Molecular analysis of PPR proteins in *Chlamydomonas reinhardtii*

DISSERTATION

zur Erlangung des Grades eines Doktors der Naturwissenschaften

an der Fakultät für Biologie

der Ludwig-Maximilians Universität München

vorgelegt von

ABDULLAH JALAL

München, März 2012

Tag der Einreichung: 27.03.2012

Erstgutachter: Prof. Dr. Jörg Nickelsen, AG Molekulare Pflanzenwissenschaften

Zweitgutachter: Prof. Dr. Jürgen Soll, AG Biochemie und Physiologie der Pflanzen

Tag der mündlichen Prüfung: 10.05.2012

ABSTRACT

Organellar biogenesis is mainly regulated by nucleus-encoded factors which act on various steps of gene expression including mRNA processing, splicing, stabilization, and translation initiation. Among these regulatory factors, PPR proteins form the largest family of RNA binding proteins in plants. Most of the PPR proteins are localized to mitochondria or chloroplasts, where they are major players in the RNA metabolism of defined transcripts. PPR domains are characterized by 2–30 tandem repeats of a degenerate 35 amino acid units typically mediating RNA binding activity. However, the mechanistic function of these proteins is largely unsolved. In higher plants the number of PPR proteins has increased dramatically during evolution complicating their systematic analysis. However, the genome of the unicellular green alga *Chlamydomonas reinhardtii* encodes only 11 PPR proteins as identified by *in silico* analysis. Since from an evolutionary point of view, *C. reinhardtii* can be considered as an ancestor of higher plants, the analysis of this small PPR protein family might reveal the ancient functions of its members.

In this study, a systematic *in silico* analysis of all *C. reinhardtii* PPR proteins has been performed. Furthermore, for one of these proteins, designated as PPR7, a detailed functional analysis was carried out. Localization and gel filtration analyses revealed that PPR7 is part of high molecular weight ribonucleoprotein complex in the chloroplast stroma. Secondary structure analysis and *in vitro* RNA binding assays referred recombinant PPR7 as a structured protein and confirmed its RNA binding property. Co-immunoprecipitations of PPR7-bound RNAs and subsequent RIP-chip analysis demonstrated the association of PPR7 with seven different chloroplast transcripts *in vivo*, namely *rrnS*, *psbH*, *rpoC2*, *rbcL, atpA*, *cemA-atpH*, *tscA* and *atpI-psaJ*. Furthermore, the investigation of PPR7 knock down mutants demonstrated a light sensitive phenotype as well as altered accumulations of target transcripts. According to that PPR7 seems to be involved in stabilization as well as in processing events of specific chloroplast transcripts. Taken together, a model is proposed which demonstrates the multiple functions of PPR7 in chloroplast gene expression of *C. reinhardtii*.

ZUSAMMENFASSUNG

Die Biogenese der Organellen wird vor allem durch kernkodierte Faktoren reguliert, die an verschiedenen Schritten der Genexpression, wie mRNA Prozessierung, Spleißen, Stabilisierung und Translationsinitiation, beteiligt sind. Zu diesen regulatorischen Faktoren gehören die PPR-Proteine, welche die größte Familie RNA-bindender Proteine in höheren Pflanzen darstellen. Die meisten PPR-Proteine sind hierbei in den Mitochondrien und Chloroplasten lokalisiert, wo ihnen eine bedeutende Rolle im RNA-Metabolismus spezifischer Transkripte zukommt. PPR-Domänen lassen sich durch 2-30 Tandemwiederholungen degenerierter 35 Aminosäure-langer Motive charakterisieren, die eine RNA-Bindung vermitteln. Hierbei ist jedoch die genaue mechanistische Funktionsweise weitestgehend unaufgeklärt. In höheren Pflanzen führte die Evolution zu einem dramatischen Anstieg der Anzahl der PPR-Proteine, wodurch eine systematische Analyse dieser Familie erschwert wird. Die einzelligen Grünalge *Chlamydomonas reinhardtii* weist hingegen lediglich 11 *in silico* identifizierte PPR-Proteine auf. Da sich *C. reinhardtii* aus evolutionärer Sicht als Vorfahr höherer Pflanzen betrachten lässt, ermöglicht die Analyse dieser hier kleinen Proteinfamilie die Charakterisierung möglicherweise ursprünglicher Funktionen der PPR-Proteine.

Im Rahmen dieser Arbeit wurde eine systematische *in silico*-Analyse aller PPR-Proteine aus *C. reinhardtii* durchgeführt. Desweiteren wurde eines dieser Proteine, bezeichnet als PPR7, im Hinblick auf seine Funktion detailliert untersucht. Lokalisierungs- und Gelfiltrationsanalysen zeigten hierbei, dass es sich bei PPR7 um eine Komponente eines hochmolekularen Ribonukleoproteinkomplexes im Stroma des Chloroplasten handelt. Sekundärstrukturanalysen und RNA-Bindungsstudien belegen, dass rekombinantes PPR7 ein strukturiertes Protein ist und bestätigten seine Fähigkeit zur RNA-Bindung *in vitro*. Co-Immunopräzipitationen PPR7-gebundener RNA und nachfolgende RIP-chip-Analysen demonstrierten die Assoziation von PPR7 mit sieben verschiedenen Chloroplasten-Transkripten *in vivo*, und zwar *rrnS*, *psbH*, *rpoC2*, *rbcL, atpA*, *cemA-atpH*, *tscA* und *atpI-psaJ*. Des Weiteren offenbarte die Untersuchung von PPR7 *knock down* Mutanten einen lichtsensitiven Phänotyp sowie veränderte Akkumulationen der Zieltranskripte. Demzufolge scheint PPR7 in die Stabilisierung und Prozessierung spezifischer plastidärer Transkripte involviert zu sein. Zusammenfassend wird ein Modell präsentiert, welches die multiplen Funktionen von PPR7 in der plastidären Genexpression darstellt.

TABLE OF CONTENTS

ABSTRACT **2**
ZUSAMMENFASSUNG **3**
TABLE OF CONTENTS **4**
ABBREVIATIONS **6**
1 INTRODUCTION **8**
1.1 Endosymbiosis and evolution of chloroplast **8**
1.2 Photosynthesis **9**
1.3 *Chlamydomonas reinhardtii* as a model organism **10**
1.3.1 Characteristics and expression of the chloroplast genome of *C. reinhardtii* 11
1.4 Role of nuclear encoded factors in chloroplast gene expression **13**
1.4.1 TPR proteins 15
1.4.2 OPR proteins 17
1.4.3 PPR proteins 18
1.4.3.1 Structure and classes of PPR proteins 18
1.4.3.2 Distribution and evolution of PPR proteins 21
1.4.3.3 Functions of PPR proteins in RNA metabolism 22
1.5 Aims of this study **29**

2 MATERIALS AND METHODS **31**
2.1 Materials **31**
2.1.1 Enzymes 32
2.1.2 Oligonucleotides 32
2.1.3 DNA-Vectors 32
2.1.4 Reaction systems (Kits) 33
2.1.5 Bacterial Stains 33
2.1.6 *C. reinhardtii* strains 33
2.2 Methods **34**
2.2.1 Growth of Bacterial strains 34
2.2.2 Growth of *C. reinhardtii* strains 34
2.2.3 Nucleic acids 34
2.2.3.1 Isolation of nucleic acids 34
2.2.3.1.1 Plasmid isolation from *E. coli* 34
2.2.3.1.2 Isolation of genomic DNA from *C. reinhardtii* 34
2.2.3.1.3 Isolation of total cellular RNA from *C. reinhardtii* 35
2.2.3.2 Determination of nucleic acid concentrations 35
2.2.3.3 Nucleid acid electrophoreses 35
2.2.3.3.1 Agarose gel electrophoresis of DNA 35
2.2.3.3.2 Agarose gel electrophoresis of RNA 35
2.2.3.4 cDNA synthesis and RT-PCR 36
2.2.3.5 Cloning 36
2.2.3.5.1 Transformation of *E. coli* 36
2.2.3.5.2 Polymerase chain reaction (PCR) 37
2.2.3.5.3 Sequencing 37
2.2.3.6 Probe labelling and transcript accumulation analyses (Northern blot) 37
2.2.4 Protein methods 39
2.2.4.1 Determination of protein concentrations 39
2.2.4.2 SDS polyacrylamide gel electrophoresis (SDS PAGE) 39
2.2.4.3 Immunoblotting 40
2.2.4.4 Isolation of total protein extracts from *C. reinhardtii* 41
2.2.4.5 Isolation of total soluble protein extracts from *C. reinhardtii* 41
2.2.4.6 Chloroplast isolation from *C. reinhardtii* 41
2.2.4.7 Chloroplast fractionation of *C. reinhardtii* 42
2.2.4.8 Mitochondria isolation from *C. reinhardtii* 42
2.2.4.9 Size exclusion chromatography 43
2.2.4.10 Expression and purification of recombinant proteins 43
2.2.4.10.1 Plasmids for expression of recombinant proteins 43
2.2.4.10.2 Expression and purification of recombinant proteins 44

2.2.4.11 Antibody production and purification 45
2.2.5 Nuclear transformation of *C. reinhardtii* 45
2.2.6 GFP based subcellular localization 45
2.2.6.1 GFP fusion constructs 45
2.2.6.2 GFP fluorescence microscopy 46
2.2.7 Generation of RNAi lines for PPR7 46
2.2.8 Co-immunoprecipitation studies 47
2.2.9 Microarray design and hybridization 48
2.2.10 UV cross-linking of RNA and recombinant PPR7 48
2.2.11 Chlorophyll fluorescence measurements 49
2.2.12 Circular dichroism measurements 49
2.2.13 Crystallization of His-PPR7 50
2.2.14 Bioinformatics sources 50
2.2.14.1 Prediction of gene models 50
2.2.14.2 Prediction of protein localization and transit peptides 50
2.2.14.3 Protein properties and repeat predictions 51
2.2.14.4 Alpha helical structure and wheel model predictions 51

3 RESULTS 52
3.1 PPR proteins in *C. reinhardtii* 52
3.2 Subcellular localization of PPR proteins in *C. reinhardtii* 54
3.3 Structure analysis of the PPR7 protein 57
3.3.1 PPR motifs in PPR7 and their helical wheel models 57
3.3.2 Circular dichroism measurements of recombinant PPR7 protein 59
3.3.3 Crystal structure analysis of recombinant PPR7 protein 60
3.4 Analysis of the function of PPR7 in *C. reinhardtii* 61
3.4.1 Analysis of PPR7 RNA interference Lines 61
3.4.2 Characterization of RNA binding property of PPR7 64
3.4.2.1 PPR7 is a component of a high molecular weight RNAse-sensitive complex 65
3.4.2.2 The recombinant PPR7 protein reveals intrinsic RNA binding activity 65
3.4.2.3 Identification of target RNAs of PPR7 by RIP-chip analysis 66
3.4.2.4 Semi quantitative RT-PCR of PPR7 co-immunoprecipitated RNAs 68
3.4.2.5 The role of PPR7 at the identified putative target RNAs 70
3.4.3 Photosynthetic stress response of PPR7 80

4 DISCUSSION 81
4.1 PPR proteins in *C. reinhardtii* 81
4.2 PPR7 is part of an RNase sensitive complex 82
4.3 PPR7 is associated with multiple chloroplast RNAs 84
4.3.1 Role of PPR7 as a stability factor 85
4.3.2 Role of PPR7 as a processing factor 87
4.4 PPR7 deficiency causes a light sensitive phenotype 91

5 REFERENCES 93
6 ANNEX 109
ANNEX A: 109
ANNEX B: 112

CURRICULUM VITAE 120
PUBLICATIONS AND CONFERENCE ABSTRACTS 121
ACKNOWLEDGMENT 122
EHRENWÖRTLICHE VERSICHERUNG/ERKLÄRUNG 123

ABBREVIATIONS

APS	Ammonium persulfate
A. thaliana	*Arabidopsis thaliana*
ATP	Adenosine triphosphate
BLAST	*Basic alignment search tool*
BSA	Bovine serum albumine
^{o}C	Degree Celsius
C. reinhardtii	*Chlamydomonas reinhardtii*
cDNA	Complementary deoxyribonucleic acid
Chl	Chlorophyll
Ci	Curie
CO_2	Carbon dioxide
CRP	Chloroplast RNA processing
cTP	Chloroplast transit peptide
Da	Dalton
ddH_2O	Double destilled water
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylene diamin tetraacetic acid
ER	Endoplasmic reticulum
g	Force of gravity
GFP	Green fluorescent protein
H_2O_2	Hydrogen peroxide
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMW	High molecular weight
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	Kilobase(s)
knt	Kilonucleotide(s)
L	Litre
LEF	Linear electron flow
LHC	Light harvesting complex
M	Mole(s) per litre
min	Minute
mol	Mole
MCS	Multiple cloning site
mRNA	Messenger RNA
$MgCl_2$	Magnesium chloride
NADPH	Nicotinamide adenine dinucleotide phosphate
NDH	NAD(P)H dehydrogenase complex
NEP	Nuclear encoded (plastidial) RNA-Polymerase
nt	Nucleotide(s)
(d)NTP	(Deoxy) nuclesidetriphosphate
OD	Optical Density
OPR	Octatricopeptide repeat
ORF	*Open reading frame*
PAGE	Polyacrylamide gel electrophoresis

PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PC	Plastocyanin
PEP	Plastid encoded (plastidial) RNA-Polymerase
pH	Negative decimal logarithm of proton activity
p*I*	Iso-electric point
PPR	Pentatricopeptide repeat
PQ	Plastoquinone
PSI	Photosystem I
PSII	Photosystem II
PVDF	Polyvinylidene difluoride
RIP	RNA Immunoprecipitation
RNA	Ribonucleic acid
RNase	Ribonuclease
RNAP	RNA polymerases
rpm	Revolutions per minute
RT-PCR	Reverse transcription polymerase chain reaction
rRNA	Ribosomal RNA
RuBisCo	Ribulose-1,5-bisphosphate carboxylase oxygenase
SDS	Sodium dodecyl sulphate
TCA	Trichloroacetic acid
TPR	Tetratricopeptide repeat
Tris	Tris(hydroxymethyl)-aminomethane
tRNA	Transfer RNA
U	Units
UTR	Untranslated region
UV	Ultra violet
v/v	Volume per volume
w/v	Weight per volume
WT	Wild type
µ	Micro

1 INTRODUCTION

Plants are the photoautotrophic organisms that are the main producers of food and fuel for almost all the living organisms on earth, directly or indirectly. The quality that makes plants as primary producers is their ability of synthesizing carbohydrates by a process known as photosynthesis. The sites of this vital process to take place are the organelles found in the plant cells known as chloroplasts.

1.1 Endosymbiosis and evolution of chloroplast

The first theory for endosymbiosis was presented in 1905 and currently well accepted that modern eukaryotic cells evolved after serial primary endosymbioses (Mereschkowsky, 1905). According to this theory, mitochondria were first derived from α-proteobacteria about 2.2 to 1.5 billion years ago, and then chloroplasts were derived from cyanobacteria when a predecessor of nowadays cyanobacteria integrated as a precursor of chloroplast to the host cell, about 1.5 to 1.2 billion years ago (Figure 1.1, Kutschera and Niklas, 2005). These series of endosymbiosis events lead to the evolution of modern plant cells, where three genomes interact with each other, (a) the nuclear genome from the early host cell (b) the genome of mitochondria and (c) the genome of chloroplasts. The circular DNA molecule of the plastid genomes represent their origin from eubacteria, and suggest that all the plastids have evolved from a single primary endosymbiosis event (McFadden and van Dooren, 2004).

With the passage of evolutionary phase, the endosymbionts lost their properties of being an individual organism and were converted to plant organelles. Most of the genes from the newly formed organelles were shifted to the nuclear genome but some of the major housekeeping and photosynthesis-related genes were retained by the plastid genome, e.g. for transcription (RNA polymerase subunits), translation (ribosomal proteins, rRNAs and tRNAs), photosynthesis (subunits for photosystem II, photosystem I, cytochrome b_6/f and ATP synthase, NDH complexes and Rubisco) and other functions (Sato et al., 1999; Richly and Leister, 2004; Timmis et al., 2004). This led to an inter-compartmental signalling and interdependence of genetic systems of chloroplasts, mitochondria and the nucleus to allow for a coordinated interplay of the three compartments (Herrmann et al., 2003). The genes that were transferred from plastids to nucleus during the evolution, have a similar mechanism of reimport of their products based on N-terminal transit peptide sequences, which provides another strong evidence for a common origin of plastids (McFadden and van Dooren, 2004). Organisms containing multi-layered or complex plastid membranes are proposed to be the result of a secondary endosymbiosis event, where a eukaryotic alga was engulfed by another

eukaryote. The engulfed eukaryote then underwent reductions and the remnants are the chloroplast and the extra membranes (McFadden, 2001).

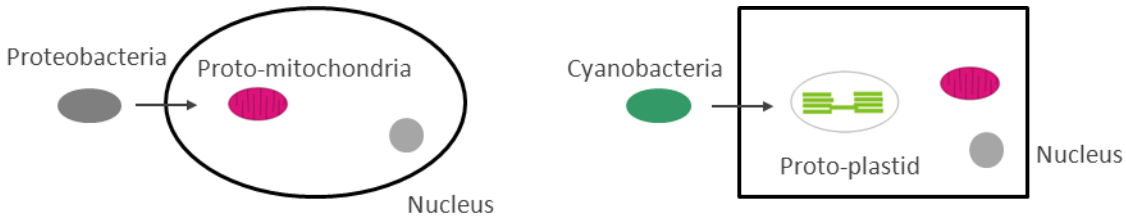


Figure 1.1: Origin of mitochondria (left) and chloroplast (right) by endosymbiosis, based on Kutschera and Niklas (2005). The two major endosymbiotic events giving rise to mitochondria and plastids which involved the transition of α-proteobacteria into proto-mitochondria and the transition from cyanobacteria into proto-plastids.

1.2 Photosynthesis

The process of photosynthesis includes the uptake of water from soil and carbon dioxide from atmosphere and converting them to organic compounds (primarily carbohydrates) with the help of light energy, hereby releasing oxygen to the atmosphere as a by-product. The organisms capable of performing photosynthesis are designated as phototrophs and include eukaryotic plants and green algae as well as cyanobacteria (Prokaryotes). The cytosol of cyanobacteria and chloroplasts of algae and plants contain structures called thylakoid membranes. The four major protein complexes of this membrane are involved in light reactions of photosynthesis. Their order of functional occurrence is: photosystem II (PSII), the cytochrome b_6/f-complex (Cyt-b_6/f), photosystem I (PSI) and the ATP synthase. In addition, light harvesting assemblies (LHCI and LHCII) are associated to the two photosystems and soluble electron carrier proteins and cofactors perform important electron shunting processes and the final reduction of NADP+ to NADPH (Figure 1.2, for a recent review see Allen et al., 2011).

A series of biochemical reactions are required to accomplish the task of photosynthesis cooperatively by pigments and protein complexes. The light energy excites chlorophyll pigments (Chl) to a higher-energy state as Chl*. The excited Chl* can either quench to the ground state by emitting fluorescence, or transfer energy to the reaction centre to drive photochemical reactions. The transferred energy is employed to split H_2O into oxygen, protons and electrons by the oxygen evolving complex (OEC) attached to PSII. Protons accumulated in the lumen generate a proton gradient across the thylakoid membrane, which can be used by the ATP synthase to produce ATP. Electrons which are transferred from PSII to PSI via the Cyt-b_6/f complex finally reduce $NADP^+$ to NADPH. All these steps are titled as linear electron flow (LEF) occurring at thylakoid membranes (Figure 1.2). Both ATP and NADPH are used in

the Calvin-Benson cycle to fix CO_2. The Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) complex plays a key role in the first step of carbon fixation (Eberhard et al., 2008).

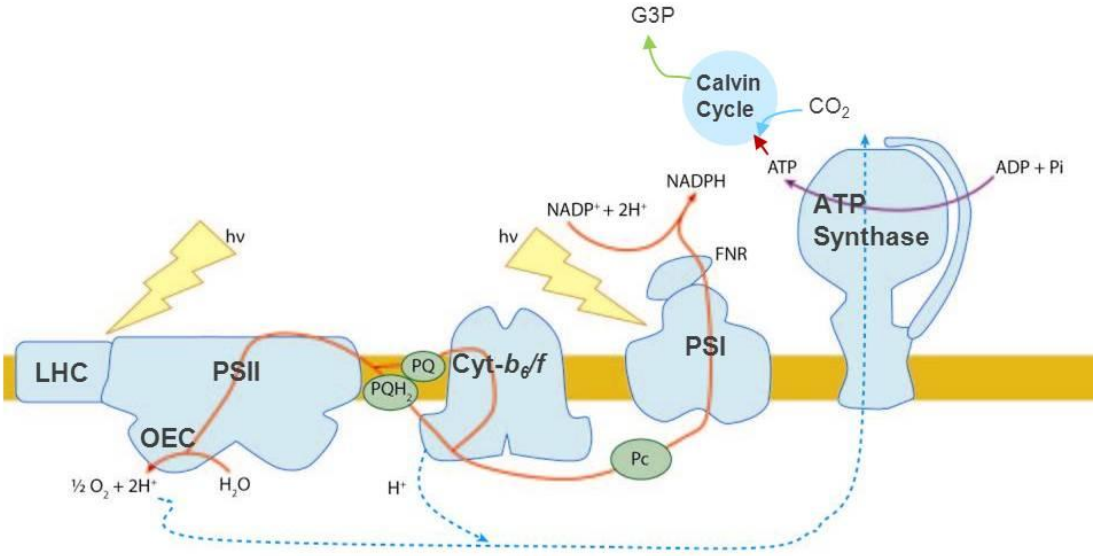


Figure 1.2: The major thylakoid membrane complexes and linear electron flow (LEF). Electrons derived from H_2O are transferred from PSII to PSI by oxidizing plastohydroquinone (PQH_2) and by reducing plastocyanin (PC) via Cyt-b_6/*f*. NADPH and ATP generated by LEF are used for carbon fixation in the Calvin cycle which results in the production of glyceraldehyde 3-phosphate (G3P, adapted from Eberhard et al., 2008).

1.3 *Chlamydomonas reinhardtii* as a model organism

For studying photosynthesis and its related machinery, the most commonly used model organisms are cyanobacteria (*Synechocystis spp.*), algae (*Chlamydomonas reinhardtii*), mosses (*Physcomitrella patens*), and higher plants (*Arabidopsis thaliana*, *Zea mays* and *Nicotiana tabacum*). *C. reinhardtii* is a unicellular biflagellate green alga measuring about 10 µm in size. It has a number of qualities that make it a very useful and popular model organism to study various aspects of cellular and molecular biology (Harris, 2001). It has a single cup shaped chloroplast which makes almost 40% of the cell volume. It can be easily grown on agar plates as well as in liquid cultures. The heterotrophic growth of *C. reinhardtii* using acetate as a carbon source provides the benefit of analysing the photosynthetic mutants (Nickelsen and Kück, 2000; Harris, 2009). *C. reinhardtii* is special for studying the molecular basis of eukaryotic cellular processes that cannot be investigated in yeast, such as photosynthesis and flagellar function (Rochaix, 1995). A big step forward for *C. reinhardtii* being used as a model organism was the sequencing of its entire nuclear, chloroplast and mitochondrial genomes (Gray and Boer, 1988; Maul et al., 2002; Merchant et al., 2007). The nuclear genome is approximately 120-megabases and haploid; due to which, any change in the geno-

type is directly evident on the phenotype (Grossman et al., 2010). All the three genomes of *C. reinhardtii* can be readily transformed, enabling specific genetic modifications of all compartments (Boynton et al., 1988; Kindle, 1990; Remacle et al., 2006). Moreover, under nitrogen starvation, haploid cells of opposite mating types fuse to form diploid zygotes resulting in four haploid progenies. Its photosynthetic apparatus is very similar to that of higher plants. The chloroplast of *C. reinhardtii* contains a photoreceptive “eye spot”, which is used for its phototactic movement. The fixation of carbon dioxide takes place inside the chloroplast at a special site known as pyrenoid (Harris, 2009).

C. reinhardtii can be grown on large scale in low cost media and short duration. For biofuel production, it has the advantage that it does not require big fields like the crop plants (Rupprecht, 2009). The RNAi (RNA interference) technique has been established as a method for target specific reduction of protein (Rohr et al., 2004). In brief, *C. reinhardtii* is an ideal model organism having many advantages over other plant models thus it is a useful tool in molecular biology.

1.3.1 Characteristics and expression of the chloroplast genome of *C. reinhardtii*

The plastid genome of higher plants and algae encodes approximately 100 to 140 genes (Sugiura, 1992). *C. reinhardtii* possesses a single chloroplast which contains about 80 copies of a circular chromosome (Rochaix, 1995; Lau et al., 2000). The plastid chromosome is completely sequenced which encompasses 203kb and is available on the web (http://www.biology.duke.edu/chlamy_genome/chloro.html). The plastid genome consists of 99 genes including 30 tRNA genes. It has two inverted repeat regions of 21.2 kb which harbour the ribosomal RNA genes and also splits the rest genome to two single-copy regions of 80 kb (Figure 1.3). The plastid genomes of higher plants show a high level of conservation in their sequence and structure while the algal plastid chromosomes show great variation (Wakasugi et al., 2001; Simpson and Stern, 2002). Unlike higher plants, the *C. reinhardtii* chloroplast genome contains various classes of short dispersed repeats (SDRs) at most of the intergenic regions. Also it contains the *tscA* gene, which encodes a small RNA required for *psaA trans*-splicing and two *C. reinhardtii*-specific large ORFs (ORF1995 and ORF2971). Furthermore, *C. reinhardtii* chloroplast genome has comparatively a small coding capacity (only 99 genes) as well as the organization of the genes encoding for the subunits of RNA polymerase show variation as compared to higher plants (Maul et al., 2002).

The transcription of plastid genes in higher plants requires two different RNA polymerases, the bacterial-type, plastid encoded plastid RNA polymerase (PEP), and the nuclear encoded plastid polymerases (NEP). NEP represents phage-type enzymes with a single polypeptide

chain encoded in the nucleus. They are closely related to DNA-dependent RNA polymerases from the bacteriophages T3/T7 (Liere and Börner, 2007).

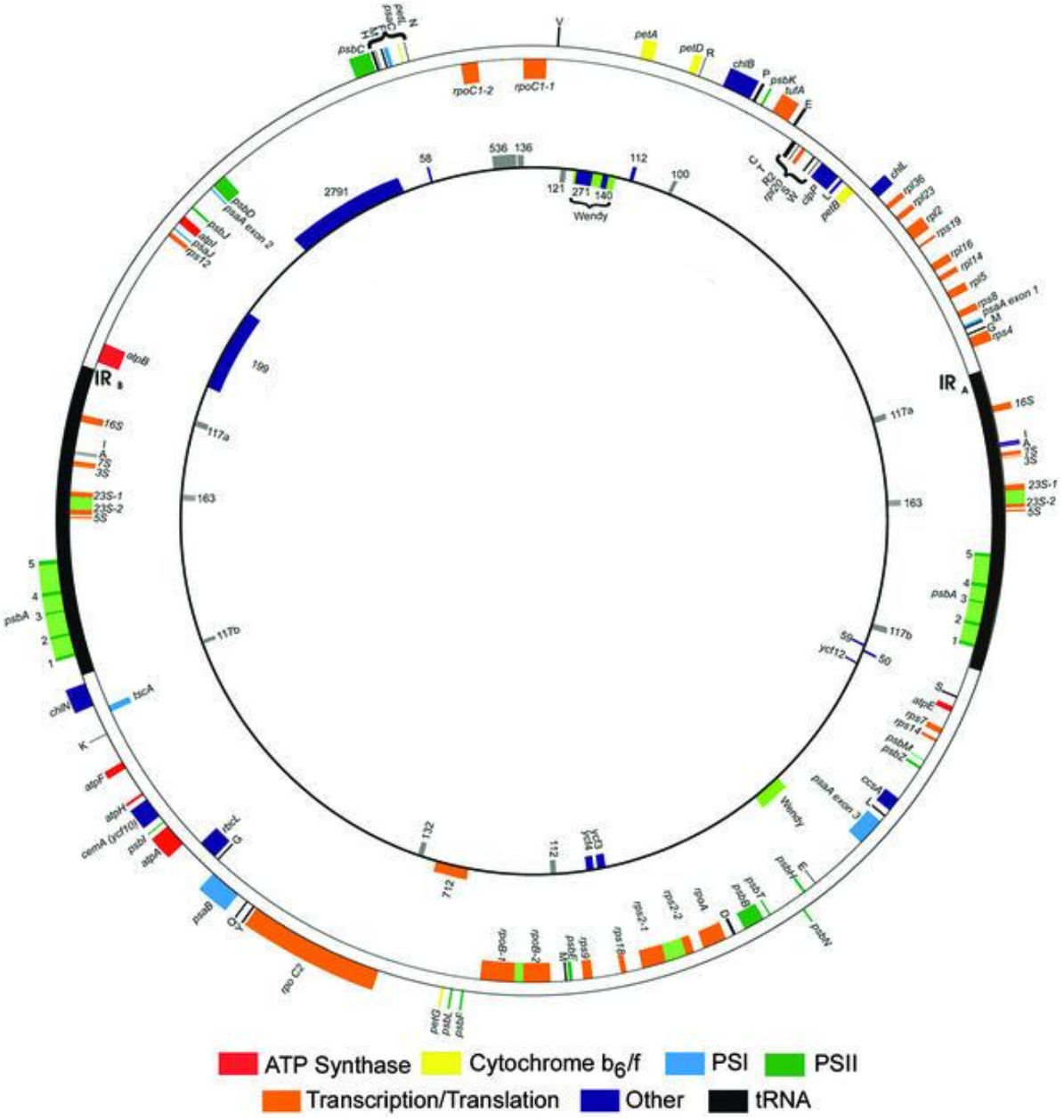


Figure 1.3: Organization of the plastid genome of *C. reinhardtii*. The 203 kb genome has two identical sequences, opposite to each other (inverted repeats, IR_A and IR_B), each of 21.2 kb in length. The outer circle shows genes of known or presumed function. Genes are colour coded by function, as shown at bottom. The genes outside of the ring are transcribed clockwise, while the genes within the ring counter-clockwise. The inner circle shows genes and ORFs of unknown function (adapted from Maul et al., 2002).

Interestingly, *C. reinhardtii* and other algae, e.g *Osteococcus tauri* and *Thalassiosira pseudonana,* seem to lack the NEP enzyme and the plastid genes are transcribed only through PEP (Armbrust et al., 2004; Derelle et al., 2006). The addition of rifampicin, an inhibitor of the

bacterial-type but not the phage-type RNA-polymerase led to complete inhibition of plastid gene transcription (Eberhard et al., 2002). Furthermore, attempts to disrupt genes encoding PEP subunits have been unsuccessful, indicating that PEP is indispensible for *C. reinhardtii* (Goldschmidt-Clermont, 1991; Fischer et al., 1996). These results suggest that all the chloroplast genes in *C. reinhardtii* are transcribed by PEP (Lilly et al., 2002; Smith and Purton, 2002).

The PEP resembles the RNA polymerase of *E. coli*, which is encoded by *rpoA*, *rpoB*, and *rpoC* genes. In chloroplast genomes and cyanobacteria, *rpoC* is divided into two separate genes, *rpoC1* and *rpoC2* (Maul et al., 2002). The *rpoB* and *rpoC1* genes are encoded by a single ORF in other chloroplast genomes while in *C. reinhardtii*, they are reported as the two closely linked ORFs *rpoB*/*rpoB2* and *rpoC1a*/*rpoC1b,* respectively, encoding the β or β'-subunit of PEP (Fong and Surzycki, 1992, Boudreau et al, 1997). Furthermore, a division of *rpoC2* gene in *C. reinhardtii* into two independent genes, i.e. *rpoC2a* and *rpoC2b* has also been discussed (Maul et al., 2002).

The expression of the chloroplast genome is regulated predominantly at post-transcriptional and translational levels. Almost all of the polycistronic transcripts are processed by endo- and exonucleases and editing events in the chloroplast (Bollenbach et al., 2007; Barkan, 2011). In contrast to higher plants no editing is observed in the *C. reinhardtii* chloroplast and only few chloroplast genes have been shown to be organized as operons that are transcribed into polycistronic primary transcripts (Rochaix, 1996; Stern et al., 2010). These polycistronic transcripts undergo post-transcriptional processes to generate the mature transcripts suitable for being translated (Sugiura, 1992). Ribonucleases act at different RNA target sites *in vivo*, but they are thought to lack cleavage specificity due to their non-specific activity *in vitro* (Yang et al., 1996; Schein et al., 2008). The specificity of ribonucleases is thought to be executed by nuclear encoded RNA-binding proteins, including members of the pentatricopeptide repeat family, which protect RNAs from non-specific nucleolytic attack by masking the sensitive sites (Stoppel and Meurer, 2011). The role of nuclear encoded regulatory factors is discussed in detail in the following section.

1.4 Role of nuclear encoded factors in chloroplast gene expression

Due to the symbiotic relationship of chloroplast and the host cell followed by the subsequent transfer of genes from chloroplast to nucleus, a coordinated expression of both nuclear and chloroplast genes is required. The protein subunits of the photosynthetic complexes are conserved between chloroplasts and its predecessor cyanobacteria. However, mRNA processing of chloroplasts has gained more complexity during integration of chloroplasts into new regulatory networks which arose between compartments of the eukaryotic cell. The chloroplast

transcripts undergo many post-transcriptional modifications including splicing, cleavage of polycistronic transcripts, stabilization with the help of stability determinants or RNA structures and editing (Barkan and Goldschmidt-Clermont, 2000). The processes of exo- and endoribonucleolytic cleavage and modulation of mRNA stability have been evolved from cyanobacteria, and in plastids these processes acquired a number of new features. Thus, gene expression of chloroplasts differs significantly from that of both prokaryotes and eukaryotes. Protein complexes of the photosynthetic membrane system as well as Rubisco consist of subunits encoded by the plastid and nuclear genomes. Therefore, biogenesis, assembly and association of these chimeric protein complexes with cofactors require the coordinated expression of chloroplast and nuclear genes. The nuclear genome encodes the majority of chloroplast localized proteins. Therefore, this genetic compartment plays a principal role in the regulation of chloroplast gene expression (Figure 1.4, Herrmann and Westhoff, 2001; Rochaix, 2004).

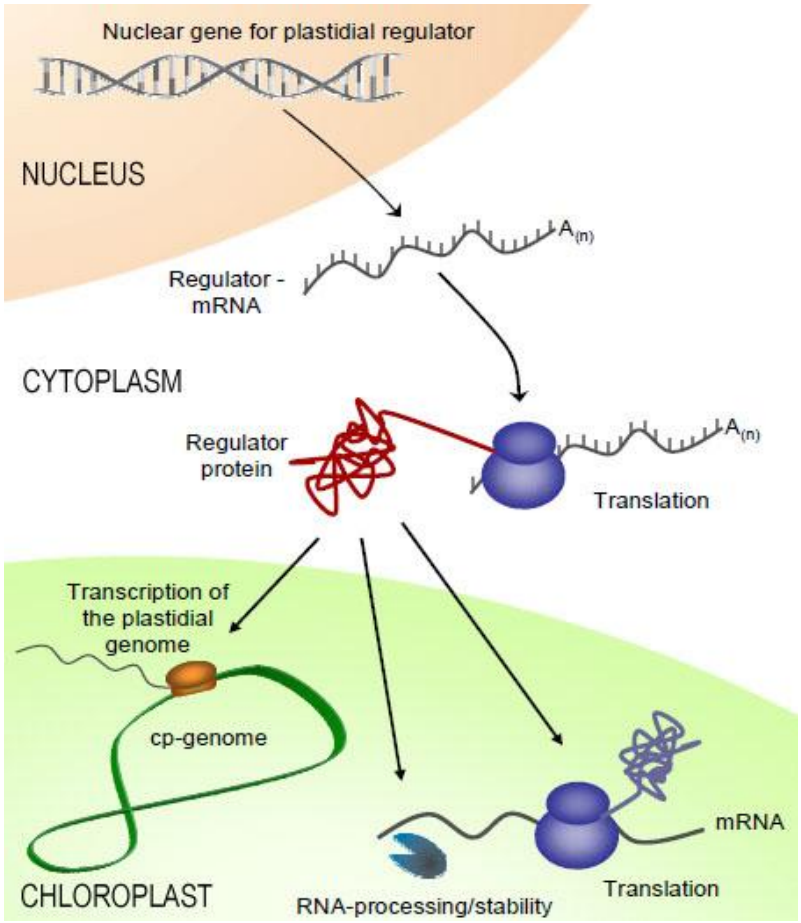


Figure 1.4: Role of nuclear encoded factors in regulation of chloroplast gene expression. Nuclear encoded regulatory factors after being expressed in the cytosol are targeted to the organelles where they are required to perform various functions at both transcriptional and post-transcriptional levels of plastidial gene expression, e.g. processing of primary transcripts, RNA stability, and RNA editing (adapted from Bohne et al., 2009).

Regulation of chloroplast gene expression mainly includes the aforementioned processes including 5'- and 3'-end processing, intercistronic cleavage, 5'- and 3'-end maturation and RNA editing (Yang et al., 1996; del Campo, 2009; Stern et al., 2010). These processes require several nucleus-encoded factors, such as endoribonucleases, exonucleases and numerous RNA-binding proteins.

Among the nucleus-encoded regulatory proteins there exist some repeat protein families that play significant roles in the mRNA metabolism of the organelles. They are characterized by the occurrence of tandem repeat motifs. The so far known families of such repeat proteins include tetratricopeptide repeat (TPR), pentatricopeptide repeat (PPR) and octatricopeptide repeat (OPR) proteins. Descriptions about these above mentioned repeat families and their characteristics are as follows.

1.4.1 TPR proteins

TPR proteins are characterized by the occurrence of 34 degenerate amino acid repeats that consist of 2-16 repeats (D'Andrea and Regan, 2003). TPR is a structural motif which is present in a wide range of proteins. TPR-containing proteins are ubiquitous and are found in bacteria, fungi, plants, insects, animals and humans (Blatch & Lässle, 1999). One TPR motif forms two anti-parallel α-helices and the tandem array of motifs generates a helical structure with an amphipathic character (Figure 1.5, Sikorski et al., 1990; Blatch and Lässle, 1999). TPR proteins are thought to form scaffolds to mediate protein–protein interactions and are part of multiprotein complexes (Das et al., 1998). They function as a chaperones, in the cell-cycle, in transcription, as well as in splicing and protein transport in the different organelles (Goebl and Yanagida, 1991).

The first solved structure of a TPR protein was presented by Das et al. in 1998 for protein phosphatase 5 which contains three TPR domains (Figure 1.5). Sequence alignments of the TPR motifs reveal the consensus sequence defined by a pattern of small and large hydrophobic amino acids (D'Andrea and Regan, 2003). The homology concerning size and hydrophobicity is high in a motif and eight amino acids are highly conserved.

The first TPR protein denoted as *nuc2*$^{+}$ was identified in yeast which functions in the cell division cycle (Hirano et al., 1990; Sikorski et al., 1990). Furthermore, TPR proteins are known to be involved in stabilization and translation of plastid transcripts. The nuclear encoded Nac2 protein is part of a high molecular weight complex that stabilizes the *psbD* transcript in *C. reinhardtii* and its absence results in degradation of the specific transcript (Boudreau et al., 2000; Schwarz et al., 2007). Two TPR proteins, encoded by the orthologous genes *hcf107* and *mbb1* in *A. thaliana* and *C. reinhardtii*, are involved in *psbH* and *psbB* transcript stability, respectively (Felder et al., 2001; Vaistij et al., 2000b). In *A. thaliana* a TPR protein, LPA1, is

proposed to be involved in PSII assembly of *de novo* synthesized subunits, while the homologue of this TPR protein in *C. reinhardtii* REP27 is found to be important during the repair cycle of PSII (Peng et al., 2006; Park et al., 2007).

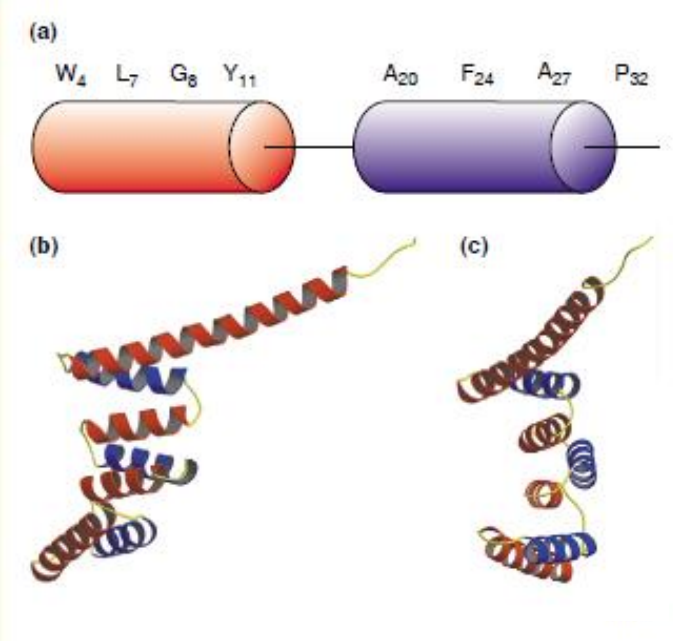


Figure 1.5: Structure of a tetratricopeptide repeat (TPR) motif (adapted from D'Andrea and Regan, 2003)**. a:** Schematic representation of the secondary structure arrangement of 34 amino acids in a TPR motif. Helix A, helix B and the loop region are shown in red, blue and black, respectively. The original consensus sequence is shown above the helices. **b:** Front and **c:** perpendicular views of the three TPRs of protein phosphatase 5.

The factor FLU in *A. thaliana*, which contains two TPR motifs, regulates the production of pigments by interacting with enzymes of the tetrapyrrol synthesis pathway (Meskauskiene et al., 2001). PratA is another TPR protein which is having a function in the maturation process of the photosystem II reaction center protein D1 in *Synechocystis* (Klinkert et al., 2004; Schottkowski et al., 2009b). In addition, Pitt is a TPR protein that is involved in the early steps of photosynthetic pigment/protein complex formation. Pitt forms a complex with the light-dependent protochlorophyllide oxidoreductase (POR) and its absence results in a three-fold decrease of POR (Schottkowski et al., 2009a). As indicated by the above mentioned examples TPR proteins are performing various functions and have different targets, which define a more diverse role of this repeat family.

1.4.2 OPR proteins

OPR proteins are the newly identified members of α-solenoid super family and are characterized by degenerate 38–40 amino acid repeats (Eberhard et al., 2011). Unlike TPR proteins, no experimental structural data is available but the secondary structure predictions suggest that the OPR motifs consist of arrayed α helices, forming super helical structures (Eberhard et al., 2011). The 38–40 amino acid repeats were first identified during 2002 and the characterization of amino acid sequence revealed a degenerate consensus sequence of five amino acid residues, PPPEW, in each repeat (Auchincloss et al., 2002). The first P and W are found to be the most conserved residues in the so far described OPRs (Figure 1.6).

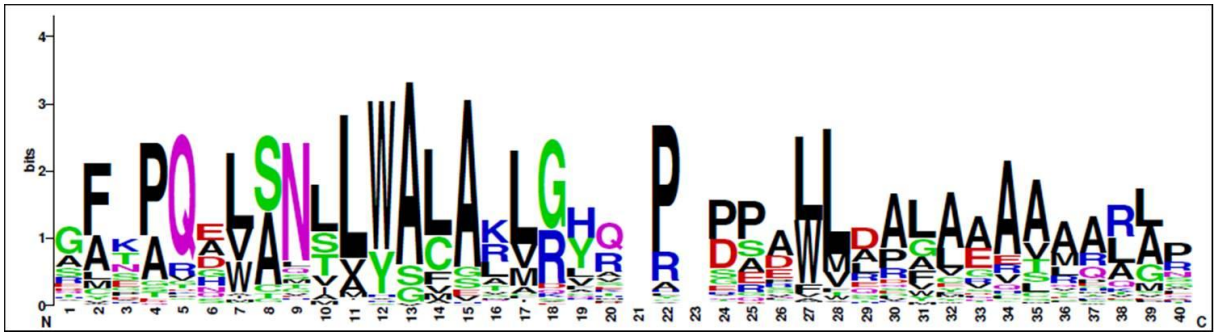


Figure 1.6 Consensus sequence of an OPR motif. (adapted from O. Vallon, A. Bohne, L. Cerutti, J.D. Rochaix, unpublished data.).

Bioinformatical analysis of *C. reinhardtii* genome revealed the presence of more than 100 OPR proteins most of which are predicted to have an organellar targeting (A. Bohne, personal communication). In contrast to the high number of OPR proteins encoded in the *C. reinhardtii* genome *A. thaliana* reveals only one OPR protein (Bohne, personal communication). The conserved sequence homology is only restricted to the OPR regions for the identified proteins. TBC2 is a protein possessing OPR repeats mostly at its C-terminus is targeted to the chloroplast and enriched in stromal fractions (Auchincloss et al., 2002). It acts specifically at the 5' UTR of *psbC* mRNA, encoding the CP43 subunit of PSII in *C. reinhardtii* and plays a role in translation. Like other repeat protein families, TBC2 is also a part of a large (~400 kDa) protein complex (Auchincloss et al., 2002). Another OPR containing protein, named RAT2, which is also localized to chloroplast stromal subfraction is found to be involved in the 3' end processing/maturation of *tscA* (Balczun et al., 2005). *tscA* RNA is a cofactor, involved in *trans*-splicing of intron 1 of *psaA* mRNA, encoding a core polypeptide of photosystem I. In *C. reinhardtii*, there are at least 14 nucleus-encoded factors essential for *psaA trans*-splicing, out of these, two are involved in splicing of both introns 1 and 2 (Goldschmidt-Clermont et al., 1990; Merendino et al., 2006). RAA1 is an OPR protein which is part of large ribonucleoprotein complex but unlike TBC2 and RAT2, RAA1 is found in chlo-

roplast membrane fractions and is involved in *trans*-splicing of both introns 1 and 2 of the *psaA* mRNA. The C-terminus of RAA1 alone was found to be sufficient for processing of *tscA* transcript and splicing of intron 1 of *psaA* transcript, while its central part is found to be involved in the *trans*-splicing of the second intron. This data indicates the presence of two functional domains in RAA1 (Merendino et al., 2006). Another example of an OPR protein, having a dual function is TDA1. The nucleus-encoded TDA1 factor is found to be specifically required for translation of the *atpA* transcript that encodes the α subunit of the ATP synthase in *C. reinhardtii* chloroplasts. The N-terminus of TDA1 is involved in trapping a subset of untranslated *atpA* transcripts into non-polysomic complexes, while the C-terminus of TDA1 protein can alone act as translational activator of this transcript. However, for TDA1, it is known that the OPR repeats are present only at the C-terminus, hence assigning the translation activation function to this repeat region (Eberhard et al., 2011). The functions of TBC2, RAT2, RAA1 and TDA1 suggest that OPR repeats interact with specific organelle transcripts and consist of RNA binding domains.

1.4.3 PPR proteins

PPR proteins are another group of repeat proteins characterized by the occurrence of a signature motif of degenerate 35 amino acid repeats. The PPR repeats are mostly found in an array ranging from 2–30 repeats in a single protein (Small and Peeters, 2000; Schmitz-Linneweber and Small, 2008). The PPR family was first described by Small and Peeters during the year 2000. They identified this protein family during a search for gene products predicted to be targeted to mitochondria or plastids in the *A. thaliana* nuclear genome (Small and Peters, 2000). In higher plants, PPR proteins constitute the largest group of RNA-binding proteins known so far (Small and Peeters, 2000; Lurin et al., 2004). Some PPR proteins were already characterized before their description as a distinct family, e.g. Pet309 in yeast, Cya-5 in *Neurospora crassa* and CRP1 in maize (Manthey and McEwen, 1995; Coffin et al., 1997; Fisk et al., 1999).

