2004-12-02

Identification of Proteins Involved in Chloroplast DNA Replication

Matthew G. Lassen
Brigham Young University - Provo

IDENTIFICATION OF PROTEINS INVOLVED IN CHLOROPLAST DNA REPLICATION

by

Matthew Gordon Lassen

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Microbiology & Molecular Biology

Brigham Young University

December 2004

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Matthew G. Lassen

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	Brent L. Nielsen, Chair
Date	William R. McCleary
Date	Craig E. Coleman

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Matthew G. Lassen in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures and tables are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Brent L. Nielsen
Chair, Graduate Committee

Accepted for the Department

Alan R. Harker
Chair, Department of Microbiology and Molecular Biology

Accepted for the College

R. Kent Crookston
Dean, College of Biology and Agriculture

ABSTRACT

IDENTIFICATION OF PROTEINS INVOLVED IN

CHLOROPLAST DNA REPLICATION

Matthew Gordon Lassen

Department of Microbiology & Molecular Biology

Master of Science

Chapter 1

Chloroplast nucleoids (ct-nucleoids) are DNA/protein complexes involved in compacting the chloroplast genome, and may play a role in regulating DNA replication. Ct-nucleoids were isolated from young soybean plants and separated by 2-D gel electrophoresis. Gel spots were excised and analyzed by MALDI-ToF mass spectrometry, resulting in several protein identifications. The proteins identified all have functions unrelated to DNA replication. While some of these proteins may be due to contamination, it is possible that some of these proteins are dual-functional, playing direct roles in the regulation of DNA replication.

Chapter 2

A 28 kDa soybean protein was isolated by sequence specific DNA affinity chromatography from total chloroplast protein isolations. Mass spectrometry analysis revealed that the 28 kDa protein contains some homology within an ssb domain of an

Arabidopsis mitochondrial-targeted SSB (mtSSB) of approximately 21 kDa. N-terminal sequencing revealed that the 28 kDa soy protein is identical to a 36 amino acid region at the N-terminus of the *Arabidopsis* mtSSB. Protein fractions containing the 28 kDa protein shift *oriA* in electrophoretic mobility shift assays (EMSAs). *Arabidopsis* mtSSB fails to shift *oriA* in EMSAs run under identical conditions. *Arabidopsis* mtSSB causes a shift of ssDNA in EMSAs, while the ability of the 28 kDa soy protein to bind ssDNA is still unclear. Importantly, the 28 kDa soy protein was identified from total protein extracts obtained from intact chloroplasts, while in-vitro targeting experiments suggest that the *Arabidopsis* mtSSB localizes only to mitochondria and not to chloroplasts. BLAST searches of the available soybean genomic and EST databases do not produce any significant homologies to the 36 amino acid N-terminal sequence.

ACKNOWLEDGEMENTS

Many people have been crucial in completing this thesis. First of all, I would like to thank Dr. Brent Nielsen for the opportunity he gave me to work in his lab, as well as for all of his patience and good advice. Drs. Bill McCleary and Craig Coleman have also provided important advice and guidance throughout the project. Dr. Craig Thulin and Katie Southwick assisted with the 2-D gels and mass spectrometry. Melisa Wall performed the targeting experiments. Andrew Edmondson provided valuable information and data concerning the *Arabidopsis* mtSSB. Luis Alvarez assisted in some of the electrophoretic mobility shift assays. Johnathon Overson assisted in various aspects throughout the project. Most of all, I would like to thank my beautiful wife, Erika, and my three children, Colby, Dallin, and Angela, who are always my biggest fans. This work was supported in part by USDA/NRICGP, by a BYU Mentoring Grant, and by funding from the Department of Microbiology and Molecular Biology.

TABLE OF CONTENTS

ABSTRACT……...iv

ACKNOWLEDGEMENTS……vi

LIST OF FIGURES/TABLE……...viii

BACKGROUND……1

CHAPTER 1: IDENTIFICATION OF DNA REPLICATION PROTEINS ASSOCIATED WITH THE CHLOROPLAST NUCLEOID
- INTRODUCTION……...9
- MATERIALS AND METHODS……...11
- RESULTS……..17
- DISCUSSION...……19

CHAPTER 2: IDENTIFICATION OF A CHLOROPLAST DNA REPLICATION INITIATION PROTEIN
- INTRODUCTION…….....22
- MATERIALS AND METHODS……...25
- RESULTS……..32
- DISCUSSION...……36

REFERENCES……..58

LIST OF FIGURES/TABLE

Figure 1: Model of ctDNA replication mechanism……..…..41
Figure 2: Sequence and structure of chloroplast *oriA* and *oriB*……42
Figure 3: Detection of *ori*-binding proteins via southwestern blot……...……....43
Figure 4: Two-dimensional gel of isolated nucleoid proteins, pH 4-7…….……....44
Figure 5: Mass spectrometry MS spectra of nucleoid proteins…….……..…45
Figure 6: Southwestern blot of nucleoid proteins separated by 2-D gel, pH 4-7…….....47
Figure 7: *OriA* EMSA of DEAE cellulose column fractions……....48
Figure 8: SDS-PAGE of *oriA* affinity column fractions……...…49
Figure 9: SDS-PAGE of affinity control column fractions…….....50
Figure 10: Mass spectrometry MS spectra of 28 kDa soy protein……..….51
Figure 11: Amino acid sequence of *Arabidopsis* mtSSB…….....52
Figure 12: Two-dimensional gel of *oriA* affinity fraction #4……..…53
Figure 13: *Arabidopsis* mtSSB targeting experiment to mitochondria……..…..54
Figure 14: *Arabidopsis* mtSSB targeting experiment to chloroplasts……..…55
Figure 15: *OriA* EMSA with *oriA* affinity column fractions and *Arab*. MtSSB.……56
Figure 16: ssDNA EMSA with 28 kDa protein and *Arabidopsis* mtSSB……...57
Table 1: Mass spectrometry protein identifications of nucleoid proteins…….....46

Background

Chloroplasts

Chloroplasts are the site of photosynthesis, resulting in carbon fixation and the production of energy and other metabolic processes vital to the plant's survival. Enclosed by a double membrane, chloroplasts contain DNA separate from the plant nucleus. Chloroplast division and DNA replication occur independent of nuclear DNA replication and cell division. Chloroplast DNA replication is highly active during the early stages of plant growth, independent of chloroplast transcription, translation, and division. During this growth stage, fully developed chloroplasts can contain 200-300 copies of the DNA, with each plant cell containing between 20 and 60 chloroplasts resulting in up to 10,000-12,000 genome copies per plant cell (reviewed in Kunnimalaiyaan and Nielsen, 1997a). Once the plant begins to mature, replication slows while chloroplast and cell division continue, resulting in segregation of the ctDNA into new chloroplasts, which are distributed among new cells. This results in a much lower ctDNA copy number in mature plant cells. However, the mechanism by which ctDNA replication and copy number are regulated is not known.

It is generally accepted that chloroplasts are the result of an endosymbiotic relationship between an early eukaryotic cell and a bacterial cell of cyanobacterial origin (Clegg et al., 1994; Dyall et al., 2004; Sheveleva and Hallick, 2004). Most chloroplast proteins are encoded in the nucleus, but the amino acid sequence of nuclear encoded chloroplast proteins is much more similar to bacterial than eukaryotic counterparts. The genetic systems of chloroplasts and bacteria are in many instances identical. Chloroplast ribosomes are indistinguishable from bacterial ribosomes and are inhibited by

antibacterial agents. Bacteria therefore provide an important model for studying the genetic systems of chloroplasts.

DNA Replication Proteins

Replication of DNA involves the cooperative interaction of a number of proteins. These proteins include the following:

- DNA polymerase, involved in the actual synthesis of the new DNA strand,
- Initiation protein that binds to an origin of replication and recruits subsequent replication proteins, similar to the *E. coli* DnaA protein,
- Helicase, which unwinds the two DNA strands making them accessible for replication,
- Primase, which synthesizes a short RNA primer that allows DNA polymerase to initiate strand elongation,
- Single-stranded binding proteins that coat the unwound DNA, protecting it from digestion and blocking reannealing, as well as attracting other replication proteins,
- Topoisomerases, whose function results in relieving the tension in the DNA caused by DNA unwinding.

Replication in chloroplasts is likely to involve proteins with functions homologous to those listed above. However, few proteins associated with ctDNA replication have been isolated. DNA polymerases have been isolated from chloroplasts in pea (*Pisum sativum*) and rice (*Oryza sativa*) (Gaikward et al., 2002; Kimura et al., 2002). Chen et al. (1996) reported a 43 kDa DNA-binding protein from pea chloroplasts that interacts with the pea ctDNA polymerase. What role the 43 kDa protein may play in ctDNA replication is still unclear.

In *E. coli*, the DnaA initiation protein binds to five 9 bp repeats in the *ori* region, stimulating the initial unwinding of the DNA and allowing subsequent proteins to bind and begin replication (Messer, 2002). Interestingly, several 9 bp repeats, with little homology to the repeats recognized by DnaA, have been found in the tobacco ctDNA *oriB* region (Kunnimalaiyaan et al. 1997), suggesting that an initiation protein with a similar function to DnaA may be present in chloroplasts. However, homologues of the *E. coli* proteins DnaA, DnaB, DnaC, and DnaG do not appear to be encoded within the *Arabidopsis* genome, suggesting that the regulatory proteins for the initiation of ctDNA replication may be unique to chloroplasts (our lab's unpublished results). Genes encoding proteins involved in replication have also not been reported within the sequenced chloroplast genomes, suggesting that any such proteins must be nuclear-encoded.

CtDNA Replication Mechanism

The DNA in chloroplasts of higher plants consists of a covalently-closed circular genome of 120-160 kilobases (kb). The number and sequence of genes in ctDNA, as well as the presence of two large inverted repeats (IR), is highly conserved among most plants. Replication in a variety of systems has been shown to begin at specific sites referred to as origins of replication (*ori*). These *ori* regions contain binding sites for proteins that initiate the replication process. At least two *ori* regions (*oriA* and *oriB*) are found in ctDNA and the location of these *oris* has been identified in several plant species, including pea, tobacco, Oenothera, maize, and soybean (reviewed in Kunnimalaiyaan and Nielsen, 1997a). Electron microscopy has led to the observation of D-loops (displacement loops in the DNA indicating a separation of the two strands with DNA

replication on one strand) from these *oris*, with one D-loop at each *ori*. According to the accepted model, the D-loops then proceed toward each other in a unidirectional manner until the two D-loops meet, at which point replication continues bi-directionally around the molecule until the two replication forks intercept each other (reviewed in Kunnimalaiyaan and Nielsen, 1997b) (Figure 1).

In our laboratory, extensive work has been done in tobacco chloroplasts to identify the location of *oriA* and *oriB*, as well as the minimal sequences needed to support replication. In tobacco chloroplasts, *oriA* and *oriB* flank the 23S rRNA gene within the inverted repeat. A recent study suggests that replication may not require a functional *oriA* region. Deletion/insertion studies of tobacco *oriA* and *oriB* by Muhlbauer et al. (2002) suggest that replication may be achieved without a functional *oriA*, but that *oriB* is essential for replication. However, in vitro replication analysis of minimal sequences for single-*ori* clones has shown that a 82 bp *oriA* template and a 243 bp *oriB* template is sufficient to support replication activity (Lugo et al., 2004; Kunnimalaiyaan and Nielsen, 1997a). Both *oriA* and *oriB* contain imperfect indirect repeats that potentially could form stem loop structures as well as direct repeats that are essential to replication, as replication activity ceases when these stem loop regions or direct repeats are deleted (Lugo et al., 2004; Kunnimalaiyaan and Nielsen, 1997a; Lu et al., 1996) (Figure 2). Large shifts were obtained when these minimal sequences were used in gel mobility shift assays to observe interactions between proteins and chloroplast *oris* (Kunnimalaiyaan et al. 1997). This observation further supports the importance of these repeated elements in ctDNA replication. Importantly, these repeated elements may be binding sites for specific replication initiation proteins. Southwestern blot analysis performed previously

in our lab has identified two tobacco chloroplast proteins that bind specifically to tobacco chloroplast *oriA* and *oriB* (Kochhar and Nielsen, unpublished results) (Figure 3).

CtDNA Replication in *Chlamydomonas*

Regulation of ctDNA replication has been best described in the green alga, *Chlamydomonas reinhardtii*. Two chloroplast *oris* have been identified in *C. reinhardtii*. In contrast to the plant chloroplast *oris*, the *C. reinhardtii* chloroplast *oris* are found in a single copy region of the genome. One chloroplast origin of replication, *oriA*, has been focused on for characterization. *OriA* is A + T rich and contains an open reading frame (ORF) within the D-loop, where replication initiates, as well as two back-to-back prokaryotic-like promoters (Wu et al. 1986). This ORF was later identified as a gene encoding the ribosomal protein Rpl16 (Lou et al. 1987). Within *oriA* are bent regions of DNA that are important in initiating replication (Hsieh et al. 1991). One *oriA*-binding protein with a molecular weight of 18 kDa was identified in *C. reinhardtii*. This protein was found to be an iron-sulfur protein that is related to a subunit of NADH dehydrogenase (Wu et al. 1989). Later work suggested that the redox state of this protein may play a role in regulating the initiation of replication (Wu et al. 1993).

DNA Replication of Animal Mitochondrial Genomes

As mentioned above, both chloroplasts and mitochondria are thought to be the result of an endosymbiotic relationship. Although there are distinct differences between these two organelles, it is possible that similarities may exist in the replication of their genomes, and thus the data available for mitochondrial DNA (mtDNA) replication is worth noting. Replication within plant mitochondria is controversial and not as well understood as in animal mitochondria. MtDNA in animals and plants differ greatly in

size, (plant mitochondrial genomes are 10-150 times larger than animal mitochondrial genomes) and plant mtDNA contain one or more large repeats that are believed to be the result of homologous recombination, so it is possible that genome maintenance may differ significantly in organelles between plants and animals (Backert et al., 1996).

Like chloroplasts, animal mtDNA contains two origins of replication (O_H and O_L). Replication initiates first at O_H (leading strand), and initiates at O_L (lagging strand) only after the replication fork of the leading strand passes O_L (Moraes, 2001). Replication initiation of animal mtDNA requires an RNA primer laid down by an RNA polymerase. The mitochondrial transcription factor A (TFA) also plays an essential role in the priming step of initiation. Priming of the origin allows DNA polymerase γ to proceed with strand elongation. DNA polymerase γ consists of two subunits: A (the catalytic subunit) and B (the accessory subunit). Rate of synthesis by *Drosophila* DNA pol γ is stimulated 40-fold by the addition of mtSSB (Moraes, 2001). This stimulation appears to result from increased primer recognition, binding, and rate of initiation. Control of mtDNA copy number is still unclear. Analysis of expression levels shows no correlation of copy number with the expression of the identified replication factors, with the exception of the accessory subunit of polymerase γ. There is some evidence to suggest that the expression of pol γ-B is associated with an increase of mtDNA replication, but further work is needed to confirm and elucidate any possible control mechanisms this may involve (Moraes, 2001). It is also thought that mtDNA copy number may be regulated by the limited expression of one of the replication factors listed above, as all are required for replication. Support for this hypothesis comes from the observation that some mtDNA molecules replicate twice before other molecules are

replicated even once, and also that replication seems to be favored in those organelles positioned closer to the nucleus (Bogenhagen and Clayton, 1977; Davis and Clayton, 1996).

