Nucleotide sequence of tobacco chloroplast gene for the α subunit of proton-translocating ATPase

Hiroshi Deno+, Kazuo Shinozaki+ and Masahiro Sugiura*

Department of Biology, Nagoya University, Chikusa, Nagoya 464, and ⁺National Institute of Genetics, Mishima 411, Japan

Received 14 January 1983; Revised and Accepted 3 March 1983

ABSTRACT

The tobacco chloroplast gene for the α subunit of protontranslocating ATPase has been cloned and sequenced. The coding region contains 1521 bp (507 codons). The nucleotide sequence and the deduced amino acid sequence show 55% and 54% homologies with those of the <u>E. coli</u> α subunit, respectively. The deduced amino acid composition is quite similar to that estimated for the spinach α subunit.

INTRODUCTION

The proton-translocating ATPase ($\text{H}^+\text{-}ATPase$) of chloroplasts is an essential component of light-driven ATP synthesis. It consists of two parts, CF₁ and CF₀. The CF₁ is located on the outer surface of the thylakoid membrane and composed of five different subunits (α , β , γ , δ and ε). The CF₀ is located in the membrane and composed of three different subunits (I, II and III). The genes for five subunits (α , β , ε , I and III) are known to be encoded on chloroplast DNA (1, 2). Recently the genes for the β and ε subunits of maize (3), spinach (4) and tobacco (5, 6) have been sequenced and found to be fused. Westhoff et al. (1) have reported that the gene for the spinach α subunit is separated by approximately 40 kbp from the β and ε genes and located near the gene for the 32 kd protein.

We show here the nucleotide sequence of the tobacco chloroplast gene for the α subunit of H⁺-ATPase and compare its deduced amino acid sequence with that of the <u>E</u>. <u>coli</u> α subunit.

MATERIALS AND METHODS

Transducing phage $\lambda asn 5$ which carries a gene cluster for the H⁺-ATPase of <u>E</u>. <u>coli</u> was a kind gift from Dr's. M. Futai and H. Kanazawa (7). Southern blot hybridization was carried out at 60°C for 16 hr in 1 x Denhardt's solution (3 x SSC, 0.02% Ficoll, 0.02% polyvinylpyroridone, 0.02% bovine serum albumin) containing 10 μ g/ml denatured calf thymus DNA and nick-translated λ <u>asn</u>5 DNA after prehybridization with 1 x Denhardt's solution at 60°C for 6 hr (8). Recombinant plasmid pTB4 containing a 5.7 Md BamHI fragment (Ba4) of <u>Nicotiana tabacum</u> (var. Bright Yellow 4) chloroplast DNA was constructed as described (9). The plasmid DNA was digested with BamHI and the 5.7 Md fragment was separated from vector pBR322 by electrophoresis in a 1% agarose gel. DNA sequence was determined by the method of Maxam and Gilbert (10).

RESULTS AND DISCUSSION

To determine which restriction fragments contain the genes for the H⁺-ATPase, restriction fragments of tobacco chloroplast DNA blotted to nitrocellulose filter were hybridized with nicktranslated $\lambda asn5$ DNA containing a gene cluster for the <u>E</u>. <u>coli</u> H⁺-ATPase. The DNA probe hybridized weakly to a 5.7 Md BamHI fragment and a 10.9 Md SalI fragment (data not shown). We cloned a 5.7 Md BamHI fragment of tobacco chloroplast DNA into pBR322. On digestion with HindIII, the 5.7 Md BamHI fragment yields 0.90, 0.03, 1.65, 1.80 and 1.32 Md sub-fragments as shown in Fig. 1. Nick-translated $\lambda asn5$ DNA hybridized weakly to the 1.65 Md sub-fragment (data not shown). We then sequenced portions of the 1.65 Md and the adjacent 0.03 Md + 0.90 Md subfragments by the strategy presented in Fig. 1.

