Université de Montréal

Expression of the plastid genome in the dinoflagellate *Gonyaulax polyedra*

Par

Yunling Wang
Département de Sciences Biologiques
Faculté des Arts et Sciences

Thèse présentée à la Faculté des études supérieures
En vue de l' obtention du grade de
Philosophiæ Doctor (Ph. D.)
Sciences Biologiques

Janvier, 2006

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée :

Expression of the plastid genome in the dinoflagellate *Gonyaulax polyedra*

Présentée par:
Yunling Wang

a été évaluée par un jury composé des personnes suivantes:

.....Mario Cappadocia.....

président rapporteur

.....David Morse.....

directeur de recherche

.....Gertraud Burger.....

membre du jury

.....Beverly R. Green.....

examinateur externe

..... Normand Brisson.....

représentant du doyen de la FES

Sommaire

Les dinoflagellés photosynthétiques sont d'importants producteurs primaires marins en plus d'être à l'origine de marées rouges toxiques. La plupart des chloroplastes des dinoflagellés typiques diffèrent des autres chloroplastes en ce sens qu'ils sont entourés de 3 membranes et que leur pigment photosynthétique accessoire est le caroténoïde péridinine. Chez la plupart des espèces ayant la péridinine, il a été proposé que le génome chloroplastique est limité à un petit nombre de mini-cercles de 2-3 kb ne portant qu'un seul gène. Les gènes que l'on ne retrouve plus dans le chloroplaste auraient été transférés dans le génome nucléaire. L'expression de ces gènes nucléaires dont le produit ciblé le chloroplaste est souvent contrôlée par l'horloge circadienne, mais très peu de choses sont connues sur la transcription de gènes chez les dinoflagellés. De plus, on a récemment mis en doute la localisation même des mini-cercles chloroplastiques. Dans cette thèse, nous faisons l'étude du génome chloroplastique de *Gonyaulax* au point de vue de l'architecture du génome, la transcription des gènes chloroplastiques et la maturation des transcrits, ainsi que le contrôle de la traduction et de la dégradation des protéines.

Un « Southern blot » contre la séquence de psbA sur l'ADN génomique non digéré de *Gonyaulax* montre la présence d'un ADN d'une taille apparente plus grande que celle des mini-cercles (>12 kb).En effet, un gel d'électrophorèse « pulse field » montre que le gène psbA est associé à un ADN de 50-150 kb. On remarque l'absence de bande par coloration au bromure d'éthidium de l'ADN AT-

riche digéré, ce qui suggère que le génome chloroplastique de *Gonyaulax* est d'une grande complexité. L'architecture du génome chloroplastique de *Gonyaulax* diffère donc de celle des autres dinoflagellés.

L'extrémité 3' des ARNm peut être déterminée en y ajoutant d'abord des résidus guanine, puis en séquençant les produits de RT-PCR amplifiés avec une amorce oligo(dC) et une amorce spécifique au gène chloroplastique. Une telle démarche à partir d'ARNm purifiés de *Gonyaulax* montre que les transcrits de psbA et atpB contiennent une queue de poly(U) non encodé par l'ADN. Ce segment poly(U) doit donc être le résultat d'une modification post-transcriptionnelle de l'ARN. Une banque d'ADNc a été préparée en utilisant une amorce oligo(dA) plutôt que l'habituelle amorce oligo(dT). Tous les gènes contenant une queue poly(U) isolés de cette banque ont été reconnus, comme dans l'autres études sur d'autres dinoflagellés, pour n'être encodés que sur des mini-cercles. Tous les gènes ont une queue de poly(U). Deux de ces gènes encodent les ARNr 23S et 16S, alors que les autres encodent des protéines impliquées dans la phase lumineuse de la photosynthèse. Ceci confirme le contenu réduit du génome chloroplastique de *Gonyaulax*.

Plusieurs protocoles d'isolation de chloroplastes ont été élaborés pour permettre des analyses protéomiques et d'activité. Nous avons constaté que les organelles sont fragiles et que leur isolation sous forme intacte et active est impossible. Les protocoles classiques de purification utilisés pour les plantes supérieures n'ont

donné que des chloroplastes brisés comme le montre la perte totale de la Rubisco, un enzyme marqueur du stroma. De petites quantités de Rubisco ont pu être retenues dans les chloroplastes si les cellules étaient traitées à la cellulose avant la lyse.Même si presque la totalité de Rubisco a été retenue dans les chloroplastes préparés à partir de cellules traitées à la chaleur, les protéines ont été dénaturées par le traitement et les chloroplastes n'étaient pas fonctionnels.

L'incorporation *in vivo* de [^{35}S]-méthionine dans les protéines de chloroplastes purifiés montre que plusieurs protéines ont été marquées en présence de cycloheximide mais pas avec du chloramphénicol. Donc, ces protéines doivent être synthétisées dans les chloroplastes. Une de ces protéines, marquée en présence de cycloheximide chez deux espèces de dinoflagellés, a été identifiée comme étant psbA par « Western blot ». Tant la synthèse que la dégradation de la protéine psbA sont induites par la lumière comme c'est le cas chez les plantes supérieures. Contrairement à plusieurs protéines encodées dans le noyau et ayant pour cible chloroplaste, ni la synthèse ni la dégradation de psbA ne sont contrôlées par l'horloge circadienne. Ces résultats suggèrent que la protéine psbA n'est pas un candidat probable pour entraîner le rythme circadien d'évolution de l'oxygène.

Mots-clés : dinoflagellés contenant de la péridinine, *Gonyaulax*, génome chloroplastique, gène sur mini-cercle, expression génique, psbA, ARN

chloroplastiques, modifications post-transcriptionnelles, polyuridylylation, synthèse protéique et dégradation.

Abstract

Photosynthetic dinoflagellates are important aquatic primary producers and the notorious cause of toxic 'red tides'. Typical dinoflagellate chloroplasts differ from all other plastids in having a combination of three envelope membranes and the carotenoid peridinin as an accessory light-harvesting pigment. In many peridinin-containing dinoflagellate species, the plastid genome has been proposed to exist as a limited number of single-gene minicircles whose size is typically only 2-3 kb. Many of the genes no longer found in the plastid genome are instead nuclear-encoded. The expression of these nuclear-encoded plastid-directed gene products is often regulated by the circadian clock, but little is known about expression of dinoflagellate chloroplast genes. Furthermore, even the plastid location of the minicircles has recently been challenged. Here, we describe our studies on the plastid genome of the dinoflagellate *Gonyaulax* in terms of genome architecture, transcription of plastid genes and transcript processing, and control of protein translation and degradation.

Undigested genomic DNA, probed on southern blots with the psbA sequence, shows the presence of *Gonyaulax* DNA with an apparent size (>12 kb) significantly larger than the minicircles usually found. Indeed, pulse field gel electrophoresis shows the psbA gene is associated with DNA of roughly 50-150 kb. Plastid genes are associated with AT-rich DNA, yet ethidium bromide staining of restriction enzyme digests of this DNA show no discrete bands. This suggests

that the *Gonyaulax* plastid genome has an unusually high complexity and that the plastid genome architecture in *Gonyaulax* is unusual even for dinoflagellates.

The 3' end tail of plastid mRNAs can be determined by first tailing the RNA with guanine residues then sequencing RT-PCR products amplified using an oligo(dC) primer and an internal primer specific for a chloroplast gene. These experiments using purified RNA from *Gonyaulax* show that both psbA and atpB transcripts contain a poly(U) stretch not encoded in the DNA. This poly(U) must thus result from post-transcriptional RNA processing. A cDNA library was prepared using an oligo(dA) primer, instead of the oligo(dT) typically used for cDNA synthesis. The only poly(U)-containing genes isolated from this library were found to be genes known from previous studies using other dinoflagellates to be encoded by minicircles. Two genes encoded the 23S and 16S rRNAs, while the remaining ten encoded membrane proteins involved in the light reactions of photosynthesis. This provides an independent measure confirming the highly reduced gene content of *Gonyaulax* plastids.

Several different plastid purification protocols were elaborated in order to isolate plastids for activity and proteomic analysis. We have found that the organelles are inherently fragile and that isolation of active unbroken organelles was impossible. Standard purification protocols, such as those used for higher plant plastids, produced only broken organelles as judged by complete loss of the stromal protein marker Rubisco. Small amounts of Rubisco could be retained in

the plastid if the cells were treated with cellulose prior to lysis. Although almost all Rubisco could be retained in plastids prepared from cells following a heat shock treatment, the proteins were denatured by the treatment and the plastids were not functional.

The *in vivo* incorporation of [^{35}S]-methionine into the proteins of purified plastids shows that several plastid proteins incorporate labels in the presence of cycloheximide but not chloramphenicol. Thus, these proteins must be synthesized in the plastids themselves. One of these proteins, labeled in the presence of cycloheximide in two different dinoflagellate species, was identified as psbA by Western blot analysis. Both synthesis and degradation of psbA protein are induced by light, similar to what is found with psbA of higher plants. Unlike the nuclear-encoded plastid-directed proteins studied to date, neither synthesis of degradation of psbA protein are regulated by the circadian clock. These results suggest that psbA protein is not a likely candidate for mediating the circadian rhythm in oxygen evolution rates.

Key words: peridinin-containing dinoflagellate, *Gonyaulax*, plastid genome, mini-circle gene, gene expression, psbA, plastid RNAs, post-transcriptional modification, polyuridylylation, protein synthesis and degradation.

Table of contents

List of Tables.

List of Figures.

List of abbreviations

2D-PAGE: 2-dimensional polyacrylamide gel electrophoresis

bp: base pair

cDNA: complementary DNA

CHAPS: 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate

CT: circadian time

cpDNA: chloroplast DNA

CTAB: hexadecylamine triammonium bromide

DTT: dithiothreitol

EDTA: ethylene diamine tetra acetic

ER: endoplasmic reticulum

GAPDH: glyceraldehydes-3-phosphate dehydrogenase

IEF: isoelectric focusing

kb: kilobase

kDa: kiloDalton

LBP: Luciferin binding protein

LD: light-dark

LDS: lithium dodecyl sulfate

LHC: light-harvesting complex

LHCP: light-harvesting chlorophyll-binding protein

mM: millimolar

mRNA: messenger RNA

ORF: open reading frame

OEE1: oxygen evolving enzyme subunit 1

PAGE: polyacrylamide gel electrophoresis

PAP: poly(A) polymerase

pg: pictogram

PCP: peridinin-chlorophyll a-protein

PCR: polymerase chain reaction

PNPase: polynucleotide phosphorylase

PSI: photosystem I

PSII: photosystem II

RBPs: RNA binding proteins

RBS: RNA binding sequence

RuBisCO: Ribulose bisphosphate carboxylase/oxygenase

rRNA: ribosomal RNA

tRNA: transfer RNA

RT-PCR: reversal transcription- PCR

SDS: sodium dodecylsulfate

SD: Shine-Dalgarno

TAIL PCR: thermal asymmetric interlaced PCR

µm: micrometer

UTR: untranslated region

Acknowledgements.

First of all I wish to express my sincere gratitude to my supervisor, Dr. David Morse, who guided this work and helped whenever I was in need. His continuous, patient, unconditional support and encouragement gave me enough confidence to get the job done. Without his help, this work would not be possible.

I must thank the students in the Morse laboratory, Nasha Nassoury, Thierry Bertomeu, Tyler MacKenzie and Mathieu Lapointe, for their support and their comradeship.

I am grateful to Dr. Mario Cappadocia and the members of his laboratory, Doan Luu, Xike Qin, Bolin Liu, for their generous assistance.

I would also like to thank the Bruneau laboratory for sequencing my samples.

Especially, I would like to give my special thanks to my husband Baoquan Lu whose patient love enabled me to complete this work.

Finally, I must acknowledge the financial support received from Les Bourses d'Excellence de la Faculté des Études Supèrieurs de l'Universitè de Montréal, and from the National Science and Engineering Research Council of Canada grant awarded to Dr. David Morse.

Chapter I. Introduction

I.1. The dinoflagellate algae

Dinoflagellates are fascinating protists that have attracted researchers from many different fields. The free-living species are major primary producers in the marine environment and the cause of the harmful algal blooms colloquially called red tides. The toxins cause a variety of poisonings that affect humans and marine wildlife and have a significant impact on coastal ecosystems throughout the world. In addition to the major ecological role played by the dinoflagellates, these organisms also display a number of highly unusual morphological features. For example, the organization of the dinophycean nucleus is a case in point, as it is unlike that found in any other eukaryotic cell. A striking morphological feature is the permanently condensed state of the chromosomes (i.e., even during interphase). Dinoflagellate nuclei can contain vast amounts of DNA compared to other eukaryotes. For example, estimates range from 3 to 250 pg per cell, while in comparison, haploid human nuclei contain a mere 3.2 pg per cell. The dinoflagellate nucleus contains such a high concentration of DNA that it exists in a liquid crystal state, which is thought to be responsible for the unique morphology (Livolant and Bouligand 1978);(Gautier, Michel-Salamin et al. 1986) .

The organization of the dinophycean chromosomes is poorly understood, as there is very little nucleohistone (Rizzo 1981). Although the nucleus does contain a small amount of basic "histone-like" protein (Chudnovsky and Li 2002) , the ratio of DNA to basic protein in dinoflagellate chromosomes has been estimated to be 10:1. This is dramatically higher than the 1:1 ratio observed in most eukaryotes and is more similar to what is found in bacteria. Indeed, phylogenetic analysis supports a common origin of these histone-like proteins with basic DNA-binding proteins of bacteria (Wong, New et al.

2003)Furthermore, the general lack of nucleosomes raises many questions about transcription and regulation of gene expression in these organisms. Evidently, the so-called histone code for regulation of genome structure and function can not be operative in an organism lacking histones (Rizzo 1981). The pattern of DNA synthesis also appear to be intermediate between eukaryotic and prokaryotic organisms, with some species reportedly undergoing continuous synthesis as found with prokaryotes (Filfilan and Sigee 1977; Rott, Zipor et al. 2003) and others discontinuous as normally observed with eukaryotes (Franker, Smith et al. 1973).

Dinoflagellate nuclear DNA usually presents a high G + C content, contains up to 70% of modified and rare bases non-randomly distributed in the genome (Triplett, Govind et al. 1993) (Rae and Steele 1978). GC rich content is reflected in EST banks from different dinoflagellates, the average GC content from 51% (*K. brevis*) to 61.7% (*A. tamarense*)(Hackett, Yoon et al. 2004; Hackett, Scheetz et al. 2005; Lidie, Ryan et al. 2005) The most frequently found of the unusual bases is hydroxymethyl uracil (Rae and Steele 1978), but it is not clear what role this may play in organizing or expressing the genome. One of the most remarkable features of the dinoflagellate genome sequences so far analyzed is the absence of TATA box or other consensus sequences in their promoters, making expression of the genome at the transcriptional level somewhat of an enigma. Also curious is the tendency toward redundancy at the DNA level, with individual genes containing repeated regions and multiple copies of genes in tandem. Many genes lack introns and produce either polycistronic RNA, like the early diverging eukaryotes, Trypanosomatidae (Elias, Marques-Porto et al. 2001), or discrete mRNAs, such as PCP, that can be translated to produce either individual protein molecules directly or long polyproteins that must be proteolytically processed to release the mature subunits(Rowan, Whitney et al. 1996; Moreno Diaz de la Espina, Alverca et al. 2005).

The most widespread plastid in dinoflagellates contains the unique photopigment peridinin. The peridinin plastid is remarkably different from chloroplasts in other eukaryotes because it lacks a typical genome. Plastids normally contain a circular genome of about 150 kb that encode 100-200 genes necessary for plastid function. In peridinin-containing dinoflagellates, however, the plastid genome has been fragmented into minicircles that encode a single, or only a few, genes per circle. However, only 18 genes have been identified thus far on minicircles(Zhang, Green et al. 1999; Barbrook, Symington et al. 2001; Nisbet, Koumandou et al. 2004)). Recent studies show that the other genes usually found in chloroplasts have been transferred to the nucleus, and indeed 15 genes found exclusively on the plastid genome in all other photosynthetic eukaryotes are encoded in the nucleus of dinoflagellates(Bachvaroff, Concepcion et al. 2004; Hackett, Yoon et al. 2004). The plastids in peridinin-containing dinoflagellates thus appear to encode the smallest number of plastid genes of any photosynthetic eukaryote, making them a model for understanding gene transfer from organelles to the nucleus (Hackett, Scheetz et al. 2005).

Gonyaulax polyedra is a dinoflagellate with peridinin-containing plastid. It has been extensively studied for over 50 years because of its bioluminescence and its abundant circadian biology. The rhythms are particularly interesting because they include daily variations in the rate of physiological processes such as photosynthesis, bioluminescence, cell division, and enzyme synthesis. In *Gonyaulax*, both oxygen evolution (Sweeney 1960) and carbon fixation (Hastings, Astrachan et al. 1961) are bona fide circadian rhythms. The synthesis of nuclear-encoded plastid proteins is also circadian, with examples including the peridinin-chlorophyll a-protein (PCP), Ribulose bisphosphate carboxylase/oxygenase (RuBisCO), oxygen evolving enzyme subunit 1 (OEE1) and glyceraldehydes-3-phosphate dehydrogenase (GAPDH). Interestingly, the distribution of RuBisCO changes in a circadian fashion (Nassoury, Fritz et al. 2001), and may be related to

changes in protein synthesis rates (Nassoury and Morse 2005). Unlike the other dinoflagellate plastids, which have minicircle genome DNA, *Gonyaulax* has a large chloroplast genome DNA. This means minicircle genome is not the only dinoflagellate plastid genome format and that *Gonyaulax* might provide clues to the ancestral state of plastid genome.

I.2. The endosymbiotic origin of chloroplasts

Plastids in plant and algal cells are usually double membrane-bounded subcellular organelles with diverse, specialist functions. The best known of these organelles, the pigmented chloroplasts, are photosynthetic and are responsible for harvesting the energy of photons is captured and catalyzing the biosynthesis of sugar from carbon dioxide. Molecular genetic studies have demonstrated the ubiquity of plastid genomes and confirmed that their replication, transcription and translation closely resemble those of bacteria. Molecular phylogenetic studies now make it abundantly clear that the closest bacterial homologs of plastids are indeed cyanobacteria(Douglas and Raven 2003) Among the algae, however, the organelles are not always surrounded by two bounding membranes. Indeed, four different groups of photosynthetically active organisms can be distinguished (Lee 1989). The first evolutionary group contains the prokaryotic algae and includes the Cyanophyta (blue-green algae) and the Prochlorophyta (prochlorophytes). These prokaryotic algae were the first organisms to function in photosynthesis and are not plastids but free-living cells. The second evolutionary group contains those algae with a chloroplast surrounded by the two membranes. This group includes Rhodophyta (red algae), Chlorophyta (green algae) and Glaucophyta. The Rhodophyta and Chlorophyta represent the completion of an evolutionary pathway linking free-living cyanobacteria with mature chloroplasts, while the Glaucophyta represent an intermediate stage, as the endosymbiotic chloroplasts retain a peptidoglycan cell wall between the two plastid envelope membranes as found in cyanobacteria

(Kies and Kremer 1990). The third evolutionary group contains the Euglenophyta (euglenoids) and Dinophyta (dinoflagellates), and these two groups of algae have three membranes surrounding the chloroplast envelope. The fourth group and most diverse group contain those algae with four membranes constituting the chloroplast envelope. This group includes Cryptophyta (cryptophytes), Chrysophyta (golden brown algae), Prymnesiophyta (prymnesiophytes or haptophytes), Bacillariophyta (diatoms), Raphidophyta (chloromonads) Xanthophyta (yellow-green algae), Eustigmatophyta, and Phaeophyta (brown algae). However clear and unequivocal these groups are, this particular classification is of limited phylogenetic use as it provides little information as to the provenance of the different groups. For this, it is necessary to perform phylogenetic reconstructions so that the evolutionary relationships between the different algal groups can be revealed. These phylogenetic relationships are best considered in the light of endosymbiotic theory.

I.2.1 Endosymbiotic theory

In 1975 the theory of endosymbiosis, which describes the origins of mitochondria and plastids from eubacterial-like cells living within a eukaryotic host, was still hanging in balance. Indeed, many scientists favored ideas pronounded by Lawrence Bogorad (Bogorad 1975) and Tom Cavalier-Smith (Cavalier-Smith 1975) that proposed an autogenous (non-endosymbiotic) origin for chloroplasts. However, during this same year Linda Bonen and Ford Doolittle provided the first qualitative measure of similarities between T1 rRNA catalogues of cyanobacteria and chloroplasts (Bonen and Doolittle 1975). Their work was first molecular data that provided strong evidence against the autogenous origin of plastids. Currently, based on both structural and molecular data, the origin of plastids theory is considered to be quite well established.

At present, it is thought that two types of endosymbiosis have occurred during the evolution of plastids. These two types are termed primary and secondary endosymbiosis. In primary endosymbiosis, a previously free-living cyanobacterium took up permanent and heritable residence in a previously non-photosynthetic eukaryote. Over time, the cyanobacteria and the host cell became interdependent, with the bacteria obtaining a safe refuge and the host cell obtaining access to photosynthetic power. The plastids we know today have resulted from massive losses and relocation of gene function from organelle to the host cell nucleus, a process that presumably happened early during the formation of plastids. Plastids descending directly from this event are called "primary plastids". Secondary endosymbiosis, on the other hand, involves the engulfment of a photosynthetic eukaryote already equipped with plastids by another eukaryote. These "secondary plastids" remain a permanently retained part of an original prey and have produced an endosymbiont by transfers and losses of large number of genes. It is thought that this cell-prey behavior responsible for the evolution of the plastids in most algal groups occurred late (Delwiche and Palmer 1997) (figure I.1).

There is universal consensus that all well-recognized types of primary plastid-containing organisms fall into three groups, each clearly monophyletic. There is also broad consensus, based on many lines of evidence, which all three of lineages "probably" trace back to the same cyanobacterial endosymbiosis. Thus, primary plastids probably arose once and only once in evolution, analogous to the evolution of mitochondria (Palmer 2003) In contrast, the number of secondary symbioses is quite unclear and remains, in some respects, controversial. Cavalier-Smith (Cavalier-Smith 1982; Cavalier-Smith 2003) has long argued that secondary plastid symbiosis is a very difficult and rare process, and thus is likely to have occurred only twice. In one group, the enslavement of a green alga as an endosymbiont has led to chloroarachniophytes and euglenophytes. In the second group, uptake and enslavement of a red alga led to the other five lineages of secondary plastids.

A contrasting viewpoint is that secondary endosymbiosis origin of chloroplast may have happened more frequently, and that there may even be tertiary endosymbiotic events in which an alga containing a secondary endosymbiont becomes engulfed by a eukaryotic cell. In recent years, tertiary endosymbiosis has received some research support from studies on the dinoflagellates. For example, phylogenetic analysis (Tengs, Dahlberg et al. 2000) based on plastid small-subunit rRNA genes (ssu rDNA) sequences of three so-called "green" 19'Hexanoyloxy-fucoxanthin-containing dinoflagellates, *Gyrodinium aureolum*, *Gymnodinium galatheanum,* and *Gymnodinium breve*, indicated the plastids were acquired via endosymbiosis of a haptophyte. In addition, two other research groups also argued for several secondary endosymbiosis and tertiary endosymbiosis of dinoflagellate fucoxanthin- and peridinin-containing plastids. Likewise, the nuclear-encoded plastid-directed oxygen-evolving enhancer 1 (PsbO) genes of several dinoflagellates most closely resembled those from haptophytes, a phylogenetically unrelated alga, and thus this nuclear encoded protein might have been obtained by tertiary endosymbiosis (Ishida and Green 2002). A previous suggestion that the green and red plastids of the dinoflagellates might be closely related phylogenetically (Yoon, Hackett et al. 2002) appears to have been the result of an artificial clustering due to codon usage heterogeneity (Inagaki, Simpson et al. 2004). Indeed, more recent studies provide strong support for the replacement of the peridinin-type plastid with a haptophyte plastid in the green dinoflagellates (Yoon, Hackett et al. 2005). Interestingly, an examination of the nuclear genome of the green dinoflagellates also indicate that the ancestral peridinin-type plastid-directed genes are absent, suggesting that extensive remodeling of the host nuclear genome occurs subsequently to the endosymbiosis.

Right now, one key question is whether plastids are monophyletic (derived from a single cyanobacterial ancestor) or polyphyletic (derived from more than one ancestor). Although on balance current evidence favors plastid

monophyly, more evidence is needed, especially as it relates to host cell (nuclear) and mitochondrial phylogeny. However many times the various events have occurred, it seems clear from ultrastructural considerations and molecular biology studies that some plastids result from primary endosymbiosis, while others result from secondary endosymbiosis.

I.2.2 The formation of plastids and the acquisitive theory of photopigments

The endosymbiosis theory interprets those chloroplasts surrounded by two membranes as having evolved from primary endosymbiosis, with the inner membrane of the chloroplast envelope derived from the primary endosymbiont, and the outer membrane of the chloroplast envelope derived from the food vesicle membrane of the host cell. Similarly, those chloroplasts surrounded by three or four membranes in the envelope were thought to have evolved from secondary endosymbiosis, with the inner two membranes corresponding to the original plastid envelope and the outermost membrane corresponding to the host food vacuole.

According to the acquisitive theory of photopigments, the ancestral plastid contained the conventional cyanobacterial photopigments chlorophyll a and phycobilins, as do the red alga and glaucophytes. Subsequently, chlorophyll c was acquired by heterokonts, haptophytes, stereotypical dinoflagellates, while green algae acquired chlorophyll b and also lost phycobilin (Bryant 1992). This theory fits reasonably well with current evidence that primary plastids probably arose from a single endosymbiotic event, and the secondary chloroplast came from the different primary lineages. The algae whose chloroplasts use chlorophyll a and b as the photopigments are thus green algal progenies, while algae whose chloroplasts use chlorophyll a and c as the photopigments are red algal progenies.

I.2.3 Gene transfer from the chloroplast to the nucleus

The first chloroplast genome sequence, from the liverwort *Marchantia polymorepha*, was determined to be a 121,024 bp circular DNA molecule almost twenty years ago (Ohyama, Fukuzawa et al. 1986). Since this time, numerous other chloroplast genomes have been sequenced, and typically, they are circles with genes organized in operons. Plastid gene content varies from about 70 in some non-photosynthetic plastids such as that of the malaria parasite up to about 200 in red algae(McFadden 2001). Indeed, more than 93% of the ancestral endosymbiont genome has been lost.

The DNA encoding necessary plastid genes has not been permanently lost however, merely transferred to the host cell nucleus. The pathway for this DNA transfer is not known but may include transient breaches in the organelle membranes during fusion and/or budding processes, terminal degradation of organelles by autophagy coupled with the subsequent release of nucleic acids to the cytoplasm, illicit use of nucleic acid or protein machinery, or fusion between heterotypic membranes. Some or all of these pathways may lead to the escape of DNA or RNA from plastid to the nucleus (Thorsness and Weber 1996; Martin 2003). Independent of the mechanism, gene transfer has occurred in large amounts (Martin 2003). Furthermore, the process is potentially still occurring, as cpDNA transfer to the nucleus has been measured in tobacco using a nucleus-specific neomycin phosphotransferase gene (encoding kanamycin resistance) integrated into the chloroplast genome, and screening for kanamycin-resistant seedlings in progeny (Huang, Ayliffe et al. 2003). The measured rate is actually quite high, as Huang and colleagues calculate that roughly one transposition event occurs in about 16,000 pollen grains. The transfer thus explains the reduced genome size of the plastid compared to their progenitor cyanobacterium (which has more than 3000 genes).

There have been many hypotheses to explain why genes should be lost from organelles to the nucleus. One possibility is that transfer confers an advantage since the genes in chloroplast or mitochondria are exposed to high levels of mutagenic species such as oxygen free radicals generated during photosynthesis or respiration. Movement of the genes into the nucleus might thus help to reduce the occurrence of deleterious mutations in organelle genes (Allen and Raven 1996). A second possibility, based on the uniparental inheritance of organelles, is that transferring the genome is advantageous since these genes would then experience segregation and recombination through sexual reproduction (Muller 1964). Still another idea is that the movement of organelle genes to the nucleus avoids the tendency of AT-rich DNA to experience unfavorable substitution (Howe, Barbrook et al. 2000) There are also many theories proposed to explain why this gene transfer has not proceeded to completion, such as mutations in the genetic code(de Grey 2005) or the role of the conserved genes in energy transduction(Allen, Puthiyaveetil et al. 2005). These theories remain largely speculative, however.

Gene retention and loss in chloroplast genomes is thought to have occurred fairly early during evolution for the primary plastids, yet the remaining genes do still reflect evolutionary patterns. Patterns of gene losses occurred different at phylum level radiations within primary symbiotic events (Grzebyk and Schofield 2003). Between 1% and 10% of the remaining plastid genes were lost at phylum level radiations within the Rhodophyta and Chlorophyta. A large number of the remaining primary plastid genes were lost at radiations linked to secondary symbiotic events. Analysis of 24 complete chloroplast genome sequences (10 from alga, 14 from land plants) has revealed a set of approximately 50 core protein-coding genes retained in all taxa (figure I.2) (Grzebyk and Schofield 2003). These genes are broadly clustered into three major function domains: 1) genes encoding for the proton-coupled ATP synthase (atp genes) 2) genes encoding for

photosynthetic processes, including reaction center proteins (psa, psb genes), as well as most of the electron transport components that connect the two photosystems (pet genes) and large subunit of RuBisCO (rbcL genes), and 3) housekeeping genes that include the plastid ribosomal proteins (rpl rps genes), RNA polymerase (rpo genes), and elongation factors (tufA genes).

I.2.4 Targeting of nuclear-encoded proteins to the plastids

The *Arabidopsis* genome project surveys indicate that as many as 2,000 nuclear genes encode proteins with plastid function(Abdallah, Salamini et al. 2000). The transfer of these genes from the chloroplast to the nucleus typically requires return passage of the gene product to its place of function. This protein targeting and transport mechanism is central to the establishment and refinement of endosymbiosis. Proteins targeted to the primary plastids from the nucleus need to cross two membranes to enter plastids. Proteins targeted to secondary plastids have an added level of difficulty, as these organelles contain one or two extra membranes surrounding the plastid. The development of a generalized protein-targeting mechanism in the chloroplast envelope could have enabled any proteins that acquired an N-terminal transit sequence to be imported; the random insertion of genes transferred to the nucleus could have ensured that eventually, an appropriate N-terminal sequence came to be associated with the transferred gene. Once any copies of a gene transferred to the nucleus acquired transit sequences, the plastid copy of the gene could be lost.

The prototypical N-terminal transit peptide sequence found in Chlorophyta and higher plants is a stretch of amino acids rich in hydroxylated amino acids such as serine and threonine, and containing some basic but few acidic amino acids (Keegstra 1989).This same transit peptide is also used for translocation into Rhodophyta (Apt, Hoffman et al. 1993) and Glaucophyta (Jakowitsch, Neumann-Spallart et al. 1996) plastids, suggesting that the gene

transfers and targeting signal acquisitions occurred prior to the divergence between the different primary plastid-containing lineage (Nassoury and Morse 2005). Furthermore, the same serine/threonine-rich sequence is also used for entry into the inner two membranes of dinoflagellate (Nassoury, Cappadocia et al. 2003) and diatom (Lang, Apt et al. 1998) plastids, suggesting that gene transfer to the nuclei of the secondary plastid-containing lineages may have occurred after the first transit sequence was already in place. For all secondary plastids, including Euglena and dinoflagellates whose plastids contain three membranes, the leader sequence is hydrophobic and is followed by a transit peptide domain rich in serine and threonine. This structure is similar to that predicted by the secretory membrane hypothesis (Killian and Kroth 2003), which proposes that passage of the additional membranes around the plastid is similar to passage across the ER membrane. Indeed, the N-terminal domain closely resembles a typical hydrophobic signal peptide. Thus the transfer of a gene, already encoding a transit peptide at its N-terminus, from the photosynthetic endosymbiont nucleus to the new host cell nucleus, could be retargeted to the secondary plastid if the translocation added a signal peptide N-terminal to the transit peptide already present.

For the triple membrane bound plastids, this model holds up quite well. The leader sequences contain an N-terminal signal peptide allowing entry to the ER, a central S/T rich region acting as a transit peptide for passage across the inner two membranes. The only unusual feature of the leader is a second hydrophobic region that follows the S/T-rich region believed to anchor the protein in the membrane of transport vesicles ferrying the protein from the ER through to the Golgi and eventually the plastid (Nassoury, Cappadocia et al. 2003; Patron, Waller et al. 2005).

While the second hydrophobic region is never found in leader sequences that target proteins to the stroma of four membrane-bound plastids, the rest

of the targeting signal (a hydrophobic ER-signal sequence followed by the S/T-rich transit peptide) is similar (Nassoury and Morse 2005). However, the signal peptide crosses over only the outermost membrane, while the transit peptide allows passage over the innermost two membranes. This four membrane-bound plastid thus has an additional complication, the membrane underlying that delimiting the exterior of the plastid (second from the outside). The mechanism is not yet completely clear, but one model has been proposed based on microscopic observation of vesicles in the space between the two middle membranes. Here, plastid-directed proteins with a signal peptide in their leader sequence exploit the secretory pathway of the host for targeting to the outer membrane. Once past this first barrier, proteins found themselves outside the host cell and at the exterior surface of a membrane topologically equivalent to the former plasma membrane of the endosymbiont (Gibbs 1979). They are taken up by the equivalent of endocytosis, and thus arrive in front of the innermost membrane where they can be transported using the transit peptide.

I.2.5 The molecular phylogeny study of plastid evolution

Molecular phylogenies can, at best, reveal the evolution of the gene being analyzed: conclusions about the evolution of the organism or organelles derived from molecular data are inferences that assume the evolution of the gene mirrors the evolution of the organism or organelle. There are been many efforts to determine the evolutionary order of endosymbiosis and evolution of the eukaryotic cell. Based on the presence or absence of cytoplasmic cell organelles and gene sequences in proteins, it has been suggested that eukaryotes developed first and only later did they engulf bacteria by phagocytosis to form organelles. The first organelles to form were the mitochondria, derived from proteobacteria more than 1.5 billon years ago. Only later were cyanobacteria engulfed and used to form chloroplasts

(Cavalier-Smith 1987; Burger, Saint-Louis et al. 1999; Gray, Burger et al. 1999; Roger 1999).

Although phylogenies based on several mitochondrial genes support a specific green plant/red algae relationship (Lang, Gray et al. 1999; Burger, Zhu et al. 2000), the phylogenetic analysis of nuclear-encoded genes yields inconclusive, sometimes contradictory results (Regan and Gutell 1995; Stiller and Hall 1997)Moreira (Moreira, Le Guyader et al. 2000) used the nuclear-encoded elongation factor 2 (flr) and 13 other genes to do their phylogenic analysis. Their analysis significantly supported sisterhood of green plants and red alga. The primordial nature of the mitochondria suggests that nuclear and mitochondrial phylogenies should mirror plastid phylogenies at least for those organisms with primary plastids. Plastid DNA is a primary source of molecular variations for phylogenetic analysis of photosynthetic eukaryotes. Plastid data, both sequence-based and idiosyncratic, overwhelmingly support a monophyletic origin of the primary plastids (red algae, green algae, and glaucophytes). For some genes, sequence information from many taxa is available. For example, the 16S rRNA phylogenies usually used in phylogenic analyses clearly support the monophyletic nature of the primary plastids, and place plastids close to the base of cyanobacteria (Nelissen, Van de Peer et al. 1995; Bhattacharya and Schmidt 1997). This monophyletic nature appears at odds with phylogenetic analysis of rbcL genes showing separate lineages for the red and green algae (Delwiche and Palmer 1996). The current view of this potential problem is that rubisco genes have undergone extensive horizontal gene transfer that has compromised the phylogeny obtained from this sequence.

During the past decade, the availability of whole genome sequences has provided a wealth of information to those interested in the evolutionary origin of plastids. The first complete plant genome sequence, of *Arabidopsis*, was reported in 2000 and the sequences of a small red alga *Cyanidoschyzon*

(The Arabidopsis genome initiative 2000; Matsuzaki, Misumi et al. 2004) and the diatom *Thalassiosira* (Armbrust, Berges et al. 2004) soon followed. As expected, the *Arabidopsis* genome contains genes whose origin is clearly cyanobacterial. Likewise, the red algal genome confirms that the origin of Calvin cycle genes (other than the rubisco genes) is essentially the same as those in *Arabidopsis*. The nuclear genome of the diatom contains genes similar to both cyanobacteria and to red algae, as expected if transfer to the nucleus came from a red algae to which cyanobacterial genes had already been transferred. Indeed, the comparative analysis of whole genomes provides the most complete information on an organism's history, and it reveals the evolutionary pathways followed by the various genes that comprise a genome. Often gene histories are highly divergent within a genome, such that the notion of a lineage may be rendered almost meaningless, because the early prokaryotic genome represents a composite of genes with different history (Doolittle 1999).

Considerably more information is available for whole chloroplast genomes, and protein sequences from 21 complete chloroplast genomes were analyzed in comparison with selected archaea, eubacteria, and eukaryotes. The distance-based analysis showed that chloroplast genomes are most closely to cyanobacteria, consistent with the endosymbiotic origin of the chloroplasts. The chloroplast genomes are separated into two major clades corresponding to Chlorophyta (green plants) and Rhodophyta. The interrelationships among the chloroplasts are largely in agreement with the current understanding on chloroplast evolution (Chu, Qi et al. 2004). For example, Douglas and Penny analysis of the complete plastid genome sequence from the cryptophyte algae, *Guillardia theta*, confirmed its common ancestry with red algae (Douglas and Penny 1999). The phylogeny of 5 genes of chromist plastids supported chromist plastids' secondary endosymbiotic origin from red algae (Yoon, Hackett et al. 2002).

I.2.6 The many types of Dinoflagellate chloroplasts

Dinoflagellates have a great variability in their life styles, with many heterotrophic, parasitic, symbiotic and autotrophic examples. Among the autotrophic lifestyles are found the photosynthetic forms, the majority of which contain triple membrane-bound plastids whose pigment complement includes chlorophylls a and c and the carotenoid peridinin. To date, the peridinin accessory pigment has only been found in dinoflagellates. It usually occurs together with the chlorophyll a in a water-soluble protein complex called peridinin-chlorophyll a-protein (PCP) (Norris and Miller 1994; Le, Markovic et al. 1997), which, as perhaps might be expected, is also unique to dinoflagellates. Interestingly, the three dimensional structure of PCP has been solved, and it is quite unlike any other light harvesting protein (Hofmann, Wrench et al. 1996). This unusual protein underscores the difficulty in determining the evolutionary ancestors of the plastid constituents.

While the majority of the dinoflagellates harbor a peridinin-type plastid, other plastid types also exist. Arguably the most fascinating of these alternative plastids are those of the fresh- and brackish-water species *Peridinium balticum*. The plastids in this species are similar to those in the Heterokontophyta, as they contain lamellae composed of three thylakoids, a girdle lamella and a ring-shaped nucleoid, instead of the dispersed nucleoid found in the typical dinophyte chloroplast. The reason for this unusual structure is that the chloroplasts in fact belong to an endosymbiotic alga that lives within the *P. balticum* cells (Thomas and Cox 1973; Keeling 2004). Ultrastructural analysis has revealed that the endosymbiont has conserved its own plasma membrane, nucleus, chloroplasts, mitochondrion and Golgi apparatus. Furthermore, the nucleus of the endosymbiont does not contain the condensed interphase chromosomes typical of the dinoflagellates, but instead has a normal nucleus whose chromosomes are extended and diffuse in interphase.

In addition to *P. balticum* there are at least four other species of the Dinophyta that have chloroplasts with structural features characteristic of different alga (Keeling 2004) *Lepidodinium* has plastids derived from a green alga, *Karenia* has a haptophyte-like plastid, *Dinophysis* has a cryptophyte-like plastid and *Kryptoperidinium* has a plastid seemingly derived from a heterokont. It is interesting that all of them have fucoxanthin as the principal accessory pigment- peridinin, the principal accessory pigment of most photosynthetic Dinophyta, is absent. These examples provide a dramatic illustration of the apparent readiness of dinoflagellates to enter into different symbiotic arrangements.

