An Intron-encoded Protein Assists RNA Splicing of Multiple Similar Introns of Different Bacterial Genes

Qing Meng, Yanfei Wang, and Xiang-Qin Liu
From the Department of Biochemistry & Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada

Published, JBC Papers in Press, September 7, 2005, DOI 10.1074/jbc.C500328200
Received for publication, July 29, 2005, and in revised form, September 7, 2005

Four group II introns were found in an unusually intron-rich dnaN gene (encoding the β subunit of DNA polymerase III) of the cyanobacterium *Trichodesmium erythraeum*, and they have strong similarities to two introns of the *RIR* gene (encoding ribonucleotide reductase) of the same organism. Of these six introns, only the RIR-3 intron encodes a maturase protein and showed efficient RNA splicing when expressed in *Escherichia coli* cells. The other five introns do not encode a maturase protein and did not show RNA splicing in *E. coli*. But these maturase-less introns showed efficient RNA splicing when the RIR-3 intron-encoded maturase protein was co-expressed from a freestanding gene in the same cell. These findings demonstrated that an intron-encoded protein could function as a general maturase for multiple introns of different genes. Major implications may include an intron-mediated co-regulation of the different genes and a resemblance of the evolutionary origin of spliceosomal introns.

Group II introns are mobile genetic elements and presumed evolutionary progenitors of nuclear spliceosomal introns (1, 2). RNA sequences of group II introns can be folded into a conserved six-domain structure with properties of a catalytic RNA (3, 4). Group II introns are found more abundantly in mitochondria and chloroplasts (5, 6), where they reside in conserved genes. In contrast, most bacterial group II introns were found in mobile DNA or outside of genes (5, 7), which may suggest barriers against intron insertion in conserved bacterial genes. To do RNA splicing in vivo, group II introns require proteins or maturase to help the intron RNA fold into a structure that is catalytically active (1, 3), although some group II introns can self-splice in vitro under non-physiological conditions. Most bacterial group II introns and some organelle group II introns encode a protein typically in the domain IV loop region of the folded intron structure. The intron-encoded protein functions as a maturase to assist the RNA splicing, in addition to assisting intron retrohoming or retrotransposition (8, 9).

Previously characterized intron-encoded maturase proteins assist specifically the intron that encodes the maturase, although it is possible that closely related maturases may cross-react. In mitochondria and chloroplasts, many group II introns do not encode a maturase protein, and at least some of the maturase-less introns have recruited nuclear proteins to assist the RNA splicing (1). In chloroplasts, indirect evidences have suggested that maturase-less group II introns may be assisted by a maturase protein encoded in another intron (10, 11), but this has not been demonstrated. In bacteria, maturase-less group II introns are rare and found recently in an Archaea and a thermophilic cyanobacterium (12). Most bacterial introns are located outside genes or conserved genes, and it is not known whether or how the maturase-less introns do RNA splicing.

Here we report an unusual case of four maturase-less group II introns in a conserved bacterial protein gene, which is the largest number of introns known in a bacterial gene. More interestingly, these maturase-less introns could do RNA splicing in *Escherichia coli* when assisted by a maturase protein encoded in another intron of a different gene. This demonstrated for the first time that an intron-encoded protein could function as a general maturase for multiple introns of different genes, which has interesting implications on intron evolution and on possible roles of intron in coordinating gene expression.

EXPERIMENTAL PROCEDURES

Gene Cloning and Sequence Analysis—The *dnaN* gene, or parts of it, was amplified from genomic DNA of *Trichodesmium erythraeum* strain IMS101, by doing PCR using the high fidelity DNA polymerase Phusion (Finzymes). The resulting DNAs were cloned in a plasmid vector pDrive (Qagen) at the T/A cloning site and DNA sequences were determined through automated DNA sequencing, GenBank searches, sequence alignments, and intron RNA folding were performed using the BLAST search program (13). The Clustal W program (14), and the Mfold program (15), respectively.

