Molecular characterization of both transesterification reactions of the group II intron circularization pathway

Félix LaRoche-Johnston[†], Caroline Monat[†], Erika Verreault and Benoit Cousineau [ID]*

Department of Microbiology and Immunology, McGill University, Montréal, QC H3A 2B4, Canada

ABSTRACT

Group II introns can self-splice from RNA transcripts through branching, hydrolysis and circularization, being released as lariats, linear introns and circles, respectively. In contrast to branching, the circularization pathway is mostly based on assumptions and has been largely overlooked. Here, we address the molecular details of both transesterification reactions of the group II intron circularization pathway *in vivo*. We show that free E1 is recruited by the intron through base pairing interactions and that released intron circles can generate free E1 by the spliced exon reopening reaction. The first transesterification reaction was found to be induced inaccurately by the 3′OH of the terminal residue of free E1 at the 3′ splice site, producing circularization intermediates with heterogeneous 3′ ends. Nevertheless, specific terminal 3′OH, selected by a molecular ruler, was shown to precisely attack the 5′ splice site and release intron circles with 3′–5′ rather than 2′–5′ bonds at their circularization junction. Our work supports a circularization model where the recruitment of free E1 and/or displacement of *cis*-E1 induce a conformational change of the intron active site from the pre-5′ to the pre-3′ splice site processing conformation, suggesting how circularization might initiate at the 3′ instead of the 5′ splice site.

INTRODUCTION

Group II introns are large ribozymes as well as versatile mobile retroelements that interrupt genetic loci in bacteria, some archaea and the organelles of certain plants, fungi and lower eukaryotes (1). They are absent from the nuclear genome of eukaryotes (2), but are nevertheless considered as the progenitors of important eukaryotic genetic elements such as nuclear introns and non-LTR retrotransposons (3,4). Altogether, group II intron-derived genetic elements are believed to encompass over half of the human genome (3,4). Despite lacking primary sequence con-

servation, they have a highly conserved secondary structure consisting of six domains (DI–DVI) rooted within a central core (Figure 1A) (5,6). During transcription, group II introns fold sequentially from DI to DVI as the intron-interrupted transcripts emerge from the RNA polymerase. DI is the largest and the first intron domain to be transcribed, serving as a scaffold for the assembly of the other five folded domains (DII–DVI) into a catalytically active 3D structure. In bacteria, group II introns often harbour a large multifunctional intron-encoded protein (IEP) within DIV required for both splicing and mobility (Figure 1A, IEP). IEPs bind their cognate introns within unspliced RNA transcripts and serve as essential maturases, assisting the introns to adopt their active 3D structures and allow self-splicing *in vivo*. Following their autocatalytic excision, released group II introns remain associated with their IEPs in the form of ribonucleoprotein (RNP) particles. Group II intron RNPs are active retromobile elements that can invade unoccupied genetic loci using a diverse array of RNA-based mobility pathways (7–9).

When group II intron self-splicing was first characterized, it was found to occur via the exact same mechanism as nuclear introns: the branching pathway (10,11). During branching, group II introns self-splice using two consecutive transesterification reactions at the 5′ and 3′ splice sites, resulting in their release as lariat molecules (Figure 1B) (12). The 5′ splice site of group IIA introns is identified through base pairing interactions between two loop regions of the intron in DI called EBS1 and EBS2, and a stretch of 11 nucleotides at the 3′ end of its upstream exon (E1) called IBS1 and IBS2 (Figure 1A). The 2′OH of a conserved bulged adenosine residue in DVI, also called branch point, serves as the nucleophile of the first transesterification reaction attacking the phosphodiester bond at the 5′ splice site to generate a lasso-like lariat (Figure 1B, step 1). Released E1 remains associated with the intron RNA through the EBS/IBS base pair interactions (Figure 1B, step 1, vertical lines). In the second transesterification reaction, the 3′OH of released E1 acts as the attacking nucleophile, targeting the phosphodiester bond between the first nucleotide of E2 and the last nucleotide of the intron at the 3′ splice site to release an intron lariat and ligated exons (Figure 1B,

---
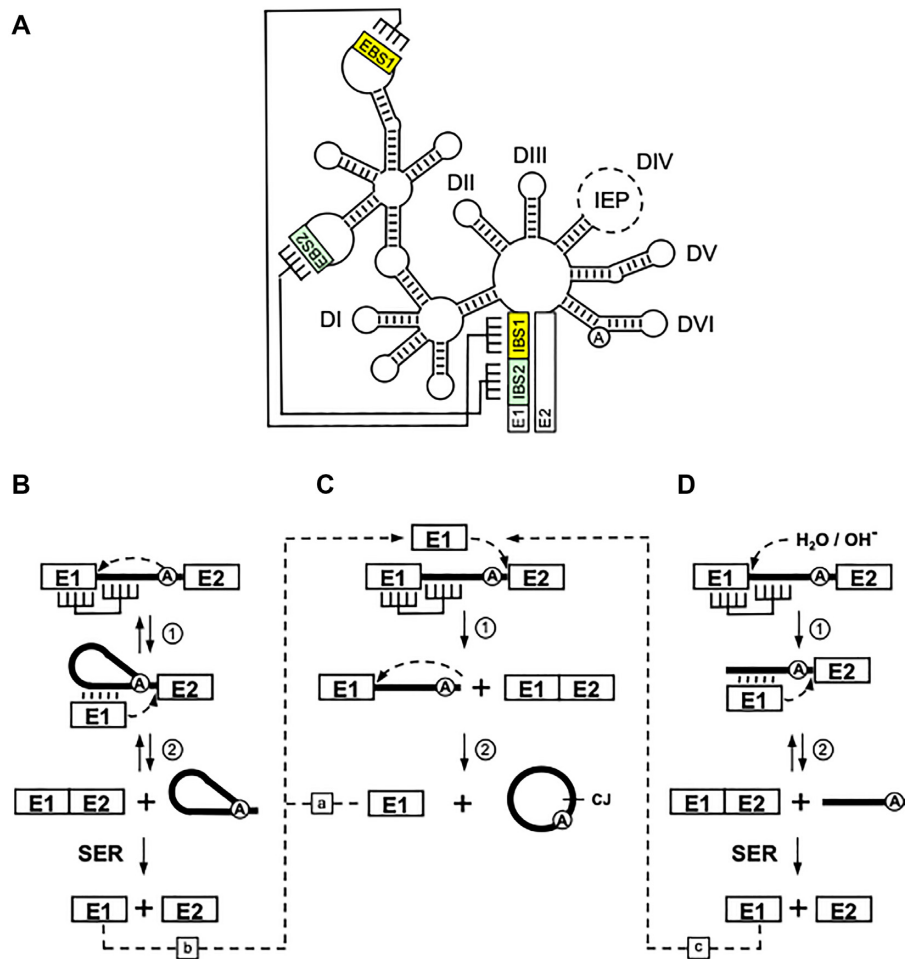
Figure 1. Group II intron secondary structure and splicing pathways. Schematics depicting the secondary structure of group II introns and the currently accepted branching, circularization and hydrolysis splicing pathways are displayed. (**A**) The consensus secondary structure of group II introns consists of six domains (DI–DVI) radiating from a central wheel. The flanking exons (E1 and E2) are shown as boxes on either side of the intron, while the bulged A nucleotide in DVI is represented as a circled A. The base pairing interactions between the intron (exon binding sites 1 and 2, EBS1 and EBS2) and the last nucleotides of E1 (intron binding sites 1 and 2, IBS1 and IBS2) involved in the recognition of the 5′ splice site during splicing and reverse splicing are also depicted (EBS1–IBS1, yellow; EBS2–IBS2, green). Most group II introns harbour an open reading frame also called IEP in DIV. (**B**) *Branching*: The 2′OH residue of the branch point A nucleotide (circled A) initiates the first nucleophilic attack at the 5′ splice site (step 1, dashed arrow). This transesterification reaction connects the 5′ end of the intron to the branch point A residue creating a 2′–5′ link and releases E1 that remains associated with the intron through base pairing interactions (EBS1/2–IBS1/2, vertical lines). The liberated 3′OH at the 3′ end of E1 then initiates a second nucleophilic attack at the 3′ splice site (step 2, dashed arrow), ligating the two exons (E1–E2) and releasing the intron as a branched lariat. (**C**) *Circularization*: The first nucleophilic attack takes place at the 3′ splice site and is initiated by the 3′OH of the last residue of free E1 (step 1, dashed arrow). This transesterification reaction generates ligated exons (E1–E2) and a circularization intermediate where the 5′ end of the linear intron is still attached to E1. Next, the 2′OH of the last intron residue initiates the second nucleophilic reaction at the 5′ splice site (step 2, dashed arrow) resulting in the release of intron circles and free E1. The position of the circularization junction (CJ) of released intron circles is depicted by a black bar. (**D**) *Hydrolysis*: A hydroxyl residue or water molecule initiates the first reaction at the 5′ splice site (step 1, dashed arrow). The second nucleophilic attack at the 3′ splice site is initiated by the liberated 3′OH of the last residue of E1 (step 2, dashed arrow), which ligates the two exons (E1–E2) and releases a linear intron. The circularization pathway is unique in that it requires free E1 while the first transesterification reaction is initiated *in trans* and occurs at the 3′ instead of the 5′ splice site. Even though the second circularization step generates free E1 (**C**, step 2) that can be recruited by the intron in a pre-mRNA to initiate another round of circularization (a), it is unclear how free E1 is initially produced. Other potential sources of free E1 are the spliced exon reopening (SER) reaction catalysed by released lariats (**B**, SER, b) and linear introns (**D**, SER, c).

step 2). Both steps of the group II intron branching pathway occur within a unique two-metal-ion catalytic centre (13,14), implying an important conformational change between the two transesterification reactions (15,16). Since the same number of phosphate bonds are generated and broken during branching, the pathway is energy neutral and readily reversible (Figure 1B, steps 1 and 2, double arrows), forming the basis for group II intron reverse splicing and mobility (17). Released intron lariats can also hydrolyse ligated

exons precisely at the junction between E1 and E2 through the SER reaction (Figure 1B, SER) (18–20).

Following the thorough characterization of branching, group II introns were also shown to self-splice via various other pathways, initially through *in vitro* observations and later supported by *in vivo* data (18,20). For example, in the hydrolysis pathway (Figure 1D), a water or hydroxyl ion is used as the attacking nucleophile, rather than the 2′OH of the branch point adenosine, during the first reaction to re-

lease E1 at the 5′ splice site (Figure 1D, step 1) (21). As in branching, the second step involves the attack of the 3′ splice site by the 3′OH of released E1, resulting in ligated exons and a linear intron (Figure 1D, step 2). Much more enigmatic is the group II intron circularization pathway, which generates full-length head-to-tail intron circles (Figure 1C). Although nuclear and group II introns both splice using the branching pathway, the circularization pathway is exclusive to group II introns. Circularization was proposed to initiate when the 3′OH of an external free E1 attacks the phosphodiester bond at the 3′ instead of the 5′ splice site (Figure 1C, step 1) (19). The first step would thus release ligated exons and a splicing intermediate consisting of upstream E1 still covalently linked to the linear intron. Next, the 2′OH of the intron's last nucleotide is believed to attack the phosphodiester bond at the 5′ splice site, releasing a 2′–5′ linked head-to-tail intron circle and free E1 that could initiate further instances of circularization (Figure 1C, step 2) (19). In contrast to branching, circularization is not considered to be a reversible splicing pathway (Figure 1C, steps 1 and 2, single arrows) and intron circles were never shown nor proposed to process ligated exons by the SER reaction.

