
Novel RNA structural features of an alternatively splicing group II intron from *Clostridium tetani*

BONNIE A. MCNEIL¹ and STEVEN ZIMMERLY^{1,2}

¹Department of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada

ABSTRACT

Group II introns are ribozymes in bacterial and organellar genomes that function as self-splicing introns and as retroelements. Previously, we reported that the group II intron *C.te.I1* of *Clostridium tetani* alternatively splices in vivo to produce five distinct coding mRNAs. Accurate fusion of upstream and downstream reading frames requires a shifted 5' splice site located 8 nt upstream of the usual 5' GUGYG motif. This site is specified by the ribozyme through an altered intron/exon-binding site 1 (IBS1–EBS1) pairing. Here we use mutagenesis and self-splicing assays to investigate in more detail the significance of the structural features of the *C.te.I1* ribozyme. The shifted 5' splice site is shown to be affected by structures in addition to IBS1–EBS1, and unlike other group II introns, *C.te.I1* appears to require a spacer between IBS1 and the GUGYG motif. In addition, the mechanism of 3' exon recognition is modified from the ancestral IIB mechanism to a IIA-like mechanism that appears to be longer than the typical single base-pair interaction and may extend up to 4 bp. The novel ribozyme properties that have evolved for *C.te.I1* illustrate the plasticity of group II introns in adapting new structural and catalytic properties that can be utilized to affect gene expression.

Keywords: ribozyme; self-splicing; bacteria; surface layer protein

INTRODUCTION

Group II introns are large catalytic RNAs and retroelements found within prokaryotic genomes, as well as in organellar genomes of some eukaryotes (Ferat and Michel 1993; Michel and Ferat 1995; Dai and Zimmerly 2003; Lambowitz and Zimmerly 2004). The ribozymes possess a conserved secondary structure of six domains (D1–6), of which domains 1 (D1) and 5 (D5) are absolutely required for catalysis (Koch et al. 1992; Michels and Pyle 1995). D5 contains the catalytic core of the introns (Jarrell et al. 1988b; Chanfreau and Jacquier 1994), while D1 contains elements that are responsible for correctly positioning the 5' and 3' exon sequences at the active site (Jacquier and Michel 1987; Michel and Jacquier 1987; Costa et al. 2000). In prokaryotes, D4 almost without exception contains an open reading frame (ORF) encoding a reverse transcriptase (RT)-containing, multidomain protein known as the intron-encoded protein (IEP).

Group II ribozymes catalyze their excision from RNA transcripts via two sequential transesterification reactions with a lariat intermediate (Peebles et al. 1986; Schmelzer and Schweyen 1986; Van der Veen et al. 1986). Typically, D6 contains a bulged adenosine residue, analogous to the branch-point residue of spliceosomal introns, which initiates the splic-

ing reaction. The 2' OH of the bulged adenosine conducts a nucleophilic attack at the 5' splice site resulting in the creation of an intron-3' exon intermediate and free 5' exon. A slight shift in the conformation of the ribozyme then positions the 3' OH of the 5' exon at the 3' splice site (Chanfreau and Jacquier 1996); a subsequent nucleophilic attack at this site results in the ligation of exons and the release of a branched intron lariat. Although the intron is typically excised as a branched lariat molecule, some natural intron variants and certain in vitro conditions favor the use of a hydrolysis reaction to initiate splicing, which results in the production of linear intron (Jarrell et al. 1988a; Vogel and Börner 2002; Li-Pook-Than and Bonen 2006; Toor et al. 2006). While many group II introns can self-splice in vitro, either the IEP or other host-encoded proteins are required to facilitate efficient splicing in vivo. After splicing in vivo, the IEP remains associated with the intron lariat RNA, forming a ribonucleoprotein (RNP) complex that is capable of inserting intron sequence into genomic sites through a retromobility reaction (Zimmerly et al. 1995; Belfort et al. 2002; Lambowitz and Zimmerly 2004).

Structurally, group II ribozymes are divided into three types, IIA, IIB, and IIC (Michel et al. 1989; Toor et al.

²Corresponding author

E-mail zimmerly@ucalgary.ca

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.042440.113>.

© 2014 McNeil and Zimmerly This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

2001), which are distinguished by their RNA secondary structures and the mechanisms of 5' and 3' exon recognition (below). An alternative classification is possible through phylogenetic analysis of the IEPs encoded within the ribozymes, which identifies eight classes of introns in bacteria: A, B, C, D, E, F, ML, and CL (Zimmerly et al. 2001; Toro et al. 2002; Simon et al. 2008; Toro and Martínez-Abarca 2013). For bacterial introns, there is a pattern of coevolution between intron RNAs and IEPs (Fontaine et al. 1997; Toor et al. 2001), such that the IEP phylogenetic classes correspond to RNA structural classes. Thus, Class ML introns possess IIA ribozyme structures, Class CL introns have typical IIB structures, Class C introns have IIC RNA structures, and the remaining classes have ribozyme structures that fall under the broad umbrella of IIB introns (Simon et al. 2009). In this study, *C.te.II* is referred to as a IIB intron of Class B; however, it should be noted that the two terms are not synonymous. IIB refers to the RNA structure, while Class B refers to the IEP; Class B introns are in fact a subset of IIB introns.

Accurate splicing function relies on precise selection of both 5' and 3' splice sites, which is achieved through base-pairing interactions between the ribozyme and the exon sequences. The mechanisms utilized differ between intron structural types. IIA and IIB introns share a common mechanism of 5' splice-site recognition, with two sets of base-pairings that form an extended recognition region of ~12 bp. IBS1–EBS1 (intron/exon-binding site 1) is a 6-bp interaction between EBS1 in D1 of the intron and IBS1 of the 5' exon (Jacquier and Michel 1987), while IBS2–EBS2 occurs between another sequence in D1 (EBS2) and the sequence immediately upstream of IBS1 in the exon (IBS2). IIC introns have only a 4-bp IBS1–EBS1 pairing and no IBS2–EBS2 pairing (Robart et al. 2007). In nearly all group II introns, the IBS1 sequence is located directly upstream of the 5' GUGYG start of the intron, with no spacer between them.

Recognition of the 3' splice site is also conferred by sequences within D1, with IIB and IIC introns sharing a common mechanism and IIA introns using a distinct mechanism. For IIB and IIC introns, there is a 1-bp interaction between the first position of the exon (IBS3) and the EBS3 nucleotide, which is located in a D1 substructure known as the coordination loop (Costa et al. 2000). In contrast, IIA introns lack the coordination loop motif and position the 3' exon through a base pair between the first position of the exon (called δ for IIA introns) and δ' , which is next to EBS1 (Michel and Jacquier 1987). The γ – γ' pairing might also be considered a component of 3' exon definition, because it positions the final nucleotide of the intron for all three intron classes (Jacquier and Michel 1990; Robart et al. 2004).

Recently, we reported an ORF-less group II intron in *Clostridium tetani* that possesses a novel genomic arrangement consisting of a full-length copy of the intron followed by three downstream copies of D5 and D6 motifs (Fig. 1A). The intron undergoes alternative splicing in vivo to form four spliced exon forms, along with an unspliced transcript

(McNeil et al. 2014). Critical to proper fusion of reading frames in upstream and downstream exons is a shifted 5' splice site located 8 nt upstream of the usual 5' GUGYG motif. The shifted site eliminates the stop codon of the 5' exon and allows correct fusion of the upstream and downstream exon reading frames. While the shifted site was shown to be due to an altered IBS1–EBS1 sequence, it was also observed that the secondary structure possesses other atypical features that might contribute to the intron's function in alternative splicing (McNeil et al. 2014). In this study, we investigate the functional significance of a variety of structural components of the intron ribozyme, and we consider how the intron evolved from a mobile retroelement to an alternatively splicing intron. Overall, we can infer that *C.te.II* is descended from a conventional IIB/Class B ribozyme and that its structural elements evolved to produce modified activities needed for alternative splicing, particularly with regard to 5' exon recognition.

