

United States Department of Energy, Form DOE F241.1, Section K. Description/Abstract  
For: Final Technical Report of DOE grant number DE-FG02-92ER20086.

Title: The plant mitochondrial *mat-r* gene/*nad1* gene complex

Principal Investigator: David R. Wolstenholme, Professor of Biology, University of Utah, Salt  
Lake City, Utah 84112

Total Grant Period: 7-10-92 to 7-9-96

RECEIVED  
JUL 02 1993  
OSTI

### Abstract

We have completed sequencing segments of the maize mitochondrial (mt) DNA that contains all five of the exons (A-E) of the gene (*nad1*) for subunit I of the respiratory chain NADH dehydrogenase. Analysis of these sequences indicates that exons B and C are joined by a continuous group II intron, but the remaining exons are associated with partial group II introns and are encoded at widely separated locations in the maize mtDNA molecule. Production of mature *nad1* gene transcripts is postulated to involve both *cis*-splicing of the b/c intron and *trans*-splicing of each of the other three introns. In the partial d/e intron upstream from exon E is encoded a gene (*mat-r*) for a protein related to fungal maturases. From comparison of the maize split d/e intron, and the soybean continuous d/e intron it was found that the location at which the maize d/e intron underwent rearrangement is marked by a repeated 8 nt sequence on each intron half. Interestingly, there is a 9 or 10 nt repeat on each of the two halves of the a/b and c/d introns. We have shown that mature transcripts of the maize *nad1* gene contain 23 edited (C → U) nucleotides and that transcripts of maize and soybean *mat-r* genes contain 15 and 14 C → U edits, respectively. The majority of edits in *nad1* transcripts result in amino acid replacements that increase similarity between the maize NAD1 protein and NAD1 proteins of other plant species and of animal species. Edits in *mat-r* transcripts result in an increase in amino acid similarity of the maize and soybean *MAT-R* proteins to each other, and of each of these proteins to fungal and bryophyte maturases.

We found that the intron between exons b and c is not edited. Edits in other introns are rare and have so far been detected only in domain 6 regions. Unspliced transcripts of exon C and of exon E are partially edited, confirming that editing can precede both *cis*- and *trans*-splicing.

From data obtained using PCR and sequencing we have shown that transcripts containing all possible exon combinations exist in maize mitochondria. This supports the view that splicing of maize *nad1* exons can occur in any order, and that *trans*-splicing of exons can occur in advance of *cis*-splicing of exons B and C. Also, we obtained evidence that the relative rates of splicing are greater for B+C and C+D exons, than for A+B and D+E exons.

Attempts made to induce *cis*-splicing *in vitro* of the B and C exons by excision of the continuous intron between them have been unsuccessful.

Comparisons of *nad1*, *nad2* and *nad5* intron sequences have been made to gain information on intron interrelationships.

We have no objection from a patent standpoint to the publication or dissemination of this material.

1  
M. P. Dvorsky  
Office of Intellectual  
Property Counsel  
DOE Field Office, Chicago  
6/29/93  
Date

## **DISCLAIMER**

**This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.**

## **DISCLAIMER**

**Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.**

A pseudo-*mat-r* gene was detected (using degenerate primers and PCR reactions), sequenced, and mapped on the maize (normal) mtDNA molecule.

United States Department of Energy, Form DOE F241.1

For: Final Technical Report of DOE grant number DE-FG02-92ER20086.

Title: The plant mitochondrial *mat-r* gene/*nad1* gene complex

Principal Investigator: David R. Wolstenholme, Professor of Biology, University of Utah, Salt Lake City, Utah 84112

Total Grant Period: 7-10-92 to 7-9-96

### Objectives of the Proposed Work

The major objective for the work outline at the time of the original application was to learn about the structure and functioning of the highly complex mitochondrial gene for NADH dehydrogenase subunit I (*nad1*) from maize and soybean. We proposed to complete sequencing of all exons and associated complete and partial group II introns. Experiments were described that would allow us to learn about RNA editing in introns and in the maturase-related (*mat-r*) gene contained in one of the introns, and how editing might influence intron splicing. We also proposed to look for conditions that would permit self-splicing *in vitro* of the various group II introns of the *nad1* gene. Following, we summarize the progress we made towards achieving these objectives. A further aim included in the original application was to synthesize in *Escherichia coli* cells, a *MAT-R* protein expressed from a maize or soybean *mat-r* gene derived from edited cDNAs. We did not make any significant progress towards accomplishing this objective due to both funding and time constraints.

### Introduction

The mitochondria of angiosperm plants, like the mitochondria of all other eukaryotes, contain their own DNA genome. Angiosperm mitochondrial (mt) genomes contain about 60 genes that show only minor differences in occurrence among species. The mt-protein genes are as follows: *coxI-III* (cytochrome *c* oxidase subunits I-III); *cob* (cytochrome *b*) *atp6* and *atp9* (subunits 6 and 9 of the  $F_0$  segment of the ATP synthetase) and *atp1* or *A* (subunits 1 or A of the  $F_1$  segment of ATP synthetase); *nad1-7, 9* and *4L* (respiratory chain NADH dehydrogenase subunits 1-7, 9 and 4L); *rps1-4, 7, 10, and 12-14* (proteins of the small subunit of the mt-ribosome) and *rpl2, 5, and 16* (proteins of the large subunit of the mt-ribosome); up to 5 genes for proteins involved in cytochrome *c* biogenesis (References in Wolstenholme and Fauron, 1995; Brennicke et al., 1996). There also occurs in angiosperm mtDNAs a number of large, as yet unidentified open reading frames. Angiosperm mtDNAs also contain genes for the three structural RNA components (5S, 18S and 26S) of mitochondrial ribosomes, and for between 16 and 22 tRNAs (References in Gray, 1992; Dietrich et al., 1992; Marechal-Drouard et al., 1993, 1995).