Despite recent progress in our understanding of group II intron circularization (22–29), most key mechanistic aspects are still based on assumptions and remain to be addressed experimentally. Although free E1 was initially proposed to provide the external nucleophile to initiate circularization *in trans* at the 3′ splice site, it was never studied *in vivo*. It is also unclear in what way E1 would be recruited by the intron-interrupted transcript and how the intron would transition between binding its upstream E1 *in cis* and potentially binding free E1 *in trans*. Moreover, the origin of free E1 remains elusive, especially since introns mutated to lack their branch point splice primarily through circularization, yet intron circles are not believed to perform the SER reaction. In addition, it is unknown why recognition of the 3′ splice site is inaccurate during the first step, and how the 5′ splice site is recognized in the second step of circularization. Finally, the means through which the intron can initiate self-splicing at the 3′ instead of the 5′ splice site remains an essential question to be addressed.

Here, we provide a full and detailed characterization of both steps of the group II intron circularization pathway at the molecular level, elucidating crucial mechanistic aspects of intron circle formation *in vivo*. Using Ll.LtrB, a model group IIA intron from the Gram-positive bacterium *Lactococcus lactis*, we present experimental evidence that circularization is initiated at the 3′ splice site *in trans* by free E1. Our data demonstrate that free E1 is recruited by the intron within the pre-mRNA through base pairing interactions and that the nucleophile of the first transesterification reaction is the 3′OH of the last residue of recruited free E1. We also show that released intron circles can generate free E1 using the SER reaction. Our work further unveils that although multiple E1–intron circularization intermediates are generated after the first transesterification reaction, the 3′OH used to precisely attack the 5′ splice site during the second transesterification reaction is specifically selected by a molecular ruler. Finally, we demonstrate that the circu-

larization junction of released intron circles is composed of a 3′–5′ rather than a 2′–5′ linkage *in vivo*. Taken together, our data describe the nature of the molecular switches that trigger the fundamental structural rearrangements between the pre-3′ and pre-5′ splice site processing conformations, providing a mechanistic rationale that suggests how group II introns might initiate the circularization pathway at the 3′ instead of the 5′ splice site unlike branching and hydrolysis.

MATERIALS AND METHODS

Bacterial strains and plasmids

Lactococcus lactis strains NZ9800Δ*ltrB*(*recA*+)(Tet^R) (7) and MMS372(*recA*−) (9) were grown in M17 media supplemented with 0.5% glucose (GM17) at 30°C without shaking. The *Escherichia coli* strain DH10β, used for cloning purposes, was grown in LB at 37°C with shaking. Antibiotics were used at the following concentrations: chloramphenicol (Cam^R), 10 μg/ml; spectinomycin (Spc^R), 300 μg/ml; ampicillin (Amp^R), 100 μg/ml; rifampicin (Rif^R), 0.5 mg/ml; and erythromycin (Erm^R), 1 mg/ml.

The following plasmids were previously described: pDL-P$_{23}{}^2$-Ll.LtrB-WT (30), pDL-P$_{23}{}^2$-Ll.LtrB-ΔA (22), pDL-P$_{23}{}^2$-Ll.LtrB-AU-rich (28), pDL-P$_{23}{}^2$-Ll.LtrB-GC-rich (28) and pDL-P$_{23}{}^2$-Ll.LtrB-IBS1-EBS1 swap (28). The plasmid pDL-P$_{23}{}^2$-Ll.LtrB-ΔSD-ΔAUG was generated by PCR amplification using a ΔSD-ΔAUG primer and Ll.LtrB as the template. Other Ll.LtrB variants were constructed by site-directed mutagenesis (Q5$^®$ Site-Directed Mutagenesis Kit, New England Biolabs): pDL-P$_{23}{}^2$-E1(CAG)-Ll.LtrB-ΔA-E2(AAT), pLE-P$_{23}{}^2$-E1(GTC)-Ll.LtrB-ΔA-E2(TTA), pDL-P$_{23}{}^2$-Ll.LtrB-ΔA(+A), pDL-P$_{23}{}^2$-Ll.LtrB-ΔA(+AA), pDL-P$_{23}{}^2$-Ll.LtrB-ΔA(−A) and pDL-P$_{23}{}^2$-Ll.LtrB-ΔA-(AC). Plasmids and primers used are shown in Supplementary Tables S1 and S2, respectively.

RNA extraction, RT-PCR and 3′ RACE

Total RNA was isolated from NZ9800Δ*ltrB* and MMS372 harbouring various plasmid constructs as previously described (25,31). RT-PCR reactions (22–26,28,32) and 3′ RACE (23,28) were performed on total RNA preparations of NZ9800Δ*ltrB* and MMS372 harbouring various intron constructs (primers in Supplementary Table S2). Some total RNA preparations (10 μg) were treated with a mixture of RNase A and ShortCut RNase III for 2 h at 37°C prior to RT-PCR amplification (Figure 2).

Ex vivo trans-splicing assay

Overnight cultures of *L. lactis* were diluted (1:12.5) and grown in GM17 for 3 h at 30°C. Some cultures were treated prior to cell lysis with rifampicin for 10 min (0.5 mg/ml) or erythromycin for 30 min (1 mg/ml) to stop transcription and translation, respectively. An aliquot of 750 μl was centrifuged and the pellet was washed once with 10 mM Tris, pH 8.0. The bacteria were resuspended in 500 μl of lysis buffer [10 mM Tris, pH 8.0, lysozyme (1 mg/ml), mutanolysin (50 U/ml)] containing or not a specific amount
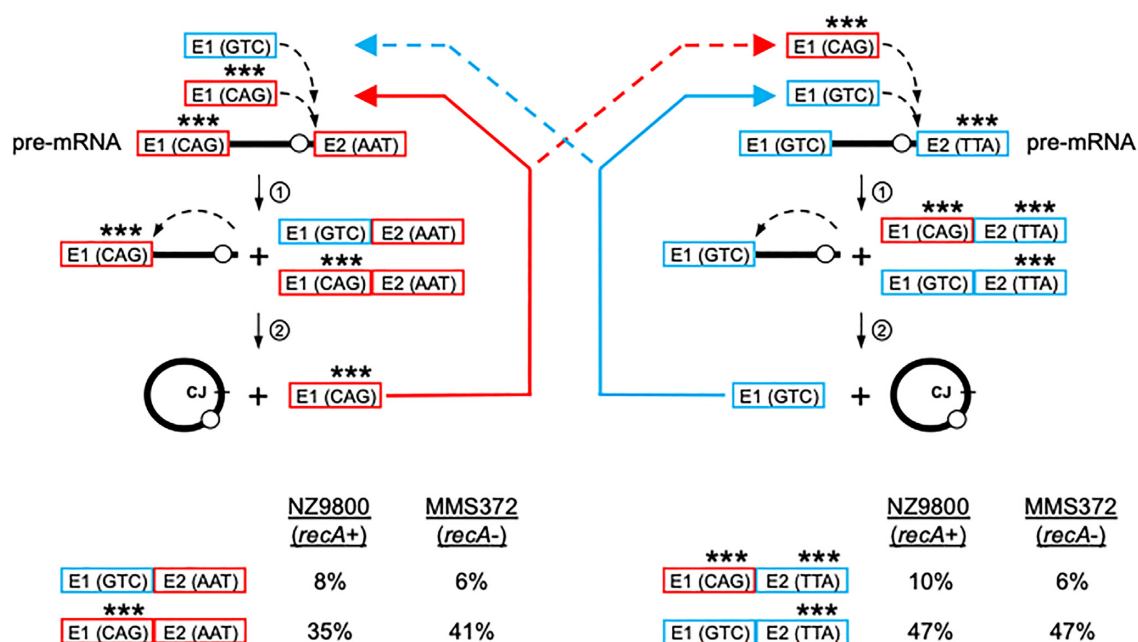
Figure 2. Free E1 initiates *in trans* the first transesterification reaction of the group II intron circularization pathway *in vivo*. Two versions of the *ltrB* gene, interrupted by identical copies of the Ll.LtrB branch point mutant (Ll.LtrB-ΔA), were expressed constitutively (P_{23} promoter) in *L. lactis* [NZ9800(*recA+*) or MMS372(*recA−*)] from two independent plasmids, pDL (red) and pLE (blue). The absence of the branch point A residue and the released intron circularization junction are illustrated by an empty circle and CJ, respectively. Splicing of Ll.LtrB-ΔA from both pre-mRNAs is represented by two schematics of the group II intron circularization pathway (red and blue). E1 was mutated from GTC to CAG (asterisks) upstream of the intron binding sequences (IBS1/2) at positions −24 to −22 in pDL-P_{23}^2-E1(CAG)-Ll.LtrB-ΔA-E2(AAT) (red pathway), while E2 was mutated from AAT to TTA (asterisks) at positions +22 to +24 in pLE-P_{23}^2-E1(GTC)-Ll.LtrB-ΔA-E2(TTA) (blue pathway). The use of cognate free E1 by each intron is represented by solid arrows, while the use of the counterpart free E1 by both introns is represented by dashed arrows. The pool of ligated exons (E1−E2) from NZ9800(*recA+*) and MMS372(*recA−*) was independently amplified by RT-PCR, cloned in pBS and 100 independent clones were sequenced. The relative abundance of the four combinations of ligated exons is shown as percentages.

of the *ex vivo* DNA oligo (1×: 3.2 μM; 2×: 6.4 μM; 4×: 12.8 μM) and incubated at 30°C for 3 h. Total RNA of the cell extract was isolated by the TRIzol method as previously described (31). The ligated exons were amplified by RT-PCR (primers in Supplementary Table S2). PCR reactions were analysed on 1% agarose gels and purified using a gel extraction kit (VWR). The PCR product was confirmed by DNA sequencing (Genome Québec).

Primer extension assay

Five micrograms of total RNA in a final volume of 8 μl of dH$_2$O was annealed with 0.5 pmol of a 5′ end-labelled primer complementary to the 5′-end of the intron along with 1 mM dNTPs in a final volume of 10 μl. After annealing of the primer (5 min, 65°C; 3 min on ice), 2 μl of 10× AMV RT buffer, 2 μl of 50 mM MgCl$_2$, 1 μl of RNase-Out (NEB) and 4 μl of dH$_2$O were added and incubated for 2 min at 42°C. The AMV RT (NEB, 5 units) was then added and the incubation continued for 2 h. The reactions were terminated by the addition of gel loading buffer (20 μl; 97.5% formamide, 10 mM EDTA, 0.3% bromophenol blue and xylene cyanol FF). The extension products were run on a 12% denaturing 8 M Urea PAGE gel, exposed on a phosphor screen and revealed with the Molecular Imager Fx (Bio-Rad). The molecular weight markers were generated by poisoned primer extension as previously described (23,28,33).

RESULTS

Free E1 is generated *in vivo* and recruited by group II introns to initiate the circularization splicing pathway

In the currently accepted model of group II intron circularization, freestanding E1, not covalently attached to the intron 5′ end, is presumably recruited by the intron within the pre-mRNA transcript to initiate the first step of the splicing pathway (Figure 1C, step 1) (19,22–28). This model implies that free E1 is somehow generated *in vivo* and readily available to be recruited by the unspliced intron. To address these assumptions, we developed an *in vivo* group II intron circularization assay using the Ll.LtrB branch point mutant, Ll.LtrB-ΔA. This Ll.LtrB construct, lacking the branch point A residue, was previously shown to be very useful to study group II intron circularization because it splices primarily through this pathway (22,23).