RESULTS

Unique secondary structure features of *C.te.II*

The secondary structure of *C.te.II* (Fig. 1) is a IIB structure belonging to Class B (McNeil et al. 2014). This was initially observed based on the closest intron relatives using a local BLASTN database search tool (Candales et al. 2012; <http://webapps2.ucalgary.ca/~groupii/index.html#>), in which the five introns most closely related to *C.te.II* all belonged to Class B. Consistent with the classification, the *C.te.II* secondary structure has features shared by essentially all Class B introns (Fig. 1), including a CGC “catalytic triad” in D5 rather than AGC, and a double stem insertion in the I(i,ii) stem near the start of the intron (dotted box). Two additional features shared by most but not all Class B introns include a short stem 3' to the α' sequence (dotted box) and the lack of a β – β' interaction (Toor et al. 2001; BA McNeil, L Wu, AM Jarding, S Zimmerly, unpubl.). Finally, and most specifically, *C.te.II* has structural features typical of a subset of Class B introns denoted the β lineage (Stabell et al. 2009), including the presence of a domain IA stem-loop motif and an extra stem located 5' to the α' sequence (dotted boxes with asterisks). Thus, it appears that *C.te.II* is derived from a mobile intron of Class B, probably from the β lineage, which then lost its IEP and acquired various RNA structural modifications. Interestingly, the 311-nt loop in domain 4 bears no remnant of the IEP sequence that might previously have resided there.

In addition to the structural features typical of IIB and Class B introns, *C.te.II* possesses features that deviate from the Class B consensus. Most of the irregularities are in a region in D1 that is responsible for binding and recognition of the 5' exon. Most notably, the EBS1 sequence of *C.te.II* is located within a loop of 24 nt (shaded dotted box), whereas other Class B introns have a loop of 8–11 nt. Additionally, the stem delimiting the EBS1 loop does not have the internal bulge

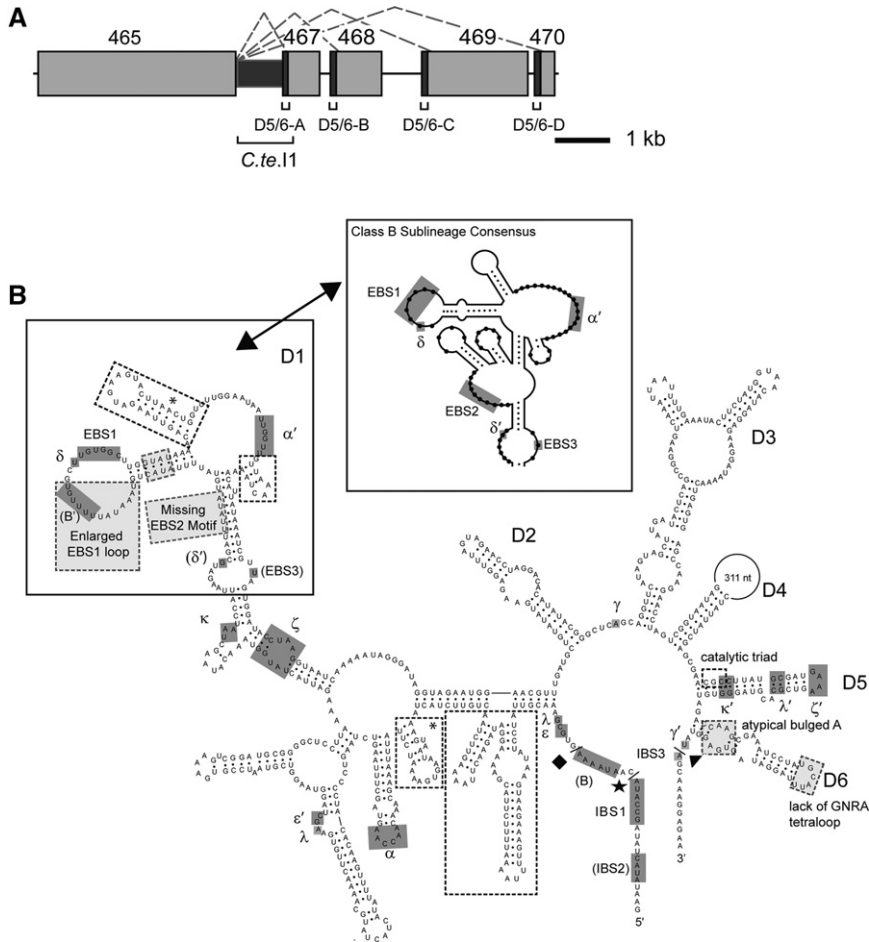


FIGURE 1. Genomic locus and secondary structure of *C.te.I1*. (A) Genomic organization of *C.te.I1* (drawn to scale) showing the intron (dark gray), upstream exon ORF, and four downstream exon ORFs (light gray, outlined boxes). Domains 1–4a of the intron are shown by the medium-sized gray box, while the four domain 5/6 motifs for the 3' exons are dark gray, outlined boxes. Thin black lines between exon sequences and intron D5/6 sequences represent noncoding spacer sequences. Dotted lines indicate the four alternative splicing reactions observed in vivo. ORFs 465–470 refer to annotated ORFs CTC00465–CTC00470. (B) Secondary structure model of *C.te.I1*. Locations of typical major tertiary interactions are noted (α – α' , γ – γ' , ϵ – ϵ' , ζ – ζ' , κ – κ' , and λ – λ'). Interactions in parentheses are potential interactions or elements not supported by mutagenesis experiments (δ' , IBS2–EBS2, EBS3, B–B'). Locations of both the expected IBS1 (based on the 5' consensus of group II introns) and actual IBS1 (experimentally supported), and the corresponding intron boundaries, are shown with a diamond and star, respectively. The 3' intron junction is shown with a triangle. All tertiary interactions are noted by medium gray shading without outlines. Light gray shading with dotted outlines indicates unexpected structural features for IIB or Class B introns, while unfilled boxes with dotted lines indicate features typical of Class B introns. Open boxes with asterisks denote features found in the β sublineage of Class B introns. The *inset* box shows a partial consensus secondary structure of the β sublineage of Class B introns, shown to indicate the deviations in the EBS1/EBS2 region of *C.te.I1*.

motif found for most other Class B introns (shaded dotted box), which is found in an analogous location as the ω' motif of IIC introns. While the structural significance of the internal bulge is unknown in Class B introns, it may function similarly to the ω – ω' interaction in IIC introns, which anchors the EBS1 stem–loop to the ribozyme via ribose zipper contacts (Toor et al. 2008). Another unusual feature of *C.te.I1* is that the EBS2 motif, which normally contributes to 5' exon recognition, appears to be missing (shaded dotted box). Two alternate

secondary structure foldings allow for the formation of a potential but modest EBS2 pairing; however, they were not supported by mutagenesis experiments (below). In other regions of the RNA structure, the D6 stem lacks the distal GNRA tetraloop that normally participates in the η – η' interaction (Chanfreau and Jacquier 1996), and the branch A motif does not have a clearly bulged A residue. Interestingly, the structural variations in D6 are present in all of the downstream copies of D5/6 that are utilized in alternative splicing (McNeil et al. 2014).

C.te.I1 self-splices robustly in vitro despite the structural irregularities

C.te.I1 was characterized for self-splicing using a construct in which the 311-nt loop of D4 was replaced with a 26-nt loop, which here is considered the wild-type intron (WT) (Materials and Methods) (McNeil et al. 2014). A construct without the deletion of the D4 loop was also tested for self-splicing in vitro and possesses very similar self-splicing properties as the 26-nt loop construct (data not shown). Initial experiments revealed that the WT intron reacted robustly, with splicing occurring even in the T7 transcription reaction that synthesized intron precursor. To eliminate the premature splicing, subsequent transcription reactions were performed with only 4 mM magnesium. Reaction optimization showed that splicing occurred at magnesium concentrations as low as 10 mM, although 100 mM was optimal (data not shown). Splicing occurred across a broad range of temperatures tested, from 38° to 63°C, but most efficiently between 53°C and 58°C (data not shown). Different concentrations (0.25–1.0 M) of monovalent salts (NH₄Cl, NaCl, KCl) had little effect on splicing (data not shown). Standard self-splicing conditions for subsequent experiments were 40 mM Tris (pH 7.5), 0.5 M NH₄Cl, and 100 mM MgCl₂. Under these conditions the intron self-spliced very rapidly, with ~40% of intron fully spliced after 30 sec and 73% spliced after 2 min (Fig. 2; Supplemental Table 1). As previously reported, *C.te.I1* self-splices through the hydrolytic pathway to produce linear intron (McNeil et al. 2014). This held true in all salt and divalent metal ion conditions tested (data not shown), although two intron

