Thesis for the Master's degree in Molecular Biosciences
Main field of study in Molecular Biology

Development of a reporter gene construct that can be used for the localization of RNA pyrophosphohydrolase (rppH) in cells of *Chlamydomonas reinhardtii* cells.

Biruk Luelseged

60 study points

Department of Molecular Biosciences
Faculty of mathematics and natural sciences

UNIVERSITY OF OSLO

December/2012

Acknowledgement

This work was carried out during the period of May 2012 - November 2012 at the University of Oslo, department of Molecular Biosciences.

I would like to thank Professor Uwe Klein for his guidance and assistance in writing of this thesis.

I would also like to thank the lab members for nice time.

Finally I would like to thank Beate, Reier, Erland and Anja for their financial support, and constant heartfelt encouragement.

Oslo, December 2012

Biruk Luelseged

Abstract

The concept that messenger RNA (mRNA) degradation in *E. coli* begins with endonucleolytic cleavage has been challenged by the recent discovery that the conversion of the 5' terminus from a triphosphate to a monophosphate is required prior to endonucleolytic activity. RNA pyrophosphohydrolase (RppH) initiates the degradation of transcripts by removing pyrophosphate from the 5'-end of mRNAs which allows binding of RNase E/ RNase J in bacteria. A putative RNA pyrophosphorylase is present in cells of the unicellular green alga *Chlamydomonas reinhardtii* and several lines of evidence suggest that the protein is involved in mRNA degradation in the chloroplast. In order to determine the location of the protein in *Chlamydomonas* cells, the 5' region of the *rppH* gene was tagged to a codon optimized green fluorescent protein and introduced into *Chlamydomonas* cells.

This study focuses on developing a reporter vector construct that can be used to transform the *Chlamydomonas reinhardtii* nuclear genome. GFP (*Zsgreen* 1) was codon optimized by using a Codon Usage Database. The reading frame of the optimized synthetic GFP was adjusted and cloned into the pBluescript-5'*rppH* SK+ vector. The 5'*rppH-GFP* gene fragment was then sequentially cloned into the pE coli-Cterm 6HN protein expression vector, intermediate vector and finally in the pChlami transformation vector. All constructs were verified by restriction cutting and sequencing. Transformants were screened for the presence of the chimeric *GFP* gene by PCR. The screening experiments have led us to choose one transformant that could be used for further work. In addition although fluorescence was not checked, the optimized synthetic GFP was expressed in *E. coli* (BL21DE3) cells confirming the functionality of the synthetic GFP construct *in vivo*.

Table of Contents

1. INTRODUCTION

1.1 *Chlamydomonas reinhardtii*

Chlamydomonas is a genus of unicellular green alga. Morphologically the cells are spherical or ellipsoidal with a size of ~ 10 µm in diameter and have two flagella of the same size that are situated at the anterior end of the cell. They are found mostly in fresh water, on damp soil, and there is a report that few occur also in the sea.

Chlamydomonas reinhardtii is one of the most widely used model organisms. Cells have the ability to grow quickly with a generation time of approximately 5 hours. They are haploid and can reproduce asexually or sexually.

It has been used as a model organism for photosynthesis and chloroplast biogenesis studies, mitochondrial biogenesis, gametogenesis and mating, assembly of flagella and motility, and cellular metabolism. It is an excellent system to study mutations as it has only a single copy of each gene and chloroplast, mitochondrial and nuclear genomes have been sequenced (Maul et al. 2002, Merchant et al. 2007, Smith and Lee 2008). Cells have the capacity to grow with light as sole energy source or on acetate in darkness, facilitating detailed examination of genes and proteins critical for photosynthetic or respiratory function. This haploid organism grows rapidly on both solid and liquid medium (Grossman et al. 2003).

1.2 *Genetics and Chlamydomonas*

Whole-genome sequencing provides a platform for identification of mutations in *Chlamydomonas* (Dutcher et al. 2012). Besides this, *C. reinhardtii* is one of the best studied alga in circadian rhythm research (Matsuo et al. 2008) which are biological rhythms that continue under constant conditions of light and temperature with a period of about 24 hour (Schulze et al. 2010). Biogenesis and the mode of action of miRNAs and siRNAs were also studied in *C. reinhardtii* (Ibrahim et al. 2010). Based on the degree of compelimentary of their small ~ 21 nt sequence and their target mRNA, miRNAs exert or regulate gene expression by cleaving or degrading mRNAs. The degradation effect of these miRNAs is highly observed in plants, in contrast animal miRNA are involved in translational repression. Despite the whole genome complexity differences between those of higher organisms and *Chlamydomonas* cells, the unicellular algal genome has been shown to constitute many of the small RNAs. Strikingly the miRNAs identified in *C. reinhardtii* share structural and functional homology with both plant and animal miRNAs (Zhao et al. 2007).

Based on the completed annotation of the *C. reinhardtii* genome, micro array platform was designed to validate the function of genes that are related to efficient light harvesting and photosynthetic electron transport (Toepel et al. 2011). Because chlorophyll synthesis also occurs in the dark in *C. reinhardtii*, the photosynthetic apparatus can be assembled in the

absence of light. Therefore it is possible to study photosynthetic reactions in light-sensitive mutants (Rochaix 2002, Nishimura 2010).

It is amenable to a diversity of genetic and molecular manipulations. In addition engineering of the plastid genome was first undergone in *C. reinhardtii* (Boynton et al. 1988). The recent sequencing of the nuclear genome and the availability of numerous molecular tools including transformation of the three (nuclear, plastid and mitochondrial) genomes, makes *C. reinhardtii* an attractive model for molecular investigations.

1.3 *Organelle Gene regulation*

Chloroplast genomes encode the most important genes for photosynthesis (Maliga and Bock 2011). The genome has uniform densities of genes, simple sequence repeats and transposable elements (Merchant et al. 2007). Chloroplast gene expression in *C. reinhardtii* is a complex process that can be influenced by mRNA processing, mRNA stability and protein turnover (Rasala et al. 2011). But unlike nuclear genes, genes encoded by chloroplast are not affected by transcriptional silencing and position effects (Michelet et al. 2011). The *Chlamydomonas* chloroplast genome has only a few introns located in the *psb*A, *psa*A, and 23S rRNA genes (Herrin and Nickelsen 2004).

Studies on higher plants have shown that there are two RNAPs (RNA polymerases), NEP (Nuclear encoded RNA polymerase) and PEP (Plastid encoded RNA polymerase). PEP has a catalytic core consisting of α, β, β′ and β′′ subunits which is thought to require σ^{70}-like factors that specifies initiation of transcription downstream of promoter sequences which resemble bacterial 10/35 promoters (consensus: -10 TATAAT,- 35 TTGACA). Chloroplast gene expression is also dependent on factors that are encoded by the NEP. Characterizing this polymerase has revealed that it is a single catalytic subunit polymerase homologous to the T7/T3 phage polymerases. Unlike those of higher plants, *Chlamydomonas* lacks NEP. A recent study has shown that *C. reinhardtii* expresses a single σ^{70}-like factor which likely functions in chloroplast transcription (Bohne et al. 2006).

1.3.1 Anterograde/Retrograde signaling

The chloroplast genome encodes proteins that are involved in transcriptional and translational apparatus. Concentration of chloroplast proteins and their function is dependent on the anterograde (nucleus to chloroplast) and retrograde (chloroplast to nucleus) signaling mechanisms. Anterograde mechanisms coordinate gene expression in organelles in response to endogenous and environmental stimuli that are perceived by the nucleus. Retrograde mechanisms transmit signals that originate in the organelles to regulate nuclear gene expression, which can then alter anterograde control (Woodson and Chory 2008).

The continuous flux of proteins into and out of the nucleus as well as among intra-nuclear compartments controls central events such as DNA replication and mRNA processing (Gorski et al. 2006). In *C. reinhardtii* many DNA-binding proteins directly or indirectly involved in the regulation of gene expression have been shown to be at least transiently located in the

nucleus (Winck et al. 2012). Nucleus encoded proteins in *Chlamydomonas* are the basis for anterograde signalling that enables the nucleo-cytoplasmic compartment to control chloroplast gene expression (Michelet et al. 2011).

The nuclear genome of *C. reinhardtii* is 100 to 110 million bp, comprising 17 genetic linkage groups (Kathir et al. 2003) presumabily corresponding to 17 chromosomes (Merchant et al. 2007), with a very high GC content (nearly 65%). *Chlamydomonas* nuclear genome is haploid, in contrast to approximately 80 copies of the chloroplast genome (Eberhard et al. 2011). The number of selection markers available for genetic engineering in *Chlamydomonas* is relatively small compared to those for other model systems. Foreign genetic material introduced to *C. reinhardtii* nuclear genome recombines randomly at any site in its genome (Kindle et al. 1989), which may result in deletion of functionally important genes. Moreover transgene expression level is very low even if the transformation reaction is highly efficient due to the following possible reasons; the first being inadquate amount of promoter and regulatory regions affecting the transcription level of the gene (Leon-Banares et al. 2004) of interest. In addition epigenetic silencing of transgenes, especially single copy transgenes can be transcriptionally silenced without detectable cytosine methylation (Jeong et al. 2002). The other being due to codon usage bias; codon frequently used in certain organisms is rarely used in others (Nakamura et al. 2000). This has hampered the applicability of *C. reinhardtii* for biotechnological expression analyses. Recently new genetically engineered resistance markers were developed (Meslet-Cladiere and Vallon 2011).

1.4 *Chloroplast mRNA Turnover*

The level of mRNA within a cell depends on the balance between transcription and degradation. mRNA is the most varied class of RNA with respect to its size and stability. The activity of RNA is determined by its structure (Zuker et al. 1999). As such the stability of a given mRNA transcript can be determined by the presence of sequences within the 5’ UTR (untranslated region) (Salvador et al. 2004), which can be bound by trans-acting RNA-binding proteins to inhibit or enhance mRNA degradation (Hollams et al. 2002). In addition the structural characteristics of 5' untranslated region, such as length, the presence of AUG upstream of the initiator greatly affect the efficiency of the translational process (Pesole et al. 2000) and thereby the longevity of the transcript. Moreover sequences that are found in the protein coding region have been shown to control promoter activity in *Chlamydomonas* chloroplast *rbcL* gene (Klein et al. 1994).

Regulated mRNA stability is achieved through fluctuations in half-lives in response to environmental stimuli like nutrient levels and temperature shifts (Guhaniyogi and Brewer 2001). In bacteria for example, mRNA half-lives can be in seconds or in hours, whereas in vertebrates they range from minutes to days. The half-lives of stable RNAs approach or even exceed the doubling times of the cells in which they are made (Meyer et al. 2004); and thus their concentration does not depend on cell growth.

mRNA turnover implys rapid synthesis and equally rapid destruction (Meyer et al. 2004). Selective degradation of messenger RNAs enables cells to regulate the levels of particular

mRNAs in response to changes in the environment (Mackie 1998). The current knowledge of chloroplast mRNA degradation begins by endonucleolytic cleavage which is followed by 3' polyadenylation of the cleavage products. Addition of poly A rich sequences to the endonucleolytic cleavage products of mRNA is required to target these molecules for rapid exonucleolytic degradation (Schuster et al. 1999). The polyadenylated cleavage products are subsequently degraded by 3'-5' exonuclease. The RNA hairpins (secondary structures) that are found at the 3' termini impede the 3'-5' exonucleases, thereby stabilizing upstream RNA (Pfalz et al. 2009). Essential factors that control mRNA turnover in chloroplast such as cis elements, proteins that bind to them, ribonuclease susceptibility, and redox status are briefly discussed below.

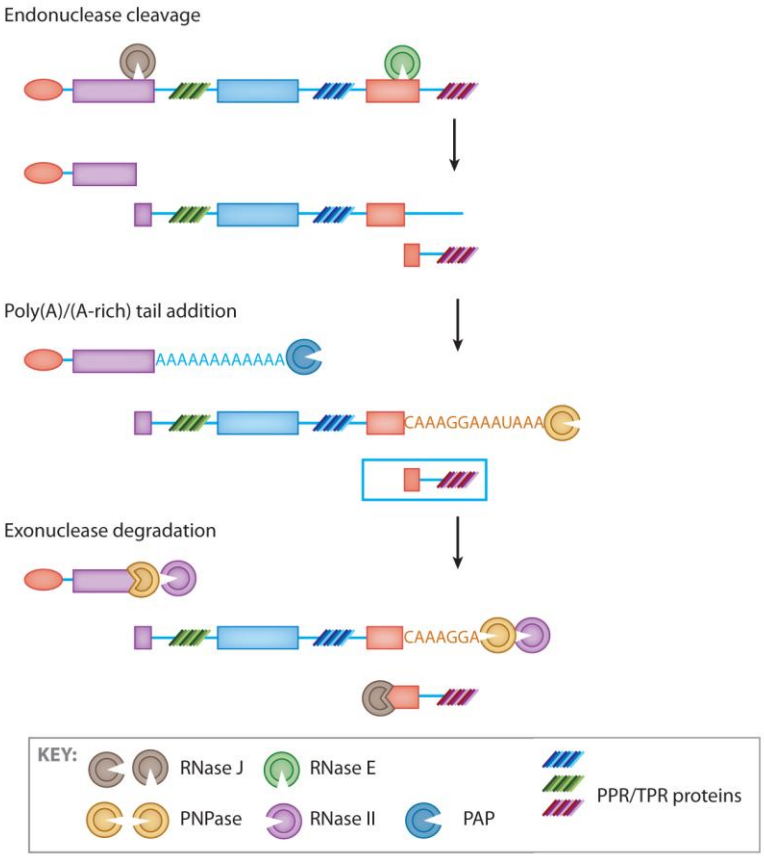


Figure 1.1. Model for chloroplast mRNA degaradation. Endonuclease cleavage is initiated by the RNase J and RNase E enzymes. The polynucleotide tail addition is facilitated by PNPase or PAP. The RNA fragment where the 3' end is blocked by a PPR protein requires a 5'→3' decay. The absence or low frequency of secondary structures at the 3' end can also mediate a 3'→5' decay by PNPase or RNase II. Image modified from (Stern et al. 2010).

1.4.1 Cis acting elements

The function of RNA can only be understood in terms of its secondary or tertiary structure (Zuker et al. 1999). Cis regulatory elements on chloroplast RNA are regions that are required for its stability or its degradation. They are usually found on the 5' and 3' UTR of mRNA. Sequences within the 5'-UTR are essential for translation (Anthonisen et al. 2001). The 5' UTR have been shown to regulate the translation of the *psbC* (Zerges et al. 1997), *psbA* (Trebitsh et al. 2000), *psbD* (Nickelsen et al. 1999) genes. In addition inverted repeat structures which are nucleotide sequences that are the reverse complement of another sequence downstream have been found in chloroplast genome of all *Chlamydomonas* species examined (Harris 2001). These inverted repeats have been shown to determine the stability of chloroplast rbcL mRNA in *C. reinhardtii* (Goldschmidt-Clermont et al. 2008). Sequence and condition-dependent 5′→3′ mRNA-degradation pathway was also shown in the chloroplast of *C. reinhardtii* (Salvador et al. 2011). The requirement of specific 5' sequence for RNA longevity is also determined by an interaction of this element with a trans-acting factor (Suay et al. 2005).

1.4.2 Trans acting factors

Trans acting factors are proteins that are mostly nuclear encoded. They are synthesized in the cytosol and subsequently imported into the chloroplast, where they interact with their target sites on either chloroplast RNAs or proteins (Nickelsen et al. 1999).

Several protein families playing essential roles in the mRNA metabolism in chloroplast are characterized by the occurrence of tandem repeat motifs. PPR members of the pentatricopeptide repeat proteins function in RNA processing, and translation (Johnson et al. 2010, Stern et al. 2010). PPR family, which comprise of 35 amino acid repeats, primarily interact with specific RNA-targets. 11 PPR protein-encoding genes have been identified in *C. reinhardtii* (**UniProt**). The stabilizing mechanism of these PPR protein is by giving a shelter to a specific region of RNA from nucleases, in the manner of a protein cap (Johnson et al. 2010). Functionally they were also implicated in protecting internal transcript sequence elements (Zhelyazkova et al. 2012). Moreover in *C. reinhardtii*, a PPR protein is required for stabilization of the rbcL mRNA (Johnson et al. 2010).

The tetratricopeptide repeat (TPR) is a degenerate 34-amino acid repeat motif, protein-protein interaction module found in multiple copies in a number of functionally different proteins (Blatch and Lassle 1999). For instance, Raa 1 with five copies of octatricopeptide repeat is thought to improve the efficiency of *psaA* mRNA maturation in *C. reinhardtii* chloroplasts (Merendino et al. 2006). NAC2 is also involved in stabilization of *psbD* transcripts (Boudreau et al. 2000).

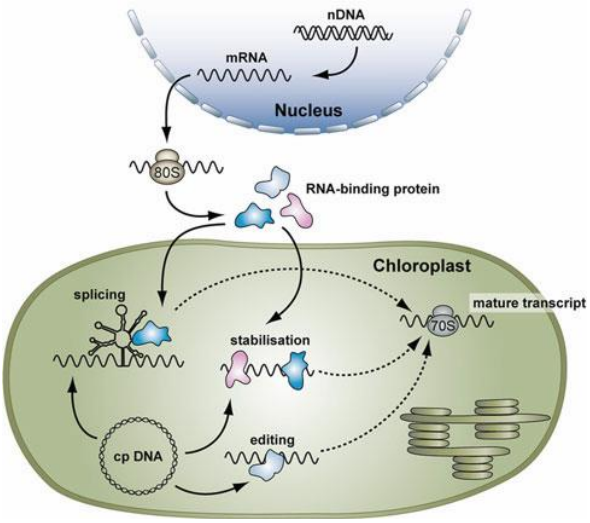


Figure 1.2. Model for post-transcriptional modifications in chloroplasts. Various nuclear encoded RNA-binding proteins participate in RNA processing events by binding directly to their target RNAs. These events involve several steps, such as splicing and stabilization which control chloroplast gene expression (Jacobs and Kuck 2011).

1.4.3 Effect of redox state

During photosynthesis, the source of electron is water. In the photosynthetic machinery, different components are involved in redox reactions. Changes in the reduction/oxidation state of these components are used as signals in gene regulation (Pfannschmidt et al. 2003). Light influences chloroplast gene expression, RNA processing, transcript stability, translation and turnover of proteins (Salvador and Klein 1999). In *C. reinhardtii*, the redox state regulates the degradation of chloroplast transcripts (Salvador and Klein 1999).

1.4.4 Ribonuclease susceptibility

As has been mentioned, the extent to which mRNA is susceptible to degradation by ribonucleases is dependent on its secondary structure. Thus secondary structures are important to prevent mRNA from degradation. Moreover the binding of trans-acting proteins to a target mRNA sequence can change its susceptibility to ribonucleases thereby promoting mRNA degradation. These ribonucleases determine the longevity of mRNA in prokaryotes such as bacteria.

Chloroplast mRNA and that of prokaryotic have certain differences. The distinguishing features are the predominance of introns and complex patterns of processing from polycistronic precursors (Stern et al. 2010). In addition, polyadenylation is associated with RNA instability in prokaryotic cells, whereas in chloroplast nucleus-encoded mRNAs, polyadenylation is to enhance their stability and promote translational initiation (Komine et al. 2000). Moreover the poly (A) tail sequences are composed of clusters of adenosines mostly bound by guanosines, and on rare occasions, by cytidines and uridines (Schuster et al. 1999) whereas in bacteria, the sequences are composed of clusters of adenosine residues.

