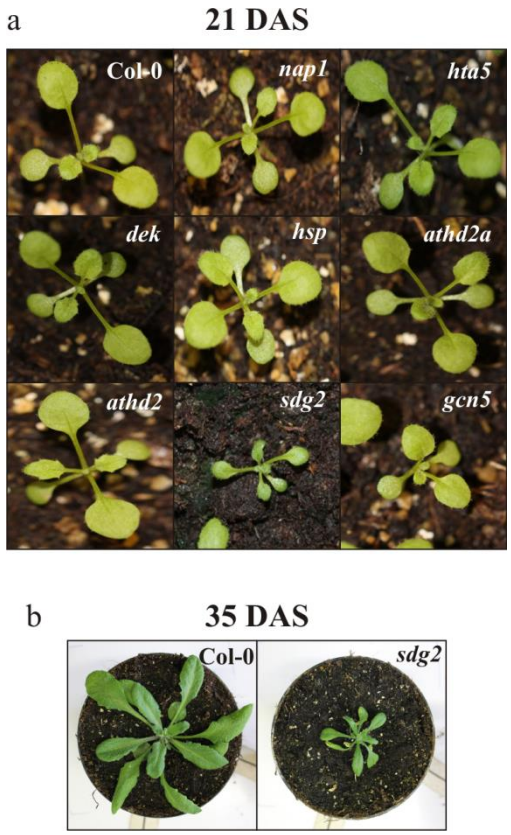

Figure 11: Phenotypes of Arabidopsis plants knockout in the eight selected SVP putative interactors. (a) Picture of the mutants taken after 21 days after sowing (DAS). Notably, *sdg2* and *gcn5* display a smaller plant size compare to WT plants, whereas all the other mutants have WT-like rosettes. (b) Comparison between WT and *sdg2* mutant plants after 35 DAS. The latter is much smaller than the WT control plant, and shows a vegetative pleiotropic phenotype, with a bushy-like rosette, due to the particular curvature of the leaves.

Among the eight selected mutants, only two showed defects at the level of flower morphology and functionality, giving rise to flowers completely or partially sterile. In particular, *sdg2* mutant plants were characterized by completely sterile flowers, whereas *gcn5* mutants did produce seeds in a small amount only on distal flowers, up in the inflorescence stem. Further investigations coupled with data available in the literature, revealed an impaired gametes formation in Arabidopsis flowers in the absence of SDG2 protein, explaining the sterile flower phenotype (Berr *et al.*, 2010). On the other hand, *gcn5* were found to have defects in flower formation, with particular regards to stamen formation. In *gcn5* plants, the stamens are shorter compared to WT, thus preventing the autofertilization of the ovary (Cohen *et al.*, 2009).

Based on the macroscopic phenotypic characteristics of the selected Arabidopsis mutants, together with literature data, the *SDG2* gene was chosen for more detailed functional genomics studies.

Flowering time in higher order *sdg2* mutants

The *sdg2* mutation was introduced into different *SVP* genetic backgrounds via manual crosses, in order to verify the existence of genetic interactions. The *svp sdg2* double mutant showed the same early flowering phenotype of *svp* single mutant plant (Table 7a), although the rosette phenotype was very similar to the *sdg2* single mutant.

However, when the *SVP* overexpression line *35S-CaMV::SVP-GFP* was combined together with the *sdg2* mutation, the late flowering phenotype, caused by the ectopic SVP expression, was completely reverted to a partially early flowering phenotype in *sdg2 35S-CaMV::SVP-GFP* plants (Table 7a). More interestingly, the same early flowering phenotype was also observed in the presence of the *SDG2*/*sdg2* heterozygous background, indicating that the absence of only one *SDG2* functional allele is able to rescue the late flowering phenotype typical of *35S-CaMV::SVP-GFP* plants (Lee *et al.*, 2007).

	Flowering time
WT	10,9 ± 1,5
svp	5,7 ± 0,5
35S-CaMV::SVP-GFP	12,3 ± 1,3
35S-CaMV::SVP-GFP - SDG2/sdg2	7,5 ± 1,2
35S-CaMV::SVP-GFP - sdg2/sdg2	7,0 ± 1,2
sdg2	10,4 ± 1,1
svp - sdg2	5,0 ± 0,0

Table 7: Flowering time phenotypes, in WT, *sdg2*, *svp*, *SVP-GFP* overexpression lines and the relative double mutant plants. Flowering time is expressed as number of rosette leaves counted at the time of bolting, together with the corresponding standard deviations.

Surprisingly, the *sdg2* mutation, both in heterozygous and homozygous backgrounds, could also rescue the aberrant flower morphology of the *SVP-GFP* overexpression line. As shown in Figure 12, in the presence of one or both *sdg2* mutant alleles, the *35S-CaMV::SVP-GFP* flowers completely reverted the leafy-like phenotype of the sepals and petals to the WT morphology. The rescue of the *SVP* overexpression flower phenotype driven by the *sdg2* genetic background, was further confirmed by independent crosses with other *35S-CaMV::SVP-GFP* lines in the presence of either one or both *sdg2* mutant alleles.

Based on these observations, it appears clear that SVP, at least when its expression is under the control of a constitutive promoter, needs the presence of both *SDG2* functional alleles to explicate its role as florigen repressor, as well as floral homeotic repressor.

Figure 12: Phenotypes of *35S-CaMV::SVP-GFP* reproductive organs in the presence of different *SDG2* genetic backgrounds. (a) *35S-CaMV::SVP-GFP* inflorescence shows its typical rounded bud phenotype, which is completely reverted to a normal WT bud morphology (see b) in absence of one or both *SDG2* functional alleles. (c) Detail of the *35S-CaMV::SVP-GFP* flower, showing the aberrant phenotype due to the *SVP* overexpression: sepals are converted into leafy-like organs, and petals become sepals. (d) The absence of one or both *SDG2* functional alleles is able to completely rescue the aberrant flower phenotype of the *SVP-GFP* overexpression to a normal WT morphology.

SVP expression in *sdg2* genetic background

Since the *sdg2* mutation was able to rescue the *SVP-GFP* overexpression phenotype both on flowering time and flower morphology, we investigated further this aspect at molecular level. In particular, the first question we tried to address was whether the rescue of the phenotype was due to changes in *SVP* gene expression.

To verify this aspect, a quantitative Real-Time PCR (qRT-PCR) was run to quantify the changes of the *SVP* expression levels in *SDG2/sdg2* heterozygous and *sdg2* homozygous genetic backgrounds. The assay evaluated the expression of endogenous *SVP*, as well as the expression of *SVP* under the control of the *35S-CaMV* promoter, both in seedlings and inflorescences (Figure 13).

In particular, in seedling tissues the *sdg2* knockout mutant showed a decrease of the endogenous *SVP* expression of about half of the control WT seedlings, whereas in presence of the heterozygous *SDG2* genetic background, *SVP* expression did not change significantly, only 0,2 times less than WT. However, looking at the *35S-CaMV::SVP-GFP* lines, the *sdg2* mutation caused a drastic decrease of *SVP* ectopic expression in both heterozygous and homozygous mutant seedlings, in a dose-dependent manner. In fact, in seedlings wild type at the *SDG2* locus, the *SVP* ectopic expression was about 70 times more than endogenous *SVP* expression, whereas in the presence of *SDG2/sdg2* heterozygous genetic background, *SVP* ectopic expression was decreased by more than 80% (only about 12 times more than WT), and a reduction by about 90% (only about 7 times more than WT) of the *SVP* ectopic expression was observed in the *sdg2* homozygous line.

The same dose-dependent trend of ectopic *SVP-GFP* expression in the presence of both heterozygous and homozygous *sdg2* mutation was also observed in the inflorescence tissues. Heterozygous *SDG2/sdg2* inflorescences showed a reduction by about 90% of *SVP* expression, and even more in *sdg*2 homozygous genetic background. Despite the trend observed in the inflorescence samples with respect to SVP ectopic expression, the *sdg2* mutation did not repress the endogenous *SVP* expression as it did in the vegetative tissues. Actually, a slight increase of endogenous *SVP* transcription could be observed both in heterozygous *SDG2/sdg2* (about 20% more than WT) and *sdg2* homozygous inflorescences (about 50% more than WT).

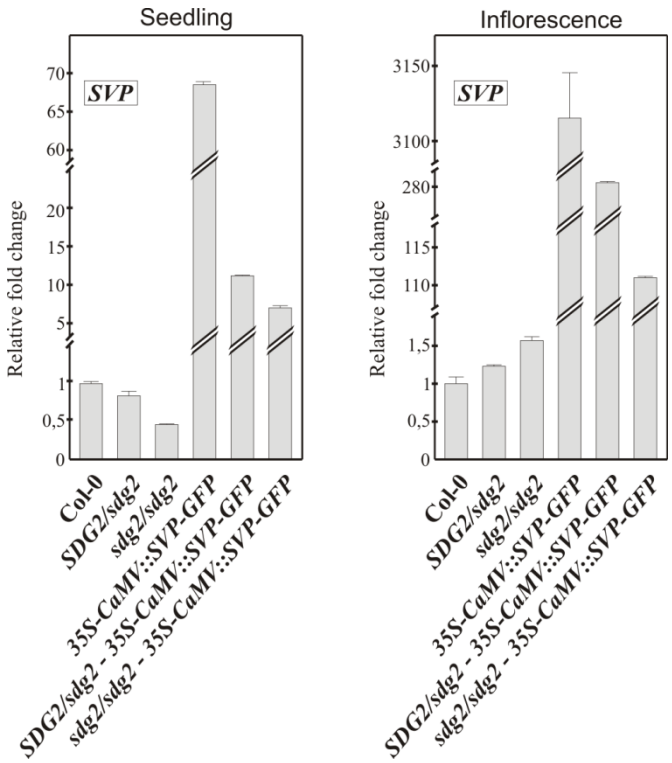

Figure 13: *SVP* expression monitored by qRT-PCR in WT (Col-0) *SDG2/sdg2*, *sdg2* and SVP overexpression lines combined with the *sdg2* heterozygous and homozygous genetic backgrounds, both from seedlings and inflorescence tissues. Relative fold enrichment was normalized on the *UBIQUITIN* housekeeping gene expression, and normalized with respect to the endogenous *SVP* expression in WT plants, that was set to 1.

Despite the data collected from inflorescence tissues of *sdg2* mutants, which are in contrast with all the others presented in this paragraph, SDG2 seems to have a major role in promoting *SVP* expression. qRT-PCR data obtained from *35S-CaMV*::*SVP-GFP* plants, both at the level of vegetative tissues and inflorescences, were indeed consistent with the phenotypic observation, i.e. rescue of flowering time and flower morphology phenotypes observed in homozygous *sdg2* and heterozygous *SDG2/sdg2* plants, carrying the *35S-CaMV::SVP-GFP* construct.

SVP **expression in the floral meristem is under the control of SDG2 enzyme**

Based on the qRT-PCR data reported above (see Figure 13), SDG2 is necessary for *SVP* endogenous expression only in vegetative tissues, whereas in the inflorescence tissues, SDG2 appears to have a repressive role. These seemingly contradictory results might be explained with the fact that *SVP* is expressed in very few cells, for a very short time window, and in the early stage of floral meristem making the precise detection of *SVP* transcripts very difficult by qRT-PCR analysis. Furthermore, *sdg2* inflorescences produce far less floral buds and their development is delayed compared to the WT plant, thus the *sdg2* amount of *SVP* cDNA could be overestimated because of the relative enrichment of the early stage of floral meristem, compared to the whole inflorescence tissues, isolated from WT samples.

To further investigate this aspect, a GUS staining was performed on *SVPp::SVP-GUS* lines, introgressed into different *SDG2* and *SVP* genetic backgrounds. The *SVP-GUS* construct showed a correct *SVP* expression pattern in WT and *svp* mutant inflorescence, at the stage 1 and 2 of floral meristem (Figure 14a and d).

However, mutation in *sdg2* alleles, were able to repress the accumulation of the SVP-GUS chimera both in presence of the endogenous *SVP* protein (Figure 14b and c) and in its absence (Figure 14e and f). In particular, the SVP-GUS chimera accumulated specifically in the right cells around the floral meristem, but the intensity of the signal in plants heterozygous at the SDG2 locus was lower than WT, and the signal completely disappeared in the *sdg2* homozygous mutant background. Most probably, that does not mean that *SVP* expression is completely silenced by the lack of SDG2 protein, certainly the signal was too low to be detected. In order to make quantitative the GUS assay, the samples were further tested with different X-Gluc concentration (from 0,5 mg/mL to 0,125mg/mL). Higher X-Gluc concentrations revealed the presence of GUS signals also in the *sdg2* mutant backgrounds, but saturated the GUS signal in WT and *svp* backgrounds, whereas lower X-Gluc concentrations made barely visible the signals in WT and *svp* backgrounds (data not shown). Thus, the best X-Gluc concentration was found to be 0,25 mg/mL, as shown in Figure 14.

Taken together, the data are consistent with the qRT-PCR results obtained from the endogenous *SVP* regulation in the vegetative tissues, as well as from *35S-CaMV::SVP-GFP* lines, pointing to a major role of SDG2 in promoting *SVP* expression both in the vegetative and inflorescence meristems.

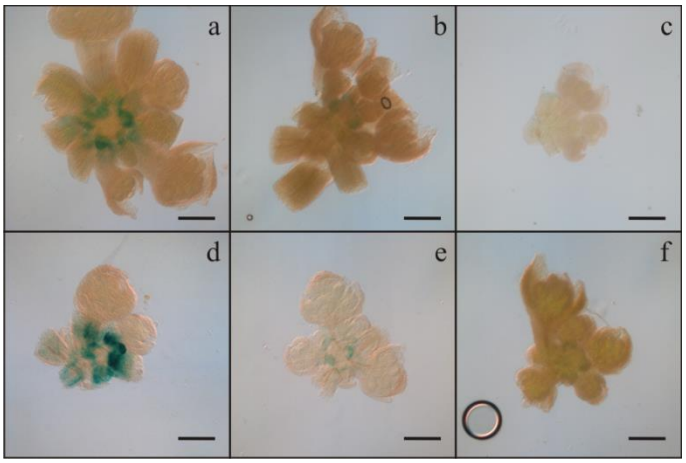

Figure 14: GUS staining assay to monitor the expression of the SVP-GUS chimera. *SVP* expression was analyzed in the inflorescence tissues of WT (a), *SDG2/sdg2* heterozygous (b), *sdg2* (c), *svp* (d), *svp SDG2/sdg2* heterozygous (e) and *svp sdg2* double (f) mutant plants containing the *SVPp::SVP-GUS* construct, to evaluate the effects of SDG2 absence on the *SVP* expression. (a) In the WT background, the SVP-GUS signal was correctly localized around the IM, at the stage 1, 2 and 3 of the FM. In *SDG2/sdg2* heterozygous (b) background, the GUS signal was still correctly localized around the IM but was much less intense than the WT control. (c) In *sdg2* knockout background the signal was even too low to be detected. (d) An intense GUS signal could be detected in *svp* knockout background in the stage 1, 2 and 3 of FM, which appeared to be also slightly diffused into the non specific cells surrounding the IM. In (e) and (f), the lack of one or both the *SDG2* functional alleles, respectively, leads to a less accumulation of the SVP-GUS protein. As in (c), also in the *svp sdg2* double knockout mutant, the GUS signal was not detectable. For the assay, a GUS solution containing 0,25mg/mL of X-Gluc was used. Bar = 100 µm. Similar results were also obtained in the seedling material, confirming the qRT-PCR data (see also Figure 13).

Pattern of H3K4 trimethylation within the *SVP* gene

In order to investigate the molecular details of the SDG2-dependent regulation of SVP expression, a Chromatin ImmunoPrecipitation (ChIP) assay was performed using an antibody specific for H3K4me3 (see also Materials and Methods). In particular, we aimed to evaluate the methylation pattern of the *SVP* genomic locus in presence or absence of SDG2 enzyme. Furthermore, the hypothesis of the involvement of an SVP-SDG2 protein

complex in the methylation of *SVP* associated histones was also verified by an identical ChIP assay performed on the *svp* mutant plants (Figure 15 and Figure 16).

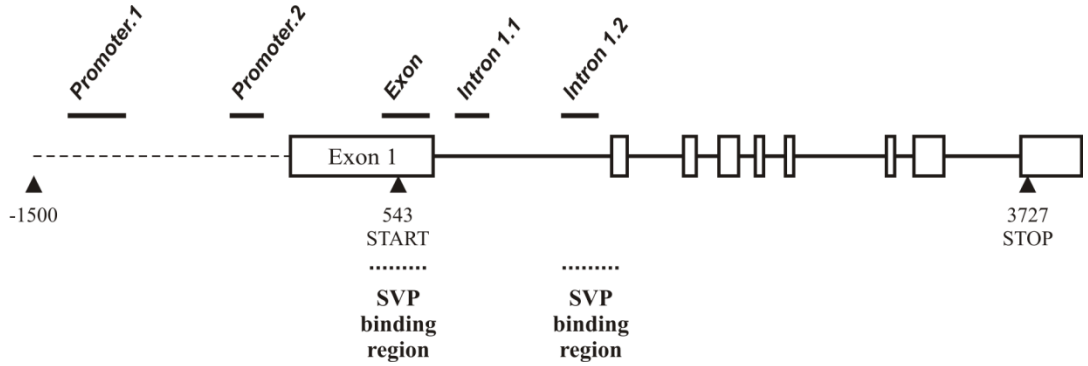

Figure 15: *SVP* genomic locus scheme together with the DNA regions tested in the ChiP assay. *SVP* regions analyzed through qRT PCR to evaluate an H3K4me3 differential enrichment by ChIP experiment, are reported with black lines, above the *SVP* genomic locus. Dashed lines, below the *SVP* genomic locus scheme, indicate the two SVP binding regions (CArG boxes) found by ChIP-seq analysis (Gregis *et al.*, 2013). Notably, regions named *Exon* and *intron 1.2* overlap those SVP binding regions.

Fourteen days old seedlings were employed for ChIP experiments, and the collected seedling chromatins were crosslinked to the histone proteins before fragmentation by sonication. Subsequently, the fragmented material was immunoprecipitated using an α-H3K4me3 antibody and then de-crosslinked (see also Materials and Methods for the detailed description of the protocol). Eluted and purified chromatin fragments were then tested by qRT-PCR to evaluate the enrichment, with respect to the input chromatin, of specific *SVP* genomic regions in the mutant IPs, *sdg2* and *svp*, with respect to the WT IPs, used as control. In particular, the *SVP* genomic locus H3K4me3 pattern was monitored by using five primer pairs spanning the promoter-first intron region of SVP (see also Figure 15). Notably, the region of the first exon, and the second region of the first intron (close to the 3' end) contain the CArG boxes found enriched by SVP-specific ChIP-seq analysis (see Gregis et al. 2013).

As described in Figure 16, qRT-PCR analyses showed that the *SVP* promoter region is not subjected to H3K4 trimethylation mediated by SDG2 enzyme, so probably *SVP* promoter region is regulated by different mechanisms. However, the *SVP* exon and the two regions of the first intron, showed a strong trimethylation pattern in WT background, supporting an active role of those genomic regions in the epigenetic *SVP* expression regulation. Moreover,

those epigenetic marks were strongly reduced, more than a half, in the absence of the SDG2 enzyme (Figure 16a), pointing to a direct role of this epigenetic factor on *SVP* regulation. Interestingly, the lack of SVP TF (Figure 16b), caused a marked decrease of H3K4 trimethylation around the first *SVP* exon, but not in the first SVP intron, indicating that the SVP-SDG2 protein complex is specifically required for the modification of histones placed within the SVP first exon. Based on these findings, it can be speculated that this modification contributes to the SDG2-mediated expression regulation of SVP at the vegetative and inflorescence meristems.

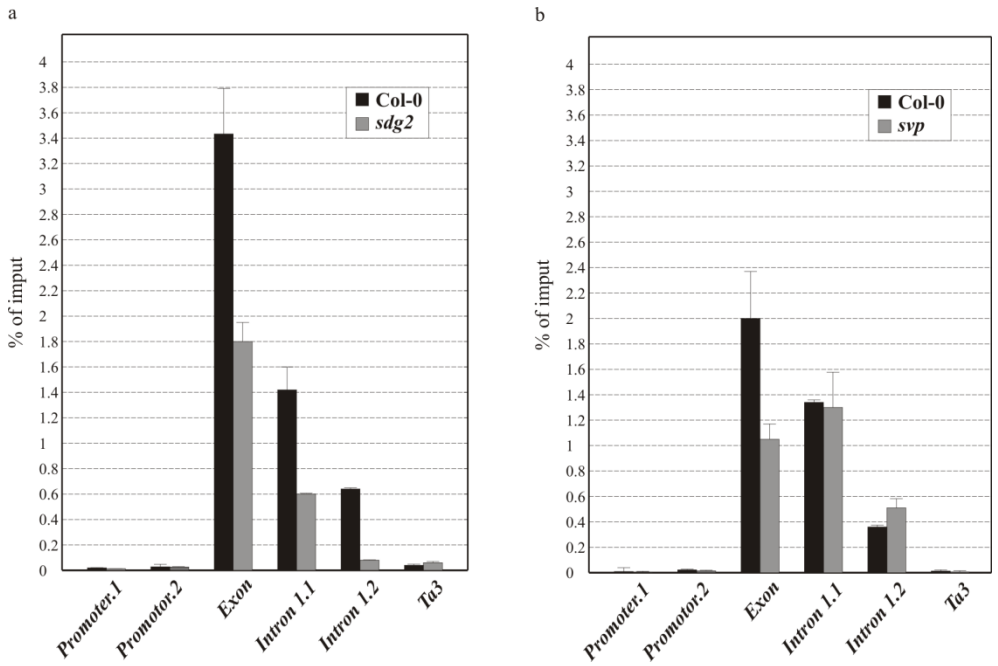

Figure 16: H3K4me3 ChiP assay on *SVP* genomic locus . Histograms refer to the IP chromatin region enrichment, with respect to the chromatin input, expressed in percentage. *ACT7* enrichment was used for the qRT-PCR data normalization. (a) *SVP* H3K4me3 pattern in *sdg2* genetic background versus Col-0. (b) *SVP* H3K4me3 pattern in *svp* genetic background versus Col-0. *Ta3* transposon genomic region was used as a negative H3K4me3 control for the ChIP experiment. Positions of the DNA fragments are indicated in Figure 15. Similar data were also observed in inflorescence material.

DISCUSSION

SVP's interactomic insight

SHORT VEGETATIVE PHASE gene encodes a nuclear MADS box transcription factor belonging to the MICK type II clades, and it has been reported having a role in different Arabidopsis developmental phases (Hartmann et al. 2000; Gregis et al. 2006; Smaczniak et al. 2012a). In particular, *SVP* acts as a strong florigen repressor gene, together with *FLC*, by the direct repression of *FT* and *SOC1* expression, in the so called vernalization flowering pathway (Jack, 2004; Gregis *et al.*, 2013). At the same time, *SVP* has been also reported to explicate its flowering repressor function in other flowering pathways, such as the GA pathway, and the thermosensory pathway (Andrés *et al.*, 2014; Hwan Lee *et al.*, 2014). Although *SVP* plays such important roles in the inflorescence meristem commitment, involving different inductive and non-inductive flowering pathways, not much is known about the molecular mechanisms and the protein partners involved in this important function. Furthermore, *SVP* has been little studied in the Arabidopsis reproductive tissues, where it shares a redundant function with other MADS box MICK proteins, such as *SOC1* and *AGL24* (Gregis *et al.*, 2009; Dorca-Fornell *et al.*, 2011). In particular, it has been reported that to ensure the FM identity and its repressor role during the early stage of flower development, SVP interacts with the repressor proteins, LEU and SEU (Gregis *et al.*, 2008).

Efforts have been made during the last decade to define the whole SVP interactor repertoire involved in those developmental processes. As many other MADS box TFs belonging to the MICK clade, also SVP needs to form protein complexes to explicate its roles in both the Arabidopsis vegetative and reproductive stages. The co-immunoprecipitation followed by the mass spectrometry analysis performed during this Ph.D. thesis, was aimed to shine further light on the SVP protein partners. The same approach has been, indeed, employed in interactome studies for other MADS box MICK TFs (Hill, Wang and Sharyn E. Perry, 2008; Immink *et al.*, 2009), allowing the discovery of various MADS-MADS proteins complexes organized in the typical MADS quaternary protein complex structure, as well as in higher order complexes containing proteins not belonging to the MADS box TF family. Due to the lack of a specific α-SVP antibody, an α-GFP antibody was employed for the immunoprecipitation of the SVP-GFP chimera. The use of GFP for the realization of functional protein chimeras is rather common also in the case

of MADS box TF functional studies (de Folter *et al.*, 2007). Indeed, the SVP-GFP protein was able to fully rescue the early flowering phenotype of *svp* mutant plants, when its expression was controlled by the SVP native promoter, *SVPp::SVP-GFP*, and the *35S-CaMV::SVP-GFP* lines showed the delayed flowering phenotype and flower morphology typical of plants overexpressing the *SVP* gene alone, indicating that the SVP-GFP chimera was fully functional and could be used for CoIP-MS studies.

Several proteins were identified to co-immunoprecipitate with SVP, however in this list no MADS box FT was present, probably due to the high stringency of the nuclei isolation buffer used during the isolation procedure (see Materials and Methods), which may have interfered with the weaker SVP-MADS box protein-protein interaction, thus preserving only stronger SVP-protein interactions. Nevertheless, the identified putative interactors were very interesting from a molecular point of view. All of them belong to classes of proteins involved in chromatin processes, which well fits with data obtained from CoIP-MS analyses of other MADS box TFs such as AP1, SEP3, AG, AP3 and PI (Smaczniak et al. 2012b). In particular, SVP CoIP data, revealed the presence of three nucleosome assembly factors (At3g53650; At5g02570; At2g19480), five histone proteins (At1g08880; At3g20670; At3g45980; At4g40030) or proteins associated to the chromatin (At5g63550) and four proteins involved in histone modification (At3g44750; At3g54610; At5g22650; At4g15180). Those findings can suggest an epigenetic function of SVP protein complexes, which may remodel the chromatin density by modifying the histone residues nearby the SVP target genes. In this model, SVP could act as a guide able to drive the SVP-containing protein complexes to specific DNA binding sequences (CArG box), where the histone modifying proteins can induce changes on chromatin density, with the aim to activate or repress gene expression via epigenetic-based molecular mechanisms. It is known, indeed, that histone modifications such as acetylation and methylation, can alter the density of nucleosomes, thus making a genomic region more or less accessible to the gene expression machinery (Jones and Sung, 2014; Lu *et al.*, 2015).

It is still unclear, whether SVP binds firstly the CArG box regions and then recruits the other protein partners, or on the contrary, the SVP protein complex is formed before the binding to DNA-specific regions. Certainly, further experiments are needed to address this aspect. Moreover, the fact that also other MADS protein (AP1, AG, SEP3 and AGL15) were found to interact with histone modifying factors, such as RELATIVE OF EARLY

FLOWERING 6 (REF6), CHROMATIN REMODELLING (CHR) protein class, SIN3, as well also PIKLE (PKL) and ATINO 80 (INO80) nucleosome remodeling factors (Smaczniak et al. 2012b; Hill, Wang & Sharyn E. Perry 2008), makes our findings on SVP interactome very interesting and supports the existence of a common regulatory mechanism within MADS box TFs. Beside nucleosome and histone modifying enzymes, SVP was also co-immunoprecipitated with other proteins acting as DNA/RNA helicases. Detailed functional analyses of this class of proteins are certainly needed to explain at molecular level the importance of their interactions with SVP.

SDG2: a necessary enzyme for the correct expression of *SVP*

Among all the putative SVP's interactor proteins, the SET DOMAIN GROUP 2 (SDG2) protein has been chosen as the best candidate for further functional analysis, based on the phenotype of *sdg2* knockout mutant plants, both at vegetative and reproductive phase.

SDG2 is a relatively large gene, about 11 kb, organized in 21 exons, that encodes the major Arabidopsis histone trimethylase, responsible of the methylation of lysine 4 (K4) of the Histone 3 (H3). *In vitro* methylation assays, revealed also a dimethylase activity on the same H3K4 residue (Guo *et al.*, 2010). This protein has been shown to be involved in several biological processes, as revealed by the pleiotropic phenotype of *sdg2* T-DNA mutant plants. Roots of *sdg2* plants are shorter in length, as well as lateral roots are strongly reduced in number compared to WT plants (Yao *et al.*, 2013). Plants lacking the SDG2 protein are also smaller than WT, with shorter curly leaves and the absence of lateral inflorescence meristems on the vegetative rosette (Guo *et al.*, 2010). Later during the reproductive phase, *sdg2* mutants are characterized by altered gamete development, both at male and female organs, leading to the formation of sterile flowers (Berr *et al.*, 2010). Furthermore, ChIP experiments using a H3K4me3 specific antibody in *sdg2* mutant background, revealed a direct regulation of *FLC* expression driven by the SDG2-mediated H3K4me3 pattern (Yun *et al.*, 2012). In addition, the SDG2 protein has been reported to play an active role in the circadian clock gene expression regulation (Malapeira and Mas, 2013). These findings

support clearly the important role of SDG2 in Arabidopsis flowering time determination, too, and help to explain its role as SVP interactor.

Notably, even if SDG2 was reported to be necessary for the activation of *FLC* florigen repressor gene, which would implicate an early flowering phenotype of *sdg2* plants, the absence of *FRIGIDA*, that together with *SDG2* is responsible for the correct expression of *FLC* in leaves, in the Col-0 ecotype used in our experiments did not allow to detect any defect in flowering time between WT and *sdg2* plants, grown under long-day conditions (see also Table 7). For the same reason, the *svp sdg2* double mutant displayed an *svp*-like early flowering phenotype under long-day conditions.

Nevertheless, *sdg2* homozygous and *SDG2/sdg2* heterozygous plants were able to rescue the late flowering phenotype and the aberrant flower morphology of the *SVP-GFP* overexpression plants. The rescue mechanism appears to be due to the need of both functional *SDG2* alleles (Figure 13) for the ectopic expression of *SVP*, both in the vegetative and inflorescence tissues. The SDG2 dose-dependent regulation of *SVP* ectopic expression suggests that a large amount of SDG2 is needed to reach the very high levels of SVP expression in the vegetative tissues (68 times more expressed than WT) and in the inflorescence (3105 times more expressed than WT). These findings also allowed to hypothesize a regulatory role of SDG2 in the expression of the *SVP* endogenous gene. Indeed, qRT-PCR and GUS staining assays confirmed the essential, dose-dependent role of SDG2 in the expression of SVP in tissues of WT seedling (see Figure 13 and Figure 14).

Surprising, it was not possible to confirm the same SDG2-dependent *SVP* endogenous regulation mechanism in the WT inflorescence tissues via qRT-PCR. The Real-time based assay, performed on these samples, revealed a slightly increased of *SVP* cDNA concentration in *sdg2* mutant background, which was in contrast with the previous findings (see Figure 13). As already speculated in the RESULT section, that could be probably due to an underestimation of *SVP* transcripts in WT inflorescence tissues, because of the very specific cell expression of SVP, which makes very difficult the sampling of inflorescence material. However, the GUS staining assay allowed to overcome this issue, since it does not need to be performed on collected floral meristem sample, as in the case of qRT-PCR, but on whole inflorescence material (see Figure 14). As a matter of fact, the GUS staining confirmed the importance of SDG2 in the spatial and temporal expression of SVP during the different flower developmental stages within the inflorescence tissues of WT plants. In addition, the

same experiment showed the *SVP* expression downregulation in *SDG2/sdg2* heterozygous background, that became even more evident in *sdg2* homozygous mutants, confirming the dose–dependent regulatory role of SDG2 on *SVP* gene expression.

SVP genomic regulatory regions

Our data support a transcriptional regulatory mechanism of *SVP* mediated by the SDG2 enzyme. Furthermore, we have shown by CoIP and MS analysis that SVP interacts physically with SDG2. Finally, literature data support the ability of SVP to bind to its own genomic regions, within the first exon and the first intron, due to the presence of CArG boxes (Gregis *et al.*, 2013). Thus, taken all the information together, it was straightforward to speculate the existence of an SVP-SDG2 containing protein complex with a key role in the transcriptional regulation of *SVP* in vegetative and inflorescence meristems. This protein complex could interact with the exon- and intron-located SVP binding regions, modulate the nucleosome density nearby the binding region, through epigenetic histone modification (H3K4me3), and then promote *SVP* expression. The ChIP experiments reported in this thesis (see Figure 16), support indeed a major regulatory role of the SVP-SDG2 protein complex through its interaction with the first exon of SVP (see also Figure 17). In fact, ChIP experiments revealed that the methylation pattern of H3 associated with the first exon of SVP is highly influenced by the presence of both SVP and SDG2 proteins, whereas SVP TF does not seem to be required for the H3K4 trimethylation of the first intron. Most probably, SDG2 can methylate the histone 3 in *SVP* intron, by forming protein complexes with other DNA binding proteins.

In conclusion, it can be inferred that the impaired H3K4me3 pattern, in close proximity to the first *SVP* exon, caused by the lack of SVP-SDG2 protein complex activity, is responsible of *SVP* expression downregulation in seedling and inflorescence tissues, implying an important role SVP-SDG2 protein complex in promoting *SVP* expression.

