Mol Gen Genet (1991) 228:89–96
002689259100215D

MGG

RNA-protein interactions at transcript 3′ ends and evidence for *trnK-psbA* cotranscription in mustard chloroplasts

Jörg Nickelsen and Gerhard Link

University of Bochum, Plant Cell Physiology & Molecular Biology, Universitätsstr. 150, W-4630 Bochum, FRG

Received January 7, 1991 / March 19, 1991

Summary. In vitro transcripts from the 3′ flanking regions of mustard chloroplast genes were tested for protein binding in a chloroplast extract. Efficient and sequence-specific RNA-protein interaction was detected with transcripts of the genes *trnK*, *rps16* and *trnH*, but not with the 3′ terminal region of *trnQ* RNA. The *trans*-acting component required for specific complex formation is probably a single 54 kDa polypeptide. The protein-binding region of the *rps16* 3′ terminal region was mapped and compared with that of the *trnK* transcript determined previously. Both regions reveal a conserved 7-mer UUUAUCU followed by a stretch of U residues. Deletion of the *trnK* 3′ U cluster resulted in more than 80% reduction in the binding activity, and after deletion of both the U stretch and the 7-mer motif no binding at all was detectable. RNase protection experiments indicate that the protein-binding regions of both the *rps16* and *trnK* transcripts correlate with the positions of in vivo 3′ ends, suggesting an essential role for the 54 kDa binding protein in RNA 3′ end formation. In the case of the *trnK* gene, evidence was obtained for read-through transcripts that extend into the *psbA* coding region, thus pointing to the possibility of *trnK-psbA* cotranscription.

Key words: Chloroplast genes – RNA binding proteins – RNA 3′ end formation – *Sinapis alba* – UV crosslinking

Introduction

With the detailed information available on the nucleotide sequence and organization of chloroplast genes and their in vivo transcripts, it has become possible to study the mechanisms involved in the transcriptional and post-transcriptional regulation of plastid gene expression (Weil 1987; Sugiura 1989; Gruissem 1989). One of the steps thought to play an essential role is the formation of transcript 3′ ends. The flanking regions downstream of many, but not all (Ruf and Kössel 1988; Neuhaus et al. 1989), chloroplast genes reveal palindromic sequences that are capable of forming stem-loop structures. Available evidence suggests that at least some of these 3′ sequence elements do not act as transcriptional terminators (Stern and Gruissem 1987; Chen and Orozco 1988; Thomas et al. 1988). Instead they seem to function as processing signals and thereby protect upstream RNA sequences (Stern and Gruissem 1987, 1989; Adams and Stern 1990) comparable to prokaryotic REP (repetitive extragenic palindromic) sequences (Higgins et al. 1988). It was shown more recently that chloroplast proteins interact in vitro with RNA from the 3′ flanking regions of the spinach *psbA*, *rbcL* and *petD* genes (Stern et al. 1989) and the mustard *trnK* gene (Nickelsen and Link 1989), suggesting a role for these proteins in RNA processing and/or transcription termination in vivo.

In the case of the mustard *trnK* gene the protein-binding RNA region was mapped. It was found to be located 70 nucleotides (nt) downstream of the *trnK* 3′ exon in a region spanning 40 nt that has no apparent secondary structure (Nickelsen and Link 1989). We now present work aimed at defining the essential features of the binding region and identifing the protein(s) that interact with it. To clarify whether the interaction of this region with chloroplast proteins is a more general mechanism, we tested several other 3′ transcripts for protein binding. These RNAs were from the adjacent mustard genes *trnH* (Nickelsen and Link 1990), *trnQ* (Neuhaus 1989), and *rps16* (Neuhaus et al. 1989), which code for $tRNA^{His}$, $tRNA^{Gln}$ and the ribosomal protein S16, respectively. To obtain more detailed information on the *cis*-acting elements involved in chloroplast RNA-protein complex formation, we analysed deletion mutants in the *trnK* binding region. Finally, RNase mapping experiments were carried out to investigate whether the positions of the in vitro binding regions correlate with those of transcript 3′ ends in vivo.

Offprint requests to: G. Link

Materials and methods

Plasmid DNA. Plasmids pSPTH80/40 and pSPTH120 have been described previously (Nickelsen and Link 1989). Plasmid pSPTES358 contains the last 48 bp of the *trnH* coding region and 310 bp of the 3′ flanking region (Nickelsen and Link 1990). It was constructed by inserting the 358 bp *Eco*RI – *Sal*I fragment of plasmid pSA452 (Link 1984) into pSPT18 (Pharmacia LKB). Plasmid pSPTT317 has the last 19 bp of the *trnQ* gene and 298 bp of the 3′ flanking region and consists of a 317 bp *Taq*I fragment inserted into the *Acc*I site of pSPT19 (Neuhaus et al. 1989; Neuhaus 1989). Plasmid pSPTS398 contains the last 174 bp of the *rps16* 3′ exon and 224 bp of the 3′ flanking region and was constructed by cloning a 398 bp *Sau*3A fragment (Neuhaus et al. 1989) into the *Bam*HI site of pSPT19. Plasmid pSPTS158 has 13 bp of the *trnK* 3′ exon and 145 bp of the 3′ flanking region extending down to the *psbA* promoter. It was constructed by inserting the 175 bp *Sma*I fragment of plasmid pSA05/A1 (Link and Langridge 1984) into pSPT18.

Production of 3′ deletion mutants was according to Henikoff (1987). In brief, pSPTH80/40 was digested with *Pst*I and *Bam*HI, 10 µg of this DNA was treated with 600 units exonuclease III in a volume of 60 µl at 25° C, and 2.5 µl aliquots removed every 30 s. Each sample was then treated with 1.75 units nuclease S1 in a volume of 10 µl at 25° C for 30 min. After fill-in with Klenow enzyme, DNA was religated and transformed. Clones were characterized by restriction and sequence analyses.

RNA and extract preparations. In vitro RNA transcribed from linearized plasmid DNA as well as plastid RNA-binding extracts were prepared as described (Nickelsen and Link 1989).

Gel retardation assays. RNA-protein binding assays with the short transcript TH80/40 were carried out as described (Nickelsen and Link 1989). Binding mixtures that contained labelled in vitro transcripts longer than 200 nt (TT317, TS398, and TES358) were subsequently treated with 20 units RNase T_1. Samples were electrophoresed in 5% non-denaturing polyacrylamide gels.

