FEBS Letters 382 (1996) 60–64 FEBS 16762

Localization and expression of the closely linked cyanelle genes for RNase P RNA and two transfer RNAs

Michael Baum, Astrid Schön*

Institut für Biochemie, Bayerische Julius-Maximilians-Universität, Biozentrum, Am Hubland, D-97074 Würzburg, Germany

Received 23 January 1996; revised version received 31 January 1996

Abstract **The genomic region encoding the RNA subunit of the cyanelle RNase P has been characterized. rnpB, which has no homologue in chloroplasts, is flanked by two tRNA genes on the complementary DNA strand. Transcriptional control elements of all three genes have been experimentally determined. Comparison of the sequenced region with the corresponding loci of chloroplast genomes from vascular plants suggests that major inversions may have led to a possible loss or severe truncation of the RNase P RNA coding region during the course of plastid evolution.**

Key words: Cyanelle; Evolution; Ribonuclease P; Promoter; $tRNA^{Arg}$; $tRNA^{His}$

1. Introduction

RNase P, the ubiquitous enzyme essential for 5′ end maturation of tRNA, consists of essential RNA and protein subunits in all cases investigated to date (reviewed in [1,2]). The data available on the subunit composition of RNase P from chloroplasts, however, suggest that the plastid enzyme might lack an RNA component entirely [3].

Photosynthetic organelles are thought to have evolved from cyanobacteria-like prokaryotes living in endosymbiosis with heterotrophic eukaryotes. One putative intermediate in plastid evolution is the unicellular photosynthetic protist *Cyanophora paradoxa*. The pigment and cell wall composition of its photosynthetic organelles or cyanelles resemble that of free-living cyanobacteria, whereas the dramatically reduced size and organization of their genome is more similar to chloroplasts [4,5]. Sequence derived phylogenetic trees position the cyanelle in a coherent group together with cyanobacteria and green plastids [6,7]; several genes that have been lost from chloroplast genomes early in plant evolution are present in the cyanelle genome. Taken together, these facts support the idea that this organelle might be considered a remnant of an early stage of plastid evolution, and thus serve as a model to study the molecular evolution of RNase P, an ancient and ubiquitous RNA enzyme, in plastids of different origin.

We have recently characterized the RNase P-associated RNA in the cyanelle of *C. paradoxa* [8]. This first description of an RNase P RNA in a photosynthetic organelle might help us understand the molecular events that led to the postulated "reduced" form of RNase P in chloroplasts, provided that the gene locus and possible genome rearrangements can be identified. To answer these questions, we have sequenced the region encoding cyanelle RNase P RNA. The complete rnpB sequence, as well as two tRNA genes located on the complementary strand, are presented. Their experimentally determined transcription start sites add substantially to the information currently available on functionally characterized promoters in cyanelles. The location and relative orientation of these tRNA genes varies widely between chloroplast genomes, suggesting that the entire region might have undergone several rearrangements during plastid genome evolution.

2. Materials and methods

2.1. Cell growth, cyanelle preparation, and nucleic acid isolation

C. paradoxa strain PCC C.7201 was obtained from the Pasteur Culture Collection and grown with aeration in Stanier's medium [4]. Preparation of cyanelles and nucleic acids was as described [8].

2.2. Oligonucleotides

The following oligonucleotides were synthesized on a Gene-assembler Plus (Pharmacia) and purified on a denaturing 20% PAA gel prior to use for sequencing or primer extension: **cprp5′**: cgaattcTTGAGGAAAGTCCGGGCTC; **cprp5′-anti**: GAGCCCGGACTTTCCTCA; **CyGSP2**: GTGTTTTTCTGTGGCACTATCCTC; **CyGSP3**: gcgaattcGGAAAGTAATTAACCTTTTTTCTTTC; **CyHis1**: CTGCCTTAACCACTTGGC; **CyHis2**: gcgaattcTAAGTATTTATGGCGATCGTAG; **CyArg1**: GTGCTCTAGTCCACTGAGC (sequences in lower case indicate restriction sites introduced for cloning purposes). Labelling at the 5′-end with T4 PNK and [γ-^{32}P]ATP was performed as described [9], and oligonucleotides were separated from unincorporated radioactivity by chromatography on Nucleobond AX columns (Macherey-Nagel).

