well as the calcium-ligation scheme are effectively identical to those seen in the crystal structure of the corresponding *N. n. atra* complex (*12, 13*).

The distal end of the calcium-binding loop bears bee venom's single oligosaccharide, which is N-glycosylated to the side chain of Asn13. Although the inhibitor's *sn-3* amino alcohol makes contact with the proximal sugar, the carbohydrate moiety must be refined before the nature of this contact can be inferred. Electron density for only the proximal three sugars can be found; the more distal saccharides extend into bulk solvent and are apparently disordered. The function of this carbohydrate in enzymatic action (if any) is unclear. Non-glycosylated bee-venom PLA_2 occurs naturally as a minor variant but retains normal activity (*14*).

When compared with its Class I/II relatives, the bee-venom crystal structure provides a clear indication of those components that are essential to PLA_2 function. In a companion article, we interpret these findings in a unifying structural model for PLA_2 catalysis (*13*).

REFERENCES AND NOTES

1. R. L. Heinrikson, E. T. Krueger, P. S. Keim, *J. Biol. Chem.* **252**, 491 (1977).
2. R. A. Shipolini, G. L. Callewaert, R. C. Cottrell, C. A. Vernon, *Eur. J. Biochem.* **48**, 465 (1974).
3. K. Kuchler, M. Gmachi, M. J. Sippl, G. Kreil, *ibid.* **184**, 249 (1989).
4. J. M. Maragnore, R. A. Poorman, R. L. Heinrikson, *J. Protein Chem.* **6**, 173 (1987).
5. A. Weber, H. Schroder, K. Ihalberg, L. Marz, *Allergy* **42**, 464 (1987).
6. G. Lin, J. Noel, W. Loffredo, H. Z. Stable, M. Tsai, *J. Biol. Chem.* **263**, 13208 (1988).
7. B. W. Dijkstra, K. H. Kalk, W. G. J. Hol, J. Drenth, *J. Mol. Biol.* **147**, 97 (1981).
8. S. Brunie, J. Bolin, D. Gewirth, P. B. Sigler, *J. Biol. Chem.* **260**, 9742 (1985).
9. E. A. M. Fleer, H. M. Verheij, G. H. de Haas, *Eur. J. Biochem.* **82**, 261 (1978). When designating a functionally important residue, the first number refers to the position in the bee-venom sequence. The second number is the corresponding sequence number in the Class I/II superfamily from the common numbering system of Renetseder *et al.* (*10*).
10. R. Renetseder, S. Brunie, B. W. Dijkstra, J. Drenth, P. B. Sigler, *J. Biol. Chem.* **260**, 11627 (1985).
11. M. Elder, P. Hitchcock, R. Mason, G. G. Shipley, *Proc. R. Soc. London Ser A* **354**, 157 (1977); R. H. Pearson and I. Pascher, *Nature* **281**, 499 (1979); H. Hauser, I. Pascher, R. H. Pearson, S. Sundell, *Biochim. Biophys. Acta* **650**, 21 (1981).
12. S. White, D. L. Scott, Z. Otwinowski, M. H. Gelb, P. B. Sigler, *Science* **250**, 1560 (1990).
13. D. L. Scott *et al.*, *ibid.*, p. 1541.
14. L. Marz, C. Mollay, G. Kreil, G. Zelger, *Insect Biochem.* **11**, 685 (1981).
15. J. E. Ladner and G. Bricogne, version 9A2 of PHARE kindly provided by the authors (1974); Z. Otwinowski, manuscript in preparation.
16. B.-C. Wang, *Methods Enzymol.* **115**, 90 (1985).
17. W. Yuan and M. H. Gelb, *J. Am. Chem. Soc.* **110**, 2665 (1988); W. Yuan, D. M. Quinn, P. B. Sigler, M. H. Gelb, *Biochemistry* **29**, 6082 (1990).
18. The coefficients of the Fourier sum are $(2F_o - F_c)\exp(i\phi_c)$ where F_o is the observed amplitude and F_c and ϕ_c are the amplitude and phase, respectively, of the structure factors calculated from the refined model.
19. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
20. The segmental correspondence of sequence and structure that relates the bee-venom PLA_2 to the canonical homologous family suggests the evolutionary process of "exon shuffling" [W. Gilbert, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 901 (1987)]. If exon shuffling is operative it is a more subtle and complicated process than normally envisioned. For example, the second exon [J. J. Seilhamer *et al.*, *J. Biol. Chem.* **264**, 5335 (1986)] includes the amino-terminal helix and calcium-binding loop. The bee-venom enzyme appears to have split the first exon. It shows no amino-terminal helix and a perfectly conserved calcium-binding loop. The α-helical segment (residues 76 to 93) that replaces the function of the amino-terminal helix has the opposite polarity and an altogether different sequence.

21. We thank L. Marz for his help early in this project in supplying glycosylated and unglycosylated enzyme and G. Kreil for access to the bee-venom PLA_2 sequence prior to publication. The research at Yale was supported by NIH grant no. GM22324 and by the Howard Hughes Medical Institute; the research at the University of Washington was supported by NIH grant no. HL36235. D.L.S. is a postgraduate fellow at Yale and a graduate student pro forma at the University of Chicago.

