Post-transcriptional Control of Chloroplast Gene Expression: Accumulation of Stable psaC mRNA is Due to Downstream RNA Cleavages in ndhD Gene

Eva M. del Campo, Bartolomé Sabater and Mercedes Martín

Department of Plant Biology. University of Alcalá. Alcalá de Henares. 28871 - Madrid. Spain

1Corresponding author, e-mail: mercedes.martin@uah.es

Address for correspondence:

Dr. Mercedes Martín
Departamento de Biología Vegetal
Universidad de Alcalá
Alcalá de Henares
28871 - Madrid. Spain
Phone: 34 91 8854934
Fax: 34 91 8855066
e-mail: mercedes.martin@uah.es

Running title: Origin of mRNAs of plastid psaC and ndhD genes
Intergenic cleavages, intron splicing and editing of primary transcripts of the plastid ndhH-D operon produce multiple overlapping RNAs of which the most abundant by far is the monocistronic 400 nt mRNA of psaC (encoding the PsaC protein of photosystem I), in contrast with the low level of transcripts of the six ndh genes. Like other plastid operons containing genes for functionally unrelated proteins, the contrasting accumulation of ndh and psaC transcripts provides a model to investigate the mechanisms of the post-transcriptional control of gene expression, a feature of chloroplast genetic machinery, with a minimum of interference by transcriptional control. In leek (Allium porrum L), the ndhD transcript (which follows the psaC gene and ends the ndhH-D operon) requires C → U editing to restore its start codon and may be used as a marker for the processing of psaC and ndhD transcripts. By determining the editing state and 5′ end sequences of specific transcripts, we demonstrated that stable monocistronic psaC mRNA results from downstream cleavages in the ndhD sequence which renders non-functional ndhD transcripts as by-products. Alternative psaC-ndhD intergenic cleavages produce complete mRNAs for both genes, but only take place in precursors containing editing-restored ndhD start codons. Hence, post-transcriptional control acts by promoting the ndhD cleavage alternative which allows the accumulation of psaC mRNA at the expense of ndhD mRNA levels.
Many chloroplast genes are expressed as polycistronic transcription units that are processed to complex sets of overlapping RNAs through steps controlled by nuclear encoded factors (1, 2). A lot remains to be learned about factors, intermediates and the order in which transcript processing occurs as well as the mechanisms responsible for the accumulation of specific mRNAs (frequent in operons containing genes for functionally unrelated proteins). This last is a feature of chloroplast genetic machinery in which post-transcriptional processing controls the differential expression of specific genes within a same operon.

The \textit{NdhH-D} operon provides a model to investigate post-transcriptional control of gene expression because it includes six \textit{ndh} genes whose protein products are present at low levels in chloroplasts, together with the \textit{psaC} gene, encoding the Psac protein (9 Kd Fe-S subunit VII) of the abundant photosystem I (PS I). The polypeptides encoded by \textit{ndh} genes are part of the thylakoid Ndh complex that has been purified from peas (3) and barley (4). The PS I/Ndh ratio in chloroplasts is estimated at 200. Accordingly, the level of monocistronic \textit{psaC} mRNA is estimated to be around two orders of magnitude higher than that of \textit{ndh} genes (5-8).

The \textit{NdhH-D} operon includes (in this order) \textit{ndhH}, \textit{ndhA}, \textit{ndhI}, \textit{ndhG}, \textit{ndhE}, \textit{psaC} and \textit{ndhD} genes (9) and produces a complex pattern of transcripts resulting from intergenic cleavages, intron splicing (within \textit{ndhA}) and C \rightarrow U editing at several specific sites (usually in \textit{ndhA} and \textit{ndhD} genes). The requirement of intron splicing prior to the editing of site III of \textit{ndhA} indicates an order of the processing steps affecting the 5’ end of the primary transcript (near 8 Knt) due to restrictions imposed by the secondary structures of RNAs (8, 10).