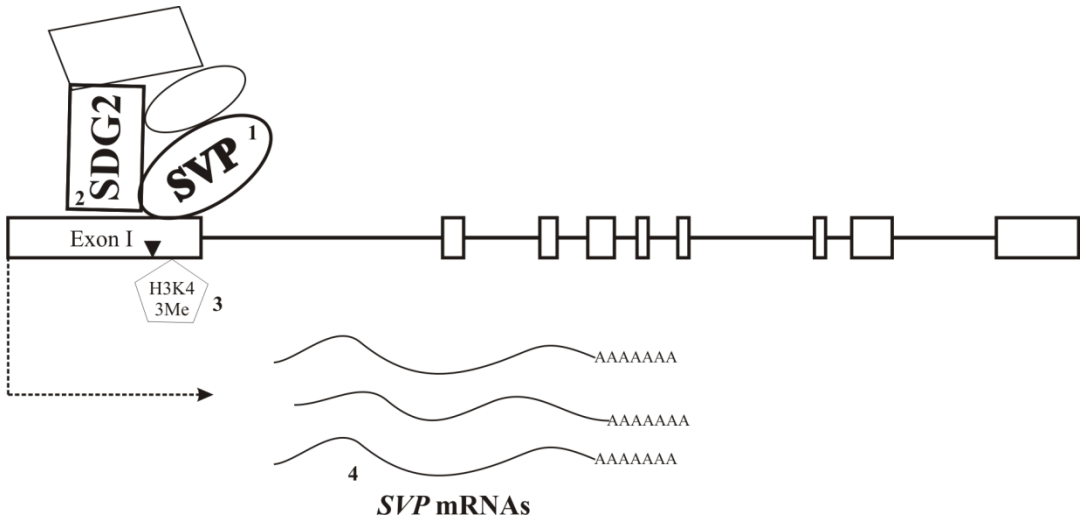

Figure 17: Model of the *SVP* gene expression regulation mediated by SDG2 enzyme in SAM and IM tissues. The *SVP* genomic locus is schematically reported, exons and introns are represented as open blocks and black lines, respectively. The black arrow within first exon box, indicates the position of the start codon. The model suggests an SVP-mediated positive loop regulation driven by the interaction of an SVP-SDG2-conteining protein complex with the first exon of SVP, in close proximity of the start codon. The presence of a CArG box element on the first *SVP* exon (see also Figure 15), allows the binding of SVP protein (1) on its own sequence. Based on our findings, the SDG2 protein (2) can be recruited by SVP forming a protein complex which can lay on *SVP* first exon and, through the SDG2 methylase activity, induce the trimethylation of the H3K4 (3) in close proximity to the *SVP* start codon. H3K4me3 modification is responsible for the active *SVP* transcription (4).

CONCLUSION

Among the MADS box TFs, SVP has two important regulatory roles in both vegetative and reproductive phases: i) responsible of flowering time determination and ii) inflorescence meristem identity. Nevertheless, not much is known about the molecular details of its functions and about its protein partners. The first part of the Ph.D. project was aimed to identify SVP protein partners through a proteomic approach, based on CoIP and MS analysis, employing a functional SVP-GFP chimera. From the list of putative SVP protein partners, SDG2, the major *Arabidopsis thaliana* trimethylase enzyme, was chosen for further functional analyses, also guided by the phenotypic features of *sdg2* knockout plants.

Gene expression analyses in combination with Chip assays revealed the existence of an SDG2-dependent *SVP* transcriptional regulation. In particular, it has been shown that H3K4me3 pattern in close proximity to the *SVP* first exon can drastically affect *SVP* expression. The trimethylation of the H3K4 residue appears to function as an activator of gene expression, when it occurs within the first exon of SVP. Such transcriptional regulatory mechanism appears to be essential for the ectopic expression of SVP in the *35S-CaMV::SVP-GFP* lines and for the expression of the endogenous gene in vegetative and inflorescence meristems.

These data together with the functional characteristics of the other SVP protein partners indicate that SVP plays its regulatory role in vegetative and reproductive tissues through the interactions with different proteins involved in nucleosome and chromatin remodeling, in agreement with data reported for other MADS box proteins. Certainly, detailed functional studies of the other SVP protein partners are needed to clarify the molecular mechanisms at the basis of SVP activities.

REFERENCES

Acevedo, F. G. *et al.* (2004) 'FLOR1, a putative interaction partner of the floral homeotic protein AGAMOUS, is a plant-specific intracellular LRR', *Plant Science*, 167(2), pp. 225–231.

Achard, P. *et al.* (2006) 'Integration of plant responses to environmentally activated phytohormonal signals.', *Science (New York, N.Y.)*, 311(5757), pp. 91–4.

Adamczyk, B. J., Lehti-Shiu, M. D. and Fernandez, D. E. (2007) 'The MADS domain factors AGL15 and AGL18 act redundantly as repressors of the floral transition in Arabidopsis.', *The Plant journal : for cell and molecular biology*, 50(6), pp. 1007–19.

Aguilar-Martínez, J. A. *et al.* (2015) 'Transcriptional, posttranscriptional, and posttranslational regulation of SHOOT MERISTEMLESS gene expression in Arabidopsis determines gene function in the shoot apex.', *Plant physiology*, 167(2), pp. 424–42.

Andrés, F. *et al.* (2014) 'SHORT VEGETATIVE PHASE reduces gibberellin biosynthesis at the Arabidopsis shoot apex to regulate the floral transition.', *Proceedings of the National Academy of Sciences of the United States of America*, 111(26), pp. E2760-9.

Andrés, F. and Coupland, G. (2012) 'The genetic basis of flowering responses to seasonal cues.', *Nature reviews. Genetics*. Nature Publishing Group, 13(9), pp. 627–39.

Aukerman, M. J. and Sakai, H. (2003) 'Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes.', *The Plant cell*, 15(11), pp. 2730–41.

Balanzà, V., Martínez-Fernández, I. and Ferrándiz, C. (2014) 'Sequential action of FRUITFULL as a modulator of the activity of the floral regulators SVP and SOC1', *Journal of Experimental Botany*, 65(4), pp. 1193–1203.

Barkan, A. and Small, I. (2014) 'Pentatricopeptide Repeat Proteins in Plants PPR: pentatricopeptide repeat', *Annu. Rev. Plant Biol*, 65, pp. 415–42.

Baudry, A. *et al.* (2010) 'F-box proteins FKF1 and LKP2 act in concert with ZEITLUPE to control Arabidopsis clock progression.', *The Plant cell*, 22(3), pp. 606–22.

Berr, A. *et al.* (2010) 'Arabidopsis SET DOMAIN GROUP2 is required for H3K4 trimethylation and is crucial for both sporophyte and gametophyte development.', *The Plant cell*, 22(10), pp. 3232–3248.

Blázquez, M. A., Ahn, J. H. and Weigel, D. (2003) 'A thermosensory pathway controlling flowering time in Arabidopsis thaliana.', *Nature genetics*, 33(2), pp. 168–71.

De Bodt, S. *et al.* (2003) 'And then there were many: MADS goes genomic.', *Trends in plant science*, 8(10), pp. 475–83.

Bond, D. M. *et al.* (2009) 'Histone acetylation, vernalization insensitive 3, flowering locus C, and the vernalization response', *Molecular Plant*, 2(4), pp. 724–737.

Bowman, J. L., Smyth, D. R. and Meyerowitz, E. M. (1991) 'Genetic interactions among floral homeotic genes of Arabidopsis', *Development*, 112(1), pp. 1–20.

Briggs, W. R. and Christie, J. M. (2002) 'Phototropins 1 and 2: versatile plant blue-light receptors.', *Trends in plant science*, 7(5), pp. 204–10.

Chan, K. X. *et al.* (2016) 'Learning the Languages of the Chloroplast: Retrograde Signaling and Beyond', *Annual Review of Plant Biology*, 67(1), pp. 25–53.

Choi, K. *et al.* (2011) 'The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in Arabidopsis, by recruiting chromatin modification factors.', *The Plant cell*. American Society of Plant Biologists, 23(1), pp. 289–303.

Cohen, R. *et al.* (2009) 'The histone acetyltransferase GCN5 affects the inflorescence meristem and stamen development in Arabidopsis', *Planta*, 230(6), pp. 1207–1221.

Colombo, M. *et al.* (2008) 'AGL23, a type I MADS-box gene that controls female gametophyte and embryo development in Arabidopsis', *The Plant Journal*. Blackwell Publishing Ltd, 54(6), pp. 1037–1048. doi: 10.1111/j.1365-313X.2008.03485.x.

Colombo, M. *et al.* (2016) 'GUN1, a Jack-Of-All-Trades in Chloroplast Protein Homeostasis and Signaling', *Frontiers in Plant Science*, 7(September).

Colville, A. *et al.* (2011) 'Role of HD2 genes in seed germination and early seedling growth in Arabidopsis', *Plant Cell Reports*, 30(10), pp. 1969–1979.

Cseke, L. J. *et al.* (2007) 'Identification of PTM5 protein interaction partners, a MADS-box gene involved in aspen tree vegetative development', *Gene*, 391(1–2), pp. 209–222.

Davies, B. *et al.* (1996) 'Multiple interactions amongst floral homeotic MADS box proteins.', *The EMBO journal*. European Molecular Biology Organization, 15(16), pp. 4330–43.

Deal, R. B. *et al.* (2007) 'Repression of Flowering in Arabidopsis Requires Activation of FLOWERING LOCUS C Expression by the Histone Variant H2A.Z', *THE PLANT CELL ONLINE*, 19(1), pp. 74–83.

Demarsy, E. and Fankhauser, C. (2009) 'Higher plants use LOV to perceive blue light.', *Current opinion in plant biology*, 12(1), pp. 69–74.

van Dijk, A. D. J. *et al.* (2010) 'Sequence motifs in MADS transcription factors responsible for specificity and diversification of protein-protein interaction.', *PLoS computational biology*, 6(11), p. e1001017.

Dorca-Fornell, C. *et al.* (2011) 'The Arabidopsis SOC1-like genes AGL42, AGL71 and AGL72 promote flowering in the shoot apical and axillary meristems', *Plant Journal*, 67(6), pp. 1006–1017.

Eriksson, S. *et al.* (2006) 'GA4 is the active gibberellin in the regulation of LEAFY transcription and Arabidopsis floral initiation.', *The Plant cell*, 18(9), pp. 2172–81.

Espinosa-Soto, C. *et al.* (2014) 'Tetramer formation in Arabidopsis MADS domain proteins: analysis of a protein-protein interaction network.', *BMC systems biology*, 8, p. 9.

Fan, H. Y. *et al.* (1997) 'Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins.', *The Plant journal : for cell and molecular biology*, 12(5), pp. 999–1010.

Favaro, R. *et al.* (2003) 'MADS-box protein complexes control carpel and ovule development in Arabidopsis.', *The Plant cell*. American Society of Plant Biologists, 15(11), pp. 2603–11.

Feng, P. *et al.* (2016) 'Chloroplast retrograde signal regulates flowering', *Proceedings of the National Academy of Sciences*, (25), p. 201521599.

de Folter, S. *et al.* (2005) 'Comprehensive interaction map of the Arabidopsis MADS Box transcription factors.', *The Plant cell*, 17(5), pp. 1424–1433.

de Folter, S. *et al.* (2007) 'Tagging of MADS domain proteins for chromatin immunoprecipitation.', *BMC plant biology*, 7, p. 47.

Fornara, F., de Montaigu, A. and Coupland, G. (2010) 'SnapShot: Control of flowering in arabidopsis', *Cell*, 141(3), pp. 3–5.

Galvão, V. C. *et al.* (2015) 'Gibberellic acid signaling is required for ambient temperature-mediated induction of flowering in Arabidopsis thaliana.', *The Plant journal : for cell and molecular biology*, 84(5), pp. 949–62.

Gandikota, M. *et al.* (2007) 'The miRNA156/157 recognition element in the 3' UTR of the Arabidopsis SBP box gene SPL3 prevents early flowering by translational inhibition in seedlings.', *The Plant journal : for cell and molecular biology*, 49(4), pp. 683–93.

Gendall, A. R. *et al.* (2001) 'The VERNALIZATION 2 Gene Mediates the Epigenetic Regulation of Vernalization in Arabidopsis', *Cell*, 107(4), pp. 525–535

Gregis, V. *et al.* (2006) 'AGL24, SHORT VEGETATIVE PHASE, and APETALA1 redundantly control AGAMOUS during early stages of flower development in Arabidopsis.', *The Plant cell*, 18(6), pp. 1373–1382.

Gregis, V. *et al.* (2008) 'AGAMOUS-LIKE24 and SHORT VEGETATIVE PHASE determine floral meristem identity in Arabidopsis', *Plant Journal*, 56(6), pp. 891–902.

Gregis, V. *et al.* (2009) 'The Arabidopsis floral meristem identity genes AP1, AGL24 and SVP directly repress class B and C floral homeotic genes', *Plant Journal*, 60(4), pp. 626–637.

Gregis, V. *et al.* (2013) 'Identification of pathways directly regulated by SHORT VEGETATIVE PHASE during vegetative and reproductive development in Arabidopsis.', *Genome biology*. BioMed Central Ltd, 14(6), p. R56.

Griffiths, J. *et al.* (2006) 'Genetic characterization and functional analysis of the GID1 gibberellin receptors in Arabidopsis.', *The Plant cell*, 18(12), pp. 3399–414.

Guo, L. *et al.* (2010) 'SET DOMAIN GROUP2 is the major histone H3 lysine [corrected] 4 trimethyltransferase in Arabidopsis.', *Proceedings of the National Academy of Sciences of the United States of America*, 107(43), pp. 18557–62.

Harding, E. W. *et al.* (2003) 'Expression and maintenance of embryogenic potential is enhanced through constitutive expression of AGAMOUS-Like 15.', *Plant physiology*, 133(2), pp. 653–63.

Hartmann, U. *et al.* (2000) 'Molecular cloning of SVP: a negative regulator of the floral transition in Arabidopsis.', *The Plant journal : for cell and molecular biology*, 21(4), pp. 351–60.

He, Y. *et al.* (2004) 'Nitric oxide represses the Arabidopsis floral transition.', *Science (New York, N.Y.)*, 305(5692), pp. 1968–71.

Helliwell, C. A. *et al.* (2006) 'The Arabidopsis FLC protein interacts directly in vivo with SOC1 and FT chromatin and is part of a high-molecular-weight protein complex.', *The Plant journal : for cell and molecular biology*, 46(2), pp. 183–92.

Heo, J. B. and Sung, S. (2011) 'Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA.', *Science (New York, N.Y.)*, 331(6013), pp. 76–9.

Hepworth, S. R. *et al.* (2002) 'Antagonistic regulation of flowering-time gene SOC1 by CONSTANS and FLC via separate promoter motifs', *The EMBO Journal*, 21(16), pp. 4327–4337.

Hill, K., Wang, H. and Perry, S. E. (2008) 'A transcriptional repression motif in the MADS factor AGL15 is involved in recruitment of histone deacetylase complex components', *The Plant Journal*, 53(1), pp. 172–185.

Hill, K., Wang, H. and Perry, S. E. (2008) 'A transcriptional repression motif in the MADS factor AGL15 is involved in recruitment of histone deacetylase complex components.', *The Plant journal : for cell and molecular biology*, 53(1), pp. 172–85.

Hisamatsu, T. *et al.* (2005) 'The involvement of gibberellin 20-oxidase genes in phytochrome-regulated petiole elongation of Arabidopsis.', *Plant physiology*. American Society of Plant Biologists, 138(2), pp. 1106–16.

Hisamatsu, T. and King, R. W. (2008) 'The nature of floral signals in Arabidopsis. II. Roles for FLOWERING LOCUS T (FT) and gibberellin.', *Journal of experimental botany*. Oxford University Press, 59(14), pp. 3821–9.

Honma, T. and Goto, K. (2001) 'Complexes of MADS-box proteins are sufficient to convert leaves into floral organs.', *Nature*, 409(6819), pp. 525–9.

Hu, Z. *et al.* (2015) 'Histone acetyltransferase GCN5 is essential for heat stress-responsive gene activation and thermotolerance in Arabidopsis', *The Plant Journal*, 84(6), pp. 1178–1191.

Huq, E. and Quail, P. H. (2002) 'PIF4, a phytochrome-interacting bHLH factor, functions as a negative regulator of phytochrome B signaling in Arabidopsis.', *The EMBO journal*, 21(10), pp. 2441–50.

Hwan Lee, J. *et al.* (2014) 'Post-translational regulation of short vegetative phase as a major mechanism for thermoregulation of flowering.', *Plant signaling & behavior*. Taylor & Francis, 9(3), p. e28193.

Immink, R. G. *et al.* (2009) 'SEPALLATA3: the "glue" for MADS box transcription factor complex formation', *Genome Biology*. BioMed Central, 10(2), p. R24.

Immink, R. G. H., Kaufmann, K. and Angenent, G. C. (2010) 'The "ABC" of MADS domain protein behaviour and interactions', *Seminars in Cell and Developmental Biology*, 21(1), pp. 87–93.

Jack, T. (2004) 'Molecular and genetic mechanisms of floral control [Review]', *Plant Cell*, 16(Suppl S), pp. S1–S17.

Jang, S., Torti, S. and Coupland, G. (2009) 'Genetic and spatial interactions between FT, TSF and SVP during the early stages of floral induction in Arabidopsis', *Plant Journal*, 60(4), pp. 614–625.

Jiang, D. *et al.* (2011) 'Arabidopsis COMPASS-like complexes mediate histone H3 lysine-4 trimethylation to control floral transition and plant development.', *PLoS genetics*, 7(3), p. e1001330.

Jing, Y. *et al.* (2013) '*Arabidopsis* Chromatin Remodeling Factor PICKLE Interacts with Transcription Factor HY5 to Regulate Hypocotyl Cell Elongation', *The Plant Cell*, 25(1), pp. 242–256.

Jones, A. L. and Sung, S. (2014) 'Mechanisms underlying epigenetic regulation in Arabidopsis thaliana.', *Integrative and comparative biology*, 54(1), pp. 61–7.

Kempin, S. A., Savidge, B. and Yanofsky, M. F. (1995) 'Molecular basis of the cauliflower phenotype in Arabidopsis', *Science*, 267(5197), p. 522 LP-525.

Kim, S.-G., Kim, S.-Y. and Park, C.-M. (2007) 'A membrane-associated NAC transcription factor regulates salt-responsive flowering via FLOWERING LOCUS T in Arabidopsis.', *Planta*, 226(3), pp. 647–54.

Kim, W.-Y. *et al.* (2013) 'Release of SOS2 kinase from sequestration with GIGANTEA determines salt tolerance in Arabidopsis.', *Nature communications*, 4, p. 1352.

Köhler, C. *et al.* (2003) 'Arabidopsis MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development.', *The EMBO journal*. European Molecular Biology Organization, 22(18), pp. 4804–14.

Koornneef, M., Hanhart, C. J. and van der Veen, J. H. (1991) 'A genetic and physiological analysis of late flowering mutants in Arabidopsis thaliana.', *Molecular & general genetics : MGG*, 229(1), pp. 57–66.

Koussevitzky, S. *et al.* (2007) 'Signals from chloroplasts converge to regulate nuclear gene expression.', *Science (New York, N.Y.)*, 316(5825), pp. 715–9.

Kumimoto, R. W. *et al.* (2010) 'NF-YC3, NF-YC4 and NF-YC9 are required for CONSTANS-mediated, photoperiod-dependent flowering in Arabidopsis thaliana.', *The Plant journal : for cell and molecular biology*, 63(3), pp. 379–91.

Laux, T. *et al.* (1996) 'The WUSCHEL gene is required for shoot and floral meristem integrity in Arabidopsis.', *Development (Cambridge, England)*, 122, pp. 87–96.

Lazaro, A. *et al.* (2015) 'Red Light-Mediated Degradation of CONSTANS by the E3 Ubiquitin Ligase HOS1 Regulates Photoperiodic Flowering in Arabidopsis.', *The Plant cell*. American Society of Plant Biologists, 27(9), pp. 2437–54.

Lee, I. *et al.* (1994) 'Isolation of LUMINIDEPENDENS: a gene involved in the control of flowering time in Arabidopsis.', *The Plant cell*, 6(1), pp. 75–83.

Lee, J. *et al.* (2008) 'SOC1 translocated to the nucleus by interaction with AGL24 directly regulates leafy.', *The Plant journal : for cell and molecular biology*, 55(5), pp. 832–43.

Lee, J. H. *et al.* (2007) 'Role of SVP in the control of flowering time by ambient temperature in Arabidopsis', pp. 397–402.

Lee, J. H. *et al.* (2013) 'Regulation of temperature-responsive flowering by MADS-box transcription factor repressors.', *Science (New York, N.Y.)*, 342(6158), pp. 628–32.

Leibfried, A. *et al.* (2005) 'WUSCHEL controls meristem function by direct regulation of cytokinin-inducible response regulators.', *Nature*, 438(7071), pp. 1172–1175.

Li, D. *et al.* (2008) 'A Repressor Complex Governs the Integration of Flowering Signals in Arabidopsis', *Developmental Cell*, 15(1), pp. 110–120.

Li, K. *et al.* (2007) 'GA signaling and CO/FT regulatory module mediate salt-induced late flowering in Arabidopsis thaliana', *Plant Growth Regulation*. Springer Netherlands, 53(3), pp. 195–206.

Li, L. *et al.* (2016) 'Flowering responses to light and temperature', *Science China Life Sciences*, 59(4), pp. 403–408.

Lim, M.-H. *et al.* (2004) 'A new Arabidopsis gene, FLK, encodes an RNA binding protein with K homology motifs and regulates flowering time via FLOWERING LOCUS C.', *The Plant cell*, 16(3), pp. 731–40.

Liu, B. *et al.* (2011) 'Arabidopsis cryptochrome 1 interacts with SPA1 to suppress COP1 activity in response to blue light.', *Genes & development*. Cold Spring Harbor Laboratory Press, 25(10), pp. 1029–34.

Liu, C. *et al.* (2008) 'Direct interaction of AGL24 and SOC1 integrates flowering signals in Arabidopsis.', *Development (Cambridge, England)*, 135(8), pp. 1481–91.

Liu, C. *et al.* (2009) 'Regulation of Floral Patterning by Flowering Time Genes', *Developmental Cell*, 16(5), pp. 711–722.

Liu, T. *et al.* (2013) 'Nitrate or NaCl regulates floral induction in Arabidopsis thaliana', *Biologia*. SP Versita, 68(2), pp. 215–222.

Liu, Z. *et al.* (2009) 'Molecular and reverse genetic characterization of *NUCLEOSOME ASSEMBLY PROTEIN1* (*NAP1*) genes unravels their function in transcription and nucleotide excision repair in *Arabidopsis thaliana*', *The Plant Journal*, 59(1), pp. 27–38.

Lu, F. *et al.* (2011) 'Arabidopsis REF6 is a histone H3 lysine 27 demethylase', *Nature Genetics*. Nature Research, 43(7), pp. 715–719.

Lu, L. *et al.* (2015) 'High-resolution mapping of H4K16 and H3K23 acetylation reveals conserved and unique distribution patterns in *Arabidopsis* and rice', *Epigenetics*, 10(11), pp. 1044–1053.

Malapeira, J. and Mas, P. (2013) 'A chromatin-dependent mechanism regulates gene expression at the core of the Arabidopsis circadian clock', *Plant Signal Behav*, 8(5), p. e24079.

Marquardt, S. *et al.* (2006) 'Additional targets of the Arabidopsis autonomous pathway members, FCA and FY.', *Journal of experimental botany*, 57(13), pp. 3379–3386.

Masiero, S. *et al.* (2011) 'The emerging importance of type I MADS box transcription factors for plant reproduction', *Plant Cell*, 23(3), pp. 865–872.

McGonigle, B., Bouhidel, K. and Irish, V. F. (1996) 'Nuclear localization of the Arabidopsis APETALA3 and PISTILLATA homeotic gene products depends on their simultaneous expression.', *Genes & development*. Cold Spring Harbor Laboratory Press, 10(14), pp. 1812–21.

Messenguy, F. and Dubois, E. (2003) 'Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development.', *Gene*, 316, pp. 1–21.

Mizzotti, C. *et al.* (2014) 'SEEDSTICK is a Master Regulator of Development and Metabolism in the Arabidopsis Seed Coat', *PLoS Genetics*. Edited by R. G. Franks. Public Library of Science, 10(12), p. e1004856.

Murase, K. *et al.* (2008) 'Gibberellin-induced DELLA recognition by the gibberellin receptor GID1', *Nature*. Nature Publishing Group, 456(7221), pp. 459–463.

Mutasa-Göttgens, E. and Hedden, P. (2009) 'Gibberellin as a factor in floral regulatory networks', *Journal of Experimental Botany*, 60(7), pp. 1979–1989.

Norman, C. *et al.* (1988) 'Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element.', *Cell*, 55(6), pp. 989–1003.

Ohto, M. *et al.* (2001) 'Effects of sugar on vegetative development and floral transition in Arabidopsis.', *Plant physiology*. American Society of Plant Biologists, 127(1), pp. 252–61.

Okada, T., Singh, M. B. and Bhalla, P. L. (2006) 'Histone H3 variants in male gametic cells of lily and H3 methylation in mature pollen', *Plant Molecular Biology*, 62(4–5), pp. 503–512.

Olinares, P. D. B., Ponnala, L. and van Wijk, K. J. (2010) 'Megadalton complexes in the chloroplast stroma of Arabidopsis thaliana characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering.', *Molecular & cellular proteomics : MCP*. American Society for Biochemistry and Molecular Biology, 9(7), pp. 1594–615.

Orlando, V. (2016) ‘Polycomb, Epigenomes, and Control of Cell Identity’, *Cell*, 112(5), pp. 599–606.

Pajoro, A. *et al.* (2014) ‘Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development.’, *Genome biology*, 15(3), p. R41.

Parenicová, L. *et al.* (2003) ‘Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world.’, *The Plant cell*, 15(7), pp. 1538–1551.

Passmore, S. *et al.* (1988) ‘Saccharomyces cerevisiae protein involved in plasmid maintenance is necessary for mating of MAT alpha cells.’, *Journal of molecular biology*, 204(3), pp. 593–606.

Pendle, A. F. (2004) ‘Proteomic Analysis of the Arabidopsis Nucleolus Suggests Novel Nucleolar Functions’, *Molecular Biology of the Cell*, 16(1), pp. 260–269.

Pin, P. A. and Nilsson, O. (2012) ‘The multifaceted roles of FLOWERING LOCUS T in plant development’, *Plant, Cell and Environment*, 35(10), pp. 1742–1755.

Pogson, B. J., Ganguly, D. and Albrecht-Borth, V. (2015) ‘Insights into chloroplast biogenesis and development’, *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1847(9), pp. 1017–1024.

Ratcliffe, O. J. *et al.* (2001) ‘Regulation of flowering in Arabidopsis by an FLC homologue.’, *Plant physiology*, 126(1), pp. 122–32.

Riboni, M. *et al.* (2013) ‘GIGANTEA enables drought escape response via abscisic acid-dependent activation of the florigens and SUPPRESSOR OF OVEREXPRESSION OF CONSTANS.’, *Plant physiology*, 162(3), pp. 1706–19.

Riboni, M. *et al.* (2014) ‘Environmental stress and flowering time: the photoperiodic connection.’, *Plant signaling & behavior*, 9(7), p. e29036.

Riechmann, J. L., Wang, M. and Meyerowitz, E. M. (1996) ‘DNA-binding properties of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA and AGAMOUS.’, *Nucleic acids research*. Oxford University Press, 24(16), pp. 3134–41.

Rizzini, L. *et al.* (2011) ‘Perception of UV-B by the Arabidopsis UVR8 Protein’, *Science*, 332(6025).

Ryu, J. Y. *et al.* (2014) ‘The Arabidopsis Floral Repressor BFT Delays Flowering by Competing with FT for FD Binding under High Salinity’, *Molecular Plant*, 7(2), pp. 377–387.

Sanda, S. L. and Amasino, R. M. (1996) ‘Ecotype-Specific Expression of a Flowering Mutant Phenotype in Arabidopsis thaliana.’, *Plant physiology*, 111(2), pp. 641–644.

Sawa, M. *et al.* (2007) ‘FKF1 and GIGANTEA complex formation is required for day-length measurement in Arabidopsis.’, *Science (New York, N.Y.)*, 318(5848), pp. 261–5.

Schultz, E. A. and Haughn, G. W. (1991) 'LEAFY, a Homeotic Gene That Regulates Inflorescence Development in Arabidopsis.', *The Plant cell*, 3(8), pp. 771–781.

Schwarz-Sommer, Z. *et al.* (1990) 'Genetic Control of Flower Development by Homeotic Genes in Antirrhinum majus.', *Science (New York, N.Y.)*, 250(4983), pp. 931–6.

Schwarz-Sommer, Z. *et al.* (1992) 'Characterization of the Antirrhinum floral homeotic MADS-box gene deficiens: evidence for DNA binding and autoregulation of its persistent expression throughout flower development.', *The EMBO journal*. European Molecular Biology Organization, 11(1), pp. 251–63.

Scortecci, K., Michaels, S. D. and Amasino, R. M. (2003) 'Genetic interactions between FLM and other flowering-time genes in Arabidopsis thaliana.', *Plant molecular biology*, 52(5), pp. 915–22.

Searle, I. *et al.* (2006) 'The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis.', *Genes & development*, 20(7), pp. 898–912.

Severing, E. I. *et al.* (2012) 'Predicting the impact of alternative splicing on plant MADS domain protein function', *PLoS ONE*, 7(1).

Shannon, S. and Meeks-Wagner, D. (1991) 'A Mutation in the Arabidopsis TFL1 Gene Affects Inflorescence Meristem Development.', *The Plant cell*, 3(September), pp. 877–892.

Sheldon, C. C. *et al.* (2009) 'Mechanisms of gene repression by vernalization in Arabidopsis.', *The Plant journal : for cell and molecular biology*, 59(3), pp. 488–98.

Simpson, G. G. (2004) 'The autonomous pathway: Epigenetic and post-transcriptional gene regulation in the control of Arabidopsis flowering time', *Current Opinion in Plant Biology*, 7(5), pp. 570–574.

Smaczniak, C. *et al.* (2012) 'Characterization of {MADS}-domain transcription factor complexes in \emph{Arabidopsis} flower development', *Proceedings of the National Academy of Sciences*, 109(5), pp. 1560–1565.

Smaczniak, C. *et al.* (2012) 'Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies', *Development*, 139(17), pp. 3081–3098.

Smaczniak, C. *et al.* (2012) 'Proteomics-based identification of low-abundance signaling and regulatory protein complexes in native plant tissues.', *Nature protocols*, 7(12), pp. 2144–58.

Sommer, H. *et al.* (1990) 'Deficiens, a homeotic gene involved in the control of flower morphogenesis in Antirrhinum majus: the protein shows homology to transcription factors.', *The EMBO journal*. European Molecular Biology Organization, 9(3), pp. 605–13.

Son, O. *et al.* (2015) 'Identification of nucleosome assembly protein 1 (NAP1) as an interacting partner of plant ribosomal protein S6 (RPS6) and a positive regulator of rDNA transcription', *Biochemical and Biophysical Research Communications*, 465(2), pp. 200–205.

Steffen, J. G. *et al.* (2008) 'AGL61 interacts with AGL80 and is required for central cell development in Arabidopsis.', *Plant physiology*. American Society of Plant Biologists, 148(1), pp. 259–68.

Sun, X. *et al.* (2011) 'A chloroplast envelope-bound PHD transcription factor mediates chloroplast signals to the nucleus', *Nature Communications*. Nature Publishing Group, 2, p. 477.

Susek, R. E., Ausubel, F. M. and Chory, J. (1993) 'Signal transduction mutants of Arabidopsis uncouple nuclear CAB and RBCS gene expression from chloroplast development.', *Cell*, 74(5), pp. 787–99.

Swiezewski, S. *et al.* (2009) 'Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target.', *Nature*, 462(7274), pp. 799–802.

Tadini, L. *et al.* (2016) 'GUN1 Controls Accumulation of the Plastid Ribosomal Protein S1 at the Protein Level and Interacts with Proteins Involved in Plastid Protein Homeostasis.', *Plant physiology*. American Society of Plant Biologists, 170(3), pp. 1817–30.

Tao, Z. *et al.* (2012) 'Genome-wide identification of SOC1 and SVP targets during the floral transition in Arabidopsis', *Plant Journal*, 70(4), pp. 549–561.

Theißen, G., Melzer, R. and Rümpler, F. (2016) 'MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution', *Development*, 143(18).

Tiwari, S. B. *et al.* (2010) 'The flowering time regulator CONSTANS is recruited to the FLOWERING LOCUS T promoter via a unique cis-element.', *The New phytologist*, 187(1), pp. 57–66.

Valentim, F. L. *et al.* (2015) 'A quantitative and dynamic model of the arabidopsis flowering time gene regulatory network', *PLoS ONE*, 10(2), pp. 1–18.

Valverde, F. *et al.* (2004) 'Photoreceptor regulation of CONSTANS protein in photoperiodic flowering.', *Science (New York, N.Y.)*, 303(5660), pp. 1003–6.

