The EMBO Journal vol.9 no.10 pp.3059-3066, 1990

Three distinct ribonucleoproteins from tobacco chloroplasts: each contains a unique amino terminal acidic domain and two ribonucleoprotein consensus motifs

Yuqing Li and Masahiro Sugiura

Center for Gene Research, Nagoya University, Nagoya 464-01, Japan

Communicated by J.H.Weil

Chloroplasts contain their own genetic system. Eighteen different split genes have been found among ~130 chloroplast genes from higher plants. However, little is known about the chloroplast splicing system. Mammalian heterogeneous nuclear ribonucleoproteins (hnRNP proteins) have been shown to be involved in splicing. We applied a purification procedure developed for HeLa cell hnRNP proteins, which uses a single-stranded DNA (ssDNA) affinity column, directly to the tobacco chloroplast lysate to isolate their chloroplast counterparts. Four proteins (mol. wt ≈ 30 kd) bound strongly to the column. The amino-terminal sequences of three of them were determined and their cDNA clones were isolated from a tobacco leaf cDNA library. Sequence analysis of these clones revealed that all three proteins contain two ribonucleoprotein consensus sequences (RNP-CS), confirming their ribonucleoprotein (RNP) nature. The presence of putative transit peptides in their predicted protein sequences, and an *in vitro* import experiment confirmed they are located in the chloroplast. This is the first report of organellar proteins containing RNP-CS. In addition, these three chloroplast proteins have a very acidic amino-terminal domain, a novel feature among RNP proteins identified so far. They are expressed both in leaves and roots; their mRNA levels showed different light modulation in mature leaves. The three proteins might be involved in splicing and/or processing of chloroplast RNAs.

Key words: acidic amino-terminal domain/chloroplast/ribonucleoprotein/RNP-CS/ssDNA-binding proteins

Introduction

Chloroplasts contain the entire machinery for photosynthesis and their own genetic system. The available sequence data and subsequent expression studies of the chloroplast genome have indicated that this genetic system has both prokaryotic and eukaryotic features (for recent review see Sugiura, 1989). For example, the *psbB* operon is transcribed from a bacterial-type promoter as a tetracistronic precursor while two splicing events and multiple processing steps are necessary to form the functional mRNAs (Westhoff *et al.*, 1986; Rock *et al.*, 1987; Tanaka *et al.*, 1987; Westhoff and Herrmann, 1988). In addition, the tobacco *rps12* and *Chlamydomonas psaA* genes are divided into three parts located far away from each other, and require a *trans*-splicing process to form functional mRNA (Kück *et al.*, 1987; Zaita *et al.*, 1987). The chloroplast genetic system seems to be very complex in terms of splicing and/or processing. Little is known about the molecular mechanism of the chloroplast splicing system.

Chloroplast introns can be classified into at least three groups (Shinozaki *et al.*, 1986a). The chloroplast group I introns can be folded into a secondary structure similar to the postulated intron structure of the autosplicable rRNA precursor of *Tetrahymena*. The chloroplast group II introns can be folded into a secondary structure similar to the postulated intron structure of the maize mitochondrial cytochrome oxidase gene and yeast cytochrome oxidase and cytochrome b genes. The boundary sequences of the chloroplast group III introns resemble those of nuclear gene introns. Therefore, chloroplast pre-mRNAs containing the group III introns are likely to be spliced by a similar mechanism to that operating for nuclear pre-mRNAs (Shinozaki *et al.*, 1986a).

In yeast and animal nuclei, most pre-mRNAs undergo polyadenylation and splicing. A group of proteins is responsible for these processes; this includes heterogeneous nuclear ribonucleoproteins (hnRNPs), small nuclear ribonucleoproteins (snRNPs) and poly (A)-binding proteins. The cloning and sequence analysis of several of these proteins lead to the identification of a highly conserved motif termed the ribonucleoprotein consensus sequences (RNP-CS) (Adam *et al.*, 1986; for a recent review see Bandziulis *et al.*, 1989, and Mattaj, 1989). The RNP-CS is an octapeptide within an RNA-binding domain of 80–100 amino acids. *In vitro* deletion analysis has demonstrated that the RNA-binding domain is a basic functional unit in RNA binding (Query *et al.*, 1989; Scherly *et al.*, 1989; Surowy *et al.*, 1989).

In the nucleus, primary transcripts bind to hnRNP proteins to form hnRNP complexes, which consists of six major proteins, A1, A2, B1, B2, C1 and C2 (for a review see Dreyfuss *et al.*, 1988). Monoclonal antibodies against the hnRNP C proteins inhibit splicing *in vitro* (Choi *et al.*, 1986) and the hnRNP A1 protein was found to bind specifically to the 3′ splice site *in vitro* (Swanson and Dreyfuss, 1988a). These results have indicated that the A1 and C proteins are directly involved in splicing.

These RNA-binding proteins also have strong affinity for ssDNA. Based on this, procedures have been developed for the preparation of HeLa cell hnRNP proteins using an ssDNA affinity column (Pandolfo *et al.*, 1987; Piñol-Roma *et al.*, 1988). We were interested to know whether there are hnRNP-like proteins in chloroplasts, and in that case, whether these proteins are involved in splicing and/or processing of chloroplast RNAs. We used an ssDNA affinity column to isolate chloroplast ribonucleoproteins. Four tobacco chloroplast proteins bound strongly to an ssDNA column. The amino acid sequences of three out of the four proteins were derived from cDNA cloning and sequencing. Sequence analysis confirmed their RNP nature because they all contain two RNP-CS motifs. All three proteins also contain one very acidic amino-terminal domain, a novel feature among RNP proteins studied so far. We also present

the expression pattern of the genes for the three chloroplast RNP proteins.

Results

Isolation of chloroplast ssDNA-binding proteins

Total soluble proteins from tobacco (*Nicotiana tabacum* var. Bright Yellow 4) chloroplasts were loaded onto a calf thymus ssDNA cellulose column. A group of 28–33 kd proteins was eluted between 0.6 M and 2 M NaCl; these proteins were selected for further study because of their strong affinity for ssDNA. They were separated by electrophoresis, blotted onto polyvinylidene difluoride membrane and their N-termini were sequenced (Figure 1). From the protein sequencing data, the group of proteins can be divided into four species which were named 28, 30, 31 and 33 kd proteins according to their molecular weights. The N-terminus of the 30 kd protein was blocked. We failed to find the N-terminal sequences of the three proteins in any frames translated from the complete sequence of tobacco chloroplast DNA (Shinozaki *et al.*, 1986b). We therefore concluded that these proteins are encoded by the nuclear genome. From the electrophoretic pattern (Figure 1), each protein species has at least two bands with a slightly different mol. wt, but sharing almost identical N-terminal sequences. *N.tabacum* is an amphidiploid and the nuclear-encoded subunits of photosystem I show polymorphism caused by alloploidy (Obokata *et al.*, 1990).

