Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 2295–2299, March 1994
Genetics

Complete RNA editing of unspliced and dicistronic transcripts of the intron-containing reading frame IRF170 from maize chloroplasts

STEPHANIE RUF, PATRIC ZELTZ, AND HANS KÖSSEL*

Institut für Biologie III, Universität Freiburg, Schänzlestrasse 1, D-79104 Freiburg, Germany

Communicated by Lawrence Bogorad, December 14, 1993

ABSTRACT **The maize plastome harbors within the *rps4–rps14* gene cluster the reading frame IRF170, which is interrupted by two introns. Although the function of the encoded peptide of 170 amino acids is not known, the conservation of IRF170 homologs in other plastomes is a strong indication that IRF170 is a functional gene. Amplification and sequence analyses of IRF170 specific cDNAs reveals two C-to-U editing events occurring within each of the first two exons. This situation allows an analysis of the temporal order between editing and splicing of a chloroplast transcript. By using intron-specific primer combinations, cDNAs derived from partially or even unspliced IRF170 transcripts could be amplified which in all cases showed complete editing. Complete editing was also observed with a cDNA derived from a transcript in which the proximal *rps4* and the 5′ half of IRF170-encoded sequences were still linked. This demonstrates that editing of the IRF170 transcript is an early processing step preceding both splicing and cleavage to monocistronic mRNA.**

The primary transcripts encoded by chloroplast protein genes are known to undergo a series of processing steps such as splicing, cleavage to monocistronic mRNAs, and trimming of terminal sequences (1, 2). More recently, editing has been detected as an additional step of chloroplast mRNA maturation (3–7). This process was originally observed in mitochondrial transcripts of trypanosomes (8, 9) but was subsequently also found in nuclear-encoded transcripts of mammals (10) and in mitochondrial transcripts of plants (11–13) and of the acellular slime mold *Physarum* (14). Depending on the different types of editing processes, the genetic information transmitted to the primary transcripts of the respective genes can be altered by nucleotide insertions and deletions (8, 9, 14) or by base substitutions (10–13). The editing events observed in chloroplast transcripts so far all result in C-to-U transitions, thereby restoring codons for conserved amino acid residues of the respective peptides (3–7).

The reading frame designated IRF170 is subdivided by two introns and is well conserved within the *rps4–rps14* gene cluster of the plastomes of higher plants (ref. 15; see Fig. 1), whereas a homologous reading frame without introns could be identified in the cyanobacterium *Synechocystis* (16). Although the function of the encoded peptide, consisting in maize of 170 amino acids, is not known, the absence of an IRF170 homolog in the plastome of the nonphotosynthetic parasitic plant *Epifagus* (17) as well as differences in the processing of IRF170 transcripts between amyloplasts and chloroplasts from maize (18) appears to suggest a function as a component of the photosynthetic apparatus. This supposition is further supported by the codon usage of IRF170-encoded mRNAs (15) and by the existence of a cyanobacterial IRF170 homolog (16).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

The existence of a large primary transcript which includes sequences of the flanking genes of the *rps4–rps14* gene cluster and the presence of two introns in the IRF170-encoded primary transcript provide an ideal situation for the study of the temporal interrelationship among editing, splicing, and cleavage to monocistronic mRNA. Having established the occurrence of two editing events in the first two exons of the transcript, we show that these events are already complete in the partially spliced and even unspliced transcripts and in a dicistronic transcript containing *rps4* and the IRF170-encoded sequences. From this we conclude that editing of the IRF170-encoded chloroplast transcript is an early processing step which precedes both splicing and cleavage to monocistronic mRNA.

MATERIALS AND METHODS

Isolation of Chloroplasts. Intact chloroplasts were isolated from 10-day-old maize seedlings by Percoll gradient centrifugation (5).

Preparation of Nucleic Acids. Nucleic acids from purified chloroplasts were prepared by guanidinium hydrochloride extraction and CsCl gradient centrifugation (5).

