Rare Group I Intron with Insertion Sequence Element in a Bacterial Ribonucleotide Reductase Gene[▽]

Qing Meng,[1,2] Yi Zhang,[3] and Xiang-Qin Liu[2]*

Institute of Biological Sciences and Biotechnology, Donghua University, Shanghai 201620, China[1]; Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, B3H 1X5, Canada[2]; and State Key Laboratory of Virology, College of Life Sciences, Wuhan University, Wuhan, Hubei 430072, China[3]

A rare group I intron in a cyanobacterial ribonucleotide reductase gene has been characterized. It contains a mobile insertion sequence element not required for RNA splicing. Ribonucleotide reductase genes were found to be hot spots for all three types of self-splicing intervening sequences, including group I and II introns and inteins.

Group I introns are ribozymes catalyzing their own RNA splicing to produce ligated exons (mature RNA) and the excised intron (4). Group I introns with dissimilar sequences can fold into similar structures consisting of up to 10 base-paired helices (stems) denoted P1 through P10, with P1 and P10 involving the 5′ exon and the 3′ exon, respectively (12, 16). Group I introns can persist in a host gene and spread across species via an intron homing process that is initiated by the intron-encoded homing endonuclease (2, 7). However, introns are extremely rare in bacterial protein-coding genes, and only one group I intron has been reported in a bacterial protein gene, which suggests strong barriers against intron insertions in such genes (5).

Identification of a rare group I intron in a bacterial protein gene. We found a group I intron in the cyanobacterial ribonucleotide reductase (*RIR*) gene of *Nostoc punctiforme* strain PCC73102 (Fig. 1). The intron boundaries were predicted by comparing the *N. punctiforme RIR* sequence to the closely related (but intron-lacking) *RIR* sequence of the cyanobacterium *Nostoc* sp. strain PCC7120. The predicted *N. punctiforme* RIR protein sequence is 83% identical and 91% similar to the *Nostoc* sp. strain PCC7120 RIR sequence, which is known as a class II B_{12}-dependent ribonucleotide reductase (6). Blast searches (1) of the recently completed genome sequence of *Nostoc punctiforme* (http://www.jgi.doe.gov/) did not reveal an additional *RIR* gene, which suggests that the group I intron of the *N. punctiforme RIR* gene must be active in RNA splicing in order to produce the functionally essential RIR protein.

We then demonstrated the intron's RNA-splicing activity in *Escherichia coli* cells (Fig. 2). The intron-coding DNA was PCR amplified from *Nostoc punctiforme* genomic DNA (ATCC product no. 29133D) and inserted in the T/A cloning site of the plasmid pDrive (QIAGEN). Short pieces of the native exon sequences (23 bp 5′ and 37 bp 3′) were included with the intron-coding sequence, which is usually needed for

efficient RNA splicing. *E. coli* cells (strain DH5α) containing the resulting plasmid (pDNR) were grown in liquid Luria broth medium at 37°C to mid-log phase (A_{600}, 0.3), and 0.8 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to induce transcription for 2 h. The cellular RNAs were extracted by using the RNAease kit (QIAGEN) and analyzed by reverse transcription (RT)-PCR using an RT-PCR kit (QIAGEN), and the resulting cDNA products were analyzed by agarose gel electrophoresis. This analysis revealed a 152-bp DNA product that corresponded to (was derived from) the spliced mRNA. Cloning and sequence determination (data not shown) of the 152-bp DNA product confirmed its identity and the predicted splice junctions.

Identification of a novel insertion sequence element (ISE) inside the group I intron. The 383-nucleotide (nt) intron sequence was predicted by the Mfold program (17), followed by manual folding, to fold into the structure of a group I intron (Fig. 3). The structure contains 9 of the 10 base-paired stems (P1 to P10) of typical group I introns but lacks a P2 stem. Its P8 loop contains an 86-nt sequence flanked by 8-nt direct repeats, which suggests an ISE. To determine whether this ISE is required for RNA splicing, a 94-nt ISE sequence (the 86-nt sequence plus one copy of the 8-nt repeat) was deleted from the intron coding sequence. The resulting mini-intron (plasmid pDNR-1) showed RNA-splicing activity comparable to that of the full-length intron (Fig. 2C).

