Curr Genet (1987) 11:251–257

Current Genetics



# The chloroplast $tRNA^{Lys}$(UUU) gene from mustard *(Sinapis alba)* contains a class II intron potentially coding for a maturase-related polypeptide

**Heike Neuhaus and Gerhard Link**

Arbeitsgruppe Pflanzliche Zellphysiologie, Ruhr-Universität Bochum, Universitätsstrasse 150, D-4630 Bochum, Federal Republic of Germany

**Summary.** The trnK gene endocing the $tRNA^{Lys}$ (UUU) has been located on mustard (*Sinapis alba*) chloroplast DNA, 263 bp upstream of the psbA gene on the same strand. The nucleotide sequence of the trnK gene and its flanking regions as well as the putative transcription start and termination sites are shown. The 5′ end of the transcript lies 121 bp upstream of the 5′ tRNA coding region and is preceded by procaryotic-type "–10" and "–35" sequence elements, while the 3′ end maps 2.77 kb downstream to a DNA region with possible stem-loop secondary structure. The anticodon loop of the $tRNA^{Lys}$ is interrupted by a 2,574 bp intron containing a long open reading frame, which codes for 524 amino acids. Based on conserved stem and loop structures, this intron has characteristic features of a class II intron. A region near the carboxyl terminus of the derived polypeptide appears structurally related to maturases.

**Key words:** trnK-UUU gene – Chloroplast promoter – Transcript mapping – Class II intron – Long open reading frame

## Introduction

Introns can be assigned to one of the following groups (Sharp 1985; Cech 1986): (1) Introns of nuclear mRNA precursors, which have invariant GU and AG dinucleotides at their boundaries; (2) introns of nuclear tRNA genes, which require the exon sequence and/or exon secondary structure for splicing; (3) mitochondrial and chloroplast genes as well as the rRNA genes of *Tetrahymena* and *Physarum*, which have conserved stem and loop structures within their introns. The latter group can be arranged into two classes, each being defined by the presence of distinct sequence elements and secondary structures (Michel and Dujon 1983) and possibly also by the circular or lariat structure of the excised intron (Cech 1986).

Amongst sequenced chloroplast genes of higher plants, examples for both class I (Steinmetz et al. 1982; Bonnard et al. 1984) and class II introns (Koch et al. 1981; Takaiwa et al. 1982) have been shown. A third class of chloroplast introns with conserved boundary sequences at the 5′ end 3′ splice junctions has been proposed (Shinozaki et al. 1986), which appears to include the intron of the trnK gene for $tRNA^{Lys}$ from tobacco (Sugita et al. 1985).

Here we present the nucleotide sequence of the split trnK gene from mustard (*Sinapis alba*). In addition, we have mapped the transcript and located the putative sites for initiation and termination of transcription. Furthermore, we present features of the intron of this chloroplast gene that were as yet unrealized. Our proposed secondary structure model reveals structural elements typical of class II introns. The trnK intron contains an open reading frame for a derived polypeptide which appears structurally related to maturases.

*Offprint requests to:* G. Link

*Abbreviations:* bp, base pairs; psbA, gene coding for the $M_r$ 32,000 herbicide binding photosystem II protein; COI, cytochrome c oxidase subunit I; trnK, gene coding for chloroplast $tRNA^{Lys}$

## Material and methods

*Plasmids and strains.* Plasmid pSA364, which contains the 12.8 kb fragment Pst5 of mustard chloroplast DNA, has been described (Link 1981; Dietrich and Link 1985). *Escherichia coli* strain JM107 (Yanish-Perron et al. 1985) was used for transformation.