Cloning and sequence analysis of cDNAs for the 28, 31 and 33 kd proteins

As it is difficult to purify these proteins further because of their low abundances, we isolated their cDNAs and thus determined their primary structures. About 1×10^5 clones of an amplified *Nicotiana sylvestris* leaf cDNA library in λgt10 were screened with oligonucleotide probes for the 28 and 31 kd proteins (*N.sylvestris* is the female progenitor of *N.tabacum*). Of these clones, seven and five positive clones were isolated for the 28 and 31 kd proteins, respectively. Those clones with the longest inserts were sequenced.

The cDNA sequence for the 28 kd protein revealed a reading frame of 276 amino acids (Figure 2a). The first 57 amino acid sequence is rich in hydroxylated residues and has an overall positive charge, and is therefore likely to be a transit peptide (Schmidt and Mishkind, 1986; Münch *et al.*, 1988). The mature protein is 219 amino acids long and has a predicted mol. wt of 24 504. Sequence analysis of a cDNA clone for the 31 kd protein revealed that there is a reading frame of 315 amino acids (Figure 2b). The first 71 amino acid sequence has the typical features of transit peptide: rich in hydroxylated residues and with a net positive charge. The mature protein is 244 amino acids long and has a predicted mol. wt of 27 385. Although the predicted mol. wt is lower than the apparent mol. wt of 31 kd, the mature protein has been expressed in *E.coli* and its apparent mol. wt is the same as that of the chloroplast one (Y.Li and M.Sugiura, unpublished data).

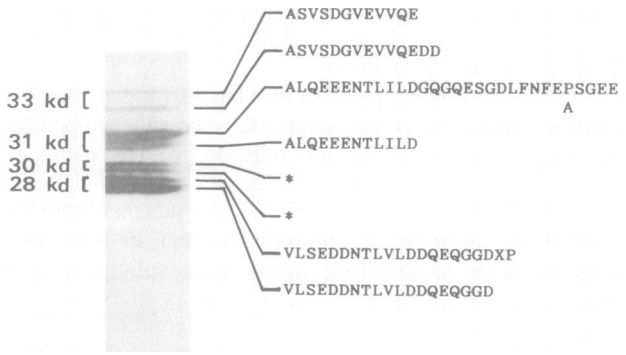

Fig. 1. Protein separation pattern on an SDS–polyacrylamide gel (transferred to polyvinylidene difluoride membrane). On the right are the determined N-terminal protein sequences. Asterisks indicate the blocked N-termini and X denotes unidentified residues. The 26th amino acid of the 31 kd protein is a mixture of alanine and proline. The proteins shown above were purified from intact chloroplasts which were isolated from ~150 g leaves.

The 33 kd protein is at least 10 times less abundant than the 28 and 31 kd proteins (Figure 1), therefore a cDNA library larger in size than the previous library was used to isolate its cDNA. An amplified *N.tabacum* leaf cDNA library was used. From $\sim 1 \times 10^6$ clones, 28 positive clones were isolated. One clone (33k-9) with the strongest hybridization signal was sequenced and found to be a partial length clone. The remaining positive clones were analyzed by PCR amplification of the inserts and restriction analysis of the PCR products. Clone 33k-1 has the longest 5′ extension from the N-terminus of the mature protein and was sequenced. The sequence contains a reading frame of 324 amino acids (Figure 2c). The first 71 amino acid sequence is likely to be a transit peptide because it is also rich in hydroxylated residues and has a net positive charge. The mature protein is 253 amino acids long with a predicted mol. wt of 27 485. The presence of putative transit peptides in the cDNAs for the 28 kd, 31 kd and 33 kd proteins indirectly supports their chloroplast location.

In vitro chloroplast import of the 31 kd protein precursor

To confirm the location of three proteins in the chloroplast, we did an *in vitro* import experiment to show that putative transit peptides are bona fide transit peptides. The 31 kd protein was chosen for this study because it has the longest 5′ untranslated sequence (Figure 2). The precursor of the 31 kd protein was made from the *in vitro* transcript using the wheatgerm system. The predicted mol. wt of the precursor from the nucleotide sequence is 35.4 kd; however, the relative mol. wt in an SDS gel was estimated to be 44 kd (Figure 3a, lane 1).

Radiolabeled precursors of 31 kd protein were used directly by incubation with the intact spinach chloroplasts, after which thermolysin was added to remove the precursors which were either present in the import medium or bound to the surface of the chloroplasts. The chloroplasts were then lysed and the total proteins were analyzed. The data in Figure 3a indicate that the 31 kd precursor was processed to the mature form with a relative mol. wt of 31 kd and was protease resistant after incubation with chloroplasts. The size was the same as the protein purified from the tobacco chloroplast (Figure 1). The results suggest that the N-terminal extension, the 71 amino acid putative transit peptide sequence, can direct the import of the 31 kd precursors into the chloroplasts and processing of the imported precursors to the mature forms. We concluded that the 71 amino acid N-terminal extension is a bona fide transit peptide and the 31 kd protein is of chloroplast location.

We have not done import experiments for the 28 kd and 33 kd proteins, partly because we thought that the 28 kd