Reverse Transcription of RNA and Amplification of DNA and cDNA. RNA was reverse transcribed with random hexanucleotide primers and avian myeloblastosis reverse transcriptase (5). Genomic DNAs and cDNAs were amplified in the presence of 1.5 mM $MgCl_2$ by a standard protocol with 32 cycles at 93°C (for 1 min), 55°C (for 1 min), and 72°C (for 1.5 min) with a 2-min extension at 93°C of the first cycle and a 5.5-min extension at 72°C of the last cycle. Amplification products were separated by electrophoresis in 1.5% agarose gel, visualized by ethidium bromide staining, and eluted onto a DEAE membrane.

Direct Sequencing of Amplification Products. Gel-purified amplification products were sequenced directly by a modified chain-termination method (19).

List of Oligonucleotides. Oligonucleotides used for PCR and/or sequencing were synthesized on a model 394 DNA synthesizer (Applied Biosystems). The positions and orientations of the primers P1–P7 are shown in Figs. 1 and 2. Their sequences were selected according to published sequences (20, 21): P1, 5′-CCATTTAGAGCATATGGATCG-3′; P2, 5′-AATTGAAGGTAGAAGAAAGGG-3′; P3, 5′-GCTTGTGTGTATAAGGCCTA-3′; P4, 5′-GAAGCCGTATGAGATGAAAAC-3′; P5, 5′-GCCCTCTTTTCTAGTATTCAC-3′; P6, 5′-CGCTGTATTCTGTAATGTAAA-3′; P7, 5′-CGAATTAAGAGCCTTAGGTCG-3′.

RESULTS AND DISCUSSION

The IRF170 Transcript Is Edited at Two Sites. As depicted in Fig. 1, the *rps4–rps14* gene cluster containing the IRF170

*To whom reprint requests should be addressed.

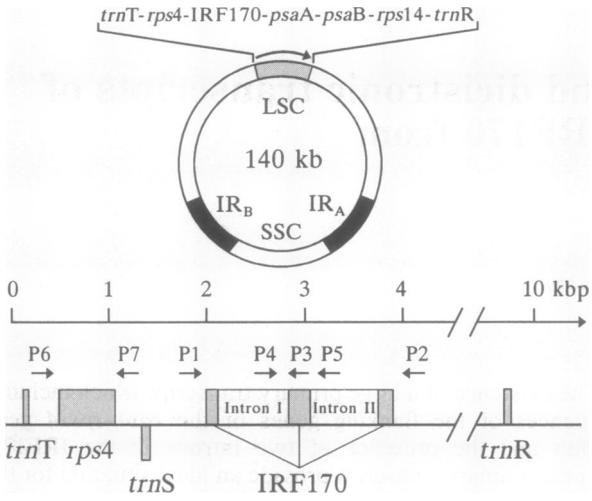

FIG. 1. Position of the intron-containing reading frame IRF170 on the maize plastome. (*Upper*) Location and orientation of the *rps4–rps14* cluster within the large single-copy region (LSC) and in relation to the inverted repeat regions (IR_A and IR_B) and to the small single-copy region (SSC). (*Lower*) Position of IRF170 in relation to other genes of the *rps4–rps14* cluster. With the exception of the *trnS*, all the genes of the cluster are transcribed from left to right. Positions and orientations of the primers P1–P7 used for PCR and sequencing are indicated by horizontal arrows.

is positioned in the central part of the large single-copy region (LSC) of the maize plastome. Its constituent genes *trnT*, *rps4*, *psaA*, *psaB*, *rps14*, and *trnR*, which had been analyzed previously by several groups (18, 20–24) (the region comprising the *rps4*, *trnS*, IRF170, and *psaA* genes had also been analyzed in our laboratory by G. L. Igloi, unpublished results), show the same polarity as IRF170, which suggests that the entire cluster constitutes a single transcriptional unit even though the *rps4*–IRF170 intergenic region does contain the *trnS* gene (23) of opposite polarity.

