Final Technical Report for DOE grant DE-FG02-02ER15352 (9/1/02- 8/31/09;
\$660,000 total; Factor-Assisted RNA Splicing and Mobile Chloroplast Introns)

(from the Chemical Sciences, Geosciences, and Energy Biosciences Division, U.S. Department of Energy, Dr. Richard Greene, Program Manager, email: richard.greene@science.doe.gov, phone: 301-903-6190)

Prepared and submitted (4/21/2011) by:

Dr. David L. Herrin, Professor
Section of Molecular Cell and Developmental Biology and Institute for Cellular and Molecular Biology, School of Biological Sciences, University of Texas at Austin, Austin, TX 78712
email: herrin@mail.utexas.edu
tel: 512-471-3843

Contents:	Page #
I. Summaries of some of the main projects supported by the grant	2
A. Group I introns in the chloroplast: light-regulated splicing, intron mobility and repair of DSBs induced by intron-encoded proteins	
B. Group II introns in the chloroplast of cold-tolerant species	4
C. Chloroplast circadian transcription: Role of a sigma factor	5
II. Complete list of publications supported by the grant	6
III. Educational and training impact of the grant	7

I. Summaries of some of the main projects supported by the grant

I-A. Group I introns in the chloroplast: light-regulated splicing, intron mobility and repair of DSBs induced by intron-encoded proteins

Due in part to DOE support, we have developed a fairly comprehensive view of the introns found in the *Chlamydomonas reinhardtii* chloroplast genome. Organellar introns have been of considerable interest since their discovery in the late 1970s. They are typically of two main types (Group I and Group II), and both are very complex ribozymes whose folding *in vivo* is assisted by proteins (which are sometimes those encoded by the intron). Many introns of both types are also mobile genetic elements, whose retention within, or invasion of, host genes is usually promoted by proteins encoded by the intron. In *C. reinhardtii*, there are 5 such introns in the chloroplast - all of which are classified as subtypes of Group I. There are 4 introns in the *psbA* gene, and 1 in the 23S LSU rRNA gene. All of these introns are bona fide ribozymes that can self-splice *in vitro* without proteins. Splicing of the *psbA* introns is strongly promoted by light, using an unknown mechanism that requires photosynthetic electron transport. Two of the *psbA* introns (2 and 4) and the LSU intron encode endonucleases which mobilize the introns *in vivo* by cleaving intronless alleles of the host gene. The Enase encoded by the LSU intron is called I-CreI and the one encoded by the *psbA4* intron is called I-CreII. We discovered and studied both of these highly-specific rare-cutting endonucleases (Kim et al. 2005; Corina et al. 2009), and have provided them to ~40 labs over the years for their own studies, or to use for chromosome engineering.

Specific summaries of some of the Group I-intron related projects performed under this grant are described below:

I-A-1. Mutagenesis of a light-regulated *psbA* intron reveals the importance of efficient splicing for photosynthetic growth (J. Lee and D.L. Herrin, *Nucleic Acids Research* 31, 4361-4372, 2003).

The chloroplast-encoded *psbA* gene encodes the D1 polypeptide of the photosystem II reaction center, which is synthesized at high rates in the light. In *Chlamydomonas reinhardtii*, the *psbA* gene contains four self-splicing group I introns whose rates of splicing *in vivo* are increased at least 6-fold to 10-fold by light. However, because *psbA* is an abundant mRNA, and some chloroplast mRNAs appear to be in great excess of what is needed to sustain translation rates, the developmental significance of light-promoted splicing has not been clear. To address this and other questions, potentially destabilizing substitutions were made in several predicted helices of the fourth *psbA* intron, *Cr.psbA4*, and their effects on *in vitro* and *in vivo* splicing assessed. Two-nucleotide substitutions in P4 and P7 were necessary to substantially reduce splicing of this intron *in vivo*, although most mutations reduced self-splicing *in vitro*. The P7-4,5 mutant, whose splicing was completely blocked, showed no photoautotrophic growth and synthesis of a truncated D1 (exons 1-4) polypeptide from the unspliced mRNA. Most informative was the P4'-3,4 mutant, which exhibited a 45% reduction in spliced *psbA* mRNA, a 28% reduction in synthesis of full-length D1, and an 18% reduction in photoautotrophic growth. These results indicate that *psbA* mRNA is not in great excess, and that highly efficient splicing of *psbA* introns, which is dependent on photosynthetic electron transport, is necessary for optimal photosynthetic growth.

