Volume 279, number 2, 193-197 FEBS 09403 February 1991

ADONIS 001457939100092J

A heptapeptide repeat contributes to the unusual length of chloroplast ribosomal protein S18

Nucleotide sequence and map position of the *rpl*33-*rps*18 gene cluster in maize*

Wolfgang Weglöhner** and Alap Raman Subramanian

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnestrasse 73, D-1000 Berlin 33 (Dahlem), Germany

Received 21 November 1990; revised version received 19 December 1990

The *rpl*33-*rps*18 gene cluster of the maize chloroplast genome has been mapped and sequenced. The derived amino acid sequence of the S18 protein shows a 7-fold repeat of a hydrophilic heptapeptide domain, S K Q P F R K, in the N-terminal region. Such a sequence is absent in the *E.coli* S18 and in the chloroplast S18 of the lower plant liverwort. In tobacco and rice chloroplast S18 it is present 2 and 6 times, repectively. Thus a long N-terminal repeat (resembling in composition the large C-terminal heptapeptide repeat in the eukaryotic pol II) appears to be characteristic of monocot cereal S18.

Chloroplast genome; Ribosomal protein; *rps*18-*rpl*33; RNA binding motif; *Zea mays*

1. INTRODUCTION

Chloroplast ribosomes contain over 60 proteins [1] which are encoded in two compartments in the plant cell. The completely sequenced plastid genomes of two angiosperms (tobacco and rice) each encode 21 r-proteins [2,3]. A similar number is also found in the completely sequenced genome of a nonvascular lower plant [4], indicating thus a 2:1 distribution of this group of genes between the nuclear and chloroplast DNA in land plants.

Each of the identified r-protein sequences in the chloroplast DNA is a homologue of a corresponding *Escherichia coli* r-protein by the criteria of degree of sequence identity [2-4,5] and the ALIGN score (reviewed in [6]). In general, they have the same/similar chain lengths [6] as the homologous bacterial proteins [7]. However, there are a few exceptions to this rule; e.g. L22 and S18 which are considerably longer than their counterparts in *E. coli* [6]. In the case of S18, there is also a remarkable difference in length within the group of higher plants: tobacco and rice (dicot versus monocot), respectively, 101 and 159 amino acid residues [2,3].

In this paper we present the nucleotide sequence of the *rpl*33-*rps*18 gene cluster in maize chloroplast DNA and discuss the occurrence of a heptapeptide repeat, which first occurs in the tobacco S18 sequence and is repeated serveral fold in the rice and maize sequences.

2. MATERIALS AND METHODS

2.1. *Materials*

Restriction endonucleases, T4 DNA ligase, Klenow fragment of *E. coli* DNA polymerase I, bacterial alkaline phosphatase and nuclease Bal 31 were purchased from Boehringer Mannheim. The T7 sequencing kit, pT7T3-18U and pT7T3-19U were from Pharmacia and [α-^{35}S]dATP was bought from Amersham.

2.2. *Methods*

Plasmid DNA was isolated by the alkaline lysis method and purified by CsCl gradient centrifugation [8]. The DNA fragment containing *rpl*33 and *rps*18 was isolated by digesting maize Bam10 clone (pZmc525-2) with *Eco*RI and *Bam*HI, separated by agarose gel electrophoresis and electroelution (Biotrap: Schleicher & Schüll). Small scale isolation of DNA fragments was also done using QIAGEN-tip 20 columns (DIAGEN) and Gene CleanII Kit (Bio-101 Inc.). The *Eco*RI/*Bam*HI fragment was cleaved with appropriate restriction enzymes or progressively digested with nuclease Bal 31 (1-6 min, 30°C) and the resulting fragments were cloned into the multiple cloning site of pT7T3-18U/19U.

