ELSEVIER

Gene 290 (2002) 95–105

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Two *RpoT* genes of *Physcomitrella patens* encode phage-type RNA polymerases with dual targeting to mitochondria and plastids ☆

Uwe Richter[a], Justine Kiessling[b], Boris Hedtke[a], Eva Decker[b], Ralf Reski[b], Thomas Börner[a], Andreas Weihe[a,*]

[a]*Institut für Biologie, Humboldt-Universität, Chausseestrasse 117, D-10115 Berlin, Germany*
[b]*Institut für Pflanzenbiotechnologie, Universität Freiburg, Schaenzlestrasse 1, D-79104 Freiburg, Germany*

Received 19 February 2002; received in revised form 1 March 2002; accepted 19 March 2002
Received by W. Martin

Abstract

Angiosperms possess a small family of phage-type RNA polymerase genes that arose by gene duplication from an ancestral gene encoding the mitochondrial RNA polymerase. We have isolated and sequenced the genes and cDNAs encoding two phage-type RNA polymerases, PpRpoT1 and PpRpoT2, from the moss *Physcomitrella patens*. *PpRpoT1* comprises 19 exons and 18 introns, *PpRpoT2* contains two additional introns. The N-terminal transit peptides of both polymerases are shown to confer dual-targeting of green fluorescent protein fusions to mitochondria and plastids. In vitro translation of the cDNAs revealed initiation of translation at two in-frame AUG start codons. Translation from the first methionine gives rise to a plastid-targeted polymerase, whereas initiation from the second methionine results in exclusively mitochondrial-targeted protein. Thus, dual-targeting of *Physcomitrella* RpoT is caused by and might be regulated by multiple translational starts. In phylogenetic analyses, the *Physcomitrella* RpoT polymerases form a sister group to all other phage-type polymerases of land plants. The two genes result from a gene duplication event that occurred independently from the one which led to the organellar polymerases with mitochondrial or plastid targeting properties in angiosperms. Yet, according to their conserved exon-intron structures they are representatives of the molecular evolutionary line leading to the *RpoT* genes of higher land plants.

Keywords: Organellar transcription; Phage-type RNA polymerase; Bryophyte; Molecular evolution

1. Introduction

The core-subunit of the mitochondrial RNA polymerase (RNAP) in the majority of eukaryotic organisms is represented by a single-subunit RNAP homologous to the RNA polymerases of bacteriophages T3, T7, K11 and SP6. It is encoded by *RpoT* genes residing in the nucleus (reviewed in Hess and Börner, 1999). The only known example in which mitochondrial transcription relies on a chondriom-encoded multi-subunit bacterial-like RNAP is the protozoan *Reclinomonas americana* (Lang et al., 1997). All other eukaryotes seem to have replaced the original α-proteobacterial transcription machinery in mitochondria by the nuclear-encoded phage-type system. Although the origin of the *RpoT* genes remains unclear (Cermakian et al., 1997), the study of the *RpoT* genes and their gene products in a wide variety of higher organisms including a number of land plants has shed some light onto the molecular evolution of this key component of organellar transcription. The nuclear genomes of angiosperms encode a small family of *RpoT* genes, which consists of three members in the dicotyledonous species studied so far (*Arabidopsis*, *Nicotiana* spp.), and of two genes in the monocotyledonous species *Zea mays* and *Triticum aestivum* (Hedtke et al., 1997, 2002; Chang et al., 1999; Ikeda and Gray, 1999; Börner et al., 1999). Analysis of their sequences and a remarkable conservation of their exon-intron structure indicate that the multiple genes are a result of gene duplication events and represent paralogs.

In addition to the gene encoding the mitochondrial RNAP (designated *RpoT1* or *RpoTm*), with orthologs beyond the plant kingdom and throughout the eukaryotic lineage, angiosperms, both monocots and dicots possess a second

Abbreviations: RpoT, bacteriophage-type RNA polymerase; GFP, green fluorescent protein; NEP, nuclear encoded polymerase; PCR, polymerase chain reaction; PEP, plastid encoded polymerase; RNAP, RNA polymerase; RACE, rapid amplification of cDNA ends

☆ EMBL accession numbers: AJ416854, AJ416855.

* Corresponding author. Tel.: +49-30-2093-8153; fax: +49-30-2093-8141.
E-mail address: andreas.weihe@rz.hu-berlin.de (A. Weihe).

0378-1119/02/$ - see front matter
PII: S0378-1119(02)00583-8

RpoT gene (*RpoT3*, or *RpoTp*) encoding a chloroplast localized phage-type RNAP (NEP, nuclear encoded polymerase) supplementing the eubacterial-like plastid transcription system relying on the PEP (plastid encoded polymerase). Subcellular localization of the gene products of *RpoTm* and *RpoTp* has been verified immunohistochemically (Ikeda and Gray, 1999), by in vitro import experiments (Hedtke et al., 1997), *in planta* through stable transformation of *Arabidopsis* with GFP fusion constructs (Hedtke et al., 1999), and in vivo through transient transformation of protoplasts (Chang et al., 1999; Hedtke et al., 2000, 2002). Surprisingly, the dicots *Arabidopsis* and *Nicotiana* harbor a third *RpoT* gene (*RpoT2*, or *RpoTmp*) encoding an RNAP which is dually targeted to both mitochondria and plastids (Hedtke et al., 2000, 2002; Kobayashi et al., 2001).

The *RpoT* genes of all angiosperms studied so far are highly similar and show a remarkable conservation of the number and position of introns and therefore can be traced to a common ancestor. However, as data on *RpoT* genes for green algae or lower plants are lacking, the early molecular evolution of the *RpoT* gene family is still unresolved. The timing of the first duplication event which gave rise to the plastid-targeted phage-type RNAP and when a third copy was generated to supply a dual-targeted RNAP as found in dicotyledonous plants is unknown.

To gain first insights into earlier phases of the molecular evolution of plant *RpoT* genes, we investigated *RpoT* genes from the moss *Physcomitrella patens*. Bryophytes are among the oldest land plants (Miller, 1984), making them ideal candidates for this evolutionary question. Furthermore, in *Physcomitrella* homologous recombination is almost as efficient as in *Saccharomyces cerevisiae*, opening the doors for using reverse genetics in a plant as routine technique (for a review see Reski, 1998).

Here, we describe the identification of two *RpoT* genes in *Physcomitrella patens*. Sequence analysis and intron positions, on the one hand, indicate that they trace back to the same ancestor gene that has given rise to the *RpoT* gene family of angiosperms. On the other hand, it is evident that the two genes arose from a duplication event that occurred independently from the one giving rise to the plastid-localized polymerase (NEP) of angiosperms. The RNAP encoded by the two *Physcomitrella* genes both are capable of dual-targeting to mitochondria and plastids, with multiple translation initiation as the underlying mechanism.

2. Materials and methods

2.1. Plant material and growth conditions

Physcomitrella patens (Hedw.) B.S.G. were grown axenically as described (Reski et al., 1994) under standard conditions (agitated liquid Knop medium, per liter: 250 mg KH_2PO_4, 250 mg $MgSO_4 \cdot 7H_2O$, 250 mg KCl, 1000 mg $Ca(NO_3)_2 \cdot 4H_2O$, 12.5 mg $FeSO_4 \cdot 7H_2O$, pH 5.8) in a growth chamber under controlled conditions (25 ± 1 °C; light provided from above by two fluorescent tubes, Philips TL-D 36W/25; light flux of 55 μmol s^{-1} m^{-2} outside the flasks, light/dark regime of 16:8 h). Plants were subcultured in 7-day intervals.