Protein Sorting

The majority of proteins that function in the chloroplast are encoded in the plant nucleus. Therefore, these proteins must be transported, following protein synthesis in the cytoplasm, to the surface of the chloroplast membrane. This protein sorting requires a target sequence on the protein that is recognized by specific receptors on the chloroplast membrane; these receptors then transport the protein across the membrane into the chloroplast, usually cleaving the target sequence in the process (Jarvis and Soll, 2001). Proteins targeted to the mitochondria undergo a similar import process. Additional target sequences may be present on the protein for further localization within the organelle. Some targeting sequences can be predicted using one of several prediction analysis tools (e.g. ChloroP 1.1 Prediction Server, http://www.cbs.dtu.dk/services/ChloroP/; Emanuelsson et al., 1999). However, a recent study of total mitochondrial proteins showed that only about 50% of mitochondrial proteins identified via mass spectrometry were predicted to localize to the mitochondria (Heazlewood et al, 2004). Therefore, while these programs may provide some direction as to the identification of organelle-targeted proteins, it appears that the predictions are limited. Confirmation of organelle localization within plant cells has been achieved via green fluorescent protein (GFP)-fusion constructs containing the putative target peptide (Khazi et al., 2003).

Thesis Objectives

Chloroplasts display a replication process that is unique when compared to other systems. The genome copy number is extremely high during the early stages of plant development. The proteins involved in ctDNA replication and the maintenance of the high copy number have yet to be identified. I sought to identify ct-targeted proteins involved in ctDNA replication by two different approaches. First, I isolated chloroplast nucleoids (ct-nucleoids) and identified some of the associated proteins via mass spectrometry. Second, I identified a chloroplast origin-binding protein using sequence-specific DNA affinity chromatography. These approaches are described in detail in the chapters that follow.

Chapter 1

Identification of DNA Replication Proteins Associated With the Chloroplast Nucleoid

Introduction

The length of DNA in a particular cell is usually many times the size of the cell itself and therefore must be highly compacted to fit within the cell. Specific basic proteins that are tightly associated with the DNA and result in complex folding of the genome accomplish this compacting (Cannon et al., 1999). In chloroplasts, protein/DNA complexes that appear to be involved in the organization of ctDNA are called nucleoids (Nakano et al., 1993). Approximately 20-25 proteins co-purify with the chloroplast nucleiod; however, little is known about these nucleoid-associated proteins (Nemoto et al., 1989). Nucleoids are observed throughout chloroplast development, while the size, number, and structure of nucleoids appear to change depending on the stage of chloroplast development (Hashimoto, 1985). It is thought that nucleoids may play a role in regulating gene expression and DNA synthesis, and that changes in nucleoid structure are correlated with the varying levels of both transcription and replication (Cannon et al. 1999). To date, seven nucleoid-associated proteins have been identified, with two potentially being the same protein. The 130 kDa plastid envelope DNA-binding (PEND) protein is involved in anchoring the ctDNA to the envelope membrane within developing plants (Sato et al. 1998). A chloroplast homolog to the ribosomal protein L19 of *Saccharomyces cerevisiae* was identified from nucleoid isolations and was named P28 (Oleskina et al. 1999). CND41, a 41 kDa nucleoid-associated protein that displays both DNA-binding abilities and proteolytic activity, is thought to play a role in the regulation

of ctDNA gene expression (Murakami et al., 2000). A 70 kDa sulfite reductase protein associated with pea chloroplast nucleoids appears to have a role in transcription regulation through a DNA compacting activity (Sato et al., 2001; Sekine et al., 2002). Interestingly, a 68 kDa ferredoxin:sulfite reductase from soybean chloroplasts was also found to associate with the chloroplast nucleoid and have a role in DNA compacting (Chi-Ham et al., 2002). MFP1 is a 82 kDa protein thought to play a role in anchoring the ctDNA to the thylakoid membrane within mature plants (Jeong et al,. 2003). Finally, a chloroplast-encoded HU-like protein has been indicated as having a role in DNA compacting within chloroplasts of the primitive red alga *Cyanidioschyzon merolae* (Kobayashi et al., 2002). The observed activities of the proteins identified from chloroplast nucleoid preparations support their role in the organization of the chloroplast genome, and suggest that nucleoids may play a major role in regulating gene expression and DNA synthesis through DNA remodeling (Sato et al., 1997). A high level of replication activity has also been observed with the proteins that co-purify with chloroplast nucleoids (Cannon et al., 1999). It is possible that proteins involved in regulating and initiating ctDNA replication associate with the ct-nucleoid during the early stages of plant development**.** The objective of this research was to separate nucleoid-associated proteins by two-dimensional gel electrophoresis and identify individual protein spots by MALDI-ToF mass spectrometry.

Materials & Methods

Isolation of intact chloroplasts

Intact chloroplasts were prepared according to the protocol outlined by Bartlett et al. (1982), with slight modifications. All centrifugation steps were performed at 4°C with no braking. 200-300 grams of leaf tissue from 8-9 day old soybean plants were homogenized in cold chloroplast grinding buffer (2 mM Na EDTA, 1 mM $MgCl_2$, 1 mM $MnCl_2$, 50 mM Hepes-KOH, pH 7.5, 330 mM sorbitol, 5 mM ascorbic acid (ascorbic acid was added fresh immediately before use)) in a chilled Waring blender containing a single razor blade using three 3-5 sec bursts on low setting and two 3-5 sec bursts on high setting. The homogenate was filtered through 4 layers of cheesecloth (drained with squeezing) and 3 layers of miracloth (Calbiochem) (drained without squeezing), and centrifuged to 4229 x g (Beckman JA-14 rotor). The drive was turned off immediately upon attainment of this speed. The pellet was resuspended in 2-3 ml of chloroplast grinding buffer and loaded onto a 40% [5.25 ml PBF Percoll (3% polyethylene glycol 3350, 1% BSA, and 1% Ficoll; adjust volume to 20 ml with Percoll (Amersham Biosciences)), 2 mM EDTA, pH 8.0, 1 mM MgCl2, 1 mM MnCl2, 50 mM Hepes, pH 7.5, 6% (w/v) sorbitol, .098% (w/v) ascorbic acid, .017% (w/v) glutathione; dH2O to 13 ml] to 85% (same as 40% but with 11.16 ml PBF Percoll) Percoll gradient (11.5 ml 85% and 12.5 ml 40%). The Percoll gradient was centrifuged at 14600 x g (Beckman JS-13.1 rotor) for 15 minutes. Intact chloroplasts were retrieved from the bottom band, just above the 40%/85% interface, and washed in chloroplast grinding buffer to remove excess percoll by centrifuging at 4300 x g (Beckman JS-13.1 rotor) for 5 minutes. Chloroplasts were washed a second time in chloroplast grinding buffer by centrifuging to 4300 x g

(Beckman JS-13.1 rotor) and immediately turning off the drive. The supernatant was decanted and chloroplast pellets were resuspended in a minimal volume of either chloroplast grinding buffer or nucleoid isolation buffer and kept on ice.

Isolation of chloroplast nucleoids

Chloroplast nucleoids were isolated according to the protocol outlined by Cannon et al. (1999), with slight modifications. All centrifugation steps were performed at 4°C with no braking. Intact chloroplast pellets were resuspended in 20-40 ml of nucleoid isolation buffer (20 mM Tris-HCl, pH 7.0, 0.5 mM EDTA, pH 8.0, 1.5 mM spermidine, 7 mM β-mercaptoethanol, 0.4 mM phenylmethyl sulfonyl fluoride (PMSF)). While stirring, Nonidet P-40 was slowly added to a final concentration of 2% (v/v). Chloroplasts were lysed by stirring at room temperature for 30 minutes. Lysed chloroplasts were centrifuged at 46,000 x g (Beckman JA-20 rotor) for 30 minutes. The pellet was resuspended in 20 ml isolation buffer/2% Nonidet P-40 and centrifuged at 46,000 x g (Beckman JA-20 rotor) for 30 minutes. The pellet was resuspended in 20 ml isolation buffer/2% NP-40, divided into two equal fractions, layered onto two sucrose gradients (15 ml of 40% and 10 ml of 80% step) and then centrifuged at 8000 rpm (Beckman JS-13.1 rotor) for 15 minutes. Nucleoids were harvested from the interface between the 40% and 80% steps. Nucleoids were diluted with nucleoid isolation buffer and centrifuged at 46,000 x g (Beckman JA-20) for 30 minutes. Nucleoids were resuspended in a minimal volume of nucleoid isolation buffer and stored at -80°C.

Two-dimensional gel electrophoresis

Nucleoid proteins (175-300 μg) were prepared for 2-D gel electrophoresis with the ReadyPrep 2-D Cleanup Kit (Bio-Rad) according to the manufacturer's instructions,

resuspending the final pellet in 125 µl of sample rehydration buffer (8 M urea, 2% (w/v) CHAPS, 1% (v/v) appropriate IPG buffer, and a few crystals of bromphenol blue dye). Proteins were rehydrated onto IPG strips (Amersham Pharmacia Biotech, pH4-7 and pH6-11) and resolved on an Amersham Multiphor II Elecrophoresis System. Resolved strips were incubated on a rocker in 5 ml equilibration buffer (50 mM Tris-Cl, pH 6.8, 6 M urea, 30% (v/v) glycerol, 1% (w/v) SDS) containing 32.4 mM dithiothreitol for 10-20 minutes at room temperature. IPG strips were then incubated on a rocker in 5 ml equilibration buffer containing 1 mM iodoacetamide and a few grains of bromphenol blue for 10-20 minutes. Proteins from the IPG strips were then separated on 10% acrylamide gels at 30 mA until the dye reached the bottom of the gel. Gels were stained with BioRad's BioSafe Coomassie stain or transferred to a PVDF membrane for southwestern blot analysis.

MALDI-ToF Mass Spectrometry

Protein spots for mass spectrometry analysis were excised from 2-D gels under sterile conditions, using sterile razor blades, and transferred to clean 1.5 ml microcentrifuge tubes. Excised gel spots were stored at -20°C. In-gel tryptic digests of protein spots were performed according to Shevchenko et al. (1996), with slight modifications. 1.5 ml of 50% acetonitrile/50 mM ammonium bicarbonate was added to each tube containing a gel spot. Samples were placed into a micro mixer and rotated for 30 minutes at room temperature. The liquid was then aspirated off, and 140 µl of 100 mM ammonium bicarbonate was added to the samples, followed by 10 µl of reduction buffer (75 mM dithiothreitol, 100 mM ammonium bicarbonate). Samples were incubated with no motion at 60°C for 45 minutes. Samples were then removed from the heat block,

cooled to room temperature, and solutions were aspirated off. 140 µl of 100 mM ammonium bicarbonate was added to the samples, followed by 10 µl of alkylation buffer (0.3 M iodoacetamide, 100 mM ammonium bicarbonate). Samples were incubated in the dark at room temperature for 30 minutes. Solutions were then aspirated off. 1 ml of 100 mM ammonium bicarbonate was added to each gel spot and tubes were rotated at room temperature for 30 minutes. Solutions were removed as before. Under sterile conditions, gel spots were cut 1-2 times (depending on size of gel spot) with a sterile razor blade and transferred to a clean 1.5 ml microcentrifuge tube. 500 µl of 50% acetonitrile/50 mM ammonium bicarbonate was added to each sample tube and rotated at room temperature for 30 minutes. Solutions were removed as before. Gel pieces were dried to completeness in a dryer. 10 µl of trypsin digestion solution (0.02 mg/ml trypsin in 25 mM ammonium bicarbonate) was added directly to each gel sample and incubated for 10 minutes at room temperature. This step was repeated until the gel pieces were completely hydrated with trypsin digestion solution. Enough 25 mM ammonium bicarbonate (usually 10-20 µl) was added to the samples to have a 0.5 mm excess of solution above the gel pieces. Samples were incubated at 37°C in a hybridization oven with rocking for 16-24 hours. Following incubation, 1 µl of 88% formic acid was added to the samples to stop the trypsin reaction. Samples were sonicated for 20 minutes in a water bath and then centrifuged briefly. Samples were then concentrated and desalted using ZipTips (Millipore). Mass spectrometry analysis was performed on an API QSTAR Pulsar I from Applied Biosystems. Protein identifications were made using the nonredundant NCBI protein database.

Biotin-labelling of 144 bp *oriA* PCR product

A 144 bp *oriA* region from tobacco ctDNA was amplified by polymerase chain reaction (Lugo et al., 2004). Each 50 µl reaction tube contained 0.5 µM of each primer (OA1F and OA144R), 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 1X PCR reaction buffer, 2.5 units of Taq polymerase (Invitrogen), and 1 µl DNA template. A control reaction containing no DNA template was also run. All PCR reactions were run on a PTC-150 MiniCycler (MJ Research) under the following conditions: 95°C for 3 minutes, 30 cycles of 95°C for 1 minute, 50°C for 1 minute, and 72°C for 1 minute, followed by 72°C for 7 minutes. A 5 µl aliquot of each reaction tube was loaded onto a 1.2% agarose gel and run at 90V for about 45 minutes. *OriA* PCR products from the remaining 45 µl of each reaction were purified using the QIAquick PCR purification kit (Qiagen) according to the manufacturer's instructions. Purified *oriA* was 3'end-labelled with biotin using the Biotin 3'end DNA Labeling Kit (Pierce) according to the manufacturer's instructions. All samples were stored at -20°C.

Southwestern blot

Protein samples were separated by either SDS-PAGE or 2-D gel electrophoresis. After electrophoresis, the gel was equilibrated in blotting buffer (25mM Tris-HCl, pH 8.3, 190mM glycine, 20% methanol, 0.01% SDS) for 30 minutes and proteins were transferred onto a PVDF membrane overnight at 90 mA in a tank transfer apparatus (Bio-Rad) with stirring and an ice pack. The blot was incubated in renaturation buffer (10mM Hepes, pH 7.5, 100mM KCl, 1mM DTT, 0.1mM EDTA, 10mM MgCl2, 0.1mM ZnSO4, and 5% nonfat milk) at 4°C for 18-20 hours. The blot was washed two times with TNE-50 buffer (10mM Tris-HCl, pH 7.5, 50mM NaCl, 1mM EDTA, 1mM DTT) for 15

minutes. The blot was then incubated in TNE-50 buffer containing biotin-labelled 144 bp *oriA* and nonspecific competitor poly(dI-dC) (10 μg/ml) for 6 hours at room temperature. The blot was washed three times with TNE-50 buffer at room temperature for 30 minutes. Biotin-labelled probe was detected using the Phototope-Star Detection Kit for Nucleic Acids (New England BioLabs) according to the manufacturer's instructions.

