NUCLEUS-ENCODED FACTORS INVOLVED IN 5' AND 3' PROCESSING OF CHLOROPLAST TRANSCRIPTS IN *CHLAMYDOMONAS REINHARDTII*

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Linda Ann Rymarquis

May 2006

NUCLEUS-ENCODED FACTORS INVOLVED IN 5' AND 3' PROCESSING OF CHLOROPLAST TRANSCRIPTS IN *CHLAMYDOMONAS REINHARDTII*

Linda Ann Rymarquis Ph. D.

Cornell University 2006

Chloroplast RNA maturation and degradation are regulated by nucleus-encoded factors that interact with sequences and structures within the RNA. Although several transcript-specific factors have been identified, those involved globally in RNA metabolism, apart from ribonucleases, have mostly remained elusive. Three pleiotropic nuclear mutations, *mcd3, mcd4,* and *mcd5*, appear to affect this global RNA metabolism, since they impact 5' end and 3' end maturation of two or more chloroplast transcripts in *Chlamydomonas reinhardtii*. These mutants were initially isolated as photoautotrophic suppressors of the 5' UTR mutations LS2 and LS6, which destabilized *petD* transcripts, but analysis of transcripts from 32 chloroplast genes showed that *mcd3* and *mcd4* displayed altered RNA transcript patterns for 17 genes, whereas three were altered in *mcd5.* The transcript patterns observed in *mcd3, mcd4, and mcd5* are consistent with defects in endonucleolytic cleavage. Since the role of endonucleolytic cleavage in *atpB* 3' maturation has been well characterized, a series of reporter strains containing the ectopically-expressed *atpB* processing determinant, which consists of an inverted repeat and endonuclease cleavage site (ECS), were used to further evaluate the role of *mcd4* in endonucleolytic cleavage as well as the sequences with which it might interact. These experiments suggested that *mcd4* suppresses endonucleolytic cleavage involved in *atpB* 3'

maturation, and that it is involved in endonucleolytic cleavages that initiate degradation. MCD4 likely interacts with the inverted repeat, the ECS and/or nucleotides 15-39 downstream of the ECS in facilitating this 3' maturation. To further characterize the role of *MCD3, MCD4,* and *MCD5*, a series of map-based cloning tools and methods were generated. Using these methods, *MCD4* has been isolated to a 1,300 kb region and candidate gene analysis is underway. Taken together, these studies suggest that MCD3, MCD4, and MCD5 may be components of multiprotein complexes responsible for RNA maturation and degradation in *Chlamydomonas* chloroplasts that are recruited by gene-specific proteins such as MCD1, or RNA sequences and structures such as those found for *atpB.*

BIOGRAPHICAL SKETCH

Linda Rymarquis was born in Edgewood, Kentucky to Barbara and Jerry Rymarquis. Her love of science blossomed throughout her youth, while her interest in molecular biology was honed by her Advanced Placement Biology class at Notre Dame Academy, where she performed her first DNA isolations and transformations. This led her to the agricultural biotechnology program at the University of Kentucky from which she graduated with a Bachelors of Science in 1999. Ithaca, New York then became her home as she embarked on a journey into the world of *Chlamydomonas* chloroplast RNA processing at Cornell University. This journey culminated in her receiving her Ph.D. in Plant Biology from Cornell University in 2006.

This thesis is dedicated to my husband and my parents who have always been supportive of my intellectual endeavors.

ACKNOWLEDGMENTS

I would like to thank David Stern for the careful nurturing of my scientific curiosity, sharpening my intellect, and his tutelage in developing my writing skills. Tom Bollenbach has been invaluable, participating in many hours of useful discussions and critically reading my manuscripts. Julien Fey and Amanda Hicks are also deserving of thanks for teaching me how to analyze RNA and cross *Chlamydomonas.* Carolyn Silflow, Laurens Mets and Susan Dutcher have graciously provided markers for mapping while Rachel Nguyen and Nancy Haas have provided BAC contig and marker information. Thanks also to the Stern lab for their help and support. This work was supported by NSF awards MCB–0235878 and MCB–0091020 to D.B.S., and an NSF Graduate Research Fellowship.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1.
INTRODUCTION

Photosynthesis is one of the driving forces of life on earth. At the center of photosynthesis is the chloroplast. It is widely accepted that the chloroplast is a descendant of a free living cyanobacterium that was engulfed by an early eukaryotic cell in an endosymbiotic event (Lopes-Juez and Pyke, 2005). Over time, the majority of the chloroplast genome was transferred to the nucleus, until only 120-200 kb encoding approximately 90-135 genes in the chloroplast remained (Simpson and Stern, 2002; Lopes-Juez and Pyke, 2005). As the vast majority of plastid-localized proteins are encoded by nuclear genes, there must be strict coordination of gene expression in both the nucleus and chloroplast in order to obtain stochiometric ratios of the photosynthetic proteins and those required for their synthesis and regulation. This coordination involves regulation at multiple steps from transcription through protein degradation.

Transcription

Chloroplast genes are transcribed by either the nucleus-encoded polymerase (NEP), or a plastid-encoded polymerase (PEP), resembling a phage-like polymerase and prokaryotic-like polymerase, respectively (Lopez-Juez and Pyke, 2005). Current models suggests that the plastid-targeted NEP transcribes the housekeeping genes, such as tRNAs, rRNAs and the PEP, whereas the PEP transcribes the photosynthetic genes (Cahoon and Stern, 2001). Plastid transcription can be regulated by light conditions, circadian rhythms, hormones such as auxins and gibberellins, and other environmental conditions (Aalen et al., 2001; Hwang et al., 1996; Mayfield et al., 1995) and at

least in part by the differential expression sigma factors that interact with the plastid-encoded polymerase. Such sigma factors have been found in maize (Beardslee *et al.*, 2002), *Chlamydomonas* (Carter *et al.*, 2004), rice (Kasai et al., 2004), and *Arabidopsis* (Kanamaru and Tanaka, 2004). Although chloroplast genes contain inverted repeats in the 3' UTR similar to the intrinsic terminators in bacteria (Henkin, 2000), transcription termination is inefficient leading to extended monocistronic pre-mRNAs and polycistronic transcripts (Stern and Gruissem, 1987; Stern and Kindle, 1993; Rott et al., 1996). These pre-mRNAs are then bound by gene-specific factors that regulate their stability (see below), splice and edit the transcript, and control 5' and 3' maturation, which is the focus of this work using *Chlamydomonas.*

Translation

Proper mRNA processing, the binding of gene-specific proteins, and the presence of the Shine-Dalgarno sequence have all been shown to stimulate translation. Although unprocessed or polycistronic RNAs can be translated (Barkan, 1988; Westhoff and Herrmann, 1988; Monde et al., 2000; Walter et al., 2002), they seem to be translated at a lower efficiency than processed RNA (Barkan et al., 1994). In *Chlamydomonas,* the mature *atpB* transcripts are preferentially associated with the polysomes (Rott et al., 1998). Taken together, RNA processing can affect the efficiency of translation, and therefore is a potential regulatory step in protein synthesis.

There is an ongoing debate over whether Shine-Dalgarno sequences, consisting of an AGGAGG sequence at -10 relative to the AUG, stimulates chloroplast translation. Some transcripts such as *psbA* require a Shine-Dalgarno sequence for translation initiation (Mayfield et al., 1994), while others

such as *petD, atpB, atpE, rps4* and *rps7* do not (Sakamoto et al., 1994; Fargo et al., 1998). Something that might influence the requirement for the Shine-Dalgarno is the binding of gene- or operon-specific proteins to 5' UTRs (reviewed in Zerges, 2000). Some examples of these are TAB2 and TBC2 which are *psaA* and *psbC* specific, respectively, in *Chlamydomonas* (Auchincloss et al., 2002; Dauvillee et al., 2003), and ATP1 and CRP1 which are *atpA* (ATP1) and *petA* and *petD* (CRP1) specific in maize (Barkan et al., 1994; McCormac and Barkan, 1999). The abundance of these proteins combined with the interaction of the base immediately preceding the initiation codon with the tRNAf(Met) anticodon at position 37 in *Chlamydomonas* may compensate for the lack of a Shine-Dalgarno (Esposito et al., 2001).

5' end maturation

Similar to translation, 5' end maturation involves the binding of gene-specific proteins. Such factors are thought to be recruited by inverted repeats (IRs), which are found in both the 5' and 3' untranslated regions (UTRs). IRs in the 5' UTR have been described for *Chlamydomonas atpB, rbcL, petD, rps7, psbB, psbC*, *psbD* and *psbA* transcripts. Deleting or mutating them destabilizes the RNA, leading to reduced transcript accumulation and translation, in some cases due to failure to interact with nucleus-encoded factors (Rochaix et al., 1989; Nickelsen et al., 1994; Fargo et al., 1999; Higgs et al., 1999; Vaistij et al., 2000; Anthonisen et al., 2001; Salvador et al., 2004; Suay et al., 2005). Many of the proteins that interact either directly or indirectly with the 5' UTR of transcripts have been initially identified by characterization of nonphotosynthetic or high chlorophyll fluorescence mutants. Factors that have been identified in this manner are MBB1, NAC2, MCD1 and MCA1 which

interact with *psbB*, *psbD, petD* and *petA*, respectively in *Chlamydomonas* (Kuchka et al., 1989; Drager et al., 1998; Vaistij et al., 2000; Lown et al., 2001), CRP1 which is involved in *petD* processing as well as translation (Barkan et al., 1994), HCF152 and HCF107 which interact with *psbB* and *psbH* (HCF107 is orthologous to MBB1), and PGR3 which interacts with *psbL* in *Arabidopsis* (Felder et al., 2001; Nakamura et al., 2003; Yamazaki et al., 2004). Mutants that lack these proteins do not accumulate their respective mature transcripts.

With the exception of MCD1, which encodes a protein with no known domains, the 5' end interacting proteins contain either pentatricopeptide (D'Andrea and Regan, 2003; Lurin *et al.*, 2004) or tetratricopeptide (D'Andrea and Regan, 2003) repeats, which are involved in RNA and protein binding (Fisk et al., 1999; Boudreau et al., 2000; Vaistij et al., 2000; Lown et al., 2001; Meierhoff et al., 2003; Sane et al., 2005). Consistent with this, MBB1, HCF107, NAC2 and CRP1 have been found in high molecular weight complexes of 300, 600-800, 500-600 and 326-362 kDa respectively (Fisk et al., 1999; Boudreau et al., 2000; Vaistij et al., 2000; Sane et al., 2005). Protein complexes involved in transcript stability, maturation and translation that bind to 5' UTRs have also been reported for *rps7, atpB, rbcL, rps12* and *psbA* (Hauser et al., 1996; Bruick and Mayfield, 1998; Fargo et al., 1999). It is tempting to speculate that such protein complexes are found at the 5' end of all chloroplast transcripts in order to confer stability and facilitate maturation and translation.

In other organisms, there are a number of enzymes that have been implicated in 5' end maturation mRNA and rRNA, such as endonucleases and 5' to 3' exonucleases, although only circumstantial evidence for these exists in

chloroplasts. Evidence for the involvement of an endonuclease in 5' maturation has been gathered from analysis of *petD* transcripts in *Chlamydomonas,* as well as the nuclear mutants *crp1* in maize and *crr2* in *Arabidopsis.* The *petD* gene in *Chlamydomonas* is transcribed from both the *petD* promoter and the upstream *petA* promoter (Sturm *et al.*, 1994). When the *petD* promoter is deleted, monocistronic *petD* transcripts still accumulate, having the same 5' end as those transcripts generated from the *petD* promoter. The most likely explanation is that an endonuclease cleaves the monocistronic *petD* and the *petA-petD* dicistronic transcripts to generate the mature 5' ends of the *petD* transcripts. In maize and *Arabidopsis*, there is also evidence that endonucleases are recruited to aid in 5' maturation. In the *crp1* mutant, no monocistronic *petD* accumulates due a lack of endonucleolytic cleavage that normally separates it from its *petB-petD* polycistronic precursor (Barkan *et al.*, 1994). Similarly, in the *Arabidopsis* mutant *crr2,* no monocistronic *ndhB* accumulates due a lack of endonuclease cleavage at the *ndhB* 5' end to separate it from its polycistronic precursor (Hashimoto et al., 2003). Taken together, these three examples suggest that an endonuclease cleavage is a common element of 5' maturation in chloroplasts.

Potential candidates for the endonucleases involved in chloroplast 5' maturation are homologues of RNase E and RNase G from *Escherichia coli,* and RNase J1 and RNase J2 from *Bacillus subtilus*. RNase E and RNase G share a N-terminal catalytic domain, but RNase E has an extra C-terminal extension for binding degradosome proteins, which will be discussed below (Wachi *et al.*, 1997). In *B. subtilis*, RNase J1 and J2 appear to have identical activities and specificities and although they share no sequence homology, are functional homologs of RNaseE/G (Even *et al.*, 2005). All four prefer 5'

monophosphorylated RNAs and cleave AU-rich single-stranded RNA (Jiang and Belasco, 2004; Even et al., 2005). Each has also been implicated in 5' maturation, since RNase E/G is required for the 5' maturation of 16S ribosomal RNA (Li *et al.*, 1999), while RNase J1/J2 cleavage of *thrS* stabilizes the RNA and allows ribosome binding (Even *et al.*, 2005). Although no RNase E/G homolog is known to be encoded in the *Chlamydomonas* genome, RNase E/G-like proteins are encoded in plant nuclear DNAs (Slomovic *et al.*, 2005). RNase J homologs, on the other hand, are encoded in *Chlamydomonas* and *Arabidopsis* and possess apparent chloroplast targeting signals (S. Zimmer, pers. comm.), making them excellent candidates for chloroplast RNA processing.

Another endonuclease that may function in the chloroplast is Rnt1p, the yeast homolog of RNase III. It specifically recognizes IRs closed by AGNN tetraloops and cleaves within the double-stranded region 13-16 nt distant from the tetraloop (Chanfreau *et al.*, 2000; Wu *et al.*, 2004). It is involved in the 5' maturation of C/D snoRNAs along with Xrn1p, a 5' to 3' exonuclease (Lee *et al.*, 2003). Although there is a homolog of Rnt1p encoded in *Arabidopsis*, none are found in the available nuclear genome of *Chlamydomonas.*

Another enzyme class potentially involved in chloroplast 5' end maturation is 5' to 3' exonucleases. Evidence of 5' to 3' exonuclease activity has been inferred from experiments where one or more polyguanosine (pG) tracts, which hinder exonuclease processivity, were inserted into the 5' UTRs of *petD, psbB* or *psbD* in *Chlamydomonas.* When these altered chloroplast genes were expressed in the *mcd1*, *mbb1* or *nac2* mutant backgrounds, respectively, which lack their respective 5' stabilizing proteins, only transcripts beginning with the pG accumulated (Drager et al., 1999; Nickelsen et al.,

1999; Vaistij et al., 2000). This suggests that a 5’ to 3’ exonuclease might be involved in 5’ processing of transcripts protected by their stabilizing proteins, while degrading unprotected portions of pre-mRNAs.

Exonucleases with 5’ to 3’ activity have been identified in both yeast and *Arabidopsis.* To date, no 5’ to 3’ exonuclease activity has been found in bacteria, although RNase E has been reported to cleave in a net 5’ to 3’ direction (Cannistraro and Kennell, 1985). The two yeast 5’ to 3’ exonucleases are Xrn1p and Xrn2p/Rat1p. Xrn1p is cytoplasmic and involved in RNA turnover, trimming the 5’ ends of rRNA, and degrading rRNA intermediates. Xrn2p/Rat1p is localized to the nucleus, processes the 5’ end rRNA and snoRNA, and its activity is enhanced by the binding of Rai1p (Xue et al., 2000; Kastenmayer et al., 2001). In *Arabidopsis*, there are three proteins homologous to Xrn2p, but none to Xrn1p (Kastenmayer and Green, 2000). AtXRN2 and AtXRN3 are nuclear and are thought to process rRNA and snoRNA while the third member, AtXRN4, localizes to the cytoplasm and is degrades the 3' products of miRNA-mediated cleavage (Kastenmayer and Green, 2000; Souret et al., 2004). There are three XRN family genes predicted in the *Chlamydomonas* genome and for one, there is EST evidence that it undergoes alternative splicing to generate a chloroplast targeted transcript (Murakami and Higgs, unpublished data). This suggests that exonucleases of the XRN family, along with endonucleases and stabilization factors, could be involved in the 5’ end maturation pathway.

Besides processing and degradation attributed to XRN proteins in yeast and *Arabidopsis*, a third function of this potential chloroplast XRN could be transcription termination. In yeast, transcription termination is paired with 3’ end maturation. After endonuclease cleavage at the 3’ end by CPF complex,

Rat1p binds to the newly-formed 5’ end and degrades the transcript until it reaches the polymerase, causing it to abort transcription (Kim *et al.*, 2004). A similar mechanism may be utilized in chloroplasts, but at the moment this is only speculation.

3’ end maturation

Although transcription termination signals are inefficient, a robust 3’ maturation pathway is functional in *Chlamydomonas* plastids. The maturation of *atpB* in *Chlamydomonas* has been particularly well-characterized. Its processing determinant resides in a 300 nt segment of the 3’ UTR containing an IR and endonuclease cleavage site (Hicks *et al.*, 2002). This maturation involves a precise endonuclease cleavage 10 nt downstream of the IR, followed by 3’ to 5’ exonucleolytic digestion to the base of the IR (Stern et al., 1989; Rott et al., 1998). Similar maturation pathways have been inferred for *rbcL* and *psaA* in *Chlamydomonas* (Blowers *et al.*, 1993), and *petD* in spinach (Hayes et al., 1996). When the endonuclease step of 3’ maturation is blocked, 3’ to 5’ exonucleolytic degradation alone can generate mature transcripts. This is seen in delta26pG which accumulates mature *atpB* transcripts even though the native IR and endonuclease cleavage site are replaced with a polyG sequence (Drager *et al.*, 1996). This indicates that there is some redundancy in 3’ maturation activities.

3’ end inverted repeats are common in cpRNAs and play the same dual roles in RNA stabilization and RNA processing as discussed above for *atpB*. Some examples of 3’ UTR IR-containing transcripts are *rbcL, petA, petD,* and *psaB* in *Chlamydomonas* (Blowers et al., 1993; Rott et al., 1998; Jiao et al., 2004), *rpoA, psbA,* and *petD* in spinach (Stern and Gruissem, 1987) and *petD*

in tobacco (Monde et al., 2000). These IRs protect their transcripts from degradation by 3' to 5' exonucleases (Stern and Gruissem, 1987; Stern et al., 1989; Drager et al., 1996; Lee et al., 1996). Unlike in mitochondria and eukaryotes where a ubiquitous AAUAAPyAUUCUU or AAUAAA and UG elements, respectively, signal endonuclease cleavage (Hofmann et al., 1993; Wahle and Ruegsegger, 1999), the IRs and surrounding regions share no obvious sequence similarity. Several studies have shown that the presence of secondary structure alone does not signal maturation since the IRs in their reverse orientation do not direct processing as efficiently or at all when compared to normally-oriented IRs (Blowers *et al.*, 1993; Drager *et al.*, 1996; Hicks *et al.*, 2002). This suggests that the IRs are orientation-specific, perhaps due to protein binding to specific regions.

In fact, IRs and their surrounding sequences have been shown to be binding sites of both general and gene-specific factors (Stern et al., 1989; Schuster and Gruissem, 1991). Some of the general factors are the 28-, 24- and 40 kDa proteins that bind to the *rbcL, petD,* and *psbA* IR-RNAs in spinach (Stern *et al.*, 1989). Genes encoding the 24 and 28 kDa proteins have been cloned, and reveal that they belong to a family of RNA-binding proteins with two binding domains. In the case of the 28 kDa protein, its binding affinity is regulated by its phosphorylation state and its abundance is developmentally regulated and varies by plant tissue (Schuster and Gruissem, 1991; Lisitsky et al., 1995; Lisitsky and Schuster, 1995). The 28 kDa protein plays a role in 3' end maturation, because chloroplast extracts depleted of it no longer process 3' ends. Surprisingly, such depleted extracts also cannot process the *rbcL* 5' end, thus implicating it in 5' maturation as well. The 40 kDa protein, later named CSP41, binds specifically to the IRs and has been implicated in mRNA

stability when in a complex with CSP29 and CSP55 (Yang and Stern, 1997). CSP41 shares homology to nucleotide-sugar epimerases and hydroxysteroid dehydrogenases and homologs have been found *Chlamydomonas, Arabidopsis,* and *Synechocystis* (Baker et al., 1998).

There are also a number of gene-specific proteins such as CSP55 and CSP29 in spinach (Stern *et al.*, 1989), 32 and 33 kDa species in barley (Memon *et al.*, 1996), and RAT1 and RAT2 in *Chlamydomonas,* that interact with specific IR sequences. CSP55 binds an AUUCAAUU sequence immediately downstream of the *petD* IR in the pre-mRNA, and upstream of the IR at a similar AUUUAAUU site in the mature transcript (Chen *et al.*, 1995). CSP29 binds to the base of the IR at a GU-rich sequence, but can only do so in the presence of CSP55 (Chen and Stern, 1991). The binding of CSP29 to the IR impedes 3' maturation (Chen and Stern, 1991), thus being a negative regulator of maturation. Conversely, RAT1 and RAT2 are specifically required for 3' end maturation of *tscA* transcripts (Balczun et al., 2005). A portion of RAT1 is homologous to the conserved NAD+-binding domain of poly(ADP-ribose) polymerases from eukaryotes, while RAT2 does not share homology to any known protein (Balczun et al., 2005). Taken together, the gene-specific proteins which interact with 3' ends modulate maturation through both positive and negative mechanisms

Potential candidates for the endonuclease involved in 3' maturation are RNase E/G, RNase J, and Rnt1p mentioned above, as well as p54 from mustard. Along with 5' maturation activity, RNase E/G have been shown to be involved in the 3' maturation and polycistronic maturation of tRNA transcripts (Li and Deutscher, 2002), and Rnt1p mutants accumulate *ADI1* transcripts and iron homeostasis mRNAs with extended 3' ends (Lee et al., 2005; Zer and

Chanfreau, 2005). The multiple roles these major enzymes play in RNA metabolism and the organisms in which they are found are summarized in Table 1.1. In mustard chloroplasts, p54 is involved in 3' end maturation of *trnK, rps16,* and *trnH.* It binds to a conserved UUUAUCU element and is positively regulated by phosphorylation and oxidation by glutathione disulfide (Liere et al., 2001). Since the gene encoding p54 has not been cloned, it is unclear whether is represents one of the endonucleases mentioned above or represents a new class.

Table 1.1. Enzymes involved in RNA maturation and degradation. The X indicates that the enzyme or activity has been discovered in that organism or has been associated with a particular pathway.

	E. coli	*B. subtilis*	yeast	*Synechocystis*	*A. thaliana* chloroplasts	Spinach chloroplasts	*C. reinhardtii*	5' maturation	3' maturation	Degradation
RNase E/G	X			X	X			X	X	X
RNase J		X			X		X	X		X
RNase III	X	X	X					X	X	X
XRN			X				X	X		X
RNase II/R	X	X	X	X	X	X	X		X	X
PNPase	X	X	X	X	X	X	X		X	X
Exosome			X						X	X
CSP41				X	X	X	X		X	X
PAP	X		X				X		X	X

In 3' maturation, trimming by a 3' to 5' exonuclease follows endonuclease cleavage. There are two candidates for the 3' to 5' exonuclease, RNase II and PNPase. RNase II is a member of the RNR superfamily of hydrolytic enzymes of which RNase II from *E. coli* is a member. *Arabidopsis* encodes three members named RNR1-3. RNR1 is co-targeted to both the mitochondria and the chloroplast where it responsible for the final

step in mitochondrial *atp9* mRNA 3' maturation (Perrin et al., 2004) and maturation of the 23S, 16S and 5S ribosomal RNA in the chloroplast (Kishine et al., 2004; Bollenbach et al., 2005). PNPase is a phosphate-dependent 3' to 5' exoribonuclease that degrades transcripts to mononucleotide diphosphates (Coburn et al., 1999). In chloroplasts, it forms a homotrimer (Baginsky et al., 2001) and its activity is stimulated by the presence of a poly(A) tail (Lisitsky et al., 1997). In *Arabidopsis* deficient for PNPase, mRNAs with extended 3' ends accumulate, as well as overall higher levels of mRNA (Walter et al., 2002). Also, in spinach, UV crosslinking experiments have found the PNPase directly bound to *petD 3'* UTR pre-mRNA (Chen and Stern, 1991). In *Chlamydomonas*, the role PNPase plays in RNA processing is currently unknown, but one clue may come from the mutant *crp3*, which has altered 3' maturation and degradation of transcripts from at least 5 genes in *Chlamydomonas* (Levy et al., 1997; Levy et al., 1999). It was hypothesized that *crp3* had reduced 3' to 5' exonuclease activity, but since *CRP3* has not been cloned, one cannot rule out that *crp3* represents a mutation in PNPase or RNase R, or in an accessory factor that regulates their function.

Degradation

When discussing the enzymes involved in 5' and 3' maturation, it is impossible to ignore the fact that they also play a role in degradation (Table 1.1). As in 3' maturation, the initiating step in chloroplast RNA degradation is thought to be endonuclease cleavage (Chen and Stern, 1991; Klaff, 1995; Kudla et al., 1996). Such endonuclease activities have been purified in two separate fractions from spinach chloroplasts and were called EndoC1 and EndoC2, which cleave in the loop and upstream of the IR, respectively (Chen

and Stern, 1991). Since the genes encoding EndoC1 and EndoC2 have not been cloned, RNase E/G and RNase J are prime candidates for these activities since both have been implicated in bulk RNA decay (Kushner, 2002; Even et al., 2005).

In addition to regulating other endonuclease activity, CSP41 can itself act as an endonuclease cleaving in the double-stranded stem of the IR and upstream of the IR when it is not in a complex (Yang et al., 1996). The binding of Mg^{2+} is required for this endonucleolytic activity, but not RNA binding so magnesium levels in the chloroplast may be a potential regulatory mechanism to switch between the RNA stability and RNA degradation functions (Bollenbach and Stern, 2003). CSP41 has also been found to be associated with ribosomes in *C. reinhardtii,* suggesting that degradation of mRNA transcripts may follow translation (Yamaguchi et al., 2003). When the cleavage site and optimal conditions for CSP41 activity is compared to those of EndoC1 and EndoC2, CSP41 appears distinct, thus it likely represents a third endonuclease activity involved in RNA degradation.

After endonuclease cleavage initiates degradation, it is likely that poly(A) is added to the transcripts (Komine et al., 2000). Unlike in the nucleus of eukaryotes where the addition of a poly(A) tail stabilizes the transcript, in prokaryotes, mitochondria and chloroplasts, it is a degradation signal most likely stimulating 3’ to 5’ exonucleases activity (Jacobson and Peltz, 1996; Blum et al., 1999; Schuster et al., 1999; Militello and Read, 2000). Organisms seem to follow two patterns of poly(A) tail synthesis, those whose tails are mainly adenosines and synthesized by PAP, and those who generate poly(A) tails and heteropolymeric tails that are not synthesized by PAP. *E. coli* and eukaryotes fall into the first class, having PAP as the polyadenylating enzyme

(Carpousis et al., 1999; Wahle and Ruegsegger, 1999). *C. reinhardtii* may fall into this class as well, since poly(A) tails in the chloroplast contain greater than 98% adenosines which is characteristic of PAP-synthesized tails (Komine et al., 2000), and preliminary data suggests that there is a chloroplast targeted PAP (Sara Zimmer, unpublished data). The second class includes *B. subtilis, Streptomyces coelicolor* and spinach chloroplasts (Yehudai-Resheff et al., 2001; Sohlberg et al., 2003; Campos-Guillen et al., 2005). Rather than PAP, it has been reported that *B. subtilis* uses a yet unknown polymerase (Campos-Guillen et al., 2005), while in *S. coelicolor* and spinach chloroplast, PNPase synthesizes the tails (Yehudai-Resheff et al., 2001; Sohlberg et al., 2003). Which class *Arabidopsis* chloroplasts belong to is currently unknown. In *Arabidopsis,* a PNPase knockout accumulated higher levels of polyadenylated RNA than in WT (Walter et al., 2002), suggesting that something other than PNPase synthesizes the poly(A) tails, but the presence of a PAP or the nature of its tails in the chloroplast have not been confirmed.

Polyadenylation is known to stimulate the activity of 3’ to 5’ exonucleases such as those contained in degradosomes, RNase II/R, PNPase, and in some cases the exosome (Lisitsky et al., 1997; Carpousis et al., 1999; Cheng and Deutscher, 2005; Milligan et al., 2005). The degradosomes are multiprotein complex scaffolded by RNase E, which cleaves at the 5’ end of transcripts as well as at the 3’ end near the poly(A) tail to initiate degradation (Kushner, 2002). The *E. coli* degradosome contains four proteins: RNase E, RhlB, enolase, and PNPase (Vanzo et al., 1998). RNase E acts as the scaffold of the degradosome having binding sites for RhlB, enolase, and PNPase in its C-terminal domain. RhlB is an ATP-dependent DEAD box helicase. It is thought to unwind secondary structure to increase the

processivity of exonucleases such as PNPase. Enolase is a glycolytic enzyme that regulates the stability of *ptsG* transcripts in response to metabolic stress (Morita et al., 2004), and PNPase has been discussed previously. The poly(A) tail recruits this complex acting as a landing pad from which the degradosome starts degrading mRNA. A second type of degradosome complex exists in *Pseudomonas syringae* Lz4W. It consists of three proteins: RNase E, RNase R and RhlE, which like RhlB is a DEAD-box helicase.

In organisms lacking degradosomes, degradation can be carried out by RNase R or PNPase alone or by a complex of 3' to 5' exonucleases called the exosome. RNase R has been implicated in degrading RNA with extensive secondary structure in *E. coli* (Cheng and Deutscher, 2005), but at least in *Arabidopsis* chloroplasts, does not appear to contribute to bulk mRNA degradation (Bollenbach et al., 2005). In *Arabidopsis*, plants deficient for PNPase accumulate overall higher levels of cpRNA (Walter et al., 2002), and purified PNPase is sufficient for transcript degradation in spinach (Yehudai-Resheff et al., 2001). *C. reinhardtii* engineered to underexpress PNPase show defects in degrading *atpB* transcripts engineered to contain poly(A) tails, further indicating its importance in chloroplast mRNA degradation (Nishimura et al., 2004). Taken together, these data suggests that PNPase play a role in chloroplast mRNA degradation.

Other 3' to 5' exonucleases are found in the yeast exosome, which can be found in both the cytoplasm and the nucleus. The cytoplasmic exosome contains ten 3'-5' exonucleases, a helicase, and two regulatory proteins, and targets unpolyadenylated mRNA in the cytoplasm (van Hoof and Parker, 1999). Of the 10 exonucleases, six, Ski6/Rrp41p, Rrp42p, Rrp43p, Rrp45p, Rrp46p and Mtr3p, have homology to the phosphate-dependent RNase PH in

E. coli (Burkard and Butler, 2000; Jacobs et al., 1998; Mitchell et al., 1997). One, Dis3p/Rrp44p, has homology to *E. coli* RNase II while three, Rrp4p, Rrp40p, and Cs14p, share homology to the S1 RNA binding domain. The helicase is Ski2p, which is in the DEVH box family of helicases. The exosome is also associated with Ski3p and Ski8p, which along with the helicases have been implicated in controlling the specificity of the complex (Jacobs et al., 1998). Curiously, the exosome forms a tertiary structure very similar to that of the PNPase trimer in spinach chloroplasts (Lorentzen et al., 2005), perhaps due to their functional similarity.

The nuclear exosome contain an eleventh exonuclease, Rrp6p, which is homologous to RNase D in *E. coli* and an additional DEVH box family helicase, Mtr4p (Burkard and Butler, 2000; Milligan et al., 2005). In addition to degrading unpolyadenylated RNA, the nuclear exosome has been reported to participate in degrading mRNAs with defective polyadenylation and removing poly(A) tails and extended 3' ends from 5.8S rRNA, snRNA and snoRNA, aiding in their maturation (van Hoof et al., 2000; Milligan et al., 2005). Perhaps the most interesting is the reported involvement of the exosome in the nuclear retention of RNA species that are unpolyadenylated or hyperpolyadenylated (Hilleren et al., 2001). This suggests that the exosome may be part of the system that prevents incorrectly processed transcripts from being transported into the cytoplasm; therefore, 3' maturation of nuclear mRNA transcripts may be required for translation as may be the case for the chloroplast *atpB* transcript.

A. thaliana may contain an exosome-like complex as well. AtRrp41p is a 3' to 5' exonuclease which is poly(A) tail-dependent (Chekanova et al., 2000). It is homologous to Rrp41p and can interact with yeast Rrp4p and

Rrp44p. Upon further sequence analysis in *Arabidopsis,* a Rrp4p homolog was found (Chekanova et al., 2002). This AtRrp4p also shows 3' to 5' exonuclease activity and is found in a 500 kDa complex that contains AtRrp41p and can bind mRNA. This strongly suggests that an exosome complex exists in higher plants.

A similar exosome like complex exists in yeast mitochondria. This mtEXO is involved in the degradation of unprocessed or aberrant RNA. It is similar to the degradosome in that it has DExH box RNA helicase, SUV3, and NTP dependent 3' to 5' exonuclease, DSS1 (Dziembowski et al., 2003). Like CSP41, mtEXO complex is found to associate with ribosomes, suggesting that degradation may follow translation (Dziembowski et al., 2003).

***Chlamydomonas* as a model system**

C. reinhardtii has been one of the key systems used in understanding chloroplast mRNA maturation and degradation. It is a unicellular alga that has a doubling time of eight hours under ideal conditions, a haploid nucleus, and a single chloroplast. Photosynthetic mutants can be maintained on media containing a reduced carbon source. Being haploid, recessive mutations are not masked, making self crossing unnecessary. Mutants can be crossed sexually to determine linkage and create double mutants, and vegetative diploids can be generated to determine dominance of a mutation. The chloroplast genome has been completely sequenced and the nuclear genome is nearing completion. Transformation of both chloroplast and nucleus is possible, making *C. reinhardtii* an important tool for answering basic biological questions such as identifying the enzymes that are involved in chloroplast mRNA maturation and degradation (Grossman, 2000; Grossman et al., 2003).

Potential pathways for RNA maturation and degradation in *C. reinhardtii* chloroplasts are shown in Figure 1.1. Although all of the activities shown in Figure 1.1 have been shown experimentally to occur, the precise enzymes involved are largely unproven with the exception of the role of PNPase in degradation (Nishimura et al., 2004). The most likely factors involved in endonuclease cleavage, polyadenylation, and exonuclease activity are shown. The goal of this work is to build on this model by identifying some of the nuclear proteins involved in the maturation of chloroplast RNA transcripts, as well as the RNA sequences involved in recruiting such proteins in *Chlamydomonas reinhardtii* chloroplasts.