DNA was sequenced by the dideoxy chain termination method using double-strand as well as single-strand templates [9]. Oligonucleotide primers were synthesized on a DNA Synthesizer (Applied Biosystems, Model 380A) and purified by reversed phase HPLC separation. Sequence alignments, homology comparisons and secon-

Correspondence address: A.R. Subramanian, Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnestrasse 73, D-1000 Berlin 33 (Dahlem), Germany

Abbreviations: r-protein, ribosomal protein; S and L designates r-proteins of the small and large ribosomal subunit, respectively; pol II, DNA-dependent RNA polymerase II

* The nucleotide sequence presented in Fig. 1 has been submitted to the EMBL database under accession no. X56673

**This work represents part of the diploma thesis of W.W. submitted to the Friendrich-Alexander University, Erlangen-Nuernberg

dary structure predictions were performed on a VAX 8600/VMS computer using the UWGCG suite of programs [10].

The pZme523-2 clone is from a library of maize chloroplast DNA [5,11].

3. RESULTS AND DISCUSSION

The Bam10 fragment of maize chloroplast DNA is a 3.8 kbp segment in the large single copy region (LSC) at the map coordinates 39-43 kbp [11]. We previously determined by sequencing that the *rpl*33 gene is located near the 43 kbp end of Bam10. By analogy with the rice genome [3] the *rps*18 gene is expected downstream of *rpl*33 on this fragment. There is a single *Eco*RI site in Bam10, and the 1.3 kbp long fragment produced by *Eco*RI digestion, covering the map coordinates 41.7 kbp to 43.0 kbp was subcloned (Fig. 1). Further subclones and deletion clones produced by gradual Bal 31 digestion, together with suitable synthetic oligonucleotide primers (based on previous sequence information), were used for sequencing.

3.1. *Identification of L33 and S18 sequence*

Computer analysis of the nucleotide sequence (Fig. 1) showed two open reading frames (ORF). The 201 bp long ORF near the *Bam*HI (43 kbp) site begins with the start codon ATG and ends with the amber stop condon TAG. It encodes a polypeptide of 66 amino acid residues (M_r ≈ 7331.1) which reveals (Fig. 2) identity of 89%, 74% and 63% to the chloroplast r-protein L33 sequences of rice [3], tobacco [2] and liverwort [4], respectively; 58% to the cyanelle L33 of *Cyanophora paradoxa* [12] and 37% to *E. coli* L33 [7,13].