Are peridinin-containing plastids the secondary plastids, and if so, from what evolutionary ancestor are they derived? This question is surprisingly difficult to answer as we have suggested above based on the nature of PCP. Furthermore, many other nuclear encoded plastid targeted gene products are similarly uninformative with respect to the origin of the dinoflagellate plastids. In perhaps the most blatant example, while cyanobacteria and all other plastids contain a form I RuBisCO, peridinin-containing dinoflagellates use a different form of the enzyme, termed a form II RuBisCO. This form II enzyme, previously found only in some species of anaerobic proteobacteria (Morse, Salois et al. 1995), shares only limited (~25%) sequence identity with the form I. How a form II RuBisCO came to replace the putatively native RuBisCO of cyanobacterial form I origin is completely conjectural, although there are two likely explanations. One is that because the mitochondrion is of proteobacterial descent, horizontal gene transfer of the proteobacterial/mitochondrial form II gene to the nuclear genome may have occurred, followed by targeting of the form II RuBisCO to the plastid origin and eventual loss of the original plastid form I gene (Morse, Salois et al. 1995; Palmer 1995; Rowan, Whitney et al. 1996). However, it is also clear that the form II RuBisCO and the presence of PCP are found together (Morse et al.,

1995), and in those species such as *Karenia* that have replaced the peridinin-containing plastid with another, the form II enzyme is no longer present (Yoon, Hackett et al. 2005). Because only the peridinin-containing dinoflagellates contain the form II RuBisCO, if RuBisCO is an indicator of the plastid's origin, then the ancestral dinoflagellate chloroplast must been an ancestor of proteobacteria and cyanobacterium. There are no extant organisms known with this combination of proteins, making them singularly useless as phylogenetic markers. The last explanation that could be advanced is that dinoflagellates undergo lateral gene transfer to a previously unsuspected extent, effectively making all phylogeny of nuclear genes suspect.

As support of this latter hypothesis, the nuclear encoded chloroplast protein Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) also appears to have an unusual origin. The phylogeny of the chloroplast form of the enzyme places it outside the clade where most of the other plastid forms are found, and in with the clade corresponding to the cytoplasmic form(Fagan, Hastings et al. 1998). The marked similarity between cryptomonad and dinoflagellate chloroplast GAPDH sequences argues against both being derived independently from their respective cytoplasmic GAPDH homologs. The origin of GAPDH in cryptomonad is proposed to be proteobacterial, despite having a chloroplast evidently derived from cyanobacteria as indicated by sequences of several plastid-encoded genes, including 16S rRNA and tufA (Douglas 1993; Giovannoni, Wood et al. 1993; Delwiche, Kuhsel et al. 1995). The most likely scenario for acquisition of this gene involves lateral transfer from a cryptomand in the course of the ancestral association with dinoflagellates (Fagan, Hastings et al. 1998). Once again, nuclear-encoded plastid genes are inadequate to the task of determining the origin of the dinoflagellate plastid.

In 1999, nine chloroplast gene sequences were obtained from a peridinin-containing dinoflagellate *Heterocapsa triquetra* (Zhang, Green et al. 1999)

and all nine sequences were found in a rather unusual genomic structure. The sequences, encoding both ribosomal RNAs and seven photosystem protein components, were each located separately on a small plasmid-like minicircular chromosome of between 2 to 3 kb ("one gene-one circle"). Each circle was found to have an unusual tripartite non-coding region highly conserved among the nine circles. Phylogenetic analysis of the seven minicircle gene encoded proteins indicated that dinoflagellate chloroplasts are related to red algal chloroplasts, thus supporting their origin by secondary symbiosis (Zhang, Green et al. 1999; Zhang, Green et al. 2000).

Since the initial discovery, the unique organization of these chloroplast genes as minicircles has been confirmed in other peridinin-containing dinoflagellate plastids such as *Ceratium horridum*, *Pyrocystis lunula*, *Protoceratium reticulatum*, four *Heterocapsa* species, two *Amphidinidum*, and several *Symbiodinium*. All the minicircles have a conserved non-coding region (250 – 500 bp in size) that is highly conserved among minicircles within a species but shows little similarity to the conserved region in other species (Zhang, Green et al. 1999; Barbrook and Howe 2000; Hiller 2001). Some minicircles are now known to carry more than a single gene, but in these cases all genes have the same orientation with respect to the core region (Barbrook and Howe 2000; Hiller 2001). In addition, a wide range of anomalous minicircles which carry genes with small insertion or deletions have been reported (Zhang, Green et al. 1999; Zhang, Cavalier-Smith et al. 2002) as well as empty minicircles which carry a normal core region but do not contain any sizeable open reading frames or RNA genes (Barbrook, Symington et al. 2001; Hiller 2001). All core regions are AT-rich and have several inverted and direct repeats, and while their function is not known it has been suggested that they could act as origins of replication or be involved in recombination and gene conversion (figure I.3) (Zhang, Green et al. 1999; Zhang, Cavalier-Smith et al. 2002; Moore, Ferguson et al. 2003).

To date, only 18 plastid genes have been reported on minicircles. Two of the genes encode the 23S and 16S rRNAs, 12 of these genes encode proteins that are involved in the light reactions of photosynthesis, including components of photosystem I and II, the cytochrome b6f complex and the ATP synthase, as well as the genes encode two ribosomal proteins (rpl28, rpl33), and two unidentified ORFs (ycf 24, yct16) (table I.2). No tRNA encoding genes have been found. All of the genes found on minicircles are in the chloroplast genome in all plant species, suggesting that the dinoflagellates have one of the most reduced plastid genomes known.

The nature of these genes and their unusual structure suggested that they should be found in the plastid. This conclusion is supported by the fact none of the minicircle genes includes sequence encoding a recognizable transit peptide as defined by the known nuclear-encoded plastid directed gene products (Nassoury, Cappadocia et al. 2003). Furthermore, transcripts for psbA have been demonstrated in chloroplasts of *Symbiodinium* by in situ hybridization (Takishita, Ishida et al. 2003). Despite this, Laatsch and coworkers (Laatsch, Zauner et al. 2004) have described in situ hybridization and organelle fractionation studies showing the minicircle DNAs were located in nucleus of *Ceratium*. The fact that the coding region still shows no N-terminal extension that could serve as a signal sequence is unexplained.

Since the chloroplast function requires many more genes than the 18 encoded in the organelle, it was expected that the remaining genes should be located in the nucleus. Two groups, working independently, have reported studies using *Gonyaulax polyedra* and *Amphidinium carterae* (Bachvaroff, Concepcion et al. 2004) or *Alexandrium tamarense* (Hackett, Yoon et al. 2004) that show exactly this. Both groups have isolated many genes from cDNA libraries that previously were always found on the plastid genome of other algae and plants. These genes are clearly nuclear-encoded, as

deduced from their high GC-content, their codon usage and the presence of a signal peptide.

Why so many more genes should have been transferred to the nucleus, and why the plastid genome should have been fragmented in dinoflagellates are still unresolved questions. It also remains to be seen how widespread this organization is. To date, the dinoflagellates that were shown to have minicircles represent a limited number of peridinin-containing species. Other species have been reported to have a more conventional genome (Boczar, Liston et al. 1991), but these species have not been reexamined. There are suggestions that some intermediate patterns containing both minicircles and large circles, such as *Adenoides eludans* that has about 5 kb unigenic circles and 10 kb dimeric circles (Green 2004). Curiously, Southern blots of total DNA from *Gonyaulax grindleyi* using a psbA probe showed strong hybridization to a large (>12 kb) DNA and only weak hybridization to bands between 1.5 and 4 kb, and 5.8 kb (Zhang, Green et al. 1999). If the heavy labeling in the region above 12 kb is not due to trapping of minicircles in the bulk DNA, it would suggest the presence of psbA genes on a large circle. It is not known if this large DNA represents a large, multigenic circle that could possible reflect an ancestral state for chloroplasts.

Independent of the mechanism used to derive the minicircular plastid genome, it is important to note that there is precedence for this unusual genome organization. For example, single gene minicircles are also found in the mitochondria of *Dicyema*, one of a little-known group of miniature parasite animals. This mesozoan has separate mitochondrial minicircles of 1.6-1.7 kb carrying the genes for subunits I-III of cytochrome oxidase, apparently without a 'master' chromosome carrying all three (Watanabe, Bessho et al. 1999). Furthermore, the mitochondria of kinetoplasts contain both large maxicircles and many small minicircles used in editing. Lastly, the mitochondrial DNA of higher plants contains numerous multigenic circles; the heterogeneous

population of DNA has been proposed to be generated by recombination events between large and small repeats (Backert, Nielsen et al. 1997).

I.3 Expression and regulation of plastid genes

About thirty years ago it was well established that chloroplast contain their own DNA and translational apparatus, but little was known of how this organelle genome is expressed. Interestingly, as a result of their endosymbiotic origin plastids contain a protein synthesizing system that displays several prokaryotic features. Its 70S ribosomes resemble those of bacteria and are sensitive to the same set of antibiotics. However, it has become apparent that the chloroplast gene-expression system is unique, differing in many respects from bacterial systems (Rochaix 2001) Gene expression is potentially amenable to regulation at different points along the linear path leading from gene to functional protein, to produce measurable differences in gene-expression, the regulatory event has to occur at a rate-limiting step. The vast majority of *C. reinhardtii* chloroplast encoded genes are involved in gene expression and photosynthesis, and are primarily regulated post-transcriptionally, including RNA processing, transcript stability, translation, protein turnover, and protein activity (Manuell, Beligni et al. 2004).

I.3.1 Chloroplast genes encoding RNAs and proteins

During the late 1970s and early 1980s, the use of techniques such as DNA-RNA hybridization and in vivo labeling, as well as 2D PAGE of soluble and membrane proteins coupled with immunological techniques identified the major RNA and protein species in chloroplasts. Depending on their function, these plastid genes can be grouped into three categories, (i) genes encoding elements of the genetic system, (ii) genes encoding photosynthesis components and (iii) other protein coding genes, including those involved in biosynthetic or other diverse functions. In higher plants, there are genes for

four different rRNAs: 16S, 23S, 4.5S and 5S. The 4.5S rRNA is unique to land plants whereas two small rRNAs, 3S and 7S were detected in *Chlamydomonas reinhardtii* (Sugiura 1996). Of the estimated 60 polypeptides in plastid ribosomes, only 21 have been found to be encoded on cpDNA of land plants (Subramanian 1993). The nucleotide sequencing of several chloroplast genomes has revealed the presence of between 27-35 tRNA genes. Angiosperm plastids contain 30 tRNA genes (Sugiura 1996), while *Euglena* and *Porphyra* possess the maximum and the minimum number of tRNA genes, respectively (Hallick, Hong et al. 1993).

Some of the major proteins coded in the chloroplast genome are listed in table I.1 (He and Malkin 2000; Lawlor 2001). The rbcL gene encodes the large subunit (LSU) of Ribulose- 1,5 –bisphosphate carboxylase/oxygenase (RuBisCO), the major enzyme component of the chloroplast (~ 40-60% of the total soluble protein of leaves). In higher plants, the enzyme is a heterohexadecameric complex of eight large and eight small subunits (LSU and SSU). The LSU is encoded by cpDNA in a sequence of 1431 base pairs encoding 477 amino acids. The group of photosynthetic genes codes for about half of the thylakoid polypeptides, and together with an equal number of nuclear-encoded gene products, form four supra-molecular complexes, photosystem I (PSI, psa genes), photosystem II (PSII, psb genes), cytochrome b6/f complex (pet genes) and the ATP synthase (atp genes). Land plants chloroplast genes encode 5-6 components of PSI, 12-13 components of PSII, 4 components of Cytb6/f complex and 6 subunits of ATP synthase complex. The ATP synthase (CF1-CF0) complex has 6 of its nine subunits encoded by the cpDNA in almost all species (table I.1).

I.3.2. Photosynthetic system protein components and their function

Photosynthesis in majority of oxygen-evolving organisms involves the sequential co-operation between the PSI, the cytochrome complex and PSII.

This so-called non-cyclic electron transport, also termed the Z-scheme, takes the electrons released from water and passes them from PSII to PSI, thus generating the strong reductant necessary for NADP+ reduction. The electron transport between the two photosystems is bridged by the cytochrome b6-f complex, which is evenly distributed throughout the thylakoid membranes, and two freely diffusing electron carriers, PQ and PC. An additional protein complex, the ATP synthase, is found only in non-appressed regions of the thylakoid membranes. Driven by the H+ gradient built up across the membrane by electron transfer reaction, the ATP synthase converts ADP into ATP. PSI and the cytochrome complex can also function independently from PSII in what is termed the cyclic electron transport pathway, where electrons circulate repeatedly through the cytochrome complex in order to augment the pH gradient.

Each photosystem can be resolved into a photoactive core complex, where light energy induces charge separation in the reactive chlorophyll pigments, and a light-harvesting complex (LHC), which increases the surface area available for light capture. The core complex contains a number of protein subunits that are involved in either binding electron carriers, other pigments or are involved in interactions with other electron transfer components. PSII is the most complex multimeric protein assembly in thylakoid membranes, with at least 15 subunits in the core complex of cyanobacteria and 18 subunits in that of higher plants. Invariably, PSII contains the reaction centre components D1 and D2 (genes psbA and psbD), subunits of cyt b559 (psbE), as well as the CP47 (psbB), CP43 (psbC) and an extrinsic 33 kDa component (psbO). All of these subunits are required for the water oxidation (table I.1 and figure I.4) (He and Malkin 2000; Lawlor 2001). The structure of PSII at atomic resolution has not yet been determined, although several groups are actively attempting to obtain high-quality crystals for structural analysis. However, biochemical fractional and chemical cross-linking studies have provided a good deal of information regarding the

superamolecular organization of the complex. This has in turn led to models for protein subunit organization that propose the D1 and D2 subunits at the core of PSII. The α- and β-subunits of cyt b559, the PsbW and PsbI subunits as well as the CP47, CP43 and the PsbO protein are tightly associated with D1-D2 (figure I.4) (Hall and Rao 1994).

PSI is also an integral membrane complex, and in prokaryotes it contains at least 11 different polypeptides. In the complex isolated from higher plants, at least three additional non-chlorophyll binding subunits are present, in addition to a substantial number of chlorophyll-binding subunits that bind antenna. All PSI core complexes contain two relatively high molecular mass subunits of approximately 83 kDa (psaA, psaB), present as single copies to form a PsaA-PsaB heterodimer, which binds the special pair of chlorophylls (P700). The heterodimer also binds A0, A1 and Fx as well as approximately 120 chlorophyll a molecules (table I.1). The terminal electron acceptors in PSI, FA and FB, are bound by a low molecular mass protein subunit of approximately 9 kDa, known as PsaC. Two subunits, PsaD and PsaF, have been implicated in the interaction of the PSI complex with its soluble electron transfer partners, Fd and Fc respectively. The PsaE subunit has been implicated in binding Fd. Other intrinsic polypeptides include products of genes psaI, J, K and I, their functions are unclear (figure I. 4) (He and Malkin 2000; Lawlor 2001).

I.3.3 Transcriptional and post-transcriptional control of plastid mRNA

As in the nucleus, one would expect the plastid gene expression to be regulated at transcriptional, post-transcriptional and translated levels. Although cpDNA codes for several basic components for gene expression, it must rely on the nuclear genes for encoding the major portion of the components involved in its biogenesis (which are largely unknown) and photosynthetic machinery. There is clearly control over plastid biogenesis, as

the number of plastids in a cell and copies of the cpDNA in each plastid can vary as a function of external or internal cues. However, faithful in vitro reproduction of the molecular processes involved in plastid gene expression has been difficult despite the fact that systems for plastid transformation and in vitro translation have been available in tobacco for about 10 years (Sexton, Christopher et al. 1990; Svab and Maliga 1993; Hirose, Fan et al. 1996; Hirose and Sugiura 1996).

In many ways, chloroplast protein-coding genes look like prokaryotic genes. They have relatively simple promoters with -10(TATAA) and -35 (GTGACA) *cis*-acting elements similar in sequence and function to bacterial promoter elements. The 3' ends of the genes are also reminiscent of bacterial genes in that they usually have a sequence able to form a 3'-terminal stem loop when transcribed into RNA. Despite these similarities, the 5' and 3' ends of chloroplast genes also have some distinctive characteristics(Zerges 2000). For example, the psbD 5' ends of barly chloroplast RNAs that arise from a single transcription unit can be heterogeneous because of multiple initiation sites and by processing of the primary transcripts (Sexton, Christopher et al. 1990).

Although in vivo determination of mRNA half-lives and transcription rate in plastids is difficult, the plastid run-on transcription assay can be used to determine the relationship between changes in the transcription activity of individual genes and the accumulation of their RNAs. The prevailing view in the late 1980s was that most of the chloroplast genes are constitutively expressed and major controls for plastid gene expression lie mainly at the post-transcriptional level. While it is now recognized that factors such as the topology of cpDNA (Stirdivant, Crossland et al. 1985), cpDNA methylation (Kobayashi, Ngernprasirtsiri et al. 1990), cpDNA phosphorylation (Tiller and Link 1993),regulation of copy number (Rapp, Baumgartner et al. 1992) and differential promoter usage by different RNA and/or specific regulatory

proteins (Sexton, Christopher et al. 1990; Kapoor, Suzuki et al. 1997) can influence plastid gene expression in response to light and developmental conditions, as a general rule transcriptional regulation is not of primary importance. For example, there are small changes in the relative transcription of the spinach psbA, rbcL and atpB genes and the gene for 16S rRNA in barley after 24 hours of illumination and during leaf expansion, but these are not sufficient to explain the differential accumulation of their mRNAs relative to ribosomal RNA. Instead, these studies support the hypothesis that post-transcriptional control of plastid gene expression is the primary mechanism by which different accumulation of most mRNA occurs during chloroplast development (Klein and Mullet 1987; Gruissem, Barkan et al. 1988). Furthermore, changes in transcriptional regulation alone can not account for the strong changes of protein amount observed (Mayfield, Yohn et al. 1995).

I.3.3.1 Regulation at the post-transcriptional level

Post-transcriptional regulation can occur at a variety of levels, including processing, degradation, splicing and editing of pre-RNAs. However, the extent to which these processes influence plastid gene expression is debatable. RNA processing is a general term to describe the modification of newly synthesized RNA molecule that form the mature 5' and 3' end. RNA degradation is important in RNA metabolism by recycling the nucleotides of non-functional molecules. Lastly, many chloroplast genes contain introns, which must be spliced.

I.3.3.1.1 RNA processing

The earliest recognized signal for mRNA processing was the highly conserved eukaryotic cleavage and polyadenylation signal AAUAAA (Proudfoot and Brownlee 1976). This vital signal allows formation of the 3'

ends of nuclear-encoded mRNA by cleavage of a longer primary transcript followed by polyadenylation of the newly liberated 3' end. Generally, the AAUAAA sequence is located between 11 and 30 nucleotides upstream of the site of poly(A) addition. The signal is needed for both cleavage and polyadenylation (Conway and Wickens 1985; Wilusz, Pettine et al. 1989; Sheets, Ogg et al. 1990), as deletion or mutation of the AAUAAA hexamer prevents polyadenylation. Generation of the proper 3' terminal structure requires a battery of enzyme activities, including an endonuclease to cleave the RNA, a poly(A) polymerase (PAP) to synthesize the poly(A) tail, and a specificity component (CPSF) that recognizes the AAUAAA sequence and coordinates the activities of the other enzymes. In addition to these factors, a stimulatory activity factor, CstF, binds to a U-rich or G-U-rich sequence downstream from the cleavage site itself (MacDonald and Redondo 2002).

Polyadenylation is an important determination of mRNA function in eukaryotes, and may affect both stability and initiation of translation (Jacobson 1996). Interestingly, poly(A) tails have also been described for several mRNAs in bacteria. For example, in the rpsO mRNA of *E. coli*, they are found at endonucleolytic and exonucleolytic cleavage sites. However, in contrast with the nuclear transcripts of eukaryotic cells, where deadenylation of the long poly(A) tail is part of the mRNA degradation pathway, the addition of poly(A) tails in bacterial mRNAs promotes their degradation. In this case, degradation is thought to be facilitating by assisting exonuclease binding. The sites of polyadenylation of bacterial mRNA are diverse and include the 3'-ends of primary transcripts, the sites of endonucleolytic processing in 3' untranslated and intercistronic regions, and sites within the coding regions of mRNA degradation products. This diversity of sites suggests that mRNA polyadenylation in prokaryotes is a relatively indiscriminate process that can occur at all mRNA's 3' ends and does not require specific consensus sequences as was the case for eukaryotes (Haugel-Nielsen, Hajnsdorf et al. 1996).

The degradation of RNA in prokaryotes relies heavily on two enzymes, RNase II and polynucleotide phosphorylase (PNPase)(Nickelsen 2003). The 3' end of many bacterial mRNAs, such as those formed by Rho-independent termination, are normally sequestered in stem-loop structures that protect them from degradation in a 3' →5' pathway. Nascent transcripts also have protected 3' ends. Endonucleolytic cleavage that removes 3' stem-loops or produces single-stranded 3' ends are substrates for the 3' →5' exonucleases. For example, the single-specific endonuclease RNase E is essential for rRNA processing in *E. coli*. Interestingly, a large multiprotein complex, now called the RNA degradosome, was discovered during the purification of *E. coli* RNase E. The major components of the RNA degradosome include RNase E, PNPase and the DEAD-box RNA helicase B. The association of RNase E and PNPase in a complex provides a direct physical link for their observed cooperation in the degradation of mRNA(Carpousis 2002).

Similar to the bacterial mRNA, many higher plant ang green algae chloroplast mRNAs also conta(Walter, Kilian et al. 2002) in an AU-rich 3'-UTR with a terminal inverted repeat that can fold into a stem-loop structure. Chloroplast 3'UTRs are important for RNA processing and the accumulation of stable mRNA(Hayes, Kudla et al. 1996). While these stem-loops function in Rho–independent termination of transcription in bacteria, in chloroplasts, termination of transcription is a relatively inefficient process that produces additional nucleotides at 3' end. These extra nucleotides are removed during 3' end processing (Monde, Greene et al. 2000) and the stem-loop structures appear to act as barriers to prevent 3'→5' degradation from encroaching on the coding sequence. They thus function as general transcript-stabilizing determinants and, in addition, serve as processing signals for correct 3' end formation. The formation of the mature 3' end is a two-step process, which involves an initial endonucleolytic cut downstream of the stem-loop structure and subsequent exonucleolytic trimming in the 3'→5' direction. Chloroplast

transcripts are similar to those in *E. coli* in that polyadenylation of the resulting cleavage products at their 3' end targets them for degradation (Hayes, Kudla et al. 1999; Schuster, Lisitsky et al. 1999). Indeed, polyadenylation has been shown to accelerate degradation of chloroplast RNAs (Kudla, Hayes et al. 1996; Lisitsky, Klaff et al. 1996). However, while a poly(A)polymerase (PAP) performs this function in *E. coli*, the exoribonuclease polynucleotide phosphorylase (PNPase) is used for polyadenylation of RNAs in the plastids (Hayes, Kudla et al. 1996). Biochemical data suggests that PNPase acts in polyadenylation in both higher plant plastids and cyanobacteria, as might be expected based on the endosymbiotic origin of plastids. Furthermore, PNPase catalysed polyadenylation results in heterogeneous poly(A)-rich tail (Yehudai-Resheff, Hirsh et al. 2001; Rott, Zipor et al. 2003) as this enzyme is less specific for adenine residues than is the PAP. Chloroplast PNPase apparently does not form a degradosome-like structure, but instead forms a homo-oligomeric complex (Baginsky, Shteiman-Kotler et al. 2001), although as for *E. coli*, the sites of polyadenylation of plastid mRNA are diverse. Reverse genetic approach reveals a more complex picture of the function of PNPase in the chloroplast of *A. thaliana*. In cosupressed plant exhibiting a drastically reduced amount PNPase, correct 3' end formation of rbcL and psbA mRNA was significantly reduced. However, their total transcript accumulation remained unaffected by the absence of PNPase. These data indicate that PNPase represents a critical determinant for the 3' processing of plastid mRNAs but not for their half-lives.

I.3.3.1.2 RNA splicing

The presence of split genes is a common feature of plastid DNA. For example in the tobacco plastome, 18 genes contain introns that vary in size from 0.5 to 2.5 kb. Clearly, intron containing plastid transcripts must undergo correct excision of the intervening sequence before they can become

translationally functional. The mechanism of plastid splicing in higher plants is largely unknown. However, most of the introns found in higher plant plastids are group II type intron, a type of intron common in organelles and in bacteria. Group II introns have the remarkable ability to excise themselves from the RNA and are termed self-splicing. Despite the fact that RNA alone can perform the splicing reaction in vitro, proteins are almost certainly required in vivo to assist with folding. Tissue specific differences in splicing have been reported in maize and rice, and to date, five nuclear genes required for chloroplast splicing have been isolated by genetic means. These include Raa1, Raa2, Raa3 ,Crs1 and crs2 (Vogel, Borner et al. 1999). However, due to lack of any plastid in vitro splicing system, delineation of the *cis* and *trans* factors involved in these processes has still not been possible and no direct contact between the proteins and respective transcripts has been reported. Thus, the precise *cis*-acting RNA element involved and, accordingly, the precise molecular working mode of these proteins remains to be elucidated (reviewed by Nickelsen, 2003). Introns with introns (Copertino and Hallick 1991) as well as ORFs (including those coding for endonucleases or RNA splicing maturases) within introns (Saldanha, Mohr et al. 1993) have also been found in plastid introns. The plastid splicing is an important post-transcriptional process; but more work will be required to ascertain its role in the regulation of plastid gene expression.

I.3.3.1.3 RNA editing

RNA editing is a post-transcriptional process involved in transcript maturation in a variety of organisms, including virus, fungi, plants, and mammals (Smith, Gott et al. 1997). It is characterized by differences between the sequence of the RNA transcript and the DNA sequence from which it was transcribed. Various types of editing are possible, and can be loosely divided into two general classes involving either insertion/deletion of nucleotides, or the substitution and modification of nucleotides. In insertion type editing,

where nucleotides are inserted into (or deleted from) the single stranded RNA molecule, the position of an insertion is not specified in the gene sequence from which the RNA has been transcribed. In trypanosome mitochondria, where in fact the phenomenon was first reported, more than half of the residues in the mRNA consist of non-coded uridine residues. Comparisons between the genomic DNA and the mRNA shows that no stretch longer than 7 nucleotides is represented in the mRNA without alteration and runs of uridine up to 7 bases long can be inserted. The mechanism involves a guide RNA, a small RNA sequence complementary to the correctly edited mRNA and flanking regions of the unedited transcript, that provides the information for the specific insertion of uridines. The insertion of U residues at sites in the mRNA defined by the guide RNA occurs by cleavage of the RNA, addition of the U, and religation of the ends. The enzyme 3' terminal uridylyl transferase (TUTase) is thought to be involved in this reaction (Benne, Van den Burg et al. 1986; Feagin, Abraham et al. 1988; Aphasizhev, Sbicego et al. 2002). Insertional RNA editing also includes the nontemplated cotranscriptional insertion of nucleotides in certain virus, which is specified by the surrounding sequence.

In substitution editing, the changes principally observed involve deaminations, either of cytosine (resulting in uridine) or of adenine (resulting in inosine). Less widely distributed changes are uridine to cytosine, and guanine to adenine conversions(Cattaneo 1991). This type of RNA editing has been identified in both mitochondria and chloroplasts of land plants, where almost all changes involve conversion from cytidine to uridine and only in rare cases from uridine to cytosine by base modification (Yu and Schuster 1995; Hirose, Fan et al. 1996). In higher plants, essentially all protein-encoding mRNA is subject to editing in mitochondria but the phenomenon is significantly less frequent in chloroplast, where it affects only some transcripts (Hirose, Fan et al. 1996). RNA editing was first documented in plant chloroplast in 1991 (Hoch, Maier et al. 1991). At present, 34 C-to-U editing

sites have been found in tobacco chloroplast transcripts (Sasaki, Yukawa et al. 2003), 21 in rice (Corneille, Lutz et al. 2000)and 19 in *Arabidopsis* (Lutz and Maliga 2001). Analysis of chloroplast transcripts from the fern *Adiantum capillus-veneris* show that editing is much more prevalent, with 350 RNA editing sites observed. Roughly 90% of these editing sites were C-to-U with the remainder U-to-C (Wolf, Rowe et al. 2004). In this type of editing, sequences surrounding editing sites share no common characteristics, except that most of the 5' neighboring residues are pyrimidines and the 3' neighbors are adenines (Bock, Hermann et al. 1996; Chaudhuri and Maliga 1996). Curiously, while RNA editing has been found in chloroplast transcripts from all major lineages of land plants, neither editing frequencies nor the editing pattern of a specific transcript correlates with the phylogenetic tree of the plant kingdom (Freyer, Kiefer-Meyer et al. 1997). RNA editing occurs mostly in protein-coding regions and generally restores evolutionarily conserved amino acid sequences (Freyer, Kiefer-Meyer et al. 1997; Bock 2001), although some of the editing events cause diversity in amino acid sequence (Inada, Sasaki et al. 2004). Recently, in vitro RNA-editing systems for both tobacco and pea chloroplast were developed, which now permit identification of the *cis*-elements and, furthermore, the proteins recognizing these RNA regions using by UV-cross linking techniques. In these editing extracts, four different RNA binding proteins (RBPs) were detected all binding to different RNA editing elements. A 25-kDa protein specifically recognized the psbL-editing site. In tobacco, proteins of 56 kDa and 70 kDa were shown to bind to the psbE and petB sites, respectively. Interestingly, pea chloroplasts have no editing capacity for the psbE site and lack the 56-kDa protein, suggesting a co-evolution of editing sites and their cognate editing factors (Nickelsen 2003).

I.3.4 Translation and degradation of chloroplast proteins

Translational regulation provides a capacity to rapidly induce high levels of protein synthesis from an existing pool of transcripts. What might be the special requirements in the mature chloroplast that made translational control a favored mechanism for regulating gene expression? First, as light is the primary energy source of plants, the chemical energy in the form of ATP and reducing power produced by various complexes in the thylakoid membranes drives all the different stromal reactions, such as carbon fixation, nitrogen reduction, and biosynthesis of amino acids and lipids. Consequently, the availability and intensity of light have profound effects on chloroplast metabolism and on chloroplast gene expression. Of all the processes involved in the regulation of gene expression, the step of translation seems to be affected most severely by light and developmental signals. The accumulation of some of the proteins may increase 100 to 10,000 fold during light-induced greening of plastids (Rochaix 1992; Gillham, Boynton et al. 1994; Mayfield, Yohn et al. 1995). Second, light absorbance by the photosystem components is accompanied by deleterious side effects that lead to photoinactivation of the photosystem II reaction center and concurrent protein turnover (Mattoo, Pick et al. 1981; Wettern and Ohad 1984; Barber and Andersson 1992). The most actively synthesized protein D1, in spite of presence of abundant respective message, are not synthesized in dark growth condition.

I.3.4.1 Translation of chloroplast proteins

In the nucleus of eukaryotes, the initiation complex forms at the m^7 G-capped 5' terminus of mRNA then scans the 5'-UTR for the correct initiation codon. The sequence surrounding the AUG, termed the context, determines the efficiency of translation initiation. The scanning process is dependent on the initiation complex intrinsic helicase activity and requires energy in the

form of ATP. Secondary structures in the 5'-UTR and binding of specific proteins to the 5'-UTR can also affect the rate of initiation of translation (Kozak 1992).

In contrast, translation initiation in prokaryotic systems is largely dependent on base pairing between the 16S rRNA and 5'-UTR of mRNAs. A specific sequence known as Shine-Dalgarno (SD) sequence in the 5'-UTR is the most important. This sequence is a polypurine stretch within 10 bases upstream of the AUG, whose consensus sequence is 5'-AGGAGG-3'. It is complementary to a highly conserved sequence close to the 3' end of 16S rRNA termed the RNA binding sequence (RBS). The SD-RBS localizes the initiation complex at the correct initiator codon and usually determines the rate of initiation of translation. In contrast to initiation of translation in eukaryotes, this process in prokaryotes is not dependent on helicase activity and does not require ATP hydrolysis. The direct RBS binding and internal initiation of translation in prokaryotes is usually independent of upstream sequence and allows for the simultaneous translation of several open reading frames (ORFS) in a polycistronic transcript (Danon 1997).

Most of the structural components involved in the plastid translation machinery, such as the 70S ribosomes, the uncapped transcripts, and the presence of SD-like sequence, are similar to those in prokaryotes. Furthermore, factors equivalent to prokaryotic initiation factors IF-2 and IF-3 have also been identified in *Euglena* chloroplasts and a chloroplast gene with sequence similarity to the gene encoding IF-1 has been found in land plants. Despite these remarkable similarities, plastid translation also has several distinct features. In particular, plastids lack any apparent coupling between transcription and translation and the spacing between the SD-like sequences and initiator codon is less stringent than in prokaryotes. Indeed, in the genome of higher plant chloroplasts, about one-third of protein coding genes do not contain SD-like sequences within the 25 nucleotides immediately

upstream from the initiation codon (Kapoor and Sugiura 2000). The absence of SD-RBS in the large fraction of chloroplast mRNA suggests that binding of the initiation complex to the mRNAs is mediated by an alternative type of cap-independent-binding domain. Since sequence complementarities with a small ribosomal subunit may not be required for binding, the binding of the initiation complex could be facilitated by proteins interacting with this domain. The finding of upstream 5'-UTR *cis*-elements that are required for translation may help to identify these putative initiation complex-binding domains. In vivo expression of chloroplast genes truncated or mutated in the 5'-UTR demonstrated that the interaction of each of the *cis*-elements with an mRNA specific nuclear encoded factor is crucial for mRNA translation. This is independent of the ORF downstream of the 5'-UTR, suggesting that initiation is the checkpoint for regulating translation (Danon 1997). To identify factors regulating plastid translation initiation at 5'UTRs, which is considered to be the rate-limiting step for polypeptide synthesis, mainly in vitro analyses have been applied in the past. During extensive RNA-binding experiments using a variety of different systems, a large number of RNA binding proteins (RBPs) were detected, which interact with chloroplast leader regions. Moreover, a promising in vitro system based on a tobacco chloroplast protein extract has been developed, which is capable of accurately translating exogenously added mRNA (Hirose and Sugiura, 1996). The initiation codon for translation of plastid mRNA is usually AUG, although GUG is used in a few cases (Sugiura, Hirose et al. 1998; Zerges 2000). Furthermore, individual mRNAs appear to differ in their requirement for AUG initiation codons. In Chlamydomonas, mutation of the initation codon of the chloroplast petD gene from AUG to ATT or ATA did not abolish translation of this mRNA in vivo, while similar changes of the initiation codon of the psbD mRNA did abolish its translation (Zerges 2000).

Light dramatically stimulates the translation of several chloroplast mRNAs in plants and algae. While the D1 protein shows the highest induction by light,

the translation of psbD, rbcL, psaA and psaB has been also shown to be light regulated. Both initiation and elongation steps of translation are probably controlled by light (Formm, Devic et al. 1985; Klein and Mullet 1987; Klein, Mason et al. 1988; Malnoe, Mayfield et al. 1988; Krupinska and Apel 1989; Berry, Breiding et al. 1990). The stimulation of *Chlamydomonas reinhardtii* chloroplast psbA translation by light has been proposed to be mediated by a 47 kDa protein called RB47, which binds to the 5' UTR of the psbA mRNA. Binding to the mRNA is regulated by two light-responsive molecular mechanisms, ADP-dependent phosphorylation and redox potential. Light thus modulates binding of RB47 to RNA through effects on photosynthesis, particularly on the levels of ADP and reduced thioredoxin in the chloroplast stroma (Danon and Mayfield 1994; Kim and Mayfield 1997). RB47 appears to act as a complex, whose other members include a 60kDa protein (RB60) whose phosphorylation state correlates with redox state and ADP levels. ADP-dependent phosphorylation of RB60 results in an inactive psbA 5'-end binding complex; this occurs at ADP levels found in chloroplasts only in the dark (Danon and Mayfield 1994). In vitro, binding to the 5'end of psbA can also be modulated by redox state, suggesting that the complex contains a redox-responsive site. Further studies showed that RB60 carries a redox-active regulatory site composed of vicinal dithiol. When the redox state of RB60 and translation of psbA mRNA were assayed in parallel in intact chloroplasts, high light intensities were found to increase both the pool of reduced RB60 and the rate of psbA translation. The specific reduction of RB60 during the light probably involves the ferredoxin-thioredoxin system, as chemical reduction of the vicinal dithiol in the dark did not activate translation (Trebitsh, Levitan et al. 2000).

The D1 protein encoded by plastid psbA gene is translated on thylakoid-bound ribosomes. Mature D1 has five transmembrane domains with the N-terminus on the stromal side of the thylakoid membrane. As one of the most photolabile photosystem components, the increased synthesis of D1 in the

light is of great importance. In addition to the control exerted at the level of initiation of translation, the elongation of D1 is also strictly regulated. Only upon illumination did the full length D1 start accumulating without considerable increase of psbA mRNA in polysomes (Klein and Mullet 1987; Klein and Mullet 1990; Mullet, Klein et al. 1990; Kim, Eichacker et al. 1994; Kim and Mayfield 1997). Effect translation elongation of psbA mRNA is dependent on optimal photosynthetic electron transport, particularly on the production of reducing power (Kuroda, Kobashi et al. 1996; Zhang, Paakkarinen et al. 2000).

I.3.4.2 Degradation of chloroplast proteins

Protein degradation is an essential component of normal cellular homeostasis. During metabolism in a cell, proteins are damaged by metabolic products or abiotic stress and must be degraded. Protein degradation is also essential to remove regulatory proteins or enzymes that are no longer required. Thus protein degradation is an essential life and death issue for a cell. The major sites of intracellular proteolysis in eukaryotic cells are the proteasomes in cytosol and the lysosomes. Proteins in semi-autonomous organelles are presumably degraded inside those organelles.

Protein degradation is a major factor regulating levels of many chloroplast proteins. Environmental stress, nutrient deficiency and senescence cause rapid degradation of chloroplast proteins. Oxidative stress causes rapid degradation of RuBisCO, while high temperature stress causes rapid protein degradation in chloroplast (Chitnis 2000). One major research focus has been the degradation and replacement of D1 protein damaged by light stress. Light stress to PSII results in photooxidative damage to the reaction center and the inactivation of electron transport. The plants survive this light stress by a repair mechanism in which damaged D1 protein is replaced by a new

protein copy. Thus, the turnover rate of the D1 protein is several orders of magnitude higher than any other chloroplast protein.

The rate of D1 protein degradation is not determined solely by the rate of PSII photodamage but also appears to be under the control of a regulatory mechanism that serves to keep the rate of D1 protein degradation in balance with the insertion of newly synthesized D1 protein into the thylakoid membrane. This mechanism may thus avoid total disassembly of PS II complex (Aro, Virgin et al. 1993). In higher plants, D1 protein degradation seems to involve a phosphorylated form. The effect of protein phosphorylation and dephosphorylation upon high-light-induced degradation of the photosystem II reaction center proteins D1 and D2 have been studied in isolated thylakoid membranes. The degradation of the phosphorylated D1 protein is drastically reduced under conditions that induce either acceptor- or donor- side photoinhibition of photosystem II. The stability of the phosphorylated D1 protein is further increased under conditions of reduced phosphatase activity, suggesting that phosphorylated and damaged D1 protein has to be dephosphorylated prior to proteolytic degradation. Degradation of the D1 and D2 proteins of photosystem II is regulated by reversible phosphorylation (Koivuniemi, Aro et al. 1995), and interestingly, the phosphorylation itself is a circadian-regulated process that can take several hours to reach completion (Booij-James, Swegle et al. 2002). Investigation of the requirements for nucleotides and cofactors during the light-induced proteolysis of the D1 protein by isolated chloroplast thylakoid membranes revealed that primary cleavage of the D1 protein is a GTP-dependent process, where the secondary cleavage steps require ATP and Zinc ions (Spetea, Hundal et al. 1999). In vivo study with chloroplast also showed that the truncated D1 protein degradation is ATP dependent (Preiss, Schrader et al. 2001).

I.3.5 Gene expression in dinoflagellate chloroplasts

Despite the fact that dinoflagellate plastid genes were only identified a few years ago, their unique organization within the chloroplast genome has rapidly become apparent. The peridinin plastids appear to have a highly reduced genome encoding only about 12-18 genes on small plasmid-like minicircles that can contain from 1 to 3 genes, in addition to minicircles encoding pseudo genes (Zhang, Green et al. 1999; Barbrook and Howe 2000; Howe, Barbrook et al. 2000; Zhang, Cavalier-Smith et al. 2002). Minicircles are unusual because plant and algal plastids generally contain a circular genome of about 150 kb in size that encodes between 100 to 200 genes. The minicircles within a species all share a common core sequence, a practical feature that has been exploited in the identification of minicircle genes. However, these common features are not conserved as one moves across taxa so their significance and role is still unlear. The known genes identified on minicircles encode the core subunits of photosystem (atpA, atpB, petB, petD, psaA, psaB, psbA-D and sometimes psbE and psbI) and rRNA (16S and 23S rRNA). The remaining genes required for plastid function appear to have been transferred to the nucleus as many of them have been recovered in cDNA libraries constructed from *Alexandrium tamarense*, *Amphidinium carterae*, and *Lingulodinium polyedrum* (Bachvaroff, Concepcion et al. 2004; Hackett, Anderson et al. 2004). These nuclear-encoded genes are GC-rich, unlike the plastid sequences and encode a targeting sequence for plastid import (Nassoury, Cappadocia et al. 2003; Bachvaroff, Concepcion et al. 2004; Hackett, Anderson et al. 2004). Curiously, the location of minicircle genes in different dinoflagellate species remains under debate, as while the minicircles have been demonstrated in the plastid of *Symbiodinium* by in situ hybridization, they have been reported in the nucleus of *Ceratium horridum* after cell fractionation (Laatsch, Zauner et al. 2004). Minicircles are provisionally also in the nucleus of *Pyrocysis lunula* (rpl28 and rpl33) and *Ceratium horridum* (ycf16, ycf24), (unpublished data; see GeneBank

accession number AF490367 and AF490364) (table I.2). In *Gonyaulax*, the synthesis of the plastid protein psbA is sensitive to chloramphenicol and not to cycloheximide, indicating that at least psbA translation occurs within the plastid (Wang, Jensen et al. 2005). Lastly, none of the minicircle genes sequenced to date encodes the characteristic leader sequence believed to be required for targeting to the triple membrane-bound plastid, so that a nuclear location for the minicircles would be expected to preclude them from encoding functional plastid proteins.