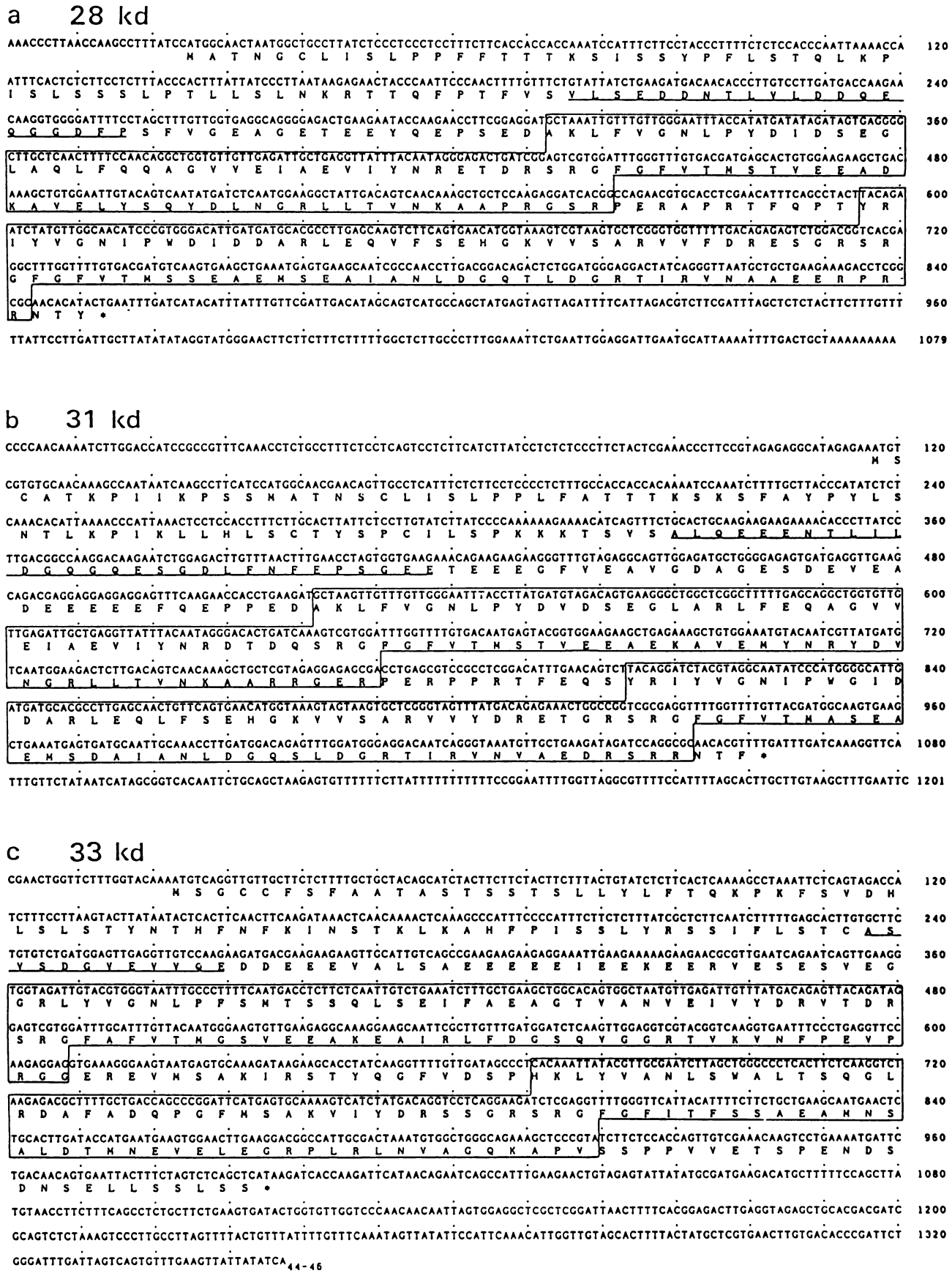

Fig. 2. cDNA sequences and predicted protein sequences of the 28, 31 and 33 kd proteins. The underlined amino acid sequences indicate those confirmed by protein sequencing. Asterisks indicate stop codons. Putative RNA-binding domains are boxed. The analyzed insert of the cDNA clone for the 31 kd protein was 2130 bp long, contained a poly(A) tract in the middle and is therefore a cloning artifact resulting from the ligation of two unrelated cDNAs in one vector. Its 3'-half (b, 1201 bp) is for the 31 kd protein and hybridized to a single RNA band of 1300 bp. The sequence of clone 33k-1 is presented (c, clone 33k-9 from nucleotide 203 to nucleotide 1343).

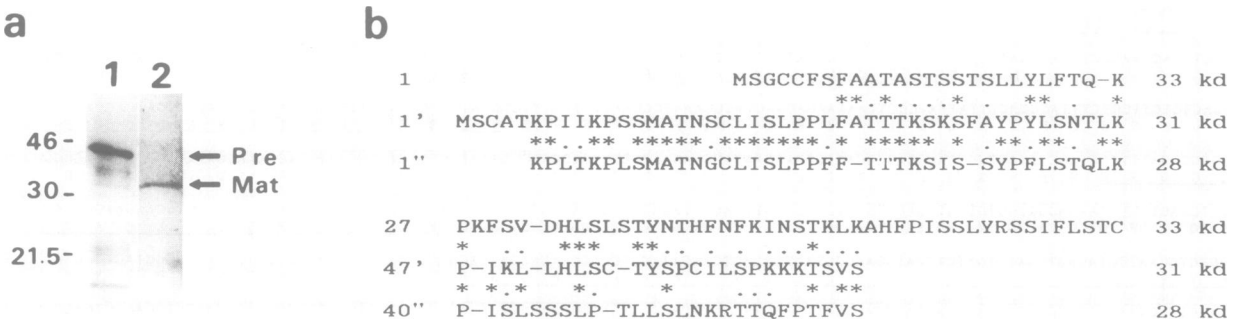

Fig. 3. (**a**) Import of the 31 kd protein precursor by intact chloroplasts. RNA was synthesized by T7 RNA polymerase from a Bluescript vector carrying the 31 kd protein's cDNA and translated in a wheatgerm extract in the presence of [^{35}S]methionine. 1: labeled products of *in vitro* translation; 2: processed 31 kd protein in protease treated chloroplasts. Mol. wt markers are marked on the left in kd. Pre: precursor; Mat: processed 31 kd mature protein. (**b**) Comparison of the putative transit peptides of the 28, 31 and 33 kd proteins. Asterisks indicate identical residues and dots indicate conservative substitution. Dashes denote gaps introduced to optimize sequence alignment. The synthesis of the cDNA for the 28 kd protein stopped prematurely and the transit peptide of the 28 kd protein shown above is a truncated one.

```
31 kd  ALQEEENTLILDGQGQESGDLFNFEPSGEETEEEGFVEAVGDAGESDEVEADEEEEEFQEPPEDAKLFVGNLPYDVDSEGLA   82
       .* *..***.**.* ...**      **            **.***.          **.***.*************.******
28 kd  VLSEDDNTLVLDDQ-EQGGDF----PSF----------VGEAGET---------EEYQEPSEDAKLFVGNLPYDIDSEGLA   57
       ^                                                        ^^
       ¦<---------        acidic amino-terminal domain        -------->¦¦<---------

31 kd  RLFEQAGVVEIAEVIYNRDTDQSRGFGFVTMSTVEEAEKAVEMYNRYDVNGRLLTVNKAARRGERPERPPRTFEQSYRIYVG  164
       .**.***************.**.***************.****.*..**.***********.**.****.****...******
28 kd  QLFQQAGVVEIAEVIYNRETDRSRGFGFVTMSTVEEADKAVELYSQYDLNGRLLTVNKAAPRGSRPERAPRTFQPTYRIYVG  137
                                                                    ^                ^
                     RNA-binding domain I                          -------->¦         ¦<----

31 kd  NIPWGIDDARLEQLFSEHGKVVSARVVYDRETGRSRGFGFVTMASEAEMSDAIANLDGQSLDGRTIRVNVAEDRSRRNTF     244
       ****.********.*************.***.***********.******.********.*********.**.*.****.
28 kd  NIPWDIDDARLEQVFSEHGKVVSARVVFDRESGRSRGFGFVTMSSEAEMSEAIANLDGQTLDGRTIRVNAAEERPRRNTY     219
                                                                                  ^
       ----                    RNA-binding domain II                          -------->¦
```

