Mitochondrial and Chloroplast Phage-Type RNA Polymerases in *Arabidopsis*

Boris Hedtke, Thomas Börner,* Andreas Weihe

In addition to the RNA polymerases (RNAPs) transcribing the nuclear genes, eukaryotic cells also require RNAPs to transcribe the genes of the mitochondrial genome and, in plants, of the chloroplast genome. The plant *Arabidopsis thaliana* was found to contain two nuclear genes similar to genes encoding the mitochondrial RNAP from yeast and RNAPs of bacteriophages T7, T3, and SP6. The putative transit peptides of the two polymerases were capable of targeting fusion proteins to mitochondria and chloroplasts, respectively, in vitro. The results indicate that the mitochondrial RNAP in plants is a bacteriophage-type enzyme. A gene duplication event may have generated the second RNAP, which along with the plastid-encoded eubacteria-like RNAP could transcribe the chloroplast genome.

In both yeast and *Neurospora*, a nuclear gene encodes the mitochondrial RNAP, which is a single-subunit enzyme that is most similar to RNAPs of the bacteriophages T3, T7, and SP6 (*1*). On the basis of isolation of short partial sequences from a variety of eukaryotes, it has been suggested that this type of polymerase represents the common type of the mitochondrial RNAP in many, if not all, eukaryotes that contain mitochondria (*2*). Plastids, on the other hand, contain the only prokaryotic genome transcribed by more than one RNAP. Subunits of one of the plastid RNAPs, similar to eubacterial RNAP, are encoded by plastid genes (*3*). Another RNAP of unknown properties is encoded by the nuclear genome (*4*). This nuclear-encoded plastid RNAP may transcribe a set of promoters different from those used by the chloroplast-encoded RNAP. We report here the sequence of two nuclear genes in the higher plant *Arabidopsis thaliana* with similarity to genes encoding RNAPs of the single-subunit type. The genes encode transit peptides that are capable of targeting fusion proteins to mitochondria and chloroplasts, indicating that they encode plant mitochondrial and plastid RNAPs, respectively.

To clone the mitochondrial RNAP from *A. thaliana*, we used a cDNA sequence encoding a putative mitochondrial RNAP from the dicotyledoneous plant *Chenopodium album* (accession number Y08067) (*5*) to screen a genomic DNA library of *A. thaliana*. Two clones exhibiting different restriction and hybridization patterns (designated Y and Z) were isolated from the *Arabidopsis* library and further analyzed. Comparison with the GeneBank and European Molecular Biology Laboratory (EMBL) databases indicated regions in both clones of sequence similarity to the mitochondrial RNAPs from *Saccharomyces cerevisiae* and *Neurospora crassa* and other phage-like RNAPs. Polymerase chain reaction (PCR) primers were designed to amplify cDNAs, corresponding to the two genomic clones, by conventional reverse transcriptase (RT)– PCR and 3′– and 5′–rapid amplification of cDNA ends (RACE). The sequence and size of the composite full-length cDNAs, established from overlapping amplification products, were confirmed by RT-PCR of the whole coding region, as well as by Northern (RNA) blot analysis (*6*). The composite full-length Y and Z cDNAs consisted of 3602 and 3482 nucleotides (nt) composing open reading frames of 976 and 993 amino acids, respectively.

Analysis of total DNA from *A. thaliana* indicated that the two phage-like RNAP-related sequences from *Arabidopsis* represented single-copy genes (*6*). Thus, the nuclear genome of *A. thaliana* carries at least two distinct genes for bacteriophage-like RNAPs, which are designated as *RpoY* and *RpoZ* (the putative gene products being RPOY and RPOZ with predicted molecular masses of 111 and 113 kD). The amino acid sequences of RPOY and RPOZ share 55% identical residues corresponding to a nucleotide sequence similarity of 62.7%. Both amino acid sequences showed extensive similarity to phage-like RNAPs. The overall degree of similarity between the two *Arabidopsis* RNAPs and the yeast enzyme decreases from the COOH-terminus to the NH_2-terminus (Fig. 1). The yeast mitochondrial RNAP has an NH_2-terminal extension of 266 amino acids that did not align with the two *A. thaliana* sequences. Furthermore, the *Arabidopsis* enzymes lack a stretch of 95 catalytically nonessential amino acids found in the *S. cerevisiae* sequence. All residues thought to be catalytically essential for polymerase activity, on the basis of molecular studies and analysis of the crystal structure of the T7 RNAP (*7*), are conserved.