1.4.3.1 Structure and classes of PPR proteins

Due to the sequence homology, PPR proteins were thought to have a similar structure as TPR proteins (Section 1.4.1). A PPR motif was predicted to consist of Helix A and B which fold into a helix-turn-helix structure similar to those found in TPR and in other ‘‘solenoid’’ proteins (Small and Peeters 2000). In a recent study from Ringel et al. (2011) on structural resolution of a mitochondrial RNA polymerase containing two PPR motifs, the prediction of a he-

lix-turn-helix structure of PPR motifs was confirmed (Figure 1.7). A difference between PPR and TPR motifs is that in PPR motifs, the residues of Helix A, facing inward of a formed groove are hydrophilic and predicted structural models of many of PPR-containing proteins show that the bottom of this groove is positively charged (Small and Peeters 2000). This provides the possibility of binding negatively charged RNA molecules by some or all the PPR motifs, therefore suggesting that PPR repeats are rather RNA-binding than protein-binding motifs. PPR proteins usually consist of twice as many repeats as TPR proteins, suggesting multiple or rather extended ligands. The width of the central groove is sufficient to hold a single RNA strand, while the positively charged surface at the bottom of the groove is able to bind the phosphate backbone (Small and Peeters 2000).

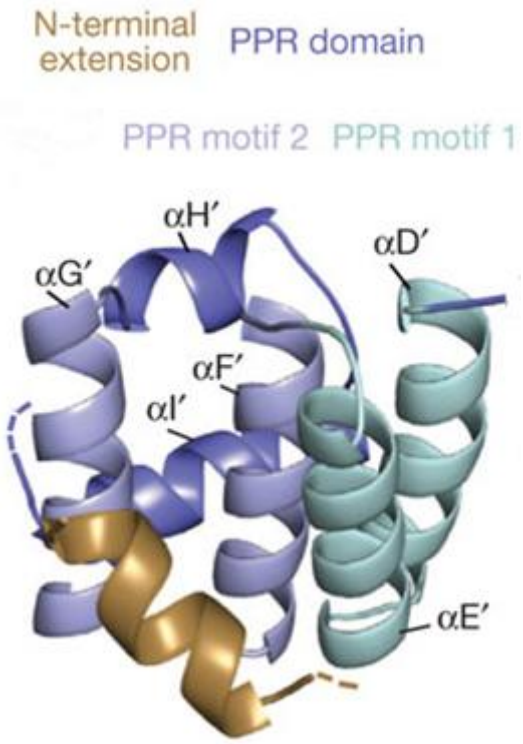


Figure 1.7: Structure of PPR domain in human mtRNA polymerase. The two PPR motifs are shown in light cyan (motif 1) and light blue colour (motif 2), consisting of helix A (αD' and αF') and helix B (αE' and αG'). Each PPR motif shows a helix-turn-helix fold (adapted from Ringel et al., 2011).

The genome-wide investigation of *A. thaliana* PPR proteins, revealed the existence of a large PPR family (Lurin et al., 2004). Out of about 450 members of PPR proteins encoded in the *A. thaliana* genome, half of them contain the canonical or direct repeats, having no gaps between them and consist of 35 amino acids. These were denoted as P subfamily (Figure 1.8). In the other half of the family, non repeat gaps of 65 to 70 amino acids between PPR motifs were observed. The analysis revealed two new motifs of 31 and 35 to 36 amino acids. These two motifs were denoted as PPR-like S (for short) and PPR-like L (for long) motifs (Figure 1.8). These sequence motifs are usually organized as PLS triplets and are not present in all

PPR proteins. The newly identified PLS motifs were found to be related to the previously identified plant combinatorial and modular protein (PCMP) family (Aubourg et al., 2000). The PPR repeats cover at least two third of the protein sequence showing homology to other proteins while the N-terminal sequences show less sequence similarity. The N-terminus usually contains a targeting sequence called as transit peptide (TP) that leads the protein to the respective organelle and is cleaved off afterwards. The C-terminal portion of PPR proteins in higher plants usually contains some extra domains unrelated to PPR domains. Some PPR proteins contain the E-domain (for extended), which is further divided into two smaller motifs E and E+. Another additional domain is defined by a C-terminal aspartic acid, tyrosine, and tryptophan-tripeptide denoted as DYW domain. This motif always follows the E and E+ motifs. The E and E+ motifs are degenerate and somewhat similar to each other while DYW domain shows a more conserved sequence. These extra C-terminal domains show no similarity to the PPR motifs and their origin is still unclear (Lurin et al., 2004). The extra C-terminal domains are only found in the PLS defined subfamily, also each domain is present once in a PPR protein. These domains appear in an order in which an E+ domain will always appear after the E domain and likewise a DYW domain will always occur after an E and E+ domain. These observations lead to the subdivision of PLS family into 4 categories: 1) PPR proteins lacking C-terminal domains, 2) PPR proteins having only E domain, 3) PPR proteins having E ad E+ domains and 4) PPR proteins containing the E, E+ and DYW domain.

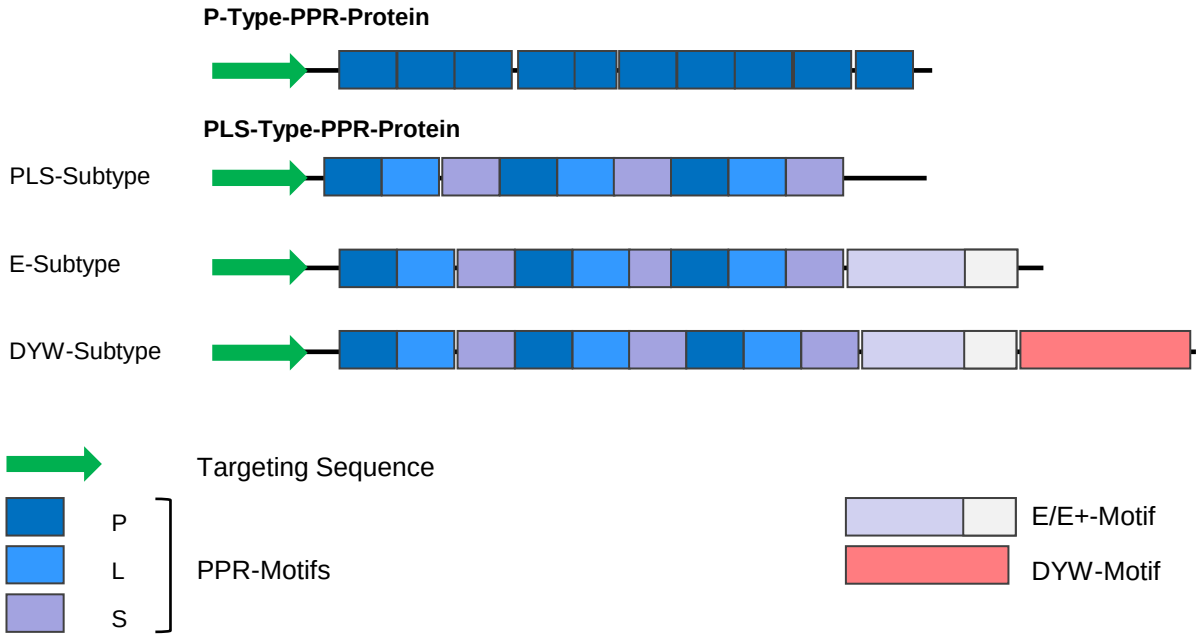


Figure 1.8: Schematic illustration of motifs in PPR proteins. The classic PPR motifs of 35 amino acid repeats are termed as P-type. In PLS type, the long (L) and short (S) motifs are present in alternate with P motifs. The green arrow represents the transit peptide found at the N-terminus of PPR proteins. The C-terminus of PLS type PPR proteins contain E/E+ and DYW domains (adapted from Lurin et al., 2004; Schmitz- Linneweber and Small, 2008).

In addition there can be some other domains in PPR proteins, such as, RRM (RNA recognition motif domains or DNA-binding small-mutS related (SMR) domains (Schmitz-Linneweber et al., 2006; Koussevitzky et al., 2007).

1.4.3.2 Distribution and evolution of PPR proteins

Genes encoding PPR proteins are widely distributed in eukaryotes but particularly in plant genomes, they constitute a big family with over 400 genes in *A. thaliana* and 477 genes in *Oryza sativa* (Figure 1.9, Small and Peeters, 2000; O'Toole et al., 2008). The majority of these are predicted to be targeted to the chloroplast or mitochondria. In contrast to higher plants, all other eukaryotes including humans, *Drosophila*, protists and algae contain a small set of PPR proteins, suggesting that the number of these factors dramatically increased during the period of land plant evolution (Lurin et al. 2004; O'Toole et al., 2008; Schmitz-Linneweber and Small 2008). In the moss *P. patens*, there are 103 PPR encoding genes. In the genome of the protozoan parasite *Trypanosoma brucei* 28 PPR motif-containing sequences were found, which is a high number as compared to other non-plant eukaryotes, e.g. human or the fruit fly *Drosophila melanogaster* which possesses only six or two identified PPR proteins, respectively (Figure 1.9, Pusnik et al., 2007). *C. reinhardtii*, as a unicellular green alga, contains a limited number (11) of PPR proteins.

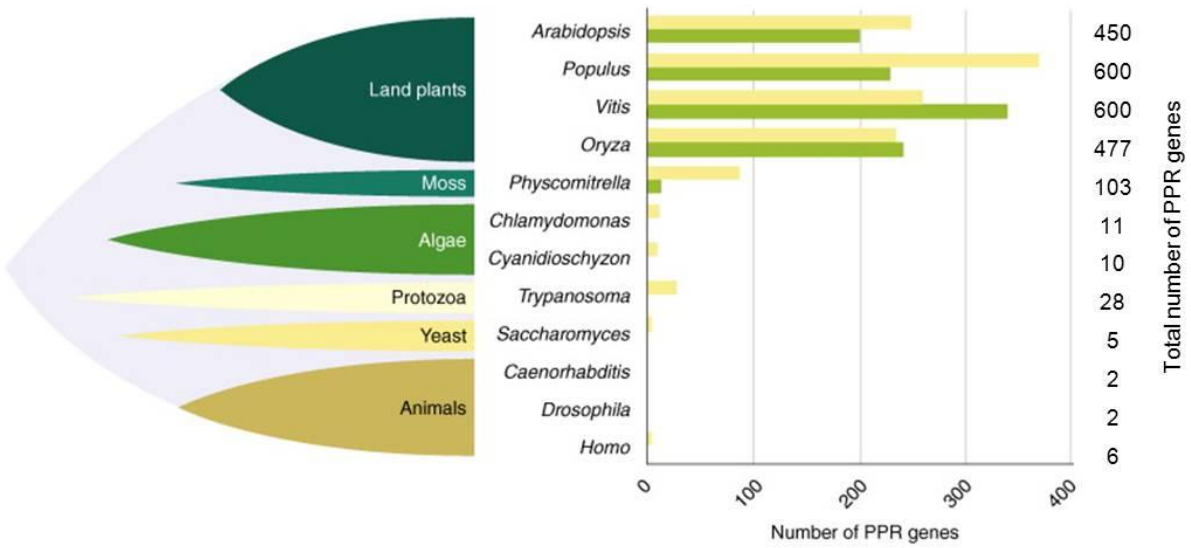


Figure 1.9: Distribution of PPR proteins. The yellow bars represent the numbers of P class and green bars represent number of PLS class PPR genes in various eukaryotes. The total number of PPR genes found in each organism is given on the right (adapted from Schmitz-Linneweber and Small, 2008).

The moss family is known to be diverged early in the evolution of land plants. The much smaller number of PPR genes encoded by the genome of *C. reinhardtii* and then *P. patens* compared with those of *A. thaliana* and rice indicates that the expansion of the PPR family occurred after the divergence of moss from the lineage leading to vascular plants.
The dicot *A. thaliana* and the monocot *O. sativa* have similar numbers of PPR genes and almost identical distribution of genes on the individual subtypes (P, PLS, E, and DYW genes). Another observation made for PPR genes in *A. thaliana* and *O. sativa* is that the majority of genes lack introns (Lurin et al., 2004; Rivals et al., 2006). Approximately 80% of *A. thaliana* and *O. sativa* PPR genes fall into this category. On the other hand, the intron containing and intron-less moss PPR genes are roughly in equal proportions. Furthermore, all the 11 PPR protein encoding genes found in *C. reinhardtii* contain introns.
Comparative genome studies for *A. thaliana*, *O. sativa* and *P. patens* revealed that intron-rich *A. thaliana* and *O. sativa* PPR genes cluster among the intron-rich PPR genes of moss. These observations together point that one of the possible mechanisms for the expansion of PPR protein family in land plants is retrotransposition, a process, in which a mature mRNA, associated with a retrotransposon, is reverse transcribed and integrated into the genome (O'Toole et al., 2008).
The wide spread of PPR protein encoding genes in eukaryotic genomes and their role in organelles points towards a mitochondrial origin but PPR genes are absent from prokaryotes (Lurin et al., 2004). Some examples in prokaryotes for genes showing homology to PPR proteins are considered as horizontal gene transfer, rather than being of prokaryotic origin (Lurin et al., 2004; Schmitz-Linneweber and Small 2008).

1.4.3.3 Functions of PPR proteins in RNA metabolism

To date, information for characterized PPR proteins reveals that they are required for a wide range of different post-transcriptional processes in plant organelles. After being translated in cytoplasm, PPR proteins are trafficked to mitochondria or chloroplasts by means of an N-terminal transit peptide, where they are found to be involved mainly in RNA metabolism (reviewed in Delannoy et al, 2007; Andrés et al, 2007; Schmitz-Linneweber and Small, 2008). Only one example of a nuclear targeted PPR protein so far, is the *A. thaliana* glutamine-rich protein 23 (GRP23) having a role in early embryogenesis by interacting with RNA polymerase II subunit III (Ding et al., 2006). Homozygous mutations in individual genes encoding PPR proteins often have strong phenotypic effects by causing the lack of expression of a specific organellar gene up to embryo lethality. Despite of large numbers of PPR proteins found in land plants, PPR genes show a remarkable conservation of function, which speaks for a kind of essential, non-redundant function of the encoded proteins (Lurin et al., 2004;

O'Toole et al., 2008; Schmitz-Linneweber and Small, 2008). The functions of PPR proteins, apparent from mutant analyses in a variety of organisms reveal their involvement in a diverse range of functions (see Table 1.1). Their role in all stages of RNA metabolism has been confirmed and interestingly, even the structurally similar PPR proteins are involved in different physiological and molecular functions (Andrés et al., 2007).

Experimental evidence that PPR motifs are responsible for RNA binding activity by forming RNA binding domains, comes from domain-swap experiments between two *A. thaliana* PPR proteins i.e. CRR21 and CRR4 which show that rather PPR motifs are responsible for RNA binding activity than the additional domains found in PPR proteins (Okuda et al., 2007). Another proof, that PPR motifs are responsible for RNA binding comes from yeast PPR protein Pet309 involved in translation of the yeast mitochondrial *cox1* mRNA. It contains 7 PPR motifs and the deletion of one PPR motif at a time abolished the RNA binding and translation of *cox1* mRNA. Each of the seven PPR motifs was found to be equally essential for the function of Pet309. (Tavares-Carreón et al., 2008). A recent study on identification of an RNA binding surface in PPR proteins, using recombinant versions containing two PPR motifs of the HCF152 protein, reveals that the 1st, 4th, 8th, 12th and 34th amino acid of a PPR motif, form the RNA interacting surface (Kobayashi et al., 2011). Furthermore, the study discovered that there are differences in RNA binding affinities among the PPR motifs of the same protein, pointing to functional differences among PPR motifs. PPR proteins involved in editing of specific site(s) can be more useful in understanding the mechanism for PPR binding to specific RNA sequences as the target site is precisely known (Schmitz-Linneweber and Small, 2008).

The RNA ligands for some of the PPR proteins have been determined, via immunoprecipitation of the respective native PPR protein and then analysing the co-precipitated RNAs by hybridizing them to a chip (RIP-chip). This technique has been successfully used to determine the target RNAs for CRP1, PPR4, PPR5 and PPR10 proteins from maize and also for *Rf592*, which is a mitochondrial targeted protein in *Petunia hybrida* (Schmitz-Linneweber et al., 2005, 2006; Gillman et al., 2007; Beick et al., 2008; Pfalz et al., 2009). Data for the identification of RNA ligands by *in vitro* approaches are available for HCF152 and CRR4 from *A. thaliana* plastid targeted proteins, rice mitochondrial Rf1, maize PPR5 and PPR10 (Nakamura et al., 2004; Okuda et al., 2006; Kazama et al., 2008; Williams-Carrier et al., 2008; Pfalz et al., 2009; Prikryl et al., 2011). The *in vitro* approaches used to identify the target RNAs for PPR proteins include UV-cross linking experiments for HCF152 and electrophoresis mobility shift assays (EMSA) for CRR4, Rf1, PPR5 and PPR10. These analyses show that PPR proteins have diverse target RNAs and the target sequences occur in 5’ UTRs, introns and in intergenic regions. PPR proteins can specifically interact with highly defined transcripts because no sequence similarities are found in the above mentioned targets.

PPR proteins have been attributed to a remarkably broad spectrum of post-transcriptional steps of organellar gene expression. The following paragraphs will give some examples for

characterized PPR proteins involved in RNA splicing, editing, and stabilization as well as translation.

Many PPR proteins have been shown to be responsible directly or indirectly for organellar RNA splicing. In *P. patens* the PPR protein PPR_38 is shown to play a role in splicing of chloroplast *clpP* mRNA encoding the ClpP protease (Hattori et al., 2007; Hattori and Sugita, 2009). HCF152 improves splicing of *petB-petD* polycistronic transcript coding for subunits of Cyt-b_6/f complex (Meierhoff et al., 2003). More specific roles of PPR proteins in splicing come from the following examples. The chloroplast localized PPR OTP51 in *A. thaliana* is required for processing of intron 2 of *ycf3* mRNA encoding the open reading frame of hypothetical protein YCF3, and probably, is involved in splicing of other introns as well (de Longevialle et al., 2008). Another example is OTP43, a mitochondrial targeted PPR which is required for splicing of the first out of four *nad1* introns. Nad1 is a subunit of mitochondrial oxidoreductase complex (de Longevialle et al., 2007). Furthermore in maize, PPR4 is essential for the *trans*-splicing of the *rps12* intron encoding chloroplast ribosomal protein S12 (Schmitz-Linneweber et al., 2006).

Since the characterization of PPR proteins as forming a large protein family in land plants, it was suggested that many of the members could play a role in editing of the organellar transcripts of plants (Small and Peeters, 2000). CRR4 was the first characterized PPR protein found to be involved in editing of *ndhD* mRNA in *A. thaliana* chloroplasts, which encodes for a subunit of NAD(P)H dehydrogenase (Kotera et al., 2005). Later on CRR21 was characterized as an editing factor of another site in the *ndhD* transcript (Okuda et al., 2007). Another editing factor which is interestingly responsible for editing of two specific and distinct sites in plastid *rpoA* transcripts (encoding subunit A of plastid encoded plastid RNA polymerase) and *clpP* transcripts is CLB19 (chloroplast biogenesis 19; Chateigner-Boutin et al., 2008) . Furthermore, MEF1 is required for RNA editing of three specific sites of different mitochondrial mRNAs, namely *rps4*, *nad7*, and *nad2* in *A. thaliana*.(Zehrmann et al., 2009). Further characterized PPR proteins involved in editing of specific organellar transcripts are described in studies from Zhou et al. (2008); Hammani et al. (2009); Okuda et al. (2009, 2010); Yu et al. (2009); Doniwa et al. (2010); Takenaka (2010); Tasaki et al. (2010) and Verbitskiy et al. (2010). It is an interesting verdict that all these editing factors belong either to the E or DYW PPR subclass and the occurrence of editing areas correlate phylogenetically strictly with the presence of DYW motifs in proteins of land plants (Salone et al., 2007; Rüdinger et al., 2008). In addition, DYW domains show similarities to cytidine deaminases. Therefore, it was suggested that the DYW domains possess editing activity (Salone et al., 2007). It is generally assumed that E domains in PPR proteins are non-catalytic and function to recruit a so far unknown editing enzyme (Shikanai, 2006; Schmitz-Linneweber and Small, 2008).

In addition, many PPR genes in different species have been identified associated with the analysis of cytoplasmic male sterility (CMS) as fertility restorer genes (*Rf* genes). CMS refers

to the failure of viable pollen production because of a genetic factor carried by mitochondria. Interestingly, all nucleus-encoded *Rf* factors except one in maize (*Rf2*), are PPR proteins which restore the male fertility by prevention of the expression of these mitochondrial gene encoding CMS-specific polypeptide (Bentolila et al., 2002; Brown et al., 2003; Desloire et al., 2003; Akagi et al., 2004). Bentolila et al. (2002) were successful to clone the first *Rf* in *Petunia*, which is responsible for the control of the expression of a CMS encoding gene. Some other examples of characterized PPR proteins associated to CMS are as follows. *Rfk1* (*Rfo*) of radish, *Rf-1* of rice, and *Rf1* and *Rf2* of sorghum (Brown et al., 2003; Desloire et al., 2003; Kazama and Toriyama, 2003; Koizuka et al., 2003; Akagi et al., 2004; Komori et al., 2004; Klein et al., 2005; Wang et al., 2006; Kato et al., 2007; Jordan et al., 2010). The precise mode of action of *Rf* factors is not fully understood but the mechanisms by which this is accomplished, are the endo-nucleolytic cleavage of aberrant transcripts, the degradation of RNA or the inhibition of their translation (Brown et al., 2003; Akagi et al., 2004; Wang et al., 2006; Kazama et al., 2008).

Several PPR proteins that participate in organellar transcript stabilization have been characterized. In *C. reinhardtii*, the only two characterized PPR proteins have been attributed to participate in post-transcriptional stabilization of specific transripts. MCA1 is a PPR protein characterized in *C. reinhardtii* which stabilizes *petA* transcript coding for cytochrome *f* apoprotein (Loiselay et al., 2008). MRL1 is another plastid localized protein in *C. reinhardtii* which binds to the 5’ UTR of *rbcL* mRNA coding for large subunit of Rubisco and is responsible for the stability of the respective transcript (Johnson et al., 2010). The MRL1 homologue in *A. thaliana* reveals a conserved function, being involved in the production/stabilization of the processed *rbcL* transcript (Johnson et al., 2010). In *Arabidopsis*, a PPR protein PGR3 stabilizes transcripts of *petL* operon (Yamazaki et al., 2004). PPR5 in maize stabilizes the *trnG*-UCC precursor by directly binding and protecting an endonuclease-sensitive site (Beick et al., 2008). In Drosophila melanogaster, a mitochondrial targeted PPR protein BSF binds to 3’ UTR of Bicoid mRNA and plays a role in its stabilization (Mancebo et al., 2001).

It is interesting that completely opposite to the stabilization function, several PPR proteins have been characterized playing a role in endonucleolytic cleavage of polycistronic transcripts. CRP1 protein in maize is required for the processing of the *petD* mRNA from a polycistronic precursor (Fisk et al., 1999; Schmitz-Linneweber et al., 2005). HCF152 is involved in the intergenic cleavage between *psbH* and *petB* within the *psbB-psbT-psbH-petB-petD* operon *in A. thaliana* (Meierhoff et al., 2003). CRR2 is another example of PPR protein that is involved in the intercistronic processing of *rps7-ndhB* transcripts in *A. thaliana* (Hashimoto et al., 2003). Furthermore, the moss PPR_38 has a role in intergenic RNA cleavage between *clpP* and 5’-*rps12* and in the splicing of *clpP* pre-mRNA resulting in affected steady state level of ClpP protease (Hattori et al., 2007).

PPR proteins are also known to assist in the translation activation directly or by recruiting the components of translational machinery (Schmitz-Linneweber et al., 2005; Wang et al., 2006). CRP1 is a chloroplast localized PPR protein in maize which activates the translation of *petA* and the *psaC* RNA encoding subunit A of ATP synthase and subunit C of photosystem I, respectively (Fisk et al., 1999; Hashimoto et al., 2003; Meierhoff et al., 2003; Schmitz-Linneweber et al., 2005). Another maize protein, PPR2, is localized to chloroplast stroma. PPR2 mutants show the barred accumulation of ribosomes in chloroplast (Williams and Barkan, 2003). PPR336 is mitochondrial targeted protein in *A. thaliana* associated with polysomes and speculated to have a role in translation (Uyttewaal et al., 2008). In yeast, the aforementioned Pet309 is also implicated in translation (Tavares-Carreón et al., 2008).

Furthermore, PPR proteins may have a role in retrograde signalling. One example is the protein genomes uncoupled 1 (GUN1), a chloroplast localized PPR protein, implicated in signalling from plastids to nucleus (Koussevitzky et al., 2007).

So far the exact working mode of the above described PPR proteins is largely unkown. One outstanding study investigating the PPR10 protein from maize, which is involved in the expression of chloroplast encoded transcripts sheds light on the molecular events (Pfalz et al., 2009). This study provides a model on the mode of action of PPR proteins in protecting the transcripts from exonucleases. The authors have mapped the processed termini in the *atpI-atpH* and *psaJ-rpl33* intergenic regions and found that the processed RNAs overlap by approximately 25 nucleotides. PPR10 associates to these 25 nucleotides found in the intergenic regions of *atpI-atpH* and *psaJ-rpl33.* According to the model, the processing of polycistronic plastid transcripts is initiated by endonucleases. In case of *atpI-atpH and psaJ-rpl33* intercistronic regions, when endonucleolytic cleavage opens the way for exonucleases, they are stalled at PPR10 RNA complex from either the 5’ or 3’ direction. This results in the accumulation of processed RNAs whose 5’ or 3’ termini are defined by the upstream or downstream bound PPR10 (Pfalz et al., 2009). These findings illustrate that PPR10 serves as a barrier to the exonucleases from either the 5’ or 3’ end and also that the bound PPR10 provides a substitute to a hairpin structure as an obstacle for 3’ exonucleases (Pfalz et al., 2009).

From the data available for PPR protein functions, it is obvious that they are essential for organellar RNA metabolism. They constitute a versatile organellar regulatory protein family in eukaryotes and are particularly found in large numbers in land plants (Lurin et al., 2004; O'Toole et al., 2008).

Table 1.1: List of cloned pentatricopeptide repeat proteins in various organisms involved in organellar gene expression. Loc, localization; Mt, mitochondria; Cp, chloroplast; N, nucleus

Organism	Protein	Function	Loc.	Reference
A. thaliana	HCF152	Processing and /or stabilization of operon *psbB-psbT-psbH-petB-petD*	Cp	Meierhoff et al., 2003; Nakamura et al., 2003 & 2004
	CLB19	RNA Editing of *rpoA* and *clpP*	Cp	Chateigner-Boutin et al., 2008
	OTP43	*Trans*-splicing of mt *nad1* intron I	Mt	de Longevialle et al., 2007
	OTP51	*Cis*-splicing of *ycf3* intron 2 and other group II introns	Cp	de Longevialle et al., 2008
	OTP82	editing of *ndhB* and *ndhG*	Cp	Okuda et al., 2010
	CRR2	RNA processing *rps7* and *ndhB*	Cp	Hashimoto et al., 2003
	CRR4	RNA editing of *ndhD*	Cp	Kotera et al., 2005; Okuda et al., 2006
	CRR21	RNA editing of *ndhD*	Cp	Okuda et al., 2007
	CRR22	Editing of *ndhB*, *ndhD* and *rpoB*	Cp	Okuda et al., 2009
	CRR28	Editing of *ndhB* and *ndhD*	Cp	Okuda et al., 2009
	PTAC2	Plastid biogenesis	Cp	Pfalz et al., 2006
	PGR3	Stabilization of *petL* operon	Cp	Yamazaki et al., 2004
	P67	Processing/ translation of RNAs	Cp	Lahmy et al., 2000
	LOJ	Lateral organ development and boundry formation	Cp	Prasad et al., 2005
	At1g53330	Embryo development	Mt	Kocabek et al., 2006
	PPR40	Regulation of cytochrome *c*	Mt	Zsigmond et al., 2008
	DG1	Regulation of PEP dependent Cp gene expression	Cp	Chi et al., 2008
	GRP23	Interacts with RNA polymerase II subunit III, involved in transcription-al regulation	N	Ding et al., 2006
	YS1	Editing of *rpoB* transcript	Cp	Zhou et al., 2008
	LPA66	Editing of *psbF* transcript	Cp	Cai et al., 2009
	MEF1	Editing *rps4*, *nad7* and *nad2* tran-scripts	Mt	Zehrmann et al., 2009
	MEF9	Editing *nad7* transcript	Mt	Takenaka, 2010
	MEF11	Editing *cox3*, *nad4* and *ccb203* transcripts	Mt	Verbitskiy et al., 2011
	AtECB2	Editing of *accD* transcript and role in early chloroplast biogenesis	Cp	Yu et al., 2009

Organism	Protein	Function	Loc.	Reference
	EMB175	Embryogenesis	Cp	Cushing et al., 2005
	PPR336	Associated with polysomes (Translation)	Mt	Uyttewaal et al., 2008
	GUN1	Retrograde signaling	Cp	Koussevitzky et al., 2007; Ruckle and Larkin, 2009
O. sativa	OsPPR1	Chloroplast biogenesis	Cp	Gothandam et al., 2005
	Rf1a	Fertility restoration and RNA cleavage	Mt	Wang et al., 2006; Kazama et al., 2008
	Rf1b	Fertility restoration	Mt	Wang et al., 2006
Zea maize	CRP1	Processing of the *petD* and translation of the *petA* and *petD* RNAs	Cp	Fisk et al., 1999; Schmitz-Linneweber et al., 2005
	PPR2	Required for plastid ribosome accumulation	Cp	Williams and Barkan, 2003
	PPR4	*Trans*-splicing of *rps12* RNA and ribosome biogenesis	Cp	Schmitz-Linneweber et al., 2006
	PPR5	Stabilization of the *trnG*-UCC tRNA precursor	Cp	Beick et al., 2008; Williams-Carrier et al., 2008
	PPR10	Stabilization by interacting with intergenic RNA regions of *atpI-atpH* and *psaJ-rpl33*	Cp	Pfalz et al., 2009; Prikryl et al., 2011
	EMP4	Unknown	Mt	Gutiérrez-Marcos et al., 2007
P. patens	PpPPR_38	Intergenic RNA cleavage between *clpP* and 5' *rps12* and the splicing of *clpP* RNA	Cp	Hattori et al., 2007 ; Hattori and sugita, 2009
	PpPPR_71	RNA editing of the *ccmFc* transcript	Mt	Tasaki et al., 2010
	PpPPR_78	RNA editing of *rps14* and *cox1* mRNA	Mt	Uchida et al., 2011
	PpPPR_79	RNA editing of *nad5-1*mRNA	Mt	Uchida et al., 2011
C. reinhardtii	MCA1	RNA stabilization of *petA* mRNA	Cp	Loiselay et al., 2008
	MRL1	RNA stabilization of *rbcL* mRNA	Cp	Johnson et al., 2010
S. cerevisae	Pet309	RNA stabilization and translation of *cox1* mRNA	Mt	Manthey and McEwen, 1995; Taveres-Carreon et al., 2008
	Aep3	RNA stabilization and maturation of Bicistronic mRNA of subunits 6 and 8 of the H^+-translocating ATP Synthase	Mt	Ellis et al., 2004
	DMR1/ Ccm1p	Splicing of fourth intron of *cob* and *cox1* pre-mRNAs and maintenance of mitochondrial 15S ribosomal RNA	Mt	Moreno et al., 2009 Puchta et al., 2010

It is generally assumed that PPR proteins act as adapters to direct the processing machinery to the correct sites at the correct time (Delannoy et al., 2007; Chateigner-Boutin et al., 2008). The gene expression system of chloroplasts has gained complexity in terms of RNA maturation as compared to their cyanobacterial ancestor (Maier et al., 2008). One explanation may be that the expansion of PPR proteins occurred due to the gain in complexity of organellar RNA metabolism. On contrary, it has been suggested that the expansion of nucleus encoded regulatory factors is due to enhanced rate of mutations in the organelle genome, which are compensated by the faster evolving nuclear genome (Maier et al., 2008). The idea for such an assumption comes from the PPR proteins that inhibit (as the *restorer of fertility* in CMS plants) the expression of toxic gene products and relapse to a functional gene expression (Chase, 2007; Maier et al., 2008). A similar explanation can be given for the editing process, which might represent the suppression of a point mutation rather than performing a regulatory function (Maier et al, 2008; Schmitz-Linneweber and Small, 2008).

1.5 Aims of this study

Organellar gene expression is characterized by complex RNA metabolism, which requires the import of many nucleus-encoded RNA-binding proteins. The eukaryote-specific family, termed as pentatricopeptide repeat (PPR) proteins has got a more significant role in almost all the steps of gene expression which are associated with RNA, i.e. transcription to translation in plant organelles (Schmitz-Linneweber and Small 2008; Stern et al., 2010; Barkan, 2011). The PPR protein family has greatly expanded in higher plants during evolution (O'Toole et al., 2008). In contrast to higher plants, the genome of the unicellular green alga *C. reinhardtii* encodes only 11 PPR proteins as identified by *in silico* analysis. Considering *C. reinhardtii* as an evolutionary ancestor of higher plants, it appears to contain a basic set of PPR proteins. The small number of algal PPR proteins compared to higher plants, facilitates the analyses of the "ancient" functions of this "basic set" of PPR proteins in a phototrophic eukaryote using a systematic approach. So far, only two PPR proteins from *C. reinhardtii*, MCA1 and MRL1 have been assigned a function i.e. the stabilization of the chloroplast *pet*A mRNA and *rbcL* mRNA (Lown et al., 2001; Loiselay et al., 2008; Johnson et al., 2010), respectively.

Considering the potential role of PPR proteins in organellar RNA metabolism the present study was designed to extend the knowledge of ancient PPR proteins found in *C. reinhardtii*.

To investigate the functions of PPR proteins, first, an *in silico* analysis of their of primary structure features was performed. For an *in vivo* approach, use of the reporter green fluorescent protein (GFP) based localization was carried out for a subset of 4 PPR proteins, including PPR1, PPR3, PPR4 and PPR7. The initial analysis of PPR proteins presented in this

study suggested PPR7 as a promising candidate for further functional analysis based on its clear annotated gene model and importantly, its localization to the chloroplast. Detailed analyses were performed for PPR7 including i) RNA co-immunoprecipitation and following RIP-chip analysis to determine putative organellar target RNAs ii) systematic down regulation of PPR7 using the RNAi technique and the phenotypical characterization of PPR7 deficient mutants iii) investigating the state of target RNAs identified, in PPR7 deficient mutants, to discern the possible function of the protein under study.

2 MATERIALS AND METHODS

2.1 Materials

All chemicals used in this study had a p.A. quality and were purchased from the following companies: Roth, Sigma, Merck and AppliChem if not indicated otherwise. The instruments used in this study are mentioned in the text. An overview of suppliers including their addresses is found in Table 2.1.

Table 2.1: List of all suppliers for chemicals, enzymes and laboratory equipment

Supplier	Address
Agrisera	Agrisera AB, Vännäs, Sweden
Alpha Innotech	Alpha Innotech Corporation, San Leandro, USA
Amersham Biosciences	Amersham Biosciences Europe GmbH, Freiburg, Germany
AppliChem	AppliChem GmbH, Darmstadt, Germany
Biometra	Biometra GmbH, Göttingen, Germany
Biozym	Biozym Diagnostik GmbH, Hameln, Deutschland
BioRad	Bio-Rad Laboratories, München, Deutschland
Epicentre biotechnology	Epicentre Biotechnologies, Madison, USA
Fermentas	Fermentas GmbH, St. Leon-Rot, Germany
Invitrogen	Invitrogen GmbH, Karlsruhe, Germany
Metabion	Metabion GmbH, Martinsried, Germany
Miltenyi Biotec	Miltenyi Biotec, Bergisch Gladbach, Germany
Millipore	Millipore Corporation, Bedford, USA
MWG Biotech	Eurofins MWG operon, Ebersberg, Germany
PeqLab	PeqLab Biotechnologie, Erlangen, Germany
Photon Systems Instruments	Photon Systems Instruments, Högrova, Czech Republic
Pierce	Pierce, Rockford, USA
Promega	Promega Corporation, Madison, USA
Qiagen	Qiagen, Hilden, Germany
Roche	Roche Diagnostics GmbH, Mannheim, Germany
Roth	Carl Roth GmbH & Co, Karlsruhe, Germany
Serva	Serva Feinbiochemika, Heidelberg, Germany
Sigma	Sigma Chemical Company, St. Louis, USA
Stratagene	Stratagene, CA, USA
Thermo Scientific	Thermo Scientific, Rockford, USA
Whatman	Whatman Paper, Maidstone, England
Zeiss	Carl Zeiss MicroImaging GmbH, Göttingen, Germany

2.1.1 Enzymes

The enzymes were used with specific buffer systems supplied by the corresponding companies. Restriction enzymes were bought from Fermentas and Promega, T4 Ligase and Protease Inhibitor cocktail (PIC) from Roche and RNase A from Roth.

2.1.2 Oligonucleotides

All oligonucleotides were ordered from Metabion or from Invitrogen. Lyophilised oligonucleotides were resuspended in sterile, ddH_2O to a final concentration of 100 pmol/µL and stored at -20°C. Sequences of used oligonucleotides are denoted in respective chapters in Methods.

2.1.3 DNA-Vectors

All DNA-vectors used in this work are outlined along with their characteristics in Table 2.2

Table 2.2: List of DNA-vectors used for this research work

Plasmid	Characteristics	Reference
pJET1.2/blunt	Cloning vector; confers ampicillin resistance in *E. coli*	Fermentas
pGEX4T-1	Overexpression vector for GST based recombinant fusion proteins under control of *lac* promoter; confers ampicillin resistance in *E. coli*	GE Healthcare
pQE30	Overexpression vector for 6 x His tagged recombinant proteins under control of *T7* promoter; confers ampicillin resistance in *E. coli*	Qiagen
pBC1-CrGFP	pBC1 expression vector containing the *C. reinhardtii* codon adapted GFP coding sequence (CrGFP) under control of the *PsaD* 5' and 3' UTR; Confers paromomycin resistance in *C. reinhardtii* by expression of the *APHVIII* gene and ampicillin resistance in *E. coli*	Neupert et al., 2009
NE537	RNAi vector used for knockdown of target gene. It contains inverted repeats of the *Maa7* gene under control of *RbcS* promoter. Confers paromomycin resistance in *C. reinhardtii* by expression of the *APHVIII* gene and ampicillin resistance in *E. coli*	Rohr et al., 2004

2.1.4 Reaction systems (Kits)

The following kits were used in this study according to the manufacturer's protocols:

- Perfectprep Gel Cleanup Kit (Eppendorf)
- TripleMaster PCR System (Eppendorf)
- CloneJET PCR Cloning Kit (Fermentas)
- Plasmid Mini and Midi Kits (Qiagen)
- DNeasy Plant Mini Kit (Qiagen)
- TriReagent (Sigma)

2.1.5 Bacterial Stains

Recombinant plasmids were propagated in *Escherichia coli* (*E. coli*) strain XL1-Blue [*endA1 gyrA96 hsdR17 lac recA1 relA1 supE44 thi-1 F´proAB lacIq Z_M15 Tn10* (*Tetr*)] (Stratagene). For overexpression of recombinant proteins, BL21-DE3 (F´, *omp*T, *hsd*SB (rB - mB-), *gal*, *dcm*; and M15 (NaI^S, Str^S, Rif^S, Thi^-, Lac^-, Ara^+, Gal^+, Mtl^-, F^-, $RecA^+$, Uvr^+, Lon^+) strains from Stratagene were used.

2.1.6 *C. reinhardtii* strains

The *C. reinhardtii* strains used in this research work are stated in the following Table 2.3.

Table 2.3: List of *C. reinhardtii* strains

Strain	Description	Reference
CC406	Cell wall deficient (*cw15)* wild type strain	Genetic Centre, Duke University, Durham, North Carolina; Davies DR, Plaskitt A, 1971
UVM4	UV mutagenized cell wall deficient strain for overexpression of transgene	Neupert et al., 2009
nac-2-26	Cell wall deficient (*cw15*) Photosystem II mutant	Boudreau et al., 2000
raa1-314B	Cell wall deficient (*cwd*) Photosystem I mutant	Merendino et al., 2006
mrl1-1	Mutant lacking stability factor MRL1 for *rbcL* and hence lacks Rubisco	Johnson et al., 2010
Δ*rbcL-1*	Cell wall deficient *(cw15) rbcL* deletion mutant	Johnson et al., 2010
XS1	Cell wall deficient (*cw15*) recipient wild type for generation of the *ΔrbcL* strain, arg7	Johnson et al., 2010

2.2 Methods

2.2.1 Growth of Bacterial strains

Bacteria were grown in LB medium (1% tryptone, 1% NaCl, 0.5% yeast extract, pH 7) under standard conditions (Sambrook and Russell 2001). The LB media for plates were solidified by adding 1% agar. The media was cooled down to approximately 60°C after autoclaving, followed by the addition of selectable antibiotics under sterile conditions.

2.2.2 Growth of *C. reinhardtii* strains

C. reinhardtii strains were maintained at 23°C on Tris-acetate-phosphate (TAP) agar medium (Harris, 2009) in medium light (30 µE/m^2/s) if not indicated otherwise. Liquid cultures were grown to a density of ~2 x 10^6 cells/mL in TAP medium containing 1% sorbitol (TAPS). For some experiments *C. reinhardtii* strains were also cultured on minimal medium (HSM) (Sager and Granick, 1953).

2.2.3 Nucleic acids

2.2.3.1 Isolation of nucleic acids

2.2.3.1.1 Plasmid isolation from *E. coli*

Plasmid DNA isolation from *E. coli* at small scale was performed by alkaline lysis of bacteria using standard protocol as described (Sambrook and Russell, 2001). For large scale and pure plasmid DNA isolation, Midi Kit (Qiagen) was used according to the manufacturer's protocol.

2.2.3.1.2 Isolation of genomic DNA from *C. reinhardtii*

The genomic DNA of *C. reinhardtii* was isolated from 50 mL (2 x 10^6 cells/mL) *cw15* CC406 liquid culture. Culture was harvested at 1100 x g, 4^{o}C for 6 min and DNA was extracted using the DNeasy Plant Mini Kit (Qiagen) according to manufacturer's protocol.

2.2.3.1.3 Isolation of total cellular RNA from *C. reinhardtii*

Liquid culture of *C. reinhardtii* strains under investigation were harvested at early log phase (~1 x 10^6–2 x 10^6 cells/mL) at 1100 x g, 4°C for 6 min and total cellular RNA was extracted using the TRI reagent (Sigma), according to the manufacturer's instructions.

2.2.3.2 Determination of nucleic acid concentrations

The quality and quantity of nucleic acids were examined optically in ethidium bromide-stained agarose gels (section 2.2.3.3). Additionally, UV absorption at 260 nm was measured in a Novaspec III photometer (Amersham Biosciences), and concentrations were calculated assuming an optical density OD260 = 1 to correspond to 50 µg/mL for double-stranded DNA and 40 µg/mL for RNA. The ratio of absorption at 260 and 280 nm was used as a measure for the purity of the sample.

2.2.3.3 Nucleid acid electrophoreses

2.2.3.3.1 Agarose gel electrophoresis of DNA

Fragments of DNA were separated in TAE (Tris-acetate-EDTA) agarose gels of varying concentrations (1% to 2%) depending on the size of the DNA fragments in a horizontal gel apparatus (i-Mupid, Advance). Agarose was dissolved in 1x TAE buffer (40 mM Tris/Acetic acid pH 8.0, 2 mM EDTA) through heating. After cooling down the gel to ~50°C, ethidium bromide was added to a final concentration of 0.5 µg/mL. Samples were supplemented with 6x loading buffer [0.25% (w/v) bromophenol blue, 30% (v/v) glycerol), 60 mM EDTA, 50 mM Tris-HCl pH 7.5] and electrophoretically separated at 100 V. For visualization, gels were analysed under UV light using the Alpha Imager station (Alpha Innotech Corporation). 5 µL of standard lambda DNA marker (Fermentas) were used to control the length and to quantify the mass of the separated DNA fragments.

2.2.3.3.2 Agarose gel electrophoresis of RNA

Total cellular RNA was electrophoretically separated from formaldehyde-agarose gels of varying concentrations (1% to 2%), depending on experiment. For this purpose, agarose was dissolved in ddH_2O through heating. Upon cooling to ~ 60°C, formaldehyde (1/40 vol) and 1 x MOPS buffer (20 mM MOPS, 5 mM Na-acetate, 1 mM EDTA pH 7) was added to final vol-

ume. RNA samples were denatured with equal volume of 2 x RNA loading dye (50% (v/v) formamide, 17.5% (v/v) formaldehyde, 20% glycerol, 1.25 mM EDTA, pH 8.0, 1.27 mM ethidium bromide; 0.2% (w/v) bromophenol blue, 0.2% (w/v); in 1 x MOPS buffer) and heating at 65°C for 5 min prior loading to the gel. Electrophoresis was carried out in a horizontal electrophoresis gel chamber (Life Technologies) at 80 V. Fractionated RNA molecules were visualized by UV *trans*-illumination by using the Alpha Imager station (Alpha Innotech Corporation). 2 µl of high range RNA marker (Fermentas) was used to control the length of the separated RNA fragments.

2.2.3.4 cDNA synthesis and RT-PCR

Reverse transcription (RT) was performed with 100–500 ng of co-immunoprecipitated RNA using Reverse Transcriptase (Epicentre) and gene-specific primers used for Northern probe generation (Table 2.5) according to the supplier's instructions. DNase I (RNase free; Promega Corp.) was used for removal of DNA from RNA preparations prior to RT-PCR reactions. The template cDNA used for PCR was 0.25 µL per reaction.

2.2.3.5 Cloning

DNA ligation by means of T4 DNA ligase enzyme, restriction cleavage by endonuclease enzymes, dephosphorylation by means of alkaline phosphatase and other manipulations were performed according to the standard procedures as described (Sambrook and Russell, 2001). Restriction enzymes, alkaline phosphatase and T4 DNA ligase were purchased from Fermentas. For direct cloning of PCR products in the pJet1.2 cloning vector, Clonejet cloning kit (Fermentas) was used according to the manufacturer's instructions.

2.2.3.5.1 Transformation of *E. coli*

Plasmid DNA or ligation products were added to 50 µL of competent XL1 blue or BL21 cells and incubated on ice for 5 min. After the incubation, the cells were transferred to a heating block at 42°C for exactly 45 seconds and then immediately cooled on ice for 2 min. After the heat shock 500 µL LB medium were added to cells and they were incubated for 20 min at 37°C and 180 rpm. The recovering cells were transferred on selective LB agar plates and incubated from 14 to 16 h at 37°C.

2.2.3.5.2 Polymerase chain reaction (PCR)

Specific DNA sequences from genomic, plasmid and cDNA were amplified using Taq DNA polymerase enzyme and the Master Cycler (Eppendorf). Standard 50 µL PCR reactions were prepared with PCR buffer (670 mM Tris-Cl pH 8.0, 67 mM $MgCl_2$ and 0.01% Tween20), 200 µM dNTPs and 10 pmol of each primer. The amount of DNA template varied from 1 ng (plasmid DNA) to 100 ng (genomic DNA). After 5 min denaturation step at 95°C, 20-40 cycles were performed (depending on experiment), which include denaturation at 94°C (1 min), annealing (depending on Tm of primer set) at 50-65°C (1 min), and extension at 72°C (1 min/kb). To minimize the presence of possibly incomplete amplification, an additional elongation step was carried out at 72°C for 10 min. The analysis of PCR products was carried out by agarose gel electrophoresis (2.2.3.3.1).