Fig. 2 shows the nucleotide sequence of the 1780 bp portion which contains an open reading frame (ORF) of 1521 nucleotides (507 codons). The deduced amino acid sequence of this ORF shows 54% homology with that of the <u>E</u>. <u>coli</u> α subunit (11-13) as shown in Fig. 3. The deduced amino acid composition of this ORF is quite similar to that reported for the spinach α subunit (14) as presented in Table 1. The calculated molecular weight is 55446 which is similar to that (59000) estimated for the spinach α subunit (14). The 5.7 Md BamHI fragment overlaps fragments B2 and B3 (19.1 Md and 12.8 Md BglI fragments, respectively) on the physical map of tobacco chloroplast DNA (15) and this ORF



Fig. 1. Physical map of the cloned 5.7 Md BamHI fragment from tobacco chloroplast DNA and the strategy for sequencing part of it. Strand B codes for the α gene (we designated a strand which codes for the large subunit of ribulose-1,5-bisphosphate carboxylase as A). Coding regions are shown by thick lines. Horizontal arrows indicate the direction and extent of DNA segments sequenced.

contains a BglI site (see Fig. 1). These observations indicate that this ORF is located approximately 40 kbp apart from the β and ε genes. We therefore concluded that this ORF codes for the α subunit. Based on the physical map of tobacco chloroplast DNA (15) and sequence analysis of the α , β and ε genes (6), these three genes were found to be oriented in the same direction on the chloroplast DNA map (strand B, see Fig. 1). Sequences AAAG, which are complementary to the 3' end of tobacco chloroplast 16S rRNA (---CUUU-3', ref. 16) and putative ribosome binding sites, are present 12 to 9 bp and 8 to 5 bp upstream from the

Nucleic Acids Research

TGTTTGAACAACGAGTTACATTAGGTACCATTG -10	01
crccaatattegecatetregeaacaateaakeaaataacteattaetetretetaetaegetattattttttttttteaaaaataat <u>aakeaaae</u> aate	7
ATGGTAACCATTCGAGGTGACGAAATTAGTAATATTATCCGTGAACGTATTGAACAATATAATAGAGAAGTAAGATTGTAAATAGGGAAGTAGGGTACCGTACTTC	00
AAGTAGGCGACGGCATTGCTCGTATTCACGGTCTTGATGAGGTAATTGGCGGGGTGAATTAGTCGAATTTGAAGAGGGTACAATAGGCATTGCTCTGAATTT	00
GGAATCAAATAATGTTGGTGTTTTATTGGGGGGGGGGGG	00
AGTGAGGCTTATTTGGGTCGTGTTATAAATGCCCTGGCTAAACCTATTGATGGTAGAGGTGAAATTTCAGCTTCTGAATTTCGATTAATCGAATCTGCCG	00
ccccgggtattattcgcgccgtrcgtatatggcctcttcaaaccgggcttattgctattgattcgatgatcctatagggcgtggtcggggagtatt 5(00
AATTATTGGGGACAGACGGTAAAACAGCAGGAGAGGACAGATACGATACGATCCTCAATCAA	00
CAAAAAGCATCTTCTGTGGCCCAGGTCGTAACTACTTTACAGGAAAGGGGAGCGATGGAATACACTATTGTGGTAGCCGAAACGGCAGATTCCCCTGCTA	00
cartacaarácetreereereeseseseseseseseseseseseseseses	00
аесесааесттатсессааатетстстаттасеааеассессеетсетеааесттатстаееаеатеттаттаттаттесаттсасесстттееаа 90	00
AGAGCCGCTÀAATTAAGTTĊTAGTTTAGGTGAAGGAAGTÀTGACCGCCTTACCAATAGTTGAAACCCAATCGGGAGAGATGTTTCGGCTTATATTCCTACTA 100	00
ATGTAATTTCCATTACTGATGGACAAATCTTCTTATCCGCCGACCTATTCAATTCTGGAATCAGGCCTGCTATTAATGTGGGGTATCTCCGTTTCCAGAGT	00
GGGGGTCCGCAGCTCAAATAAAAGCCATGAAACTAGCTGGTAAATTAAAATTAGAACTAGCACAATTCGCCAGAATTAGAAGCCTTTGCACAATTTGCT 120	00
TCTGATCTCGATAAAGCTACTCAGAATCAATTGGCAAGAGGTCAACGATTACGTGAATTGCTTAAACAATCCCAATCAGCTCCTCTCACGGTAGAAGAGGC	00
agataatga¢tattttatac¢ggaacaaa¢gctatcttgåttcattagå¢gttggacag¢taaggaaatttcttgttga¢ctacgtacttacttaaaaa¢ 140	00
TAATAAACCTCAGTTCCCAAGAAATCATATCTACCAAGACATTTACCGAGGAAGCAGGAGGCAGTTTTGAAAGAAGCTATTCAGGAACAAATGGACCGT 150	00
TTTATACTTCAGGAACAAGCATAAGAAATATTGATCACTTTTGTCTTAATAAAAAAAGGAATCAAGCGTCTTGGATTCAAATAATAACAG 16(00
TUTTUTIGUALICIAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	
Fig. 2. Nucleotide sequence of the portion of 5.7 Md BamHI fragment encoding the α subunit. The non-coding strand is presented. Possible ribosome binding sites are underlined. The coding region is boxed.	