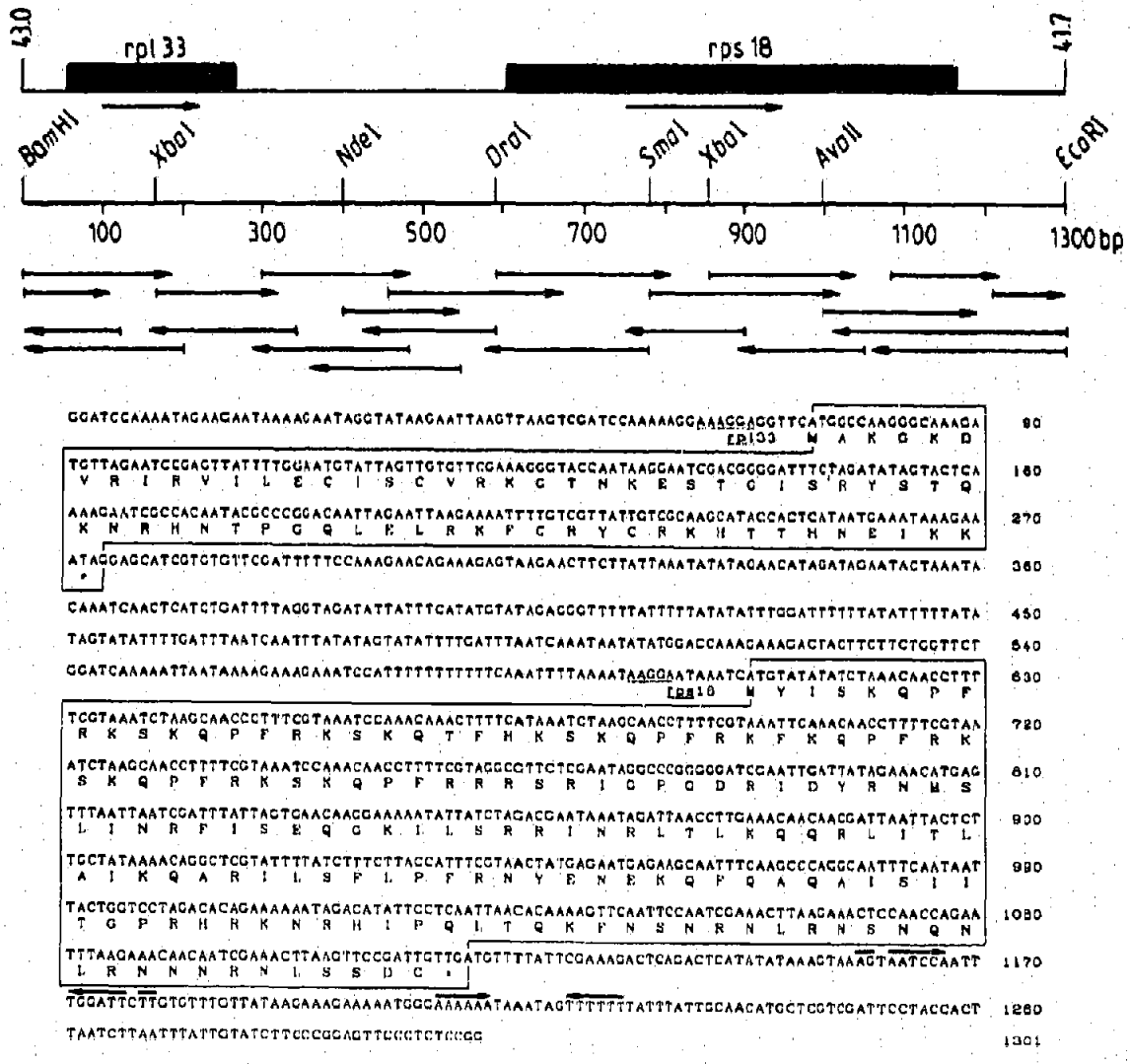

Fig. 1. Map position, sequencing strategy and nucleotide sequence of the *rpl*33-*rps*18 gene cluster, located in het Bam10 fragment [11] of maize chloroplast DNA. The coding (filled) and noncoding regions, restriction sites used for subcloning, and the extent and direction of each fragment sequenced are shown. The nucleotide sequence and the deduced amino acid sequence are given below (Shine-Dalgarno sequences are underlined and two putative termination stem-loops are shown overlined).

