Molecular tools and approaches for increasing complexity of transplastomic engineering in *Chlamydomonas reinhardtii*

Marco Larrea-Álvarez

A dissertation submitted

for the degree of

Doctor of Philosophy in Molecular Microbiology

of

University College London

15 - 02 - 2018

Declaration

I, Marco Larrea-Álvarez, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

.....

Abstract

Microalgae have been grown commercially for many years as a feedstock for aquaculture, and as a source of natural products such as pigments and omega-3 oils. More recently, the focus has turned to the development of microalgae as industrial biotechnological platforms, and several species are being explored as possible cell factories. One promising species is the green alga *Chlamydmonas reinhardtii*, which has been used extensively as a laboratory model for basic genetic and biochemistry studies. The repertoire of molecular and 'omics tools available for its genetic manipulation has increased significantly in recent years, and there have been numerous reports of recombinant proteins produced successfully in this alga. However, examples of metabolic engineering are limited, and this reflects the more challenging problem of introducing multiple transgenes, and also regulating their expression so that biomass production and redirecting metabolic flux towards a desired product can be temporally separated. The algal chloroplast is a unique sub-cellular compartment with its own genetic system, which houses core metabolic pathways for carbohydrates, fatty acids, terpenoids and tetrapyrroles. It is therefore an attractive site for metabolic engineering. Modification of the *C. reinhardtii* chloroplast genome is routine and many single proteins have been synthesised in the organelle. Thus, the challenge is to increase the complexity of chloroplast engineering to allow sophisticated metabolic engineering. To contribute to this goal, my study has focused on several aspects: i) the simultaneous expression of multiple transgenes; ii) possible novel metabolism that could be introduced into the organelle: iii) the development of a novel fluorescence reporter for assessing transgene regulation. For the first project, a combinatorial/serial approach where various expression cassettes were combined into two transformation plasmids, was demonstrated to be efficient for expressing simultaneously three recom-

binant proteins. In the second project, nitrogen fixation was investigated as a potential metabolic pathway. The advantages of the organelle to meet the nitrogenase enzymes requirements are discussed; furthermore, the expression of *nifV* from a suggested minimal set of *nif* genes, is demonstrated. Finally, the expression of a cyan fluoresecent reporter (mTurq2cp) was achieved. Emission at the expected wavelength was detected, and the fold increase in fluorescence above the background was determined. The presented results help to enlarge the repertoire of molecular tools and know-how available for chloroplast engineering in *C. reinhardtii*.

Acknowledgements

Firstly, I would like to express my gratitude to my project supervisor Prof Saul Purton, for the guidance and support during this investigation. I am very grateful for all his positive advices, and for the freedom he gave me to follow my own ideas.

I would like to thank the members of the Purton group. A special thanks to Dr Rosie Young for helping me find my way in the laboratory. Also, Dr Laura Stoffels, Dr Priscilla Rajakumar, Dr Umaima Al-Hoqani, Dr Alice Lui, Dr Max Blandshard, Dr Janet Waterhouse, Juliana Ramos, Xenia Spencer-Milnes and Fiona Li for being such a great group to work with. Also, I would like to express my gratitude to Thushi Sivagnanam for her invaluable effort in keeping the laboratory organised.

Pursuing a PhD program at UCL was possible thanks to the support of the Ecuadorian government via the Top World Universities scholarship program (2014) of the Secretariat of Higher Education, Science, Technology and Innovation (SENESCYT).

I would like to thank my close friends who have been a great influence in my life. Francisco Benítez, Juan Pablo Almeida, David Ortega, Fausto Espinosa for the years of friendship. Carlos Rodríguez, Antonio de la Piedra, Felipe Reinoso and the Brussels gang for such funny and unforgettable times. Mario Hidalgo, Ximena Coello, and the rest of the ecuadorians in London for the great memories. A very special thanks to my close friends during this journey. Diana Ochoa for taking me to great concerts! Joanna Firth for being the best housemate ever! and Enrique López for the constant and free of charge psychoanalysis!

There are no words to describe the gratitude towards my always supporting family. I would like to thank my sister Alicia and my little nephew Alonso for sparking the joy in the family. My grandmother for the constant

encouragement. My brother Marcelo for the great philosophical discussions and for teaching me how to use LaTeX, you did make my life easier bro! My loved uncle Carlos Marcelo for being always amazing to me and love me unconditionally. I love you. Finally, all my gratitude to my beloved parents, Marco and Rita; thank you for understanding me and let me pursue my dreams. Without you this would have not been possible. Les amo!

I don't know anything, but I do know that everything is interesting if you go into it deeply enough.

-Richard Feynman

Abbreviations

5-FC	5-fluorocytosine
5-FU	5-fluorouracil
AcCoA	acetyl coenzyme A
AMPS	ammonium persulfate
ATP	adenosine triphosphate
BLAST	basic local alignment search tool
BNF	biological nitrogen fixation
BSA	bovine serum albumin
CAI	codon adaption index
CFP	cyan fluorescent protein
Chl	chlorophyll
CoASH	coenzyme A
CTB	cholera toxin virus
DNA	deoxyribonucleic acid
dNTP	2' deoxynucleoside 5' triphosphate
DOPR	dark-operative protochlorophyllidae reductase
DTNB	5,5-dithiobis(2-nitrobenzoic acid)
EC	expression cassette
EDTA	ethylenediaminetetraacetic acid
FRET	fluorescence resonance energy transfer
GFP	green fluorescent protein
GOI	gene of interest
GS-GOGAT	glutamine synthase/glutamate synthase pathway
HA	human influenza hemagglutinin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hGH	human growth hormone
HSM	high salt minimal medium

IBV	infectious bronchitis virus
IPTG	isopropyl β-D-1-thiogalactopyranoside
IR	infra-red fluorescence
ITS	internal transcribed spacers
kb	kilobase
LB	lysogeny broth
mRNA	messenger ribonucleic acid
MYA	million years ago
NCBI	national center for biotechnology information
NIF	nitrogen fixation system
nm	nanometer
OD	optical density
ORF	open reading frame
PBS	phosphate-buffered saline
Pchlide	protochlorophyllidae
PCR	polymerase chain reaction
PEP	plastid-encoded polymerase
PSI	photosystem I
PSII	photosystem II
PSRP	plastid specific ribosomal proteins
rDNA	ribosomal deoxyribonucleic acid
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
SAM	S-adenosyl methionine
SDS	sodium dodecyl sulphate
SDS-PAGE	dodecyl sulfate polyacrylamide gel electrophoresis
SUF	sulphur assimilation system
TAG	triacylglyceride

TAP	tris acetate phosphate
TBS	tris buffered saline
TBS-T	tris buffer saline - tween 20
TEMED	tetramethylethylenediamine
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer ribonucleic acid
TSP	total soluble protein
UTR	untranslated region
v/v	volume for volume
w/v	weight for volume
WT	wild type

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 *Chlamydomonas reinhardtii* as an expression platform

C. reinhardtii is a single-celled eukaryotic green soil alga belonging to the Plantae kingdom, it has a nucleus, an eyespot, a pair of flagella and a chloroplast (Figure 1.1). *C. reinhardtii* has been extensively used as a model organism for investigating fundamental questions of cell biology and physiology including phototaxis, circadian rhythms and photosynthesis (Wakabayashi et al., 2011; Matsuo et al., 2008; Dent et al., 2005). *C. reinhardtii* carries a proteinaceous cell wall; nonetheless, there are several mutants that are incapable of producing it, and some have been used for genetic engineering purposes (Purton et al., 2013). Furthermore, individuals can reproduce either asexually or sexually (two types of vegetative haploid cells have been described: the so-called plus (+) and minus (-) mating types) (Figure 1.2). It has been seen that the inheritance of organelle genomes is gamete-dependent. Mitochondria are passed down to the offspring from

the minus type, while the chloroplast is inherited from the plus gamete. On the other hand, the nuclear genome appears to follow Mendelian rules (Goodenough et al., 2007). In recent decades, *C. reinhardtii* has been proposed as a platform for producing a wide range of important biomolecules; by manipulating environmental conditions the production of bio-diesel feedstock and biohydrogen has been achieved (Cakmak et al., 2012; Fouchard et al., 2005). Moreover, additional advantageous properties for molecular biology have been pointed out: its ease of growth, its ability to form colonies on agar plates, the relatively simple genetics of the nuclear, mitochondrial and chloroplastic genomes (Matsuo and Ishiura, 2011). Genetic engineering approaches applied to *C. reinhardtii* have led a proposed expression of a wide spectrum of valuable products such as biopharmaceuticals, biofuels, terpenoids or pigments (Lohr et al., 2012; Rosales-Mendoza et al., 2012; Cordero et al., 2011).

Rapid growth and short transformation-production time are fundamental for the use of this green alga as a cell factory. *C. reinhardtii* is able to reproduce asexually and it has been observed that around three and a half weeks are required to generate transformants using photosynthetic selection (Purton et al., 2013). Additionally, the ability to grow *C. reinhardtii* in containment conditions (photobioreactors) allows larger control over possible flows of contamination, and permits optimization of the process of production (Schenk et al., 2008). Based on the aforementioned features it can be argued that *C. reinhardtii* represents an efficient low cost platform for expression of recombinant proteins.

Figure 1.1: ***Chlamydomonas reinhardtii***. A. Schematic representation of the single-celled green alga showing some relevant physiological features. B. *Chlamydomonas reinhardtii*. Modified from (Merchant et al., 2007).

Figure 1.2: *C. reinhardtii* asexual and sexual reproduction. Asexual reproduction occurs when conditions are favorable; when conditions turn unfavorable the algae reproduce sexually. A and B. During asexual reproduction both mating types (mt^+ and mt^-) go through two or more rounds of mitosis and cytokinesis before the hatching of the daughter cells. C. Unfavourable conditions (e.g. nitrogen limitation) trigger sexual reproduction, and gametes are produced from both mating types. Flagellar agglutination allows mating types to form aggregates; after the cell wall is released the gametes fuse (plasmogamy and karyogamy) to generate the zygote, which matures into a dormant heavy-walled diploid zygospore. Haploid cells (two of each mating types) are generated via meiosis at the time of spore germination. (Hallmann, 2011).

1.2 The chloroplast of *C. reinhardtii* as an expression compartment

Table 1.1 shows a comparison of nuclear and chloroplast expression of transgenes. For instance, if recombinant proteins are secreted into the medium or if they require post-translational modifications, nuclear transformation is the most likely option.(Griesbeck et al., 2006). Transgene integration into the chloroplast genome is site-targeted, due mainly to the efficient process of homologous recombination taking place in the organelle. Approximately 40% of the cell volume is filled by the chloroplast, and its genome is a circular molecule around 200 kb (Figure 1.3). Each organelle has approximately 80 copies of the genome; therefore, stable transformation requires all the copies to contain the transgene (Rosales-Mendoza et al., 2012). Before detailing the process of chloroplast transformation, it is convenient to revise the mechanisms involved in native gene expression in the chloroplast. The following subsections will describe the chloroplast genome, how it is replicated, how its genes are transcribed into functional RNAs and finally translated into proteins with fundamental roles in the organelle's metabolism.

Table 1.1: Comparison of nuclear and chloroplast expression of transgenes. Modified from (Taunt, 2014)

	Nuclear	**Chloroplast**
Insertion	Random	Targeted
Gene Silencing	Transcriptional/Post-transcriptional	Not yet observed
DNA packaging	Large chromatin structures	Arranged in bacterial-like structures
Glycosylation	Similar to higher plants	None
Product location	Cytosol	Chloroplast
Product export	Chloroplast, mitochondria, endoplasmic reticulum, or secreted	None

1.2.1 The Chloroplast Genome

The earliest photosynthetic eukaryotes originated from the acquisition of a cyanobacterial endosymbiont, which eventually gave rise to glaucophytes, red algae, green algae and plants. The photosynthetic organelles of green algae and plants evolved from cyanobacterium as chloroplast, and exhibit many prokaryotic characteristics (Jensen and Leister, 2014). The chloroplast genome of *C. reinhardtii* has a circular arrangement and consists of around 200 kb containing approximately 99 genes of known or assumed function. There is a bias in nucleotides content towards A + T, which represent 65% of the total content. The structure of the plastome can be divided in four parts: two regions that are present as single-copies and two regions of inverted

repeats (Figure 1.3). The single copy regions differ in size; the larger having about 3 kb extra, which accordingly has been denoted as large single copy (LSC 81,307 bp).

Figure 1.3: ***Chlamydomonas reinhardtii*** **chloroplast genome**. The inner circle represents the open reading frames (ORFs) of undetermined function. The outer circle denotes the genes of recognized or apparent function. Modified from (Maul et al., 2002).

The other copy is obviously known as small single copy (SSC 78,088 bp). On the other hand, the inverted repeats are identical in size (22,211bp).

The single-copy sections hold the bulk of protein-coding and tRNA genes while the inverted repeats have a copy of *psbA*, some rRNA and tRNA genes (Higgs, 2009). Additionally, some genes have been recognized to be arranged in blocks. For instance, some photosystem II-coding genes (*psbL-psbF, psbB-psbT-psbN-psbH*) along with some ribosomal-coding genes (*rpl23-rpl12-rpl19, rpl16-rpl14-rps8*) are organized as gene clusters (Figure 1.3) (Higgs, 2009). Another particular feature of the *C. reinhardtii* plastome is the incidence of short dispersed repeats (SDRs), especially in intergenic regions (Maul et al., 2002). Their function is not yet solved but it is generally thought that they might have an impact on the genome evolution via homologous recombination. The plastome of *C. reinhardtii* displays another important peculiarity: it contains two copies of a 2.4 kb putative-transposable element named Wendy, which bear a 14-bp inverted repeat similar to those observed in transposons. Of the two copies (I - II), Wendy I contains two ORFs that are expressed at the RNA level, and one of them is predicted to have similarities with transposases and integrases (Higgs, 2009).

Despite having a relatively large plastome the number of genes in the *C. reinhardtii* organelle remains modest. From the overall 99 genes, 64 encode proteins, 30 for tRNAs, six for rRNAs, one for an RNA splicing sequence; also, eight open reading frames (ORFs) and one unassigned reading frame (URF) have been identified (Maul et al., 2002). Three main groups could be observed: genes involved in photosynthesis, gene expression (transcription, translation, protein processing). The third group encompasses the so-called hypothetical chloroplast open reading frames (*ycf*), which have been shown to be involved in various tasks such as photosystem stability, cytochrome b6f complex formation or chloroplast inorganic carbon uptake (Ravi et al., 2008). Chloroplast genes have typically a prokaryotic structure with a few containing group I and II introns. Introns from the former group have been found in *psbA*, or in the sequence encoding the *23S* rRNA. In contrast,

group II introns have been located in the *psaA* gene, which is made up of three exons that after being trans-spliced give rise to the mature mRNA (Nedelcu and Lee, 1998). Other genes with peculiar structure comprise those encoding the RNA polymerase subunits. These genes can be grouped into those being expressed as one unit, those consisting of segments, and those carrying non-consensus intron sequences (Higgs, 2009).

The plastome must replicate to maintain its biological inheritance. Currently, up to four origins of replication have been detected; two of them are known to be bidirectional and have been named *oriA* and *oriB*. The former is overlapping with the *rpl16* gene and the latter is found between the *petB* and the *rpl2* coding sequences. The remaining two are located between the *psbA* and the *rrnL* genes (Higgs, 2009). Two main protein complexes have been identified as having DNA polymerase activities. One of them has been observed to be composed of a major protein (75 kDa) along with two smaller ones (40 kDa, 15 kDa). Similarly, the other enzyme is a complex consisting of a protein of 116 kDa and one of 80 kDa. These two enzymes appear to be working at different stages of the life cycle (Umen and Goodenough, 2001; Sears, 1998). The plastome of *C. reinhardtii* presents, as described above, a wide range of peculiarities like the organization and dispersion of genes. These particularities are somehow reflected in the next stage of information processing: transcript production.

1.2.2 Transcription of chloroplast genes

Some chloroplast genes appear to be transcribed as single units or monocistrons, others, in contrast, have been found to be part operons with primary transcripts that more likely require further processing. Recent research has demonstrated that co-transcription is more common that previously thought, with around 70% of the chloroplast genes being arranged

in polycistronic units (Cavaiuolo et al., 2017). Due to its bacterial origins, the chloroplast transcription machinery is prokaryote-like and includes typical features such as eubacterial-type promoters and a DNA-dependent RNA polymerase with eubacterial-type subunits. Three different kinds of promoters have been recognized; the first is a characteristic sigma^{70}-type promoter that contains the -35 (TTGACA) and the -10 (TATAAT) elements present in bacterial promoters. Secondly, a variant of this type of promoter has been discovered in which the -35 sequence is missing and the -10 motif is extended (TATAATAT). The remaining class has not yet been properly scrutinized, but relates to a population of tRNA coding sequences (Klinkert et al., 2005; Stern and Drager, 1998). The only RNA polymerase that appears to be functioning in the *C. reinhardtii* chloroplast is the one encoded by the *rpo* genes in the plastid itself (known as plastid-encoded polymerase or PEP), which is a typical multisubunit eubacterial DNA-dependent RNA polymerase. The presence of just one polymerase differs drastically from what is seen in higher plants where an additional nucleus-encoded polymerase (NEP) is active in the organelle (Klein, 2009; Smith and Purton, 2002). For genes being expressed from sigma^{70}-type promoters (e.g. *rbcL*), transcription begins some base pairs downstream of the -10 motif (Anthonisen et al., 2001). One promoter belonging to this group, which drives the expression of the *23S rrn* gene, has been observed to be more powerful than the extended -10 type promoters that drives the expression of the *rbcL* and *psbA* genes (Klein, 2009). Bacterial systems utilize stem-loop structures, promoted by inverted repeats, to halt the polymerase and stimulate its release. In *C. reinhardtii*, these inverted motifs have been observed in the 3‘ untranslated region (UTR) of coding genes and are able to form RNA stem-loops; nevertheless, these structures do not always accomplish the termination of transcription. Even though the stem-loop structures seem to be ineffective in transcription termination, they do seem important in transcript maturation and need to be

employed for transgene expression. (Barnes et al., 2005; Monde et al., 2000; Stern and Drager, 1998)

As stated in the last section, some genes (*psbA, psaA* and *rrn*) possess introns that belong to either group I or group II. Group I introns, present in the *psbA* and *rrn* genes, appear to catalyze their excision by autocatalytic means, a process that requires the presence of a guanosine cofactor along with magnesium ions. Self-splicing of these types of introns is highly active *in vivo*. However, when tested *in vitro* the efficiency diminishes, suggesting the involvement of splicing factors (Stern et al., 2010). In contrast, group II introns, observed in the precursor transcript of the *psaA* mRNA, do require the presence of nuclear-encoded factors (RAA1-3 and RAT1-2) for excision. In this case, the exons are scattered in the genome but once transcribed they are assembled (along with a non-coding RNA, *tscA*) into a primary transcript that possesses two split type II introns, which require the trans-splicing factors for removal (Figure 1.4) (Herrin and Nickelsen, 2004).

Besides splicing events, chloroplast transcripts are processed at both 5‘ and 3‘ ends, and internal nucleolytic processing also occurs in polycistronic transcripts. In addition, there is a peculiar processing of the discontinuous *23S* rRNA. In this case, the *23S* pre-mRNA is divided by three sequences known as internal transcribed spacers (ITS) and one group I intron; after processing (mediated by nuclear-encoded factors), the ITSs and the intron are removed and four exons are released (*7S, 3S, 23Sγ, 23δ*) (Herrin and Nickelsen, 2004). The processing of both 5‘ and 3‘ ends of the pre-mRNA is mediated by the interaction of *cis* elements and nuclear encoded-factors. In both cases, the processing starts with an endonucleolytic cleavage followed by an exonucleolytic reaction. The exonuclease action is inhibited by the presence of a stability complex on the 5‘ end, and by the inverted repeats present at the 3‘ end. Similarly, nucleolytic processing has been determined in polycistronic mRNAs. Despite having not been thoroughly

Figure 1.4: Trans-splicing of *psaA*. The three exons of *psaA* are scattered in the plastome. Exon 1 and 2 are co-transcribed with *trnI* and *psbD*, respectively. The *tscA* (non-coding) RNA is co-transcribed together with *chlN*, whose processing is mediated by the nucleus-encoded factors (RAA1, RAT1 and RAT2). Assemble of the transcripts form two group II introns. Trans-splicing both introns involves the action of nuclear-encoded elements. (Stern et al., 2010).

investigated; it is assumed that the same machinery is involved in the processing of both mono- and polycistronic RNAs (Herrin and Nickelsen, 2004; Stern and Drager, 1998). The cell degrades RNA molecules for different purposes such as removing damaged RNA, recycling nucleotides, regulating gene expression or disposing of products of RNA processing. This activity encompasses the action of an endonuclease, which can start at the 5‘ end or other regions, followed by 5‘ to 3‘ exonuclease activity. On the other hand, if the mRNA does not possess the proper inverted repeats, characteristic of the 3‘ ends, degradation can start by means of a 3‘ to 5‘ exonucleolytic

activity (Herrin and Nickelsen, 2004). Two RNases have been recognized, the polynucleotide phosphorylase (PNP1) and the RNase II homolog RNB2 (Yehudai-Resheff et al., 2007; Nishimura et al., 2004).

Messenger RNA stability is carried out by both exogenous and endogenous factors, from the former, light and phosphate limitations have been seen to decrease or increase RNA stability, respectively. From the latter, particular sequences present in the untranslated regions are essential. The scrutiny of this phenomenon has yielded significant information; for instance, factors targeted from the nucleus have been determined as essential for stability of specific mRNAs. Such elements act on *cis* motifs, which have been located within some untranslated regions (Herrin and Nickelsen, 2004). The 3‘ end of the *atpB* gene illustrates this point. This sequence forms stem-loop structures that, as aforementioned, are involved in transcript processing but also protect it from further 3‘ to 5‘ exonucleolytic activity. Similarly, some 5‘ ends have been associated with mRNA stability. For example, in the *rbcL* 5‘ UTR a small sequence of -10 nt has been proved fundamental for mRNA accumulation. Another instance is the *psbD* 5‘ UTR which has two regions related to the production and abundance of the mature mRNA (Stern et al., 2010; Herrin and Nickelsen, 2004). These *cis* sequences mediate the action of factors imported from the cytoplasm. For example, a nuclear-encoded protein (MBD1) regulates the stability of the *psbD* mRNA. Likewise, a protein named MBB1 has been involved in *psbB* and *psbH* transcript stabilization. Both proteins bear particular sequences known as tetratricopeptide repeats, which have been identified as critical in mRNA processing, stability and translation in plants (Boudreau et al., 2000; Vaistij et al., 2000). Recently, two mRNA stabilization factors, which are members of the OctotricoPeptide Repeat (OPR) family, have been characterized. MCG1 and MBI1 are required to stabilize the mRNAs encoding a subunit of the cytochrome b_6f complex (*petG*) and a subunit of the photosystem II (*psbI*), respectively.

(Wang et al., 2015).

1.2.3 Protein production in the chloroplast

Despite its cyanobacterial origins, the chloroplast has not retained full prokaryotic features for mRNA translation. Instead, it has acquired distinctive peculiarities that have emerged during the organelle's evolution. Ribosomes have two subunits, composed of rRNA and complementary proteins, which have been classified based on their sedimentation coefficient (S). The small subunit (30S in prokaryotes and 40S in eukaryotes) comprises a unique rRNA and between 21 - 33 proteins whilst the large subunit (50S in prokaryotes and 60S in eukaryotes) consists of a triad of rRNAs and around 33 - 47 proteins. Each subunit can interact with tRNA via three binding sites. The aminoacyl (A) site is in charge of docking the acetylated tRNA, whereas the peptidyl (P) site associates the nascent peptide with the charged amino acid. Finally, the exit (E) site allows the deacylated tRNA to exit the complex. Once the machinery encounters the stop codon, translation is terminated and the final protein is detached from the complex.

C. reinhardtii chloroplast ribosomes, besides having proteins similar to those present in bacteria, possess plastid specific ribosomal proteins (PSRP), which have been seen to interact with the chloroplast-unique ribosomal proteins and RNA elements to regulate mRNA translation in the organelle (Manuell et al., 2007b). The 70S plastid ribosome has counterparts for some proteins of the 50S and the 30S subunits of *E. coli* and plants. However, PSRP have been identified in association with the LSU and the SSU. Interestingly, all these proteins appear to be encoded by the nuclear genome and imported via transit peptides. The final 70S complex carries also two additional proteins RAP38 and RAP41, which have homologs with two RNA-binding proteins from spinach, thus they might be involved in mRNA

processing or degradation (Zerges and Hauser, 2009; Marín-Navarro et al., 2007). One similarity to eubacteria is the operon-fashion organization of the rRNA genes in the plastome (*16S-tRNAIle-tRNAAla-23S-5S*) with post-transcriptional processing yielding the mature rRNA and tRNA molecules. In particular, the 16S rRNA (part of the SSU) has been directly linked with codon - anticodon pairing, hence controlling translational efficiency. This molecule is the target of various antibiotics, used as selectable agents, such as streptomycin or spectinomycin. Other antibiotics (e.g. chloramphenicol or erythromycin) target the 23S rRNA that makes part of the 50S subunit. Modifications of the targeted sequences, alteration of the ribosomal proteins, or protein-mediated inactivation of the antibiotic are the means whereby the cell generates resistance (Zerges and Hauser, 2009). General translation factors, mainly GTPases, are needed for proper initiation, elongation and termination. *C. reinhardtii* possesses homologs to eleven GTF present in *E. coli*, being predominantly encoded in the nucleus with the exception of a plastid-encoded factor (EF-Tu) involved in proper codon-anticodon matching (Manuell et al., 2004). In prokaryotic systems, translation is initiated via the interaction of 16S rRNA and the so-called Shine-Dalgarno (SD) sequences in the 5‘ UTR of the mRNA. In contrast, cytosolic translation can proceed in various ways, but generally it requires the interaction of some factors with the mRNA and the rRNA. The mechanism of ribosome positioning in the chloroplast has not yet been completely elucidated. It is certain that *cis* sequences are needed as anchor for the ribosomes; also, other factors, like RNA-binding proteins, could interact with the ribosomal subunits and alter the structure of the mRNA that promotes its positioning on the small subunit of the ribosome (Marín-Navarro et al., 2007). Various factors have been found to regulate translation by interacting mainly with the 5‘ UTR of the mRNA. The proteins (D1 and D2) that make up the reaction centre of the PSII, for instance, are translationally upregulated

by light. In this case, translation initiation appears facilitated by factors imported from the cytoplasm. Light, by altering the redox status via photosynthesis, recruits a group of proteins (RB38-47-55-60) which allow the *psbA* mRNA (encoding the D1 protein) to be loaded into the ribosome for translation. Analogously, translation of the D2 protein (encoded by the *psbD* gene) is activated by *trans* factors. Nac2 is a protein that besides providing protection from exonucleolytic degradation (by binding to the 5‘ UTR) recruits another factor that facilitates translation initiation (Zerges and Hauser, 2009; Marín-Navarro et al., 2007). Similar cases observed in other genes are summarized in (Table 1.2). In general, proper cooperation between the nuclear and chloroplast genomes, involving multiple signals and various steps, is needed for protein synthesis and, thus, for the metabolic processes that take place in the organelle.

Table 1.2: Trans factors, acting on 5‘ UTRs, involved in translation Modified from (Marín-Navarro et al., 2007)

Protein	**Gene**	**Factor**
D1	*psbA*	F35, Tba1, RB47-38-60-55
D2	*psbD*	Nac1, Acc-115, Nac2, RBP40
P6	*psbC*	TBC1-3
CP47	*psbB*	Mbb1
PsaB	*psaB*	Tab1-2
Cytf	*petA*	Sim30, Mca1, Tca1
SUIV	*petD*	Sim30, Mcd1
α subunit	*atpA*	F54

1.3 Chloroplast transformation and expression of foreign proteins

Various methods for successful chloroplast transformation in *C. reinhardtii* have been developed. Particle bombardment with DNA-coated tungsten or gold particles has been used to deliver DNA to the plastome. Electroporation has been to increase the permeability of the membrane, and allow the entrance of DNA into the chloroplast (Mussgnug, 2015; Potvin and Zhang, 2010). Additionally, simpler methods have proved successful for transformation. Agitation, using a normal laboratory vortex, of cell-wall deficient strains with glass beads and the DNA of interest has demonstrated to yield successful transformants. It is thought that the glass beads create holes in both the plasma and chloroplast membranes, allowing DNA to access the organelle's genome (Purton et al., 2013; Specht et al., 2010). Early attempts of transgene expression in the chloroplast of *C. reinhardtii* were not entirely effective; this was related to the A + T rich genome of the chloroplast. Because of that bias, certain codons are used more commonly than others. The arrival of artificial gene synthesis solved the issue, and currently *de novo* optimized genes can be used for optimal expression (Mayfield et al., 2007). Furthermore, the development of selectable markers was an essential step in the field. Knocking out sequences, and replacing them with traceable markers allowed the investigation of gene function. Also, they permitted a way to select for the insertion of foreign sequences. An example is illustrated in Figure 3.3. Among the more common markers are those coding for antibiotic resistance, fluorescent and luminescent proteins. Several endogenous genes are utilized for this purpose as well. Point mutations in particular genes confer resistance to some antibiotics (e.g. streptomycin, erythromycin). Likewise, photosynthetic genes have been used to rescue mutants to photoautotrophy (Economou et al., 2014) (Table 1.5).

Some recombinant proteins might require post-translational modifications; in this respect, chloroplasts lack the proper apparatus to post-translationally modify proteins (e.g. glycosylation). Hence, the production of properly processed mammalian proteins (e.g. antibodies) seems limited. Nevertheless, producing these sorts of proteins in the chloroplast will reveal important insights into non-glycosylated protein activity (Mayfield et al., 2007).

Equally, disulphide-bond formation denotes another important characteristic of eukaryotic proteins. Various attempts, for instance, have successfully shown the expression of mammalian proteins, which carry disulphide bonds, in the chloroplast of *C. reinhardtii*. For instance, the human IgG1 monoclonal antibody, the human glutamic acid decarboxylase and the human erythropoietin have been detected, suggesting proper folding of the proteins (Eichler-Stahlberg et al., 2009; Tran et al., 2009; Wang et al., 2008a). Finally, it is worth noting that the proof-reading mechanism, which check and remove anomalous proteins, might represent a limitation for expressing recombinant peptides (Rosales-Mendoza et al., 2012).

As research went on, several factors were determined as fundamental for improving recombinant protein production. Pairing sequence design with chloroplast codon usage, choosing proper promoters and UTR's have permitted the achievement of stable transplastomic expression. By mimicking the sequence AT-GC distribution, an improved expression of various genes like reporter proteins (e.g. luciferase), and human monoclonal antibodies has been achieved (Mayfield et al., 2007; Mayfield and Schultz, 2004; Mayfield et al., 2003). As indicated in previous sections, the flow of information from the chloroplast genome to its encoded proteome is subjected to several processes of regulation, which involve the interaction between *cis* and *trans* factors. The efficacy of transcription, stability of the messenger RNA, and efficiency of translation affect chloroplast gene expression. Promoters and

UTRs play key roles in maintaining the proper stream of information. Therefore, endogenous promoters, of highly expressed genes, have been chosen to drive transgene expression. Among the most used promoters are the ones driving the expression of the ATPase alpha subunit (*atpA*), the protein D1 (*psbA*) and the protein D2 (*psbD*) of the photosystem II. Also, promoters responsible for the expression of the large subunit of the ribulose biphosphate carboxylase (*rbcL*), and the P700 chloroplhyll a apoprotein A1 (*psaA exon 1*) have been widely used (Purton et al., 2013; Rasala and Mayfield, 2011). The use of these promoters is accompanied by the presence of their native 5‘ UTR. These structures have proved fundamental for significant transgene expression as they interact with *trans*-acting factors targeted from the nucleus. It has been reported that combining promoters and 5‘ UTRs from different sources does not generally improve expression, just a few combinations have been successful. Additionally, the utilization of 3‘ UTRs has not been regarded as fundamental (Barnes et al., 2005). The control of chloroplast gene expression is mainly post-transcriptional. mRNA processing along with its stabilization and translation are key steps in the expression of transgenes. Transgene expression in plastids can be phenotypically deleterious. Possible harmful interactions may arise when foreign proteins interact with endogenous components. Likewise, native important metabolic processes might be affected by the enzymatic activities of recombinant enzymes, or just the fact of having an extra metabolic burden might result in altered phenotypes (Bock, 2014). A plausible option to deal with this limitation is to develop an inducible transgene expression. One way to accomplish this is by controlling plastid translation using factors encoded in the nucleus. Nac2 illustrates this notion; this factor regulates the translation of any foreign gene fused to the *psbD* 5‘ UTR. Via nuclear engineering, a copper-regulated promoter could control the expression of this RNA-binding factor. Thus, the presence or absence of copper in the medium regulates indirectly the translation process

of the gene controlled by such 5‘ UTR (Figure 1.5A) (Surzycki et al., 2007). Similarly, this Nac2 protein has been used for vitamin-regulated expression of recombinant proteins from the plastome (Ramundo and Rochaix, 2015).

Another alternative relies on the *E. coli* lac repressor / lac operator system. Here, the promoter region, that is controlling the gene of interest, has been altered so it contains the bacterial operator sequences. In this way, the constitutive expression of the repressor gene blocks the processing of heterologous sequences. This inhibition can be relieved if an inducer (IPTG) is added (Figure 1.5B) (Kato et al., 2007). A different strategy involves the utilization of naturally occurring or synthetic riboswitches. Transcriptional riboswitches, in the presence of a ligand, allow the formation of a hairpin (terminator) within the mRNA that leads to the premature termination of transcription. On the other hand, translational riboswitches do not allow translation by preventing the ribosome to access the mRNA, which depends on the presence or absence of a particular compound (Figure 1.5C-D). Recently, a synthetic riboswitch was successfully introduced into the tobacco chloroplast genome, and shown to respond to theophylline (Verhounig et al., 2010). In the *C. reinhardtii* nuclear genome, short sequences have been identified as TPP (thiamine pyrophosphate) riboswitches, which appear to be involved in processing the transcripts for the genes encoding early enzymes of thiamine biosynthesis (Croft et al., 2007). Nonetheless, a functional synthetic riboswitch for *C. reinhardtii* is yet to be engineered. This is not an easy task due to the complex interplay between mRNAs and translation factors coming from the nucleus. Nevertheless, there are still other plausible options that can be assayed in *C. reinhardtii* for well-regulated protein expression. Osmotically-regulated transcriptional activation involves the use of native promoters like the proU promoter. Similarly, cold-shock activation may be accomplished by utilizing promoters derived from bacterial cold-shock genes (Rosales-Mendoza et al., 2012). The main advantage

of using inducible rather than constitutive promoters is that they would permit the expression of recombinant proteins once high cell densities have been reached. Moreover, these inducible systems represent a conceivable methodology for controlling protein expression levels when necessary. Especially if heterologous proteins have physiological influence or when they are highly toxic for the host. Therefore, optimizing these methods is fundamental if we want to construct synthetic metabolic systems that replicate the correct chronological genetic expression and the correct protein balance of important naturally-occurring metabolic pathways.

Generally, the use of these elements has helped express successfully some commercially-relevant bio-molecules, which demonstrates the practicality of this green microalga as a platform for genetic engineering. Table 1.3 summarizes some relevant examples. *C. reinhardtii* has also been used for metabolic engineering purposes, by manipulating the chloroplast genome, the expression of different important bio-molecules for biofuels, food and feed industry has been attained. For instance, (Tan et al., 2007) have successfully expressed a carotenoid gene (*crtR-B*) from *Haematococcus pluvialis* in the chloroplast. This gene encodes a β-carotene hydroxylase necessary to process the production of zeaxanthin from β-carotene and astaxanthin from canthaxanthin. Thus, a higher quantity of carotenoids was observed in the transformants when compared to the wild type. Another research has reported the expression of *de novo* synthesized genes (ferrochelatase (*hemH*) from *Bradyrhizobium japonicum*, and leghemoglobin (*Iba*) from *Glycine max*) using the chloroplast. In this case, the production of biohydrogen was higher in transformants compared to the control cultures. Similarly, the rate of respiration was higher than in the wild-type cultures (Wu et al., 2011). In summary, *C. reinhardtii* has significant potential as a protein production platform. Additionally, genetically engineering the plastome is routine, but metabolically engineering the organelle is still far from being effec-

Figure 1.5: Some common methods for trans-gene expression regulation in the chloroplast. A. Copper-regulated expression of RNA-binding factor (Nac2) that allows the expression of the gene of interest within the organelle. B. *E. coli* lac repressor/lac operator system for heterologous expression. Constitutive expression of the repressor gene blocks the processing of heterologous sequences. This inhibition can be relieved if an inducer (IPTG) is added. C. Transcriptional riboswitch. The ligand (e.g. vitamins) interacts with the sensing domain in the RNA to form a structure that triggers the development of a hairpin that terminates transcription prematurely; on the other hand, if there is no metabolite present an antiterminator is formed and proper elongation of the transcript takes place. D. Translational riboswitches. In the absence of a metabolite (e.g. theophylline) an aptamer structure is formed which conceals the ribosome-binding site (RBS). Contrarily, the aptamer structure is dissolved if the metabolite is present, allowing thus the binding of the ribosome and normal translation to take place. Modified from (Nakahira et al., 2013; Purton et al., 2013; Serganov and Patel, 2012).

tively achieved. Lastly, the trophic plasticity of this organism along with its benefits for transplastomic engineering render possible the development of conceivable strategies for altering native metabolic pathways or integrating foreign ones.

Table 1.3: Some commercially-relevant molecules produced in the chloroplast of *C. reinhardtii*. Modified from (Almaraz-Delgado et al., 2014)

Recombinant protein	Description
hGH	Human growth hormone (Wannathong et al., 2016)
VP1CTB	VP1 protein from mouth and disease virus (Sun et al., 2003)
TRAIL	Ligand related to tumor necrosis factor that induces apoptosis (Yang et al., 2006)
M-SAA	Mammary-associated serum amyloid (Manuell et al., 2007a)
hGAD65	Human glutamic decarboxylase (Wang et al., 2008b)
IBDV-VP2	VP2 protein from the infectous burial disease virus (Surzycki et al., 2009)
IPNV-VP2	VP2 protein from the infectous pancreatic necrosis virus (Surzycki et al., 2009)
PCV2	Porcine circovirus type 2 (Surzycki et al., 2009)
HC-83K7C	Heavy chain human monoclonal antibody against anthrax antigen 83 (Tran et al., 2009)
LC-83K7C	Light chain human monoclonal antibody against anthrax antigen 83 (Tran et al., 2009)
VEGF	Human vascular endothelial growth factor (Rasala et al., 2010)
acrV2, vapA2	Antigen from fish pathogen *Aereomonas salminicida* (Michelet et al., 2011)

1.4 Metabolic engineering of the chloroplast

One of the main bottlenecks of metabolically engineering the chloroplast of *C. reinhardtii* is the insertion and simultaneous expression of multiple genes. Despite some encouraging results there is no standardized approach of introducing groups of genes into the plastome (Rasala et al., 2014). In contrast, in plant-based chloroplast expression systems some techniques have been advanced for this purpose. Approaches like co-transformation, where more than one transformation vector is used, and serial transformation, which involves successive transformation events, have proved useful (Bock, 2014). The availability of techniques allowing simultaneous transgene expression has permitted researchers to remodel or enhance the production of important metabolites in plants models. For instance, (Kumar et al., 2012) reported increased levels of the products of the isoprenoid pathway in the chloroplast of tobacco; they successfully introduced a cluster consisting of six genes into the plastid genome. Likewise, the tocochromanol pathway (comprising three genes) was redesigned and effectively expressed in the leaves and fruits of *A. thaliana* through the engineering of the plastid genome (Lu et al., 2013). Likewise, the expression of the natural pigment astaxanthin (that is not synthesized in higher plants) in tobacco has been achieved (Hasunuma et al., 2008). Plastid transformation was achieved by introducing the genes encoding CrtW (β-carotene ketolase) and CrtZ (β-carotene hydroxylase) from a bacterium of the genus *Brevundimonas*. They reported, besides a higher accumulation of carotenoids, the expression of a novel carotenoid known as 4-ketoantheraxanthin. Plastid transformation has also been chosen to bring the production of polyhydrobutyric acid in tobacco. Thus, three genes (*phbC-phbB-phbA*) were placed under the control of an ethanol-inducible T7 RNA polymerase, which was targeted to the organelle. Therefore, after ethanol treatment, PHB synthesis was seen

to be stimulated in tobacco (Lössl et al., 2005). Furthermore, important pharmaceutical molecules have been synthesized through plastid engineering. Artemisinin, in particular, has proven to be an effective treatment for malaria. Twelve enzymes required to produce artemisinic acid (precursor of the antimalarial drug) in the tobacco chloroplast have been expressed (Saxena et al., 2014). In this research, an exogenous cytosolic mevalonate pathway was introduced along with the artemisinic acid pathway in order to create the flux of carbon required for the precursor production. Lipid biosynthesis has also been an important target for plastid engineering. It has been demonstrated that an overexpression of a plastid gene (*accD*), encoding the β-carboxyl transferase subunit of acetyl-CoA carboxylase, increases lipid accumulation. In addition, the plastid expression of a desaturase gene, either from wild potato or a cyanobacterium (*Anacystis nidulans*), gave rise to an altered fatty acid profile in tobacco (Maliga and Bock, 2011). These outcomes demonstrate that the chloroplast can actually be metabolically altered. Nevertheless, the knowledge of metabolic and genetic networks along with the repertoire of tools available in plant platforms contrasts sharply with the scenario present in *C. reinhardtii*. However, the methodologies used for accomplishing such results in plants can positively influence the status of metabolic engineering this green alga. As discussed, there are no available methods to couple a foreign metabolic pathway into the chloroplast of *C. reinhardtii*. However, it is valuable to investigate possible pathways that could be brought into the organelle. The bacterial nitrogenase is responsible for turning N_2 into NH_3, and render it biologically available. The assembly of this enzyme requires the coordinated expression of various genes. Nonetheless, minimal sets of genes have been proved optimal for expressing this enzyme in *E. coli* (Wang et al., 2013a,b; Temme et al., 2012). In this research, I propose the chloroplast a potential niche for biological nitrogen fixation in eukaryotic algae, the advantages of the organelle are discussed

and a minimal set of genes is proposed. Key genes from the aforementioned set were cloned individually into the plastome. The results are described in chapter 4.

The analysis of the aforementioned results confirms that metabolic engineering requires rational designing using a certain number of discrete elements (e.g. promoters, UTRs, GOIs) in order to obtain predictable outcomes. The arrival of what is known as synthetic biology has influenced this field substantially by introducing engineering approaches to address biological questions, it promises considerable advantages for industrial biotechnology. Thus, the next section surveys the field of synthetic biology, its impact on some well-established expression platforms, and how this approach could be employed for boosting metabolic engineering in the chloroplast of *C. reinhardtii*.

1.5 Synthetic biology and its contribution to metabolic engineering

In the early stages of industrial biotechnology, the production of chemicals was limited to those fabricated naturally by selected microorganisms. However, the advent of genetic manipulation opened up the possibility of obtaining heterologous products that are not produced in nature. Those products were obtained through manipulation of single genes. Genetic engineering allowed the manipulation of particular genes in natural producers, boosting the production of molecules with relevant applications (Scaife et al., 2015; Boyle and Silver, 2012; Lynch and Gill, 2012). Metabolic engineering requires the understanding of metabolic networks and the components essential for metabolite production. The capacity to manipulate metabolic fluxes has been influenced positively by the various devices developed in the field

of synthetic biology. Synthetic biology attempts to borrow principles from electronics to approach biological issues with the goal of modulating native metabolic networks or bring foreign ones. To achieve this, the transfer of biological information into methodologies embracing engineering principles was imperative. This led to the proposal of fundamental principles for engineering biological systems (Scaife et al., 2015; Boyle and Silver, 2012; Lynch and Gill, 2012). These principles make use of traditional engineering notions and some of the most relevant terms have been summarized (Scaife et al., 2015) Table 1.4.

E. coli is the most characterized expression platform nowadays, and has been widely used for developing cutting-edge synthetic biology in bacterial systems. Examples of these technologies are biosensors, which could be designed so gene expression is regulated in response to inputs such as different light wavelengths (Olson et al., 2014), metabolites (e.g. acrylate, glutarate, erythromycin) (Rogers et al., 2015) or toxic compounds (e.g. TNT) (Tan et al., 2015). Those systems make use of standardized genetic sequences that are responsive to environmental cues like environment-responsive promoters, anti-switches or sensitive elements (Khalil and Collins, 2010), which have been applied to produce a variety of industrially relevant bio-molecules such as bio-fuels and antimicrobials (Seo et al., 2013; Keasling, 2012). In the same way, *Saccharomyces cerevisae* represents a well-studied eukaryotic expression platform. The appliance of synthetic biology principles into various aspects of yeast research has, similar to the case of bacteria, increased the availability of genetic tools. Modulation of transcription, for instance, has been achieved by rationally designing transcription factors that render the expression of genes of interest tunable. Likewise, riboswitches and synthetic RNA interference (RNAi) have been utilized to post-transcriptionally regulate gene expression. Moreover, technologies using designed small protein motifs, which recognize DNA sequences, linked with nuclease domains have

Table 1.4: Relevant terms used in synthetic biology. Modified from (Scaife et al., 2015)

Abstraction	Standardization of individual parts allows predictable outcomes when combined into devices. The devices could be combined to increase the level of complexity
Boolean logic	Engineering concept where the values of the variables are binary: 1 or 0, on or off, true or false. This approach allows the designing of logical operations where the output depends on specific inputs
Chassis	Framework where the standardized parts could be assembled. Host organisms where genetic constructs could be straightforwardly inserted
Device	DNA construct bearing required parts for transgene expression
In-silico modelling	The appliance of genetic and metabolic information into experimental design to allow hypothesis-driven research
Orthogonality	The ability of standardized parts to function independently of the cellular platform or context
Parts	Genetic sequences encoding biological functions (e.g. promoters, UTR's, GOI)
Standardization	Development of methods to characterize and validate biological parts or processes. E.g. assembly protocols, selection methods, reporter assays

been utilized for genome editing (Jensen and Keasling, 2015). The arrival of these technologies has allowed the production of key molecules like antimicrobials (e.g. artemisin) (Paddon et al., 2013; Westfall et al., 2012) or potential bio-fuels (isobutanol) (Avalos et al., 2013).

It has been already suggested that synthetic biology could help tackle fundamental bottlenecks in plastid multigene engineering, and furthermore contribute towards developing a minimal synthetic plastome (Scharff and Bock, 2014; Agapakis et al., 2012; Georgianna and Mayfield, 2012). The final portion of the last section sought to illustrate the advances of metabolite engineering in plant platforms by listing some relevant results. As noted before, this panorama contrasts sharply with the one prevailing in *C. reinhardtii*, where the limitations seem to be, among others, the lack of knowledge regarding genetic, cellular and metabolic regulatory networks along with the limited variety of standardized tools to control multiple gene expression. It has been proposed that applying some of the leading concepts of synthetic biology, developed in well-established expression platforms (e.g. *E. coli*, *S. cerevisae*), could facilitate the development of *C. reinhardtii* as an industrial biotechnological platform (Scaife et al., 2015). Hence, it is worth revising the terms listed in Table 1.4 in the context of metabolically engineering the chloroplast of this green alga.

The concept of abstraction requires establishing hierarchies of the respective standardized devices, used for gene expression, to increase the level of complexity of synthetic biological systems. Standardization involves determining genetic sequences encoding biological functions, which could be utilized to render gene expression predictable. Table 1.5 summarizes some of the most relevant genetic functional parts available for engineering the chloroplast, which are used to integrate and express genes into the plastome (Figure 3.3). As could be appreciated, the insertion and expression of transgenes is straightforward. Nonetheless, increasing complexity would

require orchestrating the expression of various foreign genes simultaneously. To achieve this goal, it will be necessary to develop devices that combine various expression cassettes or tools that allow the expression of various genes in an operon-fashion.

The operational machinery of computing devices relies on Boolean logic gates. This notion has been imported into the field of synthetic biology with the goal of creating biomolecule-based computers. In fact, selected logic gates (e.g. AND, NAND) have been used to create biologically based digital devices (Miyamoto et al., 2012; Wang et al., 2011), most of this work has been done in bacterial systems. Unfortunately, the existing technology for the *C. reinhardtii* chloroplast does not meet the criteria to accomplish such sophisticated outcomes. Nevertheless, some authors have advanced some interesting proposals. For example, a hypothetical OR logic gate, whereby the presence of two inputs (light, nitrogen depletion) determine the production of the desire outcome that could be utilized to increased triacylglyceride (TAG) production has been proposed (Scaife et al., 2015). In any case, there is still a long way to go to have the catalog of genetic tools required to put these concepts in practice using the chloroplast of *C. reinhardtii*.

A biological scaffold that could be used as a platform for introducing genetic parts that allow predictable gene expression has come to be known as a chassis. As has been exemplified in previous sections, *C. reinhardtii* has been widely used for genetic, and to a lesser extent, for metabolic engineering, which proves that this green alga is a reliable microbial platform that fits for the application of synthetic biology approaches. In particular, as previously discussed the chloroplast represents the biosynthetic centre of the cell and thus harbors many metabolic reactions that are appealing for genetic programming purposes. Furthermore, the ease of inserting foreign sequences in the plastome is considered an advantage if compared with the

Table 1.5: Genetic sequences encoding biological function available for the chloroplast of *C. reinhardtii*. Modified from (Scaife et al., 2015)

Gene expression:	
Promoters / 5‘UTRs	*atpA, psaA exon 1, psbD, petB, chlL*
3‘ UTRs	*rbcL, atpB, psbA*
Selection markers:	
rrnS, *rrnL* point mutations	Resistance to spectomycin and erythromycin
aadA	Resistance to spectinomycin
aph6	Resistance to kanamycin
psbH	Photosynthetic restoration
GFP	Fluorescent protein
gusA	β-Glucoronidase activity
codA	5-fluorocytosine sensitivity
Targeting sequences (for homologous recombination):	
patpint-cg11(atpB-int)	*atpB* 3‘UTR inverted repeat
pLM7	*psbA* and 5S/23S
p322	*psbA* 16S rRNA
p71	*tscA* and inverted repeats
p72B	*psbH - psbN*

difficulties involving, for instance, nuclear transformation. Moreover, the standardized protocols for cultivation and harvest make this alga a good example of a biological chassis. Nevertheless, the convoluted interaction between *trans* and *cis* factors makes the control of transgene expression complicated, which renders the appliance of biological circuits or systems unlikely to be accomplished in the near future.

The way transgenes are delivered to the host genome is via a vector or device, which is commonly a plasmid that contains all the required information for gene insertion and expression. For chloroplast transformation, the devices must have the sequences required for homologous recombination, the gene of interest expression cassette (promoter 5‘ UTR / 3‘ UTR) and the selectable marker (Figure 3.3); some of the most utilized are listed in Table 1.5. Nonetheless, these vectors allow the insertion of a single gene at a time. Hence, it is imperative to develop devices that could allow the insertion and expression of multiple foreign sequences in a straightforward way, which seems to be vital if the level of complexity is to be increased. Relevant information could be borrowed from synthetic biology applications in other platforms like *E. coli* or *S. cerevisiae*. For instance, polycistronic expression cassettes whereby genes are separated by Shine-Dalgarno sequences have been used to produce complex heterologous enzymes in bacteria. Likewise, sequences needed for the entry of the ribosome have been utilized to arrange eukaryotic transgene expression under the control of one promoter (Bieniossek et al., 2012).

The most utilized reporter genes are listed in Table 1.5. The *gfp* reporter gene has been expressed with limited results in the chloroplast (Mayfield and Schultz, 2004); similarly, a recent study demonstrated the expression of a vivid Verde fluorescent protein, but the levels of fluorescence were low (Braun-Galleani et al., 2015). There are no current examples of cyan fluorescent proteins expressed in the organelle; however, a turquoise protein

(mTurq2cp) has been successfully expressed from the plastome of the basal plant *Marchantia polymorpha* (Boehm et al., 2015), which opens the possibility for its study in *C. reinhardtii*. Chapter 5 describes the expression of the *mturq2cp* gene from the plastome.