Data from restriction mapping analyses indicate that mtDNAs of a variety of both monocotyledons (including maize) and dicotyledons occur as a collection of circular molecules: the so-called masters circle (or master chromosome) that contains the entire

sequence complexity of mtDNA, and a series of smaller circular molecules that seem to be derived from the master circle by recombination between large directly repeated sequences, and therefore contain subsets of the master circle sequence (Lonsdale et al., 1983a, 1984; Palmer and Shields, 1984; Fauron et al., 1995).

In almost all angiosperm mt-protein gene transcripts some cytidine (C) residues are changed to uridines (U), a process termed RNA editing (Gualberto et al., 1989; Covello and Gray, 1989; Heisel et al., 1989; Gray and Covello, 1993).

One or more group II introns occurs in a number of angiosperm mt-protein genes. Angiosperm *nad1*, *nad2* and *nad5* genes each comprise five exons separated by group II introns (references in Wolstenholme and Fauron, 1995; Brennicke et al., 1996). Some of the group II introns separating these exons are continuous. However, others are split into two components, located with adjacent exons at widely separated positions in the mtDNA molecule. It is postulated that generation of mature transcripts involves a *trans*-splicing mechanism (Wolstenholme et al., 1990, 1993; Wissinger et al., 1990, 1991; Bonen et al., 1990; Conklin et al., 1990; Chapdelaine and Bonen, 1991), such as that earlier described for chloroplast genes of *Chlamydomonas*, *Nicotiana* and *Marchantia* (Choquet et al., 1988; Hildebrand et al., 1988; Goldschmidt-Clermont et al., 1991).

#### Report of Progress during the entire grant period (7-10-92 to 7-9-96)

##### Structure of the maize *nad1* gene (Wolstenholme et al. 1993, and in preparation)

During the DOE grant period we completed sequencing of segments of maize mtDNA (totalling about 19 kb, both complementary strands) that collectively encode the five exons that make up the NADH dehydrogenase subunit 1 (*nad1*) gene. Exons B and C are joined by a continuous group II intron (intron b/c), but exons A, D and E are located on the maize normal (N-B37) 570 kb mtDNA molecule at sites widely separated from each other and from exons B and C. Partial group II introns are associated with all five exons. A similar exon-intron organization has been reported for the wheat and *Petunia nad1* genes (Chapdelaine and Bonen, 1991; Conklin et al., 1991). In broad bean, soybean, *Oenothera*, and *Arabidopsis*, the D and E exons are joined by a continuous group II intron (Wahleithner et al., 1990; Wolstenholme et al., 1990; Wissinger et al., 1990, 1991; Thomson et al., 1994; Brennicke et al. 1996).

By Southern blot hybridization, using exon-specific sequences and cloned cosmid restriction fragments, all of the *nad1* exons were mapped (in collaboration with Christiane M.-R. Fauron) on the normal (N-B37) maize mtDNA molecule. There are two copies of exon A (located in the 14 (kb) A repeat) and one copy each of exons B, C, D and E. The order of exons (and approximate separation in kb) on the maize mtDNA molecule is: A1 (74 kb) A2 (152 kb) E (46 kb) B+C (222 kb) D (76 kb). The direction of transcription of exons A1, B and C is the same, but opposite to that of exons A2 and E (the relative direction of transcription of exon D has not been determined). From considerations of the exon locations, the intron sequences

associated with the various exons, and examination of cDNAs of mature (fully spliced) *nad1* transcripts (see below) the following is postulated: exons A, B+C, D and E and associated partial group II introns are separately transcribed. The two halves of each divided intron then associate, probably by hydrogen bonding of complementary regions, to form a functional group II intron that is excised to yield correctly spliced exons. The entire latter process, earlier described for split group II introns in *Chlamydomonas*, *Nicotiana* and *Marchantia* chloroplast genes (Koller et al., 1987; Hildebrand et al., 1988; Kohchi et al., 1988; Michel et al., 1989; Goldschmidt-Clermont et al. 1991), is called *trans*-splicing (compared to *cis*-splicing: the joining of exons by excision of a continuous intron). A similar mode of production of mature transcripts has also been suggested for the *nad1* genes of wheat, *Oenothera*, *Petunia*, and *Arabidopsis* (Bonen et al., 1990; Wissinger et al., 1990, 1991; Conklin et al., 1990, 1991; Chapdelaine and Bonen, 1991; Brennicke et al. 1996).

A gene designated *mat-r* (first described by us in broad bean mtDNA [Wahleithner et al., 1990]) for a protein related to yeast group II intron maturases (Carignani et al., 1983) is contained within the partial group II intron upstream from the maize *nad1* exon E. In broad bean, soybean, *Oenothera*, and *Arabidopsis*, this gene is contained in the continuous intron between exons D and E (Wahleithner et al., 1990; Wolstenholme et al., 1990, 1993; Wissinger et al., 1991; Thomson et al., 1994). In *Petunia*, the *mat-r* gene is contained in the portion of the split d/e intron linked to exon D (Conklin et al., 1991).