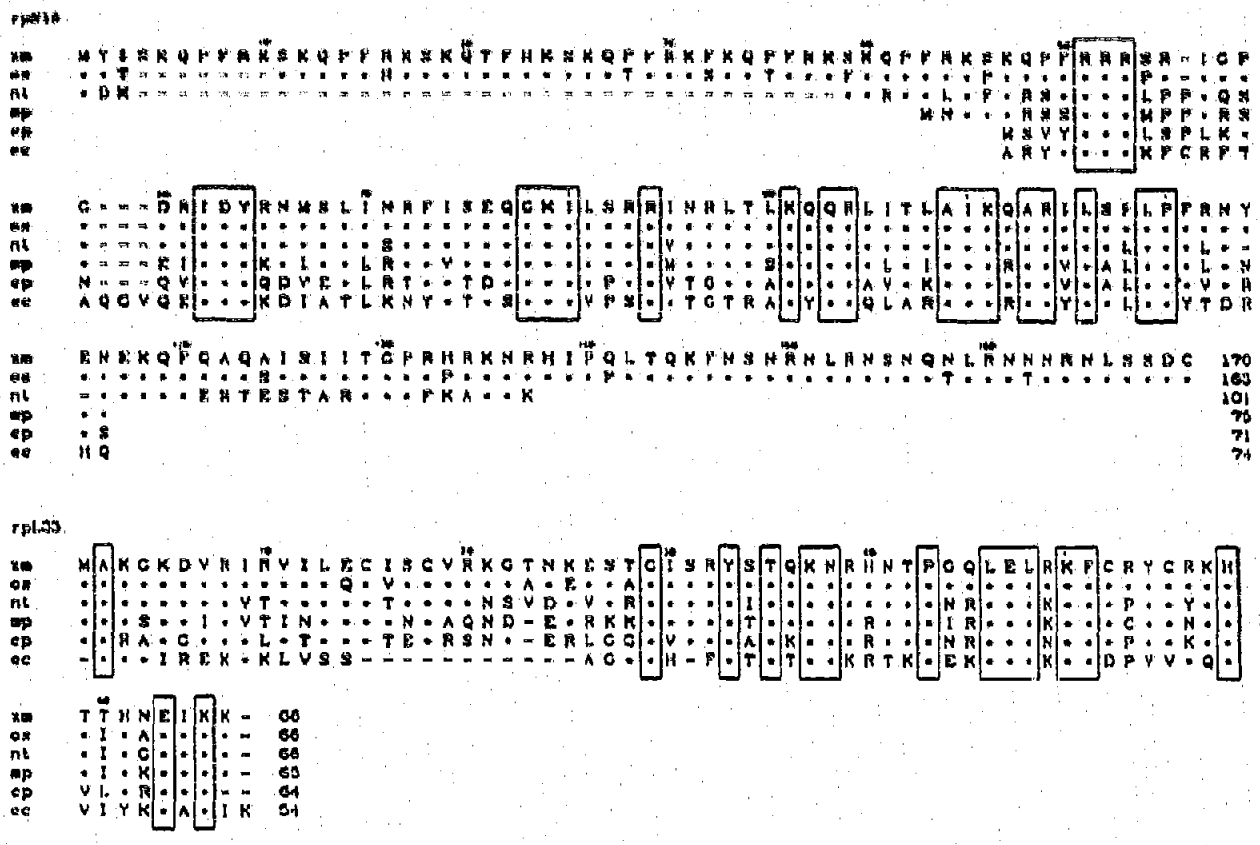

Fig. 2. Comparison between the deduced amino acid sequences of *Zea mays* chloroplast r-protein S18 and L33 and the corresponding proteins from rice (*Oriza sativa*: os), tobacco (*Nicotiana tabacum*: nt), liverwort (*Marchantia polymorpha*: mp), *Cyanophora paradoxa* (cp) and *Escherichia coli* (ec). Identical amino acids are marked by asterisks, and are boxed when invariant in all 6 sequences. The N-terminal alanine of *E. coli* S18 is acetylated [14].

The five organelle-encoded L33 sequences are of approximately the same size (64–66 residues) whereas the eubacterial L33 is significantly shorter (55 residues including fMet), mainly due to a 11 residue deletion near the N-terminal region (Fig. 2).

The second ORF, starting with ATG and ending with TGA, consists of 513 bp (171 codons) and encodes r-protein S18. The predicted 170-residue long protein (M_r = 20 601.5) shows (Fig. 2), in the homologous region, identity of 92%, 72%, 64%, 51% and 35% to the S18 sequences of rice [3], tobacco [2], liverwort [4], *Cyanophora paradoxa* [12] and *E. coli* [14], respectively.

The two ORFs of L33 and S18 are separated by a 333 bp long spacer sequence which displays a high AT-content (80%) normally found in the noncoding regions of the chloroplast genome [5]. The sequence downstream of *rps*18 contains two stretches of dyad symmetry which can be folded into stem-loop structures, indicating prokaryotic type [14] termination signals.

