A tobacco chloroplast DNA sequence possibly coding for a polypeptide similar to *E. coli* RNA polymerase β-subunit

M. Ohme, M. Tanaka, J. Chunwongse, K. Shinozaki and M. Sugiura*

Center for Gene Research, Nagoya University, Chikusa, Nagoya 464, Japan

Received 5 March 1986

DNA sequencing has revealed a long open reading frame (ORF) in the large single-copy region of tobacco chloroplast DNA. This ORF consists of 1070 codons and its deduced amino acid sequence shows about 39% homology to that of the β-subunit of *E. coli* RNA polymerase. This finding raises a possibility that some of the chloroplast RNA polymerase subunits are coded for by the chloroplast genome.

(Tobacco) Chloroplast DNA sequence RNA polymerase rpo*B gene*

1. INTRODUCTION

Chloroplasts in higher plants contain their own DNA and RNA polymerase. Chloroplast RNA polymerase in higher plants has been suggested to be nuclear encoded [1–4]. However, sequences hybridized with the β-subunit gene (*rpo*B) of *Escherichia coli* RNA polymerase have been found in *Chlamydomonas* chloroplast DNA as well as nuclear DNA [5]. We have sequenced one of the tobacco chloroplast DNA regions which hybridized with the *E. coli rpo*B probe and found an open reading frame whose amino acid sequence resembles that of the *E. coli* β-subunit. This finding raises the interesting possibility that chloroplasts in higher plants are an additional site of synthesis of their RNA polymerase subunits.

2. MATERIALS AND METHODS

Transducing phage λ rif^{d}18 was kindly provided by Dr A. Ishihama and its 6.0% *Eco*RI fragment was used as the *rpo*B probe. Southern blot hybridization was carried out as described [6]. The clone bank of tobacco (*Nicotiana tabacum* var. Bright Yellow 4) chloroplast DNA and recombinant plasmid pTB7 are described in [7]. DNA sequence was determined as described [8] and analyzed using the GENETYX program (Software Development Co., Tokyo).

3. RESULTS AND DISCUSSION

Using the clone bank of tobacco chloroplast DNA [7], we searched for the chloroplast DNA regions homologous to the *E. coli rpo*B by hybridization with the probe. Recombinant plasmid pTB7 which contains consecutive 3.0, 3.2, 1.1 and 3.6 kbp *Bam*HI fragments (Ba13, Ba12a, Ba25 and Ba10b, respectively, [7]) was one of the several clones which gave a weak signal. We sequenced portions of the cloned *Bam*HI fragments in pTB7 and found a 3210 open reading frame on the same strand as the gene for H^+-ATPase α-subunit (*atp*A) (figs 1 and 2).

The predicted polypeptide is 1070 amino acid residues long and 272 residues shorter than the *E. coli* β-subunit (1342 residues) [10]. When gaps are introduced, the amino acid sequence deduced from ORF_{1070} has about 39% homology with that of the *E. coli* β-subunit (fig.3). This homology is comparable to that (23–68%) of ribosomal proteins [6].

Recently Lerbs et al. [4] have analyzed in vitro

* To whom correspondence should be addressed

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

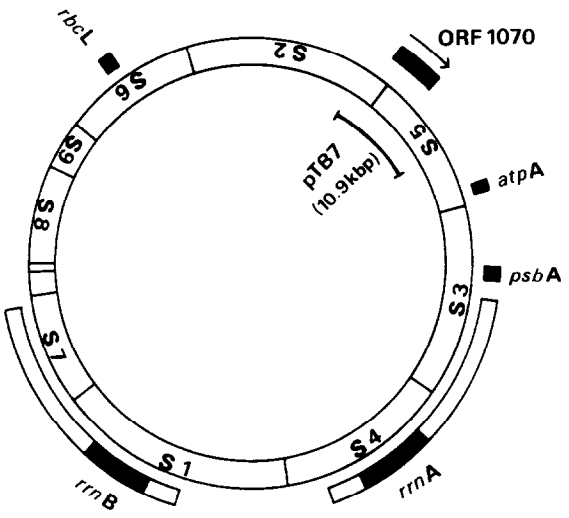

Fig.1. Position of the cloned fragment in pTB7 and of ORF_{1070} on the *SalI* map of tobacco chloroplast DNA. The arrow indicates the direction of possible transcription.

