ISSN 0006-3509, Biophysics, 2006, Vol. 51, Suppl. 1, pp. S18–S22. © Pleiades Publishing, Inc., 2006.

Translation of Some Chloroplast Genes Is Checked to Allow for Splicing and Editing

A. V. Seliverstov, O. A. Zverkov, and V. A. Lyubetsky

Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, 127994 Russia
e-mail: slvstv@iitp.ru

Abstract—The expression of some genes in the chloroplasts of algae and plants is regulated by the binding of nuclear-coded proteins to the mRNA. Previously, we have found and studied the candidate sites for protein binding to the 5'-untranslated regions (UTRs) of *atpF*, *petB*, *clpP*, *psaA*, *psbA*, and *psbB* genes in algal and plant chloroplasts, and traced the evolution of these sites. It was suggested that some of these sites are involved in a mechanism suppressing translation until the gene splicing is completed. These sites in the 5'-UTRs of *petB*, *clpP*, *psaA*, and *psbA* genes include a conserved helix, while those of the two remaining sites consist of several boxes devoid of this helix. This paper considers the assumption that long hairpins in the 5'-UTR near the ribosome-binding site are involved in translation suppression until the completion of editing the *accD* and *atpH* gene transcripts in the chloroplast of plant genera *Adiantum* and *Anthoceros*.

Key words: translation, editing, regulation, chloroplast, multiple sequence alignment

DOI: 10.1134/S0006350906070049

INTRODUCTION

Gene expression regulation based on nuclear-encoded proteins binding to mRNA has been experimentally established for some chloroplast genes of a very small number of algae and plant species [1–3]. However, recently we have predicted [4] such regulation in a large number of algae and plants.

Since many protein-coding genes of chloroplasts contain introns or require RNA editing, their translation cannot be initiated immediately after transcription. It is known that the translation mechanisms in plant chloroplasts are very similar to those in bacteria. In particular, in the absence of special mechanisms, a ribosome moves along chloroplast mRNA immediately behind the RNA polymerase. However, the ribosome must not reach the exon end before the splicing is completed, or the mRNA editing site before the editing is completed. In rare cases, the editing of the starting codon acts as such a special mechanism, whereby AUG is formed from ACG to prevents the immediate start of translation [2]. However, e.g., the well-known genome of *Marchantia polymorpha* chloroplast gives no evidence of editing.

Previously, we developed an algorithm [5] that determined the candidate protein–RNA binding sites in 5'-untranslated regions (UTRs) of *atpF*, *petB*, *clpP*, *psaA*, *psbA*, and *psbB* genes in the chloroplasts of many algae and plants. This has been confirmed by multiple sequence alignments [4] in agreement with the aforementioned experimental data.

RESULTS

Chloroplast genomes were obtained from NCBI GenBank. Previously, we have found long conserved protein-binding sites in 5'-UTRs of *atpF* (subunit of ATP synthase), *petB* (cytochrome b6), *clpP* (proteolytic subunit of ATP-dependent Clp protease), *psaA* (P700 of photosystem I), *psbA* (D1 protein of photosystem II), and *psbB* (P680 of photosystem II) genes. In the annotation of *Amborella trichopoda psbA*, a short 5'-proximal exon is probably missing, which may account for the absence of the corresponding motif. The distribution of candidate protein-binding sites in the chloroplasts of algae and plants is presented in Table 1.

There is a high correlation between the presence of introns in genes and the existence of protein-binding sites for *atpF*, *clpP*, and *petB*. In contrast, no such correlation is observed for *psaA*, *psbA*, and *psbB*. Upstream of *clpP*, *petB*, *psaA*, and *psbA*, these sites contain a conserved RNA helix. A detailed investigation of these sites was reported elsewhere [4].

