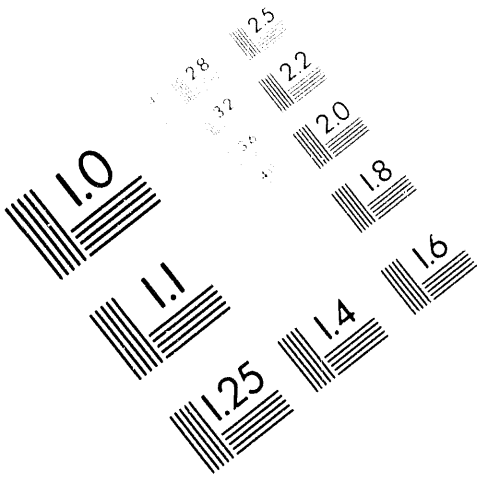
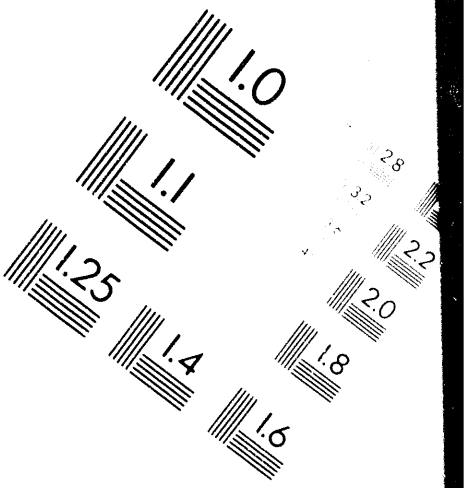
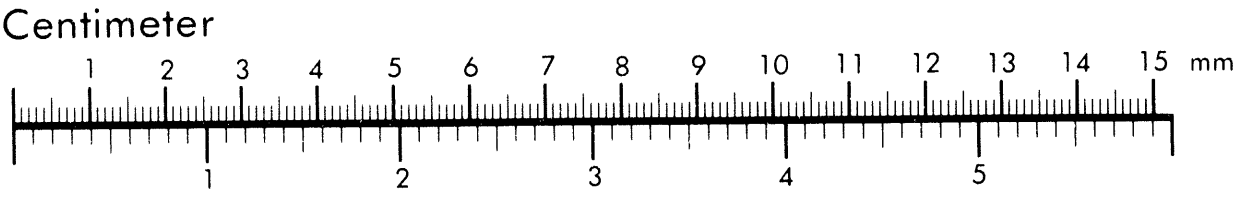
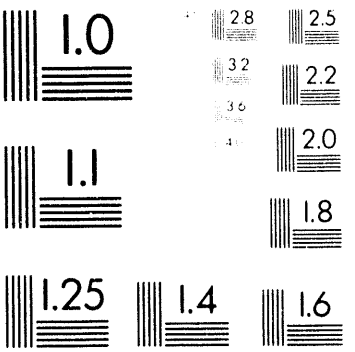
AIIM

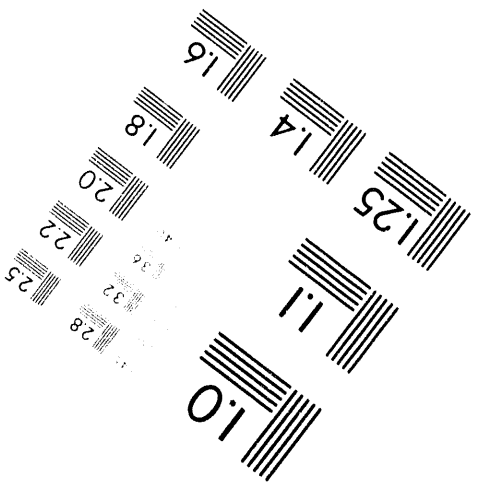
Association for Information and Image Management
1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910
301 587 8202

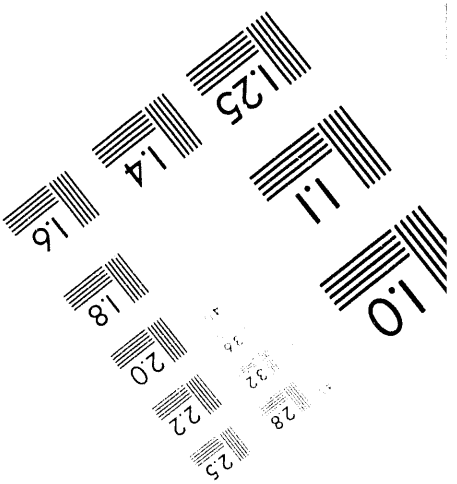












MANUFACTURED TO AIIM STANDARDS
BY APPLIED IMAGE, INC.

