

# Three new nuclear genes, *sigD*, *sigE* and *sigF*, encoding putative plastid RNA polymerase $\sigma$ factors in *Arabidopsis thaliana*

Makoto Fujiwara<sup>1</sup>, Akitomo Nagashima, Kengo Kanamaru, Kan Tanaka, Hideo Takahashi\*

*Institute of Molecular and Cellular Biosciences, The University of Tokyo, Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan*

Received 21 June 2000; revised 14 August 2000; accepted 14 August 2000

Edited by Gunnar von Heijne

**Abstract** Three new nuclear genes (*sigD*, *sigE* and *sigF*) of *Arabidopsis thaliana*, encoding putative plastid RNA polymerase  $\sigma$  factors, were identified and analyzed. Phylogenetic analysis revealed that higher plant  $\sigma$  factors fell into at least four distinct subgroups within a diverse protein family. In addition, *Arabidopsis sig* genes contained conserved chromosomal intron sites, indicating that these genes arose by DNA duplication events during plant evolution. Transcript analyses revealed two alternatively spliced transcripts generated from the *sigD* region, one of which is predicted to encode a  $\sigma$  protein lacking the carboxy-terminal regions 3 and 4. Finally, the amino-terminal sequence of the *sigF* gene product was shown to function as a plastid-targeting signal using green fluorescent protein fusions. © 2000 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.

**Key words:**  $\sigma$  factor; Multigene family; Chloroplast RNA polymerase; Transit peptide

## 1. Introduction

Plant plastids have their own genomes containing up to 150 genes that encode rRNA, tRNA and proteins for photosynthesis, translation, transcription and so on [1]. Expression of plastid genes is essential for organelle functions, such as photosynthesis in chloroplasts. Transcription in higher plant plastids is directed by at least two distinct RNA polymerases [2,3]. One enzyme is the T7 bacteriophage-type RNA polymerase, called the nuclear-encoded polymerase that transcribes housekeeping and some photosynthetic genes [2–4]. Another enzyme is the eubacteria-type RNA polymerase, called plastid-encoded polymerase (PEP). It is composed of plastid-encoded core enzyme subunits ( $\alpha$ ,  $\beta$ ,  $\beta'$  and  $\beta''$ ) and a  $\sigma$  factor, a specificity protein required for recognition of DNA promoter elements [5,6]. It has been established that PEP transcribes most photosynthesis genes and plays an essential role for chloroplast development [7–9].

Previously, we identified and analyzed nuclear-encoded chloroplast  $\sigma$  factor genes from a red alga *Cyanidium caldarium* [10,11] and a higher plant *Arabidopsis thaliana* [12]. Since then, with help from the recent progress in higher plant genome analyses, a number of putative  $\sigma$  factors have been identified from various plant species [13–19]. Because  $\sigma$  fac-

tors should be essential regulators for PEP activity, it is important to analyze the structure and function of plant  $\sigma$  factors for the understanding of the plastid transcription system. Emerging information from the *Arabidopsis* genome project revealed three new nuclear genes for  $\sigma$  factor proteins.

## 2. Materials and methods

### 2.1. Plant materials and growth conditions

*A. thaliana* Columbia was grown on Jiffy-7 (AS Jiffy Products) or Murashige–Skoog agar medium [20] at 22°C under continuous white light [21].

### 2.2. Oligonucleotides, plasmids and PCR

*Arabidopsis* DNA and RNA were extracted by DNeasy Plant Mini Kit (Qiagen) and RNeasy Plant Mini Kit (Qiagen), respectively. Pl clones were provided by Dr. Tabata.

