ISSN 0026-8933, Molecular Biology, 2009, Vol. 43, No. 4, pp. 552–556. © Pleiades Publishing, Inc., 2009.
Original Russian Text © T.A. Sadovskaya, A.V. Seliverstov, 2009, published in Molekulyarnaya Biologiya, 2009, Vol. 43, No. 4, pp. 599–604.

GENOMICS. TRANSCRIPTOMICS. PROTEOMICS

UDC 577.053

Analysis of the 5'-Leader Regions of Several Plastid Genes in Protozoa of the Phylum Apicomplexa and Red Algae

T. A. Sadovskaya[a] and A. V. Seliverstov[b]

[a] *Skryabin State Academy of Veterinary Medicine and Biotechnology, Moscow, 109472 Russia*

[b] *Kharkevich Institute of Information Transmission Problems, Russian Academy of Sciences, Moscow, 127994 Russia; e-mail: slvstv@iitp.ru*

Received May 19, 2008
Accepted for publication November 12, 2008

Abstract—Apicomplexan parasites contain so-called apicoplasts, which are similar to chloroplasts of red algae. Multiple alignments of the 5'-leader regions of plastid-encoded genes revealed several conserved noncoding regions in parasites as well as in red algae. The regions were assumed to be sites for RNA interactions with regulatory proteins. Conserved sites were found upstream of *ycf24*, which is required for [Fe-S] cluster development, and several other genes. In particular, a simultaneous regulation was predicted for *ycf24, rps4*, and *rpo*B in *Toxoplasma gondii.* The prediction agreed with the known data that apicoplasts are only required for a short time, but confer pathogenicity on *T. gondii.* Another site was predicted upstream of *rpo*B, which encodes the β subunit of RNA polymerase, in red algae *Porphyra* spp. and parasites *Eimeria tenella* and *Theileria parva.*

DOI: 10.1134/S0026893309040037

Key words: gene regulation, *ycf24, rpo*B, plastids, comparative genomics, Apicomplexa, algae

INTRODUCTION

The mechanisms of gene expression are similar in plastids and bacteria. In particular, the regulation of translation initiation is related to conserved regions in the 5'-untranslated regions of mRNAs in the vicinity of the initiator codon, where a ribosome-binding site occurs [1–3]. While the amino acid sequences of plastid-encoded proteins are highly conserved [4], the noncoding gene regions substantially vary even in closely related species, suggesting an important role in the regulation of gene expression for their conserved sites [3].

We examined the 5'-leader regions of all plastid genes in protozoa of the phylum Amicoplexa. Similar nucleotide sequences of the 5'-leader regions of plastid genes were considered for red algae and protozoa, whose plastids have related genomes [4] (table). Amixoplexan protozoa are of particular interest as animal pathogens. For instance, *Theileria* and *Babesia* ssp., which are transmitted by *Ixodes* ticks, cause several bovine diseases, including babesiosis caused by *B. bigemina* and *B. bovis*, theileriosis caused by *Th. annulata*, and East Coast fever caused by *Th. parva* [5]. *Eimeria tenella* causes eimeriosis in chicken, *Toxoplasma gondii* causes toxoplasmosis in cats and humans [6], and various *Plasmodium* species cause malaria in humans (*Pl. falciparum*) and rodents (*Pl. berghei, Pl. chabaudi*, and *Pl. yoelii*). The *B. bovis* and *Th. parva* genomes are highly similar [7].

With the exception of *Cryptosporidium parvum* [8], all of the organisms under study have plastids (apicoplasts) with a substantially reduced genome. For instance, *E. tenella* apicoplasts have certain ribosomal protein genes: *rps4* for the small-subunit protein S4; *tuf*A for the elongation factor; *rpoB*, *rpoC1*, and *rpoC2* for the β-, β'-, and β" subunits of bacterial-type RNA polymerase; *clpC*, which is important for the folding of nucleus-encoded proteins within plastids; and *ycf24*, which is orthologous to bacterial *sufB*. Other open reading frames are not conserved, and many of them have low homology to the genes for ribosomal proteins and RNA polymerase subunits.