UV crosslinking. UV crosslinking of RNA with chloroplast proteins was as described (Nickelsen and Link 1989), except that irradiated samples were treated for 10 min with RNase A (5 ng/µl) instead of RNase T_1. Bound proteins were analysed in 10% polyacrylamide-SDS gels (Laemmli 1970).

RNase protection assays. RNase protection mapping of in vivo transcripts was done with 50 µg of mustard plastid RNA and 1×10^6 cpm of ^{32}P-labelled in vitro transcripts as described (Sambrook et al. 1989).

RNase T_1 mapping of RNA-protein complexes. The protein-binding region of the in vitro transcript derived from the *rps16* 3′ region was mapped as described previously (Leibold and Munro 1988; Nickelsen and Link 1989) using labelling with [^{32}P]GTP, followed by binding, digestion with RNase T_1, and electrophoresis on denaturing 20% polyacrylamide gels.

Results

Chloroplast proteins interact with 3′ RNA sequences of the trnK, trnH *and* rps16 *genes, but not with those flanking the* trnQ *gene*

Figure 1A shows the mustard chloroplast *trnQ-trnH* region as well as the four different 3′ transcripts that were used as probes in RNA-protein binding experiments, i.e. TH80/40 (*trnK* gene), TES358 (*trnH* gene), TS398 (*rps16* gene), and TT317 (*trnQ* gene). When these ^{32}P-labelled RNAs were tested in gel retardation assays (Fried and Crothers 1981; Garner and Revzin 1981), RNase T_1-resistant RNA-protein complexes of approximately the same electrophoretic mobility were detected with TH80/40 (Fig. 1B, lane 2), TES358 (lane 4), and TS398 (lane 8). In contrast, when the *trnQ* 3′ transcript TT317 was used as probe, no complex formation was observed (Fig. 1B, lane 6).

UV crosslinking (Greenberg 1979; Wilusz and Shenk 1988) of proteins bound to 3′ RNA sequences revealed radioactively labelled polypeptides of 54 and 32 kDa in the case of TH80/40 (Fig. 1C, lane 2), TES358 (lane 4), and TS398 (lane 8). Only the 32 kDa signal was ob-

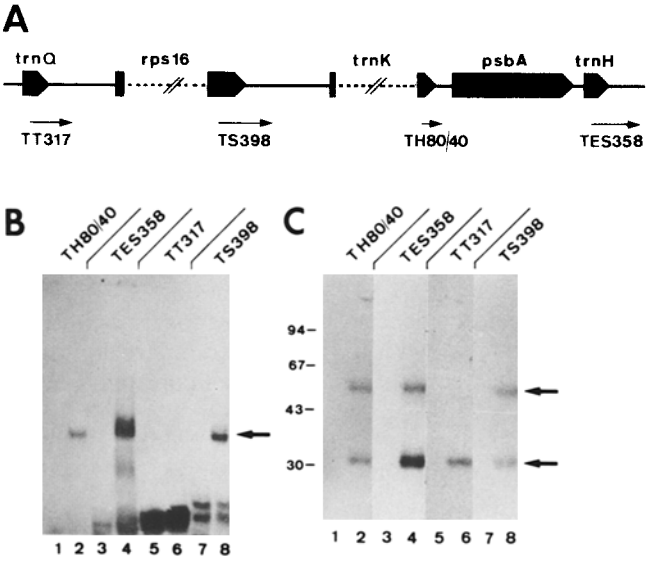

Fig. 1A–C. RNA-protein interactions in 3′ regions of chloroplast genes. **A** Schematic representation of the *trnQ-trnH* region of mustard chloroplast DNA. Coding regions are shown by *filled boxes* with a *tip* pointing into the 3′ flanking region. The in vitro transcripts used for RNA-protein binding are shown by *arrows*. **B** RNA gel retardation assays (including RNase T_1 treatment) with the ^{32}P-labelled RNA probes shown in **A**. Lanes 1, 3, 5, 7, RNA alone. Lanes 2, 4, 6, 8, RNA incubated with chloroplast extract (30 µg protein). *Arrow*, RNA-protein complexes. **C** Proteins labelled by bound RNA sequences after UV crosslinking, RNase A digestion, and separation on a 10% polyacrylamide/SDS gel. Lanes as in **B**. *Arrows*, RNA binding proteins at 54 and 32 kDa. Molecular sizes (kDa) of marker proteins are given in the left margin

served when *trnQ* 3′ RNA (TT317) was used (Fig. 1C, lane 6). These results suggest that the 54 kDa protein might be responsible for the sequence-specific RNA-protein complex formation seen in the gel retardation experiments with the transcripts from the *trnK*, *trnH*, and *rps16* 3′ regions (Fig. 1B). On the other hand, since the 32 kDa protein was visible in each UV crosslinking experiment (Fig. 1C) regardless of the probe used, its interaction with RNA appears to be sequence independent. Previous studies with TH80/40, using RNase T_1 treatment of crosslinked complexes prior to gel electrophoresis, had revealed labelled proteins of 62 and 58 kDa in size (Nickelsen and Link 1989). The present treatment with RNase A instead of T_1 was found to remove more efficiently those portions of the RNA that were not crosslinked to protein and hence is likely to give more accurate sizes for the labelled proteins.

To substantiate further the idea that the same 54 kDa protein binds to TH80/40, TES358 and TS398, competition experiments with ^{32}P-labelled TH80/40 and a 200-fold excess of unlabelled competitor RNA were carried out. As shown by gel retardation analysis (Fig. 2A, lanes 3, 4 and 6), the presence of either unlabelled TH80/40 or TES358 resulted in disappearance of the binding signal and TS398 led to decreased amounts of labelled complex. In contrast, transcript TT317 did not act as a competitor (Fig. 2A, lane 5).

