Structural accommodations accompanying splicing of a group II intron RNP

Xiaolong Dong, Srivathsan Ranganathan, Guosheng Qu, Carol Lyn Piazza and Marlene Belfort[*]

Department of Biological Sciences and RNA Institute, University at Albany, Albany, NY 12222, USA

ABSTRACT

Group II introns, the putative progenitors of spliceosomal introns and retrotransposons, are ribozymes that are capable of self-splicing and DNA invasion. In the cell, group II introns form ribonucleoprotein (RNP) complexes with an intron-encoded protein, which is essential to folding, splicing and retromobility of the intron. To understand the structural accommodations underlying splicing, in preparation for retromobility, we probed the endogenously expressed *Lactococcus lactis* Ll.LtrB group II intron RNP using SHAPE. The results, which are consistent *in vivo* and *in vitro*, provide insights into the dynamics of the intron RNP as well as RNA–RNA and RNA–protein interactions. By comparing the excised intron RNP with mutant RNPs in the precursor state, confined SHAPE profile differences were observed, indicative of rearrangements at the active site as well as disengagement at the functional RNA–protein interface in transition between the two states. The exon-binding sequences in the intron RNA, which interact with the 5′ exon and the target DNA, show increased flexibility after splicing. In contrast, stability of major tertiary and protein interactions maintains the scaffold of the RNA through the splicing transition, while the active site is realigned in preparation for retromobility.

INTRODUCTION

Group II introns are ribozymes that appear evolutionarily related to spliceosomal introns and retrotransposons in eukaryotes (1,2). These self-splicing introns interact with an intron-encoded protein (IEP) and undergo two transesterification reactions to give rise to an intron lariat and ligated exons, as do spliceosomal introns. These transesterifications can be reversed, enabling integration into genomic DNA sites (Figure 1A) (3).

Although not conserved in the primary sequence, group II introns share a highly conserved secondary structure. The RNA organizes itself into six domains that radiate from a central hub (Figure 1B). Domain I (DI) is the first domain to be transcribed as well as the largest, comprising half of the ribozyme. DI provides a folding scaffold and docks the active site. Exon binding sequences (EBSs) in DI base pair with intron binding sequences (IBSs) in the exons (Figure 1B and C) and guide the exons to properly align with the active site. EBS-IBS interactions also play key roles in retromobility of the intron (4,5). Two smaller domains DII and DIII contribute both structurally and catalytically to the RNP. DIV contains the open reading frame (ORF) that encodes the IEP, as well as a high-affinity binding site for the IEP (3). DV is a small, highly conserved domain that contains the catalytic triad, which constitutes the core of the active site and interacts with the junction of DII and DIII (J2/3) (Figure 1B). DVI harbors a bulged adenosine, the nucleophile that initiates branching to form the lariat. Essential, long-range tertiary interactions such as α–α′, β–β′, μ–μ′ and π–π′ interact to form a conserved tertiary structure, enveloping the active site (3,5–9).

IEP binding is essential to the intron's folding, splicing and retromobility in the cell (3) (Figure 1A). The well-characterized IEP for the IIA intron Ll.LtrB from *Lactococcus lactis* is termed LtrA, which has maturase, reverse-transcriptase (RT) and endonuclease activities (3). The RT domain of LtrA contains seven conserved protein motifs (RT1-7) that form the Fingers and Palm regions, which catalyze reverse transcription, whereas the Thumb domain (domain X) forms the maturase required for splicing (9–12) (Figure 1D). The IEP binds to the RNA via an N-terminal extension (NTE) that is present in group II introns but not retroviral RTs. By binding to DIV, the IEP is suggested to down-regulate its own translation (13,14). Some IEPs lack the NTE but are still capable of binding to various group II introns and facilitate splicing (14–17), suggesting the specificity of the NTE to cognate RNA targets. The IEP has additional interactions with the RNA via the Fingers/Palm and Thumb domains (9,18,19).Downstream of the Thumb domain are the DNA binding domain (DBD) and the DNA
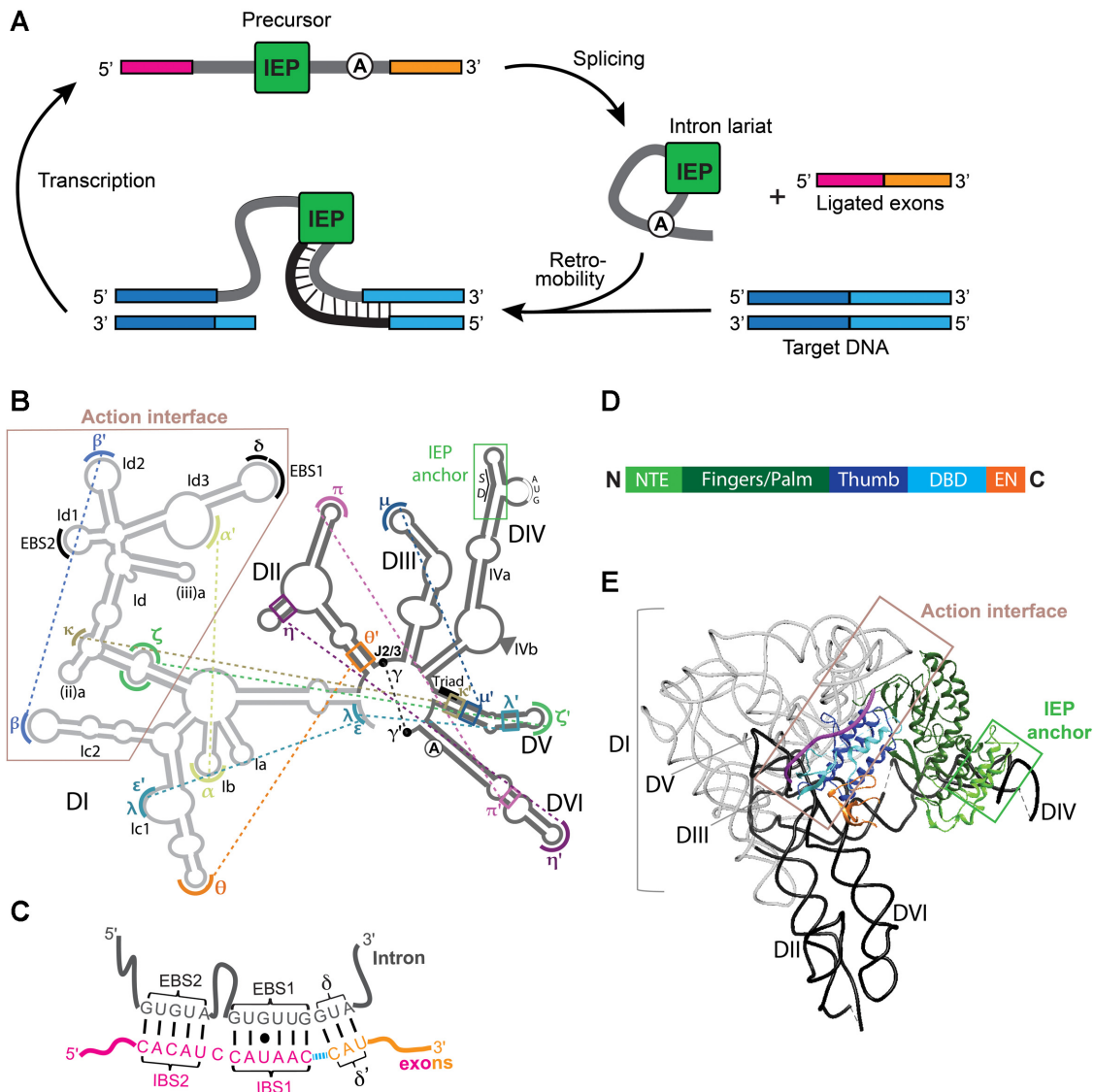
Figure 1. Activity and composition of a group II intron RNP. (**A**) Group II intron's cycle of activities. The intron RNA is shown in gray. Dark pink and gold denote exons 1 and 2, respectively. The branch point adenosine is shown as a circled 'A'. The transcribed precursor folds with help of the IEP (green) and undergoes two consecutive transesterification reactions to release the intron lariat from the exons that become ligated. The intron is then capable of recognizing and integrating into target DNA (blue) by first reverse-splicing into the top strand, followed by bottom strand cleavage and target-primed reverse transcription (TPRT, product shown in black). (**B**) Secondary structure of group II intron Ll.LtrB. DI of the intron RNA is shown in light gray and the rest of the domains are shown in dark gray. The tertiary interactions are indicated by dotted lines. The IEP anchor site and the action interface are boxed. DIVb is omitted for simplicity. (**C**) EBS-IBS interactions between the Ll.LtrB intron RNA and the exons. Blue dashed line denotes exon ligation site. The black dot denotes the G-U pair between EBS1 and IBS1. (**D**) Domain organization of the Ll.LtrB intron IEP (LtrA). NTE: N-Terminal Extension; DBD: DNA-Binding Domain; EN: Endonuclease domain. (**E**) Cryo-EM structure of Ll.LtrB intron RNP (9) (PDB: 5g2x), with the IEP anchor and action interface indicated by boxes in green and merlot, respectively. The magenta tube denotes the retained mRNA.