 Tobacco ch1.
 MYTIRADEISNIIRERIEOMAREVKIVMIGTVLOVGDIARTHGIDEVAGELVEPEEGT
 60

 <u>B. coli</u>
 IGLALNIESNINGGVIMGDILIGEGSSVRAFGRIAGTEVSVSDOVIRIHGIADCHOCEMISLPGAR
 120

 IGLALNIESNINGGVIMGDYADLAEGNEVKCTGRILEVEVGRGILGRVMMIGAPIDGRG
 120

 YALALNIERDSVGAVIMGPYADLAEGNEVKCTGRILEVEVGRGILGRVMMIGAPIDGRG
 120

 PLDHOGFSAVEAIAPGVIERGEVDOGVOGTVKCTGRILEVEVGRGILGRVMMIGAPIDGRG
 120

 TOTTILNOGGONVICVVAIGOKASVYETUTGLIAIDSMIPIGRGORELIIGDROTGKTAVA
 180

 PLDHOGFSAVEAIAPGVIERGEVDOGVOGTVRAVDSMIPIGRGORELIIGDROTGKTAVA
 180

 TOTTILNOGGONVICVVAIGOKASVAQVVTTIOBRGAMEYTIVVAETADSFATLOYLAP
 240

 VTGAALAEYTMYRERHTLIIVDDPSKOADAYROMSLLIRRPPGREAVICDVFYLHSRLLE
 300

 MPVALMGEYFRDRGEDALIIVVAIGOKASVAQVATROTSLLLRRPPGREAVICDVFYLHSRLLE
 300

 RAAKLSSSL------GE-----GSMTALPIVETOSCDVSAYIPTNVISITDGOIFISADL
 349

 FINGTIRPANNESISVSRVGGAAGTYINKULSGIRTALAQYRELAAFSOFASDLDAATON
 409

 FINGTIRPANNESISVSRVGGAAGTYINKULSGIRTALAQYRELAAFSOFASDLDDATRK
 469

 QIDHGGKVTELLKOKOGAPHENSVAODSLVLFAAERCYLADVELSKIGSTEAALLAMVDRDH
 469

 -FOFDEILSSTRTFTEERAFILKERSUGALLKEADGONDRFILQEQA
 507

Fig. 3. Comparison between the deduced smino scid sequence of the tobacco α subunit and that of the <u>E</u>. <u>coli</u> subunit (12). Homologies are indicated by boxes.

Table 1. Comparison between amino acid compositions of the estimated spinach α subunit (14) and the deduced tobacco α subunit. "Aspartic acid" includes asparagine and "Glutamic acid" includes glutamine.

amino acid residuespinach (estimated)tobacco (deduced)Lysine2018Histidine33Arginine3131
(estimated)(deduced)Lysine2018Histidine33Arginine3131
Lysine2018Histidine33Arginine3131
Lysine2018Histidine33Arginine3131
Histidine33Arginine3131
Arginine 31 31
Cysteine 2 1
Aspartic acid 38 34
Threonine 36 31
Serine 33 35
Glutamic acid 82 77
Proline 17 15
Glycine 43 39
Alanine 58 51
Valine 38 35
Methionine 11 11
Isoleucine 40 46
Leucine 53 50
Tyrosine 18 17
Phenylalanine 13 13

lst	second position								
	U		С		A		G		
υ	Phe	9	Ser	11	Tyr	14	Cys	1	U
	Phe	4	Ser	9	Tyr	3	Cys	0	C
	Leu	19	Ser	5	Term	1	Term	0	A
	Leu	8	Ser	4	Term	0	Trp	0	G
с	Leu	12	Pro	9	His	1	Arg	11	U
	Leu	3	Pro	3	His	2	Arg	3	C
	Leu	5	Pro	1	Gln	25	Arg	6	A
	Leu	3	Pro	2	Gln	11	Arg	0	G
A	Ile	26	Thr	10	Asn	14	Ser	6	U
	Ile	9	Thr	10	Asn	1	Ser	0	C
	Ile	11	Thr	7	Lys	16	Arg	9	A
	Met	11	Thr	4	Lys	2	Arg	2	G
G	Val	10	Ala	24	Asp	14	Gly	18	U
	Val	2	Ala	12	Asp	5	Gly	5	C
	Val	18	Ala	12	Glu	35	Gly	12	A
	Val	5	Ala	3	Glu	6	Gly	4	G