BLAST searches of the complete genome sequence of *Nostoc punctiforme* revealed over 50 ISE-like sequences at dispersed locations in the genome, and some representatives are shown in Fig. 4. These ISE-like sequences are ~90% identical to the intron ISE over a 60-nt core region, most are flanked by terminal direct repeats 5 to 21 nt long, and all are located in intergenic sequences and not in introns. The sequence boundaries of the ISEs are not certain, but they are most likely at the terminal direct repeats, because mobile elements often have terminal repeats and because some members (e.g., ISE-3 and ISE-4) of the ISE family show sequence conservation, including the terminal direct repeats.

The family of ISE sequences in *N. punctiforme* likely originated as mobile elements a long time ago, because some of its members have diverged extensively outside the core sequence.
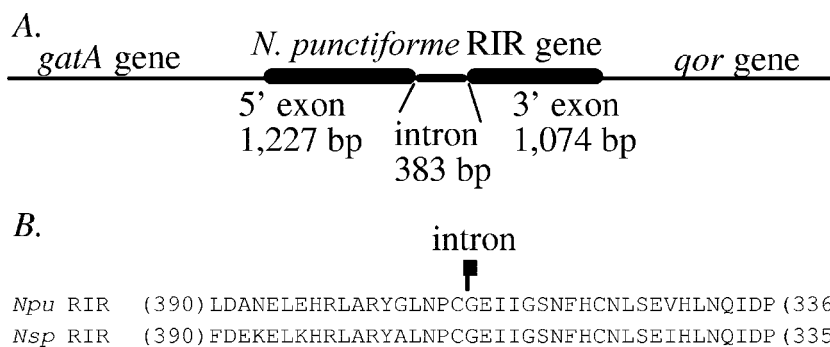

A.

FIG. 1. Sequence analysis of the *N. punctiforme RIR* gene. (A) Schematic illustration of the *N. punctiforme RIR* gene and its flanking genes. (B) Alignment of the predicted RIR protein sequence of *Nostoc punctiforme* (Npu) with the homologous sequence of *Nostoc* sp. strain PCC7120 (Nsp). Only sequences proximal to the intron insertion site (indicated) are shown, with the numbers of omitted amino acids shown in parentheses. The *N. punctiforme RIR* gene is flanked immediately upstream by the *gatA* (Glu-tRNA amino transferase subunit A) gene and downstream by the *qor* gene (encoding an alcohol dehydrogenase).
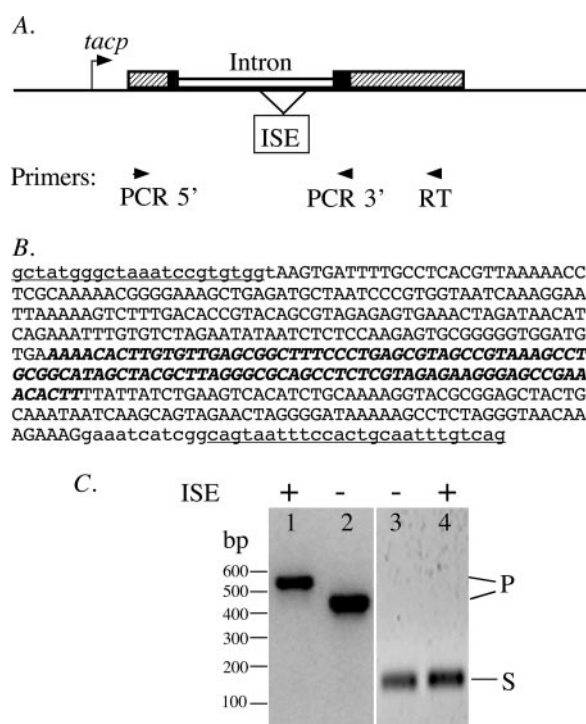