endonuclease (EN) domain (Figure 1D). The former is required for target DNA binding and the latter is a Mg^{2+}-dependent H–N–H DNA endonuclease that cleaves the bottom strand of the DNA to produce a primer for reverse transcription (Figure 1A). Both the DBD and EN domains are critical for retromobility (4,20–23).

The overall architecture of group II intron RNA is Y-shaped (8,24) with its 'arms' bridged by binding of the IEP to DI and DIVa (Figure 1E) (9). DIVa is a stem-loop structure protruding from the basal stem of DIV that consti-

tutes the IEP high-affinity binding site, which we refer to as the 'IEP anchor' (Figure 1B and E). The IEP also has weaker interactions with regions in DI (9), forming what we referred to as the 'action interface' (Figure 1B and E). In addition to helping stabilize the active RNP structure, these interactions have been proposed to mediate conformational changes in the RNP during key stages of splicing and mobility (19).

Major advances in understanding group II intron structure and function have been achieved using RNAs synthe-

sized *in vitro* (8,24–28). A recent breakthrough was provided by the 3.8-Å cryo-EM structure of endogenously spliced Ll.LtrB intron in complex with the IEP, which reveals not only the structure of the intron RNA, but also of the IEP and RNA–protein interaction sites (Figure 1E) (9). However, the structure of the group II intron precursor as a holoenzyme has not been characterized with high resolution and the dynamics of the RNA–RNA and RNA–protein interactions in the context of splicing and retromobility remain elusive. We sought to address those questions by probing the structure of endogenously spliced Ll.LtrB intron RNP and its unspliced precursor. Our results suggest consistency between the structures of purified RNPs and those within the cell. Long-range pairings and protein interactions stabilize the RNP scaffold, while structural rearrangements occur in defined regions of the intron RNA, suggesting a realigned active site as well as a less compact architecture at the action interface between the RNA and the IEP after splicing. These localized structural accommodations likely facilitate DNA access into the post-splicing RNP, in preparation for retromobility.

MATERIALS AND METHODS

RNP purification

The constructs expressing Ll.LtrB group II intron RNP each has a 27-nt 5′ exon (CUCGAGUCUA GAGAACACAUCCAUAAC) and a 54-nt 3′ exon (CAUAUCAUUUUUAAUUCUGCGGCCGCAGAA GGCCAUCCUGACGGAUGGCCUUUU), with the intron's ORF in DIVb deleted. The calculated molecular weights of the monomeric IEP, the precursor RNA and the spliced intron RNA are 70.1, 317.6 and 291.3 kDa, respectively (29).The splicing-competent construct, referred to as +A, has an additional 12XMS2 sequence inserted in the sequence following exon 2. The +A construct was expressed and purified by following previous procedures (29), with minor modifications. Briefly, after nisin induction and cell harvest, the cell suspensions were snap-frozen in liquid nitrogen and thawed three times before lysis by sonication. After the sucrose-gradient sedimentation and fractionation step, the pure RNP was applied to an Amicon Ultra-2 50-KDa MWCO filter unit for buffer exchange with RNP resuspension buffer (20 mM Tris–HCl pH 7.5, 200 mM KCl and 5 mM $MgCl_2$) before being snap-frozen in liquid nitrogen and stored at –80°C. The method used to purify the precursor RNPs is similar to that of +A and was modified from the previous work (30). Briefly, after the DTT elution, the eluate was subject to a 10% sucrose cushion ultracentrifugation. The pellets were resuspended and subject to another sucrose-gradient fractionation. Final products were concentrated and frozen as described above. The purified precursor RNPs were shown to be intact and homogeneous, as previously reported (29,30).

Primer extension analysis and RNP activity assay

Primer extension analysis was carried out following previous methods (9,30) using exon 2-specific primer IDT1072

(CTGCGGCCGCAGAATTAAAAATG) and intron 5′ primer IDT1073 (GTACCTTAAACTACTTGACTTA ACACC). The primers were labeled at the 5′ end with [γ-^{32}P] ATP (PerkinElmer) and T4 Polynucleotide Kinase (PNK, New England Biolabs). RNP activity assay was carried out following previous methods (9,30) using DNA oligonucleotides IDT3204 (CAACCCACGTCGATCGTG AACACATCCATAACCATATCATTTTTAATTCTAC) and IDT3206 (GTAGAATTAAAAATGATATGGTTA TGGATGTGTTCACGATCGACGTGGGTTG). The products were separated on a denaturing 7 M urea-8% polyacrylamide gel. The dried gel was exposed with a phosphor screen and scanned by GE Healthcare Typhoon Trio.