2.3. Cloning and sequencing procedures

Recombinant DNA manipulations were performed as described [9]. A library of *Eco*RI-*Hin*dIII digested cyanelle DNA constructed in pUC19 was screened by colony hybridization with CyGSP2 as a probe; a 510 bp *Dra*I subclone was derived from the initial 4.3 kb insert of pCyEH4.3 to facilitate sequencing (see Fig. 1). The specific primers cprp5′-anti, Cy-GSP3, CyHis1, CyHis2, CyArg1 and the commercial universal primers (forward "−40" and reverse) were used together with the Sequenase 2.0 kit (United States Biochemical) to determine the contiguous sequence of both strands.

The University of Wisconsin GCG program package (Version 8) was used for data analysis. Homology searches were performed with the programs FASTA or BESTFIT.

2.4. Primer extension analysis

5 µg total cyanelle RNA were annealed to $3{\cdot}10^5$ cpm 5′-labelled CyArg1 or cprp5′-anti, or to 1 pmol unlabelled CyHis1 in a total volume of 6 µl. After denaturation for 10 min at 85°C, hybridization proceeded for 4 h at 37°C in a buffer consisting of 20 mM HEPES pH 8, 1 mM Na_2EDTA and 400 mM NaCl. After ethanol precipitation, samples were redissolved in 19 µl RT-Mix (1×AMV reaction buffer, 0.5 mM dNTP, and 5 µCi [α-^{32}P]dATP in the case of CyHis1), supplemented with 12 U AMV-RT (Promega), and incubated for 2 h at 42°C. Reactions were terminated by phenol extraction, precipitated, dissolved in sequencing gel loading buffer and analysed by electrophoresis on 6% or 15% PAA gels. In adjacent lanes, sequencing reactions performed with pCyEH4.3 and the appropriate primers were loaded to determine the exact length of the extension products.

*Corresponding author. Fax: (49) (931) 888 4028.