2.2.3.5.3 Sequencing

For the sequencing of plasmids, samples containing 150–300 ng of DNA with 10 pmol of corresponding primer (Table 2.4) were prepared. The sequencing of DNA samples was carried out by the institute of Genetics LMU.

Table 2.4: List of oligonucleotides used for sequencing

Name	**Sequence 5' – 3'**	**Experiment**
pGEX Fw	ATCCTCCAAAATCGGATCTG	Protein overexpression
NE537 Fw	AAGGCGTGCTTTGGTGAGAC	RNAi lines generation
PsaD Fw	AGGTTTCCTCGCCGAGCAAG	GFP constructs
PsaD Rv	TCCGATCCCGTATCAATCAG	GFP constructs
CrGFP Rv	TTGTACAGCTCGTCCATGCCG	GFP constructs

2.2.3.6 Probe labelling and transcript accumulation analyses (Northern blot)

Probes for Northern blot analyses were produced by PCR (section 2.2.3.5.2) by using the respective oligonucleotides for the transcript under study denoted in Table 2.5 and DIG-11-dUTP (Roche).

Table 2.5: List of oligonucleotides used for Northern probe labelling and RT-PCR

Name	Sequence 5' – 3'
rrnS 5' Fw	TTGCGTCTGATTAGCTAGTTG
rrnS 5' Rv	GACGCTTTACGCCCAATC
Prec *rrnS* 5' Fw	GGCAGTGGTACAATAAATAAATTG
Prec *rrnS* 5' Rv	TCGGGATTTTAAACCCTTTTG
rrnL Intron Fw	GCATCAGCTATCACTCGC
rrnL Intron Rv	GTCTAGCACAGCACGAAC
trnE2-psbH Fw	CAGCTAGCCTTAACAAACAG
trnE2-psbH Rv	ACAGGAACTTCTAAAGCTAAAC
rpoC2 Fw	CTGAGCCATTTATTGCAAAAC
rpoC2 Rv	ACAAGGTTGTTTAGGAATATGT
rbcL Fw	AAGATTCAGCAGCTACAGC
rbcL Rv	CACTGCCTCTAATAAAGTCTAC
atpA Fw	GCCACTGTTCACTCCTC
atpA Rv	TCTGGAGTACGCATTGCC
cemA-atpH Fw	TACAACCAAATAGGTTTCAATAG
cemA-atpH Rv	CCATACCAGGACCAATAGC
tscA Fw	TGATCGCTCTAATATTATTACG
tscA Rv	CGGCATTACTTGTTGTTTATC
psaJ-atpI Fw	CAATTAATAAACCTGCTGTAAAAG
psaJ-atpI Rv	GGTGTATTTACTAGTGCTATCC
psbD Fw	GCCGTAGGGTTG AATG
psbD Rv	GTTGGTGTCAACTTGGTGG
ChlL Fw	GTTGTTTGTGGTGGCTTTGC
ChlL Rv	CATCCATATGAGCCGAAGTC

The procedure for Northern blot was performed as described (Sambrook and Russell, 2001). In brief, total RNA separated on 1-2% denaturing formaldehyde agarose gel, was transferred to Roti nylon^{+} membrane (Roth) having pore size of 0.45 µm, followed by UV light cross-linking (UV Crosslinker, UVC 500, Hoefer). The membrane was pre-hybridized with pre-hybridization Buffer (20% SDS, 0.25 M Na_2HPO_4 pH 7.2, 1 mM EDTA) and 0.5% blocking reagent (Roche) for at least 1 hour at 68°C before addition of DIG labelled probe. The probe was applied to a concentration of 2.5 ng/mL of pre-hybridization buffer. After hybridization with probe at 68°C overnight, the membrane was washed 3 times at 65°C for 20 min with hybridization wash buffer (20 mM Na_2HPO_4, 1 mM EDTA and 1% SDS). After washing, the blot was incubated with blocking buffer (100 mM maleic acid, 3 M NaCl, 0.3% Tween20 and 0.5% blocking reagent) for 1 hour at room temperature, followed by DIG antibody (Roche)

incubation for 30 min. Blot was washed 4 times, each 10 min with wash buffer (100 mM maleic acid, 3M NaCl, 0.3% Tween20) and signals were visualized by chemiluminescent detection using the substrate, CDP-STAR (Roche) in substrate buffer (100 mM Tris-Cl pH 9.5, 100 mM NaCl and 50 mM $MgCl_2$).

2.2.4 Protein methods

2.2.4.1 Determination of protein concentrations

Protein concentrations were measured as described by Bradford (1976), using the Roti®-Quant protein assay.

2.2.4.2 SDS polyacrylamide gel electrophoresis (SDS PAGE)

To separate proteins according to their molecular weight, SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) was carried out as described by Laemmli (1970). Proteins were separated on discontinuous polyacrylamide gels. BioRad MiniProtean II gel system (BioRad) was used to pour and run the gels. Samples for the gel electrophoresis were prepared by mixing required amount of protein with 5 x Laemmli buffer, consisting of 10% SDS, 20% glycerol, 20 % ß-mercaptoethanol and 0.1% bromophenol blue in 250 mM Tris-HCl pH 6.8. Samples were denatured either at 95ºC for 10 min (for soluble proteins) or at RT for 30 min (for membrane proteins) and loaded on the gel. The running buffer used for gel electrophoresis consisted of 25 mM Tris-Cl, 0.192 M glycine and 0.1 % SDS. Proteins were stained with Coomassie solution (50% methanol, 12% acetic acid and 0.1% Coomassie Brilliant Blue G-250 or R-250) for 30 minutes and destained with a destaining solution (10% ethanol, 10% acetic acid).
For more sensitive staining or qualitative control of highly purified proteins, polyacrylamide gels were stained by silver. For silver staining, the gels were incubated on a rotary shaker for 1 h in solution B, 15 min in solution A, 10 min in solution B, 5 min in solution D+E (1:1), 5 min in solution B, 10 min in solution C, 10 min in H_2O, 10 min in solution F, 5 min in H_2O, 2-20 min in solution G and finally for 30 min in solution B.

Silver stain solutions	
Solution A:	20% TCA, 50% Methanol, 2% $CuCl_2$ and 0.1% Formaldehyde
Solution B:	10% Ethanol, 5% Acetic acid
Solution C:	10% Ethanol

Continued from previous page

Solution D:	0.01% $KMnO_4$
Solution E:	0.01% KOH
Solution F:	0.2% $AgNO_3$
Solution G:	2% Na_2CO_3, 0,1% Formaldehyd

2.2.4.3 Immunoblotting

After SDS-PAGE (section 2.2.4.2), electrophoretically separated proteins were transferred for 1.5 h at 0.8 mA/cm^2 to a nitrocellulose membrane (pore size 0.45 µm, AppliChem) in transfer buffer (48 mM Tris, 39 mM glycine, 0.037 % SDS, 20% MeOH) as described (Towbin et al., 1979) using a semi-dry blot transfer apparatus (Peqlab). The transferred proteins were stained with Ponceau S solution (0.2% Ponceau S, 1% acetic acid) for examining proper transfer and equal loading. Finally, the nitrocellulose membrane was incubated with blocking buffer (5% milk powder in TBS-Tween 0.1%, pH 7.5) for 1-2 h at room temperature (RT). Afterwards the membrane was incubated with the primary antibody in blocking buffer at 4°C overnight (see Table 2.6 for used antibodies and their respective dilutions). Then nitrocellulose membrane was washed 3 times, each 10 min in TBS-Tween (0.1%), and afterwards incubated with the secondary antibody for 1 h at RT. Membrane was again washed 3 times in TBS-Tween (0.1%), each 10 min. The HRP conjugated antibody was detected with enhanced chemiluminescence (ECL) solution (Thermo Scientific) and quantified using AIDA Image Analyser V3.25 (Raytest Isotopenmessgeräte GmbH, Straubenhardt Germany).

Table 2.6: List of antibodies and respective titers used in this research work

Antibody	**Titer**	**Reference**
Primary:		
Anti-GFP-HRP	1:5000	Miltenyi Biotec
Anti-PPR7	1:500	this study
Anti-RbcL	1:1000	A
Anti-PsaA	1:1000	Agrisera
Anti-Cyt-*b6*	1:1000	Agrisera
Anti-AOX2	1:5000	Agrisera
Anti-HSP70B	1:10000	B
Secondary:		
anti-rabbit IgG HRP	1:10000	GE Healthcare

[a] kindly provided by G. F. Wildner (Ruhr Universität Bochum)
[b] kindly provided by Michael Schroda (Universität Freiburg Germany)

2.2.4.4 Isolation of total protein extracts from *C. reinhardtii*

For isolation of total protein extracts, cells were placed into 20 ml of liquid TAPS medium and allowed to grow till mid-log phase (~2 x 10^6–3 x 10^6 cells/mL) on a rotary shaker (125 rpm) under dark, medium light (30 µE/m^2/s) or high light (100 µE/m^2/s), depending on experiment. *C. reinhardtii* cell cultures were harvested (1100 x g, 6 min, 4 °C) and the cell pellet was re-suspended in 500 µL 2 x lysis buffer (120 mM KCl , 0.4 mM EDTA , 20 mM Tricine pH 7.8, 5 mM ß-mercaptoethanol, 0.2% Triton X 100 and Roche Complete Mini protease inhibitors). Cells were lysed mechanically, by pipetting up and down for several times. After determination of protein concentration (section 2.2.4.1), equal amounts of total protein were subjected to SDS-PAGE analysis (section 2.2.4.2).

2.2.4.5 Isolation of total soluble protein extracts from *C. reinhardtii*

For isolation of total soluble protein extracts, *C. reinhardtii* liquid cell cultures were harvested (1100 x g, 6 min, 4°C) and the cell pellet was resuspended in 500 µL hypotonic solution (10 mM Tricine/KOH, pH 7.8, 10 mM EDTA, 5 mM ß-mercaptoethanol and Roche Complete Mini protease inhibitors). Cells were first resuspended by pipetting up and down for several times and then lysed by sonication (5 pulses, 3 times, 10% available power) with Bandelin Sonopuls sonicator (Bandelin electronic, Berlin). The cell membranes and insoluble material was pelleted by centrifugation (10,000 g, 4^{o}C, 10 min). The supernatant obtained was considered as total soluble protein extract.

2.2.4.6 Chloroplast isolation from *C. reinhardtii*

Chloroplasts from a cell wall–deficient strain carrying the *cw15* mutation were isolated by a discontinuous percoll gradient (45 to 75%) in 1x isotonic solution, as described (Zerges and Rochaix, 1998). Briefly, cells (~2 x 10^6 cells/mL) were harvested by centrifugation (4000 x g, 10 min, 4°C) and the pellet resuspended in 25-45 mL/L culture in 1x isotonic buffer (300 mM sorbitol, 5 mM $MgCl_2$, 10 mM Tricine pH 7.8). Cell membrane was broken for 5 min on ice by adding 1-5% saponin (in 1x isotonic buffer) depending on strain. This suspension was centrifuged for 10 sec and the supernatant was discarded. Pellet obtained from 1 L culture was resuspended in 8 mL 1x isotonic buffer and loaded on the discontinuous percoll gradient. Chloroplasts were collected at the interphase of the two gradient steps after centrifugation (5000 x g, 20 min, 4°C) and washed with twice the volume of isotonic buffer. After centrifugation (4000 g, 10 min, 4°C) chloroplast pellet was resuspended and mechanically lysed by pipetting up and down in 2 x CP lysis buffer (120 mM KCl, 0.4 mM EDTA, 20 mM Tricine pH

7.8, 5 mM ß-mercaptoethanol, 0.2% TritonX 100 and Roche Complete Mini protease inhibitors) for total chloroplast proteins. All the steps were carried out at 4^{o}C.

2.2.4.7 Chloroplast fractionation of *C. reinhardtii*

The fractionation of chloroplast to stroma and thylakoids was carried out as described (Ossenbühl and Nickelsen, 2000). Briefly, the pelleted chloroplasts (section 2.2.4.6) were osmotically lysed in reducing hypotonic buffer (10 mM Tricine/KOH, pH 7.8, 10 mM EDTA, 5 mM ß-mercaptoethanol and Roche Complete Mini protease inhibitors) by repeated pipetting. Membranes were removed by ultracentrifugation for 30 min at 100,000 g through a 1 M sucrose cushion in hypotonic buffer. The supernatant obtained was taken as stromal fraction and the pellet as crude thylakoids for immunoblot analysis (section 2.2.4.3). For size exclusion chromatography (section 2.2.4.9) and co-immunoprecipitation experiments (section 2.2.8), chloroplasts were lysed in non-reducing hypotonic buffer (without ß-mercaptoethanol). All the steps were carried out at 4^{o}C.

2.2.4.8 Mitochondria isolation from *C. reinhardtii*

Mitochondria from a cell-wall–deficient strain carrying the *cw15* mutation were isolated as described (Eriksson et al., 1995). Briefly, cell cultures were grown under continuous light (30 µE/m^2/s) to a cell density of ~3 x 10^6 cells/mL. Cells were pelleted by centrifugation (400 x g, 5 min, 4^{o}C) and washed twice with wash buffer (20 mM Hepes-KOH, pH 7.2). Washed cells were resuspended in breaking buffer (50 mM Hepes-KOH, pH 7.2, 5 mM EDTA, 0.25 M sorbitol, 4 mM Cys, 0.5% Polyvinylpyrrolidone (PVP) 40, 0.1% BSA) to a final concentration of 0.5 g cell pellet/mL. Cells were broken with double volume of 0.45 mm glass beads at lowest vortex speed for 1 min in a 50 mL plastic centrifugation tube. The supernatant was collected in a separate tube and cell homogenate was rinsed from the glass beads by addition of 30 mL of breaking buffer and was added to the supernatant. Contaminating chloroplast membranes were pelleted by centrifugation (2000 x g, 5 min, 4^{o}C) and the supernatant was further centrifuged (12000 x g, 20 min, 4^{o}C) to obtain crude mitochondria as pellet. The mitochondria pellet was purified by resuspending in assay buffer (10 mM potassium phosphate buffer, pH 7.2, 0.1% BSA, 0.25 M sorbitol, 10 mM KCl, 5 mM $MgCl_2$) and were homogenized in a potter homogenizer by passing plunger 3 times slowly. Homogenized resuspension was layered on 30 mL of Percoll solution (20% Percoll, 0.25 M sorbitol, 10 mM MOPS-KOH, pH 7.2, 1 mM EDTA, 0.1% BSA, 0.5% PVP 40) and centrifuged (20,000 x g, 40 min, 4^{o}C). The pale hazy band at bottom (~3 mL) containing mitochondria was diluted with 40 mL of wash

buffer (10 mM potassium phosphate buffer, pH 7.2, 0.1% BSA, 0.25 M sorbitol, 1 mM EDTA) and centrifuged again (10,000 x g, 10 min, 4^{o}C). The pellet was resuspended in assay buffer and mitochondria were disrupted by sonication (5 pulses, 2 times, 10% available power) with Bandelin Sonopuls sonicator (Bandelin electronic, Berlin). All the steps were carried out at 4^{o}C.

2.2.4.9 Size exclusion chromatography

For size exclusion chromatography (SEC) of PPR7, stromal extracts were obtained from *C. reinhardtii* culture, harvested at an early log phase (~1 x 10^6–2 x 10^6 cells/mL) as described (section 2.2.4.7). The obtained stroma was concentrated in 3 kDa cutoff Amicon Ultra filtration devices (Millipore) at 4^{o}C, with or without 250 units of RNase One (Promega). Samples (5 mg protein) were loaded through an SW guard column onto a 2.15 x 30-cm G4000SW column (Tosoh), and elution was performed at 4^{o}C with SEC buffer (50 mM KCl, 5 mM $MgCl_2$, 5 mM ε-aminocaproic acid and 20 mM Tricine-KOH, pH 7.5). The flow rate of elution was kept at 2 mL/min. The elution fractions obtained were further concentrated to about 70 µL using 10 kDa cutoff Amicon Ultra devices (Millipore) and were subjected to immunoblotting (section 2.2.4.3).

For SEC of recombinant PPR7, the expressed and purified protein (section 2.2.4.10.2) was loaded onto a Superdex 75 10/300 GL column (GE Healthcare), and elution was performed at 4^{o}C with Tris buffer (10 mM Tris-HCl, and 150 mM NaCl, pH 8.0). The flow rate of elution was kept at 0.5 mL/min.

2.2.4.10 Expression and purification of recombinant proteins

2.2.4.10.1 Plasmids for expression of recombinant proteins

GST based fusion protein system (section 2.1.4) was used for raising polyclonal antibodies against PPR1, PPR3, PPR4 and PPR7. Suitable primer sites (Table 2.7) were selected from the C- terminus of genomic sequence of the *C. reinhardtii.* The amplification of the PCR fragments was achieved by using standard PCR (section 2.2.3.5.2). The amplified fragments were then inserted into the expression vector pGEX4T-1 (via the *Bam*HI and *Xho*I restriction sites present in the two primers) ligating selected fragments with an N-terminal GST-tag.

For the expression of full length recombinant PPR7 protein (His-PPR7 or GST-PPR7), the DNA sequences encoding amino acids 61-221 or 59-194 were PCR-amplified from a cDNA clone using the primers mentioned in Table 2.7 and were inserted into the plasmid pQE30

(Qiagen) or pGEX4T-1 (via *Bam*HI/*Sal*I restriction sites present in the two primers) ligating selected fragments with an N-terminal His-tag or GST-tag.

Table 2.7: List of oligonucleotides used for cloning of protein expression vectors. Lower case written nucleotides represent generated restriction sites

Name	Sequence 5' – 3'
PPR1 Fw	aaggatccGCCCGTCGCCCGTCGGTGG
PPR1 Rv	aactcgagCTACTTCCACTCTGATTCGTAATA
PPR3 Fw	aaggatccCGCACCGTCATCACCTACAG
PPR3 Rv	aactcgagCTACTGCCGCTGCAGCTTTGAGA
PPR4 Fw	aaggatccTACACGGCGCTCATCAGCGC
PPR4 Rv	aactcgagACTTGATCTCCAGCAGCCAGC
PPR7 Fw	aaggatccCTGTCCTTCAACGCGCTGCT
PPR7 Rv	aactcgagTCACAGCGTCTTGCGCA
His-PPR7 Fw2	aaggatccGAGGTCACGAAGCGGATAAAG
His-PPR7 Rv2	aactcgagTCACAGCAGCGTCTTG
His-PPR7 Fw3	aaggatcc CGCGCGGAGGTCACGAAG
His-PPR7 Rv3	aactcgag GTCAGCGGCCATGCGGTC

2.2.4.10.2 Expression and purification of recombinant proteins

GST based constructs described in 2.2.4.10.1 were used to transform the *E. coli* strain BL21 while, the His-tag contruct described in 2.2.4.10.1 was used to transform the *E. coli* strain M15, which are suitable for the overexpression of recombinant proteins.

Overexpression of recombinant proteins in bacteria by using the pGEX4T-1 system (Amersham Biosciences) or pQE30 system (Qiagen) was carried out by inoculating 1 L LB medium with 1/25 volume of the respective overnight culture. Cultures were grown under selective pressure at 37°C on a horizontal shaker with 220 rpm. At an OD_{600nm} of 0.6, the expression of the recombinant protein was induced with 0.8 mM IPTG. Expression was allowed for 4 h at 27°C and 220 rpm, before the cells were harvested by centrifugation for 10 min with 6,000 x g at 4°C.

Purification of the GST based fusion proteins was performed according to the manufacturer's protocol using Glutathione–Sepharose 4B (GE Healthcare). When required, the GST-tag was cleaved off according to manufacturer's protocol (GE Healthcare) using a site specific protease thrombin.

The His-tagged recombinant protein was purified according to the GE Healthcare protocol under native conditions using Ni-Sepharose 6 Fast Flow (GE Healthcare) and dialyzed against 30 mM ammonium hydrogen carbonate (NH_4HCO_3) buffer.

2.2.4.11 Antibody production and purification

Polyclonal antiserums were produced by immunizing rabbits with the overexpressed protein fractions (section 2.2.4.10.2) at Biogenes, Berlin.
Crude antiserum was purified by using 2 mg of antigen blotted onto nitrocellulose membrane (section 2.2.4.3). After staining with Ponceau S, the portion of blot containing the antigen was cut and incubated with 2 mL antiserum over night at 4^{o}C. The membrane was washed 2 times for 5 min with TBS buffer and the antibody was eluted with glycine buffer (0.1 M Glycine-HCl, pH 2.7). The low pH of glycine buffer was rapidly neutralized by adding 1/10 volume of 1 M Tris-HCl pH 8. To stabilize the antibody, sodium azide and BSA were added to a final concentration of 5 mM and 1 mg/mL, respectively. For long term storage, the aliquots were stored at -80^{o}C.

2.2.5 Nuclear transformation of *C. reinhardtii*

Nuclear transformation of cell-wall-deficient *C. reinhardtii* cells was carried out using the glass bead method as described previously (Kindle, 1990). Briefly, cells of *C. reinhardtii* were grown in 100 mL TAPS medium to a cell density of ~1-2 x 10^6 cells/mL. Cells were harvested by centrifugation (1100 x g, 6 min, RT) and resuspended in TAPS medium to a density of 1 x 10^8 cells/mL. 0.3 g of glass beads (0.4-0.6 mm diameter; Sigma), 300 µL cells and 7 µg plasmid DNA were mixed in 2 mL microreaction tube and vortexed at highest speed for 20 sec. After the transformation, cells were transferred to 5 ml TAPS medium in 15 mL tubes and incubated overnight with constant shaking at 120 rpm, 23^{o}C (regeneration time in case of antibiotic selection). Cells were then harvested by centrifugation (1100 g, 6 min, RT), resuspended in 300 µL TAPS medium and plated onto selection TAP agar plates. Plates were kept under dim light or dark, and scored for transformants about 2 weeks after plating.

2.2.6 GFP based subcellular localization

2.2.6.1 GFP fusion constructs

For GFP based subcellular localization of PPR proteins in *C. reinhardtii*, a synthetic gene encoding GFP adapted to the codon usage of the green alga CrGFP (Fuhrmann et al., 1999) was used in experiments.
Transit peptides (TPs) of PPR proteins PPR1, PPR3, PPR4 and PPR7 were identified by using the Target P database (http://www.cbs.dtu.dk/services/TargetP). The DNA fragments coding for transit peptides were amplified by standard PCR reactions (section 2.2.3.5.2) us-

ing primer sets selected at N-terminus of genes (Table 2.8). The amplified nucleotide sequence encoding the respective transit peptide of each PPR protein was then fused with the GFP protein coding sequence in the vector pBC1-CrGFP (pJR38, Neupert et al., 2009) in frame via an N-terminal *Nde*I site. These constructs were then transformed into *UVM4* and positive transformants were selected on TAP plates supplemented with 10 µg/mL paromomycin. As a control, the pBC1-CrGFP vector was directly transformed into *UVM4* for cytosolic GFP expression.

Table 2.8: Primers used for generation of GFP fusion proteins. Lower case written nucleotides represent generated restriction sites

Name	Sequence 5' – 3'	Experiment
PPR1 TP Fw	aacatatgACTGTGAAATACTTAGGCTCC	TP fragment of PPR1
PPR1 TP Rv	aacatatgCGTAATGCCACGAGTCCC	
PPR3 TP Fw	aacatatgGAAGTCCTCTCCGGTC	TP fragment of PPR3
PPR3 TP Rv	aacatatgGATGCTGTCGGTTCTGGG	
PPR4 TP Fw	aacatatgGCAGTTATTATTCAGAAGC	TP fragment of PPR4
PPR4 TP Rv	aacatatgGAGGTCCTTTAGCAGCAG	
PPR7 TP Fw	aacatatgCAGGCAATTCAACGGC	TP fragment of PPR7
PPR7 TP Rv	aacatatgACCCAGTGCCTTTATCCGC	

2.2.6.2 GFP fluorescence microscopy

To observe the GFP fluorescence in transformed *C. reinhardtii* cells, laser scanning confocal fluorescence microscopy (LSCFM) was used. 40 µL of liquid culture (1 x 10^6 cells/mL) was pipetted on an object holder and sealed with a cover slide. The samples were observed and confocal images were acquired with a TCS-SP5 confocal laser scanning system equipped with an inverted microscope (Leica Mikrosysteme Vertrieb GmbH, Wetzlar, Germany) and an X63 glycerin immersion objective (numerical aperture of 1.30). The 488 nm argon laser line was used for activation of GFP, with the emission window set at 496-505 nm. The laser intensity was set at 10% - 30% of the available power.

2.2.7 Generation of RNAi lines for PPR7

For generation of RNA interference lines (RNAi) of the corresponding PPR proteins, design of suitable constructs using NE537 vector was carried out as described (Rohr et al., 2004). Suitable primers were designed (Table 2.9) from the genomic sequence of *C. reinhardtii* adding *Sal*I, *EcoR*I and *BamH*I restriction sites to the sequence. Fragments of 400 and 600 bps

were amplified by PCR (section 2.2.3.5.2) and cloned as inverted repeats into the *EcoR*I site of NE537 vector (which is located in the inverted repeat of the *Maa7* gene for the ß-subunit of tryptophan synthase). The longer 600 bps fragment contains additional 200 bps which functions as a spacer in inverted repeats. The *UVM4* strain was transformed with these constructs (section 2.2.5), kept for 2 days in liquid culture (TAP + 1.5 mM L-Tryptophan) in dim light, and then plated on TAP plates containing 5 µg/mL paromomycin and 1.5 mM L-tryptophan. *UVM4* cells transformed with the empty NE537 vector served as a negative control and taken as wild type in further experiments. At intervals of 2 weeks, colonies were transferred to TAP plates containing 1, 2, 5 and then 10 µM of 5-fluoroindole in the dark.

Table 2.9: Primers used for generation of RNAi lines. Lower case written nucleotides represent generated restriction sites.

Name	Sequence 5' – 3'	Experiment
iPPR1 400a	aagtcgacgaattcGACGTGGTGGAGCTGC	RNAi construct PPR1
iPPR1 400b	aaggatccACCGCCTCACTCACCCAGC	
iPPR1 600c	aaggatccCACAGCAGCGTGTTGTAGGC	
iPPR7 400a	aagtcgacgaattcATGCAGGCAATTCAAC	RNAi construct PPR7
iPPR7 400b	aaggatccCCAGGTTAGCCAGGCCCGC	
iPPR7 600c	aaggatccTTGAGGGCATGGCCCGGCA	

2.2.8 Co-immunoprecipitation studies

For protein and RNA co-immunoprecipitation, stromal extracts were obtained from *C. reinhardtii* cultures grown to the early log phase (~1 x 10^6–2 x 10^6 cells/mL), as described (section 2.2.4.7). Co-immunoprecipitations of RNA were performed in the presence of 0.5 mg/mL yeast tRNA and 0.25 U/µL RNase inhibitor (Fermentas). Purified PPR7 antibody (100 µL/mL of stroma) was incubated with extracted stroma for 45 min on rotator (Neolab, Heidelberg, Germany) at 12 rpm and at 4^{o}C, followed by 30 min incubation with 50 µL Dyna beads-coupled protein G (Invitrogen), already equilibrated with co-immunoprecipitation buffer (150 mM NaCl, 20 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ and 0.5% (v/v) Nonidet P-40). An aliquot of 1/10 volume of supernatant stroma was taken for Western blot analysis (section 2.2.4.3) and the rest was used for RNA extraction. The pellet was washed 5 times with co-immunoprecipitation +I buffer (co-immunoprecipitation buffer + 5 µL/mL aprotinin and 1µL/mL RNase Inhibitor) and finally resuspended in 180 µL of co-immunoprecipitation + I buffer. RNA was extracted by phenol/chloroform from supernatant and pellet after the addition of SDS to 0.5%, and the RNA used either for chip analyses as described (Schmitz-Linneweber et al., 2005) or for cDNA synthesis (section 2.2.3.4).

2.2.9 Microarray design and hybridization

Labelling of PPR7-copurified RNA was carried out as described previously (Schmitz-Linneweber et al., 2005) using the Kreatech ULS labelling kit (Kreatech, Amsterdam, Netherland). A microarray was designed with 166 overlapping PCR fragments representing the complete *C. reinhardtii* chloroplast genome and 15 overlapping PCR fragments representing the complete mitochondrial genome. Total DNA extracted from *C. reinhardtii* cells served as template for PCR reactions. Fragment positions and primers used for amplification are mentioned in Annex B. Fragments were numbered (ID) according to their position on the *C. reinhardtii* chloroplast and mitochondrial genomes sequence (Acc. No. BK_000554 for chloroplast and U03843 for mitochondrial probes, see Annex B, Table 1). Each fragment was spotted in multiple copies using an Omnigrid Accent spotting device (GeneMachines, USA). For each microarray slide, PCR products were spotted in two areas. In area A, 12 replicates per PCR product were spotted, whereas in area B, only 6 replicates were spotted. Area A was hybridized with RNA from immunoprecipitations using anti-PPR7 antibody, whereas area B was hybridized with RNA from parallel immunoprecipitations using pre-immune serum as control. This parallel processing of experiment and control allowed maximum comparability during hybridization and subsequent washings. RNA labelling, hybridization of RNA on the *C. reinhardtii* chloroplast microarray, and data analysis were carried out as reported previously using a Scanarray Gx microarray scanner (Perkin Elmer, USA) and the Genepix Pro 6.0 analysis software (Axon, USA; Schmitz-Linneweber et al., 2005).

2.2.10 UV cross-linking of RNA and recombinant PPR7

The recombinant His-tagged PPR7 protein (His-PPR7) was expressed and purified as described in section 2.2.4.10. The DNA templates for *in vitro* synthesis of *rbcL*, *atpA* and *psbD* RNA probes were generated by PCR using primers mentioned in Table 2.10. RNA synthesis was catalysed by T7 RNA polymerase (Fermentas) in the presence of [α-^{32}P] UTP (3000 Ci/mmol; Hartmann Analytic), according to the manufacturer's protocol. After removal of the template by treatment with DNase I (Promega), the RNA was extracted with phenol-chloroform and precipitated with ethanol in the presence of ammonium. Binding reactions (20 µL) were performed at RT for 5 min. Each reaction contained 50-100 kilo counts per min of ^{32}P-labeled RNA probe, 50 mM NH_4HCO_3 and 3 µg of His-PPR7 protein. After irradiation, the free RNA probes were digested by treatment with 10 U RNase I (Promega) for 30 min at 37°C, and the samples were fractionated by SDS-PAGE (section 2.2.4.2), and analysed by phosphor imaging system (Fuji Photo Film, Kanagawa, Japan).

Table 2.10: Primers for *in vitro* transcription of DNA sequences used for UV cross-linking

Name	Sequence 5' – 3'
T7 rbcL5	gtaatacgactcactatagggTATGCTCGACTGATAAGAC
rbcL3	CTGCTTTAGTTTCTGTTTGTGGAACC
T7atpA 5'UTR Fw	gtaatacgactcactatagggGCCACTGTTCACTCCTC
112 atpA Rv	TCTGGAGTACGCATTGCC
T7 psbD5	gtaatacgactcactatagggCCACAATGATTAAAATTAAA
psbDUTR3	ACCGATCGCAATTGTCAT

2.2.11 Chlorophyll fluorescence measurements

The maximum quantum yield of PSII photochemistry (Qy_max) was identified using a Fluorocam (FluroCam 800MF, Photon Systems Instruments, Czech Republic). Three colonies of each genotype were analysed, average values and standard deviations were calculated. *C. reinhardtii* strains were dark adapted for 30 min and minimal fluorescence (F0) was measured. Then, pulses (0.8 s) of white light (5000 µmol photons/m^2/s^1) were used to determine the maximum fluorescence (Fm) and the ratio (Fm - F0)/Fm = Fv/Fm (maximum quantum yield of PSII) was calculated.

2.2.12 Circular dichroism measurements

His-PPR7 was expressed and purified as described in 2.2.4.10. The protein was dialyzed against circular dichroism (CD) buffer (30 mM NH_4HCO_3 at pH 7.5). His-PPR7 protein concentration for CD spectroscopy was determined as described (section 2.2.4.1). CD spectra were measured at room temperature using a Jasco J-810 spectropolarimeter as described (Kelly et al., 2005). Each spectrum was the average of three scans taken at a scan rate of 50 nm/min with a spectral bandwidth of 1 nm. After subtraction of baseline spectra, data were converted to [θ] and displayed against wavelength using JASCO Spectra Manager software. CD spectra were measured from 260 to 185 nm, using 1-mm path length of a cylindrical quartz cell at a protein concentration of 0.13 mg/mL. These measurements provided the absorbance values from 185 to 260 nm. Secondary structure contributions to the CD spectra were deconvoluted using the CDSSTR programs at DICHROWEB (Whitmore and Wallace, 2004; http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) .

2.2.13 Crystallization of His-PPR7

His-PPR7 (full length of PPR7; amino acids 61–221 or 59-194, respectively) was purified via Ni-affinity chromatography as described (2.2.4.10.2). The recombinant protein was further purified by SEC as described (section 2.2.4.9). The elution fractions obtained were further concentrated to ~20 mg/mL using Amicon Ultra devices (Millipore) and were sent to the group of Prof. Dr. Michael Groll (Department of Chemistry, Technical University Munich, Garching, Germany) for the crystallization procedure.

2.2.14 Bioinformatics sources

All software tools mentioned below were applied with default parameters unless stated otherwise.

2.2.14.1 Prediction of gene models

For gene model analysis and the obtainment of EST and genome sequences, the *C. reinhardtii* Genome Browser from the U.S. Department of Energy Joint Genome Institute (DOI JGI, (http://genome.jgi-psf.org/Chlre4/Chlre4.home.html), the University of California Los Angeles (UCLA) browser for *Chlamydomonas* 454 EST reads (http://genomes.mcdb.ucla.edu/Cre454/) and the browser of the National Center for Biotechnology Information employing the blastp and tblastn algorithms (NCBI, http://www.ncbi.nlm.nih.gov/BLAST) were used. The applied AUGUSTUS gene models (version 10.2) and gene identifiers were generated on evidence-based predictions using the v4 Chlamy genome assembly and the program AUGUSTUS (Stanke et al., 2004).

2.2.14.2 Prediction of protein localization and transit peptides

The prediction of protein localization and transit peptides was primarily based on Target-P version 1.1 (Emanuelsson et al., 2000). The Predotar (Small et al., 2004) and ChloroP (Emanuelsson et al., 1999) prediction programs were used in addition to match the prediction results.

2.2.14.3 Protein properties and repeat predictions

Basic protein property predictions, like theoretical molecular weight calculations, were done using the ProtParam tool on the ExPASy server (Gasteiger et al., 2003). The PPR repeats were predicted using the toolkit TPR-pred of the Max-Planck-Institute (Biegert et al., 2006; http://toolkit.tuebingen.mpg.de/tprpred). In addition the Prosite tool (http://prosite.expasy.org/) was used to predict known protein domains and the PPR repeats for the comparison.

2.2.14.4 Alpha helical structure and wheel model predictions

For the prediction of residues forming α helices within a protein, the secondary structure prediction server Jpred3 was used (Cole et al., 2008; http://www.compbio.dundee.ac.uk/www-jpred/). The helical wheel predictions for helix A of PPR repeats were performed using a program created by D. Armstrong and R. Zidovetzki (http://rzlab.ucr.edu/scripts/wheel/wheel.cgi) while the shape and colour code for the figures representing the amino acids was edited manually for more simple representation.

3 RESULTS

3.1 PPR proteins in *C. reinhardtii*

The bioinformatical analysis of the *C. reinhardtii* genome revealed 11 members of the PPR protein family. A schematic diagram of PPR proteins decuced from EST-based Augustus 10.2 gene models (section 2.2.14.1) is given in Figure 3.1 to present a picture of the number of PPR repeats and varying number of amino acid residues.

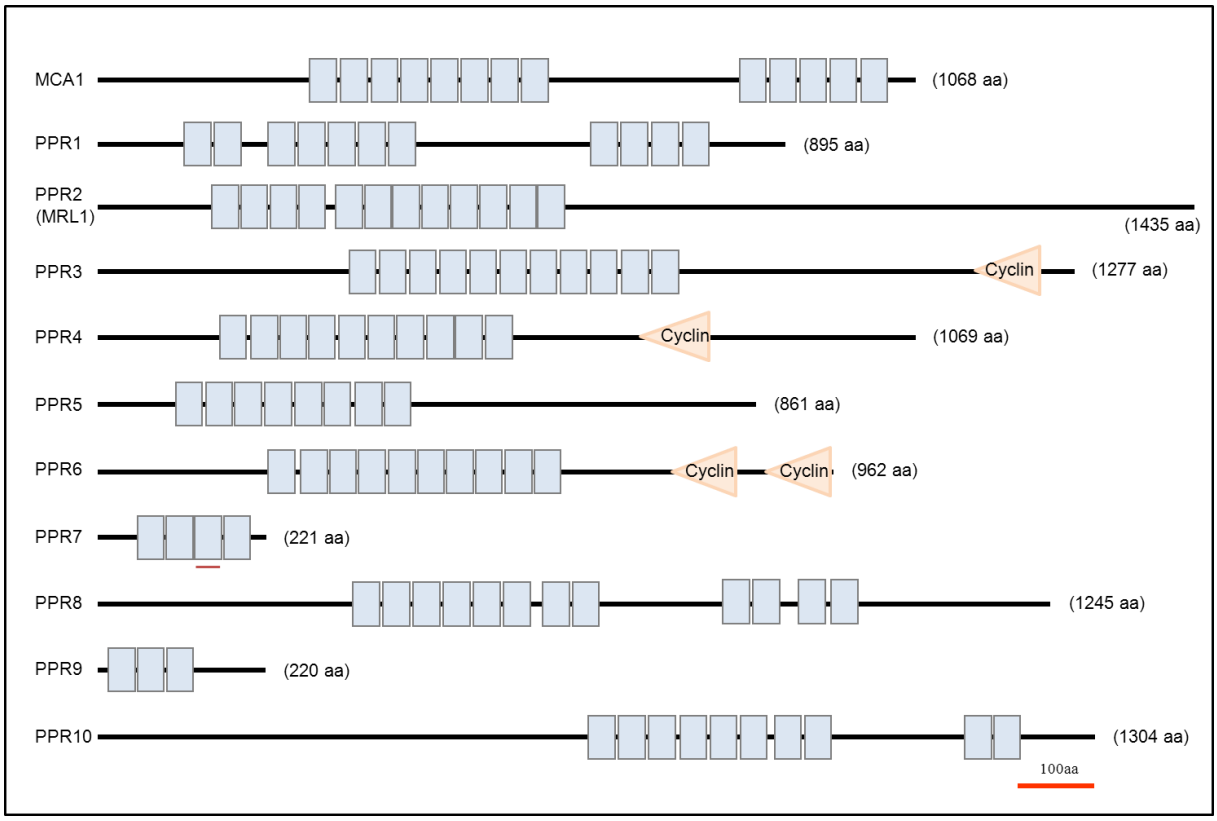


Figure 3.1: Schematic diagram of PPR proteins in *C. reinhardtii.* PPR protein sequences are represented by black lines and the blocks represent PPR repeats predicted by TPR pred (http://toolkit.tuebingen.mpg.de/tprpred) and Prosite (http://prosite.expasy.org/). The triangles represent Cyclin domains found in some of the PPR proteins. The scale bar is given as red line at the bottom. Names for PPR proteins are given on the left and total numbers of amino acids (aa) on the right hand side. The third repeat of PPR7 (indicated with red line below it) was not predicted to form α helices by Jpred prediction tool. For further details see text (section 3.3).

All the PPR proteins in *C. reinhardtii* belong to the canonical P subfamily containing direct repeats of ~35 amino acid units. They lack the C-terminal E/E+ and DYW domains which are specific to land plants (Lurin et al., 2004; see section 1.4.3.1). Interestingly, three PPR proteins (PPR3, PPR4 and PPR6) additionally possess Cyclin domains at the C-terminus. Cyclin domains are present in proteins named cyclins which are known to function in cell cycle and transcription control by regulating cyclin dependent kinases (CDKs). The decuced amino acid

sequences of the identified PPR proteins range from 220 to 1435 amino acids and according to prediction programs for PPR repeats (TPRpred and Prosite), they contain between 3 to 13 mostly tandemly arranged repeats (Figure 3.1; Table 3.1).

Table 3.1: List of *C. reinhardtii* PPR proteins and their characteristics. The assigned number/name of each PPR protein, identification number (Gene ID), number of repeats found per PPR protein, number of amino acid residues (aa) and the predicted molecular weight (MW) are shown.

PPR Protein	Gene ID[a]	No. of repeats	Length (aa)[b]	MW (kDa)[c]
MCA1	Cre08.g358250	13	1068 (1011)	109 (103.4)
PPR1	Cre01.g050500	13	895 (791)	93.2 (82.84)
PPR2 (MRL1)	Cre06.g298300	12	1435 (1416)	149.6 (147.6)
PPR3	Cre16.g649800	11	1277 (1209)	128.6 (121.7)
PPR4	Cre12.g511400	10	1069 (1042)	115.7 (112.8)
PPR5	Cre23.g766200	8	861 (765)	86.5 (76.27)
PPR6	Cre10.g437150	10	962 (-)	105.2 (-)
PPR7	Cre01.g048750	4	221 (161)	24.34 (18.8)
PPR8	Cre01.g025950	12	1245 (-)	307.3 (-)
PPR9	Cre01.g026200	3	220 (163)	23.16 (17.12)
PPR10	Cre27.g775550	10	1304 (1293)	134.4 (133.1)

[a] AUGUSTUS 10.2 gene model (section 2.2.14.1)
[b] Number of amino acids after cleavage of predicted transit peptide by TargetP (section 2.2.14.2) is given in brackets. A minus (-) indicates that no transit peptide was predicted.
[c] Predicted molecular weight after the cleavage of predicted transit peptide is given in brackets (section 2.2.14.3).

The amino acid sequence of the repeats predicted by the prediction programs are depicted in Annex A. On the basis of sequenced mRNA data obtained from the UCLA browser for *Chlamydomonas* 454 EST reads all of the identified PPR proteins in *C. reinhardtii* can be considered as expressed and seem not to represent pseudo genes. Based on the abundance of

available EST data one can assume that the PPR encoding mRNAs are expressed at low to medium levels. The identification numbers (Gene IDs) of the protein encoding genes models on the *Chlamydomonas* UCLA browser for 454 EST reads (section 2.2.14.1., http://genomes.mcdb.ucla.edu/Cre454/) and protein parameters are summarised in Table 3.1.

3.2 Subcellular localization of PPR proteins in *C. reinhardtii*

The previously characterized PPR proteins from eukaryotes are almost exclusively targeted to organelles (Lurin et al., 2004). To achieve first indications about the putative subcellular localization of PPR proteins in *C. reinhardtii*, different prediction programs (TargetP, Predotar and ChloroP, section 2.2.14.2) were used. The data suggest that five out of eleven *C. reinhardtii* PPR proteins are predicted to have plastidial targeting, while four are predicted to be targeted to mitochondria. No targeting prediction was available for PPR6 and PPR8 by the programs used (Table 3.2). To exclude alternative start codons further upstream of the predicted one, a detailed analysis was performed on the Augustus 10.2 gene models for these PPR proteins. However, no possible other translation initiation sites could be found. Nevertheless, an organellar targeting of these PPR proteins can not be excluded since the used targeting prediction programs are not optimized for *Chlamydomonas*.

Table 3.2: List of *C. reinhardtii* PPR proteins and their targeting predictions. The transit peptide (TP) length predicted by TargetP and subcellular localization of each PPR using “TargetP”, “Predotar”, and ‘‘ChloroP’’ programs is shown.

PPR	TP (aa)	Localization		
		TargetP	Predotar	ChloroP
MCA1	57	Cp	Cp	Cp
PPR1	104	Mt	Mt	Cp
PPR2 (MRL1)	19	Mt	Mt	Cp
PPR3	68	Mt	None	None
PPR4	27	Mt	Mt	Mt
PPR5	96	Mt	None	None
PPR6	-	None	None	None
PPR7	60	Mt	Mt	Cp
PPR8	-	None	None	None
PPR9	57	Mt	Mt	Mt
PPR10	11	Mt	Mt	Cp

To confirm the predicted organellar localization of *C.reinhardtii* PPR proteins, a GFP-based *in vivo* approach was chosen. Fusion of GFP to the transit peptides (TPs) of other proteins is a powerful method to investigate protein localization and dynamic processes *in vivo* by using confocal laser microscopy. Four out of eleven PPR proteins, i.e. PPR1, PPR3, PPR4 and PPR7, were selected for this analysis based on their clear annotations (AUGUSTUS 10.2 gene model, section 2.2.14.1). Due to the low expression efficiency of heterologous genes in *C. reinhardtii*, a synthetic gene encoding GFP adapted to the codon usage of the green alga was used in the experiments (CrGFP, Fuhrmann et al., 1999). The DNA sequences coding for predicted transit peptides were fused with the GFP protein coding sequence under control of the strongly expressed *psaD* promoter in the vector pBC1-CrGFP (pJR38, Neupert et al., 2009). Resulting constructs were transformed into the *UVM4* expression strain (Neupert et al. 2009) and steady state levels of GFP fusion proteins in the transformed clones were analysed.

Whereas no detectable expression was seen for the strains containing the transit peptide fusion constructs for PPR1, PPR3, and PPR4 (data not shown), successful GFP expression was observed for the construct pBC1-TP-PPR7-CrGFP (Figure 3.2 A and B). 50% of the analysed clones showed GFP expression at different levels.To investigate the localization of GFP *in vivo,* the TP-PPR7-CrGFP clone C4 which showed high GFP accumulation was further subjected to confocal laser microscopy. As PPR proteins are thought to be either targeted to mitochondria or to chloroplasts, the mitochondria of the cells were stained with Mito tracker red (CMX Ros Molecular Probes Invitrogen). As shown in Figure 3.2 C, confocal images of clone C4 revealed, the GFP signal to be clearly co-localized with auto fluorescence of the chloroplast, while no GFP fluorescence signal was observed in the mitochondrial regions (Figure 3.2 C, lower panel). These observations strongly indicate PPR7 to be a chloroplast localized protein. Wild type cells were used in parallel as control in GFP microscopy (Figure 3.2 C, upper panel).

To further confirm the finding, that PPR7 is targeted to chloroplast and get indications of the sub-plastidial localization, immunoblot analyses were performed on whole cell, chloroplast, stromal and crude thylakoid sub-fractions of intact chloroplasts and mitochondrial fractions from TP-PPR7-CrGFP expressing as well as from wild type cells (Figure 3.2 D, left panel). As shown in Figure 3.2 D, the GFP signals were detected in total cell extracts (WC), whole chloroplast (cp), and enriched in the stromal fraction (S) of TP-PPR7 CrGFP cell fractions. No GFP signal was detected in the mitochondrial fraction, thereby confirming the results of the confocal microscopy. HSP70B and AOX antibodies were used as chloroplast and mitochondrial markers, respectively.