Size-exclusion chromatography (SEC)

SEC experiments were performed to determine the Stokes radius (Rs) of the RNPs. A Superose 6 10/300 GL column (GE Healthcare) was used with a flow rate of 0.3 mL/min. The SEC experiments were run at room temperature (25°C) with RNP resuspension buffer. The following globular protein standards were used: thyroglobulin (670 kDa), γ-globulin (150 kDa), ovalbumin (44.3 kDa) and RNase A (13.7 kDa). Prior to SEC, the RNPs were subject to a 15-min incubation at 4°C, at 37°C or at 37°C followed by a 15-min incubation on ice. Additionally, in a separate experiment, the RNPs were incubated at 30°C for 3 h, which is the same condition for nisin-induced expression, prior to SEC. To obtain RNA or IEP alone for SEC experiments, the +A, Triad and ΔA RNPs were digested with protease K or RNase A in RNP resuspension buffer at 25°C or 37°C for 1 h (with or without an additional 15 min incubation on ice) prior to SEC. For the exon competition experiment, *in vitro*-synthesized exon RNA (CUCGAGUCUAGAGAACACAUCCAUAA CCAUAUCAUUUUUAAUUCUGCGGCCGCAG) was added to the RNP at a 100× molar concentration and the mixture was incubated on ice for 1 h prior to SEC.

Target DNA binding and DNA oligonucleotide probing

The DNA binding was investigated by performing the electrophoretic mobility shift assay (EMSA). The DNA oligonucleotides were labeled at the 5′ end with [γ-^{32}P] ATP. For target DNA binding, 100 ng RNP was mixed with 0.4 pmol of labeled DNA in a 10 μl reaction with RNP reverse splicing buffer (50 mM Tris–HCl, pH 7.5, 10 mM KCl, 10 mM $MgCl_2$, 5 mM DTT, 0.1% Nonidet P-40 and 0.1% Tween-20) and incubated at 37°C for 1 h. After the incubation, 5 μl 37.5% glycerol was added and the mixture was separated on a 4% native polyacrylamide gel. The native gels were run in 0.5× Tris–borate–EDTA (TBE) buffer at room temperature (25°C) at 200 V for 2 h. The dried gel was exposed with a phosphor screen and scanned on a GE Healthcare Typhoon Trio.

For the DNA probing assay, thirty-six 25-nt DNA oligonucleotides (except DNA 36, which is 21-nt long) were designed to complement the RNA sequence in a non-overlapping, head-to-tail fashion (Figure 3A and Supple-

mentary Table S1). In addition, five tiling DNA probes (each 25-nt long, numbered 37–41) were used to target specific loop regions of interest (Supplementary Figure S3A and Supplementary Table S1). Concentrations of the RNPs were determined by measuring 260/280 nm absorbance and confirmed by RNA electrophoresis with quantitative comparison to RNA standards. Equal amounts of RNP were used in binding to allow comparison of affinity to probing DNA oligonucleotides. The experimental procedures were similar to those for target DNA binding described above, except that 150 ng of RNP was used. The incubations were at 37°C for 1 h unless otherwise stated. The binding levels were quantified using ImagequantTL. For Figure 3B, all gels were exposed with a single phosphor screen and the binding levels, after subtracting the RNP-free lane, were normalized to the binding of DNA probe 27 (assigned as 1.0) across repetitions of the experiment. For Supplementary Figure S3D and E, each probe was normalized to the value for the +A RNP at 37°C.

Selective 2′-Hydroxyl acylation analyzed by primer extension (SHAPE)

SHAPE experiments were carried out at room temperature unless otherwise noted. Dimethyl sulfoxide (DMSO) was used as the solvent for the SHAPE reagents. Purified RNP (150 ng) was diluted with RNP resuspension buffer to bring the volume to 9 μl. One microliter of 50 mM SHAPE reagent, benzoyl cyanide (BzCN), *N*-methylisatoic anhydride (NMIA) or 1-methyl-6-nitro-isatoic anhydride (1M6), was added to the RNP for a final concentration of 5 mM (10% v/v for DMSO). Reaction time for BzCN, NMIA and 1M6 were 10 s, 25 min and 5 min, respectively. In parallel to SHAPE treatments, DMSO was used as solvent control. After the reactions, RNase-free water was added to increase the volume to 200 μl. The mixture was then subjected to phenol-chloroform extraction and ethanol precipitation. The extracted RNA was resuspended in RNase-free water and subjected to primer extension analysis as described above. DNA probes 4, 7, 10, 13, 16, 20, 23, 26 and 36 (described in Figure 3A and Supplementary Table S1) were used as primers so that each primer allowed data acquisition for 75–100 nucleotides, except for nucleotides in DIVb. Data for nucleotides between A2457 and the 3′ end (C2492) of the +A RNP were not obtained due to lack of priming sites. For Triad and ΔA RNPs, 3′ exon primers IDT1072 and IDT6017 (GAATTCGTAATCGCGGCCGCAGAA T) were used to examine those nucleotides. The completed primer extension reaction was combined with an equal volume of 2X Gel Loading Dye II (Ambion) and separated on a denaturing 7 M urea-6% polyacrylamide sequencing gel. The stops were mapped by separation alongside sequencing ladders made with corresponding primers. The intensities of the bands for assigned nucleotides were measured with SAFA (31) or ImagequantTL. The normalization of the SHAPE reactivity was done following previously reported methods (32,33). Briefly, in each lane, 2% of the most intensive bands were removed from the calculation and the next 8% were averaged. All reactivity values were divided by this value before further analysis.

For *in vivo* SHAPE the same *L. lactis* cells were used as for the RNP purifications. Thirty mL cultures were grown to OD 0.5, induced with nisin and concentrated down to 500 μl in fresh media. NMIA or 1M6 were added to a final concentration of 5 mM and the cell suspensions were incubated at 30°C until completion of the reactions (25 min or 5 min, respectively). The cells were then subjected to hot phenol and incubated at 65°C for 1 h with intermittent vortex followed by phenol-chloroform extraction and ethanol precipitation. Two μg of total RNA was used in each primer extension reaction and analyzed as described above.

Molecular dynamic (MD) simulations

MD simulations were performed using the GROMACS-2016 package (34). The cryo-EM structure of the intron complexed with the IEP and the exons (9) was used as the starting structure for the simulations. The unresolved regions of the protein, with the exception of Y252-L301 in Fingers/Palm, were constructed using MOE [Chemical Computing Group, Inc. (CCG, Inc.) (2016) Molecular Operating Environment (MOE), CCG, Inc., Montreal, QC, Canada]. For the RNA, the unresolved regions were modelled using the mc-fold mc-sym pipeline (35), with the exception of the truncated DIVb (Figure 6A). The 3D structures of the fragments thus obtained were incorporated into the RNA. For the protein, the missing segments were modeled as helices, turns or linear chains, and the distance between the terminal amino acid residues that closely fits the missing fragment was chosen for further energy minimization. Similarly, for the missing RNA residues, the fragment was modeled as a loop or linearly, and the structure with minimal difference in end-to-end distance was incorporated into the structure. The likely positions of magnesium ions were obtained by superposing the cryo-EM structure on the crystal structure of a group IIB intron (8) and were incorporated in the simulation. Due to larger time scales associated with equilibrating the divalent Mg^{2+} ions in solution, the system was simulated in a high concentration (1 M) of KCl to reproduce a similar ionic environment, in accord with thermodynamic stability studies of nucleic acids (36,37). The 3D periodic box size was $19.9 \times 19.9 \times 19.9$ nm³ containing the RNP complex, 1162 K^+, 484 Cl^- ions and 253 163 water molecules. The system was subjected to energy minimization to prevent any overlap of atoms. The simulations incorporated leap-frog algorithm with a 2 fs timestep to integrate equations of motion. The system was maintained at 300 K and 1 bar, using the velocity rescaling thermostat (38) and Parrinello-Rahman barostat (39), respectively. The long-range electrostatic interactions were calculated using the particle mesh Ewald (PME) (40) algorithm with a real space cut-off of 1.2 nm. LJ interactions were also truncated at 1.2 nm. TIP4P model (41) was used to represent the water molecules and the LINCS algorithm (42) was used to constrain the motion of hydrogen atoms bonded to heavy atoms. Coordinates of the RNA and protein were stored every 1 ps for further analysis. Five parallel simulations with the same starting conformation but different starting velocities were performed for 100 ns each, totaling 0.5 μs of production run. Additional details can be found in Supplementary Methods.