Results

SDS-PAGE of Isolated Nucleoids

Isolated nucleoids were first analyzed by one-dimensional SDS-PAGE (data not shown). The major protein bands were excised and identification by MALDI-ToF mass spectrometry was attempted. From these preparations, ribulose-bisphosphate carboxylase (Rubisco) was the only identification made. Since Rubisco is a major enzyme of chloroplast extracts, I determined that its presence was due to contamination. I further concluded that the lack of any other identifications may be due to the presence of more than one protein in the excised bands, as this would hinder proper identification. Isolated nucleoids were then separated by two-dimensional gel electrophoresis.

Two-dimensional Gel Electrophoresis

Nucleoids were separated by two-dimensional gel electrophoresis using IPG strips with pH gradients 4-7 and 6-11. The majority of the nucleoid proteins resolved more successfully in the pH 4-7 range (Figure 4) than in the pH 6-11 range (data not shown). The pattern of separation in the pH 4-7 range also indicated the presence of post-translational modifications in at least three of the proteins (indicated by numbers in Figure 4). Protein spots were excised for identification by mass spectrometry.

MALDI-ToF Mass Spectrometry

An example of the MS data for ferredoxin:sulfite reductase obtained by MALDI-ToF mass spectrometry is shown in Figure 5. Positive identifications of excised protein spots are listed in Table 1. Protein spots labeled in Figure 4 correspond to the protein list in Table 1. Identification of the protein ferredoxin:sulfite reductase as a nucleoid-associated protein was made previously by Chi-Ham et al. (2002). The other

identifications were mostly enzymes involved in various functions within the chloroplast. These enzymes could potentially be dual functional or present due to contamination of other chloroplast proteins not associated with nucleoids.

Detection of *OriA*-binding Nucleoid Proteins by Southwestern Blot

Southwestern blots of nucleoid proteins suggested that nucleoid complexes may contain proteins that bind specifically to *oriA* (Figure 6). The eight signals observed on the blots corresponded to approximate molecular weights of 70, 65, 31, 31, 30, 30, 30, and 26 kDa (Figure 6). Data shown represent proteins separated by two-dimensional gel electrophoresis in the pH 4-7 range. The pattern of the signals suggests that the signals corresponding to molecular weights of 31 and 30 kDa may represent the same protein or proteins with post-translational modifications.

Discussion

Several protein identifications were made from chloroplast nucleoid isolations. One of these proteins was ferredoxin:sulfite reductase, which was identified previously as a nucleoid-associated protein (Chi-Ham et al., 2002). The remaining identifications are listed in Table 1. None of these proteins were readily identifiable as proteins known or suspected as playing a direct role in DNA compacting or replication. Preliminary southwestern blot results indicated that some of these proteins in the nucleoid isolations may interact specifically with the chloroplast *oriA* (Figure 6). Signals corresponding to molecular weights of approximately 30 kDa suggest the presence of *ori*-binding proteins in the same size range as previous experiments had identified (Kochhar and Nielsen, unpublished results) (Figure 3). It is possible that some of the identified nucleoid-associated proteins may have regulatory functions in DNA replication. Multifunctional proteins, particularly in bacteria, are being identified at an increasing pace (Jeffery, 2003). Zhu et al. (2002) recently demonstrated that Clf1p, previously known to function in pre-mRNA splicing, plays an important role in initiating replication bubbles. Gonzalez et al. (2002) showed that the base complex of the 19S particle associated with the proteasome also plays a role in transcription by RNA polymerase III. Fumarate hydratase, an enzyme in the Krebs cycle, was recently identified as a tumor suppressor gene (Tomlinson et al., 2002). Thus, some of the proteins identified here from nucleoid preparations may be dual-functional proteins. Indeed, characterizations of nucleoid-associated proteins identified to date suggest that the ct-nucleoid is involved in the regulation of both replication and transcription (Chi-Ham et al., 2002; Sekine et al., 2002;

Murakami et al., 2000; Sato et al., 1998). Whether this regulatory role is direct or indirect has yet to be shown.

The two-dimensional separation patterns suggest that some of the proteins associated with the ct-nucleoid may contain posttranslational modifications (Figure 4). Since ct-targeted proteins do not pass through the Golgi or Endoplasmic Reticulum (ER), these modifications are most likely not the addition of sugars, but could be methylations, acetylations, or phosphorylations. It is possible that such modifications could play a role in functional regulation. For example, the Swi/Snf complex is required for the expression of specific genes immediately following mitosis in yeast cells (Ishov et al., 2004). Phosphorylation of Swi/Snf is thought to play a role in regulating its activity. In prokaryotes, phosphorylation of the *Caulobacter crescentus* protein CtrA inhibits DNA replication (Ryan and Shapiro, 2003). DNA replication is only initiated following CtrA proteolysis. Similar modifications may play a role in regulating the initiation of ctDNA replication.

Due to the lack of a consensus in isolating ct-nucleoids, one of the limitations of identifying and characterizing nucleoid-associated proteins is the difficulty of determining the purity of the nucleoid proteins. It may be possible to obtain a more pure isolation of nucleoids by using a co-immunoprecipitation technique. Cannon et al. (1999) demonstrated that dissociated ct-nucleoid proteins can reassociate in vitro by dialysis, and that the reformed nucleoid complexes resemble nucleoid complexes observed in intact chloroplasts. The nucleoid-associated ferrodoxin:sulfite reductase protein has been shown to have a central role in compacting the ctDNA (Chi-Ham et al., 2002). A construct of ferrodoxin:sulfite reductase with a c-myc tag could be engineered and over-

expressed. This construct could then be mixed with isolated nucleoid proteins that have been dissociated at high-salt conditions. Dialyzing the mixture against a low-salt buffer could then reform nucleoid complexes, allowing intact nucleoids to be isolated by co-immunoprecipitation using an antibody against the c-myc tag. Such a procedure would theoretically result in a highly purified preparation of nucleoid proteins.

Results presented in this thesis as well as the evidence for regulatory roles by characterized nucleoid-associated proteins merit further work on the role that chloroplast nucleoids may play in ctDNA replication (Chi-Ham et al., 2002; Sekine et al., 2002; Murakami et al., 2000; Sato et al., 1998).

Chapter 2

Identification of Chloroplast *oriA*-binding Proteins by Affinity Chromatography

Introduction

Initiation of DNA replication

Once the structure of DNA was determined, researchers soon began hypothesizing on the mechanism of replication. Jacob, Brenner, and Cuzin (1963) proposed a mechanism of initiation involving an initiator (a trans-acting protein) and a replicator (a cis-acting DNA element). Specific interactions between the initiator and the replicator were proposed to control the initiation of replication. Identification of initiators and replicators has been accomplished in a long list of both prokaryotic and eukaryotic organisms. Although the prokaryotic and eukaryotic processes differ in complexity and number of factors involved, initiation is achieved by the same basic mechanism. The replicator, or origin of replication (*ori*), contains binding sites that interact specifically with the initiator. Initiator-binding to the *ori* results in the initial unwinding of the *ori* region. The initiation protein may be a single protein (e.g. bacterial DnaA, RepA) or a complex of protein subunits that together initiate replication (e.g. eukaryotic Orc1-6), with most containing an ATPase domain, supporting the requirement for ATP hydrolysis in unwinding the *ori* region (Giraldo, 2003). Following unwinding, additional replication proteins involved in initiation are recruited to the *ori* region. In bacteria, these proteins include a helicase (DnaB) and a primase (DnaG) that together form the primosome (Messer, 2002). Priming of the DNA template allows DNA polymerase III to associate at the unwound *ori* and proceed with elongation of both the

leading and lagging strands. In eukaryotes, the process is much more complex and involves the coordinate action of cell cycle regulators interacting with multiple initiation complexes at multiple *oris* (for reviews see Lei and Tye, 2001, and Bryant et al., 2001).

Origin-binding proteins

Origin-binding proteins (OBPs) have been identified for a variety of replication systems. All characterized bacteria contain DnaA homoloques that function as the initiation protein. Many bacterial plasmids use Replication factor A (RepA) to control initiation, but also require DnaA binding (Messer et al., 1999). In eukaryotes, the origin recognition complex (ORC) is composed of 6 subunits (Orc1-6) that interact at the origins and initiate replication cooperatively with cell cycle regulators (Bryant et al., 2001). The OBP in archaea appears to be similar to the eukaryotic ORC, but in a simplified form (Giraldo, 2003).

Although there is a lack of sequence similarity between bacterial and eukaryotic OBPs, there are similarities in the types of domains they contain (Giraldo, 2003). An ATPases Associated with various cellular Activities (AAA+) domain for ATP-binding is shared among DnaA and some ORC subunits, but not RepA. ATP-binding is not required for origin-binding, but causes a conformational change within the OBP that is important in the initial unwinding of the *ori* region. OBPs contain DNA-binding domains common among many DNA-binding proteins. DnaA contains a C-terminal α-helical bundle, while RepA and Orc4 share winged-helix domains.

Two *oris* (*oriA* and *oriB*) have been identified in chloroplasts of higher plants (Kunnimalaiyaan and Nielsen, 1997a). Previous work in our lab has indicated that there are at least two chloroplast proteins, with molecular weights of approximately 20 kDa

and 30kDa, which bind specifically to both *oriA* and *oriB* (Kochhar and Nielsen, unpublished results) (Figure 3). The goal of this part of the project was to isolate an *ori*-binding protein from total chloroplast protein extracts via sequence-specific DNA affinity chromatography and identify the protein using MALDI-ToF mass spectrometry.

Materials & Methods

Total chloroplast protein isolation

All centrifugation steps were performed at 4°C with no braking. 200-250 grams of leaf tissue were harvested from 8-9 day old soybean plants and homogenized in STM buffer (0.5 M sucrose, 50 mM Tris-Cl, pH 8.0, 5 mM $MgCl_2$, 5 mM β-mercaptoethanol, 0.2 mM phenylmethyl sulfonyl fluoride) (~1/2 liter tissue in ~200 ml buffer) in a standard Waring blender with three 5 second low speed bursts and two 3 second high speed bursts. The homogenate was filtered through 4 layers of cheesecloth (drained with squeezing) and 3 layers of miracloth (drained without squeezing). The homogenate was centrifuged at 1,000 x g (Beckman JA-14 rotor) for 10 minutes. Chloroplast pellets were resuspended in STM buffer and combined with 10 ml of STM buffer/Triton X-100 (7.75 ml STM buffer and 2.25 ml Triton X-100). The volume was then adjusted to 50 ml with STM buffer. Chloroplasts were incubated on ice for 30 min, mixing gently 3-4 times during incubation. Lysed chloroplasts were centrifuged at 6,000 x g (Beckman JS-13.1) for 20 minutes. The pellets were discarded and the supernatant was saved and stored on ice.

Preparation of DNA affinity column

Preparation of *oriA*-coupled CNBr-activated Sepharose was performed according to Current Protocols in Molecular Biology (Ausubel et al. (Eds.), 1993, 12.10.6) with slight modifications. Three grams of CNBr-activated Sepharose 4B powder (Amersham Biosciences) was placed into a 15-mL polypropylene tube and hydrated with 10 mL of 1 mM HCl, mixing gently by flicking and inverting the tube. The slurry was transfered to a Coors ceramic funnel with Whatman paper #1 or a glass fiber filter and washed and

swelled by gradually pouring 500 mL of 1 mM HCl over the resin, and further washed with 100 mL of water and then with 100 mL of 10 mM potassium phosphate, pH 8.0. The resin was immediately transfered to a 15-mL polypropylene tube and 10 mM potassium phosphate, pH 8.0, was added (~1 ml) until the resin was a thick slurry. The purified 144 bp *oriA* (PCR product, as described in Chapter 1 of this thesis) was immediately added to the resin and incubated on a rotating wheel, end-over-end, at 25°C overnight (>8 hrs), transfered to a Coors ceramic funnel with Whatman paper #1 or a glass fiber filter, and washed with two 100 mL washes of water and one 100 mL wash of 1 M ethanolamine hydrochloride, pH 8.0. After transferring to a 15-mL polypropylene tube, 1 M ethanolamine hydrochloride, pH 8.0, was added (~1 ml) until the resin was a smooth slurry. The resin was incubated on a rotating wheel, end-over-end, for 2-4 hrs at room temperature, transfered to a Coors ceramic funnel with Whatman paper #1 or a glass fiber filter, and washed with 100 mL of 10 mM potassium phosphate, pH 8.0, 100 mL of 1 M potassium phosphate, pH 8.0, 100 mL of 1 M KCl, 100 mL of water, and 100 mL of column storage buffer (10mM Tris-Cl, pH 7.8, 1mM EDTA, pH 8.0, 0.3M NaCl, and 0.04% (w/v) sodium azide). The resin was transfered to a 15-mL polypropylene tube and stored in ~5 mL column storage buffer at 4°C.

DNA affinity chromatography

Total chloroplast proteins were loaded onto an equilibrated DEAE cellulose column (~70 cm^3 column volume) at 0.25 ml/min using a BioLogic LP chromatography system (Bio-Rad). The column was washed with 500 ml buffer A (50 mM Tris-Cl, pH 8.0, 10 mM β-mercaptoethanol, 0.2 mM phenylmethyl sulfonyl fluoride, 20% glycerol, 10 mM sodium metabisulfite, 10 mM benzamidine, 50 mM KCl) at 0.50 ml/min and then

eluted with 130 ml buffer A/0.6 M KCl at 0.33 ml/min. The eluate was collected in 4 ml fractions. *OriA* binding activity was detected via electrophoretic mobility shift assay (as described below). Active fractions were pooled and dialyzed in buffer Z (50 mM Tris-Cl, pH 7.8, 12.5 mM $MgCl_2$, 1.0 mM dithiothreitol, 20% glycerol, 0.1% Nonidet P-40, 100 mM KCl) overnight with 2 buffer changes. Poly(dI/dC) (10 µg/ml) was added to the dialyzed proteins which were then loaded onto an equilibrated CNBr-activated sepharose column (~ 5.5 cm^3 column volume, coupled with 144 bp *oriA* fragment) at 0.25 ml/min. The column was washed with 10 ml of buffer Z at 0.33 ml/min or with 4 - 2 ml aliquots of buffer Z at gravity flow, carefully washing the sides of the column. The bound proteins were eluted with 10 ml of a linear gradient of buffer Z from 0.1 M KCl to 0.9 M KCl at 0.25 ml/min, followed by 5 ml of buffer Z/1.0 M KCl at 0.25 ml/min, and then with 5 ml of buffer Z/2.0 M KCl at 0.25 ml/min. 1 ml fractions were collected beginning at volume 8 ml of the wash. *OriA* binding activity was detected via EMSA (as described below).