FIG. 2. RNA-splicing activity in *E. coli*. (A) Schematic illustration of the fusion gene in plasmid pDNR. The intron (open boxes) and its native exon sequences (black boxes) are inserted into the LacZα coding sequence (hatched boxes) downstream of the IPTG-inducible *tacp* promoter. Positions are marked for the ISE that was deleted in plasmid pDNR-1 and for the oligonucleotide primers used in RT-PCR. (B) Insertion sequence in plasmid pDNR. The 383-nt intron sequence is shown in uppercase letters and contains the 102-nt ISE sequence, shown in boldface italic letters. The 23-nt 5′ and the 37-nt 3′ native exon sequences are shown in lowercase letters and contain the PCR primer sequences (used for cloning the intron), which are underlined. (C) Detection of RNA-splicing products in *E. coli* cells. Lanes 1 and 2, PCR products directly from plasmids pDNR and pDNR-1, respectively; lanes 3 and 4, RT-PCR products from RNAs transcribed from plasmids pDNR-1 and pDNR, respectively. The plus and minus signs indicate the presence and absence, respectively, of an ISE in the intron. P and S mark the positions of DNA bands corresponding to the coding DNA (or precursor RNA) and the spliced RNA, respectively.

The high conservation of the core sequence may be due to a common structural constraint for a common function, such as recognition (and binding) by a common protein factor. It is formally possible that the ISEs may play regulatory roles in RNA splicing and gene expression in *N. punctiforme*, as was suggested for putative mobile elements in certain fungal mitochondria (14), although the ISEs and the mitochondrial elements do not have sequence similarity. Database searches did not reveal significant sequence similarity between the ISE and any other known sequences, and the 383-nt *N. punctiforme* RIR intron does not encode anything other than the ISE.

An intron/intein hot spot with evolutionary implications. We found the cyanobacterial *RIR* gene to be an exceptional hot spot for all three types of self-splicing intervening sequences, including a group I intron, a group II intron, and an intein (Fig. 5). The insertion sites of these intervening sequences are inside a highly conserved protein sequence motif (Fig. 5B) that is inside the catalytic center of the enzyme (15). This suggests a breakdown of barriers against intron insertion in the bacterial *RIR* gene. One suggested barrier is the fact that bacterial translation, which can start before transcription is completed, may interfere with RNA splicing (5, 13). Interestingly, there is a stop codon just 3 nt into the intron sequence, which may pause the translation until RNA splicing is completed.

The group I intron of the *N. punctiforme RIR* gene does not encode a homing endonuclease, which may be explained by the loss of the homing endonuclease through evolution. We propose an alternative hypothesis of intein-assisted intron homing (Fig. 6), which can also explain the close proximity of this intron and an intein that does have a homing endonuclease (11). This hypothesis is consistent with the known process of intein homing (3), except that it requires the presence of an intron-containing donor DNA in addition to an intein-containing donor DNA. After the intein homing endonuclease cuts the recipient DNA, the double-stranded break is expanded (by cellular exonucleases) into a gap that is expected to span both the intein insertion site and the nearby intron insertion site. During the subsequent double-stranded-break repair, the intein-containing donor DNA may serve as the template, leading to intein homing; alternatively, the intron-containing donor
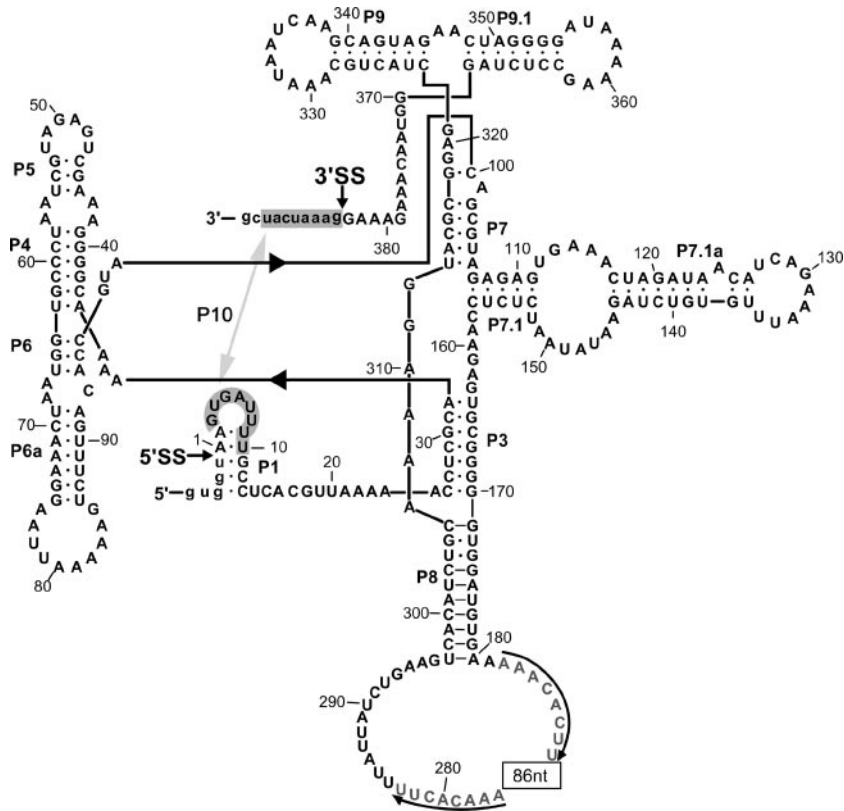
FIG. 3. Predicted intron folding. The exon-intron boundaries are marked with arrowheads and labeled as 5′ splice site (5′SS) and 3′ splice site (3′SS), with flanking exon nucleotides shown in lowercase letters. Typical group I intron features, which include the base-paired stems P1 and P3 to P10, are indicated, with P10 formed between the P1 loop and the 3′ exon (shaded). An insertion sequence element in the P8 loop consists of an 86-nt sequence (indicated by a box) and 8-nt flanking direct repeats (marked by arrows).