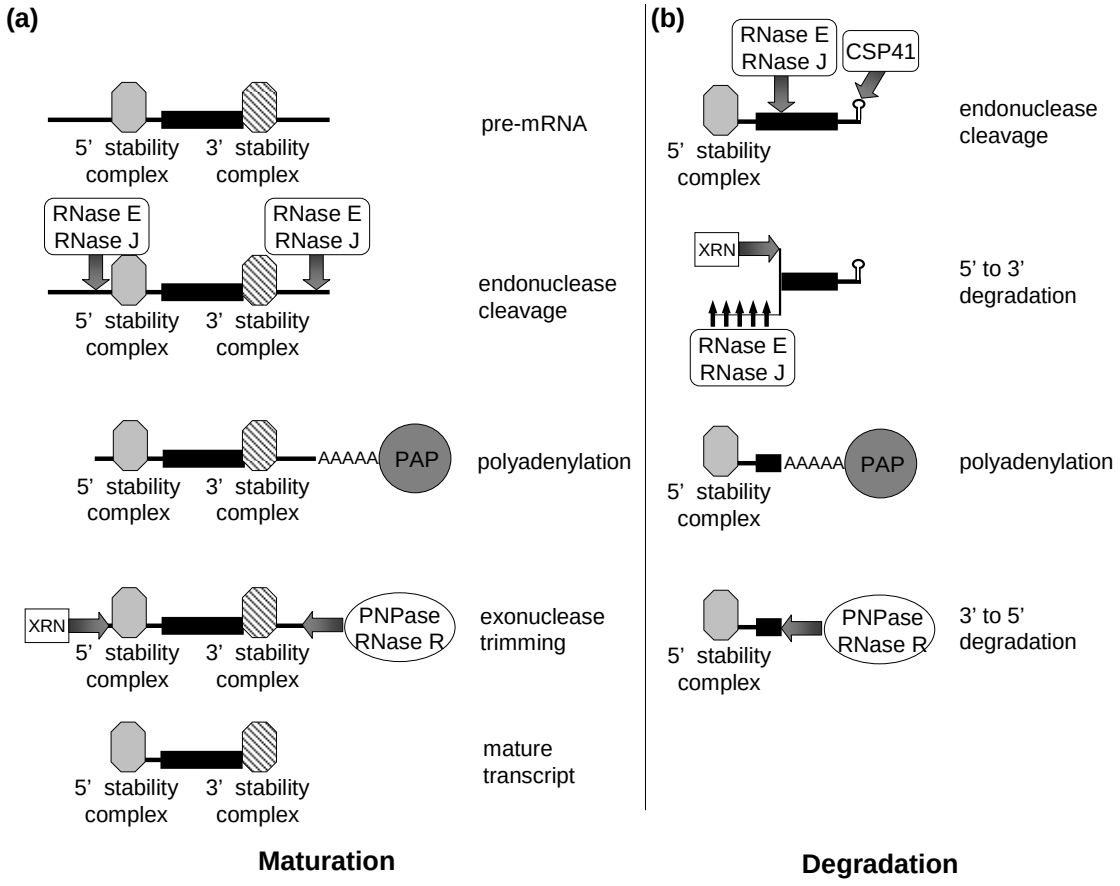


Figure 1.1. A model for chloroplast maturation and degradation pathways in *Chlamydomonas*. (a) The steps involved in RNA maturation are shown here as unordered, and for a generic RNA. Coding regions are shown as filled boxes, and inverted repeats known to stabilize the 3' ends of transcripts are shown as stem-loop structures. The 5' and 3' stability complexes have been described in the text. Endonuclease cleavages are most likely performed by an RNase E or RNase J-like enzyme. Poly(A) tails could be added by PNPase or PAP. A 5' to 3' exonuclease such as XRN and 3' to 5' exonucleases such as PNPase and RNase II complete the maturation. (b), The same enzymes are also involved in RNA degradation, which begins with endonuclease cleavages, followed by 5' to 3' degradation mediated by XRN and/or a wave of endonuclease cleavages; 3' to 5' degradation is stimulated by the addition of poly(A) tail followed by degradation by PNPase or RNase II.

REFERENCES

Anthonisen IL, Salvador ML, Klein U (2001) Specific sequence elements in the 5' untranslated regions of *rbcL* and *atpB* gene mRNA's stabilize transcripts in the chloroplast of *Chlamydomonas reinhardtii*. RNA **7:** 1024-1033

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

Baginsky S, Shteiman-Kotler A, Liveanu V, Yehudai-Resheff S, Bellaoui M, Settlage RE, Shabanowitz J, Hunt DF, Schuster G, Gruissem W (2001) Chloroplast PNPase exists as a homo-multimer enzyme complex that is distinct from the *Escherichia coli* degradosome. RNA **7:** 1464-1475.

Baker ME, Grundy WN, Elkan CP (1998) Spinach CSP41, an mRNA-binding protein and ribonuclease, is homologous to nucleotide-sugar epimerases and hydroxysteriod dehydrogenases. Biochem. Biophys. Res. Comm. **248:** 250-254

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

Barkan A (1988) Proteins encoded by a complex chloroplast transcription unit are each translated from both monocistronic and polycistronic RNAs. EMBO J. **7:** 2637-2644

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

Beardslee TA, Roy-Chowdhury S, Jaiswal P, Buhot L, Lerbs-Mache S, Stern DB, Allison LA (2002) A nucleus-encoded maize protein with sigma factor activity accumulates in mitochondria and chloroplasts. Plant J **31:** 199-209

Blowers AD, Klein U, Ellmore GS, Bogorad L (1993) Functional *in vivo* analyses of the 3' flanking sequences of the *Chlamydomonas* chloroplast *rbcL* and *psaB* genes. Mol. Gen. Genet. **238:** 339-349

Blum E, Carpousis AJ, Higgins CF (1999) Polyadenylation promotes degradation of 3'-structured RNA by the *Escherichia coli* mRNA degradosome *in vitro*. J. Biol. Chem. **274:** 4009-4016

Bollenbach TJ, Lange H, Gutierrez R, Erhardt M, Stern DB, Gagliardi D (2005) RNR1, a 3'-5' exoribonuclease belonging to the RNR superfamily, catalyzes 3' maturation of chloroplast ribosomal RNAs in *Arabidopsis thaliana*. Nucleic Acids Res. **33:** 2751-2763

Bollenbach TJ, Stern DB (2003) Divalent metal-dependent catalysis and cleavage specificity of CSP41, a chloroplast endoribonuclease belonging to the short chain dehydrogenase/reductase superfamily. Nucleic Acids Res. **31:** 4317-4325

Boudreau E, Nickelsen J, Lemaire SD, Ossenbuhl F, Rochaix J-D (2000) The *Nac2* gene of *Chlamydomonas reinhardtii* encodes a chloroplast TPR protein involved in *psbD* mRNA stability, processing and/or translation. EMBO J. **19:** 3366-3376

Bruick RK, Mayfield SP (1998) Processing of the *psbA* 5' untranslated region in *Chlamydomonas reinhardtii* depends upon factors mediating ribosome association. J. Cell. Biol. **143:** 1145-1153

Burkard KTD, Butler JS (2000) A nuclear 3'-5' exonuclease involved in mRNA degradation interacts with poly(A) polymerase and the hnRNA protein Npl3p. Mol. Cell. Biol. **20:** 604-616

Cahoon AB, Stern DB (2001) Plastid transcription: A ménage à trois? Trends Plant Sci. **6:** 45-46

Campos-Guillen J, Bralley P, Jones GH, Bechhofer DH, Olmedo-Alvarez G (2005) Addition of poly(A) and heteropolymeric 3' ends in *Bacillus subtilis* wild-type and polynucleotide phosphorylase-deficient strains. J. Bacteriol. **187:** 4698-4706

Cannistraro VJ, Kennell D (1985) The 5' ends of *Escherichia coli lac* mRNA. J. Mol. Biol. **182:** 241-248

Carpousis AJ, Vanzo NF, Raynal LC (1999) mRNA degradation. A tale of poly(A) and multiprotein machines. Trends Genet. **15:** 24-28

Carter ML, Smith AC, Kobayashi H, Purton S, Herrin DL (2004) Structure, circadian regulation and bioinformatic analysis of the unique sigma factor gene in *Chlamydomonas reinhardtii*. Photosynthesis Res. **82:** 339-349

Chanfreau G, Buckle M, Jacquier A (2000) Recognition of a conserved class of RNA tetraloops by *Saccharomyces cerevisiae* RNase III. Proc. Natl. Acad. Sci. U.S.A **97:** 3142-3147

Chen H, Stern DB (1991) Specific ribonuclease activities in spinach chloroplasts promote mRNA maturation and degradation. J. Biol. Chem. **266:** 24205-24211

Chen HC, Stern DB (1991) Specific binding of chloroplast proteins *in vitro* to the 3' untranslated region of spinach chloroplast *petD* messenger RNA. Mol. Cell. Biol. **11:** 4380-4388

Chen Q, Adams CC, Usack L, Yang J, Monde R, Stern DB (1995) An AU-rich element in the 3' untranslated region of the spinach chloroplast *petD* gene participates in sequence-specific RNA-protein complex formation. Mol. Cell. Biol. **15:** 2010-2018

Cheng ZF, Deutscher MP (2005) An important role for RNase R in mRNA decay. Mol. Cell **17:** 313-318

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

Dauvillee D, Stampacchia O, Girard-Bascou J, Rochaix JD (2003) Tab2 is a novel conserved RNA binding protein required for translation of the chloroplast *psaB* mRNA. EMBO J. **22:** 6378-6388

Drager RG, Girard-Bascou J, Choquet Y, Kindle KL, Stern DB (1998) *In vivo* evidence for 5' to 3' exoribonuclease degradation of an unstable chloroplast mRNA. Plant J. **13:** 85-96

Drager RG, Higgs DC, Kindle KL, Stern DB (1999) 5’ to 3’ exoribonucleolytic activity is a normal component of chloroplast mRNA decay pathways. Plant J. **19:** 521-532

Drager RG, Zeidler M, Simpson CL, Stern DB (1996) A chloroplast transcript lacking the 3' inverted repeat is degraded by 3' to 5' exoribonuclease activity. RNA **2:** 652-663

Dziembowski A, Piwowarski J, Hoser R, Minczuk M, Dmochowska A, Siep M, van der Spek H, Grivell L, Stepien PP (2003) The yeast mitochondrial degradosome. Its composition, interplay between RNA helicase and RNase activities and the role in mitochondrial RNA metabolism. J. Biol. Chem. **278:** 1603-1611

Esposito D, Hicks AJ, Stern DB (2001) A role for initiation codon context in chloroplast translation. Plant Cell **13:** 2373-2384.

Even S, Pellegrini O, Zig L, Labas V, Vinh J, Brechemmier-Baey D, Putzer H (2005) Ribonucleases J1 and J2: two novel endoribonucleases in *B.subtilis* with functional homology to *E.coli* RNase E. Nucleic Acids Res. **33:** 2141-2152

Fargo DC, Boynton JE, Gillham NW (1999) Mutations altering the predicted secondary structure of a chloroplast 5' untranslated region affect its physical and biochemical properties as well as its ability to promote translation of reporter mRNAs both in the *Chlamydomonas reinhardtii* chloroplast and in *Escherichia coli*. Mol. Cell. Biol. **19:** 6980-6990

Fargo DC, Zhang M, Gillham NW, Boynton JE (1998) Shine-Dalgarno-like sequences are not required for translation of chloroplast mRNAs in *Chlamydomonas reinhardtii* chloroplasts or in *Escherichia coli*. Mol. Gen. Genet. **257:** 271-282

Felder S, Meierhoff K, Sane AP, Meurer J, Driemel C, Plucken H, Klaff P, Stein B, Bechtold N, Westhoff P (2001) The nucleus-encoded *HCF107* gene of *Arabidopsis* provides a link between intercistronic RNA processing and the

accumulation of translation-competent *psbH* transcripts in chloroplasts. Plant Cell **13:** 2127-2141.

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

Grossman AR (2000) *Chlamydomonas reinhardtii* and photosynthesis: genetics to genomics. Curr. Opin. Plant Biol. **3:** 132-137

Grossman AR, Harris E, Hauser C, Lefebvre P, Martinez D, Rokhsar D, Shrager J, Silflow C, Stern D, Vallon O, Zhang Z (2003) *Chlamydomonas reinhardtii* at the crossroads of genomics. Euk. Cell **2:** 1137-1150

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

Hauser CR, Gillham NW, Boynton JE (1996) Translational regulation of chloroplast genes. Proteins binding to the 5'-untranslated regions of chloroplast mRNAs in *Chlamydomonas reinhardtii*. J. Biol. Chem. **271:** 1486-1497

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

Henkin TM (2000) Transcription termination control in bacteria. Curr. Opin. Microbiol. **3:** 149-153

Hicks A, Drager RG, Higgs DC, Stern DB (2002) An mRNA 3' processing site targets downstream sequences for rapid degradation in *Chlamydomonas* chloroplasts. J. Biol. Chem. **277:** 3325-3333

Higgs DC, Shapiro RS, Kindle KL, Stern DB (1999) Small *cis*-acting sequences that specify secondary structures in a chloroplast mRNA are essential for RNA stability and translation. Mol. Cell. Biol. **19:** 8479-8491

Hilleren P, McCarthy T, Rosbash M, Parker R, Jensen TH (2001) Quality control of mRNA 3'-end processing is linked to the nuclear exosome. Nature **413:** 538-542.

Hofmann TJ, Min J, Zassenhaus HP (1993) Formation of the 3' end of yeast mitochondrial mRNAs occurs by site-specific cleavage two bases downstream of a conserved dodecamer sequence. Yeast **9:** 1319-1330

Jacobson A, Peltz SW (1996) Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. Annu. Rev. Biochem. **65:** 693-739

Jiang X, Belasco JG (2004) Catalytic activation of multimeric RNase E and RNase G by 5'-monophosphorylated RNA. Proc. Natl. Acad. Sci. USA **101:** 9211-9216

Jiao HS, Hicks A, Simpson C, Stern DB (2004) Short dispersed repeats in the *Chlamydomonas* chloroplast genome are collocated with sites for mRNA 3' end formation. Curr. Genet. **45:** 311-322

Kanamaru K, Tanaka K (2004) Roles of chloroplast RNA polymerase sigma factors in chloroplast development and stress response in higher plants. Biosci. Biotechnol. Biochem. **68:** 2215-2223

Kasai K, Kawagishi-Kobayashi M, Teraishi M, Ito Y, Ochi K, Wakasa K, Tozawa Y (2004) Differential expression of three plastidial sigma factors, OsSIG1, OsSIG2A, and OsSIG2B, during leaf development in rice. Biosci. Biotechnol. Biochem. **68:** 973-977

Kastenmayer JP, Green PJ (2000) Novel features of the XRN-family in Arabidopsis: evidence that AtXRN4, one of several orthologs of nuclear Xrn2p/Rat1p, functions in the cytoplasm. Proc Natl Acad Sci USA **97:** 13985-13990.

Kastenmayer JP, Johnson MA, Green PJ (2001) Analysis of XRN orthologs by complementation of yeast mutants and localization of XRN-GFP fusion proteins. Methods Enzymol **342:** 269-282

Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt JF, Buratowski S (2004) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature **432:** 517-522

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

Klaff P (1995) mRNA decay in spinach chloroplasts: *psbA* mRNA degradation is initiated by endonucleolytic cleavages within the coding region. Nucleic Acids Res. **23:** 4885-4892

Komine Y, Kwong L, Anguera MC, Schuster G, Stern DB (2000) Polyadenylation of three classes of chloroplast RNA in *Chlamydomonas reinhardtii*. RNA **6:** 598-607

Kuchka MR, Goldschmidt-Clermont M, van Dillewijn J, Rochaix JD (1989) Mutation at the *Chlamydomonas* nuclear *NAC2* locus specifically affects stability of the chloroplast *psbD* transcript encoding polypeptide D2 and PS II. Cell **58:** 869-876

Kudla J, Hayes R, Gruissem W (1996) Polyadenylation accelerates degradation of chloroplast mRNA. EMBO J. **15:** 7137-7146

Kushner SR (2002) mRNA decay in *Escherichia coli* comes of age. J. Bacteriol. **184:** 4658-4665

Lee A, Henras AK, Chanfreau G (2005) Multiple RNA surveillance pathways limit aberrant expression of iron uptake mRNAs and prevent iron toxicity in *S. cerevisiae*. Mol. Cell **19:** 39-51

Lee CY, Lee A, Chanfreau G (2003) The roles of endonucleolytic cleavage and exonucleolytic digestion in the 5'-end processing of *S. cerevisiae* box C/D snoRNAs. RNA **9:** 1362-1370

Lee H, Bingham SE, Webber AN (1996) Function of 3' non-coding sequences and stop codon usage in expression of the chloroplast *psaB* gene in *Chlamydomonas reinhardtii*. Plant Mol. Biol. **31:** 337-354

Levy H, Kindle KL, Stern DB (1997) A nuclear mutation that affects the 3' processing of several mRNAs in *Chlamydomonas* chloroplasts. Plant Cell **9:** 825-836

Levy H, Kindle KL, Stern DB (1999) Target and specificity of a nuclear gene product that participates in mRNA 3' end formation in *Chlamydomonas* chloroplasts. J. Biol. Chem. **274:** 35955-35962

Li Z, Deutscher MP (2002) RNase E plays an essential role in the maturation of *Escherichia coli* tRNA precursors. RNA **8:** 97-109

Li Z, Pandit S, Deutscher MP (1999) RNase G (CafA protein) and RNase E are both required for the 5' maturation of 16S ribosomal RNA. EMBO J. **18:** 2878-2885

Lisitsky I, Kotler A, Schuster G (1997) The mechanism of preferential degradation of polyadenylated RNA in the chloroplast: The exoribonuclease 100RNP-polynucleotide phosphorylase displays high binding affinity for poly(A) sequences. J. Biol. Chem. **272:** 17648-17653

Lisitsky I, Liveanu V, Schuster G (1995) RNA-binding characteristics of a ribonucleoprotein from spinach chloroplasts. Plant Physiol. **107:** 933-941

Lisitsky I, Schuster G (1995) Phosphorylation of a chloroplast RNA-binding protein changes its affinity to RNA. Nucleic Acids Res. **23:** 2506-2511

Lopes-Juez E, Pyke K (2005) Platids unleashed: their development and their integration in plant development. Int. J. Dev. Biol. **49:** 557-577

Lopez-Juez E, Pyke KA (2005) Plastids unleashed: their development and their integration in plant development. Int. J. Dev. Biol. **49:** 557-577

Lorentzen E, Walter P, Fribourg S, Evguenieva-Hackenberg E, Klug G, Conti E (2005) The archaeal exosome core is a hexameric ring structure with three catalytic subunits. Nat. Struct. Mol. Biol. **12:** 575-581

Lown FJ, Watson AT, Purton S (2001) *Chlamydomonas* nuclear mutants that fail to assemble respiratory or photosynthetic electron transfer complexes. Biochem. Soc. Trans. **29:** 452-455

Lurin C, Andres C, Aubourg S, Bellaoui M, Bitton F, Bruyere C, Caboche M, Debast C, Gualberto J, Hoffmann B, Lecharny A, Le Ret M, Martin-Magniette ML, Mireau H, Peeters N, Renou JP, Szurek B, Taconnat L, Small I (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell **16:** 2089-2103

Mayfield SP, Cohen A, Danon A, Yohn CB (1994) Translation of the *psbA* mRNA of *Chlamydomonas reinhardtii* requires a structured RNA element contained within the 5' untranslated region. J. Cell Biol. **127:** 1537-1545

McCormac D, Barkan A (1999) A nuclear gene in maize required for the translation of the chloroplast *atpB/E* mRNA. Plant Cell **11:** 1709-1716

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

Memon AR, Meng B, Mullet JE (1996) RNA-binding proteins of 37/38 kDa bind specifically to the barley chloroplast *psbA* 3'-end untranslated RNA. Plant Mol. Biol. **30:** 1195-1205

Militello KT, Read LK (2000) UTP-dependent and -independent Pathways of mRNA turnover in *Trypanosoma brucei* mitochondria. Mol. Cell. Biol. **20:** 2308-2316

Milligan L, Torchet C, Allmang C, Shipman T, Tollervey D (2005) A nuclear surveillance pathway for mRNAs with defective polyadenylation. Mol. Cell Biol. **25:** 9996-10004

Monde RA, Greene JC, Stern DB (2000) The sequence and secondary structure of the 3'-UTR affect 3'-end maturation, RNA accumulation, and translation in tobacco chloroplasts. Plant Mol. Biol. **44:** 529-542

Monde RA, Schuster G, Stern DB (2000) Processing and degradation of chloroplast mRNA. Biochimie **82:** 573-582

Morita T, Kawamoto H, Mizota T, Inada T, Aiba H (2004) Enolase in the RNA degradosome plays a crucial role in the rapid decay of glucose transporter mRNA in the response to phosphosugar stress in *Escherichia coli*. Mol. Microbiol. **54:** 1063-1075

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

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

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

Nishimura Y, Kikis EA, Zimmer SL, Komine Y, Stern DB (2004) Antisense transcript and RNA processing alterations suppress instability of polyadenylated mRNA in *Chlamydomonas* chloroplasts. Plant Cell **16:** 2849-2869

Perrin R, Meyer EH, Zaepfel M, Kim YJ, Mache R, Grienenberger JM, Gualberto JM, Gagliardi D (2004) Two exoribonucleases act sequentially to process mature 3'-ends of *atp9* mRNAs in *Arabidopsis* mitochondria. J. Biol. Chem. **279:** 25440-25446

Rochaix J-D, Kuchka M, Mayfield S, Schirmer-Rahire M, Girard-Bascou J, Bennoun P (1989) Nuclear and chloroplast mutations affect the synthesis or stability of the chloroplast *psbC* gene product in *Chlamydomonas reinhardtii*. EMBO J. **8:** 1013-1022

Rott R, Drager RG, Stern DB, Schuster G (1996) The 3' untranslated regions of chloroplast genes in *Chlamydomonas reinhardtii* do not serve as efficient transcriptional terminators. Mol. Gen. Genet. **252:** 676-683

Rott R, Levy H, Drager RG, Stern DB, Schuster G (1998) 3'-processed mRNA is preferentially translated in *Chlamydomonas reinhardtii* chloroplasts. Mol. Cell. Biol. **18:** 4605-4611

Rott R, Liveanu V, Drager RG, Stern DB, Schuster G (1998) The sequence and structure of the 3'-untranslated regions of chloroplast transcripts are important determinants of mRNA accumulation and stability. Plant Mol. Biol. **36:** 307-314

Sakamoto W, Chen X, Kindle KL, Stern DB (1994) Function of the *Chlamydomonas reinhardtii petD* 5' untranslated region in regulating the accumulation of subunit IV of the cytochrome b_6/f complex. Plant J. **6:** 503-512

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

Sane AP, Stein B, Westhoff P (2005) The nuclear gene *HCF107* encodes a membrane-associated R-TPR (RNA tetratricopeptide repeat)-containing protein involved in expression of the plastidial psbH gene in *Arabidopsis*. Plant J. **42:** 720-730

Schuster G, Gruissem W (1991) Chloroplast mRNA 3' end processing requires a nuclear-encoded RNA-binding protein. EMBO J. **10:** 1493-1502

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

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

Slomovic S, Portnoy V, Liveanu V, Schuster G (2005) RNA polyadenylation in prokaryotes and organelles: Different tails tell different tales. Crit. Rev. Plant Sci.**:** in press

Sohlberg B, Huang J, Cohen SN (2003) The *Streptomyces coelicolor* polynucleotide phosphorylase homologue, and not the putative poly(A) polymerase, can polyadenylate RNA. J. Bacteriol. **185:** 7273-7278

Souret FF, Kastenmayer JP, Green PJ (2004) AtXRN4 degrades mRNA in Arabidopsis and its substrates include selected miRNA targets. Mol. Cell **15:** 173-183

Stern DB, Gruissem W (1987) Control of plastid gene expression: 3' inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. Cell **51:** 1145-1157

Stern DB, Jones H, Gruissem W (1989) Function of plastid mRNA 3' inverted repeats: RNA stabilization and gene-specific protein binding. J. Biol. Chem. **264:** 18742-18750

Stern DB, Kindle KL (1993) 3' end maturation of the *Chlamydomonas reinhardtii* chloroplast *atpB* mRNA is a two-step process. Mol. Cell. Biol. **13:** 2277-2285

Sturm N, Kuras R, Buschlen S, Sakamoto W, Kindle KL, Stern DB, Wollman FA (1994) The *petD* gene is transcribed by functionally redundant promoters in *Chlamydomonas reinhardtii* chloroplasts. Mol. Cell. Biol. **14:** 6171-6179

Suay L, Salvador ML, Abesha E, Klein U (2005) Specific roles of 5' RNA secondary structures in stabilizing transcripts in chloroplasts. Nucleic Acids Res. **33:** 4754-4761

Vaistij FE, Boudreau E, Lemaire SD, Goldschmidt-Clermont M, Rochaix JD (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*. Proc. Natl. Acad. Sci. USA **97:** 14813-14818

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

van Hoof A, Lennertz P, Parker R (2000) Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. Mol. Cell Biol. **20:** 441-452

van Hoof A, Parker R (1999) The exosome: a proteasome for RNA? Cell **99:** 347-350

Wachi M, Umitsuki G, Nagai K (1997) Functional relationship between Escherichia coli RNase E and the CafA protein. Mol Gen Genet **253:** 515-519

Wahle E, Ruegsegger U (1999) 3'-End processing of pre-mRNA in eukaryotes. FEMS Microbiol. Rev. **23:** 277-295

Walter M, Kilian J, Kudla J (2002) PNPase activity determines the efficiency of mRNA 3'-end processing, the degradation of tRNA and the extent of polyadenylation in chloroplasts. EMBO J. **21:** 6905-6914

Westhoff P, Herrmann RG (1988) Complex RNA maturation in chloroplasts: the *psbB* operon from spinach. Eur. J. Biochem. **171:** 551-564

Wu H, Henras A, Chanfreau G, Feigon J (2004) Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. Proc. Natl. Acad. Sci. U.S.A. **101:** 8307-8312

Xue Y, Bai X, Lee I, Kallstrom G, Ho J, Brown J, Stevens A, Johnson AW (2000) *Saccharomyces cerevisiae* RAI1 (YGL246c) is homologous to human DOM3Z and encodes a protein that binds the nuclear exoribonuclease Rat1p. Mol. Cell Biol. **20:** 4006-4015

Yamaguchi K, Beligni MV, Prieto S, Haynes PA, McDonald WH, Yates JR, 3rd, Mayfield SP (2003) Proteomic characterization of the *Chlamydomonas reinhardtii* chloroplast ribosome. Identification of proteins unique to the 70S ribosome. J. Biol. Chem. **278:** 33774-33785

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

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

Yehudai-Resheff S, Hirsh M, Schuster G (2001) Polynucleotide phosphorylase functions as both an exonuclease and a poly(A) polymerase in spinach chloroplasts. Mol. Cell. Biol. **21:** 5408-5416.

Zer C, Chanfreau G (2005) Regulation and surveillance of normal and 3'-extended forms of the yeast aci-reductone dioxygenase mRNA by RNase III cleavage and exonucleolytic degradation. J. Biol. Chem. **280:** 28997-29003

Zerges W (2000) Translation in chloroplasts. Biochimie **82:** 583-601

CHAPTER 2.

NUCLEAR SUPPRESSORS DEFINE THREE FACTORS THAT PARTICIPATE IN BOTH 5' AND 3' END PROCESSING OF MRNAS IN *CHLAMYDOMONAS* CHLOROPLASTS *

ABSTRACT

Chloroplast RNA processing and degradation are orchestrated by nucleus-encoded factors. Although several transcript-specific factors have been identified, those involved globally in RNA metabolism have mostly remained elusive. Using *Chlamydomonas reinhardtii*, three pleiotropic nuclear mutations, *mcd3, mcd4, and mcd5*, have been identified, which cause quantitative variation between polycistronic transcripts and accumulation of novel transcripts with 3' ends. The *mcd3, mcd4,* and *mcd5* mutants were initially isolated as photoautotrophic suppressors of the *petD* 5' mutants LS2 and LS6, which harbor four nucleotide linker-scanning mutations near the 5' end of the mature transcript. The LS mutants accumulate 1-3% of the wild-type (WT) *petD* mRNA level and no cytochrome b_6/f complex subunit IV, which is the *petD* gene product and required for photosynthesis. Each suppressor restores approximately 15% of the WT *petD* mRNA and subunit IV levels. Genetic analysis showed *mcd4* to be recessive, and suggested that MCD4 interacts with the *petD* mRNA stability factor MCD1. To assess the specificity of *mcd3, mcd4, and mcd5*, transcripts from 32 chloroplast genes were analyzed by RNA filter hybridizations. *mcd3* and *mcd4* displayed aberrant transcript patterns for 17 genes, whereas only three were altered in *mcd5.*

* Rymarquis, L., Higgs, D., Stern, D. Nuclear suppressors define three factors that participate in both 5' and 3' end processing of mRNAs in *Chlamydomonas* chloroplasts. *Plant Journal.* In press

Since the mutations affect multiple RNAs in a variety of ways, my data suggest that MCD3, MCD4, and MCD5 may participate in a series of multiprotein complexes responsible for RNA maturation and degradation in *Chlamydomonas* chloroplasts.

INTRODUCTION

The photosynthetic complexes are comprised of proteins encoded by both the nuclear and chloroplast genomes, necessitating coordinated expression of genes from the two compartments. Where nuclear mutants defective in chloroplast gene expression have been sought through screens for nonphotosynthetic phenotypes, most affect post-transcriptional steps including RNA processing and stability, whether isolated in higher plants (Barkan and Goldschmidt-Clermont, 2000; Stern *et al.*, 2004) or the unicellular model alga *Chlamydomonas reinhardtii* (Herrin and Nickelsen, 2004). In *Chlamydomonas*, these nucleus-encoded factors interact with *cis* elements found at the 5' and 3' ends of mRNAs (reviewed in Herrin and Nickelsen, 2004) or in one case, the coding region (Drapier *et al.*, 2002). One hallmark of such mutants is their specificity; in general a single gene cluster or RNA appears to be the primary target. Exceptions include maize *crs2*, which affects splicing of numerous introns (Jenkins *et al.*, 1997) and *crp1*, which affects expression of both *petA* and *petD* (Barkan *et al.*, 1994).

Candidates for the mRNA *cis* elements responsible for recruiting the nucleus-encoded factors are inverted repeats (IRs), which are found in both the 5' and 3' untranslated regions (UTRs). IRs in the 5' UTR have been described for *Chlamydomonas atpB, rbcL, petD, rps7, psbB, psbC*, *psbD* and *psbA.* Deleting or mutating them destabilizes the RNA, leading to reduced

transcript accumulation and translation, in some cases due to failure to interact with cognizant nucleus-encoded factors (Rochaix et al., 1989; Nickelsen et al., 1994; Fargo et al., 1999; Higgs et al., 1999; Vaistij et al., 2000; Anthonisen et al., 2001; Salvador et al., 2004; Suay et al., 2005). IRs found in the 3' UTR are associated with RNA stability, preventing 3' to 5' exonucleolytic degradation of transcripts (reviewed in Bollenbach *et al.*, 2004), and may also interact with specific nuclear factors (Levy *et al.*, 1999; Meierhoff *et al.*, 2003).

One case examined in detail is the IR which spans nucleotides (nt) 2-12 of the mature *petD* transcript of *Chlamydomonas*. Linker-scanning mutations in nt 2-5 (LS2), 6-9 (LS6) or 10-13 (LS10) decrease *petD* RNA stability (Higgs *et al.*, 1999), and LS2 and LS6 are additionally defective in *petD* translation. Given the modest effect of LS10, nt 2-9 were defined as a regulatory domain termed Element I, which was presumed to be recognized by one or more nucleus-encoded proteins. One of these proteins is MCD1 (maturation of cytochrome b6/f *petD*), which is encoded by the nuclear *MCD1* gene. Three nonphotosynthetic mutant alleles of *MCD1* were isolated in screens for high chlorophyll fluorescence after WT cells were mutagenized by 5-fluorouracil (*mcd1-1*) or UV (*mcd1-2*), or in a similar screen where a tagged allele was created by *ARG7* DNA insertion (*mcd1-3*) (Drager *et al.*, 1998; Murakami *et al.*, 2005). *mcd1-1* and *mcd1-3* are deletion mutants, whereas *mcd1-2* contains a premature stop codon, leading to a truncated protein. Despite normal transcription of *petD*, all three mutants fail to accumulate *petD* mRNA and subunit IV (SUIV), which is the *petD* gene product.

The mechanism of *petD* degradation in *mcd1* mutants was elucidated by the introduction of polyguanosine (pG) tracts at +25 and/or +165 relative to the *petD* 5' end. In *mcd1-1* strains harboring either or both pG tracts, *petD*

mRNA accumulation was restored, with the pG motif found at the 5' end of the transcripts (Drager *et al.*, 1999). Since pG tracts form a complex tertiary structure through which known exonucleases cannot progress, this accumulation suggested that the MCD1 protects *petD* transcripts from 5' to 3' degradation, presumably by binding to Element I. *MCD1* has been recently cloned and is predicted to encode a 1,553 amino acid protein with a molecular mass of 156.1 kDa, but exhibits no homology to any known motif or protein (Murakami *et al.*, 2005). Since it has no known RNA binding motifs, MCD1 is presumed to interact with other proteins in its protective function, and analogous proteins that protect *psbB* (Vaistij et al., 2000) and *psbD* (Boudreau *et al.*, 2000) mRNAs are indeed found in high molecular weight complexes.

In an attempt to isolate the proteins which potentially interacted with MCD1 and the *petD* 5' UTR, a suppressor screen for photosynthetic revertants was carried out beginning with LS2 and LS6, the two linker-scanning mutants which define Element I. In this manner three loci were identified where mutations had increased the stability of *petD* mRNA, and its translation. When these mutants, *mcd3, mcd4,* and *mcd5*, were analyzed further, it became apparent that they were not *petD*-specific. Along with affecting *petD* stability and translation*,* they collectively are involved in events occurring at the 5' and 3' ends of transcripts from 17 chloroplast genes. Their pleiotropic nature suggests that they are components of one or more multiprotein complexes responsible for endonucleolytic cleavages that affect multiple aspects of RNA metabolism.

RESULTS

mcd3, mcd4, and *mcd5* are nuclear suppressors of LS2 and LS6

The strains LS2 and LS6 are nonphotosynthetic due to mutations in nt 2-5 or 6-9, respectively, of the mature *petD* transcript, as reported previously (Higgs *et al.*, 1999). Nucleotides 2-9 are designated Element I, which is involved in RNA stability as well as translation, whereas mutations in Elements II and III cause translational defects, but do not affect transcript stability (Figure 2.1a; Higgs et al., 1999).

To identify possible *trans*-acting factors which interact with Element I, suppressor screens were carried out by plating LS2 and LS6 cells on medium lacking acetate by Dave Higgs. This screen yielded the spontaneous photosynthetic mutants *mcd3* and *mcd4,* which were derived from LS2, whereas *mcd5* was only recovered from LS6 after UV mutagenesis. These were subjected to allelism tests and backcrosses to determine that they were nuclear mutations, as discussed below, and a single representative allele of each locus was selected for detailed study. In the following experiments, as appropriate, the chloroplast genotype is given in brackets. Thus, *mcd4* [LS2] contains the *petD* LS2 mutation and the nuclear *mcd4* mutation. Where alleles are not given, the genotype is WT. “LS2” or “LS6” alone designate the progenitor strains carrying chloroplast mutations, but a WT nuclear genome.

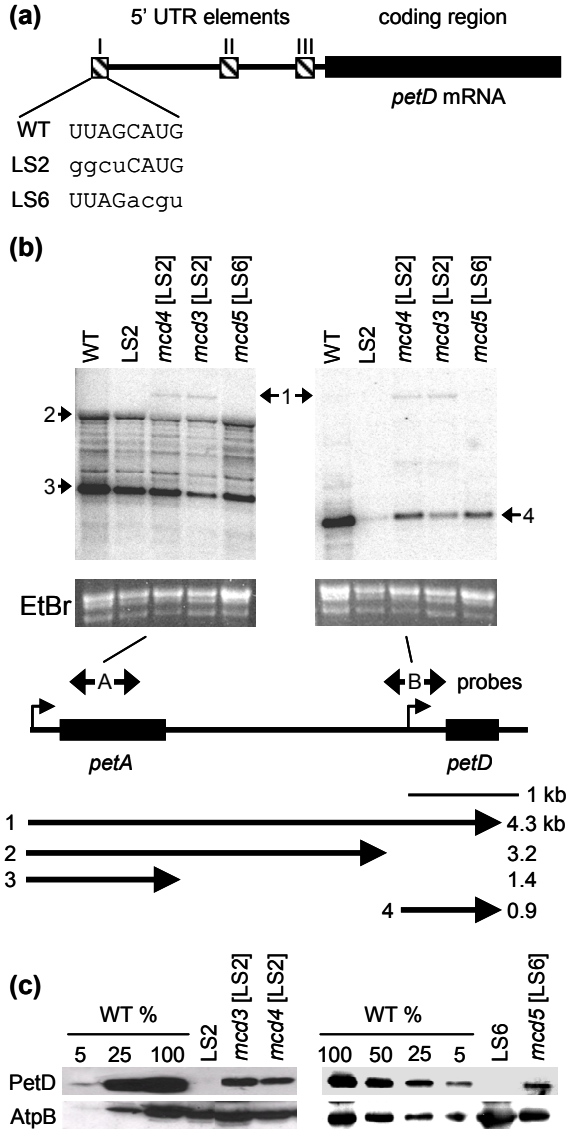


Figure 2.1. *mcd3, mcd4, and mcd5* accumulate *petD* mRNA and protein. (a) A diagram of *petD* mRNA is shown, with the 5' UTR denoted by a line and the coding region by a filled box. Hatched boxes indicate elements required for RNA stability (I) and translation (I-III). Comparison of the +2 to +9 sequence of LS2 and LS6 mutations to WT is shown. (b) RNA filters were probed with *petA* or *petD*. The EtBr-stained gels are shown as loading controls. A diagram of the *petA-petD* region is shown with the transcription start sites marked by bent arrows. The major mRNA species are shown underneath. (c) PetD protein accumulation was measured by immunoblot analysis, using AtpB as a loading control.