Orthogonality aims to circumvent network cross talk between foreign and native pathways by providing related yet independent elements for synthetic networks. The mobile circuitry present in phages, transposons or integrons have evolved to be functional regardless of the biological context of the host, and could be cited as a natural example of orthogonality. Some efforts have tried to rationally construct synthetic biological parts that are able to function independently of the context. For instance, an orthogonal ribosome-mRNA pair has been devised to control mRNA endogenous translation in bacteria (Rackham and Chin, 2005). Similarly, tunable transcription using synthetic promoters in *S. cerevisiae* has been accomplished by rationally designing the factors involved in activating transcription (Blount et al., 2012). These results demonstrate that orthogonality is attainable to some extent. However, the prospects for the chloroplast of *C. reinhardtii* do not seem favorable. This is mainly related, once more, to the intricate interaction between *cis* and *trans* factors involved in gene expression. As detailed in section 1.2, some elements encoded in the nucleus are required for proper mRNA processing and stability. Additionally, translation factors are targeted to the organelle to act on particular 5‘ UTRs and promote proper protein production. Therefore, these *cis* sequences depend on nuclear information to carry on their duties. Namely, if native promoters / 5‘ UTRs or 3‘ UTRs from the chloroplast of *C. reinhardtii* were to be used in other platforms, they could not work because they seem to require the actions of nuclear elements in order to drive heterologous protein production. Thus, without the presence of the proper *trans* information it seems unlikely that *cis* elements could drive transgene expression. Likewise, using foreign *cis*

motifs, to drive transgene expression from the plastome of *C. reinhardtii*, would not be entirely compatible with the *trans* machinery needed for expression in the organelle. Hence, substantial work must be done if orthogonal sequences are to be developed for chloroplast engineering. Nonetheless, the replacement of tryptophan codons with a stop codon (UGA) from *C. reinhardtii* prevented transgene expression in *E. coli* demonstrating that elements from the plastome could be functional in other platforms (Young and Purton, 2015).

As thoroughly discussed, many biological elements or parts (5‘ UTRs, 3‘ UTRs) are well characterized, and have been widely used for genetic and metabolic engineering purposes. Table 1.5 lists some of the most relevant. However, the next level of characterization requires to meticulously deduce the motifs on which *trans* factors act (Cavaiuolo et al., 2017), so these sequences could be, for instance, used to develop tool for transgene expression control. In addition, determining possible native sequences that could be used to process synthetic polycistronic mRNAs might help tackle the issue of simultaneous transgene expression. Besides characterizing the elements required for gene expression (promoters, 5‘ UTRs, 3‘ UTRs), it is fundamental to generate detailed knowledge of possible genes of interest for metabolic engineering. Namely, genes involved in commercially relevant metabolite production should be cloned and methodically studied in the chloroplast of *C. reinhardtii*. Although current methods do not allow to clone genes of selected metabolic pathways jointly, their characterisation as individual transgenes will prove useful when future technological efforts allow the synchronized expression of various transgenes in the organelle.

Lastly, the concept of standardization represents a fundamental component for the predictability and reliability of a synthetic biological system. This is because standardized components, whether a biological part, a device or a chassis, are needed for predictable outcomes. Biological parts need

to physically interact to give out a functional desired output. The assembly of a chloroplast transformation device illustrates this point. As could be appreciated in (Figure 3.3) the standardized biological parts are required to be set up in a particular order, sequences for homologous recombination, for instance, must flank the selectable marker as well as the expression cassette. Obviously, the promoter must be located upstream the start codon of the transgene, which must be flanked also downstream by a 3‘ UTR. In the analyzed example, photosynthetic restoration is used as a selectable marker; Thus, a correct sequence of the *psbH* gene (associated to the photosystem II, PSII) must be located downstream of the expression cassette. This physical set up allows predictable restoration of photoautotrophy and functional integration of a foreign coding sequence. Moreover, devices could be combined to increase complexity. For example, using various vectors to introduce diverse traits in a single strain might be useful for metabolic engineering purposes. In addition, the concept of standardization could be applied to other diverse contexts of algal engineering like selection methods or reporter assays. In any case, the concept of standardization is fundamental for the building of biological systems.

1.6 Objectives

It is clear that there is a crucial necessity to develop novel tools and approaches that contribute to increase complexity of genetic engineering the plastome of *C. reinhardtii*. The benefits of applying synthetic biology perspectives have been proved fundamental for developing genetic engineering in other platforms. Likewise, employing these approaches will definitively permit the development of sophisticated molecular tools for *C. reinhardtii*. As pointed out, the requirement of approaches and tools that permit simultaneous expression of recombinant proteins is imperative for metabolic engineering. However, there are no defined methodologies for such task. Additionally, it will prove crucial to study a group of genes belonging to a foreign metabolic pathway that could be engineered in the organelle. Evidently, the potential metabolite should have an important impact for algae biotechnology. Finally, expanding the palette of fluorescent proteins, expressed from the plastome, will favor chloroplast microalgal research. Production of fluorescent proteins in the chloroplast of *C. reinhardtii* has been limited to green fluorescent proteins. Therefore, evaluating novel fluorescent proteins with different range of emissions will expand the visual spectrum in the chloroplast of this green alga. To address these issues, the following objectives have been formulated for the present research:

1. To evaluate one or more approaches for simultaneously expressing various transgenes from the plastome of *C. reinhardtii*, demonstrating that such a task is attainable and providing a framework for developing a modular strategy. The results are presented in chapter 3.

2. To investigate a group of fundamental genes belonging to a foreign metabolic pathway that might have important repercussions for algae biotechnology. Biological nitrogen fixation might provide a novel alternative for microalgal nitrogen inputs. The research focused on defining a minimal

set of genes for nitrogenase, and their expression in the organelle. The results are presented in chapter 4.

3. To express a cyan fluorescent protein in the chloroplast of *C. reinhardtii*, and evaluate its performance as a reporter gene. The results are presented in chapter 5.

Chapter 2

Materials and Methods

2.1 Strains and culture conditions

2.1.1 *Chlamydomonas reinhardtii*

Two recipient cell wall deficient strains were used: cw15(mt+), and TN72, which is a mutant of the cw15 (mt+) line that has the chloroplast gene *psbH* replaced with the *aadA* selectable marker (Wannathong et al., 2016).

Tris-Acetate Phosphate (TAP) (Gorman and Levine, 1965) or High Salt Minimal (HSM) (Sueoka, 1960) medium containing 2% agar (w/v) at 25°C. Liquid cultures were done using TAP that was inoculated with a loopful of cells and incubated at 25°C with shaking at 120 rpm until reaching the mid-log phase (5 x 10^6 cells/ml). Both cultures were grown in a chamber at 50 μE m^2/s, for photosynthetic mutants the cultures were wrapped in white tissue. Chlorophyll-deficient mutants were grown in dark conditions.

2.1.2 *Escherichia coli*

All bacterial DNA manipulation was done in the DH5α *E. coli* strain, genotype: *dlacZ M15 (lacZYA-argF) U169 recA1 endA1 hsdR17(rK-mK+) supE44 thi-1 gyrA96 relA1* (Hanahan, 1983), provided by Clontech (Saint-Germain-en-Laye, France). Lysogeny broth medium was used for growth (Bertani, 1951), with the addition of ampicillin (100 μg/ml) for selection. Colonies were cultivated overnight at 37°C on LB medium containing 2% Difco agar. Inoculation from a single colony was done using a sterile toothpick. Liquid cultures were grown in a shaking incubator (200 rpm, Innova 4430, New Brunswick) at 37°C overnight.

2.2 Molecular biology

2.2.1 *Gene design and gene optimization*

The Genbank (https://www.ncbi.nlm.nih.gov/genbank/) database was used as the source for genetic sequences. The BLAST tool was utilized for sequence corroboration (Wheeler et al., 2007). All genes used in this study were codon optimized to match the requirements of the chloroplast genome using the Codon User Optimizer (CUO) (http://codonusageoptimizer.org/-download/). All the optimized genes were synthesized by GeneArt (www.life-technologies.com) and included an HA-Tag at the C-terminus for protein immunodetection.

2.2.2 *DNA extraction from Chlamydomonas reinhardtii*

In order to determine the genotype of transformants, a quick DNA extraction was performed (Chelex crude preps) (HwangBo et al., 2010). First, transformed *C. reinhardtii* TN72 cells were resuspended in 20 μl of water. Then, 20 μl of 100% ethanol and 200 μl of 5% (w/v) Chelex-100 resin (Bio-rad) were added. The mixture was boiled (99°C) for 5 min and cooled on ice. After centrifugation (13 000 g, 2 min) the supernatant was transferred to a new tube and used for amplification.

2.2.3 *DNA extraction from Escherichia coli*

Small quantities of plasmid DNA were extracted using the Qiagen Miniprep kit following the manufacturer's instructions. Firstly, 5 ml of *E. coli* culture were subjected to alkali lysis. Subsequently, potassium acetate was used to precipitate proteins and genomic DNA. Finally, a silica-based column was used to bind plasmid DNA, which after washing with ethanol was eluted in water.

Extraction of large quantities of plasmids was done using the Qiagen Midiprep kit following the manufacturer's instructions. Here, 50 ml of *E. coli* culture were disrupted using alkali lysis. Filtration was used to eliminate proteins and genomic DNA. The flow through was loaded into an ion exchange column to sequester DNA plasmids, which were later eluted in sodium chloride solution. Ethanol precipitation was used for concentration.

2.2.4 *Polymerase Chain Reaction (PCR)*

Amplification of DNA was performed using the polymerase chain reaction (PCR) (Saiki et al., 1988). Enzymatic reactions were done in a total volume

of 25 μl containing 5 μl of Phire II reaction buffer (New England Biolabs), 0.25 μl of each primer (100 pmol/μl), 0.25 μl of Phire II Hot Start DNA polymerase (40 U), 10 ng of template DNA for plasmids. Doubled-distilled water was used to top up to final volume. Amplification was executed using a Techne TC-3000X thermocycler. Initial denaturation was done at 98°C for 30 s. The reactions were subjected to 30 cycles of denaturation (98°C, 5 s), annealing (3°C above the lower Tm of the primers used, 5 s), and extension (15 s / 1 kb of amplicon). A final incubation was done at 72°C for 2 min, and samples were held at 4°C. 10 μl of the reaction were run in an agarose gel.

2.2.5 *Digestion using restriction endonucleases*

DNA extracts were digested utilizing restriction endonucleases (10 U/μg of DNA) (New England Biolabs) following the manufacturer's guidelines. Double digestions were done using an appropriate ratio between the enzymes and with the addition of a suitable buffer.

2.2.6 *Electrophoresis*

Agarose gels (1% (w/v) of TAE buffer) supplemented with ethidium bromide (0.1 μg/ml) were used to separate DNA amplicons and fragments. Gels were made to a total volume of 50 ml and loaded into electrophoresis tanks (www.sigmaaldrich.com) filled with TAE buffer (0.04 M Tris, 1 mM sodium EDTA and 17.5 mM glacial acetic acid). 6x loading dye (2.5% (w/v) Ficoll 400, 11 mM EDTA, 3.3 mM Tris-HCl, 0.017% (w/v) SDS and 0.015% (w/v) Bromophenol Blue) was added to the samples to reach a 1x concentration. 0.5 μg of size markers (New England Biolabs) were loaded per lane. Gels were run at 100 V (Gibco BRL power supply), and later

observed on an UV illuminator.

2.2.7 *Dephosphorylation of digested plasmid DNA*

In order to reduce vector background during cloning procedures, digested plasmid were treated with Antarctic phosphatase (5 U/μg of DNA) (New England Biolabs) according to the provided instructions. Incubation (37°C) was done for an hour and inactivation (75°C) for 5 min.

2.2.8 *Purification of PCR products*

Purification of amplicons and small fragments, generated by digestion, was done using the Qiagen PCR purification kit as advised by the guidelines. Binding buffer and samples were mixed in equal volumes and then loaded in a silica-based column.

2.2.9 *Ligation of DNA pieces*

Ligation reactions were done in a 10 μl final volume and incubated for 15 min at room temperature. 40 U of T4 ligase (New England Biolabs) per μg of DNA was used for the reaction.

2.2.10 *Sequencing*

Sanger sequencing (www.sourcebiosciences.com) was used to confirm plasmid constructs and PCR products. Plasmid DNA was provided at 9-16 fmoles/μl, while PCR products were supplied at 100ng/μl. Primers were supplied at 2-5 pmols/μl.

2.3 Genetic transformation

2.3.1 *Escherichia coli*

2.3.1.1 *Preparation of competent cells*

10 ml of LB medium was inoculated with 10 μl from a glycerol stock of the *E. coli* DH5α strain, and incubated overnight at 37°C. This culture was used to inoculate 100 ml of LB (100-fold dilution) and incubated for 3 hours at 37°C. After cooling on ice, centrifugation (4000 g, 5 min) was performed to pellet cells, which were then resuspended in 10 ml of cold 50 mM $CaCl_2$ and left on ice for 30 min. Centrifugation (4000 g, 5 min) was performed again but this time cells were resuspended in 1.5 ml of fresh 50 mM calcium chloride ($CaCl_2$). 3.5 ml of sterile glycerol 50% (v/v) were added. Finally, aliquots (400 μl) were made and stored at -80°C until required.

2.3.1.2 *E. coli transformation*

Stocks of competent cells were thawed on ice and 100 μl were used for transformation. Ligation mixtures (subsection 2.2.9) were added to the cells, while constructed plasmids were diluted by a factor of ten and 1 μl was used. Negative controls were set up wherein no DNA was added. Samples were left on ice for 30 min before heat-shock was performed at 42°C (1 min). Then, they were cooled on ice for 2 min. 1 ml of LB medium was added to each reaction, which were then incubated at 37°C for 60 min. For intact plasmids 100 μl of culture were spread onto LB agar plates containing 100 μg/ml of ampicilin. For ligation reactions, the 1 ml culture was centrifuged (10 000 g, 2 min) and 900 μl were discarded. Cells were resuspended in the remaining 100 μl and spread onto plates as above.

2.3.2 *Chlamydomonas reinhardtii chloroplast transformation*

2.3.2.1 *Glass beads method*

Homologous recombination between DNA flanking sequences allowed the introduction of foreign genes into targeted sites of the chloroplast genome of *C. reinhardtii*. The glass bead method was chosen for delivery of vectors into the recipient line TN72 or *cw15* (Economou et al., 2014). Cells were grown in 500 ml of TAP medium until reaching a density of 2 x 10^6 cells/ml. Cells were centrifuged (38000 g, 5 min, 16°C) for harvesting and resuspended in TAP medium to reach 2 x 10^8 cells/ml. Subsequently, a mixture of 0.3 g of glass beads (Sigma-Aldrich, 0.4-0.6 mm diameter), 10 μg of circular plasmid DNA and 300 μl of cells were vortexed in a glass tube for 15 s. For transformation involving photosynthetic mutants, high-salt minimal medium (HSM) (Kindle et al., 1991) (3.5 ml) containing 0.5% agar (w/v) was added to the mixture, which was then poured onto HSM plates containing 2% (w/v) agar. For transformations using the spectinomycin resistance marker (aminoglycoside -3-adenyltransferase -*aadA*-), TAP (3.5 ml), containing 0.5% (w/v) agar and 100 mg/ml of the drug, was added to the mixture, which was then poured onto TAP plates containing 2% agar (w/v), and the same concentration of the antibiotic as above. Incubation was performed for four weeks in a growth chamber at 25°C under 50 μE m / s light. Homoplasmicity of the transgene was checked by PCR (Wannathong et al., 2016).

2.4 *Protein analysis*

2.4.1 *Protein preparation for SDS-PAGE*

.

2.4.1.1 *Escherichia coli*

E. coli liquid cultures were grown overnight until they reached stationary phase. Cultures were normalized to an optical density (OD) of 2.5 at 600 nm. 1 ml of cultures was centrifuged (13 000 g, 2 min) and cells were resuspended using 600 μl of Solution Ab (0.8 M TrisHCl pH 8.3, 0.2 M sorbitol, 1% (v/v) β-mercaptoethanol, % (v/v) SDS, 0.08% (w/v) bromophenol blue). Samples were stored at -80°C.

2.4.1.2 *Chlamydomonas reinhardtii*

C. reinhardtii cultures (20 ml) were grown until mid-log phase (3-6 x 10^6 cells/ml), for dark anaerobic growth, once the cultures reached the mid-log phase they were place into an Oxoid™ anaerobic jar (https://www.thermofisher.com/order/catalog/product/HP0011A) covered in aluminum foil and left to grow for another 48 hours. Cells were harvested by centrifugation (4000 g, 5 min), then resuspended in Solution A (0.8 M TrisHCl pH 8.3, 0.2 M sorbitol and 1% (v/v) β-mercaptoethanol) to equal densities as measured by light scattering (OD: 750 nm). Samples were stored at -80°C.

2.4.2 *Protein preparations in non-denaturing conditions*

2.4.2.1 *Chlamydomonas reinhardtii*

Cells (1000 ml) were grown in TAP until mid-log phase (3-6 x 10^6 cells/ml) and optical density was measured at 750 nm. Cells were centrifuged (5 000 g, 5 min) and resuspended in a volume of PBS (1.37 M NaCl, 27 mM KCl, 100 mM Na_2HPO_4, 18 mM KH_2PO_4) corresponding to the measured OD multiplied by the (culture volume)/20 (for around 50 x concentrated). Centrifugation (5 000 g, 5 min) was performed again to remove the PBS. Cells were then resuspended in PBS containing a complete, EDTA-free Protease Inhibitor Cocktail (http://www.sigmaaldrich.com). Samples were frozen using liquid nitrogen and then thawed in water bath (35°C, 10 min). This step was repeated three times. Samples were then centrifuged (5 000 g, 5 min) and the supernatant was transferred to an ultracentrifugation tube and topped up to 10 ml with PBS plus protease inhibitor. Ultracentrifugation was performed at 38 000 g for 60 min at 16°C. The supernatant was removed from the tube and stored at -80°C.

2.4.3 *Protein separation using polyacrylamide gel electrophoresis (SDS-PAGE)*

2.4.3.1 *Sample preparation*

Chemically disrupted cells of *C. reinhardtii* and *E. coli* were boiled at 99°C for 1 min and then cooled on ice for 2 min. After centrifugation, the supernatant was collected and used for analysis. Mechanically broken cells of *C. reinhardtii* and *E. coli* were prepared by adding a 4x protein loading dye (200 mM TrisHCl pH 6.8, 8% (w/v) SDS, 40% (v/v) glycerol, 4% (v/v)

β-mercaptoethanol, 8% (v/v) SDS 50 mM EDTA, 0.08% (w/v) bromophenol blue) to make a 1x final concentration. The samples were prepared in the same manner as the chemically disrupted cells.

2.4.3.2 *SDS-PAGE gel preparation*

Gels were prepared in a Mini-PROTEAN Tetra cell system Bio-Rad (www.bio-rad.com) according to the manufactures instructions. The resolving gel (15% acrylamide (w/v)) was prepared to a final volume of around 8 ml, and consisted of 2.3 ml of 40% (w/v) acrylamide, 1.9 ml of 1.5 M TrisHCl pH 8.8, 75 μl of 10% (w/v) SDS, and 2.7 ml of dH_2O (Laemmli, 1970). Polymerization was achieved by adding 75 μl of 10% (w/v) ammonium persulphate and 3 μl of tetramethylethylenediamine (TEMED). 5 ml were poured for a 1.0 mm thick gel. A layer of 100% ethanol was poured above the gel to help it set and avoid any possible inhibition of polymerization. Gels were left for 30 min to set. Then, the ethanol was removed and dH_2O was used to wash several times and assure the removal of any non-polymerized acrylamide.

The stacking acrylamide gel (3.75% (w/v)) was made to a final volume of approximately 6 ml consisting of 375 μl of 40% (w/v) acrylamide, 375 μl of 1 M TrisHCl pH 6.8, 30 μl of 10% (w/v) SDS, and 2.2 ml of dH_2O (Laemmli, 1970). Polymerization was triggered by adding 30 μl of 10% (w/v) ammonium persulphate and 3 μl of TEMED. 1.5 ml were poured for a 1.0 mm thick gel. Following insertion of the suitable comb, polymerization was permitted for 30 min. After removing the comb, the gasket was assembled and the samples loaded.

2.4.3.3 *Loading and running of samples*

15 μl of the samples were loaded onto the stacking gel. Proteins were separated at 150 V for 90 min in a Tris-glycine electrophoresis buffer (0.25 M Tris, 1.92 M glycine, 1% (w/v) SDS pH 8.3). For immunodetection, gels were electroblotted onto a nitrocellulose membrane.

2.4.4 *Western blot analysis*

A BioRad Transblot semi dry cell (www.bio-rad.com) was used to transfer proteins to a Hybond ECL nitrocellulose membrane (www3.gehealthcare) at 120 V for 1 h (Towbin et al., 1979). The membrane was previously stacked between 4 sheets (above and below) of 3 MM Whatman paper, and soaked for 30 min at room temperature in Towbin buffer (25 mM Tris base, 192 mM glycine and 20% (v/v) methanol). The gel was placed above the membrane for transferring.

2.4.4.1 *Immuno-detection*

After transferring, the membrane was blocked with skimmed milk 0.5% (w/v) in TBS-T (20 mM Tris, pH 7.4, 0.1% (v/v) Tween-20) overnight at 4°C. Membranes were incubated for one hour with the primary antibody, washed with TBS-T for 30 min, and then incubated for another hour with the secondary antibody. The membranes were washed again for 30 min in TBS-T. All steps were performed at room temperature with shaking at 60 rpm. The primary antibodies were produced in rabbit and were αHA (Sigma-Aldrich product H6908 1:5000 in TBS-T). The secondary antibody for all blots was α-rabbit IgG horseradish peroxidase-linked antibody from goat (Thermo Scientific product 3557, 1:20 000 in TBS-T). For detection of the human growth hormone, primary polyclonal antibodies were produced

in chicken were used (Abcam product ab27223 1:5000 in TBS-T). The secondary antibody was α-chicken IgY horseradish peroxidase-linked antibody from donkey (Abcam product ab971315, 1:5000 in TBS-T). For detection of the large subunit of the Rubisco enzyme, which was used as a loading control, primary polyclonal antibodies produced in rabbit were used (obtained from John Gray, Cambridge University, 1:5000 in TBS-T). The secondary antibody for all blots was α-rabbit IgG horseradish peroxidase-linked antibody from goat (Thermo Scientific product 3557, 1:10 000 in TBS-T). Detection was done employing the Licor Odyssey Infrared Clx imagining system. Wet membranes were placed onto the scanner making sure that no bubbles were trapped. Images were processed using the Image Studio Lite 5.0.

2.4.4.2 *Quantification of western blot analysis*

The Odyssey Infrared Imaging System (Li-COR Biosciences) was used to determine the infra-red (IR) fluorescent signal of proteins bands in Western blot analyses, which was analyzed utilizing the Image Studio Lite 2 (using the 800 nm channel). To quantify the signal, rectangles were drawn upon each band, the values obtained were normalized by dividing them to their respective loading control values.

2.4.4.3 *Measurement of total soluble protein*

Samples were prepared for being mechanically broken as denoted in subsection 2.4.1.4. After ultracentrifugation the supernatant was used for analysis with the Bradford assay (www.sigmaaldrich.com) (Bradford, 1976). 5 μl of standard bovine serum albumin (BSA) were added to separate wells in a ninety six well plate, a blank sample was done using PBS buffer. Standards were utilized with a concentration ranging from 0.1 to 1.4 mg/ml. 250 μl of Bradford reagent were added to each plate and mix on a shaker (60

rpm) for 30 s. Incubation was done for 15 min at room temperature, and absorbance were measured at 595 nm. Plotting absorbance against protein concentration generated a standard curve, which was then used to determine the values of the tested samples.

2.5 *Proteins activity assays*

2.5.1 *5-fluorocytosine sensitivity assay*

5-fluorocytosine (5-FC, product F7129) (www.sigmaaldrich.com) was prepared as 10 mg/ml solution in TAP medium. *C. reinhardtii* cultures (20 ml) were using light scattering at 750 nm (OD: 1). 5 μl of cells were dissolved in TAP (3 ml) 0.5% (w/v) agar containing 2 mg/ml of 5-FC. This mixture was poured onto TAP medium plates containing 2 mg/ml of 5-FC, and no drug as control. Plates were incubated at 25°C under 50 μE m^2 / s light for 10 days.

2.5.2 *Fluorescence determination*

Cells were grown in TAP until the mid-log phase, then 3 ml were transferred, in triplicate, into Fisherbrand™ Disposable Cuvettes (Fisher Scientific, www.fishersci.com), and fluorescence was detected using the PerkinElmer LS 55 Fluorescence spectrometer (PerkinElmer, Waltham, MA, USA). The measurements were acquired using the optimal excitation/emission spectra (458 nm/460-495 nm). The plot was generated utilizing the normalized values. Normalized values from the control were subtracted from the transformants signals to assess the increase in fluorescence over the background. The fold increase in fluorescence was obtained by dividing the normalized

reading by the normalized values of the wild type.

2.5.3 *Fluorescence microscopy*

The crTur1 strain was used to inoculate 20 mL of TAP until reaching the mid log phase. Cells were spotted on TAP agar (1%) for image acquisition using a Leica SPS Confocal Laser Scanning Microscope (http://www.leica-microsystems.com) endowed with an HCX PL APO x 63 1.2 W objective. An argon ion laser was used to excite the fluorescent protein at 458 nm, emission was measured within the 465-495 nm spectral window. Chloroplast auto-fluorescence was determined across the 610-700 nm spectral window. Images were analysed using the Fiji image processing package.

2.5.4 *Determination of crNifV activity*

The standard assays for measuring crNifV activity of both crude extracts and purified enzymes were carried out in 100 μl that contained 0.5 mM of acetyl coenzyme A (AcCoA) and 1 mM of α-ketoglutarate (α-KG) in 100 mM of HEPES buffer (pH 7.5) (Gabriel and Milewski, 2016; Wulandari et al., 2002). The negative control did not include AcCoA. After addition of cell free extracts containing different amounts of total protein (150 μg - 0 μg), the mixtures were incubated for 10 min at room temperature. The reaction was ended by adding 100 μl of 5,5-dithiobis(2-nitrobenzoic acid) (DTNB) solutions (2 mg/ml) dissolved in 95% EtOH, which was followed by 10 min incubation at room temperature. The amount of 5-thio-2-nitrobenzoic acid produced as a consequence of the reduction of DTNB with CoASH was assessed by measuring the absorbance at 412 nm (ϵ= 13 600 M / cm).

Chapter 3

Attempts to express multiple transgenes simultaneously in the chloroplast of *C. reinhardtii*

A directed alteration of native metabolites or expression of foreign genes for novel metabolic pathways can be referred to as metabolic engineering. Metabolic engineering involves molecular biology procedures, analytic techniques and genomic methods. For algal biotechnology, this approach can be useful for developing strategies to cope with current issues such as sustainable biofuel production, enhancement of photosynthesis, biohydrogen production and nitrogen fixation (Bock, 2013; Lü et al., 2011; Mathews and Wang, 2009). One important issue of metabolic pathways is their intrinsic complexity. For example, various enzymes may have multiple functions. Also, feedback mechanisms are required for correct control of synthesis of the final product. These conditions impose constraints when trying to manipulate gene expression. Therefore, it is recommended that for such a task the arrangement of synthetic constructs must avoid native regulation, which

means the use of promoters, and UTRs from a different organism (Temme et al., 2012). Moreover, achieving the correct stoichiometric balance of the constituents of a metabolic complex is one of the most significant bottlenecks in metabolic engineering. Therefore, special emphasis has been placed in this area. In *E. coli*, for instance, polycistronic expression cassettes have been used to generate heterologous protein complexes. In this system, different genes of interest are separated by ribosome binding sites (Shine-Dalgarno sequences), and are put under the control of a single promoter. Similarly, in eukaryotic platforms coding regions have been linked together by internal ribosomal entry sites (IRESs) and stacked under one promoter. Additionally, some methods, inspired by viral proteome processing, have been developed based on proteolytic management of multiproteins. (Bieniossek et al., 2012) (Figure 3.1). In plant-based expression systems, approaches such as co-transformation or serial transformation have been applied for this purpose. The process of co-transformation involves the insertion into the genome of two or more genes simultaneously. This procedure can be done in two different ways: genes that are to be inserted can reside on the same plasmid, or they could be organized in different plasmids. These two methods can be performed using common procedures for genetic transfer (Naqvi et al., 2010). In another case of co-transformation, known as combinatorial transformation, individual plasmids contain different expression cassettes with individual transgenes. The plasmids that bear the chosen traits are mixed with an additional one that carries the marker gene for further insertion in the chloroplast genome (Bock, 2013) (Figure 3.2A). An alternative method for introducing multiple genes involves the insertion of extra genes in an already transgenic line; this procedure has been named serial transformation or super-transformation. Despite its benefits, this technique may present some limitations such as the need of additional selectable markers, and the extra steps involving their excision (Bock, 2013; Naqvi et al., 2010) (Fig-

ure 3.2B).

Figure 3.1: Approaches for expressing protein complexes. A. In *E. coli*, polycistrons are used to produce various proteins from a single promoter. The coding sequences are spaced apart by ribosome binding sites (*rbs*). A sequence (Term) for ensuring mRNA transcription termination is added. B. For eukaryotic expression, internal ribosomal entry sites (IRES) are used to link the coding sequences resulting in polycistronic mRNAs. C. Protein complexes can be generated from polyproteins. Polyproteins could be designed so they contain a highly specific protease (e.g. tobacco etch virus (TEV)Nla), which processes the polyprotenis by cleaving at the cleavages sites (*tcs*) and releasing the individual proteins. A drawback of this approach is the adornment of the preceding protein with residues from the cleavage sequence. The asterisk denotes the cleavage point, and the amino acids underlined the C-terminal overhangs. Modified from (Bieniossek et al., 2012).

These methods demonstrate that expressing foreign metabolic pathways is feasible, to some extent, in well-studied expression platforms such as bacteria or plants. Furthermore, these methods provide a solid foundation for further research in less-established expression platforms, such as microalgae. Based on the existent techniques for metabolic engineering, it seems essential to guarantee the proper flow of information during the processing of a particular genetic complex. Whether it is at transcriptional, post-transcriptional or at a translational, post-translational level. Therefore, the purpose of the first part of this research is to examine various alternatives that might prove useful to achieve optimal transcription, mRNA processing, and translation of

foreign multiple genes engineered in the plastidial genome of *C. reinhardtii*.

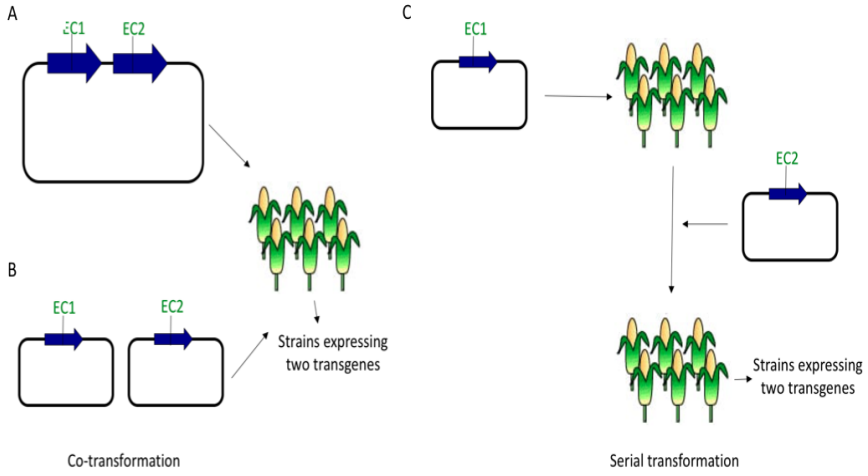


Figure 3.2: Methods for transferring multiple genes in plants. A. Co-transformation. Multiple genes can be included within a single transformation vector. B. Several genes can be stacked in multiple vectors to be simultaneously delivered. B. Inserting foreign coding sequences sequentially generates transgenic lines expressing multiple transgenes. Namely, transgenic lines are transformed with additional genes. (EC expression cassette)

The processing of chloroplastic information from DNA (Maul et al., 2002) to RNA (Stern et al., 2010) to proteins (Terashima et al., 2011), and their impact on the cell metabolism (Wienkoop et al., 2010) has been studied thoroughly in *C. reinhardtii*. Several diverse processes occur in the organelle such as the synthesis of fatty acids, isoprenoids, starch and amino acids (Harris et al., 2009), rendering it the main biosynthetic centre of the cell. It has been widely used for the expression of foreign proteins with potential commercial value (Johanningmeier and Fischer, 2010; Mayfield et al., 2007), which has prompted the development of more user-friendly approaches for generating transplastomic lines (e.g. delivery methods or selection markers based on endogenous genes). Thus, it is of necessity to develop straightforward approaches to express simultaneously multiple transgenes in the chloroplast. Such methods could then be used for expressing different important biomolecules in one cell, and for expressing foreign metabolic pathways or

altering native ones. In this chapter, I present various methods focusing on different stages of the processing of chloroplastic genomic information. First, I tested an idea based on post-translational regulation. A protease system was devised whereby two proteins were linked as a fusion protein, by the recognition site of a very specific protease with the objective that after cleavage the two components of the polyprotein are separated. Alternatively, I designed a strategy in which simultaneous expression of multiple genes could be attained by controlling correctly their transcription. To accomplish this task, each coding sequence was flanked by the proper sequences for driving gene expression, these expression cassettes could then be ultimately introduced in tandem into the plastome. Here, I describe the study of these various schemes designed to accomplished simultaneous gene expression in the chloroplast, and which could be used as a basis for expressing foreign metabolic pathways in the organelle.

A recently straightforward scheme for chloroplast transformation has been established (Economou et al., 2014), and many subsequent studies have successfully applied for analyzing various aspects of chloroplast genetic engineering (Braun-Galleani et al., 2015; Gangl et al., 2015; Young and Purton, 2015; Zedler et al., 2015; Young and Purton, 2014). This technique employs a *C. reinhardtii* mutant recipient strain (TN72) wherein an expression cassette containing the *aadA* antibiotic resistance gene has replaced the *psbH* gene, which encodes a thylakoid membrane protein that is an essential component of photosystem II (Summer et al., 1997). The vector used for inserting the gene of interest (pASapI) carries a functional version of the *psbH* gene. Upon homologous recombination, the expression cassette carrying the transgene, along with the *psbH* gene (which restores photosynthetic growth), replaces the spectinomycin-resistance expression cassette. In the pASapI vector, gene expression is driven by the *atpA* promoter / 5‘ UTR and the *rbcL* 3‘ UTR (Figure 3.3). The obvious advantage of this method

is that the gene of interest remains the only foreign genetic information introduced into the plastome. Therefore, I decided to use this approach for testing the different schemes for multiple gene expression.

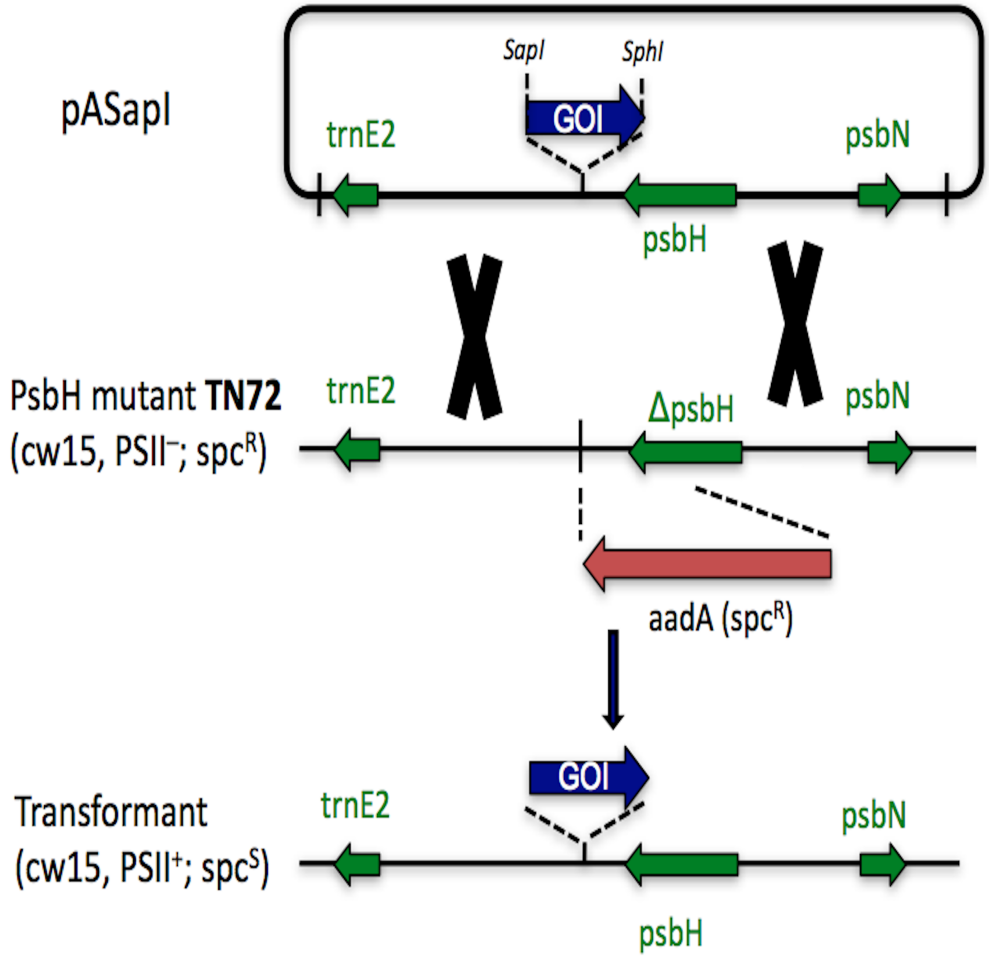


Figure 3.3: Chloroplast transformation approach. The pASapI vector has been assembled in a way so it contains a functional version the *psbH* gene, which restores a functional PSII. The expression cassette consists of the *atpA* promoter / 5‘ UTR, the gene of interest (GOI) and the *rbcL* 3‘ UTR, the GOI is coupled to the expression vector via a *Sap*I - *Sph*I digestion ligation event. Both the expression cassette and the functional version of the *psbH* gene are flanked by sequences needed for homologous recombination. In the recipient strain (TN72), the *aadA* expression cassette, which provides resistance to spectinomycin or streptomycin, has been used for disruption of the *psbH* gene. Therefore, after transformation cells are able to grow photosynthetically, and are no longer resistant to the antibiotic.

3.1 A protease-based system for processing of a recombinant polyprotein

3.1.1 Plasmid design for testing the protease system

Viral strategies could serve as a foundation for designing synthetic systems for processing multiproteins. For example, the coronavirus (responsible for the severe acute respiratory syndrome -SARS-) expresses its genome by translating two open reading frames that constitutes two polyproteins. Then, individual proteins are released from the polyproteins by a protease, which cleaves at the proteolytic sites located between each protein subunit (Vijayachandran et al., 2011). I wondered whether a similar method could be applied to generate protein complexes from polyproteins in the chloroplast of *C. reinhardtii*. In this approach, I selected a highly specific serine protease (SpIB) from *Staphylococcus aureus* based on its range of action (4 - 30°C / pH 6.5 - 9.0) (Pustelny et al., 2014), which fits perfectly with the growing conditions of *C. reinhardtii*. I used the BLAST tool to find out if any important protein component of the chloroplast contains the recognition sequence (WELQX) for this enzyme. The only native target appears to be a synthase involved in starch production (SS3). Nonetheless, this isoform is not required for optimal production of starch (Szydlowski et al., 2009). As a substrate for this enzyme I have designed a gene fusion consisting of a codon optimized version of the bacterial cystosine deaminase gene (*codA*), and a codon optimized version of the human growth hormone gene (*hGH*) interconnected by a linker encoding the amino acids (WELQP) recognized and cleaved by the protease. I decided to use the *codA* gene as the second coding sequence to be added to the plasmid. This enzyme is capable of producing a toxic product, 5-fluorouracil (5-FU) when the synthetic compound 5-fluorocytosine (5-FC) is present in the medium (Young and Purton,

2014). 5-FU can be incorporated into newly synthesized RNA, which leads to protein synthesis inhibition. Moreover, 5-FU serves as a precursor for the synthesis of a DNA synthesis inhibitor (5-fluorodeoxyuridine monophosphate) that destabilizes thymine synthesis (Vermes et al., 2000). This gene has been used as a negative marker for developing synthetic biology tools for the chloroplast of *C. reinhardtii* (Young and Purton, 2015, 2014). Similarly, the human growth hormone has been successfully used to compare promoter strength in the chloroplast of *C. reinhardtii* (Wannathong et al., 2016). With this system, the idea of producing polyproteins that could be digested using specific proteases could be tested. In this case, a fusion protein, consisting of the cytosine deaminase and the human growth hormone, should be separated into its subunits after proteolytic treatment (Figure 3.4).

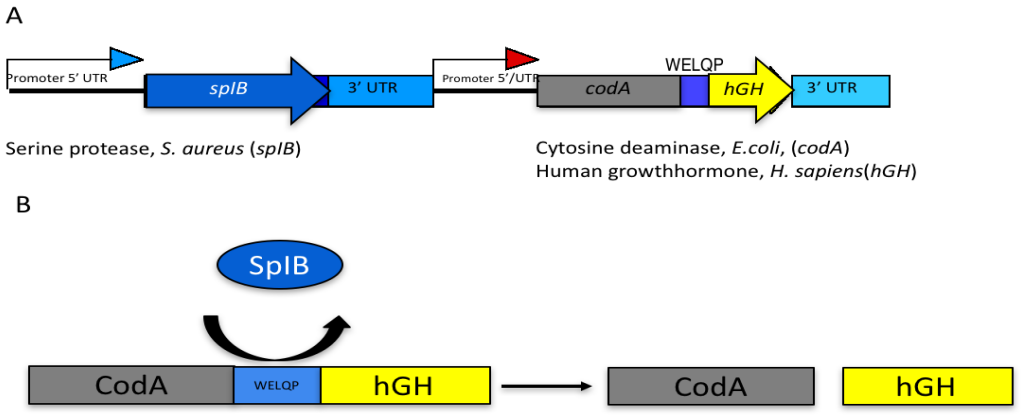


Figure 3.4: Strategy for processing polyproteins in the chloroplast of *C. reinhardtii*. Protein complexes could be produced and processed using fusion proteins. A. The coding sequence for the fusion protein (grey/yellow) must have the genes of interest spaced apart by the recognition site of the protease (WELQP-blue), and must be under the control of a single promoter (red arrow). The highly specific protease (blue) must be under the control of a different promoter (blue arrow). Both expression cassettes must have a 3‘ UTR that assures the proper processing of the mRNA. Both expression cassettes could be fit into one transformation plasmid. B. The highly specific protease ought to process the fusion protein by cutting the correspondent site (WELQP), and consequently releasing the individual sub-units.

Codon optimization for expression of SpIB in the chloroplast of *C. reinhardtii* was done using the codon usage optimizer (CUO) software, which

permits the design of a gene based on a chosen codon usage threshold. A variant of the plasmid construct (phGH) used for expressing the human growth hormone in the chloroplast of *C. reinhardtii* (Wannathong et al., 2016) was used as a backbone for generating the fusion protein, in this variant an *Nco*I site lies in front of the coding sequence (Figure 3.5A). In this construct, gene expression is driven by the *psaA exon 1* promoter 5‘ UTR and the *rcbL* 3‘ UTR. The plasmid was linearized with *Nco*I restriction for the insertion of the codon-optimized version of the cytosine deaminase (*codA*) coding sequence. The *codA* gene was amplified from the plasmid used for its expression in the chloroplast of *C. reinhardtii* (Young and Purton, 2014). The forward primer bore an *Nco*I site, while the reverse primer bore a *BspH*I site, which leave the same 3‘ overhangs as *Nco*I. The reverse primer contains also the codon-optimized sequence of the protease recognition site (WELQP) (Appendix A.1).

The *codA* gene was ligated to the *Nco*I-digested backbone (phGH) to give rise to the pF plasmid (Figure 3.5B) (Appendix A.2). This plasmid contains the genetic information for the fusion protein, consisting of the cytosine deaminase (*codA*) gene and the human growth hormone (*hGH*) gene. The codon-optimized version of the serine protease (*spIB*) was designed with a *Sap*I site at the 5‘ end and an *Sph*I site at the 3‘ end along with a HA-Tag sequence at the end of the coding region. The *codA* gene was cloned in the pASapI chloroplast expression vector, via *Sap*I-*Sph*I digestion / ligation. In the resulting plasmid (pS), SpIB expression is under the control of the *atpA* promoter and the 3‘ *rbcL* UTR (Appendix A.2). The *codA*-containing expression cassette from the pS plasmid was amplified using a forward primer bearing the *Mlu*I site and a reverse primer bearing a *BssH*II site, which leaves the same 3‘ overhangs as *Mlu*I (Appendix A.1). The resulting piece of DNA was ligated to the *Mlu*I-digested pF plasmid. The generated vector was named pF+P, and contains the genes encoding the fusion protein and

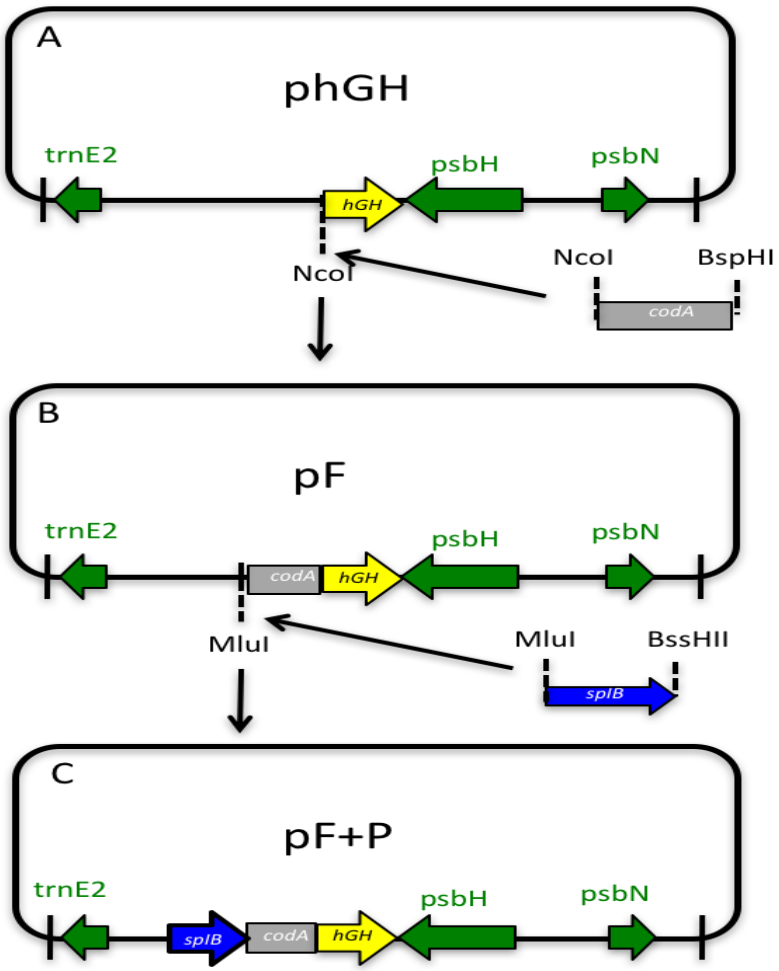


Figure 3.5: Plasmid assembly for testing the protease system. The phGH plasmid contains the gene for the human growth hormone, and was digested with *Nco*I for the addition of an extra coding sequence (A). The cytosine deaminase gene (*codA*, grey) was amplified with primers bearing the *Nco*I and the *BspH*I sites (A) and, then, inserted into the phGH backbone to generate the pF plasmid (B). The latter plasmid was then digested with *Mlu*I for the addition of the serine protease (*spIB*-blue) (accession number of the original *S. aureus* gene product: `WP_001633939`) expression cassette. The *spIB* gene was amplified with primers carrying the *Mlu*I and the *BssH*II restriction sites (B), and later ligated to the pF backbone to yield the pF+P plasmid (C). All plasmids were checked by sequencing

.

the protease (Figure 3.5C) (Appendix A.2).

In the final construct (pF+P), gene expression was driven by two different promoters 5‘ UTRs, and mRNA stability was provided by the same *rbcL*

3‘ UTR (Figure 3.6). An interesting advantage of this particular assembly was the fact that two expression cassettes were both located between the sequences for homologous recombination. Hence, beyond testing the proposed protease-based system, this genetic arrangement was useful to study the potential integration of various expression cassettes (in tandem) into the chloroplast genome.

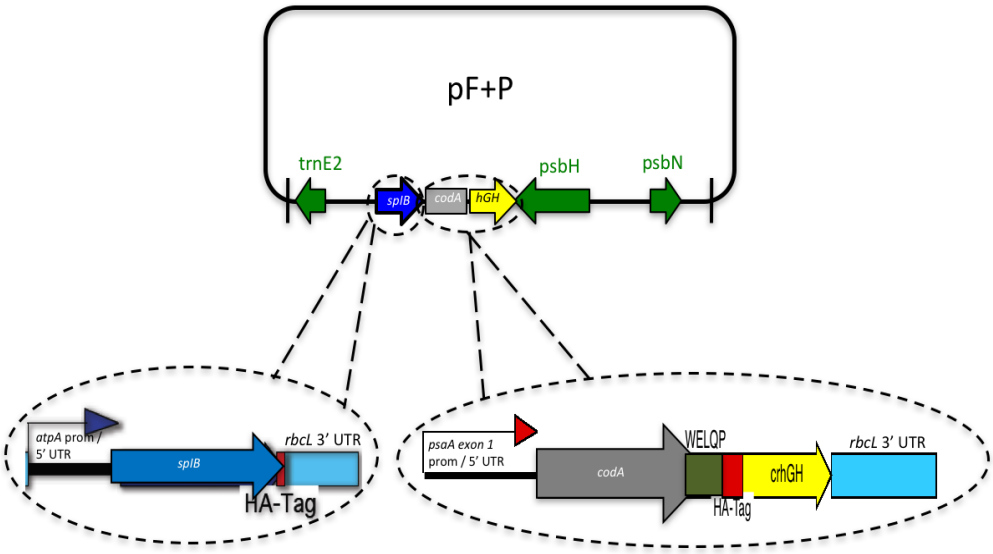


Figure 3.6: The pF+P expression vector. The pF+P plasmid is based on the pASapI vector, so it carries an operative version of the *psbH* gene along with the sequences needed for homologous recombination. Nonetheless, this vector bears two different genes of interest (*spIB*, *codA:hGH*) engulfed within their own expression cassettes. The same 3‘ UTR (from the *rbclL* gene) was used in all of them, while different promoters 5‘ UTR drive gene expression in each cassette. The sequences encoding the fusion protein, and the serine protease were HA-Tagged for immunodetection

.

3.1.2 Transformation of the recipient strain (TN72) with the plasmid bearing the protease system.

The pF+P plasmid was delivered using glass beads transformation into the chloroplast genome of *C. reinhardtii* as described in section 2.3. Primary transformants were selected on high salt minimal (HSM) medium with no

antibiotic. Then, growing colonies were streaked onto new plates to ensure homoplasmicity, which was confirmed by PCR (Appendix A.1). Primers were designed so they give different size products whether the insertion had been successful or not. Correct integration yielded a 1.2 kb amplicon for the analyzed crF+P transformants (Figure 3.7). In contrast, if homologous recombination had not happened, a 0.85 kb product would have been present, such as in the case of the host strain TN72 (Figure 3.7). The absence of an amplicon of this size from the transformants indicates that all the copies of the chloroplast genome lack the foreign *aadA* cassette, and that the functional *psbH* gene has been inserted along with the transgenes.