Table I

Selected amino acid residues in S18 sequences

rpS18	Length (AA)	Acidic (D+E)	Basic (H+K+R)	Net charge	Phe F	Pro P
Maize	170	6	47	+41	13	10
Rice	163	6	43	+37	12	11
Tobacco	101	7	26	+19	7	4
Liverwort	75	4	19	+15	2	3
Cyanelle	71	5	14	+ 9	2	4
E. coli	74	5	19	+14	3	2

Data from: maize, this paper; rice [3]; tobacco [2]; liverwort [4]; cyanelle [12]; *E. coli* [14]. fMet included in all except *E. coli* where the protein is characterized. D, Asp; E, Glu; H, His; K, Lys; R, Arg.

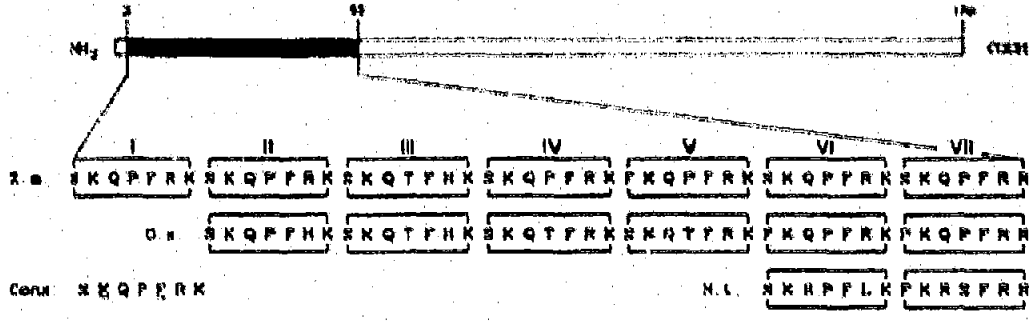

Fig. 3. The N-terminal heptapeptide repeat in maize and other higher plant S18 sequences. Cons = consensus sequence of the repeat peptide. Underlined are the two invariant residues; the other residues are represented in 10 to 13 of the 15 cases.

3.2. *Chain length increase in higher plant S18*

The S18 protein of *E. coli* [7,13] and the derived sequences from cyanelle and of the lower plant liverwort are of approximately the same size (71–75 residues) while those of the three higher plants are considerably longer (Fig. 2, Table I). The maize sequence is the longest one reported so far. This increase in size arises from variable length overhangs at both N- and C-termini. Both overhangs are particulary long for the two monocot cereal plants, rice and maize.

Inspection of the amino acid abudance in the six S18 sequences (Table I) revealed that they apparently belong to two classes. The bacterial, cyanelle and liverwort S18 sequences are similar in the numbers of acidic and basic amino acids, phenylalanine and proline. The two monocot S18 sequences on the other hand show a large, disproportionate increase in basic residues and in phenylalanine and proline. The S18 sequence of tobacco (a dicot plant) has an intermediate position between the two classes.

The increase in the basic residues is represented in both the N- and C-terminal overhangs, whereas the increase in proline and phenylalanine arises almost entirely in the N-terminal overhang.

3.3. *Heptapeptide repeat in higher plants*

The N-terminal overhang in maize S18 can be traced entirely to a 7-fold repeat of a heptapeptide with the consensus sequence S K Q P F R K. This sequence is absent in the S18s from the lower organisms. It is represented twice in tobacco and 6 times in rice S18 (Fig. 3). Secondary structure predictions show absence of α-helix and β-sheet in this domain.

The nucleotide sequence of this heptapeptide coding region also appears to be conserved. In maize it has the consensus sequence: $TC^{T}/_{C}AA^{A}/_{G}CAACCTTTTCGTAAA$.

Repeat sequences of varying lengths are found in some ribosomal proteins, the best known case being the four large (86 amino acids) repeats in het mRNA catching domain of *E. coli* r-protein S1 [16]. Also in spinach a repeat sequence has been found in the nuclear encoded chloroplast r-protein L21 [17]. This repeat is absent in the organelle encoded L21 of the lower plant liverwort.