Sequences of 9208 and 9080 nt have been determined for the genomic clones *RpoY* and *RpoZ*, respectively (Fig. 2). Comparison of the sequences of the genomic and cDNA clones revealed that both genes are organized in 19 exons, 48 to 762 nt (*RpoY*) and 48 to 846 nt (*RpoZ*) long, interrupted by 18 introns, 81 to 200 nt (*RpoY*) and 76 to 326 nt (*RpoZ*) long. All exon-intron junctions contained conserved GT and AG sequences at the 5′ and 3′ ends of the introns, respectively. The length of 16 of the 19 exons was identical in both genes. All intron insertion sites were precisely conserved relative to the aligned amino acid sequences. There was only weak similarity between corresponding intron sequences of *RpoY* and *RpoZ*. The genes contained untranslated leader sequences of 187 nt (Y gene) and 58 nt (Z gene), when one assumes that the 5′ ends of the two cDNAs correspond to the transcription start points.

Phylogenetic analysis based on full-length alignments of single-subunit RNAPs (Fig. 3) shows that the *Arabidopsis* RNAPs form a cluster together with the *Chenopodium* polymerase and the two fungal mitochondrial RNAPs, separated from both phage and linear plasmid-encoded single-subunit RNAPs found in mitochondria of many fungi and some plants. Branching patterns were robust in bootstrap analysis, and the presented tree topology was supported by other methods of tree construction (maximum parsimony and maximum likelihood algorithms).

To determine the cellular localization of the gene products, recombinant plasmids were constructed containing 131 and 124 NH_2-terminal amino acids composing the putative transit sequences of RPOY and RPOZ, respectively, fused to a green fluorescent protein (GFP) (Fig. 4A). The constructs were used to conduct in vitro import into purified mitochondria and chloroplasts (Fig. 4, B and C). The mitochondrial adenosine triphosphatase (ATPase) β presequence was used as a positive control for import into mitochondria (Fig. 4B). The 47-kD RPOY-GFP was processed into a mature 38-kD polypeptide, resistant to proteinase K. Valinomycin, an inhibitor of mitochondrial protein import, abolished import of RPOY-GFP. Binding and processing of the labeled RPOZ precursor polypeptide into proteinase-resistant products were observed at a greatly reduced level. However, in this case, processing was not inhibited by va-

Humboldt University Berlin, Institute of Biology, Chausseestrasse 117, D-10115 Berlin, Germany.

*To whom correspondence should be addressed. E-mail: Thomas=Boerner@biologie.hu-berlin.de

linomycin. As the same situation was observed with the *waxy* presequence that efficiently targets proteins to chloroplasts (8), this effect was due to a contamination of the purified mitochondria with plastids. This contamination of the mitochondrial preparation has been confirmed by light microscopy (6). Thus, from the two *A. thaliana* fusions, only RPOY-GFP was definitively imported into mitochondria in vitro. When the fusion constructs were used for import into purified chloroplasts (Fig. 4C), RPOZ-GFP (45 kD) was very efficiently processed into a proteinase-resistant mature protein of 37 kD. The processing was as efficient as with the positive control for chloroplast import carrying the *waxy* presequence. Alternatively, little binding and no processing were observed for RPOY-GFP or the mitochondrial ATPase β–GFP. Thus, the NH_2-terminal amino acid sequence of RPOZ targeted protein into chloroplasts, and the NH_2-terminus of RPOY targeted protein into mitochondria in vitro.