Varotto, C. *et al.* (2000) 'Disruption of the Arabidopsis photosystem I gene psaE1 affects photosynthesis and impairs growth.', *The Plant journal : for cell and molecular biology*, 22(2), pp. 115–24.

Verelst, W., Saedler, H. and Münster, T. (2007) 'MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific Arabidopsis promoters.', *Plant physiology*. American Society of Plant Biologists, 143(1), pp. 447–60.

Wang, J.-W., Czech, B. and Weigel, D. (2009) 'miR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana.', *Cell*, 138(4), pp. 738–49.

Wang, J. W. (2014) 'Regulation of flowering time by the miR156-mediated age pathway', *Journal of Experimental Botany*, 65(17), pp. 4723–4730.

Wang, T. *et al.* (2016) 'Histone acetyltransferase general control non-repressed protein 5 (GCN5) affects the fatty acid composition of *Arabidopsis thaliana* seeds by acetylating *fatty acid desaturase3* (*FAD3*)', *The Plant Journal*.

Wang, Y. *et al.* (2014) 'Photoperiodic control of the floral transition through a distinct polycomb repressive complex', *Developmental Cell*. Elsevier Inc., 28(6), pp. 727–736.

West, A. G. and Sharrocks, A. D. (1999) 'MADS-box transcription factors adopt alternative mechanisms for bending DNA.', *Journal of molecular biology*, 286(5), pp. 1311–23.

Wilson, R. N., Heckman, J. W. and Somerville, C. R. (1992) 'Gibberellin Is Required for Flowering in Arabidopsis thaliana under Short Days.', *Plant physiology*, 100(1), pp. 403–8.

Wu, G. *et al.* (2009) 'The sequential action of miR156 and miR172 regulates developmental timing in Arabidopsis.', *Cell*, 138(4), pp. 750–9.

Xing, J. *et al.* (2015) 'GENERAL CONTROL NONREPRESSED PROTEIN5-Mediated Histone Acetylation of *FERRIC REDUCTASE DEFECTIVE3* Contributes to Iron Homeostasis in Arabidopsis', *Plant Physiology*, 168(4), pp. 1309–1320.

Yamaguchi, S. (2008) 'Gibberellin metabolism and its regulation.', *Annual review of plant biology*, 59, pp. 225–51.

Yan, Y. *et al.* (2014) 'A MYB-Domain Protein EFM Mediates Flowering Responses to Environmental Cues in Arabidopsis', *Developmental Cell*, 30(4), pp. 437–448.

Yang, L. *et al.* (2013) 'Sugar promotes vegetative phase change in Arabidopsis thaliana by repressing the expression of MIR156A and MIR156C.', *eLife*, 2, p. e00260.

Yang, Y., Fanning, L. and Jack, T. (2003) 'The K domain mediates heterodimerization of the Arabidopsis floral organ identity proteins, APETALA3 and PISTILLATA.', *The Plant journal : for cell and molecular biology*, 33(1), pp. 47–59.

Yang, Y. and Jack, T. (2004) 'Defining subdomains of the K domain important for protein-protein interactions of plant MADS proteins.', *Plant molecular biology*, 55(1), pp. 45–59.

Yanofsky, M. F. *et al.* (1990) 'The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors.', *Nature*, 346(6279), pp. 35–9.

Yanofsky, M. F. *et al.* (2000) 'SHATTERPROOF MADS-box genes control seed dispersal in Arabidopsis', *Nature*. Nature Publishing Group, 404(6779), pp. 766–770.

Yanovsky, M. J. and Kay, S. A. (2002) 'Molecular basis of seasonal time measurement in Arabidopsis.', *Nature*, 419(6904), pp. 308–12.

Yao, X. *et al.* (2013) 'SDG2-Mediated H3K4 Methylation Is Required for Proper Arabidopsis Root Growth and Development', *PLoS ONE*, 8(2), pp. 1–11.

Yun, J. Y. *et al.* (2012) 'Arabidopsis trithorax-related3/set domain group2 is required for the winter-annual habit of arabidopsis thaliana', *Plant and Cell Physiology*, 53(5), pp. 834–846.

PART II:

PENTATRICO-PEPTIDE-REPEAT PROTEINS AND THE CHLOROPLAST-NUCLEUS COMMUNICATION

Flowering time in *Arabidopsis thaliana* has been previously shown to be dependent from the regulation of the two major florigen genes *FT* and *SOC1*, which in turn activate the gene cascade necessary to induce the expression of IM and FM identity genes to trigger the formation of flowers. Their expression is modulated by different external and internal cues, that plants perceive through different gene networks. In plants, a major environmental sensor capable to perceive external stimuli and to induce adaptive responses is the chloroplast. Through a nucleus-to-chloroplast and chloroplast-to-nucleus mutual communication, the anterograde and retrograde signaling pathways, the chloroplast can interact with the nucleus and modulate cell activities in response to different environmental stresses and stimuli, by regulating nuclear gene expression. The same communication mechanism has been recently found to be involved in high light-mediated flowering time, discovering an active role of chloroplasts in floral transition regulation. Feng et al. 2016, in fact, discovered the existence of a link between *FLC* transcriptional regulation and high light environmental conditions, mediated by the chloroplast-nucleus retrograde signaling pathway. In particular, *FLC* expression was shown to be repressed under oxidative stress conditions, thus accelerating flowering under non optimal growth conditions. The proposed model suggests the involvement of PHD TYPE TRANSCRIPTION FACTOR WITH TRANSMEMBRANE DOMAINS (PTM) protein, present on the outer chloroplast membrane, which in the presence of high light conditions can be hydrolyzed by, so far unknown, chloroplast proteases, that release the N-terminal PTM portion into the cytosol (Sun *et al.*, 2011). N-PTM, together with FVE protein, a deacethylase enzyme active also in the autonomous flowering pathway (Simpson, 2004; Jeon and Kim, 2011), can then move to the nucleus and repress *FLC* expression by removing the H3K4me3 and H3ac histone positive marks. Based on these new findings, it becomes more and more clear the importance of the chloroplast retrograde signaling as a major bridge between environmental stimuli and the nuclear gene expression machinery. Furthermore, it provides further evidences on the central role of histone modifications, such as H3K4me3, in the expression regulation of MADS-box transcription factors. Within this frame, during my Ph.D. studies, I was also involved in the elucidation of the molecular mechanisms at the basis of chloroplast-nucleus communication.

As stated in the Introduction section, Pentatrico-Peptide-Repeat (PPR) containing proteins, certainly, play a major role in exchanging information between chloroplasts and the nucleus (Barkan and Small, 2014; Tadini *et al.*, 2016). Among them, the PPR protein

GENOME UNCOUPLED 1 (GUN1), has been reported as a master regulator of chloroplast-to-nucleus retrograde signaling, and found to be part of the chloroplast nucleoids, the stromal-membrane-associated structure containing the chloroplast gene expression machinery (Koussevitzky *et al.*, 2007; Zhang *et al.*, 2015). Furthermore, portions of the GUN1 genomic sequence containing CArG boxes were found to be enriched in SVP ChiP-seq assays (Gregis et al. 2013), suggesting that GUN1 expression could be also regulated by the binding of SVP-containing protein complexes (Figure 18a). Knockout plants lacking the GUN1 protein, are altered in the retrograde signaling pathway, however the molecular mechanisms at the basis of GUN1-dependent signaling are still unknown. Below, in the section termed "Manuscript I" a review, that I authored, is reported. The manuscript proposes the multiple roles of GUN1 in chloroplast protein homeostasis, mainly based on GUN1-GFP CoIP-MS data, recently published by our lab (Tadini et al. 2016).

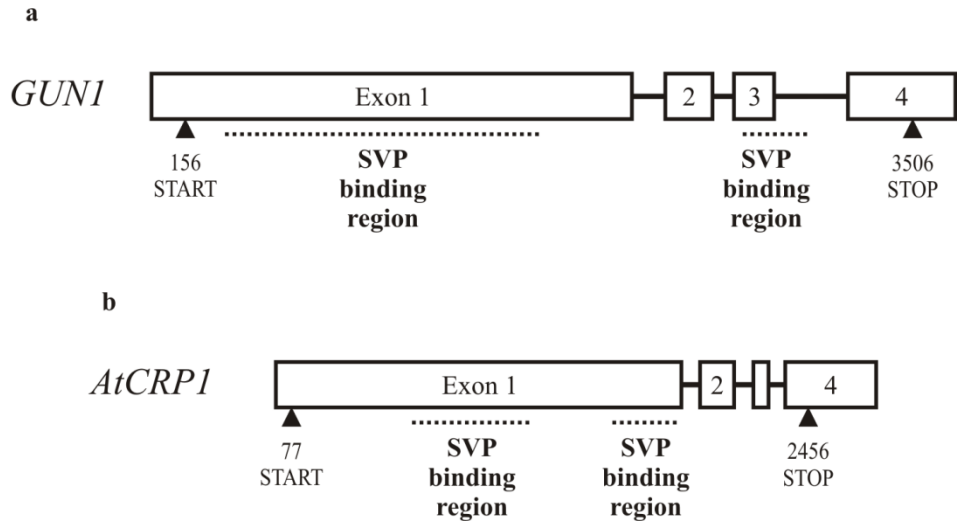

Figure 18: Schematic representation of *GUN1* and *AtCRP1* genomic loci, with regards to the relative SVP putative binding regions found by SVP ChIP-seq analysis (Gregis *et al.*, 2013). (a) *GUN1* has two SVP binding regions within the first exon and the third intron sequence. (b) *AtCRP1* shows two different SVP binding regions within its first exon. Exons and introns are represented as open blocks and black lines, respectively. Positions of the start and stop codons are indicated. SVP binding regions are reported in dashed lines.

A second Arabidopsis PPR protein, termed CHLOROPLAST RNA PROCESSING 1 (AtCRP1) was also the object of my Ph.D. research program. Like GUN1, AtCRP1 has been found to be part of megadalton protein complexes associated to the chloroplast nucleoids

(Olinares et al. 2010). Furthermore, two CArG box regions have been found within the first *AtCRP1* exon (see also Figure 18b), by SVP ChIP-seq analysis (Gregis *et al.*, 2013), meaning that *AtCRP1* expression can, also in this case, be regulated by SVP, and suggesting a possible involvement of this PPR protein in the chloroplast-dependent flowering time determination. A detailed description of AtCRP1 protein function is reported in the section termed "Manuscript II", that has been recently published (Ferrari *et al.*, 2017).

In conclusion, the data obtained in this part of the Ph.D. research program reveals a further link between chloroplasts and the timing of reproduction, and provides insight into how chloroplasts emerged as key players in plant growth and development after their integration into their host cells through endosymbiosis. Such signaling mechanisms may enable higher plants to more effectively adapt to the ever-changing environment and mitigate detrimental effects to fitness.

REFERENCES: PART II

Barkan, A. and Small, I. (2014) 'Pentatricopeptide Repeat Proteins in Plants PPR: pentatricopeptide repeat', *Annu. Rev. Plant Biol*, 65, pp. 415–42.

Feng, P. *et al.* (2016) 'Chloroplast retrograde signal regulates flowering', *Proceedings of the National Academy of Sciences*, (25), p. 201521599.

Ferrari, R. *et al.* (2017) 'CRP1 Protein: (dis)similarities between Arabidopsis thaliana and Zea mays', *Frontiers in Plant Science*, 8(February), pp. 1–18.

Gregis, V. *et al.* (2013) 'Identification of pathways directly regulated by SHORT VEGETATIVE PHASE during vegetative and reproductive development in Arabidopsis.', *Genome biology*. BioMed Central Ltd, 14(6), p. R56.

Jeon, J. and Kim, J. (2011) 'FVE, an Arabidopsis homologue of the retinoblastoma-associated protein that regulates flowering time and cold response, binds to chromatin as a large multiprotein complex', *Molecules and Cells*, 32(3), pp. 227–234.

Koussevitzky, S. *et al.* (2007) 'Signals from chloroplasts converge to regulate nuclear gene expression.', *Science (New York, N.Y.)*, 316(5825), pp. 715–9.

Olinares, P. D. B., Ponnala, L. and van Wijk, K. J. (2010) 'Megadalton complexes in the chloroplast stroma of Arabidopsis thaliana characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering.', *Molecular & cellular proteomics : MCP*. American Society for Biochemistry and Molecular Biology, 9(7), pp. 1594–615.

Simpson, G. G. (2004) 'The autonomous pathway: Epigenetic and post-transcriptional gene regulation in the control of Arabidopsis flowering time', *Current Opinion in Plant Biology*, 7(5), pp. 570–574.

Sun, X. *et al.* (2011) 'A chloroplast envelope-bound PHD transcription factor mediates chloroplast signals to the nucleus', *Nature Communications*. Nature Publishing Group, 2, p. 477.

Tadini, L. *et al.* (2016) 'GUN1 Controls Accumulation of the Plastid Ribosomal Protein S1 at the Protein Level and Interacts with Proteins Involved in Plastid Protein Homeostasis.', *Plant physiology*. American Society of Plant Biologists, 170(3), pp. 1817–30.

Zhang, Z.-W. *et al.* (2015) 'The roles of tetrapyrroles in plastid retrograde signaling and tolerance to environmental stresses', *Planta*, 242(6), pp. 1263–1276.

MANUSCRIPT I

frontiers in Plant Science

REVIEW
published: 22 September 2016
doi: 10.3389/fpls.2016.01427

GUN1, a Jack-Of-All-Trades in Chloroplast Protein Homeostasis and Signaling

Monica Colombo[1], *Luca Tadini*[2], *Carlotta Peracchio*[2], *Roberto Ferrari*[2] *and Paolo Pesaresi*[2]*

[1] Centro Ricerca e Innovazione, Fondazione Edmund Mach, San Michele all'Adige, Italy, [2] Dipartimento di Bioscienze, Università degli Studi di Milano, Milan, Italy

The *GENOMES UNCOUPLED 1 (GUN1)* gene has been reported to encode a chloroplast-localized pentatricopeptide-repeat protein, which acts to integrate multiple indicators of plastid developmental stage and altered plastid function, as part of chloroplast-to-nucleus retrograde communication. However, the molecular mechanisms underlying signal integration by GUN1 have remained elusive, up until the recent identification of a set of GUN1-interacting proteins, by co-immunoprecipitation and mass-spectrometric analyses, as well as protein–protein interaction assays. Here, we review the molecular functions of the different GUN1 partners and propose a major role for GUN1 as coordinator of chloroplast translation, protein import, and protein degradation. This regulatory role is implemented through proteins that, in most cases, are part of multimeric protein complexes and whose precise functions vary depending on their association states. Within this framework, GUN1 may act as a platform to promote specific functions by bringing the interacting enzymes into close proximity with their substrates, or may inhibit processes by sequestering particular pools of specific interactors. Furthermore, the interactions of GUN1 with enzymes of the tetrapyrrole biosynthesis (TPB) pathway support the involvement of tetrapyrroles as signaling molecules in retrograde communication.

Keywords: nucleoid, GUN1, protein homeostasis, retrograde signaling, biogenic control

OPEN ACCESS

Edited by:
Fiammetta Alagna, Research Council for Agriculture and Agricultural Economy, Italy

Reviewed by:
Lynn Richardson, Michigan State University, USA
Silvia Ramundo, University of California, San Francisco & Carnegie Institution for Science & Howard Hughes Medical Institute, USA

***Correspondence:**
Paolo Pesaresi
paolo.pesaresi@unimi.it

Specialty section:
This article was submitted to Plant Cell Biology, a section of the journal Frontiers in Plant Science

Received: *21 July 2016*
Accepted: *07 September 2016*
Published: *22 September 2016*

Citation:
Colombo M, Tadini L, Peracchio C, Ferrari R and Pesaresi P (2016) GUN1, a Jack-Of-All-Trades in Chloroplast Protein Homeostasis and Signaling. Front. Plant Sci. 7:1427. doi: 10.3389/fpls.2016.01427

INTRODUCTION

Upon illumination, proplastids differentiate into functional chloroplasts in developing photosynthetic tissues of cotyledons, leaves, and stems (Jarvis and López-Juez, 2013). Chloroplast biogenesis also occurs during the growth of young green tissues, as cells expand and mature chloroplasts undergo division by binary fission (Okazaki et al., 2010). This process is characterized macroscopically by rapid greening of the young chloroplast and microscopically by the concomitant formation of thylakoid membranes and the reorganization of nucleoids, i.e., DNA-containing structures without defined boundaries, which differ in number, size, and distribution within plastids at different developmental stages, and harbor the plastid gene expression (PGE) machinery (Pfalz and Pfannschmidt, 2013; Melonek et al., 2016).

At the molecular level, this rather complex biogenic transition is achieved by cytosolic synthesis of chloroplast-targeted proteins, followed by import, assembly, folding, and degradation of unfolded/misfolded proteins (Jarvis and López-Juez, 2013). Indeed, the plastid genome itself (the

plastome) comprises fewer than 100 protein-coding genes, and the vast majority of the 2000–3000 proteins that make up the chloroplast proteome are encoded in the nucleus (Richly and Leister, 2004). In particular, precursor proteins carrying N-terminal transit peptides initially interact with two multiprotein complexes termed Translocon at the outer envelope membrane of chloroplasts (TOC) and Translocon at the inner envelope membrane of chloroplasts (TIC), which facilitate their active transport through the chloroplast envelope, powered by an ATP import motor, consisting of the stromal heat-shock protein 93 (Hsp93), heat-shock protein 70 (Hsp70), and heat-shock protein 90 (Hsp90; Flores-Perez and Jarvis, 2013; Inoue et al., 2013; Shi and Theg, 2013a,b). Upon translocation, proteins are exposed to different proteolytic systems of prokaryotic origin, which are responsible for protein maturation, control of protein abundance, and removal of either misfolded or damaged components. Among these, the stromal protease Clp is a multimeric complex made of chaperones and serine protease subunits, which serve general housekeeping functions. In contrast, the thylakoid-associated FtsH (Filamentous temperature sensitive H) proteases are zinc-containing metalloendopeptidases that have both chaperone and proteolytic functions, and participate in the Photosystem II repair cycle, together with the DEG serine proteases (Kato and Sakamoto, 2010; Van Wijk, 2015).

Besides translation and post-translational processes, chloroplast biogenesis also requires transcriptional coordination of thousands of nuclear genes with the expression of the comparatively few plastid genes in order to meet the needs of the developing chloroplast (Chan et al., 2016; Kleine and Leister, 2016). This is achieved through extensive exchange of information between plastids and the nucleus, for instance, via biogenic retrograde signaling—a system in which developmentally relevant stimuli in plastids induce the accumulation of specific signaling molecules that relay information to the nucleus, and in turn adjust the expression of nuclear genes to the needs of the plastids (Pogson et al., 2008; Woodson and Chory, 2008; Chan et al., 2016).

During the last 30 years, experiments with the carotenoid biosynthesis inhibitor norfluorazon (NF) and the inhibitor of plastid translation lincomycin (LIN), each of which arrests chloroplast development at the proplastid stage and represses the expression of photosynthesis-associated nuclear genes (PhANGs), have provided insights into the plastid's biogenic retrograde pathways (Oelmüller and Mohr, 1986; Oelmüller et al., 1986).

Six *genome uncoupled* (*gun*) mutants have been characterized in *Arabidopsis thaliana* that fail to repress transcription of the nuclear gene *Lhcb1.2* after NF treatment, and are thus impaired in retrograde signaling (Susek et al., 1993; Mochizuki et al., 2001; Larkin et al., 2003; Koussevitzky et al., 2007; Adhikari et al., 2011; Woodson et al., 2011). Five of these genes, *GUN2-6*, were found to be involved in tetrapyrrole biosynthesis (TPB), whereas *GUN1*, which encodes a nucleoid-localized pentatricopeptide repeat protein (PPR), has been shown to have a role in PGE, and to act as an integrator of multiple retrograde signals, since *gun1* mutants are unique in exhibiting a *gun* phenotype in response to both norfluorazon and lincomycin (Gray et al., 2003; Koussevitzky et al., 2007). However, the exact molecular role of GUN1 remained enigmatic until the new insights provided by the recent identification of a set of GUN1-interacting proteins (Tadini et al., 2016; **Table 1**).

Based on the functions of these partners, GUN1 appears to take part in multiple processes essential for chloroplast biogenesis and maintenance of the chloroplast proteome. GUN1-mediated control of plastid ribosomal protein S1 (PRPS1) accumulation, together with co-immunoprecipitation (CoIP) of proteins involved in different steps of plastid translation, support the involvement of GUN1 in the regulation of plastid protein synthesis. Furthermore, the presence of several chaperones in the CoIP mixture suggests a role for GUN1 in the coordination of chloroplast protein import and protein degradation.

Intriguingly, several GUN1 interactors appear to accumulate to higher levels upon induction of the unfolded protein response (UPR) in *Chlamydomonas reinhardtii* chloroplasts, which is triggered upon conditional repression of the catalytic subunit of Clp protease (ClpP1; Ramundo et al., 2013; Ramundo and Rochaix, 2014; Rochaix and Ramundo, 2015). This finding suggests the possible involvement of GUN1 in the UPR signaling pathway.

In this review, we describe the functional roles of the different GUN1 protein partners and propose some testable hypotheses that should clarify the molecular role of GUN1 in chloroplast biogenesis and chloroplast protein homeostasis.

GUN1 IS FOUND IN PLASTID NUCLEOIDS AND INTERACTS WITH THE TRANSCRIPTIONAL MACHINERY

GUN1 encodes a member of PPR-containing protein family, which has a small MutS-related (SMR) domain at the C-terminal end and a plastid targeting signal sequence at its N terminus. PPR motifs have been shown to mediate interactions with nucleic acids, and the SMR domain is found in proteins that act in DNA repair and recombination. However, *in vivo* RNA and DNA immunoprecipitation on chip (NIP-chip), as well as one-hybrid assays, have failed to detect any stable interaction of GUN1 with nucleic acids (Tadini et al., 2016), in contrast to a previous report, in which a GUN1 fragment containing both the PPR and SMR domains was shown to bind DNA *in vitro* (Koussevitzky et al., 2007). Nevertheless, GUN1 appears to be associated with nucleoids in the chloroplast, and more specifically with the domain of active plastid transcription, as shown by the relatively large and distinct foci of a fluorescent GUN1-YFP (Yellow Fluorescence Protein) chimera that co-localize with a Plastid Transcriptionally Active Chromosome 2-Cyan Fluorescence Protein (pTAC2-CFP) fusion in chloroplasts of mesophyll cells (Koussevitzky et al., 2007). However, although the repertoire of nucleoid-associated proteins so far identified is quite extensive, the GUN1 protein is not listed in any of the chloroplast or nucleoid/pTAC proteomes published to date (for a review see Melonek et al., 2016), most probably because it accumulates in very small amounts at specific developmental stages or under

TABLE 1 | GUN1 interactors together with their functions and impacts on plant development.

Designation	AGI code	Mutant phenotype[a]	Molecular function/Defect	Nucleoid subunit[b]	Identification assay[c]	References
TRANSCRIPTION AND RNA METABOLISM						
pTAC6/PAP8	AT1G21600	Albino	Low PEP activity	+	CoIP-MS	Pfalz et al., 2006; Steiner et al., 2011; Pfalz and Pfannschmidt, 2013
RH3/EMB1138	AT5G26742	Embryo lethal	RNA splicing of group II introns, assembly of the 50S ribosomal particle	+	CoIP-MS	Asakura et al., 2012; Majeran et al., 2012
AtPPR_3g49240/ EMB1796	AT3G49240	Embryo lethal	n.d.	+	CoIP-MS	Cushing et al., 2005; Majeran et al., 2012
TRANSLATION						
rpl2	ATCG00830	n.d.	Promotes translation initiation	+	CoIP-MS	Manuell et al., 2007; Melonek et al., 2016
rps3	ATCG00800	Essential for cell survival in tobacco	Promotes translation initiation	+	CoIP-MS	Manuell et al., 2007; Fleischmann et al., 2011; Melonek et al., 2016
rps4	ATCG00380	Essential for cell survival in tobacco	Involved in the assembly of the 30S ribosomal particle; binds to16S rRNA	+	CoIP-MS	Rogalski et al., 2008; Shoji et al., 2011; Melonek et al., 2016
PRPL10/ EMB3136	AT5G13510	Embryo lethal	Part of the L12 stalk and required for translation, since it recruits auxiliary translation factors such as cpIF2	–	CoIP-MS	Baba et al., 2006; Bryant et al., 2011; Shoji et al., 2011; Pfalz and Pfannschmidt, 2013
PRPS1	AT5G30510	n.d.	Promotes translation initiation	–	Y2H; BiFC	Manuell et al., 2007; Shoji et al., 2011; Tadini et al., 2016
cpIF2/FUG1	AT1G17220	Embryo lethal	Promotes translation initiation; leaky mutant alleles suppress leaf variegation in *var* mutants	–	CoIP-MS	Miura et al., 2007
PROTEIN IMPORT, PROTEIN FOLDING, AND PROTEIN UNFOLDING/DEGRADATION						
Hsp93-III/ClpC2	AT3G48870	Single mutant identical to WT; *hsp93-III hsp93-V* double mutant is embryo lethal	Cooperates with Tic110 and Tic40 in chloroplast protein import; chaperone in the Clp protease complex	–	CoIP-MS	Inaba et al., 2003; Kovacheva et al., 2005; Chou et al., 2006; Sakamoto, 2006; Kovacheva et al., 2007; Van Wijk, 2015
Hsp93-V/ClpC1	At5g50920	Single mutant exhibits a chlorotic phenotype; *hsp93-III hsp93-V* double mutant is embryo lethal	Cooperates with Tic110 and Tic40 in chloroplast protein import; chaperone in the Clp protease complex	+	CoIP-MS	Inaba et al., 2003; Kovacheva et al., 2005; Chou et al., 2006; Sakamoto, 2006; Kovacheva et al., 2007; Van Wijk, 2015; Melonek et al., 2016
Hsp70-1	AT4G24280	Single mutant exhibits variegated cotyledons, malformed leaves, growth retardation and impaired root growth; *hsp70-1 hsp70-2* double mutant is lethal	Involved in chloroplast protein import, folding and onward guidance of newly imported polypeptide chains	+	CoIP-MS	Su and Li, 2008; Shi and Theg, 2010; Su and Li, 2010; Liu et al., 2014; Melonek et al., 2016
Hsp70-2	AT5G49910	Single mutant identical to WT; *hsp70-1 hsp70-2* double mutant is lethal	Involved in chloroplast protein import, folding and onward guidance of newly imported polypeptide chains	–	CoIP-MS	Su and Li, 2008; Shi and Theg, 2010; Liu et al., 2014; Su and Li, 2010
ptCpn60α1	AT2G28000	Albino	Involved in folding and onward guidance of newly imported polypeptide chains; essential for plastid division in *A. thaliana*; involved in Rubisco and NdhH assembly	+	CoIP-MS	Gutteridge and Gatenby, 1995; Apuya et al., 2001; Suzuki et al., 2009; Peng et al., 2011; Flores-Perez and Jarvis, 2013; Melonek et al., 2016

(Continued)

TABLE 1 | Continued

Designation	AGI code	Mutant phenotype[a]	Molecular function/Defect	Nucleoid subunit[b]	Identification assay[c]	References
ptCpn60β1	AT1G55490	Leaves of the *len1* mutant have wrinkled and irregular surfaces and display lesions due to spontaneous cell death	Involved in folding and onward guidance of newly imported polypeptide chains; essential for plastid division in *A. thaliana*; involved in Rubisco and NdhH assembly	–	CoIP-MS	Gutteridge and Gatenby, 1995; Boston et al., 1996; Kessler and Blobel, 1996; Jackson-Constan et al., 2001; Ishikawa et al., 2003; Ishikawa, 2005; Suzuki et al., 2009; Flores-Perez and Jarvis, 2013
TPB ENZYMES						
CHLD	AT1G08520	Albino	Encodes the D subunit of the Mg-chelatase enzyme, involved in chlorophyll biosynthesis	–	Y2H; BiFC	Strand et al., 2003; Tanaka et al., 2011
PBGD	AT5G08280	n.d.	Porphobilinogen deaminase activity. Enzyme in the tetrapyrrole biosynthesis pathway	–	Y2H; BiFC	Tanaka et al., 2011
UROD2	AT2G40490	n.d.	Uroporphyrinogen decarboxylase activity; Enzyme in the tetrapyrrole biosynthesis pathway	–	Y2H; BiFC	Tanaka et al., 2011
FC1	AT5G26030	No visible phenotype; overexpression of the *FC1* gene is responsible for the *gun6* phenotype	Encodes ferrochelatase I, involved in heme biosynthesis	–	Y2H; BiFC	Tanaka et al., 2011; Woodson et al., 2011
DIVERSE FUNCTIONS						
rbcL	ATCG00490	Essential for photoautotrophy	Large subunit of Rubisco	+	CoIP-MS	Phinney and Thelen, 2005; Majeran et al., 2012; Huang et al., 2013
ATP-synthase β subunit	ATCG00480	Essential for photoautotrophy	Beta subunit of the thylakoid ATP synthase complex	+	CoIP-MS	Phinney and Thelen, 2005; Pfalz et al., 2006; Majeran et al., 2012; Melonek et al., 2012; Huang et al., 2013
RER4	AT5G12470	Mutant exhibits stunted growth, weak leaf reticulation and smaller mesophyll cells	Integral component of chloroplast outer and inner envelope membranes; possibly involved in retrograde signaling, supply of metabolites, control of ROS	–	CoIP-MS	Perez-Perez et al., 2013
2-Cys PrxA	AT3G11630	Mutant exhibits increased tolerance to photo-oxidative stress	Involved in peroxide detoxification in the chloroplast; functions as a redox sensor and chaperone; controls the conversion of Mg-protoporphyrin monomethyl ester into protochlorophyllide	–	CoIP-MS	Stenbaek et al., 2008; Rey et al., 2007; Pulido et al., 2010; König et al., 2013; Dietz, 2016

Note that proteins Q9SIP7 (AT2G31610) and Q42112 (AT3G09200) reported to be identified in coimmunoprecipitates of GUN1-GFP (Tadini et al., 2016) are not listed in this Table, since they have been described as subunits of cytosolic ribosomes. Furthermore, the protein Q9C5C2 (AT5G25980) has not been included, since it localizes to the tonoplast (Agee et al., 2010).

n.d., not determined.

[a]Phenotype of knock-out mutants is described.

[b]Protein already identified as part of chloroplast nucleoid by proteomic approaches.

[c]Assays used to identify the corresponding protein as a GUN1 interactor: coimmunoprecipitation followed by mass spectrometry (CoIP-MS), yeast two-hybrid (Y2H) analysis, and Bimolecular Fluorescence Complementation (BiFC).

particular physiological conditions. This inference is supported by CoIP experiments with a Green Fluorescence Protein (GUN1-GFP) fusion and subsequent mass spectrometry (MS), which identified several nucleoid subunits as interactors with GUN1 (Tadini et al., 2016; **Table 1**).

pTAC6 is among the GUN1 interactors, and it has been reported to interact directly with the plastid-encoded RNA polymerase (PEP), building together with pTAC2 and other polymerase-associated proteins (PAPs) the soluble RNA polymerase (sRNPase) complex (Pfalz et al., 2006), a central

component of nucleoids (Steiner et al., 2011; **Figure 1**). Intriguingly, pTAC6 (also known as PAP8) contains no known domain and exhibits no homologies that could provide hints as to its function in PGE (Steiner et al., 2011). However, functional genomics analyses have indicated that homozygous *pap* knockout lines develop white cotyledons, fail to accumulate chlorophyll even under low light intensities, and do not produce primary leaves unless they are cultivated on MS medium supplemented with sucrose (for a review, see Pfalz and Pfannschmidt, 2013). Furthermore, analyses of PGE in *pap* mutants revealed strong repression of the accumulation of PEP-dependent transcripts, whereas levels of nucleus-encoded RNA polymerase (NEP)-dependent transcripts were not depleted, while some were enhanced, indicating that pTAC6/PAP8 and the other PAP proteins are essential for the activity of PEP (see **Table 1**).

GUN1 CONTROLS PLASTID TRANSLATION AND RIBOSOME BIOGENESIS

GUN1 also interacts with several ribosomal subunits, such as the plastid-encoded ribosomal proteins L2, S3, and S4 (rpl2, rps3, and rps4) and the nucleus-encoded plastid ribosomal protein L10 (PRPL10; **Figure 1**). Furthermore, yeast two-hybrid and Bimolecular Fluorescence Complementation (BiFC) assays revealed a physical interaction between GUN1 and PRPS1 (Tadini et al., 2016). Ribosomal proteins have been reproducibly detected in nucleoid and pTAC proteomes (Melonek et al., 2016), further supporting the existence of a translational subdomain within the nucleoids, as proposed by Pfalz and Pfannschmidt (2013). The homologs of PRPL10, rpl2, PRPS1, rps3, and rps4 are essential components of the protein biosynthetic machinery in *Escherichia coli* (Baba et al., 2006; Shoji et al., 2011) and the indispensability of rps3 and rps4 has been also proven in tobacco plastids (Rogalski et al., 2008; Fleischmann et al., 2011). Furthermore, PRPL10 is annotated as EMBryo defective 3136 (EMB3136) in the SeedGenes Project database (http://www.seedgenes.org/), and in its absence Arabidopsis embryo development arrests at the globular stage (Bryant et al., 2011). Mutants devoid of PRPS1 have not been described. However, given the conservation of PRPS1 function in prokaryotes and chloroplasts, it can be confidently assumed that complete lack of PRPS1 is lethal in Arabidopsis.