The same competition strategy with labelled TH80/40 was also used in UV crosslinking experiments (Fig. 2B). Binding in the presence of excess unlabelled TH80/40 led to complete disappearance of the 54 kDa signal, whereas that at 32 kDa was still visible (Fig. 2B, lane 3). The latter signal was not competed by any of the heterologous RNAs TES358 (lane 4), TT317 (lane 5) and TS398 (lane 6), again indicating that this smaller

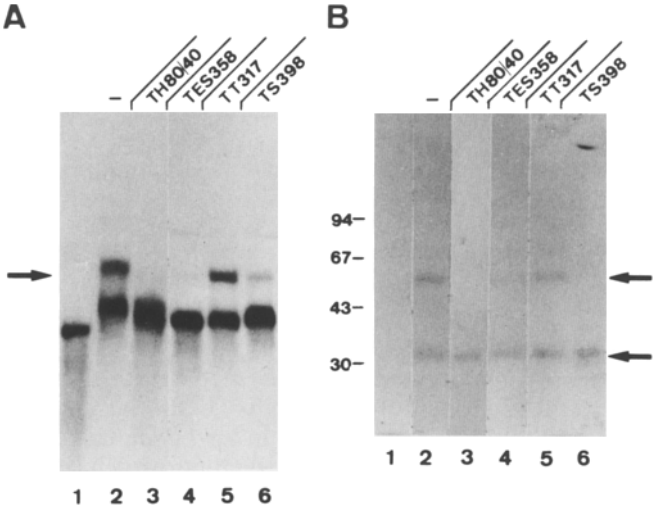

Fig. 2A and B. Competition binding experiments with various 3′ RNA segments (see Fig. 1A). Labelled *trnK* 3′ RNA (TH80/40) was incubated with 300 ng of the unlabelled competitor RNAs indicated at the top. Visualization of TH80/40-protein complexes (*arrows*) by gel retardation (without RNase T_1 treatment) (**A**) and UV crosslinking assays (**B**). Controls include binding reactions without proteins (lanes 1) or with proteins but lacking competitor RNA (lanes 2).

polypeptide does not interact with RNA in a sequence-specific way. Unlabelled TS398 from the *rps16* 3′ region completely abolished the labelled 54 kDa signal (Fig. 2B, lane 6), whereas TES358 was found only to decrease its intensity. This differs from the stronger competition effect of the *trnH* 3′ RNA seen in the gel retardation experiments (Fig. 2A, lanes 4 and 6), but could be explained by the possible existence of more than one binding site within the *trnH* 3′ region (see Fig. 7), each having a lower affinity than the site in the *rps16* 3′ transcript. TT317 again showed no competition effect at all (Fig. 2B, lane 5). Taken together, these results suggest that the 54 kDa protein interacts strongly with the 3′ in vitro transcripts of the mustard *trnK* and *rps16* genes, more weakly with the *trnH* 3′ RNA, and not at all with the *trnQ* 3′ RNA region.

RNase T_1 mapping of the protein binding region within rps16 *3′ RNA*

The *trnK* 3′ RNA region involved in RNA-protein interaction was previously located by RNase T_1 mapping (Nickelsen and Link 1989) as described by Leibold and Munro (1988). To compare the binding sites, we investigated the *rps16* 3′ RNA using the same mapping technique (Fig. 3). Following incubation of TS398 with chloroplast proteins and subsequent RNase T_1 treatment, the RNA of the T_1-resistant complex (T_1R_{rps16}) was reisolated and analysed on a 20% polyacrylamide gel. The predominant RNA species generated is 38 nt in size (Fig. 3A, lane 2). Further T_1 digestion of this RNA did not lead to smaller fragments, indicating the absence of internal G residues within this region (Fig. 3A, lane 5). The *rps16* 3′ sequence in Fig. 3B shows that only one 38 nt fragment can be expected after G-specific cleavage, indicating that the region 44 nt downstream of the *rps16* 3′ exon represents the T_1-protected segment of TS398. The in vivo 3′ end of the *rps16* transcript was previously located within this region by S1 nuclease mapping (Neuhaus et al. 1989).

In experiments similar to that shown in Fig. 3 we analysed the RNA of the T_1-resistant complex formed with the *trnH* 3′ transcript TES358. In contrast to TS398, no predominant RNA species, but multiple weak signals ranging from 15 to 3 nt in size were detected (data not shown). This is in agreement with the weak binding in UV crosslinking experiments (Fig. 2B) and indicates a more complex interaction that might involve multiple protein-binding sites (see Fig. 7).

Deletion analysis of the protein binding region within TH80/40

Sequence comparison of the T_1-protected RNA region from T_1R_{trnK} with that from T_1R_{rps16} shows many conserved nucleotide positions (Fig. 4A). Both sequences contain the 7-mer UUUAUCU as well as a stretch of U residues, which in the case of the *rps16* transcript is located immediately upstream of the in vivo 3′ end.