The present study provides initial evidence for the existence of two genes for bacteriophage-type RNAPs in a higher plant that are candidates for functional organellar enzymes. The in vitro import experiments provide evidence that the two proteins encoded in the nuclear genome of *Arabidopsis* are targeted to different organelles: mitochondria and plastids. The mitochondrial RNAP of *Arabidopsis* is with high certainty encoded by the Y gene (*RpoMt*) and thus, as in fungi and presumably other eukaryotes, is represented by a

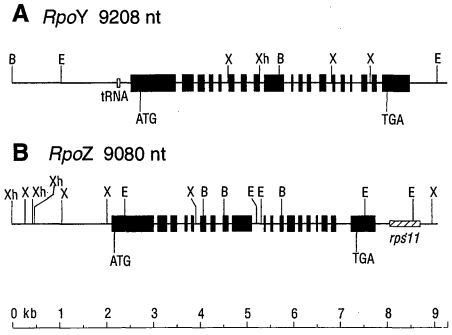

Fig. 2. Genomic organization of the two *A. thaliana* bacteriophage-like RNAP genes as deduced from cloning of genomic and cDNA clones (*12*). (**A**) *RpoY* (EMBL accession number Y09006) and (**B**) *RpoZ* (accession number Y08722). Exon sequences are indicated by filled bars. Putative translational start and stop codons and selected restriction sites (B, Bam HI; E, Eco RI; X, Xba I; Xh, Xho I) are shown. An initiator tRNA-Met gene detected on the reverse strand 480 to 409 nt upstream of the *RpoY* start codon is indicated by an open bar. An *rps11* gene found 703 to 1364 nt downstream of the *RpoZ* stop codon is shown by a hatched bar.