I-A-2. The CYT-18 homologue from *Chlamydomonas* is probably not a group I intron splicing factor (A. Yerrabelli, W. Qiu and D.L. Herrin, unpublished results)

A routine search of the *C. reinhardtii* EST databases revealed a translated cDNA sequence with similarity to the N-terminus of the CYT-18 protein from *Neurospora crassa*. CYT-18 is a bi-functional tyrosyl-tRNA synthetase that promotes splicing of several group I introns in the mitochondria, as well as the *Chlamydomonas Cr.LSU* intron *in vitro* (G. Wallweber and A. Lambowitz, pers. comm.). We obtained (from Kazusa) and sequenced the cDNA clone (~1.8 kb) for this EST, which appeared to encode all of

the protein. Bioinformatic analysis confirmed that it is a bacterial-like tyrosyl-tRNA synthetase, likely of mitochondrial origin, but as no other bacterial-like tyrosyl-tRNA synthetases were found in the *Chlamydomonas* genome (JGI version 2.0) or EST databases, it is probably targeted to both mitochondria and chloroplasts [which is not uncommon for tRNA synthetases (reviewed in Peeters and Small, 2001, *BBA* 1541: 54)]. Also, several programs (PSORT, TARGETP, etc.) predicted an ambiguous (either chloroplast or mitochondrial) targeting peptide at the N-terminus. Although secondary structure predictions indicated that there are 3 idiosyncratic N-terminal alpha helices (following the targeting peptide), just like CYT-18 [these helices are necessary for splicing stimulation by CYT-18, Mohr et al., 2001], it was not capable of acting like CYT-18 in an in vivo group I intron splicing assay [which relies on the ability of CYT-18 to suppress mutations in the thymidylate synthetase group I intron from T4 phage *T4.td* (Mohr et al., 1992)]. Thus, we think it is unlikely that the *Chlamydomonas* CYT-18 homologue also acts as a group I intron splicing factor. This work was not published but the cDNA sequence will be annotated and deposited in Genbank.

I-A-3. Mapping of nuclear suppressors of *C. reinhardtii* that promote splicing of deficient chloroplast introns (Chang et al., 2003; L. Liu and D.L. Herrin, unpublished results)

A major goal of the grant was to clone and characterize one of the nuclear suppressors of splicing-defective chloroplast introns that we isolated previously (Li et al., 2002). We isolated and characterized genetically 3 such suppressors that represented at least 2 different nuclear genes (called *css1-css2*); two of them were also shown to be dominant in diploids. These genes suppress destabilizing mutations in the *psbA4* and LSU rRNA introns, and, if isolated, could provide much insight into nuclear control of organellar splicing.

Our initial strategy to clone one of these genes was to use complementation of the primary LSU splicing mutant with a genomic cosmid library made with DNA from one of the suppressors (7151 in this case). In this approach, the LSU splicing mutant, which is also light-sensitive, is transformed with pools of cosmid DNAs from an arrayed library made with 7151 DNA. A cosmid vector was modified so that it would confer on *Chlamydomonas* transformants, resistance to bleomycin (and zeocin) (Chang et al., 2003). However, we were unable to get a decent transformation rate of the splicing-deficient mutant with genomic cosmid clones despite trying a number of variations of existing methods, and recreating a new recipient strain. Another problem was the primary intron mutant had a measurable spontaneous reversion rate that made our job even more difficult. We finally abandoned that approach in favor of map-based cloning, which has become easier in *Chlamydomonas* because of the genome project.

