The chloroplast *infA* gene with a functional UUG initiation codon

Tetsuro Hirose[a], Takashi Ideue[a], Tatsuya Wakasugi[b], Masahiro Sugiura[a,*]

[a] *Center for Gene Research, Nagoya University, Nagoya 464-8602, Japan*
[b] *Department of Biology, Faculty of Science, Toyama University, Toyama 936-8555, Japan*

Received 4 January 1999; received in revised form 18 January 1999

Abstract All chloroplast genes reported so far possess ATG start codons and sometimes GTGs as an exception. Sequence alignments suggested that the chloroplast *infA* gene encoding initiation factor 1 in the green alga *Chlorella vulgaris* has TTG as a putative initiation codon. This gene was shown to be transcribed by RT-PCR analysis. The *infA* mRNA was translated accurately from the UUG codon in a tobacco chloroplast in vitro translation system. Mutation of the UUG codon to AUG increased translation efficiency approximately 300-fold. These results indicate that the UUG is functional for accurate translation initiation of *Chlorella infA* mRNA but it is an inefficient initiation codon.

Key words: Chloroplast; Initiation codon; Initiation factor 1; *Chlorella vulgaris*; Translation

1. Introduction

The chloroplast has its own genome and genetic system. Translation machineries in chloroplasts, e.g. 70S ribosomes, tRNAs and basal translation factors, are known to be structurally similar to those in prokaryotes, suggesting the translation mechanism in chloroplasts is similar to that in *Escherichia coli* [1]. On the other hand, spacing of Shine-Dalgarno (SD)-like sequences, which is commonly found 7 ± 2 nt upstream from initiation codons in *E. coli* mRNAs, is less conserved, ranging from -2 to -29 in chloroplast mRNAs, and some chloroplast mRNAs lack the SD-like sequence in their 5′-untranslated regions (UTRs) [2,3]. In the green alga *Chlamydomonas reinhardtii*, site-directed mutation analysis of its chloroplast SD-like sequences has yielded inconclusive results for importance of this sequence on translational initiation [4,5]. These observations suggest that there is a unique mechanism(s) to recognize correct translation initiation sites by chloroplast ribosomes [3,6–8]. Initiation codons of almost all chloroplast genes are ATG. Only three genes (*rps19*, *psbC* and *ycf15*) possess GTG as an initiation codon among the 79 genes/*ycf*s in the tobacco chloroplast genome [3]. The first codon of *psbL* and *ndhD* is ACG in the genome, while the ACG codons are converted to AUG codons post-transcriptionally by RNA editing in tobacco and several other plant species [9–11]. In *E. coli*, translation initiation is directed by an initiation codon [most commonly AUG (∼90%), less frequently GUG (∼9%) or UUG (∼1%) and in one case AUU] [12]. Conversion of GUG or UUG initiation codons to AUG drastically increases the translation efficiency [13,14], indicating that GUG and UUG are inefficient initiation codons.

Among soluble translation factors, only initiation factor 1 (IF1) is encoded in the chloroplast genome [15] and other factors are thought to be encoded in the nuclear genome in most land plants (see [1]). The IF1 gene (*infA*) is a constituent of the ribosomal protein L23 (*rpl23*) operon and transcribed as a polycistronic mRNA. The tobacco chloroplast genome is an exception and contains an *infA*-like sequence but no initiation codon [16]. Sequencing of the entire chloroplast genome from the green alga *Chlorella vulgaris* revealed the existence of an ORF starting with UUG [17]. The encoded protein is similar to the *E. coli* IF1 and the location of this ORF is the same as *infA* in the *rpl23* cluster of land plant chloroplasts. However, UUG codons were not reported as the initiation codon of all *infA* genes found in prokaryotes and chloroplasts. This is the first case of UUG being a potential initiation codon in chloroplast genes. Therefore, it is interesting to see whether UUG functions as a start codon in chloroplasts or the *Chlorella infA*-like ORF is a pseudogene like the tobacco *infA*-like segment [16]. Here we demonstrate that *Chlorella infA* mRNA is translated accurately from the UUG codon in vitro to produce a polypeptide of the expected size, indicating that UUG is a functional initiation codon in chloroplasts.