Biotin-labelling of 144 bp *oriA* PCR product

A 144 bp *oriA* region was amplified by polymerase chain reaction. Each 50 µl reaction tube contained 0.5 µM of each primer (CGGGTGAGATCCAATGTAGAT (OA1F) and GCTAAACCTGTGCTCGAGAGAT (OA144R)), 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 1X PCR reaction buffer, 2.5 units of Taq polymerase (Invitrogen), and 1 µl DNA template. A control reaction containing no DNA template was also run. All PCR reactions were performed on a PTC-150 MiniCycler (MJ Research) under the following conditions: 95°C for 3 minutes, 30 cycles of 95°C for 1 minute, 50°C for 1 minute, and 72°C for 1 minute, followed by 72°C for 7 minutes. A 5 µl aliquot of each reaction tube

was loaded onto a 1.2% agarose gel and run at 90V for about 45 minutes. *OriA* PCR products from the remaining 45 µl of each reaction were purified using the QIAquick PCR purification kit (Qiagen) according to the manufacturer's instructions. The purified *oriA* fragment was 3'end-labelled with biotin using the Biotin 3'end DNA Labeling Kit (Pierce) according to the manufacturer's instructions. All samples were stored at -20°C.

Electrophoretic mobility shift assays (EMSA)

EMSAs were performed using the LightShift Chemiluminescent EMSA Kit (Pierce) according to the manufacturer's instructions with slight modifications. Each 20 µl reaction contained 2 µg poly(dIdC), 1X LightShift binding buffer, 10 mM $MgCl_2$, 5-10 µl protein fraction, and 1 µl biotinylated *oriA*. Reactions were incubated at room temperature for 30 minutes. All samples were loaded onto a 5% TBE gel and run in 0.5X TBE at 75V until the tracking dye traveled ~3/4 of the way through the gel. DNA was transferred to a nylon membrane in a tank transfer apparatus (Bio-Rad) at 380 mA for 45 minutes in cold 0.5X TBE. The nylon membrane was cross-linked at 150mJoules/cm2 in a GS Gene Linker (Bio-Rad). Biotinylated *oriA* was detected using the LightShift Chemiluminescent EMSA kit according to the manufacturer's instructions.

SDS-PAGE analysis

OriA affinity column fractions were analyzed via SDS-PAGE. Fractions were concentrated by acetone precipitation by aliquoting 60 µl into a clean 1.5 ml microcentrifuge tube, adding nine volumes of cold acetic acetone (using a disposable transfer pipet, place one drop of 6N HCl into 10 ml of acetone while on ice) and incubating at -80°C overnight. Samples were then centrifuged at 12,000 x g for 20 minutes at 4°C. The supernatant was removed and the pellets were allowed to air-dry at

room temperature for 15-20 minutes. Pellets were resuspended in 20 µl of 2X SDS-sample loading buffer and heated at 95°C for 5 minutes. Samples were then loaded onto a 8-16% Tris-glycine gel (Invitrogen) and run at 30 mA in 1X SDS-running buffer until the tracking dye reached the bottom of the gel. The gel was washed 3 times in dH_2O for 5 minutes, stained with Bio-Safe Coomassie (Bio-Rad) for 1 hour, and destained overnight in dH_2O.

MALDI-ToF Mass Spectrometry

Protein bands for mass spectrometry analysis were excised from 8-16% Tris-glycine gels (Invitrogen) under sterile conditions, using sterile razor blades, and transferred to clean 1.5 ml microcentrifuge tubes. Excised gel bands were stored at -20°C. In-gel tryptic digests of protein bands were performed according to Shevchenko et al. (1996), with slight modifications. 1.5 ml of 50% acetonitrile/50 mM ammonium bicarbonate was added to each tube containing a gel band. Samples were placed into a micro mixer and rotated for 30 minutes at room temperature. The liquid was then aspirated off, and 140 µl of 100 mM ammonium bicarbonate was added to the samples, followed by 10 µl of reduction buffer (75 mM dithiothreitol, 100 mM ammonium bicarbonate). Samples were incubated with no motion at 60°C for 45 minutes. Samples were then removed from the heat block, cooled to room temperature, and solutions were aspirated off. 140 µl of 100 mM ammonium bicarbonate was added to the samples, followed by 10 µl of alkylation buffer (0.3 M iodoacetamide, 100 mM ammonium bicarbonate). Samples were incubated in the dark at room temperature for 30 minutes. Solutions were then aspirated off. 1 ml of 100 mM ammonium bicarbonate was added to each gel spot and tubes were rotated at room temperature for 30 minutes. Solutions were

then aspirated off. Under sterile conditions, gel bands were cut 1-2 times (depending on size of gel spot) with a sterile razor blade and transferred to a clean 1.5 ml microcentrifuge tube. 500 µl of 50% acetonitrile/50 mM ammonium bicarbonate was added to each sample tube and rotated at room temperature for 30 minutes. Solutions were then aspirated off. Gel pieces were dried to completeness in a speed vac. 10 µl of trypsin digestion solution (0.02 mg/ml trypsin in 25 mM ammonium bicarbonate) was added directly to each gel sample and incubated for 10 minutes at room temperature. This step was repeated until the gel pieces were completely hydrated with trypsin digestion solution. Enough 25 mM ammonium bicarbonate (usually 10-20 µl) was added to the samples to have a 0.5 mm excess of solution above the gel pieces. Samples were incubated at 37°C in a hybridization oven with rocking for 16-24 hours. Following incubation, 1 µl of 88% formic acid was added to the samples to stop the trypsin reaction. Samples were sonicated for 20 minutes in a water bath and then centrifuged briefly. Samples were then concentrated and desalted by using ZipTips (Millipore). Mass spectrometry analysis was performed on an API QSTAR Pulsar I from Applied Biosystems. Protein identifications were made using the nonrestricted NCBI protein database.

Protein Sequencing

N-terminal sequencing was performed by ProSeq, Inc (Boxford, Massachusetts). Protein samples were immobilized onto PVDF membrane, stained with Coomassie Brilliant Blue G-250 (OmniPur), and shipped to ProSeq according to instructions obtained from ProSeq, Inc.

Two-dimensional Gel Electrophoresis

Samples (*oriA* affinity column fractions 4 & 5) were prepared and analyzed by 2-D gel electrophoresis as described in detail in Chapter 1 of this thesis. Following separation in the 2nd dimension, gels were stained with Silver Stain Plus (Bio-Rad) according to the manufacturer's instructions.

GFP-fusion Targeting Experiments

The 84-bp region corresponding to the predicted transit peptide of *Arabidopsis* mtSSB (28 amino acids: MNSLAIRVSKVLRSSSISPLAISAERGS) was amplified by PCR (forward primer: 5'-CGCAAGCTTGATGAACTCACTCGCCATTA-3'; reverse primer: 5'-ACCGGTATAGAGCCTCTCTCAGCACTTAT) and fused to green fluorescent protein (GFP) at the NcoI site of pCAMBIA-1302. Targeting experiments were performed by the Anthony Maxwell lab (Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Colney, Norwich, UK) essentially as described in Wall et al. (2004). Cultured *Arabidopsis* cells were transformed by either *Agrobacterium* or biolistic transformation. Transient GFP expression was visualized by light microscopy with an epifluorescent Nikon Eclipse 600 microscope.

Results

Detection of *OriA*-binding Proteins in Cellulose Column Fractions

EMSAs were performed to identify fractions containing *oriA*-binding proteins. Shifts of the 144 bp *oriA* fragment were observed in chloroplast protein fractions eluted from DEAE cellulose columns (Figure 7). Two major shifts appear in the fractions loaded in lanes 3-6. The observation of a double shift could be due to multiple binding sites.

OriA-affinity Chromatography

Proteins eluted from the affinity column were separated by SDS-PAGE (Figure 10). One protein band at ~28 kDa was eluted at high salt at a higher protein concentration than any other protein band. This protein band appears to be initially washed off the column (Figure 8, Fractions 1-5) and then is eluted off the column at high salt concentration (Figure 8, Fraction 15). It is notable that the molecular weight is close to the 30 kDa tobacco chloroplast protein that was identified previously (Figure 3). Importantly, this protein band was not present in a control column (no *oriA* coupled to the resin) run under the same conditions (Figure 9). The 28 kDa band was excised from fractions 2, 3, and 15 for MALDI-ToF mass spectrometry analysis.

Sequence Analysis of the 28 kDa Soy Protein

The 28 kDa protein bands were digested with trypsin for analysis by MALDI-ToF mass spectrometry. The MS spectra obtained for the 28 kDa protein is seen in Figure 10. Nearly identical spectrums were obtained for each 28 kDa band analyzed from Fractions 2, 3, and 15. Seven strong peaks (Figure 10, peaks 1-7) were chosen for MS/MS. All seven peaks produced comparable MS/MS spectrum. The resulting data were subjected

to BLAST searches of the nonredundant NCBI protein database. BLAST results from three peaks identified an *Arabidopsis thaliana* single-stranded binding protein targeted to the mitochondria (mtSSB). Four peaks did not produce any significant hits. To further determine the structure of the 28 kDa soy protein, an excised gel band containing the 28 kDa protein was submitted for N-terminal sequencing. Thirty-six amino acids from the N-terminus were obtained. These 36 amino acids were identical to a 36 amino acid region of the *Arabidopsis* mtSSB (Figure 11, sequence highlighted in blue). BLAST searches of the *Arabidopsis* genome produced no additional homologues other than the mtSSB. In addition, BLAST searches of the dynamically translated soybean EST database also did not produce any significant hits when the 36 amino acid N-terminal sequence was used as the query. However, BLAST searches of the soybean database using the sequences identified by mass spectrometry (Figure 11, sequence highlighted in red) as the query did produce some EST matches. It is possible that the SSB domain homologous to both the 28 kDa soy protein and the *Arabidopsis* mtSSB is common among a variety of plant SSBs.

Two-dimensional Gel Electrophoresis of the 28 kDa Protein

The possibility remains that the failure to make a successful identity with four of the MS peaks could be due to post-translational modifications to the protein. To address this question, protein fractions containing the 28 kDa protein were separated by two-dimensional gel electrophoresis. Figure 12 shows the results of the 2-D analysis of *oriA* affinity column fraction 4, which contains the 28 kDa protein band. The pattern that resulted from the 2-D separation suggests that post-translational modifications do occur on the 28 kDa protein and therefore may be affecting identification of some of the MS

spectra peaks. Such posttranslational modifications may play a role in regulating the activity of the 28 kDa protein. The activity of many regulatory proteins is known to be regulated by modifications such as phosphorylation.

Arabidopsis mtSSB Targeting Experiments

Although most organelle-targeted proteins localize to a single target, dual-targeted proteins have been identified that localize to both mitochondria and chloroplasts (Peeters and Small, 2001; Silva-Filho, 2003). It is possible that the 28 kDa protein is the soybean homologue to the *Arabidopsis* mtSSB and that the protein localizes to both mitochondria and chloroplasts. Because the gene sequence, and thus the target peptide, was not available for the 28 kDa soy protein, targeting experiments were performed using the target peptide from *Arabidopsis* mtSSB. Targeting experiments were performed using a green fluorescent protein (GFP)-fusion construct containing the *Arabidopsis* mtSSB target peptide (mtTP). Targeting experiments were performed by Melisa Wall of the Anthony Maxwell laboratory essentially according to Wall et al. (2004). GFP-mtTP localization to mitochondria was observed in both the *Agrobacterium*- and biolistic-transformed cultured *Arabidopsis* cells (Figures 13 and 14). Chloroplast localization of this construct was not observed in any of the transformed cultured cells (Figure 14). These results support the mitochondrial localization of *Arabidopsis* mtSSB and suggest that it is not dual-targeted to chloroplasts. However, whether the soybean 28 kDa protein is dual-targeted to both chloroplasts and mitochondria has yet to be determined, as the complete gene with the target peptide sequence will need to be identified before such experiments are performed.

Electrophoretic Mobility Shift Assays with the 28 kDa Protein

EMSAs were performed with fractions containing the 28 kDa protein (Figure 8), using the 144 bp *oriA* fragment as probe (Figure 15). Fraction 4 displays no shift (Figure 15, lane 2), while strong shifts are observed in both fractions 14 and 15 (Figure 15, lane 5), which contain a more purified elution of the 28 kDa protein. Importantly, EMSAs with purified *Arabidopsis* mtSSB (overexpressed in the mature form in *E. coli* and purified on a nickel-chelate column) performed under the same conditions do not result in a shift of *oriA* (Figure 15, lane 6), but do result in a shift of single-stranded DNA (ssDNA) (Figure 16). This indicates that the 28 kDa soy protein and *Arabidopsis* mtSSB differ in their activity, as the 28 kDa soy protein does cause a shift of *oriA*. The 28 kDa soy protein does not appear to bind nonspecific ssDNA (Figure 16).

Discussion

A 28 kDa soybean protein was isolated by sequence specific DNA affinity chromatography from total soybean chloroplast proteins. Mass spectrometry analysis revealed that the 28 kDa protein contains some homology within an SSB domain of an *Arabidopsis* mitochondrial-targeted SSB (mtSSB) of approximately 21 kDa. N-terminal sequencing revealed that the 28 kDa soy protein is identical over a 36 amino acid region at the N-terminus of both the soybean 28 kDa protein and mature *Arabidopsis* mtSSB. These sequence results suggest that the 28 kDa protein could be the soy homologue of *Arabidopsis* mtSSB. However, targeting experiments and EMSA data indicate that there are definite differences between these two proteins (see additional discussion below). *Arabidopsis* mtSSB appears to be targeted to mitochondria only and not to chloroplasts (Figures 13 and 14). Targeting experiments have not been performed with the target peptide for the 28 kDa protein due to the failure to identify the soy gene that encodes the protein.

The fact that BLAST searches of the available soybean genomic and EST databases do not produce any significant homologies to the 36 amino acid N-terminal sequence raises at least four interesting questions. First, could the presence of the 28 kDa protein in the *oriA*-affinity column be due to contamination, either from our lab or from mitochondrial contamination during the isolation procedure? *Arabidopsis* mtSSB was being overexpressed and characterized by our lab at the time that the *oriA*-affinity column was performed. However, the overexpressed form of the mtSSB protein contains a methionine residue at the N-terminus for translation purposes, and results of N-terminal sequencing of the 28 kDa protein show an alanine as the first residue on the N-terminus.

Also, soybean mitochondrial contamination during the isolation procedure is most likely not the source of the 28 kDa protein, although that possibility cannot be completely ruled out. Western blot analysis of total chloroplast proteins isolated by the same procedure used for isolating the proteins loaded onto the *oriA*-affinity column indicate that mitochondrial contamination is less than 5% (data not shown). The abundance of the 28 kDa protein in the *oriA*-affinity column argues against mitochondrial contamination as the source of the 28 kDa protein (Figure 8).

