Regulation of plastid transcription by sigma factors and anti-sense RNAs in Arabidopsis thaliana.

Malik Ghulam Mustafa

Université de Grenoble; France
Ecole Doctorale de Chimie et Science du Vivant

Regulation of plastid transcription by sigma factors and anti-sense RNAs in *Arabidopsis thaliana*.

Thesis

To obtain the grade of

Doctor of University of Grenoble, France

Discipline: Biology

Presented and defended publically by

Malik Ghulam Mustafa

on

15 December 2010

Directors of thesis : Dr. Silva Lerbs-Mache
: Dr. Livia Merendino

Composition of Jury:

Laurence Maréchal-Drouard	**Reporter**
Françoise Monéger	**Reporter**
Michel Goldschmidt-Clermont	**Examinator**
Silva Lerbs-Mache	**Director of thesis**
Livia Merendino	**Co-director of thesis**

Abstract :

Les chloroplastes, responsables de la photosynthèse chez les organismes autotrophes, possèdent un génome plastidial codant de 100 à 130 gènes dont environ 80 pour des protéines principalement impliquées dans la photosynthèse, la transcription et la traduction. L'expression de ces gènes, coordonnée entre le plaste et le noyau, implique deux types d'ARN polymérases, la NEP (Nucleus Encoded RNA Polymerase) et la PEP (plastid Encoded RNA Polymerase) laquelle s'associe à l'un des 6 facteurs sigma (SIG), codés dans le noyau pour la reconnaissance spécifique de promoteurs de transcription.

Nous avons tout d'abord analysé le rôle de ces facteurs sigma dans la régulation transcriptionnelle des deux opérons codant des sous-unités de l'ATP synthase, *atp*I/H/F/A et *atp*B/E, en précisant le rôle particulier de SIG3 dans la reconnaissance spécifique du promoteur (-418) de l'*atp*H. Nous avons identifié les promoteurs des transcrits polycistronique et ceux situés en amont des gènes *atp*H et *atp*E, et avons montré (1) que les gènes des deux opérons sont co-régulés par SIG3 et SIG2 sauf *atp*I régulé par SIG2 seul et (2), que SIG3 jouerait un rôle essentiel dans la surexpression monocistronique d'*atp*H par la reconnaissance d'un promoteur (-418) en amont de *atp*H. L'analyse systématique des transcrits plastidiaux accumulés en fonction de l'éclairement des plantes nous a permis de corréler cette surexpression à un éclairement élevé (1300 µE) de plantes matures.

SIG3 reconnaît aussi spécifiquement le promoteur de *psb*N, gène localisé sur le brin opposé de l'opéron *psb*B/T/H/*pet*B/*pet*D, produisant un ARN anti-sens de *psb*T et de la région intergénique *psb*T/*psb*H. Nos résultats montrent que l'anti-sens de *psb*T couvre la région codante, le 5'UTR et la quasi-totalité 3' UTR du transcrit sens *psb*T, pouvant ainsi réguler la production de PSBT en interférant dans la traduction par la formation d'un duplex ARN. L'anti-sens pourrait aussi intervenir dans le processing dans la région 5' UTR de *psb*H.

Chloroplasts, responsible for photosynthesis in autotrophic organisms, have a genome containing 100-130 genes, 80 of which code for proteins mainly involved in photosynthesis, transcription and translation. Gene expression, involves two types of RNA polymerases, NEP (Nucleus Encoded RNA Polymerase) and PEP (Plastid Encoded RNA Polymerase). Six nucleus encoded sigma factors participate to PEP promoter specificity.

We first have analyzed the role of sigma factors in the transcriptional regulation of the two *atp* operons, *atpI/H/F/A* and *atpB/E,* with special emphasis on the specific contribution of SIG3 to *atpH* gene expression. We identified the promoters responsible for polycistronic

transcripts and the internal promoters upstream of the *atpH* and the *atpE* genes. All genes of both *atp* operons are SIG3 and SIG2 dependent except *atpI* that is regulated by SIG2 only. The monocistronic -418 initiated *atpH* mRNA might contribute to the higher stoichiometry of *atpH*. A systematic analysis of plastid gene expression under different light conditions showed that SIG3 plays an important role in the transcript accumulation of *atpH* in high light (1300 μE) in mature plants.

Similarly, SIG3 also recognizes specifically the promoter of *psbN* located between *psbT* and *psbH* but on the opposite DNA strand and producing an anti-sense RNA to *psb*T. We showed that the anti-sense RNA covers the coding region, the 5' UTR and almost the entire 3' UTR of the *psbT* sense transcript and thus might regulate the expression of the *psbT* gene by interfering in the translation of *psbT* mRNA via duplex formation. It could also be necessary for a processing event in the 5' UTR of *psbH*.

Acknowledgements

Many may disagree but my simple answer to the question of the professional essence of ph.d period will be to "have confidence in yourself and to the fellow researchers". That's why I am highly indebted to Dr. Silva Lerbs Mache, the director of my thesis.

During this period, Dr. Silva not only provided all the lab material necessary for carrying out day to day research but also spent a lot of time for the analysis of the results. I have greatly benefitted from her scientific competence and had a chance of scientific discussions with her. I am extremely indebted and thankful to you Dr. Silva for the patience you showed and for the time you spared and spent for the improvement of the manuscript. It was exceptional. Thanks again.

But if someone asks me what Dr. Silva did best for me during ph.d period, I would say: 'The best thing she did for me was the appointment of Dr. Livia Merendino as my co-director of thesis".

It is without exaggeration that I was in dire need of a " Livia Merendino" to be what actually I am. She stunned me by telling Silva in the first meeting that Mustafa knows how to extract RNAs. When we came out of the office, I reminded Livia that my first three extractions resulted in completely degraded RNAs. She replied, 'I know but I am sure you won't spend three years to learn RNA extraction". Since that day I never looked back and don't remember if I extracted degraded RNAs. Her trust and confidence in me have always been a source of great motivation for me.

I am highly indebted to Dr. Livia Merendino for having always kept her office door open for answering to all my questions. Without doubt, she showed great patience to have scientific, religious and social discussions with me. She has always been a source of hope, courage and confidence for me during the ups and downs of my professional and personal life. She invested a lot of time for the explanation of the theoretical aspects of experiments. Above all it was a pure pleasure to work and discuss with her. I am extremely short of words to pay gratitude to you Livia for all you have done for me. I wish you a life of unlimited pleasures and success.

I am thankful to Silva and Livia for having spared a lot of time for the correction of this manuscript. I am thankful to Régis Mache for his kind help to discuss and introduce important corrections in this manuscript.

I am also grateful to Emiline Lamdert and Florence Courtois and Jean Piere Alcaraz for technical assistance in the beginning of thesis. I am grateful to Emelie Demarsy, Wafa Zghidi whose work became the base of my thesis. I am thankful to Denis Falconet, Guillaume, Frank Buhr, David Downy, Gabrielle Tichtinsky and Abder for nice working environment and company.

I am also thankful to Higher Education Commission Pakistan for providing me the scholarship of 57 months for getting higher education in France.

Along with the wonderful professional life, I have experienced five best years of my personal life during my stay in France. I dedicate all this to my friends especially 'Grenobloise' (I would like to write the individual names but the list is very long) whose company or contact had been a source of enjoyment, pleasure and peace of mind. To be honest, in their company I never missed my family. Thanks to all of you. I am thankful to Muhammad Imran Ashiq and his parents for a wonderful time in Angers.

I am extremely grateful to ' ma petite soeur' Sumaira Kousar' to take a lot of care and to be a real sister in true sense. I am highly thankful to Muhammad Asam Riaz for having waited till late in the evening to share dinner with me during my three years stay in Grenoble. I am grateful to my parents, brothers and sisters and friends in Pakistan for their best wishes, prayers and love.

Above all I am thankful to Almighty Allah for providing me the wit and health without which nothing would have been possible.

Dedication

I dedicate this humble work

To

My sweet uncle and a friend :   Prof. *Allah Dad Malik*

My high school teacher and a friend :   Prof. *Ahmad Nawaz Virk*

My university teacher and a friend :   Prof. *Naveed Irshad*

My elder brother who took great care of

the family in my absence :   Dr. *Shahid Iqbal*

And to the co-director of my thesis:

Dr. *Livia Merendino*

For her great investment of time and energy for my professional built up.

Introduction

1. Plastids

Plastids are cellular organells which are responsible for photosynthesis in autotrophic organisms like plants and algae. In contrast to many algal cells which contain a single chloroplast, mature cells of a plant leaf can have 100-150 plastids. Lens-shaped chloroplasts are 1 to 3 μm on their short axis and 5 to 8 μm on their long axis (Staehelin, 2003). Several authors (Mullet, 1988; Pyke & Leech, 1992) found 10-14 proplastids (undifferentiated plastids) of 1 μm diameter in each meristematic cell. It has been observed that the number of chloroplasts per cell depends on the size of the cell (Pyke & Leech, 1992) and that chloroplasts divide by binary fission in mature cells. Like in the bacterial cells, cytoskeletal proteins such as FtsZ form constriction rings at the mid-section of the dividing chloroplast (Osteryoung & Nunnari, 2003). In the mutants defective in chloroplast division (Robertson & Leech, 1995), there is a decreased number of chloroplasts (2 or more) but they are of large size, occupy the same volume and show similar photosynthetic capacity.

Plastids are surrounded by one envelope composed of two membranes. This envelope not only gives a shape to the plastids but also posses specific transport systems for the exchange of proteins and other metabolites. The outer membrane having larger pores is less selective while the inner membrane has a more selective and specific transport system (Ferro et al., 2002; Soll & Schleiff, 2004). Transport complexes on the outer (TOC) and inner (TIC) membranes interact with each other as the envelope membranes are tightly appressed (Nishizawa & Mori, 1989; Park et al., 1999; Sluiman & Lokhorst, 1988).

Proplastids of the seedlings grown in the dark are arrested to a transition state called etioplasts. Etioplasts have tubular network of membrane material called the prolamellar body. When etioplasts are exposed to light, they are rapidly transformed into chloroplasts. At the same time prolemellar bodies are transformed to thylakoid membrane system. During chloroplast development there is an increase in size and number of plastids, appearance and expansion of the thylakoid membrane and assembly of the energy transduction complex/processes contained within this membrane. Thylakoid membranes were found to develop by invagination of the inner membrane (von Wettstein, 2001), in rapidly greening cells of the alga *Chlamydomonas reinhardtii* (Hoober et al., 1991) and in cryofixed developing chloroplasts in rice seedlings (Bourett et al., 1999). The inner membrane invaginations produce vesicles which fuse to form thylakoid system. The thylakoid membrane, separate from the inner envelope membrane, is differentiated into cylindrical stacks of "appressed" membranes, called "grana", that are interconnected with "unappressed",

stromal membranes. The highly elaborated, folded membrane system encloses a single, continuous lumen, which is an important compartment for the process of photosynthesis. This arrangement seems to maximize efficiency of the overall process.

Plastids have their own genome of 75-200 kilo base pairs. They can have as many as 10,000 genomes per cell depending upon the physiological state of the cell. The plastid genome contains 100-130 genes which encode ribosomal and transfer ribonucleic acids (rRNAs and tRNAs) and about 80 proteins mainly involved in photosynthesis, plastid gene transcription and translation. For example, the plastid genome of arabiopsis is composed of 154,478 bp (154.5 kb). It contains 87 protein coding genes, 4 ribosomal RNA genes and 37 tRNA genes (Fig.1). Due to the gene loss/transfer from the early plastid genome, plastids lost their autonomy as they encode a rather small number of proteins (around 90) although their proteome consists of 2500 proteins.
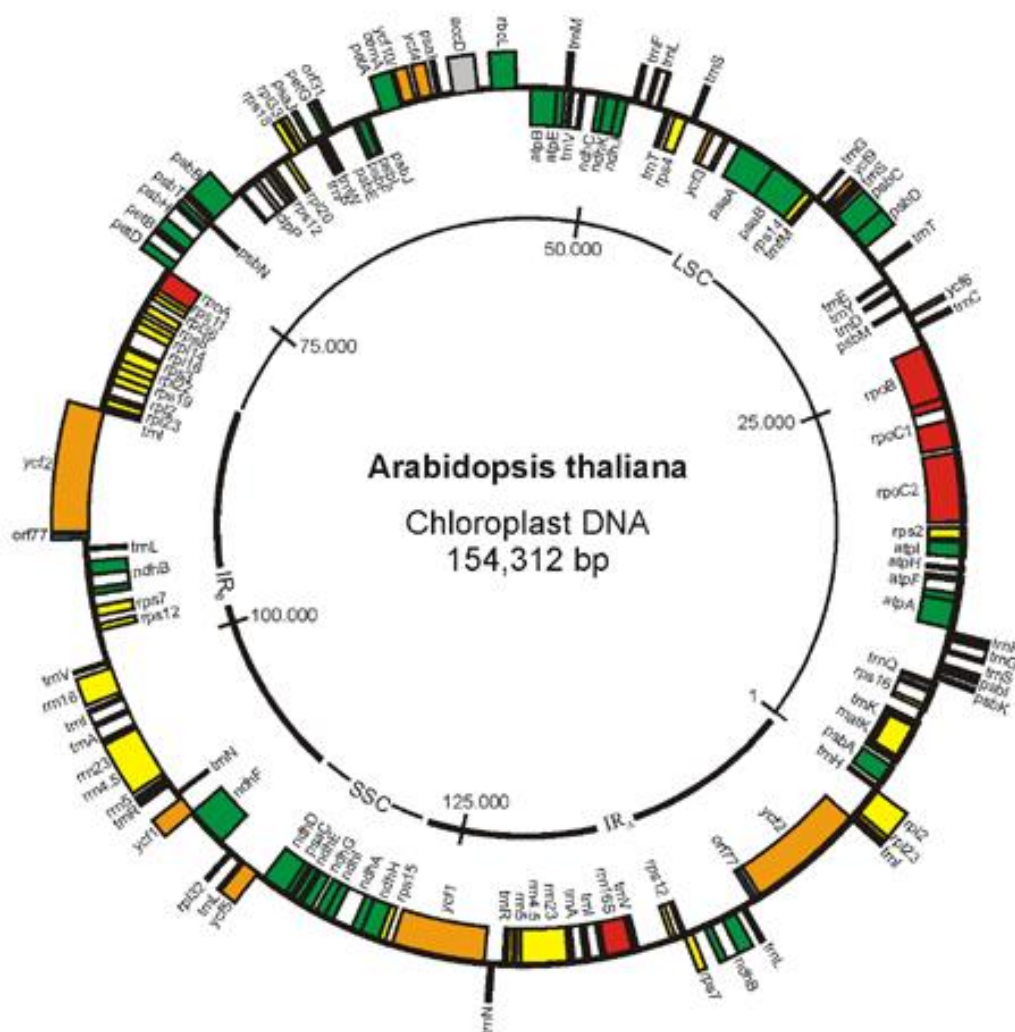
Figure 1. *A.thaliana* **chloroplast DNA** (inner circle: clock wise, outer: counter-clockwise). Function: transcription (red), translation (yellow), photosynthesis (green), tRNA (black), other (gray), unknown (orange). Sequence: AP000423. See (Sato et al., 1999).

Evidently, the plastid-encoded proteins are not sufficient for the efficient performance of the functions attributed to them. The majority of plastid proteins are encoded in the nucleus that are translated in the cytoplasm and transported to the plastids. The expression of nuclear and plastid genes need to be highly co-regulated for proper plastids' functionality.

The different types of plastids will shortly be described now. The nomenclature for different types of plastids is mainly taken from book titled " The Structure and Function of Plastids" edited by (Robert & Kenneth, 2007).

1.1. Proplastids

Proplastids are small (0.5 to 1 μm in diameter) and undifferentiated organelles. Internal membrane system of proplastids is not well defined as it has only a few tubules which are connected to the inner membrane of the proplastid envelope (Whatley, 1977). The proplastids found in meristematic tissues, embryonic tissues and tissue-cultured cells are called germinal proplastids. All other types of plastids develop as a result of differentiation of germinal plastids (Lancer et al., 1976; Pyke, 1999). Very little is known concerning the proplastid activity but a high level of gibberellic acid accumulation in meristematic tissue and developing seeds indicates that the proplastids might be involved in the biosynthesis of gibberellic acid. The discovery of colocalisation of *ent* Kaurene synthase (an important enzyme in gibberelic acid biosynthesis) with proplastids in developing wheat tissues further supported this idea (Aach et al., 1997).

A particular form of proplastids named 'nodule plastids' is found in nodule cells where nitrogen fixation takes place. They play an important role in the incorporation of the fixed nitrogen into a large number of metabolites.

1.2. Chloroplasts

Chloroplasts are green plastids (the colour is due to the high content in chlorophyll), which are present in leaves and unripe fruits of plants. Chloroplasts are essential for photosynthesis but they also take part in many metabolic reactions like biosynthesis of pigments, vitamins, plastoquinone and phylloquinone (vitamin K), fatty acids and lipids, aromatic and non aromatic amino acids, nitrogenous bases (purines and pyrmidines), isoprenoids and tetrapyroles. They are also involved in carbon oxidation via photorespiration (Ogren, 1984), in the starch synthesis and in a number of other metabolisms.

1.3. Chromoplasts

'Chromo' stands for colour. Chromoplasts are brightly coloured plastids found in fruits, flowers, and even in some root cells. They develop from chloroplasts in ripening fruits (Bouvier et al., 1998) while in other tissues they may arise from proplastids (Ljubesic, 1972). Deruere and collaborators found that chromoplasts contain fibrils, composed of supramolecular structures that contains a carotenoid core, a layer of lipid, and an outer layer composed of fibrillin (Deruere et al., 1994). Carotenoids can be of red, yellow or orange colour (Juneau et al., 2002). Chromoplasts are involved in pigment synthesis and storage. Chromoplasts lack both thylakoid membranes and photosynthetic apparatus. They help attracting the pollinators and fruit dispersing animals by their bright colour. Carotenoids such as β-carotene found in carrots and lycopene found in tomatoes act as antioxidants in the human diet (Yeum & Russell, 2002).

1.4. Etioplasts

Shoot proplastids differentiate in the presence of light. If there is no light or extremely low light, proplastids are developmentally arrested during the transition from proplastids to chloroplasts. These developmentally arrested plastids are called etioplasts. They are found in dark grown stem and leaf tissues but not in root tissues grown in dark (Newcomb, 1967). The structure of etioplasts is characterized by the presence of a complicated prolamellar body (PLB) composed of tetrahydrally branched tubules arranged in symmetry (B. E. S. Gunning, 2001; Kirk & Tilney-Bassett, 1967). Carotenoids are needed for the stabilisation of the PLBs. PLBs develop in the darkness (Park et al., 2002) and they have plastoglobuli and ribosomes trapped inside the PLBs (B. E. S. Gunning, 2004). PLBs also contain protochlorophyllide *a* which is a precursor molecule of chlorophyll *a*. When the exposed to light, PLBs are converted to thylakoid membranes and etioplasts are then converted into chloroplasts. Induction of etioplast formation in dark and the ability of etioplast to develop into chloroplast in light had made it an attractive system for the study of disruption of PLB structure, chlorophyll synthesis and thylakoid and chloroplast development (Baker & Butler, 1976; Krishna et al., 1999; Leech, 1984). Wellburn & Hampp (1979) reported that when the etioplasts are exposed to light, photosysem I (PSI) activity can be measured within 15 minutes, photosystem II (PSII) within 2 hours and water-splitting, proton pumping and ATP formation within 2 to 3 hr. As etioplast to chloroplast transition is a step-wise controlled process, the assembly of the various photochemical complexes is easy to dissect and understand Lebkuecher et al. (1999).

1.5. Leucoplasts

'Leuco' stands for 'white'. Leucoplast is the term used to categorize colourless, non-pigment-containing plastids. They store a variety of large molecules like starch in nonphotosynthetic tissues. They are usually found in roots responsible for monoterpene synthesis. They also lack both thylakoid membranes and photosynthetic apparatus.

1.6. Amyloplasts

Amyloplasts participate in starch storage having one to many large starch grains. They have a minimal internal membrane system. The presence of starch grains make them capable of graviperception. Like nodule proplastids, amyloplasts of some species like that of alfalfa (*Medicago sativa*) also contain a key enzyme of the glutamate synthase (GS-GOGAT) cycle and participate in nitrogen assimilation (Trepp et al., 1999).

1.7. Elaioplasts

They are also called as "elioplasts." They are usually small and round plastids. Presence of numerous oil droplets dominates their internal structure. Elaioplasts are oil-containing leucoplasts as they store fats. Their name is derived from "elaiov" which means olive. Among the proteins found in elaioplasts, plastoglobule associated proteins (PAP) such as fibrillin are predominating (Hernandez-Pinzon et al., 1999; Wu et al., 1997). They play an important role in pollen maturation (Hsieh & Huang, 2004; Pacini et al., 1992; Ross et al., 2000; Ting et al., 1998). The elaioplast sterol lipids coat the outside of the pollen grain, whereas the PAPs are degraded and do not appear in the coat (Hernandez-Pinzon et al., 1999).

1.8. Proteinoplasts

They are also called "proteoplasts". They are specialised for storing and modifying proteins. They were identified as plastids containing especially large and visible protein inclusions (Bain, 1968; Esau, 1944; Hurkman & Kennedy, 1976; Newcomb, 1967; Thomson & Whatley, 1980).

1.9. Gerontoplasts

'Geronto' stands for "old man". During the senescence of foliar tissue it develops from a chloroplast (Harris & Arnott, 1973; Parthier, 1988). Gerontoplasts develop as a result of unstacking of grana, loss of thylakoid membranes of chloroplasts and a massive accumulation of plastoglobuli (Harris & Schaefer, 1981). They play an important role in controlled dismantling of the photosynthetic apparatus during senescence.

Different types of plastids and their interconversions are depicted in Figure 2.
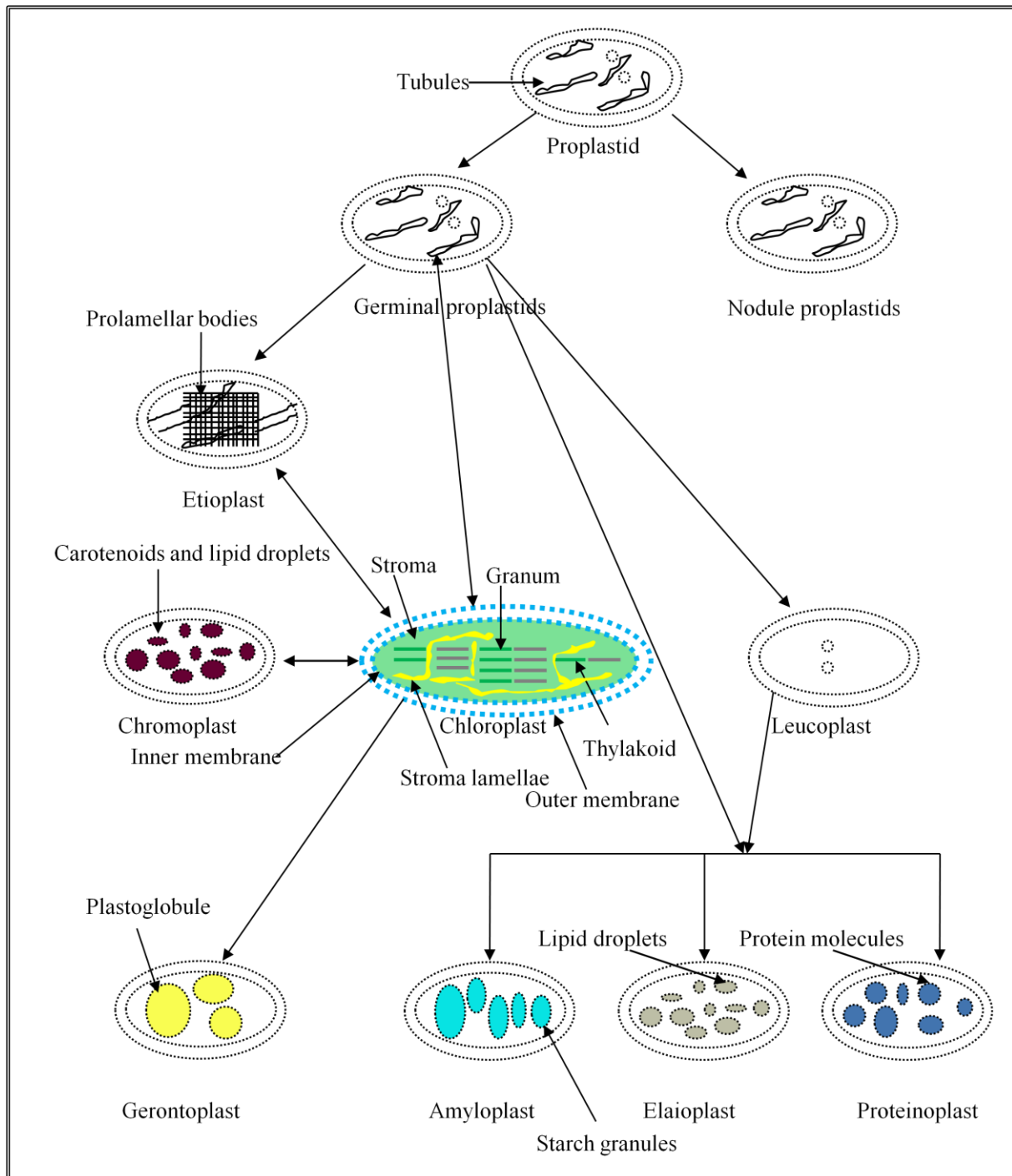
Figure 2. Types of plastids and their inter conversion (redrawn and revised from Møller (2005) and Robert & Kenneth (2007).

2. Origin of plastids

Plastids originated one billion (for some scientists 1.2-1.5 billion years) years ago as a result of an endosymbiotic event between an ancient cyanobacteria and a eukaryotic cell. (Martin & Russell, 2003; Palmer, 2003) proposed that the eukaryotic host in the endosymbiotic origin of plastid was a mitochondriate eukaryote. During the course of evolution, these endosymbionts lost most of their genome and from autonomous organisms they became semi autonomous cellular organelles. Due to the gene loss/transfer from the plastid to the nucleus, mechanisms were developed for organelle biogenesis and metabolite exchange (Dyall et al., 2004).

Like the endosymbiotic origin of plastids, mitochondria originated more than 1.5 billion years ago from α-proteobacterium like ancestor as a result of a single ancient invasion of an Archea type host (Gray et al., 1999). Some scientists believe in concurrent origin of eukaryotes and mitochondria (Martin et al., 2001; Martin & Muller, 1998; Moreira & Lopez-Garcia, 1998). There are two theories of mitochondrial origin, an anaerobic and an aerobic origin. Concerning the anaerobic origin, the methanotrophic proteobacterium provides essential compounds like hydrogen to the methanogenic archaean host (Martin & Muller, 1998). Those who believe in aerobic origin of mitochondria say that an aerobic proteobacterium invaded an anaerobic host (Andersson et al., 2003).

Schimper (1883) proposed probably for the first time that chloroplasts evolved as a result of symbiosis between non-photosynthetic host and a photosynthetic endosymbiont. This hypothesis was further supported/developed by Mereschkowsky (1905) that described plastids as " Little workers, green slaves" and reduced forms of cyanobacteria. The discovery of DNA in the chloroplast by Ris & Plaut (1962) further supported the endosymbiotic origin of plastids. This idea became generally accepted at the end of 1960s. On the basis of the different types of endosymbiosis, plastids were classified as primary, secondary and tertiary plastids which originated as a result of primary, secondary and tertiary endosymbiosis respectively. There has been a long standing debate regarding the monophylytic and polyphylytic origin of primary plastids. If we consider that a single mitochondriate eukaryote took up a single cyanobacterial host at a single moment of time and established a stable relationship with it which resulted in the gene transfer and integration of the endosymbiont as an organelle, this will be considered as monophylytic origin of primary plastid. If a single mitochondriate eukaryote established a stable relationship and integration of the endosymbiont as an organelle took place after taking up a single/multiple cyanobacterial host

at more than one moment of time, this will be considered as polyphylytic origin of primary plastid. Recently plastid origin from a single ancestor at a single point in time was viewed as an oversimplification by Larkum et al. (2007b) and a ''shopping bag model'' has been proposed for the origin of primary plastids (Howe et al., 2008; Larkum et al., 2007a; Larkum et al., 2007b). According to this model, the first endosymbiont might have persisted during certain rounds of host cell divisions and there might have been gene transfer from the endosymbiont to the host cell nucleus but finally this endosymbiont was lost. The successful integration of the endosymbiont as an organelle resulted after multiple rounds of intake and loss of the endosymbiont over a long period time.

3. Photosynthesis

It is a metabolic pathway that takes place in autotrophic organisms like plants, algae, cyanobacteria and in some aquatic animals/plants which can have both autotrophic as well as heterotrophic mode of life. Heterotrophic organisms cannot directly utilize light energy for their physiological or metabolic needs. Only the autotrophic organisms can convert light energy into chemical energy. In plants, photosynthesis takes place in green parts having chloroplasts in them. The specialized light absorbing green pigment of chloroplast is the chlorophyll which is found in thylakoids of chloroplast. The mesophyll of the leaves of higher plants is the most active photosynthetic tissue as it has large number of chloroplasts.

Plant chloroplasts use solar energy to oxidize H_2O, reduce CO_2 to synthesize organic compounds (carbohydrates; primary sugars) and at the same time to produce O_2 as a byproduct.

This complex reaction can be expressed in a simplified form in the form of following equation:

$$6\,CO_2 \quad + \quad 12\,H_2O + \quad photons \quad \rightarrow \quad C_6H_{12}O_6 \quad + 6\,O_2 \quad + 6\,H_2O$$

On the basis of utilization of light, photosynthesis could be separated in two phases:

1. **Light dependent reactions:** In these reactions sun light electromagnetic radiations are captured by chlorophyll and used to make high energy molecules. Water is oxidized to oxygen, $NADP^+$ is reduced to NADPH and ATP is produced.

2. **Calvin Cycle:** These reactions utilize the high energy products of light dependent reactions to fix CO_2 to make carbohydrates.

3.1. Light reaction

It takes place in the internal membranes, thylakoids, of the chloroplast. The light cycle is completed with the help of four major protein complexes: photosystem I (PSI), cytochrome b_6-f complex, photosystem II (PSII) and the ATP synthase. These membrane complexes are the integral membrane complexes and are found vectorially in the thylakoid membranes. Orientation of these four complexes results in the oxidation of water to oxygen in the thylakoid lumen, reduction of $NADP^+$ to NADPH on the stromal side of the membrane and the release of ATP molecules into the stroma as the protons (H^+) move from the lumen to the stroma. The photosystems (PSI and PSII) are spatially separated, physically and chemically different and each one has its own antenna pigments and photochemical reaction center. Electron transport chain links the two photosystems. The PSI reaction center, its antenna pigments, its electron transfer proteins and coupling factor enzyme are found in stroma lamellae and at the edges of the grana lamellae while PSII reaction center, its antenna chlorophylls and its electron transfer proteins are predominantly found in stacked regions of grana lamellae. Stroma lamellae are the exposed thylakoid membranes which lack stacking while grana lamellae are the stacked thylakoid membranes. The ATP synthase is located in the less curved regions of the grana end membranes and stroma lamellae (Daum et al., 2010). The antenna pigments are the pigments which collect light and transfer the energy to the photochemical reaction center where chemical reaction leads to long term energy storage. The antenna pigments consist of carotenoids, also called accessory pigments, chlorophyll b and chlorophyll a. The absorption spectrum of carotenoids is 400-500 nm while that of chlorophyll a is ≈ 430 and chlorophyll b is ≈ 450 nm. The energy transfer from carotenoids to chlorophyll is less efficient than from chlorophyll to chlorophyll. Carotenoids also play the role of photoprotection and avoid the production of nascent/ single oxygen. If the excited state of the chlorophyll is not quenched rapidly, a reaction excites the molecular oxygen to produce nascent oxygen which is hyper active and can cause damage to cellular components.

PSII, cytochrome b_6-f and PSI complexes are arranged in Z (zigzag) scheme. Transfer of excited electrons between a series of electron donors and acceptors is empowered by light energy. $NADP^+$ is the final electron acceptor and is reduced to NADPH. Proton motive force produced across the thylakoid membranes by the light energy and the resulting proton gradient are used by ATP synthase to produce ATP molecules. So the end products of light cycle, high energy molecules ATP and NADPH, are then used in the dark cycle during the CO_2 fixation.

3.1.1. Photosystem II and cytochrome b$_6$-*f* complex

Oxygenic photosynthesis in all photosynthetic organisms starts in a homodimeric multisubunit protein–cofactor complex which is embedded in the thylakoid membrane and is called PSII (Kern & Renger, 2007).

PSII is located in grana thylakoids and forms a supra molecular complex composed of polypeptides, pigments and co-factors. The protein subunits are either plastid or nucleus encoded. Water splitting complex oxidizes water and is also known as oxygen evolving complex (OEC). The reaction center is composed of several small polypeptides and homologous D1 and D2 proteins which are encoded by plastid *psbA* and *psbD* genes, respectively. The reaction center pigment (P680) receives the light energy directly or indirectly from the antenna pigments. The light energy causes an intramolecular rearrangement of the electrons and P680 changes from the ground state to the excited state (P680*). Then there is an intermolecular transfer of electrons from an electron donor (P680*) to an electron acceptor (pheophytine). Pheophytine is reduced by accepting electrons from P680*. The water splitting complex splits two molecules of H2O into 4 protons, 4 electrons and a molecule of oxygen. This reaction can be shown as under:

$$2H_2O \rightarrow 4H^+ + 4e^- + O_2$$

The electrons produced as a result of water splitting are then transferred to to a redox-active tyrosine residue, also called Z or Y_Z molecule, which reduces photoxidized P680 by transferring electrons to it. High energy electrons enter into an 'Electron Transport Chain" as pheophytine transfers these high energy electrons to the plastoquinone A (Q_A). Plastoquinone A transfers these electrons to the plastoquinone B (Q_B) which is a mobile electron carrier. The reduced Q_B transfers electron to the Cytochrome b$_6$-*f* complex which consists of a rieske iron sulfur protein (FeS$_R$), two b type cytochromes (cyt b) and a cytrochrome *f* with one covently bound heme *c* and subunit IV. Rieske iron sulfur protein has two iron atoms bridged by two sulfur atoms. Cytochrome b$_6$-*f* complex is distributed equally between the stroma lamellae and grana lamellae. But it is not the mobile carrier of electrons. Oxidised FeS$_R$ is reduced by accepting electrons from the reduced plastoquinone B and transfers theses electrons to cytrochrome *f* (cyt *f*). Cytrochrome *f* (cyt *f*) then transfers electrons to plastocyanin (PC). The reduced PC oxidizes itself by transferring electrons to the oxidized pigment (P700) of PSI. Along with the transfer of electron to oxidised FeS$_R$, plastoquinone B also transfers an electron to one of the cyt b and releases its $2H^+$ in the luminal side of the membrane. This reduced cyt b oxidizes by reducing the second cyt b which in turn transfers its electron to the

oxidized plastoquinone. The plastoquinone gets a semi-quinone form by accepting one electron and is fully reduced by the similar flow of another electron. This reduced plastoquinone (quinol) oxidizes by picking protons from the stromal side of the thylakoid membranes and changes into plastohydroquinone.

The net result of activity of the Cytochrome b_6-f complex is the transfer of two electrons to the P700 (PSI reaction center) and transfer of four proton from the stromal side of the membrane to the lumenal side. The electrochemical potential produced as a result of proton gradient is used by ATP synthase to produce high energy ATP molecules.

3.1.2. Photosystem I

PSI is located in thylakoids and consists of antenna chlorophylls, a reaction center (P700), phylloquinone, and iron-sulfur clusters. It is also a complex and highly organized trans-membrane structure. The light energy in PSI is also absorbed by the same trans-membrane proteins as the one used by PSII. But here the maximum energy absorption is at 700 nm so the pigment of chlorophyll is called P700. The overall reaction can be shown in the following way:

plastocyanin $\rightarrow$ P700 $\rightarrow$ P700* $\rightarrow$ ferredoxin $\rightarrow$ NADPH

$\uparrow$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\downarrow$

$\boldsymbol{b_6f}$ $\qquad\qquad$ $\leftarrow$ $\qquad\qquad\qquad$ **plastoquinone** b_6f
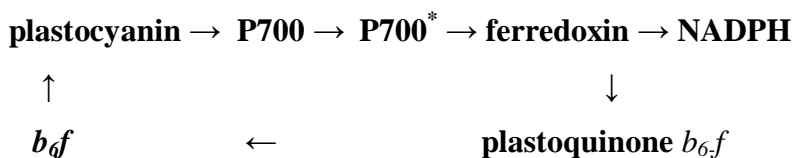
When P700 absorbs energy, it obtains excited state (P700*). The electrons from P700* are transferred to ferrodoxin, a water soluble carrier, through a number of intermediate carriers. These electrons can have two different fates. The non cyclic electron transport and the cyclic electron transport. In non cyclic transport, the electrons are carried by ferredoxin to the enzyme ferredoxin $NADP^+$ oxidoreductase which reduces $NADP^+$ to NADPH. While in cyclic electron transport, electrons are transferred to P700 via a proton pump, cytochrome b6/f. The proton gradient is then used by the ATP synthase to produce ATP. NADPH and ATP are the end products of light phase which are used in the dark phase for carbon fixation and carbohydrate precursor synthesis. Photosystem I is composed of 19-21 protein subunits, 175 chlorophyll molecules, 2 phylloqinone and 3 Fe_4D_4 clusters (Ben-Shem et al., 2003; Jensen et al., 2007). Among the protein subunits, it consists of 15 core subunits (PsaA-PsaL and PsaN-PsaP) while the peripheral antenna, LHCI, consist of six LHCa proteins (LHCa1-6)

but later on the existence of nine or ten LHCI proteins was also proposed but not yet confirmed (Dekker & Boekema, 2005). Most of the proteins of PSI are nucleus encoded and therefore should be transported from the cytosol to the plastid stroma and from stroma to be directed properly to the exact place in thylakoid membrane. This all renders it a great complexity. Some proteins (PsaA, PsaB, PsaC and PsaJ) are also encoded by chloroplast genes which add to the complexity of PSI subunit assembly process.

The electron flow from the PSII to cytochrome *b6-f* complex and PSI if schematically drawn in terms of it electric potential, forms a Z shape and hence called Z-scheme. It is depicted in Figure 3a.
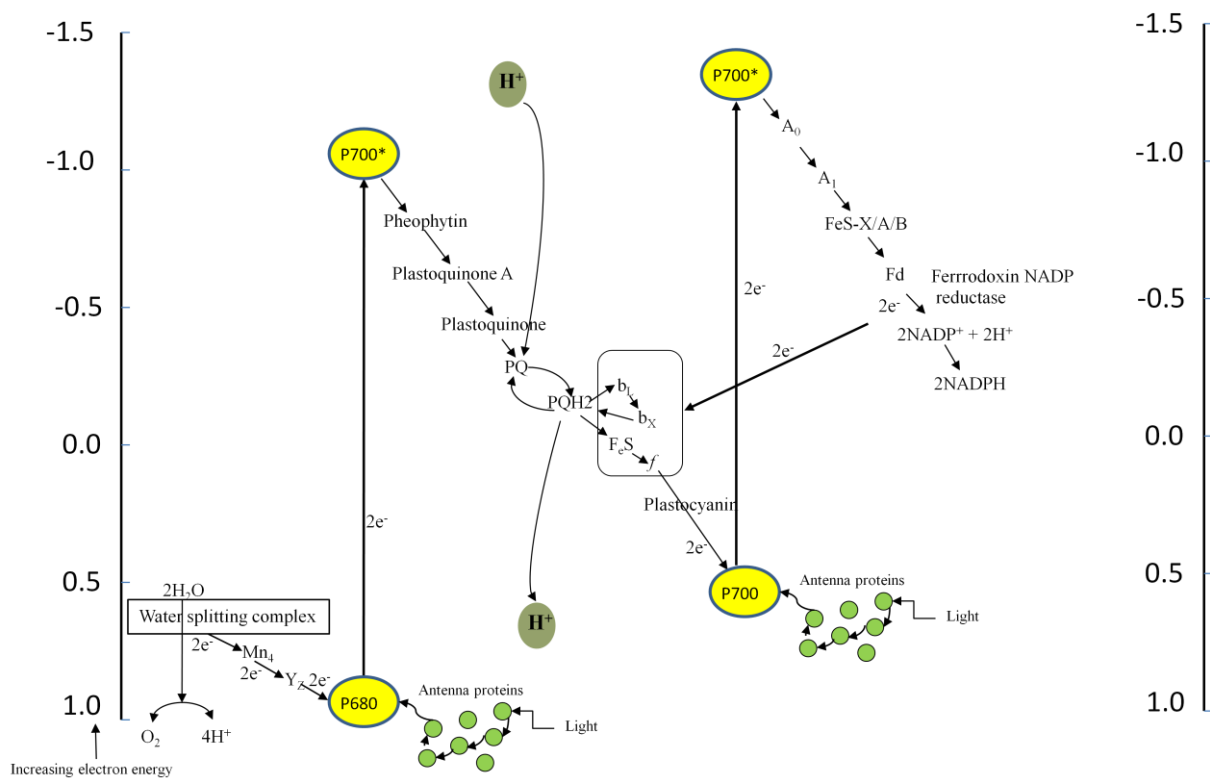
Figure 3a: Z-scheme of photosynthesis redrawn and revised from Shikanai (2007).

3.2. Calvin cycle

The Calvin cycle consists of a series of biochemical reactions that take place in the plastid stroma, are light-independent and cyclic. The substrate for carboxylation (ribulose-1,5-bisphosphate) is regenerated at the end of the cycle. The major enzyme involved in this cyclic reaction is RuBisCO (Ribulose-1,5-bisphosphate carboxylase oxygenase). It catalyses the carboxylation of a 5 carbon compound ribulose-1,5-bisphosphate by carbon dioxide in a two-step reaction. The ATP and NADPH molecules produced in the light cycle are used as energy source in this cycle.
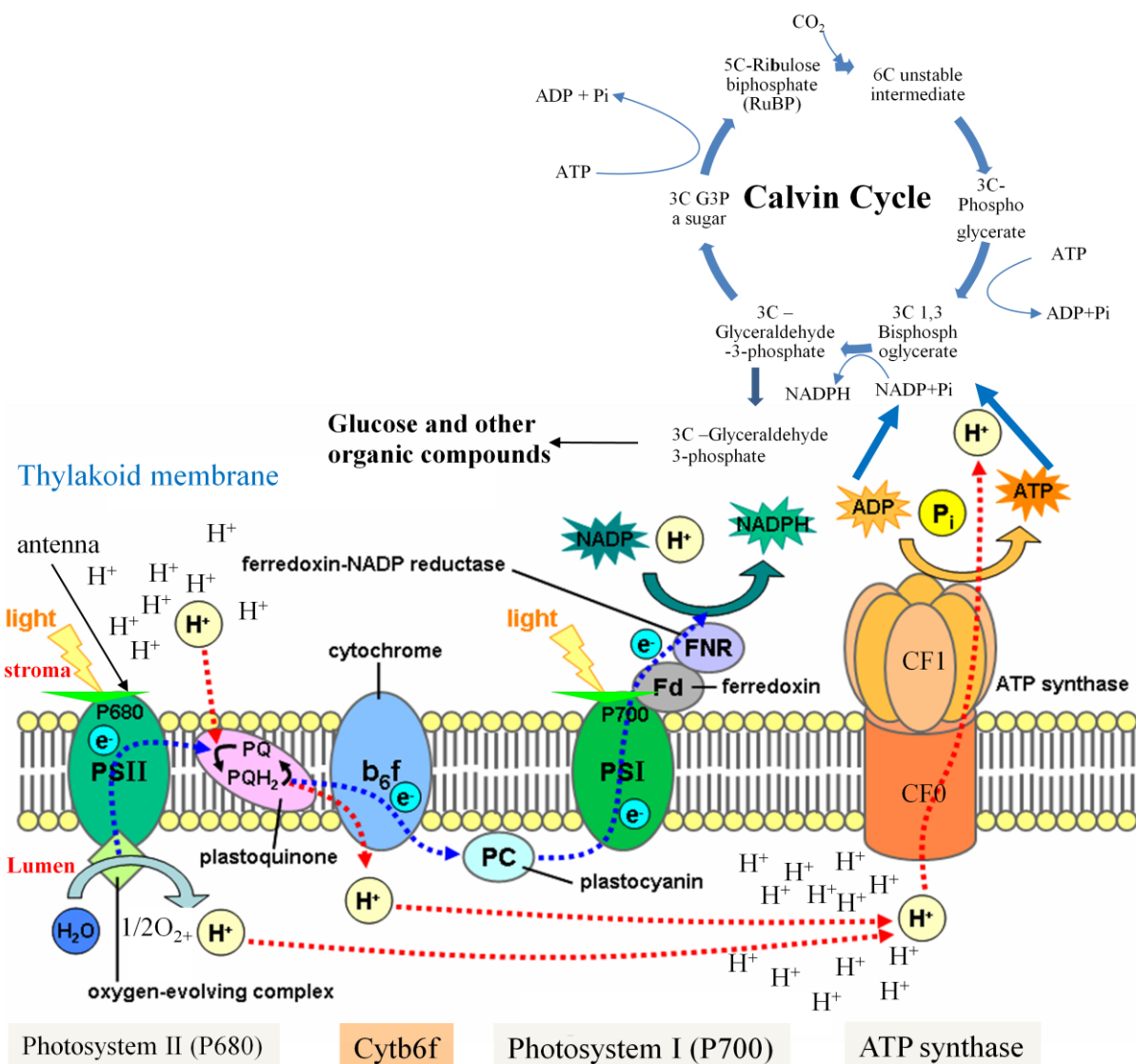
Figure 3b: Schematic presentation of the components and mechanism of photosynthesis.
(I used power point to modify the figure of **Tameeria** who created the figure in April 2007 based on: Taiz and Zeiger, "Light-dependent reactions of photosynthesis at the thylakoid membrane" *Plant Physiology*, 4th edition, ISBN 0-87893-856-7. The original figure is available on http://en.wikipedia.org/wiki/Photosynthesis).

The ADP and NADP+ produced during the dark cycle are taken up by light cycle to convert them again into ATP and NADPH molecules. The sum of reactions in the Calvin cycle is the following:

$3\ CO_2 + 6\ NADPH + 5\ H_2O + 9\ ATP \rightarrow$ glyceraldehyde-3-phosphate (G3P) $+ 2\ H^+ + 6\ NADP^+ + 9\ ADP + 8\ P_i$
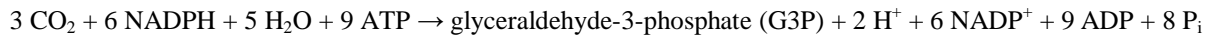
In order to produce one mole of glyceraldehyde-3-phosphate (G3P), 3 moles of CO_2 are needed which require three runs of Calvin cycle during which 9 ATP and 6 NADPH molecules are used. Glyceraldehyde-3-phosphate (G3P) is either converted to sucrose via triose phosphates or is used to regenerate ribulose-1,5-bisphosphate which again in calvin cycle. The localization of different components of photosynthesis, the flow of electrons from PSII to cyt b6-f and PSI and Calvin cycle are shown in figure 3b.

3.3. ATP synthase

It is an important enzyme involved in the ATP synthesis and hydrolysis in chloroplast thylakoid membranes during the process of photosynthesis by using proton motive force. It consists of two major portions, F1 and F0.

3.3.1. CF1

CF1 is a hydrophilic peripheral membrane protein complex attached to the thylakoid membranes and has five subunits α, β, γ, δ, ε in a stoichiometry of 3:3:1:1:1 repectively. It is almost spherical and has a width of ˷11nm and its height is ˷ 9 nm (Richter et al., 2000). Its subunits (α, β, γ, δ, ε) are encoded by atpA, atpB , atpC, atpD, atpE genes respectively where atpA, atpB and atpE are chloroplast genes while atpC and atpD are nuclear genes (Groth & Strotmann, 1999). Three αs and three βs together form a heterohexamer ring (Figure 3c).

This heterohexamer ring consists of six structural domains of an N-terminal β barrel and in its centre of an α-β domain containing the nucleotide binding site. The three catalytic sites of β subunits are β_T, β_D and β_E. Two of them are occupied by Mg.ATP and Mg-ADP while the third one is an empty site (Noji & Yoshida, 2001). Interestingly they have always different nucleotide binding state at any given moment of time. Subunits γ and ε are present in the central cavity of α, β heterohexamer ring and work as connectors between F1 and Fo (Figure 3c).

3.3.2. CFo

CFo is a hydrophobic thylakoid membrane embedded protein complex and has four subunits " a (IV), b (I), b' (II)', c (III)" in a stoichiometry of 1:1:1:14 respectively. The thickness of CFo was found to be 7.5 nm while it its diameter was estimated as 6.2 nm and 8.5 nm (Richter et al., 2000). Its subunits (I, II, III,IV) are encoded by atpF, atpG, atpH and atpI genes respectively (Figure 3c). *AtpI*, *atpF*, *atpH* are chloroplast genes while atpG is a nuclear gene.
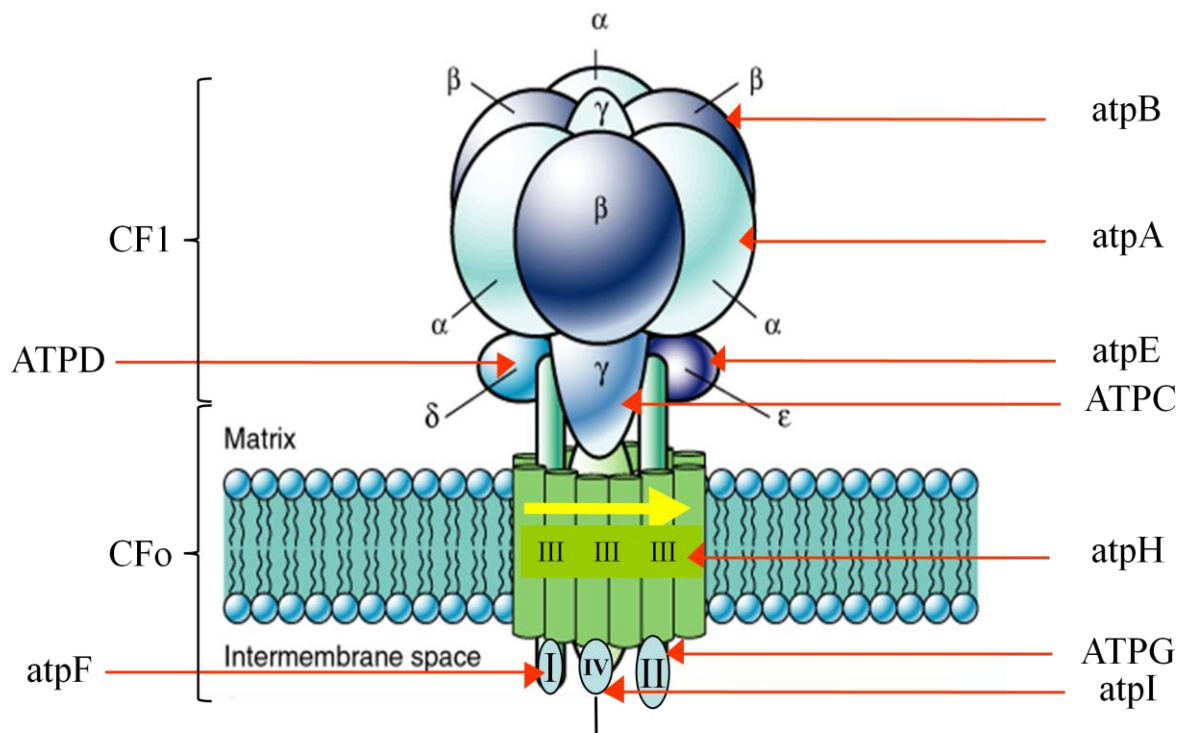
Figure 3c: Schematic presentation of the arrangement of the subunits of ATP synthase and the genes encoding these subunits.

This figure is extracted from http://www.bio.miami.edu

All the subunits have transmembrane α-helical structures. Subunits *b* and *b'* have only one transmembrane helix consisting of five helices. Subunit c has two anti-parallel transmembranes helices connected by two extramembrane polar loop (Groth and Strotmann 1999). Subunit *c* oligomer forms a ring outside where subunits *b, b'* and a are located (Neff et al., 1997). Subunits *b* and *b'* are involved in binding with CF1. They form a stalk and are connected with δ subunit of CF1 while *a* and *c* subunits are involved in H$^+$ translocation.

3.3.3. ATP synthase gene expression

The genes which encode subunits of CF0 and CF1 belong to two transcriptionally active compartments *i.e.* nucleus and plastids. The mRNAs of nuclear genes atpC, atpD and atpG are translated in the cytoplasm. The resulting soluble precursor proteins have plastid specific transit peptides. These transit peptides are proteolytically cleaved after the proteins have been transported to the thylakoid membranes of chloroplast for the assembly into the ATP synthase CFoCF1 complex. The nuclear genes involved in ATP synthase synthesis are expressed at the same time in response to light, organ specific factors and plastid derived signals (Bolle et al., 1996). In *chlamydomonas* the biosynthesis of chloroplast encoded subunits of ATP synthase is controlled at the translational level. Drapier et al. (2007) proposed a CES (*Control by Epistasy of Synthesis*) regulation for uneven stoichiometry of both the ATP synthase complexes *i.e.* CF1 and CF0. CES is a process in which the presence of a subunit is required for the sustained synthesis of another plastid subunit belonging to the same complex (Choquet & Vallon, 2000; Choquet & Wollman, 2002; Wollman et al., 1999). Drapier et al. (2007) reported that nuclear encoded subunit γ is required for a sustained synthesis of chloroplast encoded β subunit which in turn stimulates the translation of the chloroplast encoded α subunit of the CF1 complex. An important feature of CES is the negative autoregulatory feedback of beta on its own translation.

In higher plants, (Sakai et al., 1998) tested the effect of tagetitoxin (a PEP inhibitor) on the transcript accumulation of a number of genes in isolated chloroplast (from mature tobacco leaves) and proplastid nuclei (cultured cells of tobacco) by northern blot hybridization. In tobacco, they observed 95-99 % reduction in transcript accumulation of *atpA*, *atpB* in developed chloroplasts and only 40-50 % reduction in transcript accumulation of these genes in proplastids. This suggest that *atpA* and *atpB* which encode subunits of ATP synthase are transcribed by two distinct RNA polymerases in proplastids and in chloroplasts. Spinach *atpB* gene codes for five different transcripts having their 5' ends at positions -455, -275, -180, -100 and +1 from the translation initiation codon of *atpB* (L. J. Chen et al., 1990; Mullet et al., 1985). Tobacco *atpB* transcripts map at positions -610, -500, -490, -290, -225, -90 (Orozco et al., 1990; Shinozaki & Sugiura, 1982) while -610, -490, 290 and -225 mRNAs were identified as primary transcripts (Kapoor et al., 1997; Orozco et al., 1990). Kapoor et al. (1997) also reported the existence of non-consensus type promoters (NCII) (which are NEP dependent) in *atpB/E* operon responsible for its expression in non photosynthetic plastids which was

supported by the results of Sakai et al. (1998). Hirose & Sugiura (2004) reported that *atpB* mRNA needs an unstructured sequence encompassing the start codon for its translation as there is no Shine-Dalgarno (SD) sequence in the 5' UTR. *atpB* and *atpE* have overlapping stop and start codons in tobacco and arabidopsis. These overlapping codons result in the translational coupling of *atpB/atpE* transcripts in maize (Gatenby et al., 1989). In tobacco an *atpE* specific transcript of 1.3 kb was found by (Shinozaki et al., 1983), being issued from a promoter located within the coding region of the *atpB* gene (-430/-431 nt upstream of the *atpE* translation initiation codon) (Kapoor et al., 1994). Similarly an *atpE* transcript was found to be issued from a promoter located in the coding region of *atpB* (from -431 upstream of *atpE* translation initiation codon) in arabidopsis (thesis Wafa Zghidi, 2008). As *atpE* has a monocistronic transcript, the question is still open that whether in arabidopsis *atpE* is translationally uncoupled from *atpB* or not. The above mentioned results indicate that 1) majority of the transcripts for small ATP operon are issued from the promoters located in the 5' UTR of *atpB* but the promoter positions are not highly conserved among different species. 2) *atpE* is co-transcribed with *atpB* in higher plants but it is also transcriptionally uncoupled from *atpB* in tobacco and arabidopsis.

Early data indicated that spinach *atpI/H/F/A* contain almost 30 different transcripts which range from 0.5 kb to 6.0 kb in length (Cozens et al., 1986; Hudson et al., 1987; Stahl et al., 1993; Stollar & Hollingsworth, 1994). Miyagi carried out a detailed transcriptional analysis of the *atpI/H/F/A* operon in tobacco and found that *rps2* is also co transcribed with *atpI/H/F/A* (Miyagi et al., 1998). Miyagi et al., (1998) found three transcription initiation sites and four processing sites in the non coding regions of this operon. They found that one of the primary transcripts being issued from position -208 from the translation initiation codon of *atpI* is dependent on a non consensus promoter (NCII) while the other two (-131 *atpI* and -384 *atpH*) are synthesized from PEP dependent promoters.

The genes encoding for the subunits of ATP synthase complexes (CF_O and CF_1) in unicellular green alga *chlamydomonas* are dispersed throughout its plastid genome. Drapier and collaborators showed that *atpA* and *atpH* are co transcribed and are found in the same operon (*atpA-psbI-cemA-atpH*) (Drapier et al., 1998). They also showed a total of 8 transcripts for this transcriptional unit (*atpA*), three of them were described as primary transcripts and other transcripts were thought to be the result of processing events. Except these two genes none of the other *atp* genes are found in the same operon, rather minimum

distance among them is 7 kb (Woessner et al., 1987). The *atpE* is neither located on the 3' end of *atpB* nor has overlapping stop and start codon with it, rather it is located 92 Kb away from it in a single copy region (Woessner et al., 1987).

4. Plastid gene expression

4.1. Post transcriptional regulation in chloroplast

The production of a functional protein from the information stored in the plastid DNA is not simply a two step process of transcription and translation. It consists of post transcriptional and post translational modifications. The main post transcriptional modifications are listed below:

RNA processing is a process of the cleavage of polycistronic transcripts and generation of the 5' and 3' ends of the cleaved transcripts.

Editing changes the amino acids specified by the DNA sequence by converting cytosine (C) nucleotide to the uridine (U) nucleotide of an mRNA.

RNA splicing is a process in which (*cis* or *trans*) removal of group I and II introns is carried out.

Even though the plastid genome is derived from an ancient cyanobacteria, its gene expression process consists of both eukaryotic and prokaryotic features. In addition, due to the transfer of plastid genes to the nuclear genome and the need of crosstalk between the two organellar genomes, the plastid gene expression machinery has become complex and dependent on nucleus encoded proteins. Interestingly, original mechanisms of gene expression regulation have been developed in the chloroplast to optimize its integration into the eukaryotic cell. Take the example of the RNA polymerases required for transcription. Cyanobacteria which are ancestors of plastids have only one RNA polymerase consisting of core enzyme and sigma factors all encoded by their single genome. But in higher plants chloroplasts have two RNA polymerases, PEP (Plastid Encoded RNA Polymerase) and NEP (Nucleus Encoded RNA Polymerase). PEP is plastid encoded but it requires nucleus encoded sigma factors for promoter specificity and the second type of RNA polymerase is encoded by the nucleus and hence called NEP. Similarly in contrast to cyanobacteria where the polycistronis transcripts are directly translated, plastid polycistronic transcripts need to be first processed and edited and then translated. In addition, plastid mRNA require the different

trans factors for RNA stability (Stern et al., 2010), while bacterial mRNAs are stabilized by secondary structures.

4.1.1. Chloroplast ribonucleases

Ribonucleases are the enzymes which are responsible for maturation of precursor RNA and RNA degradation. Their activities depend on protein-protein interactions, protein modifications and RNA secondary or tertiary structures (Monde et al., 2000). There are two types of ribonucleases:

1- Endoribonucleases: They cleave the polynucleotide chain of RNA by breaking the phosphodiester band between two adjacent nucleotides.

2- Exoribonucleases: They cleave at the end (3') or the start (5') of a polynucleotide chain of RNA by breaking the phosphodiester band between two adjacent nucleotides.

4.1.1.1. Endoribonuclease

In chloroplast, endoribonucleases are thought to be responsible for 3' end formation (R. Hayes et al., 1996; Stern & Kindle, 1993) or initiation of RNA breakdown by breaking the stabilising stem-loop structures (H. C. Chen & Stern, 1991; Yang & Stern, 1997). The well known endoribonucleases in chloroplast are CSP41 (Chloroplast Stem-loop binding Protein, 41 kDa) (CSP41a and CSP41b), RAP 38 (Ribosome Associated Protein, 38 KDa) and RAP 41 (ribosome associated protein, 41 KDa), RNase E (Schein et al., 2008) , RNase J1(Zoschke et al., 2010), RNase P (Thomas et al., 2000), RNase Z (Canino et al., 2009) and CRR2 (Okuda et al., 2009).

4.1.1.2. Exoribonucleases

In chloroplast, exoribonucleases participate in the 3' and 5' end maturation and also in the 3'-5' polyadenylation assisted degradation pathway of mRNAs. The well known exoribonucleases in the chloroplast are Polynucleotide phosphorylase (PNP) (Hayes et al., 2006; Perrin et al., 2004), RNase R (RNR1) (Kishine et al., 2004), RNase II RNB (Yehudai-Resheff et al., 2007) and RNase J1 that is also a 5'-to -3' exoribonuclease.

4.1.2. Turn over

The neo-synthesized polycistronic mRNA transcripts in plastids are spliced, processed, stabilised and edited to produce translation competent mRNAs. Both the neo-synthesized and translation competent mRNAs enter into a degradation pathway. The difference of the rate of synthesis and degradation determines the turnover of the mRNA.

The decay or degradation of the mRNA takes place in three steps, 1) endonucleolytic cleavage, 2) polyadenylation and 3), exonucleolytic decay or turnover (Bollenbach et al., 2008; Stern et al., 2010). The translation competent mRNAs are protected at their 3' end by a stem loop structure of inverted repeat sequence (IRs) and by PPR proteins or other RNA binding proteins. The 5' end of such mRNAs are also protected by RNA binding protein like PPR proteins. When the transcripts have to be degraded, endonucleases cleave the secondary structures of the 3' end and in the 3' or 5' UTRs outside the PPR-protected regions. Endonuclease can also cleave the mRNAs internally where there are no attached ribosomes. (Stern et al., 2010) proposed endonucleolytic cleavage as the rate limiting step of the mRNA decay pathway. The endo-nucleolytically cleaved mRNAs are good substrates of polyadenylating enzymes. Polyadenylation can be catalysized either by PAP (poly (A) polymerases) or by PNPases (polynucleotide phosphorylase). PAP adds only homopolymeric poly (A) tails while PNPase adds heteropolymeric poly (A) tails. These polyadenylated transcripts are substrates for the exonuclease which degrades the RNAs in 3'-5' direction. If the exonuclease encounters a secondary structure, an endonucleolytic cleavage and polyadenylation are required further degradation of the transcripts. For 5'-3' degradation of transcripts, removal of the 5' end stability complex by an endonucleoytic activity is followed by a net 5'-3' degradation. The existence of a 5'-3' exonuclease dependent degradation activity has not yet been shown in chloroplast (Bollenbach et al., 2008; Stern et al., 2010).