SSDI S0014-5793(96)00148-2

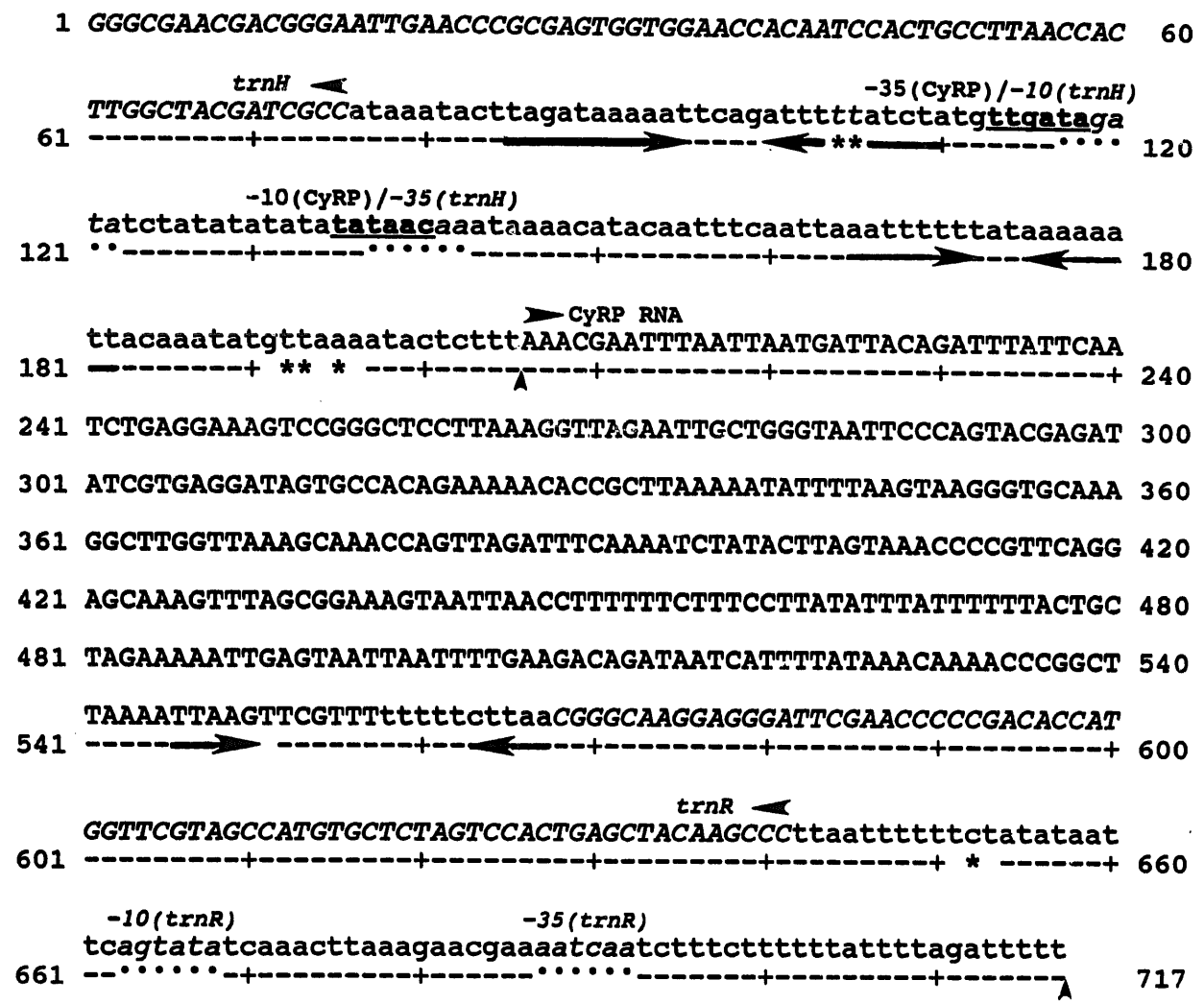

Fig. 1. Sequence of the trnH-rnpB-trnR region of the cyanelle genome. Nucleotides present in mature RNAs are in capitals, noncoding regions in lower case; the tRNA genes transcribed from the complementary strand are in *italics*. The respective gene names and direction of transcription are indicated close to the 5′ ends. Transcription initiation sites are marked by asterisks, and putative promoter elements are underlined (CyRP RNA) or identified by dotted lines (tRNAs). Inverted repeats are drawn as solid black arrows pointing towards their respective axis of symmetry. The *Dra*I restriction sites used for subcloning are marked by small vertical arrowheads.

3. Results

3.1. The gene for cyanelle RNase P RNA

Cyanelle rnpB is colinear with the RNase P RNA [8]; the coding region extends from position 206 to 554 (the top strand in Fig. 1). The 5′ end of the RNA seems to be heterogenous, as deduced from primer extension (Fig. 3A): a cluster of bands is visible between positions 204 to 209, the strongest being U_{205}. Sequence determination of the RNA 5′ terminus by RACE supports the adjacent A_{206} as the predominant

Table 1
Functional promoters in the *Cyanophora paradoxa* cyanelle

Gene	−35 box		−10 box		Reference
pcyA	TTGTAA	− 23 −	TATAAA	− 5 −	[22]
pcyA	ttgata	− 31 −	TATAAA		
apcA	TTCATA	− 18 −	TATTAT	− 6 −	[22]
apcA	ttgaac	− 31 −	TATTAT		
rps12	TTGACA	− 18 −	TATAGT	− 7 −	[23]
rnpB	TTGATA	− 16 −	TATAAC	− 50?−	This study
trnH	TTGTTA	− 18 −	TATCAA	− 7 −	This study
		− 14 −	tatcta	− 11 −	This study
trnR	TTGATT	− 18 −	TATACT	− 10 −	This study

Only promoters deduced from experimental evidence (primer extension, nuclease protection) are included. Sequences in lower case relate to elements that, although they conform to prokaryote-type consensus sequences, are thought to be nonfunctional in cyanelles because of inappropriate spacing. The two "−10" boxes possible for trnH are marked by a contiguous line in Fig. 1.