2.2. DNA isolation

Moss protonemata were ground to a fine powder under liquid nitrogen and incubated with agitation at 60 °C for 1 h in 1 volume CTAB buffer (2% CTAB, 1.4 M NaCl, 20 mM EDTA, 100 mM Tris, pH 8.0) with 2% β-mercaptoethanol and 100 μg/ml Proteinase K and 0.1 volume *N*-laurylsarcosine buffer (10% *N*-laurylsarcosine, 20 mM EDTA, 100 mM Tris–HCl, pH 8.0), the buffer/plant ratio being at least 2 ml/g. The homogenate was extracted with chloroform/isoamyl alcohol twice and nucleic acids were then precipitated with isopropanol. The pellet was resuspended in 10 mM Tris, 1 mM EDTA and precipitated with 1.2 M NaCl and ethanol. The pellets were resuspended in H_2O and incubated with 50 μg/ml RNase A overnight at 4 °C and subsequently for 1 h at 37 °C.

2.3. RNA isolation

Protonemata were ground to a fine powder under liquid nitrogen and mixed with 3 ml extraction buffer (0.15 M NaCl, 5 mM EDTA, 50 mM Tris–HCl, pH 9.0) and 0.7 ml 10% SDS per gram moss material. The suspension was incubated with 7 ml phenol for 15 min on a rotating platform. After centrifugation at $1600 \times g$ for 10 min the supernatant was again extracted twice with 1 volume phenol/chloroform/isoamylalcohol and once with chloroform. The supernatant was mixed carefully with 1 volume 5 M LiCl and RNA was precipitated overnight at 4 °C. After centrifugation at $4500 \times g$ the pellet was washed twice with 70% ethanol and resuspended in H_2O. All solutions were prepared with DEPC-treated H_2O.

2.4. PCR and library screening

Physcomitrella RpoT genes were identified by PCR on genomic DNA using degenerate primers phydeg-F and phydeg-R (for a list of primers see Table 1) as follows: 95 °C for 2 min, 40 cycles of 30 s at 94 °C, 20 s at 60 °C, 1 min at 72 °C. The PCR products were cloned into vector pGEM-T and sequenced. Derived primers were used for RACE reactions following the protocol of the GeneRacer kit (Invitrogen). cDNA was generated from *Physcomitrella* total RNA using random primers and Omniscript reverse transcriptase (Qiagen) according to the recommendations of the manufacturer. 3′- and 5′-RACE–PCR reactions were performed with the RACE-primers listed in Table 1 using Advantage2 DNA polymerase (Clontech) following the protocol of the manufacturer. Oligonucleotides for genomic PCR were derived from the contig cDNA sequences. The *PpRpoT* genes were amplified from 100 ng DNA using the

Table 1
List of oligonucleotides

Oligonucleotide(s)	Sequence (5′–3′)
Degenerate CODEHOP primers	
phydeg-F	CGAAAGGTGGTGAAGCAGACNGTNATGAC
phydeg-R	GACTCTCGTGCTTGCGGTANGGYTGNAC
RACE (RT–PCR)	
Pp1-3RACE1	GCATTGGGAGAGATGTTCAAAGAAG
Pp1-3RACE1N	GGCTTGGAGACTGTGCGAAGG
Pp1-5RACE1	TTCGCACAGTCTCCAAGCCAGC
Pp1-5RACE1N	ACCTTGGCAGCATAACAGGCA
Pp1-5RACE2	AAACGCTGGACGCAGTTCTGTCA
Pp1-5RACE2N	CGCAGGTGAACGCATAACAGAAG
Pp2-3RACE1	AAGAAGTCCCATAGCTGTTGATCCA
Pp2-3RACE1n	TGGGCATACCGTGAGATGGACA
Pp2-5RACE1	GCGGCTGCAATTATCTTGGCACA
Pp2-5RACE1n	GCATAACAAGCAGCTCGATACGTG
Pp2-5RACE2	GGAGGACGAGGTGGAGGGCTTC
Pp2-5RACE2n	GATTCAGACGCCCTATCTAATGATTG
Genomic PCR	
gPp1-F1	GAACCATCCAAGAGCAGACA
gPp1-R1	GACGATGAAAAGTGCACGAG
gPp1-F2	ACACTAACACGGATGACTCGATAG
gPp1-R2	GTCGGGACTGGTTGACTGT
gPp2-F1	ATAGGCTGAAAGAAAGAAGTCC
gPp2-R1	GACAACTCCTCGTAAGACGC
gPp2-F2	GCACAGCAGCTTGCAAGAC
gPp2-R2	AGAGAATGTGTGCGAATAACAGAT
gPp2-F3	GGCATGACATGGACTGAAGG
gPp2-R3	GTGGAGGGCTTCAAGAATCG
Probe PCR	
pP1pro-F	GAAGAGCGCAATTTCAG
pP1pro-R	CTTCCAACCTTGACTTGTA
GFP constructs	
pP1GFPmet1	CACTCTAGAATGGTAGCAATCGGTGTGTTG
pP1GFPmet2	CACTCTAGAATGTGGAGGGCGGCA
pP1GFPrev	CATGAGTCGACGGCTCGATGGCACAGAATCAGAC
pP1GFPM48I-F	GTGTGAGAGGCGGAATCTGGAGGGCGGC
pP1GFPM48I-R	GCCGCCCTCCAGATTCCGCCTCTCACAC
pP2GFPmet1	CACTCTAGAATGCCAGCTGAGGTCTG
pP2GFPmet2	CACTCTAGAATGTGGAGGTCGGCAG
pP2GFPrev	CAGCGTCGACTCAAGGAGGAAGGGGAG
pP2GFPM36I-F	CAGTCGCCGGCATCTGGAGGTCGGCAG
pP2GFPM36I-R	CTGCCGACCTCCAGATGCCGGCGACTG
Translation constructs	
pP1pre	CCCGGTATGTTGGTTGGAC
pP1atg1	ATGGTAGCAATCGGTGTGTTG
pP1atg2	ATGTGGAGGGCGGCA
pP1atg-R	TCGAGTGGCTGGCAAGTTC
pP2pre	AGAACCAACATCTCGCAG
pP2atg1	ATGCCAGCTGAGGTCTG
pP2atg2	ATGTGGAGGTCGGCAG
pP2atg-R	CGATTGAGCGAAAGGTAG

primers for genomic PCR listed in Table 1. For genomic PCR Qiataq polymerase (Qiagen) was employed. The 5′-part of *PpRpoT1* was obtained by screening a genomic *Physcomitrella* library using a *PpRpoT1* cDNA fragment generated by PCR (primers pP1pro-F and pP1pro-R) as a probe. Screening of the Lambda Fix II library (kindly provided by R. Strepp, Freiburg) and subcloning were performed according to standard protocols. Sequencing of clones and PCR products were done on an ABI373 using dyeterminator chemistry (Applied Biosystems).