Fig. 4. Comparison of the deduced amino acid sequences of the 28 and 31 kd proteins. Asterisks indicate identical residues and dots indicate conservative substitution. Dashes denote gaps introduced to optimize sequence alignment. The putative RNA-binding domains and acidic N-terminal domains are indicated.

```
protein source                                  10        20        30        40        50        60        70        80
tobacco 28 kd     I    40  AKLFVGNL-PYDIDSEGLAQLFQQAGVVEIAEVIYNRETDRSRGFGFVTMSTVEEADKAVELYSQYDLNGRLLTVNKAAPRGSR
                  II  134  YRITVGNI-PNDIDDARLEQVFSEHGKVVSARVVFDRESGRSRGFGFVTMSSEAEMSEAIANLDGQTLDGRTIRVNAAEERPRR
tobacco 31 kd     I    65  AKLFVGNL-PYDVDSEGLARLFEQAGVVEIAEVIYNRDTDQSRGFGFVTMSTVEEAEKAVEMYNRYDVNGRLLTVNKAARRGER
                  II  159  YRITVGNI-PNGIDDARLEQLFSEHGKVVSARVVYDRETGRSRGFGFVTMASEAEMSDAIANLDGQSLDGRTIRVNVAEDRSRR
tobacco 33 kd     I    43  GRLYVGNL-PFSMTSSQLSEIFAEAGTVANVEIVIDRVTDRSRGFAFVTMGSVEEAKEAIRLFDGSQVGGRTVKVNFPEVPRGG
                  II  146  HKLYVANL-SWALTSQGLRDAFADQPGFMSAKVIYDRSSGRSRGFGFITFSSAEAMNSALDTMNEVELEGRPLRLNVAGQKAPV
maize AAIP           8   YRCFVGGL-AWATSNESLENAFASTGEILDSKVITDRETGRSRGFGFVTFSSENSMLDAIENMNGKELDGRNITVNQAQSRG
yeast PABP        I   38  ASLYVGDL-EPSVSEAHLYDIFSPIGSVSSIRVCRDAITKTSLGYAYVNFNDHEAGRKAIEQLNYTPIKGRLCRIMWSQRDPSL
                  II  126  GNIFIKNL-HPDIDNKALYDTFSVFGDILSSKIATD-ENGKSKGFGFVHFEEEGAAKEAIDALNGMLLNGQEIYVAPHLSRKER
                  III 219  TNLYVKNI-NSETTDEQFQELFAKFGPIVSASLEKDAD-GKLKGFGFVNYEKHEDAVKAVEALNDSELNGEKLYVGRAQKKNER
                  IV  322  VNLFVKNL-DDSVDDEKLEEEFAPYGTITSAKVMRT-ENGKSKGFGFVCFSTPEEATKAITEKNQQIVAGKPLYVAIAQRKDVR
human hnRNP A1    I   14  RKLFIGGL-SFETTDESLRSHFEQWGTLTDCVVMRDPNTKRSRGFGFVTYATVEEV-DAAMNARPHKVDGRVVEPKRAVSREDS
                  II  105  KKIFVGGI-KEDTEEHHLRDYFEQFGKIEVIEIMTDRGSGKKRGFAFVTFDDHDSV-DKIVIQKYHTVNGHNCEVRKALSKQEM
human U1snRNP  70kd 103  KTLFVARV-NYDTTESKLRREFEVYGPIKRIHMVYSKRSGKPRGYAFIEYEHERDM-HS----AYKHADGKKIDGRRVLVDVER
human hnRNP C     16  SKVFIGNLNTLVVKKSDVEAIFSKYGKIVGCSV--------HKGFAFVQYVNERNARAAVAGEDGRMIAGQVLDINLAAEPKVN
consensus           LFVG          L   F   G V    V         RGFGFV F        AI         GR   V
                    IYI                     I               K  A   Y         V          K   I
                                                                →|RNP-CS|←
```

Fig. 5. Comparison of the RNA-binding domains from tobacco chloroplast, maize (Gómez *et al.*, 1988), yeast (Sachs *et al.*, 1986) and human (Buvoli *et al.*, 1988; Query *et al.*, 1989; Swanson *et al.*, 1987). Identical amino acids or amino acids with similar properties are boxed. Dashes denote gaps introduced to optimize sequence alignment. AAIP: abscisic acid induced glycine-rich protein. PABP: poly(A)-binding protein.

cDNA clone is probably a partial length clone and the real translation start might be 13 amino acids ahead of the putative translation start shown in Figure 2a. Comparison of the putative transit peptides of the 28 kd and 33 kd proteins with that of the 31 kd protein revealed that they share high homology (Figure 3b). If the frame of the 28 kd protein cDNA is allowed to extend from the first methionine codon, the resulted truncated transit peptide shares a global homology of 55% amino acid identity to that of 31 kd protein, while the putative transit peptide of the 33 kd protein shows 30% amino acid identity out of 47 amino acids. From the high homology we concluded that the 28 and 33 kd proteins are also of chloroplast location. In conclusion, the three proteins are encoded by the nuclear genome and are located in the chloroplast.

The three proteins contain RNP-CS and acidic amino terminal domains

Comparison of the amino acid sequences of the above three proteins revealed that they are globally homologous to each other. The 31 and 28 kd proteins are highly homologous, having 79% amino acid identity over their entire sequences

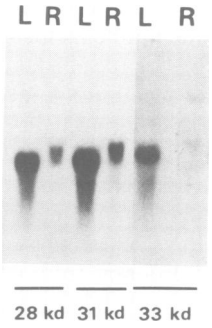

Fig. 6. Northern blot analysis of transcripts from the three protein genes in *N.sylvestris* leaf (L) and root (R) total RNA (30 μg each RNA). Probes used for the three genes were: 28 kd, positions 1−1079; 31 kd, positions 16−1110; 33 kd, positions 203−1343 (Figure 2). The films were exposed for 32 h (28 kd and 31 kd) and for 4 weeks (33 kd) at −70°C.

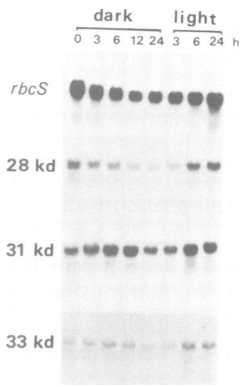

Fig. 7. Transcript levels of *rbcS* and the 28, 31 and 33 kd protein genes in light/dark-shifted tobacco plants. Tobacco plants were grown in a growth chamber for 6 weeks. At 10.30 am the light (~4000 lux) was turned off and the plants were kept in absolute darkness for 24 h, after which the light was turned on again. Leaves were harvested 0, 3, 6, 12 and 24 h after the dark treatment and 3, 6 and 24 h after the light was turned on again. Total RNAs (30 μg) were hybridized with probes for *rbcS*, 28 kd, 31 kd and 33 kd cDNAs. The probes used for the three protein genes were the same as in Figure 6. The films were exposed for 11 h (*rbcS*), 21 h (28 kd and 31 kd) and 6 days (33 kd) at −70°C.

(Figure 4) while the 33 kd protein has 37% amino acid homology to both the 31 kd and 28 kd proteins. A homology search using the SWISS-PROT database showed that the highest scores go to rat and bovine hnRNP A1 proteins, yeast poly(A)-binding protein and mouse and golden hamster nucleolins. The homology areas extend as long as 180−200 amino acids.