*Repeated sequences in complementary group II intron segments.* Comparisons of the group II intron sequences 3' to the D exon and 5' to the E exon of maize with the continuous group II intron (d/e) that joins the D and E exons in soybean mtDNA indicate that the rearrangement event that separated the two halves of the maize d/e intron occurred about 65 ntp upstream from the *mat-r* gene. At this location in the maize D exon-associated and E exon-associated partial group II intron sequences is an identical 8 ntp sequence: 5' GCAGGAGG. Examination of the pairs of sequences postulated to be the two halves of the a/b intron, and the two halves of the c/d intron, revealed in each case the presence of a repeated sequence: 5' TGCTGTTTCC in each of the a/b intron halves and 5' GCTGGCTGG in each of the c/d intron halves. It will be worthwhile to seek further evidence concerning the possibility that these latter two repeats are markers of the locations of sequence rearrangements that resulted in splitting of the two introns.

#### Editing and processing of transcripts of the maize *nad1* gene/*mat-r* gene complex

*Editing of mature nad1 transcripts* (M.C. Thomson, C.T. Beagley and D.R. Wolstenholme, in preparation). From a pool of reverse transcriptase and random hexamer-generated maize mt-cDNAs, double-stranded cDNAs of mature (fully spliced) *nad1* transcripts were synthesized and amplified using a sense primer corresponding to a sequence upstream from exon A and an antisense primer to a sequence downstream from exon E. The PCR products were cloned into

M13 vectors and sequenced. Examination of these sequences confirmed the exact location of all exon splice-sites predicted from the gene (DNA) sequences. Comparisons of multiple cloned cDNA sequences with genomic DNA sequences revealed that mature *nad1* transcripts contained 25 uridines (U) that each correspond to a cytidine (C) in the *nad1* gene sequences. There are four C→U changes (edits) in exon A, zero in exon B, ten in exon C, one in exon D and ten in exon E. As found for many other mt-proteins (Gray and Covello, 1993), the majority of these C→U changes (21) result in amino acid replacements in the *NAD1* protein predicted from the genomic (DNA) sequence. These amino acid replacements improve similarity to the *NAD1* proteins of wheat, *Oenothera* and *Marchantia*, and to the *NAD1* proteins of various animals.

*Editing of mat-r transcripts* (Thomson et al., 1994). It has been shown that the *mat-r* gene is transcribed and the predicted amino acid sequences of the *mat-r* genes of different angiosperms are highly conserved (Wolstenholme et al., 1990; Wissinger et al., 1991; Chapdelaine and Bonen, 1991) suggesting that a functional protein is encoded. To further evaluate whether this is the case, we have analyzed maize and soybean *mat-r* transcripts for editing.

From pools of maize and soybean cDNAs, generated by reverse transcriptase using random hexamers, multiple cDNAs that collectively included transcripts of the entire *mat-r* gene were selectively PCR amplified using pairs of sense and antisense primers that bracket the *mat-r* gene sequence. The PCR products were cloned, deletion-cloned and sequenced. Nucleotide differences between cDNA (RNA) sequences and corresponding genome sequences were scored as edits only if detected in two or more (up to 25) independently derived clones (this is important because PCR amplification results in sequence alterations). All edits recorded in maize and soybean *mat-r* sequences (15 and 14, respectively) were C→U changes.

Of the 15 and 14 edits in maize and soybean *mat-r* transcripts, ten occurred at identical nucleotide positions, and all but four edits in maize and two edits in soybean resulted in an amino acid replacement. When a replacement edit occurred in one sequence but not the other, the resulting amino acid was always that specified by the unedited codon in the other sequence. Also, codons in which non-replacement C→U edits occurred always specified the same amino acid in the maize and soybean *mat-r* protein. Consequently, there is a perfect match of amino acids that either result from (replacement) or are maintained by (non-replacement) edits. Therefore, editing increases amino acid sequence similarity between the maize and soybean *MAT-R* proteins.

As has been observed for many other angiosperm mt-protein genes, the different edits were not found in all *mat-r* transcripts and, with two possible exceptions (both involving two adjacent Cs), the pattern of edits between clones did not suggest editing polarity.

Three regions (I-III) of the maize *MAT-R* protein (ranging between 108 and 142 amino acids in length) have between 37% and 45%, and 23% and 35% sequence identity to corresponding consecutive regions of the maturases encoded in the first intron of the *Marchantia*

*polymorpha coxI* gene (M.p. COXI-I1; Oda et al., 1992) and the second intron of the *S. cerevisiae coxI* gene (S.c. COXI-I2; Bonitz et al., 1980). Within regions I and II of the maize (and soybean) *MAT-R* sequence occur the seven amino acid domains common to all reverse transcriptases and reverse transcriptase-like sequences (Xiong and Eickbush, 1990; Mohr et al., 1993). A conserved sequence of about 100 amino acids designated domain X has been shown to occur in all group II intron ORFs (Mohr et al., 1993). Domain X appears to be directly associated with intron splicing: two mutations in this region of the *S. cerevisiae coxI* I2 gene, that each cause an amino acid replacement, eliminate splicing (Moran et al., 1994). Region III of the maize *MAT-R* sequence contains all of domain X which includes 17 of the 19 most conserved amino acids in domain X in other group II intron ORFs (Mohr et al., 1993).