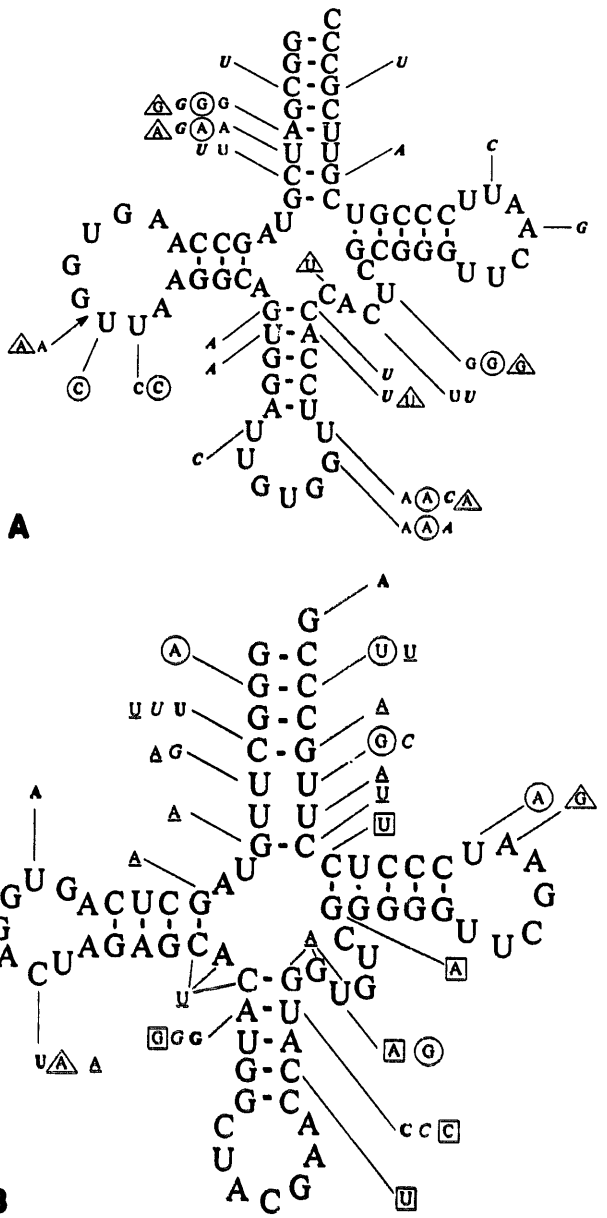

Fig. 2. Similarity of the two tRNAs to corresponding chloroplast genes. (A) $tRNA^{His}$: the cyanelle sequence was compared to trnH from chloroplasts of mono- and dicotyledoneous plants (standard lettering), *Pinus* (circled), *Marchantia* (triangles), and *Euglena* (italics). (B) $tRNA^{Arg}$(ACG) was compared to the corresponding chloroplast genes from *Marchantia* (bold), *Chlorella* (italics), *Chlamydomonas* (circled), *Cyanidium* (triangles), *Euglena* (squares), and the tRNA sequence from *Codium* (underlined). The cyanelle tRNAs are represented as their unmodified transcripts.

mature 5′ end (data not shown). The analysis also reveals longer transcripts starting around position 195. The closest prokaryote-type transcriptional control elements are a "−10 box" at nucleotide 135 and a "−35 box" at position 113 (Table 1). They are, however, separated by 50 bp from the proposed transcription start site. An inverted repeat of unknown function is located between this putative promoter region and the transcription start site. The 3′ end of rnpB is defined by a stretch of T residues, which are part of a possible stemloop structure.

