Curr Genet (1995) 28: 128–130

ORIGINAL PAPER

Karsten Liere · Maya Kestermann · Ulrike Müller
Gerhard Link

Identification and characterization of the *Arabidopsis thaliana* chloroplast DNA region containing the genes *psbA*, *trnH* and *rps19′*

Received: 21 November 1994/23 January 1995

Abstract A 1887-nucleotide chloroplast-DNA region from *Arabidopsis thaliana* was analyzed. It contains the conserved genes *psbA* for the precursor of the D1 reaction-centre protein of photosystem II, *trnH* for $tRNA^{His}$, and *rps19′* for the 6.8-kDa protein of the small ribosomal subunit. Northern hybridization and RNase protection experiments suggest co-transcription of a minor RNA fraction over the full lengths of *psbA* and the preceding *trnK*-UUU gene, but not including downstream *trnH* sequences. In front of the mapped 5′ end of the major 1.2-kb *psbA* transcript is a DNA region that shows the typical architecture of a *psbA* promoter, consisting of the prokaryotic-type '−35' and '−10' elements as well as the eukaryotic-type 'TATA' motif. The common 3′ end of *psbA* transcripts seems to be located immediately after a stem-loop structure downstream from the coding region.

Key words *Arabidopsis thaliana* · Chloroplast genes · *trnK* and *psbA* co-transcription

Introduction

The small crucifer *Arabidopsis thaliana* has become a favoured species for plant developmental genetics. Its advantages include a small chromosome number, a low complexity of the nuclear genome, and the availability of well-defined mutants (Meyerowitz 1989). However, despite its considerable contribution to the total coding capacity of the cell, surprisingly little attention has so far been given to the *Arabidopsis* chloroplast genome. Here we have analyzed a chloroplast DNA region that contains the *psbA* gene.

K. Liere · M. Kestermann · U. Müller · G. Link (✉)
Plant Cell Physiology and Molecular Biology, University of Bochum, Universitätsstrasse 150, D-44780 Bochum, Germany

Communicated by H. Kössel

Materials and methods

Plant material. Seedlings of *A. thaliana* (L.) Heynh. var. Landsberg erecta were grown on soil under greenhouse conditions for 6 weeks.

DNA analysis. In order to analyze the *psbA* DNA sequence, libraries of *PstI*/*SmaI* and *PstI*/*EcoRI* fragments of isolated *A. thaliana* DNA were constructed, from which *psbA*-containing clones were identified by colony-filter hybridization. Clones that contained the genes *trnH* and *rps19′* were screened from the *PstI*/*EcoRI* library by using a mustard *trnH*-specific probe. The resulting clones were subcloned and sequenced (Sambrook et al. 1989).

RNA, Northern hybridization, RNase protection assays. Total RNA from *A. thaliana* leaves was prepared (Hughes et al. 1987) and analyzed by Northern hybridization (Sambrook et al. 1989) with an *A. thaliana psbA*-specific in vitro-transcribed cRNA probe. RNase-protection mapping of in vitro transcripts was done according to Sambrook et al. (1989).

Results

Using *psbA*-specific probes from mustard (Link and Langridge 1984), Southern blots of DNA from the three Brassicaceae mustard, rape and *Arabidopsis* resulted in very similar hybridization patterns (data not shown), which is reflected by the nucleotide sequence of this region. The sequenced *A. thaliana* region reveals three genes, i.e. for the D1 protein of photosystem II (*psbA*), for $tRNA^{His}$ (*trnH*) and, on the opposite strand and partially overlapping, *rps19′* for a 6.8-kDa polypeptide with similarity to *E. coli* ribosomal protein S19 (*rps19′*) (data not shown, EMBL accession number X79898). In addition, upstream of the *psbA* gene, a small portion of the 3′ exon of the $tRNA^{Lys}$ (UUU) (*trnK*) is represented by the first 13 nucleotides of the sequence. Thus the gene organization of this region (*trnK–psbA- trnH–rps19′*) resembles the situation in mustard (Link 1984).