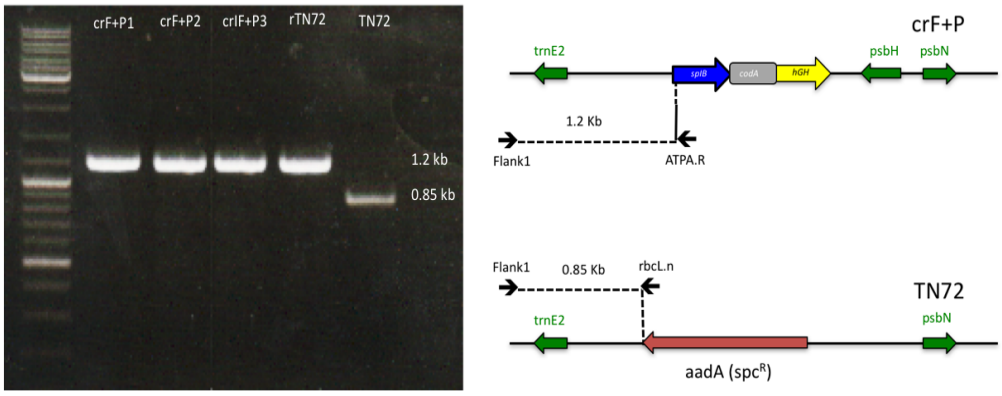


Figure 3.7: Identification of transgenic lines by colony PCR. Left-hand side: Agarose gels showing the PCR products from transformants lines and controls. Right-hand side: Diagrams illustrating the genetic arrangement of the TN72 genome and the transformants genomes. The Flank1 primer is targeted to the 5‘ upstream of the homologous recombination site. Whereas, the reverse primers target either a region of the *aadA* cassette or a section of the inserted expression cassette. The three primers were used in a single reaction. The 1.2 kb bands (crF+P lanes) demonstrates the correct integration of the serine protease (*spIB*, blue) and the cytosine deaminase:human growth hormone (*codA*:*hGH*, grey:yellow) cassettes. The rTN72 lane corresponds to the positive control, this strain has been transformed with an empty expression cassette. The 0.85 kb bands correspond to the recipient strain TN72, which was used as a negative control.

.

3.1.3 Expression of recombinant protease system in the crF+P transformant lines.

Expression of the heterologous proteins was verified by immunoblotting using suitable antibodies as described in section 2.4. Immunodetection using αHA antibodies showed the successful co-expression of the fusion protein (72 kDa), and the serine protease (27 kDa); additionally, a band of 49 kDa suggests the presence of one of the elements of the fusion protein (cytosine deaminase) (Figure 3.8). Nevertheless, immunodetection using αhGH antibodies shows no trace of the second subunit of the fusion protein. The human growth hormone should be detected at 22 kDa as seen in the cell line used as positive control (Figure 3.8). The fact the some processed CodA is observed suggested that proteolytic processing does take place. The absence of hGH could be relate to instability as a result of the extra amino acid left at the N-terminal after processing.

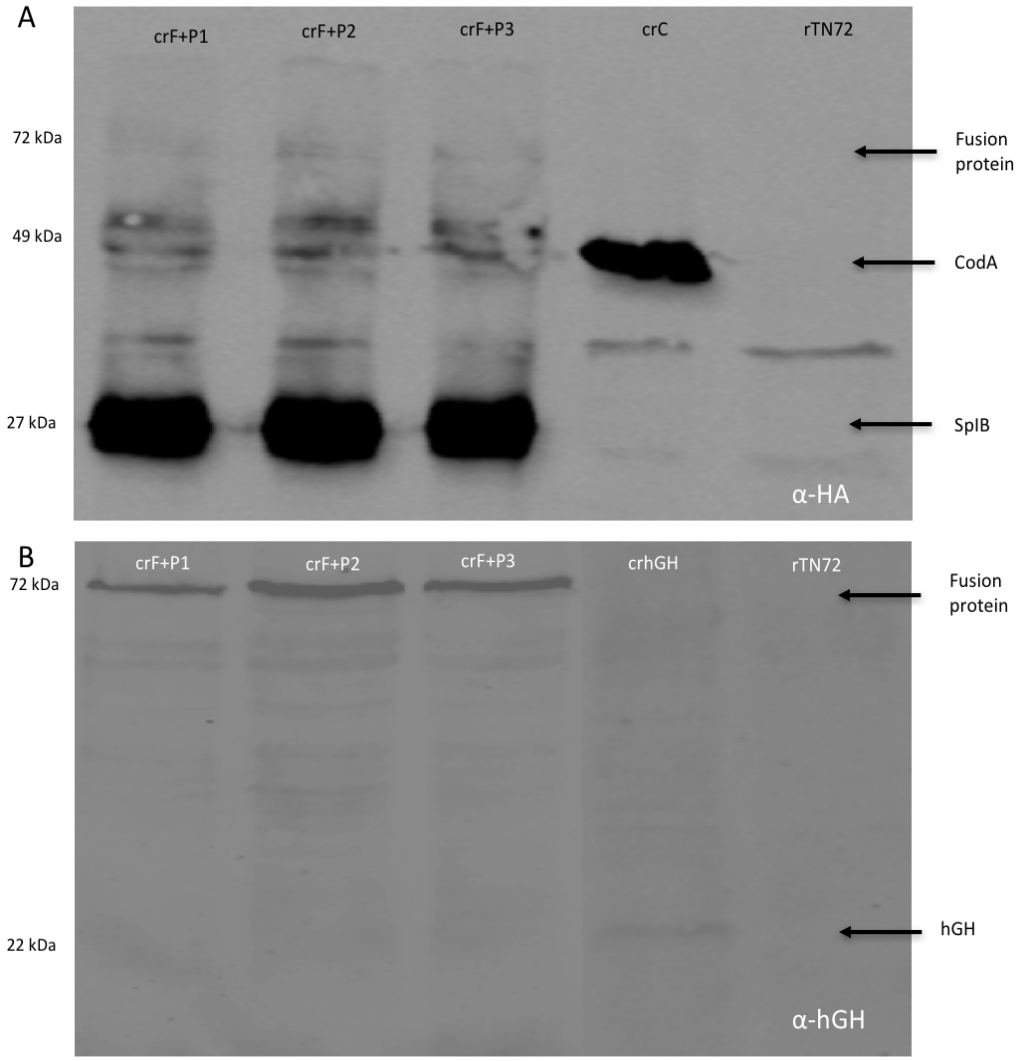


Figure 3.8: Western blot analysis of the transgenic lines expressing the protease system. Cell lysates were separated by SDS-PAGE into identical gels, blotted and incubated with appropriate antibodies. A. Detection using αHA antibodies reveals the presence of the fusion protein (72 kDa), the serine protease (27 kDa) and the cytosine deaminase (49 kDa). The crC strain, which expresses only the cytosine deaminase, was used as a positive control. B. Detection using αhGH antibodies shows the expression of the fusion protein (72 kDa). Nonetheless, there are no traces of the hGH (22 kDa) in the crF+P transgenic lines. The crhGH strain, which expresses only the human growth hormone, was used as a positive control. The rTN72 lane corresponds to the negative control, this strain has been transformed with an empty expression cassette.

3.2 A system-based on multiple expression cassettes system

As a second strategy, I chose the one that required fewer assumptions. I opted for the notion of linking together three expression cassettes expressing unrelated sequences. These cassettes were designed so that each gene is under the control of a different promoter 5‘ UTR, while having two versions of the same (*rbcL*) 3‘ UTR. In this way, I ensured the presence of the sequences required for correct processing of genetic information in the chloroplast.

3.2.1 Plasmid design for combinatorial multiple gene expression in the chloroplast of *C. reinhardtii*

I picked three different unrelated coding sequences, which produced different sized proteins. The serine protease (SpIB) from *S. aureus* (Pustelny et al., 2014), that I had previously used to test the strategy for processing polyproteins in the chloroplast. The codon optimized version of the gene (*spIB*) was cloned into the pASapI vector. Therefore, the *atpA* promoter 5‘ UTR drives its expression. The resulting plasmid, named pS, contains an *Mlu*I restriction site that was used for insertion of the next expression cassette (Figure 3.9A) (Appendix A.4).

The second coding sequence is for cystosine deaminase that catalyzes the conversion of 5-fluorocytosine (5-FC) into the toxic 5-fluorouracil (5-FU) (Dubeau et al., 2009), and has been effectively exploited to unravel the relation between nuclear mutations and chloroplast gene expression in *C. reinhardtii* (Young and Purton, 2014). It has also been used to test a codon reassignment strategy that provides informational containment of transgenes within the cell (Young and Purton, 2015). The optimized version of the gene (*codA*) had already been cloned into a variant of the pASapI

vector whereby gene expression is under the control of the *petB* promoter 5‘ UTR, instead of *atpA*. The whole expression cassette was amplified using primers bearing *Mlu*I restriction sites (Figure 3.9B) (Appendix A.3), and was inserted into the pS vector. The resulting vector was named pCS and contained an *Age*I restriction site that was used to add the third expression cassette (Figure 3.9B) (Appendix A.4).

As a third gene, I decided to use a sequence that encodes three epitopes of the Infectious bronchitis virus (IBV) and the β subunit of the cholera toxin virus (CTB). IBV is responsible for causing important losses in the poultry industry. It can elicit several negative physiological conditions like weight loss, nephritis and respiratory diseases as well as reduced and abnormal egg production (Hai-Peng et al., 2012). A multi-epitope containing a B-cell linear epitope, a T-cell linear epitope and a cytotoxic T lymphocyte epitope bound to a carboxyl terminal residue of the N protein of IBV has been found to trigger protection against IBV by eliciting antibodies and activating lymphocytes (Hai-Peng et al., 2012). On the other hand, the β subunit of the CTB virus has been shown to act as an immunologic adjuvant by binding specific receptors on the surface of gut epithelial cells, which function has been successfully demonstrated in *C. reinhardtii* (Sun et al., 2003). In fact, various species of microalga have been suggested as a safe and cheap delivery method for immunization due to various particular advantages such as rapid transformation and proper protein folding (Rosales-Mendoza et al., 2016). The optimized sequence containing the IBV multi-epitope fused to the β subunit of the CTB virus had been previously cloned into the pSRSapI vector and successfully expressed in the chloroplast of *C. reinhardtii* (Rajakumar, 2016). In the pSRSapI vector, gene expression is driven by the *psaA exon 1* promoter 5‘ UTR, while mRNA stability is ensured by using the *rbcL* 3‘ UTR. This whole expression cassette (*ibt-ctb*) was amplified with primers bearing the *Age*I restriction site. (Appendix A.3), and then cloned

into the pCS vector. The resulting plasmid was named pICS (Figure 3.9C) (Appendix A.4), and contained three different coding sequences arranged within their own expression cassettes.

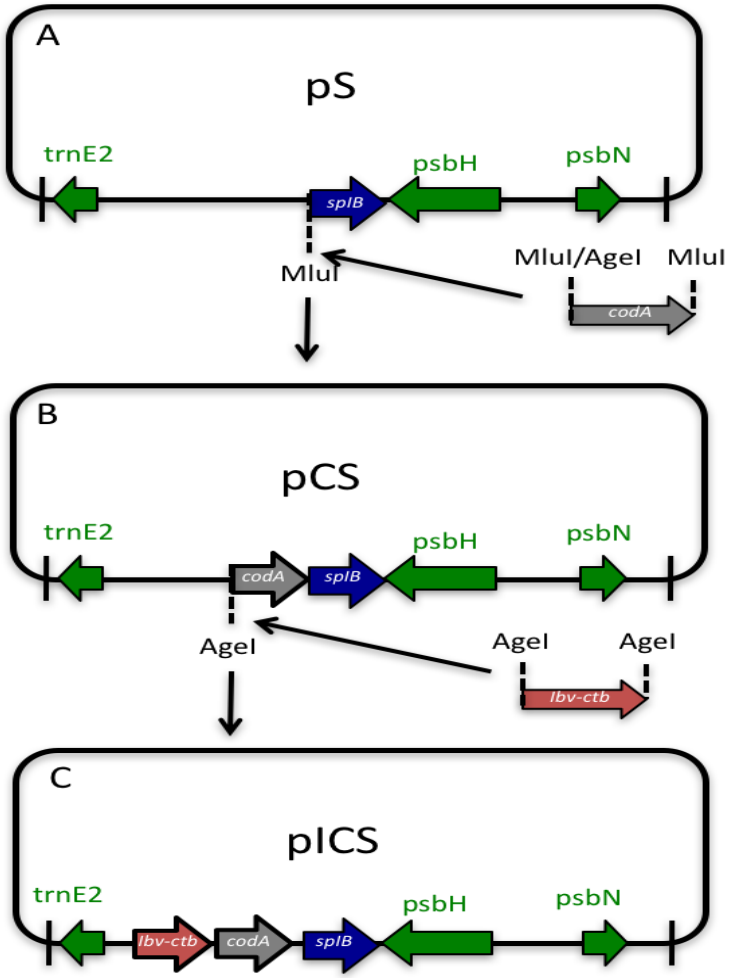


Figure 3.9: Plasmid assembly for combinatorial transformation. The pS plasmid was generated by inserting the serine protease expression cassette (*spIB*, blue) into the pASapI vector. The pS plasmid was digested with *Mlu*I and served as a backbone for the addition of the second expression cassette. The expression cassette bearing the cytosine deaminase gene (*codA*, grey) was amplified with primers bearing the *Mlu*I restriction site (A) and, then, inserted into the pS backbone to generate the pCS plasmid. The latter plasmid was then digested with *Age*I for the addition of the third expression cassette. The expression cassette containing the IBV-CTB (*ibt-ctb*, red) immunogenic construct was amplified with primers carrying the *Age*I restriction site (B), and later ligated to the pCS backbone to yield the pICS plasmid (C). All plasmids were checked by sequencing.

In the final construct, gene expression was driven by three different pro-

moters 5‘ UTR’s, and mRNA stability was provided by the same *rbcL* 3‘ UTR (all genes encoded an HA-Tag at the C terminus for protein detection) (Figure 3.10). The 3‘ UTR used to provide stabilization of the *codA* mRNA was shorter in size than the other two. The *ibv-ctb* and the *spIB* expression cassettes have a 0.4 kb version whereas the *codA* cassette has one that is 0.2 kb. An interesting advantage of this particular arrangement is that it allows us to study potential recombination events between the repeats (subsection 3.2.5).

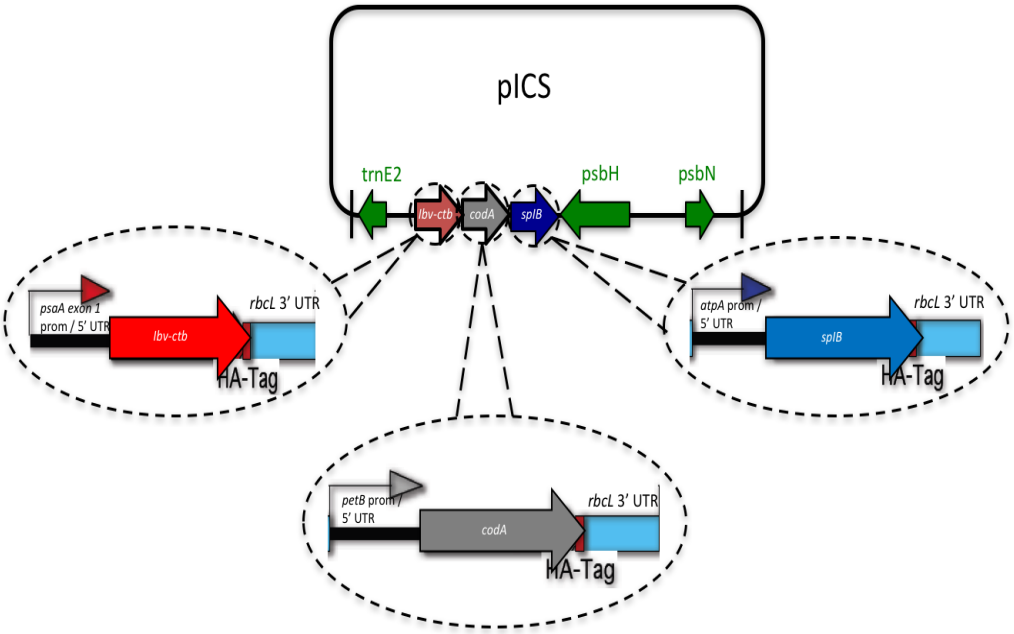


Figure 3.10: The pICS expression vector. The pICS plasmid is based on the pAsapI vector, so it carries an operative version of the *psbH* gene along with the sequences needed for homologous recombination. Nonetheless, this vector bears three different genes of interest (*ibv-ctb, codA, spIB*) engulfed within their own expression cassettes. The *rbclL* 3‘ UTR was used in all of them (a shorter version was used in the *codA* expression cassette); while different promoters / 5‘ UTRs drive gene expression in each cassette. All genes are HA-Tagged for immunodetection.

3.2.2 Transformation of the recipient strain (TN72) with the multi-gene containing plasmid

The pICS plasmid was used for *C. reinhardtii* chloroplast transformation using rapid agitation with glass beads as described in section 2.3. Transformants were re-streaked on high salt minimal medium, and homoplasmicity was confirmed by PCR using three primers (Appendix A.3). As before, primers were designed so they give a different product whether the insertion had been successful or not. Correct integration yielded a 1.2 kb amplicon for the crICS transformants. On the other hand, if homologous recombination had not happened, a 0.85 kb product would result, such as in the case of the host strain TN72 (Figure 3.11). The absence of a band of this size in the transformants indicates that all the copies of the chloroplast genome lack the foreign *aadA* cassette, and that the transgenes have been inserted along with the functional *psbH* gene.

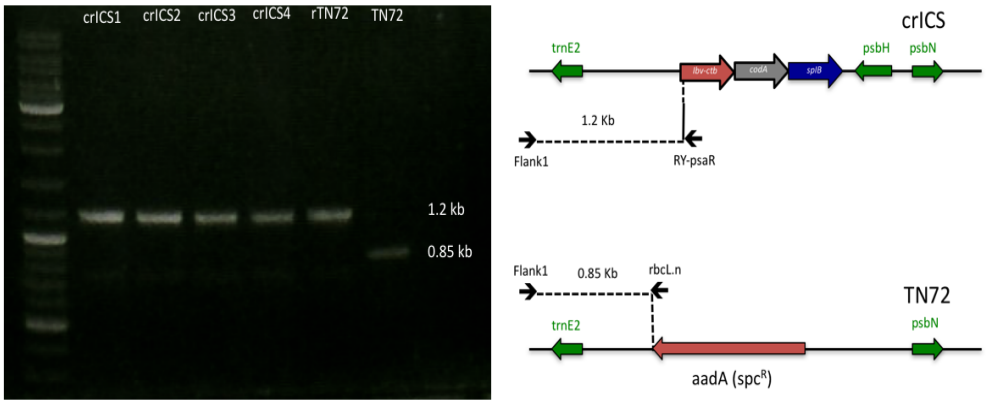


Figure 3.11: Identification of transgenic lines by colony PCR Left-hand side: Agarose gels showing the PCR products from transformants lines and controls. Right-hand side: Diagrams illustrating the genetic arrangement of the TN72 genome and the transformants genomes. The Flank1 primer is targeted upstream of the homologous recombination site; whereas, the reverse primers target either a region of the *aadA* cassette or a section of the inserted expression cassette. The three primers were used in a single reaction. The 1.2 kb amplicons (crICS lanes) demonstrates the correct integration of the three expression cassettes corresponding to the immunogenic complex (*ibv-ctb*, red), the cytosine deaminase (*codA*, grey) and the serine protease (*spIB*, blue) genes. The rTN72 lane corresponds to the positive control; this strain has been transformed with an empty expression cassette. The 0.85 kb amplicons correspond to the recipient strain TN72, which was used as a negative control.

3.2.3 Expression of recombinant proteins in the crICS transformant lines

Expression of the heterologous proteins in crude extracts was verified by immunoblotting using antibodies targeted to the HA-tag, as described in section 2.4. The four transformant lines analyzed (crICS1 - 4) show the expression of the three heterologous proteins (Figure 3.12A). The combined expression of the cytosine deaminase (49 kDa), the IBV multiple epitope linked to CTB (37 kDa), and the serine protease (27 kDa) demonstrates that the chloroplast of *C. reinhardtii* could sustain simultaneous heterologous expression. As positive controls, cell lines expressing the individual proteins (CodA, IBV-CTB, SpIB) were used. As a negative control, a strain that has been photosynthetically restored with an empty plasmid (rTN72) was utilized.

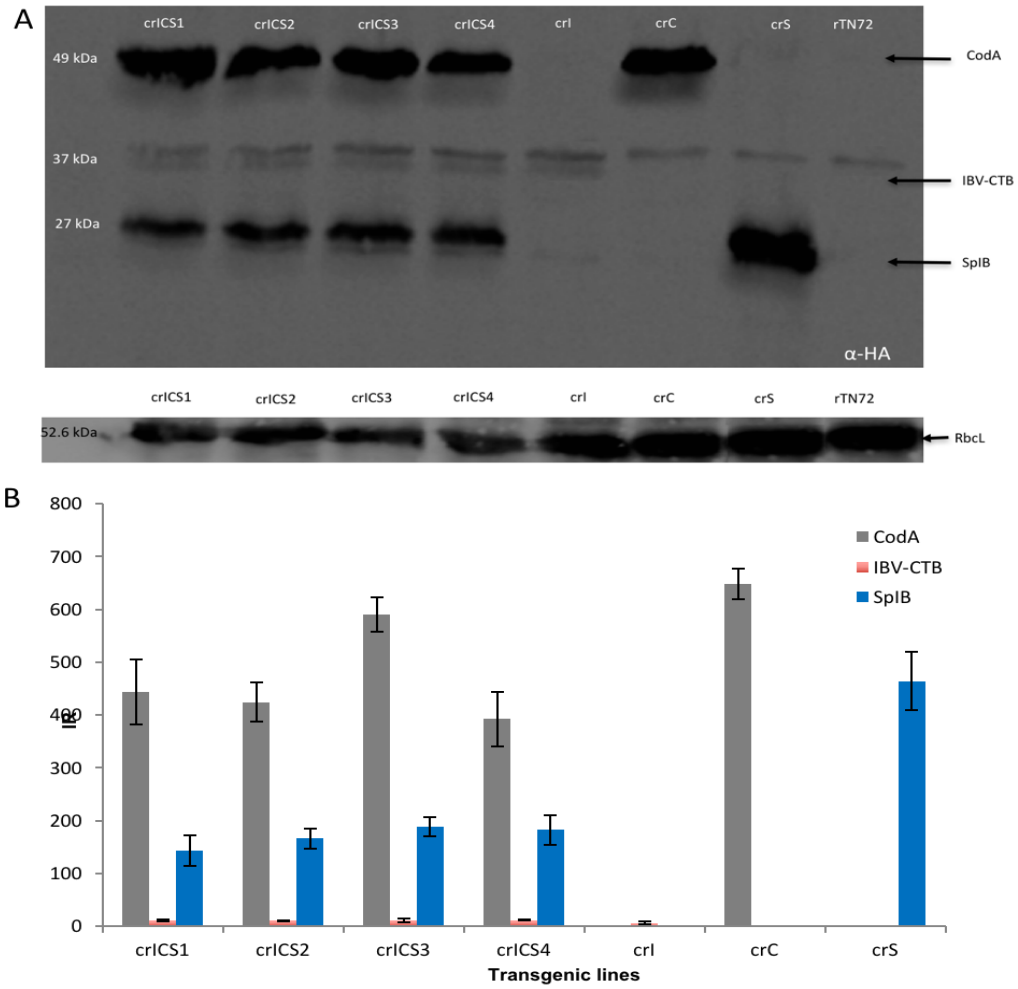


Figure 3.12: Western blot analysis of the transgenic lines containing the heterologous expression cassettes. Cell lysates were separated by SDS-PAGE, blotted and incubated with α-HA-Tag antibodies. A. The simultaneous expression of the heterologous proteins could be detected in the four studied transformants (crICS1 - 4). The CodA, IBV-CTB, and SpIB proteins could be observed at 49 kDa, 37 kDa and 27 kDa, respectively. Cell lines expressing these proteins individually (crI, crC, crS) were utilized as positive controls; the restored TN72 was employed as negative control. The large subunit of the Rubisco enzyme (52.6 kDa) was used as a loading control. An identical gel was probed and blotted with antibodies against the RbcL. B. Quantification of the heterologous proteins based on the infrared signal (IR) of their bands on the western blot assays. Expression of the cytosine deaminase (CodA) does not vary importantly between three cell lines (crICS 1 - 2 - 4). Nonetheless, the crICS3 shows higher levels of these proteins. On the other hand, the four cell lines indicated a similar expression of the IBT-CTB multi-epitope and the serine protease (SpIB). Nevertheless, heterologous proteins produced separately (in the strains crI, crC and crS) seem to have higher protein expression than the ones produced alongside others.

3.2.4 Comparing protein expression levels

The Odyssey Infrared Imaging System was used to calculate the relative bands intensities produced by the recombinant proteins. Samples along with controls were prepared as detailed in section 2.4, and then run for Western blot analysis. The quantitative examination was performed using LiCor system (Appendix A.5). The levels of protein, within the same strain (crICS), varied among the three polypeptides tested. The serine protease and the cytosine deaminase were substantially more abundant than the vaccine complex. This characteristic was observed in the four transformants tested. However, the crICS3 showed higher levels of the cytosine deaminase (Figure 3.12B). Nevertheless, if protein levels are compared to the strains expressing one heterologous polypeptide at a time, the differences seem representative. For instance, cytosine deaminase levels measured in the crC strain were higher in comparison to the levels obtains in the crICS strains. Similarly, the levels of the serine protease appeared to be higher in the crS strain than in the crICS strains (Figure 3.12B). The vaccine complex, contrarily, showed similar levels of production when expressed individually or along with the other transgenes. These outcomes appear to be related to the heteroplasmic state of the transgenes as discussed below, which will be revealed when exposing the cells to 5-FC.

3.2.5 Testing multiple-gene construct stability when exposed to toxic conditions

The crICS strains express simultaneously three recombinant proteins (Figure 3.12). From these I have seen that the serine protease works partially in an *in vivo* system to process polyproteins (section 3.1). On the other hand, the *C. reinhardtii*-produced IBV multi-epitope has been successful in generating an immune response in chickens that have been fed

with such crude extracts (Rajakumar, 2016). However, it is not possible to rapidly test the functionality of the multi-epitope produced in the crICS strain. On the contrary, it is straightforward to test the CodA enzyme *in vivo*. As mentioned earlier, this enzyme is able to convert a harmless compound (5-FC) into a toxic one (5-FU) that affects the cell's viability. Hence, I decided to subject one of the transgenic lines (crICS3) to 5-FC treatment, the test was carried out by plating in 5-FC-containing media as described in section 2.5. As a positive control I used a strain expressing CodA alone, and as a negative one I used a mutant strain transformed with an empty vector (rTN72). Additionally, I grew the strains on normal TAP plates without any drug. The plates were left 10 days for incubation. Within the first four days cell growth occurred as expected. Colonies of *C. reinhardtii* appeared in all plates that do not contain 5-FC. Similarly, colonies of the rTN72 strain were observed on the plates containing 5-FC, this is because this strain lacks the enzyme to process 5-FC into 5-FU (Figure 3.13A). In contrast, colonies from the crICS3 and the crC strains could not grow, which demonstrates that the CodA enzyme is functional in these cell lines (Figure 3.13A) . There were no visible differences between the positive control and the tested strain, which reveals that the action of the enzyme seems not to be altered whether the colonies are expressing only one heterologous enzyme or three of them. However, after a few more days of incubation colonies from the crICS3 strain started to appear in the plate containing 5-FC. This do not happen with the crC cells plated in 5-FC, and in the other controls cell growth continued as normal (Figure 3.13B) .

It seems likely that some genetic rearrangements in the plastome influenced negatively the expression of the cytosine deaminase enzyme. As the positive control (crC) appeared to be genetically stable for CodA expression, this phenomenon could be related to the arrangement of the heterologous expression cassettes localized in tandem upstream *psbH*. In order to detect

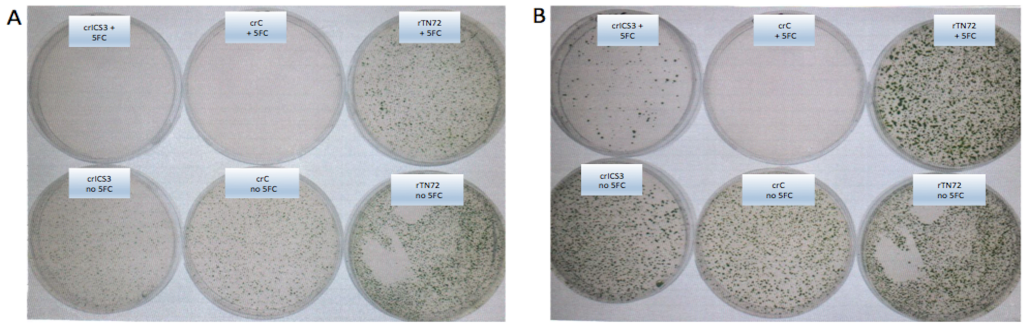


Figure 3.13: Sensitivity of the crICS3 strain to 5-fluorocytosine. Cells were adjusted using optical density and plated on TAP agar containing 5-FC or not. A strain expressing the cytosine deaminase (crC) alone, and the restored TN72 (rTN72) strain were used as positive and negative control, respectively. A. After 4 days of incubation, cells producing the cytosine deaminase were not able to grow in the presence of 5-FC. B. After 10 days of incubation, colonies from the crICS3 strain began to appear, which demonstrates that the cells went through some genetic rearrangement that influenced negatively the function of the cytosine deaminase. In contrast, the positive control plate remained in the same conditions, and in the negative controls cell growth was maintained as expected.

any possible genetic alteration that affected the CodA enzyme, six resistant colonies were re-streaked in 5-FC-containing medium to be further analyzed by PCR. To determine if the *codA* gene was still integrated into the plastome, two primers were designed targeting the flanking regions of the *codA* expression cassette (Figure 3.14A) (Appendix A.3). The crICS3 strain (positive control) yielded the expected amplicon of 2.4 kb, which corroborates the presence of the *codA* expression cassette. On the other hand, the analyzed strains (crICS3 1-6) did not yield any amplicons, which suggests that the expression cassette is missing in these strains explaining why colonies were able to grow in 5-FC containing media (Figure 3.14A). However, since the reaction did not produce any product it was necessary to confirm the presence of the other two transgenes (*ibv-ctb, spIB*). For this purpose, a forward primer targeting the IBV-CTB multi-epitope was used along with a reverse primer directed to the flanking region upstream the transgenic cassettes (Appendix A.3). As can be observed in Figure 3.14B, had the *codA*

cassette been the only one excised, a band of 2.8 kb would have confirmed the presence of the serine protease (*spIB*) expression cassette, but instead a band of 1.0 kb was observed in all colonies. This outcome can be explained if both the cytosine deaminase and the serine protease containing-cassettes were excised. This means that all six colonies (crICS31-6) have lost both cassettes while maintaining the one containing the IBV-CTB multi-epitope (Figure 3.14B). One observation is that the positive control (crICS3) also yielded a 1.0 kb band, Although, a band of around 5.0 kb (which would include the *codA* and the *spIB* expression cassettes) was expected. Apparently, some of the 80 copies of the plastome have lost both cassettes without being subjected to 5-FC treatment. Thus, the amplification of the smaller portion (1.0 kb) was favored.

In normal conditions the state of the three transgenes is heteroplasmic. Namely, in the crICS strains, from the around 80 copies of the genome some have the three transgenes upstream *psbH* (Figure 3.14A) (Figure 3.15), whereas others have only the *ibv-ctb* gene (Figure 3.14B) (Figure 3.15), which show that active recombination occurs between the 0.4 kb *rbcL* 3‘ UTRs. The colonies that appeared after 5-FC treatment showed that they no longer have the *codA* and the *spIB* genes in their plastomes (Figure 3.14), (Figure 3.15). Recombination between the *rbcL* repeats was favored once 5-FC was used as a selection. In this way, *codA* could be excised from all the plastomes, and colonies could grow in the presence of the drug. The toxic environment could have triggered recombination between the *rbcL* repeats in the genomes that still contain the *codA* and *spIB* genes. Alternatively, since 5-FC treatment imposes a negative selection for the genomes that still have *codA*, replication of the ones containing only the *ibv-ctb* upstream *psbH* would be favored. In this way, cells are no longer able to process the drug and colonies are able to grow. Furthermore, the heteroplasmic condition of these genes could explain the variation observed in recombinant

protein production among transformants, and provide an explanation for the reduced accumulation of heterologous proteins in comparison to the controls (Figure 3.12).

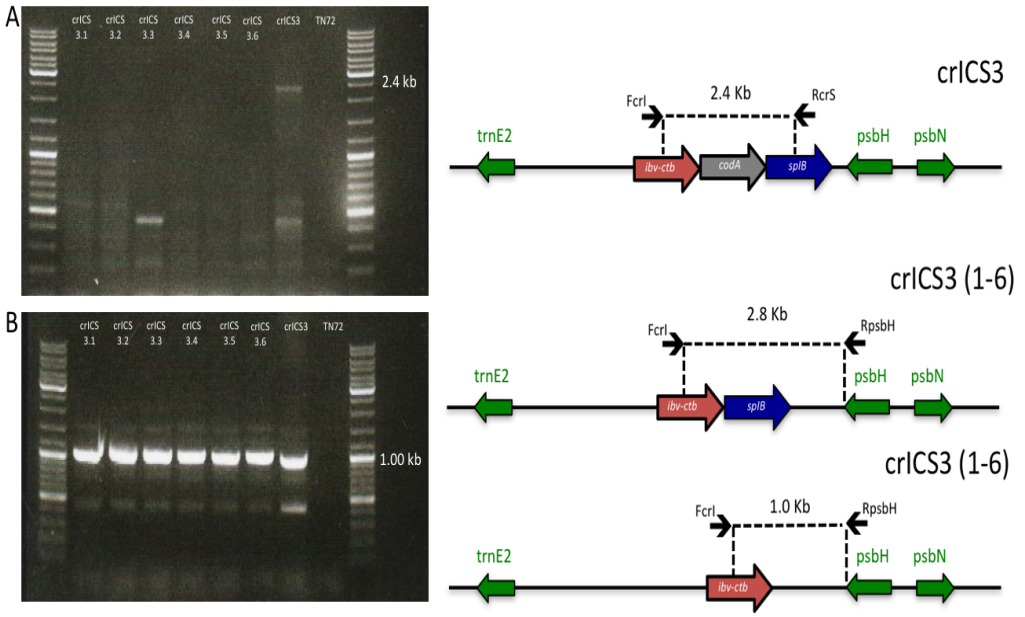


Figure 3.14: Identification of genetic rearrangements in the crICS3 strain after 5-FC treatment by colony PCR. Left-hand side: Agarose gels showing the PCR products from transformants lines and controls. Right-hand side: Diagrams illustrating the genetic arrangement of the crICS3 genome and the transformants genomes. A. The FcrI primer is targeted to the *ibv-ctb* multi-epitope; whereas, the reverse primer is targeted to the serine protease gene. The 2.4 kb band (crICS3 lane) demonstrates the correct integration of the expression cassette bearing the cytosine deaminase (grey). Conversely, the lack of PCR products in the strains treated with 5-FC (crICS 3.1 - 3.6) shows that this cassette has been excised from these cell lines. B. To confirm the state of the serine protease cassette (blue) in the strains treated with 5-FC, a new PCR reaction was carried out using the same FcrI and a reverse primer directed to the *pbsH* gene. The 1.0 kb band demonstrates that in the crICS 3.1 - 3.6 strains both the *codA* and the *spIB* cassettes have been excised from the plastomes. Had the *spIB* cassette remained integrated into the plastome, the PCR reaction would have produced a 2.8 kb band. Using this combination of primers, the crICS3 strain should have yielded a 5.0 kb band, but instead a 1.0 kb was obtained, which demonstrates the genetic arrangement has not reached homoplasmicity (see main text for details). The TN72 strain was used as the negative control.

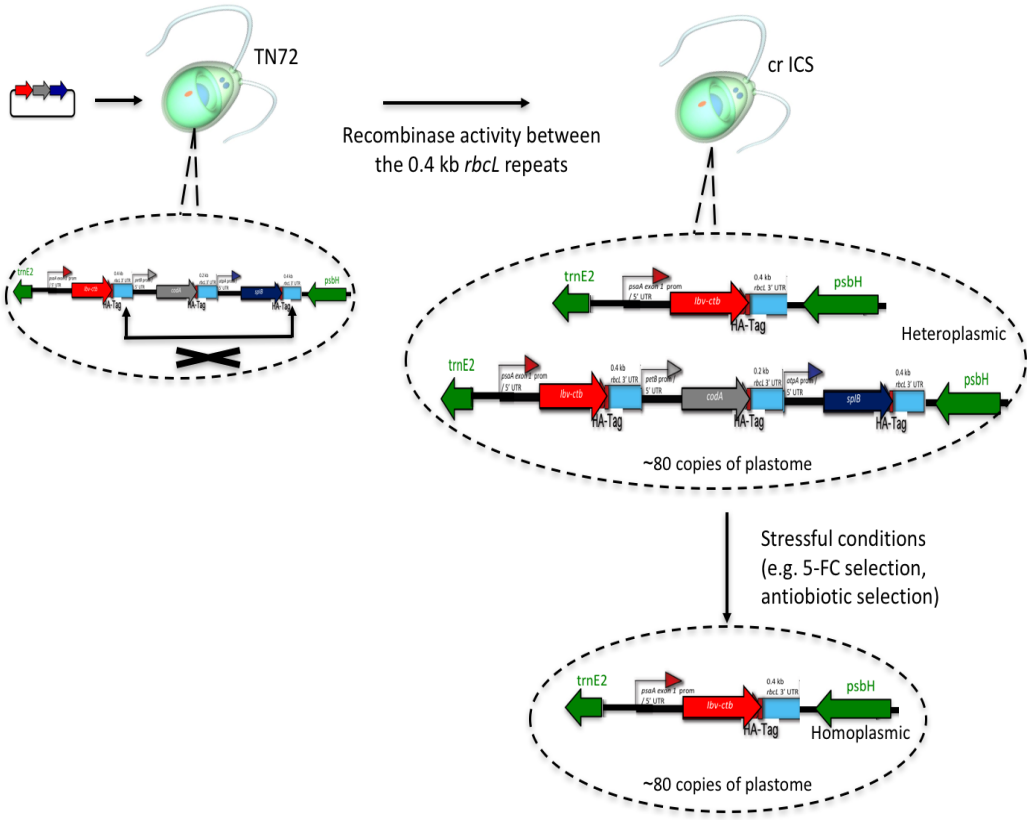


Figure 3.15: Schematic representation of homologous recombination occurring between the *rbcL* repeats. A transformation vector was used to deliver three transgenes in tandem upstream *psbH* into the TN72 strain. In normal conditions, homologous recombination takes places between the 0.4 kb repeats, in the crICS lines some plastomes contain three transgenes (*ibv-ctb, codA, spIB*) and some only one (*ibv-ctb*), namely the character is heteroplasmic. If conditions become stressful (e.g. 5-FC, antibiotic selection) all genomes readily lose the genes flanked by the 0.4 kb 3‘ UTR.

3.3 Serial transformation approach for multiple transgene expression

3.3.1 Plasmid design for serial multiple gene expression in the chloroplast of *C. reinhardtii*

After confirming that the combinatorial approach (various expression cassettes within one transformation plasmid) was successful for generating a strain (crICS) containing three different foreign genes, I considered whether

this could be taken further to generate a transformant with at least five transgenes. Serial transformation was chosen as a methodology. Thus, two additional expression cassettes had to be inserted in the already transplastomic strain (crICS). I decided to target an insertion site containing a gene that is not crucial for algal survival (when grown under controled conditions). Since strains are kept under constant light, genes required to generate chlorophyll in the dark are no longer of vital importance. One of the enzymes involved in this pathway is known as dark-operative protochlorophyllidae (Pchlide) reductase (DPOR). Pchlide is a precursor of chlorophyllide which is reduced to chlorophyllide, which itself is reduced to chlorophyll. The DPOR is a multisubunit enzyme containing three subunits encoded by the chloroplast genes *chlL*, *chlN* and *chlB*. Knockouts of these genes present a typical yellow in the dark phenotype, and the *chlL* gene has been used for studying metabolic aspects of chloroplast biology. (Cheng et al., 2005; Cahoon and Timko, 2000). Therefore, I decided to target the two extra expression cassettes to the site of *chlL*. To achieve this goal, a transformation plasmid had to be created. First, the sequences needed for homologous recombination were amplified, from genomic DNA of the cell wall-less strain *cw-15*, using primers flanking *chlL* (Appendix A.6). The product was cloned into the pJet vector, and the resulting plasmid was named pchlL (Figure 3.16A) (Appendix A.7). It was subsequently digested with *Bst*BI and *Nde*I restriction enzymes to allow the insertion of the fourth expression cassette. I decided to use the aminoglycoside -3-adenyltransferase (*aadA*) expression cassette, which confers resistance to spectinomycin and has been widely used for selection of transformants (Sun et al., 2003; Tan et al., 2007; Goldschmidt-Clermont, 1991). The expression cassette was amplified using primers bearing the *Ac1*I and *Nde*I restriction sites (Appendix A.6). The product was inserted into the pchlL plasmid and the resulting construct was named pAchlL (Figure 3.16B) (Appendix A.7). Subsequently, this plasmid

was digested with *Xma*I , which permitted the incorporation of the second expression cassette. As the fifth transgene, I chose a gene encoding an endolysin (Cpl-1) that targets the bacterial pathogen *Streptococcus pneumonia*. Delivering this endolysin into mice has proved to rescue individuals affected with fatal pneumococcal pneumoniae, which demonstrates the potential of this enzyme to be used as an alternative to traditional antibiotics (Doehn et al., 2013; Witzenrath et al., 2009). *C. reinhardtii* has been suggested as a platform for producing therapeutic proteins given the low cost of cultivation and the fact that *C. reinhardtii* has generally been recognised as safe (Rosales-Mendoza et al., 2016; Jones et al., 2013; Mayfield et al., 2007). In fact, recent research has demonstrated that *C. reinhardtii* crude extracts, expressing the Cpl-1 endolysin, are effective in lysing bacterial culture *in vitro* (Stoffels et al., 2017). Hence, a whole expression cassette containing the Cpl-1-encoding gene was amplified using primers carrying the *Xma*I restriction site (Appendix A.6). The resulting amplicon was cloned into the pAchlL plasmid to yield the pAPchlL construct (Figure 3.16C)(Appendix A.7) that was used to transform the crICS strain of *C. reinhardtii*.

The final construct consisted of two expression cassettes containing the aminoglycoside -3-adenyltransferase (*aadA*) gene, and the gene encoding Cpl-1 endolysin, which were under control of the *atpA* promoter 5‘ UTR and the *psaA exon 1* promoter 5‘ UTR, respectively. Both cassettes bear the *rbcL* 3‘ UTR (Figure 3.17). These elements were flanked by the sequences needed for the integration of the cassette via homologous recombination (Figure 3.17). This transformation plasmid was used to disruption the *chlL* gene by the insertion of two additional transgenes, selection was based on resistance to spectinomycin (Figure 3.18).

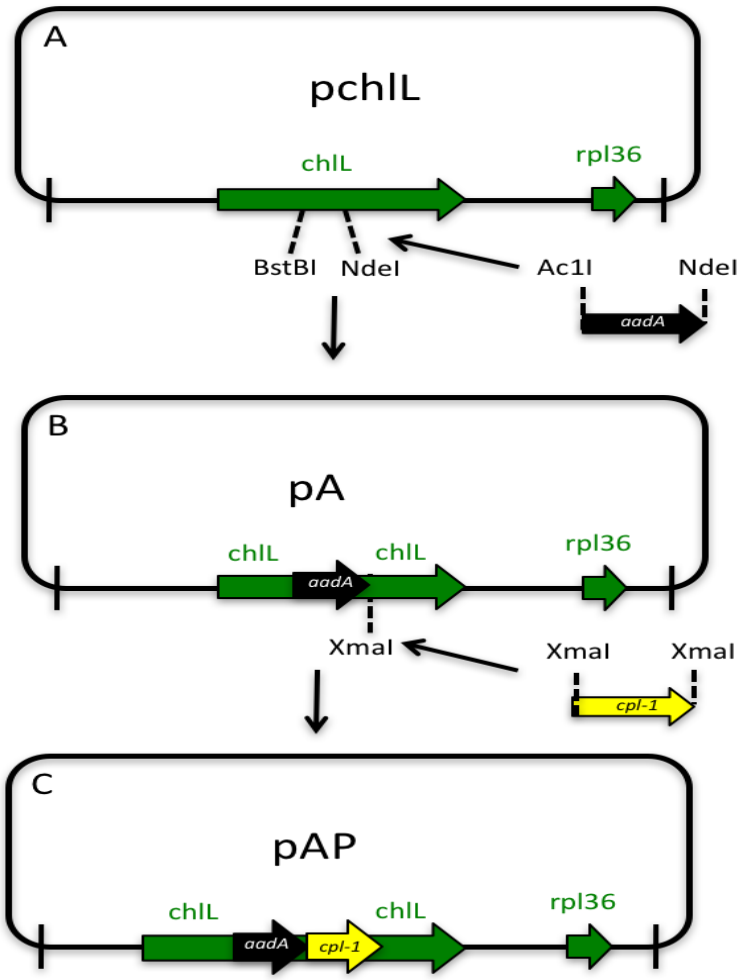


Figure 3.16: Plasmid assembly for serial transformation. The pchlL plasmid was generated by introducing the sequences, belonging to the *chlL* gene, into the commercial plasmid pJet. The pchlL plasmid served as a backbone for the insertion of the *aadA* expression cassette (black), after being digested with the *BstB*I and *Nde*I restriction enzymes. The selection marker expression cassette was amplified with primers bearing the *Ac1*I and *Nde*I restriction sites for proper integration into the pchlL plasmid A. The resulting plasmid was named pA, and was later digested to serve as a backbone for the endolysin expression cassette (*cpl-1*, yellow), which was amplified using primers bearing *Xma*I restriction sites B. After *Xma*I digestion and ligation the pAP plasmid was created, which bears all the required information to insert two additional transgenes in the crICS line. All plasmids were checked by sequencing.

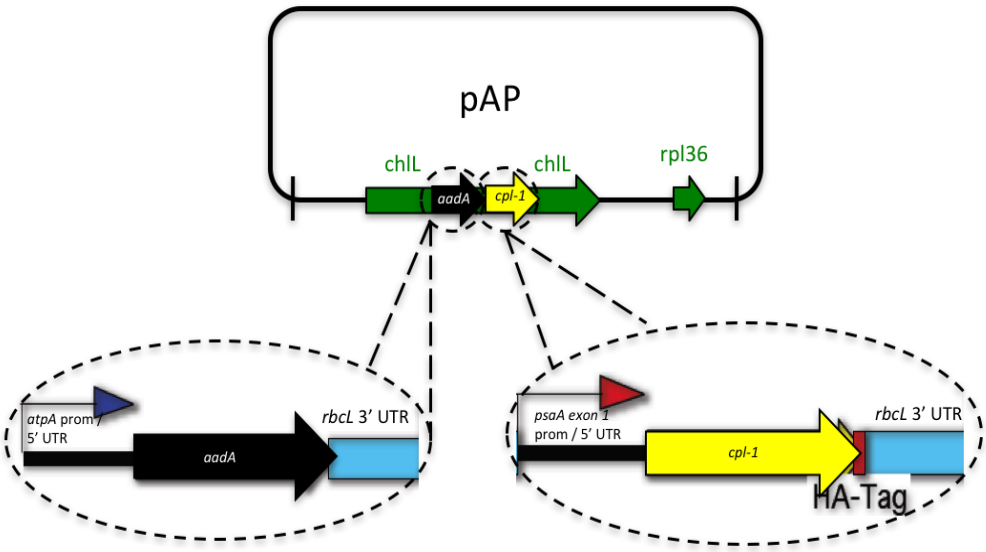


Figure 3.17: The pAP expression vector. The pAP plasmid is based on the pJet commercial plasmid and carries the sequences needed for homologous recombination (chlL, green). Furthermore, this vector bears two different genes of interest (*aadA, cpl-1*) engulfed within their own expression cassettes. The same 3‘ UTR (from *rbclL* gene) was used in both of them, while different promoters 5‘ UTR's drive gene expression. The sequence encoding the endolysin was HA-Tag for immunodetection.

3.3.2 Transformation of the transplastomic strain (crICS) with the multi-gene containing plasmid

The already transplastomic strain crICS3, which is heteroplasmic for a triad of transgenes located upstream *psbH* (Figure 3.15), was used for another round of transformation using the pAPchL vector (Figure 3.18) (section 2.3). After selection for spectinomycin resistance, the surviving colonies were exposed to extra rounds of drug selection. Homoplasmicity was established by PCR (Appendix A.6). Primers were designed so the correct insertion of the cassettes could be checked by the different sizes of the amplicons. Correct integration produced a 1.3 kb amplicon for the the crAPICS transformants, and an amplicon of 1.0 kb for the untransformed strain (Figure 3.19). Therefore, the absence of 1.0 kb for the crAPICS also indicates that all the genomes in the chloroplast have correctly integrated

the expression cassettes. Moreover, since the *chlL* gene is disrupted, no chlorophyll could be produced in these transformants when grown in the absence of light (Appendix A.8). From the several transformants obtained three were chosen for further scrutiny.

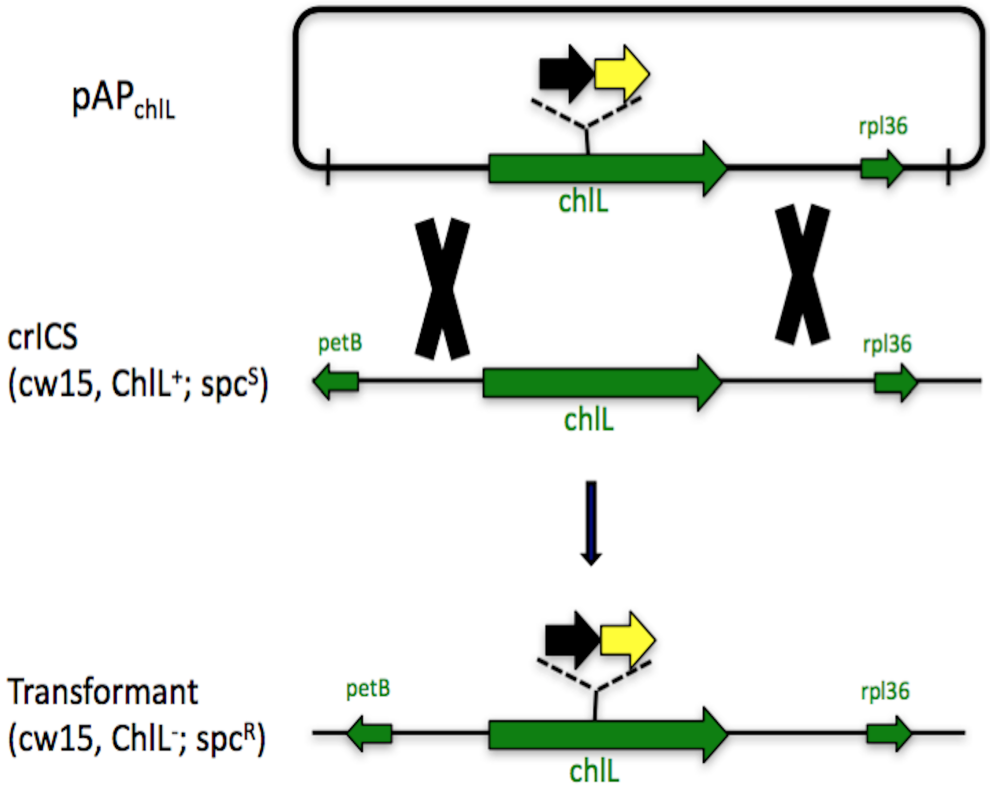


Figure 3.18: Chloroplast serial transformation approach. The pAP vector has been manufactured in a way so it contains sequences encoding biological functions required for integrating transgenes into the plastome. The *aadA* gene, which confers spectinomycin/streptomycin resistance was used for transformant selection (black arrow). The yellow arrow represents the expression cassette containing the gene encoding the endolysin. Both the expression cassettes are flanked by the sequences needed for homologous recombination. In the recipient strain (crICS), the *chlL* gene is undisrupted. Nonetheless, after transformation cells are able to grow in the presence of the mentioned antibiotic and no longer display chlorophyll production in the dark, which is typical of the *chlL* mutants (See main text for details).