cDNAs of *sigD*, *sigE* and *sigF* were amplified by PCR from an *Arabidopsis* cDNA library [12]. Two overlapping *sigD* cDNA fragments, one amplified between D4-N2 (5'-GACCAAAAAGAAGATAGATAAC(-3') and D4-TP-C TCTCCATGGACTGAATCCC-CATTTCTC, and the other between D4-NdeI CCCACATATGTCA-GAAAAGAGAAGGAAGAG and D4-BamHI GAGGATCCAATT-TATTTGCTTTTATTTCAC, a *sigE* cDNA between D5-TP-N1 TGGGATCCCAAGATTGTGATGGGAGTTGTG and D5-BamHI3' GGGGATCCAGTTCGGTTCGGTTTGGT, and a *sigF* cDNA between D6-SalI CTTGTGCGACTTTGGTGATTCTGTGAAATGG and D6-EcoRI TGAATTCAGAAACCAATATCTCCATGAGC, were each cloned into pBluescript vectors (Stratagene). RT-PCR was performed with RT-PCR kit (Toyobo, Japan). 5  $\mu$ g of total RNA was used for each reaction. Oligonucleotides used were: D4#1-1 (for *sigD*) 5'-AATCTAGACTTGTCAAATTGCCGGG, D4#1-2 (for *sigD*\*) AATCTAGACTTGTCAAATTGCCGCT and D4#2 CTTCAAACGACATAGGAGTCTCTC.

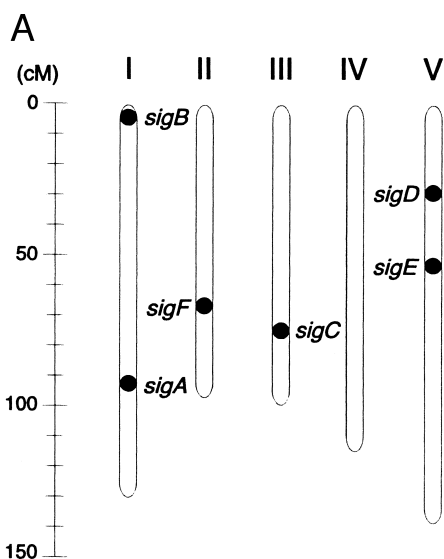
To examine the subcellular localization of SigF, green fluorescent protein (GFP) was used as a reporter [14,22]. A cDNA fragment corresponding to the amino-terminal 80 amino acids of SigF was amplified between D6-SalI and D6-NcoI ACACCATGGATGTTCTATCGTCTCTTA, and introduced into *SalI*–*NcoI* sites of CaMV35S-sGFP(S65T)-NOS vector [14] (provided by Dr. Niwa) to yield pF-TP-GFP.

### 2.3. Phylogenetic analysis

Sigma factor sequences were obtained from GenBank database and CyanoBase (<http://www.kazusa.or.jp/cyano/>), and aligned with CLUSTAL W version 1.7 [23]. Amino acid positions of ambiguous alignment were omitted from subsequent phylogenetic analyses. The phylogenetic tree was constructed by the neighbor-joining method [24] with PHYLIP package version 3.5c (J. Felsenstein, University of Washington). Bootstrap analyses for 1000 replicates were performed to provide confidence estimates for tree topologies. The analyzed proteins are as follows: *A. thaliana* SigA (D89993), SigB (AB004293), SigC (D89994), SigD (AB021119), SigE (AB021120), SigF (AB029916); *Sinapis alba* Sig1 (Y15899), Sig2 (AJ276656), Sig3 (AJ276657); *Nicotiana tabacum* SigA1 (AB023571), SigA2 (AB023572); *Oryza sativa* Os-SigA (AB005290); *Triticum aestivum* SigA (AJ132658); *Zea mays* Sig1 (AF058708 [17]), Sig2 (AF058709 [17]), Sig1 (AF099110 [19]), Sig2 (AF099111 [19]), Sig3 (AF099112 [19]); *Sorghum bicolor* Sig1 (Y14276); *Cyanidium caldarium* SigA

\*Corresponding author. Fax: (81)-3-5841 8476.  
E-mail: htakaha@imcbns.iam.u-tokyo.ac.jp

<sup>1</sup> Present address: Plant Functions Laboratory, Institute of Physical and Chemical Research (RIKEN), Wako-shi, Saitama 351-0198, Japan.