translated products of both chloroplast RNA and poly(A)$^+$ nuclear RNA in spinach and suggested that all chloroplast RNA polymerase polypeptides are coded for by the nuclear genome [4]. From their study, ORF_{1070} could be a pseudogene. Another possibility is that chloroplasts have two RNA polymerase species, one coded for by nuclear genomes and the other by chloroplast genomes which is expressed in restricted conditions, e.g. in etiochloroplasts. Two types of RNA polymerase activity have been isolated from *Euglena* chloroplasts [9].

```
ATGCTCGGGGATGGAAATGAGGGAATATCTACAATACCTGGATTTAATCAGATACAATTTGAAGGATTTTGTAGGTTCATTGATCAAGGTTTGACGGAAG  100
 M  L  G  D  G  N  E  G  I  S  T  I  P  G  F  N  Q  I  Q  F  E  G  F  C  R  F  I  D  Q  G  L  T  E  E

AACTTTATAAGTTTCCAAAAATTGAAGATACAGATCAAGAAATTGAATTTCAATTATTTGTGGAAACATATCAATTGGTCGAACCCTTGATAAAGGAAAG  200
    L  Y  K  F  P  K  I  E  D  T  D  Q  E  I  E  F  Q  L  F  V  E  T  Y  Q  L  V  E  P  L  I  K  E  R

AGATGCTGTGTATGAATCACTCACATATTCTTCTGAATTATATGTATCCGCGGGATTAATTTGGAAAAACAGTAGGGATATGCAAGAACAAACAATTTTT  300
 D  A  V  Y  F  S  L  T  Y  S  S  E  L  Y  V  S  A  G  L  I  W  K  N  S  R  D  M  Q  E  Q  T  I  F

ATCGGAAACATTCCTCTAATGAATTCCCTGGGAACTTCTATAGTCAATGGAATATATAGAATTGTGATCAATCAAATATTGCAAAGTCCCGGTATTTATT  400
I  G  N  I  P  L  M  N  S  L  G  T  S  I  V  N  G  I  Y  R  I  V  I  N  Q  I  L  Q  S  P  G  I  Y  Y

ACCGATCAGAATTGGACCATAACGGAATTTCGGTCTATACCGGCACCATAATATCAGATTGGGGAGGAAGATCAGAATTAGAAATTGATAGAAAAGCAAG  500
  R  S  E  L  D  H  N  G  I  S  V  Y  T  G  T  I  I  S  D  W  G  G  R  S  E  L  E  I  D  R  K  A  R

GATATGGGCTCGTGTAAGTAGGAAACAAAAAATATCTATTCTAGTTCTATCATCAGCTATGGGTTTGAATCTAAGAGAAATTCTAGAGAATGTTTGCTAT  600
 I  W  A  R  V  S  R  K  Q  K  I  S  I  L  V  L  S  S  A  M  G  L  N  L  R  E  I  L  E  N  V  C  Y

CCTGAAATTTTTTTGTCTTTTCTGAGTGATAAGGAGAGAAAAAAAATTGGGTCAAAAGAAAATGCCATTTTGGAGTTTTATCAACAATTTGCTTGTGTAG  700
P  E  I  F  L  S  F  L  S  D  K  E  R  K  K  I  G  S  K  E  N  A  I  L  E  F  Y  Q  Q  F  A  C  V  G

GTGGCGATCCGGTATTTTCTGAATCCTTATGTAAGGAATTACAAAAGAAATTCTTTCAACAAAGATGTGAATTAGGAAGGATTGGTCGACGAAATATGAA  800
  G  D  P  V  F  S  E  S  L  C  K  E  L  Q  K  K  F  F  Q  Q  R  C  E  L  G  R  I  G  R  R  N  M  N

CCGAAGACTGAACCTTGATATACCCCAGAACAATACATTTTTGTTACCACGAGATATATTGGCAGCCGCCGATCATTTGATTGGGCTGAAATTTGGAATG  900
 R  R  L  N  L  D  I  P  Q  N  N  T  F  L  L  P  R  D  I  L  A  A  A  D  H  L  I  G  L  K  F  G  M