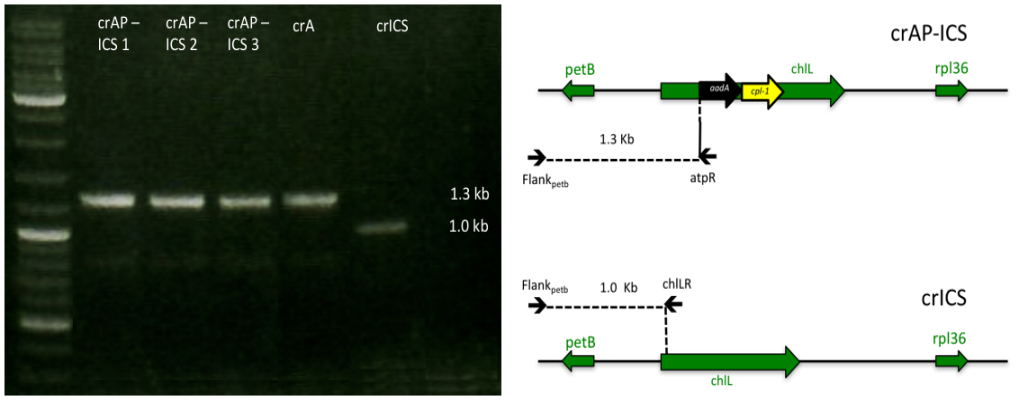


Figure 3.19: Identification of transgenic lines by colony PCR Left-hand side: Agarose gels showing the PCR products from transformants lines and controls. Right-hand side: Diagrams illustrating the genetic arrangement of the crICS genome and the transformants genomes. The Flank$_{petB}$ primer is targeted to the 5‘ upstream of the homologous recombination site; whereas, the reverse primers target either a region of the *aadA* cassette or a section of the intact *chlL* gene. The three primers were used in a single reaction. The 1.3 kb band demonstrates the correct integration of the two expression cassettes corresponding to the selectable marker (*aadA*, black) and the endolysin (*cpl-1*, yellow). The crA lane corresponds to the positive control. This strain has been transformed with a vector containing only the selectable marker. The 1.0 kb band correspond to the recipient strain crICS, which was used as a negative control.

3.3.3 Expression of recombinant proteins in the crAP ICS transformant lines

Immunobloting with the appropriate antibodies, as described in section 2.4, was used to identify the heterologous proteins in the crude extracts of the crAP ICS *C. reinhardtii* cell lines. As positive controls, cell lines expressing the heterologous proteins collectively (crICS3) or individually (crP, crC, crI, crS) were used. As a negative control, a strain that has been photosynthetically restored with an empty plasmid (rTN72) was utilized. The three strains scrutinized (crAP ICS 1-3) show the presence, at 43 kDa, of the endolysin targeting *S. pneumoniae* (Figure 3.20A); however, two of the original three heterologous proteins were lost; it can be appreciated that the cytosine deaminase as well as the serine protease have disappeared from these transformants. The expression of the vaccine could not be observed properly due

to the closeness in size of the endolysin. Nonetheless, further PCR analyses confirmed the presence of the gene in the transgenic lines (Figure 3.21). The recipient strain crICS3 has some plastomes containing the *ibv-ctb-codA-spIB* genes upstream *psbH*, and some others with *ibv-ctb* uniquely (Figure 3.15). After antibiotic selection, recovered colonies no longer show traces of *codA* and *spIB*. Analogous to what happened after 5-FC treatment, once the conditions become stressful recombination between the 0.4 kb *rbcL* repeats is favored and thus *codA* and *spIB* are excised from all plastomes (Figure 3.15). Despite the loss of the heterologus proteins, the crAP ICS strains do indeed express multiple transgenes at the same time. Namely, these cell lines have within their plastome the vaccine, the endolysin and the *aadA* coding sequences.

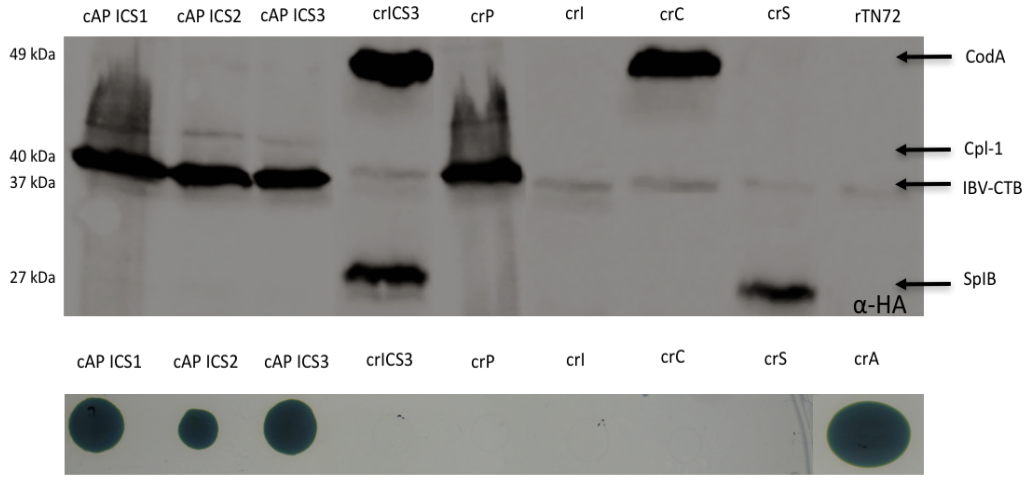


Figure 3.20: Analysis of the transgenic lines containing the additional heterologous expression cassettes. A. Western blot analysis of cell lysates that were separated by SDS-PAGE, blotted and incubated with α-HA-Tag antibodies. A. The simultaneous expression of the additional heterologous proteins could be detected in the three studied transformants (crAP ICS1 - 3). The endolysin (Cpl-1) could be observed at 40 kDa; nevertheless, the CodA and SpIB proteins could not be observed at 49 kDa, 37 kDa and 27 kDa, respectively. Cell lines expressing the studied proteins individually (crP, crI, crC, crS) were utilized as positive controls. The restored TN72 was employed as negative control. B. Growth tests. Cells were spotted on TAP agar containing spectynomicin. Plates were incubated at 25°C under 50 μE m^{-2} s^{-1} light for 5 days. Resistance to the drug is associated with the presence of the *aadA* gene as can be observed in lanes crAP ICS1 - 3. A cell line containing only the selectable marker was used as a positive control (crA).

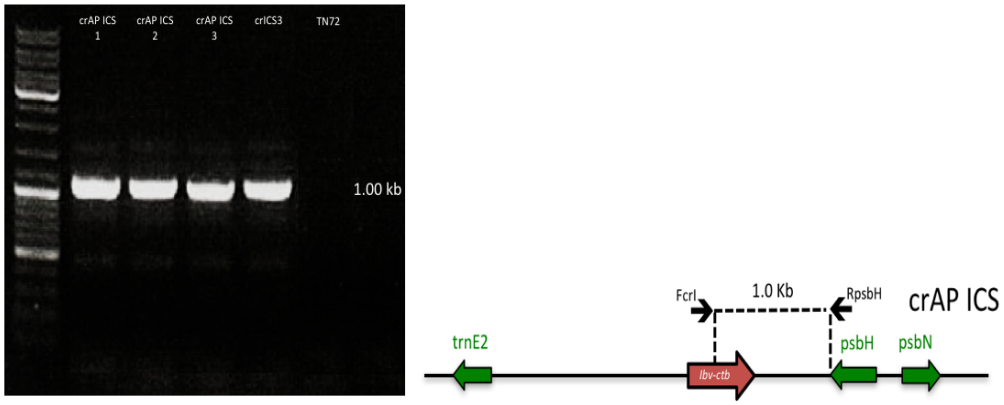


Figure 3.21: Identification of genetic rearrangements in the crAP ICS strains after selection on antibiotics by colony PCR. Left-hand side: Agarose gels showing the PCR products from transformants lines and controls. Right-hand side: Diagrams illustrating the genetic arrangement of the crAP ICS genome. To confirm the state of the serine protease and the cytosine deaminase-containing cassettes in the strains recovered after selection, a PCR reaction was carried out using a primer targeting the vaccine sequence (red) and a reverse primer directed to the *pbsH* gene. The 1.0 kb amplicon demonstrates that in the crAP ICS 1 - 3 strains the only expression cassettes that is firmly located correspond to the one carrying the sequence encoding the vaccine. The TN72 strain was used as the negative control.

3.4 Effects on growth of multiple transgene expression

The crICS3 strain, which expresses the IBV-CTB epitope, the CodA and the SpIB enzymes, and the crAP ICS strain, which expresses the Cpl-1, the AadA and the IBV-CTB epitope were evaluated in order to determine the impact on specific growth rate and maximum cell density. The cells were grown in mixotrophic conditions. The values are shown in Table 3.1, and Figure 3.22 shows the growth curve for each strain. As shown in Table 3.1, the specific growth rate and the maximum cell density values were marginally lower than those obtained from the control.

Table 3.1: Specific growth rate and maximum cell density obtained for the strains expressing multiple transgenes. Cells were grown in 1 L flasks at 120 rpm of agitation and at 50 μE m^{-2} s^{-1}. Specific growth rates at stationary phase are an average from three independent measurements.

Strain	Specific growth rate (min^{-1})	Maximun cell density OD_{750}
crICS3	0.048 +/- 0.006	3.05 +/- 0.04
crAP ICS1	0.044 +/- 0.004	3.00 +/- 0.04
rTN72	0.051 +/- 0.004	4.06 +/- 0.09

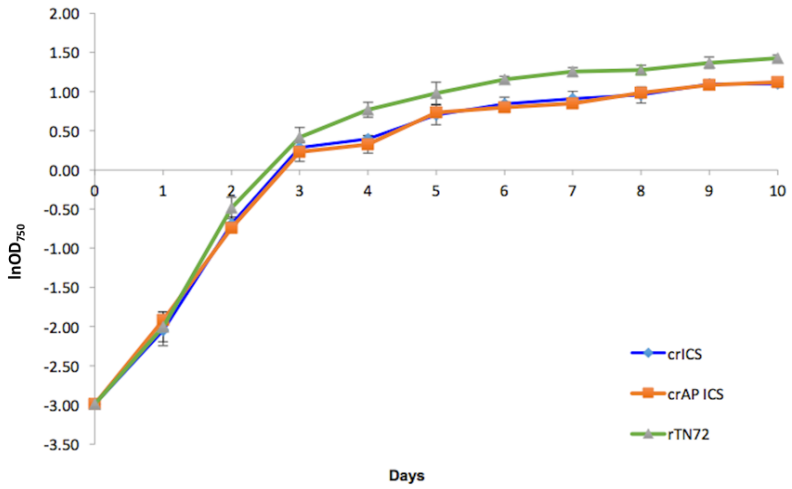


Figure 3.22: Effects on culture growth of multiple transgene expression from the plastome. Expression of three heterologous proteins does have a minor effect on growth. The transgenes-containing strains crICS3, crAP ICS1 and the control rTN72 were grown mixotrophically. The data presented are averages of three independent measurements.

3.5 Discussion

The potential of microalgae for metabolic engineering lies principally on the algae's abilities to grow under control conditions, as well as on the capacities to engineer all subcellular compartments. Economic benefits, in particular, could be drawn from the photosynthetic production of relevant metabolites. The chloroplast possesses many important metabolic pathways essential for cell fitness such as photosynthesis, hydrogen production or formation of precursors of triacylglycerol (TAG) (Dubini and Ghirardi, 2015; Fan et al., 2011). There are plenty of prospective metabolites the production of which could be enhanced by engineering. Additionally, many novel molecules could be derived from these native pathways. However, in order to get to such a level of sophistication is necessary to have several transgenes expressed simultaneously. Currently, the tools available for transplastomic engineering of *C. reinhardtii* are not developed enough for such tasks. Major impediments have been the shortage of tools permitting the synchronised production of multiple heterologous proteins, along with an insufficient number of selectable markers, or the lack of straightforward methods to recycle them. Therefore, one of the biggest challenges remains the development of chloroplast engineering tools that require no usage of antibiotic selection and require a minimal number of transformation events for inserting foreign metabolic information. As has been reviewed already and is diagrammatically represented in Figure 3.3, a convenient method for stable expression of heterologous proteins has been put forward. This technique relies on photoautotrophic restoration as a mechanism for transformant selection, positive transformants carry only the gene of interest as imported information. This chapter presented some attempts to demonstrate that this technology could be scaled for the insertion and expression of multiple transgenes whether relying on proteolytic processes or on the presence of ge-

netic elements encoding biological functions. Additionally, I envisaged the design and utilization of supplementary devices that allow the coupling of additional genetic sequences in an already transplastomic strain.

3.5.1 Protease system

In many well-characterized expression platforms (e.g. *E. coli*, *S. cerevisae*) techniques have been developed to achieve a coordinated simultaneous expression of genes for several enzymes involved in important pathways (Bieniossek et al., 2012; Vijayachandran et al., 2011). However, approaches to perform such research are not available for the chloroplast of *C. reinhardtii*. I outlined a system designed to express and proteolytically process a synthetic fusion protein made up of two units. This method uses a two-component device for delivering the information, consisting of the fusion protein and the enzyme in charge of its processing. N^{α} terminal acetylation has been noticed to be operational in the stroma of *C. reinhardtii* chloroplasts, which indicates that post translational modifications exist in the organelle of this green alga. In addition, proteolytic events have also been noticed. These involve the cleavage of N-terminal sequences from imported proteins (Bienvenut et al., 2011). Interestingly, proteolytic events have been seen to be involved in the processing of some viruses proteomes. For instance, the Middle East respiratory syndrome coronavirus (MERS-CoV) has been shown to carry at least two proteases to process the polyproteins encoded in its RNA genome (Yang et al., 2014b). This principle has been utilized for synthetic biology approaches. To attain proper balances in protein expression, polyproteins have been designed which could be processed by specific proteases (Kim et al., 2011; Chen et al., 2010). I wondered if this idea could be applied to develop an expression system that parallels this post-translational processing of polyproteins. For this purpose, I chose a commercialized protease

from *Staphylococcus aureus* known as WELQut. (www.thermofisher.com). It is highly specific and active across a broad range of temperature and pH. It cuts after the glutamine of its recognition site (WELQX). Consequently, after digestion the first unit of our fusion protein will carry a motif of four amino acids at the C-terminal site, which might or might not be detrimental for protein stability and function. As modules for the polyprotein, I opted for two proteins that were robustly-expressed in the chloroplast of *C. reinhardtii*. The optimized version of the *E. coli* cytosine deaminase gene was linked, by the protease recognition site, to the sequence encoding the human growth hormone. The genes encoding the fusion protein and the protease were put under the control of different promoters 5‘ UTRs and the same 3‘ UTR. Both expression cassettes were positioned in the same plasmid (Figure 3.6). Three healthy transformants were recovered, and PCR analyses confirmed the proper insertion into the plastome (Figure 3.7). Lysates extracted from cell lines expressing the protease system were analysed by SDS gel electrophoresis and further Western blot analyses. At 72 kDa the fusion protein could be observed. Likewise, at 27 kDa the protease can be identified (Figure 3.8). While I expected some undigested polyprotein, I anticipated that evidence of the two products of proteolysis could be detected. As can be observed in Figure 3.8A at 49 kDa some traces of the processed cytosine deaminase could be identified. Nevertheless, its accumulation level is importantly reduced if compared to the control, which seems reasonable since some of the cytosine deaminase remains as the polyprotein. However, this finding confirms that the protease does indeed reach its recognition site. On the other hand, no traces of the human growth hormone could be detected. The proline amino acid that remains on the N-terminal region of the human growth hormone could be one reason that might explain the lack of detectable processed protein. In fact, proteolytic degradation is one of the many factors that might interfere with recombinant protein production in

the chloroplast (Rosales-Mendoza et al., 2012). Chloroplast proteases are homologous to bacterial proteases. Their intrinsic role is to remove faulty or atypical proteins but also they have been implicated in chloroplast gene expression and in plastid signalling with the nucleus (Olinares et al., 2011; Adam et al., 2006). Nevertheless, I have seen that the system works partially, and that this type of post-translational processing depends strongly on the properties of the proteins modules. In other words, the processed protein's capacity to fold properly is correlated with its ability to carry some extra amino acids on either terminal sites. This fact imposes limitations for metabolic engineering ambitions, since not all proteins of interest might be suitable. Moreover, in this system the genes of interest are expressed from the same promoter 5‘ UTR, which would have produced, if working properly, the same amount of both proteins. This complicates any attempt to balance the stoichiometry of co-expressed proteins. Furthermore, it will be necessary to always insert the gene for the serine protease alongside the genes of interest, which increases the amount of foreign information added to the plastome. Despite the limited results obtained from this experiments, some relevant information could still be drawn. For instance, I observed that the transformation plasmid used for photosynthetic restoration could handle the presence of more than one expression cassette, which opened the possibility for a combinatorial serial approach that involved the stacking of various cassettes in different devices, and their further insertion and simultaneous expression in the chloroplast.

3.5.2 Combinatorial approach

Using a combinatorial strategy I created a device that permitted the simultaneous integration of three genes from diverse organisms into the plastome. The genetic elements employed gave robust simultaneous expression

in the organelle of a vaccine against avian viruses, an enzyme from *E. coli* and one from *S. aureus*. As was demonstrated in the last section, stacking two expression cassettes in one plasmid (Figure 3.6) can be used as a strategy to express two proteins at the same time (Figure 3.8). Consequently, I decided to use the same approach and test if the plasmid could include a third expression cassette. Transformation was successful and four transformants were chosen for analysis, the PCR profile (Figure 3.11) indicated that the three foreign sequences were correctly integrated into the plastome. Furthermore, a normal cloning plasmid is around 6.0 kb (Appendix C.2), whereas the pICS device consists of around 11 kb (Appendix A.4C). The next step was to detect the simultaneous presence of the heterologous proteins. Western blot analyses revealed that all the transformants were able to produce the heterologous proteins (Figure 3.12A). Showing that the vaccine, the cytosine deaminase and the serine protease could be stacked within the organelle. The infra-red (IR) fluorescence system was employed to assess differences in protein levels among transformants. Some variation was found among the transformants, with crICS3 the cell line that showed the highest levels (Figure 3.12A). Important differences, though, were observed when the transformants were compared to the positive controls. That is to say, when the transgenes are expressed individually their protein production is higher (Figure 3.12B). The standard values for the cystosine deaminase were higher than the ones acquired for the crICS3 cell line (Appendix A.5). Similarly, the average accumulation of the serine protease was lower in the crICS cell lines than in the positive control. On the other hand, accumulation of the vaccine apparently did not vary significantly. All transgenes were stacked within the same expression cassettes as the positive controls, so it could be assumed that accumulation of the heterologous proteins should stay at the same level. The use of appropriate flanking elements assured the proper transcriptional processing of the coding sequences. Moreover, the

assembling of the transgenes within their own expression cassettes sought to avoid any alteration in transcriptional and translational processes of the transgenes. This appears to be related to the heteroplasmic state of the transgene assembly downstream of *psbH* gene (Figure 3.14). That is to say, that the crICS lines cannot accumulate the heterologous proteins at the same levels because some of the plastomes do not contain the *codA* and *spIB* genes. Therefore, having less copies of these trangenes leads to a reduction in protein accumulation. On the other hand, if homoplasmicity would have been achieved, the accumulation of recombinant protein should be similar to those of the controls. In fact, the levels of IBV-CTB from the crICS lines do not seem to vary from the control (Figure 3.12), that is because in crICS all the plastomes carry a copy of *ibv-ctb*. It is clear that in normal conditions recombination is taking place between the *rbcL* 3‘ UTRs. It has been shown that sequences consisting of around 300 bp are able to induce recombination events in *Chlamydomonas* plastids. However, in plants plastids recombination could be triggered by sequences of only 84 bp (Mudd et al., 2014). One observation is that recombinases seem to make use of the 0.4 kb *rbcL* repeats, since the PCR profiles reveal that when recombination happens both *codA* and *spIB* are excised from the plastomes of the crICS3 strains (Figure 3.14B) (Figure 3.15). If recombination had happened between the 0.4 kb repeat (used in the *ibv-ctb* expression cassette) and the 0.2 kb repeat (used in the *codA* expression cassette) a different amplicon would have been obtained due to the presence of *spIB* (Figure 3.14B). Moreover, if *spIB* had not been removed, along with *codA*, from the genomes, a similar protease accumulation to the controls would have been observed. It is clear that this technical complication could be circumvented by utilizing different 3‘ UTRs. Nonetheless, these results show that different versions of the *rbcL* repeat do not promote recombination. The functionality of the cytosine deaminase is straightforward to examine. It involves growing the algae in

the presence of 5-FC, if the cell is able metabolize such component into 5-FU it will no longer be viable and no cell growth would be observed. The heterologous protein-containing cell lines could not form colonies, just as the positive controls, after exposed to 5-FC which demonstrates that the enzyme could effectively deaminate its substrate (Figure 3.13A). Nevertheless, after 10 days of incubation colonies from the crICS3 line started to emerge (Figure 3.13B). Via PCR analyses I scrutinized the genetic arrangements of the transgenes. As could be seen on Figure 3.14, the colonies selected after growth on 5-FC did not show trace of the presence of the expression cassettes as the positive control did (Figure 3.14A). Subsequent PCR tests showed that the only transgene that remained in place was the one containing the vaccine (Figure 3.14B). This genetic readjustment is obviously the why these colonies survived in the presence of 5-FC. The absence of a 5.0 kb band in the crICS3 strain reflects the heteroplasmic conditions of this arrangement. Namely, due to PCR efficiency the amplification of the 1.0 kb amplicon was favoured. These observations confirmed that recombination happens between the two 0.4 kb repeats, and that the 0.2 kb repeats are not involved. As schematically depicted in Figure 3.15, once the the crICS lines are subjected to 5-FC treatment all the genomes readily lose the *codA* and the *spIB* genes that were present in some of the plastomes. This loss allows the colonies to survive in the presence of the drug. Another alternative, is that 5-FC negatively selects the genomes that contain *codA*. Thus, genomes that do not contain the toxic gene have a replicative advantage over the rest. Overall, these results show that simultaneous recombinant protein accumulation is attainable in the chloroplast. Furthermore, the instability observed could be potentially circumvented by diversifying the elements used to drive transgene expression. More importantly, they demonstrate also that the removal of *codA* is possible by setting up stressful conditions. Moreover, not only the cytosine deaminase containing cassette could be excised but the

expression cassette, located to its right hand side, was also eliminated from the plastome (Figure 3.15). This peculiarity could be exploited for designing systems that allow the removal of sequences based on the presence of 5-FC in the medium. For example, if the *aadA* cassette were located downstream the cytosine deaminase it could be removed, using 5-FC treatment, after being used for transformant selection. It was observed, furthermore, that the vaccine expression cassette was left untouched after 5-FC exposure. Analogously, any gene of interest could be located in such position.

One important notion that could be drawn from these results is the feasibility to insert three different transgenes using a unique device and with no antibiotic resistance as a selection marker. This experimental design and set up could be used as a foundation to design, say, innovative devices that ensure multiple gene expression and protein stability regardless of the conditions. To achieve such a task, it will be necessary to use different components in each expression cassette to avoid possible recombination effects. Also, it would be ideal to have a device whereby genes could be inserted directly and with no additional steps of amplification and ligation of expression cassettes. Nevertheless, these results demonstrate that the concept of combinatorial transformation could be effectively exploited for delivering various transgenes into the plastome with only one transformation event.

3.5.3 Serial approach

Thirdly, I described the use of an additional device to incorporate more foreign information in the already transplastomic cell line crICS3. Two more foreign sequences were added in this serial transformation. One of the new sequences I opted for was the *aadA* gene that confers resistance to spectinomycin and streptomycin (Tan et al., 2007; Sun et al., 2003; Goldschmidt-Clermont, 1991). An endolysin targeting *Streptococcus pneumoniae* was

chosen as the additional gene for insertion (Stoffels et al., 2017). It was decided to target these two expression cassettes to the *chlL* region encoding a reductase enzyme involved in the production of chlorophyll precursors in dark conditions. This mechanism allows the algae to assemble the photosynthetic apparatus in the dark. However, such a mechanism is no longer required if the algae are grown in conditions of constant light. Hence, the regions encoding this enzyme could be utilized as target regions for transgene insertion. The three genes involved (*chlL, chlB* and *chlN*) are relatively large and could be replaced with large expression cassettes. Thus, these genes could be considered disposable for strains growing under constant light. The *chlL* gene has been previously targeted to study various aspects of chloroplast biology with successful transformations (Cheng et al., 2005; Cahoon and Timko, 2000). I decided, therefore, to construct a new expression vector based on *chlL*. The vector was engineered with two expression cassettes where gene expression is driven by different promoters 5‘ UTRs, and two different versions of the *rbcL* 3‘ UTR (Figure 3.17). This plasmid was employed to transform the already transplastomic strain crICS3. Consequently, as shown in Figure 3.18, the device allowed the disruption of the *chlL* gene by the two additional transgenes, with selection based on resistance to spectinomycin. Three transformants were recovered in which, as confirmed by PCR, the additional transgenes were successfully inserted as anticipated (Figure 3.19). These strains in principle should have transgenes located at different sites of the plastome. Either one (*ibv-ctb*) or three (*ibv-ctb, codA, spIB*) downstream of *psbH*, and two (*aadA, cpl-1*) at *chlL*. As can be appreciated, the functionality of the fourth transgene is essential since it allowed the generation of strains resistant to the antibiotic. The next phase was to detect the presence of the proteins that were HA-tagged. After Western blot analyses, it was observed that the endolysin was expressed in the same manner as in the controls (Figure 3.20). On the other hand, there are

no traces of accumulation of the cytosine deaminase and the serine protease. This phenomenon could be attributed again to the instability brought in by the presence of the *rbcL* 3‘ UTR repeats used to drive the expression of the three original transgenes. This process parallels that depicted in Figure 3.15. The presence of the vaccine is barely observed in the Western blot analysis since the endolysin band is covering the weak signal from the vaccine. Thus, the same primers used to track genetic modifications in the crICS3 strains, after 5-FC exposure, were employed to see if the expression cassette containing the vaccine was still present in the plastome. As indicated in Figure 3.21, the 1.0 kb band shows that the sequence encoding the vaccines is still localized downstream *psbH*, and confirms the lost of *codA* and *spIB*.

Notably, however, the cell lines recovered after the second round of transformation (crAP ICS1-3) were healthy when grown, which demonstrates the reliability of transplastomic strains to go through additional rounds of transformation. These cell lines do possess three heterologous genes expressed at the same time: the sequence encoding the vaccines, the *aadA* gene and the gene encoding the endolysin. Serial transformation, as demonstrated, represents a plausible strategy for generating strains with various transgenes stacked within the chloroplast. Furthermore, taking into account that photosynthetic restoration was used for inserting the first group of transgenes, antibiotic selection could then be used to increase the amount of foreign information within the cell. Clearly, it would be optimal to have at our disposal a methodology that allow the insertion of heterologous genetic information in an already transplastomic strain without the use of antibiotic markers. Subsequently, I decided to determine the impact on growth of having three extra heterologous proteins produced from the plastome. Negative effects on growth, under mixotrophic conditions, have been observed when overexpressing a nuclear gene from the plastome or when expressing an enzyme that interfere with some native metabolites. In the former case,

overexpression of a fructose 1,6-biphosphatase did affect the growth importantly (Dejtisakdi and Miller, 2016). On the latter example, analysis of the expression of a bifunctional diterpene synthase from the plastome showed that cells grow slower. The authors suggest that the enzyme might affect the pool of geranylgeranyl diphosphate, which would be the cause of the slower growth rate (Zedler et al., 2015). It is observed that a higher cell density and specific growth rate are obtained from the strain that has been transformed with an empty vector (rTN72). The production of heterologous proteins, in the crICS and the crAP ICS strains, does seem to slightly affect such parameters (Table 3.1). As can be observed in Figure 3.22, the control does grow quicker than the analyzed strains. Nonetheless, the difference does not seem so acute. In the aforementioned examples, the proteins expressed interact with the native metabolic pathways and thus could have an influence on growth. In our case, the heterologous proteins do not interact with any metabolic pathway, so the growth rate does not seem to be affected when having to produce additional proteins. Interestingly, as seen in Table 4.3 the values for cell density and specific growth rate are rather similar to those obtained for the strains expressing just one heterologous protein (crNifV). Overall, the data generated from this chapter shows that simultaneous transgenes expression from the chloroplast genome is attainable. First, I showed that post-translational processing could be possible, but it entailed many limitations for metabolic engineering purposes. Second, I demonstrated that devices could be constructed in a way that they could insert three foreign coding sequences simply. Moreover, it was confirmed that the plasmid and the methodology, based on photoautotrophic restoration, could be scaled to carry additional genetic information without being affected by the burden of extra transgenes. The drawback in transgene instability, brought in by the *rbcL* 3‘ UTR, could be circumvented by utilizing different 3‘ UTRs in each expression cassette. In addition, physical

stability of the transgene-containing cassettes could have resulted in an ideal serial transformation. As detected, the incorporation of the *aadA* gene and the gene encoding the endolysin against *S. pneumoniae* were successful and both seemed to be stably expressed in the organelle. Microalgae, as stated earlier, represent a novel expression platform with several advantages over others. The molecular genetic instruments and procedures described in this chapter could be used as a foundation for improving the intricacy of genetically engineering the chloroplast of this microalga. Undoubtedly, employing synthetic biology approaches such as rational design and standardization of elements encoding biological functions has proved valuable for obtaining predictable outcomes. Additionally, the results obtained here, by the application of some synthetic biology approaches, have revealed many vital particularities that should be taken into account when designing strategies for multi gene engineering.

Chapter 4

Chlamydomonas reinhardtii as a model organism for a genetic analysis of biological nitrogen fixation in green algae

4.1 Introduction

4.1.1 The nitrogen issue

In phototrophs nitrogen (N) is an essential component of nucleotides, amino acids, vitamins and chlorophylls. It is involved, also, in response mechanisms to environmental changes as well as being involved in strategies against oxidative stress (Giordano and Raven, 2014). Despite the fact that the atmosphere is rich in nitrogen, eukaryotes cannot directly utilize this molecule for metabolic processes. Since dinitrogen (N_2) possesses a triple bond, it is necessary to reduce this molecule to render it biologically accessible. This conversion can be done by some prokaryotes (known as di-

azotrophs), and has been referred to as biological nitrogen fixation (BNF). The enzymes carrying out this reduction event are collectively known as nitrogenases. They catalyze the production of ammonia (NH_3) from N_2. Some of these diazotrophs are free living while others have developed symbiotic relationships with some plants. Also, an obligate endosymbiotic relationship between a nitrogen-fixing cyanobacterium and an eukaryotic alga (*Rhopalodia gibba*) has been observed. In this case, the endosymbiont has been observed to live inside the host (Swain and Abhijita, 2013; Cheng, 2008; Kneip et al., 2008).

Along with the industrial revolution came an overwhelming increase in human population. In order for food production to keep up with such rapid growth, the use of nitrogen fertilizers needed to increase substantially. This was aided by the development of what is known as the Haber-Bosch reaction, in which NH_3 is produced from N_2 and H_2 (Figure 4.1). The chemical fixing of nitrogen opened the possibility for generating synthetic fertilizers at an industrial level. Nonetheless, the usage of these fertilizers is directly related to high cost production, high energy consumption and negative effects on human health and the environment (Bouwman et al., 2009; Erisman et al., 2007; Townsend et al., 2003). Synthetic sources of nitrogen are also utilized as nutrients for microalgae cultivation. For instance, algal-based biofuel production requires important inputs of fertilizers. Indeed, it has been suggested that around 40% of the energy input is by this route. (Peccia et al., 2013; Clarens et al., 2010). The likelihood of using eukaryotic microalgae for producing commercially relevant metabolites has been boosted by the application of genetic and metabolic engineering approaches. Nonetheless, the production cost for industrial purposes has been shown to be too high for a profitable production of commodities using this platform (Wijffels et al., 2013).

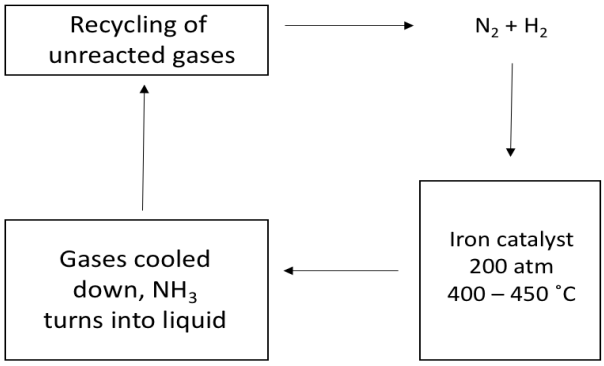


Figure 4.1: Schematic representation of the Haber-Bosch reaction. Nitrogen from the air and hydrogen gases are mixed and piped into a converter. The converter is heated and pressurized, iron is used as a catalyst of the reaction. Nonetheless, only some percentage of the mixture is turned into ammonia, the mixture is cooled down so the ammonia liquefies and could be collected. The remaining gases are recycled and piped into the converter for another round of reaction. The liquified ammonia is stored under pressure. Modified from (Modak, 2011).

Consequently, researchers have explored strategies to reduce cost production and boost the sustainability of microalgae cultivation. Some approaches have focused on reducing synthetic nitrogen inputs (Table 4.1). For instance, the use of wastewater as a source of nitrogen has been proposed. Although, the requirement for aseptic conditions, when producing food or pharma products, reduces the advantages of this option (Peccia et al., 2013). Nutrient recycling has also been used to reduce fertilizer usage. The waste portion (containing nitrogen and phosphorus) generated after lipid extraction has been used as an input for other cultures (Sialve et al., 2009). Conversely, whole cells can be subjected to thermal catalytic conversion for nutrient recycling, which avoids extra steps of extraction and separation (Stephenson et al., 2010). Additionally, microalgae growth has been sustained with the ammonia fixed by genetically modified strains of the diazotroph *Azotobacter vinelandii* (Ortiz-Marquez et al., 2014, 2012). The algae get access to the inaccessible N_2 through the action of the bacterium.

Table 4.1: Methods for reducing external nitrogen requirements.

Method	Approach
Domestic wastewater	N removal; uptake of NH_4^+ and NO_3^-; non applicable when sterile conditions required. Growth competition from other microorganisms (Peccia et al., 2013).
Nutrient recycling	TAG feedstock from microalgae do not contain N or P. Post-lipid extraction waste is rich in N, which could be extracted by: Anaerobic digestion, bacteria are used to transform organic waste into biogas; when organic material is turned into methane NH_4^+ is released (Sialve et al., 2009). or Thermal conversion Pyrolysis, gasification (Stephenson et al., 2010).
Synthetic microalgae-bacteria consortia	*Azotobacter vinelandii* mutants excrete NH_4^+ that sustain microalgae growth. Impairment of *A. vinelandii* NH_4^+-dependent homeostatic control; NH_4^+ released by cells (Ortiz-Marquez et al., 2014, 2012).

An alternative way of expanding BNF for microalgae cultivation would involve direct access to the N_2 atmospheric reservoirs. For this purpose, a nitrogenase enzyme must be assembled within the algal cell. The idea of engineering this enzyme in eukaryotes (plants) has already been put forward (Curatti and Rubio, 2014; Oldroyd and Dixon, 2014). Nevertheless, despite the feasibility of introducing the nitrogenase genes into plant genomes, there are major limitations regarding the environmental conditions required for the enzyme to work such as the enzyme's sensitivity to oxygen or the levels of ATP and reducing power needed. However, other more metabolically diverse eukaryotes, like microalgae, could tackle the demanding environmental requirements for the reduction of N_2 by nitrogenases, making this notion more likely to be relevant to these platforms. Applying such advantages for the study of nitrogenases could help devise strategies to get direct access to atmospheric N_2, which eventually might contribute to reduced fertilizer usage and thus diminish the production cost of commercially relevant commodities. In this chapter, I propose the utilization of *C. reinhardtii* as a model organism for the genetic study of nitrogenase expression in green algae. This model alga is metabolically diverse and its genetics are well established. Thus, it could be utilized as a platform to demonstrate the validity of the system, which could be later be transferred to important oleaginous species or be used to reduce the impact of nitrogen inputs for the synthesis of high value products. The chloroplast, in particular, appears to provide a suitable environment that could harbour nitrogenase activity. A brief description of the enzyme and its function is given as an introduction. Then, the key challenges for engineering the enzyme are described. Subsequently, the pairing of this character with the plastid metabolism is discussed, and finally a laid out of work is presented.

4.1.2 Properties of nitrogenases

Nitrogenase genes are arranged in clusters that encode the enzyme subunits, electron transporters, regulators and metallocluster enzymes. The assembly of the enzyme is transcriptionally regulated depending on oxygen levels, fixed nitrogen and temperature. The best-studied cluster is that from *Klebsiella oxytoca*, and comprises twenty genes encoded in 23.5 kb (Temme et al., 2012). The molybdenum-based nitrogenase enzyme consists of two metallocomponents referred to as dinitrogenase (FeMo-co protein) and dinitrogenase reductase (Fe protein). The former contains the active site for reduction and is organized as a heterotetramer ($\alpha_2\beta_2$), whilst the latter forms a homodimer made of γ subunits. The Fe protein is encoded by *nifH*, while the α and β subunits of the FeMo-co protein are the product of the *nifD* and *nifK* genes. NifH bears a MgATP binding site in each subunit along with a single (4Fe-4S) iron sulphur cluster located at the subunit interface, whereas the NifDK element contains a pair of complex clusters in each subunit known as P-cluster and FeMo-co. P clusters are made of iron atoms and sulfides (8Fe - 7S), and the FeMo-co contains an organic moiety, molybdenum, iron, sulphur and a single carbon atom (Mo Fe7-S9 C-homocitrate) (Figure 4.2).

These clusters provide a channel for electron transfer towards the substrate. The reduction of dinitrogen requires energy in the form of ATP along with high levels of reducing power. Electrons reach the Fe protein via electron donors (flavodoxin or ferrodoxins). Two molecules of MgATP are hydrolyzed in order to transfer an electron from the reduced Fe protein to the FeMo-co protein. This transference must occur eight times to achieve the reduction of the N_2 molecule.

The total stoichiometric reaction is represented as follows:

$N_2 + 8e^- + 16$ MgATP $+ 8\ H^+ = 2NH_3 + H_2 + 16$ MgADP $+ 16$ Pi (Curatti and Rubio, 2014; Swain and Abhijita, 2013; Rubio and Ludden,

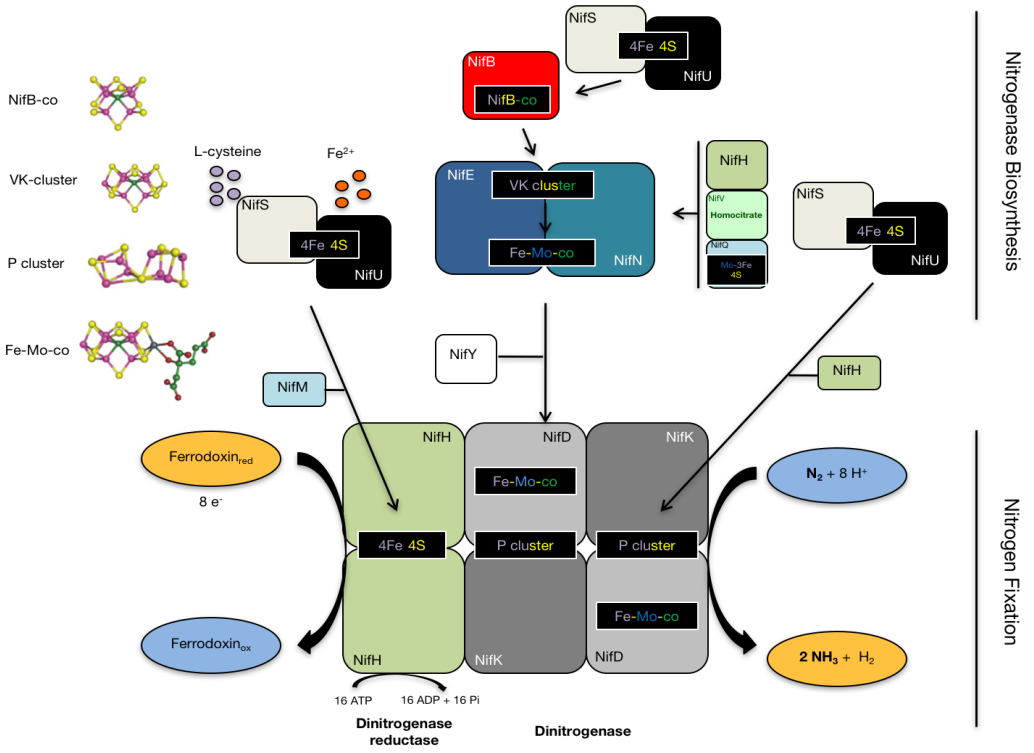


Figure 4.2: Nitrogenase biosynthesis. The enzyme is made of two components, the dinitrogenase reductase (Fe protein) and the dinitrogenase itself (FeMo-co protein). All catalytic cofactors contain Fe-S clusters, which are provided by the NifU/S system. Sulphur (S) atoms are taken from L-cysteine by the action of NifS. These atoms are assembled into clusters along with Fe ion in the NifU scaffold protein. These Fe-S clusters are then transferred to the proper apoproteins. The Fe protein (NifH) acquires these clusters after the mediation of a maturation factor (NifM). The same system provides the clusters upon which the catalytic cofactors of the dinitrogenase element will be assembled. A functional FeMo-co protein (NifDK) requires the synthesis and assembly of the P and the FeMo-co clusters. The assemblage of the P cluster relies on the reduction of two 4Fe-4S clusters by NifH into a single 8Fe-7S cluster. On the other hand, the FeMo-co cofactor is assembled outside the NifDK component. In early steps, 4Fe-4S clusters are transferred to the NifB protein, which is a S-adenosylmethionine (SAM)-dependent enzyme that inserts the C atom into the cluster to form the Nif-co cofactor. This newly formed structure is then passed (via NifX or directly) to another scaffold protein made by the product of *nifE* and *nifN* genes. Here, the NifB-co cluster is transformed into a structure known as VK-cluster (8Fe-9S-C). The NifEN protein contains structural 4Fe-4S clusters, and bears also a Mo-3Fe-4S cluster. The latter is formed in NifQ, where a molybdenum atom is assembled with an Fe-S cluster provided by the NifU/NifS system, and then transferred to the NifEN. Upon interaction between NifH and NifEN, the Mo atom and the homocitrate molecule (provided by NifV) are incorporated to the VK-cluster giving rise to the FeMo-co cofactor, which is then transferred by a chaperone (NifY), or directly, to the apo-NifDK to generate the final holoenzyme. (Yellow: S atoms; Purple: S atoms; Blue: Molybdenum atom; Light green: Carbon atom: Green: Homocitrate).

2008).

As is depicted in Figure 4.3, the number of genes encoding the Mo-based nitrogenase varies depending on the species. It has been observed that *Clostridium acetobutylicum* and *Methanococcus maripaludis* require nine genes while *Paenibacillus* sp. requires ten to produce a functional nitrogenase (Wang et al., 2013a). Thus, analogous enzymes could replace some biosynthetic elements of the nitrogenase assembly machinery. Actually, recent successful expressions of nitrogenase (Wang et al., 2013a,b; Temme et al., 2012) along with reports of its *in vitro* synthesis (Curatti et al., 2007) (Table 4.2) have shown that there is a group of fundamental genes needed for nitrogenase assembly. Hence, a minimal set of proteins must be present, which must include NifB (NifB-co synthesis), NifE, NifN (scaffold protein for FeMo-co synthesis), NifHDK (apoenzyme), and NifV (homocitrate synthase) (Figure 4.2) (Figure 4.3).

All diazotrophs, studied so far, carry a molybdenum-dependent nitrogenase but some also carry additional alternative versions that are dependent on either vanadium or iron (Yang et al., 2014a; Lee et al., 2010). These enzymes are encoded by the *vnf* (vanadium nitrogen fixation) and *anf* (alternative nitrogen fixation) clusters, respectively. These isoforms seem to be activated depending on metal availability (Jouogo Noumsi et al., 2016); It is believed that they follow the same mode of action as the FeMo-co dependent enzyme. Moreover, it has been seen that primary sequences, cluster composition and component proteins share a high degree of sequence similarity (Yang et al., 2014a; Lee et al., 2010, 2009). Nevertheless, there are striking differences between these enzymes. Alternative nitrogenases, for instance, have the structural components arranged as a hexamer, which means that they have an extra pair of σ subunits, encoded by *vnfG* and *anfG*. Compared to the Mo-dependent enzyme, they exhibit a low turnover number. It is worth noting that these alternative versions do not work alone and require

Table 4.2: Genetic engineering examples of biological nitrogen fixation

Component protein	Genes employed	Remarks	References
Iron-only nitrogenase from *A. vinelandii*	(*nifJ, H, D, G, K, U, S, V, F, B*)	*E. coli* able to sustain nitrogenase activity	(Yang et al., 2014a)
Mo-based nitrogenase from *Paenibacillus* sp	(*nifB, H, D, K, E, N, X, V, hesA*)	Synthesis of catalytically active nitrogenase in *E. coli*	(Wang et al., 2013a)
Mo-based nitrogenase from *K. oxytoca*	(*nifJ, L, A, H, D, K, T, Y, E, N, X, U, S, V, N, Z, M, B, Q*)	Nitrogenase activity measured in *E. coli* strains	(Wang et al., 2013b)
Mo-based nitrogenase from *K. oxytoca*	(*nifH, D, K, Y, E, N, J, B, Q, F, U, S, V, W, Z, N*)	*nif* operon under control of synthetic regulators in the native host	(Temme et al., 2012)
Iron-Molybdenum cofactor from *K. oxytoca*	(*nifB, E, N, H, D, K*)	*in vitro* synthesis of FeMo-co, and activity of apo-dinitrogenase	(Curatti et al., 2007)
NifH from *K. pneumoniae*	*nifH*	Detection of NifH by immunoblot analysis from *S. cerevisae* extracts	(Berman et al., 1985)
nif cluster of *K. pneumoniae*	Not specified	Stable chromosomal integration of *nif* cluster in *S. cerevisae*	(Zamir et al., 1981)
NifH from *K. pneumoniae*	*nifH*	NifH restores chlorophyll production in the dark by substituting the ChlL enzyme in *C. reinhardtii*	(Cheng et al., 2005)
NifV from *K. oxytoca*	*nifV*	NifV detected by Western blot analyses; NifV accumulation in *C. reinhardtii* does not disturb growth rate	This report

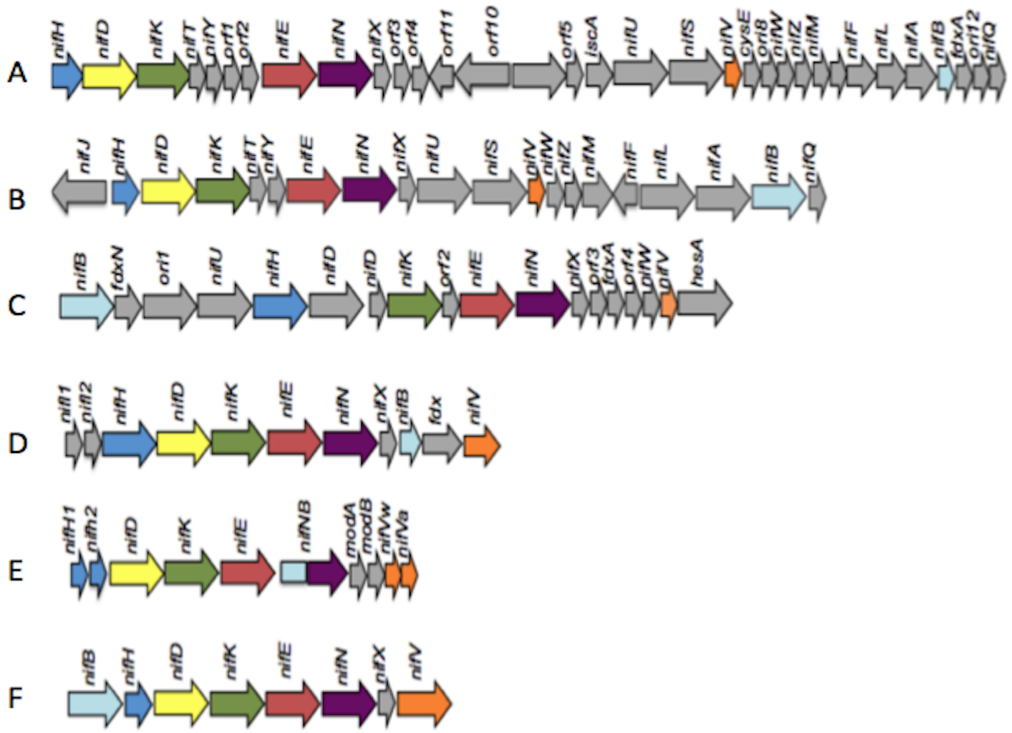


Figure 4.3: Comparison of various representative *nif* clusters. A. *Azotobacter vinelandii*, B. *Klebsiella oxytoca*, C. *Anabaena variabilis* D. *Heliobacterium chlorum* E. *Clostridium acetobutylicum* W5, F. *Paenibacillus* sp. WLY78.

the function of some Nif proteins (NifU, NifS, NifB, and NifV) for assembling the clusters (Yang et al., 2014a; Lee et al., 2010, 2009; Hallenbeck and Benemann, 2002).

4.1.3 Limitations for expressing nitrogenase genes in eukaryotic platforms

One of the most significant barriers for transferring the nitrogenase trait into eukaryotic organisms is the enzyme's sensitivity to molecular oxygen. In early evolution, iron and sulphur atoms were easily combined to form Fe-S clusters. Nonetheless, the eventual oxygenation of the atmosphere brought a new environment that was not suitable for Fe-S-containing enzymes. Uni-

valent oxidants such as superoxide, hydrogen peroxide or molecular oxygen disrupt the Fe-S clusters by means of oxidization (Imlay, 2006). Some diazotrophs cope with oxygen production and nitrogen fixation by means of temporal separation (N_2 fixation at night) or spacial separation (N_2 fixation occurs in special compartments named heterocysts). Therefore, the platform devised to study heterologous nitrogenase must be capable of growth in low levels of oxygen. The complex arrangement and redundant function of the nitrogenase components (Figure 4.2) represent another limitation for their engineering in eukaryotes. The various attempts of engineering nitrogenase in bacteria have demonstrated that the most efficient approach is to design synthetic constructs (using promoters, UTR's and terminators from a different organism), which permits the avoidance of native regulation or feedback mechanisms (Yang et al., 2014a; Wang et al., 2013a,b; Temme et al., 2012). There are several methodologies for expressing and processing multigene complexes in eukaryotes (Bieniossek et al., 2012; Vijayachandran et al., 2011; Naqvi et al., 2010). Nonetheless, the efficiency of these methods depends on the host being used. The potential build up of ammonia and ammonium (formed by the protonation of ammonia) produced by the transgenic nitrogenase would represent an important limitation when coupling the enzyme into the host. It has been argued that, in plants, toxic levels could limit growth and reduced yield, on the grounds of carbon depletion, pH elevation, shortage of mineral cations, imbalanced chlorophyll and lipid biosynthesis, activation of oxidative stress responses, and wasted cycling of ammonium in the plasma membrane (Curatti and Rubio, 2014). As seen earlier, the process of nitrogen fixation is metabolically demanding and requires significant amounts of energy in the form of ATP as well as high levels of reducing power. Additionally, for metabolic engineering purposes the number of genes to be inserted must be reduced to a minimum in order to avoid metabolic load. Thus, it is necessary to target the enzyme

to a place that provides potential substitutes for some of the biosynthetic elements of the pathway. Additionally, significant levels of Mo, V or Fe atoms must be present in the host in order to sustain the assembly of the catalytic cofactor. All these limitations seem hard to overcome. However, various potential locations have been suggested for targeting nitrogenases into eukaryotes. Cellular organelles, in particular, have been highlighted as prospective compartments capable of sustaining the assembly and function of nitrogenases.