In this strategy, the suppressor mutants (which are mating type + and carry the splicing-deficient chloroplast LSU rRNA intron) are crossed with a highly polymorphic strain (S1D2), and the progeny are analyzed using molecular markers and PCR. Because the S1D2 strain is highly polymorphic compared to the common lab strains (e.g., 137c), the cross with our 7151 suppressor yielded multiple phenotypes. Fortunately, crosses of S1D2 to the 7120 and 71N1 suppressors yielded 2 phenotypes similar to the backcrosses with the parental strain. Liming Luo mapped both of these mutants to the same region on chromosome III. Unfortunately, the suppressor mutations mapped to a region that is devoid of recombination, and yet is quite large, ~450,000 bp. Over 70 genes are in this region, including 15-20 that might be organelle targeted. She sequenced about 10 candidate genes but did not find the mutation.

I-A-4. Using the protein from the *psbA4* intron, I-CreII, to study chloroplast DNA repair and recombination (Odom et al., 2008; Kwon et al., 2010)

Two of the *psbA* introns and the 23S rRNA intron are homing introns – so called because they invade intron-minus alleles with high efficiency. Homing is initiated by the intron's own endonuclease that makes a double-strand break (DSB) in the recipient DNA. The free DNA ends are highly recombinogenic

and can invade homologous, intron-donor DNA or can interact intramolecularly to repair the DSB. We completed the first systematic study of the repair and consequences of a DSB in chloroplast DNA (Odom et al., 2008), by using I-CreII to cleave the chloroplast genome at its native target in the *psbA* gene. The main conclusions were that the chloroplast of *Chlamydomonas* can use 3 homologous recombination (HR) pathways to repair a DSB, and that it can also use short dispersed repeats (SDRs) located proximally, distally or even on separate molecules, to repair the break. The findings also provided a new rationale for the extensive repertoire of repeated sequences in this genome.

In the aforementioned study (MS in Appendix), the Enase was co-transformed into the chloroplast with a selectable marker and cpDNA was analyzed to detect repair events. This approach would not be straightforward in most plants, because of the difficulty in getting chloroplast transformants. It would be preferable then to express the gene in the nucleus and target the protein to the chloroplast; it would also be nice if the Enase was inducible. We developed such a system in *Arabidopsis* using the transit peptide of *rbcS* to target I-CreII to the chloroplast, where it cut the *psbA* gene (Kwon et al., 2010). The *Arabidopsis* cpDNA is very repeat-poor compared to *Chlamydomonas*. The results provided evidence of alternative repair pathways in this chloroplast that resemble the MMEJ (Microhomology-mediated end joining) and NHEJ pathways of the nucleus, in terms of their minimal or no homology requirement.

I-B. Group II introns in the chloroplast of cold-tolerant *Chlamydomonas* spp.

Group II introns are found in diverse genomes, including bacteria, archaea, mitochondria, and especially land plant chloroplasts. Group II introns have contributed to our understanding of large ribozymes, nuclear mRNA splicing, and mobile retroelements. However, primary publications on the proteins encoded by group II introns are sparse compared to the ribozyme component. The *Lactococcus* LtrA protein is the only group II intron-encoded protein that has been studied biochemically with purified protein. The *Lactococcus* LtrA protein has reverse transcriptase (RT), maturase (Mat), and endonuclease (En) domains, in that order (RT-Mat-En). When reconstituted with intron RNA, the RNP catalyzed reverse splicing into the sense strand of double-stranded target DNA (fused Ltr exons), followed by copying of the RNA into cDNA using the antisense 3'-exon as primer; the primer is created by the endonuclease activity. LtrB is thus a mobile intron related to retrotransposons. However, invasion of an intronless allele (a.k.a. homing) is much more efficient than transposition.

Much less was known about the abilities of group II introns found in chloroplasts, or in the mitochondria of plants. We discovered group II introns in the chloroplast *psbA* gene of two cold-tolerant *Chlamydomonas* spp. that we have suggested invaded from bacteria.