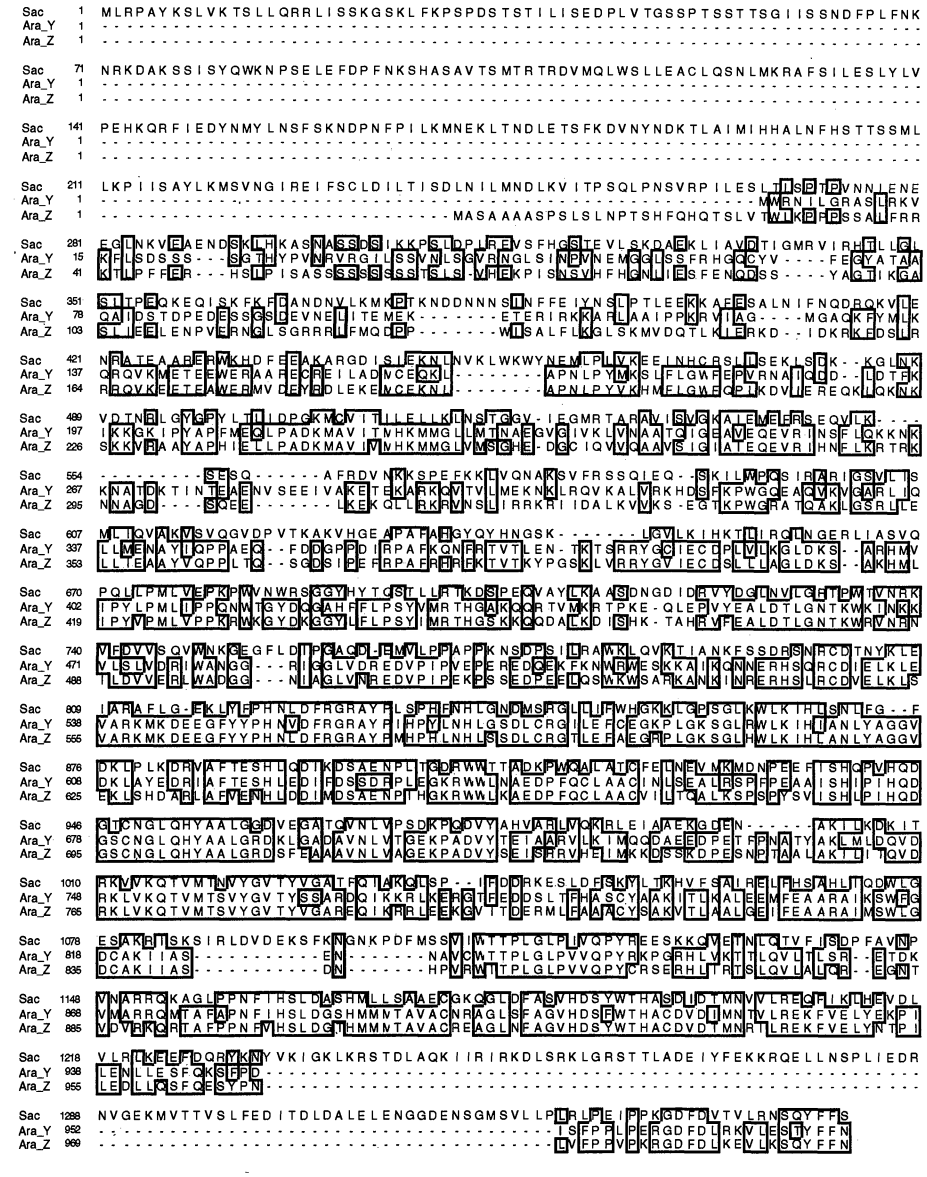

Fig. 1. Alignment of RPOY (Ara_Y; accession number 08137) and RPOZ (Ara_Z; accession number 08463) amino acid sequences with mitochondrial RNAP from yeast (Sac; accession number P13433). The alignment was constructed with the ClustalW algorithm (*11*) and refined manually. Amino acids that are positionally identical are highlighted by boxes.

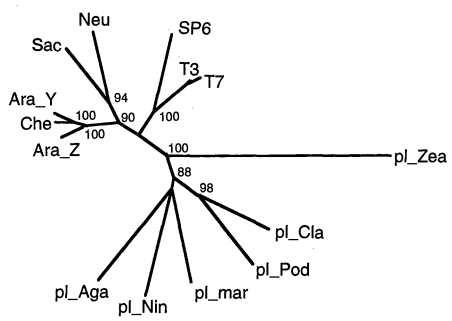

Fig. 3. Phylogenetic tree of bacteriophage-like RNAPs based on maximum likelihood distances calculated with the PUZZLE program (*13*) and the Jones model (*14*) of sequence evolution. Support values for branches are shown at nodes when above 90% (1000 puzzling steps performed). Distances are proportional to relative sequence divergence. Abbreviations are as in Fig. 1, with additional ones as follows (SWISS-PROT accession numbers in parentheses): T7, bacteriophage T7 (P00573); T3, bacteriophage T3 (P07659); SP6, bacteriophage SP6 (P06221); Neu, mitochondrial RNAP from *N. crassa* (P38671); and Che, putative mitochondrial RNAP from *C. album* (EMBL Y08067). Mitochondrial plasmid-encoded single-subunit RNAPs: pl_mar, plasmid maranhar from *N. crassa* (P33540); pl_Aga, plasmid from *Agaricus bitorquis* (P33539); pl_Pod, plasmid pAL2-1 from *Podospora anserina* (505556); pl_Nin, plasmid kalilo from *Neurospora intermedia* (P33541); pl_Cla, plasmid pCIK1 from *Claviceps purpurea* (P22372); pl_Zea, plasmid S-2 from *Z. mays* (P10581).