To discover the mechanism behind the restoration of photosynthesis in these suppressors, RNA filter hybridizations and immunoblot analyses were performed. These revealed that *mcd3, mcd4, and mcd5* accumulate 15% of the WT monocistronic *petD* mRNA level (Figure 2.1b, transcript 4) and 15-20% of the WT SUIV level (Figure 2.1c). This indicates that *mcd3, mcd4,* and *mcd5* stabilize *petD* transcripts. In *mcd3* and *mcd4,* dicistronic *petA-petD* transcripts also accumulate (Figure 2.1b, transcript 1). Thus, *mcd3* and *mcd4* play a role in accumulation of the *petA-petD* dicistronic message, whereas all three suppressors affect *petD* RNA stability and translation.

To determine whether *mcd3, mcd4,* and *mcd5* defined chloroplast or nuclear mutations, *mcd3* [LS2] mt+, *mcd4* [LS2] mt+, and *mcd5* [LS6] mt+ were crossed to a WT mt- tester strain, CC-3681 (Table 2.1, crosses 1-3). *C. reinhardtii* progeny usually inherit the cpDNA of the mt+ strain. Thus, if *mcd3, mcd4,* and *mcd5* were chloroplast mutations, all of the progeny would inherit the suppressor mutation along with the LS2 or LS6 chloroplast, and thus possess a photosynthetic (PS+) phenotype. If they were nuclear mutations, only two progeny per tetrad would inherit the suppressor mutation, giving a 2 PS+:2 PS- (PS-; nonphotosynthetic and acetate-requiring) segregation. When the photosynthetic ability of random progeny was monitored by growth on minimal medium, PS- progeny were produced 21-50% of the time for the three crosses, indicating that *mcd3, mcd4,* and *mcd5* are nuclear mutations.

To ascertain whether *mcd3* and *mcd4* represented unique loci or multiple alleles of the same locus, *mcd4* [LS2] mt+ was crossed to *mcd3* [LS2] mt- (Table 2.1, cross 4). In this cross, PS- progeny represent the genotype *MCD3 MCD4* [LS2], i.e. not carrying either suppressor mutation, and can only

be produced if the mutant loci are separable by recombination. When random progeny were analyzed for photosynthetic competence, 24% of them were PS-, revealing that *mcd3* and *mcd4* are unlinked. Recombination tests with *mcd5* were not conducted in the same manner, since *mcd4* does not suppress LS6 (Table 2.1, cross 7). Instead, since molecular mapping had shown that *mcd4* is 3 cM from *PF12* (Rymarquis *et al.*, 2005), the linkage of *mcd3* and *mcd5* to *PF12* was determined by crossing them to CC-610 mt-, which cannot swim due to a *pf12* mutation (Table 2.1, crosses 5-6). These crosses revealed that *mcd3* lies 12 cM from *pf12*, and that *mcd5* is unlinked to *pf12*. Thus, *mcd3, mcd4, and mcd5* are mutations at different loci, although *mcd4* and *mcd3* appear to lie on the same linkage group.

Table 2.1. Genetic analysis of *mcd3, mcd4, and mcd5* mutations.

Cross	Cross mt+ X mt-	Number analyzed	Segregation[ab]	Distance in cM[c]
1	*mcd3* [LS2] X CC-3681	24 random progeny	19/5	
2	*mcd4* [LS2] X CC-3681	10 random progeny	6/4	
3	*mcd5* [LS6] X CC-3681	16 random progeny	8/8	
4	*mcd4* [LS2] X *mcd3* [LS2]	34 random progeny	26/8	
*	*mcd4* [LS2] X CC-610	33 tetrads	31 PD, 2 T, 0 NPD	3 cM
5	*mcd3* [LS2] X CC-610	18 tetrads	14 PD, 4 T, 0 NPD	12 cM
6	*mcd5* [LS6] X CC-610	8 tetrads	3 PD, 4 T, 1 NPD	46 cM
7	LS6 X *mcd4* [LS2]	4 tetrads	0/16	
8	*mcd3* [LS2] X F16 (*mcd1-1)*	20 tetrads	2:2-5 (PD), 1:3-10 (T), 0:4-5 (NPD)	
9	*mcd4* [LS2] X F16 (*mcd1-1)*	8 tetrads	2:2-4 (PD), 1:3-3 (T), 0:4-1 (NPD)	

* Cross reported in Rymarquis *et al.* (2005).
[a]Segregation is listed as the ratio of photosynthetic/nonphotosynthetic progeny for crosses 1-4 and 7. Cross *, 4-5 both photosynthesis and swimming proficiency were measured. Three photosynthetic ratios were seen in crosses 8-9. Each ratio is followed by a dash and the number of tetrads containing that ratio.
[b] PD, parental diatypes; T, tetratypes; NPD, nonparental ditypes where 0, 1, or 2 double mutants are produced respectively.
[c]The distance between each of the *mcd* loci and PF12 in centimorgans (cM) is based on the number recombinants generated.

mcd4 is a recessive mutation

To establish the dominance of *mcd4,* six vegetative diploids were constructed by performing crosses 1-12 (Table 2.2). These diploids carried

either LS2 or WT cpDNA, and were homozygous WT, heterozygous, or homozygous mutant at the *MCD4* locus. As shown in the three left lanes of Figure 2.2a, *MCD4/MCD4* [WT], *MCD4/mcd4* [WT] and *mcd4/mcd4* [WT] accumulated similar levels of *petD* mRNA and cytochrome *f* (PetA) protein. PetA was used as an indicator of the SUIV level because an anti-SUIV antibody was no longer available. In the absence of SUIV, PetA is present at only 10% of the WT level (Kuras and Wollman, 1994). The results in Figure 2.2b thus indicate that the *mcd4* mutation has little if any effect on expression of the WT monocistronic *petD* transcript. On the other hand, both *mcd4/mcd4* [LS2] and *mcd4/mcd4* [WT] accumulated the dicistronic *petA-petD* transcript found in the haploid *mcd4* [LS2] strain (Figure 2.1b), showing that its accumulation is not due to the LS2 mutation, but rather is characteristic of the *mcd4* mutation.

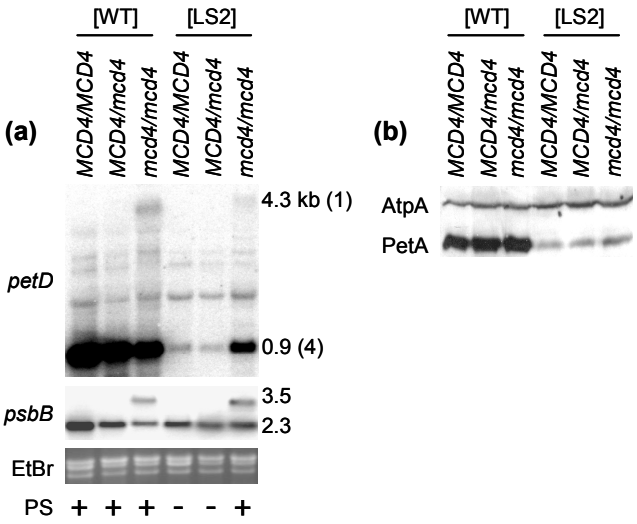


Figure 2.2. *mcd4* is a recessive mutation. (a) An RNA gel blot was probed sequentially with *petD* and *psbB*. The EtBr-stained gel serves as a loading control. (b) Immunoblot analysis was performed with a PetA antibody, as a measure of SUIV accumulation. AtpA was used as a loading control. (1) and (4) indicate that these transcripts correspond to transcripts 1 and 4 in Figure 2.1.

The three right lanes of Figure 2.2a show that *MCD4/MCD4* [LS2] and *MCD4/mcd4* [LS2] diploids were PS-, and that they had *petD* mRNA and PetA protein levels similar to LS2, whereas *mcd4/mcd4* [LS2] was PS+ and resembled *mcd4* [LS2]. Since *MCD4/mcd4* [WT] and *MCD4/mcd4* [LS2] posses the same RNA, protein and photosynthetic phenotypes as *MCD4/MCD4* [WT] and *MCD4/MCD4* [LS2], respectively, *mcd4* is recessive for suppression of LS2. Interestingly, when the RNA filter was probed with *psbB*, a novel transcript was seen in both *mcd4/mcd4* [WT] and *mcd4/mcd4* [LS2]. This transcript was later determined to be a *psbB-psbT* dicistronic transcript with an extended 3' end (Figure 2.4, transcript 1). As well as further verification that *mcd4* is recessive, the results with *psbB* showed that *mcd4* had pleiotropic effects on chloroplast transcripts. Since *mcd4* is recessive for stabilization of *petD* mRNA*,* accumulation of dicistronic *petA-petD* mRNA, and the novel *psbB* transcript, it is likely to harbor a loss of function mutation.

The ability of *mcd3* and *mcd4* to stabilize *petD* LS2 transcripts is dependent on MCD1

Since mutations in *mcd3,* and *mcd4* suppress LS2, and *mcd5* suppresses LS6, it was conceivable that their gene products interact with MCD1, which is predicted to bind specifically to Element I (Erickson *et al.*, 2005). To determine their interaction, *mcd3* and *mcd4* were crossed to *mcd1-1* (Table 2.1, crosses 8 and 9). Four genotypes were produced in each of these crosses since *mcd1-1* and *mcd3* or *mcd4* are unlinked based on their map positions. For example in cross 8, progeny with genotypes *mcd1-1 MCD3* [LS2], *MCD1 MCD3* [LS2], *MCD1 mcd3* [LS2], and *mcd1-1 mcd3* [LS2] would be expected. The genotypes *mcd1-1 MCD3* [LS2] and *MCD1 MCD3* [LS2] will

confer a PS- phenotype due to the presence of the *mcd1-1* mutation or an unsuppressed LS2 mutation, respectively. Since *MCD1 mcd3* [LS2] progeny are genetically equivalent to the *mcd3* [LS2] parent and therefore PS+, the only unknown photosynthetic phenotype was that of the *mcd1-1 mcd3* [LS2] progeny. Tetrads with three distinct photosynthetic ratios, 2 PS+:2 PS-, 1 PS+:3 PS- and 0 PS+:4 PS-, were observed for crosses 8 and 9, corresponding to parental ditypes, tetratypes, and nonparental ditypes where 0, 1 or 2 *mcd1-1 mcd3* [LS2] or *mcd1-1 mcd4* [LS2] progeny were produced, respectively. The existence of three PS+:PS- ratios indicated that *mcd1-1 mcd3* [LS2] and *mcd1-1 mcd4* [LS2] are PS-, because if they were PS+, the parental ditypes, tetratypes and nonparental ditypes would all yield a 2 PS+:2 PS- ratio. Taken together, these results indicate that in the absence of MCD1 (*mcd1-1* background), *mcd3* and *mcd4* are unable to suppress LS2, and thus are not bypass suppressors.

To study the interaction of MCD4 and MCD1 on WT *petD* transcripts, I performed crosses to generate *mcd1-2 mcd4* [WT] and *mcd1-2 mcd2 mcd4* [WT] (Table 2.2, crosses 14 and 15). As mentioned in the Introduction, *mcd1-2* harbors a premature stop codon at position 113 of the *MCD1* gene, and fails to accumulate *petD* mRNA. *mcd2*, which suppresses *mcd1-2* by encoding an amber suppressor tRNA, was subsequently isolated (Murakami *et al.*, 2005). These two strains, *mcd1-2* and *mcd1-2 mcd2*, were useful for subsequent studies of the interaction of MCD4 with MCD1, because *mcd1-2* contains no MCD1, while *mcd1-2 mcd2* contains an intermediate amount, estimated at 10% of the WT level based on the level of *petD* mRNA accumulation in the suppressed strain (Esposito *et al.*, 2001). The MCD1 protein produced in *mcd1-2 mcd2* differs by only 1 amino acid from the WT version, having Ser in

place of Leu at position 113. In order to interpret the crosses to the *mcd1-2*-containing strains, I had to ensure that the amber suppressor *mcd2* did not itself suppress *mcd4* (Table 2.2, cross 13). The four left lanes of Figure 2.3 show this result, where lanes 3 and 4 compare strains differing only by lacking (lane 3) or possessing (lane 4) the *mcd2* mutation. These two strains accumulate equivalent amounts of *petD* transcript, PetA protein and the novel *psbB* band found in *mcd4*, indicating that *mcd2* does not suppress *mcd4*

Table 2.2. Crosses performed to generate *mcd4*-related strains used in this study. The name of the strain generated from each cross is shown as "Genotype chosen," and in certain cases, correspond to strains in Table 2.1.

Cross	Cross [mt+ X mt-][a]	Genotype chosen[b]	Phenotypes chosen[c]
1	CC-3680 X *mcd4* [LS2]	*mcd4 arg7* [WT] mt+	PS+, arginine requiring
2	CC-88 X *mcd4* [LS2]	*mcd4 arg2* [WT] mt-	PS+, arginine requiring
3	*mcd4* [LS2] X CC-3681	*mcd4 arg7* [LS2] mt-	PS+, arginine requiring
4	*mcd4* [LS2] X CC-1930	*mcd4 arg2* [LS2] mt+	PS+, arginine requiring
5	LS2 X CC-1930	*arg2* [LS2] mt+	arginine requiring
6	LS2 X CC-3681	*arg7* [LS2] mt-	arginine requiring
7	*arg2* [LS2] X *arg7* [LS2]	*MCD4/MCD4* [LS2]	arginine autotrophy
8	*arg2* [LS2] X *mcd4 arg7* [LS2]	*MCD4/mcd4* [LS2]	arginine autotrophy
9	*mcd4 arg2* [LS2] X *mcd4 arg7* [LS2]	*mcd4/mcd4* [LS2]	arginine autotrophy
10	CC-3680 X CC-1930	*MCD4/MCD4* [WT]	arginine autotrophy
11	*mcd4 arg7* [WT] X CC-1930	*MCD4/mcd4* [WT]	arginine autotrophy
12	*mcd4 arg7* [WT] X *mcd4 arg2* [WT]	*mcd4/mcd4* [WT]	arginine autotrophy
13	*mcd4* [LS2] X Sup670	*mcd2 mcd4* [LS2]	PS+, *atpA* RNA, dCAPS
14	670 X *mcd4* [LS2]	*mcd1-2 mcd4* [WT]	PS-, *atpA* RNA
15	670R1 X *mcd4* [LS2]	*mcd1-2 mcd2 mcd4* [WT]	PS+, *atpA* RNA, dCAPS, AFLP

[a] *arg2* and *arg7* mutants require arginine in the growth medium.
[b] Diploids (crosses 7-12) contain both the mt+ and mt- loci.
[c] PS, photosynthesis; *atpA* RNA, displays *mcd4* RNA pattern shown in Figure 2.3; dCAPS (*mcd2)* and AFLP (*mcd1-2)* markers were described in Murakami *et al.* (2005).

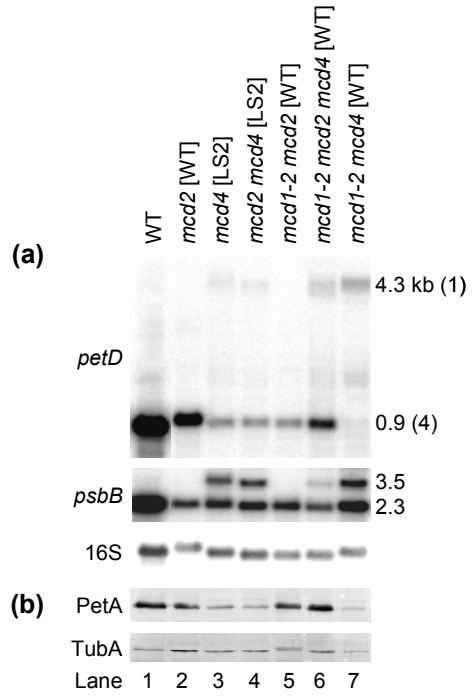


Figure 2.3. MCD4 may interact with MCD1. (a) An RNA gel blot was probed sequentially with *petD, psbB* and 16S rRNA. (b) An immunoblot was incubated with PetA and TubA (tubulin) antibodies to estimate the accumulation of subunit IV.

I then compared *mcd1-2 mcd2* to *mcd1-2 mcd2 mcd4,* as shown in lanes 5-6 of Figure 2.3. The results showed that more *petD* mRNA and slightly more PetA protein accumulated in the latter strain, indicating that the *mcd4* mutation has a positive effect on *petD* accumulation when MCD1 is limiting. In contrast, *mcd4* does not convey a positive effect in the absence of MCD1, because in *mcd1-2 mcd4* (i.e. in a *mcd1* mutant background) no monocistronic *petD* accumulates (lane 7). Taken together, it shows that the MCD1 protein is absolutely required for monocistronic *petD* stability, while the presence of *mcd4* increases monocistronic *petD* stability only when MCD1 is present. This stabilization of *petD* mRNA by *mcd4* is independent of whether the strain

contains a WT or LS2 chloroplast. Finally, I noted that compared to 16S rRNA, the dicistronic transcripts (*petA-petD* and *psbB-psbT*) may accumulate to a higher level in *mcd1-2 mcd4* than in *mcd1-2 mcd2 mcd4,* indicating a possible effect of MCD1 on dicistronic mRNA accumulation. Since the dicistronic messages do not accumulate in *mcd1-2 mcd2* (lane 5) or *mcd1-2* (Esposito *et al.*, 2001), I conclude that their accumulation is dependent on *mcd4,* whereas WT MCD4 is sufficient to prevent their accumulation.

mcd3, mcd4, and *mcd5* have pleiotropic effects

As shown in Figures 2.2 and 2.3, *mcd4* strains contain a novel *psbB* transcript. To determine if *mcd3* and *mcd5* also accumulated this transcript, RNA filter hybridizations were performed as shown in Figure 2.4. The left panel shows that a *psbB* coding region probe also detected the novel transcript 1 in *mcd3* and *mcd5,* although transcript 1 is less abundant in *mcd5* [LS6] than in *mcd3* [LS2] and *mcd4* [LS2]. A 5' extension is possible for this transcript since two 5' ends have been reported for *psbB* (Vaistij et al., 2000), but they only differ by 112 nt, and thus cannot explain the 1.2 kb extension. To determine whether transcript 1 was a 3' extension of the dicistronic *psbB-psbT* (transcript 2), a second probe approximately 400 bp downstream of the *psbT* stop codon was used (right panel). This probe hybridized to transcript 1, but not transcript 2, confirming the 3' extension. The accumulation of the extended *psbB-psbT* species in *mcd3, mcd4,* and *mcd5* defines these mutants as pleiotropic, i.e. they are not *petD*-specific, and also implicates *mcd3, mcd4, and mcd5* in the formation of 3' ends.

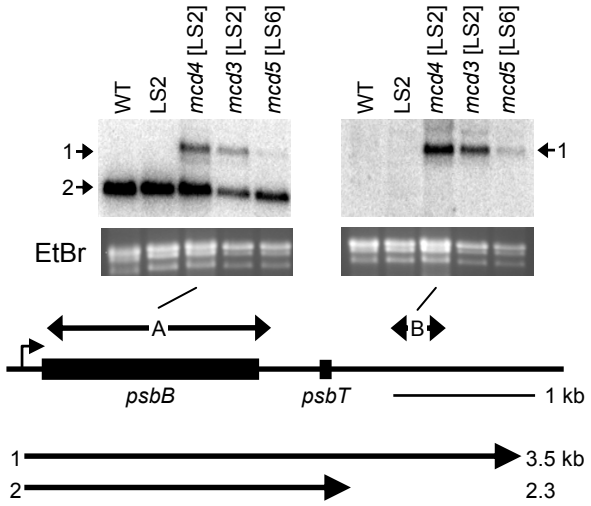


Figure 2.4. *mcd3, mcd4, and mcd5* accumulate *psbB-psbT* transcripts with extended 3' ends. RNA gel blots were probed with the *psbB* coding region (A) or the PstI-SpeI fragment from P-72 (B). The EtBr-stained gels are shown as loading controls. A diagram of the *psbB-psbT* region is shown with the transcription start site marked by a bent arrow. The extents of the major mRNA species and their sizes are shown underneath.

Further RNA filter hybridizations were performed to determine the extent to which *mcd3, mcd4, and mcd5* affected accumulation of other chloroplast RNAs. The *atpA* cluster, containing the *atpA, psbI, cemA* and *atpH* genes, was examined next because this complex transcription unit has been well-characterized (Drapier *et al.*, 1998). When RNA gel blots were probed with the *atpA* coding region (Figure 2.5, probe A), the expected transcripts (4-7) were observed, however three larger species of 6.2-9.0 kb (transcripts 1-3) were observed in *mcd3* and *mcd4* that were not seen in WT, LS2 or *mcd5*. These transcripts probably originate from the *atpA* promoter and extend downstream of *atpH*, since they hybridize to probes B-D. Based on their estimated sizes, transcripts 1 and 2 should extend through *atpF*, or begin upstream of *atpA,* but they were not detected by the *atpF* probe (probe E) or two probes upstream of *atpA* (data not shown). Additional probes could not be

developed to resolve this issue, due to the prevalence of repeated sequences in the intergenic regions.

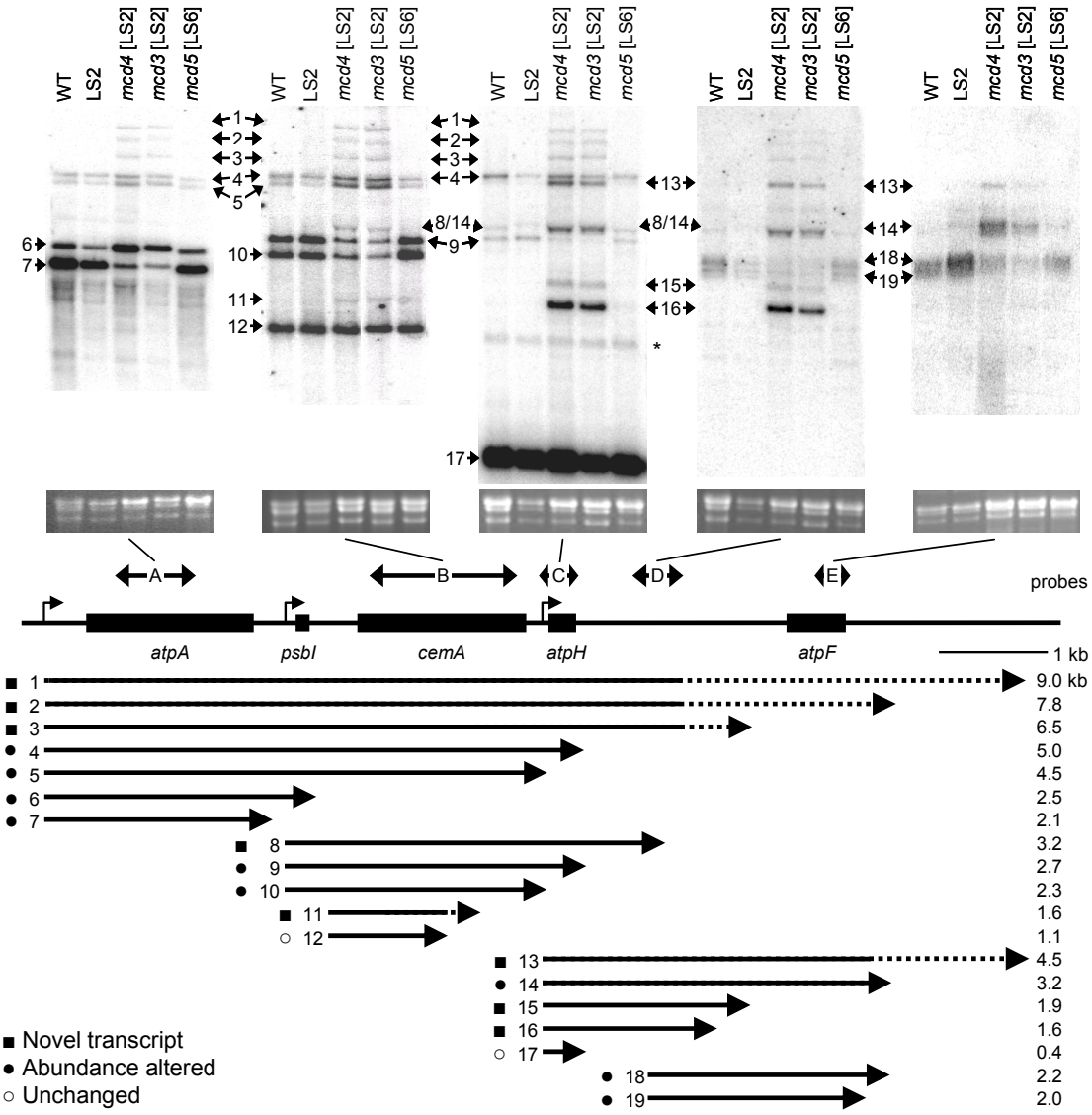
The *atpA* probe also revealed differences in ratio of accumulation for monocistronic *atpA* (transcript 7) and dicistronic *atpA-psbI* (transcript 6). In *mcd3* and *mcd4*, the dicistronic species accumulates to a higher level than in WT, LS2 and *mcd5*, with monocistronic *atpA* being commensurately less abundant. The simplest interpretation is a defect in converting transcript 6 to transcript 7, although an effect on transcription termination at the end of *atpA* cannot be excluded.

The second gene in the *atpA* cluster is *psbI.* It is a small gene (112 nt coding region) whose probe hybridizes to five transcripts (Drapier *et al.*, 1998). Since four of these also encode *cemA,* the *cemA* probe was used to represent both genes. The *cemA* probe (B) revealed two additional novel transcripts in *mcd3* and *mcd4*. Transcript 8 is most likely a tricistronic *psbI-cemA-atpH* mRNA with an extended 3' end that terminates somewhere in probe D. Appearance of transcript 8 correlates with the decrease in RNAs 9 and 10, which are the WT tricistronic *psbI- cemA-atpH* and the dicistronic *psbI-cemA* messages, respectively. Again, this suggests that transcript 8 may be an incompletely processed precursor of transcripts 9 and 10. The second novel transcript is numbered 11. Transcripts 11 and 12 are most likely degradation intermediates, being too small to span the entire *cemA* coding region. Transcript 12 was not detected by the *cemA* 3' probe used in Drapier *et al*. (1998), whereas it is reproducibly detected by probe B in Figure 2.5, which extends approximately 550 bp upstream of the Drapier *et al.* probe. It is therefore likely that the 3' end of transcript 12 is near the 5' end of the Drapier *et al.* probe. The results with the *cemA* probe confirm that *mcd3* and *mcd4* are

involved in the accumulation of transcripts from the *atpA* gene cluster, and extends this finding to those that do not contain the *atpA* coding region.

The fourth gene in the cluster is *atpH*, which is transcribed from its own promoter (transcripts 14 and 17), as well as from the upstream *psbI* (transcript 9) and *atpA* (transcript 4) promoters. In WT cells, these four transcripts (4, 9, 14, 17) accumulate. In *mcd3* and *mcd4*, 11 transcripts hybridize to the *atpH* probe (C). Of these, four emanate from the *atpA* promoter (1-4) and two from the *psbI* promoter (8-9), all of which have been described above. Of the five which are transcribed from the *atpH* promoter, transcripts 13 and 14 extend downstream of *atpF* and hybridize to an *atpH-atpF* intergenic probe as well as *atpF* (probes D and E; transcripts 8 and 14 are indistinguishable in size). Accumulation of these two transcripts correlates with a decrease in the abundance of transcripts 18 and 19, the monocistronic *atpF* transcripts. Transcripts 15 and 16 are monocistronic *atpH* transcripts with extended 3' ends, since they hybridize to *atpH* as well as the *atpH-atpF* intergenic probe, while transcript 17 is the mature *atpH* transcript. Of transcripts 13-17, only transcript 17 is found in WT cells, the others being unique to *mcd3* and *mcd4* and having extended 3' ends. Altogether, 7 out of the 11 transcripts detected by the *atpH* probe have extended 3' ends in *mcd3* and *mcd4*, suggesting their global involvement in 3' end formation in this gene cluster.

Figure 2.5. Transcripts from the *atpA* gene cluster show altered abundance in *mcd3* [LS2] and *mcd4* [LS2]. RNA filters were probed with *atpA* (A) *cemA* (B), *atpH* (C), the *atpH-atpF* intergenic region (D) or *atpF* (E). The EtBr-stained gels are shown as loading controls. A diagram of the *atpA* gene cluster is shown with the transcription start sites marked by bent arrows. The positions of the major mRNA species and their sizes are shown underneath. Solid lines mark regions of transcripts that were confirmed experimentally while dotted lines mark regions predicted by size. Each transcript is marked at left with a filled square if it is a novel transcript in *mcd3* and *mcd4,* a filled circle if the abundance is altered, and an open circle if it is unchanged in *mcd3* and *mcd4.* The asterisk indicates an unspecific transcript detected by the *atpH* probe.


WT
LS2
mcd4 [LS2]
mcd3 [LS2]
mcd5 [LS6]
probes
A
B
C
D
E
atpA
psbI
cemA
atpH
atpF
1 kb
Novel transcript
Abundance altered
Unchanged

mcd3 and *mcd4* affect additional gene clusters

Two clusters containing ribosomal protein genes were examined, and alterations were found for *mcd3* and *mcd4*, but not *mcd5*. One cluster consists of *psbJ, atpI, psaJ,* and *rps12*, and is most likely transcribed from a promoter upstream of *psbJ* (Liu *et al.*, 1989). As shown in Figure 2.6, a complex transcript pattern is evident in WT cells and unlike the *atpA* gene cluster, the nature of these species has not been thoroughly defined. However, 3 transcripts (1-3) clearly showed accumulation differences in *mcd3* and *mcd4* when compared to WT, LS2 and *mcd5*. Each of these is a tetracistronic *psbJ-atpI-psaJ-rps12* species, since they hybridized to probes A, B and C. Transcripts 1 and 2 have increased abundance in *mcd3* and *mcd4* compared to the WT, whereas transcript 3 is reduced in abundance. Processing of transcripts 1 and 2 to yield transcript 3 is predicted to occur from the 5' end rather than the 3' end, because probes 600 or 800 bp downstream of *rps12* do not detect any transcripts (Liu *et al.*, 1989). On the other hand, the weaker hybridization of transcripts 1 and 2 vs. transcript 3 with probe C, suggests that the former species may differ at their 3' ends as well. A probe 5' of the *psbJ* coding region was not attempted to confirm that transcripts 1-3 have extended 5' ends because BLAST analysis revealed that this area is highly repetitive, making finding a unique probe unfeasible. Based on the assumption that transcripts 1 and 2 accumulate at the expense of transcript 3, and that they differ at their 5' ends, these results implicate *mcd3* and *mcd4* in the formation or stabilization of 5' ends of these RNAs, in addition to that of *petD.*

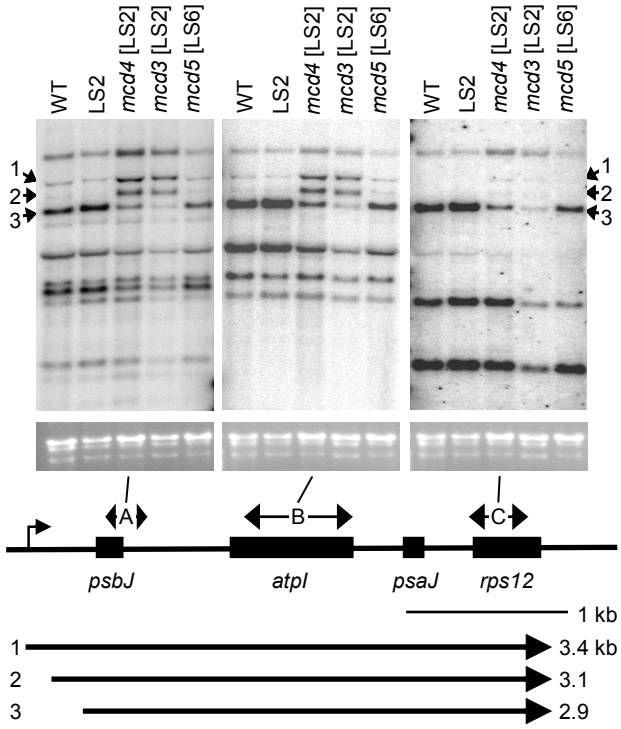


Figure 2.6. *mcd3* [LS2] and *mcd4* [LS2] have altered mRNA accumulation from the *psbJ* gene cluster. RNA filters were probed with *psbJ* (A), *atpI* (B), or *rps12* (C). The EtBr stained gels are shown as loading controls. A diagram of the gene cluster is shown with the putative transcription start site marked by a bent arrow. The position of the mRNA species affected in *mcd3* and *mcd4* and their sizes are shown underneath.

The last cluster examined includes *rpl36*, *rpl23*, *rpl2* and *rps19*, and results are shown in Figure 2.7. Some of the transcripts from this region have been described elsewhere (Cui *et al.*, 2006), but most have not been defined in detail. When RNA gel blots were probed with *rpl36* (probe A), an altered pattern was revealed for *mcd3* and *mcd4*. Transcript 1, the putative tricistronic *rpl36-rpl23-rpl2* species predicted by size to contain an incomplete *rpl2* gene, has increased abundance in *mcd3* [LS2] and *mcd4* [LS2]. Transcript 2 is only found in WT, LS2 and *mcd5* and is a *rpl36-rpl23* dicistronic transcript. In *mcd3* [LS2] and *mcd4* [LS2], transcript 2 seems to be replaced by the smaller

transcript 3, potentially indicating an increase in processing or degradation. Transcript 4 is the tricistronic *rpl23-rpl2-rps19* species found in WT, LS2 and *mcd5*, which either has decreased stability or increased processing in *mcd3* [LS2] and *mcd4* [LS2], perhaps leading to the accumulation of transcript 5, which is also a tricistronic *rpl23-rpl2-rps19* message. The increase in transcript 5 in *mcd3* and *mcd4* also correlates with a decease in transcript 6, a dicistronic *rpl2-rps19* message. While the interpretation of the patterns seen here is difficult, in part due to the overall low accumulation of the mRNAs, it is clear that the *mcd3* and *mcd4* mutations have substantial and pleiotropic effects on the accumulation of transcripts from this region.

DISCUSSION

The *mcd3, mcd4,* and *mcd5* (*mcd3/4/5*) mutants were originally isolated as suppressors of *petD* mRNA instability, increasing its abundance and restoring its translation in the LS2 and LS6 strains. Although the screen targeted *petD*-specific factors, each of these mutants was subsequently revealed to be pleiotropic. Including the examples shown above, transcripts from 32 genes were analyzed using RNA filter hybridizations, to determine the extent of *mcd3/4/5* effects on chloroplast RNA processing. Overall, *mcd3* and *mcd4* affect the accumulation of transcripts from 17 genes residing in 5 gene clusters, whereas *mcd5* only affects the accumulation of *petD, psbB,* and *psbT* transcripts. The effects of the *mcd3/4/5* mutations fell into two categories: quantitative variation between polycistronic transcripts mostly consistent with decreased 3' end processing and in two cases 5' end processing/stability, and qualitative changes in the form of novel transcripts with extended 3' ends.

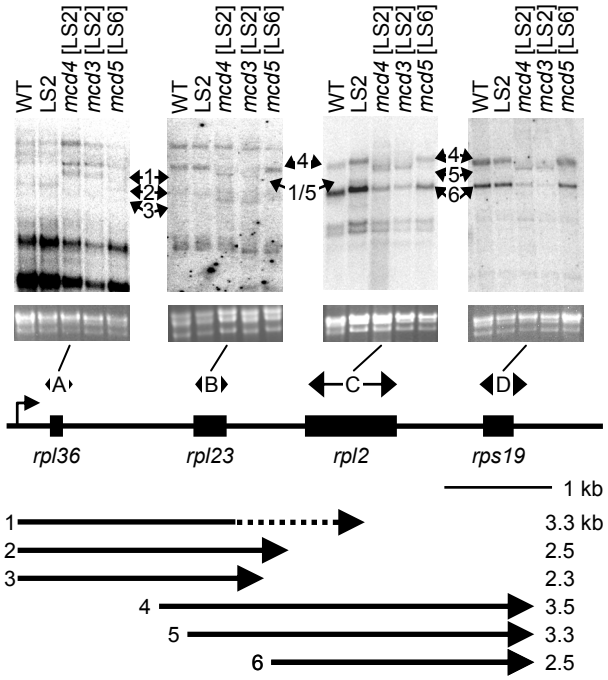


Figure 2.7. Altered RNA patterns from the *rpl36* gene cluster are seen in *mcd3* [LS2] and *mcd4* [LS2]. RNA filters were probed with *rpl36* (A) *rpl23* (B), *rpl2* (C) or *rps19* (D). The EtBr-stained gels are shown as loading controls. A diagram of the *rpl36* gene cluster is shown with the putative transcription start site marked by a bent arrow. The position of the mRNA species affected in *mcd3* and *mcd4* and their sizes are shown underneath. Solid lines mark regions of transcripts that were confirmed by the RNA filter hybridizations, while dotted lines mark regions predicted only by size.