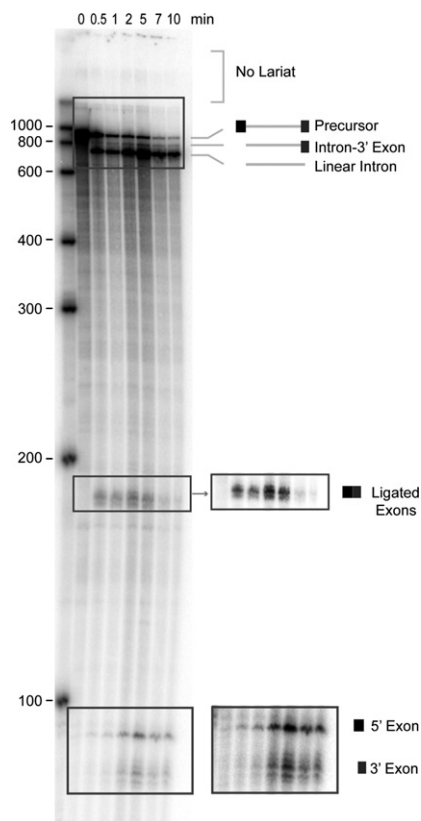


FIGURE 2. Self-splicing assay of the wild-type *C.te.II* construct. Splicing proceeds through a hydrolysis pathway with ~40% of precursor being converted into linear intron within 30 sec, and without intron lariat being observed. The regions containing ligated exons, 5' exon, and 3' exon are shown at increased intensity to the right to visualize the products more distinctly. Evidence for a post-splicing SER (spliced exon reopening) reaction is seen by the disappearance of ligated exon bands at the 7- and 10-min time points.

mutants induced formation of a slow migrating product (below). Hydrolysis of ligated exons indicates that a substantial level of SER (spliced exon reopening) (Jarrell et al. 1988a) occurs for *C.te.II* following the splicing reaction (7, 10 min) (Fig. 2).

Testing the secondary structure model

Mutagenesis and compensatory mutagenesis experiments were done to test some of the secondary structure elements of *C.te.II*. First, the I(ii) stem was tested, as it essentially bisects the ribozyme. Consistent with the structure model, disruption of either strand eliminated splicing, while compensatory mutations restored function (Stem 1 in Fig. 3A, B; quantitation in Supplemental Table 2). Similarly, the α - α' pairing was supported by mutagenesis. Disruption of the α' sequence eliminated splicing, while the α disruption had a modest effect (Fig. 3C; Supplemental Table 2); the compensatory mutations rescued splicing and confirmed the identity of the important interaction.

Stem 2, which delimits the EBS1 sequence, was tested through the simultaneous mutation of four base pairs (Fig. 3A,D). In this case, mutation of the 5' strand had a modest effect on splicing, while the mutation of the 3' strand blocked splicing after the first step. Thus, both sequences play a role in splicing; however, the compensatory mutations abolished all reactivity of the ribozyme, failing to support the proposed pairing (Supplemental Table 3). Although these mutagenesis experiments fail to support the hypothesized base pairs, further experiments argue against the possibility of alternate structures (e.g., structure 1 in Fig. 3A). The experiments show that the single-stranded nucleotides in the EBS1 loop cannot be involved in important pairings, because these sequences can be mutated and deleted with little effect (mB', mEBS1loop, Δ EBS1loop; below). As such, this leaves little choice for the modeling of the EBS1 region other than a structure that includes Stem 2. We rationalize the data as being due to disruption of critical tertiary interactions for Stem 2, in addition to the disruption of its helicity. In the crystal structure of the *Oceanobacillus iheyensis* IIC intron, the analogous stem is involved in formation of the ω' motif, which docks the stem to helix ID through a ribose zipper interaction (Toor et al. 2008). If a similar tertiary interaction exists for class B introns, and in particular *C.te.II*, then more subtle mutations in Stem 2 may be required to verify the hypothesized pairing.

In intron Classes D, E, and F, the α' sequence is contained within a stem-loop, and such a structure can be found in alternate foldings of *C.te.II* (Stem 4 in Fig. 3A). This feature is not typical of Class B introns, but it was tested as a possible secondary structure feature for *C.te.II*. None of the mutations affected splicing at discernible levels (Supplemental Fig. 1; Supplemental Table 3). Although this gives no information about the hypothesized pairing, it does show that neither of the mutated sequences plays a critical role in splicing. In order to be consistent with other Class B structures, we model the α' sequence in an open loop.

Similarly, the stem located immediately 5' to α' in the β lineage of class B introns was tested (Fig. 3A, Stem 3). Again, self-splicing assays showed no effect for either of the disruptive mutations, or the compensatory mutation (Supplemental Fig. 1; Supplemental Table 3), revealing a lack of importance for either mutated sequence. We choose to model Stem 3 in the secondary structure despite the lack of supporting evidence, as it is consistent with the consensus of the β lineage of Class B introns, and because there are no other obvious candidate foldings for the sequence.

Finally, the possibility of an EBS2 sequence was tested (Fig. 3E). While a typical EBS2 motif was not initially evident in the secondary structure, an alternate pairing allows a modest 4-bp IBS2-EBS2 interaction (rather than the usual 6-bp interaction) (Fig. 3A). Mutation of individual strands in the hypothetical helix had no discernible effect on splicing. The lack of splicing disruption caused by the IBS2 mutation suggests the absence of any IBS2-EBS2 pairing, not just the one tested

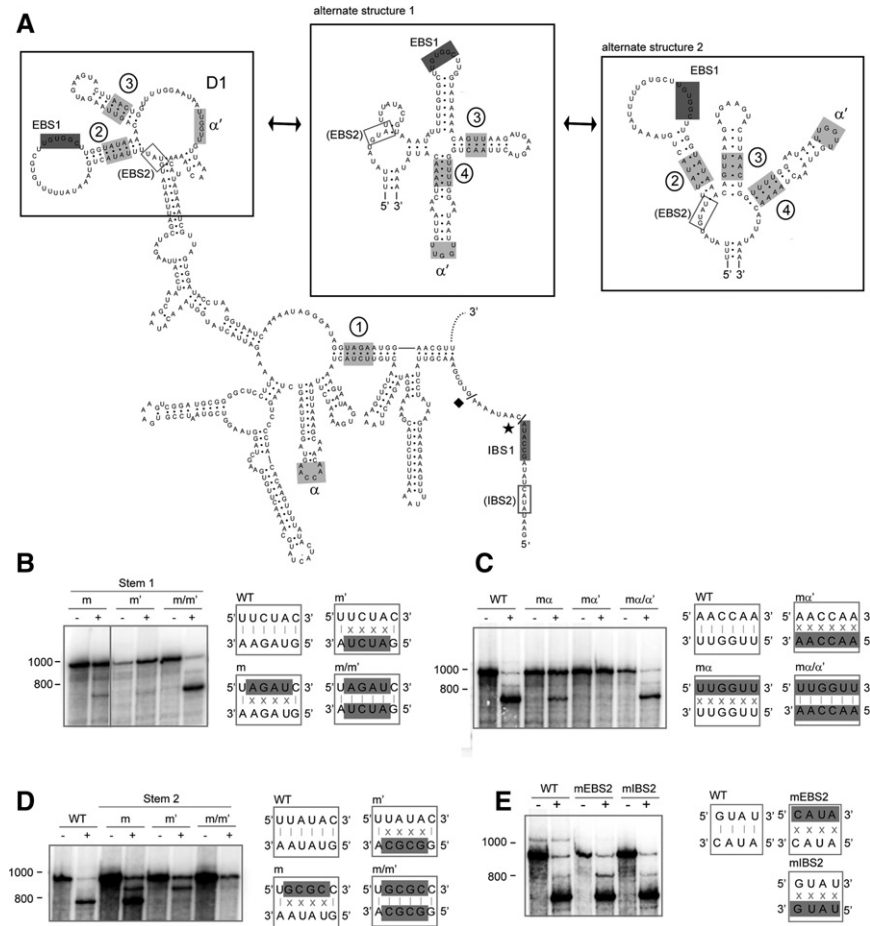


FIGURE 3. Testing secondary structure features by mutagenesis. (A) Locations of the hypothesized pairings in the secondary structure, showing domain 1 only. The tested stems 1, 2, 3, and 4 are labeled with circled numbers. Mutagenesis data for Stems 3 and 4 are shown in Supplemental Figure 1. Two alternate secondary structures for the EBS1 region are shown in boxes. The unsupported EBS2 and IBS2 sequences are in outlined gray boxes. (B–E) For each pairing interaction tested, the wild-type sequence is shown along with individual mutations for each strand (m and m') and the compensatory mutation (m/m'). The hypothesized pairing is shown to the right of each gel, with mutated residues shaded in gray and X's indicating mispairings. The lanes “+” and “–” denote incubation in self-splicing buffer or in self-splicing buffer lacking magnesium, respectively. Only regions of the gels corresponding to precursor, intron-3' exon, and linear intron are shown; the remainder of the gels do not contain bands different from those in Figure 2. (B) Confirmation of the I(ii) pairing (Stem 1). (C) Confirmation of the α - α' pairing. Self-splicing conditions of 30 mM MgCl₂ (rather than 100 mM) were used to reveal the disruption and rescue. (D) Testing of the EBS1 stem sequence (Stem 2). The two mutations are disruptive to different degrees, but activity is not restored with the simultaneous mutation, failing to give support to the hypothesized pairing. (E) Mutations of the potential EBS2 and IBS2 sequences do not impact self-splicing, giving no evidence for the hypothesized pairing.

in Figure 3E. Together, in light of all of the mutagenesis data, we consider the structure presented in Figure 1B to be the most likely secondary structure of *C.te.I1*, despite the remaining ambiguities in some of the data.