4 June 1990; accepted 27 September 1990

Bacterial Origin of a Chloroplast Intron: Conserved Self-Splicing Group I Introns in Cyanobacteria

MING-QUN XU, SCOTT D. KATHE, HEIDI GOODRICH-BLAIR, SANDRA A. NIERZWICKI-BAUER, DAVID A. SHUB*

A self-splicing group I intron has been found in the gene for a leucine transfer RNA in two species of *Anabaena*, a filamentous nitrogen-fixing cyanobacterium. The intron is similar to one that is found at the identical position in the same transfer RNA gene of chloroplasts of land plants. Because cyanobacteria were the progenitors of chloroplasts, it is likely that group I introns predated the endosymbiotic association of these eubacteria with eukaryotic cells.

THE DISCOVERY OF INTRONS IN EUkaryotic genes has stimulated considerable discussion of their origin and evolution. One hypothesis assumes a relatively recent, eukaryotic origin for introns. According to this view, introns are mobile elements that can be transferred between genes, between nuclear and organelle genomes, and horizontally between organisms (*1*). A second hypothesis proposes that introns arose very early, perhaps in the "progenote" ancestor of all living organisms (*2*). According to this scheme, rapidly dividing microorganisms have lost their introns as a result of selection for rapid DNA replication. On the other hand, eukaryotes, under reduced pressure to streamline their genomes, have retained their introns.

Only circumstantial support has been found for either of these views. The demonstration that the group I intron of *Tetrahymena* is able to act as a true enzyme (*3*) added support for the antiquity of introns: self-splicing introns could have been present in the earliest genes of a precellular "RNA world" (*4*). Furthermore, it has been suggested that group I, group II, and nuclear mRNA introns—all of which splice through a series of transesterification reactions—may be evolutionarily related (*5*). However, no trace could be found in contemporary bacteria of these once ubiquitous, primordial introns. Rather, the only prokaryotic introns were in tRNA and ribosomal RNA (rRNA) genes of archaebacteria. The splicing mechanism of these introns is entirely different, being catalyzed by protein enzymes that seem to recognize structural features of the precursor RNA, and is similar to that of introns of eukaryotic nuclear tRNA genes (*6*).

The discovery of group I self-splicing introns in bacteriophages of both Gram-negative and Gram-positive eubacteria (*7, 8*) did not resolve this controversy, because the origin and evolution of viruses are themselves unclear. A convincing argument has been made that the large, tailed, DNA phages are ancient, originating in the earliest eubacteria. However, viruses are genetic mosaics and any particular virus might be a relatively recent assemblage of genes from different sources (*9*). Indeed, some introns (including two in phage T4) encode proteins that mobilize the transfer of the intron to "homing" sites in DNA (*10*), providing further support for the spread of introns through horizontal gene transfer.

In considering a likely source of introns in contemporary eubacteria, we have been impressed by the abundance of introns in

M.-Q. Xu, H. Goodrich-Blair, D. A. Shub, Department of Biological Sciences and Center for Molecular Genetics, State University of New York at Albany, Albany, NY 12222.
S. D. Kathe and S. A. Nierzwicki-Bauer, Plant Research Group, Biology Department, Rensselaer Polytechnic Institute, Troy, NY 12181.

*To whom correspondence should be addressed.

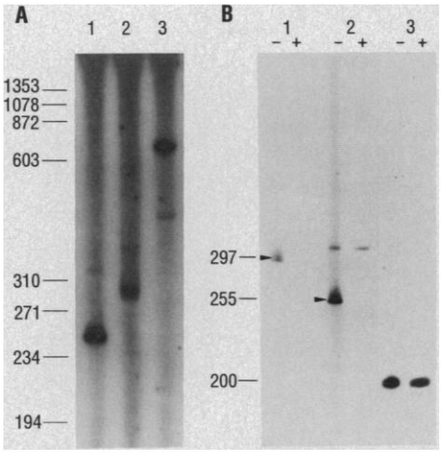

Fig. 1. End-labeling of cyanobacterial RNA with [α-^{32}P]GTP. (**A**) RNAs were incubated with [α-^{32}P]GTP (*19*) and separated on a 5% polyacrylamide–8 M urea gel. The autoradiogram is shown. RNAs were from *Anabaena* PCC7120 (10 μg) (lane 1), *A. azollae* (10 μg) (lane 2), and *Synechocystis* PCC6803 (17.5 μg) (lane 3) (*34*). (**B**) [α-^{32}P]GTP-labeled RNAs (5 μg), from (lanes 1) *A. azollae* and (lanes 2) *Anabaena* PCC7120, or (lanes 3) a 200-nt transcript (from cloned phage T4 DNA) uniformly labeled in vitro by transcription with phage T7 RNA polymerase were incubated with (+) or without (−) alkaline phosphatase (2 units; Boehringer Mannheim) in 50 mM tris-HCl (pH 8.0), and 0.1 mM EDTA at 50°C for 30 min. Sample 3 was included to rule out contamination of alkaline phosphatase with endonuclease activity. The radioactivity in a minor RNA species from *Anabaena* PCC7120 was resistant to alkaline phosphatase; the calculated size of this RNA is consistent with it being the splicing intermediate formed by cleavage at the 5′ splice site. The highly structured 3′ exon positioned near the 5′ end of the intron probably interferes sterically with the enzyme. On longer exposure, an equivalent resistant band is also seen with *A. azollae*. Sizes in (A) and (B) are shown in nucleotides.

chloroplast genomes (*11*). In particular, a group I intron is found in the gene for a leucine tRNA with a UAA anticodon [tRNA^{Leu} (UAA)] in the chloroplasts of many land plants—including bean, liverwort, maize, rice, and tobacco (*12, 13*)—as well as in the cyanelle of the red alga *Cyanophora paradoxa* (*14*). Phylogenetic analysis of the sequences of the rRNAs of these obligate endosymbionts has shown that their origins can be traced to the eubacteria, within the cyanobacterial group (*15*). The cyanobacteria and plastids comprise a monophyletic group (*16*). Therefore, evidence for a chloroplast-like group I intron in the genomes of cyanobacteria would demonstrate that this intron arose in eubacteria and already existed at the time of establishment of endosymbiosis.