In brief mRNA turn over depends on the extent of transcriptional activity, mRNA stabilising structures or proteins (stem loops of IRs at the 3' end, mRNA binding proteins like PPR) and the activity of enzymes responsible for mRNA degradation/decay (endonuclease, polyadenylation enzymes, exonuclease).

4.1.3. 5' End maturation

On the basis of their origin, transcripts found in chloroplasts of higher plants are divided in two types. The ones which originate directly from transcriptional events are called primary transcripts while the others which are produced as a result of processing of the

primary transcripts are called processed transcripts. Processed transcripts are the predominant form of transcripts found in chloroplast of higher plants. Even in *Chlamydomonas* no primary transcripts can be detected which indicate that most if not all of the transcripts undergo processing and 5' end maturation and that this event is very fast (Stern et al., 2010). Primary transcripts have a tri-phosphate on their 5' ends while the processed transcripts have only one phosphate on their 5' end. It is possible to discriminate between primary and processed transcripts using a combination of TAP treatment and RT-PCR techniques.

Two pathways are proposed for the 5' end maturation of the transcripts. The first one is an endoribonuclease pathway while the second one is a 5'-3' exoribonuclease pathway. Endoribonuclease starts maturation by site-specific cleavage of intercistronic transcripts or by dephosphorylation of primary transcripts. Once a free 5' end is available, the mRNA becomes a substrate for the 5'-3' exonuclease actually. The extent of this 5' maturation is determined by the presence of sequence specific binding proteins like PPR proteins which bind to the specific elements/sequence or by secondary structures. In addition RNA binding proteins can also guide endoribonuclease for site specific cleavage.

RNAse J is thought to play a role in maize chloroplast mRNAs maturation (Pfalz et al., 2009). RNAse E which is another exoribonuclease was found to be involved in the 5' end maturation of plastid RNAs (Mudd et al., 2008). *Arabidopsis thaliana* possess both RNAse J and RNAse E while *Chlamydomonas* has RNAse J but lacks RNAse E. RNase J and E are thought to possess both the endoribonucleolytic and exoribonucleolytic activities and are the best candidates for 5' end maturation of plastid transcripts (de la Sierra-Gallay et al., 2008).

4.1.4. 3' End maturation

Most of the mature 3' ends are produced as a result of processing and not by transcription termination which is inefficient in chloroplast. 3' end maturation is also carried out by exoribonucleases, endoribonucleases and RNA-binding proteins. In the most frequent situation, a newly synthesized mRNA is taken as a substrate by PNPase (polynucleotide phosphorylase) for a 3'-5' exoribonuclease activity. PNPase continues its activity until it encounters an IR (inverted repeat) secondary structure. The same pathway is observed if a free 3' end is obtained upon endonucleolytic cleavage inside a polycistronic transcript. In this context, Pfalz and colleagues (Pfalz et al., 2009) proposed that endonucleolytic cleavages inside polycistronic transcripts are random and are followed by 5'-to 3' and 3'-to 5' exonucleolytic activity in order to define overlapping 5' and 3' ends respectively. Sequence specific binding proteins as PPR10 play their role in stabilising the 5' and 3' ends and limiting

the 5' and 3' maturation. An example of this PPR10 hypothesis is a PPR protein HCF 152 that controls the intercistronic processing of *psbH* and *petB* in Arabidopsis (Meierhoff et al., 2003). In Chlamydomonas chloroplasts, the mature 3' end of atpB lies at the end of a stem loop forming inverted repeat (IR) sequence. The maturation of the 3' end of atpB pre-mRNA takes place by a two step process. As a first step, endonuclease cleaves 3' end of the pre-mRNA till 8-10 nt downstream of the mature 3' end. In the second step this intermediate product is trimmed by a $3' \rightarrow 5'$ exonuclease to produce a mature 3' end (Hicks et al., 2002). It has also been proposed a synergetic relationship between the IR and ECS (endonuclease cleavage site).

4.1.5. Splicing

Splicing is the process that consists in removal of introns from the precursor mRNAs. There are two types of introns found in precursor mRNAs, group I and group II introns which are 'self splicing' under non-physiological conditions *in vitro* (Saldanha et al., 1993) but not in land plants and *in vivo* need protein factors (Herrin et al., 1998).

Splicing of group I introns needs two times *trans* ester bond formation. The first bond is formed between 5' end of the intron and a nucleophilic guanine nucleotide and the second ester bond is formed between 3' OH of the upstream exon and 5' end of the downstream exon (Stern et al., 2010). Land plants have only one group I intron in *trnL* gene. Unicellular alga *C. reinhardtii* plastid gene *rrnL* has one intron while *psbA* has four introns (Turmel et al., 1993). All group I introns of *chlamydomonas* are autocatalytic *in vitro* but *trnL* of land plants has lost this ability (Goldschmidt-Clermont, 2008).

Six helical domains connected to a central core are conserved in group II introns (Fedorova & Zingler, 2007). Group II introns are divided into two subgroups on the basis of intron structure, group IIA and group IIB (Michel et al., 1989). Like group I introns, two steps of *trans* esterification are needed for splicing of this group of introns (Jarrell et al., 1988; Peebles et al., 1987). But here the 5' end of the intron is attacked by 2' OH (a nucleophile) of the bulged adenosine of the helix of the sixth domain (D6). The reaction results in the release of 5' exon and of the intron in the form of lariat. Then the 3' OH of the released 5' exon attacks the 3' splice site, reaction results in the union of 5' and 3' exons (spliced product) and release of the lariat intron (Fedorova & Zingler, 2007). The other mechanism of group II intron splicing is the hydrolytic splicing. This method is adopted in the absence of D6 free adenosine and OH$^-$ group of H_2O is used as a nucleophile (Daniels et al., 1996; Jarrel et al.,

1988; Vanderveen et al., 1987). The second step of *trans* esterification results in the release of ligated exons and lariat intron. *Arabidopsis* and tobacco have 20 group II introns in the plastid genomes while maize has 17 (Fedorova & Zingler, 2007).

Splicing of group I and II introns requires the involvement of nucleus-encoded factors. For example, 14 nuclear genes were found to be involved in trans-splicing of *psaA* precursor RNA (Goldschmidt-Clermont et al., 1990). The mature mRNA of *psaA* gene of chloroplast of *C. reinhardtii* is formed by three separate precursor RNAs in two steps of *trans*-splicing as it consists of three exons spread in the whole chloroplast genome (Herrin et al., 1998). Among the nucleus encoded factors four genes, *Raa1, Raa2, Raa3* and *Tr72* have been cloned. *Raa2* encodes a protein which belongs to pseudouridine synthase family while the rest of three encode novel proteins. Interestingly, a small non-coding RNA, *tscA*, which is plastid encoded, is required for splicing of exon 1 and 2 to occur (Goldschmidt-Clermont et al., 1991). Among the land plants, products of nuclear genes *crs1* and *crs2* were found to be involved in splicing of group II introns of maize chloroplast (Jenkins et al., 1997). *Crs1*(chloroplast RNA splicing-1) is involved in splicing of *atpF* introns (Jenkins et al., 1997), while *crs2* (chloroplast RNA splicing-2) is involved in splicing of many introns belonging to the subgroup B of the group II introns (Vogel et al., 1999).

Involvement of nuclear encoded proteins in splicing of group I and group II introns has been established. Using both genetic and biochemical approaches, 12 different nucleus-encoded proteins were found to be directly involved in chloroplast splicing (Kroeger et al., 2009; Schmitz-Linneweber & Barkan, 2007). Majority of these 12 nucleus encoded proteins which belong to a family of proteins containing four RNA-binding domains called CRMs (chloroplast RNA splicing and ribosome maturation) (Ostheimer et al., 2003). CFM2, a CRM domain protein, is required for *trnL* intron splicing. CFM3 is thought to participate in small ribosomal subunit biogenesis. PPR proteins also participate to RNA splicing. PPR4 contains 16 PPR repeats and one RRM domain and it is involved in the trans splicing of the *rps12-1* in maize and *Arabidopsis* (Schmitz-Linneweber et al., 2006). De Longevialle and colleagues showed that OTP51, having 7 PPR repeats and two RRM domains, is involved specifically in intron splicing of *ycF3-2* and in general for group IIA introns (de Longevialle et al., 2008). PPR 5 saves unspliced precursor of *trnG* from degradation (Beick et al., 2008; Kroeger et al., 2009). Kroeger et al., 2009 found the association of WTF 1 protein with the CAF1 (chromatin assembly factor 1) and CAF2 ribonucleoproteins (RNPs) in maize. WTF1 has DUF860 or PORR (plant organellar RNA recognition) conserved domain. Wtf1 mutant plants of maize showed defects in splicing of introns.

Along with nuclear gene products, subgroup IIA introns also need plastid gene products as there was failure of splicing of subgroup IIA chloroplast introns in nuclear mutants which lacked plastid ribosomes (Jenkins et al., 1997; Vogel et al., 1999). One good candidate for plastid gene products is the product of *matK* gene. MatK protein is encoded by an intron ORF inside the *trnK* gene and believed to act as an intron maturase. All the group IIA introns except *clpP* intron 2 are associated with *matK* (Zoschke et al., 2010). *Chlamydomonas* group I introns were found to contain open reading frames. Some group I and group II introns encode maturases which splice and stabilize introns. Protein encoded by these genes can act as endoribonucleases in some group I introns but not in group II introns (Goldschmidt-Clermont, 2008).

In brief, plastid introns of land plants require the assistance of proteins, which are both nucleus and plastid encoded. Their function is believed to be the stabilization of the catalytic RNA structure and /or assistance for correct folding of such structures.

4.1.6. Editing

The process of conversion of a specific nucleotide of RNA from C to U and less frequently from U to C in plant organelles is called editing. The apparent purpose of editing seems to convert an amino acid to a more conserved amino acid or sometimes to generate a start or a stop codon. Both *cis* and *trans* acting elements are required in the process of editing. Approaches using plastid transformation and editing-competent chloroplast extracts *in vitro* have been used to study the editing mechanism in different plastid mRNAs. The members of the PPR protein family are also known to be involved in RNA editing. PPR proteins are thought to bind to the *cis* acting elements and guide the recruitement of the editing enzyme on the precursor mRNA. A member of the PPR family, CRR4 was discovered as a first chloroplast editing *trans* factor by (Kotera et al., 2005) in an *Arabidopis* mutant which was defective in NAD(P)H dehydrogenase (NDH) activity. Okuda et al. (2006) later on, found that it is involved in the editing of *ndhD*-C2 which converts the ACG codon to AUG start codon. Okuda et al. (2009) and Okuda et al. (2007) discovered editing *trans* factors as CRR 21, CRR22, CRR28. Till now almost 17 different PPR and RRM proteins all encoded by the nucleus have been discovered which control the editing events in plastid transcripts in *Arabidopsis* (Stern et al., 2010), two in maize and one in rice plastid (Shikanai, 2007). Interestingly, till date, no editing event has been reported in the unicellular alga *chlamydomonas*. Editing not only occurs in open reading frames (ORFs) or coding regions but also it has been reported to occur in introns as well. Hammani and collaborators found the

C58 nucleotide of *rps12* intron-1 to be edited. The trans factor OTP81 is involved in recruiting enzyme for this editing site (Hammani et al., 2009).

Editing of tobacco *psbF* is necessary for its protein function (Bock et al., 1996) and editing of the *accD* mRNA to convert the serine to leucine codon is essential for a functional enzyme activity (Sasaki et al., 2001). These are a few examples where editing plays an important role in plastid gene expression but there can be silent editing sites as well which do not play any significant role in gene expression (Inada et al., 2004). Even though editing has been reported to be essential for RNA maturation and correct translation of certain gene transcripts (Kotera et al., 2005; Stern et al., 2010), in arabidopsis most of the editing sites were found to be non essential. It is not yet clear if the role of these non essential editing sites becomes important in certain unsuitable environmental conditions for plants.

Among the *atp* genes, editing in *atpA* and *atpF* was analysed in Tobacco, Arabidopsis, Pea, Rice and Maize. *atpA* C-264 nucleotide (cCc) is edited to (cUc) in pea and tobacco but not edited in *Arabidopsis*, Rice and Maize as they already have T in DNA. *atpF* C-31 (cCa) was edited to (cUa) in *Arabidopsis* and tobacco but not edited in Pea, Rice and Maize as they already have a 'T' in DNA at this position, reviewed in (Sugiura, 2008).

Gene expression pathways especially the post transcriptional regulation is shown in Figure 4.
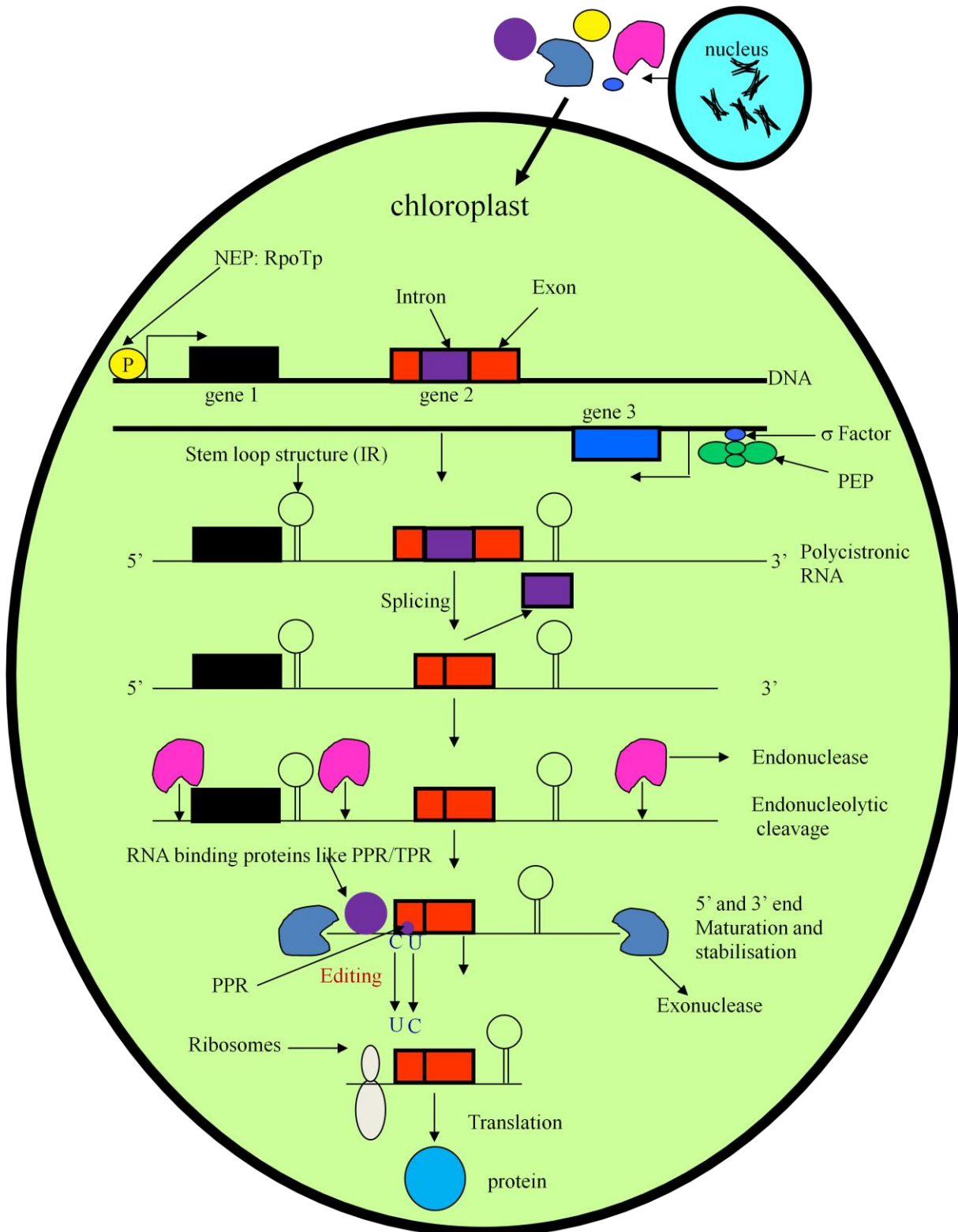
Figure 4: Gene expression pathways in chloroplast.

PEP (Plastid Encoded RNA Polymerase). NEP (Nucleus Encoded RNA Polymerase). PPR (Pentatrichopetide Repeat). TPR (Tetratrichopeptide Repeat). IR (Inverted Repeat). UTR (Untranslated Region).

4.2. Transcriptional regulation in chloroplast

Transcriptionally active compartments of green cells are nucleus, mitochondria and chloroplasts. Chloroplasts are semiautonomous, photosynthetic organelles having their own genome. It has been believed for a long time that the majority of plastid DNA is found in a circular form. In this regards, probably the most influential article was that of Kolodner & Tewari (1972) followed by confirmation of their results by Herrmann et al. (1975). Since then, even if some scientists found linear form of chloroplast DNA, no one dared to challenge the broken circle theory established by Kolodner & Tewari (1972). The chloroplast genome sequencing projects and restriction fragment mapping also resulted in circular maps of the chloroplast genomes. It was only in 2004 when Oldenburg & Bendich (2004) provided evidence that the most common form of DNA in chloroplast is linear, even though the circular form does exist.

The genome size in chloroplasts of higher plants is much reduced in comparison to the genome of their cyanobacterial ancestor. During the course of evolution genes were either lost or transferred to the nucleus. The genome of *Arabidopsis thaliana* chloroplast consists of almost 120 genes which encode for ribosomal RNAs (rRNAs), transfer RNAs (tRNAs) and proteins engaged in transcription, translation and physiological processes. Even though the chloroplast genome is reduced, the transcription machinery is very complex. In higher plants, the chloroplast genome is transcribed by two types of RNA polymerases, a Plastid Encoded RNA Polymerase (PEP) which is a multimeric eubacterial type RNA polymerase and two Nucleus Encoded RNA Polymerase (NEP), which are monomeric phage type RNA polymerases (Figure 5).
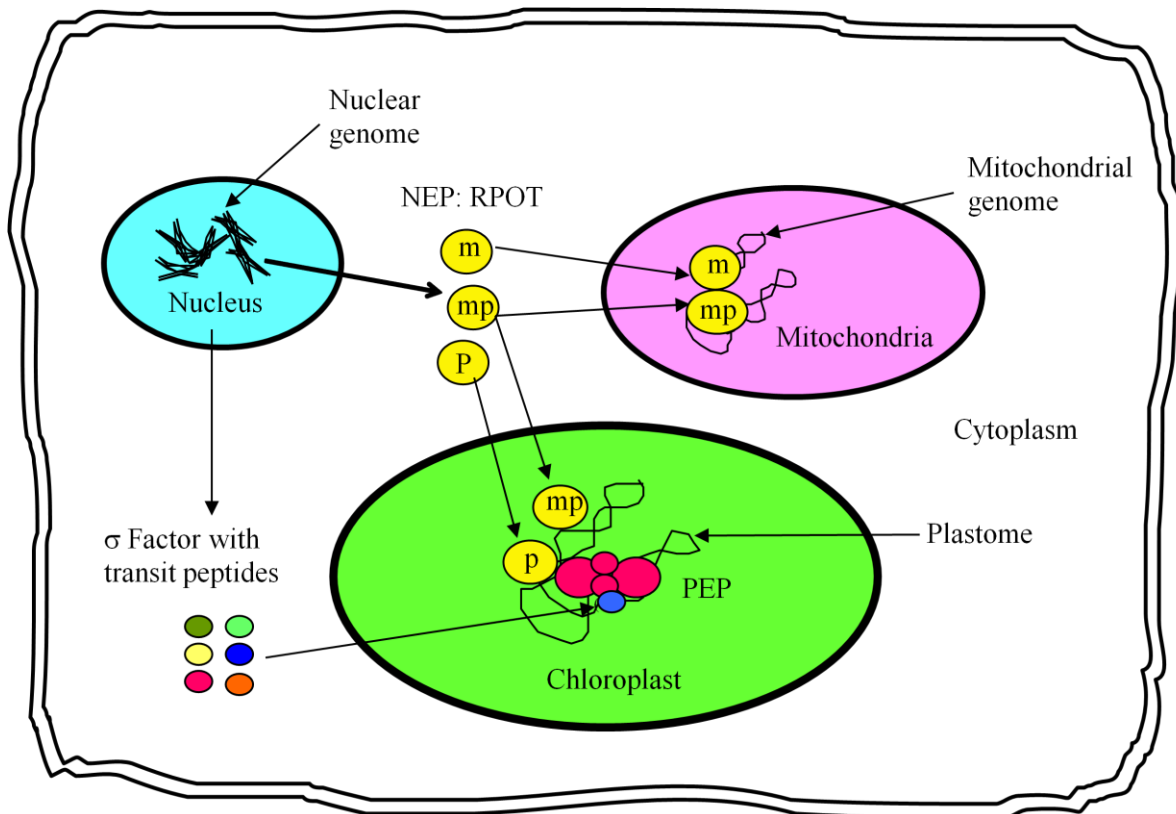
Figure 5: Mechanism of transcription in chloroplasts.
NEP stands for Nucleus Encoded RNA Polymerase. It is a monomeric phage type (T3 or T7) RNA Polymerase. RPOT (RNA polymerase of the phage T3/T7) is the name of the genes which encode NEP. RPOTm is exclusively targeted to mitochondria while RPOTp is targeted to plastids (chloroplast). RPOTmp is targeted to both the mitochodria and chloroplast. PEP stands for Plastid Encoded RNA Polymerase which needs nucleus encoded sigma factors for promoter specificity.

On the basis of the type of RNA polymerases needed to transcribe a gene, plastid genes are divided in three classes:

1. Class I: Genes transcribed exclusively by PEP (*psbB, psbD, psbE, petB, ndhA, rps14, atpH*)

2. Class II: Genes transcribed by both PEP and NEP (*atpB, ndhB, rrn, rps16, rpl33, rps33, rps18, clpP, ycf1*).

3. Class III: Genes transcribed Exclusively by NEP (*rpoB, accD, ycf2*).

4.2.1. Promoters

There are two types of promoters found in plastid genome, NEP and PEP promoters.

4.2.1.1. NEP promoter

On the basis of the architecture, NEP promoters are divided into the following subgroups.

Type Ia: NEP promoters having the conserved core motif i.e. YRTA where Y stands for (Pyrmidine-Cytosine/Thymine/Uracil) while R stands for (Purines-Adenine/Guanine).

Type Ib: NEP promoters having the conserved core motif YRTA and an additional GAA-box.

Type II: NEP promoters having un-conserved core motifs with no resemblance to any of the others consensus promoters.

4.2.1.2. PEP promoter

Standard plastidial PEP promoters are similar to the σ^{70} type promoters of *E. coli*. Plastidial PEP promoters have a conserved core sequence consisting of a -35 (TTGACA) box and a -10 (TATAAT) box. The distance between the -35 and the -10 elements is 17-19 nucleotides. Generally, the presense of the -35 and the -10 elements is necessary for the PEP dependent transcription initiation. Some promoters like the psbA promoter, have an extended the -10 box. This extended -10 box of *psbA* consists in an upstream 'TGn' motif. The extended -10 box is required for transcription initiation by the σ^{70} type RNA polymerase from the promoters which lack -35 element. In addition to the -35 and the -10 boxes an 'AAG' box has been described which is located upstream of the -35 box of the *psbD* BLRP promoter (**b**lue **l**ight **r**esponsive **p**romoter). AAG box is required for transcription from the *psbD* BLRP and is proposed to interact with RNA polymerase through AGF (AAG box factor) (Kim et al., 1999).

4.2.2. RNA polymerases

4.2.2.1. Nucleus Encoded RNA Polymerase (NEP)

At the beginning of the transcriptional analysis of chloroplast, it was believed that the chloroplast genome is transcribed by only one type of RNA polymerase *i.e.* Plastid Encoded RNA Polymerase (PEP). But this hypothesis was put under question by Morden et al. (1991) who showed that plastid transcription occurred in the parasitic plant *Epifagus*, which lacked PEP. Similar results were observed by Han et al. (1992) in maize mutants and by Hess et al. (1993) in ribososome deficient mutants of barley. In the coming years it was found that in the absence of PEP, a nucleus encoded RNA polymerase plays an important role in the transcription of plastid genes (Allison, 2000; De Santis-Maciossek et al., 1999; Hajdukiewicz

et al., 1997; Krause et al., 2000; Legen et al., 2002; Serino & Maliga, 1998). Another important finding by Lerbs-Mache (1993) was the discovery of a 110-kDa single subunit RNA polymerase which recognized the T7 promoter but not the σ^{70} promoters in *in vitro* transcription experiments. In the following, Genes transcribed by the nucleus encoded RNA polymerase were found to have non-consensus type promoters in their 5' UTRs (Allison et al., 1996; Hajdukiewicz et al., 1997; Kapoor et al., 1997). Furthermore, the transcription of these genes was not influenced by Tagetin, a specific inhibitor of the PEP (Kapoor et al., 1997; Sakai et al., 1998). This 'Nuclear Encoded RNA Polymerase' was called NEP. The genes which code for NEP were first discovered in *Arabidopsis* and *Chenopodium* and they were named RpoT (RNA polymerase of the phage T3/T7 type). Three types of Nuclear Encoded RNA Polymerases are found in Arabidopsis. Their recent nomenclature is given below:

RpoTp (RpoZ or RpoT;3 in former articles**):** Transcribed in the nucleus, translated in the cytoplasm and imported into the plastids.

RpoTm (RpoY or RpoT;1 in former articles**):** Transcribed in the nucleus, translated in the cytoplasm and imported into the mitochondria.

RpoTmp (RpoX or RpoT;2 in former articles**)** Transcribed in the nucleus translated in the cytoplasm and imported into both mitochondria and plastids (Hedtke et al., 1997, 2000; Hedtke et al., 2002; Hedtke et al., 1999; Kobayashi et al., 2002; Kobayashi, Dokiya, & Sugita, 2001; Kobayashi, Dokiya, Sugiura et al., 2001).

4.2.2.2. Plastid Encoded RNA Polymerase (PEP)

PEP is the major plastid RNA polymerase. It is believed to have its origin from the ancestral cyanobacteria. It consists of a core catalytic complex composed of the α 2, β, β' and β'' subunits encoded by plastid genes. PEP generally recognizes the prokaryotic-type consensus sequences (-35/-10). The -35 consensus sequence is TTGACA and the -10 consensus sequence is TATAAT. Promoter specificity of PEP is rendered by sigma factors (Allison, 2000; Liu & Troxler, 1996; Tanaka et al., 1996). Sigma factors are encoded by the nucleus, translated in the cytoplasm and imported into chloroplasts. The signal of chloroplast localization of sigma factors is present in the N-terminal transit peptides. In the *Arabidopsis* nuclear genome, six different sigma factors (SIG1-SIG6) have been found by sequencing studies.

All these sigma factors belong to the σ^{70} type sigma factors which can be divided into two groups *i.e.* important or essential sigma factors and non essential sigma factors. AtSIGMA3, AtSIGMA5 and AtSIGMA4 are thought to be non essential but specialized

sigma factors as knock out mutants of the corresponding genes do not show obvious visible phenotypes. AtSIG5 transcribes specifically from the light dependent *psbD* BLRP (Tsunoyama et al., 2004) and AtSIG4 from the ndhF promoter (Favory et al., 2005; Tsunoyama et al., 2004). AtSIGMA3 specifically transcribes the *psbN* and the *atpH* genes (Zghidi et al., 2007). AtSIGMA1, 2 and 6 act as more general but essential sigma factors as knock out mutants of the AtSIG2 and 6 have visible pale-green phenotypes that are all based on chlorophyll deficiency. All the genes coding sigma factors of Arabidopsis are unlinked and found on different chromosomes.

4.2.2.3. Expression of genes and division of labour between NEP and PEP in Plastids

The first model that was established concerning the functioning of NEP and PEP in plastids was the so-called "Cascade Model" which stated that at the early stage of chloroplast development, NEP transcribes house keeping genes *i.e*. genes encoding ribosomal proteins and genes encoding subunits of PEP in order to synthesize ribosomes and the PEP polymerase. According to this model, NEP is preferentially active during early stages of plastid development. With the advancement of chloroplast development, PEP takes over transcription and transcribes preferentially the photosynthesis related genes. It was further proposed that, switching from NEP to PEP activity during chloroplast development implicates inactivation of RpoTp by transfer RNA glutamate ($tRNA^{glu}$) (Hanaoka et al., 2005) and sequestration of RpoTmp to thylakoid membranes (Azevedo et al., 2008). During the establishment of photosynthetically competent chloroplasts, PEP transcribes photosynthesis related genes that are then differentially regulated by light (Mullet, 1993).

The "Cascade Model" was later challenged by the findings that 1) NEP and PEP are present in mature chloroplasts (Bligny et al., 2000), 2) the activities of both the NEP and PEP increase with chloroplast development in maize and 3) PEP is active even during germination *i.e*. in differentiating protoplasts and is important for efficient germination of seeds of Arabidopsis (Demarsy et al., 2006). From these results our group has proposed a new model of "Co-existence of PEP and NEP" during germination.

4.2.3. The transcriptional apparatus of *Chlamydomonas reinhardti* chloroplast

The chloroplast of *Chlamydomonas* has a different transcriptional machinery and arrangement of genes in transcriptional units when compared to higher plant plastids. Complete absence of transcription in PEP inhibited plastids (Guertin & Bellemare, 1979), the

failure to obtaine PEP knock out mutants (Fischer et al., 1996; Goldschmidt-Clermont, 1991) and the lack of tagetitoxin resistant transcription in plastids (Lilly et al., 2002) indicated that PEP is indispensable for plastid transcription and that there is no phage type nucleus encoded RNA polymerase. Only one nuclear *RpoT* gene was found which codes for a mitochondrion localized RNA polymerase. Most of the plastidial promoters do not have the -35 box but posses an elongated – 10 box. Surzycki & Shellenbarger (1976) at first reported a σ-like activity in chloroplasts. Immunological evidence of σ -like factors was then obtained by Troxler et al. (1994). Later on it was found that chlamydomonas posses only one gene (*CrRpoD*) encoding a σ -like factor (Bohne et al., 2006; Carter et al., 2004). Lilly and collaborators carried out a detailed northern blot, microarray and inhibitor analysis of the genome wide transcription in chlamydomonas chloroplasts and found that 1) sulfate deficiency results in 2-10 fold reversible decrease in transcription. 2) phosphate limitation results in 2-3 fold increase in transcript accumulation and 3) chloroplast lacks a nucleus encoded RNA polymerase (Lilly et al., 2002).

4.2.4. Role of sigma factors in plastid transcription of higher plants

4.2.4.1. SIGMA 1

AtSIG1 was at first characterized by Privat et al. (2003) in Arabidopsis by using an anti sense approach. The mutant plants did not show any visible phenotype. A T-DNA insertion mutant for sigma1 were later on characerised in O*ryza sativa* (Tozawa et al., 2007). OsSIG1 plays an important role in the chloroplast transcription of the genes of three different operons, *psaA*, *psbB*, and *psbE*. The reduction of transcripts was 68-89% for the *psaA* operon that consists of *psaA*, *psaB* and *rps14*; 41-48 % for the *psbB* operon (*psbB*, *psbT*, *psbH*, *petB* and *petD*) and 15-25 % for the *psbE* (*psbE*, *psbT*, *psbL* and *psbJ*) operon in *sig1* plants. Thus sigma1 regulates PSI and PSII gene expression in rice. The genes whose transcript accumulation was increased in Os*sig1* plants were *rpl22*, *rpoA*, *rpoB*, *rpoC1*, *petE*, *petA*, *psbG*, *ORF159*, *psbZ* and *psbI*. Northern blot analysis showed that the 5.2 Kb long transcript of *psaA* which is a tri-cistronic transcript was 46-49 % reduced in mutant plants. Similarly, the 2 Kb transcript of *psbB* was 23-34 % reduced in mutant plants while the 0.9 Kb transcript of *psbE* was reduced for 10-18 %. In contrast, the 2.6 Kb transcript of *atpB* (di-cistronic) was increased by 25-40%. The 1.9 Kb transcript of *rbcL* was present in the same quantities in wild type and mutant plants.

The western blot analysis showed that the PSI reaction center complex (PSAA-PSAB) was reduced by 26% in Os*sig1* plants and also the activity of PSI was reduced. However, mutant plants have a functional PSII and the reduction in the electron transfer through PSII is probably due to the defect in PSI. Mature leaves of *sig1* mutants have 1/3rd reduction in chlorophyll contents.

Using the yeast two hybrid system, a protein binding to the region 4.0 of SIGMA1 was found. This protein was named SibI (Morikawa et al., 2002). As it is known that region 4.2 is important for recognition of the -35 elements (Campbell et al., 2002), it was speculated that SibI may play an important role in modifying SIG1 promoter preference or regulating its activity. But experimental proof for this hypothesis is still lacking.

Recently, it was found that redox signals regulate the phosphorylation of SIG1 which in turn inhibits specifically the transcription of *psaA* gene. It was found that Thr-170 of SIG1 is phosphorylated. Under oxidative conditions of plastoquinone (PQ) the amount of phosphorylated SIG1 is increased (Shimizu et al., 2010, Lerbs-Mache, 2011). The authors found that phosphorylation of SIG1, through *psaA* gene expression, plays an important role in regulating the stoechiometry of PS-I and PS-II.

4.2.4.2. SIGMA 2

The expression of SIGMA2 is induced by red light as well as blue light (Mochizuki et al., 2004). A T-DNA insertion mutant for SIG2 shows pale green cotyledons and poor growth (Kanamaru et al., 2001). There was 15 % reduction in chlorophyll content in the mutant plants. The chloroplast number was the same as in wild type plants but plastids were smaller in size and the internal structures were poorly developed. *Sig2* plants showed reduced level of transcript accumulation of four out of the six tRNAs that had been analysed *i.e.* trnV-UAC, trnM-CAU, trnE-UUC, trnD-GUC. The transcript level of trnG-GCC and trnW-CCA remained unchanged. On the protein level, accumulation of the D1 subunit of PSII, cytochrome f, the β subunit of chloroplast coupling factor 1, Rubisco Large subunit, a catalytic subunit of clpP protease and an acetyle co-enzyme A Carboxylase subunit was reduced in *sig2-1* plants as compared to that in wild type plants (Kanamaru et al., 2001). The above described decrease in tRNAs might explain the strong reduction in protein accumulation. In addition to tRNAs, one of the three primary transcripts (starting at position -256 from the ATG translation initiation codon) of *psbD* was also highly reduced in *sig2* plants. However, the -946 primary transcript of *psbD* was more accumulated in *sig2-1* plants

as compared to that in wild type plants (Kanamaru et al., 2001). A SIG2 transcriptome analysis was performed by microarray and showed that on the mRNA level only the *psaJ* transcript was reduced in *sig2-1* plants. Transcripts of the other 47 genes were increased (Nagashima et al., 2004a). Most of the genes whose expression/transcript accumulation was increased in *sig2* plants were predominantly transcribed by NEP.

Sequence alignment of plant sigma factors with the *E.coli* primary sigma factor σ^{70} showed that SIG2 and SIG6 have the highest homology (Privat et al., 2003). AtSIG2 was also characterized in Arabidopsis by using an anti sense approach (Privat et al., 2003). SIG2 anti sense plants were deficient in chlorophyll during early stages of development (white cotyledons) and there was an efficient recovery of the wild type phenotype in later developmental stages. At later developmental stages, the phenotype of *sig2* anti sense plants was different from that observed in T-DNA knock out plants (Kanamaru et al., 2001). In SIG2 anti sense plants, chlorophyll deficiency was restricted to cotyledons while in T-DNA knock out plants the chlorophyll deficiency is found in leaves as well *i.e.* it continues till later stages of development. Two hypotheses were raised by the authors to explain the recovery of wild type phenotype in mutant plants. 1) Expression of SIG2 is regulated at the post transcriptional level and SIG2 deficiency is compensated by protein stability in later stages of development (Privat et al., 2003). 2) Phenotype recovery at later stages of development may be explained also because one of the other sigma factors may take over the role of SIG2 for promoter specificity and transcription. SIG3 was proposed as a strong candidate for substitution of SIG2 as it has a similar promoter specificity (Hakimi et al., 2000; Privat et al., 2003) and SIG3 protein is increased in the absence of SIG2 (Privat et al., 2003).

SIG2 is a soluble protein. Interestingly SIG2 protein is more abundant in cotyledons than in leaves while SIG3 protein is more abundant in leaves than in cotyledons. This observation could explain why SIG3 can substitute SIG2 more easily in leaves than in cotyledons (Privat et al., 2003). So, SIG2 might play a key role in cotyledons while SIG3 is required for transcription in leaves.

4.2.4.3. SIGMA 3

An anti sense mutant of AtSIG3 was characterized in arabidopsis by Privat et al., (2003) and no visible phenotype was detected. Later on, two T-DNA insertion mutants in SIG3 (*sig3-2*, *sig3-4*) from the SALK collection were characterised (Zghidi et al., 2007). *Sig3-2* has an insertion at the border of intron 1 and exon 2 while *sig3-4* has an insertion within exon 4. Both of these mutants showed no visible phenotype, thus confirming the result

of anti sense plants. Microarray analysis showed a strong reduction of the transcript level of *psbN* and a moderate reduction of the transcript level of all *atp* genes except for *atpI*. The *psbN* gene is located on the opposite strand of the *psbB* operon in the intergenic region between *psbT* and *psbH*. Primer extension and TAP RACE analysis showed that the promoter of *psbN* is located at position -32 from the ATG translation initiation codon of *psbN* and transcription initiation is totally under control of PEP SIG3 holoenzyme.

It was further shown that the *psbN* initiated transcript extends at its 3' end over the *psbT* gene. Thus the PEP SIG3 dependent *psbN* promoter not only controls the synthesis of the sense transcript of *psbN* but produces also an anti sense transcript to *psbT* mRNA. This observation suggests that SIG3 might be involved in the gene expression regulation of the *psbB* operon by anti sense RNA production (Zghidi et al., 2007). SIGMA3 holoenzyme was also found to be implicated in the transcription of *atp* genes and in particular of the *atpH* gene. Primer extension analysis showed that accumulation of the -413 *atpH* transcript was strongly reduced in *sig3* plants (Zghidi et al., 2007).

Interestingly, *atpH* codes for subunit III of the ATP synthase F0 complex; an integral thylakoid membrane complex (see Fig. 3C on p.16), PSBN has been localized to the thylakoid membranes (Ikeuchi et al., 1995) and SIG3 was also found to be membrane-bound (Privat et al., 2003). So far, SIG3 is the only sigma factor that has been shown to be membrane-bound. Altogether, these data might suggest transcription of SIG3 dependent genes in the vicinity of the final localization of the corresponding protein products.

Another peculiarity of SIG3 is that this factor seems to be inactive in its full-length form and functionality might require proteolytic cleavage to remove the N-terminal part of the factor (Hakimi et al., 2000; Homann & Link, 2003). In addition, SIG3 expression is independent of light (Homann and Link, 2003; Privat et al., 2003) and, as shown by in vitro transcription assays, this factor is able to recognize many promoters in the absence of competition with other sigma factors. Thus, in vivo SIG3 might serve to rescue the function of other sigma factors in case of problems.

4.2.4.4. SIGMA 4

A T-DNA knock out mutant for SIG4 was characterized which showed no visible phenotype at any stage of development (Favory et al., 2005). The T-DNA tagged line was selected from the INRA (Versailles) collection by PCR screening of DNA pools. The T-DNA was found to be inserted in the first exon of the SIG4 gene at 210 base pairs downstream of

the ATG translation initiation codon. It was shown by microarray and primer extension analysis that SIGMA4 mediates transcription of the sole transcript of *ndhF* being issued from position -320 upstream of the ATG translation initiation codon. The NDHF protein is a component of NAD(P)H dehydrogenase or NAD(P)H-plastoquinone oxidoreductase (also called NDH complex). Wester blot analysis showed that the absence of NDHF in *sig4* knock out plants causes a strong reduction (almost undetectable) in the accumulation of another subunit of the NDH complex *i.e.* the NDHH protein. The authors suggested that SIG4 controls the assembly of the NDH complex by specifically controlling the *ndhF* expression at the transcriptional level (Favory et al., 2005).

4.2.4.5. SIGMA 5

The expression of AtSIG5 is highly expressed in young leaves and less expressed in roots. Its expression is developmental stage specific as it is not expressed in 4 days old plants but highly expressed in 8 days and 22 days old plants.

A T-DNA knock out mutant (*sig5.1*, SALK049021) for SIG5 was characterized and showed no visible phenotype at any stage of development (Tsunoyama et al., 2004). The T-DNA insertion is in the last exon of AtSIG5 (Tsunoyama et al., 2004). The insertion site was found to be 1931 bp downstream to the translation initiation codon and adjacent to the conserved region 4.2 which is necessary for -35 recognition.

Onda et al., (2008) found that the only sigma factor that is essential for the transcription from the psbD blue light responsive promoter (BLRP) is AtSIG5 as in the mutant *sig5.1* transcription of *psbD* from BLRP is highly affected (Tsunoyama et al., 2004). In addition, it was found that the expression of AtSIG5 is induced by blue light (Mochizuki et al., 2004; Onda et al., 2008) suggesting that AtSIG5 acts as a mediator of blue light signalling (Tsunoyama et al., 2004).

In addition to *Atsig5.1*, there exists another T-DNA insertion mutant *sig5-2*, having the T-DNA insertion in exon 2. This mutant showed embryonic lethality. Indicating that AtSIG5 might act as an essential factor for plant reproduction (Yao et al., 2003). The different phenotypes of the two mutants are not yet understood.

4.2.4.6. SIGMA 6

Two SIGMA6 mutants for *Arabidopsis thaliana* were characterized. *sig6-1* having a T-DNA insertion in exon 4 (Ishizaki et al., 2005) and *sig6-2* having an insertion in exon 5

(Loschelder et al., 2006). The *sig6-1* mutant has a cotyledon specific pale green phenotype (< 20 % chlorophyll of wild type plants). The mutant phenotype is restored in 8 days old seedlings. Electron micrographs of chloroplasts of wild type and *sig6-1* mutant cotyledons (in 3 days old seedlings) showed that the mutant chloroplasts had smaller spongiform plastids and lesser internal thylakoid membranes. However, the chloroplasts in 7 days old mutant seedlings were indistinguishable from wild type plants. In addition, greening of etiolated seedlings is retarted in *sig6-1* mutants. It can be concluded that SIG6 is involved in light dependant differentiation of proplasts and etioplasts into chloroplast (Ishizaki et al., 2005).

Microarray and northern blot analyses of 4 days old *sig6-1*and *sig6-2* mutant seedlings showed a great reduction in transcript level of most of the PEP dependent gene transcripts and a general increase in the NEP dependent gene transcripts. These data suggest that SIG6 acts as a major general sigma factor at early stages of development. Most of the SIG6 recognised promoters posses -35/-10 (TATAAT/ TTGACA) conserved elements. Microarray and northern analyses of 4 days old *sig6-1* plants showed a decrease in class I genes (*rbcL*, *psbA*, *psbB*, *psbC*, *psbD*, *psbH*, *psbN* and *psbT*) and class II genes (*rrn16*, *rrn23*, *rrn5*, *rrn4.5*) (Ishizaki et al., 2005; Loschelder et al., 2006). Transcript levels of NEP dependant genes of class II and class III including *clpP*, *rps15*, *ndhB*, *ycf1* and *rpoB*, *rpoC1*, *rpoC2* respectively were increased. But this reduction was restored in 8 days old plants where the pale green phenotype is also restored. The expression of the *trnE* operon and *trnQ* (UUG) was reduced in *sig6-1* mutants of 4 days while trnV (UAC) did not decrease.

Northern blot analysrs indicate that SIGMA 6 controls the transcription of the long transcript of 2.6 Kb of the *atpB/atpE* operon from the fourth day of development onward. The 2.0 Kb long transcript of *atpB*/*atpE* and 0.7 Kb monocistronic transcript of *atpE* are independent of SIG6. In the absence of SIG6 there is a promoter switch in which an additional promoter controlling transcription of a 4.8 Kb RNA is activated (Loschelder et al., 2006).

From the above mentioned results, it seems that SIG6 has a dual role: it is an early general sigma factor for PEP dependent genes as *psbA, rbcL* and *trnV* while it plays an important role as a late sigma factor for *atpB*.

Another interesting feature of SIG6 is its phosphorylation state. Tiller & Link (1993) described a serine/Thrionine Kinase (cpCK2) that was found to be associated to the PEP polymerase (Baginsky et al., 1997). Very recently, Schweer et al. (2010) used computational tools to identify potential sites for cpCK2 phosphorylation in SIG6 factor. They observed changes in *in vitro* binding of SIG6 to *atpB* promoter sequence and *in vivo* alteration of the *atpB* transcript profile when they disrupted the putative cpCK2 phosphorylation sites of the

SIG6 factor. These data represent the first strong evidence for a role of phosphorylation of SIGMA factors in the regulation of plastid gene transcription.
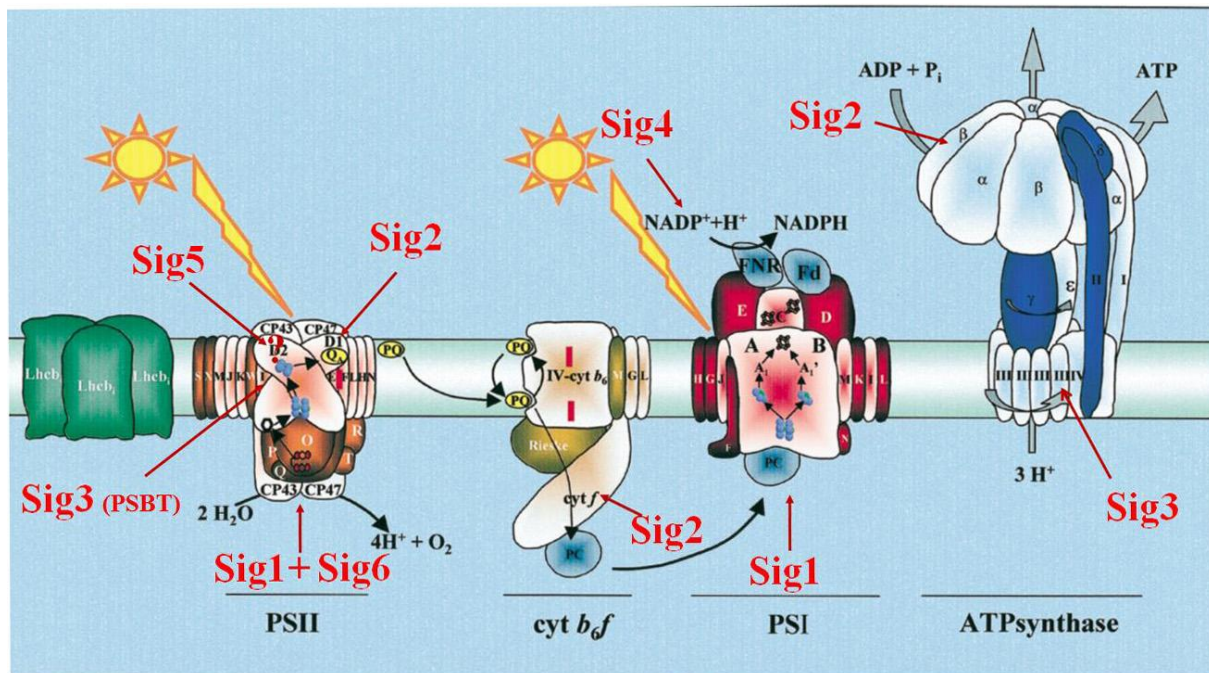
Figure 6: Summary of the role of sigma factors in controlling different components of photosynthesis.
The Sigma factors are indicated on the figure which is taken from (Choquet & Vallon, 2000).

The summary of the role of different sigma factors in plastid transcription is given below in figure 6 and in table 1.

Table1: Summary of the role of sigma factors in plastid transcription.

Sigma factor	Role in Transcription
SIG1	- controls transcription of *psaA* operon-*PSI* (68-89 %), *psbB* operon-*PSII* (41-48%), *psbE* operon-*PSII* (15-25%). - expression of following gene increased in *Os sig1: rpl22, rpoA, rpoB, rpoC1, petE, petA, psbG, ORF159, psbZ and psbI* - in *Os sig1* 2.6 kb *atpB* transcript increased by 25-40 %. - in *Os sig1* PSI reaction center complex (psaA-psaB) reduced by 26 %. - in *Os sig1* PSI activity decreased. - in *Os sig1* there is functional PSII.
SIG2	- 15 % reduction in chlorophyll content in the mutant plants. - In *sig2* mutant plants there is decrease in the transcript accumulation of : (trnV-UAC,

	trnM-CAU, trnE-UUC, trnD-GUC) and the transcript issued from -256 *psbD and psaJ* transcript.
	
	- In *sig2* mutant plants transcript accumulation of 47 genes was increased.
	- In *sig2* mutant plants there was reduction in the protein accumulation of:
	  -D1 subunit of PSII, cytochrome f, β subunit of chloroplast coupling factor 1,
	  - Rubisco Large subunit, a catalytic subunit of *clpP* protease and
	  -an acetyle co-enzyme A Carboxylase subunit.

SIG3	- Controls specifically the transcription of the -413 *atpH* transcript. - Controls transcription of the -32 *psbN* and thereby produces and anti sense transcript of *psbT*.
SIG4	- Controls the assembly of the NDH complex by controlling the transcription of *ndhF* gene.
SIG5	- It specifically controls the transcription from blue light responsive promoter of *psbD*.

SIG6	- In *sig6* mutant chlorophyll is < 20 % of wild type plants).
	- It is involved in light dependant differentiation of proplasts and etioplasts into chloroplast
	- *sig6* mutant has reduced transcript accumulation of :
	  -*rbcL*, *psbA*, *psbB*, *psbC*, *psbD*, *psbH*, *psbN*, *psbT, rrn16, rrn23, rrn5, rrn4.5, clpP,*
	  rps15, ndhB, ycf1, rpoB, rpoC1, rpoC2, trnQ genes
	  (UUG) and *trnE* operon.
	  -2.6 Kb transcript of the *atpB/atpE*

4.2.5. The transcriptional Active Complex (TAC)

It has recently been found that along with the sigma factors, transcriptional active chromosome (TAC) associated proteins are also required for the PEP activity.

Till today almost 35 PEP associated or PEP co-eluted proteins have been found (for review see Lerbs-Mache, 2011). Although knock out mutants of (pTAC2, -6 and -12) have shown that these TAC proteins influence plastid gene expression their exact function in the plastid gene expression has not yet been elucidated. Some of these proteins are targeted to plastids and to nuclei and their influence on plastid gene expression could be indirect via regulation of other nucleus-encoded genes.

Project

In the last decades the chloroplast has become an attractive tool for biotechnology. The possibility to transform the chloroplast of an increasing number of plant species, the tolerance of the chloroplast for accumulation of high levels of foreign or endogenous proteins and the ease to engineer its genome *via* homologous recombination make the chloroplast an ideal site to engineer resistance traits, to improve metabolic pathways and to produce pharmaceuticals in plants (Bock, 2007). As I would like to work later in my carrier on the problem of producing substances of pharmaceutical/medical interest in plastids, I am strongly interested in finding an endogenous specific and inducible transcription system for plastid transgene expression in dependence on the tissue, organ, developmental stage and/or external stimuli.

As already mentioned in the introduction section of this manuscript, the plastid genome in Arabidopsis is transcribed by three different types of RNA polymerases, two NEPs (RPOTp and RPOTmp) and one PEP. RPOTp represents the principal RNA polymerase to transcribe NEP controlled mRNA genes (mainly housekeeping) during early developmental stages while RPOTmp transcribes specifically the *rrn* operon from the PC promoter during seed stratification (Courtois et al., 2007). The PEP transcribes mainly and differentially photosynthetic genes in association with one out of six sigma factors in later developmental stages. Thus, RPOTmp and RPOTp are active only in early developmental stages excluding the possibility of a large production of engineered proteins. In this complex scenario, the best candidate for the establishment of an endogenous specific transcription system is the PEP system in combination with a specific sigma factor.

In general, eubacterial sigma factors are classified into primary and alternative sigma factors (Lonetto et al., 1992; Wosten, 1998). Primary sigma factors control transcription initiation of a large number of plastid genes and they are essential for growth and survival. In contrast, alternative sigma factors have more specific functions. For the higher plant chloroplast system we can consider SIG2 and SIG6 as primary sigma factors. They transcribe many genes and Arabidopsis knock-out mutants of *SIG2* and *SIG6* show strong chlorophyll deficient phenotypes. For SIG5 the question is not completely clear. SIG5 was supposed to play an important (primary) role during seed development as a T-DNA insertion in exon 2 was shown to be embryo-lethal (Yao et al., 2003). However, the same insertion mutant has been characterized as non-lethal by Nagashima et al. (2004b) placing SIG5 among the

alternative sigma factors. Thus, alternative sigma factors are SIG1, SIG3, SIG4 and SIG5 where knock-out mutants do not show visible phenotypes.

Considering the alternative sigma factors of the PEP transcriptional system, SIG3 seems to be the most interesting candidate for three different reasons:

- First of all, the recognition of the -413 *atpH* promoter was shown to be absolutely specific to SIG3 (Zghidi et al., 2007) thus excluding minimal transcription from any other sigma factor. In contrast to SIG1, SIG4 and SIG5 specific transcripts, not even traces of SIG3 dependent -413 initiated *atpH* transcripts are detectable in the corresponding knock-out mutants. Using such a promoter in plastid transformation experiments would provide an absolute specific expression of the transgene by SIG3-PEP.

- The second reason is related to the regulation of SIG3 activity that might imply proteolytic cleavage for activation in analogy to the activation of sigma K of *E. coli* (Hakimi et al., 2000; Homann and Link, 2003). In contrast to most of the other sigma factors, SIG3 is also expressed in darkness (Homann and Link, 2003; Privat et al., 2003). However, ATPH and PSBN are a priori proteins engaged in photosynthetic complexes and their synthesis should be light dependent. Thus, proteolytic activation of SIG3 might be triggered by light and/or under specific stress conditions.

- The third reason concerns the regulation of plastid gene expression by anti sense RNAs. It has been shown that transcription from the SIG3 specific *psbN* promoter produces anti sense RNA to *psbT* mRNA (Zghidi et al., 2007) which is likely involved in the regulation of *psbT* mRNA translation (thesis of Wafa Zghidi, 2008) thus providing an ideal system to analyze this new path of plastid gene-expression regulation.

According to these three reasons, I focused work during my thesis on

(a) the characterization of the expression of the *atpH* gene at the -413 SIG3 specific promoter and the gene expression analysis of all the other *atp* genes which code for the subunits of the *atp* synthase complex.

(b) the effect of different light conditions, from low intensity to photo-oxydative light, on the accumulation of SIG3-PEP specific transcripts, *i. e. psbN* and *atpH* mRNAs. Evidently, the analysis of *psbN* transcripts needs to be extended to the analysis of *psbT* sense and anti sense transcripts (see point c).

(c) the exact mapping of *psbT*/*psbN* sense/anti-sense transcripts and their expression levels.

In detail, these problems have been addressed as follows:

(a) The question of the specificity by SIG3 from the -413 *atpH* promoter is connected to the general question of expression regulation of all plastid genes coding for proteins of the ATP synthase complex. The ATP synthase complex is formed by different proteins that are either nucleus or plastid encoded. The plastid genes that encode the subunits of ATP synthase are organized in two different operons, the large operon consisting of the *atpI/H/F/A* genes and the small operon consisting of the *atpB/E* genes. The *atpH* gene product is present in a much higher stoichiometry than all the other subunits. The -413 SIG3 specific *atpH* promoter might therefore be necessary to reach a high level of *atpH* mRNAs for producing high levels of protein. In the first chapter, I have analyzed the expression of the two ATP synthase operons in detail in order to find out whether there is a co-regulation of the expression of the two ATP synthase operons by SIG3 or by any other sigma factor and whether SIG3 plays an important role in the higher stoichiometry of *atpH* at transcript or protein level.

(b) In the second chapter I have analyzed the effect of different light conditions on the mRNA levels of the plastid transcriptome. Although the most interesting part for me corresponded to the mRNAs that are made by SIG3-PEP, I have used the general approach of macroarray analyses to get an overview on all plastid mRNAs at once. As light probably will influence many other mRNA levels (to remind that all sigma factors are expressed in a light-dependent manner) it is important to distinguish between general (or other) and SIG3-specific light effects. This can be done by a complete transcriptome analysis where we can "filter" the SIG3-specific effect from the general effect if we know exactly which genes are transcribed by SIG3-PEP. In addition, if SIG3 activity is regulated by light-triggered proteolytic cleavage we would expect that the light conditions for this activation are different from the light conditions that induce the expression of other sigma factors. Exposure of plantlets to different light conditions should therefore result in different plastid transcriptome patterns. Some selected mRNAs have been characterized in more detail by primer extension analyses.

(c) In the last chapter (chapter 3) of my thesis I have characterized the transcripts of the *psbT* sense and the *psbT* anti-sense RNA in more detail in order to get a better idea on the formation of possible sense/anti-sense RNA hybrids. Understanding the sense and anti-sense RNA hybrid formation of *psbT* will help to better explain the role of anti-sense RNA in gene expression regulation. We can expect that *psbT* mRNA that is trapped in a *psbT*/*psbN* hybrid is not free for translation. Results of this chapter need to be evaluated in connection with results described in chapter 2. As the PSBT protein is involved in the biogenesis of PSII (Ohnishi *et al*., 2008) and in efficient repair of photo-damaged PSII reaction centers (Ohnishi

et al., 2001, 2007) we should expect an increase in the levels of PSBT protein (required for PSII biogenesis) and correspondingly in free translatable *psbT* mRNA during the greening of etiolated plants. On the other hand, the formation of *psbT*/*psbN* sense/anti-sense hybrids might serve photo-protection of the *psbT* mRNA to survive a high light stress under which single-stranded mRNA might be degraded. Thus, *psbT* mRNA might be immediately available after the stress to produce PSBT protein for repair supposed that strand separation is included in the repair mechanism. On the contrary, strong light that leads to photo-damage of PSII might also induce over-expression of PSBT protein by reducing the amount of anti-sense RNA in the plant and thus augmenting free single-stranded mRNA for translation. Results obtained from the light experiments (chapter 2) are interpreted corresponding to these two options.

A deep comprehension of the mechanism of action of the *psbT* anti-sense RNA in the control of *psbT* expression might contribute to develop new strategies for gene-expression control. For instance, the anti-sense mediated regulation could be exploited for a fine-tuning control of the expression of a specific plastid trans-gene even inside of an operon.

Results: Chapter 1

1. Expression analysis of the two plastid ATP operons: the large ATPI/H/F/A and the small ATPB/E operon.

The plastid *atp* genes are arranged in two *atp* operons on the chloroplast genome. The large *atp* operon consists of *atpI, atpH, atpF* and *atpA* genes (Fig. 7a-A) which code for IV, III, I and α subunits of the ATP synthase respectively. The small *atp* operon consists of *atpB* and *atpE* genes (Fig. 7a-B) which code for β and ε subunits respectively.
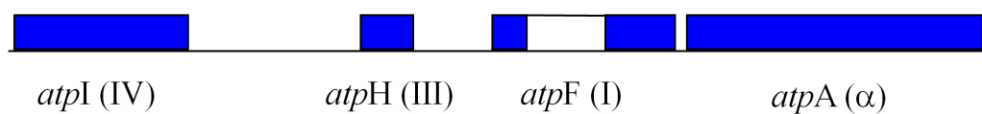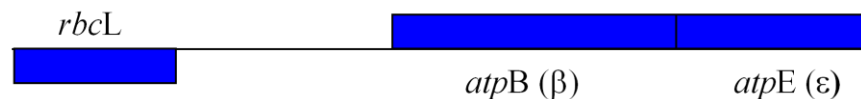
Figure7a. Schematic presentation of the arrangement of genes of the two *atp* operons.

Levels for all plastid-encoded atp mRNAs were more or less reduced in *sig3* plants, except for the *atp*I transcript (Zghidi et al., 2007). I first have participated to the analysis of the role of SIG3 in the transcriptional regulation of the *atp* genes in collaboration with Wafa Zghidi who completed the thesis in our laboratory in 2008 (results included in her thesis). We found that SIG3 controls the transcription from the -413 *atpH* promoter specifically and completely (Fig. 7b-A). It also partially controls transcription from the -467 *atpB* promoter (Fig.7b-B and in thesis of Wafa Zghidi, 2008). All the other promoters of *atp* operons (the -229 *atpI*, the -520 *atpB* and the -431 *atpE*) were SIG3 independent (Fig.7b and thesis of Wafa Zghidi 2008). These data indicated that, although SIG3 specifically regulates the transcription from the -413 *atpH* promoter and may play an important role in the higher *atpH* stoichiometry at both the transcript and protein level, it does not co-regulate the transcription of the two *atp* operons. A summary of the results obtained in collaboration with Wafa Zghidi is given below in Figure 7b.

A-Large operon

-229/225 (PEP-SIG3 independent)

-413 (SIG3-PEP)

-45   -30

rps2   atpI   atpH   atpF   atpA

B-Small operon

-520 (PEP- SIG3 independent)   -431 (PEP-SIG3 independent)

-467 (PEP-partially SIG3 dependent)

84

rbcL   atpB   atpE

Figure 7b. Summary of SIG3 dependent transcription initiation sites of the *atp* genes.
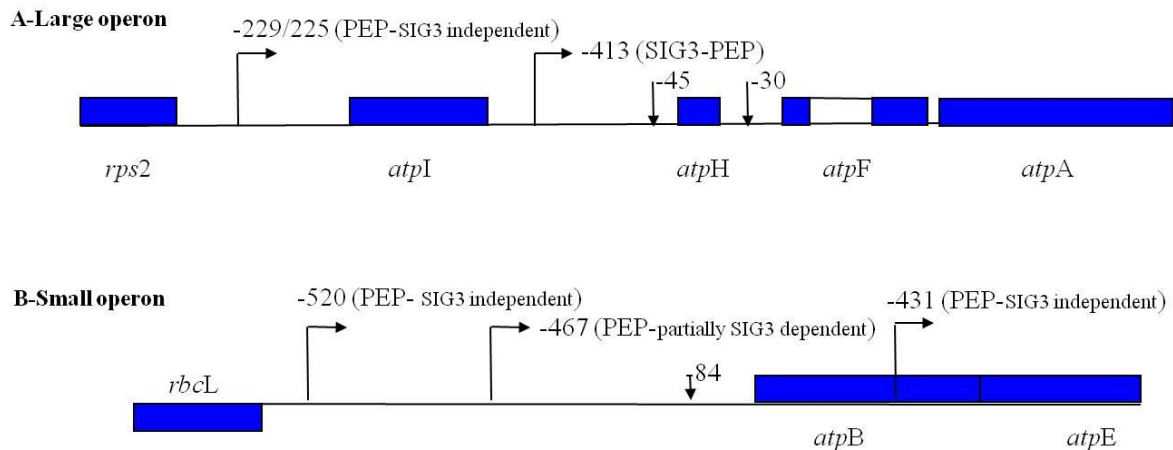(Results of the experiments carried out in collaboration with Wafa Zghidi).