C.te.I1 interactions with the 5' splice site

We previously identified the 6-bp EBS1-IBS1 interaction depicted in Figure 1 and showed that the intron-exon pairing specifies the shifted 5' splice site that is required for correct fu-

sion of reading frames during alternative splicing (McNeil et al. 2014). Figure 4A repeats that observation but also shows that when either the IBS1 or EBS1 sequence is mutated, other bands appear that probably represent artifactual reactions. Each mutant shows two slowly migrating bands (asterisks), whose migration is consistent with a branched or circular form. In addition, the EBS1 mutant (m) shows a band that migrates slightly slower than intron-3' exon (black dot), and the IBS1 mutant (m') shows two bands that do not migrate as either excised intron or intron-3' exon (black dots). All of these bands can be considered side reactions due to IBS1-EBS1 mispairings or improper interactions with the exons. This inference is supported by the fact that the bands are not formed when pairing is restored (m/m').

To characterize the products formed, ligated exons generated by the IBS1 mutant (m') were amplified by RT-PCR to detect ligation products not detectable in the splicing gel (Fig. 4A). Sequencing showed the use of a cryptic 5' splice site located 2 nt downstream from the wild-type splice site, which has partial complementarity to EBS1 (data not shown). Inverse RT-PCR of total reaction products for the IBS1 mutant revealed a junction consisting of the 3' portion of the intron followed by the last 38 nt of the 5' exon, followed by the 5' end of the intron (data not shown). This junction suggests that the 3' end of the intron is linked to a cryptic splice site located 38 nt upstream of the actual splice site of the intron, and conceivably could correspond to a circular form visible in Figure 4A.

For the EBS1 mutant (m), a distinct splicing product is seen that migrates slightly above the position of intron-3' exon. This band may be intron-3' exon formed when splicing uses a cryptic upstream 5' splice site, which would generate a product longer than the normal intron-3' exon size. Another explanation is that it could be a 5' exon-intron, since a 3' exon product is visible in the lane, but not a 5' exon. This product assignment would be consistent with the slow-migrating bands (asterisks) being excircles, because a 5' exon-intron RNA is hypothesized to be an intermediate in the intron circle-forming reaction (Murray et al. 2001). Altogether, it is clear that IBS1-EBS1 mispairings severely impair splicing and cause artifactual reactions to take

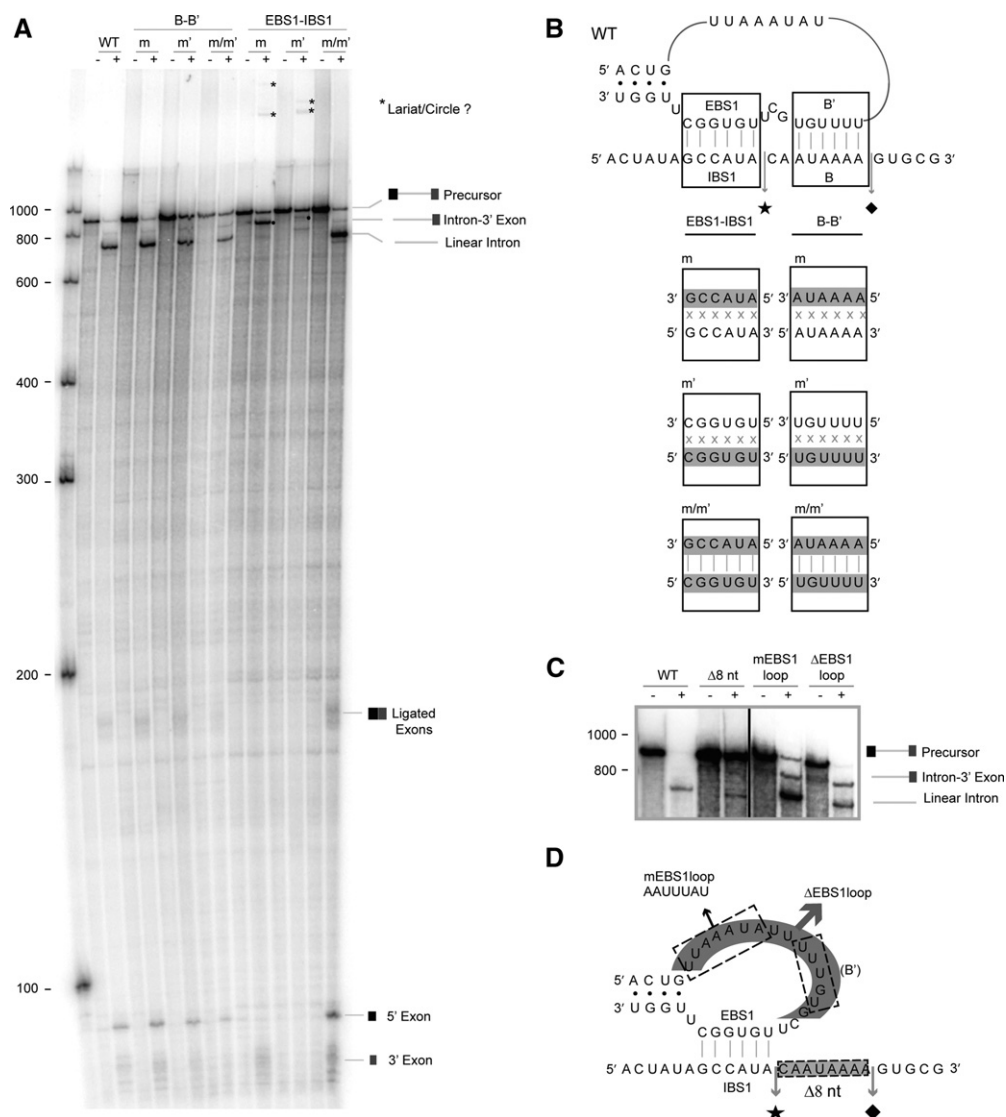


FIGURE 4. Mutagenesis of potential 5' splice-site recognition elements. (A) Self-splicing assays for mutations of EBS1-IBS1 and the hypothesized pairing B-B'. Lanes “+” and “-” denote incubation in self-splicing buffer or in self-splicing buffer lacking magnesium, respectively. For each of the EBS1-IBS1 mutations, slow-migrating product bands are formed that could potentially be intron lariat or circle (asterisks); however, the bands are not seen for the compensatory mutations. (B) The sequences of mutations and compensatory mutations for the constructs tested in A. The full EBS1 stem-loop is shown on top, paired to the 5' exon-intron junction. The star and diamond denote the actual and predicted intron boundaries, respectively. (C) Self-splicing assays of the mutations ΔEBS1loop, mEBS1loop, and Δ8nt IBS1. (D) Schematic of the mutations in the same format as B. The sequence mutated in mEBS1loop is shown by the dotted box and arrows, and the sequences deleted in the ΔEBS1loop and Δ8nt IBS1 mutations are denoted by gray shading. In addition, the positions mutated for B' are indicated.

place when the exon substrates are not properly recognized or positioned by the ribozyme.

The large size of the 24-nt EBS1 loop is quite unusual and its function could be understood in several ways. The large loop might contain additional functional elements involved in self-splicing, or it might be a flexible linker whose length is required but whose sequence is unimportant, or it might have little or no function. In considering the possibility of functional elements in this loop, we noticed a potential 6-bp pairing that could extend the EBS1-IBS1 pairing, which we call B-B' (Fig. 4B). Notably, the B sequence

is part of the spacer between IBS1 and the 5' GUGYG motif. Mutation of the B' sequence in the loop had no discernible effect on self-splicing (Fig. 4; Supplemental Table 4), failing to support the B-B' pairing, and showing that B' does not have an important function, at least with regard to its sequence. In contrast, mutating the B sequence next to IBS1 showed a modest decrease in splicing (m', m/m'), suggesting that its sequence has some significance; however, we could identify no other sequence in the intron that could pair with B, and so the reason for the sequence preference is unclear.