2. Materials and methods

Chlorella vulgaris C-27 cells were grown in M-4NA medium [18,19] at 28°C under the continuous light condition for 5 days with gentle rotation (120 rpm). Reverse transcription (RT)-PCR was carried out to detect *infA* transcripts. Total RNA from *Chlorella* cells were prepared as described [20] and treated with DNase I. Its cDNA was synthesized according to the instruction manual of the cDNA cycle kit (Invitrogen) using random primers. An *infA* portion (342 bp, positions -206–$+117$ with regard to UUG and an extra 19 bp from primers) was amplified by PCR and detected by 1% agarose gel electrophoresis using appropriate primers. The amplified product was then purified by 1% agarose gel electrophoresis and sequenced using the DNA cycle sequencing kit (Stratagene) as described [21]. For in vitro translation, a fragment of the *Chlorella infA* gene (positions -206–$+117$) (see Fig. 2A) was amplified by PCR and cloned into pBluescriptII SK$^+$ to construct an *infA-lacZ* fusion gene. Site-directed mutagenesis to produce M-AUG was carried out according to the instruction manual of Transformer Site Directed Mutagenesis kit Version 2 (Clontech) using primer pairs and the construct UUG as a template. Plasmid DNAs were linearized with *Bgl*II, extracted with phenol/chloroform and precipitated with ethanol. mRNA templates were synthesized from the plasmid DNA using T3 MEGASCRIPT (Ambion). Preparation of tobacco chloroplast extracts and in vitro translation reaction were carried out as described [22].

3. Results and discussion

3.1. *Chlorella chloroplast infA-like ORF starts from UUG and is transcribed*

The amino acid sequence deduced from the *infA* homologue in *Chlorella* chloroplasts has 56–66% identity with those from

*Corresponding author. Fax: (81) (52) 789 3081.
E-mail: h44979a@nucc.cc.nagoya-u.ac.jp