The sequence of events involved in the processing of the 3’ end of the \textit{ndhH-D} operon primary transcript are less known and affects \textit{psaC} and \textit{ndhD} mRNAs (246 and 1503 nt, respectively, in most plants), including editing of several sites in the latter. All dicots investigated, as well as the monocots Liliaceae and Aloaceae (11, 12), require editing to restore the \textit{ndhD} AUG start codon from the cryptic start sequence ACG. Other monocots, such as Poaceae, have the correct ATG start codon and, therefore, do not require editing at this position. The \textit{ndhD} transcripts of several plants require editing at one or two additional sites (8, 12). Pertinent to this work is position +59 which is C \rightarrow U edited in \textit{Allium cepa} (onion), \textit{Allium porrum} (leek), \textit{Allium sativa} (garlic) and \textit{Aloe vera} (aloe) converting the genomic encoded codon UCA for Ser to UUA for Leu (12). These four monocots also require restoration of the start codón from ACG to AUG by C \rightarrow U editing and the appropriate edited transcripts were found for garlic and aloe (12). However, although RT-PCR and sequencing of RNA in onion and leek demonstrated editing of position +59, it failed to detect AUG start codons restored by
editing. This raised the questions of whether the ndhD message is finally translated to NDH-D polypeptide in leek and onion and if the editing status of the ndhD start codon is related to processing and accumulation of functional monocistronic mRNA of psaC. Thus, leek and onion may provide a good model to investigate the differential accumulation of psaC and ndhD mRNAs.

Leek plastid DNA has a psaC - ndhD intergenic sequence of 120 nt; the 351 nt sequence spanning from the last 45 nt of psaC to the first 186 nt of ndhD has been described (12) (accession no. AJ278352). This work is concerned with the identification of processing steps and RNA structural features responsible for the accumulation of monocistronic psaC mRNA as a model of post-transcriptional control mechanisms in chloroplasts. Firstly, we demonstrate the presence of NDH-D polypeptide in leek and identify transcripts containing psaC and/or ndhD messages. By sequencing the RACE amplified 5´ends of different ndhD-containing transcripts and using the editing-recovered ndhD start codon as a marker, we demonstrated that stable monocistronic psaC mRNA results from 5´ cleavages at the downstream ndhD sequence, with non-functional ndhD messages as a by-product. Alternative psaC- ndhD intergenic cleavages produce complete gene messages but not stable psaC mRNAs.

EXPERIMENTAL PROCEDURES

Plant Material - Leek (Allium porrum L) was obtained from a local grocery store. Green non-damaged leaves were used as source of chloroplast DNA and RNA.

List of Oligonucleotides -

psaC2: GCACTCAATGTGTACGAGCT
psacC5: TAAGATAGACCCATGCTGC
NDC1: GTTTTGCCGATATTTGCCGG
NDC2: CTGGTAACCACGTATGTAGG
NDC8: AGATGTGAATCCGCCTGTCC
NDC7: TCCATCGATTCCCAATCTCC
NDC9: GGATCGTCCAATTGGAAATG
NDC4: TCCTGGTAAGGCAAGAGAAG
NDC11: TGAGGAACCCGCGAATATTG
NDC12: TGAAACAACTCGCAGCATGG
NA1: GCTCAATCTATTAGTTATGAAA
INTA3: GGATGAGATCAATTCGGAAGC
Oligo d(T) anchor:   GACCACGCTATCGATGTCGAC(T)\textsubscript{16}V
PCR anchor:   GACCACGCTATCGATGTCGAC

**Isolation of Chloroplast DNA and RNA** - Total DNA from leek leaves was isolated by slight modifications (8) of the method described by Ausubel et al. (13). For RNA isolation, leaf segments (1 g) were ground in liquid nitrogen and RNA preparation was performed by phenol/SDS extraction and selective precipitation with LiCl (8,14). RNA yields were approximately 0.25 mg/1 g leaves.

**RNA Electrophoresis, Band Isolation and Northern-blot Hybridisation** - After treatment of RNA samples with formamide, transcripts were separated by electrophoresis in non-denaturing pre-cast (Reliant-Gel System) FMC gels and isolated and northern blotted as described previously (8). Hybridisations on nylon membranes (Zeta-Probe, Bio-Rad) were carried out using the non-radioactive Boehringer Mannheim (Germany) DNA labelling and detection kit and PCR-generated probes.

**Reverse Transcription, Amplification and Sequencing of DNA and cDNA** - RNA bands isolated from agarose gels were reverse transcribed and amplified using the Access RT-PCT System (Promega) as previously described (8), with the following cycling conditions: initial reverse transcription at 48°C for 45 min; one cycle at 94°C, 2 min; then 94°C, 30 s; 50°C, 90 s and 68°C, 3 min for 50 cycles followed by a final extension at 68°C for 7 min. Reverse transcription and amplification of transcripts longer than 5 Knt was performed using the Superscript One-Step RT-PCR System with Platinum Taq DNA Polymerase (Gibco BRL) combined with the Elongase Amplification System (Gibco BRL), with the following cycling conditions: initial reverse transcription at 48°C for 30 min; one cycle at 94°C for 2 min; then 94°C for 30 s, 50°C for 1 min and 68°C for 5 min for 40 cycles followed by a final extension at 68°C for 7 min. Primer pairs were based on the sequence of the barley ndhH-D operon (8) and (when available) on the leek psaC-ndhD sequence (12) and are shown in Experimental Procedures. Amplification products were subjected to electrophoresis through 1.5% agarose gels, purified with DNA gel extraction kit (Quiaex II of Quiagen) and used for sequence analysis. Genomic DNA and cDNAs were sequenced in an Applied Biosystems automatic sequencer. Three to five different RT-PCR sequencing assays from each specific RNA preparation were carried out to identify editing sites. The Zuker and Stiegler program (15) (GCG Inc.) was used to determine secondary structures of RNA.