Second, could the 28 kDa protein be chloroplast encoded? NCBI contains complete sequences for chloroplast genomes from 36 species. Soybean (*Glycine max*) has not been sequenced. Soy is part of the legume family and chloroplast genomes from only two legumes have been sequenced: *Medicago truncatula* and *Lotus japonicus*. Although chloroplast genomes appear to be fairly conserved, and no identification of replication proteins has been made in the sequenced chloroplast genomes, there are sequence differences between chloroplasts of different species (Palmer et al., 1988; Palmer et al., 1987; Palmer et al., 1983). The soy genome may contain a novel gene not found in the chloroplast genomes sequenced to date. The fact that a match is not found in the soy EST databases suggests that the 28 kDa protein may not be nuclear encoded, though a similar protein has not been identified in any of the 36 chloroplast genomes sequenced to date.

Third, if the 28 kDa protein is involved in ctDNA replication initiation, is the window of expression of this protein sufficiently short to elude EST isolations thus far? If ctDNA replication is controlled by a mechanism similar to the hypothesized mechanism for animal mtDNA copy number control (reviewed in Background section), a

protein involved in initiation of ctDNA replication may be actively expressed at high levels for a short time during the early stages of plant development and then turned off. The mRNA may be present in the cytoplasm for a short enough time as to elude the RNA isolations for EST library construction that have been previously performed.

Fourth, is the mRNA encoding the 28 kDa protein expressed at sufficiently high concentrations to be detected in mRNA isolations, and if so, has it simply eluded detection? Although the probability is low that the mRNA encoding one particular protein would not show up, the fact remains that no matches have been found in any of the available EST databases. Future mRNA isolations may yet reveal an EST match. The identification of a SSB by mass spectrometry raised additional questions. Why would a single-stranded binding protein elute at high salt concentration from a double-stranded DNA affinity column? One possibility is that the activity of another protein present in the column could have unwound *oriA*, allowing an SSB to bind. It is also possible that the 28 kDa protein has an SSB domain as well as some double-stranded DNA binding ability. However, EMSAs performed with the 28 kDa protein and *Arabidopsis* mtSSB show differences in their binding abilities. The 28 kDa protein shifts *oriA*, while *Arabidopsis* mtSSB fails to shift *oriA* (Figure 15). The sequence homology identified with the *Arabidopsis* mtSSB (Figure 11, sequence highlighted in red) by mass spectrometry is found within a conserved SSB domain. *Arabidopsis* mtSSB causes a strong shift of ssDNA (Figure 16). If the two proteins are homologous, why does the 28 kDa soybean protein fail to shift ssDNA (Figure 16)? The answer to that question is still unclear. Double-stranded *oriA* EMSAs indicate that the 28 kDa protein binds specifically to *oriA*. It is possible that the 28 kDa protein may also bind ssDNA in a sequence

specific manner. It is also interesting to consider possible effects that post-translational modifications may have on the function of the 28 kDa protein. Separation by two-dimensional gel electrophoresis suggests that there may be modified forms (eg. phosphorylated) of the 28 kDa protein within chloroplasts. Such modifications could potentially be involved in regulating the DNA-binding activity of the 28 kDa protein. This could be one explanation for why the 28 kDa protein appeared to be first washed off of the column and then was eluted at higher salt concentrations (Figure 10). Alternatively, post-translational modifications may regulate protein-protein interactions with accessory proteins involved in ctDNA replication. Further work is needed to elucidate the DNA-binding abilities of the 28 kDa protein.

Two-dimensional gel electrophoresis of the 28 kDa protein indicates that (1) post-translational modifications may be occurring and thus interfering with identification of all the mass spectrometry peaks, and (2) the isoelectric point (pI) of the 28 kDa protein is between 5.5 and 6.0 (Figure 12). If the 28 kDa protein is involved in ctDNA replication initiation, posttranslational modifications may be a way of regulating or modifying its activity. The pI of the 28 kDa protein was at first confusing. DNA binding proteins are typically more basic in their charge. However, *E. coli* SSB has a pI of 6.0, and *Synechococcus* SSB has a predicted pI of 5.0 (Meyer and Laine, 1990; *Synechococcus* SSB pI predicted by EMBL WWW Gateway to Isoelectric Point Service, http://www.embl-heidel). The C-terminus of *E. coli* SSB, which does not appear to play a direct role in DNA binding, contains several negatively charged amino acids that when cleaved leaves the SSB with a pI of 8.9 (Kinebuchi et al., 1997; Meyer and Laine, 1990). Thus, while the 28 kDa protein has an overall pI of between 5.5 and 6.0, it may contain

DNA-binding domains with a much more basic pI. Also, it is worth noting that the predicted pI of the mature *Arabidopsis* mtSSB is 8.2, which is significantly different than that of the 28 kDa protein (mtSSB pI predicted by EMBL WWW Gateway to Isoelectric Point Service, http://www.embl-heidel).

In summary, these results suggest that the 28 kDa soy protein interacts specifically with the chloroplast *oriA* and therefore may play a role in ctDNA replication during the initiation stage. While the 28 kDa protein shows some homology to *Arabidopsis* mtSSB, there are some critical differences in the localization and activity of these proteins in vitro, suggesting that they have different functions in vivo. Further work is needed to confirm chloroplast localization of the 28 kDa protein and to elucidate the role that the 28 kDa protein may play in ctDNA replication. An antibody raised against the N-terminus of the 28 kDa protein is currently in production. This will aid in determining localization. Identification of the soy gene encoding the 28 kDa protein will also be critical in determining the target sequence and the extent to which the 28 kDa protein is homologous to *Arabidopsis* mtSSB. Identification of the gene will also aid in determining gene expression levels in various tissues during different stages of plant development that may give some indication of its possible role in ctDNA replication.

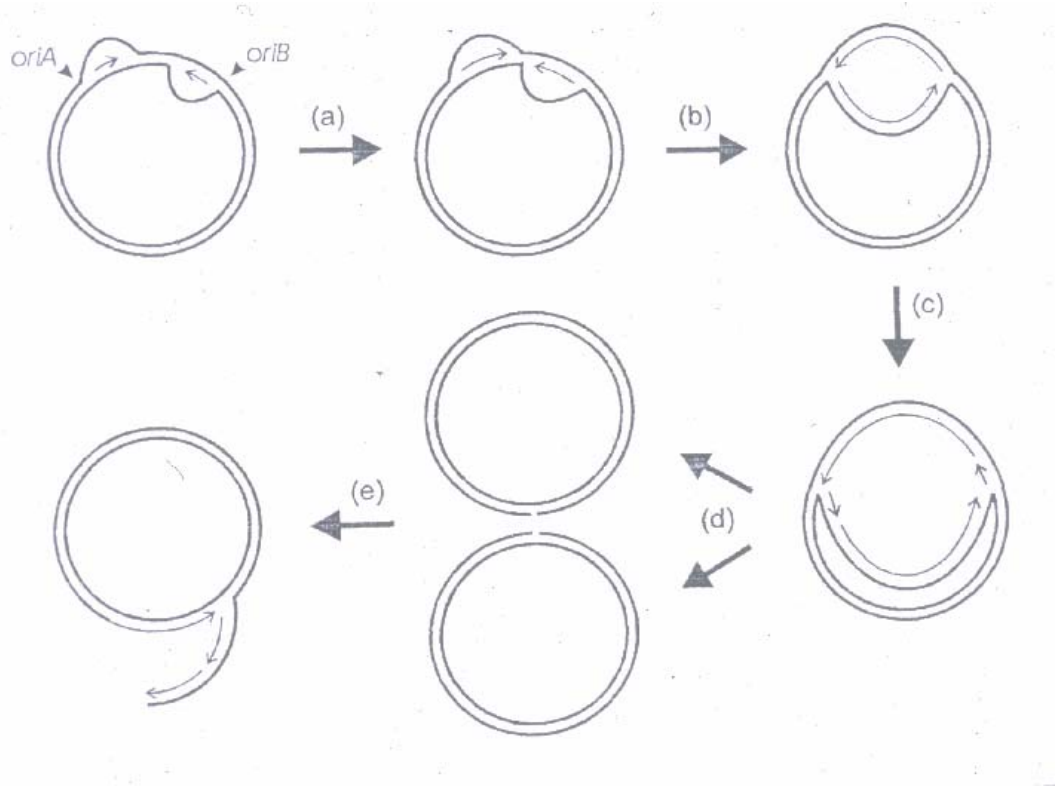


Figure 1 Model of ctDNA replication mechanism. Replication initiates at two origins of replication, *oriA* and *oriB*. Displacement-loops (D-loops) move toward each other in a unidirectional manner. Once the D-loops meet, replication proceeds bidirectionally in a theta mechanism. Once the entire DNA molecule is replicated and before the nick is sealed, replication can continue by a rolling circle mechanism.

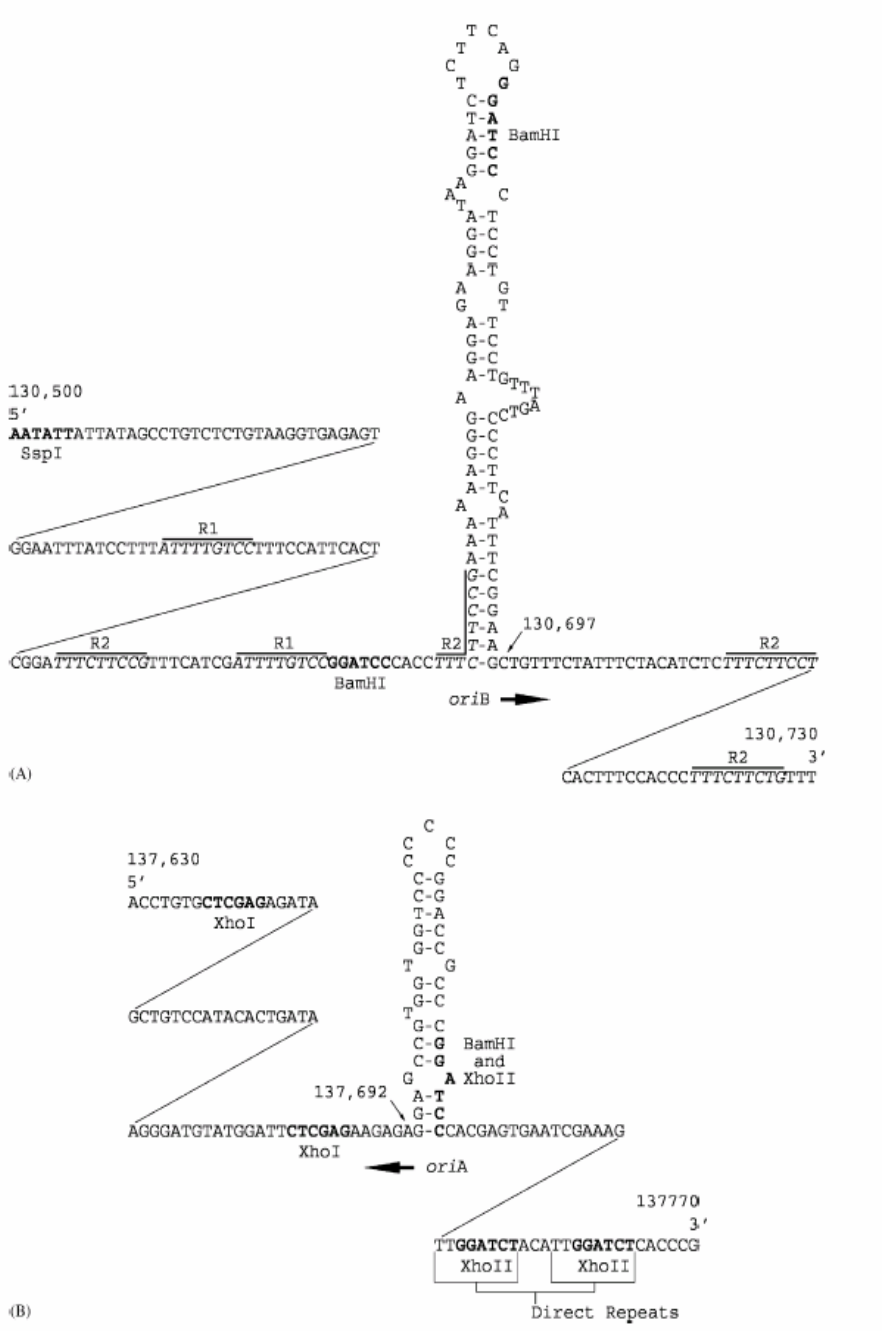


Figure 2 Potential stem loop structure and location of direct repeats for *oriB* (A) and *oriA* (B). Direct repeats in *oriB* are indicated by R1 (repeat 1) or R2 (repeat 2); *oriA* direct repeats are labeled near the 3' end. Arrows indicate the direction of replication initiation from each *ori*. Labeled restriction sites were used in the creation of deletion mutants for identification of essential elements required for replication (Lugo et al., 2004). Sequence coordinates were taken from Shinozaki et al. (1986).

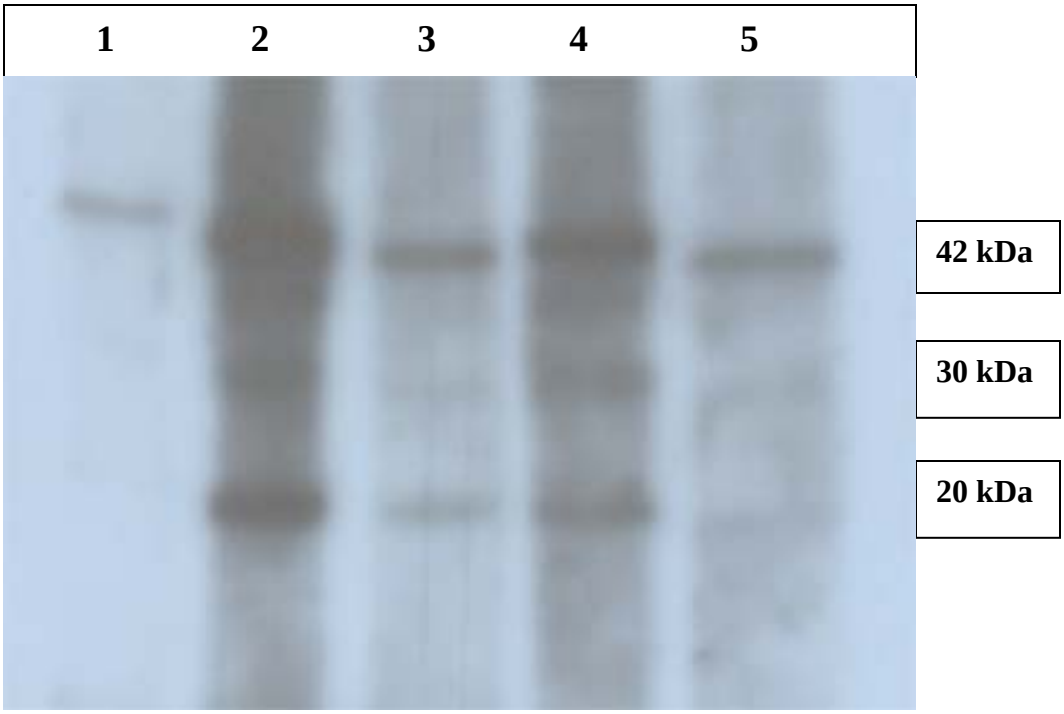


Figure 3 Detection of *ori*-binding proteins from total chloroplast extracts via southwestern blot. Lane 1 was probed with pUC19 DNA; lane 2 was probed with a 168 bp *oriA* fragment; lane 3 was probed with a 1.3 kb fragment containing *oriA*; lane 4 was probed with a 248 bp *oriB* fragment; lane 5 was