RESULTS

The group II intron Ll.LtrB RNP can dimerize in the pre-catalytic state

We used the same intein-based purification strategy as for our cryo-EM studies to isolate endogenously spliced Ll.LtrB group II intron RNP from the native host *L. lactis*. This spliced intron RNP (+A) with the ORF deleted (ΔORF) is active in substrate DNA binding, DNA integration and bottom-strand DNA cleavage (9,29). To enable direct comparison of RNPs in pre- and post-splicing states, we used a branch-point mutant with deletion of the bulged adenosine in DVI (ΔA) and a catalytic triad mutant (AGC>AAU, Triad) (Figure 2A). Both mutants are trapped in the precursor state (9,30). Whereas the ΔA construct is catalytically competent but lacks the adenosine nucleophile that initiates splicing and shows a small amount of splicing activity, the triad mutant is catalytically inert (Figure 2B), likely due to inability to bind metal ions at the active site (26). In contrast to the +A RNP, the mutant RNPs are unable to bind substrate DNA (Figure 2C) and they behave similarly in all of the following experiments.

We first sought to examine and compare the sizes of the RNPs. Interestingly, as shown by SEC, at temperatures below 30°C the ΔA and Triad RNPs had a larger size compared to the +A RNP (Figure 2D and Supplementary Figure S1A), which we previously interpreted as a conformational change upon splicing (29,30). However more recently, this size discrepancy had been attributed to precursor dimerization (two intron RNAs and two IEPs) by preliminary cryo-EM study on both precursor RNPs (Hong-Wei Wang and Nan Liu, personal communication). Strikingly, the dimer dissociated to monomer upon short incubation at 37°C and reverted to the dimer form after brief incubation on ice (Figure 2D). With extended incubation at physiological temperature (30°C), the precursor remained a dimer (Figure 2D and Supplementary Figure S1A). In spite of dimerization, RNA structure profiling showed that the precursor RNPs in the monomer and dimer forms share identical modification profiles, both within the intron RNA and the exons (Supplementary Figure S2). Dimerization is therefore likely to occur via the IEP. Accordingly, with an excess of competing *in vitro*-synthesized exons, the precursor RNP remained a dimer (Supplementary Figure S1B), making it unlikely that the dimerization occurs via the exons.

DNA oligonucleotide probing of the intron RNPs suggests exposure of DId after splicing

We then assessed the differences between +A and the precursor RNPs by probing the accessibility of ensembles of nucleotides in the RNA. We initially used thirty-six 25-nt DNA oligonucleotides anti-sense to segments of the intron RNA to specifically target the RNA through Watson–Crick base pairing, as described above (Figure 3A, Supplementary Figure S3A and Supplementary Table S1). Although the DNA probes used have different T_m values (Supplementary Table S1), the DNA oligonucleotides were shown to exhibit generally strong binding to *in vitro*-synthesized, unfolded Ll.LtrB intron RNA (data not shown) and the bind-ing of DNAs to the RNA was confirmed to be sequence-specific by RNase H cleavage at precisely the site of the anticipated DNA-RNA hybrid (Supplementary Figure S3B). Provided that the corresponding RNA regions are accessible, the DNA oligonucleotides are expected to bind the RNP through base pairing with unpaired nucleotides and/or compete for weakly paired regions. Binding of the probes to both +A and the precursor RNPs occurred mainly in DIV, which extends away from the RNP's main mass and appears largely unstructured (9) (Figure 3B, Supplementary Figure S3A and C). Nucleotide ensembles in domains II, III, V and VI generally showed similarly low levels of binding to corresponding anti-sense DNA oligonucleotides for both +A and the precursor RNPs (Figure 3C), suggesting the compact, well-folded nature of the RNPs.

In contrast, probes anti-sense to regions in DI, particularly to nucleotides encompassing EBS1, showed significantly increased binding to the +A RNP compared to the precursors. Moderate post-catalytic affinity increases were also found in regions encompassing DI(iii)a, β and β′ (Figure 3B and C and Supplementary Figure S3C). In addition to the 36 end-to-end DNA oligonucleotides, loop-specific tiling DNA oligonucleotides were used (Supplementary Figure S3A). These tiling oligonucleotides showed consistent results for key regions in DI, with the EBS2-specific probe showing different binding levels between the precursors and the spliced RNP (Supplementary Figure S3D). The elevated affinity to DNA oligonucleotides in the +A RNP suggests increased local nucleotide accessibility post-catalytically in those regions in DI. Compared to that of EBS1, the EBS2 region showed a more subtle increase in anti-sense DNA oligonucleotide binding for the spliced RNP (Figure 3B, Supplementary Figure S3C and D), suggesting that EBS1 is more accessible in the post-catalytic state. In addition, the binding patterns were largely unchanged by temperature-dependent stoichiometry differences (Supplementary Figure S3E).