But the molecular biology of plastids is basically prokaryotic (Salvador et al. 2011). Sequences composed of -35 and -10 elements resembling promoters of *E. coli* genes direct transcription *in vitro* as chloroplast promoters *in vivo* (Blowers et al. 1990). Translation regulation in chloroplasts has been also reported to be very similar to prokaryotes (Kozak 2005). The poly (A) tails found in *Chlamydomonas* chloroplasts are similar in length to those of *E. coli*, being mostly between 20 and 50 nt (Komine et al. 2000). Chloroplast ribosomes and that of bacteria share similarities on the mechanism of protein synthesis initiation and inhibition (Ellis 1970). In addition higher plants plastid genome encodes subunits of an *E. coli*-like RNA polymerase (PEP) which initiates transcription from *E. coli* σ^{70}-type promoters (Hajdukiewicz et al. 1997). Furthermore the ribonucleases involved in chloroplast mRNA turnover are derived from bacteria (Barkan 2011).

1.5 *Bacterial gene regulation*

RNA degradation is a major component of RNA metabolism in determining its intracellular levels during which rapid decay serves as a mechanism to adjust to environmental change (Deutscher 2006).

Most of the mRNA found in bacteria is polycistronic where its genetic information is translated to several polypeptides that are functionally related to each other. The genetic information consists of functional gene clusters, termed as operon. Bacterial operons share similarity wth the polycistronic transcription units of the chloroplast genome (Barkan 2011).

1.6 *mRNA decay in bacteria*

RNA stability, degradation and processing is determined by ribonucleases in bacteria. There are two main classes of ribonucleases; endoribonuclease and exoribonuclease. Endoribonuclease cleave RNA segment at the internal phosphodiester bonds. Whereas the exoribonucleases cleave at the end of RNA chain. Exoribonucleases can act either on the 5’ end or 3’ end of an RNA segment.

1.6.1 Bacterial mRNA stability

Stabilizing elements such as the 5’-triphosphate of primary transcripts and strong secondary structures are crucial in bacterial mRNA stability (Evguenieva-Hackenberg and Klug 2009, Evguenieva-Hackenberg and Klug 2011). 5' secondary stem-loop structurs are able to protect mRNA from attack by cellular ribonuclease that initiates mRNA degradation (Emory et al. 1992, Sharp and Bechhofer 2005). Ribosome binding at Shine–Dalgarno elements adjacent to the translation initiation codon (collectively called ribosome binding site) also serves to protect specific endonucleolytic target sites required for the initiation of decay (Petersen 1992).

1.6.2 Bacterial mRNA processing and degradation

RNA processing and degradation determine transcript accumulation and thus are key processes in the control of gene expression. RNA processing is used to describe RNase-catalyzed events leading to the generation of a functional RNA, while RNA degradation is used to describe its decay, in which RNA processing leads to an increase in RNA half-life relative to the primary transcript whereas RNA degradation leads to a decrease in half life. RNA processing is relevant in the 5' maturation of transcripts. The mRNA processing depends usually, but not always, on the activity of RNases (Lehnik-Habrink et al. 2012).

RNA turnover in *E. coli* requires pathways that involve endoribonucleases and 3'-exoribonucleases that modify RNA or affect its conformation (Carpousis et al. 2009). The decay of most mRNAs in *E. coli* is initiated by RNase E, a 1061-amino acid enzyme which is specific for single stranded regions in mRNA. The endoribonuclease RNase E is conserved among proteobacteria and is essential for viability due to its role in turnover of mRNA in *E. coli* (Anupama et al. 2011) where it acts on mRNA and small non-coding RNA. RNase E is organized into a number of catalytic core domains that are arranged to different sub-domains and has a long non-catalytic C-terminal extension (Carpousis 2007). The larger subdomain is composed of RNA binding domain and 5' sensor. Previous study (Garrey et al. 2009) where the RNA binding subdomain and the 5′ sensor interact to form a certain compartment that enhances 5'monophosphate binding was shown to have increased the RNase E activity. RNase E is important in mRNA decay which cleave single stranded AU rich RNA (Mudd et al. 1990). A 5' end-independent mRNA decay pathway which requires endonuclease activity is also reported (Deikus and Bechhofer 2011).

Exoribonucleases can act on mRNA either in the 5'→3' or 3'→5' direction in bacteria. Some are involed in the maturation and stability of mRNA for instance, RNase J1 (Mathy et al. 2007) whereas others are involved in quality control, or degradation of defective RNAs such as RNase R (Vincent and Deutscher 2009), PNPase or RNase II. *E. coli* cells contain multiple 3'→5' acting exoribonucleases (Arraiano et al. 2010). The 3' end of bacterial mRNAs is protected from 3'-to-5' exonuclease attack by secondary structure (Condon 2007). Thus, the action of endonucleases generate new unprotected 3' ends that are rapidly degraded by exonucleases. Hence since *E. coli* lacks 5' exoribonucleases (Kaberdin et al. 2011), it is believed that endonucleolytic cleavage generally precedes 3′ exonucleolytic degradation (Belasco 2010).

Similarly stem loop (hairpin loop) which is the building block for a secondary structure, at the 5' end of *E. coli* transcripts can inhibit mRNA degradation by RNase E (Mackie 2000). RNase E endoribonucleolytic activity is increased by the 5' terminus state of its target RNA. mRNA stabilization by a 5'-terminal stem-loop has also been reported in *B.subtilis* (Hambraeus et al. 2002). Similar to *E. coli*, the presence of a 5′-triphosphate have stabilizing effects on the downstream genes in *B.subtilis* (Lehnik-Habrink et al. 2012). In both organisms RNA degradation begins with the conversion of the 5'-terminal triphosphate to a monophosphate. The hydrolysis of a variety of nucleoside triphosphates is catalyzed *in vivo* by the Nudix hydrolases.

1.7 *Nudix Hydrolase*

The terminal RNA sequences are modified by various enzymes of different gene families. These modifications are present in organelles. And since RNA metabolism occurs in these organelles as well as in cytosol, characterizing the role that these enzymes have in mRNA turnover is important.

"Nudix" Hydrolase is a family of widely distributed pyrophosphohydrolase enzyme found in all classes of organisms. Catalysis depends on Nudix motif which is 23 amino acids long. The domain structure is formed by two β-sheets packed between α-helices that is well adapted to binding and hydrolysis of a wide range of nucleoside (McLennan 2006). The defining characteristic of Nudix enzymes is their ability to catalyze the hydrolysis of a variety of nucleoside diphosphate derivatives with varying degrees of specificity (Kraszewska 2008). Nudix hydrolases could be considered "housecleaning" genes whose function is to cleanse the cell of potentially deleterious endogenous metabolites and to modulate the accumulation of intermediates in biochemical pathways (Bessman et al. 1996). In *E. coli* hydrolysis of 5' terminal triphosphate to monophosphate is catalyzed by a member of the Nudix hydrolase family called RppH (RNA pyrophosphohydrolase).

1.8 *RNA pyrophosphohydrolase (RppH)*

RNA pyrophosphohydrolase (RppH) formerly designated as NudH/YgdP is a recently discovered protein that has RNA pyrophosphohydrolase activity in *E. coli*. It is a member of the Nudix family with the Nudix signature motif GX5EX7REUXEEXGU where U is a hydrophobic residue and X is any residue (McLennan 2006). It catalyzes the conversion of a 5' triphosphosphate primary transcript to a 5'monophosphate RNA, assisting in the 5'→3' mRNA degradation (Deana et al. 2008). Like in *E. coli*, RNA pyrophosphohydrolase in *B. subtilis* that catalyzes this event is a Nudix protein that prefers unpaired 5' ends. However, in *B. subtilis*, this modification exposes transcripts to rapid 5' exonucleolytic degradation by RNase J, which is absent in *E. coli* but present in most bacteria lacking RNase E (Richards et al. 2011). Yet *Chlamydomonas* appears to encode RNase J (Stern et al. 2010).

Recent findings that goes along with previous works show that the 5'-3' rppH mediated degradation can be regulated by ribosome binding close to the translation start site (Richards et al. 2012). Some evidences suggest that rppH is involved in mRNA degradation in the chloroplast. Recently an *in vivo* study done on degradation of transcripts in a specific reporter gene constructs in *Chlamydomonas* cells shows that initiation of degradation starts at the 5′ end (Salvador et al. 2011). The same study has indicated that it is likely that an exposed 5′ terminal nucleotide with its phosphate group as the binding site of a protein (trans acting factor) that initiates mRNA breakdown in chloroplasts (Salvador et al. 2011) from the 5' end.

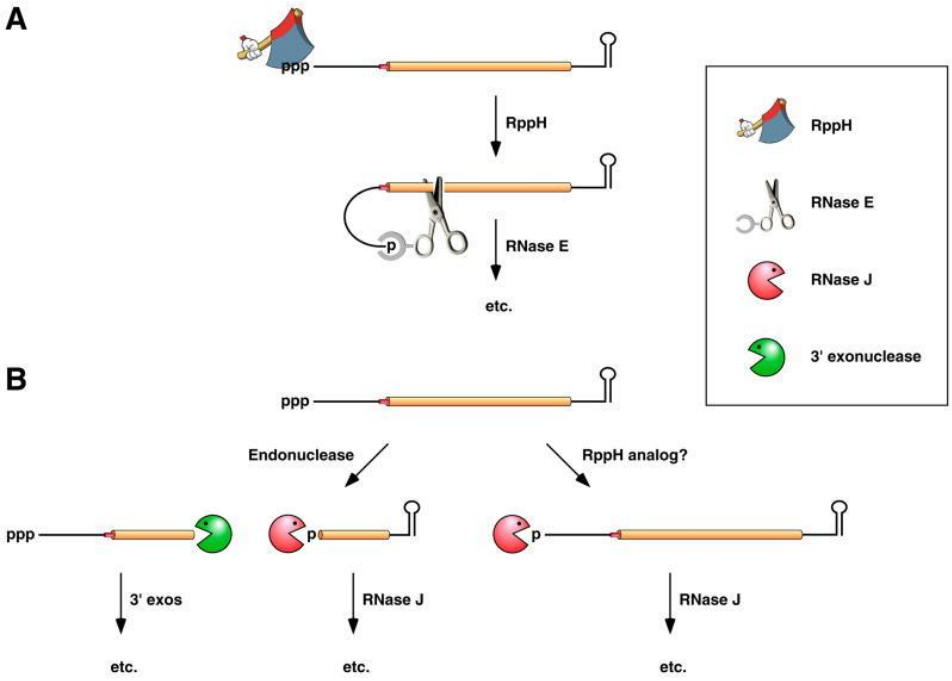


Figure 1.3. Bacterial RNA decay model. A) The removal of pyrophosphate by rppH enhances the the RNase E compartment binding to 5' monophosphorylated mRNA thereby activating degradation of mRNA in *E. coli*. B) RNA decay in bacteria containing the exonuclease RNase J (*B. subtilis*), endonuclease activity exposes the primary transcripts to 3' exonuclease or RNase J attack. In addition pyrophosphate removal by rppH analogue results in degradation of monophosphorylated transcript by 5' exonulease activity of RNase J (Belasco 2010).

1.9 *Sub-cellular localization*

Transcripts are thought to contain localization signals, consisting of discrete stem loop structures found at the 5' end of a gene which are recognized by transacting factor for the screening of localized transcripts among the general population of RNAs .

Cellular asymmetry in *Chlamydomonas* is determined by the anterior and posterior position of the flagella and chloroplast respectively with the nucleus in between these two organelles. The nuclear architecture of *Chlamydomonas* is organized as a series of concentric spheroids, the innermost being the nucleolus that serves as the site for the ribosomal RNA synthesis. Surrounding the nucleolus is a spherical compartment that presumably contains the sites of synthesis and processing of pre-messenger RNAs in *Chlamydomonas.* Nuclear organization is linked to cytoplasmic events such as transcript targeting (Colon-Ramos et al. 2003).

1.10 *GFP*

Bioluminiscence in cnidaria is due to oxidation of luciferin via luciferase or Ca^{2+}-activated photoprotein that excites a class of proteins called green-fluorescent proteins (GFPs) (Prasher et al. 1992). GFP, first named (Shimomura et al. 1962) as "Aequorin", was shown to have

emission of light on addition of Ca^{2+} ions. Unlike the other fluorescent proteins *A. victoria* wild type GFP has chromophores in protonated state.

The green fluorescent protein (GFP) from the jellyfish *A. victoria* is a 238 amino acid protein that exhibits bright green fluorescence under ultraviolet blue light exposure. Neutral wild type GFP has a dual absorption wave length at 395nm and 475nm and emits green light at 509nm (Brejc et al. 1997). Crystallographic structural analyses on the wild type and mutant S65T GFP has shown that, this dual absorption spectra is caused by a change in the ionization state of the chromophore, which is the structural unit of GFP.

GFP tagging permit analyses of proteins in living cells easier and offer distinct advantages over conventional immunofluorescence. Among these are lower background, higher resolution, and avoidance of fixation artifacts (Michaelson and Philips 2006). Various GFP variants have been mutated to give reporters of different spectral properties so that multicolor labeling is now a practical option (Paddock 2008). Transgenic expression of GFP within any given cell requires simply placing the optimized versions of GFP sequence under the transcriptional control of appropriate regulatory sequences and in the correct reading frame (Chytilova et al. 1999). The spectral properties of the recombinant GFP suggest that production of the fluorescence is not species-specific (Chalfie et al. 1994).

1.11 *Aim of the project*

The aim of this project was to develop a reporter gene construct that could be used to localize RNA pyrophosphohydrolase in *C. reinhardtii* cells. The project had also sub-aims;

- ✓ Selection of a reporter Green fluorescent protein (*GFP).*
- ✓ Codon optimization of Z*s Green 1* from *Zoanthus Sp.* to codon usage in the nucleus of *C. reinhardtii.*
- ✓ Linking the codon optimized *GFP* to the 5'*rppH* region which is thought to have the localization signal for RNA pyrophosphohydrolase.
- ✓ Construction of pE coli -5'*rppH-GFP* vector to express *GFP* construct in *E. coli*.
- ✓ Construction of transformation vector pChlami-5'RppH-GFP that could be used to transform *C. reinhardtii*.
- ✓ Nuclear transformation of *C. reinhardtii.*
- ✓ Selection and screening of positive *C. reinhardtii* transformants by PCR.

2. Materials and methods

2.1 *PCR*

By using thermocycler (Biometra), standard PCR procedure was followed to amplify DNA fragments by using Taq polymerase. Amplification of fragments for cloning experiments was carried out in three steps, two denaturation cycles at 98˚C for 1 min and 30 sec respectively, 30 annealing cycles at 65˚C for 30 sec, and two elongation steps at 72˚C for 1 min in 30 reaction cycle. In all PCRs, the primer concentrations were 10 pMol/µL. For genomic DNA (1 ng/ µL) amplifications, the annealing temperature were different for different reactions involving different primers (see Appendix V). Since different primers that were used have different melting temperatures, the annealing temperature was adjusted accordingly.

2.2 *Construction of fusion proteins and vectors*

Standard molecuar biology techniques were used as explained in (Sambrook and Russell 2001), and all restriction enzymes for vector digestion was done according to the manufacturers general guide lines (New England Biolabs). Furthermore, for all restriction site check up; for ligation or specific sequence analyses, was done by NEB cutter V 2.0 (New Englands BioLabs). The following vectors were constructed for the expression of the fusion protein in which the 5'-*rppH* region (207), was fused to the N-terminus of the optimized *GFP*. An extra amino acid (GLY) was added inorder to have the correct reading frame of GFP by including an extra nucleotide [C] just upstream of the start codon of GFP. The synthesized lyophilized plasmid containing *GFP* gene was manipulated according to the manufacturers instruction (Life technologies) for further use. *Sma*I and *Eag*I sites were used for cloning of *GFP* to a pBluescript II SK+-5'*rppH* vector received from Uwe Klein. *GFP* which is flanked by *Sma*I and *Eag*I sites was cloned into pBluescript II SK+-5'*rppH* at the *Sma*I and *Eag*I sites which was digested with *Sma*I and *Not*I restriction enzymes. The 5'*rppH-GFP* gene was then amplified by one PCR thereby introducing the *Nco*I restriction site in the PCR product. See Appendix V for primers.

The PCR product was purified by using gel purification kit (GE Healthcare), cut with restriction enzymes *Nco*l/*Not*l and was cloned into pE coli vector (Clontech) after digesting the vector with *Nco*I and *Not*I enzymes. The 5'*rppH-GFP* fragment was then amplified by 3 PCR reactions to incorporate a new *NdeI* site and replace the *Nco*l site in pE coli vector. Three *Nde*l forward primers and *Xba*l reverse primer were used for the oligonucleotide [CAT] synthesis (see Appendix V). At each interval of the PCR, amplified PCR products were run on agarose gel (1%) and purified. Subsequently each purified PCR products were diluted to meet the starting template amount required for the next PCR.

After purification of the final PCR product from a gel, the *NdeI*/*XbaI* digested PCR product was cloned into the pBluescript II SK+ vector (Uwe Klein). The vector contains *Xho*I-*Xba*I

gene fragment that is taken from pChlamiRNA3int vector (Molnar et al. 2009); for clarity this vector is named as an 'intermediate vector' from now on.

To form the transformation vector, the *Xho*I-*Xba*I fragment which was digested from the intermediate vector was cloned into pChlami vector (Molnar et al. 2009) which had released its original *Xho*I-*Xba*I fragment region after restriction cutting to form the pChlami-*5'rppH-GFP* vector. In the next sections the term pChlami vector is used to refer to the pChlamiRNA3int vector that has released the *rbcS*2-intron sequence.

2.3 *Agarose gel electrophoresis*

Determination of the size of DNA was done by running DNA samples on a 1% agarose gel containing ethidium bromide [0.17 µg/mL] in TAE buffer *(*see Appendix I*)*. A 1Kb plus standard ladder was used to estimate the size of the DNA fragments or PCR products. The gel was run at 90 V for 30 min, and for longer durations when separation of smaller bands were needed. Pictures were taken for further analyses.

2.4 *Ethanol precipitation*

All DNA samples were precipitated by adding 1/10 volume of Na acetate [3M] and 2× volume of 96% ethanol on ice for 30 min. Centrifugation at 4˚ C, washing with 70% ethanol was respectively done afterwards.

2.5 *Bacterial cells*

TB1 *E. coli* strain was the host bacteria that was used for all plasmid and vector transformation in this study except for recombinant protein expression analyses. Procedures for *E. coli* growth, transformation, maxipreps and minipreps was done by following the lab manual and coarse book (Uwe Klein). All *E. coli* transformants were grown on agar plates containing ampicillin (60 µg/mL) as a selection marker, and overnight growth of cultures was done for all on LB medium (See Appendix I) supplemented with ampicillin (100 µg/mL).

2.6 *Transformation of Chlamydomonas*

Nuclear transformation of the cell wall-less *Chlamydomonas* strain *cw*-15 cells, was done by Uwe Klein, by following the protocol of (Kindle et al. 1989). Cells were vortexed in the presence of linearized DNA (transformation vector), glass beads and polyethylene glycol (PEG). Cells were plated out on paromomycin containing agar plates and incubated under sterile condition.