While *ycf24* is positionally linked to *ycf16* in algae, the latter is absent from apicoplasts of Apicomplexa. The two genes code for SufB and SufC, which are involved in iron metabolism and the formation of FeS clusters [9, 10]. Many FeS proteins encoded by the nucleus are known to reach apicoplasts. For instance, this group includes ferredoxin (with a Fe2S2-type cluster) in *Pl. falciparum* [11], as well as GcpE and LytB, which are enzymes involved in isoprenoid synthesis, and LipA, which is involved in lipoic acid synthesis, in *T. gondii* [12]. The proteins are transferred unfolded across the plastid membrane, and their FeS clusters are formed within plastids. In red algae, the

Plastid genomes and genome regions extracted from the NCBI and Sanger Institute databases

Division, class	Species	NCBI	Sanger Institute
Apicomplexa	*Cryptosporidium parvum*	No plastids	
Coccidia	*Eimeria tenella*	NC_004823.1	
	Toxoplasma gondii	NC_001799.1	
Apicomplexa	*Plasmodium chabaudi*	CAAJ01003137.1	
Aconoidasida	*Pl. falciparum FCC-2/Hainan*	ABGW01002254.1	
	Pl. falciparum Santa Lucia	ABHA01004340.1	
	Pl. yoelii yoelii 17XNL	AABL01000014.1	
	Pl. berghei		Contig4648
	Babesia bovis T2Bo	NC_011395.1	
	Theileria annulata		Contig1014
	Theileria parva	NC_007758.1	
Bacillariophyta	*Heterosigma akashiwo*	NC_010772.1	
	Odontella sinensis	NC_001713.1	
	Phaeodactylum tricornutum	NC_008588.1	
	Thalassiosira pseudonana	NC_008589.1	
Cryptophyta	*Guillardia theta*	NC_000926.1	
	Rhodomonas salina	NC_009573.1	
Glaucocystophyceae	*Cyanophora paradoxa*	NC_001675.1	
Haptophyceae	*Emiliania huxleyi*	NC_007288.1	
Rhodophyta	*Cyanidioschyzon merolae*	NC_004799.1	
	Cyanidium caldarium	NC_001840.1	
	Gracilaria tenuistipitata	NC_006137.1	
	Porphyra purpurea	NC_000925.1	
	Porphyra yezoensis	NC_007932.1	
Xanthophyceae	*Vaucheria litorea*	NC_011600.1	

chloroplast genome contains genes for several other proteins with FeS clusters, such as the ChlL subunit (with a Fe4S4 cluster) of protochlorophyllide reductase, PetF ferredoxin (with a Fe2S2 cluster), apoprotein A1 of photosystem I PsaC, and ferredoxin-dependent glutamate synthase GltB. Note that cyanobacteria have two different glutamate synthases [13].

EXPERIMENTAL

Genome sequences. We examined the complete plastid genomes and genome fragments that contained genes orthologous to *ycf24*. The sequences were extracted from the NCBI and Sanger Institute databases by the BLAST program. The genomes and their fragments are listed in the table.

An *ycf24* ortholog was not found in the nuclear and plastid genomes of *Th. parva, B. bovis*, and *C. parvum*. The *Th. annulata* genome includes a pseudogene, which potentially codes for a protein with a low amino acid sequence homology to the N end of SufB (39 out of 140 amino acids (27%) are identical) and has an in-frame UAG codon. In addition, the C-terminal domain (Pfam accession no. PF01458), characteristic of SufB, was not found in *Th. annulata.*

Conserved regions were sought using a multiple sequence alignment program that consecutively constructs pairwise alignments with the use of a species tree known a priori (L.I. Rubanov, unpublished data). The resulting alignments were then corrected manually. Multiple sequence alignments of proteins were constructed using the ClustalX program [14]. The

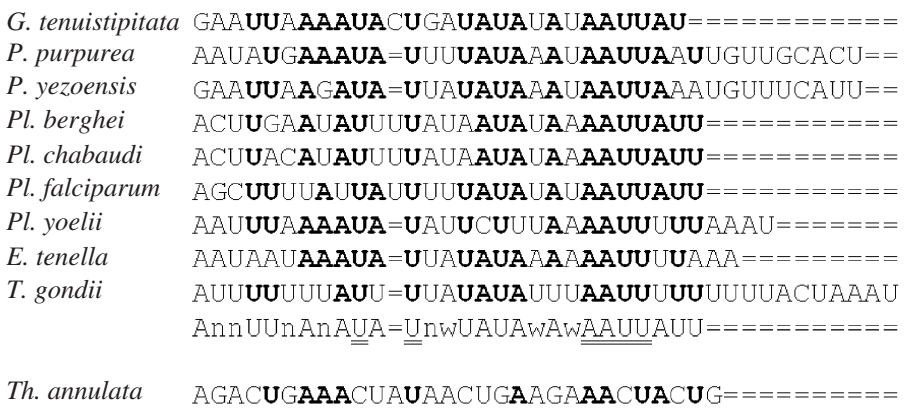

Fig. 1. Multiple sequence alignment of the 5'-leader regions of *ycf24* (at the top) and the corresponding *Th. annulata* pseudogene (at the bottom). The sequences are immediately upstream of the initiator codon. Conserved positions are in bold. The nucleotides found in more than 65% of the sequences are in capital letters. The absolutely conserved nucleotides (with the exception of *Th. annulata*) are underlined. The region under study is the same in *Pl. falciparum* FCC-2/Hainan and *Pl. falciparum* Santa Lucia.

annotation was additionally verified using the Pfam database.