## B

[illegible]



Fig. 2. Phylogenetic analysis of eubacterial and plant  $\sigma$  factors. The tree was constructed by the neighbor-joining method [24] on the basis of 195 aligned residues corresponding to the conserved regions 2.1–4.2. *E. coli* RpoD was used as the outgroup. The scale bar represents 0.1 substitutions per amino acid position. Numbers at nodes indicate bootstrap values out of 1000 bootstrap resamplings. Abbreviations: At, *A. thaliana*; Sa, *S. alba*; Os, *O. sativa*; Ta, *T. aestivum*; Zm, *Z. mays* (to avoid confusion of the same designations for the different genes, gene products are marked with reference number [17] or [19]); Sb, *S. bicolor*; RK1, *C. caldarium* RK-1; Gal, *G. sulphuraria*; 7002, *Synechococcus* PCC7002; 7942, *Synechococcus* PCC7942; 6803, *Synechocystis* PCC6803; Ec, *E. coli*.

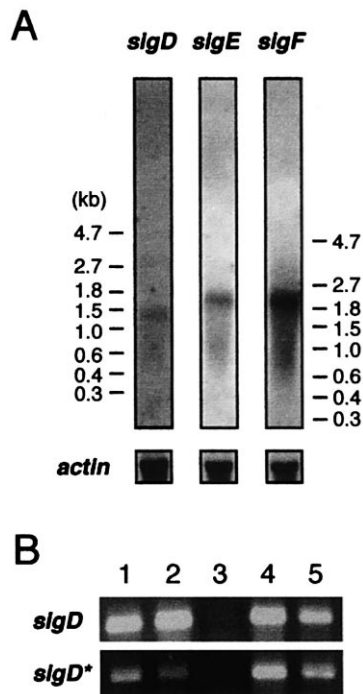


Fig. 3. Hybridization and RT-PCR analyses. A: Northern blot analysis. 10  $\mu$ g of total RNAs from *Arabidopsis* 12-day-old seedlings were hybridized with DIG-labelled antisense ssDNA of *sigD*, *sigE*, *sigF* or *actin* gene. Chemiluminescent signals of CSPD were exposed to an X-ray film for 10 h (*sigD*, *sigE* and *sigF*) or 3 min (*actin*). B: RT-PCR analysis. Distribution of *sigD* transcripts in plant organs, cotyledon (lane 1), stem (lane 2), root (lane 3), cauline leaf (lane 4) and rosette leaf (lane 5). PCR products were electrophoresed on a 1.2% agarose gel and stained with ethidium bromide.

are two possibilities for its calculated molecular weight. Interestingly, we also isolated two cDNA species of *sigD*, both of which would express a polypeptide translationally interrupted within region 2.4 (see Section 3.4).

### 3.2. Phylogenetic relationship of plant $\sigma$ factors

To understand the relationship among plant  $\sigma$  factors, representative eubacterial group 1 and group 2  $\sigma$  factors and known plastid  $\sigma$  factors were analyzed phylogenetically. The results showed that higher plant  $\sigma$  factors formed a monophyletic group (Fig. 2). However, probably because of lower conservation, branches of plastid and cyanobacterial group 2  $\sigma$  factors were not clearly separated. Interestingly, many of the higher plant  $\sigma$  factors fell into four subgroups, each including either *Arabidopsis* SigA, SigB, SigC or SigF. In addition, at least SigA subgroup members were further divided into monocot and dicot groups. Higher plant  $\sigma$  factors appear to be classified into structurally conserved groups beyond plant species.

In a comparison of the intron sites of *Arabidopsis* plastid *sig* genes, all three in *sigD* and four out of eight in *sigF* turned out to be identical to the corresponding intron sites of *sigA*, *sigB* and *sigC* (Fig. 1B). Especially, intron sites of *sigB*, *sigC* and *sigF* corresponding to the conserved regions through 1.2–4.2 are completely identical. In contrast, none of the *sigE* intron sites appears to correspond to those in any of the other genes. These observations indicate that gene duplication events from one or a few origins and sequential insertion of intron may have led to multiple *sig* genes during plant evolution, and that higher plant  $\sigma$  factors may share a common origin(s).