With the exception of two amino acids specified by codons that have a silent edit, all amino acids specified by edited codons in maize and soybean *mat-r* transcripts occur in domains 1-6 and X. Most striking is the observation that of the twelve amino acids that result from a replacement edit, eight occur in domain X. Of the total of 17 amino acids that result from (replacement) or are maintained by (silent) editing in the maize and/or soybean *mat-r* transcript sequences, twelve are conserved in either the *S.c* COXI-I2 or the *M. p* COXI-I1 maturase sequences. These data establish that the majority of edits in *mat-r* gene transcripts serve to increase amino acid sequence similarity to distantly related, as well as more closely related organisms. The increase in amino acid sequence similarity to a yeast maturase contrasts with the findings of Wissinger et al., 1991. These authors examined part of the *Oenothera mat-r* gene transcript for edits, and concluded that the ten C→U changes they recorded did not result in increased amino acid sequence similarity to fungal maturases.

Our data support the argument that the *MAT-R* protein is functional and suggest that this protein has both reverse transcriptase and RNA splicing activity. However, in the angiosperm *MAT-R* sequences, 19 or 23 of the 26 amino-terminal-proximal residues characteristic of reverse transcriptase domain 1 (Xiong and Eickbush, 1990) are missing. Also, angiosperm *MAT-R* sequences lack the upstream Z and Protease regions, and the downstream Zn<sup>2+</sup> finger-like region that occur in *S. cerevisiae* COXI I1 and I2 (Mohr et al., 1993; Zimmerly et al., 1995a, b).

*Translation initiation of mat-r transcripts* (also, Thomson et al., 1994). The first ATG from the 5' end of the *MAT-R*-encoding ORF is either codon 23 (broad bean and *Oenothera*) or codon 27 (maize, wheat, soybean). We were hesitant to interpret these ATGs as translation initiation codons because the amino acid sequences predicted from the region upstream of these codons are almost completely conserved among the five angiosperms, and are highly conserved relative to the amino acid sequences of fungal maturases (Wahleithner et al., 1990). However, examination of sequences of cDNAs of transcripts of regions upstream from, and including the 5' end of the ORF, and sequences of cDNAs primed within the 5' end region of the ORF, rule out the possibilities that in transcripts, either a new AUG is created or the ORF is extended upstream as the result of either RNA editing (nucleotide substitution or deletion/insertion) or of RNA

splicing. As an alternative to the above mentioned AUGs acting as initiator codons we have considered the possibility that, as has been postulated for *Drosophila* mitochondrial COI genes (de Bruijn, 1983; Clary and Wolstenholme, 1983), a four nucleotide sequence (5'ATAA or 5'ATGA) located immediately upstream from the first codon of the *mat-r* ORF could act as a translation initiation codon.

Editing in introns, splicing order, and the relationship of editing to exon splicing (M.C. Thomson, J.L. Macfarlane, C.T. Beagley and D.R. Wolstenholme, in preparation).

We have examined cDNAs of domains 5 and 6 of each of the four maize *nad1* introns for edits in order to evaluate the possibility that excision of each intron and accompanying exon splicing might be edit-dependent. Domains 5 and 6 (and 1) of yeast group II introns have been shown to be essential for splicing (Jarrell et al., 1988a; Boulanger et al., 1995; Peebles et al., 1995) and evidence has been obtained supporting the view that an edit in domain 6 of the *Oenothera* c/d intron is necessary for intron excision (Börner et al., 1995). We generated cDNAs from primers antisense to the exon downstream from each maize *nad1* intron. Upstream, non-spliced intron-exon cDNA sequences were selected by PCR amplifying the cDNAs, using sense primers corresponding to sequences 5' to domain 5 of the intron. The amplification products were cloned and sequenced. Examination of multiple, independently derived cloned cDNAs failed to detect any edited nucleotides in the domain 5 and domain 6 regions of intron b/c (although some expected edits occurred in the C exon regions of the cDNAs -- see below). In contrast, a single C→U edit was found in domain 6, between 3 and 9 nt from the 3' ends of *trans*-spliced introns a/b, c/d and d/e. However, it is not clear how any of these edits might affect splicing. Two of the edits, in the c/d and d/e introns, each change a predicted G-C pair to a G•U pair in a helical region of domain 6. The edit in the c/d intron, 3 nt from the 3' end, involves a C that is not predicted to be paired. The edit in the a/b intron does not correspond to either of two edits in the *nad1* a/b intron of *Oenothera* (Wissinger et al., 1991), and the edit in the d/e intron does not correspond to an edit in the *nad1* d/e intron of *Petunia* (Conklin et al., 1991). No other edits were detected from cDNA analysis of the portion of the maize and soybean d/e introns between the *mat-r* gene and the E exon.

We next examined transcripts of the entire b/c intron for editing. Because our results indicated that this intron is rapidly excised following transcription (see below), we prepared b/c intron-containing cDNAs as follows. First, using PCR, two kinds of cDNAs were generated: the first cDNA contained the last 24 ntp of the a/b intron, all of the B exon (not edited in maize), and the first 214 ntp of the b/c intron. The second cDNA contained all but the first 104 ntp of the b/c intron (and, therefore, overlapped the first cDNA by 110 ntp), the entire C exon and the first 30 ntp of the c/d intron. Examination of multiple copies of each of these cDNAs failed to identify any edits in the b/c intron. However, in eight separately derived cDNAs, the C exon contained variable numbers (average, 44%) of the ten edits detected in this exon in complete, mature *nad1* transcripts.