The minicircle genes display a number of anomalous features. Several have unusual predicted translational start codons, including GUA in *Amphidinium* and AUA in *Heterocapsa,* and these are apparently not corrected through post-transcriptional modification such as mRNA editing (Koumandou, Nisbet et al. 2004). Furthermore, in many minicircles there is hardly any intervening sequence between the ends of the conserved core region and the ends of the coding sequence. This suggests that the core region may play a role in transcription initiation, although no promoter has been identified. It appears probable that minicircle genes are transcribed and translated, as in situ hybridization with a psbA probe in *Symbiodinium* indicated that psbA mRNA was on a ribosome bound to thylakoid membranes. Plastids contain both thylakoid membrane-bound and free ribosomes, and since the D1 protein encoded by psbA must be synthesized on membrane-associated ribosomes in order to be cotranslationally inserted into thylakoid membranes, a thylakoid location of the message is strong presumptive evidence for its translation (Takashita, Ishikura et al. 2003). To date, however, there is little direct experimental evidence to show the translation of other minicircle gene transcripts.

The reports concerning post-transcriptional RNA modifications are limited to RNA editing as detected by RT-PCR, and are very divergent. While no mRNA editing or other changes in *Amphidinium operculatum* minicircle gene

transcripts were detected (Barbrook, Symington et al. 2001), extensive substitutional editing of plastid transcripts was reported in the dinoflagellate *Ceratium horridum*. In this case, editing changes were changes from adenine to guanine, although a previously unreported adenine to cytosine transversion was also reported (Zauner, Greilinger et al. 2004). Since the prokaryotic nature of the chloroplast transcriptional and translational machinery require mechanisms different from eukaryotic nature of the nucleus, the research on minicircle genes transcription and translation would help answer this question. An elegant and conclusive resolution can be obtained by determining the antibiotic sensitivity of protein synthesis derived from the minicircle genes. The different tail of chloroplast RNAs caused by post-transcriptional RNA processing could be a further helper.

1.4 Research Goals

The most generally found type of dinoflagellate chloroplasts are the so-called peridinin-containing plastids, and differ from all other known organelles because of a combination of unusual features: the use of the carotenoid peridinin as an accessory light-harvesting pigment and the presence of a unique peridinin-binding protein, the use of a form II Rubisco and the presence of three envelope membranes. Furthermore, in recent years, plastid genes have been found on small (2-3 kb), generally unigenic, minicircles instead of a large (about 100-200 kb) circle that encodes about 100 -200 genes as found in higher plant plastids. To date, a total of 18 minicircle genes have been reported (although not all have been found in any given species), and these include two ATP synthase subunits, two photosystem I proteins, six photosystem II proteins, two ribosomal RNAs, two ribosomal proteins and two unidentified open reading frame proteins. No tRNAs have yet been reported in any of the minicircles.

The unusual features of the plastids have led to the formulation of three major questions:

1. Is the minicircle the only genome structure found in peridinin-containing dinoflagellate plastids?
2. Are additional dinoflagellate plastid protein coding genes found in non-minicircle formats?
3. Are dinoflagellate plastid genes really located and translated in the plastids ?

I have addressed these questions using the peridinin-contaning dinoflagellate *Gonyaulax polyedra* (now *Lingulodinium polyedrum*) as research material. To answer the question of the plastid genome architecture, I used a combination of pulse field electrophoresis and Southern blotting with a plastid DNA probe, and was able to demonstrate that at least one *Gonyaulax* plastid gene is found not on a mini-circle but on a DNA of roughly 100-150 kb, more typical of higher plant plastid genomes. The results of these experiments are reported in Chapter 2.

To answer the second question, I have identified and characterized an unusual 3' end modification specific for plastid encoded gene transcripts. This modification, polyuridylylation, has allowed the plastid transcriptome to be charcacterized. My results demonstrate that the only genes expressed in the plastid are likely to be those found encoded on minicircles in other dinoflagellate species. The results of this series of experiments are reported in Chapter 3.

The last question has been addressed by a combination of organelle purification and *in vivo* labelling with 35S-methionine. By performing this metabolic labelling in the presence of protein synthesis inhibitors chloramphenicol (inhibitor of synthesis on 70S ribosomes) or cycloheximide (inhibitor of synthesis on 80S ribosomes), I was able to demonstrate that the

psbA gene product is synthesized in the plastid. By extension therefore, this is where the psbA gene must be located. The plastid purification is reported in chapter 4, while the labelling studies are reported in chapter 5.

Figure Legends

Figure I.1. Schematic representation of plastid evolution showing the endosymbiosis origin of chloroplasts (Nassoury and Morse 2005). A typical eukaryote, containing both a nucleus and mitochondria, is proposed to have engulfed a cyanobacterial-like cell in the primary endosymbiotic event that eventually gave rise to all plastids.This original endosymbiont, surrounded by two membranes and a peptidoglycan wall (dottes line), presumably contained thylakoids dotted with phycobilisomes. The three groups of extant primary plastids are those of green alga and higher plants (which have lost their phycobilisomes), those of the red alga, and those of the glaucophytes (which have retained the peptidoglycan wall). In the different secondary endosymbioses, a new eukaryotic host engulfed a photosynthetic eukaryote (either a red or a green alga). Extant secondary plastids are surrounded by either three (dinoflagellates and euglenids) or four membranes (Chlorarachniophytes, Haptophytes, Heterokonts, Cryptophytes, Apicomplex), and in some cases contain a residual nucleus termed a nucleomorph derived from the original eukaryotic host (Cryptophytes, Chlorarachniophytes, Haptophytes, Heterokonts). In several cases the outer membrane of the secondary plastid is contiguous with the ER of the new host (Cryptophytes, Haptophytes and Heterokonts). No secondary plastids have phycobilisomes, although cryptophytes do have phycobilin pigments inside the thylakoid lumen. Not shown are the genes transfered from the endosymbiont to host nucleus in the two endosymbiosis.

Secondary plastids

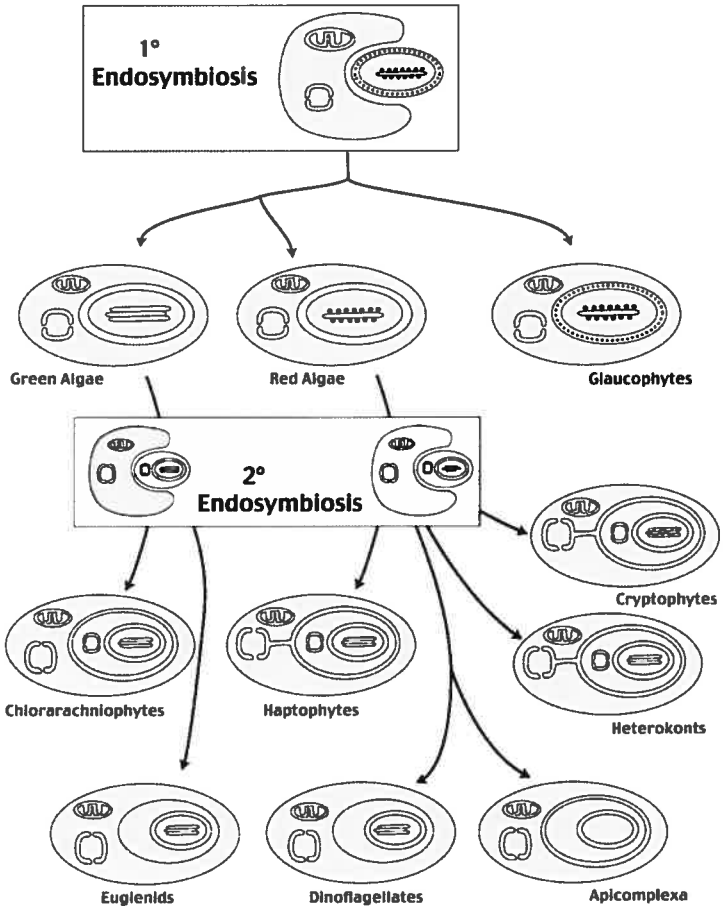


Figure I. 2. Distribution of the 190 first identified and 66 hypothetical (ycf) protein-coding genes in the nine photosynthetic algal plastid genome (Grzebyk and Schofield 2003). Primary endosymbiotic genomes (Glaucophyta, Rhodophyta, Chlorophyta, square boxes) are widely overlapping, in accordance with the monophyletic origin of all plastids. The secondary endosymbiotic genomes (Euglenophyta, Cryptophyta, Bacillariophyta, round boxes) are included in the primary lineages from which they emerged. Parentheses indicate genes not found in all rhodophyte plastid genomes; square brackets indicate genes not found in all chlorophyte plastid endosymbiont. Genes involved in the three main plastid functions represented in the core set are highlighted: ATP synthase genes (yellow), photosynthetic processes (green), and housekeeping genes (blue). Genes involved in protein regulatory pathways (transcriptional and post-transcriptional regulation) in the complementary gene set of red plastid are shown in pink.

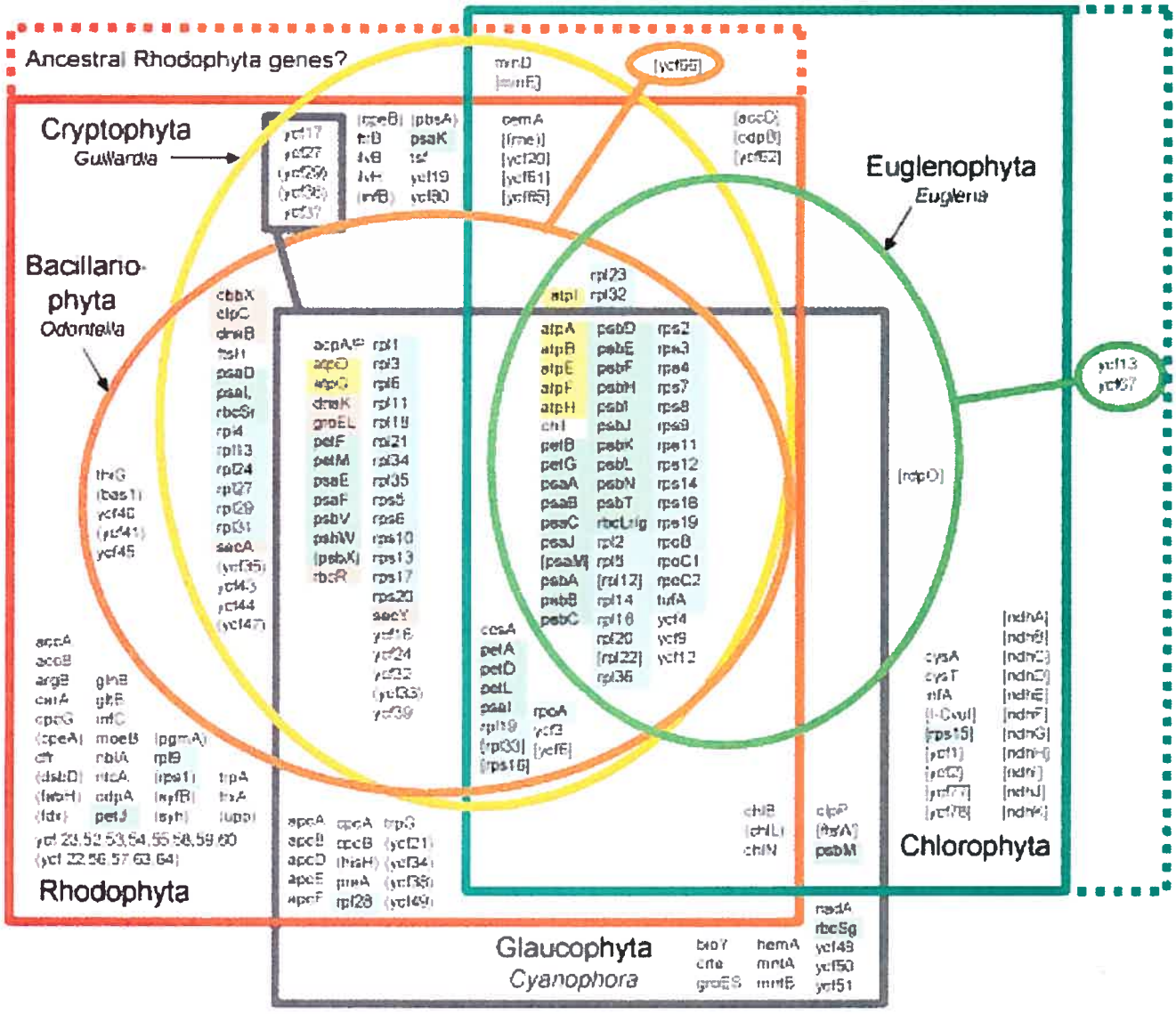

Ancestral Rhodophyta genes?
Cryptophyta
Guillardia
Bacillario-
phyta
Odontella
Euglenophyta
Euglena
Chlorophyta
Rhodophyta
Glaucophyta
Cyanophora

Figure I.3. Schematic view of a dinoflagellate plastid minicircle (modified from Zhang et al., 1999 and Koumandou et al., 2004). One circle (2-3kb) generally carries 1 gene, although in some case, 2-3 genes with same orientation have been observed. The dark green box represents the coding region; light pink, light green and light yellow boxes indicate the three well conserved regions (C1, C2, C3), which separated by two less well conserved regions (grey boxes). Three regions in the core are of particular interest. Region A (blue green box) represents a 9-bp A-rich region, a feature shared between *Amphidinium operculatum* and *Heterocapsa triquetra*, while regions B and C (two blue boxes) represent the 9-bp elements that are conserved among all *Amphidinium* minicircles (in *Heterocapsa triquetra* they are 9Gs).

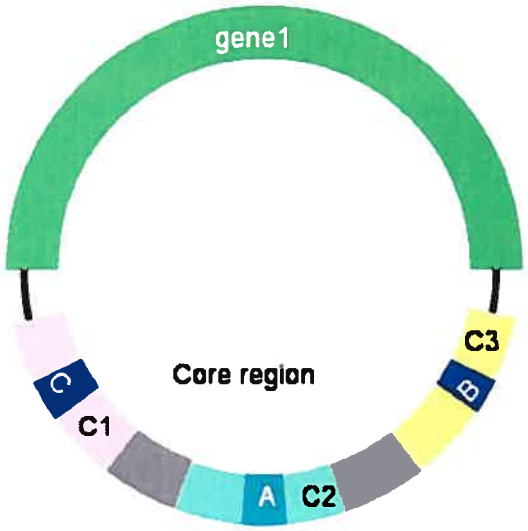

gene1
C3
B
Core region
C
C1
A
C2

Figure I.4. Schematic arrangement of polypeptides in barley thylakoid membranes (Hall and Rao 1994). PSII, Photosystem II; PSI, Photosystem I; CF0 and CF1, coupling factors.

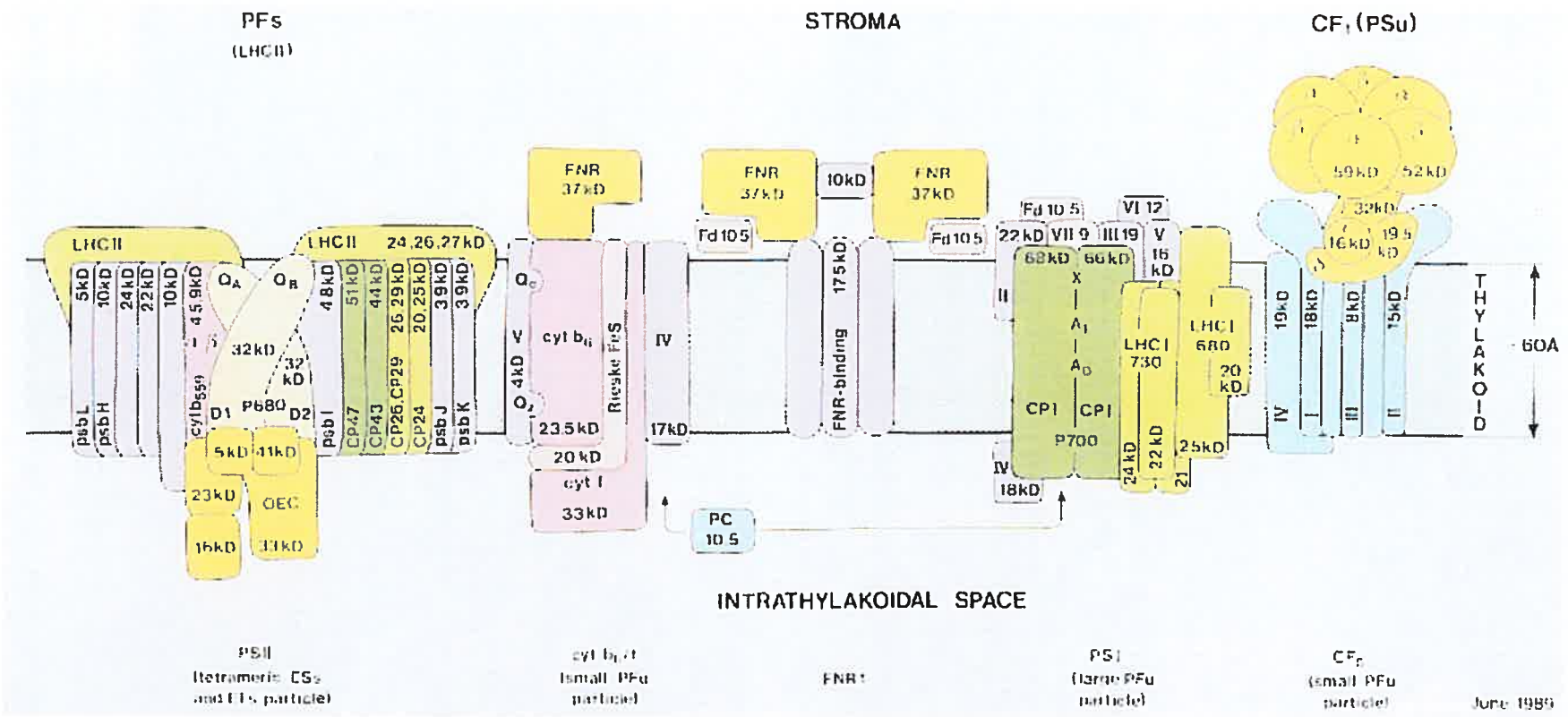

PFs
(LHCII)
STROMA
CF1 (PSu)
LHCII
LHCII 24,26,27kD
5kD
10kD
24kD
22kD
10kD
45.9kD
32kD
32 kD
48kD
51kD
44kD
26.29kD
20.25kD
39kD
39kD
psbL
psbH
psbI
CP47
CP43
CP26.CP29
CP24
psbJ
psbK
D1
P680
D2
cytb559
5kD
41kD
23kD
OEC
16kD
33kD
QA
QB
QC
O2
V
4kD
cyt b6
23.5kD
20 kD
cyt f
33kD
Rieske FeS
IV
17kD
FNR
37kD
FNR
37kD
10kD
FNR
37kD
Fd 10.5
Fd 10.5
Fd 10.5
175kD
FNR-binding
II
22 kD
VII 9
III 19
V
16
kD
VI 12
68kD
66kD
X
A1
A0
CP1
CP1
P700
LHC I
730
LHC I
680
20
kD
25kD
24kD
22 kD
21
IV
18kD
PC
10.5
59kD
52kD
32kD
16kD
19.5
kD
IV
19kD
I
18kD
III
8kD
II
15kD
THYLAKOID
60A
INTRATHYLAKOIDAL SPACE
PSII
(tetrameric CCs
and LHCs particle)
cyt b6,f
(small PEu
particle)
FNR
PSI
(large PEu
particle)
CF0
(small PEu
particle)
June 1989

Table I.1. Major chloroplast coded genes in higher plants (modified from Lawlor, 2001; He and Malkin, 2000).

Chloroplast	Gene	Gene Product
ATP synthesis Complex	atpA	CF1α subunit
	atpB	CF2β subunit
	atpE	CF1ε subunit1
	atpF	CF0 subunit III
	atpH	CF0 subunit IV
	atpI	CF0 subunit I
Cyctochrome b-f complex	petA	Cytochrome f
	petB	Cytochrome b6
	petD	Cytochromeb6-f
Photosystem I	psaA	P700-Chl a protein
	psaB	P700-Chl a protein
	psaC	FA and FB binding protein
	psaJ	4.9 kDa protein
	psaI	4.0 kDa protein
Photosystem II	psbA	D1 protein
	psbB	Chl-a protein
	psbC	Chl-a protein
	psbD	D2 protein
	psbE	Cyt b559
	psbF	Cyt b559
	psbH	Phospho-protein
	psbI	Reaction centre protein
RuBisCO	rbcL	RuBisCO large subunit
Ribosomal RNA		
	23S rRNA	Large subunit
	16S rRNA	Small subunit
	5S rRNA	
	4.5SrRNA	
Transfer RNA	about 30 different components of protein synthesis trn genes	
Ribosomal proteins	20 different 30S + 50S ribosomal protein rpl genes	
RNA polymerase	rpoA	subunit α
	rpoB	subunit β1
	rpoC	subunit β
Initiation factor 1 proteins	infA	components of protein synthesis

Table I.2. A total of 18 genes have been identified on dinoflagellate minicircles (data from gene bank, http://www.ncbi.nlm.nih.gov). Asterisk (*) indicates a minicircle gene reportedly transferred to the nucleus.

Chloroplast	Gene	Gene product	Species where found
ATP synthesis Complex	atpA atpB	CF1α subunit CF2β subunit	*Amphidinium, Heterocapsa* *Amphidinium*
Cytochrome b-f complex	petB petD	Cytochrome b6 Cytochromeb6-f	*Amphidinium,Heterocapsa, Ceratium** *Amphidinium*
Photosystem I	psaA psaB	P700-Chl a protein P700-Chl a protein	*Amphidinium, Heterocapsa, Ceratium** *Amphidinium, Heterocapsa, Ceratium**
Photosystem II	psbA psbB psbC psbD psbE psbI	D1 protein Chl-a protein Chl-a protein D2 protein cytochrome b559 alpha subunit Reaction centre protein	*Amphidinium, Heterocapsa, Symbiodinium, Pyrocystis** *Amphidinium, Heterocapsa, Ceratium** *Amphidinium, Heterocapsa, Ceratium*, Pyrocystis** *Amphidinium, Ceratium** *Amphidinium* *Amphidinium*
Ribosomal RNA	23S rRNA 16S rRNA		*Amphidinium, Heterocapsa, Protoceratium* *Heterocapsa, Ceratium**
Ribosomal protein	rpl28 rpl33	50S ribosomal protein L28 50S ribosomal protein L33	*Pyrocystis** *Pyrocystis**
Unidentified ORF	ycf16 ycf24	putative ATP-dependent transport protein putative ABC transporter membrane protein	*Ceratium** *Ceratium**

Chapter II. The plastid-encoded psbA gene in the dinoflagellate *Gonyaulax* is not encoded on a minicircle

Objectives

1) **To investigate the plastid genome architecture in the dinoflagellate *Gonyaulax*.**
2) **To obtain sequences of genes encoded by the plastid of dinoflagellate *Gonyaulax*.**

Publication of chapter II: Accepted in Gene, Nov. 29, 2005 .

All the experimental procedures were done by me.

The plastid-encoded psbA gene in the dinoflagellate *Gonyaulax* is not encoded on a minicircle

Running title: *Gonyaulax* psbA gene

Yunling Wang and David Morse*

Département de Sciences Biologiques, Université de Montréal

4101 Sherbrooke est, Montreal, Quebec, Canada H1X 2B2

*To whom correspondence should be addressed.

Tel: 514 872 9975; Fax: 514 872 9406

[redacted]

Abstract

In all dinoflagellate species studied to date, the plastid genome is highly reduced, with many genes normally found in the plastid genome found instead encoded by the nucleus. Furthermore, those genes still remaining in the plastid are found as primarily single gene minicircles whose size is typically only 2-3 kb. We show here that the plastid genome architecture in the dinoflagellate *Gonyaulax polyedra* is unusual for this class of organism. In particular, the psbA gene is associated with DNA of roughly 50-150 kb and appears to have an unusually high complexity.

Key words: plastid genome; peridinin-containing dinoflagellate; plastid evolution; psbA

Introduction

The plastid genomes of all dinoflagellates that have been studied to date are characterized by two peculiar features. First, plastid genes are found on small, plasmid-like minicircles. These minicircles have been observed in *Heterocapsa triqueta* (Zhang, Green et al. 1999), *H. pygmeae* and *Protoceratium reticulatum* (Zhang, Cavalier-Smith et al. 2002), *Amphidinium operculatum* (Barbrook and Howe 2000; Barbrook, Symington et al. 2001; Nisbet, Koumandou et al. 2004), *A. carterae* (Hiller 2001) and *Symbiodinium* sp. (Moore, Ferguson et al. 2003). The minicircles from any given species have a common core conserved among the different genes within a species, and some motifs of the core are also found conserved between species, but no elements are conserved between genera. Typically, the minicircles contain a single gene (Zhang, Green et al. 1999; Barbrook and Howe 2000) although exceptions containing two or three genes and "empty" minicircles lacking any recognizable protein coding sequence (Barbrook, Symington et al. 2001; Hiller 2001; Nisbet, Koumandou et al. 2004) have been documented. Those minicircles containing several genes do not appear to transcribe them as a polycistronic message (Barbrook, Symington et al. 2001).

The second peculiar feature of the plastid genome is the paucity of genes. Indeed, very few of the genes normally found in plastids are found on the minicircles. While it might initially be thought that the remaining plastid genes might be found in an alternate plastid genome structure, recent EST sequencing projects have discovered that many of the expected proteins are actually encoded by the dinoflagellate nucleus. Indeed, a number of proteins encoded by the plastids of higher plants and other algae are nuclear encoded in *Alexandrium tamarense* (Hackett, Yoon et al. 2004), *Amphidinium carterae* and *Gonyaulax polyedra* (Bachvaroff, Concepcion et al. 2004). The nuclear location for these plastid-directed proteins was confirmed by the presence of a unique N-terminal extension required for plastid targeting in dinoflagellates (Nassoury, Cappadocia et al. 2003), the presence of a poly(A) tail, and the

expected high GC-content of the sequences (Spector 1984). The only genes remaining in the plastid are the minicircle genes encoding the 16S and 23S rRNA, eight photosystem components (*psaA-B*, *psbA-E*, *psbI*), two ATP synthase subunits (*atpA-B*), two cytochrome b_6f subunits (*petB*, *petD*) as well as two unidentified ORFs (*ycf16, ycf24*) and two ribosomal proteins (*rpl28, rpl23*). Dinoflagellate plastid genomes thus contain a highly reduced gene complement distributed amongst many small circular elements.

One possible caveat to the idea that these unusual features reflect a unique plastid genome is the suggestion that the minicircles may in fact be encoded by the nucleus (Laatsch, Zauner et al. 2004). However, in *Symbiodinium,* psbA transcripts have been observed in the plastids by *in situ* hybridization (Takashita, Ishikura et al. 2003). Furthermore, none of the minicircle genes sequences to date encode the characteristic N-terminal leader sequence required for plastid targeting (Nassoury, Cappadocia et al. 2003). Lastly, we have recently demonstrated that psbA synthesis *in vivo* is inhibited by chloramphenicol but not by cycloheximide in *Gonyaulax* and *Amphidinium* (Wang, Jensen et al. 2005). Taken together, these studies provide strong presumptive evidence for the presence of minicircles in the plastid. However, since few dinoflagellate species have been examined, it is still a formal possibility that different dinoflagellates may utilize different strategies for encoding and expressing the plastid genome.

We have been interested for some time in the genes encoded by the plastid of the dinoflagellate *Gonyaulax*, but have so far been unsuccessful in cloning these genes based on sequence similarity to minicircle sequences of other dinoflagellate species. We report here a surprising structural difference in the plastid genome architecture of *Gonyaulax*. In particular, we find that psbA from this species is not found on a minicircle. The plastid genome organization in *Gonyaulax* thus appears quite different from that described for the other dinoflagellates.

Materials and Methods

Cell culture

Amphidinium carterae (CCMP 1314) and *Gonyaulax polyedra* (now *Lingulodinium polyedrum*; strain 70) were obtained from the Provasoli-Guillard Culture Center for Marine Phytoplancton (Boothbay Harbor, Maine) and grown in a modified seawater medium (f/2) (Guillard and Ryther 1962) at constant temperature (18 ± 1°C) in 12-h light/12-h dark cycles using cool white fluorescent light at an intensity of 50 µmol photons·m^{-2}·s^{-1}. The beginning of light period is defined as time 0 (LD 0), and the beginning of the dark period as LD 12. Cultures were grown to a cell density of 12-14,000 cells/mL (*Gonyaulax*) or 500,000 cells/mL (*Amphidinium*) as measured by counting using a hemocytometer. Both cultures were unialgal but only *Amphidinium* is axenic.

Nucleic acid isolation

Two g wet weight cells, harvested by filtration on Whatman 541 paper, were ground in liquid nitrogen and resuspended in 10 mL lysis buffer (20 mM Tris pH 8 containing 50 mM EDTA and 100 mM NaCl). One mL 20% SDS and 0.5 mL 20 mg/mL proteinase K were added, mixed well and incubated at 50°C. After two hours, an additional 0.5 mL 20 mg/mL proteinase K was added and left overnight at room temperature. The mixture was extracted twice with an equal volume or phenol/chloroform and the nucleic acids precipitated from the aqueous fraction by addition of two volumes 100% ethanol. DNA was air-dried and redissolved in 2.5 mL TE, and 1.1 volumes of CTAB buffer (250 mM Tris HCl, pH 8, 100 mM EDTA, 55 mM hexadecylamine triammonium bromide, and 2.5 M NaCl) added and incubated for 45 min at 65 °C with vortexing every 5 min. The solution was

extracted with one volume chloroform, and 5.5 g CsCl and 5 µL 10 mg/mL benzimide were added to the aqueous solution. Following centrifugation at 45,000 rpm in a Beckman 70.1 Ti rotor for 24 hours at 20°C, DNA bands were visualized by long wave UV, collected with a large bore needle, and precipitated by centrifugation after addition of three volumes TE.

Pulse field gel electrophoresis was carried out on 1 % agarose gels in 0.5X TBE. Typically, samples were run at 6 V/cm for 15 hours with a pulse ramp from 1 to 30 seconds. Samples included purified DNA obtained from CsCl gradients or total DNA from cells ground in liquid nitrogen. For the latter, powder from 0.5 g wet weight ground cells was resuspended in 5 mL of lysis buffer (10 mM Tris, pH 7.2, 50 mM NaCl, 0.2% sodium deoxycholate, 0.5% sodium lauryl sarcosine) and incubated at 50°C for 10 minutes. The suspension was combined with an equal volume of 2% low melting temperature agarose and gently mixed. This mixture transferred to plug molds at 50°C using sterile transfer pipettes. The agarose plugs were solidified at 4°C, incubated in lysis buffer for 1 hour at 37°C, rinsed with wash buffer (20 mM Tris, pH 8.0, 50 mM EDTA) and incubated with proteinase K at 50°C overnight without agitation. After the protease treatment, the plugs were washed 4 times at room temperature using 50 ml of wash buffer for 1 hour each. The washed DNA plugs were cut to a size 90% of the well height and gently pressed to the bottoms of the wells. Each sample well was filled with agarose at the same concentration as that in the gel, and the agarose allowed to set for 10 to 15 min before electrophoresis.

Nucleic acid analysis

Gene probes for psbA and 23S rDNA were amplified from *Gonyaulax* DNA using PCR (Wang, Jensen et al. 2005). All sequences were analyzed using MacVector software (Accelrys). For high stringency hybridization of nucleic acids, high temperature (65°C) in 0.5 M Phosphate/5% SDS/2% BSA method was used (Church and Gilbert 1984), while for low stringency hybridizations,

the temperature was reduced to 55°C. Other standard molecular techniques were performed as described (Sambrook, Fritsch et al. 1989).

Results

DNA extracted from *Gonyaulax polyedra* contains a psbA gene (encoding the photosystem D1) similar to that found in the dinoflagellate *Amphidinium carterae* and other dinoflagellates (Fig. II. 1). Curiously, PCR reactions using the reverse complement of the oligonucleotides used to amplify the psbA coding sequence did not amplify a minicircle sequence, as expected from the plastid genome structure known for other dinoflagellates. To test directly for the presence of a psbA minicircle in *Gonyaulax*, we probed undigested genomic DNA on Southern blots with the psbA sequence. The *Gonyaulax* DNA hybridizing with psbA appears significantly larger (> 20 kb) than the roughly 3 kb DNA fragment observed using undigested DNA from *Amphidinium* (Fig. II.2). The *Amphidinium* psbA is found only as a minicircle when probed with a homologous probe at high stringency, and the slight hybridization signal to larger DNA fragments observed here with is due to the low stringency hybridization with our heterologous probe. *E. coli* DNA, chosen as a negative control lacking psbA, also hybridizes weakly under these conditions. We conclude that the psbA in *Gonyaulax* is not encoded by minicircular DNA.

The psbA sequence is AT-rich, unlike all nuclear-encoded genes from *Gonyaulax* studied to date. To determine if the psbA was found in an AT-rich context, we centrifuged total DNA extracts from *Gonyaulax* on CsCl gradients in the presence of bisbenzimide. Of the two bands resolved using this technique, only the upper AT-rich DNA was found to hybridize to our psbA probe (Fig. II.3). In contrast, the lower GC-rich DNA was found to hybridize preferentially to a probe for the nuclear-encoded plastid-directed peridinin-chlorophyll a-protein (PCP) gene (Fig. II.3). The small degree of PCP hybridization to DNA in the upper band may reflect difficulty in completely purifying away the large amount of PCP genes, found in roughly 5000 copies per *Gonyaulax* genome (Le, Markovic et al. 1997) and the large amount of

nuclear DNA, estimated at 200 pg per nucleus (Spector 1984). We conclude that our psbA sequence is found in an AT-rich context.

We were curious as to the physical size of the DNA encoding the psbA gene, since agarose gels do not allow fragments greater than 20 kb to be resolved. In a first approach, we attempted to estimate the size of the plastid genome by restriction fragment analysis of the AT-rich psbA-containing DNA. For typical plastids, this approach estimates plastid genome sizes of between 100 to 200 kb. However, no distinct bands were ever observed by HindIII digestion of AT-rich DNA (Fig. II. 3 panel iii). DNA extracted from *Gonyaulax* is always difficult to digest completely, and thus the restriction of the DNA was verified by hybridization to psbA (the sequence contains a single internal HindIII site). In the samples shown, at least half the psbA sequences are digested. Despite this, no distinct DNA bands are visible by ethidium bromide staining. We conclude from this that the AT-rich DNA from *Gonyaulax* may be more complex than that usually found in plastids.

In a second approach to determining the physical size of the DNA encoding the psbA gene, DNA from whole cells ground in liquid nitrogen and embedded in agarose plugs was analyzed using pulse field gel electrophoresis. Control DNA fragments ranging in size from 50 to 1000 kb are easily resolved using these conditions. *Gonyaulax* DNA (Fig. II.4) is found in two size classes, one ≥ 1000 kb (open triangle) and one of 50 – 150 kb (closed triangle). The large DNA presumably reflects nuclear DNA not resolved into individual chromosomes under these experimental conditions. *Gonyaulax* contains roughly 10^{11} bp of DNA in a reported 220 chromosomes (Spector 1984), suggesting that individual chromosomes may be as large as ~500,000 kb. In any event, a clear hybridization signal with the psbA gene probe is found in the region of 50-150 kb (Fig. II.4). The *Gonyaulax* psbA gene is associated with a DNA molecule whose size is significantly larger than the minicircles reported for other dinoflagellate species. It is noteworthy that the AT-rich DNA isolated from the bisbenzimide-CsCl gradients is of

similar size (Fig. II.4), although some breakage of these samples is likely to have occurred due to breakage of the DNA during the purification procedure.

Discussion

It is now generally accepted that dinoflagellate plastid genomes are highly unusual, and are comprised of a limited number of genes encoded on small plasmid-like minicircles. We find that the plastid genome architecture in *Gonyaulax* is quite different from that typical of dinoflagellates. Instead of the small 2-3 kb minicircles found in other species, we find that the psbA gene in *Gonyaulax* associated with DNA of roughly 50-150 kb. While this amount of DNA appears on the surface to be similar to the genome size of higher plant plastids, several lines of evidence suggest to us that the *Gonyaulax* plastid genome may in fact be even larger. First, hybridization of our pulse field gels with a psbA probe shows a smear with undigested DNA (Fig. II.4) but two small bands after HindIII digestion (Fig. II.3). The hybridization to the digested DNA is consistent with the presence of a single HindIII in the psbA sequence, but the smeared hybridization signal observed with undigested DNA might be interpreted either as several conformations of the same molecule or as several different molecules. These characteristics of hybridization to a psbA probe are reminiscent of the hybridization of a cox1 probe to mitochondrial DNA from the dinoflagellate *Crypthecodinium*: a smear between 6 and 9 kb is observed when undigested DNA is probed, and discrete bands are observed when a restriction enzyme digest is probed (Norman and Gray 2001). These observations have led to the suggestion that the mitochondria may contain many different molecules, and an analogous situation may thus also occur in the *Gonyaulax* plastid genome.

A second line of evidence is based on a potentially large amount of non-coding DNA in the plastid genome. For example, random sequencing of mechanically sheared DNA fragments recovered no identifiable genes in ~100 kb of sequence. This suggests that either large amounts of non-coding DNA are present or that the plastid DNA is heavily edited. While editing in the dinoflagellate mitochondrial genome appears widespread and extensive (Lin, Zhang et al. 2002), and editing has been reported for the minicircles in the dinoflagellate *Ceratium* (Zauner, Greilinger et al. 2004) there is no evidence

of editing in *Amphidinium* minicircles (Barbrook and Howe 2000). While we cannot conclusively rule out in editing in *Gonyaulax*, our psbA sequence amplified from DNA was identical to the sequence amplified from RNA, suggesting this possibility is unlikely.

A third line of evidence supporting a high complexity plastid genome is that no discrete bands are observed by ethidium bromide staining of restriction enzyme digests of the AT-rich DNA. Plastid genomes in general, and curiously enough, the first reports of dinoflagellate plastid genomes (using one species of *Glenodinium* and two species of *Protogonyaulax*), generally produce a limited number of gene fragments after restriction enzyme digestion (Boczar, Liston et al. 1991). While neither *Glenodinium* nor *Protogonyaulax* have been reexamined for the presence of a minicircular plastid genome, our results with digestion of *Gonyaulax* AT-rich DNA argues for a high complexity DNA sample.

Lastly, although we have so far only identified one other gene fragment belonging to the *Gonyaulax* plastid genome (a 0.9 kb region with homology to dinoflagellate 23 rRNA), long range PCR combining primers from both psbA and 23S genes did not amplify any DNA. This suggests that either the two sequences are too far apart to be amplified (> 40 kb) or are located on different molecules. Taken together, then, our results point for an unusual organization of the plastid genome in the dinoflagellate *Gonyaulax*: two plastid genes (psbA and 23S) are not encoded by minicircles, and it appears as though a large proportion of the genome may be non-coding. At this point, it is impossible to determine whether the *Gonyaulax* plastid genome will be ancestral or derived. Clearly, it will be necessary to isolate many more plastid genes from this species before this aspect to the evolution of dinoflagellate plastids can be resolved.

The reddish-colored peridinin-containing dinoflagellate plastids are thought to have originated from a secondary endosymbiotic event, and are distinct from the so-called "green" dinoflagellates whose plastids contain fucoxanthin (Yoon, Hackett et al. 2005). However, the identity of the ancestor

to the peridinin-type plastid is still enigmatic. *Gonyaulax*, for example, contains a form II ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) (Morse, Salois et al. 1995), an enzyme found previously only in anaerobic proteobacteria. Furthermore, the major light harvesting protein in this organism, the soluble peridinin-chlorophyll *a*-protein (Lee, Mittag et al. 1993) is found in no other extant class of organisms, and interesting, has a three-dimensional structure unrelated to any other light harvesting protein (Hofmann, Wrench et al. 1996). Lastly, the nuclear-encoded plastid-directed glyceraldehyde-3-phosphate dehydrogenase (GAP), plastid gene appear to be derived from lateral gene transfer from cryptomonads (Fagan, Hastings et al. 1998). Although much further work will be required to fully understand the complex evolution of the dinoflagellate plastid, the *Gonyaulax* plastid genome in particular deserves attention because of its dissimilarity with the plasmid-like circles found in other species.