RNA structure probing by SHAPE in the context of the cryo-EM structure

To probe the RNPs at nucleotide resolution, we applied SHAPE reagents, which form acylation adducts at 2′-hydroxyl positions on the backbone of exposed RNA nucleotides. Nucleotides with SHAPE adducts were located via primer extension, where reverse transcriptase disengages at modification sites, allowing identification of nucleotides with higher local flexibility (43). We first used 1M6 on the purified +A RNP described in the cryo-EM study (9). The modifications were located to the nucleotide level (Figure 4A and Supplementary Figure S4) and a SHAPE profile for the +A RNP was generated (Figure 4B). To validate the profile, we repeated the experiments for key regions in the RNA using BzCN and NMIA, which are SHAPE reagents with faster or slower chemistry, respectively, compared to 1M6. Our data showed consistency among profiles generated by different SHAPE reagents (Supplementary Figure S5). SHAPE revealed a spectrum of modification levels on different nucleotides, depending on their local nucleotide exposure. Strong, moderate and weak modifications (relative reactivity of 0.7–1.0, 0.4–0.7 and 0.1–0.4, respectively)
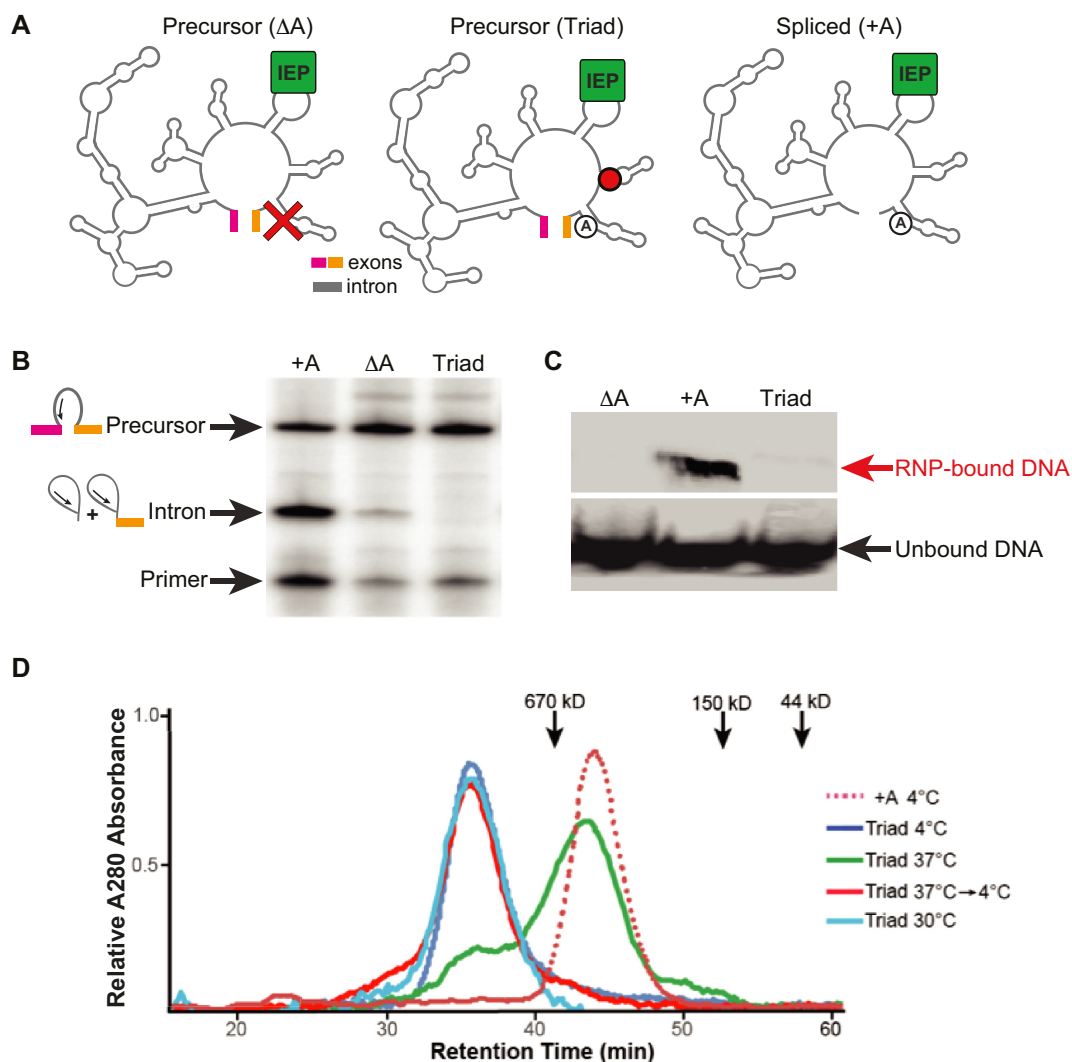
Figure 2. Group II intron Ll.LtrB RNP in pre- and post-catalytic states. (**A**) Spliced intron RNP (+A, right) and the two precursor RNPs. ΔA and Triad mutants are trapped in the pre-catalytic state by deleting the bulged branch-point adenosine in DVI or mutating the catalytic triad sequences in DV (AGC>AAU), respectively. (**B**) Splicing analysis by primer extension of total RNA isolated from the three constructs. The +A construct is capable of splicing whereas ΔA and Triad mutants generate little (ΔA) or no observable (Triad) excised intron. (**C**) Native gel showing substrate DNA binding to the purified RNPs. The +A RNP is able to bind the target DNA, whereas ΔA and Triad RNPs are not capable of DNA binding. (**D**) Size-Exclusion Chromatography (SEC) characterization of the Triad and +A RNPs. SEC was carried out at 25°C. The precursor RNP was incubated on ice or at 37°C for 15 min (with or without another 15 min on ice) or at 30°C for 3 h. Since equilibrium measurements were not done, the SEC data do not reflect the actual ratio of dimers and monomers in a given condition. The color code for the treatments is shown on the right. Retention times for globular protein standards are indicated with arrows on the top. Results for the ΔA RNP are similar to Triad (Supplementary Figure S1A).

were mostly confined to unpaired nucleotides including unprotected loops and domain junctions. The SHAPE profile is generally consistent with the canonical secondary structure. Nevertheless, unexpected modifications scattered in supposedly base-paired nucleotides in DIc2, DII, DIII, DIVa and DV (Figure 4B) and may be attributable to flexible configurations of the RNA (44). For instance, stem nucleotides adjacent to π in DII and those at the 3′ end of DIVa showed high SHAPE reactivity, likely reflecting the flexibility of the corresponding stem regions (Figure 4B).

The SHAPE profile was generally in accord with the packing of the RNA. For example, nucleotides in the basal part of DI are surrounded by neighboring RNA domains

and showed little SHAPE reactivity. In contrast, the distal part of DII protrudes from the tightly packed RNA and the nucleotides here were mostly SHAPE-sensitive (Figure 4B).