2.7 *Selection of transformants*

Individual colonies were picked once their sizes were clearly visible on the paromomycin (50mg/mL) containing agar plate. Each colony that was picked was marked at the back of the plate.

2.8 *Cell culture*

The mutant strain *cw*-15 of *C. reinhardtii* was obtained from cell culture in *Chalmaydomonas* genetic center at duke university NC, USA. Transformants of the alga grown on paromomycin containing agar plates were inoculated in 100-mL liquid culture. Cells were maintained on HS (high salt) (Sueoka 1960) medium supplimented with [salt stock, phosphate stock and trace elements, see Appendix I] at ~24˚ C on light intensity ~ [500 W/m^2]. Cells were supplimented with 2% CO_2 in bubbling water chamber (water- bath supplimentd with air) until the growth confluency is about ~ 2×10^6 cells/mL at 32˚C in 12 hours light and 12 hours dark cycle. For genomic DNA isolation purposes, cells were collected from the cell culture according to laboratory manual (Uwe Klein).

2.9 *Selective marker*

Different marker genes are used for developing of transformation in *Chlamydomonas*. For this experiment, the *aphVIII* gene (encodes for aminoglycoside 3'-phosphotransferase) was used which confers paromomycin resistance to the *Chlamydomonas* cells. *aphVIII* has a codon usage similar to *Chlamydomonas* genes (GC content of 68.9%), thereby expression of this gene in *Chlamydomonas* cells doesn't creat random mutations (Sizova et al. 2001). Usually expression of a gene from chimeric gene constructs requires strong, constitutive or inducible promoter system. Previously a linearized plasmid used to express *aphVIII* under the control of *RBCS-HSP70* chimeric promoter and the *RBCS* terminator generated ~22,000 *Chlamydomonas* transformants (Gonzalez-Ballester et al. 2005). Moreover high sensitivity of *C. reinhardtii* cells to paromomycin and the efficient inactivation of paromomycin by *aphVIII,* made the *aphVIII* gene a good candidate (Sizova et al. 2001). Concentration of marker gene will impact the number of transformants obtained per transformation event and the number of integrated marker gene copies per transformant.

2.10 *Codon optimization*

Codons used frequently in a certain organism are often not used in others. This causes a change in the number of tRNAs and significantly affects the translation efficiency. GFP (Clontech) which was derived from *Zoanthus Sp*. was opimized to *C.reinherdtii* nuclear codon sequence (Nakamura et al. 2000). Graphical analyses of codon usage was done by using Graphical Codon Usage Analyzer (Fuhrmann et al. 2004).

2.11 *Screening of transformants with PCR*

PCR was used to check whether *C.reinherdtii* wild type cells have integrated the transgene. Template DNAs from *C.reinhardtii* genome of selected transformants were amplified by different primers to analyze the PCR product. The list of primers and the melting temperatures can be reffered (Appendix V).

2.12 *Recombinant protein expression in E. coli (BL21DE3)*

The pE coli expression vector system (Clontech) contains the hybrid T7 *lac* promoter as well as a *lacI* gene, which encodes Lac repressor. In lac hybrid promoter system (Clontech), basal expression of the protein of interest is repressed by the Lac repressor (*lacI*), that binds to the lac operator, preventing expression of genes from the promoter p lac in the absence of IPTG. Isopropyl β-D-1-thiogalactopyranoside (IPTG) is a lactose metabolite in which up on addition, causes allosteric conformational change in the repressor (*lacI*), allowing T7 RNA polymerase to gain access to the promoter and initiate transcription of gene coding for beta-galactosidase.

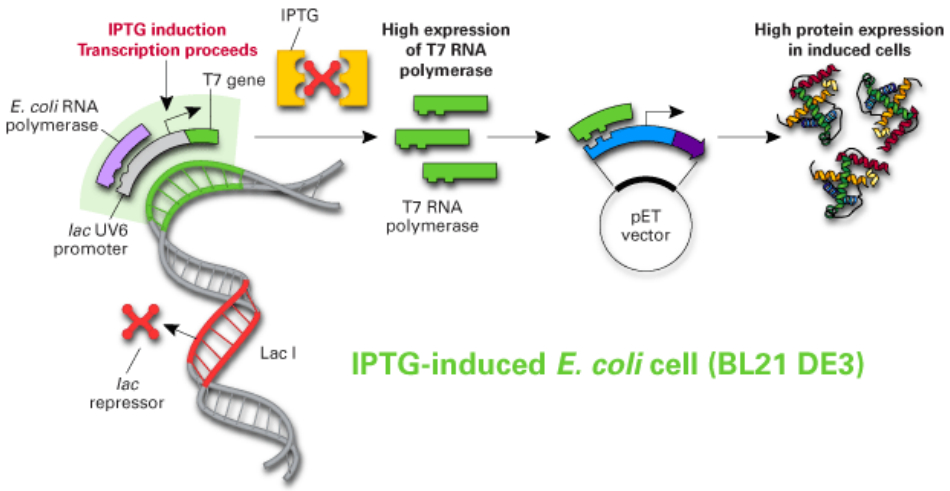


Figure 2.1. Recombinant protein expression in *E. coli* (BL21DE3) transformed with pE coli-C-term 6HN vector. Induction of *E. coli* (BL21 DE3) cells with IPTG enables the T7 RNA polymerase machinery to initiate transcription and translation. Basal protein expression is inhibited by *lacI* in the absence of IPTG (Clontech).

The vector also contains an ampicillin resistance gene (Ampr) and a pBR322 origin of replication, which maintains the vector at a low-copy-number to reduce basal levels of the protein of interest. An origin of replication is sequence of DNA at which replication is initiated in plasmids. It determines the vector copy number. The pBR322 replication origin is a site from where two RNAs (RNA I and II) are transcribed. RNA I serves as the primer initiating transcription which is attached to the 5' RNA II sequences which is stabilized by *rop* gene product. Hence the stabilization prevents change in the conformation of RNA II which would otherwise lead to RNAse H cleavage. Therefor this stabilized bond between the

two RNAs maintain replication initiation from a single site or origin and inturn the mechanism should also maintain the number of copies of the vector at low level.

E. coli strain (BL21DE3) was transformed with pE coli-5'*rppH-GFP* vector. The transformants were grown on agar plates containing ampicillin at 37˚C. A 3 mL LB/Amp culture was inoculated and was grown overnight. The next day, 2 mL of the overnight culture was inoculated in 100mL fresh LB/Amp medium. The culture was incubated on a shaker (200 rpm) for 2 hours (until the OD reaches 0.6-0.8). The cell culture was induced by adding IPTG to a final concentration of [0.22 ng/µL]. Cells were continously induced for 6 hours on a shaker (200 rpm). Samples were centrifuged at 5000 rpm/5 min, supernatant was descarded and pellet was resuspended in buffer A (Appendix II).

2.12.1 Purification of Histidine-tagged GFP

BL21DE3 cells expressing His-tagged GFP (always kept on ice) suspended in low imidazole buffer A (Binding/wash buffer, see Appendix II) were thawed, sonicated three times 5/10 sec burst on /off cycle while maintaining the samples on ice, then centrifuged at 4˚C for 10 min, at 10,000 rpm. Sequential extraction of samples and washing was done according to the laboratory manual (Uwe Klein).

Preparation of TALON Co resins (Clontech), was done according to the manufacturers guide lines. 200 µL TALON Co resin was centrifuged at 1500 rpm for 5 min. Supernatant was discarded and the resin was resuspended in 1.5 mL dH_2O and centrifuged once more at 1500 rpm for 5 min. The procedure was repeated with buffer A instead of dH_2O. The resin was then added to Empty PD-10 Desalting column (Biorad).

2.13 *SDS PAGE*

In sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), proteins are separated according to their molecular mass under denaturing conditions. Treatment with the anionic detergent SDS in heating environment (92˚C) destabilizes secondary and tertiary structures, thus denatured proteins gain a uniform net negative charge thereby movement of proteins on the gel depends on their size.

A 12% polyacrylamide solution was prepared by using solutions; dH_2O, A, B and APS (Appendix II) by carefully mixing. The gel was poured into a casting strand between glass plates leaving 1 cm of space between the top of the short plate and the resolving gel level. The top part was filled with water. After 30 min polymerization, the floating water was removed from the top. Stacking solution was prepared by carefully mixing solutions dH_2O, A, B', and APS (Appendix II). The stacking solution (4%) was quickly added to the top of the 12% polyacrylamide solution. Then the comb was inserted on top before the stacking solution polymerizes. After 30 min polymerization, the comb was taken out and the space between the glass plates was rinsed with water to get rid of bubbling. The electrophoresis was done at 90 V for 15 min and susequently at 120 V for 30 min.

3. Results

3.1 *Optimization of GFP*

Sequencing of *C. reinhardtii* nuclear genome revealed a high GC content (65%). Usage of codons specific for *Chlamydomonas* nuclear genes is essential. Previous works have shown that adaptation of codon usage of transgenes greatly enhances the expression of transgenes because of differences in tRNA abundance (Leon-Banares et al. 2004). It significantly improves translation and even promotes efficient integration into the genome (Meslet-Cladiere and Vallon 2011).

This study focused mainly on development of a reporter construct for the localization of the putative RNA pyrophosphohydrolase in *C. reinhardtii*. We used *ZsGreen* 1 GFP as reporter molecule from (Clontech), which has a size of 699 bp because of its bright green fluorescence with excitation, and emission maxima at 493 and 505 nm, respectively. *GFP* was codon optimized using the Codon Usage Database (Nakamura et al. 2000). Standard format was used for code selection. The database lists *C. reinhardtii*'s codon usage in genes and sum of codon; used. The codon usage table lists the relative frequency of each codon for a particular amino acid. Triplet codons are grouped according to the fraction of each triplet in the total nuclear genomic sequences and the frequency of each triplet per thousand bp of genomic sequences (see Appendix IV).

By using the GCUA tool (Fuhrmann et al. 2004), the *GFP* sequence was split into triplet , and the frequency of each codon was compared to that of 420,455 nuclear *C. reinhardtii* codons from the codon usage table. Thus, the relative adaptiveness defined by the GCUA tool shows the percentage of each codon as compared to that of the most preferred *C. reinhardtii* codon. Each GFP codon was manually optimized to codons that are most preferred and frequently used in the total nuclear *C. reinhardtii* genome. Almost all codons of GFP were optimized to the most preferred *C. reinhardtii* nuclear codons as shown in fig. 3.1.

The sequence was further modified prior to *GFP* synthesis for cloning into the pE coli-Cterm 6xHN (Clontech) vector. This vector encodes a 6xHN tag composed of 6 repeating His-Asn subunits. For the purpose of GFP expression and purification in *E. coli* cells, the 6xHN tag is incorporated in our gene construct. The stop codon for GFP was removed and an extra [C] nucleotide was included just upstream to the start codon of GFP. The optimized and adjusted *GFP* sequence was then synthesized (Life technologies).

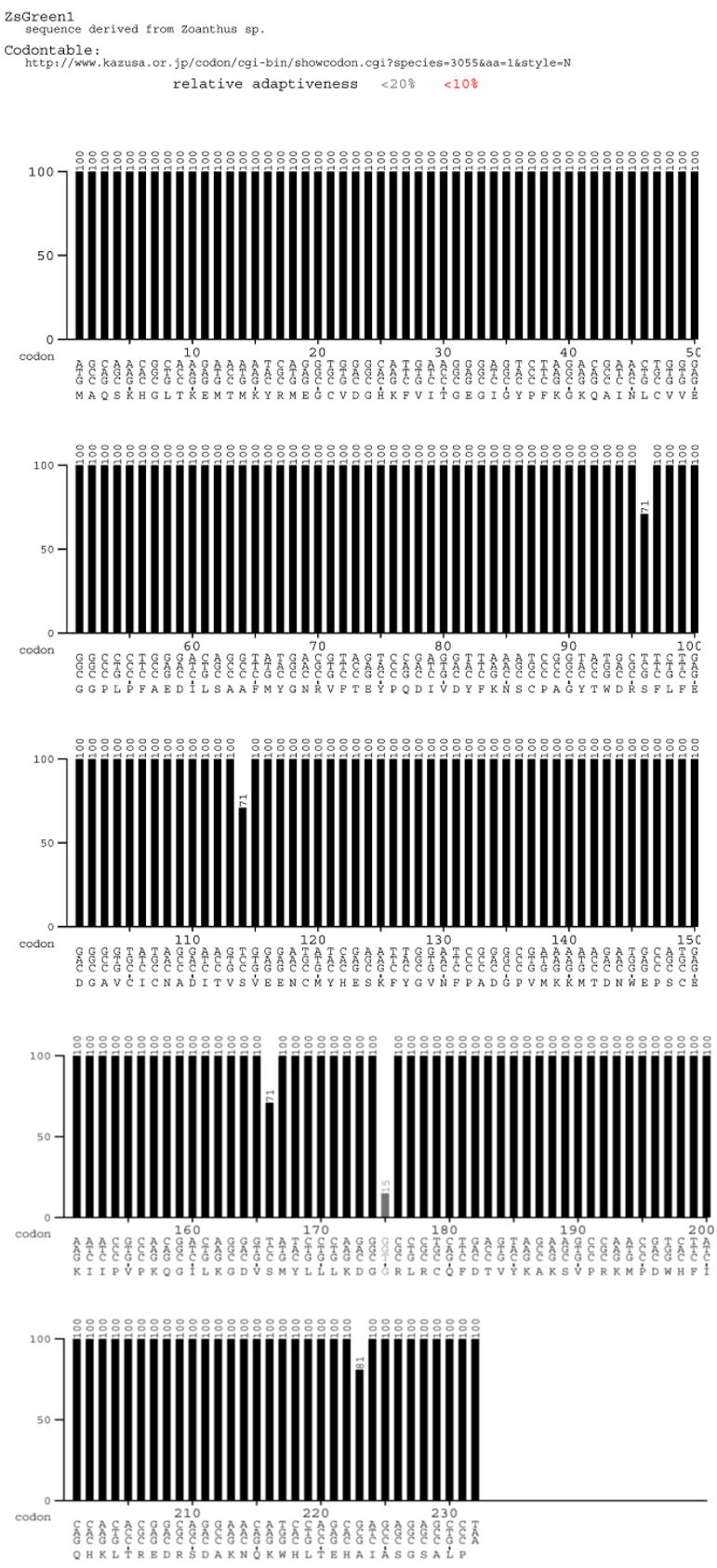
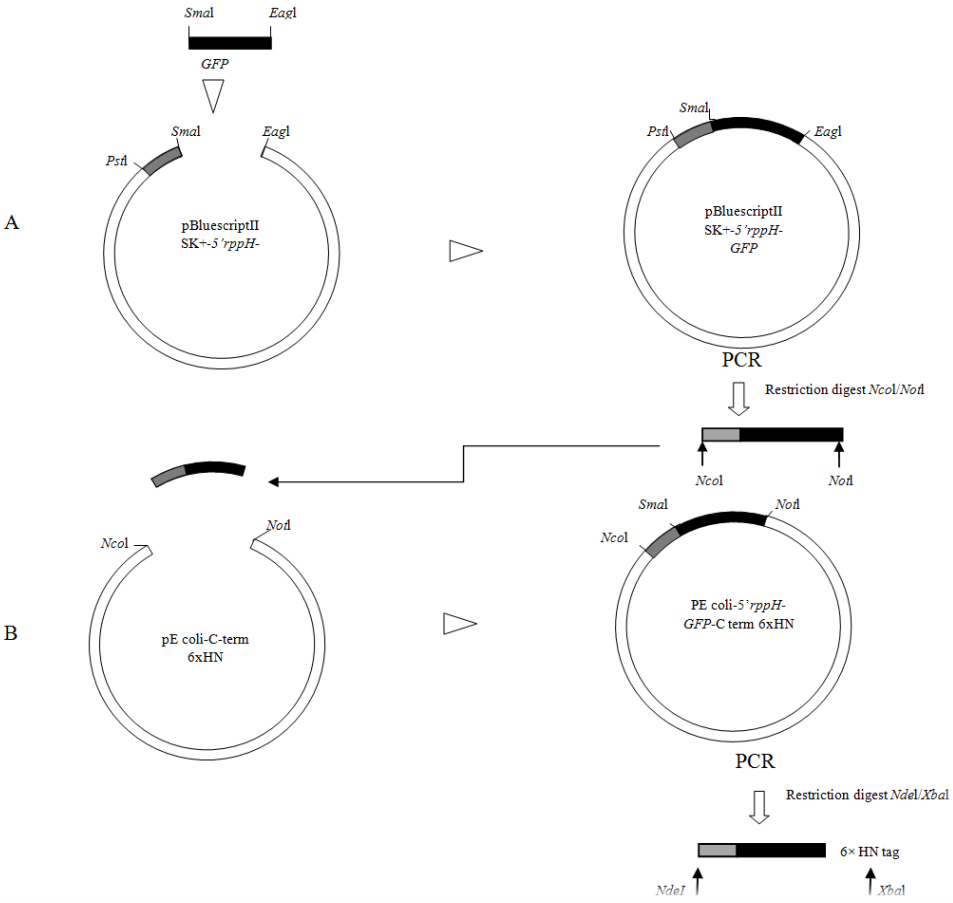


Figure 3.1. Graphical Codon Usage Analyses. GFP (Zsgreen1) has 699 total codons. Most codons were optimized to a relative adaptiveness (vertical line) of > 75% (black bar). Codons that have relative adaptiveness of < 20% are also shown (grey bar). The analyses was done by (GCUA) (Fuhrmann et al. 2004).

3.2 *Cloning*

Step by step cloning of the synthesized GFP is shown in fig. 3.2. There are four cloning steps designated as A, B, C and D.



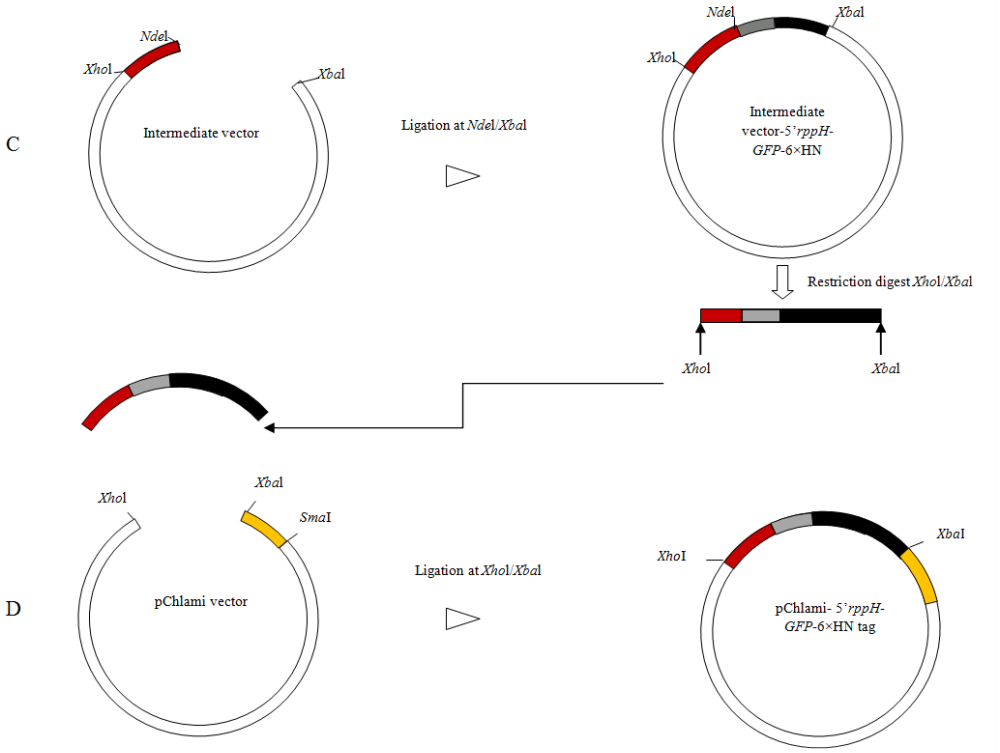


Figure 3.2. Cloning of GFP. Four steps (A-D) of cloning procedures and the restriction sites used are shown. Step A; *GFP* was cloned into pBluescript-5'*rppH* vector. Step B; 5'*rppH-GFP* was cloned into pE coli-C-term vector. Step C; 5'*rppH-GFP* was cloned into intermediate vector. Step D; 5'*PSAD*-5'*rppH-GFP* was cloned into pChlami vector (transformation vector). Red color-5'*PSAD* promoter sequences; Grey color-5'*rppH* region; Black color-*GFP*; Yellow color; 3'*PSAD* terminator sequences.