4.1.4 Organelles as environments for nitrogenase expression

Mitochondria and plastids have been proposed as potential locations for nitrogenase expression in plants (Curatti and Rubio, 2014; Oldroyd and Dixon, 2014). These organelles provide features that can be exploited for recombinant nitrogenase production. Mitochondria, for instance, offer potential protection against oxygen sensitivity, since respiration consumes the O_2 reservoirs within these organelles, which renders feasible the assembly of O_2- labile enzymes. In fact, Fe-S clusters are part of several mitochondrial enzymes. Moreover, the assembly machinery of these enzymes has been observed to be similar to the one involved in nitrogenase biosynthesis (NifU/S) (Lill, 2009). Similarly, chloroplasts harbour many enzymes based on Fe-S clusters. Again, the assembly system present in the chloroplast shares some biosynthetic principles with the Nif system. For example, metalloclusters are assembled in scaffold proteins and then transferred to the apoprotein (Lill, 2009). These organelles represent important biosynthetic centres of the cell, they harbour several energetically demanding pathways, which require high levels of ATP and reducing power. In this respect, these subcellular compartments represent a potential functional host for express-

ing nitrogen-fixing proteins. Both organelles are derived from bacteria, thus, they harbour prokaryotic-like machinery for processing genetic information, which opens the possibility of operon-like gene arrangements. Despite the advantages provided by these alternatives, there are some essential drawbacks. For instance, the coupling of nitrogen fixation and oxygen-producing photosynthesis within a chloroplast is still a limitation for plant platforms. Since the enzyme is irreversibly affected by oxygen, the process of nitrogen fixation must take place at night, which would create an imbalance of the amount of energy allocated for chloroplast metabolism (Curatti and Rubio, 2014). Nevertheless, other eukaryotic platforms could fit better the nitrogenase profile of a host. Yeast as well as the microalga *Chlamydomonas reinhardtii* have been widely used for mitochondrial genetic manipulation (Bonnefoy et al., 2007). *C. reinhardtii*, as detailed before, is widely utilized for chloroplast transformation, and it can be grown under various conditions. Some attempts at expressing *nif* genes in yeast have been carried out (Berman et al., 1985; Zamir et al., 1981). The presence of the polypeptides was confirmed but no nitrogen reduction qualities were observed in the transformants (Table 4.2). In contrast, the sequence coding for the structural part of the dinitrogenase reductase (*nifH*), from the *K. oxytoca nif* cluster, was inserted into the chloroplast genome of *C. reinhardtii*. The functionality of the enzyme was confirmed by rescuing the yellow in the dark phenotype typical of mutants unable to produce chlorophyll pigments in the dark (Cheng et al., 2005). This phenotypic rescue, mediated by the bacterial dinitrogenase reductase (*nifH*), shows that genes from the nitrogenase cluster could be expressed in the chloroplast of *C. reinhardtii*. Moreover, it is obvious that some native chloroplast machinery is providing ATP, reducing power and surrogate scaffold proteins for this enzyme to work. Interestingly, *C. reinhardtii* presents additional genetic and metabolic features that could be exploited to potentially sustain nitrogen fixation within its plastid.

4.1.5 Integrating nitrogen fixation into the chloroplast of *C. reinhardtii*

4.1.5.1 *Nitrogenase sensitivity to oxygen*

The enzyme's sensitivity to oxygen represents an important limitation for engineering a heterologous nitrogenase into photosynthetic platforms. Nevertheless, some diazotrophs have been able to couple the presence of the oxygen-producing photosystem II and the ability to fix nitrogen directly from the atmosphere. Some cyanobacteria, for instance, are able to combine these two apparent mutually exclusive processes by two main strategies. On the one hand, during dark conditions (day/night cycle) oxygen production by the PSII decreases and the remaining oxygen is consumed by respiration. In such conditions, nitrogenases are able to carry out their reactions (Zehr, 2011). On the other hand, physical separation between nitrogen fixation and photosynthesis has been observed in a number of diazotrophic cyanobacteria. Special chambers, known as heterocysts are generated, which develop thick cell walls (hydrophobic barrier) and down regulate the activity of PSII. These conditions prompt the expression of nitrogenase genes (Berman-Frank et al., 2007). Evidently, the strategy for testing nitrogenases in eukaryotes must involve a period of dark anaerobic conditions. In this respect, *C. reinhardtii* seems to fit adequately as a potential host (Kosourov et al., 2007; Mus et al., 2007). In fact, this green alga is able to sustain the activity of oxygen-sensitive enzymes that work under these conditions. For example, green algae possess an enzyme system responsible for the production of a chlorophyll precursor in the absence of light. This reductase is structurally related to nitrogenase and it is activated in the dark. Dark-operative Pchlide oxidoreductase (DPOR) reduces protochlorophyllide (Pchlide) to chlorophyllide, which is then reduced to chlorophyll. The enzyme is made of structural components (ChlL, ChlN, ChlB) whose

amino acid sequences are highly similar to those of nitrogenase (NifH, NifD, NifK, respectively). (Moser et al., 2013; Reinbothe et al., 2010). Likewise, hydrogenases, which catalyse the reduction of protons, are assembled in the chloroplast once anaerobic conditions are sensed (Dubini and Ghirardi, 2015; Grossman et al., 2011). This opens the possibility of utilizing promoters and UTRs of genes activated in such conditions. Nevertheless, expression levels driven by these sequences must be assessed first. The ability of *C. reinhardtii* to grow in such conditions is based on fermentation circuits that allow the generation of energy in the absence of oxygen (Yang et al., 2015; Grossman et al., 2011). Furthermore, photosynthetic mutants, lacking photosystem II (PSII), have been used as recipient strains for the introduction of foreign genes (Economou et al., 2014; Young and Purton, 2014; Cheng et al., 2005), which demonstrates the ability of *C. reinhardtii* to grow in the absence of a functional PSII when fixed forms of carbon such as acetate are provided. This ability is based on the coupling of acetate to the cell metabolism via the glyoxylate pathway after being converted into acetyl CoA (Yang et al., 2015). Generally, the ease of expressing heterologous proteins in the chloroplast and the trophic plasticity of this green alga make it possible to envisage strategies for avoiding the disassembly of the metal clusters caused by oxygen damage.

4.1.5.2 *Providing metal atoms, surrogate scaffold proteins, energy and reducing power for nitrogenase expression*

The assembly of nitrogenase requires the coordinate expression of various genes (Figure 4.2); with a core group of fundamental genes being essential (Figure 4.3). In fact, successful attempts of nitrogenase expression in *E. coli* have made use of minimal sets of genes, which encode for the structural components as well as some biosynthetic enzymes (Wang et al., 2013a,b) (Table 4.2). The fact that the metalloclusters had been assembled shows

that some native machinery is interacting with the heterologous components of nitrogenase. The three types of metalloclusters, present in a functional enzyme, are composed of iron-sulphur clusters and are assembled in scaffold proteins (Figure 4.2). *C. reinhardtii* have been observed to harbor genes showing high similarity with the bacterial SUF (sulphur assimilation) system (Godman and Balk, 2008). The SUF system, also present in plant plastids, consists of a group of scaffold and biosynthetic proteins (Figure 4.4) that act in a similar way to the NIF system found in diazotrophs (Lill, 2009; Tokumoto et al., 2004). This system may be responsible for the assembly of Fe-S cofactors of the photosystems as well as of hydrogenases (Esselborn et al., 2016; Rochaix, 2002). In any case, the fact that NifH reductase from *K. oxytoca* could restore pigment synthesis in dark-grown *C. reinhardtii* demonstrates that the chloroplast native machinery could assemble its Fe-S cluster. Furthermore, these results show that ATP could be coupled into the NifH reductase.

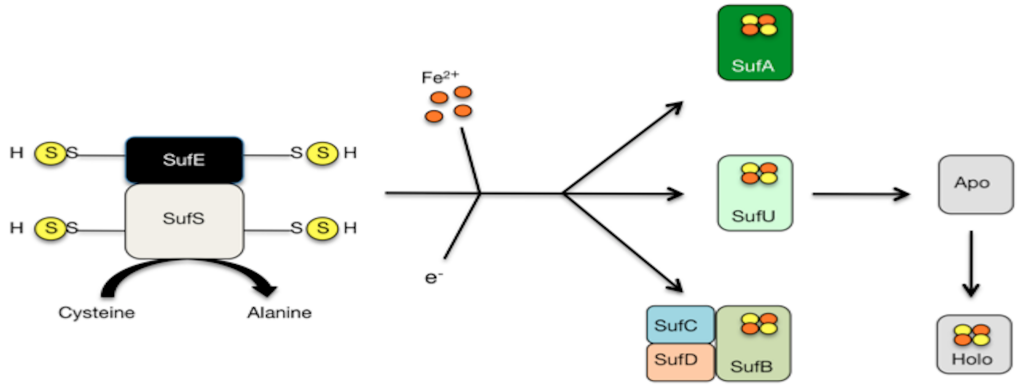


Figure 4.4: Schematic representation of the SUF model for Fe-S protein biogenesis. The genesis of Fe-S clusters involves their *de novo* assembly in scaffold proteins, and their subsequent transference to the respective apopoteins to finally yield the functional forms. Sulphur atoms are transferred from cysteine via a cysteine desulphurase (SufU) to the sulphur shuttle system (SufU, SufE). These atoms are reduced using electrons of yet unknown origin to produce persulphide. Donors not specified so far provide iron ions, which are used to assemble the clusters in scaffold proteins. SufA is related to both 2Fe-2S and 4Fe-4S clusters, while SufU is associated with 2Fe-2S. Also, another scaffold protein (SufB) has been seen to interact with an ATPase (SufC) and another protein (SufD) for the formation of Fe-S clusters. Once assembled, these clusters are transferred to their respective apoproteins to generate the final holo form. Modified from (Lill, 2009).

The scarcity of molybdenum atoms in chloroplasts (Chamizo-Ampudia et al., 2011) might pose a limitation for the assembly of the catalytic cluster. Nonetheless, the recent expression of an iron-only nitrogenase in *E. coli* (Yang et al., 2014a) paves a way for its further research in plastids. Overall, the chloroplast metabolic network could provide the required elements for an iron-based nitrogenase to work. Nevertheless, considering the amount of electrons and energy needed to sustain nitrogenase activity, it is almost impossible to predict if the redirection of the electron flow would affect native pathways. Additionally, if the algae should be grown in dark anaerobic conditions, high levels of energy would not be able to be maintained. As previously mentioned, *C. reinhardtii* possess a wide spectrum of metabolic capacities. When the cells sense low levels of oxygen, they trigger the catabolism of polysaccharides or sugars to produce ATP via the mito-

chondrial electron transport chain (Yang et al., 2015). Moreover, it has been noted that ATP and reducing agents can be imported from mitochondria into chloroplasts with the purpose of setting up proper conditions for photosynthesis before the onset of light (Yang et al., 2015). This observed cross talk between organelles could provide an extra source of energy and reducing power towards nitrogenase. As noted earlier, temporal separation between nitrogen fixation and photosynthesis might benefit the coupling of these two pathways in the chloroplast. Cells could be grown under light conditions, so they can accumulate energy and reducing power, to be later exposed to dark/anoxic conditions suitable for nitrogenase functioning. Additionally, the fact that photosynthetic (PSII) mutants are viable under fixed sources of carbon demonstrates that light could be used to increase the ATP pool, in anaerobic conditions, by means of cyclic photophosphorylation. Despite all the potential alternatives, such a demanding enzyme will obviously impact the dynamics of the cell. Thus, the regulated expression of nitrogenase components will be imperative.

4.1.5.3 *Ammonia Ammonium toxicity*

Chlorophytes have been shown to be more tolerant to high ammonium concentrations than diatoms, dinoflagellates or raphidophytes (Collos and Harrison, 2014). Nonetheless, ammonium toxicity has been seen to have a negative effects on the photosynthetic apparatus of some green algae (Markou and Nerantzis, 2013). The assimilation of inorganic nitrogen in *C. reinhardtii* is based on nitrates, nitrites and ammonium. Nitrates and nitrites enter the cell via transporters located in the plasma membrane. Nitrites enter directly to the chloroplast via specific transporters located in the membrane. Nitrates, on the contrary, are reduced first to nitrite in the cytoplasm and then enter the chloroplast via the already mentioned transporters.

Once nitrites are inside the organelles they are reduce to ammonium, which is taken up by the glutamine synthase - glutamate synthase pathway (GS-GOGAT). Ammonium present in the media enters the cell via a group of transporters located both in the plasma and chloroplast membranes. Once within the organelle, ammonium is incorporated into carbon skeletons by the GS-GOGAT pathway (Fernandez and Galvan, 2007). Hence, it could be speculated that the extra ammonium provided by the nitrogenase could be incorporated into amino acids to avoid toxic concentrations. Furthermore, it has been shown that *C. reinhardtii* prefers ammonium as a source of nitrogen, and once the cell detects it, transcriptional repression of nitrate assimilation is activated (de Montaigu et al., 2010). In any case, it is evident that the ammonium produced by the nitrogenase will have an influence on the normal carbon-nitrogen ratios.

4.1.5.4 *Minimal set of genes for nitrogenase expression*

The genes required for diazotrophy vary among species, this divergence depends upon physiological and ecological requirements of the host. Figure 4.3 depicts a comparison of various gene clusters of some nitrogen-fixing organisms. It has been proposed that a minimal group of six genes (*nifH, nifD, nifK, nifE, nifB* and *nifN*) is conserved in almost all diazotrophs. These genes encode the apoenzyme (*nifHDK)* and the scaffolding structures responsible for the synthesis of the Mo cofactor (Curatti and Rubio, 2014; Wang et al., 2013b). As previously mentioned, various attempts have partially succeeded in generating *E. coli* strains that are able to express a functional Mo-dependent nitrogenase using genes from the *K. oxytoca* cluster (Wang et al., 2013a). Similarly, alternative efforts have managed to produce an *E. coli* strain capable of reducing atmospheric nitrogen by using the reduced *nif* cluster from *Paenibacillus* sp. (Wang et al., 2013a). Similarly,

(Yang et al., 2014a) achieved a similar result by using the cluster of the alternative nitrogen fixation pathway belonging to *A. vinelandii*. In this research, they produced an artificial system in *E. coli* consisting of genes from the *anf* and *nif* clusters. In this case, the structural components of the enzyme were encoded by the *anfHDKG* genes, while the peptides needed for the biosynthesis of the FeFe cofactor were encoded by the *nifBUSV* genes. Moreover, the *nifF* and *nifJ* sequences were also employed to provide a flux of electrons towards the reductive chain. This assembly of ten genesshowed to generate a functional Mo-independent nitrogenase. As mentioned in the previous section, those results represent an important foundation for further research in eukaryotic platforms. In fact, I suggest that the version of nitrogenase to be tested must be based on the 10-gene cluster used by (Yang et al., 2014a). Additionally, it might be possible to reduce the number of genes from this minimal set. As discussed before, the partially successful expression of *nifH* in the chloroplast of *C. reinhardtii* demonstrates that electrons flow through the heterologous subunit (NifH), which means that electron donors from other pathways could be hijacked for different purposes. Thus, I could exclude the genes (*nifJ; nifF*) encoding electron donors. The presence of the SUF system, as highlighted previously, could allow us to exclude the genes (*nifU; nifS*) required for Fe-S cluster formation. Hence, the suggested minimal set of genes would comprise the sequences encoding for the apoenzyme (*anfHDGK*), the NifB-cofactor (*nifB*) and the organic moiety (*nifV*) subunit (Figure 4.5). If the plastid machinery could not be coupled, genes encoding for electrons donors (NifJ, NifF) and the NifU/S system will have to be added as optimization steps. As detailed before, homologous recombination allows the site-specific insertion of foreign genes into the chloroplast genome, and the expression of multiple genes has been achieved to some extent. Metabolic engineering efforts have achieved the expression of single enzymes involved in pathways with commercial promi-

nence. To date, there are no accounts regarding the simultaneous expression of more than two heterologous proteins (including the selectable marker) in the chloroplast of *C. reinhardtii*. However, current efforts in our laboratory have helped devise a procedure to express three heterologous proteins simultaneously with one transformation event and no trace of antibiotic markers (section 3.2).

4.1.5.5 *Strategy for investigating nitrogenase gene engineering in the chloroplast of Chlamydomonas reinhardtii*

The lack of a standardized strategy for multiple gene expression in the chloroplast of *C. reinhardtii* limits the ability to metabolically engineer this organelle. Nevertheless, testing the individual expression of key enzymes, involved in commercially relevant pathways, remains fundamental as a proof of the chloroplast's metabolic versatility. Therefore, by using the already detailed tools for genetically engineering of the chloroplast of *C. reinhardtii*, I decided to individually test the genes encoding the biosynthetic elements of the suggested minimal cluster for nitrogenase expression (Figure 4.5). As discussed before, the functionality of nitrogenase depends on the proper assembly and further incorporation of the metal cofactors into their correspondent apoproteins. I have already discussed that some native machinery of the chloroplast could provide the means for proper cluster assembly. Nonetheless, the catalytic cluster of nitrogenase bears a complicated arrangement of atoms that can only be attained by means of Nif enzymes. The hypothetical minimal set of genes is composed of six elements, three of them encode the structural part of the enzyme (AnfGDK), while two the biosynthetic elements (NifB, NifV) of the pathway, the remaining constituent corresponds to the reductase component (AnfH), which play vital roles in both instances (Figure 4.5). These biosynthetic elements are crucial for the proper assembly of the enzyme. For instance, the *nifB* gene encodes

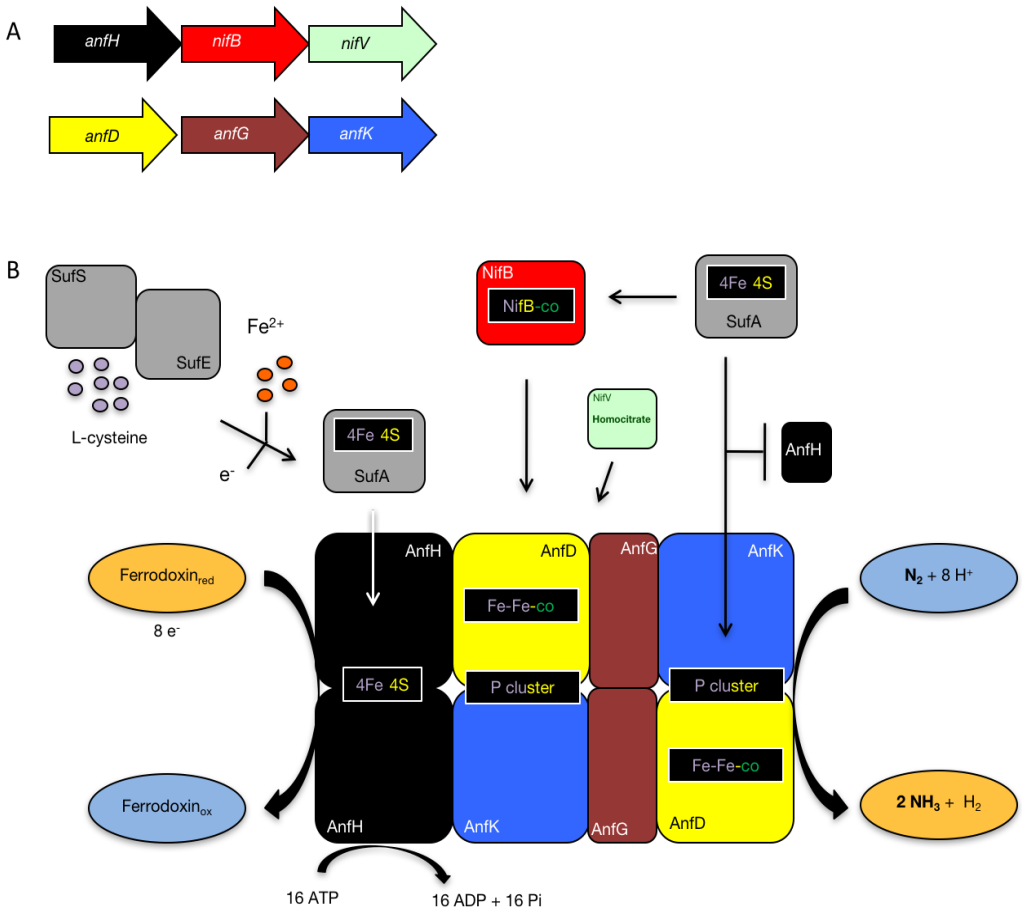


Figure 4.5: Proposed nitrogenase for expression in the chloroplast of *C. reinhardtii*. A. Suggested minimal set of genes codifying for an iron-only nitrogenase. The sequences encoding for the Fe reductase (AnfH) and the dinitrogenase (AnfHDGK) components belong to *A. vinelandii*, while the ones encoding for biosynthetic elements (NifB, NifV) belong to *K. oxytoca*. B. Schematic representation of the hypothetic biosynthetic pathway of the iron-only nitrogenase in the chloroplast. The native (SUF) system for assembling Fe-S clusters could provide the catalytic cofactors for the Fe reductase (AnfH). Similarly, the cofactors (P clusters, Fe-Fe-co) present in the dinitrogenase (AnfHDGK) could be assembled based on Fe-S clusters provided by the same system. The P clusters are generated via the action of AnfH. The (SAM)-dependent NifB is required for the addition of the carbon atom to the 4Fe-4S metal structure, while NiV provides the homocitrate molecule. (Yellow S atoms; Purple S atoms; Light green Carbon atom: Green Homocitrate).

a SAM enzyme that is in charge of sequestering the carbon atom from the methyl group and transferring it to yield a transitional cluster (Figure 4.2) (Wiig et al., 2012). In the canonical description, this structure is then passed to the NifEN protein complex where it is transformed into what is known as

the VK-cluster. Subsequently, the Mo atom along with the homocitrate are adapted to the latter cluster, which is then transferred to the apo-NifDK (Figure 4.2). However, this process of cluster maturation has not been observed in other versions of the molybdenum-based nitrogenase. For example, *Paenebacillus* is able to fix nitrogen without the presence of maturations proteins like NifEN, which suggests that the assembly of the cluster, in this case, must be carried out in apo-NifDK directly (Wang et al., 2013a). Likewise, heterologous versions of the enzyme as well as alternative versions of it do not seem to make use of these maturation proteins (Yang et al., 2014a; Wang et al., 2013a) Nevertheless, the NifB SAM enzyme must be present to catalyze the early formation of the metal cluster, and, thus, the *nifB* gene has been cataloged as essential for diazotrophy. Likewise, the presence of the *nifV* gene is essential for nitrogen-fixing capacities (Figure 4.3). The NifV enzyme is a homocitrate synthase whose product (Homocitrate: (R) -2-hydroxy-1,2,4-butanetricarboxylic acid) is coupled to the catalytic cluster via its 2-hydroxy and 2-carboxyl groups. Homocitrate synthases catalyze the condensation of some keto acids (e.g. α-keto-glutarate) with acetyl-CoA to produce homocitrate and coenzyme A. This organic moiety has been shown to contribute to the global structural assembly of the cluster and to its overall redox properties (Gabriel and Milewski, 2016; Wulandari et al., 2002; Zheng et al., 1997). Similarly, the AnfH reductase has a fundamental role in the assembly of the P clusters and is the component responsible for the coupling of ATP to the redox reaction. This component also plays a structural role whereby it serves as a scaffold for an iron-sulphur cluster that allows the passing of electrons towards the catalytic site (Yang et al., 2014a). As mentioned earlier, NifH has been shown to be functional in the chloroplast of *C. reinhardtii* (Cheng et al., 2005). Percentages of similarities, produced using the BLAST tool (https://blast.ncbi.nlm.nih.gov), showed that AnfH had more similarity (37%) than NifH (34%) to the ChlL en-

zyme. Furthermore, protein alignments, done using the ClustalW2 software (http://www.ebi.ac.uk/Tools/msa/clustalw2/) demonstrated that the AnfH possesses consensus regions serving as Fe-S cluster ligands, and nucleotide-binding motifs (Figure 4.6).

Figure 4.6: Comparison between protein sequences of the ChlL, NifH and AnfH reductases. Sequence alignment of the Fe protein of the *nif* cluster from *K. oxytoca*, the AnfH reductase from *A. vinelandii*, and the ChlL protein from *C. reinhardtii* display the conserved cysteine groups (purple) that are required as ligands for the Fe-S clusters. Whereas, in green the motif that is in charge of nucleotide binding (Lee et al., 2010).

These components, as noted, are fundamental for the assembly of the necessary machinery that allow the reduction of atmospheric nitrogen. Thus, the investigation of their behavior in the chloroplast of *C. reinhardtii* will provide crucial insights about how this pathway might be integrated into the organelle's metabolism.

4.2 Plasmid design for insertion of *nif/anf* genes into the chloroplast of *C. reinhardtii*

The chloroplast transformation scheme already described in chapter 3 was used to transform viable cells of *C. reinhardtii*. In this case, the vector used for transformation (pSRSapI) is a variant of the original pASapI vector whereby gene expression is under the control of the *psaA exon 1* promoter 5' UTR, rather than that of *atpA*. This new expression vector has been recently used for the study of various aspects regarding the engineering of the chloroplast of *C. reinhardtii* (Wannathong et al., 2016; Young and Purton, 2015, 2014). The gene sequence encoding the NifB and NifV enzymes from *K. oxytoca* and the AnfH from *A. vinelandii* were codon optimized and synthesized by GeneArt. An HA-tag was added at the C terminus for immunodetection. The *nifB*, *nifV* and *anfH* genes were cloned into the pSRSapI vectors via a *Sap*I - *Sph*I digestion ligation event. The resulting plasmids were named pnifB, pnifV and panfH, respectively (Figure 4.7) (Appendix B.2).

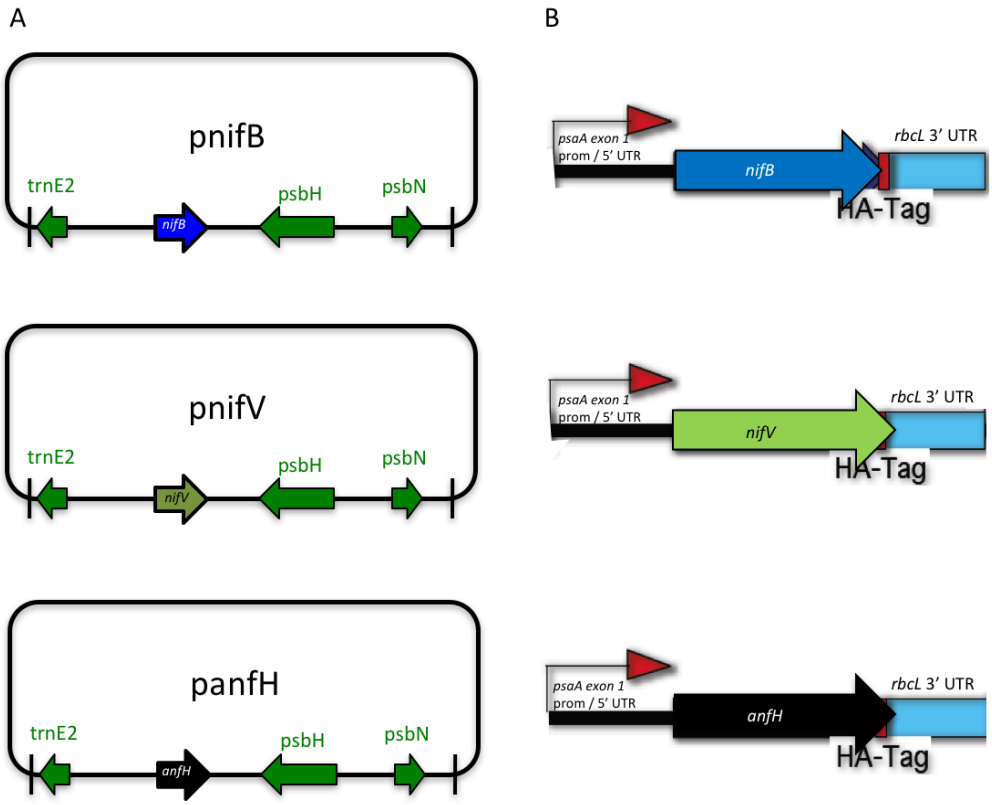


Figure 4.7: Plasmids used for insertion of *nif/anf* genes into the chloroplast genome of *C. reinhardtii*. A. The three plasmids used for foreign gene insertion into the plastome, which are based on the pSRSapI vector, so they carry a functional version of the *psbH* gene along with the sequences needed for homologous recombination. B. The three genes are arranged into the same expression cassette, so expression is driven by the *psaA exon 1* promoter 5' UTR and the *rbclL* 3' UTR. All genes were HA-Tag for immunodetection.

4.3 Transformation of the recipient strain (TN72) with plasmids bearing *nif/anf* genes

The three constructs generated (pnifB, pnifV and panfH) were used to transform the recipient strain TN72. DNA delivery was attained using agitation with glass beads, as described in section 2.3. After various rounds of photoautotrophic selection, transformants were confirmed by PCR (Appendix B.1). The size of the products obtained corroborates the proper

integration of the heterologous genes. A 1.2 kb amplicon can be observed in the crnifB, crnifV and cranfH lanes, while a 0.85 kb amplicon can be seen in the TN72 lane (Figure 4.8). The lack of a 0.85 kb amplicon in the transformant lines indicated that all the copies of the genome have the foreign expression cassette.

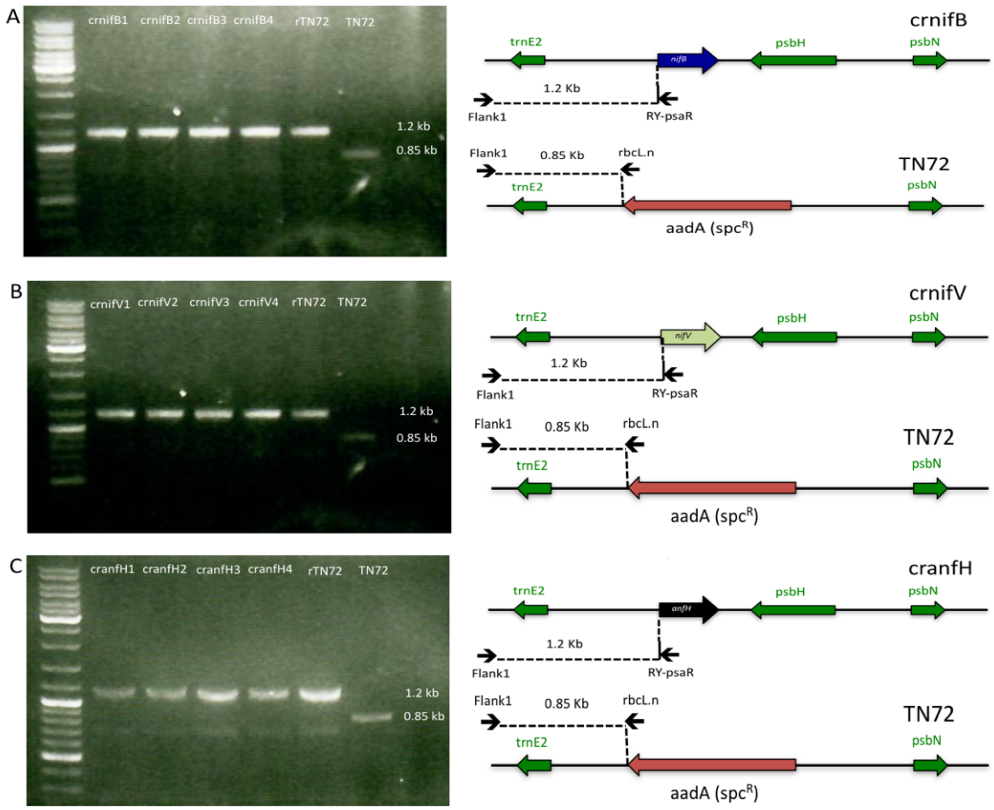


Figure 4.8: Identification of transgenic lines by colony PCR. Left-hand side: Agarose gels showing the PCR products from transformant lines and controls. Right-hand side: Diagrams illustrating the genetic arrangement of the TN72 and the transformant genomes. The Flank1 primer is targeted to the 5' upstream of the homologous recombination site. Whereas, the reverse primers target either a region of the *aadA* cassette or a section of the inserted expression cassette. A. A 1.2 kb amplicon confirms the insertion of the *nifB* gene (blue) into the chloroplast genome. B. A 1.2 kb amplicon shows the proper incorporation of the *nifV* gene (green) into the plastome. C. The 1.2 kb amplicon demonstrates the correct integration of the *anfH* gene (black). The rTN72 lane corresponds to the positive control. This strain has been transformed with an empty expression cassette. The 0.85 kb amplicons correspond to the recipient strain TN72, which was used as a negative control.

4.4 Expression of recombinant nitrogen-fixation proteins in the transformant lines

Expression of heterologous nitrogen fixing proteins in crude extracts was confirmed by immunoblotting using antibodies targeted to the HA-tag, as described in section 2.4. Strains were grown and harvested under both light and dark anaerobic conditions as explained in section 2.1. Under light conditions, the expression of the NifB SAM enzyme could not be detected (52 kDa) (Figure 4.9A). On the other hand, the expression of the homocitrate synthase NifV could be observed at 43 kDa (Figure 4.9B). Similar to crNifB, the reductase AnfH could not be detected (31 kDa) (Figure 4.9C). Under dark anaerobic conditions, (Figure 4.9) the same results were observed. The only heterologous protein that could be detected was the homocitrate synthase (NifV) (Figure 4.9B). The lack of protein accumulation might be related to deficient folding, which triggers degradation or could be that codon optimization influenced negatively the secondary structure of mRNA and compromised translation, more details are given in the discussion section.

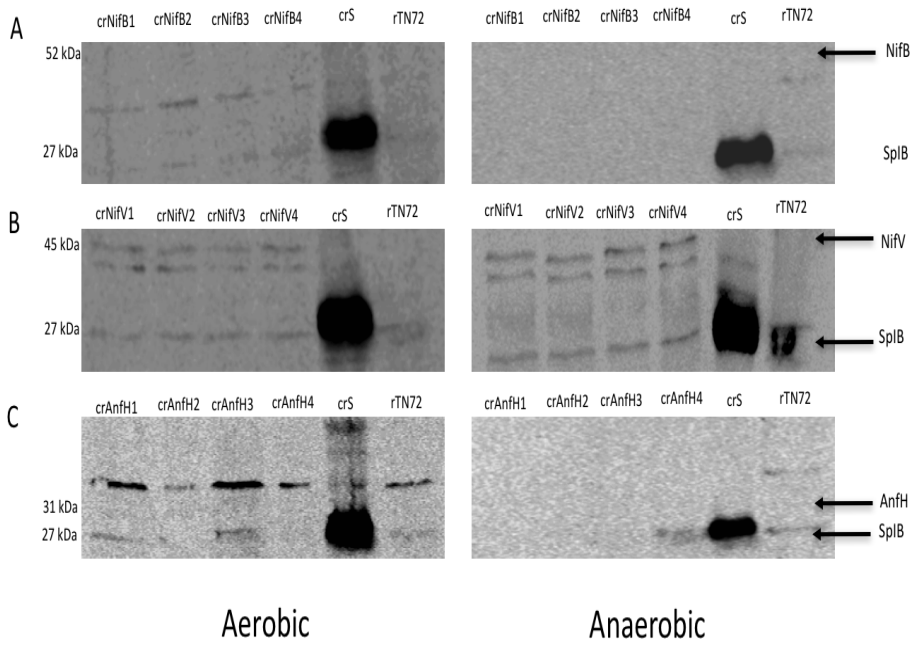


Figure 4.9: Western blot analysis of the transgenic lines containing the *nif/anf* genes. Cell lysates were separated by SDS-PAGE on identical gels, blotted and incubated with α-HA-Tag antibodies. Left handside: normal conditions. Right handside: dark anaerobic conditions. A. Transformants that have the *nifB* gene integrated into the plastome. B. Transformants expressing the NifV enzyme (43 kDa). C. Transformants that have the *anfH* gene integrated into the plastome. The crS strain, expressing a serine protease (27 kDa) was used as positive control for α-HA blotting. The rTN72 lane corresponds to the negative control. This strain has been transformed with an empty expression cassette.

4.5 Effects of NifV expression on culture - growth

One of the four transformants (crNifV3) was chosen to determine if there were any effects of homocitrate synthase expression on the specific growth rate and maximum cell density. These values are shown in Table 4.3, and Figure 4.10 shows the growth curve for the evaluated strain and the negative control (rTN72), which is a strain that has been photosynthetically restored using an empty pSRSapI vector. During the exponential phase, the strain expressing the homocitrate synthase appears to grow slightly slower than the control, this difference in growth is maintained throughout the linear

phase. As shown in Table 4.3 the values of maximum cell density and specific growth rate were somewhat lower than those obtained from the control, which indicates that there is some burden in producing NifV, which may suggest that the enzyme is interacting with an important pathway in the chloroplast. The next step was to determine if the heterologous enzyme is active in cell-free extracts.

Table 4.3: Specific growth rate and maximum cell density obtained for the strains expressing NifV. Cells were grown in 1 L flasks at 120 rpm of agitation and at 50 μE m^{-2} s^{-1}. Specific growth rates at stationary phase are an average from three independent measurements.

Strain	**Specific growth rate (min^{-1})**	**Maximun cell density OD_{750}**
crNifV3	0.048 +/- 0.003	3.15 +/- 0.05
rTN72	0.051 +/- 0.002	4.50 +/- 0.12

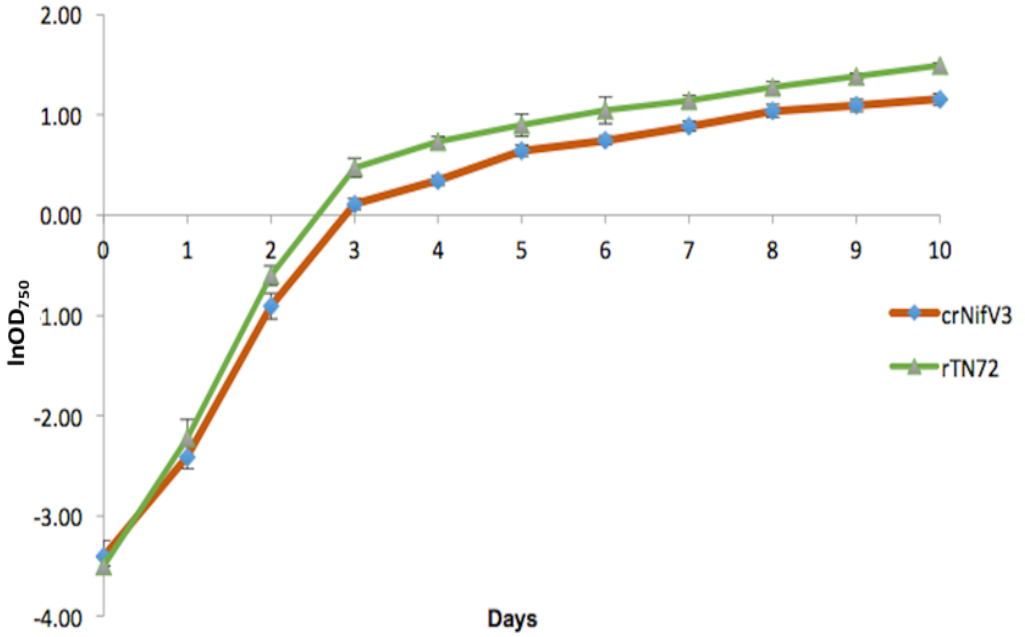


Figure 4.10: Effects on culture growth of NifV expressed from the plastome. Expression of NifV proteins does not have a negative effect on growth. Although, a maximun OD value is attained by the control. The transgene-containing strain crNifV3, and the control rTN72 were grown mixotrophically. The data presented are averages of three independent measurements.

4.6 *C. reinhardtii* NifV activity in cell-free extracts

Since accumulation of NifV was detected in the transformants, I decided to analyze if the enzyme activity could be detected in cell-free extracts. Cells from the four strains expressing the homocitrate synthease NifV (Figure 4.9B) were mechanically broken as explained in section 2.4. The amount of total soluble protein was determined using the Bradford assay as denoted in section 2.4 (Appendix B.3). Homocitrate synthase catalyzes the formation of homocitrate from acetyl-CoA and α-ketoglutarate with the concomitant production of CoASH (Figure 4.11). The majority of procedures used for

assessing homocitrate synthase activity are based on the amount of CoASH formed, which can be detected indirectly using 5,5-dithiobis(2-nitrobenzoic acid) (DTNB). This method has been widely used to determine the function of homocitrate synthases (Gabriel and Milewski, 2016; Wulandari et al., 2002; Zheng et al., 1997). The assay was carried out as detailed in section 2.5 (Appendix B.4). Results from quantification are shown in Figure 4.12. As can be observed, the amount of CoAsH formed (nmol min^{-1} μg $protein^{-1}$) did not change significantly between the transgenic lines (crnifV 1 - 4) and the control (rTN72) regardless of the concentrations of total protein (150 μg - 0 μg) in the cell extracts. As a negative control, reaction mixtures using cell extracts from crnifV1 but lacking AcCoA were set up, they yielded similar values as the control and transgenic lines. These results show clearly that the cell extracts have compounds that are reactive to the components of the DTNB test. Protein purification was envisaged from one of the crNifV strains. Nonetheless, the amounts recovered after purification were not enough to perform further tests, as protein accumulation of NifV is low (Figure 4.9B).

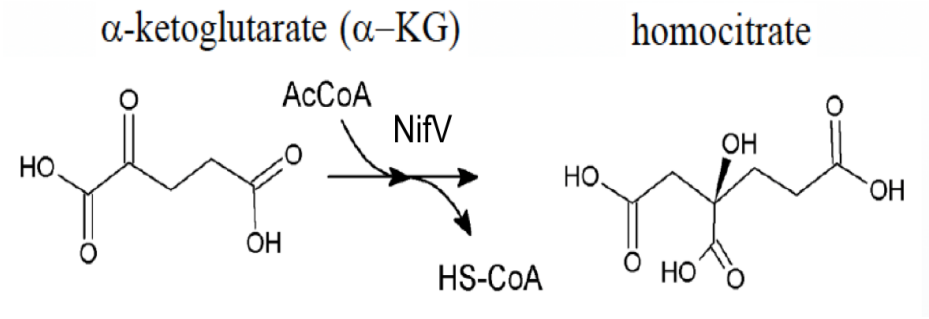


Figure 4.11: Reaction catalysed by NifV. Homocitrate is generated after the condensation of alpha-ketoglutarate with acetyl-CoA. Modified from (Gabriel and Milewski, 2016).

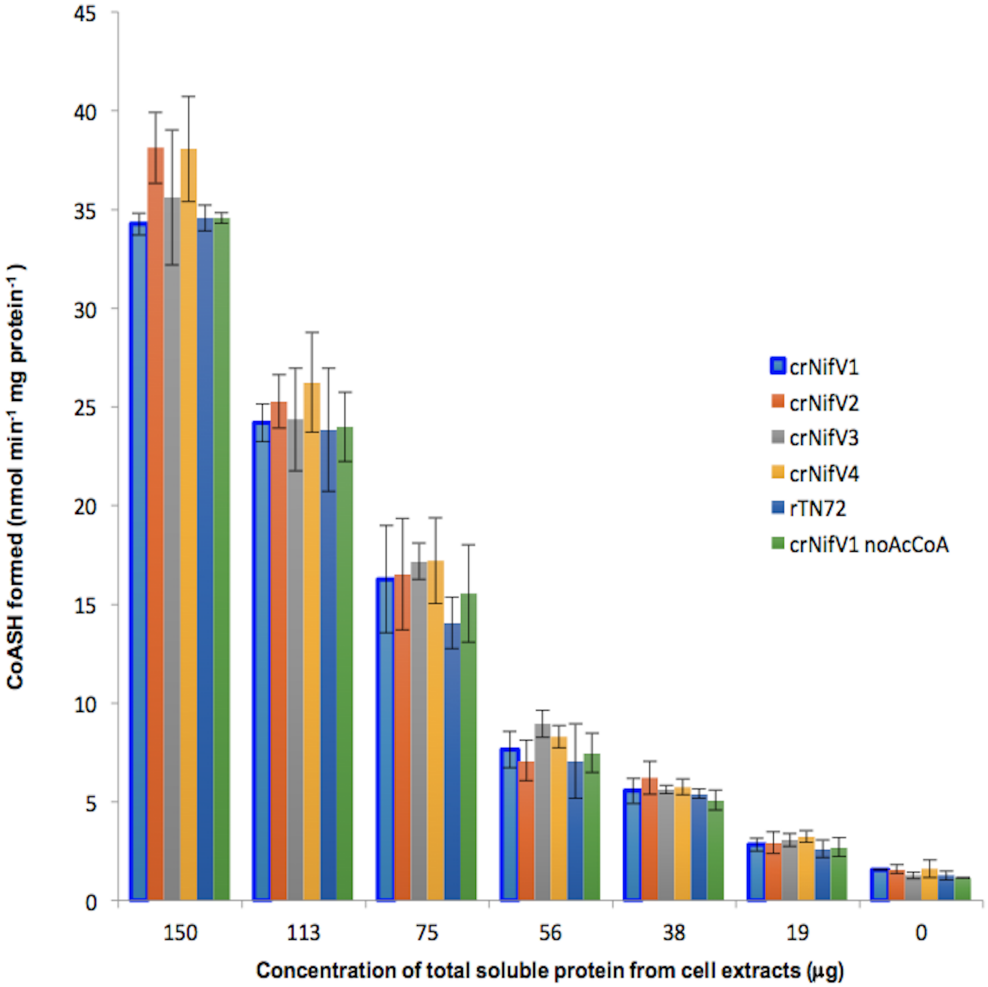


Figure 4.12: Homocitrate synthase activity test in cell free extracts from crNifV-expressing lines. The quantity of CoAsH detected using the DTNB reagent (absorbance measured at 420 nm), which depended on the amount of total protein present in the crude extracts. The rTN72 corresponds to the negative control. This strain has been transformed with an empty expression cassette. Also, another negative control was set up whereby one of the components of the reaction is missing (AcCoA).

4.7 Discussion

Synthetic fertilizers used in both agriculture and microalgal cultivation are expensive and generally produced via the energy-demanding Haber-Bosch reaction. Therefore, finding alternative ways of generating accessible forms of nitrogen is imperative to help develop a more sustainable

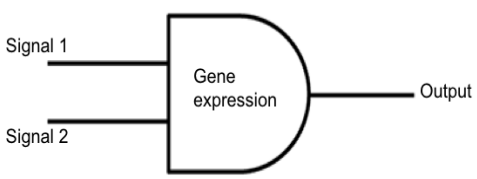
application of these platforms. Biological nitrogen fixation represents a potential option, and eukaryotic organelles have been proposed as potential locations for nitrogenase expression (Curatti and Rubio, 2014; Oldroyd and Dixon, 2014). However, the group of nitrogenase enzymes responsible for atmospheric nitrogen reduction possess several constraints for their proper heterologous expression in eukaryotes. As discussed in the introduction for this chapter, nitrogenases are arranged as binary complexes. Namely, they are made of two reductases that transfer electrons to the substrate. The Fe protein passes the incoming electrons towards the FeMo-co protein, which carries out the reduction of N_2. All the metallic clusters required for transporting the electrons towards N_2 are assembled in scaffold proteins before being finally coupled to the holoenzyme (Swain and Abhijita, 2013; Temme et al., 2012) (Figure 4.2). It has been observed that the number of genes required for diazotrophy depends on the bacterial species. For instance, *A. vinelandii* and *K. oxytoca* require more than twenty coding sequences to make a functional nitrogenase. In contrast, in species such as *C. acetobutylicum* and *Paenibacillus* less than a dozen genes are involved in biological nitrogen fixation (Wang et al., 2013a) (Figure 4.3). Similarly, the genes involved in nitrogen fixation do not only encode the canonical version that has a molybdenum atom in its catalytic site, but they also encode alternative versions. These variants are characterized by the presence of different atoms (V, Fe) in the nitrogen-reducing cluster. The gene clusters encoding these alternatives have been named accordingly: *vnf* and *anf* (Yang et al., 2014a; Lee et al., 2010). The information on how the natural process of biological nitrogen fixation takes place and what are the genetic elements involved has permitted the heterologous expression of *nif/anf* genes. *E. coli*, primarily, has been used to generate strains that are capable of N_2 reduction to some extent (Yang et al., 2014a; Wang et al., 2013a; Temme et al., 2012). These results have highlighted a core of genes essential for

nitrogenase functionality, which include the structural parts and some elements of the assembly machinery (Figure 4.3). These promising results have prompted the notion of using this group of essential genes to introduce BNF into eukaryotes. Nevertheless, these enzymes require environmental conditions that seem hard to find in eukaryotic expression platforms. For example, they require high levels of energy and reducing power as well as the presence of the proper atoms for cluster formation. These situations seem complicated but may be manageable to some extent. The major consideration involves the enzyme's liability to molecular oxygen, which render its function in aerobic systems virtually impossible. Despite this adverse scenario, some authors have proposed the compartmentalization of the enzyme as a conceivable solution (Curatti and Rubio, 2014). Organelles could provide an adequate niche for BNF. Oxygen-sensitive enzymes are known to function in both chloroplasts and mitochondria; moreover, the Fe-S assembly machinery of these organelles resembles the one in charge of nitrogenase formation (Lill, 2009). For example, the sequence encoding the Fe protein has been functionally expressed from the plastome of *C. reinhardtii* (Cheng et al., 2005), which demonstrates that the chloroplast does indeed provide some building blocks for the parts of the enzyme to function. Furthermore, both genetic and metabolic characteristics that are present in *C. reinhardtii* make this green alga particularly attractive for BNF study within its plastid. Therefore, I decided to employ this expression platform in an attempt to help develop of a technology that could be used ultimately to directly incorporate atmospheric nitrogen into algal biomass. *C. reinhardtii* offers a suitable option on the basis of its trophic plasticity and on the advances made regarding transplastomic engineering. First, I decided to rationally analyse the advantages present in this platform, which could let us explore the integration of the nitrogenase pathway. Mainly, how the cell could provide the necessary conditions for having a stable heterologous enzyme within

the chloroplast. Second, I categorized a minimal set of genes that would be required to bring BNF to this green alga. Finally, I have proposed an approach to study individually some elements of the nitrogenase biosynthetic machinery in the chloroplast. As reviewed before, nitrogenase sensibility to oxygen is the main obstacle for coupling this trait into aerobic organisms. In nature, nonetheless, some organisms have adapted these pathways into a single cell. Cyanobacteria, for instance, could use either temporal or spatial separation to avoid the interference of oxygen in the reduction of dinitrogen (Zehr, 2011). Both processes benefit from a reduced action of the PSII, which diminishes the amount of molecular oxygen generated from water splitting. Obviously, the host must be able to grow under dark anaerobic conditions. In this regard, *C. reinhardtii* fits perfectly, since it is able to grow in dark anaerobic conditions and power the function of several oxygen-sensitive enzymes (Moser et al., 2013; Grossman et al., 2011; Reinbothe et al., 2010). This is mainly due to the various fermentation pathways within the algal chloroplast. Furthermore, PSII-deficient mutants have been shown to be viable if maintained on acetate (Harris et al., 2009). Nonetheless, energy levels are importantly reduced and growth is slower when cells are deprived of light. This represents a pivotal problem for putting such a demanding enzyme to work. Hence, it would be imperative to develop genetic circuits that allow the downregulation of the PSII and the concomitant activation of the nitrogenase genes.

Logic gates have been widely used for creating biological circuits, these circuits rely on sensing signals whose presence could define a particular state, which could be denoted in binary (e.g. 1 or 0). These strategies have been of significant importance for developing metabolic engineering in platforms like *S. cerevisae* or *E. coli*. However, these technologies are not yet available in *C. reinhardtii*. Recently, it has been proposed that an OR logic gate could increase the production of triacylglycerol (Scaife et al., 2015). Analogously,

a logic gate device could be hypothesized that would permit the production of NH_3 from N_2. For such a purpose, an AND logic gate seems suitable. The structure of an AND gate comprises two inputs and one output, whose values could be defined as 1s or 0s. The outcome of this system depends on the state of the inputs. Namely, to obtain the desired output is necessary that the two inputs are in the same state (state: 1). If any of them is in a different state (state: 0), the outcome will not be produced (Miyamoto et al., 2012). In this case, NH_3 is considered the outcome and the inputs the signals that will activate a system that induces the expression of nitrogenase genes along with repression of others, this activation will ultimately produce a proper environment for N_2 reduction (Figure 4.13).



Signal 1	Signal 2	Output	
0	0	0	Normal conditions
1	0	0	*nif* genes expression
0	1	0	Keep anaerobic conditions
1	1	1	N_2 fixation

Figure 4.13: Schematic representation of a two-input AND logic gate. When information reaches or leaves the gate the values are stated as 1. If nothing is sensed or released from the gate the values are denoted as 0. In order to have the desired outcome the two inputs must always be present. A hypothetical nitrogen fixation-controlling scheme by an AND Boolean logic gate for the chloroplast of *C. reinhardtii*. Here, the presence of the two signals (e.g. Low oxygen levels and higher incubation temperatures) will trigger N_2 reduction by the heterologous nitrogenase.