2.5. Generation of targeting constructs and transient expression

The RpoT sequences for the full-length transit peptide GFP constructs (*PpRpoT1[M1]GFP* and *PpRpoT2[M1]GFP*) were amplified using primers Pp1GFPmet1 and pP1GFPrev (for RpoT1) and Pp2GFPmet1 and Pp2GFPrev (for RpoT2). The shortened leader peptide sequences, for constructs *PpRpoT1[M48]GFP* and *PpRpoT2[M36]GFP*, were generated using primers Pp1GFPmet2 and Pp1GFPrev (RpoT1) and Pp2met2 and Pp2GFPrev (RpoT2). In the mutated constructs (*PpRpoT1[mut]GFP* and *PpRpoT2[mut]GFP*), amino acid Met48 and Met36, respectively, were substituted by Ile using primers Pp1GFPM49I-F and Pp1GFPM49I-R in combination with primers Pp1GFPrev and Pp1GFPmet1, and accordingly primers Pp2GFPM36I-F and Pp2GFPM36I-R in combination with Pp2GFPrev and Pp2GFPmet, to introduce the base change. PCR products were ligated into vector pGEM-T and cut using *Xba*I and *Sal*I. Fragments were inserted into vector pOL-GFPS65 (Peeters et al., 2000) digested with *Spe*I and *Sal*I. Mitochondrial and plastid targeting control constructs containing coxIV and recA transit peptide sequences were kindly provided by I. Small, INRA (Versaille).

Protoplasts were isolated as described elsewhere (Hohe et al., 2001). 3×10^5 protoplasts were transiently transfected with 30 μg of circular plasmid DNA as described in Strepp et al. (1998). Localization of GFP constructs was analyzed two days after transfection of protoplasts by confocal laser scanning microscopy with a Leica TCS 4D using 488 nm excitation and two-channel measurement of emission from 510 to 580 nm (green/GFP) and > 590 nm (red/chlorophyll).

2.6. In vitro translation

DNA fragments containing 283 bp (*PpRpoT1*) or 62 bp (*PpRpoT2*) of the untranslated 5′-region and 870 bp (*RpoT1*) or 825 bp (*RpoT2*) of the coding region were amplified from *Physcomitrella* cDNA. As controls, constructs starting at the first and second in-frame AUG, respectively, were generated. The PCR products were cloned into vector pGEM-T (Promega). In vitro coupled transcription–translation was performed using the TNT Quick Coupled T7 reticulocyte lysate system (Promega) according to the instructions of the manufacturer. Translation products were separated on 12% SDS–polyacrylamide gels and detected by a Phosphor Imager (BioRad).

2.7. Phylogenetic analysis

Protein sequences were aligned using ClustalW. Based on described conserved blocks for phage-type RNAP (McAllister and Raskin, 1993; Cermakian et al., 1997), regions of high similarity were carefully evaluated. Manual refinement of the ClustalW results gave rise to 12 conserved regions (see Fig. 1). The newly defined blocks extend the domains described by Cermakian et al. (1997), and include additional upstream motifs (blocks I, II). Cut out and merged blocks I–XII were subjected to maximum likelihood analysis using Tree-Puzzle (Strimmer and von Haeseler, 1996).

3. Results

3.1. Molecular cloning of two Physcomitrella RpoT genes

In a first attempt to identify *RpoT* sequences in *Physcomitrella*, PCR using degenerate primers was performed (see Table 1 for an overview of primers). Primers were designed using the CODEHOP procedure (Rose et al., 1998) on the basis of a multiple alignment of the derived amino acid sequences of all known land plant and fungal *RpoT* genes. PCR products spanning a region corresponding to conserved blocks VII–IX (see Fig. 1) were generated and cloned. Sequencing identified two different sequences both homologous to land plant *RpoT* genes. Full length cDNAs were obtained by 5′- and 3′-RACE (rapid amplification of cDNA ends) using derived specific primers. The two *Physcomitrella RpoT* sequences were designated *PpRpoT1* and *PpRpoT2*, respectively. The full-length *PpRpoT1* cDNA sequence (3727 bp) was obtained by aligning fragments produced by RT–PCR and RACE, and constitutes 293 bp of 5′ untranslated leader, the protein coding sequence of 3264 bp, and a 3′ untranslated sequence of 170 bp. The predicted PpRpoT1 protein comprises 1087 amino acids and shows 43–47.7% identity with the three *Arabidopsis* RpoTs. For *RpoT2*, a cDNA of 3403 nt was established by aligning RT–PCR and RACE products. The cDNA contains a 62 bp 5′ leader sequence, 3198 bp of coding sequence, and 143 bp 3′ untranslated sequence. The derived protein is composed of 1065 amino acids and exhibits an identity of 43–45.6% compared with the Arabidopsis RpoT proteins. The two *Physcomitrella* RpoT polypeptides share 59% identity. Fig. 1 shows an alignment of the *Physcomitrella* RpoT polymerases with RpoT2 from *Arabidopsis thaliana* demonstrating the high degree of similarity and the conservation of functionally essential residues and motifs.

To elucidate the gene structure of the two *RpoT* genes, primers were derived from the cDNAs and used for amplification of the genomic sequences. The 5′ part of the *RpoT1* gene was isolated as a Lambda clone from a genomic library. The two *RpoT* genes are 7.3 kb (*RpoT1*) and 8.7 kb (*RpoT2*) in length. *PpRpoT1* consists of 19 exons and 18 introns, *PpRpoT2* of 21 exons and 20 introns. Compared with *Arabidopsis*, *PpRpoT1* contains one and *PpRpoT2* two additional introns in the 5′ part of the gene, intron 16 (as of *Arabidopsis*) is missing in *PpRpoT1* (see Fig. 2). The insertion sites of their common introns are precisely conserved relative to the aligned amino acids sequences between the two *Physcomitrella* genes and those of *Arabidopsis*. All exon–intron junctions contain conserved GT and AG sequences at the 5′-and 3′-ends of the introns, respectively. On average, introns are longer in *PpRpoT2* compared to *PpRpoT1*. We do not have any evidence for the presence of further copies of *RpoT* genes in *Physcomitrella patens*, an observation that is supported by a recent database entry reporting two Physcomitrella *RpoT* cDNA sequences (accession nos. AB055214 and AB055215), which exactly coincide with the coding sequences reported in the present study.

3.2. Subcellular localization of RpoT1 and RpoT2

To determine the subcellular localization of the *PpRpoT1*- and *PpRpoT2*-encoded polypeptides, putative N-terminal transit peptides were fused to the GFP coding region and the constructs were used to transfect *Physcomitrella* protoplasts. The transformed protoplasts were examined by confocal laser scanning microscopy. Both *PpRpoT1* and *PpRpoT2* contain in their 5′-parts two in-frame AUG codons, separated by 144 nt (*PpRpoT1*) and 108 nt (*PpRpoT2*), respectively. Computer analysis of the sequences for organellar targeting properties (Emanuelsson et al., 2000) produced scores far below the classification threshold (0.04–0.22) for the first encoded methionine of each of the sequences, whereas the second encoded methionines produced high scores (0.84–0.91) for mitochondrial localization. Therefore, for each of the two RpoT sequences three different constructs were examined: one starting translation at the first encoded methionine, one starting at the second in-frame AUG, and a construct, in which this second AUG had been mutated (Fig. 3). Along with the RpoT–GFP constructs we used two controls in which GFP targeting could be predicted. Fig. 4 shows the results of these experiments. The panels at left show the GFP (green) channel, those at right show merged images of the green and red (chlorophyll autofluorescence) channels, with yellow color indicating colocalization of the GFP and chlorophyll. In Fig. 4A, showing a GFP construct lacking a specific targeting sequence, the expected non-specific localization in the cytoplasm and the nucleus is seen. A mitochondrial control construct, containing the coxIV transit peptide (Fig. 4B) shows punctuate GFP fluorescence distinct from that of chloroplasts and typical of mitochondrial localization in plant cells (Köhler et al., 1997). Fig. 4C shows a chloroplast control, in which the GFP was fused to a plastid transit peptide (recA), resulting in coincident green and red channel images. All PpRpoT–GFP constructs exhibited the same characteristic subcellular localization: The constructs containing the full-length transit peptide (using the first