We carried out experiments to determine whether there are limits to the order in which exons can splice. Sense and antisense primer pairs that recognize maize *nad1* intron sequences flanking each of the nine possible multiple, incomplete, *nad1* exon combinations (A+B, A+B+C, A+B+C+D, B+C, B+C+D, B+C+D+E, C+D, C+D+E, D+E) were used to generate double-stranded cDNAs. These cDNAs were then PCR amplified, sized in gels, cloned, and sufficiently sequenced to confirm their exon content. The results obtained indicated that RNAs containing all possible *nad1* exon combinations exist in maize mitochondria, and therefore support the view that splicing of the maize *nad1* exons can occur in any order. The finding of RNA molecules containing only spliced exons A+B, B+C, C+D and C+D+E indicates that *trans*-splicing of exons A and B; C and D; and C, D and E can all occur in advance of *cis*-splicing of exons B and C. [These arguments are based on the assumption that all observed transcripts are intermediates in the generation of mature transcripts.]

As noted above, of the ten Cs in exon C that are edited to Us in the mature *nad1* transcript, an average of only 44% are edited in primary transcripts of exon C. We have, therefore, examined the frequencies of edits in exon-C in immature transcripts that contain this exon spliced to other exons in the seven intermediate combinations. The frequencies observed were as follows: C only, 44%; B+C, 69%; C+D, 87%; A+B+C, 89%; C+D+E, 89%; B+C+D, 93%; A+B+C+D, 97%; B+C+D+E, 98%, indicating a progression of editing with increased numbers of splicing events. Data obtained for exon E (that also contains ten edits in mature *nad1* transcripts) also indicates that the frequency of C→U changes increases as splicing progresses, but from an initial frequency of 88% in primary transcripts (that is, E only): D+E, 90%; B+C+D+E, 97%.

The following further conclusions can be drawn from the above data. As noted previously for other plant mtDNAs, C→U editing is a post-transcriptional event, and editing can precede splicing (Sutton et al., 1991; Yang and Mulligan, 1991). Full editing of either the C or E exons is not a prerequisite for splicing. However, our data leave open the possibility that specific edits may be necessary for exon C and for exon E to splice to other exons. Further, as noted previously (Gray and Covello, 1993), there is a lack of editing polarity in unspliced C and E exons as is the general case for *mat-r* gene transcripts (above).

*Relative rates of nad1 exon splicing.* Four antisense oligonucleotides were made, complementary to the sense strands of the *nad1* B, C, D and E exons, various distances downstream from the exon's 5' splice site. In separate experiments the different oligonucleotides were annealed to whole mtRNA and used with reverse transcriptase in dideoxy-sequencing reactions. Synthesis primed from the C exon, and from the D exon, each yielded a sequence ladder indicating that in most or all of the transcripts the C and D exons are spliced to the upstream exons (B and C, respectively). In contrast, for each of the synthesis reactions primed from the B exon and from the E exon, the resulting sequence ladder indicated that there are two approximately equally abundant kinds of B exon-containing, and E exon-containing

transcripts. In one kind of transcript, the exon is spliced to the upstream exon (B to A, and E to D) and in the second kind, the exon remains continuous with the upstream group II intron. To confirm this interpretation for exon E we carried out two further dideoxy-sequencing reactions: one using a 20 nt primer complementary to the last 10 nt of the d/e intron and the first 10 nt of the E-exon (the intron-exon junction) and a second using a 20 nt primer complementary to the D exon-E exon junction. The sequences obtained confirmed that spliced and unspliced exon E-containing transcripts occur (with approximately equal frequency) in maize mitochondria.

Assuming that all exons and their associated group II intron sequences are continuously transcribed, the most straight forward explanation of these findings is that on average exon B is spliced to exon C and exon C is spliced to exon D sooner after transcription than exon A is spliced to exon B and exon D is spliced to exon E.

#### Attempts to induce the b/c intron to self-splice *in vitro*

We have attempted to determine conditions that permit the (normally cis-spliced) b/c intron to self-splice *in vitro*. As natural transcripts of this intron appear not to contain any edits, we prepared *in vitro* transcripts from the gene sequence. A construct, containing about one half of the B exon, all of the b/c intron, all of the C exon (unedited) and about 23 nt of the c/d intron was introduced into the T7 RNA polymerase transcription vector pTZ19 (USB), and run-off transcripts were generated. All attempts to induce self-splicing *in vitro* were unsuccessful. These included using the conditions that others have shown to be successful to induce self-splicing of the *S. cerevisiae* aI5c intron (van der Veen et al., 1986; Peebles et al., 1986) and the *S. cerevisiae* bI1 intron (Schmelzer and Schweyen, 1986), and also the improved reaction conditions used to optimize *in vitro* self-splicing of the *S. cerevisiae* aI5g intron (Jarrell et al., 1988a, b).