Self-splicing group I introns can be identified by incubating unspliced RNA with [^{32}P]GTP (guanosine triphosphate) in vitro (*17*). The first step in splicing is attack by the 3′ hydroxyl group of guanosine at the 5′ splice site, with the guanosine derivative remaining covalently attached to the 5′ end of the linear excised intron (*18*). We have used this end-labeling reaction to detect self-splicing introns in bacteriophage genes (*8, 19*). Although the group I introns of plant chloroplasts and cyanelles do not self-splice (*13, 14, 20*), we showed that RNA prepared from cyanobacteria can be end-labeled with GTP in vitro (Fig. 1). After incubation with [α-^{32}P]GTP and separation on a denaturing polyacrylamide gel, RNA samples from the cyanobacteria *Anabaena azollae (caroliniana)*, *Anabaena* PCC7120, and *Synechocystis* PCC6803 contained several labeled species (Fig. 1A), a finding that is diagnostic of group I introns. The major RNA species from both *A. azollae* [~297 nucleotides (nt)] and *Anabaena* PCC7120 (~255 nt) were shown to be end-labeled because radioactivity could be removed by alkaline phosphatase treatment (Fig. 1B) (the *Synechocystis* sample was not treated).

The sizes of these putative cyanobacterial introns were similar to that of the group I intron of the tRNA^{Leu}(UAA) gene in chloroplasts of the liverwort *Marchantia polymorpha* (315 nt) (*21*). Indeed, when a lambda library of *A. azollae* DNA was screened with a radioactive transcript of the *M. polymorpha* chloroplast intron, a positive clone was found. The DNA inserted into this phage contained a 2.7-kb Eco RI–Hind III restriction fragment (also present in digests of genomic DNA), which was detected by Southern hybridization with the same probe (*22*). The 2.7-kb fragment was subcloned into pBSM13− (Stratagene) that had been digested with Eco RI and Hind III, to create the 5.9-kb plasmid pAAC1, which contains promotors for T3 and T7 phage RNA polymerases flanking the cloned insert (Fig. 2A).

After pAAC1 DNA was cut with Hind III and transcribed with T7 polymerase, the size of the major product was found to be that predicted for the run-off transcript. However, when the same DNA was cut with Eco RI and transcribed with T3 polymerase, additional major products appeared, including a species of ~300 nt (*23*)—the same size as the end-labeled RNA in Fig. 1. To determine the boundaries of the intron, pAAC1 DNA was digested with various restriction enzymes and T3 transcription products of these templates were analyzed by gel electrophoresis (Fig. 2). Multiple products of T3 transcription, consistent with the removal of an intron of approximately 300 nt, were seen only when DNA was truncated with Hind II (Fig. 2B) or at more distal sites (*23*). Because the distance between the Dra I and Hind II restriction sites was estimated to be 100 to 200 base pairs (bp), the intron should be located within the DNA immediately preceding the Hind II site. The similarity in size between the cloned intron and the GTP–end-labeled species in Fig. 1 implies that they may be the same, although the band from *A. azollae* is clearly a doublet and may contain more than one intron.

The sequence of the DNA from the region containing the putative intron (Fig. 3) is very similar, both in the intron and exon sequences, to the tRNA^{Leu}(UAA) genes of *M. polymorpha* chloroplasts and *C. paradoxa* cy-

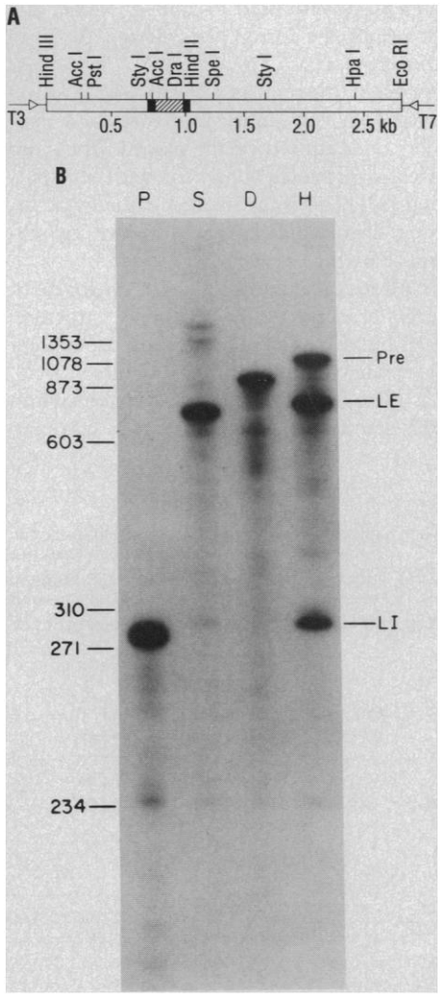

Fig. 2. (**A**) Restriction map of the 2.7-kb insert of pAAC1. Arrows indicate direction of transcription from T3 and T7 phage promoters in the vector. Putative tRNA (solid boxes) and intron (hatched box) sequences are shown. (**B**) In vitro transcription of the 2.7-kb insert of pAAC1. DNA was truncated with Pst I (P), Sty I (S), Dra I (D), or Hind II (H), and 1 μg of the resulting product was transcribed with phage T3 RNA polymerase in the presence of [α-^{32}P]UTP (uridine triphosphate) (*8*). The presence of GTP in the reaction mixtures allows group I intron self-splicing to occur. The reaction products were separated on a 5% polyacrylamide–8 M urea gel, and the autoradiogram is shown. Labels indicate positions expected for the runoff transcript (Pre), ligated exons (LE), and linear intron (LI). Sizes are in nucleotides.