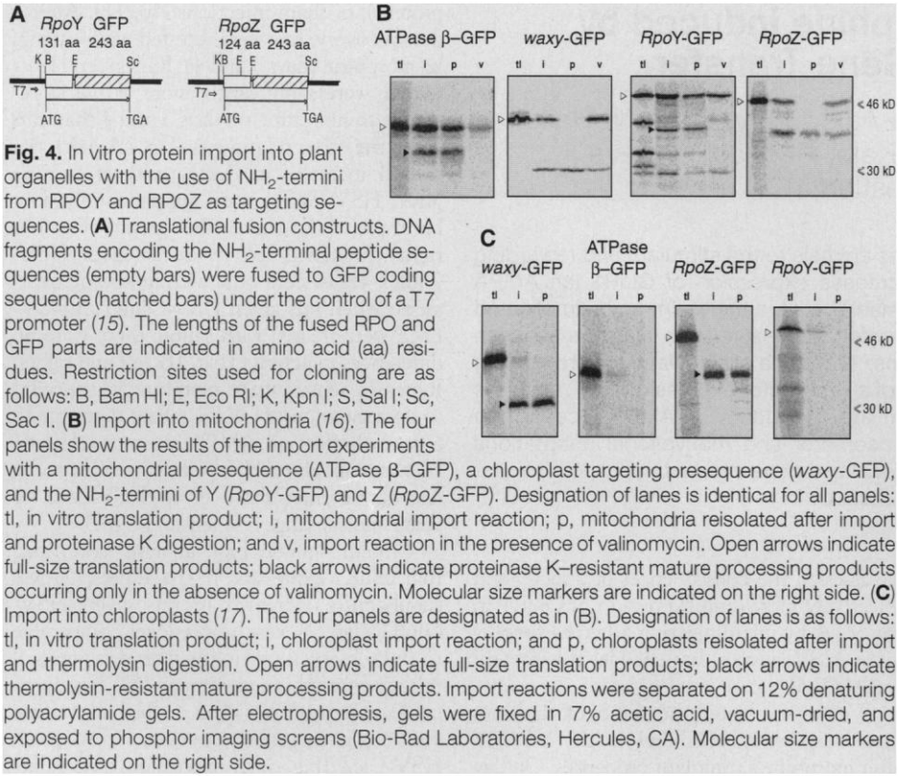

Fig. 4. In vitro protein import into plant organelles with the use of NH_2-termini from RPOY and RPOZ as targeting sequences. (**A**) Translational fusion constructs. DNA fragments encoding the NH_2-terminal peptide sequences (empty bars) were fused to GFP coding sequence (hatched bars) under the control of a T7 promoter (*15*). The lengths of the fused RPO and GFP parts are indicated in amino acid (aa) residues. Restriction sites used for cloning are as follows: B, Bam HI; E, Eco RI; K, Kpn I; S, Sal I; Sc, Sac I. (**B**) Import into mitochondria (*16*). The four panels show the results of the import experiments with a mitochondrial presequence (ATPase β–GFP), a chloroplast targeting presequence (*waxy*-GFP), and the NH_2-termini of Y (*RpoY*-GFP) and Z (*RpoZ*-GFP). Designation of lanes is identical for all panels: tl, in vitro translation product; i, mitochondrial import reaction; p, mitochondria reisolated after import and proteinase K digestion; and v, import reaction in the presence of valinomycin. Open arrows indicate full-size translation products; black arrows indicate proteinase K–resistant mature processing products occurring only in the absence of valinomycin. Molecular size markers are indicated on the right side. (**C**) Import into chloroplasts (*17*). The four panels are designated as in (B). Designation of lanes is as follows: tl, in vitro translation product; i, chloroplast import reaction; and p, chloroplasts reisolated after import and thermolysin digestion. Open arrows indicate full-size translation products; black arrows indicate thermolysin-resistant mature processing products. Import reactions were separated on 12% denaturing polyacrylamide gels. After electrophoresis, gels were fixed in 7% acetic acid, vacuum-dried, and exposed to phosphor imaging screens (Bio-Rad Laboratories, Hercules, CA). Molecular size markers are indicated on the right side.