Although I have based the analysis on mutant strains, I hypothesize that in WT cells, MCD3/4/5 interact with 5' ends of transcripts, exemplified by *petD* and *psbJ*, and also with 3' ends of transcripts, exemplified by the *psbB* and *atpA* gene clusters. This combination of 5' end and 3' end interactions must be considered when generating models for MCD3/4/5 action. These models must also take into account that *mcd3/4/5* define different genetic loci, even though their RNA phenotypes are similar, remarkably so for *mcd3* and *mcd4*. In addition, although *mcd5* yields a weaker phenotype for RNA other than *petD* when compared to *mcd3* and *mcd4*, one cannot distinguish whether

MCD5 has a minor or redundant role in these processes, or whether *MCD5* retains significant function in the *mcd5* background.

In order to speculate how *mcd3/4/5* restore *petD* mRNA stability, one must first appreciate why the LS2 and LS6 mutations cause instability. Element I, which is defined by LS2 and LS6, is thought to be the binding site of MCD1, or a putative complex of proteins which includes MCD1 ($MCD1_C$). A complete disruption or weakened binding of $MCD1_C$ would leave transcripts vulnerable to degradation by the 5' to 3' exonuclease activity found in the chloroplast (Drager *et al.*, 1998; Drager *et al.*, 1999), or to endonucleases such as RNase E or RNase J–like proteins (Mackie, 1998; Even et al., 2005). A disruption of $MCD1_C$ binding would also explain the translational inhibition in LS2 and LS6, because MCD1 has been implicated in SUIV translation (Drager *et al.*, 1999).

There are two alternate pathways by which *mcd3/4/5* might stabilize *petD* transcripts: they could increase the binding of $MCD1_C$, or otherwise decrease the rate of *petD* mRNA degradation. The former mechanism would be consistent with the facts that *mcd3* and *mcd4* cannot suppress LS2 in a *mcd1-1* background, and that *mcd4* cannot suppress the *mcd1-2* mutation (*mcd1-2 mcd4* [WT]; Figure 2.3, lane 7), since in both cases there would not be a complex whose binding could be enhanced. It is also consistent with my observation that when MCD1 is limiting (*mcd1-2 mcd2 mcd4* [WT]; Figure 2.3), *mcd4* confers additional stability even in a WT chloroplast background, since a more efficiently-binding complex should protect mRNA more effectively. Since multiprotein complexes similar to $MCD1_c$ have been reported to interact with the 5' ends of several other chloroplast transcripts in *Chlamydomonas* (Hauser et al., 1996; Bruick and Mayfield, 1998; Fargo et al., 1999; Boudreau et al.,

2000; Vaistij et al., 2000), it may be that MCD3/4/5 are factors common to each of these 5' end complexes, leading to their pleiotropic effects. Given that *mcd4* is recessive, it would suggest that MCD4 and perhaps MCD3/5, negatively regulate the activity of these complexes.

Alternatively, *mcd3/4/5* might act to decrease the rate of *petD* degradation. The strains might encode mutant versions of the ribonuclease(s) which are responsible for degradation, or encode factors which activate the ribonuclease(s), or recruit them to the RNA. Any of the three possibilities would be consistent with the recessive nature of *mcd4*. Although *a priori* any exo- or endonuclease might play a role in degrading *petD-LS2/LS6* transcripts, a 5' to 3' exonuclease is a good candidate given that the LS2 and LS6 mutations are suspected to mimic the *mcd1* mutant. Rough mapping of *MCD5* to LG XI, and *MCD3* and *MCD4* to LG II precludes them from encoding any of the four annotated *XRN1/RAT1*-like (5' to 3' exonuclease) proteins found in the available *Chlamydomonas* nuclear genome sequence. This does not rule out the possibility that *mcd3/4/5* encode unknown 5' to 3' exonucleases or accessory factors that regulate *XRN1/RAT1* activity, analogous to *RAI1* (Xue *et al.*, 2000). Also, *XRN1/RAT1* mutants accumulate *ADI1* transcripts with extended 3' ends (Zer and Chanfreau, 2005), as well as iron homeostasis mRNAs *FIT3*, *ARN2* and *ARN3* with 5', 3' and polycistronic differences (Lee *et al.*, 2005), consistent with the *mcd3/4/5* RNA phenotypes. Although this similarity exists, inactivation of *XRN1/RAT1* also leads to an overall increase in several RNAs due to accumulation of RNA that is normally degraded (Lee *et al.*, 2005; Zer and Chanfreau, 2005). In *mcd3/4/5,* overall RNA detected by any probe did not seem to change; it just appeared to be distributed differently among transcripts than in WT. Also, if RNAs such as the *atpA-psbI* dicistronic

species (Figure 2.5) accumulated in *mcd3* and *mcd4* due a reduction in normal degradation, the *atpA* monocistronic mRNA would not decrease, which is however what I observed. Instead, it seems more likely that the RNA phenotypes seen in *mcd3/4/5* are due to decreased transcription termination or processing rather than a reduction in degradation.

Extended 3' ends or ratio changes between mono- and dicistronic transcripts cannot *a priori* be ascribed to altered processing vs. reduced termination. However, decreased transcription termination is unlikely to be the cause of the 3' extended transcripts observed in *mcd3/4/5,* because termination is already highly inefficient in WT cells. Chloroplast 3' IRs have been shown both *in vivo* (Rott *et al.*, 1996) and *in vitro* (Stern and Gruissem, 1987) to terminate less than 50% of the time, and sometimes not at all. As a specific example, *petA* transcription terminates at the mature 3' end only 40% of the time (Rott *et al.*, 1996). The longer transcripts formed the remaining 60% of the time are sufficient to produce the dicistronic *petA-petD* transcript seen in *mcd3* and *mcd4*, as previously demonstrated by the accumulation of both the dicistronic and also monocistronic *petD* mRNA in a strain engineered to lack the *petD* promoter (Sturm *et al.*, 1994). Also, a lack of transcription termination cannot explain the accumulation of the extended 5' ends seen for *psbJ* (Figure 2.6); I conclude that a reduction in processing most likely causes the accumulation of transcripts with 5' or 3' end extensions.

RNA processing involves a variety of enzymes such as the 5' to 3' exonucleases mentioned above, endonucleases (Nickelsen and Link, 1993; Stern and Kindle, 1993; Sakamoto et al., 1994), 3' to 5' exonucleases such as polynucleotide phosphorylase (Walter et al., 2002; Nishimura et al., 2004), and other factors such as *CRP3*, a *Chlamydomonas* nuclear gene which is

involved in 3' end processing and degradation of multiple chloroplast transcripts (Levy *et al.*, 1997; Levy *et al.*, 1999). The common link between *petD* mRNA degradation, 5' processing of *psbJ* polycistronic transcripts and 3' processing of mRNA from the *psbB* and *atpA* gene clusters could well be the involvement of endonucleases. Of the endonucleases found in chloroplasts or prokaryotes, two have activities which merit mention here. One, RNase E, has been implicated in the 5' processing of 16S rRNA (Li *et al.*, 1999), 3' processing and polycistronic processing of tRNA transcripts (Li and Deutscher, 2002), as well as mRNA degradation (Cheng and Deutscher, 2005). Although no RNase E homolog has yet been identified in the *Chlamydomonas* genome, RNase E-like proteins are encoded in plant nuclear DNA (Slomovic *et al.*, 2005). Also, an ORF which could encode RNase J1, a functional homolog of RNase E from *Bacillus subtilus*, is present on linkage group XVIII in *Chlamydomonas*. While *MCD3, MCD4,* and *MCD5* cannot encode RNase J (or PNPase) due to their locations on the genome, they may encode accessory factors analogous to the *E. coli* degradosome components (Carpousis, 2002) or proteins like RraA and RhlB, which regulate RNase E (Lee et al., 2003; Khemici et al., 2005).

To provide a framework for our past results and future dissections of chloroplast mRNA metabolism, I have developed the models shown in Figure 2.8, either for maturation (a) or degradation (b). Given that a defect in an endonuclease activity could explain the suppression of *petD* mRNA instability and the pleiotropic effects seen in *mcd3/4/5,* I parsimoniously depict MCD3/4/5 as being part of an endonucleolytic complex. The same or related interactions feed into either 5' end (a^1) or 3' end maturation (a^2), or into the degradation pathway by aiding in the initial endonuclease cleavage step (b^3), or perhaps by

stimulating 5' to 3' degradation in a wave of endonuclease cleavages (b^4). I postulate one of several roles: catalysis, directing the catalytic proteins to their sites of action, or regulating the kinetics of the reactions (e.g. as a helicase). It would be interesting to determine whether *mcd3/4/5* could re-stabilize other *Chlamydomonas* chloroplast mRNAs where mutations in the 5' UTR have destabilized them (Nickelsen *et al.*, 1999). It is even possible that one or more of the three could be allelic to mutations isolated in a similar screen using *psbD*, although that set of mutations was not examined for pleiotropy other than for *psbA* (Nickelsen, 2000).

METHODS

Strains, culture conditions and crosses

mcd3, mcd4, and mcd5 were generated in suppressor screens using LS2 and LS6 as the progenitors. *mcd3* and *mcd4* were generated in a spontaneous screen of LS2, while *mcd5* was generated in a UV mutagenesis screen of LS6. Both screens were carried out as previously described (Esposito *et al.*, 2001). The strains listed in Table 2.3 were grown in Tris-Acetate-Phosphate (TAP) medium or TARG (TAP + arginine) medium (Harris, 1989) under 23 hr light and 1 hr dark at 25°C. Crosses in Table 2.s 1 and 2 were carried out as described previously (Gorman and Levine, 1965). Progeny were tested for photoautotrophic growth by plating on minimal medium lacking acetate (Harris, 1989) and segregation of *mcd3, mcd4,* or *mcd5* was determined in this manner when progeny inherited their respective LS2 or LS6 chloroplasts. In cases where WT cpDNA was inherited, segregation was determined by *atpA* or *psbB* RNA phenotypes. Segregation of the *arg2* and *arg7* mutations was determined by the ability to grow on TARG, but not TAP

medium. Progeny from crosses to the *pf12* mutants were examined by microscopy as previously described (Rymarquis *et al.*, 2005). Distances in cM was calculated as previously described (Harris, 1989).

Diploids were generated by mating the strains listed in Table 2.2 (crosses 7-12). The resulting mating reaction was plated on TAP medium and grown in the light. The resulting colonies were inspected under 100X magnification. Those that appeared larger than haploid cells were streaked on fresh TAP medium. Colonies underwent three rounds of single-colony purification before being tested by PCR to determine if they contained both mating type loci, which is diagnostic of a diploid strain (Werner and Dieter, 1998).

RNA extraction and filter hybridizations

Cells were grown in TAP medium and RNA was extracted using Tri-reagent (Molecular Resource Center, Cincinnati, OH, USA). For filter hybridizations, 5 µg of total RNA was fractionated in 1.2% agarose, 6% formaldehyde gels, transferred to Genescreen (Perkin Elmer, Boston, MA, USA), and crosslinked by UV irradiation. Double-stranded PCR products generated using the primers listed in Table 2.S1 or the *Pst*I-*Spe*I fragment from P-72 (*Chlamydomonas* stock center) representing the *psbT* 3' region were labeled by random priming (Feinberg and Vogelstein, 1983), and filters were prehybridized, hybridized and washed as described previously (Church and Gilbert, 1984). All hybridizations were analyzed using the Storm system (Amersham Biosciences Piscataway, NJ, USA).

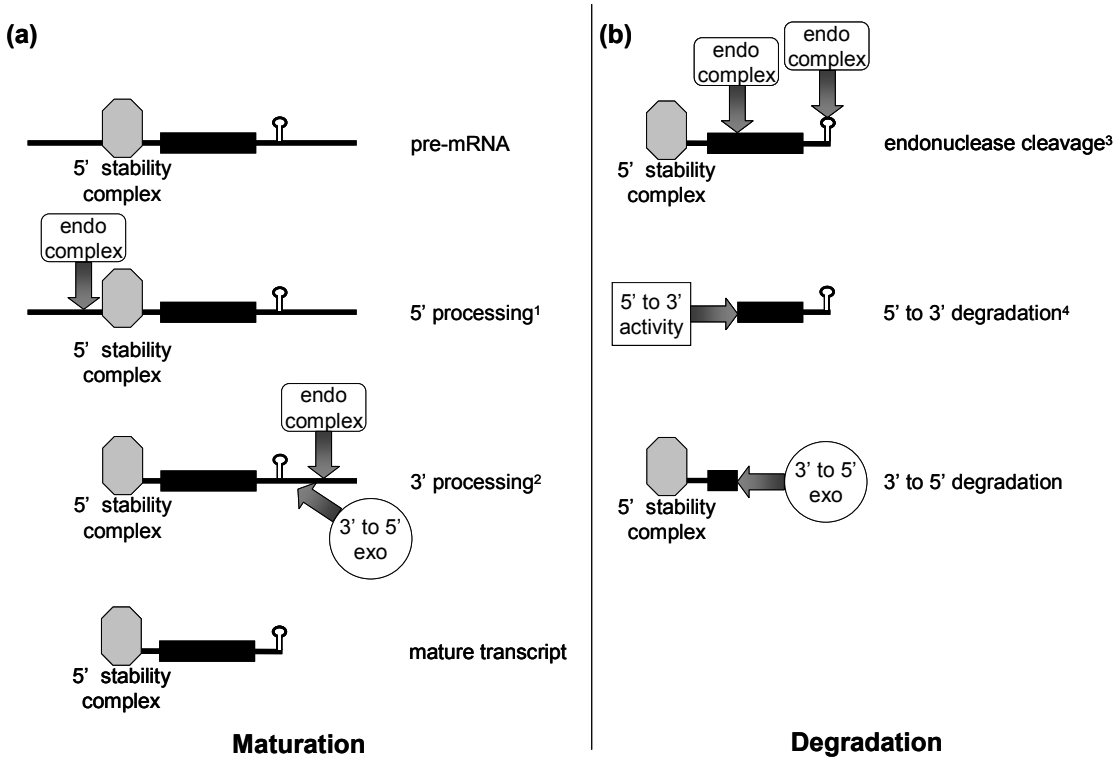


Figure 2.8. Chloroplast maturation and degradation pathways in *Chlamydomonas*. (a) The steps involved in RNA maturation are shown here as unordered, and for a generic RNA. Coding regions are shown as filled boxes, and inverted repeats known to stabilize the 3' ends of transcripts are shown as stem-loop structures. The 5' stability complexes have been described in the text. MCD3, MCD4, and MCD5 are predicted to be part of the endonucleolytic complex (endo complex) involved in both 5' (a^1) and 3' processing (a^2). In 3' processing, the endonuclease cleavage is followed by 3' to 5' exonucleolytic trimming (3' to 5' exo) as occurs for *atpB* (Stern and Kindle, 1993). (b) The same endo complex as in (a) could also participate in degradation initiation (b^3), leading to or stimulating the 5' to 3' pathway involving a 5' to 3' exonuclease or a wave of endonuclease cleavages by the MCD-associated endo complex (b^4). The 3' to 5' pathway is similar to 3' processing, except the secondary structure has been removed, allowing the 3' to 5' exonuclease to completely degrade the transcript.

Table 2.3. Strains used in this study. Nuclear genotypes are followed by chloroplast genotypes in brackets. Photosynthetic phenotype and the source of the strain are indicated.

Strain	Genotype	Phenotype[b]	Source
CC-3681	*arg7*, mt- [WT]	PS+	*Chlamydomonas* Stock Center
CC-124	WT, mt+ [WT]	PS+	*Chlamydomonas* Stock Center
CC-1930	*arg2,* mt- [WT]	PS+	*Chlamydomonas* Stock Center
CC-3680	*arg7*, mt+ [WT]	PS+	*Chlamydomonas* Stock Center
CC-88	*arg2,* mt+ [WT]	PS+	*Chlamydomonas* Stock Center
LS2	WT, mt+ [LS2]	PS-	Higgs *et al*. 1999
LS6	WT, mt+ [LS6]	PS-	Higgs *et al*. 1999
mcd3 [LS2] mt+	*mcd3,* mt+ [LS2]	PS+	This study
mcd3 [LS2] mt-	*mcd3,* mt- [LS2]	PS+	This study
mcd4 [LS2] mt+	*mcd4,* mt+ [LS2]	PS+	This study
mcd4 [LS2] mt-	*mcd4,* mt- [LS2]	PS+	This study
mcd5 [LS6] mt+	*mcd5,* mt+ [LS6]	PS+	This study
mcd5 [LS6] mt-	*mcd5,* mt- [LS6]	PS+	This study
F16	*mcd1-1* mt- [WT]	PS-	Drager *et al.* 1998
670	*mcd1-2,* mt+ [WT]	PS-	Drager *et al*. 1998
Sup670	*mcd2-1,* mt- [WT]	PS+	Esposito *et al.* 2001
670R1	*mcd1-2, mcd2-1,* mt+ [WT]	PS+	Esposito *et al.* 2001
mcd2 mcd4 [LS2]	*mcd2-1, mcd4* [LS2]	PS+	This study
mcd1-2 mcd4 [WT]	*mcd1-2, mcd4* [WT]	PS-	This study
MCD4/MCD4 [LS2]	*MCD4 arg7* [LS2] *MCD4 arg2*	PS-	This study
MCD4/mcd4 [LS2]	*MCD4 arg7* [LS2] *mcd4 arg2*	PS-	This study
mcd4/mcd4 [LS2]	*mcd4 arg7* [LS2] *mcd4 arg2*	PS+	This study
MCD4/MCD4 [WT]	*MCD4 arg7* [WT] *MCD4 arg2*	PS+	This study
MCD4/mcd4 [WT]	*MCD4 arg7* [WT] *mcd4 arg2*	PS+	This study
mcd4/mcd4 [WT]	*mcd4 arg7* [WT] *mcd4 arg2*	PS+	This study
CC-610	*pf12* mt- [WT]	PS+	*Chlamydomonas* Stock Center

[a] *arg2* and *arg7*, requires arginine in growth medium *pf12,* a mutant with paralyzed flagella
[b] Photosynthesis as determined by growth on medium lacking acetate

Protein preparation and immunoblotting

Total protein was isolated and analyzed by immunoblotting as previously described (Drager *et al.*, 1998; Higgs *et al.*, 1998). Blots were reacted with an antibody against SUIV (Chen *et al.*, 1993), PetA (Chen *et al.*, 1995), AtpB (Stern *et al.*, 1991) or TubA (Sigma-Aldrich, St. Louis, MO, USA). Proteins were visualized using ECL (Amersham Biosciences Piscataway, NJ) and exposing on film or ECL Plus (Amersham Biosciences) and detection with the Storm system.

REFERENCES

Anthonisen IL, Salvador ML, Klein U (2001) Specific sequence elements in the 5' untranslated regions of *rbcL* and *atpB* gene mRNA's stabilize transcripts in the chloroplast of *Chlamydomonas reinhardtii*. RNA **7:** 1024-1033

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

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

Bollenbach TJ, Schuster G, Stern DB (2004) Cooperation of endo- and exoribonucleases in chloroplast mRNA turnover. Prog. Nucleic Acid Res. Mol. Biol. **78:** 305-337

Boudreau E, Nickelsen J, Lemaire SD, Ossenbuhl F, Rochaix J-D (2000) The *Nac2* gene of *Chlamydomonas reinhardtii* encodes a chloroplast TPR protein involved in *psbD* mRNA stability, processing and/or translation. EMBO J. **19:** 3366-3376

Bruick RK, Mayfield SP (1998) Processing of the *psbA* 5' untranslated region in *Chlamydomonas reinhardtii* depends upon factors mediating ribosome association. J. Cell. Biol. **143:** 1145-1153

Carpousis AJ (2002) The *Escherichia coli* RNA degradosome: structure, function and relationship in other ribonucleolytic multienzyme complexes. Biochem. Soc. Trans. **30:** 150-155

Chen X, Kindle K, Stern D (1993) Initiation codon mutations in the *Chlamydomonas* chloroplast *petD* gene result in temperature-sensitive photosynthetic growth. EMBO J. **12:** 3627-3635

Chen X, Kindle KL, Stern DB (1995) The initiation codon determines the efficiency but not the site of translation initiation in *Chlamydomonas* chloroplasts. Plant Cell **7:** 1295-1305

Cheng ZF, Deutscher MP (2005) An important role for RNase R in mRNA decay. Mol. Cell **17:** 313-318

Church G, Gilbert W (1984) Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:** 1991-1995

Cui L, Leebens-Mack J, Wang L, Tang J, Rymarquis L, Stern D, dePamphilis C (2006) Adaptive evolution of chloroplast genome structure. BMC Evol. Biol.**:** in press

Drager RG, Girard-Bascou J, Choquet Y, Kindle KL, Stern DB (1998) *In vivo* evidence for 5' to 3' exoribonuclease degradation of an unstable chloroplast mRNA. Plant J. **13:** 85-96

Drager RG, Higgs DC, Kindle KL, Stern DB (1999) 5’ to 3’ exoribonucleolytic activity is a normal component of chloroplast mRNA decay pathways. Plant J. **19:** 521-532

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

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

Erickson B, Stern DB, Higgs DC (2005) Microarray analysis confirms the specificity of a *Chlamydomonas reinhardtii* chloroplast RNA stability mutant. Plant Physiol. **137:** 534-544

Esposito D, Higgs DC, Drager RG, Stern DB, Girard-Bascou J (2001) A nucleus-encoded suppressor defines a new factor which can promote *petD* mRNA stability in the chloroplast of *Chlamydomonas reinhardtii*. Curr. Genet. **39:** 40-48

Even S, Pellegrini O, Zig L, Labas V, Vinh J, Brechemmier-Baey D, Putzer H (2005) Ribonucleases J1 and J2: two novel endoribonucleases in *B.subtilis* with functional homology to *E.coli* RNase E. Nucleic Acids Res. **33:** 2141-2152

Fargo DC, Boynton JE, Gillham NW (1999) Mutations altering the predicted secondary structure of a chloroplast 5' untranslated region affect its physical and biochemical properties as well as its ability to promote translation of reporter mRNAs both in the *Chlamydomonas reinhardtii* chloroplast and in *Escherichia coli*. Mol. Cell. Biol. **19:** 6980-6990

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:** 6-13

Gorman DS, Levine RP (1965) Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardii*. Proc. Natl. Acad. Sci. USA **54:** 1665-1669

Harris EH (1989) The *Chlamydomonas* Sourcebook: A comprehensive guide to biology and laboratory use. Academic Press, San Diego

Hauser CR, Gillham NW, Boynton JE (1996) Translational regulation of chloroplast genes. Proteins binding to the 5'-untranslated regions of chloroplast mRNAs in *Chlamydomonas reinhardtii*. J. Biol. Chem. **271:** 1486-1497

Herrin DL, Nickelsen J (2004) Chloroplast RNA processing and stability. Photosynth. Res. **82:** 301-314

Higgs DC, Kuras R, Kindle KL, Wollman FA, Stern DB (1998) Inversions in the *Chlamydomonas* chloroplast genome suppress a *petD* 5' untranslated region deletion by creating functional chimeric mRNAs. Plant J. **14:** 663-671

Higgs DC, Shapiro RS, Kindle KL, Stern DB (1999) Small *cis*-acting sequences that specify secondary structures in a chloroplast mRNA are essential for RNA stability and translation. Mol. Cell. Biol. **19:** 8479-8491

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

Khemici V, Poljak L, Toesca I, Carpousis AJ (2005) Evidence *in vivo* that the DEAD-box RNA helicase RhlB facilitates the degradation of ribosome-free mRNA by RNase E. Proc. Natl. Acad. Sci. USA **102:** 6913-6918

Kuras R, Wollman F-A (1994) The assembly of cytochrome b_6/f complexes: an approach using genetic transformation of the green alga *Chlamydomonas reinhardtii*. EMBO J. **13:** 1019-1027

Lee A, Henras AK, Chanfreau G (2005) Multiple RNA surveillance pathways limit aberrant expression of iron uptake mRNAs and prevent iron toxicity in *S. cerevisiae*. Mol. Cell **19:** 39-51

Lee K, Zhan X, Gao J, Qiu J, Feng Y, Meganathan R, Cohen SN, Georgiou G (2003) RraA. a protein inhibitor of RNase E activity that globally modulates RNA abundance in *E. coli*. Cell **114:** 623-634

Levy H, Kindle KL, Stern DB (1997) A nuclear mutation that affects the 3' processing of several mRNAs in *Chlamydomonas* chloroplasts. Plant Cell **9:** 825-836

Levy H, Kindle KL, Stern DB (1999) Target and specificity of a nuclear gene product that participates in mRNA 3' end formation in *Chlamydomonas* chloroplasts. J. Biol. Chem. **274:** 35955-35962

Li Z, Deutscher MP (2002) RNase E plays an essential role in the maturation of *Escherichia coli* tRNA precursors. RNA **8:** 97-109

Li Z, Pandit S, Deutscher MP (1999) RNase G (CafA protein) and RNase E are both required for the 5' maturation of 16S ribosomal RNA. EMBO J. **18:** 2878-2885

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

Mackie GA (1998) Ribonuclease E is a 5'-end-dependent endonuclease. Nature **395:** 720-723

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

Murakami S, Kuehnle K, Stern DB (2005) A spontaneous tRNA suppressor of a mutation in the *Chlamydomonas reinhardtii* nuclear *MCD1* gene required for stability of the chloroplast *petD* mRNA. Nucleic Acids Res. **33:** 3372-3380

Nickelsen J (2000) Mutations at three different nuclear loci of *Chlamydomonas* suppress a defect in chloroplast *psbD* mRNA accumulation. Curr. Genet. **37:** 136-142

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

Nickelsen J, Link G (1993) The 54 kDa RNA-binding protein from mustard chloroplasts mediates endonucleolytic transcript 3' end formation *in vitro*. Plant J. **3:** 537-544

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

Nishimura Y, Kikis EA, Zimmer SL, Komine Y, Stern DB (2004) Antisense transcript and RNA processing alterations suppress instability of polyadenylated mRNA in *Chlamydomonas* chloroplasts. Plant Cell **16:** 2849-2869

Rochaix J-D, Kuchka M, Mayfield S, Schirmer-Rahire M, Girard-Bascou J, Bennoun P (1989) Nuclear and chloroplast mutations affect the synthesis or stability of the chloroplast *psbC* gene product in *Chlamydomonas reinhardtii*. EMBO J. **8:** 1013-1022

Rott R, Drager RG, Stern DB, Schuster G (1996) The 3' untranslated regions of chloroplast genes in *Chlamydomonas reinhardtii* do not serve as efficient transcriptional terminators. Mol. Gen. Genet. **252:** 676-683

Rymarquis LA, Handley JM, Thomas M, Stern DB (2005) Beyond complementation. Map-based cloning in *Chlamydomonas reinhardtii*. Plant Physiol. **137:** 557-566

Sakamoto W, Sturm NR, Kindle KL, Stern DB (1994) *petD* mRNA maturation in *Chlamydomonas reinhardtii* chloroplasts: The role of 5' endonucleolytic processing. Mol. Cell. Biol. **14:** 6180-6186

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

Slomovic S, Portnoy V, Liveanu V, Schuster G (2005) RNA polyadenylation in prokaryotes and organelles: Different tails tell different tales. Crit. Rev. Plant Sci.**:** in press

Stern DB, Gruissem W (1987) Control of plastid gene expression: 3' inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. Cell **51:** 1145-1157

Stern DB, Hanson MR, Barkan A (2004) Genetics and genomics of chloroplast biogenesis: maize as a model system. Trends Plant Sci. **9:** 293-301

Stern DB, Kindle KL (1993) 3' end maturation of the *Chlamydomonas reinhardtii* chloroplast *atpB* mRNA is a two-step process. Mol. Cell. Biol. **13:** 2277-2285

Stern DB, Radwanski ER, Kindle KL (1991) A 3' stem/loop structure of the *Chlamydomonas* chloroplast *atpB* gene regulates mRNA accumulation *in vivo*. Plant Cell **3:** 285-297

Sturm N, Kuras R, Buschlen S, Sakamoto W, Kindle KL, Stern DB, Wollman FA (1994) The *petD* gene is transcribed by functionally redundant

promoters in *Chlamydomonas reinhardtii* chloroplasts. Mol. Cell. Biol. **14:** 6171-6179

Suay L, Salvador ML, Abesha E, Klein U (2005) Specific roles of 5' RNA secondary structures in stabilizing transcripts in chloroplasts. Nucleic Acids Res. **33:** 4754-4761

Vaistij FE, Boudreau E, Lemaire SD, Goldschmidt-Clermont M, Rochaix JD (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*. Proc. Natl. Acad. Sci. USA **97:** 14813-14818

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

Walter M, Kilian J, Kudla J (2002) PNPase activity determines the efficiency of mRNA 3'-end processing, the degradation of tRNA and the extent of polyadenylation in chloroplasts. EMBO J. **21:** 6905-6914

Werner R, Dieter M (1998) Mating type determination of *Chlamydomonas reinhardtii* by PCR. Plant Mol. Biol. Rep. **16:** 295-299

Xue Y, Bai X, Lee I, Kallstrom G, Ho J, Brown J, Stevens A, Johnson AW (2000) *Saccharomyces cerevisiae* RAI1 (YGL246c) is homologous to human DOM3Z and encodes a protein that binds the nuclear exoribonuclease Rat1p. Mol. Cell Biol. **20:** 4006-4015

Zer C, Chanfreau G (2005) Regulation and surveillance of normal and 3'-extended forms of the yeast aci-reductone dioxygenase mRNA by RNase III cleavage and exonucleolytic degradation. J. Biol. Chem. **280:** 28997-29003

CHAPTER 3.

THE NUCLEUS-ENCODED FACTOR MCD4 MEDIATES 3' END MATURATION OF *atpB* mRNA IN *CHLAMYDOMONAS* CHLOROPLASTS

ABSTRACT

3' maturation of chloroplast RNA creates defined 3' termini, is coupled to degradation of downstream sequences, and may stimulate translation. In the case of the *Chlamydomonas atpB* transcript, which encodes the F_0F_1-ATPase β-subunit, this maturation involves endonucleolytic cleavage, followed by exonucleolytic trimming of the upstream fragment, and rapid degradation of the downstream fragment. A 300 nucleotide region containing the *atpB* 3' untranslated region inverted repeat (IR) and endonuclease cleavage site (ECS) was previously shown to be sufficient to promote maturation, and was defined as the *atpB* 3' processing determinant (PD). To characterize this element further, the processing efficiencies of 16 mutant variants of the *atpB* PD were examined *in vivo.* These studies revealed that the presence, but not the sequence of the IR is important for processing efficiency, as is the sequence of the ECS and nucleotides 15-39 following it. A nucleus-encoded factor, MCD4, was found to functionally interact with the *atpB* PD, and has been previously implicated in 5' and 3' processing events of many other transcripts. The *mcd4* mutant is partially defective in the endonucleolytic cleavage step, allowing *atpB* pre-mRNA to accumulate whether in its normal location or in one of several ectopic contexts. Taken together, these data suggest that MCD4 most likely facilitates the endonucleolytic cleavage step in *atpB* processing, potentially interacting either with RNA downstream of the IR or with a protein bound there. In doing so, it functions to accelerate 3' end

maturation, and also prevents the accumulation of potentially deleterious antisense RNA.

INTRODUCTION

The majority of chloroplast genes in both higher plants and algae lack strong transcription termination signals (Stern and Gruissem, 1987; Rott et al., 1996), generating transcripts with extended 3' ends or polycistronic transcripts (Sugita and Sugiura, 1996). Maturation of such transcripts requires endonucleolytic cleavage, which is followed by exonucleolytic trimming in some species (Blowers et al., 1993; Nickelsen and Link, 1993; Hayes et al., 1996; Levy et al., 1999). The maturation of *atpB* in *Chlamydomonas reinhardtii* has been particularly well characterized. Termination of *atpB* transcription is at most 50% efficient, although it contains a 3' inverted repeat (IR) structure similar to those that facilitate termination in bacteria (Henkin, 2000). When not terminated, transcription continues to an undefined downstream site, leading to pre-mRNAs with extended 3' ends (Rott *et al.*, 1996). *In vitro* studies have shown that processing of *atpB* pre-mRNAs is initiated by rapid endonucleolytic cleavage 10 nucleotides (nt) downstream of the IR, followed by 3' to 5' exonucleolytic digestion to the base of the IR, forming the mature transcript (Stern and Kindle, 1993). Polyadenylation by polynucleotide phosphorylase or poly(A) polymerase may stimulate this 3' to 5' exonuclease, since transcripts terminating at the endonuclease cleavage site (ECS) have been found to be polyadenylated (Komine *et al.*, 2000). 3' processing may stimulate translation since mature RNA, rather than the pre-mRNA, is preferentially associated with polysomes (Rott *et al.*, 1998).

The initial endonuclease cleavage generates two products. One is the upstream mRNA that will become the mature transcript, and the other is the downstream product which is quickly degraded both *in vitro* and *in vivo*. This degradation likely involves sequential and vectorial endonuclease cleavages, since it cannot be blocked by secondary structure such as IRs or polyguanosine tetraloops (Hicks *et al.*, 2002).

The degradation of the downstream cleavage product has proved a powerful tool for studying the 3' processing efficiency of *atpB* in reporter gene constructs. Two such reporter constructs are DG2 and DAAD (Sakamoto et al., 1993; Rott et al., 1996), which contain the *petD* promoter and 5' UTR region, the *uidA* or *aadA* coding regions, respectively, and the *rbcL 3'* stabilizing element. The *petD* 5' UTR has been engineered to contain a *Bgl*II site at +25 relative to transcription start, which is neutral for transcription, but may inhibit translation to some degree (Drager *et al.*, 1998). Five variants of the *atpB* processing determinant (PD) were inserted into this +25 site, and the accumulation of the reporter gene mRNA examined *in vivo*. Strains containing the complete *atpB* PD (a 300 nt segment containing the IR and ECS) at +25 accumulated no *uidA* or *aadA* mRNA, presumably due to cleavage of the transcripts at the ECS, and subsequent degradation of the downstream *uidA* or *aadA* coding regions. Variants lacking the ECS, or in the opposite orientation, accumulated full-length RNA beginning at the *petD* transcription start and terminating in the *rbcL* moiety, consistent with a lack of endonuclease cleavage at the *atpB* PD. The ECS alone gave an intermediate result. In summary, the IR contributes to efficient processing and ECS is absolutely required, and the ECS and possibly IR are orientation-dependent, suggesting they interact with specific factors.

The focus of this work was to further define the *atpB* processing determinant, and to identify the nucleus-encoded factors that that may interact with it in the 3' maturation pathway. Using the DG2 and DAAD reporter constructs mentioned above, the processing efficiency of 15 new mutant variants of the *atpB* PD were examined. The results indicated that the structure of the IR, the sequence of the ECS, and presence of bases 15-39 after the ECS all contributed to processing efficiency. We further found that *mcd4,* a nuclear mutant implicated in both 5' and 3' end processing, failed to recognize the *atpB* PD, defining a role for it in endonuclease cleavage and 3' end maturation. MCD4, along with the exonuclease polynucleotide phosphorylase (Walter et al., 2002), is the first protein to be associated *in vivo* with a specific facet of mRNA 3' end maturation in chloroplasts.

RESULTS

Defining the *atpB* processing determinant

The *atpB* processing determinant had previously been isolated to a 300 nt region, containing an IR and an ECS, both of which contribute to processing efficiency (Hicks *et al.*, 2002) To further define their contributions to the PD, a series of reporter gene constructs was assembled and introduced into *Chlamydomonas* chloroplasts via biolistic transformation. These constructs contained the *atpB* PD with mutations or deletions in the loop of the IR, the stem of the IR, the ECS, regions upstream and downstream of the IR and ECS, or a combination of these, inserted into the +25 site of the *petD* 5' UTR as part of the DG2 reporter. Although DG2 is a translational fusion and expresses GUS, insertions into the +25 position inhibit translation, thus the reporter can be used to measure transcript abundance, but not translatability.

Following confirmation of the desired transformation events, the processing efficiency of the *atpB* PD was monitored by RNA gel blot hybridizations, using a double-stranded probe from the *uidA* coding region. Because processing conferred by the *atpB* PD leads to degradation of the *uidA* coding region, constructs with lower processing efficiencies were expected to accumulate higher amounts of *uidA* mRNA. As previously reported (Hicks *et al.*, 2002), the double-stranded *uidA* probe detects two transcripts: the sense transcripts originating at the *petD* promoter, for example in DG2 (Figure 3.1c), and in strains containing the IR at the +25, a second transcript of approximately 4.4 kb, corresponding to RNA derived from the opposite strand (Figure 3.1b). This RNA is predicted to be a degradation product from the adjacent and convergently transcribed *atpB* gene. It is most likely derived from a transcript that initiates from the *atpB* promoter that undergoes endonucleolytic cleavage in the *atpB* coding region rather than the in the ECS. It will hereafter be called the antisense transcript, whereas transcripts originating at the *petD* promoter of the *uidA* cassette will be called the sense transcript.