Two different plasmids were co-transformed in *L. lactis*, each harbouring identical copies of Ll.LtrB-ΔA that were constitutively expressed using P_{23} promoters (Figure 2, red and blue). Both introns were flanked by a unique combination of *ltrB* exons, which were wild type or harboured three consecutive point mutations in either E1 [GTC to CAG; positions −24 to −22 (red)] or E2 [AAT to TTA; positions +22 to +24 (blue)] (Figure 2, asterisks). The point mutations were engineered to identify the origin of E1 and E2 in the pool of ligated exons (E1–E2). Ligated exons were amplified by RT-PCR from total RNA extracts of

L. lactis co-expressing the E1(CAG)-Ll.LtrB-ΔA-E2(AAT) and E1(GTC)-Ll.LtrB-ΔA-E2(TTA) *ltrB* genes (Figure 2, red and blue) (23–26). The RT-PCR amplicon was purified, cloned in pBS and 100 independent clones were sequenced. We first observed the expected contiguous ligated exons for both pre-mRNA transcripts at proportions of 35% [E1(CAG)–E2(AAT), red–red] and 47% [E1(GTC)–E2(TTA), blue–blue]. We additionally detected chimeric ligated exons where E1 was precisely *trans*-spliced to E2 of its counterpart transcript at 8% [E1(GTC)–E2(AAT), blue–red] and 10% [E1(CAG)–E2(TTA), red–blue], respectively (Figure 2). Very similar results were obtained when the assay was carried out in an *L. lactis* strain deficient for homologous recombination (MMS372(*recA*−)). The small variations obtained between both strains [Figure 2, cf. MMS372(*recA*−) and NZ9800(*recA*+)] are not significant and most likely due to the nature of the assay performed. Moreover, the mixed amplicon of ligated exons was not observed when the total RNA preparations from both *L. lactis* strains were subjected to RNase treatment before the RT-PCR was performed. Taken together, our data confirm that the chimeric ligated exons amplified by RT-PCR are generated at the RNA level and rule out the possibility that they might originate through homologous DNA recombination.

These results demonstrate that free E1 from both interrupted pre-mRNA transcripts is available and independently recruited by the intron of its counterpart transcript to initiate the group II intron circularization pathway *in trans* and generate chimeric ligated exons. The lower proportion of chimeric ligated exons compared to the ligated exons coming from the same pre-mRNA transcript is most probably due to the segregation of the co-transformed plasmids into distinct foci within *L. lactis* cells. By extension, our data indicate that E1 is precisely released from the 5′ splice site of both copies of the Ll.LtrB-ΔA-interrupted pre-mRNAs *in vivo*. These results also strongly suggest that both introns use the same pathway to generate their contiguous ligated exons, recruiting and *trans*-splicing their respective cognate free E1 to their own downstream E2.

Unspliced group II introns recruit free E1 through specific base pairing interactions

To study how free E1 is recruited by the intron to initiate the first nucleophilic attack, we elaborated an *ex vivo* group II intron circularization assay (Figure 3A). The assay consisted of providing short DNA oligos of 54 nt as external sources of free E1 to *L. lactis* cell lysates that expressed the *ltrB* gene interrupted by different Ll.LtrB variants. Next, ligated exons were amplified by RT-PCR from total RNA extracts and the mixed amplicons were directly sequenced to determine whether the external source of free E1 could be found *trans*-spliced to E2 in the pool of ligated exons (23–26). Three consecutive point mutations (GTC to CAG; positions –24 to –22) (Figure 3A, asterisks) were engineered within the various free DNA E1 oligos [E1(CAG)] used to distinguish them from the internal source of RNA E1 [E1(GTC)] coming from the pre-mRNA. Sequence chromatograms of ligated exons spanning nucleotides –26 to –20 are shown (Figure 3A–E).

We first added various amounts of free E1(CAG) to a cell lysate that expressed wild-type Ll.LtrB (Ll.LtrB-WT) (Figure 3B). The external source of free E1(CAG) can be found in significant amounts along with cognate E1(GTC) precisely *trans*-spliced to E2 in the pool of ligated exons at all three oligo concentrations [E1(1×), E1(2×) and E1(4×)] when compared to the no oligo control (Figure 3B). The external source of free E1(CAG) reached ∼50% of total E1 in the population of ligated exons at the highest oligo concentration. This ratio was corroborated by cloning the RT-PCR amplicon of ligated exons at the highest oligo concentration and sequencing individual clones. The same phenotype was observed when the cells were treated with the RNA polymerase inhibitor rifampicin for 10 min prior to cell lysis (Figure 3C, Ll.LtrB-WT, rifampicin). These results show that active transcription is not required for both the recruitment of free E1 by the intron and the nucleophilic attack during the initial step of circularization (Figure 3A). In contrast, preventing translation of E1 within the pre-mRNA by using a construct missing the Shine–Dalgarno (ΔSD) sequence and the first AUG codon (ΔAUG) (Figure 3D, Ll.LtrB-ΔΔ) or by inhibiting translation with erythromycin for 30 min prior to cell lysis (Figure 3E, Ll.LtrB-WT, erythromycin) almost completely abrogated the use of free E1(CAG) by the intron. Only traces of external E1(CAG) were detected within the pool of ligated exons at the highest oligo concentration in both approaches [Figure 3D and E, E1(4×)].

We then hypothesized that free E1 is recruited by the intron through base pairing interactions to initiate the circularization pathway (Figure 4A). We thus used, in addition to Ll.LtrB-WT, an intron variant where the IBS1 and EBS1 sequences were swapped between E1 and the intron (24). Ll.LtrB-IBS1-EBS1 swap (Ll.LtrB-SW) was previously shown to splice accurately and efficiently, using both branching and circularization, and to generate significant amounts of lariats and circles, similarly to Ll.LtrB-WT (28). Using E1(CAG) oligos harbouring either wild-type IBS1 [E1-WT(4×)] or the complementary swapped sequence [E1-SW(4×)] at their 3′ end, we saw that only the complementary oligos could be specifically found in ligated exons for both introns [Figure 4B, E1-WT(4×) and E1-SW(4×)]. This recruitment of E1 was just as specific for the Ll.LtrB-ΔA intron, since only E1(CAG) harbouring the wild-type IBS1 sequence [E1-WT(4×)] was found in the pool of ligated exons, but not the E1(CAG) swap [E1-SW(4×)] that carries a non-complementary IBS1 sequence (Figure 4B, Ll.LtrB-ΔA). The importance of base pairing was corroborated by adding external sources of E1(CAG) harbouring either AU-rich or GC-rich IBS1/2 sequences at their 3′ end to cell lysates that expressed introns with complementary AU-rich (Ll.LtrB-AU-rich) or GC-rich (Ll.LtrB-GC-rich) EBS1/2 sequences, respectively (28). Despite having modified EBS sequences, both introns were able to exclusively recruit and *trans*-splice the complementary DNA oligo (Supplementary Figure S1).