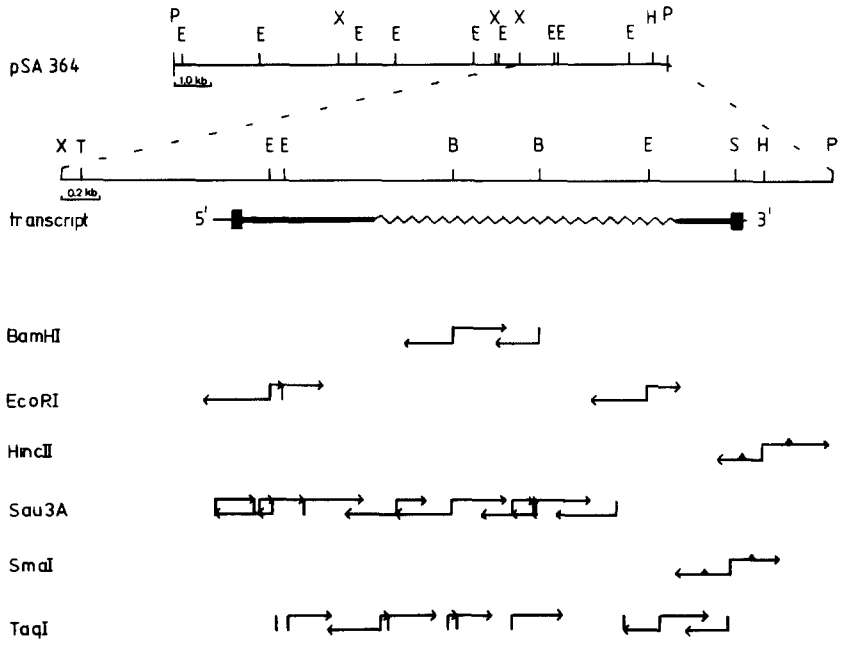


Fig. 1. Map position, structure and orientation of the mustard trnK-UUU gene and transcript, and the sequencing strategy. Physical maps of mustard chloroplast fragment Pst5, the insert of plasmid pSA364 (Link 1981) (*upper line*), and of the region containing the trnK gene (*second line*) are shown. The position and 5′ to 3′ orientation of the trnK transcript are indicated. *Filled boxes* represent the two tRNA coding regions and the *zigzag line* marks the unidentified open reading frame within the trnK intron. Sequencing strategy: *vertical lines* mark the positions of the restriction sites, *horizontal arrows* represent the length and direction of sequences determined by using individual M13 clones. Sequences determined by the chemical technique (Maxam and Gilbert 1977) are marked by *triangles*

*DNA sequencing.* The 3.4 kb SmaI/XhoI region of pSA364 was sequenced by using the dideoxy chain termination method (Sanger et al. 1977) with either alpha $^{32}$P-dATP or $^{35}$S-dATP (Biggin et al. 1983) as radiolabels and M13 mp18/mp19 as vectors (Yanish-Perron et al. 1985). Nucleotide sequence analysis was performed by using the Beckman MicroGenie programs (Queen and Korn 1984).

*Isolation of chloroplast RNA, radioactive labelling and S1 mapping.* Chloroplast RNA was purified as described by phenol and chloroform extraction and DNase I treatment (Link 1982), except for inclusion of isopropanol precipitation to remove low-molecular weight RNA (Maniatis et al. 1982). DNA fragments were fractionated by polyacrylamide gel electrophoresis, eluted, and purified. Fragments were then labelled either at their 5′ ends with gamma $^{32}$P-dATP and T4 polynucleotide kinase or at their 3′ ends, using alpha $^{32}$P-dNTPs and the Klenow fragment of DNA polymerase I. For S1 mapping, 10 ng of labelled DNA fragments were hybridized to chloroplast RNA (50 μg) and digested with 2,000 units nuclease S1 (Berk and Sharp 1977) essentially as described (Link and Langridge 1984). S1-resistant products were coelectrophoresed with pBR322 restriction fragments as size markers (Sutcliffe et al. 1978). In high-resolution mapping experiments, sequence ladders obtained by the dideosy technique were used for calibration. Digestion with EcoRI provided a common end of the labelled DNA strand with the S1 products.

## Results

### *Sequence analysis*

Figure 1 shows the position and physical map of the sequenced chloroplast DNA region as well as details of the sequencing strategy. The sequenced region spans the 2,837 bp from the 3′ SmaI site to a position 368 bp upstream from a 5′ EcoRI site.

To the right of the SmaI site it extends into a previously sequenced region that contains the psbA gene (Link and Langridge 1984). The entire nucleotide sequence preceding the psbA coding region is shown in Fig. 2. This region harbors two sequence stretches, at positions 1–37 and 2,613–2,646, which represent the two portions of a split trnK-UUU coding region (Sugita et al. 1985). The 5′ portion is identical for the mustard and tobacco tRNA coding region, whereas in the 3′ portion of the mustard coding region a G is replaced by A at position 2,632. In both genes a long intron interrupts the tRNA coding region at equivalent positions, two nucleotides 3′ from the anticodon.