Indicates the 5' end of the primary transcript while   indicates cleavage sites of the primary transcripts. These arrows are used for all primary transcripts and cleavage sites respectively in the following figures.

Keeping in view the above data/results, in the current chapter, I have included results that address the following questions:

1- Which sigma factor/factors controls the transcription from the SIG3 independent promoters of *atpI* (the -229), *atpB* (the -520 and -467), and *atpE* (the -431)?

2- What is the mechanism of expression of the *atpF* and *atpA* genes ?

3- Is transcription of the plastid *atp* genes, which are located in two distinct operons but code for subunits of the same complex, co-regulated by a single transcription factor?

4- What is the role of SIG3 in obtaining the higher stoichiometry of ATPH? Is the uneven stoichiometry of the *atpH* mRNA, which corresponds to the high level of ATPH protein, obtained at the transcriptional and/or post transcriptional level?

In order to address the above mentioned questions, I have analyzed, in collaboration with Dr. Livia Merendino, the expression of these two *atp* operons at RNA and protein level in detail using all the six sigma mutants that have been available in the laboratory. In addition, I have also carried out detailed analysis of extremities of transcripts of the *atp* genes. For better and easy understanding of the role of sigma factors in transcriptional regulation of the *atp* genes, the results are arranged according to the order of the genes present in the two *atp* operons starting from the first gene of the large *atp* operon (*atpI*).

Note:

1- The RNAs used for the primer extension and northern blot analyses were first quantified by spectrophotometer and then their quantitative and qualitative analysis was carried out by 1% agarose gel (data not shown for all the experiments).

2- The proteins used for western blot analysis were carefully quantified by Bradford assay or a method established in the laboratory taken from (Esen, 1978)(see materials and methods) and then again the quantitative and qualitative analysis was carried out by staining the SDS PAGE (for proteins of higher molecular weight) or by TRICINE-SDS-PAGE (for proteins of lower molecular weight) with Coomassie blue. The quantity of proteins among the wild type and mutants was also compared on the membrane after transfer of protein from the gel by staining them with "Ponceau Rouge". In brief extra care was taken for the quantitative and qualitative analysis of RNAs and proteins.

3- Some of the results were confirmed by complementary techniques.

4- Mapping of the 5' extremities of transcripts was carried out by running 5' labelled primer extension product in parallel with the sequence prepared with the same oligonucleotide used for primer extension on the polyacrylamide gel. For repetitions of primer extension analyses the gels were only run with molecular weight markers instead of sequence ladders because this was sufficient to identify the corresponding cDNAs. Mapping of the 5' extremities was also obtained by 5'RACE and by circular RT-PCR and could be compared to primer extension results to verify or correct the results obtained by primer extension.

1.1. Transcriptional analysis of *atpI* transcripts.

It was recently shown that the *atpI* gene has two PEP dependent promoters responsible for the transcription initiation from positions -229 and -225 from the ATG translation initiation codon of *atpI* (Swiatecka-Hagenbruch et al., 2007). We have analyzed the *atpI* promoter region in *Arabidopsis* at two stages of development (2 days and 6 days old plants) (data not shown for two days old plants) but in our culture conditions (see material and methods) we found only -229 initiated transcripts (Fig. 8C and 8E). Macroarray data showed that among all the plastid encoded ATP synthase genes, transcript accumulation of only the *atpI* gene remained unchanged in *sig3* plants (Zghidi et al., 2007). This was also confirmed by primer extension analysis that showed that the -229 *atpI* promoter is SIG3 independent (thesis Wafa Zghidi, 2008).

In order to find out which of the sigma factors is involved in the transcription initiation of *atpI* transcripts, we carried out primer extension analysis (Fig. 8C and 8D) by using wild type Columbia (Co) and Wassilewskija (Ws) plants and *sig1* (Co), *sig2* (Ws), *sig3* (Co), *sig4*

(Ws), *sig5* (Co) and *sig6* (Co) T-DNA insertion mutants. Primer extension is a technique used to map the 5' extremities of both the primary and the processed transcripts. A 5' end radioactively labeled (P^{32}) transcript specific primer is used for cDNAs synthesis from mRNAs. These radioactively labeled cDNAs are loaded and run on 6% polyacrylamide gel. For mapping the 5' extremity of the transcript under study, the sequence prepared with the same primer as used for primer extension was loaded and run on the same gel together with molecular weight markers. Only the molecular weight markers are shown in Fig. 8. We found that the -229 primary transcript is strongly reduced (almost absent) only in *sig2* plants (Fig. 8C, compare lanes 2' to 7' and Fig. 8E, lanes 8' and 11'). This indicates that SIGMA2 (SIG2) regulates *atpI* initiated transcription of the large *atp* operon. It also indicates that the large *atp* operon is co-regulated by SIG3 and SIG2. It would be interesting to analyze the contribution of this SIG2 dependent transcript of *atpI* to the transcript levels of the downstream genes of this operon (*atpH*, *atpF* and *atpA*). In addition, this finding also raises the question, whether SIG2 co-regulates both *atp* operons (discussed later).

If we carefully observe, it seems that the -229 initiated transcript is also slightly reduced in the *sig1*, *sig3* (Fig. 8C compare lane 3' and 5' to 2'), *sig4* (Fig. 8C compare lane 7' to 6') and *sig6* mutant (Fig. 8E, compare lane 11' to 10'). These sigma factors may play a minor role in the transcription from the -229 promoter. As the observed reduction was not always reproducible and it is much less important than that observed in *sig2* plants, the conclusion remains the same: The -229 is SIG2 dependent.
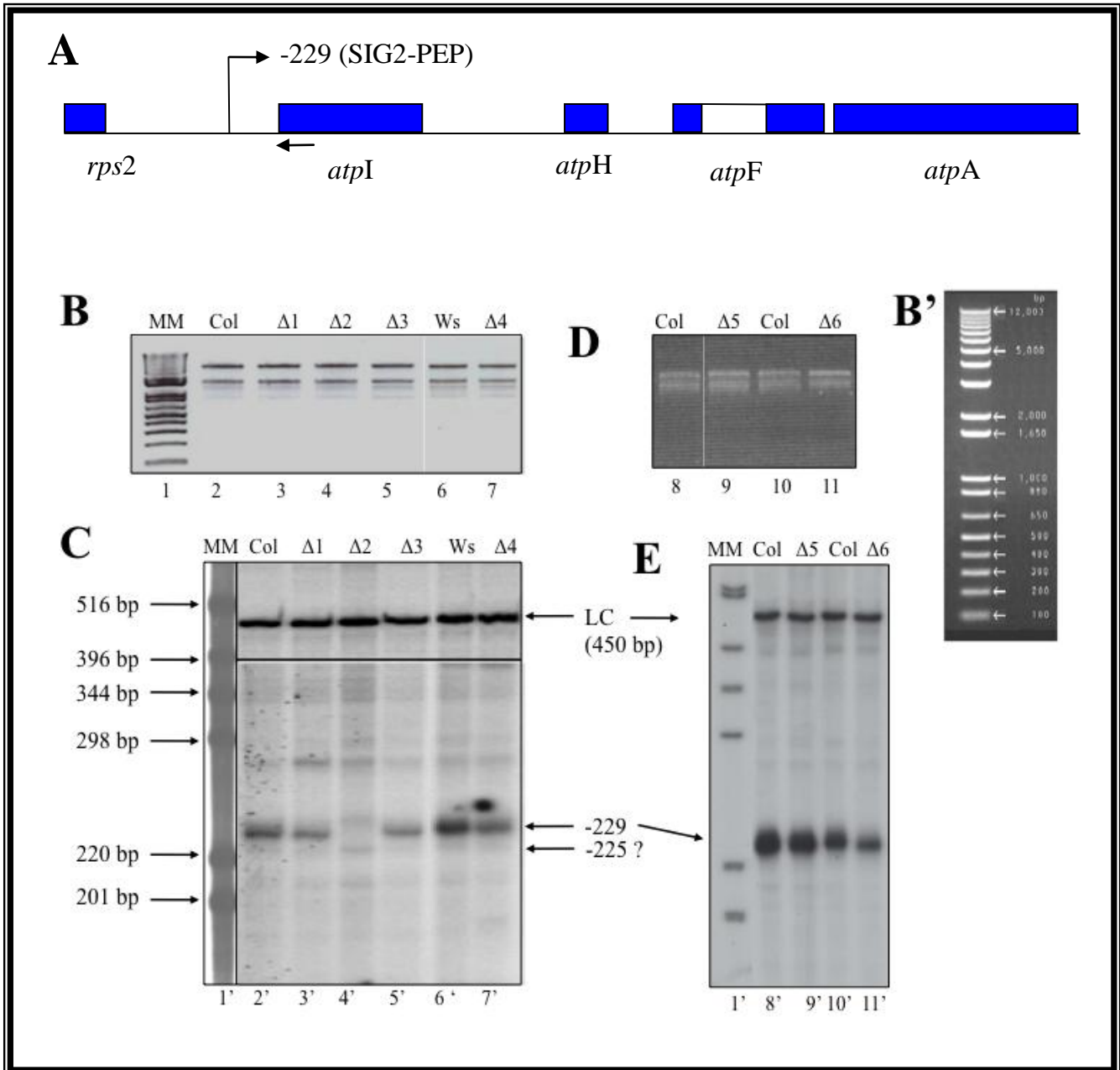
Figure 8. Characterization of the -229 *atp*I promoter.

(A) Schematic presentation of the *atpI/H/F/A* operon. ⌐➞ Indicates the 5' end of the primary transcript while ◄──── indicates the oligonucleotide used in primer extension analysis.

(B, D) Qualitative and quantitative analysis of total RNAs. RNAs were first quantified by the spectrophotometer and equal amounts were then analyzed on 1% agarose gels. Total RNA from the following Arabidopsis plants was analyzed: Wild type Columbia (lane 2, 8, 10) with *sig1* (lane 3), *sig3* (lane 5), *sig5* (lane 9), *sig6* (lane 11) and wild type Wassilewskija (lane 6) with *sig2* (lane 4) and *sig4* (lane 7). Lane 1 shows the 1Kb plus molecular weight marker detailed in (B'). The same RNAs were then used for primer extension analyses.

(C, E) Primer extension analysis of -229 *atp*I transcripts. 5 µg of total RNA extracted from 6 days old plants of wild type Columbia (lanes 2', 8' and 10') were analyzed with *sig1* (lane 3'), *sig3* (lane 5'), *sig5* (lane 9'), *sig6* (lane 11') and wild type Wassilewskija (lane 6') was analyzed with *sig2* (lane 4') and *sig4* (lane7'). Lane 1' indicates the molecular weight marker used for identification of the transcripts. LC indicates the loading control which corresponds to the PCR amplified product of 800 bp added to the mixture just before phenol chloroform extraction after cDNA synthesis.

In the absence of the -229 transcript (Fig. 8C, lane 4'), another band just below the -229 appears, at position -225. This indicates that a -225 initiated transcript does exist and might only be produced in conditions where the -229 transcript is either absent or not sufficiently produced.

The loading control in Figure 8C and E helps to determine a possible loss of labeled cDNAs in any of the samples during the procedure of purification, precipitation and loading.

1.2. Transcriptional analysis of *atpH* transcripts.

Two types of transcripts have been revealed for the *atpH* gene: one starting at position -413 and another one starting at position -45 from the ATG translation initiation codon (Fig. 9B lane 2') (Zghidi et al., 2007). Using the 5' RACE technique on TAP (Tobacco Acid Pyrophosphatase: which dephosphralises the 5' ends of the primary transcripts) treated and untreated transcrits, it was previously shown that the -413 is a primary transcript and -45 is a processed transcript (Thesis of Wafa Zghidi, 2008). Sequencing of 5' Race products of the -413 transcripts showed that the transcripts (data not shown) indeed start 5 nucleotides upstream of the -413 position that was determined by primer extension experiments (Zghidi et al., 2007). So further in the text they will be called -418 *atpH* transcripts. In the transcript analysis of *sig3* mutant plants by macroarray analysis, *atpH* was the most affected *atp* gene (Zghidi et al., 2007). Primer extension analysis showed that the -418 *atpH* transcripts are SIG3 dependent and the processed transcripts (-45) are at least partially SIG3 dependent (Zghidi et al., 2007) (Fig. 9B, compare lane 5' to 2'). This means that the -45 transcripts are produced as a result of processing of both the *atpI* (-229) and the *atpH* (-418) initiated transcripts, the -229 being SIG2 dependent.

In order to explore the role of the other sigma factors on *atpH* transcription and the contribution of the SIG2 dependent *atpI* -229 transcript to the -45 *atpH* transcript, we carried out primer extension on the total RNAs extracted from 6 days old wild type plants (Columbia and Wassilewskija) and *sig1* (Co), *sig2* (Ws), *sig3* (Co), *sig4* (Ws), *sig5* (Co) and *sig6* (Co) mutants (Fig. 9B and C lanes 2' to 7' and 8 to 11). The localization of the oligonucleotide used for primer extension is shown in Fig. 10B as horizontal arrow. No clear change in the accumulation of -418 and -45 transcripts was observed in *sig2* plants (Fig. 9B, compare lanes 4' and 6'). This result can be explained by functional compensation of SIG2 by SIG3. SIG3 might compensate the absence of SIG2, which regulates the transcription of the -229 *atpI* transcript, by over transcribing the large *atp* operon from the -418 *atpH* promoter, the -418

transcript being then very rapidly processed at the -45 site. Over-expression of SIG3 in *sig2* plants had been observed previously (Privat et al., 2003). Both transcripts, the -418 and the -45 transcripts, remain unchanged in *sig1*, *sig4*, *sig5* and *sig6* plants. This shows that SIGMA1 (SIG1), SIGMA4 (SIG4), SIGMA5 (SIG5) and SIGMA6 (SIG6) do not play any role in the transcription of neither the -418 transcript nor the -45 transcript.

In order to know whether the *atpH* transcripts exist as monocistronic or polycistronic transcripts and to determine 3' extremities of the -418 and the -45 transcripts, we used the circular RT-PCR technique. In this technique transcripts are treated with TAP (Tobacco Acid Pyrophosphatase), an enzyme that dephosphoralizes the primary transcripts and makes them available for auto ligation. Thereafter, mRNAs are reverse transcribed to prepare cDNAs by using a transcript specific primer. To differentiate between the -418 and the -45 mRNAs, we designed two different primers for reverse transcription (R1 and R2 respectively, schematically demonstrated in Fig. 10A).

cDNAs were then PCR amplified by using transcript specific primers R1+F or R2+F. Cloning and sequencing of the cDNAs gave the results presented in paragraphs 1.2.1 and 1.2.2.
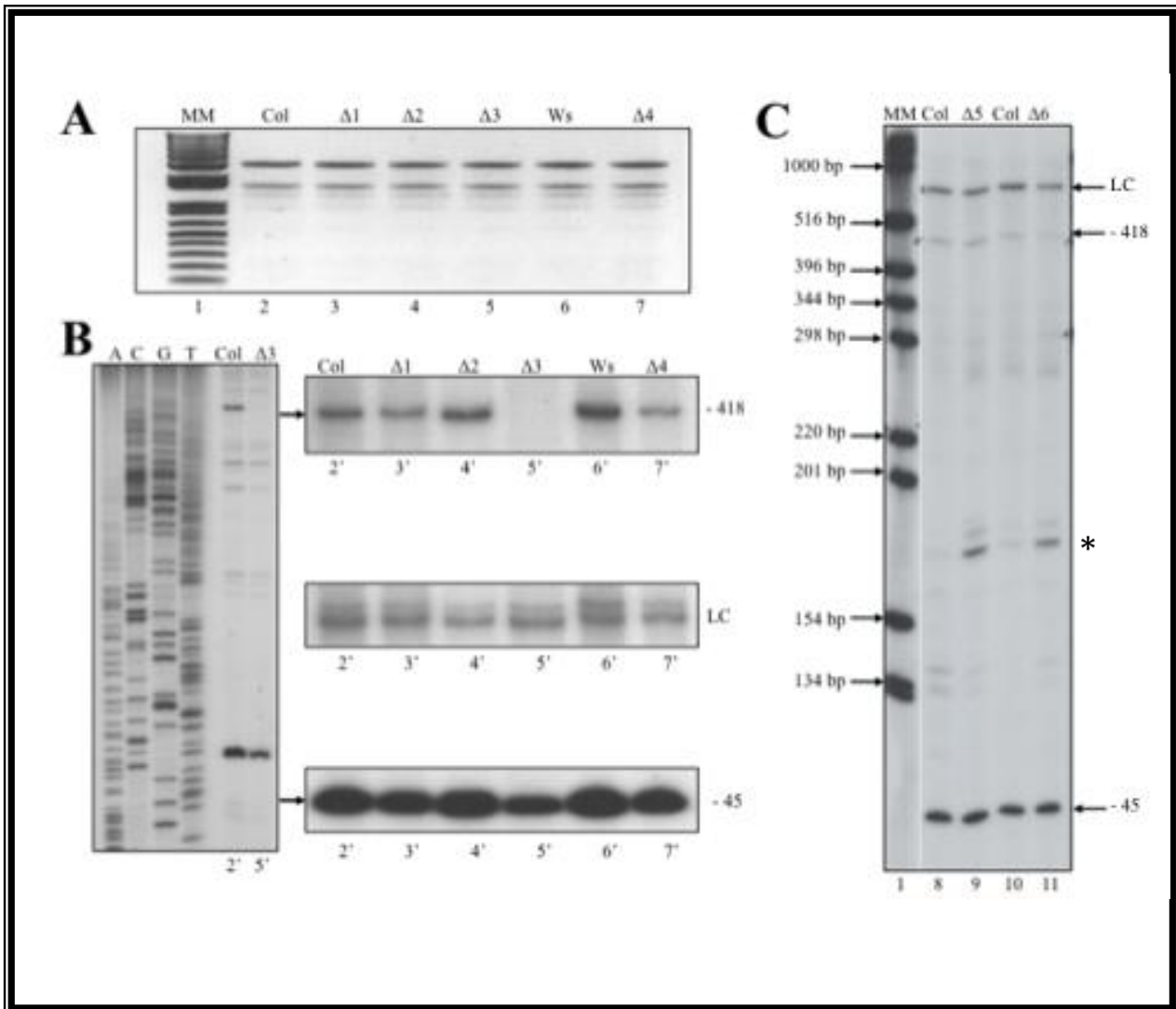
Figure 9. Characterization of the -418 and -45 *atpH* mRNAs.
(A) Qualitative and quantitative analysis of total RNAs of the wild type Columbia (lane 2), *sig1* (lane 3), *sig2* (lane 4), *sig3* (lane 5), and wild type Wassilewskija (lane 6) and *sig4* (lane 7), on the 1% agarose gel. Lane 1 shows the 1Kb plus molecular weight marker whose map is shown in (Fig. 8 B'). RNAs were first quantified by the spectrophotometer and then analyzed on agarose gel. These RNAs were used for the primer extension analysis shown in (B) while the RNAs used for primer extension presented in (C) are shown in Fig. 8D.
(B, C) Primer extension analysis of *atpH* transcripts. 5 μg of total RNA extracted from 6 days old plants of wild type Columbia (lanes 2', 8 and 10), *sig1* (lane 3'), *sig2* (lane 4'), *sig3* (lane 5'), *sig5* (lane 9), *sig6* (lane 11) and wild type Wassilewskija (lane 6') and *sig4* (lane 7') plants have been used for analysis. Lane 1' indicates the molecular weight marker used for mapping and identification of the transcripts. LC indicates the loading control which corresponds to the PCR amplified product of 800 bp added to the mixture just before phenol chloroform extraction after cDNA synthesis. ∗, processing intermediates or degradation products.
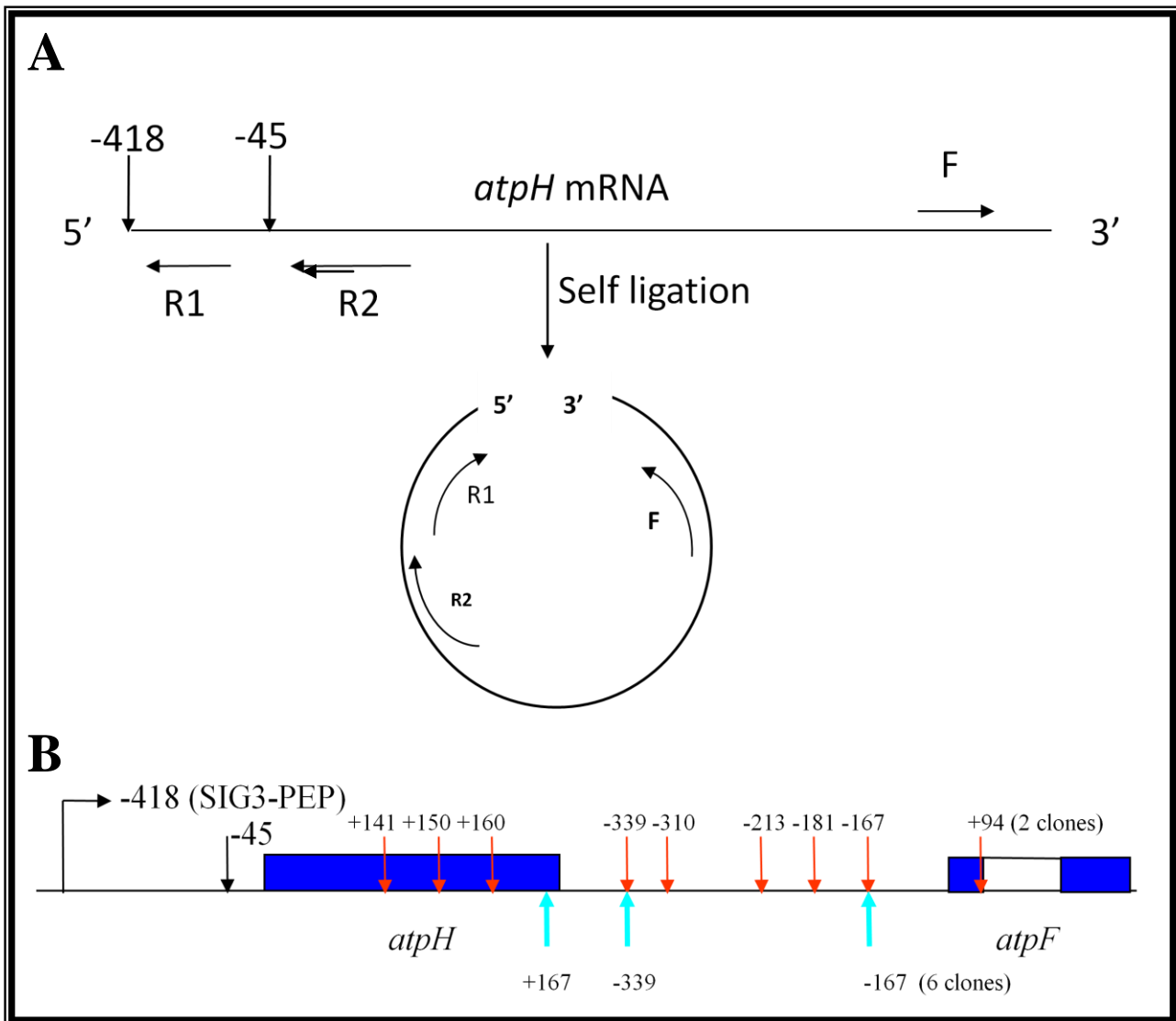
Figure 10. Mapping of *atp*H 3' ends by circular RT PCR.

(A) Schematic presentation of the circular RT PCR technique on *atpH* transcripts.
(B) Schematic presentation of the *atpH/atpF* genes and location of the 5' and 3' ends of *atp*H transcripts.
The 5' end of the -418 *atpH* primary transcript and the -45 cleavage site of the *atpH* primary transcripts are reported.

Indicates the distribution of 3' ends of -418 *atpH* transcripts while ↑ indicates the distribution of 3' ends of -45 *atpH* transcripts. ⟵ indicates the oligonucleotide used in primer extension analysis.

1.2.1. 3' end mapping of -418 initiated *atpH* transcripts.

For selection of -418 initiated transcripts, RT-PCR was performed with primer R1 that is located upstream of the -45 cleavage site (Fig. 10A) and thus excludes all -45 transcripts from the analysis. All clones 5' ended at position -418 from the ATG translation initiation codon of *atpH*. On the contrary, the 3' ends of the clones vary considerably (listed below and shown in Fig. 10B as red arrows). They are found within the coding sequence of *atpH,* in the

intergenic region of *atpH/atpF* and in the coding region of *atpF*. This indicates that the -418 transcripts, despite of having a stable 5' end, posses a highly unstable 3' end.

1 clone	+141 from *atpH* ATG translation initiation start codon
1 clone	+150 from *atpH* ATG translation initiation start codon
1 clone	+160 from *atpH* ATG translation initiation start codon
1 clone	-167 from *atpF* ATG translation initiation start codon
1 Clone	- 181 from *atpF* ATG translation initiation start codon
1 clone	-213 from *atpF* ATG translation initiation start codon
1 clone	-310 from *atpF* ATG translation initiation start codon
1 clone	-339 from *atpF* ATG translation initiation start codon
2 clones	+94 from *atpF* ATG translation initiation start codon

1.2.2. 3' end mapping of -45 processed *atpH* transcripts.

For study of the -45 *atpH* transcripts, RT-PCR was performed with primer R2 (see Fig. 10A) that is located downstream to the -45 cleavage site. All clones had the same 5' end at position -45 from the ATG translation initiation codon of *atpH*. This indicates that the -45 transcript has a very stable 5' end. The majority of the clones (six out of eight) had the same 3' end in the intergenic region between *atpH* and *atpF* genes (listed below and shown in Fig. 10B as light blue arrows).

6 clones   -167 from atpF ATG translation initiation start codon

1 clone   -339 from *atpF* ATG translation initiation start codon

1 clone   +167 from ATG translation initiation start codon of *atpH*.

The above mentioned results indicate that the *atpH* transcripts do exist as monocistronic transcripts and that the 3' end of the -418 initiated transcripts are unstable and dispersed while 3' end of -45 transcripts are highly stable. We can assume that -418 initiated transcripts are stabilized after cleavage (processing) at the -45 position. This stabilization is probably brought about by binding of a so-called pentatricopeptide repeat (PPR) protein, PPR10, to the region spanning the -45 position up to the ATG translation start codon as recently shown in maize (Pfalz et al., 2009). In this hypothesis the higher protein stoichiometry of the ATPH in our culture conditions (see material and methods) might be controlled on the transcriptional level by SIG3, and in addition on the post-transcriptional level by a PPR10 like protein.

1.2.3. Regulation of the higher stoichiometry of ATPH.

The next question regarding the higher stoichiometry of the ATPH protein is to which extent the higher stoichiometry of the ATPH protein is determined on the transcript level. In order to answer this question, we carried out a macroarray analysis of the total RNAs extracted from the 6 days old plants. For macroarray analysis, mRNAs have been reverse transcribed in the presence of a radiolabelled deoxynucleotide by using specific primers to each plastid gene chosen as close as possible to the corresponding probes deposited on the macroarray. This trait in addition with the selection of 60-mer probes near the ATG start sites contribute to the semi-quantitative property of the macroarray developed in our laboratory (Patent FR 0607168). Hybridization signals on the macroarray were detected by phosphor imager analysis and were quantified using the Array gauge (Fuji) software. We found that the transcript accumulation of *atpH* mRNA is 16-30 times higher than that of other *atp* genes (Fig. 11). The value of *atpI* was considered to be one for the calculation of the relative amounts of transcript accumulation of other *atp* genes. The data presented in figure 11 is the average of three independent experiments.

The data presented in figure 11 show that a higher stoichiometry for ATPH does exist at the transcript level. Run-on experiments should be carried out to determine whether this higher *atp*H mRNA level is achieved at the transcriptional level by SIG3-PEP specific transcription of the a*tp*H gene or at the post transcription level by stabilization of the processed -45 *atp*H mRNA. But it can be inferred by the data presented in section 1.2a and 1.2b of this chapter that the higher stoichiometric accumulation of *atpH* mRNA observed by macroarray analysis is achieved by specific stabilization of the -45 monocistronic transcript which is produced as a result of cleavage of both the SIG2 (-229 *atpI*) and the SIG3 (-418 *atpH*) dependent transcripts. So, the higher stoichiometric accumulation of *atpH* mRNA is not solely dependent on SIG3.
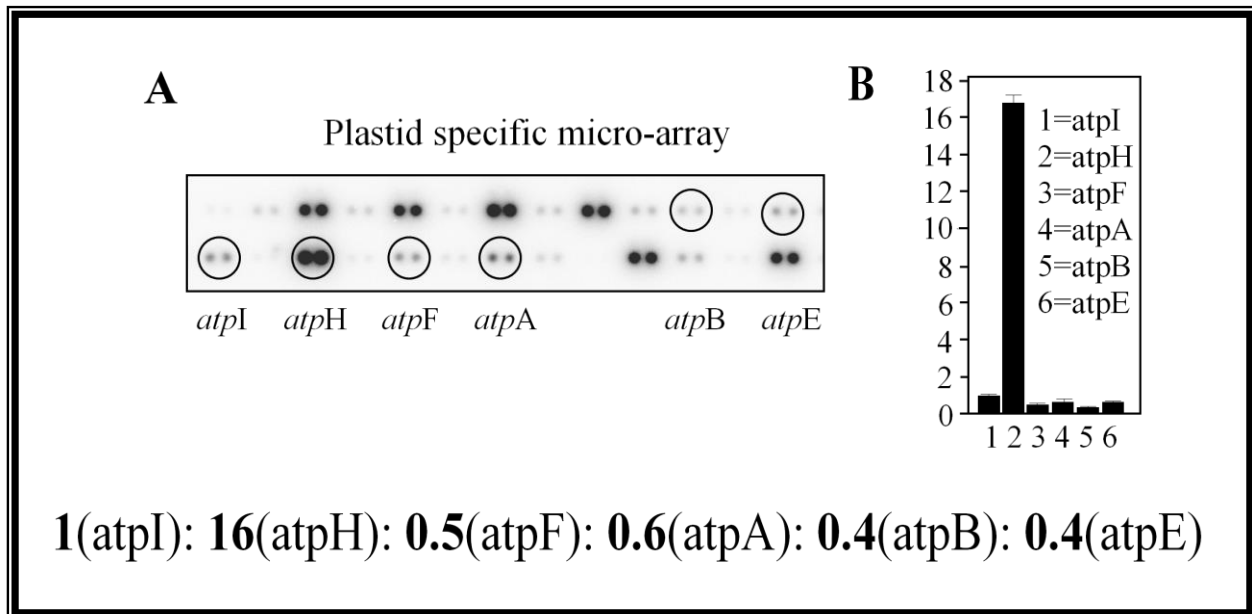
Figure 11. Stoichiometry of plastid encoded ATP synthase mRNAs.
A) A part of the plastid macroarray showing the transcript accumulation of *atp* genes. B) The relative quantities of transcript accumulation, as determined by macroarray analysis of plastid genes encoding the subunits of ATP synthase in 6 days old green Wt plants are represented by columns. The quantity of accumulated transcripts of *atpI* was taken as one. The RNAs were qualitatively and quantitatively analyzed before being used in macroarray analysis (data not shown).

If a higher stoichiometry is present at the transcript level, which is evident from the results presented in the figure 11, then the next question is which sigma factor (SIG2 or SIG3) plays a more important role in the regulation of this higher stoichiometry at the transcript level.

In order to answer this question, we carried out Northern blot analysis of wild type, *sig2* and *sig3* plants. This Northern blot analysis shows that in *sig3* mutants the -418 transcripts are highly reduced while the -45 transcripts are only slightly reduced (Fig. 12B). In contrast, in *sig2* plants both the -418 and the -45 transcripts are accumulated in relation to wild type plants (Figure 12A). These results are coherent with the results of primer extension analysis discussed in the section 1.2 of this chapter. Primer extension analysis showed no changes in -418 and -45 transcripts in the absence of SIG2 and this result was interpreted by transcriptional compensation at the -418 *atpH* promoter by SIG3-PEP. It shows that SIG3 plays a more important role in the maintenance of the transcript stoichiometry in the absence of SIG2. This raises the question whether SIG3 also plays a similar important role in the conditions of stress in maintaining higher transcript level (discussed in chapter 2).
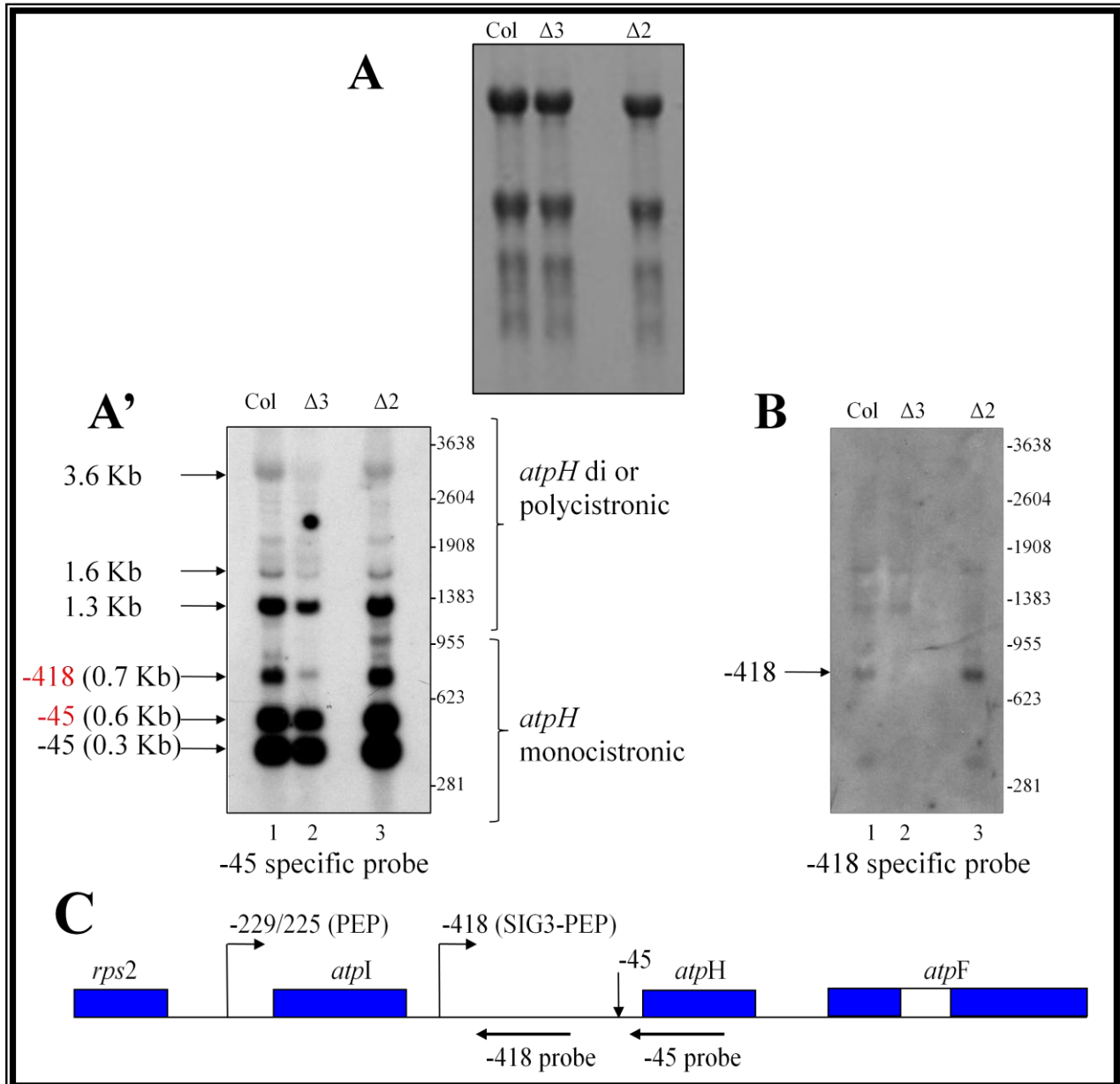
Figure 12. Analysis of expression of *atpH* precursor RNAs by Northern Blot analysis.

(A) 5 μg of total RNAs extracted from 6 days old plants of wild type Columbia (lane 1), *sig2* (lane 3), *sig3* (lane 2) were separated on denaturing agarose gel, transferred onto a nitrocellulose membrane and the membrane was colored with methylene blue.

The membrane was hybridized either to the -418 specific *atpH* probe (A') or to the -45 transcript specific *atpH* probe (B). The molecular weight marker, visible on the nitrocellulose membrane under UV because of Ethidium Bromide, was marked on the membrane with the lead pencil.

Note: Quantitative and qualitative analysis of RNAs used in this experiment are also shown in figure 9A.

(C) Localisation of the -45 and -418 specific probes used for northen blots.

Supposed that the transcript level of *atpH* mRNA determines the protein level then we would expect a decrease in ATPH protein in *sig3* mutants and an increase in *sig2* mutants. In order to verify this hypothesis and to determine whether the compensatory role of SIG3 at the transcript level is reflected on the protein level, we carried out a western blot analysis of proteins extracted from the 6 days old wild type plants (Columbia and Wassilewskija) and *sig2* (Ws) and *sig3* (Co) mutants. This experiment was done in collaboration with Frank Buhr.

Unexpectedly, western blot analysis shows that ATPH protein levels remain constant in *sig3* and are reduced in *sig2* plants (Figure 13). It is difficult to conclude on the reduction of ATPH in *sig2* mutants. In *sig2* mutant a general reduction of plastid protein accumulation should be observed because of reduction in tRNAs (Kanamaru et al., 2001). Keeping in mind the constant level of *atpH* transcripts in *sig2* plants, it can be inferred that this reduction is due to absence or reduction of tRNAs necessary for translation.
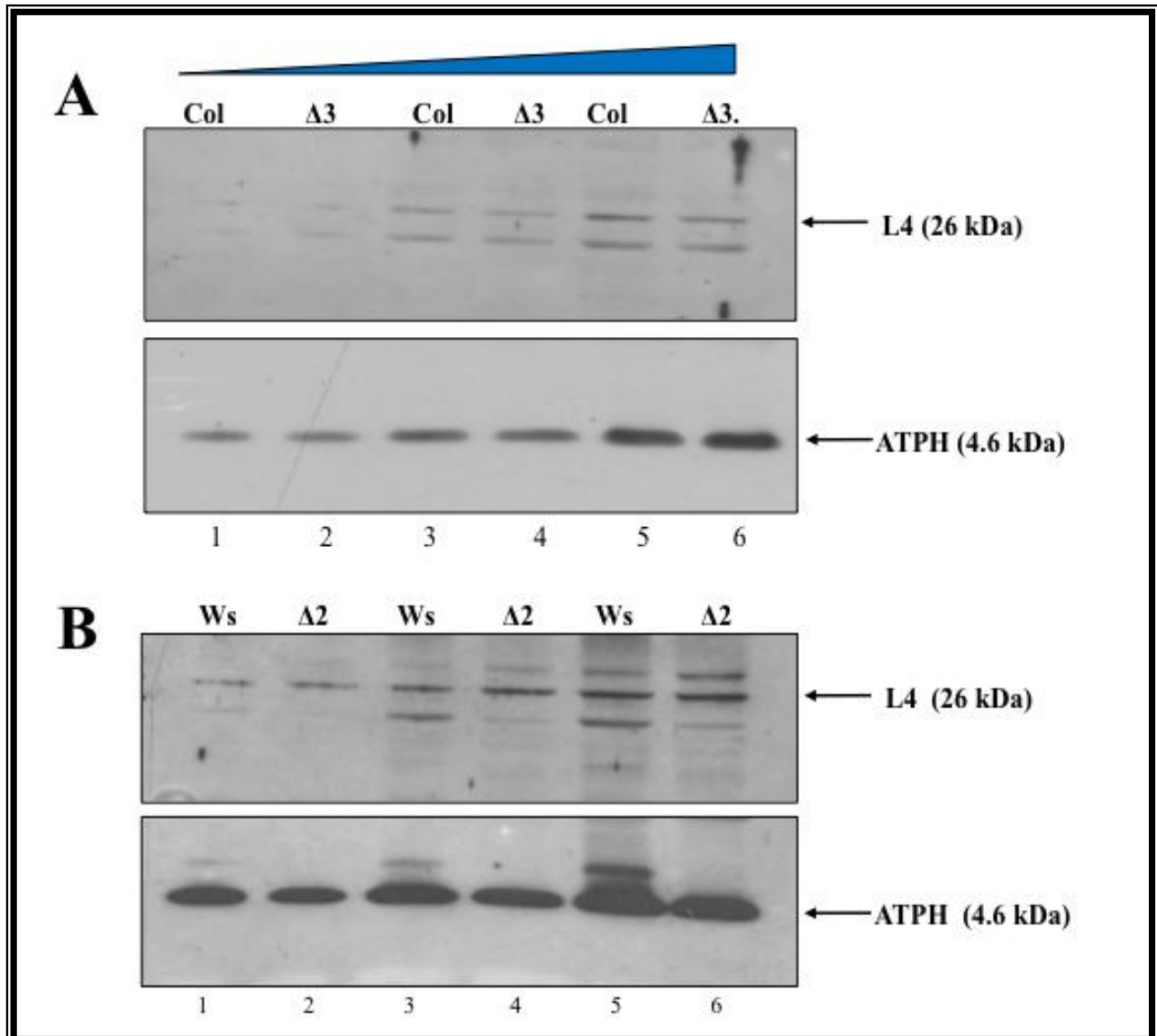
Figure 13. Analysis of ATPH protein levels in *SIG2* and *SIG3* knock out plants by Western blot analysis.
5µg (lanes 1 and 2), 10 µg (lanes 3 and 4) and 20 µg (lanes 5 and 6) of total proteins extracted from the 6 days old plants of wild type and *sig3* plants (A) and wild type and *sig2* plants (B) were separated on Tricine-SDS-PAGE (see materials and methods) and then transferred to nitrocellulose membrane. The antibody reactions were done with antibodies dressed against ATPH and RPL4 (a nuclear encoded plastid ribosomal protein used as a control).

From the result with the *sig3* plants we must conclude that SIG3-PEP dependent *atp*H transcription, although monocistronic SIG3 dependent *atp*H mRNA is detectable in Northern blots, does not contribute significantly to the ATPH protein level, at least not under the growth conditions that we have used. However, another Northern analysis that was done under similar conditions shows much less SIG3 dependent 0,7 kb RNA as detectable in Fig. 12 and the Western blot (Fig. 13) has been done with plant material that had not been analyzed in parallel at the RNA level. Because of the variability of the amount of the SIG3-dependent monocistronic *atp*H RNA in different preparations, it is difficult to conclude something concerning the protein level. RNAs and proteins need to be analyzed from the same plant material grown under conditions that activate SIG3. During my work I could not find the condition(s) where SIG3-dependent transcription of the *atp*H gene becomes essential to establish the high *atp*H mRNA stoichiometry.

In conclusion, higher stoichiometric accumulation of ATPH is also present at the transcript level. Both the SIG3 and SIG2 play partial and compensatory roles in transcription but the conditions that determine their mutual interactions are not yet elucidated, and a master control of the *atp*H monocistronic mRNA level exists probably on the post-transcriptional level by a PPR10 like protein. Last but not least, the activation of the SIG3 dependent promoter in the absence of SIG2 dependent transcription indicates a "safeguarding" role of SIG3 in the transcription of *atp* genes and we wonder if SIG3 does not become specifically important under some well defined stress conditions that unfortunately I could not find during my thesis.

1.3. Transcriptional analysis of *atpF* transcripts

Primer extension analysis shows that there are three *atpF* transcripts (Fig.14B). The two long transcripts by their size and their dependence on SIG3 correspond to the -418 and the -45 *atpH* transcripts (Fig. 14C, lanes 2 and 3). By comparison with the corresponding sequence, which was run on the same gel (Fig. 14 B), the 5' end of the third short transcript was mapped at position -30 from the ATG translation initiation codon of the *atpF* gene. In order to determine whether the -30 transcript is a primary or processed transcript, we carried out 5' RACE RT–PCR. In this techniques mRNA transcripts are treated with TAP (Tobacco Acid Pyrophosphatase), an enzyme that dephosphorylises the primary transcripts and makes them available for ligation with an adapter. Thereafter, mRNAs are reverse transcribed to prepare cDNAs by using a transcript specific primer. The produced cDNAs are then amplified by PCR by using a 5' adapter specific primer and a 3' transcript specific primer (external

primer). A second PCR, which renders more specificity for the cDNA amplification, is performed by using internal primers which are adapter specific and transcript specific. The obtained PCR products are then used for cloning and sequencing. However, we failed to amplify a band of the expected size for the -30 *atpF* transcripts.
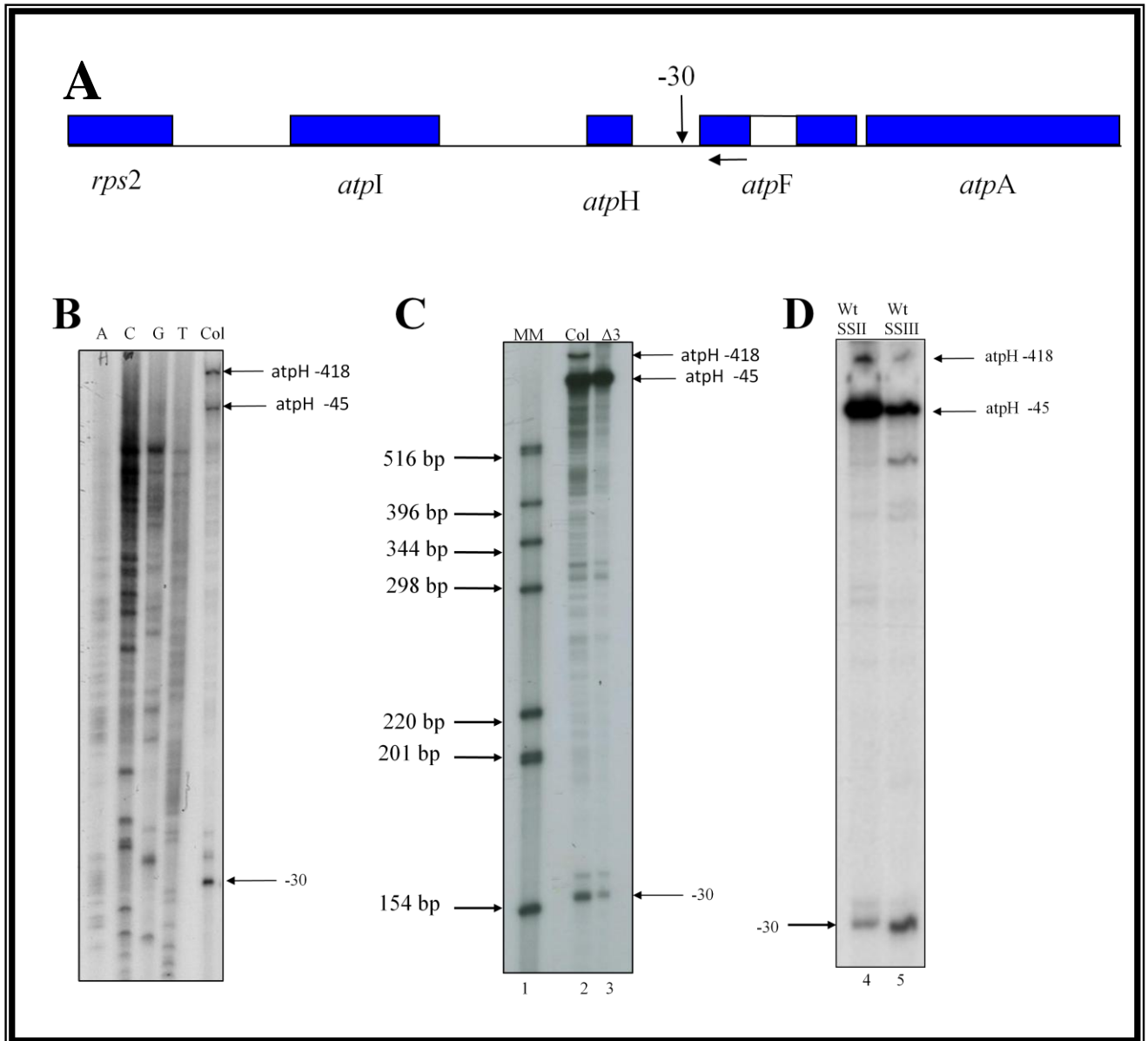
Figure 14. Expression analyses of the *atpF* gene.
(A) Schematic presentation of the *atpI/H/F/A* operon with respect to *atpF* transcripts. ↓ indicates the -30 cleavage site. ← indicates the oligonucleotide used for primer extension.
(B) Mapping of the small *atpF* precursor RNA by primer extension accompanied by the sequence ladder made with the same primer as used for primer extension.
(C) Primer extension analysis of *atpF* transcripts in WT and *sig3* plants. 5 µg of total RNA extracted from 6 days old wild type Columbia (lane 2) and *sig3* plants (lane 3) were used for analysis. Lane 1 shows the 1Kb plus molecular weight marker whose map is shown in (Fig. 8 B'). The experiment was reproduced several times giving identical results.
(D) Primer Extension analysis of *atpF* transcripts using total RNAs of 6 days old wild type plants and SSII (lane 4) and SSIII (lane 5) reverse transcriptases at 42 °C (lane 4) and 55 °C (lane 5) respectively.

In order to know whether in the primer extension study, the signal corresponding to -30 was due to an artifactual stop of the retro transcription caused by a secondary structure present, we carried out the primer extension at 42°C and 55°C by using SSII and SSIII reverse transcriptases, respectively (Fig. 14D). Higher temperature (55°C) helps to reduce secondary structures. If the band was an artifact due to secondary structures then there should be absence or strong reduction of signal at 55°C. Conversely, we found that the signal is even stronger (Fig. 14D, lanes 4 and 5). Hence it is clear from this experiment that -30 *atpF* transcripts do exist. It is known that retro transcription is less efficient for longer transcripts at 55°C and SSIII preferentially retro transcribes smaller transcripts. Therefore it is not surprising that we observe a reduction in longer cDNAs (-45 and -418 *atpH*) at 55°C (Fig. 14D, lane 5) and an increase in short cDNAs (-30 *atpF*).
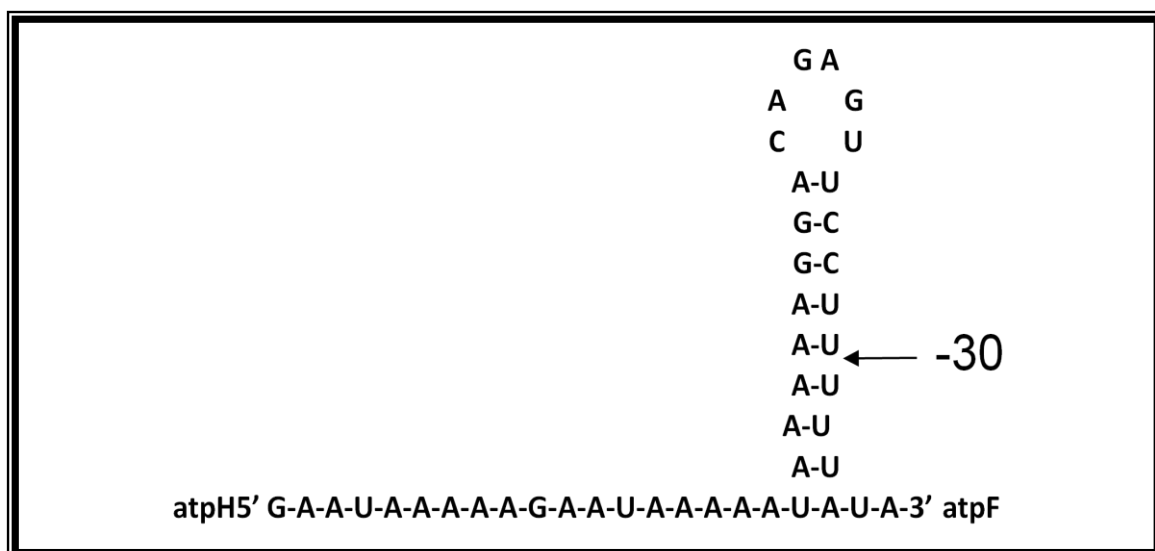
Figure 15. Hypothetical hairpin structure in the *atpH/atpF* intergenic region and position of the 5' end of the -30 *atpF* transcript.

Analysis of the hypothetical secondary structures of *atpH* and *atpF* intergenic region by RNA fold (http://bioweb2.pasteur.fr) showed that the -30 site lies in the stem of a hypothetical hairpin structure (Fig. 15). This type of hairpin is a typical substrate for endonucleases. This observation suggests that the -30 *atpF* transcript originates from a processing event.

Reduction of the -30 *atpF* transcript accumulation in *sig3* mutant plants when compared to wild type plants in primer extension analysis indicates that this transcript originates

preferentially from cleavage of the SIG3 initiated -418 *atpH* transcript. In addition, as there is only reduction but not complete absence of the -30 transcript in *sig3* plants, we conclude that this transcript is produced as a result of the processing of both the polycistronic transcripts which are SIG2 dependent (-229 *atpI* initiated) and SIG3 dependent (-418 *atpH* initiated).

1.4. Transcriptional analysis of *atpA* transcripts.

We carried out primer extension analysis of *atpA* transcripts. We observed very long transcripts that are difficult to measure on 6 % acryl amide gels. We suppose that these are co-transcripts of *atpA* with *atpF*, *atpH* and / or *atpI*. Along with these long transcripts, we also observed one small transcript (Fig. 16B, lanes 1' and 2'). 5' RACE showed that this transcript starts at position -74 from ATG translation initiation codon of *atpA*.

5' ends of 11 out of 19 clones ended at this position (Fig. 16A) that corresponds to the first nucleotide of the last codon of *atpF*.

5' RACE with and without TAP treatment showed that this transcript is a processed transcript (Fig. 16B, lanes 4 and 5). No primary transcripts for *atpA* could be detected by 5' RACE. As reduction in accumulation of the -74 *atpA* transcripts was observed in *sig3* plants (Fig. 16B, lane 2), we propose that this transcript is produced as a result of cleavage from both the -418 *atpH* initiated polycistronic mRNA and the -229 *atpI* initiated polycistronic mRNA.
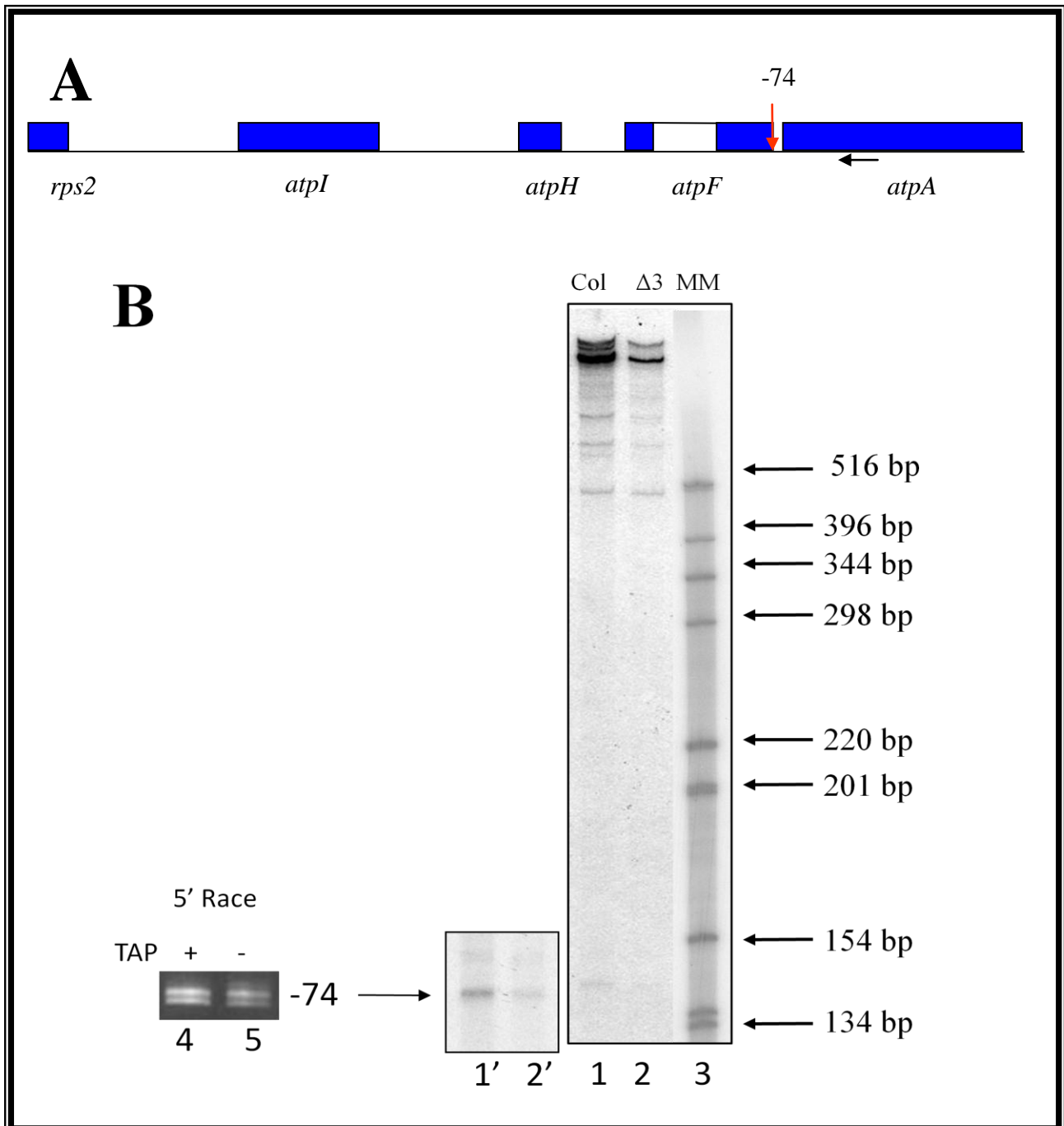
Figure 16. **Transcript characterization of the *atpA* gene.**
 (A) Schematic presentation of the *atpI/H/F/A* operon with respect to *atpA* transcripts. ↓ indicates the cleavage site at-74. ⟵ indicates the oligonucleotide used for primer extension.
(B) Primer extension analysis of the *atpA* transcripts. 5 µg of total RNAs extracted from wild type Columbia (lane 1) and *sig3* (lane 2) plants were used for analysis. Wt RNA was also analyzed by 5' RACE. TAP + and TAP - (lane 3, 4) correspond to the RT-PCR products obtained from RNAs treated with tobacco acid pyrophosphatase or without treatment. Lane 3 shows the 1Kb plus molecular weight marker. Lanes 1' and 2' correspond to longer exposure of lanes 1 and 2.

1.5. Transcriptional analysis of *atpB* transcripts.

In arabidopsis, a Northern blot analysis of the *atpB* transcripts showed that there are two types (2.6 kb, 2.0 kb) of dicistronic transcripts (Schweer et al., 2006). Later on two transcripts had been 5' mapped at -520 and -318 from the *atpB* ATG and found to be primary transcripts (Swiateka-Hagenbruch et al., 2007). The -520 transcripts were shown to be dependent on PEP and the -318 transcripts to be dependent on NEP. Our own analysis by primer extension and 5' RACE shows three different *atpB* transcripts with their 5' ending at positions -520, -467 and -84 (see Fig. 17A, B and thesis of Wafa Zghidi, 2008) where the -84 transcript results from processing. We conclude that the 2.6 kb and 2 kb *atpB* transcripts observed in Northern experiments correspond to the -520/-467 and -84 transcripts that were detected by primer extension respectively. The NEP dependent -318 transcripts that were identified by 5' RACE by Swiateka and colleagues (2007) in non treated as well as spectinomycin treated plant materials, were, in our conditions, only observed in plants that were treated with spectinomycin (a PEP inhibitor) (thesis Wafa Zghidi, 2008).

In order to identify which sigma factor was involved in the recognition of the two *atpB* promoters, we carried out primer extension on the total RNAs of 6 days old wild type plants (Columbia and Wassilewskija) and *sig1* (Co), *sig2* (Ws) *sig3* (Co) *sig4* (Ws), *sig5* (Co) and *sig6* (Co) plants (Fig. 17C). Our analysis shows that the -467 transcripts are predominantly made by SIG2 and the -520 transcripts are predominantly made by SIG6 as they are strongly reduced in *sig2* and *sig6* plants, respectively (Fig. 17, compare lanes 4 and 6 and lanes 11 and 12). On the other hand, the -84 transcripts are not reduced in *sig2* and *sig6* but surprisingly, these transcripts are reduced in *sig1* and *sig3* plants (compare lane 2 to lanes 3 and 4). How can this be explained?