I-B-1. A horizontally acquired group II intron in the chloroplast *psbA* gene of a psychrophilic *Chlamydomonas*: *In vitro* self-splicing and genetic evidence for maturase activity (O.W. Odom, D.L. Shenkenberg, J. A. Garcia and D.L. Herrin, *RNA* 10:1097-1107)

Although the majority of known group II introns were from chloroplast genomes, the first self-splicing group II intron from a chloroplast gene was reported just a week or so before this paper was submitted; the *Euglena* work was published by Sheveleva and Hallick (2004, *NAR* 32: 803). Briefly, we discovered a large (2.6-kb) group II intron (*Chs.psbA1*) in the *psbA* gene of a psychrophilic *Chlamydomonas* sp. that self-splices accurately *in vitro*. Remarkably, this intron, which also encodes an ORF with putative reverse transcriptase (RT), maturase (Mat), and endonuclease (En) domains, is in the same location, and is related to the *E. myxocylindracea* intron discovered by Sheveleva and Hallick (2004), as well as to group IIB2 introns from cyanobacteria. *In vitro* self-splicing of *Chs.psbA1* occurred via a lariat, and required Mg^{2+} (>12 mM) and NH_4^+ . Self-splicing was improved by deleting most of the ORF, and by using pre-RNAs directly from transcription reactions, suggestive of a role for folding during transcription. Self-splicing of *Chs.psbA1* showed a temperature optima of ~44EC, but was nearly absent at 50EC, indicative of thermolability. Splicing of *Chs.psbA1* also occurred in *E. coli*, but not when the ORF was disrupted by mutations, providing genetic evidence that it has maturase activity. This work provided the first

description of a ribozyme from a psychrophilic organism. It also appeared to provide a second instance of inter-kingdom horizontal transfer of this group IIB2 intron (or a close relative) from cyanobacteria to chloroplasts.

I-B-2. Functional analysis of the protein encoded by the group II intron in the *psbA* gene of the psychrophilic *C. rauidensis* from Antarctica (Odom, O.W. and D.L. Herrin, manuscript in preparation)

Based on multiple sequence alignments, the 600-amino acid *psbA1*-encoded protein (PsbA1p) seemed to have all 3 of the domains found in the longest forms of the protein encoded by group II introns, RT-Mat-En (reverse transcriptase-maturase-endonuclease). We endeavored to study the full-length protein expressed and purified from *E. coli* even though it was expected to be difficult to do. Early attempts using an intein tag (as was used for the bacterial LtrA protein) resulted in an insoluble protein regardless of induction conditions, and GroEL chaperone co-overexpression did nothing. We solubilized the inclusion bodies and tried to refold the protein *in vitro* using a number of refolding methods, but to no avail; the protein always precipitated when the refolding reagents were diluted. By turning to maltose-binding protein (MBP) as an affinity tag, we were able to get a small fraction (1-2%) of the fusion protein soluble. This protein, purified by affinity chromatography, had a robust reverse transcriptase (RT) activity (assayed with a synthetic RNA-DNA hybrid substrate) that was stable at 37°C (under assay conditions). Maturase activity was tested using a variety of conditions with precursor RNA, but was not clearly detected. We did obtain some evidence of specific binding to the precursor RNA, however, using filter-binding assays. Gel filtration chromatography indicated that the protein was much larger than a dimer. We also successfully expressed PsbA1p as a fusion with the trigger factor chaperone (in pColdTF). Although this fusion gave a much greater fraction of the protein in soluble form, its RT activity was less than the MBP fusion. This was the first chloroplast protein shown to have RT activity. This result also supports our hypothesis that this intron, following its successful invasion of the *psbA* gene from a bacterial donor, catalyzed the removal of the group I introns usually found in the *Chlamydomonas psbA* gene. This work has not yet been published, as we have been acquiring other types of data in order to make a stronger picture of intron evolution in the chloroplast.