To further examine the role of the 24-nt loop, a 15-nt deletion was made to convert the loop to the size of other Class B introns (Δ EBS1loop). Strikingly, the large deletion only modestly affected splicing, resulting in elevated levels of intron-3' exon intermediate (Fig. 4C,D; Supplemental Table 5). Another sizable mutation of 7 nt in the loop (mEBS1loop) had a small effect on splicing (Fig. 4C,D). Together, the three mutations of B', mEBS1loop, and Δ EBS1loop indicate that the loop does not have a critical role in splicing or 5' exon choice, because nearly all of the "extra" sequence of the loop can be mutated or deleted. On the other hand, the loop still has a small effect on splicing of *C.te.II*, through undefined means, because of modest effects of the mutations on self-splicing. An important point regarding the secondary structure model is that the dispensability of the loop sequence (Fig. 4C,D) argues strongly against alternate secondary structure 1 (Fig. 3A), or against other folding possibilities that do not include Stem 2, because the loop sequence cannot be a part of any important structural element.

Finally, we tested the intron's ability to splice at the canonical location next to the 5' GUGYG motif, by deleting the 8-nt spacer between the actual splice site and the 5' GUGCG (Fig. 4C,D). Despite the fact that the deletion mutation (Δ 8 nt) does not disturb IBS1–EBS1 complementarity, this mutation blocked splicing almost completely. A faint band appears in the reaction products that might be linear intron, but it is clear that splicing efficiency is at most a small fraction of the wild-type ribozyme (Supplemental Table 5). We conclude that the 8-nt spacer is critical for efficient splicing, reflecting an alteration of the ribozyme properties compared with conventional group II introns. The sequence of the spacer appears to be less critical, although it has a modest significance (B mutation in Fig. 4A,B). Thus, the data demonstrate that the *C.te.II* ribozyme is differentiated from the ancestral, retro-mobile form, which would have spliced directly next to the canonical 5' GUGYG motif.

***C.te.II* uses a IIA-like mechanism for 3' splice-site selection**

Because *C.te.II* is a IIB intron, it is expected to recognize the 3' exon through an IBS3–EBS3 base pair between IBS3 of the exon and EBS3 of the coordination loop (Fig. 5B). In addition, it is expected to have a δ – δ' pairing that helps stabilize the coordination loop relative to EBS1 (Fig. 5B). A motif resembling the coordi-

nation loop can be found in the *C.te.II* secondary structure (Figs. 1, 5B) and consequently the IBS3–EBS3 and δ – δ' pairings were tested by mutagenesis. Mutation of IBS3 in the 3' exon indeed caused a deficiency for the second step of splicing, with 30% of the products accumulating as intron-3' exon intermediates (Fig. 5C; Supplemental Table 6); however, mutation of the putative EBS3 nucleotide in the coordination loop had a negligible effect on splicing. Surprisingly, the compensatory IBS3/EBS3 mutation had an even greater splicing defect than the single EBS3 mutation (Fig. 5C), arguing against the hypothesized IBS3–EBS3 pairing. Mutation of δ and δ' also failed to support the IIB-like mechanism. Mutation of δ caused ~50% of the reaction products to accumulate as intron-3' exon intermediate, while the δ' mutation showed very little effect on splicing, with 80% of the products fully spliced, and only 9% of the products detectable as intron-3' exon intermediates (Fig. 5D; Supplemental Table 6).

In the absence of a IIB-like mechanism, we considered the possibility of a IIA-like mechanism. Indeed, the combined mutations of δ and IBS3 restored the splicing defects seen for the individual δ and IBS3 mutations (m δ , mEBS3, Fig. 5C,D), indicating that the two nucleotides form a base pair.

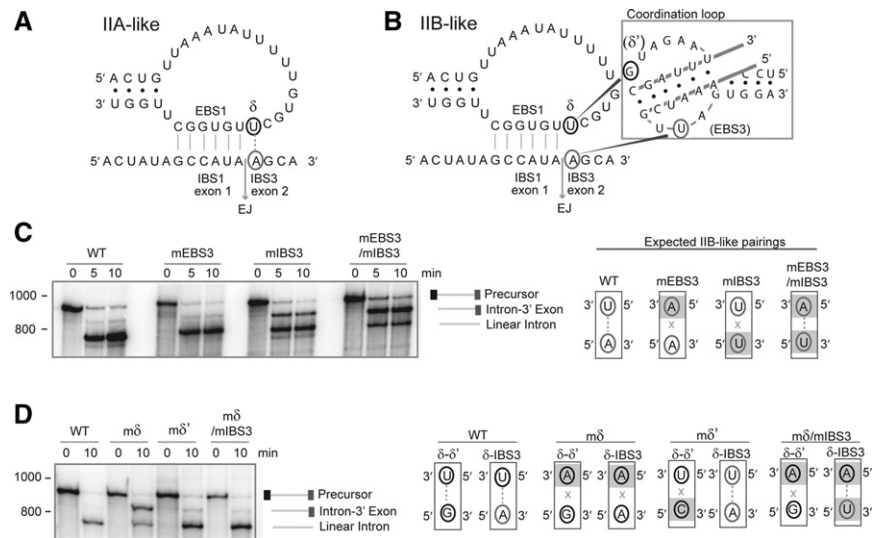


FIGURE 5. Mutagenesis and splicing assays of 3' exon recognition elements. (A) Summary of the mechanisms for 3' exon recognition for IIA introns, as applied to the *C.te.II* intron sequence. For IIA introns, the nucleotide directly upstream of EBS1 base-pairs with the first position of the 3' exon, and is called the δ – δ' pairing, although we call it the δ –IBS3 pairing here to avoid confusion. (B) For IIB introns, the nucleotide directly upstream of EBS1 base-pairs with a position in the "coordination loop" of domain I (δ – δ'), while a nearby residue pairs with the first nucleotide of the 3' exon (EBS3–IBS3). (C) Self-splicing assays of EBS3–IBS3 mutants, showing that mutation of EBS3 does not affect splicing, while mutation of IBS3 partially disrupts the second step of splicing (greater amount of intron-3' exon intermediate). Mutant sequences shown to the right of the self-splicing gel indicate the pairings and disruptions predicted for the IIB mechanism of 3' exon recognition. Mutated nucleotides are indicated by gray shading. (D) Self-splicing assays of δ – δ' mutants, showing that mutation of δ' has no impact on self-splicing, while mutation of δ disrupts the second step of splicing. The compensatory mutations of δ and IBS3 together rescue the second step deficiency, supporting a IIA-like mechanism of 3' exon recognition. Shown adjacent to the splicing gel are the expected pairings or disruptions for each of the mutants, according to either the IIB-like (δ – δ') or IIA-like (δ –IBS3) mechanism. Mutated nucleotides are indicated by gray shading.

Thus, *C.te.I1* appears to use a IIA-like mechanism to recognize its 3' exon.

When one examines the sequences of the four 3' exon sequences spliced in vivo, it is evident that all four exons begin with the sequence AGCA (Fig. 6C). Furthermore, the GCA nucleotides (exon positions 2, 3, 4) are complementary to the nucleotides adjacent to δ in the EBS3 loop, raising the possibility that the δ -IBS3 pairing may extend up to four base pairs (Fig. 6B). To test this possibility, positions 2, 3, and 4 were mutated. Mutation of three bases in either strand led to a modest but reproducible increase of intron-3' exon intermediate, going from a level of ~10% of products in the WT construct to ~20% in both the mUGC and m3'exon constructs.