**Amplification of cDNA Ends by 5’ RACE** - Selective amplification of the cDNA 5’ end of the monocistronic ndhD mRNA was performed using the 5’/3’ RACE Kit (Roche). First strand cDNA was synthesised using the NDC4 primer incubating for 60 min at 48°C. This cDNA was purified using the high Pure
PCR Product Purification Kit (Roche). The tailing reaction of cDNA was carried out using a terminal transferase that appended a polyA tail to the 3' end of the first strand cDNA. After denaturation at 94°C for 3 min, the mixture was incubated for 30 min at 37°C followed by 10 min at 70°C to heat-inactivate the terminal transferase. This dA-tailed cDNA was amplified by PCR using an oligo dT-anchor primer and the NDC2 primer using Taq DNA polymerase of Amersham-Pharmacia. Cycling conditions were: one cycle at 93°C for 5 min, 93°C for 2 min, 50°C for 90 s and 72°C for 3 min for 40 cycles, followed by a final extension at 72°C for 7 min. We did a second PCR round with the oligo PCR-anchor primer and the NDC9 nested primer to obtain a visible PCR product which was sequenced with the NDC9 primer.

Zymographic and Immunoblotting Assays - Activities and proteins were assayed in chloroplast extracts. Intact chloroplasts were isolated as described (16). NADH dehydrogenase was solubilized in 20 mM Tris-HCl, pH 7.5, 5 mM EDTA, by adding 10% (w/v) Triton X-100 to a final concentration of 2% (approximately 5 mg detergent/mg of protein) as described (4). Protein concentration was quantified by the Bradford method (17) with a Protein Assay Kit (Bio-Rad) using bovine serum albumin as a standard. Gel electrophoresis, zymograms, and immunoassays were performed according to Casano et al. (4). NADH-D antibody was produced by Sigma Genosys Co. (Cambridge, UK) using a synthetic peptide as antigen, which was established by the barley NDH-D protein sequence analysis (8). The amino acid sequence of the antigen peptide is TWLPDTHGEAHYSTC and corresponds to amino acids 232-246 of the NDH-D protein.
RESULTS

Plastid Ndh Complex of Leek Contains NDH-D Polypeptide – In order to determine the presence of NDH-D polypeptide in leek leaves we analyzed the protein preparation solubilised from thylakoids. Two percent Triton (mild detergent conditions) solubilised some polypeptides and protein complexes from barley thylakoid membranes (4), including the Ndh complex. The solubilised preparation from leek thylakoids showed several NADH dehydrogenase bands in zymograms revealed after non-denaturing PAGE (Fig. 1A). Western-blot (Fig. 1B) with antibody against barley NDH-D polypeptide showed that NDH-D polypeptide was also present in leek thylakoids and co-migrated in native PAGE, as in barley (4), with the low migrating NADH dehydrogenase activity corresponding to the Ndh complex. Clearly, unless ACG can act as a start codon, leek chloroplasts must contain fully functional ndhD mRNA with the appropriate AUG start codon which is translated to NDH-D polypeptide. However, sequencing of the RT-PCR amplification product of whole RNA preparation of leek leaves detected only an unedited (ACG) cryptic start sequence in ndhD mRNA (12).

ndhH-D Operon Transcripts Containing psaC and ndhD mRNAs - In barley, the ndhH-D operon produces numerous transcripts that are probably processing products of a full-length primary transcript (8). Frequently, due to competition among the different RNAs present, RT-PCR fails to amplify cDNAs of specific transcripts in the mixture of the whole RNA preparation. It may be that in total RNA of leek leaves, transcripts with editing-restored initiation codons are poor templates for RT-PCR due to competition with unedited transcripts. Hence, we investigated the status of the different transcripts of leek leaves containing the ndhD message taking advantage of the possibility to RT-PCR amplify specific RNA bands separated by non-denaturing electrophoresis (8).