The transcripts of *accD* (beta-subunit of acetyl-CoA carboxylase) and *atpH* (subunit of ATP synthase) are subject to editing in the chloroplasts of plants (*Anthoceros* and *Adiantum*) [6]. This significantly changes the

Table 1. Distribution of conserved protein-binding sites in 5'-UTRs of six chloroplast genes

Section	Species	Gene					
		atpF	*clpP*	*petB*	*psaA*	*psbA*	*psbB*
Bacillariophyta	*Odontella sinensis*	–	–	–	+	+	–
Cryptophyta	*Guillardia theta*	–	–	–	+	+	–
Rhodophyta	*Cyanidioschyzon merolae*	–	–	–	–	+	–
	Cyanidium caldarium	–	–	–	–	–	–
	Porphyra purpurea	–	–	–	+	+	+
	Gracilaria tenuistipitata	–	–	–	–	+	–
Chlorophyta	*Chlamydomonas reinhardtii*	–	–	–	–s	+s	–
	Nephroselmis olivacea	–	–	–	+	+	+
Streptophyta	*Chaetosphaeridium globosum*	–	+s	–s	+	+	+
	Chara vulgaris	–s	–s	+s	+	–	+
	Mesostigma viride	–	–	–	+	–	–
	Zygnema circumcarinatum	–	–	+	–	+	–
Anthocerophyta	*Anthoceros formosae*	+s	+s	+s	+	+	+
Bryophyta	*Physcomitrella patens*	+s	+s	+s	+	+	+
Hepatophyta	*Marchantia polymorpha*	+s	+s	+s	+	+	+
Lycopodiophyta	*Huperzia lucidula*	+s	+s	+s	+	+	+
Pteridophyta	*Adiantum capillus-veneris*	+s	+s	–s	+	+	+
Psilophyta	*Psilotum nudum*	+s	+s	+s	+	+	+
Pinophyta	*Pinus thunbergii*	+s	+	+s	+	+s	+
Magnoliophyta	*Amborella trichopoda*	+s	+s	+s	+	–	+
	Arabidopsis thaliana	+s	+s	+s	+	+	+
	Atropa belladonna	+s	+s	+s	+	+	+
	Calycanthus floridus	+s	+s	+s	+	+	+
	Cucumis sativus	+s	+s	+s	+	+	+
	Epifagus virginiana	n	+s	n	n	n	n
	Lotus corniculatus	+s	+s	+s	+	+	+
	Nicotiana tabacum	+s	+s	+s	+	+	+
	Nymphaea alba	+s	+s	+s	+	+	+
	Panax ginseng	+s	+s	+s	+	+	+
	Spinacia oleracea	+s	+s	+s	+	+	+
	Oryza nivara	+s	+s	+s	+	+	+
	Oryza sativa	+s	+s	+s	+	+	+
	Triticum aestivum	+s	+s	+s	+	+	+
	Zea mays	+s	+s	+s	+	+	+

Note: (+) presence and (–) absence of candidate protein–mRNA binding sites; (s) gene contains introns; (n) gene is absent. The sites are presented in [4].

Table 2. Presence of powerful hairpins upstream of *accD* and *atpH* genes

Species	Gene			
	accD		*atpH*	
Anthoceros formosae	–7.0	++	–5.1	++
Adiantum capillus-veneris	–7.2	++	–5.2	++
Huperzia lucidula	–4.8	–	–2.9	–
Psilotum nudum	–0.8	–	–2.9	–
Pinus thunbergii	–3.6	–	–2.8	–

Note: Columns 2 and 4 give the minimum free energy of hairpins [kcal/mol] in a 40-nt-long mRNA region in 5'-UTR upstream of starting; columns 3 and 5 indicate the (++) presence and (–) absence of editing for the corresponding gene transcript.

codons of these transcripts, so that their correct translation is impossible until the completion of editing.

Since a gene immediately upstream of *accD* in plants belonging to genera *Anthoceros*, *Adiantum*, *Huperzia*, *Psilotum*, and *Pinus* encodes tRNA and the leader regions of *accD* in these plants have approximately equal lengths, it can be suggested that these regions experienced only point mutations in the course of evolution. In contrast, *accD* in all flower plants having common ancestry with the aforementioned five genera is immediately preceded by a protein-coding *rbcL*, which probably indicates that a chromosomal rearrangement in the 5'-UTR of *accD*. This gene is not found in the chloroplasts of *Marchantia polymorpha* and monocots. Therefore, it can be suggested that these five genera constitute a unique group that can be used for studying the mechanism of editorial stalling before *accD* and *atpH*.