The IR was the first region examined by mutational analysis, and broken down into loop and stem regions. The two variants of the loop contained a four nt mutation to create a *Sac*I site ($IR_{Sac}ECS$), or a 32 nt deletion ($IR_{\Delta L}ECS$), respectively. Neither mutation reduced processing efficiency (Figure 3.1b), since they did not accumulate sense *uidA* transcripts, but like +IRECS-DG (the full-length WT PD), they both accumulated the 4.4 kb antisense transcript. The lack of sense RNA accumulation in the loop mutants suggests that the loop region is not required for efficient processing.

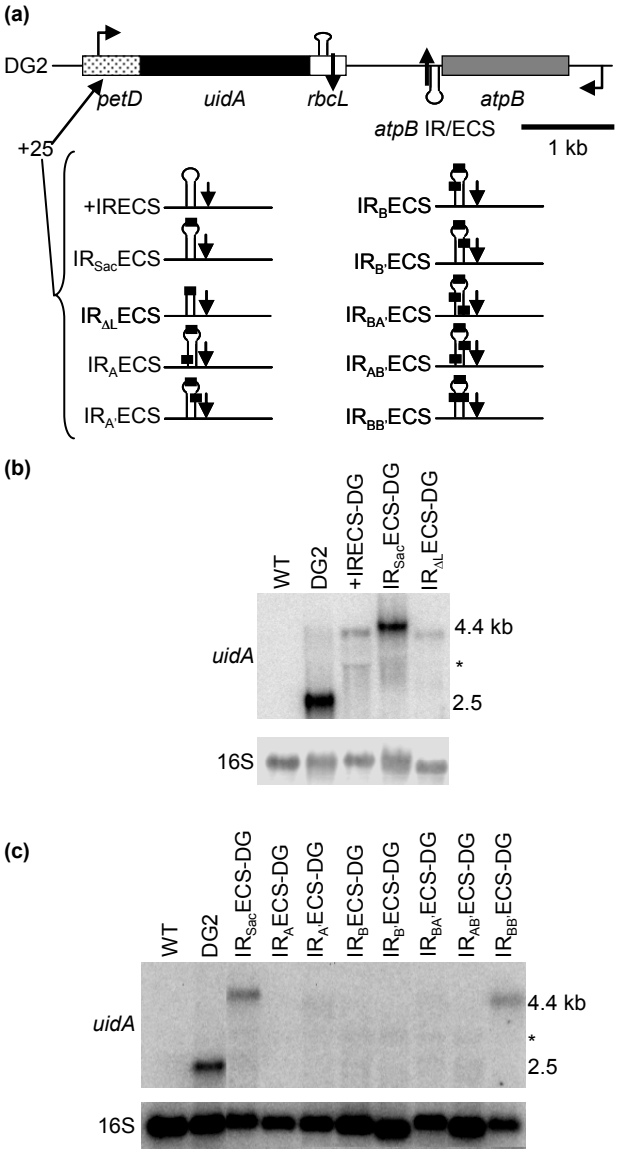


Figure 3.1. The stem and loop play minimal role in processing. (a) Representations of the chimeric *uidA* genes used to study the IR of the *atpB* PD. Bent arrows show the promoters for the *uidA* cassettes and the endogenous *atpB* gene. The *rbcL* and *atpB* IRs are shown as stem-loop structures, and vertical arrows represent the endonuclease cleavage sites. Below the map, names and representations of insertions into the *petD* 5' UTR are shown. (b and c) RNA gel blots were probed sequentially with a double-stranded *uidA* probe, and then 16S rDNA to estimate loading. DG2 is the progenitor strain lacking an insertion, and WT lacks any *uidA* gene. The sizes of the transcripts are indicated, and the asterisk indicates an artifactual hybridization to rRNA.

Next, the stem of the IR was examined by replacing the native sequences with *Spe*I sites (Figure 3.1a). The purpose of these mutants was to test whether the sequence or the structure of the IR is important for RNA processing. $pIR_{Sac}ECS$ was their progenitor to facilitate the creation of the double mutants $IR_{AB'}ECS-DG$, $IR_{BA'}ECs-DG$ and $IR_{BB'}ECS-DG$ (Figure 3.1a). All of the single and double mutants, except $IR_{BB'}ECS-DG$, were expected to destabilize the IR to varying extents, although alternate secondary structures were predicted when the mutated sequences were examined by RNAstructure 4.2 (Mathews *et al.*, 2004). When examined *in vivo*, none of the seven stem mutations led to accumulation of sense *uidA* transcripts, and only $IR_{BB'}ECS-DG$ accumulated the 4.4 kb antisense transcript seen in the $IR_{Sac}ECS-DG$ progenitor to a significant level, although $IR_{A'}ECS-DG$ and $IR_{BA'}ECS-DG$ appeared to accumulate trace amounts. In previous work, the accumulation of antisense transcripts appeared to correlate with the presence of a stable stem-loop in the +25 site (Hicks *et al.*, 2002). This suggests that the +25 insertions in $IR_AECS-DG$, $IR_BECS-DG$, $IR_{B'}ECS-DG$, and $IR_{AB'}ECS-DG$ may not form strong secondary structures *in vivo*. Overall, it appears that the requirement for the IR is not sequence-specific.

Next, a series of mutants containing deletions of the 5' and/or 3' ends of the original 300 nt *atpB* PD was constructed (Figure 3.2a). The ΔIRECS-DG strain contains a deletion of the first 60 nt at the *atpB* PD 5' end, and a 122 nt deletion at the 3' end, although the IR and ECS remain intact. No sense *uidA* accumulates in this strain, indicating that the deleted regions are not required for efficient processing (Figure 3.2b). All of the variants based on ECS-DG, contain longer 5' deletions, having 5' ends in the loop region of the IR. As reported previously, ECS-DG accumulates sense *uidA*, but no antisense

transcripts (Hicks *et al.*, 2002). To detect slight changes in transcripts levels, RNA accumulation was quantified and normalized to 16S accumulation. This analysis revealed that ECS_{39}-DG, which contains the same 3' end deletion as ΔIRECS-DG, accumulates the same level of sense RNA as ECS-DG. This supports the conclusion from ΔIRECS-DG that the region beginning 39 nt downstream of the ECS is dispensable for processing. ECS_{15}-DG, which harbors only 15 bases after the ECS, accumulates more sense *uidA* than ECS-DG, signifying a decrease in processing efficiency. This indicates that nt 15-39 following the ECS contribute to *atpB* PD function.

The sequence requirements of the ECS was examined next. The ECS site was previously identified by RNase protection as a four nt GUCA site that is cleaved three consecutive times by the endonuclease (Stern and Kindle, 1993). The ECS was first mutated from the GUCA to a GGCC by insertion of a *Not*I site because this change was shown to affect the site of cleavage *in vitro* (Rott *et al.*, 1999). Again, RNA levels were quantified and normalized to 16S accumulation. While $IRECS_{Not}$-DG showed no accumulation of sense *uidA,* ECS_{Not}-DG showed a slight increase in accumulation when compared to ECS (Figure 3.2b). This increase held true when comparing ECS_{15}-DG to ECS_{Not15}-DG, which are identical other than the substitution of a *Not*I site for the ECS. When ECS_{Not}-DG and ECS_{Not15}-DG were compared, more sense *uidA* was seen again, emphasizing that the region 15-39 downstream of the ECS plays a role in processing. Taken together, these data suggest that the sequence of the ECS plays a minor role in processing when the *atpB* PD is compromised by 5' or 3' deletions, but appears to be irrelevant when the atpB PD is intact.

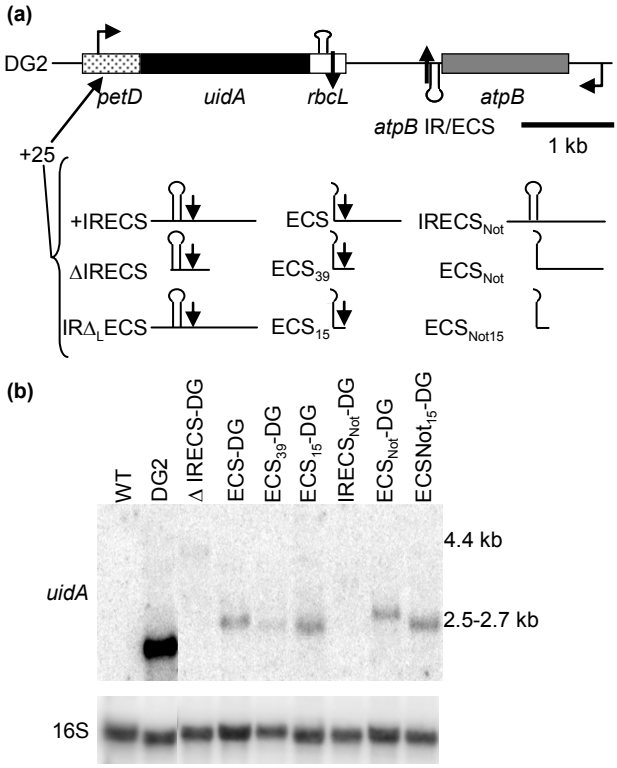


Figure 3.2. The ECS and bases 15-39 after the ECS include PD elements. (a) Chimeric *uidA* genes used in this Figure. All notations remain the same as in Figure 3.1. (b) An RNA gel blot was probed sequentially with a double-stranded *uidA* probe and 16S rRNA. Transcript sizes are indicated.

3' maturation of *atpB* and *rbcL* PDs disrupts *psaA* splicing in *tscA*$_{+RA}$

The analysis above identified certain regions that contribute to processing efficiency, but left open the question of how the PD is recognized *in vivo*. To address this question, I began to develop an experimental system which would implicate known nucleus-encoded factors, and/or detect others which might interact specifically with *the atpB* PD or participate more globally in 3' end maturation. To do this, I created a strain in which the function of a chloroplast gene essential for photosynthesis would be disrupted through processing of an internal *atpB* 3' PD, permitting genetic screens for the restoration of photosynthesis, and therefore impaired *atpB* 3' processing.

The *tscA* transcript is a non-coding RNA that is essential for photosynthesis because it facilitates *trans*-splicing of exons I and II of *psaA* mRNA, which encodes the PSI P700 apoprotein 1a (Goldschmidt-Clermont et al., 1990). An *EcoR*V site, neutral for *tscA* function (Goldschmidt-Clermont *et al.*, 1991), was chosen as the PD insertion site. The antibiotic resistance cassette from the pUC-atpX-AAD (Goldschmidt-Clermont, 1991) plasmid was chosen as the carrier of the PDs, since it already contained the *rbcL* PD, which is processed in a manner similar to *atpB* (Blowers *et al.*, 1993), acting as 3' stability determinant for the encoded *aadA* mRNA. Also, this construct had already been successfully used to disrupt *tscA* (M. Goldschmidt-Clermont, pers. comm.). The *aadA* coding region in this cassette is driven by the *atpA* promoter, conferring spectinomycin and spectinomycin resistance. The plasmid was altered so that it contained the *atpB* PD in tandem with the pre-existing *rbcL* PD, or as a replacement for the *rbcL* PD. These cassettes, along with the progenitor cassette, were inserted in either of two orientations into the *tscA* gene at the *EcoR*V site (Figure 3.3a). These constructs were introduced into the *Chlamydomonas* chloroplast genome, and the resulting strains were named according to the PDs and their orientations. For example, $tscA_{+RA}$ contains the *rbcL* (R) and *atpB* (A) PDs in their functional orientation, whereas in $tscA_{-AR,}$ they are in their antisense (nonfunctional) orientation.

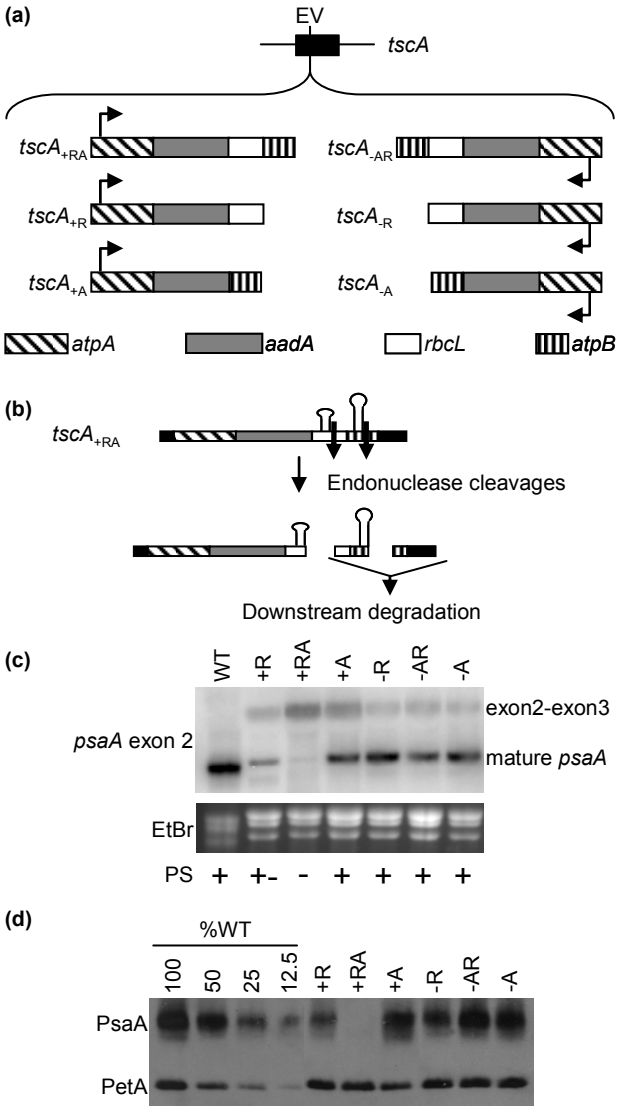


Figure 3.3. Introduction of the *rbcL* and *atpB* PDs into *tscA* disrupts photosynthesis. (a) The filled rectangle represents the *tscA* gene, and the position of the *EcoR*V site (EV) is indicated. Six strains were generated by the integration of constructs containing the *atpA* promoter and 5' UTR (diagonal hatching), *aadA* coding region (grey rectangle), the *rbcL* PD (white rectangle) and/or the *atpB* PD (vertically hatching) into the EV site. (b) The initial transcript in the $tscA_{+RA}$ strain with the *rbcL* and *atpB* IRs and ECS's denoted by stem-loop structures and vertical arrows. The pre-mRNA undergoes processing at the PDs and the 3' segment of *tscA* is degraded. (c) RNA gel blot analysis using a probe for *psaA* exon 2, with an ethidium bromide-strained gel as a loading control. The photosynthetic (PS) phenotypes of the strains are indicated where +, +- and - indicate rapid growth, slow growth or no growth on medium lacking acetate, respectively. (d) PsaA protein accumulation was measured by immunoblot analysis, with PetA as a loading control.

Strains containing the *atpB* and/or *rbcL* PDs in the (+) orientations were predicted (M. Goldschmidt-Clermont, pers. comm.) to be nonphotosynthetic due to PD processing and degradation of the tscA domain required for *psaA* exon 2 binding, downstream of the PDs. To determine if this was the case, the *tscA* insertion strains were monitored for photoautotrophic growth, accumulation of mature *psaA* mRNA, and PsaA protein accumulation (Figures 3.3c and d). Insertion of the *aadA* cassette had a slight impact on *tscA* function, since *tscA*$_{-R}$, *tscA*$_{-AR}$, and *tscA*$_{-A,}$ which contain the *atpB* and/or *rbcL* PD in their non-functional (-) orientations, accumulated a reduced level of mature *psaA.* This reduction in mature *psaA* corresponds to an accumulation of exon2-exon3 transcripts, which is not seen in WT. Despite the reduction in mature *psaA* mRNA*,* these strains were photosynthetic, accumulating 50%-100% of the WT PsaA protein level.

With the PDs are in their functional orientation, varying effects were observed. These ranged from the weakly photosynthetic phenotype of *tscA*$_{+R}$ to the nonphotosynthetic phenotype of tscA$_{+RA}$, and the robust photosynthetic growth of *tscA*$_{+A}$. These phenotypes correlated with the accumulation of mature *psaA* mRNA and PsaA protein, since *tscA*$_{+A}$ (and *tscA*$_{-A}$) accumulated 50% or greater of the WT levels of *psaA* and PsaA, *tscA*$_{+R}$ accumulated approximately 10% of the WT mature *psaA* mRNA level and 25% of PsaA protein, while *tscA*$_{+RA}$ accumulated trace amounts of mature *psaA* mRNA but no detectable PsaA protein. These results show that in this tandem context, the *rbcL* PD and *atpB* PD have additive effects, and that *tscA*$_{+RA}$ could best be used to identify nucleus-encoded factors common to the two processing events.

To obtain spontaneous photosynthetic suppressors, approximately 10^8 cells of the PS- *tscA*$_{+RA}$ strain were plated on medium lacking acetate. This screen yielded three colonies, all of which contained deletions of both the *rbcL* and *atpB* PDs and a portion of the *aadA* gene, leading to PS+ and spectinomycin-sensitive phenotypes (data not shown). In an attempt to circumvent cpDNA deletions, *tscA*$_{+RA}$ cells were again plated on medium lacking acetate, but supplemented with 200 μg/ml spectinomycin to select for *aadA* function. In this screen, however, hundreds of colonies were obtained, which were therefore likely once again to be chloroplast mutants. While we were unable to obtain useful suppressors with this approach, it does illustrate the plasticity of cpDNA when selection is applied, as we (Nishimura et al., 2004) and others have observed previously.

MCD4 facilitates 3' maturation of *rbcL* and *atpB*

To complement the forward genetic approach outlined above, a candidate gene approach was attempted. The *tscA*$_{+RA}$ mt+ strain was crossed to *mcd4* mt-, which contains a recessive mutation in a nuclear gene implicated in both 5' and 3' processing of chloroplast mRNAs (Rymarquis *et al.*, 2006). It is important to note that the *mcd4* mutation itself does not confer any PS growth phenotype except in certain *petD* mutant contexts. Because progeny inherit the chloroplast of the mt+ parent in *Chlamydomonas reinhardtii* crosses, all progeny inherited the *tscA*$_{+RA}$ chloroplast, while the *mcd4* mutation was inherited in a Mendelian fashion. Ten random progeny from this cross were assayed for photosynthetic growth and mature *psaA* mRNA abundance, as indirect measures of processing efficiency of the *rbcL* and *atpB* PDs (Figure 3.4). RNA gel blots were also probed with *atpA* to show the segregation of the

mcd4 allele, since *mcd4* displays an altered ratio of dicistronic *atpA-psbI* to monocistronic *atpA* transcripts, as compared to WT cells (Rymarquis *et al.*, 2006). We found that PS+ progeny harbored *psaA* and *atpA* mRNAs patterns resembling the *mcd4* parent, whereas those of the PS- progeny resembled *tscA*$_{+RA}$. This suggests that the MCD4 is involved in the processing of both the *rbcL* and *atpB* PDs, because in the *mcd4* background, processing efficiency is reduced, allowing accumulation of functional *tscA* RNA and restored splicing of *psaA* mRNA.

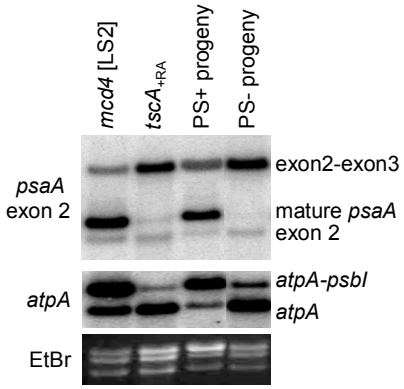


Figure 3.4. *mcd4* suppresses the *psaA* mRNA splicing defect in *tscA*$_{+RA}$. An RNA gel blot for the crossing parents and representative progeny was sequentially probed with *psaA* exon 2 and *atpA.* The transcripts are defined at right, and the ethidium bromide-stained gel is used as a measure of loading.

Antisense DG and DA transcripts are stabilized in *mcd4* background

To verify the interaction of *mcd4* with the *atpB* PD, and to map the region with which MCD4 might interact, *mcd4* mt- was crossed to the four *uidA* reporter constructs shown in Figure 3.5 (a). It was expected that the *mcd4* mutation would increase the accumulation of sense RNA in strains that retained the *atpB* PD elements necessary for its recruitment. Therefore, any *mcd4* DG strain that resembled the corresponding MCD4 DG strain would

have a mutation or deletion in the RNA that binds *mcd4* either directly or through another protein.

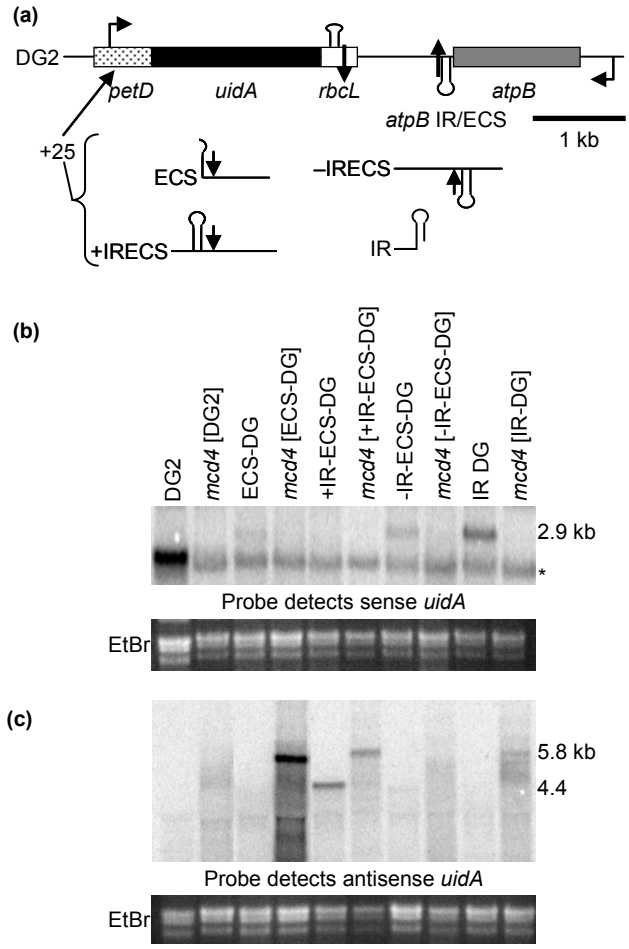


Figure 3.5. The *mcd4* mutation stabilizes *uidA* antisense transcripts (a) The strains used in this figure, with symbols as in Figure 3.1. (b) An RNA gel blot hybridized with a probe to detect *uidA* sense transcripts. Ethidium bromide staining was used to estimate loading. The asterisk indicates an artifactual hybridization to rRNA. (c) A probe to detect antisense *uidA* transcripts was hybridized to an RNA gel blot as in panel (b).

In the following experiments, the chloroplast genotype is given in brackets for strains carrying *mcd4,* and the name of the chloroplast genotype alone is used for strains containing the WT *MCD4* allele, as for the previous figures. For example, *mcd4* [DG2] contains the DG2 reporter gene in the

chloroplast, and the nuclear *mcd4* mutation. To my surprise, however, no *uidA* sense RNA was observed in *mcd4* [DG2], although it is abundant in the DG2 parent (Figure 3.5b). This phenomenon occurred in every case where strains contained both *mcd4* and the *uidA* coding region, suggesting that the presence of the *mcd4* mutation destabilizes the *uidA* transcript. On the other hand, all *mcd4* strains accumulated antisense *uidA* RNA, even DG2, ECS-DG, and IR-DG, which do not accumulate it in an *MCD4* background. In *mcd4* [+IRECS-DG] and *mcd4* [-IRECS-DG], the antisense transcript is 1.4 kb larger than it is in *MCD4* +IRECS-DG and -IRECS-DG, indicating that it either extends upstream of the *uidA* cassette, or includes more of the *atpB* gene. Taken together, *mcd4* appears to destabilize sense *uidA* transcripts, while stabilizing the antisense transcripts.

Because the degradation of sense *uidA* RNA seen in the DG2 derivatives is uninformative, the *atpB* PD variants shown in Figure 3.6 (a) were inserted into the DAAD plasmid. This plasmid is identical to DG2 except that the *aadA* coding region replaces the *uidA* coding region (Rott *et al.*, 1996). The first five lanes in Figure 3.6 (b) shows that the change of coding region does not affect the processing at the *atpB* PD, as the DA strains resemble their DG counterparts. After crossing the new strains to *mcd4,* a reduction of the sense *aadA* was observed in *mcd4* [DAAD], *mcd4* [ECS-DA] and *mcd4* [ECS_{Not}-DA]. The only strain that did not show degradation of the sense RNA was *mcd4* [ECS_{Not15}], which accumulated more sense *aadA* RNA than $ECS_{Not15.}$ This increase may be due to the reduction of endonucleolytic cleavage in the *mcd4* background, but since the cause of the degradation in *mcd4* [DAAD], *mcd4* [ECS-DA] and *mcd4* [ECS_{Not}-DA] is unknown, this

conclusion is tentative. Overall, the *mcd4* mutation reduced accumulation of sense RNA whether the coding region was *aadA* or *uidA.*

Blots were next hybridized with a probe to detect antisense transcripts (Figure 3.6c). All *mcd4*-containing strains except *mcd4* [ECS_{Not15}] accumulated antisense *aadA* transcripts, while only +IR-ECS-DA accumulated antisense RNA in the *MCD4* background. As was seen for antisense RNA in the DG strains, a 1.4 kb size increase was seen when comparing the +IR-ECS-DA antisense transcript (3.4 kb) to the smallest *mcd4* [+IR-ECS-DA] antisense transcript (4.8 kb). In addition, two other antisense transcripts of 5.7-6.0 kb and 7.2-7.5 kb were observed in *mcd4* [DAAD], *mcd4* [+IR-ECS-DA], and *mcd4* [ECS-DA], and a low abundance transcript of 3.1 kb was detected in *mcd4* [DAAD] and *mcd4* [ECS-DA]. The *mcd4* [ECS-DA] strain accumulated the most antisense RNA, similar to *mcd4* [ECS-DG], although *mcd4* [ECS-DG] did not accumulate the 5.7 and 7.2 kb antisense transcripts. These 4.7-7.5 kb antisense transcripts either extends upstream of the *uidA* cassette, or includes more of the *atpB* gene, and was examined in a subsequent experiment (Figure 3.6). In summary, the *mcd4* mutation enhances accumulation of *aadA* antisense transcripts, as it does for analogous ones containing *uidA*.

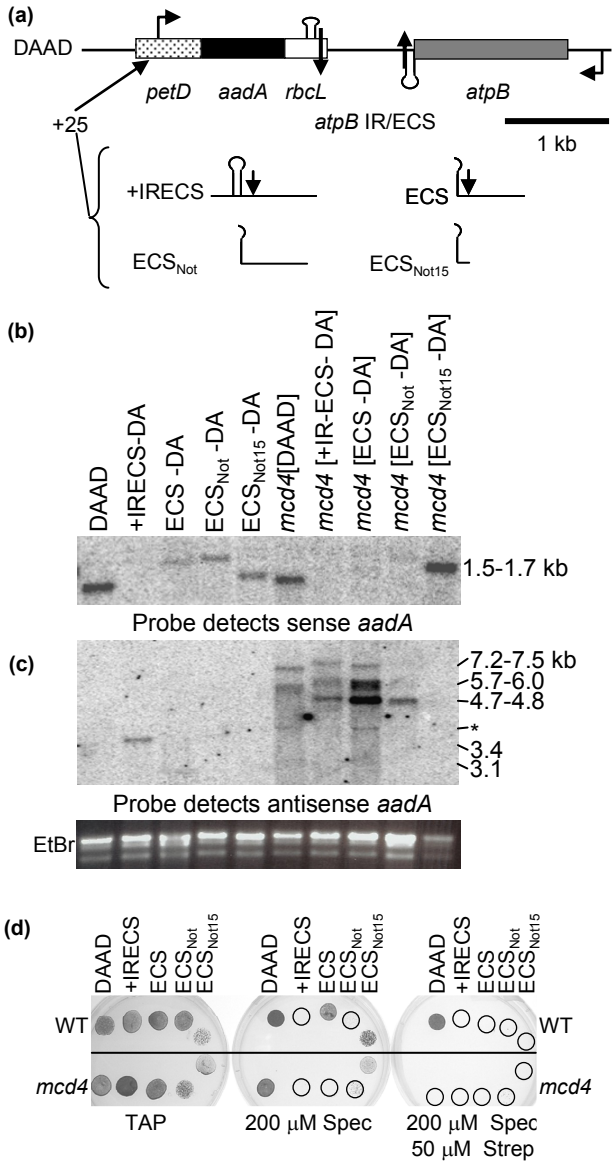


Figure 3.6. *mcd4* has opposite effects on *aadA* sense and antisense transcripts. (a) The strains used in this figure, with symbols as in Figure 3.1. (b) An RNA gel blot was hybridized with a probe to detect sense *aadA* transcripts, with ethidium bromide staining used to estimate loading. DA indicates that the strains contain the *atpB* PD within the *aadA* reporter gene context. (c) A probe to detect antisense *uidA* transcripts was used. (d) Antibiotic resistance of strains containing a WT nucleus or the *mcd4* mutation was tested by growth on TAP, TAP + 200 μg/ml spectinomycin and TAP + 200 μg/ml spectinomycin and 50 μg/ml streptomycin.

To the determine if the loss of sense transcript in the *mcd4* corresponded to a loss of antibiotic resistance conferred by *aadA* expression, strains were replica plated on TAP medium, TAP + 200 μg/ml spectinomycin and TAP + 200 μg/ml spectinomycin + 50 μg/ml streptomycin (Figure 3.6d). As expected, all strains grew on TAP, a nonselective rich medium. DAAD grew on both spectinomycin and streptomycin-containing media, while *mcd4* [DAAD] grew in the presence of spectinomycin, but not under the stronger selection where both antibiotics were used. Both +IRECS-DA strains failed to grow on antibiotic-containing media, consistent with the fact that no sense RNA was observed for these strains. ECS-DA was less resistant than DAAD, only growing on TAP-spectinomycin, while *mcd4* [ECS-DA] only grew on TAP. Although it accumulated sense RNA, ECS_{Not}-DA was not resistant to antibiotics, presumably because its *aadA* transcripts cannot be translated. On the other hand, ECS_{Not15}-DA grew on TAP-spectinomycin in both WT and *mcd4* backgrounds, but *mcd4* [ECS_{Not15}-DA] did not grow as robustly, even though it accumulates more sense RNA. This difference may also reflect a reduction in translation. Overall, the *mcd4* mutation negatively impacts the antibiotic resistance of all DA strains, either by reducing sense RNA levels and/or by interfering with translation.

Because *mcd4* alters both the size and number of antisense transcripts in the DA strains, an effort was made to determine their origin. Double-stranded probes corresponding to sequences upstream of the *aadA* cassette transcription start site (probe A, Figure 3.7a), the *aadA* coding region (B), the *atpB* coding region (C), or the region upstream of the *atpB* transcription start site (D) were hybridized to gel blots containing *mcd4* [ECS-DA] RNA, chosen because this strain accumulated the most antisense RNA. Four antisense

transcripts (1-4) and one sense transcript (7) were detected with a double-stranded *aadA* probe (probe B). When these were compared to results with probe A, only transcripts 1 and 2 were detected by both. The other transcripts detected by probe A may arise from cryptic promoter elements in the adjacent chloroplast genome inverted repeat, or may correspond to unlinked area(s) of the genome, since this region is highly repetitive. When blots were probed with the *atpB* coding region (probe C), transcripts 1-4, as well as two other transcripts, were observed. Transcript 6 represents the mature *atpB* RNA, while transcript 5 is probably an *atpB* RNA that terminates between *atpB* and the *aadA* cassette. No transcripts were detected with the probe upstream of *atpB* (probe D). Based on their sizes and hybridization patterns, we conclude that the antisense RNAs seen in *mcd4* [ECS-DA] most likely initiate at the *atpB* promoter, and are stabilized by both the *mcd4* mutation and the insertion of the *aadA* cassette into the genome. Their accumulation further depends on the lack of processing at the *atpB* PD as was seen for $tscA_{+RA}$, and defective endonucleolytic cleavage which is normally responsible for forming the antisense transcript seen in +IR-ECS.

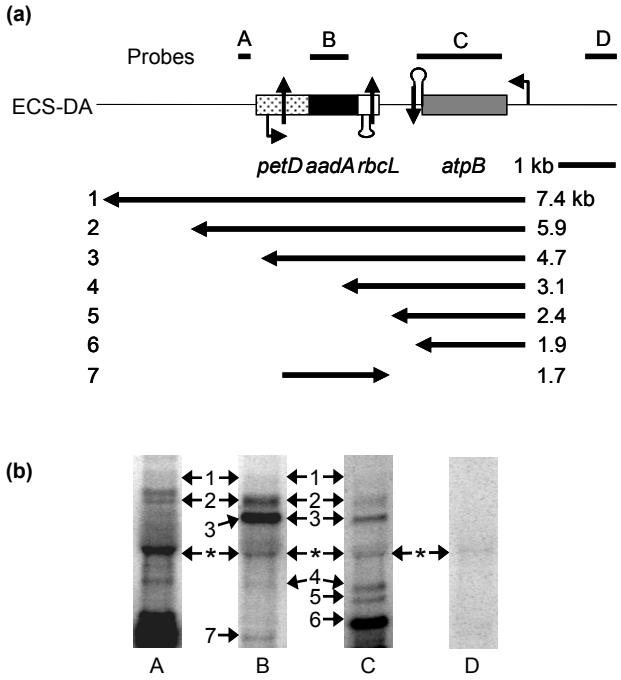


Figure 3.7. *mcd4* accumulates unprocessed *atpB* transcripts in DA reporter gene contexts. (a) A map of the genomic configuration of ECS-DA, with symbols as in Figure 3.1. The probes used to map the antisense transcripts are shown above the map and the deduced transcripts which accumulate in *mcd4* [ECS-DA] are shown below. (b) RNA gel blots were hybridized with probes A-D. The asterisk indicates an artifactual hybridization.

DISCUSSION

The regions of the *atpB* PD involved in mediating its function in 3' processing were identified using a reporter gene approach. Although previous data showed that a stem-loop is required for efficient *atpB* 3' processing (Hicks *et al.*, 2002), neither deletions of the loop nor mutations of the sequence of the stem appeared to affect processing efficiency. Analysis of the mutated IRs with RNAstructure 4.2 (Mathews *et al.*, 2004) showed that secondary structures could form. One possibility is that some of the proteins involved in 3' maturation recognize and bind to IR structures and not specific sequences.

The ECS and the 15-39 bases downstream were both found to contribute to processing efficiency, although the sequence of the ECS only proved important when the *atpB* PD lacked a stem-loop. This suggests that although the ECS contributes 3' processing, the presence of the IR can suppress mutations in the ECS. Together with the IR data, this suggests that the *atpB* 3' stem-loop, the ECS, and the bases 15-39 downstream of it contribute to the binding site(s) of one or more proteins which stimulate or catalyze efficient processing.

Insertion of *atpB* and *rbcL* PDs into *tscA* served to verify that the *rbcL* PD is processed in a manner similar to the *atpB* PD, to demonstrate that the *atpB* PD could act autonomously in a location far from the endogenous *atpB* gene, and to generate a tool to test or discover genes potentially involved in 3' processing. The fact that $tscA_{+R}$ did not accumulate mature *psaA* mRNA suggests that processing at the *rbcL* PD disrupted *tscA* RNA through endonucleolytic cleavage and downstream degradation, analogous to processing of the *atpB* PD. Similarly, when the *rbcL* PD is placed in tandem upstream of a second PD in a reporter gene context, the second PD is not present in the mature mRNA (Blowers et al., 1993).

I had expected that insertion of the *atpB* PD alone would also disrupt photosynthesis by stimulating degradation of the 3' moiety of *tscA* RNA. Since this was not the case, it might be that secondary structures in the chimeric context might cause incorrect folding of the *atpB* IR and ECS, thereby inhibiting its processing. A second possibility is that *tscA* RNA binds *psaA* mRNA exons I and II, and the splicing complex assembles before the *atpB* PD can be processed. This would sterically exclude the proteins required for *atpB* processing. If this scenario were correct, it would argue that the *rbcL* PD is

processed more rapidly in this context than the *atpB* PD. In any event, the *atpB* and *rbcL* PDs had an additive effect in the *tscA*$_{+RA}$ construct, effectively disrupting photosynthesis. Although screens for photosynthetic revertants yielded only deletions of the PDs, the strain could be adapted by simultaneously inserting the PDs into a second gene required for photosynthesis. This should prevent selection of strains with chloroplast deletions in a single PD, and allow selection of mutations in nuclear genes encoding processing enzymes, or in genes encoding proteins involved in their recruitment or activation.