Table 2. Codon usage in the α subunit.

initiation codon.

The nucleotide sequence of the tobacco α gene shows 55% homology with that of the <u>E</u>. <u>coli</u> α gene. The longest identical sequence between the tobacco and <u>E</u>. <u>coli</u> α genes was found in positions 484-494. This sequence is GC-rich (72%) and located in the 1.65 Md sub-fragment. This seems to be a reason why λ <u>asn5</u> DNA hybridized to the 1.65 Md sub-fragment but not to the 0.90 Md sub-fragment. λ <u>asn5</u> DNA did not hybridize to the tobacco β and ε genes in our conditions.

Table 2 shows the codon usage in the α gene. This resembles the codon usage in the β and ε genes and in the large subunit gene of ribulose-1,5-bisphosphate carboxylase of tobacco chloroplasts (6, 8). Tobacco chloroplast DNA contains two genes for tRNA^{Val} (GUC codon) and one gene for tRNA^{Val} (GUA codon) which has a 571 bp intron (17, 18). It is therefore interesting that GUA codon is used much more than GUC codon in the four chloroplast polypeptides.

* To whom correspondence should be addressed.

ACKNOWLEDGEMENTS

We thank Mrs. Mie Kusuda for technical assistance. This work was supported in part by Grants-In-Aid from the Ministry of Education, Science and Culture of Japan.

REFERENCES

- 1. Westhoff, P., Nelson, N., Bünemann, H. and Herrmann, R.G. (1981) Current Genetics 4, 109-120.
- 2. Nelson, N., Nelson, H. and Schatz, G. (1980) Proc. Natl. Acad. Sci. USA 77, 1361-1364.
- 3. Krebbers, E.T., Larrinua, I.M., McIntosh, L. and Bogorad, L. (1982) Nucleic Acids Res. <u>10</u>, 4985-5002.
- 4. Zurawski, G., Bottomley, W. and Whitefeld, P.R. (1982) Proc. Natl. Acad. Sci. USA 79, 6260-6264.
- 5. Shinozaki, K. and Sugiura, M. (1982) Nucleic Acids Res. 10, 4923-4934.
- 6. Shinozaki, K., Deno, H., Kato, A. and Sugiura, M. submitted. 7. Kanazawa, H., Miki, T., Tamura, F., Yura, T. and Futai, M.
- (1979) Proc. Natl. Acad. Sci. USA 76, 1126-1130.
- 8. Shinozaki, K. and Sugiura, M. (1982) Gene 20, 91-102.
- 9. Sugiura, M. and Kusuda, J. (1979) Mol. Gen. Genet. 172, 137-141.
- 10. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA <u>74</u>, 560-564.
- 11. Kanazawa, H., Mabuchi, K., Kayano, T., Tamura, F. and Futai, M. (1981) Biochem. Biophys. Res. Commun. 100, 219-225.
- 12. Kanazawa, H., Kayano, T., Mabuchi, K. and Futai, M. (1981) ibid. 103, 604-612.
- 13. Gay, N.J. and Walker, J.E. (1981) Nucleic Acids Res. 9, 2187-2194.
- 14. Nelson, N. (1976) Biochim. Biophys. Acta 456, 314-338.
- 15. Seyer, P., Kowallik, K.V. and Herrmann, R.G. (1981) Current Genetics <u>3</u>, 189-204. 16. Tohdoh, N. and Sugiura, M. (1982) Gene <u>17</u>, 213-218.
- 17. Tohdoh, N., Shinozaki, K. and Sugiura, M. (1981) Nucleic Acids Res. 9, 5399-5406.
- 18. Deno, H., Kato, A. Shinozaki, K. and Sugiura, M. (1982) ibid. 10, 7511-7520.