Taking into consideration the function of these ribosomal proteins, it can be argued that their interaction with GUN1 has a dual purpose. On the one hand, GUN1 modulates protein synthesis by controlling the abundance of PRPS1, which, together with rps3 and rps2, has been reported to form the domain responsible for the interaction of the 30S ribosomal subunit with mRNA, promoting translation initiation (Manuell et al., 2007; Tadini et al., 2016). This role is supported further by the stable interaction of GUN1 with the chloroplast translation initiation

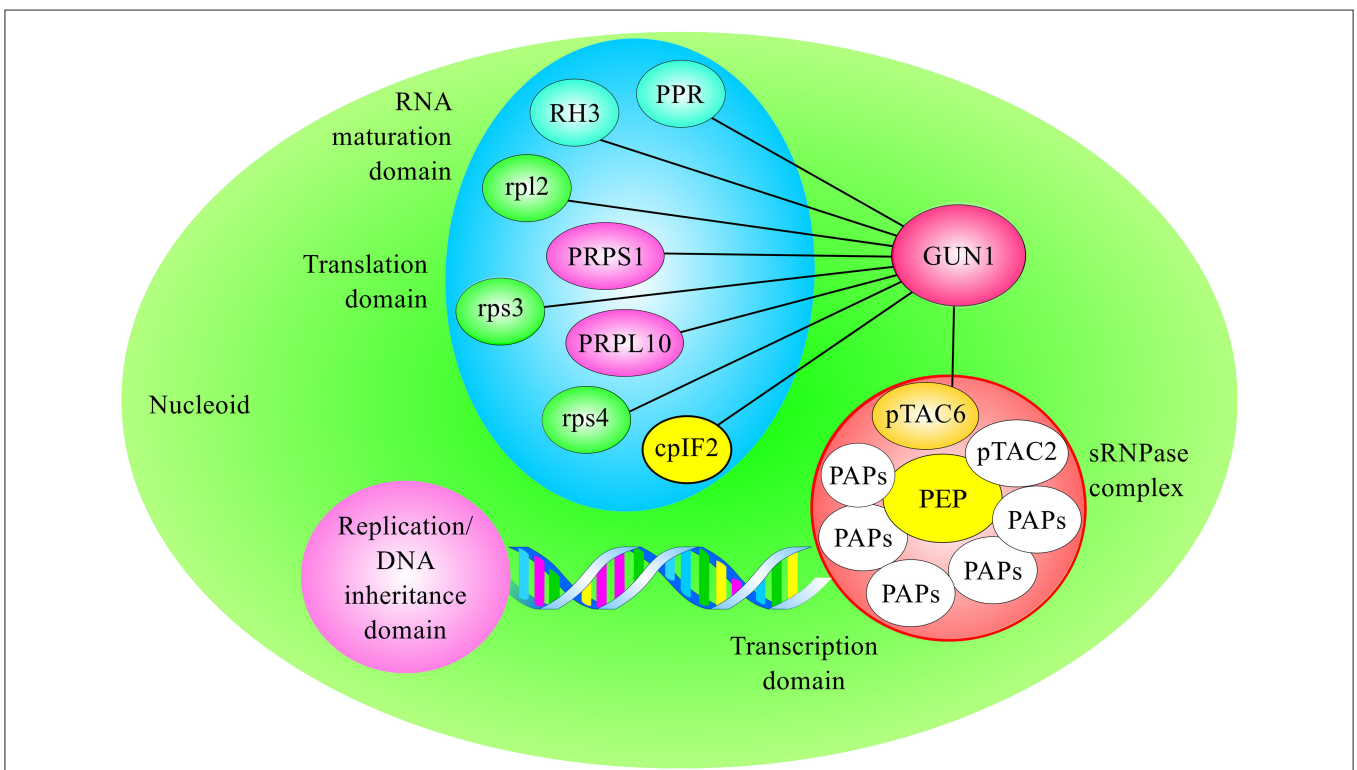

FIGURE 1 | Schematic overview of GUN1 protein interactors involved in gene transcription, ribosome biogenesis and plastid translation. The scheme takes into account the partition of nucleoids into functional subdomains proposed by Pfalz and Pfannschmidt (2013). PPR refers to AtPPR_3g49240, also known as EMB1796, as reported in **Table 1**.

factor 2 (cpIF2; Tadini et al., 2016), also known as FUG1, and reported to be essential for chloroplast biogenesis (Miura et al., 2007).

On the other hand, GUN1 seems to be involved in the process of ribosome biogenesis too, since nucleoid-associated ribosomes are thought to be in various stages of assembly, with several rRNA maturation steps occurring in a co-transcriptional and assembly-assisted manner, as in prokaryotic systems (Bohne, 2014). For instance, the DEAD-box-containing, ATP-dependent RNA helicase 3 (RH3), which has been functionally linked to the chloroplast nucleoid (Majeran et al., 2012), is among the proteins that interact with GUN1 (Tadini et al., 2016; see also **Figure 1** and **Table 1**). RH3 is directly involved in the splicing of group II introns in the *rpl2*, *trnA*, *trnI*, and *rps12* transcripts and could be coimmunoprecipitated with immature and mature 23S rRNA (Asakura et al., 2012). Furthermore, the PPR protein At3g49240 also known as AtPPR_3g49240, according to the PPR protein database (http://www.plantenergy.uwa.edu.au/applications/ppr/ppr.php), is also part of GUN1's interactors, and its maize ortholog, GRMZM2G074599_P01, has been identified in the chloroplast nucleoid (Majeran et al., 2012). The gene is annotated as embryo defective 1796 (EMB1796) in the SeedGenes database, since the complete lack of AtPPR_3g49240 leads to the arrest of embryonic development at the globular stage (Cushing et al., 2005), further supporting the essential role of GUN1 interactors in chloroplast biogenesis.

GUN1 AND THE IMPORT OF CHLOROPLAST PROTEINS

Almost a quarter of the GUN1 interactors identified by CoIP-MS are chaperones (see **Table 1**), a relatively high proportion when compared with the extensive repertoire of protein functions found within the nucleoid (Melonek et al., 2016). The stromal Hsp93 and Hsp70 chaperones mediate different steps in protein import into the chloroplast stroma, whereas the 60 KD chaperonin Cpn60 is thought to be involved in the folding of newly imported mature proteins and to function downstream of Hsp93 and Hsp70 (Kessler and Blobel, 1996; Jackson-Constan et al., 2001; Flores-Perez and Jarvis, 2013). Furthermore, the two genes most highly co-regulated with *GUN1* encode the proteins TIC110 and TOC159 (Tadini et al., 2016), which are part of the outer and inner chloroplast translocons, respectively, suggesting a role of GUN1 in chloroplast protein import (**Figure 2**).

The Hsp93 Chaperones

In Arabidopsis, there are two nearly identical isoforms of Hsp93, termed Hsp93-V and Hsp93-III (or ClpC1 and ClpC2, respectively) and both interact with GUN1. The two proteins are highly homologous, but Hsp93-V is expressed at much higher levels than Hsp93-III (Kovacheva et al., 2005, 2007), and only Hsp93-V has been reported as a component of the nucleoid proteome (Phinney and Thelen, 2005; Majeran et al., 2012; Melonek et al., 2012; Huang et al., 2013). Furthermore, both *hsp93* single mutants are viable whereas *hsp93-III hsp93-V* double mutant is embryo-lethal, indicating that the two proteins have redundant functions in Arabidopsis chloroplasts (Constan et al., 2004; Sjögren et al., 2004; Kovacheva et al., 2007).

The current model for chloroplast protein import assumes that the preprotein transit peptide interacts with the TOC, and is subsequently transported through the TIC in an energy-dependent process (Shi and Theg, 2013b). In particular, the Tic110–Tic40 interaction is proposed to trigger the release of the transit peptide from Tic110 and enable the association of the preprotein with Hsp93 (Inaba et al., 2003). Tic40 then stimulates ATP hydrolysis by Hsp93, which harnesses the energy released to draw the preprotein into the stroma (Chou et al., 2006).

The Hsp70 Chaperones

Recent work has also demonstrated the involvement of Hsp70 in protein translocation into chloroplasts, as part of the translocon energy-dependent engine together with Hsp93 and Hsp90 (Inoue et al., 2013; Liu et al., 2014). Like Hsp93, Hsp70 proteins occur in two isoforms, Hsp70-1 and Hsp70-2, in the chloroplasts of Arabidopsis (Su and Li, 2008) and only Hsp70-1 was found in the proteomes of pTAC and crude nucleoids (for a review see Melonek et al., 2016). However, both Hsp70 proteins have been identified as GUN1 interactors (Tadini et al., 2016). Protein import assays using chloroplasts isolated from the Arabidopsis Hsp70 knockout mutants *hsp70-1* and *hsp70-2* showed that stromal Hsp70s are important for the import of both photosynthetic and non-photosynthetic precursor proteins, especially in early developmental stages (Su and Li, 2010). Furthermore, no *hsp70-1 hsp70-2* double mutant has ever been isolated. Thus, the two Hsp70s are likely to have redundant functions that are essential for plant development and chloroplast biogenesis.

The Cpn60 Chaperonins

After preproteins delivered to the stroma have been processed, they may require accessory factors to enable them to fold into their functional conformation, or to reach their final intra-organellar destination. The stromal molecular chaperones Hsp70, Cpn60, and Cpn10 are all believed to mediate the folding or onward guidance of newly imported polypeptide chains (Boston et al., 1996; Jackson-Constan et al., 2001). In particular, immunoprecipitation experiments have revealed that Cpn60 operates in close proximity with Tic110 (Kessler and Blobel, 1996), while import experiments have shown a transient association of mature, newly imported proteins with the Cpn60-Tic110 complex, suggesting that Tic110 can recruit Cpn60 in an ATP-dependent manner for the folding of proteins upon their arrival in the stroma. It has also been suggested that stromal Hsp70 and Cpn60 act sequentially to facilitate the maturation of imported proteins, particularly those destined for the thylakoid membranes (Madueno et al., 1993; Tsugeki and Nishimura, 1993; Peng et al., 2011). The Arabidopsis genome encodes two members of the Cpn60α family, denoted ptCpn60α1 and ptCpn60α2, and four members of Cpn60β, known as ptCpn60β1–β4 (Suzuki et al., 2009). Two of them, ptCpn60α1 and ptCpn60β2, have been linked to the nucleoid proteome (Melonek et al., 2016), and ptCpn60α1 and ptCpn60β1 are among the GUN1 interactors identified via the CoIP-MS strategy (see **Table 1**). The complete loss of ptCpn60α1, in the

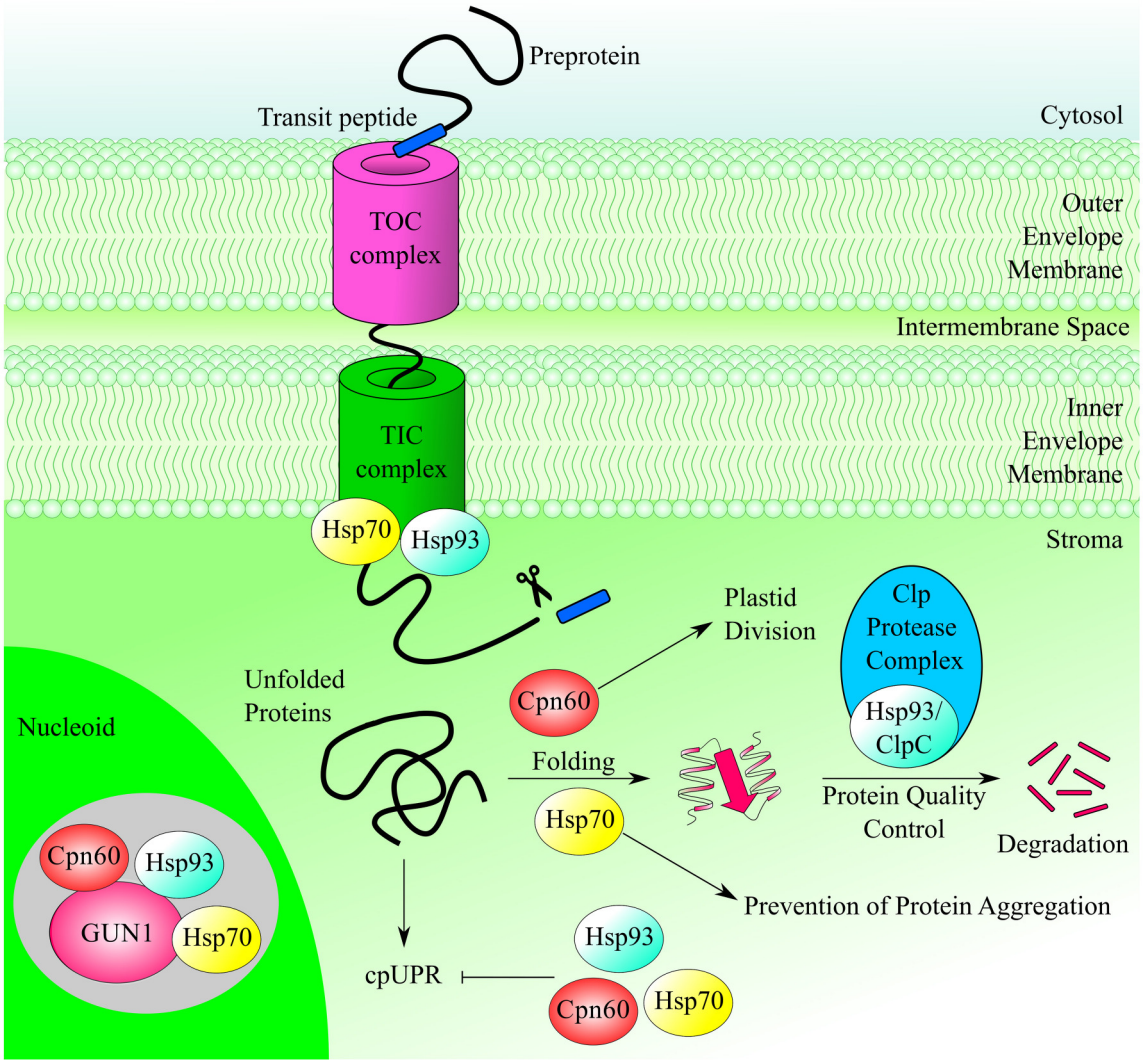

FIGURE 2 | GUN1 interacts with different plastid chaperones. The chaperones Hsp93, Hsp70, and Cpn60 participate in different processes within the chloroplast, such as protein import, protein folding/unfolding, prevention of protein aggregation, and regulation of plastid division, and they might play a key role in the chloroplast Unfolded Protein Response (cpUPR). CoIP-MS analysis has shown that they are also part of GUN1-containing protein complexes.

mutant termed *schlepperless* (*slp*), causes retardation of embryo development before the heart stage and an albino seedling phenotype, indicating that ptCpn60α1 is essential for chloroplast biogenesis (Apuya et al., 2001). Conversely, plants devoid of ptCpn60β1, also known as *lesion initiation 1* (*len1*), have leaves with wrinkled and irregular surfaces and undergo localized, spontaneous cell death in the absence of pathogen attack, i.e., lesion formation, under short-day conditions (Ishikawa et al., 2003).

OTHER FUNCTIONS OF PLASTID CHAPERONES

Besides their roles in plastid protein import, all GUN1-interacting chaperones are present in the stroma at significant amounts relative to their association with the chloroplast import apparatus and perform various other functions together with different protein complexes (**Figure 2**). For instance Hsp93, also termed ClpC, acts as a regulatory chaperone in the Clp protease complex, the most abundant stromal protease with general household functions (Sakamoto, 2006; Van Wijk, 2015). Clp substrates are selected through various signals intrinsic to amino acid sequences and the ATP-dependent ClpC chaperone activity helps to progressively unfold selected substrates that are delivered to the ClpPR core for degradation into small peptides (~8–10 amino acids long; Olinares et al., 2011).

Similarly, Cpn60 forms a large oligomeric protein complex (>600 KDa) that promotes the assembly of Rubisco (Gutteridge and Gatenby, 1995). In particular, it has been observed that the large subunit of Rubisco (RbcL) is specifically associated with

Cpn60 before assembly into the holoenzyme and that the Cpn60-RbcL complex is an obligatory intermediate. Furthermore, Cpn60 proteins have been shown to be essential for plastid division in *A. thaliana* (Suzuki et al., 2009). Thus, mesophyll cells in *ptcpn60α1-1* (a missense mutant) and *ptcpn60β1-1* (a protein null) plants, contain fewer and larger chloroplasts, indicating that normal levels of plastid Cpn60 are required for the correct folding of the stromal plastid division proteins and/or regulation of FtsZ (Filamentous temperature-sensitive Z) polymer dynamics (Suzuki et al., 2009).

The same holds true for the Hsp70 proteins, which are also involved in modulation of protein activity, regulation of protein degradation and prevention of irreversible protein aggregation when they are free in the stroma (Su and Li, 2008). Potentially GUN1 can be involved in a multitude of activities, besides plastid protein import, thus further investigations are needed to clarify the functional significance of GUN1–chaperone interactions.

GUN1 AND THE CHLOROPLAST UNFOLDED PROTEIN RESPONSE (cpUPR)

Chaperones, together with enzymes that process and degrade proteins, are also necessary to maintain protein folding homeostasis in the various compartments of eukaryotic cells. Distinct signal transduction pathways, known as unfolded protein responses (UPRs), have evolved to couple the unfolded/misfolded protein load to the expression of specific chaperones and enzymes that promote folding and the disposal of misfolded proteins in each compartment.

The unfolded protein response was first discovered in the endoplasmic reticulum (ER) in yeast, where inhibition of protein folding leads to the transcriptional up-regulation of several chaperones (Cox et al., 1993), and subsequently in mitochondria, where accumulation of unfolded proteins in the mitochondrial matrix stimulates the expression of nuclear gene transcripts coding for mitochondrial chaperones (Aldridge et al., 2007; Lin and Haynes, 2016). Compared to yeast and metazoans, studies of plant UPRs are less advanced, and molecular details are known mainly for the ER-dependent UPR, which shows certain similarities with the process in multicellular eukaryotes, as well as plant-specific features (Ruberti et al., 2015). Recently, the possible existence of a chloroplast UPR (cpUPR) has been investigated in the green alga *Chlamydomonas reinhardtii*. Taking advantage of a repressible chloroplast gene expression system (Rochaix et al., 2014), Ramundo et al. (2014) induced the selective gradual depletion of the essential stromal Clp protease, in order to follow the early and late events caused by the decrease in its abundance. Temporal profiles of gene expression and protein accumulation revealed a marked increase in levels of chaperones, including Hsp70B, upon Clp depletion. Similar data have also been reported for Arabidopsis, where up-regulation of chloroplast chaperones and protein-sorting components occurred upon constitutive repression of Clp (Rudella et al., 2006; Zybailov et al., 2009). In particular, characterization of total leaf proteomes of WT and *clpr2-1* highlighted differential expression of 768 proteins. The largest functional category quantified (with 205 proteins) comprised proteins involved in translation, folding and degradation. Strikingly, all the chaperones interacting with GUN1, including Hsp93, Hsp70, Cpn60, as well as the DEAD box RNA helicase RH3, are among those up-regulated (by between 1.6- and 8.5-fold) in *clpr2-1* leaves, whereas no significant change in the chloroplast ribosomal protein population was observed (Zybailov et al., 2009).

Taken together, these findings suggest that disruption of protein homeostasis in organelles can be sensed and transduced to the nucleus to induce the expression of a specific set of factors responsible for promoting folding and monitoring protein quality control (Ramundo and Rochaix, 2014; Rochaix and Ramundo, 2015). After entering the higher plant chloroplast, these factors are able to interact with the nucleoid-associated GUN1 protein (**Figure 2**), which might therefore play a role in the cpUPR process.

GUN1 AND CHLOROPLAST METABOLISM

The large subunit of ribulose bisphosphate carboxylase (RbcL) and the β subunit of the ATP synthase are also among the interactors of GUN1 identified by CoIP-MS analysis (Tadini et al., 2016). Because of their relatively high abundance in the chloroplast proteome, it is tempting to assume that these proteins are simply contaminants. However, RbcL and subunits of the ATP synthase have been repeatedly identified in the pTAC/nucleoid proteomes, even though different procedures were employed for isolation of crude nucleoid fractions and highly purified pTAC complexes (for a review see Melonek et al., 2016), thus suggesting these proteins might have a dual localization to the chloroplast stroma and nucleoids. The nucleoid association of RbcL and ATP synthase, i.e., of proteins that are not directly involved in core nucleoid functions, might also indicate that nucleoids also monitor photosynthesis and energy metabolism and respond appropriately to any perturbations (**Figure 3**).

Unlike RbcL and the ATP synthase β subunit, RETICULATA-RELATED 4 (RER4), an integral component of the chloroplast envelope membranes with three transmembrane α-helices, has never been identified in the pTAC/nucleoid proteome, although it appears to be an interactor of GUN1 (**Table 1**). The mutant *rer4-1* exhibits leaf reticulation, having green veins that stand out against paler intervein tissue, with fewer and smaller mesophyll cells than those of the wild type leaves (Perez-Perez et al., 2013). The molecular function of RER4 remains to be established. However, some hints as to its role in the chloroplast can be derived from features of the *rer4-1* mutant phenotype. A possible involvement of RER4 in retrograde signaling is suggested by the altered growth and development of mesophyll cells. Alternatively, the absence of RER4 might deplete the supply of essential metabolites during early stages of leaf development, which could explain the aberrant mesophyll structure. Furthermore, RER4 has been suggested to be involved in the control of reactive oxygen species (ROS), since the reticulated pigmentation of the *rer4-1* mutant grown under long-day conditions can be rescued

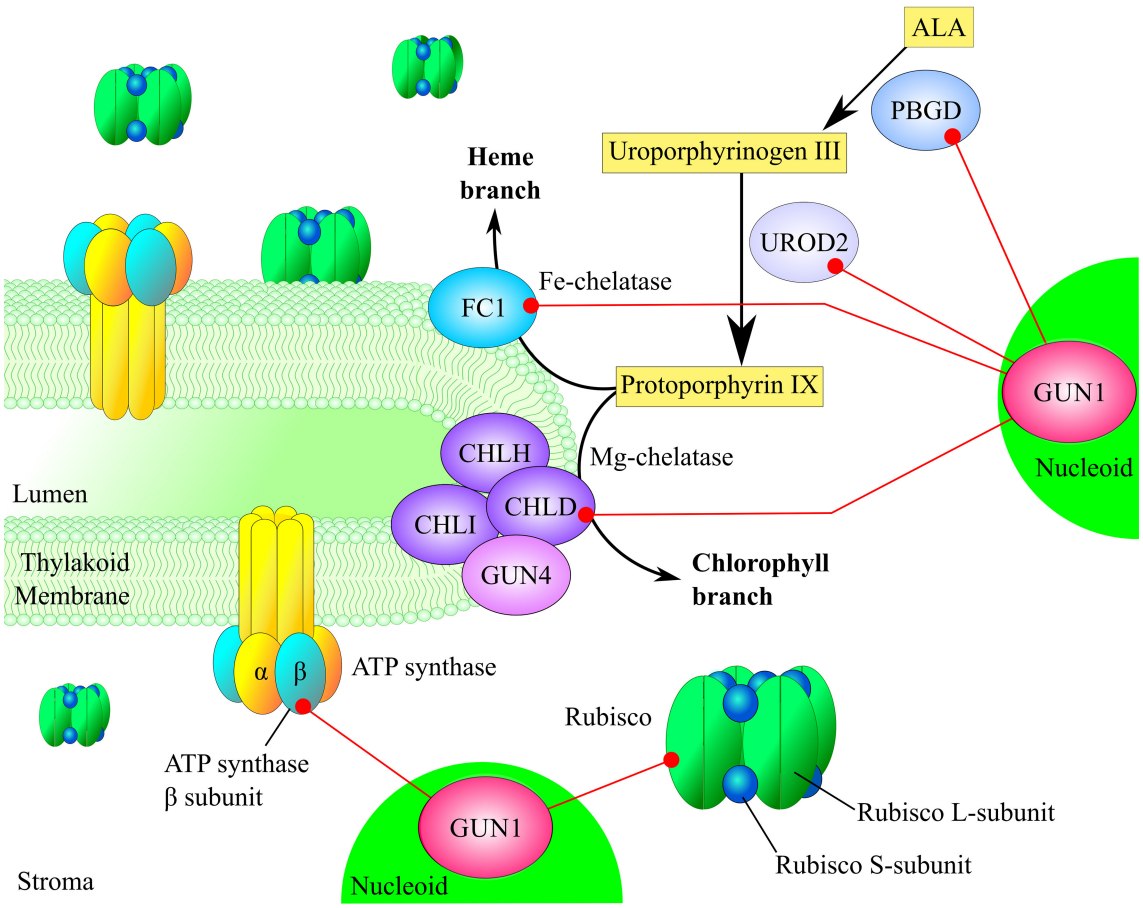

FIGURE 3 | GUN1 is involved in photosynthesis and tetrapyrrole biosynthesis. The large subunit of Rubisco and the β-subunit of the thylakoid ATP synthase have been coimmunoprecipitated with GUN1, supporting a role for GUN1 in coordinating nucleoid activities with chloroplast metabolism. GUN1 also interacts with four enzymes of the tetrapyrrole biosynthesis pathway, i.e., the D subunit of Mg chelatase (CHLD), porphobilinogen deaminase (PBGD), uroporphyrinogen III decarboxylase (UROD2), and ferrochelatase I (FC1), as shown by yeast two-hybrid and Bimolecular Fluorescence Complementation. Note that the proteins RER4 and 2-Cys PrxA have not been included in this scheme for reasons of clarity.

by a short-day photoperiod, which markedly dampens ROS accumulation.

The 2-Cys peroxiredoxin A (2-Cys Prx A; see also **Table 1**), another interactor with GUN1, appears also to have a role in ROS scavenging (Rey et al., 2007; Pulido et al., 2010; Dietz, 2016) and, like RER4, it has never been reported to be part of the pTAC/nucleoid proteome (Pfalz et al., 2006; Majeran et al., 2012; Huang et al., 2013). 2-Cys Prx A and the highly homologous 2-Cys Prx B function as redox sensors and chaperones, thanks to the flexibility of their protein structure (König et al., 2013), and they have been shown to control the conversion of Mg-protoporphyrin monomethyl ester into protochlorophyllide (Stenbaek et al., 2008).

The involvement of GUN1 in TPB is further supported by its interaction with four TPB enzymes, namely subunit D of Mg chelatase (CHLD), porphobilinogen deaminase (PBGD), uroporphyrinogen III decarboxylase (UROD2), and ferrochelatase I (FC1), as demonstrated by both yeast two-hybrid and BiFC assays (Tadini et al., 2016; **Figure 3**). Interestingly, mutants defective in three of these GUN1 interactors—CHLD, PBGD, and FC1—have themselves been described as *gun* mutants (Strand et al., 2003; Huang and Li, 2009; Woodson et al., 2011), but have never been identified in crude nucleoid preparations, unlike subunit I of Mg chelatase (CHLI; Melonek et al., 2012; Huang et al., 2013).

GUN1 AND PLASTID PROTEIN HOMEOSTASIS: SOME TESTABLE HYPOTHESES

The recent identification of the GUN1 protein's partners in chloroplasts of Arabidopsis by means of CoIP-MS studies as well as in yeast two-hybrid and BiFC assays (Tadini et al., 2016) strongly suggests a major role for GUN1 in plastid protein homeostasis (**Figure 4**). This regulatory role involves proteins that are, in most cases, members of multimeric protein complexes and whose functions are often context-dependent. Furthermore, most GUN1 interactors appear to participate in four major processes:

Chloroplast Protein Synthesis

A wealth of evidence accumulated during the last two decades supports the primacy of plastid protein synthesis in the control of chloroplast gene expression (Choquet and Wollman, 2002; Manuell et al., 2007; Tiller and Bock, 2014; Sun and Zerges, 2015). In this context, GUN1 has been suggested to regulate translation in plastids by modulating the abundance and binding affinity of PRPS1 (Tadini et al., 2016). In particular, PRPS1 is the only ribosomal protein that shuttles between ribosome-bound and ribosome-free forms (Merendino et al., 2003; Delvillani et al., 2011), the latter being more abundant in plants that lack GUN1. Based on observations in *E. coli*, where the unbound form is thought to inhibit translation by competing with ribosomes for mRNAs (Delvillani et al., 2011), it can be argued that the GUN1-dependent equilibrium between the two PRPS1 states has an important role in controlling polysome assembly and protein synthesis in chloroplasts (**Figure 4A**). However, further investigations are needed to clarify this issue. For instance, lines characterized by the ectopic expression of *PRPS1* or carrying *PRPS1* constructs under the control of inducible promoters, coupled with assays aimed to measure the translation rate in plastids, should allow us to verify the role of PRPS1 in modulating protein synthesis. Furthermore, GUN1 controls the abundance of PRPS1 at the post-transcriptional level. This suggests the involvement of an as yet unidentified plastid protease in this aspect of GUN1 function. In addition, the significance of the interaction of GUN1 with other ribosomal proteins, factors involved in ribosome biogenesis and regulators of plastid protein synthesis remains to be elucidated.

Chloroplast Protein Import and Degradation

Based on the observations reported above, it appears that GUN1 may well control the interactions of a sub-set of chaperones, promoting plastid protein import when their association with the TIC complex is favored, and stimulating protein degradation, folding/unfolding when they interact with proteases or other protein complexes in the stroma or in the thylakoid membranes (**Figure 4B**). Such a regulatory mechanism would enable GUN1 to coordinate protein translocation across the chloroplast envelope with protein degradation in the stroma,

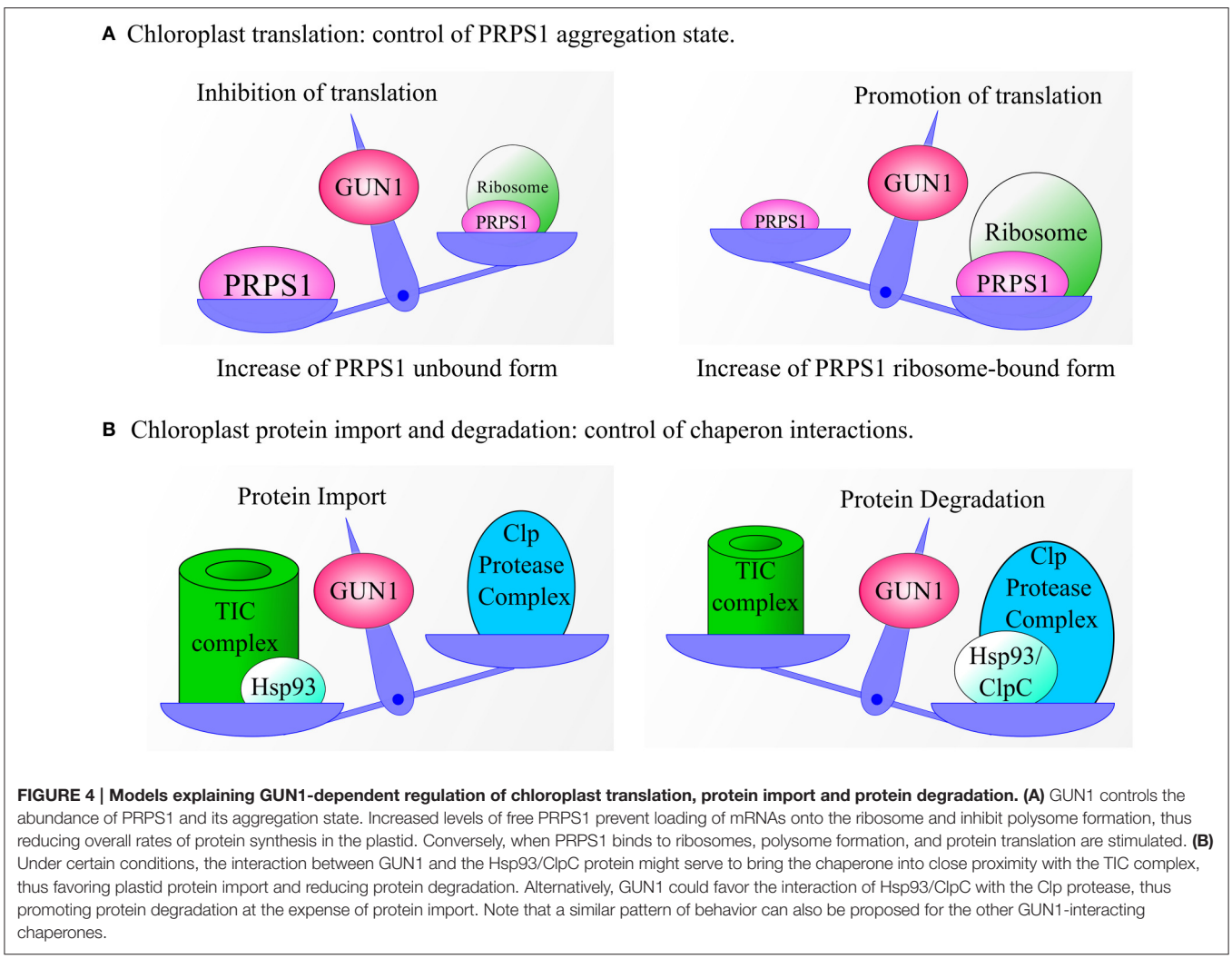

FIGURE 4 | Models explaining GUN1-dependent regulation of chloroplast translation, protein import and protein degradation. (A) GUN1 controls the abundance of PRPS1 and its aggregation state. Increased levels of free PRPS1 prevent loading of mRNAs onto the ribosome and inhibit polysome formation, thus reducing overall rates of protein synthesis in the plastid. Conversely, when PRPS1 binds to ribosomes, polysome formation, and protein translation are stimulated. **(B)** Under certain conditions, the interaction between GUN1 and the Hsp93/ClpC protein might serve to bring the chaperone into close proximity with the TIC complex, thus favoring plastid protein import and reducing protein degradation. Alternatively, GUN1 could favor the interaction of Hsp93/ClpC with the Clp protease, thus promoting protein degradation at the expense of protein import. Note that a similar pattern of behavior can also be proposed for the other GUN1-interacting chaperones.

as well as with plastid division, thus modulating the protein content of the chloroplast in accordance with physiological requirements.