Two overlapping cDNA sequences, one containing a portion of the a/b intron, the B exon and a 5' portion of the b/c intron, and a second containing the 3' portion of the b/c intron, a completely edited C exon and 25 nt of the c/d intron were joined using PCR to provide the expected continuous b/c intron sequence. Run-off transcripts of this construct were generated, and again, under various conditions that permit *in vitro* self-splicing of some *S. cerevisiae* group II introns, evidence for splicing of the maize *nad1* b/c intron was not found.

#### *nad1* gene intron interrelationships (Wolstenholme et al. 1993)

To gain information on the interrelationships of group II introns in plant *nad1* genes, we made sequence comparisons involving the domain 5 and domain 6 sequence of the four mt-*nad1* introns, other mt-*nad1* introns and two chloroplast (from rice and *Marchantia polymorpha*) *nad1* introns. Data obtained from these comparisons suggest that the four angiosperm mt-*nad1* introns differed in sequence at the time(s) of original insertion, that the time(s) of insertions predated the divergence of the monocotyledon and dicotyledon ancestral lines, and that at least the domain 5 and domain 6 sequences of the introns have been highly conserved throughout

angiosperm evolution. Our data further indicate that the chloroplast *nad1* introns have been highly conserved during plant evolution, but do not suggest a close relationship between the chloroplast *nad1* introns and any of the four mt-*nad1* introns. Comparisons involving *nad2* and *nad5* introns of maize and *Oenothera* supports the general view that homologously located mt-intron sequences have been highly conserved during angiosperm evolution.

#### A pseudo-*mat-r* gene sequence in maize mtDNA (Wolstenholme et al. 1993)

A radio-labelled degenerate oligonucleotide, based on the amino acid sequence YA/VRYADD (a motif found in all angiosperm *mat-r*-encoded proteins so far reported, and highly conserved in fungal mt-group II introns), was annealed to electrophoretically separated maize mtDNA that had been cleaved (separately) with four different restriction enzymes. In autoradiographs of each hybridization, two bands of different sizes were observed for each digest. One band was of a size expected for fragments containing the sequence of the *nad1 mat-r* gene. A clone containing the fragments in the second band was selected (using the degenerate primer) from a maize mtDNA *EcoRI* library, and a 2,994 nt segment recognized by the probe was sequenced. Near the 5' end of this sequence is a 30 codon ORF that predicts an amino acid sequence including the motif YARYADD (but does not begin with AUG). Amino acid sequence comparisons indicate that this sequence is more closely related to the maize *mat-r* gene product (*MAT-R*) than any mt-group II maturases of *Marchantia polymorpha* and *Saccharomyces cerevisiae*, or the protein predicted from the unique reverse transcriptase gene in *Oenothera* mtDNA (Schuster et al. 1989). We have therefore designated this short maize sequence  $\Psi\textit{mat-r-2}$ . Sequence analysis indicated that  $\Psi\textit{mat-r-2}$  is closely linked to the mt-tRNA<sup>f-Met</sup> gene and exon 2 of the *nad2* gene.

#### Summary Statement

The results of the RNA editing studies obtained during the grant period has provided substantial evidence that the angiosperm mitochondrial *mat-r* gene encodes a functional protein, and have increased our overall knowledge regarding nucleotide edits that occur in angiosperm mitochondrial introns. Other studies have advanced our understanding of the splicing of multiple exons of the maize mitochondrial *nad1* gene, and of how *trans*-splicing of split intron sequences occurs.

#### Publications Attributable to DOE Funding

Wolstenholme DR, Macfarlane JL, Beagley CT, Thomson MC, Okada NA, and Fauron CM-R (1993) Maize mitochondrial DNA: the *nad1* gene/*mat-r* gene complex, a maturase-related pseudogene linked to a *nad2* exon, and *nad* gene intron interrelationships. In: Plant Mitochondria. Brennicke A and Kuck U, eds. VCH Publishers, Weinheim, Germany. Pp. 151-161.

- Thomson MC, Macfarlane JL, Beagley CT, and Wolstenholme DR (1994) RNA editing of *mat-r* transcripts in maize and soybean increases similarity of the encoded protein to fungal and bryophyte group II intron maturases: evidence that *mat-r* encodes a functional protein. *Nucleic Acids Res.* 22:5705-5752.
- Wolstenholme DR and Fauron CM-R (1995) Mitochondrial genome organization. In: *Advances in Cellular and Molecular Biology of Plants. Vol. 3: The Molecular Biology of Plant Mitochondria.* Eds. C.S. Levings III and I.K. Vasil. Kluwer Academic Publishers. Dordrecht, The Netherlands. pp 1-59.