PII: S0014-5793(99)00123-4

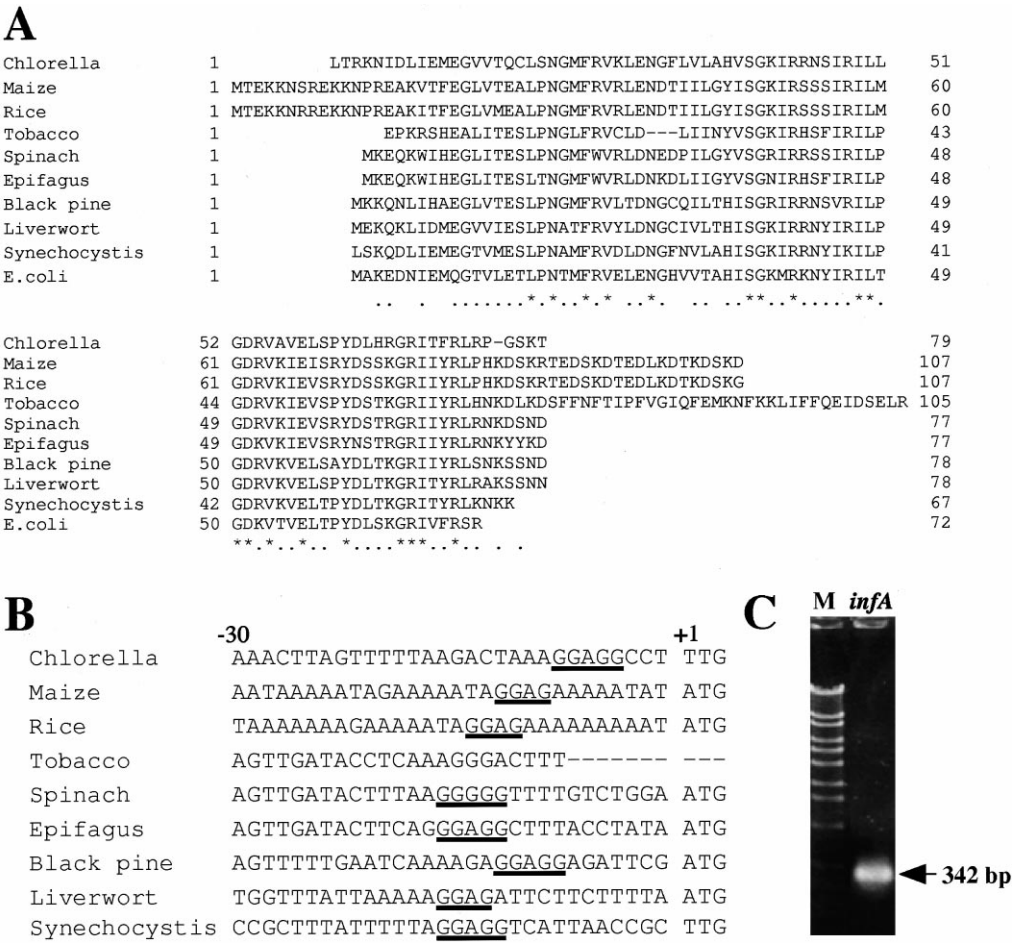

Fig. 1. A: Comparison of the deduced amino acid sequences of IF1 proteins. The first amino acid in *Chlorella* and *Synechocystis* is indicated as L derived from TTG codons. Dashes represent gaps introduced to maximize similality. The *infA*-like sequence in tobacco chloroplasts is a pseudogene (no initiation codon) [16]. Dots and asterisks under the alignment represent residues with conserved characters and identical residues, respectively. B: Comparison of the nucleotide sequences adjacent to putative initiaition codons of *infA*. SD-like sequences are inderlined. C: Detection of *infA* transcripts in *Chlorella* by RT-PCR (lane IF1). Lane M, size marker (λDNA digested with *Eco*T14I). Sequences were from the DDBJ databases.

E. coli and land plant chloroplasts (Fig. 1A). A typical SD-like sequence (GGAGG) is located 4 nt upstream from the UUG codon of *Chlorella infA* homologue (Fig. 1B), strongly suggesting that it uses UUG as the initiation codon. In order to examine whether the *infA* homologue is transcribed, RT-PCR analysis was carried out using the total RNA isolated from *Chlorella* cells. A single band of the expected size (342 bp) was clearly detected, indicating that this sequence is actually transcribed in *Chlorella* (Fig. 1C).

In all lineages of land plants, RNA editing occurs in chloroplasts [23] and C-to-U editing creates an initiation codon AUG from ACG in several chloroplast gene transcripts [9,10,24,25]. Although RNA editing has not been found so far in any algal chloroplasts, we checked the possibility that the UUG codon is converted to AUG post-transcriptionally by RNA editing. Direct sequencing of the above RT-PCR fragment indicated that no editing occurs in *infA*-like mRNAs (data not shown), indicating that mRNAs with the UUG codon are accumulated in *Chlorella* chloroplasts.

3.2. *Chlorella chloroplast infA mRNA is translated from the UUG initiation codon*

In order to analyze the role of UUG in the *Chlorella infA* mRNA, in vitro translation experiments were carried out using our chloroplast in vitro translation system prepared from tobacco chloroplasts [22]. The potential *infA* coding region (the first 39 codons) was extended by fusing a *lacZ* portion (87 codons) so as to facilitate the detection of an in vitro translation product (14.0 kDa if translation starts from UUG) by SDS-polyacrylamide gel electrophoresis (UUG in Fig. 2A). in vitro translation analysis revealed that the *infA-lacZ* mRNA is translated to produce a polypeptide of the expected size of 14 kDa (Fig. 2B). This result indicates that the UUG codon acts as the initiation codon of *infA* mRNA and hence the *infA* sequence is a genuine gene encoding IF1.

As mentioned above, *infA* genes in other organisms possess AUG as the initiation codon. The *Chlorella* UUG codon was substituted with AUG by site-directed mutagenesis (M-AUG in Fig. 2A). In vitro translation of the mutant *infA-lacZ* mRNA showed that translation takes place approximately 300-fold more efficiently than that of the wild-type mRNA with UUG (Fig. 2B, M-AUG). The level of translation from the mutant *infA* mRNA is roughly similar to those from several photosynthetic gene mRNAs of tobacco chloroplasts (e.g. *psbA* mRNA) which are efficiently translated in our in vitro translation system [22]. This seems to be due to