In Fig. 1 the positions and orientations of the primers P1–P7 used for amplification reactions and sequencing are indicated schematically, whereas in Fig. 2 the exact positions of the primers P1–P5 used for amplification of IRF170-specific genomic and cDNA sequences are given. By using the primer pair P1/P2, amplification products comprising the entire IRF170-encoded sequence can be obtained (Fig. 3, lanes a and b). Their sizes are in accordance with the calculated lengths of 2168 and 701 bp for the genomic DNA and cDNA, respectively, with the size difference reflecting the loss of the two introns in the cDNA. Subsequent sequence analyses led to the detection of two positions in which the cDNA sequence deviated from the genomic sequence by a C-to-T transition (corresponding to a G-to-A transition where the complementary strand is analyzed). The sequence autoradiograms representing the two editing events I and II are depicted in Fig. 4. The corresponding positions within the first and second exon of the IRF170-encoded sequence and the codon changes caused by the two editing events are indicated in Fig. 2. An alignment of the IRF170-encoded amino acid sequences, as presented in Fig. 5, shows that the editing events cause restoration of codons for amino acid residues which are conserved at the gene level in liverwort and other species. The phenylalanine residue restored at editing site I is conserved even in the IRF170-encoded homolog of the more distantly related cyanobacterium *Syn-*

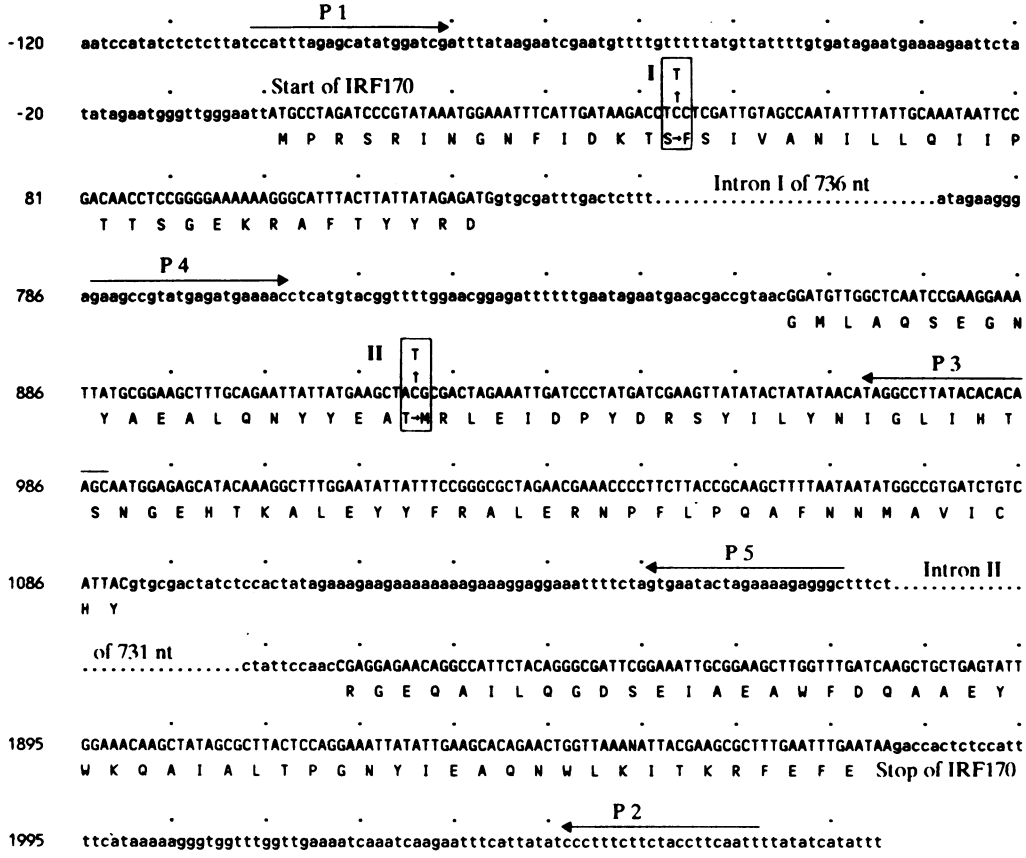

FIG. 2. Nucleotide sequence and deduced amino acid sequence of the three exons of the maize chloroplast IRF170. The 5′ and 3′ noncoding regions and the exon flanking sequences of the two introns are also presented. Positions and orientations of the primers P1–P5 used for PCR and sequencing are indicated by horizontal arrows. The two editing positions I and II, where C-to-U transitions alter the amino acid sequence, are boxed.