#### References cited in the Report

- Bonen L, Chapdelaine Y and Garber P (1990) Effects of DNA rearrangements of wheat mitochondrial gene structure and expression. In: *Proceedings of the Fourth International Workshop on Plant Mitochondria*, Cornell University, Ithaca, New York.
- Bonitz SG, Coruzzi G, Thalenfeld BE and Tzagologg A (1980) Assembly of the mitochondrial membrane system. Structure and nucleotide sequence of the gene coding for subunit 1 of yeast cytochrome oxidase. *J. Biol. Chem.* 255: 11927-11941.
- Börner GV, Mörl M, Wissinger B, Brennicke A and Schmelzer C (1995) RNA editing of a group II intron in *Oenothera* as a prerequisite for splicing. *Mol. Gen. Genet.* 246: 739-744.
- Boulanger SC, Belcher SM, Schmidt U, Dib-Hajj SD, Schmidt T, and Perlman PS (1995) Studies of point mutants define three essential paired nucleotides in the domain 5 substructure of a group II intron. *Mol. Cell. Biol.* 15:4479-4488.
- Brennicke A, Klein M, Binder S, Knoop V, Grohmann L, Malek O, Marchfelder A, Marienfeld J, Unseld M (1996) Molecular biology of plant mitochondria. *Naturwissenschaften* 83:339-346.
- Carignani G, Groudinsky O, Frezza D, Schiavon E, Bergantino E and Slonimski PP (1983) An mRNA maturase is encoded by the first intron of the mitochondrial gene for the subunit 1 of cytochrome oxidase in *S. cerevisiae*. *Cell* 35: 733-742.
- Chapdelaine Y and Bonen L (1991) The wheat mitochondrial gene for subunit I of the NADH dehydrogenase complex: a *trans*-splicing model for this gene-in-pieces. *Cell* 65: 465-472.
- Choquet Y, Goldschmidt-Clermont M, Girard-Bascou J, Kuck U, Bennoun P and Rochai JD (1988) Mutant phenotypes support a *trans*-splicing mechanism for the expression of the tripartite *psaA* gene in the *C. reinhardtii* chloroplast. *Cell* 52: 903-913.
- Clary DO and Wolstenholme DR (1983) Genes for cytochrome c oxidase subunit I, URF2 and three tRNAs in *Drosophila* mitochondrial DNA. *Nucl. Acids Res.* 11: 6859-6872.
- Conklin PL, Hampton J, Conley C and Hanson M (1990) Evidence for *trans*-splicing of the RNA encoding NADH dehydrogenase subunit 1 (NAD1) in *Petunia* mitochondria. In: *Proceedings of the Fourth International Workshop on Plant Mitochondria*, Cornell University, Ithaca, New York.

- Conklin PL, Wilson RK and Hanson MR (1991) Multiple *trans*-splicing events are required to produce a mature *nad1* transcript in a plant mitochondrion. *Genes Dev.* 5: 1407-1415.
- Covello PS and Gray MW (1989) RNA editing in plant mitochondria. *Nature* 341: 662-666.
- de Bruijn MHL (1983) *Drosophila melanogaster* mitochondrial DNA: a novel organization and genetic code. *Nature* 304: 234-241.
- Dietrich A, Weil JH and Maréchal-Drouard L (1992) Nuclear-encoded transfer RNAs in plant mitochondria. *Ann. Rev. Cell Biol.* 8: 115-131.
- Fauron CM-R, Casper M, Gao Y and Moore B (1995) The maize mitochondrial genome: dynamic, yet functional. *Trends in Genetics* 11:228-235.
- Goldschmidt-Clermont M, Choquet Y, Girard-Bascon J, Michel f, Schirmer-Rahire M and Rochaix R-D (1991) A small chloroplast RNA may be required for *trans*-splicing in *Chlamydomonas reihardtii*. *Cell* 65:135-143.
- Gray MW and Covello PS (1993) RNA editing in plant mitochondria and chloroplasts. *FASEB J.* 7: 64-71
- Gray MW (1992) The endosymbiont hypothesis revisited. *Int. Rev. Cytol.* 141:233-357.
- Gualberto JM, Lamattina L, Bonnard G, Weil J-H and Grienenberger J-M (1989) RNA editing in wheat mitochondria results in the conversion of protein sequences. *Nature* 341: 660-662.
- Hiesel R, Wissinger B, Schuster W and Brennicke A (1989) RNA editing in plant mitochondria: plant mitochondria use the standard genetic code in translation. *Science* 246: 1632-1634.
- Hildebrand M, Hallick RB, Passavant CW and Bourque DP (1988) *Trans*-splicing in chloroplasts: The *rps12* loci of *Nicotiana tabacum*. *Proc. Natl. Acad. Sci. USA* 85: 372-376.
- Jarrell KA, Dietrich RC and Perlman PS (1988a) Group II intron domain 5 facilitates a *trans*-splicing reaction. *Mol. and Cell Biol.* 8: 2361-2366.
- Jarrell KA, Peebles CL, Dietrich RC, Romiti SL and Perlman PS (1988b) Group II intron self-splicing: Alternative reaction conditions yield novel products. *J. Biol. Chem.* 263: 3432-3429.
- Kohchi T, Umesono K, Ogura Y, Komine Y, Nakahigashi K, Komano T, Yamada Y, Ozeki H and Ohyama K (1988) *Nucl. Acids Res.* 16: 10025-10036.
- Koller B, Fromm H, Galun E and Edelman M (1987) Evidence for *in vivo trans*-splicing of pre-mRNAs in tobacco chloroplasts. *Cell* 48: 111-119.
- Lonsdale DM, Hodge TP, Fauron CM and Flavell RB (1983a) A predicted structure for the mitochondrial genome from the fertile cytoplasm of maize. In: RB Goldberg (Ed) *Plant Molecular Biology. ICN-UCLA Symposia on Molecular and Cellular Biology, New Series, Vol. 12.* (pp. 445-456)
- Lonsdale DM, Hodge TP and Fauron CM (1984) The physical map and organization of the mitochondrial genome from the fertile cytoplasm of maize. *Nucl. Acids Res.* 12: 9249-9261.