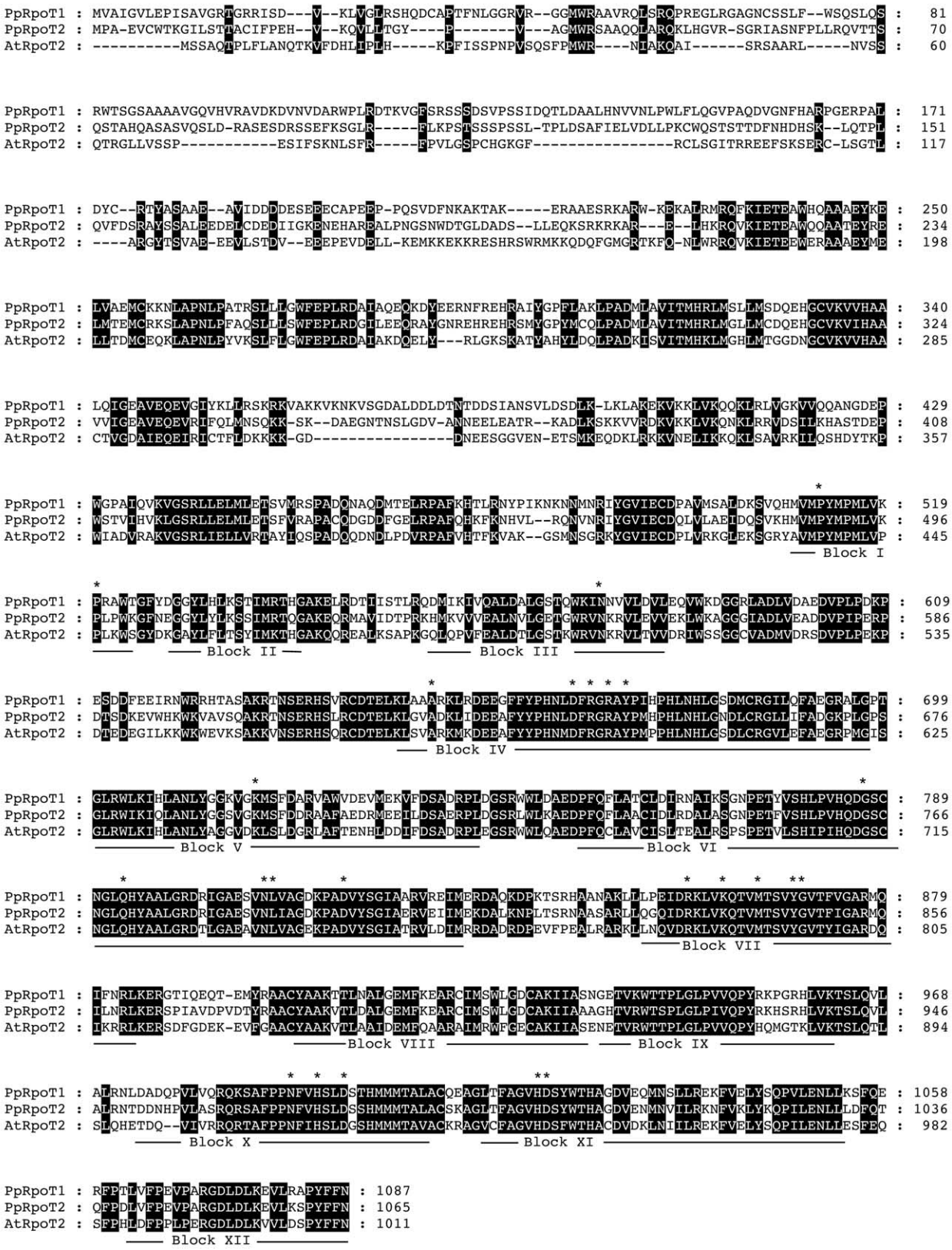

Fig. 1. Amino acid sequence alignment. Amino acid sequences were compared among the two *Physcomitrella* RpoT polymerases and *Arabidopsis* RpoT2 by using the ClustalW algorithm. Sequences with accession numbers: AtRpoT2 (CAC17120); PpRpoT1 (CAC95163); PpRpoT2 (CAC95164). Black lines indicate conserved blocks in the RpoT polymerase family; functionally crucial residues (McAllister and Raskin, 1993; Sousa et al., 1993) are indicated by asterisks.

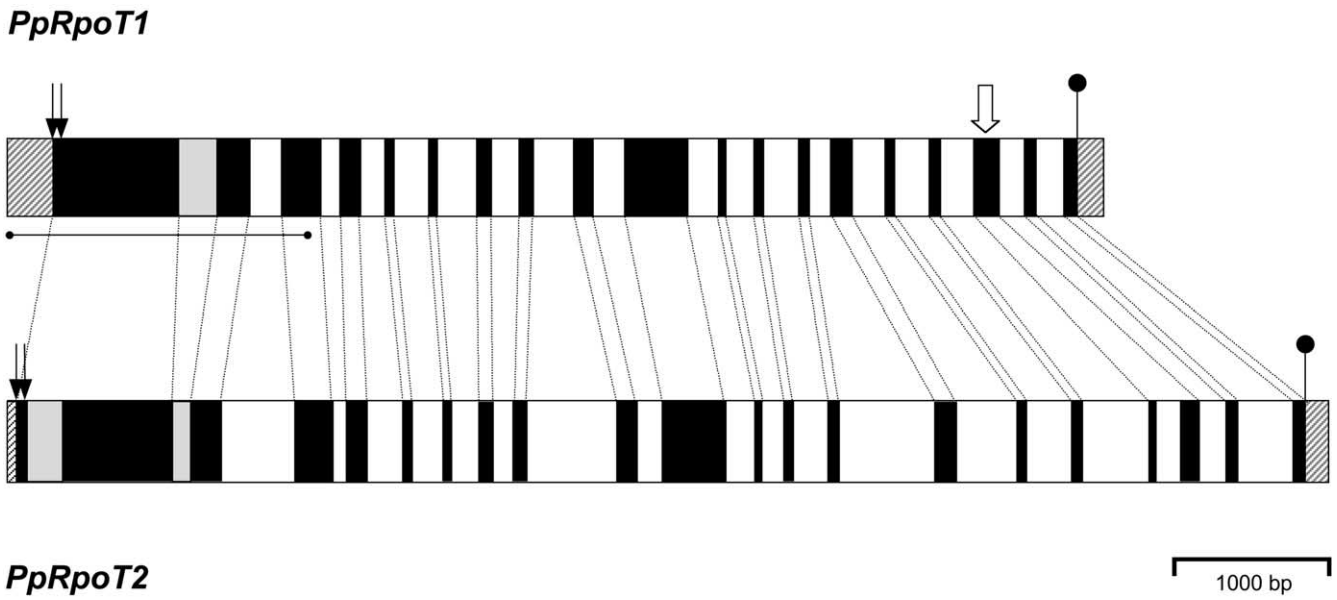

Fig. 2. Molecular map of *PpRpoT1* and *PpRpoT2*. The filled parts indicate the coding regions, with the translation initiation sites (AUG) designated with arrows, stop codons with black circles. The 5′ and 3′ untranslated regions are shown hatched. The thin bar under the *PpRpoT1* map indicates the portion of the gene which was obtained by screening a Lambda library (see Section 2.2). Introns specific for *Physcomitrella* and lacking in *Arabidopsis RpoT* genes are shown in light gray, the position of an intron present in *Arabidosis* and tobacco, but missing in the *Physcomitrella RpoT1* gene is indicated by a white arrow.