The *tscA*$_{+RA}$ strain did provide a useful tool for determining the role of *mcd4* in 3' processing. *mcd4* was previously implicated in processing of 5' and 3' ends of transcripts from 17 genes, and was hypothesized to be involved in endonucleolytic cleavage (Rymarquis *et al.*, 2006). This is consistent with the fact that *mcd4* restores photosynthesis and *psaA* mRNA splicing to *tscA*$_{+RA}$, either by inhibiting endonucleolytic cleavage, inhibiting degradation downstream of the *rbcL* and *atpB* PDs, or both. Although the *mcd4* mutation appears to suppress processing of the *atpB* and *rbcL* PDs in the context of *tscA*$_{+RA}$, no change was seen in transcripts from the endogenous *atpB* and *rbcL* genes (data not shown). One might expect to see transcripts with extended 3' ends, consistent with those seen for *psbB* and *atpH* in *mcd4* (Rymarquis *et al.*, 2006). The lack of them may be ascribed to a redundant mechanism that bypasses the endonucleolytic cleavage step, coupled with a lack of alternative sites for 3' end formation immediately downstream. Such a mechanism has been observed in the strain *atpB*Δ26pG, where the *atpB* PD was replaced with a polyguanosine tract (Drager *et al.*, 1996). Although this strain does not contain the sequences required to promote endonucleolytic

cleavage, mature transcripts containing polyG at their 3' ends accumulated. When corresponding pre-mRNAs were synthesized and incubated *in vitro* in a chloroplast protein extract, the results were consistent with the mature transcripts being generated by 3' to 5' exonucleolytic activity. The same 3' to 5' pathway would be blocked in transcripts from the *tscA* region due to its inherent secondary structure, unmasking deficiency of the endonuclease cleavage step.

Because *mcd4* is involved in the 3' processing of *atpB,* the mutation was crossed into strains possessing several DG and DA reporter genes in an attempt to define the RNA elements with which it interacts. Instead of the expected increase in sense *uidA* or *aadA* RNA due to missing *atpB* PD *cis* elements, *mcd4* strains accumulated lower levels of sense transcripts than those found in the corresponding WT (*MCD4*) strains. Because *mcd4* has been shown to affect the stability and processing of transcripts, it is very likely that decreased stability was the major factor in the lower accumulation. Such increased rates of degradation or processing are not unprecedented in *mcd4,* because transcripts from the *rpl36* operon display transcript patterns consistent with increased processing, degradation or both in the *mcd4* mutant when compared to WT (Rymarquis *et al.*, 2006).

Examining the effects of *mcd4* on the transcripts containing *uidA* and *aadA* antisense sequences (antisense transcripts) proved more enlightening. For both reporter genes, an increase in size and/or number of antisense transcripts was observed in the *mcd4* background. The antisense transcripts, such as those in +IR-ECS, were previously shown to emanate from the adjacent *atpB* gene, and were most likely degradation products having 5' ends in the *atpB* coding region (Hicks *et al.*, 2002). Our data are consistent with the

interpretation that all of the antisense transcripts in *mcd4* mutant backgrounds begin at the *atpB* transcription initiation site, since the 4.7 kb transcript hybridizes to the *atpB* coding region probe, but not to probes upstream of *atpB* or the ECS-DA cassette (Figure 3.7b). Also, The 1.4 kb increase in size between this transcript in WT and *mcd4* backgrounds is consistent with inclusion of the *atpB* coding region and 5' UTR.

In order for such transcripts to accumulate, the *mcd4* mutation must not only cause a defect in cleavage at the *atpB* PD, but also in the cleavage that forms the 5' end of the antisense transcript in +IRECS-DG. This implicates MCD4 in endonucleolytic cleavages involved in degradation, as well as in processing. Such an effect on degradation is consistent with that fact that the *mcd4* mutation allows fivefold increased accumulation of the otherwise unstable *petD* transcript in the LS2 strain, which carries a 4 nt mutation in the 5' UTR (Higgs et al., 1999). Taken together, the *mcd4* mutation confers two seemingly opposite phenotypes: it stabilizes some transcripts that normally would not accumulate, and is involved in degradation of others derived from chimeric reporter genes.

Figure 3.7 also shows *atpB*-hybridizing transcripts of 3.1 and 2.4 kb, which are most likely *atpB* transcripts with extended 3' ends. The 3.1 kb species (transcript 4) has a size consistent with termination in the *aadA* coding region just upstream of where the *aadA* probe anneals, while the 2.4 kb species (transcript 5) appears to terminate in the intergenic region between *atpB* and the ECS-DA cassette. The 2.4 kb transcript might be expected to independently of the DA transgene, since the intergenic spacer is upstream of the DA cassette, and thus be detectable in WT chloroplasts with an *mcd4* nuclear background. This is not the case, however, leading us to speculate

that this transcript is stabilized by the complementary transcripts initiated at the *petD* promoter. Stabilization of *atpB* transcripts by antisense RNA has been observed in two strains, the mutants *spa19/23*, and ∆26ADaB (Nishimura et al., 2004). *spa19/23* and ∆26ADaB produce *atpB* transcripts that are unstable due to an engineered poly(A) tail or lack of the 3' IR, respectively. In *spa19/23*, antisense *atpB* RNA is generated by a chloroplast genome rearrangement, while ∆26ADaB contains a transgene that transcribes the complement to 121 nt of the *atpB* coding region. A similar pairing of sense and antisense transcripts may inhibit degradation of extended *atpB* mRNAs by 3' to 5' exonucleases in *mcd4* [ECS-DA].

Mutational analysis has narrowed the *atpB* RNA 3' processing determinant to a 150 nt region containing the IR, the ECS, and up to 39 nt downstream of the ECS. This region is likely involved in recruiting nucleus-encoded factors such as MCD4 to aid in transcript maturation. Transcripts that escape processing seem an endonucleolytic cleavage in the coding region, that in the *Chlamydomonas* situation would form the antisense (or extended *atpB*) transcripts observed in +IRECS-DG and +IRECS-DA. The *mcd4* mutation causes a failure to cleave the *atpB* PD in both the native *atpB* mRNA and $tscA_{+RA}$, leading to accumulation of transcripts with extended 3' ends, and lacks certain degradation activities, as seen by the size sift of the antisense transcripts in +IRECS-DA and +IRECS-DG carrying *mcd4*. This indicates that MCD4 is involved in this processing and degradation, although its exact role remains to be determined. Cloning of *MCD4* and using $tscA_{+RA}$ to identify other mutants will unravel the mystery of 3' end processing.

Methods

Strains, culture conditions, and crosses

CC-125 and CC-373 from the *Chlamydomonas* stock center were used as transformation recipients. +IR-ECS-DG2, IR-DG2, -IR-DG2, ECS-DG, +IR-ECS-DA and ECS-DA were described previously (Hicks *et al.*, 2002), and *mcd4* was described in Rymarquis *et al.* (2006). All strains were grown in Tris-Acetate-Phosphate (TAP) medium (Harris, 1989), or medium lacking acetate under 23 hr light and 1 hr dark at 25°C . All crosses were carried out as described previously (Gorman and Levine, 1965).

Reporter gene constructs

atpB variants were amplified by PCR using the templates and primers listed in Table 3.1. For some constructs, two sets of primers (Set 1 and Set 2) were used to introduce *Sac*I or *Spe*I site as a mutation into the *atpB* PD. A flow chart illustrating this cloning strategy is shown for $pIR_{Sac}ECS$ in Figure 3.8. All PCR products were inserted into the pGEM-T Easy vector (Promega, Madison, WI) and verified by DNA sequening. To generate $pIR_{Sac}ECS$ and $pIR_{\Delta L}ECS$, plasmids containing the products from Set 1 were digested with *Bgl*II and *Sac*I, and Set 2 with *Sac*I and *BamH*I, to release the PCR products. These were combined with the *Bgl*II-*BamH*I-digested vector backbone of pΔIRECS in a triple ligation to obtain $pIR_{sac}ECS$ and $pIR_{\Delta L}ECS$. For pIR_AECS, $pIR_{A'}ECS$, pIR_BECS, $pIR_{B'}ECS$, Set 1 PCR products were released from pGEM-T Easy plasmids by *Bgl*II-*Spe*I digestion, and Set 2 products with *Spe*I-*BamH*I. These were combined in a triple ligation with the *Bgl*II-*BamH*1 digested vector backbone of pΔIRECS to generate their respective plasmids. $pIR_{AB'}ECS$ was generated by ligating the ~200 bp *Sac*I-*BamH*I fragment

containing the ECS site from $pIR_{B'}ECS$ into the similarly digested 3.1 kb vector backbone of pIR_AECS. $pIR_{BA'}ECS$ and $pIR_{BB'}ECS$ were generated in a similar fashion using the pIR_BECS backbone and the ECS-containing fragment of $pIR_{A'}ECS$, and the pIR_BECS backbone plus ECS-containing fragment of $pIR_{B'}ECS$, respectively. For insertion into DG2 or DAAD, the *atpB* PDs were released by digestion with *Bgl*II and *BamH*I and ligated into *Bgl*II-digested DG2 or DAAD; *Bgl*II cuts uniquely at the +25 position of the *petD* 5' UTR.

For the *tscA*-based strains, the *Pst*I-*Sph*I fragment from plasmid P-71 (*Chlamydomonas* stock center) was subcloned into pGEM5Z (Promega) at the *Pst*I and *Sph*I sites to generate p-71pGEM. The 2 kb *EcoR*V-*Sma*I fragment containing the resistance cassette from pUC-atpX-AAD (Goldschmidt-Clermont, 1991) was inserted into *EcoR*V-digested p-71pGEM in both orientations to obtain $tscA_{+R}$ and $tscA_{-R.}$ Two variants of pUC-atpX-AAD, atpX-aadA+RA and atpX-aadA+A, were generated by first amplifying the *atpB* PD using the primers in Table 3.1. For atpX-aadA+RA, the PCR products were digested with *Sac*I and *Spe*I and ligated into the *Sac*I and *Spe*I sites of pUC-atpX-AAD, whereas the PCR product atpX-aadA+A were digested with *Pst*I and *Sph*I and ligated into *Pst*I and *Sph*I digested atpX-aadA to replace the *rbcL* PD. As before, the *aadA* cassettes from atpX-aadA+RA and atpX-aadA+A were excised by *EcoR*V-*Sma*I digestion and ligated in either orientation into p-71pGEM, generating $tscA_{+RA}/tscA_{-RA}$ and $tscA_{+A}/tscA_{-A}$ respectively.

Table 3.1. Templates and primers used to generate variants of the *atpB* processing determinant. Plasmids indicate the name of the final plasmid after insertion of the *atpB* PD.

Plasmids	Template	Forward Primer	Reverse Primer
pΔIRECS	WT *atpB*	attagtaaagatctttcattaaaa	acgtttaacgggatcctttattttaaag
pECS39	WT *atpB*	ggacaccattaagatcttttcttctt	acgtttaacgggatcctttattttaaag
pECS15	WT *atpB*	ggacaccattaagatcttttcttctt	ggatccattaaatagctaatatgacatat
pIRNotECS	pΔ26-Not	ggacaccattaagatcttttcttctt	aaagaaactggatcccgtggacgac
pECSNot	pΔ26-Not	ggacaccattaagatcttttcttctt	aaagaaactggatcccgtggacgac
pECSNot15	pΔ26-Not	ggacaccattaagatcttttcttctt	ggatccattaaatagctgcggccgcatat
pIRSacECS	WT *atpB*	Set 1 gtgaattagatcttttaccagaa	Set 1 ctttaagaagagagctccttaatggtg
		Set 2 caccattaaggagctctcttcttaaag	Set 2 aaagaaactggatcccgtggacgac
$pIR_{\Delta L}ECS$	WT *atpB*	Set 1 gtgaattagatcttttaccagaa	Set 1 aacttaatggagctcaaatattta
		Set 2 gttttcttcttagagctccaatttat	Set 2 aaagaaactggatcccgtggacgac
pIR_AECS	$IR_{Sac}ECS$	Set 1 gtgaattagatcttttaccagaa	Set 1 ccaaatatttattactagtacacataaa
		Set 2 tttatgtgtactagtaataaatatttgg	Set 2 aaagaaactggatcccgtggacgac
$pIR_{A'}ECS$	$IR_{Sac}ECS$	Set 1 gtgaattagatcttttaccagaa	Set 1 tttatgtgtactagtaataaatatttgg
		Set 2 aagagccaaactagtttaattacac	Set 2 aaagaaactggatcccgtggacgac
pIR_BECS	$IR_{Sac}ECS$	Set 1 gtgaattagatcttttaccagaa	Set 1 gtgtccaaataactagtttaattacac
		Set 2 gtgtaattaaactagttatttggacac	Set 2 aaagaaactggatcccgtggacgac
$pIR_{B'}ECS$	$IR_{Sac}ECS$	Set 1 gtgaattagatcttttaccagaa	Set 1 gtgtaattaaactagtttggctctt
		Set 2 ccaatttattactagtacacataca	Set 2 aaagaaactggatcccgtggacgac
atpx-aadA+RA	WT *atpB*	agaaactggagctcgtggacga	tgatttactagtacaagcattctactt
atpx-aadA+A	WT *atpB*	ttaccagaacctgcagtctacttag	agaaactgtagcatgcggacgacaaa

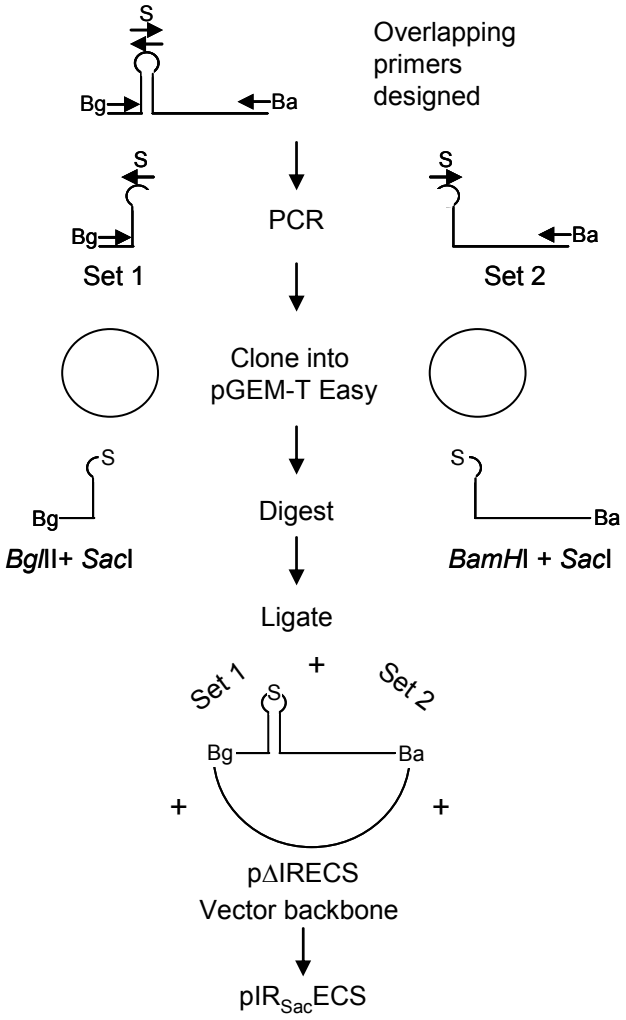


Figure 3.8 Cloning strategy used to generate the pIR$_{Sac}$ECS. Set 1 and Set 2 refer to PCR products generated using the primers in Table 3.1. S, Bg, and Ba represent a *Sac*I, *Bgl*II, and *BamH*I sites, respectively.

Chloroplast genome transformation

All transformations were carried out using biolistics as described previously (Boynton *et al.*, 1988). CC-373, which carries a deletion of most of the *atpB* gene, was the transformation recipient for the DG2- and DAAD-based constructs, while CC-125 served as the recipient for the *tscA* constructs. Transformants were selected on medium lacking acetate for DG2 and DAAD constructs, and TAP+200 μg/ml spectinomycin for the *tscA* constructs.

Transformants underwent single colony purification by streaking on medium lacking acetate (DG2 and DAAD constructs) or TAP+500 μg/ml spectinomycin (*tscA* constructs) until they became homoplasmic for the transgene. For the DG2- and DAAD-based construct transformants, homoplasmy was confirmed by PCR using the primers Pr-atpB/159, Pr-d031306-rev-bp1381 and gus 5' for DG reporter genes, and Pr-atpB/159, Pr-d031306-rev-bp1381 and SZ-aada(F2) for DAAD reporter genes. A ~520 bp product can be amplified from strains containing the reporter genes, whereas a ~250 bp product is amplified from transformants not containing the reporter gene, which is sometimes lost due to copy correction between the two chloroplast genome inverted repeats. The homoplasmic state of the *tscA* constructs was confirmed by gel blots of *Pst*I-digested total DNA hybridized with a *tscA* probe generated using the primers lar18 Forward *tscA* and lar29 Reverse *tscA*.

Table 3.2. Primers used to verify homoplasmic state and generate probes.

Primer	Sequence
Pr-atpB/159	ccactgccactaaaatttatttgcctcctaacg
SZ-aada(F2)	gtcatacttgaagctagacaggc
gus 5'	gcgacctcgcaaggc
Pr-d031306-rev-bp1381	gaaggaggacgtccccttacg
lar18 Forward *tscA*	gtccattggcgttccgtaagg
lar19 Reverse *tscA*	ttggtgaagtatagcagcgagatg
gusF	tagcgggactttgcaagtg
gusR	cagcagcagtttcatcaatca
Cr *psaA* exon2 F	aaaaacaaatcctacctgactaaacc
Cr *psaA* exon2 R	tgtcgcacccacttaaccaaatg
atpB 5' F	gttttgtacggcttgaagg
atpB 5' R	ggatggagacgttttgagttc
DAAD 5' F	tactccgaaggaacttgttagc
DAAD 5' R	catccccttacagtattcttgc

RNA isolation and filter hybridizations

Cells were grown in TAP medium and RNA was extracted using Tri-Reagent (Molecular Resource Center, Cincinnati, OH). For filter hybridizations, 10 µg of total RNA was fractionated in 1.2% agarose, 6% formaldehyde gels, transferred to nylon filters (Genescreen, Perkin Elmer, Boston, MA), and cross-linked by UV irradiation. Double-stranded PCR products were generated using primers gusF and gusR for *uidA,* lar005 and lar006 for *aadA* (Hicks *et al.*, 2002), and Cr *psaA* exon2 F and Cr *psaA* exon2 R for *psaA* exon 2. The probes upstream of *atpB* and DAAD were generated using primers *atpB* 5' F and *atpB* 5' R and DAAD 5' F and DAAD 5' R, respectively. Probes for 16S rRNA and *atpA* were described previously (Rymarquis *et al.,* 2006). Single-stranded probes for *uidA* and *aadA* were generated as previously described (Hicks *et al.*, 2002). Filters were prehybridized, hybridized and washed as previously described (Church and Gilbert, 1984). All hybridizations were analyzed using the Storm system (Amersham Biosciences Piscataway, NJ).

REFERENCES

Blowers AD, Klein U, Ellmore GS, Bogorad L (1993) Functional *in vivo* analyses of the 3' flanking sequences of the *Chlamydomonas* chloroplast *rbcL* and *psaB* genes. Mol. Gen. Genet. **238:** 339-349

Boynton JE, Gillham NW, Harris EH, Hosler JP, Johnson AM, Jones AR, Randolph-Anderson BL, Robertson D, Klein TM, Shark KB, Sanford JC (1988) Chloroplast transformation in *Chlamydomonas* with high velocity microprojectiles. Science **240:** 1534-1538

Church G, Gilbert W (1984) Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:** 1991-1995

Drager RG, Girard-Bascou J, Choquet Y, Kindle KL, Stern DB (1998) *In vivo* evidence for 5' to 3' exoribonuclease degradation of an unstable chloroplast mRNA. Plant J. **13:** 85-96

Drager RG, Zeidler M, Simpson CL, Stern DB (1996) A chloroplast transcript lacking the 3' inverted repeat is degraded by 3' to 5' exoribonuclease activity. RNA **2:** 652-663

Goldschmidt-Clermont M (1991) Transgenic expression of aminoglycoside adenine transferase in the chloroplast: A selectable marker for site-directed transformation of *Chlamydomonas*. Nucleic Acids Res. **19:** 4083-4090

Goldschmidt-Clermont M, Choquet Y, Girard-Bascou J, Michel F, Schirmer-Rahire M, Rochaix JD (1991) A small chloroplast RNA may be required for *trans*-splicing in *Chlamydomonas reinhardtii*. Cell **65:** 135-144

Goldschmidt-Clermont M, Girard-Bascou J, Choquet Y, Rochaix JD (1990) *Trans*-splicing mutants of *Chlamydomonas reinhardtii*. Mol. Gen. Genet. **223:** 417-425

Gorman DS, Levine RP (1965) Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardii*. Proc. Natl. Acad. Sci. USA **54:** 1665-1669

Harris EH (1989) The *Chlamydomonas* Sourcebook: A comprehensive guide to biology and laboratory use. Academic Press, San Diego

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

Henkin TM (2000) Transcription termination control in bacteria. Curr. Opin. Microbiol. **3:** 149-153

Hicks A, Drager RG, Higgs DC, Stern DB (2002) An mRNA 3' processing site targets downstream sequences for rapid degradation in *Chlamydomonas* chloroplasts. J. Biol. Chem. **277:** 3325-3333

Higgs DC, Shapiro RS, Kindle KL, Stern DB (1999) Small *cis*-acting sequences that specify secondary structures in a chloroplast mRNA are essential for RNA stability and translation. Mol. Cell. Biol. **19:** 8479-8491

Komine Y, Kwong L, Anguera MC, Schuster G, Stern DB (2000) Polyadenylation of three classes of chloroplast RNA in *Chlamydomonas reinhardtii*. RNA **6:** 598-607

Levy H, Kindle KL, Stern DB (1999) Target and specificity of a nuclear gene product that participates in mRNA 3' end formation in *Chlamydomonas* chloroplasts. J. Biol. Chem. **274:** 35955-35962

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc. Natl. Acad. Sci. USA **101:** 7287-7292

Nickelsen J, Link G (1993) The 54 kDa RNA-binding protein from mustard chloroplasts mediates endonucleolytic transcript 3' end formation *in vitro*. Plant J. **3:** 537-544

Nishimura Y, Kikis EA, Zimmer SL, Komine Y, Stern DB (2004) Antisense transcript and RNA processing alterations suppress instability of polyadenylated mRNA in *Chlamydomonas* chloroplasts. Plant Cell **16:** 2849-2869

Rott R, Drager RG, Stern DB, Schuster G (1996) The 3' untranslated regions of chloroplast genes in *Chlamydomonas reinhardtii* do not serve as efficient transcriptional terminators. Mol. Gen. Genet. **252:** 676-683

Rott R, Levy H, Drager RG, Stern DB, Schuster G (1998) 3'-processed mRNA is preferentially translated in *Chlamydomonas reinhardtii* chloroplasts. Mol. Cell. Biol. **18:** 4605-4611

Rott R, Liveanu V, Drager RG, Higgs DC, Stern DB, Schuster G (1999) Altering the 3' UTR endonucleolytic cleavage site of a *Chlamydomonas* chloroplast mRNA affects 3' end maturation *in vitro* but not *in vivo*. Plant Mol. Biol. **40:** 676-686

Rymarquis L, Higgs D, Stern D (2006) Nuclear suppressors define three factos that participate in both 5' and 3' end processing of mRNAs in *Chlamydomons* chloroplasts. Plant J.**:** In Press

Sakamoto W, Kindle KL, Stern DB (1993) *In vivo* analysis of *Chlamydomonas* chloroplast *petD* gene expression using stable transformation of β-glucuronidase translational fusions. Proc. Natl. Acad. Sci. USA **90:** 497-501

Stern DB, Gruissem W (1987) Control of plastid gene expression: 3' inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. Cell **51:** 1145-1157

Stern DB, Kindle KL (1993) 3' end maturation of the *Chlamydomonas reinhardtii* chloroplast *atpB* mRNA is a two-step process. Mol. Cell. Biol. **13:** 2277-2285

Sugita M, Sugiura M (1996) Regulation of gene expression in chloroplasts of higher plants. Plant Mol. Biol. **32:** 315-326

Walter M, Kilian J, Kudla J (2002) PNPase activity determines the efficiency of mRNA 3'-end processing, the degradation of tRNA and the extent of polyadenylation in chloroplasts. EMBO J. **21:** 6905-6914

CHAPTER 4.

BEYOND COMPLEMENTATION : MAP-BASED CLONING IN *CHLAMYDOMONAS REINHARDTII* *

ABSTRACT

Chlamydomonas reinhardtii is an excellent model system for plant biologists because of its ease of manipulation, facile genetics, and the ability to transform the nuclear, chloroplast and mitochondrial genomes. Numerous forward genetics studies have been performed in *Chlamydomonas*, in many cases to elucidate the regulation of photosynthesis. One of the resultant challenges is moving from mutant phenotype to the gene mutation causing that phenotype. To date, complementation has been the primary method for gene cloning, but this is impractical in several situations, for example when the complemented strain cannot be readily selected, or in the case of recessive suppressors which restore photosynthesis. New tools including a molecular map consisting of 506 markers, and an 8X draft nuclear genome sequence are now available, making map-based cloning increasingly feasible. Here we discuss advances in map-based cloning developed using the strains *mcd4* and *mcd5,* which carry nuclear suppressors restoring photosynthesis to chloroplast mutants. Tools which have not been previously applied to *Chlamydomonas*, such as bulked segregant analysis and marker duplexing, are being implemented to increase the speed at which one can go from mutant phenotype to gene. In addition to assessing and applying current resources, we outline anticipated future developments in map-based cloning in the context of the newly-extended *Chlamydomonas* genome initiative.

*Adapted from: Rymarquis LA, Handley JM, Thomas M, Stern DB (2005) Beyond complementation. Map-based cloning in *Chlamydomonas reinhardtii*. Plant Physiol. **137** 557-566.

INTRODUCTION

Sometimes called “green yeast” (Goodenough, 1992), the unicellular, eukaryotic green alga *Chlamydomonas reinhardtii* (hereafter called “*Chlamydomonas*”) is a venerable model system for plant biology, as well as for cell motility. The tag “green yeast” refers to its haploid vegetative state, the existence of two mating types, and the general similarity in applicable genetic techniques. These aspects of *Chlamydomonas* biology have been previously reviewed (Rochaix, 1995).

Like many microorganisms, screening of *Chlamydomonas* strains for rare mutations is straightforward, since large numbers of cells can be plated on an appropriate selective medium, or non-swimmers, for example, can be selected from large numbers of swimming cells. At the same time, the ease of nuclear transformation in *Chlamydomonas*, coupled with the plant-like nonhomologous integration of transforming DNA, facilitates the creation of insertional mutant collections. Taken together, the assortment of techniques useable for *Chlamydomonas* indulges both the amateur and experienced geneticist, yielding sometimes overwhelming collections of mutant strains. In this report, we focus on mutants affecting photosynthesis, in keeping with the thrust of this journal, and the emphasis of the newly-renewed and NSF-supported *Chlamydomonas* genome project[1]. However, the map-based cloning tools described here are generally applicable to *Chlamydomonas* biology.

The use of *Chlamydomonas* to study the elaboration and regulation of the photosynthetic apparatus is long-established and was recently reviewed (Dent *et al.*, 2001; Grossman, 2000). Key to this is the ability to maintain nonphotosynthetic (PS-) mutants on acetate-containing media, as well as the

[1] http://www.chlamy.org/

ability to use replica plating ±acetate and/or chlorophyll fluorescence to identify such mutants (Bennoun and Béal, 1997; Niyogi et al., 1997). Furthermore, numerous photosynthetic (PS+) suppressors have been recovered from screening of PS- strains (e.g. Bernd and Kohorn, 1998; Esposito *et al.*, 2001; Girard-Bascou *et al.*, 1992; e.g. Levy *et al.*, 1997; Li *et al.*, 2002; Nickelsen, 2000).

Recovery of wild-type alleles of genes mutated in PS- strains has been successful, since both genomic complementation with selection on medium lacking acetate, or identification of DNA flanking an insertional mutant site, are relatively straightforward although perhaps tedious methods (e.g. Auchincloss *et al.*, 2002; e.g. Boudreau *et al.*, 2000; Dauvillee *et al.*, 2003; Gumpel *et al.*, 1995; Vaistij *et al.*, 2000). However, in cases where PS+ suppressors have been recovered from PS- strains, or where a trait otherwise cannot be selected on medium lacking acetate, recovery of the mutation requires other methods. In the case of PS+ suppressors, for example, a recessive suppressor would yield a PS- phenotype upon complementation, and even dominant suppressors such as *mcd2*, which suppresses the nuclear *mcd1-2* mutation responsible for instability of the chloroplast *petD* mRNA (Esposito et al., 2001), require construction of a genomic library from the suppressed strain if they are to be cloned by complementation. Another example is the xanthophyll cycle mutant *npq1*, which is defective in nonphotochemical quenching (Niyogi et al., 1997). Although *npq1* was generated in an insertional mutant population, its defect is not linked to the *ARG7* insertional mutagen. In each of these cases, isolation of the gene of interest could be achieved through map-based cloning in a suitably developed system.

Map-based cloning relies on two basic principles; namely, the existence of a genetic/physical map, and the ability to generate progeny of sexual crosses which segregate for the trait of interest as well as phenotypic and/or molecular markers. In higher plants, such resources are most fully advanced in *Arabidopsis* and rice which not coincidentally, have complete nuclear genome sequences. Furthermore, interfertile polymorphic ecotypes (*Columbia* and Landsberg *erecta*; and *indica* and *japonica*, respectively) have been utilized as sources of genetic variation to introduce into selected mutant backgrounds. Resources for Arabidopsis and rice have been extensively publicized and have been utilized in numerous examples of successful gene isolation (Chen et al., 2002; Garcia-Hernandez et al., 2002; Torjek et al., 2003); maize is another subject of intensive efforts (Coe et al., 2002; Cone et al., 2002).

Here we present a case-based study to describe existing and projected resources for map-based cloning in *Chlamydomonas reinhardtii*. Mutations generated in this species can be mapped by crossing to the interfertile strain known as *C. grossii*, S1-D2 or its culture collection designation of CC-2290, which has a suitable profusion of STS (Sequence Tagged Site), CAPS (Cleavable Amplified Polymorphic Sequence), SNP (Single Nucleotide Polymorphism), and RFLP (Restriction Fragment Polymorphic Sequence) markers (Gross *et al.*, 1988; Grossman *et al.*, 2003; Vysotskaia *et al.*, 2001). Beginning with laborious RFLP-based mapping (Gross et al., 1988), *Chlamydomonas* mapping has moved towards a PCR-based method (Kathir et al., 2003), and now is poised to incorporate more high-throughput methods. This, in concert with an increasingly complete nuclear genome sequence

(Grossman et al., 2003), provides the necessary tools for studies of all classes of mutations.

The nuclear mutants *mcd4* and *mcd5* were derived from strains LS2 and LS6 respectively, in which mutations engineered into the 5' untranslated region of the chloroplast *petD* gene caused RNA instability and thus a PS- phenotype (Higgs et al., 1999). Both *mcd4* and *mcd5* are PS+ mutants which do not carry a molecular tag, and genetic analysis showed *mcd4* to be recessive (Rymarquis *et al.*, 2006). Thus, complementation of *mcd4* or possibly *mcd5* with the wild-type genes would revert the PS+ phenotype to PS-, making genomic complementation an inappropriate approach. We therefore decided to map *mcd4* and *mcd5* by using available genomic resources, and by developing new ones as opportunities or needs arose.

RESULTS AND DISCUSSION

A Composite Map of Molecular Markers

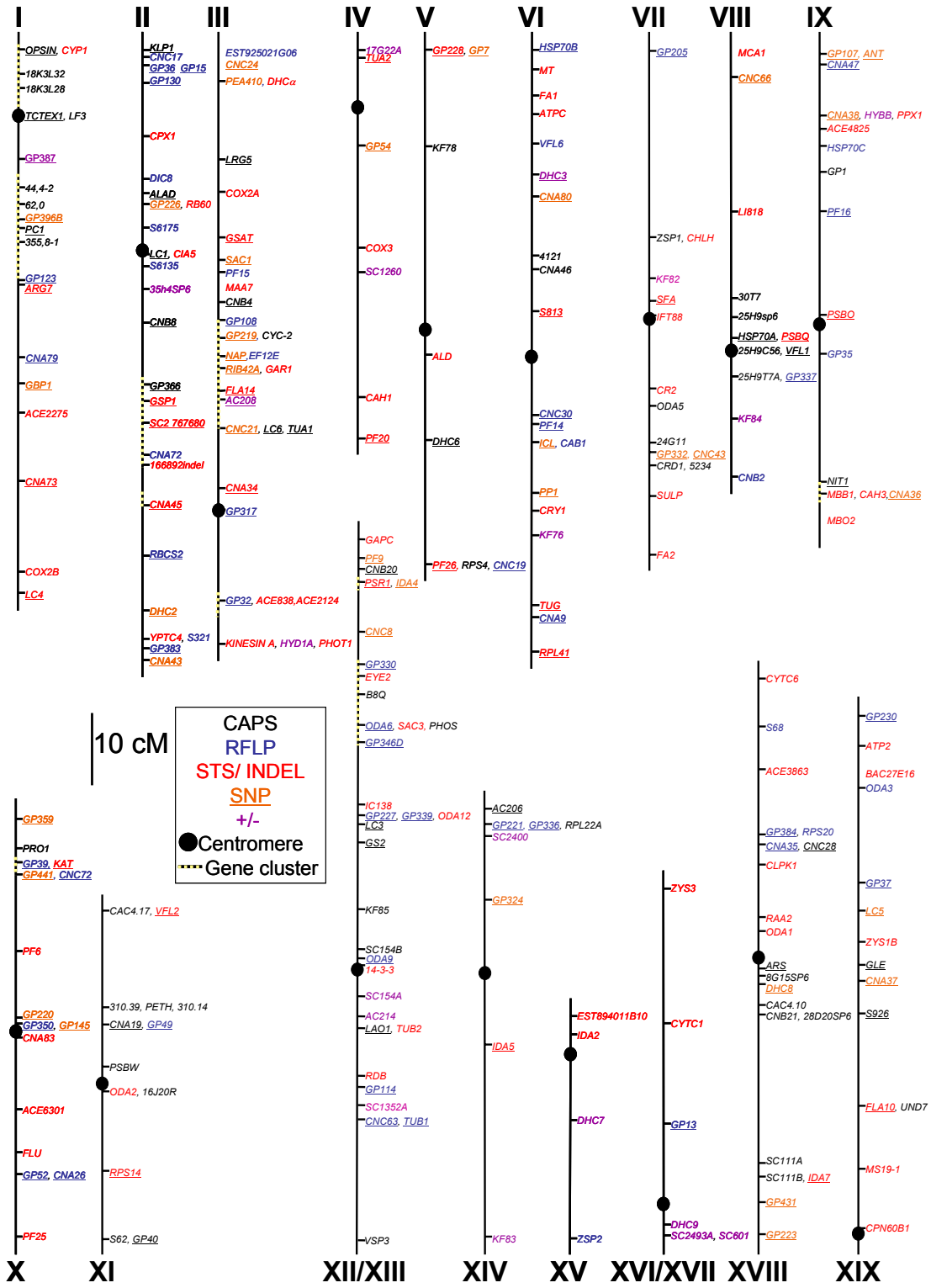
As noted above, *Chlamydomonas* molecular markers are derived from comparisons of laboratory strains and the interfertile polymorphic strain S1-D2. To date, four laboratories have generated at least 506 STS, InDel (insertion/deletion), CAPS, SNP, +/- (a PCR fragment amplifiable in one polymorphic strain, but not the other), and RFLP markers, whose respective utilities are discussed below. The combined molecular maps shown in Figure 4.1 contain 385 markers, representing 266 loci arranged on 17 linkage groups (LG), and differ from previously published maps (Kathir et al., 2003) by including markers generated by multiple laboratories. Fifteen markers could be assigned to a LG, but not a specific location, whereas 107 are not displayed on the map because they are within marker-dense gene clusters or because

precise marker order could not be determined due to insufficient data. Complete information on the markers can be found at http://www.chlamy.org/kit.html.