Acknowledgements

We thank B. Green (University of British Columbia) for critical analysis of our initial experiments and helpful suggestions for further work. We gratefully acknowledge the technical expertise of B.F. Lang (University of Montreal) in cloning and random sequencing the AT-rich DNA. This work has been supported by a grant from the National Science and Engineering Research Council of Canada.

Figure Legends

Figure II.1 Sequence similarities of psbA from various dinoflagellate species. Sequence identities are shown for four dinoflagellate psbA sequences at the DNA (upper right) and protein (lower left) levels.

Sequence similarities of psbA from various dinoflagellate species.

DNA / Protein	Alexandrium	Amphidinium	Heterocapsa	Gonyaulax
Alexandrium		66%	72%	84%
Amphidinium	78%		79%	68%
Heterocapsa	83%	82%		75%
Gonyaulax	86%	82%	88%	

Figure II.2 The psbA gene in *Gonyaulax* is not a minicircle.

Low stringency Southern blot analysis of total DNA extracted from *Gonyaulax polyedra* (lane 2), *Amphidinium carterae* (lane 3) and *Escherichia coli* (lane 4) electrophoresed on a standard 1.2 % agarose gel. Transfers were hybridized to either a 900 bp fragment of *Gonyaulax* 23S RNA or to a 700bp fragment of *Gonyaulax* psbA. Molecular weight markers (lane 1) are a HindIII digest of lambda DNA; the open triangles indicate ~20 kb and the closed triangles 3 kb. The ethidium bromide stained gel is at top, and the two Southern blots with the indicated probes below.

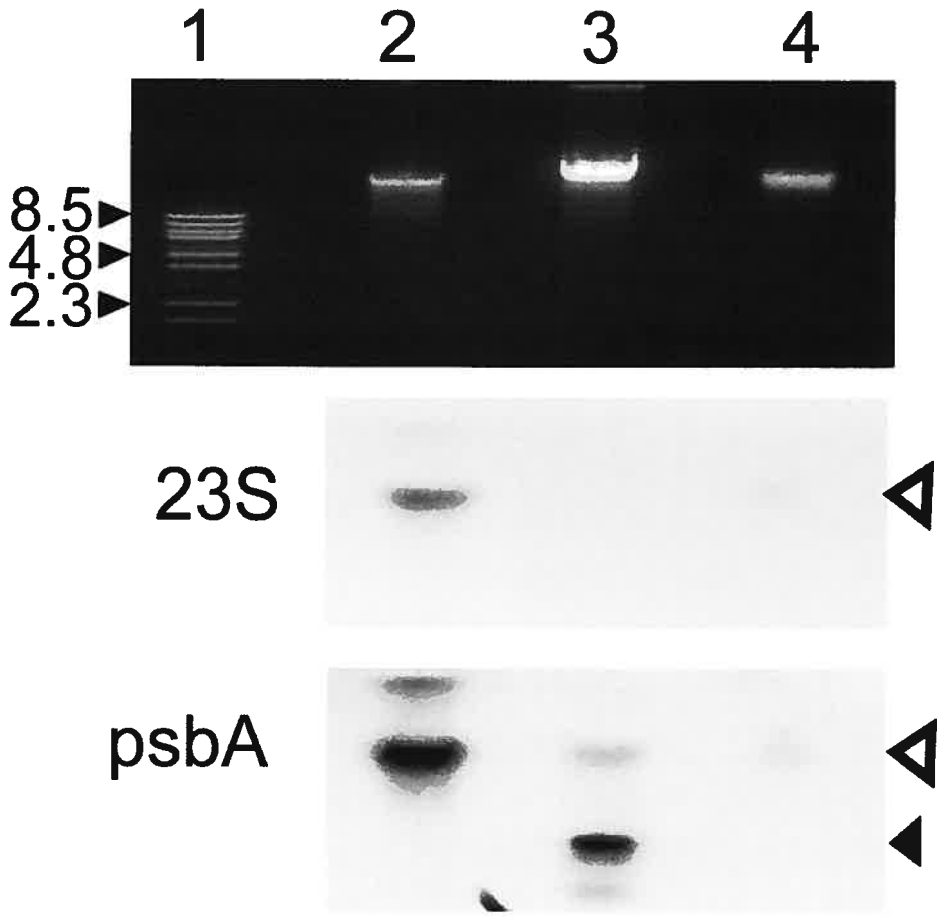

1
2
3
4
8.5
4.8
2.3
23S
psbA

Figure II.3 The psbA gene in *Gonyaulax* is in an AT-rich context.
(i) Total DNA extracts were resolved into GC-rich (lane 2) and AT-rich fractions (lane 3) by CsCl gradient centrifugation in the presence of bisbenzimide and hybridized to a 700 bp fragment of *Gonyaulax* psbA. Molecular weight markers (lane 1) are as in figure 2. Ethidium bromide staining is at left and the corresponding Southern blot at right. (ii) As in panel i, except that the nuclear-encoded PCP gene was used a probe. (iii) AT-rich DNA alone (lane 2) or following digestion with HindIII (lane 3) were electrophoresed and hybridized to a 0.7 kb fragment of *Gonyaulax* psbA. Molecular weight markers (lane 1) are as in figure 2.

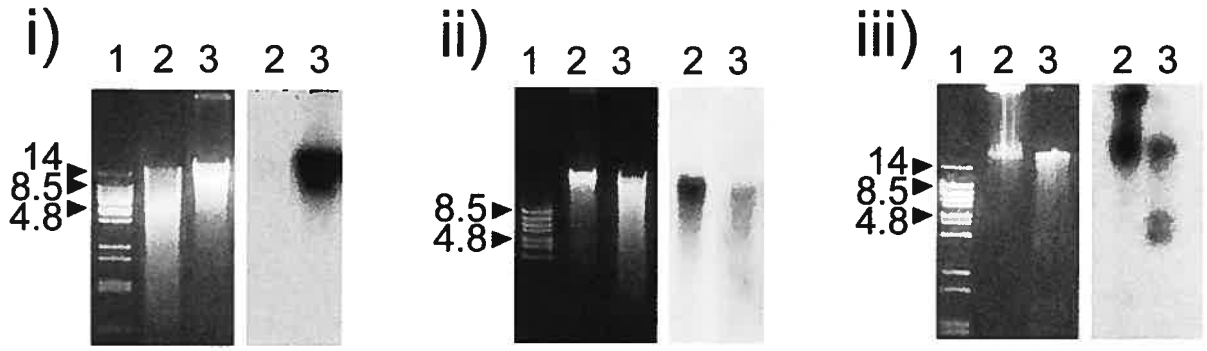
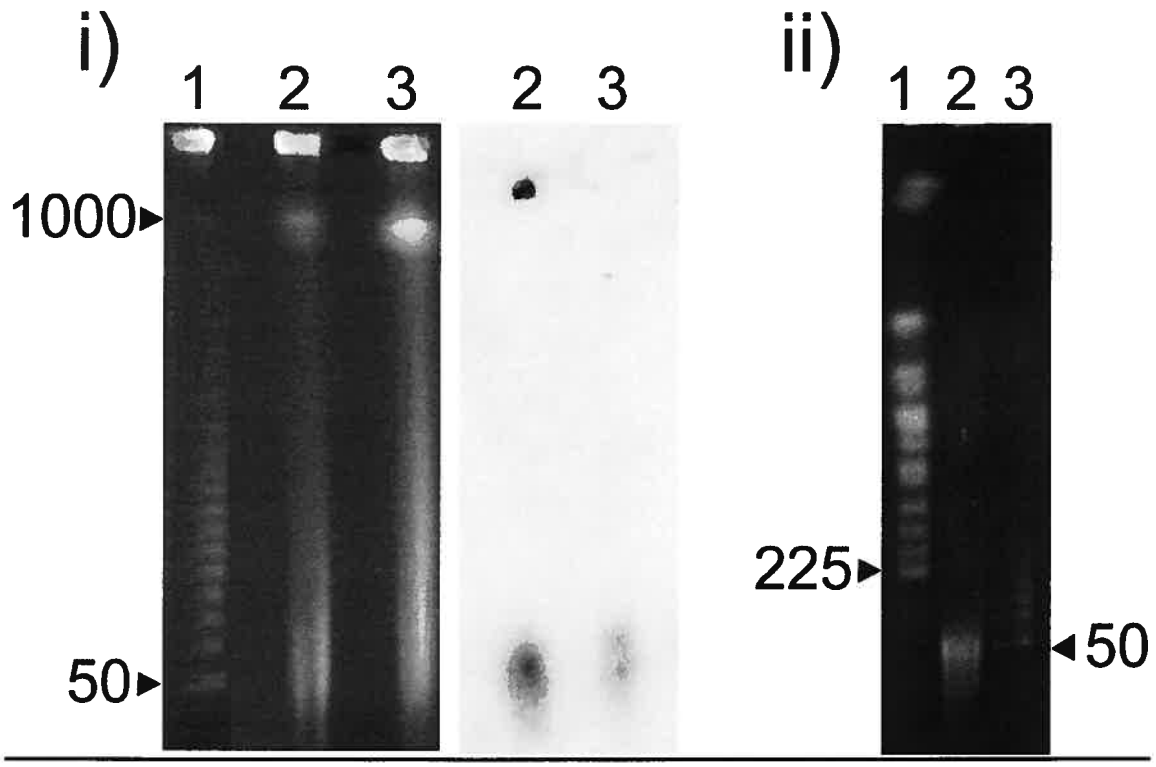

i)
1 2 3 2 3
14
8.5
4.8
ii)
1 2 3 2 3
8.5
4.8
iii)
1 2 3 2 3
14
8.5
4.8

Figure II.4 The psbA gene is encoded by DNA of between 50-150 kb.
(i) Pulse field electrophoresis of two independent preparations of total *Gonyaulax* DNA (lanes 2, 3) after ethidium bromide staining (left) or hybridization to a 700bp fragment of *Gonyaulax* psbA (right). The lambda ladder (Promega) from 50 – 1000 kb are used as molecular weight markers (lane 1). The open triangles indicate sizes > 1000 kb and the closed triangle 50 kb. (ii) Pulse field electrophoresis of AT-rich DNA (lane 2) prepared as in figure 3. Molecular weight markers are PGF yeast chromosome markers (Promega) from 225 to 1,900 kb (lane 1) or a lambda ladder (lane 3). The closed triangle indicates 50 kb.


i)
1 2 3 2 3
1000
50
ii)
1 2 3
225
50

Chapter III. Rampant polyuridylylation of plastid gene transcripts in the dinoflagellate Lingulodinium

Objectives

1) To characterize dinoflagellate plastid RNA 3' end processing

2)To obtain EST sequencing data from a poly(U) cDNA library

Publication of chapter III : This manuscript has been published in Nucleic Acids Res. 2006 ;34(2):613-9.

All the experimental procedures were done by me.

Rampant polyuridylylation of plastid gene transcripts in the dinoflagellate *Lingulodinium*

Yunling Wang and David Morse*

Institut de Recherche en Biologie Végétale

Département de Sciences Biologiques, Université de Montréal

4101 Sherbrooke est, Montreal, Quebec, Canada H1X 2B2

Running title: Polyuridylylated plastid transcripts

* To whom correspondence should be addressed.
Tel: 514 872 9975; Fax: 514 872 9406;
[redacted]

Key Words: Dinoflagellate, Plastid mRNA, post-transcriptional modification, poly(U) tail

Summary

Dinoflagellate plastid genes are believed to be encoded on small generally unigenic plasmid-like minicircles. The minicircle gene complement has reached saturation with an incomplete set of plastid genes (18) compared to typical functional plastids (60-200). While some of the missing plastid genes have recently been found in the nucleus, it is still unknown if additional genes, not located on minicircles, might also contribute to the plastid genome. Sequencing of tailed RNA showed that transcripts derived from the known minicircle genes *psbA* and *atpB* contained a homogenous 3' polyuridine tract of 25-40 residues. This unusual modification suggested that random sequencing of a poly(dA) primed cDNA library could be used to characterize the plastid transcriptome. We have recovered only 12 different polyuridylylated transcripts from our library, all of which are encoded on minicircles in several dinoflagellate species. The correspondence of all polyuridylylated transcripts with previously described minicircle genes thus supports the dinoflagellate plastid as harboring the smallest genome of any functional chloroplast. Interestingly, Northern blots indicate that the majority of transcripts are modified, suggesting that polyuridylylation is unlikely to act as a degradation signal as do the heterogeneous poly(A)-rich extensions of transcripts in cyanobacteria and other plastids.

Introduction

Although dinoflagellates are best known as the notorious cause of toxic red tides, they are also important contributors to the ocean's primary production. Photosynthesis in these organisms is typically carried out in plastids surrounded by three membranes (Dodge 1975), an evolutionary footprint reflecting their origin through secondary endosymbiosis (Keeling 2004). The evolutionary ancestor of the peridinin-containing plastids is suggested from molecular phylogenetic reconstructions using plastid-encoded genes to be a red alga (Zhang, Green et al. 2000; Yoon, Hackett et al. 2002), a conclusion supported by phylogeny of nuclear-encoded plastid-directed genes (Fast, Kissinger et al. 2001). In general, these findings are also consistent with phylogenetic reconstructions determined from non plastid-directed host genes (Baldauf, Roger et al. 2000; Harper, Waanders et al. 2005).

Despite widespread acceptance of a red-algal origin for the peridinin-containing plastid, these organelles display a number of peculiar characteristics that share no homology with any known extant plastids. For example, the carotenoid peridinin itself (Jeffrey, Seilicki et al. 1975) and the unusual light harvesting peridinin-chlorophyll *a*-protein to which it binds (Hofmann, Wrench et al. 1996) are found in no other organisms. Furthermore, the Rubisco in peridinin-containing plastids, an unusual form II enzyme, is dissimilar from the form I protein employed by all other plastids (Morse, Salois et al. 1995). The evolutionary provenance of these proteins is thus unknown, and as they are derived from nuclear-encoded genes, it seems possible that their history may be distinct from that the plastid itself.

An additional issue having bearing on plastid evolution, and one more likely to reflect the plastids themselves, is the number and arrangement of the genes within the genome. Genome architecture may be different from the phylogeny of the plastid genes themselves, and in the case of the dinoflagellates, quite remarkably so. Indeed, the only known plastid genes so far identified have been located on generally unigenic plasmid-like minicircles.

These minicircles have been found in at least five genera of peridinin-containing dinoflagellates including *Amphidinium* (Barbrook and Howe 2000), *Ceratium* (Laatsch, Zauner et al. 2004), *Heterocapsa* (Zhang, Green et al. 1999), *Protoceratium* (Zhang, Cavalier-Smith et al. 2002) and *Symbiodinium* (Moore, Ferguson et al. 2003). Each minicircle has regions conserved within a species, and extensive PCR amplification of the genes located between these conserved regions has been performed in several species. Taken together, these studies have led to the conclusion that the known minicircle gene complement has reached saturation (Koumandou, Nisbet et al. 2004) with a total of sixteen protein encoding genes (*atpA-B*; *petB* and *petD*; *psaA-B*; *psbA-E* and *psbI*; *ycf16* and *ycf24*; *rpl28* and *rpl33*) in addition to the large (23S) and small (16S) ribosomal RNA. The identification of this highly reduced set of plastid genes as comprising the plastid genome is also supported by recent results demonstrating that at least some of the missing genes (i.e., ones normally found in plastids) are instead nuclear-encoded in several species (Bachvaroff, Concepcion et al. 2004; Hackett, Yoon et al. 2004; Patron, Waller et al. 2005). These experiments suggest that dinoflagellates do not obey for the rules normally governing plastid gene transfer to the nucleus (Martin and Herrmann 1998).

However, it is still a formal possibility that additional genes may be encoded in the plastid in a form different from the minicircle format found so far. One method potentially applicable to the characterization of the plastid genome is to determine the spectrum of genes expressed, as gene expression should be independent from the form in which the genes are found. However, it is generally difficult to discriminate between organelle- and nuclear-encoded transcripts, as both can be modified by addition of a poly(A) tail at their 3' termini (Rott, Zipor et al. 2003). We have found that, unlike other transcripts in the dinoflagellates, those encoded by known minicircle genes carry a homogenous polyuridine tract at their 3' termini. We have taken advantage of this unusual feature to characterize the dinoflagellate plastid transcriptome, and find that our analysis fully supports a highly reduced

plastid genome for the peridinin-containing dinoflagellates. Furthermore, as it seems likely that polyuridylylation may be common in dinoflagellate plastids, it may be possible to rapidly characterize the transcriptome of many different species using the method described here.

Materials and Methods

Amphidinium carterae (CCMP 1314) and *Lingulodinium polyedrum* (CCMP 1936, formerly *Gonyaulax polyedra*) were obtained from the Provasoli-Guillard Culture Center for Marine Phytoplancton (Boothbay Harbor, Maine) and cultured as described (Wang, Jensen et al. 2005). Poly(A) RNA was purified from *Lingulodinium* and *Amphidinium* using oligo(dT) chromatography (Sambrook, Fritsch et al. 1989) and hybridized to a psbA and 23S probes as described (Wang, Jensen et al. 2005). The 16S probe was prepared against an EST sequence from our library (see below), while the atpB probe was prepared using previously described amino acid microsequence data (Wang, Jensen et al. 2005) to design two degenerate oligonucleotides 5'- TTYTICARGCIGGIWSIGARGT -3' and 5'-ACYTCIGCIACRAARAAIGGYTG -3'; the 500 bp PCR product was confirmed as atpB by sequence analysis.

The 3' end sequences of *atpB* and *psbA* transcripts were obtained from cDNA synthesized from poly(A) RNA tailed *in vitro* with rGTP by poly(A) polymerase. Specific sequences were amplified using a $d(C)_{10}$ oligonucleotide and *5'*-TATCCAATTGGACAAGGAAG-3' (*Lingulodinium psbA*) 5'- GTGTAGCACAAGATGTAAGC-3' (*Lingulodinium atpB*), 5'-GTATTCGGTCAAGAGGATG-3' (*Amphidinium psbA*) and 5'-CTATCTCAGCCGTTCTTTG-3' (*Amphidinium atpB*). The genomic *psbA* sequences from *Lingulodinium* were obtained using TAIL-PCR (Liu, Chen et al. 2005) using three nested internal *psbA* oligonucleotides (5'-GCTGCTTGGCCAGTTATTGGTATCTG-3'; 5'-TCTGGTTTACAGCACTTGGTGTTAG -3'; 5'-GTCCATAATTGATTCATCAGGTCATC -3') to amplify a fragment from AT-rich DNA purified by bisbenzimide-CsCl gradients. To obtain the 3' end of genomic sequences from *Amphidinium*, minicircle DNA was amplified by

inverse PCR using outwardly directed oligonucleotides 5'-GTATTCGGTCAAGAGGATG-3' and 5'-CAGGAGCAAGGAAGAAAG-3' (*psbA*) or 5'-GTGGTCGTAAGATTGAGAGG-3' and 5'-GATGAGAGCGTTGGCATAC-3' (*atpB*).

For cDNA preparation, 10 □g poly(A)-enriched RNA was used as a template for first strand synthesis using a commercial kit (Stratagene) but replacing the usual oligo(dT) primer with 5'-$(GA)_{10}$ACTAGTCTCGAG$(A)_{18}$-3'. Sequences were identified using the BLAST algorithm (www.ncbi.nlm.nih.gov) and have been deposited in GenBank under the accession numbers DQ264844 through DQ264867. All sequence alignments and analyses were performed using MacVector (Accelrys) except for RNA secondary predictions that were made using a web server (bioinfo.rpi.edu/applications/mfold). Statistical analysis using rarefaction (Hurlbert 1971) to determine the likelihood that our sample size was sufficient to detect all different cDNAs in the library was made on a web server (www2.biology.ualberta.ca/jbrzusto/rarefact.php).

Results

We began our analysis of the dinoflagellate plastid transcriptome by fractionation of RNA into poly(A)-enriched and -depleted fractions by oligo(dT) chromatography. The plastid-encoded *psbA* (Wang, Jensen et al. 2005) in RNA extracted from the dinoflagellates *Lingulodinium* (Figure III.1A) and *Amphidinium* (Figure III.1B), was found enriched by ~ 10 fold in poly(A^+) fractions, while 23S rRNA remained in the poly(A-) fraction. The small size of the 23S rRNA signal is suggestive of processing and has been previously observed in dinoflagellates (Zhang, Green et al. 1999). The 16S RNA was similarly processed, but the full-length form was found to a greater extent in the polyadenylated fraction than was the 23S RNA. The abundance of these transcripts in a poly(A) rich fraction was unexpected, as usually only a small fraction of plastid messages are polyadenylated (Rott, Zipor et al. 2003). Furthermore, it seemed at odds with the lack of minicircle genes found in

dinoflagellate EST libraries (Bachvaroff, Concepcion et al. 2004; Hackett, Yoon et al. 2004; Patron, Waller et al. 2005).

We originally thought that the tails might be heterogeneous, similar to the 3' termini in chloroplasts and cyanobacteria RNA formed by polynucleotide phosphorylase (PNP) (Rott, Zipor et al. 2003), since the presence of nucleotides other than adenine in the 3' tails might inhibit cDNA synthesis more than oligo(dT) chromatography. To test this, we guanylated our poly(A)-enriched RNA using poly(A) polymerase, and performed RT-PCR with a primer pair allowing specific amplification of the psbA 3' end (Fig. III.2A). Fourteen different psbA clones were sequenced, and all contained a homogeneous 3' terminal stretch of thymidine residues. At least five of the sequences obtained were independent clones based on differences in tail length, which varied between 25 and 40 thymidine residues (Fig. III.2B). A comparison to the genomic DNA sequence, obtained by thermal asymmetric interlaced (TAIL) PCR (Liu, Chen et al. 2005), shows that these residues are added post-transcriptionally, and all appeared to have been added at the same site (between arrows, Fig. III.2B). Poly(T) tracts of similar length were found on the 3' termini of atpB transcripts from *Lingulodinium* (not shown) and on atpB transcripts from the dinoflagellate *Amphidinium* (Fig. III.2C). We propose that binding of these polyuridylylated mRNAs to the longer and more prevalent poly(A) tails of nuclear transcripts is responsible for their presence in poly(A)-enriched RNA fractions obtained by oligo(dT) cellulose chromatography.

The unusual 3' terminal polyuridylylation, if a common feature of dinoflagellate plastid transcripts, suggested that analysis of cDNA synthesized using an oligo(dA) instead of the usual oligo(dT) primer would provide a straightforward method to catalog the plastid gene complement. Roughly 60 ng of cDNA was synthesized from 10 □g poly(A)-enriched RNA, a yield thirty-fold less than that obtained with an oligo(dT) primer in similar experiments. Polyuridylylated transcripts are thus a small but significant proportion of cellular RNA. To characterize the library, several hundred

clones were selected at random and sequenced. Some GC-rich and unidentified sequences were found, but none were polyuridylylated and were presumably derived from hairpin priming of the predominant GC-rich nuclear-encoded transcripts. In contrast, all polyuridylylated sequences were AT-rich and were identified as transcripts from the minicircle genes found in other dinoflagellates (Koumandou, Nisbet et al. 2004) (Table III.1). The majority of the transcripts correspond to photosystem II components (*psbA-D*) and 16S RNA. This latter may appear more frequently in our oligo(dA)-primed library than the 23S RNA because the full length 16S RNA appears more abundant in a poly(A) enriched sample (Fig. III.1A).

To assess the likelihood that other low abundance polyuridylylated transcripts might be present in the library, we performed rarefaction analysis (Hurlbert 1971). Originally developed to compare species richness in biodiversity collections of different sizes, this technique estimates the number of different cDNAs that would be obtained if smaller sample sizes were taken. The analysis of the data in Table I indicates that the identification of twelve different sequences could have been possible with only 120 random clones sequenced; the three hundred clones reported here represent a significant excess of this minimum value. The calculations can also be used to illustrate the progression in the number of random sequences required to reveal each additional sequence: while 10 different cDNAs are expected in 65 random sequences, almost twice as many sequences are required to uncover the eleventh sequence, and roughly six times more sequences to yield the twelfth cDNA (Fig. III.3 inset). This analysis suggests that to recover a new plastid transcript, well over a thousand random clones would have to be sequenced.

To further test the contention the twelve genes recovered are likely to represent saturation coverage of the oligo(dA) primed library, we employed a virtual subtraction protocol (Germain, Rudd et al. 2005). Here, roughly a thousand different cDNA clones were randomly selected and streaked on a new Petri plate in a grid configuration. Colony lifts were then hybridized with a probe prepared against a mixture of *psbA-D* or 16S sequences. From this, 50

cDNAs that hybridized weakly or not at were selected and sequenced. Only 14 sequences among the 50 sequenced were AT-rich and polyuridylylated, and these included two *atpA*, four 23S RNA, a *petB*, three *psbB*, three *psbC*, and a *psbD*. These later three photosystem II components may have been recovered following the virtual subtraction because of poor bacterial growth or poor transfer to the membrane. More importantly, no new polyuridylylated genes were identified, again suggesting that our coverage of the library had reached saturation. As it seems unlikely that alternative 3' modifications might be found within the same organelle, we conclude that the genome of this photosynthetically active chloroplast is likely to encode only the transcripts identified here.

To address the mechanism underlying site selection for poly(U) addition, we evaluated a potential role for both secondary structure and primary structure elements. Secondary structure is the principal determinant of polyadenylation site selection in prokaryotes (Sarkar 1997). However, computer generated secondary structure predictions from RNA complementary to the genomic sequence surrounding the polyuridylylation site of either *Lingulodinium psbA* or *Amphidinium atpB* did not show any reasonably stable stem-loop structures (not shown). However, two loosely conserved primary structure motifs are located within 50 bp of the modification site in all twelve transcripts (Fig. III.4). These motifs (AGAAA and AAUUA) might thus constitute primary structure elements signaling the polyuridylylation site in a manner similar to the use of AAUAAA in determining polyadenylation sites in nuclear encoded transcripts (Edmonds 2002).

We were also curious about the identity of the enzyme that might be used to catalyze the polyuridylylation reaction. As polyuridylylation of protein coding transcripts has been observed in organelles undergoing extensive uridine insertion editing (Adler, Harris et al. 1991; Horton and Landweber 2000), we thus checked for uridine insertion in comparisons of genomic and cDNA sequences using *Lingulodinium psbA*, *petB* and 16S RNA (~2.5 kb total sequence). Our data reveals no evidence for uridine insertion although

numerous examples of substitutional editing in dinoflagellate plastid transcripts (principally A to G) were observed (Table III.2, Supplementary Figures III.S1, S2). These results agree with a previous report of substitutional editing in the dinoflagellate *Ceratium* (Zauner, Greilinger et al. 2004).

Discussion

We have found that the plastid transcripts in two species of peridinin-containing dinoflagellates are characterized by an unusual 3' polyuridylylation. This modification differs dramatically from the poly(A) tails of nuclear-encoded transcripts, and so provides a facile method for cataloging the plastid transcriptome. We report here that an oligo(dA)-primed cDNA library has a remarkably low complexity, with only 12 different clones identified in 300 randomly selected polyuridylylated sequences (Table III.1). Interestingly, all the sequences identified from the library were previously identified as minicircle genes in other dinoflagellate species. This concordance between two independent methods (characterization of the minicircular genome and the transcriptome analysis reported here) suggests that the minicircular gene format is likely to be the only genome architecture in the plastid. Our results thus strongly support the contention that the dinoflagellate genome is the smallest of any functional chloroplast. Furthermore, since the dinoflagellate *Amphidinium* also contains polyuridylylated transcripts, it is possible that the technique described here could be used to catalog the plastid transcriptome from a range of other species.

Is it likely that the transcriptome contains other low abundance transcripts that were undetected in our relatively small sample size? The chance of finding a specific transcript in a single random sample is a function of its relative proportion, or frequency of occurrence within the bank. However, if many other sequences were present, then the chance of finding any other sequence would also depend on the number of additional sequences present and on their relative proportions within the library. To estimate our coverage

of the library, we used rarefaction analysis to determine if smaller sample sizes would also have recovered the same twelve genes. Our analysis suggests that it is likely the same twelve genes would have been recovered using a smaller sample size, indicating that the number of clones sequences was in excess of that required. Furthermore, the progression in the number of clones sequenced as a function of the number of different clones identified (Fig. III.3) suggests that well over a thousand random clones would have to be sequenced to find an additional clone if it were indeed present in the library. Taken together with the virtual subtraction of *psbA-D* and 16 S RNA sequences, these results strongly suggest that coverage of the library has reached saturation.

The selective forces that serve to maintain genes in the chloroplast are hotly debated, and thought to reflect either difficulties in targeting or transport of proteins that are extremely hydrophobic or a relationship between control of gene expression and the redox state of the organelle (Timmis, Ayliffe et al. 2004). With respect to the former, the twelve protein encoding genes found in our library are not the most hydrophobic of known thylakoid proteins. Indeed, a better explanation for the retention of this particular gene set in the plastid may lie in the length of the protein rather than hydrophobicity. As recently shown by analysis of *Arabidopsis* thylakoid proteins (Friso, Giacomelli et al. 2004), the proteins encoded by the dinoflagellate plastid genes are generally the longest of the thylakoid proteins. However, even a combination of length and hydrophobicity is insufficient to explain transfer of some genes to the nucleus, such as the shorter yet relatively hydrophobic protein encoded by the *atpI* gene recently reported in an oligo(dT) primed EST bank from the dinoflagellate *Alexandrium*. If instead of hydrophobicity, the redox control of gene expression were the determining factor, this would suggest the set of proteins encoded by the dinoflagellate plastids are most sensitive to changes in redox potential. This hypothesis could potentially be tested in other plastids. The proposal that it is genes requiring editing may that are conserved in plastids (Bungard 2004) is supported by analysis of the *petB* gene of

Lingulodinium, as editing removes a stop codon in the middle of the derived protein sequence (Supplementary Figure III.S2).

With respect to the unusual nature of the modification itself, there is precedent for polyuridylylation in organelles experiencing extensive editing, such as the mitochondria of trypanosomes or myxomycetes. In some cases, 3' polyuridine tails have been observed not only for the guide RNAs used to determine the site of uridine insertion but for rRNA and mRNA as well (Adler, Harris et al. 1991; Horton and Landweber 2000). However, a comparison of ~2.5 kb genomic and cDNA sequences showed only substitutional editing (Table III.2). Our experiments thus provide the first example of extensive polyuridylylation occurring in the absence of RNA editing. It is tempting to speculate that the poly(U) tracts may result from a terminal uridylyltransferase (TUT) (Aphasizhev 2005) acting in the absence of guide RNA to define a site for uridine insertion. However, an alternative possibility is that the plastids may contain a poly(A) polymerase specific for uridine residues instead of adenine.

Recently, it has also been shown that microRNA-directed cleavage can result in addition of non-encoded oligonucleotides (mostly uridine) to 3' termini (Shen and Goodman 2004). However, in this case the polyuridylylated transcripts are intermediates in an RNA degradation pathway. It seems unlikely that the extensive polyuridylylation of the plastid transcripts is used as a signal for RNA turnover, since the majority of psbA transcripts appear modified as judged by their copurification on oligo(dT) chromatography. In contrast to the full-length protein coding transcripts, the small fragments hybridizing to 16S and 23S RNA on Northern analyses appear unmodified by these criteria (Fig. III.1). Interestingly, the apparent stability of the polyuridylylated plastid transcripts thus suggests that this particular 3' end modification has a different function from that in either cyanobacteria or the plastids of higher plants, where transcripts are marked for degradation by 3'end polyadenylation (Rott, Zipor et al. 2003).

Despite extensive molecular phylogenetic studies pointing to a common evolutionary origin for both host cells (Harper, Waanders et al. 2005) and plastids (Yoon, Hackett et al. 2005) of the chromalveolates, no other plastids or cyanobacteria are known to polyuridylylate transcripts. Thus the mechanism and function of the dinoflagellate plastid transcript 3' end modification are as unique as their form II Rubisco (Morse, Salois et al. 1995) and peridinin-chlorophyll a-protein (Hofmann, Wrench et al. 1996). The major challenge for plastid phylogeny underscored by our results is to reconcile the many unique features of the dinoflagellate plastids with their phylogenetic relationships to red algae. In addition, given the concurrence of several lines of evidence supporting the highly reduced nature of the dinoflagellate plastid genome, it will be of interest to reinvestigate the nature of the selective forces maintaining the current plastid genome size in higher plants.

Acknowledgements

We thank M. Lapointe, A. Lukombo and A. Tran for technical assistance, Drs. B. F. Lang for helpful discussion, P. Legendre for suggesting the rarefaction analysis, and M. Hijri and M. Cappadocia for critical reading of the manuscript. The present work was funded by the National Science and engineering Research Council of Canada (171382-03 to DM). The Open Access publication charges for this article were waived by Oxford University Press.

Figure legends

Figure III.1 Dinoflagellate plastid messages are located in poly(A)-enriched RNA

Total RNA samples (T) from the dinoflagellates *Lingulodinium* (A) and *Amphidinium* (B) were resolved into fractions enriched (A+) and depleted (A-) for poly(A) RNA by chromatography on oligo(dT) cellulose. RNA blots were challenged with gene probes for either psbA, 23S RNA or 16S. Lower panels show the ethidium bromide stained gels.

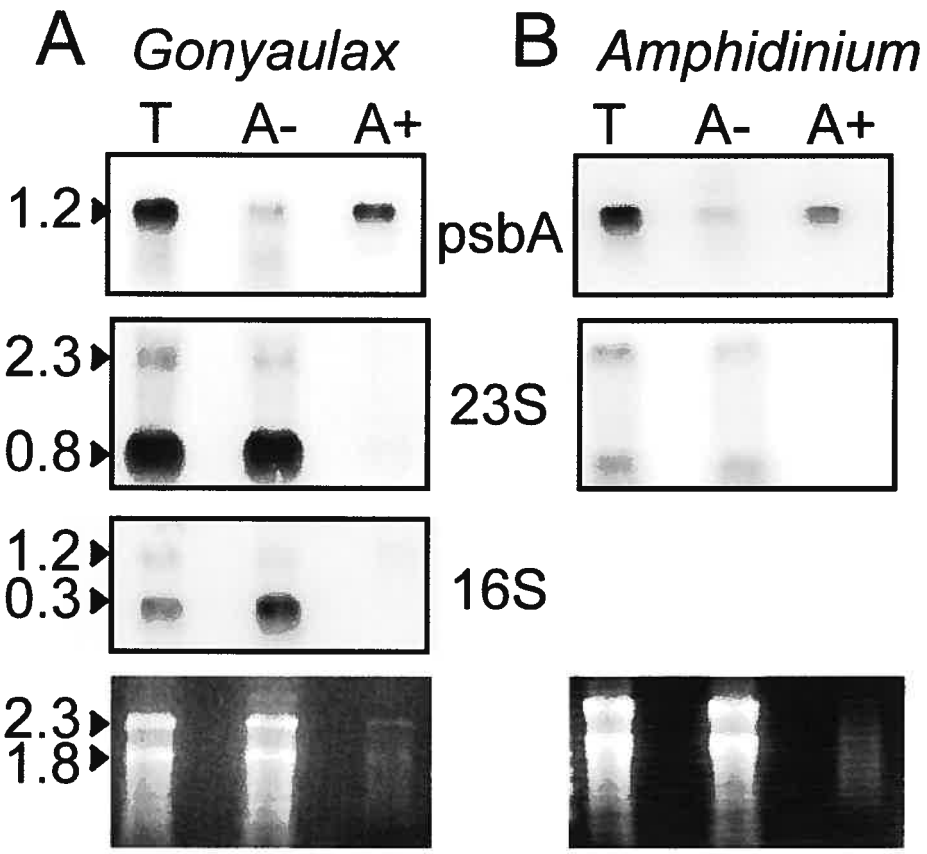

A
Gonyaulax
B
Amphidinium
T
A-
A+
T
A-
A+
1.2
psbA
2.3
0.8
23S
1.2
0.3
16S
2.3
1.8

Figure III.2 Plastid transcripts contain a homogenous poly(U) tail at a specific site

Lingulodinium RNA samples enriched for poly(A) RNA were tailed with guanine residues, and transcript 3' end sequences amplified using a specific internal oligonucleotide and an oligo(dC) oligonucleotide. (B) Sequences of the 3' end of fourteen *psbA* cDNAs yielded five clones with different numbers of thymidine residues. The cDNA sequences are aligned with genomic *psbA* sequences obtained by TAIL-PCR. The polyuridylylation site is defined to the stretch of thymidine residues encoded by the genomic sequence (arrows). The asterisk indicates the termination codon. (C) *Amphidinium* RNA samples were treated similarly except that *atpB* specific primers were used in the amplification. The cDNA sequences are aligned with minicircle *atpB* sequences obtained by inverse PCR.


A
gggg
cccc
B
Gonyaulax psbA
gDNA GATTTAGCTTGACTTTTCTTGTGTTTTTGGTTGGATCCAGACATGAAGTGCCCTGTCTC
cDNA
GATTTAGCTTGACTTTTCTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTT
GATTTAGCTTGACTTTTCTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTT
GATTTAGCTTGACTTTTCTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
GATTTAGCTTGACTTTTCTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
GATTTAGCTTGACTTTTCTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
C
Amphidinium atpB
gDNA GCGATGTCCGATTAGTTGAATTGTGTAGGTAAATAATATTATTAAGAAACGAAACGTAC
cDNA
GCGATGTCCGATTAGTTGAATTGTGTTTTTTTTTTTTTTTTTTTTTTTTTT
GCGATGTCCGATTAGTTGAATTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTT
GCGATGTCCGATTAGTTGAATTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

Figure III.3 The library is likely to contain only 12 different cDNA sequences.

Estimates of the number of different sequences expected for different numbers of random clones sequenced were made by rarefaction analysis of the data in Table I. The estimated number of different cDNA sequences expected with smaller sample sizes shows a statistical possibility that the twelve different clones could have been identified with only 120 different sequences. The progression of the number of random sequences as a function of the number of different clones (inset) suggests over a thousand sequences would be required to identify any potentially new clone in the library.

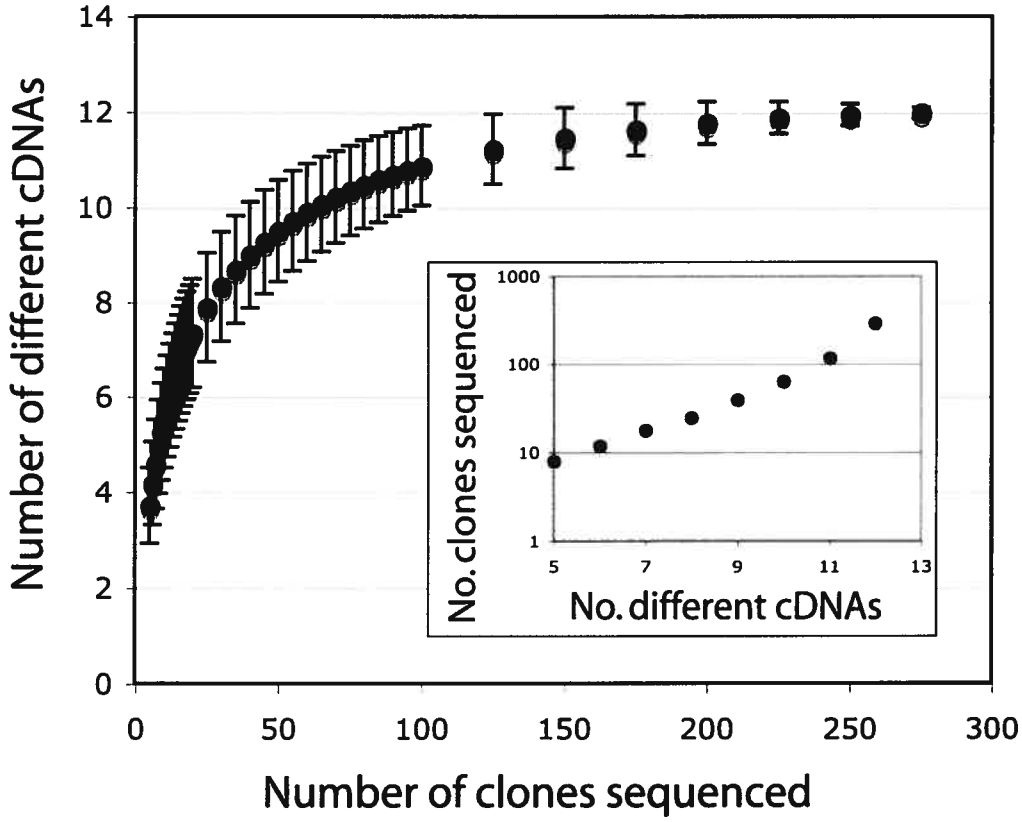
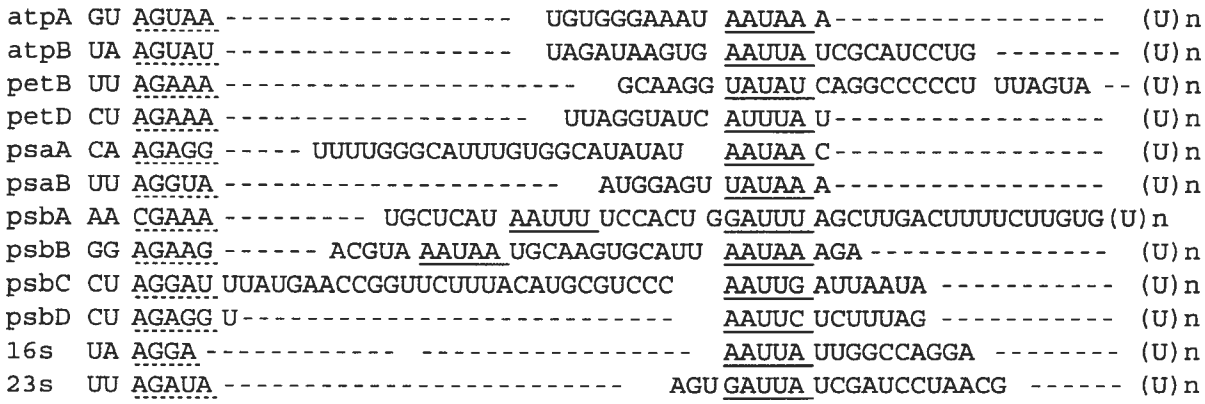

Number of different cDNAs
Number of clones sequenced
No. clones sequenced
No. different cDNAs

Figure III.4 Plastid transcripts may contain primary structure motifs for polyuridylylation.