3.2. The gene for $tRNA^{His}$

trnH is located on the DNA strand complementary to rnpB, between positions 75 and 1 (Fig. 1). If compared to the known chloroplast genes for $tRNA^{His}$, a number of differences are evident (Fig. 2A): two striking peculiarities are the presence of a U–U "pair" in the acceptor stem and of G_{37}, which is A in all other cases [10]. The expression of trnH was verified by primer extension. The 5′ end of the mature tRNA is located at G_{75} (Fig. 3B), giving rise to an 8 bp acceptor stem as found in all $tRNA^{His}$ species. Longer transcripts start around A_{105}, downstream from a promoter present at nucleotides 113 (or 117) to 137 (−10 and −35 element, respectively; see Table 1). Similar to the rnpB gene, a stemloop structure may form between promoter and coding region; in the case of trnH, the transcription start site is located within the upstream half of the stem. Transcription termination likely occurs within a stretch of Ts immediately downstream of the trnH coding region (not included in Fig. 1).

3.3. The gene for $tRNA^{Arg}(ACG)$

The structural gene extends from position 642 to 568 (Fig. 1) on the same strand as trnH. trnR differs at several positions from the corresponding plastid sequences, with the highest similarity to the gene from the red alga *Cyanidium* (Fig. 2B). The most peculiar feature of the cyanelle gene, two consecutive U–U "pairs" in the acceptor stem, is shared only by that sequence; in all chloroplast genes, the base pair 5–68 is a standard Watson-Crick or wobble pair [10]. Transcription initiates at G_{652} (Fig. 3C); possible "−10" and "−35" elements are found at positions 663 and 687, respectively. The 3′ end of the $tRNA^{Arg}$ coding region is in close proximity to the end of RNase P RNA; a stem-loop structure may thus also be formed by the $tRNA^{Arg}$ primary transcript.

4. Discussion

4.1. Structure and localization of the gene locus coding for cyanelle RNase P RNA: implications for the evolution of plastid RNase P

Chloroplasts are the only organelles where the composition of RNase P from both an RNA and a protein subunit is still controversial [3,11]. Cyanobacteria, the closest living relatives of the bacterial ancestors of plastids, contain RNase P RNAs which are similar to those of other bacteria and exhibit ribozyme activity in vitro [12–14]. RNase P from *Cyanophora paradoxa* cyanelles contains an RNA component resembling the cyanobacterial counterparts; however, it deviates from a proposed consensus sequence at several conserved positions, is not catalytically active on its own, and may thus represent an ancient intermediate in plastid RNase P evolution [8]. As no longer sequence homologies to rnpB can be identified in chloroplast genomes [15], we have addressed the question of molecular events that might have led to a loss of the rnpB gene during evolution of plastids, by determining the sequence and genomic position of cyanelle rnpB and comparison to sequenced chloroplast genomes. The cyanelle RNase P RNA gene is flanked, on the complementary DNA strand, by the single gene for $tRNA^{His}$ about 150 bp upstream of the rnpB coding region, and the $tRNA^{Arg}$-ACG gene immediately downstream thereof (Fig. 1). Hybridization studies had localized rnpB on restriction fragments that had previously been shown to hybridize to $tRNA^{His}$ and $tRNA^{Arg}$ ([8,16], and

unpublished data). Cyanelle RNase P RNA and the flanking tRNA genes thus reside between 37.7 and 40.4 kbp on the genomic map, a region that also contains the petB and petD genes, as concluded from hybridization data [5,17].