Of the six types of markers mentioned above, five are PCR-based (STS, InDel, CAPS, SNP, +/-), whereas RFLPs rely on DNA gel blots. STS markers can be generated when sequences of both strains are known. In *Chlamydomonas*, most STS markers have one primer conserved in both *C. reinhardtii* and S1-D2, and one that is specific for each strain. The three primers are used in a single reaction, generating products that differ in size for the *C. reinhardtii* or the S1-D2 allele. InDels have primers conserved in both strains, but yield PCR products of different sizes due to insertions or deletions. Both STS and InDels only require visualization on an agarose gel to distinguish alleles. CAPS markers, on the other hand, require an additional step, since polymorphisms are revealed upon restriction enzyme digestion of the PCR product. SNPs are easy to find since only a single nucleotide difference is required, and on average 2.7 base substitutions were found in every 100 bp of sequence when S1-D2 and the laboratory strain were compared (Kathir et al., 2003). A disadvantage is that the reagents and technology for SNP detection are relatively expensive and some methods require special equipment. +/- markers are the last PCR-based marker shown on the map. These only amplify a PCR product from one parent, and thus have a higher error rate since failed PCR reactions would be scored as the non-amplifying parent, skewing recombination-based distance calculations. Use of RFLP markers is comparatively arduous, requiring isolation of genomic DNA and gel blot hybridizations.

Because each laboratory whose markers are represented in Figure 4.1 used different mapping populations to determine either centromere linkage or recombination-based distances, such values could not be directly compared. Approximate distances were calculated based on the combined data and where available, the genome sequence. Markers whose recombination data were not available were placed by alignment with the nuclear genome sequence, and their distances were calculated based on the assumption by Kathir *et al.* (2003) that 1 centimorgan (cM) equals 100 kb. This estimate was based on the total number of cM in the genome and an approximate genome size of 10^8 base pairs. To determine the accuracy of this assumption, 23 marker pairs were examined for their actual kb:cM ratio. This ratio, not surprisingly, varied widely, from 0.860 kb/cM between *TPX* and *FA1* on LG VI, to 511 kb/cM between *ARG7* and GP123 on LG I. Varied rates of recombination are also seen in Arabidopsis, where the relationship between genetic and physical distance ranges, for example, from 30 kb/cM to >550 kb/cM on chromosome 4 (Schmidt et al., 1995). Due to this variance, most distances must be regarded as rough estimates, and are annotated as such in the Updated Excel Table at http://www.chlamy.org/kit.html. Nonetheless, gene/marker order is expected to be accurate, especially where it has been confirmed by the genome sequence.

Figure 4.1. Molecular map of *Chlamydomonas reinhardtii*. 506 molecular markers have been arranged on the 17 LGs; 385 are shown here. Markers are color coded as to their type: CAPS markers are in black, RFLP markers are in blue, STS and InDel makers are in red, +/- markers are in purple. All SNP markers are underlined; those which can be assayed using an alternative method are indicated by a color other than orange. Broken lines indicate gene clusters, where only representative markers are shown.


I
II
III
IV
V
VI
VII
VIII
IX
OPSIN, CYP1
18K3L32
18K3L28
TCTEX1, LF3
GP387
44,4-2
62,0
GP396B
PC1
355,8-1
GP123
ARG7
CNA79
GBP1
ACE2275
CNA73
COX2B
LC4
KLP1
CNC17
GP36
GP15
GP130
CPX1
DIC8
ALAD
GP226, RB60
S6175
LC1, CIA5
S6135
35h4SP6
CNB8
GP366
GSP1
SC2 767680
CNA72
166892indel
CNA45
RBCS2
DHC2
YPTC4, S321
GP383
CNA43
EST925021G06
CNC24
PEA410, DHCα
LRG5
COX2A
GSAT
SAC1
PF15
MAA7
CNB4
GP108
GP219, CYC-2
NAP,EF12E
RIB42A, GAR1
FLA14
AC208
CNC21, LC6, TUA1
CNA34
GP317
GP32, ACE838,ACE2124
KINESIN A, HYD1A, PHOT1
17G22A
TUA2
GP54
COX3
SC1260
CAH1
PF20
GAPC
PF9
CNB20
PSR1, IDA4
CNC8
GP330
EYE2
B8Q
ODA6, SAC3, PHOS
GP346D
IC138
GP227, GP339, ODA12
LC3
GS2
KF85
SC154B
ODA9
14-3-3
SC154A
AC214
LAO1, TUB2
RDB
GP114
SC1352A
CNC63, TUB1
VSP3
GP228, GP7
KF78
ALD
DHC6
PF26, RPS4, CNC19
HSP70B
MT
FA1
ATPC
VFL6
DHC3
CNA80
4121
CNA46
S813
CNC30
PF14
ICL, CAB1
PP1
CRY1
KF76
TUG
CNA9
RPL41
AC206
GP221, GP336, RPL22A
SC2400
GP324
IDA5
KF83
EST894011B10
IDA2
DHC7
ZSP2
GP205
ZSP1, CHLH
KF82
SFA
IFT88
CR2
ODA5
24G11
GP332, CNC43
CRD1, 5234
SULP
FA2
ZYS3
CYTC1
GP13
DHC9
SC2493A, SC601
MCA1
CNC66
LI818
30T7
25H9sp6
HSP70A, PSBQ
25H9C56, VFL1
25H9T7A, GP337
KF84
CNB2
CYTC6
S68
ACE3863
GP384, RPS20
CNA35, CNC28
CLPK1
RAA2
ODA1
ARS
8G15SP6
DHC8
CAC4.10
CNB21, 28D20SP6
SC111A
SC111B, IDA7
GP431
GP223
GP107, ANT
CNA47
CNA38, HYBB, PPX1
ACE4825
HSP70C
GP1
PF16
PSBQ
GP35
NIT1
MBB1, CAH3,CNA36
MBO2
GP230
ATP2
BAC27E16
ODA3
GP37
LC5
ZYS1B
GLE
CNA37
S926
FLA10, UND7
MS19-1
CPN60B1
10 cM
CAPS
RFLP
STS/ INDEL
SNP
+/-
Centromere
Gene cluster
GP359
PRO1
GP39, KAT
GP441, CNC72
PF6
GP220
GP350, GP145
CNA83
ACE6301
FLU
GP52, CNA26
PF25
CAC4.17, VFL2
310.39, PETH, 310.14
CNA19, GP49
PSBW
ODA2, 16J20R
RPS14
S62, GP40
X
XI
XII/XIII
XIV
XV
XVI/XVII
XVIII
XIX

Mapping of *MCD4*

To develop a population for development and application of mapping methods, the mutant *mcd4* [LS2] mt+ was crossed to S1-D2 mt-[2]. The LS2 chloroplast genotype confers a nonphotosynthetic (PS-) phenotype due to instability of the *petD* mRNA (Higgs et al., 1999), and this PS- phenotype is suppressed by the nuclear mutation *mcd4*. Thus, photosynthetic (PS+) progeny of the cross carry the mutant *mcd4* allele, whereas PS- progeny carry the S1-D2 wild-type allele *MCD4*. To minimize analysis of duplicate recombination events, one PS+ progeny was chosen from each of 54 tetrads to create the mapping population for the mutant allele, *mcd4*. With a population of 54 progeny, a 17.7 cM (Kosambi units) resolution can be expected with a confidence interval of 95% (p>.05). At this resolution, 83% of the progeny should contain the marker allele from the *mcd4* parent, although any marker giving greater than 70% *C. reinhardtii* alleles was deemed significant by chi squared analysis and the region examined with additional markers. 60 markers were scored for the percent of progeny that contained the *mcd4* allele, resulting in an unambiguous assignment of *mcd4* to LG II. The STS marker with the highest linkage was *GSP1*, for which 42/53 (80%) progeny displayed the PCR product size associated with the *mcd4* parent. Another tightly linked marker was the CAPS marker *DHC4*, shown as an example in Figure 4.2. Since 41/52 (79%) of the progeny contained the *mcd4* allele in this example, *MCD4* can be calculated as 22 cM distant from *DHC4*. Linkage was also seen to *CIA5* and CNA45, which based on PCR data are both 26 cM from *MCD4* (data not shown).

[2] By convention, the chloroplast genotype is given in brackets, except where it is wild-type.

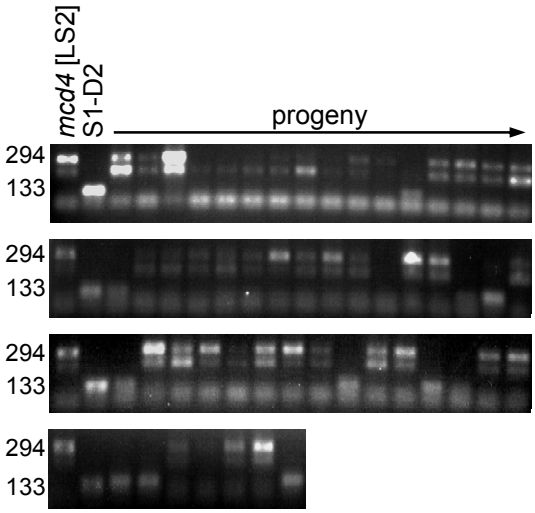


Figure 4.2. *DHC4* is linked to *mcd4*. DNAs isolated from the mapping population described in the text were amplified with the *DHC4* CAPS primers and digested with *Pst*I. Digests were analyzed in a 3% agarose gel. The U indicates the undigested PCR product of 235 bp. Upon digestion the *mcd4* allele yields 191 and 44 bp products, whereas the S1-D2 (*MCD4*) allele yields ~170, 44, and ~30 bp products. Forty-five out of 53 progeny tested are shown here, and the parental controls are at the left of the top row.

When comparing the map to the linkage shown by markers in the *mcd4* mapping population, it became apparent the mapping population overestimated the distance between markers on LG II. An example of this is that the mapping population suggested that *CIA5* and CNA45 are 52 cM apart, based on the fact that they are both 26 cM from *MCD4,* while previous mapping in *Chlamydomonas* suggest that they are only 30 cM apart. This combined with the fact that no marker displayed more than 80% linkage, indicated that our mapping population was biased. Two possible sources of bias are suppressed recombination between the *mcd4* and S1-D2 genomes in the region of interest, or failure to inherit the chloroplast genome containing the LS2 mutation. The latter would cause a bias because if the WT *petD* gene was inherited from the S1-D2 parent, the strain would be PS+ whether or not it

also inherited the nuclear *mcd4* allele. Although the chloroplast is usually inherited from the mt+ parent, it has been reported that the mt- chloroplast can be inherited up to 5% of the time (Harris, 1989).

To determine if S1-D2 cpDNA inheritance was confounding the mapping efforts, a marker needed to be generated to distinguish S1-D2 cpDNA from the LS2 cpDNA. LS2 contains a unique *Not*I site providing a convenient CAPS marker to verify the chloroplast genome, but upon amplification we found unexpectedly that the S1-D2 product was 400 bp larger than the *mcd4* [LS2] product, even before digestion. Sequencing revealed that the size shift was due to a TTTATATACTCCAAA tandem repeat in the 5' UTR of *petD* that is not found in the *C. reinhardtii* sequence. When the mapping population was tested for this polymorphism, 17 out of 54 progeny were found to have inherited S1-D2 cpDNA. A similarly high proportion of mt- chloroplast inheritance was found in the *mcd5* mapping population generated by Jocelyn Handley. It is not yet known whether this high rate of mt- chloroplast inheritance is conferred by unknown nuclear alleles of S1-D2, or is conditioned by the *mcd4* and *mcd5* mutations. We were unable to find anyone in the *Chlamydomonas* community who had previously examined or encountered this issue.

We replaced the 17 unusable progeny following a second cross between *mcd4* [LS2] mt+ and S1-D2, giving a final mapping population of 64 progeny. This population gave 100% linkage to *mcd4* for both *GSP1* and *DHC4*, rather than the 80% and 79% found previously. These data confirmed unambiguously that *mcd4* is on LG II in or very near the *GSP1* gene cluster.

Bulked segregant analysis and duplexing

In order to place markers at 20-30 cM intervals, as well as within 10 cM of the end of every linkage group, 57-72 markers would be needed to span the 1,107 cM genome. Markers placed at each end of the linkage group ensure adequate coverage in cases where the genetic maps are longer than the molecular maps, indicating potentially missing sequence. A subset of 67 markers, all of which are currently known, could provide this coverage. Doing so with the 50 progeny required to achieve 20-30 cM coverage would entail 3,350 PCR reactions, a daunting number. In order to reduce the time and expense of mapping, bulked segregant analysis (BSA) and marker duplexing were evaluated. Although these techniques have been used in other systems, for example *Arabidopsis* (Lukowitz et al., 2000), they have not been tested systematically for *Chlamydomonas*. In BSA, DNAs from multiple segregating progeny are combined, and results from PCR-based markers are examined for significant bias from a roughly equal contribution from each parent.

For our markers, BSA was evaluated by Jocelyn Handley by creating defined mixtures of *mcd5* DNA and S1-D2 DNA, where the total DNA amount was held constant. The strain *mcd5* is a laboratory (*C. reinhardtii*)-derived PS+ suppressor of a chloroplast mutation LS6 (Higgs et al., 1999), and analogous to *mcd4*. 40 STS markers have been examined for compatibility with BSA, and three examples are shown in Figure 4.3. Examination of the reactions using 1:1 DNA ratios shows that the longer product stained more brightly in two cases, as would be expected for a roughly equal amplification of the two alleles. The relative prominence of products also changed as expected, although in some reactions one product or the other was stronger or weaker than anticipated. We have observed this variability for numerous primer sets

and it is to some extent unavoidable. Importantly, however, even at the most extreme ratios (4-fold excess of *mcd5* or 5-fold excess of S1-D2), the diluted allele was still visible. This suggests that BSA of 5-6 bulked progeny will still reveal a rare allele for these primer sets, and that the overall banding pattern is to some degree indicative of the relative number of each allele. In particular, staining corresponding to the 1:1 ratio would be observed for unlinked markers to the gene of interest, which would avert the need to deconvolute the bulked DNAs for those particular markers. This technique was used successfully with much larger numbers of individuals from an *Arabidopsis* mapping population (Lukowitz et al., 2000), perhaps being facilitated by the lower [G+C] content and better sequence information.

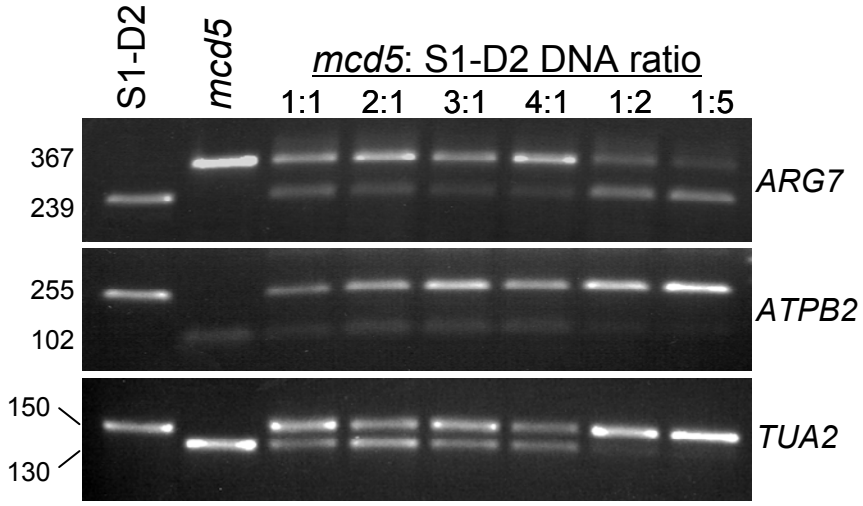


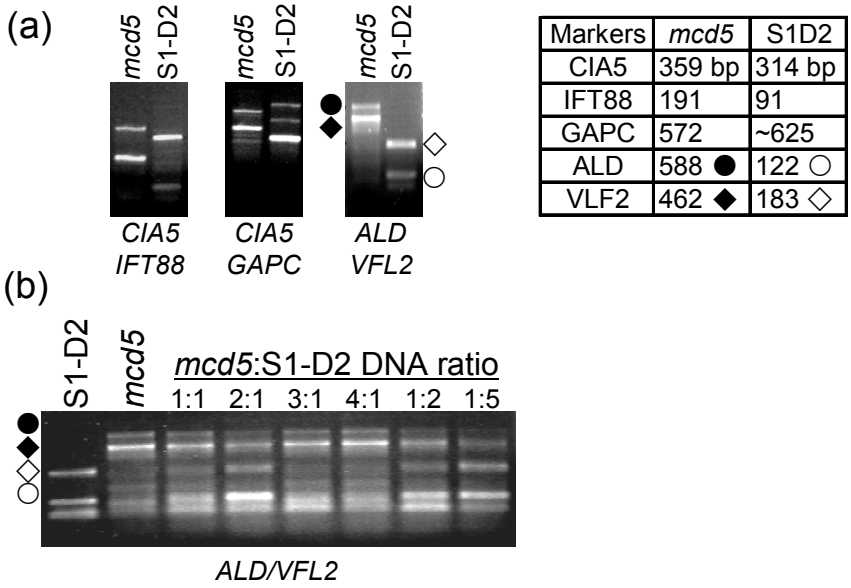
Figure 4.3. DNA of up to six progeny can be combined for BSA. The indicated ratios of *mcd5*: S1-D2 DNA were used to simulate bulked progeny. PCR products for the markers indicated at right were analyzed in 1% agarose gels, and product sizes in bp are shown at the left.

Out of the 40 primer sets tested, 34 primer sets yielded good results with 4:1 or 1:5 *mcd5*:S1-D2 DNA ratios. Five primer sets failed to generate

both parental bands at a 1:1 *mcd5*: S1-D2 ratio, thus in these cases, individual progeny must be analyzed. The remaining primer sets worked for up to 4 progeny. Since some difficulties could be resolved with gradient PCR or other adjustments, it might be possible with further optimization to implement BSA for additional markers. Assuming that 34 primer sets can be used with 5 bulked progeny, one with 4 bulked progeny, and 5 with single samples, the 2,000 PCR reactions required for the analysis of 40 markers (the number we tested for BSA) with 50 progeny (a typical mapping population) would be reduced to only 603. A complete set of amplification instructions and actual results can be found at the *Chlamydomonas* genome web site, http://www.chlamy.org.

To determine whether BSA caused a loss of mapping resolution due to incorrect interpretation of gel band intensities, DNA from *mcd4* mapping progeny were bulked in groups of four by Mabel Thomas. Four markers were blind tested using the bulked samples, and results were compared to those using single progeny. Two such comparisons are shown in Figure 4.4. Previous BSA experiments (*ARG7* and GP228 from Figure 4.3 and others) showed that when bulks contain parental DNAs at equal concentrations, the products generated do not appear to have equal intensities. In light of this, results from the bulked progeny were compared amongst themselves to determine which of the five ratios each lane represented. Differences between predicted (deduced from BSA gels) and actual (known from single progeny measurements) frequency of the *mcd4* allele ranged from 4% to 15% of the total progeny, with the average being 9%. In principle, this uncertainty could prevent detection of markers >20 cM distant from a gene of interest, although this was not the case for CNA45, which is 26 cM away from *mcd4* (Figure 4.4).

More experience scoring BSA, fewer progeny per PCR reaction, and/or the addition of a control lane with 1:1 ratio of *mcd4*/ S1-D2 would aid in reducing the margin of uncertainty to acceptable levels.



Markers	*mcd5*	S1D2
CIA5	359 bp	314 bp
IFT88	191	91
GAPC	572	~625
ALD	588 ●	122 ○
VLF2	462 ◆	183 ◇

Figure 4.4. BSA affects mapping resolution. DNAs from four randomly selected *mcd4* progeny were bulked per reaction and analyzed for the markers indicated on the left, with separation of PCR products in 1% agarose gels. Products were scored for number of progeny which contained the *mcd4* allele. Actual *mcd4* allele frequencies were determined by scoring individual unbulked samples (data not shown). Bulks from 48 out of the 53 progeny tested are shown here.

Combining two primer sets in one PCR reaction (duplexing) is a second method to reduce the amount of work needed for mapping. In choosing marker pairs for duplexing, annealing temperature and product size need to be considered. The four PCR products should be similar in size to avoid overly preferential amplification of the smaller bands, yet the products must be distinguishable by gel electrophoresis. Also, annealing temperatures should be close enough to prevent nonspecific amplification. Jocelyn Handley tested six paired markers were tested and 4 proved successful, as shown for 3 examples

in Figure 4.5a. When duplexing and BSA were combined, all 4 parental bands were visible at DNA ratios of 1:2 or 2:1 *mcd5*:S1-D2 for *ALD*/*VFL2*, whereas >3 DNAs did not offer sufficient sensitivity to detect the rare allele (Figure 4.5b). These results suggest that at least in some cases, combining duplexing and BSA holds promise for reducing labor while retaining resolution.

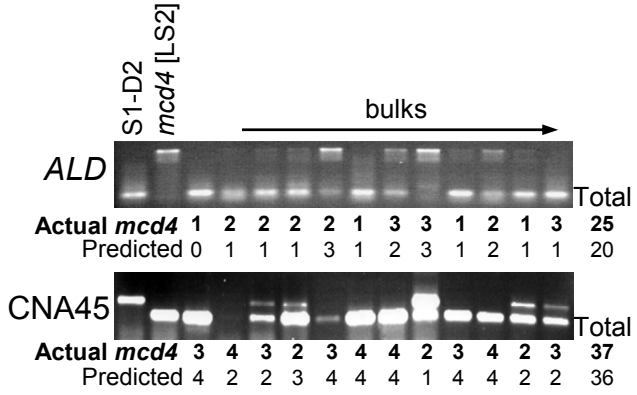


Figure 4.5. Markers can be duplexed and combined with BSA. (**a)** PCR products from three duplexed sets are shown, analyzed in 1% agarose gels. (**b)** The *ALD/VFL2* duplexed primers were used in combination with BSA. PCR products were analyzed in a 1% agarose gel. The Table at right shows expected product sizes for each marker. The symbols denoting the respective products from amplification of *ALD* or *VFL2* from *mcd5* or S1-D2 are used to mark product positions on the relevant gels. S1-D2 generates a second, artifactual band with *ALD* primers that migrates at approximately 75 bp (below 122 bp S1-D2 product).

From marker to genome: Combining mapping with genome data

The above data show the application of well-established molecular mapping technologies to *Chlamydomonas*. In plants, very high resolution mapping is needed prior to final identification of the gene of interest, because testing of numerous candidates by transformation is impractical. Since *Chlamydomonas* is easily transformed, the considerations are somewhat different. In the case of *mcd4*, using the mapping population to determine the

possible region that might contain *MCD4,* followed by a gene candidate search seemed to be ideal.

Since *mcd4* had been assigned to LG II, it was crossed to strains carrying two relevant phenotypic markers: *pf12,* a paralyzed flagella mutant (Frey *et al.*, 1997; McVittie, 1972), and *act1,* a cycloheximide resistant mutant (Sager and Tsubo, 1961) to aid in mapping. *PF12* is tightly linked to the RFLP marker GP225, while *ACT1* is 3.5 cM from *PF12* (Harris, 1989), and closely linked to CNA72 (Rochaix, personal communication). Progeny from a *mcd4* [LS2] mt+ by *pf12* mt- cross were scored for segregation of ability to swim and PS growth, and the results showed that *pf12* and *mcd4* were within 3.3 cM. In a cross using *act1* as the mt- parent, *act1* was estimated to be 9 cM from *mcd4.* When combined, phenotypic marker data suggested that *MCD4* is between *PF12* and CNA45.

In order to place *MCD4* more accurately, additional markers were required. CAPS, STS and SSR markers were all possible choices. CAPS markers can be made by converting existing RFLP markers, which entails generating primers flanking the polymorphic restriction site. If RFLP markers are not available in the region of interest, one can use a “brute force” method for any restriction site found in the nuclear genome sequence. This is because S1-D2 has 2.7 base substitutions per 100 bp compared to the laboratory strain (Kathir et al., 2003), leading to an average of one useable marker per 7 primer sets tested (for 6-base unique sequence restriction enzymes). When 15 primer sets were generated from scaffolds 5 and 79 (genome version 3), on which *PF12* and CNA45 are found, respectively, using this concept, five sets produced a PCR product only for the *C. reinhardtii* allele, two sets yielded PCR products that differed in size between *mcd4* and S1-D2 and could be

used as InDel markers, while three contained polymorphisms when digested and could be used as CAPS markers. When the S1-D2 products from two non-polymorphic markers were sequenced, one was converted to an STS marker after exhibiting enough sequence divergence to develop a S1-D2-specific primer. This demonstrates that the brute force method can be a practical way to generate useful primer sets.

A second method for marker generation exploits the collection of S1-D2 EST sequences to create STS markers. 1,616 S1-D2 sequences have been deposited in the NCBI database, and additional sequencing has been performed recently, generating an additional 45,312 sequences. Because approximately 165,000 cDNA sequences from the laboratory strain are present in dbEST[3], it is nearly always possible to align an S1-D2 EST with a corresponding laboratory strain sequence. Inspection of such alignments can directly suggest new markers based on sequence polymorphisms, and in fact 12 markers for scaffold 5 were generated in this manner.

A third method of generating markers utilizes the ubiquitous repeat elements of the *Chlamydomonas* genome, which can be converted into short sequence repeat (SSR) markers. The advantages of SSR markers are that they are available in all regions of the genome, and are already annotated (Repeat masker track). Repeat motifs range from a single base repeat up to a 10 bp repeat. Primer sets encompassing the annotated SSRs can be generated from the genome sequence. In most cases, SSR markers can be separated in 2 or 4% agarose gels, although some markers might require separation through a polyacrylamide gel or identification by capillary

[3] http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html

electrophoresis to distinguish differences of only 1 or 2 repeats between alleles.

Using the markers I generated as a template, our lab collaborated with Qi Sun[4], a Senior Research Associate at the Cornell Theory Center, to generate five new theoretical marker sets based on the S1-D2 EST sequencing and *Chlamydomonas* genome version 2, to significantly expand the available molecular markers. Each of the five sets represented a different type of marker which included SSRs, InDels, STS, CAPS and dCAPS (Derived Cleavable Amplified Polymorphic Sequences). dCAPS markers utilize SNPs and primers containing a single base mutation to generate a restriction site in the PCR product from one of the strains. After digestion, one PCR product will be 20-25 bp shorter than the other, revealing the polymorphism when the digests are analyzed in 4% agarose gels. When April Mevin tested subsets of each of these markers, she found that 62% produced viable markers. Extrapolating that success rate to the total number of markers generated suggests that 8,775 useable markers were generated. These, in addition to previously tested markers, gives an estimate of 9,300 markers which will perform correctly in mapping experiments using S1-D2 polymorphisms with laboratory strains, significantly expanding the tools available for mapping a given gene.

[4]To whom primer generation correspondence should be addressed; email: qisun@tc.cornell.edu.

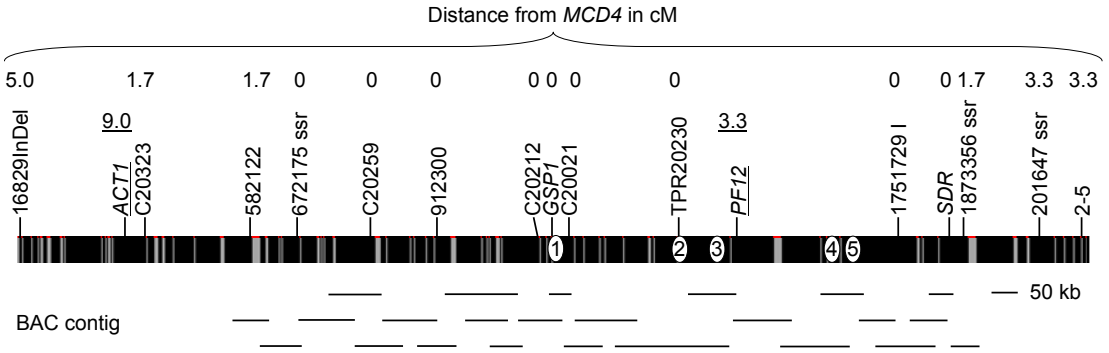


Figure 4.6. Summary of mapping data, showing the region of LG II surrounding *mcd4*. Available sequence from scaffold 5 (genome version 3) is represented by filled portions, with grey representing gaps in the sequence. The names of the markers are indicated above the black and grey bar. With the exception of *ACT1* and *PF12*, which are phenotypic markers, the markers are PCR-based. Distances between selected markers and *mcd4* are given in cM above the map. A BAC contig developed for the MCD4 region is shown to illustrate typical coverage where such contigs have been anchored. The candidate genes are represented by white ovals and the number corresponds to those in Table 4.2.

The generation of the theoretical marker set aided in fine mapping, and revealed that *mcd4* resides in a 1,300 kb region between markers 5821222 and 1873356 ssr (Figure 4.6). Since fine mapping was completed, gene isolation naturally became the focus. Several methods were used in this endeavor, the first being BAC transformation, which was performed concurrently with the fine mapping. It was desirable to generate BAC contigs to scaffold 5 to facilitate complementation of *mcd4*. Several BAC libraries have been constructed for *Chlamydomonas*, two of which are available through the Clemson Genomics Institute[5]. In addition, BAC contigs[6] have been assembled for most of the STS and RFLP markers. A 41-BAC contig already exists for scaffold 5 covering approximately 800 kb (Nguyen *et al.*, 2005). This was extended using the genome sequence to perform a virtual BAC walk. In this

[5] http://www.genome.clemson.edu/groups/bac/
[6] http://www.biology.duke.edu/chlamy_genome/BAC/index.html

method, a scaffold end or segment is compared using BLAST to the BAC end database, and any BACs containing that sequence are identified. Then, the other end sequence associated with that BAC is used in BLAST searches with the genome sequence. BAC contigs can also be elongated by the traditional method of hybridization of single-copy BAC end probes to a filter representing the entire BAC library, followed by informatic analysis as described just above. This would be the preferred method if a BAC end lands in a hole between scaffolds (i.e. is not represented in available genome sequence). Eventually such contigs will rejoin the genome sequence, in effect bridging two scaffolds. The virtual BAC walk alone sufficed to extend the existing BAC contig 500 kb to span the entire area between 5821222 and 1873356 ssr markers. A combination of the existing BAC scaffold and virtual BAC scaffold is shown in Figure 4.6.

Our strategy for complementation is generally applicable for a loss-of-function restorer of photosynthesis, or any other situation where direct selection of complementation is not practicable. This was to co-transform each BAC into *mcd4* [LS2] with pCB797 (Schroda *et al.*, 2002), a construct which confers resistance to the antibiotic zeocin. Once selected on zeocin-containing medium, transformants were screened for complementation of *mcd4* by the WT allele. Since *mcd4* confers a PS+ phenotype in the LS2 chloroplast background (Rymarquis *et al.*, 2006), expression of a WT copy of *MCD4* will disrupt photosynthesis by destabilizing *petD*-LS2 mRNA. Zeocin resistant transformants were replica-plated onto TAP and medium lacking acetate to determine their photosynthetic growth phenotype. Any PS- and zeocin-resistant transformants would be candidates for containing BAC DNA including *MCD4.* These PS- strains were then underwent the secondary screening

described below to verify that their phenotypes matched that of the *MCD4*/*mcd4* [LS2] diploids (Rymarquis *et al.*, 2006).

Several factors need to be taken into consideration in deriving a number of zeocin-resistant transformants to be screened. One is the efficiency of co-transformation of pCB797 (5 kb) and the much larger BAC clones (average insert size of 75 kb). This efficiency was estimated by a proof-of-concept experiment where the plasmids pARG7.8 (15 kb) was co-transformed with pCB797 into the *arg7* mutant strain CC-3680. This strain requires added arginine for growth, and thus is unable to grow on medium lacking arginine. Transformants were first selected for zeocin resistance, then plated on medium lacking arginine to determine which incorporated a functional *ARG7* gene. We found that 14% of the zeocin-resistant transformants also has an Arg+ phenotype. Since in the *mcd4* complementation experiment we expected a certain background of PS- colonies (see below), we decided to calculate the number of colonies to screen in order to obtain at least 5 heterozygous *MCD4/mcd4* colonies, which was presumed to exceed the PS- background. Since the BACs are on average 5-fold larger than the pARG7.8, the following equation was derived: [Number of colonies needed = 5*BAC size/(15 kb*0.14)]. With a BAC size of 75 kb, 180 transformants would need to be screened in order to find five *MCD4* complementation events. With larger BACs, more colonies would be required, but smaller BACs would presumably yield a greater number of useful events.

Other factors to consider in transformation is the length of DNA incorporated, and the silencing of introduced genes. *ARG7* is a rather large gene, with a total size of 8 kb, and thus a contiguous integration of this region is required for complementation. Should *MCD4* (or another gene of interest)

be significantly smaller, more frequent complementation would be observed, or vice versa for larger genes. In other words, many of the zeocin-resistant colonies may have incorporated non-functional segments of pARG7.8, but only insertion of functional *ARG7* or *MCD4* are of interest when complementing mutations. Another issue is gene silencing, which is commonly observed when transgenes are not selected in *Chlamydomonas* (Schroda *et al.*, 2002). Given that the BACs contain native (non-engineered) DNA, this might not prove to be a significant factor. Finally, one must consider the background of PS- transformants, since any introduced DNA is a potential insertional mutagen, and indeed *ARG7* has been used to generate numerous PS- mutants in gene tagging approaches (Lown *et al.*, 2001). Based on our actual experimental data (Table 4.1), a background of approximately 4% PS- transformants can be anticipated. If 5 complemented PS- transformants were obtained, an approximate difference of 9 vs. 4 PS- colonies from a given co-transformation would be reproducibly observed. Since this difference is not large, all PS- transformants were screened for lower SUIV levels, which is the *petD* gene product, by monitoring the PetA protein levels. PetA is an indicator of the SUIV levels because in the absence of SUIV, PetA is present at only 10% of the WT level (Kuras and Wollman, 1994). Any transformants that showed 10% PetA levels were assayed for reduced *petD*-LS2 RNA and loss of the altered *atpA* or *psbB* mRNA phenotypes found in *mcd4* (Rymarquis *et al.*, 2006). As of yet, no transformant has passed all four screens.

Table 4.1 Results from BAC transformations. Transformants were screened for photosynthesis, PetA protein levels and *petD* mRNA levels.

BAC	Transformants screened	PS- colonies	Reduced PetA	Reduced *petD*
3N20	240	3	4	0
8F3	940	5	0	0
8O4	102	2	0	0
9L10	240	1	0	0
11J2	180	2	0	0
23P12	205	5	0	0
38J17	240	1	0	0
39c12	ND			
13G22	185	3	0	0
21A6	240	0	0	0
5p20	240	1	0	0
1d19	54	0	0	0
39m2	91	0	0	0
33k1	258	16	0	0
19f12	320	4	0	0
34f11	235	19	0	0
3b19	240	1	0	0
28n4	240	1	0	0

Candidate gene analysis

Since the BAC transformation strategy did not yield *MCD4*, several other methods are possible. These include cloning the candidate genes and transforming single genes into *mcd4* [LS2] to disrupt photosynthesis, cloning the candidate gene from *mcd4* for sequencing to look for mutations, RNAi disruption of the candidate genes' expression, and generating a third mapping population to try to reduce the size of the potential *MCD4* region by finding new recombination events. With the completion of mapping and recent release version 3 of the genome, candidate gene searches were performed on the *MCD4* region. Given that *mcd4* suppresses an RNA instability phenotype and is recessive, one possibility was a loss-of-function in a chloroplast-targeted ribonuclease (reviewed in Bollenbach et al., 2004), however no such genes were found in this region. Among the other possibilities are tetratricopeptide

repeat (TPR) proteins and pentatricopeptide repeat (PPR) proteins, which are often found as post-transcriptional regulators of organelle gene expression, including in *Chlamydomonas* chloroplasts (Boudreau *et al.*, 2000; Lurin *et al.*, 2004; reviewed in Small and Peeters, 2000), as well as RNA-binding proteins.

Five candidate genes were found and are labeled 1-5 in Figure 4.6. Gene candidates 1 and 2 are TPR-containing predicted proteins. Candidate 1 is predicted to be chloroplast-localized based on N-terminal sequence motifs while candidate 2 is not, however there are several reasons why this could be misleading (see Future Perspectives). Candidates 3 and 5 have RNA-recognition motifs (RRM), but neither is predicted to be chloroplast-targeted. In addition, candidate 3 has an La motif, which is most commonly associated with nuclear RNA transcription and processing (Kenan and Keene, 2004), thus it is less likely to be *MCD4,* but since it has RRM motifs, it was not completely discounted. Candidate gene 5 is a hypothetical chloroplast targeted protein, which was considered because MCD1, which also affects *petD* stability, does not contain any known motifs (Murakami *et al.*, 2005). All of the candidate genes have EST support indicating that they represent expressed genes.