bacteriophage-type enzyme. Genes of a eubacterial-like RNAP have been detected in a mitochondrial genome (*9*). This observation raises the question whether the organism, the freshwater protozoon *Reclinomonas americana*, additionally has a mitochondrial RNAP of the phage type. If so, the situation would resemble transcription in chloroplasts. To date, there has been only circumstantial evidence for the existence of a nuclear-encoded plastid RNAP, which has been demonstrated by transcriptional analysis of plastids lacking the chloroplast-encoded enzyme (*4*). Our results strongly suggest that the postulated nuclear-encoded plastid RNAP is encoded by the Z gene, for which we therefore propose the designation *RpoPt*. The proposed subcellular localization and function of the gene product gain support from the observation of RNAP activity associated with a 110-kD protein in preparations from spinach chloroplasts (*10*). The size of this RNAP fits well with the size (113 kD) predicted for the *RpoPt* product. The high conservation of the overall organization of the two *Arabidopsis Rpo* genes, as well as their position in phylogenetic trees, suggests that *RpoPt* may have originated from a recent duplication event from *RpoMt*. Gene duplication could have been followed by the gene product retargeting from the mitochondria to the plastids, leading to the elaboration of a dual RNAP system of transcription in the plastids, which originally relied solely on transcription by the organelle-encoded eubacterial-like RNAP.