1 of 1

DOE/ER/20086--1

David R. Wolstenholme
Project title: The plant mitochondrial *mat-r* gene/*nad1* gene complex
DOE Grant No.: DE-FG02-92ER20086

Progress Report

Within the first two months of our initial DOE grant period, we completed sequencing the segments (totalling 19 kb, both complementary strands) of the maize mtDNA molecule that encode the entire NADH dehydrogenase subunit 1 (*nad1*) gene. This gene comprises five exons (A-E). Exons B and C are joined by a continuous group II intron, but exons A, D and E are at sites widely separated from each other and from exons B and C. Partial group II introns are associated with all five exons. There are two copies of exon A (located in the 14A repeat) and one copy each of exons B, C, D and E. 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) maize, 570 kb mtDNA molecule (see Fig. 1 of Wolstenholme et al. 1993 - reprint attached). 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 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 divided introns 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, first described for split group II introns in *Chlamydomonas* chloroplast genes, is called *trans*-splicing (compared to *cis*-splicing, the process by which continuous introns, such as that separating the B and C exons, occurs). Similar structure and mode of production of mature transcripts have also been suggested for the *nad1* genes of wheat, *Oenothera* and *Petunia*. Within the partial group II intron upstream from the maize *nad1* exon E, is encoded the *mat-r* (maturase related) gene (first described by us in broad bean mtDNA). In broad bean, soybean and *Oenothera*, the *mat-r* gene-containing intron is continuous between exons D and E. Following is a summary of the results we have obtained since the beginning of DOE support (July, 1992).

We have identified nucleotides in mature transcripts of the *nad1* gene that are edited. Using primers complementary to sequences downstream from exon E and upstream from exon A, we generated clones of cDNAs of entire mature (fully spliced) *nad1* transcripts. Analysis of these sequences indicated that mature *nad1* transcripts contained 23 uridines (U) that each correspond to a cystine (C) in the *nad1* gene sequences. These C→U changes result in 20 amino acid replacements in the *NAD1* protein predicted from the genomic (DNA) sequence. These 20 replacement amino acids match perfectly (20/20) the amino acids found in corresponding positions in the *NAD1* protein predicted from edited *Oenothera* RNAs.

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: GCAGGAGG. Examination of the pairs of partial group II intron 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: TGCTGTTTCC and GCTGGCTGG, respectively. Data from experiments currently being carried out to determine the 5′ and 3′ end locations of *nad1* exon-containing transcripts should help further elucidate whether these repeated sequences define sites of sequence

rearrangement.

We have examined the relative rates of splicing in transcripts of the four *nad1* gene group II introns. Four antisense oligonucleotides were made, complementary to regions in the E, D, C and B exons, various distances downstream from the exon's 5′ splice site. In separate experiments the oligonucleotides were annealed to whole mtRNA and used with reverse transcriptase in dideoxy sequencing reactions. Synthesis reactions primed from the D exon, and from the C exon, each yielded a sequence ladder indicating that in most or all of the transcripts the D and C exons are spliced to the upstream exons (C and B, respectively). In contrast, for each of the synthesis reactions primed from the E exon and from the B exon, the resulting sequence ladder indicated that there are two approximately equally abundant kinds of E exon-containing, and B exon-containing transcripts. In one kind of transcript, the exon is spliced to the upstream exon (E to D, B to A) and in the second kind, the exon remains continuous with the upstream group II intron. Assuming that all exons and their associated group II intron sequences are transcribed at about the same rate, the most straightforward explanation of these findings is that 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. Regarding the order of splicing of the five exons, at this time we have evidence for the following: exons B and C can splice before B splices to A or C splices to D; exons D and E can splice before D splices to C; and exons C, D and E can be spliced together before B splices to C. Therefore, it remains possible that there are no restrictions on splicing order of exons in the maize *nad1* gene.