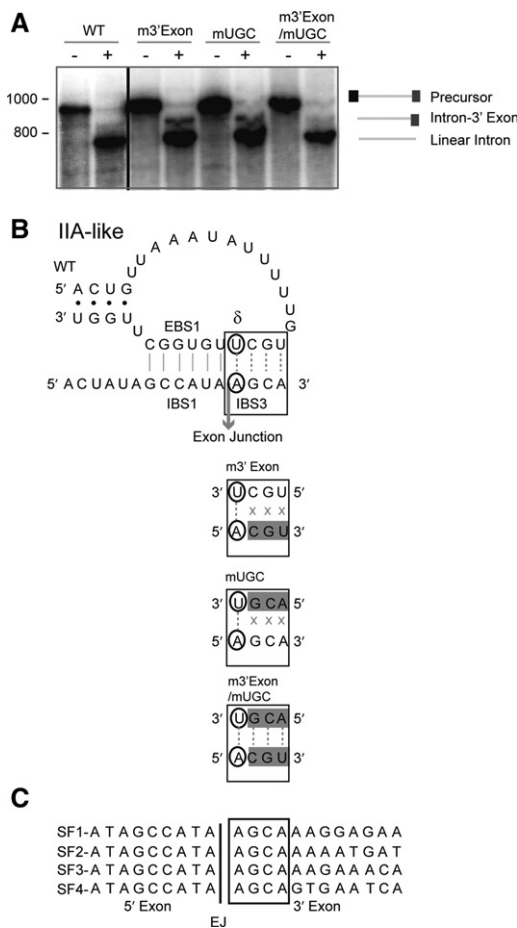


FIGURE 6. Three nucleotides adjacent to δ also affect 3' exon recognition and appear to form an extended δ -IBS3 pairing. (A) Self-splicing assays of the *C.te.I1* intron mutated for the three potential base pairs adjacent to δ -IBS3. Splicing intermediate accumulates when nucleotides are mutated on either strand; however, the compensatory mutations rescue the second step deficiency, consistent with the hypothesized pairing. The black line separates two exposures of the gel that show that intermediate is not evident for the WT construct. (B) Diagram of the potentially extended δ -IBS3 pairing and the mutations tested in A. In each of the mutations, the δ -IBS3 pairing is maintained. (C) Sequences of the exon junctions corresponding to each of the four splice forms observed in vivo. The first four nucleotides of the 3' exons are identical for all four alternatively splicing reactions and are indicated by a box.

This increase in accumulation was reversed by the compensatory mutation (Fig. 6A; Supplemental Table 7). The data indicate that while the single base pair between δ and IBS3 is the main component of 3' exon recognition, there are additional contributions by adjacent base pairs.

DISCUSSION

C.te.I1 is a fascinating example of a mobile DNA that has become domesticated by a bacterium to play a role in gene expression. The intron's alternative splicing reaction has overall similarity to alternative splicing in eukaryotic genomes, because it produces multiple mRNAs encoding five protein isoforms (McNeil et al. 2014). However, the mechanistic details differ and the alternative splicing mechanisms evolved independently.

Evolution of *C.te.I1* from a group II intron retroelement

The present study focuses on RNA structural changes that occurred during the genesis of *C.te.I1* and how these changes affect the self-splicing reaction. It is of interest as well to consider these changes in the larger context of the intron's conversion from a mobile intron to an alternatively splicing intron. The retromobile ancestor of *C.te.I1* can be concluded to be a IIB intron of the Class B type and probably from the β sublineage (Stabell et al. 2009). Interestingly, two previous examples of Class B introns were reported to utilize shifted splice sites, which may help explain why a Class B intron was capable of evolving the shifted splice site used by *C.te.I1*. The intron *B.a.I2* from *B. anthracis* splices at either position +1 or +4 in the 3' exon, which is specified by two sets of potential γ - γ' and IBS3-EBS3 pairings (Robart et al. 2004). In a more dramatic example, eight Class B introns splice at distant 3' splice sites due to an extension called domain 7, which causes splicing to occur as far away as 72 nt downstream from D6 (Stabell et al. 2007, 2009). With *C.te.I1* being a third example of a Class B intron using a shifted splice site, it is possible that splice-site selection is more flexible for Class B than for other classes.

A second major change that occurred in the development of *C.te.I1* was the loss of the intron-encoded ORF. This would have caused both loss of retromobility and loss of a splicing factor for the intron. Because group II introns exist in bacteria as selfish DNAs, the loss of mobility would inevitably result in loss from the genome unless the intron provided a benefit to its host. In the case of *C.te.I1*, it is evident how alternative splicing of a surface layer protein gene might benefit the bacterium and hence the intron could be retained as a non-mobile genetic element. On the other hand, loss of the IEP's splicing function would require compensation either by a new host-encoded splicing factor or by increased efficiency of ribozyme splicing, such that no protein splicing factors are needed. Several bacterial group II introns are known

that do not encode IEPs and presumably lost them (Meng et al. 2005; Van der Auwera et al. 2005; Simon et al. 2008). In those cases, an IEP is encoded elsewhere in the genome that could act in *trans*. To our knowledge, *C.te.II* is the first example of an ORF-less intron in bacteria in a genome that does not encode an IEP or IEP-related protein. While it is unknown whether other splicing factors are encoded in the *Clostridium tetani* genome, the robust splicing activity of *C.te.II* in vitro raises the possibility that in vivo splicing might occur without protein splicing factors.

The third major modification marking the origin of *C.te.II* is the gain of three D5/6 sequences along with their associated downstream ORFs. One plausible mechanism for this is that each D5/6 copy corresponds to an independent insertion of a mobile intron, followed by deletion of all of the intron except D5/6 and the downstream exon (McNeil et al. 2014). Along these lines, it is possible that the 311-bp D4 loop sequence is not a remnant of an intron but is derived from other genomic sequences.

Adaptations of the *C.te.II* ribozyme for a role in alternative splicing

Of the RNA structural changes that occurred during the formation of *C.te.II*, most are in the EBS1 region of domain 1, a region known to be involved in recognition of the 5' and 3' splice sites, and as well in D6, which is involved in the branching reaction. The loop of EBS1 became enlarged from 8 to 9 nt (typical of the β lineage) to 24 nt, with the insertion occurring directly at the 5' side of EBS1. The internal bulge motif in the EBS1 stem was apparently lost, which is inferred from the fact that it is absent from *C.te.II* but is found in 36/36 introns of the α sublineage of Class B, and 11/13 introns in the β sublineage (see Fig. 1B, inset; <http://webapps2.ucalgary.ca/~groupii/index.html#>; BA McNeil, L Wu, AM Jarding, S Zimmerly, unpubl.). *C.te.II* appears to have lost the EBS2 motif and IBS2–EBS2 interaction as well, and two stem–loop structures adjacent to EBS2. Finally, the D6 motif appears to have lost the branch A motif, and the terminal η' tetraloop along with the η – η' interaction.

Functionally, the most significant adaptation of the *C.te.II* ribozyme is the shifted 5' splice site, because it allows elimination of the stop codon of the upstream ORF and the correct fusion of reading frames for all four splicing reactions. This study shows that while IBS1–EBS1 is the main determinant for the shifted splice site, additional ribozyme elements contribute. Namely, the 8-nt spacer at the 5' end of the intron is essential to accommodate the shifted splicing, because its deletion abolishes splicing (Fig. 4C,D). To a lesser degree, there is a detectable but modest influence on 5' splice site usage by the sequence of the spacer (B–B'/m'), the presence of the enlarged loop containing EBS1 (Δ EBS1), and the loop's sequence (mEBS1loop). Another potential adaptation that may promote the shifted splice site is the loss of the IBS2–EBS2 interaction. Although not tested experimentally here,

it has been noted that other group II introns that use a shifted 5' splice site have also lost their IBS2–EBS2 motifs (Li et al. 2011; see below).

Splicing by the hydrolytic pathway is a functional change from the retromobile ancestor and can be attributed to the alterations in D6 structure, as the disrupted branch site makes the branching pathway unavailable. For other introns, mutation of the branch point structure has been shown to lead to use of the hydrolysis pathway, as the ribose of the bulged A normally provides the nucleophile for the first step of splicing (Van der Veen et al. 1987; Jacquier and Jacquesson-Breuleux 1991; Daniels et al. 1996). D6 of *C.te.II* also lacks the GNRA tetraloop that usually caps the stem and participates in η – η' , an interaction that helps guide the transition between through the two steps of splicing (Chanfreau and Jacquier 1996). Thus, the structural anomalies of D6 readily rationalize the use of the hydrolytic splicing pathway by *C.te.II*. On the other hand, it is not immediately clear whether the hydrolytic pathway is beneficial for the intron's role in alternative splicing or whether the branching reaction was simply lost because it was no longer needed for the mobility reaction.

A third ribozyme-based modification is the mechanism of 3' exon recognition, which changed from the IIB mechanism of the Class B ancestor to a IIA-like mechanism. Curiously, there are no clues in the secondary structure for loss of the IIB mechanism, because candidate EBS3 and δ' motifs can be found in the *C.te.II* secondary structure folding (Fig. 1). However, the mutagenesis experiments clearly support the IIA-like mechanism and not the IIB mechanism (Fig. 5). The reasons why *C.te.II* evolved a IIA-like mechanism rather than a IIB mechanism are not clear. One possibility is that it allows for an expanded pairing with the 3' exon beyond the single nucleotide pairing of the IIB mechanism, and perhaps up to 4 bp. An extended base pairing with the 3' exon may be beneficial in providing better recognition of the distantly located alternative 3' exon sequences.