Evidently, before applying these kind of systems, it will be necessary to have at our disposal a diverse inventory of tools that permit controlled expression of transgenes. Despite some existing methods (Purton et al., 2013), there is still a need for a simple, highly reliable technique. Nevertheless, for the purpose of evaluating the proposed logic gate, I am going to make use of conjectural examples for controlled expression of transgenes. The use of synthetic anaerobic promoters, for instance, would allow the activation of genes when levels of oxygen drop significantly. If the *nif/anf* genes were put under

the control of these promoters, the lack of oxygen would trigger nitrogenase expression and thus could be considered as the first signal (Figure 4.14). Nonetheless, oxygen will still be generated by the splitting of water taking place in the oxygen evolving complex of the PSII. One direct solution would be growing the cultures in the dark to stop the activity of the photosystems. However, this situation would entail a reduction of ATP synthesis fuelled by light. Hence, less chemical energy would be available for nitrogen fixation. One approach to circumvent this energetic constraint would be to imitate the strategy used by cyanobacteria when fixing nitrogen. Once the cell has turned into a heterocyst, whereby a thick wall has appeared blocking the entrance of O_2, it degrades the PSII to assure the maintenance of anaerobic conditions, and still generate ATP from light via the process known as cyclic photophosphorylation. Energy is captured by the antenna complex and passed to the PSI reaction centre. From this point electrons are passed via a ferrodoxin and a plastoquinone to the cytochrome *b6f*, then the electrons are returned via a plastocyanin to the PSI. The cytochrome allows the passage of protons towards the thylakoid lumen, which generates a gradient that fuels the ATP synthase (Allen, 2003) (Figure 4.15). Therefore, a signal is needed that allows the repression of a functional PSII. Recent work in our laboratory has revealed interesting results regarding a methodology, based on temperature, that permits the regulated expression by means of altering translation of desired genes. Hence, temperature could be used to repress genes encoding components of the PSII, and could be considered as the second signal of the system (Figure 4.14). Consequently, our logic gate would comprise two signals, being the lack of oxygen, which would turn the *nif* genes on, and a change in growth temperature, which will maintain anaerobic conditions. Under these circumstances, ammonia production (outcome) would be possible by means of nitrogenase, which would not be inhibited by O_2 and it could be powered by the ATP produced by cyclic photophospho-

rylation (Figure 4.14). Admittedly, current techniques do not allow for such type of research, but it is always constructive to postulate possible synthetic biology approaches. In fact, this idea could be extended for production of other bio-commodities such as H_2. For instance, some of the methodologies involve the down regulation of PSII to keep anaerobic conditions and enhance biohydrogen production (Dubini and Ghirardi, 2015). In this example, genes encoding the hydrogenase enzyme could be activated by anaerobic conditions (signal 1), which will be maintained by means of PSII degradation that is under temperature control (signal 2). Once these conditions are met, hydrogenases would be capable of reducing protons (output).

The fact that reduced synthetic gene clusters could be employed to bring nitrogen fixation to non-diazotrophs shows that Nif proteins could interact with the native machinery of the host (Yang et al., 2014a; Wang et al., 2013a,b; Temme et al., 2012). The NifH protein was proved to be functional in the chloroplast of *C. reinhardtii*, which showed that the Fe-S clusters that make part of it were assembled and coupled to the final holoprotein by means of the chloroplast's native machinery (Cheng et al., 2005). This could be explained on the basis of a plastid system in charge of assembling these atoms into clusters which are passed then to the final structure. Such a system has been named the SUF system and is schematically depicted in Figure 4.4. It parallels the machinery of the NifU/S system (Figure 4.2). Evidently, the genes *nifU* and *nifS* are omitted from the group of essential genes needed to N_2 fixation. Moreover, the chloroplast contains many enzymes that carry Fe-S clusters in their catalytic sites, demonstrating the availability of Fe atoms, which would suit perfectly for an Fe only nitrogenase. Taking into account that synthetic fertilizers are used for microalgae cultivation, incorporating atmospheric N_2 directly into algal biomass would significantly reduced the need for this costly component. Due to the genetic and metabolic features of *C. reinhardtii*, it could be argued that it would fit

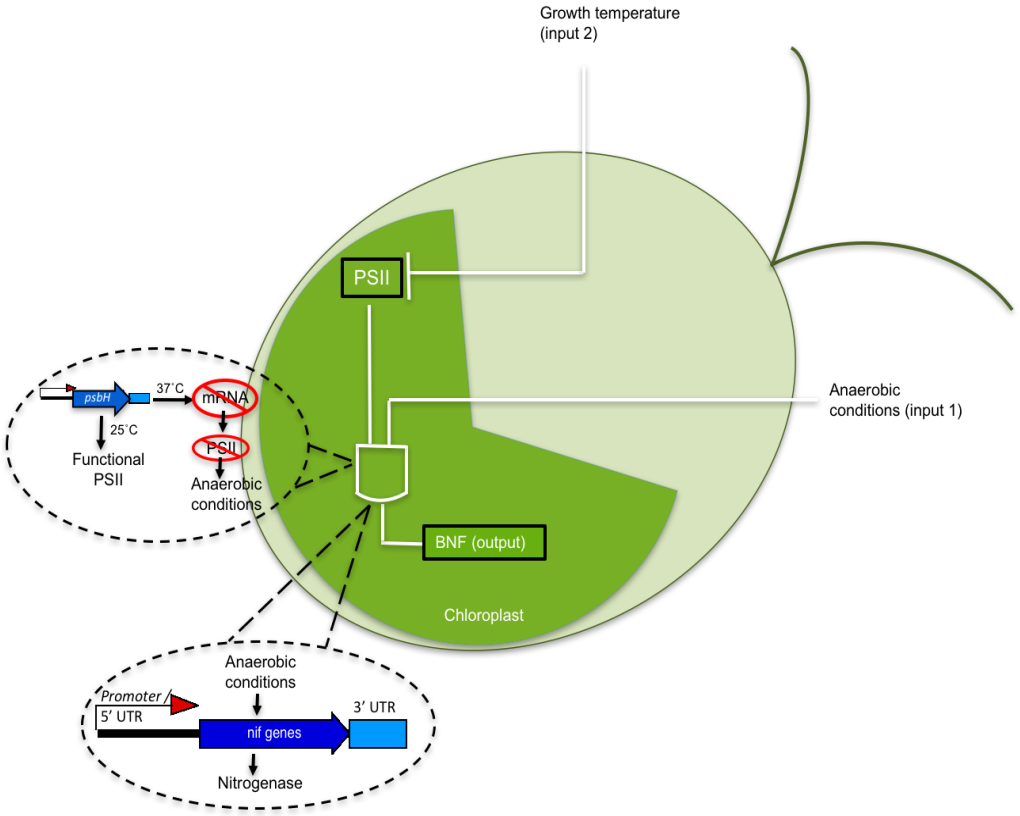


Figure 4.14: Schematic representation of an AND logic gate for biological nitrogen fixation in the chloroplast of *C. reinhardtii*. Ammonia production from heterologous nitrogenase in an input-controlled manner through an AND logic gate. Anaerobic conditions and temperature regulation (inputs) trigger the expression of *nif* genes and the down regulation of the PSII (for eluding production of O_2 from H_2O), respectively. Both inputs must be present for N_2 to be directly fixed. If anaerobic conditions are not met, promoters controlling expression of *nif* genes would not be activated. On the other hand, if these conditions are met, *nif* genes are expressed from the plastome. However, even if all the O_2 is consumed from the cultures, more will still be produced by means of water splitting driven by the PSII. Therefore, repressing the PSII is needed to maintain anaerobic conditions, this could be achieved via a system that permits to inhibit gene expression using temperature regulation. Hence, once *nif* genes are expressed (input 1), cells must be put under temperature regulation (input 2), which will assure that the nitrogenase will not be affected by O_2, and that ammonia will be generated from N_2.

as a model system for this sort of investigation. Here, I propose a minimal cluster of genes that could be used as a foundation for a bottom up strategy for engineering BNF into the chloroplast of this green alga. A basic group of genes must be included to assure the assembly of the main components of the enzyme. Genes encoding replaceable secondary characters (e.g. electron

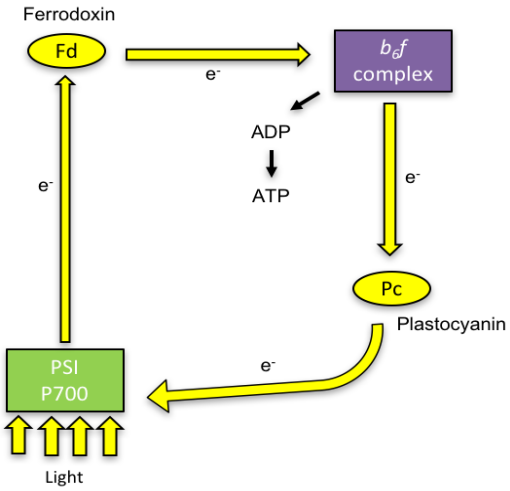


Figure 4.15: Cyclic Photophosphorylation. Light induces the release of electrons by the reaction centre of PSI. Electrons are then transferred to ferrodoxin, plastoquinone, cytochrome *b6f* complex, plastocyanin, and ultimately to P700 again in a cyclic fashion. Electrons lose their potential energy as they move down the electron transport chain, ATP molecules are formed concomitantly (Allen, 2003).

donors) could be added as optimization steps. Moreover, since the chloroplasts do lack important levels of molybdenum, the iron-based version of the enzyme, relying only on iron as part of the reducing cofactor, appears to be the most suitable option. By utilizing genes from both *nif* and *anf* clusters, the expression of this type of enzyme could be brought into the organelle. As could be seen in Figure 4.5, the structural elements of the hypothesized enzyme must be encoded by the *anf* genes, whereas the biosynthetic parts by *nif* genes. One of the main approaches utilized for engineering metabolic pathways is refactoring, which consists of restructuring the informational code (genetic code), without altering the external behaviour of the system (metabolic outcome). This computer science-borrowed principle has been useful to circumvent the genetic complexity involved in typical multigene systems. Evolution has molded native gene clusters to be actively regulated, redundant and particularly efficient in processing information. Hence, it is imperative to insulate the elements to avoid normal regulation, by putting

the *nif/anf* genes under the control of non-native genetic regulatory elements, Table 1.5 lists some of the most utilized for transgene expression in the organelle. Consequently, the genes suggested in Figure 4.5 have to be put under the control of these biological elements. Unfortunately, as discussed in chapter 3, there are no standardized methods for achieving multiple transgene expression in the chloroplast. Thus, I took an approach that involves cloning and investigating these genes individually, and made use of a device that allows the integration of foreign genetic information in the chloroplast genome and its further tracking by photoautotrophy (Figure 3.1). The genes involved in the biosynthetic assembly of the clusters are of fundamental importance for the holoenzyme to function. Thus, I decided to generate codon optimized versions of the *nifB*, and *nifV* genes from *K. oxytoca*, and the *anfH* from *A. vinelandii* to couple them into transformation devices to be introduced into the plastome (Figure 4.7). The *nifB* gene encodes for a SAM enzyme that is in charge of creating the NifB.co cluster, which will be then modified to create the catalytic cluster (Figure 4.2). The purpose of this reaction is to add the carbon atom to the centre of the forming cluster. This is achieved by a SAM-dependent mechanism, this co-substrate will be the source of the methyl group that will be converted into methylene and then further deprotonated; the generated carbon atom will make part of the N_2-reducing cluster (Wiig et al., 2012). The *nifB* gene from *K. oxytoca* was cloned into the plastome. In both aerobic and dark anaerobic conditions the expression of the heterologous enzyme could not be observed at 52 kDa (Figure 4.9A). The *anfH* gene along with *nifH* and *vnfH*, encode the enzymes in command of reducing the catalytic cluster, and is in charge of the ATP hydrolysis that powers ammonia production. (Cheng et al., 2005) demonstrated that the *nifH* gene from *K. oxytoca* could successfully replace the functionality of a similar reductase (ChlL). The codon optimized *anfH* gene was cloned into the plastome, and despite the similarities shown among the

reductases (Figure 4.6), no trace of the protein was observed after Western blot analyses (Figure 4.9C). The results showing that *nif/anf* gene clusters could bring N_2 fixation to *E. coli* have been confirmed, mainly, by identifying the integration and expression (RT-PCR) of the transgenes, along with two common functionality tests: acetylene reduction and $^{15}N_2$ incorporation (Yang et al., 2014a; Wang et al., 2013a,b; Temme et al., 2012). Nonetheless, only some of these demonstrate the presence of the proteins by means of Western blot analyses, primarily they make use of antiserum against the NifH protein, which shows clearly the presence of the heterologous enzyme (Yang et al., 2014a; Wang et al., 2013b). They do not make use of foreign tags for protein identification. Since no antibodies raised against these proteins were at our disposal, HA-tags were added to the C-terminal end of each coding sequence, as is common in our procedures (Figure 4.7). It could be argued that this additional group of amino acids might disrupt proper folding of the proteins, and target them to the active proteolytic system that has been observed to eliminate defective proteins (Rosales-Mendoza et al., 2012). Nevertheless, it could be suggested that codon optimization of the sequences may bring some disruptive effects on protein expression. Codon composition have been observed to affect translation efficiency by influencing the secondary structure of mRNA, Equally, translation elongation and protein folding have been seen to be induced by codon arrangement (Novoa and de Pouplana, 2012; Spencer and Barral, 2012). Therefore, for metabolic engineering purposes the standardization of desired genes is imperative. In other words, as could be observed, it would be preferable to avoid using tags for protein (NifB/AnfH) detection. Comparably, it would be advisable to test both optimized and non-optimized versions of the coding sequences.

The *nifV* gene that encodes a homocitrate synthase, which is in charge of generating the organic moiety present in the N_2 reducing cluster (Figure 4.2) (Figure 4.5). Homocitrate contributes to the stability of the cluster assembly

as well as to its overall redox properties (Gabriel and Milewski, 2016; Wulandari et al., 2002), and thus is considered vital for BNF (Figure 4.3). The *nifV* gene was cloned into the plastome, and the homocitrate synthase could be observed at 45 kDa under anaerobic conditions, as expected (Figure 4.9B). When cells grown in normal conditions were scrutinized, the presence of the NifV enzyme could still be detected (Figure 4.9B). The utilization of foreign promoters 5' UTRs relieves the genes from the action of native regulators. The *nif* system is responsive to fixed nitrogen and oxygen levels via a complex pathway that controls the production of a transcriptional activator (NifA), which itself is oxygen sensitive. This protein functions on the so called upstream activator sequences (UAS) of other *nif* genes (Wang et al., 2013b). Under normal conditions oxygen would disrupt NifA and no other *nif* gene could be activated. In contrast, if *nif* genes are put under control of other promoters 5' UTRs (Figure 4.7B), this regulation by O_2 is omitted and the *nif* genes, *nifV*, could be expressed. Nonetheless, the nitrogenase enzyme is affected by O_2 regardless of the properties of its individual elements.

NifV does not have a major impact on cell density and specific growth rate (Figure 4.10). Crude extracts containing NifV were tested using DTNB, which can detect the CoASH produced from the action of the homocitrate synthase on acetylCoA and α-ketoglutarate. No significant differences could be observed among the extracts from the various strains studied (crNifV1 - 4) (Figure 4.12). Importantly, one of the negative controls (rTN72), which does not carry any heterologous information, showed similar values as the tested cell lines. This reflects the possible presence of thiols groups in the extracts, which react to the DNTB. Purification was tried as a supplementary option. Nonetheless the amount of protein recovered was insufficient to perform any further tests.

Demonstrating the expression of the essential nitrogenase genes set im-

portant foundations for a further coordinated expression of each protein of the pathway. Envisaging possible limitations from single expression experiments would help to develop more accurate strategies to couple this pathway into the chloroplast metabolism. Here, I have described a hypothetical minimal set of genes that could be used to assemble a nitrogenase in the chloroplast. I have also demonstrated that a codon optimized version of homocitrate synthase (*nifV*) from *K. oxytoca* could be successfully cloned and expressed from the organelle's genome. Moreover, the recombinant protein does not seem to have an important impact on cell growth. Further steps must involve other elements of the suggested minimal cluster of genes. Evidently, when molecular tools for attaining simultaneous gene expression become available for *C. reinhardtii*, the building up of the pathway and the determination of protein activity *in vivo* must be assessed. The requirement of low-cost and especially safe expression platforms is regarded as essential for industrial biotechnology. Envisaging ideas to reduce nitrogen inputs from chemically synthetic sources could help reduce the cost of algae cultivation. The green alga *C. reinhardtii* has suitable characteristics on the grounds of growth rate, containment and minimal requirements for growth. Furthermore, the bacterial-like environment of the chloroplast provides some of the fundamental building blocks required to put the binary system of nitrogenase to work.

Chapter 5

Expression of a cyan fluorescent protein in the plastid of *C. reinhardtii*

Many genes can be exploited as reporters and have been widely used for unravelling various aspects of cell biology in a wide range of organisms. Particularly, the utilization of fluorescent proteins has influenced positively many areas of life sciences. Among the vast list of applications are the study of protein localization (Giepmans et al., 2006), promoter function (Zaslaver et al., 2006), development of biosensors (Ibraheem and Campbell, 2010) and cell-cycle progression (Pauklin and Vallier, 2013). Some reporters have been adapted successfully for expression in plant platforms. These reporters include luciferases (Chen et al., 2008), along with green fluorescent proteins (GFP) and derivatives (Wei et al., 2011; Saito et al., 2009; Shcherbo et al., 2007; Seibel et al., 2007; Buhot et al., 2006). In contrast, the development of fluorescent reporters for microalgae has not been as successful as in other biotechnological platforms. GFP was the first fluorescent protein expressed from the nuclear genome of *C. reinhardtii* (Fuhrmann et al., 1999), and more

recent research has demonstrated the viability of expressing various fluorescent proteins from the nuclear genome of this green alga. The fluorescent signal of cyan, green, yellow, and red variants were detected, with the fluorescent proteins localized to the cytoplasm (Lauersen et al., 2015; Rasala et al., 2013). Significant effort has been made to express these reporter genes also in the algal chloroplast. Codon optimization and various 5‘ UTRs have been utilized to improve expression. However, these attempts have been met with limited success. The main proteins studied have been GFP and luciferases (Barnes et al., 2005; Franklin et al., 2002; Minko et al., 1999). Recently, a coral vivid Verde Fluorescent Protein (VFP) has been investigated in the organelle. Despite being detected by Western blot analysis, the fluorescence levels detected were notably low. The complications of working with these proteins in the chloroplast might be related to autofluorescence interference or quenching of the signal mediated by the various pigments present in the algal cell (Braun-Galleani et al., 2015). Recently, the livewort *Marchantia polymorpha* has been put forward as an important basal plant model. Particularly, the characteristics of its plastome along with the relative ease of obtaining transplastomic lines have made this organism relevant for chloroplast-related studies (Bock, 2015; Ueda et al., 2014; Ohyama et al., 2009). Recent research has demonstrated that the complications of working with fluorescent proteins, such as the GFP, in the chloroplast could be circumvented by using a variant called cyan fluorescent protein (mTurq2cp), which was designed using the amino acid sequence of mTurquoise2 as a basis. This protein's range of emission is within the blue-green spectrum (Figure 5.1). The sequence encoding the mTurq2cp was fused to endogenous *cis* elements and inserted into the plastome of *M. polymorpha* via homologous recombination. The heterologous protein proved to be useful for primary screening of transformants, imaging of plastid structures and evaluation of promoter activity (Boehm et al., 2015). These results are encouraging for

the investigation of this marker as a tool for other photosynthetic platforms. The fact that the emission peak is in the blue-green region would increase the chances that the fluorescence from the protein would not be completely masked by the action of the different pigments present in the cell. I, therefore, decided to study the functionality of the *mturq2cp* gene in the chloroplast of *C. reinhardtii*. I present the design of the plasmid and the insertion of the gene into the plastome, its expression in the organelle and the fluorescence measurements within the blue-green spectrum.

Figure 5.1: Properties of mTurqoise2. A. Normalized absorption and fluorescence emission spectra (λmax= 474 nm). Fluorescence microscopy images of human cell lines (HeLa) expressing mTurqoise2 when fused to an alpha-tubulin (B) and a mitochondrial targeting signal (C) (Goedhart et al., 2012).

5.1 *Plasmid design*

As discussed early, codon optimization has led to a better expression of foreign genes by matching the codon usage distribution found in native genes in the plastome. Hence, in the original study (Boehm et al., 2015), the sequence encoding the cyan fluorescent protein was codon optimized for its expression in the livewort *Marchantia polymorpha*. The mTurqoise2 was chosen as it has a higher quantum yield compared to other fluorescent proteins (Goedhart et al., 2012), thereby expanding the reporter emission range to the blue-green ($\lambda = 474$ nm). I decided to test if this reporter was able to work in the chloroplast of *C. reinhardtii*. The sequence encoding the reporter was kindly provided by Jim Haseloff from the University of Cambridge. I decided to use the optimized version for *M. polymorpha* due to the similarities of AT-GC distribution of the two plastomes. The *mturq2cp* gene was amplified from the plasmid pCS CL0*b (Boehm et al., 2015), and an HA-tag was added for immunodetection. The primers used bear *Sap*I and the *Sph*I sites at the 5‘ and 3‘ end, respectively (Appendix C.1). The sequence was cloned into the pSRSapI vector, which uses the *psaA exon 1* promoter / 5‘ UTR and the *rbcL* 3‘ UTR for gene expression (Appendix C.2) (Young and Purton, 2015, 2014). Subsequently, the whole expression cassette containing *mturq2cp* was amplified using primers carrying the *Xma*I restriction sites (Appendix C.1). The resulting amplicon was introduced into the pA plasmid (Figure 5.2A) which contains the aminoglycoside -3-adenyltransferase (*aadA*) expression cassette for selection, this vector targets the *chlL* gene which encodes the reductase element of the dark-operative protochloprophyllidae (Pchlide) reductase (DPOR) enzyme responsible for chlorophyll production in the absence of light. The resulting device was named pAT and allows delivering of both the selectable marker and the reporter to the plastome of *C. reinhardtii* (Figure 5.2) (Appendix C.2).

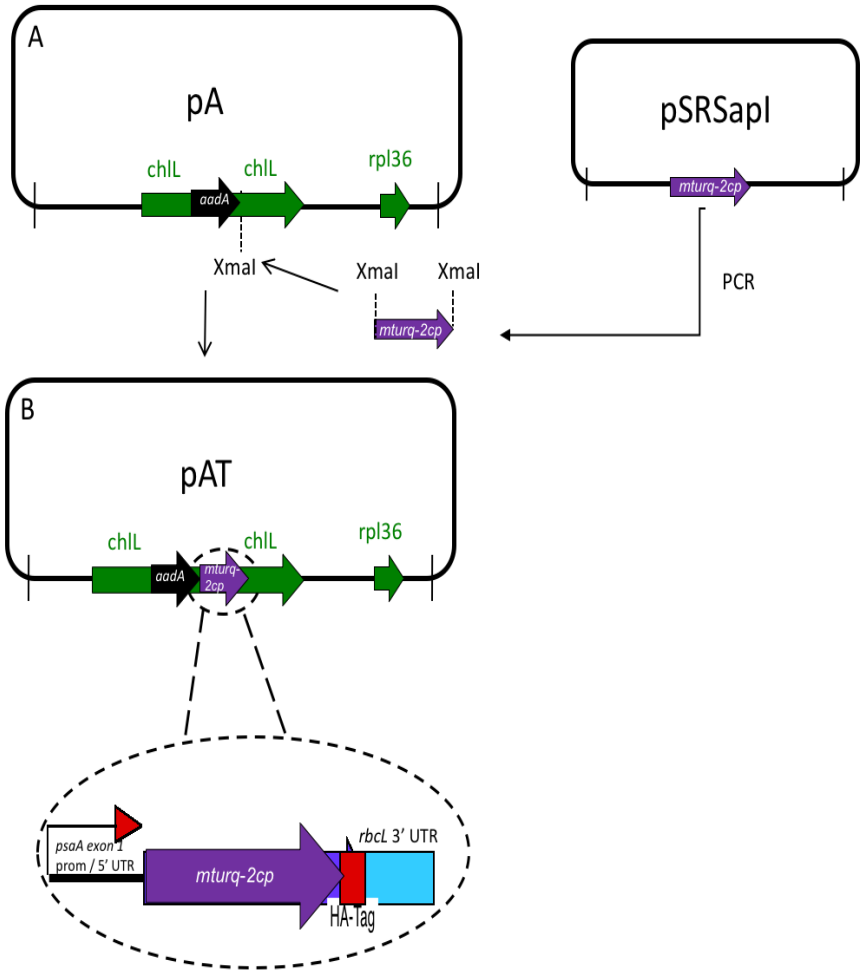


Figure 5.2: Plasmid assembly for cloning the *mturqcp* gene. The pA plasmid was generated by introducing the sequences, belonging to the *chlL* gene, into the pJet vector. The pA plasmid contains the *aadA* expression cassette (black) A. The resulting plasmid was named pA, and was later digested to serve as a backbone for the mTurqcp expression (purple) (accession number: KT364744). The cassette was amplified with primers bearing *Xma*I sites. After enzymatic digestion and ligation the pAT plasmid was created, which bears all the required information to insert the foreign information into the plastome. The gene of interest is under the control of the the *psaA exon 1* promoter / 5‘ UTR and the *rbcL* 3‘ UTR. All plasmids were checked by sequencing.

5.2 *Transformation of cw-15 strains with the mturq-2cp-containing plasmid*

The pAT plasmid was used to transform the plastome of *C. reinhardtii* using the glass bead method described in section 2.3. Selection was on specti-

nomycin, and surviving colonies were spotted on TAP agar and grown in the dark (Figure 5.3). As mentioned before, the construct was directed to the *chlL* gene, which encodes a reductase element that makes part of an enzyme responsible for producing chlorophyll precursors in the dark. Therefore, if the foreign DNA was correctly inserted, colonies should have a yellow phenotype when grown in the dark. Therefore, by growing the transformants in dark it is possible to reduce the background of false positive spectinomycin transformants (Figure 5.3B). Four colonies were selected for further analysis, homoplasmicity was assessed by PCR (Appendix C.1). Primers were designed so if a correct integration has happened, the amplicons produced from the transformants and the controls must differ in size. A correct incorporation yields a 1.3 kb amplicon. Whereas with the untransformed plastome, an amplicon of 1.0 kb is produced. As can be seen in Figure 5.4 the three transformants analyzed show correct integration and appear to be homoplasmic.

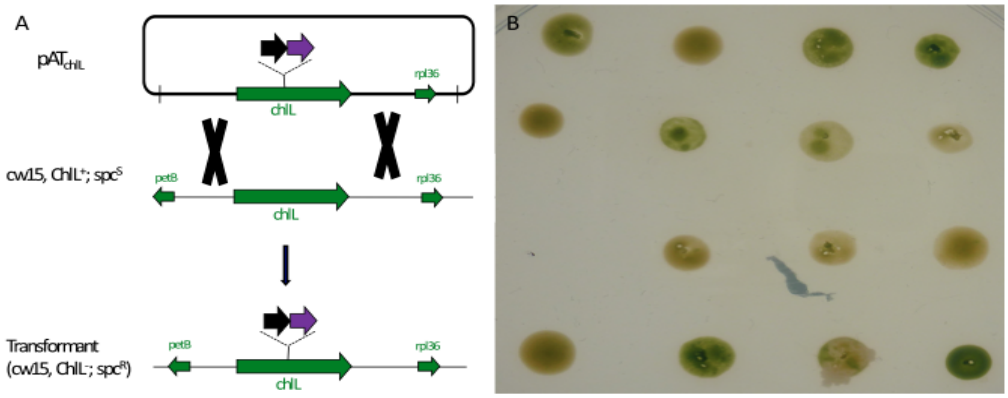


Figure 5.3: Chloroplast transformation approach for expressing the *mturqcp* gene from the plastome of *C. reinhardtii* A. The pAP vector has been manufactured in a way so it contains sequences encoding biological functions required for integrating foreign sequences into the plastome. The *aadA* gene, which confers spectinomycin/streptomycin resistance was used for transformants selection (black arrow). The purple arrow represents the expression cassette containing the gene encoding the mTurqcp2. Both the expression cassettes are flanked by the sequences needed for homologous recombination. B. Transformants were spotted in TAP agar containing the proper antibiotic and were grown in the dark, positive colonies display a pale phenotype, whereas false positives maintain the ability of producing chlorophyll in the absence of light.

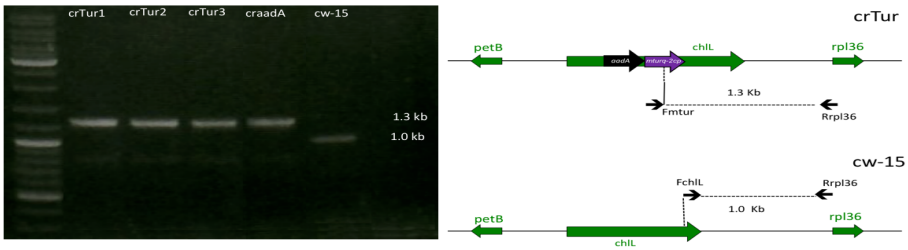


Figure 5.4: Identification of transgenic lines by colony PCR. Left-hand side: Agarose gels showing the PCR products from transformants lines and controls. Right-hand side: Diagrams illustrating the genetic arrangement of the cw-15 genomes and the transformants genomes. The Fmtur primer is targeted to the *mturq2cp* gene, and the FchlL primer to a region in *chlL* that would not be present if transformation was successful. Whereas, the reverse primer targets a region of the *rpl36* gene. The 1.3 kb amplicon demonstrates the correct integration of the two expression cassettes corresponding to the selectable marker (black) and the *mturq2cp* gene (purple). The crA lane corresponds to the positive control, this strain has been transformed with a vector containing only the selectable marker. The 1 kb amplicon corresponds to the recipient strain cw-15, which was used as a negative control.

5.3 *Expression of mTurq2cp from the plastome of C. reinhardtii*

The mTurq2cp protein was detected in crude extracts of the transformant lines by immunoblotting using antibodies targeting the HA-Tag epitope (section 2.4). As shown in Figure 5.5, the cyan fluorescent protein could be detected at the expected size (28 kDa) in all transformants. Controls were also run. Untransformed cw-15 was used as a negative control and a cell line expressing an HA-tagged endolysin (Cpl-1) was used as positive control.

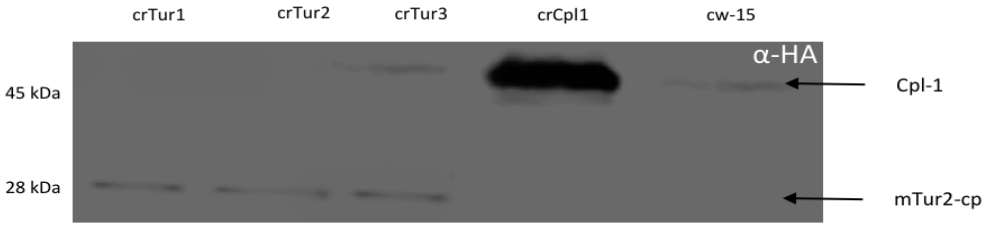


Figure 5.5: Western blot analysis of the transgenic lines bearing the *mturq2cp* gene. Cell lysates were separated by SDS-PAGE gels, blotted and incubated with α-HA-Tag antibodies. The three transformants analyzed have the *mturq2cp* gene integrated and expressed into the plastome at 28 kDa. The crCpl1 strain, expressing an endolysin against *S. pneumoniae* (45 kDa) was used as positive control for α-HA blotting. The cw-15 was used as a negative control.

Transformant line crTur1 was subsequently used for studying the accumulation of the fluorescent protein at different stages of cultivation. Cells were grown as described in section 2.1, with samples taken each 24 hours. Western blot analysis revealed similar levels of protein accumulation in the crTur1 strain through the exponential and the stationary phase (Figure 5.6A). The crTur1 strain was used to assess the impact on specific growth rate of having the mTurq2-cp accumulating in the organelle. As show in (Figure 5.6B), growth rates of the strains carrying the cyan fluorescent protein do not seem to differ significantly to the control, although, final

cell concentrations seem to be slightly lower in the crTur1 strain. This implies that having this fluorescent protein in the organelle do not drastically affect growth rate.

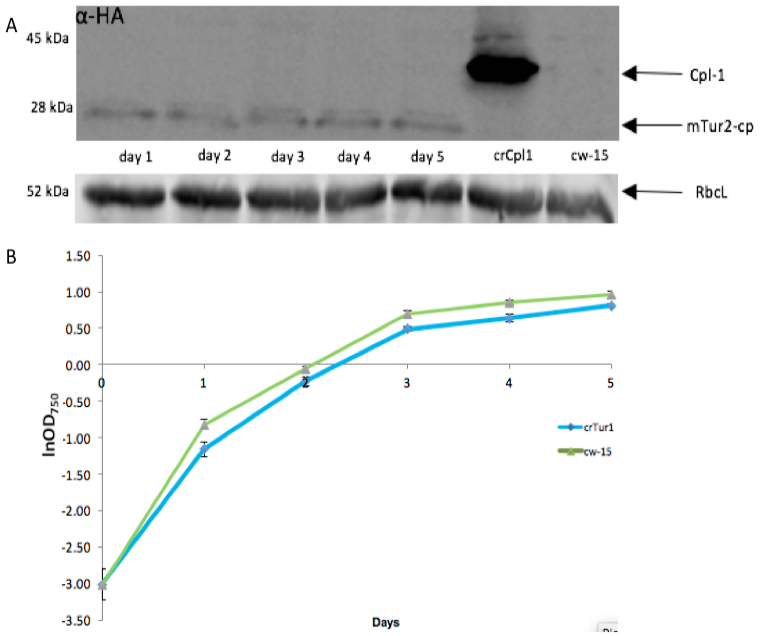


Figure 5.6: mTurq2-cp accumulation, fold increase in fluorescense, and effects on growth in the crTur1 strain A. Accumulation of the mTurq2-cp cyan fluorescent protein at different stages of cultivation. Cell lysates were separated by SDS-PAGE gels, blotted and incubated with α-HA-Tag antibodies. As a positive control, a transformant expressing a 45 kDa HA-tagged protein (crCpl-1) was used, as a negative control, a cell wall-less strain was employed (cw-15). The large subunit of Rubisco was utilized as a loading control (52 kDa). B. Effects on growth of mTurq2-cp. Accumulation of the cyan fluorescent protein does not have a negative effect on growth. The crTur1 strain and the control cw-15 were grown mixotrophically. The data presented are averages of three independent measurements.

5.4 *Fluorescence microscopy of the crTur1 transformant*

Fluorescence microscopy was chosen to determine if the mTurq2cp could be detected using live-imaging techniques (section 2.5). Cells from the crTur1 strain were grown in TAP until mid-log phase. They were then spotted onto TAP agar (1% (w/v)) for image acquisition. Disappointingly, the fluorescent protein was not detectable in live cells. Figure 5.7 shows the chloroplast-localized fluorescence at the cyan fluorescent protein (CFP) gain settings (excitation: 458 nm / emission: 475 - 499 nm) in the transformant line. No differences could be observed when the wild type was analyzed under the same conditions (Figure 5.7). Taken together these data show that when expressed from the plastome of *C. reinhardtii* the mTurqcp accumulates in the organelle but the signal is not sufficient to be distinguished from the background using live-imaging techniques.

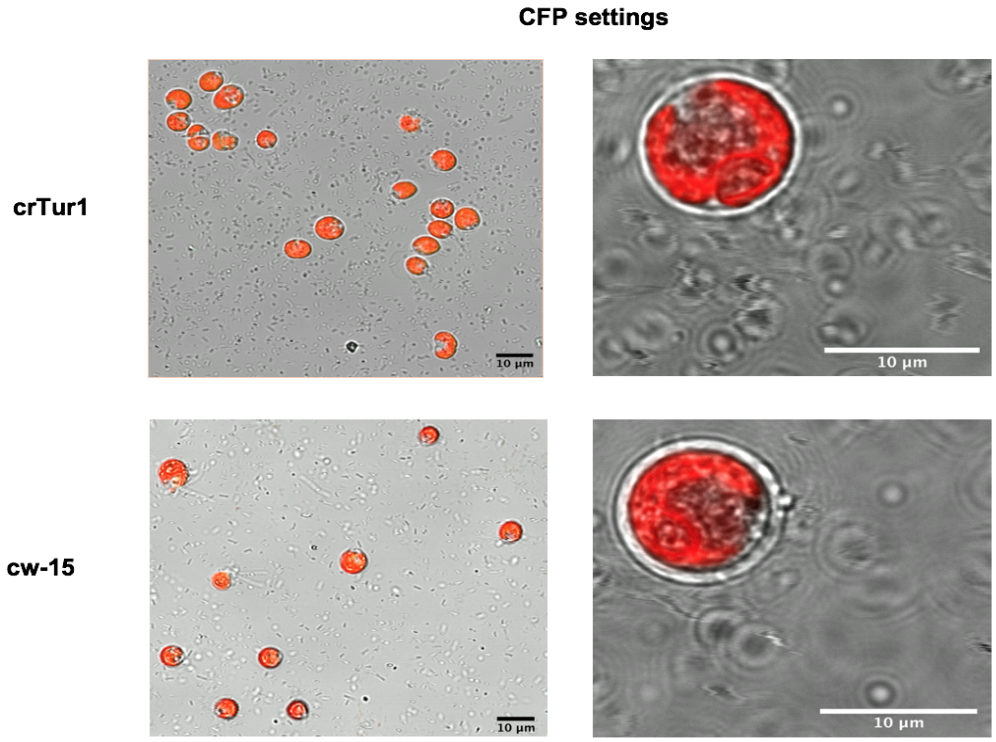


Figure 5.7: Live-cell fluorescence confocal microscopy of the mutant and control strain. Cells were spotted on 1% TAP agar and were subjected to confocal microscopy. Images were taken under CFP channel emission settings. Scale bar = 10 μm .

5.5 *Discussion*

The usage of fluorescent proteins has helped substantially to unravel various processes of biological systems such as gene expression, localization and interaction of proteins as well as cellular dynamics (Pauklin and Vallier, 2013; Ibraheem and Campbell, 2010; Giepmans et al., 2006; Zaslaver et al., 2006). The first fluorescent protein expressed in *C. reinhardtii* was GFP and recent research has demonstrated the successful expression of a palette of fluorescent proteins emitting from the blue to the red window of the spectrum (Lauersen et al., 2015; Rasala et al., 2013) Notably, the sequences encoding the fluorescent proteins were inserted into the nuclear genome and, hence, their products were localized in the cytoplasm or targeted to the flagella.

The application of these proteins when expressed from the plastome of *C. reinhardtii* has not achieved the same level of success. GFP and luciferases have been studied with moderate results (Barnes et al., 2005; Franklin et al., 2002; Minko et al., 1999). Moreover, a recent report has revealed that a vivid Verde Fluorescent Protein (VFP) from the coral *Cyphastrea microphthalma* could be successfully expressed from the plastome, but the levels of fluorescence were markedly low (Braun-Galleani et al., 2015). Nevertheless, a cyan fluorescent protein (mTurq2cp) have been successfully expressed from the plastome of *M. polymorpha*, and has been utilized for assessing promoter activity, screening transformants and imaging plastid structures (Boehm et al., 2015). Versions of cyan fluorescent proteins have been successfully expressed from the nucleus of *C. reinhardtii* but no records of their expression from the organelles genome have been reported. Therefore, the results acquired from *M. polymorpha* are encouraging for developing a similar approach for the chloroplast of *C. reinhardtii*. Thus, I decided to test if this protein could be expressed and its fluorescence measured when expressed from the plastome of this green alga. The gene was targeted to the *chlL* locus. Three pale spectinomycin-resistant colonies were picked and subjected PCR examination (Figure 5.4). Western blot analyses reveal that the cell lines were capable to accumulate the heterologous protein, which could be observed at the expected size of 28 kDa (Figure 5.5). The next goal was to determine if the were any difference in mTur accumulation at different growth stages. Recombinant protein levels appear to be independent of the growth stage. Namely, the intensity of the band does not seem to vary at different cell concentrations (Figure 5.6A). This has been observed to happen to another fluorescent protein expressed in the chloroplast (Braun-Galleani et al., 2015).

In order to determine the fluorescence emission from mTurq2cp, fluorescence confocal microscopy was used . Disappointingly, no differences in signal could be perceived between the analyzed transformant and the cw-

15 strain (Figure 5.7). Confocal microscopy is known for generating vastly detailed fluorescent images of cells. However, absolute fluorescence sensitivity is reduced since the image is created by utilizing measurements of each location across the cell and, thus, the total time for detecting the signal is diminished. This methodology is compatible with applications that involve visualization of sub-cellular features but is not well-suited for detecting faint signals (Basiji et al., 2007), which explains to some extent the reason why the difference between the mutant and the control line could not be perceived. The reasons might be related with autofluorescence interference from pigments present in the cell. On the contrary, when this protein was expressed from the plastome of *M. polymorpha*, the difference in fluorescence between the control and the mutant was significant, so fine tubules extending from plastids were visualized using confocal microscopy. Also, they reported a 4-fold increase in fluorescence above the background (Boehm et al., 2015). Evidently, the thalli analyzed did not show the same level of auto-fluorescence when using the CFP channel emission settings. As mentioned before, another study expressed an mCerulean protein from the nuclear genome of *C. reinhardtii*. The fluorescent protein could be observed in the cytoplasm, and when fused to an alpha-tubulin protein the flagella and some microtubule-like structures were visualized with no apparent alteration of the cell's normal functioning (Rasala et al., 2013). They reported, nevertheless, high levels of auto-fluorescense in the range of the blue/cyan spectrum (450/486), which is similar with what I observed (Figure 5.7); however, their protein is localized in the cytoplasm while ours is confined within the chloroplast and, seemingly, auto-fluorescence comes mainly from the organelle, which harbors a myriad of pigments (Grossman et al., 2004). Thus, it could be assumed that these light-emitting elements might interfere or quench the signal of the studied fluorescent protein. Auto-fluorescence, along with quenching have been argued before as possible reasons why a vivid

Verde Fluorescent Protein (VFP) could not be detected when expressed from the plastome of *C. reinhardtii* (Braun-Galleani et al., 2015). Additionally, the authors advocate possible problems with post-translational formation of the chromophore. I do not discard that these features might influence negatively the activity of this fluorescent protein as well. What seems evident is that the blue/cyan spectrum in *C. reinhardtii* seem to be troubling for the study of fluorescent proteins. Despite the optimistic results obtained in plant models, the level of auto-fluorescence of the algal cell makes it difficult to detect the mTurqcp protein by live-imaging techniques such as confocal microscopy. It has been reported that the levels of auto-fluorescence in *C. reinhardtii* for the blue/cyan spectrum are important. In fact, only the auto-fluorescence in the green region is higher (Rasala et al., 2013). Thus, as has been suggested in the cited paper the region in the red section appears to be a better option for studying fluorescent proteins. Such study was performed using nuclear transformants. Nevertheless, regardless of the drawback for live-imaging techniques, I have demonstrated that the sequence encoding the *mturq2cp* gene could be easily cloned and expressed from the plastome of *C. reinhardtii*.

Recent work in our laboratory has demonstrated that a chimeric promoter, consisting of the *16 S* promoter and the *psaA exon 1* 5‘ UTR, yields higher levels of transgene expression than the native *psaA exon 1* promoter 5‘ UTR. Therefore, a plasmid has been constructed whereby the *mturq2cp* gene is under the control of mentioned chimeric promoter.(Figure 5.8). If higher levels of expression of mTurq2cp could achieved, it is possible that the signal of the fluorescent protein would be detected.

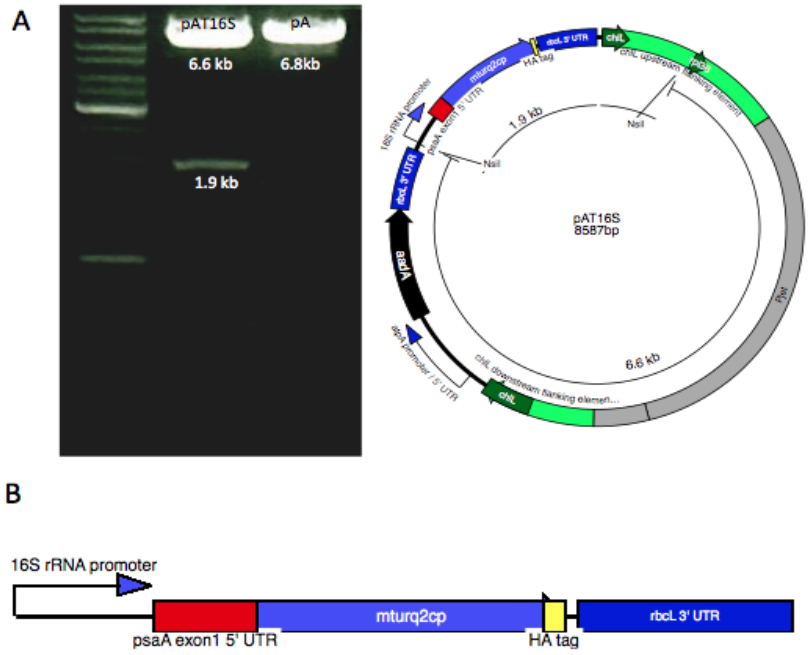


Figure 5.8: Identification of correct plasmid assembly for expression of the mTurq2cp protein under the control of a chimeric promoter. Left-hand side: Agarose gels showing the digestion products from transformants. Right-hand side: Diagrams illustrating the genetic arrangement of the plasmids. A. Proper assembly of the pAT16S plasmid yields a fragment of 6.6 kb and one of 1.9 kb after digestion with *Nsi*I restriction enzyme. The empty vector (pA plasmid) yields a fragment of 6.8 kb. The plasmid has been confirmed by sequencing. B. Schematic representation of the expression cassette containing the mTurq2cp coding sequence. Note that the sequence upstream the coding region is composed of the *16 S* promoter and the *psaA exon 1* 5‘ UTR.

Chapter 6

General Discussion

As presented in the introduction *C. reinhardtii* has been extensively used as a model organism for many years to investigate fundamental questions of cell biology. In more recent years, this green alga has been employed for producing a wide variety of commodities either by manipulating environmental conditions or via genetic engineering. Rapid growth, a short timeframe to produce transformant lines and the easiness to express heterologous genes have made this microalga a promising platform for biotechnology. The chloroplast, especially, has been employed for such purposes since its genetics are extensively studied and transgene integration into the plastome is site-specific. Various technologies have been put forward to deliver DNA into the plastome. Particle bombardment, electroporation or agitation with glass beads have been extensively exploited for transformant generation. Additionally, the utilization of UTR's from highly expressed genes along with codon optimization have boosted recombinant protein expression (Purton et al., 2013). Despite the success of these genetic engineering approaches, metabolic engineering is still in its infancy, there are no current examples of elaborate metabolic pathways engineered in the organelle. Nonetheless,

there are few examples of single metabolic enzymes introduced in the algal chloroplast (Wu et al., 2011; Tan et al., 2007). Recently, is has been suggested that applying synthetic biology principles might accelerate the development of sophisticated tools compatible with the ones available in other established platforms such as *E. coli* and *S. cerevisae* (Scaife et al., 2015). The authors have pointed out some of the relevant terms used in this field (Table 1.4). Having sophisticated molecular tools at our disposal would increase the intricacy and efficiency of the genetic engineering approaches applied to this green alga. Therefore, the current challenge is to identify key issues and develop novel tools and approaches to address them. As mentioned before, simultaneous production of multiple recombinant proteins has not been achieved for more than two genes of interest (Noor-Mohammadi et al., 2012; Tran et al., 2009), which imposes a limitation to metabolically engineer the organelle. Moreover, a comprehensive investigation of a possible metabolic pathway to be engineered in the cell will prove crucial. Defining the key components of the pathway, and inserting them in the plastome will reveal essential information of the plausibility to couple the desired metabolite into the organelle's metabolism. In addition, fluorescent proteins have been proved to be invaluable in cell biology research in general Nonetheless, the expression of this class of proteins in the *C. reinhardtii* chloroplast has been limited in success and with only a few examples. Therefore, testing novel fluorescent proteins will be beneficial for enhancing chloroplast genetic engineering. Thus, in this research I decided to address the aforementioned issues by first, testing different approaches for simultaneously express various recombinant proteins from the plastome. Secondly, evaluating a group of fundamental genes encoding for a nitrogenase enzyme, and, finally, expressing a cyan fluorescent protein from the plastome.

6.1 Attempts to express multiple transgenes simultaneously in the chloroplast of *C. reinhardtii*

6.1.1 Research

I have tested two different methodologies for expressing various heterologous genes simultaneously from the plastome of *C. reinhardtii*. Firstly, I opted for a system whereby poly-proteins could be processed post-translationally into individual subunits. Secondly, I took an approach that involves stacking various expression cassettes in a single transformation vector in a combinatorial fashion. Moreover, I tried also to introduce additional transgenes in a transgenic strain in a serial manner. Finally, I showed that simultaneous expression of transgenes is attainable in the chloroplast.

In the first strategy, I wondered if I could imitate the strategy used by some viruses for producing their proteome. Namely, design a system where a polyprotein could be produced and then processed into subunits by the action of a specific protease. I have shown the design and the assembly of the transformation plasmids, and the analysis of the recovered transformants. As discussed in chapter three, the processing of the generated polyprotein was detectable but did not go to completion. The first module was observed but the second component of the polyprotein could not be detected after protease treatment. Thus, post-translational processing of polyproteins does not seem to be an easily attainable approach for heterologous proteins expressed in the organelle. Furthermore, this approach requires that the subunits making part of the polyproteins can tolerate extra amino acid residues in their terminal regions, which imposes a fundamental constraint to this idea. Secondly, I have shown the design, assembly and simultaneous expression of three transgenes from the plastome of *C. reinhardtii*. The sequences

encoding the IBT-CTB antigen, the cytosine deaminase from *E. coli* and a serine protease from *S. aureus* were assembled together in a single plasmid within their own expression cassettes in a combinatorial fashion. A standard plasmid for single transgene expression was upgraded to contain two more expression cassettes, and the plasmid was demonstrated to be reliable in delivering the genetic information into the genome. The expression of the heterologous proteins demonstrated that clustering three transcription units together does function well for multiple gene expression. Nevertheless, when expression is compared to the cell lines expressing only one gene, it is clear that there is a reduction in the heterologous expression level, which is revealed to be linked to the unstable state of the plastome. The 0.4 kb repeats were observed to promote recombination. The sequences located between these repeats are excised from some of the plastomes, which reduced the copy number of the transgenes. The presence of 5-FU in the culture imposed a selection pressure for the replication of those genomes that do not contain the gene encoding the deaminase enzyme. This particularity opens the possibility to exploit it as an application in which CodA could be used in combination with the 0.4 *rbcL* 3' UTR repeats to selectively removed parts from the genome (e.g. selectable markers).

The cell line expressing three heterologous proteins was used for another round of transformation in a serial approach. Therefore, a novel plasmid was designed containing the sequences for targeting the foreign information to the *chlL* gene. Two additional coding sequences were selected, the *aadA* gene which confers resistance to spectinomycin and the *cpl-1*, which encodes for an endolysin targeting *S. pneumoniae*. These genes were coupled into their own expression cassettes and assembled into the new transformation device, which was shown to be suitable for delivering the additional transgenes into the plastome. Transformants were recovered on spectinomycin, which demonstrated the functionality of the *aadA* gene, and further analysis

demonstrated the expression of the endolysin. These results confirm that the serial approach is practical. However, two of the original heterologous proteins were lost after the second round of transformation, due to the issue of instability discussed above. This drawback could easily be circumvented by utilizing different 3' UTRs. Finally, two different cell lines expressing three unrelated heterologous proteins simultaneously were generated. The effects on growth of the extra protein load were revealed not to be significant. The transgenic lines do lag behind the control in terms of growth. Nonetheless, this difference does not differ from the one obtained from cell lines expressing only one heterologous protein. Conclusively, these results have shown that a combinatorial and serial approach seem to be a worthy route for expressing multiple transgenes simultaneously in the chloroplast of *C. reinhardtii*.

6.1.2 Future work

Despite the limited results obtained with the protease system, the fact that the fusion protein is expressed demonstrates the viability of generating proteins made of multi-subunits from the plastome. The usage of appropriate linkers that allow efficient cleavage by the protease could help producing proteins with various functions, in this way different coding sequences could be connected by linkers and assembled in single expression cassettes, reducing the number of genetic elements used for producing heterologous proteins in the chloroplast. In fact, the notion of a fusion protein has been widely applied for expressing a variety of antigens in the chloroplast of *C. reinhardtii* (Gregory et al., 2013). On the other hand, a fusion protein consisting of a Xylanase and a selectable marker (zeomicin resistance) has been successfully expressed from the nuclear genome (Rasala et al., 2012). A similar approach could be applied for chloroplast engineering whereby a selectable

marker could be linked to another enzyme of interest. The downside of this notion remains the ability of the subunits to handle extra amino acids at their terminal regions.