### 3.3. Transcriptional expression of *sigD*, *sigE* and *sigF* genes

Northern blot analyses were performed to detect transcripts of *sigD*, *sigE* and *sigF* using total RNA of 12-day-old seedlings grown under continuous illumination (Fig. 3A). Only one weak signal was observed for every *sig* gene. The estimated transcript sizes were 1.4 kb for *sigD*, 1.6 kb for *sigE* and 1.7 kb for *sigF*, and roughly corresponded to the isolated cDNA lengths. Transcript levels for *sigD* were the lowest among the three genes, whereas transcript levels for *sigF* were highest.

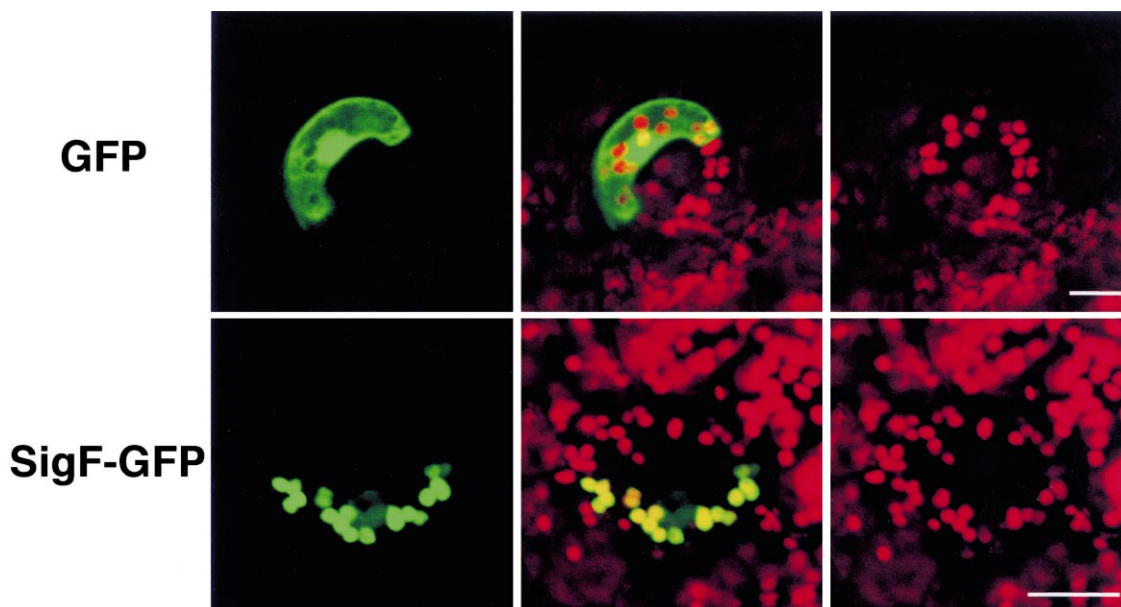


Fig. 4. Chloroplast localization of GFP fused with the amino-terminal region of SigF. Fluorescence images of GFP (left, green), chlorophyll (right, red), and merged images (middle) for non-fused GFP (GFP) or the chimeric GFP with SigF (SigF-GFP) are shown. Bars, 20  $\mu$ m.

### 3.4. Alternative splicing of *sigD*

We have isolated another *sigD* cDNA species that contains a 14-base upstream shift of the intron 2 splice donor site. Since this alternative cDNA species was frequently isolated from the library, we concluded that this transcript was generated by an alternative splicing mechanism. No other splicing pattern was found by RT-PCR analyses (data not shown). The transcript encoded a 29.8-kDa polypeptide lacking the carboxy-terminal region of the full-length *sigD* product (SigD) by a translational frameshifting. The 29.8-kDa polypeptide (SigD\*) contains only 10 extended residues, -L-V-V-C-R-E-A-C-G-S, from the frameshifting site within region 2.4 (Fig. 1B). Subsequently we investigated plant organ specificity of expression of the two alternatively spliced transcripts by RT-PCR using specific primer sets for each (Fig. 3B), and both were detected in cotyledons, stems, cauline leaves and rosette leaves. However, relative amounts of the two transcripts appeared to differ in each organ.