No conserved helices in 5'-UTRs of *accD* and *atpH* were found in plants of all these genera. Then, it was suggested that a delay necessary for mRNA editing is provided by a certain mRNA hairpin structure. Accordingly, we have calculated the minimum free energy of such hairpins [kcal/mol] and identified them in a 40-nt-long mRNA region preceding the start codons of *accD* and *atpH*, as described in [7]. Long hairpins with low energies were found in this region for some *Anthoceros* and *Adiantum* plants, which covered the entire site proposed for ribosome binding, while the transcripts of these genes were subject to editing [6]. The minimum energy values are presented in Table 2 and the corresponding hairpins are depicted in Figs. 1–4. To check whether the region was chosen correctly, its length was varied up to 70 nt, but the results concerning the presence of low-energy hairpins covering the entire site of ribosome binding remained unchanged.

DISCUSSION

The conserved regions of 5'-UTRs of *atpF*, *petB*, *clpP*, *psaA*, *psbA*, and *psbB* chloroplast genes are prob-

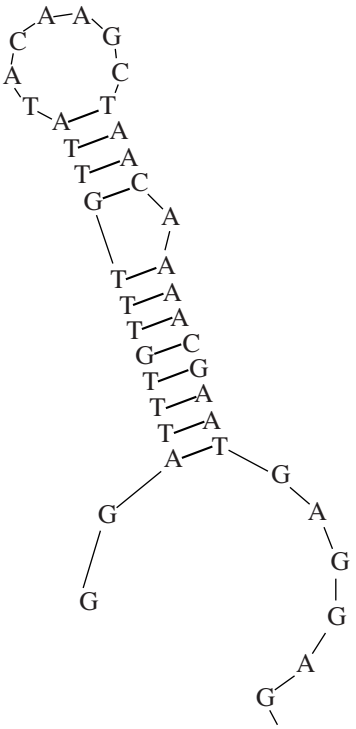

Fig. 1. mRNA structure upstream of *accD* gene in plants of the *Adiantum* genus.

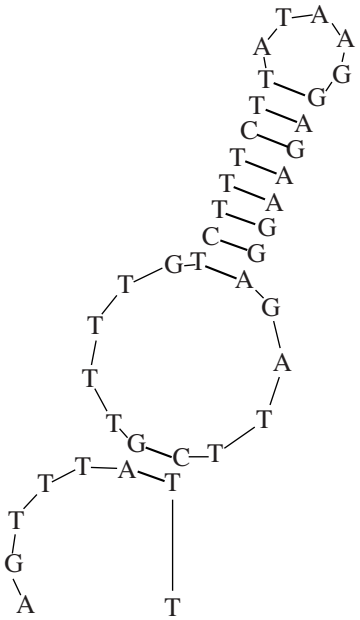

Fig. 2. mRNA structure upstream of *accD* gene in plants of the *Anthoceros* genus.

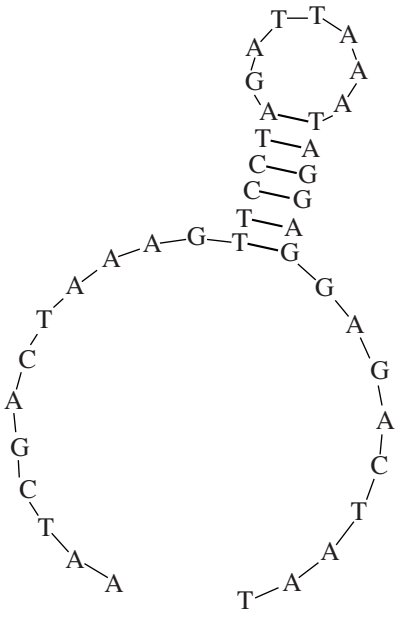

Fig. 3. mRNA structure upstream of *atpH* gene in plants of the *Adiantum* genus.

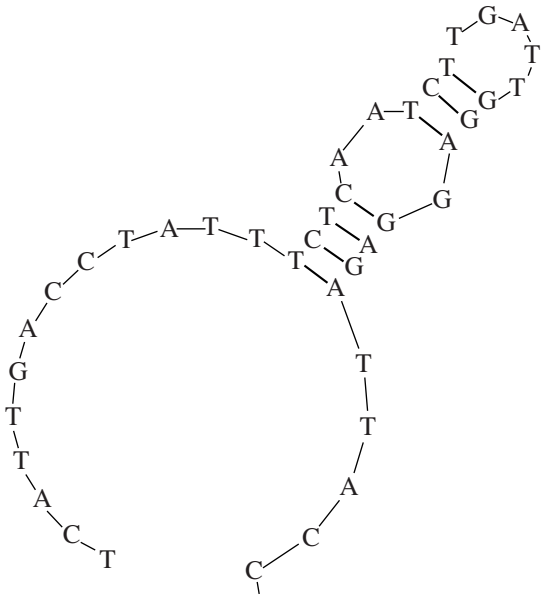

Fig. 4. mRNA structure upstream of *atpH* gene in plants of the *Anthoceros* genus.