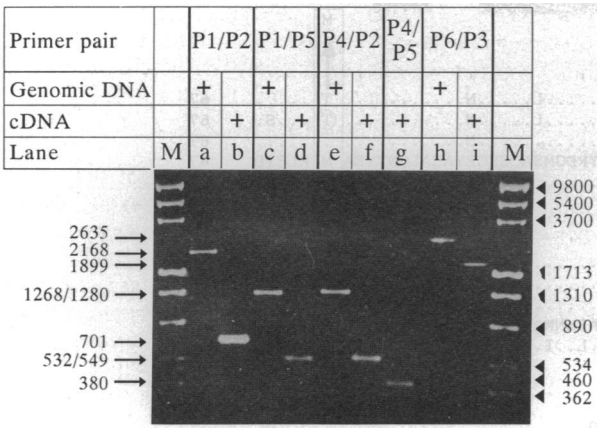

FIG. 3. Amplification products obtained by PCR with pairs of primers, P1–P6, specific for the maize chloroplast *rps4*–IRF170 region. Amplifications were carried out either with DNA from maize chloroplasts or with cDNA obtained by reverse transcription of RNA from maize chloroplasts. Positions and expected lengths in base pairs of the amplification products are marked by horizontal arrows at left. Positions of the length markers (lanes M) are indicated by arrowheads at right.

echocystis. Previously, the ACG-to-ATG codon transition has been observed for the restoration of initiating methionine codons (3, 4). With editing site II this transition is now also observed for the restoration of an internal methionine codon. The presence of an ACG threonine codon in the homologous position of the rice IRF170 suggests that editing of site II occurs also in rice (Fig. 5), which, however, remains to be verified experimentally. The occurrence of two editing events by which conserved amino acid residues are restored supports the view that the IRF170-encoded peptide is subject to stringent structural constraints which do not allow amino acid substitutions at these two sites. From this we further conclude that the IRF170-encoded peptide is indispensable for chloroplast function and, hence, that IRF170 is not degenerated to a pseudogene.

In addition to editing at position II, the autoradiogram on the right half of Fig. 4 shows a further deviation between the cDNA and the genomic DNA sequence. This deviation (marked by a star only three bases upstream of editing site II) seems to indicate a C/G ambiguity in the genomic sequence which apparently is resolved into two distinct positions (G and C) in the cDNA sequence. However, both the genomic DNA and the cDNA can be cleaved at this position with the restriction endonuclease *Alu* I (data not shown), thus confirming the AGCT sequence also for the genomic DNA. The ambiguity of the genomic DNA could also be resolved more directly by separating a product mixture obtained with 7-deaza-dGTP during the sequencing reaction in a formamide/urea gel (ref. 25; data not shown). Apparently, a stable minihelix (and hence a compression) can be formed with the sequence GCGTAGC of the genomic DNA but not with the GCATAGC sequence of the cDNA, in spite of the G-to-A exchange not altering the helical stem. This interpretation is strongly supported by a recent systematic study in which the stabilities and compressions during sequence analysis of a series of similar minihelices have been measured (26). Thus, we conclude that the apparent difference between the genomic DNA and the cDNA is only indirectly caused by the base change at the adjacent editing site II, just resolving a sequence compression, but that no additional editing event is reflected by this sequence deviation. The absence of similar minihelices in the vicinity of other editing sites also excludes the possibility that such structures act as determinants for the editing process.

Editing of the IRF170 Transcript Precedes Splicing. Application of the intron-specific primers P4 and P5 allows amplification of cDNAs which originate from partially spliced or even from unspliced IRF170 transcripts. As shown in Fig. 3 (lanes c and d), the primer pair P1/P5 leads to a genomic DNA product corresponding to the expected size of 1268 bp which still contains sequences of the first and second intron, whereas the cDNA obtained with the same primer pair has a size of 532 bp (Fig. 6*a*) due to the absence of intron I. Sequence analysis of the cDNA shows that editing at both editing sites is complete (autoradiograms of Fig. 6*a*). Con-