RESULTS

Conserved sequences of the 5'-leader regions of *ycf24* were found in red algae of the genera *Porphyra* and *Gracilaria* and many Apicomplexa species (Fig. 1). The 5'-leader regions had nucleotide substitutions even in species of the same genus of *Porphyra.* It was unfeasible to extend the above alignments of the 5'-leader gene regions for the red algae *C. merolae* and *C. caldarium*; cryptophyte algae *Guillardia theta* and *Rhodomonas salina*; diatoms *Odontella sinensis, Phaeodactylum tricornutum*, and *Thalassiosira pseudonana; Vaucheria litorea; Cyanophora paradoxa*; and *Emiliania huxleyi.*

The 5'-leader sequences of *ycf24* were found to have a conserved AU-rich region with the consensus UUnAnAUA=UnwUAUAwAwAAUUAUU (Fig. 1). The 5'-leader region almost entirely covers the intergene region in the *Pl. chabaudi, E. tenella*, and *T. gondii* genomes, while an extended nonconserved region that does not overlap other genes is upstream of the 5'-leader region in the genomes of *Pl. yoelii* and red algae of the genera *Prophyra* and *Gracilaria.* The genome context and the length of the 5'-leader region substantially differ in red algae, *E. tenella*, and *T. gondii.* On the other hand, the conserved region is close to the initiator codon in *E. tenella* and *G. tenuistipitata* and is 13 and 11 nt away from the initiator codon in *T. gondii* and *Porphyra* spp., respectively. Hence, conservation of the region cannot be explained by a general similarity of the loci under study.

The putative protein-binding site upstream of the *ycf24* pseudogene substantially differs from the consensus in the *Th. annulata* genome.

When the consensus of the 5'-leader regions upstream of *ycf24* was used as a query, similar regions were found upstream of other genes. In *Porphyra*, the only new candidate for a putative regulation is *glt*B, whose 5'-leader region differs from the *ycf24* consensus in three positions. The 5'-leader regions differ from the consensus in at least four positions in all genes found for *C. merolae* and *G. tenuistipitata* (with the exception of *G. tenuistipitata ycf24*) and in three positions in *C. caldarium* (upstream of the *odp*AB operon, which codes for pyruvate dehydrogenase). This indicates that the putative regulatory mechanism is associated exclusively with *ycf24* in red algae.

The large-subunit rRNA and threonine tRNA genes are upstream of *rps4* and *ycf24* in the *T. gondii* and *E. tenella* apicoplast genomes. However, a difference in one nucleotide close to the coding region was observed in the 5'-leader regions upstream of *rps4* and *ycf24* only in *T. gondii.* In *E. tenella*, the 5'-leader regions upstream of *rps4* and *ycf24* differ in both length and composition. A similar nucleotide sequence with the identity in 25 out of the 40 nt shown in Fig. 2 was additionally found upstream of *rpo*B in the *T. gondii* apicoplast genome. However, the positions that are conserved in the 5'-leader regions of *T. gondii ycf24* and *rpo*B only poorly agree with the *ycf24* consensus of various species (Fig. 2).

A low-conserved region with the consensus UUAUUnAUnUAG=AwUAUwnAAAAnwAnU was found in the 5'-leader regions of *glt*B in the chloroplast genomes of the red algae *C. merolae, C. caldarium, P. purpurea*, and *P. yezoensis.* The *G. tenuistipitata* genome lacks a homologous region upstream of *glt*B. The plastid genomes of other species lack *glt*B. We did not detect any extended conserved sequences in the 5'-leader regions of *pet*F and *psa*C, which code for FeS proteins, in the chloroplast genomes of red algae.

ycf24 **AUUUUUUUAUUUUAUAUAUUUAAUUUUUUUUU**=**ACUAAAU**
rps4 **AUUUUUUUAUUUUAUAUAUUUAAUUUUUUUUUU**U**ACUAAAU**
rpoB **AUUUU**A**UUA**A**UUU**U**U**UA**U**U**UUAAU**AAAA**UUUU**UUAA**AA**U**U**

Fig. 2. Sequence alignment of the 5'-leader regions of *T. gondii* apicoplast genes. The sequences are immediately upstream of the initiator codon.