The ability of Ll.LtrB to successfully recruit external DNA oligonucleotides lacking OH groups at the 2′ positions to generate ligated exons with perfect splice junctions indicates that the first transesterification reaction of the circularization pathway is initiated by the nucleophilic attack of the 3′OH of the terminal residue of E1. Accordingly, we
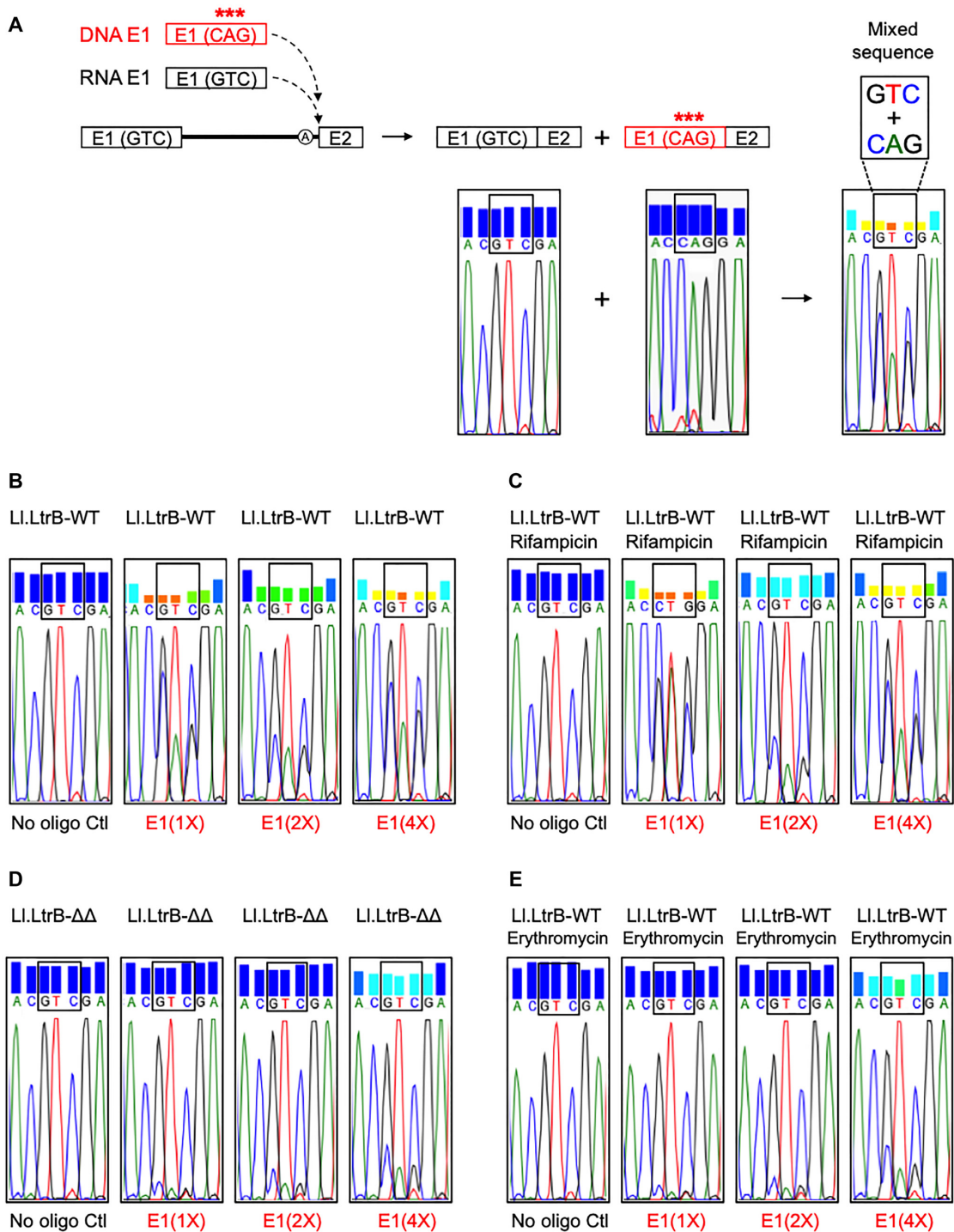
Figure 3. External sources of free E1 can be recruited to initiate the group II intron circularization pathway. (**A**) Schematic of the *ex vivo* group II intron circularization assay. The assay consists of adding various amounts [No oligo Ctl, E1(1×), E1(2×) or E1(4×)] of a short DNA oligo (DNA E1, red) as an external source of free E1 to *L. lactis* cell lysates expressing various constructs of the Ll.LtrB intron to assess whether the intron can recruit free E1 to initiate circularization and generate ligated exons. The internal source of E1 coming from the pre-mRNA contains the wild-type GTC sequence at positions −24 to −22 [RNA E1: E1(GTC)], upstream of the intron binding sequences (IBS1/2), whereas the external source of E1 (red) contains CAG (asterisks) [DNA E1: E1(CAG)]. The schematic of the assay shows that the intron can use *in trans* either the internal or external source of E1 (dashed curved arrows) to produce two types of ligated exons: E1(GTC)–E2 or E1(CAG)–E2. A portion of the sequence chromatogram of the purified PCR amplicons spanning nucleotides −26 to −20 from the end of E1 is shown where nucleotides −24 to −22 are boxed. Nucleotides GTC or CAG are present at positions −24 to −22, depending on whether the internal or external source of E1 is found ligated to E2, respectively. A mixed sequence at positions −24 to −22 is thus expected in the chromatogram if both sources of ligated exons are generated. The colour code above each base is a visual representation of the Phred score calculated for base calling from 0 to 60 (red to blue). Ligated exons were amplified by RT-PCR from total RNA extracts containing Ll.LtrB-WT (Ll.LtrB-WT) (**B**, **C**, **E**) and Ll.LtrB-WT-ΔSDΔAUG (Ll.LtrB-ΔΔ) (**D**). Some bacterial cultures were treated with rifampicin (**C**) or erythromycin (**E**) before cell lysis to block transcription and translation, respectively.
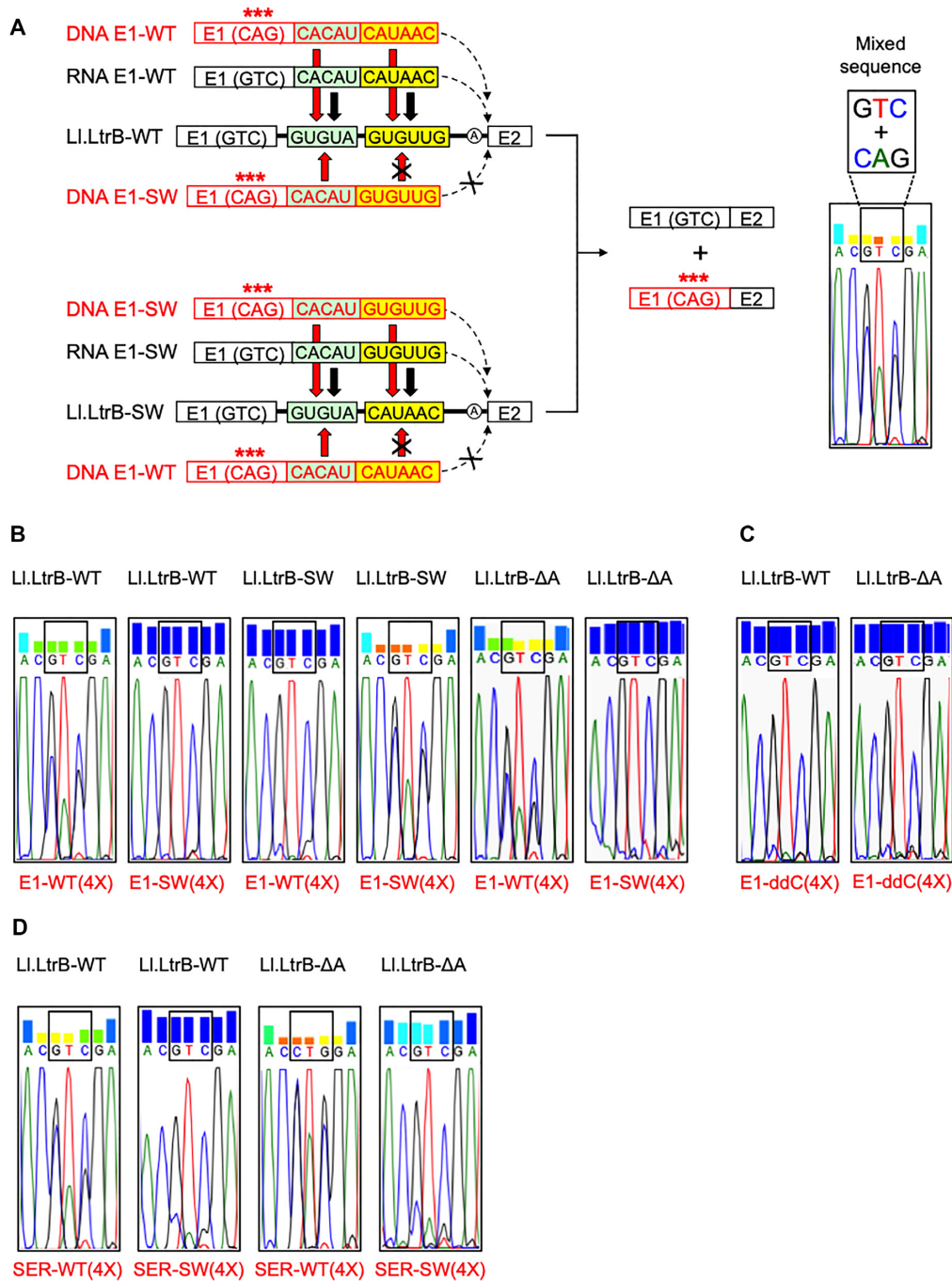
Figure 4. Free E1 is recruited by base pairing interactions to initiate the group II intron circularization pathway. (**A**) Schematic of the *ex vivo* group II intron circularization assay. The assay consists of adding short DNA oligos (DNA E1, red) as external sources of free E1 to *L. lactis* cell lysates expressing various constructs of the Ll.LtrB intron to assess how the intron recruits free E1 to initiate circularization and generate ligated exons. Ll.LtrB-WT (top left) and a variant with swapped IBS1/EBS1 sequences between E1 and the intron [Ll.LtrB-IBS1-EBS1 swap (Ll.LtrB-SW)] (bottom left) were used. The EBS1–IBS1 (yellow) and EBS2–IBS2 (green) base pairing interactions between the intron and E1 are represented by black [cognate *trans* E1 (E1(GTC))] and red [external *trans* E1 (E1(CAG))] arrows. The internal source of E1 coming from the pre-mRNA contains the wild-type GTC sequence at positions −24 to −22 [RNA E1: E1(GTC)], upstream of the intron binding sequences (IBS1/2), whereas the external source of E1 (red) contains CAG (asterisks) [DNA E1: E1(CAG)]. The schematic of the assay shows that the intron can use *in trans* either the internal or external source of E1 (dashed curved arrows) to produce two types of ligated exons: E1(GTC)–E2 or E1(CAG)–E2. A portion of the sequence chromatogram of the purified PCR amplicons spanning nucleotides −26 to −20 from the end of E1 is shown where nucleotides −24 to −22 are boxed. Nucleotides GTC or CAG are present at positions −24 to −22, depending on whether the internal or external source of E1 is found ligated to E2, respectively. A mixed sequence at positions −24 to −22 is thus expected in the chromatogram if both sources of ligated exons are generated. The colour code above each base is a visual representation of the Phred score calculated for base calling from 0 to 60 (red to blue). Ligated exons were amplified by RT-PCR from total RNA extracts containing Ll.LtrB-WT (Ll.LtrB-WT) (**B**, **C**, **D**), Ll.LtrB-ΔA (**B**, **C**, **D**) and Ll.LtrB-IBS1-EBS1 swap (Ll.LtrB-SW) (**B**). While the E1(CAG) DNA oligos (E1-WT and E1-SW) were 54 nt in length (**B**, **C**), the E1(CAG)-CACCCCCCCC oligos used for the SER assay (SER-WT and SER-SW) were 64 nt long (**D**, SER). The nucleotides at positions 3–10 of E2 were replaced by a stretch of C residues to distinguish the ligated exons provided *in trans* from the cognate ligated exons.

saw that modification of the last residue of free E1(CAG) from a deoxy C to a dideoxy C completely abrogated its use by both Ll.LtrB-WT and Ll.LtrB-ΔA [Figure 4C, E1-ddC(4×)].

Overall, these results corroborate our previous data (Figure 2) and further demonstrate that base pairing interactions, notably between the EBS1 and IBS1 regions, are used by unspliced Ll.LtrB to specifically recruit free E1 to the pre-mRNA. Moreover, they demonstrate that the first step of the group II intron circularization pathway is initiated *in trans* by the 3′OH of the last residue of recruited free E1.

Intron circles generate some free E1 through the SER reaction

During the proposed group II intron circularization pathway, free E1 is released by the second transesterification reaction (Figure 1C, step 2) that most likely serves to perpetuate the pathway (Figure 1C, a). However, an alternative mechanism must also exist in order to generate the initial source of free E1 that starts the first rounds of circularization. One of the proposed pathways is the SER reaction catalysed by excised lariats (Figure 1B, b) and linear introns (Figure 1D, c) (18,19,22,34).

To address SER as a potential source of free E1, we provided *L. lactis* cell extracts that expressed Ll.LtrB-WT with DNA oligos mimicking ligated exons [E1(CAG)–E2(CACC CCCCC)]. To distinguish the ligated exons provided *in trans* from the cognate ligated exons, nucleotides at positions 3–10 of E2 were replaced by a stretch of C residues. The two oligos used harboured either the complementary wild-type IBS1 [SER-WT(4×)] or the non-complementary swapped sequence [SER-SW(4×)] at the end of E1 just upstream of the splice junction. In agreement with our previous data, we detected in the pool of ligated exons the presence of the external source of E1(CAG) precisely ligated to the internal source of E2 exclusively for the oligo that contained wild-type IBS1 (Figure 4D, Ll.LtrB-WT). Unexpectedly, we saw the same phenotype for Ll.LtrB-ΔA (Figure 4D, Ll.LtrB-ΔA) despite the fact that this construct only uses the circularization pathway and that released intron circles were not previously considered able to perform SER (Figure 1C).