- Maréchal-Drouard L, Dietrich A and Grienenberger J.-M. (1995b) Mitochondrial transfer RNAs and RNA editing. In: Levings CS III and Vasil IK. Kluwer (eds) *Advances in Cellular and Molecular Biology of Plants. Vol. 3: The Molecular Biology of Plant Mitochondria.* pp 93-130 Academic Publishers. Dordrecht, The Netherlands.
- Maréchal-Drouard L, Ramamonjisoa D, Cosset A, Weil JH and Dietrich A (1993a) Editing corrects mispairing in the acceptor stem of bean and potato mitochondrial phenylalanine transfer RNAs. *Nucl. Acids Res.* 21: 4909-4914.
- Maréchal-Drouard L, Weil JH and Dietrich A (1993) Transfer RNAs and transfer RNA genes in plants. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 44: 13-32.
- Michel F, Umesono K and Ozeki H (1989) Comparative and functional anatomy of group II catalytic introns -- A review. *Gene* 82: 5-30.
- Mohr G, Perlman PS and Lambowitz AM (1993) Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucl. Acids Res.* 21: 4991-4997.
- Moran JV, Mecklenburg KL, Sass P, Belcker SM, Mahnke D, Lewin A and Perlman P (1994) Splicing defective mutants of the *cox1* gene of yeast mitochondrial DNA: Initial definition of the maturase domain of the group II intron A12. *Nucl. Acids Res.* 22: 2057-2064.
- Oda K, Yamato K, Ohata E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T and Ohyama K (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA: a primitive form of plant mitochondrial genome. *J. Mol. Biol.* 223: 1-7.
- Palmer JD and Shields CR (1984) Tripartite structure of the *Brassica campestris* mitochondrial genome. *Nature (London)* 307: 437-440.
- Peebles CL, Perlman PS, Mecklenburg KL, Petrillo ML, Tabor JH, Jarrell KA and Cheng H-L (1986) A self-splicing RNA excises an intron lariat. *Cell* 44: 213-223.
- Peebles CL, Zhang M, Perlman PS, Franzen JS (1995) Catalytically critical nucleotides in domain 5 of a group II intron. *Proc. Nat. Acad. Sci. U.S.A.* 92:4422-4426.
- Schmelzer C and Schweyen RJ (1986) Self-splicing of group II introns *in vitro*: Mapping of the branch point and mutational inhibition of lariat formation. *Cell* 46: 557-565.
- Sutton CA, Conklin PL, Pruitt KD and Hanson MR (1991) Editing of pre-mRNAs can occur before *cis*- and *trans*-splicing in *Petunia* mitochondria. *Mol. Cell Biol.* 11: 4274-4277.
- Thomson MC, Macfarlane JL, Beagley CT and Wolstenholme DR. (1994) RNA editing of *mat-r* transcripts in maize and soybean increases similarity of the encoded protein to fungal and bryophyte group II intron maturases: evidence that *mat-r* encodes a functional protein. *Nucl. Acids Res.* 22:5745-5752.
- van der Veen R, Arnberg AC, van der Horst G, Bonen L, Tabak HF and Grivell LA (1986) Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing *in vitro*. *Cell* 44: 225-234.

- Wahleithner JA, Macfarlane JL and Wolstenholme DR (1990) A sequence encoding a maturase-related protein in a group II intron of a plant mitochondrial *nad1* gene. Proc. Natl. Acad. Sci. U.S.A. 87: 548-552.
- Wissinger B, Schuster W and Brennicke A (1990) *Trans*-splicing of *nad1* transcripts in plant mitochondria. In: Proceedings of the Fourth International Workshop on Plant Mitochondria, Cornell University, Ithaca, New York.
- Wissinger B, Schuster W and Brennicke A (1991) *Trans*-splicing in *Oenothera* mitochondria: *nad1* mRNAs are edited in exon and *trans*-splicing group II intron sequences. Cell 65: 473-482.
- Wolstenholme DR, Macfarlane JL, Beagley CT, Thomson MC, Okada NA and Fauron CM-R. (1993) Maize mitochondrial DNA: the *nad1* gene - *mat-r* gene complex, a maturase-related pseudogene linked to a *nad2* exon, and *nad* gene intron interrelationships. In: Brennicke A, Kück U (eds.) Plant Mitochondria. (pp. 151-161) VCH Publishers, Weinheim, Germany.
- Wolstenholme DR, Macfarlane JL, Fauron CM-R, Okada NA and Cahill MA (1990) The *mat-r* gene/*nad1* gene complex. In: Proceedings of the Fourth International Workshop on Plant Mitochondria, Cornell University, Ithaca, New York.
- Wolstenholme DR and Fauron CM-R (1995) In: Levings CS III and Vasil IK (eds.) Advances in Cellular and Molecular Biology of Plants. Vol. 2: Molecular Biology of the Mitochondria. Kluwer Academic Publishers, Dordrecht, The Netherlands. pp 1-59.
- Xiong Y and Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 9: 3353-3362.
- Yang AJ and Mulligan RM (1991) RNA editing intermediates of *cox2* transcripts in maize mitochondria. Mol. Cell Biol. 11: 4278-4281.
- Zimmerly S, Guo H, Eskes R, Yang J, Perlman PS, Lambowitz AM (1995b) A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. Cell 83:529-538.