Relatively simple biochemical analyses can be used to verify the importance of GUN1 in influencing the interactions of the stromal chaperones, such as protein complex fractionation via sucrose-gradient ultracentrifugation and/or Blue-Native PAGE coupled with two-dimensional (2D) SDS-PAGE, and immunoblot analyses. Furthermore, the interactions of GUN1 with chaperones should be shown to occur at the plastid envelope and protein import efficiency should be tested in chloroplasts isolated from *gun1* and WT seedlings in order to implicate GUN1 in regulating plastid protein import.

Retrograde Signal Induction

GUN1 may well be a master regulator of plastid-to-nucleus communication in *A. thaliana*, as it appears to integrate signals derived from perturbations in PGE, TPB, and redox state, in order to modulate nuclear gene expression. Indeed, components of all three pathways have been shown to interact with GUN1, suggesting that signal integration might take place through physical interaction.

Due to the limited abundance of GUN1, as indicated by the fact that the protein has yet to be detected in plastid proteome studies, it is tempting to disregard the idea that its physical interaction with PGE-, TPB-, and redox-related proteins could lead to protein sequestration and directly to differences in protein translation, TPB, and redox balance (Koussevitzky et al., 2007; Pogson et al., 2008; Woodson and Chory, 2008; Kleine and Leister, 2016). Nevertheless, a direct association with GUN1 could control protein abundance through post-transcriptional mechanisms, as in the case of PRPS1 and CHLD (Tadini et al., 2016). Thus, control of CHLD and possibly of FC1 levels could alter the tetrapyrrole flux and influence the abundance of the tetrapyrrole intermediate Mg-protoporphyrin IX (Mg-ProtoIX), or the tetrapyrrole product Fe-protoporphyrin IX (heme), which have been reported to act as negative and positive retrograde signals, respectively (for a review, see Chan et al., 2016). Alternatively, the interaction of GUN1 with the near-identical paralogs ClpC1 and ClpC2 could contribute to the coordination of plastid protein content with tetrapyrrole biosynthesis. Indeed, the activity of the stromal Clp protease has been shown to modulate tetrapyrrole flux by controlling (i) the accumulation of chlorophyll a oxygenase, which converts chlorophyll a into chlorophyll b (Nakagawara et al., 2007), and (ii) the level of glutamyl-tRNA reductase (GluTR), thus regulating the rate-limiting reaction in tetrapyrrole synthesis—the conversion of glutamate-1-semialdehyde into 5-aminolevulinic acid (Apitz et al., 2016).

Therefore, accurate determination of tetrapyrrole intermediates should be performed in *gun1* mutant and WT backgrounds. The analyses should be restricted to young seedlings or even to different developmental stages of the chloroplast, in line with the roles of tetrapyrrole and GUN1-mediated signaling in chloroplast development.

CONCLUDING REMARKS

In the past decade, substantial progress has been made in elucidating retrograde signaling, with the identification of multiple retrograde pathways and more than 40 components involved at different levels in chloroplast-to-nucleus communication. Nevertheless, the molecular function of GUN1 has remained unclear until the recent identification of the GUN1 protein's partners. Based on the functional roles of GUN1 interactors and the embryo lethal or albino phenotypes of most of the corresponding knock-out mutants, we have learned that GUN1 plays a role in chloroplast biogenesis, possibly by controlling protein turnover and protein import, and through the coordination of plastid and nuclear gene expression. Furthermore, GUN1 could have a role in the cpUPR process. Nonetheless, the involvement of GUN1 in plastid biogenesis and protein homeostasis is only just beginning to be understood. For instance, other approaches will be needed to validate the GUN1's protein partners identified by CoIP-MS. The use of a GUN1-GFP protein chimera, expressed under the control of a strong constitutive promoter such as the Cauliflower Mosaic Virus 35S (35S-CaMV), is indeed prone to the identification of false interactors. CoIP-MS studies using a GUN1 specific antibody appears to be the ideal strategy to identify protein partners. Alternatively, the use of GUN1 chimeras under the control of GUN1 native promoter is also practicable. Moreover, we do not know whether all these activities take place within one GUN1-containing nucleoid or if there are different nucleoids/locations for each GUN1-dependent function. The developmental stages of the chloroplast itself may even show distinct patterns of compartmentalization of the different functions. In addition, GUN1's interactions with its diverse partners might have quite different functional consequences: (i) promote specific functions, by bringing enzymes into close proximity with their own substrates and, ultimately, controlling the enzyme abundance, (ii) inhibit processes by sequestering sub-pools of specific proteins and, also in this case, controlling their abundance.

We are confident that future work, based on the exciting breakthroughs discussed in this Review, will shed new light on the molecular functions of GUN1 and its involvement in chloroplast biogenesis and protein homeostasis.

AUTHOR CONTRIBUTIONS

MC, LT, CP, RF, and PP participated to the organization of the manuscript. MC and PP designed and conceived the pictures. PP wrote the manuscript.

ACKNOWLEDGMENTS

This work was supported by ERA-NET Cofund FACCE SURPLUS (BarPLUS grant id. 93). We thank Paul Hardy for critical reading of the manuscript. The authors wish to acknowledge the reviewers for the detailed and helpful comments to the manuscript.

REFERENCES

Adhikari, N. D., Froehlich, J. E., Strand, D. D., Buck, S. M., Kramer, D. M., and Larkin, R. M. (2011). GUN4-porphyrin complexes bind the ChlH/GUN5 subunit of Mg-Chelatase and promote chlorophyll biosynthesis in Arabidopsis. *Plant Cell* 23, 1449–1467. doi: 10.1105/tpc.110.082503

Agee, A. E., Surpin, M., Sohn, E. J., Girke, T., Rosado, A., Kram, B. W., et al. (2010). MODIFIED VACUOLE PHENOTYPE1 is an Arabidopsis myrosinase-associated protein involved in endomembrane protein trafficking. *Plant Physiol.* 152, 120–132. doi: 10.1104/pp.109.145078

Aldridge, J. E., Horibe, T., and Hoogenraad, N. J. (2007). Discovery of genes activated by the mitochondrial unfolded protein response (mtUPR) and cognate promoter elements. *PLoS ONE* 2:e874. doi: 10.1371/journal.pone.0000874

Apitz, J., Nishimura, K., Schmied, J., Wolf, A., Hedtke, B., Van Wijk, K. J., et al. (2016). Posttranslational control of ALA synthesis includes GluTR degradation by Clp protease and stabilization by GluTR-binding protein. *Plant Physiol.* 170, 2040–2051. doi: 10.1104/pp.15.01945

Apuya, N. R., Yadegari, R., Fischer, R. L., Harada, J. J., Zimmerman, J. L., and Goldberg, R. B. (2001). The Arabidopsis embryo mutant schlepperless has a defect in the chaperonin-60alpha gene. *Plant Physiol.* 126, 717–730. doi: 10.1104/pp.126.2.717

Asakura, Y., Galarneau, E., Watkins, K. P., Barkan, A., and Van Wijk, K. J. (2012). Chloroplast RH3 DEAD box RNA helicases in maize and Arabidopsis function in splicing of specific group II introns and affect chloroplast ribosome biogenesis. *Plant Physiol.* 159, 961–974. doi: 10.1104/pp.112.197525

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2, 2006.0008. doi: 10.1038/msb4100050

Bohne, A. V. (2014). The nucleoid as a site of rRNA processing and ribosome assembly. *Front. Plant Sci.* 5:257. doi: 10.3389/fpls.2014.00257

Boston, R. S., Viitanen, P. V., and Vierling, E. (1996). Molecular chaperones and protein folding in plants. *Plant Mol. Biol.* 32, 191–222. doi: 10.1007/BF00039383

Bryant, N., Lloyd, J., Sweeney, C., Myouga, F., and Meinke, D. (2011). Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in Arabidopsis. *Plant Physiol.* 155, 1678–1689. doi: 10.1104/pp.110.168120

Chan, K. X., Phua, S. Y., Crisp, P., McQuinn, R., and Pogson, B. J. (2016). Learning the languages of the chloroplast: retrograde signaling and beyond. *Annu. Rev. Plant Biol.* 67, 25–53. doi: 10.1146/annurev-arplant-043015-111854

Choquet, Y., and Wollman, F. A. (2002). Translational regulations as specific traits of chloroplast gene expression. *FEBS Lett.* 529, 39–42. doi: 10.1016/S0014-5793(02)03260-X

Chou, M. L., Chu, C. C., Chen, L. J., Akita, M., and Li, H. M. (2006). Stimulation of transit-peptide release and ATP hydrolysis by a cochaperone during protein import into chloroplasts. *J. Cell Biol.* 175, 893–900. doi: 10.1083/jcb.200609172

Constan, D., Froehlich, J. E., Rangarajan, S., and Keegstra, K. (2004). A stromal Hsp100 protein is required for normal chloroplast development and function in Arabidopsis. *Plant Physiol.* 136, 3605–3615. doi: 10.1104/pp.104.052928

Cox, J. S., Shamu, C. E., and Walter, P. (1993). Transcriptional induction of genes encoding endoplasmic reticulum resident proteins requires a transmembrane protein kinase. *Cell* 73, 1197–1206. doi: 10.1016/0092-8674(93)90648-A

Cushing, D. A., Forsthoefel, N. R., Gestaut, D. R., and Vernon, D. M. (2005). Arabidopsis emb175 and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. *Planta* 221, 424–436. doi: 10.1007/s00425-004-1452-x

Delvillani, F., Papiani, G., Dehò, G., and Briani, F. (2011). S1 ribosomal protein and the interplay between translation and mRNA decay. *Nucleic Acids Res.* 39, 7702–7715. doi: 10.1093/nar/gkr417

Dietz, K. J. (2016). Thiol-based peroxidases and ascorbate peroxidases: why plants rely on multiple peroxidase systems in the photosynthesizing chloroplast? *Mol. Cells* 39, 20–25. doi: 10.14348/molcells.2016.2324

Fleischmann, T. T., Scharff, L. B., Alkatib, S., Hasdorf, S., Schöttler, M. A., and Bock, R. (2011). Nonessential plastid-encoded ribosomal proteins in tobacco: a developmental role for plastid translation and implications for reductive genome evolution. *Plant Cell* 23, 3137–3155. doi: 10.1105/tpc.111.088906

Flores-Pérez, U., and Jarvis, P. (2013). Molecular chaperone involvement in chloroplast protein import. *Biochim. Biophys. Acta* 1833, 332–340. doi: 10.1016/j.bbamcr.2012.03.019

Gray, J. C., Sullivan, J. A., Wang, J. H., Jerome, C. A., and Maclean, D. (2003). Coordination of plastid and nuclear gene expression. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 135–144. discussion 144–145. doi: 10.1098/rstb.2002.1180

Gutteridge, S., and Gatenby, A. A. (1995). Rubisco synthesis, assembly, mechanism, and regulation. *Plant Cell* 7, 809–819. doi: 10.1105/tpc.7.7.809

Huang, M., Friso, G., Nishimura, K., Qu, X., Olinares, P. D., Majeran, W., et al. (2013). Construction of plastid reference proteomes for maize and Arabidopsis and evaluation of their orthologous relationships; the concept of orthoproteomics. *J. Proteome Res.* 12, 491–504. doi: 10.1021/pr300952g

Huang, Y. S., and Li, H. M. (2009). Arabidopsis CHLI2 can substitute for CHLI1. *Plant Physiol.* 150, 636–645. doi: 10.1104/pp.109.135368

Inaba, T., Li, M., Alvarez-Huerta, M., Kessler, F., and Schnell, D. J. (2003). atTic110 functions as a scaffold for coordinating the stromal events of protein import into chloroplasts. *J. Biol. Chem.* 278, 38617–38627. doi: 10.1074/jbc.M306367200

Inoue, H., Li, M., and Schnell, D. J. (2013). An essential role for chloroplast heat shock protein 90 (Hsp90C) in protein import into chloroplasts. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3173–3178. doi: 10.1073/pnas.1219229110

Ishikawa, A. (2005). Tetrapyrrole metabolism is involved in lesion formation, cell death, in the Arabidopsis lesion initiation 1 mutant. *Biosci. Biotechnol. Biochem.* 69, 1929–1934. doi: 10.1271/bbb.69.1929

Ishikawa, A., Tanaka, H., Nakai, M., and Asahi, T. (2003). Deletion of a chaperonin 60 beta gene leads to cell death in the Arabidopsis lesion initiation 1 mutant. *Plant Cell Physiol.* 44, 255–261. doi: 10.1093/pcp/pcg031

Jackson-Constan, D., Akita, M., and Keegstra, K. (2001). Molecular chaperones involved in chloroplast protein import. *Biochim. Biophys. Acta* 1541, 102–113. doi: 10.1016/S0167-4889(01)00148-3

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

Kato, Y., and Sakamoto, W. (2010). New insights into the types and function of proteases in plastids. *Int. Rev. Cell Mol. Biol.* 280, 185–218. doi: 10.1016/S1937-6448(10)80004-8

Kessler, F., and Blobel, G. (1996). Interaction of the protein import and folding machineries of the chloroplast. *Proc. Natl. Acad. Sci. U.S.A.* 93, 7684–7689. doi: 10.1073/pnas.93.15.7684

Kleine, T., and Leister, D. (2016). Retrograde signaling: organelles go networking. *Biochim. Biophys. Acta.* 1857, 1313–1325. doi: 10.1016/j.bbabio.2016.03.017

König, J., Galliardt, H., Jütte, P., Schäper, S., Dittmann, L., and Dietz, K. J. (2013). The conformational bases for the two functionalities of 2-cysteine peroxiredoxins as peroxidase and chaperone. *J. Exp. Bot.* 64, 3483–3497. doi: 10.1093/jxb/ert184

Koussevitzky, S., Nott, A., Mockler, T. C., Hong, F., Sachetto-Martins, G., Surpin, M., et al. (2007). Signals from chloroplasts converge to regulate nuclear gene expression. *Science* 316, 715–719. doi: 10.1126/science. 1140516

Kovacheva, S., Bédard, J., Patel, R., Dudley, P., Twell, D., Ríos, G., et al. (2005). *In vivo* studies on the roles of Tic110, Tic40 and Hsp93 during chloroplast protein import. *Plant J.* 41, 412–428. doi: 10.1111/j.1365-313X.2004.02307.x

Kovacheva, S., Bédard, J., Wardle, A., Patel, R., and Jarvis, P. (2007). Further *in vivo* studies on the role of the molecular chaperone, Hsp93, in plastid protein import. *Plant J.* 50, 364–379. doi: 10.1111/j.1365-313X.2007.03060.x

Larkin, R. M., Alonso, J. M., Ecker, J. R., and Chory, J. (2003). GUN4, a regulator of chlorophyll synthesis and intracellular signaling. *Science* 299, 902–906. doi: 10.1126/science.1079978

Lin, Y. F., and Haynes, C. M. (2016). Metabolism and the UPR(mt). *Mol. Cell* 61, 677–682. doi: 10.1016/j.molcel.2016.02.004

Liu, L., McNeilage, R. T., Shi, L. X., and Theg, S. M. (2014). ATP requirement for chloroplast protein import is set by the Km for ATP hydrolysis of stromal Hsp70 in *Physcomitrella patens*. *Plant Cell* 26, 1246–1255. doi: 10.1105/tpc.113.121822

Madueno, F., Napier, J. A., and Gray, J. C. (1993). Newly imported rieske iron-sulfur protein associates with Both Cpn60 and Hsp70 in the chloroplast stroma. *Plant Cell* 5, 1865–1876. doi: 10.1105/tpc.5.12.1865

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

Manuell, A. L., Quispe, J., and Mayfield, S. P. (2007). Structure of the chloroplast ribosome: novel domains for translation regulation. *PLoS Biol.* 5:e209. doi: 10.1371/journal.pbio.0050209

Melonek, J., Matros, A., Trösch, M., Mock, H. P., and Krupinska, K. (2012). The core of chloroplast nucleoids contains architectural SWIB domain proteins. *Plant Cell* 24, 3060–3073. doi: 10.1105/tpc.112.099721

Melonek, J., Oetke, S., and Krupinska, K. (2016). Multifunctionality of plastid nucleoids as revealed by proteome analyses. *Biochim. Biophys. Acta.* 1864, 1016–1038. doi: 10.1016/j.bbapap.2016.03.009

Merendino, L., Falciatore, A., and Rochaix, J. D. (2003). Expression and RNA binding properties of the chloroplast ribosomal protein S1 from *Chlamydomonas reinhardtii*. *Plant Mol. Biol.* 53, 371–382. doi: 10.1023/B:PLAN.0000006941.56233.42

Miura, E., Kato, Y., Matsushima, R., Albrecht, V., Laalami, S., and Sakamoto, W. (2007). The balance between protein synthesis and degradation in chloroplasts determines leaf variegation in Arabidopsis yellow variegated mutants. *Plant Cell* 19, 1313–1328. doi: 10.1105/tpc.106.049270

Mochizuki, N., Brusslan, J. A., Larkin, R., Nagatani, A., and Chory, J. (2001). Arabidopsis genomes uncoupled 5 (GUN5) mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. *Proc. Natl. Acad. Sci. U.S.A.* 98, 2053–2058. doi: 10.1073/pnas.98.4.2053

Nakagawara, E., Sakuraba, Y., Yamasato, A., Tanaka, R., and Tanaka, A. (2007). Clp protease controls chlorophyll b synthesis by regulating the level of chlorophyllide a oxygenase. *Plant J.* 49, 800–809. doi: 10.1111/j.1365-313X.2006.02996.x

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

Oelmüller, R., and Mohr, H. (1986). Photooxidative destruction of chloroplasts and its consequences for expression of nuclear genes. *Planta* 167, 106–113. doi: 10.1007/BF00446376

Okazaki, K., Kabeya, Y., and Miyagishima, S. Y. (2010). The evolution of the regulatory mechanism of chloroplast division. *Plant Signal. Behav.* 5, 164–167. doi: 10.4161/psb.5.2.10461

Olinares, P. D., Kim, J., and van Wijk, K. J. (2011). The Clp protease system; a central component of the chloroplast protease network. *Biochim. Biophys. Acta* 1807, 999–1011. doi: 10.1016/j.bbabio.2010.12.003

Peng, L., Fukao, Y., Myouga, F., Motohashi, R., Shinozaki, K., and Shikanai, T. (2011). A chaperonin subunit with unique structures is essential for folding of a specific substrate. *PLoS Biol.* 9:e1001040. doi: 10.1371/journal.pbio.1001040

Pérez-Pérez, J. M., Esteve-Bruna, D., González-Bayón, R., Kangasjarvi, S., Caldana, C., Hannah, M. A., et al. (2013). Functional redundancy and divergence within the Arabidopsis RETICULATA-RELATED gene family. *Plant Physiol.* 162, 589–603. doi: 10.1104/pp.113.217323

Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K. J., and Oelmüller, R. (2006). pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell* 18, 176–197. doi: 10.1105/tpc.105.036392

Pfalz, J., and Pfannschmidt, T. (2013). Essential nucleoid proteins in early chloroplast development. *Trends Plant Sci.* 18, 186–194. doi: 10.1016/j.tplants.2012.11.003

Phinney, B. S., and Thelen, J. J. (2005). Proteomic characterization of a triton-insoluble fraction from chloroplasts defines a novel group of proteins associated with macromolecular structures. *J. Proteome Res.* 4, 497–506. doi: 10.1021/pr049791k

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

Pulido, P., Spínola, M. C., Kirchsteiger, K., Guinea, M., Pascual, M. B., Sahrawy, M., et al. (2010). Functional analysis of the pathways for 2-Cys peroxiredoxin reduction in *Arabidopsis thaliana* chloroplasts. *J. Exp. Bot.* 61, 4043–4054. doi: 10.1093/jxb/erq218

Ramundo, S., Casero, D., Mühlhaus, T., Hemme, D., Sommer, F., Crèvecoeur, M., et al. (2014). Conditional depletion of the chlamydomonas chloroplast ClpP protease activates nuclear genes involved in autophagy and plastid protein quality control. *Plant Cell* 26, 2201–2222. doi: 10.1105/tpc.114.124842

Ramundo, S., Rahire, M., Schaad, O., and Rochaix, J. D. (2013). Repression of essential chloroplast genes reveals new signaling pathways and regulatory feedback loops in chlamydomonas. *Plant Cell* 25, 167–186. doi: 10.1105/tpc.112.103051

Ramundo, S., and Rochaix, J. D. (2014). Chloroplast unfolded protein response, a new plastid stress signaling pathway? *Plant Signal. Behav.* 9:e972874. doi: 10.4161/15592316.2014.972874

Rey, P., Becuwe, N., Barrault, M. B., Rumeau, D., Havaux, M., Biteau, B., et al. (2007). The *Arabidopsis thaliana* sulfiredoxin is a plastidic cysteine-sulfinic acid reductase involved in the photooxidative stress response. *Plant J.* 49, 505–514. doi: 10.1111/j.1365-313X.2006.02969.x

Richly, E., and Leister, D. (2004). An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of Arabidopsis and rice. *Gene* 329, 11–16. doi: 10.1016/j.gene.2004.01.008

Rochaix, J. D., and Ramundo, S. (2015). Conditional repression of essential chloroplast genes: evidence for new plastid signaling pathways. *Biochim. Biophys. Acta* 1847, 986–992. doi: 10.1016/j.bbabio.2014.11.011

Rochaix, J. D., Surzycki, R., and Ramundo, S. (2014). Tools for regulated gene expression in the chloroplast of Chlamydomonas. *Methods Mol. Biol.* 1132, 413–424. doi: 10.1007/978-1-62703-995-6_28

Rogalski, M., Schöttler, M. A., Thiele, W., Schulze, W. X., and Bock, R. (2008). Rpl33, a nonessential plastid-encoded ribosomal protein in tobacco, is required under cold stress conditions. *Plant Cell* 20, 2221–2237. doi: 10.1105/tpc.108.060392

Ruberti, C., Kim, S. J., Stefano, G., and Brandizzi, F. (2015). Unfolded protein response in plants: one master, many questions. *Curr. Opin. Plant Biol.* 27, 59–66. doi: 10.1016/j.pbi.2015.05.016

Rudella, A., Friso, G., Alonso, J. M., Ecker, J. R., and Van Wijk, K. J. (2006). Downregulation of ClpR2 leads to reduced accumulation of the ClpPRS protease complex and defects in chloroplast biogenesis in Arabidopsis. *Plant Cell* 18, 1704–1721. doi: 10.1105/tpc.106.042861

Sakamoto, W. (2006). Protein degradation machineries in plastids. *Annu. Rev. Plant Biol.* 57, 599–621. doi: 10.1146/annurev.arplant.57.032905.105401

Shi, L. X., and Theg, S. M. (2010). A stromal heat shock protein 70 system functions in protein import into chloroplasts in the moss *Physcomitrella patens*. *Plant Cell* 22, 205–220. doi: 10.1105/tpc.109.071464

Shi, L. X., and Theg, S. M. (2013a). The chloroplast protein import system: from algae to trees. *Biochim. Biophys. Acta* 1833, 314–331. doi: 10.1016/j.bbamcr.2012.10.002

Shi, L. X., and Theg, S. M. (2013b). Energetic cost of protein import across the envelope membranes of chloroplasts. *Proc. Natl. Acad. Sci. U.S.A.* 110, 930–935. doi: 10.1073/pnas.1115886110

Shoji, S., Dambacher, C. M., Shajani, Z., Williamson, J. R., and Schultz, P. G. (2011). Systematic chromosomal deletion of bacterial ribosomal protein genes. *J. Mol. Biol.* 413, 751–761. doi: 10.1016/j.jmb.2011.09.004

Sjögren, L. L., Macdonald, T. M., Sutinen, S., and Clarke, A. K. (2004). Inactivation of the clpC1 gene encoding a chloroplast Hsp100 molecular chaperone causes growth retardation, leaf chlorosis, lower photosynthetic activity, and a specific reduction in photosystem content. *Plant Physiol.* 136, 4114–4126. doi: 10.1104/pp.104.053835

Steiner, S., Schröter, Y., Pfalz, J., and Pfannschmidt, T. (2011). Identification of essential subunits in the plastid-encoded RNA polymerase complex reveals building blocks for proper plastid development. *Plant Physiol.* 157, 1043–1055. doi: 10.1104/pp.111.184515

Stenbaek, A., Hansson, A., Wulff, R. P., Hansson, M., Dietz, K. J., and Jensen, P. E. (2008). NADPH-dependent thioredoxin reductase and 2-Cys peroxiredoxins are needed for the protection of Mg-protoporphyrin monomethyl ester cyclase. *FEBS Lett.* 582, 2773–2778. doi: 10.1016/j.febslet.2008.07.006

Strand, A., Asami, T., Alonso, J., Ecker, J. R., and Chory, J. (2003). Chloroplast to nucleus communication triggered by accumulation of Mg-protoporphyrinIX. *Nature* 421, 79–83. doi: 10.1038/nature01204

Su, P. H., and Li, H. M. (2008). Arabidopsis stromal 70-kD heat shock proteins are essential for plant development and important for thermotolerance of germinating seeds. *Plant Physiol.* 146, 1231–1241. doi: 10.1104/pp.107.114496

Su, P. H., and Li, H. M. (2010). Stromal Hsp70 is important for protein translocation into pea and Arabidopsis chloroplasts. *Plant Cell* 22, 1516–1531. doi: 10.1105/tpc.109.071415

Sun, Y., and Zerges, W. (2015). Translational regulation in chloroplasts for development and homeostasis. *Biochim. Biophys. Acta* 1847, 809–820. doi: 10.1016/j.bbabio.2015.05.008

Susek, R. E., Ausubel, F. M., and Chory, J. (1993). Signal transduction mutants of Arabidopsis uncouple nuclear CAB and RBCS gene expression from chloroplast development. *Cell* 74, 787–799. doi: 10.1016/0092-8674(93)90459-4

Suzuki, K., Nakanishi, H., Bower, J., Yoder, D. W., Osteryoung, K. W., and Miyagishima, S. Y. (2009). Plastid chaperonin proteins Cpn60 alpha and Cpn60 beta are required for plastid division in *Arabidopsis thaliana*. *BMC Plant Biol.* 9:38. doi: 10.1186/1471-2229-9-38

Tadini, L., Pesaresi, P., Kleine, T., Rossi, F., Guljamow, A., Sommer, F., et al. (2016). GUN1 controls accumulation of the plastid ribosomal protein S1 at the protein level and interacts with proteins involved in plastid protein homeostasis. *Plant Physiol.* 170, 1817–1830. doi: 10.1104/pp.15.02033

Tanaka, R., Kobayashi, K., and Masuda, T. (2011). Tetrapyrrole metabolism in *Arabidopsis thaliana*. *Arabidopsis Book* 9:e0145. doi: 10.1199/tab.0145

Tiller, N., and Bock, R. (2014). The translational apparatus of plastids and its role in plant development. *Mol. Plant* 7, 1105–1120. doi: 10.1093/mp/ssu022

Tsugeki, R., and Nishimura, M. (1993). Interaction of homologues of Hsp70 and Cpn60 with ferredoxin-NADP+ reductase upon its import into chloroplasts. *FEBS Lett.* 320, 198–202. doi: 10.1016/0014-5793(93)80585-I

Van Wijk, K. J. (2015). Protein maturation and proteolysis in plant plastids, mitochondria, and peroxisomes. *Annu. Rev. Plant Biol.* 66, 75–111. doi: 10.1146/annurev-arplant-043014-115547

Woodson, J. D., and Chory, J. (2008). Coordination of gene expression between organellar and nuclear genomes. *Nat. Rev. Genet.* 9, 383–395. doi: 10.1038/nrg2348

Woodson, J. D., Perez-Ruiz, J. M., and Chory, J. (2011). Heme synthesis by plastid ferrochelatase I regulates nuclear gene expression in plants. *Curr. Biol.* 21, 897–903. doi: 10.1016/j.cub.2011.04.004

Zybailov, B., Friso, G., Kim, J., Rudella, A., Rodríguez, V. R., Asakura, Y., et al. (2009). Large scale comparative proteomics of a chloroplast Clp protease mutant reveals folding stress, altered protein homeostasis, and feedback regulation of metabolism. *Mol. Cell. Proteomics* 8, 1789–1810. doi: 10.1074/mcp.M900104-MCP200

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

MANUSCRIPT II

frontiers in Plant Science

ORIGINAL RESEARCH
published: 15 February 2017
doi: 10.3389/fpls.2017.00163

CRP1 Protein: (dis)similarities between *Arabidopsis thaliana* and *Zea mays*

*Roberto Ferrari[1†], Luca Tadini[1†], Fabio Moratti[2†], Marie-Kristin Lehniger[3], Alex Costa[1], Fabio Rossi[4], Monica Colombo[5], Simona Masiero[1], Christian Schmitz-Linneweber[3] and Paolo Pesaresi[6]**

[1] Dipartimento di Bioscienze, Università degli studi di Milano, Milano, Italy, [2] Max-Planck-Institut für Molekulare Pflanzenphysiologie, Potsdam-Golm, Germany, [3] Molecular Genetics, Institute of Biology, Humboldt University of Berlin, Berlin, Germany, [4] Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università degli studi di Milano, Milano, Italy, [5] Centro Ricerca e Innovazione, Fondazione Edmund Mach, San Michele all'Adige, Italy, [6] Dipartimento di Scienze Agrarie e Ambientali - Produzione, Territorio, Agroenergia, Università degli studi di Milano, Milano, Italy

OPEN ACCESS

Edited by:
Federico Valverde, Consejo Superior de Investigaciones Científicas (CSIC), Spain

Reviewed by:
Jean-David Rochaix, University of Geneva, Switzerland
Alexandra-Viola Bohne, Ludwig Maximilian University of Munich, Germany

***Correspondence:**
Paolo Pesaresi
paolo.pesaresi@unimi.it

† These authors have contributed equally to this work.

Specialty section:
This article was submitted to Plant Evolution and Development, a section of the journal Frontiers in Plant Science

Received: *23 November 2016*
Accepted: *26 January 2017*
Published: *15 February 2017*

Citation:
Ferrari R, Tadini L, Moratti F, Lehniger M-K, Costa A, Rossi F, Colombo M, Masiero S, Schmitz-Linneweber C and Pesaresi P (2017) CRP1 Protein: (dis)similarities between Arabidopsis thaliana and Zea mays. Front. Plant Sci. 8:163. doi: 10.3389/fpls.2017.00163

Biogenesis of chloroplasts in higher plants is initiated from proplastids, and involves a series of processes by which a plastid able to perform photosynthesis, to synthesize amino acids, lipids, and phytohormones is formed. All plastid protein complexes are composed of subunits encoded by the nucleus and chloroplast genomes, which require a coordinated gene expression to produce the correct concentrations of organellar proteins and to maintain organelle function. To achieve this, hundreds of nucleus-encoded factors are imported into the chloroplast to control plastid gene expression. Among these factors, members of the Pentatricopeptide Repeat (PPR) containing protein family have emerged as key regulators of the organellar post–transcriptional processing. PPR proteins represent a large family in plants, and the extent to which PPR functions are conserved between dicots and monocots deserves evaluation, in light of differences in photosynthetic metabolism (C3 vs. C4) and localization of chloroplast biogenesis (mesophyll vs. bundle sheath cells). In this work we investigated the role played in the process of chloroplast biogenesis by At5g42310, a member of the Arabidopsis PPR family which we here refer to as *At*CRP1 (Chloroplast RNA Processing 1), providing a comparison with the orthologous *Zm*CRP1 protein from *Zea mays*. Loss-of-function *atcrp1* mutants are characterized by yellow-albinotic cotyledons and leaves owing to defects in the accumulation of subunits of the thylakoid protein complexes. As in the case of *Zm*CRP1, *At*CRP1 associates with the 5′ UTRs of both *psaC* and, albeit very weakly, *petA* transcripts, indicating that the role of CRP1 as regulator of chloroplast protein synthesis has been conserved between maize and Arabidopsis. *At*CRP1 also interacts with the *petB-petD* intergenic region and is required for the generation of *petB* and *petD* monocistronic RNAs. A similar role has been also attributed to *Zm*CRP1, although the direct interaction of *Zm*CRP1 with the *petB-petD* intergenic region has never been reported, which could indicate that *At*CRP1 and *Zm*CRP1 differ, in part, in their plastid RNA targets.