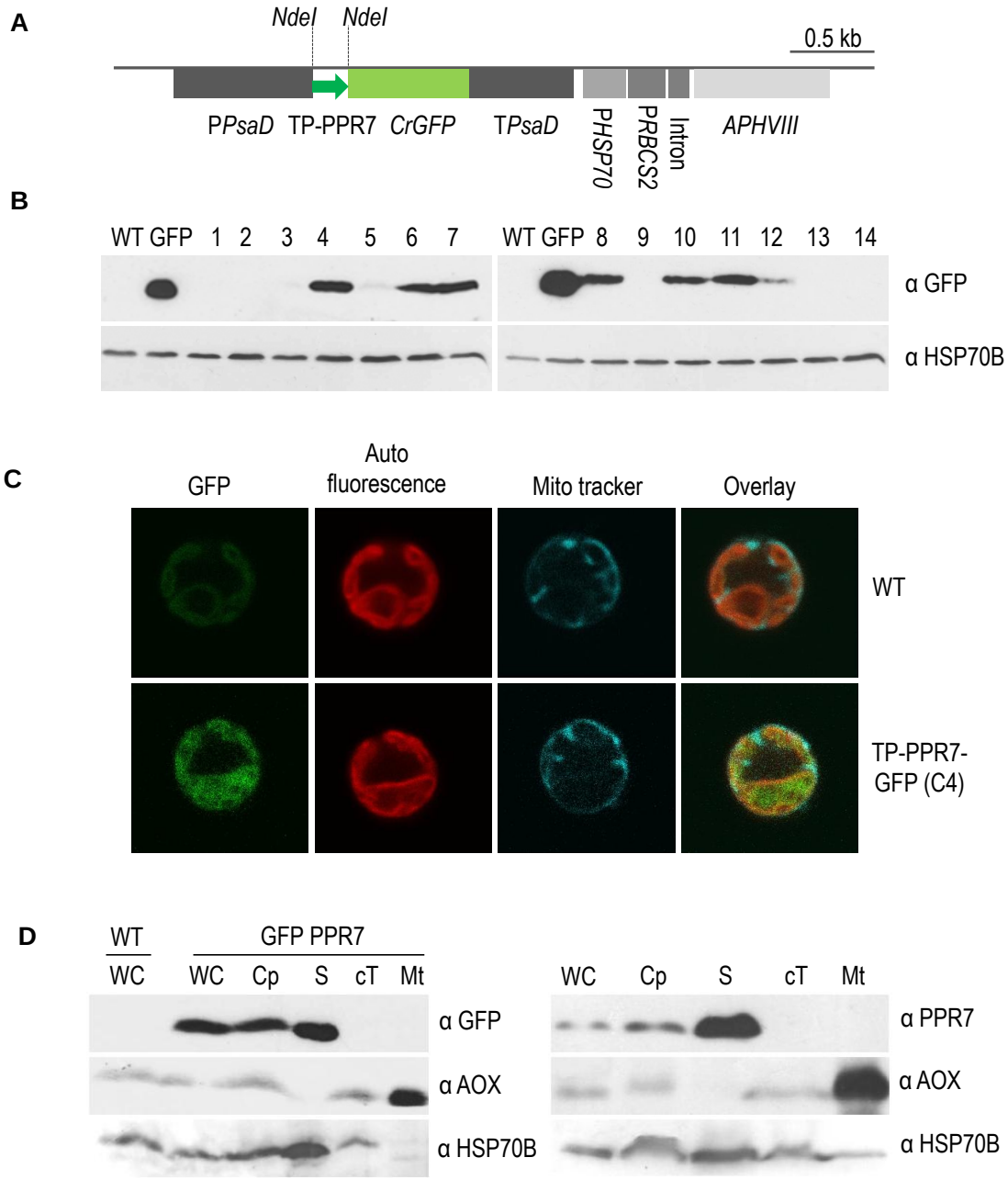


Figure 3.2: Subcellular localization of PPR7. A: Schematic representation of GFP fusion construct used in subcellular localization assay. The TP-PPR7–CrGFP construct contains the putative chloroplast transit peptide (TP) fused to codon adapted GFP via *NdeI* site. **B: TP-PPR7-CrGFP immunoblots after incubation with a GFP specific antibody (Miltenyi Biotec).** 50 µg of whole cell proteins of TP-PPR7-CrGFP transformants (1-14) were analysed via immunoblot using a GFP specific antibody. The non-transformed *UVM4* strain (WT) served as negative control, and a CrGFP plasmid transformed clone (GFP) as a positive control. HSP70 was used as a loading control (lower panel). **C: Fluorescence microscopy of TP-PPR7-CrGFP transformed cell.** Laser scanning confocal fluorescence microscopy (LSCFM) images of *UVM4* (Wild type) and TP-PPR7-CrGFP expressing *UVM4* cell line (C4). GFP fluorescence (GFP), the auto fluorescence (Auto fluorescence) of *C. reinhardtii* chloroplast, the mitochondrial staining (Mito tracker) and overlays of the three can be seen. **D: Immunoblot analysis of cell fractions of TP-PPR7-CrGFP and wild type cells.** 50 µg of whole cell proteins (WC), chloroplast (Cp), stroma (S), crude thylakoids (cT), and mitochondrial proteins (Mt) of TP-PPR7-CrGFP (left panel) or wild type cells (right panel) were analysed via immunoblot using antibodies stated at the right of each panel. Whole cell proteins of wild type (WT) cells served as negative control in TP-PPR7-CrGFP fractions blot.

To validate these data and for further biochemical analyses a polyclonal antibody against PPR7 was raised (section 2.2.4.11). In a comparable fractionation of *Chlamydomonas* wild type cells a protein of ~18.8 kDa corresponding to the molecular weight of mature PPR7 was detected (Figure 3.2 D, right panel, compare Table 3.1). Hereby, the same distribution and enrichment was observed for native PPR7 protein in cell fractions, confirming that PPR7 is a soluble chloroplast localized protein. The pattern of PPR7 is similar to chloroplast proteins such as HSP70B, while it was undetectable in the crude thylakoid (cT) and mitochondrial fractions (Mt).

Since no GFP data could be obtained for PPR1, PPR3, PPR4, and based on the finding that PPR7 is a soluble chloroplast localized protein, this thesis focusses on in-depth characterization of the PPR7 protein encoded by the gene locus Cre01.g048750.

3.3 Structure analysis of the PPR7 protein

3.3.1 PPR motifs in PPR7 and their helical wheel models

A typical PPR protein consists of tandem arrays of PPR motifs forming a superhelix enclosing a groove, which is likely to be an RNA-binding site (Small and Peeters, 2000; Delannoy et al., 2007). One PPR motif forms two helices, i.e. helix A and B (Figure 3.3 A). Helix B in the PPR motif is on the outside of the superhelix and is thought to mainly contribute to the structure, whereas helix A is towards the interior, and thus most of the amino acid side chains projecting into the putative RNA-binding site are from helix A. The conserved residues in a PPR motif are mostly hydrophobic and are positioned on one side of the helix, probably for the helix formation, while the putative RNA binding charged residues come on the contrasting side of the helix (Small and Peeters, 2000; Kobayashi et al., 2011). To investigate for the helical structure and positioning of the amino acids in helix A of PPR7 repeats, first, the alignments of the four PPR motifs predicted by TPRpred and Prosite and the consensus sequence for PPR repeats generated by the pfam database were created (Figure 3.3 A). To identify putative helix forming residues the aligned residues of each repeat were subjected to the secondary structure prediction tool Jpred (Figure 3.3 B). It was observed that repeat 1, 2 and 4 are predicted to form two α helices but not the 3rd repeat of PPR7. Also this repeat was not predicted by all programs used for the prediction of PPR repeats, indicating it as a non-PPR repeat. Hence, the residues 1-12 of helix A from each repeat except for repeat 3 were plotted on a helical wheel model using the rzlab helical wheel projection program (Figure 3.3 C, section 2.2.13.4). For all the three repeats (1, 2 and 4) analysed, the hydrophobic residues were found to be positioned mostly on one side, while the hydrophilic and charged amino acids were found on the other side (Figure 3.3 C). These observations suggest that

the helix A of PPR7 repeats can provide the putative RNA-binding surfaces by having the charged arginine and lysine at one side of the helix.

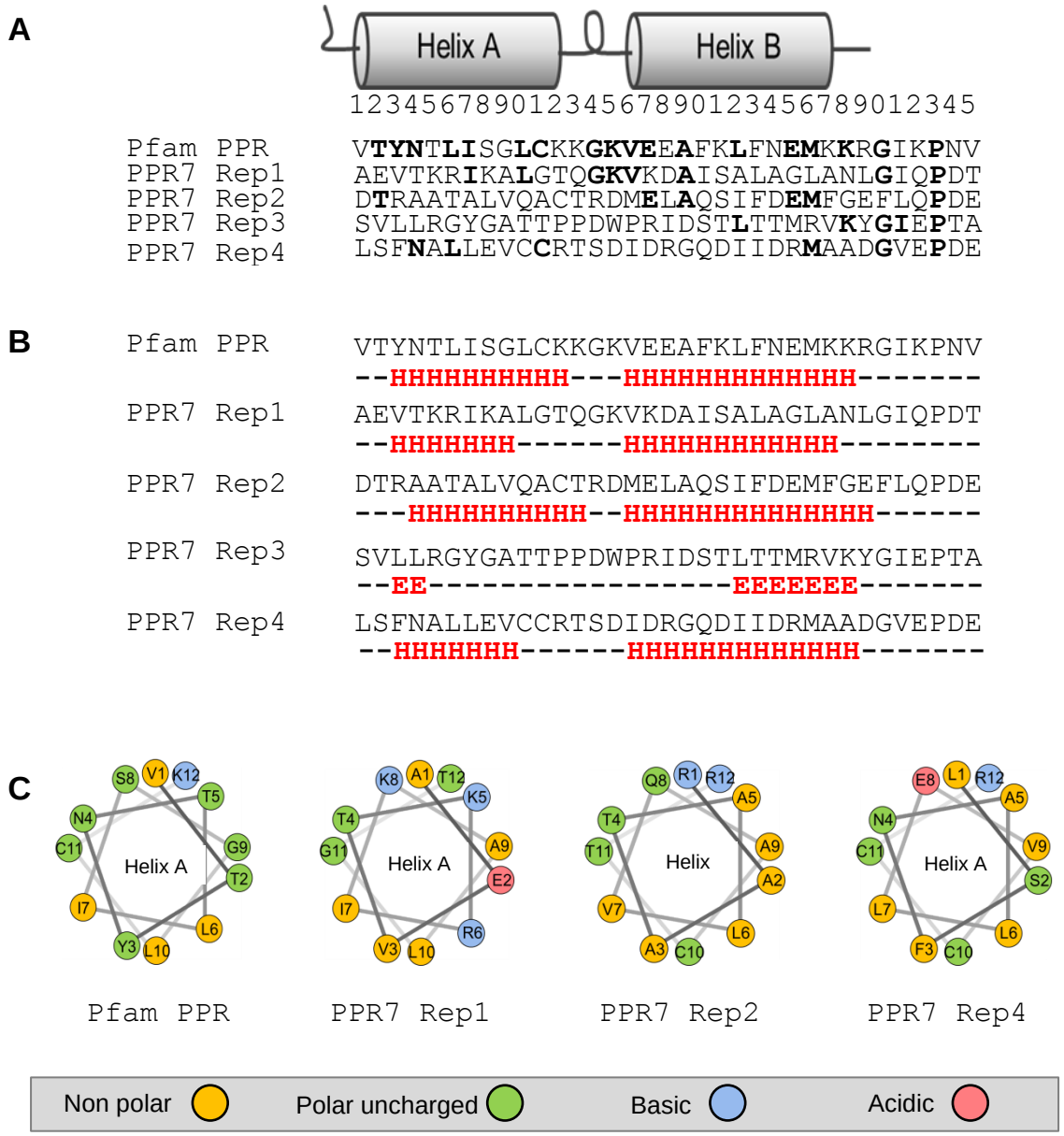
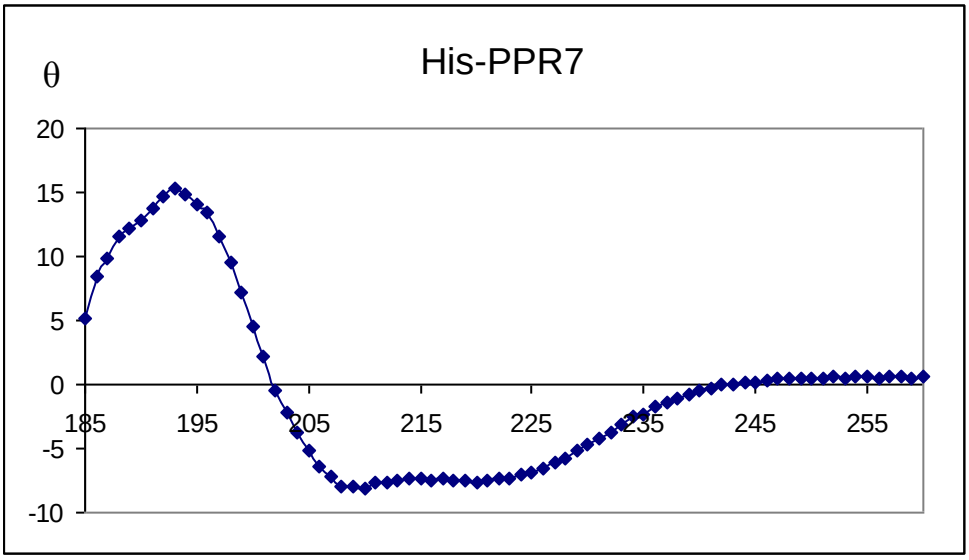


Figure 3.3: Primary and secondary structural features of PPR motifs found in PPR7. A: The predicted PPR repeats of PPR7 were aligned with the consensus sequence of the PPR consensus sequence generated by Pfam (pfam.sanger.ac.uk). Above the alignment helix A, B and the loop structure of a PPR motif are schematically represented. Identical residues in the Pfam motif and the PPR7 repeat are shown in bold. **B:** The residues predicted by Jpred3 (www.compbio.dundee.ac.uk/www-jpred) to form helices are shown by underneath red 'H' letter. **C:** Helical wheel models and the positions of amino acids in helix A of consensus sequence and PPR repeats in PPR7 are depicted with the colour codes given below. The colour code is manually added to the predicted wheels for more simple representation of the amino acid properties.

3.3.2 Circular dichroism measurements of recombinant PPR7 protein

To analyse the secondary structure and physical organization of the PPR7 protein, circular dichroism (CD) spectroscopy of recombinant hexahistidine-tagged PPR7 (His-PPR7, section 2.2.4.10) was performed. CD spectroscopy is a spectroscopic technique where the CD of molecules is measured over a range of wavelengths. Secondary structure composition (% helix, sheet, turns, etc.) can be determined by CD spectroscopy in the "far-UV" spectral region (180-260 nm). At these wavelengths the chromophore is the peptide bond, and the signal arises when it is located in a regular, folded environment. Alpha-helix, beta-sheet, and random coil structures each give rise to a characteristic shape and magnitude of CD spectrum (Kelly et al., 2005).

The parameters and procedure for CD spectroscopy of His-PPR7 are as described in section 2.2.12. The measurements provided the absorbance values from 185 to 260 nm as shown in Figure 3.4. The UV CD spectra obtained shows two minimal peaks at 208 and 222 nm. These minimal peaks are associated with the proteins having α-helices (Kelly et al., 2005).



Figue 3.4: Secondary structure analysis of His-PPR7. Circular Dichroism spectroscopy of His-PPR7 was performed using 0.13 mg/mL proteins. The analysis of one typical spectrum (an average of three scans with a spectra band width of 1 nm is shown. X axis shows UV spectra in nm and Y axis shows theta machine units θ.

The analysis of secondary structure was performed using the CDSSTR program (protein reference set 3) from the DichroWeb server (http:// dichroweb. cryst.bbk.ac.uk/ html/process. shtml) by providing the absorbance values, molecular weight and number of amino acids of His-PPR7. The percentage of α-helices, β-strands, turns and unordered regions is summa-

rized in Table 3.3. The CD data obtained show that His-PPR7 is a properly folded and structured protein containing α helices and β-strands.

Table 3.3: Secondary structure data for the His-PPR7 protein. The UV absorbance values for His-PPR7 from 185 to 260 nm obtained by CD measurement were used for the online protein structure prediction program CDSSTR.

His-PPR7 structural data				
α-Helices %	**β-Strands %**	**Turns %**	**Other structures %**	**Total**
23	27	20	29	99

3.3.3 Crystal structure analysis of recombinant PPR7 protein

To understand the functions of proteins at a molecular level, their three dimensional structure is of great importance. The techniques employed for determinig the three dimensional structure include X-ray crystallography and NMR (Nuclear magnetic resonance) spectroscopy. As the CD data for His-PPR7 revealed that it is a properly folded and structured protein, it was further subjected to the crystallization procedure as described (section 2.2.13). Unfortunatelly, the His-PPR7 precipitated soon after preparation and was not able to crystallize. In further attempts, different salt concentrations (150-300 mM NaCl) in the dialysis buffer were applied; however the protein solubility was not increased. In another attempt, the protein was dialysed against CD buffer (30 mM NH_4HCO_3, pH 7.5), where the protein remained in soluble form but upon the concentration procedure of protein, it precipitated again. Further efforts to improve the protein solubility included the removal of C-terminus of recombinant PPR7 protein as this contained some residues that were supposed to interfere in the crystallization of protein e.g. a proline residue at amino acid position 198 which might lead to a high flexiblity of protein domains (Ferdinand Alte, AG Groll, personal communication). Additionally the removal of tag was also recommended. For this purpose, the sequence encoding for 59 till 194th aminoacids was fused to GST-tag as described (section 2.2.4.10.1). After the purification of protein, the tag was cleaved with the help of thrombin as described (section 2.2.4.10.2). However, this attempt to crystallize the shorter form of recombinant PPR7 protein was also not successful and the protein precipitated soon after its purification by gel filtration (section 2.2.4.9).

3.4 Analysis of the function of PPR7 in *C. reinhardtii*

3.4.1 Analysis of PPR7 RNA interference Lines

To investigate the role of PPR7, RNA interference (RNAi) lines were generated according to Rohr et al. (2004). As described in methods (section 2.2.7), an inverted repeat structure comprising of PPR7 coding region was cloned into the NE537 RNAi vector (Rohr et al., 2004). After transformation of the PPR7 RNAi construct into the expression strain *UVM4,* about 1000 transformed colonies were obtained on the paromomycin containing selection media. These transformants were passed through a regular selection process that included a gradual increase in the concentration of 5-fluoroindole, from 1 µM/mL up to 10 µM/mL. Out of these, 60 colonies were found to be resistant to 10 µM 5-fluoroindole, indicating efficient silencing of the vector-encoded Maa7 gene in the non-resistant clones. The RNAi PPR7 lines (iPPR7), that survived on the maximum selection were analysed further by Western blot to determine the level of reduction of PPR7 protein accumulation. In parallel, *UVM4* cells transformed with the empty NE537 vector served as a negative control and taken as wild type for further analyses. Finally, three knock down mutants were selected and analysed along with a dilution series of wild type (Figure 3.5). iPPR7 lines N7 and N21 showed about 60 to 50% knock down of the PPR7 protein. The greatest reduction was observed in line CC19, which contained <20% of PPR7 as compared to the wild type level (Figure 3.5). These three mutants were finally selected for further analyses.

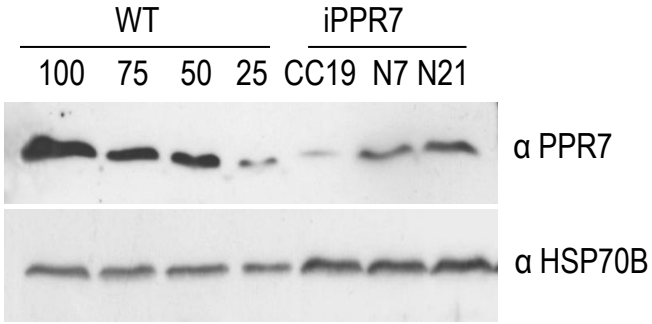


Figure 3.5: PPR7 protein accumulation in RNAi PPR7 lines. Immuno-blot analysis of total soluble proteins (100 µg) isolated from the indicated iPPR7 lines (CC19, N7, N21) and the wild type (WT) dilution series from 100 to 25% was performed by using antibodies raised against indicated proteins. HSP70B served as loading control for stromal proteins.

To investigate the iPPR7 mutants with regard to phenotype, their growth under different light conditions and media was analysed. Drops (1 x 10^6 cells/mL) of the selected iPPR7 mutants along with the recipient *UVM4* strain containing the empty NE537 vector considered as wild type (WT), were spotted on tris-acetate-phosphate (TAP) and high salt minimal (HSM) agar plates under high light (HL, 100 µE/m^2/s), moderate light (ML, 30 µE/m^2/s), and dark (D) con-

ditions. In the absence of acetate in the media, cells are dependent on photoautotrophic growth, whereas they grow mixotrophically or heterotrophically in the presence of acetate depending on the light conditions. As a result, the line CC19, which reveals the strongest reduction in PPR7 protein accumulation, showed inhibited growth under photoautotrophic and mixotrophic conditions in HL (Figure 3.6). The inhibited growth pattern seen under mixotrophic conditions in HL, points to a light sensitive phenotype of PPR7 depleted cells. Other knockdown lines N7 and N21 also showed poor photoautotrophic growth as compared to wild type while growth under mixotrophic in HL was found to be unaffected (Figure 3.6). No substantial differences were observed in the growth pattern of all iPPR7 lines under mixotrophic conditions in moderate light and under heterotrophic growth conditions, in comparison to wild type (Figure 3.6). Furthermore chlorophyll fluorescence measurements for iPPR7 lines were carried out to observe any abnormality in photosystem II (PSII) due to PPR7 knock down. The Fv/Fm (Qy-Max) values obtained for iPPR7 lines were same as wild type, which indicated no clear effect on the PSII function.

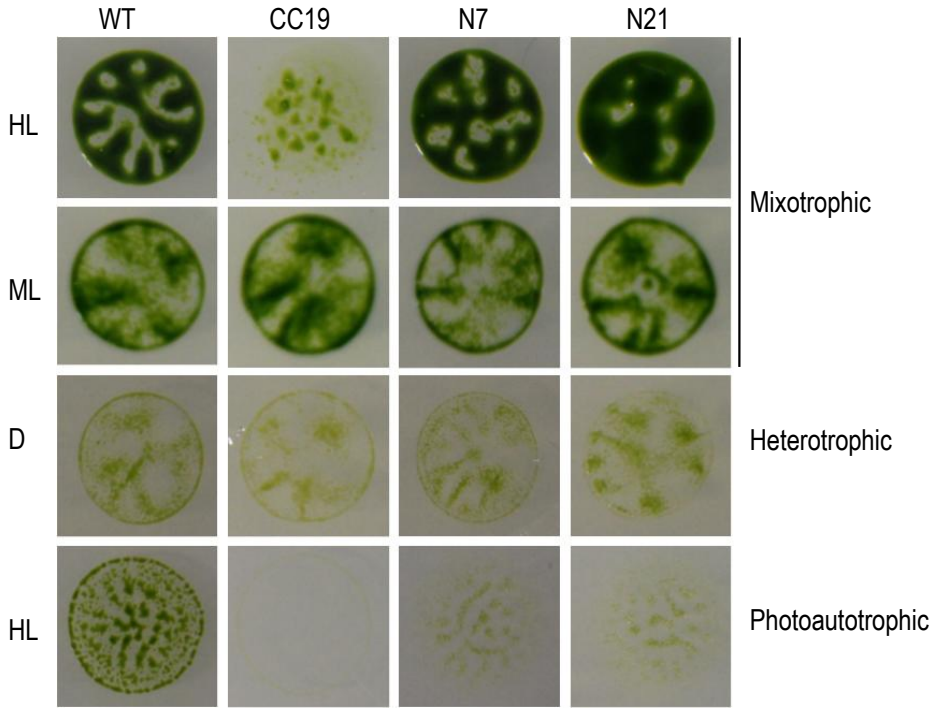


Figure 3.6: Phenotypical characterization of PPR7 knock down lines under different growth conditions. Drop tests were performed by spotting 50 µL of cultures resuspended in HSM to approx 1 x 10^6 cells/mL. The three indicated iPPR7 mutants, CC19, N7, and N21 along with wild type were grown on indicated media under high light (HL), moderate light (ML) and dark (D) conditions for seven days. See text for more details.

The reason for the light sensitive phenotype observed for iPPR7 lines in the growth tests described above could be a less efficient photosynthesis, including proteins which work fur-

ther downstream, like the ATP synthase or Rubisco (Rochaix et al., 2000; Majeran et al., 2001; Johnson, 2011). Therefore, the accumulation of core proteins from photosynthetic complexes, the ATP synthase, and large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) was investigated in iPPR7 lines by immunoblot analysis. The recipient *UVM4* strain containing the empty NE537 vector was selected as wild type (WT). Since a stronger phenotype of the PPR7 knock down lines was observed under high light conditions, cells from both, high and moderate light conditions were analysed (Figure 3.7 A). Whereas no changes in protein accumulation could be observed under moderate light conditions (Figure 3.7, right panels), under high light conditions (Figure 3.7, left panels), a reduction in the signal intensity of the PSI subunit PsaA quantified by image analyser (section 2.2.4.3) was observed to ~42% in the mutant line CC19 as compared to wild type (Figures 3.7 A and B). The other two mutant lines, N7 and N21 showed a PsaA reduction of about 20 and 17%, respectively, as compared to wild type levels (Figures 3.7 A and B). This degree of reduction of PsaA correlates with the levels of PPR7 protein decrease in iPPR7 lines (Figure 3.5). The PSII core subunit D2 and large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) were found unaltered in all the three iPPR7 lines as compared to the wild type levels (Figure 3.7 A). Also, no change in steady state levels of ß-subunit of ATP synthase and Cyt-b_6 was observed in the mutants (Figure 3.7 A). As the assembly of chloroplast multiprotein complexes i.e. Cyt-b_6/*f*, PSII, PSI, as well as the ATPase in *C. reinhardtii* occurs according to control by epistasy of synthesis (CES) where the presence of one subunit is required for continued synthesis of another chloroplast encoded subunit from the same protein complex (Choquet et al., 1998, 2003; Wostrikoff et al., 2004; Minai et al., 2006; Drapier et al., 2007). So according to the CES principle one can assume that the investigated protein complexes Rubisco, ATP synthase, and PSII accumulate to normal levels in the iPPR7 mutants. A normal functionality of the latter one is further supported by chlorophyll fluorescence measurements (section 2.2.11), since no increased fluorescence was detected by determination of the Fv/Fm (Qy-Max) values for the iPPR7 lines which would be characteristic for a defect in PSII. However, the only detected change in protein levels, observed for the accumulation of the PSI subunit PsaA, implies that the light sensitive phenotype of RNAi PPR7 lines could be a result of a decrease in PSI levels.

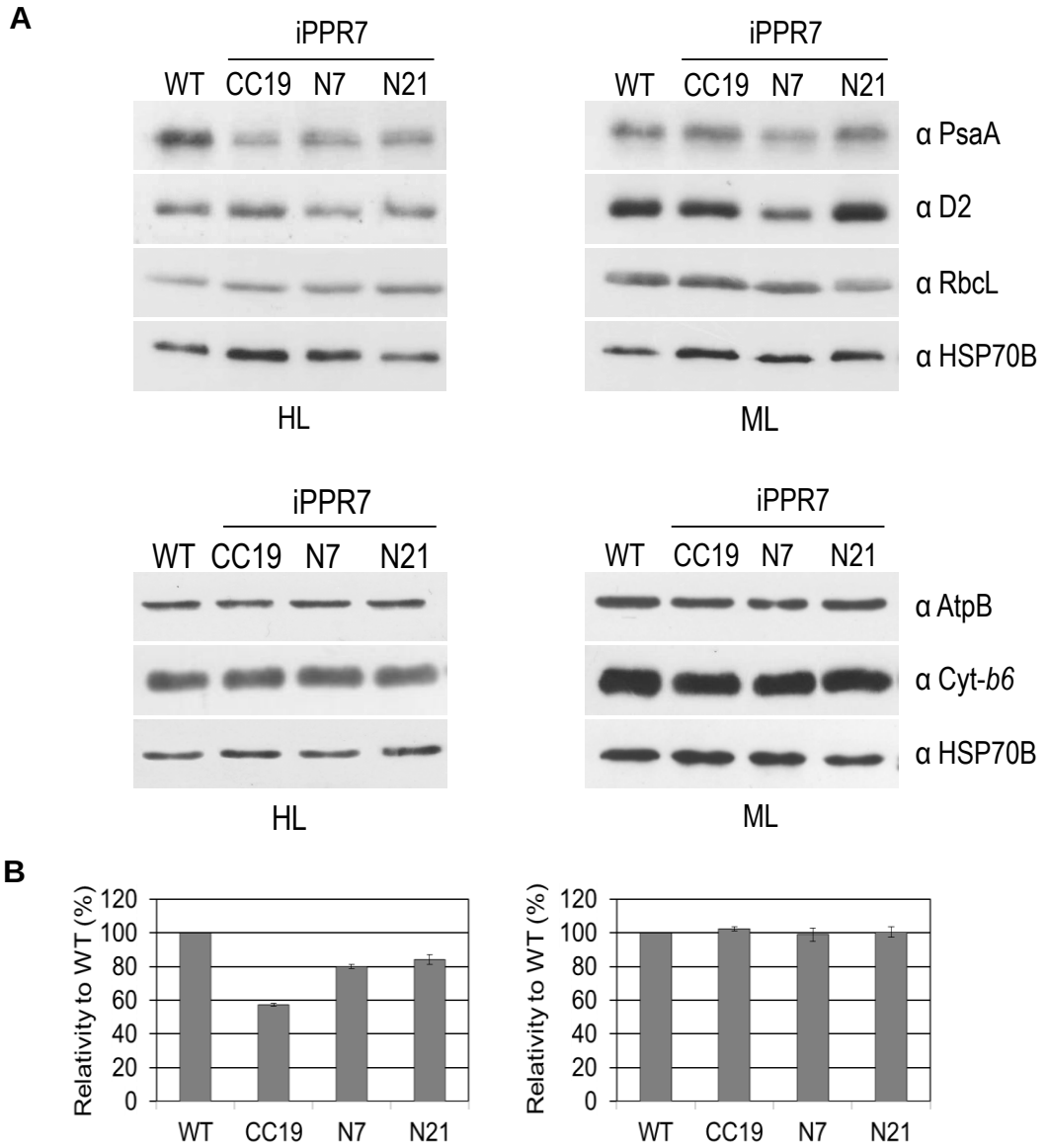


Figure 3.7: Immunoblot analysis of representative photosynthesis-related protein complexes. A: Total proteins (15 µg) from iPPR7 lines and wild type (WT) strain grown in TAPS under high light (HL; 100 µE/m^2/s) and moderate light (ML; 30 µE/m^2/s) were fractionated by SDS-PAGE, and blots were probed with antibodies raised against individual subunits of PSII (D2), PSI (PsaA), Cyt-b_6/*f* complex (Cyt-b_6) and chloroplast ATP synthase (ß-subunit). HSP70B served as a loading control. **B:** Quantification of signal intensity for PsaA in A. Error bars represents the standard deviations of values from two independent experiments.

3.4.2 Characterization of RNA binding property of PPR7

For many PPR proteins characterized so far an RNA binding activity was described, which is important for their functionality in RNA metabolism (Schmitz-Linneweber et al., 2005, 2006; Gillman et al., 2007; Beick et al., 2008; Pfalz et al., 2009). Therefore, the following chapters focus on the analysis of the RNA binding properties of PPR7 and the identification of its specific RNA target(s).

3.4.2.1 PPR7 is a component of a high molecular weight RNAse-sensitive complex

PPR proteins are mostly found in RNase-sensitive, high molecular weight (HMW) complexes (Fisk et al., 1999; Meierhoff et al., 2003; Williams and Barkan, 2003; Schmitz-Linneweber 2006; Gillman et al., 2007; Johnson et al., 2010). To find out whether PPR7 forms part of an RNA/protein complex, size exclusion chromatography (SEC) was performed. Chloroplast stroma was isolated as described in section 2.2.4.7 and fractionated by SEC, with or without previous treatment with RNase I (section 2.2.10). PPR7 eluted in fractions 2 to 9, which indicates that it is part of a high molecular weight complex in the range of 440-1,000 kDa (Figure 3.8). The peak of PPR7 elution was observed in fractions 4-7 which corresponds to a complex size of 500 to 800 kDa. In the presence of RNase, however, this complex size was reduced to approximately $\geq$ 300 kDa suggesting that PPR7 is part of an RNase-sensitive high molecular weight complex (Figure 3.8).

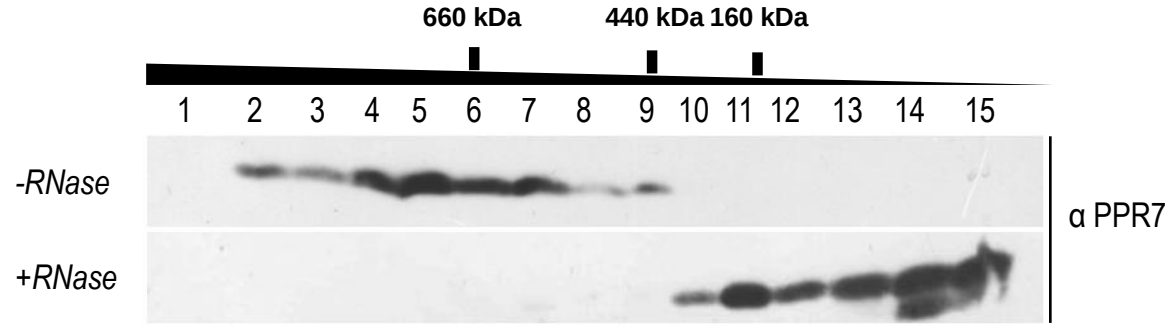


Figure 3.8: PPR7 is a component of a high molecular mass ribonucleoprotein complex. Stromal proteins (~2 mg) from wild type strain CC406, were separated by size exclusion chromatography in the absence (-RNase) or presence (+RNase) of 250 U RNase One. Fractions were subjected to immunoblot analysis using PPR7 specific antibody. Molecular weights were calculated by parallel analysis of HMW calibration markers. Elution profiles of marker proteins (kDa) are given at the top, together with the respective fraction numbers.

3.4.2.2 The recombinant PPR7 protein reveals intrinsic RNA binding activity

To achieve indications for an intrinsic ability of PPR7 to bind RNA, *in vitro* RNA binding assays were performed. Therefore, full-length hexahistidine-tagged PPR7 protein (His-PPR7) was heterologously expressed in *E. coli*, purified by affinity chromatography and subjected to UV crosslinking experiments as described (sections 2.2.4.10.2 and 2.2.10). Since no specific RNA targets were known, His-PPR7 was tested for its ability to bind the 5' UTRs of *in vitro* transcribed radiolabelled *rbcL, atpA,* and *psbD* RNAs (section 2.2.10). As shown in Figure 3.9, His-PPR7 exhibits a clear RNA binding activity for all transcripts tested (Lane 4, 5 and 6) The RNA bound His-PPR7 signals were detected at ~24 kDa. In the first three lanes the radiolabelled *in vitro* transcribed RNAs were loaded without incubation with His-PPR7 protein as

negative control, where no signals were detected in the range of His-PPR7. The signals detected at ~35 kDa in all the six lanes are probably from binding of radiolabelled nucleotides to the RNase, used in the experiment to degrade the free RNAs as described in section 2.2.10. These results indicate that PPR7, like other characterized PPR proteins, has an RNA binding moiety even though under conditions used in this assay no specificity for a certain RNA target was seen.

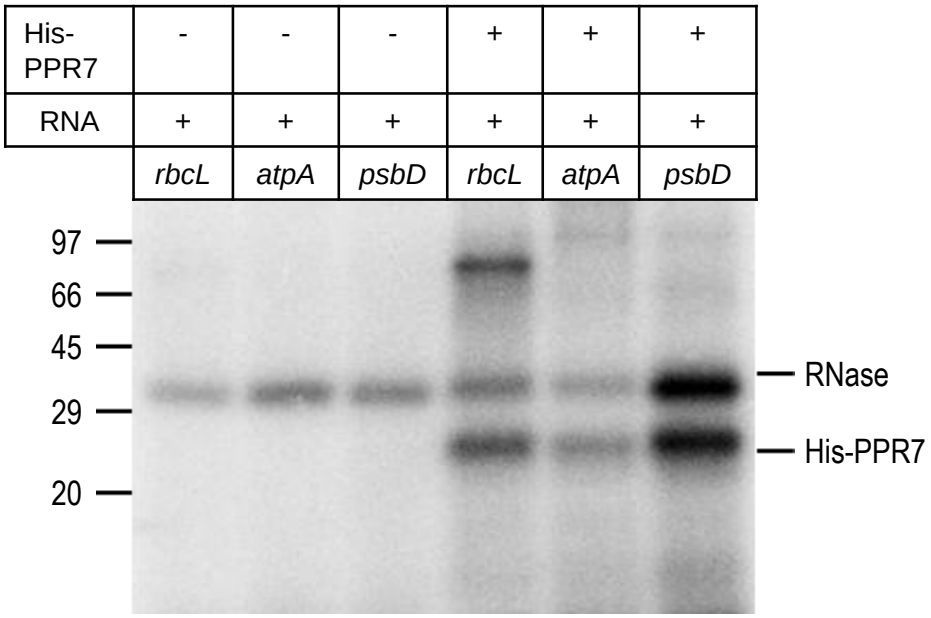


Figure 3.9: RNA binding activity of full length recombinant PPR7 protein (His-PPR7). 3µg His-PPR7 was UV-crosslinked to radiolabelled *rbcL, atpA,* and *psbD* 5´UTR RNA probes (50-100 kcpm), fractionated by SDS-PAGE and subjected to autoradiography (sections 2.2.10). In the first three lanes, radiolabelled RNA was added without incubation with His-PPR7 as control. Lane 4 (*rbcL*), 5 (*atpA*), and 6 (*psbD*) shows the RNA probes incubated with His-PPR7.

3.4.2.3 Identification of target RNAs of PPR7 by RIP-chip analysis

To investigate the association of PPR7 with specific chloroplast transcripts, RNA co-immunoprecipitation and chip hybridization (RIP-chip) experiments were performed. RIP-chip is a microarray-based strategy for identifying the *in vivo* RNA ligands of organellar RNA binding proteins (Schmitz-Linneweber et al., 2005). First, chloroplasts from a cell wall–deficient strain were isolated as described (section 2.2.4.6). The isolated chloroplasts were further fractionated to stroma and crude thylakoids (section 2.2.4.7). The stromal fraction obtained, was used for co-immunoprecipitation of PPR7 containing complex and bound RNAs, by using a purified antibody raised against PPR7 (αPPR7, section 2.2.8). The pre-immune serum was used as negative control. 10% of pellet and supernatant fractions from both, pre-immune

and αPPR7 samples, were analysed by immunoblot, to determine whether the PPR7 antibody is able to precipitate the native PPR7 protein (Figure 3.10 A).

In the pellet fraction from αPPR7, a strong signal showing precipitated PPR7 was detected, while the signal became weaker in the supernatant fraction, indicating that about 75% of PPR7 from the input stroma fraction is precipitated, based on signal quantification. In the negative control, using pre-immune serum, no signal was detected in the pellet fraction and a strong signal was obtained in the supernatant indicating no unspecific binding of PPR7 (Figure 3.10 A). This result designates that PPR7 can be successfully precipitated by using the PPR7 antibody.

In parallel, RNA was isolated from immunoprecipitated pellets and supernatants of the αPPR7 and pre-immune fractions (section 2.2.8) and subjected to RIP-chip hybridizations as described by Schmitz-Linneweber et al. (2005). For RIP-chip analysis, a microarray was designed with 166 overlapping PCR products representing the complete *C. reinhardtii* chloroplast genome and 15 overlapping PCR fragments representing the complete mitochondrial genome (see section 2.2.9) in collaboration with the group of Prof. Dr. Christian Schmitz-Linneweber (Institute for Genetics, Humboldt University, Berlin). During RIP-chip analysis, the co-immunoprecipitated RNA and the supernatant RNA was labelled with red and green dyes, respectively and hybridized onto the full genome chloroplast microarray in a tiling fashion (section 2.2.9). Interestingly, the signal intensities obtained from RIP-chip analysis by differential enrichment of the signal intensities obtained for the RNAs precipitated by pre-immune and antiserum provided an unusually high number of putative PPR7 binding sites. The following binding regions of the PPR7 protein were obtained and genomic positions as well as PCR product IDs according to Annex B are given in brackets: 5’ UTR, coding region and 3’ UTR of *rrnS* (ID 35, 36, and 37, 37708-40995), intergenic region of *trnE2* and *psbH* (ID 70, 75794-77003), coding region of *rpoC2* (ID 103, 113327-114901), 5’ UTRs of *atpA* and *rbcL* (ID 112, 123734-124902), intergenic region of *cemA* and *atpH* (ID 116, 128314-129509), coding region of *tscA* (ID 122, 134706-135910) and intergenic region of *psaJ* and *atpI* (ID 136, 170490-171665) (Figure 3.10 B; see Annex B). These data suggest that PPR7 bind to the above mentioned target RNAs where it probably plays a role in RNA metabolism. The positions of the mentioned PCR products within respective regions identified by RIP-chip are designated under A) in Figures 3.12-3.16.

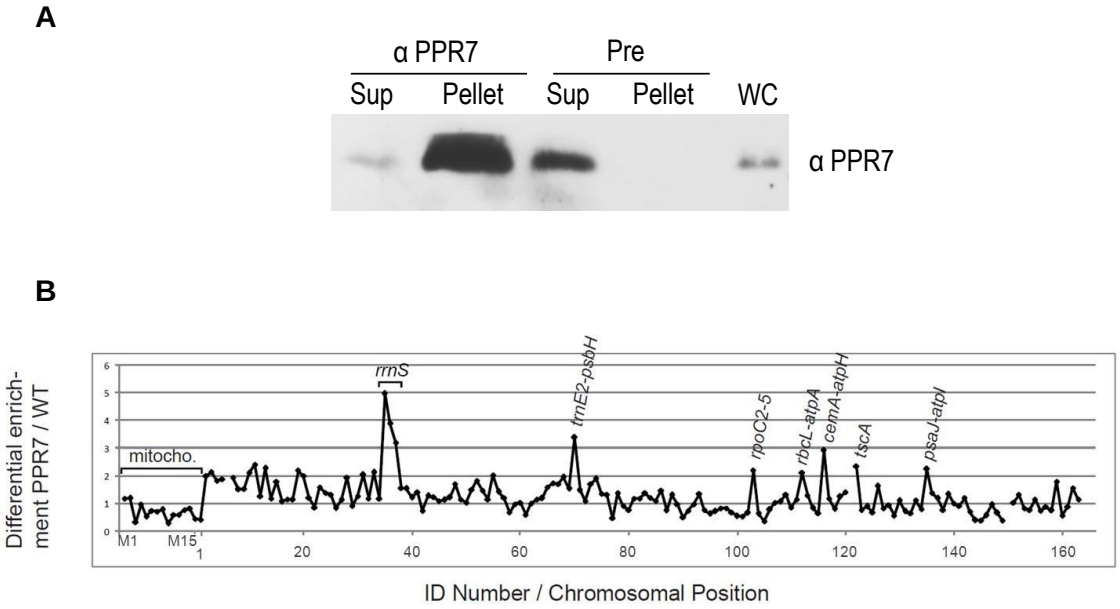


Figure 3.10: Co-immunoprecipitation and RIP-chip for PPR7. A: Immunoprecipitation of the native PPR7 protein. 10% from pre-immune (Pre) and αPPR7 supernatants (Sup) and precipitates (pellet) were subjected to immunoblot analysis using a purified PPR7 specific antibody. 50 µg of whole cell proteins (WC) were loaded alongside the immunoprecipitation samples for identification of native PPR7 signal. **B: Association of PPR7 with chloroplast transcripts.** PPR7 was enriched by immunoprecipitation. Co-precipitated RNA was labelled with Cy5 and unbound RNA in the supernatant was labelled with Cy3. Labelled RNAs were hybridized onto a microarray representing the chloroplast genome of *C. reinhardtii* in a tiling fashion. After scanning, the differential enrichment ratios (FCy5:FCy3) were normalized between four independent assays using antibody against PPR7 and two control assays using pre-immune serum. The median normalized values for replicate spots from the PPR7 immunoprecipitations were divided by those from pre-immune controls and plotted according to fragment number on the *C.reinhardtii* organellar genome microarray. Fragments are numbered according to chromosomal position (the first 15 fragments represent mitochondrial genes). Holes in the graph indicate PCR products that did not meet quality standards as defined in Schmitz-Linneweber et al. (2005). The data used to generate this figure are provided in Annex B, Tables 1 and 2.

3.4.2.4 Semi quantitative RT-PCR of PPR7 co-immunoprecipitated RNAs

To confirm the putative RNA targets of PPR7 as found by RIP-chip experiments (section 3.4.2.3), RT-PCRs of PPR7 co-immunoprecipitated RNAs were performed. Therefore, cDNAs from pellet and supernatant RNAs of αPPR7 and pre-immune serum co-immunoprecipitations were generated as described in section 2.2.3.4. Gene specific primers for possible targets identified by the RIP-chip analysis as well as primers for control genes which are not supposed to be bound by PPR7 were used for reverse transcription (RT+) and semi quantitative PCRs using the generated cDNA as template. In parallel, reactions without reverse transcriptase were carried out as negative controls to exclude an amplification of contaminating DNA fragments (RT-). As shown in Figure 3.11, the stronger signals in the RT+ PCR reactions from αPPR7 pellets compared to those of the pre-immune pellet for the putative RNA target regions of PPR7 (*rrnS*, *trnE2-psbH*, *rpoC2, rbcL, atpA*, *cemA-atpH*, *tscA* and

psaJ-atpI) indicate indeed a binding of PPR7 to these RNAs. The weak PCR products obtained in pre-immune pellets are most probably due to the unspecific binding of RNAs to the magnetic beads used for co-immunoprecipitations (Figure 3.11). For controls, sequences from *chlL* and *psbD* were amplified using cDNA from pellets and supernatants as these regions seem not to be bound by PPR7 based on the RIP-chip results (Figure 3.10 B). No obvious differences in copy number of PCR products were observed for the controls (Figure 3.11). These results imply that the enrichment of putative target RNAs in the pellet fraction from PPR7 antibody are due to specific binding of PPR7 to these regions. No PCR product was obtained in RT-, indicating that there was no genomic DNA contamination in αPPR7 co-immunoprecipitated RNA. For PCR products obtained in αPPR7 supernatants for *trnE2-psbH*, *atpA*, *cemA-atpH*, *tscA* and *psaJ-atpI*, a reduction in copy number was observed as compared to that in pre-immune supernatants, which indicates reduction in transcript in αPPR7 supernatant due to co-immunoprecipitation. These results verified that RNAs from all putative target regions are highly enriched in PPR7 immunoprecipitates but not from pre-serum, thereby supporting the binding of PPR7 to these RNAs.