The three proteins contain two tandem repeats corresponding to RNA-binding domains in their sequences. Each protein contains two RNP-CS motifs in the two putative RNA-binding domains (Figure 5). Both the RNP-CS motif and RNA-binding domain are highly conserved among tobacco chloroplast, maize, yeast and human proteins. Maize abscisic acid-induced glycine-rich protein (AAIP) is the only plant protein so far published that contains a putative RNA-binding domain (Gómez *et al.*, 1988; Mortenson and Dreyfuss, 1989). Based on their strong affinity for ssDNA and their high homology to typical RNP proteins, we concluded that the three tobacco proteins are RNP proteins. The region of the putative RNA-binding domains (Figure 4; 180 amino acids starting from each C-terminus) is highly conserved between the 28 and 31 kd proteins. About 85% of the amino acids are identical, substitution of the amino acids with the same properties accounts for another 3.3%.

In addition to the two putative RNA-binding domains, each of the three tobacco proteins contains an acidic N-terminal domain. The first 39 amino acids of the 28 kd protein (or 64 for the 31 kd protein) contains 36% acidic amino acids (42% for the 31 kd protein) while there are no basic amino acids in the region. For the 33 kd protein, the overall net negative charge is still as high as 43% although there are two basic amino acids in the first 42 amino acids.

The genes for the three proteins are expressed both in mature leaves and roots and have different light-modulated expression patterns

As an approach towards understanding the function of these tobacco RNP proteins, we studied the expression of their genes in different tobacco tissues and their response to light in mature leaves. The three proteins are each encoded by a single-copy nuclear gene in *N.sylvestris* (Y.Li and M.Sugiura, unpublished data). Northern blot analysis of the leaf and root total RNA revealed that the three genes were expressed both in leaves and roots (Figure 6), but the levels in leaves were 6- to 10-fold higher than in roots. We then used a similar experimental design to that described by Sugita and Gruissem (1987) to study the light-modulation of transcript levels of the three genes in tobacco. The transcript level of the 28 kd protein gene decreased after the light was turned off and increased when the light was turned on again (Figure 7). However the transcript levels of the 31 and 33 kd protein genes were unaffected by light (a slight increase was observed with time).

Discussion

The development of *in vitro* chloroplast splicing systems and the characterization of components involved in the splicing of chloroplast RNAs are necessary for the elucidation of the molecular mechanism and regulation of the chloroplast splicing system. Since no *in vitro* chloroplast splicing systems are available at present, an alternative approach is to characterize the proteins involved in this process. As a first step, using the purification procedure for HeLa hnRNP proteins, we isolated the chloroplast ribonucleoproteins. Sequence analysis of the cDNA clones from the 28, 31 and 33 kd proteins established their RNP nature. In addition, we have shown that their putative transit peptides are functional by an *in vitro* import experiment, therefore confirming their chloroplast location. To our knowledge, these tobacco chloroplast proteins are the first RNP proteins analyzed from organelles.

RNP proteins or RNA-binding proteins are generally composed of two parts: one to four RNA-binding domains arranged in tandem and one or several auxiliary domains (Bandziulis *et al.*, 1989). The RNA-binding domain is responsible for RNA-binding. Figure 5 shows a current collection of the typical RNA-binding domains as well as the six chloroplast putative RNA-binding domains. Positively charged amino acids and aromatic amino acids have been suggested to be important in RNA−protein interaction through sequential intercalation of the aromatic residues with nucleotide bases and the interaction of the positively charged

amino acids with the negatively charged phosphodiester backbone (Adam *et al.*, 1986). The amino acids at positions 4, 22, 45 and 47 are exclusively aromatic. Residues 4 and 47 have been shown to be indispensable for RNA-binding (Scherly *et al.*, 1989) and to form part of RNA-binding pockets by UV-crosslinking experiments (Merrill *et al.*, 1988). Some other amino acids are also highly conserved, e.g. branched chain aliphatic residues (residues 3, 5, 8, 18, 28, 33, 48, 59, 60, 68, 73 and 75) and basic residues (residues 43, 71 and 81) as well as highly conserved glycines (residues 26, 44 and 70). In the six putative chloroplast RNA-binding domains, five out of the six amino acids at the last position of the RNP-CS octapeptide (positions 43–50) are methionines instead of aromatic residues. Among RNP-CS motifs so far identified, only the third RNP-CS motif of the *Drosophila elav* gene product has a methionine in the same position (Robinow *et al.*, 1988). The finding that the chloroplast proteins contain the highly conserved putative RNA-binding domains found in nuclear RNP proteins strongly supports the hypothesis that RNP proteins including those from organelles have a common evolutionary origin (Adam *et al.*, 1986; Dreyfuss *et al.*, 1988).

The auxiliary domain in RNA-binding proteins is suggested to function primarily in protein–protein interaction. The three chloroplast RNP proteins have one auxiliary acidic domain in their N-termini. Over one third of the residues in their N-terminal domains are acidic amino acids. The mammalian hnRNP C protein contains an acidic domain of ~100 amino acids (~25% net negative charge) at its C-terminus, which is postulated to be involved in protein–protein interaction when forming a spliceosome by neutralizing the positive charge of basic proteins (Preugschat and Wold, 1988).

The three chloroplast RNP proteins are structurally distinct from any known RNA-binding proteins (Figure 8). They are different from the hnRNP C proteins because they have two RNA-binding domains (compared to one in hnRNP C) and have a more acidic domain (36–43% net negative charge) not in the C-termini but in the N-termini. The three chloroplast RNP proteins are also different from nucleolin, which has four putative RNA-binding domains. In the N-terminal part of nucleolin, there are areas with long strings of acidic amino acids, but they are divided by strings of residues rich in positively charged residues.

Though the three chloroplast proteins were eluted from the ssDNA column with 2 M NaCl, we detected a small amount of the 31 kd protein and a substantial amount of the 33 kd protein but not the 28 kd protein in a 0.6 M NaCl eluted fraction. The three proteins therefore have different affinities for ssDNA in the order 28 kd > 31 kd > 33 kd. In addition, the steady-state mRNA level in mature leaves showed that there were different light-modulation patterns for the 28 and 31 kd proteins. These observations suggest that the 28 and 31 kd proteins play different roles in the chloroplast RNA metabolism, despite their high homology. In conclusion, the three chloroplast RNP proteins are probably functionally different from one another.

During the preparation of this manuscript, the primary structures of HeLa cell hnRNPs A2, B1 and C2 were published (Burd *et al.*, 1989). Their results indicate that the diversity of hnRNP proteins is generated by small peptide inserts. The A2 and B1 proteins are very similar to the A1 protein: the RNA-binding domains of A2 and B1 have ~ 80% amino acid identity with those of A1. The structural similarity between A2/B1 and A1 is reminiscent of the similarity between the 28 and 31 kd proteins (Figure 3). From the similarities between the chloroplast RNP proteins and the hnRNP A and B proteins, we speculate that the three chloroplast proteins play the same role as the hnRNP proteins and are involved in splicing (and/or processing) of chloroplast RNAs. Our speculation is supported by our recent observation that the 28 and 31 kd proteins selectively bind to certain ribonucleotide homopolymers and that this binding is salt- and heparin-resistant (Y.Li and M.Sugiura, unpublished data). The affinities for different ribonucleotide homopolymers have been suggested to be a useful means of comparing proteins from other sources with HeLa cell hnRNP proteins (Swanson and Dreyfuss, 1988b). The N-terminal acidic domains might be involved in protein–protein interaction when forming a spliceosome. The different affinities for ssDNA led us to speculate that their RNA-binding specificities are also different.