Similarities in the evolution of *C.te.II* and a subset of mitochondrial rRNA group II introns

Apart from the interest in how a mobile group II intron evolved to be an alternatively splicing intron, the changes in *C.te.II* RNA structure and ribozyme properties are intriguing because a very similar set of changes has been observed in mitochondrial rRNA group II introns (Li et al. 2011). The similarities suggest that this set of ribozyme adaptations exemplify an inherent dynamic for the evolution of group II introns.

In the previous study, a set of rRNA introns were identified that utilize 5' splice sites located 1–33 nt upstream of the 5' GUGYG motif, with correspondingly altered IBS1–EBS1 pairings (Li et al. 2011). The secondary structures lack an EBS2 motif and IBS2–EBS2 pairing, and have a shortened D6 stem that lacks the branch A motif. All of these introns lost the RT IEP reading frame in D4, although four of 10

had acquired homing endonuclease ORFs of the LAGLIDADG family. One intron was tested for self-splicing and, like *C.te.II*, spliced exclusively through the hydrolysis pathway (Li et al. 2011).

Why should *C.te.II* and these mitochondrial rRNA introns have acquired the same set of structural and splicing anomalies? One likely factor noted for the rRNA introns is loss of retromobility (Li et al. 2011). Most of the rRNA introns as well as *C.te.II* lost mobility completely, although the four introns with LAGLIDADG ORFs may be mobile through the DNA-based, homing endonuclease pathway that has been studied for group I introns (Lambowitz and Belfort 1993; Chevalier and Stoddard 2001). Due to the fact that retromobility no longer occurs for the introns, the ribozyme properties that are mobility related might be lost, most particularly the ability of the intron to reverse splice into the DNA target site. Because the lariat topology is required for the two transesterification reactions of reverse splicing (Zimmerly et al. 1995; Belfort et al. 2002; Lambowitz and Zimmerly 2004), the branching reaction cannot be lost for mobile introns but may be lost for immobile introns. Similarly, it is possible that mobile introns require IBS1 to be located directly adjacent to the GUGYG motif in order to effectively reverse splice into DNA. Obviously, the juxtaposition is not needed for the forward splicing reaction (e.g., this study). A fascinating speculation has been made for the mitochondrial rRNA introns that the structural anomalies are causally linked (Li et al. 2011). It was suggested that the loss of the EBS2 motif correlates with the loss of the branching pathway, and that loss of the IBS2–EBS2 pairing preceded and might even be necessary for the loss of the branching reaction (Li et al. 2011). The co-occurrence in *C.te.II* of the same set of structural and splicing irregularities strengthens the case for a functional and causal interrelation among this set of modifications.

In closing, *C.te.II* provides an interesting example of a derived group II intron in which the ribozyme structure has differentiated along with its activities, leading to altered mechanisms of both 5' and 3' splice-site recognition and the emergence of a new biological role. The properties of the intron illustrate the capacity of group II intron ribozymes to evolve functional variants that can be used for gene expression in bacteria. It seems likely that additional derived forms of group II introns are undiscovered in bacterial genomes, and that alternative splicing is an underappreciated mechanism of gene expression and regulation in bacteria.

MATERIALS AND METHODS

Clostridium tetani growth and gDNA extraction

Clostridium tetani strain ATCC10779 (Designation 43415—Harvard Strain) was obtained from the American Type Culture Collection (ATCC). *C. tetani* cultures were grown in Brain Heart Infusion Medium (Oxoid CM1135) at 37°C under anaerobic conditions using the GasPak EZ anaerobic container system (BD Biosciences).

Genomic DNA was prepared from a 1.5-mL culture, and pelleted cells were washed and resuspended in 0.45 mL of TE (10 mM Tris–HCl at pH 7.4, 1 mM EDTA). Lysis was by the addition of 0.1 mg proteinase K and SDS to a final concentration of 1% with incubation at 37°C for 45 min. The sample was repeatedly extracted with an equal volume of phenol–CIA (25:24:1 of phenol:chloroform:isoamyl alcohol) until there was a clear interface upon centrifugation, followed by ethanol precipitation in the presence of 0.3 M NaOAc (pH 5.2). The pellet was washed with cold 75% ethanol, air-dried, and dissolved in TE (10 mM Tris–HCl at pH 7.4 and 1 mM EDTA).

Cloning and mutagenesis

The *C.te.II* intron was PCR amplified from *C. tetani* genomic DNA in two separate pieces using the primers S-5SSC-Bam, AS-5SSC-Eco, S-3SSC-Eco, and AS-3SSC-Cla (see Supplemental Material for a listing of all oligonucleotide sequences). The two pieces were ligated by recombinant PCR and cloned into the BamHI and ClaI sites of pBluescript KS+ (Stratagene). The resulting plasmid (pWT-SSC) contains 91 nt of the 5' exon (CTC00465) and 80 nt of the 3' exon (CTC00467). A total of 311 bp of sequence in domain 4 were deleted and replaced by a 26-nt loop. Mutant constructs were created using site-directed mutagenesis performed according to Byrappa et al. (1995) using 5' phosphorylated oligos and Pfu DNA polymerase. The Stem 3 mutants are an exception to this and were created through recombinant PCR. The resultant PCR products were gel extracted and ligated with T4 polynucleotide kinase (Invitrogen). Oligos used in mutagenesis are provided in Supplemental Table 8. All plasmid constructs were confirmed by sequencing (a list of constructs is provided in Supplemental Table 9).

RT-PCR

cDNA synthesis was performed in 20 μ L using 10 pmol primer AS-3SSC-Cla and 200 units Superscript II Reverse Transcriptase (Invitrogen) with 1 μ g of RNA as template, according to manufacturer's protocol. No RT controls in which no enzyme was added were also performed. PCR reactions were performed with 2 μ L RT- reaction as template, AS-3SSC-Cla, and S-5SSC-Bam, and Pfu DNA polymerase in 10 mM Tris–HCl (pH 8.8), 2.5 mM MgCl₂, 50 mM KCl, and 0.1% Triton X-100.

In vitro transcription and self-splicing

Transcription reactions were performed at 37°C for 30 min in a volume of 20 μ L of 40 mM Tris–HCl (pH 8.0), 4 mM MgCl₂, 50 mM NaCl, 1 mM each NTP, 5 mM DTT, and 0.05% Triton X-100, and with 500 ng plasmid template (linearized with XhoI) and 2 μ L T7 RNA polymerase. The reaction was extracted with phenol–chloroform isoamyl alcohol (25:24:1) and precipitated with ethanol and 2.5 M NH₄OAc. For radiolabeled transcripts, reactions contained 1 μ L of [α -³²P]UTP (10 mCi/mL; 3000 Ci/mmol; MP Biomedicals). For self-splicing reactions, ³²P-labeled (100,000 cpm) or cold transcript (200 ng) was resuspended in TE and folded by the following incubations: 90°C for 1 min, 75°C for 5 min, and slow cooling to 45°C over 15 min. Self-splicing buffer was added to produce a volume of 50 μ L containing 100 mM MgCl₂, 0.5 M NH₄Cl, and 40 mM Tris–HCl (pH 7.5). Reactions were incubated at 53°C for

5 min, followed by ethanol precipitation and resolution on a 4% polyacrylamide (19:1 acrylamide:bisacrylamide ratio)/8 M urea gel. For the unspliced control, reactions were done in parallel with the omission of MgCl₂.

Gels were exposed overnight to phosphor plates and imaged using a Storm 860 PhosphorImager (Amersham). Quantification was done with the manufacturer's ImageQuant TL 2005 software and is provided in Supplemental Tables 1–7.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

This work was supported by the Canadian Institutes of Health Research (grant MOP-93662 to S.Z.) B.A.M. was supported by a PGS-D studentship from the Natural Sciences and Engineering Research Council (Canada).

Received October 1, 2013; accepted March 3, 2014.