Figure 3 shows that the 2,573 nucleotide intron of the mustard trnK-UUU gene has characteristic features of class II introns (Michel and Dujon 1983). These include e.g. a conserved sequence motif "UGCGAC" at the 5′ junction and the hairpin structure with its terminal sequence "GAAA" near the 3′ splice junction.

The mustard trnK intron contains a long open reading frame, which codes for 524 amino acids. Comparison of the derived amino acid sequence with the tobacco sequence (Sugita et al. 1985) reveals overall homology of 66% with a similar degree of conservation along the entire region. Hydropathy analysis (Kyte and Doolittle 1982) shows only two regions which seem to vary with regard to their polarity among the mustard and tobacco sequences, i.e. at amino acid positions 280–300 and 380–400 of the mustard sequence (data not shown). A homology search indicates that a segment near the carboxyl terminus of the derived mustard trnK polypeptide (at amino acid positions 369–471) is structurally related to portions of maturase-like polypeptides of mitochon-

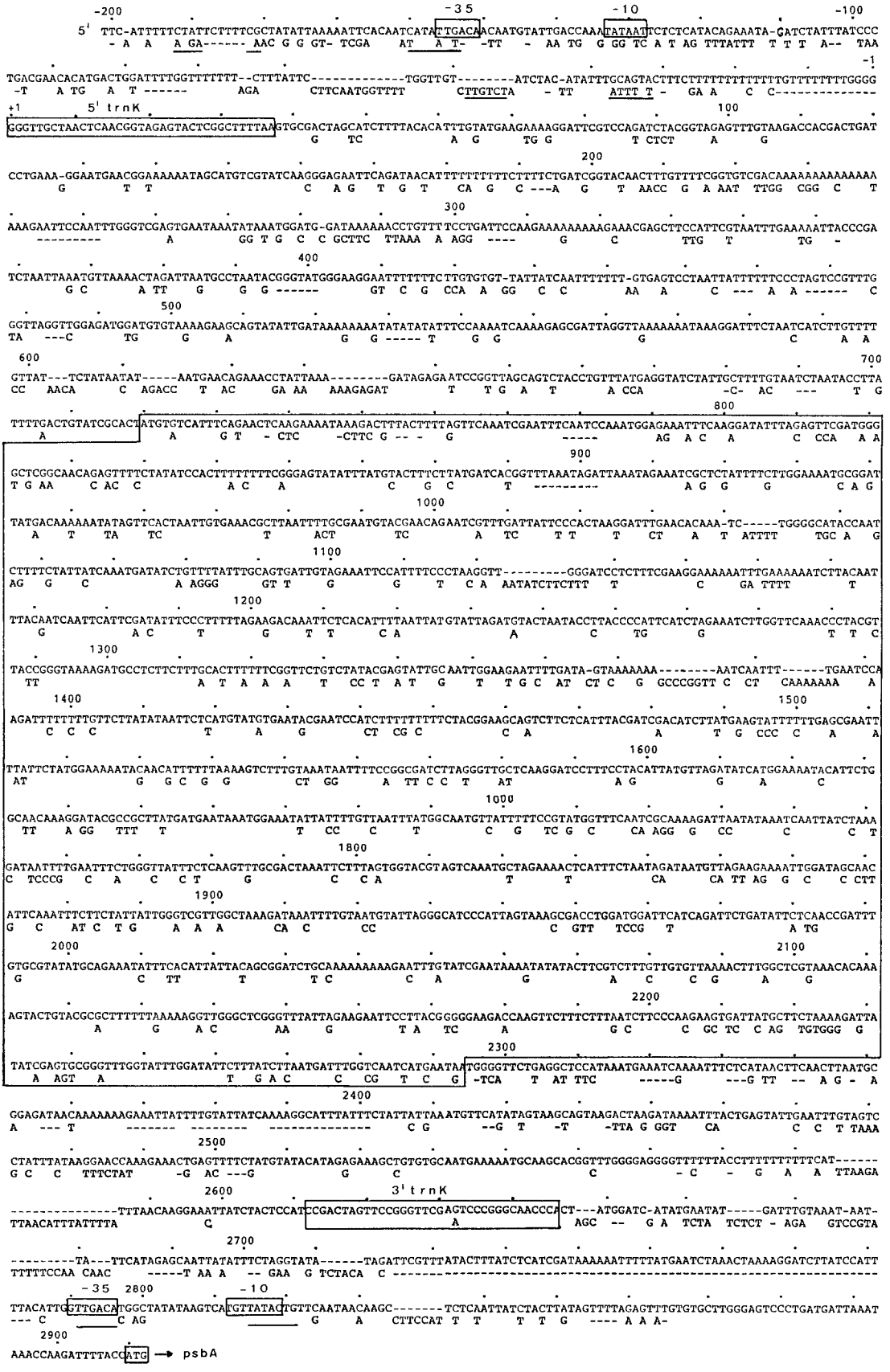


**Fig. 2.** Nucleotide sequence (non-coding strand) of the trnK gene of mustard (*upper line*) in comparison with the homologous gene of tobacco (Sugita et al. 1985). Diverging bases are shown in the *lower line*, the hyphens indicate deleted bases. The mustard trnK coding region, open reading frame and putative "−10" and "−35" promotor elements are *boxed*. The suggested promoter elements of the tobacco trnK gene are *underlined*

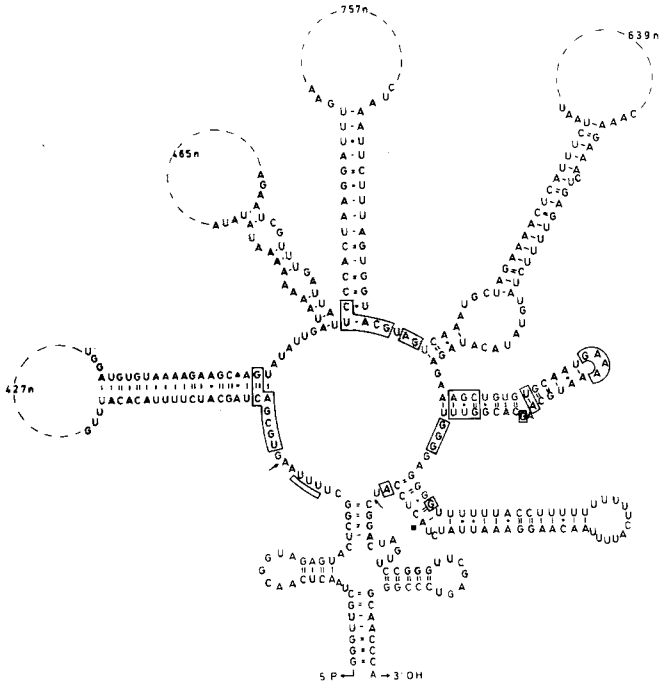