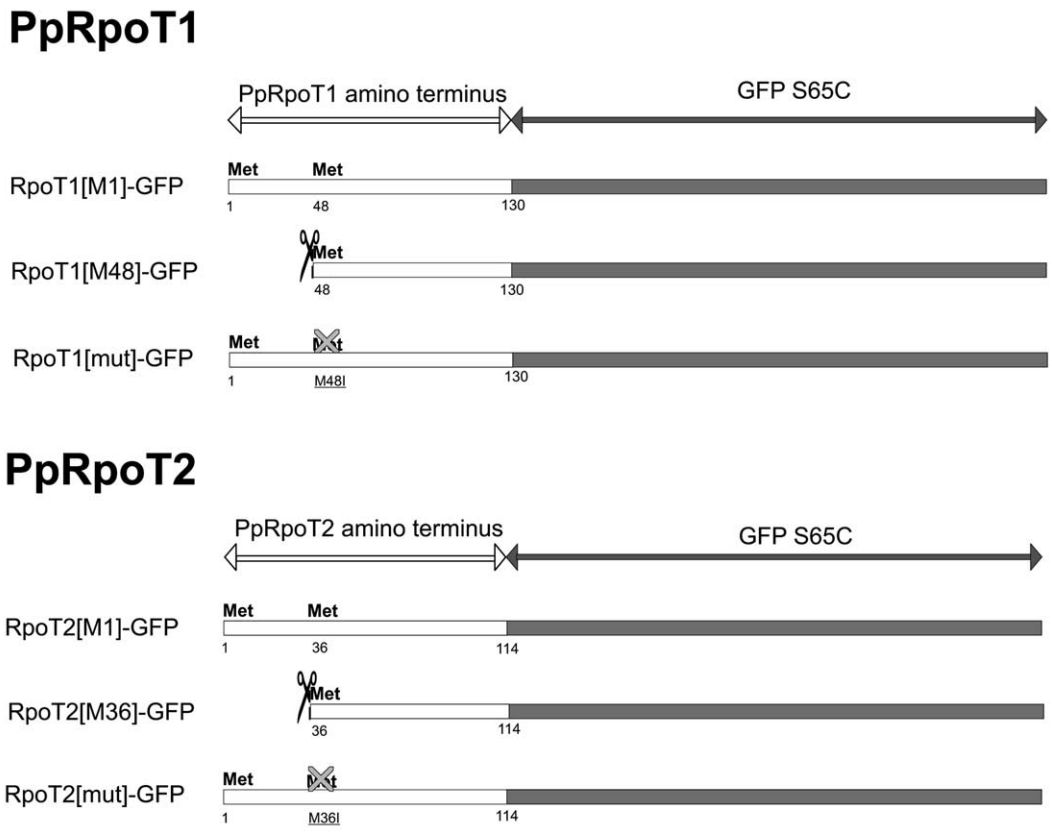

Fig. 3. GFP fusion constructs used for targeting experiments. Amino-terminal RpoT sequences (white bars) were translationally fused to GFP S65C (gray bars) in plasmid pOL (see Section 2.5). The length of the fragments are given by amino acid numbers. The second, in-frame ATG triplet is indicated; the crossed Met position designates the mutation introduced at that position to prevent initiation of translation.

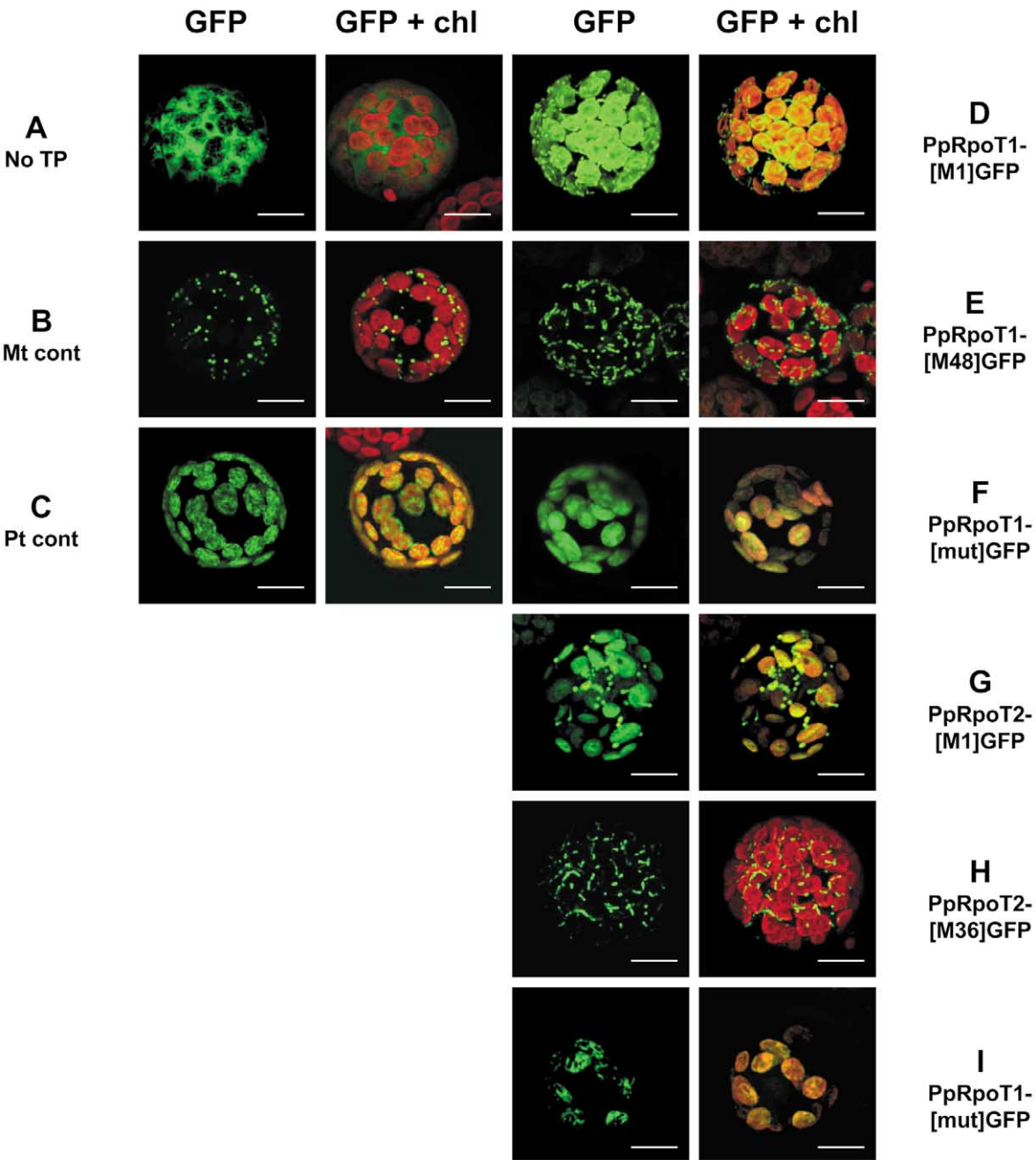

Fig. 4. Subcellular localization of *PpRpoT1* and *PpRpoT2* gene products. Confocal laser scanning microscopy of transformed *Physcomitrella* protoplasts. The first three images depict fluorescence patterns of non-transit peptide (A), mitochondrial (B) and plastid (C) control constructs. Results of targeting experiments using the *Physcomitrella* RpoT-GFP constructs are shown for RpoT1[M1]GFP (D); RpoT1[M48]GFP (E); RpoT1[mut]GFP (F); RpoT2[M1]GFP (G); RpoT2¢M36]GFP (H); and RpoT2[mut]GFP (I). Scale bar: 10 μm.