### 3.5. Chloroplast localization of SigF

Although the amino-terminal regions of plastid  $\sigma$  factors revealed limited conservation, they are relatively rich in serine and threonine residues. Nuclear-encoded plastid protein precursors generally contain serine- and threonine-rich amino-terminal extensions that function as plastid-targeting signals [29]. Therefore, we fused GFP [14,22] with the amino-terminal region of SigF, and co-introduced the corresponding constructs into tobacco leaf cells by the particle bombardment method. The fused GFP gene, which was controlled under the CaMV35S promoter, was expressed transiently and localized in chloroplasts, while non-fused GFP was observed in the cytoplasm and the nucleus (Fig. 4). Thus the amino-terminus of SigF was shown to function as a plastid-targeting signal. We could not confirm properties of the amino-terminal sequences of SigD and SigE by similar experiments (data not shown).

## 4. Discussion

In this report, we described three new genes for putative plastid  $\sigma$  factors, and therefore in *A. thaliana*, six  $\sigma$  factor genes have been identified thus far. Since almost 90% of the whole genome sequence data have already been identified, it is probable that few unidentified  $\sigma$  factor genes remain. Such  $\sigma$  heterogeneity is observed in dicots [12,14] as well as in monocots [15,17,19], and appears to be a general characteristics in higher plant plastids. What is the significance of plastid  $\sigma$  factor heterogeneity? In eubacteria, modulation of RNA polymerase specificities by  $\sigma$  heterogeneity is a major mode of regulating transcription. Because  $\sigma$  factors are the specificity factor for eubacterial RNA polymerases, PEP should be a target for regulation by plastid  $\sigma$  factors. Based on the structures of plastid  $\sigma$  factors, corresponding promoters are expected to contain the 'TATAAT'-type -10 box, and differential activation of these promoters has been observed in many cases [30]. In our previous study [21], the expression of *sigA* and *sigB* genes was analyzed, and found to be differentially regulated, depending on the organ and developmental stage. It is likely, therefore, that  $\sigma$  heterogeneity is involved in differential promoter usage during plastid development.

Of the six *Arabidopsis sig* gene products, SigA, SigB, SigC [12,21] and SigF (this study) were shown to be targeted into

plastids using GFP fusions. In our preliminary study, SigD- and SigE-GFP fusions also tended to localize around plastids (data not shown). However, these fusions also localized to the nucleus as well as to the cytosol. Therefore, a more detailed study is required to provide conclusive results.

We have found an alternative *sigD* transcript due to differential splicing, and this transcript was predicted to encode a SigD protein lacking the conserved regions 3 and 4 (SigD\*). Region 4 is involved in the recognition of the promoter -35 element, and thus, this particular splicing might generate changes in promoter specificity. It should be noted that the molecular weight of a  $\sigma$ -like factor (SLF29 [31]) is similar to SigD\*. However, the functional significance of this observation must await analyses at the protein level.

**Acknowledgements:** We thank Dr. Nobuyoshi Mochizuki (Kyoto University), Dr. Satoshi Tabata (Kazusa DNA Research Institute) and Dr. Yasuo Niwa (University of Shizuoka) for providing DNA materials. This work was supported by Grants-in-Aid for Scientific Research in Priority Areas (No. 1017207) to K.T., and Grants-in-Aid for Scientific Research (No. 09460046) to H.T. from the Ministry of Education, Science, and Culture of Japan, by Special Coordination Funds for Promoting Science and Technology to K.T. from the Science and Technology Agency. M.F. was a Research Fellow of the Japan Society for the Promotion of Science.