The combinatorial serial approach seems suitable for multiple transgene expression. The next step would be to standardized the assembly of devices containing various expression cassettes, which could then be directed to various sites of the plastome. An optimal method would be one that allows the modular combination of various genetic elements in a straightforward way, without the necessity of multiple steps of amplification, digestion and ligation. The Golden Gate cloning system allows the straight assemble and exchange of various characterized genetic elements in such a way that transcriptional units are generated and could be eventually combined into one single device. This technology is based on a particular characteristic of the so-called Type IIs restriction enzymes (e.g. *Bsa*I). These enzymes recognize non-palindromic sequences and cut outside the recognition motif (Figure 6.1).

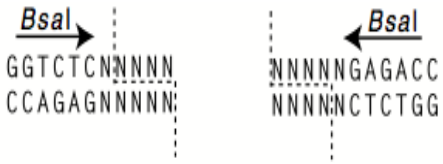


Figure 6.1: Type IIs restriction enzymes. These enzymes (e.g. *Bsa*I) recognize asymmetric sequences and cut outside the their restriction sites. Modified from (Patron et al., 2015).

This fact has allowed their use for engineering purposes. Since the overhangs left by this type of enzyme could be designed at will, it is possible to flank genetic elements with such sequences. After a simple digestion ligation reaction, parts are freed from carrier plasmids and assembled into acceptor ones (Patron et al., 2015; Engler et al., 2009). At least five points have been emphasized as fundamental for this approach: 1: Genetic elements must

be flanked by convenient Type IIs recognition sequences. 2: The Acceptor plasmid must have suitable recognition sites so the genetic element could be organized between them. 3: All genetic parts must be free of recognition sites belonging to the Type IIs enzymes employed. 4: A different antibiotic resistance must be used between the storage plasmids and the acceptor plasmid. 5: The created overhangs must be unique and non-palindromic. Recently, many existing protocols have been adapted to fit within the requirements of the Golden Gate assembly by using tuned overhangs (Engler et al., 2014; Binder et al., 2014; Emami et al., 2013). Various approaches have been demonstrated to be successful in plants. In particular (Weber et al., 2011) have developed the so-called MoClo strategy, which allows the assembly of multigene constructs (Figure 6.2).

Figure 6.2: The modular cloning (MoClo) system. Level 0 encompasses the genetic sequences required for gene expression, these elements are stored in the so-called level 0 modules. Transcription units (TU) are assembled in a one step digestion ligation reaction and stored in level 1 modules. Likewise, multigene constructs are put together by combining various TU in another one step digestion ligation reaction. P1-a/b, promoter fragment a or b; U, 5' untranslated region; SP, signal peptide; CDS, coding sequence; T, terminator; CRed, red color selectable marker; LacZ, lacZa fragment, blue color selectable marker; L2E, end-linker 2; Ap^R ampicillin resistance; Km^R; kanamycin resistance; Sp^R, spectinomycin resistance. Modified from (Weber et al., 2011).

This strategy is divided into different levels (0, 1 and 2). The lowest level

or level 0 corresponds to all the modules that individually contain the basic genetic sequences required for gene expression. Namely, promoters, 5', 3' UTRs, GOI, signal peptides sequences, etc. These sequences are integrated in their level 0 plasmids via a single digestion ligation reaction using a designed Type IIs restriction enzyme (*Bpi*I). These genetic elements could be combined into transcription units in the level 1 modules. Special overlapping sequences were chosen to minimize possible alterations. These transcription units (TU) are assembled via a single digestion ligation reaction using a different Type IIs restriction enzyme (*Bsa*I). All level zero elements of the same kind are flanked by the same recognition sequences so they could be arranged in different combinations. In other words, they are interchangeable so any desired transcription unit could be created. The level 1 plasmids contain particular restriction sites (*Bpi*I) that will allow the assembly of various expression cassettes. The resulting modules are the highest in the hierarchy and have been dubbed level 2 plasmids. These modules bear particular *Bpi*I sites that flank the area where the various transcription units will be assembled, which could be done in various combinations. The three kind of recipient plasmids carry different antibiotic resistance genes for selection: spectinomycin, ampicillin and kanamycin for level 0, 1 and 2 modules, respectively. Additionally, level zero and level 1 plasmids carry the *lacZalpha* fragment for blue/white selection, while the level two plasmid encodes a red color selection based on an artificial operon responsible for producing the pigment canthaxanthin. The functionality of these constructs was tested by expressing the GFP gene in the leaves of *N. benthamiana*. The authors, in this way, demonstrated that this method is highly modular and requires relatively simple steps of digestion and ligation. Furthermore, they emphasize on the easiness to automate such strategy and its useful application for metabolic engineering. In *C. reinhardtii*, a recent approach has been developed whereby a gene of interest could be assembled with different UTRs

and then inserted into the plastome and selected for phototrophic growth. The assembly of the transcription units into the transformation device is mediated by a *Bsa*I digestion ligation reaction (Bertalan et al., 2015). Another toolbox, based on the GoldenBraid method has also been put forward for similar purposes (Vafaee et al., 2014). Despite being useful approaches, they just allow the assembly of level 0 parts into level 1 units, they do not permit the assembly of level 2 multigene constructs, which is imperative for improving metabolic engineering. Therefore, I reason that an approach that matches the characteristics set by (Weber et al., 2011) would be the most suitable for *C. reinhardtii*. Genetic sequences could be designed to match the requirements of level 0 and level 1 plasmids. It is clear that the level 2 plasmids, besides containing the sequences needed for transgene integration into the plastome, must have an algae selectable marker (Table 1.5). As demonstrated in chapter 3, the pASapI-based vector pICS (Figure 3.8) could be used to deliver three independent expression cassettes downstream of the *psbH* gene (Figure 3.9). The simultaneous expression of the heterologous proteins was observed in four different strains (Figure 3.10). Therefore, this vector could be used as a foundation to develop a suitable backbone for level 1 and level 2 plasmids. The next stage of research is to develop a Golden Gate system for multiple transgene expression from the plastome of *C. reinhardtii*. As a starting point, I have created a small library of zero level parts (Table 6.1). I have designed them according to the requirements of the MoClo method (Weber et al., 2011) to maintain a similar format (Figure 6.3). Further work must be carried out in designing the backbone for level one and level two plasmids. If the technique proves routine, then it would help to take transplastomic engineering in this green alga to a higher level.

Table 6.1: Summary of genetic elements encoding biological function used for generating the level 0 plasmids

Element	Function	Size	Reference
psaA exon 1 promoter / 5' UTR	Drives the expression of the gene for photosystem I P700 chlorophyll a apoprotein A1	285 bp	(Young and Purton, 2015, 2014)
psbA 3' UTR	mRNA stabilization of the Photosystem II protein D1 mRNA	346 bp	(Bertalan et al., 2015; Demurtas et al., 2013)
mTurq2cp	Encodes a cyan fluorescent protein	745 bp	This report
petB promoter / 5' UTR	Drives the expression of the Cytochrome b gene	371 bp	This report
rbcL 3' UTR	mRNA stabilization of the Ribulose bisphosphate carboxylase large chain transcript	258 bp	(Wannathong et al., 2016) (Zedler et al., 2015)

Figure 6.3: Level 0 destination vectors used for a prospective Golden Gate approach for *C. reinhardtii*. The devices are generated via a *BpiI*-mediated ligation of amplified sequences and the destination vectors. The pL0-PU plasmid is used to clone promoter / 5' UTR sequences while and the pL0-SC and the pL0-T to clone coding sequences and the 3' UTRs, respectively. Modified from (Weber et al., 2011).

6.2 *Chlamydomonas reinhardtii* as a model organism for a genetic analysis of biological nitrogen fixation in green algae

6.2.1 Research

Synthetic fertilizers are still an important and costly input in microalgal cultivation, some biological alternatives have been proposed to deal with this issue, but none of them allow direct access to the reservoir of atmospheric N_2 (Peccia et al., 2013; Stephenson et al., 2010; Ortiz-Marquez et al., 2012). In order to achieve this, a nitrogenase enzyme has to be assembled in the algal cell. The assembly of this enzyme requires the coordinated expression of various genes. Once the holoenzyme is formed, N_2 can be reduced to NH_3. Nevertheless, the environment required for this reaction to occur is highly demanding. The observed limitations seem to make nitrogenase expression in photoautotrophs very unlikely (Oldroyd and Dixon, 2014; Curatti and Rubio, 2014). However, I have argued that *C. reinhardtii* might

provide the necessary requirements to test the heterologous expression of this enzyme. For instance, nitrogenase requires a completely anaerobic environment. In nature, some organisms have evolved strategies to cope with the two mutually-exclusive reactions of photosynthesis and nitrogen fixation. Cyanobacteria, for instance, use temporal or spatial separation to achieve both photosynthesis and nitrogen fixation (Zehr, 2011; Berman-Frank et al., 2007). Thus, the model organism used to study nitrogenase must be able to sustain growth in the absence of oxygen. In this respect, *C. reinhardtii* shows a wide range of trophic plasticity. As I have summarized, it can be grown anaerobically if the proper conditions are estabilished. Providing enough energy, reducing power and the proper metal atoms for nitrogenase might impose another limitation. The chloroplast, in particular, represents the biosynthetic centre of the cell which harbours demanding reactions requiring high levels of ATP and reducing power. Hence, some of these elements could be commandeered for nitrogenase function. The fact that reduced *nif* clusters have been used to bring nitrogen fixation to non-diazotrophic bacteria, shows that some assembly factors of the host could interact with the Nif enzymes (Yang et al., 2014a; Wang et al., 2013a,b). In fact, NifH from *K. oxytoca* was shown to be functional in the chloroplast of *C. reinhardtii*, which demonstrates that some native machinery present in the chloroplast is able to assemble and then transfer the Fe-S cluster that this enzyme requires (Cheng et al., 2005). In fact, it has been observed that the chloroplast has elements similar to the bacterial SUF (sulphur assimilation) system, which acts in a similar way to the Nif system found in diazotrophs. Essentially, it relies on scaffold and biosynthetic proteins, and maybe be responsible for the assembly of the metal cofactors of some chloroplast enzymes (Godman and Balk, 2008). Therefore, some elements of the nitrogenase pathway could be replaced by the chloroplast native machinery. The constraint of the lack of Mo or V atoms could be circumvented by utilizing genes from the *anf*

cluster. The fact that an iron-only nitrogenase could be heterologously expressed in *E. coli* paves the way for its further investigation in eukaryotes, in particular green algae (Yang et al., 2014a). I have proposed a minimal set of genes, whereby only six coding sequences might be necessary to catalyse the reduction of atmospheric N_2 in the chloroplast. I hypothesize that a mixture of *anf* and *nifH* genes could do the job. The apoenzyme has to be encoded by the *anfHDKG* genes, whereas the essential biosynthetic components have to be encoded by the *nifBV* genes. However, since techniques to metabolically engineer the chloroplast of *C. reinhardtii* are yet to be developed, it is still difficult to test the simultaneous expression of various transgenes. Nonetheless, I decided to investigate individually some key elements of the suggested cluster. As explained before, the sequences belonging to the *nifB/V* of *K. oxytoca* and *anfH* from *A. vinelandii* were inserted in the chloroplast genome. The transformants generated were grown in both aerobic and dark anaerobic conditions. However, when detecting the recombinant proteins, only the NifV was identified. It was unexpected not to detect the NifB and AnfH enzymes. In particular, because the chloroplast does harbour similar elements in its metabolic network. For instance, other SAM-dependent enzymes, like NifB, have been linked to the assembly of metalloclusters related to hydrogenases (Dubini and Ghirardi, 2015; Grossman et al., 2011). Likewise, the nitrogenase reductase element (AnfH) shows high levels of homology with a native chloroplast subunit (ChlL). Thus, I speculate that using an HA-Tag might have disrupted the proper folding of the enzymes and targeted them for degradation. Moreover, codon rearrangement might have brought instability to the secondary structure of the mRNA and thus impair protein production. On the other hand, NifV could be observed at the expected size when cells were grown aerobically and anaerobically. I decided to use one of the strains to determine if the enzyme imposed a load on cell growth, it was observed that transformants

grew slightly slower than the control, but the difference was not substantial. Cell-free extracts were used to test the enzyme *in vitro*. Unfortunately, the measurements obtained showed a signal, but no difference between the transformants and the controls, which indicates that other compounds normally present in the algal cell, is interacting with the test reagents. Protein purification was tried, but insufficient levels of protein were recovered to carry out further tests. Overall, I have presented a suggested minimal set of genes that could be used as a bottom up strategy to engineer a nitrogenase enzyme in the chloroplast of *C. reinhardtii*, and in this way provide a novel alternative to reduce inputs from synthetic fertilizers. I have cloned into the plastome three of the six recommended genes. However, only the expression of the homocitrate synthase (NifV) was detected. These results show that the expression of *nif* genes is attainable, but extensive optimization of its elements must be carried out.

6.2.2 Future work

Different versions of *nifB* should be tested, I used the sequence from the *nif* cluster of *K. oxytoca*, but genes from other diazotrophs have been utilized to heterologously express nitrogenase in *E. coli*. Hence, it is imperative to investigate alternative genes that might be expressed in the chloroplast. It is unfortunate that the AnfH enzyme was not detected, since it makes part of the iron-only cluster that appears to be an optimal option for expression in the organelle. It remains to be seen if combinations of other nitrogenase reductases (NifH, VnfH) could work in collaboration with the AnfHDKG reductase. Critically, this should be resolved utilizing *E. coli* as a platform. Nonetheless, the option of utilizing the non-optimized versions of genes is recommendable, as well as testing version with no tags on them. The homocitrate synthase (NifV) of *K. oxytoca* is expressed from the plas-

tome and was seen to have negligible effects on growth. However, it was not possible to detect its function in the crude extracts, and the levels of expression were not enough to retrieve sufficient protein after purification. Utilizing different promoters 5' UTR's could increase the expression of this protein, so purification could be carried out effectively. The remaining elements of the suggested cluster should be tested in a similar way to assess the possibility of their expression in the chloroplast. Once molecular biology tools have been developed to efficiently expressed various recombinant proteins in the organelle, the group of six genes has to be tested in tandem to determine if this combination is enough to drive N_2 reduction. As has been explained, the metabolic impact is expected to be important. Thus, mechanisms that allow the controlled expression of transgenes must be used. I have suggested a theoretical logic gate that could be used for such purposes. An AND logic gate was chosen. In this case, both inputs should be present to generate a predictive outcome. I propose low levels of oxygen and temperature as potential clues for activating the *nif* genes and creating a proper environment to sustain nitrogenase function. Although this idea is highly speculative, it is still critical to apply current approaches so we can gain a better appreciation of the principles involved.

6.3 Expression of a cyan fluorescent protein from the plastome of *C. reinhardtii*

6.3.1 Research

Expression of the mTurq2cp protein from the plastome of *C. reinhardtii* was shown to be successful. Fluorescent measurements demonstrated that the transformants do emit at the expected wavelength. I have shown the

design and assembly of a plasmid for inserting the *mturq2cp* gene into the algal plastome. Transformants recovered were homoplasmic and expressed successfully the protein at 27 kDa. Confocal microscopy was chosen because it was used for detecting fluorescence in *M. polymorpha*. Lamentably, this methodology did not permit us to observed major differences between the transgenic line and the control, which could be related to the low expression of the protein. Confocal microscopy has been observed not to be optimal when the signals are faint (Basiji et al., 2007). Despite the high levels of auto-fluorescence observed in the blue/cyan range, signals from cyan proteins have been detected in the cytoplasm. As was noted, the main source of fluorescence appears to be the chloroplast, which is known to enclose a variety of pigments. Thus, I assume that such molecules could either interfere or quench the signal emitted from the turquoise protein.

6.4 Concluding remarks

I have made use of synthetic biology principles to develop approaches that might enhance transplastomic engineering using this green alga. First, I determined that it is possible to use a single cloning device to deliver three transgenes in a single transformation event, and with no use of antibiotic resistance as a selectable marker. Additionally, I established another device that could be used to transform an already transplastomic strain. The next step is to develop a modular way of combining the expression cassettes, the selectable markers and the sequences for homologous recombination. It has been suggested that the modular cloning (MoClo) system could be used as a foundation. Secondly, I showed that, in principle, the chloroplast might represent a plausible niche for developing BNF in eukaryotic microalgae, and I demonstrated that a key element of the pathway (NifV) could be expressed in the organelle. Lastly, the stable expression of the cyan fluorescent protein was shown. Lamentably, no fluorescent emission could be detected from the mutant strains. Despite the technical limitations encountered, the results obtained reveal critical information that could eventually help devise schemes to metabolically engineer the chloroplast of this green alga.

References

Abbyad, P., Childs, W., Shi, X., and Boxer, S. G. (2007). Dynamic stokes shift in green fluorescent protein variants. *Proceedings of the National Academy of Sciences*, 104(51):20189–20194.

Adam, Z., Rudella, A., and van Wijk, K. J. (2006). Recent advances in the study of Clp, FtsH and other proteases located in chloroplasts. *Current opinion in plant biology*, 9(3):234–240.

Agapakis, C. M., Boyle, P. M., and Silver, P. A. (2012). Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nature chemical biology*, 8(6):527–535.

Allen, J. F. (2003). Cyclic, pseudocyclic and noncyclic photophosphorylation: new links in the chain. *Trends in plant science*, 8(1):15–19.

Almaraz-Delgado, A. L., Flores-Uribe, J., Pérez-España, V. H., Salgado-Manjarrez, E., and Badillo-Corona, J. A. (2014). Production of therapeutic proteins in the chloroplast of *Chlamydomonas reinhardtii*. *AMB express*, 4(1):57.

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

Apel, W. and Bock, R. (2009). Enhancement of carotenoid biosynthesis in transplastomic tomatoes by induced lycopene-to-provitamin a conversion. *Plant Physiology*, 151(1):59–66.

Avalos, J. L., Fink, G. R., and Stephanopoulos, G. (2013). Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nature biotechnology*, 31(4):335–341.

Barnes, D., Franklin, S., Schultz, J., Henry, R., Brown, E., Coragliotti, A., and Mayfield, S. P. (2005). Contribution of 5'-and 3'-untranslated regions of plastid mRNAs to the expression of *Chlamydomonas reinhardtii* chloroplast genes. *Molecular Genetics and Genomics*, 274(6):625–636.

Basiji, D. A., Ortyn, W. E., Liang, L., Venkatachalam, V., and Morrissey, P. (2007). Cellular image analysis and imaging by flow cytometry. *Clinics in laboratory medicine*, 27(3):653–670.

Bayle, V., Nussaume, L., and Bhat, R. A. (2008). Combination of novel green fluorescent protein mutant TSapphire and DsRed variant mOrange to set up a versatile in planta FRET-FLIM assay. *Plant physiology*, 148(1):51–60.

Berman, J., Zilberstein, A., Salomon, D., and Zamir, A. (1985). Expression of a nitrogen-fixation gene encoding a nitrogenase subunit in yeast. *Gene*, 35(1):1–9.

Berman-Frank, I., Quigg, A., Finkel, Z. V., Irwin, A. J., and Haramaty, L. (2007). Nitrogen-fixation strategies and fe requirements in cyanobacteria. *Limnology and Oceanography*, 52(5):2260–2269.

Bertalan, I., Munder, M. C., Weiß, C., Kopf, J., Fischer, D., and Johanningmeier, U. (2015). A rapid, modular and marker-free chloroplast expression system for the green alga *Chlamydomonas reinhardtii*. *Journal of biotechnology*, 195:60–66.

Bertani, G. (1951). Studies on lysogenesis i.: The mode of phage liberation by lysogenic *Escherichia coli*. *Journal of bacteriology*, 62(3):293.

Bieniossek, C., Imasaki, T., Takagi, Y., and Berger, I. (2012). Multibac: expanding the research toolbox for multiprotein complexes. *Trends in biochemical sciences*, 37(2):49–57.

Bienvenut, W. V., Espagne, C., Martinez, A., Majeran, W., Valot, B., Zivy, M., Vallon, O., Adam, Z., Meinnel, T., and Giglione, C. (2011). Dynamics of post-translational modifications and protein stability in the stroma of *Chlamydomonas reinhardtii* chloroplasts. *Proteomics*, 11(9):1734–1750.

Binder, A., Lambert, J., Morbitzer, R., Popp, C., Ott, T., Lahaye, T., and Parniske, M. (2014). A modular plasmid assembly kit for multigene expression, gene silencing and silencing rescue in plants. *PLoS One*, 9(2):e88218.

Blount, B. A., Weenink, T., Vasylechko, S., and Ellis, T. (2012). Rational diversification of a promoter providing fine-tuned expression and orthogonal regulation for synthetic biology. *PLoS One*, 7(3):e33279.

Bock, R. (2013). Strategies for metabolic pathway engineering with multiple transgenes. *Plant molecular biology*, 83(1-2):21–31.

Bock, R. (2014). Genetic engineering of the chloroplast: novel tools and new applications. *Current opinion in biotechnology*, 26:7–13.

Bock, R. (2015). Engineering plastid genomes: methods, tools, and applications in basic research and biotechnology. *Annual Review of Plant Biology*, 66:211–241.

Boehm, C. R., Ueda, M., Nishimura, Y., Shikanai, T., and Haseloff, J. (2015). A cyan fluorescent reporter expressed from the chloroplast genome of *Marchantia polymorpha*. *Plant and Cell Physiology*, 57(2):291–299.

Bonnefoy, N., Remacle, C., and Fox, T. D. (2007). Genetic transformation of *Saccharomyces cerevisiae* and *Chlamydomonas reinhardtii* mitochondria. *Methods in cell biology*, 80:525–548.

Boudreau, E., Nickelsen, J., Lemaire, S. D., Ossenbühl, F., and Rochaix, J.-D. (2000). The *nac2* gene of *Chlamydomonas* encodes a chloroplast TPR-like protein involved in *psbd* mRNA stability. *The EMBO Journal*, 19(13):3366–3376.

Bouwman, A., Beusen, A., and Billen, G. (2009). Human alteration of the global nitrogen and phosphorus soil balances for the period 1970–2050. *Global Biogeochemical Cycles*, 23(4).

Boyle, P. M. and Silver, P. A. (2012). Parts plus pipes: synthetic biology approaches to metabolic engineering. *Metabolic engineering*, 14(3):223–232.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry*, 72(1-2):248–254.

Brandner, K., Sambade, A., Boutant, E., Didier, P., Mély, Y., Ritzenthaler, C., and Heinlein, M. (2008). Tobacco mosaic virus movement protein interacts with green fluorescent protein-tagged microtubule end-binding protein 1. *Plant physiology*, 147(2):611–623.

Braun-Galleani, S., Baganz, F., and Purton, S. (2015). Improving recombinant protein production in the *Chlamydomonas reinhardtii* chloroplast using vivid verde fluorescent protein as a reporter. *Biotechnology journal*, 10(8):1289–1297.

Buhot, L., Horvàth, E., Medgyesy, P., and Lerbs-Mache, S. (2006). Hybrid transcription system for controlled plastid transgene expression. *The Plant Journal*, 46(4):700–707.

Cahoon, A. B. and Timko, M. P. (2000). Yellow-in-the-dark mutants of chlamydomonas lack the *chlL* subunit of light-independent protochlorophyllide reductase. *The Plant Cell*, 12(4):559–568.

Cakmak, T., Angun, P., Demiray, Y. E., Ozkan, A. D., Elibol, Z., and Tekinay, T. (2012). Differential effects of nitrogen and sulfur deprivation on growth and biodiesel feedstock production of *Chlamydomonas reinhardtii*. *Biotechnology and bioengineering*, 109(8):1947–1957.

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

Cavaiuolo, M., Kuras, R., Wollman, F.-A., Choquet, Y., and Vallon, O. (2017). Small RNA profiling in *Chlamydomonas*: insights into chloroplast RNA metabolism. *Nucleic Acids Research*, 45(18):10783–10799.

Chamizo-Ampudia, A., Galvan, A., Fernandez, E., and Llamas, A. (2011). The *Chlamydomonas reinhardtii* molybdenum cofactor enzyme CRARC has a zn-dependent activity and protein partners similar to those of its human homologue. *Eukaryotic cell*, 10(10):1270–1282.

Chen, H., Zou, Y., Shang, Y., Lin, H., Wang, Y., Cai, R., Tang, X., and Zhou, J.-M. (2008). Firefly luciferase complementation imaging assay for protein-protein interactions in plants. *Plant physiology*, 146(2):368–376.

Chen, X., Pham, E., and Truong, K. (2010). Tev protease-facilitated stoichiometric delivery of multiple genes using a single expression vector. *Protein Science*, 19(12):2379–2388.

Cheng, Q. (2008). Perspectives in biological nitrogen fixation research. *Journal of integrative plant biology*, 50(7):786–798.

Cheng, Q., Day, A., Dowson-Day, M., Shen, G.-F., and Dixon, R. (2005). The *Klebsiella pneumoniae* nitrogenase Fe protein gene (*nifh*) functionally substitutes for the *chlL* gene in *Chlamydomonas reinhardtii*. *Biochemical and biophysical research communications*, 329(3):966–975.

Clarens, A. F., Resurreccion, E. P., White, M. A., and Colosi, L. M. (2010). Environmental life cycle comparison of algae to other bioenergy feedstocks. *Environmental science & technology*, 44(5):1813–1819.

Collos, Y. and Harrison, P. J. (2014). Acclimation and toxicity of high ammonium concentrations to unicellular algae. *Marine pollution bulletin*, 80(1):8–23.

Cordero, B. F., Couso, I., León, R., Rodríguez, H., and Vargas, M. Á. (2011). Enhancement of carotenoids biosynthesis in *Chlamydomonas reinhardtii* by nuclear transformation using a phytoene synthase gene isolated from *Chlorella zofingiensis*. *Applied microbiology and biotechnology*, 91(2):341–351.

Croft, M. T., Moulin, M., Webb, M. E., and Smith, A. G. (2007). Thiamine biosynthesis in algae is regulated by riboswitches. *Proceedings of the National Academy of Sciences*, 104(52):20770–20775.

Curatti, L., Hernandez, J. A., Igarashi, R. Y., Soboh, B., Zhao, D., and Rubio, L. M. (2007). In vitro synthesis of the iron–molybdenum cofactor of nitrogenase from iron, sulfur, molybdenum, and homocitrate using purified proteins. *Proceedings of the National Academy of Sciences*, 104(45):17626–17631.

Curatti, L. and Rubio, L. M. (2014). Challenges to develop nitrogen-fixing cereals by direct *nif*-gene transfer. *Plant Science*, 225:130–137.

Day, R. N. and Davidson, M. W. (2009). The fluorescent protein palette: tools for cellular imaging. *Chemical Society Reviews*, 38(10):2887–2921.

de Montaigu, A., Sanz-Luque, E., Galván, A., and Fernández, E. (2010). A soluble guanylate cyclase mediates negative signaling by ammonium on expression of nitrate reductase in *Chlamydomonas*. *The Plant Cell*, 22(5):1532–1548.

Dejtisakdi, W. and Miller, S. M. (2016). Overexpression of calvin cycle enzyme fructose 1, 6-bisphosphatase in *Chlamydomonas reinhardtii* has a detrimental effect on growth. *Algal Research*, 14:116–126.

Demurtas, O. C., Massa, S., Ferrante, P., Venuti, A., Franconi, R., and Giuliano, G. (2013). A *Chlamydomonas*-derived human papillomavirus 16 e7 vaccine induces specific tumor protection. *PLoS One*, 8(4):e61473.

Dent, R. M., Haglund, C. M., Chin, B. L., Kobayashi, M. C., and Niyogi, K. K. (2005). Functional genomics of eukaryotic photosynthesis using insertional mutagenesis of *Chlamydomonas reinhardtii*. *Plant physiology*, 137(2):545–556.

Doehn, J. M., Fischer, K., Reppe, K., Gutbier, B., Tschernig, T., Hocke, A. C., Fischetti, V. A., Löffler, J., Suttorp, N., Hippenstiel, S., et al.

(2013). Delivery of the endolysin Cpl-1 by inhalation rescues mice with fatal pneumococcal pneumonia. *Journal of Antimicrobial Chemotherapy*, 68(9):2111–2117.

Dubeau, M.-P., Ghinet, M. G., Jacques, P.-É., Clermont, N., Beaulieu, C., and Brzezinski, R. (2009). Cytosine deaminase as a negative selection marker for gene disruption and replacement in the genus *Streptomyces* and other actinobacteria. *Applied and environmental microbiology*, 75(4):1211–1214.

Dubini, A. and Ghirardi, M. L. (2015). Engineering photosynthetic organisms for the production of biohydrogen. *Photosynthesis research*, 123(3):241–253.

Economou, C., Wannathong, T., Szaub, J., and Purton, S. (2014). A simple, low-cost method for chloroplast transformation of the green alga chlamydomonas reinhardtii. *Chloroplast Biotechnology: Methods and Protocols*, pages 401–411.

Eichler-Stahlberg, A., Weisheit, W., Ruecker, O., and Heitzer, M. (2009). Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*. *Planta*, 229(4):873–883.

Emami, S., Yee, M.-c., and Dinneny, J. R. (2013). A robust family of golden gate agrobacterium vectors for plant synthetic biology. *Frontiers in plant science*, 4.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PloS one*, 4(5):e5553.

Engler, C., Youles, M., Gruetzner, R., Ehnert, T.-M., Werner, S., Jones, J. D., Patron, N. J., and Marillonnet, S. (2014). A golden gate modular cloning toolbox for plants. *ACS synthetic biology*, 3(11):839–843.

Erisman, J., Bleeker, A., Galloway, J., and Sutton, M. (2007). Reduced nitrogen in ecology and the environment. *Environmental Pollution*, 150(1):140–149.

Esselborn, J., Muraki, N., Klein, K., Engelbrecht, V., Metzler-Nolte, N., Apfel, U.-P., Hofmann, E., Kurisu, G., and Happe, T. (2016). A structural view of synthetic cofactor integration into [FeFe]-hydrogenases. *Chemical Science*, 7(2):959–968.

Fan, J., Andre, C., and Xu, C. (2011). A chloroplast pathway for the de novo biosynthesis of triacylglycerol in *Chlamydomonas reinhardtii*. *FEBS letters*, 585(12):1985–1991.

Fernandez, E. and Galvan, A. (2007). Inorganic nitrogen assimilation in *Chlamydomonas*. *Journal of experimental botany*, 58(9):2279–2287.

Fouchard, S., Hemschemeier, A., Caruana, A., Pruvost, J., Legrand, J., Happe, T., Peltier, G., and Cournac, L. (2005). Autotrophic and mixotrophic hydrogen photoproduction in sulfur-deprived *Chlamydomonas* cells. *Applied and environmental microbiology*, 71(10):6199–6205.

Franklin, S., Ngo, B., Efuet, E., and Mayfield, S. P. (2002). Development of a gfp reporter gene for *Chlamydomonas reinhardtii* chloroplast. *The Plant Journal*, 30(6):733–744.

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

Gabriel, I. and Milewski, S. (2016). Characterization of recombinant homocitrate synthase from *Candida albicans*. *Protein expression and purification*, 125:7–18.

Gangl, D., Zedler, J. A., Włodarczyk, A., Jensen, P. E., Purton, S., and Robinson, C. (2015). Expression and membrane-targeting of an active plant cytochrome p450 in the chloroplast of the green alga *Chlamydomonas reinhardtii*. *Phytochemistry*, 110:22–28.

Georgianna, D. R. and Mayfield, S. P. (2012). Exploiting diversity and synthetic biology for the production of algal biofuels. *Nature*, 488(7411):329–335.

Giepmans, B. N., Adams, S. R., Ellisman, M. H., and Tsien, R. Y. (2006). The fluorescent toolbox for assessing protein location and function. *science*, 312(5771):217–224.

Giordano, M. and Raven, J. A. (2014). Nitrogen and sulfur assimilation in plants and algae. *Aquatic Botany*, 118:45–61.

Godman, J. and Balk, J. (2008). Genome analysis of chlamydomonas reinhardtii reveals the existence of multiple, compartmentalized iron–sulfur protein assembly machineries of different evolutionary origins. *Genetics*, 179(1):59–68.

Goedhart, J., Von Stetten, D., Noirclerc-Savoye, M., Lelimousin, M., Joosen, L., Hink, M. A., Van Weeren, L., Gadella Jr, T. W., and Royant, A. (2012). Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nature communications*, 3:751.

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

Goodenough, U., Lin, H., and Lee, J.-H. (2007). Sex determination in *Chlamydomonas*. In *Seminars in cell & developmental biology*, volume 18, pages 350–361. Elsevier.

Gorman, D. and Levine, R. (1965). TAP and Tris-minimal medium recipes. *Proceedings of the National Academy of Sciences USA*, 54:1665–1669.

Gregory, J. A., Topol, A. B., Doerner, D. Z., and Mayfield, S. (2013). Alga-produced cholera toxin-Pfs25 fusion proteins as oral vaccines. *Applied and environmental microbiology*, 79(13):3917–3925.

Griesbeck, C., Kobl, I., and Heitzer, M. (2006). *Chlamydomonas reinhardtii*. *Molecular biotechnology*, 34(2):213–223.

Grossman, A. R., Catalanotti, C., Yang, W., Dubini, A., Magneschi, L., Subramanian, V., Posewitz, M. C., and Seibert, M. (2011). Multiple facets of anoxic metabolism and hydrogen production in the unicellular green alga *Chlamydomonas reinhardtii*. *New Phytologist*, 190(2):279–288.

Grossman, A. R., Lohr, M., and Im, C. S. (2004). *Chlamydomonas reinhardtii* in the landscape of pigments. *Annu. Rev. Genet.*, 38:119–173.

Hai-Peng, C., Hong-Ning, W., Zhang, A.-Y., Meng-Die, D., Si-Tong, L., Cheng, H., Ying-Shun, Z., and Xin, L. (2012). Expression of avian infectious bronchitis virus multi-epitope based peptide epic in *Lactococcus lactis* for oral immunization of chickens. *Bioscience, biotechnology, and biochemistry*, 76(10):1871–1876.

Hallenbeck, P. C. and Benemann, J. R. (2002). Biological hydrogen production; fundamentals and limiting processes. *International Journal of Hydrogen Energy*, 27(11):1185–1193.

Hallmann, A. (2011). Evolution of reproductive development in the volvocine algae. *Sexual plant reproduction*, 24(2):97–112.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of molecular biology*, 166(4):557–580.

Harris, E., Stern, D., and Witman, G. (2009). The *Chlamydomonas* sourcebook, 2nd edn oxford.

Hasunuma, T., Miyazawa, S.-I., Yoshimura, S., Shinzaki, Y., Tomizawa, K.-I., Shindo, K., Choi, S.-K., Misawa, N., and Miyake, C. (2008). Biosynthesis of astaxanthin in tobacco leaves by transplastomic engineering. *The Plant Journal*, 55(5):857–868.

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

Higgs, D. (2009). The chloroplast genome. *The Chlamydomonas sourcebook*, 2:871–892.

HwangBo, K., Son, S. H., Lee, J. S., Min, S. R., Ko, S. M., Liu, J. R., Choi, D., and Jeong, W. J. (2010). Rapid and simple method for DNA extraction from plant and algal species suitable for PCR amplification using a chelating resin chelex 100. *Plant Biotechnology Reports*, 4(1):49–52.

Ibraheem, A. and Campbell, R. E. (2010). Designs and applications of fluorescent protein-based biosensors. *Current opinion in chemical biology*, 14(1):30–36.

Imlay, J. A. (2006). Iron-sulphur clusters and the problem with oxygen. *Molecular microbiology*, 59(4):1073–1082.

Jensen, M. K. and Keasling, J. D. (2015). Recent applications of synthetic biology tools for yeast metabolic engineering. *FEMS yeast research*, 15(1):1–10.

Jensen, P. E. and Leister, D. (2014). Chloroplast evolution, structure and functions. *F1000prime reports*, 6.

Johanningmeier, U. and Fischer, D. (2010). Perspective for the use of genetic transformants in order to enhance the synthesis of the desired metabolites: engineering chloroplasts of microalgae for the production of bioactive compounds. In *Bio-Farms for Nutraceuticals*, pages 144–151. Springer.

Jones, C. S., Luong, T., Hannon, M., Tran, M., Gregory, J. A., Shen, Z., Briggs, S. P., and Mayfield, S. P. (2013). Heterologous expression of the c-terminal antigenic domain of the malaria vaccine candidate Pfs48/45 in the green algae *Chlamydomonas reinhardtii*. *Applied microbiology and biotechnology*, 97(5):1987–1995.

Jouogo Noumsi, C., Pourhassan, N., Darnajoux, R., Deicke, M., Wichard, T., Burrus, V., and Bellenger, J.-P. (2016). Effect of organic matter on nitrogenase metal cofactors homeostasis in *Azotobacter vinelandii* under diazotrophic conditions. *Environmental microbiology reports*.

Kada, S., Koike, H., Satoh, K., Hase, T., and Fujita, Y. (2003). Arrest of chlorophyll synthesis and differential decrease of photosystems I and II in a cyanobacterial mutant lacking light-independent protochlorophyllide reductase. *Plant molecular biology*, 51(2):225–235.

Kato, K., Marui, T., Kasai, S., and Shinmyo, A. (2007). Artificial control of transgene expression in *Chlamydomonas reinhardtii* chloroplast using the Lac regulation system from *Escherichia coli*. *Journal of bioscience and bioengineering*, 104(3):207–213.

Keasling, J. D. (2012). Synthetic biology and the development of tools for metabolic engineering. *Metabolic engineering*, 14(3):189–195.

Kerppola, T. K. (2008). Bimolecular fluorescence complementation (bifc) analysis as a probe of protein interactions in living cells. *Annu. Rev. Biophys.*, 37:465–487.

Khalil, A. S. and Collins, J. J. (2010). Synthetic biology: applications come of age. *Nature Reviews Genetics*, 11(5):367–379.

Kim, J. H., Lee, S.-R., Li, L.-H., Park, H.-J., Park, J.-H., Lee, K. Y., Kim, M.-K., Shin, B. A., and Choi, S.-Y. (2011). High cleavage efficiency of a 2A peptide derived from porcine Teschovirus-1 in human cell lines, zebrafish and mice. *PloS one*, 6(4):e18556.

Kindle, K. L., Richards, K. L., and Stern, D. B. (1991). Engineering the chloroplast genome: techniques and capabilities for chloroplast transformation in *Chlamydomonas reinhardtii*. *Proceedings of the National Academy of Sciences*, 88(5):1721–1725.

Klein, U. (2009). Chloroplast transcription. *The Chlamydomonas sourcebook*, 2:893–914.

Klinkert, B., Schwarz, C., Pohlmann, S., Pierre, Y., Girard-Bascou, J., and Nickelsen, J. (2005). Relationship between mrna levels and protein accumulation in a chloroplast promoter-mutant of *Chlamydomonas reinhardtii*. *Molecular Genetics and Genomics*, 274(6):637–643.

Kneip, C., Voβ, C., Lockhart, P. J., and Maier, U. G. (2008). The cyanobacterial endosymbiont of the unicellular algae rhopalodia gibba shows reductive genome evolution. *BMC evolutionary biology*, 8(1):30.

Kosourov, S., Patrusheva, E., Ghirardi, M. L., Seibert, M., and Tsygankov, A. (2007). A comparison of hydrogen photoproduction by sulfur-deprived *Chlamydomonas reinhardtii* under different growth conditions. *Journal of Biotechnology*, 128(4):776–787.

Kumar, S., Hahn, F. M., Baidoo, E., Kahlon, T. S., Wood, D. F., McMahan, C. M., Cornish, K., Keasling, J. D., Daniell, H., and Whalen, M. C. (2012). Remodeling the isoprenoid pathway in tobacco by expressing the cytoplasmic mevalonate pathway in chloroplasts. *Metabolic engineering*, 14(1):19–28.

Laemmli, U. (1970). SDS-page Laemmli method. *Nature*, 227:680–5.

Lauersen, K. J., Kruse, O., and Mussgnug, J. H. (2015). Targeted expression of nuclear transgenes in *Chlamydomonas reinhardtii* with a versatile, modular vector toolkit. *Applied microbiology and biotechnology*, 99(8):3491–3503.

Lee, C. C., Hu, Y., and Ribbe, M. W. (2009). Unique features of the nitrogenase VFe protein from *Azotobacter vinelandii*. *Proceedings of the National Academy of Sciences*, 106(23):9209–9214.

Lee, C. C., Hu, Y., and Ribbe, M. W. (2010). Vanadium nitrogenase reduces co. *Science*, 329(5992):642–642.

Lee, L.-Y., Fang, M.-J., Kuang, L.-Y., and Gelvin, S. B. (2008). Vectors for multi-color bimolecular fluorescence complementation to investigate protein-protein interactions in living plant cells. *Plant Methods*, 4(1):24.

Lehtimäki, N., Koskela, M. M., and Mulo, P. (2015). Posttranslational modifications of chloroplast proteins: an emerging field. *Plant physiology*, 168(3):768–775.

Lill, R. (2009). Function and biogenesis of iron–sulphur proteins. *Nature*, 460(7257):831–838.

Lohr, M., Schwender, J., and Polle, J. E. (2012). Isoprenoid biosynthesis in eukaryotic phototrophs: a spotlight on algae. *Plant science*, 185:9–22.

Lössl, A., Bohmert, K., Harloff, H., Eibl, C., Mühlbauer, S., and Koop, H.-U. (2005). Inducible trans-activation of plastid transgenes: expression of the *R. eutropha phb* operon in transplastomic tobacco. *Plant and cell physiology*, 46(9):1462–1471.

Lü, J., Sheahan, C., and Fu, P. (2011). Metabolic engineering of algae for fourth generation biofuels production. *Energy & Environmental Science*, 4(7):2451–2466.

Lu, Y., Rijzaani, H., Karcher, D., Ruf, S., and Bock, R. (2013). Efficient metabolic pathway engineering in transgenic tobacco and tomato plastids with synthetic multigene operons. *Proceedings of the National Academy of Sciences*, 110(8):E623–E632.

Lynch, S. A. and Gill, R. T. (2012). Synthetic biology: New strategies for directing design. *Metabolic engineering*, 14(3):205–211.

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

Manuell, A., Beligni, M., Yamaguchi, K., and Mayfield, S. (2004). Regulation of chloroplast translation: interactions of RNA elements, RNA-binding proteins and the plastid ribosome.

Manuell, A. L., Beligni, M. V., Elder, J. H., Siefker, D. T., Tran, M., Weber, A., McDonald, T. L., and Mayfield, S. P. (2007a). Robust expression of a bioactive mammalian protein in *Chlamydomonas* chloroplast. *Plant biotechnology journal*, 5(3):402–412.

Manuell, A. L., Quispe, J., and Mayfield, S. P. (2007b). Structure of the chloroplast ribosome: novel domains for translation regulation. *PLoS biology*, 5(8):e209.

Marín-Navarro, J., Manuell, A. L., Wu, J., and Mayfield, S. P. (2007). Chloroplast translation regulation. *Photosynthesis Research*, 94(2-3):359–374.

Markou, G. and Nerantzis, E. (2013). Microalgae for high-value compounds and biofuels production: a review with focus on cultivation under stress conditions. *Biotechnology advances*, 31(8):1532–1542.

Mathews, J. and Wang, G. (2009). Metabolic pathway engineering for enhanced biohydrogen production. *International journal of hydrogen energy*, 34(17):7404–7416.

Matsuo, T. and Ishiura, M. (2011). *Chlamydomonas reinhardtii* as a new model system for studying the molecular basis of the circadian clock. *FEBS letters*, 585(10):1495–1502.

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

Maul, J. E., Lilly, J. W., Cui, L., Miller, W., Harris, E. H., Stern, D. B., et al. (2002). The *Chlamydomonas reinhardtii* plastid chromosome islands of genes in a sea of repeats. *The Plant Cell*, 14(11):2659–2679.

Mayfield, S. P., Franklin, S. E., and Lerner, R. A. (2003). Expression and assembly of a fully active antibody in algae. *Proceedings of the National Academy of Sciences*, 100(2):438–442.

Mayfield, S. P., Manuell, A. L., Chen, S., Wu, J., Tran, M., Siefker, D., Muto, M., and Marin-Navarro, J. (2007). *Chlamydomonas reinhardtii* chloroplasts as protein factories. *Current opinion in biotechnology*, 18(2):126–133.

Mayfield, S. P. and Schultz, J. (2004). Development of a luciferase reporter gene, *luxct*, for *Chlamydomonas reinhardtii* chloroplast. *The Plant Journal*, 37(3):449–458.

Meinecke, L., Alawady, A., Schroda, M., Willows, R., Kobayashi, M. C., Niyogi, K. K., Grimm, B., and Beck, C. F. (2010). Chlorophyll-deficient mutants of *Chlamydomonas reinhardtii* that accumulate magnesium protoporphyrin IX. *Plant molecular biology*, 72(6):643–658.

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., Maréchal-Drouard, L., et al. (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science*, 318(5848):245–250.

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

Minko, I., Holloway, S., Nikaido, S., Carter, M., Odom, O., Johnson, C., and Herrin, D. (1999). Renilla luciferase as a vital reporter for chloroplast gene expression in *Chlamydomonas*. *Molecular and General Genetics MGG*, 262(3):421–425.

Miyamoto, T., Razavi, S., DeRose, R., and Inoue, T. (2012). Synthesizing biomolecule-based boolean logic gates. *ACS synthetic biology*, 2(2):72–82.

Modak, J. M. (2011). Haber process for ammonia synthesis. *Resonance*, 16(12):1159–1167.

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

Moser, J., Lange, C., Krausze, J., Rebelein, J., Schubert, W.-D., Ribbe, M. W., Heinz, D. W., and Jahn, D. (2013). Structure of ADP-aluminium fluoride-stabilized protochlorophyllide oxidoreductase complex. *Proceedings of the National Academy of Sciences*, 110(6):2094–2098.

Mudd, E. A., Madesis, P., Avila, E. M., and Day, A. (2014). Excision of plastid marker genes using directly repeated DNA sequences. *Chloroplast Biotechnology: Methods and Protocols*, pages 107–123.

Mus, F., Dubini, A., Seibert, M., Posewitz, M. C., and Grossman, A. R. (2007). Anaerobic acclimation in *Chlamydomonas reinhardtii* anoxic gene expression, hydrogenase induction, and metabolic pathways. *Journal of Biological Chemistry*, 282(35):25475–25486.

Mussgnug, J. H. (2015). Genetic tools and techniques for *Chlamydomonas reinhardtii*. *Applied microbiology and biotechnology*, 99(13):5407–5418.

Nakahira, Y., Ogawa, A., Asano, H., Oyama, T., and Tozawa, Y. (2013). Theophylline-dependent riboswitch as a novel genetic tool for strict regulation of protein expression in cyanobacterium *Synechococcus elongatus* pcc 7942. *Plant and Cell Physiology*, 54(10):1724–1735.

Naqvi, S., Farre, G., Sanahuja, G., Capell, T., Zhu, C., and Christou, P. (2010). When more is better: multigene engineering in plants. *Trends in plant science*, 15(1):48–56.

Nedelcu, A. M. and Lee, R. W. (1998). Modes and tempos of mitochondrial and chloroplast genome evolution in *Chlamydomonas*: A comparative analysis. In *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas*, pages 63–91. Springer.

Nishimura, Y., Kikis, E. A., Zimmer, S. L., Komine, Y., and Stern, D. B. (2004). Antisense transcript and rna processing alterations suppress instability of polyadenylated mRNA in *Chlamydomonas* chloroplasts. *The Plant Cell*, 16(11):2849–2869.

Noor-Mohammadi, S., Pourmir, A., and Johannes, T. W. (2012). Method to assemble and integrate biochemical pathways into the chloroplast genome of *Chlamydomonas reinhardtii*. *Biotechnology and bioengineering*, 109(11):2896–2903.

Novoa, E. M. and de Pouplana, L. R. (2012). Speeding with control: codon usage, tRNAs, and ribosomes. *Trends in Genetics*, 28(11):574–581.

Ohyama, K., Takemura, M., Oda, K., Fukuzawa, H., Kohchi, T., Nakayama, S., Ishizaki, K., Fujisawa, M., and Yamato, K. (2009). Gene content, organization and molecular evolution of plant organellar genomes and sex chromosomes—insights from the case of the liverwort *Marchantia polymorpha*. *Proceedings of the Japan Academy, Series B*, 85(3):108–124.

Oldroyd, G. E. and Dixon, R. (2014). Biotechnological solutions to the nitrogen problem. *Current Opinion in Biotechnology*, 26:19–24.

Olinares, P. D. B., Kim, J., and van Wijk, K. J. (2011). The Clp protease system; a central component of the chloroplast protease network. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1807(8):999–1011.

Olson, E. J., Hartsough, L. A., Landry, B. P., Shroff, R., and Tabor, J. J. (2014). Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nature methods*, 11(4):449–455.

Ortiz-Marquez, J. C. F., Do Nascimento, M., and Curatti, L. (2014). Metabolic engineering of ammonium release for nitrogen-fixing multi-species microbial cell-factories. *Metabolic engineering*, 23:154–164.

Ortiz-Marquez, J. C. F., Do Nascimento, M., de los Angeles Dublan, M., and Curatti, L. (2012). Association with an ammonium-excreting bacterium allows diazotrophic culture of oil-rich eukaryotic microalgae. *Applied and environmental microbiology*, 78(7):2345–2352.

Paddon, C. J., Westfall, P., Pitera, D., Benjamin, K., Fisher, K., McPhee, D., Leavell, M., Tai, A., Main, A., Eng, D., et al. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 496(7446):528–532.

Patron, N. J., Orzaez, D., Marillonnet, S., Warzecha, H., Matthewman, C., Youles, M., Raitskin, O., Leveau, A., Farré, G., Rogers, C., et al. (2015). Standards for plant synthetic biology: a common syntax for exchange of dna parts. *New Phytologist*, 208(1):13–19.

Pauklin, S. and Vallier, L. (2013). The cell-cycle state of stem cells determines cell fate propensity. *Cell*, 155(1):135–147.

Peccia, J., Haznedaroglu, B., Gutierrez, J., and Zimmerman, J. B. (2013). Nitrogen supply is an important driver of sustainable microalgae biofuel production. *Trends in biotechnology*, 31(3):134–138.

Potvin, G. and Zhang, Z. (2010). Strategies for high-level recombinant protein expression in transgenic microalgae: a review. *Biotechnology Advances*, 28(6):910–918.

Purton, S., Szaub, J., Wannathong, T., Young, R., and Economou, C. (2013). Genetic engineering of algal chloroplasts: progress and prospects. *Russian journal of plant physiology*, 60(4):491–499.

Pustelny, K., Stach, N., Wladyka, B., Dubin, A., and Dubin, G. (2014). Evaluation of p1'substrate specificity of *Staphylococcal* SplB protease. *Acta Biochim Pol*, 61:149–152.

Rackham, O. and Chin, J. W. (2005). Cellular logic with orthogonal ribosomes. *Journal of the American Chemical Society*, 127(50):17584–17585.

Rajakumar, P. (2016). *The chloroplast of Chlamydomonas reinhardtii as a platform for recombinant vaccine production*. PhD thesis, UCL (University College London).

Ramundo, S. and Rochaix, J.-D. (2015). Chapter thirteen-controlling expression of genes in the unicellular alga *Chlamydomonas reinhardtii* with a vitamin-repressible riboswitch. *Methods in enzymology*, 550:267–281.

Rasala, B. A., Barrera, D. J., Ng, J., Plucinak, T. M., Rosenberg, J. N., Weeks, D. P., Oyler, G. A., Peterson, T. C., Haerizadeh, F., and Mayfield, S. P. (2013). Expanding the spectral palette of fluorescent proteins for the green microalga *Chlamydomonas reinhardtii*. *The Plant Journal*, 74(4):545–556.

Rasala, B. A., Chao, S.-S., Pier, M., Barrera, D. J., and Mayfield, S. P. (2014). Enhanced genetic tools for engineering multigene traits into green algae. *PloS one*, 9(4):e94028.

Rasala, B. A., Lee, P. A., Shen, Z., Briggs, S. P., Mendez, M., and Mayfield, S. P. (2012). Robust expression and secretion of Xylanase1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the Fmdv 2A peptide. *PloS one*, 7(8):e43349.

Rasala, B. A. and Mayfield, S. P. (2011). The microalga *Chlamydomonas reinhardtii* as a platform for the production of human protein therapeutics. *Bioengineered bugs*, 2(1):50–54.