I-C. Structure, circadian regulation and bioinformatic analysis of the unique sigma factor gene in *Chlamydomonas reinhardtii* (Carter, M.L., A. Smith, H. Kobayashi, S. Purton and D.L. Herrin (2004) *Photosynthesis Research* 82, 339-349; Kawazoe, R., K. Mahan, M. Carter and D.L. Herrin, manuscript in preparation)

In higher plants, the transcription of plastid genes is mediated by at least two types of RNA polymerase (RNAP); a plastid-encoded eubacterial-type RNAP in which promoter specificity is conferred by sigma factors encoded by a nuclear multigene family, and a single-subunit bacteriophage-type RNAP encoded in the nucleus. In contrast, green algae such as *Chlamydomonas reinhardtii* appear to possess only the bacterial-type enzyme. As we have shown previously that transcription of much, if not most, of the chloroplast genome in *C. reinhardtii* is regulated by the circadian clock, we sought to identify sigma factor genes that may be responsible for this nuclear regulation. In this paper, we described the characterization of a nuclear gene (*rpoD*) encoding a putative sigma factor. In addition to a predicted chloroplast transit peptide at the N-terminus, the C-terminal half of the RPOD protein contains the conserved motifs (2.1-4.2) diagnostic of bacterial sigma-70 factors. We also identified two motifs not previously recognized for sigma factors, adjacent PEST sequences and a leucine zipper, both suggested to be involved in protein-protein interactions. PEST sequences were also found in ~40% of sigma factors examined, indicating they may be of general significance. Genomic southern blots, and BLAST searches of the draft genome and EST databases, suggested that there is only one sigma factor-like gene in *C. reinhardtii*, rather than the multiple genes found in higher plants. The expression of *RPOD* mRNA showed a circadian rhythm with the maximum level occurring in the mid-to-late dark period, prior to the

peak in chloroplast transcription. The results suggested that circadian clock-regulated expression of the nuclear gene, *rpoD*, contributes to the clock's control of chloroplast transcription.

Subsequently, we made an antibody to a synthetic peptide of RPOD and used it to determine whether the protein level fluctuated, or if it might be phosphorylated, in a cyclic manner. The answer to both questions was no, although we have not published this yet. We also developed a procedure for extracting the chloroplast RNA polymerase from cells and assaying its core activity with denatured heterologous DNA. This was done during a light-dark cycle, and the results will be published together with the RPOD antibody analysis.

II. Publications resulting from this grant (and acknowledged in the respective pubs)

1. Chang, M.P., F. Li, O.W. Odom, J. Lee and D.L. Herrin. 2003. A Cosmid Vector Containing a Dominant Selectable Marker for Cloning *Chlamydomonas* Genes by Complementation. *Plasmid* 49, 75-78.
2. Lee, J. and D.L. Herrin. 2003. Mutagenesis of a Light-Regulated *psbA* Intron Reveals the Importance of Efficient Splicing for Photosynthetic Growth. *Nucleic Acids Research* 31, 4361-4372.
3. Odom, O.W., D. Shenkenberg, J. Garcia and D.L. Herrin. 2004. A Horizontally Acquired Group II Intron in the Chloroplast *psbA* Gene of a Psychrophilic *Chlamydomonas*: In Vitro Self-Splicing and Genetic Evidence for Maturase Activity. *RNA* 10, 1097-1107.
- *4. Herrin, D.L. and J. Nickelsen. 2004. Chloroplast RNA Processing and Stability. *Photosynthesis Research* 82, 301-314.
5. Carter, M.L., A. Smith, H. Kobayashi, S. Purton and D.L. Herrin. 2004. Structure, Circadian Regulation and Bioinformatic Analysis of the Unique Sigma Factor Gene in *Chlamydomonas reinhardtii*. *Photosynthesis Research* 82, 339-349.
6. Kim, H-H., L. Corina, J-K. Suh, and D.L. Herrin. 2005. Expression, Purification, and Biochemical Characterization of the Intron-Encoded Endonuclease, I-CreII. *Protein Expression and Purification* 44, 162-172.
- *7. Misquitta, R. and D.L. Herrin. 2005. Circadian Regulation of Chloroplast Transcription: A Review. *Plant Tissue Culture* 15, 83-101.
8. Mahan, K.M., O.W. Odom, and D.L. Herrin. 2005. Controlling Fungal Contamination in *Chlamydomonas reinhardtii* cultures. *Biotechniques* 39, 457-458.
9. Kuo, T-C., O.W. Odom and D.L. Herrin. 2006. Unusual Metal Specificity and Structure of the Group I Ribozyme from *Chlamydomonas reinhardtii* 23S rRNA. *FEBS Journal* 273, 2631-2644.
10. Bao, Y. and D.L. Herrin. 2006. Mg⁺² Mimicry in the Promotion of Group I Ribozyme Activities by Aminoglycoside Antibiotics. *Biochemical and Biophysical Research Communications* 344, 1246-1252.
11. Zicker, A.A., C.S. Kadakia and D.L. Herrin. 2007. Distinct Roles for the 5' and 3' Untranslated Regions in the Degradation and Accumulation of Chloroplast *tufA* mRNA: Identification of an Early Intermediate in the In Vivo Degradation Pathway. *Plant Molecular Biology* 63, 689-702.
12. Odom, O.W., K-H. Baek, R.N. Dani and D.L. Herrin. 2008. *Chlamydomonas* Chloroplasts Can Use Short Dispersed Repeats (SDRs) and Multiple Pathways to Repair a Double-Strand Break in the Genome. *Plant Journal* 53, 842-853.
- *13. Herrin, D.L. 2009. Chloroplast RNA Processing and Stability. In: *The Chlamydomonas Sourcebook*, 2nd Ed., Vol. 2 (D. Stern, E. Harris, eds.), Academic Press, pp. 937-965.
14. Corina, L.E., W. Qiu, A. Desai and D.L. Herrin. 2009. Biochemical and mutagenic analysis of I-CreII reveals distinct but important roles for both the H-N-H and GIY-YIG motifs. *Nucleic Acids Research* 37, 5810-5821.
15. Odom, O.W. and D.L. Herrin. 2010. An Unprecedented Group II Intron from *Chlamydomonas Subcaudata* (Chlorophyceae) that has Two Large Open Reading Frames – Potentially Encoding a