## References

- [1] Sugiura, M. (1992) Plant Mol. Biol. 19, 149–168.
- [2] Maliga, P. (1998) Trends Plant Sci. 3, 4–6.
- [3] Hess, W.R. and Börner, T. (1999) Int. Rev. Cytol. 190, 1–59.
- [4] Liere, K. and Maliga, P. (1999) EMBO J. 18, 249–257.
- [5] Helmann, J.D. and Chamberlin, M.J. (1988) Annu. Rev. Biochem. 57, 839–872.
- [6] Igloi, G.L. and Kössel, H. (1992) Crit. Rev. Plant Sci. 10, 525–558.
- [7] Allison, L.A., Simon, L.D. and Maliga, P. (1996) EMBO J. 15, 2802–2809.
- [8] Serino, G. and Maliga, P. (1998) Plant Physiol. 117, 1165–1170.
- [9] De Santis-Maclossek, G., Kofer, W., Bock, A., Schoch, S., Maier, R.M., Wanner, G., Rudiger, W., Koop, H.U. and Herrmann, R.G. (1999) Plant J. 18, 477–489.
- [10] Tanaka, K., Oikawa, K., Ohta, N., Kuroiwa, H., Kuroiwa, T. and Takahashi, H. (1996) Science 272, 1932–1935.
- [11] Oikawa, K., Tanaka, K. and Takahashi, H. (1998) Gene 210, 277–285.
- [12] Tanaka, K., Tozawa, Y., Mochizuki, N., Shinozaki, K., Nagatani, A., Wakasa, K. and Takahashi, H. (1997) FEBS Lett. 413, 309–313.
- [13] Liu, B. and Troxler, R.F. (1996) Proc. Natl. Acad. Sci. USA 93, 3313–3318.
- [14] Isono, K., Shimizu, M., Yoshimoto, K., Niwa, Y., Satoh, K., Yokota, A. and Kobayashi, H. (1997) Proc. Natl. Acad. Sci. USA 94, 14948–14953.
- [15] Tozawa, Y., Tanaka, K., Takahashi, H. and Wakasa, K. (1998) Nucleic Acids Res. 26, 415–419.
- [16] Kestermann, M. (1998) Nucleic Acids Res. 26, 2747–2753.
- [17] Tan, S. and Troxler, R.F. (1999) Proc. Natl. Acad. Sci. USA 96, 5316–5321.
- [18] Morikawa, K., Ito, S., Tsunoyama, Y., Nakahira, Y., Shiina, T. and Toyoshima, Y. (1999) FEBS Lett. 451, 275–278.
- [19] Lahiri, S.D., Yao, J., McCumbers, C. and Allison, L.A. (1999) Mol. Cell. Biol. Res. Commun. 1, 14–20.
- [20] Murashige, T. and Skoog, F. (1962) Physiol. Plant. 15, 473–497.
- [21] Kanamaru, K., Fujiwara, M., Seki, M., Katagiri, T., Nakamura, M., Mochizuki, N., Nagatani, A., Shinozaki, K., Tanaka, K. and Takahashi, H. (1999) Plant Cell Physiol. 40, 832–842.
- [22] Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. and Prasher, D.C. (1994) Science 263, 802–805.
- [23] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.

- [24] Saitou, N. and Nei, M. (1987) *Mol. Biol. Evol.* 4, 406–425.
- [25] Kim, J.W. and Minamikawa, T. (1996) *Plant Sci.* 117, 131–138.
- [26] Wösten, M.M. (1998) *FEMS Microbiol. Rev.* 22, 127–150.
- [27] Lonetto, M., Gribskov, M. and Gross, C.A. (1992) *J. Bacteriol.* 174, 3843–3849.
- [28] Hakimi, M.A., Privat, I., Valay, J.G. and Lerbs-Mache, S. (2000) *J. Biol. Chem.* 275, 9215–9221.
- [29] Keegstra, K., Olsen, L.J. and Theg, S.M. (1989) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 40, 471–501.
- [30] Satoh, J., Baba, K., Nakahira, Y., Tsunoyama, Y., Shiina, T. and Toyoshima, Y. (1999) *Plant J.* 18, 407–415.
- [31] Tiller, K., Eisermann, A. and Link, G. (1991) *Eur. J. Biochem.* 198, 93–99.