anelles (*14, 21*). The tRNA portion of the *A. azollae* sequence can be folded into a typical cloverleaf structure (Fig. 4), which is identical at 73 of 85 positions (86%) to that of *M. polymorpha* (*20*). All differences are conservative in the secondary structure, either occurring in loops or preserving (or enhancing) base-pairing in the stem regions. The 291-nt intron interrupts the tRNA gene at exactly the same location, between the first and second anticodon nucleotides.

Polymerase chain reaction (PCR) primers based on the sequences of the tRNAs shown in Fig. 3 were used to amplify, clone, and sequence the intron we inferred to be in *Anabaena* PCC7120 (*24, 25*) (Fig. 1). As expected, the sequence of this 249-nt intron (Fig. 5) is very similar to that of *A. azollae*, and transcription of the cloned DNA also yielded products consistent with self-splicing (*23*). This is the smallest naturally occurring intron that has been shown to self-splice in vitro (*13*).

All group I introns share a phylogenetically conserved core secondary structure, and highly conserved sequences are present at certain positions in the structure (*13, 26, 27*). The *Anabaena* introns can be folded into this standard secondary structure representation (Fig. 5). In addition to the short sequences that are conserved among all group I introns, the *Anabaena* introns also bear a strong resemblance in primary sequence to the chloroplast and cyanelle introns throughout the core secondary structure (Fig. 3). At 156 homologous positions in the intron core, *A. azollae* is identical to *C. paradoxa* at 127 (81%) and to *M. polymorpha* at 114 positions (73%). As expected, the *A. azollae* and *Anabaena* PCC7120 introns show even greater similarity, differing at only 17 positions (89% identity) within these same regions (Figs. 3 and 5).

The extremely high conservation of bacterial and plastid introns can be appreciated in the context of the closely related *td* and *nrdB* introns of bacteriophage T4 (*28*). These introns are in closely linked genes and their sequences are sufficiently similar that deletions resulting from homologous recombination between them are readily observed in the laboratory, yielding functional hybrid introns (*29*). Yet only 52% of the 260 residues in structurally homologous regions of these introns (*28*) are identical.

The $tRNA^{Leu}$ introns also share peculiarities in both their sequences and their structures. For example, the two bulged A residues in P8 are shared by all four introns. In addition, a base pair in P7 (the first one on the left in Fig. 5) is A-U in most group I introns. However, the *Anabaena*, cyanelle, and land plant chloroplast introns all have C-G in this position (*12, 13*). For such similarities to have been maintained, it is likely that the introns interact with other macromolecular components. These might include the splicing factors required by the plastid introns (*14, 20*).

```
                              P1[5']    P1[3']   P2[5']        P2[3']
            tRNA[5']
C.p. ggggatgtggtggaattggtagacgcaacggacttAAAA-ATTTGAGCTCTTAATTAAGAAAT-TATTAAGTT
     **** ***** ***** ********* ********* ** *  ***** *** **  ****** * ******
A.a. gggggtgtggcggaat-ggtagacgctacggacttAGAA-AACTGAGC-CTTGATAGAGAAATCTTTTAAGTG
     ****** ***** *** ********** ********    ** *****  **  * ******* *  *** **
M.p. gggggtatggcgaaattggtagacgctgcggacttAATTTAATTGAGC-TTTAGTTGAGAAATTTACTAAATG

     P3[5']     P4[5']    P5[5']       P5[3']   P4[3']P6[5']P6a[5']    P6a[3']
C.p. GCAAACTCTCA|AATTCAGGGAAA|CC — 38— GGC|AATCCTGAGC|CAAGACAAATCATTTTGTAT
     * ** ******|************|**       ***|**********|******            * *
A.a. G-AAGCTCTCA|AATTCAGGGAAA|CC — 19— GGC|AATCCTGAGC|CAAGCCC - 69 - GGGAA
         * * ***|************|**       ** |**********|***                 *
M.p. -ATTGTTTTCA|AATTCAGGGAAA|CC — 27— GGT|AATCCTGAGC|CAAATTT - 14 - AAAGA
                     P                          Q

     P6[3'] P7[5']    P3[3']   P8[5']            P8[3']        P7[3']  P9[5']
C.p. G|GTGCAGAGACTCGA|TGAGAGTTATCCTAACATTT——ATAATGAGGATA|AAGGTAGAGTCC|AATTCTT
     *|**************| * *** ** ****** *         **** **|**** *******|******
A.a. G|GTGCAGAGACTCGA|CGGGAGCTACCCTAACGTAA——AGCCGAGGGTA|AAGGGAGAGTCC|AATTCTC
     *|************ *| *  * *** ******* **   ** ***** **|***  *******|  ** *
M.p. G|GTGCAGAGACTCAA|AGAAAACTATCCTAACGAAA-84-AGACGAGGATA|AAGATAGAGTCC|GTTTTTA
              R                                               S

              P9[3']
                            tRNA[3']
C.p. — 27 — AAGAATGaaaatccgtcgattgtataga-tcgtgagggttcaagtccctccgtcccca
             *************** ** **** * * ***** ************* **  *****
A.a. — 46 — GAGAATGaaaatccgttgactgtaaaaagtcgtgggggttcaagtccccccacccca
             * *************** ** **** *  **** ************* * *******
M.p. — 29 — TAAAATGaaaatccgttggctttaaaga-ccgtgagggttcaagtccctctacccca
```