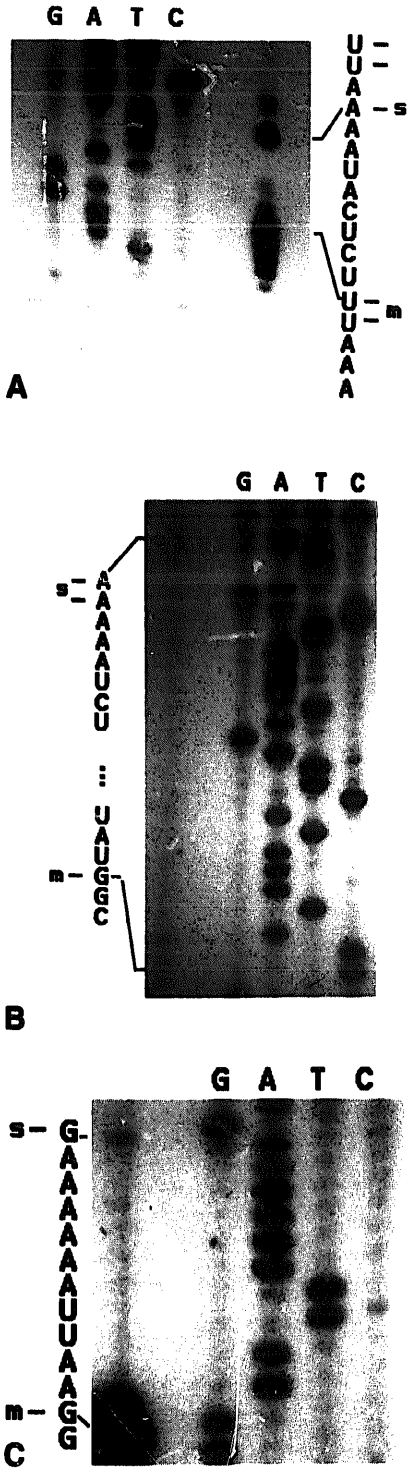

Fig. 3. Determination of transcription start sites for rnpB (panel A), trnH (B), and trnR (C). Sequencing reactions used as size standards are identified by their specificities (GATC); the RNA sequence corresponding to the end of the transcript is given adjacent to the lane containing the primer extension product. Positions of mature RNA ends (m) and transcription start sites (s) are indicated. The signals relating to trnH and trnR appear slightly shifted relative to the sequencing reactions due to uneven heat distribution in the gel.

Comparison of the relative orientation and map position of these cyanelle genes to the location of the corresponding tRNA genes and the petB–petD operons on chloroplast genomes of most vascular plants reveals a striking variability ([18–21]; Fig. 4): whereas trnR(ACG) is close to the gene for 5s rRNA within the IR region and thus present in duplicate on the chloroplast genomes examined, $tRNA^{His}$ is encoded in the central part of the LSC, close to the gene for psbA, in the liverwort *Marchantia*. In *Nicotiana*, psbA and trnH have moved — possibly by an inversion — close to the border of IR_A; in monocots, trnH is incorporated into the repeat unit, whereas psbA has remained outside the border of IR_A. The petB–petD operons have retained their position relative to IR regions in all chloroplast genomes; psbA, which is close to trnH in chloroplasts, is located almost 50 kbp away on the LSC region in cyanelles. Thus, we must assume several major recombination events after the divergence [7] of ancestral cyanelles and chloroplasts, which finally led to the apparent loss in chloroplasts of an essential RNA gene that is well preserved in cyanelles.

4.2. *Transcription start sites and the identification of promoter and terminator elements*

Transcript analysis of RNase P RNA yields, in addition to the predominant mature 5′ end, several weaker signals between T_{192} and A_{195} (Fig. 3A). However, no "prokaryote-type" promoter sequences are in close proximity: the nearest elements resembling transcriptional control sites are located between positions 135 and 113 ([22,23]; Table 1). As the distance of this possible promoter from the 5′ end of the detected transcripts is about 50 nucleotides, it can not be decided at present whether these signals correspond to primary transcripts or to processed, longer variants of RNase P RNA. A similarly pronounced heterogeneity of RNA 5′ ends has been described before in RNase P from yeasts [24,25]. Transcription termination of rnpB likely occurs within a run of uridines contained within a stemloop, a structure frequently described near the end of cyanelle genes and thought to function as a transcription terminator [26,27].