GGTGCACTTGACGATATGAATCATTTGAAAAATAAACGTATTCGTTCTGTAGCAGATCTTTTACAAGATCAATTCGGATTGGCTCTGGTTCGTTTAGAAA 1000
G  A  L  D  D  M  N  H  L  K  N  K  R  I  R  S  V  A  D  L  L  Q  D  Q  F  G  L  A  L  V  R  L  E  N

ATGTGGTTCGGGGGACTATATGTGGAGCAATTCGGCATAAATTGATACCGACACCTCAGAATTTGGTAACCTCAACTCCATTAACAACTACTTATGAATC 1100
  V  V  R  G  T  I  C  G  A  I  R  H  K  L  I  P  T  P  Q  N  L  V  T  S  T  P  L  T  T  T  Y  E  S

CTTTTTCGGTTTACACCCATTATCTCAAGTTTTGGATCGAACTAATCCATTGACACAAATAGTTCATGGGAGAAAATTAAGTTATTTGGGCCCTGGAGGA 1200
 F  F  G  L  H  P  L  S  Q  V  L  D  R  T  N  P  L  T  Q  I  V  H  G  R  K  L  S  Y  L  G  P  G  G

CTGACAGGGCGCACTGCTAGTTTTCGGATACGAGATATCCATCCTAGTCACTATGGACGTATTTGCCCAATTGACACATCTGAAGGAATCAATGTTGGAC 1300
L  T  G  R  T  A  S  F  R  I  R  D  I  H  P  S  H  Y  G  R  I  C  P  I  D  T  S  E  G  I  N  V  G  L

TTATTGGATCCTTAGCAATTCATGCGAGGATTGGTCATTGGGGATCTCTAGAAAGCCCTTTTTATGAAATTTCTGAGAGGTCAACCGGGGTACGGATGCT 1400
  I  G  S  L  A  I  H  A  R  I  G  H  W  G  S  L  E  S  P  F  Y  E  I  S  E  R  S  T  G  V  R  M  L

TTATTTATCACCAGGTAGAGATGAATACTATATGGTAGCGGCAGGAAATTCTTTAGCCTTAAATCAGGATATTCAGGAAGAACAGGTTGTTCCAGCTCGA 1500
 Y  L  S  P  G  R  D  E  Y  Y  M  V  A  A  G  N  S  L  A  L  N  Q  D  I  Q  E  E  Q  V  V  P  A  R

TACCGTCAAGAATTCTTGACTATTGCATGGGAACAGGTTCATCTTCGAAGTATTTTTCCTTTTCAATATTTTTCTATTGGAGCTTCCCTCATTCCTTTTA 1600
Y  R  Q  E  F  L  T  I  A  W  E  Q  V  H  L  R  S  I  F  P  F  Q  Y  F  S  I  G  A  S  L  I  P  F  I

TCGAACATAATGATGCGAATCGAGCTTTAATGAGTTCTAATATGCAACGTCAAGCAGTTCCTCTTTCTCGCTCCGAGAAATGCATTGTTGGAACTGGGTT 1700
  E  H  N  D  A  N  R  A  L  M  S  S  N  M  Q  R  Q  A  V  P  L  S  R  S  E  K  C  I  V  G  T  G  L

GGAACGACAAGCAGCTCTAGATTCGGGGGCTCTTGCTATAGCCGAACGCGAGGGAAGGGTCGTTTATACCAATACTGACAAGATTCTTTTAGCAGGTAAT 1800
 E  R  Q  A  A  L  D  S  G  A  L  A  I  A  E  R  E  G  R  V  V  Y  T  N  T  D  K  I  L  L  A  G  N
```

```
GGAGATATTCTAAGCATTCCATTAGTTATATATCAACGTTCCAATAAAAATACTTGTATGCATCAAAAACTCCAGGTTCCTCGGGGTAAATGCATTAAAA 1900
G  D  I  L  S  I  P  L  V  I  Y  Q  R  S  N  K  N  T  C  M  H  Q  K  L  Q  V  P  R  G  K  C  I  K  K

AGGGACAAATTTTAGCGGATGGTGCTGCTACGGTTGGTGGCGAACTTGCTTTGGGGAAAAACGTATTAGTAGCTTATATGCCGTGGGAGGGTTACAATTC 2000
   G  Q  I  L  A  D  G  A  A  T  V  G  G  E  L  A  L  G  K  N  V  L  V  A  Y  M  P  W  E  G  Y  N  S

TGAAGATGCAGTACTTATTAGCGAGCGTTTGGTATATGAAGATATTTATACTTCTTTTCACATACGGAAATATGAAATTCAGACTCATGTGACAAGCCAA 2100