Rasala, B. A., Muto, M., Lee, P. A., Jager, M., Cardoso, R. M., Behnke, C. A., Kirk, P., Hokanson, C. A., Crea, R., Mendez, M., et al. (2010). Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of *Chlamydomonas reinhardtii*. *Plant biotechnology journal*, 8(6):719–733.

Rasala, B. A., Muto, M., Sullivan, J., and Mayfield, S. P. (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.

Ravi, V., Khurana, J., Tyagi, A., and Khurana, P. (2008). An update on chloroplast genomes. *Plant Systematics and Evolution*, 271(1-2):101–122.

Reinbothe, C., El Bakkouri, M., Buhr, F., Muraki, N., Nomata, J., Kurisu, G., Fujita, Y., and Reinbothe, S. (2010). Chlorophyll biosynthesis: spotlight on protochlorophyllide reduction. *Trends in plant science*, 15(11):614–624.

Richard, C., Ouellet, H., and Guertin, M. (2000). Characterization of the Li818 polypeptide from the green unicellular alga *Chlamydomonas reinhardtii*. *Plant molecular biology*, 42(2):303–316.

Rizzo, M. A., Davidson, M. W., and Piston, D. W. (2009). Fluorescent protein tracking and detection: fluorescent protein structure and color variants. *Cold Spring Harbor Protocols*, 2009(12):pdb–top63.

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

Rogers, J. K., Guzman, C. D., Taylor, N. D., Raman, S., Anderson, K., and Church, G. M. (2015). Synthetic biosensors for precise gene control and real-time monitoring of metabolites. *Nucleic acids research*, doi.org/10.1093/nar/gkv616.

Rosales-Mendoza, S., Angulo, C., and Meza, B. (2016). Food-grade organisms as vaccine biofactories and oral delivery vehicles. *Trends in biotechnology*, 34(2):124–136.

Rosales-Mendoza, S., Paz-Maldonado, L. M. T., and Soria-Guerra, R. E. (2012). *Chlamydomonas reinhardtii* as a viable platform for the production of recombinant proteins: current status and perspectives. *Plant cell reports*, 31(3):479–494.

Rubio, L. M. and Ludden, P. W. (2008). Biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Annu. Rev. Microbiol.*, 62:93–111.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239(4839):487–491.

Saito, Y., Kishida, K., Takata, K., Takahashi, H., Shimada, T., Tanaka, K., Morita, S., Satoh, S., and Masumura, T. (2009). A green fluorescent protein fused to rice prolamin forms protein body-like structures in transgenic rice. *Journal of Experimental Botany*, 60(2):615–627.

Saxena, B., Subramaniyan, M., Malhotra, K., Bhavesh, N. S., Potlakayala, S. D., and Kumar, S. (2014). Metabolic engineering of chloroplasts for artemisinic acid biosynthesis and impact on plant growth. *Journal of biosciences*, 39(1):33–41.

Scaife, M. A., Nguyen, G. T., Rico, J., Lambert, D., Helliwell, K. E., and Smith, A. G. (2015). Establishing *Chlamydomonas reinhardtii* as an industrial biotechnology host. *The Plant Journal*, 82(3):532–546.

Scharff, L. B. and Bock, R. (2014). Synthetic biology in plastids. *The Plant Journal*, 78(5):783–798.

Schenk, P. M., Thomas-Hall, S. R., Stephens, E., Marx, U. C., Mussgnug, J. H., Posten, C., Kruse, O., and Hankamer, B. (2008). Second generation biofuels: high-efficiency microalgae for biodiesel production. *Bioenergy research*, 1(1):20–43.

Sears, B. B. (1998). Replication, recombination, and repair in the chloroplast genetic system of *Chlamydomonas*. In *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas*, pages 115–138. Springer.

Seibel, N. M., Eljouni, J., Nalaskowski, M. M., and Hampe, W. (2007). Nuclear localization of enhanced green fluorescent protein homomultimers. *Analytical biochemistry*, 368(1):95–99.

Seidel, T., Seefeldt, B., Sauer, M., and Dietz, K.-J. (2010). In vivo analysis of the 2-cys peroxiredoxin oligomeric state by two-step fret. *Journal of biotechnology*, 149(4):272–279.

Seo, S. W., Yang, J., Min, B. E., Jang, S., Lim, J. H., Lim, H. G., Kim, S. C., Kim, S. Y., Jeong, J. H., and Jung, G. Y. (2013). Synthetic biology: tools to design microbes for the production of chemicals and fuels. *Biotechnology advances*, 31(6):811–817.

Serganov, A. and Patel, D. J. (2012). Molecular recognition and function of riboswitches. *Current opinion in structural biology*, 22(3):279–286.

Shcherbo, D., Merzlyak, E. M., Chepurnykh, T. V., Fradkov, A. F., Ermakova, G. V., Solovieva, E. A., Lukyanov, K. A., Bogdanova, E. A., Zaraisky, A. G., Lukyanov, S., et al. (2007). Bright far-red fluorescent protein for whole-body imaging. *Nature methods*, 4(9):741.

Sialve, B., Bernet, N., and Bernard, O. (2009). Anaerobic digestion of microalgae as a necessary step to make microalgal biodiesel sustainable. *Biotechnology advances*, 27(4):409–416.

Smith, A. and Purton, S. (2002). The transcriptional apparatus of algal plastids. *European Journal of Phycology*, 37(3):301–311.

Specht, E., Miyake-Stoner, S., and Mayfield, S. (2010). Micro-algae come of age as a platform for recombinant protein production. *Biotechnology letters*, 32(10):1373–1383.

Spencer, P. S. and Barral, J. M. (2012). Genetic code redundancy and its influence on the encoded polypeptides. *Computational and structural biotechnology journal*, 1(1):1–8.

Stephenson, A. L., Kazamia, E., Dennis, J. S., Howe, C. J., Scott, S. A., and Smith, A. G. (2010). Life-cycle assessment of potential algal biodiesel production in the united kingdom: a comparison of raceways and air-lift tubular bioreactors. *Energy & Fuels*, 24(7):4062–4077.

Stern, D. B. and Drager, R. G. (1998). Chloroplast RNA synthesis and processing. In *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas*, pages 165–181. Springer.

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

Stoffels, L., Taunt, H. N., Charalambous, B., and Purton, S. (2017). Synthesis of bacteriophage lytic proteins against *Streptococcus pneumoniae* in the chloroplast of *Chlamydomonas reinhardtii*. *Plant Biotechnology Journal*.

Su, Z.-L., Qian K.-X., Tan, C.-P., Meng, C.-X., and Qin, S. (2005). Recombination and heterologous expression of allophycocyanin gene in the chloroplast of *Chlamydomonas reinhardtii*. *Acta biochimica et biophysica Sinica*, 37(10):709–712.

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

Summer, E. J., Schmid, V. H., Bruns, B. U., and Schmidt, G. W. (1997). Requirement for the h phosphoprotein in photosystem II of *Chlamydomonas reinhardtii*. *Plant physiology*, 113(4):1359–1368.

Sun, M., Qian, K., Su, N., Chang, H., Liu, J., and Shen, G. (2003). Foot-and-mouth disease virus vp1 protein fused with cholera toxin B subunit expressed in *Chlamydomonas reinhardtii* chloroplast. *Biotechnology letters*, 25(13):1087–1092.

Surzycki, R., Cournac, L., Peltier, G., and Rochaix, J.-D. (2007). Potential for hydrogen production with inducible chloroplast gene expression in *Chlamydomonas*. *Proceedings of the National Academy of Sciences*, 104(44):17548–17553.

Surzycki, R., Greenham, K., Kitayama, K., Dibal, F., Wagner, R., Rochaix, J.-D., Ajam, T., and Surzycki, S. (2009). Factors effecting expression of vaccines in microalgae. *Biologicals*, 37(3):133–138.

Swain, H. and Abhijita, S. (2013). Nitrogen fixation and its improvement through genetic engineering. *J Global Biosci*, 2:98–112.

Szydlowski, N., Ragel, P., Raynaud, S., Lucas, M. M., Roldán, I., Montero, M., Muñoz, F. J., Ovecka, M., Bahaji, A., Planchot, V., et al. (2009). Starch granule initiation in *Arabidopsis* requires the presence of either class IV or class III starch synthases. *The Plant Cell*, 21(8):2443–2457.

Tan, C.-P., Zhao, F.-Q., Su, Z.-L., Liang, C.-W., and Qin, S. (2007). Expression of β-carotene hydroxylase gene (*crtr-b*) from the green alga *Haematococcus pluvialis* in chloroplasts of *Chlamydomonas reinhardtii*. *Journal of Applied Phycology*, 19(4):347–355.

Tan, J., Kan, N., Wang, W., Ling, J., Qu, G., Jin, J., Shao, Y., Liu, G., and Chen, H. (2015). Construction of 2, 4, 6-trinitrotoluene biosensors with novel sensing elements from *Escherichia coli* k-12 mg1655. *Cell biochemistry and biophysics*, 72(2):417–428.

Taunt, H. N. (2014). *The synthesis of novel antibacterial proteins in the Chlamydomonas reinhardtii chloroplast*. PhD thesis, UCL (University College London).

Temme, K., Zhao, D., and Voigt, C. A. (2012). Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proceedings of the National Academy of Sciences*, 109(18):7085–7090.

Terashima, M., Specht, M., and Hippler, M. (2011). The chloroplast proteome: a survey from the *hlamydomonas reinhardtii* perspective with a focus on distinctive features. *Current genetics*, 57(3):151–168.

Tokumoto, U., Kitamura, S., Fukuyama, K., and Takahashi, Y. (2004). Interchangeability and distinct properties of bacterial Fe-S cluster assembly systems: functional replacement of the Isc and Suf operons in escherichia coli with the *nifsu*-like operon from *Helicobacter pylori*. *Journal of biochemistry*, 136(2):199–209.

Towbin, M., Staehelin, T., and Gordon, J. (1979). Western blotting: Electrophoretic transfer of proteins from sodium dodecyl sulphate–polyacrilamide gels to unmodified nitrocellulose and radiographic detection with antibodies and radio-iodinated protein. *Proceedings of the National Academy of Sciences of the United States of America*, 76:43–50.

Townsend, A. R., Howarth, R. W., Bazzaz, F. A., Booth, M. S., Cleveland, C. C., Collinge, S. K., Dobson, A. P., Epstein, P. R., Holland, E. A., Keeney, D. R., et al. (2003). Human health effects of a changing global nitrogen cycle. *Frontiers in Ecology and the Environment*, 1(5):240–246.

Tran, M., Zhou, B., Pettersson, P. L., Gonzalez, M. J., and Mayfield, S. P. (2009). Synthesis and assembly of a full-length human monoclonal antibody in algal chloroplasts. *Biotechnology and bioengineering*, 104(4):663–673.

Troxler, R. F., Zhang, F., Hu, J., and Bogorad, L. (1994). Evidence that [sigma] factors are components of chloroplast RNA polymerase. *Plant physiology*, 104(2):753–759.

Ueda, M., Tanaka, A., Sugimoto, K., Shikanai, T., and Nishimura, Y. (2014). chl b requirement for chlorophyll biosynthesis under short photoperiod in marchantia polymorpha l. *Genome biology and evolution*, 6(3):620–628.

Umen, J. G. and Goodenough, U. W. (2001). Chloroplast DNA methylation and inheritance in *Chlamydomonas*. *Genes & development*, 15(19):2585–2597.

Vafaee, Y., Staniek, A., Mancheno-Solano, M., and Warzecha, H. (2014). A modular cloning toolbox for the generation of chloroplast transformation vectors. *PloS one*, 9(10):e110222.

Vaistij, F. E., Boudreau, E., Lemaire, S. D., Goldschmidt-Clermont, M., and Rochaix, J.-D. (2000). Characterization of MBB1, a nucleus-encoded tetratricopeptide-like repeat protein required for expression of the chloroplast *psbb/psbt/psbh* gene cluster in *Chlamydomonas reinhardtii*. *Proceedings of the National Academy of Sciences*, 97(26):14813–14818.

Verhounig, A., Karcher, D., and Bock, R. (2010). Inducible gene expression from the plastid genome by a synthetic riboswitch. *Proceedings of the National Academy of Sciences*, 107(14):6204–6209.

Vermes, A., Guchelaar, H.-J., and Dankert, J. (2000). Flucytosine: a review of its pharmacology, clinical indications, pharmacokinetics, toxicity and drug interactions. *Journal of Antimicrobial Chemotherapy*, 46(2):171–179.

Vijayachandran, L. S., Viola, C., Garzoni, F., Trowitzsch, S., Bieniossek, C., Chaillet, M., Schaffitzel, C., Busso, D., Romier, C., Poterszman, A., et al. (2011). Robots, pipelines, polyproteins: enabling multiprotein expression in prokaryotic and eukaryotic cells. *Journal of structural biology*, 175(2):198–208.

Violot, S., Carpentier, P., Blanchoin, L., and Bourgeois, D. (2009). Reverse ph-dependence of chromophore protonation explains the large stokes shift of the red fluorescent protein mkeima. *Journal of the American Chemical Society*, 131(30):10356–10357.

Wakabayashi, K.-i., Misawa, Y., Mochiji, S., and Kamiya, R. (2011). Reduction-oxidation poise regulates the sign of phototaxis in *Chlamydomonas reinhardtii*. *Proceedings of the National Academy of Sciences*, 108(27):11280–11284.

Wang, B., Kitney, R. I., Joly, N., and Buck, M. (2011). Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature communications*, 2:508.

Wang, F., Johnson, X., Cavaiuolo, M., Bohne, A.-V., Nickelsen, J., and Vallon, O. (2015). Two chlamydomonas Opr proteins stabilize chloroplast mRNAs encoding small subunits of photosystem II and cytochrome b6f. *The Plant Journal*, 82(5):861–873.

Wang, J., Zhang, X., Chen, Y., Sommerfeld, M., and Hu, Q. (2008a). Toxicity assessment of manufactured nanomaterials using the unicellular green alga *Chlamydomonas reinhardtii*. *Chemosphere*, 73(7):1121–1128.

Wang, L., Zhang, L., Liu, Z., Zhao, D., Liu, X., Zhang, B., Xie, J., Hong, Y., Li, P., Chen, S., et al. (2013a). A minimal nitrogen fixation gene cluster from *Paenibacillus* sp. wly78 enables expression of active nitrogenase in *Escherichia coli*. *PLoS Genet*, 9(10):e1003865.

Wang, W.-Y. (1979). Photoconversion of photochlorophyllide in the y-1 mutant of *Chlamydomonas reinhardtii*. *Plant physiology*, 63(6):1102–1106.

Wang, X., Brandsma, M., Tremblay, R., Maxwell, D., Jevnikar, A. M., Huner, N., and Ma, S. (2008b). A novel expression platform for the production of diabetes-associated autoantigen human glutamic acid decarboxylase (hgad65). *Bmc Biotechnology*, 8(1):87.

Wang, X., Yang, J.-G., Chen, L., Wang, J.-L., Cheng, Q., Dixon, R., and Wang, Y.-P. (2013b). Using synthetic biology to distinguish and overcome regulatory and functional barriers related to nitrogen fixation. *PloS one*, 8(7):e68677.

Wannathong, T., Waterhouse, J. C., Young, R. E., Economou, C. K., and Purton, S. (2016). New tools for chloroplast genetic engineering allow the synthesis of human growth hormone in the green alga *Chlamydomonas reinhardtii*. *Applied microbiology and biotechnology*, 100(12):5467–5477.

Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PloS one*, 6(2):e16765.

Wei, Z., Liu, Y., Lin, C., Wang, Y., Cai, Q., Dong, Y., and Xing, S. (2011). Transformation of alfalfa chloroplasts and expression of green fluorescent protein in a forage crop. *Biotechnology letters*, 33(12):2487–2494.

Westfall, P. J., Pitera, D. J., Lenihan, J. R., Eng, D., Woolard, F. X., Regentin, R., Horning, T., Tsuruta, H., Melis, D. J., Owens, A., et al. (2012). Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proceedings of the National Academy of Sciences*, 109(3):E111–E118.

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S., et al. (2007). Database resources of the national center for biotechnology information. *Nucleic acids research*, 36(suppl_1):D13–D21.

Wienkoop, S., Weiß, J., May, P., Kempa, S., Irgang, S., Recuenco-Munoz, L., Pietzke, M., Schwemmer, T., Rupprecht, J., Egelhofer, V., et al. (2010). Targeted proteomics for *Chlamydomonas reinhardtii* combined with rapid subcellular protein fractionation, metabolomics and metabolic flux analyses. *Molecular BioSystems*, 6(6):1018–1031.

Wiig, J. A., Hu, Y., Lee, C. C., and Ribbe, M. W. (2012). Radical SAM-dependent carbon insertion into the nitrogenase M-cluster. *Science*, 337(6102):1672–1675.

Wijffels, R. H., Kruse, O., and Hellingwerf, K. J. (2013). Potential of industrial biotechnology with cyanobacteria and eukaryotic microalgae. *Current opinion in biotechnology*, 24(3):405–413.

Witzenrath, M., Schmeck, B., Doehn, J. M., Tschernig, T., Zahlten, J., Loeffler, J. M., Zemlin, M., Müller, H., Gutbier, B., Schütte, H., et al. (2009). Systemic use of the endolysin cpl-1 rescues mice with fatal *Pneumococcal pneumonia*. *Critical care medicine*, 37(2):642–649.

Wu, S., Xu, L., Huang, R., and Wang, Q. (2011). Improved biohydrogen production with an expression of codon-optimized *hemh* and *lba* genes in the chloroplast of *Chlamydomonas reinhardtii*. *Bioresource technology*, 102(3):2610–2616.

Wulandari, A. P., Miyazaki, J., Kobashi, N., Nishiyama, M., Hoshino, T., and Yamane, H. (2002). Characterization of bacterial homocitrate synthase involved in lysine biosynthesis. *FEBS letters*, 522(1-3):35–40.

Yang, J., Xie, X., Wang, X., Dixon, R., and Wang, Y.-P. (2014a). Reconstruction and minimal gene requirements for the alternative iron-only nitrogenase in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 111(35):E3718–E3725.

Yang, W., Catalanotti, C., Wittkopp, T. M., Posewitz, M. C., and Grossman, A. R. (2015). Algae after dark: mechanisms to cope with anoxic/hypoxic conditions. *The Plant Journal*, 82(3):481–503.

Yang, X., Chen, X., Bian, G., Tu, J., Xing, Y., Wang, Y., and Chen, Z. (2014b). Proteolytic processing, deubiquitinase and interferon antagonist activities of middle east respiratory syndrome coronavirus papain-like protease. *Journal of General Virology*, 95(3):614–626.

Yang, Z., Chen, F., Li, D., Zhang, Z., Liu, Y., Zheng, D., Wang, Y., Shen, G., et al. (2006). Expression of human soluble trail in *Chlamydomonas reinhardtii* chloroplast. *Chinese Science Bulletin*, 51(14):1703–1709.

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

Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G., and Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. *Molecular biology and evolution*, 21(5):809–818.

Young, R. E. and Purton, S. (2014). Cytosine deaminase as a negative selectable marker for the microalgal chloroplast: a strategy for the isolation of nuclear mutations that affect chloroplast gene expression. *The Plant Journal*, 80(5):915–925.

Young, R. E. and Purton, S. (2015). Codon reassignment to facilitate genetic engineering and biocontainment in the chloroplast of *Chlamydomonas reinhardtii*. *Plant biotechnology journal*.

Zadran, S., Standley, S., Wong, K., Otiniano, E., Amighi, A., and Baudry, M. (2012). Fluorescence resonance energy transfer (fret)-based biosensors: visualizing cellular dynamics and bioenergetics. *Applied microbiology and biotechnology*, 96(4):895–902.

Zamir, A., Maina, C. V., Fink, G. R., and Szalay, A. A. (1981). Stable chromosomal integration of the entire nitrogen fixation gene cluster from *Klebsiella pneumoniae* in yeast. *Proceedings of the National Academy of Sciences*, 78(6):3496–3500.

Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M. G., and Alon, U. (2006). A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nature methods*, 3(8):623.

Zedler, J. A., Gangl, D., Hamberger, B., Purton, S., and Robinson, C. (2015). Stable expression of a bifunctional diterpene synthase in the chloroplast of *Chlamydomonas reinhardtii*. *Journal of Applied Phycology*, 27(6):2271–2277.

Zehr, J. P. (2011). Nitrogen fixation by marine cyanobacteria. *Trends in microbiology*, 19(4):162–173.

Zelazny, E., Borst, J. W., Muylaert, M., Batoko, H., Hemminga, M. A., and Chaumont, F. (2007). FRET imaging in living maize cells reveals that plasma membrane aquaporins interact to regulate their subcellular localization. *Proceedings of the National Academy of Sciences*, 104(30):12359–12364.

Zerges, W. and Hauser, C. (2009). Protein synthesis in the chloroplast. *The Chlamydomonas sourcebook*, 2:967–1026.

Zheng, L., White, R. H., and Dean, D. R. (1997). Purification of the *Azotobacter vinelandii nifv*-encoded homocitrate synthase. *Journal of bacteriology*, 179(18):5963–5966.

Appendix A

A.1 Primers used to generate strains bearing the protease system

Table A.2: List of primers used to generate strains bearing the protease system

Primer	Purpose	Sequence (5' - 3')	Features
FcodAfusion	Amplifying the coding sequence of the cytosine deaminase and add the WELQP recognition sequence for the serine protase	ATTCGGC**TCATGA**C-TAACAACGCTTTA-CAAAC	Bold: *BspH*I restriction site
RcodAfusion- WELQP		AGGCAGC**TCATGA** ATTGTAATTCC CAAGCGTAATCTG-GAACATC	Bold: *BspH*I restriction site
			Underlined: WELQP recognition site
FSpIB	Amplifying serine protease expression cassette	AGGTA**ACGCGT**CT-CCAATATAGTAGAC	Bold: *Mlu*I restriction site
RSpIB		GCG**GCGCGC**ACGC-GCTACATCCGCTTT	Bold: *BssH*II restriction site
Flank1	Screening *C. reinhardtii* transformants (*psbH* reconstitution)	GTCATTGCGAAAATA-CTGGTGC	
rbcL.n		CGGATGTAACTCAATCGGTAG	
ATPA.R		ACGTCCACAG- GCGTG-GTAAGC	

A.2 Identification of correct plasmid assembly for testing the protease system

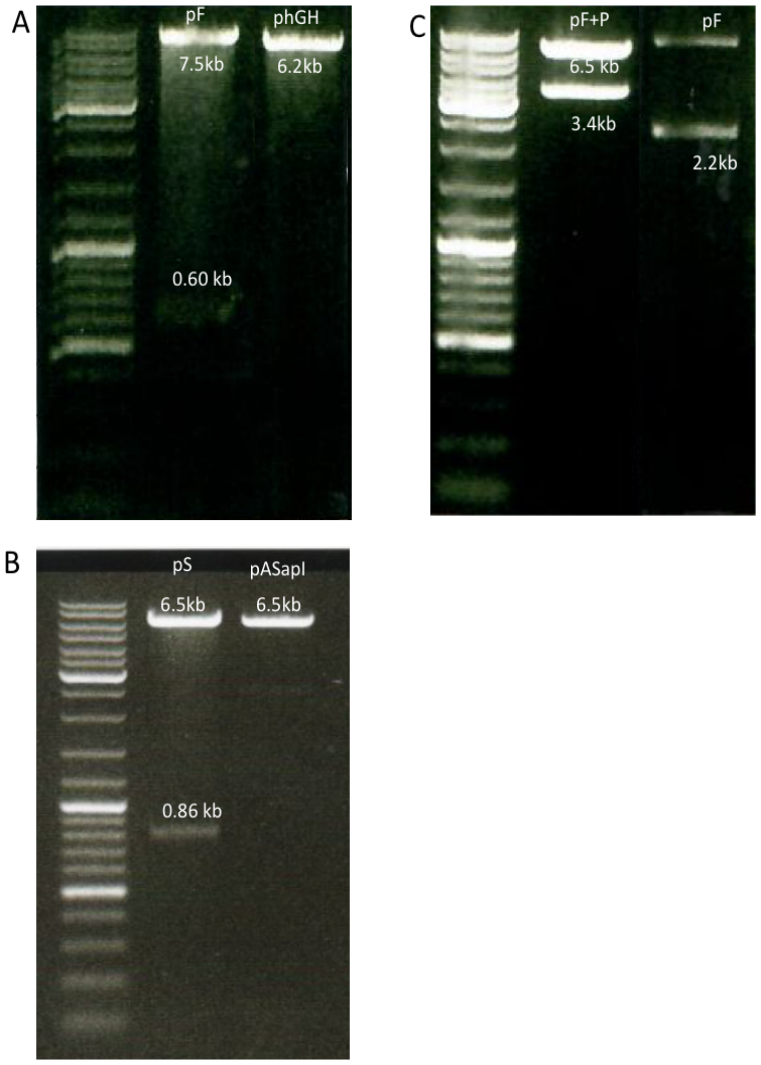


Figure A.4: Left-hand side: Agarose gels showing the digestion products from transformants. Right-hand side: Diagrams illustrating the genetic arrangement of the plasmids. A. Proper assembly of the pF plasmid yields a fragment of 7.5 kb and one of 0.60 kb after digestion with *Xcm*I and *Sph*I restriction enzymes. The empty vector (phGH) yields a fragment of 6.2 kb. B. Proper assembly of the pS plasmid yields a fragment of 6.5 kb and one of 0.86 kb after digestion with *Xho*I and and *Hind*III restriction enzymes. The pASapI plasmid yields a fragment of 6.5 kb. C. Proper assembly of the pF+P plasmid yields a fragment of 6.5 kb and one of 3.4 kb after digestion with *Mlu*I and *Xcm*I restriction enzymes. The pF plasmid yields a fragment of 6.5 kb and one of 2.2 kb.

A.3 List of primers used to generate strains expressing the combinatorial scheme

Table A.3: List of primers used to generate strains expressing the combinatorial scheme

Primer	Purpose	Sequence (5' - 3')	Features
FcodAMluI/Age	Amplifying CodA expression cassette	TTGGCT**ACGCG**TCCTTGACCGGTATTGTACCACAATAAATAAATTTGTC	Bold: *Mlu*I restriction site
RcodAMluI		TTGGCT**ACGCGT**TACATCCGCTTTAGTATGTTACTATTTC	Underlined: *Age*I restriction site
Fibv-ctbAgeI	Amplifying IBV-CTB expression cassette	CCTTG**ACCGGT**AAGCTTTCTTAATTCAACATTTT	Bold: *Age*I restriction site
Ribv-ctbAgeI		TTGGCTACCGGTTACATCCGCTTTAGTATGTTACTAT	
Flank1	Screening *C. reinhardtii* transformants (*psbH* reconstitution)	GTCATTGCGAAAATACTGGTGC	
rbcL.n		CGGATGTAACTCAATCGGTAG	
RY-psaR		ATAGGCTCTTCTCATGGATTTCTCCTTATAATAAC	
FcrI	Identifiy gene rearrangements in the crICS3 strain due to 5-FC treatment	CGGTGCTACATTTCAAGTAG	
RcrS		ATAAAGTCTACTATATTGGAGACGCG	
RpsbH		GCTTTAGAAGTTCCTGTTGC	

A.4 Identification of correct plasmid assembly for combinatorial transformation

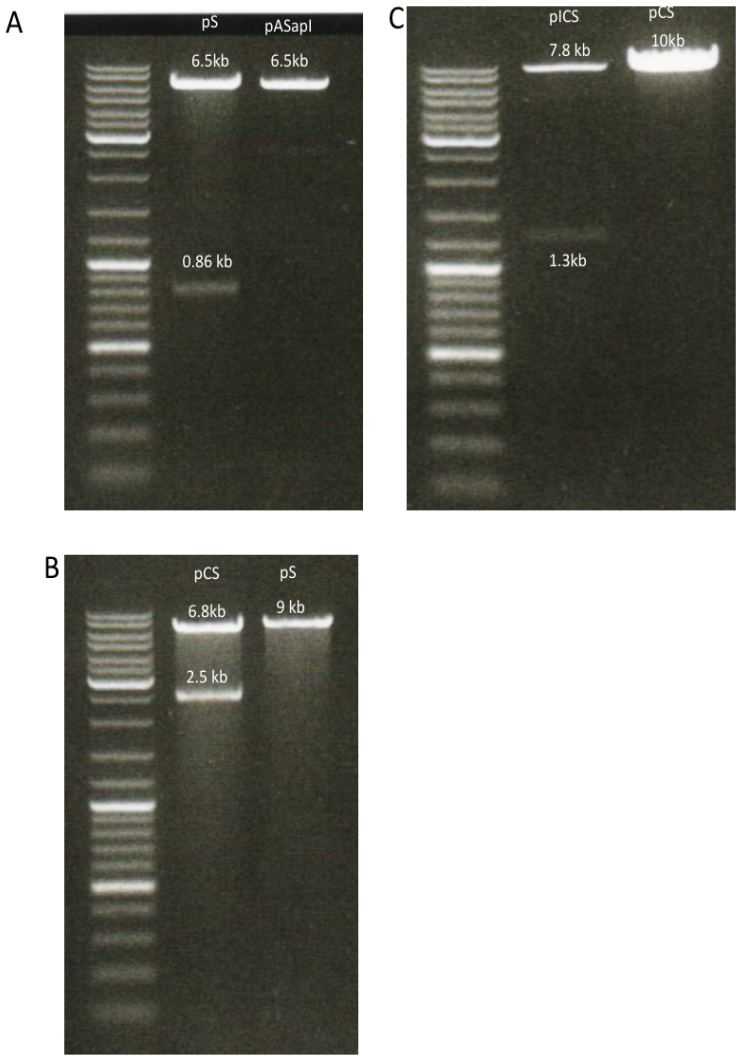


Figure A.5: Left-hand side: Agarose gels showing the digestion products from transformants. Right-hand side: Diagrams illustrating the genetic arrangement of the plasmids. A. Proper assembly of the pS plasmid yields a fragment of 6.5 kb and one of 0.86 kb after digestion with *Xho*I and *Hind*III restriction enzymes. The empty vector (pASapI) yields a fragment of 6.5 kb. B. Proper assembly of the pCS plasmid yields a fragment of 6.8 kb and one of 2.5 kb after digestion with *Age*I and *Xho*I restriction enzymes. The pS plasmid yields a fragment of 9.0 kb. C. Proper assembly of the pICS plasmid yields a fragment of 7.8 kb and one of 1.3 kb after digestion with *Sna*BI restriction enzyme. The pCS plasmid yields a fragment of 10.0 kb.

A.5 Western blot quantification of heterologous proteins in the crICS strains

Table A.4: Normalized values of the IR signal from the transformants analyzed

crICS1			**IR values**		
Protein	Sample1	Sample2	Sample3	Average IR	S.D.
CrC	423	395	513	443.67	61.65
CrI	9.51	13.00	11.50	11.34	1.75
CrS	112	149	168	143.00	28.48
crICS2					
CrC	382	445	446	424.33	36.67
CrI	11.00	11.60	8.72	10.44	1.52
CrS	144	174	180	166.00	19.29
crICS3					
CrC	555	599	618	590.67	32.32
CrI	11.27	6.93	15.42	11.21	4.25
CrS	173	208	184	188.33	17.90
crICS4					
CrC	333	420	425	392.67	51.73
CrI	12.41	10.16	11.72	11.43	1.15
CrS	174	160	213	182.33	27.47
crI					
CrI	3.99	8.67	5.87	6.18	2.36
crC					
CrC	619	646	677	647	29.02
crS					
CrS	417	452	524	464.33	54.56

A.6 List of primers used to generate strains expressing the serial scheme

Table A.5: List of primers used to generate strains expressing the serial scheme

Primer	Purpose	Sequence (5' - 3')	Features
FchlL	Amplifying *chlL* sequence	TGTCACCTCTTACC-ATATTTCTATACTCCAA	
RchlL		CCGATTATGTGCA-AGCTTAACAGCC	
FAc1F	Amplifying *aadA* expression cassette	CTTCCT**AACGTT**AAGC-TTATCGATGACTT-TATTAGAGGCAGTGTT	Bold: *Ac1F* restriction site
RNdeI		TTGGTG**CATATG**TTGC-ATGCCTGCAGGTC-GACTCTA	Bold: *Nde*I restriction site
FXmaI	Amplifying Cpl-1 expression cassette	CCTTG**CCCGGG**AA-GCTTTCTTAATTCAA-CATTTT	Bold: *Xma*I restriction site
RXmaI		TTGGCT**CCCGGG** TACA- TCCGCTTTAG-TATGTTACTAT	Bold: *Xma*I restriction site
F(chlLpetB)	Screening *C. reinhardtii* transformants (*chlL* disruption)	GGTGGAACATATTTA-CTTGTAATATC	
R(aadAcassette)		GATAGCATAAGCATTAAAATCAT	
R(ChlLuntransformed)		AATAACATCG-TATTCGAAAAAAGC	

A.7 Identification of correct plasmid assembly for serial transformation

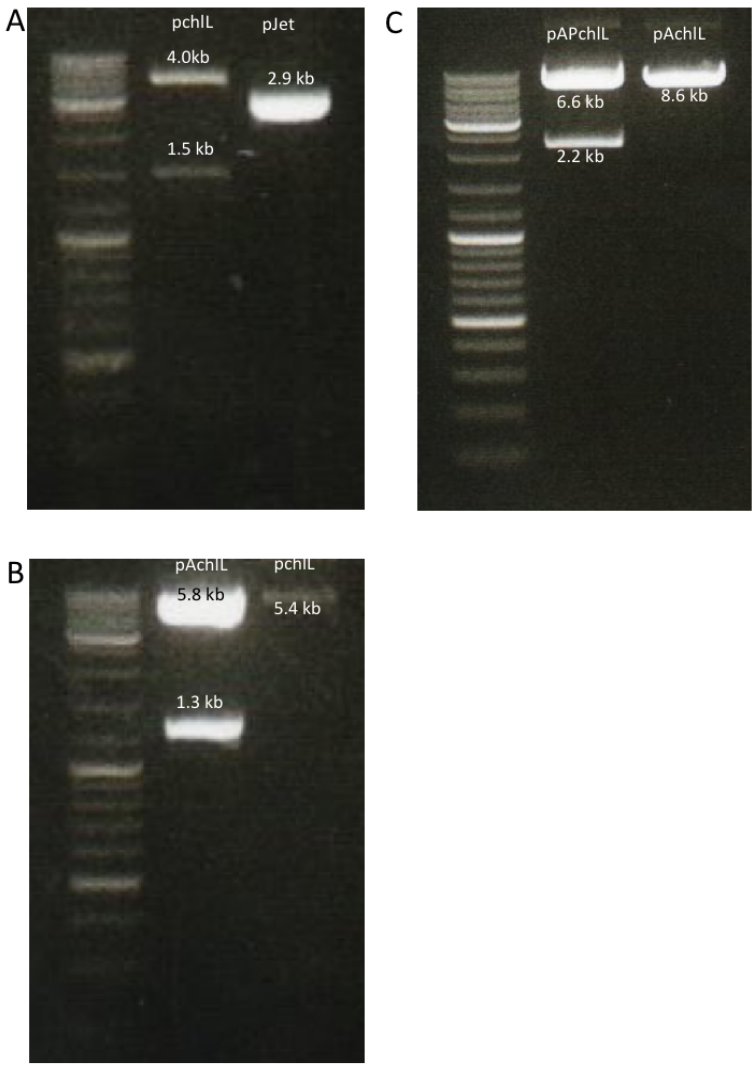


Figure A.6: Left-hand side: Agarose gels showing the digestion products from transformants. Right-hand side: Diagrams illustrating the genetic arrangement of the plasmids. A. Proper assembly of the pchlL plasmid yields a fragment of 4.0 kb and one of 1.50 kb after digestion with *Psi*I and *Xba*I restriction enzymes. The empty vector (pJet) yields a fragment of 2.9 kb. B. Proper assembly of the pA plasmid yields a fragment of 5.8 kb and one of 1.3 kb after digestion with *Nae*I and *Nsi*I restriction enzymes. The pchlL plasmid yields a fragment of 5.4 kb. C. Proper assembly of the pAP plasmid yields a fragment of 6.6 kb and one of 2.2 kb after digestion with *NsiI* restriction enzyme. The pA plasmid yields a fragment of 8.6 kb.

A.8 Phenotype of ΔChL strains

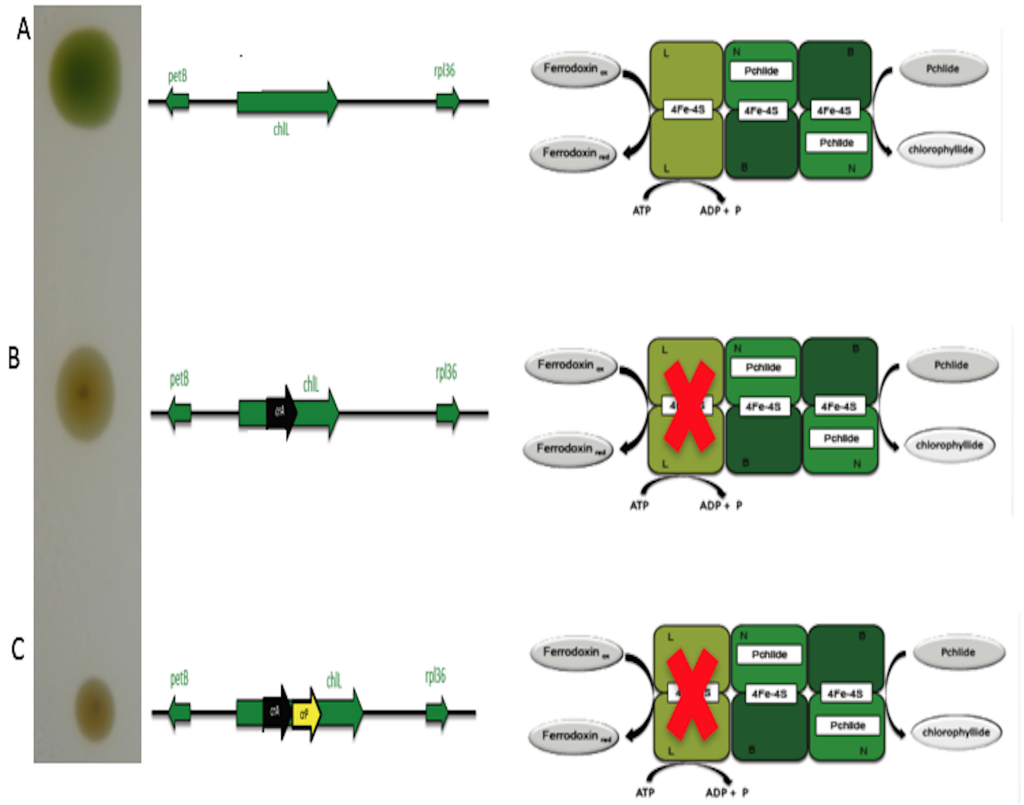


Figure A.7: The enzyme involved in the production of a precursor of chlorophyll, in the absence of light, is named DPOR. This complex is encoded by a triad of genes known as *chlL*, *chlB* and *chlN*. The ChlL protein constitutes a pivotal component for producing chlorophyllide, which will be metabolized into chlorophyll. A. Wild-type conditions. B. The crA transformant where the *chlL* gene is disrupted by the *aadA* cassette. C. The crAP transformant where the *chlL* gene is mutated by both the *aadA* cassette and the endolysing-containing cassette. The last two cell lines have lost the ChlL enzyme, therefore are unable to produce chlorophyll when light is missing.

Appendix B

B.1 Primers used for selecting algae transformants

Table B.1: Primers used for selecting algae transformants

Primer	Purpose	Sequence (5' - 3')
Flank1	Screening *C. reinhardtii* transformants (*psbH* reconstitution)	GTCATTGCGAAAAT-ACTGGTGC
rbcL.n		CGGATGTAACTCAA-TCGGTAG
RY-psaR		ATAGGCTCTTCTCAT-GGATTTCTCCT-TATAATAAC

B.2 Confirmation of insertion of *nif/anf* genes into the pSRSapI vector via test digestion

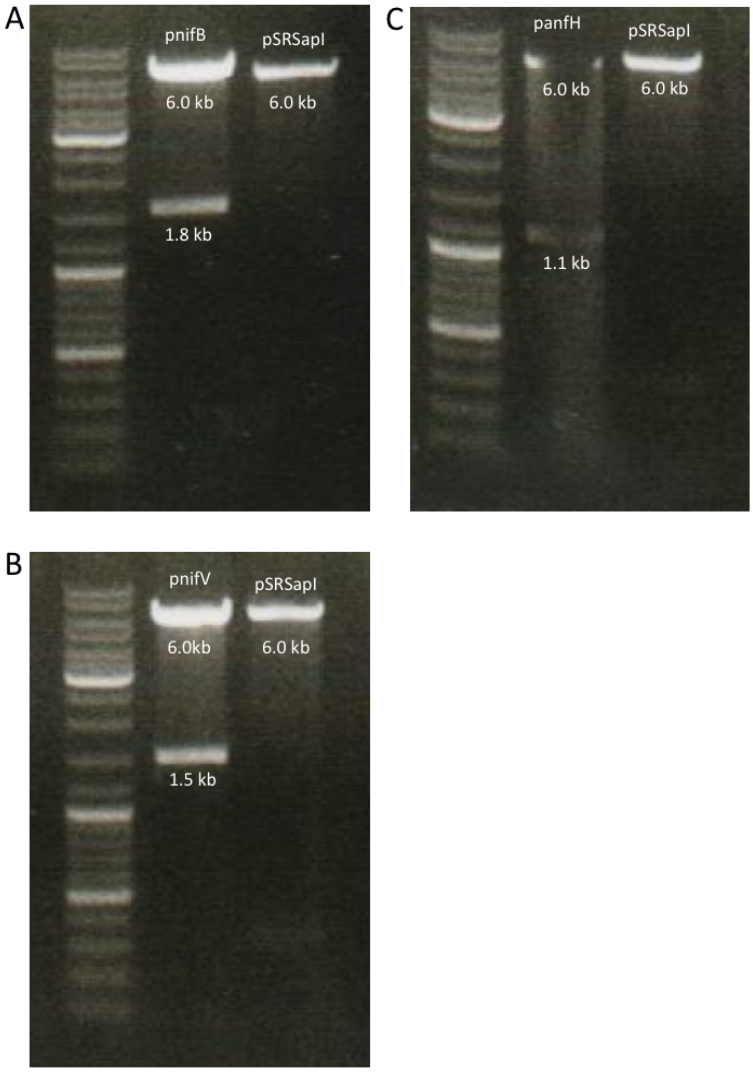


Figure B.1: A. Confirmation of correct insertion of crnifB into the plasmid via *Mlu*I - *Sph*I digestion, which generates products of 6.0 kb and 1.8 kb. B. Confirmation of correct insertion of crnifV into the plasmid via *Mlu*I - *Sph*I digestion, which generates products of 6.0 kb and 1.5 kb. C. Confirmation of correct insertion of cranfH into the plasmid via *Mlu*I - *Sph*I digestion, which generates a product of 6.0 kb and 1.1 kb. As a negative control, the empty pSRSapI vector (6.0 kb) was used.

B.3 Total Soluble protein determination in cell free extracts

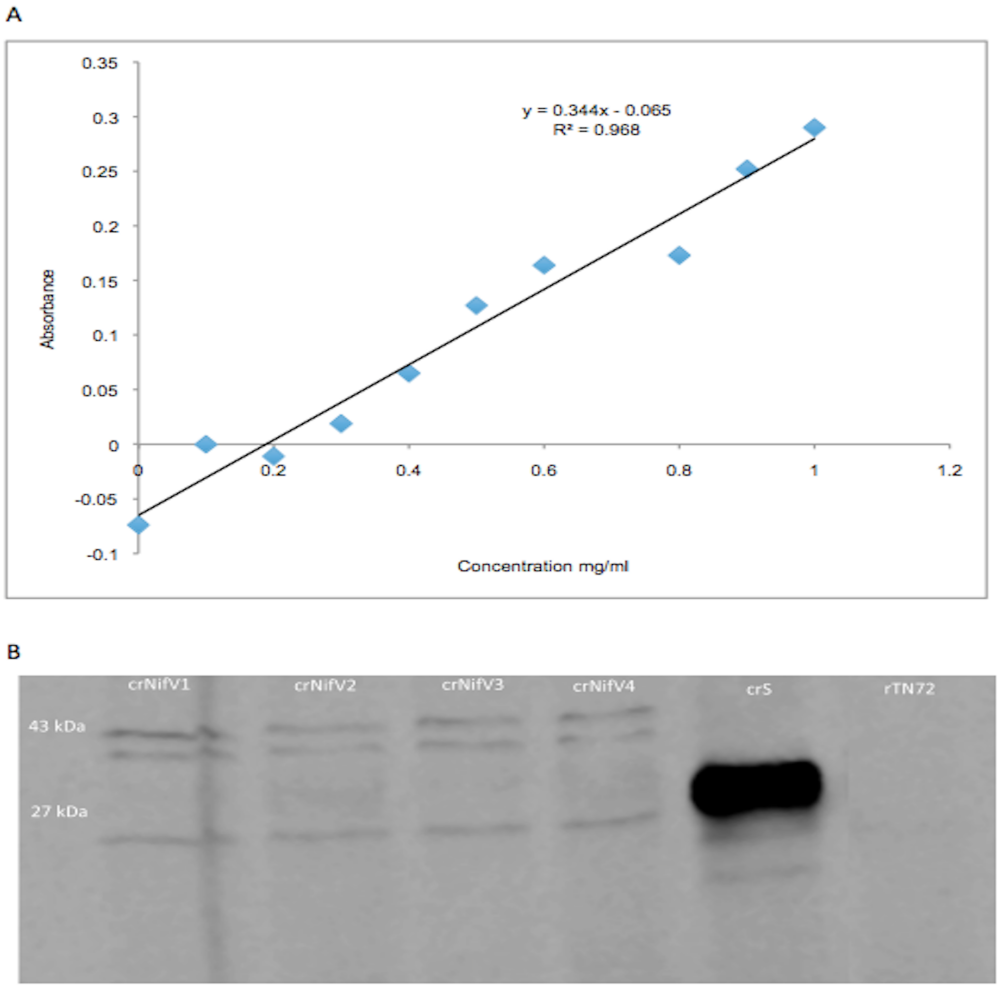


Figure B.2: A. Calibration curve assessed using Bovin serum albumin (BSA) (1 - 0 mg /ml) measured at 595 nm. B. Western blot analysis of the ultracentrifuged samples containing the crNifV enzyme (43 kDa), which were separated by SDS-PAGE gels, blotted and incubated with α-HA-Tag antibodies. The crS strain, expressing a serine protease (27 kDa) was used as positive control for α-HA blotting. The rTN72 lane corresponds to the negative control, this strain has been transformed with an empty expression cassette.

Table B.2: Values of total soluble protein. Values of three independent determinations were obtained from the four transformant lines. Mean values were used as reference for performing the DTNB test

	Concentration mg/ml				
Transformants	**1**	**2**	**3**	**Mean**	**S.D.**
crnifV1	2.41	2.52	2.01	2.31	0.22
crnifV2	2.36	2.21	2.39	2.32	0.08
crnifV3	2.51	2.22	2.03	2.25	0.20
crnifV4	1.75	2.57	2.38	2.23	0.43

B.4 Optical density based on the DTNB assay using crNifV-containing crude extract

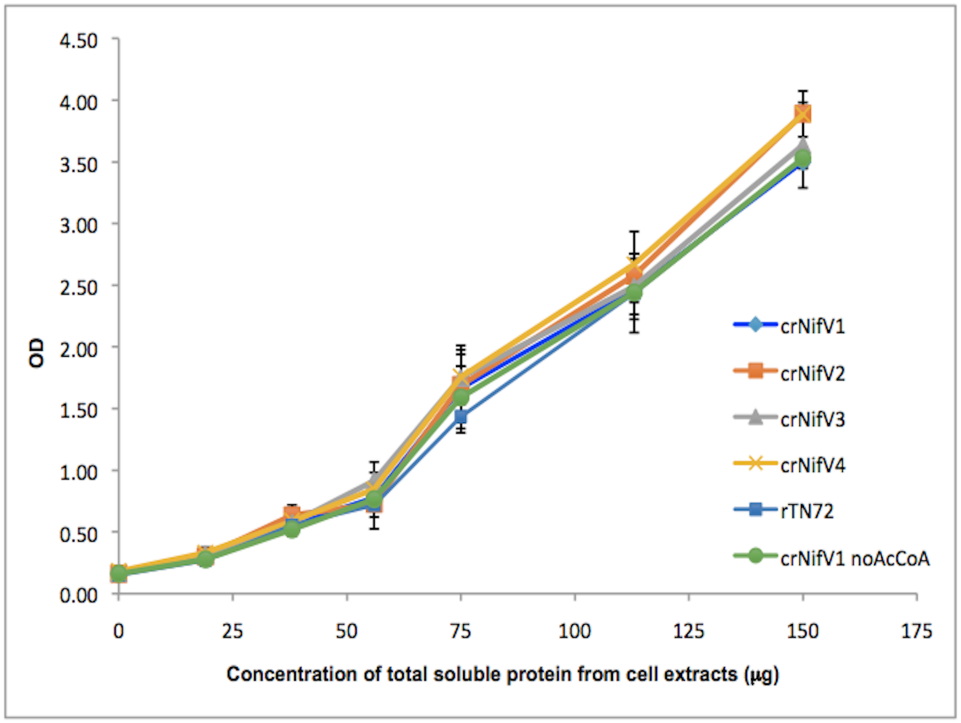


Figure B.3: OD values obtained when using different amounts of total protein from cell extracts.

Appendix C

C.1 Primers used to generate strains expressing the cyan fluorescent protein

Table C.1: List of primers used to generate strains expressing the cyan fluorescent protein

Primer	Purpose	Sequence (5' - 3')	Features
TurqSapIF	Amplifying the *mturq2cp* coding sequence	TGGTAT**GCTCTTCT**A-TGGTTTCTAAAGGA-GAAGAACTTTT	Bold: *Sap*I restriction site
TurqSphIR- HA-Tag		GTAT**GCATGC**-TTATTAAGCGTAATCTGGAACATCGTATGGGTA-TTTGTACAATTCATCCATTCC	Bold: *Sph*I restriction site Underlined: HA-Tag
FXmaI	Amplifying mTurq2cp expression cassette	CCTTG**CCCGGG**AA-GCTTTCTTAATTCAA-CATTTT	Bold: *Xma*I restriction site
RXmaI		TTGGCT**CCCGGG** TACA- TCCGCTTTAG-TATGTTACTAT	Bold: *Xma*I restriction site
Fmtur	Screening *C. reinhardtii* transformants (*chlL* disruption)	CACAATCAAAATTAT-CCAAAGATCCAAA	
Rrpl36		TGATTATTGTATTAT-TGTAACTGATAATG	
FchlL		TTTTTAACCGAAGCA - CGTACTTTCAT	

C.2 Identification of correct plasmid assembly for expression of the mTurq2cp protein

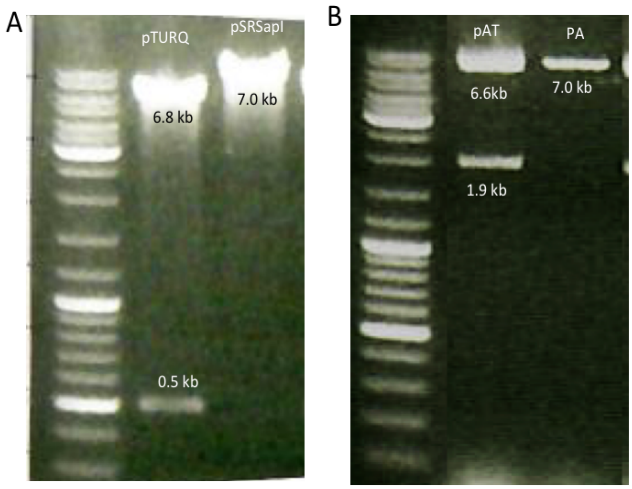


Figure C.1: Left-hand side: Agarose gels showing the digestion products from transformants. Right-hand side: Diagrams illustrating the genetic arrangement of the plasmids. A. Proper assembly of the pT plasmid yields a fragment of 6.4 kb and one of 0.5 kb after digestion with *Mlu*I restriction enzyme. The empty vector (pSRSapI) yields a fragment of 6.8 kb. B. Proper assembly of the pAT plasmid yields a fragment of 6.6 kb and one of 1.9 kb after digestion with *Nsi*I restriction enzyme. The empty vector (pA plasmid) yields a fragment of 6.8 kb.