Although we have no direct evidence that these chloroplast RNP proteins are involved in the splicing of chloroplast RNAs, our findings may imply that the splicing machinery is also conserved between chloroplasts and nuclei. Recently, a protein splicing factor associated with U5 snRNP has been shown to be conserved between yeast and man (Anderson *et al.*, 1989).

One of the central questions as to the function of these proteins may be their binding specificities. Tobacco is one of the best systems for solving this problem. The complete sequence of tobacco chloroplast DNA has been determined and we have mapped two-thirds of the transcription units

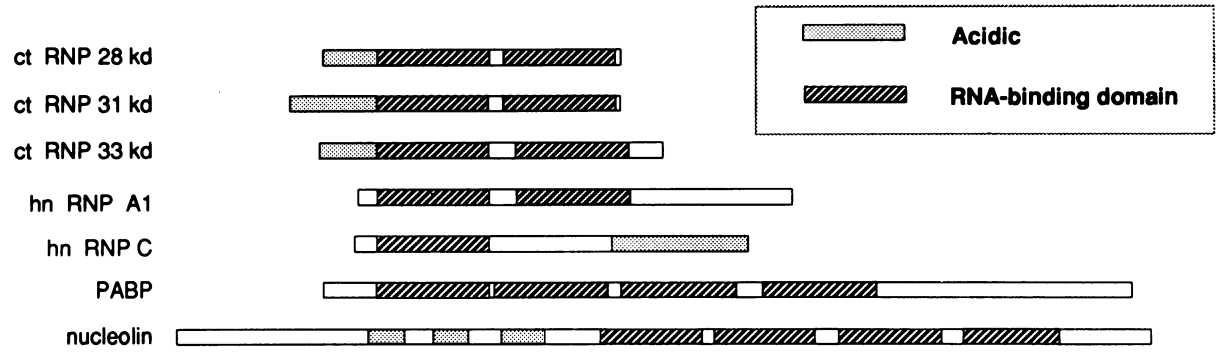

Fig. 8. Domain arrangement of the three chloroplast RNP proteins, human hnRNP A1 and C proteins (Buvoli *et al.*, 1988; Swanson *et al.*, 1987), yeast poly(A)-binding protein (PABP, Sachs *et al.*, 1986) and mouse nucleolin (Bourbon *et al.*, 1988). Hatched boxes indicate RNA-binding domains and stippled boxes acidic domains or regions.

on the chloroplast genome. All the available data will facilitate the quick identification of the RNAs to which they bind and enable us to determine the exact binding specificity of these proteins and study the chloroplast RNA−protein interaction in more detail.

Materials and methods

Plant materials

Tobacco, *N.tabacum* var. Bright Yellow 4, was grown in a greenhouse for protein extraction; for RNA extraction, *N.sylvestris* was grown in a growth chamber at 25°C with 12 h light/12 h dark periods. Roots were harvested from 4 week-old hydroponically grown tobacco plants. Spinach was purchased at a local market.

Preparation of intact chloroplasts

The method for isolation of tobacco and spinach intact chloroplasts was modified from that of Bartlett *et al.* (1982). Briefly, 200 g leaves in 600 ml MCB1 (0.30 M mannitol, 50 mM HEPES−NaOH, pH 8, 2 mM EDTA, 5 mM β-mercaptoethanol) with 0.1% bovine serum albumin (BSA) and 0.6% polyvinylpyrrolidone were ground with a Polytron homogenizer. The resultant homogenate was filtered through 4 layers of cotton gauze. The filtrate was centrifuged at 3000 r.p.m. (Tomy No. 9N rotor) and with immediate stop. The green pellet was resuspended in 12 ml of MCB1 with 0.1% BSA and loaded onto four 10−80% Percoll gradients in MCB1. After centrifugation at 8000 r.p.m. (Tomy TS-2 rotor) for 30 min, the lower green band was collected and mixed with 3−5 volumes of MCB2 (0.32 M mannitol, 50 mM HEPES−NaOH, pH 8, 2 mM EDTA). Chloroplasts were collected by centrifugation at 3000 r.p.m. (Tomy No. 4N rotor) for 30 s and washed once with MCB2.

Purification and sequencing of ssDNA-binding proteins

The intact chloroplasts from 1 kg tobacco leaves were prepared as above and stored at −80°C. The frozen chloroplasts were thawed and homogenized with 40 ml of buffer A (20 mM Tris−HCl, pH 8.0, 2 mM PMSF, 2 mM DTT) and centrifuged for 30 min at 30 000 *g*. Solid $(NH_4)_2SO_4$ was added to the supernatant to 80% saturation. The pellet was dissolved in 60 ml buffer B (10 mM Tris−HCl, pH 8.0, 0.1 M NaCl, 10% glycerol, 1 mM PMSF, 1 mM EDTA, 0.1 mM DTT) and precipitated again with $(NH_4)_2SO_4$. After centrifugation the pellet was dissolved in 15 ml buffer B and applied to a 1 ml ssDNA cellulose (Sigma) column equilibrated with the same buffer. The column was washed with 10 column volumes of buffer B and the bound proteins were eluted with 6 column volumes each of 0.3 M, 0.6 M and 2 M NaCl in buffer B. The 2 M fraction was pooled, concentrated as described by Wessel and Flügge (1984) and separated on a 7.5−20% polyacrylamide gradient gel containing 0.1% SDS (Laemmli, 1970). The proteins were then transferred to polyvinylidene difluoride membrane, stained with Coomassie R 350 and sequenced using an Applied Biosystems 470A gas-phase sequencer (Matsudaira, 1987).

Screening cDNA libraries

N.sylvestris leaf cDNA library was a generous gift from Dr A.Kato. An *N.tabacum* leaf cDNA library was constructed by Dr R.F.Whittier using the cDNA synthesis kit (Pharmacia). A λgt10 vector was used for both libraries. Mixed probes were synthesized according to the determined amino acid sequences using an Applied Biosystems 381A synthesizer. The sequences and the hybridization temperatures of the probes used were:

28 kd protein,	probe PR6:	AGNACNAGNGTGTTGTCG TCTTC	51°C
	probe PR7:	GTCNCCNCCTTGTTCTTGG TCGTC	55°C
31 kd protein,	probe PR2:	TAITYICAYAARAGRAGRA GRACITYTCG	55°C
	probe PR4:	TTCTTGTTCTTGNCCGTCN AGGAT	52°C
33 kd protein,		TTCYTGIACIACYTCIACICC RTCI(G/C)(A/T)IACI(G/C)(A/T)IGC	50°C

(N = G,A,T and C; R = A and G; Y = T and C; I = Inosine)

Screening of the libraries was essentially according to the instruction manual from Amersham (cDNA synthesis kit, λgt10) and others (Wallace and Miyada, 1987). Positive clones were purified and their inserts were amplified by PCR. Primers used for PCR amplification were: 1233 ACCGTTTCGAGCTGCTCTATAGACT and 1234 CTTGTGAAGATTGGGGGTAAATAACA. The PCR was performed for 25 cycles. Each cycle consisted of treatments at 92°C for 1 min, 55°C for 2 min and 72°C for 4 min (for 15 min in the last cycle).