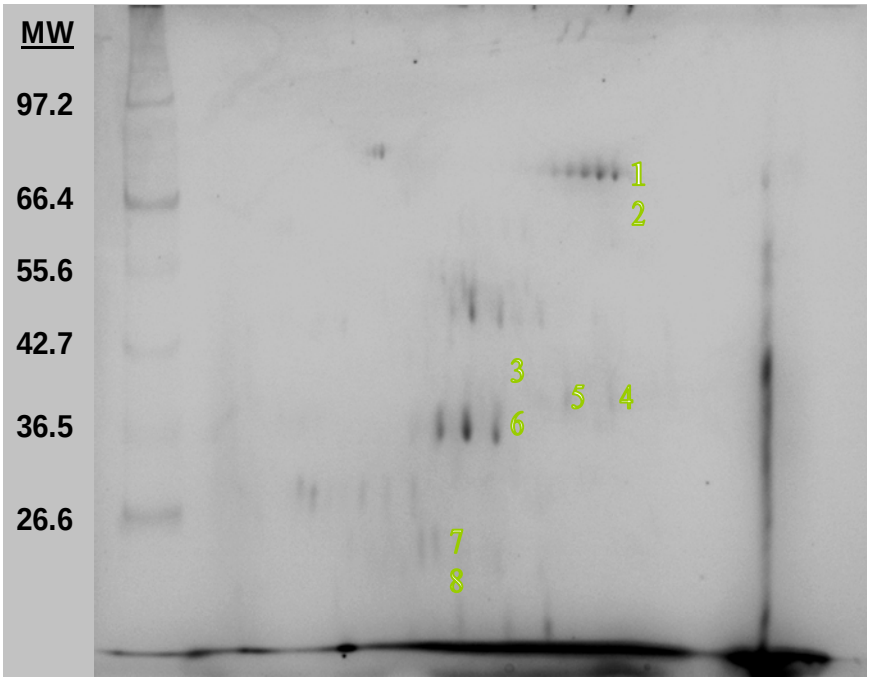
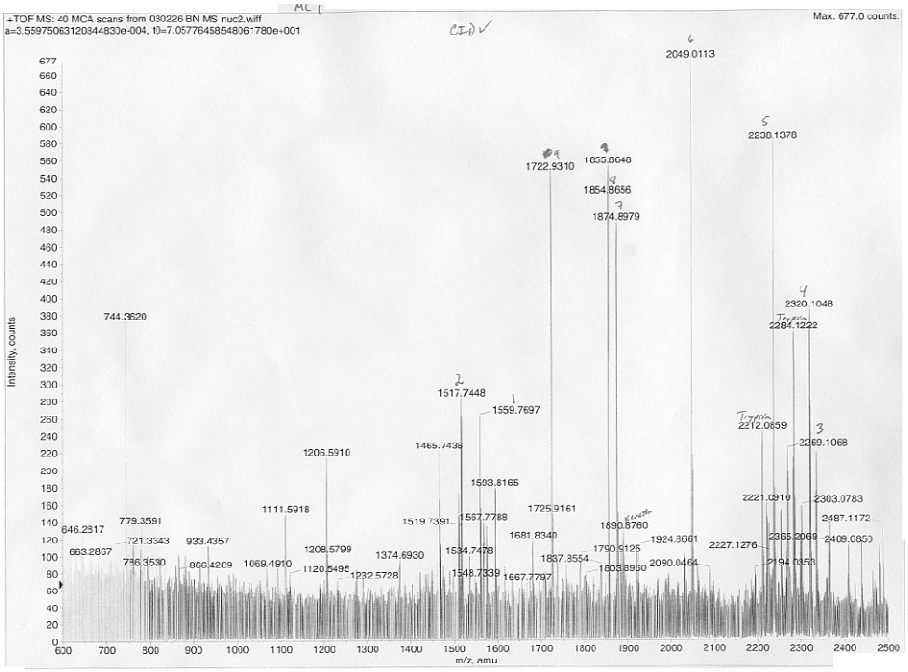


Figure 4 Separation of isolated nucleoid proteins by two-dimensional gel electrophoresis, pH 4-7. Protein bands on left side of gel are a marker. The streak on the right side of the gel contains unresolved proteins. Virtually all of the spots visible on the gel were excised for mass spectrometry analysis. Labelled spots correspond to the protein identifications listed in Table 1. Identifications were made from several gels, so, due to variation in isolations, not all of the spots are represented on this gel. Also, ferredoxin:sulfite reductase (#2) protein sample was excised from nucleoid proteins separated by one-dimensional SDS-PAGE, so the location indicated on the gel represents only the approximate molecular weight.



e 5 Example of MS spectra obtained from MALDI-ToF mass spectrometry analysis of nucleoid proteins. Typically, peaks with a mass-to-charge ratio between 1300 and 2200 were chosen for MS/MS.

Protein	Approx. Molecular Weight on Gel
1) Ct import-associated channel protein	66 kDa
2) Ferredoxin:sulfite reductase precursor	64 kDa
3) Glyceraldehyde-3-phosphate dehydrogenase	41 kDa
4) ATP Synthase, beta subunit	38 kDa
5) Malate dehydrogenase	37 kDa
6) Oxygen-evolving enhancer protein 1	36 kDa
7) Carbonic anhydrase	23 kDa
8) Chlorophyll a/b binding protein	20 kDa

Table 1 List of protein identifications made via MALDI-ToF mass spectrometry analysis of isolated nucleoid proteins. Ferredoxin:sulfite reductase was identified previously as a nucleoid-associated protein. See Figure 4 for corresponding protein spot excised from 2-D gel (protein identifications were obtained from gel spots from several gels, thus, due to variation in isolations, not all spots are visible on the gel represented in Figure 4).

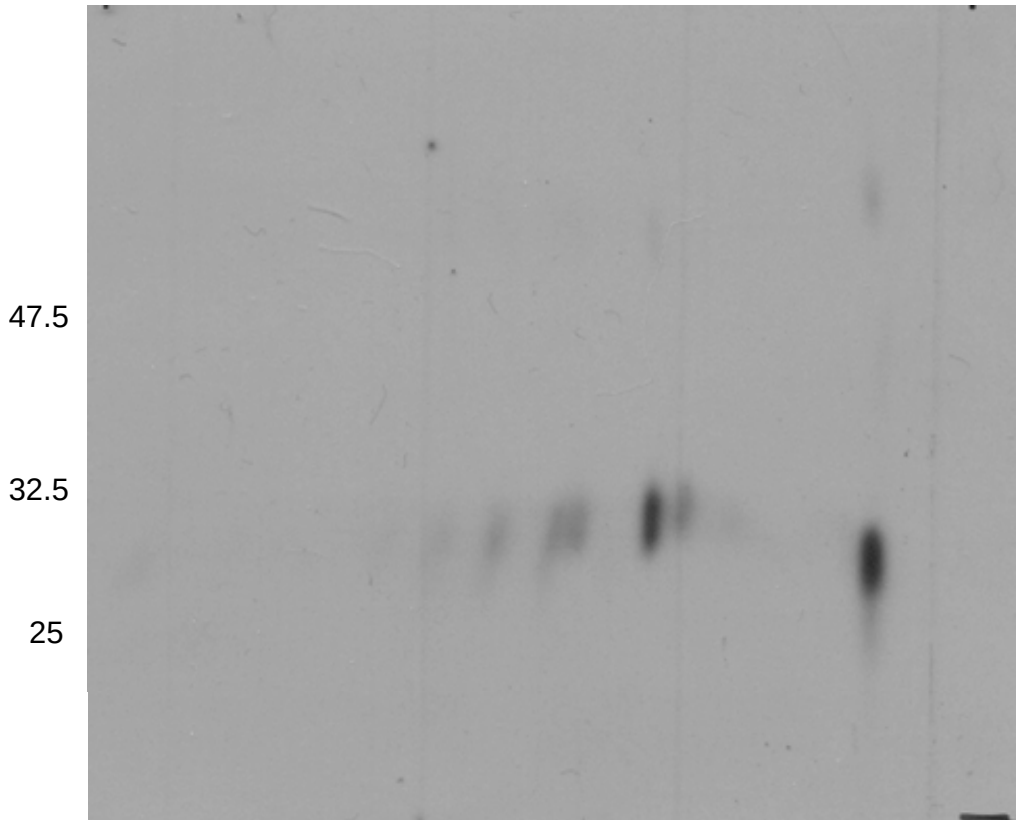


Figure 6 Detection of *oriA*-binding proteins by southwestern blot. Spots correlate with approximate molecular weights of 70, 65, 31, 31, 30, 30, 30, and 26 kDa. Blot was performed with transferred nucleoid proteins separated by two-dimensional gel electrophoresis, pH 4-7.

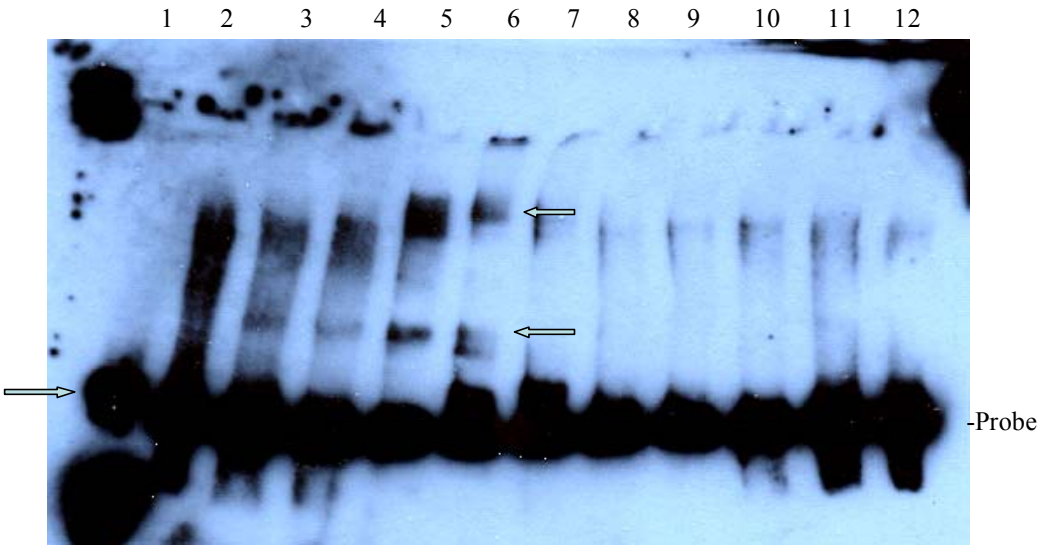


Figure 7 *OriA* electrophoretic mobility shift assay of DEAE cellulose column fractions. Lane 1 is a kit positive control. Lane 2 contains no protein. Lanes 3-12 contain column fractions. Arrow pointing to the right indicates a shift in the kit control. Arrows pointing to the left indicate shifts of *oriA*. Free *oriA* is marked as probe.

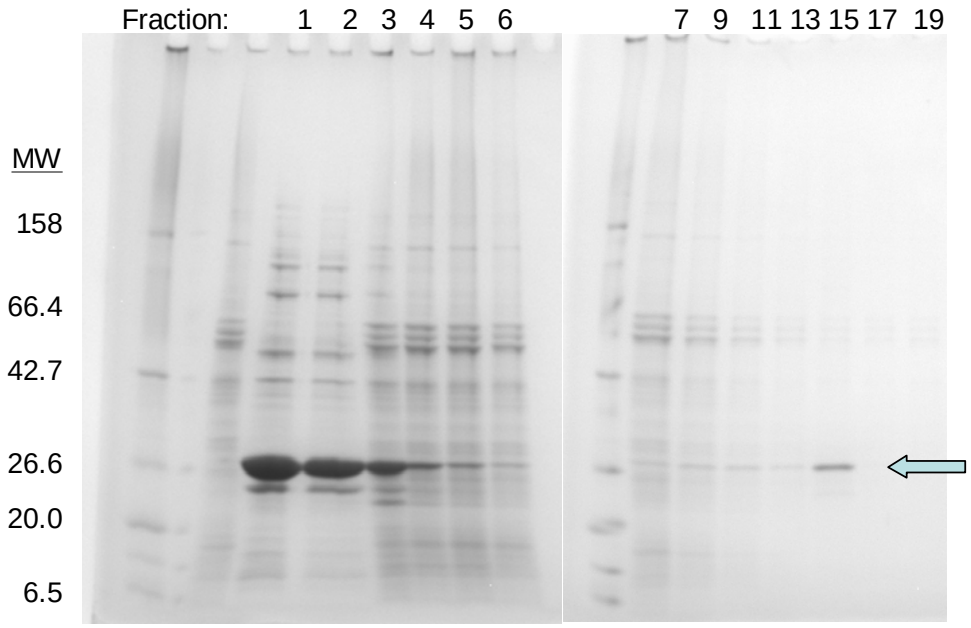


Figure 8 *OriA* affinity column fractions. The lane on the far left of each gel is the marker; the second lane from the far left of the gel on the left is the total dialyzed protein loaded onto the column; lanes 1-7, 9, 11, 13, 15, 17, and 19 refer to the affinity column fraction loaded. Arrow indicates the 28 kDa protein band.

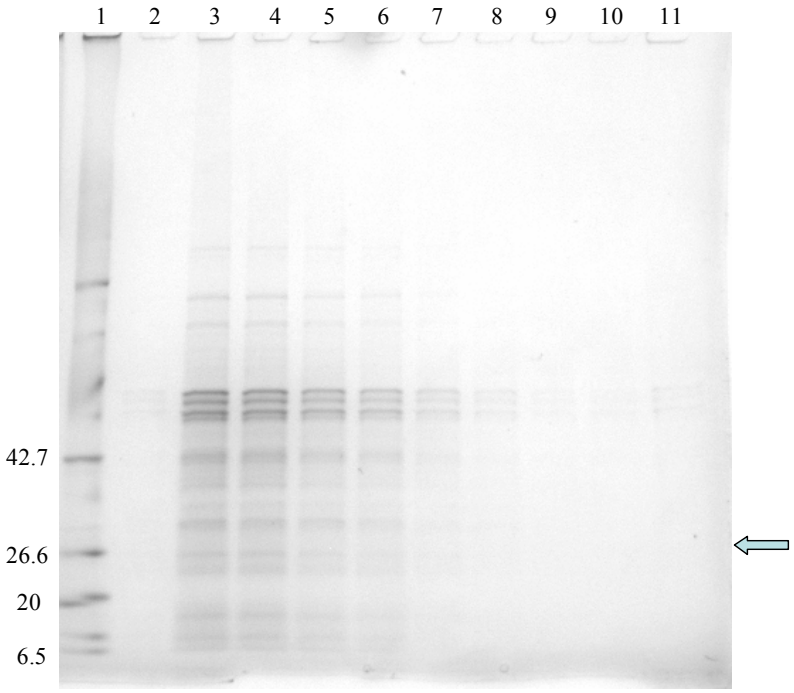


Figure 9 Affinity control column fractions. CNBR-activated sepharose resin was prepared identically as the *oriA*-coupled affinity resin, except that no *oriA* was added to the resin. Lane 1 is a marker. Lane 2 is the last wash fraction. Lanes 3-11 are even fractions 4-20. The arrow indicates approximately where 28 kDa is on the gel.