P. purpurea **A**AU**AUUAAA**C**U**C**UU**CA**A**UUU**C**A**GA**A**U**U**U**GCU**AU**A**AA**GGAG**AU**C**U**=
P. yezoensis **A**GU**AUUAAA**C**U**C**UU**CG**A**UUU**C**A**AA**A**U**U**U**GUU**AU**A**AA**GGAG**AU**C**U**=
E. tenella **A**UA**AUUAAA**U**U**A**UU**UA**A**AAUA**A**UU**A**A**U**A**UUA**AU**U**AA**UUUU**AU**A**U**A**
Th. parva **A**AUU**UUAAA**UAU**UU**AAGAG**U**UUUA**A**A**U**UUAA**AA**U**AA**UUUUUU**AA=
AnUAUUAAAyUnUUnAAwnUnAnAAwUUknwAUwAAkkwkAUmU=

Fig. 3. Multiple sequence alignment of the 5'-leader regions of *rpoB*. The sequences are immediately upstream of the initiator codon. The conserved nucleotides are in bold. The absolutely conserved positions in the consensus sequences are underlined.

A conserved region of another composition was found upstream of *rpo*B, whose 5'-leader regions in *Porphyra* and *E. tenella* are similar and share 22 out of 44 nt upstream of the initiator codon. A homologous region was detected upstream of *rpoB* in *Th. parva* (Fig. 3). The similarity of the corresponding regions is appreciably lower in other algae and *T. gondii*. We did not find any conserved sequence in the 5'-leader regions of other apicoplast genes.

DISCUSSION

A conserved region was found in all Apicomplexa species that have *ycf24* in our sample. The presence of conserved sequences in the 5'-leader regions upstream of the *ycf24* and *rpoB* plastid genes in Apicomplexa and red algae agrees well with a high homology of many of their proteins [4]. This finding supports the hypothesis that Apicomplexa are secondary symbionts and that their plastids originate from red algae related to *Porphyra*.

The conserved regions found upstream of *ycf24* probably serve as binding sites for unknown regulatory proteins. Indeed, their proximity to the initiator codon makes it possible to assume that the regions are involved in the regulation at the translation initiation level. Such regulation is characteristic of plant and algal chloroplasts [1–3]. Moreover, the stop codon of *ycf24* and the initiator codon of *ycf16* overlap by two nucleotides in *Porphyra*, suggesting translation reinitiation. Thus, translation of *ycf24* and *ycf16*, which code for SufB and SufC, respectively, is probably simultaneously regulated in *Porphyra*. A lack of SufB reduces the activity of its interaction partner SufC [9]. The translational regulation thus simultaneously saves translation expenditures and suppresses the relevant metabolic pathway.

Since Apicomplexa plastids lack genes for potential regulatory proteins, the predicted regulation is performed by nucleus-encoded proteins. The absence of such regulation in other secondary symbionts—cryptophytes, diatoms, *C. paradoxa*, and *E. huxleyi*—may reflect the lack of the regulation in some red algae.

It is possibly due to this regulation, which initially appeared in red algae, that *ycf24* remains in the apicoplast genome of some Apicomplexa species, while its homologs occur in the nucleus in *Arabidopsis* [15].

The putative common regulation of *ycf24* and *rps4* expression makes it possible to assume that translational suppression involves not only *ycf24*, but also all other mRNAs in *T. gondii* apicoplasts because of a decrease in ribosomal protein S4. If the common regulation extends, even if less efficiently, to *rpoB*, then transcription is also suppressed. It is known that apicoplasts are absolutely essential for *T. gondii* to enter a new host cell [16]. This observation is supported by experiments with the nonpathogenic *T. gondii* mutant that differs from the wild type in lacking the nucleus-encoded enzyme that reaches the apicoplast and plays a role in fatty acid synthesis [17]. In other words, *T. gondii* pathogenicity is associated with synthetic processes in apicoplasts. On the other hand, *T. gondii* has only one pyruvate dehydrogenase complex, which is located in apicoplasts and, unlike in plants, is absent from mitochondria [18].

The putative binding site upstream of the *ycf24* pseudogene in the *Th. annulata* genome substantially differs from the consensus, but most of the substitutions are G for A and C for U. This makes it possible to assume that the regulation has disappeared only recently, so that mutations have not altered the site as of yet.