REFERENCES AND NOTES

1. J. L. Kelly, A. L. Greenleaf, I. R. Lehmann, *J. Biol. Chem.* **261**, 10348 (1986); B. Chen, A. R. Kubelik, S. Mohr, C. A. Breitenberger, *ibid.* **271**, 6537 (1996); B. S. Masters, L. L. Stohl, D. A. Clayton, *Cell* **51**, 89 (1987).
2. N. Cermakian, T. M. Ikeda, R. Cedergren, M. W. Gray, *Nucleic Acids Res.* **24**, 648 (1996).
3. L. Bogorad, in *The Molecular Biology of Plastids*, L. Bogorad and I. K. Vasil, Eds. (Academic Press, San Diego, CA, 1990), pp. 93–124.
4. W. R. Hess, A. Prombona, B. Fieder, A. R. Subramanian, T. Börner, *EMBO J.* **12**, 563 (1993); S. C. Ems *et al.*, *Plant Mol. Biol.* **29**, 721 (1995); L. A. Allison, L. D. Simon, P. Maliga, *EMBO J.* **15**, 2802 (1996).
5. A. Weihe, B. Hedtke, T. Börner, *Nucleic Acids Res.* **25**, 2319 (1997).
6. B. Hedtke, T. Börner, A. Weihe, data not shown.
7. R. Sousa, Y. J. Chung, J. P. Rose, B.-C. Wang, *Nature* **364**, 593 (1993); D. Patra, E. M. Lafer, R. Sousa, *J. Mol. Biol.* **224**, 307 (1992).
8. R. B. Klösgen, H. Saedler, J.-H. Weil, *Mol. Gen. Genet.* **217**, 155 (1989).
9. B. F. Lang *et al.*, *Nature* **387**, 493 (1997).
10. S. Lerbs-Mache, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5509 (1993).
11. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994).
12. Screening of an *A. thaliana*, ecotype Columbia, genomic library in Lambda GEM-11 (established by J. T. Mulligan and R. W. Davis) and isolation of clones were performed according to standard methods. RT-PCR and RACE were conducted with adaptor-ligated double-stranded cDNA (Marathon cDNA Amplification Kit; Clontech, Palo Alto, CA) synthesized from polyadenylated RNA from green leaves of *A. thaliana*, ecotype Columbia. In PCR reactions, Klentaq polymerase mix (Clontech) was used for improved fidelity of amplification.
13. K. Strimmer and A. von Haeseler, *Mol. Biol. Evol.* **13**, 964 (1996).
14. D. T. Jones, W. R. Taylor, J. M. Thornton, *Comput. Appl. Biosci.* **8**, 275 (1992).
15. *Rpo*Y and *Rpo*Z cDNA sequences corresponding to the first 131 and 124 amino acids, respectively, of the predicted gene products were amplified with the gene-specific primer pairs cggatccaaggagatataacaATGTGGAGGAACATTCTGGG and cgaattcCTTCTGTGCTCCCATTCCCG for *Rpo*Y and cggatccaaggagatataacaATGGCTTCCGCTGCGGCTTC and cgaattcAAACAGTCTTCTCCGACCAGAAAGTCC for *Rpo*Z (gene-specific sequences are indicated by capital letters). PCR products were ligated to Bam HI–Eco RI–restricted DNA of pBIN-mGFP5-ER [J. Haseloff, K. R. Siemering, D. Prasher, S. Hodge, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2122 (1997)], thereby replacing the endoplasmic reticulum (ER) targeting signal of the parental vector by the NH_2-terminal peptides of *Rpo*Y and *Rpo*Z. GFP fusions were excised from the pBIN plasmids by Sac I–Bam HI digestion and ligated into pBluescriptSK to yield constructs *Rpo*Y-GFP and *Rpo*Z-GFP, respectively. A GFP fusion with the 59 NH_2-terminal residues of mitochondrial ATPase β from *Nicotiana plumbaginifolia* [F. Chaumont, M. de Castro Silva Filho, D. Thomas, S. Leterme, M. Boutry, *Plant Mol. Biol.* **24**, 631 (1994)] served as a control for mitochondrial import (construct ATPase β–GFP). The chloroplast import control construct (*waxy*-GFP) consisted of a 470–base pair fragment of *Zea mays* glycogen synthase (*8*) encoding 115 NH_2-terminal amino acids of the protein fused to GFP. We carried out in vitro translation with T7 Coupled Reticulocyte Lysate System (Promega, Madison, WI) using 1 μg of plasmid, 40 U of RNasin, and 100 μCi of ^{35}S-labeled methionine (3000 Ci/mmol) in a 50-μl translation reaction.
16. For mitochondrial import studies, pea seedlings (*Pisum sativum* L., var. progress number 9) were grown in the dark for 7 to 10 days. Isolation of mitochondria and in vitro import was performed essentially as described [F. Chaumont, V. O'Riordan, M. Boutry, *J. Biol. Chem.* **265**, 16856 (1990)] with the following modifications: mitochondria equivalent to 100 μg of protein and 10 μl of in vitro translation mix were added to 160 μl of uptake medium. Import reactions were incubated at 23°C for 40 min. Proteinase K (Boehringer Mannheim, Mannheim, Germany) treatment at a final concentration of 50 μg/ml was done for 20 min on ice. In control reactions, valinomycin was added at a concentration of 20 μM.
17. Chloroplasts were isolated from a commercial cultivar of spinach (*Spinacia oleracea* L.). Organelle isolation and in vitro import followed the protocol of B. D. Bruce, S. Perry, J. Froehlich, and K. Keegstra [in *Plant Molecular Biology Manual*, S. B. Gelvin and R. A. Schilperpoort, Eds. (Kluwer Academic, Dordrecht, Netherlands, 1994), pp. J1, 1–15]. Import reactions consisted of 62 μl of import buffer, 3 μl of 100 mM adenosine-5′-triphosphate, 10 μl of in vitro translation mix, and 25 μl of purified chloroplasts (equivalent to a chlorophyll concentration of 1 mg/ml). Import was allowed to take place for 30 min at room temperature. Thermolysin (Sigma) treatment (0.2 mg/ml) was performed after import for 30 min on ice.

18. We thank J. Haseloff for the vector pBIN-mGFP5–ER; the *Arabidopsis* DNA Stock Centre (Max Planck Institut für Züchtungsforschung, Köln, Germany) for the genomic *A. thaliana* library; M. Boutry for a ATPase β presequence clone; R. B. Klösgen and R. Lührs for *waxy* constructs; C. Stock, I. Wagner, and J. Müller for technical assistance; W. Schuster, Berlin, for sharing unpublished data of an *Arabidopsis* DNA sequence [accession number Y09432 (W. Schuster *et al.*, unpublished data)] that is nearly identical to clone *Rpo*Y; and B. Neilan for critical reading of the manuscript. Part of this work was funded by grants from the Deutsche Forschungsgemeinschaft and the Fond der Chemischen Industrie to T.B.

28 March 1997; accepted 24 June 1997