E  D  A  V  L  I  S  E  R  L  V  Y  E  D  I  Y  T  S  F  H  I  R  K  Y  E  I  Q  T  H  V  T  S  Q

GGCCCTGAAAAAGTAACTAATGAAATACCGCATTTAGAAGCCCATTTACTCCGCAATTTAGATAAAAATGGAATTGTGATGCTGGGATCTTGGGTAGAGA 2200
G  P  E  K  V  T  N  E  I  P  H  L  E  A  H  L  L  R  N  L  D  K  N  G  I  V  M  L  G  S  W  V  E  T

CAGGTGATATTTTAGTAGGTAAATTAACACCCCAGGTCGTGAAAGAATCGTCGTATGCCCCGGAAGATAGATTGTTACGAGCTATACTTGGTATTCAGGT 2300
   G  D  I  L  V  G  K  L  T  P  Q  V  V  K  E  S  S  Y  A  P  E  D  R  L  L  R  A  I  L  G  I  Q  V

ATCTACTTCAAAAGAAACTTGTCTAAAACTACCTATAGGTGGCAGGGGTCGGGTTATTGATGTGAGGTGGATCCAGAAGAGGGGGTGGTTCTAGTTATAAT 2400
S  T  S  K  E  T  C  L  K  L  P  I  G  G  R  G  R  V  I  D  V  R  W  I  Q  K  R  G  G  S  S  Y  N

CCCGAAACGATTCGTGTATATATTTTACAGAAACGTGAAATCAAAGTAGGCGATAAAGTAGCTGGAAGACACGGAAATAAAGGTATCATTTCCAAAATTT 2500
P  E  T  I  R  V  Y  I  L  Q  K  R  E  I  K  V  G  D  K  V  A  G  R  H  G  N  K  G  I  I  S  K  I  L

TGCCTAGACAAGATATGCCTTATTTACAAGATGGAAGATCCGTTGATATGGTCTTTAACCCATTAGGAGTACCTTCACGAATGAATGTAGGACAGATATT 2600
   P  R  Q  D  M  P  Y  L  Q  D  G  R  S  V  D  M  V  F  N  P  L  G  V  P  S  R  M  N  V  G  Q  I  F

TGAATGTTCACTAGGGTTAGCAGGGAGTCTGCTAGACAGACATTATCGAATAGCACCTTTTGATGAGAGATATGAACAAGAAGCTTCGAGAAAACTTGTG 2700
E  C  S  L  G  L  A  G  S  L  L  D  R  H  Y  R  I  A  P  F  D  E  R  Y  E  Q  E  A  S  R  K  L  V

TTTTCTGAATTATATGAAGCCAGTAAGCAAACAGCGAATCCATGGGTATTTGAACCCGAATATCCAGGAAAAAGCAGAATATTTGATGGAAGGACGGGGA 2800
F  S  E  L  Y  E  A  S  K  Q  T  A  N  P  W  V  F  E  P  E  Y  P  G  K  S  R  I  F  D  G  R  T  G  N

ATCCTTTTGAACAACCCGTTATAATAGGAAAGCCTTATATCTTGAAATTAATTCATCAAGTTGATGATAAAATCCATGGGCGCTCCAGTGGACATTATGC 2900
   P  F  E  Q  P  V  I  I  G  K  P  Y  I  L  K  L  I  H  Q  V  D  D  K  I  H  G  R  S  S  G  H  Y  A

GCTTGTTACACAACAACCCCTTAGAGGAAGAGCCAAACAGGGGGGACAGCGGGTAGGAGAAATGGAGGTTTGGGCTCTAGAAGGGTTTGGGGTTGCTCAT 3000
L  V  T  Q  Q  P  L  R  G  R  A  K  Q  G  G  Q  R  V  G  E  M  E  V  W  A  L  E  G  F  G  V  A  H

ATTTTACAAGAGATGCTTACTTATAAATCGGATCATATTAGAGCTCGCCAGGAAGTACTTGGTACTACGATCATTGGGGGAACAATACCTAATCCCGAAG 3100
I  L  Q  E  M  L  T  Y  K  S  D  H  I  R  A  R  Q  E  V  L  G  T  T  I  I  G  G  T  I  P  N  P  E  D

ATGCTCCAGAATCTTTTCGATTGCTCGTTCGAGAACTACGATCTTTAGCTCTGGAACTGAATCATTTCCTTGTATCTGAGAAGAACTTCCAGATTAATAG 3200
   A  P  E  S  F  R  L  L  V  R  E  L  R  S  L  A  L  E  L  N  H  F  L  V  S  E  K  N  F  Q  I  N  R

GAAGGAAGCT
K  E  A
```