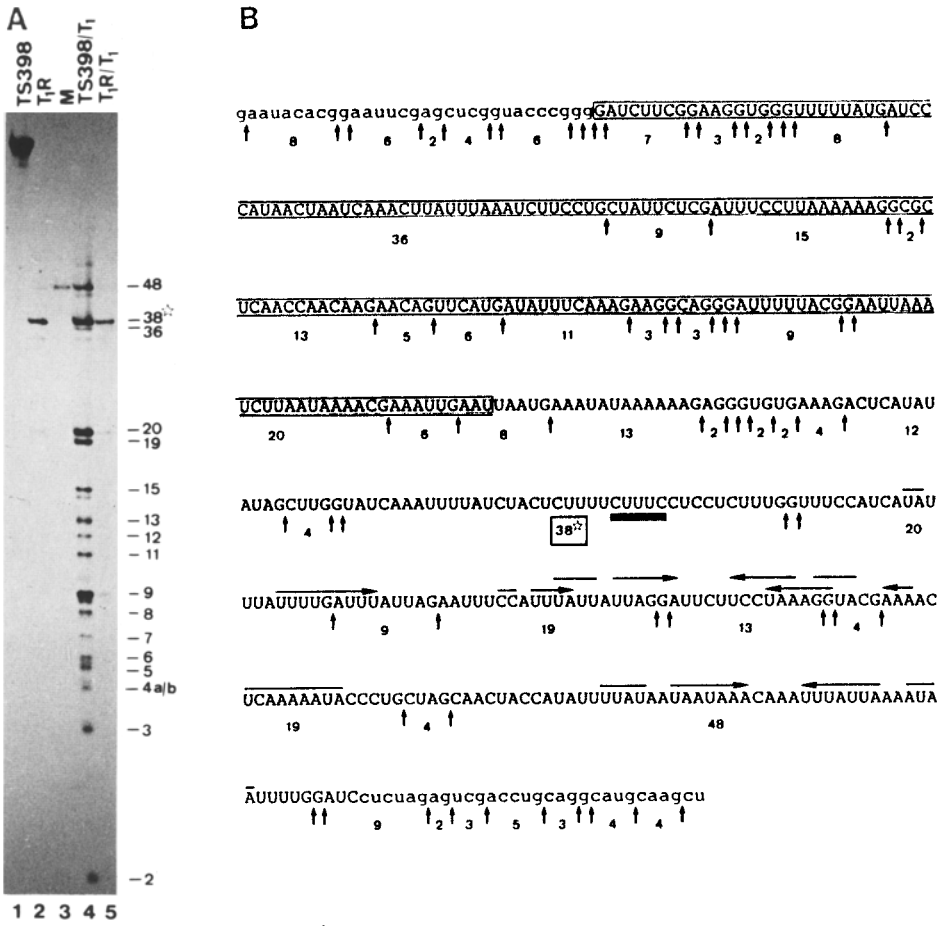

Fig. 3A and B. RNase T_1 mapping of the TS398-protein complex. **A** Separation of T_1-resistant RNA fragments on a 20% denaturing polyacrylamide gel. Lane 1, intact TS398; lane 2, isolated T_1R_{rps16} RNA; lane 3, 48 nucleotide (nt) marker transcript from the pSPT18/19 polylinker; lane 4, RNase T_1 digestion products of TS398; lane 5, RNase T_1 digestion product of T_1R_{rps16} RNA. Fragment lengths (nt) are given in the right margin. *Asterisk*, 38 nt fragment of the T_1R_{rps16} complex. **B** Nucleotide sequence of TS398, with the bases of the polylinker portion indicated by *lower case* letters. The rps16 3′ exon is *boxed* and potential stem-loop forming regions are *overlined*. The mapped 3′ end of the *rps16* in vivo transcript (Neuhaus et al. 1989) is *underlined* by the *heavy bar*. Expected RNase T_1 cutting sites at G residues are marked by *arrows*. The *numbers below* indicate sizes (nt) of the resulting oligonucleotide fragments. *Asterisk*, fragment of the T_1R_{rps16} complex

An additional conserved 7-mer motif,UAUAUAG, is present in both sequences immediately 5′ upstream of the T_1R region. The cluster of five U residues within TH80/40 (Fig. 4A) was previously suggested to be essential for the interaction of TH80/40 with chloroplast protein(s), since the binding activity with this RNA is strongly competed by poly(U) (Nickelsen and Link 1989).

To test the possible significance of regions within T_1R_{trnK} RNA for protein binding 3′ deletion mutants of TH80/40 were constructed (Fig. 4B) and their protein-binding activity was tested. Deletion of the 18nt portion including the U cluster at the 3′ side of T_1R_{trnK} (Δ21) resulted in reduced binding activity in gel retardation experiments (Fig. 4C, lane 4) down to less than 20% of that seen with wild-type TH80/40 (Fig. 4C, lane 2) and also led to decreased labelling of both the 54 and 32 kDa protein after UV crosslinking (Fig. 4D, lane 4). No binding activity at all was observed in gel retardation experiments (Fig. 4C, lanes 6 and 8) with two other mutants, one lacking 30 nt of the 3′ terminal portion of T_1R_{trnK} RNA including the conserved UUUAUCU element (Δ33), and the other completely lacking the entire T_1R_{trnK} RNA sequence (Δ47). Following UV crosslinking, the 54 kDa protein was undetectable, whereas there was no further decrease in the intensity of the signal at 32 kDa as compared with Δ21 (Fig. 4C, D, lanes 6 and 8). In competition experiments (Fig. 5A, B), none of the three mutant RNAs was found to interfere significantly with the protein binding of TH80/40. These results together indicate that the region previously mapped by RNase T_1 protection (T_1R_{trnK}) indeed corresponds to the protein-binding region, and that the 18 nt portion at the 3′ side of this region contains a necessary signal (probably the track of five U residues) for efficient RNA-protein interaction. The further decrease in binding activity of the mutant RNAs Δ33 and Δ47 as compared with Δ21 suggests that sequences within the 5′ portion of the T_1R_{trnK} region are also involved in RNA-protein interaction.

The T_1R_{trnK} RNA region is expressed in vivo

The protein-binding region of the *rps16* in vitro 3′ transcript (Fig. 3) roughly corresponds to the position of the 3′ end of the precursor RNA in vivo (Neuhaus et al. 1989), indicating that the RNA binding protein(s) detected by UV crosslinking might be involved in 3′ end formation. In the case of the *trnK* 3′ region, however, no evidence for such a correlation between the in vitro and in vivo data had previously been obtained. In vivo transcript mapping of the *trnK* region (Neuhaus and Link 1987), showed a single 2.8 kb precursor with a 3′ end approximately 70 nt upstream of the in vitro binding region. Using a sensitive RNase protection assay (Sam-