Fig. 3. Secondary structure model of the trnK-UUU intron. The structural arrangement into stems and loops is based on the criteria for class II introns (Michel and Dujon 1983). Conserved sequences are *boxed* and splice points are marked by *arrows*. The *dot* indicates the position of a conserved A, which in yeast mitochondria has been implicated in lariat formation (Van der Veen et al. 1986)

```
S.a.  trnK intron:   369  ENS--FLIDNVRRKLDSNIQ---ISSI
P.a.  COI 1.intron:  518  EDKAYFLGTEIQRISSVKGEIKRFKNI
S.c.  COI 1.intron:  511  ETPARFLGYNISITPLKRMPTV-TKTI
S.c.  COI 2.intron:  534  EGVS-FLGYDVKVTPWEKRPYRMIKKG

S.a.  391  IGSHAKD--KFCNVLGHPISK-----AT-------------
P.a.  546  KGHPQRIP-TTSTYMNAPISKLVTKLADKGIVIWKSKALNE
S.c.  539  RGKTIRSRNTTRPIINAPIRDIINKLATNGYC-KHNK--NG
S.a.  562  DNFIRVRH-HTSLVVNAPIRSIVMKLNKHGYC---SH---G

S.a.  412  ----------WMDSSDSDILNRFVRICRNISHYYSGSA-KK
P.a.  585  DNLIPQPILKWVNLPIRDIILRYKMIWNGYINYYSF-ADNK
S.c.  575  RMGVPTRVGRWTYEEPRTIINNYKALGRGILNYYKL-ATNY
S.c.  595  ILGKPRGVGRLIHEEMKTILMHYLAVGRGIMNYYRL-ATNF

S.a.  442  KNLY-RIKYILRLCCVKTLARKHKS-TVRAFLK
P.a.  625  PRLV-LIYWILRKSLAKTLATKLKLGTVRKVYL
S.c.  615  KRLRERIYYVLYYSCVLTLASKYRLKTMSKTIK
S.c.  635  TTLRGRITYILFYSCCLTLASKFKLNTVKKVIL
```