Keywords: PPR, anterograde signaling, chloroplast, biogenesis, RNA metabolism

INTRODUCTION

In land-plants, nuclear-encoded pentatricopeptide repeat (PPR) containing proteins constitute a large family, which regulates organelle gene expression at the RNA level (Lurin et al., 2004; O'Toole et al., 2008; Barkan and Small, 2014). They are, indeed, a major constituent of the genome-coordinating anterograde signaling pathway that evolved to adapt the expression of the organellar genomes in response to endogenous and environmental stimuli that are perceived by the nucleus (Woodson and Chory, 2008).

A typical PPR motif is characterized by a degenerate 35-amino acid repeat that folds into two antiparallel alpha helices (Small and Peeters, 2000). PPR proteins contain a tandem array of 2–30 PPR motifs, which stack together to form a superhelix with a central groove that allows the protein to bind RNA (Lurin et al., 2004; Rivals et al., 2006). According to the characteristics of their repeats, PPR proteins are generally classified into P and PLS sub-families. The P-type proteins are implicated in the determination and stabilization of 5′ and/or 3′ RNA termini, RNA splicing and translation of specific RNAs in chloroplasts and mitochondria, while PLS-type proteins are generally involved in RNA editing (Barkan and Small, 2014). Higher plants harbor several hundreds of PPR proteins, which generally have distinct, non-redundant functions in organelle biogenesis, plant growth and development and adaptation to environmental cues (Barkan and Small, 2014; Manna, 2015), as revealed by the high number of *ppr* mutants with distinct phenotypes. This is due to their ability to recognize primary RNA sequences, with each protein having different target sites, thus implying that the elucidation of the primary role of each PPR protein is greatly facilitated by the identification of its RNA targets.

The detection of few native PPR-RNA interactions through RNA immunoprecipitation on microarray (RIP-Chip) analyses and *in vitro* binding assays using PPR recombinant proteins, together with PPR crystal structures indicate that PPR proteins bind their cognate RNA targets in a sequence specific manner (Meierhoff et al., 2003; Schmitz-Linneweber et al., 2005, 2006; Williams-Carrier et al., 2008; Yin et al., 2013; Okuda et al., 2014; Shen et al., 2016). The code describing how PPR proteins recognize specific nucleotides of their RNA targets relies primarily on two amino acids that are within a single PPR motif, specifically the fifth residue in the first helix and the last residue on the loop interconnecting adjacent motifs (Barkan et al., 2012; Yin et al., 2013; Cheng et al., 2016). However, the current understanding of the code does not allow accurate large-scale computational predictions of PPR targets (Takenaka et al., 2013; Kindgren et al., 2015; Hall, 2016; Harrison et al., 2016). Predictive power is constrained by the fact that the code is degenerate and by the low accuracy of current methods used for the identification of PPR domains, which in turn leads to mismatches in the amino acid/nucleotide alignments. However, a more robust annotation of PPR domains has recently been conducted and made available at the PlantPPR database[1] (Cheng et al., 2016). Furthermore, more PPR-RNA interactions as well as crystal structures of PPR-RNA complexes need to be characterized in different species in order to improve the understanding of the code. This would also help to determine if the amino acid sequences of the PPR domains coevolved with the nucleotide sequences of their RNA targets and ultimately to determine whether there is functional conservation of PPR proteins among land plants.

The function of PPR proteins, and more generally the function of the nuclear gene complement involved in organellar RNA metabolism, have been primarily studied in maize, since the large seed reserves of maize support rapid heterotrophic growth of non-photosynthetic mutants and provide ready access to non-photosynthetic tissues for molecular biology and biochemical studies (Belcher et al., 2015). However, the degree of functional conservation of PPR proteins between maize and other species, including *Arabidopsis thaliana*, has yet to be investigated. The question is of particular interest since the elaboration of the thylakoid membrane system and the biogenesis of the multi-subunit photosynthetic complexes appear to have major differences between monocotyledonous and dicotyledonous plants (Pogson et al., 2015). Indeed in maize, and more generally in monocots, the process of chloroplast development from the proplastid to functional chloroplasts can be observed as a gradient along the leaf blade, whereas in dicots, such as *Arabidopsis thaliana*, the development of chloroplasts differs between developmental stages, plant organs – i.e., chloroplast development is different in cotyledons and leaves – and plant tissues (Pogson and Albrecht, 2011; Jarvis and Lopez-Juez, 2013).

The majority of PPR proteins are conserved at sequence level between dicots (Arabidopsis) and monocots (rice) (O'Toole et al., 2008). Orthologous pairs can readily be identified and in a number of cases, primary sequence conservation can be traced back to the roots of all embryophytes (O'Toole et al., 2008). As a matter of fact, functional differences between orthologous PPR proteins of maize and Arabidopsis have been observed. For example, the molecular phenotypes resulting from loss of the orthologous PPR proteins ATP4 (maize) and SVR7 (Arabidopsis) differ substantially (Liu et al., 2010; Zoschke et al., 2012, 2013a,b), as do the molecular defects in maize and Arabidopsis mutants lacking the PGR3 protein (Yamazaki et al., 2004; Cai et al., 2011; Belcher et al., 2015). Thus, the extent to which lessons on PPR proteins learnt from maize can be extrapolated to dicots, such as Arabidopsis, and more broadly to other organisms, needs further investigation.

In this context, we investigated here the function of and identified the RNA targets of the PPR protein At5g42310 from *Arabidopsis thaliana*, that shares high similarity with the well-characterized CRP1 (Chloroplast RNA Processing 1) protein from maize (*Zm*CRP1), and which we here refer to as *At*CRP1. Our findings indicate that *At*CRP1, like the orthologous *Zm*CRP1 (Barkan et al., 1994; Fisk et al., 1999; Schmitz-Linneweber et al., 2005), is essential for plant autotrophy since it plays a direct role in the accumulation of the cytochrome b_6/f (Cyt b_6/f) complex and of the PsaC subunit

[1] http://www.plantppr.com

of photosystem I (PSI). Furthermore *At*CRP1, similarly to *Zm*CRP1, is required for the accumulation of *petB* and *petD* monocistronic RNAs, indicating that the functional roles of CRP1 proteins are highly conserved between monocots and dicots.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Arabidopsis thaliana atcrp1-1 (SALK_035048) (Alonso et al., 2003) and *atcrp1-2* (SAIL_916A02) (Sessions et al., 2002) T-DNA insertion lines were identified by searching the T-DNA Express database[2]. For promoter analyses, the putative *AtCRP1* promoter region (*AtCRP1p*, −1062 to −2 upstream the translation starting codon) was cloned into pBGWFS7 destination vector and introduced into Arabidopsis wild type background, ecotype Columbia-0 (Col-0), by *Agrobacterium tumefaciens*-mediated transformation. *At*CRP1-GFP transgenic lines were obtained by transformation of *AtCRP1*/*atcrp1-1* heterozygous plants with either the *At*CRP1 coding sequence fused to GFP under the control of *35S-CaMV* promoter, cloned into pB7FWG2 vector, or the genomic locus fused to GFP under the control of the native promoter, cloned into a modified pGreenII vector (Gregis et al., 2009). The GUN1 coding sequence, devoid of the stop codon, was cloned into pB7RWG2 vector, carrying an RFP reporter gene. pB7FWG2, pBGWFS7, and pB7RWG2 plasmids were obtained from Flanders Interuniversity Institute for Biotechnology of Gent (Karimi et al., 2002). Primers used for amplification of the DNA fragments cloned into the vectors, reported above, are listed in **Supplementary Table S2**. Arabidopsis Col-0 and mutant plants were grown on soil under controlled growth chamber conditions with a 16 h light/8 h dark cycle at 22°C/18°C. In the case of mesophyll protoplast preparation, Arabidopsis plants were also grown on soil in a growth chamber under the above reported conditions. Moreover, phenotypic characterization and molecular biology analyses were also conducted on plants grown on Murashige and Skoog (MS) medium (Duchefa)[3], supplemented with or without 1% (w/v) sucrose. Tobacco plants, employed for transient gene expression, were cultivated for 5–6 weeks in a greenhouse under a 12 h light/12 h dark cycle at 22°C/18°C.

Protoplast Transformation

Mesophyll protoplasts of *Arabidopsis thaliana* (Col-0) were isolated and transiently transformed according to Yoo et al. (2007) and Costa et al. (2012). Briefly, well-expanded rosette leaves from 3-to-5 week-old plants were cut into strips of 0.5–1 mm with a fresh razor blade. Leaf tissue was digested using an enzyme solution containing 1.25% cellulase Onozuka R-10 (Duchefa) and 0.3% Macerozyme R-10 (Duchefa) for 3 h at 23°C in the dark. The protoplast suspension was filtered through a 50 μm nylon mesh washed three times with W5 solution (154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 2 mM MES, pH 5.7 adjusted with KOH) and used for PEG-mediated transformation. For each protoplast transformation 10 μg of a MidiPrep purified DNA (QIAGEN) plasmid harboring the *35S-CaMV*::*AtCRP1-GFP* cassette was used. Protoplasts were maintained for 16–24 h at 23°C in the dark, before performing epifluorescent microscopy.

Transient Expression in *Nicotiana benthamiana* Leaves

Tobacco leaf infiltration was performed using *A. tumefaciens* strain GV3101/pMP90 carrying the specified constructs (see results for details) together with the p19-enhanced expression system (Voinnet et al., 2003), according to the method described by Waadt and Kudla (2008). The final OD_{600} for *A. tumefaciens* strains harboring *35S-CaMV*::*AtCRP1-GFP* and *35S-CaMV*::*GUN1-RFP* was 0.2 and 0.3, respectively. After infiltration, plants were incubated for 3–5 days under the conditions described above.

Confocal Microscopy Analysis

Confocal Scanning Laser Microscopy analyses were performed using an inverted microscope, Leica DMIRE2, equipped with a Leica TCS SP2 laser scanning device (Leica). For the simultaneous detection of GFP and chlorophyll auto-fluorescence the cells were excited (Arabidopsis mesophyll protoplasts or tobacco leaf cells) with the 488 nm line of the Argon laser and the emissions were collected between 515/535 and 650/750 nm, respectively. For RFP detection the cells were excited at 561 nm from a He/Ne laser and the emission was collected between 575/625 nm. Image analyses were performed with Fiji[4]: an open-source platform for biological-image analysis (Schindelin et al., 2012).

Nucleic Acid Analyses

Arabidopsis DNA was isolated according to Ihnatowicz et al. (2004). Isolation of total RNA from homozygous *atcrp1-1* plants at four-leaf rosette stage and RNA gel blot analyses were performed as described by Meurer et al. (2002), using 10 μg of total RNA for each sample. For the RNA slot blot hybridization experiments, one-fourth of the RNA purified from each immunoprecipitation pellet and one-tenth of the RNA purified from the corresponding supernatant were applied to a nylon membrane with a slot-blot manifold and hybridized to specific radiolabeled probes (see **Supplementary Table S2**). ^{32}P-labeled DNA probes, complementary to chloroplast genes, were amplified using the primer pairs listed in **Supplementary Table S2**. Four micrograms of total RNA, treated with TURBO DNA-*free* (Ambion by Life Technologies), were employed for first-strand cDNA synthesis using GoScript Reverse Transcription System (Promega) according to the supplier's instructions. Quantitative Real-Time PCR (qRT-PCR) was carried out on an CFX96 Real-Time system (Bio-Rad), using the primer pairs

[2]http://signal.salk.edu/cgi-bin/tdnaexpress

[3]http://www.duchefa.com

[4]https://fiji.sc/

reported in **Supplementary Table S2**. The *SAND* (Remans et al., 2008) and *ubiquitin* transcripts were used as internal references. Data from three biological and three technical replicates were analyzed with Bio-Rad CFX Manager software (V3.1).

Immunoblot Analyses

For immunoblot analyses, total proteins were prepared as described by Martinez-Garcia et al. (1999). Total proteins, corresponding to 5 mg of leaf fresh-weight (100% of WT and *atcrp1-1* samples) and isolated from plants at four-leaf rosette stage, were fractionated by SDS–PAGE (12% acrylamide [w/v]; (Schagger and von Jagow, 1987). Proteins were then transferred to polyvinylidene difluoride (PVDF) membranes (Ihnatowicz et al., 2004) and replicate filters were immunodecorated with antibodies specific for PSI (PsaA, PsaC, and PsaD), PSII (D1, PsbO) Cyt b_6/f (PetA, PetB, and PetC), ATPase (ATPase-β) subunits, PSI (Lhca1, Lhca2) and PSII (Lhcb2, Lhcb3) antenna proteins, all obtained from Agrisera[5]. The GFP antibody was purchased from Life Technologies[6].

Chloroplast Stromal Preparation and Protein Immunoprecipitation

Intact chloroplasts were isolated from 11 days old Arabidopsis plants, according to Kunst (1998), and Kupsch et al. (2012) with some modifications. Chloroplasts were directly resuspended in 300–400 µl of extraction buffer [2 mM DTT, 30 mM HEPES-KOH, pH 8.0, 60 mM KOAc, 10 mM MgOAc and proteinase inhibitor cocktail (Sigma–Aldrich-P9599)]. Two independent stromal preparations were carried out and one of them was performed in the presence of 2% sodium deoxycholate in order to solubilize the membrane-attached *At*CRP1 protein fraction. Chloroplasts were then disrupted by pulling them through a syringe (0.55 mm × 40 mm) 30–40 times. The solution was centrifuged at 21,000 × *g* at 4°C to separate the stromal from the membrane fraction.

The isolated stromal fraction was diluted with one volume of coimmunoprecipitation (CoIP) buffer (150 mM NaCl, 20 mM Tris-HCl pH 7.5, 2 mM $MgCl_2$, 0.5% Nonidet P-40 and 0.5 µg/mL Aprotinin). Five microliters of mouse anti-GFP antibody (Roche, No. 11814460001) were added to the stromal fraction and incubated for 1 h at 4°C and 13 rpm on an overhead shaker. Thereafter the coimmunoprecipitation was performed as described by Kupsch et al. (2012). Successful precipitation of *At*CRP1-GFP was confirmed by immunoblot analyses, using the same GFP antibody.

RNA Extraction and Labeling for RIP-Chip Assay

RNA immunoprecipitation-chip analyses were performed using a tiling microarray covering the complete Arabidopsis chloroplast genome (Kupsch et al., 2012). The coimmunoprecipitated RNA was isolated from pellet and supernatant fractions either by phenol-chloroform extraction or using the Direct-zol™ RNA MiniPrep kit (Zymo Research). For the phenol-chloroform extraction, RNA samples were incubated in 1% SDS and 5 mM EDTA at room temperature for 5 min to dissociate RNA-protein complexes. RNA was phenol-chloroform extracted, ethanol precipitated with the addition of Glycoblue™ Coprecipitant (Thermo Fisher Scientific), washed twice with 75% ethanol, air-dried and resuspended in 20 µl RNase-free water. For the replicate, RNA was extracted using the Direct-zol™ RNA MiniPrep kit (Zymo Research) according to the manufacturer's instructions. Before the extraction 2 µg yeast RNA was added to the coimmunoprecipitated RNA pellet. The entire RNA of the pellet fraction and 2 µg RNA of the supernatant fraction were used for labeling. The pellet and supernatant RNA were labeled with 0.5 µl Cy5 and 1 µl Cy3 dye, respectively (aRNA labeling kit, Kreatech Diagnostics). Labeling reaction, microarray hybridization, scanning, and evaluation were performed as described in Kupsch et al. (2012). Only PCR products for which more than half of all replicate spots (24 per PCR product spanning two experiments) passed our quality assessment (Kupsch et al., 2012) and were used in this analysis (**Supplementary Table S1**).

In silico Prediction of AtCRP1 Binding Sites

The putative *At*CRP1 binding motif, i.e., the nucleotide preference for each of the amino acid pairs at the fifth and last position of PPR domains, was predicted *in silico* using the reported weighting schemes (Barkan et al., 2012; Barkan and Small, 2014; Harrison et al., 2016). The software FIMO[7], which analyzes sequence databases for occurrences of known motifs (Grant et al., 2011), was employed to identify the potential binding sites of *At*CRP1 within the regions enriched in our RIP-Chip experiment. Furthermore, the same regions were searched for the presence of sRNA native footprints, by consulting the JBrowse sRNA database[8] (Ruwe et al., 2016). Numbers that delimit the native footprints refer to the chloroplast genome of *Arabidopsis thaliana* (NC_000932.1).

β-Glucuronidase (GUS) Assay

For GUS histochemical detection, plant material was fixed in 90% acetone at −20°C for 1 h. Samples were then washed three times with NaPi buffer (NaH_2PO_4 50 mM, Na_2HPO_4 50 mM; pH 7.0) and stained overnight at 37°C with *X*-gluc solution [1 mM 5-bromo-4-chloro-3-indolyl-β-D-glucoronide, 2 mM $K_3/K_4Fe(CN)_6$, 0,1% Triton (v/v), 10 mM EDTA, 50 mM NaPi pH 7.0]. 70% EtOH (v/v) was used as washing solution. Stained samples were then stored at 4°C and observed using a Zeiss Axiophot D1 microscope equipped with differential interference contrast (DIC) optics. Images were recorded with an Axiocam MRc5 camera (Zeiss) using the Axiovision program (v.4.1).

[5]http://www.agrisera.com/en/artiklar/plantalgal-cell-biology/index.html
[6]http://www.thermofisher.com
[7]http://meme-suite.org/tools/fimo
[8]https://www.molgen.hu-berlin.de/projects-jbrowse-athaliana.php

FIGURE 1 | Primary amino acid sequence alignment of *At*CRP1 and *Zm*CRP1 proteins. The amino acid sequence of the Arabidopsis CRP1 (*At*CRP1, At5g42310) was compared with CRP1 from *Zea mays* (*Zm*CRP1), using ClustalW2. Black boxes indicate strictly conserved amino acids, and gray boxes closely related ones. The predicted chloroplast transit peptides (ChloroP, http://www.cbs.dtu.dk/services/ChloroP/) are indicated in italics, and the PPR motives (P0-to-P14), identified using the PlantPPR database (http://www.plantppr.com), are marked with gray bars. The specificity determining amino acids in each PPR motif at position 5 and 35 are indicated by black and gray stars, respectively. Note that P0 motif was not considered to contribute to the identification of RNA targets, as previously reported by Barkan et al. (2012). P0 is composed of 30 aa, whereas all other P motifs are of 35 aa, with the exception of P2, wich contains 37 aa in Arabidopsis and 38 in maize.

RESULTS

*At*CRP1 Is a PPR Protein Imported into the Chloroplast

The Maize Genetics and Genomics Database (Lawrence et al., 2004)[9] was used to identify the At5g42310 gene as the Arabidopsis ortholog of *ZmCRP1* (see also Belcher et al., 2015). At5g42310 encodes a polypeptide of 709 amino acids with a calculated molecular mass of 80 kDa. Intron number (three) and position are conserved between the two genes, and BLASTP query of public Arabidopsis sequence database with *Zm*CRP1 amino acid sequence detected At5g42310 protein as the top hit with 55% sequence identity and 72% sequence similarity (**Figure 1**).

*At*CRP1 is annotated as a PPR protein and shares with *Zm*CRP1 15 PPR tandem repeats, which were predicted by using the PlantPPR database (Cheng et al., 2016). All PPR motifs are of 35 aa, with the exception of P0 which consists of 30 aa and P2 of 37 aa in Arabidopsis and 38 aa in maize. The fifth and the last residue of each PPR domain form the amino acid pairs that specify the RNA target molecules (Cheng et al., 2016), and are labeled with gray and black stars in **Figure 1**. The ChloroP server (Emanuelsson et al., 1999)[10] predicted the presence of a cTP of 54 residues (see amino acid residues in italics in **Figure 1**), indicating that *At*CRP1, like *Zm*CRP1, could be imported into the chloroplast. To corroborate the *in silico* prediction, the *At*CRP1-GFP fusion protein was expressed in transiently transformed Arabidopsis protoplasts (**Figure 2**). In agreement with the ChloroP prediction, the chimeric protein (GFP fluorescence) accumulated within the chloroplast in distinct fluorescent foci (CHL, autofluorescence of chloroplast chlorophylls, **Figure 2A**), resembling the nucleoid complexes. Indeed, *At*CRP1-GFP chimera co-localized perfectly with the GUN1-RFP fusion protein, used as a nucleoid marker in this assay (RFP fluorescence, **Figure 2B**), (Koussevitzky et al., 2007; Colombo et al., 2016; Tadini et al., 2016), in tobacco leaf cells. To further localize *At*CRP1, chloroplasts were fractionated to separate the stroma and thylakoid compartments. Immunoblot analysis, using a GFP specific antibody, allowed detection of *At*CRP1-GFP specific signal in total chloroplasts, as well as in thylakoids and in the stromal fraction, indicating that the nucleoid *At*CRP1 protein is both associated to membranes and soluble in the stroma (**Figure 2C**). These findings are in agreement with the identification of *At*CRP1 as part of Megadalton complexes in the chloroplast stroma (Olinares et al., 2010), as well as in the grana of thylakoid membranes (Tomizioli et al., 2014).

*At*CRP1 Is Essential for Plant Autotrophy

To investigate the role that *At*CRP1 plays in Arabidopsis, two lines carrying T-DNA insertions into the coding sequence of At5g42310, renamed *atcrp1-1* (Salk_035048) and *atcrp1-2*

[9]http://www.maizegdb.org/

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

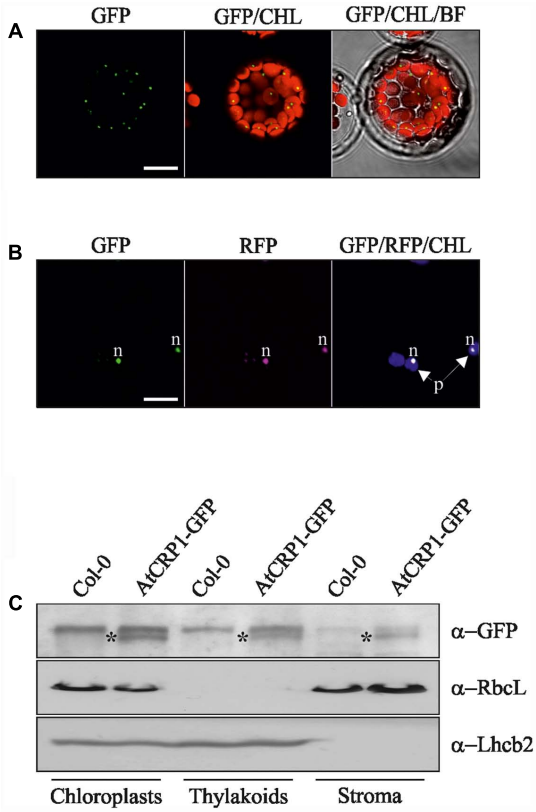

FIGURE 2 | Subcellular localization of *At*CRP1 in Arabidopsis mesophyll protoplasts and leaf cells. (A) Series of Lasers Scanning Confocal images (CLSM) of the subcellular localization of the *At*CRP1-GFP fusion protein (indicated as GFP) expressed in transiently transformed Arabidopsis (ecotype Col-0) leaf mesophyll protoplasts. The GFP signal accumulates in distinct spots within the chloroplasts, visualized by the red chlorophyll autofluorescence (CHL), resembling the pattern of chloroplast nucleoids. BF, Bright Field. **(B)** Series of CLSM images of the subcellular localization of *At*CRP1-GFP and GUN1-RFP [indicated as RFP and used as a marker of chloroplast nucleoids (Koussevitzky et al., 2007)] fusion proteins upon transient co-expression in tobacco leaf cells. The green fluorescence (GFP) co-localizes perfectly with the purple fluorescence (RFP) inside the chloroplasts (violet autofluorescence of chlorophylls, CHL), indicating that *At*CRP1 protein is part of the chloroplast nucleoids. Images are representative of three independent experiments. Bar = 10 μm; *p* = chloroplast; *n* = nucleoid. **(C)** Immunoblot analyses of proteins from Col-0 and Arabidopsis transgenic lines containing the *AtCRP1-GFP* construct under the control of *AtCRP1* native promoter (approximately 1 kb upstream of the translation start codon, see also Materials and Methods). Equal protein amounts isolated from total chloroplasts, thylakoids and stroma were loaded. Filters were immunolabeled with a GFP specific antibody to detect the localization of the *At*CRP1-GFP chimera. An antibody specific for the large subunit of RUBISCO (RbcL) was used as a marker of chloroplast stroma, whilst an Lhcb2 specific antibody was used as a marker of thylakoid membranes. Asterisks indicate the position of the AtCRP1-GFP fusion protein. One out of three immunoblots for each antibody is shown. Note that the *At*CRP1-GFP chimera is fully functional, since it was able to rescue the *atcrp1-1* mutant phenotype (see also **Figure 4**).

(Sail_916A02), were obtained from the T-DNA Express Arabidopsis mutant collection (**Figure 3A**; see also Materials and Methods).

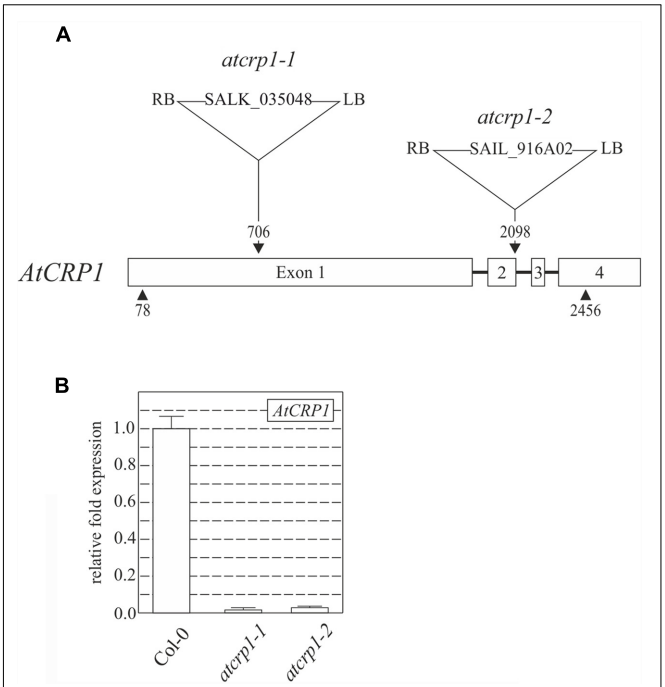

FIGURE 3 | T-DNA tagging and expression levels of *AtCRP1* gene. **(A)** Schematic representation of *AtCRP1* gene, where exons are indicated as numbered white boxes, while introns are shown as black lines. Arrowheads indicate the positions of translation initiation and stop codons. The locations, designations and orientations of T-DNA insertions are indicated (RB, right border; LB, left border). Note that the T-DNA insertions are not drawn to scale. **(B)** Levels of *AtCRP1* gene expression was ascertained by real-time PCR of cDNA obtained from leaves of WT (Col-0) and *atcrp1-1*, *atcrp1-2* mutant plants. Gene expression was normalized to the level of *AtCRP1* transcripts in Col-0 plants, and *SAND* and *ubiquitin* were used as internal references. The bars indicate standard deviations.

Both T-DNA insertions completely suppressed the accumulation of the corresponding transcripts in homozygous mutant seedlings (**Figure 3B**), which were characterized by a paler pigmentation of cotyledons, visible even at the fully mature embryo stage (**Figure 4A**), and leaves (**Figures 4B,C**), and found to be seedling lethal under autotrophic growth conditions on soil and MS medium without sucrose, but able to develop yellow-albinotic rosette leaves and sterile inflorescence when sucrose was provided in the medium (**Figure 4C**). The mutant phenotype could be rescued by *Agrobacterium tumefaciens*-mediated transformation of heterozygous plants with either the appropriate coding sequence fused to the 35S promoter of cauliflower mosaic virus (*35S-CaMV*::*AtCRP1-GFP*), or the genomic sequence including a 1-Kbp fragment of the promoter region (*AtCRP1p*::*AtCRP1-GFP*), corroborating a direct correspondence between genotype and phenotype, and indicating that the AtCRP1-GFP chimera was fully functional, in both cases (**Figure 4D**). Interestingly, complemented plants carrying the *AtCRP1-GFP* construct under the control of the native promoter showed a fivefold increase in *AtCRP1* gene expression (**Figure 4E**), most probably as consequence of the T-DNA insertion in a highly expressed

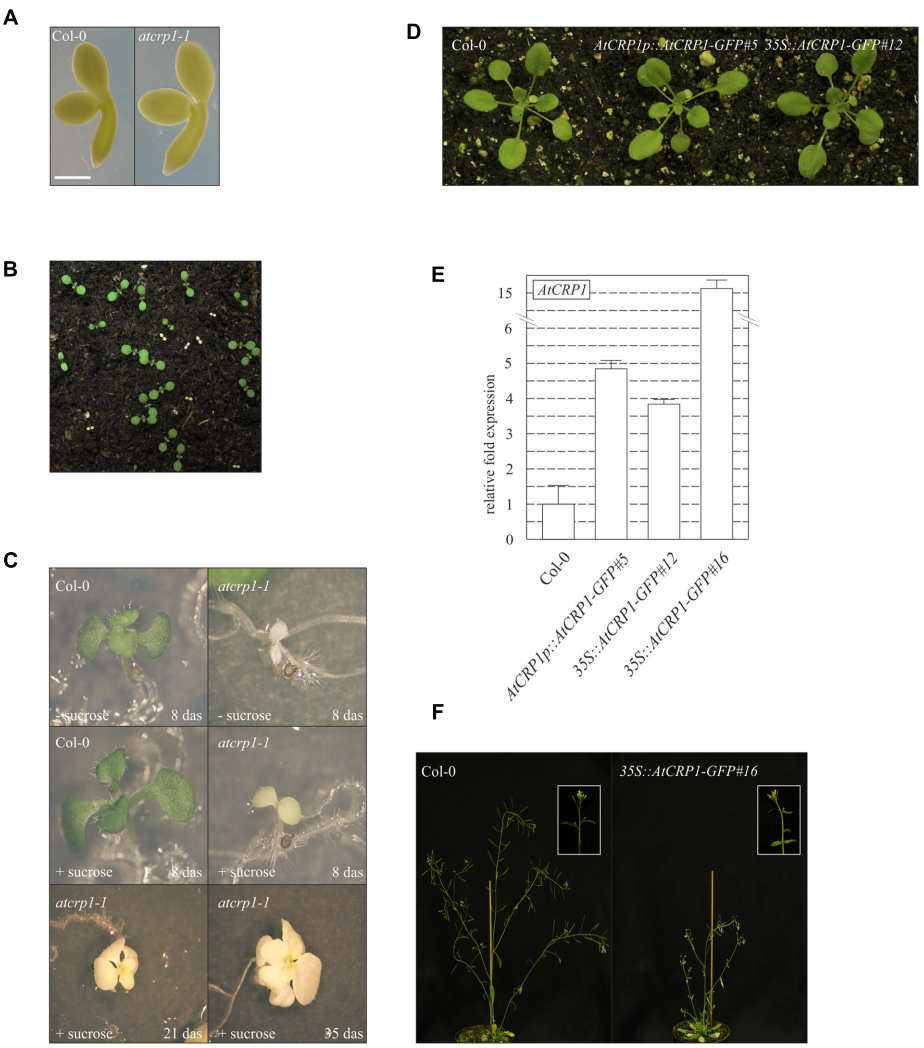

FIGURE 4 | Effects of loss of *At*CRP1 on plant development. (A) Images of isolated fully mature embryos (bent cotyledon stage) from WT (Col-0) and *atcrp1-1* seeds. The lack of *At*CRP1 protein did not alter embryo development, although mutant embryos were slight larger and paler than Col-0. **(B)** *atcrp1-1* seeds were able to germinate on soil, giving rise to yellow seedlings that accounted for about one-quarter of all seedlings, indicative of a monogenic recessive trait. Mutant seedlings did not survive past the cotyledon stage. **(C)** Mutant seedlings showed albino cotyledons when grown on MS medium without sucrose and arrested at the cotyledon stage as in **(B)**. However, when *atcrp1-1* seedlings were grown on MS medium supplement with 1% sucrose, they showed yellow-albinotic cotyledons at 8 das (days after sowing) and were able to develop up to 8–10 true leaves after 35 das. **(D)** The *atcrp1-1* seedling lethal phenotype could be fully rescued by *Agrobacterium tumefaciens*-mediated transformation of *AtCRP1*/*atcrp1-1* heterozygous plants with either the *AtCRP1* coding sequence fused to GFP under the control of 35S-CaMV promoter (*35S::AtCRP1-GFP#12*), or the genomic sequence fused to GFP under the control of native promoter (*AtCRP1p::AtCRP1-GFP#5*). **(E)** Real-time PCR to monitor the expression of *AtCRP1* gene in WT and complemented plants. Gene expression was normalized with respect to the level of *AtCRP1* transcripts in Col-0, and *SAND* and *ubiquitin* were used as internal references. The bars indicate standard deviations. **(F)** Col-0 and *35S::AtCRP1-GFP#16* transgenic line with about 15-folds more *AtCRP1* transcripts than WT. In this case the transgenic line shows WT-like rosette, but it is characterized by shorter and paler stems, with bleached cauline leaves and sterile flowers. A detail of the stem and inflorescence is shown in the inset. Note that the detailed molecular characterization of *At*CRP1 function was conducted on *atcrp1-1* plants, since the *atcrp1-2* seedlings showed an identical phenotype.

euchromatin region of the nuclear genome. Furthermore, a complete rescue of mutant plant phenotype could only be observed in *35S::AtCRP1-GFP* transgenic lines with a limited accumulation of *AtCRP1* transcripts (**Figures 4D,E**). Higher *AtCRP1* expression levels (around 15-folds in comparison to WT) led to transgenic plants with WT-like rosette but shorter and paler stems, bleached cauline leaves, together with sterile flowers (**Figures 4E,F**).