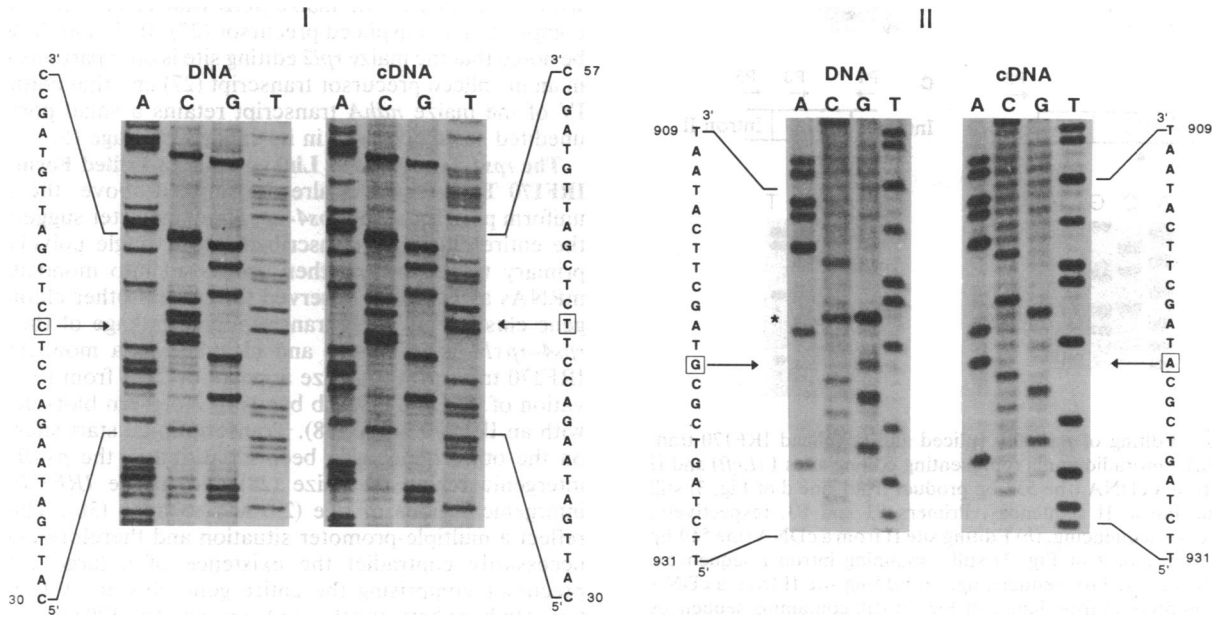

FIG. 4. Identification of editing positions by comparison of IRF170 nucleotide sequences obtained either with amplified genomic DNA or with cDNA. The 2168- and 701-bp amplification products (Fig. 3, lanes a and b) were isolated and sequenced directly by a modified chain-termination method using primer P1 for editing position I and primer P3 for editing position II. The editing positions where chloroplast DNA and cDNA differ are marked by arrows. Due to the polarity of primer P3 the sequences of the editing site II are complementary to the IRF170 transcript. A sequence compression close to editing site II in the genomic DNA is marked by a star.