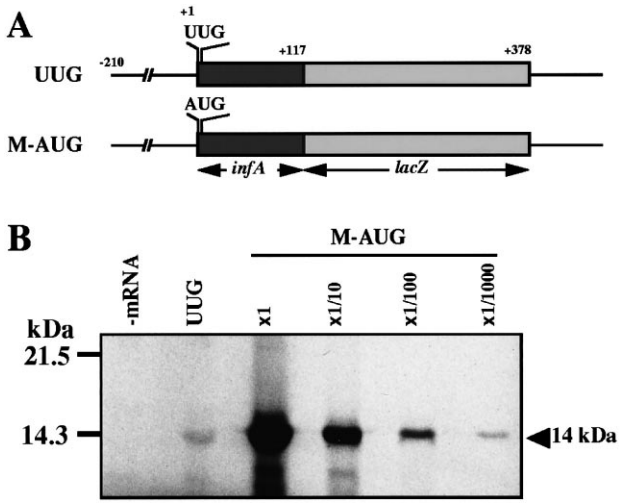

Fig. 2. Translation in vitro of *Chlorella infA* mRNAs. A: Schematic representation of two mRNA templates. B: Each mRNA template was incubated for 30 min at 30°C with the tobacco chloroplast in vitro translation system (25 μl). ^{35}S-labeled products were separated by 18% PAGE and visualized by Bioimaging analyzer BAS2000. Size markers are trypsin inhibitor (21.5 kDa) and lysozyme (14.3 kDa). In vitro translation samples from the M-AUG template were loaded with indicated dilution.

the presence of a typical SD-like sequence (GGAGG) in the appropriate position (−8 to −4) in its 5′-UTR (Fig. 1B). Typical SD-like sequences are found in the 5′-UTR of *infA* mRNAs in other plant species which possess AUG initiation codons (Fig. 1B), suggesting that *infA* mRNAs from other plant species are expected to be efficiently translated. The biological significance of the UUG initiation codon of *infA* gene in *Chlorella* chloroplasts is still unclear.

As the *Chlorella infA* mRNA is extremely low in translational initiation in tobacco chloroplast extracts, its efficient translation may be regulated by a protein factor(s) specific for *Chlorella* chloroplasts. The *E. coli* mRNA for RNase D possesses an UUG initiation codon and its translation is enhanced by interaction of ribosomal protein S1 to the upstream U-rich stretch [26]. In cyanobacteria, the *infA* homologue is present in the gene cluster corresponding to the chloroplast *rpl23* operon in *Synechocystis* PCC6803 [27] but not in *Synechococcus* PCC6301 [28]. The putative initiation codon of the *Synechocystis infA* homologue was assigned to be ATG at the 11th position with respect to *Chlorella* positions [27]. The alignment shown in Fig. 1A discloses the existence of an in-frame TTG eight codons upstream from this ATG. Based on the alignment together with our present result, this TTG is most likely to be the initiation codon for the *Synechocystis infA* gene. If this is the case, a factor(s) may be necessary to enhance its translation. A database search of possible initiation codons of *infA* homologues revealed one TTG (*Lactococcus lactis*) and three GTG (*Treponema pallidum*, *Haemophilus influenzae* and *Streptomyces coelicolor*) among 14 eubacterial species.

The *infA* gene in the green alga *Chlorella* may be an intermediate in the evolutionary process between functional genes and pseudogenes. The *infA* gene has been lost from chloroplast genomes in another green alga *Chlamydomonas*, and other algae *Odontella*, *Porphyra*, *Cyanophora*, *Cyanidosyzon* and *Euglena*, and the corresponding gene is thought to be translocated to the nuclear genome in these species [1]. It cannot therefore be excluded that an additional copy of the *infA* gene is present in the nuclear genome and IF1 is transported into chloroplasts in *Chlorella*. When corresponding genes are present both in nuclear and organelle genomes, those in organelle genomes tend to be silenced at the step of mRNA accumulation by an unknown mechanism [29]. In the case of *Chlorella*, *infA* in the *rpl23* operon was believed to be transcribed as a polycistronic mRNA and its mRNA was demonstrated to be accumulated. Therefore, the silencing of the organelle gene, if it is necessary, should be carried out at the post-transcriptional level. Replacement of the initiation codon from AUG to UUG may be a novel gene silencing mechanism.

In order to understand the biological significance of the UUG codon of *infA* gene, in vivo analysis using chloroplast transformation technology to *Chlorella* cells and searching for the counterpart in the nuclear genome will be necessary.