All different polyuridylylated 3' terminal sequences corresponding to sequences identified in the library were aligned at the site of the poly(U) modification and at each of two potential conserved sequences (underlined).


atpA GU AGUAA ---------- UGUGGGAAAU AAUAA A---------- (U)n
atpB UA AGUAU ---------- UAGAUAAGUG AAUUA UCGCAUCCUG -------- (U)n
petB UU AGAAA ---------- GCAAGG UAUAU CAGGCCCCCU UUAGUA -- (U)n
petD CU AGAAA ---------- UUAGGUAUC AUUUA U---------- (U)n
psaA CA AGAGG ----- UUUUGGGCAUUUGUGGCAUAUAU AAUAA C---------- (U)n
psaB UU AGGUA ---------- AUGGAGU UAUAA A---------- (U)n
psbA AA CGAAA --------- UGCUCAU AAUUU UCCACU GGAUUU AGCUUGACUUUUCUUGUG (U)n
psbB GG AGAAG ------ ACGUA AAUAA UGCAAGUGCAUU AAUAA AGA---------- (U)n
psbC CU AGGAU UUAUGAACCGGUUCUUUACAUGCGUCCC AAUUG AUUAAUA ---------- (U)n
psbD CU AGAGG U---------- AAUUC UCUUUAG ---------- (U)n
16s UA AGGA ---------- ---------- AAUUA UUGGCCAGGA -------- (U)n
23s UU AGAUA ---------- AGU GAUUA UCGAUCCUAACG ------ (U)n

Supplemental Figure legends

Figure III.S1

Editing sites in psbA defined by differences in genomic and cDNA sequences.

The genomic sequence of the psbA gene and its derived amino acid sequence are shown in full, and only differences are shown for the cDNA sequences. An asterisk shows the only position in the sequenced where editing (C to T) does not result in an amino acid replacement.

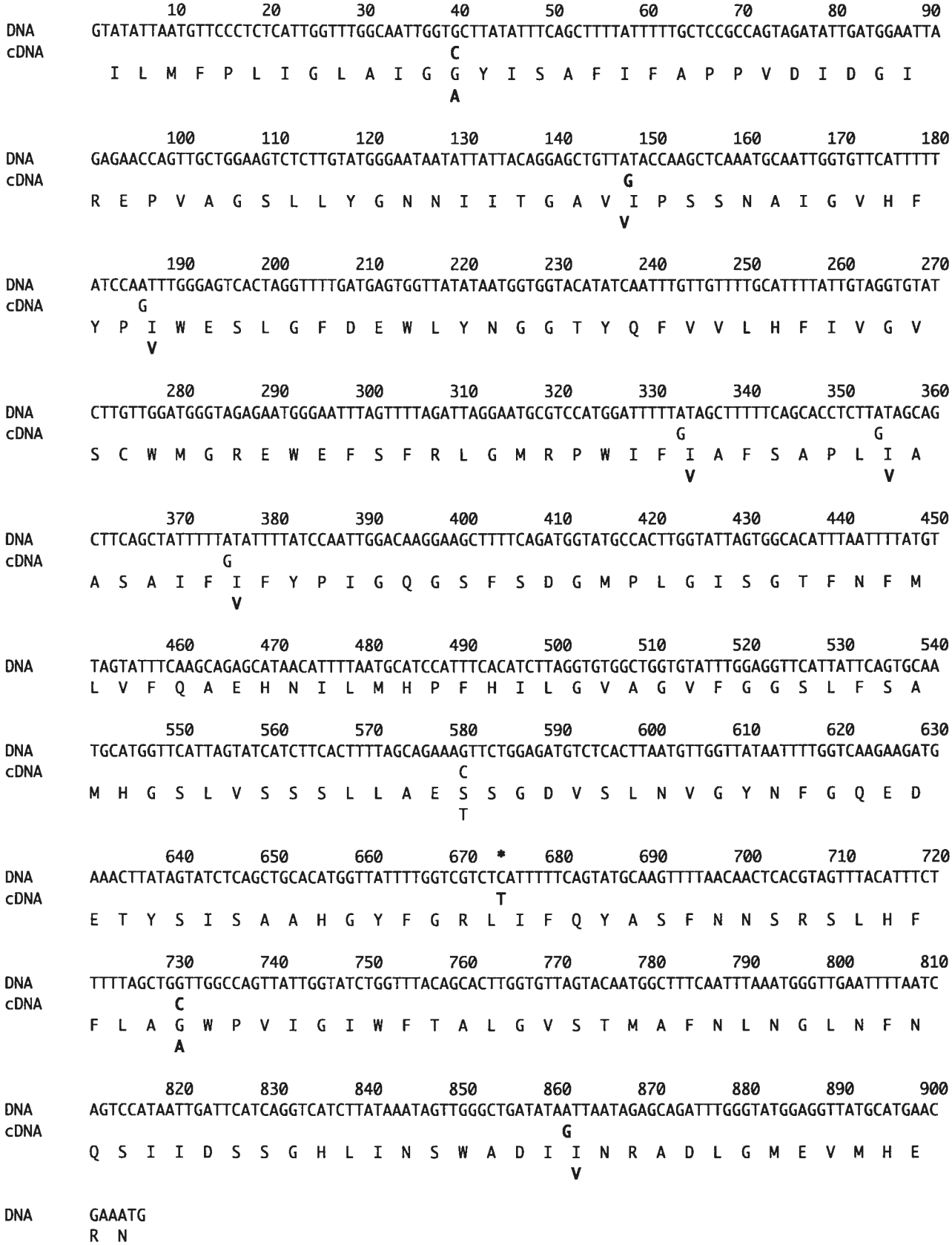

10 20 30 40 50 60 70 80 90
DNA GTATATTAATGTTCCCTCTCATTGGTTTGGCAATTGGTGCTTATATTTCAGCTTTTATTTTTGCTCCGCCAGTAGATATTGATGGAATTA
cDNA C A
I L M F P L I G L A I G G Y I S A F I F A P P V D I D G I
100 110 120 130 140 150 160 170 180
DNA GAGAACCAGTTGCTGGAAGTCTCTTGTATGGGAATAATATTATTACAGGAGCTGTTATACCAAGCTCAAATGCAATTGGTGTTCATTTTT
cDNA G
R E P V A G S L L Y G N N I I T G A V I P S S N A I G V H F
V
190 200 210 220 230 240 250 260 270
DNA ATCCAATTTGGGAGTCACTAGGTTTTGATGAGTGGTTATATAATGGTGGTACATATCAATTTGTTGTTTTGCATTTTATTGTAGGTGTAT
cDNA G
Y P I W E S L G F D E W L Y N G G T Y Q F V V L H F I V G V
V
280 290 300 310 320 330 340 350 360
DNA CTTGTTGGATGGGTAGAGAATGGGAATTTAGTTTTAGATTAGGAATGCGTCCATGGATTTTTATAGCTTTTTCAGCACCTCTTATAGCAG
cDNA G G
S C W M G R E W E F S F R L G M R P W I F I A F S A P L I A
V V
370 380 390 400 410 420 430 440 450
DNA CTTCAGCTATTTTTATATTTTATCCAATTGGACAAGGAAGCTTTTCAGATGGTATGCCACTTGGTATTAGTGGCACATTTAATTTTATGT
cDNA G
A S A I F I F Y P I G Q G S F S D G M P L G I S G T F N F M
V
460 470 480 490 500 510 520 530 540
DNA TAGTATTTCAAGCAGAGCATAACATTTTAATGCATCCATTTCACATCTTAGGTGTGGCTGGTGTATTTGGAGGTTCATTATTCAGTGCAA
L V F Q A E H N I L M H P F H I L G V A G V F G G S L F S A
550 560 570 580 590 600 610 620 630
DNA TGCATGGTTCATTAGTATCATCTTCACTTTTAGCAGAAAGTTCTGGAGATGTCTCACTTAATGTTGGTTATAATTTTGGTCAAGAAGATG
cDNA C
M H G S L V S S S L L A E S S G D V S L N V G Y N F G Q E D
T
640 650 660 670 * 680 690 700 710 720
DNA AAACTTATAGTATCTCAGCTGCACATGGTTATTTTGGTCGTCTCATTTTTCAGTATGCAAGTTTTAACAACTCACGTAGTTTACATTTCT
cDNA T
E T Y S I S A A H G Y F G R L I F Q Y A S F N N S R S L H F
730 740 750 760 770 780 790 800 810
DNA TTTTAGCTGGTTGGCCAGTTATTGGTATCTGGTTTACAGCACTTGGTGTTAGTACAATGGCTTTCAATTTAAATGGGTTGAATTTTAATC
cDNA C
F L A G W P V I G I W F T A L G V S T M A F N L N G L N F N
A
820 830 840 850 860 870 880 890 900
DNA AGTCCATAATTGATTCATCAGGTCATCTTATAAATAGTTGGGCTGATATAATTAATAGAGCAGATTTGGGTATGGAGGTTATGCATGAAC
cDNA G
Q S I I D S S G H L I N S W A D I I N R A D L G M E V M H E
V
DNA GAAATG
R N

Figure S1

Figure III.S2

Editing sites in petB defined by differences in genomic and cDNA sequences.

The genomic sequence of the petB gene and its derived amino acid sequence are shown in full, and only differences are shown for the cDNA sequences. The asterisk shows the only position in the sequence where editing changes a stop codon to instead encode an amino acid (lysine).

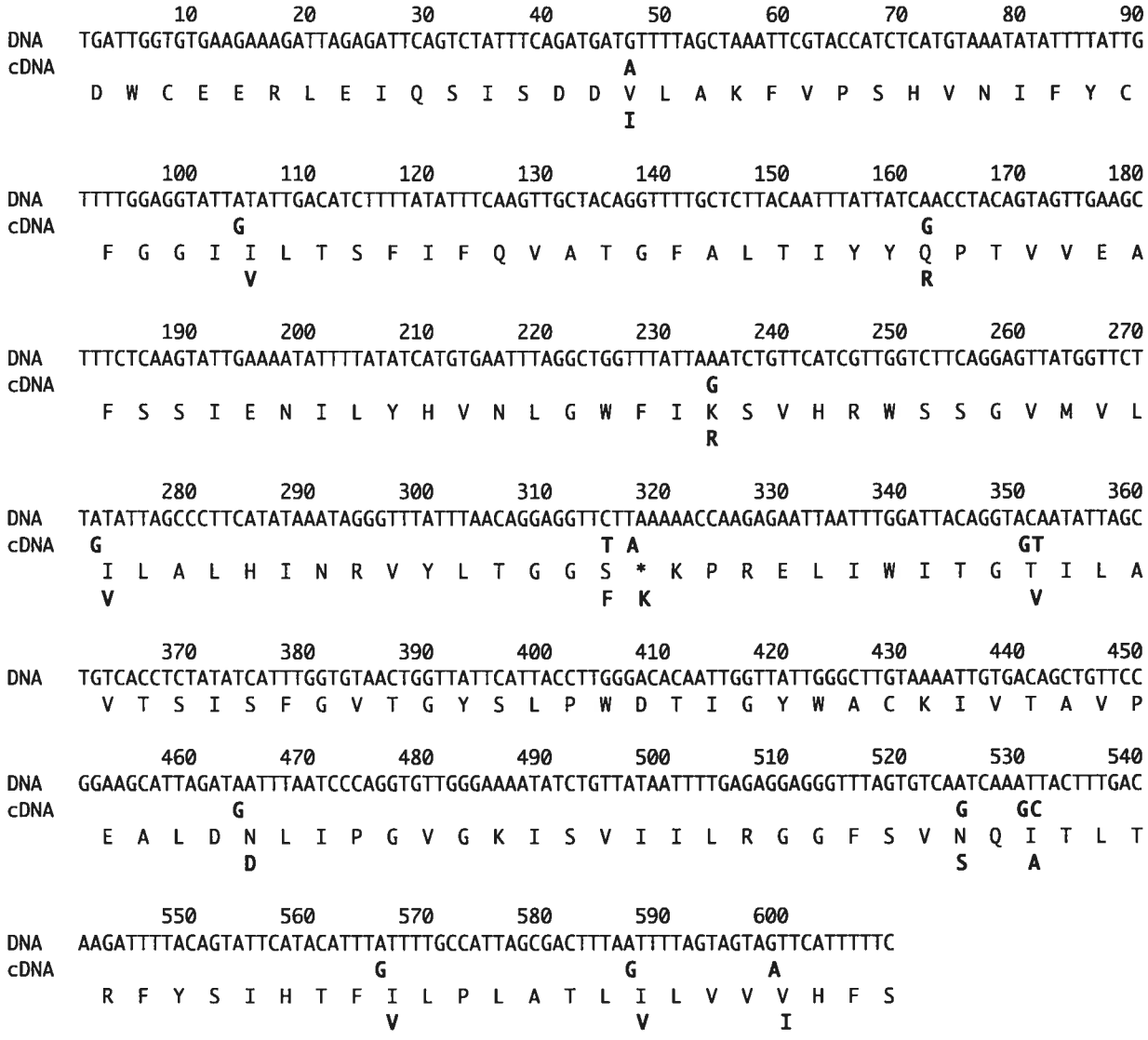

10 20 30 40 50 60 70 80 90
DNA TGATTGGTGTGAAGAAAGATTAGAGATTCAGTCTATTTCAGATGATGTTTTAGCTAAATTCGTACCATCTCATGTAAATATATTTTATTG
cDNA A
D W C E E R L E I Q S I S D D V L A K F V P S H V N I F Y C
I
100 110 120 130 140 150 160 170 180
DNA TTTTGGAGGTATTATATTGACATCTTTTATATTTCAAGTTGCTACAGGTTTTGCTCTTACAATTTATTATCAACCTACAGTAGTTGAAGC
cDNA G G
F G G I I L T S F I F Q V A T G F A L T I Y Y Q P T V V E A
V R
190 200 210 220 230 240 250 260 270
DNA TTTCTCAAGTATTGAAAATATTTTATATCATGTGAATTTAGGCTGGTTTATTAAATCTGTTCATCGTTGGTCTTCAGGAGTTATGGTTCT
cDNA G
F S S I E N I L Y H V N L G W F I K S V H R W S S G V M V L
R
280 290 300 310 320 330 340 350 360
DNA TATATTAGCCCTTCATATAAATAGGGTTTATTTAACAGGAGGTTCTTAAAAACCAAGAGAATTAATTTGGATTACAGGTACAATATTAGC
cDNA G T A GT
I L A L H I N R V Y L T G G S * K P R E L I W I T G T I L A
V F K V
370 380 390 400 410 420 430 440 450
DNA TGTCACCTCTATATCATTTGGTGTAACTGGTTATTCATTACCTTGGGACACAATTGGTTATTGGGCTTGTAAAATTGTGACAGCTGTTCC
V T S I S F G V T G Y S L P W D T I G Y W A C K I V T A V P
460 470 480 490 500 510 520 530 540
DNA GGAAGCATTAGATAATTTAATCCCAGGTGTTGGGAAAATATCTGTTATAATTTTGAGAGGAGGGTTTAGTGTCAATCAAATTACTTTGAC
cDNA G G GC
E A L D N L I P G V G K I S V I I L R G G F S V N Q I T L T
D S A
550 560 570 580 590 600
DNA AAGATTTTACAGTATTCATACATTTATTTTGCCATTAGCGACTTTAATTTTAGTAGTAGTTCATTTTTC
cDNA G G A
R F Y S I H T F I L P L A T L I L V V V H F S
V V I

Figure S2

Table III.1

Only transcripts from known minicircle genes are polyuridylylated in *Lingulodinium*

Gene	Size (kb)	% Clones (n)
atpA	1.5	5.7 % (17)
atpB	1.4	0.7 % (2)
psaA	2.5	3.7 % (11)
psaB	2.3	2.0 % (6)
psbA	1.2	23.0 % (69)
psbB	1.6	9.3 % (28)
psbC	1.4	12.7 % (38)
psbD	1.3	17.0 % (51)
petB	0.9	0.7 % (2)
petD	0.5	2.3 % (7)
16S rRNA	1.2	21.0 % (63)
23S rRNA	2.3	2.0 % (6)

Table III.2

Various patterns of substitutional editing are found in dinoflagellate plastid transcripts

	Number of modifications			
Modification	petB	psbA	16S	Total
A -> G	10	6	16	32
G -> A	2	0	1	3
C -> T	2	1	1	4
T -> C	1	0	0	1
T -> G	0	0	1	1
G -> C	0	3	0	3

Chapter IV. Purification of Plastids from the Dinoflagellate *Lingulodinium*

Objectives.

1) **To purify chloroplasts from dinoflagellate *Gonyaulax* (now *Lingulodinium*).**

2) **To analyze the integrity of purified plastids using antibodies.**

3) **To analyze the possibility of using purified plastid for organelle proteomic studies and potentially genomic studies.**

Publication of chapter IV: This manuscript has been published online in Marine Biotechnology, 23 August 2005.

All the experimental procedures were done by me except for oxygen evolution rates that were measured by Tyler MacKenzie.

Purification of Plastids from the Dinoflagellate *Lingulodinium*

Yunling Wang, Tyler MacKenzie and David Morse*

Département de Sciences Biologiques, Université de Montréal, 4101 Sherbrooke est, Montreal, Quebec, H1X 2B2, Canada

***Corresponding author**

David Morse

[redacted]

Phone:514-872-9975

Fax: 514-872-9406

Abstract

Peridinin-containing dinoflagellates are a group of generally marine and photosynthetic protists whose plastids display a number of unusual features. In particular, the plastid genome may be reduced to as few as a dozen genes, and it is not clear if all these genes are expressed. To begin to characterize the plastid proteins, we attempted to purify chloroplasts from the dinoflagellate *Lingulodinium polyedrum*. We tested several different protocols and found that the organelles were inherently fragile and difficult to isolate intact. In particular, standard purification protocols as described for higher plants produced only broken plastids, as judged by complete loss of the stromal protein RuBisCO. We found that small amounts of RuBisCO could be retained in the plastids if the cells were treated with cellulase prior to lysis. Finally, we report that almost all RuBisCO was retained in plastids prepared from cells subjected to a heat shock treatment, although cellular proteins were denatured by the treatment.

Keywords chloroplast - dinoflagellate - purification - RuBisCO - PCP

Introduction

The peridinin-containing plastids found in most photosynthetic dinoflagellates have a number of particular morphologic and biochemical characteristics. For example, these plastids are surrounded by 3 membranes (*Gibbs, 1981*) rather than the more typically found 2 or 4. The peridinin itself is an allenic oxi-carotenoid (Jeffrey, Seilicki et al. 1975), which has been found to date only in photosynthetic dinoflagellates, and is found together with chlorophyll *a* bound to peridinin–chlorophyll *a* proteins (PCPs) (Prezelin and Haxo 1976). The principal light-harvesting protein in the algae is a water-soluble PCP whose structure is clearly unrelated to any other light-harvesting proteins (Norris and Miller 1994; Hofmann, Wrench et al. 1996). An intrinsic membrane protein (iPCP) related to the chlorophyll *ab*-binding protein of higher plants is also found (Grossman, Bhaya et al. 1995), but is thought to be of lower abundance than the soluble form. Perhaps even more unusual is the finding that the plastids contain a form II ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), previously found only in some species of anaerobic proteobacteria (Morse, Salois et al. 1995; Rowan, Whitney et al. 1996). The form II enzyme differs from the typical form I enzyme in that it is composed only of large subunits that share limited sequence identity with those of the form I enzyme (Narang, McIntosh et al. 1984). Importantly, oxygen competes with CO_2 more successfully for binding to the active site of the form II enzyme than to that of the form I enzyme (Whitney and Andrews 1998), suggesting that special mechanisms might be required to allow this enzyme to function efficiently (Nassoury, Fritz et al. 2001). Finally, the sequence of a nuclear-encoded plastid-directed glyceraldehyde-3-phosphate dehydrogenase (GAPDH) forms a monophyletic group with the plastid isoforms of cryptomonads, within an otherwise cytosolic clade, and is thus distinct from all other plastid GAPDHs (Fagan, Hastings et al. 1998).

Molecular studies have demonstrated that the genome of at least some peridinin–containing dinoflagellate plastids is also quite unconventional, as it

appears to be encoded on small single-gene or dual-gene minicircles (Zhang, Green et al. 1999; Barbrook and Howe 2000; Barbrook, Symington et al. 2001; Hiller 2001). It has been suggested that the isolation of minicircle sequences based on polymerase chain reaction (PCR) may have reached saturation (Howe, Barbrook et al. 2003), and if so, the dinoflagellate plastid would contain one of the most reduced plastid genomes known. To date, the genes found in minicircles include the 16S and 23S rRNAs, 2 photosystem I (PSI) components (*psaA* and *psaB*), 6 photosystem II (PSII) genes (*psbA*, *psbB*, *psbC*, *psbD*, *psbE* and *psbI*), 2 subunits of ATP synthase (*atpA* and *atpB*) and two cytochrome b_6f subunits (*petB* and *petD*).

The strong emphasis on dinoflagellate plastid genes has been driven by an interest in the evolutionary history of dinoflagellate plastids (Zhang, Green et al. 2000). In contrast, our understanding of the plastid proteome is still at a very basic level. Indeed, the primary interest in plastid proteins has been the relation between the circadian rhythms of carbon fixation and oxygen evolution and clock-controlled regulation of plastid protein synthesis (Fagan, Morse et al. 1999; Le, Jovine et al. 2001; Nassoury, Fritz et al. 2001). A number of plastid proteins have been identified and characterized, but so far all have been found to be nuclear-encoded (Hiller, Wrench et al. 1995; Morse, Salois et al. 1995; Le, Markovic et al. 1997; Fagan, Morse et al. 1999). Nuclear-encoded plastid-directed proteins can be readily identified by virtue of an unusual leader sequence containing 2 hydrophobic regions surrounding a region rich in hydroxylated amino acids, apparently a requirement of the targeting mechanism into triple membrane bound plastids (Nassoury, Cappadocia et al. 2003).

Plastid purification from dinoflagellates has been reported only once (Asano, Okamoto et al. 1998). This report showed that purified organelles had retained protein kinase activity, but it did not evaluate plastid integrity or the degree of purification. We have found the issue of plastid integrity to be the most important during our attempts to develop a purification protocol. We have therefore developed a sensitive immunologic assay for plastid integrity

using measurements of known protein marker retention. To date the only molecular markers identified in the *Lingulodinium* chloroplast are RuBisCO (Morse, Salois et al. 1995), PCP (Le, Markovic et al. 1997) and GAPDH (Fagan, Morse et al. 1999). We have exploited the previously prepared anti-RuBisCO and anti-PCP antibodies (*Nassoury et al., 2001*) to follow retention of the stromal marker RuBisCO and the thylakoid lumen marker PCP during purification. The results of these immunologic assessments were also confirmed by 2-dimensional polyacrylamide gel electrophoresis (2D-PAGE), the type of analysis likely to be of particular utility in proteomic characterization of soluble stromal as well as lumenal and peripheral thylakoid proteins in plastid preparations (Peltier, Friso et al. 2000). Our studies underscore the importance of organelle integrity and document the plastid protein profiles resulting from several different purification protocols.

Materials and Methods

Cell Culture

Lingulodinium polyedrum (previously *Gonyaulax polyedra*) was obtained from the Provasoli-Guillard Culture Center for Marine Phytoplankton (strain 1936, Booth Bay Harbor, Maine) and grown in a modified seawater medium (f/2) (Guillard and Ryther 1962) at constant temperature (16°C) in 12:12 light-dark cycles using cool white fluorescent light at an intensity of 40 µmol $m^{-2}s^{-1}$. The beginning of the light period was defined as time 0 (LD 0), and the beginning of the dark period as LD 12. Cultures were grown to a cell density of 5000 to 6000 cells ml^{-1} as measured by counting using a hemocytometer. The cell cultures are unialgal but not axenic, and the prokaryotes in the cultures have not been defined.

Plastid isolation protocols

All plastid purification protocols used cells harvested by filtration on Whatman 541 paper and washed 3 times with sterile culture medium before use to remove bacteria. The standard plastid purification protocol, (*protocol 1*) as used for higher plants (Schuler and Zielinski 1989), involved resuspension of the cells in isolation buffer A (100 mM Tris-HCl buffer, pH 7.5, containing 0.3 M sucrose, 10 mM EDTA, and 14 mM ß-mercaptoethanol) and a 5-minute incubation at 2000 psi in an ice-cold N_2 bomb (Parr Instruments). The cells were lysed as the pressure was released slowly enough to produce discrete drops of liquid. This crude extract was centrifuged at 120*g* for 3 minutes to remove unbroken cells and cell debris. Five milliliters of this clarified lysate was then loaded on a 15-ml 10% to 30% linear Percoll (Amersham-Pharmacia Biotech) gradient in isolation buffer atop a 60% Percoll cushion and centrifuged at 6000*g* for 25 minutes at 4°C. The pigmented bands were isolated and the Percoll was removed by centrifugation after dilution with 10 volumes of isolation buffer. These preparations were analyzed using

antibodies to monitor retention of marker proteins and 2-dimensional gel electrophoresis to characterize the protein complement as described below. Several variations on a "stirred plastid" preparation were developed that purified plastids with low but detectable levels of RuBisCO (protocol 2). In one version of this protocol, cells were harvested by filtration and washed as described above. Cells from 4 L of culture medium were resuspended in 10 ml isolation buffer B (0.626 M Sorbitol, 0.1 M KCl, 0.1 M Hepes-NaOH, pH 7.2, 20 mM ß-mercaptoethanol, 0.1 % bovine serum albumin [BSA]) containing 0.6 g cellulase. The suspension was shaken gently at 16°C for 1 hour, and then transferred to a prechilled nitrogen bomb (Parr Instruments). The nitrogen pressure was brought to 1100 psi and released immediately. The suspension was centrifuged at 120*g* for 3 minutes, and the pellets containing still unbroken cells were resuspended in a low osmotic strength buffer (0.25 M Sorbitol, 0.1 M Hepes-NaOH, pH7.2, 0.1 M KCl, 0.1 % BSA) and stirred vigorously with a magnetic stirrer for 20 minutes in an ice bath. This stirred homogenate was centrifuged at 120 rpm for 3 minutes to remove unbroken cells, and the supernatant was centrifuged at 1000*g* for 3 minutes. This final pellet represented the crude plastid preparation and was resuspended in 10 ml isolation buffer B. As an alternative to this procedure, cells were resuspended in 10 ml isolation buffer B containing 0.05% Saponin (ICN) instead of the low ionic strength buffer, and the suspension was stirred vigorously for 15 minutes on ice. This stirred homogenate was centrifuged as described above to produce the plastid-containing pellet.

Plastids were purified from the crude preparations by density gradient centrifugation. A stock solution of PBF-Percoll (3% [w/v]) PEG 8000, 1% (w/v) BSA and 1% [w/v] Ficoll) was first prepared by dissolving the dry powders in 100% Percoll. A 15-ml 10% to 30% linear PBF-Percoll gradient in isolation buffer was laid on top of a 5-ml 40% PBF-Percoll in isolation buffer cushion. Five milliliters of the crude plastid preparation was layered on top of the Percoll gradients and centrifuged at 6,000 *g* for 25 minutes at 4°C (Beckman JS13.1 rotor). Two distinct highly pigmented bands were found in Percoll, and

the lower band was collected with a Pasteur pipette. These plastid preparations, examined microscopically, typically contained more than 90% aggregated and pigmented plastids. To remove the Percoll from the plastid preparation, 3 volumes of cold washing buffer (isolation buffer without BSA) were added, gently mixed, and the plastid preparation was centrifuged at 1400*g* (Beckman JS13.1 rotor) for 15 minutes at 4°C. This process was repeated 3 times, each time with fresh washing buffer and gentle resuspension of the plastid pellet using a Pasteur pipette.

The plastids that retained the most RuBisCO were prepared using a 1-hour treatment with cellulase at 40°C to degrade the cell wall (*protocol 3*). For this "heat shock" preparation, cells were resuspended in 10 ml cellulase digestion buffer (0.4 M sucrose, 0.02 g/ml cellulase) per 4 L of cell culture. The suspension was incubated with gentle shaking at 40°C for 1 hour to digest the cell wall. The cells were then centrifuged and resuspended in 15 ml isolation buffer A (0.4 M sucrose, 0.1 M KCl, 0.1 M Hepes-NaOH [pH 7.2], 0.03 M ß-mercaptoethanol, 0.1% BSA), and transferred to a prechilled nitrogen bomb (Parr Instruments). The nitrogen pressure was brought to 500 psi then immediately and slowly released. All subsequent steps were carried out at 4°C. The cell homogenates were centrifuged at 120*g* for 10 minutes to pellet down unbroken cells and cell debris. The pellets were resuspended in 10 ml isolation buffer, and the bomb treatment was repeated 3 times. The combined supernatants constituted the crude plastid preparations from which plastids were purified as for the stirred plastid protocol described above.

Immunologic Assay for Plastid Integrity

At each step of the purification, 0.2% of the volume of each sample was removed for gel analysis, including the initial total protein extract, the crude extract after centrifugation at 120*g* (to remove whole cells and large cellular debris), the PBF-Percoll purified plastids, and the supernatant remaining above PBF-Percoll gradient. These samples were dissolved in sodium dodecylsulfate (SDS) sample buffer (Laemmli 1970) to a final volume of 100 µl, from which 10 µl was electrophoresed on 10% polyacrylamide gels and

transferred electrophoretically to nitrocellulose membranes. The different fractions are directly comparable because they contain similar percentages of the total fraction. This method was chosen instead of measurements of protein concentration in each fraction, as many of the fractions contained BSA, added to the extraction buffer as a stabilizing agent.

The protein blots were incubated with a mixture containing a 1:5000 dilution of rabbit anti-form II RuBisCO, raised against a *Lingulodinium* RuBisCO cDNA expressed in *Escherichia coli*, and a 1:5000 dilution of rabbit anti-PCP, raised against PCP purified from *Lingulodinium* by a combination of column chromatography and preparative electrophoresis (Nassoury, Fritz et al. 2001). Antibody binding was visualized using a peroxidase-linked goat antirabbit secondary antibody and chemiluminescence (Amersham).

Oxygen Evolution Measurements

Oxygen evolution rates were measured from whole cells and purified plastid preparations essentially as described (Dionisio-Sese, Maruyama et al. 2001). Briefly, whole cells were concentrated by centrifugation and resuspended at 10 µg/ml chlorophyll *a* in f/2 medium containing 5 mM $NaHCO_3$. Purified plastid preparations were measured at similar chlorphyll concentrations in plastid isolation buffer containing 5 mM $NaHCO_3$, and 2.5 mM potassium ferricyanide ($FeK_3(CN)_6$) and 1 mM methyl viologen were also added as artificial electron acceptors to facilitate sustained O_2 evolution. Measurements were made at 18°C under illumination from an IR-filtered 300-W quartz halogen bulb (Kondo) with intensities controlled by neutral density filters (Lee Colortran).

2D-PAGE

To prepare samples for 2D-PAGE, cells from 1 L of culture medium were broken in a minibead beater (BioSpec Products) with 500 µl 4% aqueous 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), and the proteins were precipitated overnight at -20°C by addition of 3 volumes of acetone. The precipitate was washed 6 to 8 times with water-acetone 1:3, dried under vacuum, and redissolved in the strip rehydration buffer (7 M urea,

2 M thiourea, 4% CHAPS, 0.02 M dithiothreitol, 0.5% pH 4–7 immobilized pH gradient [IPG] buffer; Amersham-Pharmacia Biotech) overnight at room temperature. The insoluble material was removed by centrifugation at 13000*g* for 30 minutes. Protein concentrations were measured using the Bio-Rad protein assay and BSA as standard. Typically, approximately 600 µg protein in 250 µl rehydration buffer was applied to an isoelectric focusing (IEF) strip (Amersham-Pharmacia Biotech) and left overnight at 20°C to rehydrate. The first dimension was run at 500 V for 1 hour, 1000 V for 1 hour and 8000 V for 6 hours (maximum current, 50 µA). After IEF the strip was equilibrated for 15 minutes in 10 ml SDS equilibration buffer (50 mM Tris-HCl, pH 8.8, 6 M urea, 30% glycerol, 2% SDS, bromophenol blue), 15 minutes with 10 mg ml^{-1} dithiothreitol, followed by a 15-minute equilibration in 10 ml SDS equilibration buffer with 250 mg iodoacetamide. The strips were then placed on top of a standard 10% polyacrylamide gel and electrophoresed at constant voltage. Following electrophoresis the gel was stained with Coomassie blue for 1 hour at 55°C, and destained overnight at room temperature.

Electron Microscopy

The isolated plastids were fixed separately with 2% glutaraldehyde and 2% OsO_4 in isolation buffer for 1 hour each. The fixation medium was removed by washing with saline buffered with 0.1 M phosphate (100 mM Na_2HPO_4/KH_2PO_4, pH 7.2, 150 mM NaCl) in a microfuge. The plastid pellets were dehydrated in a graded ethanol series, and embedded in either LR white or Epon. Ultrathin sections were viewed in a JEOL JEM-100S transmission electron microscope at 80 kV. All EM supplies were obtained from Ted Pella (Redding, Calif).

Results

The plastids in *Lingulodinium* are highly reticulate in electron micrographs of whole cells, as plastid profiles often branch out and merge with one another (Schmitter 1971) (Fig.IV.1, A). It is thus possible that what are often observed as small discrete plastids in fact represent sections cutting through an end of a more reticulate structure. This makes purification of the organelles problematic, as considerable care must be used to reduce breakage of the organelles when the cell wall is disrupted. The standard protocols developed for purification of plastids from higher plants (protocol 1) (*Schuler and Zielinski, 1989*) produced preparations containing discrete pigmented bodies when examined using the light microscope (data not shown). When examined by transmission electron microscopy (Fig. IV.1, B), these preparations contained plastids surrounded by multiple membranes and containing dark and folded thylakoid membranes inside a light gray stroma (Fig. IV.1, C). The thylakoids must fold if long cylinders become spheres when released from the cell, although we note the spherical plastids in the purified samples appear smaller than the plastids observed in the intact cells. Thus it appears likely that many of the plastids have broken and resealed during cell lysis or purification.

Electron microscopy clearly shows that thylakoid membranes are purified by this procedure. In agreement with this, we found that roughly half the total cellular pigment (estimated by chlorophyll absorbance at 680 nm) was retained in these preparations (data not shown). However, when these plastid preparations were tested with antibodies against our marker proteins PCP and RuBisCO, the plastid preparations (Fig. IV.1, D, lane 4) contained less PCP than the supernatant that did not enter the Percoll gradient (lane 3). More importantly, no RuBisCO was detected in the purified plastid

preparations (Fig. IV.1, D, lane 4). These observations are consistent with a breakage of the plastid membranes and complete loss of soluble stromal contents. They also suggest partial breakage of the thylakoid membranes and partial loss of the lumenal contents. These observations thus underscore the importance of assessing organelle integrity during purification.

The protein profile of the purified plastids after 2D-PAGE is much simpler than the profile in whole cells (Fig. IV.1, E, F). There is no indication of cytoplasmic contamination as judged by the presence of luciferin binding protein (LBP; box at 72 kDa). However, these analyses also confirm the absence of RuBisCO (box at 55 kDa) noted by the immunologic assays. In terms of the complexity of the protein pattern observed, this analysis superficially resembles that of pea thylakoid lumenal proteins (*Peltier et al., 2000*). Taken together, our results suggest that this protocol purifies thylakoids away from both cytoplasmic and chloroplast stroma proteins, and that some of the thylakoid lumen proteins have been lost.

Several previous studies have shown that plastid ultrastructure varies over the course of a day (Herman and Sweeney 1975; Rensing, Taylor et al. 1980). More importantly, these morphologic changes correspond to movements of RuBisCO, which becomes sequestered in small regions of the chloroplast, called pyrenoids, during the early day when photosynthesis rates are high (Nassoury, Fritz et al. 2001). Pyrenoids are readily identifiable in TEM images of whole cells by their more widely spaced thylakoid stacks (Fig. IV.1, A), but have never been observed in any of the purified organelles, consistent with the lack of RuBisCO in the preparations. We thought it possible that the loss of RuBisCO reflected a particularly sensitive time during the daily cycle of plastid structure and function. To test this, we also prepared plastids from early night-phase cells, which do not contain pyrenoids and show RuBisCO distributed over the entire plastid (*Nassoury et al., 2001*). We found no difference in marker protein retention of plastids prepared from cells taken at several different times throughout the Light-dark cycle (data not

shown). These results indicate that RuBisCO retention is unaffected by the chloroplast ultrastructure at the time of extraction.

The "stirred plastid" method (protocol 2) produced organelles that did not appear spherical. Instead, these preparations contained irregularly shaped organelles as well as considerable quantities of broken plastids and isolated thylakoid membranes (Fig. IV.2, A). These plastids also contained roughly half the cellular pigments as measured by absorbance at 680 nm (data not shown). The plastid preparations contained roughly half the total PCP, as judged by the similar signals found for the purified plastids (Fig. IV.2, B, lane 3) and the fraction that did not enter the Percoll (lane 2). Unlike the standard method, however, these preparations also retained a small and variable fraction of the stromal marker RuBisCO (Fig. IV.2, B). The protein profile after 2D-PAGE can be seen to contain a greater number of proteins than observed using the standard protocol (compare Fig. IV.2, C, with Fig. IV.1, F), suggesting that lumenal integrity is more highly preserved. These 2-dimensional gels also confirm the presence of small amounts of RuBisCO and the absence of the cytoplasmic marker LBP. We conclude from these data that the stirred plastid protocol produced relatively clean preparations enriched primarily in both lumenal and membrane-bound thylakoid proteins.

To test if these plastids retained function after isolation, chlorophyll-specific O_2 evolution rates were measured under a range of light intensities and compared with O_2 evolution rates in intact cells. The artificial electron acceptors ferricyanide and methyl viologen were added to replace the electron sinks normally maintained by stromal proteins in whole cells. The facilitated O_2 evolution rates from the plastid preparations were nearly as high (>80%) as the rates measured from whole cells (Fig. IV.2, D), showing that the photosynthetic complexes were indeed functionally active in the purified fractions. In control experiments we observed no major difference in addition of the artificial electron acceptors to whole cells (Table IV.1). However, plastids were photosynthetically competent only in the presence of artificial

electron acceptors (Table IV.1), consistent with our observation that the stromal proteins mediating CO_2-dependent electron transport were absent from the prepared plastids.

To attempt to increase the yield of intact plastids, we also tested the addition of cellulase to degrade the cell wall. However, the optimal temperature for the cellulase is 40°C, and this heat shock produced effects on the ultrastructure of the intact cell that we interpret as consistent with denaturated protein (Fig. IV.3, A); indeed, the cells did not remain viable after a heat shock. Plastids purified following a heat shock (protocol 3) appeared to be the least pure and least intact. As viewed by electron microscopy, the plastids were primarily spherical but only approximately 50% were intact, and the preparations also contained considerable non-membrane-bound material (Fig. IV.3, B). Thus the non-membrane-bound material observed in the purified plastids might represent denatured protein released from the organelles as they rupture. Interestingly, immunologic analyses showed that the purified organelles retained the majority of the stromal marker RuBisCO (Fig. IV.3, C, lane 4), although considerable PCP had still been lost. Quantitative analyses of these immunoblots clearly showed this selective loss of PCP (Table IV.2), which we interpret to reflect a greater stability and increased solubility of PCP at the increased temperature of the heat shock. As in all previous protocols, there were no differences between day and night-phase plastids, showing that plastid ultrastructure had no effect on the extraction protocol. Finally, the presence of RuBisCO was confirmed on 2-dimensional gels (Fig. IV.3, D; box at 55 kDa). The protein profiles obtained by this protocol were the most complex of all the methods tested, and yet remained free from detectable levels of the cytoplasmic marker LBP (box at 72 kDa). These data suggest that the heat shock protocol is most likely to produce plastids containing the full gamut of stromal and thylakoid proteins. The caveat, as based on electron microscopic examination, is that nonplastid impurities might still be present. These preparations appeared to contain high levels of denatured protein, precluding analysis of the level of contaminants

using marker enzyme activity. At present we lack the immunologic tools to adequately determine the level of contaminants by Western blot analysis.