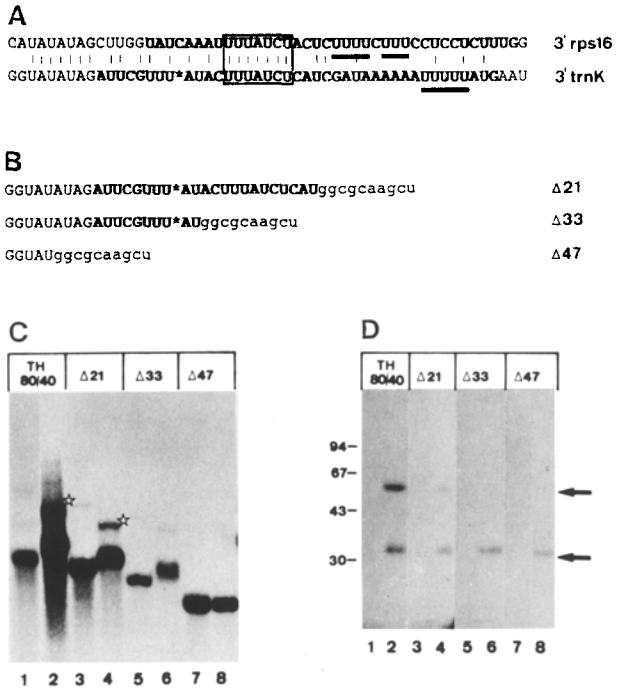

Fig. 4A–D. Analysis of *trnK* 3′ RNA deletion mutants. **A** RNA sequence comparison of T_1R_{rps16} and T_1R_{trnK}. A single nucleotide gap (*asterisk*) was inserted into the 3′ *trnK* sequence to allow for optimal alignment. Conserved nucleotides are marked by *vertical lines*, the UUUAUCU 7-mer is *boxed*, and U stretches are *underlined*. Bases within the RNase T_1-resistant binding regions are given in *bold letters*. **B** Sequences of 3′ RNA deletion mutants of the *trnK* 3′ region (*Δ*21, *Δ*33, *Δ*47). *Lowercase letters* indicate polylinker sequences. **C** Gel retardation assays (without RNase T_1 treatment) with either ^{32}P-labelled full-size TH80/40 or 3′ deleted RNAs. Lanes 1, 3, 5, 7, RNA alone. Lanes 2, 4, 6, 8, RNA incubated with chloroplast extract (30 µg protein). *Asterisks*, RNA-protein complexes. **D** UV crosslinking of chloroplast proteins to TH80/40 and 3′ deleted RNAs. Lanes as in **C**. *Arrows*, RNA binding proteins of 54 and 32 kDa

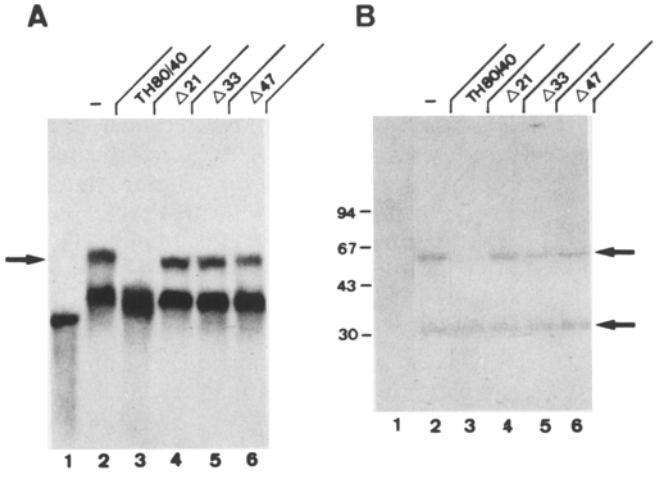

Fig. 5A and B. Competition binding experiments with *trnK* 3′ deleted RNAs. ^{32}P-labelled TH80/40 was incubated with 300 ng of the unlabelled competitor RNAs indicated at the top (see Fig. 4B). Visualization of TH80/40-protein complexes (*arrows*) by gel retardation (without RNase T_1 treatment (**A**) and UV crosslinking assays (**B**). Lanes 1, no protein; lanes 2, no competitor RNA

brook et al. 1989), we tested whether there might be minor *trnK* transcripts that extend into the T_1R_{trnK} region. When the ^{32}P-labelled transcript cTS158 (Fig. 6B) was hybridized to chloroplast RNA and subsequently treated with RNases, two resistant fragments of 158 and 95 nt were detected (Fig. 6A, lane 3), which were not observed after hybridization of cTS158 to *Escherichia coli* tRNA (Fig. 6A, lane 2). These data provide evidence that in vivo RNA sequences do indeed exist downstream of the previously mapped 2.8 kb *trnK* precursor (Neuhaus and Link 1987), which ends shortly after the 3′ exon and is not detected by the probe used here. The 95 nt fragment would correspond to a minor *trnK* transcript ending at approximately the position of the conserved 7-mer sequence within T_1R_{trnK}. A 158 nt RNA fragment would be expected if the entire mustard se-

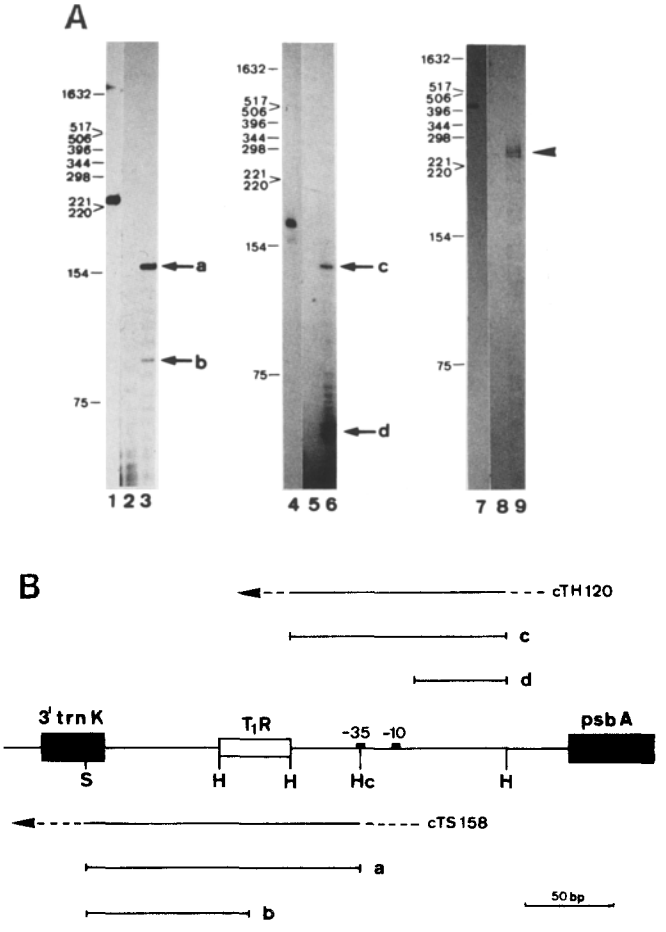

Fig. 6A and B. RNase protection mapping of in vivo transcripts spanning the *trnK* and *rps16* 3′ regions. **A** Following RNA-RNA hybridization and treatment with RNase, resistant products were separated on a 6% denaturating polyacrylamide gel. Probes, cTS158 (lanes 1–3), cTH120 (lanes 4–6), cTS398 (lanes 7–9). Lanes 1, 4, 7, RNA alone without RNase treatment. Lanes 2, 5, 8, RNA probes hybridized to 50 µg *Escherichia coli* tRNA and subsequently treated with RNase. Lanes 3, 6, 9, RNase-resistant products after hybridization to 50 µg chloroplast RNA. **B** Scheme of the *trnK-psbA* intergenic region. Coding regions are shown as *filled boxes*. The *open box* marks the T_1R_{trnK} sequence. Also indicated is the *psbA* promoter (−35/−10). *Arrows*, ^{32}P-labelled in vitro RNA probes with *broken lines* representing polylinker sequences. *Lines below*, protected RNA fragments. H, *Hin*fI; Hc, *Hin*cII; S, *Sma*I

quence within cTS158 were protected by chloroplast RNA and only linker sequences (79 nt) were digested by the RNases. This suggests that in vivo at least a fraction of the *trnK* transcripts extends further downstram into the *psbA* promoter region.