Fig. 3. Comparison of the sequence of the *A. azollae* $tRNA^{Leu}$ gene with cyanelle and plant chloroplast genes. Transfer RNA sequences are in lowercase and intron sequences in uppercase letters. Brackets indicate phylogenetically conserved structural elements of group-I introns (P1 to P9). Conserved primary sequence elements (P, Q, R, S) are boxed. Asterisks indicate pairwise identity. Dashes indicate adjacent nucleotides. Numbers between dashes represent nucleotides omitted from the sequence. C.p., *C. paradoxa* cyanelle; A.a, *A. azollae*; and M.p., *M. polymorpha* chloroplast. Sequence was determined by the dideoxy chain termination method with T7 DNA polymerase (Sequenase; U.S. Biochemical). Universal primers complementary to vector sequences were used to sequence both strands of the DNA.

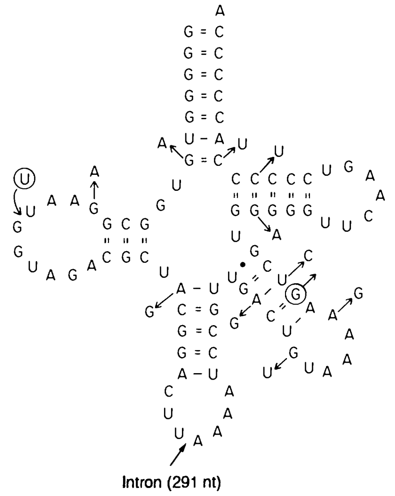

Fig. 4. Secondary structure of *A. azollae* $tRNA^{Leu}$ (UAA). The sequence presented in Fig. 3 is folded into the standard tRNA secondary structure. Arrows represent base changes, and circles show insertions or deletions compared with *M. polymorpha*. The position of the intron is indicated.

It has been suggested that the group I intron ends are brought together by base-pairing between a sequence near the 5′ end of the intron (the internal guide sequence) and the exon sequences flanking both splice boundaries, forming helices P1 and P10 (*26*). A recent study concluded that for cleavage to occur in vitro, the P1 helix of the *Tetrahymena* group I intron requires a minimum of 4 bp preceding the 5′ splice site (*30*). Thus, it is surprising to note that the *Anabaena* introns, with only 3 bp in P1 and completely lacking a recognizable P10, are able to self-splice in vitro. Perhaps the conformation required for splicing is stablized by an additional helix that brackets the splice boundary (the tRNA anticodon stem), and by an exceptionally stable association (P9) that brings the 3′ end of the intron close to the catalytic site (*31*). These features are all shared with the *Cyanophora* and *Marchantia* introns and cannot, therefore, be responsible for the absence of self-splicing of the $tRNA^{Leu}$(UAA) introns of these plastids.

Group I introns have now been detected in a variety of chloroplasts and cyanobacteria. Because intron loss and insertion are common events in evolution, retention of a

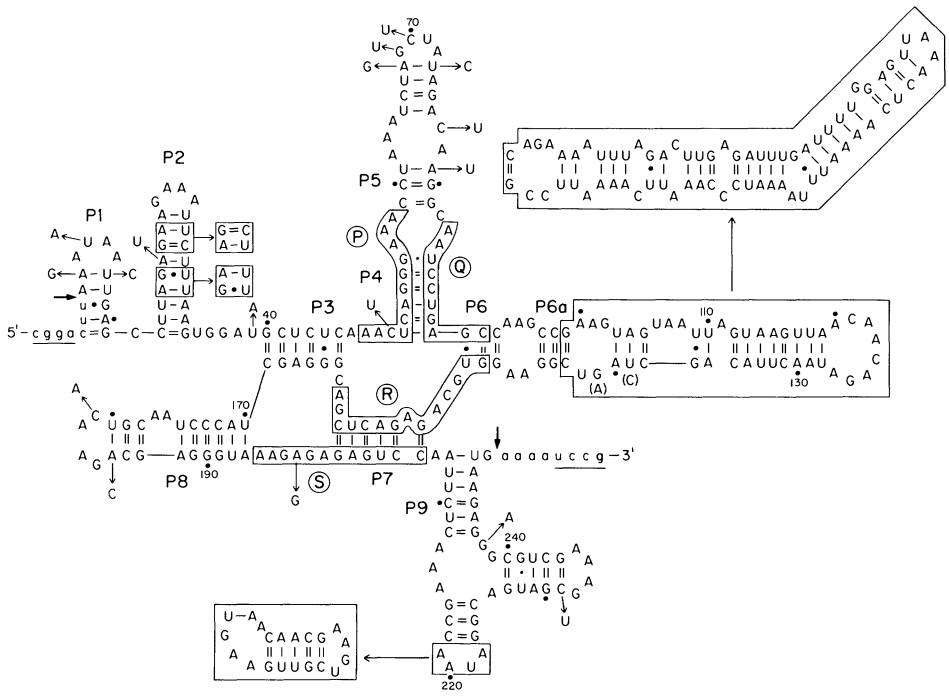

Fig. 5. Secondary structure of the *Anabaena* PCC7120 intron. Exon sequences are in lowercase and intron sequences in uppercase letters. Putative splice sites are indicated by bold arrows. Conserved secondary structure (P1 to P9) elements are shown and conserved primary sequence elements (P, Q, R, S) are boxed. Sequence differences between two PCR clones, at positions 139 and 141, are indicated by parentheses. Nucleotide changes in the *A. azollae* intron are indicated by arrows. Nucleotides in the tRNA anticodon stem are underlined.