Discussion

We describe here the purification of chloroplast fractions from the dinoflagellate *Lingulodinium polyedrum*. The plastids from this organism appear to be more sensitive to rupture than those from higher plants, as normal protocols for the purification of plant plastids produce dinoflagellate plastids completely lacking the stromal marker enzyme RuBisCO (Fig. IV.1, D). In part, the fragility of the organelles might be related to problems of osmotic strength, as we observed that higher than usual concentrations of sucrose or other osmolytes were beneficial (compare protocols 1 and 2, for example). However, a more important issue may be the highly reticulate nature of the chloroplasts observed in electron micrographs (Fig. IV.1, A). Clearly purification of these structures is expected be difficult, and it remains to be seen if other dinoflagellates will prove to be more tractable. Our preliminary experiments, which have tested the stirred plastid protocol for purification of plastids from the dinoflagellate *Amphidinium carterae*, suggest that plastid fragility may be common in the dinoflagellates, as these plastids showed no noticeable improvement in RuBisCO retention over what was observed with the *Lingulodinium* plastids.

The complete lack of RuBisCO in the plastids purified by traditional methods was surprising, and underscores the importance of methods that can be used to monitor organelle integrity. For example, plastids prepared using protocol 1 evolve oxygen (data not shown), leaving the erroneous impression that the organelles have been isolated intact. It is only when the retention of the stromal enzyme RuBisCO and the thylakoid protein PCP are directly monitored that the true situation is revealed. The method we have developed to assess marker protein retention simply compares the amount of immunoreactive protein in a sample representing 0.2% of each fraction. This differs from traditional biochemical measurements of organelle purity, in which specific activity (in terms of enzyme activity per unit protein) would be followed in successive stages of purification until no further increases could be observed. There are two advantages to our method in the present

instance. First, the method does not require the proteins to be functional, and can be used to follow specific marker proteins through the purification process even though the sample has been heavily denatured (for example, as in protocol 3). Second, our method does not require general protein measurements, thus allowing inclusion of proteins such as BSA to the extraction buffers as a stabilizing agent.

Given the novelty of our assessment procedure, it was thought important to confirm our observations by determining the overall protein pattern by 2D-PAGE. We observed that the 2D-PAGE protein pattern was simplified in all plastid preparation protocols compared with the total cell extract. This, and in particular the lack of the cytoplasmic marker protein LBP, suggests that the plastids may have been purified away from other organelles and the bulk of the cytoplasmic components. However, the more important issue revealed by our studies is that of organelle integrity. The thylakoid membranes were clearly present in our preparations, as they were readily visualized by electron microscopy and pigments were retained in the purified plastid fractions. However, our more sensitive immunologic methods clearly demonstrated that the plastids had broken and resealed during the purification.

In general, we found that a higher osmotic strength medium was beneficial, as suggested by comparison of protocols 1 and 2. Furthermore, the addition of both KC1 and BSA appeared helpful, as also found during purification of plastids from higher plants(Jones 1995). Thus these changes should be considered first if plastid purification protocols are being developed using other dinoflagellate species. The 3 different purification protocols described here reflect the full range of RuBisCO retention inside the purified plastids that we have observed. We believe that none of the procedures delivered intact organelles, presumably as a consequence of the plastid structure. However, the heat shock treatment that appears to have denatured most of the cellular proteins seems most promising for proteomic analyses. These preparations were the only ones in which plastid stromal contents co-

purified with the plastid pigments. It is also possible that plastid DNA can be isolated from these preparations. However, it must be noted that the heat shock treatment precludes functional analyses of the plastids in vitro.

For researchers interested primarily in thylakoid proteins, such as the proteins encoded by plastid minicircles, our recommendation would be to employ the stirred plastid protocol described here. Examination of the electron micrographs and the protein patterns observed after 2D-PAGE suggests that this protocol should provide the most highly purified thylakoid preparations.

In summary, we present here the range of results that were obtained with respect to the purification of plastids from the dinoflagellate *Lingulodinium*. We found that traditional protocols yielded plastids from *Lingulodinium* that typically contained intact thylakoid stacks as judged from their morphology under the electron microscope. However, while retaining significant amounts of the thylakoid marker PCP, these preparations had lost all RuBisCO. The 2-dimensional electrophoretic patterns were roughly comparable to those observed with pea thylakoids. We conclude that this protocol purified essentially thylakoid membranes and that the plastids had broken and resealed during cell lysis or purification. This suggests that *Lingulodinium* plastids from this organism are remarkable for their fragility, in particular with regard to the level of RuBisCO retained in the purified fractions.

Nonetheless, we found several techniques that improved the purification when compared with the standard methods usually employed with higher plant plastid preparations. One modified version of the basic method was found to produce plastids that do retain small amounts of RuBisCO and appear reasonably pure as judged by the 2D-PAGE profile and electron microscopy. The most important improvements include an increase in the osmotic strength of the extraction buffer, and the addition of various stabilizing compounds to the extraction buffers. Finally, we report a protocol that retained almost all the RuBisCO, but electron micrographs of the purified fractions suggest it may not be as highly purified. In this case the application

of a heat shock treatment to the algae in conjunction with a cellulase treatment was found to be the most important feature. These preparations, if used with caution, appear suitable for whole organelle proteomic studies and potentially genomic studies as well.

Acknowledgements

We thank N. Nassoury for the electron micrograph of plastids in the intact cell and Dr. B. Green (University of British Columbia) for critical comments and helpful suggestions. We are also grateful to all members of the Morse and Cappadocia laboratories for their contributions during development of the different plastid purification protocols. T.M. is the recipient of a postdoctoral fellowship (National Science and Engineering Research Council of Canada, NSERC). This work has been supported by a grant from NSE.

Figure Legends.

Fig. IV.1 Analysis of plastids prepared by protocol 1. **A**: Transmission electron micrograph of *Lingulodinium*. One large highly reticulate and several small individual plastids (P) can be seen. The pyrenoid (Py), characterized by more widely spaced thylakoid stacks, is near the center of the cell away from the cell wall (CW). Arrowheads indicate plastid DNA with a fibrous structure. **B,C**: Representative TEM of purified plastids. Pyrenoids are not present in any of the purified organelles, consistent with a lack of RuBisCO. **D**: Representative Western blot using 0.2% of the sample from lane 1, total protein extract; lane 2, crude plastid preparation; lane 3, sample above the Percoll; and lane 4, plastid band in Percoll using a mixture of anti-RuBisCO and anti-PCP. Total protein load cannot be measured owing to the presence of BSA in the buffers. **E:** Representative 2-dimensional electrophoresis of total *Lingulodinium* extract (0.6 mg). F. Representative 2D-PAGE of purified plastid protein (~0.6 mg). Molecular mass markers (kDa) are shown at left, and pH extremes (linear gradient) are at the top of each gel. The boxed areas in each 2D-PAGE gel indicate the position of LBP (molecular weight, 72 kDa) and RuBisCO (molecular weight, 55 kDa). Scale bars: 1 μm.

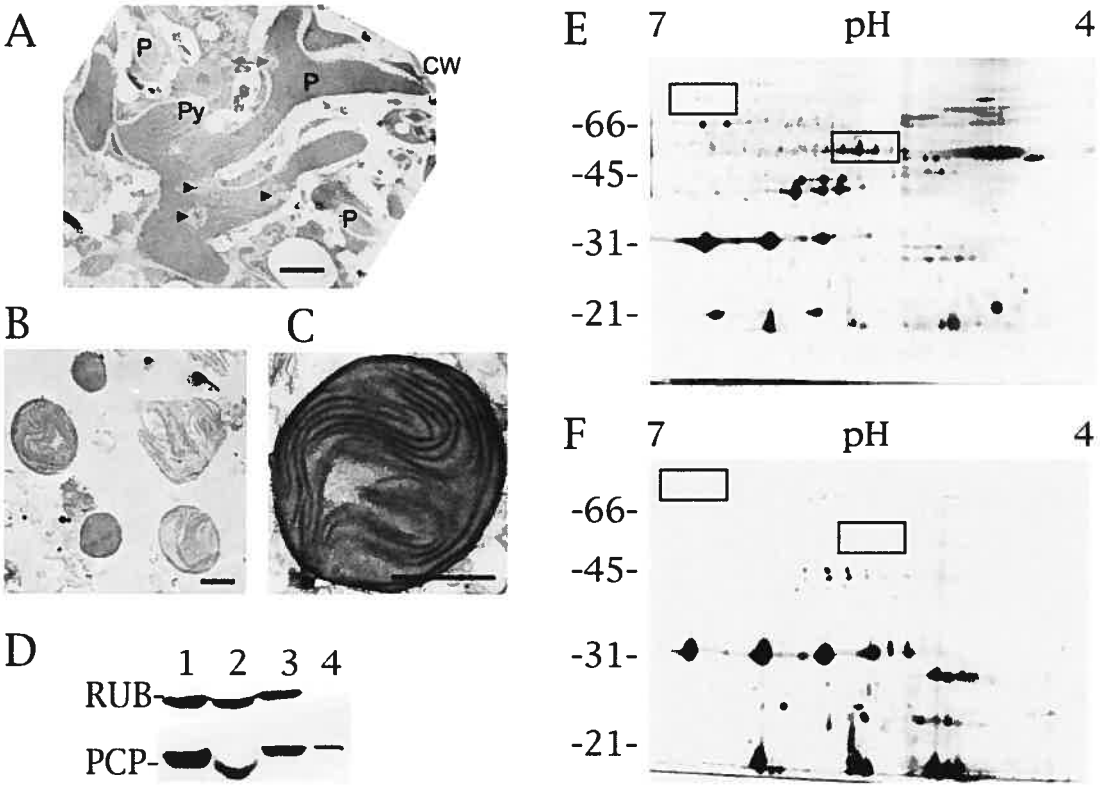

A
P
P
Py
CW
P
B
C
D
1 2 3 4
RUB-
PCP-
E
7
pH
4
-66-
-45-
-31-
-21-
F
7
pH
4
-66-
-45-
-31-
-21-

Fig. IV.2 Analysis of plastids prepared by protocol 2. **A**: TEM of purified plastids. Scale bars: 1 µm. **B**: Western blot using 0.2% of the sample from lane 1, crude plastid preparation; lane 2, sample above the Percoll; and lane 3, plastid band in Percoll using a mixture of anti-RuBisCO and anti-PCP. **C**: 2D-PAGE of purified plastid protein (~0.6 mg). Molecular mass markers (kDa) are shown at left, and pH extremes (linear gradient) at the top of the gel, and the positions of LBP and RuBisCO are boxed. The group of 6–7 protein isoforms at 66 kDa is BSA from the extraction buffer and not a true plastid component. **D**: Net photosynthetic O_2 evolution rates from *Lingulodinium* whole cells (closed squares) and purified plastids (open circles) measured over a range of light intensities.

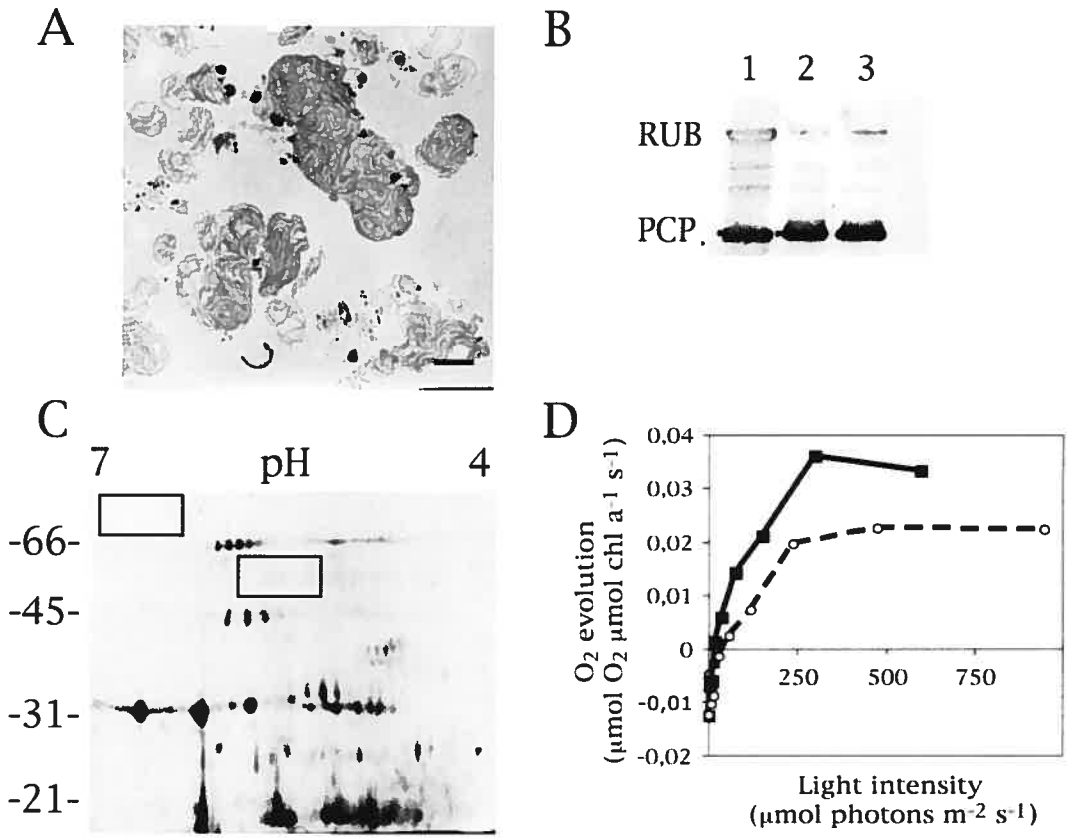

A
B
1 2 3
RUB
PCP
C
7 pH 4
-66-
-45-
-31-
-21-
D
0,04
0,03
0,02
0,01
0
-0,01
-0,02
250
500
750
O2 evolution
(µmol O2 µmol chl a-1 s-1)
Light intensity
(µmol photons m-2 s-1)

Fig. IV.3 Analysis of plastids prepared using protocol 3. **A:** TEM of *Lingulodinium* plastids in situ in heat-shocked cells. **B**: Representative TEM of purified plastids. **C**: Representative Western blot using 0.2% of the sample from lane 1, total protein extract; lane 2, crude plastid preparation; lane 3, sample above the Percoll; and lane 4, plastid band in Percoll using a mixture of anti-RuBisCO and anti-PCP. **D**: Representative 2D-PAGE of purified plastid protein (~0.6 mg). Molecular mass markers (kDa) are shown at left, and pH extremes (linear gradient) at the top of the gel. The positions of LBP and RuBisCO are boxed. The group of 6–7 protein isoforms at 66 kDa is BSA from the extraction buffer.

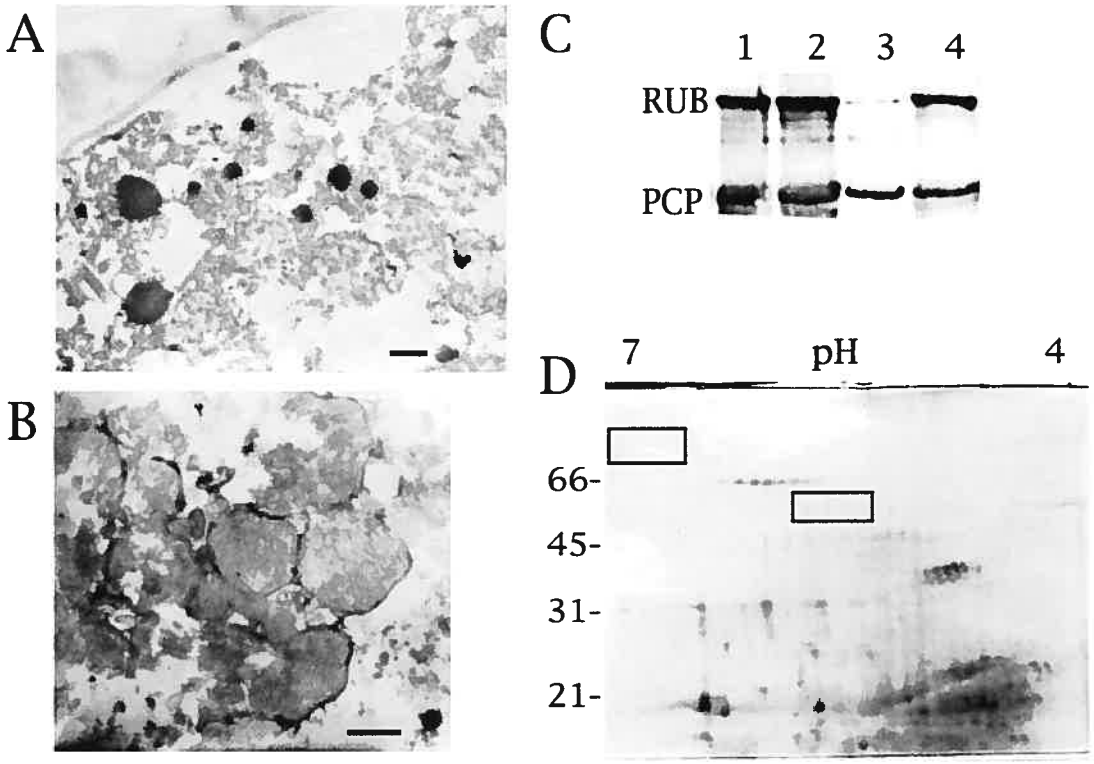

A
B
C
1
2
3
4
RUB
PCP
D
7
pH
4
66-
45-
31-
21-

Table IV.1 Maximum Gross Chlorophyll-Specific O_2 Evolution Rates Relative to Whole Cells with Added CO_2[a]

Fraction	Added CO_2 (%)	Added MV + $FeK_3(CN)_6$ (%)
Whole cells	100	89
Plastids	0.5	87

[a]CO_2 was supplied as 5 mM $NaHCO_3$; MV and ferricyanide concentrations were as described in "Materials and Methods."

Table IV.2 Subcellular Fractionation of RuBisCO, PCP, and Plastid Pigments After Heat Shock[a]

	Day-phase cells (%)			Night-phase cells (%)		
Fraction	**OD_{680}**	**RuBisCO**	**PCP**	**OD_{680}**	**RuBisCO**	**PCP**
Crude	100	100	100	100	100	100
Clarified	80	90	96	77	102	83
Supernatant	1	4	53	1	6	42
Plastids	43	86	69	40	78	57

[a]Aliquots of each fraction (0.2% total volume) were analyzed for absorption at 680 nm, and response to anti-RuBisCO and anti-PCP on Westerns blots. Western blots were quantitated by densitometric scans and compared with the values observed in crude extracts. The data are representative of 4 different experiments.

Chapter V. **Synthesis and degradation of dinoflagellate plastid-encoded psbA proteins are light-regulated, not circadian-regulated**

Objectives.

1) **To test if dinoflagellate minicircle genes encoding plastid proteins are translated in dinoflagellate plastids.**
2) **To study the character of psbA (the plastid encoded photosystem gene) synthesis and degradation.**
3) **To demonstrate the feasibility of using a proteomic approach to characterization of the dinoflagellate plastid genome.**

Publication of chapter V: This manuscript has been published in Proc. Natl. Acad. Sci. USA (2005) 102: 2: 44-2849.

All the experimental procedures were done by me except for protein microsequencing which was performed by Lene Jensen and Peter Højrup (University of Southern Denmark).

Synthesis and degradation of dinoflagellate plastid-encoded psbA proteins are light-regulated, not circadian-regulated

Yunling Wang [*], Lene Jensen [†], Peter Højrup [†] and David Morse [*,‡]

[*]Department of Biological Science, University of Montreal, 4101 Sherbrooke est, Montreal, QC, Canada H1X 2B2; and [†]Department of Biochemistry, University of Southern Denmark, Campusvej 55, DK-5230 Odense, Denmark

[‡] To whom correspondence should be addressed. [redacted]

Abstract

In many dinoflagellate species, the plastid genome has been proposed to exist as a limited number of single-gene minicircles, and many genes normally found in the plastid genome are nuclear-encoded. Unlike the nuclear-encoded plastid-directed gene products whose expression is often regulated by the circadian clock, little is known about expression of minicircle genes. Furthermore, even the plastid location of the minicircles has recently been challenged. We have examined the incorporation in vivo of [^{35}S]methionine into the proteins of purified plastids, and we find that several plastid proteins are labeled in the presence of cycloheximide but not chloramphenicol. One of these proteins, labeled in two different dinoflagellate species, was identified as psbA by Western blot analysis. Furthermore, this psbA has the expected physiological characteristics, because both synthesis and degradation are induced by light. We find no evidence for circadian control over either synthesis or degradation of psbA, unlike the several nuclear-encoded plastid-directed proteins studied. Finally, we find that levels of psbA protein or RNA do not change over a 24-h light-dark cycle, suggesting that this protein may not be involved in mediating the circadian rhythm in oxygen evolution rates. This demonstration is the first, to our knowledge, that minicircle genes encoding plastid proteins are translated in dinoflagellate plastids, and it suggests that a proteomic approach to characterizing the dinoflagellate plastid genome is feasible.

gene expression | plastid genome | *Gonyaulax* | photosynthesis

The majority of photosynthetic dinoflagellates harbor plastids that are surrounded by three membranes (Gibbs 1981), rather than the more typically found two or four membranes, suggesting that these plastids are derived from a secondary endosymbiotic event followed by loss of a single membrane. However, the identity of the endosymbiont that gave rise to the plastids is not firmly established and, based on the identity of nuclear-encoded plastid-directed proteins, may have had some unusual characteristics. For example, neither the unusual light-harvesting protein peridinin-chlorophyll a protein (PCP) (Hofmann, Wrench et al. 1996) nor the allenic oxicarotenoid peridinin to which it binds is found in other organisms (Prezelin and Haxo 1976). Furthermore, the peridinin-type plastids contain a form II ribulose-1,5-bisphosphate carboxylase/oxygenase, previously observed only in some species of anaerobic proteobacteria (Morse, Salois et al. 1995; Whitney and Yellowlees 1995).

Clearly, the true dinoflagellate plastid phylogeny is more likely to be revealed by using analysis of plastid-encoded genes. However, many proteins usually encoded by the plastid genome are found as nuclear-encoded genes in dinoflagellates (Bachvaroff, Concepcion et al. 2004; Hackett, Anderson et al. 2004; Koumandou, Nisbet et al. 2004), suggesting that the organelle genome is highly reduced. The best candidates for the plastid-encoded genes include the 16S and 23S rRNA genes, eight photosystem components (psaA-B, psbA-E, and psbI), two ATP synthase subunits (atpA-B), two cytochrome b_6f-subunits (petB and petD), two unidentified ORFs (ycf16 and ycf24), and two ribosomal proteins (rpl28 and rpl23). These genes form the most reduced set of plastid genes known (with the exception of the Apicomplexan plastids). Furthermore, in a feature unique among plastid genomes, they are all encoded by small single-gene minicircles (Zhang, Green et al. 1999; Koumandou, Nisbet et al. 2004). One

question of major biological interest is how and why dinoflagellates have evolved and maintained this reduced set of plastid genes.

If the minicircles are to be considered as a novel form of a true plastid genome, a second and more fundamental question is whether the minicircle genes are indeed expressed in the plastid. To date, although transcripts derived from minicircles have been observed, there is no direct experimental evidence to demonstrate their translation. Furthermore, whereas psbA transcripts have been observed in the plastids of *Symbiodinium* by in situ hybridization (Takashita, Ishikura et al. 2003), recent cell fractionation studies using the dinoflagellate *Ceratium* showed that minicircle DNAs were located in the nucleus (Laatsch, Zauner et al. 2004). Because none of the minicircle genes sequenced to date encodes the characteristic leader sequence believed to be required for targeting to the triple membrane-bound plastids (Nassoury, Cappadocia et al. 2003), a nuclear location for the minicircles would preclude them from encoding functional plastid proteins. An elegant and conclusive resolution to issues of both expression and location of the minicircles can be obtained by determining the antibiotic sensitivity of protein synthesis derived from the minicircle genes.

We show here that psbA is one of several plastid proteins synthesized in the presence of cycloheximide by two different dinoflagellate species, *Amphidinium* and *Gonyaulax*. We also find that, in *Gonyaulax*, the regulation of psbA synthesis and degradation is light-regulated, as has been observed in higher plants. Our results provide compelling evidence for placing the psbA gene in the plastids of dinoflagellates and for ruling out circadian control of psbA levels. Dinoflagellate plastid-encoded proteins may be unlike the nuclear-encoded plastid-directed gene products, whose synthesis is often clock-controlled (Markovic, Roenneberg et al. 1996; Fagan, Morse et al. 1999; Le, Jovine et al. 2001).

Materials and methods

Cell Culture. *Amphidinium carterae* [Culture Center for Marine Phytoplankton (CCMP) no. 1314] and *Gonyaulax* polyedra (now Lingulodinium polyedrum; CCMP no. 1936) were obtained from the Provasoli–Guillard Culture Center for Marine Phytoplankton (Boothbay Harbor, ME) and grown in a modified seawater medium (f/2) (Guillard and Ryther 1962) at a constant temperature (18 ± 1°C) in 12-h light/12-h dark cycles by using cool white fluorescent light at an intensity of 50 μmol of photons·m^{-2}·s^{-1}. The beginning of the light period in the light-dark (LD) cycle is defined as time 0 (LD 0), and the beginning of the dark period is defined as LD 12. Cultures were grown to a cell density of 12–14,000 cells per ml (*Gonyaulax*) or 500,000 cells per ml (*Amphidinium*) as measured by counting using a hemocytometer. Both cultures were unialgal, but only *Amphidinium* is axenic.

Plastid Isolation Protocols. Cells from 4 liters of culture medium were harvested by filtration, washed extensively with seawater to remove bacterial contaminants, resuspended in 10 ml of isolation buffer (0.626 M sorbitol/0.1 M KCl/0.1 M Hepes-NaOH, pH 7.2/20 mM 2-mercaptoethanol/0.1% BSA) containing 0.05% saponin (ICN), and stirred vigorously for 15 min at room temperature. This stirred homogenate was centrifuged at 150 x g for 3 min to remove unbroken cells, and the supernatant was centrifuged at 1,500 x g for 3 min to produce a plastid-containing pellet. Plastids were purified from these crude preparations by density gradient centrifugation on a 15-ml 10–30% linear PEG- Percoll (Amersham Pharmacia Biotech, Piscataway, NJ) gradient in isolation buffer. A stock solution of PBF-Percoll [3% (wt/vol) polyethylene glycol 8000, 1% (wt/vol) BSA, and 1% (wt/vol)] Ficoll) was prepared by

dissolving the dry powders in 100% Percoll. A 15-ml 10-30% linear PBF-Percoll gradient in isolation buffer B (0.626 M sorbitol/0.1 M KCl/0.1 M Hepes-NaOH, pH 7.2/20 mM 2-mercaptoethanol/0.1% BSA) was laid on top of a 5-ml 40% PEF-Percoll in isolation buffer cushion. Five milliliters of the crude plastid preparation was layered on top of the Percoll gradients and centrifuged at 6,000 × g for 25 min at 4°C. Two distinct highly pigmented bands were found in Percoll, and the lower band was collected with a Pasteur pipette. These plastid preparations, examined microscopically, typically contained >90% aggregated and pigmented plastids. Percoll was removed from the plastid preparation by centrifugation at 1,500 × g for 15 min at 4°C after addition of three volumes of cold washing buffer (isolation buffer without BSA). This process was repeated three times, each time with fresh washing buffer and gentle resuspension of the plastid pellet by using a Pasteur pipette.

Electrophoretic Analysis. Samples for 2D electrophoresis were prepared from cells broken in a bead beater with 4% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) by overnight precipitation at –20°C after addition of three volumes of acetone. The precipitate was washed 6-8 times with water: acetone 1:3, dried under vacuum, and redissolved in the strip rehydration buffer [7 M urea/2 M thiourea/4% CHAPS/0.02 M DTT/0.5%, pH 4-7/immobilized pH gradient buffer (Amersham Pharmacia-Pharmacia Biotech)] overnight at room temperature. The insoluble material was removed by centrifugation at 14,000 rpm for 30 min, and protein concentrations were measured using the Bio-Rad protein assay. Typically, 600 μg of protein redissolved in 250 μl of rehydration buffer [7 M urea/2 M thiourea/4% CHAPS/0.02 M DTT/0.5%, pH 4–7/immobilized pH gradient buffer (Amersham Pharmacia Biotech)] was applied to an isoelectric focusing (IEF) strip (Amersham Pharmacia Biotech) and electrophoresed at 500 V for 1 h, at 1,000 V for 1 h, and at 8,000 V for 6

h. After IEF, the strip was equilibrated for 15 min in 10 ml of SDS equilibration buffer (50 mM Tris·HCl, pH 8.8/6 M urea/30% glycerol/2% SDS/10 mg/ml DTT/Bromophenol blue) and for 15 min in 10 ml of SDS equilibration buffer containing 250 mg of iodoacetamide. The strips were then placed on top of a 10% polyacrylamide gel and electrophoresed at constant voltage. Following electrophoresis, the gel was stained with Coomassie blue for 1 h at 55°C, and destained overnight at room temperature.

The 1D lithium dodecyl sulfate (LDS) gels were run after substitution of SDS with LDS in all buffers. Plastid pellets were dissolved in two volumes of LDS sample buffer (100 mM Tris·HCl, pH 6.8/2% LDS/50 mM DTT/10% glycerol) at room temperature for 1 h, and 10 µl was electrophoresed on 10% polyacrylamide gels before electrophoretic transfer to nitrocellulose.

For Western analyses, all gels were transferred electrophoretically to nitrocellulose membranes. The protein blots were blocked with 5% powdered milk in Tris-buffered saline containing 0.05% Tween 20 before incubation with either a 1:5,000 dilution of chicken anti-psbA (Agri-Sera) or a 1:5,000 dilution of rabbit anti-PCP (Nassoury, Fritz et al. 2001) to control for protein loading. Antibody binding was visualized by using a 1:10,000 dilution of commercial peroxidase-linked secondary antibodies and chemiluminescence (Amersham Pharmacia Biotech).

In Vivo Labeling. Exponentially growing cells were treated with 100 µM chloramphenicol or 100 µM cycloheximide for 1–2 h before the start of labeling to ensure inhibition of protein synthesis on 70S or 80S ribosomes, respectively. At a time corresponding to the start of the light period, 1 liter of cells was concentrated to ≈10 ml by filtration by using a 20-µm nylon filter and incubated with 2 mCi (1 Ci = 37 GBq) of [^{35}S]methionine (>1,000 Ci/mmol; ICN) for 40 min. After labeling, the cell suspension was repeatedly washed to remove unincorporated radiolabel, and the plastids were purified as above. Plastid proteins were then extracted and analyzed by electrophoresis and

autoradiography. For labeling by using high light, cells were exposed to a sodium arc lamp light (≈2,500 µmol of photons·m^{-2}·s^{-1}) passed through 1 cm of cold flowing water to filter out the infrared wavelengths. Cell samples did not increase in temperature during the labeling period.

Protein Microsequence. Sequencing of tryptic fragments prepared from LDS/PAGE gel-purified proteins (Fig. V.6) were performed essentially as described (Schevchenko, Wilm et al. 1996; Gevaert, Demol et al. 2001). Gel bands were cut out from LDS/PAGE gels and the protein was digested in-gel as described (Schevchenko, Wilm et al. 1996), except that 25 ng/µl trypsin (Wako) was used for digestion. Before MS, the samples were purified by using microcolumns packed with Poros R2 beads (Applied Biosystems) as described (Goboom, Nordhoff et al. 1999). For MS peptide mapping, samples were eluted directly onto the mass spectrometric steel targets by a solution of alpha-cyano-4-hydroxy cinnamic acid (4-HCCA), in 70% acetonitrile in 0.1% TFA (18 µ g/µ l). Before analysis, samples for MS/MS were derivatized by 4-sulfophenyl isothiocyanate (SPITC) (Aldrich) to improve fragmentation and facilitate interpretation (Plastids were purified from these crude preparations by density gradient centrifugation on a 15-ml 10–30% linear Percoll (Gevaert, Demol et al. 2001). A total of 2-10 µ l of peptide sample dissolved in 20 mM $NaHCO_3$, pH 8.6, was mixed with 5-8 µ l of SPITC (10 µ g/µ l in 20 mN $NaHCO_3$, pH 8.6) and kept at 50°C for 30 min before purification on microcolumns and elution directly onto the mass spectrometric steel target by using a 4-HCCA concentration of 7 µ g/µ l.

MS analysis was performed on an Applied Biosystems Voyager Elite operated in the positive reflector mode and MS/MS analysis was performed on a Micromass QTOF MALDI instrument. For the evaluation of mass spectra, GPMAW software (Lighthouse Data, Odense, Denmark) was used in addition to the instrument specific programs (Protein Explorer and MASSLYNX V. 4.0). Peptide sequences were either individually searched against the National Center for Biotechnology Information (NCBI) database

by using the NCBI BLAST search engine (www.ncbi.nlm.nih.gov), or each group of sequences were searched, using the MS BLAST search engine (dove.embl-heidelberg.de/Blast2/msblast.html)(Schevchenko, Wilm et al. 1996).

MS analysis was performed on an ABI Voyager Elite operated in the positive reflector mode, and MS/MS analysis was performed on a Micromass QTOF MALDI instrument.

Sequencing of tryptic fragments prepared from 2D gel-purified proteins was performed by the Harvard Microchemistry Facility (Cambridge, MA), using microcapillary reverse-phase HPLC ion-electrospray tandem MS on a Finnigan LCQ DECA XP Plus quadruple ion-trap mass spectrometer.

Nucleic Acid Analysis. PCR using oligonucleotides (5'-TCAGCTTTTATTTTTGCTCCG-3' and 5'-CCTGATGAAT-CAATTATGGACTG-3'), based on GenBank accession no. AB025588 [GenBank] , were designed to amplify a psbA sequence, and our sequence was identical to that reported. The sequence was analyzed by using MACVECTOR software (Accelrys). The techniques for extraction add hybridization of nucleic acids were performed as described (Sambrook et al., 1989).

Results

To identify plastid-encoded proteins in *Gonyaulax*, we first examined purified plastids for proteins that incorporated [^{35}S]methionine in vivo in the presence of cycloheximide. A first series of experiments was designed to examine plastid proteins globally by using 2D gel analysis. However, none of the proteins incorporating radiolabeled methionine in the presence of cycloheximide (circled proteins in Fig. V.1A) was present in levels appropriate for microsequencing (Fig. V.1B). This finding suggested that either the plastid proteins in these preparations did not incorporate significant levels of methionine or only nuclear-encoded proteins were extracted by using this procedure.

To address the possibility that labeling of dinoflagellate plastid proteins was inefficient in vivo, a complementary series of experiments was performed to identify proteins whose synthesis was blocked by chloramphenicol. The majority of plastid proteins observed on the gels did incorporate radiolabel in the presence of chloramphenicol and act as a complement to the cycloheximide inhibition (Fig. V.1 C and D). However, the four circled proteins were not labeled during this experiment. These were isolated from preparative gels and identified by microsequencing (Table V.1). The protein at 45 kDa was identified as BSA and presumably corresponds to a proteolytic fragment of the protein added to stabilize the plastids during the purification, because all of the sequences obtained were derived from the C-terminal two-thirds of the protein. The other three proteins were identified as oxygen-evolving enhancer protein subunit 1 (OEE1) (p33, similar to OEE1 of the dinoflagellate *Heterocapsa*), PCP p31, and light-harvesting polyprotein (LHP) precursor (p24, similar to LHP from the dinoflagellate *Amphidinium*). The molecular masses of the proteins on SDS/PAGE agree with those predicted

from the sequence except for LHP, which is larger than the 19 kDa predicted from the *Amphidinium* sequence.

PCP and OEE1 have been previously identified in a screen for proteins whose synthesis is under circadian control in *Gonyaulax* (Markovic et al., 1996; Le et al., 1997)(Markovic, Roenneberg et al. 1996; Le, Markovic et al. 1997; Lang, Gray et al. 1999). It thus seemed possible that little radiolabel was incorporated into these proteins because the time of in vivo labeling was inappropriate, rather than because their synthesis was inhibited by chloramphenicol. In agreement with this hypothesis, when the in vivo labeling protocol was repeated at midday, PCP, OEE1, and light-harvesting chlorophyll-binding protein (LHCP) were observed to incorporate radiolabel (data not shown and Fig. V.3). These proteins are thus indeed nuclear-encoded.

The absence of minicircle-encoded proteins in this global screen was surprising. However, because many of the minicircle genes encoding plastid proteins should produce hydrophobic thylakoid membrane components, it was possible their extraction in the buffers used for isoelectric focusing was inefficient. We therefore tested a protocol employing room temperature LDS extractions that can be used for electrophoresis of hydrophobic proteins on 1D gels. Under these conditions, several proteins in the purified plastids were observed to incorporate radiolabel in the presence of cycloheximide (Fig. V.2A). In particular, a band at 34 kDa is particularly notable (arrow) for the amount of label incorporated. Of all of the thylakoid proteins, the psbA is expected to have the highest turnover rate (Mattoo, Marder et al. 1989), suggesting that the highly labeled protein at 34 kDa might be psbA. To test this hypothesis, a commercial anti-psbA was used to determine the electrophoretic mobility of psbA in our conditions. This antibody identifies psbA as a protein of 34 kDa in whole-cell extracts of two dinoflagellates and two higher plants (Fig.V.2C). An additional band at 50 kDa in the *Gonyaulax*

extract is not present in extracts from purified plastids (data not shown). Western blot analyses of the labeled plastid proteins shows that the labeled protein at 34 kDa comigrates with psbA (Fig. V.2D). Furthermore, one of the labeled proteins in whole-cell extracts from the dinoflagellate *Amphidinium* comigrates with a protein reacting with the anti-psbA (arrow in Fig. V.2E). Taken together, these results suggest that the 34-kDa protein incorporating label in the presence of cycloheximide is psbA.

To confirm our identification of the 34-kDa band as psbA, we examined the synthesis and degradation rates of the protein under different conditions. In higher plants and green algae, psbA translation is light-regulated (Danon and Mayfield 1994). We find that psbA synthesis is also light-regulated in *Gonyaulax*, because the amount of radiolabel incorporated into the protein is ≈20-fold greater during the light phase than during the dark phase, as determined by PhosphorImager quantification of the amount of radiolabel incorporated (Fig. V.3A). This behavior is light-dependent rather than circadian, because synthesis is observed during the subjective night phase of cells kept in constant dim light. Furthermore, radiolabel incorporation into psbA in night-phase cells can be induced by brief exposure to light (Fig. V.3B). This behavior is clearly different from that of nuclear-encoded plastid proteins, whose synthesis is circadian and not directly light-responsive (Fig. V.3C).

In higher plants, psbA degradation is also light-dependent (Tyystjarvi and Aro 1996). To measure degradation rates, we took advantage of the ability of chloramphenicol to block psbA synthesis. Under normal cell culture lighting (≈50 μmol of photons·m^{-2}·s^{-1}), the degradation rate is not measurable in a 2-h period, and night- and day-phase cells respond in a similar manner (Fig. V.4A). However, psbA in day-phase cells is rapidly degraded if the cells are exposed to light intensities similar to that of bright sunlight (≈2,500 μmol photons·m^{-2}·s^{-}) (Fig. V.4B). Interestingly, the psbA in cells taken from the dark phase appears much more resistant to this light-induced degradation. Once again, this behavior is not circadian, because the rate of light-induced

degradation is high in the subjective night of cells kept in constant dim light (Fig. V.4C). Finally, psbA is degraded more slowly in the absence of chloramphenicol, suggesting that synthesis rates largely counterbalance the degradation rates under these conditions (Fig. V.4D). Taken together, the physiological responses to light in terms of induced synthesis and degradation all support the identification of the labeled protein as psbA.

The balance between light-induced synthesis and degradation suggested that psbA would not show a significant variation in the level of either immunoreactive psbA. Indeed, this finding agrees well with measurements of psbA taken over the course of a daily LD cycle (Fig. V.5A). We also find no change in levels of psbA mRNA (Fig. V.5 B and C) over the course of the daily cycle. These results are also observed when samples are taken from cells kept in constant dim light (data not shown). We conclude that psbA does not represent a probable regulatory site for the ≈3-fold changes in circadian O_2 evolution rates (Nassoury, Fritz et al. 2001).

Our data unambiguously place psbA synthesis in the plastid. It was thus of interest to determine whether other plastid-encoded genes could potentially be identified by protein microsequencing of radiolabeled bands. Although these analyses are complicated by the limited resolution of proteins separated by only 1D PAGE, 4 bands from a sample of 10 excised from LDS/PAGE gels at positions corresponding to plastid-encoded proteins were found to contain peptide sequences similar to those encoded by the known minicircle genes atpA, atpB (found twice), and psbD from *Amphidinium* (Table V.2 and Table 3). Of the remaining samples, two were not identifiable, and four contained peptides from an abundant nuclear-encoded LHCP. Clearly, although difficult, proteomic analysis of plastid-encoded proteins using this procedure seems feasible.