Conserved tertiary interactions are usually essential for the structure and function of large RNAs. Those interactions include Watson–Crick base pairings, tetraloop-receptor interactions and non-Watson–Crick interactions (Figure 1B). The SHAPE profile of the +A RNP was consistent with the cryo-EM structure in terms of tertiary interactions. For instance, nucleotides involved in direct tertiary interactions were mostly unmodified (Figures 1B and 4B), very likely due to protection by interacting distant motifs. For example, within DI, nucleotides forming α–α′ and
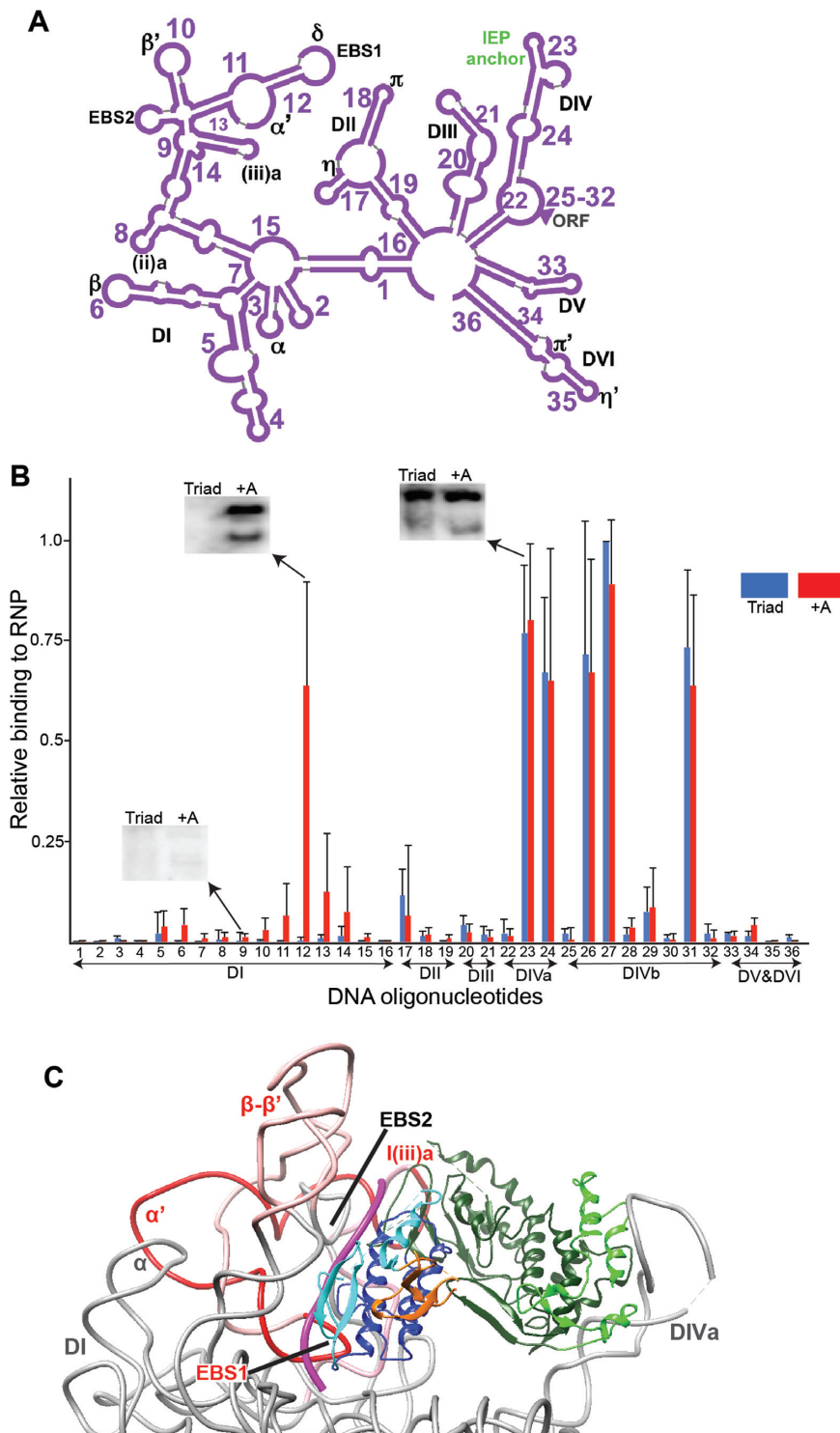
Figure 3. DNA oligonucleotide binding shows increased nucleotide accessibility in DId of the post-catalytic RNP. (**A**) Schematic showing RNA regions to which the DNA oligonucleotide probes are complementary. The DNA oligonucleotides were designed to cover the entire RNA from 5′ to 3′. Briefly, the ^{32}P-labeled oligonucleotides were co-incubated with the RNP at 37°C for 1 h and separated on a native 4% polyacrylamide gel for binding comparison. (**B**) Quantitative analysis of DNA oligonucleotide binding to Triad and +A RNPs. The binding experiment, exhibiting different levels of binding affinity in DI, was conducted three times and the data were averaged. Representative gels are shown for DNA oligonucleotides covering key regions of the RNA. Results of the ΔA RNP are similar to the Triad RNP and the binding patterns for all three RNPs are shown in Supplementary Figure S3C. The results using loop-specific tiling probes and data for experiments carried out at different temperatures are shown in Supplementary Figure S3D and E, respectively. (**C**) Spatial locations of the nucleotide accessibility gain in DId as reflected by DNA oligonucleotide probing. RNA nucleotides with moderately (pink) and dramatically (red) increased accessibility after splicing are highlighted. Key motifs are indicated. The magenta tube denotes the retained mRNA. For the IEP, light green denotes NTE, dark green denotes the Fingers/Palm domain, dark blue denotes the Thumb domain, cyan denotes the DBD and orange denotes the EN domain.
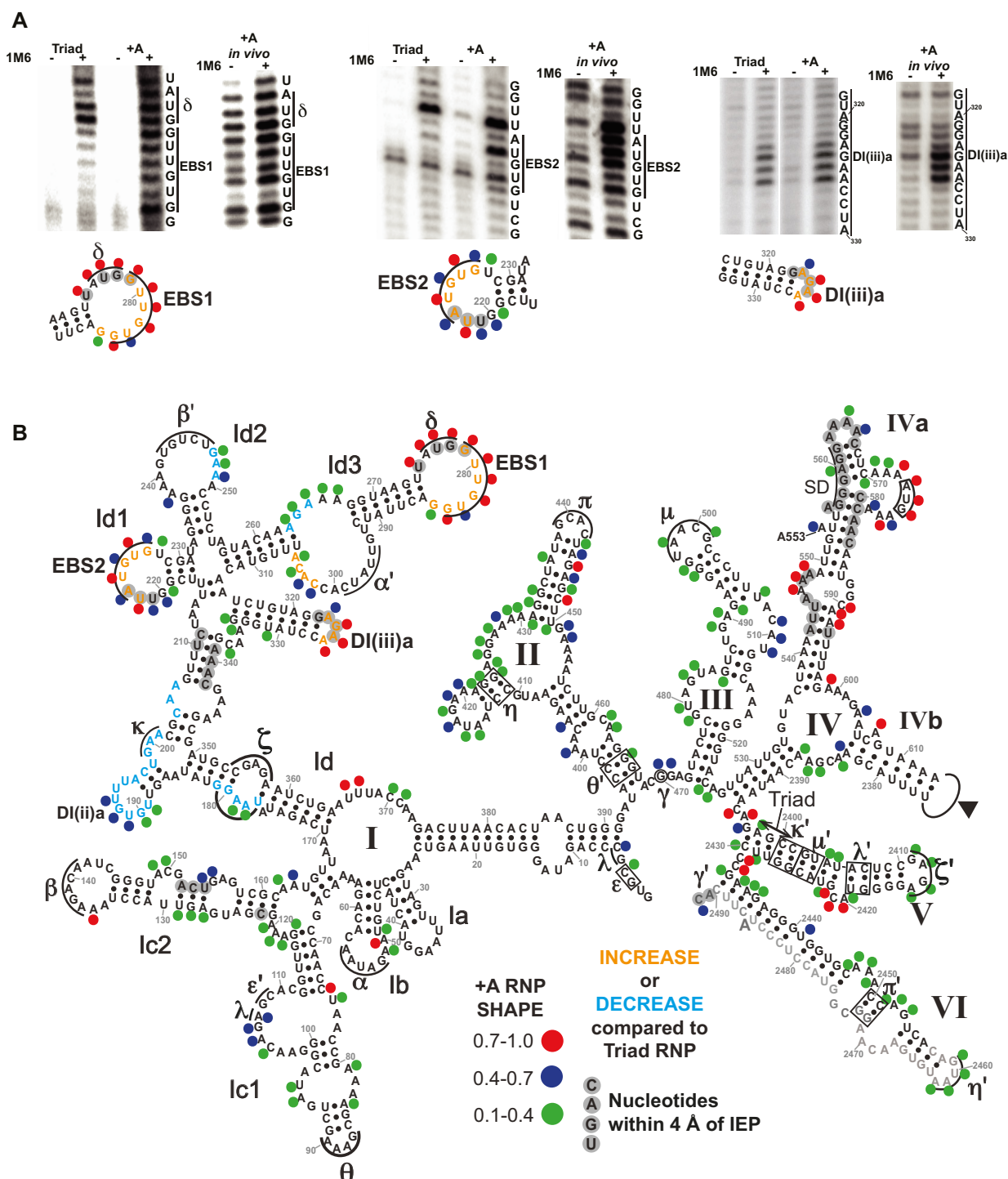
Figure 4. SHAPE profile of the +A RNP and comparison to the Triad RNP. (**A**) Representative gels showing *in vitro* SHAPE reactivity of the EBSs and DI(iii)a nucleotides in the +A RNP and the Triad RNP, as well as data for the +A RNP *in vivo*. Structures and colored dots are as in panel B. (**B**) Secondary structure of the Ll.LtrB intron showing *in vitro* SHAPE profile of the +A RNP and the comparison to that of the Triad RNP. Colored dots are assigned to nucleotides to denote the levels of SHAPE modifications. SHAPE reactivity comparison is shown by coloring nucleotides with either orange or cyan in the +A profile to represent increase or decrease compared to that of the Triad RNP, respectively. DVI nucleotides in light gray denote the site where the +A RNP was not profiled. The truncated DIVb is mostly omitted. Representative data for all three RNPs are shown in Supplementary Figure S4. Shaded nucleotides are postulated protein-interaction sites within an arbitrary 4-Å distance from the protein in the cryo-EM structure (9) (PDB: 5g2x).