The *psbA* coding region itself (1062 bp) as well as the deduced amino-acid sequence (353 residues) each have 99% homology to those from mustard. Eighty-five base

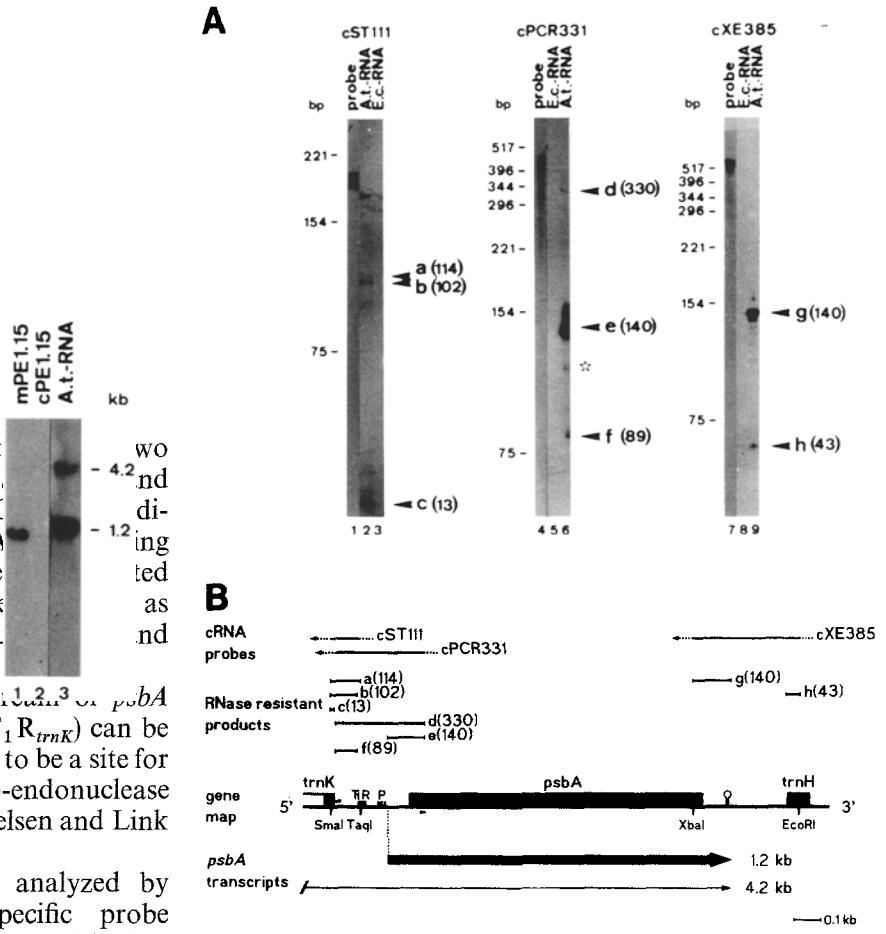

Fig. 2A, B RNase-protection mapping of in vivo transcripts of the *psbA* 5′ and 3′ regions. **A** following RNA–RNA hybridization and treatment with RNases A and T1, resistant products were separated on a 6% denaturing polyacrylamide gel. The probes (shown in **B**) were cST111 (*lanes* 1–3), cPCR331 (*lanes* 4–6) and cXE385 (*lanes* 7–9). *Lanes* 1, 4, 7: ^{32}P-labelled RNA probes without RNase treatment. *Lanes* 3, 5, 8: RNA probes hybridized to 50 µg of *E. coli* tRNA and subsequently treated with RNases. *Lanes* 2, 6, 9: RNase-resistant products following hybridization with 30 µg of total RNA from *A. thaliana*. The *asterisk* denotes a band of unknown origin (*lane* 6). **B** Scheme showing the organization of the *psbA* gene and its flanking regions, including the *trnK* 3′ exon and *trnH*. Coding regions are depicted as *filled boxes*. The *lines* above the gene map indicate the putative position of the protected RNA fragments shown in **A**. *Arrow on top*, ^{32}P-labelled in vitro cRNA probes with *broken lines* representing polylinker and vector sequences. Below the gene map, an alignment by the mapped in vivo *psbA* transcripts, as deduced from the RNase protection and Northern hybridization experiments (Fig. 1)

Fig. 1 Northern hybridization analysis of *psbA* transcripts; 30 µg of total *A. thaliana* RNA were hybridized with the *psbA*-specific RNA probe cPE1.15 (*lane* 3). As controls, 50 ng of unlabelled in vitro transcripts of plasmid pATPE1.15 were used (*lane* 1, mRNA; *lane* 2, cRNA = antisense strand)

pairs upstream of *psbA* is a region that wo
sequence motifs resembling the *E. co* nd
′ − 10′ elements (′TTGACA′ ... ′TATAC di-
tion, a TATA-like element (′TATATAA ing
the eukaryotic-type core promoter seque ted
between the prokaryotic-type sequence as
described in tobacco (Sugita and Sugiu nd
mustard (Link 1984).