Discussion

To our knowledge, this is the first demonstration of protein synthesis in dinoflagellate plastids. The observation that plastid proteins incorporate radiolabel in the presence of cycloheximide provides strong support for a plastid location of the *Gonyaulax* psbA gene and several others. There is still a formal possibility that mRNA synthesized in the cytoplasm is relocated to the plastid, as are tRNAs in some mitochondria (Schneider and Marechal-Drouard 2000), but this possibility seems unlikely because there is no precedent for mRNA translocation into the plastid.

Our experiments using 2D electrophoresis to identify plastid gene product synthesis were unsuccessful, and two factors may have contributed to this result. First, plastids from *Gonyaulax* are remarkably fragile, and an analysis of the purified plastid preparations indicates that they contain primarily thylakoid membranes (Y.W. and D.M., unpublished work). This finding suggests that most of the stromal proteins normally located in the plastid may have been lost during the purification. Second, the hydrophobic photosystem components found encoded as minicircle genes might have been resistant to extraction by the usual 2D gel electrophoresis buffers. Indeed, when the extractions protocols were changed to include LDS, a number of radioactive bands could be detected in purified plastid preparations from cells labeled in vivo.

We have identified a 34-kDa protein, labeled in the presence of cycloheximide but not chloramphenicol, as psbA by using several criteria. First, the labeled protein comigrated with immunoreactive psbA on LDS/PAGE, as determined by Western blot analysis. Second, the synthesis rate of the protein is light-induced, as is psbA from higher plants. Last, exposure to high light intensities increases the degradation rate of the protein

when synthesis is blocked by chloramphenicol. These results, taken together, indicate that psbA is indeed synthesized in the plastid. We observed psbA synthesis in the presence of cycloheximide in both *Gonyaulax* and *Amphidinium*, strongly supporting a general plastid location for this gene in dinoflagellates.

The degradation rates measured for the *Gonyaulax* psbA in response to high light intensities are similar to those described for the psbA of higher plants (Tyystjarvi and Aro 1996). It seems initially surprising that the light-dependent degradation of psbA is not observed in dark-phase cells but requires prior exposure to light. In part, this result is due to a slow response of the system to changes between dark and light phases. When cell extracts were examined only 1 h after lights on or lights off, intermediate degradation rates were observed (data not shown), suggesting that adaptation to either light or dark phase requires several hours of exposure to the new conditions. This result could then explain why dark-phase cells exposed to high light for only 1 h (the measurement period) did not degrade psbA efficiently as observed (Fig. V.4B). The mechanism underlying the reduced susceptibility of psbA to degradation in dark-phase cells is also unlikely to involve a redox-based degradation mechanism, because the redox state might be the same in low- and high-light intensity. Instead, it seems possible that a different mechanism may influence this process. It is known that degradation of psbA in higher plants is controlled by reversible phosphorylation (Koivuniemi, Aro et al. 1995) and that the circadian-regulated phosphorylation in higher plants can take several hours to reach completion (Booij-James, Swegle et al. 2002). We propose that a slow, light-induced psbA protein phosphorylation in *Gonyaulax* might account for our observations.

The induction of both protein synthesis and degradation by light suggests that little diurnal variation in psbA levels should be expected, which is in agreement with what was observed experimentally (Fig. V.5A). However, degradation rates increase as a function of light intensity (Fig. V.4), whereas

synthesis rates appear to be related to only to the presence or absence of light (Fig. V.3). Whereas our cells are difficult to grow for long periods under the high-intensity laboratory lights, preliminary experiments suggest that steady-state levels of psbA RNA may increase during high light (data not shown). Thus, cells under natural light intensities may balance degradation with synthesis because higher psbA RNA levels allow for a greater rate of translational initiation.

The lack of diurnal or circadian variation in psbA was surprising, however. There are numerous examples of circadian regulation of nuclear-encoded plastid-directed protein synthesis in *Gonyaulax* (Markovic, Roenneberg et al. 1996; Fagan, Morse et al. 1999; Le, Jovine et al. 2001), and the ability of nuclear-encoded gene products to regulate plastid gene expression is well established in other systems (Zerges 2002). Clearly, more extensive studies will be required to determine whether nuclear contribution to regulation of plastid-encoded message translation will be reduced in dinoflagellates. In any event, it seems unlikely that psbA will be involved in mediating clock control over the circadian rhythym of oxygen evolution.

The demonstration that several plastid proteins can be synthesized in the presence of cycloheximide is an important step that may allow a more complete characterization of the plastid proteome. It is interesting in this regard that only a limited number of proteins in crude extracts of both *Amphidinium* and *Gonyaulax* incorporate radiolabel in the presence of cycloheximide, suggesting that the dinoflagellate plastid genome may indeed encode only a small number of genes. Our initial data suggest that it will be possible to identify plastid-encoded proteins by combining our in vivo labeling protocol with mass spectrometric microsequencing analysis (Table V.2), although these analyses may be limited by the resolution of proteins on 1D LDS/PAGE.

Acknowledgements

We thank all members of the D.M. and Cappadocia laboratories for contributing helpful suggestions during development of the plastid purification protocol. We also thank Drs. J. Rivoal (Université de Montréal), B. Green (University of British Columbia, Vancouver), and U. Maier (Universitat Marburg, Marburg, Germany) for their critical analysis of our initial experiments and helpful suggestions for further work. This work has been supported by a grant from the National Science and Engineering Research Council of Canada.

Footnotes

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: PCP, peridinin-chlorophyll a protein; LD, light-dark; CT, circadian time; LDS, lithium dodecyl sulfate; OEE1, oxygen-evolving enhancer protein subunit 1; IEF, isoelectric focusing; LHCP, light-harvesting chlorophyll-binding protein.

Figure legends

Fig. V.1. The 2D electrophoresis of in vivo-labeled plastid proteins identifies only nuclear-encoded gene products. Proteins (0.6 mg) extracted from plastids purified after in vivo labeling with [^{35}S]methionine in the presence of 100 µM cycloheximide (A and B) or 100 µM chloroamphenicol (C and D) were electrophoresed on 2D gels. (A and C) Autoradiographs. (B and D) Coomassie blue-stained gels. Molecular mass markers (kDa) are shown on the left, and pH extremes (linear gradient) are shown at the top of each gel. Circles represent potential plastid-encoded proteins.

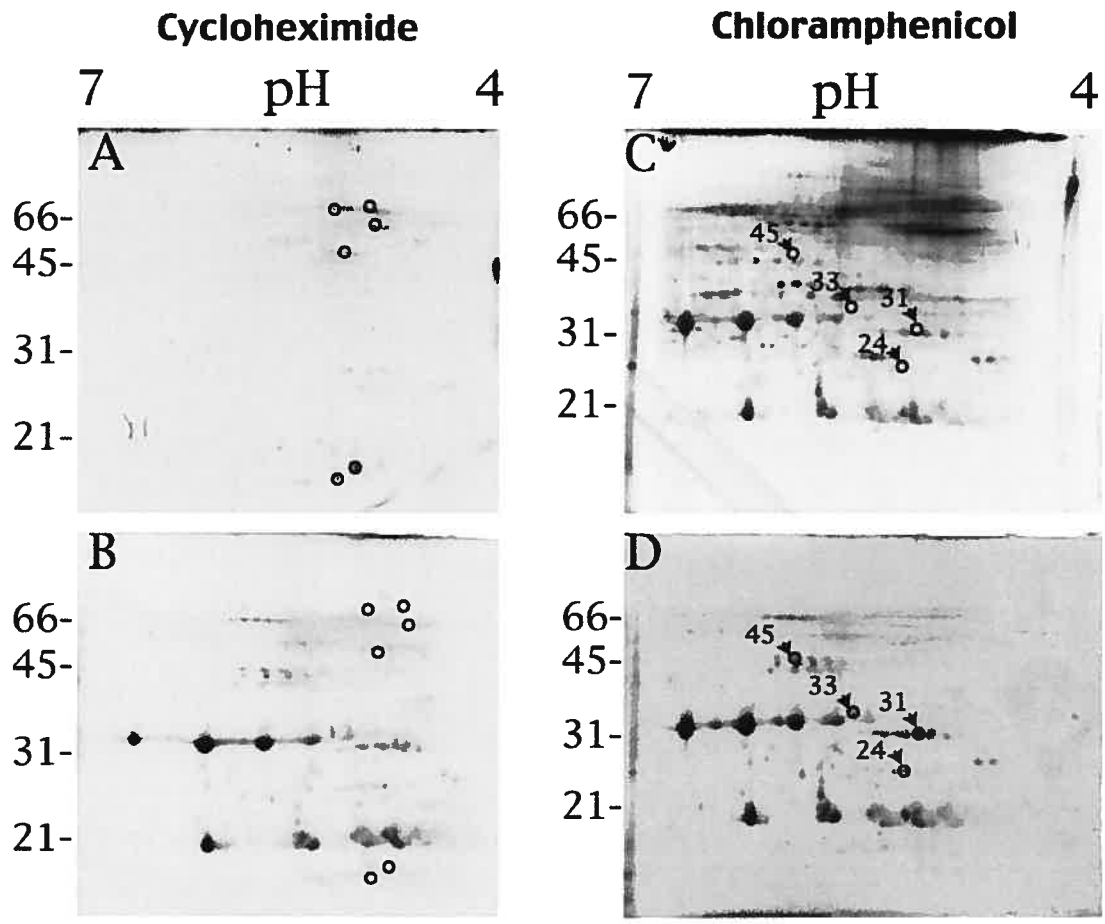

Cycloheximide
Chloramphenicol
7
pH
4
A
B
C
D
66-
45-
31-
21-
45
33
31
24

Fig. V.2. In vivo labeling of psbA is sensitive to chloramphenicol. Autoradiograph (A) and Coomassie blue staining (B) of purified plastid proteins (0.1 mg) electrophoresed on 1D LDS gels after in vivo labeling with [^{35}S]methionine in the presence of either 100 µM chloramphenicol (CHL) or 100 µM cycloheximide (CHX) are shown. Molecular mass markers (kDa) are shown on the left of B.(C) Western blot analysis with anti-psbA and chemiluminescence detection of total protein extracts (50 µg) from the dinoflagellates *G. polyedra* (Gp) and *A. carterae* (Ac), and two higher plants, *Spinacia oleracea* (spinach, Sp) and *Solanum chacoense* (wild potato, Sc). A band at 34 kDa in *Gonyaulax* whole-cell extracts corresponds to the psbA gene product. (D) Western blot analysis of the labeled proteins in A by using anti-psbA identifies the position of psbA (arrow). (E) Autoradiography (AR) and western analysis of whole-cell extract (50 µg) from A. carterae labeled in vivo with [^{35}S]methionine in the presence of 100 µM cycloheximide also shows labeled psbA (arrow).

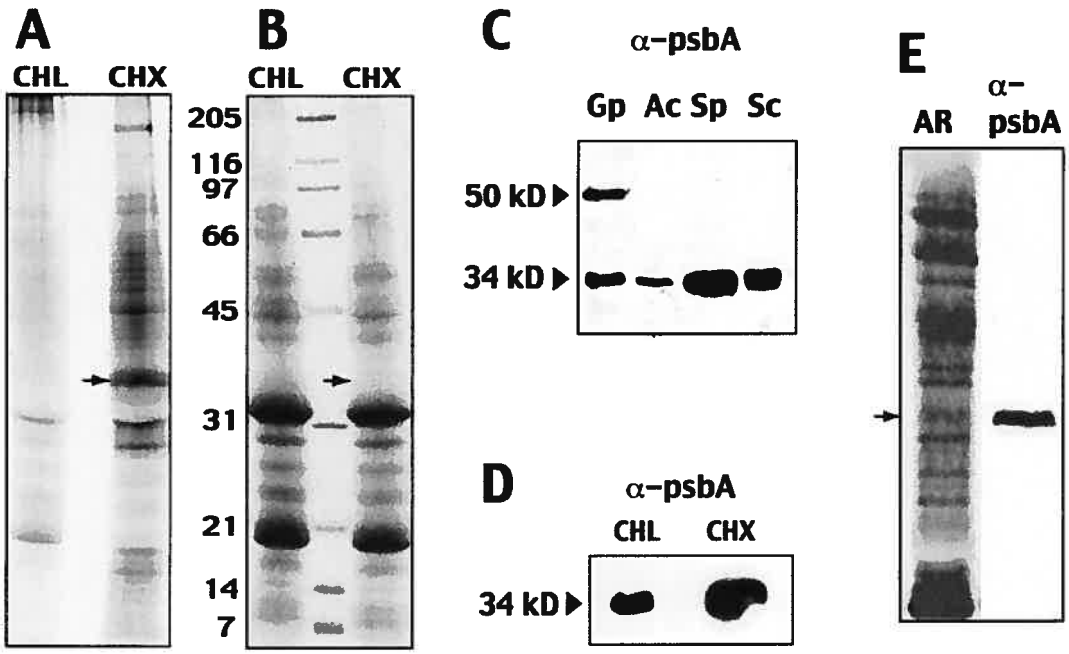

A
CHL
CHX
B
CHL
CHX
205
116
97
66
45
31
21
14
7
C
α-psbA
Gp Ac Sp Sc
50 kD
34 kD
D
α-psbA
CHL
CHX
34 kD
E
AR
α-
psbA

Fig. V.3. Light induces labeling of psbA in vivo.(A) Cells taken at various times from an LD (white and black bars) or a light-light (gray bar) light regime were labeled with [^{35}S]methionine in the presence of 100 µM cycloheximide. Total protein extracts (≈ 100 µg) were prepared by using LDS, electrophoresed on LDS gels, and transferred to nitrocellulose. (Middle and Bottom) Shown are autoradiograms and Ponceau-stained membranes, respectively. Arrowhead indicates position of psbA. (B) Cells taken at the indicated times from an LD cycle were labeled with [^{35}S]methionine in the presence of 100 µM cycloheximide in darkness (LD21, left lane), normal culture room lights (LD3, left lane; 50 µmol of photons·m^{-2}·s^{-1}), or exposed to high-light intensity (HL; right lanes ≈2,500 µmol of photons·m^{-2}·s^{-1}). (Upper and Lower) Shown are show autoradiograms and Ponceau-stained membranes, respectively. (C) Cells taken at LD 18, LD 6, and circadian time (CT) 18 were labeled with [^{35}S]methionine in the presence of 100 µM chloramphenicol, and the radiolabeled nuclear-encoded proteins were resolved by 2D electrophoresis as a control for the light induction of psbA synthesis. The spots corresponding to ribulose bisphosphase carboxylase/oxygenase (Rub; 55 kDa), glyceraldehyde-3-phosphate dehydrogenase (Gap, 45 kDa), OEE1 (33 kDa), and PCP (32 kDa) are shown.

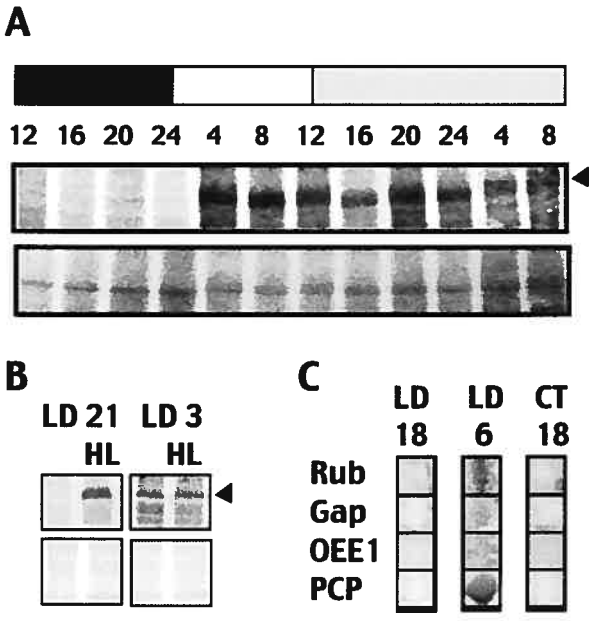

A
12 16 20 24 4 8 12 16 20 24 4 8
B
LD 21 HL
LD 3 HL
C
LD 18
LD 6
CT 18
Rub
Gap
OEE1
PCP

Fig. V.4. Degradation of psbA can be induced by high-light intensity in cells previously exposed to light. (A) Cells in either late-night phase (LD 22) or midday phase (LD 6) were treated with 100 μM chloramphenicol to block new psbA synthesis, and cell extracts were prepared at various times thereafter (time in minutes after chloramphenicol addition) by using normal lighting. Cell extracts were electrophoresed on LDS, transferred to nitrocellulose and psbA, visualized by using anti-psbA and chemiluminescence. (B) Cells in either late-night phase (LD 22) or early day phase (LD 3) were treated with 100 μM chloramphenicol and exposed to high-light intensity ($\approx$ 2,500 μmol of photons·m^{-2}·s^{-1}), and cell extracts were examined for psbA as above. Membranes were stripped and reprobed with anti-PCP as a control for protein load. (C) Cells in either late subjective night phase (CT 21) or early subjective day phase (CT 3) were treated with 100 μM chloramphenicol and exposed to high-light intensity ($\approx$2,500 μmol of photons·m^{-2}·s^{-1}), and cell extracts were examined for psbA and PCP as above. (D) Cells in early day phase (LD 3) were exposed to high-light intensity ($\approx$2,500 μmol of photons·m^{-2}·s^{-1}) in the absence of chloramphenicol treatment, and cell extracts were examined for psbA and PCP as above.

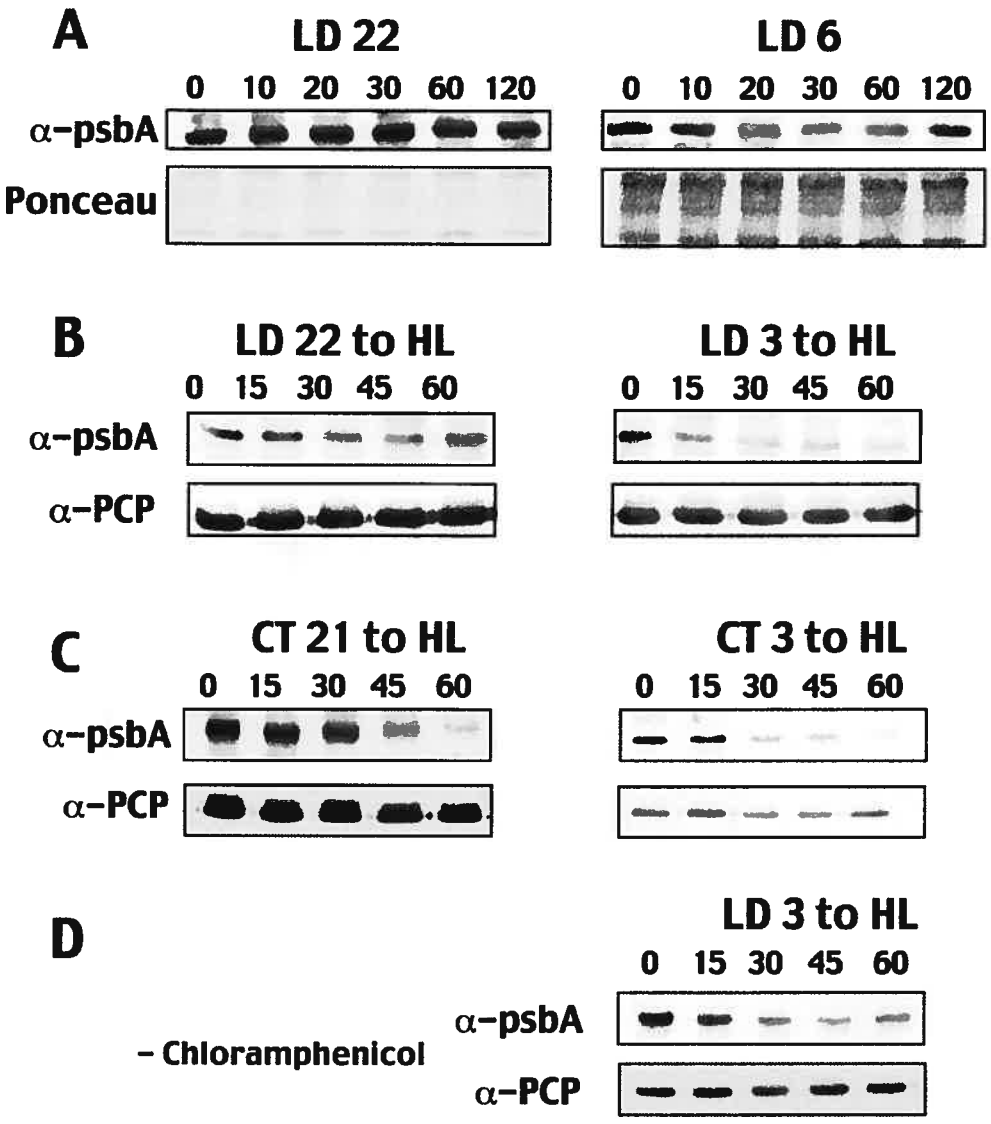

A
LD 22
LD 6
0 10 20 30 60 120
0 10 20 30 60 120
α-psbA
Ponceau
B
LD 22 to HL
LD 3 to HL
0 15 30 45 60
0 15 30 45 60
α-psbA
α-PCP
C
CT 21 to HL
CT 3 to HL
0 15 30 45 60
0 15 30 45 60
α-psbA
α-PCP
D
LD 3 to HL
0 15 30 45 60
- Chloramphenicol
α-psbA
α-PCP

Fig. V.5. psbA transcripts and protein levels are constant under LD. (A) The psbA in total protein extracts prepared by using LDS from cells at the indicated times was determined by using Western blot analysis. (B) Samples of total RNA extracted from cells at the indicated times were transferred to nylon membranes and were first probed with psbA, and then with a 23S rRNA. (C) The signals obtained by using psbA and 23S rRNA probes were quantitated using a PhosphorImager and averaged for three experiments over the 24-h period. LDT, light-dark time.

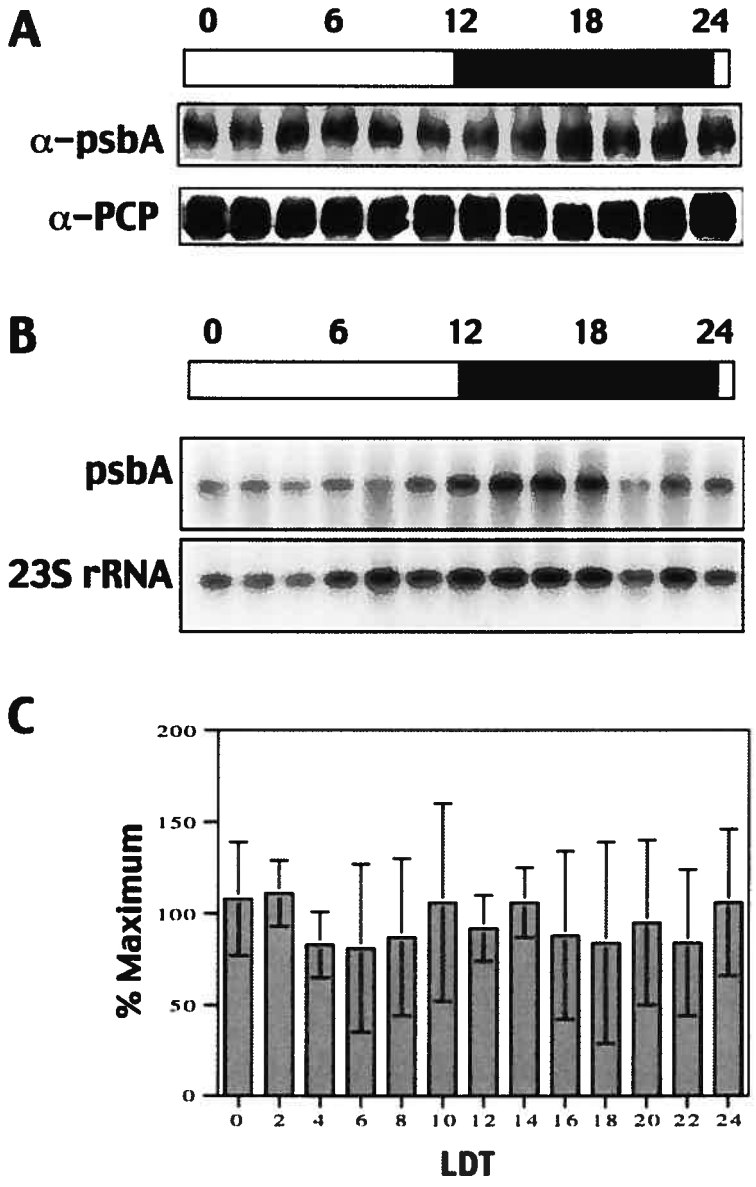

A
0
6
12
18
24
α-psbA
α-PCP
B
0
6
12
18
24
psbA
23S rRNA
C
200
150
100
50
0
% Maximum
0 2 4 6 8 10 12 14 16 18 20 22 24
LDT

Fig. V.6. Electrophoretic mobility of sequenced proteins on LDS/PAGE. Purified plastid samples (0.1 mg) were prepared and electrophoresed as described in Materials and Methods, and protein bands were identified for sequence analysis (numbered boxes at right) based on comparison to radiolabeled samples. Molecular mass markers (on the left) are, in descending order, 116, 97, 66, 45, 31, 21, 14, and 7 kDa.

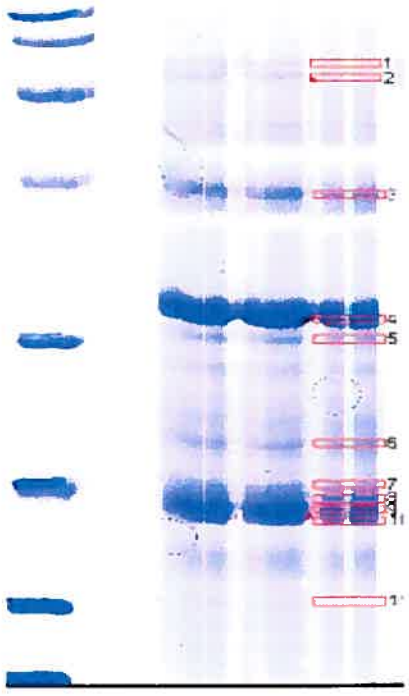

1
2
3
4
5
6
7
8
9
10
11
12

Table V.1. Plastid proteins identified by microsequencing from 2D PAGE

Protein identification no.		No. of MS/MS spectra	No. of amino acids	Predicted molecular mass, kDa
P45	BSA	66	311*	41*
P33	*Heterocapsa* OEE1	10	91	33
P31	*Gonyaulax* PCP	20	164	32
P25	*Amphidinium* LHP	5	38	19

*Recovered from the C-terminal 359 amino acids.

Table V.2. Plastid proteins identified by microsequencing from LDS/PAGE

Protein identification no.		No. of peptides identified/peptids sequenced	No. conserved amino acids/total	Predicted molecular mass, (kDa)
P70	ATP synthase β-chain	3/6	41/46	59.1
P45	ATP synthase α-chain	2/5	14/19	49.4
P31	Photosystem II D2 protein	2/6	18/25	39.6

Table V.3. Peptides sequenced from the bands indicated in Fig. V.6

Band	Peptide sequence	Identity	No. of amino acids identified
P70	TTPLELLLR		
	VTPYSFELSELR		
	FLQAGSEVSTLLGR	ATP-synthase β -chain	14/14
	N(LAV)EHGLGSLFAGVGER	ATP-synthase β -chain	14/18
P68	VTPYSFLESELR		
	FLSQPFFVAEVFTR	ATP-synthase β -chain	13/14
	NWPEHGGLSLFARGER		
P45	ELYLLLR	ATP synthase α -chain	7/7
	SVFPSLVVPPR	Actin	
	GFSYTTTAER	Actin	
	E(AF/SM)(PVS)LFFVHSR	ATP synthase α -chain	9/13
	S(YE)(DP)GQVLTLGNER	Actin	
P33	ADWDSVNAALGR	PCP	12/12
	LXXLSGPDGVTSR	PCP	8/13
	ACPAAHHQAALGSLDANGV CSLADYSAVNAALGR	PCP	27/33
P31	YFENVPR		
	NLLLNEGLR	Photosystem 2 protein D2	8/9
	QYVAAFAFYR		
	NCAQQFSQLTYAQVR	Oxygen evolving enhancer	12/15

	AXXQXXETYSFVTANR	Photosystem 2 protein D2	10/16
	LSPNXTFQEEDXDFAATTVQ LPDWR	Oxygen evolving enhancer	16/23
P24	TPNAVPWFR		
	YTEAVAYFR		
	MRPSELADGR		
	DPAGFTSDGDAQEFYR		
	ESTGVNLPADSETRR		
	VSEWGLTVENAGDYR		
	DFGLDSEPGNYGVGFSVGY G		
	PQFTSDGDAGEFYR		
	LXXXAFNALTQAALTR		
P22	LNAELANGR	LHCP	9/9
	LAANGSK	NAD ribosyl transferase	7/7
	QFDVQPHFSSAR		
P21	DRDTVFR	Cytochrome p450	6/7
	LNAELANGR	LHCP	9/9
	DTAVATFGGR	NADH dehydrogenase M	7/10
	GFQPPFLATDDAETR	LHCP	10/15
	QLLGLA(Q)GGYLAAEYFR		
P20	VGDEEGFR		
	LNAELANGR	LHCP	9/9
	GLGLGVGDPST	LHCP	8/11
	AXXAQLAELSSQR		

	PAFGTADGDVENFQR	LHCP	9/15
	NXLGFPXXGSGPLSDPEER		
	E(SPS)FGDPLGLGFYDEETR	NADH dehydrogenase 3	10/13
P19	LSAELANGR	LHCP	8/9
	RXPSELANGR	LHCP	6/10
	EVXXAVLGFFVPESYR		
	EPXFDGPVGLGSYDSETR		
	NXLGFPDGSGPLSDPEER		
P14	VTAVSLPR		
	XXXAEALGGSGR		
	LYEALGGSGR		
	STRFTLPGLXR		
	RTHVEPEGYR		
	QEVNVNQXXAR		
	RLLPESYWFQER	Photosystem I chain IV	7/11
	XXDTYFDVELR		
	GGYVLAESDLPR		

Chapter VI. Discussion and Perspectives

VI.1 The dinoflagellate plastid genome and post-transcriptional modifications

The plastid genome architecture in *Gonyaulax* is quite different from that typical of dinoflagellates. Instead of the small 2-3 kb minicircles found in other peridinin-containing plastids, the psbA gene in *Gonyaulax* associated with DNA roughly 50-150 kb. Thus minicircular genomes are not the only format available for the dinoflagellate plastid genome. Possibly, this larger size may represent an ancestral state, and if so, could be recovered by sequencing *Gonyaulax* plastid DNA. However, random sequencing of mechanically sheared DNA fragments recovered as AT-rich DNA from bisbenzimide CsCl gradients revealed no identifiable genes in ~100 kb of sequence. While this sample may have included both mitochondrial and plastid DNA, the potential risk to sequencing projects is that the *Gonyaulax* plastid genome may contain a large amount of non-coding DNA. No discrete bands are observed by ethidium bromide staining of restriction enzyme digest of the AT-rich DNA supporting the contention that this sample is of high complexity. Lastly, we find considerable polymorphism in both PCR (genomic DNA) and cDNA sequences of some plastid encoded genes (data not shown), also suggesting a complexity exceeding what is usually found.

Despite differences in genome organization, the polyuridylylation of plastid transcripts appears to be a common feature in dinoflagellate plastids. Polyuridylylation occurs at a defined position in transcripts of each protein-encoding gene and 16S rRNA, although the majority of 23S rRNA appears to be unmodified. The ubiquitous nature of the polyuridylylation suggests that poly(U)-tails may be required for either transcript translation or stability, unlike the heterogeneous poly(A)-rich extensions in cyanobacteria and other plastids that act as signals for degradation. The sequencing of over three hundred clones from a poly(A) primed cDNA library recovered 12 genes known to be present as minicircles in other species. Thus, the plastid genome in *Gonyaulax* contains psbA-D, psaA-B, atpA-B, petB, petD, as well as 23S and 16S rRNA. Indeed, these 12 chloroplast genes may comprise the total plastid coding gene sequences in *Gonyaulax* which would indicates that these plastids have the smallest genome ever described for any functional plastid. A total of only 18 minicircle plastid genes have been found in dinoflagellates, indicating that different species may have slightly different plastid gene complements.

The typical dinoflagellate plastid minicircles are superficially similar to the minicircles found in trypanosome mitochondria. This is interestingly because the trypanosme mitochondria edit their transcripts by encoded guide RNAs, and these guide RNAs are themselves polyuridylylated. Furthermore, in organelles experiencing extensive editing, such as the mitochondria of myxomycetes, 3' polyuridylylation coexist with RNA editing.(Adler, Harris et al.

1991; Horton and Landweber 2000) However, the ability to edit transcripts through uridine insertion does not seem to be a property shared by *Gonyaulax* plastids. In *Ceratium horridum*, comparisons of genomic and cDNA gene has provided evidence for substitutional editing of plastid transcripts, primarily from A to-G, but no evidence for either U insertion or deletion. There is no RNA editing found in *Amphidinium* plastid minicircle transcripts despite the 3' end polyuridylylation. This suggests that poly(U) addition to the 3' end of plastid transcript may be mechanistically different from that in *Trypanosoma* and *Myxomycetes* mitochondria, where 3' end polyuridylylation and RNA editing are found together (Adler, Harris et al. 1991; Horton and Landweber 2000). Our experiments thus provide the first example of extensive polyuridylylation occurring in the absence of RNA editing, and may result from a terminal uridylyltransferase (TUT) acting in the absence of guide RNA. It is an intriguing possibility that an enzyme other than the terminal uridylyltransferase (TUT) used to add uridine residues to transcripts undergoing editing and to guide RNAs may modify the transcripts in dinoflagellate plastids.

Independent from which enzyme conducts the 3' end polyuridylylation tailing of dinoflagellate plastid RNAs, the source of this particular activity is enigmatic since no other plastids nor cyanobacteria have yet been found to polyuridylate transcripts. Perhaps a clue may be derived from the problem of the RuBisCO form II origin in dinoflagellate plastids. This enzyme, while absent from all other chloroplasts and cyanobacteria, is found in

proteobacterial cells which may be similar to the mitochondrial origin. This suggests the RuBisCO form II may have been derived by gene transfer from the mitochondrial genome. However, the mitochondrial ancestor is not believed to have polyuridylylated the 3' end of its RNAs. The origin of this unique feature of peridinin-containing dinoflagellate plastids is another topic to need further research.

The availability of the sequence for twelve plastid genes could now allow the *Gonyaulax* plastid genomic arrangement to be dissected. The use of long distance PCR products to amplify between primers from all 12 genes could potentially determine the genomic architecture despite a large amount of non-coding DNA and genomic complexity. The organization of *Gonyaulax* plastid genes may in turn provide clues to the ancestral state of dinoflagellate plastid genome and perhaps open the way to uncovering the reason behind the choice of a minicircular genome format. Lastly, phylogenic analysis of the 12 genes sequenced could also be used to confirm the evolutionary relationship between *Gonyaulax* and other dinoflagellates.

VI.2 About *Gonyaulax* plastid encoded protein translation and degradation

The highly reticulate nature of the dinoflagellate chloroplasts and their apparently inherently fragile nature when isolated from the cell makes purification problematic. Exhaustive tests with different protocols demonstrate

that it is difficult to prepare intact and physiologically active chloroplast, although preparations keep proteins found in or inside the thylakoid membranes. Since most of dinoflagellate chloroplast minicircle encoded genes are thylakoid proteins involved in photosynthesis, this may be why some of these at least were recovered during our proteomic analysis of the isolated chloroplasts.

Biochemical analyses of protein translation and degradation have proven to be powerful tools for characterization of *Gonyaulax* chloroplast encoded genes and the regulation of their expression. The psbA synthesized in presence of cycloheximide by two different dinoflagellates *Gonyaulax* and *Amphidinium* demonstrated that dinoflagellate plastids have their own gene translating system, even though no plastid tRNAs were found. These genes, characterised as plastid encoded by this procedure, could then be identified by protein sequencing in any dinoflagellate. In crude extracts of both *Amphidinium* and *Gonyaulax* few proteins incorporate radiolabel in presence of cycloheximide, providing another line of evidence supporting the presence of only a small number of genes in the dinoflagellate plastid genome. With respect to the regulation of gene expression, dinoflagellate psbA synthesis and degradation is light regulated as it is higher plants. This is surprising, as synthesis of nuclear-encoded plastid-directed genes is often clock-controlled. Furthermore, in *Gonyaulax* both photosynthesis and oxygen evolution are bone fide circadian rhythms and the oxygen evolution is conducted by photosystem II. Clearly, those photosystem components encoded by the

plastid genome are less likely to be involved in circadian control than their nuclear counterparts. It is still an open question as to what elements in PSII constitute the control valve for this physiological phenomenon. Thus, further biochemical research on *Gonyaulax* plastid PSII proteins is a potentially very valuable line of experimentation. The mechanism whereby light directly regulates plastid gene expression will also be of interest in dinoflagellates as it is in higher plants and other alga.

The light-dependent degradation of psbA exhibits intermediate degradation rates, suggesting that adaptation to either light or dark phase requires several hours of exposure to the new conditions. This phenomenon could be caused by the slow, light-induced transition between phosphorylated and non-phosphorylated forms of psbA protein. In higher plants, the phosphorylated forms found during night phase are stable, so it may be that light activates either a protein phosphatase or inactivates a kinase. At present it is not possible to distinguish between these two possibilities, and further research about the relationship between light and photosystem protein phosphorylation and degradation will be required.

There are already suggestions from sequencing analysis that the 3' UTR region of plastid genes might contain a motif for regulating post-transcriptional RNA processing. Further sequencing of genomic DNA is likely to be able to shed further light on this. The alternative would involve a biochemical approach for the identification of plastid factors controlling post-

transcriptional processes, such as polyuridylylation, by application of in vitro RNA-binding assays, which include gel-shift and UV-cross linking techniques. However, these types of studies will likely be made extremely complicated by the difficulties encountered in plastid purification.

References

Abdallah, F., F. Salamini, et al. (2000). "A prediction of the size and evolutionary origin of the proteome of chloroplasts of Arabidopsis." Trends Plant Sci **5**(4): 141-2.

Adler, B. K., M. E. Harris, et al. (1991). "Modification of Trypanosoma brucei mitochondrial rRNA by posttranscriptional 3' polyuridine tail formation." Mol Cell Biol **11**(12): 5878-84.

Allen, J. F., S. Puthiyaveetil, et al. (2005). "Energy transduction anchors genes in organelles." Bioessays **27**(4): 426-35.

Allen, J. F. and J. A. Raven (1996). "Free-radical-induced mutation vs redox regulation: costs and benefits of genes in organelles." J Mol Evol **42**(5): 482-92.

Aphasizhev, R. (2005). "RNA uridylyltransferases." Cell Mol Life Sci.

Aphasizhev, R., S. Sbicego, et al. (2002). "Trypanosome mitochondrial 3' terminal uridylyl transferase (TUTase): the key enzyme in U-insertion/deletion RNA editing." Cell **108**(5): 637-48.

Apt, K. E., N. E. Hoffman, et al. (1993). "The gamma subunit of R-phycoerythrin and its possible mode of transport into the plastid of red algae." J Biol Chem **268**(22): 16208-15.

Armbrust, E. V., J. A. Berges, et al. (2004). "The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism." Science **306**(5693): 79-86.

Aro, E. M., I. Virgin, et al. (1993). "Photoinhibition of Photosystem II. Inactivation, protein damage and turnover." Biochim Biophys Acta **1143**(2): 113-34.

Asano, C. S., O. K. Okamoto, et al. (1998). "Diurnal patterns of incorporation of gammaP32-ATP in isolated chloroplasts from G. polyedra." Biol. Rhythm Research **29**(5): 488-493.

Bachvaroff, T. R., G. T. Concepcion, et al. (2004). "Dinoflagellate expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome." Protist **155**(1): 65-78.

Backert, S., B. L. Nielsen, et al. (1997). "The mystery of the rings: structure and replication of mitochondrial genomes from higher plants " Trends Plant Sci. **2**: 470-483.

Baginsky, S., A. Shteiman-Kotler, et al. (2001). "Chloroplast PNPase exists as a homo-multimer enzyme complex that is distinct from the Escherichia coli degradosome." Rna **7**(10): 1464-75.

Baldauf, S. L., A. J. Roger, et al. (2000). "A kingdom-level phylogeny of eukaryotes based on combined protein data." Science **290**(5493): 972-7.

Barber, J. and B. Andersson (1992). "Too much of a good thing: light can be bad for photosynthesis." Trends Biochem Sci **17**(2): 61-6.