Temporal and spatial expression patterns of *At*CRP1, monitored by fusing the promoter region of the gene upstream of the GUS reporter gene (see also Materials and Methods), support further the key role played by *At*CRP1 during early stages of seedling and leaf development (**Figure 5**). The GUS staining could, indeed, be detected in young cotyledons and in the upper portion of the hypocotyl (**Figure 5A**). Furthermore, intense GUS signals were observable in young developing leaves

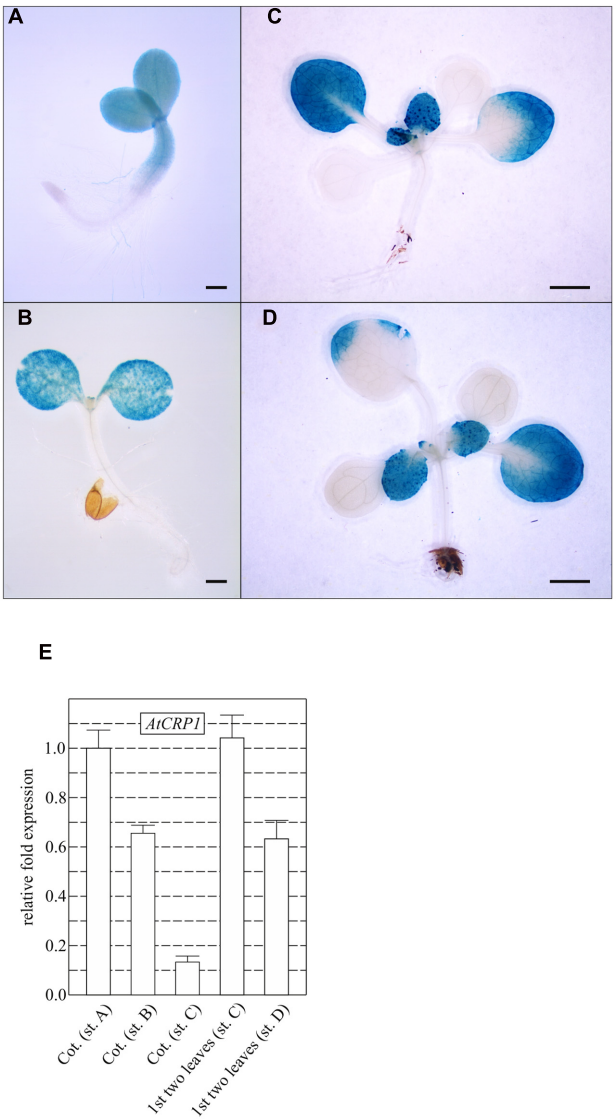

FIGURE 5 | *AtCRP1* promoter-driven β-glucuronidase (GUS) activity in cotyledons and rosette leaves. Histochemical GUS staining was conducted on seedlings at the two cotyledon stage **(A)**, at the onset of the first true leaves **(B)**, at four leaves rosette-stage **(C)**, and at the onset of the third pair of true leaves **(D)**. In general, GUS staining in younger leaves was stronger than in older leaves and the activity of *AtCRP1* promoter was below the limit of detection in cotyledons after the development of the first true leaves. **(A)** and **(B)** Bar = 1 mm, **(C)** and **(D)** bar = 1 cm. **(E)** Real-time PCR analyses were conducted with cDNA obtained from cotyledons at the developmental stages reported in **(A–C)** (Cot. st. A, Cot. st. B and Cot. st. C) and on the first pair of true leaves at stages C–D (st. C and st. D) to monitor the accumulation of *AtCRP1* transcripts. Gene expression was normalized with respect to the level of *AtCRP1* transcripts in cotyledons at stage A, and *SAND* and *ubiquitin* were used as internal references. The bars indicate standard deviations.

(**Figures 5C,D**), whereas the GUS coloration tended to decrease in old cotyledons and leaves (**Figures 5B–D**). Similar results were also obtained by monitoring the expression of *AtCRP1* in cotyledons and leaves using quantitative Real-Time PCR (qRT-PCR). In general, a high level of expression of *AtCRP1* was observed in green developing tissues, such as young cotyledons and leaves, whereas the expression decreased in older tissues (**Figure 5E**).

atcrp1 Mutant Chloroplasts Fail to Accumulate Cytochrome b_6/f Protein Complex and the PsaC Subunit of PSI

The albino pigmentation of *atcrp1* seedlings, together with their inability to grow under autotrophic conditions, indicated a defect in the thylakoid-associated photosynthetic apparatus. To verify this assumption, immunoblot analyses with antibodies specific for single subunits of the four major thylakoid protein complexes were performed on total leaf proteins. Leaf samples were harvested from *atcrp1* plants at the four-leaf rosette stage and grown on MS-medium supplemented with 1% sucrose (**Figure 6**; see also Materials and Methods). Under standard light conditions (50 μmol photons m^{-2} s^{-1}), subunits of Photosystem I (PsaA, PsaC, and PsaD), Photosystem II (D1, PsbO), Light harvesting complexes (Lhca1, Lhca2, Lhcb2, and Lhcb3) and ATPase (ATPase-β) accumulated to levels lower than 10% with respect to wild type plants. Furthermore, subunits of the Cyt b_6/f (PetA, PetB, and PetC) and PSI (PsaC) were below the limits of immunoblot detection.

In summary, these results indicate a general reduction of thylakoid protein complex subunits in *atcrp1* leaves, with a particularly severe effect on the accumulation of the Cyt b_6/f complex and PsaC.

AtCRP1 Is Associated *In vivo* with *psaC* and *petB-petD* Transcripts

*Zm*CRP1 has been previously demonstrated to associate with the *psaC* and *petA* mRNAs *in vivo* by RIP-Chip analyses (Schmitz-Linneweber et al., 2005). To investigate whether *At*CRP1 shares with *Zm*CRP1 the RNA targets, the same RIP-Chip approach employed in maize was used here. Stroma from plants expressing *At*CRP1-GFP, under the control of the native promoter (*AtCRP1p*::*AtCRP1-GFP*), was isolated and the fusion protein was immunoprecipitated using an anti-GFP serum. As a control, we performed mock precipitations with stroma extracted from WT plants, using the same GFP antibody. RNA was purified from the immunoprecipitation pellets and supernatants and was labeled with Cy5 (red) and Cy3 (green) fluorescent dyes, respectively. The two RNA fractions from *At*CRP1-GFP immunoprecipitations (IPs) and from mock IPs were competitively hybridized to a chloroplast genome tiling microarray (Kupsch et al., 2012). Enrichment of RNA is reflected in the ratio of red to green fluorescence for each spot on the microarray. Two biological replicate experiments were performed with stroma from *At*CRP1-GFP expressing plants and two with WT stroma. Data from the four assays were normalized and used to calculate median enrichment ratios of the red and green fluorescence signals for each PCR product among the 24 replicate spots on two arrays (**Supplementary Table S1**). To identify enrichment of RNA species specifically in the *At*CRP1-GFP immunoprecipitation,

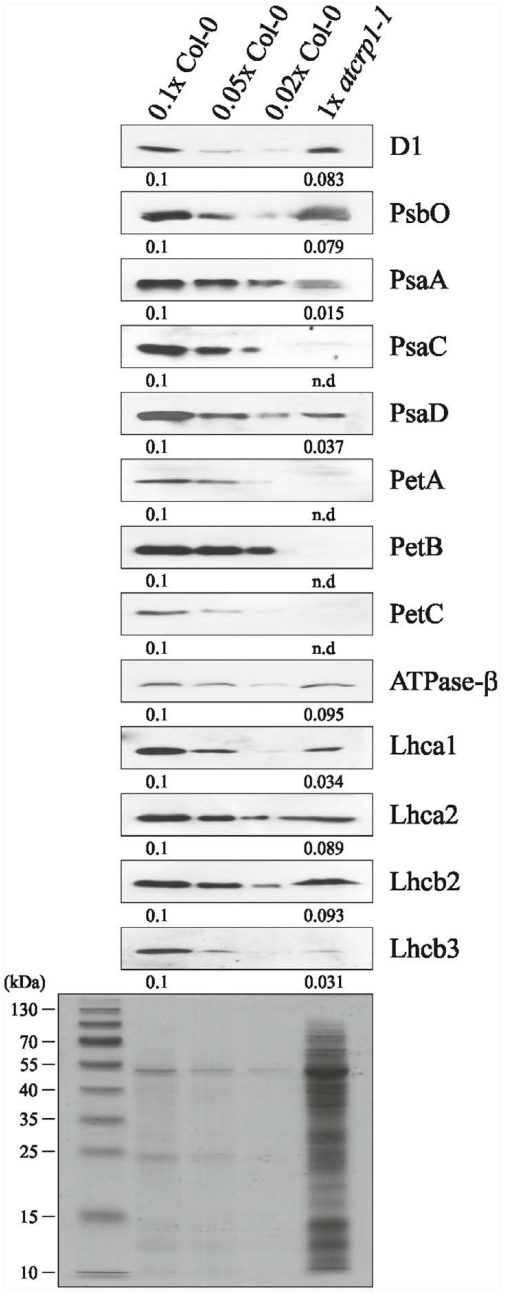

FIGURE 6 | Immunoblot analyses of thylakoid protein complexes in Col-0 and *atcrp1-1* mutant leaves. PVDF filters bearing fractionated total proteins, isolated at the four-leaf rosette stage from Col-0 and *atcrp1-1* plants grown on MS medium supplemented with 1% sucrose (see also **Figure 4**), were probed with antibodies raised against individual subunits of PSII (D1, PsbO), PSI (PsaA, PsaC, and PsaD), Cyt b_6f (PetA, PetB, and PetC), ATPase (ATPase-β), LHCI (Lhca1, Lhca2) and LHCII (Lhcb2, Lhcb3). Reduced levels of Col-0 total proteins were loaded in the lanes marked 0.1x Col-0, 0.05x Col-0, and 0.02x Col-0 in order to obtain signals from Col-0 proteins within the range of mutant protein signals (1x *atcrp1-1*). A replica SDS-PAGE stained with Coomassie-brilliant-blue is shown as loading control. Averaged relative protein abundance is given below each immunoblot and standard deviation was less than 10%. One out of three immunoblots for each antibody is shown. Note that the complete lack of Cyt b_6f and PsaC subunits was also observed in *atcrp1-2* leaves. n.d., not detected.

we plotted the difference in median enrichment ratio for each DNA fragment between the *At*CRP1-GFP and mock experiment against the position of the product on the plastid chromosome (**Figures 7A,B**).

Four prominent peaks of differential enrichment were observed. One of them corresponds to the 5′UTR of *psaC* transcript, a target already recognized as a ligand of *Zm*CRP1 in RIP-Chip assays (Schmitz-Linneweber et al., 2005). A second RNA target is represented by the *petB-petD* intergenic region. This RNA was not identified to interact with *Zm*CRP1 by RIP-Chip analysis, however, *Zm*CRP1 is known to aid in maturation of this particular intergenic region (Barkan et al., 1994). Interestingly, the observed enrichment of *rps15* transcripts might uncover a further, novel target of *At*CRP1, whereas the enrichment of *psbM/trnD* transcripts is often observed in RIP-Chip experiments, thus this peak was considered an artifact.

To corroborate the RIP-Chip data, the *At*CRP1-associated RNAs were analyzed by slot blots (**Figure 7C**). RNA purified from immunoprecipitation pellets and supernatants were probed with the PCR fragments that detected the most highly enriched sequences in the RIP-Chip assay. The data confirmed that the *psaC* and *petB-petD* transcripts are highly enriched in the *At*CRP1-GFP immunoprecipitates, but not the *rps15* RNA. *Zm*CRP1 was also reported to be associated with RNAs of the *petA* region (Schmitz-Linneweber et al., 2005; Williams-Carrier et al., 2008), however, no enrichment of *petA* transcripts could be observed in the *At*CRP1-GFP RIP-Chip assay (**Figure 7A**) and a low enrichment was detected in the slot blot assay (**Figure 7C**), possibly indicating that the interaction of *At*CRP1 with *petA* transcripts is not very stable. In general, our analysis cannot exclude the possibility that CRP1 binds to additional target RNAs, for example when interactions take place at chloroplast membranes. Since we are not using cross-linked material, weak RNA-protein interactions might be lost during our assay.

To support further the RIP-Chip findings, *At*CRP1 target RNAs were interrogated for the presence of native footprints at the JBrowse database[11]. The JBrowse database provides annotations of *Arabidopsis thaliana* organellar short RNA (sRNA), thought to be generated from protein-mediated temporary protection of target RNAs against exonucleolytic degradation (Ruwe et al., 2016; see also **Figure 8**). sRNAs were found within the 5′UTR of *psaC* (corresponding to the 117633–117597 region of chloroplast genome) and the *petB-petD* intergenic region (region 76318–76358), and an sRNA was also annotated in the 5′UTR of *petA* (region 61615–61643). Furthermore, *At*CRP1 predicted RNA binding motifs were shown to co-map with the native footprints, when the corresponding sequences were searched for the occurrence of the consensus binding motif with the FIMO program in the MEME suite[12] (**Figure 8B**; Takenaka et al., 2013). A short RNA has been also mapped upstream of *rps15*, but this region was not enriched in the RIP-Chip assay and the match with the predicted binding site of *At*CRP1 is weaker than for the *psaC*, *petB-petD,* and *petA* sRNAs.

[11]https://www.molgen.hu-berlin.de/projects-jbrowse-athaliana.php
[12]http://meme-suite.org/tools/fimo

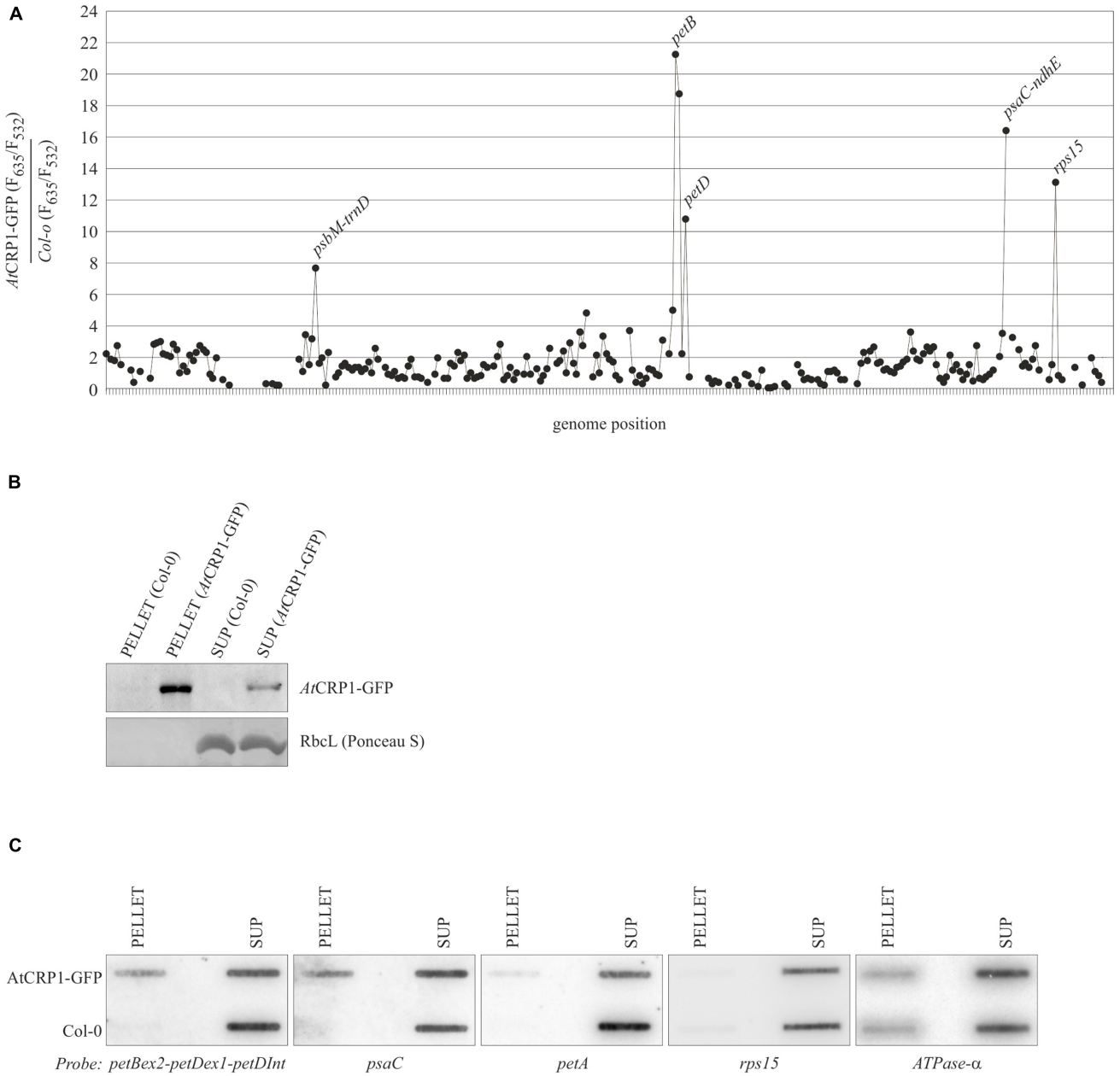

FIGURE 7 | *At*CRP1 RIP-Chip data plotted according to gene order within the plastid genome. (A) Differential enrichment ratios obtained by RNA immunoprecipitation (RIP)-Chip analysis. The enrichment ratios (F_{635}/F_{532}) obtained from an assay of *AtCRP1p::AtCRP1-GFP* chloroplast stroma extract were normalized with respect to a control assay that used WT (Col-0) chloroplast stroma extract (both assays were performed in duplicate). The median-normalized values for replicate spots were plotted according to gene order within the plastid genome. Fragments for which fewer than 13 spots per experiment (*AtCRP1p::AtCRP1-GFP*/WT) passed our manual quality control and/or yielded an F532 signal below background were excluded and appear as gap in the curve. The enrichment of *psaC* 5′UTR is in agreement with previous findings obtained by RIP-Chip analysis on CRP1 from maize (Schmitz-Linneweber et al., 2005). **(B)** Immunoblot analysis of protein fractions obtained from immunoprecipitation experiments using the anti-GFP mouse antibody and stroma material from Col-0 and *At*CRP1-GFP plants. Equal volumes of supernatant and pellet preparations were loaded onto the gel. Note that the pellet from *At*CRP1-GFP immunoprecipitation gave a stronger signal than the corresponding supernatant, implying quantitative precipitation of *At*CRP1-GFP. The fact that no signal was obtained with Col-0 extracts demonstrates the specificity of the antibody. The RbcL migration region of the Ponceau S stained nylon membrane, after transfer from SDS-PAGE, was used to verify equal loading. **(C)** Verification of *At*CRP1 RNA targets. Coimmunoprecipitations and RNA extractions from *At*CRP1-GFP and Col-0 samples were performed as for RIP-Chip assays. The RNAs were then analyzed by slot-blot hybridization with the indicated probes (see also Materials and Methods and **Supplementary Table S2**). The *ATPase-α* probe hybridization was included as a control. SUP, supernatant.

A

motif	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14
AtCRP1 aa pos. 5	N	S	N	V	N	S	S	N	N	N	T	N	N	T
AtCRP1 aa pos. 35	D	D	K	R	D	N	D	D	C	N	S	S	D	D
Predicted RNA b.s.	*U*	*G*	*Y*	-	***U***	***A***	*G*	***U***	***Y***	***Y***	***R***	***Y***	***U***	***G***
ZmCRP1 aa pos. 5	N	S	S	T	N	S	S	N	N	N	T	H	N	T
ZmCRP1 aa pos. 35	D	D	R	R	D	S	D	D	G	N	S	S	D	D
Predicted RNA b.s.	*U*	*G*	*R*	*R*	***U***	***A***	***G***	***U***	***Y***	***Y***	***R***	-	***U***	***G***

B

***petB-petD Intergenic Region* (+ strand):**

At 76318-CUUACUUAUUACUUGGUGAAGGAACGAUAG**UA**U**UUUAUUG**C-76358
Zm 74582-AUAU**CGGGUAG**GUUGUGG**UA**U**UU**C**AUUG**CU-74611

***5' UTR psaC* (-strand):**

At 117633-UUUUAAUAUACCAUUCAGUUAGA**AGUUUACU**AGAUUG-117597

***5' UTR petA* (+strand):**

At 61615-GCUA**AC**U**UUUAUUG**UAGAAAUUUUCGGGAU-61643

FIGURE 8 | *At*CRP1 RNA binding sites and the chloroplast *in vivo* footprints. (A) PPR motifs in *At*CRP1 were identified with the aid of PlantPPR database [www.plantppr.com, (Cheng et al., 2016). Amino acid residues in the 5th and last position of PPR motifs have been considered critical for sequence-specific RNA recognition, as previously reported (Barkan et al., 2012; Barkan and Small, 2014; Cheng et al., 2016; Harrison et al., 2016); see also **Figure 1**]. When the code developed for the different amino acid pairs is applied to the *At*CRP1 repeats, the sequence UGYNUAGUYYRYUG emerges as predicted RNA binding sequence (b.s.), whereas the sequence UGRRUAGUYYRNUG is predicted for *Zm*CRP1, in agreement with Barkan et al. (2012). **(B)** The sequences of *in vivo* footprints identified in the *petB-petD* intergenic region (Arabidopsis and Maize) and 5′UTR *psaC* region that co-map with *At*CRP1 binding sites (p-value < 0.01, highlighted in bold on a gray background) are shown. The Arabidopsis 5′UTR of *petA* transcripts shows also the presence of a native footprint that co-maps with *At*CRP1 binding site, however, this region was only enriched in the slot blot, but not in the AtCRP1 RIP-Chip assay (see **Figure 6**). There is no published sRNA within the *psaC* or *petA* 5′UTR of maize. A predicted binding site for maize CRP1 in the 5′-UTR of *psaC* (UGGAUAAACCAUUG; Barkan et al., 2012) is not similar in sequence to the Arabidopsis prediction shown here. Moreover, nucleic acid binding assay showed a direct interaction of *Zm*CRP1 with the 5′-UTR of *petA* (UUAGCUACCUAUCU**AA****UUUAUUG**UAGAAAUU; Williams-Carrier et al., 2008), that shows high similarity with the corresponding Arabidopsis sequence (see predicted binding site highlighted in bold). Note that no *At*CRP1-specific *in vivo* footprint could be identified in the other RIP-Chip enriched regions, *rps15* and *psbM* (see also **Figure 7**).

In summary, the RIP-Chip and slot blot data together with the colocalization of native footprints and *At*CRP1 RNA binding motifs indicate that *At*CRP1 likely binds directly to the 5′UTR of *psaC* and the *petB-petD* intergenic region and possibly to the 5′UTR of *petA*. On the contrary, the absence of an *At*CRP1-specific footprint within the *rps15* RNA, together with the failure of slot blot enrichment, makes any *At*CRP1-*rps15* interaction unlikely.

*At*CRP1 Is Required for the Correct Processing of *psbB-psbT-psbH-petB-petD* Transcripts

To assess whether the lack of Cyt b_6/f complex and PsaC subunit, together with the marked reduction of all protein complex subunits observed in *atcrp1-1* thylakoids, was caused by deficiencies in transcript accumulation and *At*CRP1-dependent transcript processing, we probed the identified *At*CRP1 RNA targets and other plastid transcripts by gel blot hybridization (**Figure 9**).

We investigated the transcripts encoding the subunits CP47 (*psbB*), T (*psbT*), and H (*psbH*) of photosystem II (PSII), subunits A (*psaA*) and C (*psaC*) of PSI, Cyt *f* (*petA*), Cyt b_6 (*petB*) and subunit IV (*petD*) of cytochrome b_6/f and the alpha subunit of ATPase (*ATPase-α*). All these transcripts accumulated in *atcrp1-1* plastids to levels lower than WT, indicating that global plastid gene expression is affected by the *atcrp1-1* mutation, and explaining the marked reduction of thylakoid protein accumulation observed in *atcrp1-1* leaves.

Furthermore, the plastid polycistronic transcription unit *psbB-psbT-psbH-petB-petD* showed some striking alteration of transcript pattern in *atcrp1* samples (**Figure 9**). In particular, the monocistronic *petB* (band #4, 0.8 Kb), the dicistronic *psbH-petB* (band #3; 1.2 Kb) and the unspliced *petB* (band #2, 1.6 Kb) transcripts were barely detectable in the mutant, whereas the *petB*-unspliced *petD*-spliced dicistronic transcript (band #1, 2.2 Kb), detected with probes D, E, F, and H, accumulated to even higher levels in *atcrp1* plastids, presumably due to the failure of *At*CRP1-dependent processing between the *petB* and *petD* coding regions, as also shown in *zmcrp1* mutant plants (Barkan et al.,

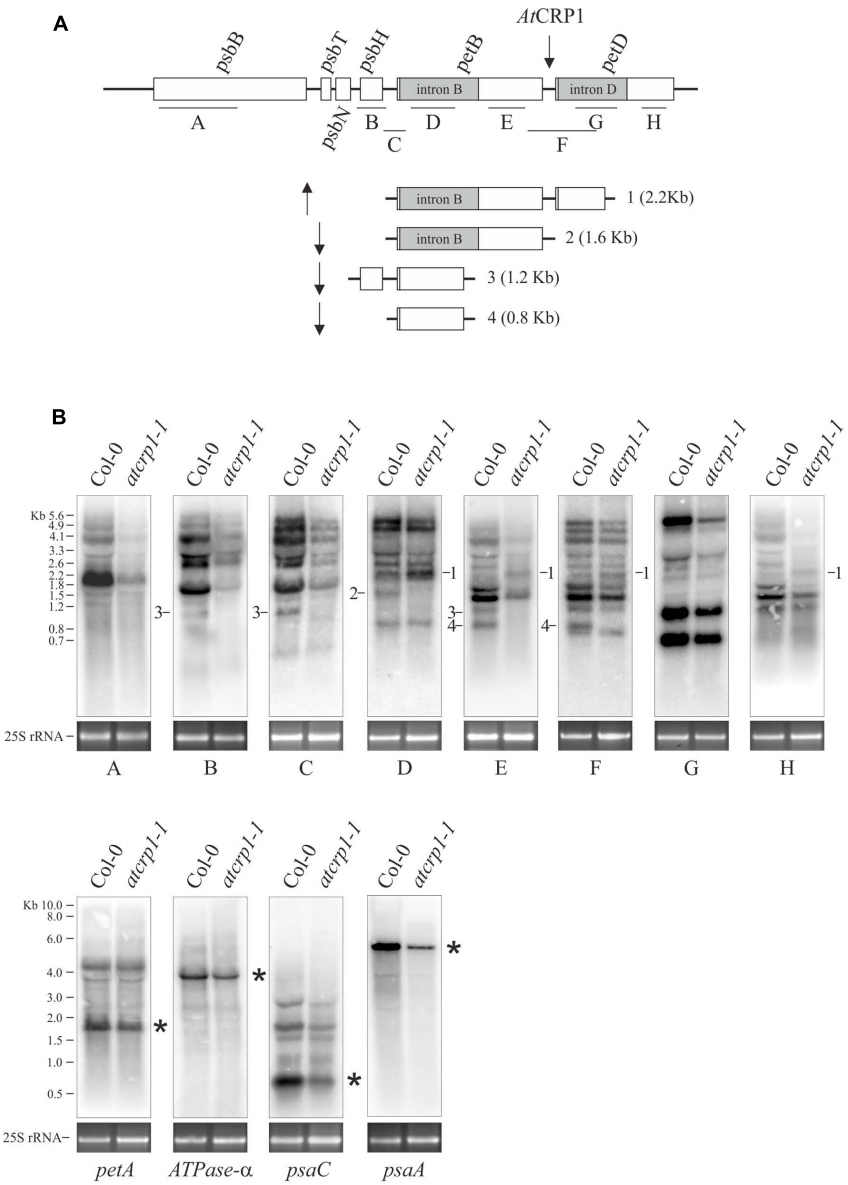

FIGURE 9 | Transcript patterns of chloroplast genes in Col-0 and *atcrp1-1* mutant leaves. (A) The structure of the *psbB* gene cluster and probes A to H used in RNA gel blots analysis in **(B)** are shown. Furthermore, processed and spliced transcripts that accumulate differentially between Col-0 and mutant chloroplasts are drawn to scale and numbered from 1 to 4. Upward arrow indicates transcripts that accumulate to higher levels in *atcrp1-1* than Col-0 chloroplasts, whilst the downward arrow is used for transcripts less abundant or absent in mutant samples. The putative binding site of *At*CRP1 within the *petB-petD* intergenic region is also indicated. **(B)** RNA gel blot analysis of the *psbB* gene cluster were performed using probes indicated as A to H, whilst *petA*, *ATPase-A*, *psaC*, and *psaA* specific probes are described in section "Materials and Methods." The identity of labeled transcripts (1–4), shown in **(A)** together with their size, was established based on the hybridization pattern, transcript size and on data reported in Meierhoff et al. (2003) and Stoppel et al. (2011). Asterisks indicate the mature transcript forms. A portion of the ethidium bromide stained Agarose gels, containing the cytosolic 25S rRNA, is included, as loading control, below each filter. One out of three Northern-blots for each transcript-specific probe is shown.

1994; Fisk et al., 1999). In contrast with maize, monocistronic and spliced *petD* transcripts of ∼600 nucleotides do not accumulate to significant levels in Arabidopsis, and thus its absence was not observed in *atcrp1* plastids (Barkan et al., 1994; Barkan, 2011).

Moreover, the lack of the PsaC and PetA subunits could be the consequence of the simultaneous decrease of transcript accumulation and a possible defect in *At*CRP1-dependent activation of *psaC* and *petA* translation, as shown in *Zea mays* (Barkan et al., 1994; Schmitz-Linneweber et al., 2005). However, the specific regulatory role of *AtCRP1* in plastid protein translation is difficult to verify, owing to the general and pleiotropic decrease of mature plastid *rRNA* in *atcrp1-1* leaves, in spite of WT-like accumulation of *rrn23* and *rrn4.5* precursor forms (**Figure 10**). This rRNA accumulation pattern

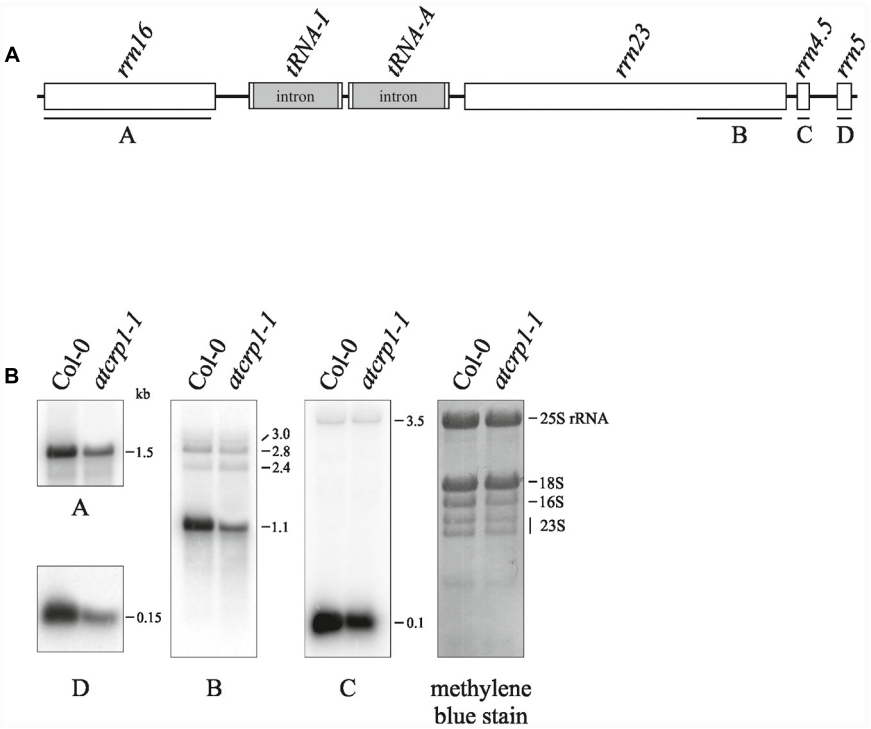

FIGURE 10 | Plastid rRNA accumulation in Colo-0 and *atcrp1-1* mutant leaves. (A) Schematic representation of the chloroplast *rrn* operon. Probes used in Northern blot analysis are indicated as black bars under each rRNA gene (A–D). **(B)** RNA gel blot analysis of plastid rRNAs were performed using the probes A-to-D described above. For loading control, a methylene blue stained filter is shown. One out of three Northern-blots for each transcript-specific probe is shown.

is very similar to the ones of mutants with impaired chloroplast translation and has been interpreted as a secondary consequence of reduced plastid protein synthesis (Tiller et al., 2012; Tadini et al., 2016).