In higher plants, psaC and ndhD genes are (in this order) the last 3’ end genes of the ndhH-D operon. Thus, we investigated transcripts of the operon containing psaC and/or ndhD messages. Northern blots with psaC and ndhD probes (S₁ and S₂, respectively) revealed a large number of transcripts in leek containing the message of one or both genes (Fig. 2). Despite uncertainties due to the low band resolution after non-denaturing electrophoresis, the transcript pattern was very similar to that described for barley (8). Hence, by comparing the results shown in Fig. 2 with former northern blots obtained with several other probes in barley and taking into account the probable structure of the leek ndhH-D operon, a map of transcripts containing leek psaC and/or ndhD messages is shown in Fig. 3. Monocistronic transcripts of around 0.4 and 1.7 Knt of psaC and ndhD genes, respectively, were detected in addition to many other transcripts. The largest, low abundant, transcript of 7.8 Knt
must contain the message of the seven genes of the operon including the ndhA intron and was detected with both psaC and ndhD probes. In fact, most of the transcripts larger than 1.7 Knt were detected with both probes.

Despite uncertainties related to the presence of polycistronic transcripts of similar size but different message content (see transcripts of 6.7, 5.4, 2.8 and 2.0 Knt in Fig. 3) and to the different signal intensities produced by S₁ and S₂ probes, the relative abundance of some transcripts could be estimated. While the signal of monocistronic 1.7 Knt ndhD transcripts was very low when compared with other transcripts detected with ndhD probe (S₂, Fig. 2), the monocistronic 400 nt psaC transcript was by far the most abundant (S₁, Fig. 2). This agrees with the higher amount of PsaC polypeptide (and in general photosystem I) relative to NDH polypeptides (and Nd complex) in chloroplasts (3, 4) and suggests that the monocistronic transcript was the main translatable psaC mRNA. It is noteworthy that the S₂ probe detects a 400 nt transcript of similar size to the monocistronic psaC (Fig. 2). Taking into account the short intergenic region between psaC and ndhD (120 nt in leek (12)) it is possible that the 3′ end of the abundant monocistronic psaC transcript includes part of the 5′ end of the ndhD gene. Since the S₂ probe spans from +34 to +706 nt of ndhD, it could detect (although more weakly than the S₁ probe) monocistronic psaC transcript. In this case, a truncated non-functional (and probably quickly degraded) ndhD transcript is produced and the large amount of translatable mRNA of psaC would explain the low amount of monocistronic ndhD transcript.

On the other hand, the low amount of the monocistronic 1.7 Knt transcript of ndhD gene raised doubts as to whether it significantly contributes to translated NDH-D polypeptide. To shed light on these questions, we investigated the editing state of sites I and II in different transcripts of leek ndhD and the 5′ end of the presumed ndhD monocistronic transcript.

*Editing Restores Initiation Codon in Some Transcripts of the Leek ndhD Gene* - After electrophoresis of RNA in non-denaturing pre-prepared FMC gels, RNA bands of the four gel sections shown in Fig. 2 were isolated. Respectively and approximately, the sections contained transcripts of 6.0 to 8.0, 2.0 to 4.0, 0.8 to 1.9 and 0.1 to 0.8 Knt. cDNA synthesis, amplification with specific DNA primers and sequencing were performed to determine whether sites I and II were edited in transcripts containing the ndhD message. Figure 4 shows the genomic sequence flanking the two potential editing sites of the ndhD gene as well as the sequence profiles determined after RT-PCR amplification with primers NDC8/NDC7 that span from position -160 to +252 with regard to the first nucleotide of the ndhD gene. The cDNA sequence profiles of Fig. 4 were obtained from RNA solubilised from gel section 1 (Fig. 2) containing transcripts of 6.0 to 8.0 Knt and similar results were obtained with each of the other three sections. It was clear that the nucleotide at position +59 was always C to U edited in
transcripts of all four sections (II of Fig. 4). This indicates an early editing of site II during transcript processing and coincides with the full editing of site II found when RT-PCR was performed with whole unfractionated RNA (12). However, while RT-PCR with whole RNA failed to detect editing of site I (12), assays with fractionated RNAs showed that all slices tested contained, at least, one transcript with both sites I and II edited. In fact, the profile sequence (I of Fig. 4) showed ambiguity, with both C and T nt at the second position which indicated a mixed population of site I non-edited and site I edited transcripts in leek chloroplasts. Obviously, the first ones would account for the mRNAs translated to functional NDH-D polypeptide. ndhD mRNAs edited at site I were present even in large transcripts, hence a population of transcripts must be edited at sites I and II at an early stage during or immediately after transcription. Sequencing of RT-PCR amplifications with appropriate primers (e. g.: INTA3/NDC2 primers for intron-containing and NA1/NDC2 for intron-lacking ndhA transcripts) showed that there are site I edited and site I non-edited in both ndhA intron-containing and intron-lacking transcripts (e.g.: the two 6.7 Knt transcripts of Fig. 3). Editing occurs early at site II in all transcripts, but some of them are edited at a later stage or not at all at site I. The latter possibility is intriguing and one may wonder if transcripts not edited at site I are precursors of the highly abundant psaC monocistronic transcript. We have mentioned above that northern-blots (Fig. 2 and discussion herein) suggested that monocistronic psaC mRNA could be enlarged at its 3’ end to the first nt of ndhD, which would imply that only a fraction of primary transcripts would render full-length functional monocistronic ndhD mRNA. In this regard, Hirose and Sugiura (18) reported that site I editing and intergenic cleavage between the psaC and ndhD genes are required for translation to NDH-D polypeptide in tobacco.