encoded methionine) clearly exhibit a dual targeting to both chloroplasts and mitochondria as indicated by green fluorescence in both organelles (Fig. 4D,G). In contrast, the shorter constructs lacking the part up to the second, in-frame AUG (Fig. 4E,H) show exclusively mitochondrial targeting, and thus confirm the computer-based prediction. When full-length transit peptides were used with the second methionine mutated (Fig. 4F,I), GFP fluorescence was no longer detectable in mitochondria, but rather localized exclusively to chloroplasts. To test the hypothesis of dual translational starts, we performed in vitro translation of wild-type mRNA, derived from partial cDNAs containing wild-type 5′ upstream sequences beyond the first AUG start codon. As controls, mRNAs with forced translational initiation at the first or second methionine, respectively, were examined. In vitro translation of the synthesized wild-type mRNAs produced two translation products for each of the two *PpRpoT* cDNAs, which coincided in size with the

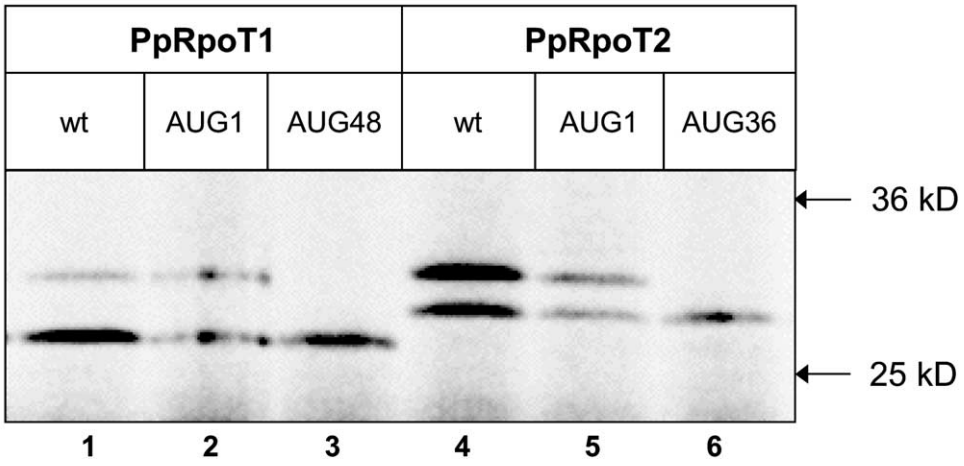

Fig. 5. In vitro translation of *RpoT* mRNAs. ^{35}S-Labeled polypeptides synthesized in a coupled transcription–translation system using cDNAs containing the 5′-untranslated leader (wt, lanes 1,4) and control cDNAs with forced translation from the first (lanes 2,5) or second (lanes 3,6) in-frame AUG. Lanes 1–3, *PpRpoT1*; lanes 3–6, *PpRpoT2*. Molecular masses of marker proteins are given at the right.

products translated by initiation from the first and second AUG start codon, respectively (Fig. 5). These data indicate that translation of the *RpoT* mRNAs in *Physcomitrella* is realized by dual translational starts.

3.3. Phylogenetic analyses

In order to evaluate their phylogenetic relationship, the *Physcomitrella* RpoT proteins were compared to all known nuclear-encoded organellar RpoT polymerases. Following ClustalW alignments, twelve conserved amino acid blocks were dissected and fused to give a length of approximately 425 amino acids which were subjected to tree reconstruction using the maximum likelihood algorithm of Tree-Puzzle (Strimmer and von Haeseler, 1996) (Fig. 6). Essentially the same topology of the phylogenetic tree was obtained using maximum parsimony (PAUP, data not shown). The higher plant RpoT sequences are well separated from the fungal and animal RpoTs encoding mitochondrial polymerases, and form two clearly distinguishable groups: one comprising all exclusively plastid-targeted polymerases described so far, and a second one with sequences which are either mitochondrial or dual-targeted. The two *Physcomitrella* RpoT polymerases form a sister group apart from the two angiosperm groups. This separation reflects rather the evolutionary conservation of the bryophytes which belong to the oldest land plants, than a divergence based on a different ancestor (see also Section 4).

4. Discussion

Two genes encoding phage-type RNA polymerases have been identified in the nuclear genome of the moss *Physcomitrella patens*. The encoded polymerases are highly similar to the organellar RNAP encoded by the *RpoT* gene family in angiosperms. A number of motifs and amino acid positions has been shown to be crucial for the enzymatic activity of phage-type RNA polymerases (McAllister and Raskin, 1993; Sousa et al., 1993). All functional motifs identified in T7 RNA polymerase are conserved in both *Physcomitrella* polymerases. Residues shown to be essential and located within the catalytical pocket of the polymerase (D537, K631, Y639, G640, D812, numbers as given for T7 RNA polymerase) as well as further amino acid residues found to be conserved in the RpoT family (P270, P275, N331, A405, D421, R423, R425, Y427, K472, Q544, N560, L561, D569, R627, M635, N781, H784, D787, H811) are conserved in the deduced polypeptide sequences of *PpRpoT1* and *PpRpoT2* (see Fig. 1). Thus, the derived amino acid sequences are consistent with both *Physcomitrella* RpoT gene products performing RNA polymerase functions.

Comparison of the introns between the two *RpoT* genes reveals a vast variation in length and nucleotide sequences, an indication that both genes have considerably evolved following the gene duplication event. The conservation of intron positions as compared to higher plants *RpoT* genes supports the idea that the ancestor gene giving rise to the phage-type mitochondrial and plastid RNAp of angiosperms originates from a position in the phylogeny of plants that must be placed before the separation of the vascular plant lineage from the bryophytes. On the other hand, the origin of the plastid-localized *RpoTp* genes of higher land plants should be assigned to a duplication which occurred independently of that having given rise to the *Physcomitrella RpoT* sister group and which took place after the separation of the mosses from the branch leading to the flowering plants.

We have used GFP fusion constructs and transformation of *Physcomitrella* protoplasts to trace the subcellular localization of the RNAP encoded by the two genes. Our data clearly show that the transit peptides of both *Physcomitrella RpoT* gene products are capable of targeting the encoded polymerases to both chloroplasts and mitochondria. As the

mutation of the second methionine in RpoT1 and RpoT2 results in an exclusive plastid targeting, it can be proposed that dual targeting is the result of dual translation initiation at the two in-frame AUG codons in both genes. No indication of multiple transcription initiation sites was detected in numerous 5′-RACE reactions (results not shown). The in vitro translation experiments prove that dual targeting of the two *Physcomitrella* RpoT polymerases is the consequence of multiple translational initiation which has been repeatedly reported to occur in other plants (Mireau et al., 1996; Small et al., 1998; Souciet et al., 1999; Watanabe et al., 2001). Dual targeting of RpoT polymerases has been reported so far for the RpoT2 polymerases from *Arabidopsis* and *Nicotiana* (Hedtke et al., 2000, 2002; Kobayashi et al., 2001). There is, however, a remarkable difference in the properties of the transit peptides responsible for the dual targeting in, at least, *Arabidopsis* versus *Physcomitrella*. While showing a similar structure of the extreme N-terminus, i.e. an in-frame AUG codon separated from the first AUG by 100–50 nt, the dual targeting properties of the *Arabidopsis* presequence seem to be caused by the ambiguous nature of the full-length transit peptide, since mutation of the in-frame AUG does not prevent dual targeting (Hedtke et al., 2000). In contrast, mutation of the second methionine in the *Physcomitrella* sequences leads to the loss of mitochondrial targeting, and thus dual targeting in this case seems to be the consequence of dual translational starts. An alignment of the N-termini of RpoT2 from *Arabidopsis* with those of the two *Physcomitrella* RpoT polymerases (see Fig. 1) reveals that the context of the second methionine is conserved between the species, an indication of a functional relevance of the two alternative amino-termini.