trnH and trnR(ACG) are quite similar to the corresponding sequences from other plastids (Fig. 2). Strikingly, both genes lead to tRNAs with U–U mismatches in the acceptor stem, at positions where most of the other genes contain Watson-Crick or wobble base pairs. Interestingly, the sequence from the green alga *Codium*, the only sequence determined on the tRNA level, possesses a completely paired acceptor stem [28]. Considering the strict requirements known to rule tRNA recognition, it is tempting to speculate whether those mismatched acceptor stems are "repaired" by editing, a process that has been described for plant mitochondria and chloroplasts, but seems to be confined to mRNAs in the latter [21]. In mitochondria, however, tRNA editing has frequently been found to restore base pairing [29,30]. Alternatively, the tRNAs might be modified at the corresponding positions to pseudouridine "pairs", a feature that has been described in the anticodon stem of eukaryotic $tRNA_m^{Met}$ [10] and seems to contribute to the stability of helical structures.

The promoters governing expression of the two tRNAs have been deduced from the transcriptional start sites (Fig. 3B,C). For trnR, the −10/−35 boxes conform to elements

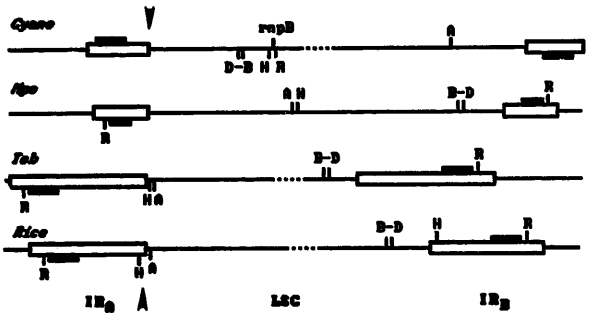

Fig. 4. Genomic localization of trnR(ACG), trnH, and rnpB on the cyanelle and comparison to chloroplast genomes. The cyanelle genome (Cyano) is compared to chloroplast genomes of *Marchantia* (Mpo), *Nicotiana* (Tob), and rice. IR regions are drawn as boxes, rRNA operons as black bars. Genomes are aligned such that the junction between IR_A and LSC is at the same position (marked by an arrow). Transcription is from left to right for the genes written above, and from right to left for those below the line representing the chromosome which has been truncated within LSC (dotted line) to accomodate the different genome sizes. The orientation of the trnH-rnpB-trnR region and of petB-petD relative to the genome has been chosen arbitrarily. Gene abbreviations are: A (psbA), B–D (petB-petD), H (trnH), R (trnR), rnpB.

determined in previous studies, in both sequence and spacing ([22,23]; Table 1). The $tRNA^{His}$ gene allows for some ambiguity in the exact position of the −10 element: it may be located at nucleotide 117 or 113, the latter option being in better agreement with the spacing found in other cases (Table 1). Interestingly, this position overlaps significantly with the promoter postulated for rnpB.

The proximity of trnR and rnpB 3′ ends implies that transcription possibly continues into the complementary strand of the rnpB coding region. Convergent transcription has been described for protein coding genes of the cyanelle [26,27] and seems to be widespread in the cyanelle genome. As the available data on gene expression in cyanelles are restricted to a few protein genes, our characterization of the trnH-rnpB-trnR region renders it unique as the only known structural RNA genes that are convergently transcribed. The experimental data leading to the identification of sequence elements for transcriptional control of RNA genes substantially increase our yet scarce knowledge on functionally determined promoters in these photosynthetic organelles.

Acknowledgements: This work was supported by a grant from the Deutsche Forschungsgemeinschaft to A.S. Part of the CyRP RNA sequence has been presented at the 16th International tRNA Workshop, Madison 1995. The sequence described in this paper is accessible under the Genbank number X89853.