To analyse whether the region transcribed in vivo proceeds even further into the *psbA* gene, RNA probe cTH120 (Fig. 6B) was used in a similar RNase protection experiment. Electrophoresis of the nuclease-resistant products revealed signals at 120 and 50 nt (Fig. 6A, lane 6). The latter is likely to reflect the transcript initiated at the *psbA* promoter in vivo, whereas the larger fragment is of the size expected if there were read-through transcripts that cover the entire chloroplast RNA-specific portion of the probe. Based on the intensity of the two bands in Fig. 6A, lane 6, the fraction of read-through transcripts is much smaller than that of transcripts initiated at the *psbA* promoter. Likewise, the in vivo concentration of transcripts of the *trnK-psbA* intergenic region is at least one order of magnitude lower than that of total *trnK* precursor transcripts as assessed by the intron-specific RNA probe cTBX140 (Nickelsen and Link 1989) (data not shown).

To test if read-through also occurs at the *rps16* gene, we used probe cTS398, which covers the region containing the in vivo 3′ end of the *rps16* transcript (Fig. 3B). A single 241 nt nuclease-resistant product was generated (Fig. 6A, lane 9), which matches the position of the in vivo 3′ end, indicating that no significant read-through transcripts are present.

Discussion

As a step towards defining the mechanisms of transcript 3′ end formation in chloroplasts, we investigated the in vitro protein binding activity of transcripts from 3′ flanking regions of plastid genes, and tried to correlate the binding sites with the positions of the mapped 3′ ends of the corresponding in vivo transcripts. The results of our binding assays suggest a close relationship between sites of RNA-protein complex formation and authentic 3′ ends for the *trnK* and *rps16* transcripts, but not for the *trnH* and *trnQ* 3′ RNAs. The sequence-specific RNA-protein interaction evident in gel retardation experiments is paralleled by the radioactive labelling of a 54 kDa protein in UV crosslinking experiments. Binding assays using transcripts of *trnK* 3′ deletion mutants indicate that the 54 kDa protein indeed binds to the region (T_1R_{trnK}) that was previously shown to interact with plastid proteins (Nickelsen and Link 1989). The 32 kDa protein that also becomes labelled in UV crosslinking experiments appears to bind in a sequence-independent way, reminiscent of the 28–33 kDa single-stranded DNA binding proteins that have recently been purified from tobacco chloroplasts (Li and Sugiura 1990).

Sequence comparison of T_1R_{trnK} RNA with the putative binding region of the *rps16* transcript (T_1R_{rps16} RNA) reveals a high degree of nucleotide sequence conservation. Apart from the overall high content of U resi-

```
3' trnK

TTTATCT.............TTTTT          S.a.    (Neuhaus and Link 1987)
TTTCTGT.............TTTTT          N.t.    (Sugita et al. 1985)
TTTCACT.............TTTTTT         O.s.    (Hiratsuka et al. 1989)
TTTATTT.............TTTTTT         M.p.    (Ohyama et al. 1986)


3' rps16

TTTATCT....TTTTCTTT               S.a.    (Neuhaus et al. 1989)
TTTCTCT....TTTTTTT                N.t.(a)
TTTCTAT.....TTTTTT                N.t.(b) (Shinozaki et al. 1986)
TCTATCT..TTTTTTTCTT               O.s.    (Hiratsuka et al. 1989)


3' trnH

TTTCTCT.TTTTT                     S.a.(a)
TTTCTCT - 40bp - TTTTTTCCTTTTTT   S.a.(b) (Nickelsen and Link 1990)


3' trnQ

TTTCTCCT.....................TTTTTT  S.a.  (Neuhaus et al. 1989,
                                            Neuhaus and Link 1990)
```

Fig. 7. Comparison of plastid 3′ flanking regions. From top to bottom: DNA sequence elements within the 3′ flanking regions of *trnK* and *rps16* genes from different species that appear equivalent to the 7-mer motif and U cluster of $T_1R_{trnK/rps16}$ from mustard; sequence elements within the 3′ flanking regions of the mustard *trnH* and *trnQ* genes, resembling the conserved motifs. S.a. *Sinapis alba*, N.t. *Nicotiana tabacum*, O.s. *Oryza sativa*, M.p. *Marchantia polymorpha*

dues (47.5% for T_1R_{trnK} and 54.4% for T_1R_{rps16} RNA), a notable feature of both regions is the conserved 7-mer UUUAUCU, which is followed by a stretch of several U residues, whereas conserved secondary structure elements appear to be absent. A search for similar conserved sequences downstream of the *trnK* and *rps16* genes from tobacco, rice and liverwort (the latter for *trnK* only) further substantiates the notion that these motifs might be functional *cis*-acting elements (Fig. 7). The 3′ regions of the mustard *trnH* and *trnQ* genes also reveal related, but variant, motifs. This might explain the different binding efficiencies observed with the latter RNAs (Figs. 1 and 2). It has recently been reported (Stern et al. 1989) that a 55 kDa spinach chloroplast protein is capable of binding to the unprocessed in vitro transcript of the *petD* 3′ region. This protein did not bind to RNA processed in vitro and therefore was implicated in the processing reaction itself. Interestingly, the sequence UUUCAUCU followed by a cluster of 6 U residues is present within the *petD* 3′ region immediately downstream of a conserved stem-loop structure that acts as a transcript stabilizing element.