A.

```
ISE      AAACACTTGTGTTGAGCGGCTTTCCCTGA-GCGTAGCCGTAAAGCCTGCGGC
ISE-2    CAAACAATGCGATGTTCGCAATGGGCTATGACGCTCGATGACTCG-CTCTAAGCGTAAAGCCTACGGC
ISE-3    TGGGGAATTGGGAATTGGGGAACTTGTACTGA-GCGCAGTCGTAAAGCCTGCGGC
ISE-4    TGGGTAGTAGGGGAATGGGGAACTTGTACTGA-GCGCAGTCGTAAAGCCTGCGGC
ISE-5    ACTTGTACTGAGTTTCGACTGCGCTCAACTACCGCGCAGTCGTAAAGCCTGCGGC
ISE-6    GAATTACGAATTGCG-AACTTGTACTGA-GCGTAGTCGTAAAGCCTGCGGC
ISE-7    GGAAGACACCTTGTCTCTTCTTCATGCCCAATTCCCAATTCCCACTTG-CCCTGAGCGTAAAGCCTGCGGC
                             #**    ######  ************#****

ISE      ATAGCTACGCTTAGGGCGCAGCCTCTCGT-AGAGAAGGGAGCCGAAACACTT            102 bp
ISE-2    ATAGCTTCGTTTAGGGCGCAGCCTCTCAT-AGAGACGCTACGCAAACGCTA            117 bp
ISE-3    ATAGCTACGCTTAGGGCGTAGCCTCTCGTTAGAGAAGTATGGGGAATGGGGAATTGGGAA   114 bp
ISE-4    ATAGCTACGCTTAGGGCGTAGCCTCTCGTTAGAGAAGTATGGGGAATGGGGAATTGGGAA   114 bp
ISE-5    ATAGCTACGCTTAGGGCGCAGCCTCTCGT-AGAGAAGTATGGGGAATGGGGAA          107 bp
ISE-6    ATAGCTACGCTTAGGGCGCAGCCTCTCGT-AGAGAAGTATTGCTAATTG              98 bp
ISE-7    ATAGCTTCGCTTAACGCGTAGCGTCTCGT-AGAGAAGGCGTTCGCATAGCGTCCCGCAGGGAAG  132 bp
         *******#**#****##****#**#*****#*  *****#*
```

B.

Upstream Gene	Space (bp)	ISE-like sequence	Space (bp)	Downstream Gene
Transcriptional termintor protein	287	**ISE-2**	31	*mmsB*, 3-hydroxyisobutyrate dehydrogenase
Zn-dependent protease	6	**ISE-3**	48	<--- Calcineurin-like phosphoesterase
<--- *ATH1*, trehalose & maltose hydrolase	4	**ISE-4**	189	<--- *susB*, sucrose synthase
MoeA, molybdopterin biosynthesis	6	**ISE-5**	18	<--- *csdB*, selenocysteine lyase
<--- Glyoxalase	96	**ISE-6**	2	<--- Bacteriophytochrome
Transcription regulator	16	**ISE-7**	104	Conserved ORF