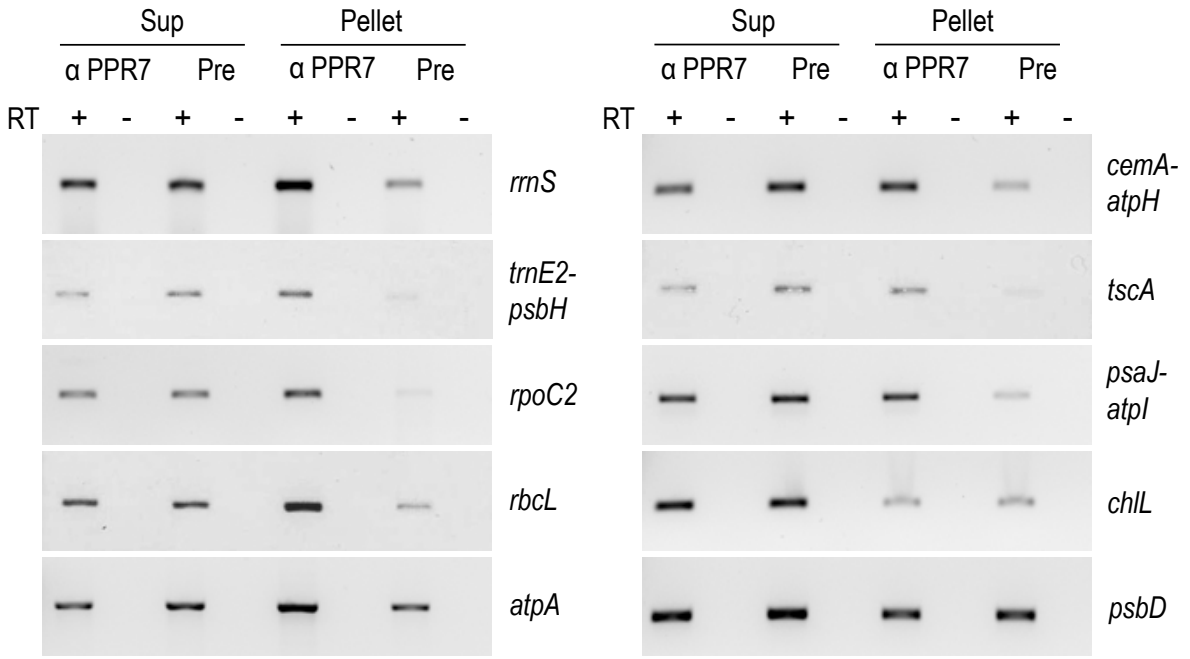


Figure 3.11: RT-PCR of PPR7-copurified RNA. RT-PCR was performed using specific primer sets of the genes indicated at the right of each panel. RNA used for RT-PCR was obtained from co-immunoprecipitation of PPR7 supernatant (Sup) and pellet, using PPR7 specific antibody (αPPR7) as described (section 2.2.3.4). Pre-immune serum (Pre) was used as a control. RT- indicates the absence of reverse transcriptase in the RT reaction. The number of cycles used to amplify specific sequences is 20 for each mentioned product. *chlL* and *psbD* were used as controls. PCR product (7 µL) was subjected to DNA electrophoresis. Each RT-PCR reaction was performed three times to ensure reproducibility.

3.4.2.5 The role of PPR7 at the identified putative target RNAs

Many PPR proteins characterized so far are known to be involved in processing and/or stabilization of their target RNAs. To investigate the function of PPR7 at the identified putative targets (section 3.4.2.3), RNA gel blot analyses were performed on PPR7 deficient RNAi lines decribed above (section 3.4.1). Therefore, total cellular (WC) RNA of wild type (*UVM4* with empty NE537 vector) and the three selected iPPR7 lines were used in the transcript analysis. Suitable Digoxigenin (Roche) labelled PCR probes (section 2.2.3.6) at the putative binding sites were designed to detect the respective transcripts of *rrnS*, *trnE2-psbH*, *rpoC2, rbcL, atpA*, *cemA-atpH*, *tscA,* and *psaJ-atpI*. The analysis of each putative target transcript of PPR7 is described separately in the following section.

The first hit from RIP-chip data was a region covering 5' UTR, coding region and 3' UTR of small subunit of ribosomal RNA (*rrnS*). As depicted in Figure 3.12 A, the plastid rRNA genes (*rrn*) exist in an operon in the chloroplast genome. The genetic order of the *rrn* genes is 16S, 7S, 3S, 23S, and 5S, respectively (Harris et al., 1994). Additionally two tRNA genes (*trnI* and *trnA*) are located between the 16S and 7S genes.

To analyse the size, composition, and relative abundance of transcripts from the chloroplast *rrn* gene cluster in iPPR7 lines in comparison to wild type cells, RNA gel blot analysis was performed (Figure 3.12 B). For a more comprehensive view on the accumulation of respective transcripts, three different probes were used as indicated in Figure 3.12 A. The probe upstream of the 5' end of *rrnS* (P1) labelled a high molecular weight (HMW) precursor transcript of ~7.35 knt (a) containing all sequences from 16S to 5S and a partially processed *rrnS* transcript of ~1.6 knt (b) in iPPR7 and wild type strains (Figure 3.12 B). The relative abundance of the large precursor (a) was found to be increased in iPPR7 lines as compared to wild type (Figure 3.12 B, C, left panels). iPPR7 line CC19, showed a pronounced effect of increase in the HMW *rrn* precursor (~70%) than in N7 and N21 (~20%) indicating its relation to the extent of PPR7 down regulation. Due to increased accumulation of the precursor, the partially processed *rrnS* transcript (b) showed a decrease in abundance as observed in PPR7 deficient mutants compared to wild type cells (Figure 3.12 B, C, left panels). To verify if there are changes in the accumulation of the mature 16S transcript a second probe (P2) designed in the coding region of mature *rrnS* (c) was used in an RNA gel blot analysis (Figure 3.12 B, C, middle panels). However, no obvious alterations were observed between iPPR7 and wild type RNAs.

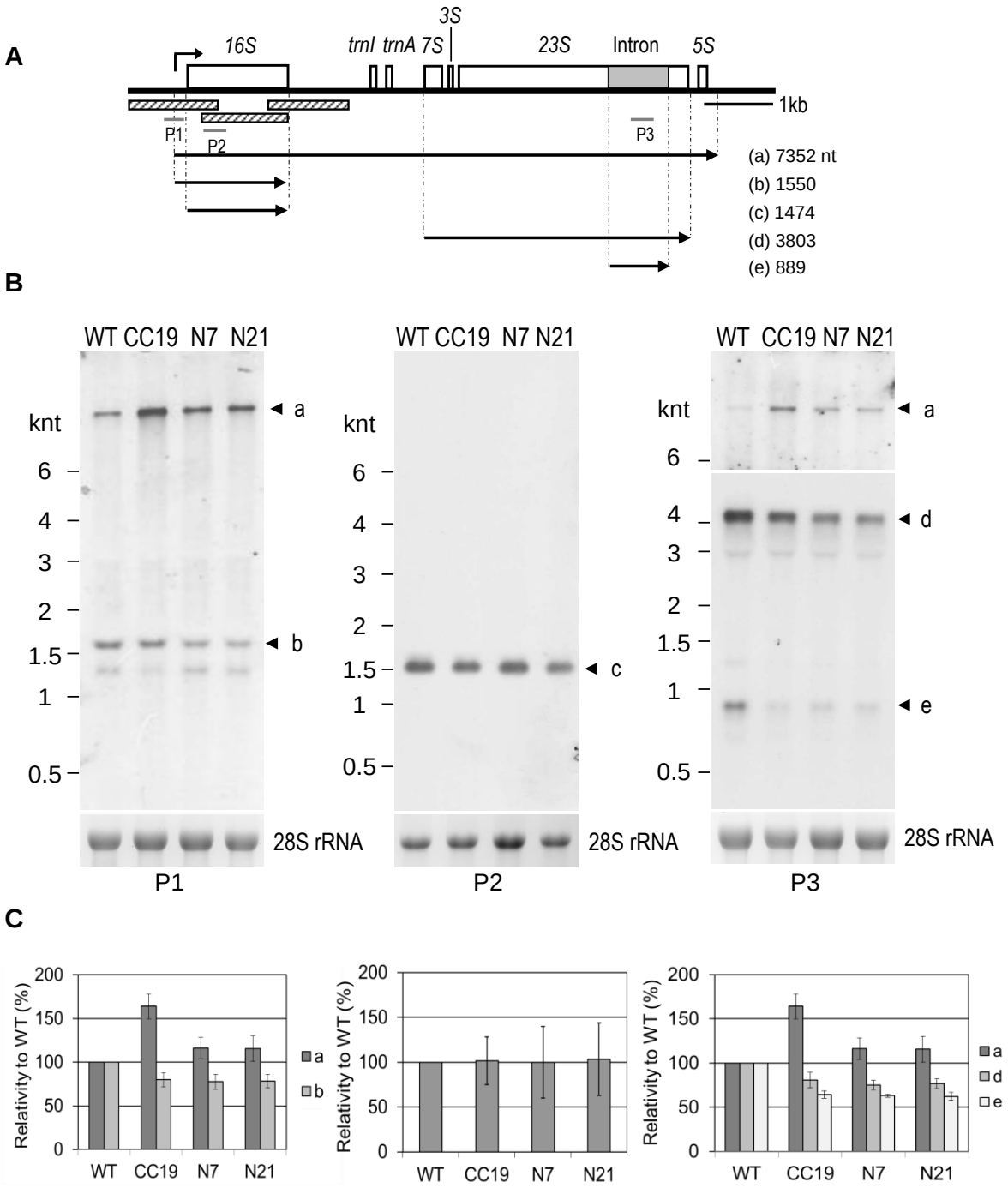


Figure 3.12: Northern analysis of *rrn* transcripts. A: Schematic diagram of *rrn* gene cluster in the *C. reinhardtii* chloroplast genome. The open boxes represent the genes present in the cluster. The PCR products at which RIP-chip for PPR7 supports binding are indicated as shaded grey boxes. Positions of DIG labelled probes P1-P3 designed are indicated as grey bars. Transcripts and respective sizes are indicated by black arrows according to Holloway and Herrin (1998). The promoter mapped upstream of the 16S gene is indicated by a bent arrow (Holloway and Herrin, 1998). **B:** Total cellular RNA extracted from wild type (WT) and iPPR7 lines (CC19, N7 and N21) was subjected to Northern blot analysis. DIG labelled probe used to label the transcripts are indicated at the bottom of respective blot. 10 µg RNA was used for blots probed with P1 and P3 while, 1 µg RNA was used for blot probed with P2. The upper portion of blot probed with P3 is a prolonged exposure to visualize the precursor transcript. Ethidium bromide-stained RNA gels are shown as loading controls (28S rRNAs). The positions of RNA size markers are indicated on the left. **C:** Quantification of signal intensity for transcripts detected in B, with error bars representing standard deviations calculated from three independent experiments. Signal intensities obtained for the WT were set to 100%.

A third probe (P3) at the intron region of *rrn* large subunit (23S) was applied to a similar blot which was supposed to detect amongst others the HMW precursor. The probe labelled the HMW precursor (a), a transcript consisting of the 7S, 3S, and 23S sequences (d), and the spliced intron sequence of ~0.88 knt (e). A similar transcript pattern was also observed by studies of Holloway and Herrin (1998) using a 23S intron probe, even though the HMW transcript could not be detected in their analysis. As expected, the same pattern of relative abundance of the large precursor transcript (a) was observed as seen for the probe P1. As a consequence, the cleaved 7S-23S precursor as well as the spliced intron sequence showed a decrease in abundance in PPR7 deficient mutants as compared to wild type cells (Figure 3.12 C). These observations confirm an important role of PPR7 in processing of this *rrn* gene cluster.

The second hit from RIP-chip data was a region covering the *psbH* gene, downstream to it *trnE2* gene and their intergenic region. *psbH* is part of the *psbB-psbT-psbH* gene cluster and codes for a small PSII subunit (Johnson and Schmidt, 1993; Hong et al., 1995; Vaistij et al., 2000a). It has been found necessary for phototrophic growth and has a putative role in assembly or stability of the PSII complex (Summer et al., 1997; O'Connor et al., 1998). For the stability of the *psbB-psbT-psbH* transcription unit and also for the mature *psbH* transcript, a nuclear factor Mbb1 has been characterized (Monod et al., 1992; Vaistij et al., 2000b) which binds at the 5' UTR of the precursor transcript. Upon investigating the role of PPR7 on the *psbB-psbT-psbH* transcript the RNA gel blot analysis was performed by using a probe designed at the coding region of *psbH* gene (Figure 3.13 A, left panel). As shown in Figure 3.13 B (left panel) the probe detected four *psbH* related transcripts i.e. 0.9 knt (a), 0.8 knt (b), 0.5 knt (c), and 0.4 knt (d). The detected transcripts were also previously reported by Monod et al. (1992) and Vaistij et al. (2000a). The analysis revealed a reduction in all *psbH* related transcripts in PPR7 deficient mutants, where iPPR7 line CC19 showed a reduction of about 20% as compared to wild type (Figure 3.13 C, left panel). These observations reveal that PPR7 is likely a stability factor for monocistronic *psbH* transcript, where its depletion causes compromised stability of *psbH* transcripts.

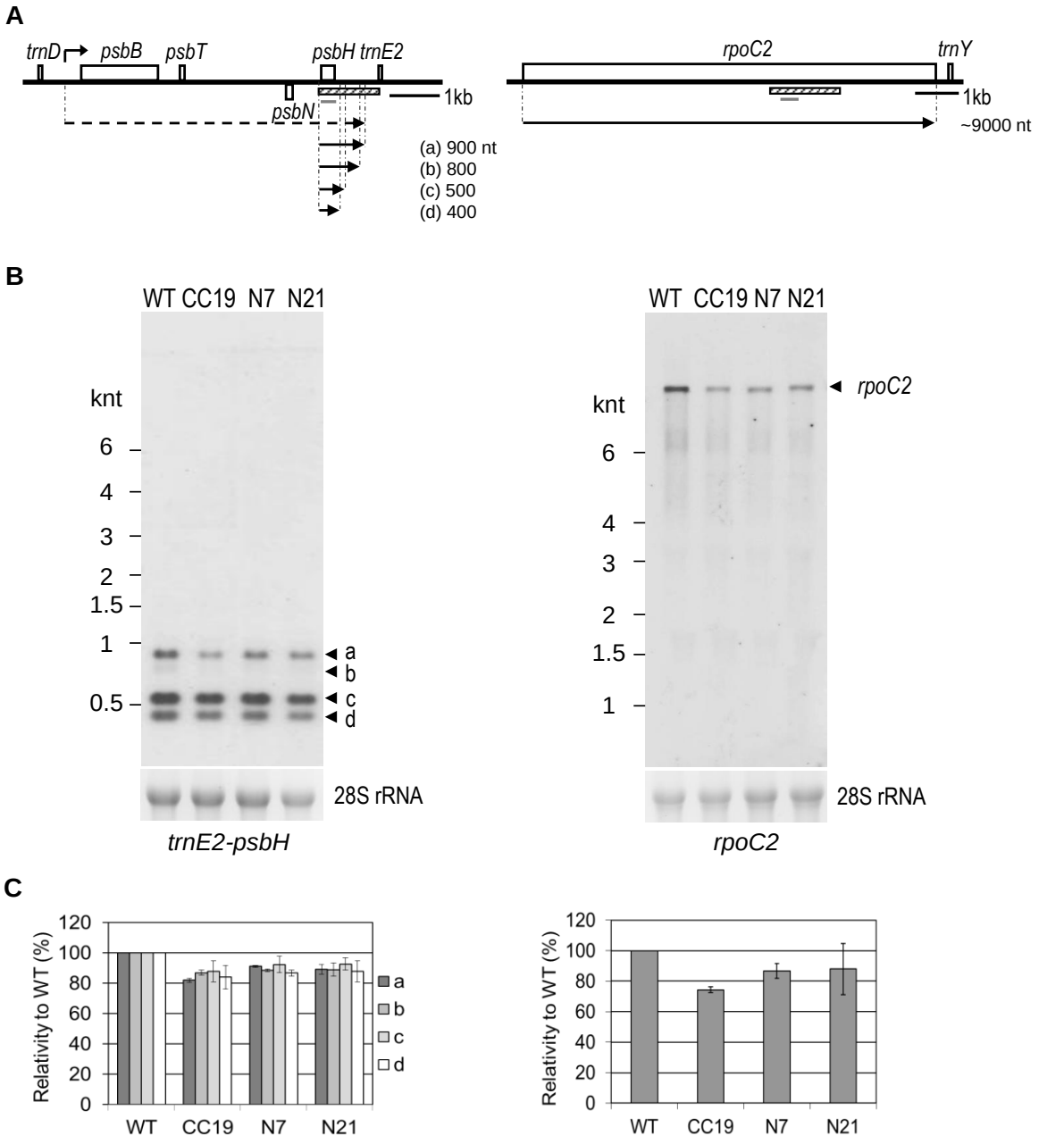


Figure 3.13: Northern analysis of *psbH* and *rpoC2* transcripts. A: Schematic diagrams of the *trnE2-psbH* region and *rpoC2* gene in the *C. reinhardtii* chloroplast genome. The open boxes represent the genes. Genes above the line are transcribed from left to right, genes below the line from right to left. PCR products at which RIP-chip analysis provided hits for binding are indicated as shaded grey boxes. Positions of DIG labelled probes designed are indicated as grey bars. Transcripts and respective sizes for *psbH* are indicated by black arrows according to Monod et al. (1992). The dotted arrow represents the polycistronic transcript as described by Vaistij et al. (2000a). Mapped promoter for *psbH* containing transcript is indicated by bent arrow (Vaistij et al., 2000a). Promoter for *rpoC2* is not mapped, therefore not indicated. **B:** 10 µg of total cellular RNA extracted from wild type (WT) and iPPR7 lines (CC19, N7 and N21) was subjected to Northern blot analysis. DIG labelled probes used to label the transcripts are indicated at the bottom of respective blots. Ethidium bromide-stained RNA gels are shown as loading controls (28S rRNAs). The positions of RNA size markers are indicated on the left of each blot. **C:** Quantification of signal intensity for the transcripts detected in B, with error bars representing standard deviations from three independent experiments. Signal intensities obtained for the WT were set to 100%.

The third hit from RIP-chip data showing a putative binding site for PPR7 protein was a region covering almost the middle part (more towards 3') of *rpoC2* gene, encoding a subunit of the bacterial type RNA polymerase (see section 1.3.1). Genes of plastidial genomes are transcribed by DNA-dependent RNA polymerases (Fong and Surzycki, 1992; Troxler et al., 1994). Plastid genomes contain *rpoA, rpoB, rpoC1*, and *rpoC2* coding for *α, ß, ß',* and ß'' RNA polymerase subunits.

The RNA gel blot analysis using a probe designed at the coding region of *rpoC2* gene (Figure 3.13 A, right panel) labelled a HMW weak band of ~9 knt. This corresponds to the size of *rpoC2* gene in chloroplast genome. As shown in Figure 3.13 B, the abundance of *rpoC2* transcript is altered in PPR7 deficient mutants as compared to wild type cells. The reduction level in CC19 strain is higher (> 20%) as compared to N7 and N21 strains which reveal about 15% reduction (Figure 3.13 C, right panel). The extent of reduction in *rpoC2* transcript is in accordance to the knockdown level of iPPR7 lines. This result shows that like in the case of the *psbH* transcript, PPR7 seems to play a role of a stability factor for *rpoC2* transcript as well.

The fourth putative target of PPR7 as revealed by RIP-chip analysis is a region that covers 5' UTRs of the *rbcL* gene as well as the *atpA* gene cluster. *rbcL* encodes for Rubisco large subunit and has an opposite orientation in the genome as the *atpA* cluster (Figure 3.14 A).

For the stability of mRNAs of prokaryotic origin and organelles, the essential factors are known to be located at 5' UTR (Salvador et al., 1993; Grunberg-Manago, 1999). In *C. reinhardtii, rbcL* transcripts are known to fold at their 5' end into two stem–loop structures. These two stem loops are separated by a short single-stranded region, which is known as a sequence necessary for stabilizing the transcript (Anthonisen et al., 2001). Furthermore a PPR protein MRL1 has been characterized to bind at the 5' UTR of *rbcL* transcript which is required for its stability (Johnson et al., 2010).

To investigate the putative role of PPR7 at the 5' UTR of *rbcL*, RNA gel blot analysis using a probe designed at the 5' UTR of *rbcL* gene (Figure 3.14 A) labelled a non-processed 2.9 kb transcript (a) and the processed mature *rbcL* mRNA (b, Figure 3.14 B). In contrast to the WT RNA the 2.9 kb transcript appeared at a much lower level in iPPR7 lines, indicating that the 2.9 kb transcript is either less transcribed or is quickly degraded (Figure 3.14 B, left panel). A second blot with less amount of RNA was hybridized with the same probe, to detect quantifiable signals of the highly abundant mature *rbcL* transcript so that any difference for mature *rbcL* transcript in iPPR7 lines can be observed as compared to wild type (Figure 3.14 B, right panel). After normalizing the signals to ethidium bromide stained 28 S rRNA signals, quantification of signals was carried out. According to the data obtained, it was observed that the mature *rbcL* transcript level is slightly reduced in iPPR7 line CC19 (~15%) as compared to wild type cells (Figure 3.14 C, right panel). In the other two lines, N7 and N21, negligible reduction was observed as compared to wild type cells. These results indicate that like in the

above mentioned cases of *psbH* and *rpoC2* transcripts, PPR7 could act as a stability factor for *rbcL* as well by acting on its 5' UTR.

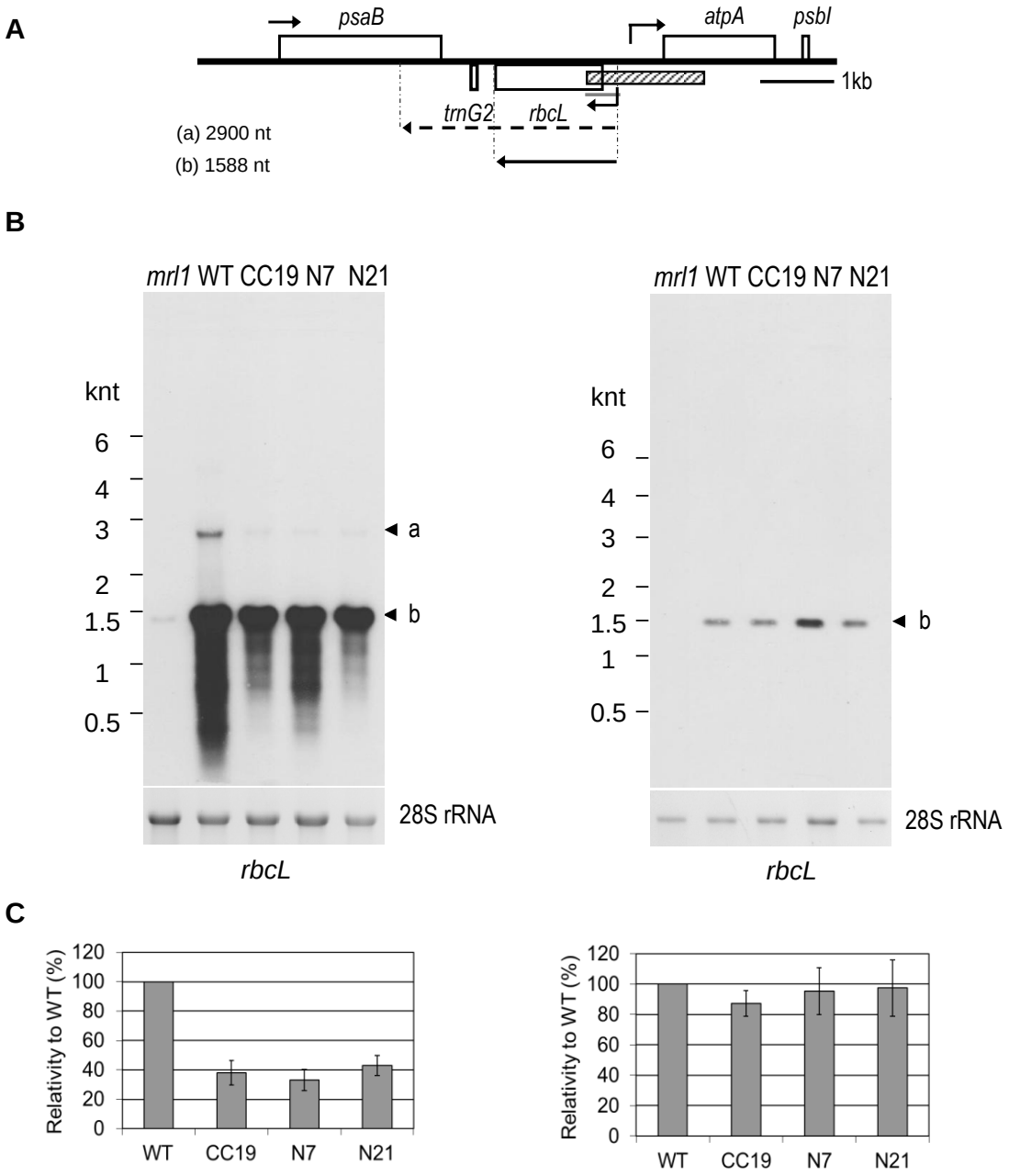


Figure 3.14: Northern blot analysis of *rbcL* transcripts. A: Schematic diagram of *rbcL/atpA* region in the *C. reinhardtii* chloroplast genome. The open boxes represent the genes. PCR products at which RIP-chip for PPR7 provided hit for binding is indicated as shaded grey box. Positions of DIG labelled probe designed is indicated as grey bar. Mature transcript and respective size for *rbcL* is indicated by black arrow according to Johnson et al. (2010). The precursor transcript for *rbcL* observed in this study is shown as doted arrow. Mapped promoters are indicated by bent arrows (Salvador et al., 1993) . **B:** 5 µg (left) and 200 ng (right) of total cellular RNA extracted from wild type (WT) and iPPR7 lines (CC19, N7 and N21) were subjected to Northern blot analysis. DIG labelled probe used to label the transcripts is indicated at the bottom of respective blot. Ethidium bromide-stained RNA gels are shown as loading controls (28S rRNAs). The positions of RNA size markers are indicated on the left. **C:** Quantification of signal intensity for the transcripts detected in B, with error bars representing standard deviations from three independent experiments. Signal intensities obtained for the WT were set to 100%.

The *atpA* gene cluster in *C. reinhardtii* chloroplast genome has been characterized in detail (Drapier et al., 1998). It contains four genes and three promoter regions (Drapier et al., 1998; Figure 3.15 A). The genes include *atpA*, *psbI*, *cemA* and *atpH*, encoding for α-subunit of ATP synthase (Dron et al., 1982; Hallick, 1984; Leu et al., 1992; Drapier et al., 1998), a small PSII subunit (Boudreau et al., 1994; Kunstner et al., 1995), a putative envelope membrane protein involved in inorganic carbon uptake (Rolland et al., 1997) and subunit III of the chloroplast ATP synthase (Lemaire and Wollman, 1989), respectively.

RIP-chip data for PPR7 revealed the region including 5' UTR of *atpA* and also a sequence that covers the 3' end of *cemA* and the whole coding region of *atpH* as putative binding sites of PPR7. To investigate the putative role of PPR7 at the 5' UTR of *atpA*, RNA gel blot analysis was performed using a probe designed at the 5' UTR of *atpA* gene (P1, Figure 3.15 A). Same as for highly abundant *rbcL* transcript, a blot with less amount of RNA was used to detect quantifiable signals. The di- and monocistronic *atpA* transcripts were labelled in wild type and iPPR7 lines (Figure 3.12 B, left panel). The quantification data indicates only negligible differences for mutants as compared to wild type (Figure 3.12 C, left panel). When RNA gel blots were applied with a probe covering the 3' end of *cemA* and *atpH* coding region (P2, Figure 3.15 B, right panel), the ~5.3 knt tetracistronic (a), ~3 knt tricistronic (d) and ~0.4 knt monocistronic transcript of *atpH* (e), same as described previously (Drapier et al., 1998), were observed in wild type and iPPR7 lines (Figure 3.15 B, right panel). As shown in Figure 3.15 B (right panel) the relative abundance of the tetracistronic transcript in iPPR7 lines is altered and an increase in this transcript is observed as compared to wild type. iPPR7 line CC19 which has maximum knockdown of PPR7, shows a pronounced effect in increase of this transcript (~30%) compared to N7 (~10%) and N21 (~2%), indicating its relation to the extent of down regulation of PPR7 (Figure 3.15 C, right panel). As a consequence, a decrease in tricistronic transcript can be observed in iPPR7 lines. There was no change observed in the mature *atpH* transcript level in PPR7 RNAi as compared to wild type transcripts.

These observations lead to the point that PPR7 seems to be involved in the processing of polycistronic *atpA* transcript where it has two putative binding sites. The relative increase in tetracistronic transcript in mutants indicates that the processing of tetracistronic transcript to tricistronic transcript is taking place at less frequency as compared to wild type cells.

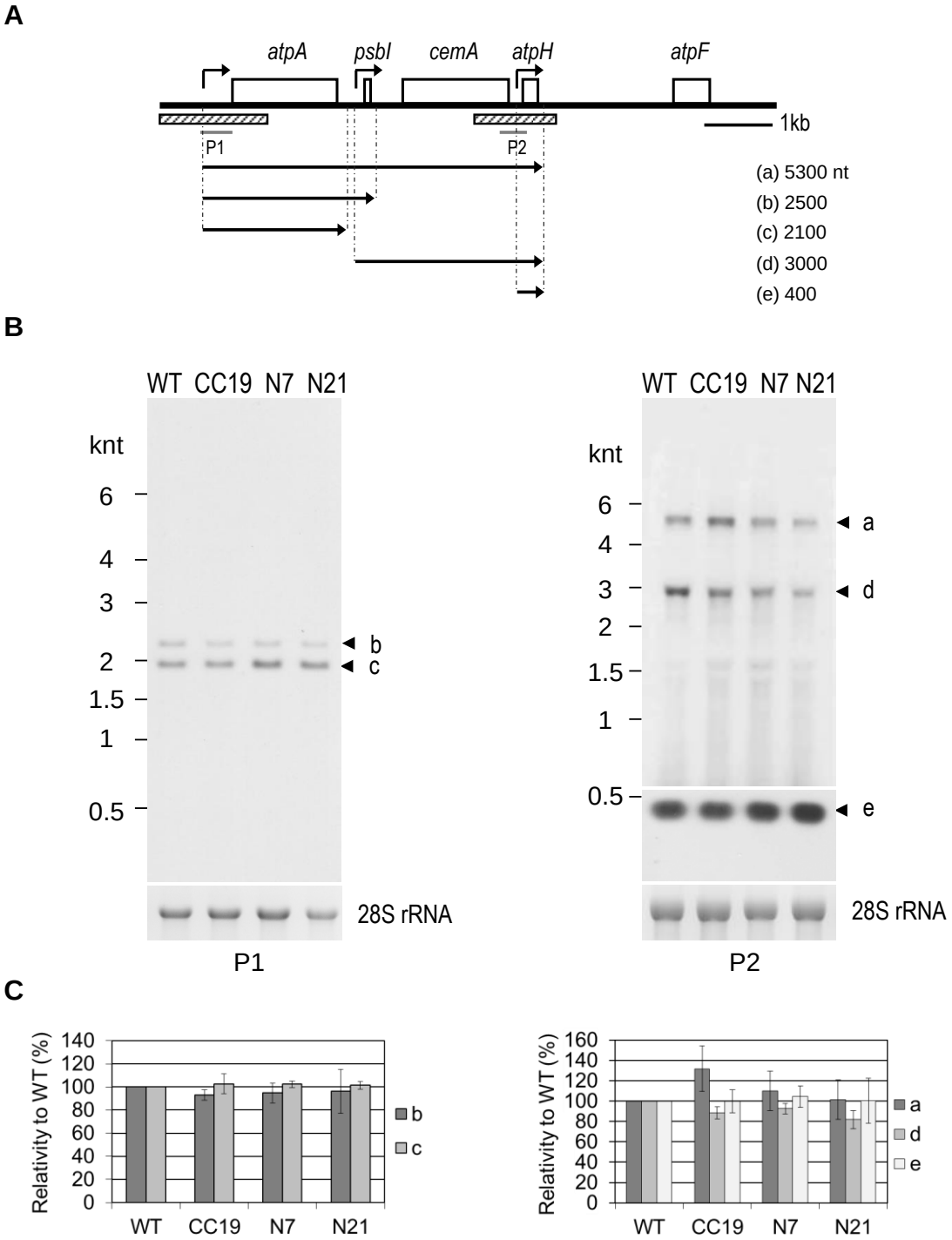


Figure 3.15: Northern analysis of *atpA* and *cemA-atpH* transcripts. A: Schematic diagram of the *atpA gene cluster* in the *C. reinhardtii* chloroplast genome. The open boxes represent the genes. PCR products at which RIP-chip for PPR7 provided hits for binding are indicated as shaded grey boxes. Positions of DIG labelled probes designed are indicated as grey bars. Transcripts and respective sizes are indicated by black arrows according to Drapier et al. (1998). Mapped promoters are indicated by bent arrows (Drapier et al., 1998). **B:** 2 µg (left) and 10 µg (right) of total cellular RNA extracted from wild type (WT) and iPPR7 lines (CC19, N7 and N21) was subjected to Northern blot analysis. DIG labelled probe used to label the transcripts is indicated at the bottom of respective blot. The lower portion of blot probed with P2 is a short time exposure to detect quantifiable signals for highly abundant mature *atpH* transcript. Ethidium bromide-stained RNA gels are shown as loading control (28S rRNAs). The positions of RNA size markers are indicated on the left. **C:** Quantification of signal intensity for the transcripts detected in B, with error bars representing standard deviations from three independent experiments. Signal intensities obtained for the WT were set to 100%.

The *psaA* gene in the chloroplast genome of *C. reinhardtii*, is composed of three exons located at different positions in the chloroplast genome and are transcribed separately as precursors (Kück et al., 1987; Choquet et al., 1988). These precursors assemble and then the introns between exon 1 and 2 as well as between exon 2 and 3, are *trans*-spliced to generate the mature *psaA* mRNA. Intron 2 is formed by joining of 3' flanking region of exon 2 and 5' flanking region of exon 3 only. On the other hand, intron 1 consists of three main components including the precursors of exon 1 and of exon 2, and the short *tscA* RNA (Goldschmidt-Clermont et al., 1991). *tscA* is a non-coding RNA which becomes a part of intron 1 of *psaA* pre-transcript and plays role in *trans*-splicing of intron 1 of the *psaA* transcript (Goldschmidt-Clermont et al., 1991).

tscA was found as the sixth hit of RIP-chip data for PPR7 precipitates. To investigate the probable role of PPR7, a probe at the coding region of *tscA* was generated (Figure 3.16 A, left panel) to carry out RNA gel blot analysis using wild type and iPPR7 lines. The mature *tscA* transcript at the size of about 0.45 knt was detected (Figure 3.16 B, left panel). After normalization with the loading control signals, the quantification data for *tscA* signals reveals that *tscA* transcript level has increased in iPPR7 lines CC19 and N7, as compared to wild type cells. The level of wild type and N21 for *tscA*, are comparable to each other and only negligible difference can be observed (Figure 3.16 C, left panel). These results suggest that PPR7 acts as a suppressor for *tscA* expression, where a decrease in PPR7 levels caused its upregulation. It is also a possibility that PPR7 may be involved in integration of *tscA* to intron 1 of *psaA* pre-transcript.

The last fragment in RIP-chip analysis for PPR7, which provided a hit for being a putative binding site of PPR7, contains the coding and intergenic region of *psaJ* and *atpI* genes. These genes are transcribed in a cluster, which contains *psbJ*, *atpI*, *psaJ* and *rps12* (Figure 3.16 A, right panel). The transcription begins from an upstream promoter of *psbJ* (Liu et al., 1989). The transcript pattern for this gene cluster is complex and not completely defined as for the *atpA* gene cluster (Rymarquis et al., 2006).

To investigate about the putative role of PPR7 in *psaJ-atpI* region, RNA gel blot analysis was performed using a probe covering the 3' region of *psbI* coding sequence, intergenic region of *psaJ-atpI* and coding region of *psaJ* gene. The 2.9 knt tetracistronic (*a*) transcript as described previously (Rymarquis et al., 2006) and some unknown transcripts were observed (Figure 3.16 B, right panel). Upon quantification of 2.9 knt transcript's signal, it was observed to be slightly increased in iPPR7 lines compared to wild type (Figure 3.16 C, right panel). These results show that, same as for *rrnS* and *atpA* gene clusters, PPR7 seems to be involved in processing of *psaJ-atpI* containing cluster as well. Furthermore, the probe also labelled a 0.45 knt transcript, corresponding to the size of mature *psaJ* (b). In this case, the *psaJ* transcript is reduced in abundance in iPPR lines CC19 and N7 (Figure 3.16 B and C, right panel), which might be due to the processing defect of the precursor form.

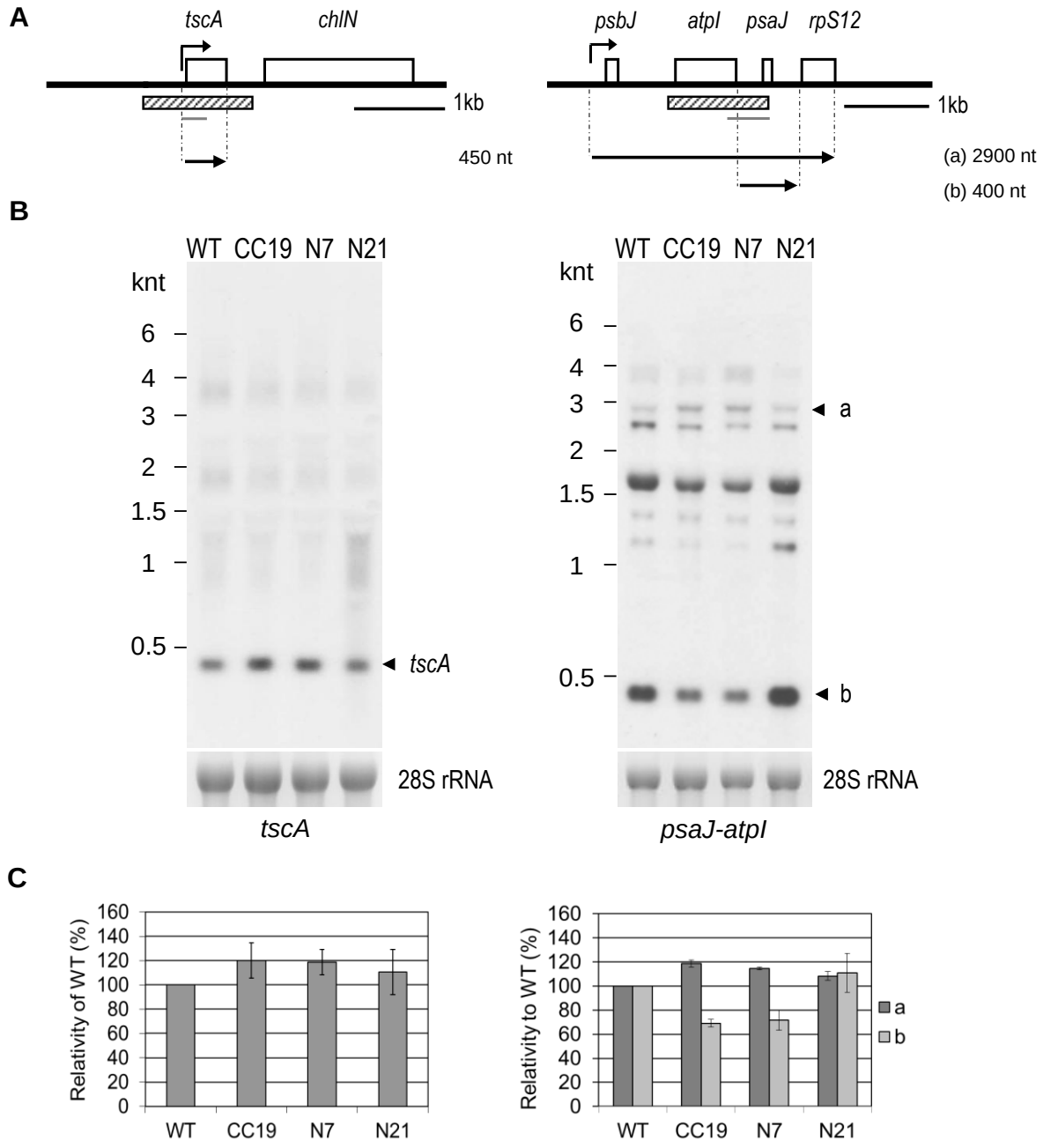


Figure 3.16: Northern analysis of *tscA* and *psaJ-atpI* transcripts. A: Schematic diagrams of *tscA and psaJ-atpI containing gene cluster* in *C. reinhardtii* chloroplast genome. The open boxes represent the genes. PCR fragments at which RIP-chip for PPR7 provided hits for binding are indicated as shaded grey boxes. Positions of DIG labelled probes designed are indicated as grey bars. Transcripts and respective sizes are indicated by black arrows according to Goldschmidt-Clermont et al. (1991) and Liu et al. (1989) for *tscA* and *psaJ-atpI* cluster, respectively. Mapped promoters for *tscA* and *psaJ-atpI* are indicated by bent arrows (Liu et al., 1989; Goldschmidt-Clermont et al., 1991). **B:** 10 µg of total cellular RNA extracted from wild type (WT) and iPPR7 lines (CC19, N7 and N21) was subjected to Northern blot analysis. DIG labelled probes used to label the transcripts are indicated at the bottom of respective blots. Ethidium bromide-stained RNA gels are shown as loading controls (28S rRNAs). The positions of RNA size markers are indicated on the left. **C:** Quantification of signal intensity for the transcripts detected in B, with error bars representing standard deviations from three independent experiments. Signal intensities obtained for the WT were set to 100%.

3.4.3 Photosynthetic stress response of PPR7

According to RIP-chip and RNA gel blot analysis data, PPR7 is found to have a function in RNA metabolism of important photosynthesis-related complex components. To investigate the steady state level of PPR7 in some available photosynthetic mutants, immunoblot analysis was performed using total soluble protein extracts of two wild type strains (XS1 and CC406; section 2.1.6) and four photosynthetic mutants. The mutants for Rubisco include m*rl1,* which lacks MRL1 factor required for the stability of *rbcL* mRNA and hence lacks large subunit of Rubisco and *ΔrbcL* strain which is a deletion mutant for the gene encoding the large subunit of Rubisco. The photosystem mutants include a nuclear mutant *nac2-26*, which lacks the Nac2 factor required for the stability of *psbD* mRNA and hence lacks D2 core subunit of PSII and a PSI mutant *raa1-314B,* which lacks part of the Raa1 factor required for *trans*-splicing of *psaA* precursor transcripts and hence lacks PsaA core subunit of PSI (Boudreau et al., 2000; Merendino et al., 2006; Johnson et al., 2010; see Table 2.3). In addition to CC406, another wild type strain XS1 is included for the comparison because this strain was used to generate the *ΔrbcL* strain (Johnson et al., 2010; Table 2.3). Interestingly it was observed that PPR7 level has increased in the photosynthetic mutants as compared to wild type cells (Figure 3.17 A and B). The maximum increase was observed for the PSII mutant where the increase is ~50% as compared to wild type, while Rubisco and PSI mutants show an increase of ~30% (Figure 3.17 A and B).

This result reveals that PPR7 is up regulated or its stability is enhanced due the photosynthetic stress conditions caused by the absence of any of the photosynthesis-related complex.

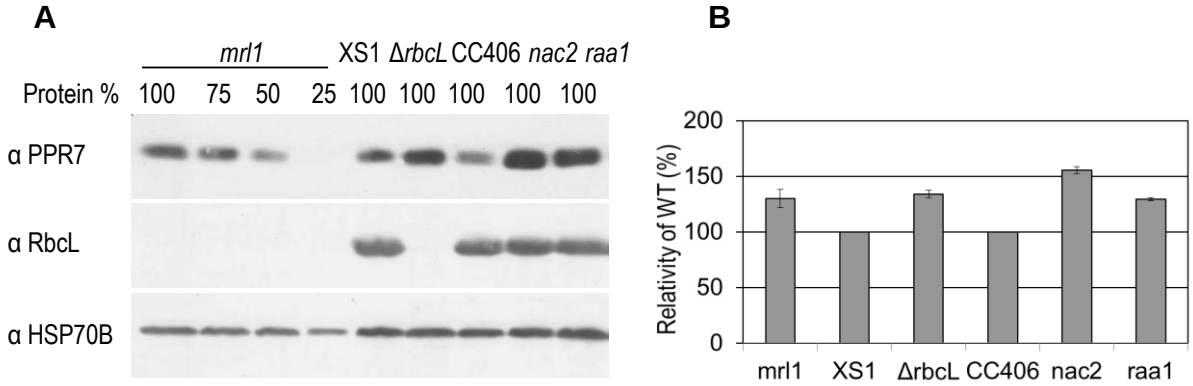


Figure 3.17: Immunoblot analysis of PPR7 in photosynthetic stress. A: Total soluble proteins from *mrl1, ΔrbcL, nac2-26, raa1-314B* and wild types (XS1 and CC406) strains grown under continuous moderate light conditions (30 µE/m^2/s) were fractionated by SDS-PAGE, and blots were probed with antibodies raised against proteins indicated at the right. RbcL antibody is used as control for RbcL mutants. HSP70B served as a loading control for stromal proteins. **B:** Quantification of signal intensity for PPR7 in A. Error bars represents the standard deviations of values from two independent experiments.

4 DISCUSSION

The complex organellar RNA metabolism requires nucleus-encoded factors which play a role in various steps of gene expression. A plethora of these nucleus-encoded RNA binding factors have to be imported into the organelles (Maier et al., 2008). Members of the nucleus-encoded PPR protein family are found to be imported into organelles and bind organellar transcripts. These eukaryote specific proteins form the largest family of RNA binding proteins in plants and are required for organellar RNA metabolism (Small and Peeters, 2000; Lurin et al, 2004). Other than higher plants, the eukaryotic organisms like humans, *Drosophila*, protists and algae contain a small set of PPR proteins. These observations suggest the expansion of PPR proteins during the period of land plant evolution (Lurin et al. 2004; O'Toole et al., 2008; Schmitz-Linneweber and Small 2008).

4.1 PPR proteins in *C. reinhardtii*

In contrast to higher plants, *C. reinhardtii* reveals only a very small number of PPR proteins. The reason for this discrepancy is under debate. Two alternative hypotheses try to explain the expansion of PPR proteins and their fundamental role in plant organelles (compare section 1.4.3.3). According to one hypothesis, the expansion of PPR proteins occurred due to the gain in complexity of organellar RNA metabolism (Schmitz-Linneweber and Small, 2008, section 1.4.3.3). This would suggest that the PPR gene family expanded in concert with the development of advanced multicellular plants during evolution while, *C. reinhardtii* is a unicellular organism and has simple organization. Furthermore, no editing is described for *C. reinhardtii* organellar transcripts, while many PPR proteins are involved in editing events in higher plants (Stern et al., 2010). Therefore, this small set of PPR proteins in *C. reinhardtii* might be sufficient to meet all required functions. The second hypothesis suggests that PPR proteins compensate for the mutations occurring in the organellar genomes and they regain the functionality of the gene products by processing, splicing, editing or by masking an RNase sensitive site (Maier et al., 2008). Thus the expansion of PPR proteins occurred to correct the increased mutations in the organellar genome during evolution. However, it is also possible that another family of RNA binding proteins is performing the necessary functions in *C. reinhardtii* that are fullfilled by PPR proteins in higher plants. One example of such family is represented by OPR proteins (see Introduction section 1.4.2). This protein family consists of more than 100 members in *C. reinhardtii*, which include several characterized members involved in chloroplast RNA processing (see section 1.4.2). In contrast, there is only one OPR protein found in *A. thaliana* where PPR proteins are abundant (O. Vallon, A. Bohne, L. Cerutti, J.D. Rochaix, unpublished data).