RNA Splicing Analysis—Intron-containing DNAs were PCR-amplified and inserted in a ColE1-derived pDrive plasmid vector behind a *P_{lac*} promoter. This plasmid has an ampicillin resistance gene, and the PBAD promoter is inducible with isopropyl 1-thio-β-D-galactopyranoside (IPTG). The coding sequence of the maturase protein was PCR-amplified and inserted in a modified pAR plasmid vector behind a *P_{lac*} promoter. This plasmid was derived from a pACYC184 plasmid that is compatible with ColE1-derived plasmids. In fact, it has a chloramphenical resistance gene, and the *P_{lac*} promoter is inducible with isopropyl 1-thio-β-D-galactopyranoside (IPTG). For RNA splicing assays, *E. coli* cells containing the specified plasmid were grown in liquid Luria Broth medium at 37 °C to mid-log phase (A600=0.3). IPTG was then added to a final concentration of 0.8 mM to induce transcription of the intron-containing gene, and the induction was continued for 2 h. For cells containing also the maturase-encoding plasmid, l-arabinose was added to a final concentration of 0.2% to induce production of the maturase protein 30 min before adding IPTG. Total RNAs of the induced cells were extracted and treated with RNase-free DNase using the RNAsese kit (Qagen). Reverse transcription and subsequent PCR were carried out using specified primers and the reverse transcription (RT)-PCR kit (Qagen). The resulting DNA products were identified by their predicted sizes and confirmed through cloning and DNA sequencing.

RESULTS

Identification and Sequence Analysis of *dnaN* Gene and Its Introns—The recent findings of group II introns in an *RIR* gene (16) prompted us to find similar introns in related genes in the nearly complete genome sequence of the oceanic *N₂* fixing cyanobacterium *T. erythraeum* strain IMS101 (www.igi doe. gov), which had been determined at the United States Department of Energy Joint Genome Institute. This search revealed an intron-rich *dnaN* gene as illustrated in Fig. 1, whose 6,510-bp sequence consists of five exons and four group II introns. The five exons are 30, 55, 158, 303, and 615 bp long, respectively, and together they predict a 387-amino acid protein sequence that is 38% identical and 70% similar to previously characterized group II introns. The five exons are 30, 55, 158, 303, and 615 bp long, respectively, and together they predict a 387-amino acid protein sequence that is highly similar to DNA sequences of related organisms. As shown in Fig. 1, the predicted DNA protein sequence is 54–57% identical and 70% similar to intron-less DNA sequences of other cyanobacteria. It is also 23% identical and 42% similar to the functionally characterized *E. coli* DNA protein sequence (data not shown).

The four introns in the *dnaN* gene, referred to as *dnaN*-1, -2, -3, and -4 introns, are 1408, 1128, 1835, and 978 nt long, respectively. Boundaries of these introns were easily predicted through sequence comparisons with intron-less *dnaN* genes of closely related organisms (Fig. 1). As shown in Fig. 2, the four introns are similar to one another and also to two other introns previously identified in other cyanobacteria. These two introns were previously identified in other cyanobacteria. These two introns were previously identified in other cyanobacteria.

* This work was supported by a research grant from the National Science and Engineering Research Council of Canada. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 To whom correspondence should be addressed. Tel.: 902-494-1208; Fax: 902-494-1355; E-mail: pxqliu@dal.ca.

2 The abbreviations used are: IPTG, isopropyl 1-thio-β-D-galactopyranoside; nt, nucleotide; RT, reverse transcription.
FIGURE 1. The dnaN gene and its introns. A, schematic illustration of the dnaN gene and its neighboring dnaA gene. In the dnaN gene, black boxes represent the five exons, and gray boxes represent the four introns (dnaN-1 through dnaN-4). B, DnaN protein sequence comparison. Predicted DnaN protein sequence of T. erythraeum (Ter) is aligned with DnaN sequences of Nostoc sp. strain PCC7120 (Nsp) and Synechocystis sp. strain PCC6803 (Ssp). Each intron insertion site in the Ter sequence is marked with an arrowhead and the name of the intron. Symbols: - represents gaps introduced to optimize the alignment; * marks the positions of identical residues, and dots mark the positions of similar residues.