Fig. 4. Alignment of a region of the derived trnK polypeptide with maturase-like proteins encoded by mitochondrial introns. Amino acid sequences shown are derived from the first intron of the COI gene from *Podospora anserina* (Osiewacz and Esser 1984) and from the first and second intron of the *Saccharomyces cerevisiae* COI gene (Bonitz et al. 1980). Amino acid position 369 of the mustard sequence corresponds to nucleotide position 1,824 in Fig. 2. Homologies, including conservative amino acid replacements, are *boxed;* conservative replacements are marked by *dots*

drial introns (Lazowska et al. 1980) (Fig. 4). This region shows homology with the derived polypeptides of the first intron (amino acid positions 511–648) and second intron (positions 534–662) of the *Saccharomyces cerevisiae* cytochrom c oxidase subunit I gene (Bonitz et al. 1980) and of the first intron (positions 518–648) of the *Podospora anserina* COI gene (Osiewacz and Esser 1984). Inspection of the derived polypeptide of the tobacco trnK gene (Sugita et al. 1985) at amino acid positions 354–456 also reveales such structural relationship (data not shown).

### *5′ and 3′ ends of the trnK transcript*

The 5′ end of the trnK transcript was determined by S1 nuclease mapping of the 5′ labelled 1.1 kb XhoI/EcoRI fragment (see Fig. 1). As shown in Fig. 5A, left panel, lane 2, two S1 resistant fragments with sizes of 290 and 180 bp are generated. To locate the 5′ end of the transcript more precisely, the resistant DNA fragments were electrophoresed alongside the products of dideoxy chain terminating reactions (Sanger et al. 1977). Based on the sizes of S1 resistant bands in Fig.