3.2.1 pBluescript II SK+ 5' *rppH-GFP* vector: Cloning step A

For the purpose of localization of RppH in *C. reinhardtii*, the 5' region of RppH was tagged to GFP to be able to follow the localization of the *GFP*-tagged gene product. Although not proven, a signal peptide that directs localization of the RppH protein was assumed to be present in the 5' region of RppH.

The pBluescript II SK+-*5'rppH* vector was received from Uwe Klein. It is a pBluescript II SK+ with 2961 bp (stratagene) that has the 5' region of RppH. *GFP* was cloned into pBluescript II SK+-*5'rppH* at the *Sma*l and *Eag*l sites to form pBluescript II SK+-*5'rppH-GFP* as shown in fig. 3.2 A and fig. 3.3. The restriction site for *Sma*l (CCCGGG) is found one nucleotide upstream of the start codon of GFP *i.e.* at the end of the 5'*rppH* sequence, and *Eag*l (CGGCCG) is found at the end of the coding region of *GFP* (see Appendix III). By restriction cutting, the vector fig. 3.4 (pSK+-5'*rpph-GFP*), at *Eco*RV/*Eag*l restriction sites and subsequent gel-electrophoresis analyses, the vector released a fragment which is ~ 953 bp in size as expected (fig. 3.4). The vector used for cloning pBluescript II SK+-*5'rppH* fig. 3.4

(C) was also cut with the same restriction enzymes. From the gel electrophoresis analyses, it can be seen that the restriction digested vector released a band (lower) of ~ 250 bp corresponding to the 5' *rppH* region (see fig. 3.4).

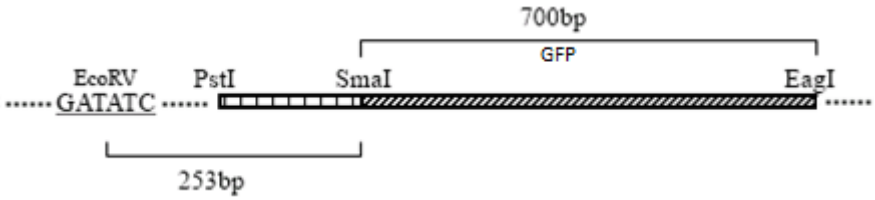


Figure 3.3. Restriction map for pBluescript II SK+-5'*rppH-GFP*. The 5' *rppH* is represented by vertical lines. GFP is represented with diagonal lines. The different restriction sites present in 5' *rppH-GFP* gene fragment (*Sma*I, *Eag*I) as well as in the pBluescript SK+ vector (*Eco*RV, *Pst*I) are shown. *Eco*RV site is found 253 bp upstream to the start codon of *GFP* in the pBluescript II SK+ vector. *Sma*I and *Eag*I flanked GFP region has a size of 700 bp. Dotted line represents sequences in the pBluescript II SK+ vector.

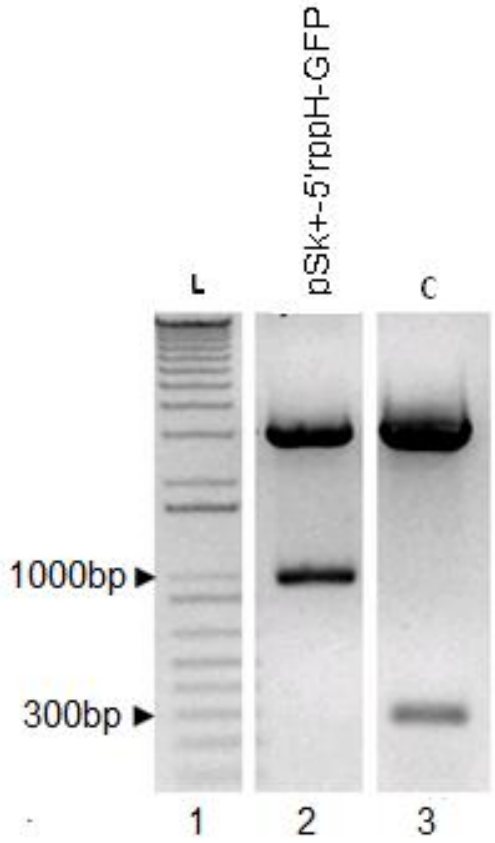


Figure 3.4. Gel electrophoresis analyses of pBluescriptII SK+-5'*rppH-GFP* vector cut with *EcoR*V/*Eag*I. Lane 1 (L); is 1kb+ ladder. Lane 2 (pSK+-5'*rppH-GFP*); is restriction digested pBluescriptII SK+-5'*rppH-GFP* construct. Lane 3 (C); is pBluescript II SK+-5'*rppH* used for cloning as a control.

3.2.2 pE coli-C-term-5'*rppH-GFP 6×HN* vector: Cloning step B

The 5'*rppH-GFP* gene was amplified by PCR thereby introducing an *Nco*I restriction site in the PCR product. After cutting the PCR product with *Nco*l/*Not*l restriction enzymes, the 5'*rppH-GFP* gene flanked by *Nco*l/*Not*l sites was cloned into *Nco*l/*Not*l restriction sites of pE coli-C-term 6×HN vector with the same enzymes as shown in fig. 3.2B. The ligation was checked with restriction digest of the ligated vector (pE coli-5'*rppH-GFP*-6HN) by using *Nco*l/*Xba*l, that releases a weak fragment which is ~ 990 bp as shown in fig. 3.6.

In the control experiment fig. 3.6 (C), the restriction digest releases a fragment that corresponds to the multiple cloning sites (MCS) in the pE coli-C-term 6×HN vector (Clontech) between *Nco*I/*Xba*I sites. But since the size is very small and the band is too weak, it is not seen in the gel electrophoresis analyses.

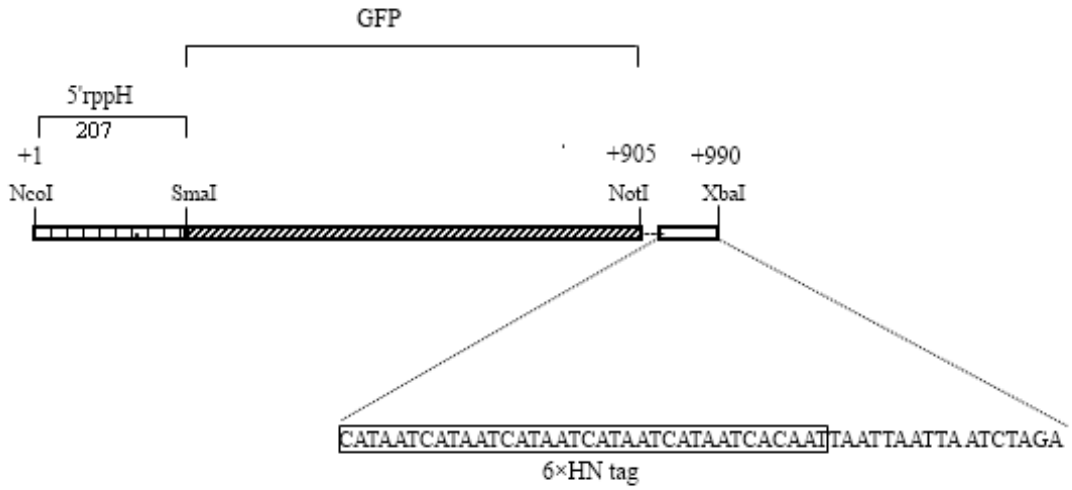


Figure 3.5. Restriction map of 5'*rppH-GFP* in pE coli vector. The map shows the different restriction sites *Nco*I, *Sma*I, *Not*I, *Xba*I. +1 refers to the translation start site of the fusion protein [5'*rppH-GFP-6×HN* (boxed)] until the *Xba*I site. Single dot line represent 39 nucleotide sequences that are present in the pE coli-C-term 6×HN vector (MCS). The vertical lines represent 5'rppH (207 bp) and the diagonal lines represent the GFP region.

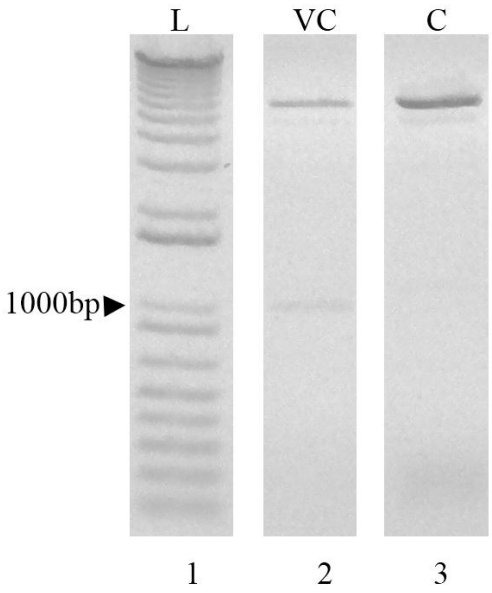


Figure 3.6. Gel-electrophoresis analyses of pE coli 5'*rppH-GFP* vector digested with *Nco*l/*Xba*l. Lane 1 (L) is 1 kb+ ladder; Lane 2 (VC) and Lane 3 (C); are restriction digested vector construct and pE coli C-term 6×HN vectors (used for cloning) respectively with *Nco*l/*Xba*l.

3.2.3 Intermediate 5*' rppH-GFP-6×HN* vector: Cloning step C

The intermediate vector which was received from Uwe Klein, is a pBluescriptII-SK+ vector that has a 1233 bp *Xho*l/*Xba*l fragment taken from the pChlamiRNA3int vector (Molnar et al. 2009). This fragment has a *Chlamydomonas* 5'-*PSAD* promoter region from the *Xho*I site until the *Nde*I site as can be seen in fig. 3.8.

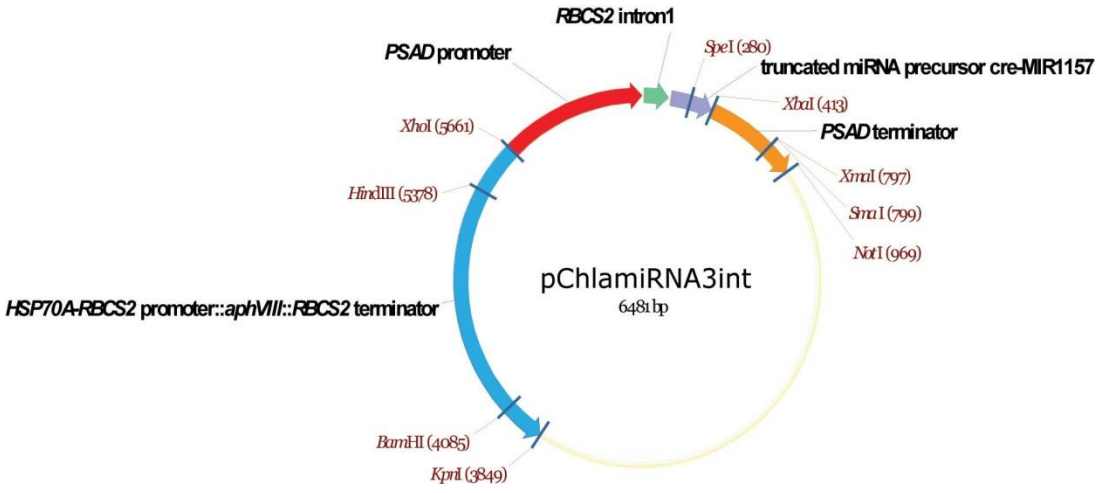


Figure 3.7. pChlamiRNA3int vector background. Vector map showing the restriction sites, selective marker (*aphVIII*) and the regulatory promoter and terminator regions (Molnar et al. 2009).

One of the characterstic features of nuclear genes is the presence of introns in the coding region. Transgene expression was shown to significantly increase by the presence of an intron in constructs used for ectopic expression in *Chlamydomonas* (Lumbreras et al. 1998). The *RBCS*2 intron has been widely used in transgene expression in *C. reinhardtii*. Although this system is possible, it takes time to insert the *Rbc*S2 intron into the coding sequence. The nuclear encoded PsaD protein which has a size of 20 kDa, encodes an abundant chloroplast protein found in the stromal side of the photosystem I complex. The *psaD* reading frame doesn't have any introns implying that the regulatory sequences should reside in the 5' and 3' untranslated regions.

For cloning of 5*'rppH-GFP* gene into the intermediate vector, a restriction site (*Nde*l) was introduced by PCR at the translation start site of pE coli *5'rppH-GFP* (see Appendix V). The *Nde*l site was introduced in 3 PCRs (section 2.2). Each PCR amplification gave a product that changed a single nucleotide. Three different 5' primers and a single 3' primer at the *Xba*l site were used for oligonucleotide synthesis. The PCR product which is flanked by the *Nde*l site at 5' end and *Xba*l site at the 3' end was then cloned into the *Nde*l/*Xba*l digested intermediate vector. Cloning of the PCR product was checked by digesting the vector with *Pvu*ll which cuts *GFP* at two sites that are 186 bp apart (see fig. 3.8) and at two other sites in the vector background (see fig 3.9). The restriction digest was run on a 1% agarose gel (fig. 3.9).

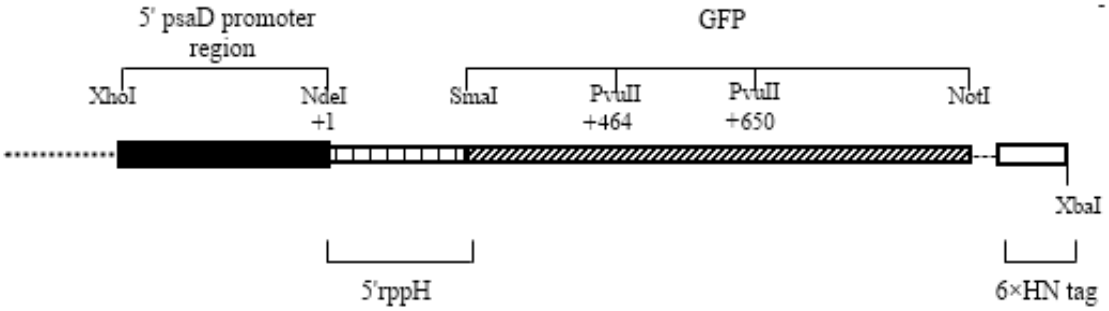


Figure 3.8. Restriction map of 5'*PSAD*-5'*rppH-GFP*-6×HN gene segment in the intermediate vector. Shown here are restriction sites (*Xho*I, *Nde*I, *Sma*I, *Pvu*II, *Not*I and *Xba*I) relative to +1 which is the translation start codon. The 5' *PSAD* promoter region is shown in black, vertical and diagonal lines represent 5' *rppH* and *GFP* regions respectively. Bold dots represent sequences that are part of the vector.

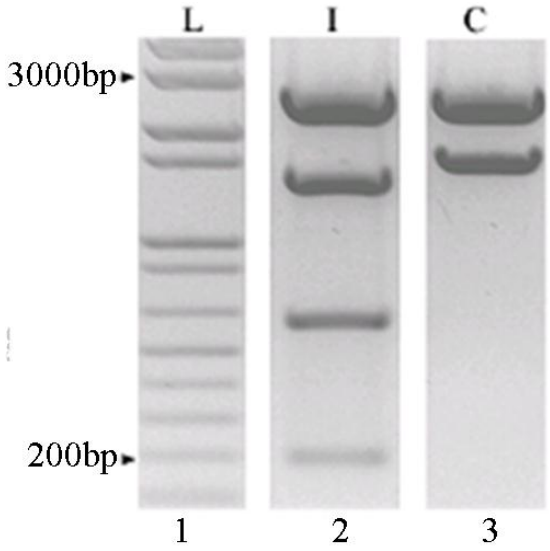


Figure 3.9. Gel electrophoresis analyses of PvuII digested Intermediate vector. Lane 1 (L); is 1kb+ ladder, Lane 2 (I); is an intermediate vector with gene construct (5'*rppH-GFP*) and Lane 3 (C); is an intermediate vector without the gene construct (5'*rppH-GFP*) used as a control.

3.2.4 Transformation vector pChlami-*5'rppH-GFP-6×HN*: Cloning step D

Finally, the gene fragment flanked by *Xho*l/*Xba*l was cloned into pChlami vector. First pChlamiRNA3int vector was digested by cutting the vector at *Xho*l/*Xba*l sites. The gene fragment flanked by *Xho*l/*Xba*l was also cut out from the intermediate vector and ligated into the opened transformation vector at the *Xho*l/*Xba*l sites. It is important to note here that the transformation vector has no longer the *rbcS*2-intron sequence which is present in the original pChlamiRNA3int vector (fig 3.7).

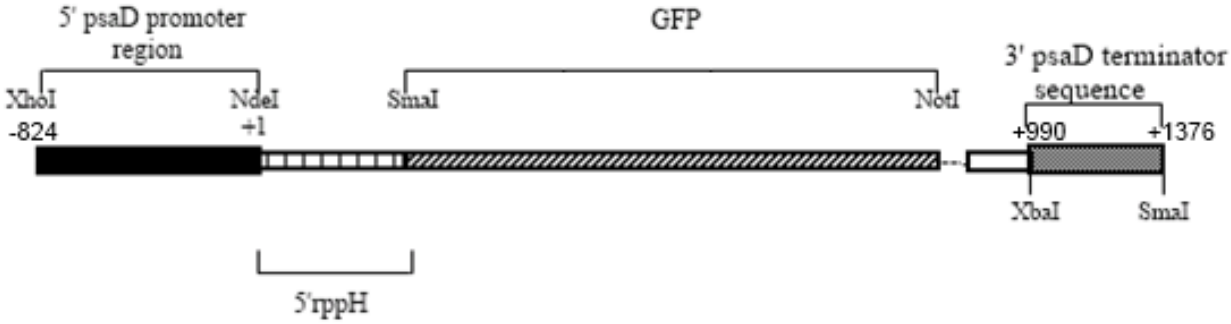


Figure 3.10. Restriction map of 5'*PSAD*-5'*rppH-GFP*-6×*HN* in pChlami vector. Shown here are positions of restriction sites *Xho*I, *Nde*I, *Sma*I, *Xba*I and *Sma*I relative to +1 which is the translation start codon. The 5' *PSAD* promoter region is shown in black; 5' *rppH* and *GFP* regions are represented in vertical and diagonal lines respectively. The 3' *PSAD* terminator sequence is shown in dotted gray color.