β–β′ Watson–Crick pairings showed no SHAPE reactivity, but nucleotides adjacent to them exhibited sensitivity (Figures 4B and 5A). In addition, nucleotides involved in θ–θ′ and π–π′ tetraloop-receptor interactions between DIc1 and DII, and between DII and DVI, respectively, exhibited no SHAPE chemistry (Figures 4B and 5B), suggesting strong interactions (Table 1). It is likely that such interactions are structurally essential by acting as clamps to stabilize the RNP throughout the precursor-spliced intron transition.

However, numerous interactions including η–η′, between DII and DVI (Figures 4B and 5B), and μ–μ′, between DIII and DV (Figure 4B), showed weak SHAPE modifications in the loop or receptor nucleotides or both, suggesting relatively more fragile or dynamic associations. Interestingly, despite the fact that the ligated exons are retained in the +A RNP and seem to base pair with the EBSs in the cryo-EM structure (9), the exon binding sequences (EBSs) were highly sensitive to SHAPE in the +A RNP (Figure 4), suggesting weakened base pairing with IBS nucleotides of the retained mRNA. Another SHAPE-sensitive region was the start codon (AUG) loop (A571-A579) in DIVa, which is required to be exposed for ribosome access (Figure 4B).

In addition, SHAPE data corroborated the cryo-EM model with respect to protein binding. Most of the nucleotides proximal to the protein showed little or no SHAPE reactivity (Figure 4B, gray shading). For instance, nucleotides at the IEP anchor site in DIVa, which interact with the IEP's NTE with high affinity (45), exhibited little or no SHAPE reactivity (Figures 4B and 5C). Protease K digestion of the IEP resulted in dramatically increased SHAPE sensitivity at the anchor site (Supplementary Figure S6), confirming that the protection is conferred by protein-RNA interaction. However, there were a few exceptions. For instance, nucleotides A547-A548 are close to the IEP's NTE but showed very strong SHAPE reactivity, suggesting flexibility of those nucleotides (Figure 4B). Also, nucleotides U221-A223, adjacent to or within EBS2, as well as the δ nucleotides, showed strong SHAPE reactivity despite close proximity to the IEP.

Another interesting exception is DI(iii)a, which is a group IIA intron-specific stem-loop previously suggested as an optimal motif for protein binding (19). It is arranged in close proximity to an auxiliary RNA-binding element (insertion 4a) in the Fingers/Palm of the IEP, which was largely unresolved but appeared to extend towards the DI(iii)a stem-loop (9). However, SHAPE data showed that the four nucleotides in the DI(iii)a loop were strongly modified both *in vitro* and *in vivo,* indicative of high flexibility and/or lack of direct protection from the IEP for the loop nucleotides. It is possible that interactions are transient or that this auxiliary RNA-binding motif binds to DI(iii)a primarily via the stem region, possibly at the groove facing away from the RNA mass (Figure 5D). Together, the SHAPE data provide structural insights into the configuration of the RNA architecture as well as RNA–protein interactions.

In-cell SHAPE suggests structural consistency *in vivo*

In a living cell, RNA structures are constantly under the influence of local solution conditions, interacting cellular factors and nascent transcription and translation events. Thus RNA structures *in vivo* are subject to more complexity and undergo shifts in their structure compared to that *in vitro.* To address this question, in-cell SHAPE was performed on the cells used to produce +A or the precursor RNPs to examine the modification profiles *in vivo* (46). Probing was done with SHAPE reagents 1M6 and NMIA, which readily permeated the *L. lactis* cell membrane and reacted with the RNA at fast or slow rates, respectively. Our results showed that the landscape of the SHAPE profiles stays largely unchanged *in vivo* in key regions (Figure 4A and Supplementary Figure S7). In addition, the precursor RNPs showed *in vivo-in vitro* consistency, and SHAPE profiles generated from stationary phase and aerobically cultured cells were similar as well (data not shown), implying that the RNP maintains its stable architecture in different cellular circumstances.

Structure rearrangements accompanying splicing

Whereas the two precursor RNPs showed similar SHAPE profiles to that of the +A RNP for most of the nucleotides (Figure 4 and Supplementary Figure S4A–D), confined SHAPE profile differences in discrete regions in DId between both the precursors and spliced RNPs were observed (Figure 4 and Supplementary Figure S4E and F). In the post-catalytic RNP, EBS1, EBS2 and the DI(iii)a loop nucleotides were more sensitive to SHAPE compared to the precursors (Figure 4). The pre-catalytic RNP showed little or no SHAPE reactivity at the EBSs. In contrast, the SHAPE profiles suggested that both of the EBSs become less protected by base pairing after splicing, in spite of the apparent proximity to the IBSs in the retained mRNA (Figure 1E). Interestingly, the δ nucleotides, which are an extension of EBS1, showed heavy modifications at both catalytic stages. In addition, the four nucleotides in the DI(iii)a loop that is in close proximity to the IEP at insertion 4a, showed increased SHAPE reactivity in the +A RNP (Figures 4 and 5D), implying a change in binding of the protein. Together, the results suggest disengagements at the action interface with respect to both RNA–RNA and RNA–protein interactions.