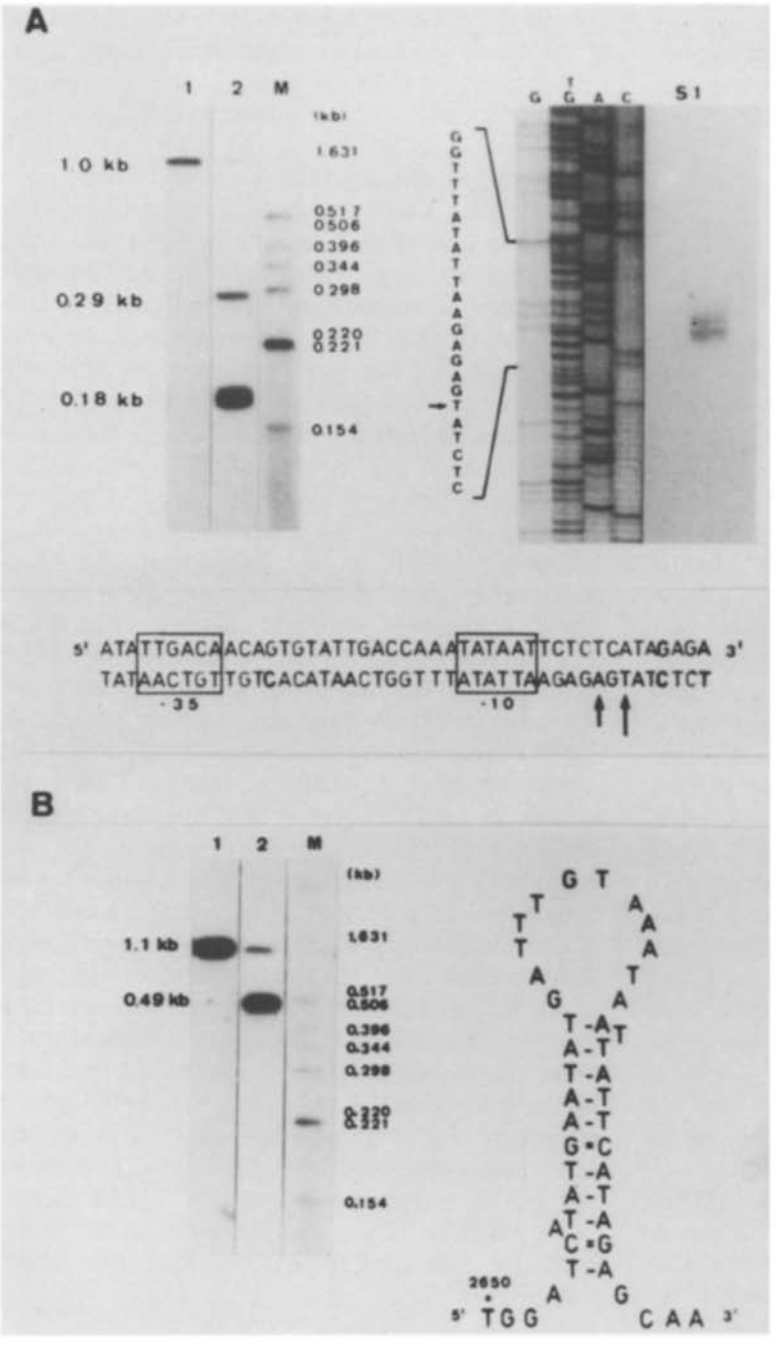


**Fig. 5A, B.** Nuclease S1 mapping of the trnK transcript. **A** Mapping of the 5′ end. *Left panel,* the 1.1 kb XhoI/EcoRI fragment, 5′ labelled at the EcoRI site (*lane 1*); S1-resistant fragments, following hybridization with unlabelled cpRNA and nuclease S1 treatment (*lane 2*). Size markers (*M*) are HinfI fragments of pBR322. *Right panel,* high-resolution S1 mapping of the RNA 5′ end. To measure the length of the S1 resistant products, the 1.0 kb TaqI/EcoRI fragment was cloned in M13mp19 and sequenced by the dideoxy technique (Sanger et al. 1977). The newly sythesized double stranded DNA segments were cut with EcoRI and the sequence ladder (coding strand) was used for calibration. *Lower panel,* nucleotide sequence of the DNA region surrounding the RNA 5 end (*arrows*). The *boxed regions* are the putative "−10" and the "−35" promotor elements. **B** S1 mapping of the RNA 3′ end. *Left panel, lane 1,* the 989 bp EcoRI/PstI fragment, 3′ labelled at the EcoRI site. *Lane 2,* S1 protected 480 bp fragment upon hybridization with chloroplast RNA and S1 treatment. Marker (*M*), HinfI fragments of pBR322. *Right panel,* proposed stem-loop structure of the DNA region at the 3′ end of the trnK transcript (non-coding strand)

5A, right panel, the RNA 5′ end corresponds to nucleotide positions −122 to −124 upstream of the tRNA coding region (see Fig. 2). It appears likely that the upper band at −124 is due to partial S1 digestion and the lower band at the A at −122 represents the authentic 5′ end of the transcript. The additional, 0.18 kb, S1 resistant band in Fig. 5A, left panel, lane 2, corresponds to a cutting position within a long A-T stretch in front of the tRNA coding region and is possibly due to non-specific cleavage by the nuclease (data not shown). It could, however, also reflect an authentic 5′ end of a tRNA processing intermediate.

S1 mapping of the 3′ end of the trnK transcript was performed with the 3′ labelled 989 bp EcoRI/PstI fragment (see Fig. 1). A single S1 resistant band of 490 bp is generated in addition to the full-size material (Fig. 5B, lane 2). This defines the 3′ end of the transcript near nucleotide position 2,650 (see Fig. 2). As shown in Fig. 5B, this region contains sequences capable of forming stem-loop structures, which could serve as termination

signals (Rosenberg and Court 1979). The distance between the mapped 5′ and 3′ end is 2,770 nucleotides, in agreement with 2.8 kb derived from Northern analysis (data not shown).