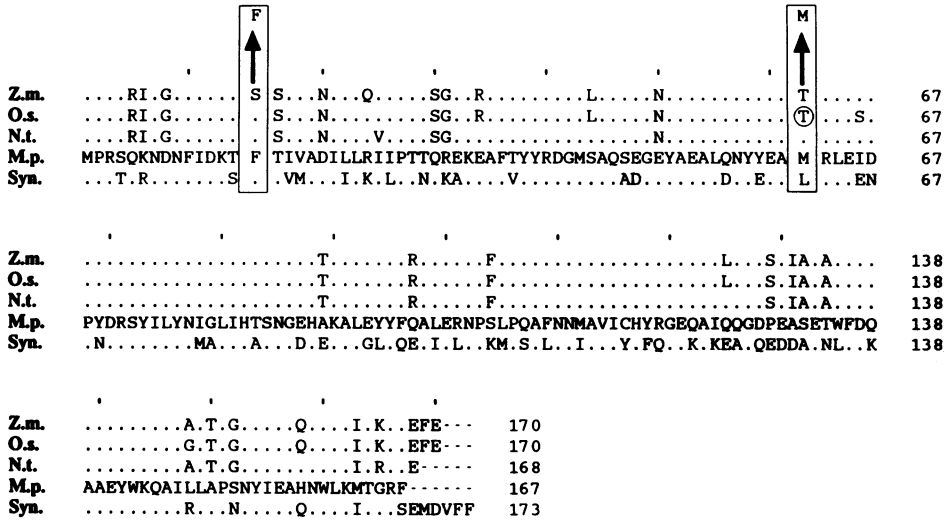

FIG. 5. Alignment of the amino acid sequences encoded by the maize (*Zea mays*, Z.m.) chloroplast IRF170 with homologous sequences from rice (*Oryza sativa*, O.s.), tobacco (*Nicotiana tabacum*, N.t.), and liverwort (*Marchantia polymorpha*, M.p.) chloroplasts and from a cyanobacterium (*Synechocystis* sp. PCC6803, Syn.). Only amino acid residues deviating from the liverwort sequence are indicated. The two amino acid substitutions resulting from editing are marked by vertical arrows. Circling indicates an amino acid position in rice that is most likely to be restored to the conserved amino acid residue by editing of the respective codon.

versely, the primer pair P4/P2 leads to a genomic DNA product of 1280 bp still containing sequences of the first and second intron, whereas the cDNA obtained with the same primer combination, due to the absence of intron II, shows the smaller size of 549 bp (Fig. 3, lanes e and f). Sequence analysis of the cDNA, in this case containing only editing site II, again shows complete editing (Fig. 6*b*). This allows the conclusion that the partially spliced IRF170 transcripts still containing one of the two introns are already fully edited.

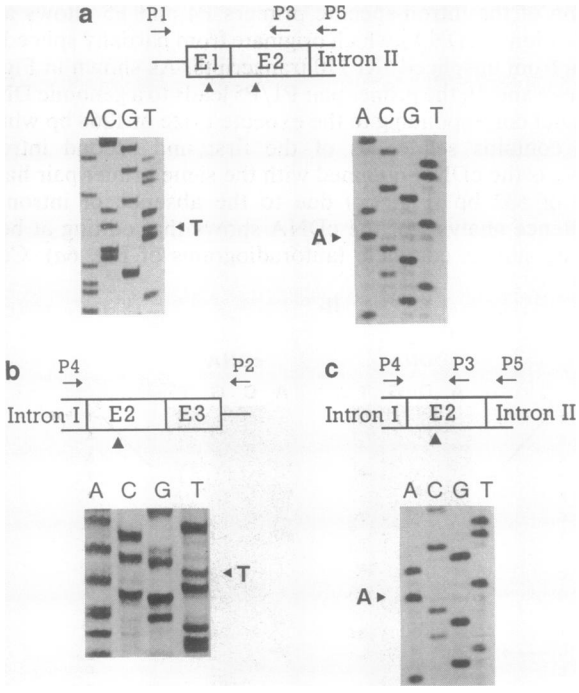

FIG. 6. Editing of partially spliced and unspliced IRF170 transcripts. (*a*) Autoradiograms representing editing sites I (*Left*) and II (*Right*) from a cDNA (the 532-bp product from lane d of Fig. 3) still containing intron II sequences. Primers P1 and P3, respectively, were used for sequencing. (*b*) Editing site II from a cDNA (the 549-bp product from lane f of Fig. 3) still containing intron I sequences. Primer P4 was used for sequencing. (*c*) Editing site II from a cDNA (the 380-bp product from lane g of Fig. 3) still containing sequences of both introns. Primer P3 was used for sequencing. Above the autoradiograms the cDNA amplification products obtained with the primer combinations P1/P5 (*a*), P4/P2 (*b*) and P4/P5 (*c*) are presented schematically. E1, E2, and E3 symbolize the exon sequences. Arrowheads mark the positions of the editing sites in the schematic representations as well as in the sequence autoradiograms.

An attempt was then made to isolate a cDNA which is derived from an unspliced IRF170 transcript by using the primer pair P4/P5. This yielded a 380-bp cDNA (Fig. 3, lane g) which was indistinguishable in size from the corresponding genomic DNA (data not shown). As in the previous case, only the second exon with its editing site II is contained in this cDNA. Sequence analysis again shows that this cDNA carries a fully edited site II (Fig. 6*c*). Thus, editing of at least this site must be a very early process which is finished before the onset of both the splicing events. This is in agreement with the editing of the maize *petB* transcript, which is also complete in its unspliced precursor (27). It should, however, be noted that the maize *rpl2* editing site is only partially edited in an unspliced precursor transcript (27) and that editing site IV of the maize *ndhA* transcript retains a small portion of unedited sequence even in its spliced message (5).