Northern blot analysis performed with *sig6* plants showed that the 2.6 kb *atpB* transcript is SIG6 dependent in 5 days old plants and that in the absence of this transcript an upstream 4.8 kb transcript (whose 5' end was proposed to map in the intergenic region of *accD* and *rbcL*) is activated. This 4.8 kb transcript was proposed to be processed to the 2.0 kb (-84) transcript which was found not to be reduced in the *sig6* plants (Schweer et al., 2006). Our primer extension analysis of *sig2* and *sig6* plants shows that the -84 processed transcripts do not change considerably in the *sig6* plants but they change in *sig1* and *sig3* plants (Fig. 17C). Thus it might be that the 4.8 transcript is under control of SIG1 and SIG3. However, we never have observed the 4.8 kb transcript described by this group. In our case this RNA might be more rapidly processed to give a -84 *atp*B transcript. The -84 transcript pool should then

results from at three different promoters and the observed variations in the amount of the -84 transcripts should be due to the interplay of transcription at these 3 different promoters and thus difficult to predict.
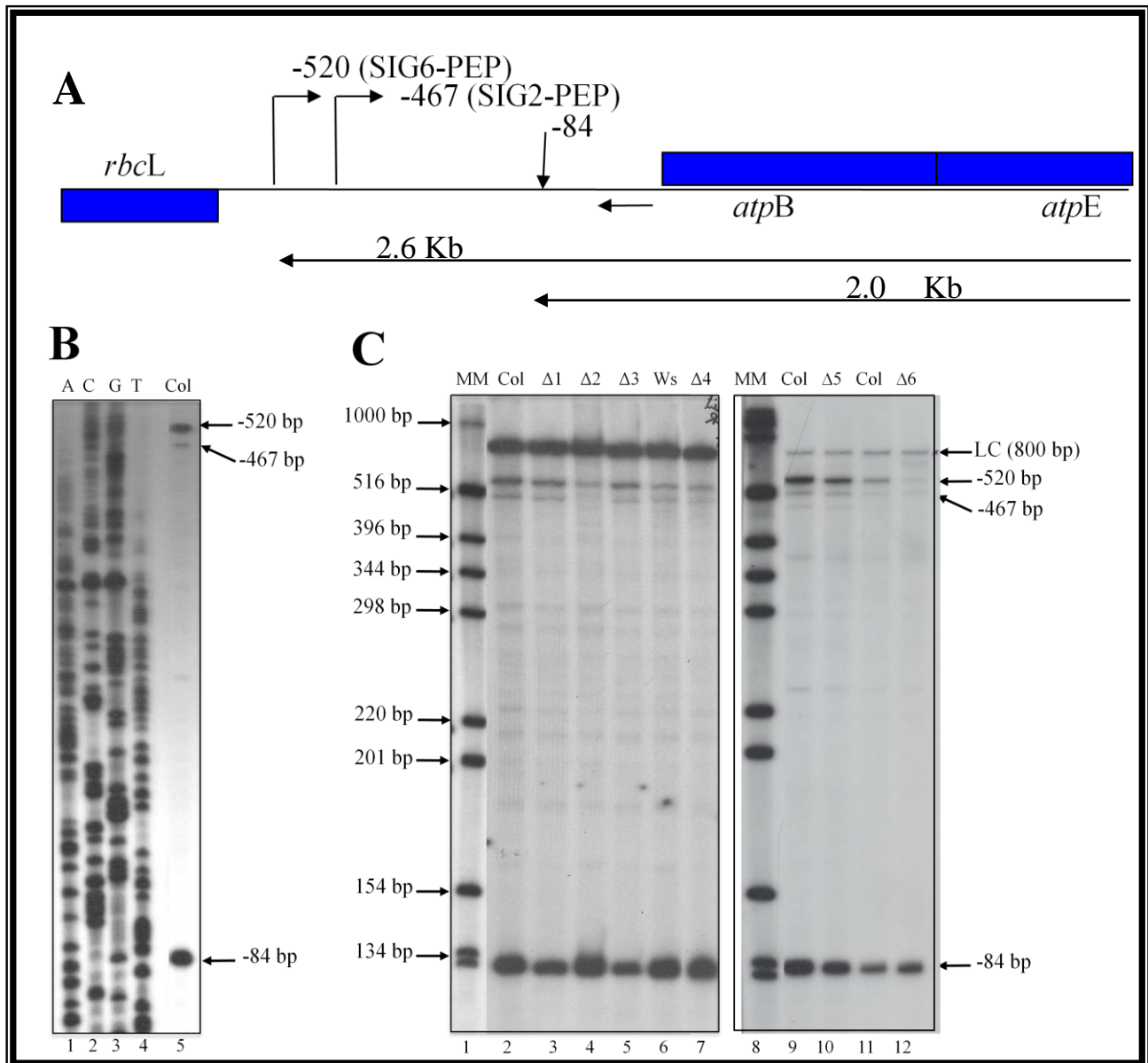
Figure 17. Transcript characterization of the *atp*B gene.

(A) Schematic presentation of the transcripts of *atpB/E*. ⌐ indicates the 5' end of primary transcripts while ↓ indicates the site of cleavage of the primary transcripts. ← indicates the oligonucleotide used in primer extension analysis for *atpB* mRNAs.

(B) Mapping of the 5' ends of the *atpB* transcripts in Wt Columbia (lane5) while using the sequence prepared with the same oligonucleotide (lanes 1-4).

(C) Primer extension analysis of *atpB* transcripts. 5 μg of total RNA extracted from 6 days old plants of wild type Columbia (lanes 2, 9 and 11), *sig1* (lane 3), *sig2* (lane 4), *sig3* (lane 5), and wild type Wassilewskija (lane 6), *sig4* (lane7), *sig5* (lane 10) and *sig6* (lane 12) plants have been used for analysis. Lane 1 indicates the molecular weight marker used for mapping and identification of the transcripts. LC indicates the loading control which corresponds to the PCR amplified product of 800 bp added to the mixture just before phenol chloroform extraction after cDNA synthesis.

Note: RNAs used for C (lane 2-7) part of this figure are shown in figure 8 B while the RNAs used for the C (lane 9-12) are shown in figure 8 D.

1.6. Expression analysis of the *atpE* gene.

A primary transcript being issued from position -430/431 from ATG translation initiation codon of *atpE* was found to exist in Tobacco (Kapoor et al., 1994). This transcript had its 5' end in the coding region of *atpB*. Also in *Arabidopsis* a transcript starting at position -431 from ATG translation initiation codon of *atpE* was identified by primer extension (Fig. 18B). This transcript was shown to be primary and PEP dependent (Wafa Zghidi thesis, 2008) in arabidopsis. This transcript is a monocistronic *atpE* transcript.

In order to find out the sigma factor involved in the recognition of *atpE* PEP promoter, we carried out primer extension on the total RNAs of 6 days old wild type plants (Columbia and Wassilewskija) and *sig1* (Co), *sig2* (Ws) *sig3* (Co) *sig4* (Ws), *sig5* (Co) and *sig6* (Co) plants (Fig. 18C). The a*tp*B/*atp*E genes and the localization of the oligonucleotide that was used for primer extension analysis of the *atp*E transcript are schematically indicated in Fig. 18A. The exact mapping of the -431 transcript by primer extension with an accompanying sequence ladder is shown in Fig. 18B. The primer extension analyses of all arabidopsis sigma mutants indicate that the -431 *atpE* monocistronic RNA is strongly SIG2 dependent (Fig. 18C compare lane 3 and 5).

To get an idea on the quantitative relation between *atpB/atpE* cotranscripts and *atpE* monocistronic transcripts, we performed also a Northern blot analysis using an *atpE* specific probe. The analysis was carried out using total RNAs extracted from the 6 days old wild type, *sig2* and *sig3* plants (Fig. 19A). The result shows a higher quantity of *atpB/atpE* cotranscripts than of monocistronic *atpE* transcripts (compare the 2.6 kb + 2.0 kb transcripts with the 0.7 kb transcript). In addition, the Northern experiment confirms the results of primer extension analyses of *atpB* and *atpE* mRNAs. It shows the absence of the -431 (0.7Kb) *atpE* transcript, the reduction in the 2.6 Kb *atpB/atpE* transcripts (-520 and -467) and a slight increase in the 2.0 Kb *atpB/atpE* (-84) transcript in the *sig2* plant. It also shows a reduction in the 2.0 Kb (-84 *atpB*) transcript in the *sig3* plant. All this is coherent with the primer extension results.
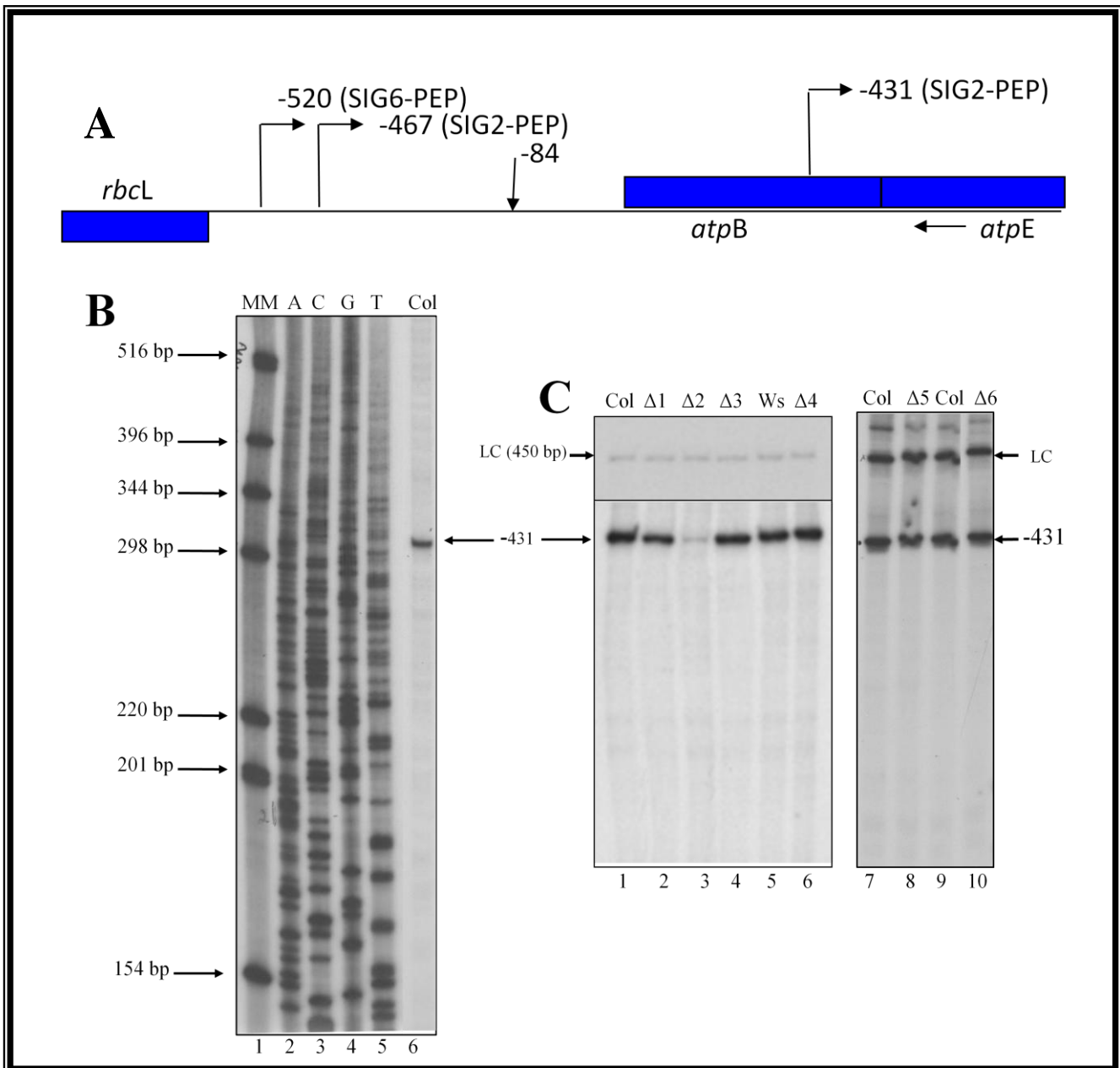
Figure 18. Transcript analysis of the *atpE* gene.

(A) Schematic presentation of the *atpE* gene. ⌐→ indicates the primary transcript while ← indicates the oligonucleotide used in primer extension analysis.

(B) Mapping of the 5' end of the *atpE* transcript in Wt Columbia (lane 6) by running the cDNA together with the sequence prepared with the same oligonucleotide that was used for primer extension (lanes 2-5). Lane 1 indicates the molecular weight marker used for mapping and identification of the transcripts in further experiments.

(C) Primer extension analysis of *atpE* transcripts. 5 µg of total RNA extracted from 6 days old plants of wild type Columbia (lanes 1, 7 and 9), *sig1* (lane 2), *sig2* (lane 3), *sig3* (lane 4), *sig5* (lane 8), *sig6* (lane 10) and wild type Wassilewskija (lane 5) and *sig4* (lane6) plants have been used for analysis. LC indicates the loading control which corresponds to the PCR amplified product of 450 bp added to the mixture just before phenol chloroform extraction after cDNA synthesis. The samples were run with molecular weight marker (not shown) on acrylamide gel.

Note: RNAs used for C (lane 1-6) part of this figure are shown in figure 8 B while the RNAs used for the C (lane 7-10) are shown in figure 8 D
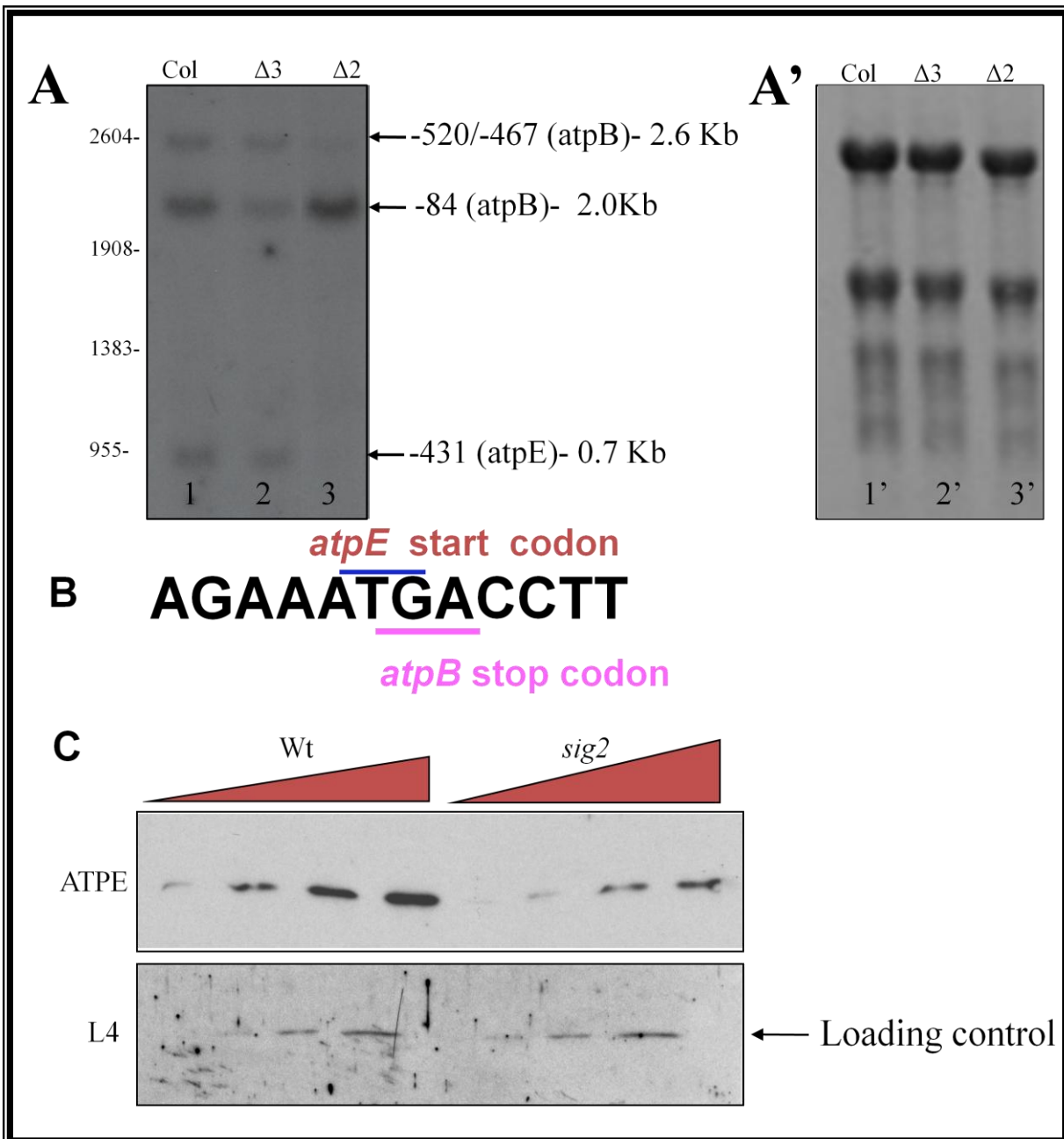
Figure 19. Expression analysis of the *atpE* gene.
 (A) Analysis of expression of *atpE* gene by Northern Blot analysis. 5 µg of total RNA extracted from 6 days old plants of wild type Columbia (lane 1), *sig2* (lane 3), *sig3* (lane 2) were separated on a denaturing agarose gel and then visualized for qualitative and quantitative analysis under the UV with the help of Ethidium Bromide added in the RNA samples. Then the RNAs were transferred on a nitrocellulose membrane and hybridized to the *atpE* specific probe. The molecular weight marker, visible on the nitrocellulose membrane under UV because of Ethidium Bromide, was marked on the membrane with the lead pencil. A' indicate the loading control.
(B) Sequence showing the overlapping *atpB* stop and *atpE* start codons.
(C) Immunoblot detection of ATPE and L4 (as loading control) proteins in Wt and *sig2* mutant plants. Four different protein concentrations have been used for this analysis.
Note: Quantitative and qualitative analysis of RNAs used in this experiment are shown in figure 9A. The membrane used for the northern blot analysis of -45 *atpH* transcripts was de-hybridized and then re-hybridized with *atpE* specific labeled probe.

The two genes of the small *atp* operon (*atpB/atpE*) have overlapping stop and start codons (Fig. 19B). We investigated whether the *atpE mRNA* could also be efficiently translated from the *atpB/atpE* dicistronic transcript. Translation of the *atpE* mRNA from the *atpB/atpE* dicistronic RNA would require a specific mechanism, the so called translational coupling (Yukawa & Sugiura, 2008). For this purpose we analyzed the levels of ATPE protein by immunoblot in Wt and in *sig2* plants that lack the -431 *atpE* monocistronic transcripts.

For the western blot analysis, ATPE specific antibodies were used. We observed the presence of ATPE protein also in *sig2* plants (Fig.19C). This indicates that even in the absence of monocistronic transcripts of *atpE*, the protein is synthesized as a result of translational coupling of the *atpB/atpE* dicistronic transcript. We used antibodies against RpL4, a nuclear encoded plastid ribosomal protein, as loading control. The question of whether the strong reduction in the ATPE protein levels reflects an important role played by SIG2 in *atpE* gene expression or whether it is at least in part due to reduced translation because of reduced tRNA levels in *sig2* plants cannot be answered at the moment.

Conclusion:

In brief, to answer the questions asked at the beginning of this chapter, we can conclude that

1- It is SIG2 that recruits PEP at the -229 *atpI,* the -467 *atpB* and the -431 *atpE* promoters, while SIG6 also plays an important role in the transcription from the -520 and the -467 *atpB* promoters in 6 days old plants. SIG3 recruits PEP at the -418 *atpH* promoter.

2- The *atpF* and *atpA* genes do not have any internal own promoters. The transcripts of both of them are processed transcripts produced as a result of cleavage of primary transcripts that are initiated from the upstream promoters of *atpH* (-418) and *atpI* (-229).

3- The transcription of the large *atp* operon (*atpI*, *atpH, atpF* and *atpA*) is co-regulated by SIG2 and SIG3. SIG3 specifically regulates transcription from the -418 *atpH* promoter and is activated to play a compensatory role in the transcript accumulation of *atpH* in the absence of SIG2. SIG2 transcribes the small *atp* operon (*atpB, atpE*) together with SIG6 and SIG2 co-regulates both *atp* operons.

4- The higher stoichiometry of *atpH* is also present at the transcript level. SIG3 might contribute to the higher stoichiometric accumulation of *atpH* transcripts under specific stress conditions and/or in other sigma mutants. However, the major role in *atpH* transcript accumulation is probably brought about at the post-transcriptional level by stabilization of the -45 transcripts by a PPR10 like protein.

5- In the absence of SIG2, SIG3 has a compensatory role in the maintenance of steady state levels of the mRNAs of the large *atp* operon (except for *atpI*). This indicates that SIG3 may play an important role in *atpH* transcription under some specific physiological and/or stress conditions.

Chapter 2

2. Do light conditions influence the expression of SIG3 dependent genes?

In the preceding chapter, the transcription of the *atp* operon has been studied. It has been shown, in particular, that the expression of the second gene in this operon, *atpH*, is activated by the presence of a specific SIG3-dependent promoter that is present in the intergenic region between *atpI* and *atpH*. This means that the *atpH* cistron could be transcribed independently from the whole ATP synthase operon. We have shown by Northern analysis that a monocistronic *atpH* mRNA exists and that this RNA is more or less strongly diminished in *sig3* plants (Fig. 12B). The observed differences in the amount of SIG3-dependent transcripts relative to all other transcripts of the large *atp* synthase operon made an interpretation concerning the importance of SIG3 for *atpH* expression impossible (see chapter 1.2c). In principal, all plant material has been grown under the same or similar conditions, *i. e*. under 12h light/12h dark cycle with light intensities at 50-60 µE. Therefore, changes in SIG3 activity are difficult to explain.

The first idea that came to our mind was the possibility that *atpH* expression is under circadian control. To verify this hypothesis, we harvested the plant material at different time points of the day and we analysed the two *atpH* transcripts (-418 and the -45) during the circadian cycle by primer extension analysis. The plants were grown under a light intensity of 50-60 µE for 7 days in a 12 hours light/12 hours dark cycle. The RNAs were extracted from plants frozen at the end of the dark cycle (Fig. 20, lane 1), at the middle of the light cycle (6 hours light) (Fig. 20, lane 2), at the end of the light cycle (12 hour ligth) (Fig. 20, lane 3), at the middle of the dark cycle (6 hours darkness) (Fig. 20, lane 4) and again at the end of the dark cycle but after 24 hours (12 hours darkness) (Fig. 20, lane 5). The data show perhaps a little more mRNA accumulation of the -45 and the -418 transcripts at the end of the dark cycle (Fig. 20 B, lanes 2' and 6'), but differences are so small that they cannot explain the differences in *atpH* monocistronic mRNA levels that we observed in our previously described experiments (see chapter 1). From this result we conclude that it is not the time point of daily harvesting that determines the differences in the *atpH* mRNA level.
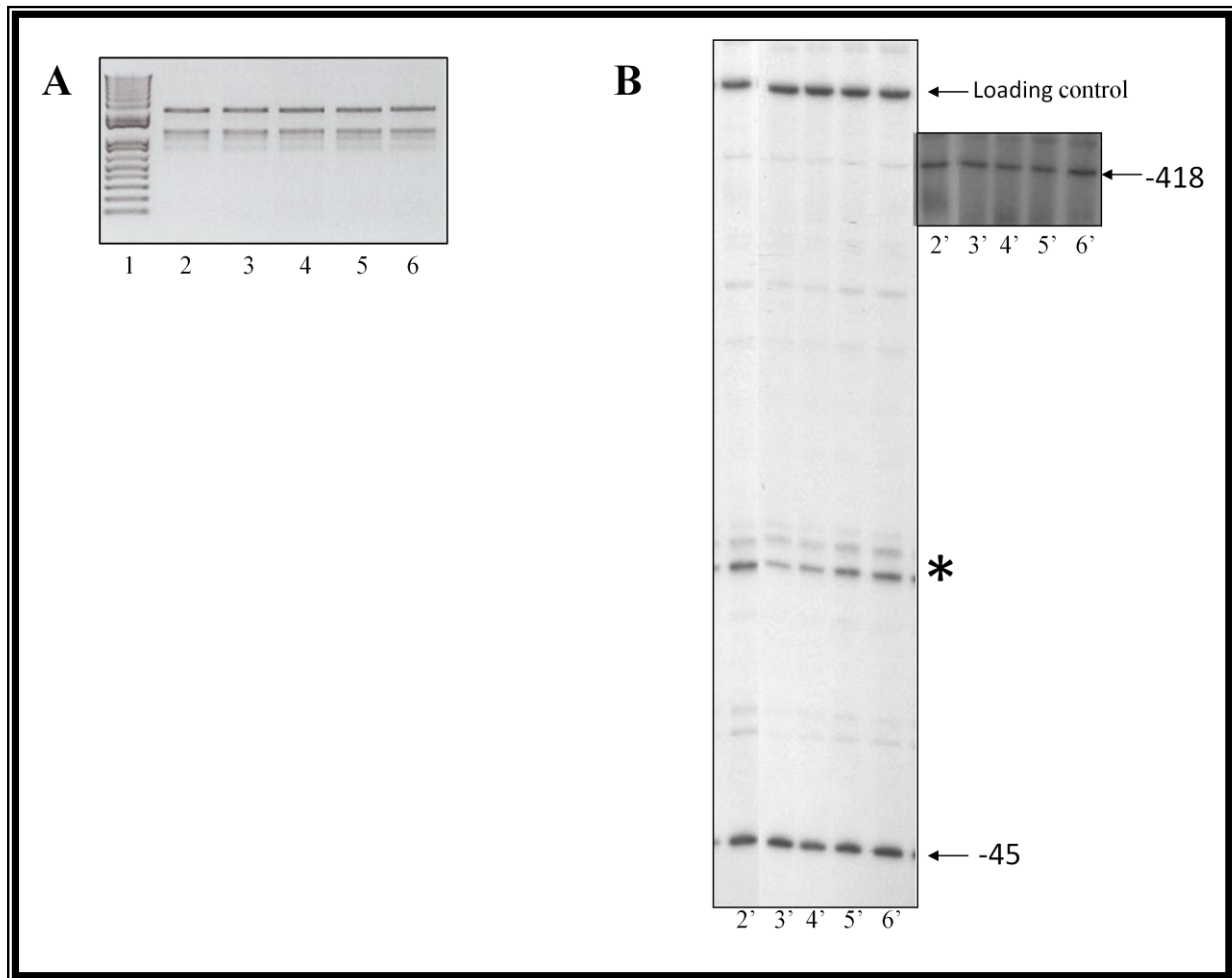
Figure 20. Primer extension analysis of both the transcripts of *atpH* gene (the -45 and the -418) in the circadian rhythm.
(A) Qualitative and quantitative analysis of total RNAs extracted from 6 days old WT plantlets at the end of the dark cycle (lane 2), at the middle of the light cycle (6 hours light) (lane 3), at the end of the light cycle (12 hour ligth) (lane 4), at the middle of the dark cycle (6 hours darkness) (lane 5) and again at the end of the dark cycle but after 24 hours (12 hours darkness) (lane 6). Lane 1 shows the 1Kb plus molecular weight marker whose map is shown in (Fig. 8 B'). RNAs were first quantified by the spectrophotometer and then analyzed on 1% agarose gel.
(B) Primer extension analysis of *atpH* transcripts. 5 µg of total RNA shown in (A) have been analysed. The small insert on the right side represents a larger exposure of the autoradiogram.   ∗, processing intermediates or degradation products.

On the other hand, we found an overexpression of the SIG3-specific *atpH* mRNA in *sig2* knock-out mutants (see Fig. 12) pointing to a supposed "rescue" role of SIG3 under adverse conditions (see also role of SIG3 in the introduction). With regard to such an eventual role of SIG3 as "rescue factor" we expected to obtain perhaps an activation of SIG3 under certain stress conditions. To verify this hypothesis we next analysed the relative expression of the *atpH* gene in regard of the other *atp* genes that are dispersed in two operons under different environmental conditions. To start with, we have at first limited our research to the action of light considering that ATPsunthase is an essential enzyme of the photosynthetic

process. Because this factor is expected to influence the expression of many genes, we decided to start with a transcriptomic analysis that permits to analyze the expression of all plastid genes at once. This approach should give an overall view and should permit to detect if the *atpH* expression is singular, or if it is linked to the expression of particular genes involved or not in the ATP synthase complex. By using a transcriptomic approach we hoped also to include the SIG3-dependent *psbN* mRNA in our analyses, *i. e.* the expression of *psbN*, *psbT* and *psbN*-initiated *psbT* antisense RNA. As shown for *Chlamydomonas*, PSBT is required for efficient biogenesis of the photosystem II (PSII) complex and for repair of photodamaged PSII reaction centres (Ohnishi et al., 2007; Ohnishi & Takahashi, 2008). According to these results we expected to find activation of expression of SIG3-dependent RNAs either under the condition of chloroplast biogenesis and/or photodamage of PSII, *i. e.* under high light stress conditions. Therefore, we have used two different experimental systems to study light effects:

1. Illumination of etiolated plantlets, *i. e.* we analyzed the influence of different light intensities on the plastid transcriptiome pattern during plastid differentiation when photosystems are built up.

2. Light stress on green plantlets, *i. e.* plantlets have been grown for 7 days at 50 μE under 12h light/darl cycle and were then exposed to 1300 μE photooxydative light conditions for several time periods.

The expression of some selected plastid genes was further studied by primer extension.

2.1. Illumination of etiolated plantlets.

Sterilized seeds (see Materials and Methods) of *Arabidopsis thaliana* wild type (Columbia) were stratified at 4°C in complete darkness for 48 hours. *In vitro* culture medium having sugar in it was used. To assure efficient germination stratified seeds were shortly (two hours) exposed to light (60 μE). We did so because if Arabidopsis seeds are not shortly exposed to light after stratification, there is either very poor or no seed germination.

After the light treatment of stratified seeds, they were transferred to a growth chamber in complete darkness for 5 days. Thereafter, we observed characteristic etiolated plants each one with long hypocotyl, yellow-coloured and closed cotyledons preceded by a typical hypocotyl curvature. The cotyledons lacked chlorophyll as their cells had etioplasts (not yet differentiated into chloroplasts). These etiolated plantlets were then exposed to different light

intensities (50, 200 or 1300 µE) for different exposure time (4, 8 or 16 h) while control plants were maintained in the dark for corresponding periods of time.

The appearance of plants after exposition to light for 16 h is shown in Fig. 21 (exposure to 4 h is not reported). The de-etiolation process is slightly more advanced with 16 h than with 4 h light, as expected.

a   b   c   d
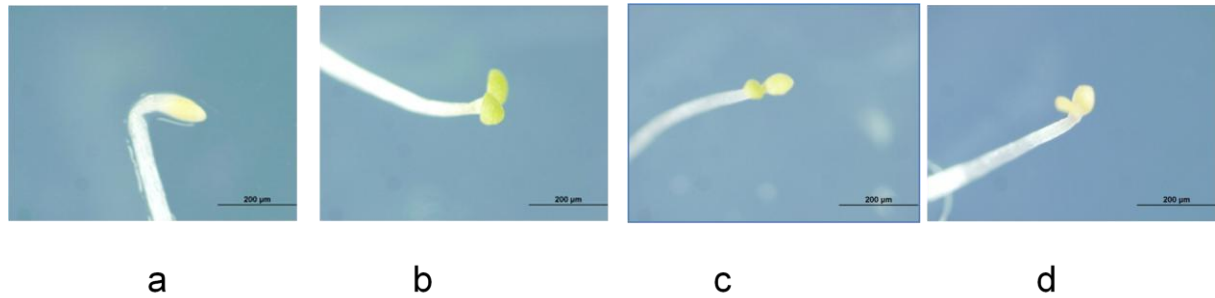
Fig. 21. Phenotypes of etiolated plantlets after light exposure.
Five days old etiolated plants were either maintained in darkness (a) or exposed to different light conditions as 50 µE (b), 200 µE (c) or 1300 µE (d) for 16 h.

At the end of the treatment, plant material was frozen in liquid nitrogen and total RNA was then extracted and used for either macroarray analysis of the plastid transcriptome or for primer extension analysis of several specific mRNAs.

2.1.1. Macroarray analyses

For macroarray analysis, mRNAs have been reverse transcribed in the presence of a radiolabelled deoxynucleotide by using specific primers to each plastid gene chosen as close as possible to the corresponding probes deposited on the macroarray. This trait in addition with the selection of 60-mer probes near the ATG start sites contribute to the semi-quantitative property of the macroarray developed in our laboratory (Patent FR 0607168). A second peculiarity of this macroarray consists in the separate detection of plastid sense and anti sense RNAs. Therefore, it permits also to detect the expression of *psbT*/ *psbN* sense/ anti sense RNAs. The spotting scheme of the array is shown in Fig. 22.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
10	psbT	psbT	rbcL	rbcL	ycf2	ycf2	ndhF.	ndhF	rpoA	rpoA	ndhC	ndhC	rpoC1	rpoC1	ycf1-1	ycf1-1
9	psbH	psbH	psaI	psaI	rps12	rps12	ndhD	ndhD	rps11	rps11	atpE	atpE	rpoB.	rpoB	psbA	psbA
8	rpoC2	rpoC2	rps2	rps2	atpI	atpI	atpH	atpH	atpF	atpf	atpA	atpA	rps16	rps16	matK	matK
7	ndhK	ndhK	ndhJ	ndhJ	rps4	rps4	ycf3	ycf3	psaA	psaA	psaB	psaB	rps14	rps14	psbM	psbM
6	psbN	psbN	clpP	clpP	rpl20	rpl20	psbE	psbE	psbF	psbF	psbL	psbL	psbJ	psbJ	atpB	atpB
5	ysf1	ycf1	rps19	rps19.	rpl22	rpl22	rps3	rps3	rpl16	rpl16	rpl14	rpl14	rps8	rps8	rpl36	rpl36
4	orf77	orf77	rps15	rps15	ndhH	ndhH	ndhA	ndhA	ndhI	ndhI	ndhG	ndhG	ndhE	ndhE	psaC	psaC
3	psbZ	psbZ	psbC	psbC	psbD	psbD	petN	petN	psbI	psbI	psbK	psbK	rps18.	rps18	accD	accD
2	psbB	psbB	rpl33	rpl33	psaJ	psaJ	petG	petG	petL	petL	petA	petA	cemA	cemA	ycf4	ycf4
1	pl23	rpl23	ndhB	ndhB	rps7	rps7	ycf5	ycf5	rpl32	rpl32	rpl2	rpl2	petD	petD	petB	petB

Figure 22. Organization of the 60-mer Arabidopsis plastid gene probes on the macroarray.
Spots (probes) detecting sense transcripts are coloured in red and spots detecting the anti-sense transcripts are coloured in blue. To facilitate the visual identification of genes, numbers are in the same order as on the 3D representation in the following figures.

Hybridization signals on the macroarray were detected by phosphoImager analysis and were quantified using the Array gauge (Fuji) software. The quantitative values were introduced into an Excel programme in the order of the spotting schema (shown in Fig. 22) and visualized with the graphic option (see Figs. 23 and 24). In addition, for easy visualization and interpretation of the results, we have presented the values of different functional groups in different colours, as detailed below.

ATP Synthase (colored in red)**:** *atpA, atpB, atpE, atpF, atpH, atpI.*

Photosystem I and Electron Transfer System (colored in light green)**:** *petA, petB, petD, petG, petL, petN, psaA, psaB, psaC, psaI and psaJ.*

Photosystem II and *rbcL* (colored in dark green)**:** *psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ and rbcL.*

Translation (colored in yellow)**:** *rpL 14, rpL 16, rpL12, rpL20, rpL22, rpL23, rpL32, rpL33, rpL36, rps11, rps12, rps14,rps15, rps16, rps18, rps19, rps2, rps3, rps4, rps7, rps8.*

Transcription (colored in blue)**:** *rpoA, rpoB, rpoC1, rpoC2.*

Chlororespiration (colored in violet)**:** *ndhA, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK.*

Genes of other Function (colored in greenish brown)**:** *accD, cemA, clpP, matK, orf77, ycf1, ycf1-1, ycf2, ycf3, ycf4, ycf5.*

In summary, the different functional groups are colored as follows in the macroarray graphs of Fig. 23:

- [green] ⟶ PSI
- [dark green] ⟶ PSII
- [orange] ⟶ ATP synthase
- [yellow] ⟶ Translation
- [blue] ⟶ Transcription
- [dark red] ⟶ Chlororespiration
- [olive] ⟶ Others

Five days etiolated plantlets were put under light of 50 µE, 200 µE or 1300 µE, for 4 and 16 hours. Control plants were kept in darkness for the same time periods. The accumulation of gene transcripts were analysed on the macroarray by using the phosphorimager and the Array Gauge software, as already indicated. Data obtained for sense and anti-sense transcripts are reported on 3D graphs in Fig. 23.

Illumination for 4h (Fig. 23, A-D): In general, RNA levels (reflected by the height of each cone) are very low in etiolated plants and after illumination with 50 or 200 µE. In the following, we will focus only on significant changes in mRNA levels. After exposure of plantlets to 50 µE for 4 hours, the mRNA expression profile is still very similar to the control in darkness, except for *psbM* mRNA that increased (position 16/7, compare Fig. 23A and B). This early increase in *psbM* mRNA is surprising because PSBM seems not to be required for biogenesis of PSII complexes (Umate et al., 2007). After exposition to 200 µE for 4 h (Fig. 23C) a slight but clear raise in RNA levels is observed for several gene groups, especially for the PSI, PSII and ATP synthase groups. Interestingly, the most increased mRNAs correspond to the *psaJ* and *psbM* genes (positions 5/2 and 16/7 on the array) and not to the well-known photogene *psbA* (position 16/9) that is thought to belong together with *rbcL* to the highest expressed plastid genes. To note also, for a number genes sense and antisense transcripts are present in rather equal amounts, look to genes as *psaA*, *psbB*, *psbZ, orf77, ycf1, ndhK* and *psbN*. For *psaA*, antisense expression is even higher than sense expression.

After exposition of plants at 1300 µE for 4h (Fig. 23D) the mRNA levels of a number of genes are noticeably increased. This is the case for several genes of the PSI and PSII complex, for *rbcL* and for genes coding plastid ribosomal proteins. On the other hand, the
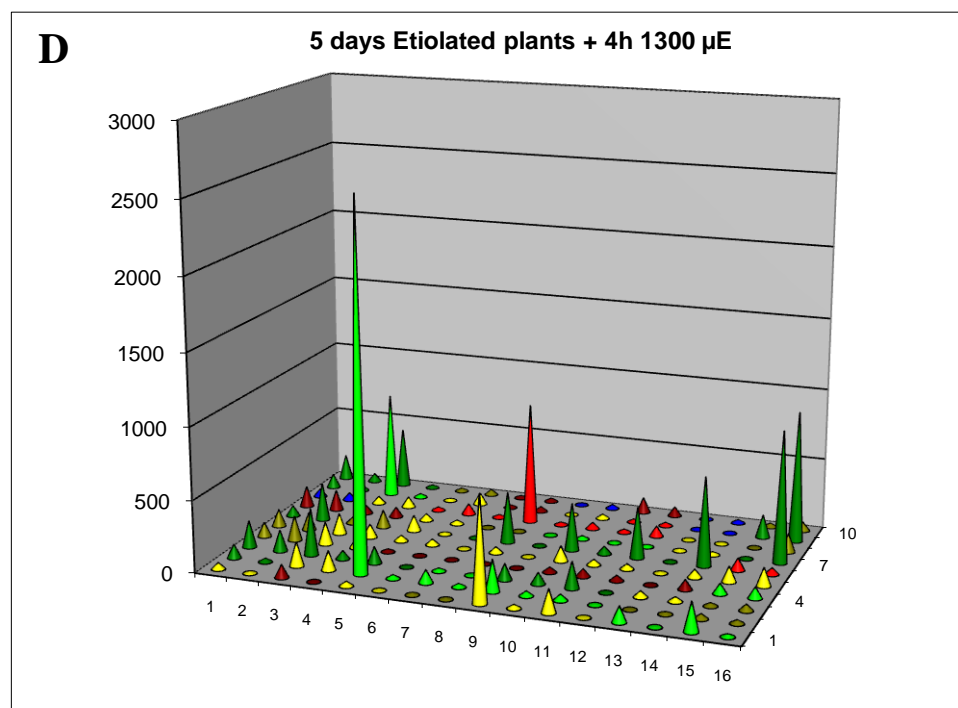
antisense expression observed under 200 µE light exposure is diminished under 1300 µE. A particular strong activation is observed for the *psbA* gene (position 16/9) that reaches now the level of *psbM* (position 16/7). However, even now *psbA* is not the most expressed gene. The highest mRNA level is detected for the *psaJ* gene (position 5/2). PSAJ is a small hydrophobic subunit of the photosystem I complex (PSI) that is important for the formation of the plastocyanin-binding domain of PSI and plays an important role for the stability or assembly of the complex (Hansson et al., 2007). Of particular interest to us, a high increase is also observed for *atpH* mRNA (position 8/8) compared to the other mRNAs of the *atp* sythase complex.

Altogether, the short exposure of etiolated plants to light (4h) shows that many genes are light dependent, especially those that are necessary to establish the photosystems in developing young plastids. This latter observation is not surprising, but an important observation is that the activation of gene expression is differentiated. All the plastid genes are not evenly turned on by light. This implies either a regulation at the transcription level or differences in the stability of transcripts. As the observations are made at the very early steps of plastid differentiation, results are more likely to be interpreted with the first hypothesis.

Figure 23. Schematic representation of macroarray results of the plastid transcriptome after 4 and 16 h illumination of etiolated platlets with different light intensities.
Five-days-etiolated plants were exposed to 50 µE (B and F), 200 µE (C and G) and 1300 µE (D and H) for 4h (B-D) and for 16 h (F-H). Control plants kept in darkness for 4h (A) and for 16h (E). Afterwards, total RNA was isolated, reverse transcribed in the presence of radiolabelled dATP, hybridized to plastid-specific macroarrays and hybridization sigbnals were quantified with a phosphoImager as described above.

The X and Z axis indicate the position of the gene probes on the macroarray.
The Y-axis corresponds to the hybridization signal intensity
.

A 5 Days Etiolated Plants + 4h darkness

B 5 Days Etiolated plants + 4h 50 µE

C 5 Days Etiolated plants + 4h 200µE

D 5 days Etiolated plants + 4h 1300 µE

E Etiolated plants, 5 days darkness + 16h darknss

F Etiolated Plants 5days darkness + 16h 50 µE

G Etiolated Plants 5 days darkness + 16h 200 µE

H Etiolated plants 5days darkness + 16 h 1300 µE
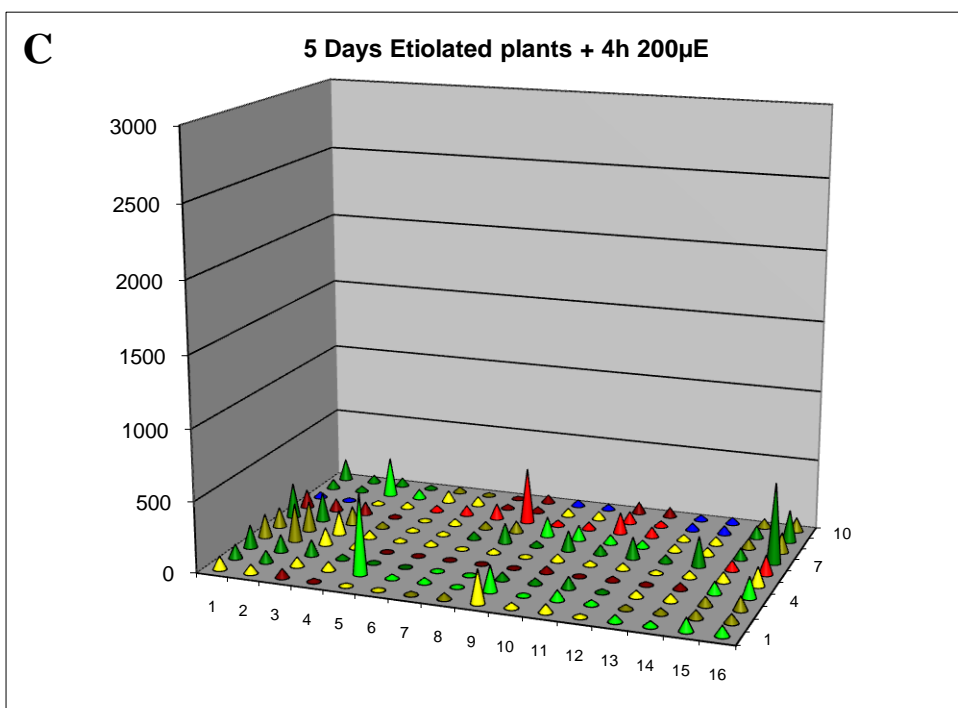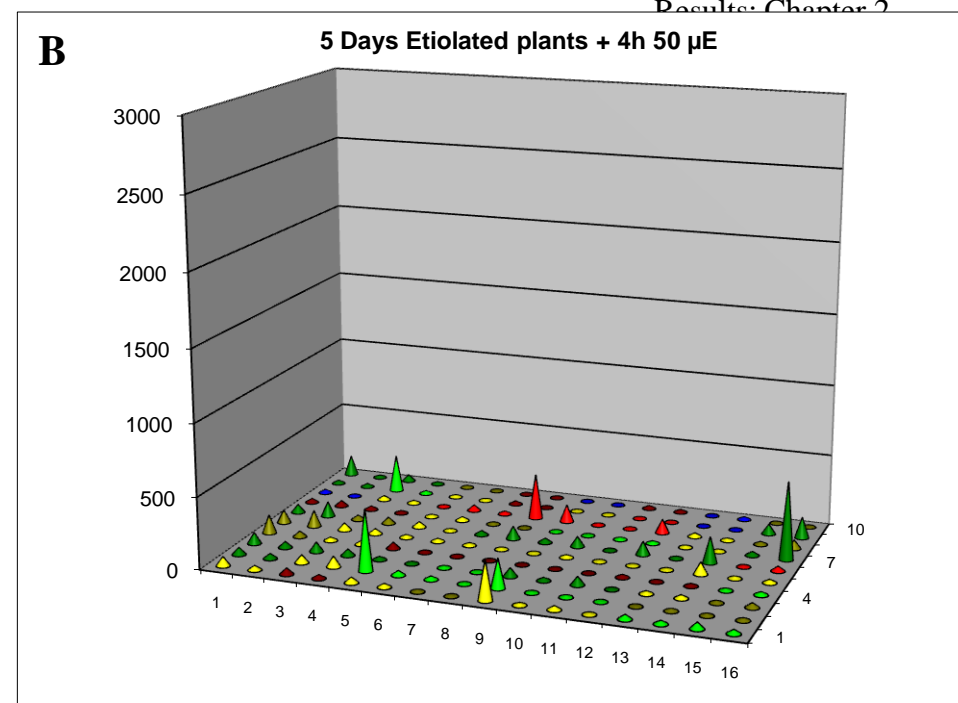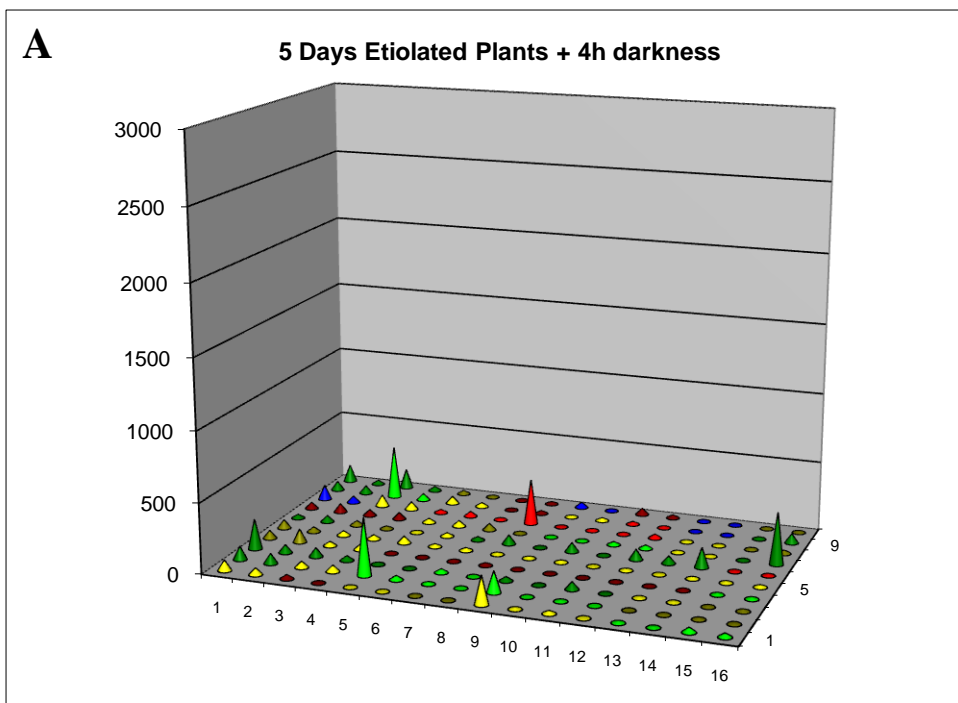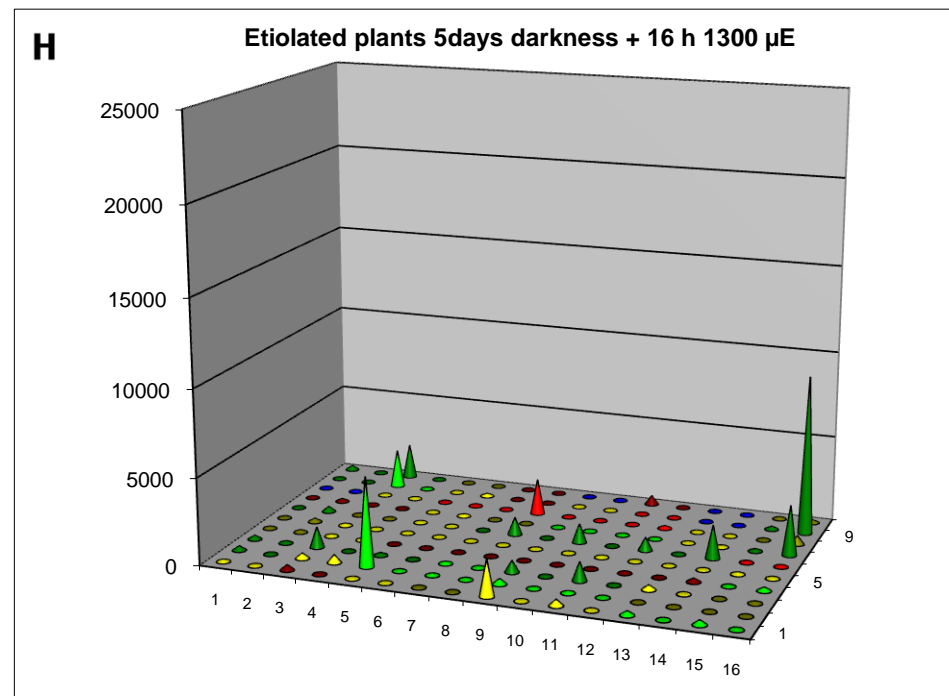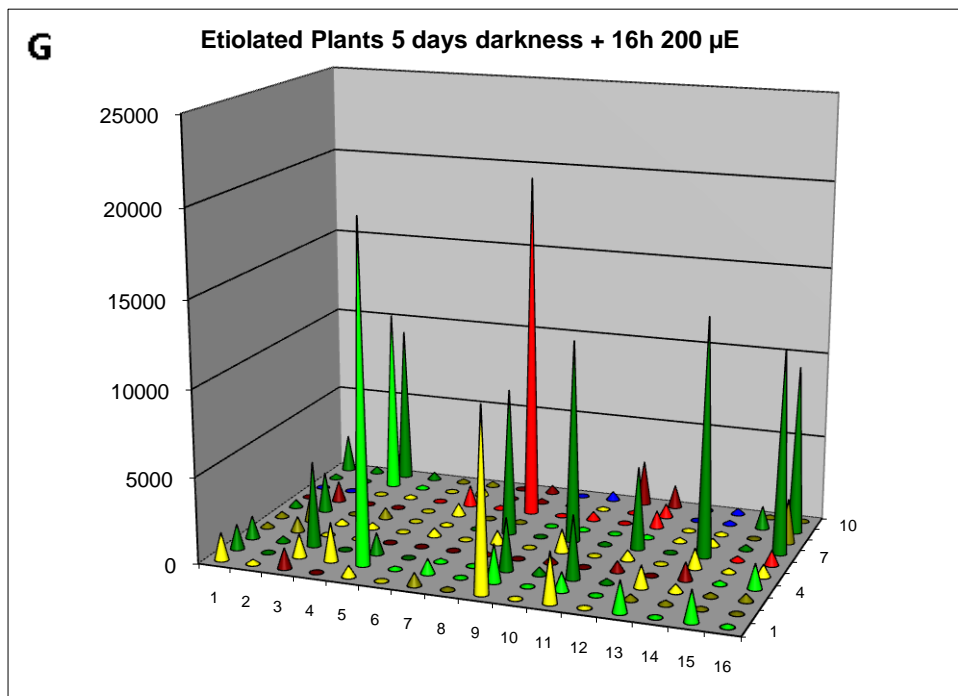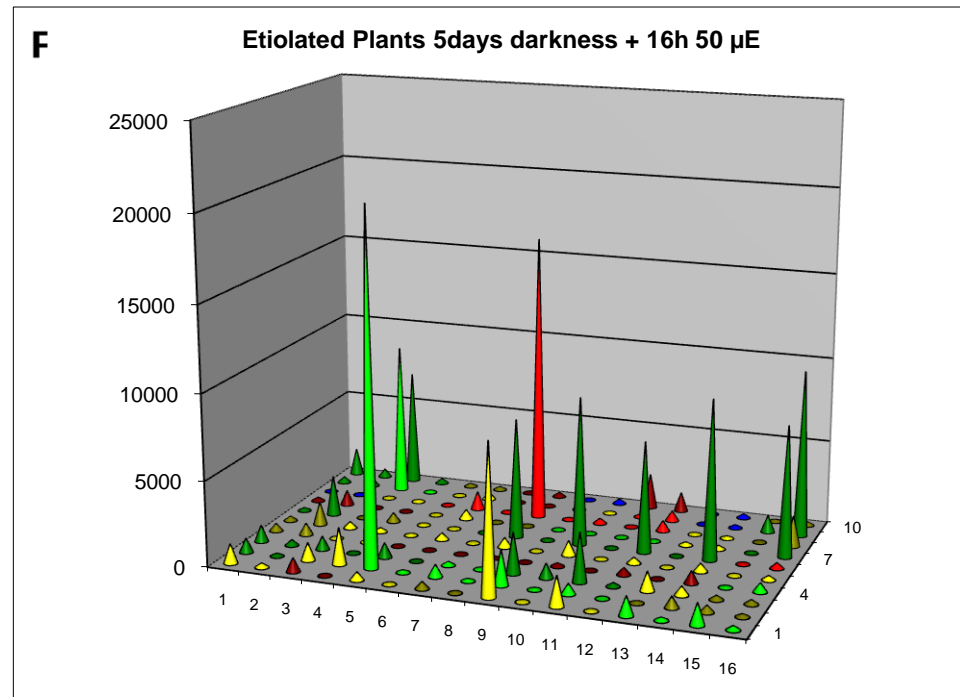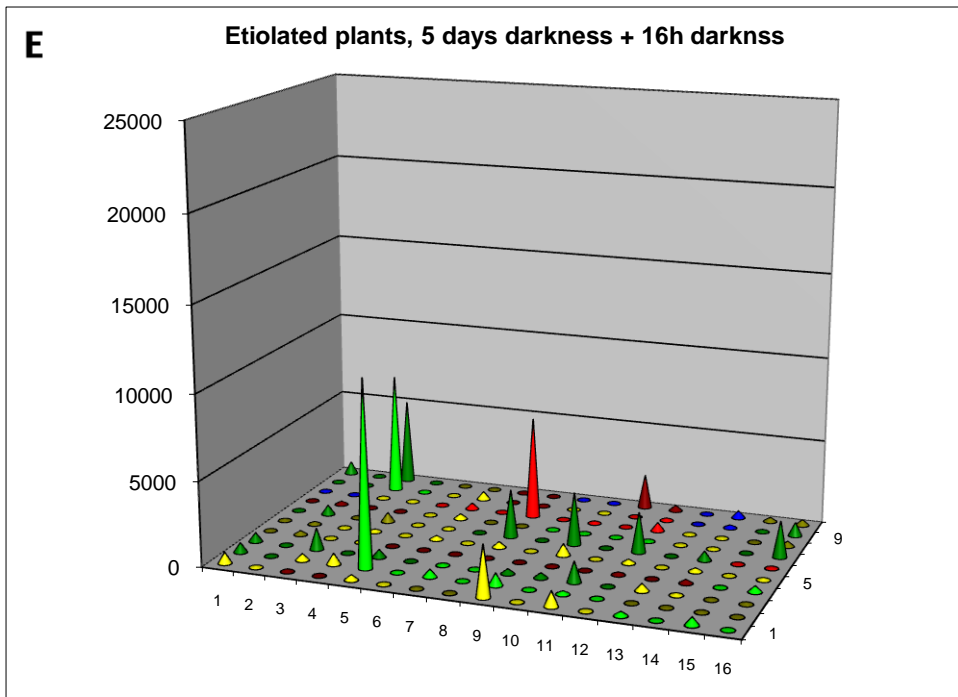
1Illuminated for 16h (Fig. 23, E-H): After exposure of plantlets to different light conditions for 16 h, a time period that correspond to a long day photoperiod, the mRNA expression profiles changes considerably. We observe a strong increase of the mRNA levels for many but not all genes after illumination at 50 or 200 μE of light (Fig. 23, compare F and G to E). To notice is the high expression of the *rpl32* gene among the genes coding for ribosomal proteins. Taken together with the results obtained for 4h illumination we observe that for the three groups of genes, i. e. PSI, ATP synthase and r-proteins, always one gene is much stronger expressed than the others belonging to the same group. For PSI it is *psaJ*, for the ATP synthase it is *atpH* and for the group of ribosomal proteins it is *rpl32*. The meaning of these different mRNA levels is not yet clear. Interestingly, *psaJ* and *atpH* mRNAs seem to be stabilized by PPR10 in maize chloroplasts (Pfalz et al., 2009). It would be interesting to know whether also the *rpl32* mRNA is also specifically stabilized by a PPR protein.

In contrast to the general augmentation of mRNA levels after illumination at 50 and 200 μE, exposure to photooxydative light of 1300 μE provokes a strong diminution of mRNA levels (up to 10 times), falling even below dark levels, of all the genes that are activated with the lower light intensities (Fig. 23, compare H to E). This overall reduction of transcript accumulation after high light exposure could be interpreted as a photoinhibition. The expression of *psbA* is a noticeable exception as this gene is expressed at about the same level under a lower light brilliance.

Altogether, we notice very strong differences between the mRNA levels of the different plastid genes. Not all mRNAs accumulate after light exposure. Many mRNAs are always present at very low levels only, so low that a quantitative exploration of the results for these genes would not give reliable results. Among these mRNAs are unfortunately most of the antisense RNAs and finally we cannot conclude on quantitative values between *psbT* sense and *psbT* antisense RNAs from the results of the macroarray analyses. However, the mRNA levels of the SIG3 controlled *atpH* and *psbN* genes are high enough to permit conclusion on their changes in response to different light conditions.

2.1.2. Action of light on SIG3 dependent gene expression

Transcription of the *atpH* and *psbN* genes is controlled by the SIG3/PEP holoenzyme. Therefore, we would expect similar behaviour of these genes when plants are exposed to different light intensities. Relatively to the background level, *atpH* transcript levels are much

higher than *psbN* transcript levels (see in particular Fig. 23F and G). However, this relation is found already in darkness (Fig. 23A and E). To know whether the expression of *psbN* and *atpH* is similarly activated by light, we calculated the difference of transcript levels between dark and light growth for both of these genes. The ratios thus obtained are reported in Fig. 24.
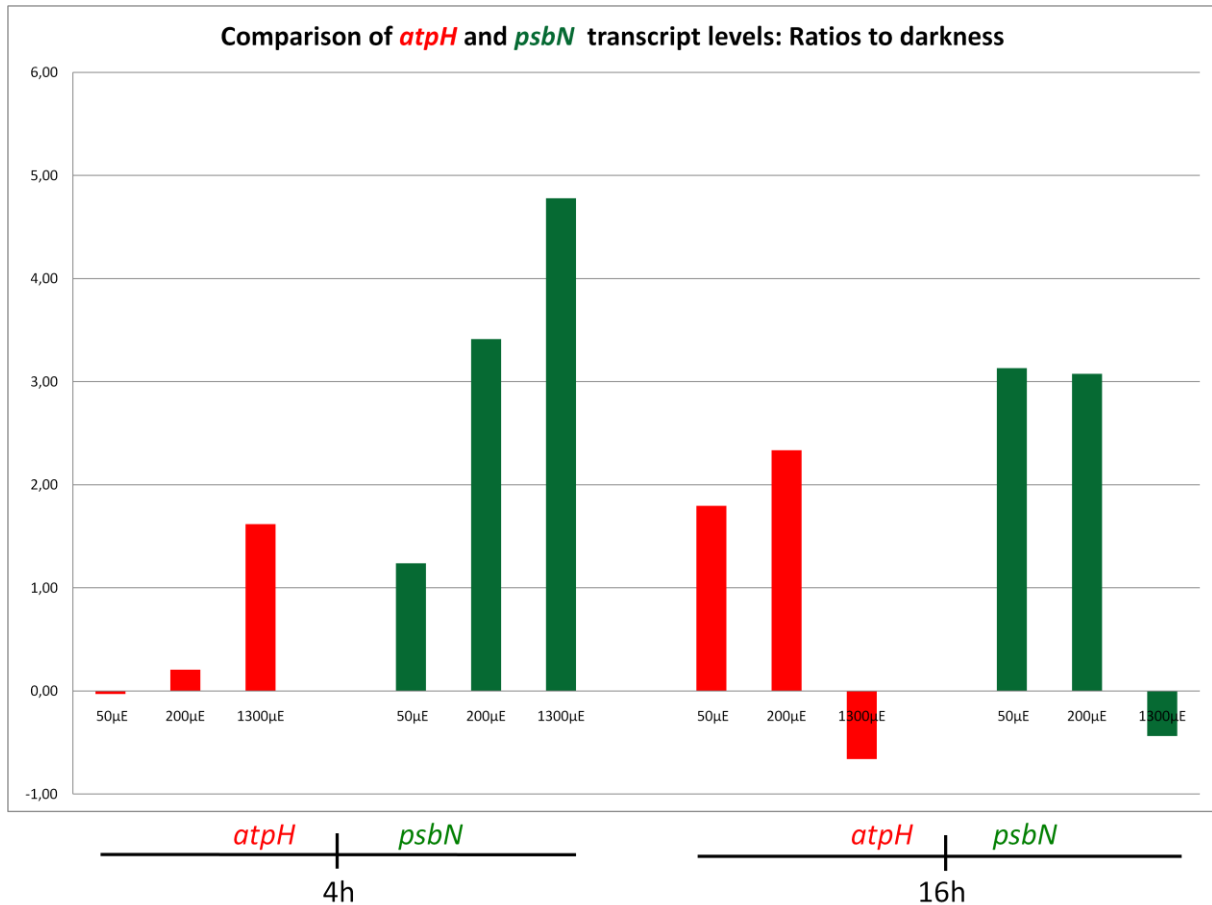
Figure 24. Accumulation of *atpH* and *psbN* transcripts in light relative to darkness.
In order to get 0 for "no change" and negative values for ratios where the dark level exceeds the light level "ratios *minus* 1" are reported in the ordinate of the graph. In this way, activation (ratio > 0) and inhibition (ratio < 0) by light is positively or negatively represented. Values for the *atpH* are in red and values for *psbN* in green. Data are taken from the macroarrays analyses of etiolated plants (Fig. 23) and are reported for 4h (left hand side) and 16h (right hand side) in the order of increasing light intensities as indicated in the abscissa.