## Discussion

The present sequence analysis of the mustard trnK upstream region has revealed "–35" and "–10" sequence elements which are identical with the "TTGACA" and "TATAAT" consensus sequences of procaryotic promotors (Rosenberg and Court 1979; Hawley and McClure 1983). Available evidence suggests that many chloroplast genes have typical procaryotic-type promoters (Whitfeld and Bottomley 1983; Kung and Lin 1985), which is in line with the proposed procaryotic origin of these organelles (Margulies 1970). Consensus motifs "TTGACA" for the "–35" and "TATAAT" for the "–10" region have been established (Whitfeld and Bottomley 1983; Kung and Lin 1985; Strittmatter et al. 1985). By using functional assays, it was demonstrated for several chloroplast genes that the upstream region containing the putative promotor elements is required for efficient in vitro transcription (Link 1984; Hanley-Bowdoin et al. 1985; Gruissem and Zurawsky 1985; Bradley and Gatenby 1985; Boyer and Mullett 1986) S1 mapping of the mustard trnK gene has shown that the in vivo 5′ end of the transcript is directly downstream of the "–10" element. It thus appears likely that ths region serves as functional promotor for transcription of the trnK gene. Further proof of the RNA 5′ end as the authentic transcription start site will depend on functional assays (Link 1984) and RNA capping experiments (Strittmatter et al. 1985).

The mustard trnK gene contains a 2,574 bp intron in the anticodon-loop. Sequence comparison with the equivalent $tRNA^{Lys}$ gene from tobacco (Sugita et al. 1985) reveals 73% sequence conservation for the tRNA coding region and the intron, but only 50% for the 5′ and 3′ non-coding regions. Within the intron, sequence stretches with up to 90% homology are found (e.g. positions 2,515–2,584). Sequence elements within these stretches are parts of the core and six base-paired stem sequences typical for the secondary structure of class II introns (Michel and Dujon 1983; Keller and Michel 1985). The folding of the intron might facilitate a transesterification event between a conserved adenine at position 2,605 near the 3′ end of the guanine at position 38, leading to the cleavage of the 5′ exon-intron boundary and lariat formation (Van der Veen et al. 1986). The tobacco trnK intron was classified as class III intron (Shinozaki et al. 1986), based on conserved splice junctions resembling those of nuclear mRNA precursors (Cech 1983). Our data suggest that both junction sequences and secondary structure are conserved freatures of the mustard trnK intron, which seem to justify classification as a class II intron. Similar structural is features are also noticeable within the tobacco trnK intron (not shown).

The derived amino acid sequence of the long open reading frame, which in case of the mustard gene has 524 and in case of the tobacco gene 509 amino acids, shows 66% homology among both species. Structural conservation is reflected by the hydropathy profile (Kyte and Doolittle 1982) of the two amino acid sequences, indicating a similar pattern of alternating hydrophilic and hydrophobic regions (data not shonw).

Although many mitochondrial and chloroplast introns contain open reading frames, only fewintron encoded proteins have been tentatively identified. The 2,295 bp class I intron of the *Neurospora crassa* mitochondrial 24S rRNA gene (Burke and RajBhandary 1982) contains a reading frame of 426 amino acids, which has been related to ribosomal protein S5. Several mitochondrial class I and class II introns are known to encode a maturase, which is involved in the splicing process of its own precursor (Lazowska et al. 1980; Weis-Brummer et al. 1983). The open reading frames of four mitochondrial introns (Bonitz et al. 1980; Osiewacz and Esser 1984; Lang et al. 1985) seem to encode a maturase and, in addition, contain putative coding capacity for a reverse transcriptase (Michel and Lang 1985). Among known chloroplast introns only the open reading frame of the 23S rRNA class I intron in *Chlamydomonas reinhardtii* (Rochaix et al. 1985) was found to be structurally related to mitochondrial reading frames that potentially code for maturases. It is interesting to note that the derived amino acid sequence of the mustard trnK reading frame has a stretch of residues near the C-terminal end which appears structurally related to maturases. Thus, both a chloroplast class I and class II intron are now known to bear this relationship. Further experiments will show if a maturase-like activity is in fact involved in splicing of these and possibly other chloroplast introns.

*Acknowledgements.* We are grateful to Claudia Wittig and Bärbel Lipka for skilled technical assistance. We thank Andrea Scholz for initial help and discussion, T. Karsch and W. Oettmeier for providing their hydropathy analysis programs, and J. Hughes and U. Kück for critical reading of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (Schwerpunkt Molekularbiologie der Pflanzen) and the Fonds der Chemischen Industrie, FRG.