FIGURE 2. Intron sequence comparison. Compared sequences include four introns (dnaN-1 through dnaN-4) of the dnaN gene and two introns (RIR-2 and RIR-3) of the RIR gene. Only sequences proximal to points of interest are shown, omitted nucleotides are shown as numbers in parentheses, and the total length of each intron is shown at the end. Exon sequences proximal to the intron are shown in lowercase letters, in which the intron-binding sequences (IBS1 and IBS2) are underlined. In the intron, the folded regions correspond to the conserved catalytic core, the domain IV loop region has either a maturase coding sequence (in RIR-3 intron) or remnants of it (in all others), and the exon-binding sequences (EBS1 and EBS2) are underlined.
identified in the same organism. The previously identified introns are the T. eri and T. eri4 introns in the RIR gene encoding a ribonucleotide reductase (16), and they are referred to as RIR-2 and RIR-3 introns, respectively, to be consistent with the naming of the dnaN1, -2, -3, and -4 introns and to allow easier identification of the host gene of each intron. These six introns of the dnaN and RIR genes are 75–85% identical in a ~680-nucleotide catalytic core region, which can be folded into a conserved six-domain structure typical of group II introns. The folded structures are not shown, because they are very similar to published structures of previously identified introns of this organism (17).

Outside the folded catalytic core, the six introns differ greatly in size in the domain IV loop region (Fig. 2), where a maturase-coding open reading frame is usually located. Among the six introns, only the RIR-3 intron has a complete maturase-coding sequence in domain IV of its folded structure. None of the others has a maturase-coding open reading frame, although various amounts of apparent remnants of maturase-coding sequences could be recognized. Exon-binding sequences (EBS1 and EBS2) and intron-binding sequences (IBS1 and IBS2) could also be predicted for each of the six introns (Fig. 2), which are known to facilitate RNA splicing by forming base-pairs between the intron and the exon (18). As expected, EBS1 and EBS2 are in the loop sequence of Domain I, the corresponding IBS1 and IBS2 are in the 5′-exon immediately before the intron, with EBS1 and EBS2 base-paired with IBS1 and IBS2, respectively. The EBS1 and EBS2 of the six introns are located at the same corresponding position of the different introns, but their sequences are not similar (Fig. 2).

**RNA Splicing Activity of the Intron**—To determine whether the six similar introns of dnaN and RIR genes are capable of RNA splicing in E. coli, each intron coding sequence was inserted in a plasmid expression vector to construct an intron-containing recombinant gene (Fig. 3A). Various lengths of the native exon sequences were included with the introns, because this is usually required for efficient RNA splicing. After transcription of the recombinant gene and RNA splicing in E. coli, total RNAs were isolated and reverse-transcribed, and the resulting cDNAs were PCR-amplified, using specific oligonucleotide primers. This RT-PCR procedure produces DNA fragments corresponding to the precursor RNA and the spliced RNA, whose predicted sizes are listed in Fig. 3A. The predicted sizes of different precursors vary according to the different sizes of the introns. The predicted sizes of the spliced products also differ with different introns, because the sizes of the native exons that were included with the introns differ with different introns.

The RT-PCR products corresponding to individual introns were resolved by agarose gel electrophoresis and shown in Fig. 3B. DNA bands corresponding to the precursor and spliced RNAs were identified initially by matching their observed sizes with the predicted sizes listed in Fig. 3A. DNA bands corresponding to the spliced RNAs were inserted into cloning plasmids and subjected to sequence determination (data not shown), which confirmed their identities as well as the predicted splice junctions. For the complete RIR-3 intron that encodes a maturase protein, only a band corresponding to the spliced RNA was observed (lane 2), indicating complete RNA splicing. To see whether the intron-encoded maturase is required for RNA splicing, we deleted the maturase coding sequence except the first 15 codons. The resulting intron (RIR-3d) produced only the precursor band (lane 3), indicating that the maturase coding sequence was required for the splicing. We then expressed the maturase-coding sequence as a freestanding gene on a separate plasmid (lane 4). This indicated that the maturase protein can act in trans, and the maturase coding sequence does not need to be a part of the intron.