Since candidates 4 and 5 are between *PF12* and 1873356 ssr (Figure 4.6), the region to which the combined molecular and phenotypic data localized *MCD4,* they were analyzed first. Taking into account that the *mcd4* mutation could be in the promoter or regulatory elements, the combined coding regions and 2 kb upstream and downstream totaled 12 kb for each gene. Since direct amplification of 12 kb by PCR can be difficult, the genes were amplified as two separate pieces sharing a unique restriction site. The two products were then joined at the unique site to generate a full-length candidate gene. Primers were optimized using BAC DNA as the templates.

This generates the WT copy of the gene, and proved relatively facile. When the same conditions were applied with *mcd4* DNA as the template, a high abundance of product could not be generated even when using nested PCR. Thus, the transformation of WT copies of the candidate genes into *mcd4* [LS2] appeared to be the easiest method to test them. Candidate gene 5 was co-transformed with pCB797 into *mcd4* [LS2], but all 240 transformants were PS+. This suggests that *mcd4* does not represent a mutation in this hypothetical protein.

In conjunction with candidate gene transformation, an approach of RNAi knockout is in progress, since it was previously shown to be highly effective in reducing PNPase expression (S. Zimmer, pers. comm.). If the candidate gene is *MCD4*, reducing its expression by RNAi should restore photosynthesis to LS2, assuming *mcd4* is a loss-of-function mutation. As a control, the RNAi constructs will be transformed into WT to ensure that reduced expression of the candidate gene does not disrupt photosynthesis. To date, Jocelyn Handley has inserted portions of candidate genes 3 and 5 into the RNAi construct, and analysis of the transformants is underway.

A summary of candidate gene analysis is shown in Table 4.2. Candidate 5 is unlikely to be *MCD4,* since none of the 940 BAC transformants or 240 candidate gene transformants conferred a PS- phenotype on *mcd4* [LS2]. Candidate 4 has a low probability of being *MCD4* because it resides on the same BAC as candidate 5, and there should have been 25 transformants receiving a functional *MCD4,* out of the 940 transformants screened using the equation for BAC transformation efficiency shown above. Since there were only 5 non-photosynthetic transformants, and none had reduced PetA protein levels, it is likely that *MCD4* is not present on BAC 83F. Since candidates 4

and 5 were the only ones between *PF12* and 1873356 ssr, it could mean that either *MCD4* is between *PF12* and 1873356 ssr and does not resemble a protein with RNA binding, TPR, or PPR motifs, or that the phenotypic mapping was misleading and that *MCD4* is not between *PF12* and 1873356 ssr, but between *PF12* and *ACT1.* To determine if the second is true, candidates 1-3 are being analyzed. If they too are not *MCD4*, a larger mapping population will be generated in an attempt to narrow the 1,300 kb *MCD4* region.

Table 4.2 Summary of candidate gene analysis

Number	Name	Description	BAC	CP Targeting[a]	RNAi[b]	Cloned	Gene Transformation[c]
1	C20214	TPR	34F11	Y	ND	ND	ND
2	C20230	TPR	8O4	N	ND	ND	ND
3	C20239	RRM motif	39C12	N	Y	ND	ND
4	C20067	RRM motif	8F3	N	ND	Y	ND
5	C20072	hypothetical protein	8F3	Y	Y	Y	All PS+

[a] Chloroplast targeting as determine by GENOPLANTE™ PREDOTAR and iPSORT
[b] Whether the candidate genes have been cloned into the RNAi vector.
[c]Results from transformation of a WT copy of the gene into *mcd4* [LS2]. ND= not done.

FUTURE PERSPECTIVES

In the past year, many tools to streamline map-based cloning such as the updated molecular map (Figure 4.1), genome release version 3, new S1-D2 ESTs, new *Chlamydomonas* EST contigs, the theoretical marker set and BAC contigs have increase the attractiveness of *Chlamydomonas*. An overview of the map-based cloning process from mutant to gene, as applied to *Chlamydomonas*, is shown in Figure 4.7. Once a linked marker has been identified, the genome sequence associated with that marker is accessible. Although the nuclear genome is not complete, 120 Mb has been sequenced and assembled into 1,557 scaffolds; roughly half of this in 24 scaffolds of at least 1.6 Mb in length in genome version 3. Genome finishing is underway,

and completion in 2006 or 2007 is the current goal. Nuclear genome annotation has been performed on an automated basis and has been supplemented by manual annotation. This, combined with EST data displayed on the genome browser, allow for reasonably efficacious candidate gene searches.

Once a candidate is found, the gene sequence may need to be confirmed before complementation proceeds. This is particularly true for poorly annotated genes and/or those lacking EST support, or where sequence gaps are present. In addition, the gene prediction programs used have a significant error frequency, both in defining splice sites and in the recognition of mitochondrial and chloroplast transit peptides. Annotation has been aided by the combined EST and protein homolog data displayed on the genome browser, but only about 60% of the predicted genes have EST coverage and 42% have protein homologies (Li et al., 2003). These considerations, of course, are not unique to *Chlamydomonas*.

With existing tools, rough mapping in *Chlamydomonas* is straightforward. A kit of 67 PCR-based marker primers, allowing assignment of a mutant locus to a chromosome arm, is available to the community via the Stern Lab. Using 50 progeny from a single cross and the marker kit, the entire genome can be screened at 20-30 cM resolution in less than three months. BSA and duplexing of markers can decrease the time and effort needed for mapping, while reducing the cost. One can also anticipate the completion of the genome, facilitating candidate gene searches and determining which BACs should be used for complementation. This in addition to the 9,300 markers recently generated will allow the mapping of interesting mutations to be a routine endeavor.

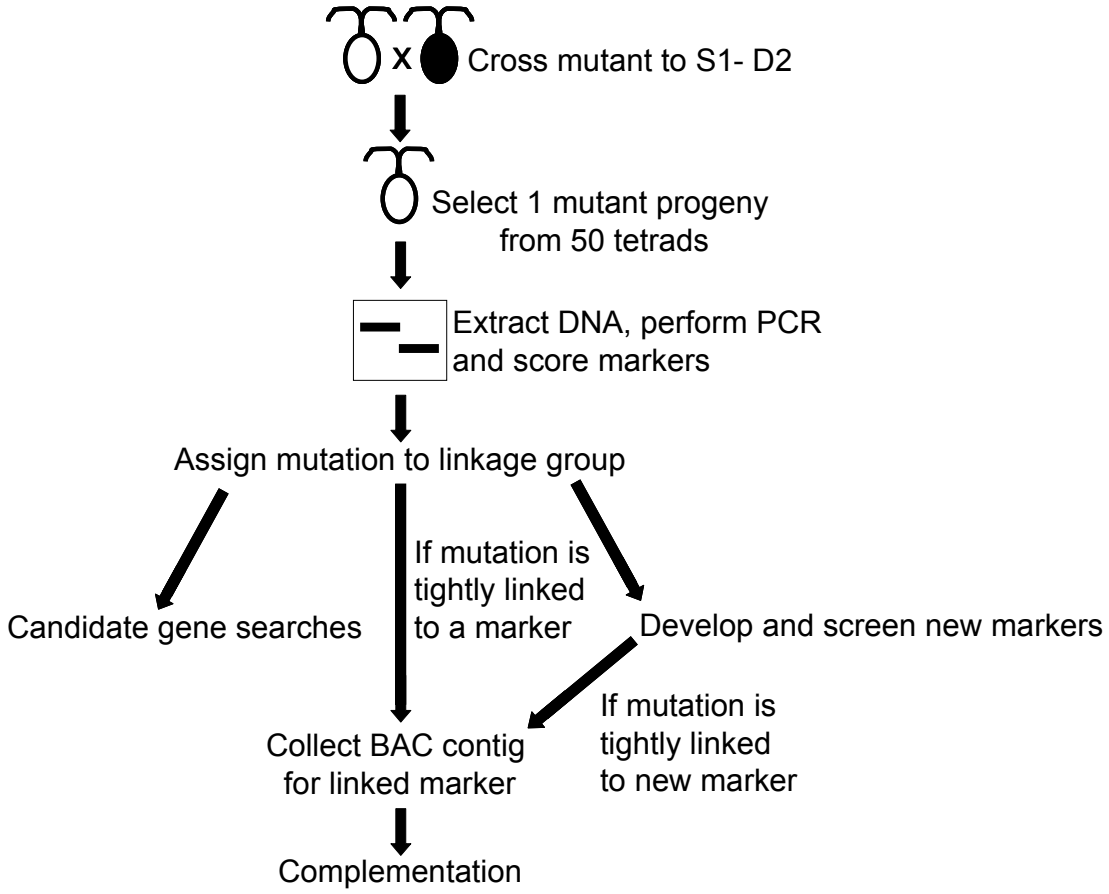


Figure 4.7. Flow chart for map-based cloning in *Chlamydomonas*. The scheme begins with a cross between the strain carrying the mutation of interest, and S1-D2. In the second step, one clone showing the mutant phenotype is selected from each tetrad, although in principle the S1-D2 (non-mutant) allele could be mapped by selecting one WT clone per tetrad. The initial mapping phase identifies a linkage group and chromosome arm, and depending on location, phenotype, and linkage, several subsequent steps may be pursued.

METHODS

Strains, media and genetic crosses

The *C. reinhardtii* laboratory strains *mcd4* [LS2] mt+ or *mcd5* [LS6] mt+ and S1-D2 mt- (CC-2290) were used as parents. Before crossing to S1-D2, backcrosses to a WT laboratory strain were conducted to ensure that the *mcd4* and *mcd5* phenotypes were caused by single, nuclear mutations. Then, *mcd4* and *mcd5* were crossed to S1-D2 and progeny were dissected as described previously (Levine and Ebersold, 1960). Progeny were tested for photoautotrophic growth by plating on medium lacking acetate lacking acetate (Harris, 1989), and otherwise maintained on TAP (Gorman and Levine, 1965) under 23 hours light and 1 hour dark at 25°C. One photosynthetic progeny from each of 63 tetrads was used to create the *mcd4* mapping population; the current *mcd5* population is 64 individuals.

For crosses to phenotypic markers, *mcd4* [LS2] mt+ was crossed to *pf12* mt- (CC-610) or *act1* mt- (CC-2953). Progeny were dissected and assayed for photosynthesis and swimming ability (*pf12* cross) or cycloheximide resistance (*act1* cross). For the *pf12* cross, swimming ability was determined by light microscopy at 100X magnification. The *act1* genotype was determined by growth on TAP medium containing 10 ng/ml cycloheximide. Genetic distance was determined as described (Harris, 1989).

DNA preparation and PCR conditions

Total DNA from the mapping population was prepared using a protocol adapted from Steve Pollock[7]. A toothpick of cells from a TAP plate was resuspended in 50 µl of 10 mM NaEDTA in a 1.5 ml microfuge tube. The tube

[7] http://www.biology.duke.edu/chlamy/methods/quick_pcr.html

was vortexed and incubated at 100°C for 5 min. Then, the tube was centrifuged at 12,000 rpm for 1 min. The supernatant was retained as the DNA sample, and its concentration was measured on a spectrophotometer, then diluted to 20 ng/µl.

Markers used were reported previously and are listed http://www.chlamy.org/kit.html (Kathir et al., 2003; Torjek et al., 2003). PCR conditions were 8.5% glycerol, 0.83% formamide, 7 µl GoTaq buffer (Promega), 1 µl of each primer (10 mM stock), 0.5 µl 10 mM dNTPs, 20 ng *Chlamydomonas* DNA and 0.25 µl GoTaq (Promega) in a final volume of 30 µl. The PCR program began with denaturation at 94°C for 2 min, followed by 40 cycles of 94°C for 1 min, annealing for 1 min at primer-specific temperatures, and 72°C for 1 min. A final extension was performed at 72°C for 10 min. PCR products were analyzed in 1% or 3% agarose gels, and visualized using ethidium bromide.

For BSA, the *mcd5* and S1-D2 DNAs were combined so that total amount of DNA in the PCR reaction remained at 20 ng. Ratios ranged from 4:1 to 1:5 *mcd5:*S1-D2. PCR reactions were performed as specified above. For duplexing reactions, markers were chosen that had sizes which were distinguishable in a 3% agarose gel and had similar annealing temperatures. When the total DNA amount was not held constant, reactions containing DNA from usually >3 progeny did not yield any PCR products. Thus, the total DNA amount is a key parameter.

Calculating co-transformation efficiency and BAC transformation

arg7 (CC-3680) cells were grown in TARG (TAP + 1 µg/ml arginine) liquid until they entered stationary phase. One ml of this culture was used to

inoculate 100 ml of TARG. After the culture had reached a density of 5 X 10^7 cells/ml, cells were harvested by centrifugation, resuspended in 2 ml autolysin, and placed in high light. After 1 hr, cells were collected by centrifugation, resuspended in 500 µl TAP containing 60 mM sucrose, and transformation was carried out as preciously described (Shimogawara *et al.*, 1998). One µg of pCB797 (Schroda *et al.*, 2002) and 4 µg of pARG7.8 (Debuchy *et al.*, 1989) were used per transformation. After a 24 hr recovery period, cells were collected by centrifugation, and resuspended in 400 µl TAP containing 60 mM sucrose instead of the corn starch mixture described by Shinogawara et al. (1998). And spread on TARG containing 5 µg/ml zeocin. After one week transformants were replica plated onto TARG and medium lacking acetate. Transformants that grew on both media were presumed to contain a functional *ARG7*.

BAC DNA was prepared using PureYield™ Plasmid Midiprep System (Promega, Madison, WI, USA). The BAC transformations were carried out in a manner similar to the *ARG7* transformation except TAP medium was used instead of TARG. Any transformant that did not grow on medium lacking acetate was tested for PetA protein levels, *petD* mRNA levels and *atpA* RNA pattern as previously described (Rymarquis *et al.*, 2006).

Candidate gene amplification

PCR conditions were 4 µl betaine, 4 µl High GC iProof buffer (Bio-Rad, Hercules, CA, USA), 1 µl of each primer (10 mM stock), 2 µl 10 mM dNTPs, 25 ng BAC DNA and 0.2 µl iProof High-Fidelity DNA Polymerase (Bio-Rad, Hercules, CA, USA) in a final volume of 20 µl. The PCR program began with denaturation at 98°C for 3 min, followed by 35 cycles of 98°C for 10 sec.,

annealing for 15 sec. at primer-specific temperatures, and 72°C for 3 min. A final extension was performed at 72°C for 10 min. PCR products were gel extracted. To generate A-tails on the PCR product, 7 µl of gel purified PCR product was incubated with 1 µl of 2 mM ATP, 1 µl 10X Taq polymerse buffer, and 1 µl Taq polymerase buffer (Promega, Madison, WI, USA) for 30 min at 70°C. The TOPO TA cloning kit was used to clone the product according to manufactures instructions overnight at 4°C. The reaction was then transformed into electrocompetent DH10α and grown at 28°C. The two cloned PCR products for each gene candidate were then ligated together using the unique internal site and a unique site in the vector. The list of primers and restriction enzymes used for candidate gene cloning are shown in Table 4.3.

Table 4.3. Primers and restriction enzymes used in candidate gene cloning

Candidate gene	5' Product	3' product	Enzymes[a]	
			5'	3'
C20214	acacgtgagagcgaagagtg cggtagacccttggtgatgag	agcagctcataaagctcaaggtac cgaagaccagggataatactgttg	*Bgl*II	*Bgl*II
C20067	gttttaggcaggcgagcagcatgagacttg gcaacccacggccggacctgtagc	gcaaggccgaccaggacgcagatgacgac actagtcgccgatgtgaacgcccagacc	*Spe*I *Xho*I	*Xho*I *Xba*I
C20072	agtccctatcaccctggatatg agacaggaggtgtcggtttac	cgatcgctcacgcttccttc tccgaggcagatctatcatagc	*Nhe*I *Sca*I	*Nhe*I *Sca*I

[a]Restriction enzymes used to digest 5' PCR product clone (5') or 3' PCR product clone (3')

REFERENCES

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

Bennoun, P. and Béal, D. (1997) Screening algal mutant colonies with altered thylakoid electrochemical gradient through fluorescence and delayed luminescence digital imaging. *Photosynthesis Res.*, **51**, 161-165.

Bernd, K.K. and Kohorn, B.D. (1998) *Tip* loci: six *Chlamydomonas* nuclear suppressors that permit the translocation of proteins with mutant thylakoid signal sequences. *Genetics*, **149**, 1293-1301.

Bollenbach, T.J., Schuster, G. and Stern, D.B. (2004) Cooperation of endo- and exoribonucleases in chloroplast mRNA turnover. *Prog. Nucleic Acid Res. Mol. Biol.*, **78**, 305-337.

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

Chen, M., Presting, G., Barbazuk, W.B., Goicoechea, J.L., Blackmon, B., Fang, G., Kim, H., Frisch, D., Yu, Y., Sun, S., Higingbottom, S., Phimphilai, J., Phimphilai, D., Thurmond, S., Gaudette, B., Li, P., Liu, J., Hatfield, J., Main, D., Farrar, K., Henderson, C., Barnett, L., Costa, R., Williams, B., Walser, S., Atkins, M., Hall, C., Budiman, M.A., Tomkins, J.P., Luo, M., Bancroft, I., Salse, J., Regad, F., Mohapatra, T., Singh, N.K., Tyagi, A.K., Soderlund, C., Dean, R.A. and Wing, R.A. (2002) An integrated physical and genetic map of the rice genome. *Plant Cell*, **14**, 537-545.

Coe, E., Cone, K., McMullen, M., Chen, S.S., Davis, G., Gardiner, J., Liscum, E., Polacco, M., Paterson, A., Sanchez-Villeda, H., Soderlund, C. and Wing, R. (2002) Access to the maize genome: an integrated physical and genetic map. *Plant Physiol.*, **128**, 9-12.

Cone, K.C., McMullen, M.D., Bi, I.V., Davis, G.L., Yim, Y.S., Gardiner, J.M., Polacco, M.L., Sanchez-Villeda, H., Fang, Z., Schroeder, S.G., Havermann, S.A., Bowers, J.E., Paterson, A.H., Soderlund, C.A., Engler, F.W., Wing, R.A. and Coe, E.H., Jr. (2002) Genetic, physical, and informatics resources for maize. On the road to an integrated map. *Plant Physiol.*, **130**, 1598-1605.

Dauvillee, D., Stampacchia, O., Girard-Bascou, J. and Rochaix, J.D. (2003) Tab2 is a novel conserved RNA binding protein required for translation of the chloroplast *psaB* mRNA. *EMBO J.*, **22**, 6378-6388.

Debuchy, R., Purton, S. and Rochaix, J.-D. (1989) The argininosuccinate lyase gene of *Chlamydomonas reinhardtii*: an important tool for nuclear transformation and for correlating the gentic and molecular maps of the ARG7 locus. *EMBO J.*, **8**, 2803-2809.

Dent, R.M., Han, M. and Niyogi, K.K. (2001) Functional genomics of plant photosynthesis in the fast lane using *Chlamydomonas reinhardtii*. *Trends Plant Sci.*, **6**, 364-371.

Esposito, D., Higgs, D.C., Drager, R.G., Stern, D.B. and Girard-Bascou, J. (2001) A nucleus-encoded suppressor defines a new factor which can promote *petD* mRNA stability in the chloroplast of *Chlamydomonas reinhardtii*. *Curr. Genet.*, **39**, 40-48.

Frey, E., Brokaw, C.J. and Omoto, C.K. (1997) Reactivation at low ATP distinguishes among classes of paralyzed flagella mutants. *Cell Motil Cytoskeleton*, **38**, 91-99.

Garcia-Hernandez, M., Berardini, T.Z., Chen, G., Crist, D., Doyle, A., Huala, E., Knee, E., Lambrecht, M., Miller, N., Mueller, L.A., Mundodi, S., Reiser, L., Rhee, S.Y., Scholl, R., Tacklind, J., Weems, D.C., Wu, Y., Xu, I., Yoo, D., Yoon, J. and Zhang, P. (2002) TAIR: a resource for integrated Arabidopsis data. *Funct. Integr. Genomics*, **2**, 239-253.

Girard-Bascou, J., Pierre, Y. and Drapier, D. (1992) A nuclear mutation affects the synthesis of the chloroplast *psbA* gene product in *Chlamydomonas reinhardtii*. *Curr. Genet.*, **22**, 47-52.

Goodenough, U.W. (1992) Green yeast. *Cell*, **70**, 533-538.

Gorman, D.S. and Levine, R.P. (1965) Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardii*. *Proc. Natl. Acad. Sci. USA*, **54**, 1665-1669.

Gross, C.H., Ranum, L.P.W. and Lefebvre, P.A. (1988) Extensive restriction fragment length polymorphisms in a new isolate of *Chlamydomonas reinhardtii*. *Curr. Genet.*, **13**, 503-508.

Grossman, A.R. (2000) *Chlamydomonas reinhardtii* and photosynthesis: genetics to genomics. *Curr. Opin. Plant Biol.*, **3**, 132-137.

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

Gumpel, N.J., Ralley, L., Girard-Bascou, J., Wollman, F.A., Nugent, J.H. and Purton, S. (1995) Nuclear mutants of *Chlamydomonas reinhardtii* defective in the biogenesis of the cytochrome b_6f complex. *Plant Mol. Biol.*, **29**, 921-932.

Harris, E.H. (1989) *The Chlamydomonas Sourcebook: A comprehensive guide to biology and laboratory use*. San Diego: Academic Press.

Higgs, D.C., Shapiro, R.S., Kindle, K.L. and Stern, D.B. (1999) Small *cis*-acting sequences that specify secondary structures in a chloroplast mRNA are essential for RNA stability and translation. *Mol. Cell. Biol.*, **19**, 8479-8491.

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

Kenan, D.J. and Keene, J.D. (2004) La gets its wings. *Nat. Struct. Mol. Biol.*, **11**, 303-305.

Kuras, R. and Wollman, F.-A. (1994) The assembly of cytochrome b_6/f complexes: an approach using genetic transformation of the green alga *Chlamydomonas reinhardtii*. *EMBO J.*, **13**, 1019-1027.

Levine, R.P. and Ebersold, W.T. (1960) The genetics and cytology of *Chlamydomonas*. *Annu. Rev. Microbiol.*, **14**, 197-216.

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

Li, F., Holloway, S.P., Lee, J. and Herrin, D.L. (2002) Nuclear genes that promote splicing of group I introns in the chloroplast 23S rRNA and *psbA* genes in *Chlamydomonas reinhardtii*. *Plant J.*, **32**, 467-480.

Li, J.B., Lin, S., Jia, H., Wu, H., Roe, B.A., Kulp, D., Stormo, G.D. and Dutcher, S.K. (2003) Analysis of *Chlamydomonas reinhardtii* genome structure using large-scale sequencing of regions on Linkage Groups I and III. *J. Euk. Microbiol.*, **50**, 145-155.

Lown, F.J., Watson, A.T. and Purton, S. (2001) *Chlamydomonas* nuclear mutants that fail to assemble respiratory or photosynthetic electron transfer complexes. *Biochem. Soc. Trans.*, **29**, 452-455.

Lukowitz, W., Gillmor, C.S. and Scheible, W.R. (2000) Positional cloning in *Arabidopsis*. Why it feels good to have a genome initiative working for you. *Plant Physiol.*, **123**, 795-805.

Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M.L., Mireau, H., Peeters, N., Renou, J.P., Szurek, B., Taconnat, L. and Small, I. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, **16**, 2089-2103.

McVittie, A.C. (1972) Flagellum mutants of *Chlamydomonas reinhardtii*. *J. Gen. Microbiol.*, **71**, 525–540.

Murakami, S., Kuehnle, K. and Stern, D.B. (2005) A spontaneous tRNA suppressor of a mutation in the *Chlamydomonas reinhardtii* nuclear *MCD1* gene required for stability of the chloroplast *petD* mRNA. *Nucleic Acids Res.*, **33**, 3372-3380.

Nguyen, R.L., Tam, L.W. and Lefebvre, P.A. (2005) The *LF1* gene of *Chlamydomonas reinhardtii* encodes a novel protein required for flagellar length control. *Genetics*, **169**, 1415-1424.

Nickelsen, J. (2000) Mutations at three different nuclear loci of *Chlamydomonas* suppress a defect in chloroplast *psbD* mRNA accumulation. *Curr. Genet.*, **37**, 136-142.

Niyogi, K.K., Bjorkman, O. and Grossman, A.R. (1997) *Chlamydomonas* xanthophyll cycle mutants identified by video imaging of chlorophyll fluorescence quenching. *Plant Cell*, **9**, 1369-1380.

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

Rymarquis, L., Higgs, D. and Stern, D. (2006) Nuclear suppressors define three factos that participate in both 5' and 3' end processing of mRNAs in *Chlamydomons* chloroplasts. *Plant J.*, In Press.

Sager, R. and Tsubo, Y. (1961) Genetic analysis of streptomycin resistance and dependence in Chlamydomonas. *Z. Vererbungsl.*, **92**, 430-438.

Schmidt, R., West, J., Love, K., Lenehan, Z., Lister, C., Thompson, H., Bouchez, D. and Dean, C. (1995) Physical map and organization of *Arabidopsis thaliana* chromosome 4. *Science*, **270**, 480-483.

Schroda, M., Beck, C.F. and Vallon, O. (2002) Sequence elements within an HSP70 promoter counteract transcriptional transgene silencing in *Chlamydomonas*. *Plant J.*, **31**, 445-455.

Shimogawara, K., Fujiwara, S., Grossman, A. and Usuda, H. (1998) High-efficiency transformation of *Chlamydomonas reinhardtii* by electroporation. *Genetics*, **148**, 1821-1828.

Small, I.D. and Peeters, N. (2000) The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.*, **25**, 46-47.

Torjek, O., Berger, D., Meyer, R.C., Mussig, C., Schmid, K.J., Rosleff Sorensen, T., Weisshaar, B., Mitchell-Olds, T. and Altmann, T. (2003) Establishment of a high-efficiency SNP-based framework marker set for *Arabidopsis*. *Plant J.*, **36**, 122-140.

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*. *Proc. Natl. Acad. Sci. USA*, **97**, 14813-14818.

Vysotskaia, V.S., Curtis, D.E., Voinov, A.V., Kathir, P., Silflow, C.D. and Lefebvre, P.A. (2001) Development and characterization of genome-wide single nucleotide polymorphism markers in the green alga *Chlamydomonas reinhardtii*. *Plant Physiol.*, **127**, 386-389.

CHAPTER 5.
FUTURE PERSPECTIVE

Chloroplast gene expression is regulated at many levels from transcription to protein degradation. RNA maturation is one level of this regulation, affecting transcript stability and translation (Herrin and Nickelsen, 2004). RNA maturation involves the binding of nucleus-encoded gene-specific factors to specific structures and sequences in the RNA to direct endonucleolytic cleavage and trimming by exonucleases. This study identified three genes, *MCD3*, *MCD4*, and *MCD5* that are involved in both the maturation of RNA and at least in the case of MCD4, its degradation. The RNA patterns produced in these mutants when 32 chloroplast transcripts were analyzed, as well as reporter gene constructs containing the *atpB* 3' processing determinant. These indicate that MCD3, MCD4, and MCD5 are most likely involved in endonucleolytic cleavage potentially involved in RNA maturation and degradation.

Many possible avenues of study are available while cloning efforts are underway. These include crosses to generate double *mcd* mutants to test for redundancy or potential interactions (see appendix for preliminary results), and crosses to *crp3,* which appears to contain defects in 3' to 5' exonuclease activity (Levy et al., 1997; Levy et al., 1999). Other possibility include determining whether *mcd3/4/5* can re-stabilize other *Chlamydomonas* chloroplast mRNAs where mutations in the 5' UTR have destabilized them (Nickelsen *et al.*, 1999) or determining whether any of *mcd3/4/5* are allelic to, or regulate, suppressors isolated in an analogous screen using *psbD* (Nickelsen, 2000). Analysis of all of these crosses could add to our

understanding of MCD3/4/5 function as well as aid in assembling the pathway for RNA processing.

Lastly, numerous tools were generated to aid in mapping *mcd3, mcd4,* and *mcd5* such as a new molecular map and the generation of a theoretical marker set. These will benefit the *Chlamydomonas* community as a whole, allowing for isolation of numerous genes that cannot be cloned by complementation. Specifically, cloning of *MCD3, MCD4,* and *MCD5* will enable epitope tagging and pull-down assays to be conducted. These will reveal whether they act in a complex with each other, and gene-specific proteins such as MCD1, MBB1 and NAC2 (Drager et al., 1998; Boudreau et al., 2000; Vaistij et al., 2000). Such complexes, if they exist, would also help identify other factors involved in global RNA metabolism.

REFERENCES

Boudreau E, Nickelsen J, Lemaire SD, Ossenbuhl F, Rochaix J-D (2000) The *Nac2* gene of *Chlamydomonas reinhardtii* encodes a chloroplast TPR protein involved in *psbD* mRNA stability, processing and/or translation. EMBO J. **19:** 3366-3376

Drager RG, Girard-Bascou J, Choquet Y, Kindle KL, Stern DB (1998) *In vivo* evidence for 5' to 3' exoribonuclease degradation of an unstable chloroplast mRNA. Plant J. **13:** 85-96

Herrin DL, Nickelsen J (2004) Chloroplast RNA processing and stability. Photosynth. Res. **82:** 301-314

Levy H, Kindle KL, Stern DB (1997) A nuclear mutation that affects the 3' processing of several mRNAs in *Chlamydomonas* chloroplasts. Plant Cell **9:** 825-836

Levy H, Kindle KL, Stern DB (1999) Target and specificity of a nuclear gene product that participates in mRNA 3' end formation in *Chlamydomonas* chloroplasts. J. Biol. Chem. **274:** 35955-35962

Nickelsen J (2000) Mutations at three different nuclear loci of *Chlamydomonas* suppress a defect in chloroplast *psbD* mRNA accumulation. Curr. Genet. **37:** 136-142

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

Vaistij FE, Boudreau E, Lemaire SD, Goldschmidt-Clermont M, Rochaix JD (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*. Proc. Natl. Acad. Sci. USA **97:** 14813-14818

APPENDIX

Using *Chlamydomonas reinhardtii*, three pleiotropic nuclear mutations, *mcd3, mcd4, and mcd5*, have been identified, which cause quantitative variation between polycistronic transcripts and accumulation of novel transcripts with 3' ends. While the RNA phenotype of *mcd5* is distinct from *mcd3* and *mcd4*, *mcd3* and *mcd4* share identical RNA phenotypes when transcripts from 32 genes were surveyed by RNA filter hybridization (Rymarquis et al., 2006). Since *mcd3* and *mcd4* represent separate loci, this suggests that their proteins might interact to stimulate RNA maturation and possibly degradation in *Chlamydomonas* chloroplasts.

In order to test for the potential interactions and/or redundancy between MCD3 and MCD4*, mcd3 mcd4* [LS2] was generated by crossing *mcd3* [LS2] mt+ to *mcd4* [LS2] mt-. Since *mcd3* and *mcd4* represent separate loci, four genotypes, *MCD3 MCD4* [LS2], *MCD3 mcd4* [LS2], *mcd3 MCD4* [LS2], and *mcd3 mcd4* [LS2], were expected. The *MCD3 MCD4* [LS2] genotype will confer a PS- phenotype due to the presence of an unsuppressed LS2 mutation and have mRNA patterns similar to WT, with the exception of reduced *petD* mRNA (Higgs et al., 1999). The *mcd3 MCD4* [LS2] and *MCD3 mcd4* [LS2] progeny, are genetically equivalent to the *mcd3* [LS2] and *mcd4* [LS2] parents, respectively, and would therefore be PS+ and harbor the same pleiotropic mRNA patterns as the *mcd3* and *mcd4* parents (Rymarquis et al., 2006). The only unknown photosynthetic and mRNA phenotype is that of *mcd3 mcd4* [LS2].

Tetrads with 4 PS+:0 PS- phenotype and 2 *mcd3 MCD4* [LS2]: 2 *MCD3 mcd4* [LS2] genotype were predominantly produced from this cross. This is due to the fact that the *mcd3* and *mcd4* loci are within 9-15 cM (Rymarquis et

al., 2005), thus recombination between them occurs at a correspondingly reduced frequency. Despite this, one tetrad with a 2 PS+:2 PS- phenotype was produced. This tetrad could either be a tetratype or nonparental ditype, where 1 or 2 of the progeny with the genotype *mcd3 mcd4* [LS2] were produced, respectively. The tetrad would be a tetratype if *mcd3 mcd4* [LS2] is PS-, and a nonparental ditype if it is PS+.

RNA and protein were analyzed from this tetrad in an attempt to determine the phenotype of the double mutant (Appendix 1). Progeny 1 and 4 were PS- and featured similar *petD* and *psbB* mRNA patterns, and accumulated similar PetA levels to LS2, while progeny 2 and 3 most closely resembled *mcd3* [LS2] and *mcd4* [LS2], respectively (Rymarquis et al., 2006), although progeny 3 lacked the 4.3 kb dicistronic *petA-petD* transcript. Since none of the progeny showed a phenotype substantially different from the *mcd3* and *mcd4* parents or LS2, it was impossible to determine which progeny had the *mcd3 mcd4* [LS2] genotype based on the RNA and protein data alone.

To genotype the progeny, they could be crossed to the tester strains listed in Table A1. Table A2 shows the expected results from these crosses for each potential genotype. Any cross producing tetrads with 4 PS+:0 PS- phenotype, indicates that the parental strain had the *mcd3 mcd4* [LS2] genotype, irrespective of whether *mcd3 mcd4* [LS2] is PS+ or PS-. Another method to genotype the progeny would be to create a molecular tag for the *mcd3* or *mcd4* alleles after they have been cloned.

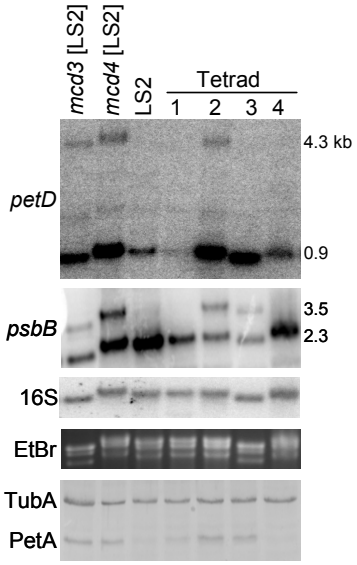


Figure A1. *mcd3 mcd4* [LS2] progeny are either PS+ or PS-. (a) An RNA gel blot was probed sequentially with *petD, psbB* and 16S rRNA. EtBr staining and 16S rRNA were used as a measure of loading. (b) An immunoblot was incubated with a PetA antibody to estimate the accumulation of subunit IV, and TubA (tubulin) was used as a loading control.

Table A1. Crosses required to genotype the progeny shown in Figure A1

Progeny	Photosynthetic phenotype	Potential genotypes		Mating type	Tester strain
1	PS-	WT [LS2]	*mcd3mcd4* [LS2]	-	LS2 mt+
2	PS+	*mcd3* or *4* [LS2]	*mcd3mcd4* [LS2]	+	WT mt-
3	PS+	*mcd3* or *4* [LS2]	*mcd3mcd4* [LS2]	-	LS2 mt+
4	PS-	WT [LS2]	*mcd3mcd4* [LS2]	+	WT mt-

Table A2. Expected cross results for each genotype. The ratios will be the same whether the tester strain is LS2 or WT because all progeny will inherit the LS2 chloroplast.

Genotype	Photosynthetic Phenotype	Parental ditype	Tetratype	Nonparental ditype
WT [LS2]	PS-	No segregating loci, all progeny will be PS-		
mcd3 or *4* [LS2]	PS+	The *mcd* locus will segregate 2:2, leading to 2 PS+:2 PS- with either tester		
mcd3mcd4 [LS2]	PS-	0 PS+:4 PS-	2 PS+:2 PS-	4 PS+:0 PS-
mcd3mcd4 [LS2]	PS+	2 PS+:2 PS-	3 PS+:1 PS-	4 PS+:0 PS-

REFERENCES

Higgs DC, Shapiro RS, Kindle KL, Stern DB (1999) Small *cis*-acting sequences that specify secondary structures in a chloroplast mRNA are essential for RNA stability and translation. Mol. Cell. Biol. **19:** 8479-8491

Rymarquis L, Higgs D, Stern D (2006) Nuclear suppressors define three factos that participate in both 5' and 3' end processing of mRNAs in *Chlamydomons* chloroplasts. Plant J.**:** In Press

Rymarquis LA, Handley JM, Thomas M, Stern DB (2005) Beyond complementation. Map-based cloning in *Chlamydomonas reinhardtii*. Plant Physiol. **137:** 557-566