The in vitro binding regions T_1R_{trnK} and T_1R_{rps16} both match the positions of 3′ ends mapped in vivo. There are differences, however, with regard to the abundance of the corresponding transcripts. Whereas T_1R_{rps16} marks the 3′ end of the single *rps16* precursor transcript that was previously defined by nuclease S1 analyses (Neuhaus et al. 1989), T_1R_{trnK} does not correspond to the 3′ end of the major 2.8 kb *trnK* precursor, which was shown to end shortly after the *trnK* 3′ exon (Neuhaus and Link 1987; Boyer and Mullet 1986). We were able to detect the minor transcript 3′ end within T_1R_{trnK} only by RNase protection mapping, which is more sensitive than the S1 assay by at least one order of magnitude (Sambrook et al. 1989). Although ineffi-

cient transcription of this region cannot be excluded, the low concentration of *trnK* 3′ RNA sequences is more likely to be the result of rapid processing. The degradation of 3′ sequences after endonucleolytic cleavage at the 3′ exon appears to be a general mechanism in the precursor processing of chloroplast tRNA genes (Yamaguchi-Shinozaki et al. 1987; Wang et al. 1988; Marion-Poll et al. 1988).

An unexpected result of the RNase protection assays in our present work is that a fraction of *trnK* transcripts appears to extend further downstream beyond the T_1R_{trnK} region into the *psbA* coding region, suggesting cotranscription of the two genes. We were not able to detect a band of the size expected for the dicistronic transcript (at least 4.3 kb) in Northern experiments using cTS158 as hybridization probe (data not shown). Obviously, however, this failure could have been due to low transfer efficiency and/or instability and does not preclude the existence of such a transcript.

An interesting consequence of the read-through transcripts detected by the RNase protection assay is that at least two *psbA* transcripts with different 5′ leader regions exist in vivo. Multiple RNA 5′ ends have been reported for a number of chloroplast genes (Crossland et al. 1984; Poulsen 1984; Mullet et al. 1985; Hanley-Bowdoin et al. 1985; Rock et al. 1987; Tanaka et al. 1987; Westhoff and Herrmann 1988; Neuhaus et al. 1989; Haley and Bogorad 1990). Evidence has been presented that the 5′ region of chloroplast mRNA contains the target site for a translational activator (Rochaix et al. 1989), and it has been suggested that different 5′ termini might be involved in the regulation of plastid translation. Indeed with regard to the *psbA* gene, much of the control of its expression appears to occur at the translational level (Fromm et al. 1985; Klein and Mullet 1986, 1987; Deng and Gruissem 1987; Gamble and Mullet 1989).

Recently, Callahan et al. (1990) reported the existence of a novel form of the *psbA* gene product. This minor form of the 32 kDa D1 protein, called 32*, has a slightly different mobility on SDS-polyacrylamide gels and different in vivo turnover rates. The authors suggested that 32* might originate from conversion of one form of the protein into the other. However, in view of our RNase protection data, another possible explanation would be that the two versions of the D1 protein might be the result of different transcription and RNA maturation pathways and hence each may represent a distinct translation product. It will be interesting to test this possibility by using appropriate in vitro systems.

Acknowledgements. We wish to thank H. Gehrke and C. Wittig for their expert technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

References

Adams CC, Stern DB (1990) Control of mRNA stability in chloroplasts by 3′ inverted repeats: effects of stem and loop mutations on degradation of *psbA* mRNA *in vitro*. Nucleic Acids Res 18:6003–6010

Boyer SK, Mullet JE (1986) Characterization of *P. sativum* chloroplast *psbA* transcripts produced *in vivo*, *in vitro* and in *E. coli*. Plant Mol Biol 6:229–244

Callahan FE, Ghirardi ML, Sopory SK, Mehta AM, Edelman M, Mattoo AK (1990) A novel metabolic from of the 32k Da-D1 protein in the grana-localized reaction center of photosystem II. J Biol Chem 256:15357–15360

Chen LJ, Orozco EM (1988) Recognition of prokaryotic terminators by spinach chloroplast RNA polymerase. Nucleic Acids Res 16:8411–8431

Crossland LD, Rodermel SR, Bogorad L (1984) Single gene for the large subunit of ribulosebisphosphate carboxylase in maize yields two differentially regulated mRNAs. Proc Natl Acad Sci USA 81:4060–4064

Deng XW, Gruissem W (1987) Control of plastid gene expression during development: The limited role of transcriptional regulation. Cell 49:379–387

Fried M, Crothers DM (1981) Equilibria and kinetics of lac repressor-operator interaction by polyacrylamide gel electrophoresis. Nucleic Acids Res 9:6505–6525

Fromm H, Devic M, Fluhr R, Edelman M (1985) Control of *psbA* gene expression: In mature chloroplasts, light regulation of 32 kDa protein synthesis is independent of transcript level. EMBO J 4:291–295

Gamble PE, Mullet JE (1989) Translation and stability of proteins encoded by the plastid *psbA* and *psbB* genes are regulated by a nuclear gene during light-induced chloroplast development in barley. J Biol Chem 264:7236–7243

Garner MM, Revzin A (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. Nucleic Acids Res 9:3047–3060

Greenberg JR (1979) Ultraviolet light-induced crosslinking of mRNA to proteins. Nucleic Acids Res 6:715–732

Gruissem W (1989) Chloroplast gene expression: how plants turn their plastids on. Cell 56:161–170

Haley J, Bogorad L (1990) Alternative promoters are used for genes within maize chloroplast polycistronic transcription units. Plant Cell 2:323–333

Hanley-Bowdoin L, Orozco EM, Chua NH (1985) Transcription of chloroplast genes by homologous and heterologous RNA polymerases. In: Steinback KE, Bonitz S, Arntzen CJ, Bogorad L (eds) Molecular biology of the photosynthetic apparatus. Cold Spring Harbor Laboratory Press, New York, pp 311–318

Henikoff S (1987) Unidirectional digestion with exonuclease III in DNA sequence analysis. Methods Enzymol 155:156–165

Higgins CF, McLaren RS, Newbury SF (1988) Repetitive extragenic polindromic sequences, mRNA stability and gene expression: evolution by gene conversion? Gene 72:3–14

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun C-R, Meng B-Y, Li Y-Q, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid inversion during the evolution of cereals. Mol Gen Genet 217:185–194