Cloning of the insert in the transformation vector was checked by restriction cutting the vector at *Xho*l/*Xba*l site and subsequent agarose gel electrophoresis analyses. The expected insert size from the *Xho*l site to the *Nde*l site and further to the *Xba*l site is 1814 bp see fig. 3.10.

This fragment (approximate size) was released when analyzed by gel electrophoresis as can be seen in fig.3.11.

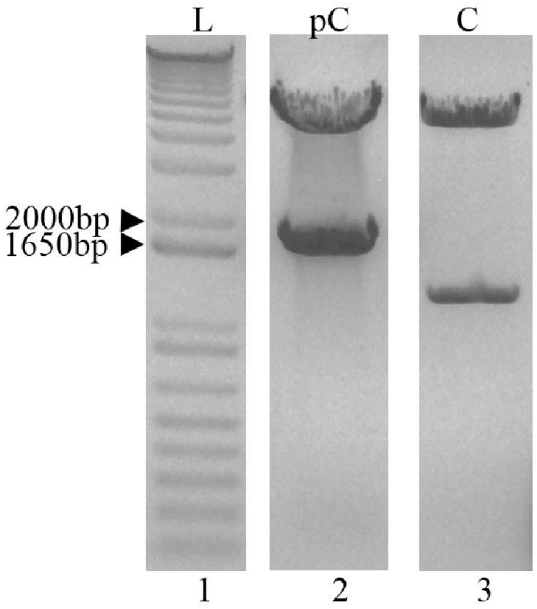


Figure 3.11. Gel electrophoresis analyses of pChlami- 5'*rppH-GFP*-6×HN vector digested with *Xho*I/*Xba*I restriction enzymes. Lane 1(L) is 1kb+ ladder; Lane 2 (pC) is digested pChlami- 5'*rppH-GFP*-6×HN vector releasing a fragment of ~ 1810 bp (lower band); Lane 3 (C) is digested pChlamiRNA3int vector (control) used for cloning releasing a fragment of ~ 1300 bp (lower band). The difference between the two lower bands corresponds to the difference between the two (pC and C) total vector sequences (581 bp).

3.3 *Transformation and selection of transformants*

The cell wall-less *Chlamydomonas* strain cw-15 cells were transformed by Uwe Klein as explained in section 2.6. Cells were grown on agar plates containing paromomycin (section 2.7). Cells which have integrated the marker gene (*aphVIII*) were selected. It is impossible to conclude that cells that have exhibited resistance to paromomycin are all positive transformants for the GFP construct. This is because of random integration of sequences into the *Chlamydomonas* nuclear genome. Thus further screening for GFP positive transformants at the genomic DNA level was needed.

3.4 *Analyses of transformants by PCR*

Genomic DNA was isolated as described in section 2.8. Identification of cells harboring the introduced gene at the genomic level was done by PCR. More than 15 transformants that were resistant to paromomycin were analyzed. And we proceeded with five transformants that we thought were GFP positive. A combination of different primers was used to amplify specific sequences of genomic DNA at different sites as can be seen in fig.3.12.

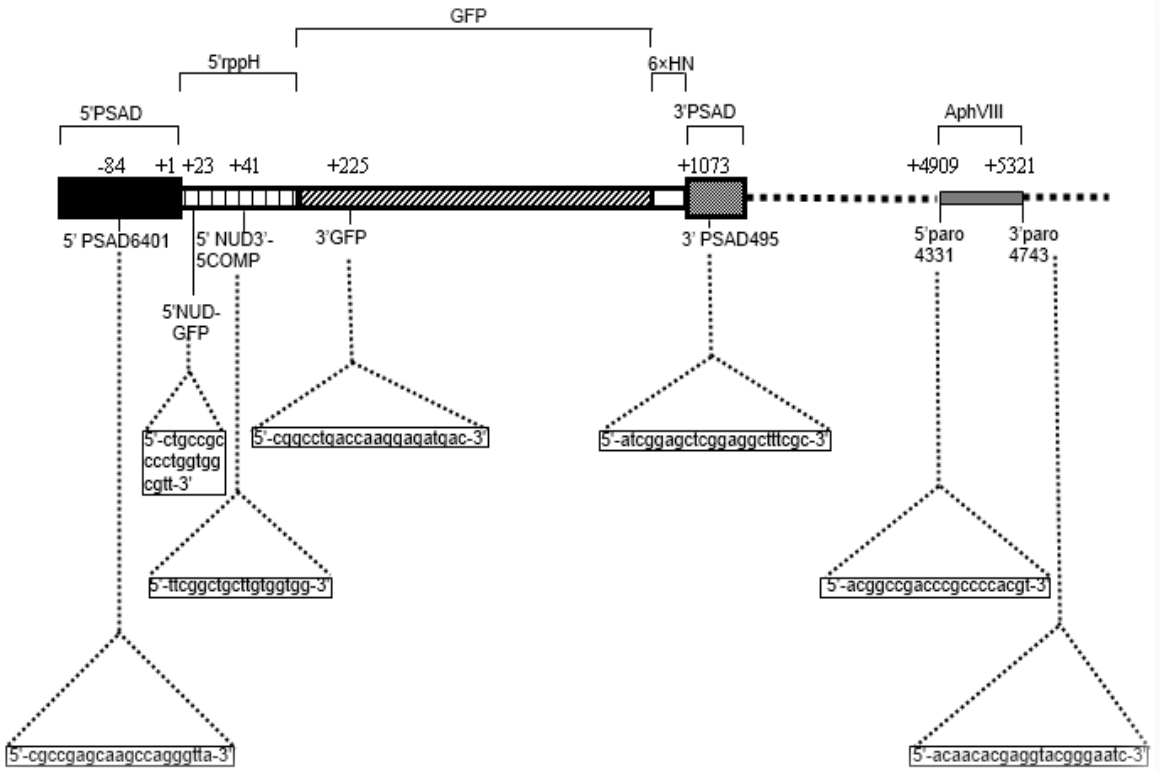


Figure 3.12. Map of the transformation vector (pChlami 5'*PSAD*-5'*rppH*-*GFP*-6×*HN*) with the different primers that were used to amplify specific sequences by PCR. Each primer (boxed) and the respective starting nucleotide positions in each strand depicted are shown relative to +1 which is the translation start codon. The 5' *PSAD* promoter region is shown in black; 5' *rppH* and *GFP* regions are represented in vertical and diagonal lines respectively; 6×HN until the *Xba*I site (see fig. 3.5) is shown as a background color. The 3' *PSAD* terminator sequence is shown in dotted gray color. The selective marker gene (*aphVIII*) is shown in gray. Dotted lines represent sequences in the pChlami vector.

Genomic DNA from the five transformants were amplified by **5' PSAD6401** and **3' PSAD495** (see fig. 3.13). As a positive control, template gene fragment from the tranformation vector was amplified by the same primers. Similarly, as a negative control, *Chlamydomonas* wild type genome was used as a template.

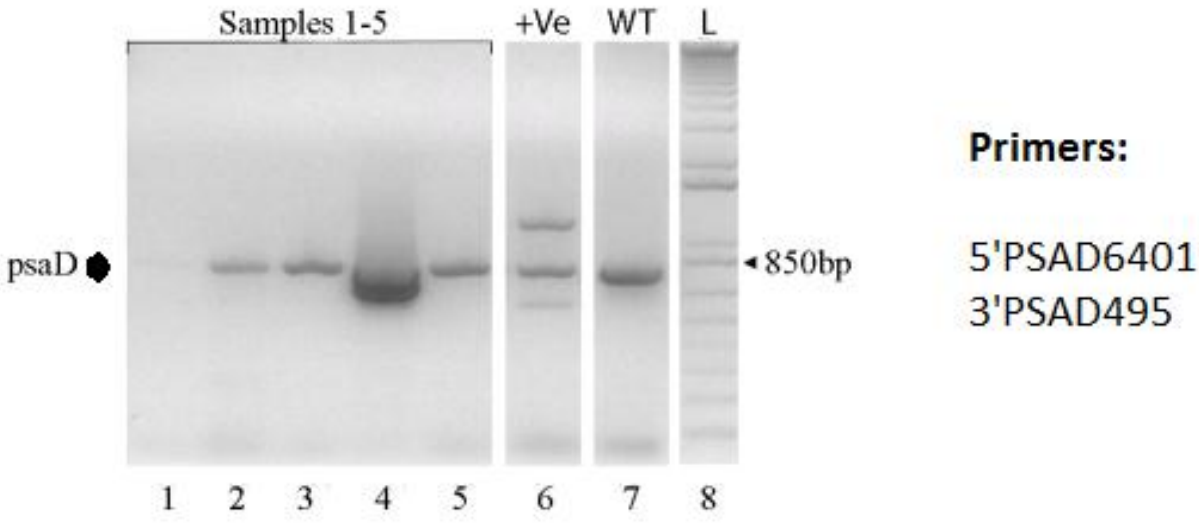


Figure 3.13. Gel electrophoresis analyses of PCR products using genomic DNA as a template. Lanes 1-5; represent transformant's (samples 1-5 respectively) genomic DNA amplifications. The PCR gave a product which has a size of ~ 780 bp that corresponds to the endogenous *psaD*. Lane 6 (+Ve) positive control; is amplified template gene fragment from the transformation vector; Lane 7 (WT); is the negative control (amplified wild type, *Chlamydomonas* genomic DNA as a template). The positive control (+Ve) shows amplification at ~ 780 bp (lower band) and at 1178 bp (upper band). The latter is the expected band size from this PCR. The negative control also shows amplification as expected at ~780 bp (*psaD*) gene size. Lane 8 (L); is ladder used (1Kb+).

The above figure shows that the genomic DNA of the five transformants have a PCR product which is ~ 780 bp that corresponds to a fragment from the endogenous *psaD* gene. The expected PCR product ~ 1178 bp, in the positive control (amplified template from transformation vector), upper band is also shown in fig. 3.13. On the negative control, the primers have amplified a gene fragment that has a similar size to the PCR products from the genomic DNA of the transformants and to the positive control (lower band). In the negative control (amplified wild type, *Chlamydomonas* genomic DNA as a template), the expected PCR product is 778 bp that corresponds to the wild type endogenous *psaD* gene from the primers used for this amplification.

Next, a specific genomic DNA fragment was amplified by using **NUD3'-5COMP** and **3'GFP** primers (see fig. 3.14). These primers amplify a gene fragment that has a size of 205 bp (see fig. 3.12). The amplification reaction gives PCR products that are 205 bp in size. This is expected from this PCR reaction. The positive (amplified template from transformation vector) and negative controls (amplified wild type, *Chlamydomonas* genomic DNA as a template) were also included in the PCR. The positive control has a PCR product that is similar in size to that of the transformants (samples 1-5 respectively) genomic DNA amplification. In the negative control, the DNA amplification has a size of ~ 650 bp. This is unexpected in a way that one of the primers (3'GFP) is not supposed to amplify the wild type genomic DNA used as a template, since GFP is not present in the *Chlamydomonas* genomic DNA.

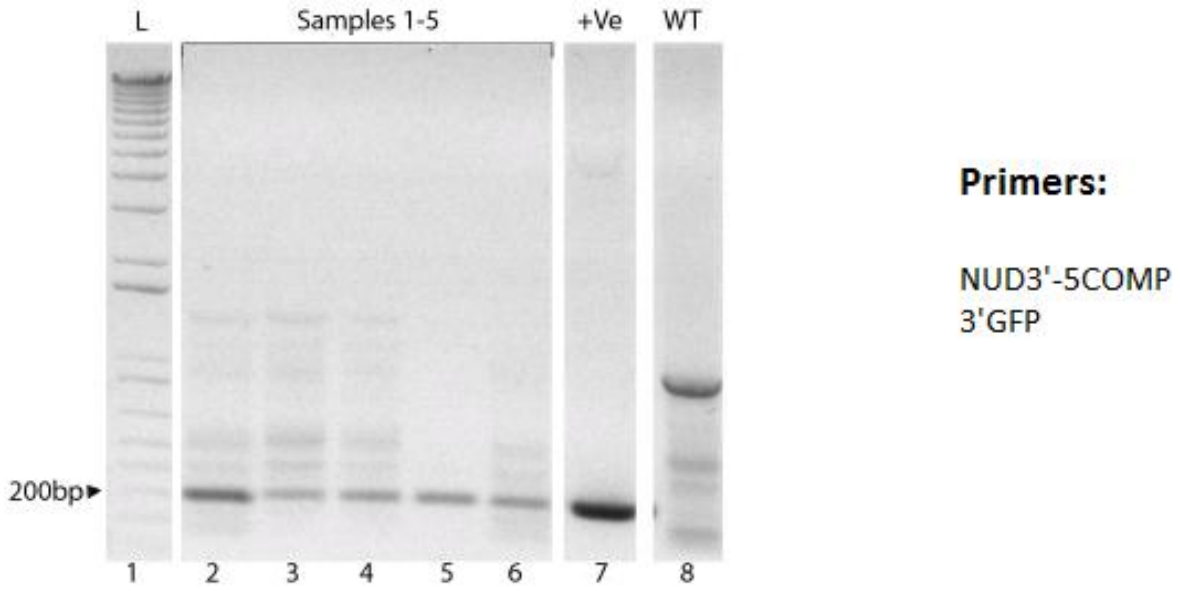


Figure 3.14. Gel electrophoresis analyses of PCR products using genomic DNA as a template. Lane 1(L); is 1kb+ ladder used. Lane 2-6 (sample 1-5 respectively); represent transformant's genomic DNA amplifications. Lane 7 (+Ve); is amplified transformation vector Lane 8- (WT); is the negative control (amplified wild type, *Chlamydomonas* genomic DNA as a template). Samples 1-5 have PCR products at 205 bp which is similar to the positive control. Amplification product in the negative control has a size of ~ 650 bp.

The transformants were also screened for the gene that confers paromomycin resistance which is present in the transformation vector but absent in the wild type genomic DNA. Figure 3.15 shows PCR products of gene fragments amplified by using **5'paro4331** and **3'paro4743** primers (see fig 3.15). The PCR products from all samples (1-5 respectively) have size of 433 bp. The positive control (amplified template from transformation vector) also has a PCR product of approximately the same size. The wild type *Chlamydomonas* lacks the gene that confers

paromomycin resistance, *aphVIII*. Thus, no amplification is seen in the negative control as expected.

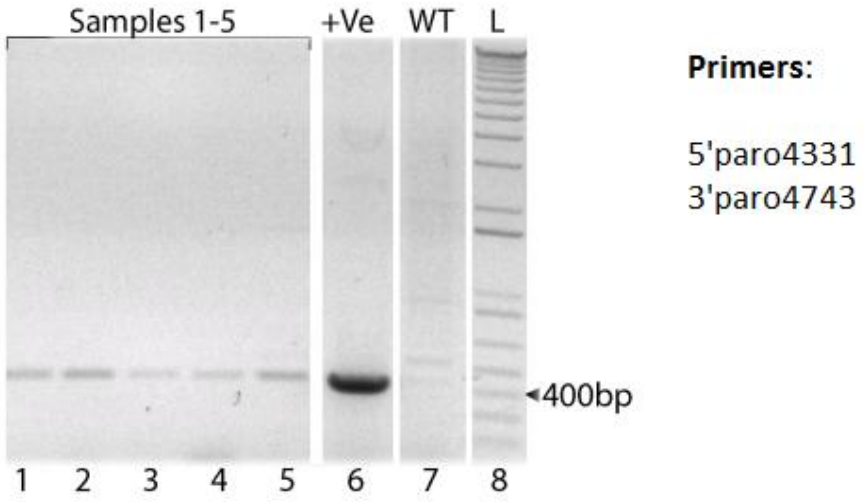


Figure 3.15. Gel electrophoresis analyses of PCR products using genomic DNA as a template. Lanes 1-5 (samples 1-5 respectively); show PCR amplification at ~ 430 bp as expected. Lane 6 (+Ve); is transformation vector that shows amplified fragment, which has a size of ~ 430 bp. Lane 7 (WT); is negative control (amplified wild type, *Chlamydomonas* genomic DNA as a template) which shows no amplification since it lacks the *aphVIII*. Lane 8 (L); is 1Kb+ ladder.

Finally transformants (samples 1-5 respectively) were screened by using **5'NUD GFP** and **3' PSAD 495** (see fig. 3.16). From the gel electrophoresis result, only transformant number 5 (sample 5) gives a PCR product that has a size of ~1070 bp as expected. The amplified gene fragment of the positive control (transformation vector) also has the same size (1070 bp). The negative control (amplified wild type, *Chlamydomonas* genomic DNA as a template) shows weak products that have sizes of ~ 200 and another at ~ 1000 bp.

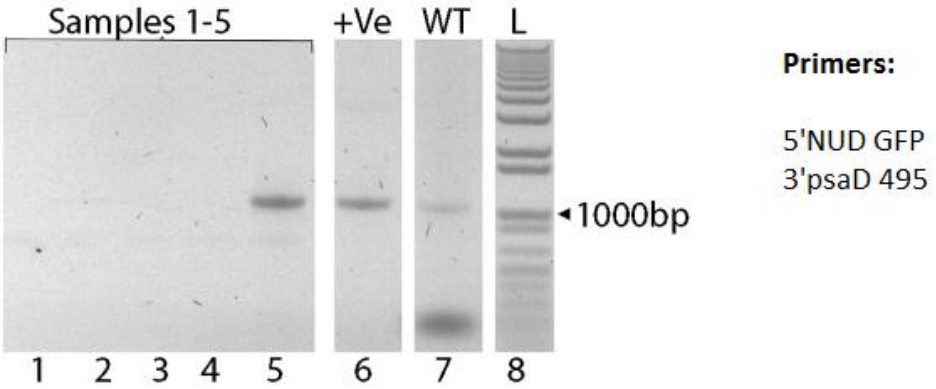


Figure 3.16. Gel electrophoresis analyses of PCR products using genomic DNA as a template. Lanes 1-5 (samples 1-5 respectively). Lane 5 (transformant number 5); shows a PCR product that has a size of ~ 1070 bp. Lane 6 (+Ve); is amplified transformation vector that gives a PCR product which has a size of ~ 1070 bp similar to PCR product from transformant number 5. Lane 7 (WT); is a negative control (amplified wild type, *Chlamydomonas* genomic DNA as a template) that shows weak bands at ~200 bp and at ~ 1000 bp. Lane 8 (L); is 1Kb+ ladder.

3.5 *Recombinant protein expression and purification*

Recombinant expression of GFP from the jellyfish *A. victoria* was first accomplished in *E. coli* (Chalfie et al. 1994). To determine if the GFP construct transformed into *Chlamydomonas* is capable of producing functional GFP protein (fluorescence) *in vivo*, the construct was expressed in *E. coli* cells as described in section 2.12. But fluorescence microscopy was not done due to shortage of time. Rather *E. coli* BL21DE3 lysates were purified by His-tag affinity chromatography using cobalt-sepharose. We examined *E. coli* BL21DE3 cell lysates for the expression of the GFP fusion protein.