particular intron in the same gene suggests the possibility that splicing is regulated in these organisms, providing a selection for intron retention. The chloroplast $tRNA^{Leu}$ introns cannot splice in vitro, even in the presence of extracts made from chloroplasts (*32*). It will be of interest to determine if cyanobacteria retain the splicing factors for these introns.

Our data support a prokaryotic origin—rather than invasion from eukaryotic nuclear genes—for the presence of group I introns in chloroplasts of land plants. A highly conserved intron is present in (i) the free-living laboratory strain *Anabaena* PCC7120, (ii) *A. azollae*, a bacterium that is associated with the water fern *Azolla*, and (iii) the highly degenerate and obligately endosymbiotic chloroplasts.

We have also shown that group I introns are present in cyanobacteria that are distantly related. *Anabaena* grows as filaments and is capable of forming specialized cells for fixation of atmospheric nitrogen (heterocysts) and specialized resting cells (akinetes). *Synechocystis*, on the other hand, is unicellular and is not capable of fixing nitrogen. Kuhsel *et al.* (*25*) have described a group I intron in three additional genera. Thus, it seems that self-splicing group I introns are prevalent in this deeply branching eubacterial phylum (*15, 16*) and may be at least as old as the divergence of the cyanobacteria, which occurred 2.0 to 3.5 billion years ago (*33*). Further studies will reveal whether introns occur only in this eubacterial phylum, or are more generally distributed among the bacteria.