After 4h of light exposure, *atpH* and *psbN* mRNA levels are increasing with the intensity of light. However, values differ for the two genes. While *atpH* mRNA accumulation is most enhanced between 200 and 1300 µE, *psbN* mRNA accumulation is most stimulated between 50 and 200 µE. Activation value are higher for *psbN* than for *atpH*, i. e. *psbN* mRNA is more photosensitive than *atpH* mRNA. After 16 h illumination, mRNAs levels are considerably enhanced at 50 µE for both genes. Illumination at 200 µE does not result in remarkable additional enhancement of mRNA levels and at 1300 µE mRNA levels are

diminished when compared to dark levels. We concluded that the activation/inhibition of the expression of the two genes is quite similar but not identical. The highest accumulation of *psbN* mRNA is found at 1300 µE illumination for 4h while the highest accumulation of *atpH* mRNA is observed at 200 µE illumination for 16h (Fig. 24).

Concerning *atpH* transcripts, we were looking for conditions in which the *atpH* mRNA levels are highest in relation to the mRNA levels of the other ATP synthase genes. For this reason, the values for all mRNA levels of the two ATP synthase genes are reported in Table I. High levels of *atpH* mRNA might correspond to mRNA stabilization or to supplementary transcription of the mRNA by SIG3/PEP holoenzyme. Although we cannot decide immediately between these two possibilities, at first we were looking to find conditions of high *atpH* mRNA levels. The first thing to notice is the already relatively high *atpH* mRNA level in etiolated plantlets that explains why the *psbN* expression is more sensitive to light than *atpH* expression (see Fig. 24).

Genes	4h darkness	/atpI	4h 50µE	/atpI	4h 200µE	/atpI	4h 1300µE	/atpI
atpI	37	1,0	41	1,0	65	1,0	75	1,0
atpH	323	8,7	313	7,6	389	6,0	845	11,3
atpF	16	0,4	13	0,3	52	0,8	46	0,6
atpA	30	0,8	97	2,4	69	1,1	74	1,0
atpB	14	0,4	35	0,8	135	2,1	47	0,6
atpE	12	0,3	2	0,0	40	0,6	26	0,3

Genes	16h darkness	/atpI	16h 50µE	/atpI	16h 200µE	/atpI	16h 1300µE	/atpI
atpI	363	1,0	1057	1,0	1184	1,0	99	1,0
atpH	5987	16,5	16749	15,8	19959	16,9	2041	20,6
atpF	154	0,4	311	0,3	484	0,4	13	0,1
atpA	501	1,4	615	0,6	971	0,8	175	1,8
atpB	94	0,3	345	0,3	861	0,7	118	1,2
atpE	166	0,5	581	0,5	763	0,6	54	0,5

Table I. Signal intensity values of *atp* mRNAs in etiolated plants exposed to different light conditions (darkness, 50 µE, 200 µE and 1300 µE) for 4h and 16h. The signal intensity value of *atpI* (the first gene of the large ATP synthase operon) was taken as reference. The values of other *atp* mRNAs were divided by the *atpI* value to get the stoichiometric transcript accumulation of each *atp* mRNA in relation to *atpI*. Values were extracted from the macroarray analyses. For each growth condition, only one macroarray was used.

Interestingly, *atpH* mRNA accumulation in relation to the mRNA levels of all other ATP synthase genes seems to be independent from light, except for 1300 µE photooxydative stress. Although all values are generally diminished after photooxydative light, *atpH* mRNA levels are significantly higher when compared to the other genes.

Conclusion on the macroarray results:

We have used a macroarray approach to analyse the effect of light on the mRNA levels of SIG3 dependent plastid genes (Fig. 23). Etiolated plantlets were exposed to different light intensities for short (4h) and longer (16h) time. We found very large differences in the expression levels of different gene groups. Very lowly expressed genes are *ndH* genes, genes coding for transcriptional components and many genes that were grouped "genes of other functions". For these genes, values obtained for mRNA levels were too low to conclude on changes in transcript levels after light treatment.

The genes of photosystem I, photosystym II and ATP synthase are the most expressed plastid genes during the de-etiolation process under light exposure. But even within one group of genes, not all mRNA levels are equal. Highest levels were found for *psaJ* (PSI), *psbM* (PSII) and *atpH* (ATP synthase). In the case of *psaJ* and *atpH* these high mRNA levels might be determined by special stabilization of these mRNAs by a PPR10-like protein. PPR10 stabilisation was only discovered very recently during my thesis and it makes the interpretation of the macroarray results more difficult because we cannot decide between stabilization and *de novo* synthesis of *atpH* mRNA. Highest mRNA amounts for *atpH* are observed after 4h illumination of etiolated plantlets at 1300 µE and after 16h illumination at 200 µE (Fig. 24). After 16h illumination, the *atpH* mRNA level is even higher than the *psaJ* mRNA level.

2.1.3. Primer extension analyses

In order to confirm the macroarray results on *atpH* expression by another method and to distinguish between SIG3 initiated transcripts (-418 *atpH* mRNA) and *atpI/atpH* initiated processed transcripts (-45 *atpH* mRNA), we performed also primer extension analysis of total RNAs isolated from etiolated plantlets after 16h illumination with different light intensities (Fig. 25). The experiment was done using the same RNAs which were used for macroarray analysis. Primer extension analysis shows an increase in both *atpH* transcripts (the -45 and the -418) after exposure of etiolated plants to 50µE and 200 µE light (Fig. 25B, compare lane 2' with 3' and 4'). A strong decrease in mRNA is observed for both transcripts in plants exposed to 1300 µE for sixteen hours (Fig. 25 B, compare lanes 2' and 5'). Thus, primer extension analysis of *atpH* transcripts of 5 days etiolated plants exposed to different light intensities confirmed the results obtained by macroarray analysis. However, both types of RNA, *atpI*

initiated and -418 *atpH* initiated transcripts, augment or decrease concomitantly after the different light treatments and no specific SIG3 related effect could be discerned.
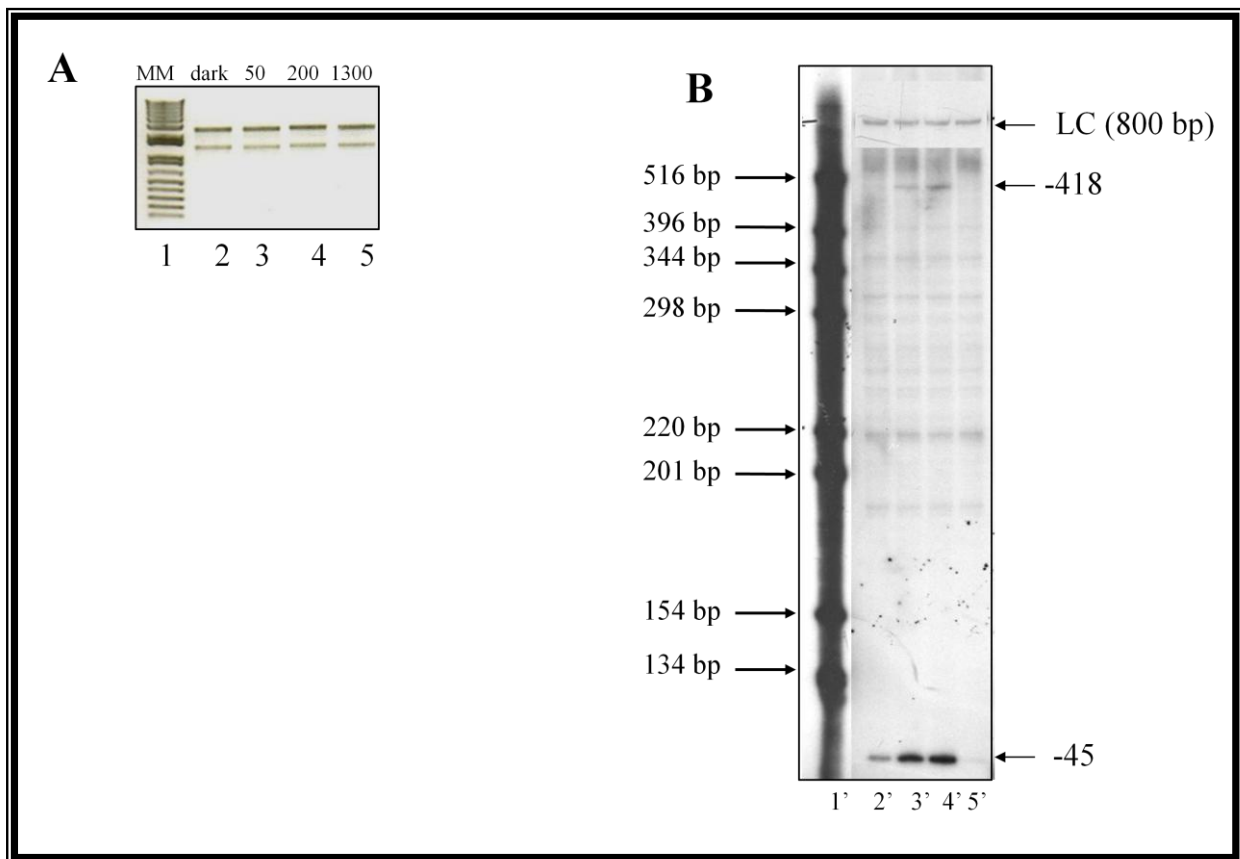
Figure 25. Primer Extension analysis of *atpH* mRNAs.
(A) Qualitative and quantitative analysis of total RNAs extracted from the 5 days etiolated plants exposed to darkness (lane 2), 50 (lane 3), 200 (lane 4) and 1300 µE (lane 5) for 16 hours. Lane 1 shows the 1Kb plus molecular weight marker whose map is shown in (Fig. 8 B'). RNAs were first quantified by the spectrophotometer and then analyzed on 1% agarose gel.
 (B) Primer extension analysis of *atpH* mRNA. 5 µg of the same RNAs as shown in (A) have been analysed by primer extension. Lane 1 represents the molecular weight marker used for mapping and identification of the transcripts. LC indicates the loading control which corresponds to the PCR amplified product of 800 bp added to the reaction just before phenol chloroform extraction after cDNA synthesis. -418 is the SIG3 dependent primary *atpH* transcript while -45 is a processed transcript being cleaved from -418 *atpH* and -229 *atpI* primary transcripts.

Conclusion on the primer extension result:

The primer extension result indicates that results of macroarray analysis of etiolated plants are at least semi-quantitative and reliable. Nevertheless, the macroarray experiments need to be repeated biologically, using more RNA and higher radioactivity for labelling. This might permit to get reliable values also for the lowly expressed genes that could not be considered here.

The here presented results show in each case when *atpH* mRNAs augment, both *atpH* transcripts (the -45 and the -418) augment in about the same relation. In fact, when starting these experiments, we hoped to find conditions under which the SIG3 dependent -418 transcripts are specifically enhanced. This would have allowed us to conclude on the function of SIGMA 3 and the -418 initiated monocistronic *atpH* transcript. Because both transcripts, the -45 processing product that includes also *atpI* initiated co-transcripts and the -418 transcripts, augment we cannot conclude on a specific activation of SIG3 dependent transcription at the -418 promoter under different light conditions. Experiments need to be repeated by comparing WT and *sig3* plants. This would allow concluding whether the augmentation of the -45 transcripts is due to transcription by SIG3, SIG2 or both sigma factors.

2.2. Light stress of green plants (photoinhibition of chloroplasts).

As seen in the chapter before, under short time illumination (4h) highest augmentation of many mRNA levels was obtained for 1300 µE of light intensity (Fig. 23, A-D). On the other hand, after treatment with high light for 16h mRNAs do not augment, on the contrary mRNA levels fall below the dark values, except for *psbA* (Fig. 23, compare E and H). In the former chapter, all experiments were done using young etiolated plantlets in which photosystems are not yet established. High light treatment might be more dangerous for etiolated tissues than for green tissues where photosystems are already well established. To verify this point we used plants now that were grown for 7 days under 12h-light/dark cycle, with 50-60 µE light intensity during the light period. This time, we also included *sig3* plants for better interpretation of changes in the -45 *atpH* mRNA. 7 days old WT and *sig3* plantlets were exposed to 1300 µE for 4, 8 and 16 hours. Using the RNAs extracted from these plants we carried out primer extension analysis for the two transcripts of *atpH* (the -45 and the -418) (Fig. 26).
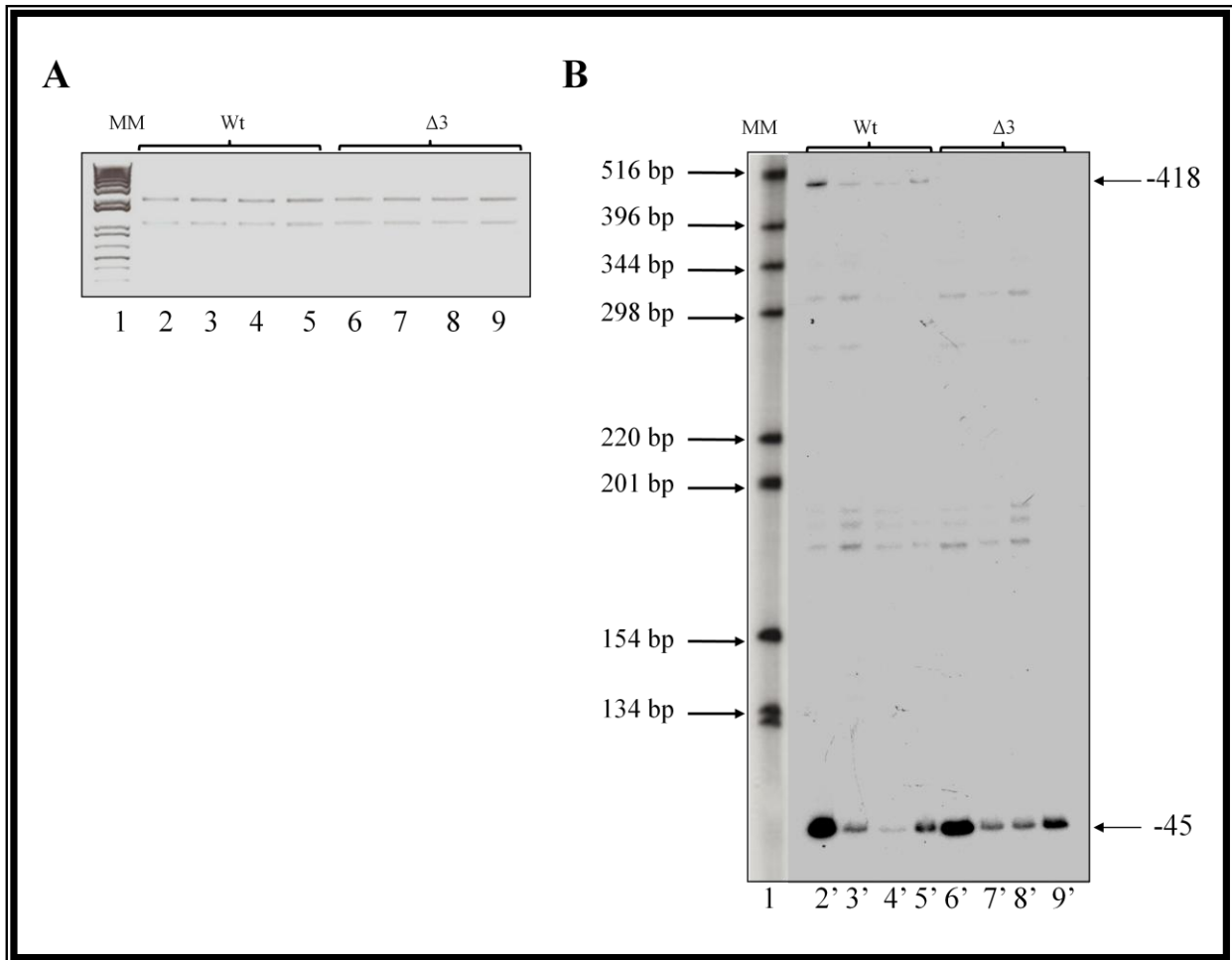
Figure 26. Primer extension analysis of *atpH* mRNAs of 7 days old plants grown under 50-60 µE in 12 hours light/12 hours dark cycle.
(A) Qualitative and quantitative analysis of total RNAs extracted from 7 days wild type plants kept under 50-60 µE for 4 hours (lane 2) or exposed to 1300 µE for 4, 8 and 16 hours (lanes 3, 4, 5 respectively). Similarly 7 days *sig3* plants were kept under 50-60 µE for 4 hours (lane 6) or exposed to 1300 µE for 4, 8 and 16 hours (lanes 7, 8 and 9 respectively) Lane 1 shows the 1Kb plus molecular weight marker whose map is shown in (Fig. 8 B'). RNAs were first quantified by the spectrophotometer and then analyzed on a 1% agarose gel.
(B) 5 µg of the same RNAs as shown in (A) were analysed by primer extension to reveal *atpH* precursor RNAs. Lane 1 indicates the molecular weight marker used for mapping and identification of the transcripts.

Unexpectedly, we observe a strong reduction in accumulation of the -45 and the -418 *atpH* transcripts already after 4 h exposure to high light in WT plants (compare lane 2' to lanes 3'-5'). A similar strong reduction is observed for the -45 transcripts in *sig3* plants that lack the -418 transcripts (Fig. 26 B, compare lane 6' to lanes 7'-9' and lane 3'-5'). In general, the -45 transcripts are not more diminished in *sig3* than in WT plants after light stress. This means that only very few of -45 transcripts originate from the -418 SIG3 dependent transcripts. Even a little more -45 transcripts seem to exist in *sig3* plants when compared to WT plants after 8 and 16h light exposure. This might be due to over expression of SIG2 in *sig3* plants. To get a higher expression of SIG3 initiated *atpH* transcripts perhaps we should try to analyse *SIG2* knockout plants.

In fact, the effect of 1300 µE light on 7 days old green plants is different to what was observed in etiolated plants. In etiolated plants there was at first an accumulation of *atpH* transcripts (4h light) and a strong diminution after 16h light exposure (Figs. 23 and 24 and Table I) but in green plants there is strong reduction in transcript accumulation of both the -45 and the -418 transcripts already after 4h of high light treatment (Fig. 26B).

In order to verify whether this effect of reduction in transcript accumulation in the plants grown under normal light conditions (green plants) on exposure to high light 1300 µE is specific to *atpH* or it is a general phenomena, we carried out primer extension analysis of some genes of the same operon of *atpH* and genes of other operons as well (Figs. 27 and 28). For these analyses we used only RNAs extracted from the plants exposed to 1300 µE for 4 hours as the effect of 8 and 16 hours exposure of 1300 µE was similar to what was observed after 4 hours exposure.

Primer extension analysis of the *atpI* -229 primary transcripts indicates that there is a strong reduction in accumulation of these transcripts when the 7 days old green plants are exposed to 1300 µE for 4 h. This reduction was observed in both the arabidopsis wild type and *sig3* mutant plants (Fig. 27B, compare lane 2'with 3' and lane 4' with 5'). Similar results were observed in primer extension analysis of these plants for *clpP* gene transcripts that are made by the nucleus encoded NEP RNA polymerase (Fig. 27C).
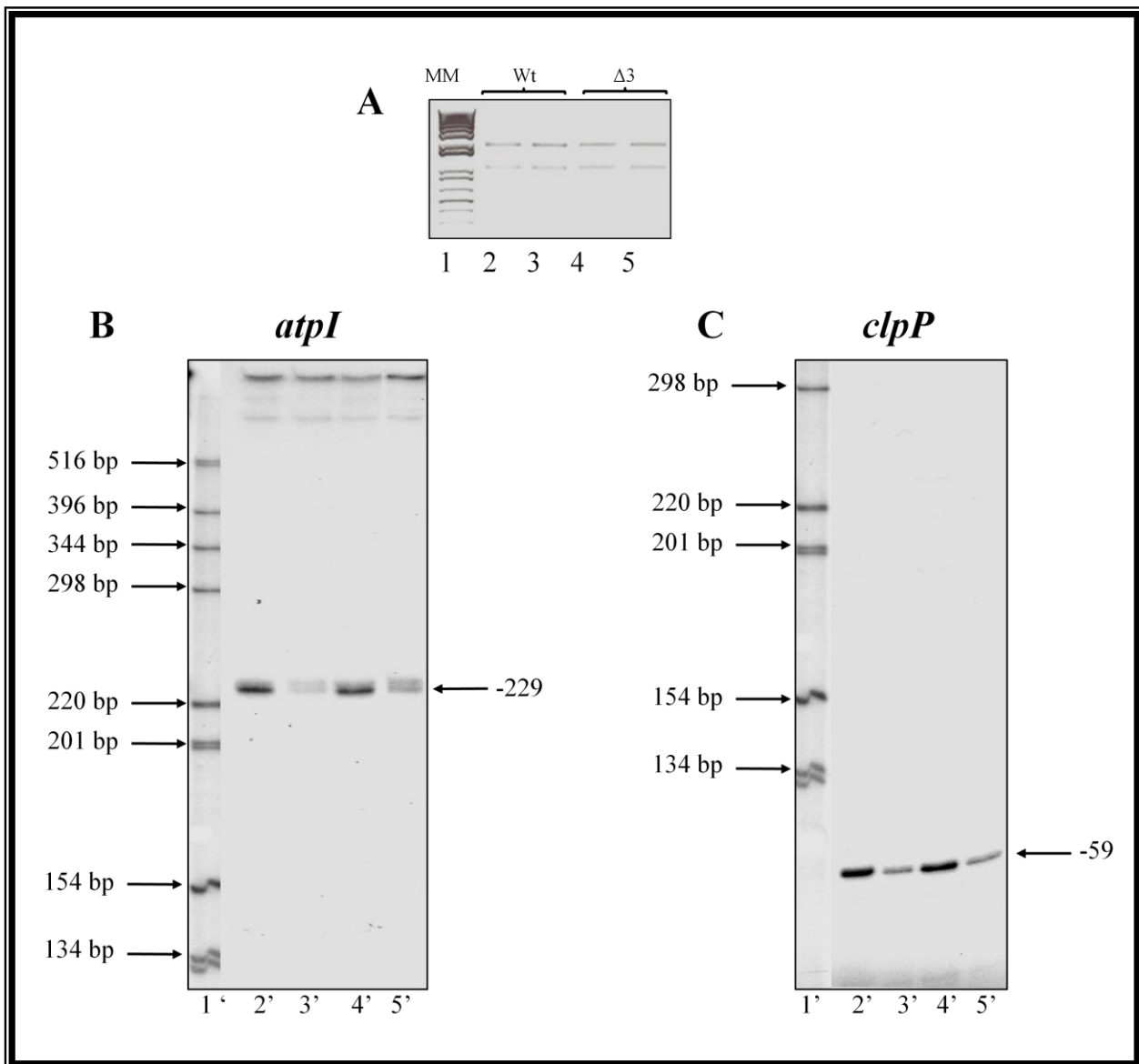
Figure 27. Primer extension analysis of *atpI* and *clpP* mRNAs of 7 days old Arabidopsis plants grown under 50-60 µE under 16h light/8h dark cycle.
(A) Qualitative and quantitative analysis of total RNAs extracted from wild type plants and *sig3* plants kept either under 50-60 µE for additional four hours (lane 2, 4 respectively) or exposed to 1300 µE for 4 hours (lanes 3, 5 respectively). Lane 1 shows the 1Kb plus molecular weight marker whose map is shown in (Fig. 8 B'). RNAs were first quantified by the spectrophotometer and then analyzed on a 1% agarose gel.
(B) 5 µg of the same RNAs as shown in (A) have been analysed by primer extension to reveal *atpI* (B) and *clpP* (C) precursor RNAs. Lane 1 indicates the molecular weight marker used for mapping and identification of the transcripts.

Equally, 16S ribosomal precursor transcripts being issued from the PC (NEP) and the P2 (PEP) promoters (Fig. 28 B) are diminished after light treatment. This shows that the transcripts of all analysed genes, *atpI*, *clpP* and *16S,* are diminished after high light treatment, independent whether they are transcribed by NEP or by PEP.
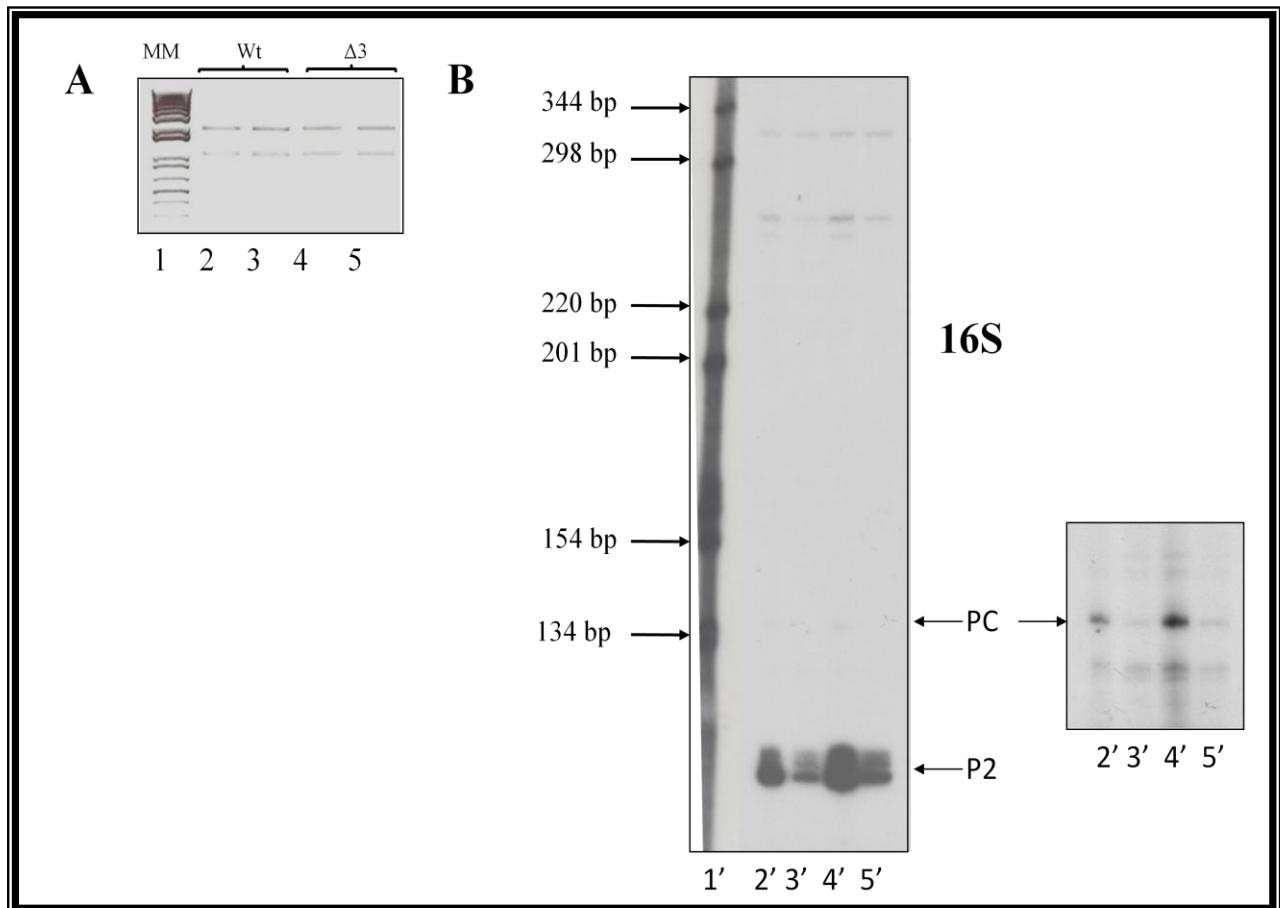
Figure 28. Primer extension analysis of 16S *rrn* precursor transcripts using 7 days old Arabidopsis plants grown under 50-60 µE under 16h light/8h dark cycle.

(A) Wild type plants and *sig3* plants were kept either under 50-60 µE for additional four hours (lanes 2, 2' and 4, 4' respectively) or have been exposed to 1300 µE for 4 hours (lanes 3, 3' and 5, 5' respectively).

(B) Total RNA has been extracted and analysed either by agarose gel electrophoresis as quality control (A) or by primer extension to reveal 16 S ribosomal precursor RNAs. The small insert at the right side of (B) represents a long exposure time of the autoradiogram. Lane 1 indicates the molecular weight marker used for mapping and identification of the transcripts.

Altogether, we observed large differences in changes of mRNA levels in response to different light conditions in the two different experimental conditions that had been used. To further explore the importance of the physiological stage of the plant material for light responses, we analysed finally the effect of 1300 µE light on the transcript accumulation of *atpH* and *atpI* in 34 days old mature plants (Fig. 29). Arabidopsis wild type and *sig3* plants were grown under 50-60 µE in a 16 hours light and 8 hours dark cycle and exposed to high light condition for 2h30.
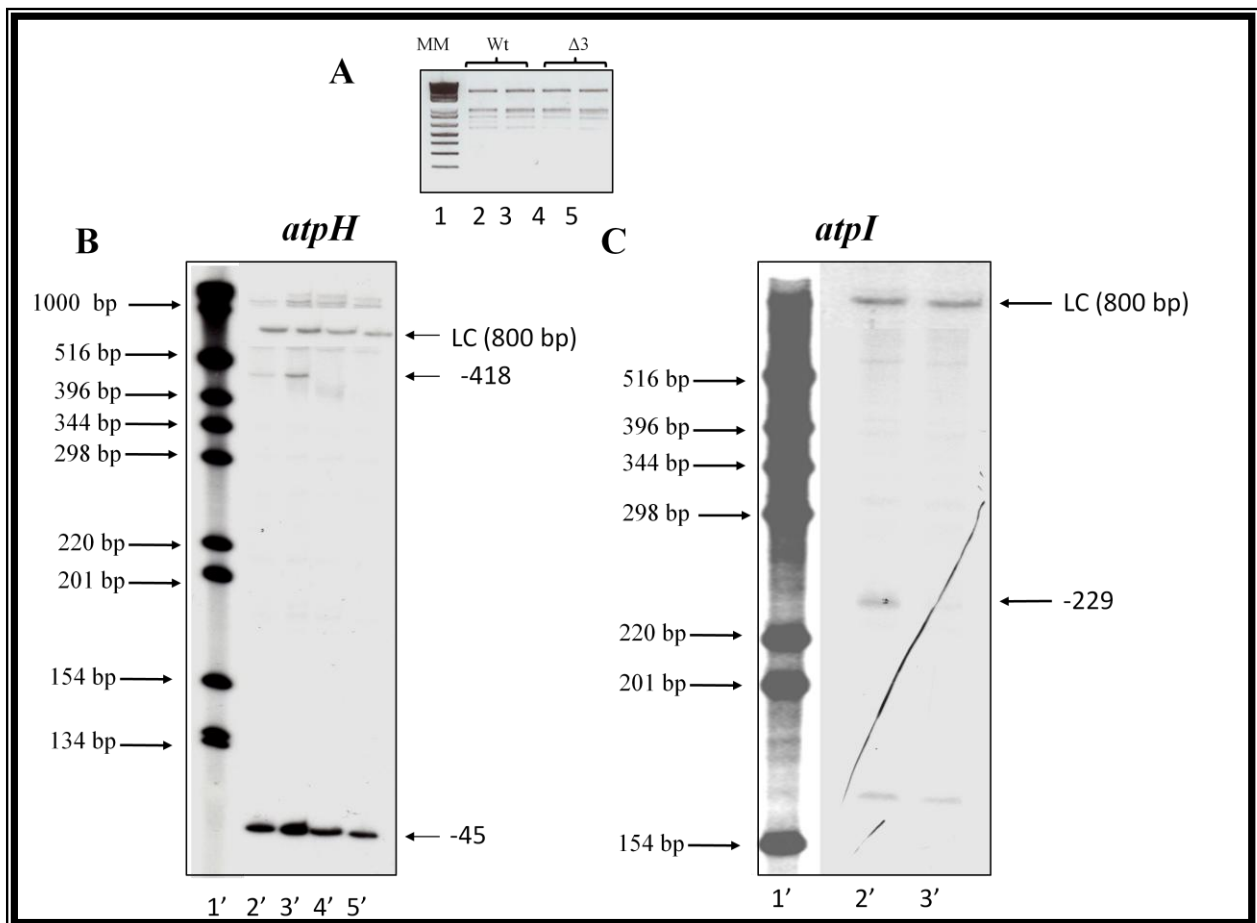
Figure 29. Primer extension analysis of *atpH* and *atpI* transcripts using 34 days old Arabidopsis plants grown in 50-60 µE in 16h/8h light cycle.
34 days old Wt (lanes 2 and 3) and *sig3* (lanes 4 and 5) plants were either kept under 50-60 µE for 2h30 (lanes 2, and 4, respectively) or exposed to 1300 µE for 2h30 (lanes 3, and 5 respectively). Total RNA was extracted and analysed either on a 1% agarose gel for quality control (A) or by primer extension to reveal *atpH* (A) or by primer extension to reveal *atpH* (B) and *atpI* (C) precursor RNAs.
LC indicates the loading control which corresponds to the PCR amplified product of 800 bp added to the mixture just before phenol chloroform extraction after cDNA synthesis.

In contrast to young green plantlets (Figs. 26-28) we found an increase in accumulation of both the -45 and the -418 transcripts of *atpH* when these plants were exposed

to 1300 μE (Fig. 29B, compare lanes 2' and 3'). Interestingly, the -45 transcripts remained unchanged in *sig3* plants (Fig. 29B compare lanes 4' and 5'). This indicates that the increase of -45 transcripts in WT plants after high light treatment is due to SIG3 initiated –418 transcripts. In order to exclude that we had also an increase in PEP SIG2 initiated *atpI* -229 transcripts in these conditions, we carried out primer extension analysis for *atpI* (Fig. 29C). Instead of an increase in transcript accumulation we observe a reduction/inhibition of the *atpI* mRNA level. This clearly shows that 1300 μE light exposure of 34 days old green plants for 2h30 specifically enhances the *sig3* dependent *atpH* -418 transcripts.

Conclusion:

In conclusion, the consequences of how photooxydative light influences the mRNA levels of plastid genes, depends highly on the physiological stage of the plant. For instance, short time (4h) exposure to high light (1300 μE) enhances the *atpH* mRNA level in etiolated plantlets, reduces the transcript accumulation of *atpH* in green plantlets and augments the *atpH* mRNA level in 32 days old mature plants. Only in mature plants the augmentation of *atpH* mRNA seems to be SIG3 dependent. In addition, this accumulation of transcripts in mature plants seems to be specific for SIG3 dependent genes because the accumulation of SIG2 dependent transcripts (*atpI* -229) is reduced in these plants (see Fig. 29). Thus, high light treatment of mature plants might represent a good model system to analyse the contribution of SIG3 dependent *atpH* transcription to ATP synthase stoichiometry by comparing for instance *atpH* mRNA and protein levels in young and mature plants (WT and *sig3*). Unfortunately, the experiment with the mature plant was made rather at the end of my thesis and I had no time any more to exploit this promising experimental system further.

Chapter 3

3. Expression analysis of the *psbT* sense/antisense RNAs.

SIG3-PEP holoenzyme transcribes specifically the plastid *atpH* and *psbN* genes. In the first chapter we have analysed the expression of the two ATPsythase operons in detail to get a better understanding of how the expression of all the ATPsythase genes is coordinated. Results did not give a clear answer to the question under which conditions SIG3 dependent *atpH* transcription is of special importance. In order to find conditions of high *atpH* (and also *psbN*) expression we have then analysed the influence of different light conditions on the plastid transcriptome. This was done by macroarray analyses (chapter 2). From the macroarray results, we could conclude on the expression of the ATPsynthase genes, but the values obtained for *psbT* sense/antisense RNAs have been too low to conclude on changes in gene expression. Our next idea to analyse *psbT* sense/antisense transcripts was then to use semi-quantitative RT-PCR, but this idea was dropped down by the appearance of a paper describing artefacts of RT-PCR analysis in case of sense/antisense transcripts (Haddad et al., 2007). To advance nevertheless in our characterization of *psbT* sense/antisense RNAs we decided therefore to map these transcripts carefully by 5'-RACE and circular RT-PCR to get an idea which types of RNA/RNA double strands could form. 5'-RACE and circular RT-PCR include systematic sequencing of the obtained amplification products, a fact that should exclude wrong interpretations due to artefacts in PCR amplification.

3.1 Previous results obtained in the laboratory.

Recently in the laboratory, analysis of the transcriptome in *sig3* mutant plants has shown that the expression of the *psbN* gene is strongly affected by the absence of the SIG3 factor (Zghidi et al., 2007). The *psbN* gene is located on the opposite strand in the inter-genic sequence between the *psbT* and *psbH* genes in the *psbB* operon. The *psbB* operon contains the genes *psbB, psbT, psbH, petB* and *petD*, coding for subunits of the photosystem II (*psb* genes) and of the cytochrome b6/f complex (*pet* genes). The gene order of the operon and the so far revealed transcripts of *psbT* sense and antisense RNAs (Zghidi et al., 2007 and unpublished results) are schematically demonstrated in Figure 30. It has been shown by primer extension that the *psbN* mRNA extends at its 3' end and covers the entire *psbT* gene as anti-sense RNA and 5' RACE experiments revealed that there is only one promoter for *psbN* (SIG3-PEP

P*psbN*), i. e. the *psbT* antisense transcript is produced as a result of processing of the *psbN*/*psbT* antisense transcript (Zghidi et al., 2007 and detailed in Fig. 30).
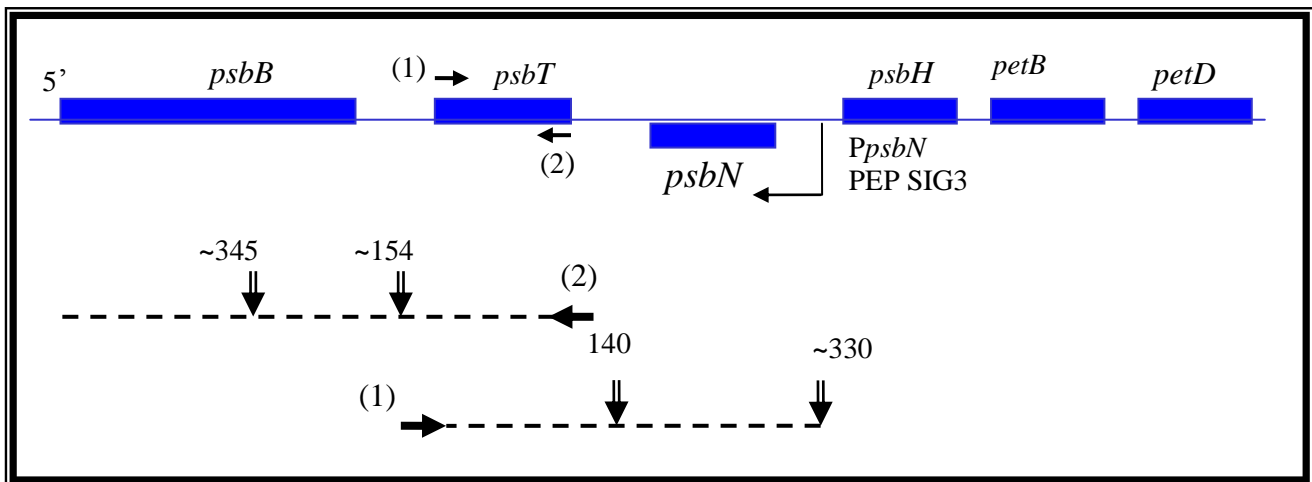
Figure 30. Schematic presentation of the *psbB* operon, the *psbN* gene and the *psbT* and *psbN* transcripts as described by Zghidi et al. (2007).

Oligonucleotides (1) and (2) used for two primer extension of *psbT* anti sense and *psbT* sense transcripts, respectively, are represented as horizontal arrows (↔). The *psbN* promoter is indicated by (⌐→) while processing sites are indicated by (⇓). The revealed sense and antisense transcripts are represented as dotted lines. The sizes of the cleavage products are represented by numbers.

Noteworthy, the question of whether PSBN protein belongs indeed to PSII is still controversial (Kashino et al., 2002). The product of the *psbN* gene could be a non-coding RNA.

In order to get more insight into the mechanisms of sense/antisense RNA expression, we decided to map the 5' and 3' extremities of all *psbT* and *psbN* sense and antisense transcripts (this work was performed in collaboration with Dr. Livia Merendino).

For this purpose, we performed circular RT PCR. In all these experiments we used self-ligated RNAs and RT reactions were performed by using specific oligonucleotides to discriminate between sense and antisense transcripts. Primers for RT and PCR reactions were designed in a way that the amplification product could include the ligated 5' and 3' extremities. In addition, by comparing TAP (Tobacco Acid Pyrophosphatase) treated to untreated RNAs, we differentiated the 5' ends of the primary and processed transcripts (for details see materials and methods). Controls without RT reaction were carried out for detection of DNA contamination. PCR products were then cloned and sequenced.

3.2 Mapping of *psbT* anti sense RNA extremities.

We first mapped the 5' and 3' extremities of *psbT* anti sense transcripts (Fig. 31).

A

RT   -   -   +   +
TAP -   +   -   +

← ~ 950 bps

1   2   3   4

B

PpsbB

psbT

(RT)

5'

psbB

(F)   (R)

634 bp

33 bp

psbN

PpsbN

3'

953 bps

3'   5'

psbT anti sense

C

ORF *psbT* sense

5' t**ATGGAAGCATTGGTTTATAC**ATTTC**TCTTAGTATCCACTTTAGGGAT**AA
TTTTTTTCGCTATTTTTTTTCGG**GAACCACCTAAAATTTCAAC**TAAAAAA
TGA **3'**

RT oligo   :   5' **ATGGAAGCATTGGTTTATAC**   **3'**
Forward oligo **:**   **5' ATCCCTAAAGTGGATACTAAGA 3'**
Reverse oligo **:**   **5' GAACCACCTAAAATTTCAAC**   **3'**

D

```
                        a
 uaacguaaucagccuccaa      u
 auugcauuagucggagguu      a
                        a
```

Figure 31. Mapping of 5' and 3' ends of *psbT* anti sense RNA.
(A) Circular RT PCR on anti sense transcripts of *psbT*. Lanes (1, 2) – RT control was carried out to check for DNA contamination. Lane (3) indicates RT PCR on TAP untreated circularised RNAs. Lane (4) indicates RT PCR on TAP treated circularised RNAs.
(B) *psbT* anti sense transcript with its 5' and 3' ends.
(C) Sequence of the coding region of *psbT* sense strand with the oligos used for cRT-PCR to map extremities of *psbT* anti sense transcript.
(D) Hairpin structure localized at the 3' end of *psbN* mRNA. The cleavage site that liberates *psbT* antisense RNA is indicated in a bold letter.

Due to difficulties in the growth of colonies having the RT-PCR product as insert, we directly sequenced the product of circular RT-PCR along with a few colonies. We found that both the 5' and 3' end of the antisense transcript of *psbT* is well defined. The 5' end maps in the intergenic region between *psbN* and *psbT* genes, 33 nucleotides downstream to the stop codon (UAG) of *psbN* sense RNA. This corresponds to the band of 140 bps in primer extension analysis of Zghidi et al., (2007) (see also Fig. 30). The 3' end maps 634 nucleotides upstream to the stop codon inside the *psbB* coding region but on the opposite strand.

By comparing TAP treated and untreated RNAs (Fig. 31A, lanes 3 and 4) we confirmed that the 5' end of the *psbT* anti sense RNA was obtained by processing of *psbN* sense/ *psbT* antisense transcript (Zghidi et al., 2007). Indeed, we observe that the 5' end maps inside a solid hairpin (Fig. 31D, in bold is indicated the 5' end position) just downstream to the *psbN* coding region.

3.3 Mapping of *psbT* sense RNA extremities.

We then mapped the extremities of *psbT* sense transcripts (Fig. 32). Our results show the existence of multiple 5' ends (as indicated by the smeary appearance of the RT-PCR product, Fig. 32A). Most of the 5' extremities were dispersed in the coding region of *psbB* while a few were found in the intergenic region between *psbB* and *psbT*, upstream to the processing site at position -18 from the *psbT* ATG codon found by Zghidi et al. (2007). Conversely, the 3' end was well defined in a range of 8 nucleotides starting 60 base pairs downstream to the *psbT* stop codon in the intergenic region between *psbT* and *psbN*. By comparing TAP treated to TAP untreated RNAs, no primary transcripts could be detected.
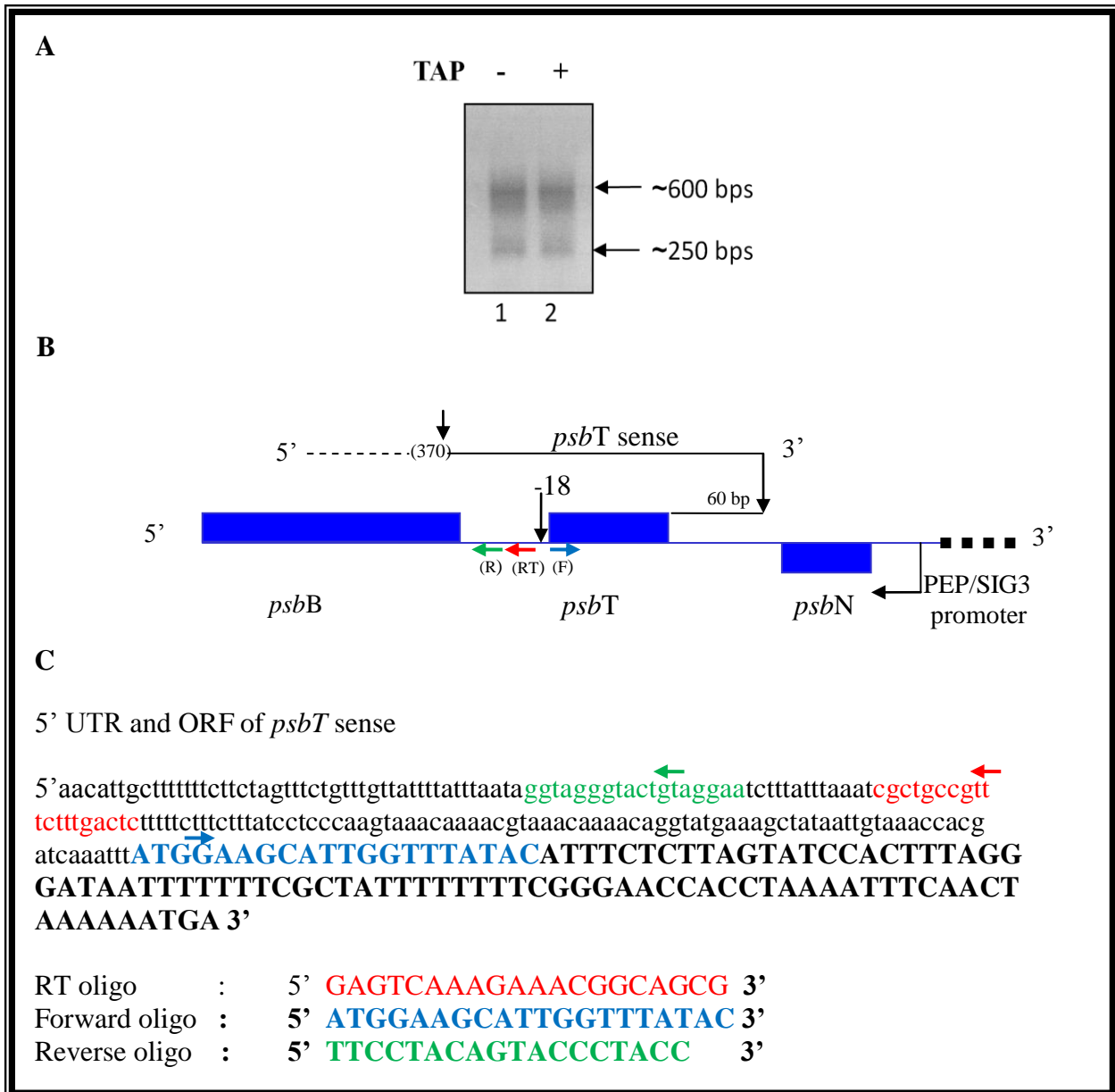
Figure 32. Mapping of 5' and 3' ends of *psbT* sense RNA by circular RT PCR.
(A) Lane (1) indicates RT PCR on TAP untreated circularised RNAs. Lane (2) indicates RT PCR on TAP treated circularised RNAs.
(B) *psbT* sense transcript with its 5' and 3' ends. The processing site at -18 from *psbT* ATG is indicated shown by Zghidi et al., (2007).
(C) Sequence of the coding region of *psbT* gene and its 5' UTR with the oligos used for cRT-PCR to map extremities of *psbT* sense treanscript.

In order to determine if the *psbT* sense RNA also extends its 3' end over the *psbN* gene to produce a *psbN* antisense RNA, we performed a new RT-PCR (Fig. 33A and B) with a more downstream forward oligonucleotide. Reverse and forward oligonucleotides were located in the ORFs of *psbT* and *psbN* (on the opposite strand) respectively (Fig.33B and C). We found multiple 5' ends that were clustered in two regions, upstream (15 clones) and downstream (11 clones) to the processing site at position -18 from the ATG translation

initiation codon of *psbT* (indicated as dotted line in Fig. 33B). The -18 site corresponds to the position where the ~154 band found by Zghidi et al. (2007) maps (see also Fig. 30). We found no clone that mapped exactly at -18 or -28 as described by Zghidi et al. (2007) (thesis of Wafa Zghidi, 2008). Conversely, the 3' end is very well defined and located 25 nucleotides upstream to the *psbN* ATG codon but in the opposite strand (Fig. 33B). If the -18 RNA is completely hybridized with antisens RNA it could not be revealed by circularization.

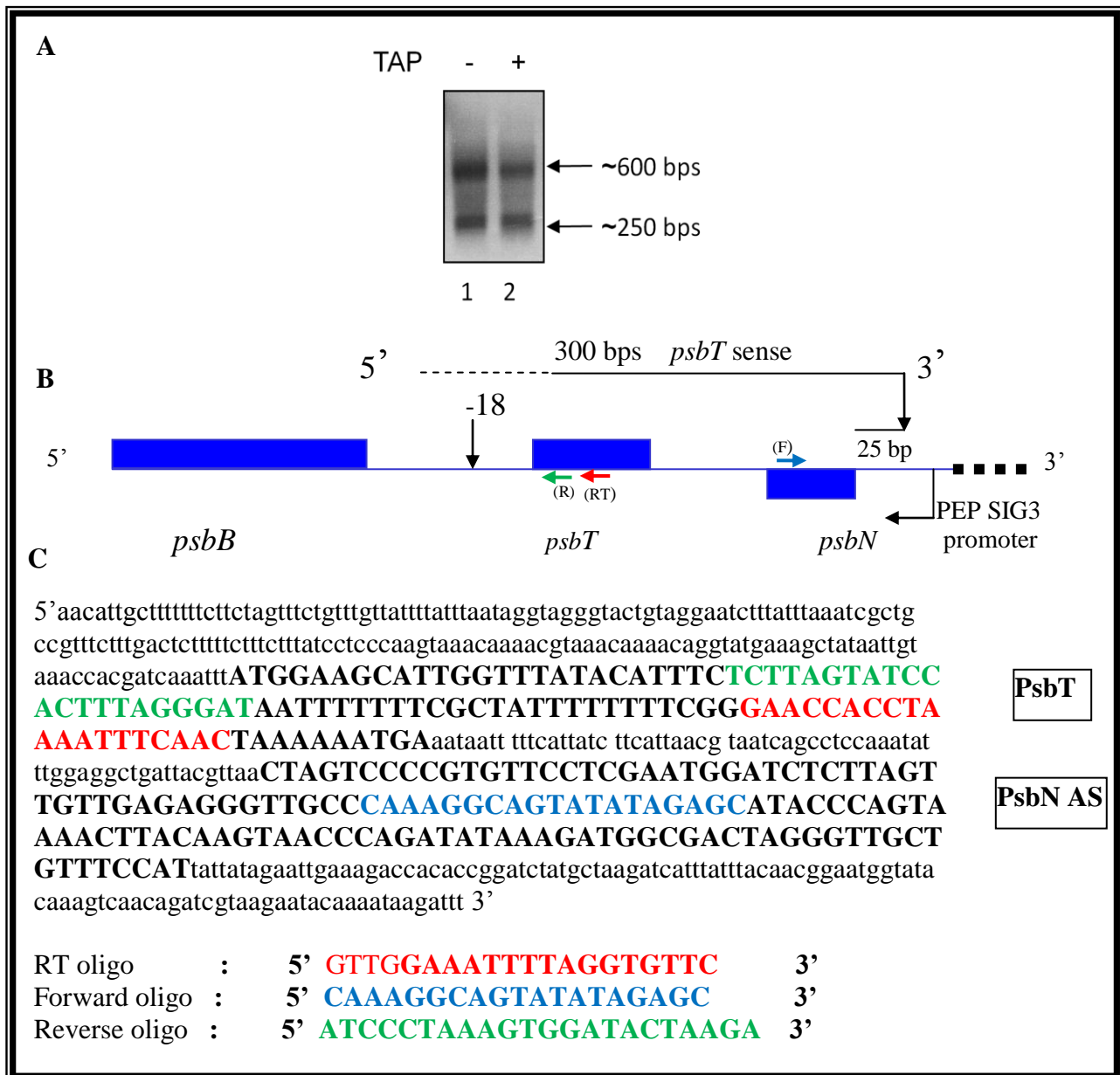**Figure 33. Mapping of 5' and 3' ends of *psbN* anti sense RNA by c**ircular RT PCR on sense transcripts of *psbT*.
(A) Lane (1) indicates RT PCR on TAP untreated circularised RNAs. Lane (2) indicates RT PCR on TAP treated circularised RNAs.
(B) Scheme of the *psbT* sense transcripts with its 5' and 3' ends.
(C) Sequence of the coding region of *psbT* gene with the oligos used for cRT-PCR.

The results of the above two experiments indicate that there are two types of *psbT* sense transcripts both having dispersed 5' ends and different, but stable 3' ends. One type of transcripts (Fig. 32B) covers only the *psbT* gene while the other one (Fig. 33B) covers part of the *psbT* gene and the entire *psbN* gene as anti sense RNA. Results show the existence of a cotranscript *psbT/psbN* antisense RNA whose role has not yet been investigated. Indeed, the 3' end of the monocistronic *psbT* sense RNA and the 5' end of the *psbN* antisense RNA map both in a solid hairpin (Fig. 31D shows the hairpin for the antisense RNA). In conclusion, the results obtained during my thesis by circular RT PCR in combination with previous analysis by primer extension (Zghidi et al., 2007) and TAP (5' RACE) indicate that the antisense transcripts of *psbT* overlap the entire *psbT* sense RNA including the coding region and both 5' and 3' UTRs. Overlapping of the *psbT* antisense transcript to the 5' UTR is particularly interesting, as in chloroplast transcripts the 5' UTR contains the signals to recruit the translation initiation complex. Thus, duplex formation of sense and antisense transcripts in the 5' UTR should prevent the ribosome recruitment suggesting the existence of a translational regulation mechanism by antisense RNAs.

3.4 Putative role of *psbN* expression on processing of *psbB* operon.

A primer extension analysis on *psbH* transcripts indicates that a processing event occurs at position -79 from *psbH* ATG in (Fig. 34, lane 5, compare also Fig. 30 for localization of *psbH).* The same position was found by 3'-end mapping of *psbT* sense RNA by circular RT PCR (Fig. 33). Notably, the 5' end of the transcript that is synthesized from the PEP/SIG3 promoter of *psbN* overlaps with a small region of the 5' UTR of *psbH* transcripts. Interestingly in the absence of SIG3, the -79 *psbH* RNA was not detected by primer extension (Fig. 34 compare lane 5 to 6).

These data suggest that duplex formation of sense and antisense RNAs may be required for this processing event. However, this difference in *psbT/psbH* intergenic processing events does not change the translation competence of the monocistronic *psbH* mRNA because the PSBH protein level is not diminished in *SIG3* knock-out mutants (Zghidi et al., data not shown). Westhoff & Herrmann found two 5' ends (42 and 72 bps upstream to the translation initiation codon of *psbH*) for the *psbH* transcripts in spinach by S1 nuclease protection assays and northern blot analysis (Westhoff & Herrmann, 1988). Our primer extension analysis of Arabidopsis Wt confirmed the presence of these two processing sites showing conservation of processing sites between spinach and Arabidopsis (Fig. 34).

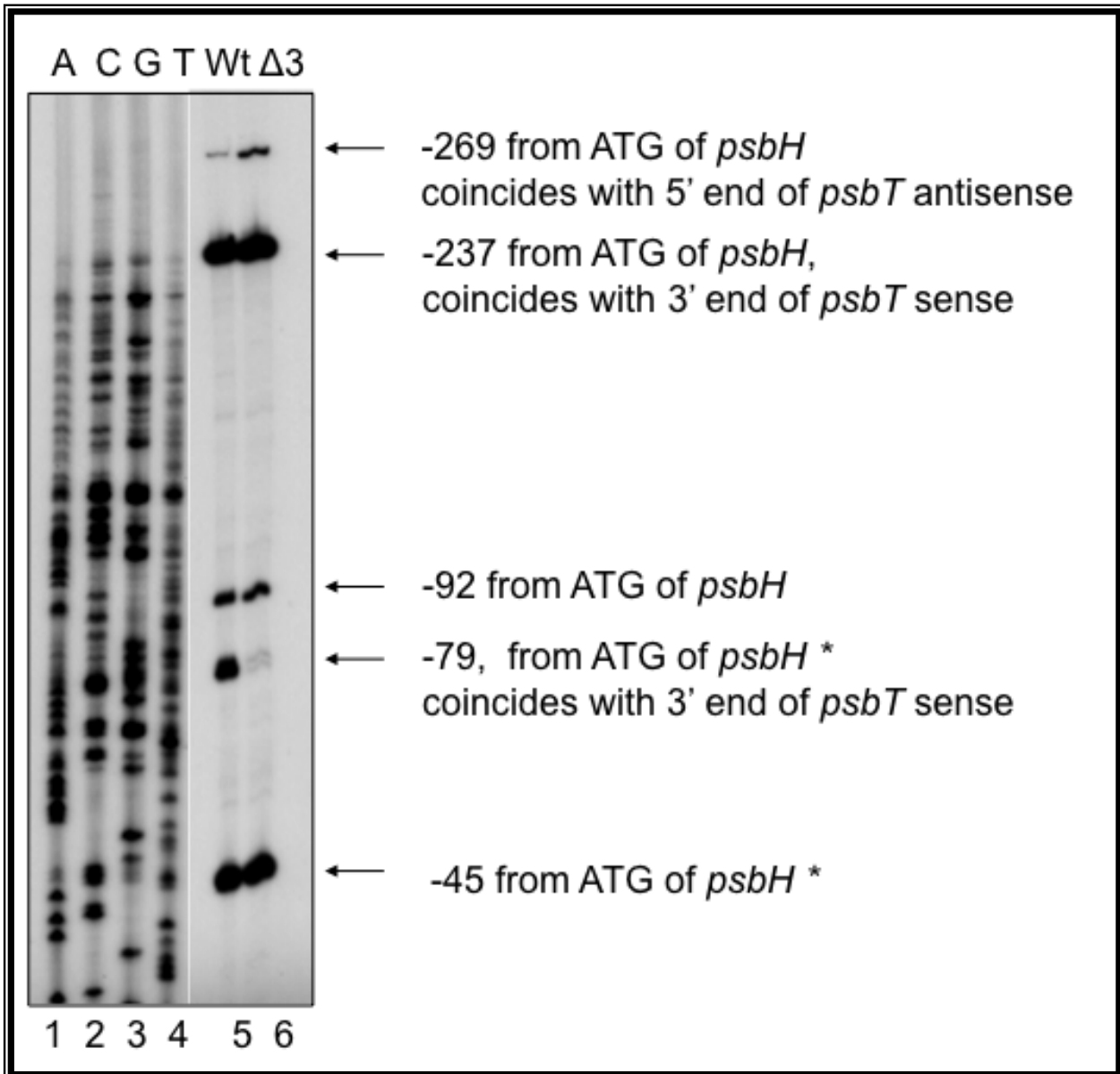**Figure 34. Primer Extension analysis of *psbH* transcripts.**
5 µg of total RNA extracted from 6 days old plants of wild type Columbia (lane 5) and *sig3* (lane 6) plants have been used for analysis. Lanes 1-4 indicate the sequence used to map the 5' extremities of the transcripts. Sequence was prepared with the same oligo used for primer extension. * These transcripts were also described by Westhoff and Herrmann (1988).

A

+RT   wt

TAP   -   +

← ~700/750 bps

1   2

B

-860

-18

PCR2

RT, PCR1

psbB

psbT

psbN

3'

PEP SIG3 promoter

C

5' UTR and ORF of *psbT*

5'aacattgctttttttcttctagtttctgtttgttatttt atttaataggtagggtactgtaggaatctttatttaaat cgctgccgtttctttgactcttttttcttctttatcctcccaagtaaacaaaacgtaaacaaaacaggtatgaaagctataattgtaaaccacgatcaaattt **ATG GAAGCATTGGTTTATACATTTCTCTTAGTATCCACTTTAGGGATAATTTTTT TCGCTATTTTTTTTCGGGAACCACCTAAAATTTCAACTAAAAAATGA 3'**

RT oligo	:	5' **GAAATTTTAGGTGGTTC**	**3'**
PCR 1	:	5' **GAAATTTTAGGTGGTTC**	**3'**
PCR 2	:	**5'** GAGTCAAAGAAACGGCAGCG	**3'**

Figure 35. A: 5' RACE on sense transcripts of *psbT*.
(A) Circular RT PCR. Lane (1) indicates RT PCR on TAP untreated circularised RNAs. Lane (2) indicates RT-PCR on TAP treated RNAs.
(B) Summary of *psbT* and *psbN* sense/anti sense transcripts with their 5' and 3' ends.
(C) Sequence of the coding region of *psbT* gene and its 5' UTR with the oligos used for RT and the two PCR reactions (PCR1, PCR2) to map 5' end of *psbT* transcript.

3.5 Existence of an internal *psbT* promoter within *psbB* gene.

Primer extension analysis of *psbT* transcripts showed a transcript starting with in the coding region of *psbB* (see Fig. 30, ~345). Up to now, we could not find a clearly defined transcript by circular RT-PCR (see Fig. 32). Therefore, we performed a 5' RACE analysis by using an oligonucleotide for PCR2 located in the intergenic region of *psbB-psbT* upstream to

the -18 position (Fig. 35B, C). The PCR reaction revealed multiple transcripts indicating extensive instability of the transcripts (Fig. 35A). However a single band present only in the TAP treated samples when compared to untreated samples showed the presence of a primary transcript (Fig. 35A, ~700/750). Cloning and sequencing of the corresponding RT-PCR product (~700 bps) showed that the 5' end maps in the *psbB* coding region at position -860 from the AUG translation initiation codon of *psbT* and 674 bp upstream from the *psbB* stop codon. We conclude that transcription of *psbT* can be uncoupled from *psbB* by a *psbB* internal promoter.