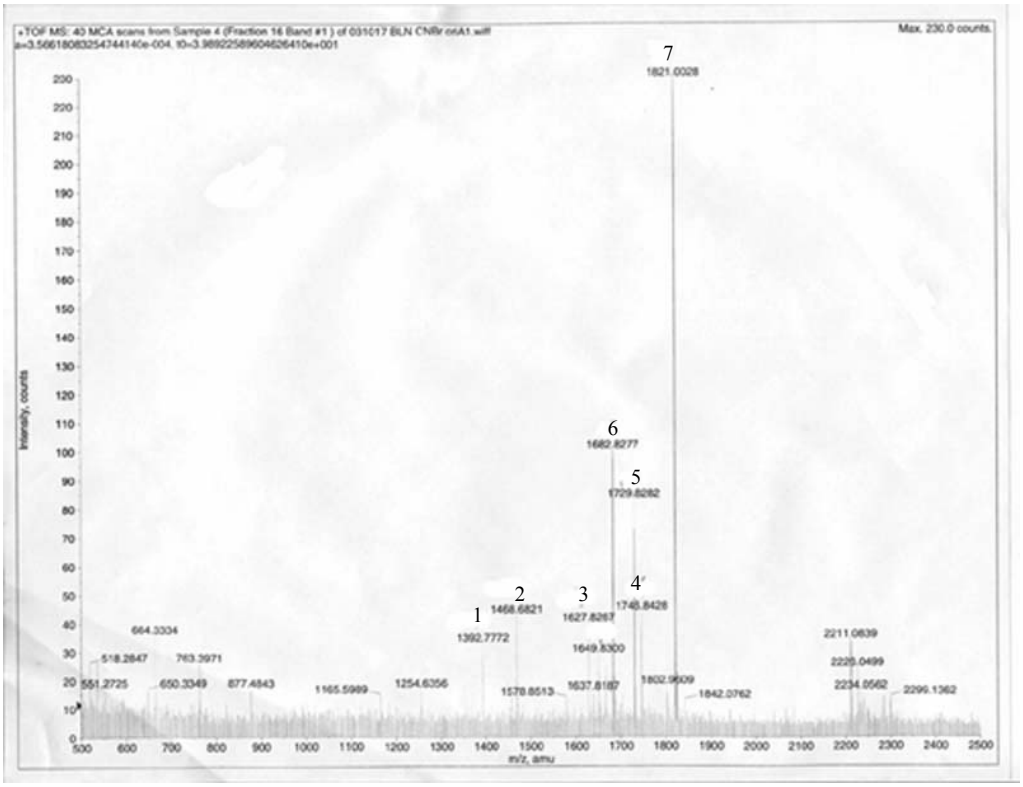


Figure 10 Mass spectrometry data (MS spectra) obtained from 28 kDa protein eluted off *oriA* affinity column. Seven peaks between 1300-1900 m/z were chosen for MS/MS analysis. Three peaks (2, 3, and 5) identified an *Arabidopsis thaliana* mitochondrial-targeted single-stranded binding protein (mtSSB). The remaining four peaks (1, 4, 6, and 7) had comparable MS/MS spectra, but did not result in any significant hits.

1 MNSLAIRVSKVLRSSSISPLAISAERGSKSWFSTGPIDEGVEEDFEEN
49 VTERPELQPHGVDPRKGWGFRGVHRAIICGKVGQAPLQKILRNGR
94 TVTIFTVGTGGMFDQRLVGATNQPKPAQWHRIAVHNEVLGSYAV
148 QKLAKNSSVYVEGDIETRVYNDSISSEVKSIPEICVRRDGKIRMIKY
195 GESISKISFDELKEGLI

Figure 11 Amino acid sequence of *Arabidopsis* mtSSB. Highlighted sections indicate sequences homologous to soy 28 kDa protein. Blue – N-terminal sequence obtained from ProSeq, Inc. Red – sequence identified by mass spectrometry.

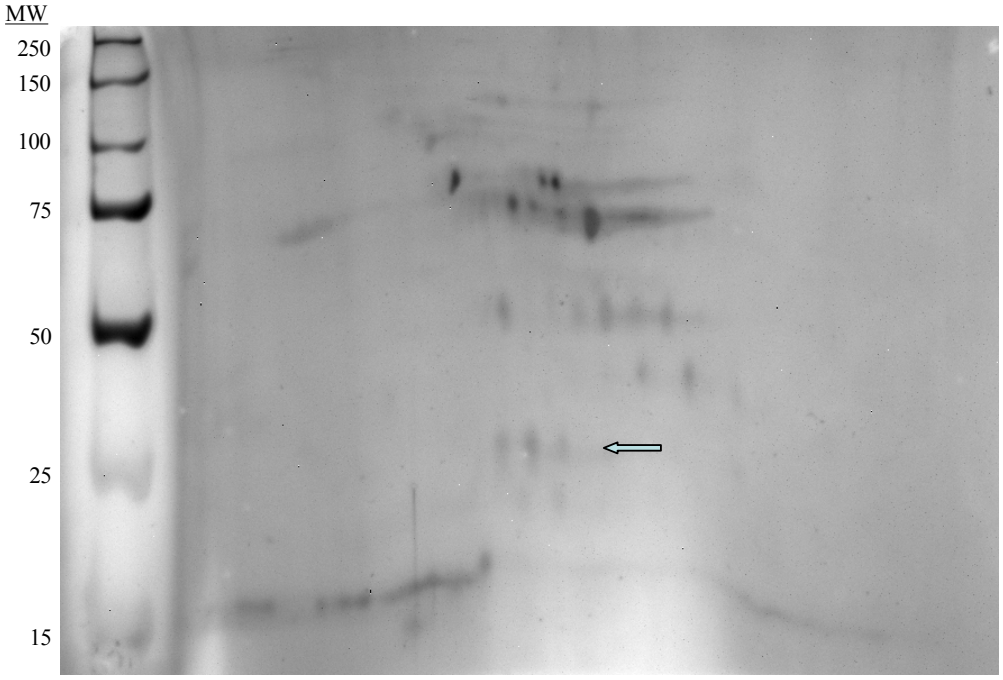


Figure 12 Two-dimensional gel separation of *oriA* affinity fraction 4 (a previous wash fraction containing the 28 kDa protein was used due to the higher concentration of the 28 kDa band), stained by silver staining. Arrow approximates 28 kDa on the gel.

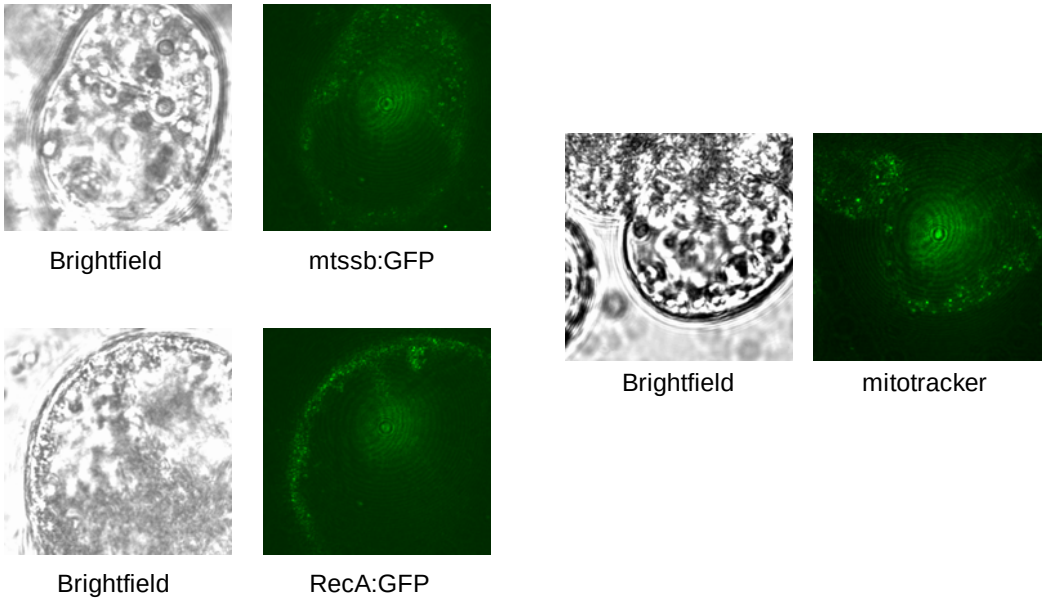


Figure 13 Agrobacterium-transformed *Arabidopsis* cells with pCAMBIA:targetpeptideGFP. GFP was either fused with the target peptide of *Arabidopsis* mtSSB or *Arabidopsis* RecA homologue. Mitotracker indicates the location of the mitochondria within the cell. Both mtssb:GFP and RecA:GFP appear to localize to mitochondria.

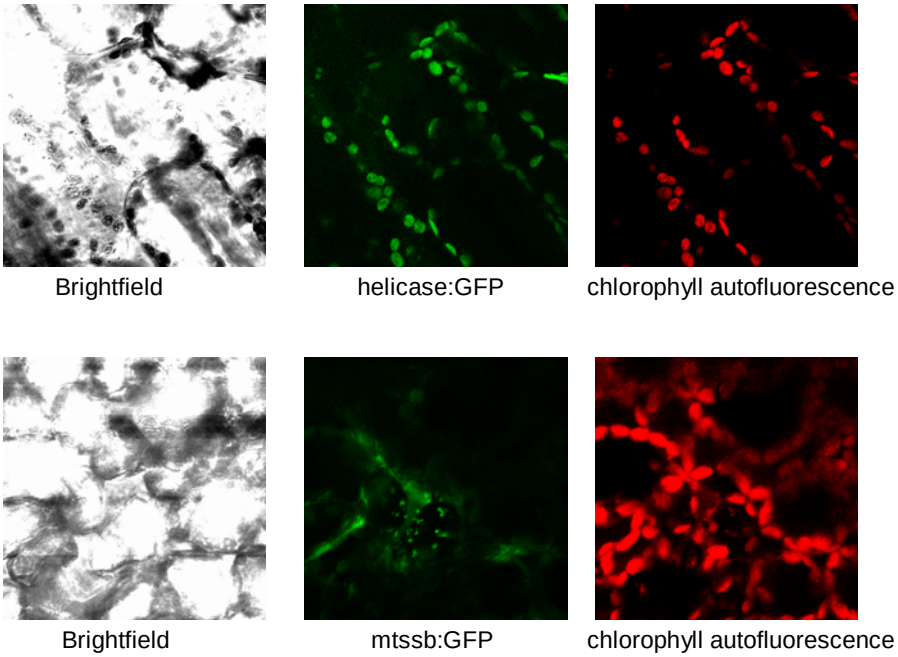


Figure 14 *Arabidopsis* mesophyll cells bombarded with pCAMBIA:targetpeptideGFP. GFP was either fused with the target peptide of *Arabidopsis* mtSSB or *Arabidopsis* putative helicase. The putative helicase appears to localize within chloroplasts, while mtSSB does not.

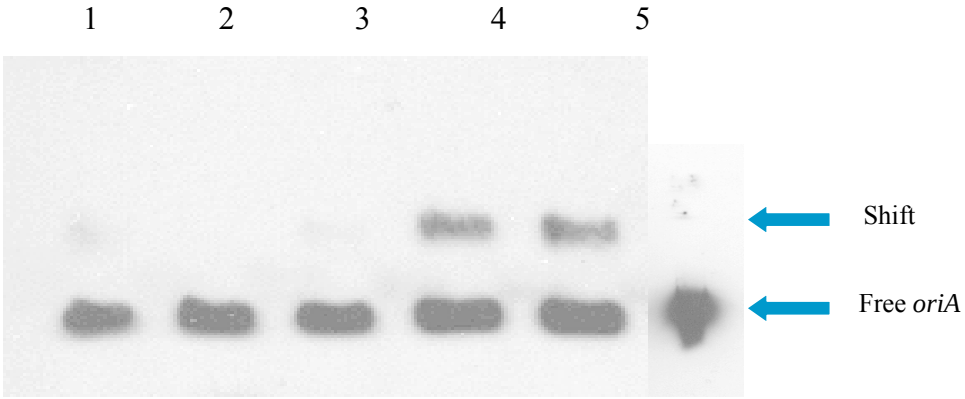


Figure 15 *OriA* electrophoretic mobility shift assays with *oriA* affinity column fractions. Lane 1 contains no protein; lane 2 contains *oriA* affinity fraction 4; lane 3 contains *oriA* affinity fraction 5; lane 4 contains *oriA* affinity fraction 14; lane 5 contains *oriA* affinity fraction 15; lane 6 contains *Arabidopsis* mtSSB.

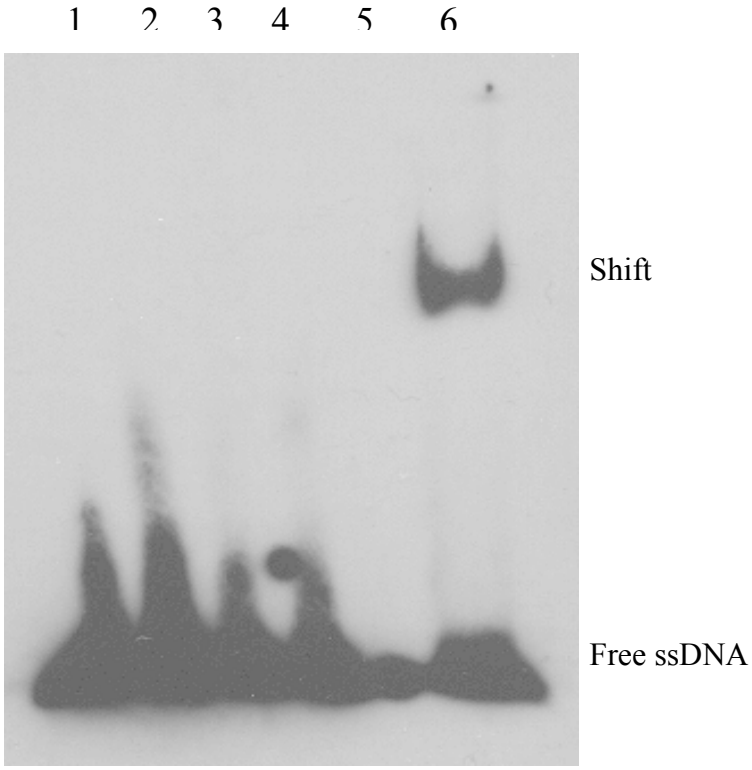


Figure 16 Single-stranded DNA electrophoretic mobility shift assay with the 28 kDa soy protein and *Arabidopsis* mtSSB. Lane 1 contains no protein; lanes 2-5 contain the 28 kDa protein (*oriA* affinity fraction 15) with decreasing amounts of ssDNA; lane 6 contains *Arabidopsis* mtSSB.