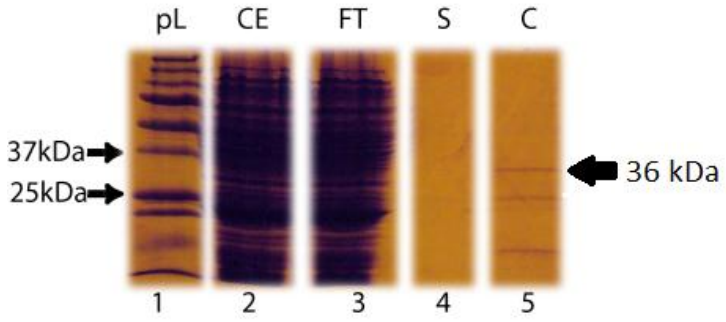


Figure 3.17. SDS-PAGE analyses of *E. coli* BL21DE3 lysates purified by Co sepharose affinity chromatography. Lane 1; is prestained protein ladder. Lane 2 (CE); is crude extract showing all proteins that are present in *E. coli* cells. Lane 3 (FT); is the flow through, all proteins that are present in the crude extract except for the HN tagged GFP protein (5'*rppH-GFP*). Lane 4 (S); is the purified cell lysate sample (transformed with pE coli-*5'rppH-GFP-6×HN*) that was made in this experiment. Lane 5 (C); is a purified 5'*rppH-GFP*-6×HN protein as a control. The purified sample (S) does not show any proteins that are purified. The purified sample (C) shows the expected 5'*rppH-GFP*-6×*HN* tagged protein size which is 36 kDa.

The SDS-PAGE analyses shows *E. coli* BL21DE3 lysates purified by affinity chromatography. Crude extracts (CE) in Lane 2 in fig. 3.17, are all proteins in the cell. The flow throw (FT), are polypeptides that have no affinity to the metal (cobalt) therefore directly pass through the column. Due to the high affinity of histidine which is tagged to the N-terminal region of GFP, to the metal ion, the tagged protein remains bound to the resin. The tagged protein is eluted with elution buffer which has a higher imidazole concentration than the washing buffer. The size of the fusion protein (5'rppH-GFP-6×HN) is 36 kDa. We have tried to express this protein in *E. coli*. From fig. 3.17, the construct that was made in this experiment (S) didn't show any purified protein signal. But other already induced *E. coli* cells received from Uwe Klein represented as (C) on fig. 3.17 shows a band that is ~ 36 kDa in size which is the expected band size for the fusion GFP protein.

4. Discussion

To develop a reporter gene for expression in the *C. reinhardtii* nucleus, a codon optimized *GFP* gene was synthesized. The reporter *GFP* fused to the 5*'* region of the *rppH* gene was used to transform *C. reinhardtii* cells. Transformants were then screened at the genomic level for *GFP* positives.

Codon optimization of synthetic GFP

Nuclear-based ectopic expression in *C. reinhardtii* is highly codon biased with G or C preferred at the third position. Events causing gene silencing in *C. reinhardtii* are related to inappropriate nucleotide sequences (Fuhrmann et al. 1999). In addition, previous reports have shown that codon usage affects the level of expression of recombinant proteins in the *C. reinhardtii* (Franklin et al. 2002). Other experiments have also shown that the cellular tRNA abundance is correlated with the number of tRNA genes and is adjusted to the codon usage to optimize translation efficiency in *C. reinhardtii* (Cognat et al. 2008). Moreover the extent of codon bias for each gene is related to the protein production level (Nakamura et al. 1999).

Previously synthetic GFP tagged to a protein reflecting the *C. reinhardtii* nuclear codon usage was expressed in *C. reinhardtii* which allowed the visualization of the recombinant protein. We have optimized *GFP* according to the nuclear *C. reinhardtii* codon usage preference by using Codon Usage Database. The fraction of each codon usage from the *GFP* was computed to that of *C. reinhardtii* nuclear genome codon usage. Among the total 233 codons in GFP, 198 were optimized to the most frequently used codons in *C. reinhardtii i.e.* 100% relative adaptiveness as defined by the GCUA tool. 4 of the GFP codons were optimized to those that are not as most frequently used, but moderately used in the *C. reinhardtii* nuclear genome. There is one codon in the *GFP* sequence that codes for glycine (GGT) which is not optimized at all (relative adaptiveness <20%). Genes encoding highly expressed proteins tend to utilize codons whose levels of tRNAs are particularly abundant. *rbcL* which is highly expressed in the *C. reinhardtii* genome was computed in the same way as was done for *GFP*. From the analyses, it is evident that around 7% of the total codon sequence have relative adaptiveness (as defined by GCUA) below 20%.

Expression of GFP in *E. coli* BL21DE3

GFP was expressed in *E. coli* cells to validate its function (fluorescence) *in vivo*. Due to shortage of time, fluorescence couldn't be analyzed. But ***E. coli*** BL21DE3 cell lysates were purified by Co sepharose affinity chromatography to detect the chimeric GFP protein. SDS-PAGE analyses of the synthesized GFP expressed in *E. coli* cells shows that the expression level of the protein is low as seen in the control experiment. This is expected in a way that the *GFP* codon sequences were not optimized to that of bacterial cells genome codon usage rather to *C. reinhardtii* nuclear genome.

Nuclear transformation of *C. reinhardtii*

Nuclear transformation in *C. reinhardtii* is random. Although billions of cells were used to transform *C. reinhardtii*, we could only proceed with 5 transformants that integrated the paromomycin resistance gene. We primarily used PCR for transformant screening at the genomic level. Since results from PCR do not differentiate between genomic and ectopic fragment amplification, we used our construct (transformation vector) as a positive control and wild type *C. reinhardtii* genomic DNA as a negative control to determine positive transformants.

Transformants were first screened for gene fragment that is amplified by primers that anneal at the 5' and 3' *PSAD* regulatory regions. The expected PCR product amplification for the transgene is 1178 bp. This could not be verified from this PCR for all the transformants. In addition the wild type *C. reinhardtii* genomic DNA has *psaD* gene flanked by 5' and 3' UTR. Amplification product of the negative control by the primers used gives a PCR product which is ~ 780 bp, expected product for the endogenous amplification. In the same experiment the positive control has amplification products of two sizes at ~ 1180 and at ~ 780 bp. This could either be contamination with the wild type genomic DNA or that the PCR was not optimal to amplify a template gene fragment that has a larger size. The possibility that the positive control is contaminated is less likely that all the PCR reagents were prepared in master mix first and that the template DNA was added to the respective reaction tubes. In addition the fact that only the endogenous genomic DNA amplification from the transformants imply that the PCR was not optimal (concentration, annealing temperature and elongation time) for larger size template DNA. The PCR was optimized by altering the annealing temperature, but valid results could not be obtained. Additional experiment can be done to optimize the PCR more by altering the primer concentrations and adjust the concentration of magnesium ions.

We further screened the transformants with primers that amplify a gene fragment from the 5' *rppH* region to the 5' end of *GFP*. The PCR product for all genomic DNA from the transformants and the positive control (transformation vector) has a size of 205 bp as expected. The negative control (wild type genomic DNA) on the other hand has a PCR product which was not expected in this PCR since there is no GFP in the wild type genome. But other results (not shown) have implicated that the NUD 3'-5 COMP primer binds non-specifically. Transformants were also screened for *aphVIII* gene which encodes for aminoglycoside 3'-phosphotransferase that confers resistance to paromomycin. All the specific transformant genomic DNA amplification with primers that amplify the *aphVIII* gene sequence have product sizes of ~ 430 bp similar to the amplification product from the transformation vector (positive control). The wild type genomic DNA lacks *aphVIII* therefore no amplification is seen in the negative control.

Finally all transformants were screened with primers that amplifies a gene fragment that has a size of 1072 bp. We found one transformant (№ 5) that is positive for this screening that has a similar amplification product with the positive control. The wild type genomic DNA amplification shows that there is non-specific binding of the primers. Further optimization of

the primers can be applied by altering the annealing temperatures and changing the concentration of the magnesium ions in the PCR.

5. Conclusion

Vector constructs made for this experiment were analyzed by restriction cutting analyses. All analyses show that the vectors have incorporated the 5' *rppH-GFP* gene fragment. The transformation vector, which is the reporter vector construct was as well validated by restriction cutting analyses.

The screening of transformants for positives has led us to choose one transformant (№ 5) that seems to harbor the *GFP* transgene. Further analyses needs to be done to verify expression of the transgene at the mRNA and protein levels. From these experiments it is possible to conclude that transformants that have incorporated the paromomycin resistance gene don't necessarily have the *GFP* transgene.

Further work

Even though there was an attempt to express GFP by the pE coli vector construct which was made in this study, it was not possible to detect GFP when the purified sample was run on SDS-PAGE. Normally *E. coli* cells transformed with a protein expression vector induced with IPTG (depending on the concentration, timing and length of induction) should be able to express the protein of interest. Even though the cells were induced by following the protocol and later with slight modification, such as various timing of induction (result not shown), it was impossible to detect the fusion protein in all cases. Moreover, the construct has been validated by restriction cutting as shown in fig. 3.6. Nevertheless the vector was not sequenced. So further work can be done to sequence the vector construct and also detect the fluorescence of the protein in *E. coli* cells.

In addition in the control experiment, the purified sample shows other purified proteins in addition to what is expected for the fusion GFP protein. Thus to be certain that GFP is indeed expressed in these *E. coli* cells and that the result is not an artifact of certain experimental error, the sample could be analyzed by mass spectrometry (MS). By applying prior washing and elution conditions, the presence of imidazole and high salt concentration can be adjusted unless and otherwise would increase mass spectrometry backgrounds. Furthermore by using specific antibody against GFP or HN tag, immunoprecipitation technique can be applied to precipitate the fusion protein.

Conventional PCR provides limited information on the number of inserts (copy number). In addition false negatives can result from problems with long-range PCR or low amounts of DNA. Therefore further validation of the transformant can be done to have further information on the copy number of the transgene before proceeding with the transcription analyses. Even though Southern analyses is laborious, time consuming and requires large amounts of high-quality DNA when compared to conventional PCR, screening with Southern

blot analyses identifies targeted transformants and permits analyses of the copy number of the transgene.

Furthermore, before concluding that transformant (№ 5) could be used for the localization of RNA pyrophosphohydrolase in *C. reinhardtii* cells, it is advisable to confirm the transcription of the transgene. RNA can be isolated and further hybridization based experiments (e.g Northern blot) can be done by designing specific probes. Alternatively, the RNA template can be reversibly transcribed to synthesize cDNA, and gene expression could be analyzed by using real-time reverse transcription polymerase chain reaction (RT-PCR) which uses fluorescent reporter molecules to monitor the production of amplification products during each cycle of the PCR reaction.

References

Anthonisen, I. L., M. L. Salvador and U. Klein (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**(7): 1024-1033.

Anupama, K., J. K. Leela and J. Gowrishankar (2011). "Two pathways for RNase E action in *Escherichia coli* in vivo and bypass of its essentiality in mutants defective for Rho-dependent transcription termination." Molecular microbiology **82**(6): 1330-1348.

Arraiano, C. M., J. M. Andrade, S. Domingues, I. B. Guinote, M. Malecki, R. G. Matos, R. N. Moreira, V. Pobre, F. P. Reis, M. Saramago, I. J. Silva and S. C. Viegas (2010). "The critical role of RNA processing and degradation in the control of gene expression." Fems Microbiology Reviews **34**(5): 883-923.

Barkan, A. (2011). "Expression of plastid genes: Organelle-specific elaborations on a prokaryotic scaffold." Plant physiology **155**(4): 1520-1532.

Belasco, J. G. (2010). "All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay." Nature Reviews Molecular Cell Biology **11**(7): 467-478.

Bessman, M. J., D. N. Frick and S. F. O'Handley (1996). "The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes." The Journal of biological chemistry **271**(41): 25059-25062.

Blatch, G. L. and M. Lassle (1999). "The tetratricopeptide repeat: a structural motif mediating protein-protein interactions." Bioessays **21**(11): 932-939.

Blowers, A. D., G. S. Ellmore, U. Klein and L. Bogorad (1990). "Transcriptional analysis of endogenous and foreign genes in chloroplast transformants of *Chlamydomonas*." The Plant cell **2**(11): 1059-1070.

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

Boudreau, E., J. Nickelsen, S. D. Lemaire, F. Ossenbuhl and J. D. Rochaix (2000). "The Nac2 gene of *Chlamydomonas* encodes a chloroplast TPR-like protein involved in *psbD* mRNA stability." The EMBO journal **19**(13): 3366-3376.

Boynton, J. E., N. W. Gillham, E. H. Harris, J. P. Hosler, A. M. Johnson, A. R. Jones, B. L. Randolph-Anderson, D. Robertson, T. M. Klein, K. B. Shark and et al. (1988). "Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles." Science **240**(4858): 1534-1538.

Brejc, K., T. K. Sixma, P. A. Kitts, S. R. Kain, R. Y. Tsien, M. Ormo and S. J. Remington (1997). "Structural basis for dual excitation and photoisomerization of the *Aequorea victoria*

green fluorescent protein." Proceedings of the National Academy of Sciences of the United States of America **94**(6): 2306-2311.

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

Carpousis, A. J., B. F. Luisi and K. J. McDowall (2009). "Endonucleolytic initiation of mRNA decay in *Escherichia coli*." Progress in molecular biology and translational science **85**: 91-135.

Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward and D. C. Prasher (1994). "Green fluorescent protein as a marker for gene-expression." Science **263**(5148): 802-805.

Chytilova, E., J. Macas and D. W. Galbraith (1999). "Green fluorescent protein targeted to the nucleus, a transgenic phenotype useful for studies in plant biology." Annuals of Botany **83**(6): 645-654.

Cognat, V., J. M. Deragon, E. Vinogradova, T. Salinas, C. Remacle and L. Marechal-Drouard (2008). "On the evolution and expression of *Chlamydomonas reinhardtii* nucleus-encoded transfer RNA genes." Genetics **179**(1): 113-123.

Colon-Ramos, D. A., J. L. Salisbury, M. A. Sanders, S. M. Shenoy, R. H. Singer and M. A. Garcia-Blanco (2003). "Asymmetric distribution of nuclear pore complexes and the cytoplasmic localization of beta2-tubulin mRNA in *Chlamydomonas reinhardtii*." Developmental cell **4**(6): 941-952.

Condon, C. (2007). "Maturation and degradation of RNA in bacteria." Current opinion in microbiology **10**(3): 271-278.

Deana, A., H. Celesnik and J. G. Belasco (2008). "The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal." Nature **451**(7176): 355-358.

Deikus, G. and D. H. Bechhofer (2011). "5' End-independent RNase J1 endonuclease cleavage of *Bacillus subtilis* model RNA." The Journal of biological chemistry **286**(40): 34932-34940.

Deutscher, M. P. (2006). "Degradation of RNA in bacteria: comparison of mRNA and stable RNA." Nucleic acids research **34**(2): 659-666.

Dutcher, S. K., L. Li, H. Lin, L. Meyer, T. H. Giddings, Jr., A. L. Kwan and B. L. Lewis (2012). "Whole-genome sequencing to identify mutants and polymorphisms in *Chlamydomonas reinhardtii*." G3 (Bethesda) **2**(1): 15-22.

Eberhard, S., C. Loiselay, D. Drapier, S. Bujaldon, J. Girard-Bascou, R. Kuras, Y. Choquet and F. A. Wollman (2011). "Dual functions of the nucleus-encoded factor TDA1 in trapping and translation activation of *atpA* transcripts in *Chlamydomonas reinhardtii* chloroplasts." The Plant journal : for cell and molecular biology **67**(6): 1055-1066.

Ellis, R. J. (1970). "Further similarities between chloroplast and bacterial ribosomes." Biochemical Journal **116**(4): P28-&.

Emory, S. A., P. Bouvet and J. G. Belasco (1992). "A 5'-Terminal stem loop structure can stabilize messenger-RNA in *Escherichia.Coli*." Genes & development **6**(1): 135-148.

Evguenieva-Hackenberg, E. and G. Klug (2009). "RNA degradation in Archaea and Gram-negative bacteria different from *Escherichia coli*." Progress in molecular biology and translational science **85**: 275-317.

Evguenieva-Hackenberg, E. and G. Klug (2011). "New aspects of RNA processing in prokaryotes." Current opinion in microbiology **14**(5): 587-592.

Franklin, S., B. Ngo, E. Efuet and S. P. Mayfield (2002). "Development of a GFP reporter gene for *Chlamydomonas reinhardtii* chloroplast." The Plant journal : for cell and molecular biology **30**(6): 733-744.

Fuhrmann, M., A. Hausherr, L. Ferbitz, T. Schodl, M. Heitzer and P. Hegemann (2004). "Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene." Plant molecular biology **55**(6): 869-881.

Fuhrmann, M., W. Oertel and P. Hegemann (1999). "A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in *Chlamydomonas reinhardtii*." Plant Journal **19**(3): 353-361.

Garrey, S. M., M. Blech, J. L. Riffell, J. S. Hankins, L. M. Stickney, M. Diver, Y. H. Hsu, V. Kunanithy and G. A. Mackie (2009). "Substrate binding and active site residues in RNases E and G: role of the 5'-sensor." The Journal of biological chemistry **284**(46): 31843-31850.

Goldschmidt-Clermont, M., M. Rahire and J. D. Rochaix (2008). "Redundant cis-acting determinants of 3' processing and RNA stability in the chloroplast *rbcL* mRNA of *Chlamydomonas*." The Plant journal : for cell and molecular biology **53**(3): 566-577.

Gonzalez-Ballester, D., A. de Montaigu, J. J. Higuera, A. Galvan and E. Fernandez (2005). "Functional genomics of the regulation of the nitrate assimilation pathway in *Chlamydomonas*." Plant physiology **137**(2): 522-533.

Gorski, S. A., M. Dundr and T. Misteli (2006). "The road much traveled: trafficking in the cell nucleus." Current opinion in cell biology **18**(3): 284-290.

Grossman, A. R., E. E. Harris, C. Hauser, P. A. Lefebvre, D. Martinez, D. Rokhsar, J. Shrager, C. D. Silflow, D. Stern, O. Vallon and Z. D. Zhang (2003). "*Chlamydomonas reinhardtii* at the crossroads of genomics." Eukaryotic cell **2**(6): 1137-1150.

Guhaniyogi, J. and G. Brewer (2001). "Regulation of mRNA stability in mammalian cells." Gene **265**(1-2): 11-23.

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

References

Hambraeus, G., K. Karhumaa and B. Rutberg (2002). "A 5' stem-loop and ribosome binding but not translation are important for the stability of *Bacillus subtilis aprE* leader mRNA." Microbiology **148**(Pt 6): 1795-1803.

Harris, E. H. (2001). "*Chlamydomonas* as a Model Organism." Annual review of plant physiology and plant molecular biology **52**: 363-406.

Herrin, D. L. and J. Nickelsen (2004). "Chloroplast RNA processing and stability." Photosynthesis research **82**(3): 301-314.

Hollams, E. M., K. M. Giles, A. M. Thomson and P. J. Leedman (2002). "mRNA stability and the control of gene expression: implications for human disease." Neurochemical research **27**(10): 957-980.

Ibrahim, F., L. A. Rymarquis, E. J. Kim, J. Becker, E. Balassa, P. J. Green and H. Cerutti (2010). "Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in *Chlamydomonas*." Proceedings of the National Academy of Sciences of the United States of America **107**(8): 3906-3911.