The summary of the promoters, processing sites and extremities of *psbT/psbN* sense and antisense transcripts and the number of clones found for the 5' and 3'ends of *psbT* sense transcript are illustrated in Fig. 36. We see that the -860 initiated sense RNA overlapps the 3' end of the long antisense RNA. The long sense RNA seems to be rapidly degraded from its 5' end while the 3' ends as well as the antisense RNA seems to be stable.
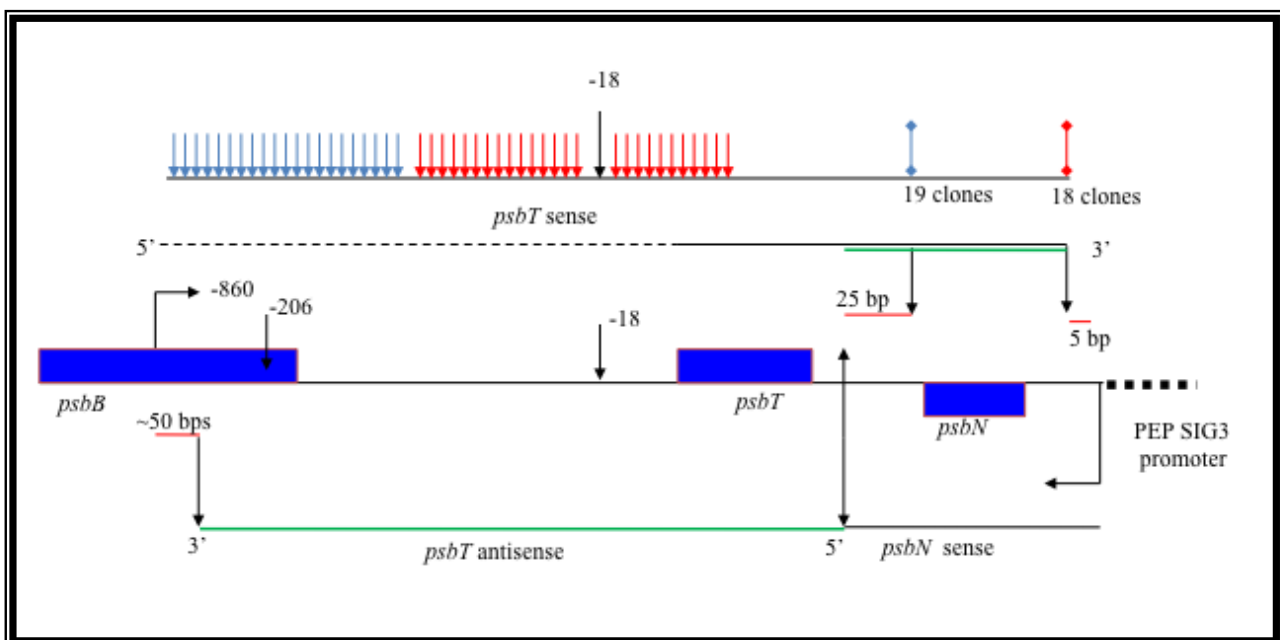

Figure 36. Summary of extremities of *psbT/psbN* sense and antisense transcripts.
↓↓ Indicate the 5' ends while ↕↕ indicate the 3' ends of *psbT* sense transcripts. The number of sequenced clones is given under the 3' end indicating vertical arrows. Green lines show the regions where perfect RNA/RNA double-strand formation is possible while the red lines indicate overhanging regions.

Conclusion:

Taken together with the unpublished results of Zghidi et al., we can conclude on the existence of two different *psbT* antisense RNAs, one covering only *psbT* mRNA and another one covering *psbT* and a part of *psbB* mRNA. *PsbN* transcription seems to be necessary for cleavage of the *psbB* multicistronic mRNA at the position -79 from the *psbH* ATG codon (Fig. 34), but this cleavage is not important for the PSBH protein level. *PsbT* sense RNA can be initiated at the *psbB* promoter of the operon or at a *psbB* internal promoter (-860, Fig. 36). Primer extension and 5'-RACE analyses show the existence of still another *psbT* specific promoter, localized in the intergenic region between *psbB* and *psbT* (Zghidi et al., unpublished result) and cloning of the *psbT* mRNA 5' and 3' ends by circular RT-PCR reveals many cDNAs (Fig. 32A) indicating rapid 5' end degradation of *psbT* mRNAs. Surprisingly, the 3' ends of the mRNAs are rather stable. Sense and antisense RNAs will probably form RNA/RNA double strands and it is perhaps the strong sensitivity of the *psbT* sense RNA to 5' end nuclease digestion that makes double-strand formation necessary to protect *psbT* mRNA from rapid degradation.

Results obtained by performing RT-PCR reactions with RNAs that were previously treated with low RNase A/T1 concentrations show that the *psbT* sense and antisense RNAs exist *in vivo* in a stable double stranded RNA form (Zghidi et al., unpublished results) and in *sig3* plants where the *psbT* antisense RNA is absent, the level of PSBT protein is highly increased (Zghidi et al., unpublished results). Thus, by forming RNA/RNA hybrids, antisense RNA might regulate the expression of the *psbT* gene at the level of translation and/or by nucleolytic protection of the sense RNA. Both mechanisms are not exclusive but the final understanding of sense/antisense regulation needs still further analyses.

General Discussion

Expression analysis of the two plastid encoded ATPsynthase operons: the large *ATPI/H/F/A* and the small *ATPB/E* operon.

We have analyzed the expression of the two plastid encoded ATPsynthase operons, *atpI/atpH/atpF/atpA* and *atpB/atpE* in WT plants (6 days old) of *Arabidopsis thaliana* by macroarray analyses (Lambert, Malik Ghulam and Lerbs-Mache, unpublished result). These analyses showed a much higher level (about 16 folds) of *atpH* transcripts when compared with the other plastid ATP mRNAs (see table 1). The *atpH, I* and *F* genes code for subunits III, I and IV of the CF0 complex. Our macroarray results show that the observed protein stoichiometry of the CF0 complex, defined as 1:1:6-14:1 for subunits I, II, III and IV respectively, is also approximately found on the mRNA level in *Arabidopsis.* In some cases, the accumulation of *atpH* mRNA versus the other mRNAs is even higher. This suggests a specific regulation of *atpH* gene expression when compared to the other plastid genes that encode subunits of ATPsynthase and are in the same transcriptional unit.

A SIG3 specific promoter had been identified at position -418 upstream of the *atpH* translation start codon (Zghidi et al. 2007) and when I started my thesis, we thought that the high level of *atpH* mRNA is achieved by additional transcription of the *atpH* gene by SIG3/PEP holoenzyme at this promoter. Transcription from this promoter could decouple transcription of *atpH* and/or *atpH/atpF/atpA* from co-transcription with the first gene of the operon, i. e. *atpI*. Furthermore, with the exception of the *atpI* mRNAs, all ATPsynthase mRNAs were diminished in *SIG3* mutants when compared to WT plants (Zghidi et al., 2007). These results pointed to a specific regulation of the two ATPsythase operons by sigma factor 3 and one of the aims of my work was to elucidate this putative role of SIG3 in the expression of the two ATPsynthase operons in detail.

For this aim, during my PhD work I have analyzed the expression of each gene of the two plastid encoded ATPsynthase operons independently in WT plants and in all sigma factor mutants (i.e. *sig1 − sig6*) of *Arabidopsis thaliana*. Results show further that the large ATPsynthase operon is transcribed from two different PEP promoters, one is under control of SIG2 (*atpI*-229/-225, Figure 8) and the other promoter (*atpH*-418) is specifically recognized by a PEP/SIG3 holoenzyme (Figure 9). At least some of the SIG3/PEP initiated *atpH* transcripts are part of polycistronic mRNAs that cover the *atpH, atpF* and *atpA* reading frames (Figs. 14 and 16). Both of the *atpH* precursor RNAs, initiated either at the -229 *atpI* promoter or at the -418 *atpH* promoter, are cleaved at position -45 from the *atpH* translation start codon. This is shown by the presence of the -45 transcript in both, the *sig2* and the *sig3,*

mutants (Fig. 9). In the absence of SIG2 the -229 *atpI* transcript disappears (Figure 8C, compare lanes 2' and 4') but the -418 *atpH* mRNA accumulates slightly (Figure 9B, compare lanes 2' and 4'). This suggests a higher activity of SIG3 in *sig2* plants, a result that might reflect the raise of SIG3 protein in SIG2 under-expressing plants as already previously reported (Privat et. al., 2003). This hypothesis is also supported by our Northern analysis of SIG3 dependent *atpH* mRNA (Fig. 12A, -418 specific probe) that shows an enhancement of *atpH* mRNA in the *sig2* mutant when compared to WT plants.

The -45 *atpH* processing intermediate should be more stable than the -418 primary transcript as suggested from the quantitative difference between these two transcripts (Fig. 9B). Specific stabilization of *atpH* mRNA by PPR10 protein has recently been described in maize (Pfalz et al., 2009). Interestingly, the stabilizing PPR10 protein was shown to bind to the 5' terminus of the *atpH* mRNA in the region between -46 and -20 upstream of the translation initiation codon, *i. e*. it should only protect the -45 *atpH* mRNA, but not the -418 mRNA. Our analyses to determine 5' and 3'ends of *atpH* mRNA termini (Fig. 10) are in good agreement with such an interpretation. The complexity of the mapped -418 *atpH* mRNA 3' termini suggests that -418 *atpH* initiated transcripts are rapidly degraded from their 3' end. The -45 *atpH* mRNA 3' ends that have been mapped just downstream to the *atpH* coding sequence are less dispersed indicating that this RNA is more stable than the -418 initiated RNA. Thus, the stabilization of the -45 mono-cistronic transcript should be mainly responsible for the high RNA level of the *atpH* mRNA that is observed by microarray analysis. During my thesis, I could not clearly define conditions under which either stabilization of *atpH* mRNA or SIG3-dependent transcription of the *atpH* gene is more important to reach the high *atpH* mRNA level.

The results of the primer extension and TAP RT PCR experiments indicate that *atpF* and *atpA* genes are co-transcribed with *atpI* and *atpH* (Figs. 14, 15 and 16). In the case of *atpF*, long transcripts are observed that should correspond to *atpH/AtpF* cotranscripts starting either at the - 418 SIG3 dependent promoter or at the -45 *atpH* processing site (Fig. 14B). A large part of the -30 *atpF* transcript that has been mapped by primer extension should arise from cleavage of the -418 initiated *atpH/atpF* co-transcript. This can be concluded because this transcript is diminished in *sig3* plants (Fig. 14C). Cleavage of double stranded RNA, *e. g.* cleavage in each strand of a hairpin stem, is made by RNase III in *E. coli* (Pertzev & Nicholson, 2006). The *Arabidopsis* genome encodes four RNAse III like proteins RTL2, RNC1, At1g55140 and At3g13740. Such type of enzyme could cleave on both sites of the hairpin structure containing the -30 *atpF* site and liberate an *atpH* mRNA ending with a

poly(A) tail. This RNA, in turn, could be substrate to immediate 3' exonuclease degradation, the reason why we could not find 3'ends corresponding to this site by ligation mediated 3'-5' cloning and sequencing. Our suggestion that the *atpH/atpF* co-transcripts might be substrates for RNase III cleavage is further supported by the fact that the sequence of the double helical structure of the -30 *atpF* hairpin is identical to the proximal box (pb) of R1.1[WC] RNA which has been shown to be important for cleavage by RNase III enzymes of different origin (Meng et al., 2008). For future experiments, it would be interesting to analyze the stability of *atpH* transcripts in RNase III knock-out plants. On the other hand, a 3'end localized within a hairpin structures might be difficult or impossible to ligate providing another reason why we could not find 3'of *atpH* transcripts ending at the -30 *atpF* site by circular RT PCR. Lack of ligation might also be the reason why we could not find 5' ends of the *atpF* mRNA corresponding to the predicted termination structure by 5' RACE.

The small ATPsynthase operon, consisting of the *atpB* and *atpE* genes, is transcribed from two *atpB* PEP promoters in young *Arabidopsis* plantlets, initiating transcription at positions -520 and -467 relative to the ATG codon (Fig. 17). The -467 promoter has not yet been described before. The -520 promoter is preferentially recognized by SIG6/PEP holoenzyme (Fig.17, compare lanes 11 and 12) and the -467 promoter is recognized by SIG2/PEP holoenzyme (Fig. 17C, compare lanes 4 and 6). Transcription of the small ATPsythase operon is also dependent on SIG1 and SIG3, although we do not know exactly which of the promoters is recognized by these factors. In both mutants, the -84 transcript level is diminished without considerable changes in the -520 and -467 precursor transcripts (Fig. 17B, compare lane 2 to lanes 3 and 5). The promoter recognized by SIG1 and SIG2 might correspond to the 4.8 kb transcript that is initiated far upstream of the *atpB* gene (Schweer et al., 2006). Altogether, *atpB* transcription is very complex implicating at least 4 of the 6 plastid sigma factors and we cannot discern a specific role for SIG3 in this interplay.

Transcription from these two *atpB* promoters results probably in a dicistronic RNA of about 2.6 kb that is cleaved at position -84 to produce a 2 kb transcript (Fig. 17A). Transcription of the *atpE* gene can be decoupled from transcription of *atpB* by usage of an *atpB* internal PEP promoter that is recognized by SIG2 (Fig. 18). The part of the SIG3 initiated transcripts in the total amount of *atpB/atpE* transcripts is low and concerns only the -84 RNA (Fig. 19). We could not detect changes of ATPE and ATPB proteins in *sig3* plants by Western immunoblotting but we observed a reduction (not absence) of ATPE in *sig2* mutant plants (Fig. 19C and not shown). Interestingly, the existence of ATPE protein in *sig2* plants (where the *atpE* monocistronic RNA is absent) shows that *atpE* mRNA is translated

from the dicistronic mRNA. This observation is particularly surprising because the *atpE* start codon is localized upstream to the *atpB* stop codon, within the *atpB* coding region (Fig. 19B). A stop codon dependent mechanism of translation of the downstream overlapping transcript of a dicistronic mRNA has recently been described for *ndhC-ndhK* using an *in vitro* chloroplast translation system (Yukawa and Sugiura, 2008). Our result suggests that this mechanism should also exist in chloroplast *in vivo*. Altogether, our results indicate transcriptional as well as post-transcriptional (stability) regulation of the expression of chloroplast ATP synthase genes, in contrast to Chlamydomonas where regulation occurs especially at the translational level. In *Arabidopsis*, all genes coding subunits of the ATP synthase could be regulated at the transcriptional level by SIG2 and SIG3. The only exception is the *atpI* gene that is under control of SIG2 only and does not harbor a SIG3 dependent promoter. Coordinated transcriptional regulation of all plastid ATPsynthase genes is probably necessary to keep stoichiometry of the mRNAs. The SIG2 initiated transcripts are generally more abundant than SIG3 initiated transcripts. This is not surprising since SIG2 is considered as one of the principal sigma factors in chloroplasts. However, this does not exclude that under specific conditions modulation of transcript levels by SIG3 becomes very important.

Influence of light on the expression of SIG3 dependent transcripts.

From our own analyses on *atpH* gene transcription in *Arabidopsis* (chapter I) and from the recent finding on *atpH* mRNA stabilisation by PPR10 in maize (Pfalz et al., 2008) we suggest that *atpH* gene expression in *Arabidopsis* is regulated at least in two different ways:

(1) Transcription of the *atpH* gene can be decoupled from the transcription of the *atpI* gene by an *atpH* specific promoter, localized 418 bp upstream of the *atpH* coding region and specifically recognized by SIG3/PEP holoenzyme.

(2) *AtpH* mRNA can be specifically stabilized by a PPR10 like protein.

As indicated in the first chapter of my thesis, the quantity of SIG3 dependent *atpH* mRNA varied in different mRNA preparations, although plants had been grown under the same conditions, i. e. in the same growth chamber under the same light cycle. This was difficult to understand for us and at first we thought that these differences might be due to changes in diurnal expression of the *atpH* gene, because harvesting of plant material was not always done at the same time on the day. This idea was not confirmed by the analysis of *atpH* transcripts from plants harvested in 6h intervals during one circadian cycle (Fig. 20).

The next hypothesis that we tested was a specific influence on SIG3 dependent gene expression during light stress or in relation to changes in light intensities. This hypothesis was

supported by the supposed "rescue" function of SIG3 (see introduction) that might include rescue during and/or after photooxydative conditions, by the fact that the function of the ATPsynthase is related to photosynthesis and *atpH* expression should be influenced by different light conditions and finally by the importance of the PSBT protein for recovery of PSII after light stress (further supposed that *psbT* gene expression is influenced by SIG3 dependentt *psbN* transcription and antisense RNA production). For these reasons, we next analysed the mRNA levels of *psbN*, of *atpH* and of all other ATPsynthase genes in plantlets that were grown under different light conditions. Especially, we were looking for conditions in which *atpH* mRNA levels are highest in relation to the other ATPsynthase mRNAs and also in relation to all other plastid mRNAs. Experiments were carried out by plastid transcriptome analyses using a plastid specific macroarray, a technique that allows analysing all plastid mRNAs at once. We also wanted to analyse the relation of *psbT* sense/antisense RNAs and *psbN* sense/antisense RNAs. However, macroarray values for these RNAs were in general very low and do not permit to draw solid conclusions.

As expected, macroarray results show that the *atpH* mRNA level is highly influenced by different light conditions. Highest *atpH* mRNA levels are obtained when etiolated plantlets were exposed to 1300 µE for 4h or to 50 and 200 µE for 16h (Figs. 23 and 24). Under short time light exposure (4h) mRNA levels of both genes, *atpH* and *psbN*, increase with increasing light intensity. Under long time light exposure (16h) mRNA levels of both genes increase when low (50 µE) or moderate (200 µE) light intensities are applied, however they fall under dark levels when etiolated plantlets were exposed to high photooxydative light (1300 µE, Fig. 24). Thus, it seems that both SIG3 dependent mRNAs behave in the same way in response to light. However, it has to be mentioned again that values for *psbN* mRNA are very low and to consider with care.

When *atpH* mRNAs were compared to the other mRNAs of all plastid encoded ATPsynthase genes we found that the relation was rather constant, indicating about 7 fold higher *atpH* mRNA levels after 4h of light exposure and about 16 fold higher *atpH* mRNA levels after 16h light exposure (Table I). Interestingly, after exposure to 1300 µE photooxydative light *atpH* mRNA levels are significantly enhanced when compared to all other ATPsynthase mRNAs (Table I) indicating another behavior of *atpH* mRNAs under stress conditions. The question of whether this higher level of *atpH* mRNA relies on specific transcription from the -418 *atpH* promoter by SIG3/PEP holoenzyme or by specific stabilization of *atpH* mRNA by a PPR10 like protein remains to be elucidated in the future. In order to distinguish between these two possibilities run-on transcription experiments need to

be performed. This method has not yet been applied to macroarray analyses and I had not enough time to elaborate this method.

A new and interesting observation coming out from these experiments was the general and strong diminution of the levels of rather all mRNAs after long time high light exposure (Fig. 23, compare E with H). Therefore, this phenomenon has been further analysed by using also green plantlets that had been obtained by growing under 50-60 µE light intensity in light/dark cycle and that were then exposed to 1300 µE for 4, 8 and 16h (Fig. 26). Our guiding idea was that high light might be more dangerous to etiolated tissues in which photosystems are not yet established than to green tissues. To our surprise, the result of the experiment showed the contrary (Fig. 26). In green plantlets, the *atpH* mRNA level is already diminished after only 4h of high light exposure. The same light sensitivity was found for other mRNAs like *atpI* and *clpP* (Fig. 27) and also for ribosomal precursor RNAs (Fig. 28). The only exception that we found was the *psbT* antisense RNA. This RNA was not diminished after high light exposure (not shown). This fact was of special interest because it points to a protection of RNAs against high light induced degradation by RNA/RNA double strand formation. This result provided also one of the reasons that guided us to characterize the *psbT* sense/antisense RNAs in great detail (see chapter 3).

However, before treating the problem of sense/antisense RNAs, I want to finish the problem of (single-stranded) mRNA degradation under light stress. Having realized the differences in mRNA behavior in etiolated and green plantlets that are both at the cotyledon stage (Figs. 26 and 27), we also wanted to know what happens in green leaves. Therefore, we have also analyzed at least two mRNAs, *atpH* and *atpI,* by using mature *Arabidopsis* plants (Fig. 29). Again, we observed a different behavior of mRNA levels. While the *atpH* mRNA level increases after high light treatment (Fig. 29A) the *atpI* mRNA level decreases (Fig. 29B). It is of special interest to notice that the increase in the *atpH* mRNA level is due to transcription by SIG3/PEP. This can be concluded from the fact that both the -45 and the -418 *atpH* mRNA precursors augment after high light treatment in wild type plants and the -45 transcript does not change in sig3 plants (Fig. 29A).

In conclusion, we observed different behaviors of mRNA levels in response to high light when analyzing RNAs prepared from either 5 days old etiolated plantlets, 7 days old green plantlets or mature plants. Our results indicate that mRNA metabolism varies strongly dependent on the physiological stage of the plants. In consequence, the answer to the question of which mechanism is more important to reach a high level of *atpH* mRNA, SIG3 dependent transcription or RRP10 dependent mRNA stabilization, will strongly depend on the

physiological stage of the plant and on the growth conditions, probably including yet other parameters than only light that has been analyzed here.

Expression analysis of the *psbT* sense/antisense RNAs.

Many different regulatory pathways mediated by antisense RNA base pairing have been discovered during the last years in eukaryots as well as in bacteria (Lapidot & Pilpel, 2006; Repoila & Darfeuille, 2009; Werner & Swan, 2010). Much less is known on regulatory antisense mechanisms in organelles like mitochondria and chloroplasts. So far, two RNAs oriented in antisense direction to known plastid genes have been revealed by cloning of small nc RNAs (Lung et al., 2006) and a long antisense RNA, complementary to the reading frame of the *ndh*B gene but starting within the reading frame of the *ndh*B, has recently been transcribed (Georg et al., 2010). Until now, no function has been attributed to these RNAs.

A special case of antisense RNA had been observed in our Laboratory where it was shown that in *Arabidopsis* the transcription initiation factor SIGMA 3 is responsible for production of *psbT* antisense RNA by specific transcription of the *psbN* gene (Zghidi et al., 2007). This *psbT* antisense RNA seemed to us of particular interest because it was shown to cover the entire *psbT* coding region. In bacteria, ncRNAs generally base pair to mRNAs in their 5' or 3' UTR and RNA duplex formation modifies translation efficiency and/or stability of the corresponding RNA. Transcription of entire genes in sense and antisense orientation has been described only for RNA polymerase II where it was shown that polymerases collide and stall each other on the DNA template or that transcription from a strong promoter hinders the transcription from a weaker promoter. This mechanism is known as "transcriptional interference" (Hongay et al., 2006; Yazgan & Krebs, 2007). Such a mechanisms does not seem to hold true for the *psb*T gene because the absence of *psbN* transcription in *sig3* plants does not change the *psb*T transcript levels (Zghidi et al., 2007). Therefore, we assume that *psb*T sense/antisense RNA transcription is connected to yet another mechanism of regulation.

In order to better understand the mechanism by which the antisense RNAs regulate *psbT* gene expression, we mapped the 5' and 3' extremities of the *psbT* and *psbN* sense and antisense transcripts. Our results obtained by RT-PCR on circularized RNAs and 5' RACE TAP experiments show that the *psbT* antisense transcript covers not only the reading frame of *psbT* but extents far into the reading frame of *psbB* and covers a large part of the *psbT/psbH* intergenic region (Fig. 31B). A second, shorter, *psbT* antisense RNA has been demonstrated by Zghidi et al. (unpublished results). This RNA has the same 5' end as the one described

above, but the 3' end is located in the intergenic region between *psbB* and *psbT*. Thus, there are two *psbT* antisense RNA of different length that could hybridize to *psbT* sense RNA.

In chapter 3 we have in addition mapped 5' and 3' ends of *psbT* sense RNAs (Figs. 32 and 33). These experiments show a principal difference between sense and antisense RNAs: Antisense RNAs have stable 3' and 5' ends, but sense RNAs have only stable 3' ends while their 5' ends are exposed to extensive nucleolytic degradation. This difference might be explained in the following way: (1) Results showing the survival of *psbT* RNA to RNase A/T1 digestion (i.e. its protection by RNA/RNA double strand formation) confirm the existence of double stranded RNA *in vivo* (Zghidi et al., unpublished result). (2) From our macroarray results we can conclude that antisense RNAs are present at much lower levels than sense RNAs. (3) Taken together, we can assume that all antisense RNA is hybridized with sense RNA in RNA/RNA hybrids and this hybrid formation stabilizes the RNA. In contrast, a large part of *psbT* sense RNA exists in single stranded form and this RNA is quickly degraded from its 5' end.

If we remember now the strong diminution of plastid mRNAs after high light treatment in young plantlets (chapter 2) together with the fact that only *psbT* antisense RNA does not diminish (not shown), it becomes clear that double stranded RNA should also be protected from degradation during high light stress. Thus, one of the possible functions of *psbT* antisense RNA might consist in the protection of a part of the *psbT* sense RNA during photooxydative stress conditions to provide intact RNA for PSBT production immediately during recovery from the stress.

Another possibility of regulation by antisense RNAs concerns the translational level. At the translational level mRNA duplex formation with antisense RNAs should prevent the recruitment of the translational machinery at the mRNA 5' UTR. That *psbT* douplex RNA is not translated becomes evident from the fact that in the absence of antisense RNA (i.e. in *sig3* plants) we observe much higher levels of PSBT protein (Zghidi, Buhr and Lerbs-Mache, unpublished result).

However, at the moment, both hypothesis of *psbT* antisense RNA function seem to be contradictory. If double stranded RNA cannot be translated, what is the meaning to protect them during stress? It needs to suppose a mechanism that liberates the sense RNA after the stress and makes it competent for translation. In order to analyse possible mechanisms of translational regulation by antisense RNA we propose to analyse the *psbT* sense and antisense RNAs and PSBT protein accumulation during and after photooxidative conditions by using

methods like toe printing, polysome analysis in WT and *sig3* plants and the establishment of an *in vitro* translation system using single and double stranded RNAs.

Preliminary experiments by RT-PCR show that in tobacco as well as in *Chlamydomonas psbT* antisense RNA may exist (Merendino, unpublished results). These data suggest that an antisense-mediated mechanism of *psbT* gene expression control in chloroplast might be conserved in other organisms.

Yet another function of *psbN* initiated antisense RNA might be related to processing of the polycistronic transcript of the *psbB* operon. Primer extension analysis of *psbH* processing intermediates showed the absence of one of the processing event in the 5' UTR of *psbH* in *sig3* mutant plants (Fig. 34). Our results indicated that duplex formation is also necessary for at least one of the processing events, although the importance of this specific event for PSBH protein expression is not clear because the protein level does not change in the *sig3* plants (Zghidi et al. unpublished result).

Finally, results obtained using the oligonucleotide based macroarray approach also indicate the existence of other antisense RNAs in addition to the *psbT* antisense RNA (e. g. *psbA* and *rbcL*) in *Arabidopsis* chloroplasts (Fig. 23). These data suggest that the cell might use antisense-mediated regulation as a general mechanism to control plastid gene expression.

To advance further in our understanding of antisense RNA function, we propose to analyse the expression of the antisense RNAs in plants grown under different physiological conditions, in wild type plants of different developmental stages (germinating seeds, young seedlings and mature plants) and in different plant organs.

Yet another approach to investigate the function of plastid antisense RNAs consists in plastid transformation. Transplastomic tobacco plants could be a mean to overexpress *in trans* some of the antisense RNAs and to analyse the stability and the translation efficiency of the corresponding sense RNAs subsequently by Northern blot and Western blot experiments, respectively. An eventual effect of the antisense RNAs on translation should be confirmed by analysing polysomes association of the sense and antisense transcripts in the transgenic plants.

Material and Methods :

In vitro cultivation and origin of plants :

The *Arabidopsis* seeds of Wild type Columbia (Wt), Wild type Wassilevskija (Wt ws), *sig1* (GK-758B02, Co), *sig2* (Kanamaru et al., 2001; Ws), *sig3* (SALK_081321, Co), *sig4* (Favory et al., 2005; Ws), *sig*5 (Tsunoyama et al., 2004) and *sig6* (Ishizaki et al., 2005) mutants plants were always surface sterilised before *in vitro* cultivation. Surface sterilisation of the seeds was carried out in a 50 ml solution of 0.75% hypochlorite and 80% Ethanol. Seeds were dipped into the solution for five minutes and continuously agitated by inverting the 50 ml Falcon tube. They were washed three times with absolute Ethanol. They were kept under hood for at least two hours for drying.

The sterilized seeds were cultivated in petri dishes on 25 ml of a growth medium containing 4.3 g MS salt (SIGMA); 0.5 g of MES pH 5.7 (KOH); 10 g of bacto agar and 10 g sucrose per litre. Stratification of the seeds was done at 4°C for 72 hours in darkness in order to remove dormancy and to have a synchronised germination. Then the seeds were transferred to a growth chamber with a cycle of 16 hours light and 8 hours darkness for germination and growth. The light intensity in the growth chamber was 60 photons $m^{-2} s^{-1}$, temperature 23-24 °C. Petri dishes were scotched with urgopore scotch to avoid contamination but to ensure passage of air for respiration of growing seedlings.

In vitro Cultivation of etiolated plants was carried out in the same growth chamber but in complete darkness for five days. However seeds were exposed to light for two hours after stratification as *Arabidopsis thaliana* seeds need light for germination.

Cultivation of plants in soil:

The seeds were sown on sterilised soil beneath a plastic film (until the appearance of the first pair of leaves). Seeds were stratified for three days at 4°C in darkness. The seeds were spread over the surface of the sterilised soil and were covered with a plastic film to avoid dryness which was removed after the appearance of first pair of leaves.

The plants were grown in a growth room (25 °C, light intensity 60 photons $m^{-2} s^{-1}$, 16h/8h light and dark cycle). The plants for the stress experiments were grown under the same conditions but with a different light-dark cycle, 12h/12h light and dark cycle.

Extraction of RNA:

Plants grown *in vitro* for five, six or seven days were collected and frozen into liquid nitrogen which ensures quick inactivation of ribonucleases. Then the entire plants (roots, stems and leaves) were ground into a fine powder in liquid nitrogen with the help of pestle and mortar. A volume of plant powder was used for extraction and the rest was stored at −20°C. All the necessary measures were taken to avoid melting of frozen powder as it activates ribonucleases which in turn cause degradation of RNAs. 1000 µl of RNA extraction solution (Tris HCl 0.2 M pH 9, LiCl 0.4 M, EDTA 25 mM, SDS 1%) and two volumes of phenol chloroform iso-amyl alcohol for removal of proteins were added and mixture was vortexed. Centrifugation was carried out in a microcentrifuge at room temperature at 13000 rpm for 10 minutes. Aqueous phase was taken into a new appendorf tube and extraction was repeated. It was followed by a chloroform extraction. Ribonucleic acids were preferentially (selectively) precipitated by LiCl (2M final concentration) at −20°C over night. Centrifugation was done at 4°C for 20 minutes and supernatant was removed. Pelleted RNAs were dissolved in water and re-precipitated at -20°C over night with 1/10 volume of 10X TNE (200mM Tris pH 7.5, 10mM EDTA, 1M NaCl) and three volume of Absolute Ethanol. Centrifugation was done at 4°C for 15 minutes and pellets were dried under vacuum. The pellet was dissolved in 20 µl water. RNAs were diluted to measure the exact quantity on a spectrophotometer at 260 nm wavelength and were run on a 1% agarose gel to check quantification again and the quality of RNAs. RNA aliquots were stored at −20°C.

Treatment of RNAs with DNase:

In order to efficiently remove DNA (as required for RT-PCR reaction) 20ug of RNA were treated with 3U of Turbo DNase (Ambion) for 30 minutes at 37°C and the treatment was once again repeated. Sterile Water was added to increase the volume to 200µl and phenol acid extraction was done with one volume of phenol acid followed by one chloroform extraction. The aqueous phase was precipitated with 1/10 volume of 10X TNE (200mM Tris pH 7.5, 10mM EDTA, 1M NaCl) and 2 volumes of absolute Ethanol for minimum two hours to over night at −20°C followed by centrifugation at 13000rpm at 4°C for 20 minutes. The pellet was washed by 70% ethanol and dried by vacuum drier. RNAs were quantified by spectrophotometer and gel electrophoresis.

Northern Blot analysis:

Principal

Total RNA are extracted from a homogenized plant sample. RNA samples are then separated by gel electrophoresis. The RNA samples separated by size are transferred to a nylon membrane through a capillary system. A nylon membrane with a positive charge is used in northern blotting as the negatively charged nucleic acids (RNAs) have a high affinity for them. Formamide is used in the transfer buffer to lower the annealing temperature of the probe-RNA interaction which prevents RNA degradation by high temperatures. RNAs are immobilized through covalent linkage to the membrane by UV light or heat. The labelled probe is hybridized to the RNA on the membrane. Ionic strength, viscosity, duplex length, mismatched base pairs and base composition affect the efficiency and specificity of hybridization. To avoid background signals the membrane is washed. The hybrid signals are then detected by X-ray film.

Probe preparation:

Hybridisation probes were PCR amplified by using following (F) forward and reverse (F) primers specific to each chloroplast transcript.

-418 *atpH* probe

5' GATATTGCCTAGGTATATATG 3' (F)

5' GTAATCGCTAAGATTAATCCAGCC 3' (R)

CDS -45 *atpH* probe

5' ATGAATCCACTG GTTTCTGCTG 3' (F)

5' GGCTTAAACAAAAGGATTCGC 3' (R)

***atpE* probe**

5' GTGTACTGACTCCGAATC 3' (F)

5' GCCTCAATTGTCTGTCTC 3' (R)

The PCR amplified products were purified by using PCR purification kit ( ). The size of each amplified fragment (probe) was verified on 1% agarose gel. using the random priming kit (New England Biolabs). 30 ng of fragment was radioactively labelled by α-dATP following the hexamer random priming method. Shortly, The amplified DNA fragment is denatured at 65°C for 5 minutes and put on ice for five minutes. 50 μCi of α-^{32}P dATP and 5

units of DNA pol I- Klenow Fragment 3'>5' exo are added to denatured DNA along with dCTP, dGTP and dTTP (2 µl of each). Polymerization reaction is carried out for one hour at 37°C. The probe is purified by one ml column G50 by gravity.

Gel electrophoresis:

RNA denaturing gel contained 1X MOPS pH 8.0, 2.2 M Formaldehyde and 1.3 % agarose. 12 µg of quantitatively and qualitatively analysed RNAs (total) of Wild type Columbia (Wt), Wild type Wassilevskija (Wt ws), *sig2* (Kanamaru et al., 2001; Ws) and *sig3* (SALK_081321, Co) were mixed with 1X sample buffer containing deionised formamide (18.5 % - half volume), formaldehyde (13 % - 1/6 volume), MOPS pH 8.0 (1X), Bromophenol blue (0.25 %), and Xylene cyanol (0.25 %) .

Formamide was deionised by adding 10 % amberlite 150 L IRN and put at slow agitation for one hour at room temperature and then filtered by 0.2 µm filter. 1µl of 400 µg/ml Ethedium Bromide was added to each sample before loading on the gel in order to observe the relative RNA quantification under UV light.

The gel was pre-run for 30 minutes at 100 V. Upon sample loading, the gel was run for 8 hours at 100 V in 1X MOPS and 1/10 th volume of Formaldehyde. Recycling of running buffer was done by magnetic stirrer. Photos of the gel were taken under UV light. The gel was then rinsed in two changes of ddH$_2$O and washed for 5 minutes in 5X SSC.

RNA Transfer

RNAs were transferred to Nylon membrane (Hybond +) from the gel in 5X SSC following the capillary method (Sambrook & Russell, 2005). Transfer was done for three nights. During this period tissue papers were changed and transfer buffer was added to the reservoir dish or apparatus.

The RNAs were cross linked to the membrane by incubation at 80°C for two hours. The membrane was then coloured in methylene bleu solution (0.03% methylene blue, 0.3M NaOAc, pH 5.2) for 60 seconds and washed in ddH$_2$O for two minutes. The membrane was scanned for photo. A photo of the membrane under UV light was also taken. The membrane was decoloured in ddH$_2$O for 10 minutes.

Hybridisation

The membrane was prehybridised at 65°C for two hours in 7% SDS, 0.5M Na Phosphate pH 7, 1mM EDTA and then overnight hybridisation was performed in the same solution in presence of the probe (which was denatured at 95°C for 10 minutes). The hybridised membrane was then quickly rinsed in 0.2XSSC, 0.2 % SDS and washed in the same solution for 20 minutes at 65°C. The membrane was wrapped in a plastic cover and exposed to films to acquire the appropriate signals.

Primer Extension:

Primer extension is a technique used to map the initiation sites of the primary or processed transcripts. The use of a proper loading controlallows the technique to be quantitative.

3 µl of 5' ^{32}P labelled oligo (see below) was added to 4-10 µg of total RNA; depending on the expression level of the genes. RNAs were denatured at 65°C for 10 minutes and immediately put on ice at 4°C for 5 minutes for the RNAs to maintain their denatured form. The hybridization was done at 5°C less than melting temperature of the oligo (Tm − 5) calculated by the formula (2AT + 4GC) for 20 minutes. Again the tubes were put on ice for 5 minutes. Reverse transcription was done in presence of 1µl dNTPs (10 mM), 5 µl first strand buffer 5X, 2.5 µl DTT (0.1M), 1 µl Rnase inhibitor and 0.5 µl superscript II 200 units/µl (invitrogen) at 42°C for 50 minutes. In case of strong secondary structures present on the RNAs, superscript III was used and RT was done at 55°C. The reaction was stopped by inactivating the enzymes at 70°C for 15 minutes. RNA was removed by adding 1 µl RNase A (10mg/ml) at 37°C for 20 minutes. Proteins were removed by 2 phenol chloroform and one chloroform extractions. Precipitation of purified cDNA was done in ethanol 1/10 volume of 10X TNE (200mM Tris pH 7.5, 10mM EDTA, 1M NaCl) 3 volumes of absolute ethanol over night at −20°C followed by centrifugation at 13000rpm at 4°C for 15 minutes. The pellet was washed by 70% ethanol and dried by vacuum drier. The pellet was dissolved in 5µl stop buffer and denaturation was done by heating at 95°C for 10 minutes. cDNA were run on 6% acryl amide gel. The oligos used for primer extension are given below:

atpI:	5' CATATTGCCCTCTGACAG 3'	(ta:49°C),
atpH:	5' GTCCAATAGAAGCAAGC 3'	(ta:50°C),
atpF:	5' TCAATACACCGAAAACTACAC 3'	(ta:55°C),
atpA:	5' GGTACCGGTATTTACAATCG 3'	(ta:53°C),
atpB:	5' TTTTTCACGTATCGAAACCTCTGG 3'	(ta:43°C),
atpE:	5' CTAATTGTTTCCGCTAGACC 3'	(ta: 53°C),

clpP :   5' GGAGAAGGAGATACATC 3'   (ta: 40°C),
16S :   5' TTCATAGTTGCATTACT 3'   (ta: 40°C),
psbH;   5' GCTTCTTGGACCAGATC 3'   (ta: 40°C),

Before phenol chloroform extraction equal quantiy of the loading control is added into each reaction mixture. Loading control is the labeled PCR product. The size of the loading control is such that it does not co-migrate with the band of interest on the polyacrylamide gel (see below). For labeling the purified PCR product, it is denatured at 95°C for 10 minutes and then the protocol for oligo labeling is followed as such (see below). For our primer extension analyses, we have used loading controls of 450 and 800 bps size depending on the size of the band of interest.

Preparation of polyacrylamide gel:

For the analysis of primer extension products 6% denaturing polyacryamide gel was used. The gel contained 6% polyacrylamide solution [bis-acrylamide/ acrylamide; 8 M Urea and TBE-S 1X (45 mM Tris-Borate pH 8.3, 0.5 mM EDTA)]. Upon addition of APS (amonium persulfate) TMED (Tetramethylethylenediamine) the gel was polymerised at room temperature in 30 minutes. The samples were run on the gel for one and half hour at 30W, for long gels time was increased to two hours. The gel was transferred to a 3MM paper, dried and exposed to photographic films at -80°C or at room temperature.

Sequencing:

T7 sequencing kit (usb) was used for sequencing of genomic fragments. 2 µg of the plasmid DNA containing the genomic fragments of interest and 10 pmoles of the corresponding oligo were added in a tube with a final volume of 12 µl. Denaturation was done at 65°C for 10 minutes in the presence of 1.5 µl NaOH (1N) solution. After denaturation 1.5 µl of HCl was immediately added to the mixture for neutralisation. For hybridisation of primer with the genomic fragment 2 µl of annealing buffer was added and incubation was done at 37°C for 10 minutes and at room temperature for five minutes. Then 5.3 µl of the premix (containing 3 µl labelling buffer A, 2 µl dilution buffer, 0.5 µl T7 DNA polymerase 10U/µl and 1 µl of α-P^{32}-dATP) was added followed by polymerisation at room temperature for five minutes. 4.5 µl of this reaction was added to 2.5 µl of each ddNTP and incubated for five minutes at 37°C. 5µl of stop buffer (97.5 % formamide, 10mM EDTA, 0.3 % bromophenom blue, 0.3 % xylène cyanol) was added to the reaction in each tube to stop the

reaction. Before loading 3 µl of reaction on 6% denaturing polyacrylamide gel, denaturation was done at 95°C for 10 minutes.

Oligo labelling:

Oligo labeling which is required for primer extension technique is performed by phosphorylation of the 5' end with γ- P^{32} -ATP. 100 ng of the oligo is incubated for 10 minutes at 37°C with 5 µl γ- P^{32} –ATP (10 u Ci/ µl, 3000 Ci/mmol), 5µl of 5X buffer and 10 units of T4 polynucleotide Kinase (invitrogen) in a total volume of 25 µl. Labelled oligo-nucleotide is then purified to remove unincorporated nucleotides by chromatography using G-25 columns (GE healthcare).

PCR Amplification:

PCR (polymerase chain reaction) is the technique used for amplification of a specific DNA/ cDNA fragment from total DNA.

The reaction was carried out in a final volume of 25 µl contaning 1 µl of DNA, 5 µl Taq DNA polymerase buffer (5X), 1.5 µl of 25 mM $MgCl_2$, 0.5 µl of dNTPs mix (10 mM each), 1 µl of each (forward and reverse) primer (10 µM) and 0.1 µl of Taq DNA Polymerase enzyme.

Denaturation of double stranded DNA was done at 94°C for 5 minutes and it was followed by n cycles of amplification which generally included 1 minute at 94°C, 30 sec at Ta (Tm-5) and 1 minute at 72°C. For the fragments longer than 1kb extended time at 72°C was increased at the rate of 30 sec for each 500 bps. If the fragment was amplified for cloning then one cycle at 72°C for 10 minutes was added at the end.

Cloning of the DNA Fragment:

Principle:

When required, DNA fragments amplified by PCR or RT-PCR were directly inserted in TOPO 2.1 PCR vectors by the standard protocol given by the supplier (invitrogen). These vectors contain gene for resistance to ampiciline. The insertion site for the fragment is in LacZ gene which codes for $\beta-$galactosidase. When the DNA fragment is inserted in it, the gene becomes inactive and there is no production of $\beta-$galactosidase enzyme. Bacteria were allowed to grow in liquid lauria broth (LB 1 % trypton p/v, 0.5 % Yeast Extract p/v, 0.5 % NaCl) for one hour. Selection is done on agar plates containing both the ampiciline and X-gal.

Bacteria transformed with vectors containing the insert grow into white colonies due to the absence of β−galactosidase enzyme while the bacteria transformed with empty vectors grow into blue colonies as a result of cleavage of X-gal due to the production of β−galactosidase.

Protocol:

Commercially availanle XL-2 bleu competent bacterial cells provided in the Kit (invitrogen) were transformed with the ligated vectors by thermal shock. These bacteria were first cultured in LB liquid for one hour at 37°C and this culture is then spread on LB agar (1% tryptophane; 0.5% yeast extract, 0.5% NaCl and 15 g/l bacto-agar) plates containing ampicilline (50 μg/ μl) and X-gal over night at 37°C. The white colony bacteria are then cultured again in liquid LB medium over night at 37°C in the presence of ampicilline (50 μg/ml) at continuous horizontal agitation at 200 rpm. The bacteria are pelleted at 9000 rpm for 5 minutes at room temperature and DNA is then extracted.

Miniprep; Plasmid DNA Extraction:

Plasmid Extraction from bacteria was carried out by alkaline Lysis method (Sambrook et al., 1989). Bacteria cultured over night at 37°C were pelleted by centrifugation at 9000 rpm for 5 minutes at room temperature in a 2 ml tube. The pellet is re-dissolved in 100 μl of solution I (glucose 50 mM ; EDTA 10 mM; Tris.HCl pH 8, 25 mM). Lysis of the bacteria and denaturation of DNA is done by 200μl of alkaline solution II (1% SDS, O.2 N NaOH) and incubated for 5 minutes at room temperature. 150 μl of Solution III (sodium Acetate 5M, Galacial acetic acid, pH 5.2) is added for renaturation of the plasmid DNA. The mix is agitated gently by inverting the tubes and incubated for 10 minutes at 4°C on ice. Centrifugation is done at 13200 rpm for 15 minutes which precipitate the chromosomal DNA while plasmid DNA remains in supernatant. The supernatant is taken in a new tube and precipitated with two volumes of cold absolute ethanol at -20°C for two hours. The plasmids are pelleted by centrifugation at 13200 rpm for 15 min. The pellet is washed with 70% ethanol, dried under vacuum and re-dissolved in 20 μl sterile water.

RNA Analysis by RT-PCR:

Upon DNase treatment total RNA is denatured at 65°C for 5 minutes in the presence of 1 μl of gene specific primer (2 μM), reaction buffer (50mM Tris HCl; pH 8.5, 75 mM KCl, 3 mM $MgCl_2$ and 10 mM EDTA) and 0.5 mM dNTPs. The reverse transcription is done in the presence of 200 U of Superscript II H-RT (invitrogen) and 100 U of RNase inhibitor at

42°C during 50 min. The reaction is stopped by heating at 70°C for 15 min. 1-2 μl of cDNA is utilised to carry out the PCR reaction as mentioned above.

Analysis of plastidial transcript profile expression by cDNA macroarray:

cDNA macroarray was used for plastid transcriptome analysis. The technique is carried out in the following steps:

- cDNA synthesis: Labelling of plastid mRNAs was obtained by reverse transcription in the presence of α P^{32}-dATP and gene specific primers.
- Hybridisation of cDNA on the membrane where 80 gene-specific primers were spotted.
- Analysis of results.

cDNA synthesis:

6 μg of DNAase treated total mRNAs are denatured at 65°C for 5 min in the presence of 0.2 pmol of oligonucleotide mix (plastidial gene specific primers), 1.25 mM of d CTP, 1.25 mM of dGTP, 1.25 mM of dTTP and 10 μl of α P^{32} –dATP (10 μCu/μl). Reverse transcription is done with 600 U of superscript II H-RT (invitrogen) in the presence of 100 U of RNase inhibitor during 50 min at 42°C. Reaction is stopped by heating at 70°C for 15 minutes. The labelled cDNAs are purified by 1 ml column of G-50 Sephadex columns. In order to check the quantity of incorporated radioactivity, 1 μl labeled and purified cDNA is run on denaturing acrylamide gel.

Hybridisation of labelled cDNAs on membrane:

Macro-array membranes made of nitrocellulose, were imbibed in SSC 2X (0.3 M Na Citrate, pH 7). The membranes were then pre-hybridised for one hour in hybridisation solution (NaHPO$_4$ pH 7.2, 1mM EDTA, 7 % SDS, 1 % BSA) at 65°C and then hybridized with labeled cDNAs (upon denaturation at 95°C for 5 minutes) in 25 ml hybridisation solution for 72 hours at 65°C. The membranes are then washed with 25ml washing buffer (40 mM NaHPO$_4$ pH 7.2, 1 mM EDTA, 1% SDS) for 5 min at 65°C and then at room temperature for 10 min in 100 ml washing buffer. The membranes are then covered with a plastic film and exposed to photo-sensible (fujifilms imaging plates) films during two weeks at room temperature. The signal intensities from the screens are converted in digital signals by phosphor-imager (fujifilm FLA 8000).

Analysis of Results:

The digital signals were then processed with the Array guage software. After background substraction, the values were transferred to Microsoft excel sheets.

TAP (Tobacco Acid Pyrophosphate) treatment and 5' RACE:

Principle:

The plastidial primary transcripts have three phosphate groups on the 5' extremity. The processed transcripts which are issued as a result of cleavage of primary transcripts have one phosphate group attached on the 5' extremity. An RNA adaptor can only ligate with the 5' extremity of processed transcripts and not to primary transcripts. But if the transcripts are treated with TAP (tobacco acid pyrophosphate) which removes two phosphate groups (β, γ phosphates) and the adaptor ligation is possible also for primary transcripts. In order to differentiate between the primary and processed transcripts TAP treated and TAP untreated transcripts were amplified by RT-PCR by using RLM-RACE kit (ambion). Cloning and sequencing of products of RACE-PCR can also be used to map exactly the 5' extremities. The principle of the process in shown in Fig 36.

Protocol:

10 µg of DNase treated RNAs were treated with 2 µl TAP enzyme (first choice RLM- RACE; Ambion) in presence of the proper buffer and 40 units of RNase inhibitor in a final volume of 20 µl for 1h at 37°C. A control reaction without TAP was carried out. 4 µl of each + TAP and – TAP reactions were ligated to 300 ng of 5' adaptor with the help of 5 U of T4 RNA ligase in presence of its buffer at 37°C for 1h in a final volume of 10 µl. Reverse transcription of 2 µl of ligated RNAs was done with a gene specific oligo (2 µM).
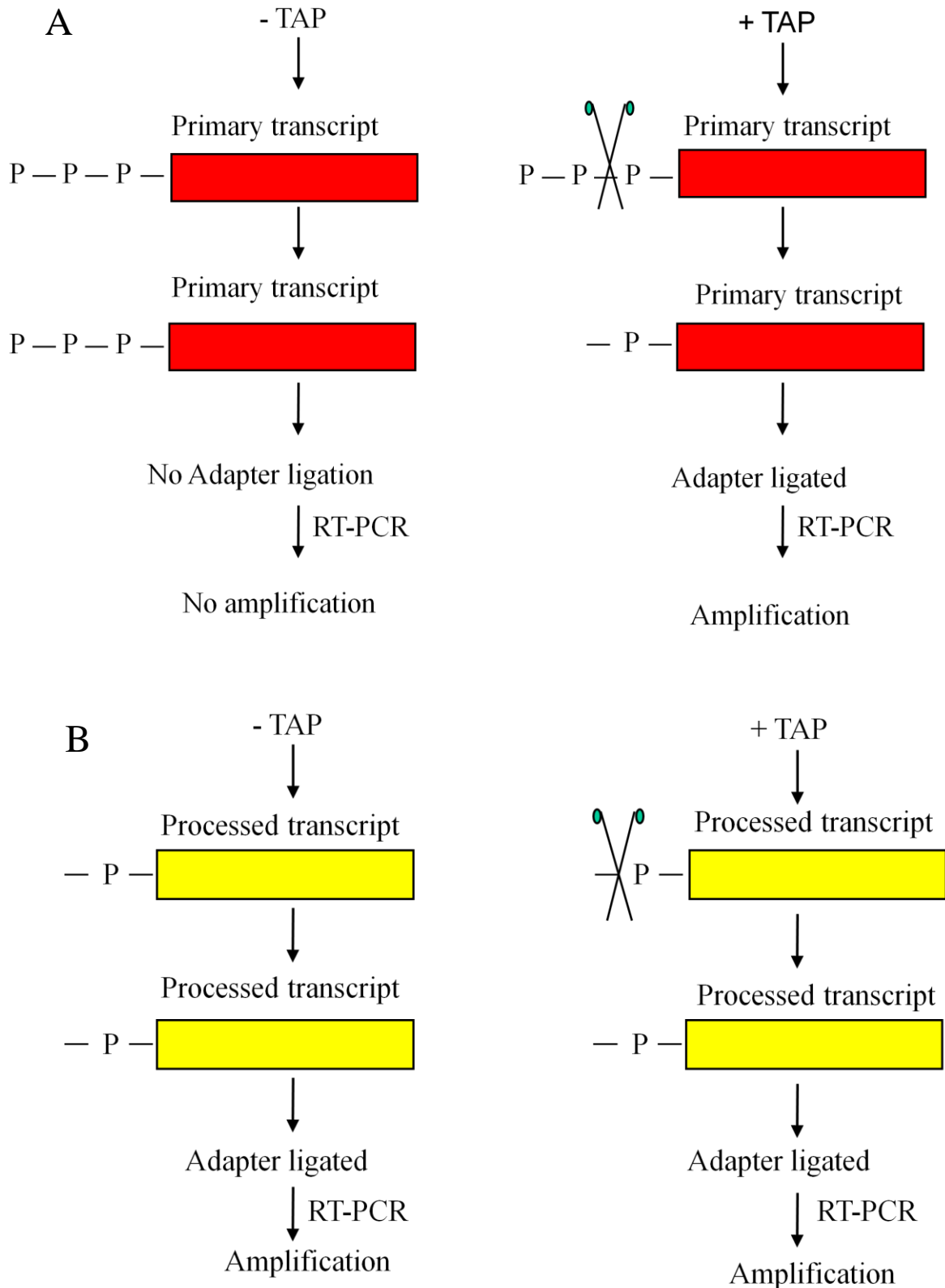
Figure 37: Schematic presentation of the principle of 5' RACE TAP PCR. A) primary transcript. B) processed transcripts.

In order to enhance the specificity of amplification two successive PCRs were carried out on the obtained cDNA. First PCR was carried out with the help of an adaptor specific external oligo and transcript specific external oligo. While the second PCR was carried out with an adaptor specific internal oligo and transcript specific internal oligo. PCR products were run on agarose gel. Presence of band (amplification) in un-treated (-TAP) transcripts shows that the RNA is a processed transcript while absence of band in treated transcripts (+TAP) shows that the RNA is a primary transcript.

Circular RT- PCR:

Principle:

This technique is used to determine the 5' and 3' extremities of a given transcript. For the purpose RNAs are TAP treated and then auto ligated, reverse transcribed and PCR amplified with specific oligos situated near the 3', 5' junction. PCR product is cloned. Many clones are then sequenced and analysed to determine 5' and 3' extremities.

Protocol:

5-6 µg of total RNAs treated with DNase and TAP are denatured at 70°C for 5 min and immediately transferred to ice for 5 min. Auto-ligation is done with 50 U of T4 RNA ligase (fermentas) in the presence of RNase inhibitor (40 U) for 1h at 37°C. Auto-ligated RNAs are then purified with two phenol chloroform extractions followed by ethanol precipitation. Upon centrifugation at 13000 rpm at 4°C for 15 minutes, the pellet is washed with 70% ethanol and once dried, it is re-suspended in sterile water. Reverse transcription was done in presence of 1 µl of transcript specific oligo (2 µM) and 200 U of superscript II H-Rt (invitrogen) in a final volume of 20 µl for 1h at 42°C. 2 µl of this cDNA mixture are used for PCR. PCR product was cloned and many clones were sequenced.

Western Blot Analysis:

Principal

Total proteins are extracted from an homogenized plant sample. The proteins are then separated by size by gel electrophoresis and transferred to a membrane (nitrocellulose or PVDF). The transferred proteins are then detected by using antibodies specific to the target protein.

Protein extraction

Plants grown *in vitro* for seven days were collected and frozen into liquid nitrogen which ensures quick inactivation of proteases. Then the entire plants (roots, stems and leaves) were ground into a fine powder in liquid nitrogen with the help of pestle and mortar.

Equal quantities (100 µg) of the powder of Wild type Columbia (Wt), *sig2* (Kanamaru et al., 2001), *sig3* (SALK_081321) plants were suspended in the 1V of 4X loading buffer (200 mM Tris HCL, pH 6.8; 5 % (v/v) β-mercapto ethanol; 0.2 % bleu de Bromophenol ; 20 % glycerol and 4 % SDS). The solution was vortexed and denatured at 100 °C for 10 minutes. After denaturation it was centrifuged at maximum speed (13200 g) for 15 minutes. The supernatant was taken into a new appendorf. This supernatant can directly be used for quantification and migration by electrophoresis. For the western blot analysis of smaller plastid proteins (ATPH and PSBT), a portion of extracted proteins (100 µg) was precipitated in acetone 80 % at -80 °C for about half an hour. The samples were centrifuged for 12 minutes at room temperature at 12 000 rpm. The pellet was washed with 200-250 µl 100 % acetone and the pellet was dried on the bench at room temperature. The pelleted proteins were dissolved in an appropriate volume of the loading buffer by automatic pipette and used for quantification and migration by electrophoresis. The proteins were always denatured at 95 °C for 10 minutes and put at ice for 5 minutes before loading on the gel.

Protein quantification

The protein quantification was carried out by Esen Method as following:

2 µl of a protein extract (in SDS sample buffer, heat-denaturated at 95°C for 5 min) were spotted on a square of Whatman paper (1 x 1 cm). For the zero spot SDS sample buffer was used while for the fixed point spot BSA-solution of known concentration was used. The Whatman paper was dried at room temperature. Proteins fixing on the paper was done by incubating it in Fixing-Buffer (25% (v/v) Isopropyl alcohol, 10% (v/v) acetic acid) for 5 minutes with gentle agitation. Proteins were coloured by incubation in Colour-Buffer [25% (v/v) Isopropyl alcohol, 10% (v/v) acetic acid, 0,1% (w/v) Coomassie Brilliant Blue R250)] for 15 min with gentle agitation. The paper was rinsed one to two times with cold water. The paper was then decoloured by putting it in boiling water for two minutes. The paper was rinsed one time with cold water. The paper was dried at room temperature. The squares were cut

and each one was put in an eppendorf-tube. 1 ml of Elution-Buffer (0,5 % SDS) was added and the blue colour was elute over night at RT or for 2 h at 50°C. The absorbance was measured at 578 nm against the zero-probe and the protein-concentration was calculated by comparing to the fixed point value.

Gel preparation

The gel apparatus was set and sealed with 1 % agarose gel at the base and sides to avoid leakage of the gel. The gel for the proteins of >10 kDa consisted of two parts like stacking (Tris HCL 0.5 M pH 6.8, SDS 0.4 %) and separation gel (Tris HCL 1.5 M pH 8.8, SDS 10%).

20 µl of 10 % ammonium per sulphate (APS) and 5 µl of TEMED were added to 20 ml separation gel solution. Out of it, 5-6 ml of separation gel solution was poured. 1 ml of either H_2O or isopropanol was added to it to keep the surface levelled. After 30 minutes water or ethanol was removed with the help of whatman paper and 1-2 ml of the stacking gel solution having appropriate quantities of APS and TEMED were added to it. The comb of required sized was placed on it. After 30 minutes the comb was removed and the denatured samples of protein and molecular weight marker were loaded to it.

The gel was run in an electrophoretic buffer (25 mM Tris, 200mM Glycine and 0.05 % SDS) for 2 hours at 100 V.

Protein Transfer

Normally two gel were prepared, loaded and run in parallel. One was used for coomassie blue staining for verification of quantity and quality of the proteins and the other was used for western blot analysis.

Electric transfer of the proteins to the membrane (nitocellulose) was done in the transfer solution (25 mM Tris, 200 mM Glycine, 0.04 % SDS, 20 % EtOH) for sixty minutes at 100 V. The current flows from cathode to anode, so the proteins being negatively charged due to SDS move from the gel (cathode side) to the membrane (anode side) and are trapped in the membrane.

The membrane is stained in "Ponceau Rouge" to confirm that the protein transfer was successful. Membrane blocking was done in TBS-Tween 0.1 % and 5 % milk for one hour with agitation at room temperature. Then the incubation with the first antibody was

done at room temperature for one hour in TBS-Tween 0.1 % and 5 % milk. The membrane was washed 3-4 times with TBS-Tween 0.1 %. The membrane was again incubated with the 2nd antibody for one hour at room temperature in TBS-Tween 0.1 % and 1 % milk. The membrane was washed 3-4 times in the TBS-Tween 0.1 %. 1 ml of ECL solutions was added on the membrane and exposed to films.