Acknowledgements: We thank Dr. M. Sugita for his support and useful discussions. This work was supported in part by a Grant-in-Aid from the Ministry of Education (Japan).

References

[1] Sugiura, M. (1992) Plant Mol. Biol. 18, 149–168.
[2] Bonham-Smith, P.C. and Bourque, D.P. (1989) Nucleic Acids Res. 17, 2057–2080.
[3] Sugiura, M., Hirose, T. and Sugita, M. (1998) Annu. Rev. Genet. 32, 437–459.
[4] Sakamoto, W., Chen, X., Kindle, K.L. and Stern, D.B. (1994) Plant J. 6, 503–512.
[5] Mayfield, S.P., Cohen, A., Danon, A. and Yohn, C.B. (1994) J. Cell Biol. 127, 1537–1545.
[6] Gillham, N.W., Boynton, J.E. and Hauser, C.R. (1994) Annu. Rev. Genet. 28, 71–93.
[7] Stern, D.B., Higgs, D.C. and Yang, J. (1997) Trends Plant Sci. 2, 308–315.
[8] Danon, A. (1997) Plant Physiol. 115, 1293–1298.
[9] Kudla, J., Igloi, G.L., Metzlaff, M., Hagemann, R. and Kössel, H. (1992) EMBO J. 11, 1099–1103.
[10] Neckermann, K., Zeltz, P., Igloi, G.L., Kössel, H. and Maier, R.M. (1994) Gene 146, 177–182.
[11] Hirose, T. and Sugiura, M. (1997) EMBO J. 16, 6804–6811.
[12] McCarthy, J.E.G. and Brimacombe, R. (1994) Trends Genet. 10, 402–407.
[13] Gold, L. (1988) Annu. Rev. Biochem. 57, 199–233.
[14] Zhang, J. and Deutscher, M.P. (1989) J. Biol. Chem. 264, 18228–18233.
[15] Sijben-Muller, G., Hallick, R.B., Alt, J., Westhoff, P. and Herrmann, R.G. (1986) Nucleic Acids Res. 14, 1029–1044.
[16] Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chungwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., 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.
[17] Wakasugi, T., Nagai, T., Kapoor, M., Sugita, M., Ito, M., Ito, S., Tsudzuki, J., Nakashima, K., Tsudzuki, T., Suzuki, Y., Hamada, A., Ohta, T., Inamura, A., Yoshinaga, K. and Sugiura, M. (1997) Proc. Natl. Acad. Sci. USA 94, 5967–5972.
[18] Tamiya, H., Iwamura, T., Shibata, K., Hase, E. and Nihei, T. (1953) Biochim. Biophys. Acta 12, 23–40.
[19] Yoshinaga, K., Ohta, T., Suzuki, Y. and Sugiura, M. (1988) Plant Mol. Biol. 10, 245–250.
[20] Shizadegan, M., Christie, P. and Seemann, J.R. (1991) Nucleic Acids Res. 19, 6055.
[21] Hirose, T., Fan, H., Suzuki, J.Y., Wakasugi, T., Tsudzuki, T., Kössel, H. and Sugiura, M. (1996) Plant Mol. Biol. 30, 667–672.
[22] Hirose, T. and Sugiura, M. (1996) EMBO J. 15, 1687–1695.

[23] Freyer, R., Kiefer-Meyer, M.C. and Kössel, H. (1997) Proc. Natl. Acad. Sci. USA 94, 6285–6290.
[24] Hoch, B., Maier, R.M., Appel, K., Igloi, G.L. and Kössel, H. (1991) Nature 353, 178–180.
[25] Wakasugi, T., Hirose, T., Horihata, M., Tsudzuki, T., Kössel, H. and Sugiura, M. (1996) Proc. Natl. Acad. Sci. USA 93, 8766–8770.
[26] Zhang, J. and Deutscher, M.P. (1992) Proc. Natl. Acad. Sci. USA 89, 2605–2609.
[27] Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. and Tabata, S. (1996) DNA Res. 3, 109–136.
[28] Sugita, M., Sugishita, H., Fujishiro, T., Tsuboi, M., Sugita, C., Endo, T. and Sugiura, M. (1997) Gene 195, 73–79.
[29] Covello, P.S. and Gray, M.W. (1992) EMBO J. 11, 3815–3820.