Jacobs, J. and U. Kuck (2011). "Function of chloroplast RNA-binding proteins." Cellular and Molecular Life Sciences **68**(5): 735-748.

Jeong, B. R., D. Wu-Scharf, C. M. Zhang and H. Cerutti (2002). "Suppressors of transcriptional transgenic silencing in *Chlamydomonas* are sensitive to DNA-damaging agents and reactivate transposable elements." Proceedings of the National Academy of Sciences of the United States of America **99**(2): 1076-1081.

Johnson, X., K. Wostrikoff, G. Finazzi, R. Kuras, C. Schwarz, S. Bujaldon, J. Nickelsen, D. B. Stern, F. A. Wollman and O. Vallon (2010). "MRL1, a conserved pentatricopeptide repeat protein, is required for stabilization of *rbcL* mRNA in *Chlamydomonas* and Arabidopsis." The Plant cell **22**(1): 234-248.

Kaberdin, V. R., D. Singh and L. C. Sue (2011). "Composition and conservation of the mRNA-degrading machinery in bacteria." Journal of Biomedical Science **18**.

Kathir, P., M. LaVoie, W. J. Brazelton, N. A. Haas, P. A. Lefebvre and C. D. Silflow (2003). "Molecular map of the *Chlamydomonas reinhardtii* nuclear genome." Eukaryotic cell **2**(2): 362-379.

Kindle, K. L., R. A. Schnell, E. Fernandez and P. A. Lefebvre (1989). "Stable nuclear transformation of *Chlamydomonas* Using the *Chlamydomonas* gene for nitrate reductase." Journal of Cell Biology **109**(6): 2589-2601.

Klein, U., M. L. Salvador and L. Bogorad (1994). "Activity of the *Chlamydomonas* chloroplast *rbcL* gene promoter is enhanced by a remote sequence element." Proceedings of the National Academy of Sciences of the United States of America **91**(23): 10819-10823.

Komine, Y., L. Kwong, M. C. Anguera, G. Schuster and D. B. Stern (2000). "Polyadenylation of three classes of chloroplast RNA in *Chlamydomonas reinhadtii*." RNA **6**(4): 598-607.

Kozak, M. (2005). "Regulation of translation via mRNA structure in prokaryotes and eukaryotes." Gene **361**: 13-37.

Kraszewska, E. (2008). "The plant Nudix hydrolase family." Acta Biochimica Polonica **55**(4): 663-671.

Lehnik-Habrink, M., R. J. Lewis, U. Mader and J. Stulke (2012). "RNA degradation in *Bacillus subtilis*: an interplay of essential endo- and exoribonucleases." Molecular microbiology **84**(6): 1005-1017.

Leon-Banares, R., D. Gonzalez-Ballester, A. Galvan and E. Fernandez (2004). "Transgenic microalgae as green cell-factories." Trends in biotechnology **22**(1): 45-52.

Lumbreras, V., D. R. Stevens and S. Purton (1998). "Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron." Plant Journal **14**(4): 441-447.

Mackie, G. A. (1998). "Ribonuclease E is a 5'-end-dependent endonuclease." Nature **395**(6703): 720-723.

Mackie, G. A. (2000). "Stabilization of circular rpsT mRNA demonstrates the 5 '-end dependence of RNase E action in vivo." Journal of Biological Chemistry **275**(33): 25069-25072.

Maliga, P. and R. Bock (2011). "Plastid biotechnology: food, fuel, and medicine for the 21st century." Plant physiology **155**(4): 1501-1510.

Mathy, N., L. Benard, O. Pellegrini, R. Daou, T. Wen and C. Condon (2007). "5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA." Cell **129**(4): 681-692.

Matsuo, T., K. Okamoto, K. Onai, Y. Niwa, K. Shimogawara and M. Ishiura (2008). "A systematic forward genetic analysis identified components of the *Chlamydomonas* circadian system." Genes & development **22**(7): 918-930.

Maul, J. E., J. W. Lilly, L. Cui, C. W. dePamphilis, W. Miller, E. H. Harris and D. B. Stern (2002). "The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats." The Plant cell **14**(11): 2659-2679.

McLennan, A. G. (2006). "The Nudix hydrolase superfamily." Cellular and molecular life sciences : CMLS **63**(2): 123-143.

Merchant, S. S., S. E. Prochnik, O. Vallon, E. H. Harris, S. J. Karpowicz, G. B. Witman, A. Terry, A. Salamov, L. K. Fritz-Laylin, L. Marechal-Drouard, W. F. Marshall, L. H. Qu, D. R. Nelson, A. A. Sanderfoot, M. H. Spalding, V. V. Kapitonov, Q. Ren, P. Ferris, E. Lindquist, H. Shapiro, S. M. Lucas, J. Grimwood, J. Schmutz, P. Cardol, H. Cerutti, G. Chanfreau, C. L. Chen, V. Cognat, M. T. Croft, R. Dent, S. Dutcher, E. Fernandez, H. Fukuzawa, D. Gonzalez-Ballester, D. Gonzalez-Halphen, A. Hallmann, M. Hanikenne, M. Hippler, W. Inwood, K. Jabbari, M. Kalanon, R. Kuras, P. A. Lefebvre, S. D. Lemaire, A. V. Lobanov, M. Lohr, A. Manuell, I. Meier, L. Mets, M. Mittag, T. Mittelmeier, J. V. Moroney, J. Moseley, C. Napoli,

References

A. M. Nedelcu, K. Niyogi, S. V. Novoselov, I. T. Paulsen, G. Pazour, S. Purton, J. P. Ral, D. M. Riano-Pachon, W. Riekhof, L. Rymarquis, M. Schroda, D. Stern, J. Umen, R. Willows, N. Wilson, S. L. Zimmer, J. Allmer, J. Balk, K. Bisova, C. J. Chen, M. Elias, K. Gendler, C. Hauser, M. R. Lamb, H. Ledford, J. C. Long, J. Minagawa, M. D. Page, J. Pan, W. Pootakham, S. Roje, A. Rose, E. Stahlberg, A. M. Terauchi, P. Yang, S. Ball, C. Bowler, C. L. Dieckmann, V. N. Gladyshev, P. Green, R. Jorgensen, S. Mayfield, B. Mueller-Roeber, S. Rajamani, R. T. Sayre, P. Brokstein, I. Dubchak, D. Goodstein, L. Hornick, Y. W. Huang, J. Jhaveri, Y. Luo, D. Martinez, W. C. Ngau, B. Otillar, A. Poliakov, A. Porter, L. Szajkowski, G. Werner, K. Zhou, I. V. Grigoriev, D. S. Rokhsar and A. R. Grossman (2007). "The *Chlamydomonas* genome reveals the evolution of key animal and plant functions." Science **318**(5848): 245-250.

Merendino, L., K. Perron, M. Rahire, I. Howald, J. D. Rochaix and M. Goldschmidt-Clermont (2006). "A novel multifunctional factor involved in trans-splicing of chloroplast introns in *Chlamydomonas*." Nucleic acids research **34**(1): 262-274.

Meslet-Cladiere, L. and O. Vallon (2011). "Novel shuttle markers for nuclear transformation of the green alga *Chlamydomonas reinhardtii*." Eukaryotic cell **10**(12): 1670-1678.

Meyer, S., C. Temme and E. Wahle (2004). "Messenger RNA turnover in eukaryotes: pathways and enzymes." Critical reviews in biochemistry and molecular biology **39**(4): 197-216.

Michaelson, D. and M. Philips (2006). "The use of GFP to localize Rho GTPases in living cells." Method Enzymol **406**: 296-315.

Michelet, L., L. Lefebvre-Legendre, S. E. Burr, J. D. Rochaix and M. Goldschmidt-Clermont (2011). "Enhanced chloroplast transgene expression in a nuclear mutant of *Chlamydomonas*." Plant biotechnology journal **9**(5): 565-574.

Molnar, A., A. Bassett, E. Thuenemann, F. Schwach, S. Karkare, S. Ossowski, D. Weigel and D. Baulcombe (2009). "Highly specific gene silencing by artificial microRNAs in the unicellular alga *Chlamydomonas reinhardtii*." The Plant journal : for cell and molecular biology **58**(1): 165-174.

Mudd, E. A., H. M. Krisch and C. F. Higgins (1990). "RNase E, an endoribonuclease, has a general role in the chemical decay of *Escherichia coli* mRNA: evidence that *rne* and *ams* are the same genetic locus." Molecular microbiology **4**(12): 2127-2135.

Nakamura, Y., T. Gojobori and T. Ikemura (1999). "Codon usage tabulated from the international DNA sequence databases; its status 1999." Nucleic acids research **27**(1): 292.

Nakamura, Y., T. Gojobori and T. Ikemura (2000). "Codon usage tabulated from international DNA sequence databases: status for the year 2000." Nucleic acids research **28**(1): 292.

Nickelsen, J., M. Fleischmann, E. Boudreau, M. Rahire and J. D. Rochaix (1999). "Identification of cis-acting RNA leader elements required for chloroplast *psbD* gene expression in *Chlamydomonas*." The Plant cell **11**(5): 957-970.

Nishimura, Y. (2010). "Uniparental inheritance of cpDNA and the genetic control of sexual differentiation in *Chlamydomonas reinhardtii*." Journal of plant research **123**(2): 149-162.

Pesole, G., G. Grillo, A. Larizza and S. Liuni (2000). "The untranslated regions of eukaryotic mRNAs: structure, function, evolution and bioinformatic tools for their analysis." Briefings in bioinformatics **1**(3): 236-249.

Petersen, C. (1992). "Control of functional mRNA stability in bacteria: multiple mechanisms of nucleolytic and non-nucleolytic inactivation." Molecular microbiology **6**(3): 277-282.

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

Pfannschmidt, T., K. Schutze, V. Fey, I. Sherameti and R. Oelmuller (2003). "Chloroplast redox control of nuclear gene expression - A new class of plastid signals in interorganellar communication." Antioxidants & redox signaling **5**(1): 95-101.

Prasher, D. C., V. K. Eckenrode, W. W. Ward, F. G. Prendergast and M. J. Cormier (1992). "Primary structure of the *Aequorea victoria* green-fluorescent protein." Gene **111**(2): 229-233.

Rasala, B. A., M. Muto, J. Sullivan and S. P. Mayfield (2011). "Improved heterologous protein expression in the chloroplast of *Chlamydomonas reinhardtii* through promoter and 5' untranslated region optimization." Plant biotechnology journal **9**(6): 674-683.

Richards, J., Q. Liu, O. Pellegrini, H. Celesnik, S. Yao, D. H. Bechhofer, C. Condon and J. G. Belasco (2011). "An RNA pyrophosphohydrolase triggers 5'-exonucleolytic degradation of mRNA in *Bacillus subtilis*." Molecular cell **43**(6): 940-949.

Richards, J., D. J. Luciano and J. G. Belasco (2012). "Influence of translation on RppH-dependent mRNA degradation in *Escherichia coli*." Molecular microbiology.

Rochaix, J. D. (2002). "*Chlamydomonas*, a model system for studying the assembly and dynamics of photosynthetic complexes." FEBS letters **529**(1): 34-38.

Salvador, M. L. and U. Klein (1999). "The redox state regulates RNA degradation in the chloroplast of *Chlamydomonas reinhardtii*." Plant physiology **121**(4): 1367-1374.

Salvador, M. L., L. Suay, I. L. Anthonisen and U. Klein (2004). "Changes in the 5'-untranslated region of the *rbcL* gene accelerate transcript degradation more than 50-fold in the chloroplast of *Chlamydomonas reinhardtii*." Current genetics **45**(3): 176-182.

Salvador, M. L., L. Suay and U. Klein (2011). "Messenger RNA degradation is initiated at the 5' end and follows sequence- and condition-dependent modes in chloroplasts." Nucleic acids research **39**(14): 6213-6222.

Sambrook, J. and D. W. Russell (2001). Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.

Schulze, T., K. Prager, H. Dathe, J. Kelm, P. Kiessling and M. Mittag (2010). "How the green alga *Chlamydomonas reinhardtii* keeps time." Protoplasma **244**(1-4): 3-14.

Schuster, G., I. Lisitsky and P. Klaff (1999). "Polyadenylation and degradation of mRNA in the chloroplast." Plant physiology **120**(4): 937-944.

Sharp, J. S. and D. H. Bechhofer (2005). "Effect of 5'-proximal elements on decay of a model mRNA in *Bacillus subtilis*." Molecular microbiology **57**(2): 484-495.

Shimomura, O., F. H. Johnson and Y. Saiga (1962). "Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*." Journal of cellular and comparative physiology **59**: 223-239.

Sizova, I., M. Fuhrmann and P. Hegemann (2001). "A Streptomyces rimosus *aphVIII* gene coding for a new type phosphotransferase provides stable antibiotic resistance to Chlamydomonas reinhardtii." Gene **277**(1-2): 221-229.

Smith, D. R. and R. W. Lee (2008). "Nucleotide diversity in the mitochondrial and nuclear compartments of *Chlamydomonas reinhardtii*: investigating the origins of genome architecture." BMC evolutionary biology **8**: 156.

Stern, D. B., M. Goldschmidt-Clermont and M. R. Hanson (2010). "Chloroplast RNA metabolism." Annual review of plant biology **61**: 125-155.

Suay, L., M. L. Salvador, E. Abesha and U. Klein (2005). "Specific roles of 5' RNA secondary structures in stabilizing transcripts in chloroplasts." Nucleic acids research **33**(15): 4754-4761.

Sueoka, N. (1960). "Mitotic replication of deoxyribonucleic acid in *Chlamydomonas reinhardtii*." Proceedings of the National Academy of Sciences of the United States of America **46**(1): 83-91.

Toepel, J., S. P. Albaum, S. Arvidsson, A. Goesmann, M. la Russa, K. Rogge and O. Kruse (2011). "Construction and evaluation of a whole genome microarray of *Chlamydomonas reinhardtii*." BMC genomics **12**: 579.

Trebitsh, T., A. Levitan, A. Sofer and A. Danon (2000). "Translation of chloroplast *psbA* mRNA is modulated in the light by counteracting oxidizing and reducing activities." Molecular and cellular biology **20**(4): 1116-1123.

Vincent, H. A. and M. P. Deutscher (2009). "The roles of individual domains of RNase R in substrate binding and exoribonuclease activity. The nuclease domain is sufficient for digestion of structured RNA." The Journal of biological chemistry **284**(1): 486-494.

Winck, F. V., D. M. Riano-Pachon, F. Sommer, J. Rupprecht and B. Mueller-Roeber (2012). "The nuclear proteome of the green alga *Chlamydomonas reinhardtii*." Proteomics **12**(1): 95-100.

Woodson, J. D. and J. Chory (2008). "Coordination of gene expression between organellar and nuclear genomes." Nature Reviews Genetics **9**(5): 383-395.

Zerges, W., J. GirardBascou and J. D. Rochaix (1997). "Translation of the chloroplast *psbC* mRNA is controlled by interactions between its 5' leader and the nuclear loci *TBC1* and *TBC3* in *Chlamydomonas reinhardtii*." Molecular and cellular biology **17**(6): 3440-3448.

Zhao, T., G. Li, S. Mi, S. Li, G. J. Hannon, X. J. Wang and Y. Qi (2007). "A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*." Genes & development **21**(10): 1190-1203.

Zhelyazkova, P., K. Hammani, M. Rojas, R. Voelker, M. Vargas-Suarez, T. Borner and A. Barkan (2012). "Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts." Nucleic acids research **40**(7): 3092-3105.

Zuker, M., D. H. Mathews and D. H. Turner (1999). "Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide." Nato Asi 3 High Technology **70**: 11-43.

APPENDIX I

Agarose gel electrophoresis:

TAE (50×)/L *Loading buffer*

242g Tris base 0.25% bromphenol blue

57.1 mL glacial acetic acid……0.25% xylene cyanol FF

100 mL 0.5M EDTA (pH 8.0) 30% glycerol

LB medium/L: **LBAmp/L Plates**:

10 g tryptone 10 g tryptone

5 g yeast extract 5 g yeast extract

10 g NaCl 10 g NaCl

15 g agar

100 µg Amp

psaD gene

HS media

Salt stock (50×)/0.5 L: *Phosphate stock* (50×)/0.5 L: *Hutner trace elements*/L:

12.5g NH_4Cl 47g $k_2HPO_4.3H_2O$

0.50g $MgSO_4.7H_2O$ 18g KH_2PO_4

0.25g $CaCl_2.2H_2O$

compound	amount	water
EDTA disodium salt	50 g	250 mL
$ZnSO_4$. 7 H_2O	22 g	100 mL
H_3BO_3	11.4 g	200 mL
MnCl2 . 4 H_2O	5.06 g	50 mL
CoCl2. 6 H_2O	1.61 g	50 mL
$CuSO_4$. 5 H_2O	1.57 g	50 mL
$(NH_4)6Mo7O_{24}$. 4 H_2O	1.10 g	50 mL
$FeSO_4$. 7 H_2O	4.99 g	50 mL

APPENDIX II

SDS page buffers:

Solution A (Acrylamide)-[30%]

Solution B (Tris-HCl)-[1.5M]

Solution B' (Tris-HCL)- [0.5M]

12% polyacryl amide gel:

dH_2O-1.425 mL

Solution A -1.65 mL

Solution B-1.03mL

Ammonium persulfate [10%]-16.25 µL

4% Stacking gel:

dH_2O-1.8 mL

Solution A-0.4 mL

Solution B'-0.75 mL

Ammonium persulfate [10%]-30 µL

1×electrode buffer:

Tris base 3 g

Glycine 14.6 g

SDS 1 g

Affinity chromatography buffers:

Buffer A (*Binding/wash buffer*): Sodium phosphate-[20 mM]
NaCl-[500 mM]
Imidazole, pH 7.3-[20 mM]

Buffer B (*Elution buffer*): Sodium phosphate-[20 mM]
NaCl-[500 mM]
Imidazole, pH 7.3-[500 mM]