For the RIR-2 intron that does not encode a maturase protein, only a precursor band corresponding to the precursor RNA was observed (lane 5), indicating a lack of RNA splicing. However, efficient splicing of the RIR-2 intron was observed when the maturase protein from the RIR-3 intron was co-produced in the same cell from a freestanding gene on a separate plasmid (lane 6). This indicated that the maturase protein from the RIR-3 intron acted in trans on the maturase-less RIR-2 intron to assist its RNA splicing. Similarly, the four introns (dnaN1, dnaN2, dnaN3, and dnaN4) of the dnaN gene do not encode a maturase protein, and each produced only a precursor band corresponding to the precursor RNA (lanes 7, 9, 11, and 13), indicating a lack of RNA splicing. However, all showed efficient RNA splicing when the maturase coding sequence from the RIR-3 intron was co-expressed in the same cell as a freestanding gene on a separate plasmid (lanes 8, 10, 12, and 14). This indicated that the maturase protein from the RIR-3 intron acted in trans on each of these maturase-less introns to assist their RNA splicing. To confirm that the maturase protein, not its coding sequence, acted on the maturase-less introns, we disrupted the maturase protein synthesis by introducing a frameshift mutation (single nucleotide insertion) at nucleotide position 1178 of the 1770-bp maturase gene, and this abolished the RNA splicing activity of all the maturase-less introns (data not shown).

In group II introns, the domain IV loop region is outside the catalytic core and generally not conserved. Among the six introns studied here, the domain IV loop region varies greatly in size (Fig. 2), but the terminal parts of this region show a high level of sequence conservation (Fig. 4A). To test whether the domain IV loop region is required for RNA splicing, we deleted from three introns (RIR-3, dnaN3, and dnaN4) the entire domain IV loop sequence except the five terminal nucleotides AATTG. The resulting deletion introns (RIR-3D2, dnaN3D, and dnaN4D) were tested for RNA splicing in comparison with the original introns (Fig. 4B). Unlike the original introns, none of the deletion introns showed RNA splicing whether in the presence or in the absence of the maturase from the RIR-3 intron, indicating that the domain IV loop region of these introns is required for the RNA splicing.

**DISCUSSION**

We have demonstrated for the first time that an intron-encoded protein could function as a general maturase for the RNA splicing of multiple maturase-less introns of different genes. This suggests that the RIR-3 intron-en-
The general maturase assisting multiple introns of different genes may suggest a novel and intron-mediated mechanism of coordinating gene expressions, in which the production or function of the general maturase could control the functional expression (transcript maturation) of both the RIR gene and the dnaN gene. The two genes are functionally related in nucleotide synthesis and DNA replication, with the DNA polymerase III and the dnaN gene encoding a ribonucleotide reductase. The dnaN and RIR genes are expected to be functional in vivo despite their multiple introns, because each gene predicts a protein highly conserved with homolog proteins of other bacteria, and because no additional copy of the gene was predicted from the genome sequence. The RNA splicing activities of the dnaN and RIR introns were studied in E. coli; therefore it is not certain that our findings will apply in the native organism. We have not been able to study the RNA splicing in the native organism T. erythraeum, due to extreme difficulties of culturing this organism in the laboratory. It is also not known whether the general maturase acts on additional group II introns of other genes in this organism; therefore the above hypothesis of intron-mediated gene regulation remains to be proven.

The finding of a general maturase for multiple group II introns may also have implications on the evolution of nuclear spliceosomal introns. It is thought that group II introns have given rise to spliceosomal introns (2). A likely early step may be the emergence of a general maturase that assisted the RNA splicing of multiple introns, which would eventually lead to the evolution of a general splicing enzyme like the spliceosome. A general maturase might emerge more easily from a preexisting intron-specific maturase by recognizing conserved structures of the group II introns and not needing idiosyncratic structures of specific introns. The presence of multiple similar introns in the same organism, like the six similar introns in T. erythraeum, may facilitate this evolution process. Once the maturase from one intron can assist the RNA splicing of other introns, the maturase coding sequences of the other introns can be lost through mutation. The resulting maturase-less introns become dependent on the maturase-containing intron for their RNA splicing and therefore host gene function. This dependence shall prevent a loss of the maturase-containing intron, except when the maturase coding sequence becomes a freestanding gene. Freestanding maturase-like genes have been found in plants and thought to have derived from group II introns (20), although their biochemical function remains to be determined. A freestanding maturase gene has not been found in T. erythraeum.