A full description of the site I editing status of each transcript would require precise gel slicing that is technically difficult. More affordable were questions concerning the editing status of monocistronic ndhD mRNA and, consequently, the equilibrium between the maturation of monocistronic mRNAs of psaC and ndhD in leek chloroplasts.

Site I is always Edited in Monocistronic mRNA of ndhD – Both the editing status and the sequence of the 5’ end of monocistronic ndhD mRNA was investigated by sequencing the RACE amplified 5’end (19). RNAs eluted from the third gel section (Fig. 2) were used as templates to synthesize first strand cDNA with the NDC4 primer, which spans from position +710 to +730 of the barley ndhD, and very probably spans the same sequence in leek. A poly (dA) tail was attached to the 5’ end of the cDNA by a terminal transferase, and this was then amplified using a 5’ oligo(dT) anchor primer and a nested 3’ primer, NDC2, which spans from position +687 to +706 of the ndhD sequence. This step eliminated a portion of the 3’ end but included the complete 5’
end. This product was again amplified using a PCR anchor primer and nested primer NDC9 which spans from position +160 to +179 of barley ndhD in order to obtain an adequate amount for sequencing. Agarose gel electrophoresis of all repetitions showed two amplification bands of 250 and 150 nt (Fig. 5). Obviously, the larger band must correspond to ndhD mRNA with a complete 5’ end, and the shorter band to a truncated, non-functional RNA. Sequencing of the amplified bands confirmed this assumption, revealing heterogeneity of monocistronic ndhD mRNA, and also confirmed that the cryptic start sequence was restored by C to U editing in all RNAs of the large band.

The sequences of both bands shown in Fig. 5 were determined four (large band) and two (short band) times after independent RACE amplifications. Sequences deduced for the amplified RNAs agreed with the predicted genomic sequence (except for editing as indicated below) and all of their 3’ ends terminated at the boundaries of the NDC9 primer sequence. However, as shown in Fig. 6, the 5’ ends of the transcripts, which start from the end of the oligo(dT) anchor - poly(dA) tail, varied among the different experiments. In the large band, transcript 5’ ends were found at positions –56, -45 and –31 (two transcripts) upstream of the start codon of the ndhD gene. All had been edited at sites I and II. No trace of unedited site I was detected (Fig. 6A). In fact, no ambiguity was found at any position in the 5’ end, indicating that in each experiment, only one type of transcript is cDNA copied and/or amplified. Using a different approach, Hirose and Sugiura (18) also found multiple 5’ ends of monocistronic functional ndhD transcripts. The short band (150 nt, Fig. 5) transcript 5’ ends were found at position +14 and + 75 downstream of the ndhD gene start codon (Fig. 6B). Therefore, one can easily assume that these transcripts are not translatable to NDH-D polypeptide although the former showed an edited site II (not shown).

Probe S2 (which spans from position +34 to +706 of the ndhD gene) weakly detected a transcript of approximately 400 nt (Fig. 2), which we tentatively assumed to be monocistronic psaC. It would therefore be possible that the transcripts producing the short band of Fig. 5 be non-functional by-products of the processing pathway of polycistronic transcripts to monocistronic psaC mRNA. In fact, RT-PCR amplification of RNA from the 100-800 nt gel section 4 (Fig. 2) with primers NDC8/ND11, which span from position +157 of psaC to +59 of ndhD produced a cDNA fragment of the expected length and sequence (not shown). The nucleotide profile showed C/T ambiguity at position +2 of the included ndhD region, indicating the existence of both edited and unedited site I transcripts. Taking into account both the gel section utilized as a source of RNA and the primers, the amplified and sequenced transcripts must correspond to monocistronic psaC mRNA which includes a portion of the 5’ end of ndhD due to a cleavage alternative during processing. It can not be excluded that additional,
shorter monocistronic psaC transcripts with 3' ends terminating in the psaC-ndhD intergenic region were also present in the 400 nt RNA gel section (Fig. 2). In summary, these results indicate that alternative cleavage of polycistronic transcripts around psaC – ndhD could take place, either in the intergenic region, between positions –30 and –60 upstream of the ndhD message, or in the ndhD coding region between positions +10 and +80 downstream of the start codon. The first alternative produces potentially translatable monocistronic mRNAs of both psaC and ndhD, with the latter completely edited at sites I and II. The second alternative produces functional and stable psaC mRNA only.