FIG. 4. Comparison of ISEs. (A) The ISE of the group I intron in the *N. punctiforme RIR* gene is aligned with sequences (ISE-2 through ISE-7) from other locations of the genome, with the length of each sequence indicated at the end and the terminal direct repeats underlined. Dashes represent gaps introduced to optimize the alignment, asterisks mark the positions of identical bases, and # marks the positions of just two different bases. (B) Names (or descriptions) of flanking genes for each ISE, with the distance (space) between the ISE and the flanking genes indicated. The genes marked by arrows are on the DNA strand opposite to the ISE. ORF, open reading frame.
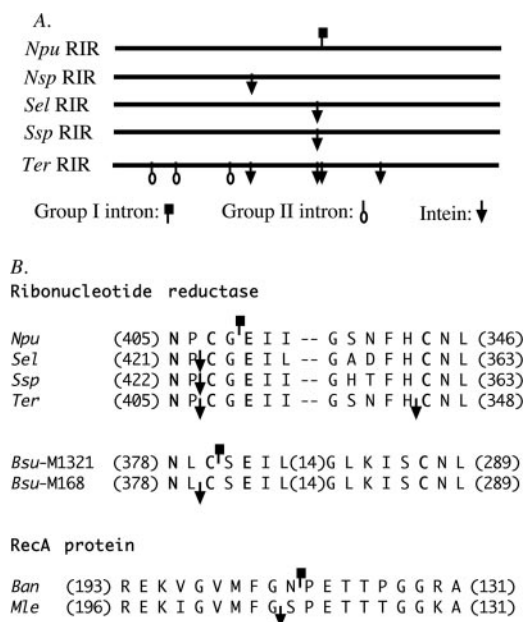
FIG. 5. Comparison of intron and intein insertion sites. (A) Schematic comparison of *RIR* genes. The symbols mark the insertion sites of introns and intein-coding sequences. (B) Protein sequence alignment showing insertion sites of introns and inteins close to each other in the ribonucleotide reductase genes and the RecA protein genes. Introns and inteins are represented by the same symbols as in panel A. Only conserved sequences near the intron and intein insertion sites are shown, with the numbers of omitted residues shown in parentheses and with dashes representing gaps introduced to optimize the alignment. Residues with known functions in ribonucleotide reductase are shown in boldface letters. The sequences compared are from *Nostoc punctiforme* (*Npu*), *Nostoc* sp. strain PCC7120 (*Nsp*), *Synechococcus elongatus* (*Sel*), *Synechococcus* sp. (*Ssp*), *Trichodesmium erythraeum* (*Ter*), *Bacillus subtilis* phages M1321 (*Bsu*-M1321) and M168 (*Bsu*-M168), *Bacillus anthracis* (*Ban*), and *Mycobacterium leprae* (*Mle*).
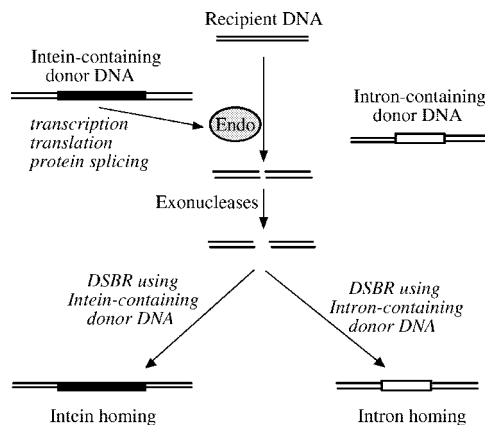

FIG. 6. A hypothetical model of intein-assisted intron homing. A bacterial cell with a recipient DNA acquires from its environment two homologous genes (donor DNAs) containing either an intein (black box) or an intron (open box). Expression of the intein-containing gene produces the intein homing endonuclease (Endo), which makes a cut on the recipient DNA at or near the intein insertion site that is expanded into a gap by cellular exonucleases. In the subsequent double-stranded-break repair (DSBR), either the intein-containing donor DNA or the intron-containing donor DNA may serve as the template. As a result, either the intein-coding sequence or the intron-coding sequence is copied into the recipient DNA, which results in intein homing or intron homing, respectively.