Western Blot analysis for smaller proteins

We carried out the western blot analysis very small protein (ATPH- ~5 kDa) as well. According to the protocol described by (Schagger, 2006). For the gel preparation two mother solutions like Tris/ Tricine 3X solution (3M Tris, SDS 0.3 % pH 8.45) and Acryl amide/ Bisacrylamide 06 (AB 06) (46,5 g acrylamide (49,5%), 3 g bisacrylamide (6%), into 100 ml H$_2$O) were prepared. Separation gel of 5-6 cm (7,5 ml 14% acrylamide, 2.3 ml H2O, 2.5 ml Tris/Tricine buffer 3x, 2,1 ml Acryl amide/ Bisacrylamide 06, 0,6 ml Glycerol, 37,5 μl APS 10%, 3,75 μl TEMED) was followed by 1 cm intermediate gel (1 ml 10% acrylamide, 67 μl H$_2$O, 333 μl Tampon Tris/Tricine 3x, 200 μl Acryl amide/ Bisacrylamide 06, 5 μl APS 10%, 0,5 μl TEMED). 1 ml of H$_2$O was added on top of them and were allowed to polymerize for 30 minutes. The water was taken off by whatman paper and 1cm of stacking gel (3 ml 4% acrylamide, 2 ml H2O, 750 μl Tampon Tris/Tricine 3x, 250 μl Acryl amide/ Bisacrylamide 06, 22,5 μl APS 10%, 2,25 μl TEMED) was deposited on them and allowed to polymerize with the comb in it. The arrangement of different parts of the gel is shown in the following figure.
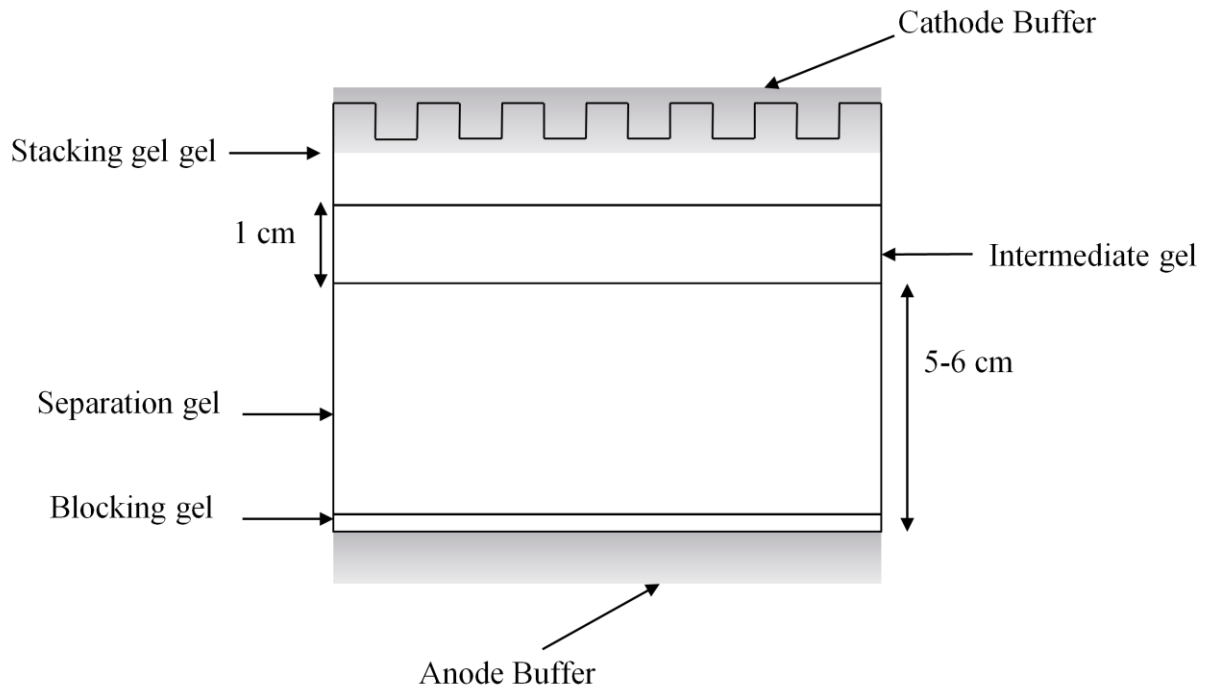
Figure 38: Schematic presentation of the formation of the gels for loading and migration of low molecular weight proteins. (Drawn in power point)

Procedure

The cathode buffer (0.1 M Tris, 0.1 M Tricine, 0.1 % SDS) was added in the upper part while the anode buffer (0.1 m Tris-HCl pH 8.9) at the base. The migration was started with very low voltage (20 V). When the samples entered in the intermediate gel, the voltage was increased to 40V: when the samples entered in the separation gel, the voltage was increased to 100 V. This migration procedure takes almost 6 hours. The transfer of proteins from the gel to the membrane (nitrocellulose 0.2 μm) was carried out at 50 V for 50 minutes in transfer solution (2.6 mM Tris, 0.19 glycine, EtOH 20 %). After colouring the membrane with the "ponceau rouge", the normal procedure was followed.

References

Aach, H., Bode, H., Robinson, D. G., & Graebe, J. E. (1997). Ent-kaurene synthase is located in proplastids of meristematic shoot tissues. *Planta, 202*(2), 211-219.

Allison, L. A. (2000). The role of sigma factors in plastid transcription. *Biochimie, 82*(6-7), 537-548.

Allison, L. A., Simon, L. D., & Maliga, P. (1996). Deletion of rpoB reveals a second distinct transcription system in plastids of higher plants. *Embo Journal, 15*(11), 2802-2809.

Andersson, S. G. E., Karlberg, O., Canback, B., & Kurland, C. G. (2003). On the origin of mitochondria: a genomics perspective. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 358*(1429), 165-177.

Azevedo, J., Courtois, F., Hakimi, M. A., Demarsy, E., Lagrange, T., Alcaraz, J. P., et al. (2008). Intraplastidial trafficking of a phage-type RNA polymerase is mediated by a thylakoid RING-H2 protein. *Proceedings of the National Academy of Sciences of the United States of America, 105*(26), 9123-9128.

Baginsky, S., Tiller, K., & Link, G. (1997). Transcription factor phosphorylation by a protein kinase associated with chloroplast RNA polymerase from mustard (Sinapis alba). *Plant Molecular Biology, 34*(2), 181-189.

Bain, J. M. (1968). A crystalline inclusion in chloroplasts of outer hypodermal cells of banana fruit. *Australian Journal of Biological Sciences, 21*(3), 421-&.

Baker, N. R., & Butler, W. L. (1976). Development of primary photochemical apparatus of photosynthesis during greening of etiolated bean-leaves. *Plant Physiology, 58*(4), 526-529.

Beick, S., Schmitz-Linneweber, C., Williams-Carrier, R., Jensen, B., & Barkan, A. (2008). The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. *Molecular and Cellular Biology, 28*(17), 5337-5347.

Ben-Shem, A., Frolow, F., & Nelson, N. (2003). Crystal structure of plant photosystem I. *Nature, 426*(6967), 630-635.

Bligny, M., Courtois, F., Thaminy, S., Chang, C. C., Lagrange, T., Baruah-Wolff, J., et al. (2000). Regulation of plastid rDNA transcription by interaction of CDF2 with two different RNA polymerases. *Embo Journal, 19*(8), 1851-1860.

Bock, R., Hermann, M., & Kossel, H. (1996). In vivo dissection of cis-acting determinants for plastid RNA editing. *Embo Journal, 15*(18), 5052-5059.

Bohne, A. V., Irihimovitch, V., Weihe, A., & Stern, D. B. (2006). *Chlamydomonas reinhardtii* encodes a single sigma(70)-like factor which likely functions in chloroplast transcription. *Current Genetics, 49*(5), 333-340.

Bolle, C., Kusnetsov, V. V., Herrmann, R. G., & Oelmuller, R. (1996). The spinach AtpC and AtpD genes contain elements for light-regulated, plastid-dependent and organ-specific expression in the vicinity of the transcription start sites. *Plant Journal, 9*(1), 21-30.

Bollenbach, T. J., Schuster, G., Portnoy, V., & Stern, D. B. (2008). Processing, degradation, and polyadenylation of chloroplast transcripts. *Topics in Current Genetics, R. Bock (Ed.): Cell and Molecular Biology of Plastids, 19*.

Bourett, T. M., Czymmek, K. J., & Howard, R. J. (1999). Ultrastructure of chloroplast protuberances in rice leaves preserved by high-pressure freezing. *Planta, 208*(4), 472-479.

Bouvier, F., Backhaus, R. A., & Camara, B. (1998). Induction and control of chromoplast-specific carotenoid genes by oxidative stress. *Journal of Biological Chemistry, 273*(46), 30651-30659.

Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., et al. (2002). Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Molecular Cell, 9*(3), 527-539.

Canino, G., Bocian, E., Barbezier, N., Echeverria, M., Forner, J., Binder, S., et al. (2009). Arabidopsis encodes four tRNase Z enzymes. *Plant Physiology, 150*(3), 1494-1502.

Carter, M. L., Smith, A. C., Kobayashi, H., Purton, S., & Herrin, D. L. (2004). Structure, circadian regulation and bioinformatic analysis of the unique sigma factor gene in *Chlamydomonas reinhardtii. Photosynthesis Research, 82*(3), 339-349.

Chen, H. C., & Stern, D. B. (1991). Specific ribonuclease-activities in spinach-chloroplasts promote messenger-RNA maturation and degradation. *Journal of Biological Chemistry, 266*(35), 24205-24211.

Chen, L. J., Rogers, S. A., Bennett, D. C., Hu, M. C., & Orozco, E. M. (1990). An invitro transcription termination system to analyze chloroplast promoters - identification of multiple promoters for the spinach atpB gene. *Current Genetics, 17*(1), 55-64.

Choquet, Y., & Vallon, O. (2000). Synthesis, assembly and degradation of thylakoid membrane proteins. *Biochimie, 82*(6-7), 615-634.

Choquet, Y., & Wollman, F. A. (2002). Translational regulations as specific traits of chloroplast gene expression. *Febs Letters, 529*(1), 39-42.

Courtois, F., Merendino, L., Demarsy, E., Mache, R., & Lerbs-Mache, S. (2007). Phage-type RNA polymerase RPOTmp transcribes the rrn operon from the PC promoter at early developmental stages in arabidopsis. *Plant Physiology, 145*(3), 712-721.

Cozens, A. L., Walker, J. E., Phillips, A. L., Huttly, A. K., & Gray, J. C. (1986). A 6th subunit of atp synthase, an f0 component, is encoded in the pea chloroplast genome. *Embo Journal, 5*(2), 217-222.

Daniels, D. L., Michels, W. J., & Pyle, A. M. (1996). Two competing pathways for self-splicing by group II introns: A quantitative analysis of in vitro reaction rates and products. *Journal of Molecular Biology, 256*(1), 31-49.

Daum, B., Nicastro, D., Austin II, J., McIntosh, R. J., & Ku¨ hlbrandta, W. (2010). Arrangement of photosystem II and ATP synthase in chloroplast membranes of spinach and pea. *The Plant Cell 22*, 1299–1312.

de la Sierra-Gallay, I. L., Zig, L., Jamalli, A., & Putzer, H. (2008). Structural insights into the dual activity of RNase J. *Nature Structural & Molecular Biology, 15*(2), 206-212.

de Longevialle, A. F., Hendrickson, L., Taylor, N. L., Delannoy, E., Lurin, C., Badger, M., et al. (2008). The pentatricopeptide repeat gene OTP51 with two LAGLIDADG motifs is required for the cis-splicing of plastid ycf3 intron 2 in Arabidopsis thaliana. *Plant Journal, 56*(1), 157-168.

De Santis-Maciossek, G., Kofer, W., Bock, A., Schoch, S., Maier, R. M., Wanner, G., et al. (1999). Targeted disruption of the plastid RNA polymerase genes rpoA, B and C1: molecular biology, biochemistry and ultrastructure. *Plant Journal, 18*(5), 477-489.

Dekker, J. P., & Boekema, E. J. (2005). Supramolecular organization of thylakoid membrane proteins in green plants. *Biochimica Et Biophysica Acta-Bioenergetics, 1706*(1-2), 12-39.

Demarsy, E., Courtois, F., Azevedo, J., Buhot, L., & Lerbs-Mache, S. (2006). Building up of the plastid transcriptional machinery during germination and early plant development. *Plant Physiology, 142*(3), 993-1003.

Deruere, J., Romer, S., Dharlingue, A., Backhaus, R. A., Kuntz, M., & Camara, B. (1994). Fibril assembly and carotenoid overaccumulation in chromoplasts - a model for supramolecular lipoprotein structures. *Plant Cell, 6*(1), 119-133.

Drapier, D., Rimbault, B., Vallon, O., Wollman, F. A., & Choquet, Y. (2007). Intertwined translational regulations set uneven stoichiometry of chloroplast ATP synthase subunits. *Embo Journal, 26*(15), 3581-3591.

Drapier, D., Suzuki, H., Levy, H., Rimbault, B., Kindle, K. L., Stern, D. B., et al. (1998). The chloroplast atpA gene cluster in *Chlamydomonas reinhardtii* - Functional analysis of a polycistronic transcription unit. *Plant Physiology, 117*(2), 629-641.

Dyall, S. D., Brown, M. T., & Johnson, P. J. (2004). Ancient invasions: From endosymbionts to organelles. *Science, 304*(5668), 253-257.

Esau, K. (1944). Anatomical and cytological studies on beet mosaic. *Journal of Agricultural Research, 69*, 0095-0117.

Esen, A. (1978). A simple method for quantitative, semiquantitative and qualitative assay of protein. . *Analytical Biochemistry, 89*(1), 264-273.

Favory, J. J., Kobayshi, M., Tanaka, K., Peltier, G., Kreis, M., Valay, J. G., et al. (2005). Specific function of a plastid sigma factor for ndhF gene transcription. *Nucleic Acids Research, 33*(18), 5991-5999.

Fedorova, O., & Zingler, N. (2007). Group II introns: structure, folding and splicing mechanism. *Biological Chemistry, 388*(7), 665-678.

Ferro, M., Salvi, D., Riviere-Rolland, H., Vermat, T., Seigneurin-Berny, D., Grunwald, D., et al. (2002). Integral membrane proteins of the chloroplast envelope: Identification and subcellular localization of new transporters. *Proceedings of the National Academy of Sciences of the United States of America, 99*(17), 11487-11492.

Fischer, N., Stampacchia, O., Redding, K., & Rochaix, J. D. (1996). Selectable marker recycling in the chloroplast. *Molecular & General Genetics, 251*(3), 373-380.

Gatenby, A. A., Rothstein, S. J., & Nomura, M. (1989). Translational coupling of the maize chloroplast atpB and atpE genes. *Proceedings of the National Academy of Sciences of the United States of America, 86*(11), 4066-4070.

Georg, J., Honsel, A., Voss, B., Rennenberg, H., & Hess, W. R. (2010). A long antisense RNA in plant chloroplasts. *New Phytologist, 186*(3), 615-622.

Goldschmidt-Clermont, M. (1991). Transgenic expression of aminoglycoside adenine transferase in the chloroplast - a selectable marker for site-directed transformation of chlamydomonas. *Nucleic Acids Research, 19*(15), 4083-4089.

Goldschmidt-Clermont, M. (2008). Chloroplast splicing. In The Chlamydomonas Sourcebook. *Organellar and Metabolic Processes, ed. D Stern, . San Diego, CA: Academic, Vol 2: *, pp. 915–936.

Goldschmidt-Clermont, M., Choquet, Y., Girardbascou, J., Michel, F., Schirmerrahire, M., & Rochaix, J. D. (1991). A small chloroplast RNA may be required for transsplicing in *chlamydomonas-reinhardtii*. *Cell, 65*(1), 135-143.

Goldschmidt-Clermont, M., Girard Bascou, J., Y., C., & D., R. J. (1990). Trans-splicing mutants of *Chlamydomonas reinhardtii*. *Mol. Gen. Genet., 223*, pp. 417–425.

Gray, M. W., Burger, G., & Lang, B. F. (1999). Mitochondrial evolution. *Science, 283*(5407), 1476-1481.

Groth, G., & Strotmann, H. (1999). New results about structure, function and regulation of the chloroplast ATP synthase (CF0CF1). *Physiologia Plantarum, 106*(1), 142-148.

Guertin, M., & Bellemare, G. (1979). Synthesis of chloroplast ribonucleic-acid in *Chlamydomonas-reinhardtii* toluene-treated cells. *European Journal of Biochemistry, 96*(1), 125-129.

Gunning, B. E. S. (2001). Membrane geometry of "open" prolamellar bodies. *Protoplasma, 215*(1-4), 4-15.

Gunning, B. E. S. (2004). Plant Cell Biology on CD, details at www.plantcellbiologyonCD.com.

Haddad, F., Qin, A. Q. X., Giger, J. M., Guo, H. Y., & Baldwin, K. M. (2007). Potential pitfalls in the accuracy of analysis of natural sense-antisense RNA pairs by reverse transcription-PCR. *Bmc Biotechnology, 7*.

Hajdukiewicz, P. T. J., Allison, L. A., & Maliga, P. (1997). The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *Embo Journal, 16*(13), 4041-4048.

Hakimi, M. A., Privat, I., Valay, J. G., & Lerbs-Mache, S. (2000). Evolutionary conservation of C-terminal domains of primary Sigma(70)-type transcription factors between plants and bacteria. *Journal of Biological Chemistry, 275*(13), 9215-9221.

Hammani, K., Okuda, K., Tanz, S. K., Chateigner-Boutin, A. L., Shikanai, T., & Small, I. (2009). A study of new arabidopsis chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell, 21*(11), 3686-3699.

Han, C. D., Coe, E. H., & Martienssen, R. A. (1992). Molecular-cloning and characterization of *iojap* (ij), a pattern striping gene of maize. *Embo Journal, 11*(11), 4037-4046.

Hanaoka, M., Kanamaru, K., Fujiwara, M., Takahashi, H., & Tanaka, K. (2005). Glutamyl-tRNA mediates a switch in RNA polymerase use during chloroplast biogenesis. *Embo Reports, 6*(6), 545-550.

Hansson, A., Amann, K., Zygadlo, A., Meurer, J., Scheller, H. V., & Jensen, P. E. (2007). Knock-out of the chloroplast-encoded PSI-J subunit of photosystem I in Nicotiana tabacum - PSI-J is required for efficient electron transfer and stable accumulation of photosystem I. *Febs Journal, 274*(7), 1734-1746.

Harris, J. B., & Arnott, H. J. (1973). Effects of senescence on chloroplasts of tobacco leaf. *Tissue & Cell, 5*(4), 527-544.

Harris, J. B., & Schaefer, V. G. (1981). Some correlated events in aging leaf tissues of tree tomato and tobacco. *Botanical Gazette, 142*(1), 43-54.

Hayes, Reed, M. L., Hegeman, C. E., & Hanson, M. R. (2006). Sequence elements critical for efficient RNA editing of a tobacco chloroplast transcript in vivo and in vitro. *Nucleic Acids Research, 34*(13), 3742-3754.

Hayes, R., Kudla, J., Schuster, G., Gabay, L., Maliga, P., & Gruissem, W. (1996). Chloroplast mRNA 3'-end processing by a high molecular weight protein complex is regulated by nuclear encoded RNA binding proteins. *Embo Journal, 15*(5), 1132-1141.

Hedtke, B., Borner, T., & Weihe, A. (1997). Mitochondrial and chloroplast phage-type RNA polymerases in Arabidopsis. *Science, 277*(5327), 809-811.

Hedtke, B., Borner, T., & Weihe, A. (2000). One RNA polymerase serving two genomes. *Embo Reports, 1*(5), 435-440.

Hedtke, B., Legen, J., Weihe, A., Herrmann, R. G., & Borner, T. (2002). Six active phage-type RNA polymerase genes in Nicotiana tabacum. *Plant Journal, 30*(6), 625-637.

Hedtke, B., Meixner, M., Gillandt, S., Richter, E., Borner, T., & Weihe, A. (1999). Green fluorescent protein as a marker to investigate targeting of organellar RNA polymerases of higher plants in vivo. *Plant Journal, 17*(5), 557-561.

Hernandez-Pinzon, I., Ross, J. H. E., Barnes, K. A., Damant, A. P., & Murphy, D. J. (1999). Composition and role of tapetal lipid bodies in the biogenesis of the pollen coat of *Brassica napus. Planta, 208*(4), 588-598.

Herrin, L. H., Kuo, T. C., & Goldschmidt-Clermont, M. (1998). The molecular biology of chloroplasts and mitochondria inChlamydomonas (RNA splicing in the chloroplast). 451–458.

Herrmann, R. G., Bohnert, H. J., Kowallik, K. V., & Schmitt, J. M. (1975). Size, conformation and purity of chloroplast dna of some higher-plants. *Biochimica Et Biophysica Acta, 378*(2), 305-317.

Hess, W. R., Prombona, A., Fieder, B., Subramanian, A. R., & Borner, T. (1993). Chloroplast rps15 and the rpoB/C1/C2 gene-cluster are strongly transcribed in ribosome-deficient plastids - evidence for a functioning non-chloroplast-encoded RNA-polymerase. *Embo Journal, 12*(2), 563-571.

Hicks, A., Drager, R. G., Higgs, D. C., & Stern, D. B. (2002). An mRNA 3 ' processing site targets downstream sequences for rapid degradation in Chlamydomonas chloroplasts. *Journal of Biological Chemistry, 277*(5), 3325-3333.

Hirose, T., & Sugiura, M. (2004). Multiple elements required for translation of plastid atpB mRNA lacking the Shine-Dalgarno sequence. *Nucleic Acids Research, 32*(11), 3503-3510.

Homann, A., & Link, G. (2003). DNA-binding and transcription characteristics of three cloned sigma factors from mustard (Sinapis alba L.) suggest overlapping and distinct roles in plastid gene expression. *European Journal of Biochemistry, 270*(6), 1288-1300.

Hongay, C. F., Grisafi, P. L., Galitski, T., & Fink, G. R. (2006). Antisense transcription controls cell fate in Saccharomyces cerevisiae. *Cell, 127*(4), 735-745.

Hoober, J. K., Boyd, C. O., & Paavola, L. G. (1991). Origin of thylakoid membranes in *Chlamydomonas-reinhardtii* Y-1 at 38-degrees-C. *Plant Physiology, 96*(4), 1321-1328.

Howe, C. J., Barbrook, A. C., Nisbet, R. E. R., Lockhart, P. J., & Larkum, A. W. D. (2008). The origin of plastids. *Philosophical Transactions of the Royal Society B-Biological Sciences, 363*(1504), 2675-2685.

Hsieh, K., & Huang, A. H. C. (2004). Endoplasmic reticulum, oleosins, and oils in seeds and tapetum cells. *Plant Physiology, 136*(3), 3427-3434.

Hudson, G. S., Mason, J. G., Holton, T. A., Koller, B., Cox, G. B., Whitfeld, P. R., et al. (1987). A gene-cluster in the spinach and pea chloroplast genomes encoding one CF1 and 3 CF0 subunits of the h+-ATP synthase complex and the ribosomal protein-S2. *Journal of Molecular Biology, 196*(2), 283-298.

Hurkman, W. J., & Kennedy, G. S. (1976). Fine-structure and development of proteoplasts in primary leaves of mung bean. *Protoplasma, 89*(1-2), 171-184.

Ikeuchi, M., Inoue, Y., & Vermaas, W. (1995). Characterization of photosystem II subunits from the cyanobacterium Synechocystis sp. PCC6803. In Phototsynthesis: from light to biospere. *Kluwer Academic Publisheres, Edited by Mathis, P. Vol. III,*, pp. 297-300.

Inada, M., Sasaki, T., Yukawa, M., Tsudzuki, T., & Sugiura, M. (2004). A systematic search for RNA editing sites in pea chloroplasts: an editing event causes diversification from the evolutionarily conserved amino acid sequence. *Plant and Cell Physiology, 45*(11), 1615-1622.

Ishizaki, Y., Tsunoyama, Y., Hatano, K., Ando, K., Kato, K., Shinmyo, A., et al. (2005). A nuclear-encoded sigma factor, Arabidopsis SIG6, recognizes sigma-70 type chloroplast promoters and regulates early chloroplast development in cotyledons. *Plant Journal, 42*(2), 133-144.

Jarrel, K. A., Dietrich, R. C., & Perlman, P. S. (1988). Group-II intron domain-5 facilitates a trans-splicing reaction. *Molecular and Cellular Biology, 8*(6), 2361-2366.

Jarrell, K. A., Peebles, C. L., Dietrich, R. C., Romiti, S. L., & Perlman, P. S. (1988). Group-II intron self-splicing - alternative reaction conditions yield novel products. *Journal of Biological Chemistry, 263*(7), 3432-3439.

Jenkins, B. D., Kulhanek, D. J., & Barkan, A. (1997). Nuclear mutations that block group II RNA splicing in maize chloroplasts reveal several intron classes with distinct requirements for splicing factors. *Plant Cell, 9*(3), 283-296.

Jensen, P. E., Bassi, R., Boekema, E. J., Dekker, J. P., Jansson, S., Leister, D., et al. (2007). Structure, function and regulation of plant photosystem I. *Biochimica Et Biophysica Acta-Bioenergetics, 1767*(5), 335-352.

Juneau, P., Le Lay, P., Boddi, A., Samson, G., & Popovic, R. (2002). Relationship between the structural and functional changes of the photosynthetic apparatus during chloroplast-chromoplast transition in flower bud of *Lilium longiflorum*. *Photochemistry and Photobiology, 75*(4), 377-381.

Kanamaru, K., Nagashima, A., Fujiwara, M., Shimada, H., Shirano, Y., Nakabayashi, K., et al. (2001). An Arabidopsis sigma factor (SIG2)-dependent expression of plastid-encoded tRNAs in chloroplasts. *Plant and Cell Physiology, 42*(10), 1034-1043.

Kapoor, S., Suzuki, J. Y., & Sugiura, M. (1997). Identification and functional significance of a new class of non-consensus-type plastid promoters. *Plant Journal, 11*(2), 327-337.

Kapoor, S., Wakasugi, T., Deno, H., & Sugiura, M. (1994). An atpE-specific promoter within the coding region of the atpB gene in tobacco chloroplast DNA. *Current Genetics, 26*(3), 263-268.

Kashino, Y., Koike, H., Yoshio, M., Egashira, H., Ikeuchi, M., Pakrasi, H. B., et al. (2002). Low-molecular-mass polypeptide components of a photosystem II preparation from the thermophilic cyanobacterium Thermosynechococcus vulcanus. *Plant and Cell Physiology, 43*(11), 1366-1373.

Kern, J., & Renger, G. (2007). Photosystem II: Structure and mechanism of the water : plastoquinone oxidoreductase. *Photosynthesis Research, 94*(2-3), 183-202.

Kim, M., Thum, K. E., Morishige, D. T., & Mullet, J. E. (1999). Detailed architecture of the barley chloroplast psbD-psbC blue light-responsive promoter. *Journal of Biological Chemistry, 274*(8), 4684-4692.

Kirk, J. T. O., & Tilney-Bassett, R. A. E. (1967). kirk,jto - plastids - their chemistry structure growth and inheritance. *Scientific American, 217*(3), 302-&.

Kishine, M., Takabayashi, A., Munekage, Y., Shikanai, T., Endo, T., & Sato, F. (2004). Ribosomal RNA processing and an RNase R family member in chloroplasts of Arabidopsis. *Plant Molecular Biology, 55*(4), 595-606.

Kobayashi, Y., Dokiya, Y., Kumazawa, Y., & Sugita, M. (2002). Non-AUG translation initiation of mRNA encoding plastid-targeted phage-type RNA polymerase in Nicotiana sylvestris. *Biochemical and Biophysical Research Communications, 299*(1), 57-61.

Kobayashi, Y., Dokiya, Y., & Sugita, M. (2001). Dual targeting of phage-type RNA polymerase to both mitochondria and plastids is due to alternative translation initiation in single transcripts. *Biochemical and Biophysical Research Communications, 289*(5), 1106-1113.

Kobayashi, Y., Dokiya, Y., Sugiura, M., Niwa, Y., & Sugita, M. (2001). Genomic organization and organ-specific expression of a nuclear gene encoding phage-type RNA polymerase in Nicotiana sylvestris. *Gene, 279*(1), 33-40.

Kolodner, R., & Tewari, K. K. (1972). Molecular-size and conformation of chloroplast deoxyribonucleic acid from pea leaves. *Journal of Biological Chemistry, 247*(19), 6355-&.

Kotera, E., Tasaka, M., & Shikanai, T. (2005). A pentatricopetide repeat protein is essential for RNA editing in chloroplasts. *Nature, 433*, 326–330.

Krause, K., Maier, R. M., Kofer, W., Krupinska, K., & Herrmann, R. G. (2000). Disruption of plastid-encoded RNA polymerase genes in tobacco: expression of only a distinct set of genes is not based on selective transcription of the plastid chromosome. *Molecular and General Genetics, 263*(6), 1022-1030.

Krishna, K. B., Joshi, M. K., Vani, B., & Mohanty, P. (1999). Structure-function correlation during the etioplast-chloroplast transition in cucumber cotyledonary leaves. *Photosynthetica, 36*(1-2), 199-212.

Kroeger, T. S., Watkins, K. P., Friso, G., van Wijk, K. J., & Barkan, A. (2009). A plant-specific RNA-binding domain revealed through analysis of chloroplast group II intron splicing. *Proceedings of the National Academy of Sciences of the United States of America, 106*(11), 4537-4542.

Lancer, H. A., Cohen, C. E., & Schiff, J. A. (1976). Changing ratios of phototransformable protochlorophyll and protochlorophyllide of bean seedlings developing in dark. *Plant Physiology, 57*(3), 369-374.

Lapidot, M., & Pilpel, Y. (2006). Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *Embo Reports, 7*(12), 1216-1222.

Larkum, A., Lockhart, P., & Howe, C. (2007a). The origin of plastids: A shopping bag model. *Photosynthesis Research, 91*(2-3), PS1815.

Larkum, A., Lockhart, P. J., & Howe, C. J. (2007b). Shopping for plastids. *Trends in Plant Science, 12*(5), 189-195.

Lebkuecher, J. G., Haldeman, K. A., Harris, C. E., Holz, S. L., Joudah, S. A., & Minton, D. A. (1999). Development of photosystem-II activity during irradiance of etiolated Helianthus (Asteraceae) seedlings. *American Journal of Botany, 86*(8), 1087-1092.

Leech, R. M. (1984). Chloroplast development in angiosperms: current knowledge and future prospects., . *In: Baker NR*

and Barber J (eds) Topics in Photosynthesis, Chloroplast Biogenesis, Elsevier Science, Amsterdam, Vol 5, pp 1–21.

Legen, J., Kemp, S., Krause, K., Profanter, B., Herrmann, R. G., & Maier, R. M. (2002). Comparative analysis of plastid transcription profiles of entire plastid chromosomes from tobacco attributed to wild-type and PEP-deficient transcription machineries. *Plant Journal, 31*(2), 171-188.

Lerbs-Mache, S. (1993). The 110-KDa polypeptide of spinach plastid DNA-dependent RNA-polymerase - single-subunit enzyme or catalytic core of multimeric enzyme complexes. *Proceedings of the National Academy of Sciences of the United States of America, 90*(12), 5509-5513.

Lerbs-Mache, S. (2011). Function of plastid sigma factors in higher plants: regulation of gene expression or just preservation of constitutive transcription? *Plant Mol. Biol*. In press.

Lilly, J. W., Maul, J. E., & Stern, D. B. (2002). The *Chlamydomonas reinhardtti* organellar genomes respond transcriptionally and post-transcriptionally to abiotic stimuli. *Plant Cell, 14*(11), 2681-2706.

Liu, B., & Troxler, R. F. (1996). Molecular characterization of a positively photoregulated nuclear gene for a chloroplast RNA polymerase sigma factor in *Cyanidium caldarium*. *Proceedings of the National Academy of Sciences of the United States of America, 93*(8), 3313-3318.

Ljubesic, N. (1972). Ultrastructural changes of plastids during the yellowing of the fruit of *Cucurbita pepo* var pyriformis. . *Acta Bot Croatia 31*, 47–53.

Lonetto, M., Gribskov, M., & Gross, C. A. (1992). The sigma-70 family - sequence conservation and evolutionary relationships. *Journal of Bacteriology, 174*(12), 3843-3849.

Loschelder, H., Schweer, J., Link, B., & Link, G. (2006). Dual temporal role of plastid sigma factor 6 in Arabidopsis development. *Plant Physiology, 142*(2), 642-650.

Lung, B., Zemann, A., Madej, M. J., Schuelke, M., Techritz, S., Ruf, S., et al. (2006). Identification of small non-coding RNAs from mitochondria and chloroplasts. *Nucleic Acids Research, 34*(14), 3842-3852.

Martin, W., Hoffmeister, M., Rotte, C., & Henze, K. (2001). An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle. *Biological Chemistry, 382*(11), 1521-1539.

Martin, W., & Muller, M. (1998). The hydrogen hypothesis for the first eukaryote. *Nature, 392*(6671), 37-41.

Martin, W., & Russell, M. J. (2003). On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 358*(1429), 59-83.

Meierhoff, K., Felder, S., Nakamura, T., Bechtold, N., & 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*(6), 1480-1495.

Meng, W. Z., Nicholson, R. H., Nathania, L., Pertzev, A. V., & Nicholson, A. W. (2008). New approaches to understanding double-stranded RNA processing by ribonuclease III: Purification and assays of homodimeric and heterodimeric forms of RNAse III from bacterial extremophiles and mesophiles. In *Rna Turnover in Bacteria, Archaea and Organelles* (Vol. 447, pp. 119-129).

Mereschkowsky, C. (1905). Uber natur und ursprung der chromatophoren im pflanzenreiche. *biol. centrabl, 25*, 593-604, 689-691.

Michel, F., Umesono, K., & Ozeki, H. (1989). Comparative and functional-anatomy of group-II catalytic introns - a review. *Gene, 82*(1), 5-30.

Miyagi, T., Kapoor, S., Sugita, M., & Sugiura, M. (1998). Transcript analysis of the tobacco plastid operon rps2/atpI/H/F/A reveals the existence of a non-consensus type II (NCII) promoter upstream of the atpI coding sequence. *Molecular and General Genetics, 257*(3), 299-307.

Mochizuki, T., Onda, Y., Fujiwara, E., Wada, M., & Toyoshima, Y. (2004). Two independent light signals cooperate in the activation of the plastid psbD blue light-responsive promoter in Arabidopsis. *Febs Letters, 571*(1-3), 26-30.

Møller, S. G. (2005). Plastids. *Annual Plant Reviews, Blackwell publishing, 13*.

Monde, R. A., Schuster, G., & Stern, D. B. (2000). Processing and degradation of chloroplast mRNA. *Biochimie, 82*(6-7), 573-582.

Morden, C. W., Wolfe, K. H., Depamphilis, C. W., & Palmer, J. D. (1991). Plastid translation and transcription genes in a nonphotosynthetic plant - intact, missing and pseudo genes. *Embo Journal, 10*(11), 3281-3288.

Moreira, D., & Lopez-Garcia, P. (1998). Symbiosis between methanogenic archaea and delta-proteobacteria as the origin of eukaryotes: The syntrophic hypothesis. *Journal of Molecular Evolution, 47*(5), 517-530.

Morikawa, K., Shiina, T., Murakami, S., & Toyoshima, Y. (2002). Novel nuclear-encoded proteins interacting with a plastid sigma factor, Sig1, in *Arabidopsis thaliana*. *Febs Letters, 514*(2-3), 300-304.

Mudd, E. A., Sullivan, S., Gisby, M. F., Mironov, A., Kwon, C. S., Chung, W. I., et al. (2008). A 125 kDa RNase E/G-like protein is present in plastids and is essential for chloroplast development and autotrophic growth in Arabidopsis. *Journal of Experimental Botany, 59*(10), 2597-2610.

Mullet, J. E. (1988). Chloroplast development and gene-expression. *Annual Review of Plant Physiology and Plant Molecular Biology, 39*, 475-502.

Mullet, J. E. (1993). Dynamic regulation of chloroplast transcription. *Plant Physiology, 103*(2), 309-313.

Mullet, J. E., Orozco, E. M., & Chua, N. H. (1985). Multiple transcripts for higher-plant rbcl and atpb genes and localization of the transcription initiation site of the rbcl gene. *Plant Molecular Biology, 4*(1), 39-54.

Nagashima, A., Hanaoka, M., Motohashi, R., Seki, M., Shinozaki, K., Kanamaru, K., et al. (2004a). DNA microarray analysis of plastid gene expression in an Arabidopsis mutant deficient in a plastid transcription factor sigma, SIG2. *Bioscience Biotechnology and Biochemistry, 68*(3), 694-704.

Nagashima, A., Hanaoka, M., Shikanai, T., Fujiwara, M., Kanamaru, K., Takahashi, H., et al. (2004b). The multiple-stress responsive plastid sigma factor, SIG5, directs activation of the psbD blue light-responsive promoter (BLRP) in Arabidopsis thaliana. *Plant and Cell Physiology, 45*(4), 357-368.

Neff, D., Tripathi, S., Middendorf, K., Stahlberg, H., Butt, H. J., Bamberg, E., et al. (1997). Chloroplast F0F1 ATP synthase imaged by atomic force microscopy. *Journal of Structural Biology, 119*(2), 139-148.

Newcomb, E. H. (1967). Fine structure of protein-storing plastids in bean root tips. *Journal of Cell Biology, 33*(1P1), 143-&.

Nishizawa, N., & Mori, S. (1989). Ultrastructure of the thylakoid membrane in tomato leaf chloroplast revealed by liquid-helium rapid-freezing and substitution-fixation method. *Plant and Cell Physiology, 30*(1), 1-7.

Noji, H., & Yoshida, M. (2001). The rotary machine in the cell, ATP synthase. *Journal of Biological Chemistry, 276*(3), 1665-1668.

Ogren, W. L. (1984). Photorespiration - pathways, regulation, and modification. *Annual Review of Plant Physiology and Plant Molecular Biology, 35*, 415-442.

Ohnishi, N., Kashino, Y., Satoh, K., Ozawa, S., & Takahashi, Y. (2007). Chloroplast-encoded polypeptide PsbT is involved in the repair of primary electron acceptor Q(A) of photosystem II during photoinhibition in Chlamydomonas reinhardtii. *Journal of Biological Chemistry, 282*(10), 7107-7115.

Ohnishi, N., & Takahashi, Y. (2008). Chloroplast-encoded PsbT is required for efficient biogenesis of photosystem II complex in the green alga Chlamydomonas reinhardtii. *Photosynthesis Research, 98*(1-3), 315-322.

Okuda, K., Chateigner-Boutin, A. L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F., et al. (2009). Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in arabidopsis chloroplasts. *Plant Cell, 21*(1), 146-156.

Okuda, K., Myouga, F., Motohashi, R., Shinozaki, K., & Shikanai, T. (2007). Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proceedings of the National Academy of Sciences of the United States of America, 104*(19), 8178-8183.

Okuda, K., Nakamura, T., Sugita, M., Shimizu, T., & Shikanai, T. (2006). A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *Journal of Biological Chemistry, 281*(49), 37661-37667.

Oldenburg, D. J., & Bendich, A. J. (2004). Changes in the structure of DNA molecules and the amount of DNA per plastid during chloroplast development in maize. *Journal of Molecular Biology, 344*(5), 1311-1330.

Onda, Y., Yagi, Y., Saito, Y., Takenaka, N., & Toyoshima, Y. (2008). Light induction of Arabidopsis SIG1 and SIG5 transcripts in mature leaves: differential roles of cryptochrome 1 and cryptochrome 2 and dual function of SIG5 in the recognition of plastid promoters. *Plant Journal, 55*(6), 968-978.

Orozco, E. M., Chen, L. J., & Eilers, R. J. (1990). The divergently transcribed rbcL and atpB genes of tobacco plastid DNA are separated by 19 base-pairs. *Current Genetics, 17*(1), 65-71.

Osteryoung, K. W., & Nunnari, J. (2003). The division of endosymbiotic organelles. *Science, 302*(5651), 1698-1704.

Ostheimer, G. J., Williams-Carrier, R., Belcher, S., Osborne, E., Gierke, J., & Barkan, A. (2003). Group II intron splicing factors derived by diversification of an ancient RNA-binding domain. *Embo Journal, 22*(15), 3919-3929.

Pacini, E., Taylor, P. E., Singh, M. B., & Knox, R. B. (1992). Development of plastids in pollen and tapetum of rye-grass, lolium-perenne L. *Annals of Botany, 70*(2), 179-188.

Palmer, J. D. (2003). The symbiotic birth and spread of plastids: How many times and whodunit? *Journal of Phycology, 39*(1), 4-11.

Park, H., Eggink, L. L., Roberson, R. W., & Hoober, J. K. (1999). Transfer of proteins from the chloroplast to vacuoles in Chlamydomonas reinhardtii (Chlorophyta): A pathway for degradation. *Journal of Phycology, 35*(3), 528-538.

Park, H., Kreunen, S. S., Cuttriss, A. J., DellaPenna, D., & Pogson, B. J. (2002). Identification of the carotenoid isomerase provides insight into carotenoid biosynthesis, prolamellar body formation, and photomorphogenesis. *Plant Cell, 14*(2), 321-332.

Parthier, B. (1988). Gerontoplasts - the yellow end in the ontogenesis of chloroplasts. *Endocytobiosis and Cell Research, 5*(2), 163-190.

Peebles, C. L., Benatan, E. J., Jarrell, K. A., & Perlman, P. S. (1987). Group-II intron self-splicing - development of alternative reaction conditions and identification of a predicted intermediate. *Cold Spring Harbor Symposia on Quantitative Biology, 52*, 223-232.

Perrin, R., Lange, H., Grienenberger, J. M., & Gagliardi, D. (2004). AtmtPNPase is required for multiple aspects of the 18S rRNA metabolism in Arabidopsis thaliana mitochondria. *Nucleic Acids Research, 32*(17), 5174-5182.

Pertzev, A. V., & Nicholson, A. W. (2006). Characterization of RNA sequence determinants and antideterminants of processing reactivity for a minimal substrate of Escherichia coli ribonuclease III. *Nucleic Acids Research, 34*(13), 3708-3721.

Pfalz, J., Bayraktar, O. A., Prikryl, J., & Barkan, A. (2009). Site-specific binding of a PPR protein defines and stabilizes 5 ' and 3 ' mRNA termini in chloroplasts. *Embo Journal, 28*(14), 2042-2052.

Privat, I., Hakimi, M. A., Buhot, L., Favory, J. J., & Lerbs-Mache, S. (2003). Characterization of Arabidopsis plastid sigma-like transcription factors SIG1, SIG2 and SIG3. *Plant Molecular Biology, 51*(3), 385-399.

Pyke, K. A. (1999). Plastid division and development. *Plant Cell, 11*(4), 549-556.

Pyke, K. A., & Leech, R. M. (1992). Chloroplast division and expansion is radically altered by nuclear mutations in arabidopsis-thaliana. *Plant Physiology, 99*(3), 1005-1008.

Repoila, F., & Darfeuille, F. (2009). Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. *Biology of the Cell, 101*(2), 117-131.

Richter, M. L., Hein, R., & Huchzermeyer, B. (2000). Important subunit interactions in the chloroplast ATP synthase. *Biochimica Et Biophysica Acta-Bioenergetics, 1458*(2-3), 326-342.

Ris, H., & Plaut, W. (1962). Ultra structure of DNA containing areas in the chloroplast of *chlamydomonas. . Cell Biology, 13*, 593-604.

Robert, R. W., & Kenneth, j. H. (2007). Advances in photosynthesis and respiration. 'The Structure and Function of Plastids'. *springer 23*.

Robertson, E. J., & Leech, R. M. (1995). Significant changes in cell and chloroplast development in young wheat leaves (Triticum-aestivum cv hereward) grown in elevated CO2. *Plant Physiology, 107*(1), 63-71.

Ross, J. H. E., Milanesi, C., Murphy, D. J., & Cresti, R. (2000). Rapid-freeze fixation and substitution improves tissue preservation of microspores and tapetum in *Brassica napus*. *Sexual Plant Reproduction, 12*(4), 237-240.

Sakai, A., Saito, C., Inada, N., & Kuroiwa, T. (1998). Transcriptional activities of the chloroplast-nuclei and proplastid-nuclei isolated from tobacco exhibit different sensitivities to tagetitoxin: Implication of the presence of distinct RNA polymerases. *Plant and Cell Physiology, 39*(9), 928-934.

Saldanha, R., Mohr, G., Belfort, M., & Lambowitz, A. M. (1993). Group-I and group-II introns. *Faseb Journal, 7*(1), 15-24.

Sambrook, & Russell. (2005). Northern blotting: transfer of denatured RNA to membranes. *Nature Methods, 2*(12), 997-998.

Sasaki, Y., Kozaki, A., Ohmori, A., Iguchi, H., & Nagano, Y. (2001). Chloroplast RNA editing required for functional acetyl-CoA carboxylase in plants. *Journal of Biological Chemistry, 276*(6), 3937-3940.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., & Tabata, S. (1999). Complete Structure of the Chloroplast Genome of Arabidopsis thaliana. *DNA Research, 6*, 283-290.

Schagger, H. (2006). Tricine-SDS-PAGE. *Nature Protocols, 1*(1), 16-22.

Schein, A., Sheffy-Levin, S., Glaser, F., & Schuster, G. (2008). The RNase E/G-type endoribonuclease of higher plants is located in the chloroplast and cleaves RNA similarly to the E-coli enzyme. *Rna-a Publication of the Rna Society, 14*(6), 1057-1068.

Schimper, A. F. W. (1883). Ueber die entwickelung der chlorophyllkoerner und farbkoerper. *Botanische Zeitung, 41*, 105-113.

Schmitz-Linneweber, C., & Barkan, A. (2007). RNA splicing and RNA editing in chloroplasts. . *In Cell and Molecular Biology of Plastids, ed. R Bock, Heidelberg, Germany: Springer*, pp. 213–248.

Schmitz-Linneweber, C., Williams-Carrier, R. E., Williams-Voelker, P. M., Kroeger, T. S., Vichas, A., & Barkan, A. (2006). A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast rps12 pre-mRNA. *Plant Cell, 18*(10), 2650-2663.

Schweer, J., Loschelder, H., & Link, G. (2006). A promoter switch that can rescue a plant sigma factor mutant. *Febs Letters, 580*(28-29), 6617-6622.

Schweer, J., Turkeri, H., Link, B., & Link, G. (2010). AtSIG6, a plastid sigma factor from Arabidopsis, reveals functional impact of cpCK2 phosphorylation. *Plant Journal, 62*(2), 192-202.

Serino, G., & Maliga, P. (1998). RNA polymerase subunits encoded by the plastid rpo genes are not shared with the nucleus-encoded plastid enzyme. *Plant Physiology, 117*(4), 1165-1170.

Shikanai, T. (2007). Cyclic electron transport around photosystem I: Genetic approaches. *Annual Review of Plant Biology, 58*, 199-217.

Shimizu, M., Kato, H., Ogawa, T., Kurachi, A., Nakagawa, Y., & Kobayashi, H. (2010). Sigma factor phosphorylation in the photosynthetic control of photosystem stoichiometry. *Proceedings of the National Academy of Sciences of the United States of America, 107*(23), 10760-10764.

Shinozaki, K., Deno, H., Kato, A., & Sugiura, M. (1983). Overlap and cotranscription of the genes for the beta-subunit and epsilon-subunit of tobacco chloroplast ATPase. *Gene, 24*(2-3), 147-155.

Shinozaki, K., & Sugiura, M. (1982). The nucleotide-sequence of the tobacco chloroplast gene for the large subunit of ribulose-1,5-bisphosphate carboxylase oxygenase. *Gene, 20*(1), 91-102.

Sluiman, H. J., & Lokhorst, G. M. (1988). The ultrastructure of cellular division (autosporogenesis) in the coccoid green alga, Trebouxia aggregate, revealed by rapid freeze fixation and freeze substitution. *Protoplasma, 144*(149–159).

Soll, J., & Schleiff, E. (2004). Protein import into chloroplasts. *Nature Reviews Molecular Cell Biology, 5*(3), 198-208.

Staehelin, L. A. (2003). Chloroplast structure: from chlorophyll granules to supra-molecular architecture of thylakoid membranes. *Photosynthesis Research, 76*(1-3), 185-196.

Stahl, D. J., Rodermel, S. R., Bogorad, L., & Subramanian, A. R. (1993). Co-transcription pattern of an introgressed operon in the maize chloroplast genome comprising 4 ATP synthase subunit genes and the ribosomal rps2. *Plant Molecular Biology, 21*(6), 1069-1076.

Stern, D. B., Goldschmidt-Clermont, M., & Maureen, R. H. (2010). Chloroplast RNA metabolism. *Annu Rev Plant Biol, Vol. 61*, 125-155.

Stern, D. B., & Kindle, K. L. (1993). 3' end maturation of the chlamydomonas-reinhardtii chloroplast-atpB messenger-RNA is a 2-step process. *Molecular and Cellular Biology, 13*(4), 2277-2285.

Stollar, N. E., & Hollingsworth, M. J. (1994). Expression of the large ATP synthase gene-cluster from spinach-chloroplasts. *Journal of Plant Physiology, 144*(2), 141-149.

Sugiura, M. (2008). RNA editing in chloroplasts. . *In RNA Editing. Nucleic Acids and Molecular Biology, ed. HU Goringer, Berlin Heidelberg: Springer-Verlag*, pp. 123–142.

Surzycki, S. J., & Shellenbarger, D. L. (1976). Purification and characterization of a putative sigma factor from *chlamydomonas-reinhardi. Proceedings of the National Academy of Sciences of the United States of America, 73*(11), 3961-3965.

Swiatecka-Hagenbruch, M., Liere, K., & Börner, T. (2007). High diversity of plastidial promoters in Arabidopsis thaliana. *Mol Genet Genomics, 277*, 725–734.

Tanaka, K., Oikawa, K., Ohta, N., Kuroiwa, H., Kuroiwa, T., & Takahashi, H. (1996). Nuclear encoding of a chloroplast RNA polymerase sigma subunit in a red alga. *Science, 272*(5270), 1932-1935.

Thomas, B. C., Li, X. Q., & Gegenheimer, P. (2000). Chloroplast ribonuclease P does not utilize the ribozyme-type pre-tRNA cleavage mechanism. *Rna-a Publication of the Rna Society, 6*(4), 545-553.

Thomson, W. W., & Whatley, J. M. (1980). Development of nongreen plastids. : . *Annu Rev Plant Physiol 31*, 375–394.

Tiller, K., & Link, G. (1993). Phosphorylation and dephosphorylation affect functional-characteristics of chloroplast and etioplast transcription systems from mustard (sinapis-alba l). *Embo Journal, 12*(5), 1745-1753.

Ting, J. T. L., Wu, S. S. H., Ratnayake, C., & Huang, A. H. C. (1998). Constituents of the tapetosomes and elaioplasts in Brassica campestris tapetum and their degradation and retention during microsporogenesis. *Plant Journal, 16*(5), 541-551.

Tozawa, Y., Teraishi, M., Sasaki, T., Sonoike, K., Nishiyama, Y., Itaya, M., et al. (2007). The plastid sigma factor SIG1 maintains photosystem I activity via regulated expression of the psaA operon in rice chloroplasts. *Plant Journal, 52*(1), 124-132.

Trepp, G. B., Plank, D. W., Gantt, J. S., & Vance, C. P. (1999). NADH-glutamate synthase in alfalfa root nodules. Immunocytochemical localization. *Plant Physiology, 119*(3), 829-837.

Troxler, R. F., Zhang, F., Hu, J., & Bogorad, L. (1994). Evidence that sigma-factors are components of chloroplast RNA-polymerase. *Plant Physiology, 104*(2), 753-759.

Tsunoyama, Y., Ishizaki, Y., Morikawa, K., Kobori, M., Nakahira, Y., Takeba, G., et al. (2004). Blue light-induced transcription of plastid-encoded psbD gene is mediated by a nuclear-encoded transcription initiation factor, AtSig5. *Proceedings of the National Academy of Sciences of the United States of America, 101*(9), 3304-3309.

Turmel, M., Gutell, R. R., Mercier, J. P., Otis, C., & Lemieux, C. (1993). Analysis of the chloroplast large subunit ribosomal-rna gene from 17 chlamydomonas taxa - 3 internal transcribed spacers and 12 group-I intron insertion sites. *Journal of Molecular Biology, 232*(2), 446-467.

Umate, P., Schwenkert, S., Karbat, I., Dal Bosco, C., Mlcochova, L., Volz, S., et al. (2007). Deletion of PsbM in tobacco alters the Q(B) site properties and the electron flow within photosystem II. *Journal of Biological Chemistry, 282*(13), 9758-9767.

Vanderveen, R., Kwakman, J., & Grivell, L. A. (1987). Mutations at the lariat acceptor site allow self-splicing of a group-II intron without lariat formation. *Embo Journal, 6*(12), 3827-3831.

Vogel, J., Borner, T., & Hess, W. R. (1999). Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants. *Nucleic Acids Research, 27*(19), 3866-3874.

von Wettstein, D. (2001). Discovery of a protein required for photosynthetic membrane assembly. *Proceedings of the National Academy of Sciences of the United States of America, 98*(7), 3633-3635.

Wellburn, A. R., & Hampp, R. (1979). Appearance of photochemical function in prothylakoids during plastid development. *Biochim Biophys Acta, 547*, 380–397.

Werner, A., & Swan, D. (2010). What are natural antisense transcripts good for? *Biochemical Society Transactions, 38*, 1144-1149.

Westhoff, P., & Herrmann, R. G. (1988). Complex RNA maturation in chloroplasts - The psbB operon from spinach. *European Journal of Biochemistry, 171*(3), 551-564.

Whatley, J. M. (1977). Variations in basic pathway of chloroplast development. *New Phytologist, 78*(2), 407-&.

Woessner, J. P., Gillham, N. W., & Boynton, J. E. (1987). Chloroplast genes encoding subunits of the H+-ATPase complex of *chlamydomonas-reinhardtii* are rearranged compared to higher-plants - sequence of the atpE gene and location of the atpF and atpI genes. *Plant Molecular Biology, 8*(2), 151-158.

Wollman, F. A., Minai, L., & Nechushtai, R. (1999). The biogenesis and assembly of photosynthetic proteins in thylakoid membranes. *Biochimica Et Biophysica Acta-Bioenergetics, 1411*(1), 21-85.

Wosten, M. (1998). Eubacterial sigma-factors. *Fems Microbiology Reviews, 22*(3), 127-150.

Wu, S. S. H., Platt, K. A., Ratnayake, C., Wang, T. W., Ting, J. T. L., & Huang, A. H. C. (1997). Isolation and characterization of neutral-lipid-containing organelles and globuli-filled plastids from Brassica napus tapetum. *Proceedings of the National Academy of Sciences of the United States of America, 94*(23), 12711-12716.

Yang, J. J., & Stern, D. B. (1997). The spinach chloroplast endoribonuclease CSP41 cleaves the 3'-untranslated region of petD mRNA primarily within its terminal stem-loop structure. *Journal of Biological Chemistry, 272*(19), 12874-12880.

Yao, J. L., Roy-Chowdhury, S., & Allison, L. A. (2003). AtSig5 is an essential nucleus-encoded Arabidopsis sigma-like factor. *Plant Physiology, 132*(2), 739-747.

Yazgan, O., & Krebs, J. E. (2007). Noncoding but nonexpendable: transcriptional regulation by large noncoding RNA in eukaryotes. *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire, 85*(4), 484-496.

Yehudai-Resheff, S., Zimmer, S. L., Komine, Y., & Stern, D. B. (2007). Integration of chloroplast nucleic acid metabolism into the phosphate deprivation response in Chlamydomonas reinhardtii. *Plant Cell, 19*(3), 1023-1038.

Yeum, K. J., & Russell, R. M. (2002). Carotenoid bioavailability and bioconversion. *Annual Review of Nutrition, 22*, 483-504.

Yukawa, M., & Sugiura, M. (2008). Termination codon-dependent translation of partially overlapping ndhC-ndhK transcripts in chloroplasts. *Proceedings of the National Academy of Sciences of the United States of America, 105*(49), 19550-19554.

Zghidi, W., Merendino, L., Cottet, A., Mache, R., & Lerbs-Mache, S. (2007). Nucleus-encoded plastid sigma factor SIG3 transcribes specifically the psbN gene in plastids. *Nucleic Acids Research, 35*(2), 455-464.

Zito, F., Kuras, R., Choquet, Y., Kossel, H., & Wollman, F. A. (1997). Mutations of cytochrome b(6) in *Chlamydomonas reinhardtii* disclose the functional significance for a proline to leucine conversion by petB editing in maize and tobacco. *Plant Molecular Biology, 33*(1), 79-86.

Zoschke, R., Nakamura, M., Liere, K., Sugiura, M., Borner, T., & Schmitz-Linneweber, C. (2010). An organellar maturase associates with multiple group II introns. *Proceedings of the National Academy of Sciences of the United States of America, 107*(7), 3245-3250.

the plant journal

Multiple regulatory mechanisms are implicated in the expression of the two plastid ATP synthase operons

scholarONE™
Manuscript Central

Multiple regulatory mechanisms are implicated in the expression of the two plastid ATP synthase operons

Mustafa Malik Ghulam[1], Livia Merendino[1]*, Ouafa Zghidi-Abouzid, Emeline Lambert and Silva Lerbs-Mache*

Laboratoire de Physiologie Cellulaire Végétale, Université Joseph Fourier, Centre National de la Recherche Scientifique, CEA-Grenoble, UMR 5168, 38054 Grenoble cedex 9, France

[1]The first two authors participated equally to the work

*co-corresponding authors

*corresponding author during the submission process:

Silva Lerbs-Mache, Tel.: (33) (0)4 76 63 57 44; Fax: (33) (0)4 76 63 55 86;

E-mail: Silva.Lerbs-Mache@ujf-grenoble.fr

E-mail adresses : mgmustafa2005@hotmail.com, livia.merendino@ujf-grenoble.fr, ouafa.zghidi-abouzid@insa-lyon.fr, emeline.lambert@ujf-grenoble.fr

Running title : transcription of plastid ATP synthase genes

Key words : arabidopsis, chloroplast, gene expression, regulation, ATP synthase, sigma factors

Total word count : 6 901

Summary : 245, Introduction : 955, Results : 2 247, Discussion : 1 282, Materials and Methods : 940, Figure legends : 1 232, References : 1 025

1

SUMMARY

We have analysed the expression of the two plastid ATP synthase operons *atpI/atpH/atpF/atpA* and *atpB/atpE* in arabidopsis wild type plants and different sigma factor mutants (*sig1, sig2, sig3, sig4*). Results indicate regulation on the transcriptional and post-transcriptional level. Both operons are under control of SIG2 and SIG3 and could be coregulated *via* these two sigma factors. Transcription of the large ATP synthase gene cluster starts either at the *atpI* gene that is under SIG2 control or from a SIG3 specific promoter that is located in the intergenic region between *atpI* and *atpH*. Besides existing as (atpI)/atpH/atpF/atpA cotranscript, the *atpH* mRNA exists as monocistronic RNA that accumulates to rather high levels. Two different mechanisms might contribute to the accumulation of the *atpH* mRNA: a) Specific transcription of the *atpH* gene, starting at the SIG3 specific *atpH* promoter and terminating at a potential terminator that is located at the end of the *atpH* gene. b) Stabilisation of processed *atpH* transcripts originating either from SIG3/*atpH* initiated polycistronic transcripts or from SIG2/*atpI* initiated transcripts. Regarding the dicistronic *atpB/atpE* operon, two PEP promoters are found upstream of the *atpB* gene. One is recognized by SIG2/PEP holoenzyme as well as by SIG3/PEP. *AtpE* transcription could be decoupled from *atpB* transcription by an *atpB* internal, SIG2 dependent, promoter. However, the *atpE* mRNA is also translated from dicistronic *atpB/atpE* mRNA for which overlapping stop/start codons have been described. This is shown by the existence of the ATPE protein in *SIG2* knock-out plants.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCTION

H^+-ATPases (ATP synthases) catalyse proton transport coupled with ATP synthesis. They exist in bacteria, chloroplasts and mitochondria. ATP synthases are composed of the membrane integrated F_0 complex (CF_0) and the surface exposed F_1 complex (CF_1). The CF_1 part consists of five different subunits (α, β, γ, δ, ϵ) present in a 3:3:1:1:1 stoichiometry. The CF_0 part consists of four different subunits (I, II, III, IV) with a stoichiometry defined as 1:1:6-12:1 (Böttcher and Gräber, 2000). In photosynthetic organisms subunit α (*atpA*), β (*atpB*), ϵ (*atpE*), I (*atpF*), III (*atpH*) and IV (*atpI*) are plastid encoded. The question of how the stoichiometry is achieved is not yet completely answered. Considerable advance to solve this problem has been made with *Chlamydomonas reinhardtii*, a unicellular green alga that is amenable to a diversity of genetic manipulations. Studies with mutants have shown that assembly controls the stability of newly synthesized ATPase polypeptides (Choquet and Vallon, 2000) as well as translation rates (Drapier et al., 2007). Generally, post-transcriptional/translational regulation represents the prevailing mode of gene expression regulation in Chlamydomonas (Barkan et al., 2000). This is probably connected to the fact that only one RNA polymerase and only one sigma factor exist in chloroplasts (Bohne et al., 2006) thus limiting the option of transcriptional regulation.

Unlike Chlamydomonas, the transcriptional apparatus of the plastid genome of higher plants is more complex. Transcription is performed by two different transcription systems, either plastid or nucleus encoded. The nucleus encoded plastid RNA polymerase (NEP) is monomeric and of the phage type. The plastid encoded RNA polymerase (PEP) is a multimeric, eubacterial type, RNA polymerase that is

complemented with nucleus encoded transcription initiation factors of the sigma type.

Six different genes coding sigma factors have been revealed in arabidopsis (Shiina et al., 2005; Liere et al., 2006). SIG4 has been proposed to determine the quantity of NDH complexes by regulating the expression of the *ndhF* gene (Favory et al. 2005). SIG5 has been shown to regulate blue-light dependent expression of the *psbD* gene (Tsunoyama et al., 2004). These and other examples indicate that transcriptional regulation plays an important role in plastid gene expression of higher plants. Our recent analyses of plastid gene expression patterns in arabidopsis WT and *sig3* T-DNA insertion mutants showed, beside a strong reduction of *psbN* mRNA, moderate reduction of several mRNAs of the ATP synthase complex in the *SIG3* mutant (Zghidi et al., 2007). Among these, the *atpH* mRNA was most reduced, indicating that also transcriptional regulation, especially by sigma factor 3 (SIG3), participates to the establishment of the observed stoichiometry of the CF_0 subunits at the protein level.

In higher plants the plastid-encoded ATP synthase genes are localized in two different transcriptional units. One of them comprises *rps2, atpI, atpH, atpF* and *atpA* (Henning and Herrmann, 1986; Stahl et al., 1993; Stollar and Hollingsworth, 1994a; Miyagi et al., 1998). The other transcription unit contains the *atpB* and *atpE* genes (Zurawski et al., 1982; Shinozaki et al., 1983). The *atpB* promoter region is not well conserved between different plant species. For instance, in tobacco the *atpB* operon is transcribed from three PEP promoters and one NEP promoter (Orozco et al., 1990; Hajdukiewicz et al., 1997), while in maize and in arabidopsis only one NEP and one PEP promoter have been described (Silhavy and Maliga, 1998; Swiatecka-Hagenbruch *et al.*, 2007). Independent of the exact number of transcription start sites, the presence of several promoters suggests expression regulation of the *atpB* operon on the

transcriptional level by differential promoter usage. The same holds true for the large

ATP synthase operon where different transcription start sites have been determined,

located upstream of the *rps2*, *atpI* and/or *atpH* genes (Stahl et al., 1993; Stollar et al.,

1994a; Miyagi et al., 1998). The *atpI* gene of tobacco is transcribed from one NEP and

one PEP promoter (Miyagi et al., 1998). In arabidopsis, the *atpI* gene is solely

transcribed from one PEP promoter (Swiatecka-Hagenbruch et al., 2007).

The expression of both operons is regulated by several mechanisms. The two

genes (*atpB* and *atpE*) of the small ATP synthase operon have overlapping translation

stop/start signals (Zurawski et al., 1982). They are transcribed as dicistronic mRNA for

which translational coupling has been demonstrated in *E. coli* (Gatenby et al., 1998).

Whether this mechanism functions also in higher plants is not clear because an *atpE*

specific promoter is localized within the coding region of the *atpB* gene and

transcriptional coupling is not necessary for *atpE* expression (Kapoor et al., 1994).

Northern analysis shows the existence of two dicistronic transcripts of different sizes,

one of them is under control of SIG6 (Schweer et al., 2006).

In the case of the large ATP synthase gene cluster, the different transcripts

starting either upstream of *rps2*, *atpI* and/or *atpH* terminate downstream of *atpA* (Stahl

et al., 1993; Stollar et al., 1994a; Miyagi et al., 1998). Co-transcription of *atpH, atpF*

and *atpA* should result in an initial RNA ratio of 1:1:1 for the three genes and the

question arises how the protein stoichiometry of 6-12 molecules ATPH against 1

molecule ATPF in the CF_0 complex is achieved. Different mechanisms like ribosome

pausing and cleavage-induced alteration of mRNA stability have been proposed as

explanation (Stollar et al., 1994b; Hotchkiss and Hollingsworth, 1997).