References

Ausubel F.M., Brent R., Kingston R.E., Moore D.D., Seidman J.G., Smith J.A., and Struhl K. (Eds.) (1993) Purification of Sequence-Specific DNA-Binding Proteins by Affinity Chromatography. Current Protocols in Molecular Biology, 12.10.6.
Backert S., Dorfel P., Lurz R., and Borner T. (1996) Rolling-Circle Replication of Mitochondrial DNA in the Higher Plant Chenopodium album (L.). *Mol. Cell. Biol.* **16**(11):6285-6294.
Bartlett S.G., Grossman A.R., Chua N.H. (1982) In vitro synthesis and uptake of cytoplasmically-synthesized chloroplast proteins, in: Edelman M., Hallick R.B., Chua N.H. (Eds.), Methods in Chloroplast Molecular Biology. Elsevier, Amsterdam, pp 1081-1091.
Bogenhagen D., and Clayton D.A. (1977) Mouse L cell mitochondrial DNA molecules are selected randomly for replication throughout the cell cycle. *Cell* **11**:719-727.
Bryant J.A., Moore K., and Aves S.J. (2001) Origins and complexes: the initiation of DNA replication. *J. Exp. Bot.* **52**(355):193-202.
Cannon G.C., Ward L.N., Case C.I., and Heinhorst S. (1999) The 68 kDa DNA compacting nucleoid protein from soybean chloroplasts inhibits DNA synthesis in vitro. *Plant Mol. Biol.* **39**(4):835-845.
Chen W., Gaikwad A., Mukherjee S.K., Choudhary N.R., Kumar D., and Tewari K.K. (1996) A 43 kDa DNA binding protein from the pea chloroplast interacts with and stimulates the cognate DNA polymerase. *Nucleic Acids Research* **24**(20):3953-3961.
Chi-Ham C.L., Keaton M.A., Cannon G.C., and Heinhorst S. (2002) The DNA-compacting protein DCP68 from soybean chloroplasts is ferredoxin:sulfite reductase and co-localizes with the organellar nucleoid. *Plant Mol. Biol.* **49**:621-631.
Clegg M.T., Gaut B.S., Learn G.H., and Morton B.R. (1994) Rates and patterns of chloroplast DNA evolution. *Proc. Natl. Acad. Sci.* **91**:6795-6801.
Davis A.F. and Clayton D.A. (1996) *In situ* localization of mitochondrial DNA replication in intact mammalian cells. *J. Cell Biol.* **135**:883-893.
Dyall S.D., Brown M.T., and Johnson P.J. (2004) Ancient Invasions: From Endosymbionts to Organelles. *Science* **304**:253-257.
Emanuelsson O., Nielsen H., and von Heijne G. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science* **8**:978-984.
Gaikwad A., Hop D.V., and Mukherjee S.K. (2002) A 70-kDa chloroplast DNA polymerase from pea (Pisum sativum) that shows high processivity and displays moderate fidelity. *Mol. Genet. Genomics* **267**:45-56.
Giraldo R. (2003) Common domains in the initiators of DNA replication in Bacteria, Archaea and Eukarya: combined structural, functional and phylogenetic perspectives. *FEMS Micro. Rev.* **26**:533-554.
Gonzalez F., Delahodde A., Kodadek T., and Johnston S.A. (2002) Recruitment of a 19S proteasome subcomplex to an activated promoter. Science 296(5567):548-550.
Hashimoto H. (1985) Changes in Distribution of Nucleoids in Developing and Dividing Chloroplasts and Etioplasts of Avena sativa. *Protoplasma* **127**:119-127.
Heazlewood J.L., Tonti-Filippini J.S., Gout A.M., Day D.A., Whelan J., and Millar A.H. (2004) Experimental Analysis of the Arabidopsis Mitochondrial Proteome

Highlights Signaling and Regulatory Components, Provides Assessment of Targeting Prediction Programs, and Indicates Plant-Specific Mitochondrial Proteins. *Plant Cell* **16**:241-256.

Hsieh C.-H., Wu M., and Yang J. (1991) The sequence-directed bent DNA detected in the replication origin of Chlamydomonas reinhardtii chloroplast DNA is important for the replication function. *Mol. Gen. Genet.* **225**:25-32.

Ishov A.M., Vladimirova O.V., and Maul G.G. (2004) Heterochromatin and ND10 are cell-cycle regulated and phosphorylation-dependent alternate nuclear sites of the transcription represser Daxx and SWI/SNF protein ATRX. *J. of Cell Sci.* **117**:3807-3820.

Jacob F., Brenner S., Cuzin F. (1963) On the regulation of DNA replication in Bacteria. *Cold Spring Harbor Symp. Quant. Biol.* 28:329-438.

Jarvis P. and Soll J. (2001) Toc, Tic, and chloroplast protein import. *Biochim. Biophys. Acta* **1541**:64-79.

Jeong S.Y., Rose A., Meier I. (2003) MFP1 is a thylakoid-associated, nucleoid-binding protein with a coiled-coil structure. Nucleic Acids Res. 31(17):5175-5185.

Jeffery C.J. (2003) Moonlighting proteins: old proteins learning new tricks. *Trends in Genetics* **19**(8):415-417.

Khazi F.R., Edmondson A.C., Nielsen B.L. (2003) An Arabidopsis homologue of bacterial RecA that complements an E. coli recA deletion is targeted to plant mitochondria. *Mol. Genet. Genomics* **269**(4):454-463.

Kimura S., Uchiyama Y., Kasai N., Namekawa S., Saotome A., Ueda T., Ando T., Ishibashi T., Oshige M., Furukawa T., Yamamoto T., Hashimoto J., and Sakaguchi K. (2002) A novel DNA polymerase homologous to Escherichia coli DNA polymerase I from a higher plant, rice (Oryza sativa L.). *Nucleic Acids Research* **30**(7):1585-1592.

Kinebuchi T., Shindo H., Nagai H., Shimamoto N., and Shimizu M. (1997) Functional Domains of Escherichia coli Single-Stranded DNA Binding Protein As Assessed by Analysis of the Deletion Mutants. *Biochemistry* **36**:6732-6738.

Kobayashi T., Takahara M., Miyagishima S., Kuroiwa H., Sasaki N., Ohta N., Matsuzaki M., and Kuroiwa T. (2002) Detection and Localization of a Chloroplast-Encoded HU-Like Protein That Organizes Chloroplast Nucleoids. *Plant Cell* **14**:1579-1589.

Kunnimalaiyaan M. and Nielsen B.L. (1997a) Fine mapping of replication origins (oriA and oriB) in Nicotiana tabacum chloroplast DNA. *Nucleic Acids Research* **25**(18):3681-3686.

Kunnimalaiyaan M. and Nielsen B.L. (1997b) Chloroplast DNA Replication: Mechanism, Enzymes and Replication Origins. *J. Plant Biochem. Biotech*. **6**:1-7.

Kunnimalaiyaan, M., Shi F., and Nielsen B.L. (1997) Anaysis of a tobacco chloroplast DNA replication origin (*ori*B) downstream of the 23S rRNA gene. *J. Mol. Biol.* **268**:273-283.

Lei M. and Tye B.K. (2001) Initiating DNA synthesis: from recruiting to activating the MCM complex. *J. Cell Sci.* **114**:1447-1454.

Lou J.K., Wu M., Chang C.H., Cuticchia A.J. (1987) Localization of a r-protein gene within the chloroplast DNA replication origin of Chlamydomonas. *Curr. Genet*. **11**(6-7):537-541.

Lu Z., Kunnimalaiyaan M., Nielsen B.L. (1996) Characterization of replication origins flanking the 23S rRNA gene in tobacco chloroplast DNA. *Plant Mol. Biol.* **32**:693-706.

Lugo S.K., Kunnimalaiyaan M., Singh N.K., and Nielsen B.L. (2004) Required sequence elements for chloroplast DNA replication activity in vitro and in electroporated chloroplasts. *Plant Science* **166**:151-161.
Messer W. (2002) The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Micro. Rev.* **26**:355-374.
Messer W., Blaesing F., Majka J., Nardmann J., Sigrid S., Schmidt A., Seitz H., Speck C., Tungler D., Wegrzyn G., Weigel C., Welzeck M., and Zakrzewska-Czerwinska J. (1999) Functional domains of DnaA proteins. *Biochimie* **81**:819-825.
Meyer R.R. and Laine P.S. (1990) The Single-Stranded DNA-Binding Protein of Escherichia coli. *Microbiol. Reviews* **54**(4):342-380.
Moraes C.T. (2001) What regulates mitochondrial DNA copy number in animal cells? *Trends in Genetics* **17**(4):199-205.
Muhlbauer S.K., Lossl A., Tzekova L., Zou Z., Koop H.U. (2002) Functional analysis of plastid DNA replication origins in tobacco by targeted inactivation. *Plant J.* **32**(2):175-184.
Murakami S., Kondo Y., Nakano T., and Sato F. (2000) Protease activity of CND41, a chloroplast nucleoid DNA-binding protein, isolated from cultured tobacco cells. *FEBS Letters* **468**:15-18.
Nakano T., Sato F., and Yamada Y. (1993) Analysis of Nucleoid-Proteins in Tobacco Chloroplasts. *Plant Cell Physiol.* **34**(6):873-880.
Nemoto Y., Nagata T., and Kuroiwa T. (1989) Studies on Plastid-Nuclei (Nucleoids) in Nicotiana tabacum L. II. Disassembly and Reassembly of Proplastid-Nuclei Isolated from Cultured Cells. *Plant Cell Physiol.* **30**(3):445-454.
Oleskina Y.P., Yurina N.P., Odintsova T.I., Egorov Ts.A., Otto A., Wittmann-Liebold B., and Odintsova M.S. (1999) Nucleoid Proteins of Pea Chloroplasts: Detection of a Protein Homologous to Ribosomal Protein. *Biochem. Mol. Biol. Inter.* **47**(5):757-763.
Palmer J.D., Singh G.P., and Pillay D.T.N. (1983) Structure and Sequence Evolution of Three Legume Chloroplast DNAs. *Mol. Gen. Genet.* **190**:13-19.
Palmer J.D., Osorio B., Aldrich J., and Thompson W.F. (1987) Chloroplast DNA evolution among legumes: Loss of a large inverted repeat occurred prior to other sequence rearrangements. *Curr. Genet.* **11**:275-286.
Palmer J.D., Osorio B., and Thompson W.F. (1988) Evolutionary significance of inversions in legume chloroplast DNAs. *Curr. Genet.* **14**:65-74.
Peeters N. and Small I. (2001) Dual targeting to mitochondria and chloroplasts. *Biochim. Biophys. Acta* **1541**:54-63.
Ryan K.R. and Shapiro L. (2003) Temporal and spatial regulation in prokaryotic cell cycle progression and development. *Annu. Rev. Biochem.* **72**:367-394.
Sato N., Misumi O., Shinada Y., Sasaki M., and Yoine M. (1997) Dynamics of localization and protein composition of plastid nucleoids in light-grown pea seedlings. *Protoplasma* **200**:163-173.
Sato N., Ohshima K., Watanabe A., Ohta N., Nishiyama Y., Joyard J., and Douce R. (1998) Molecular Characterization of the PEND Protein, a Novel bZIP Protein Present in the Envelope Membrane That Is the Site of Nucleoid Replication in Developing Plastids. *Plant Cell* **10**:859-872.

Sato N., Nakayama M., and Hase T. (2001) The 70-kDa major DNA-compacting protein of the chloroplast nucleoid is sulfite reductase. *FEBS Letters* **487**:347-350.
Sekine K., Hase T., and Sato N. (2002) Reversible DNA Compaction by Sulfite Reductase Regulates Transcriptional Activity of Chloroplast Nucleoids. *J. Biol. Chem.* **277**(27):24399-24404.
Shevchenko A., Wilm M., Vorm O., Mann M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**(5):850-858.
Sheveleva E.V. and Hallick R.B. (2004) Recent horizontal intron transfer to a chloroplast genome. *Nucleic Acids Research* **32**(2):803-810.
Shinozaki K., Ohme M., Tanaka M., Wakasugi T., Hayashida N., Matsubayashi T., Zaita N., Chunwongse J., Obokata J., Yamaguchi-Shinozaki K., Ohto C., Torazawa K., Meng B.Y., Sugita M., Deno H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A., Tohdoh N., Shimada H., Suguira M. (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.* **5**:2043-2049.
Silva-Filho M.C. (2003) One ticket for multiple destinations: dual targeting of proteins to distinct subcellular locations. *Curr. Opin. Plant Biol.* **6**:589-595.
Tomlinson I.P., Alam N.A., Rowan A.J., Barclay E., Jeager E.E., Kelsell D., Leigh I., Gorman P., Lamlum H., Rahman S., Roylance R.R., Olpin S., Bevan S., Barker K., Hearle N., Houlston R.S., Kiuru M., Lehtonen R., Karhu A., Vilkki S., Laiho P., Eklund C., Vierimaa O., Aittomaki K., Hietala M., Sistonen P., Paetau A., Salovaara R., Herva R., Launonen V., Aaltonen L.A.; Multiple Leiomyoma Consortium (2002) Germline mutations in FH predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer. *Nat. Genet.* **30**(4):406-410.
Wall M.K., Mitchenall L.A., and Maxwell A. (2004) Arabidopsis thaliana DNA gyrase is targeted to chloroplasts and mitochondria. *Proc. Natl. Acad. Sci. USA* **101**(20):7821-7826.
Wu M., Lou J.K., Chang D.Y., Chang C.H., and Nie Z.Q. (1986) Structure and function of a chloroplast DNA replication origin of Chlamydomonas reinhardtii. *Proc. Natl. Acad. Sci. USA* **83**:6761-6765.
Wu M., Nie Z.Q., and Yang J. (1989) The 18-kD Protein That Binds to the Chloroplast DNA Replicative Origin Is an Iron-Sulfur Protein Related to a Subunit of NADH Dehydrogenase. *Plant Cell* **1**:551-557.
Wu M., Chang C.-H., Yang J., Zhang Y., Nie Z.Q., and Hsieh C.-H. (1993) Regulation of chloroplast DNA replication in Chlamydomonas reinhardtii. *Bot. Bull. Acad. Sin.* **34**:115-131.
Zhu W., Rainville I.R., Ding M., Bolus M., Heintz N.H., and Pederson D.S. (2002) Evidence that the pre-mRNA splicing factor Clf1p plays a role in DNA replication in Saccharomyces cerevisiae. *Genetics* **160**(4):1319-1333.