ably involved in translation regulation. A conserved region in the *atpF* 5'-UTR is much longer than the AG-rich motif characteristic of the ribosome-binding site and contained in this conserved region. This fact suggests that the ribosome-binding site overlaps with a still longer site of regulatory protein binding, which prevents the translation onset and ensures its suppression until the completion of splicing.

The 5'-UTR of *petB* contains no typical sites of ribosome binding, but we have found a conserved RNA helix that can be related to processing of the 5'-UTR of mRNA. In all plants, *petB* genes have introns and the conserved regions revealed by our analysis, which suggests that this processing ensures suppression of the translation until splicing is completed. Except for *Adiantum*, where the conserved RNA region is absent, this delay in the translation onset can be explained by the fact that the start codon appears only after completion of mRNA editing.

The translation regulation of *psbA* was experimentally studied in *Clamydomonas reinhardtii*, where the transcription is constituitive and the translation is activated under illumination by a 47-kDa protein that forms a complex with some other proteins not bound directly to mRNA [1]. This complex breaks apart in the dark. The conservation observed in the leader regions for many algae and plants suggests that the translational regulation of *psbA* by this protein appeared long before the penetration of introns into this gene.

The conserved regions in 5'-UTRs of *psaA* and *psbA* were found at almost all orthologous genes (including those free of introns); hence, this regulation existed before the appearance of introns. It can be suggested that the penetration of introns into these genes became possible only because the mechanism of translation suppression was already operative and there were no obstacles to splicing.

The conserved regions in 5'-UTRs of *petB*, *clpP*, *psaA*, *psbA*, and *psbB* contain hairpins surrounded by conserved nucleotides, which is typical of many regulatory systems in bacteria [8]. It should be noted that this situation does not hold for *ycf3*, which (though containing introns) has a long 5'-UTR without a conserved region.

In the 5'-UTRs of *accD* and *petH* in *Anthoceros* and *Adiantum*, 40-nt regions contain long low-energy hairpins, and these genes are subject to editing. The low energy of these hairpins ensures their stability for a long time, and, according to our hypothesis, they prevent the onset of translation before the completion of editing. Table 2 gives the free energies of the same mRNA regions in representatives of all five genera. In *Huperzia lucidula*, these genes are not edited. Simultaneously (despite the presence of a hairpin with sufficiently low energy), the site of ribosome binding in the *accD* 5'-UTR is in the loop of this hairpin, and all other hairpins in this region have much higher energies (exceeding –1.7 kcal/mol). In the remaining two lines of Table 2, there are no low-energy hairpins in this region and these genes are not edited.

ACKNOWLEDGMENTS

The authors are grateful to M.S. Gelfand, K.Yu. Gorbunov, and L.Yu. Rusin for their help and fruitful discussions.

This study was supported in part by the ISTC (grant no. 2766).

REFERENCES

1. C. R. Hauser, N. W. Gillham, and J. E. Boynton, J. Biol. Chem. **271**, 1486 (1996).
2. W. Zerges, Biochemie, **82**, 583 (2000).
3. J. Nickelsen, Current. Gen., **43**, 392 (2003).
4. A. V. Seliverstov and V. A. Lyubetsky, J. Bioinform. Comput. Biol., **4** (4), 783 (2006).
5. V. A. Lyubetsky and A. V. Seliverstov, Inform. Protsessy, **4**, 241 (2004).
6. P. G. Wolf, C. A. Rowe, and M. Hasebe, Gene, **339**, 89 (2004).
7. M. Zuker, Nucl. Acids Res., **31**, 3406 (2003).
8. A. V. Seliverstov, H. Putzer, M. S. Gelfand, and V. A. Lyubetsky, BMC Microbiol., **5** (54), 14 (2005).