Subcloning and DNA sequencing

Recombinant λ DNAs were prepared by a mini-lysate method (Manfioletti and Schneider, 1988). The inserts were cut with *Eco*RI, subcloned into Bluescript M13(+) plasmid and sequenced. Sequencing and computer-assisted analysis were essentially as described by Hiratsuka *et al.* (1989).

In vitro transcription/translation and import assays

The plasmid used to make *in vitro* transcripts of the 31 kd protein was deletion clone 8A81. In this plasmid, most of the artificial 5′ half except for 13 A residues upstream of the authentic cDNA has been removed (Figure 2). The plasmid was linearized by cutting with *Eco*RI, and was then treated with proteinase K and extracted with phenol−chloroform. 5 μg of template were transcribed in the presence of 50 U of T7 RNA polymerase, 0.45 mM each of ATP, CTP and UTP, 50 μM GTP and 500 μM of the cap structure analog, $m^7G(5')ppp(5')G$. The subsequent treatment was essentially according to Sambrook *et al.* (1989).

Wheatgerm extract was a generous gift from Dr A.Watanabe and Mr T.Konishi; it was prepared essentially as in Marcu and Dudock (1974). About 1 μg of capped transcripts was utilized to program the [^{35}S]methionine-based wheatgerm cell-free translation system. Translation reactions were done in a final volume of 50 μl in the presence of 70 μCi [^{35}S]methionine (specific activity > 1000 Ci/mmol) (Marcu and Dudock, 1974). Following incubation at 30°C for 1 h, 1 μg RNase A was added and incubated for 30 min at 37°C. One-fourth volume of 5 × sorbitol−HEPES buffer (1.65 M sorbitol; 250 mM HEPES−KOH, pH 7.5) and 3 μl 100 mM cold methionine were added to the translation mixture, mixed well and stored at −70°C until use.

Import was performed essentially as described by Bartlett *et al.* (1982) and Minami *et al.* (1986). Intact spinach chloroplasts were prepared as described in the *'Preparation of intact chloroplasts'* section above, except that the final wash was done with sorbitol−Tricine (0.33 M sorbitol, 50 mM Tricine−KOH, pH 8.4). The chloroplast pellet was then resuspended in 1 ml 1 × sorbitol−HEPES and its chlorophyll concentration was determined. The translation products were centrifuged at 15 000 *g* for 10 min. and half of the supernatant was diluted with 175 μl of 1 × sorbitol−HEPES in a 5 ml glass test tube, and 100 μl intact chloroplasts were added (125 μg total chlorophyll). The suspension was incubated for 1 h on a rotary shaker at 110 r.p.m. inside a 25°C growth chamber with light intensity of ~4000 lux. The sample was transferred into a 1.5 ml plastic tube and centrifuged at 5000 r.p.m. for 1 min. The chloroplasts were resuspended in 0.3 ml ice-cold 1 × sorbitol−HEPES. $CaCl_2$ and thermolysin were added at final concentrations of 1 mM and 0.17 mg/ml, respectively. The sample was stood on ice for 30 min and then EGTA was added at a final concentration of 2.5 mM to inactivate the protease. The mixture was centrifuged at 5000 r.p.m. for 1 min and the pellet was washed twice with 0.3 ml ice-cold sorbitol−Tricine containing 1 mM EGTA. Sixty μl SDS−PAGE loading buffer were added to the pellet, heated at 100°C for 2 min. and centrifuged. The supernatant was concentrated as described by Wessel and Flügge (1984) and applied to a 15% SDS−PAGE (Laemmli, 1970). Fluography was done according to Skinner and Griswold (1983).

RNA isolation and Northern blot analysis

Total leaf and root RNA were obtained from *N.sylvestris* according to the method of Vries *et al.* (1988). The poly(A)$^+$ RNA was isolated by purification twice on oligo-dT cellulose (BRL) columns according to the instruction manual (cDNA synthesis kit, Pharmacia). RNA electrophoresis and Northern blotting to Hybond-N membrane were essentially according to Davis *et al.* (1986). The DNA fragments from the three cDNAs were labeled by the random primer method to specific activities of 5×10^8 to 2×10^9 c.p.m./μg (Feinberg and Vogelstein, 1984). Hybridization and washing of the membrane were done essentially according to Amersham's protocol except that the stringent wash was conducted at 65°C in 0.1 × SSPE with 0.1% SDS for 20 min. The sequence of a probe for *rbcS* is CTTCTTGTTAATTGGTGGCCACACCTGCAT, which is located in the transit peptide (Pinck *et al.*, 1984). The probe was labeled at its 5′ end by polynucleotide kinase to a specific activity of 1.2×10^7 c.p.m./μg. Hybridization and washing were done essentially according to Angelini *et al.* (1986) except that the membrane was finally washed with 6 × SSC containing 0.1% SDS at 65°C for 30 min. The used blot was put into a boiled 0.1% SDS solution and allowed to cool to room temperature for repeated use.

Acknowledgements

We thank Dr A.Kato and Dr R.F.Whittier for tobacco leaf cDNA libraries, Dr J.H.Weil for correcting English, Dr N.Hayashida and Dr K.Kobayashi for protein sequencing, Dr R.F.Whittier for screening the 33 kd protein cDNA clones, Miss S.Nagayoshi for sequencing the 33k-1 clone, Dr A.Watanabe and Mr T.Konishi for the wheatgerm extract and advice on the import experiment, Dr J.Obokata and Dr T.Matsubayashi for establishing the modified version for isolation of intact chloroplasts. We also thank Dr M.Sugita, Dr T.Wakasugi and Miss L.Ye for valuable discussions and continuous encouragement. We are grateful to Dr W.Gruissem and Dr D.B.Stern for their valuable suggestions and preprints. Y.Li was supported by a Monbusho fellowship. This research was supported by a Grant-in-Aid from the Ministry of Education, Science and Culture (Japan).