References

[1] Altman, S., Kirsebom, L. and Talbot, S. (1993) FASEB J. 7, 7–14.
[2] Pace, N.R. and Brown, J.W. (1995) J. Bacteriol. 177, 1919–1928.
[3] Wang, M.J., Davis, N.W. and Gegenheimer, P. (1988) EMBO J. 7, 1567–1574.
[4] Herdman, M. and Stanier, R.Y. (1977) FEMS Microbiol. Lett. 1, 7–12.
[5] Löffelhardt, W. and Bohnert, H.J. (1994) Molecular biology of cyanelles. in: The Molecular Biology of Cyanobacteria (Bryant, D.A. ed.), vol. 4, pp. 65–89, Kluwer Academic Publishers.
[6] Giovannoni, S.J., Turner, S., Olsen, G.J., Barns, S., Lane, D.J. and Pace, N.R. (1988) J. Bacteriol. 170, 3584–3592.
[7] Turner, S., Burger-Wiersma, T., Giovannoni, S.J., Muur, L.R. and Pace, N.R. (1989) Nature 337, 380–382.
[8] Baum, M., Cordier, A. and Schön, A. (1996) J. Mol. Biol., in press.
[9] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: a Laboratory Manual, 2nd Ed., Cold Spring Harbor Press.
[10] Sprinzl, M., Dank, N., Nock, S. and Schön, A. (1991) Nucleic Acids Res. 19 (suppl.), 2127–2171.
[11] Yamaguchi-Shinozaki, K., Shinozaki, K. and Sugiura, M. (1987) FEBS Lett. 215, 132–136.
[12] Banta, A.B., Haas, E.S., Brown, J.W. and Pace, N.R. (1992) Nucleic Acids Res. 20, 911.
[13] Vioque, A. (1992) Nucleic Acids Res. 20, 6331–6337.
[14] Pascual, A. and Vioque, A. (1994) Biochim. Biophys. Acta 1218, 463–465.
[15] Schön, A. (1996) Mol. Biol. Rep., in press.
[16] Kuntz, M., Crouse, E.J., Mubumbila, M., Burkard, G., Weil, J.-H., Bohnert, H.J., Mucke, H. and Löffelhardt, W. (1984) Mol. Gen. Genet. 194, 508–512.
[17] Bohnert, H.J., Michalowski, C., Bevacqua, S., Mucke, H. and Löffelhardt, W. (1985) Mol. Gen. Genet. 201, 565–574.
[18] Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S.I., Inokuchi, H. and Ozeki, H. (1986) Nature 322, 572.
[19] Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chungwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torozawa, 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.
[20] 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., Shinozaki, K. and Sugiura, M. (1989) Mol. Gen. Gen. 217, 185–194.
[21] Maier, R.M., Neckermann, K., Igloi, G.L. and Kössel, H. (1995) J. Mol. Biol. 251, 614–628.
[22] Lemaux, P.G. and Grossman, A.R. (1985) EMBO J. 4, 1911–1919.
[23] Kraus, M., Götz, M. and Löffelhardt, W. (1990) Plant Mol. Biol. 15, 561–573.
[24] Krupp, G., Cherayil, B., Frendewey, D., Nishikawa, S. and Söll, D. (1986) EMBO J. 5, 1697–1703.
[25] Zimmerly, S., Gamulin, V., Burkard, U. and Söll, D. (1990) FEBS Lett. 271, 189–193.
[26] Michalowski, C.B., Pfanzagl, B., Löffelhardt, W. and Bohnert, H.J. (1990) Mol. Gen. Genet. 224, 222–231.
[27] Neumann-Spallart, C., Brandtner, M., Kraus, M., Jakowitsch, J., Bayer, M.G., Maier, T.L., Schenk, H.E.A. and Löffelhardt, W. (1990) FEBS Lett. 268, 55–58.
[28] Francis, M.A., Suh, E.R. and Dudock, B.S. (1989) J. Biol. Chem. 264, 17243–17249.
[29] Marechal-Drouard, L., Ramamonjisoa, D., Cosset, A., Weil, J.H. and Dietrich, A. (1993) Nucleic Acids Res. 21, 4909–4914.
[30] Binder, S., Marchfelder, A. and Brennicke, A. (1994) Mol. Gen. Genet. 244, 67–74.