These results show that Ll.LtrB-WT and Ll.LtrB-ΔA can specifically recognize IBS1/2 regions with sequences complementary to their EBS1/2 sequences at the E1–E2 splice junction and precisely hydrolyse external sources of ligated exons by SER at the splice junction to contribute to the pool of free E1. Subsequently, the external source of processed free E1 can be recruited by the unspliced intron, through specific base pairing interactions, to initiate the first transesterification reaction of the circularization pathway.
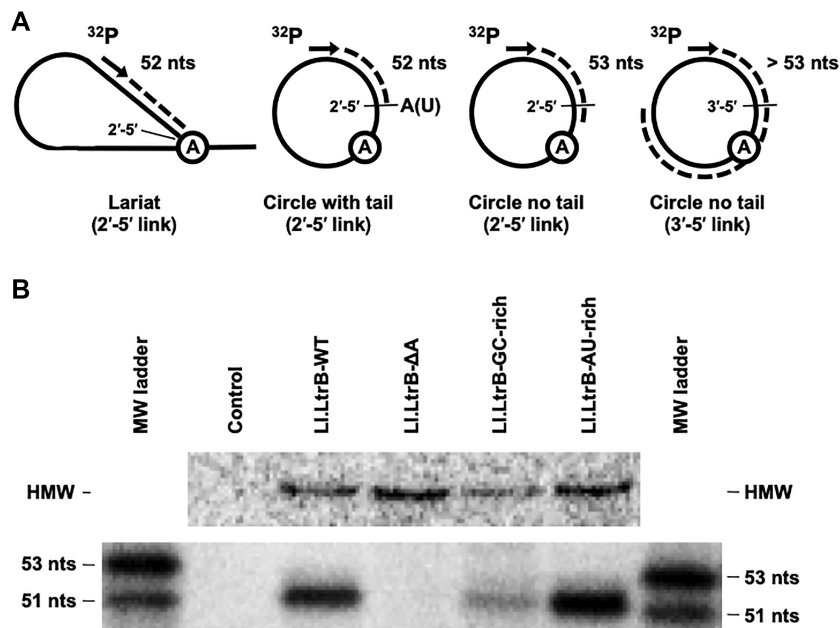
The first and last residues of released intron circles are joined by a 3′–5′ phosphodiester bond

Based on the branching pathway and *in vitro* data (19), the 2′OH of the last nucleotide of circularization intermediates was initially proposed to be the nucleophile of the second transesterification reaction of the group II intron circularization pathway, generating a 2′–5′ link between the first and the last nucleotide of the intron at the circularization junction (Figure 1C, step 2, CJ). Following the detection of E1-Ll.LtrB-ΔA circularization intermediates harbouring the first 2 nt (CA) or 3 nt (CAU) of E2 at the intron 3′ end (23,28) along with the presence of the first residue of E2 (C) at the circularization junction of released Ll.LtrB-ΔA circles (22,23), we previously proposed that excised intron circles harbour lariat-like tails of 1 nt (A) or 2 nt (AU) at their 2′–5′ circularization junction (23). However, several unsuccessful attempts to detect lariat-like tails attached to the extra C residue at the intron circularization junction led us to reassess both the presence of lariat-like tails and the nature of the phosphodiester bond at the intron circularization junction *in vivo*.

It was previously shown that the AMV RT pauses when it encounters 2′–5′ linkages in RNA templates, generating premature stops during primer extension (35). The enzyme pauses before reading through 2′–5′ linkages in RNA lariats and 1 nt after 2′–5′ bonds in substrates without 3′ extensions (19,35). Using the AMV RT, we performed primer extension assays across the released intron branching and circularization junctions from total RNA extracts of *L. lactis* expressing different Ll.LtrB variants (Figure 5). The primer was labelled with ^{32}P at the 5′ end and positioned to generate bands of 52 nt for intron lariats and circles with lariat-like tails, and 53 nt for full-length circles harbouring a 2′–5′ bond at the circularization junction (Figure 5A). Ll.LtrB-ΔA, which primarily splices through circularization (22,23), showed no bands (Figure 5B). Ll.LtrB-GC-rich, which produces significantly more circles than lariats (∼85:15) (28), showed only a faint band at 52 nt that corresponds to the small amount of released intron lariats and no band corresponding to circles at 53 nt. Furthermore, only one band of 52 nt, corresponding specifically to lariats, was detected for Ll.LtrB-WT and Ll.LtrB-AU-rich even though these constructs were previously shown to produce circle to lariat ratios of ∼50:50 and ∼15:85, respectively, *in vivo* (28). Thus, no signs of RT pausing were detected at the circularization junctions of full-length intron circles for all four Ll.LtrB constructs. The lack of signals corresponding to 2′–5′ linkages at full-length intron circularization junctions suggests that this bond is rather made of a 3′–5′ phosphodiester linkage not causing the RT enzyme to pause. An alternative explanation for the lack of pausing signal at the Ll.LtrB-ΔA circularization junction was that the overall amount of excised introns may be much lower for Ll.LtrB-ΔA compared to Ll.LtrB-WT. As a control, we mixed equal amounts of total RNA extracts of *L. lactis* expressing either Ll.LtrB-ΔA or Ll.LtrB-WT from identical constitutive promoters, amplified the released intron junctions by RT-PCR (22–25,28), cloned the mixed amplicon and sequenced independent clones. We detected ∼6 times more Ll.LtrB-ΔA circles compared to the total amount of Ll.LtrB-WT circles and lariats combined, showing that the absence of a signal corresponding to RT pausing for Ll.LtrB-ΔA is not due to lower amounts of released intron circles compared to released Ll.LtrB-WT. In addition, as expected, we detected similar levels of lariats and circles for Ll.LtrB-WT confirming that the AMV RT only pauses at the branching junction and not at the circularization junction of Ll.LtrB-WT full-length circles. Finally, a much higher molecular weight band

Figure 5. The first and last residues of released intron circles are joined by a 3′–5′ phosphodiester bond. Primer extension assays were performed across the Ll.LtrB branching and circularization junctions using the AMV RT. (**A**) Schematic of the various potential pause sites of the AMV RT at the branching and circularization junctions of different released intron lariats and circles. Extension (dashed line) of the ^{32}P-labelled primer (arrow) across the branching and circularization junctions may lead to bands of 52 nt [lariat (formed following branching), 2′–5′ link with a 3′ tail; circle with tail (formed following circularization where the attacking nucleophile is a 2′OH upstream of terminal nucleotide), 2′–5′ link with a 3′ tail], 53 nt [circle with no tail (formed following circularization where the attacking nucleophile is the 2′OH of terminal intron nucleotide), 2′–5′ link without a 3′ tail] or no bands (circle with no tail, 3′–5′ link). (**B**) Denaturing 12% polyacrylamide gel of primer extensions across the branching and circularization junctions of different Ll.LtrB variants. The two bands used as molecular weight markers (MW ladder) were generated using poisoned primer extension by adding high concentrations of dideoxy CTP during cDNA synthesis (23,33) and correspond to unspliced precursor mRNA (53 nt) and ligated exons (51 nt) of Ll.LtrB-WT expressed from pDL-P$_{23}$²-Ll.LtrB-WT. The top of the gel reveals the presence of much higher molecular weight bands (HMW).

was observed at the top of the gel for all intron variants studied including Ll.LtrB-ΔA (Figure 5B, HMW). This suggests that intron circles were indeed used as substrates by the AMV RT, but that the enzyme did not pause at the full-length junction because it encountered a 3′–5′ instead of a 2′–5′ link (35).

Taken together, these results demonstrate that the phophodiester bond at the circularization junction of released intron circles is a 3′–5′ link and that Ll.LtrB-ΔA circles do not harbour lariat-like tails nor a free 3′OH at their circularization junction *in vivo*. By extension, our data also show that the nucleophile of the second transesterification reaction of the circularization pathway is the 3′OH of the last nucleotide of circularization intermediates.

Group II introns use a molecular ruler to specifically select the nucleophile of the second transesterification reaction of the circularization pathway

We previously demonstrated that Ll.LtrB circularization intermediates with different 3′ ends are generated by the first transesterification step (23,28) and that some intron circles harbour additional nucleotides at their circularization junction (22,23). Having shown that the phosphodiester bond at the circularization junction of intron circles is a 3′–5′ link (Figure 5), we next wanted to assess what rule dictates which 3′OH is used by the intron as the nucleophile of the second transesterification reaction.

We first analysed the 3′ ends of E1–intron circularization intermediates generated after the first transesterification reaction for Ll.LtrB-ΔA and a series of variants missing one [Ll.LtrB-ΔA(−A)] or two [Ll.LtrB-ΔA(−AC)] nucleotides and having one [Ll.LtrB-ΔA(+A)] or two [Ll.LtrB-ΔA(+AA)] additional nucleotides in the linker region between DVI and E2 (Figure 6A). Next, we identified the nucleotides present at the intron circularization junction of these five constructs after the second transesterification reaction (Figure 6B).

We initially used 3′ RACE assays to identify the 3′OH residues available at the end of the various intron circularization intermediates (Figure 6A) (23,28). Because several A and U residues are present at the 3′ splice site, the 3′ RACE assay was done by the addition of poly(A) and poly(U) tails independently in order to identify all the free 3′ ends available for the five Ll.LtrB-ΔA variants (23,28). As previously described for Ll.LtrB-ΔA (23), the 3′ splice site is misrecognized during the first nucleophilic attack of the circularization pathway generating various 3′ ends for the five constructs (Supplementary Figure S2). Overall, we detected a series of free 3′OH at positions +3 to +6 from the base of the DVI stem (Figure 6C, vertical lines).