The nuclear genome of *C. reinhardtii* encodes for eleven PPR proteins, most of which are predicted to localize to mitochondria or chloroplasts and possess 3-13 tandemly arranged repeats (Figure 3.1, Tables 3.1 and 3.2). All PPR proteins from *C. reinhardtii* belong to the P-subfamily, whereas the PLS subfamily is only prevalent in land plants (Lurin et al., 2004; Schmitz-Linneweber and Small, 2008; section 3.1). The P subfamily constitutes the biggest subgroup of PPR proteins in eukaryotic organisms and is phylogenetically the oldest, which supports the ancient origin of *C. reinhardtii* PPRs (Lurin et al., 2004). In terms of extra C-terminal domains described in higher plants, no E/E+ or DYW domains were found in PPR proteins from *C. reinhardtii* (Figure 3.1, section 1.4.3.1). The E/E+ and DYW domains are mainly found responsible for editing of organellar transcripts in higher plants. To date, no other function for the E/E+ domains has been assigned, but the editing of specific organellar RNAs (Chateigner-Boutin et al., 2008). The absence of editing events in *C. reinhardtii* and the lack of the C-terminal E/E+ and DYW domains in *C. reinhardtii* PPR proteins further strengthens the idea that these domains are responsible for editing reactions to occur (Stern et al., 2010). Furthermore, studies in which domains between two editing PPR factors from *A. thaliana*, namely CRR4 and CRR21, were exchanged, revealed that the E/E+ domain is required for editing while the PPR motifs are more responsible for specific RNA binding (Okuda et al., 2007). Three out of eleven PPR proteins found in *C. reinhardtii* additionally contain a cyclin domain at the C-terminal end, whose function in RNA metabolism is uncertain but might point to a role in cell division cycle.

This study focussed on the detailed functional analysis of one so far uncharachterized PPR protein from *C. reinhardtii*, designated as PPR7. PPR7 contains 4 tandem PPR repeats and an N-terminal transit peptide, which mediates chloroplast localization as shown by GFP import studies and cell fractionation experiments (Figures 3.2 C and 3.2 D). PPR7 thereby, was mainly found in the soluble chloroplast stroma fraction (Fig 3.2 D).

The role of PPR7 in plastidial RNA metabolism is discussed in detail in the following sections.

4.2 PPR7 is part of an RNase sensitive complex

PPR proteins have been previously shown to bind directly to their target RNAs and are often described to be part of HMW ribonucleoprotein complexes (Nakamura et al., 2003; Schmitz-Linneweber et al., 2005; Loiselay et al., 2008; Pfalz et al., 2009; Johnson et al., 2010). Similar findings are obtained for PPR7 in the present study. By using SEC it was observed that PPR7 is part of a HMW complex and the size shifts to smaller fractions upon degradation of RNA with RNases (Figure 3.8). The complex size upon RNase treatment still ranged in a size of up to ~300 kDa which indicates the association of PPR7 with other proteins or a multimeric occurrence of PPR7.

The RNA binding property of PPR7 was further confirmed by *in vitro* RNA binding assays where the recombinant PPR7 protein (His-PPR7) was found to unspecifically bind *in vitro* transcribed RNAs (Figure 3.9). However, even though an intrinsic RNA binding ability of His-PPR7 was observed, the specificity of this protein-RNA interaction remains to be shown. Future experiments for determining the recognition sequence and also the nucleobase specificity of PPR7, could be achieved e.g. by a strategy based on systematic evolution of ligands by exponential enrichment (SELEX) which has been described for HCF152 from *A. thaliana* (Kobayashi et al., 2011). Once exact binding sites of PPR7 have been determined, further studies can be designed involving competition assays using excess amounts of unlabelled specific target RNA to determine additionally an intrinsic specificity of the protein. Similar assays could then be used to assign the amino acids within the PPR repeats which are responsible for RNA binding and specificity. The use of recombinant PPR proteins for identification of RNA binding ligands have been performed for HCF152 and CRR4 from *A. thaliana*, Rf1 from rice and PPR5 and PPR10 from maize (Nakamura et al., 2004; Okuda et al., 2006; Williams-Carrier et al., 2008; Kazama et al., 2008; Pfalz et al., 2009; Prikryl et al., 2011).

To determine the possible RNA binding residues in PPR7, the putative substrate binding surface was modelled. Each PPR repeat is known to form pair of α helices consisting of helix A and B. The consecutive helices stack upon one another to form a superhelix. The concave or inner groove of the helix A forms the binding surface while, the helix B plays a more structural role (Small and Peeters 2000; Kobayashi et al., 2011). The amino acid arrangement from helices A of PPR7 was predicted using helical wheel models. In all the three helices, the upper edge is found to contain highly charged residues, while the lower edge consists of the polar and uncharged residues (Figure 3.3). The amino acids, arginine, serine and lysine are the most commonly found at RNA/protein interfaces (Treger and Westhof, 2001). It is most likely that the arginines and lysines at the upper edge interact with phosphate moieties in the RNA backbone (Williams and Barkan, 2003; Kobayashi et al., 2011). Similar findings were observed for the sequences used from CRP1 and PPR2 proteins in maize and for the PPR2 ortholog in *A. thaliana* where the upper edge of helix A for each PPR repeating unit was found to contain charged residue while at the lower edge, the uncharged residues are found (Williams and Barkan, 2003). Therefore, it would be interesting to analyse RNA binding properties of recombinant PPR7 in which the putative RNA binding residues have been replaced by other amino acids. Furthermore, to gain insight into the physical organization of a PPR tract, full length recombinant PPR7 protein (His-PPR7) was analysed by CD spectroscopy. The analysis revealed that His-PPR7 is a structured protein and contains α helices (section 3.3.2). These results are consistent with the helical-hairpin model for PPR tracts. However, the attempt to crystallize recombinant PPR7 protein was not successful in this study. Three dimensional structure of a “pure” PPR protein which contain tandem arrays of PPR motifs is yet to be reported. However, recently, the three dimensional structure of a human mitochon-

drial polymerase was resolved, which contains two PPR motifs (Ringel et al., 2011). This study provided an insight to the structure of PPR motifs and the previously predicted helix-turn-helix structure of PPR motifs was confirmed. It is possible that crystallization of human mitochondrial polymerase was possible due to special domain architecture while, in a pure PPR protein there are certain features which inhibit the proteins to crystallize. Future experiments to resolve the structure of a PPR protein complexed with an RNA ligand will be of great help to answer about the specific target recognition of PPR proteins.

4.3 PPR7 is associated with multiple chloroplast RNAs

Previous studies investigating the function of PPR proteins reveal an important role in the organellar RNA metabolism and that they are required for a number of processing steps. The analyses of PPR mutants suggest that PPR proteins bind specific target transcripts and affect their processing or accumulation. However, only in few cases, the direct association of PPR proteins to their ligands has been shown *in vivo* (Schmitz-Linneweber et al., 2005, 2006; Gillman et al, 2007; Beick et al., 2008). To identify RNAs associated with the chloroplast localized PPR7 protein, a microarray-based strategy has been used in this study (section 3.4.2.3). The data shown in this thesis supports the association of PPR7 with seven different chloroplast transcripts, namely *rrnS*, *psbH*, *rpoC2*, *rbcL, atpA*, *cemA-atpH*, *tscA* and *atpI-psaJ*. The binding sites thereby included 5' UTRs, intergenic regions and also coding regions (Figure 3.10 B). This is a first example of a PPR protein for which an association with seven different RNAs was revealed. The previous *in vivo* studies have shown the association of a PPR protein at maximum with three transcripts in chloroplasts and mitochondria (Schmitz-Linneweber et al., 2005; Zehrmann et al., 2009; Verbitskiy et al., 2011). The enrichment of the identified PPR7 associated transcripts in PPR7 dependent precipitates were confirmed by RT-PCRs, which provided similar pattern of enrichment, as found in RIP-chip experiments (Figure 3.11). Therefore, a third reason for the relatively low number of PPR proteins in *C. reinhardtii* can be assumed: PPR proteins found in *C. reinhardtii* might represent ancient PPRs which are required to perform multiple tasks, whereas in higher plants, the large number of PPR proteins can accommodate these tasks by acting only on one or two specific transcripts.

The observations based on immunoprecipitation and RIP-chip studies were supplemented by characterization of the identified transcripts in PPR7 deficient mutants via Northern blot analysis. These analyses and roles of PPR7 are discussed in the following sections.

4.3.1 Role of PPR7 as a stability factor

Several PPR proteins have been described to be involved in stabilizing their respective target RNAs (section 1.4.3.3, Table 1.1). It is generally thought that stability factors bind to 5' or 3' UTR regions in a sequence specific manner, and thus protecting the RNA from exonucleolytic cleavage. In agreement with this, the reduced accumulation of transcripts in RNA gel blot analyses indicates PPR7 to act as a stability determinant of *psbH*, *rpoC2* and *rbcL* messages. The level of reduction in the *rpoC2* transcript in PPR7 deficient mutants (Figure 3.13, right panel) was consistent to the knock down levels of PPR7 protein in the RNAi mutants. According to the RIP-chip data, PPR7 interestingly does not bind to the 5' or 3' UTR of the *rpoC*2 message, but rather in the middle of coding region of *rpoC2* transcript which indicates that PPR7 prevents the endonucleolytic cleavage of *rpoC2* transcript. Even though PPR proteins have been characterized as stability factors that generally bind to the 5' UTR of the transcripts and play a role in the stability, so far, no PPR protein involved in stability of a transcript has been reported to bind in the coding region as PPR7 (Yamazaki et al., 2004; Loiseley et al., 2008; Johnson et al., 2010). However, one of the examples for a nucleus-encoded stability factor for chloroplast mRNA comes from MDA1, which acts on the coding region of *atpA* transcript in *C. reinhardtii* and plays a role in determining its stability (Drapier et al., 2002). Furthermore, also all four *psbH* related transcripts are reduced in PPR7 deficient mutants suggesting PPR7 being a stability factor for these transcripts. It is also possible that PPR7 is involved in stabilization of only the longer precursor form (0.9 knt), and hence the decrease in the resulting processed transcripts could be due to a decrease in the longer transcript (Figure 3.13, left panel). To which region of the *psbH* transcript PPR7 exactly binds remains to be determined since the identified target region in the RIP-chip analysis comprises the full length transcript including the coding region as well as 5' and 3' UTRs.

More distinct is the binding region of PPR7 within the 5' UTR of the *rbcL* transcript. Here, the PPR7 effect was found to be more intense for a precursor transcript whose stability was compromised severely due to deficiency of PPR7 (Figure 3.14). This longer precursor transcript was not reported previously but it is completely missing in the *mrl1* mutant. MRL1 is a stability factor for the *rbcL* transcript and its absence results in degradation of *rbcL* transcript soon after transcription (Johnson et al., 2010). This observation argues that it is a precursor *rbcL* transcript which probably extends further downstream. However, there was an almost negligible effect observed on the mature *rbcL* transcript in PPR7 deficient mutants, which might be due to two reasons. One possible reason can be that PPR7 is not strictly required for the mature *rbcL* transcript as another PPR protein MRL1 is performing this function. The second reason can be that, after processing of the available precursor transcript, the stability of mature *rbcL* transcript is enhanced due to a feedback mechanism and they accumulate to near normal levels.

The 3’ UTR of chloroplast transcripts also adds to their stabilization by forming stem-loop structures. The transcripts of *atpB* and *psaB* are shown to contain these structures (Stern et al., 1991; Lee et al., 1996). These stem loop structures are known to block the 3’ to 5’ exonuclease activity that acts after an endonucleolytic cut has primed the 3’ end maturation process (Stern and Kindle, 1993). However, the examples of nucleus-encoded factors acting at the 3’ ends of transcripts are limited. One example of a nucleus-encoded regulatory factor that affects 3’ end formation of several mRNAs in *C. reinhardtii* comes from CRP3 (Levy et al., 1997, 1999). The 5’ and 3’ untranslated regions have been commonly known as controlling RNA stability while, there have been few reports of regulatory elements acting on the coding regions of chloroplast genes that play such a role. In prokaryotes, mammalian cells and in fungal cells however, there exist examples of such factors that act on the coding region of transcripts (Ross, 1995; Hennigan and Jacobson, 1996; Kulkarni and Golden, 1997). The exact mechanism by which PPR proteins act in various steps in gene expression has yet to be reported. Among the characterized PPR proteins, the study of PPR10 in maize provides a model on the mode of action of PPR proteins in protecting the transcripts from exonucleases. According to this study, PPR10 binds at the intergenic regions of *atpI-atpH* and *psaJ-rpl33* having similar sequences and is responsible for the accumulation of processed RNAs by stalling of exonucleases at both 5’ and 3’ termini (Pfalz et al., 2009; see section 1.4.3.3). However, the targets defined for PPR7 in this study, where it helps to stabilize the transcripts are monocistronic. In case of *rpoC2*, the model presented for a stabilizing function of PPR proteins presented for maize PPR5 resembles more the situation for PPR7, where PPR proteins are thought to protect nuclease-sensitive sites simply by masking them (Beick et al., 2008). Also, if PPR7 has a binding site at the coding region of *psbH* transcript, the same model suits to describe a stabilizing function. However, PPR7 has a binding site at the 5’ UTR in case of *rbcL* transcript, which is similar to the function of previously characterized PPR proteins MRL1 and MCA1 in *C. reinhardtii*, where they protect the transcripts from 5’ to 3’ exonucleases (Loiseley et al., 2008; Johnson et al., 2010). In the current study, preliminary analyses for PPR7 suggest a stabilizing role for *rpoC2*, *psbH* and *rbcL* transcripts where *rpoC2* is an essential gene; encoding a subunit of the only RNA polymerase in *C. reinhardtii* chloroplasts (section 1.3.1). The attempt to disrupt *rpoC2* (ORF472) resulted in the production of a heteroplasmic population of chloroplast DNA molecules, containing both the wild type and the inactivated allele revealing its essential function for cell viability (Goldschmidt-Clermont, 1991). If PPR7 is strictly required for the stability of *rpoC2*, then PPR7 is also indispensible for cell viability. Future experiments, like run on transcription assays for iPPR7 lines could be usefull to find out whether the reduction of *rpoC2* has an effect on overall transcription rates or the reduced transcript is still sufficient to code for adequate amounts of ß” subunit of PEP.

Based on the findings, proposed models for the role of PPR7 as a stability factor have been depicted in Figure 4.1. In model A, PPR7 is shown to protect the transcript from endonucleases while, in model B, PPR7 is shown to protect the transcript from 5' to 3' exonucleases.

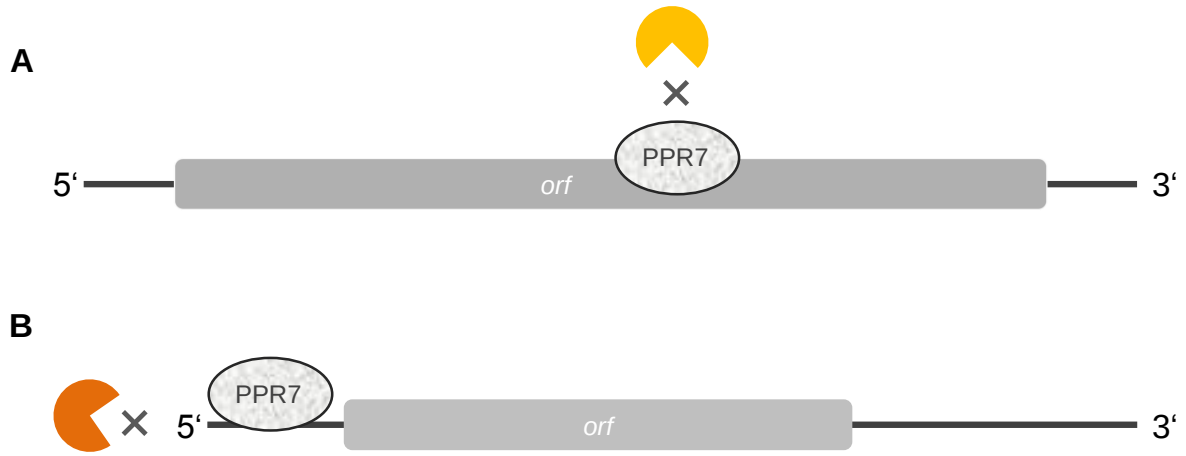


Figure 4.1: Proposed model explaining the stability role of PPR7 in *C. reinhardtii* chloroplast gene expression. A: PPR7 acts as a stability factor by masking an RNase sensitive site from endonucleases (depicted in orange). **B:** PPR7 acts as a stability factor by blocking the exonucleases (depicted in red) from 5' to 3' end. In both cases, it acts as a barrier for the nucleases and protects the transcript from degradation.

4.3.2 Role of PPR7 as a processing factor

The polycistronic transcripts of plastid encoded genes have to be processed to serve as proper templates for translation. These processing steps include endonucleolytic cleavages of intercistronic regions, exonucleolytic trimming of RNA 5' and 3' ends and the splicing of intron sequences (Sugita and Sugiura, 1996; Barkan and Goldschmidt-Clermont, 2000). Once these mature transcripts are produced by processing, the specific stabilization factors are required to ensure their protection from nucleases till their translation. Many of the nucleus-encoded factors are known to play a role in these processes and studies on mutants for such regulatory proteins provide a suitable mean to identify the processing steps taking place in the chloroplast transcripts (Barkan and Goldschmidt-Clermont, 2000; Meierhoff et al., 2003). Upon the transcript analyses of the putative PPR7 associated RNAs in this study; it was found that PPR7 is not only involved in stabilization but also in processing of specific chloroplast transcripts. The deficiency of PPR7 protein resulted in aberrant accumulation of the *rrn*, *cemA-atpH* and *psaJ-atpI* containing tetracistronic primary transcripts from chloroplast.

The *rrn* genes of chloroplasts exist in an operon and resemble their prokaryotic ancestors in the coding sequences (Harris et al., 1994). The primary transcript is processed in many steps to generate the RNAs found in plastid ribosomes. As shown in Figure 3.12, the intergenic

RNA cleavage of the *rrn* precursor, containing the 16S, 7S, and 23S rRNAs, is reduced in PPR7 deficient mutants and the precursor form overaccumulates. According to the studies from Holloway and Herrin (1998), the cleavage of 16S rRNA (*rrnS*) is the first step in processing and then 23S rRNA (*rrnL*) is further processed in different steps. The overaccumulation of the *rrn* precursor was further confirmed by RNA gel blot analysis with a probe specific for the intron region of *rrnL.* This region did not show an association with PPR7 in the RIP-chip analysis but the probe designed at this region is able to detect the *rrn* precursor transcript. A decrease in the accumulation of *rrnS* and *rrnL* partially processed precursor forms was observed, which is probably the cause of reduced cleavage and overaccumulation of the full length *rrn* precursor in PPR7 deficient mutants (Figure 3.12). In *C. reinhardtii*, a mutant termed as *ac20* has been characterized in which the processing of *rrnL* was impaired, while, no processing factor for *rrnS* has been characterized to date (Holloway and Herrin, 1998). However, in maize, the nuclear protein HCF7 was found to be involved in the processing of *rrnS*. Despite the defect in *rrnS* processing in *hcf7* mutants, the *rrnL* precursor was found unaffected (Barkan, 1993). Based on these observations, the results of the *rrn* RNA gel blot analysis indicate that PPR7 is involved in the processing of *rrn* precursor where it has a role in the cleavage while, it seems that PPR7 is not involved in the further processing of partially processed *rrnS* precursor (transcript b, Figure 3.12, left panel), as in case of *rrnS* processing mutant in maize, the *rrnL* transcript accumulated at wild type levels. The mature transcript of *rrnS* was also expected to be reduced but unexpectedly it accumulated to wild type levels. Possibly it is due to the reason that the *rrnS* precursor is processed normally and accumulates stably.

An almost similar function of PPR7 was observed in processing the transcripts of the *atpA* gene cluster as well as the *psaJ-atpI* containing gene cluster. The *atpA* cluster contains four genes, namely *atpA, psbI, cemA,* and *atpH*, where at immediately upstream of the *psbI* and *atpH* genes additional independent promoters are described (Drapier et al., 1998). Interestingly, two putative binding sites of PPR7 are found within this cluster. One lies at the 5’ UTR of *atpA* and the second at the intergenic region of *cemA* and *atpH* genes. The *psaJ-atpI* containing gene cluster also consists of four genes but is transcribed from a single promoter (Liu et al., 1989). The intergenic RNA cleavage of tetracistronic precursor transcripts of both gene clusters is reduced in PPR7 deficient mutants and as a consequence, the precursor form accumulated at higher levels as compared to wild type. This was accompanied by a decrease in tricistronic transcript in case of *atpA* gene cluster (Figure 3.15 right panel). The obvious effect according to these results is a reduced nucleolytic cleavage that processes the tetracistronic transcript to tricistronic transcript. In case of the *atpA* gene cluster, it is not clear, whether the binding of PPR7 at one or at both positions (5’ UTR or intergenic region of *cemA* and *atpH*) contribute to this cleavage event. No effect on the accumulation of mature *atpH* transcript was observed which may be explained by the observation that the predomi-

nant *atpH* transcript is monocistronic and is transcribed from its own promoter (Drapier et al., 1998). In contrast, the mature *psaJ* transcript was reduced in abundance which can not be compensated by its transcription from multiple promoters (Figure 3.16, right panel). PPR proteins having roles in processing of polycistronic transcripts by endonucleolytic cleavage have been characterized in higher plants and as well in moss *P. patens* (Fisk et al., 1999; Lahmy et al., 2000; Hashimoto et al., 2003; Meierhoff et al., 2003; Schmitz-Linneweber et al., 2005; Hattori et al., 2007; see Table 1.1). In *P. patens*, mutant lacking PPR_38 showed a severe phenotype and aberrant RNA cleavage between *clpP* and 5'-*rps12* and also splicing of *clpP* pre-mRNA. Due to the processing defect, the primary dicistronic transcript of *clpP and rps12* overaccumulated to substantial levels in the mutant. This was accompanied by a reduction in the mature form of *clpP* mRNA (Hattori et al., 2007). PPR7 may perform the processing of the target RNAs directly or it may be indirectly involved in site specific cleavage by recruiting other factors having the enzymatic activity. The later model is preferred here because PPR7 consists of simple array of PPR motifs and lacks any extra domain, so it is unlikely that it holds any enzymatic activity for RNA cleavage. From results obtained in this study, it is obvious that PPR7 plays a role in the endonucleolytic cleavage of the unprocessed polycistronic transcripts. No further effect of PPR7 deficiency was observed on the processing of already cleaved small precursor forms. The proposed enzymes for the initial cleavage of precursor RNAs include RNase E, J, CSP41a and CSP41b. RNase E is also known to play a major role in the processing and maturation of rRNAs and tRNAs (reviewed in Stoppel and Meurer, 2011). In *E. coli*, RNase E is found in multiprotein complexes along with PNPase and other nucleases (Carpousis, 2007). As PPR7 is also part of a HMW complex, it can be assumed that it recruits multiprotein complexes containing RNase E or another likewise complex to the precise cleavage site. A similar model has been proposed for the above mentioned P type PPR protein PPR_38 in *P. patens* that it might interact with other factors to facilitate the target mRNA processing (Hattori et al., 2007). Further studies are required to identify the interacting RNases using co-immunoprecipitation and mass spectrometry techniques to understand the mechanism of PPR proteins on RNA cleavage and degradation.

One model for the intercistronic cleavage of polycistronic RNAs, based on maize PPR10 is that the cleavages between genes in precursor transcripts occur randomly and are not specific while, the RNA binding proteins define overlapping 5' and 3' ends together with 5' to 3' and 3' to 5' exonucleolytic trimming of exonucleases (Pfalz et al., 2009). However, in case of PPR7, the overaccumulation of the large precursor transcripts indicates the lack of such random cleavage. Such observations were also made for CRR2 protein from *A. thaliana*, where the random cleavage of transcript was not observed. CRR2 is found to be involved in a specific cleavage at an intergenic region between *rps7-ndhB* transcripts without producing the overlapping termini and the RNA processing appears to be essential for *ndhB* translation (Hashimoto et al., 2003). This protein was first described in mutants known as "chlororespira-

tory reduction mutants," with reduced chloroplast NDH activity. Therefore, the preferred model in this study for PPR7 is to act as sequence specific adaptor and bringing catalytic proteins to the correct site and on the correct transcript.

The findings of RNA gel blot analysis for the *tscA* transcript in PPR7 deficient mutants were more perplexing. *tscA* is a monocistronic non-coding RNA whose accumulation was found to be increased in PPR7 deficient mutants (Figure 3.16, left panel). The possible explanation for such an effect is that PPR7 acts in *psaA* processing since *tscA* RNA is required for correct *trans*-splicing of *psaA* mRNA (section 1.3.1). It is plausible that PPR7 regulates the *tscA* transcript accumulation by recruiting degradation machinery to the transcript. Here, the role of PPR7 in leading another complex to the target site would be in accordance with the working mode for PPR7 hypothesized for *rrn*, *atpA* and *psaJ-atpI* precursor transcripts. Another possibility can be that PPR7 is responsible for integration of *tscA* into intron 1 of the *psaA* pre-transcript. After the successful *trans*-splicing event, *tscA* might be degraded promptly. In contrast, in PPR7 deficient mutants, some of the *tscA* transcripts without PPR7 binding could not integrate into intron 1 which hence decreases *tscA* degradation. However, degradation of the *tscA* transcript after *trans*-splicing is a mere speculation and is not experimentally proved.

The roles of PPR7 discussed here demonstrate that PPR proteins function normally as adapters, which mediate the interaction of specific RNA targets and other proteins with enzymatic activity (Hashimoto et al., 2003; Hattori et al., 2007; Schmitz-Linneweber and Small 2008). Also the stabilization of organellar transcripts is one of the essential functions of PPR proteins where they stabilize the transcripts by masking a nuclease sensitive site (Beick et al., 2008). So in all probabilities, PPR7 doesn't require any additional domains or enzymatic activity and acts passively to fulfil all the functions it is involved in. It is interesting to mention that all known stabilizing PPR proteins are of the P type. They represent the phylogenetically oldest and most widespread PPR type in the eukaryotes (Lurin et al., 2004; see section 1.4.3.1). It is most likely that these simple modes of PPR proteins like stabilization and acting as adapters to facilitate recruitment of catalytic proteins to the correct site on the respective transcript correspond to the ancient PPR protein functions.

According to the data obtained, models for the probable processing role of PPR7 have been depicted in Figure 4.2. In model A, PPR7 is shown to facilitate the intercistronic cleavage while in model B, a possibility of PPR7 being involved in integration of *tscA* into intron 1 of the *psaA* pre-transcript is illustrated.

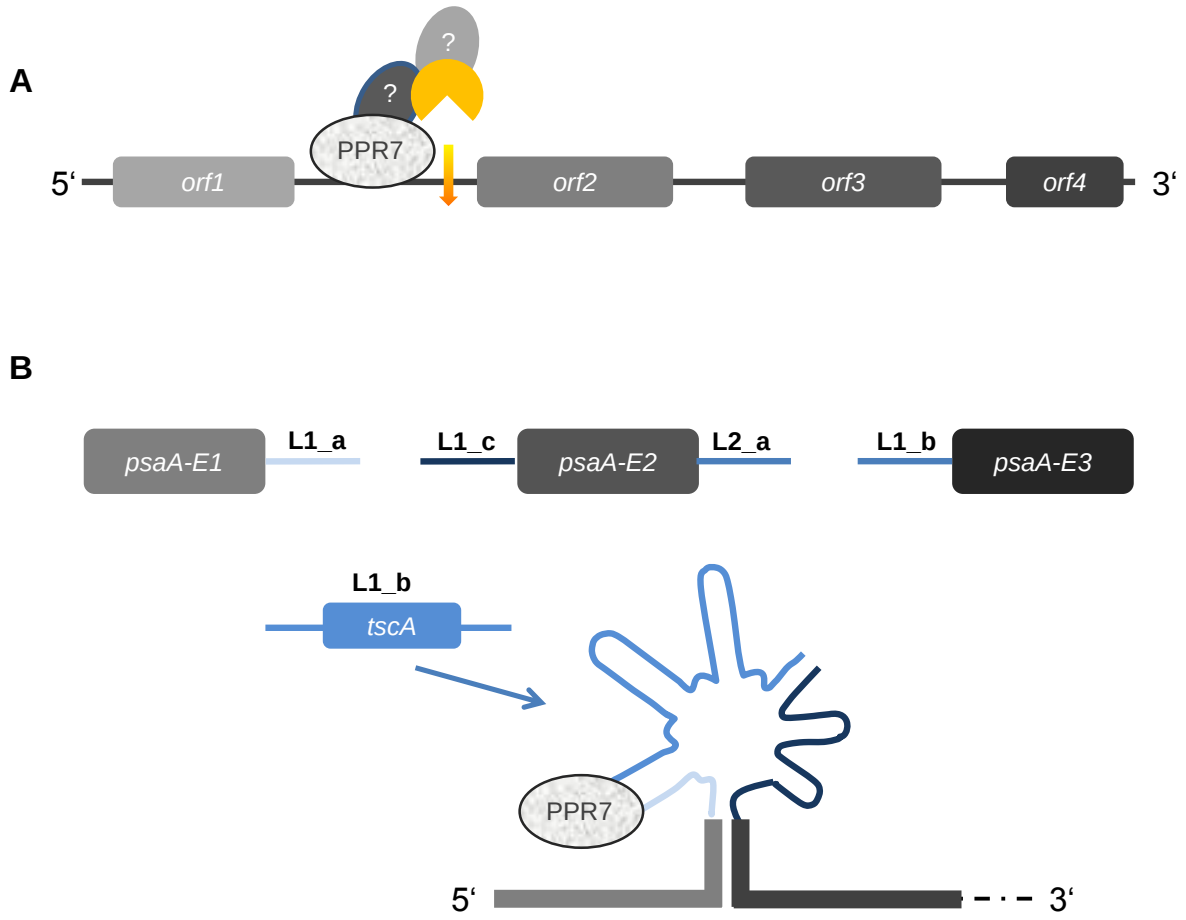


Figure 4.2: Proposed models explaining the role of PPR7 in processing of specific transcripts of *C. reinhardtii* chloroplast gene expression. A: PPR7 facilitates the intercistronic cleavage of *rrn*, *atpA* and *atpI-psaJ* containing polycistronic transcripts by recruiting the catalytic proteins (depicted in orange) to the precise cleavage site. **B:** A possibility for PPR7 is presented that it may be responsible for integration of *tscA* into intron 1 of the *psaA* pre-transcript (Model ‘‘C’’ adapted from Moreira et al., 2012).

4.4 PPR7 deficiency causes a light sensitive phenotype

The lack of a particular PPR protein often results in the lack of expression of specific target organelle transcripts and causes severe phenotype (Fisk et al., 1999; Hashimoto et al., 2003; Meierhoff et al., 2003; Williams and Barkan, 2003; Schmitz-Linneweber 2006; Gillman et al., 2007; Beick et al., 2008; Chateigner-Boutin et al., 2008; Loiselay et al., 2008; Pfalz et al., 2009; Johnson et al., 2010; reviewed in Schmitz-Linneweber and Small, 2008; see Table 1.1). The deficiency of PPR7 caused a complex phenotype including poor growth rate under photoautotrophic conditions and also light sensitivity for the RNAi line CC19, which reveals the strongest reduction in PPR7 protein accumulation (Figures 3.5 and 3.6). RNA gel blot analyses of PPR7 deficient mutants discovered alterations in the stability and processing of the identified target RNAs. Furthermore, immunoblot analysis showed the reduction of PsaA levels in PPR7 deficient mutants. The reduced level of PsaA in PPR7 deficient lines can be

due to the inhibited integration of *tscA* RNA into *psaA* pre-transcript, hence causing light sensitivity, as PSI mutants are highly light sensitive (Rochaix et al., 2000). Another reason which might cause light sensitive phenotype can be the involvement of PPR7 in the processing of *cemA-atpH* region. A similar phenotype was observed for *cemA* mutants where the photoheterotrophic and photoautotrophic growth under low light was unaffected. However, under high light, the mutant did not grow photoautotrophically and a slow growth under photoheterotrophic conditions was observed (Rolland et al., 1997). Furthermore, a cumulative effect of the altered stability and processing of the target RNAs can also be a cause for the phenotype of PPR7 deficient lines.

In a recent study by Ruwe and Schmitz-Linneweber (2011), they identified short RNA sequences (sRNA) from chloroplast of higher plants by using the deep sequencing method. They suggest that a large number of such sRNAs are the sequences known for binding regions of PPR proteins and they accumulate being protected by PPR proteins against nucleases. However, a minority of other RNA binding proteins (RBPs) are also proposed to generate such sRNAs. Regarding the future directions in studying PPR proteins, a similar study on the sRNAs of *C. reinhardtii* chloroplast will provide useful information about the recognition sequences of the RBPs and in particular for the PPR proteins. Furthermore, focussing on the sRNAs corresponding to the remnant sequences from the putative targets of PPR7 will help to narrow down the binding region of PPR7 and also will further enhance the knowledge about the role of PPR7 in RNA stabilization and processing.

5 REFERENCES

Akagi, H., Nakamura, A., Yokozeki-Misono, Y., Inagaki, A., Takahashi, H., Mori, K., and Fujimura, T. (2004). Positional cloning of the rice *Rf-1* gene, a restorer of BT-type cytoplasmic male sterility that encodes a mitochondria-targeting PPR protein. Theor. Appl. Genet. **108,** 1449-1457.

Allen, J.F., de Paula, W.B.M., Puthiyaveetil, S., and Nield, J. (2011). A structural phylogenetic map for chloroplast photosynthesis. Trends Plant Sci. **16,** 645-655.

Andrés, C., Lurin, C., and Small, I.D. (2007). The multifarious roles of PPR proteins in plant mitochondrial gene expression. Physiol. Plant. **129,** 14-22.

Anthonisen, I.L., Salvador, M.L., and Klein, U. (2001). Specific sequence elements in the 5' untranslated regions of *rbcL* and *atpB* gene mRNAs stabilize transcripts in the chloroplast of *Chlamydomonas reinhardtii*. RNA **7,** 1024-1033.

Armbrust, E.V., Berges, J.A., Bowler, C., Green, B.R., Martinez, D., Putnam, N.H., Zhou, S., et al. (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science **306,** 79-86.

Aubourg, S., Boudet, N., Kreis, M., and Lecharny, A. (2000). In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. Plant Mol. Biol. **42,** 603-613.

Auchincloss, A.H., Zerges, W., Perron, K., Girard-Bascou, J., and Rochaix, J.-D. (2002). Characterization of Tbc2, a nucleus-encoded factor specifically required for translation of the chloroplast *psbC* mRNA in *Chlamydomonas reinhardtii*. J. Cell. Biol. **157,** 953-962.

Balczun, C., Bunse, A., Hahn, D., Bennoun, P., Nickelsen, J., and Kück, U. (2005). Two adjacent nuclear genes are required for functional complementation of a chloroplast *trans*-splicing mutant from *Chlamydomonas reinhardtii*. Plant J. **43,** 636-648.

Barkan, A. (1993). Nuclear mutants of maize with defects in chloroplast polysome assembly have altered chloroplast RNA metabolism. Plant Cell **5,** 389-402.

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

Barkan, A., and Goldschmidt-Clermont, M. (2000). Participation of nuclear genes in chloroplast gene expression. Biochimie **82,** 559-572.

Beick, S., Schmitz-Linneweber, C., Williams-Carrier, R., Jensen, B., and Barkan, A. (2008). The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. Mol. Cell. Biol. **28,** 5337.

Bentolila, S., Alfonso, A.A., and Hanson, M.R. (2002). A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proc. Natl. Acad. Sci. USA **99,** 10887-10892.

Biegert, A., Mayer, C., Remmert, M., Söding, J., and Lupas, A.N. (2006). The MPI Bioinformatics Toolkit for protein sequence analysis. Nucleic Acids Res. **34,** W335-W339.

Blatch, G.L., and Lässle, M. (1999). The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. Bioessays **21,** 932-939.

Bohne, A.-V., Schwarz, C., Jalal, A., Ossenbühl, F., and Nickelsen, J. (2009). Control of organellar gene expression in *Chlamydomonas reinhardtii* – future perspectives. Endocyt. Cell Res. **19,** 70-80.

Bollenbach, T., Schuster, G., Portnoy, V., and Stern, D. (2007). Processing, degradation, and polyadenylation of chloroplast transcripts. In Cell and molecular biology of plastids, Bock, R., ed (Springer Berlin / Heidelberg), pp. 175-211.

Boudreau, E., Otis, C., and Turmel, M. (1994). Conserved gene clusters in the highly rearranged chloroplast genomes of *Chlamydomonas moewusii* and *Chlamydomonas reinhardtii*. Plant Mol. Biol. **4,** 585-602.

Boudreau, E., Nickelsen, J., Lemaire, S.D., Ossenbühl, F., and Rochaix, J.D. (2000). The *Nac2* gene of *Chlamydomonas* encodes a chloroplast TPR-like protein involved in *psbD* mRNA stability. EMBO J. **19,** 3366-3376.

Boudreau, E., Turmel, M., Goldschmidt-Clermont, M., Rochaix, J.D., Sivan, S., Michaels, A., and Leu, S. (1997). A large open reading frame (*orf1995*) in the chloroplast DNA of *Chlamydomonas reinhardtii* encodes an essential protein. Mol. Gen. Genet. **253,** 649-653.

Boynton, J.E., Gillham, N.W., Harris, E.H., Hosler, J.P., Johnson, A.M., Jones, A.R., Randolph-Anderson, B.L., et al. (1988). Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. Science **240,** 1534-1538.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72,** 248-254.

Brown, G.G., Formanová, N., Jin, H., Wargachuk, R., Dendy, C., Patil, P., Laforest, M., et al. (2003). The radish *Rfo* restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. Plant J. **35,** 262-272.

Bruick, R.K., and Mayfield, S.P. (1998). Processing of the *psbA* 5′ untranslated region in *Chlamydomonas reinhardtii* depends upon factors mediating ribosome association. J. Cell. Biol. **143,** 1145-1153.

Cai, W., Ji, D., Peng, L., Guo, J., Ma, J., Zou, M., Lu, C., et al. (2009). LPA66 is required for editing *psbF* chloroplast transcripts in *Arabidopsis*. Plant Physiol. **150,** 1260-1271.

Carpousis, A.J. (2007). The RNA degradosome of *Escherichia coli*: an mRNA-degrading machine assembled on RNase E. Annu. Rev. Microbiol. **61,** 71-87.

Chase, C. (2007). Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. Trends Genet. **23,** 81 - 90.

Chateigner-Boutin, A.-L., and Small, I. (2007). A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. Nucleic Acids Res. **35,** e114.

Chateigner-Boutin, A.-L., Ramos-Vega, M., Guevara-Garcia, A., Andres, C., de la Luz Gutierrez-Nava, M., Cantero, A., Delannoy, E., et al. (2008). CLB19, a

pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. Plant J. **56,** 590-602.

Chi, W., Ma, J., Zhang, D., Guo, J., Chen, F., Lu, C., and Zhang, L. (2008). The pentratricopeptide repeat protein DELAYED GREENING1 is involved in the regulation of early chloroplast development and chloroplast gene expression in *Arabidopsis*. Plant Physiol. **147,** 573-584.

Choquet, Y., Zito, F., Wostrikoff, K., and Wollman, F. (2003). Cytochrome *f* translation in *Chlamydomonas* chloroplast is autoregulated by its carboxyl-terminal domain. Plant Cell **15,** 1443 - 1454.

Choquet, Y., Goldschmidt-Clermont, M., Girard-Bascou, J., Kück, U., Bennoun, P., and Rochaix, J.D. (1988). Mutant phenotypes support a *trans*-splicing mechanism for the expression of the tripartite *psaA* gene in the *C. reinhardtii* chloroplast. Cell **52,** 903–913.

Choquet, Y., Stern, D.B., Wostrikoff, K., Kuras, R., Girard-Bascou, J., and Wollman, F.A. (1998). Translation of cytochrome *f* is autoregulated through the 5′ untranslated region of *petA* mRNA in *Chlamydomonas* chloroplasts. Proc. Natl. Acad. Sci. USA **95,** 4380-4385.

Coffin, J.W., Dhillon, R., Ritzel, R.G., and Nargang, F.E. (1997). The *Neurospora crassa Cya-5* nuclear gene encodes a protein with a region of homology to the *Saccharomyces cerevisiae* Pet309 protein and is required in a post-transcriptional step for the expression of the mitochondrially encoded COXI protein. Curr. Genet. **4,** 273-280.

Cole, C., Barber, J.D., and Barton, G.J. (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Res. **36,** W197-W201.

Cushing, D.A., Forsthoefel, N.R., Gestaut, D.R., and Vernon, D.M. (2005). *Arabidopsis emb*175 and other *ppr* knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. Planta **221,** 424-436.

D'Andrea, L.D., and Regan, L. (2003). TPR proteins: the versatile helix. Trends Biochem. Sci. **28,** 655-662.

Das, A.K., Cohen, P.T.W., and Barford, D. (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. EMBO J. **17,** 1192-1199.

Davies, D., and Plaskitt, A. (1971). Genetical and structural analyses of cell-wall formation in *Chlamydomonas reinhardi.* Genet. Res. **17,** 33-43.

de Longevialle, A.F., Meyer, E.H., Andrés, C., Taylor, N.L., Lurin, C., Millar, A.H., and Small, I.D. (2007). The Pentatricopeptide repeat gene OTP43 Is required for *trans*-splicing of the mitochondrial *nad1* intron 1 in *Arabidopsis thaliana*. Plant Cell **19,** 3256-3265.

de Longevialle, A.F., Hendrickson, L., Taylor, N.L., Delannoy, E., Lurin, C., Badger, M., Millar, A.H., 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 J. **56,** 157-168.

del Campo, E.M. (2009). Post-transcriptional control of chloroplast gene expression. Gene Regul. Syst. Biol. **3,** 31-47.

Delannoy, E., Stanley, W.A., Bond, C.S., and Small, I.D. (2007). Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. Biochem. Soc. Trans. **35,** 1643-1647.

Derelle, E., Ferraz, C., Rombauts, S., Rouzé, P., Worden, A.Z., Robbens, S., Partensky, F., et al. (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. Proc. Natl. Acad. Sci. USA **103,** 11647-11652.

Desloire, S., Gherbi, H., Laloui, W., Marhadour, S., Clouet, V., Cattolico, L., Falentin, C., et al. (2003). Identification of the fertility restoration locus, *Rfo*, in radish, as a member of the pentatricopeptide-repeat protein family. EMBO Rep. **4,** 588-594.

Ding, Y.-H., Liu, N.-Y., Tang, Z.-S., Liu, J., and Yang, W.-C. (2006). Arabidopsis GLUTAMINE-RICH PROTEIN23 is essential for early embryogenesis and encodes a novel nuclear PPR motif protein that interacts with RNA polymerase II subunit III. Plant Cell **18,** 815-830.

Doniwa, Y., Ueda, M., Ueta, M., Wada, A., Kadowaki, K.-i., and Tsutsumi, N. (2010). The involvement of a PPR protein of the P subfamily in partial RNA editing of an *Arabidopsis* mitochondrial transcript. Gene **454,** 39-46.

Drager, R.G., Girard-Bascou, J., Choquet, Y., Kindle, K.L., and Stern, D.B. (1998). *In vivo* evidence for 5′→3′ exoribonuclease degradation of an unstable chloroplast mRNA. Plant J. **13,** 85-96.

Drapier, D., Girard-Bascou, J., Stern, D.B., and Wollman, F.A. (2002). A dominant nuclear mutation in *Chlamydomonas* identifies a factor controlling chloroplast mRNA stability by acting on the coding region of the *atpA* transcript. Plant J. **31,** 687-697.

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

Drapier, D., Suzuki, H., Levy, H., Rimbault, B., Kindle, K.L., Stern, D.B., and Wollman, F.A. (1998). The chloroplast *atpA* gene cluster in *Chlamydomonas reinhardtii* - functional analysis of a polycistronic transcription unit. Plant Physiol. **117,** 629-641.

Dron, M., Rahire, M., and Rochaix, J.D. (1982). Sequence of the chloroplast DNA region of *Chlamydomonas reinhardtii* containing the gene of the large subunit of ribulose bisphosphate carboxylase and parts of its flanking genes. J. Mol. Biol. **162,** 775–793.

Eberhard, S., Drapier, D., and Wollman, F.A. (2002). Searching limiting steps in the expression of chloroplast-encoded proteins: relations between gene copy number, transcription, transcript abundance and translation rate in the chloroplast of *Chlamydomonas reinhardtii*. Plant J. **31,** 149-160.

Eberhard, S., Finazzi, G., and Wollman, F.A. (2008). The dynamics of photosynthesis. Annu. Rev. Genet. **42,** 463-515.

Eberhard, S., Loiselay, C., Drapier, D., Bujaldon, S., Girard-Bascou, J., Kuras, R., Choquet, Y., et al. (2011). Dual functions of the nucleus-encoded factor TDA1 in trapping and translation activation of *atpA* transcripts in *Chlamydomonas reinhardtii* chloroplasts. Plant J. **67,** 1055-1066.

Ellis, T.P., Helfenbein, K.G., Tzagoloff, A., and Dieckmann, C.L. (2004). Aep3p stabilizes the mitochondrial bicistronic mRNA encoding subunits 6 and 8 of the H^+-translocating ATP synthase of *Saccharomyces cerevisiae*. J. Biol. Chem. **279,** 15728-15733.

Emanuelsson, O., Nielsen, H., and Heijne, G.V. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci. **8,** 978-984.

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. **300,** 1005-1016.

Eriksson, M., Gardeström, P., and Samuelsson, G. (1995). Isolation, purification and characterization of mitochondria from *Chlamydomonas reinhardtii*. Plant Physiol. **107,** 479-483.

Felder, S., Meierhoff, K., Sane, A.P., Meurer, J., Driemel, C., Plücken, H., Klaff, P., et al. (2001). The nucleus-encoded HCF107 gene of *Arabidopsis* provides a link between intercistronic RNA processing and the accumulation of translation-competent *psbH* transcripts in chloroplasts. Plant Cell **13,** 2127-2141.

Fischer, N., Stampacchia, O., Redding, K., and Rochaix, J. (1996). Selectable marker recycling in the chloroplast. Mol Gen Genet **251,** 373-380.

Fisk, D., Walker, M., and Barkan, A. (1999). Molecular cloning of the maize gene *crp1* reveals similarity between regulators of mitochondrial and chloroplast gene expression. EMBO J. **18,** 2621-2630.

Fong, S.E., and Surzycki, S.J. (1992). Chloroplast RNA polymerase genes of *Chlamydomonas reinhardtii* exhibit an unusual structure and arrangement. Curr. Genet. **21,** 485-497.

Fuhrmann, M., Oertel, W., and Hegemann, P. (1999). A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in *Chlamydomonas reinhardtii.* Plant J. **3,** 353-361.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. **31,** 3784-3788.

Gillman, J.D., Bentolila, S., and Hanson, M.R. (2007). The petunia restorer of fertility protein is part of a large mitochondrial complex that interacts with transcripts of the CMS-associated locus. Plant J. **49,** 217-227.

Goebl, M., and Yanagida, M. (1991). The TPR snap helix: a novel protein repeat motif from mitosis to transcription. Trends Biochem. Sci. **5,** 173-177.

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

Goldschmidt-Clermont, M., Girard-Bascou, J., Choquet, Y., and Rochaix, J.D. (1990). *Trans*-splicing mutants of *Chlamydomonas reinhardtii*. Mol. Gen. Genet. **3,** 417-425.

Goldschmidt-Clermont, M., Choquet, Y., Girard-Bascou, J., Michel, F., Schirmer-Rahire, M., and Rochaix, J.-D. (1991). A small chloroplast RNA may be required for *trans*-splicing in *Chlamydomonas reinhardtii*. Cell **65,** 135-143.

Gothandam, K.M., Kim, E.-S., Cho, H., and Chung, Y.-Y. (2005). OsPPR1, a pentatricopeptide repeat protein of rice is essential for the chloroplast biogenesis. Plant Mol. Biol. **58,** 421-433.