DNA may serve as the template, leading to intron homing. This model predicts that the recipient DNA may acquire either the intein or the intron, but not both, because the insertion of either the intein or the intron disrupts the recognition sequence of the intein homing endonuclease and prevents further cutting. This prediction is consistent with all three known cases of closely positioned inteins and introns in homologous genes (Fig. 5B), and no exception has been found. The three cases include cyanobacterial *RIR* genes in which the intein and intron are 6 bp apart, bacteriophage *RIR* genes (distantly related to the cyanobacterial *RIR* gene) in which the intein and intron are 3 bp apart (9, 10), and bacterial RecA genes in which the intein and intron are 3 bp apart (8). In all three cases, the intron does not encode a homing endonuclease, the intein has a homing endonuclease domain, and each individual gene has either the intron or the intein, but not both.

Nucleotide sequence accession number. The GenBank accession number of the group I intron in the cyanobacterial *RIR* gene of *Nostoc punctiforme* strain PCC73102 is AAAY02000101.

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
2. **Belfort, M., V. Derbyshire, B. Cousineau, and A. M. Lambowitz.** 2002. Mobile introns: pathways and proteins, p. 761–783. *In* N. Craig, R. Craigie, M. Gellert, and A. M. Lambowitz (ed.), Mobile DNA II. ASM Press, Washington, DC.
3. **Belfort, M., and R. J. Roberts.** 1997. Homing endonucleases: keeping the house in order. Nucleic Acids Res. **25:**3379–3388.
4. **Cech, T. R.** 1990. Self-splicing of group I introns. Annu. Rev. Biochem. **59:**543–568.
5. **Edgell, D. R., M. Belfort, and D. A. Shub.** 2000. Barriers to intron promiscuity in bacteria. J. Bacteriol. **182:**5281–5289.
6. **Gleason, F. K., and N. E. Olszewski.** 2002. Isolation of the gene for the B₁₂-dependent ribonucleotide reductase from *Anabaena* sp. strain PCC 7120 and expression in *Escherichia coli*. J. Bacteriol. **184:**6544–6550.
7. **Haugen, P., D. M. Simon, and D. Bhattacharya.** 2005. The natural history of group I introns. Trends Genet. **21:**111–119.
8. **Ko, M., H. Choi, and C. Park.** 2002. Group I self-splicing intron in the *recA* gene of *Bacillus anthracis*. J. Bacteriol. **184:**3917–3922.
9. **Landthaler, M., and D. A. Shub.** 1999. Unexpected abundance of self-splicing introns in the genome of bacteriophage Twort: introns in multiple genes, a single gene with three introns, and exon skipping by group I ribozymes. Proc. Natl. Acad. Sci. USA **96:**7005–7010.
10. **Lazarevic, V.** 2001. Ribonucleotide reductase genes of *Bacillus* prophages: a refuge to introns and intein coding sequences. Nucleic Acids Res. **29:**3212–3218.
11. **Liu, X. Q., J. Yang, and Q. Meng.** 2003. Four inteins and three group II introns encoded in a bacterial ribonucleotide reductase gene. J. Biol. Chem. **278:**46826–46831.
12. **Michel, F., and E. Westhof.** 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J. Mol. Biol. **216:**585–610.
13. **Ohman-Heden, M., A. Ahgren-Stalhandske, S. Hahne, and B. M. Sjoberg.**

1993. Translation across the 5′-splice site interferes with autocatalytic splic-
ing. Mol. Microbiol. **7:**975–982.

14. **Paquin, B., M. J. Laforest, and B. F. Lang.** 2000. Double-hairpin elements in
    the mitochondrial DNA of allomyces: evidence for mobility. Mol. Biol. Evol.
    17:1760–1768.

15. **Sintchak, M. D., G. Arjara, B. A. Kellogg, J. Stubbe, and C. L. Drennan.** 2002.
    The crystal structure of class II ribonucleotide reductase reveals how an allos-
    terically regulated monomer mimics a dimer. Nat. Struct. Biol. **9:**293–300.

16. **Vicens, Q., and T. R. Cech.** 2006. Atomic level architecture of group I introns
    revealed. Trends Biochem. Sci. **31:**41–51.

17. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization
    prediction. Nucleic Acids Res. **31:**1–10.