REFERENCES AND NOTES

1. T. Cavalier-Smith, in *Endocytobiology III*, J. J. Lee and J. F. Fredrick, Eds. (New York Academy of Sciences, New York, 1987), pp. 17–71.
2. J. E. Darnell and W. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1271 (1986).
3. A. J. Zaug and T. R. Cech, *Science* **231**, 470 (1986).
4. W. Gilbert, *Nature* **319**, 618 (1986).
5. T. R. Cech, *Cell* **44**, 207 (1986).
6. L. D. Thompson *et al.*, *Can. J. Microbiol.* **35**, 36 (1989).
7. F. K. Chu, G. F. Maley, D. K. West, M. Belfort, F. Maley, *Cell* **45**, 157 (1986); K. Ehrenman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5875 (1986).
8. H. Goodrich-Blair, V. Scarlato, J. M. Gott, M.-Q. Xu, D. A. Shub, *Cell* **63**, 417 (1990).
9. D. C. Reanney and H.-W. Ackermann, *Adv. Virus Res.* **27**, 205 (1982); A. Campbell, in *The Bacteriophages*, R. Calender, Ed. (Plenum, New York, 1988), vol. 1, pp. 1–14.
10. I. G. Macreadie, R. M. Scott, A. R. Zinn, R. A. Butow, *Cell* **41**, 395 (1985); L. Colleaux, L. d'Auriol, F. Galibert, B. Dujon, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6022 (1988); S. M. Quirk, D. Bell-Pedersen, M. Belfort, *Cell* **56**, 455 (1989).
11. J. D. Palmer, in *The Molecular Biology of Plastids*, L. Bogorad and I. K. Vasil, Eds. (Academic Press, San Diego, in press).
12. J. Hiratsuka *et al.*, *Mol. Gen. Genet.* **217**, 185 (1989).
13. T. R. Cech, *Gene* **73**, 259 (1988).
14. J.-L. Evrard, M. Kuntz, N. A. Straus, J.-H. Weil, *ibid.* **71**, 115 (1988). The term "cyanelle" refers to a plastid-like endosymbiont that has lost most of its genome but retains a cell wall and other ultrastructural features common to cyanobacteria.
15. C. R. Woese, *Microbiol. Rev.* **51**, 221 (1987).
16. S. J. Giovannoni *et al.*, *J. Bacteriol.* **170**, 3584 (1988).
17. G. Garriga and A. M. Lambowitz, *Cell* **39**, 631 (1984).
18. K. Kruger *et al.*, *ibid.* **31**, 147 (1982).
19. J. M. Gott *et al.*, *ibid.* **47**, 81 (1986).
20. H. Ozeki *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 791 (1987). We have confirmed that in vitro transcripts containing the *Marchantia* intron do not self-splice.
21. K. Umesono *et al.*, *J. Mol. Biol.* **203**, 299 (1988).
22. Transcripts were prepared as described (Fig. 2, legend) from plasmid pT7-23, which contains an *M. polymorpha* chloroplast *trnL*(UAA) gene next to a phage T7 promoter. The EMBL3 library, which contains inserts of 9 to 20 kb from a Sau 3A digest of *A. Azollae* DNA (1.6×10^7 plaque-forming units/ml), was constructed and amplified by J. S. Gebhardt in 1987 [T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982)]. The host was *Escherichia coli* strain P2392. For hybridization to λ plaques or Southern blots, DNA was transferred to a nitrocellulose filter, which was then incubated in 6× SSPE (saline, sodium phosphate, EDTA), 0.5% SDS, 5× Denhardt's solution, and yeast RNA (200 μg/ml) for 5 hours at 45°C (Southern blots) or 50°C (λ plaques). The filter was probed under the same conditions for 12 to 16 hours, and then washed in 6× SSPE and 0.5% SDS twice at 45°C and twice at 50°C, for 30 min each time.
23. M.-Q. Xu *et al.*, unpublished data.
24. R. K. Saiki *et al.*, *Science* **239**, 487 (1988). The PCR primers, 5'GAATTCACA(G/A)TC(G/A)ACGGATTTT3' and 5'AAGCTTGACGC(T/A)(A/G)CGGACTT3', are complementary to residues 35 to 50 and identical to residues 21 to 34 of the tRNA sequence, respectively. Each has a restriction site added to its 5' end. PCR was performed with 100 ng of DNA and 0.3 μM of each oligonucleotide in 10 mM tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, 200 μM of each deoxynucleotide triphosphate, and 2.5 units of *Thermus aquaticus* DNA polymerase (U.S. Biochemical). Reaction mixtures were heated to 92°C for 2 min, 72°C for 2.5 min, and 37°C for 3 min, for 30 cycles, with a final extension step of 72°C for 10 min, in a thermal cycler (Coy Labs). After separation by electrophoresis on a 5% polyacrylamide nondenaturing gel, a single distinct species of ~295 nt was apparent. The PCR product was ligated to Hinc II–digested pBSM13− and introduced into *E. coli* by electroporation. Two plasmids, containing the insert in opposite orientations, were isolated and sequenced as described (Fig. 3, legend) with the same primers.
25. M. G. Kuhsel *et al.*, *Science* **250**, 1570 (1990).
26. R. W. Davies, R. B. Waring, J. A. Ray, T. A. Brown, C. Scazzocchio, *Nature* **300**, 719 (1982).
27. F. Michel *et al.*, *Biochimie* **64**, 867 (1982).
28. D. A. Shub *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1151 (1988).
29. D. H. Hall *et al.*, *Nature* **340**, 574 (1989).
30. J. A. Doudna, B. P. Cormack, J. W. Szostak, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7402 (1989).
31. F. Michel, M. Hanna, R. Green, D. P. Bartel, J. W. Szostak, *Nature* **342**, 391 (1989); J. M. Burke, J. S. Esherick, W. R. Burfeind, J. L. King, *ibid.* **344**, 80 (1990); F. Michel, P. Netter, M.-Q. Xu, D. A. Shub, *Genes Dev.* **4**, 777 (1990).
32. R. B. Hallick, personal communication.
33. A. H. Knoll and J. Bauld, *Trans. R. Soc. Edinburgh: Earth Sci.* **80**, 209 (1989).
34. The free-living cultures of *Anabaena* PCC7120 and *Synechocystis* PCC6803 were grown on BG-11 medium, without (N−) and with (N+) sodium nitrite, respectively [R. Y. Stanier, R. Kunisawa, M. Mandel, G. Cohen-Bazire, *Bacteriol. Rev.* **35**, 171 (1971)]. Cultures were grown in cotton-stoppered flasks on a gyratory shaker platform in a phytotron with a photoperiod of 16 hours light and 8 hours dark at ~24°C. The small aquatic fern *Azolla caroliniana* Willdenow (International Rice Research Institute Collection 3001), which is in symbiotic association with the filamentous nitrogen-fixing bacterium *A. azollae*, was grown as described [E. B. Braun-Howland and S. A. Nierzwicki-Bauer, *Planta* **180**, 361 (1990)]. The symbiotically associated

cyanobacteria were separated as described [S. A. Nierzwicki-Bauer and R. Haselkorn, *EMBO J.* 5, 29 (1986)]. The isolated cyanobacteria were examined microscopically to ensure the purity of the preparation. Total RNA and DNA was isolated from free-living and symbiotic cyanobacteria as described [J. W. Golden, S. J. Robinson, R. Haselkorn, *Nature* 314, 419 (1985)]. RNA was separated from DNA by CsCl density gradient centrifugation, with a 5.7 M CsCl cushion. The DNA was centrifuged as described by Maniatis *et al.* (*22*). RNA samples extracted from isolated *A. azollae* lacked eukaryotic rRNA bands, providing additional confirmation that these preparations were free of contaminating eukaryotic organelles.

35. We thank M. G. Kuhsel and J. D. Palmer for communicating their manuscript (*25*) prior to submission, and for comments on our manuscript; and H. Ozeki for providing pT7-23. Supported by NIH grant GM37746 (D.A.S.), NSF grant DMB-9011698 (S.A.N.-B.), and, in part, by a Burke Graduate Fellowship (H.G.B.).

17 July 1990; accepted 17 October 1990

An Ancient Group I Intron Shared by Eubacteria and Chloroplasts

MARIA G. KUHSEL, RODERICK STRICKLAND, JEFFREY D. PALMER*

Introns have been found in the genomes of all major groups of organisms except eubacteria. The presence of introns in chloroplasts and mitochondria, both of which are of eubacterial origin, has been interpreted as evidence either for the recent acquisition of introns by organelles or for the loss of introns from their eubacterial progenitors. The gene for the leucine transfer RNA with a UAA anticodon [tRNA^{Leu}(UAA)] from five diverse cyanobacteria and several major groups of chloroplasts contains a single group I intron. The intron is conserved in secondary structure and primary sequence, and occupies the same position, within the UAA anticodon. The homology of the intron across chloroplasts and cyanobacteria implies that it was present in their common ancestor and that it has been maintained in their genomes for at least 1 billion years.