The initiation of translation is dependent on the sequence context surrounding the AUG (Kozak, 1991); in plants, purines at positions −3 and +4 relative to the first nucleotide of the AUG codon increase the initiation efficiency (Joshi et al., 1997). Both *PpRpoT1* and *PpRpoT2* each display a pyrimidine at one of the critical positions. Thus, ribosome slippage could be expected to occur, which as a consequence would lead to partial initiation at the second AUG as observed in the in vitro translation assay (Fig. 5). As it has been shown that the translation efficiency of a given AUG context may vary among different cell types (Lukaszewicz et al., 2000), one could speculate that the

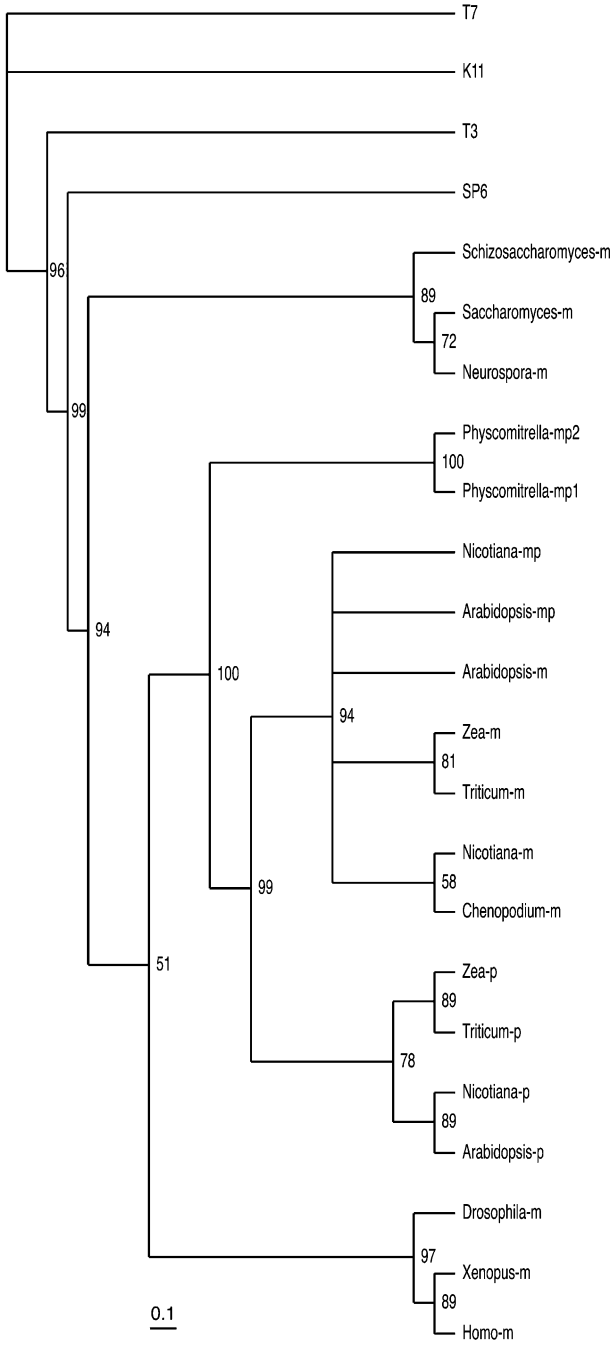

Fig. 6. Phylogenetic analysis of RpoT sequences. Maximum likelihood tree of RpoT protein sequences based on an alignment of conserved blocks (see Section 2.7). Sequences and their accession numbers (-m designates mitochondrial enzyme, -p plastid enzyme, -mp enzyme with dual targeting): T7, bacteriophage T7 RNAP (M38308); T3, bacteriophage T3 RNAP (X02981); K11, bacteriophage K11 RNAP (X53238); SP6, bacteriophage SP6 RNAP (Y00105); Saccharomyces-m, *Saccharomyces cerevisiae* (M17539); Neurospora-m, *Neurospora crassa* (L25087); Schizosaccharomyces-m, *Schizosaccharomyces pombe* (T38431); Homo-m, *Homo sapiens* (O00411); Xenopus-m, *Xenopus laevis* (AAF19376); Drosophila-m, *Drosophila melanogaster* (AAF51421); Arabidopsis-m, *Arabidopsis thaliana* RpoT1 (P92969); Arabidopsis-p, *Arabidopsis thaliana* RpoT3 (O24600); Arabidopsis-mp, *Arabidopsis thaliana* RpoT2 (CAC17120); Nicotiana-m, *Nicotiana sylvestris* RpoT1 (AJ416568); Nicotiana-p, *Nicotiana sylvestris* RpoT3 (AJ302020); Nicotiana-mp, *Nicotiana sylvestris* RpoT2 (AJ302019); Chenopodium-m, *Chenopodium album* (CAA69305); Zea-p, *Zea mays* RpoT1 (AAD22977); Zea-m, *Zea mays* RpoT2 (AAD22976); Triticum-m, *Triticum aestivum* RpoT-G (AAF32492);Triticum-p, *Triticum aestivum* RpoT-C (AAB01085); Physcomitrella-mp1, *Physcomitrella patens* RpoT1 (CAC95163); Physcomitrella-mp2, *Physcomitrella patens* RpoT2 (CAC95164). Essentially the same topology was obtained in parsimony analyses. Branch support values from 1000 replicates are indicated at the nodes. The bar denotes 0.1 substitutions per site.

portion of mitochondrial and chloroplast targeted *RpoT* gene products might be subject to regulation in a tissue- or life cycle stages-dependent manner in the moss. It is also interesting to note that dual targeting had been observed at a much lower frequency when tobacco protoplasts were transfected with the *Physcomitrella* GFP fusion constructs (data not shown), with a predominance of mitochondrial targeting. This could be the consequence of co-factors involved in the in vivo recognition of AUG codons missing in the tobacco cells. It could also reflect an altered import mechanism of the moss chloroplasts versus that of tobacco.

Dual-targeted forms of RpoT have so far been shown to be present in two eudicotyledonous genera, *Arabidopsis* and *Nicotiana* (Hedtke et al., 2000, 2002; Kobayashi et al., 2001). The two *Physcomitrella* RpoT polymerases, for which in this study we also demonstrated dual-targeting properties, do not cluster with the group combining mitochondrial and dual-targeted angiosperm RNAP (see Fig. 5), but rather form a sister group to all angiosperm RpoT proteins. The tree topology as shown in Fig. 6 suggests also that the gene duplication leading to the dual-targeted RNAP in angiosperms occurred only after the separation of the dicots from the monocots. The fact that none of the *Physcomitrella* RpoT polymerases is an exclusively plastid-localized enzyme indicates that the event giving rise to the *RpoTp* genes, encoding the plastid NEP, occurred after the separation of bryophytes from the vascular plant lineage. On the other hand, intron positions are conserved between the moss and higher plant *RpoT* genes. The conserved gene structure indicates that the *RpoT* sequences of *Physcomitrella* originate from an ancestor common to all *RpoT* genes in mosses and higher plants. Thus, the relative separation of the *Physcomitrella* RpoT sister group from the higher plants might rather reflect the exceptional phylogenetic position and conservation of the moss than a completely different molecular evolution of its *RpoT* genes. The presence of two *RpoT* genes in *Physcomitrella* might also simply reflect the fact that autopolyploidy, allopolyploidy and amphidiploidy may have contributed to the evolution of bryophytes (Newton, 1984). Recently, it has been shown that in a species that evolved by allotetraploidization, *Nicotiana tabacum*, the *RpoT* genes from both progenitor species have been retained and are functional (Hedtke et al., 2002). The fact that both *Physcomitrella* RpoT polymerases are capable of dual-targeting is of particular interest. Analysis of *PpRpoT* mutants may resolve further questions concerning the dual-targeting of nuclear-encoded RNA polymerases of organelles.