Klein RR, Mullet JE (1986) Regulation of chloroplast-encoded chlorophyll-binding protein translation during higher plant chloroplast biogenesis. J Biol Chem 261:11138–11145

Klein RR, Mullet JE (1987) Control of gene expression during higher plant chloroplast biogenesis. J Biol Chem 9:4341–4348

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T_4. Nature 227:680–685

Leibold EA, Munro HN (1988) Cytoplasmic protein binds *in vitro* to a highly conserved sequence in the 5′ untranslated region of ferritin heavy- and light-subunit mRNAs. Proc Natl Acad Sci USA 85:2171–2175

Li Y, Sugiura M (1990) Three distinct ribonucleoproteins from tobacco chloroplasts: each contains a unique amino terminal acidic domain and two ribonucleoprotein consensus motifs. EMBO J 9:3059–3066

Link G (1984) Hybridization study of developmental plastid gene expression in mustard (*Sinapis alba* L.) with cloned probes for most plastid DNA regions. Plant Mol Biol 3:243–248

Link G, Langridge U (1984) Structure of the chloroplast gene for the precursor of the M_r 32000 photosystem II protein from mustard (*Sinapis alba*). Nucleic Acids Res 12:945–957

Marion-Poll A, Hibbert CS, Radebaugh CA, Hallick RB (1988) Processing of mono-, di- and tricistronic transfer RNAs precursors in a spinach or pea chloroplast soluble extract. Plant Mol Biol 11:45–56

Mullet JE, Orozco EM, Chua N-H (1985) Multiple transcripts for higher plant *rbcL* and *atpB* genes and localization of the transcription initiation site of the *rbcL* gene. Plant Mol Biol 4:39–54

Neuhaus H (1989) Nucleotide sequence of the chloroplast genes for $tRNA^{Gln}$ and the 4 kD K polypeptide of photosystem II from mustard (*Sinapis alba*). Nucleic Acids Res 17:444

Neuhaus H, Link G (1987) The chloroplast $tRNA^{Lys}$(UUU) gene from mustard (*Sinapis alba*) contains a class II intron potentially coding for a maturase-related polypeptide. Curr Genet 11:251–257

Neuhaus H, Link G (1990) The chloroplast *psbK* operon from mustard (*Sinapis alba* L.): multiple transcripts during seedling development and evidence for divergent overlapping transcription. Curr Genet 18:377–383

Neuhaus H, Scholz A, Link G (1989) Structure and expression of a split chloroplast gene from mustard (*Sinapis alba*): ribosomal protein gene *rps16* reveals unusual transcriptional features and complex RNA maturation. Curr Genet 15:63–70

Nickelsen J, Link G (1989) Interaction of a 3′ RNA region of the mustard *trnK* gene with chloroplast proteins. Nucleic Acids Res 17:9637–9648

Nickelsen J, Link G (1990) Nucleotide sequence of the mustard chloroplast genes *trnH* and *rps19*′. Nucleic Acids Res 18:1051

Ohyama K, Fukuzawa H, Kohchi T, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S-I, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322:572–574

Poulsen C (1984) Two mRNA species differeing by 258 nucleotides at the 5′ end are formed from the barley chloroplast *rbcL* gene. Carlsberg Res Commun 48:57–80

Rochaix JD, Kuchka M, Mayfield S, Schirmer-Rahire M, Girard-Bascou J, Bennoun P (1989) Nuclear and chloroplast mutations affect the synthesis or stability of the chloroplast *psbC* gene product in *Chlamydomonas reinhardtii*. EMBO J 8:1013–1021

Rock CD, Barkan A, Taylor WC (1987) The maize plastid *psbB-psbF-petB-petD* gene cluster: spliced and unspliced *petB* and *petD* RNAs encode alternative products. Curr Genet 12:69–77

Ruf M, Kössel H (1988) Structure and expression of the gene coding for the alpha-subunit of DNA-dependent RNA polymerase from the chloroplast genome of *Zea mays*. Nucleic Acids Res 16:5741–5753

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning, a laboratory manual, second edition. Cold Spring Harbor Laboratory Press, New York

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng B-Y, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. EMBO J 5:2043–2049

Stern DB, Gruissem W (1987) Control of plastid gene expression: 3′ inverted repeats acts as mRNA processing and stabilizing elements, but do not terminate transcription. Cell 51:1145–1157

Stern DB, Gruissem W (1989) Chloroplast mRNA 3′ end maturation is biochemically distinct from prokaryotic mRNA processing. Plant Mol Biol 13:615–625

Stern DB, Jones H, Gruissem W (1989) Function of plastid mRNA 3′ inverted repeats. RNA stabilization and genespecific protein binding. J Biol Chem 264:18742–18750

Sugita M, Shinozaki K, Sugiura M (1985) Tobacco chloroplast tRNA Lys(UUU) genes contain a 2.5 kilobase pair intron: An open reading frame and a conserved boundary sequence in the intron. Proc Nat Acad Scu USA 82:3557–3561

Sugiura M (1989) The chloroplast genome. In: Marcus A (ed) The biochemistry of plants, vol 15. Academic Press, New York, pp 133–150

Tanaka M, Obokata J, Chunwongse J, Shinozaki K, Suguira M (1987) Rapid splicing and stepwise processing of a transcript from the *psbB* operon in tobacco chloroplasts: Determination of the intron sites in *petB* and *petD* . Mol Gen Genet 209:427–431

Thomas F, Zeng GQ, Mache R, Briat JF (1988) Transcription study of the genes encoded in the region of the junction between the large single copy and the inverted repeat A of spinach chlooroplast DNA. Plant Mol Biol 10:447–457

Wang MJ, Davis NW, Gegenheimer P (1988) Novel mechanisms for maturation of chloroplast transfer RNA precursors. EMBO J 7:1567–1574

Weil JH (1987) Organization and expression of the chloroplast genome. Plant Sci 49:149–157

Westhoff P, Herrmann RG (1988) Complex RNA maturation in chloroplasts. Eur J Biochem 171:551–564

Wilusz J, Shenk T (1988) A 64 kd nuclear protein binds to RNA segments that include the AAUAAA polyadenylation motif. Cell 52:221–228

Yamaguchi-Shinozaki K, Shinozaki K, Sugiura M (1987) Processing of precursor tRNAs in a chloroplast lysate. FEBS Lett 215:132–136

Communicated by H. Saedler