## References

Berk AJ, Sharp PA (1977) Cell 12:721–732
Biggin MD, Gibson TJ, Hong GF (1983) Proc Natl Acad Sci USA 80:3963–3965

Bonitz SG, Coruzzi G, Thalenfeld BE, Tzagoloff A (1980) J Biol Chem 255:11927–11941
Bonnard G, Michel F, Weil JH, Steinmetz A (1984) Mol Gen Genet 194:330–336
Boyer SK, Mullett JE (1986) Plant Mol Biol 6:229–244
Bradley D, Gatenby AA (1985) EMBO J 4:3641–3648
Burke JM, RajBhandary UL (1982) Cell 31:509–520
Cech T (1983) Cell 34:713–716
Cech T (1986) Cell 44:207–210
Dietrich G, Link G (1985) Curr Genet 9:683–692
Gruissem W, Zurawski G (1985) EMBO J 4:3375–3383
Hanley-Bowdoin L, Orozco EM, Chua N-H (1985) Mol Cell Biol 5:2733–2745
Hawley D, McClure WR (1983) Nucleic Acids Res 11:2237–2255
Keller M, Michel F (1985) FEBS Lett 179:69–73
Koch E, Edwards K, Kössel H (1981) Cell 25:203–213
Kung SD, Lin CM (1985) Nucleic Acids Res 13:7543–7549
Kyte J, Doolittle RF (1982) J Mol Biol 157:105–132
Lang BF, Ahne F, Bonen L (1985) J Mol Biol 184:353–366
Lazowska J, Jacq C, Slonimski PC (1980) Cell 22:333–348
Link G (1981) Nucleic Acids Res 15:3681–3694
Link G (1982) Planta 154:81–86
Link G (1984) EMBO J 3:1697–1704
Link G, Langridge U (1984) Nucleic Acids Res 12:945–957
Maniatis T, Fritsch EF, Sambrook J (eds) (1982) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Margulies L (ed) (1970) Origin of eukaryotic cells. Yale University Press, New Haven
Maxam AM, Gilbert W (1977) Proc Natl Acad Sci USA 74:560–564
Michel F, Dujon B (1983) EMBO J 2:33–38
Michel F, Lang BF (1985) Nature (London) 316:641–643
Osiewacz HD, Esser K (1984) Curr Genet 8:299–305
Przybyl D, Fritzsche E, Edwards K, Kössel H, Falk H, Thompson JA, Link G (1984) Plant Mol Biol 3:147–158
Queen C, Korn LJ (1984) Nucleic Acids Res 12:581–599
Rochaix JD, Rahire M, Michel F (1985) Nucleic Acids Res 13:975–984
Rosenberg M, Court D (1979) Annu Rev Genet 13:319–353
Sanger F, Nicklen S, Coulson ARC (1977) Proc Natl Acad Sci USA 74:5463–5467
Sharp P (1985) Cell 42:397–400
Shinozaki K, Deno H, Sugita M, Kuramitsu S, Sugiura M (1986) Mol Gen Genet 202:1–5
Steinmetz A, Gubbins EJ, Bogorad L (1982) Nucleic Acids Res 10:3027–3037
Strittmatter G, Gozdzicka-Josefiak A, Kössel H (1985) EMBO J 4:599–604
Sugita M, Shinozaki K, Sugiura M (1985) Proc Natl Acad Sci USA 82:3557–3561
Sutcliffe JG (1978) Nucleic Acids Res 5:2721–2728
Takaiwa F, Sugiura M (1982) Nucleic Acids Res 10:2665–2676
Van der Veen R, Arnberg AC, Van der Horst G, Bonen L, Tabak HF, Grivell LA (1986) Cell 44:225–234
Weis-Brummer B, Holl J, Schweyen RJ, Rodel G, Kaudewitz F (1983) Cell 33:195–202
Whitfeld PR, Bottomley W (1983) Annu Rev Plant Physiol 34:279–310
Yanisch-Perron C, Vieira J, Messing J (1985) Gene 33:103–119

Communicated by H. Kössel

Received July 9, 1986