We have begun examining *nad1* intron cDNAs to determine the extent and distribution of RNA edits in introns, in order to evaluate the possibility that intron excision and exon splicing might be editing dependent. As it has been shown by others that in yeast the domain 5 and domain 6 sequences of group II introns are essential for splicing, we first examined these regions of each of the four *nad1* introns. This was done by generating cDNAs from primers within the exon downstream from each intron, and then selecting non-spliced sequences by amplifying the cDNAs using primers within the intron upstream from the 5′ end of domain 5. Examination of multiple, independently derived cDNAs failed to detect any edited nucleotides in the domain 5 and domain 6 regions of intron b/c (although some edits occurred in the C exon regions of the cDNAs -- see below). However, a single C→U change was found in domain 6, between 3 and 9 nucleotides from the 3′ ends of introns a/b, c/d and d/e. Two of these edits, in the c/d and d/e introns, 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 nucleotides from the 3′ end, involves a C that does not appear to be paired. Interestingly, in the latter case the corresponding nucleotide in the soybean sequence (not examined for editing) would be U. The edit in the a/b intron does not correspond to either of two edits in the *nad1* a/b intron of *Oenothera*, and the edit in the d/e intron does not correspond to an edit in the *nad1* d/e intron of *Petunia* examined by others.

Comparisons of nucleotide and predicted amino acid sequences among *mat-r* genes of broad bean, soybean and maize (sequenced previously by us) and wheat and *Oenothera* (sequenced by others), indicate that the encoded protein has been highly conserved during evolution. However, because the beginning of the 5′-proximal region of the open reading frame is highly conserved, and the first AUG (methionine) codon present in the open reading frame of all species is 22 or 26 codons from the beginning of the open reading frame, it is not clear at which position in the sequence translation is initiated. Also, a protein product of the *mat-r* gene has not been demonstrated. Because of this, to further elucidate whether it is likely that a functional protein is encoded by the *mat-r* gene, we have made a detailed study of editing in maize and soybean *mat-r* gene transcripts. The rationale for doing this is based on the findings of others (and our own for *nad1* exon regions - see above) that C→U editing is a characteristic feature of transcripts of plant mt-protein genes. The majority of these edits result in amino acid replacements which increase overall amino acid sequence similarity of homologous genes in different species.

The results we obtained (given in detail in an accompanying preprint (Thomson et al.)) are summarized as follows. From examination of cDNAs generated from multiple transcripts of the *mat-r* genes of maize and soybean, we established that in these transcripts thirteen and fourteen uridines correspond in location to cytosines in the *mat-r* gene sequences. Of these C→U changes, nine involve corresponding nucleotides in the maize and soybean transcripts, and the remainder are at positions corresponding to Us in the *mat-r* gene transcripts of the other species. The majority of the C→U changes result in amino acid replacement. Between maize and soybean there is a perfect match of amino acids that are encoded by codons containing edited nucleotides in one or both species, and a high degree of similarity of these amino acids to the corresponding amino acids predicted from edited *mat-r* transcripts of *Oenothera* (determined by others). Our data indicate that during evolution of angiosperms there has been strong selection to conserve the *mat-r* gene-encoded protein and, therefore, that this *MAT-R* protein is likely to be functional. Data from these experiments also indicate that RNA editing neither creates a translation initiation codon near the 5′ end of the open reading frame, nor does it extend the open reading frame. Because of this, we have considered the possibility that translation of the *mat-r* gene may initiate with a four nucleotide codon (ATAA or ATGA), such as has been suggested by us (and by others) for translation of transcripts of the mitochondrial COI gene of *Drosophila.*

In maize and soybean cDNAs that include the portion of the d/e intron between the *mat-r* gene and the E exon, edits other than the single conserved C→U change in domain 6 were not detected (see Thomson et al., enclosed preprint). We have also examined transcripts of the entire b/c intron for editing. Because our results indicated that the intron is rapidly excised following transcription, 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 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 of the nine edits detected in this exon in complete, mature *nad1* transcripts. One cDNA contained all nine edits and the remaining seven cDNAs contained between zero and six edits. These data indicate that editing can be complete before splicing, but leaves open the question as to whether exon editing must be complete before splicing can occur.