The eukaryotic DNA-dependent RNA polymerase, pol II, contains the many-fold repeat of a heptapeptide, Y S P T S P S, at its C-terminal domain (CTD), which is essential for the polymerase function [18]. The S18 repeat has a resemblance to the latter in size and in the content of certain amino acids (P, S/T, and F/Y). The S18 repeat does not appear to resemble the conserved RNA-binding octapeptide motif (RNP-CS) found in many RNA-binding proteins [19].

The high sequence conservation of the S18 repeat (also on nucleotide level) would imply an as yet unkown function for this domain. It has been shown that sequential aromatic amino acids and positively charged residues are important in the binding to single-stranded nucleic acids, through intercalation of the aromatic residues with the nucleotide bases and the interaction of the positively charged groups with the negatively charged phosphodiester backbone [19,20]. Also proline seems to be involved in possible RNA binding motif. Serine/threonine, proline and one aromatic residue are common in the heptapeptide repeats in pol II and S18. Whether the repeats in chloroplast S18 have a specific function (i.e. binding to mRNA or single-strand regions of rRNA/chloroplast DNA) and, indeed, whether the repeats are actually present in the S18 protein from chloroplast ribosomes remains to be seen

Acknowledgements: We thank Prof. L. Bogorad (Harvard University) for the maize chloroplast library, and K. Giese and J. Schmidt for helpful discussions.

REFERENCES

[1] Subramanian, A.R., Stahl, D. and Prombona, A. (1990) in: The Molecular Biology of Plastids and the Photosynthetic Apparatus (Bogorad, L. and Vasil, I.K. eds) Academic Press, New York (in press).

[2] Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.-Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) EMBO J. 5, 2043–2049.

[3] Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.-R., Meng, B.-Y., Li, Y.-Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K. and Sugiura, M. (1989) Mol. Gen. Genet. 217, 185–194.
[4] Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. (1986) Nature 322, 572–574.
[5] Subramanian, A.R., Steinmetz, A. and Bogorad, L. (1983) Nucleic Acids Res. 11, 5277–5286.
[6] Subramanian, A.R., Smooker, P.M. and Giese, K. (1990) in: The Ribosome: Structure, Function and Evolution (Hill, W.E., Dahlberg, A., Garrett, R.A., Moore, P.B., Schlessinger, D. and Warner, J.R. eds) pp. 655–663, ASM Publications, Washington, DC.
[7] Wittmann-Liebold, B. (1986) in: Structure, Function and Genetics of Ribosomes (Hardesty, B. and Kramer, G. eds) pp. 326–361, Springer-Verlag, Berlin.
[8] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[9] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[10] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[11] Larrinua, I.M., Muskavitch, K.M.T., Gubbins, E.J. and Bogorad, L. (1983) Plant Mol. Biol. 2, 129–140.
[12] Evrard, J.L., Kuntz, M. and Weil, J.H. (1990) J. Mol. Evol. 30, 16–25.
[13] Wittmann-Liebold, B. and Pannenbecker, R. (1976) FEBS Lett. 68, 115–118.
[14] Yaguchi, M. (1975) FEBS Lett. 59, 217–220.
[15] Adhya, S. and Gottesman, M. (1978) Annu. Rev. Biochem. 47, 967–996.
[16] Subramanian, A.R. (1984) Trends Biochem. Sci. 9, 491–494.
[17] Smooker, P.M., Kruft, V. and Subramanian, A.R. (1990) J. Biol. Chem. 265, 16699–16703.
[18] Corden, J.L. (1990) Trends Biochem. Sci. 15, 383–387.
[19] Bandziulis, R.J., Swanson, M.S. and Dreyfuss, G. (1989) Genes and Development 3, 431–437.
[20] Helene, C. and Maurizot, J.-C. (1981) Crit. Rev. Biochem. 10, 213–258.