Next, we amplified by RT-PCR the circularization junction of the five Ll.LtrB-ΔA constructs to identify the nucleotide composition at these junctions (Figure 6B) (22–25,28). We first corroborated our previous results showing that removal of the branch point A residue results mainly
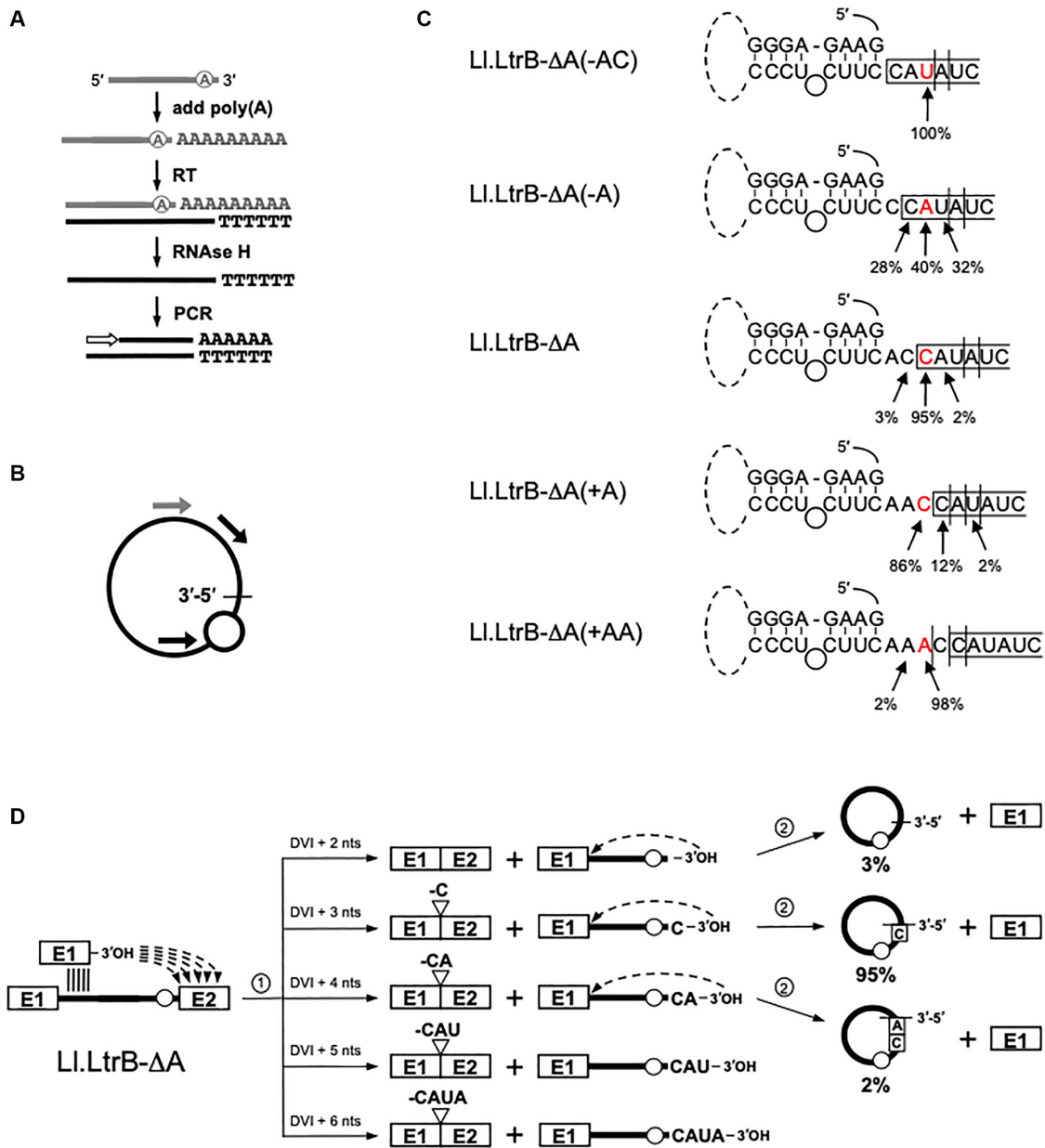
Figure 6. The 3′OH that initiates the second nucleophilic attack of the group II intron circularization pathway is selected by a molecular ruler. (**A**) The 3′ ends of circularization intermediates were identified by 3′ RACE. Free intron 3′ ends were amplified by RT-PCR from *L. lactis* total RNA extracts. Intron 3′ ends were identified by first extending the intron RNA with a poly(A) tail followed by the synthesis of cDNA with an oligo dT. The RNA strand was removed by an RNAse H treatment and the single-strand DNA amplified by PCR. The same procedure was repeated for all constructs but extending a poly(U) instead of a poly(A) tail at the 3′ end of the intron (Supplementary Figure S2). (**B**) The nucleotides present at the circularization junction of released intron circles were identified by RT-PCR. The primers used are represented by grey (RT) and black (PCR) arrows. The absence of the branch point A residue is illustrated by an empty circle. (**C**) The 3′ splice site of five Ll.LtrB-ΔA constructs is depicted showing the last stem of DVI attached to E2 (boxed sequence). One or two nucleotides were removed [Ll.LtrB-ΔA(−A), Ll.LtrB-ΔA(−AC)] or added [Ll.LtrB-ΔA(+A), Ll.LtrB-ΔA(+AA)] to the linker between DVI and E2 of Ll.LtrB-ΔA. Vertical lines represent free 3′ ends identified by 3′ RACE (**A**), while the nucleotides found at circularization junctions (**B**) are shown as percentages. The most represented sequences at the circularization junction mostly correspond for all the intron variants studied to the third nucleotide from the bottom of the DVI stem (red nucleotide). (**D**) Molecular details of both steps of the group II intron circularization pathway for the Ll.LtrB-ΔA construct. Prior to the initiation of the first step of the circularization pathway, the unspliced intron recruits free E1 through specific base paring interactions between the intron EBS1/2 and the E1 IBS1/2 sequences (vertical lines). The first step of the pathway is initiated at the 3′ splice site of the pre-mRNA by the 3′OH of the recruited free E1 (step 1). The first transesterification reaction is not always accurate (step 1, dashed arrows) and generates circularization intermediates with different 3′ ends along with their corresponding ligated exons. Circularization intermediates with 5 and 6 nt from the base of the DVI stem (DVI + 5 nt, DVI + 6 nt) were identified by 3′ RACE (**C**, vertical lines). Next, the favoured nucleophile to initiate the second transesterification reaction is the 3′OH of the third nucleotide after the base of the DVI stem (DVI + 3 nt) (step 2, dashed arrow). The great majority of the Ll.LtrB-ΔA released circles (95%) have an extra C at the circularization junction and were generated by the circularization intermediate with 3 nt after DVI. The 3′OH of the third nucleotide seems to be at the optimal distance to be positioned within the intron active site. A small number of Ll.LtrB-ΔA released circles were generated from circularization intermediates with 2 nt (3%) and 4 nt (2%) after DVI. The 3′OH of the detected intermediates with 5 nt (CAU) and 6 nt (CAUA) after DVI is not used by the intron to initiate the second step of the pathway.

in the addition of the first nucleotide of E2 (+C, 95%) at the circularization junction (Figure 6C and D; Ll.LtrB-ΔA) (22,23). The addition of 1 nt [Ll.LtrB-ΔA(+A)] and 2 nt [Ll.LtrB-ΔA(+AA)] in the linker between DVI and E2 mostly led to the generation of full-length circles (86%) and circles missing the last nucleotide of the intron (−C, 98%) respectively. On the other hand, the removal of 1 nt [Ll.LtrB-ΔA(−A)] generated a mix of intron circles harbouring one (+C, 28%), two (+CA, 40%) or three (+CAU, 32%) additional nucleotides from E2 at the circularization junction. Finally, removing 2 nt of the linker region [Ll.LtrB-ΔA(−AC)] gave rise exclusively to intron circles with the first 3 nt of E2 (+CAU, 100%) at the circularization junction.

These results show the presence of circularization intermediates with different 3′ ends as well as different nucleotide compositions at the circularization junction for all the Ll.LtrB-ΔA variants studied. The nucleotide composition at the circularization junctions further reveals that, in contrast to the 3′ splice site, the 5′ splice site is precisely and accurately recognized during the second transesterification reaction. Taken together, our data demonstrate that the 3′OH residue used as the nucleophile of the second step of circularization is chosen using a molecular ruler, since the nucleophilic attack at the 5′ splice site is initiated most of the time by the 3′OH of the third nucleotide from the base of the DVI stem by the five Ll.LtrB-ΔA constructs (Figure 6C, red nucleotides).

DISCUSSION

Bacterial group II introns were recently shown to use, in addition to branching, a second significant splicing pathway called circularization (22–29). Released intron lariats and circles were both found at substantial levels *in vivo*. Various factors including the strength of the base pairing interactions between E1 and the intron, the expression level of the interrupted gene, the host environment and growth temperature were also demonstrated to greatly influence the relative proportion of released intron circles and lariats (28). The group II intron circularization pathway was shown to be involved in the generation of genetic diversity by exon shuffling in bacterial cells leading in some cases to the production of chimeric enzymes with modified specificities (24,26). Nevertheless, in contrast to branching, the molecular details of the proposed circularization pathway were mostly based on assumptions and had largely remained overlooked. Importantly, the molecular switch that determines whether the intron initiates splicing through the branching or the circularization pathway was not known. Here, we provide a detailed and complete characterization of the two transesterification reactions of the group II intron circularization pathway at the molecular level *in vivo*. Our data also provide insights about the molecular switches that trigger the structural rearrangements between the pre-5′ and pre-3′ splice site processing conformations and enable the initiation of the circularization pathway at the 3′ splice site.

We first demonstrated that the group II circularization pathway is initiated by free E1 at the 3′ splice site. Free E1 was shown to be recruited by the unspliced intron within pre-mRNAs and precisely *trans*-spliced to E2 leading to the release of ligated exons. These results strongly suggest that E1 is precisely released from the 5′ splice junction of group II intron-interrupted pre-mRNA transcripts and is freely available *in vivo*. Our data also demonstrated that the EBS1/2–IBS1/2 base pairing interactions are absolutely required for the specific recruitment of free E1 by the intron in order to initiate the first step of the circularization pathway. The ability of the intron to use a short DNA oligonucleotide as an external source of free E1, which was completely lost when the oligo harboured a dideoxy C residue as the last nucleotide, showed that the nucleophile of the first transesterification reaction is the 3′OH of the terminal residue of free E1. However, this does not completely rule out the potential use of the 2′OH as the nucleophile of this reaction. Recruitment of free E1 by the intron was found to be significantly reduced when *cis*-E1 is not translated. This suggests that free E1 can compete more efficiently with *cis*-E1 to base pair with the EBS1/2 sequences of the intron when *cis*-E1 is intermittently disengaged from EBS1/2 by the passage of ribosomes during translation. Alternatively, since naked mRNAs may be more prone to degradation this could contribute to the overall reduction of free E1 recruitment. Our data also revealed that active transcription is required neither for the recruitment of free E1 nor for the initiation of the first step of group II intron circularization.

We also showed that intron circles can generate free E1 from ligated exons by the SER reaction. Lariats and linear introns were previously shown to produce free E1 by SER but not released intron circles (Figure 1, SER) (18,19,34). Our results demonstrate that released group II intron circles are also able to specifically recognize the spliced junction of ligated exons through base pairing and liberate E1 by hydrolysing the phosphodiester bond precisely at the junction between E1 and E2. Our data further showed that free E1, released by SER, can also be specifically recruited by base pairing and ligated to E2. This two-step process of SER followed by E1 *trans*-splicing was previously observed *in vitro* for the mitochondrial aI5γ group II intron from *Saccharomyces cerevisiae* (36).

Our work also demonstrated that recruited free E1 initiates the first nucleophilic attack at more than one position in the vicinity of the 3′ splice site, generating circularization intermediates with heterogenous 3′ ends. Using five Ll.LtrB-ΔA variants with linkers of different lengths (0–4 nt) between the base of DVI and E2, we identified circularization intermediates harbouring at their 3′ end between 3 and 6 nt from the base of the DVI stem. Nevertheless, the great majority (84%) of intron circles identified for the five Ll.LtrB-ΔA variants used the 3′OH of the third nucleotide from the base of the DVI stem as the nucleophile of the second transesterification reaction, leading to different nucleotide compositions at their circularization junctions. Since the only variation we observed in nucleotide composition at the circularization junction originated from the 3′ end of the circularization intermediates, we concluded that recognition of the 5′ splice site is specific and accurate. Taken together, this indicates that circularization intermediates use a molecular ruler to position the nucleophile of the second transesterification reaction in the intron active site to precisely attack the phosphodiester bond at the 5′ splice site. For example,

Figure 6D illustrates the various circularization intermediates detected along with the released intron circles generated specifically for Ll.LtrB-ΔA. Two circularization intermediates with five (DVI + 5 nt) and six (DVI + 6 nt) residues from the base of the DVI stem were detected. These intermediates presumably accumulate *in vivo* since they do not lead to the generation of released intron circles. Released intron circles with none (3%), one (C) (95%) or two (CA) (2%) additional nucleotides at the circularization junction were identified. However, the corresponding circularization intermediates with two (DVI + 2 nt), three (DVI + 3 nt) and four (DVI + 4 nt) nucleotides from the base of the DVI stem were not detected by 3′ RACE, presumably because they are in lower amounts and/or proficient substrates to initiate the second transesterification reaction and thus actively circularize, making them more difficult to detect. The same molecular ruler principle also applies to Ll.LtrB-WT, leading mostly to the generation of full-length intron circles and accurately ligated exons (22,23). Because of the presence of the branch point A residue in the last stem of DVI, the nucleophile of the second step of circularization for Ll.LtrB-WT is the 3′OH of the second nucleotide from the base of the DVI stem, compared to the third nucleotide for the five Ll.LtrB-ΔA constructs, which all lack the branch point residue. This shows that the last stem of DVI is an element of the molecular ruler. Ll.LtrB was recently shown to circularize from a population of intron-interrupted mRNAs rather than its own cognate mRNA, thus generating chimeric mRNA molecules (24). From a statistical perspective, only a minority (33%) of such *trans*-spliced products would yield in-frame mRNAs that can be translated. However, the inaccuracy of the first transesterification reaction at the 3′ splice site, leading to the production of some out of frame ligated exons shown here, likely increases the chances for each invaded transcript to yield viable chimeric mRNAs through circularization, albeit at low levels.