Acknowledgements

The technical assistance of C. Stock and E. Buena Vera is gratefully acknowledged. This work was supported by grants from the Deutsche Forschungsgemeinschaft (SFB 429, WE 1595/6-1) to T.B. and A.W.

References

Börner, T., Hedtke, B., Hess, W.R., Legen, J., Herrmann, R.G., Weihe, A., 1999. Phage-type RNA polymerases in higher plants. In: Argyroudi, J.H., Senger, H. (Eds.). The Chloroplast: From Molecular Biology to Biotechnology, Kluwer Academic Publishers, Dordrecht, pp. 73–78.

Cermakian, N., Ikeda, T.M., Miramontes, P., Lang, B.F., Gray, M.W., Cedergren, R., 1997. On the evolution of the single-subunit RNA polymerases. J. Mol. Evol. 45, 671–681.

Chang, C.C., Sheen, J., Bligny, M., Niwa, Y., Lerbs-Mache, S., Stern, D.B., 1999. Functional analysis of two maize cDNAs encoding T7-like RNA polymerases. Plant Cell 11, 911–926.

Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., 2000. Predicting subcellular localization of proteins based on their N- terminal amino acid sequence. J. Mol. Biol. 300, 1005–1016.

Hedtke, B., Börner, T., Weihe, A., 1997. Mitochondrial and chloroplast phage-type RNA polymerases in Arabidopsis. Science 277, 809–811.

Hedtke, B., Meixner, M., Gillandt, S., Richter, E., Börner, T., Weihe, A., 1999. Green fluorescent protein as a marker to investigate targeting of organellar RNA polymerases of higher plants in vivo. Plant J. 17, 557–561.

Hedtke, B., Börner, T., Weihe, A., 2000. One RNA polymerase serving two genomes. EMBO Rep. 1, 435–440.

Hedtke, B., Legen, J., Weihe, A., Herrmann, R.G., Börner, T., 2002. Six active phage-type RNA polymerase genes in *Nicotiana tabacum*. Plant J. in press.

Hess, W.R., Börner, T., 1999. Organellar RNA polymerases of higher plants. Int. Rev. Cytol. 190, 1–59.

Hohe, A., Schween, G., Reski, R., 2001. Establishment of a semicontinuous bioreactor culture of Physcomitrella patens for mass production of protoplasts. Acta Horticult. 560, 425–428.

Ikeda, T.M., Gray, M.W., 1999. Identification and characterization of T3/T7 bacteriophage-like RNA polymerase sequences in wheat. Plant Mol. Biol. 40, 567–578.

Joshi, C.P., Zhou, H., Huang, X., Chiang, V.L., 1997. Context sequences of translation initiation codon in plants. Plant Mol. Biol. 35, 993–1001.

Kobayashi, Y., Dokiya, Y., Sugita, M., 2001. Dual targeting of phage-type RNA polymerase to both mitochondria and plastids is due to alternative translation initiation in single transcripts. Biochem. Biophys. Res. Commun. 289, 1106–1113.

Köhler, R.H., Zipfel, W.R., Webb, W.W., Hanson, M.R., 1997. The green fluorescence protein as a marker to visualize plant mitochondria in vivo. Plant J. 11, 613–621.

Kozak, M., 1991. Structural features in eukaryotic mRNAs that modulate the initiation of translation. J. Biol. Chem. 266, 19867–19870.

Lang, B.F., Burger, G., O'Kelly, C.J., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M., Gray, M.W., 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature 387, 493–497.

Lukaszewicz, M., Feuermann, M., Jérouville, B., Stas, A., Boutry, M., 2000. In vivo evaluation of the translation initiation codon in plants. Plant Sci. 154, 89–98.

McAllister, W.T., Raskin, C.A., 1993. The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. Mol. Microbiol. 10, 1–6.

Miller, N.D., 1984. Tertiary and quaternary fossils. In: Schuster, R.M. (Ed.). New Manual of Bryology, Vol. 2. Hattori Bot. Lab., Nichinan Miyazaki, Tokyo, pp. 1194–1332.

Mireau, H., Lancelin, D., Small, I., 1996. The same Arabidopsis gene encodes both cytosolic and mitochondrial alanyl-tRNA synthetases. Plant Cell 8, 1027–1039.

Newton, M.E., 1984. The cytogenetics of bryophytes. In: Dyer, A.F., Duckett, J.G. (Eds.). The Experimental Biology of Bryophytes, Academic Press, London, pp. 65–96.

Peeters, N.M., Chapron, A., Giritch, A., Grandjean, O., Lancelin, D., Lhomme, T., Vivrel, A., Small, I., 2000. Duplication and quadruplica-

tion of *Arabidopsis thaliana* cysteinyl- and asparaginyl-tRNA synthetase genes of organellar origin. J. Mol. Evol. 50, 413–423.

Reski, R., 1998. Molecular genetics of Physcomitrella. Planta 208, 301–309.

Reski, R., Faust, M., Wang, X.-H., Wehe, M., Abel, W.O., 1994. Genome analysis of the moss *Physcomitrella patens* (Hedw.) B.S.G. Mol. Gen. Genet. 244, 352–359.

Rose, T.M., Schultz, E.R., Henikoff, J.G., Pietrokovski, S., McCallum, C.M., Henikoff, S., 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly-related sequences. Nucleic Acids Res. 26, 1628–1635.

Small, I., Wintz, H., Akashi, K., Mireau, H., 1998. Two birds with one stone: genes that encode products targeted to two or more compartments. Plant Mol. Biol. 38, 265–277.

Souciet, G., Menand, B., Ovesna, J., Cosset, A., Dietrich, A., Wintz, H., 1999. Characterization of two bifunctional Arabdopsis thaliana genes coding for mitochondrial and cytosolic forms of valyl-tRNA synthetase and threonyl-tRNA synthetase by alternative use of two in-frame AUGs. Eur. J. Biochem. 266, 848–854.

Sousa, R., Chung, Y.J., Rose, J.P., Wang, B.C., 1993. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. Nature 364, 593–599.

Strepp, R., Scholz, S., Kruse, S., Speth, V., Reski, R., 1998. Plant nuclear gene knockout reveals a role in plastid division for the homologue of the bacterial cell division protein FtsZ, an ancestral tubulin. Proc. Natl. Acad. Sci. USA 95, 4368–4373.

Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964–969.

Watanabe, N., Che, F.S., Iwano, M., Takayama, S., Yoshida, S., Isogai, A., 2001. Dual targeting of spinach protoporphyrinogen oxidase II to mitochondria and chloroplasts by alternative use of two in-frame initiation codons. J Biol. Chem. 276, 20474–20481.