We have begun experiments to determine whether the continuous *nad1* group II introns can self *cis*-splice *in vitro*, as has been demonstrated for some fungal group II introns. Because the maize b/c intron is not edited, we have attempted to achieve *in vitro* splicing with this intron, using run-off transcripts generated in a T7 polymerase transcription system from constructs derived from the gene (DNA) sequence. No splicing of an intron-containing transcript was achieved, using the various methods that have proven successful for fungal introns.

In light of our finding that unspliced C exons are variably edited, we considered it worthwhile to make a construct containing the B exon (not edited in mature transcripts), the b/c intron and a fully edited C exon. This was done by combining, using PCR techniques, the cDNA that includes a C exon with all nine edits, and a cDNA containing the B exon and the portion of the b/c intron not included in the C exon-containing cDNA. Run-off transcripts of this construct will now be made in an *in vitro* T7 polymerase transcription system, and tested for their ability to self-splice *in vitro.*

We have begun a series of experiments designed to test whether the protein encoded in the *mat-r* gene can stimulate splicing activity in intron-containing transcripts, and whether this protein (*MAT-R*) has reverse transcriptase activity. We plan to obtain sufficient amounts of the *MAT-R* protein to carry out the various assays, by synthesizing it in an *in vitro* translation system.

The data obtained from our editing studies strongly suggest that the functional *MAT-R*

protein can only be generated from a fully edited transcript. From a cDNA pool of maize mtRNA, we selected and amplified eight cDNAs containing the entire *mat-r* gene. Each DNA was cloned in both orientations in M13mp18 and M13mp19. The occurrence of twelve of the thirteen edited nucleotides in each of the eight cDNAs was determined by directly sequencing the two ends of the molecule. The presence or absence of the remaining, centrally-located edit was determined by sequencing from an oligonucleotide complementary to a downstream region of the *mat-r* gene. One cDNA (designated D8) was found to contain all but two of the expected edits. These two edits were therefore introduced into the D8 cDNA by recombination with a restriction fragment of a cDNA that contained them, but not all other edits. To the fully edited cDNA was added, using PCR techniques, a sequence containing a ribosomal binding site and a downstream AUG in-frame with the *mat-r* gene. Next, this construct will be inserted into a T7 promoter-containing PUC plasmid, and run-off transcripts will be generated in an *in vitro* transcription system. The resulting transcripts will then be used in both an *in vitro* rabbit reticulocyte lysate translation system, and in an *in vitro* wheat germ translation system to produce the MAT-R protein.

Research Plans for the Coming Year

In general, the research goals and plans remain intact relative to the original application. A major technical change is that we now propose to use *in vitro* rabbit reticulocyte lysate and *in vitro* wheat germ lysate translation systems to synthesize the protein (MAT-R) encoded in the *mat-r* gene, rather than the *E. coli* expression system mentioned in the original application. During the coming year, we will continue to carry out PCR experiments of *nad1* exon and intron-containing cDNAs to further test the hypothesis that there is no set order of exon splicing. The exact nucleotide locations of 5′ and 3′ ends of *nad1* exon-containing transcripts will be determined using PCR-RACE (rapid amplification of cDNA ends) to further elucidate which sequences are involved in associations of separately transcribed RNAs during *trans*-splicing. We will continue our attempt to achieve cis-splicing *in vitro* of the b/c intron using transcripts that include fully edited (and, if successful, partially edited) C exons. Also, the possibility that the continuous d/e (*mat-r* gene-containing) intron of the soybean *nad1* gene can self-splice *in vitro* will be examined. Attempts at achieving self *trans*-splicing *in vitro* will also be made, using the separately transcribed halves of the *nad1* a/b and d/e introns. A major part of our planned work for the coming year concerns the MAT-R protein. Using the T7 polymerase *in vitro* transcription system, we will generate transcripts of our fully edited maize *mat-r* gene construct, and of a corresponding soybean *mat-r* gene-containing construct we plan to make, and use these transcripts in *in vitro* rabbit and wheat translation systems to synthesize the MAT-R proteins. We will then assay these proteins for reverse transcriptase activity, and for their ability to induce or enhance *in vitro* splicing of the d/e and b/c introns.

DATE

FILMED

7/13/94

END