3' End Secondary Structure of psaC mRNA - One immediate question was the relationship between the alternative cleavages and the large amount of psaC monocistronic transcript when compared with low level of ndhD monocistronic transcript (Fig. 2). The difference could indicate that ndhD cleavage was far more frequent than intergenic cleavage. This poses the question of why such a large primary transcript is produced if most of the ndhD message will be finally lost. Another factor to be considered is the relative stability of the alternative psaC mRNAs that are produced by each of the two cleavages. It is well established (20, 21) that plastid mRNA stability is considerably improved when large (some 15-20 nt) secondary base-pairing occurs near their 3’ end. Figure 7A shows the estimated 3’ end secondary structure of psaC mRNA including the intergenic psaC – ndhD sequence. The most prominent feature is a large base-pairing branch spanning from 51 to 92 nt downstream of the termination codon of psaC which hybridises with the sequence –70 to –18 nt upstream of the ndhD initiation codon. Clearly, cleavages detected at the intergenic region (between –30 and –60 nt upstream of ndhD), while rending translatable monocistronic ndhD mRNA will not produce stable psaC mRNA because the essential stabilising component, the base-pairing branch, is absent (60 to 90 nt downstream of psaC, Fig. 7). Stable monocistronic psaC mRNA requires a longer 3’ end that is close to, if not within, the ndhD coding region, which would result in a poorly translatable or non-functional ndhD message.

Transcripts with a 5’ end between +10 and +80 nt downstream of the ndhD initiation codon (those giving rise to the short amplification band of Fig. 5) are probably non-translatable by-products of ndhD cleavages that produce mature and stable monocistronic psaC mRNA. On the other hand, monocistronic translatable ndhD mRNA would require cleavages far upstream in the intergenic region (–70 to –18 nt upstream of the ndhD initiation codon equivalent to 51 to 92 nt downstream of the psaC termination codon, see Fig. 7) to include the GAG Shine-Dalgarno motif (Fig. 7). This motif lies precisely within the otherwise stabilizing base-pairing branch of the monocistronic psaC mRNA. The secondary structure stabilizing psaC mRNA seems to be
due to a relatively conserved sequence within the psaC – ndhD intergenic region. As shown in Fig. 7, the large base-pairing branch is also predicted from the intergenic barley sequence. The GAG Shine-Dalgarno motif presumably required for ndhD translation is part of the stabilizing base-pairing branch of psaC mRNA in barley as well. Thus, our results of alternative cleavages in leek to produce monocistronic mRNAs of psaC and ndhD are probably a general feature in higher plants.

**DISCUSSION**

When whole extracted RNA is used for RT-PCR amplification, failure to detect the editing-dependent recovery of the ndhD start codon (12) is probably an artifact due to competition among the different transcripts. Non-denaturing electrophoresis produced medium-quality band separation, but isolated bands were suitable for the synthesis of the corresponding cDNAs, their amplification and sequencing (8) without interference from other bands. In addition, fractionation of RNA is required for unambiguous RT-PCR amplification of specific low-sized transcripts, such as monocistronic transcripts, which is a common problem in the multiple transcript operon of chloroplasts. As expected for a necessary protein (22, 4), NDH-D polypeptide is present in leek (Fig. 1) which requires editing to restore the cryptic ndhD start codon at the RNA level. This is particularly patent in leek, where the ndhD gene lacks an alternative in-frame start codon in the region. Accordingly, partial editing of the ndhD start codon was found in all RNA fractions (Figs. 2-4).

Northern assays with different probes allowed us to extend the 3′ end of the transcript map of the ndhH-D operon previously deduced for barley (8) and to confirm a very similar pattern in the two monocots. The map includes a large number of polycistronic transcripts containing essentially all possible combinations of contiguous gene messages (with and without the ndhA intron) as well as monocistronic transcripts corresponding to each gene. Although multiple transcription starts can not be conclusively excluded, these results indicate that intergenic cleavage and the splicing of the ndhA intron take place unspecifically at any stage of the processing pathway of primary transcripts. Results on the editing status of the different sites demonstrated that editing takes place early during processing except for site III of ndhA in barley which requires previous intron splicing (8) and site I of ndhD in leek which will be discussed bellow.