Reverse Transcriptase-Maturase-Endonuclease and a Novel Protein. *Journal of Phycology* 46, 969-973.

16. Kwon, T., E. Huq and D.L. Herrin. 2010. Microhomology-Mediated and Nonhomologous Repair of a Double-Strand Break in the Chloroplast Genome of *Arabidopsis*. *Proceedings of the National Academy of Sciences USA* 107, 13954-13959.

* Invited review article or book chapter

III. Educational and training impact of the grant

Students provided a research experience and training:

High school students

Crystal Kadakia, Texas Academy of Science & Math, 2004

Juan Belman, Austin High School, 2009 (in ACS-sponsored Project SEED, for under-represented groups)

Undergraduates (in chronological order)

Jerry Chen, Rhyne Simon, Meredith Allen, David Shenkenberg, Josh Garcia, Amy Desai, Keren Hilgendorf, Rosina Amua-Sekyi, Chantal Lucia, Crystal Kadakia, John Liao, Kristina Mahan, Grace Tsieu, David Fudman, Sean Grunsten, Julie Kim, Smitha Shivaprasad, Brad Venghaus, Wayne Bowden, John Nguyen, David Kim

Graduate students (degree, year received, and current position)

Jaesung Lee, PhD Plant Biology, 2003 (Research Scientist, UC-San Diego)

Rachel Misquitta, MA Microbiology, 2004 (Graduate student, NYU Medical College)

Anitha Yerrabelli, MA Microbiology, 2004 (Technician, NIH, Bethesda)

Alicia Zicker, PhD Biology, 2005 (Applications Specialist, Thermo Scientific)

Laura Corina, PhD Plant Biology, 2005 (Assoc. Dean, Quincy College)

Matthew Carter, MA Biology, 2006 (Production Supervisor, Agilent)

Weihua Qiu, PhD Plant Biology, 2010 (Postdoctoral Fellow, Columbia U)

Liming Luo, PhD Plant Biology, 2010 (Research Associate, Texas Tech Medical School)

Taegun Kwon, PhD Plant Biology, 2011 (accepted postdoc at UC-Berkeley w/S. Luan)

Postdoctoral Research Associate:

Dr. O.W. Odom

Visiting Scholar:

Ananya Roy, M.S., U of Texas-Dallas, 2009