References

Adam,S.A., Nakagawa,T., Swanson,M.S., Woodruff,T.K. and Dreyfuss,G. (1986) *Mol. Cell Biol.*, **6**, 2932–2943.
Anderson,G.J., Bach,M., Lührmann,R. and Beggs,J.D. (1989) *Nature*, **342**, 819–821.
Angelini,G., de Preval,C., Gorski,J. and Mach,B. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4489–4493.
Bandziulis,R.J., Swanson,M.S. and Dreyfuss,G. (1989) *Genes Dev.*, **3**, 431–437.
Barlett,S.G., Grossman,A.R. and Chua,N.-H. (1982) In Edelman,M., Hallick,R.B. and Chua,N.-H. (eds), *Methods in Chloroplast Molecular Biology*. Elsevier Press, NY, pp. 1081–1091.
Bourbon,H., Lapeyre,B. and Amalric,F. (1988) *J. Mol. Biol.*, **200**, 627–638.
Burd,C.G., Swanson,M.S., Görlach,M. and Dreyfuss,G. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9788–9792.
Buvoli,M., Biamonti,G., Ghetti,A., Riva,S., Bassi,M.T. and Morandi,C. (1988) *Nucleic Acids Res.*, **16**, 3751–3770.
Choi,Y.D., Grabowski,P.J., Sharp,P.A. and Dreyfuss,G. (1986) *Science*, **231**, 1534–1539.
Davis,L.G., Dibner,M.D. and Battey,J.F. (1986) *Methods in Molecular Biology*. Elsevier Press, NY, pp. 139–143.
Dreyfuss,G., Swanson,M.S. and Pinõl-Roma,S. (1988) *Trends Biochem. Sci.*, **13**, 86–91.
Feinberg,A.P. and Vogelstein,B. (1984) *Anal. Biochem.*, **137**, 266–267.
Gómez,J., Sanchez-Martínez,D., Stiefel,V., Rigau,J., Puigdomenèch,P. and Pagès,M. (1988) *Nature*, **334**, 262–264.
Hiratsuka,J., *et al.* (1989) *Mol. Gen. Genet.*, **217**, 185–194.
Kück,U., Choquet,Y., Schneider,M., Dron,M. and Bennoun,P. (1987) *EMBO J.*, **6**, 2185–2195.
Laemmli,U.K. (1970) *Nature*, **227**, 680–686.
Manfioletti,G. and Schneider,C. (1988) *Nucleic Acids Res.*, **16**, 2873–2884.
Marcu,K. and Dudock,B. (1974) *Nucleic Acids Res.*, **1**, 1385–1397.
Matsudaira,P. (1987) *J. Biol. Chem.*, **262**, 10035–10038.
Mattaj,I.W. (1989) *Cell*, **57**, 1–3.
Merrill,B.M., Stone,K.L., Cobianchi,F., Wilson,S.H. and Williams,K.R. (1988) *J. Biol. Chem.*, **263**, 3307–3313.
Minami,E., Shinohara,K., Kuwabara,T. and Watanabe,A. (1986) *Arch. Biochem. Biophys.*, **244**, 517–527.
Mortenson,E. and Dreyfuss,G. (1989) *Nature*, **337**, 312.
Münch,S., Ljungberg,U., Steppuhn,J., Schneiderbauer,A., Nechushtai,R., Beyreuther,K. and Herrmann,R.G. (1988) *Curr. Genet.*, **14**, 511–518.
Obokata,J., Mikami,K., Hayashida,N. and Sugiura,M. (1990) *Plant Physiol.*, **92**, 273–275.
Pandolfo,P., Valentini,O., Biamonti,G., Rossi,P. and Riva,S. (1987) *Eur. J. Biochem.*, **162**, 213–220.
Pinck,M., Guilley,E., Durr,A., Hoff,M., Pinck,L. and Fleck,J. (1984) *Biochimie*, **66**, 539–545.
Piñol-Roma,S., Choi,Y.D., Matunis,M.J. and Dreyfuss,G. (1988) *Genes Dev.*, **2**, 215–227.
Preugschat,F. and Wold,B. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 9669–9673.
Query,C.C., Bentley,R.C. and Keene,J.D. (1989) *Cell*, **57**, 89–101.
Robinow,S., Campos,A.R., Yao,K.-M. and White,K. (1988) *Science*, **242**, 1570–1572.
Rock,C.D., Barkan,A. and Taylor,W.C. (1987) *Curr. Genet.*, **12**, 69–77.
Sachs,A.B., Bond,M.W. and Kornberg,R.D. (1986) *Cell*, **45**, 827–835.
Sambrook,J., Maniatis,T. and Fritsch,E.F. (1989) *Molecular Cloning*. 2nd edition. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.
Scherly,D., Boelens,W., van Venrooij,W.J., Dathan,N.A., Hamm,J. and Mattaj,I.W. (1989) *EMBO J.*, **8**, 4163–4170.
Schmidt,G.W. and Mishkind,M.L. (1986) *Annu. Rev. Biochem.*, **55**, 879–912.
Shinozaki,K., Deno,H., Sugita,M., Kuramitsu,S. and Sugiura,M. (1986a) *Mol. Gen. Genet.*, **202**, 1–5.
Shinozaki,K., *et al.* (1986b) *EMBO J.*, **5**, 2043–2049.
Skinner,M.K. and Griswold,M.D. (1983) *Biochem. J.*, **209**, 281–284.
Sugita,M. and Gruissem,W. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 7104–7108.
Sugiura,M. (1989) *Annu. Rev. Cell Biol.*, **5**, 51–70.
Surowy,C.S., van Santen,V.L., Scheib-Wixted,S.M. and Spritz,R.A. (1989) *Mol. Cell. Biol.*, **9**, 4179–4186.
Swanson,M.S. and Dreyfuss,G. (1988a) *EMBO J.*, **7**, 3519–3529.
Swanson,M.S. and Dreyfuss,G. (1988b) *Mol. Cell. Biol.*, **8**, 2237–2241.
Swanson,M.S., Nakagawa,T.Y., Levan,K. and Dreyfuss,G. (1987) *Mol. Cell. Biol.*, **7**, 1731–1739.
Tanaka,M., Obokata,J., Chunwongse,J., Shinozaki,K. and Sugiura,M. (1987) *Mol. Gen. Genet.*, **209**, 427–431.
Vries,S. de, Hoge,H. and Bisseling,T. (1988) In Gelvin,S.B., Schilperoort,R.A. and Verma,D.P.S. (eds), *Plant Molecular Biology Manual*. Kluwer Academic Publishers, Dordrecht, B-6.
Wallace,R.B. and Miyada,C.G. (1987) *Methods Enzymol.*, **152**, 432–442.
Wessel,D. and Flügge,U.I. (1984) *Anal. Biochem.*, **138**, 141–143.
Westhoff,P., Farchaus,J.W. and Herrmann,R.G. (1986) *Curr. Genet.*, **11**, 165–169.
Westhoff,P. and Herrmann,R.G. (1988) *Eur. J. Biochem.*, **171**, 551–564.
Zaita,N., Torazawa,K., Shinozaki,K. and Sugiura,M. (1987) *FEBS Lett.*, **210**, 153–156.

Received on April 5, 1990; revised on June 19, 1990

Note added in proof

Since this paper was submitted, Buvoli *et al.*, (*EMBO J.*, **9**, 1229–1235, 1990) reported the sequence of human hnRNP B2 protein. The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X53932 (33 kd), X53933 (28 kd) and X53942 (31 kd).