Barbrook, A. C. and C. J. Howe (2000). "Minicircular plastid DNA in the dinoflagellate Amphidinium operculatum." Mol Gen Genet **263**(1): 152-8.

Barbrook, A. C., H. Symington, et al. (2001). "Organisation and expression of the plastid genome of the dinoflagellate Amphidinium operculatum." Mol Genet Genomics **266**(4): 632-8.

Benne, R., J. Van den Burg, et al. (1986). "Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA." Cell **46**(6): 819-26.

Berry, J. O., D. E. Breiding, et al. (1990). "Light-mediated control of translational initiation of ribulose-1, 5-bisphosphate carboxylase in amaranth cotyledons." Plant Cell **2**(8): 795-803.

Bhattacharya, D. and H. D. Schmidt (1997). Division Glaucocystophyta. New York, Springer Wien.

Bock, R. (2001). RNA editing in plant mitochondria and chloroplast. RNA editing. B. L. Bass. Oxford U. K., Oxford University Press: 38-60.

Bock, R., M. Hermann, et al. (1996). "In vivo dissection of cis-acting determinants for plastid RNA editing." Embo J **15**(18): 5052-9.

Boczar, B., J. Liston, et al. (1991). "Characterization of satellite DNA from three marine dinoflagellates (Dinophyceae): Glenodinium sp. and two members of the toxic genus, Protogonyaulax." Plant Physiol. **97**: 613-618.

Bogorad, L. (1975). "Evolution of organelles and eukaryotic genomes." Science **188**(4191): 891-8.

Bonen, L. and W. F. Doolittle (1975). "On the prokaryotic nature of red algal chloroplasts." Proc Natl Acad Sci U S A **72**(6): 2310-4.

Booij-James, I. S., W. M. Swegle, et al. (2002). "Phosphorylation of the D1 photosystem II reaction center protein is controlled by an endogenous circadian rhythm." Plant Physiol **130**(4): 2069-75.

Bryant, D. A. (1992). "Puzzles of chloroplast ancestry." Curr Biol **2**(5): 240-2.
Bungard, R. A. (2004). "Photosynthetic evolution in parasitic plants: insight from the chloroplast genome." Bioessays **26**(3): 235-47.

Burger, G., D. Saint-Louis, et al. (1999). "Complete sequence of the mitochondrial DNA of the red alga Porphyra purpurea. Cyanobacterial introns and shared ancestry of red and green algae." Plant Cell **11**(9): 1675-94.

Burger, G., Y. Zhu, et al. (2000). "Complete sequence of the mitochondrial genome of Tetrahymena pyriformis and comparison with Paramecium aurelia mitochondrial DNA." J Mol Biol **297**(2): 365-80.

Carpousis, A. J. (2002). "The Escherichia coli RNA degradosome: structure, function and relationship in other ribonucleolytic multienzyme complexes." Biochem Soc Trans **30**(2): 150-5.

Cattaneo, R. (1991). "Different types of messenger RNA editing." Annu Rev Genet **25**: 71-88.

Cavalier-Smith, T. (1975). "The origin of nuclei and eukaryotic cells." Nature **265**: 464-468.

Cavalier-Smith, T. (1982). "The origins of plastids." Biol. J. Linn. Soc. **17**: 289-306.

Cavalier-Smith, T. (1987). "The simultaneous symbiotic origin of mitochondria, chloroplasts, and microbodies." Ann N Y Acad Sci **503**: 55-71.

Cavalier-Smith, T. (2003). "Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae)." Philos Trans R Soc Lond B Biol Sci **358**(1429): 109-33; discussion 133-4.

Chaudhuri, S. and P. Maliga (1996). "Sequences directing C to U editing of the plastid psbL mRNA are located within a 22 nucleotide segment spanning the editing site." Embo J **15**(21): 5958-64.

Chitnis, P. R. (2000). Import, assembly and degradation of chloroplast proteins. Photosynthesis: a comprehensive treatise. A. S. Raghavendra. Cambridge, U. K., Cambridge University Press.

Chu, K. H., J. Qi, et al. (2004). "Origin and phylogeny of chloroplasts revealed by a simple correlation analysis of complete genomes." Mol Biol Evol **21**(1): 200-6.

Chudnovsky, Y. and J. F. Li (2002). "Cloning, expression, and characterization of a histone-like protein from the marine dinoflagellate *Ligulodinium polyedra* (dinophyceae)." J. Phycol. **38**: 543-550.

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

Conway, L. and M. Wickens (1985). "A sequence downstream of A-A-U-A-A-A is required for formation of simian virus 40 late mRNA 3' termini in frog oocytes." Proc Natl Acad Sci U S A **82**(12): 3949-53.

Copertino, D. W. and R. B. Hallick (1991). "Group II twintron: an intron within an intron in a chloroplast cytochrome b-559 gene." Embo J **10**(2): 433-42.

Corneille, S., K. Lutz, et al. (2000). "Conservation of RNA editing between rice and maize plastids: are most editing events dispensable?" Mol Gen Genet **264**(4): 419-24.

Danon, A. (1997). "Translational regulation in the chloroplast." Plant Physiol **115**(4): 1293-8.

Danon, A. and S. P. Mayfield (1994). "ADP-dependent phosphorylation regulates RNA-binding in vitro: implications in light-modulated translation." Embo J **13**(9): 2227-35.

Danon, A. and S. P. Mayfield (1994). "Light-regulated translation of chloroplast messenger RNAs through redox potential." Science **266**(5191): 1717-9.

de Grey, A. D. (2005). "Forces maintaining organellar genomes: is any as strong as genetic code disparity or hydrophobicity?" Bioessays **27**(4): 436-46.

Delwiche, C. and J. Palmer (1996). "Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plasmids." Mol. Biol. Evol. **13**: 873-882.

Delwiche, C. F., M. Kuhsel, et al. (1995). "Phylogenetic analysis of tufA sequences indicates a cyanobacterial origin of all plastids." Mol Phylogenet Evol **4**(2): 110-28.

Delwiche, C. F. and J. D. Palmer (1997). The origin of plastids and their spread via secondary endosymbiosis. The origins of algae and their plastids. D. Bhattacharya. Vienna, Springer-Verlag: 53-86.

Dionisio-Sese, M. L., T. Maruyama, et al. (2001). "Photosynthesis of Prochloron as affected by environmental factors." Mar Biotechnol (NY) **3**(1): 74-9.

Dodge, J. D. (1975). "A survey of chloroplast ultrastructure in Dinophyceae." Phycologia **14**: 253-263.

Doolittle, W. F. (1999). "Lateral genomics." Trends Cell Biol. **9**: M5-M8.

Douglas, A. E. and J. A. Raven (2003). "Genomes at the interface between bacteria and organelles." Philos Trans R Soc Lond B Biol Sci **358**(1429): 5-17; discussion 517-8.

Douglas, S. E. (1993). Probable evolutionary history of cryptomonad algae. . NewYork, Chapman andHall press.

Douglas, S. E. and S. L. Penny (1999). "The plastid genome of the cryptophyte alga, Guillardia theta: complete sequence and conserved synteny groups confirm its common ancestry with red algae." J Mol Evol **48**(2): 236-44.

Edmonds, M. (2002). "A history of poly A sequences: from formation to factors to function." Prog Nucleic Acid Res Mol Biol **71**: 285-389.

Elias, M. C., R. Marques-Porto, et al. (2001). "Transcription rate modulation through the Trypanosoma cruzi life cycle occurs in parallel with changes in nuclear organisation." Mol Biochem Parasitol **112**(1): 79-90.

Fagan, T., J. W. Hastings, et al. (1998). "The phylogeny of glyceraldehyde-3-phosphate dehydrogenase indicates lateral gene transfer from cryptomonads to dinoflagellates." J Mol Evol **47**(6): 633-9.

Fagan, T., D. Morse, et al. (1999). "Circadian synthesis of a nuclear encoded chloroplast Glyceraldehyde-3-phosphate dehydrogenase in the dinoflagellate *Gonyaulax polyedra* is translationally controlled." Biochemistry **38**: 7689-7695.

Fast, N. M., J. C. Kissinger, et al. (2001). "Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids." Mol Biol Evol **18**(3): 418-26.

Feagin, J. E., J. M. Abraham, et al. (1988). "Extensive editing of the cytochrome c oxidase III transcript in Trypanosoma brucei." Cell **53**(3): 413-22.

Filfilan, S. A. and D. C. Sigee (1977). "Continuous DNA replication in the nucleus of the dinoflagellate Prorocentrum micans (Ehrenberg)." J Cell Sci **27**: 81-90.

Formm, H., M. Devic, et al. (1985). "Control of psbA gene expression: In mature spirodela chloroplast: light regulation of 32 k Da protein synthesis is independant of transcript leve." EMBOJ. **4**(291-295).

Franker, C. K., L. F. Smith, et al. (1973). "Morphotypic transition during synchronous growth of Cryptothecodinium (Gyrodinium) cohnii." Arch Mikrobiol **90**(3): 255-62.

Freyer, R., M. C. Kiefer-Meyer, et al. (1997). "Occurrence of plastid RNA editing in all major lineages of land plants." Proc Natl Acad Sci U S A. **94**(12): 6285-90.

Friso, G., L. Giacomelli, et al. (2004). "In-depth analysis of the thylakoid membrane proteome of Arabidopsis thaliana chloroplasts: new proteins, new functions, and a plastid proteome database." Plant Cell **16**(2): 478-99.

Gautier, a., L. Michel-Salamin, et al. (1986). "Electron microscopy of the chromosome of dinoflagellates in situ: confirmation of Bouligand's liquid crystal hypothesis." *J. Ultrastructure Mol. Structure Res.* **97**: 10-30.

Germain, H., S. Rudd, et al. (2005). "A 6374 unigene set corresponding to low abundance transcripts expressed following fertilization in Solanum chacoense Bitt, and characterization of 30 receptor-like kinases." Plant Mol Biol **59**(3): 515-32.

Gevaert, K., H. Demol, et al. (2001). "Protein identification based on matrix assisted laser desorption/ionization-post source decay-mass spectrometry." Electrophoresis **22**(1645-51).

Gibbs, S. (1981). "The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae." Ann. N.Y. Acad. Sci. **361**: 193-208.

Gibbs, S. P. (1979). "The route of entry of cytoplasmically synthesized proteins into chloroplasts of algae possessing chloroplast ER." J Cell Sci **35**: 253-66.

Gillham, N. W., J. E. Boynton, et al. (1994). "Translational regulation of gene expression in chloroplasts and mitochondria." Annu Rev Genet **28**: 71-93.

Giovannoni, S. J., N. Wood, et al. (1993). Molecular phylogeny of oxygenic cells and organelles based on small subunit ribosomal RNA sequences. New York, Chapman and Hall Press.

Goboom, J., E. Nordhoff, et al. (1999). "Sample purification and preparation technique based on nano-scale reverse-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry." J. Mass Spectrum **34**: 105-116.

Gray, M. W., G. Burger, et al. (1999). "Mitochondrial evolution." Science **283**(5407): 1476-81.

Green, B. R. (2004). "The chloroplast genome of dinoflagellates--a reduced instruction set?" Protist **155**(1): 23-31.

Grossman, A. R., D. Bhaya, et al. (1995). "Light-harvesting complexes in oxygenic photosynthesis: diversity, control, and evolution." Annu Rev Genet **29**: 231-88.

Gruissem, W., A. Barkan, et al. (1988). "Transcriptional and post-transcriptional control of plastid mRNA levels in higher plants." Trends Genet **4**(9): 258-63.

Grzebyk, D. and O. Schofield (2003). "The mesozoic radiation of eukaryotic algea: the portable plastid hypothesis." J. Phycol. **39**: 259-267.

Guillard, R. R. L. and J. H. Ryther (1962). "Studies on marine planktonic diatoms: *Cyclotella nana* Hufstedt and *Denotula confervacea* (Cleve) Gran." Can. J. Microbiol. **8**: 229-239.

Hackett, J. D., D. M. Anderson, et al. (2004). "Dinoflagellates: a remarkable evolutionary experiment." Am. J. Bot. **91**: 1523-1534.

Hackett, J. D., T. E. Scheetz, et al. (2005). "Insights into a dinoflagellate genome through expressed sequence tag analysis." BMC Genomics **6**(1): 80.

Hackett, J. D., H. S. Yoon, et al. (2004). "Migration of the plastid genome to the nucleus in a peridinin dinoflagellate." Curr Biol **14**(3): 213-8.

Hall, D. O. and O. O. Rao (1994). Light absorption and the two photosystems Photosynthesis. Great Britain, Cambridge University Press: 66.

Hallick, R. B., L. Hong, et al. (1993). "Complete sequence of Euglena gracilis chloroplast DNA." Nucleic Acids Res **21**(15): 3537-44.

Harper, J. T., E. Waanders, et al. (2005). "On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes." Int J Syst Evol Microbiol **55**(Pt 1): 487-96.

Hastings, J. W., L. Astrachan, et al. (1961). "A persistent daily rhythm in photosynthesis." J Gen Physiol **45**: 69-76.

Haugel-Nielsen, J., E. Hajnsdorf, et al. (1996). "The rpsO mRNA of Escherichia coli is polyadenylated at multiple sites resulting from endonucleolytic processing and exonucleolytic degradation." Embo J **15**(12): 3144-52.

Hayes, R., J. Kudla, et al. (1999). "Degrading chloroplast mRNA: the role of polyadenylation." Trends Biochem Sci **24**(5): 199-202.

Hayes, R., J. Kudla, et al. (1996). "Chloroplast mRNA 3'-end processing by a high molecular weight protein complex is regulated by nuclear encoded RNA binding proteins." Embo J **15**(5): 1132-41.

He, W. Z. and R. Malkin (2000). Photosystem I and II. Photosynthesis. A. S. Raghavendra. United Kingdom, Cambridge University Press: 29-43.

Herman, E. M. and B. M. Sweeney (1975). "Circadian rhythm of chloroplast ultrastructure in Gonyaulax polyedra, concentric organization around a central cluster of ribosomes." J.Ultrastruct. Res. **50**: 347-354.

Hiller, R., P. Wrench, et al. (1995). "The light harvesting chlorophyll a-c binding protein of dinoflagellates: a putative polyprotein." FEBS letters **363**: 175-178.

Hiller, R. G. (2001). "'Empty' minicircles and petB/atpA and psbD/psbE (cytb559 alpha) genes in tandem in Amphidinium carterae plastid DNA." FEBS Lett **505**(3): 449-52.

Hirose, T., H. Fan, et al. (1996). "Occurrence of silent RNA editing in chloroplasts: its species specificity and the influence of environmental and developmental conditions." Plant Mol Biol **30**(3): 667-72.

Hirose, T. and M. Sugiura (1996). "Cis-acting elements and trans-acting factors for accurate translation of chloroplast psbA mRNAs:

development of an in vitro translation system from tobacco chloroplasts." Embo J **15**(7): 1687-95.

Hoch, B., R. M. Maier, et al. (1991). "Editing of a chloroplast mRNA by creation of an initiation codon." Nature **353**(6340): 178-80.

Hofmann, E., P. M. Wrench, et al. (1996). "Structural basis of light harvesting by carotenoids: peridinin- chlorophyll-protein from *Amphidinium carterae*." Science **272**(5269): 1788-91.

Horton, T. L. and L. F. Landweber (2000). "Mitochondrial RNAs of myxomycetes terminate with non-encoded 3' poly(U) tails." Nucleic Acids Res **28**(23): 4750-4.

Howe, C. J., A. C. Barbrook, et al. (2003). "Evolution of the chloroplast genome." Philos Trans R Soc Lond B Biol Sci **358**(1429): 99-106; discussion 106-7.

Howe, C. J., A. C. Barbrook, et al. (2000). "Organelle genes--do they jump or are they pushed?" Trends Genet **16**(2): 65-6.

Huang, C. Y., M. A. Ayliffe, et al. (2003). "Direct measurement of the transfer rate of chloroplast DNA into the nucleus." Nature **422**(6927): 72-6.

Hurlbert, S. H. (1971). "The nonconcept of species diversity: a critique and alternative parameters." Ecology **52**(4): 577-586.

Inada, M., T. Sasaki, et al. (2004). "A systematic search for RNA editing sites in pea chloroplasts: an editing event causes diversification from the evolutionarily conserved amino acid sequence." Plant Cell Physiol **45**(11): 1615-22.

Inagaki, Y., A. Simpson, et al. (2004). "Phylogenetic artifacts can be caused by leucine, serine, and arginine codon usage heterogeneity: dinoflagellate plastid origins as a case study." Syst Biol **53**(4): 582-93.

Ishida, K. and B. R. Green (2002). "Second- and third-hand chloroplasts in dinoflagellates: phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont." Proc Natl Acad Sci U S A **99**(14): 9294-9.

Jacobson, A. (1996). Translation control. Translation control. J. W. B. Hershey, M. B. Mathews and N. Sonrnberg. Plainview, New York, Cold Spring Harbor Lab. Press.

Jakowitsch, J., C. Neumann-Spallart, et al. (1996). "In vitro import of pre-ferredoxin-NADP+-oxidoreductase from Cyanophora paradoxa into cyanelles and into pea chloroplasts." FEBS Lett **381**(1-2): 153-5.

Jeffrey, S., M. Seilicki, et al. (1975). "Chloroplast pigment patterns in dinoflagellates." J. Phycol. **11**: 374-384.

Jones, H., Ed. (1995). Plant gene transfer and expression protocols. Methods in Plant Molecular Biology. Totowa, N.J., Humana Press.

Kapoor, S. and M. Sugiura (2000). Expression and regulation of plastid genes. Photosynthesis. A. S. Raghavendra. Cambridge U. K., Cambridge University Press: 72-86.

Kapoor, S., J. Y. Suzuki, et al. (1997). "Identification and functional significance of a new class of non-consensus-type plastid promoters." Plant J **11**(2): 327-37.

Keegstra, K. (1989). "Transport and routing of proteins into chloroplasts." Cell **56**(2): 247-53.

Keeling, P. J. (2004). "Diversity and evolutionary history of plastids and their hosts." Am. J. Bot. **91**: 1481-1493.

Kies, L. and B. P. Kremer (1990). Phylum Glaucocystosphyta. Boston, Jones & Bartlett.

Killian, O. and P. G. Kroth (2003). "Evolution of protein targeting into "complex" plastids: the "secretory transport hypothesis"." Plant Biol. **5**: 350-358.

Kim, J., L. A. Eichacker, et al. (1994). "Chlorophyll regulates accumulation of the plastid-encoded chlorophyll proteins P700 and D1 by increasing apoprotein stability." Plant Physiol **104**(3): 907-16.

Kim, J. and S. P. Mayfield (1997). "Protein disulfide isomerase as a regulator of chloroplast translational activation." Science **278**(5345): 1954-7.

Klein, R. R., H. S. Mason, et al. (1988). "Light-regulated translation of chloroplast proteins. I. Transcripts of psaA-psaB, psbA, and rbcL are associated with polysomes in dark-grown and illuminated barley seedlings." J Cell Biol **106**(2): 289-301.

Klein, R. R. and J. E. Mullet (1987). "Control of gene expression during higher plant chloroplast biogenesis. Protein synthesis and transcript levels of psbA, psaA-psaB, and rbcL in dark-grown and illuminated barley seedlings." J Biol Chem **262**(9): 4341-8.

Klein, R. R. and J. E. Mullet (1990). "Light-induced transcription of chloroplast genes. psbA transcription is differentially enhanced in illuminated barley." J Biol Chem **265**(4): 1895-902.

Kobayashi, H., J. Ngernprasirtsiri, et al. (1990). "Transcriptional regulation and DNA methylation in plastids during transitional conversion of chloroplasts to chromoplasts." Embo J **9**(2): 307-13.

Koivuniemi, A., E. M. Aro, et al. (1995). "Degradation of the D1- and D2-proteins of photosystem II in higher plants is regulated by reversible phosphorylation." Biochemistry **34**(49): 16022-9.

Koumandou, V. L., R. E. Nisbet, et al. (2004). "Dinoflagellate chloroplasts--where have all the genes gone?" Trends Genet **20**(5): 261-7.

Kozak, M. (1992). "Regulation of translation in eukaryotic systems." Annu Rev Cell Biol **8**: 197-225.

Krupinska, K. and K. Apel (1989). "Light-induced transformation of etioplasts to chloroplast of barley without transcriptional control of plastid gene expression." Mol. Gen. Gent. **219**: 467-473.

Kudla, J., R. Hayes, et al. (1996). "Polyadenylation accelerates degradation of chloroplast mRNA." Embo J **15**(24): 7137-46.

Kuroda, H., K. Kobashi, et al. (1996). "Possible involvement of a low redox potential component(s) downstream of photosystem I in the translational regulation of the D1 subunit of the photosystem II reaction centre in isolated pea chloroplast." Plant Cell physiol. **37**: 754-761.

Laatsch, T., S. Zauner, et al. (2004). "Plastid-Derived Single Gene Minicircles of the Dinoflagellate Ceratium horridum Are Localized in the Nucleus." Mol Biol Evol **21**(7): 1318-22.

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

Lang, B. F., M. W. Gray, et al. (1999). "Mitochondrial genome evolution and the origin of eukaryotes." Annu Rev Genet **33**: 351-97.

Lang, M., K. E. Apt, et al. (1998). "Protein transport into "complex" diatom plastids utilizes two different targeting signals." J Biol Chem **273**(47): 30973-8.

Lawlor, D. L. (2001). Molecular biology f the photosynthetic system. Photosynthesis. New York, Springer-verlag: 242-277.

Le, Q., R. Jovine, et al. (2001). "PCP is not implicated in the photosynthesis rhythm of Gonyaulax despite circadian regulation of its translation." Biol. Rhythm Research **32**(5): 579-594.

Le, Q. H., P. Markovic, et al. (1997). "Sequence and genomic organization of the peridinin-chlorophyll a-protein from *Gonyaulax polyedra*." Mol. Gen. Genet. **255**: 595-604.

Lee, D.-H., M. Mittag, et al. (1993). "Molecular cloning and genomic organization of a gene for luciferin binding protein from the dinoflagellate *Gonyaulax polyedra*." J. Biol. Chem. **268**: 8842-8850.

Lee, R. E. (1989). Phycology, Cambriage University Press: 41-43.

Lidie, K. B., J. C. Ryan, et al. (2005). "Gene expression in Florida red tide dinoflagellate Karenia brevis: analysis of an expressed sequence tag library and development of DNA microarray." Mar Biotechnol (NY) **7**(5): 481-93.

Lin, S., H. Zhang, et al. (2002). "Widespread and extensive editing of mitochondrial mRNAS in dinoflagellates." J Mol Biol **320**(4): 727-39.

Lisitsky, I., P. Klaff, et al. (1996). "Addition of destabilizing poly (A)-rich sequences to endonuclease cleavage sites during the degradation of chloroplast mRNA." Proc Natl Acad Sci U S A **93**(23): 13398-403.

Liu, Y. G., Y. Chen, et al. (2005). "Amplification of genomic sequences flanking T-DNA insertions by thermal asymmetric interlaced polymerase chain reaction." Methods Mol Biol **286**: 341-8.

Livolant, F. and Y. Bouligand (1978). "New observation on the twisted arrangement of dinoflagellate chromosomes." Chromosome **68**: 21-44.

Lutz, K. A. and P. Maliga (2001). "Lack of conservation of editing sites in mRNAs that encode subunits of the NAD(P)H dehydrogenase complex in plastids and mitochondria of Arabidopsis thaliana." Curr Genet **40**(3): 214-9.

MacDonald, C. C. and J. L. Redondo (2002). "Reexamining the polyadenylation signal: were we wrong about AAUAAA?" Mol Cell Endocrinol **190**(1-2): 1-8.

Malnoe, P., S. P. Mayfield, et al. (1988). "Comparative analysis of the biogenesis of photosystem II in the wild-type and Y-1 mutant of Chlamydomonas reinhardtii." J Cell Biol **106**(3): 609-16.

Manuell, A., M. V. Beligni, et al. (2004). "Regulation of chloroplast translation: interactions of RNA elements, RNA-binding proteins and the plastid ribosome." Biochem Soc Trans **32**(Pt 4): 601-5.

Markovic, P., T. Roenneberg, et al. (1996). "Phased protein synthesis at several circadian times does not change protein levels in *Gonyaulax*." J Biol Rhythms **11**(1): 57-67.

Martin, W. (2003). "Gene transfer from organelles to the nucleus: frequent and in big chunks." Proc Natl Acad Sci U S A **100**(15): 8612-4.

Martin, W. and R. G. Herrmann (1998). "Gene transfer from organelles to the nucleus: how much, what happens, and Why?" Plant Physiol **118**(1): 9-17.

Matsuzaki, M., O. Misumi, et al. (2004). "Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D." Nature **428**(6983): 653-7.

Mattoo, A. K., J. B. Marder, et al. (1989). "Dynamics of the photosystem II reaction center." Cell **56**(2): 241-6.

Mattoo, A. K., U. Pick, et al. (1981). "The rapidly metabolized 32,000-dalton polypeptide of the chloroplast is the "proteinaceous shield" regulating photosystem II electron transport and mediating diuron herbicide sensitivity." Proc Natl Acad Sci U S A **78**(3): 1572-6.

Mayfield, S. P., C. B. Yohn, et al. (1995). "Regulation of chloroplast gene expression." Annu. Rev. Plant Physiol. **46**: 147-166.

McFadden, G. I. (2001). "Chloroplast origin and integration." Plant Physiol **125**(1): 50-3.

Monde, R. A., J. C. Greene, et al. (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**(4): 529-42.

Moore, R. B., K. M. Ferguson, et al. (2003). "Highly organized structure in the non-coding region of the psbA minicircle from clade C *Symbiodinium*." Int. J. Syst. Evol. Microbiol. **53**: 1725-1734.

Moreira, D., H. Le Guyader, et al. (2000). "The origin of red algae and the evolution of chloroplasts." Nature **405**(6782): 69-72.

Moreno Diaz de la Espina, S., E. Alverca, et al. (2005). "Organization of the genome and gene expression in a nuclear environment lacking histones and nucleosomes: the amazing dinoflagellates." Eur J Cell Biol **84**(2-3): 137-49.

Morse, D., P. Salois, et al. (1995). "A nuclear encoded form II rubisco in dinoflagellates." Science **268**: 1622-1624.

Muller, H. J. (1964). "Th of recombination to mutational advance.e relation." Mutat. Res. **106**: 2-9.

Mullet, J. E., P. G. Klein, et al. (1990). "Chlorophyll regulates accumulation of the plastid-encoded chlorophyll apoproteins CP43 and D1 by increasing apoprotein stability." Proc Natl Acad Sci U S A **87**(11): 4038-42.

Narang, F., L. McIntosh, et al. (1984). "Nucleotide sequence of the ribulose bisphosphate carboxylase gene from Rhodospirillum rubrum." Mol. Gen. Genet. **193**: 220-224.

Nassoury, N., M. Cappadocia, et al. (2003). "Plastid ultrastructure defines the protein import pathway in dinoflagellates." J Cell Sci **116**(Pt 14): 2867-74.

Nassoury, N., L. Fritz, et al. (2001). "Circadian changes in ribulose-1,5-bisphosphate carboxylase/oxygenase distribution inside individual chloroplasts can account for the rhythm in dinoflagellate carbon fixation." Plant Cell **13**(4): 923-34.

Nassoury, N. and D. Morse (2005). "Protein targeting to the chloroplasts of photosynthetic eukaryotes: getting there is half the fun." Biochim Biophys Acta **1743**(1-2): 5-19.

Nelissen, B., Y. Van de Peer, et al. (1995). "An early origin of plastids within the cyanobacterial divergence is suggested by evolutionary trees based on complete 16S rRNA sequences." Mol Biol Evol **12**(6): 1166-73.

Nickelsen, J. (2003). "Chloroplast RNA-binding proteins." Curr Genet **43**(6): 392-9.

Nisbet, R. E. R., V. L. Koumandou, et al. (2004). "Novel plastid gene minicircles in the dinoflagellate Amphidinium operculatum." Gene **331**: 141-147.

Norman, J. E. and M. W. Gray (2001). "A complex organization of the gene encoding cytochrome oxidase subunit 1 in the mitochondrial genome of the dinoflagellate, Crypthecodinium cohnii: homologous recombination generates two different cox1 open reading frames." J Mol Evol **53**(4-5): 351-63.

Norris, B. J. and D. J. Miller (1994). "Nucleotide sequence of a cDNA clone encoding the precursor of the peridinin-chlorophyll a-binding protein from the dinoflagellate Symbiodinium sp." Plant Mol Biol **24**(4): 673-7.

Ohyama, K., H. Fukuzawa, et al. (1986). "Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA " Nature **322**: 572-574.

Palmer, J. (1995). "Rubisco rules broken; gene transfer triumphs." Bioessays **17**: 1005-1008.

Palmer, J. F. (2003). "The symbiotic birth and spread of plastids: how many times and whodunit?" J. Phycol. **39**: 4-11.

Patron, N. J., R. F. Waller, et al. (2005). "Complex protein targeting to dinoflagellate plastids." J Mol Biol **348**(4): 1015-24.

Peltier, J. B., G. Friso, et al. (2000). "Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins." Plant Cell **12**(3): 319-41.

Preiss, S., S. Schrader, et al. (2001). "Rapid, ATP-dependent degradation of a truncated D1 protein in the chloroplast." Eur J Biochem **268**(16): 4562-9.

Prezelin, B. and F. Haxo (1976). "Purification and characterization of peridinin-chlorophyll a-proteins from the marine dinoflagellates *Glenodinium* and *Gonyaulax*." Planta **130**: 251-256.

Proudfoot, N. J. and G. G. Brownlee (1976). "3' non-coding region sequences in eukaryotic messenger RNA." Nature **263**(5574): 211-4.

Rae, P. M. and R. E. Steele (1978). "Modified bases in the DNAs of unicellular eukaryotes: an examination of distributions and possible roles, with emphasis on hydroxymethyluracil in dinoflagellates." Biosystems **10**(1-2): 37-53.

Rapp, J. C., B. J. Baumgartner, et al. (1992). "Quantitative analysis of transcription and RNA levels of 15 barley chloroplast genes.

Transcription rates and mRNA levels vary over 300-fold; predicted mRNA stabilities vary 30-fold." J Biol Chem **267**(30): 21404-11.

Regan, M. and R. Gutell (1995). "Are red algae plants?" J. Linn. Soc. **118**: 81-105.

Rensing, L., W. R. Taylor, et al. (1980). "The effects of protein synthesis inhibitors on the *Gonyaulax* clock II: the effect of cycloheximide on ultrastructural parameters." J. Comp. Physiol. **138**: 9-18.

Rizzo, P. J. (1981). "Comparative aspects of basic chromatin proteins in dinoflagellates." Biosystems **14**(3-4): 433-43.

Rochaix, J. D. (1992). "Post-transcriptional steps in the expression of chloroplast genes." Annu Rev Cell Biol **8**: 1-28.

Rochaix, J. D. (2001). "Posttranscriptional control of chloroplast gene expression. From RNA to photosynthetic complex." Plant Physiol **125**(1): 142-4.

Roger, A. J. (1999). "Reconstructing Early Events in Eukaryotic Evolution." Am Nat **154**(S4): S146-S163.

Rott, R., G. Zipor, et al. (2003). "RNA polyadenylation and degradation in cyanobacteria are similar to the chloroplast but different from Escherichia coli." J Biol Chem **278**(18): 15771-7.

Rowan, R., S. M. Whitney, et al. (1996). "Rubisco in marine symbiotic dinoflagellates: form II enzymes in eukaryotic oxygenic phototrophs encoded by a nuclear multigene family." Plant Cell **8**(3): 539-53.

Saldanha, R., G. Mohr, et al. (1993). "Group I and group II introns." Faseb J **7**(1): 15-24.

Sambrook, J., E. Fritsch, et al. (1989). Molecular cloning: a laboratory manual., Cold Spring Harbor Laboratory Press.

Sarkar, N. (1997). "Polyadenylation of mRNA in prokaryotes." Annu. Rev. Biochem. **66**: 173-197.

Sasaki, T., Y. Yukawa, et al. (2003). "Identification of RNA editing sites in chloroplast transcripts from the maternal and paternal progenitors of

tobacco (Nicotiana tabacum): comparative analysis shows the involvement of distinct trans-factors for ndhB editing." Mol Biol Evol **20**(7): 1028-35.

Schevchenko, A., M. Wilm, et al. (1996). "Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels." Anal. Chem. **68**: 850-858.

Schmitter, R. (1971). "The fine structure of *Gonyaulax polyedra*, a bioluminescent marine dinoflagellate." J Cell Sci **9**: 147-173.

Schneider, A. and L. Marechal-Drouard (2000). "Mitochondrial tRNA import: are there distinct mechanisms?" Trends Cell Biol **10**(12): 509-13.

Schuler, M. and R. Zielinski (1989). Methods in Plant Molecular Biology. San Diego, Academic Press.

Schuster, G., I. Lisitsky, et al. (1999). "Polyadenylation and degradation of mRNA in the chloroplast." Plant Physiol **120**(4): 937-44.

Sexton, T. B., D. A. Christopher, et al. (1990). "Light-induced switch in barley psbD-psbC promoter utilization: a novel mechanism regulating chloroplast gene expression." Embo J **9**(13): 4485-94.

Sheets, M. D., S. C. Ogg, et al. (1990). "Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro." Nucleic Acids Res **18**(19): 5799-805.

Shen, B. and H. M. Goodman (2004). "Uridine addition after microRNA-directed cleavage." Science **306**(5698): 997.

Smith, H. C., J. M. Gott, et al. (1997). "A guide to RNA editing." Rna **3**(10): 1105-23.

Spector, D., Ed. (1984). Dinoflagellate nuclei. Dinoflagellates. New York, Academic Press.

Spetea, C., T. Hundal, et al. (1999). "GTP bound to chloroplast thylakoid membranes is required for light-induced, multienzyme degradation of the photosystem II D1 protein." Proc Natl Acad Sci U S A **96**(11): 6547-52.

Stiller, J. W. and B. D. Hall (1997). "The origin of red algae: implications for plastid evolution." Proc Natl Acad Sci U S A **94**(9): 4520-5.

Stirdivant, S. M., L. D. Crossland, et al. (1985). "DNA supercoiling affects in vitro transcription of two maize chloroplast genes differently." Proc Natl Acad Sci U S A **82**(15): 4886-90.

Subramanian, A. R. (1993). "Molecular genetics of chloroplast ribosomal proteins." Trends Biochem Sci **18**(5): 177-81.

Sugiura, M. (1996). Structure and replication of chloroplast DNA. Molecular genetics of photosynthesis. B. Anderson and J. Barber. Oxford, Oxford Press: 58-74.

Sugiura, M., T. Hirose, et al. (1998). "Evolution and mechanism of translation in chloroplasts." Annu Rev Genet **32**: 437-59.

Svab, Z. and P. Maliga (1993). "High-frequency plastid transformation in tobacco by selection for a chimeric aadA gene." Proc Natl Acad Sci U S A **90**(3): 913-7.

Sweeney, B. M. (1960). "The photosynthetic rhythm in single cells of Gonyaulax polyedra." Cold Spring Harb Symp Quant Biol **25**: 145-8.

Takashita, K., M. Ishikura, et al. (2003). "Comparison of phylogenies based on nuclear-encoded SSU rDNA and plastid-encoded psbA in the symbiotic dinoflagellate genus Symbiodinium." Phycologia **42**: 285-291.

Takishita, K., K. Ishida, et al. (2003). "An enigmatic GAPDH gene in the symbiotic dinoflagellate genus Symbiodinium and its related species (the order Suessiales): possible lateral gene transfer between two eukaryotic algae, dinoflagellate and euglenophyte." Protist **154**(3-4): 443-54.

Tengs, T., O. J. Dahlberg, et al. (2000). "Phylogenetic analyses indicate that the 19'Hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin." Mol Biol Evol **17**(5): 718-29.

The Arabidopsis genome initiative (2000). "Analysis of the genome sequence of the flowering plant Arabidopsis thaliana." Nature **408**(6814): 796-815.

Thomas, R. N. and E. R. Cox (1973). "Observations on the symbiosis of peridinium balticum and its intracellular alga I. Ultrastructure." J. Phycol. **9**: 304-323.

Thorsness, P. E. and E. R. Weber (1996). "Escape and migration of nucleic acids between chloroplasts, mitochondria, and the nucleus." Int Rev Cytol **165**: 207-34.

Tiller, K. and G. Link (1993). "Phosphorylation and dephosphorylation affect functional characteristics of chloroplast and etioplast transcription systems from mustard (Sinapis alba L.)." Embo J **12**(5): 1745-53.

Timmis, J. N., M. A. Ayliffe, et al. (2004). "Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes." Nat Rev Genet **5**(2): 123-35.

Trebitsh, T., A. Levitan, et al. (2000). "Translation of chloroplast psbA mRNA is modulated in the light by counteracting oxidizing and reducing activities." Mol Cell Biol **20**(4): 1116-23.

Triplett, E. L., N. S. Govind, et al. (1993). "Characterization of the sequence organization of DNA from the dinolagellate *Heterocapsa pygmea* (*Glenodinium sp.*)." Mol.Mar.Biol.Biotech. **2**: 239-245.

Tyystjarvi, E. and E. M. Aro (1996). "The rate constant of photoinhibition, measured in lincomycin-treated leaves, is directly proportional to light intensity." Proc Natl Acad Sci U S A **93**(5): 2213-8.

Vogel, J., T. Borner, et al. (1999). "Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants." Nucleic Acids Res **27**(19): 3866-74.

Walter, M., J. Kilian, et al. (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**(24): 6905–6914.

Wang, Y., L. Jensen, et al. (2005). "Synthesis and degradation of dinoflagellate plastid-encoded psbA proteins are light-regulated, not circadian-regulated." Proc Natl Acad Sci U S A **102**(8): 2844-9.

Watanabe, K. I., Y. Bessho, et al. (1999). "Mitochondrial genes are found on minicircle DNA molecules in the mesozoan animal Dicyema." J Mol Biol. **286**(3): 645-50.

Wettern, A. K. and I. Ohad (1984). "Light-induced turnover of thylakoids polypeptides in *chlamydominas reinhardtii*." Isral. J. Botany **33**: 253-263.

Whitney, S. and T. Andrews (1998). "The CO2/O2 specificity of single-subunit ribulose-bisphosphate carboxylase from the dinoflagellate *Amphidinium carterae*." Aust. J. Plant Physiol. **25**: 131-138.

Whitney, S. and D. Yellowlees (1995). "Preliminary investigations into the structure and activity of ribulose bisphosphate carboxylase from two photosynthetic dinoflagellates." J. Phycol. **31**: 138-146.

Wilusz, J., S. M. Pettine, et al. (1989). "Functional analysis of point mutations in the AAUAAA motif of the SV40 late polyadenylation signal." Nucleic Acids Res **17**(10): 3899-908.

Wolf, P. G., C. A. Rowe, et al. (2004). "High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern Adiantum capillus-veneris." Gene **339**: 89-97.

Wong, J. T., D. C. New, et al. (2003). "Histone-like proteins of the dinoflagellate Crypthecodinium cohnii have homologies to bacterial DNA-binding proteins." Eukaryot Cell **2**(3): 646-50.

Yehudai-Resheff, S., M. Hirsh, et al. (2001). "Polynucleotide phosphorylase functions as both an exonuclease and a poly(A) polymerase in spinach chloroplasts." Mol Cell Biol **21**(16): 5408-16.

Yoon, H. S., J. D. Hackett, et al. (2002). "The single, ancient origin of chromist plastids." Proc Natl Acad Sci U S A **99**(24): 15507-12.

Yoon, H. S., J. D. Hackett, et al. (2005). "Tertiary Endosymbiosis Driven Genome Evolution in Dinoflagellate Algae." Mol Biol Evol **22**(5): 1299-1308.

Yu, W. and W. Schuster (1995). "Evidence for a site-specific cytidine deamination reaction involved in C to U RNA editing of plant mitochondria." J Biol Chem **270**(31): 18227-33.

Zauner, S., D. Greilinger, et al. (2004). "Substitutional editing of transcripts from genes of cyanobacterial origin in the dinoflagellate Ceratium horridum." FEBS Lett **577**(3): 535-8.

Zerges, W. (2000). "Translation in chloroplasts." Biochimie **82**(6-7): 583-601.

Zerges, W. (2002). "Does complexity constrain organelle evolution?" Trends Plant Sci **7**(4): 175-82.

Zhang, L., V. Paakkarinen, et al. (2000). "Biogenesis of the chloroplast-encoded D1 protein: regulation of translation elongation, insertion, and assembly into photosystem II." Plant Cell **12**(9): 1769-82.

Zhang, Z., T. Cavalier-Smith, et al. (2002). "Evolution of dinoflagellate unigenic minicircles and the partially concerted divergence of their putative replicon origins." Mol Biol Evol **19**(4): 489-500.

Zhang, Z., B. R. Green, et al. (1999). "Single gene circles in dinoflagellate chloroplast genomes." Nature **400**(6740): 155-9.

Zhang, Z., B. R. Green, et al. (2000). "Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: a possible common origin for sporozoan and dinoflagellate plastids." J Mol Evol **51**(1): 26-40.