Using the AMV RT that pauses before 2′–5′ links that have 3′ extensions and after 2′–5′ links without 3′ extensions, we demonstrated that the circularization junction of released intron circles consists of a 3′–5′ rather than a 2′–5′ link *in vivo*. In accordance with our data on both circularization intermediates and nucleotide composition at the circularization junctions, these results confirm that the nucleophile of the second transesterification is the 3′OH of the last nucleotide of circularization intermediates and that group II intron circles do not harbour 3′ tail extensions as previously proposed (23). Accordingly, we performed several unsuccessful assays to detect the 2′–5′ link and identify the previously proposed lariat-like tails of 1 nt (A) and 2 nt (AU) presumably attached at the 3′OH of the first C residue of E2 present at the Ll.LtrB-ΔA circularization junction (23). In agreement with the data described here, various assays that combined 3′ RACE or 3′ end oligo ligation followed by the use of a 2′–5′ debranching enzyme (Dbr1) were all unsuccessful in detecting 2′–5′ linkages and lariat-like tails for Ll.LtrB-ΔA. However, we cannot completely rule out that the 2′OH of the terminal residue of circularization intermediates is sometimes used as the nucleophile of the second transesterification reaction since we detect a very faint band at 53 nt for both Ll.LtrB-WT and Ll.LtrB-AU-rich when we significantly overexpose the gel. In addition,

in vitro studies using the yeast mitochondrial group II intron aI5γ previously identified the circularization junction as a 2′–5′ linkage (19). Nevertheless, no faint bands at 53 nt were detected for Ll.LtrB-ΔA and Ll.LtrB-GC-rich at higher exposures, both of which self-splice predominantly as circles *in vivo*. While circularization itself has been shown to increase genetic diversity by generating chimeric mRNAs (24) and to produce novel enzymes with gain-of-function phenotypes (26), the function of released intron circles has remained elusive. The accumulation of intron RNA circles in bacterial cells could serve a specific function yet to be discovered or represent the sole consequence of their increased stability.

During transcription of group II intron-interrupted genes, the intron folds as the RNA emerges from the RNA polymerase (37). The folding of DI, harbouring the EBS1/2 sequence motifs, allows for the docking of *cis*-E1 through base pairing interactions with the IBS1/2 sequence motifs (16). Following the complete folding of DI, the other five intron domains (DII–DVI) sequentially fold and assemble onto the DI structural scaffold (15,16). Crystal structures of DI show that the base pairing interaction between *cis*-E1 and the intron occurs at the surface of the DI platform within the central cavity. The *cis*-E1–intron base pairing interactions remain accessible until DV and DVI dock onto DI to complete the assembly of the catalytic core, totally surrounding the 5′ splice site (15,16). Even if DI fluctuates between the open and closed structures before branching is ultimately initiated, the EBS1/2 sequence motifs remain accessible in both states.

In this dynamic structural context, our data suggest that free E1 initiates the group II intron circularization pathway by displacing *cis*-E1 from the EBS1/2 binding site after proper folding of the intron instead of competing with *cis*-E1 to base pair with the EBS1/2 motifs during the initial folding steps of DI. Accordingly, using a splicing-deficient Ll.LtrB variant with a mutated EBS1 motif (Ll.LtrB-EBS1) that cannot recognize the IBS1 motif of *cis*-E1 at its 5′ splice site, we found that an external source of E1 [E1(CAG)SW], harbouring a complementary IBS1 sequence, cannot rescue its splicing by circularization. This suggests either that the intron is unable to recruit free E1 during folding or that the initial binding of *cis*-E1 by the intron may be required for the subsequent formation of the active site, such that the early recruitment of free E1 instead of *cis*-E1 would lead to improper folding and/or splicing-incompetent structures. In support of that hypothesis, we found that preventing translation of *cis*-E1 almost completely abrogated the use of free E1 by the intron, showing that disruption of base pairing between an initial *cis*-E1–intron interaction facilitates the recruitment of free E1 within a properly folded splicing-competent structure. We previously demonstrated that various factors influence the circle to lariat ratios of Ll.LtrB *in vivo* (28). For example, the relative proportion of released intron circles compared to lariats consistently increased with temperature as well as with the strength of the base pairing interactions between the 3′ end of E1 and the intron (28). In light of our data, this suggests that free E1 can compete more efficiently with *cis*-E1 to get access to the EBS1/2 motifs of the intron at higher temperature and when the EBS1/2–IBS1/2 interactions are stronger. Using
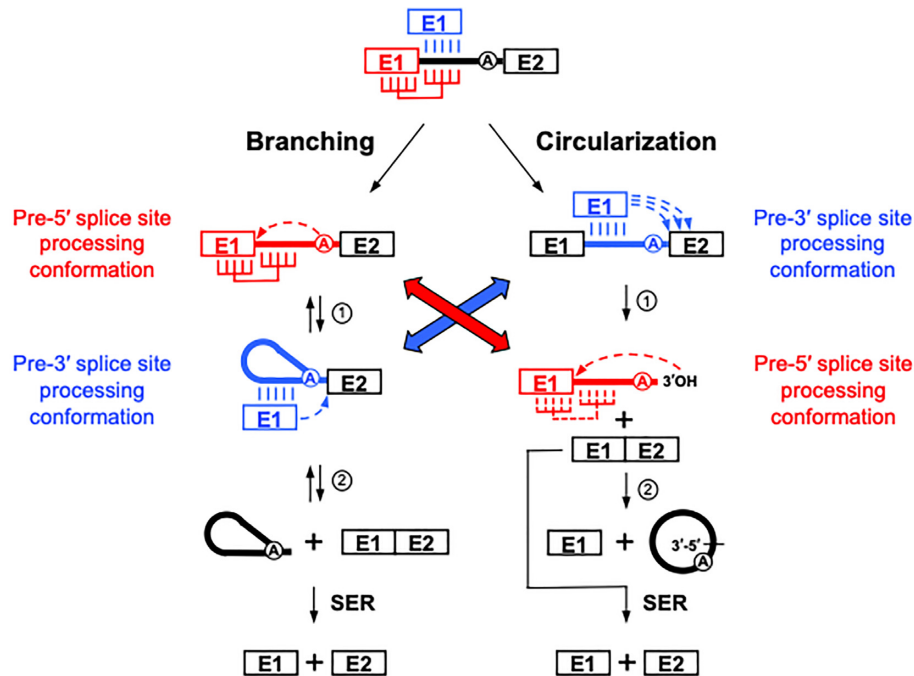
Figure 7. Model of the active site conformations between both steps of the circularization and branching splicing pathways. During the intron folding pathway, the IBS1/2 sequences of *cis*-E1, covalently attached to the intron 5′ end, bind to the EBS1/2 of the intron through base pairing interactions (top, red vertical lines). Following complete intron folding, free E1 competes with *cis*-E1 to base pair with the EBS1/2 motifs of the intron (top, blue vertical lines). *Branching*: The intron adopts the pre-5′ splice site processing conformation (red) when *cis*-E1 is base paired to the intron (red vertical lines). This leads to the nucleophilic attack of the 5′ splice site by the 2′OH of the branch point nucleotide (step 1, red dashed arrow). Cleavage of the 5′ splice site during the first step of branching was proposed to trigger a significant rearrangement of the catalytic site putting the intron in the pre-3′ splice site processing conformation (blue) and allowing for the second reaction to occur at the 3′ splice site (step 2, blue dashed arrow). *Circularization*: The displacement of *cis*-E1 in conjunction with the binding of free E1 to the intron EBS1/2 motifs (blue vertical lines) causes a significant rearrangement of the catalytic site, putting the intron in the pre-3′ splice site processing conformation (blue). This allows for the first circularization reaction to occur at the 3′ splice site (step 1). Recognition of the intron–E2 3′ splice site is imprecise (blue dashed arrows) producing a series of E1–intron splicing intermediates with different 3′ ends. The first step of circularization is equivalent to the second step of branching (blue double-headed arrow): E1 is base paired to the intron but not covalently attached to its 5′ end, while the 3′OH of E1 attacks the 3′ splice site leading to the release of ligated exons. Cleavage of the 3′ splice site and the release of ligated exons during the first step of circularization trigger a significant rearrangement of the catalytic site putting the intron in the pre-5′ splice site processing conformation (red) and allowing for the second reaction to occur at the 5′ splice site (step 2). The nucleophile is chosen using a molecular ruler and the recognition of the E1–intron 5′ splice site is accurate (red dashed arrow). The second step of circularization is equivalent to the first step of branching (red double-headed arrow): the attack at the 5′ splice site closes the intron structure in lariat or circle form and releases free E1. However, in the second step of circularization, E1 is most likely not base paired to the intron (dashed red vertical lines) since it is not the 2′OH of the branch point nucleotide but rather the 3′OH of the last nucleotide of the intron, 6 nt further away, that induces the nucleophilic attack. Our work also demonstrated that released intron circles harbour a 3′–5′ link at their circularization junctions and can contribute to the generation of free E1 by processing ligated exons (E1–E2) through the SER reaction similarly to released intron lariats.

our *ex vivo* splicing assay, we indeed demonstrated that the Ll.LtrB-GC-rich construct (28), harbouring stronger EBS1/2–IBS1/2 interactions than Ll.LtrB-WT, efficiently initiates the circularization pathway even when translation is inhibited (Supplementary Figure S3). This is in contrast to Ll.LtrB-WT that was shown to be significantly affected when translation was prevented (Figure 3D and E). Collectively, these data support the hypothesis that docking of *cis*-E1 to DI through base pairing is a prerequisite for proper intron folding and that free E1 can only be functionally recruited by the intron to initiate circularization once it is properly folded. The various factors influencing the circle to lariat ratios of Ll.LtrB *in vivo* (28) are most likely modulating the efficiency of free E1 recruitment and *cis*-E1 displacement. More in-depth studies are required to determine the difference in free energy between *cis*-E1 and free E1, as well as how the various external factors described previously (28) and in this study influence these intermolecular interactions from an energetic perspective.