The presence of psaC in the ndhH-D operon with apparently unrelated functional genes has for long attracted attention (5), especially due to the large difference in the level of polypeptide products relative to the other genes. On the other hand, editing to restore the start codon of the contiguous ndhD gene is not complete in many plants (partial editing, 11, 8, 12) which suggests an interplay in the processing of psaC and ndhD mRNAs.
Translatability may vary among the different transcripts of one operon (23). By using an in vitro translation system for transcripts generated from a dicistronic construction of \(psaC\) and \(ndhD\) of tobacco, Hirose and Sugiura (18) concluded that both editing and cleavage of RNA are required for translation of \(ndhD\) mRNA. Only monocistronic RNA for each \(psaC\) and \(ndhD\) are translatable in their system and RNA protection assays revealed that cleavage occurs at multiple sites in the intergenic region (18). Our results in leek confirm multiple transcripts (differing at their 5’ ends) containing translatable mRNA of \(ndhD\) and identify additional alternative cleavages within the \(ndhD\) gene to produce the translatable and stable \(psaC\) mRNA.

RNA bands of 0.8 to 1.9 Knt (that include monocistronic \(ndhD\)) separated by non-denaturing electrophoresis were successfully amplified by RACE in order to sequence their 5’ends (Fig. 5 and 6). Three types of translatable mRNA for \(ndhD\) were identified that differed at their 5’end, respectively at positions –56, -45 and –31 upstream of the initiation codon. All had edited sites I and II and could potentially be translated to NDH-D polypeptide. Obviously, the two additional types of transcripts with 5’ends at positions +14 and +75 downstream of the initiation codon cannot be translated to NDH-D. Very probably these last two transcripts are by-products of at least one type of mature translatable \(psaC\) mRNA. In fact, northern assays (Fig. 2) and amplification and sequencing with appropriate primers demonstrated that part of, if not all, monocistronic 400 nt transcripts containing the \(psaC\) message had 3’ends extending within the \(ndhD\) sequence. Only partial editing of the \(ndhD\) start codon was found in these transcripts. Significantly, a stabilizing 3’end secondary structure (Fig. 7) may be possible in \(psaC\) mRNAs whose 3’ end include a portion of the \(ndhD\) 5’ end but not in those produced by the intergenic cleavages rendering translatable \(ndhD\).

Post-transcriptional control must necessarily account for the different levels required of each translatable transcript of operons with functionally unrelated coding sequences (such as the \(ndhH-D\) operon). We conclude from our results (Fig. 8) that two alternative cleavages are involved in the processing of precursor RNA to functional \(psaC\) and \(ndhD\) mRNAs in leek. Cleavage in the \(ndhD\) coding region probably produces stable \(psaC\) mRNA but nonfunctional \(ndhD\) mRNA with major frequency. This cleavage is independent of the restoration of the cryptic start codon of \(ndhD\) by editing and can occur in both site I edited and site I non-edited transcripts. Site II is always edited. In the other less frequent processing alternative, cleavage takes place within the intergenic sequence involved in base-pairing stabilization of \(psaC\) mRNA (Fig. 7) and only in transcripts where site I has been edited to allow translation of \(ndhD\) mRNA. The latter alternative cleavage may potentially produce translatable mRNAs of both \(psaC\) and \(ndhD\) genes but it is likely that this last \(psaC\) mRNA is not as stable as that produced by the first alternative cleavage. By contrast, the second alternative cleavage produces the
ndhD mRNA with the full-length Shine-Dalgaro sequence. Therefore, the production of stable psaC RNA implies the loss of functional ndhD mRNA. Different rates of the alternative cleavages and stabilization of the psaC mRNA resulting from ndhD cleavage explain the far higher level of psaC mRNA than that of ndhD mRNA (Fig. 2). Post-transcriptional control is carried out by preferential cleavage within ndhD.

Acknowledgements- We thank Mrs. P. Hauke for kindly lecture and suggestions. This work was supported by Grant BFI2000-0780 of the Spanish DGICT (Ministerio Educación y Cultura).
REFERENCES


Figure Legends

**Fig. 1.** Zymograms of NADH dehydrogenase activities and immunoblot identification of NDH-D protein. Chloroplast extracts (50 µg of protein) were separated by native PAGE electrophoresis and revealed for NADH dehydrogenase activity (A) and, after membrane blotting, immunoassayed with antibody against NDH-D (B).