In the *trnK* 3′ flanking region upst... of *psbA* a conserved U-rich sequence element (T_1R_{trnK}) can be recognized, which was shown in mustard to be a site for binding and cleavage by a 54-kDa ribo-endonuclease involved in *trnK* 3′-end formation (Nickelsen and Link 1991, 1993).

When total *Arabidopsis* RNA was analyzed by Northern hybridization, the *psbA*-specific probe (cPE1.15) detected two different transcripts with sizes in the range of 4.2 kb and 1.2 kb (Fig. 1, lane 3). While the 1.2-kb transcript is likely to represent the main *psbA* transcript, the 4.2-kb RNA species is of a size sufficient to reflect transcripts that cover both the *trnK* and the *psbA* genes (Nickelsen and Link 1991; Lidholm and Gustafsson 1992).

To locate transcript ends in the 5′ region, RNase protection assays were carried out using ^{32}P-labelled RNA probes cPCR331 and cST111 (Fig. 2 B). The intensive band ′e′ at 140 nt (Fig. 2 A, lane 6) is in agreement with a *psbA* 5′ transcript end 80 nt upstream of the *psbA* coding region, and might represent the 5′ portion of the 1.2-kb main *psbA* transcript. Conserved promoter motifs located directly upstream of this potential 5′ end suggest that this might be a transcription start site as in mustard (Link and Langridge 1984). Protected products of 114 nt (′a′; Fig. 2 A, lane 2) and 330 nt (′d′; Fig. 2 A, lane 6) are of the size expected for in vivo transcripts that cover the entire chloroplast RNA-specific portion of the probes cST111 and cPCR331, respectively (Fig. 2 B). These weak bands thus point to another more-upstream 5′ end and might represent the portion of the 4.2-kb *trnK-psbA* dicistronic transcripts inferred from the Northern experiment (Fig. 1).

The 102-nt fragment ′b′ (Fig. 2 A, lane 2) corresponds to the 89-nt band ′f′ (Fig. 2 A, lane 6) and places an in vivo 3′ end of *trnK* precursors approximately 90 bp downstream from the 3′ exon, in a region containing the conserved sequence element T_1R_{trnK}. The 13-nt band ′c′ (Fig. 2 A, lane 2) most likely correlates to the 3′ *trnK* exon, reflecting a portion of the mature tRNA.

To locate the 3′ ends of *psbA* transcripts, an RNA probe that contained 41 nt of the *psbA* coding region, 307 nt of the *psbA-trnH* intergenic region, and 43 nt of the downstream *trnH* gene, was used (Fig. 2 B, probe cXE385). As shown in Fig. 2 A (lane 9), two RNase-resistant products were generated. The 43-nt band ′n′

coincides with the size of the *trnH* coding region and therefore seems to represent the mature tRNA. Fragment 'g' is 140 nt in length and might reflect the common 3′ end of *psbA* in vivo transcripts, hence placing it 98 nt downstream from the *psbA* coding region.

Discussion

In this work we have focused on the *A. thaliana* chloroplast DNA region containing the *psbA* gene. The nucleotide sequence revealed that this region is composed of the genes *trnK*, *psbA*, *trnH* and *rps19′*. This gene order resembles the situation in mustard, spinach and petunia (Link and Langridge 1984; Zurawski et al. 1984; Aldrich et al. 1988), where this region is located at the junction between the large single-copy region and the right-hand inverted repeat and points to a classification of the chloroplast DNA from *A. thaliana* into group II (Sugiura 1992).

Our Northern and RNase protection experiments revealed a major 1.2-kb *psbA* transcript spanning the entire distance between the presumed transcription start size and the mapped 3′ end (Figs. 1 and 2). The putative transcription start site is suggested to map 80 bp upstream of the *psbA* coding region. The *psbA* promoter architecture, with the elements 'TTGACA' (− 35), 'TATATAA' (TATA-like) and 'TATACT' (− 10) found directly upstream, is well conserved throughout many examined plant species. It has been shown to be critical for promoter activity (Link 1984; Gruissem-Zurawski 1985) and its sequence elements were found to be differentially used in mustard chloroplasts versus etioplasts (Eisermann et al. 1990).