The *rps4* Transcript Is Linked with an Edited Form of the IRF170 Transcript. As already outlined above, the almost uniform polarity of the *rps4–rps14* gene cluster suggests that the entire cluster is transcribed into a single polycistronic primary transcript and then processed into monocistronic mRNAs as has been observed for several other chloroplast gene clusters (1, 2). A transcriptional linkage of the entire *rps4–rps14* gene cluster and cleavage to a monocistronic IRF170 transcript in maize appears evident from the observation of 9.5- and 0.95-kb bands in Northern blots detected with an IRF170 probe (18). Transcriptional start sites have, on the other hand, also been identified in the *psaB–rps14* intergenic region of maize (28) and in the *IRF170–psaA* intergenic regions of rice (29) and tobacco (30). This may reflect a multiple-promoter situation and therefore does not necessarily contradict the existence of a large common precursor comprising the entire gene cluster. A transcriptional linkage between the *rps4* gene and IRF170 had not been tested experimentally.

By using the primer pair P6/P3, a genomic DNA and a cDNA comprising the entire *rps4* coding region and the 5′-terminal half of the IRF170-encoded sequence can be amplified (Fig. 3, lanes h and i). Their sizes are in accordance

with the lengths of 2635 and 1899 bp calculated for the genomic DNA and cDNA, respectively, from published sequences (20, 21, 23) (a gap of about 60 bp between the *trnS* gene and IRF170 was sequenced in our laboratory). Their size difference, reflecting the loss of intron I of the IRF170 transcript, clearly ensures an RNA origin of the 1899-bp amplification product. The formation of this cDNA at the same time indicates a transcriptional linkage between the *rps4* gene and IRF170, and this implies also a transcription of the *rps4*–IRF170 intergenic region in opposite polarity with respect to the *trnS* gene. Thus, the intergenic transcript contains an antisense tRNA (or an antisense tRNA precursor). It is not clear how inhibitory effects that one might expect for the function (or maturation) of the serine tRNA are avoided, but similar situations have been reported for other polycistronic chloroplast transcripts containing antisense tRNA sequences (31, 32). Our observation of the *rps4*–IRF170 linkage supports the supposition that the entire gene cluster constitutes a single transcriptional unit (18), although it still remains to be determined whether the flanking tRNA genes (*trnT* and *trnR*) have to be included in this unit.

Sequence analysis of the *rps4*-containing cDNA with P6 and P7 as sequencing primers reveals no differences with respect to the *rps4* genomic sequence, which indicates that the *rps4* transcript is not subject to editing. On the other hand, the cDNA obtained with the primer pair P6/P3 includes also the two editing sites identified in the IRF170 transcript (see above), and analysis of the IRF170 cDNA portion reveals complete editing at both sites (data not shown). From this we conclude that editing of the IRF170 transcript is an early event also with respect to the cleavage of the *rps4*–IRF170 precursor to its monocistronic mRNAs. The absence of the IRF170 intron sequences in dicistronic transcripts demonstrates that splicing can precede processing to the monocistronic transcripts. Thus, a temporal order of the three processing steps emerges with editing being the first, splicing the second, and cleavage to monocistronic mRNAs as probably the last event in the processing cascade. It remains to be seen whether this order is determined merely by different rates of the three processes or whether it is dictated by mechanistic links compatible only with this order of events.

We wish to thank Drs. Gabor L. Igloi and Rainer M. Maier for critical reading of the manuscript, Kai Neckermann for computer work, Elfi Schiefermayr for primer syntheses, and Sigrid Krien for technical assistance. This work was supported by grants from the Deutsche Forschungsgemeinschaft (SFB 206) and the Fonds der Chemischen Industrie to H.K.