DISCUSSION

In this study we have investigated the role of *At*CRP1 in the biogenesis of dicotyledonous-C3 chloroplasts and compared its function to the already characterized monocotyledonous-C4 chloroplast counterpart, *Zm*CRP1. Both proteins are essential for chloroplast biogenesis and photosynthetic activity, since they are required for the processing and translation of specific plastid transcripts encoding subunits of the thylakoid protein complexes. Our results indicate that *At*CRP1 and *Zm*CRP1 have very similar RNA targets and the main functional divergences are most likely due to the distinct localization of the two proteins inside the chloroplast and the partially different affinity for the RNA targets (see **Table 1**).

CRP1 Proteins Are Part of Chloroplast Nucleoids

We detected *At*CRP1 in the stroma and associated with thylakoid membranes (see **Figure 2**; **Table 1**), whereas *Zm*CRP1 is a stromal protein with no detectable association with chloroplast membranes (Fisk et al., 1999). The dual localization of *At*CRP1 within the chloroplast is supported by proteomic studies that detected *At*CRP1 in the grana-fraction of Arabidopsis thylakoids (Tomizioli et al., 2014) and in the stroma proteome, as part of Megadalton complexes (Olinares et al., 2010). In particular, *At*CRP1 appeared to be highly enriched in fractions that contained ribosomal proteins, translation factors, RNA helicases and other PPR proteins, suggesting a major role of *At*CRP1 in chloroplast gene expression. These data, together with the co-localization with GUN1 protein (see **Figure 2**), indicate that *At*CRP1 is integral to chloroplast nucleoids (Koussevitzky et al., 2007; Colombo et al., 2016; Tadini et al., 2016), i.e., the DNA-containing structures without defined boundaries that harbor the plastid gene expression machinery (Pfalz and Pfannschmidt, 2013; Melonek et al., 2016). Similarly, *Zm*CRP1 was found to be highly enriched in the nucleoid fractions of maize plastids, together with proteins involved in DNA replication, organization and repair as well as transcription, mRNA processing, splicing and editing (Majeran et al., 2012), further supporting the involvement of CRP1 proteins in plastid gene expression.

CRP1 Proteins Are Required for the Biogenesis of the Photosynthetic Apparatus

The yellow-albinotic and seedling lethal phenotype exhibited by *atcrp1* is very similar to the chlorophyll deficient and lethal phenotype of *zmcrp1* plants (Barkan et al., 1994; Fisk et al.,

TABLE 1 | Overview of the phenotypes of Arabidopsis and maize *crp1* mutants and comparison of their molecular roles in chloroplast biogenesis.

atcrp1[a]				***zmcrp1***[b]			
Plant phenotype							
Seedling lethal with yellow-albinotic cotyledons and leaves. Plants are able to develop mature leaves and sterile flowers when grown on MS medium supplemented with sucrose				Seedling lethal with pale-green cotyledon and leaves. Plants are able to develop mature non-photosynthetic leaves thanks to the large reserves of maize seeds			
CRP1 protein localization							
*At*CRP1 is a component of plastid nucleoids and it is found associated to thylakoid membranes and in the stroma				*Zm*CRP1 has been reported to be highly enriched in plastid nucleoids and to localize exclusively in the chloroplast stroma			
Thylakoid protein content							
PSI	PSII	Cyt b_6f	ATPase	PSI	PSII	Cyt b_6f	ATPase
− (/PsaC)	−	/	−	−	=	/	=
RNA targets							
RIP-Chip	Slot-Blot	*In vivo* footprint		RIP-Chip	Slot-Blot	*In vivo* footprint	
psaC	*psaC*	*psaC*		*psaC*	*psaC*	n.r.	
petB-petD	*petB-petD*	*petB-petD*		/	/	*petB-petD*	
/	*petA*	*petA*		*petA*	*petA*	n.r.	
rps15	/	/		/	/	n.r.	
Metabolism of chloroplast RNAs							
	Accumulation	Processing defects		Accumulation		Processing defects	
psaC	−	No		=		No	
petB-petD	/	Yes		/		Yes	
petA	−	No		=		No	

[a]*Data are obtained from the present manuscript* [b]*Data are obtained from Barkan et al. (1994, 2012), Fisk et al. (1999), Schmitz-Linneweber et al. (2005). −, marked reduction; /, complete absence; =, no changes; +, increase; n.r., not reported.*

1999). Arabidopsis mutants die at the two-cotyledon stage after germination on soil, but can overcome seedling lethality on sucrose-containing media, where they develop mature leaves and sterile flowers (see **Figure 4**; **Table 1**). Similarly, non-photosynthetic *zmcrp1* plants die at about 3 weeks after germination when seed reserves are exhausted. Furthermore, the *atcrp1* phenotype appears to be typical of Arabidopsis mutants lacking components of the photosynthetic apparatus and not of the gene expression machinery or of the protein import apparatus, since the latter usually result in the premature arrest at the globular-to-heart stage of embryo development, when chloroplast biogenesis begins (Ruppel and Hangarter, 2007; Romani et al., 2012; Beeler et al., 2014). Nevertheless, the pale-green pigmentation of the mutant embryo at bent-cotyledon stage (see **Figure 4**) and the β-glucuronidase (GUS) activity observed in young developing cotyledons and rosette leaves, but not in older tissues (see **Figure 5**), indicate that *AtCRP1* gene expression and protein accumulation is required during the very early stages of the photosynthetic apparatus assembly. Immunoblot data indicate, indeed, that *At*CRP1, like *Zm*CRP1, might act as a nuclear-encoded anterograde regulatory component responsible for coordination of the accumulation of Cyt b_6/f and PSI protein complexes (see **Figure 6**). Besides their role in linear electron transport (LET), Cyt b_6/f and PSI indeed play a key role in Cyclic Electron Transport (CET), which has been reported to be enhanced in Arabidopsis green seeds and to be required for optimal seed vigor and seed germination rate (Allorent et al., 2015).

In contrast to *zmcrp1* plants (Barkan et al., 1994), the absence of *At*CRP1 destabilized the entire photosynthetic apparatus, as shown by the marked reduction of PSII core, ATPase and LHC protein levels. The general down-regulation of thylakoid complexes owing to defects in the intersystem electron transport chain appears to be a common feature of Arabidopsis photosynthetic mutants and provides clear evidence of a different adaptive response between monocot and dicot plants (Meurer et al., 1996; Varotto et al., 2000, 2002; Maiwald et al., 2003; Weigel et al., 2003; Ihnatowicz et al., 2004, 2007; Belcher et al., 2015). Furthermore, the *atcrp1-1* phenotype, both in terms of plastid transcript and plastid protein accumulation, appears to be much more drastic than the one of other *ppr* mutants required for the processing and expression of *psbB-psbT-psbH-petB-petD* operon, such as *hcf152* (Meierhoff et al., 2003), suggesting that the absence of *At*CRP1 protein might affect the activity of other factors essential for plastid gene expression. As a matter of fact, rRNA abundance is markedly reduced in *atcrp1-1* plastids, indicating a general reduction of protein synthesis, as consequence of pleiotropic effects.

RNA Targets: Commonalities and Divergences between *At*CRP1 and *Zm*CRP1 Proteins

RNA immunoprecipitation-Chip and slot blot data suggest a physical interaction between *At*CRP1 and the transcripts of

psaC, *petB-petD* and possibly *petA*, even though it is not known whether these interactions are direct or mediated by other factors (see **Figure 7**). However, all of these RNAs harbor a region where a native footprint is annotated, raising the tempting hypothesis that *At*CRP1 is in fact the RNA-binding factor responsible for that footprint (see **Figure 8**). Furthermore, when these enriched fragments were searched for occurrences of the predicted binding motif of *At*CRP1, each of them proved to contain a hit inside the footprint region, strongly suggesting that *At*CRP1 could be the factor leaving those footprints. Nevertheless, the observation that the footprints identified in Arabidopsis *psaC, petA,* and *petB-petD* transcripts are larger than the 14 nucleotide size of the predicted *At*CRP1 footprint (37, 29, and 41 nucleotides in *psaC, petA,* and *petB-petD,* respectively) supports the view that the binding of *At*CRP1 to its targets *in vivo* could be stabilized by other protein partners. For instance, the peptide chain release factor B3 (PrfB3) has been also shown to be required for Arabidopsis autotrophic growth and for the stability of 3′ processed *petB* transcripts to adjust cytochrome b_6 levels (Stoppel et al., 2011), thus possibly being an *At*CRP1 specific protein partner. Similarly, PPR proteins involved in RNA stabilization and editing have been shown to interact with RNA Recognition Motif (RRM) proteins and other factors, indicating that larger protein complexes assembled around a PPR protein are likely to occur (Kupsch et al., 2012; Takenaka et al., 2014; Shi et al., 2015).

The interactions with the 5′UTR of *psaC* and *petA* have also been reported in the case of *Zm*CRP1 (Schmitz-Linneweber et al., 2005; Williams-Carrier et al., 2008), indicating that this feature of CRP1 function is conserved between Arabidopsis and maize. *Zm*CRP1 was also shown to bind directly to the 5′-UTR of *petA* transcripts by electrophoresis mobility shift assay (Williams-Carrier et al., 2008), favoring the possibility of a direct binding of CRP1 proteins to the corresponding RNA targets (see also **Figure 8**). Furthermore, *Zm*CRP1 has been proposed to directly control the translation of *petA* and *psaC* transcripts (Barkan et al., 1994), as shown through pulse labeling and polysome loading (in the case of *petA*), or deduced from the reduced association of *psaC* RNAs with ribosomes. Interestingly, the PsaC subunit of PSI and the PetA subunit of Cyt b_6/f could not be detected in *atcrp1* thylakoids, despite the accumulation of the corresponding transcripts with no processing defects (see also **Figure 9**), suggesting that *At*CRP1 plays a major role in translation regulation also in Arabidopsis. Unfortunately, the specific requirement of *At*CRP1 in plastid protein translation cannot be verified by comparing Col-0 and *atcrp1-1* leaves, due to the marked reduction of *rRNA* accumulation in *atcrp1-1* plastids.

In addition to the defects in *petA* translation, the complete absence of Cyt b_6/f protein complex observed in *atcrp1* thylakoids can also be attributed to processing alterations of the *psbB-psbT-psbH-petB-petD* polycistronic transcription unit. The lack of the monocistronic *petB*, the dicistronic *psbH-petB*, and the unspliced *petB* transcripts, together with the direct binding of *At*CRP1 to the *petB-petD* intergenic region, strongly support the role of *At*CRP1 in the metabolism of *petB* and *petD* transcripts. PPR protein-derived RNA-footprints are considered to arise due to exonucleolytic activity (Ruwe et al., 2016). Since sRNAs corresponding to predicted binding sites of *At*CRP1 are identified here, the most likely role for *At*CRP1 is to block exonucleases from degrading the *petB* and *petD* transcripts. A similar defect in *petB-petD* maturation has been reported in *zmcrp1* mutant plants (Barkan et al., 1994; Fisk et al., 1999; Schmitz-Linneweber et al., 2005), although no association was detected between *Zm*CRP1 and the *petB-petD* intergenic region (Schmitz-Linneweber et al., 2005), so it is still uncertain whether the role of *Zm*CRP1 is direct or indirect.

CONCLUSION

Taken together, the characterization of the functional role of *At*CRP1 in chloroplast biogenesis has highlighted several features in common with the *Zm*CRP1. Both proteins appear to control, directly or indirectly, the expression of plastid genes encoding subunits of Cyt b_6/f and PSI protein complexes. The coordination of the accumulation of these two protein complexes is fundamental to guarantee optimal photosynthesis in mature plants, but appears also to be important during seed germination, when cyclic electron transport is highly enhanced relative to LET.

Differences in RNA targets observed by immunoprecipitation and hybridization assays between *At*CRP1 and *Zm*CRP1 might be explained by a broad affinity for RNA targets, but may also have technical reasons (GFP antibody for Arabidopsis versus direct anti-ZmCRP1 antibody in maize). Evidence in favor of conservation of PPR protein activity between different species has been reported for the PLS and P subfamilies (Choury et al., 2005; Bolle and Kempken, 2006; Choury and Araya, 2006); for instance, the maize MPPR6 protein can complement loss-of-function Arabidopsis mutants lacking the orthologous protein (Manavski et al., 2012). However, functional divergence has been also observed, as in the case of orthologous PPR proteins ATP4 (maize) and SVR7 (Arabidopsis) (Liu et al., 2010; Zoschke et al., 2012, 2013a,b). Further studies aimed to verify the degree of protein activity conservation between monocots and dicots are needed to extend our knowledge of PPR protein functions and the degree of protein function conservation. The parallel characterization of PPR orthologs, including the relationship between their protein structures and the corresponding target RNA species, may represent an underestimated and powerful strategy to precisely determine the PPR code, essential for a fast and accurate large scale prediction of PPR targets, and for the functional characterization of the PPR-mediated nucleus-to-chloroplast anterograde signaling pathway.

AUTHOR CONTRIBUTIONS

RF, LT, FM, FR, SM, M-KL, CS-L, and PP participated to the organization of the manuscript. RF, LT, FM, FR, SM, MC, AC,

and PP designed and carried out the experiments related to the molecular biology and biochemical characterization of atcrp1 mutants. RF, FM, LT, M-KL, and CS-L were involved in RIP-Chip and slot blot assays, as well as in the *in silico* identification of native footprints and prediction of AtCRP1 binding motif. PP wrote the manuscript.

FUNDING

This work was supported by ERA-NET Cofund FACCE SURPLUS (BarPLUS grant id. 93) to PP. Work in the lab of CS-L was supported by DFG grant SCHM 1698/5-1.

ACKNOWLEDGMENT

We thank Hannes Ruwe and Gongwei Wang for helpful comments on the sRNA analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2017.00163/full#supplementary-material

TABLE S1 | RNA immunoprecipitation (RIP-Chip) data.

TABLE S2 | Oligonucleotides employed for cloning, genotyping, northern blot, slot blot and qRT-PCR assays.

REFERENCES

Allorent, G., Osorio, S., Vu, J. L., Falconet, D., Jouhet, J., Kuntz, M., et al. (2015). Adjustments of embryonic photosynthetic activity modulate seed fitness in *Arabidopsis thaliana*. *New Phytol.* 205, 707–719. doi: 10.1111/nph.13044

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653–657. doi: 10.1126/science.1086391

Barkan, A. (2011). Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol.* 155, 1520–1532. doi: 10.1104/pp.110.171231

Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y. S., Bond, C. S., et al. (2012). A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.* 8:e1002910. doi: 10.1371/journal.pgen.1002910

Barkan, A., and Small, I. (2014). Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* 65, 415–442. doi: 10.1146/annurev-arplant-050213-040159

Barkan, A., Walker, M., Nolasco, M., and Johnson, D. (1994). A nuclear mutation in maize blocks the processing and translation of several chloroplast mRNAs and provides evidence for the differential translation of alternative mRNA forms. *EMBO J.* 13, 3170–3181.

Beeler, S., Liu, H. C., Stadler, M., Schreier, T., Eicke, S., Lue, W. L., et al. (2014). Plastidial NAD-dependent malate dehydrogenase is critical for embryo development and heterotrophic metabolism in *Arabidopsis*. *Plant Physiol.* 164, 1175–1190. doi: 10.1104/pp.113.233866

Belcher, S., Williams-Carrier, R., Stiffler, N., and Barkan, A. (2015). Large-scale genetic analysis of chloroplast biogenesis in maize. *Biochim. Biophys. Acta* 1847, 1004–1016. doi: 10.1016/j.bbabio.2015.02.014

Bolle, N., and Kempken, F. (2006). Mono- and dicotyledonous plant-specific RNA editing sites are correctly edited in both in organello systems. *FEBS Lett.* 580, 4443–4448. doi: 10.1016/j.febslet.2006.07.011

Cai, W., Okuda, K., Peng, L., and Shikanai, T. (2011). PROTON GRADIENT REGULATION 3 recognizes multiple targets with limited similarity and mediates translation and RNA stabilization in plastids. *Plant J.* 67, 318–327. doi: 10.1111/j.1365-313X.2011.04593.x

Cheng, S., Gutmann, B., Zhong, X., Ye, Y., Fisher, M. F., Bai, F., et al. (2016). Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J.* 85, 532–547. doi: 10.1111/tpj.13121

Choury, D., and Araya, A. (2006). RNA editing site recognition in heterologous plant mitochondria. *Curr. Genet.* 50, 405–416. doi: 10.1007/s00294-006-0100-3

Choury, D., Farre, J. C., Jordana, X., and Araya, A. (2005). Gene expression studies in isolated mitochondria: *Solanum tuberosum* rps10 is recognized by cognate potato but not by the transcription, splicing and editing machinery of wheat mitochondria. *Nucleic Acids Res.* 33, 7058–7065. doi: 10.1093/nar/gki1017

Colombo, M., Tadini, L., Peracchio, C., Ferrari, R., and Pesaresi, P. (2016). GUN1, a jack-of-all-trades in chloroplast protein homeostasis and signaling. *Front. Plant Sci.* 7:1427. doi: 10.3389/fpls.2016.01427

Costa, A., Gutla, P. V., Boccaccio, A., Scholz-Starke, J., Festa, M., Basso, B., et al. (2012). The *Arabidopsis* central vacuole as an expression system for intracellular transporters: functional characterization of the Cl-/H+ exchanger CLC-7. *J. Physiol.* 590, 3421–3430. doi: 10.1113/jphysiol.2012.230227

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

Fisk, D. G., Walker, M. B., and Barkan, A. (1999). Molecular cloning of the maize gene crp1 reveals similarity between regulators of mitochondrial and chloroplast gene expression. *EMBO J.* 18, 2621–2630. doi: 10.1093/emboj/18.9.2621

Grant, C. E., Bailey, T. L., and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018. doi: 10.1093/bioinformatics/btr064

Gregis, V., Sessa, A., Dorca-Fornell, C., and Kater, M. M. (2009). The *Arabidopsis* floral meristem identity genes AP1, AGL24 and SVP directly repress class B and C floral homeotic genes. *Plant J.* 60, 626–637. doi: 10.1111/j.1365-313X.2009.03985.x

Hall, T. M. (2016). De-coding and re-coding RNA recognition by PUF and PPR repeat proteins. *Curr. Opin. Struct. Biol.* 36, 116–121. doi: 10.1016/j.sbi.2016.01.010

Harrison, T., Ruiz, J., Sloan, D. B., Ben-Hur, A., and Boucher, C. (2016). aPPRove: an HMM-based method for accurate prediction of RNA-pentatricopeptide repeat protein binding events. *PLoS ONE* 11:e0160645. doi: 10.1371/journal.pone.0160645

Ihnatowicz, A., Pesaresi, P., and Leister, D. (2007). The E subunit of photosystem I is not essential for linear electron flow and photoautotrophic growth in *Arabidopsis thaliana*. *Planta* 226, 889–895. doi: 10.1007/s00425-007-0534-y

Ihnatowicz, A., Pesaresi, P., Varotto, C., Richly, E., Schneider, A., Jahns, P., et al. (2004). Mutants for photosystem I subunit D of *Arabidopsis thaliana*: effects on photosynthesis, photosystem I stability and expression of nuclear genes for chloroplast functions. *Plant J.* 37, 839–852. doi: 10.1111/j.1365-313X.2004.02011.x

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

Karimi, M., Inze, D., and Depicker, A. (2002). GATEWAY vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci.* 7, 193–195. doi: 10.1016/S1360-1385(02)02251-3

Kindgren, P., Yap, A., Bond, C. S., and Small, I. (2015). Predictable alteration of sequence recognition by RNA editing factors from *Arabidopsis*. *Plant Cell* 27, 403–416. doi: 10.1105/tpc.114.134189

Koussevitzky, S., Nott, A., Mockler, T. C., Hong, F., Sachetto-Martins, G., Surpin, M., et al. (2007). Signals from chloroplasts converge to regulate nuclear gene expression. *Science* 316, 715–719. doi: 10.1126/science.%201140516

Kunst, L. (1998). Preparation of physiologically active chloroplasts from *Arabidopsis*. *Methods Mol. Biol.* 82, 43–48.

Kupsch, C., Ruwe, H., Gusewski, S., Tillich, M., Small, I., and Schmitz-Linneweber, C. (2012). *Arabidopsis* chloroplast RNA binding proteins CP31A and CP29A associate with large transcript pools and confer cold stress tolerance by influencing multiple chloroplast RNA processing steps. *Plant Cell* 24, 4266–4280. doi: 10.1105/tpc.112.103002

Lawrence, C. J., Dong, Q., Polacco, M. L., Seigfried, T. E., and Brendel, V. (2004). MaizeGDB, the community database for maize genetics and genomics. *Nucleic Acids Res.* 32, D393–D397. doi: 10.1093/nar/gkh011

Liu, X., Yu, F., and Rodermel, S. (2010). An *Arabidopsis* pentatricopeptide repeat protein, SUPPRESSOR OF VARIEGATION7, is required for FtsH-mediated chloroplast biogenesis. *Plant Physiol.* 154, 1588–1601. doi: 10.1104/pp.110.164111

Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., et al. (2004). Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16, 2089–2103. doi: 10.1105/tpc.104.022236

Maiwald, D., Dietzmann, A., Jahns, P., Pesaresi, P., Joliot, P., Joliot, A., et al. (2003). Knock-out of the genes coding for the Rieske protein and the ATP-synthase delta-subunit of *Arabidopsis*. Effects on photosynthesis, thylakoid protein composition, and nuclear chloroplast gene expression. *Plant Physiol.* 133, 191–202. doi: 10.1104/pp.103.024190

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

Manavski, N., Guyon, V., Meurer, J., Wienand, U., and Brettschneider, R. (2012). An essential pentatricopeptide repeat protein facilitates 5′ maturation and translation initiation of rps3 mRNA in maize mitochondria. *Plant Cell* 24, 3087–3105. doi: 10.1105/tpc.112.099051

Manna, S. (2015). An overview of pentatricopeptide repeat proteins and their applications. *Biochimie* 113, 93–99. doi: 10.1016/j.biochi.2015.04.004

Martinez-Garcia, J. F., Monte, E., and Quail, P. H. (1999). A simple, rapid and quantitative method for preparing *Arabidopsis* protein extracts for immunoblot analysis. *Plant J.* 20, 251–257. doi: 10.1046/j.1365-313x.1999.00579.x

Meierhoff, K., Felder, S., Nakamura, T., Bechtold, N., and Schuster, G. (2003). HCF152, an *Arabidopsis* RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs. *Plant Cell* 15, 1480–1495. doi: 10.1105/tpc.010397

Melonek, J., Oetke, S., and Krupinska, K. (2016). Multifunctionality of plastid nucleoids as revealed by proteome analyses. *Biochim. Biophys. Acta* 1864, 1016–1038. doi: 10.1016/j.bbapap.2016.03.009

Meurer, J., Lezhneva, L., Amann, K., Godel, M., Bezhani, S., Sherameti, I., et al. (2002). A peptide chain release factor 2 affects the stability of UGA-containing transcripts in *Arabidopsis* chloroplasts. *Plant Cell* 14, 3255–3269. doi: 10.1105/tpc.006809

Meurer, J., Meierhoff, K., and Westhoff, P. (1996). Isolation of high-chlorophyll-fluorescence mutants of *Arabidopsis thaliana* and their characterisation by spectroscopy, immunoblotting and northern hybridisation. *Planta* 198, 385–396. doi: 10.1007/BF00620055

Okuda, K., Shoki, H., Arai, M., Shikanai, T., Small, I., and Nakamura, T. (2014). Quantitative analysis of motifs contributing to the interaction between PLS-subfamily members and their target RNA sequences in plastid RNA editing. *Plant J.* 80, 870–882. doi: 10.1111/tpj.12687

Olinares, P. D., Ponnala, L., and Van Wijk, K. J. (2010). Megadalton complexes in the chloroplast stroma of *Arabidopsis thaliana* characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering. *Mol. Cell. Proteom.* 9, 1594–1615. doi: 10.1074/mcp.M000038-MCP201

O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., et al. (2008). On the expansion of the pentatricopeptide repeat gene family in plants. *Mol. Biol. Evol.* 25, 1120–1128. doi: 10.1093/molbev/msn057

Pfalz, J., and Pfannschmidt, T. (2013). Essential nucleoid proteins in early chloroplast development. *Trends Plant Sci.* 18, 186–194. doi: 10.1016/j.tplants.2012.11.003

Pogson, B. J., and Albrecht, V. (2011). Genetic dissection of chloroplast biogenesis and development: an overview. *Plant Physiol.* 155, 1545–1551. doi: 10.1104/pp.110.170365

Pogson, B. J., Ganguly, D., and Albrecht-Borth, V. (2015). Insights into chloroplast biogenesis and development. *Biochim. Biophys. Acta* 1847, 1017–1024. doi: 10.1016/j.bbabio.2015.02.003

Remans, T., Smeets, K., Opdenakker, K., Mathijsen, D., Vangronsveld, J., and Cuypers, A. (2008). Normalisation of real-time RT-PCR gene expression measurements in *Arabidopsis thaliana* exposed to increased metal concentrations. *Planta* 227, 1343–1349. doi: 10.1007/s00425-008-0706-4

Rivals, E., Bruyere, C., Toffano-Nioche, C., and Lecharny, A. (2006). Formation of the *Arabidopsis* pentatricopeptide repeat family. *Plant Physiol.* 141, 825–839. doi: 10.1104/pp.106.077826

Romani, I., Tadini, L., Rossi, F., Masiero, S., Pribil, M., Jahns, P., et al. (2012). Versatile roles of *Arabidopsis* plastid ribosomal proteins in plant growth and development. *Plant J.* 72, 922–934. doi: 10.1111/tpj.12000

Ruppel, N. J., and Hangarter, R. P. (2007). Mutations in a plastid-localized elongation factor G alter early stages of plastid development in *Arabidopsis thaliana*. *BMC Plant Biol.* 7:37. doi: 10.1186/1471-2229-7-37

Ruwe, H., Wang, G., Gusewski, S., and Schmitz-Linneweber, C. (2016). Systematic analysis of plant mitochondrial and chloroplast small RNAs suggests organelle-specific mRNA stabilization mechanisms. *Nucleic Acids Res.* 44, 7406–7417. doi: 10.1093/nar/gkw466

Schagger, H., and von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* 166, 368–379. doi: 10.1016/0003-2697(87)90587-2

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. doi: 10.1038/nmeth.2019

Schmitz-Linneweber, C., Williams-Carrier, R., and Barkan, A. (2005). RNA immunoprecipitation and microarray analysis show a chloroplast Pentatricopeptide repeat protein to be associated with the 5' region of mRNAs whose translation it activates. *Plant Cell* 17, 2791–2804. doi: 10.1105/tpc.105.034454

Schmitz-Linneweber, C., Williams-Carrier, R. E., Williams-Voelker, P. M., Kroeger, T. S., Vichas, A., and Barkan, A. (2006). A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast rps12 pre-mRNA. *Plant Cell* 18, 2650–2663. doi: 10.1105/tpc.106.046110

Sessions, A., Burke, E., Presting, G., Aux, G., Mcelver, J., Patton, D., et al. (2002). A high-throughput *Arabidopsis* reverse genetics system. *Plant Cell* 14, 2985–2994. doi: 10.1105/tpc.004630

Shen, C., Zhang, D., Guan, Z., Liu, Y., Yang, Z., Yang, Y., et al. (2016). Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. *Nat. Commun.* 7:11285. doi: 10.1038/ncomms11285

Shi, X., Hanson, M. R., and Bentolila, S. (2015). Two RNA recognition motif-containing proteins are plant mitochondrial editing factors. *Nucleic Acids Res.* 43, 3814–3825. doi: 10.1093/nar/gkv245

Small, I. D., and Peeters, N. (2000). The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* 25, 46–47. doi: 10.1016/S0968-0004(99)01520-0

Stoppel, R., Lezhneva, L., Schwenkert, S., Torabi, S., Felder, S., Meierhoff, K., et al. (2011). Recruitment of a ribosomal release factor for light- and stress-dependent regulation of petB transcript stability in *Arabidopsis* chloroplasts. *Plant Cell* 23, 2680–2695. doi: 10.1105/tpc.111.085324

Tadini, L., Pesaresi, P., Kleine, T., Rossi, F., Guljamow, A., Sommer, F., et al. (2016). GUN1 controls accumulation of the plastid ribosomal protein S1 at the protein level and interacts with proteins involved in plastid protein homeostasis. *Plant Physiol.* 170, 1817–1830. doi: 10.1104/pp.15.02033

Takenaka, M., Verbitskiy, D., Zehrmann, A., Hartel, B., Bayer-Csaszar, E., Glass, F., et al. (2014). RNA editing in plant mitochondria-connecting RNA target sequences and acting proteins. *Mitochondrion* 19(Pt B), 191–197. doi: 10.1016/j.mito.2014.04.005

Takenaka, M., Zehrmann, A., Brennicke, A., and Graichen, K. (2013). Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS ONE* 8:e65343. doi: 10.1371/journal.pone.0065343

Tiller, N., Weingartner, M., Thiele, W., Maximova, E., Schöttler, M. A., and Bock, R. (2012). The plastid-specific ribosomal proteins of *Arabidopsis thaliana* can be divided into non-essential proteins and genuine ribosomal proteins. *Plant J.* 69, 302–316. doi: 10.1111/j.1365-313X.2011.04791.x

Tomizioli, M., Lazar, C., Brugiere, S., Burger, T., Salvi, D., Gatto, L., et al. (2014). Deciphering thylakoid sub-compartments using a mass spectrometry-based approach. *Mol. Cell. Proteom.* 13, 2147–2167. doi: 10.1074/mcp.M114.040923

Varotto, C., Pesaresi, P., Jahns, P., Lessnick, A., Tizzano, M., Schiavon, F., et al. (2002). Single and double knockouts of the genes for photosystem I subunits G, K, and H of *Arabidopsis*. Effects on photosystem I composition, photosynthetic electron flow, and state transitions. *Plant Physiol.* 129, 616–624. doi: 10.1104/pp.002089

Varotto, C., Pesaresi, P., Meurer, J., Oelmuller, R., Steiner-Lange, S., Salamini, F., et al. (2000). Disruption of the *Arabidopsis* photosystem I gene psaE1 affects photosynthesis and impairs growth. *Plant J.* 22, 115–124. doi: 10.1046/j.1365-313x.2000.00717.x

Voinnet, O., Rivas, S., Mestre, P., and Baulcombe, D. (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* 33, 949–956. doi: 10.1046/j.1365-313X.2003.01676.x

Waadt, R., and Kudla, J. (2008). In planta visualization of protein interactions using Bimolecular Fluorescence Complementation (BiFC). *CSH Protoc* 2008, dbrot4995.

Weigel, M., Varotto, C., Pesaresi, P., Finazzi, G., Rappaport, F., Salamini, F., et al. (2003). Plastocyanin is indispensable for photosynthetic electron flow in *Arabidopsis thaliana*. *J. Biol. Chem.* 278, 31286–31289. doi: 10.1074/jbc.M302876200

Williams-Carrier, R., Kroeger, T., and Barkan, A. (2008). Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. *RNA* 14, 1930–1941. doi: 10.1261/rna.1077708

Woodson, J. D., and Chory, J. (2008). Coordination of gene expression between organellar and nuclear genomes. *Nat. Rev. Genet.* 9, 383–395. doi: 10.1038/nrg2348

Yamazaki, H., Tasaka, M., and Shikanai, T. (2004). PPR motifs of the nucleus-encoded factor, PGR3, function in the selective and distinct steps of chloroplast gene expression in *Arabidopsis*. *Plant J.* 38, 152–163. doi: 10.1111/j.1365-313X.2004.02035.x

Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., et al. (2013). Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* 504, 168–171. doi: 10.1038/nature12651

Yoo, S. D., Cho, Y. H., and Sheen, J. (2007). *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* 2, 1565–1572. doi: 10.1038/nprot.2007.199

Zoschke, R., Kroeger, T., Belcher, S., Schottler, M. A., Barkan, A., and Schmitz-Linneweber, C. (2012). The pentatricopeptide repeat-SMR protein ATP4 promotes translation of the chloroplast atpB/E mRNA. *Plant J.* 72, 547–558. doi: 10.1111/j.1365-313X.2012.05081.x

Zoschke, R., Qu, Y., Zubo, Y. O., Borner, T., and Schmitz-Linneweber, C. (2013a). Mutation of the pentatricopeptide repeat-SMR protein SVR7 impairs accumulation and translation of chloroplast ATP synthase subunits in *Arabidopsis thaliana*. *J. Plant Res.* 126, 403–414. doi: 10.1007/s10265-012-0527-1

Zoschke, R., Watkins, K. P., and Barkan, A. (2013b). A rapid ribosome profiling method elucidates chloroplast ribosome behavior in vivo. *Plant Cell* 25, 2265–2275. doi: 10.1105/tpc.113.111567

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