REFERENCES

1. Lambowitz, A. M., and Zimmerly, S. (2004) Annu. Rev. Genet. 38, 1–35
2. Cavalier-Smith, T. (1991) Annu. Rev. Biochem. 60, 383–402
3. Edgell, D. R., Belfort, M., and Shub, D. A. (2000) Annu. Rev. Biochem. 42, 386–3874
4. Michel, F., and Ferat, J. L. (1995) Annu. Rev. Biochem. 64, 435–461
5. Tian, L., and Garcia, A. (1999) DNA polymerase III and the dnaN gene encoding the sliding DNA clamp of E. coli, due to extreme difficulties of culturing this organism in the laboratory. It is also not known whether the general maturase acts on additional group II introns of other genes in this organism; therefore the above hypothesis of intron-mediated gene regulation remains to be proven.

The finding of a general maturase for multiple group II introns may also have implications on the evolution of nuclear spliceosomal introns. It is thought that group II introns have given rise to spliceosomal introns (2). A likely early step may be the emergence of a general maturase that assisted the RNA splicing of multiple introns, which would eventually lead to the evolution of a general splicing enzyme like the spliceosome. A general maturase might emerge more easily from a preexisting intron-specific maturase by recognizing conserved structures of the group II introns and not needing idiosyncratic structures of specific introns. The presence of multiple similar introns in the same organism, like the six similar introns in T. erythraeum, may facilitate this evolution process. Once the maturase from one intron can assist the RNA splicing of other introns, the maturase coding sequences of the other introns can be lost through mutation. The resulting maturase-less introns become dependent on the maturase-containing intron for their RNA splicing and therefore host gene function. This dependence shall prevent a loss of the maturase-containing intron, except when the maturase coding sequence becomes a freestanding gene. Freestanding maturase-like genes have been found in plants and thought to have derived from group II introns (20), although their biochemical function remains to be determined. A freestanding maturase gene has not been found in T. erythraeum.

REFERENCES

1. Lambowitz, A. M., and Zimmerly, S. (2004) Annu. Rev. Genet. 38, 1–35
2. Cavalier-Smith, T. (1991) Annu. Rev. Biochem. 60, 383–402
3. Edgell, D. R., Belfort, M., and Shub, D. A. (2000) Annu. Rev. Biochem. 42, 386–3874
4. Michel, F., and Ferat, J. L. (1995) Annu. Rev. Biochem. 64, 435–461
5. Tian, L., and Garcia, A. (1999) DNA polymerase III and the dnaN gene encoding the sliding DNA clamp of E. coli, due to extreme difficulties of culturing this organism in the laboratory. It is also not known whether the general maturase acts on additional group II introns of other genes in this organism; therefore the above hypothesis of intron-mediated gene regulation remains to be proven.

The finding of a general maturase for multiple group II introns may also have implications on the evolution of nuclear spliceosomal introns. It is thought that group II introns have given rise to spliceosomal introns (2). A likely early step may be the emergence of a general maturase that assisted the RNA splicing of multiple introns, which would eventually lead to the evolution of a general splicing enzyme like the spliceosome. A general maturase might emerge more easily from a preexisting intron-specific maturase by recognizing conserved structures of the group II introns and not needing idiosyncratic structures of specific introns. The presence of multiple similar introns in the same organism, like the six similar introns in T. erythraeum, may facilitate this evolution process. Once the maturase from one intron can assist the RNA splicing of other introns, the maturase coding sequences of the other introns can be lost through mutation. The resulting maturase-less introns become dependent on the maturase-containing intron for their RNA splicing and therefore host gene function. This dependence shall prevent a loss of the maturase-containing intron, except when the maturase coding sequence becomes a freestanding gene. Freestanding maturase-like genes have been found in plants and thought to have derived from group II introns (20), although their biochemical function remains to be determined. A freestanding maturase gene has not been found in T. erythraeum.
An Intron-encoded Protein Assists RNA Splicing of Multiple Similar Introns of Different Bacterial Genes
Qing Meng, Yanfei Wang and Xiang-Qin Liu

J. Biol. Chem. 2005, 280:35085-35088.
doi: 10.1074/jbc.C500328200 originally published online September 7, 2005

Access the most updated version of this article at doi: 10.1074/jbc.C500328200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 19 references, 5 of which can be accessed free at http://www.jbc.org/content/280/42/35085.full.html#ref-list-1