In *E. coli*, the translational regulation of the *rpoB* mRNA depends on region –29/+70 relative to the

rpoB initiator codon [19]. The conservation of the 5'-leader region of *rpoB* in plastids suggests a similar regulation. As in the case of *rpos4*, the regulatory region substantially differs between *T. gondii* and *E. tenella* apicoplasts.

To summarize, our findings predict a conserved translational regulation for plastids of red algae and Apicomplexa. The regulation differs between *T. gondii* and *E. tenella*, indicating that the role of apicoplasts may differ even in closely related species. The difference may be important for developing new drugs that affect the apicoplasts of parasites.

ACKNOWLEDGMENTS

This work was supported by the International Science and Technology Center (ISTC grant no. 3807).

REFERENCES

1. Zerges W. 2000. Translation in chloroplasts. *Biochimie*. **82**, 583–601.
2. Nickelsen J. 2003. Chloroplast RNA-binding proteins. *Curr. Genet*. **43**, 392–399.
3. Seliverstov A.V., Lyubetsky V.A. 2006. Translation regulation of intron containing genes in chloroplasts. *J. Bioinformat. Comput. Biol*. **4**, 783–793.
4. Lemieux C., Otis C., Turmel M. 2007. A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. *BMC Biology*. **5**, 1–17.
5. Balashov Yu.S. 1998. *Iksodovye kleshchi – parazity i perenoschiki infektsii* (Ixodid Ticks: Parasites and Infection Vectors), St. Petersburg: Nauka.
6. Beyer T.V. 1992. Opportunistic infections of protozoan nature. *Tsitologiya*. **34**, 26–27.
7. Brayton K.A., Lau A.O.T., Herndon D.R., et al. 2007. Genome sequence of *Babesia bovis* and comparative analysis of apicomplexan hemoprotozoa. *PLoS Pathogens*. **3**, e148.
8. Zhu G., Marchewka M.J., Keithly J.S. 2000. *Cryptosporidium parvum* appears to lack a plastid genome. *Microbiology.* **146**, 315–321.
9. Rangachari K., Davis C.T., Eccleston J.F., Hirst E.M.A., Saldanha J.W., Strath M., Wilson R.J.M. 2002. SufC hydrolyzes ATP and interacts with SufB from *Thermotoga maritima*. *FEBS Letters*. **514**, 225–228.
10. Eccleston J.F., Petrovic A., Davis C.T., Rangachari K., Wilson R.J.M. (Iain). 2006. The kinetic mechanism of the SufC ATPase. *J. Biol. Chem*. **281**, 8371–8378.
11. Vollmer M., Thomsen N., Wiek S., Seeber F. 2001. Apicomplexan parasites possess distinct nuclear-encoded, but Apicoplast-localized, plant-type ferredoxin-NADP+ reductase and ferredoxin. *J. Biol. Chem*. **276**, 5483–5490.
12. Thomsen-Zieger N., Schachtner J., Seeber F. 2003. Apicomplexan parasites contain a single lipoic acid synthase located in the plastid. *FEBS Letters*. **547**, 80–86.
13. Muro-Pastor M.I., Florencio F.J. 2003. Regulation of ammonium assimilation in cyanobacteria. *Plant Physiol. Biochem*. **41**, 595–603.
14. Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G. 1997. The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. **25**, 4876–4882.
15. Xu X.M., Adams S., Chua N.-H., Moller S.G. 2005. AtNAP1 represents an atypical SufB protein in *Arabidopsis* plastids. *J. Biol. Chem*. **280**, 6648–6654.
16. Wilson R.J.M. (Iain), Rangachari K., Saldanha J.W., Rickman L., Buxton R.S., Eccleston J.F. 2003. Parasite plastids: Maintenance and functions. *Phil. Trans. R. Soc. Lond*. **B. 358**, 155–164.
17. Mazumdar J., Wilson E.H., Masek K., Hunter C.A., Striepen B. 2006. Apicoplast fatty acid synthesis is essential for organelle biogenesis and parasite survival in *Toxoplasma gondii*. *Proc. Nat. Acad. Sci. USA*. **103**, 13 192–13 197.
18. Fleige T., Fischer K., Ferguson D.J.P., Gross U., Bohne W. 2007. Carbohydrate metabolism in the *Toxoplasma gondii*Apicoplast: localization of three glycolytic isoenzymes, the single pyruvate dehydrogenase complex, and a plastid phosphate translocator. *Eucaryotic Cell*. **6**, 984–996.
19. Passador L., Linn T. 1992. An internal region of *rpo*B is required for autogenous translational regulation of the subunit of *Escherichia coli* RNA polymerase. *J. Bacteriol.* **174**, 7174-7179.