**Fig. 2.** Northern blots of transcripts of *ndhH-D* operon and gel sections used for transcript elution. After non-denaturing electrophoresis of leek RNA, lanes were northern-blotted with probes S₁ and S₂ (*psaC* and *ndhD* coding regions, respectively). The gel sections used to isolate RNA are indicated at the left of northern lanes. Sizes (Knt) of the main transcripts detected by northern blot and of RNAs from gel sections are also indicated.

**Fig. 3.** Gene and transcript maps of the *ndhH-D* operon of leek. The gene map of the leek operon (top) indicates the location of the edited sites discussed in this study with Roman numerals on vertical arrows. S₁ and S₂ probes are represented under the gene map. The map of transcripts (bottom) was deduced from northern assays and comparison with the barley transcript map. The estimated sizes are indicated (Knt) at the right of each transcript.

**Fig. 4.** Sequences of cDNA flanking editing sites I and II of leek *ndhD*. The cDNA sequences were obtained from the 396 nt product of RT-PCR amplification using RNA isolated from gel section 1 (Fig. 1) as template. NDC8/NDC7 primers were used for amplification and sequencing. Arrows mark positions showing C/T differences between DNA and cDNA (editing sites I and II). Both C (blue) and T (red) peaks are present at the nucleotide position of site I.

**Fig. 5.** Agarose gel electrophoresis of the second PCR round of RACE products. 5′ RACE reaction was carried out using RNA isolated from gel section 3 (Fig. 1) as template, NDC4 primer for first strand of cDNA synthesis, oligo-d(T)-anchor/nested and NDC2 primers for first PCR amplification and PCR-anchor/nested NDC9 for the second amplification (lane 1). Lane 2 is a RT-PCR control without cDNA template. Lane M contains size markers indicated at the left (nt). Approximate sizes (nt) of amplified bands are indicated at the right.

**Fig. 6.** 5′ end of transcripts deduced from sequencing of RACE amplified bands. A) 5′ end sequences determined from four complete assays with the large band (Fig. 5). In box, nucleotide sequence profile of cDNA
flanking the editing site I of the transcripts with the 5′ end at –31 position. A similar sequence was found for other transcripts of the large band. B) 5′ end sequences determined from two complete assays with short band.

Fig. 7. **Predicted secondary structure of the 3′ end of psaC mRNA.** Secondary structures were predicted from the psaC stop codon to the ndhD start codon (black boxes) of the intergenic sequences of leek and barley. The Shine-Delgarno sequence is boxed in thin lines. Arrows indicate cleavage sites that would produce complete ndhD mRNA.

Fig. 8. **Alternative cleavages producing psaC and ndhD mRNAs.** The editing status of sites I and II are indicated. Stable psaC transcripts show a 3′ loop structure not present in unstable psaC transcripts. Truncated mRNAs of ndhD are shown as bars with a ragged left side.
<table>
<thead>
<tr>
<th>Gel section (Knt)</th>
<th>psaC ($S_1$)</th>
<th>ndhD ($S_2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.0 - 8.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>7.8</td>
<td>7.8</td>
</tr>
<tr>
<td>2.0 - 4.0</td>
<td>6.7/6.1</td>
<td>6.7</td>
</tr>
<tr>
<td>3</td>
<td>5.4</td>
<td>5.4</td>
</tr>
<tr>
<td>0.8 - 1.9</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>4</td>
<td>2.8</td>
<td>2.8</td>
</tr>
<tr>
<td>0.1 - 0.8</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>1.3</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>0.4</td>
</tr>
</tbody>
</table>
I

initiation codon of ndhD mRNA

II

20th codon of ndhD mRNA
**A**

**LARGE BAND 5' ENDS**

```
5'-guaaaaccccagccgacgggcaaggccagaaacaaaccgucgcucGAGUUAUULCCCUUGGUU
5'-guaaaaccccagccgacgggcaaggccagaaacaaaccgucgcucGAGUUAUULCCCUUGGUU
5'-guaaaaccccagccgacgggcaaggccagaaacaaaccgucgcucGAGUUAUULCCCUUGGUU
5'-guaaaaccccagccgacgggcaaggccagaaacaaaccgucgcucGAGUUAUULCCCUUGGUU
```

C

I

start codon
of *adhD* mRNA

**B**

**SHORT BAND 5' ENDS**

```
5'-CUUGGUAACAAUUACUGGUAGUUGCCCAAUUUCUGCGGGUUCCUCAUUUUUUUUUCUCCCCCAUAAA
5'-UCAUAAA
```