REFERENCES

- Belfort M, Derbyshire V, Parker MM, Cousineau B, Lambowitz AM. 2002. Mobile introns: pathways and proteins. In *Mobile DNA II* (ed. Craigie R, et al.), pp. 761–783. ASM Press, Washington, DC.
- Byrappa S, Gavin DK, Gupta KC. 1995. A highly efficient procedure for site-specific mutagenesis of full-length plasmids using *Vent* DNA polymerase. *Genome Res* 5: 404–407.
- Candales MA, Duong A, Hood KS, Li T, Neufeld RAE, Sun R, McNeil BA, Wu L, Jarding AM, Zimmerly S. 2012. Database for bacterial group II introns. *Nucleic Acids Res* 40: D187–D190.
- Chanfreau G, Jacquier A. 1994. Catalytic site components common to both splicing steps of a group II intron. *Science* 266: 1383–1387.
- Chanfreau G, Jacquier A. 1996. An RNA conformational change between the two chemical steps of group II self-splicing. *EMBO J* 15: 3466–3476.
- Chevalier BS, Stoddard BL. 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res* 29: 3757–3774.
- Costa M, Michel F, Westhof E. 2000. A three-dimensional perspective on exon binding by a group II self-splicing intron. *EMBO J* 19: 5007–5018.
- Dai L, Zimmerly S. 2003. ORF-less and reverse-transcriptase-encoding group II introns in archaeobacteria, with a pattern of homing into related group II intron ORFs. *RNA* 9: 14–19.
- Daniels DL, Michels WJ Jr, Pyle AM. 1996. Two competing pathways for self-splicing by group II introns: a quantitative analysis of *in vitro* reaction rates and products. *J Mol Biol* 256: 31–49.
- Ferat J-L, Michel F. 1993. Group II self-splicing introns in bacteria. *Nature* 364: 358–361.
- Fontaine JM, Goux D, Kloareg B, Loiseaux-de Goër S. 1997. The reverse-transcriptase-like proteins encoded by group II introns in the mitochondrial genome of the brown alga *Pylaiella littoralis* belong to two different lineages which apparently coevolved with the group II ribozyme lineages. *J Mol Evol* 44: 33–42.
- Jacquier A, Jacquesson-Breuleux N. 1991. Splice site selection and role of the lariat in a group II intron. *J Mol Biol* 219: 415–428.
- Jacquier A, Michel F. 1987. Multiple exon-binding sites in class II self-splicing introns. *Cell* 50: 17–29.
- Jacquier A, Michel F. 1990. Base-pairing interactions involving the 5' and 3'-terminal nucleotides of group II self-splicing introns. *J Mol Biol* 213: 437–447.
- Jarrell K, Peebles C, Dietrich R, Romiti S, Perlman P. 1988a. Group II intron self-splicing. Alternative reaction conditions yield novel products. *J Biol Chem* 263: 3432–3439.
- Jarrell KA, Dietrich R, Perlman P. 1988b. Group II intron domain 5 facilitates a *trans*-splicing reaction. *Mol Cell Biol* 8: 2361–2366.
- Koch J, Boulanger S, Dib-Hajj S, Hebbar S, Perlman P. 1992. Group II introns deleted for multiple substructures retain self-splicing activity. *Mol Cell Biol* 12: 1950–1958.
- Lambowitz AM, Belfort M. 1993. Introns as mobile genetic elements. *Annu Rev Biochem* 62: 587–622.
- Lambowitz AM, Zimmerly S. 2004. Mobile group II introns. *Annu Rev Genet* 38: 1–35.
- Li CF, Costa M, Bassi G, Lai YK, Michel F. 2011. Recurrent insertion of 5'-terminal nucleotides and loss of the branchpoint motif in lineages of group II introns inserted in mitochondrial preribosomal RNAs. *RNA* 17: 1321–1335.
- Li-Pook-Than J, Bonen L. 2006. Multiple physical forms of excised group II intron RNAs in wheat mitochondria. *Nucleic Acids Res* 34: 2782–2790.
- McNeil BA, Simon DM, Zimmerly S. 2014. Alternative splicing of a group II intron in a surface layer protein gene in *Clostridium tetani*. *Nucleic Acids Res* 42: 1959–1969.
- Meng Q, Wang Y, Liu X-Q. 2005. An intron-encoded protein assists RNA splicing of multiple similar introns of different bacterial genes. *J Biol Chem* 280: 35085–35088.
- Michel F, Ferat J. 1995. Structure and activities of group II introns. *Annu Rev Biochem* 64: 435–461.
- Michel F, Jacquier A. 1987. Long-range intron-exon and intron-intron pairings involved in self-splicing of class II catalytic introns. *Cold Spring Harb Symp Quant Biol* 52: 201–212.
- Michel F, Umeson K, Ozeki H. 1989. Comparative and functional anatomy of group II catalytic introns—a review. *Gene* 82: 5–30.
- Michels WJ, Pyle AM. 1995. Conversion of a group II intron into a new multiple-turnover ribozyme that selectively cleaves oligonucleotides: elucidation of reaction mechanism and structure/function relationships. *Biochemistry* 34: 2965–2977.
- Murray HL, Mikheeva S, Coljee VW, Turczyk BM, Donahue WF, Bar-Shalom A, Jarrell KA. 2001. Excision of group II introns as circles. *Mol Cell* 8: 201–211.
- Peebles CL, Perlman PS, Mecklenburg KL, Petrillo ML, Tabor JH, Jarrell KA, Cheng HL. 1986. A self-splicing RNA excises an intron lariat. *Cell* 44: 213–223.
- Robert AR, Montgomery NK, Smith KL, Zimmerly S. 2004. Principles of 3' splice site selection and alternative splicing for an unusual group II intron from *Bacillus anthracis*. *RNA* 10: 854–862.
- Robert AR, Seo W, Zimmerly S. 2007. Insertion of group II intron retroelements after intrinsic transcriptional terminators. *Proc Natl Acad Sci* 104: 6620–6625.
- Schmelzer C, Schweyen RJ. 1986. Self-splicing of group II introns *in vitro*: mapping of the branch point and mutational inhibition of lariat formation. *Cell* 46: 557–565.
- Simon DM, Clarke NAC, McNeil BA, Johnson I, Pantuso D, Dai L, Chai D, Zimmerly S. 2008. Group II introns in eubacteria and archaea: ORF-less introns and new varieties. *RNA* 14: 1704–1713.
- Simon DM, Kelchner SA, Zimmerly S. 2009. A broadscale phylogenetic analysis of group II intron RNAs and intron-encoded reverse transcriptases. *Mol Biol Evol* 26: 2795–2808.
- Stabell FB, Tourasse NJ, Ravnum S, Kolstø AB. 2007. Group II intron in *Bacillus cereus* has an unusual 3' extension and splices 56 nucleotides downstream of the predicted site. *Nucleic Acids Res* 35: 1612–1623.
- Stabell FB, Tourasse NJ, Kolstø AB. 2009. A conserved 3' extension in unusual group II introns is important for efficient second-step splicing. *Nucleic Acids Res* 37: 3202–3214.
- Toor N, Hausner G, Zimmerly S. 2001. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* 7: 1142–1152.
- Toor N, Robert AR, Christianson J, Zimmerly S. 2006. Self-splicing of a group IIC intron: 5' exon recognition and alternative 5' splicing

- events implicate the stem-loop motif of a transcriptional terminator. *Nucleic Acids Res* **34**: 6461–6471.
- Toor N, Keating KS, Taylor SD, Pyle AM. 2008. Crystal structure of a self-spliced group II intron. *Science* **320**: 77–82.
- Toro N, Martínez-Abarca F. 2013. Comprehensive phylogenetic analysis of bacterial group II intron-encoded ORFs lacking the DNA endonuclease domain reveals new varieties. *PLoS One* **8**: e55102.
- Toro N, Molina-Sánchez MD, Fernández-López M. 2002. Identification and characterization of bacterial class E group II introns. *Gene* **299**: 245–250.
- Van der Auwera GA, Andrup L, Mahillon J. 2005. Conjugative plasmid pAW63 brings new insights into the genesis of the *Bacillus anthracis* virulence plasmid pXO2 and of the *Bacillus thuringiensis* plasmid pBT9727. *BMC Genomics* **6**: 103.
- Van der Veen R, Arnberg AC, van der Horst G, Bonen L, Tabak HF, Grivell LA. 1986. Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing *in vitro*. *Cell* **44**: 225–234.
- Van der Veen R, Kwakman J, Grivell L. 1987. Mutations at the lariat acceptor site allow self-splicing of a group II intron without lariat formation. *EMBO J* **6**: 3827–3831.
- Vogel J, Börner T. 2002. Lariat formation and a hydrolytic pathway in plant chloroplast group II intron splicing. *EMBO J* **21**: 3794–3803.
- Zimmerly S, Guo H, Perlman PS, Lambowitz AM. 1995. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* **82**: 545–554.
- Zimmerly S, Hausner G, Wu X. 2001. Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Res* **29**: 1238–1250.