APPENDIX III

pChlami-5'*rppH-GFP 6HN* transformation vector sequence

```
   1                                                  C ATATGGACGA GGGAGACGCG
  22  CCCGCTGCCG CCCCTGGTGG CGTTCGGCTG CTTGTGGTGG TGGGTGTGGT CCTGCTGGAC
  82  GACCCACTGT GGGACCACGA GACGGGCGAG CCCAGCAAGG GCGCCGCCGC CGCGGATCGG
 142  CCTGTGCGGG TGCTGCTGGC TCGGCGGCCT GTGGGCAAGA GCAACGCGGG GCTGTGGGAG
 202  TTCCCGGGCA TGGCCCAGAG CAAGCACGGC CTGACCAAGG AGATGACCAT GAAGTACCGC
 262  ATGGAGGGCT GCGTGGACGG CCACAAGTTC GTGATCACCG GCGAGGGCAT CGGCTACCCC
 322  TTCAAGGGCA AGCAGGCCAT CAACCTGTGC GTGGTGGAGG GCGGCCCCCT GCCCTTCGCC
 382  GAGGACATCC TGAGCGCCGC CTTCATGTAC GGCAACCGCG TGTTCACCGA GTACCCCCAG
 442  GACATCGTGG ACTACTTCAA GAACAGCTGC CCCGCCGGCT ACACCTGGGA CCGCTCGTTC
 502  CTGTTCGAGG ACGGCGCCGT GTGCATCTGC AACGCCGACA TCACCGTGTC CGTGGAGGAG
 562  AACTGCATGT ACCACGAGAG CAAGTTCTAC GGCGTGAACT TCCCCGCCGA CGGCCCCGTG
 622  ATGAAGAAGA TGACCGACAA CTGGGAGCCC AGCTGCGAGA AGATCATCCC CGTGCCCAAG
 682  CAGGGCATCC TGAAGGGCGA CGTGTCCATG TACCTGCTGC TGAAGGACGG CGGTCGCCTG
 742  CGCTGCCAGT TCGACACCGT GTACAAGGCC AAGAGCGTGC CCCGCAAGAT GCCCGACTGG
 802  CACTTCATCC AGCACAAGCT GACCCGCGAG GACCGCAGCG ACGCCAAGAA CCAGAAGTGG
 862  CACCTGACCG AGCACGCGAT CGCCAGCGGC AGCGCCCTGC CCGGCGGCCG CCTGGTTCCG
 922  CGTGGCTCTC CGGGCGCTGC AGGTCATAAT CATAATCATA ATCATAATCA TAATCACAAT
 982  TAATTAATTA ATCTAGAtgg cagcagctgg accgcctgta ccatggagaa gagctttact
1042  tgccgggatg gccgatttcg ctgattgata cgggatcgga gctcggaggc tttcgcgcta
1102  ggggctaggc gaagggcagt ggtgaccagg gtcggtgtgg ggtcggccca cggtcaatta
1162  gccacaggag gatcaggggg aggtaggcac gtcgacttgg tttgcgaccc cgcagttttg
1222  gcggacgtgc tgttgtagat gttagcgtgt gcgtgagcca gtggccaacg tgccacaccc
1282  attgagaaga ccaaccaact tactggcaat atctgccaat gccatactgc atgtaatggc
1342  caggccatgt gagagtttgc cgtgcctgcg cgcgccccgg gggcgcagtt tagctgacca
1402  gccgtgggat gatgcacgca tttgcaagga cagggtaatc acagcagcaa catggtgggc
1462  ttaggacagc tgtgggtcag tggacggacg gcaggggagg gacggcgcag ctcgggagac
1522  agggggagac agcgtgactg tgcaatgcgg ccgccaccgc ggtggagctc caattcgccc
1582  tatagtgagt cgtattacgc gcgctcactg gccgtcgttt tacaacgtcg tgactgggaa
1642  aaccctggcg ttacccaact taatcgcctt gcagcacatc ccccttcgc cagctggcgt
1702  aatagcgaag aggcccgcac cgatcgccct tcccaacagt tgcgcagcct gaatggcgaa
1762  tgggacgcgc cctgtagcgg cgcattaagc gcggcgggtg tggtggttac gcgcagcgtg
1822  accgctacac ttgccagcgc cctagcgccc gctcctttcg ctttcttccc ttcctttctc
1882  gccacgttcg ccggctttcc ccgtcaagct ctaaatcggg ggctcccttt agggttccga
1942  tttagtgctt tacggcacct cgaccccaaa aaacttgatt agggtgatgg ttcacgtagt
2002  gggccatcgc cctgatagac ggtttttcgc cctttgacgt tggagtccac gttctttaat
2062  agtggactct tgttccaaac tggaacaaca ctcaacccta tctcggtcta ttcttttgat
2122  ttataaggga ttttgccgat ttcggcctat tggttaaaaa atgagctgat ttaacaaaaa
2182  tttaacgcga attttaacaa aatattaacg cttacaattt aggtggcact tttcggggaa
2242  atgtgcgcgg aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca
2302  tgagacaata accctgataa atgcttcaat aatattgaaa aaggaagagt atgagtattc
2362  aacatttccg tgtcgccctt attccctttt ttgcggcatt ttgccttcct gtttttgctc
2422  acccagaaac gctggtgaaa gtaaaagatg ctgaagatca gttgggtgca cgagtgggtt
2482  acatcgaact ggatctcaac agcggtaaga tccttgagag ttttcgcccc gaagaacgtt
2542  ttccaatgat gagcacttttt aaagttctgc tatgtggcgc ggtattatcc cgtattgacg
2602  ccgggcaaga gcaactcggt cgccgcatac actattctca gaatgacttg gttgagtact
2662  caccagtcac agaaaagcat cttacggatg gcatgacagt aagagaatta tgcagtgctg
2722  ccataaccat gagtgataac actgcggcca acttacttct gacaacgatc ggaggaccga
2782  aggagctaac cgcttttttg cacaacatgg gggatcatgt aactcgcctt gatcgttggg
2842  aaccggagct gaatgaagcc ataccaaacg acgagcgtga caccacgatg cctgtagcaa
2902  tggcaacaac gttgcgcaaa ctattaactg gcgaactact tactctagct tcccggcaac
2962  aattaataga ctggatggag gcggataaag ttgcaggacc acttctgcgc tcggcccttc
3022  cggctggctg gtttattgct gataaatctg gagccggtga gcgtgggtct cgcggtatca
3082  ttgcagcact ggggccagat ggtaagccct cccgtatcgt agttatctac acgacgggga
3142  gtcaggcaac tatggatgaa cgaaatagac agatcgctga gataggtgcc tcactgatta
```

```
3202 agcattggta actgtcagac caagtttact catatatact ttagattgat ttaaaacttc
3262 atttttaatt taaaaggatc taggtgaaga tccttttttga taatctcatg accaaaatcc
3322 cttaacgtga gttttcgttc cactgagcgt cagaccccgt agaaaagatc aaaggatctt
3382 cttgagatcc ttttttttctg cgcgtaatct gctgcttgca aacaaaaaaa ccaccgctac
3442 cagcggtggt ttgtttgccg gatcaagagc taccaactct ttttccgaag gtaactggct
3502 tcagcagagc gcagatacca aatactgtcc ttctagtgta gccgtagtta ggccaccact
3562 tcaagaactc tgtagcaccg cctacatacc tcgctctgct aatcctgtta ccagtggctg
3622 ctgccagtgg cgataagtcg tgtcttaccg ggttggactc aagacgatag ttaccggata
3682 aggcgcagcg gtcgggctga acgggggttt cgtgcacaca gcccagcttg gagcgaacga
3742 cctacaccga actgagatac ctacagcgtg agctatgaga aagcgccacg cttcccgaag
3802 ggagaaaggc ggacaggtat ccggtaagcg gcagggtcgg aacaggagag cgcacgaggg
3862 agcttccagg gggaaacgcc tggtatcttt atagtcctgt cgggtttcgc cacctctgac
3922 ttgagcgtcg atttttgtga tgctcgtcag gggggcggag cctatggaaa aacgccagca
3982 acgcggcctt tttacggttc ctggcctttt gctggccttt tgctcacatg ttctttcctg
4042 cgttatcccc tgattctgtg gataaccgta ttaccgcctt tgagtgagct gataccgctc
4102 gccgcagccg aacgaccgag cgcagcgagt cagtgagcga ggaagcggaa gagcgcccaa
4162 tacgcaaacc gcctctcccc gcgcgttggc cgattcatta atgcagctgg cacgacaggt
4222 ttcccgactg gaaagcgggc agtgagcgca acgcaattaa tgtgagttag ctcactcatt
4282 aggcacccca ggctttacac tttatgcttc cggctcgtat gttgtgtgga attgtgagcg
4342 gataacaatt tcacacagga aacagctatg accatgatta cgccaagcgc gcaattaacc
4402 ctcactaaag ggaacaaaag ctgggtaccc gcttcaaata cgcccagccc gcccatggag
4462 aaagaggcca aaatcaacgg aggatcgtta caaccaacaa aattgcaaaa ctcctccgct
4522 ttttacgtgt tgaaaaagac tgatcagcac gaaacgggga gctaagctac cgcttcagca
4582 cttgagagca gtatcttcca tccaccgccg ttcgtcaggg ggcaaggctc agatcaacga
4642 gcgcctccat ttacacggag cggggatccc aacgtccaca ctgtgctgtc acccacgcga
4702 cgcaacccta cccagccacc aacaccatca ggtccctcag aagaactcgt ccaacagccg
4762 gtaaaacgcc agcttttcct ccgataccgc cccatcccac ccgcgcccgt actcccgcag
4822 gaacgccgcg gaacactccg gcccgaacca cgggtcctcc tcgtgggcca gctcgcgcag
4882 caccagcgcg agatcggagt gccggtccgc acggccgacc cgccccacgt cgatcagccc
4942 ggtcacctcg caggtacgag ggtcgagcag cacgttgtcc gggcacaggt gaccgtggca
5002 aaccgccaga tcctcgtccg caggccgagt ccgctccagc tcggcgagaa gccgctcccc
5062 cgaccacccc ttccgctcct cgtccagatc ctccaagtcg acgctccctt cagcgacagc
5122 acgggccgcc tgcggcaccg tcaccgcgag actgcgatcg aacggacacc gctcccagtc
5182 cagcgcgtgc agcgaacgag cgagccccgc gagcgccacc gccacgtcca gccgctgctc
5242 ccgcggccac cgcgcactgg ccggacgccc cggaaccgct tcggtgacca accaggcgac
5302 cctctcgtcc ccaccaccct ccacaacacg aggtacggga atcccacct ccgccaacca
5362 caccagccgc tcagcctcac ccaacaagcc caccccggcc cccagagctg ccaccttgac
5422 aaacaactcc cgcccaccac cccgaagccg ataaacacca gcccccgagg ccccatcctc
5482 cacaacaacc cactcacaac cgggataccg accccgcagt gcacgcaacg catcgtccat
5542 gcttcgaaat tcttcagcac cggggagggc ggagtggcca tcctgcaaat ggaaacggcg
5602 acgcagggtt agatgctgct tgagacagcg acagaggagc caaaagcctt cgtcgacaca
5662 atgcgggcgt tgcaagtcaa atctgcaagc acgctgcctg atccgccggg cttgctcgtc
5722 gactcacctg gccattttaa gatgttgagt gacttctctt gtaaaaaagt aaagaacata
5782 ggccccctgg ccggtttatc aggagggcac cgctccaggg gctgcatgcg aactgcttgc
5842 attggcgcct agcctttgtg ggccaggggg cttccggata agggttgcaa gtgctcaaat
5902 accccatcaa acatcatcct ggtttggctg cgctccttct ggcgcgcccg gcatgcaagc
5962 ttgatgggat cttaagctag ctgagtggtt atgtatagcg gcagaatagt cgcgtatgta
6022 taagtgctcg tttgtcgctg aaagtggagg tcaccgttcg gggtcgcggg cttttatacc
6082 ggatgggtgc cgccagcggg ccgtatggcg ccttctggac gccgcgcgcc ccatcgcggc
6142 ccttccagag cttcccgcgc cctcatagcc cgccaaatca gtcctgtagc ttcatacaaa
6202 catacgcacc aatcatgtca agcctcagcg agctccccgc tcgagcacac acctgcccgt
6262 ctgcctgaca ggaagtgaac gcatgtcgag ggaggcctca ccaatcgtca cacgagccct
6322 cgtcagaaac acgtctccgc cacgctctcc ctctcacggc cgaccccgca gccctttgc
6382 cctttcctag gccaccgaca ggacccaggc gctctcagca tgcctcaaca acccgtactc
6442 gtgccagcgg tgcccttgtg ctggtgatcg cttggaagcg catgcgaaga cgaaggggcg
6502 gagcaggcgg cctggctgtt cgaagggctc gccgccagtt cgggtgcctt tctccacgcg
6562 cgcctccaca cctaccgatg cgtgaaggca ggcaaatgct catgtttgcc cgaactcgga
6622 gtccttaaaa agccgcttct tgtcgtcgtt ccgagacatg ttagcagatc gcagtgccac
6682 ctttcctgac gcgctcggcc ccatattcgg acgcaattgt catttgtagc acaattggag
6742 caaatctggc gaggcagtag gcttttaagt tgcaaggcga gagagcaaag tgggacgcgg
6802 cgtgattatt ggtatttacg cgacggcccg gcgcgttagc ggcccttccc ccaggccagg
```

```
6862 gacgattatg tatcaatatt gttgcgttcg ggcactcgtg cgagggctcc tgcgggctgg
6922 ggaggggggat ctgggaattg gaggtacgac cgagatggct tgctcggggg gaggtttcct
6982 cgccgagcaa gccagggtta ggtgttgcgc tcttgactcg ttgtgcattc taggacccca
7042 ctgctactca caacaagccc a
```

Features:

Purple-**5'rppH** (4..210).

Green-**GFP** (210..906).

Light Blue-**Histidine tag** (946..982).

Orange-**3'PSAD terminator region** (994..1379).

Dark blue- **HSP70A-RBCS2_promoter::aphVIII::RBCS2_terminator** (4430..6241).

Red-**5'PSAD promoter region** (6242..7062).

Double Underlined-**Different primers used** .

Single Underlined-***Nde*I/*Xba*I/*Not*I/*Xho*I/*Sma*I/*Eag*I restriction sites**.

TT-The 5' and 3' end of second and first primers in their respective strand.

APPENDIX IV

Codon Usage Table

***Chlamydomonas reinhardtii* [gbpln]: 846 CDS's (420455 codons)**

fields: [triplet] [amino acid] [fraction] [frequency: **per thousand**] ([number])

UUU	F	0.16	5.0	(2110)	UCU	S	0.07	4.7	(1992)	UAU	Y	0.10	2.6	(1085)	UGU	C	0.10	1.4	(601)
UUC	F	0.84	27.1	(11411)	UCC	S	0.25	16.1	(6782)	UAC	Y	0.90	22.8	(9579)	UGC	C	0.90	13.1	(5498)
UUA	L	0.01	0.6	(247)	UCA	S	0.05	3.2	(1348)	UAA	*	0.52	1.0	(441)	UGA	*	0.27	0.5	(227)
UUG	L	0.04	4.0	(1673)	UCG	S	0.25	16.1	(6763)	UAG	*	0.22	0.4	(183)	UGG	W	1.00	13.2	(5559)
CUU	L	0.05	4.4	(1869)	CCU	P	0.13	8.1	(3416)	CAU	H	0.11	2.2	(919)	CGU	R	0.09	4.9	(2071)
CUC	L	0.15	13.0	(5480)	CCC	P	0.47	29.5	(12409)	CAC	H	0.89	17.2	(7252)	CGC	R	0.62	34.9	(14676)
CUA	L	0.03	2.6	(1086)	CCA	P	0.08	5.1	(2124)	CAA	Q	0.10	4.2	(1780)	CGA	R	0.04	2.0	(841)
CUG	L	0.73	65.2	(27420)	CCG	P	0.33	20.7	(8684)	CAG	Q	0.90	36.3	(15283)	CGG	R	0.20	11.2	(4711)
AUU	I	0.22	8.0	(3360)	ACU	T	0.10	5.2	(2171)	AAU	N	0.09	2.8	(1157)	AGU	S	0.04	2.6	(1089)
AUC	I	0.75	26.6	(11200)	ACC	T	0.52	27.7	(11663)	AAC	N	0.91	28.5	(11977)	AGC	S	0.35	22.8	(9590)
AUA	I	0.03	1.1	(443)	ACA	T	0.08	4.1	(1713)	AAA	K	0.05	2.4	(1028)	AGA	R	0.01	0.7	(287)
AUG	M	1.00	25.7	(10796)	ACG	T	0.30	15.9	(6684)	AAG	K	0.95	43.3	(18212)	AGG	R	0.05	2.7	(1150)
GUU	V	0.07	5.1	(2158)	GCU	A	0.13	16.7	(7030)	GAU	D	0.14	6.7	(2805)	GGU	G	0.11	9.5	(3984)
GUC	V	0.22	15.4	(6496)	GCC	A	0.43	54.6	(22960)	GAC	D	0.86	41.7	(17519)	GGC	G	0.72	62.0	(26064)
GUA	V	0.03	2.0	(857)	GCA	A	0.08	10.6	(4467)	GAA	E	0.05	2.8	(1172)	GGA	G	0.06	5.0	(2084)
GUG	V	0.67	46.5	(19558)	GCG	A	0.35	44.4	(18688)	GAG	E	0.95	53.5	(22486)	GGG	G	0.11	9.7	(4087)

Coding GC 66.30% 1st letter GC 64.80% 2nd letter GC 47.90% 3rd letter GC 86.21%
Genetic code 1: Standard

(Nakamura et al. 2000).

APPENDIX V

Primer List:

Genomic DNA amplification:

PCR at annealing temperature of 63 ˚C

5’PSAD6401 : 5’- CGCCGAGCAAGCCAGGGTTA-3’ T_m (63.5˚C)

3’PSAD495 : 5’-GCGAAAGCCTCCGAGCTCCGAT-3’ T_m (65.8˚C)

PCR at annealing temperature of 63.5 ˚C

NUD3’-5COMP: 5’-TTCGGCTGCTTGTGGTGGTGGG-3’ T_m (54˚C)

3’GFP: 5’-GTCATCTCCTTGGTCAGGCCG-3’ T_m (64 ˚C)

PCR at annealing temperature of 61 ˚C

5’paro4331:5’-ACGGCCGACCCGCCCCACGT-3’ T_m (69.6 ˚C)

3’paro4743-:5’-GATTCCCGTACCTCGTGTTGT-3’ T_m (59.8 ˚C)

PCR at annealing temperature of 61 ˚C

5’NUD GFP:5’-CTGCCGCCCCTGTGGCGTT-3’ T_m (67.6 ˚C)

3’ PSAD 495: 5’-GCGAAAGCCTCCGAGCTCCGAT-3’ T_m (65.8˚C)

Plasmid DNA amplification:

PCR at annealing temperature of 65 ˚C, pBluescriptII SK+

5'NUD-22 Forward : 5’-ACCTTTGCCACCATGGACGAG-3’

5'NUD-22 Reverse : 5’-GGAAACAGCTATGACCATG-3’

PCR PCR at annealing temperature of 65 ˚C ,NcoI→NdeI

Forward : 5’-AAGAAGGAGATATACTATGGACGA-3’

5’-AAGAAGGAGATATAATATGGACGA-3’

5’-AAGAAGGAGATATCATATGGACGA-3’

Reverse : 3’-CCAACTCAGCTTCCTCTAGAT-5’

APPENDIX VI

Abbreviations:

Ampr-Ampiciline resistant

AphVIII- Aminoglycoside 3'-phosphotransferase type VIII

APS- Ammonium persulfate

MCS-Multiple cloning sites

NEP- Nuclear-encoded plastid RNA polymerase

ORF-Open reading frame

PEP- Plastid encoded RNA polymerase

PPR- Pentatricopeptide repeat proteins

RBCS2-HSP 70- Ribulose bisphosphate carboxylase small chain 2-Heat shock prtein 70

Rpm-Rounds per minute

SDS-PAGE-Sodium dodecyl sulfate polyacrylamide gel electrophoresis

TAE- Tris-acetate-EDTA

UTR- Untranslated region

APPENDIX VII

Bioinformatic tools:

C.U.D (http //www.kazusa.or.jp/codon)

Expasy (http://www.expasy.org/)

GCUA (http://gcua.schoedl.de/)

KEGG pathway (http://www.genome.jp/kegg/pathway.html)

NCBI (http://www.ncbi.nlm.nih.gov/)

NEB cutter V2.0 (http://tools.neb.com/NEBcutter2/)

Uniprot (http://www.uniprot.org/)