Gray, M.W., and Boer, P.H. (1988). Organization and expression of algal (*Chlamydomonas reinhardtii*) mitochondrial DNA. Philos. Trans. R. Soc. Lond. B Biol. Sci. **319,** 135-147.

Grossman, A., Karpowicz, S., Heinnickel, M., Dewez, D., Hamel, B., Dent, R., Niyogi, K., et al. (2010). Phylogenomic analysis of the *Chlamydomonas* genome unmasks proteins potentially involved in photosynthetic function and regulation. Photosynth. Res. **106,** 3-17.

Grunberg-Manago, M. (1999). Messenger RNA stability and its role in control of gene expression in bacteria and phages. Annu. Rev. Genet. **33,** 193-227.

Hallick, R.B. (1984). Identification and partial DNA sequence of the gene for the alpha-subunit of the ATP synthase complex of *Chlamydomonas reinhardii* chloroplasts. FEBS Lett. **2,** 374-376.

Hammani, K., Okuda, K., Tanz, S.K., Chateigner-Boutin, A.-L., Shikanai, T., and 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,** 3686-3699.

Harris, E.H. (2001). *Chlamydomonas* as a model oganism. Annu. Rev. Plant Physiol. Plant Mol. Biol. **52,** 363-406.

Harris, E.H. (2009). The *Chlamydomonas* source book, 2nd edition, Vol. 1. (San Diego: Academic Press).

Harris, E.H., Boynton, J.E., and Gillham, N.W. (1994). Chloroplast ribosomes and protein synthesis. Microbiol. Rev. **58,** 700-754.

Hashimoto, M., Endo, T., Peltier, G., Tasaka, M., and Shikanai, T. (2003). A nucleus-encoded factor, CRR2, is essential for the expression of chloroplast *ndhB* in *Arabidopsis*. Plant J. **36,** 541-549.

Hattori, M., and Sugita, M. (2009). A moss pentatricopeptide repeat protein binds to the 3' end of plastid *clpP* pre-mRNA and assists with mRNA maturation. FEBS J. **276,** 5860-5869.

Hattori, M., Miyake, H., and Sugita, M. (2007). A pentatricopeptide repeat protein is required for RNA processing of *clpP* Pre-mRNA in moss chloroplasts. J. Biol. Chem. **282,** 10773-10782.

Hennigan, A.N., and Jacobson, A. (1996). Functional mapping of the translation-dependent instability element of yeast *MATalpha1* mRNA. Mol. Cell. Biol. **16,** 3833-3843.

Herrmann, R.G., and Westhoff, P. (2001). Thylakoid biogenesis and dynamics: The result of a complex phylogenetic puzzle. In: Regulation of photosynthesis. (Aro, E.-M. and Andersson, B., eds). Dordrecht: Kluwer Academic Publishers, **11,** 1-28.

Herrmann, R.G., Maier, R.M., and Schmitz-Linneweber, C. (2003). Eukaryotic genome evolution: rearrangement and coevolution of compartmentalized genetic information. Philos. Trans. R. Soc. Lond. B Biol. Sci. **358,** 87-97.

Hirano, T., Kinoshita, N., Morikawa, K., and Yanagida, M. (1990). Snap helix with knob and hole: essential repeats in *S. pombe* nuclear protein *nuc2*$^{+}$. Cell **60,** 319-328.

Holloway, S.P., and Herrin, D.L. (1998). Processing of a composite large subunit rRNA: studies with *Chlamydomonas* mutants deficient in maturation of the 23S-like rRNA. Plant Cell **10,** 1193-1206.

Hong, L., Stevenson, J.K., Roth, W.B., and Hallick, R.B. (1995). *Euglena gracilis* chloroplast *psbB*, *psbT*, *psbH* and *psbN* gene cluster: regulation of *psbB-psbT* pre-mRNA processing. Mol. Gen. Genet. **2,** 180-188.

Johnson, C.H., and Schmidt, G.W. (1993). The *psbB* gene cluster of the *Chlamydomonas reinhardtii* chloroplast: sequence and transcriptional analyses of *psbN* and *psbH*. Plant Mol. Biol. **4,** 645-658.

Johnson, X. (2011). Manipulating RuBisCO accumulation in the green alga, *Chlamydomonas reinhardtii*. Plant Mol. Biol. **76,** 397-405.

Johnson, X., Wostrikoff, K., Finazzi, G., Kuras, R., Schwarz, C., Bujaldon, S., Nickelsen, J., et al. (2010). MRL1, a conserved pentatricopeptide repeat protein, is required for stabilization of *rbcL* mRNA in *Chlamydomonas* and *Arabidopsis*. Plant Cell **22,** 234-248.

Jordan, D., Mace, E., Henzell, R., Klein, P., and Klein, R. (2010). Molecular mapping and candidate gene identification of the *Rf2* gene for pollen fertility restoration in sorghum [*Sorghum bicolor* (L.) Moench]. Theor. Appl. Genet. **120,** 1279-1287.

Kato, H., Tezuka, K., Feng, Y.Y., Kawamoto, T., Takahashi, H., Mori, K., and Akagi, H. (2007). Structural diversity and evolution of the *Rf-1* locus in the genus *Oryza*. Heredity **99,** 516-524.

Kazama, T., and Toriyama, K. (2003). A pentatricopeptide repeat-containing gene that promotes the processing of aberrant *atp6* RNA of cytoplasmic male-sterile rice. FEBS Lett. **544,** 99-102.

Kazama, T., Nakamura, T., Watanabe, M., Sugita, M., and Toriyama, K. (2008). Suppression mechanism of mitochondrial ORF79 accumulation by Rf1 protein in BT-type cytoplasmic male sterile rice. Plant J. **55,** 619-628.

Kelly, S.M., Jess, T.J., and Price, N.C. (2005). How to study proteins by circular dichroism. Biochim. Biophys. Acta **1751,** 119-139.

Kindle, K.L. (1990). High-frequency nuclear transformation of *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. USA **87,** 1228-1232.

Klein, R., Klein, P., Mullet, J., Minx, P., Rooney, W., and Schertz, K. (2005). Fertility restorer locus *Rf1* of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. Theor. Appl. Genet. **111,** 994-1012.

Klinkert, B., Ossenbühl, F., Sikorski, M., Berry, S., Eichacker, L., and Nickelsen, J. (2004). PratA, a periplasmic tetratricopeptide repeat protein involved in biogenesis of photosystem II in *Synechocystis* sp. PCC 6803. J. Biol. Chem. **279,** 44639-44644.

Kobayashi, K., Kawabata, M., Hisano, K., Kazama, T., Matsuoka, K., Sugita, M., and Nakamura, T. (2011). Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein. Nucleic Acids Res.**,** DOI 10.1093/nar/gkr1084.

Kocabek, T., Repkova, J., Dudova, M., Hoyerova, K., and Vrba, L. (2006). Isolation and characterization of a novel semi-lethal *Arabidopsis thaliana* mutant of gene for pentatricopeptide (PPR) repeat-containing protein. Genetica **128,** 395-407.

Koizuka, N., Imai, R., Fujimoto, H., Hayakawa, T., Kimura, Y., Kohno-Murase, J., Sakai, T., et al. (2003). Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosena radish. Plant J. **34,** 407-415.

Komori, T., Ohta, S., Murai, N., Takakura, Y., Kuraya, Y., Suzuki, S., Hiei, Y., et al. (2004). Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). Plant J. **37,** 315-325.

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

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

Kück, U., Choquet, Y., Schneider, M., Dron, M., and Bennoun, P. (1987). Structural and transcription analysis of two homologous genes for the P700 chlorophyll a-apoproteins in *Chlamydomonas reinhardii*: evidence for *in vivo trans*-splicing. EMBO J. **6,** 2185-2195.

Kulkarni, R.D., and Golden, S.S. (1997). mRNA stability is regulated by a coding-region element and the unique 5′ untranslated leader sequences of the three *Synechococcus psbA* transcripts. Mol. Microbiol. **24,** 1131-1142.

Kunstner, P., Guardiola, A., Takahashi, Y., and Rochaix, J.-D. (1995). A mutant strain of *Chlamydomonas reinhardtii* lacking the chloroplast photosystem II *psbI* gene grows photoautotrophically. J. Biol. Chem. **270,** 9651-9654.

Kutschera, U., and Niklas, K.J. (2005). Endosymbiosis, cell evolution, and speciation. Theory Biosci. **124,** 1-24.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227,** 680-685.

Lahmy, S., Barneche, F., Derancourt, J., Filipowicz, W., Delseny, M., and Echeverria, M. (2000). A chloroplastic RNA-binding protein is a new member of the PPR family. FEBS Lett. **2-3,** 255-260.

Lau, K.W.K., Ren, J., and Wu, M. (2000). Redox modulation of chloroplast DNA replication in *Chlamydomonas reinhardtii*. Antioxid. Redox Signal. **2,** 529–535.

Lee, H., Bingham, S., and Webber, A. (1996). Function of 3' non-coding sequences and stop codon usage in expression of the chloroplast *psaB* gene in *Chlamydomonas reinhardtii*. Plant Mol. Biol. **2,** 337-354.

Lemaire, C., and Wollman, F.A. (1989). The chloroplast ATP synthase in *Chlamydomonas reinhardtii.* II. Biochemical studies on its biogenesis using mutants defective in photophosphorylation. J. Biol. Chem. **264,** 10235-10242.

Leu, S., Schlesinger, J., Michaels, A., and Shavit, N. (1992). Complete DNA sequence of the *Chlamydomonas reinhardtii* chloroplast *atpA* gene. Plant Mol. Biol. **3,** 613-616.

Levy, H., Kindle, K.L., and Stern, D.B. (1997). A nuclear mutation that affects the 3' processing of several mRNAs in *Chlamydomonas* chloroplasts. Plant Cell **9,** 825-836.

Levy, H., Kindle, K.L., and Stern, D.B. (1999). Target and specificity of a nuclear gene product that participates in mRNA 3′-end formation in *Chlamydomonas* chloroplasts. J. Biol. Chem. **274,** 35955-35962.

Liere, K., and Börner, T. (2007). Transcription and transcriptional regulation in plastids. In Cell and molecular biology of plastids, Bock, R., ed (Springer Berlin / Heidelberg), pp. 121-174.

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

Liu, X.Q., Gillham, N.W., and Boynton, J.E. (1989). Chloroplast ribosomal protein gene *rps12* of *Chlamydomonas reinhardtii.* Wild-type sequence, mutation to streptomycin resistance and dependence, and function in *Escherichia coli*. J. Biol. Chem. **264,** 16100-16108.

Loiselay, C., Gumpel, N.J., Girard-Bascou, J., Watson, A.T., Purton, S., Wollman, F.A., and Choquet, Y. (2008). Molecular identification and dunction of *cis*- and *trans*-acting determinants for *petA* transcript stability in *Chlamydomonas reinhardtii* chloroplasts. Mol. Cell. Biol. **28,** 5529-5542.

Lurin, C., Andrés, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyère, C., Caboche, M., et al. (2004). Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell **16,** 2089-2103.

Maier, U., Bozarth, A., Funk, H., Zauner, S., Rensing, S., Schmitz-Linneweber, C., Börner, T., et al. (2008). Complex chloroplast RNA metabolism: just debugging the genetic programme? BMC Biol. **6,** 36.

Majeran, W., Olive, J., Drapier, D., Vallon, O., and Wollman, F.A. (2001). The light sensitivity of ATP synthase mutants of *Chlamydomonas reinhardtii*. Plant Physiol. **1,** 421-433.

Mancebo, R., Zhou, X., Shillinglaw, W., Henzel, W., and Macdonald, P.M. (2001). BSF binds specifically to the *bicoid* mRNA 3′ untranslated region and contributes to stabilization of *bicoid* mRNA. Mol. Cell. Biol. **21,** 3462-3471.

Manthey, G.M., and McEwen, J.E. (1995). The product of the nuclear gene *PET309* is required for translation of mature mRNA and stability or production of intron-containing RNAs derived from the mitochondrial *COX1* locus of *Saccharomyces cerevisiae*. EMBO J. **16,** 4031-4043.

Maul, J.E., Lilly, J.W., Cui, L., dePamphilis, C.W., Miller, W., Harris, E.H., and Stern, D.B. (2002). The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats. Plant Cell **14,** 2659-2679.

McFadden, G.I. (2001). Chloroplast origin and integration. Plant Physiol. **125,** 50-53.

McFadden, G.I., and van Dooren, G.G. (2004). Evolution: red algal genome affirms a common origin of all plastids. Curr. Biol. **14,** R514-516.

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

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., et al. (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science **318,** 245-250.

Merendino, L., Perron, K., Rahire, M., Howald, I., Rochaix, J.-D., and Goldschmidt-Clermont, M. (2006). A novel multifunctional factor involved in *trans*-splicing of chloroplast introns in *Chlamydomonas*. Nucleic Acids Res. **34,** 262-274.

Mereschkowsky, C. (1905). Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. Biol. Centralbl. **25,** 593-604.

Meskauskiene, R., Nater, M., Goslings, D., Kessler, F., op den Camp, R., and Apel, K. (2001). FLU: A negative regulator of chlorophyll biosynthesis in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA **98,** 12826-12831.

Minai, L., Wostrikoff, K., Wollman, F., and Choquet, Y. (2006). Chloroplast biogenesis of photosystem II cores involves a series of assembly-controlled steps that regulate translation. Plant Cell **18,** 159 - 175.

Monod, C., Goldschmidt-Clermont, M., and Rochaix, J.D. (1992). Accumulation of chloroplast *psbB* RNA requires a nuclear factor in *Chlamydomonas reinhardtii.* Mol. Gen. Genet. **3,** 449-459.

Moreira, S., Breton, S., and Burger, G. (2012). Unscrambling genetic information at the RNA level. Wiley Interdiscip. Rev. RNA **3,** 213-228.

Moreno, J.I., Buie, K.S., Price, R.E., and Piva, M.A. (2009). Ccm1p/Ygr150cp, a pentatricopeptide repeat protein, is essential to remove the fourth intron of both *COB* and *COX1* pre-mRNAs in *Saccharomyces cerevisiae*. Curr. Genet. **55,** 475-484.

Nakamura, T., Meierhoff, K., Westhoff, P., and Schuster, G. (2003). RNA-binding properties of HCF152, an *Arabidopsis* PPR protein involved in the processing of chloroplast RNA. Eur. J. Biochem. **270,** 4070-4081.

Nakamura, T., Schuster, G., Sugiura, M., and Sugita, M. (2004). Chloroplast RNA-binding and pentatricopeptide repeat proteins. Biochem. Soc. Trans. **32,** 571-574.

Neupert, J., Karcher, D., and Bock, R. (2009). Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. Plant J. **57,** 1140-1150.

Nickelsen, J., and Kück, U. (2000). The unicellular green alga *Chlamydomonas reinhardtii* as an experimental system to study chloroplast RNA metabolism. Naturwissenschaften **87,** 97-107.

Nickelsen, J., van Dillewijn, J., Rahire, M., and Rochaix, J.-D. (1994). Determinants for stability of the chloroplast *psbD* RNA are located within is short leader region in *Chlamydomonas reinhardtii*. EMBO J. **13,** 3182-3191.

O'Connor, H.E., Ruffle, S.V., Cain, A.J., Deak, Z., Vass, I., Nugent, J.H.A., and Purton, S. (1998). The 9-kDa phosphoprotein of photosystem II. Generation and characterisation of *Chlamydomonas* mutants lacking PSII-H and a site-directed mutant lacking the phosphorylation site. Biochim. Biophys. Acta **1364,** 63-72.

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

Okuda, K., Nakamura, T., Sugita, M., Shimizu, T., and Shikanai, T. (2006). A pentatricopeptide repeat protein is a site-recognition factor in chloroplast RNA editing. J. Biol. Chem. **281,** 37661.

Okuda, K., Myouga, F., Motohashi, R., Shinozaki, K., and Shikanai, T. (2007). Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. Proc. Natl. Acad. Sci. USA **104,** 8178-8183.

Okuda, K., Chateigner-Boutin, A.L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F., Motohashi, R., 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,** 146-156.

Okuda, K., Hammani, K., Tanz, S.K., Peng, L., Fukao, Y., Myouga, F., Motohashi, R., et al. (2010). The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid *ndhB* and *ndhG* transcripts. Plant J. **61,** 339-349.

Ossenbühl, F., and Nickelsen, J. (2000). *Cis*- and *trans*-acting determinants for translation of *psbD* mRNA in *Chlamydomonas reinhardtii*. Mol. Cell. Biol. **20,** 8134-8142.

Park, S., Khamai, P., Garcia-Cerdan, J.G., and Melis, A. (2007). REP27, a tetratricopeptide repeat nuclear-encoded and chloroplast-localized protein, functions in D1/32-kD reaction center protein turnover and photosystem II repair from photodamage. Plant Physiol. **143,** 1547-1560.

Peng, L., Ma, J., Chi, W., Guo, J., Zhu, S., Lu, Q., Lu, C., et al. (2006). LOW PSII ACCUMULATION1 is involved in efficient assembly of photosystem II in *Arabidopsis thaliana*. Plant Cell **18,** 955-969.

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. **28,** 2042-2052.

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

Prasad, A.M., Sivanandan, C., Resminath, R., Thakare, D.R., Bhat, S.R., and Srinivasan. (2005). Cloning and characterization of a pentatricopeptide protein encoding gene

(*LOJ*) that is specifically expressed in lateral organ junctions in *Arabidopsis thaliana*. Gene **353,** 67-79.

Prikryl, J., Rojas, M., Schuster, G., and Barkan, A. (2011). Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. Proc. Natl. Acad. Sci. USA **108,** 415-420.

Puchta, O., Lubas, M., Lipinski, K.A., Piatkowski, J., Malecki, M., and Golik, P. (2010). *DMR1* (*CCM1*/YGR150C) of *Saccharomyces cerevisiae* encodes an RNA-binding protein from the pentatricopeptide repeat family required for the maintenance of the mitochondrial 15S ribosomal RNA. Genetics **184,** 959-973.

Pusnik, M., Small, I., Read, L.K., Fabbro, T., and Schneider, A. (2007). Pentatricopeptide repeat proteins in *Trypanosoma brucei* function in mitochondrial ribosomes. Mol. Cell. Biol. **27,** 6876-6888.

Remacle, C., Cardol, P., Coosemans, N., Gaisne, M., and Bonnefoy, N. (2006). High-efficiency biolistic transformation of *Chlamydomonas* mitochondria can be used to insert mutations in complex I genes. Proc. Natl. Acad. Sci. USA **103,** 4771-4776.

Richly, E., and Leister, D. (2004). An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of *Arabidopsis* and rice. Gene **329,** 11-16.

Ringel, R., Sologub, M., Morozov, Y.I., Litonin, D., Cramer, P., and Temiakov, D. (2011). Structure of human mitochondrial RNA polymerase. Nature **478,** 269-273.

Rivals, E., Bruyère, C., Toffano-Nioche, C., and Lecharny, A. (2006). Formation of the *Arabidopsis* pentatricopeptide repeat family. Plant Physiol. **141,** 825-839.

Rochaix, J.-D. (2004). Genetics of the biogenesis and dynamics of the photosynthetic machinery in eukaryotes. Plant Cell **16,** 1650-1660.

Rochaix, J.-D., Fischer, N., and Hippler, M. (2000). Chloroplast site-directed mutagenesis of photosystem I in *Chlamydomonas*: Electron transfer reactions and light sensitivity. Biochimie **82,** 635-645.

Rochaix, J.D. (1995). *Chlamydomonas reinhardtii* as the photosynthetic yeast. Annu. Rev. Genet. **29,** 209-230.

Rochaix, J.D. (1996). Post-transcriptional regulation of chloroplast gene expression in *Chlamydomonas reinhardtii*. Plant Mol. Biol. **32,** 327-341.

Rohr, J., Sarkar, N., Balenger, S., Jeong, B.-R., and Cerutti, H. (2004). Tandem inverted repeat system for selection of effective transgenic RNAi strains in *Chlamydomonas*. Plant J. **40,** 611-621.

Rolland, N., Dorne, A.-J., Amoroso, G., Sultemeyer, D.F., Joyard, J., and Rochaix, J.-D. (1997). Disruption of the plastid *ycf10* open reading frame affects uptake of inorganic carbon in the chloroplast of *Chlamydomonas*. EMBO J. **16,** 6713-6726.

Ross, J. (1995). mRNA stability in mammalian cells. Microbiol. Rev. **59,** 423-450.

Ruckle, M.E., and Larkin, R.M. (2009). Plastid signals that affect photomorphogenesis in *Arabidopsis thaliana* are dependent on GENOMES UNCOUPLED 1 and cryptochrome 1. New Phytol. **182,** 367-379.

Rüdinger, M., Polsakiewicz, M., and Knoop, V. (2008). Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in Jungermanniid but not in Marchantiid Liverworts. Mol. Biol. Evol. **25,** 1405-1414.

Rupprecht, J. (2009). From systems biology to fuel--*Chlamydomonas reinhardtii* as a model for a systems biology approach to improve biohydrogen production. J. Biotechnol. **142,** 10-20.

Rymarquis, L.A., Higgs, D.C., and Stern, D.B. (2006). Nuclear suppressors define three factors that participate in both 5′ and 3′ end processing of mRNAs in *Chlamydomonas* chloroplasts. Plant J. **46,** 448-461.

Sager, R., and Granick, S. (1953). Nutritional studies with *Chlamydomanas reinhardi*. Ann. N. Y. Acad. Sci. **56,** 831-838.

Salone, V., Rüdinger, M., Polsakiewicz, M., Hoffmann, B., Groth-Malonek, M., Szurek, B., Small, I., et al. (2007). A hypothesis on the identification of the editing enzyme in plant organelles. FEBS Lett. **581,** 4132-4138.

Salvador, M.L., Klein, U., and Bogorad, L. (1993). 5' sequences are important positive and negative determinants of the longevity of *Chlamydomonas* chloroplast gene transcripts. Proc. Natl. Acad. Sci. USA **90,** 1556-1560.

Salvador, M.L., Suay, L., Anthonisen, I.L., and Klein, U. (2004). Changes in the 5'-untranslated region of the *rbcL* gene accelerate transcript degradation more than 50-fold in the chloroplast of *Chlamydomonas reinhardtii*. Curr. Genet. **45,** 176-182.

Sambrook, J., and Russell, D. (2001). Molecular Cloning: A laboratory manual. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., and Tabata, S. (1999). Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res. **6,** 283-290.

Schein, A., Sheffy-Levin, S., Glaser, F., and 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 **14,** 1057-1068.

Schmitz-Linneweber, C., and Small, I. (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci. **13,** 663-670.

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

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

Schottkowski, M., Ratke, J., Oster, U., Nowaczyk, M., and Nickelsen, J. (2009a). Pitt, a novel tetratricopeptide repeat protein involved in light-dependent chlorophyll biosynthesis and thylakoid membrane biogenesis in *Synechocystis* sp. PCC 6803. Mol. Plant **2,** 1289-1297.

Schottkowski, M., Gkalympoudis, S., Tzekova, N., Stelljes, C., Schünemann, D., Ankele, E., and Nickelsen, J. (2009b). Interaction of the periplasmic PratA factor and the PsbA (D1) protein during biogenesis of photosystem II in *Synechocystis* sp. PCC 6803. J. Biol. Chem. **284,** 1813-1819.

Schwarz, C., Elles, I., Kortmann, J., Piotrowski, M., and Nickelsen, J. (2007). Synthesis of the D2 protein of photosystem II in *Chlamydomonas* is controlled by a high molecular mass complex containing the RNA stabilization factor Nac2 and the translational activator RBP40. Plant Cell **19,** 3627-3639.

Shikanai, T. (2006). RNA editing in plant organelles: machinery, physiological function and evolution. Cell. Mol. Life Sci. **63,** 698-708.

Sikorski, R.S., Boguski, M.S., Goebl, M., and Hieter, P. (1990). A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. Cell **60,** 307-317.

Simpson, C.L., and Stern, D.B. (2002). The treasure trove of algal chloroplast genomes. Surprises in architecture and gene content, and their functional implications. Plant Physiol. **129,** 957-966.

Small, I., Peeters, N., Legeai, F., and Lurin, C. (2004). Predotar: A tool for rapidly screening proteomes for *N*-terminal targeting sequences. Proteomics **4,** 1581-1590.

Small, I.D., and Peeters, N. (2000). The PPR motif - a TPR-related motif prevalent in plant organellar proteins. Trends Biochem. Sci. **25,** 45-47.

Smith, A.C., and Purton, S. (2002). The transcriptional apparatus of algal plastids. Euro. J. Phycol. **37,** 301-311.

Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. **32,** W309-W312.

Stern, D.B., and Kindle, K.L. (1993). 3'end maturation of the *Chlamydomonas reinhardtii* chloroplast *atpB* mRNA is a two-step process. Mol. Cell. Biol. **13,** 2277-2285.

Stern, D.B., Radwanski, E.R., and Kindle, K.L. (1991). A 3' stem/loop structure of the *Chlamydomonas* chloroplast *atpB* gene regulates mRNA accumulation *in vivo*. Plant Cell **3,** 285-297.

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

Stoppel, R., and Meurer, J. (2011). The cutting crew - ribonucleases are key players in the control of plastid gene expression. J. Exp. Bot.**,** DOI 10.1093/jxb/err1401.

Sugita, M., and Sugiura, M. (1996). Regulation of gene expresion in chloroplast of higher plants. Plant Mol. Biol. **32,** 315-326.

Sugiura, M. (1992). The chloroplast genome. Plant Mol. Biol. **19,** 149-168.

Summer, E.J., Schmid, V., Bruns, B.U., and Schmidt, G.W. (1997). Requirement for the H phosphoprotein in photosystem II of *Chlamydomonas reinhardtii*. Plant Physiol. **113,** 1359-1368.

Takenaka, M. (2010). MEF9, an E-subclass pentatricopeptide repeat protein, is required for an RNA editing event in the *nad7* transcript in mitochondria of *Arabidopsis*. Plant Physiol. **152,** 939-947.

Tasaki, E., Hattori, M., and Sugita, M. (2010). The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial *ccmFc* transcript. Plant J. **62,** 560-570.

Tavares-Carreón, F., Camacho-Villasana, Y., Zamudio-Ochoa, A., Shingú-Vázquez, M., Torres-Larios, A., and Pérez-Martínez, X. (2008). The pentatricopeptide repeats present in Pet309 are necessary for translation but not for stability of the mitochondrial *COX1* mRNA in yeast. J. Biol. Chem. **283,** 1472-1479.

Tillich, M., Beick, S., and Schmitz-Linneweber, C. (2010). Chloroplast RNA-binding proteins: Repair and regulation of chloroplast transcripts. RNA Biol. **7,** 172-178.

Timmis, J.N., Ayliffe, M.A., Huang, C.Y., and Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. **5,** 123-135.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. Proc. Natl. Acad. Sci. USA **76,** 4350-4354.

Treger, M., and Westhof, E. (2001). Statistical analysis of atomic contacts at RNA–protein interfaces. J. Mol. Recog. **14,** 199-214.

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

Tsuchiya, N., Fukuda, H., Sugimura, T., Nagao, M., and Nakagama, H. (2002). LRP130, a protein containing nine pentatricopeptide repeat motifs, interacts with a single-stranded cytosine-rich sequence of mouse hypervariable minisatellite Pc-1. Eur. J. Biochem. **269,** 2927-2933.

Uchida, M., Ohtani, S., Ichinose, M., Sugita, C., and Sugita, M. (2011). The PPR-DYW proteins are required for RNA editing of *rps14, cox1* and *nad5* transcripts in *Physcomitrella patens* mitochondria. FEBS Lett. **585,** 2367-2371.

Uyttewaal, M., Mireau, H., Rurek, M., Hammani, K., Arnal, N., Quadrado, M., and Giegé, P. (2008). PPR336 is associated with polysomes in plant mitochondria. J. Mol. Biol. **375,** 626-636.

Vaistij, F.E., Goldschmidt-Clermont, M., Wostrikoff, K., and Rochaix, J.-D. (2000a). Stability determinants in the chloroplast *psbB/T/H* mRNAs of *Chlamydomonas reinhardtii*. Plant J. **21,** 469-482.

Vaistij, F.E., Boudreau, E., Lemaire, S.D., Goldschmidt-Clermont, M., and Rochaix, J.D. (2000b). Characterization of Mbb1, a nucleus-encoded tetratricopeptide-like repeat protein required for expression of the chloroplast *psbB/psbT/psbH* gene cluster in *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. USA **97,** 14813-14818.

Verbitskiy, D., Zehrmann, A., Van Der Merwe, J.A., Brennicke, A., and Takenaka, M. (2011). The PPR protein encoded by the LOVASTATIN INSENSITIVE 1 gene is involved in RNA editing at three sites in mitochondria of *Arabidopsis thaliana*. Plant J. **61,** 446-455.

Wakasugi, T., Tsudzuki, T., and Sugiura, M. (2001). The genomics of land plant chloroplasts: Gene content and alteration of genomic information by RNA editing. Photosynth. Res. **70,** 107-118.

Wang, Z., Zou, Y., Li, X., Zhang, Q., Chen, L., Wu, H., Su, D., et al. (2006). Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. Plant Cell **18,** 676 - 687.

Whitmore, L., and Wallace, B.A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res. **32,** W668-W673.

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

Williams, P.M., and Barkan, A. (2003). A chloroplast-localized PPR protein required for plastid ribosome accumulation. Plant J. **36,** 675-686.

Wostrikoff, K., Girard-Bascou, J., Wollman, F., and Choquet, Y. (2004). Biogenesis of PSI involves a cascade of translational autoregulation in the chloroplast of *Chlamydomonas*. EMBO J. **23,** 2696 - 2705.

Yamazaki, H., Tasaka, M., and Shikanai, T. (2004). PPR motifs of the nucleus-encoded factor, PGR3, function in the selective and distinct steps of chloroplast gene expression in *Arabidopsis*. Plant J. **38,** 152-163.

Yang, J., Schuster, G., and Stern, D.B. (1996). CSP41, a sequence-specific chloroplast mRNA binding protein, is an endoribonuclease. Plant Cell **8,** 1409-1420.

Yu, Q.-B., Jiang, Y., Chong, K., and Yang, Z.-N. (2009). AtECB2, a pentatricopeptide repeat protein, is required for chloroplast transcript *accD* RNA editing and early chloroplast biogenesis in *Arabidopsis thaliana*. Plant J. **59,** 1011-1023.

Zehrmann, A., Verbitskiy, D., van der Merwe, J.A., Brennicke, A., and Takenaka, M. (2009). A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of *Arabidopsis thaliana*. Plant Cell **21,** 558-567.

Zerges, W., and Rochaix, J.D. (1998). Low density membranes are associated with RNA-binding proteins and thylakoids in the chloroplast of *Chlamydomonas reinhardtii*. J. Cell. Biol. **140,** 101-110.

Zhou, W., Cheng, Y., Yap, A., Chateigner-Boutin, A.-L., Delannoy, E., Hammani, K., Small, I., et al. (2008). The *Arabidopsis* gene *YS1* encoding a DYW protein is required for editing of *rpoB* transcripts and the rapid development of chloroplasts during early growth. Plant J. **58,** 82-96.

Zsigmond, L., Rigo, G., Szarka, A., Szekely, G., Otvos, K., Darula, Z., Medzihradszky, K.F., et al. (2008). *Arabidopsis* PPR40 connects abiotic stress responses to mitochondrial electron transport. Plant Physiol. **146,** 1721-1737.

6 ANNEX

ANNEX A: Amino acid sequences of PPR repeats predicted by TPR-pred and Prosite. The numbers on left side indicate the position of amino acid from start methionine where the PPR repeat starts and numbers on the right side indicate the end of the repeat.

MCA1

	Begin	Alignment	End
PPR	246	KLDKLVLSLSANKATWRRSLLLFEWLKAAGHQLDD	280
PPR	281	RLCTTLIRVCSDHGDAVSALAVYDWMTGSTASGGA	315
PPR	322	YTYTAAMRAALAGGLTDRALSIWNEAWRRHSAGRL	356
PPR	361	RLCITYLELCTRLGLTDQALAMYAAMRAAPAGSRM	395
PPR	400	HAYTAAMRAATEGGRWYRALDIWADMRSAGCEPTG	434
PPR	435	HAYSAAISACAAAGDWRRAVALFDEMTGPGGIRPD	469
PPR	471	VSCTALITALAAGGEADRAEAVVAWMLSNSVRPNA	505
PPR	506	RTYTALMAALGNAKRWARAVEVLGRMQTPEWGGVQ	540
PPR	544	YTYSALLKSLGEHGQWQLAEAVFSSIERQVLGPAA	578
PPR	854	VVCGALMLAYERAGKWQEAVAVLLRALNLGITPNT	888
PPR	920	VTHETMIAAYGMAGLPERAEAVFSAMTSAGLRPRD	954
PPR	955	FAFCGLIAAHSLAGNWEGAMRVRARMRRAGVQPSV	989
PPR	990	HVYNALLAACERAGQPDRALELLGAMRREGVEPNT	1024

PPR1

	Begin	Alignment	End
PPR	122	SGTRGITALCQAGERWRDVVQVYNDLRGAGVQLDL	156
PPR	157	ASYRMLMDAAIEAKQGLDVVELLRHCEEAHGPAPP	191
PPR	193	LYCSLISRLLKHPGRGTPAKQAAYGVWRALRASGQ	227
PPR	232	VAYRTGMNLCVELGHIGEARRLMDAMRVAGFRPGW	266
PPR	267	GAYHILMKYHASRGDMDGARRLFAQLRAYRGGKPL	301
PPR	304	SAYNTLLCGFVRVGDLTMARAVLDKARREGARPDA	338
PPR	339	VSYSAYAAGLAAVGRLDEAEELLGEMAEAEGLRPG	373
PPR	375	HVYGALLDGSVRVHDWARVDRLLNRMRSEGLRPNL	409
PPR	615	GAAAYAYTSGVGGLAGEHVDALIAEMRAAGLRPDA	649
PPR	650	VTYGTLIHAAVRSGSVDGALGVLSAMRLEGVAPDA	684
PPR	685	AVFTSLMKLFRGQGMQAQALEFFQQLSGSRSAVVD	719
PPR	721	WSLTCLTAVHASGGNMEEAEGAARRAHELAAEQGL	755
PPR	760	EAAYALVQGYGQQRRLRNALLAFRRFLASGGRPHR	794
PPR	795	KLCEYTYRLCLAHFDFAAAGQVLRAMRLMRGLPLR	829

PPR2 (MRL1)

	Begin	Alignment	End
PPR	155	NTIMATLSGLSKRGEFDTLLRLLQELNKPMLEVVC	189
PPR	191	RWHYSSFIEEVCLTMKAKPAVVFAKIAAPYVSKPA	225
PPR	226	ALFMAVLKACARCRDLDAGMTVVELARKSGVTVDV	260
PPR	261	QMMTTLIKVCKAVGNVDKAYRVYLEMRAAQLRIDS	295
PPR	304	TCAEAMKRDLTVVHERKDQYVLLERAFQYVADAEA	338
PPR	346	PVWNSLMVCAGRSGELNRAFEVLTMMQQRGIGASA	380
PPR	381	TTYGSLIESCVCARQPEKALRVFEVALHKGFESEV	415
PPR	418	YTQALSACMLPFPGAWDRAQAIYSALQRCTSVRPD	452
PPR	454	KFFACFMAVAGRCGRMEVVFELLTEMAAEGIRPSS	488
PPR	489	TSVSGIIHACLDQGNVALARRVYDLCAKQSVYPVP	523
PPR	524	SQFNRMMDVYASEFRFGEVVSLLCDMVAAGRQPNL	558
PPR	559	NTYRIIINACEVTDQAGLAFQVFALMHANKVHILQ	593

PPR3

	Begin	Alignment	End
PPR	338	HVDTLIHQLFMSVRQLPRGQPAHEAVASRLDGLDG	372
PPR	373	RAVAALMKQLSTNGLAAAAWDIFDWLRGLEPGVDL	407
PPR	414	YLYTAMISLCSSNRRDLDVALSLSREMAARGVPRN	448

PPR	450	HTFSALMNVCIKAGQHQAALDVWRELQEAGCRPNV	484
PPR	485	TFNTLIDVYGKTGQWAEALKVLARMKEEGMAPVTR	519
PPR	520	RTYNTLMIACNSSNQWHEALSVYAQLVAAGQAPNT	554
PPR	555	TTYNALITAYSKAGRLEKVLETLREMEAAGCERTV	589
PPR	590	ITYSALISACERNGQWELALQMFGQMLREGCNPNI	624
PPR	625	ITYNSLITALAAGAQWERAADVFAKLQRQGCEPDV	659
PPR	660	VTYTALINAYEKGGQWRRALQAFKRLLQQGCRTDH	694
PPR	698	AAITDVLWDSGTASAQARAARIYRAAVGSGAIKPV	732

PPR4

	Begin	Alignment	End
PPR	170	RAVALLLKDLSRLGKDRRAMELFDWLRSANERSPL	204
PPR	211	YSYTATISLCIYSQDVDRAMELMNEMRQRNIERNV	245
PPR	246	HTFTALMNVCIKCGKLPLALEIYNNMRAANCMPNV	280
PPR	281	VTYNTLVDVYGKLGRWERAIHVLDLMKQEGVEPVL	315
PPR	316	RTYNTLIIACNMCNQPREALAVYQRLLSDGYTPNS	350
PPR	351	TTYNALISAYGKTMQLGKALEVYQEMLRQNMERSV	385
PPR	386	ITYSSLISACEKAGQWETALRIFNEMQQDNCVPNT	420
PPR	421	VTYNSLVTACAQGGQWEKATEVFEQMTAHGCTPDV	455
PPR	456	VTYTALISAYERGGQWQKALQAFGKMCMQGCKPDA	490
PPR	491	IVYNAIIDTLWETGIIWAQGRALQLFLTAVQQGHF	525

PPR5

	Begin	Alignment	End
PPR	114	GLFDSALLVCAAEYQDERCVALYRRMRQAGLAADV	148
PPR	149	GSLNLALTGLCCRGRLEEAVELLEQEVEDSGGVIK	183
PPR	189	VLMILQACNYKRRGAYREAIDAVRALESWGRPANE	223
PPR	228	ALLCVCEAAMHKAPSFEAAVEVFSALAELQLADST	262
PPR	263	RVYNGLLGAAGRAGRWREAQALYIAMQADDIPASL	297
PPR	298	ETHTALIQACVVGRALDHALDIFEFLVAGRAAHEL	332
PPR	338	QTYNHLIHACHQAGMLEKALEIAAWVQKTGVEFND	372
PPR	378	LMATIEVAQLWDEKAMKQALKRHPAVLPKHLRPAP	412

PPR6

	Begin	Alignment	End
PPR	237	SATAALLKELAKQGYLKRAVEIFDWLRSLAPGDEL	271
PPR	278	YTYTTMISQCGSHQQLRRALELVAEMRSRGIDCNV	312
PPR	313	HTYSALMNVCIKANELDLAQDVYKQMLEEGCSPNL	347
PPR	348	VTYNILIDVYVKRCQWEEAVRVLDTLEKQGIQAEV	382
PPR	383	RTYNTVISACNKSGQPEQALKVYEKMLAAGVKPSA	417
PPR	418	TTYTALISAYGKKGQVEKALDIFRDMIRRGCERNV	452
PPR	453	ITYSSLISACEKAGRWEMALELFSKMHKENCKPNV	487
PPR	488	VTYNSLIAACSHGGHWEKASELFEQMQTQGCKPDS	522
PPR	523	ITYCGLITAYERGGQWRRALKAFEQMQSQGCHPDA	557
PPR	558	AVFNSLMEVLWQSGVLLAQAKALQLWTLANRSGHF	592

PPR7

	Begin	Alignment	End
PPR	59	RAEVTKRIKALGTQGKVKDAISALAGLANLGIQPD	93
PPR	94	DTRAATALVQACTRDMELAQSIFDEMFGEFLQPDE	128
PPR	130	FSVLLRGYGATTPPDWPRIDSTLTTMRVKYGIEPT	164
PPR	166	LSFNALLEVCCRTSDIDRGQDIIDRMAADGVEPDE	200

PPR8

	Begin	Alignment	End
PPR	347	CNALLAAYARATPTQWQRAIRLLELMWQIGGELCP	381
PPR	384	VSYNTVIKACGNAGQVDLGFKVYGLMRRRGIEPSV	418
PPR	419	ATYGTLVCIAADAGASGRVIEVWGWLRASGLEVHV	453
PPR	454	TCANAYLAALIKQGEWDAAVGFFRSLLRGGSPCRP	488
PPR	491	ITFNTLMAGCLDRGQPEQVLTLFHELQSVCGLAPS	525
PPR	527	SSYGHVVAAYGGLGRWEDAFNVVAFVCRPDSSVRP	561
PPR	587	AAGDGATTAAAAAALLALARHMHDLRNLLKMYPHL	621
PPR	626	ACHRAMIGAFAAVGEVDVCVELFNVLLLLPPPQPP	660

PPR	834	EVATQVLRAICMHRAWDKAVQLLKVLQDHGPPLPA	868
PPR	869	SALRVVLEAMALDAAWTAAYSVLQMLCSPLLPGSV	903
PPR	925	LLFLALLAAKDEAGAWGEALSMFNWLLREARPDAL	959

PPR9

	Begin	Alignment	End
PPR	14	AAQEVAVVSYARGGQWQRAEALYKLLQGRGHVAAG	48
PPR	49	ATTQELVSALAAAGKGGEAAALLAAAAAQAGADAG	83
PPR	87	PAHCVVARALARQGQLSEAYAMLNRMMREEMRIDP	121

PPR10

	Begin	Alignment	End
PPR	643	HTYASAFRGLYRAGRLGMALAAFEEACRSGRDLGP	677
PPR	678	VACSALLHVMSQERNLRLAWQLFDQMVAARMTLNR	712
PPR	713	YAYNCMAHLASLHGALDDTLLLYNMMRSEARQQQQ	747
PPR	753	QQQQQLQQAAATATAGTAAGGAAAQREEAEPEAAA	787
PPR	789	ESVGAAFSGSQGGHSSHNGAAYRSHAPELTLDCSP	823
PPR	826	YTYSALVRAAVTAGRGDLLPALFNEMAASQRAADR	860
PPR	882	EVWGHFISAASRTGQPELAMRFFDAGIAELGLLPT	916
PPR	918	PIYNAALAAMARAGRPLPELMSLYREMVGGCSRPL	952
PPR	1138	TSYICMAALYASARQPGDVLRVVREMVAGGVVADG	1172
PPR	1173	SSWEFLVAAAEDAGLYGLVPQLQRAAEAQACAAEQ	1207

ANNEX B: Table 1

The data for the microarray designed for RIP-chip experiments is depicted in Annex B, Table 1. The microarray contains 15 overlapping PCR fragments representing the complete *C. reinhardtii* mitochondrial genome and 166 overlapping PCR fragments representing the complete chloroplast genome. Names, positions and primer pairs used for amplification of the PCR products are mentioned in Table 1. Fragments were numbered (ID) according to their position on the *C. reinhardtii* chloroplast and mitochondrial genomes sequence (Acc. No. BK_000554 for chloroplast and U03843 for mitochondrial probes).