In the present paper, we have analysed the expression of the two ATP synthase operons in detail by using different sigma factor knock-out mutants and by paying special attention to the question of whether transcription of the *atpH* gene by SIG3/PEP holoenzyme might contribute to the stoichiometry of the CF_0 complex. We show that both ATP synthase operons are under control of SIG2 and SIG3 and that several mechanisms like differential promoter usage, mRNA stability (*atpH*) and translational coupling (*atpB*/*atpE*) contribute to expression regulation.

RESULTS

Analyses of *atp* mRNA levels by microarray hybridisation

ATP mRNA levels were quantified using data obtained with our recently developed plastid microarray. Figure 1a, left hand side, shows the hybridisation signals of all plastid genes coding ATP synthase subunits. Signal quantification and standard deviations obtained from three independent experiments are diagrammed on the right hand side. If the *atpI* RNA level is set to 1 the stoichiometry of the different mRNAs is 1(*atpI*):16(*atpH*):0,5(*atpF*):0,6(*atpA*):0,4(*atpB*):0,6(*atpE*), *i. e.* the level of *atpH* mRNA corresponds well to the high number of ATPH subunits in the CF_0 complex. The observed mRNA pattern suggests that *atpH* gene expression is specifically regulated and we wanted to know whether the recently described SIG3 dependent *atpH* promoter plays a significant role in achieving this high *atpH* mRNA level. Also, *atpE* mRNA seems to be more abundant than *atpB* mRNA. To better understand these differences in mRNA accumulation we have analysed the expression of the large and the small ATP

synthase operon genes (schematically presented in Figures 1b and 6a, respectively) in arabidopsis WT plants and several sigma factor knock-out mutants by using different methods like primer extension, 5' and 3'end mapping, Northern hybridisation and Western immunoblotting.

Analysis of *atpI* transcripts

Transcription initiation sites of the arabidopsis *atpI* gene have been mapped recently using either green leaves from normally grown plants or chlorophyll-deficient plantlets grown in the presence of spectinomycin (Swiatecka-Hagenbruch et al., 2007). Two different transcripts having 5'ends at positions -229 and -225 had been detected in RNA preparations from green leaves only, suggesting that both transcripts were made from PEP dependent promoters.

In order to confirm transcription by PEP and to determine sigma factors that are engaged in *atpI* transcription we performed primer extension analyses using RNAs of WT plants grown in the presence of spectinomycin and *SIG1, SIG2, SIG3* and *SIG4* mutants (Figure 1c). We could not clearly distinguish two different transcripts, but transcripts are not present in spectinomycin treated plantlets (compare lanes 5 and 6) confirming that transcription initiation occurs exclusively at PEP promoters. The analyses of arabidopsis *sig1, sig2, sig3* and *sig4* plants show strong diminution of *atpI* transcripts only in *sig2* plants (lanes 7-12). This indicates a special importance of sigma factor 2 for *atpI* transcription.

Analysis of *atpH* transcripts

 Our previous analyses revealed two different *atpH* transcripts, ending at positions -45 and -413 relative to the ATG translation initiation codon (Zghidi et al., 2007). We had further shown that the -413 transcript results from transcription initiation by SIG3/PEP holoenzyme. In order to know more on the synthesis of the two different transcripts, we have analysed these two RNAs by primer extension during germination and early seedling development and we have verified the origin of the -45 transcript by 5'-RACE without and after treatment of mRNAs with tobacco acid pyrophosphatase (TAP). Results show that the -45 transcript accumulates before the -413 transcript during early plant development (Figure 2b, lanes 1-5) and that only the -413 transcript results from transcription initiation (Figure 2b, lanes 6-9), *i. e.* the -45 transcript should result mainly from cotranscription of *atpH* with *atpI*, cleavage occurring rapidly after transcription. This assumption is supported by the fact that the -45 *atpH* transcript is only slightly diminished in *sig3* plants compared to WT (Figure 2c, compare lanes 1 and 4).

The quantitative relation between the -45 and the -413 RNAs is much in favour of the -45 transcripts, indicating rather low transcription from the -413 promoter and a strong stabilisation of -45 processed RNAs. In *sig2* plants both *atpH* transcripts accumulate slightly, pointing to an enhancement of transcription from the -413 SIG3 specific *atpH* promoter in order to compensate for the lack of *atpI* initiated transcripts (Figure 2c, compare lanes 3 and 5). Sequencing of the 5'-RACE product of the -413 transcript shows that the RNA is 5 nucleotides longer than estimated from the primer extension results, *i. e.* the transcription starts with the G nucleotide located at position -418 upstream of the ATG translation start codon (not shown). From now, we will name the -413 transcript as -418 transcript.

Analysis of *atpF* transcripts

Next we have analysed the expression of the *atpF* gene by primer extension using primer 2 (Figure 3a). Three different transcripts are revealed (Figure 3b). One ends up in the *atpH*/*atpF* intergenic region, at position -30 from the *atpF* translation initiation codon. The two longer transcripts are too long to be mapped on a 6% polyacryamide sequencing gel. It is very likely that they correspond to *atpH*/*atpF* cotranscription, starting at positions -418 or -45 of the *atpH* gene. This assumption is supported by the facts that the longest transcript (supposed to be the -418 transcript) is absent in *sig3* (compare lanes 2 and 3 of Figure 3b). In addition, this transcript is a primary transcript as shown by 5'-RACE (lanes 6 and 7). The smaller one of the two large transcripts (supposed to be the -45 transcript) is only slightly diminished in *sig3* (lanes 2 and 3) and originates from RNA processing (lanes 6 and 7). Both transcripts disappear after spectinomycin treatment, indicating that they are made by the eubacterial-type RNA polymerase, PEP (Figure 3b, lanes 4 and 5).

Interestingly, the 5'end of the -30 *atpF* transcript (Figure 3b, lanes 8-12) is localised within the stem of a hypothetical hairpin structure. RNA cleavage at position -30 would liberate an *atpH* mRNA with a stretch of 10 successive A residues near its 3'end (Figure 3c). Such oligo(A) tails are believed to create 'loading pads' for nucleases that will then initiate mRNA degradation. Thus, cleavage of the *atpH*/*atpF* cotranscript at position -30 might direct the *atpH* mRNA into the polyadenylation-assisted RNA degradation pathway that acts in prokaryotes and organelles (Slomovic et al., 2006). The -30 transcript is reduced in the *SIG3* mutant (Figure 3b, lanes 2 and 3) indicating that a considerable part of the SIG3 initiated transcripts is cleaved at this position. We were not able to reveal the -30 *atpF* transcript by 5'-RACE but we could confirm its

existence by primer extension using a second, different, primer (Figure 3a, primer 1 and Figure 3b, lanes 8-12).

Next, we have mapped 3'ends of *atpH* transcripts by self-ligation of 5'-3'ends followed by sequencing of RT-PCR products of the circularised RNAs (Perrin et al., 2004). Especially, we wanted to know whether -418 and -45 *atpH* mRNAs have the same or different 3'ends. For this reason, two different 5'primers have been used in order to distinguish between -45 and -418 *atpH* circularised transcripts. Figure 4 shows the predicted secondary structure of the entire *atpH*/*atpF* intergenic nucleotide sequence (rnafold; http://bioweb2.pasteur.fr). Besides the -30 stem-loop structure another region of dyad symmetry is found immediately downstream of the *atpH* coding region (labelled as T). This structure is followed by a poly(U) tract, thus representing a structure reminiscent for transcription termination signals (Chen et al., 1995). We expected to find two different 3'ends of *atpH* mRNAs corresponding to these two structures.

Ten clones have been analysed having 5'ends at position -418 and 8 clones having 5'ends at position -45. No clone was obtained showing other 5'ends than at the -418 and the -45 position. All of the -418 clones had different 3'ends; three of them ended at different positions within the reading frame of *atpH* mRNA and two ended at different sites in the reading frame of the *atpF* mRNA (not shown). This 3'end dispersion indicates rapid degradation of -418 SIG3/PEP initiated transcripts. The 3'ends of the remaining 5 clones are located within the *atpH*/*atpF* intergenic region. These ends are indicated in Figure 4 by arrows and as (-418). In contrast to the -418 transcripts, the 3'ends of the -45 transcripts are less randomly distributed. Only one of the -45 mRNA 3'ends is found within the reading frame of the *atpH* mRNA (not

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

shown). Seven transcripts end within the *atpH*/*atpF* intergenic region, six of them, *i. e.* 75%, end at the same position (Figure 4, *). We could not detect any clone whose 3'end corresponds to the -30 *atpF* cleavage site neither to the supposed *atpH* terminator.

Analysis of *atpA* transcripts

Primer extension analysis of *atpA* transcripts reveals two different transcripts. One of them is much longer than 500 bases and ends within or upstream of the *atpF* coding region. This transcript should result from cotranscription of *atpA* and *atpF* (Figure 5a, lane 1). The smaller transcript locates near the 3'end of the *atpF* gene. Both transcripts are reduced in the *sig3* indicating their origin from the -418 *atpH* promoter as well as from the *atpI* promoter (Figure 5a, lanes 1 and 2). 5'-RACE of *atpA* transcripts reveals several transcripts, all of them result from processing (Figure 5a, lanes 4 and 5). Sequencing of these RACE products show that 11 out of 19 transcripts (i. e. 58 %) end up with the first nucleotide of the last codon of the *atpF* coding region. The 5'ends of the other transcripts are scattered over the *atpF*/*atpA* intergenic region (not shown).

To confirm cotranscription of *atpH*/*atpF*/*atpA* and to design processing intermediates of the transcripts we performed Northern analysis of WT and *sig3* RNAs using an *atpH* specific probe (Figure 5b, lanes 1 and 2). The ~3500 bases RNA corresponds in length to *atpH*/*atpF*/*atpA* cotranscripts (schematically demonstrated in Figure 5c). This is confirmed by hybridization with an *atpA* specific probe (lane 3). The *atpF* intron should already have been removed from the transcripts because hybridisation with an intron specific probe does not reveal this ~3500 bases transcript (Figure 5b, lane 4). Processing intermediates of the large ATP synthase gene cluster are schematically presented in Figure 5c. Results are conceivable by most of the *atpH*

mRNAs ending either at the supposed terminator or at the 3'end that had been determined by circular RT-PCR (Figure 4, *).

Analysis of transcripts of the *atpB/atpE* operon

Northern analysis of *atpB* operon transcripts reveals three different RNAs of about 2.6; 2.0 and 0.7 kb (Schweer et al., 2006; schematically represented in Figure 6a). Primer extension analysis of the *atpB* transcripts shows three different transcripts ending at positions -520, -462 and -84 apart from the *atpB* translation initiation codon (Figure 6b, left hand side, lane 5 and right hand side, lane 1). The -318 NEP transcript, that has been recently described in green control and spectinomycin-treated chlorophyll-deficient tissues (Swiatecka-Hagenbruch et al., 2007), is not detectable. To confirm our result, we analysed also spectinomycin-treated plant material (Figure 6b, left hand side, lanes 7 and 8). All three transcripts (-520, -462 and -84) diminish in spectinomycin-treated plantlets, showing that they originate from PEP. After long exposure, a faint band is visible that could correspond to the -318 RNA (not shown). The differences in the abundance of this transcript might be due to the different physiological stages that have been analysed in the two experiments (one and three weeks old plants).

5'-RACE experiments performed after and without TAP treatment show that the two larger transcripts (-520 and -462) result from transcription initiation (Figure 6b, right hand side, lanes 7 and 8). The -84 transcript represents a processing product (Figure 6b, left hand side, lanes 9 and 10). The -462 transcript has not yet been described before. This transcript is diminished in the *SIG3* mutant, indicating that SIG3 plays a role in transcription from the -462 *atpB* promoter (compare lanes 5 and 6 on the left hand side and lanes 1 and 2 on the right hand side). The -84 transcript is also

The Plant Journal

diminished in *sig3* plants suggesting that this transcript results to some extent from cleavage of the -462 transcript. Taken together, we suggest that the 2.6 kb RNA that is observed in the Northern experiment, corresponds to both, the -520 and the -462 transcripts. These transcripts cannot be distinguished by Northern. The -84 transcript should correspond to the 2.0 kb transcript that has been described from Northern studies.

Analysis of the -520 and -462 transcripts using *sig1* to *sig4* plants shows that the -462 transcript is also under strong control of SIG2 (Figure 6c, compare lanes 3 and 5), i. e. the -462 promoter is recognized by SIG2/PEP as well as by SIG3/PEP holoenzyme. At least one of these two promoters should be also recognized by SIG6/PEP holoenzyme because the 2.6 kb RNA diminishes in *sig6* plants (Schweer et al., 2006).

Analysis of *atpE* monocistronic transcripts

Primer extension analysis of *atpE* mRNAs reveals one transcript at a position around -431 relative to the *atpE* translation start codon (Figure 7a, lane 1). Such a transcript has previously been described in tobacco (Kapoor et al., 1994). This transcript is made by transcription initiation as shown by 5'-RACE (lanes 7 and 8) and sequencing of the RACE product determines the 5'end at position -431, inside the coding region of the *atpB* gene. This monocistronic *atpE* RNA is absent in spectinomycin treated plant material (Figure 7a, lane 2), *i. e.* this RNA is made by PEP. More precisely, this RNA is made by SIG2/PEP holoenzyme as shown by primer extension analysis of *sig1* to *sig4* plants (Figure 7b, lanes 1-6). Transcription by SIG2/PEP holoenzyme from this *atpB* internal *atpE* promoter is also confirmed by Northern analyses as revealed by the absence of the 0.7 kb RNA in *sig2* plants (Figure 7c). The protein level of ATPE is

13

reduced in *sig2* when compared to WT plants (Figure 7d). However, the protein is not completely absent. This shows that the *atpE* mRNA is translated *in vivo* from the *atpB/atpE* cotranscript, *i. e.* the overlapping stop/start signals do not prevent translation.

DISCUSSION

We have analysed the expression of the two plastid encoded ATP synthase operons, *atpI/atpH/atpF/atpA* and *atpB/atpE* in WT and *SIG1, 2, 3* and *4* mutants of *Arabidopsis thaliana*. We have been interested in these studies for two different reasons. First of all, microarray analyses of the plastid transcriptome showed a much higher level of *atpH* transcripts when compared with the mRNAs of the other plastid encoded subunits of ATP synthase. This suggests specific/differential regulation of *atpH* gene expression when compared to the other ATP synthase genes. Secondly, with the exception of the *atpI* mRNA, all ATP synthase mRNAs were diminished in *SIG3* mutants when compared to WT plants, suggesting a SIG3 coordinated regulation of the transcription of the two ATP synthase operons (Zghidi et al., 2007).

By transcript quantification, we observe a 16-fold higher *atpH* mRNA level compared to the other mRNAs of the large ATP synthase operon (Figure 1a). The *atpH* gene codes for subunits III and the *atpI/atpF* genes code for subunits I and IV of the CF_0 complex. Our results show that the observed protein stoichiometry of the CF_0 complex, defined as 1:1:6-12:1 for subunits I, II, III and IV respectively, is also approximately found on the mRNA level in arabidopsis. Results show further that the large ATP

14

synthase operon is transcribed from two different PEP promoters, one is under control of SIG2 (*atpI*-229/-225, Figure 1c). The other promoter (*atpH*-418) is specifically recognized by a PEP/SIG3 holoenzyme (Figure 2b and c). During early seedling development, the -45 *atpH* mRNA accumulates before the -418 *atpH* transcript (Figure 2b). Both transcripts (*atpI*-229 and *atpH*-418 initiated) are cleaved at position -45 upstream of the *atpH* translation start codon (Figures 2c and 3b). In the absence of SIG2 the -229 *atpI* transcript disappears (Figure 1c, compare lanes 11 and 9) but the -418 *atpH* mRNA and the -45 *atpH* cleavage product accumulate slightly (Figure 2c, compare lanes 3 and 5). This suggests a higher activity of SIG3 in *sig2* plants, a result that might reflect the raise of SIG3 protein in SIG2 under-expressing plants (Privat et. al., 2003).

The -45 *atpH* processing intermediate should be more stable than the -418 primary transcript as suggested from the quantitative difference between these two transcripts. Specific stabilisation of *atpH* mRNA by PPR10 protein has recently been described in maize (Pfalz et al., 2009). Interestingly, the stabilizing PPR10 protein was shown to bind to the 5'terminus of the *atpH* mRNA in the region between -46 and -20 upstream of the translation initiation codon, *i. e.* it should only protect the -45 *atpH* mRNA, but not the -418 mRNA. Our analyses to determine 5' and 3'ends of *atpH* mRNA termini (Figures 2 and 4) are in good agreement with such an interpretation. The complexity of the mapped -418 *atpH* mRNA 3'termini suggests that -418 *atpH* initiated transcripts are rapidly degraded from their 3'end. The -45 *atpH* mRNA 3'ends are less dispersed indicating that this RNA is more stable than the -418 initiated RNA. Thus, specific stabilisation of -45 processed, monocistronic, *atpH* mRNA at the 5' as well as

at the 3'ends should be mainly responsible for the high RNA level of the *atpH* mRNA that is observed by microarray analysis.

The *atpF* and *atpA* genes do not have independent promoters. Both genes are cotranscribed with *atpI* and *atpH* (Figures 3 and 5). In the case of *atpF*, long transcripts are observed that should correspond to *atpH/AtpF* cotranscripts starting either at the - 418 SIG3 dependent promoter or at the -45 *atpH* processing site (Figure 3b). A large part of the -30 *atpF* transcript that been mapped by primer extension should arise from cleavage of the -418 initiated *atpH/atpF* cotranscript. This can be concluded because this transcript is diminished in *sig3* plants (Figure 3b). Cleavage of double stranded RNA, *e. g.* cleavage in each strand of a hairpin stem, is made by RNase III in *E. coli* (Pertzev and Nicholson, 2006). The arabidopsis genome encodes two RNase III homologues with putative transit peptides for chloroplast localization (At3g20420 and At4g37510). Such type of enzyme could cleave on both sites of the hairpin structure and liberate an *atpH* mRNA ending with a poly(A) tail. This RNA, in turn, could be substrate to immediate 3' exonuclease degradation, the reason why we could not find 3'ends corresponding to this site by ligation mediated 3'-5' cloning and sequencing. Our suggestion that the *atpH/atpF* cotranscripts might be substrates for RNase III cleavage is further supported by the fact that the sequence of the double helical structure of the -30 *atpF* hairpin is identical to the proximal box (pb) of R1.1[WC] RNA which has been shown to be important for cleavage by RNase III enzymes of different origin (Meng et al., 2008). For future experiments, it would be interesting to analyse the stability of *atpH* transcripts in RNase III knock-out plants. On the other hand, a 3'end localized within a hairpin structures might be difficult or impossible to ligate providing another reason why we could not find the -30 *atpF* transcript by 5'-RACE. Lack of

ligation might also be the reason why we could not find 3'ends corresponding to the predicted termination structure (Figure 4).

The small ATP synthase operon, consisting of the *atpB* and *atpE* genes, is transcribed from two *atpB* PEP promoters in young arabidopsis plantlets, initiating transcription at positions -520 and -462 relative to the ATG codon (Figure 6b). Both promoters are recognized by SIG2/PEP holoenzyme (Figure 6c). Transcription from the -462 promoter is also dependent on SIG3. The -462 promoter has not yet been described before. Transcription from these two *atpB* promoters results probably in a dicistronic RNA of about 2.6 kb that is cleaved at position -84 to produce a 2 kb transcript (Figure 6a and b). Transcription of the *atpE* gene can be decoupled from transcription of *atpB* by activation of an *atpB* internal PEP promoter that is recognized by SIG2 (Figure 7a-c). The part of the SIG3 initiated transcripts in the total amount of *atpB/atpE* transcripts is very low (Figure 6b) and we could not detect changes of ATPE and ATPB proteins in *sig3* plants by Western immunoblotting (not shown). However, the existence of ATPE protein in *sig2* plants shows that *atpE* mRNA is translated from the dicistronic mRNA although translation must start upstream of the *atpB* stop codon within the *atpB* coding region (Figure 7d). A stop codon dependent mechanism of translation of the downstream overlapping transcript of a dicistronic mRNA has recently been described for *ndhC-ndhK* using an *in vitro* translation system (Yukawa and Sugiura, 2008). Our result suggests that this mechanism should also exist *in vivo*.

Altogether, our results indicate transcriptional as well as post-transcriptional regulation of the expression of chloroplast ATP synthase genes, *i. e.* in contrast to Chlamydomonas, regulation can also occur on the transcriptional level in arabidopsis. In arabidosis, all genes coding subunits of the ATP synthase could be regulated at the

transcriptional level by SIG2 and SIG3. The only exception is the *atpI* gene that is under control of SIG2 only and does not harbour a SIG3 dependent promoter.

Coordinated transcriptional regulation of all plastid ATP synthase genes is probably necessary to keep stoichiometry of the mRNAs. The SIG2 initiated transcripts are generally more abundant than SIG3 initiated transcripts. This is not surprising since SIG2 is considered as one of the principal sigma factors in chloroplasts. However, this does not exclude that under specific conditions modulation of transcript levels by SIG3 becomes very important. We have not yet found under which conditions SIG3 dependent transcription will be important. To analyse different development and growth conditions to find differences in the importance of SIG3 dependent transcription will be a challenge for the future.

EXPERIMENTAL PROCEEDURES

Plant growth conditions

Arabidopsis (*Arabidopsis thaliana*) seeds of wild-type (ecotypes Wassilewskija and Columbia) and *sig1* (GK-758B02, Co), *sig2* (Kanamaru et al., 2001; Ws), *sig3* (SALK_081321, Co) and *sig4* (Favory et al., 2005; Ws) mutants were surface-sterilized for *in vitro* culture. Seeds (0) were spread on MS agar plates, kept for 72 h at 4°C in darkness (0+), and then grown for up to 7 days at 23°C under a 16h/8h light/dark cycle at 70 μmol of photons m^{-2} s^{-1}. Where indicated, plates contained spectinomycin (0.5 mg/ml). Mutant plants were grown until they have reached the same developmental stage as wild type.

Selection of arabidopsis *sig1* plants

Arabidopsis seeds of the *sig1* line (GK-758B02) have been obtained from the GABI-Kat collection (http://www.gabi-kat.de). Homozygous plants were PCR-selected using primers 5'-gctatcactgaggagtggtgc-3' and 5'-tgcatttcctttgaatgg-3'. The T-DNA insertion into the 8th exon of the *sig1* gene was confirmed by T-DNA border sequencing with primers 5'-tgcatttccttctttgaatgg-3' and 5'-gatttcccggacatgaagcc-3'.

RNA purification

Frozen material of plants was ground in a mortar and resuspended in 3 volumes of solution A (10 mM Tris-HCl pH 8 ; 100 mM NaCl ; 1 mM EDTA ; 1% SDS). Upon two phenol-chloroform and one chloroform extractions, RNAs were precipitated first in 2M LiCl and then in ethanol. When stages 0 and 0+ were included into the analyses, RNAs from all samples were purified according to the protocol described in Suzuki *et al*. (2004).

Primer extension

Upon denaturation at 65°C, 5-10 μg of total RNAs were annealed with 200 nmols of 5' ^{32}P-labeled primer at the primer annealing temperature (ta) and then retro-transcribed in presence of 100 U of SuperScript II (Invitrogen) at 42°C according to manufacturer's protocol. RNAs were digested with 10μg of RNase A and cDNAs were phenol/chloroform purified and precipitated with ethanol. As loading control a radio-labeled PCR product was added to the reactions before phenol/chloroform treatment. cDNAs were separated in a 6 % acrylamide denaturing gel in parallel to sequencing reactions performed with the same primer. The primers used for primer extension are

the following: *atpI*: 5'CATATTGCCCTCTGACAG3' (ta:49°C), *atpH*: 5'GTCCAATAGAAGCAAGC3' (ta:50°C), *atpF1*: 5'GGTATTAAATCCGAAACTCCC3' (ta:55°C) *atpF2*: 5'TCAATACACCGAAAACTACAC3' (ta:55°C), *atpA*: 5'GGTACCGGTATTTACAATCG3' (ta:53°C), *atpB2*: 5'TTTTTCACGTATCGAAACCTCTGG3' (ta:43°C), *atpB1*: 5'ATAGGAATAGGCAAGCC3'(ta: 45°C), *atpE*: 5'CTAATTGTTTCCGCTAGACC3' (ta: 53°C).

5'-RACE

In order to distinguish between primary and secondary transcripts, the first choice RLM-RACE kit was used (Ambion). In short, Turbo DNase (Ambion)-treated RNAs were first incubated with TAP (Tobacco acid pyrophosphatase) and then ligated to the 5'-RACE adapter. 400 ng of RNAs were retro-transcribed using a gene specific primer and the Super Script II enzyme. The first PCR was performed in presence of the 5'-RACE outer primer (as forward primer) and a backward gene specific outer primer. The second nested PCR was performed in presence of the 5'-RACE inner primer and a backward gene specific inner primer. Control reactions were carried out without TAP and without RT. Gene specific primers for reverse transcription are as follows: for the -45 and −418 *atpH* transcripts: 5'GTCCAATAGAAGCAAGC3'; for *atpF* transcripts: 5'TCAATACACCGAAAACTACAC3'; for *atpE* transcripts: 5'GCCTCAATTGTCTGTCTC'; for the −467/−520 *atpB* transcripts: 5'GTAAGCACTCGATTTCGTTGGTCC3'; for the −84 *atpB* transcripts: 5'GGAAGGCTACATCCAGTACC3'; for the *atpA* transcripts:

5'AGGGCAATACCTATAGTACC3'. Gene specific outer primers for the first PCR are the same as used for reverse transcription, except for the *atpA* transcripts for which the following oligo was used: 5'GCCATTACTTCATCAAGACC3'. Gene specific inner primers for the second PCR are as follows: for the -45 *atpH* transcripts: 5'CAACAGCCAACCCAGCAGC3'; for the -418 *atpH* transcripts: 5'CGCTAAGATTAATCCAGCC3'; for the *atpF* transcripts: 5'GGTATTAAATCCGAAACTCCC3'; for the *atpA* transcripts: 5'GGTACCGGTATTTACAATCG3'; for the *atpE* transcripts: 5'TACACCAATTTGTCCAC3'; for the –84 *atpB* transcripts: 5'TTTTTCACGTATCGAAACCTCTGG3'; for the –467/-520 *atpB* transcripts: 5'AAGCTCAACTAACTGAAACCTAG3'. The 5'-RACE inner and outer primers are provided by the kit.

Northern

10 μg of total RNA were loaded in 1.3% agarose-formaldehyde gel (containing MOPS pH 8) and separated in MOPS pH 7 buffer containing 1/10th vol of formaldehyde solution. RNAs were then transferred to nylon membranes that were then pre-hybridized and hybridized in phosphate buffer, as already described (Zghidi *et al.*, 2007). After quick rinsing in pre-warmed (65°C) solution 0.2 x SSC, 0.2% SDS, membranes were washed in the same buffer for 20 minutes at 65°C. Probes were amplified from total DNA using the following primers: *atpH*- 5'GAATCCACTGGTTTCTGCTG3' and 5'GGCTTAAACAAAAGGATTCGC3'; *atpE* - 5'GTGTACTGACTCCGAATC3' and 5'GCCTCAATTGTCTGTCTC3', *atpF* intron - 5'CAAGAATAGGCTGGATTCA3' and 5'CATTTGGCTCTCATGCTC3';

atpA - 5'GTAACCATTAGAGCCGACG3' and 5'GAGCTTAATTTAGCGGCTC3'.

Probes were radio-labeled using the random priming kit (New England Biolabs).

Plastid microarray

Description of the *A. thaliana* plastid-specific DNA microarray (patent FR06.07168), labeling of plastid RNAs by retro-transcription and array analysis have been detailed in Zghidi *et al*. (2007).

Protein purification and Western blot analysis

200 mg of plant material were frozen-grounded and resuspended in 200 µl of protein loading dye. After boiling for 10 min, protein extracts were cleared by full speed-centrifugation in a micro-centrifuge. Equal amounts of protein extracts were separated by SDS-PAGE and transferred to nitrocellulose membranes. For immunodetection antibodies were diluted 1:1000 and revealed by the ECL detection kit (GE Healthcare).

Circular RT-PCR

In order to map the extremities of the *atpH* transcripts, the first choice RLM-RACE kit (Ambion) was used. Total RNAs were first incubated with TAP and then self-ligated. 400 ng of RNAs were retro-transcribed using 5'GTCCAATAGAAGCAAGC3' as gene specific primer and the Super Script II enzyme. In order to determine the ends of either the –45 or the –418 transcripts, PCR was performed in presence of a -418 specific primer (5'TGATAGTAGTTCCTATCCGC3') or a -45 specific primer (5'CAACAGCCAACCCAGCAGC3') with the *atpH* oligonucleotide

(5'TCAAGGTACAGCTGCGGG3') as the second primer. -TAP and -RT reaction were

carried out as controls.

ACKNOWLEDGEMENTS

REFERENCES

Barkan, A. and Goldschmidt-Clermont, M. (2000) Participation of nuclear genes in

chloroplast gene expression. *Biochimie*, **82**, 559-572.

Bohne, A.V., Irihimovitch, V., Weihe, A. and Stern, D. (2006) *Chlamydomonas reinhardtii* encodes a single sima[70]-like factor which likely functions in chloroplast transcription. *Curr*. *Genet*. **49**, 333-340.

Böttcher, B. and Gräber, P. (2000) The structure of the H+-ATP synthase from chloroplasts and its subcomplexes as revealed by electron microscopy. *Biochim*. *Biophys*. *Acta,* **1458**, 404-416.

Chen, L.J., Liang, Y.J., Jeng, S.T., Orozco, E.M., Gumport, R.I., Lin, C.H. and Yang, M.T. (1995) Transcription termination at the *E. coli thra* terminator by

spinach chloroplast RNA polymerase *in vitro* is influenced by downstream NA sequences. *Nucleic Acids Res*. **23**, 4690-4697.

Choquet, Y. and Vallon, O. (2000) Synthesis, assembly and degradation of thylakoid membrane proteins. *Biochimie*, **82**, 615-634.

Drapier, D. Rimbault, B. Vallon, O. and Wollmann, F.A. (2007) Intertwined translational regulations set uneven stoichiometry of chloroplast ATP synthase subunits. *EMBO J*. **26**, 3581-3591.

Favory, J.J., Kobayshi, M., Tanaka, K., Peltier, G., Kreis, M., Valay, J.G. and Lerbs-Mache, S. (2005) Specific function of a plastid sigma factor for *ndhF* gene transcription. *Nucleic Acids Res*. **33**, 5991-5999.

Gatenby, A.A., Rothstein, S.J. and Nomura, M. (1989) Translational coupling of the maize chloroplast *atpB* and *atpE* genes. *Proc. Natl. Acad. Sci. USA*, **86**, 4066-4070.

Hajdukiewicz, P.T.J., Allison, L.A. and Maliga, P. (1997) The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J.* **16**, 4041-14048.

Henning, J. and Herrmann, R.G. (1986) Chloroplast ATP synthase of spinach contains nine nonidentical subunit species, six of which are encoded by plastid chromosomes in two operons in a phylogenetically conserved arrangement. *Mol. Gen. Genet*. **203**, 117-128.

Hotchkiss, T. and Hollingsworth, M.J. (1997) RNA processing alters open reading frame stoichiometry from the large ATP synthase gene cluster of spinach chloroplasts. *Plant Mol. Biol*. **33**, 635-640.

Kanamaru, K., **Nagashima, A., Fujiwara, M., Shimada, H., Shirano, Y., Nakabayashi, K., Shibata, D., Tanaka, K. and Takahashi, H.** (2001) An Arabidopsis Sigma Factor (SIG2)-Dependent Expression of Plastid-Encoded tRNAs in Chloroplasts. *Plant Cell Physiol*. **42,** 1034-1043.

Kapoor, S., Wakasugi, T., Deno, H. and Sugiura, M. (1994) An *atpE*-specific promoter within the coding region of the *atpB* gene in tobacco chloroplast DNA. *Curr*. *Genet*. **26**, 263-268.

Liere, K. and Börner, T. (2006) Transcription of plastid genes. In *Regulation of Transcription in Plants* (Grasser, K.Q., ed). Oxford: Blackwell Publishing Ltd., pp. 184-224.

Meng, W., Nicholson, R.H., Nathania, L., Pertzev, A.V. and Nicholson, A.W. (2008) New approaches to understanding double-stranded RNA processing by Ribonuclease III: Purification and assays of homodimeric and heterodimeric forms of RNase III from bacterial extremophiles and mesophiles. *Methods Enzymol*. **447**, 119-129.

Miyagi, T., Kapoor, S., Sugita, M. and Sugiura, M. (1998) Transcript analysis of the tobacco plastid operon *rps2/atpI/H/F/A* reveals the existence of a non-consensus type II (NCII) promoter upstream of the *atpI* coding sequence. *Mol. Gen. Genet*. **257**, 299-307.

Orozco, E.M., Chen, L.J. and Eilers, R.J. (1990) The divergently transcribed rbcL and atpB genes of tobacco plastid DNA are separated by nineteen base pairs. *Curr Genet*. **17**, 65-71.

Perrin, R., Meyer, E.H., Zaepfel, M., Kim, Y.J., Mache, R., Grienenberger, J.M., Gualberto, J.M. and Gagliardi, D. (2004) Two exoribonucleases act

sequentially to process mature 3'-ends of atp9 mRNAs in *Arabidopsis* mitochondria. *J. Biol. Chem.* **279**, 25440-25446.

Pertzev, A.V. and Nicholson, A.W. (2006) Characterization of RNA sequences determinants and antideterminants of processing reactivity for a minimal substrate of *Escherichia coli* ribonuclease III. *Nucleic Acids Res*. **34**, 3708-3721.

Pfalz, J., Bayraktar, O.A., Prikryl, J. and Barkan, A. (2009) Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO J*. **22**, 2042-2052.

Privat, I., **Hakimi, M.-A., Buhot, L., Favory, J.-J. and Lerbs-Mache, S.** (2003) Characterization of *Arabidopsis* plastid sigma-like transcription factors SIG1, SIG2 and SIG3. *Plant Mol. Biol*. **55**, 385-399.

Schweer, J., Loschelder, H. and Link, G. (2006) A promoter switch that can rescue a plant sigma factor mutant. *FEBS Lett*. **580**, 6617-6622.

Shiina, T., Tsunoyama, Y., Nakahira, Y., Khan and M.S. (2005) Plastid RNA Polymerases, Promoters, and Transcription Regulators in Higher Plants. *Int. Rev. Cytol*. **244**, 1-68.

Shinozaki, K., Deno, H., Kato, A. and Sugiura, M. (1983) Overlap and cotranscription of the genes for the beta and epsilon subunits of tobacco chloroplast ATPase. *Gene*, **24**, 147-155.

Silhavy, D. and Maliga, P. (1998) Mapping of the promoters for the nucleus-encoded plastid RNA polymerase (NEP) in the iojap maize mutant. *Current Genetics,* **33**, 340-344.

Slomovic, S., Portnoy, V., Liveanu, V. and Schuster, G. (2006) RNA polyadenylation in prokaryotes and organelles; different tails tell different tales. *Crit. Rev. Plant Sci.* **25**, 65-77.

Stahl, D.J., Rodermel, S.R., Bogarad, L. and Subramanian, A.R. (1993) Co-transcription pattern of an introgressed operon in the maize chloroplast genome comprising four ATP synthase subunit genes and the ribosomal *rps2*. *Plant Mol. Biol.* **21**, 1069-1076.

Stollar, N.E. and Hollingsworth, M.J. (1994a) Expression of the large ATP synthase gene cluster from spinach chloroplasts. *J. Plant Physiol*, **144**, 141-149.

Stollar, N.E., Kim, J.K. and Hollingsworth, M.J. (1994b) Ribosomes pause during the expression of the large ATP synthase gene cluster in spinach chloroplasts. *Plant Physiol*. **105**, 1167-1177.

Suzuki, Y., Kawazu, T. and Koyama, H. (2004) RNA isolation from siliques, dry seeds, and other tissues of Arabidopsis thaliana. *Biotechniques*, **37**, 542-544.

Swiatecka-Hagenbruch, M., Liere, K. and Börner, T. (2007) High diversity of plastid promoters in Arabidopsis thaliana. *Mol. Genet. Genomics*, **277,** 725-734.

Tsunoyama, Y., Ishizaki, Y., Morikawa, K., Kobori, M., Nakahira, Y., Takeba, G., Toyoshima, Y. and Shiina, T. (2004) Blue light-induced transcription of plastid-encoded *psbD* gene is mediated by a nuclear-encoded transcription initiation factor, AtSig5. *Proc Natl Acad Sci USA*, **101**, 3304-3309.

Yukawa, M. and Sugiura, M. (2008) Termination codon-dependent translation of partially overlapping ndhC-ndhK transcripts in chloroplasts. *Proc. Natl. Acad. Sci. USA*, **105**, 19550-19554.

Zghidi, W., Merendino, L., Cottet, A., Mache, R. and Lerbs-Mache, S. (2007) Nucleus-encoded plastid sigma factor SIG3 transcribes specifically the *psb*N gene in plastids. *Nucleic Acids Res.* **35**, 455-464.

Zurawski, G., Bottomley, W. and Whitfeld, P.R. (1982) Structures of the genes for the b and e subunits of spinach chloroplast ATPase indicate a dicistronic mRNA and an overlapping translation stop/start signal. *Proc. Natl. Acad. Sci. USA*, **79**, 6260-6264.

FIGURE LEGENDS

Figure 1. Microarray analysis of plastid ATP synthase mRNAs and characterisation of *atpI* transcripts.

(a) Relative transcript levels of all plastid encoded ATP synthase genes are visualized by microarray analysis of RNAs extracted from seven days old plantlets (left hand side). Quantification of the surrounded spots is presented in a diagram on the right hand side. Values correspond to three independent experiments. (b) Schematic presentation of the large ATP synthase operon. ↪ Localisation of the principal *atpI* promoter (-229/-225). ← Localisation of the primer extension oligonucleotide. (c) *AtpI* transcripts have been analysed by primer extension using total RNA extracted from wild type (lanes 5-7, lane 11) or *SIG1* (lane 8), *SIG2* (lane 9), *SIG3* (lane 10) and *SIG4* (lane 12) knock-out plants 6 days after stratification. Lanes 5 and 6 correspond to primer extension analyses of RNAs extracted from 6 days old plantlets that have been grown either in the presence

(lane 6) or in the absence (lane 5) of spectinomycin. The accompanying sequence ladder (lanes 1-4) was established by using the same primer oligonucleotide as for primer extension.

Figure 2. Characterization of *atpH* transcripts.

(a) Schematic presentation of *atpH* transcripts and primer localisation. ↱ Localisation of the operon internal SIG3 dependent *atpH* promoter. ↓ Localisation of the -45 processing site. ← Localisation of the primer extension oligonucleotide. (b) *AtpH* transcripts have been analysed by primer extension using total RNAs extracted from dry seeds (lane 1), seeds after imbibition (lane 2) and plantlets 2, 4 and 7 days after germination (lanes 3-5). The 5'ends of the -418 and -45 transcripts have been characterized by 5'-Race with (lanes 6 and 8) and without (lanes 7 and 9) prior TAP treatment. (c) The -418 and -45 *atpH* transcripts have been analysed by primer extension using RNAs from WT (lanes 1 and 5) and *SIG1* (lane 2), *SIG2* (lane 3), *SIG3* (lane 4) and *SIG4* (lane 6) knock-out plants 6 days after germination. Load corresponds to a radiolabelled PCR product that was added to the primer extension reaction as loading control before phenol/chloroform treatment.

Figure 3. Characterization of *atpF* transcripts.

(a) Schematic presentation of the *atpF* transcripts and primer localisation. ↓ Localisation of the -30 processing site. ← Localisation of the primer extension oligonucleotides. (b) *AtpF* transcripts have been analysed by primer extension using primer 2 and RNAs extracted from 6 days old WT (lanes 2, 4 and 5) and *ΔSIG3* (lane 3) plants grown without (lanes 2 - 4) or in the presence of spectinomycin (lane 5). The

29

5'ends of the two long transcripts have been characterised by 5'-Race with (lane 6) and

without (lane 7) prior TAP treatment. The insert at the right hand side shows the exact

localisation of the 5' end of the short transcript by re-analysis of primer extension

products using primer 1 (lane 8) together with the corresponding sequence ladder (lanes

9-12). The sequence was established with the same primer that was used for primer

extension. (c) Secondary structure of the sequence surrounding the -30 *atpF* processing

site.

Figure 4. Secondary structure of the atpH/atpF intergenic region.

AtpH 3' ends that have been determined by retrotranscription, PCR amplification,

cloning and sequencing of circularized RNAs (circular RT-PCR) starting either at

position -418 or at position -45 relative to the *atpH* translation initiation codon, are

marked by arrows. The number of (-418) and (-45) clones that have been found are

indicated below the arrows. (*) corresponds to the most frequently found 3'end, T

corresponds to a hypothetical terminator and -30 indicates the position of the 5'end of

atpF RNA as determined by primer extension.

Figure 5. Characterization of *atpA* transcripts and processing intermediates.

(a) *AtpA* transcripts have been analysed by primer extension using RNAs from 6 days

old WT (lane 1) and *sig3* (lane 2) plantlets. The asterisk shows a 450 bases PCR product

that was added to the reactions before phenol/chloroform treatment and served as

loading control. Lane 3 corresponds to molecular weight standards. Transcripts have

been further characterized by 5'-Race with (lane 4) and without (lane 5) prior TAP

treatment. (b) RNAs of WT (lanes 1, 3 and 4) and *sig3* plants (lane 2) have been

analysed by Northern hybridisation using probes corresponding to the *atpH* gene (lanes 1 and 2), to the *atpA* gene (lane 3) and to the *atpF* intron (lane 4). The different transcripts are labelled by arrows and letters corresponding to schema (c). (c) Schematic presentation of the transcripts. Probes that have been used in Northern experiments are marked as double lines. The putative *atpH* terminator is shown as stem-loop structure. ← Localisation of the *atpA* primer extension oligonucleotide, ↓ Localisation of processing sites at positions -45 *atpH*, -30 *atpF* and *atpH* 3'end (*) as determined by circular RT-PCR (Figure 4, *).

Figure 6. Characterisation of *atpB* transcripts.

(a) Schematic presentation of the small ATP synthase operon. ↱ Transcription start sites. ↓ Processing sites. ← Primer extension oligonucleotides. = Northern hybridisation probes. → Transcripts. (b) Total RNA extracted from 6 days old WT (lanes 5, 7 and 8, left hand side and lane 1, right hand side) and *ΔSIG3* plantlets (lane 6, left hand side and lane 2, right hand side) have been analysed by primer extension using either primer 2 (lanes 5 and 6, left hand side) or primer 1 (lanes 1 and 2, right hand side) for cDNA synthesis. Lanes 7 and 8 on the left part correspond to primer extension analysis (primer 2) of RNA extracted from 6 days old plantlets grown either without (lane 7) or in the presence of spectinomycin (lane 8). The accompanying sequence ladders are made with the same primers as used for primer extension. The -84 (lanes 5 and 6, left hand side) and the -462/-520 transcripts (lanes 1 and 2, right hand side) have been further characterized by 5'-RACE without (lane 10 left hand side and lane 8 right hand side) or with prior TAP treatment (lane 9 left hand side and lane 7 right hand side). (c) RNAs from 6 days old WT plants (lanes 1 and 5) and from *sig1* (lane 2), *sig2* (lane 3), *sig3*

(lane 4) and *sig4* (lane 6) plants have been analysed by primer extension. Load corresponds to a radiolabelled PCR product that was added to the reactions before phenol/chloroform treatment and served as loading control.

Figure 7. Characteristion of the *atpE* transcripts.

(a) Total RNAs extracted from WT plantlets grown without (lane 1) or in the presence of spectinomycin (lane 2) have been analysed by primer extension using the *atpE* primer indicated in Figure 6a. The -431 transcript has been further characterized by 5'-RACE without (lane 7) or after (lane 8) TAP treatment. The sequence ladder (lanes 3-6) was established with the same primer as used for primer extension. (b) Total RNAs extracted from 6 days old WT (lanes 1 and 5), *sig1* (lane 2), *sig2* (lane 3), *sig3* (lane 4) and *sig4* (lane 6) plants have been analysed by primer extension. Load corresponds to a PCR loading control. (c) RNAs from WT (lane 1), *sig3* (lane 2) and *sig2* (lane3) plants have been analysed by Northern blotting using *atpE* as probe (see Figure 6A). (d) 1,5; 2; 2,5 and 3 μg of protein from WT (lanes 1-4) and *sig2* plantlets (lanes 5-8) have been analysed using antibodies made against ATPE. Antibodies that had been raised against the chloroplast ribosomal protein L4 (RPL4) have been used as loading control.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(a)

(b)

(c)

Figure 1
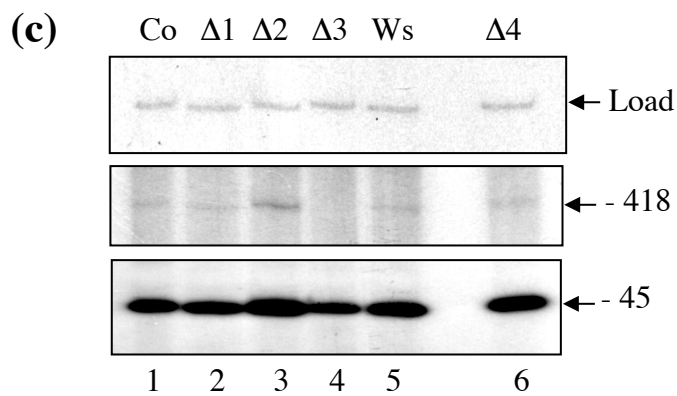
Figure 2

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(a)

- 418 (SIG3/PEP)
- 45
- 30

*atp*H   (1)   *atp*F   *atp*A
(2)

(b)

Primer extension   5'-Race

M   WT Δ3   -   +   +   -

- 418 (*atp*H)   1636
- 45 (*atp*H)   1018

506

6   7

516
396
344
298

220
201

WT T G C A

- 30

154

1   2   3   4   5   8   9 10 11 12

(c)

```
                    GA
                  A    G
                  C    U
                  A-U
                  G-C
                  G-C
                  A-U
                  A-U  ✂ -30
                  A-U
                  A-U
                  A-U
atpH 5'-G-A-A-U-A-A-A-A-A-G-A-A-U-A-A-A-A-A-U-A-U-A-3' atpF
```

Figure 3

*atp*H-*atp*F intergenic region

T *atp*H

*atp*F
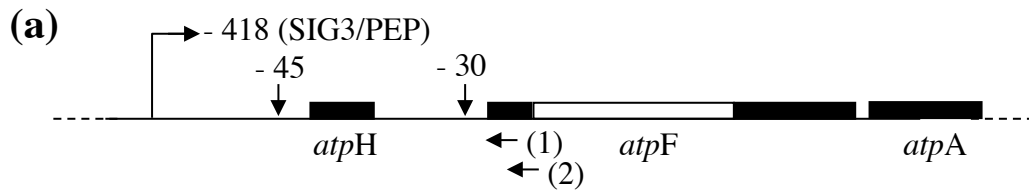
-30

(-418)

(-418)

(-418)

*

(-418)

(-418)
(-45)

(-418)
(-45)
(-45)
(-45)
(-45)
(-45)
(-45)
(-45)

Figure 4

Figure 5

Figure 6

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
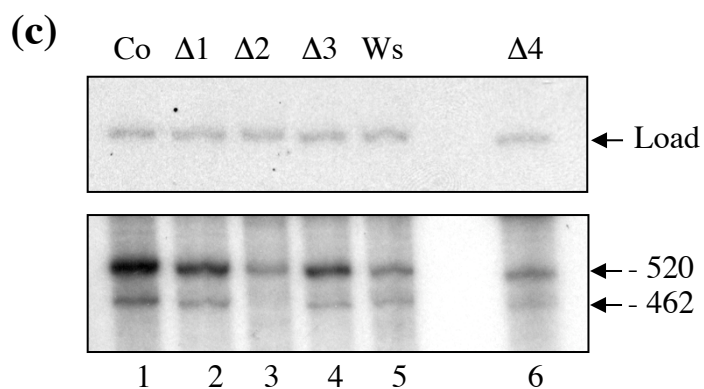41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
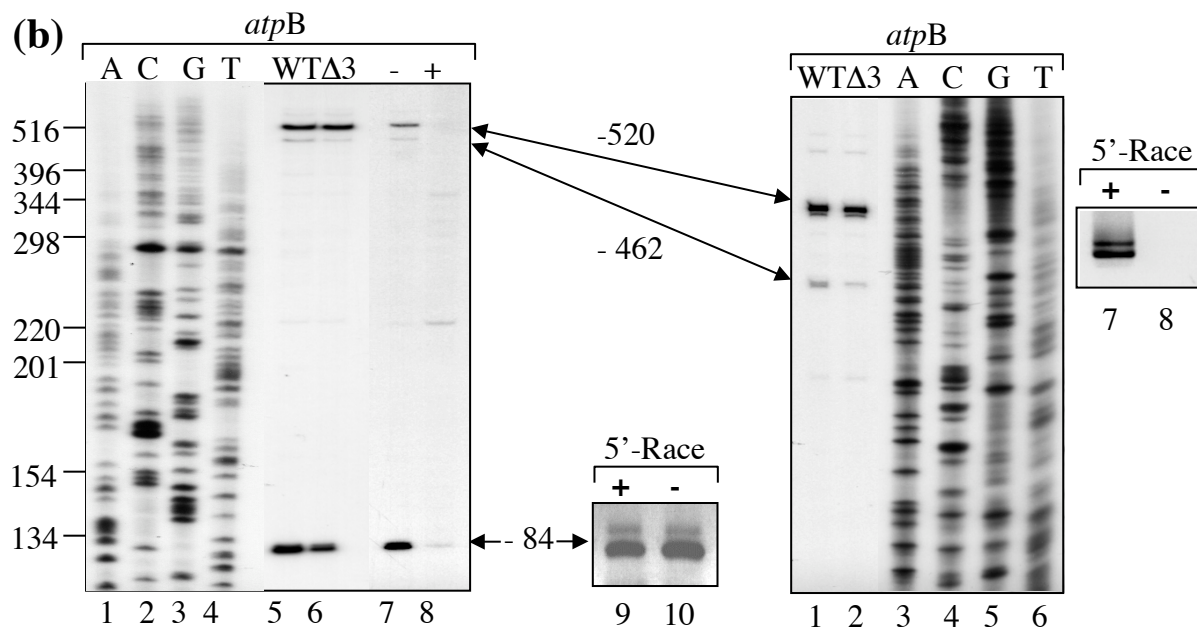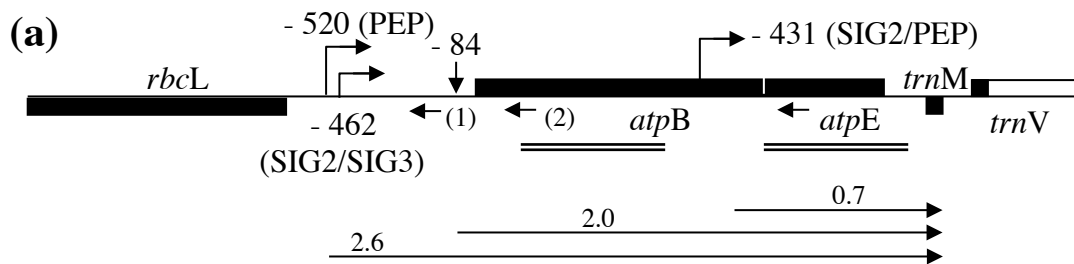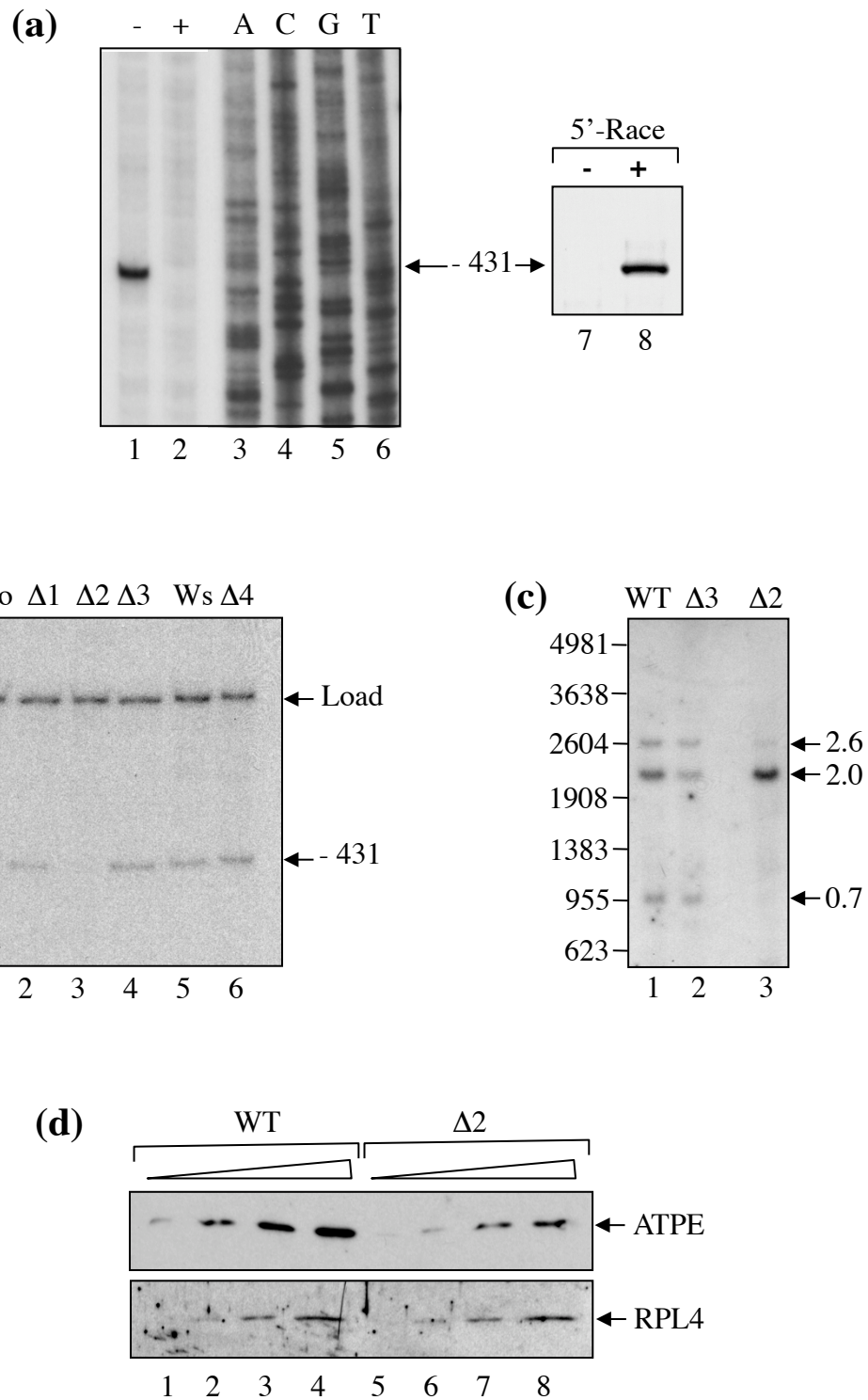60
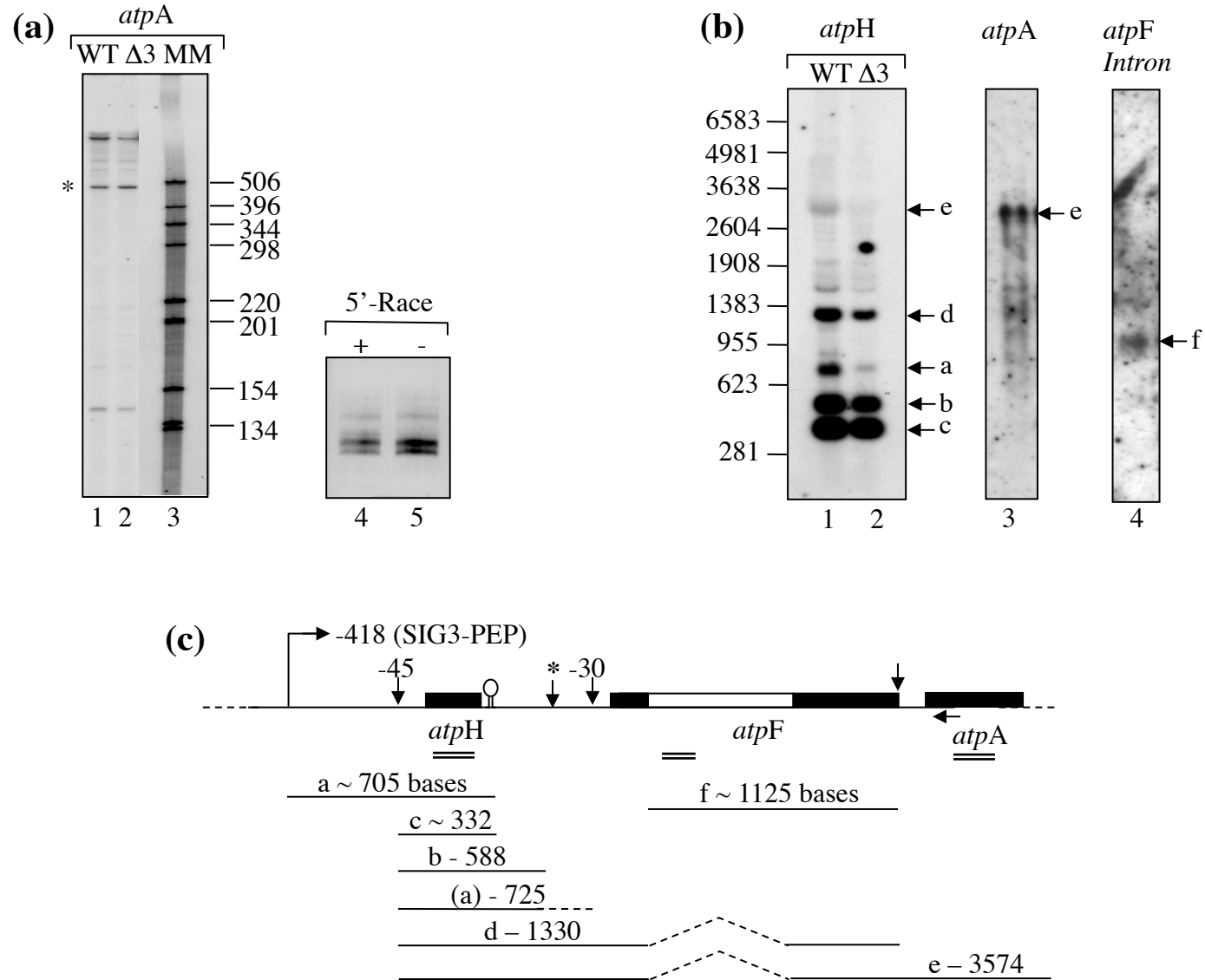
Figure 7

Abstract :

Les chloroplastes, responsables de la photosynthèse chez les organismes autotrophes, possèdent un génome plastidial codant de 100 à 130 gènes dont environ 80 pour des protéines principalement impliquées dans la photosynthèse, la transcription et la traduction. L'expression de ces gènes, coordonnée entre le plaste et le noyau, implique deux types d'ARN polymérases, la NEP (Nucleus Encoded RNA Polymerase) et la PEP (plastid Encoded RNA Polymerase) laquelle s'associe à l'un des 6 facteurs sigma (SIG), codés dans le noyau pour la reconnaissance spécifique de promoteurs de transcription.

Nous avons tout d'abord analysé le rôle de ces facteurs sigma dans la régulation transcriptionnelle des deux opérons codant des sous-unités de l'ATP synthase, *atp*I/H/F/A et *atp*B/E, en précisant le rôle particulier de SIG3 dans la reconnaissance spécifique du promoteur (-418) de l'*atp*H. Nous avons identifié les promoteurs des transcrits polycistronique et ceux situés en amont des gènes *atp*H et *atp*E, et avons montré (1) que les gènes des deux opérons sont co-régulés par SIG3 et SIG2 sauf *atp*I régulé par SIG2 seul et (2), que SIG3 jouerait un rôle essentiel dans la surexpression monocistronique d'*atp*H par la reconnaissance d'un promoteur (-418) en amont de *atp*H. L'analyse systématique des transcrits plastidiaux accumulés en fonction de l'éclairement des plantes nous a permis de corréler cette surexpression à un éclairement élevé (1300 µE) de plantes matures.

SIG3 reconnaît aussi spécifiquement le promoteur de *psb*N, gène localisé sur le brin opposé de l'opéron *psb*B/T/H/*pet*B/*pet*D, produisant un ARN anti-sens de *psb*T et de la région intergénique *psb*T/*psb*H. Nos résultats montrent que l'anti-sens de *psb*T couvre la région codante, le 5'UTR et la quasi-totalité 3' UTR du transcrit sens *psb*T, pouvant ainsi réguler la production de PSBT en interférant dans la traduction par la formation d'un duplex ARN. L'anti-sens pourrait aussi intervenir dans le processing dans la région 5' UTR de *psb*H.

Chloroplasts, responsible for photosynthesis in autotrophic organisms, have a genome containing 100-130 genes, 80 of which code for proteins mainly involved in photosynthesis, transcription and translation. Gene expression, involves two types of RNA polymerases, NEP (Nucleus Encoded RNA Polymerase) and PEP (Plastid Encoded RNA Polymerase). Six nucleus encoded sigma factors participate to PEP promoter specificity.

We first have analyzed the role of sigma factors in the transcriptional regulation of the two *atp* operons, *atpI/H/F/A* and *atpB/E,* with special emphasis on the specific contribution of SIG3 to *atpH* gene expression. We identified the promoters responsible for polycistronic transcripts and the internal promoters upstream of the *atpH* and the *atpE* genes. All genes of both *atp* operons are SIG3 and SIG2 dependent except *atpI* that is regulated by SIG2 only. The monocistronic -418 initiated *atpH* mRNA might contribute to the higher stoichiometry of *atpH*. A systematic analysis of plastid gene expression under different light conditions showed that SIG3 plays an important role in the transcript accumulation of *atpH* in high light (1300 µE) in mature plants.

Similarly, SIG3 also recognizes specifically the promoter of *psbN* located between *psbT* and *psbH* but on the opposite DNA strand and producing an anti-sense RNA to *psbT*. We showed that the anti-sense RNA covers the coding region, the 5' UTR and almost the entire 3' UTR of the *psbT* sense transcript and thus might regulate the expression of the *psbT* gene by interfering in the translation of *psbT* mRNA via duplex formation. It could also be necessary for a processing event in the 5' UTR of *psbH*.