Fig.2. DNA sequence of ORF_{1070}. The RNA-like strand (strand B) is presented. The deduced amino acid sequence is shown below the DNA sequence.

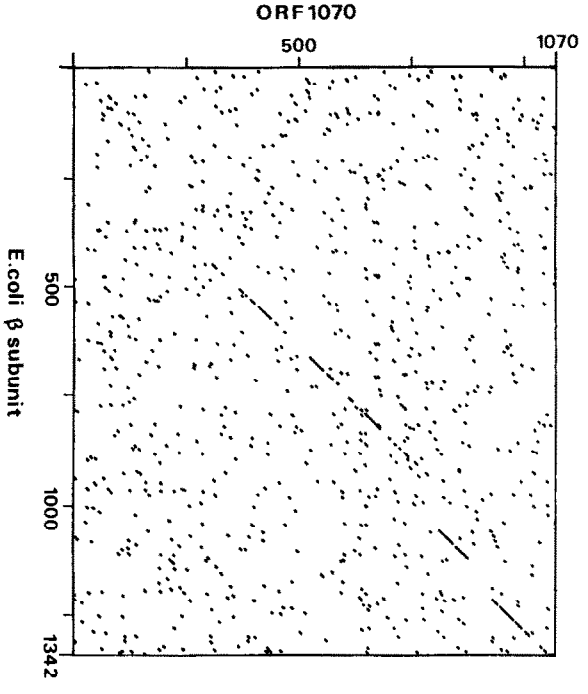

Fig.3. Dot matrix analysis between the deduced amino acid sequences of ORF_{1070} and *E. coli* β-subunit. The matrix program used for analysis scores a dot for 3 out of 4 identities.

ACKNOWLEDGEMENT

This work was supported in part by a Grant-in-Aid from the Ministry of Education, Science and Culture (Japan).

REFERENCES

[1] Ellis, R.J. and Hartley, M.R. (1971) Nat. New Biol. 233, 193–196.
[2] Bunger, W. and Feierabend, K. (1980) Planta 149, 163–169.
[3] Siemenroth, A., Wollgiehn, R., Neumann, D. and Borner, T. (1981) Planta 153, 547–555.

[4] Lerbs, S., Brautigam, E. and Parthier, B. (1985) EMBO J. 4, 1661–1666.
[5] Watson, J.C. and Surzycki, S.J. (1983) Curr. Genet. 7, 201–210.
[6] Tanaka, M., Wakasugi, T., Sugita, M., Shinozaki, K. and Sugiura, M. (1986) submitted.
[7] Sugiura, M., Shinozaki, K., Zaita, N., Kusuda, M. and Kumano, M. (1986) submitted.
[8] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[9] Greenberg, B.M., Narita, J.O., DeLuca-Flaherty, C., Gruissem, W., Rushlow, K.A. and Hallick, R.B. (1984) J. Biol. Chem. 259, 14880–14887.
[10] Ovchinnikov, Y.A., Monastyrskaya, G.S., Gubanov, V.V., Guryev, S.O., Chertov, O.Y., Modyanov, N.N., Grinkevich, V.A., Makarova, I.A., Marchenko, T.V., Polovnikova, I.N., Lipkin, V.M. and Sverdlov, E.D. (1981) Eur. J. Biochem. 116, 621–629.