Name	ID	Sequence of primer 1	Sequence of primer 2	Start of probe[1]	End of probe[1]
cob 3'	M1	GTG TCA ATGCAAAATATTAGTTG	CTTTCGGTTCTTACTTTGGTG	529	1032
nad4 3', cob 5'	M2	GGA TTAGGTTGTCTGGGTG	CTTCAGCGGTCTAGGTG	916	2031
nad4 5'	M3	TAAGCAAGCAACTCGTGTAC	GAACGTAGGTAGCTAAGAC	1882	3165
nad5 3'	M4	CATATGATAAAATAATAGGATGCTC	TTTCTATCTTGATTGGTGCTATG	3042	4332
nad5 5', cox1 5'	M5	TACAAGCAAGTATACACCAGC	GTTCAGCCAGAAACTAATGTTG	4190	5353
cox1	M6	ACGGTATTATCATGCTATTGTTC	CTAATGCTAGCACCGAAGC	5220	6405
cox1 3', nad2 5'	M7	CTTGGGTTTGGCTGGTATG	GTTAATAGTACTATAAGCAAACAAG	6313	7441
nad2 3', nad6, trnW 5'	M8	GCTTCTTCTTGGCACCATG	GAACCCTCGATTTCAAGATTTG	7319	8517
trnW 3', rrnL6, rrnS1, rrnL5, trnQ	M9	GGTTTTGACTGTAGCTCTG	CATTTTGCCGAGTTCCTTTAAC	8378	9521
rrnL7, rrnS2, nad1 5'	M10	CCACTTTTAGAAAAACAATTTGTTAC	GAAAGAAGCATCGGAAATGC	9402	10633
nad1 3', rrnL3a, rrnL3b 5'	M11	GCTTCTATGCAACGCCG	GCAGAAAACCAGCTATTACC	10430	11518
rrnL3b 3', rtl 5'	M12	GATCGATCTAGCGACGTG	GAATGGTCCATACTTGTGG	11393	12600
rtl 3', rrnL8, rrnS3 5'	M13	CTAGGGCACAAGGTACTG	GTATCACCAGGCGTTATAG	12507	13607
rrnS3 3', trnM, rrnL4, rrnS4, rrnL1, rrnL2a5'	M14	GAAAGCCGAAGGCGAAAG	AAAGGTACGCTAAATTATAGCTG	13467	14737
rrnS4 3', rrnL1, rrnL2a, rrnL2b	M15	CGGATTGGAGTAGCAAAAC	GCCCTTATGGATCATCAAC	14215	15278
ORF 271 5', ORF 140, Wendy	1	CACCGCAAAAGGTTGGG	CTAAAGAATAACATTCAACTATAAAAC	6	1110
wendy	2	GTAACTTCTTTTGAACGATCC	CCGAAGGACGTCAGTG	921	2138
petA 5'	3	GCAAATAAATTTTATTCTATATTTATATAC	GAAACGTTTTTATTTTTAGCAGG	2051	3415
petA 3'	4	GAACAAAAAAATATTTTAGTTGTTGG	CTGCCTATGTTAATACTCC	3312	4608
IG petA-petD	5	AAATATAGAATGTTTACATACTCCG	GCTCTTCGGAGTATTAAAATAG	4360	5798
petD	6	CAAATTTATTTATTGTATATAAATATCGG	AGGGTTGCAAATCGTATTG	5589	6913
petD-3'-trnR1	7	GCTACTATCTTATTCCTTTTAG	CTAGTATTAAATCCTATATTTATATAC	6766	7980
IG trnR1-chlB	8	GTCACTGTCACGCAATTTG	TAATTTCATAAAATGTTATTTTGTAGTC	7822	8922
chlB-3'-trnP	10	GACTGCTATGTATTTAAAATCC	TATCGGGATGACAGGATTC	9882	11220
trnP 3' -psbK	11	AGGTTCGAATCCTGTCATC	AAATAGGTGGTATATTTATCTTTTTC	11196	12252
tufA-5'	12	GAAAGGGTTTAAAAACTCCC	CCATTAATTGGTAAATTTTGTCAAC	12095	13318
tufA-3'-trnE1 5'	13	GTTGAATTAGAAGTACGTGAAAC	CGTTTTCTCCGTGAAAGG	13156	14511
trnE1-trnC-trnT-trnR2	14	CACTCAACATGTAACTATTATTTAG	TAAGTGGTTTTTAAAAGTTAATATATC	14361	15694
trnR2-rpl20	15	AAGAAAACACATTAATTAATCCCG	CTAAGCTTTAATGGTGTACTC	15514	16743
trnS1-trnW-clpP-3'	16	CTACTAAACTGAACTAAATTTTCG	GTTATCATGATAGTAAACAAGAAC	16595	17724
clpP-3'	17	TAGTGATATCCCTTAATATAAAAAC	GCACCTCAAAATTATAATTTAGAAC	17524	18535
clpP-5'-trnL1	18	AATCTATTCATTAATTCTTTGTGTG	GATATTATAAAAGTAATATTTTTTAC	18390	19534
trnL1-petB 3'	19	GATAAATTCAAACAAATATTGGCTC	AAGAACGTTTAGAAATTCAAGC	19387	20629
petB 5'-chlL-5'	20	TAGTAGAAAGTCATAGCAAAAC	CGATTAAACCCGCTAAACG	20482	21820
chlL-3'-rpl36	21	GTAACTGATAATGGTTTTGATGC	GCGTAAATAAATATCAACTGGC	21718	23113
rpl23	22	ACGTCCTAAGTTTACTTGCC	ATCTTTTGTTGTTCTTGAGAGG	22883	24301
rpl23 3'-rpl2 5'	23	AGATGTAATTTCAGTAAATACGC	ATTGAATAATATCACCAATATTTAAAC	24087	25318
rpl2-3'	24	GGTATTATTACATGTCGTCATC	CAATTAACGTTAAAAATTTATCAACC	25105	26254
rps19	25	CCACTAAAATTTATTCTTTCTCG	CACATAAAACTAATATATATAATTATTC	26093	27338
rpl16	26	AAGGGATTTACTACTATAAAATTAATG	AACAATACTTATGCTCTTAATTTTC	27205	28531
rpl14	27	TTGTTTAGAAAGAATTAATGATTTAAC	ATTTTATATGGAAATCCTATTCGC	28423	29815
rpl15	28	CTCCGAAGAGGCAGTTG	CGTCCACTACAATTTATTTGC	29658	30843
rps8-5'	29	GAAATTTGTTAGCCCCCATC	GCTTCACGATCAGTCATTAG	30687	32018
rps8-3'-psaAex1-trnM1	30	GTATTACTAACTTAAAACGCATTAG	CAAGTATTAAAAACTAATACTCTATC	31883	33315
trnG1-rps4	31	CCTACTCCGGTTTTCTAAG	CAATAAATAAATTGTATGTAAACCAG	33194	34609
rps4-3'	32	GAAAACGCCAAATTATATTTGTC	TTGCAAACCTTACGGGAC	34273	35362
IG rps4-rrnS-1	33	CGGCAAGGGCAGTTGG	CATATAGCTTAAACAAAAAATAAGACC	34919	36860

Name	ID	Sequence of primer 1	Sequence of primer 2	Start of probe[1]	End of probe[1]
IG rps4-rrnS-2	34	GCTTGTCAAAGTTTACTAAAAAAAAG	CTGTCGTCTATTTTAACACTC	36442	37995
rrnS-5'	**35**	GGTCACAAGAAAATGTAAGAC	GACGCTTTACGCCCAATC	37708	39079
rrnS-3'-1	**36**	GTAACGGCCTCCCAAG	AGTTGAGTTTTATAGCTAAAACTC	38785	40075
rrnS-3'-2	**37**	CAGTCAGCTATATGGCGG	GCTCCTAAGTTTACCGGC	39835	40995
trnI-trnA	38	CAGCAGGTTTACATACTCC	GCTCTTAATATTTAGTGGCTAC	40703	41759
rrn7-rrn3-rrnLex1	39	GCTGCGTTAAACTTTAGTTAC	GAAACAGTAGCTTTACCCC	41564	42723
rrnL ex1	40	GGGTAACACCAAGTGGAG	GGTCGTCAGCTTTTTTCTTAC	42559	43723
rrnL ex1-in	41	CCAAAAACACAGGTCTCC	GCCAAATAATTTTTAAAACTAAATTTG	43613	44943
rrnLin-ex2-rrn5	42	GGTAGCGTTTCGGATTATATTC	TCCTAATTTAATACTCCGAAGG	44823	46504
IG rrn5-psbA-1	43	GAAACCAAAGAAGCTCCAAC	GCTGGATCCGTATCCATG	45851	47252
IG rrn5-psbA-2	44	TAAGGGAAAGTCGTGGAG	CGTAAGACGTCCTCCTTC	47087	48414
psbA ex5-in4	45	GGTTTTTTGTGGCAGTTGC	AAGCTTTACCACCATGTAAAG	48149	49428
psbA in4	46	TTGGGACTCAAGCTCTAG	CATCCAAACACCTTCTCAAG	49205	50225
psbA in4 ex4	47	CAGTTTTCACTTGTTGTTTTTGC	TCTGGTACTTTCAACTTCATG	50085	51125
psbA ex4 in3	48	GATGAAGTAACTAAAGAACCG	TCTGACGGTATGCCTTTAG	50989	52248
psbA in3 ex3 in2	49	GAGGTAGCAAAGGGCTG	CAAAACGTATTTATTTAGGTGGC	52098	53327
psbA in2 ex2 in1	50	GTAAAGGTACAAAAAAGAAATCAG	CAAAAATTGCTAATAAAGCAAGTC	53166	54524
psbA in1 ex1	51	GGATCTTTATATACCCCGAAG	CGGAGTATATAAATATAGAATGTTTAC	54273	55871
psbAex1 5'	52	CACGACGTTCTAAAATTGCTG	CCGACACGCATTCCTTAG	55389	56879
IG psbA-trnS2-2	53	GGGGAATTCGCACAAAAAG	TGGTTCTATTTTATAAAAATAGTGAC	56712	58020
IG psbA-trnS2-3	54	CCGTAAGGGATTAAAAGAGG	TAGCTCCAATGGAGAATAGG	57852	59062
trnS2-ycf12-atpE-3'	55	CACACAAACAATTTTATTGTATGG	CTAAAGATGCTTTTGAAACAGC	58922	60088
atpE	56	TTATTAGGGATTTGTTGATAAGTTG	CCACTGCCTAGGATTAAAATC	59927	61128
rps7	57	GGAATTGCAATGCTCCTATTG	TCAAAGAGGCAGTTGGCAG	60996	62390
rps7-5'-rps14	58	CATTGGTTGACCTGAGCG	TGCCTATTTTAATACAATAAATAATTTG	61765	63283
psbM	59	GAAGTAAAATTAGGAATAAAGACTG	CGTAGTTGGATCATTGTCAG	63082	64516
psbZ	60	GAATATTGACATACCGTCTAG	CATTCTATATTTATATACTGCAGC	64375	65717
IG psbZ-ccsA	61	CTTAAGAAATTTTATTTACCCTATG	GGGTAGCAAACATTTTTAGTC	65546	66716
ccsA-3'	62	GGAAAAAAATTGAATTTAATAAATGACC	GTTATGCCACATTAATTATTGGATC	66592	67922
ccsA 5'-rrnL2 3'	63	CCTAATCCAATAATGCGATAAC	CTACGGACTTAAAATCCGTTC	67784	69055
trnL2-psaAex3-1	64	CTTACAGGGATGTCTTTTAAC	TGTTTTATTTGCTCGTAGCTC	68856	70117
psaA ex3-2	65	GAAAAGACCTAAGAATACGTG	AATTGGTGGTTTAGTAATGGC	69985	71287
psaA ex3-3	66	CACCTAAGTGGTGGTTTAAC	CCGATATTTTAATACTCCGAAG	71166	72564
wendy disabled 3'	67	CAATAAATAAATTGTATTAAAATATTAATC	GATTTCTAACCAACTCAAGGC	72176	73605
wendy disabled	68	GCCAAAATTGTTTGGATAAATCC	CACCGCAACAAAAAGTTGC	73494	74705
wendy disabled 5'-trnE2 3'	69	GGATCAACTTGTAACTGCAAG	GGCACTTTCAACTAAGTGTAC	74565	75939
trnE2 5'- psbH	**70**	GAATTCGAATCCGCGTTTTC	GAACTTCTAAAGCTAAACCATC	75794	77003
psbH 5'-psbN	71	ACGTCATCTAAAATTAAAGAACTG	CATCGCAGTATATAAATATCCAC	76784	78425
IG psbN-psbT	72	CTTCGGGCAAATGCATTTTAG	AAGGTCTATTAAGACTCACTTTC	78208	79411
psbT-psbB 3'	73	CATTTATATACGAAAAGAGGAAAG	ATCTAAATATAGTATTGAACAAGTAG	79264	80643
psbB	74	GAACGGTCAAATTCAAAAATTTCAC	GGTATCGTGTACATACAGTAG	80500	81790
psbB 5'-trnD	75	GTTTAATACTGGGTCAGATGG	CATGGGAGTATATAAATCTCC	81645	82956
rpoA 3'-1	76	GCTGGTAATTTCTTTTAGGAC	CAAGAAAATTTTAAATTAAAACCATTAC	82833	84283
rpoA 3'-2	77	CTAAATCTGATTTAAACATGCTAC	CAACTTTAAAAAAACGCAATATTTTAC	84106	85369
rpoA-5'	78	ATACGTTTAAAGCTTTTTTTAAAACG	CCCTTTAGAATATTTAAATATATAATAC	85213	86496
rps2C-3'	79	ATAATTTTTGCAGAGTTTTTACTTTC	GTAATATTATTCGTGTTTTACGTC	86329	87648
rps2C-5'-rps2 like 3'	80	CATTTTAGGATCCACAAATTTTTTAAC	GCTAAAATTATTAAATTAATCCTCTTTC	87491	88949
rps2 like 5'	81	TTTTACGTGTCCATACGTAATTC	CAAGATATTTATATACATATTAATACTC	88796	90146
rps18	82	GGATATTTATATACTATTTATTTATTGTAC	AAGCTGCTTCAAATTTTATAAAAAAC	89914	91189
rps18-5'-ycf3	83	GATAAAACTTTACGACGGAATTTAC	CTTTTTACAAGTTTCTTTAGAAGCC	91029	92194
ycf3 5'-ycf4-rps9 3'	84	TTGGCATTAAAAATAAAATCAGCG	GAAGCTTTAGCTATTCCAAATG	91981	93305
rps9	85	GTTAAAATTTAGAAGTTAATAAAATTC	ACGTACAAGAAATAATAATATGATAG	93117	94418
psbE-M2	86	GCAAATAAACACAAGGGAATTTAG	CATTCTTTAGTTTTTTGCATATTCG	94130	95983
rpoB2-3'	87	CGTAAAGAACGTGCCTTC	GTGTTCCATCACGTATGAATG	95428	96904
rpoB2-5'	88	ATAAAGTTTAGATAAAACAAAACTGC	CTTTGGATATGCAATTTCTTTCG	96756	97954
rpoB2-5'-rpoB1 3'	89	GAATTAAGCAAGTATCTTGATTTG	AATAGCAACATCATTAATTC	97673	98855
rpoB1-3'	90	GACATGACCAGAATCACTAAC	AAAACCTTTACCTTATGTAAAAACAC	98696	100073
rpoB1-5'	91	CTGCTGGAGTTTTACCTTTAG	GAATCACAAAAAAATAAAACAGAACG	99856	101032
rpoB1 5'-psbF 5'	92	CTGGCAGTTAAATCAATTTGG	ACCAACGAACTGTGAAAATAG	100881	102205
psbF-psbL	93	CGAGTTAGCTTAATACAAAAGG	CTCGTATAAATAATATCTAACGTG	102097	103501
petG-rps3 5'	94	AGTACATTATCATTATATGGTACC	GAATTTGAATACCAATTTCATATGG	103354	104740
rps3-5'	95	TGTTTTGCACCTTGAAATGGC	CAAAATTATTTACTACTTTTGTGTAATG	104357	105847
rps3-3'	96	CGTAAAGATTTTATTACATTTGGTAC	ACCCTAGTATATAAATATCGCG	105595	106784

Name	ID	Sequence of primer 1	Sequence of primer 2	Start of probe[1]	End of probe[1]
IG rps3-rpoC2	97	CTAACATCAAACCTTTTGTTACG	CAACAATCTAAAATAATAGTTAGCTC	106627	107923
rpoC2-5'	98	GATTCTATTATCTGTTTTTTTAGAAG	GTTTATAAATATCACGTGCTAAAAC	107779	109227
rpoC2-1	99	GCGCAACATGTAATTATTTCAC	CCGAATAACTGTCGACTTC	109086	110388
rpoC2-2	100	CTTCTATAGTACGTCGATATC	CACTAGCAGGTAAATAAATTTTATC	110245	111375
rpoC2-3	101	CAGACGGAGTACCAGAC	GCTTTTGCTAGCATATGTGTG	111217	112486
rpoC2-4	102	CGTTAGGCGATAAGCAAGC	TATAATAAAAACAGGTCTAGCGTAG	112237	113576
rpoC2-5	**103**	TGTTAAAACTATTAAAATAAATCCAAC	GTTGAAGCGTTTATAACCAG	113327	114901
rpoC2-6	104	GGTCACGTTTTAATTTAATTTTAAC	CATTATACAGAACAAGTTCGTG	114589	115782
rpoC2-7	105	GAAGATTGATTGGCCTTAAGG	CTGTTCTTTTAGTAGTGATAACG	115624	116925
rpoC2 3'-trnY	106	GGTGAAATGCCTGAAACTG	TCCCAGTGAGGATACTAG	116763	118117
trnQ	107	GTCACAGATTTGCCTATAAAG	ATCATGGCTTTCAAAGTCATG	117983	119708
psaB-5'	108	GAGTAAGACAATTTTATGGCAAC	GCATATGGTGGTAATGAGTAC	119586	120679
psaB 3'	109	CTAACTTTGGTATTGGTCAC	GCTGATCTGATGAGGTAG	120481	121844
psaB 3'-trnG2-rbcL 3'	110	CAGTTGCTCTTTCTATTGTAC	ACGTTGAAAAAGACCGTAGC	121693	122860
rbcL	111	GGTGACCTAGAGTACCAC	GCAGGTGCTGGATTCAAAG	122690	123893
rbcL 5'-atpA 5'	**112**	AAGATTCAGCAGCTACAGC	CCATAAATACGAGCAATACCG	123734	124902
5' atpA	113	GGCAATGCGTACTCCAG	CAAGGTCTGAAGCGAATTG	124774	125981
atpA-3'-psbI	114	TAGGTATTTCAGTATCACGTG	GAATGTGTCGAAATCGATAC	125850	1E+05
cemA 5'	115	GGAAACGCCATTTTATAAATAAAC	CTATCGTCTAAACCAAAAAATAC	127038	128477
cemA 3'-atpH	**116**	GAATTAGCAACATACTATAATAATC	AAGAGACTACCAGAATATCAAC	128314	129509
IG atpH-atpF	117	TTACTATAGAGATTTGTTAGACTG	AATAAATGCCTGTCTTACCTTG	129382	130381
atpF 5'	118	CTTACTTTCGAAGTTGCAAAG	CTTGTTTAAATTCTTGTAAACGAG	130252	131535
atpF 3'	119	GCTGCTGTTGTTGGTATTG	ATTTCCGGCAAAAGTTTG	131259	132769
IG atpF-rps11-trnK	120	CCTTTAAGGTCTGCTAAAAGC	TAAAAATTAGGCAGTCAGCAG	132377	133800
trnK	121	CCACCATAATACAATTGTAACC	GTTTATTTGCTCCATACTTTGC	133033	134968
tscA	**122**	GATATTTAATACTCCGAAGCAG	CGAGATGCTACACCCATC	134706	135910
chlN 5'	123	ACTACTTGATTGCACTTTTGG	CTGTCCCATTAACTGAAGAAG	135678	136820
chlN 3'	124	GATTTAGAAGGGATGGCTC	AGGCGGTTATATACTCCAAAG	136595	138053
chlN 3'	125	GTGAATTAAAACCTGATTTAACG	CACATTTATGTACCTACGTC	137448	138659
atpB-3'	126	GGAATTTTACCGATTAAATTAACC	GTGGTAAAATTGGTCTTTTCG	159813	161147
atpB 5'- ORF1995 3'	127	GTGGTTCGTTCATTTGACC	CGTTTAGAAATTGTTGAAATGCC	160914	162333
ORF1995-1	128	CACTTTAGTTAATTGTTGTGTTG	CGACTACGTTACCATTTTATG	162148	163367
ORF1995-2	129	CGTTTAAACCTTGAATTGGAG	ATTGCTGTTGTTAATAAAAAATTAGC	163215	164316
ORF1995-3	130	GGTAAGTAAGCTTGATTATCTTG	GCTATTACACCAAAACAAGGC	164126	165318
ORF1995-4	131	GTTGTTGAGTTTACACTTATATTAG	TAATGTGAATACGCGTTTAAATC	165146	166372
ORF1995-5	132	GTTGAATTTTTGAAAAAATGGATTTC	CACTTGCAGGTAATTTTTTATGG	166251	167633
ORF1995-5'	133	CATACGACGTTCTTTTGAAC	CAAAATCCATTTACTTTGTCTTAC	167482	168699
IG ORF1995-rps12	134	GTTAACTTTAATTAACTGCAAACG	GATGCGTATCATAAACGCC	168537	169612
rps12-psaJ	135	CGTAAGGACAAATTTATTTATTGC	CTAAATTATGCATTCTTTGAACTAG	169483	170649
psaJ 5'-atpI	**136**	CGGTTAATTTCAATTAATAAACCTG	CCAATAATTGGTAAGAATTTTATTAC	170490	171665
atpI 5'-psbJ	137	CAAGAAGTAATTAGTACTTGACC	AGTTTATCGCAGTATATAAATATCC	171533	173120
IG psbJ- psaA ex2	138	CAACTGACTCCGAAGGAG	GTATAAGCACTTTTAAAAATCAAAGC	172822	174140
psaA ex2-psbD 3'	139	GATGTCTTACAAGGATTTAGG	ATTCCACATGATGGGTGTTG	173907	175048
psbD 5'	140	GAGTATGTTTCTTCAGCCTG	GATTGACACCCTTTAAGTAAAC	174899	176102
ORF2971-5'	141	CCATGAATAAAGGTTTTATTATTATG	GTTTCAATAGGTCTTGCTTG	175859	177270
ORF2971-5' 1	142	TTAGATATACCATAATGTTTTTAACC	GTAAATAAATTTCAACTGACGTC	177117	178364
ORF2971-2	143	TACTTACACCAACTTTATTAACAC	CCATCTTTATTAATAATAGTGTTTG	178197	179364
ORF2971-3	144	CGAGACTTATGCACAATTCG	GTTTAGGGCTTTATGTATTCC	179207	180364
ORF2971-4	145	ACAAGGTTTTTCACTTCCAAG	GATTTTTGATGATTCATATAAAAGTAAC	180163	181575
ORF2971-5	146	CTTTAGTGAGTATTAGCCAAG	GACATAGTTAAATTGTCAAAATCG	181392	182628
ORF2971-6	147	GGTACATTATTTTGGGACATTTG	CTTCATCACGCTGACTACC	182480	183545
ORF2971-7	148	CCTTAAATACACCTTGTATCTTC	GAATTAGGTAAGTGTTGTTGAAG	183399	184668
ORF2971-8	149	ACTCAATGTTTCTAAATTGATTGG	TTGTCACAATTTCTTCTGAATAC	184540	185876
ORF2971-3'	150	CTGATGGACTTGATGTATTATTAG	ACTTTGCATTACCTCCGTAC	185736	187026
psbC-5'	151	TACAAGAGGATTTTTGTTGTTAAAG	GAAACGCATAGTTTCACCAC	186894	188062
psbC 3'- trnH-trnfM	152	GAAGCAAGTCAATCTCAAGC	GTAATAAAGTGCCTGTTGCC	187910	189099
trnF-psaC	153	CATTGGATAGTGATTCTTTTAATTG	CGGAAGGTAATTTTTATCTTCTC	188951	190263
petL- trnN	154	GCTTTATTCTATCCCATTTAGG	CGGGTAACTAAAGTTTATCGC	190124	191590
IG N-rpoC1-2-1	155	GAGTATATAAATATCCTGCCAAC	CAACTGCAACTAAATTAGATTTAC	191262	192760
IG N-rpoC1-2-2	156	CAGAGTATGTTGAGTTTACAC	CCATTTCTTTACTCCTCTTTG	192587	193837
rpoC1b-3'-1	157	ATGCATTATGGTTACAAGTTTAG	CATAATATTTAACTGCACCACC	193686	194831
rpoC1b-3'-2	158	CCTGAGTGGAATTTGATATTTAG	GATGGCGTTAGTTTTATATTCG	194686	195940
IG rpoC1- rpoC2	160	GAGTCAATCTGTAAATAGTAATAAAG	GGACGCAGTTCTATAAAATCC	196883	198217

Name	ID	Sequence of primer 1	Sequence of primer 2	Start of probe[1]	End of probe[1]
rpoC1-1-1	161	CTAACTGACGAAATGTCTTTAC	CCTGTCTCACCATCAAGC	198069	199362
rpoC1-1-2	162	GACGTATTACTATTCGGTTGAC	GCAGCTTCTGCTGAAATGG	199214	200391
rpoC1-1 5' -trnV	163	CTCACAGTTAGTTGAGCTTC	GGACGTCCTATATTTATATACTTC	200255	201444
ORF271 3'	165	GAGCATCCACCGACATC	CTTAACACAAGCTGTTAGGAAG	202275	203642
ORF271 5'-ORF140 3', Wendy	166	CTTGATTGAGCTCTCTAATTTTC	AACTTATTGTGTTTATTAGATGGG	203120	612

[1] According to gene bank acc. no. BK000554 for plastid probes and according to U03843 for mitochondrial probes; probes with less than one third of all spots on the array passing the quality control criteria detailed in Schmitz-Linneweber et al., (2005) were excluded from the analysis; this eliminated three PCR products from the analysis.

ANNEX B: Table 2

Summary of RIP-Chip data containing normalized values from 8 experiments (four co-immunoprecipitations of PPR7 and four co-immunoprecipitations with pre-immune serum). Background-subtracted median values from the indicated total spot numbers per probe for both fluorescence channels are shown.

Name	Median(F635 Median - B635)_ PPR7 [2]	Median(F635 Median - B635)_ pre	Median(F532 Median - B532)_ PPR7	Median(F532 Median - B532)_ pre	Median(Median of Ratios (635/532)_ PPR7	Median(Median of Ratios (635/532)_ pre	Diff. Enrich. [3]	ttest [4]
cob 3'	184	128,5	994	899,5	0,1905	0,1611225	1,1823302	0,2297248
nad4 3', cob 5'	156	96	631	373,5	0,302	0,252681	1,1951829	0,1573749
nad4 5'	163,5	348	503	318	0,336	1,032207	0,3255161	1,36E-04
nad5 3'	169	100	700,5	540,5	0,2575	0,2613765	0,9851689	0,1881739
nad5 5', cox1 5'	168	201,5	862	457,5	0,212	0,396924	0,5341073	8,86E-04
cox1	88	96,5	316	255	0,334	0,45012	0,7420243	0,0035303
cox1 3', nad2 5'	82	52	231,5	171,5	0,4135	0,575949	0,7179455	0,0058738
nad2 3', nad6, trnW 5'	140	207,5	597	374,5	0,29	0,36828	0,7874443	0,0089529
trnW 3', rrnL6, rrnS1, rrnL5, trnQ	301	603	4920,5	2834,5	0,0615	0,2153415	0,2855929	7,66E-05
rrnL7, rrnS2, nad1 5'	127,5	170	696	445	0,2145	0,363165	0,5906406	1,40E-04
nad1 3', rrnL3a, rrnL3b 5'	284,5	456	1477,5	1044,5	0,178	0,298716	0,5958837	3,24E-04
rrnL3b 3', rtl 5'	96	71	238	178,5	0,504	0,656766	0,7673966	0,0042897
rtl 3', rrnL8, rrnS3 5'	135,5	132	1042	629,5	0,1545	0,187209	0,8252808	0,0102332
rrnS3 3', trnM, rrnL4, rrnS4, rrnL1, rrnL2a5'	330,5	670	2124	1375	0,164	0,3636765	0,4509502	4,56E-04
rrnS4 3', rrnL1, rrnL2a, rrnL2b	162,5	270,5	401,5	226	0,408	0,9498555	0,429539	1,10E-04
ORF 271 5', ORF 140, Wendy	447	137	1650	909	0,2815	0,140151	2,0085479	0,0576352
wendy	758	190	3620,5	2583	0,2015	0,094116	2,140975	0,0617301
petA 5'	696	182,5	2022	833	0,35	0,1918125	1,8246986	0,0060103
petA 3'	517	135	1143	595	0,498	0,263934	1,8868353	0,0037705
IG petA-petD	117	NaN	46	NaN	1,516			-1
petD	681,5	223	1326	1104,5	0,4985	0,257796	1,9336995	0,0520958
petD-3'-trnR1	172	95	682	649	0,2775	0,1826055	1,5196695	0,1073136
IG trnR1-chlB	336	154	719	466	0,527	0,3421935	1,5400643	0,718594
chlB-3'-trnP	717	150	17649	8772	0,043	0,02046	2,1016618	0,0051958
trnP 3' -psbK	524	98	949	507	0,542	0,224037	2,4192432	4,89E-04
tufA-5'	279	151	1312	609,5	0,246	0,193347	1,2723239	0,1194369
tufA-3'-trnE1 5'	660	419,5	8380	11429	0,085	0,0373395	2,2764097	3,94E-07
trnE1-trnC-trnT-trnR2	372	161,5	4862,5	2978,5	0,083	0,0700755	1,1844368	0,6925313
trnR2-rpl20	716,5	209,5	12485,5	6675,5	0,0665	0,036828	1,8056913	0,0185596
trnS1-trnW-clpP-3'	485	229	7765	4264,5	0,065	0,060357	1,0769256	0,6082921
clpP-3'	272,5	169,5	455,5	326,5	0,612	0,5314485	1,1515697	0,087794
clpP-5'-trnL1	307	195,5	580	481,5	0,5495	0,4741605	1,1588903	0,5133168
trnL1-petB 3'	850	190	18904	9640,5	0,045	0,02046	2,1994135	6,05E-09
petB 5'-chlL-5'	1134	343,5	1845,5	1015,5	0,6075	0,3043425	1,9961064	0,2423658
chlL-3'-rpl36	408,5	214	1059	727	0,3675	0,3012735	1,2198219	0,36982
rpl23	111,5	88	449	300,5	0,2995	0,34782	0,8610776	0,0167936
rpl23 3'-rpl2 5'	422	166,5	1571	969	0,2875	0,1805595	1,5922729	0,2877617
rpl2-3'	395	159,5	1190	680,5	0,3665	0,264957	1,3832433	0,457055
rps19	296	189,5	450	448	0,715	0,536052	1,3338258	0,0910881
rpl16	211	123,5	599	309	0,364	0,431706	0,8431664	0,0047146
rpl14	222	136	1129	591	0,219	0,1928355	1,135683	0,0477791
rpl15	505	183,5	1680	955,5	0,297	0,152427	1,9484737	0,2770182
rps8-5'	231	141	944	533,5	0,277	0,2992275	0,9257171	0,0188034
rps8-3'-psaAex1-trnM1	236,5	100,5	1737	1224,5	0,1465	0,1150875	1,2729445	0,3428742
trnG1-rps4	270	115,5	755	600,5	0,4065	0,198462	2,0482511	0,3032434
rps4-3'	180	70	544	358,5	0,352	0,2961585	1,1885528	0,3525569
IG rps4-rrnS-1	755,5	216	2353,5	1398,5	0,3435	0,1611225	2,1319183	0,1745625
IG rps4-rrnS-2	595	486,5	1071,5	940	0,587	0,496155	1,183098	0,2015881
rrnS-5'	12091,5	2262,5	52126,5	37675,5	0,2925	0,0588225	4,972587	0

Name	Medi-an(F635 Median - B635)_ PPR7 [2]	Medi-an(F635 Median - B635)_ pre	Medi-an(F532 Median - B532)_ PPR7	Medi-an(F532 Median - B532)_ pre	Medi-an(Median of Ratios (635/532)_ PPR7	Medi-an(Median of Ratios (635/532)_ pre	Diff. Enrich. [3]	ttest [4]
rrnS-3'-1	1776	445	8889	6147,5	0,2465	0,063426	3,8864188	0
rrnS-3'-2	508	132,5	2115,5	1680	0,272	0,084909	3,2034296	0,002431
trnI-trnA	409,5	193	3723	2602	0,1085	0,069564	1,5597148	0,2084981
rrn7-rrn3-rrnLex1	4106	1852,5	38711	28568,5	0,099	0,0639375	1,5483871	0,7452699
rrnL ex1	3932	2692	35936,5	25513	0,113	0,09207	1,227327	0,4912798
rrnL ex1-in	1204	450	11083	6406,5	0,113	0,079794	1,4161466	0,7406813
rrnLin-ex2-rrn5	164	167	897	728,5	0,186	0,251658	0,7390983	0,0430607
IG rrn5-psbA-1	186	60,5	735	408	0,261	0,203577	1,2820702	0,7504445
IG rrn5-psbA-2	415	268,5	783,5	536	0,5425	0,4404015	1,2318305	0,4639997
psbA ex5-in4	169	89,5	487,5	380,5	0,3915	0,359073	1,0903075	0,1963068
psbA in4	448,5	270	1476,5	986	0,315	0,2777445	1,1341359	0,4604607
psbA in4 ex4	488	319	1920,5	1147,5	0,2805	0,2286405	1,2268168	0,2576914
psbA ex4 in3	1880	629	11415	7060,5	0,154	0,0905355	1,7009902	0,0064965
psbA in3 ex3 in2	470,5	230,5	2214	1457,5	0,2215	0,1928355	1,1486474	0,7443097
psbA in2 ex2 in1	507	309	2842,5	1816,5	0,18	0,17391	1,0350181	0,0629065
psbA in1 ex1	297,5	104	1531	780	0,211	0,1396395	1,5110338	0,9451379
psbAex1 5'	762,5	282	7060,5	4481	0,113	0,0618915	1,8257758	0,2220616
IG psbA-trnS2-2	242	150,5	500,5	392,5	0,5295	0,3595845	1,4725329	0,2256916
IG psbA-trnS2-3	370	206,5	1347,5	804	0,2785	0,2419395	1,1511142	0,3687081
trnS2-ycf12-atpE-3'	699,5	190	13847,5	8623	0,05	0,024552	2,036494	7,93E-05
atpE	421,5	205,5	2295	1482	0,196	0,137082	1,4298011	0,4780914
rps7	273	122	678	422,5	0,409	0,3391245	1,2060467	0,2549178
rps7-5'-rps14	140,5	134	467	388,5	0,3365	0,4976895	0,6761244	0,006796
psbM	447	197,5	1476	730	0,303	0,308946	0,9807539	0,0820849
psbZ	169	129,5	565	345	0,358	0,3432165	1,0430734	0,0677019
IG psbZ-ccsA	897,5	785	3993	2037,5	0,2355	0,3964125	0,5940781	0,002419
ccsA-3'	215,5	152,5	947	679	0,2315	0,2306865	1,0035264	0,0690788
ccsA 5'-rrnL2 3'	462,5	208	1030,5	622	0,4035	0,3524235	1,1449293	0,331455
trnL2-psaAex3-1	255,5	117	2126,5	1489	0,132	0,110484	1,1947431	0,5910671
psaA ex3-2	568,5	266,5	2950,5	2114	0,2035	0,1283865	1,5850576	0,1212163
psaA ex3-3	837,5	293	5512,5	3280	0,149	0,0854205	1,744312	0,1765402
wendy disabled 3'	358,5	83	837,5	418	0,4285	0,249612	1,7166643	0,3891083
wendy disabled	332	108	1438	742	0,243	0,1242945	1,9550342	0,7093176
wendy disabled 5'-trnE2 3'	314,5	182,5	1384,5	1149	0,2335	0,1508925	1,5474593	0,9780224
trnE2 5'- psbH	908	214	16889	15879	0,052	0,015345	3,388726	0
psbH 5'-psbN	598	274,5	2719,5	1975,5	0,215	0,1447545	1,4852733	0,881707
IG psbN-psbT	224	107,5	841	520,5	0,28	0,2552385	1,0970132	0,5533193
psbT-psbB 3'	413	269	1270	1010,5	0,332	0,193347	1,71712	0,6923771
psbB	966	310,5	7761	5324,5	0,131	0,0690525	1,8971073	0,0067257
psbB 5'-trnD	875	480	6706	3726,5	0,141	0,1038345	1,3579302	0,7969903
rpoA 3'-1	305	157,5	1460,5	856	0,2405	0,182094	1,3207464	0,2699298
rpoA 3'-2	233	267,5	561	307	0,469	0,9989595	0,4694885	8,11E-04
rpoA-5'	283,5	171	724	568,5	0,439	0,318153	1,3798393	0,3866909
rps2C-3'	255	201	624	378,5	0,423	0,463419	0,9127809	0,02281
rps2C-5'-rps2 like 3'	329	237,5	953,5	564	0,3315	0,425568	0,7789589	0,0116287
rps2 like 5'	509	254	2107	1138,5	0,2445	0,205623	1,1890693	0,1674048
rps18	217,5	132	495	295,5	0,497	0,4178955	1,1892925	0,130976
rps18-5'-ycf3	469	185,5	1080	563,5	0,45	0,326337	1,3789426	0,4234681
ycf3 5'-ycf4-rps9 3'	396,5	172,5	1340,5	630	0,3085	0,256773	1,2014503	0,1006493
rps9	294	199	805,5	548,5	0,345	0,316107	1,0914026	0,04248
psbE-M2	596,5	270,5	2243,5	1329	0,252	0,1713525	1,4706526	0,482395
rpoB2-3'	248,5	184,5	607,5	356	0,41	0,544236	0,7533496	0,0098367
rpoB2-5'	371,5	204	795,5	473,5	0,528	0,403062	1,3099722	0,1372561
rpoB2-5'-rpoB1 3'	394	236	933	693	0,418	0,427614	0,9775171	0,6283004
rpoB1-3'	257	237	569	308	0,463	0,923769	0,5012076	0,0033186
rpoB1-5'	212	206,5	421	270	0,529	0,7053585	0,7499732	0,0076552
rpoB1 5'-psbF 5'	791	447,5	1533	858	0,49	0,500247	0,9795161	0,1956453
psbF-psbL	1095	481,5	1396,5	714,5	0,777	0,5744145	1,3526817	0,9878205
petG-rps3 5'	321	248,5	940	450	0,358	0,4680225	0,7649205	0,005851
rps3-5'	142	159	347,5	218	0,51	0,801009	0,636697	0,0032366

Name	Median(F635 Median - B635)_ PPR7 [2]	Median(F635 Median - B635)_ pre	Median(F532 Median - B532)_ PPR7	Median(F532 Median - B532)_ pre	Median(Median of Ratios (635/532)_ PPR7	Median(Median of Ratios (635/532)_ pre	Diff. Enrich. [3]	ttest [4]
rps3-3'	130	70	359	173,5	0,378	0,518661	0,7287997	0,0064201
IG rps3-rpoC2	157,5	188,5	530	505,5	0,325	0,3923205	0,8284043	0,0194878
rpoC2-5'	213	115,5	470,5	298	0,4585	0,5539545	0,8276853	0,0080366
rpoC2-1	181	186,5	244,5	177	0,798	1,164174	0,6854645	0,0104444
rpoC2-2	151	211,5	457,5	290,5	0,396	0,718146	0,5514199	0,0038613
rpoC2-3	257	305	557	336,5	0,528	0,9784995	0,5396017	0,0025204
rpoC2-4	201	135	397	206	0,5195	0,772365	0,6726095	0,0014987
rpoC2-5	876	284,5	10201,5	6814	0,09	0,04092	2,1994135	3,00E-07
rpoC2-6	243,5	296,5	646,5	428,5	0,417	0,6521625	0,6394112	0,0033314
rpoC2-7	204	322,5	381	196	0,576	1,56519	0,3680064	7,43E-04
rpoC2 3'-trnY	322	198,5	7188	3989	0,0445	0,056265	0,7909002	0,0299774
trnQ	444,5	234,5	2674	1210,5	0,18	0,17391	1,0350181	0,0477948
psaB-5'	521	269,5	2773	1583	0,197	0,182094	1,0818588	0,5482733
psaB 3'	253	93	1286,5	808,5	0,1855	0,1396395	1,3284207	0,2085547
psaB 3'-trnG2-rbcL 3'	197,5	193,5	1203,5	1066,5	0,181	0,212784	0,8506279	0,1472769
rbcL	386	197,5	807,5	481	0,4745	0,4107345	1,1552475	0,4389928
rbcL 5'-atpA 5'	3332,5	1019	16108	9867	0,2105	0,0997425	2,1104344	1,39E-04
5' atpA	354	185	2711	1798,5	0,143	0,1099725	1,3003251	0,6772686
atpA-3'-psbI	781,5	443,5	1711	788	0,4445	0,528891	0,8404378	0,05291
cemA 5'	244	251	753,5	460	0,3455	0,5334945	0,6476168	0,0060413
cemA 3'-atpH	2613,5	542,5	8238	5244	0,3245	0,1109955	2,9235419	5,32E-12
IG atpH-atpF	722,5	408	2888	1319	0,2475	0,209715	1,1801731	0,4590557
atpF 5'	231	218	1480,5	820	0,173	0,2112495	0,8189368	0,0117878
atpF 3'	1092,5	606,5	8215	4844,5	0,14	0,111507	1,2555266	0,3915297
IG atpF-rps11-trnK	212	101	857,5	494,5	0,267	0,189255	1,410795	0,7681328
trnK	125	NaN	63	NaN	1,738	-	-	-1
tscA	652,5	174	3524,5	2458,5	0,185	0,078771	2,3485801	3,77E-06
chlN 5'	184,5	123	363,5	253	0,523	0,670065	0,7805213	0,0227086
chlN 3'	262	136,5	859	470	0,401	0,4424475	0,9063222	0,0102033
chlN 3'	-	-	-	-	-	-	-	-
atpB-3'	225	234	464	387,5	0,4655	0,6848985	0,6796628	0,0161861
atpB 5'- ORF1995 3'	665	263,5	2417	1327,5	0,2805	0,171864	1,6321045	0,9715122
ORF1995-1	227	177,5	380,5	257,5	0,614	0,739629	0,8301459	0,039283
ORF1995-2	148	99	210	138,5	0,7415	0,785664	0,9437877	0,0190437
ORF1995-3	185	172	453,5	226	0,413	0,7268415	0,5682119	0,0028294
ORF1995-4	331,5	225,5	619,5	429	0,5655	0,5089425	1,1111275	0,1978812
ORF1995-5	161,5	195,5	327	252	0,588	0,8127735	0,7234488	0,0066615
ORF1995-5'	225,5	293,5	645,5	420,5	0,3785	0,575949	0,6571762	0,0106748
IG ORF1995-rps12	254,5	139,5	712	485	0,3655	0,3278715	1,114766	0,032569
rps12-psaJ	194,5	129	329,5	210,5	0,5845	0,722238	0,80929	0,0107487
psaJ 5'-atpI	1565	420	2287	1333	0,64	0,284394	2,2503991	3,33E-05
atpI 5'-psbJ	896,5	348	2133,5	1074	0,4055	0,290532	1,3957154	0,6925954
IG psbJ- psaA ex2	461,5	284	949,5	648	0,492	0,4117575	1,1948781	0,6732737
psaA ex2-psbD 3'	290,5	216,5	1070	736,5	0,2555	0,330429	0,7732372	0,0124169
psbD 5'	1632,5	795	13975	7651,5	0,1295	0,096162	1,3466858	0,5376746
ORF2971-5'	239,5	169,5	778	499,5	0,3465	0,3432165	1,0095668	0,028377
ORF2971-5' 1	237,5	172,5	423,5	291	0,58	0,63426	0,9144515	0,0372745
ORF2971-2	1264,5	572,5	14463,5	7771,5	0,088	0,0731445	1,203098	0,8871409
ORF2971-3	241	218	354	239	0,69	0,976965	0,7062689	0,0086461
ORF2971-4	217	260	338	205	0,6035	1,476189	0,408823	8,93E-04
ORF2971-5	199	274	347	202	0,5625	1,425039	0,394726	0,0015723
ORF2971-6	244	318	401	246	0,637	1,0879605	0,5854992	0,0017255
ORF2971-7	185	123	345,5	227	0,5545	0,577995	0,9593509	0,0218586
ORF2971-8	220,5	216,5	402	294	0,531	0,7933365	0,6693251	0,0056261
ORF2971-3'	630	1076,5	5769,5	3078	0,1205	0,309969	0,3887486	0,0012706
psbC-5'	119	NaN	38	NaN	1,379			-1
psbC 3'- trnH-trnfM	402	206,5	11171,5	7748,5	0,036	0,034782	1,0350181	0,5385624
trnF-psaC	807,5	274,5	4066,5	2488	0,206	0,1560075	1,3204493	0,0685475
petL- trnN	259,5	203,5	2980,5	1828,5	0,085	0,1017885	0,8350649	0,0275162
IG N-rpoC1-2-1	188	162	635	401	0,3445	0,4434705	0,7768273	0,0129344

Name	Median(F635 Median - B635)_ PPR7 [2]	Median(F635 Median - B635)_ pre	Median(F532 Median - B532)_ PPR7	Median(F532 Median - B532)_ pre	Median(Median of Ratios (635/532)_ PPR7	Median(Median of Ratios (635/532)_ pre	Diff. Enrich. [3]	ttest [4]
IG N-rpoC1-2-2	220,5	158,5	431,5	306,5	0,5535	0,4802985	1,1524083	0,0893856
rpoC1b-3'-1	218	147	627	460	0,3935	0,524799	0,7498109	0,0317377
rpoC1b-3'-2	193,5	162	270	184	0,754	0,856251	0,8805829	0,0161861
IG rpoC1- rpoC2	249	204,5	405	294,5	0,597	0,790779	0,7549518	0,0100541
rpoC1-1-1	269	103,5	667	614	0,411	0,229152	1,7935693	5,80E-04
rpoC1-1-2	207	195	484,5	263	0,436	0,7646925	0,5701638	0,001687
rpoC1-1 5' -trnV	366,5	205,5	3088	2071	0,119	0,1345245	0,8845972	0,3950361
ORF271 3'	198	70	420,5	326	0,488	0,3125265	1,5614676	0,9131703
ORF271 5'-ORF140 3', Wendy	156	74	343	261,5	0,454	0,3964125	1,1452717	0,4150178

[2] Experiments with pre-immune serum were carried out with a microarray that has only a six-fold replication of each PCR-product, while the experiments with anti-PPR7 antibody were done with an array of 12-times replicated PCR-products; thus the different maximum spot counts (24 versus 48).

[3] Data also shown in Figure 3.10 B.

[4] Two-tailed, unpaired t-test, median (median of ratios [635/532]) as in Schmitz-Linneweber et al. (2005).

CURRICULUM VITAE

Name: ABDULLAH JALAL

Born: October 10, 1979; Peshawar, Pakistan

Professional and research experience:

Oct 2007 - Mar 2012	PhD thesis at Ludwig Maximilians University Munich, Prof. Dr. Jörg Nickelsen
July 2007 - Sept 2007	Field research at Cereal Crops Research institute (CCRI) Nowshera, Pakistan
Dec 2003 - June 2007	Leaf Asistant Manager at Lakson Tobacco Company Ltd. Mardan, Pakistan
Mar 2002 - Mar 2003	M.Sc. (Hons) thesis at KPK, Agricultural University, Peshawar, Prof. Dr. Hidayat-ur-Rahman

Education::

Mar 2001 - Mar 2003	Masters (Hons) Agriculture at KPK, Agricultural University, Peshawar.
Feb 1997 - Feb 2000	Bachelors (Hons) Agriculture at KPK, Agricultural University, Peshawar.
Sep 1994 - Sep 1996	F.Sc. (Pre medical) Board of Intermediate & Secondary Education Peshawar
Mar 1983 - Mar 1994	Secondary School Certificate (Matriculation) at F.G boys pulic high school, Peshawar

Scholarships/Awards:

Oct 2007 - Sep 2011	HEC-DAAD scholarship award for PhD studies at Germany
Mar 2002 - Mar 2003	Merit scholarship award for Master’s research at KPK Agricultural University, Peshawar

PUBLICATIONS AND CONFERENCE ABSTRACTS

Conference-Poster

Jalal, A., Bohne, A.-V., Schwarz, C., Vallon, O., Schmitz-Linneweber, C. and Nickelsen, J. (Sept. 2009). Functional Analysis of PPR (Pentatricopeptide repeat) Protein Family in *Chlamydomonas reinhardtii.* LEOPOLDINA-SYMPOSIUM, Institute of Biology, Free University Berlin, Germany.

Jalal, A., Bohne, A.-V., Schwarz, C., Vallon, O., Schmitz-Linneweber, C. and Nickelsen, J. (Sept. 2009). Functional Analysis of PPR (Pentatricopeptide repeat) Protein Family in *Chlamydomonas reinhardtii.* XIIIth Annual Meeting of the German Section of the International Society for Endocytobiology (ISE-G-Meeting 2009), FU-Biologie, Berlin, Germany.

Review-Paper

Bohne, A.V., Schwarz, C., **Jalal, A.**, Ossenbühl, F. and Nickelsen, J. (2009) Control of organellar gene expression in *Chlamydomonas reinhardtii* – future perspectives. Endocyt. Cell Res. 19, 70-80

ACKNOWLEDGMENT

First and foremost, I lay down my grateful estrangement before Allah Almighty let me with encouragement to take this job to its end, and the respect for his last Prophet Muhammad (peace be upon him) for enlightening and guiding the mankind to the true path of life.

From the core of my heart, deepest sense of gratitude is forwarded to my supervisor Prof. Dr. Jörg Nickelsen, for providing the opertunity to conduct this exciting research work under his supervision. His dynamic supervision, lasting mood of keen interest and constructive criticism brought the material exposed in this thesis to its final shape. Even in his countless engagements he was present all the time to help me.

Special thanks are extended to Prof. Dr. Jürgen Soll for taking over the second position in the exam commission.

I feel my vocabulary limited in finding words for thanking Dr. Alexandra-Viola Bohne for useful ideas and genuine help in conducting the research. Her critical suggestions and consistent advice throughout the entire PhD project helped me a lot in my proceedings.

I would like to thank Prof. Dr. Christian Schmitz-Linneweber, Humboldt University Berlin, for carrying out the RIP-chip analysis and Dr. Olivier Vallon for Bioinformatical information and also for their valuable suggestions in my research project.

Heartiest thanks to all my former and current lab group members, Anna, Birgit, Christian Schwarz, Christian Stelljes, Fei, Karin, Lin, Marco, Michael and Xia, who always gave me hands in difficult times and maintained an excellent atmosphere in the lab.

I am greatly thankful to Higher Education Commission of Pakistan (HEC) and Deutscher Akademischer Austausch Dienst (DAAD) for providing financial support for the project. I would like to acknowledge all my friends and colleagues at Germany and Pakistan for providing the company, help and moral support during my PhD studies.

At the end, I express my deepest feelings of gratitude to my kind and respectable parents whose care, support and valauble advice remained everlasting assets throughout my career. I am inspired to acknowledge my amazing family members, available all the time to uphold me. Final thanks to Zelle for everything.

EHRENWÖRTLICHE VERSICHERUNG/ERKLÄRUNG

Ehrenwörtliche Versicherung

Ich versichere hiermit ehrenwörtlich, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Beihilfen angefertigt wurde.

München, den 27.03.2012 ABDULLAH JALAL

Erklärung

Hiermit erkläre ich, dass diese Dissertation weder ganz noch in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt wurde. Weiterhin habe ich mich nicht anderweitig einer Doktorprüfung ohne Erfolg unterzogen.

München, den 27.03.2012 ABDULLAH JALAL