1. Sugiura, M. (1991) in *Cell Culture and Somatic Cell Genetics of Plants*, eds. Bogorad, L. & Vasil, I. K. (Academic, London), Vol. 7A, pp. 125–137.
2. Gruissem, W. & Tonkyn, J. C. (1993) *Crit. Rev. Plant Sci.* **12**, 19–55.
3. Hoch, B., Maier, R. M., Appel, K., Igloi, G. L. & Kössel, H. (1991) *Nature (London)* **353**, 178–180.
4. Kudla, J., Igloi, G. L., Metzlaff, M., Hagemann, R. & Kössel, H. (1992) *EMBO J.* **11**, 1099–1103.
5. Maier, R. M., Hoch, B., Zeltz, P. & Kössel, H. (1992) *Plant Cell* **4**, 609–616.
6. Maier, R. M., Neckermann, K., Hoch, B., Akhmedov, N. B. & Kössel, H. (1992) *Nucleic Acids Res.* **20**, 6189–6194.
7. Zeltz, P., Hess, W. R., Neckermann, K., Börner, T. & Kössel, H. (1993) *EMBO J.* **12**, 4291–4296.
8. Benne, R. (1990) *Trends Genet.* **6**, 177–181.
9. Stuart, K. (1991) *Annu. Rev. Microbiol.* **45**, 327–344.
10. Chan, L. (1993) *BioEssays* **15**, 33–41.
11. Covello, P. S. & Gray, M. W. (1989) *Nature (London)* **341**, 662–666.
12. Gualberto, J. M., Lamattina, L., Bonnard, G., Weil, J.-H. & Grienenberger, J. M. (1989) *Nature (London)* **341**, 660–662.
13. Hiesel, R., Wissinger, B., Schuster, W. & Brennicke, A. (1989) *Science* **246**, 1632–1634.
14. Mahendran, R., Spottswood, M. R. & Miller, D. L. (1991) *Nature (London)* **349**, 434–438.
15. Shimada, H. & Sugiura, M. (1991) *Nucleic Acids Res.* **19**, 983–995.
16. Vörös, K., Hübschmann, T. & Börner, T. (1992) *Endocyt. Cell Res.* **9**, 71–76.
17. Wolfe, K. H., Morden, C. W. & Palmer, J. D. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10648–10652.
18. McCullough, A. J., Kangasjärvi, J., Gengenbach, B. G. & Jones, R. J. (1992) *Plant Physiol.* **100**, 958–964.
19. Bachmann, B., Lüke, W. & Hunsmann, G. (1990) *Nucleic Acids Res.* **18**, 1309.
20. Subramanian, A., Steinmetz, A. & Bogorad, L. (1983) *Nucleic Acids Res.* **11**, 5277–5286.
21. Kangasjärvi, J., McCullough, A. & Gengenbach, B. G. (1991) *Plant Mol. Biol.* **17**, 513–515.
22. Fish, L. E., Kück, U. & Bogorad, L. (1985) *J. Biol. Chem.* **260**, 1413–1421.
23. Steinmetz, A. A., Krebbers, E. T., Schwarz, Z., Gubbins, E. J. & Bogorad, L. (1983) *J. Biol. Chem.* **258**, 5503–5511.
24. Rodermel, S., Orlin, P. & Bogorad, L. (1987) *Nucleic Acids Res.* **15**, 5493.
25. Griersen, D. (1982) in *Gel Electrophoresis of Nucleotides—A Practical Approach*, eds. Rickwood, D. & Hames, B. D. (IRL, Oxford, U.K.), pp. 1–38.
26. Hirao, I., Nishimura, Y., Tagawa, Y., Watanabe, K. & Miura, K. (1992) *Nucleic Acids Res.* **20**, 3891–3896.
27. Freyer, R., Hoch, B., Neckermann, K., Maier, R. M. & Kössel, H. (1993) *Plant J.* **4**, 621–629.
28. Srinivasa, B. R. & Subramanian, A. R. (1987) *Biochemistry* **26**, 3188–3192.
29. Chen, M., Cheng, M. & Chen, S. G. (1993) *Plant Cell Physiol.* **34**, 577–584.
30. Meng, B. Y., Tanaka, M., Wakasugi, T., Ohme, M., Shinozaki, K. & Sugiura, M. (1988) *Curr. Genet.* **14**, 395–400.
31. Haley, J. & Bogorad, L. (1990) *Plant Cell* **2**, 323–333.
32. Neuhaus, H. & Link, G. (1990) *Curr. Genet.* **18**, 377–383.