Evidence for co-transcription of *trnK* with photosystem genes was obtained for mustard (Nickelsen and Link 1991), two pine species (Lidholm and Gustafsson 1992), and *Euglena gracilis* (Stevenson and Hallick 1994). Whereas the *Euglena psaA* operon seems to be transcripted exclusively from a single promoter 5′ upstream of the unsplit *trnK* gene, *trnK* and *psbA* co-transcripts in mustard and pine represent only a fraction of the total *psbA* transcripts. The 4.2-kb transcript detected in our Northern experiments (Fig. 1) might reflect *trnK-psbA* co-transcripts, as was confirmed by the RNase mapping experiments (Fig. 2). A number of chloroplast genes have more than one promoter (for a review see Link 1994). Utilization of alternative promoters was shown for the *psbD-psbC* operon from barley and has been viewed as a means for differential regulation of gene expression in response to varying light conditions (Christopher and Mullet 1994).

Acknowledgements We wish to thank A. Ingensand for her expert technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie, FRG.

References

Aldrich J, Cherney BW, Merlin E, Christopherson L (1988) The role of insertions/deletions in the evolution of the intergenic region between *psbA* and *trnH* in the chloroplast genome. Curr Genet 14:137–146

Christopher DA, Mullet JE (1994) Separate photosensory pathways co-regulate blue light/ultraviolet-A-activated *psbD-psbC* transcription and light-induced D2 and CP43 degradation in barley (*Hordeum vulgare*) chloroplasts. Plant Physiol 104:1119–1129

Eisermann A, Tiller K, Link G (1990) In vitro transcription and DNA-binding characteristics of chloroplast and etioplast extracts from mustard (*Sinapis alba*) indicate differential usage of the *psbA* promoter. EMBO J 9:3981–3987

Gruissem W, Zurawski G (1985) Analysis of promoter regions of the spinach chloroplast *rbcl*, *atpB*, and *psbA* genes. EMBO J 4:3375–3383

Hughes JE, Neuhaus H, Link G (1987) Transcript levels of two adjacent chloroplast genes during mustard (*Sinapis alba* L.) seedling development are under differential temporal and light control. Plant Mol Biol 9:355–363

Lidholm J, Gustafsson P (1992) A functional promoter shift of a chloroplast gene: a transcriptional fusion between a novel *psbA* gene copy and the *trnK*(uuu) gene in *Pinus contorta*. Plant J 2:875–886

Link G (1984) DNA sequence requirements for the accurate transcription of a protein-coding plastid gene in a plastid in vitro system from mustard (*Sinapis alba* L.). EMBO J 3:1697–1704

Link G, Langridge U (1984) Structure of the chloroplast gene for the precursor of the Mr 32,000 photosystem II protein from mustard (*Sinapis alba* L.). Nucleic Acids Res 12:945–958

Link G (1994) Plastid differentiation: organelle promoters and transcription factors. In: Nover L (ed) Plant promoters and transcription factors. Springer, Berlin Heidelberg New York, pp 65–85

Meyerowitz EM (1989) Arabidopsis, a useful weed. Cell 56:263–269

Nickelsen J, Link G (1991) RNA-protein interactions at transcript 3′ ends and evidence for *trnK-psbA* co-transcription in mustard chloroplasts. Mol Gen Genet 228:89–96

Nickelsen J, Link G (1993) The 54-kDa RNA-binding protein from mustard chloroplasts mediates endonucleolytic transcript 3′-end formation in vitro. Plant J 3:537–544

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

Stevenson JK, Hallik RB (1994) The *psbA* operon pre-mRNA of the *Euglena gracilis* chloroplast is processed into photosystem I and II mRNAs that accumulate differentially depending on the conditions of cell growth. Plan J 5:247–260

Sugita M, Sugiura M (1984) Nucleotide sequence and transcription of the gene for the 32000-dalton thylakoid membrane protein from *Nicotiana tabacum*. Mol Gen Genet 195:308–313

Sugiura M (1992) The chloroplast genome. Plant Mol Biol 19:149–168

Zurawski G, Bottomley W, Whitfeld PR (1984) Junctions of the large single-copy region and the inverted repeats in *Spinacea oleracea* and *Nicotiana debneyi* chloroplast DNA: sequence of the genes for tRNA His and the ribosomal proteins S19 and L2. Nucleic Acids Res 12:6547–6557