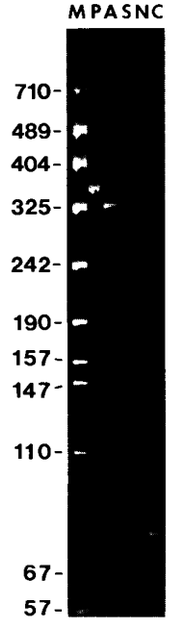

Fig. 1. Size determination of tRNA^{Leu} genes from cyanobacteria and chloroplasts. PCR products from *Phormidium* N182 (P), *Anacystis* R2 (A), *Scytonema* PCC7110 (S), *Nicotiana tabacum* (N), and *Chara hispida* (C), and Hpa II fragments from Bluescript SK+ (M) (sizes marked in base pairs), were subjected to electrophoresis on a 5% polyacrylamide gel and stained with ethidium bromide. Sequence of the 5′ PCR primer (exon positions 8 to 33): TGG(CT)GAAAT (CT)GGTAGACGC(AT) (AG)CGGAC. Sequence of the 3′ PCR primer (positions 51 to 85): TGGG-GATAGAGGGACTTGA-ACCCTCACGATTTTTA. PCR was performed with 100 ng of template DNA, 100 pmol of each primer, 20 nmol of each deoxynucleotide triphosphate, 2 units of Taq polymerase (Promega), and 1× Taq buffer (Promega). The reaction mixture was heated at 94°C for 2.0 min and then subjected to 40 cycles consisting of 1.5 min at 92°C, 1.5 min at 50°C, and 2.0 min at 72°C, with the last step being increased by 5 s at each cycle and the final 72°C step lasting for 12 min.

INTRONS ARE WIDESPREAD, OCCURring in nuclear, chloroplast, and mitochondrial genomes of eukaryotes (*1*), in archaebacterial genomes (*2*), and in viruses of eukaryotes and eubacteria (*3*). The "introns early" view of intron evolution maintains that most introns were present in the common ancestor of eukaryotes, eubacteria, and archaebacteria, and that the general direction of their evolution has been toward loss (*4*). These arguments are based on the self-splicing properties displayed by certain introns—which are interpreted to be remnants of an ancient "RNA world"—on the broad distribution of certain nuclear introns among diverse lineages of eukaryotes, and on the conserved structure among eukaryotes and bacteria of protein genes thought to have been assembled by the intron-mediated process termed "exon shuffling" (*5*). The "introns late" view holds that primordial genes lacked introns, which were instead acquired independently, and relatively recently, in various lineages of life (*6*). This view is based on the lateral mobility of certain types of introns (*7*) and on the sporadic distribution of introns across all of life (*1–3*).

Critical to the debate over intron evolution is the complete absence of introns from eubacterial genes. Although introns have been found in several eubacteriophage genes, these are mobile group I introns that are thought to be derived from lateral gene transfer (*3, 7*). Introns-late advocates interpret the absence of introns from eubacterial genes as evidence that their common ancestor lacked introns (*6*), whereas introns-early proponents argue that selection for rapid bacterial growth has led to a streamlining of the genome and the elimination of introns (*4*). Chloroplasts and mitochondria are derived from eubacteria by endosymbiosis (*8*), yet in many lineages they contain numerous introns. Introns-early adherents view this as evidence that the selection to eliminate introns and streamline the genome is lessened in organelles compared to bacteria (*4*), whereas introns-late adherents contend that organelles acquired their introns after endosymbiosis (*6, 9, 10*).

A nearly identical set of 20 introns has been conserved throughout the evolution of land plant chloroplasts (*11*); however, these introns are largely absent from the few algae that have been examined (*10*), and phylogenetic survey shows that two tRNA introns were acquired specifically in the green algal ancestors of land plants (*12*). In contrast, a third tRNA intron, a group I intron in a gene for the leucine tRNA with a UAA anticodon [tRNA^{Leu}(UAA)], is found in chloroplasts of land plants and the distantly related alga *Cyanophora paradoxa* (*13*). This is therefore the best candidate for being an ancient intron potentially of eubacterial origin.

The distribution of this group I intron in tRNA^{Leu}(UAA) genes of chloroplasts and cyanobacteria—the progenitors of chloroplasts (*8*)—was assessed by determining the size and sequence of the products of polymerase chain reaction (PCR) amplification that was performed with primers specific for the exons of sequenced tRNA^{Leu}(UAA) genes of land plants and *Cyanophora* (*13, 14*). The group I intron splits these genes into exons of 35 and 50 base pairs (bp). Amplification with a 26-nucleotide (nt) 5′ primer (positions 8 to 33) and a 35-nt 3′ primer (positions 51 to 85) should yield a PCR product of ~80 bp for tRNA^{Leu} genes lacking introns, and for genes containing an intron, a product larger than 80 bp by the size of the intron. Control amplification with DNA from *Nicotiana tabacum* gave a product estimated to be 640 bp in size (Fig. 1), in reasonable agreement with the expected size of 581 bp (*14*). The DNA from three phylogenetically diverse (*15*) cyanobacteria—*Phormidium* N182, *Anacystis* R2, and *Scytonema* PCC7110—also yielded PCR products substantially larger (330 to 410 bp) (Fig. 1) than the 80 bp expected for an

Department of Biology, Indiana University, Bloomington, IN 47405.

*To whom correspondence should be addressed.