Both steps of the group II intron branching pathway at the 5′ and 3′ splice sites occur within a unique catalytic centre, implying an important conformational change between the two transesterification reactions (15,16). Moreover, this splicing pathway is completely reversible (Figure 1A, steps 1 and 2, double arrows) where reverse splicing is used by released intron lariats to invade DNA substrates during retrohoming and retrotransposition. Recent structural studies revealed the conformation of the active site before the first and second steps of branching as well as the conformational changes occurring between these two steps (38–40). Upon proper intron folding, *cis*-E1 is docked to the intron through its interaction with DI (Figure 7, top, red vertical lines), guiding the intron to adopt by default the pre-first step conformation of branching. However, our

work unveiled that free E1 is produced, available and in competition with *cis*-E1 to base pair with the intron at the same EBS1/2 sequence motifs (Figure 7, top, blue vertical lines). The recruitment of free E1 by the intron within the pre-mRNA, through base pairing, leads to the displacement of *cis*-E1 from the EBS1/2 motifs, which most likely triggers a significant conformational change of the intron from the pre-first step (pre-5′ splice site processing) to the pre-second step (pre-3′ splice site processing) conformation of branching. This would allow for the first transesterification reaction of the circularization pathway to occur at the 3′ splice site rather than at the 5′ splice site. The initial step of circularization (Figure 7, circularization, step 1) is equivalent to the second step of the branching pathway (Figure 7, branching, step 2) where *cis*-E1 is no longer covalently attached to the 5′ end of the intron and only associated through base pairing interactions (Figure 7, double-headed blue arrow). It was previously proposed that the conformational change during the first step of branching is triggered by the cleavage of the 5′ splice site (15). However, in the context of the circularization pathway, our data indicate that displacement of *cis*-E1 by free E1 is sufficient to induce the structural change from the pre-5′ splice site processing to the pre-3′ splice site processing conformation. We hypothesize that the inaccurate recognition of the 3′ splice site by the 3′OH of free E1 during the first step of circularization (Figure 7, circularization, step 1, blue dashed arrows) is due to the presence of *cis*-E1 still attached to the 5′ end of the intron. Cleavage of the 3′ splice site and the release of ligated exons would then induce a second conformational change allowing the intron to adopt the alternative catalytically active conformation, the pre-5′ splice site processing conformation, allowing the second step of the circularization reaction to occur at the 5′ splice site. This second transesterification reaction (Figure 7, circularization, step 2) is equivalent to the first step of the branching pathway (Figure 7, branching, step 1) where *cis*-E1 is covalently attached to the 5′ end of the intron and associated with DI through base pairing interactions (Figure 7, double-headed red arrow). However, in the second step of circularization, *cis*-E1 is most likely not base paired to the intron (Figure 7, circularization, dashed red vertical lines) because it was previously displaced, prior to step 1, by the recruitment of free E1. Our data also show that it is not the 2′OH of the branch point nucleotide, even when present for Ll.LtrB-WT, but rather the 3′OH of the last nucleotide of the intron, 6 nt further downstream, that induces this nucleophilic attack (Figure 7, circularization, step 2). In contrast to the first step, the nucleophilic attack of the second step at the E1–intron 5′ splice site, controlled by a molecular ruler, is accurate (Figure 7, circularization, step 2, red dashed arrow).

In summary, our work reveals that the initial processing of the 3′ splice site followed by the 5′ splice site during group II intron circularization is catalytically equivalent to complete reverse splicing of the branching pathway. It also suggests that the conformational shift around a unique active site from the pre-5′ to the pre-3′ splice site processing conformation within the pre-mRNA enables group II introns to be spliced out as circles instead of lariats. Our work further implies that the molecular switch between the branching and circularization pathways *in vivo* is largely based on

the way in which E1 is associated with the intron and influenced by a variety of biological determinants. If E1 is covalently bound to the intron 5′ end (*cis*-E1) and base paired to the EBS1/2 motifs (Figure 7, top, red vertical lines), the pre-5′ splice site processing conformation would be adopted and the first transesterification would occur at the 5′ splice site to initiate branching (Figure 7, branching, step 1). However, the displacement of *cis*-E1 triggered by the recruitment of an additional free E1, not covalently linked to the intron but base paired with the EBS1/2 sequences (Figure 7, top, blue vertical lines), would induce a conformational change of the active site to the pre-3′ splice site processing conformation and rather initiate the circularization pathway at the 3′ splice site (Figure 7, circularization, step 1). Our findings thus provide a mechanistic framework to explain how alternative splicing pathways and products can arise from a complex ribozyme with a unique versatile catalytic centre.

DATA AVAILABILITY

Data needed to evaluate the conclusions of this study are all present in the paper and the Supplementary Data.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

REFERENCES

1. Lambowitz,A.M. and Zimmerly,S. (2011) Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.*, **3**, a003616.
2. Chalamcharla,V.R., Curcio,M.J. and Belfort,M. (2010) Nuclear expression of a group II intron is consistent with spliceosomal intron ancestry. *Genes Dev.*, **24**, 827–836.
3. Lambowitz,A.M. and Belfort,M. (2015) Mobile bacterial group II introns at the crux of eukaryotic evolution. *Microbiol. Spectr.*, **3**, MDNA3-0050-2014.
4. McNeil,B.A., Semper,C. and Zimmerly,S. (2016) Group II introns: versatile ribozymes and retroelements. *Wiley Interdiscip. Rev. RNA*, **7**, 341–355.
5. Michel,F. and Ferat,J.L. (1995) Structure and activities of group II introns. *Annu. Rev. Biochem.*, **64**, 435–461.
6. Michel,F. and Dujon,B. (1983) Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. *EMBO J.*, **2**, 33–38.
7. Ichiyanagi,K., Beauregard,A., Lawrence,S., Smith,D., Cousineau,B. and Belfort,M. (2002) Retrotransposition of the Ll.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. *Mol. Microbiol.*, **46**, 1259–1272.
8. Cousineau,B., Lawrence,S., Smith,D. and Belfort,M. (2000) Retrotransposition of a bacterial group II intron. *Nature*, **404**, 1018–1021.
9. Cousineau,B., Smith,D., Lawrence-Cavanagh,S., Mueller,J.E., Yang,J., Mills,D., Manias,D., Dunny,G., Lambowitz,A.M. and Belfort,M. (1998) Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell*, **94**, 451–462.
10. Peebles,C.L., Perlman,P.S., Mecklenburg,K.L., Petrillo,M.L., Tabor,J.H., Jarrell,K.A. and Cheng,H.L. (1986) A self-splicing RNA excises an intron lariat. *Cell*, **44**, 213–223.

11. van der Veen,R., Arnberg,A.C., van der Horst,G., Bonen,L., Tabak,H.F. and Grivell,L.A. (1986) Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing *in vitro*. *Cell*, **44**, 225–234.

12. Padgett,R.A., Konarska,M.M., Grabowski,P.J., Hardy,S.F. and Sharp,P.A. (1984) Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science*, **225**, 898–903.

13. Steitz,T.A. and Steitz,J.A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl Acad. Sci. U.S.A.*, **90**, 6498–6502.

14. Toor,N., Keating,K.S., Taylor,S.D. and Pyle,A.M. (2008) Crystal structure of a self-spliced group II intron. *Science*, **320**, 77–82.

15. Zhao,C. and Pyle,A.M. (2017) Structural insights into the mechanism of group II intron splicing. *Trends Biochem. Sci.*, **42**, 470–482.

16. Zhao,C., Rajashankar,K.R., Marcia,M. and Pyle,A.M. (2015) Crystal structure of group II intron domain 1 reveals a template for RNA assembly. *Nat. Chem. Biol.*, **11**, 967–972.

17. Robart,A.R. and Zimmerly,S. (2005) Group II intron retroelements: function and diversity. *Cytogenet. Genome Res.*, **110**, 589–597.

18. Jarrell,K.A., Peebles,C.L., Dietrich,R.C., Romiti,S.L. and Perlman,P.S. (1988) Group II intron self-splicing. Alternative reaction conditions yield novel products. *J. Biol. Chem.*, **263**, 3432–3439.

19. Murray,H.L., Mikheeva,S., Coljee,V.W., Turczyk,B.M., Donahue,W.F., Bar-Shalom,A. and Jarrell,K.A. (2001) Excision of group II introns as circles. *Mol. Cell*, **8**, 201–211.

20. Fedorova,O. and Zingler,N. (2007) Group II introns: structure, folding and splicing mechanism. *Biol. Chem.*, **388**, 665–678.

21. Daniels,D.L., Michels,W.J. Jr and Pyle,A.M. (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.

22. Monat,C., Quiroga,C., Laroche-Johnston,F. and Cousineau,B. (2015) The Ll.LtrB intron from *Lactococcus lactis* excises as circles *in vivo*: insights into the group II intron circularization pathway. *RNA*, **21**, 1286–1293.

23. Monat,C. and Cousineau,B. (2016) Circularization pathway of a bacterial group II intron. *Nucleic Acids Res.*, **44**, 1845–1853.

24. LaRoche-Johnston,F., Monat,C., Coulombe,S. and Cousineau,B. (2018) Bacterial group II introns generate genetic diversity by circularization and *trans*-splicing from a population of intron-invaded mRNAs. *PLoS Genet.*, **14**, e1007792.

25. LaRoche-Johnston,F., Monat,C. and Cousineau,B. (2020) Detection of group II intron-generated chimeric mRNAs in bacterial cells. *Methods Mol. Biol.*, **2079**, 95–107.

26. LaRoche-Johnston,F., Bosan,R. and Cousineau,B. (2021) Group II introns generate functional chimeric relaxase enzymes with modified specificities through exon shuffling at both the RNA and DNA level. *Mol. Biol. Evol.*, **38**, 1075–1089.

27. Li-Pook-Than,J. and Bonen,L. (2006) Multiple physical forms of excised group II intron RNAs in wheat mitochondria. *Nucleic Acids Res.*, **34**, 2782–2790.

28. Monat,C. and Cousineau,B. (2020) The circle to lariat ratio of the Ll.LtrB group II intron from *Lactococcus lactis* is greatly influenced by a variety of biological determinants *in vivo*. *PLoS One*, **15**, e0237367.

29. Molina-Sanchez,M.D., Martinez-Abarca,F. and Toro,N. (2006) Excision of the *Sinorhizobium meliloti* group II intron RmInt1 as circles *in vivo*. *J. Biol. Chem.*, **281**, 28737–28744.

30. Matsuura,M., Saldanha,R., Ma,H., Wank,H., Yang,J., Mohr,G., Cavanagh,S., Dunny,G.M., Belfort,M. and Lambowitz,A.M. (1997) A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev.*, **11**, 2910–2924.

31. Belhocine,K., Mak,A.B. and Cousineau,B. (2007) Trans-splicing of the Ll.LtrB group II intron in *Lactococcus lactis*. *Nucleic Acids Res.*, **35**, 2257–2268.

32. Belhocine,K., Mak,A.B. and Cousineau,B. (2008) Trans-splicing versatility of the Ll.LtrB group II intron. *RNA*, **14**, 1782–1790.

33. Plante,I. and Cousineau,B. (2006) Restriction for gene insertion within the *Lactococcus lactis* Ll.LtrB group II intron. *RNA*, **12**, 1980–1992.

34. Qu,G., Piazza,C.L., Smith,D. and Belfort,M. (2018) Group II intron inhibits conjugative relaxase expression in bacteria by mRNA targeting. *eLife*, **7**, e34268.

35. Lorsch,J.R., Bartel,D.P. and Szostak,J.W. (1995) Reverse transcriptase reads through a 2′–5′ linkage and a 2′-thiophosphate in a template. *Nucleic Acids Res.*, **23**, 2811–2814.

36. Podar,M., Perlman,P.S. and Padgett,R.A. (1995) Stereochemical selectivity of group II intron splicing, reverse splicing, and hydrolysis reactions. *Mol. Cell. Biol.*, **15**, 4466–4478.

37. Su,L.J., Waldsich,C. and Pyle,A.M. (2005) An obligate intermediate along the slow folding pathway of a group II intron ribozyme. *Nucleic Acids Res.*, **33**, 6674–6687.

38. Chan,R.T., Peters,J.K., Robart,A.R., Wiryaman,T., Rajashankar,K.R. and Toor,N. (2018) Structural basis for the second step of group II intron splicing. *Nat. Commun.*, **9**, 4676.

39. Liu,N., Dong,X., Hu,C., Zeng,J., Wang,J., Wang,J., Wang,H.W. and Belfort,M. (2020) Exon and protein positioning in a pre-catalytic group II intron RNP primed for splicing. *Nucleic Acids Res.*, **48**, 11185–11198.

40. Manigrasso,J., Chillón,I., Genna,V., Vidossich,P., Somarowthu,S., Pyle,A.M., De Vivo,M. and Marcia,M. (2020) Visualizing group II intron dynamics between the first and second steps of splicing. *Nat. Commun.*, **11**, 2837.