Furthermore, the active site underwent changes in the +A RNP. U194 and U195 in DI(ii)a interact with A206 and A205, respectively, forming a loop encompassing κ, according to the cryo-EM structure (9). The nucleotides involved in the UU–AA pairing were accordingly inert to SHAPE (Figure 4B). However, these UU-AA pairing nucleotides were SHAPE-sensitive in the pre-catalytic RNP (Figure 4B), suggesting that the κ loop forms during splicing and is a characteristic of the post-catalytic RNP. In addition, numerous nucleotides at the DI(ii)a loop, as well as those in the κ and ζ motifs showed decreased SHAPE reactivity (Figure 4B). Meanwhile, nucleotides forming the μ–μ′ interaction between DIII and DV showed no change in SHAPE profile, implying that the μ–μ′ interaction plays a role in stabilizing the DV active site throughout the splicing reaction. More generally, the strong tertiary interactions, including α–α′, β–β′, θ–θ′ and π–π′, are mostly unchanged in the transition between the two states, suggesting that they maintain stability of the RNP structure throughout splicing (Table 1).
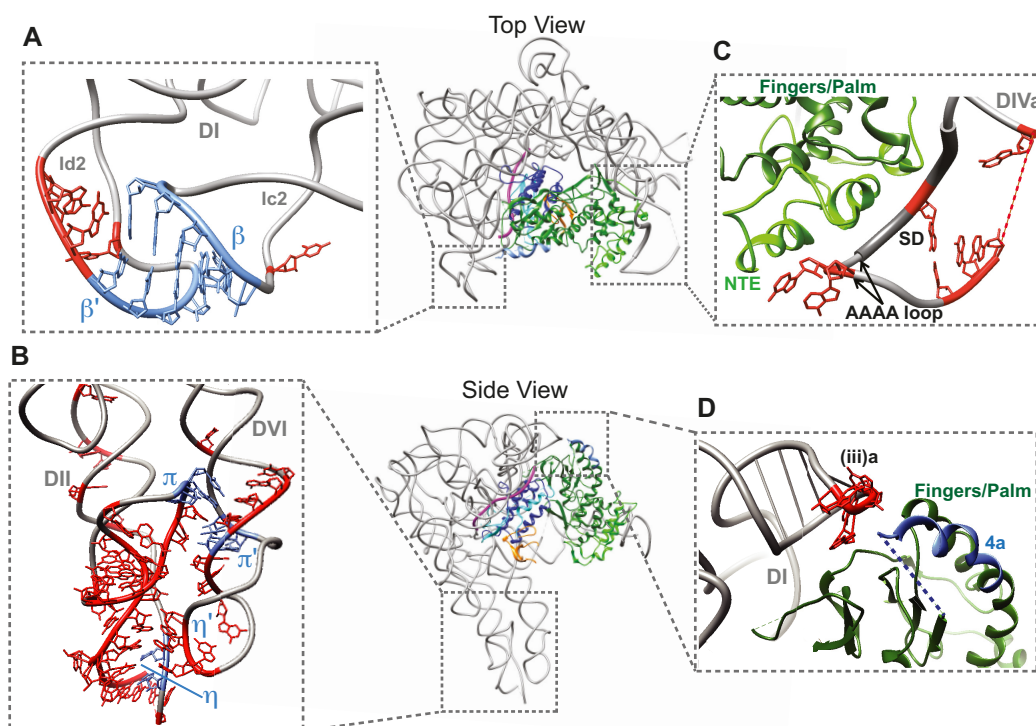
Figure 5. SHAPE profile of +A RNP is largely in accord with cryo-EM structure (9) (PDB: 5g2x). Bases are shown for nucleotides involved in key tertiary pairings and/or those with SHAPE reactivity (colored in red). (**A**) Tertiary base pairing β–β′ (blue) is protected from SHAPE chemistry while adjacent nucleotides are modified. (**B**) SHAPE modifications near the two tetraloop-receptor interactions between DII and DVI. The π–π′ interaction nucleotides (upper pair) show no SHAPE reactivity, whereas nucleotides involved in the η–η′ interaction (lower pair) show sensitivity to SHAPE. (**C**) IEP anchor site in DIVa is largely protected by protein interaction. The Shine-Dalgarno (SD) sequence is shown as a bold tube. The AAAA loop is indicated by black arrows. (**D**) DI(iii)a is sensitive to SHAPE despite being vicinal to the Fingers/Palm domain of the IEP. Insertion 4a is shown in blue with its unresolved part denoted as a dotted line.

Table 1. Levels of SHAPE reactivity of nucleotides at key tertiary interaction motifs in the two states

	Interacting domains	Pre-catalytic SHAPE	Post-catalytic SHAPE change
EBS1-IBS1	DId3 & exon1	none	Increased
EBS2-IBS2	DId1 & exon1	none	Increased
α–α′	DIb & DId3	none	Unchanged
β–β′	DIc2 & DId2	none	Unchanged
γ-γ′	J2/3 & 3′ splice site	moderate	Unchanged
δ-δ′	DId3 & exon2	moderate	Unchanged
ε–ε′	5′ splice site & DIc1	none	Unchanged
ζ–ζ′	DId & DV	strong	Decreased
η–η′	DII & DVI	weak	Unchanged
θ–θ′	DIc1 & DII	none	Unchanged
κ–κ′	DId & DV	strong	Decreased
λ–λ′	DId, 5′ splice site & DV	moderate	Unchanged
μ–μ′	DII & DV	weak	Unchanged
π–π′	DII & DVI	none	Unchanged

Levels of pre-catalytic and post-catalytic SHAPE reactivity are categorized as shown in Figure 4B.

The intron RNA has a dynamic scaffold that is stabilized by tertiary and protein interactions

We next carried out molecular dynamics (MD) simulations to investigate the role that long-range tertiary interactions and RNA–protein associations play in stabilization of the RNP scaffold. The MD simulation model was constructed based on the cryo-EM structure of the +A RNP (9) (Supplementary Figure S8A). The root-mean square fluctuations (RMSF) were measured for the intron RNA (Figure 6A), the IEP (Supplementary Figure S8B) and the exons (Sup-

plementary Figure S8C). The IEP and the exons showed general stability and small range of motion (Supplementary Figure S8B and C). In comparison, the intron RNA exhibited large motions, especially towards the distal nucleotides of the domains (Figure 6A). Despite the highly fluctuating RNA domains, the movements of the domains were highly correlated via tertiary or RNA–protein interactions (Figure 6B), as supported by the SHAPE studies (Figure 4). For example, the α–α′ interaction anchored DIb to the internal loop in DId3, resulting in high motion correlation
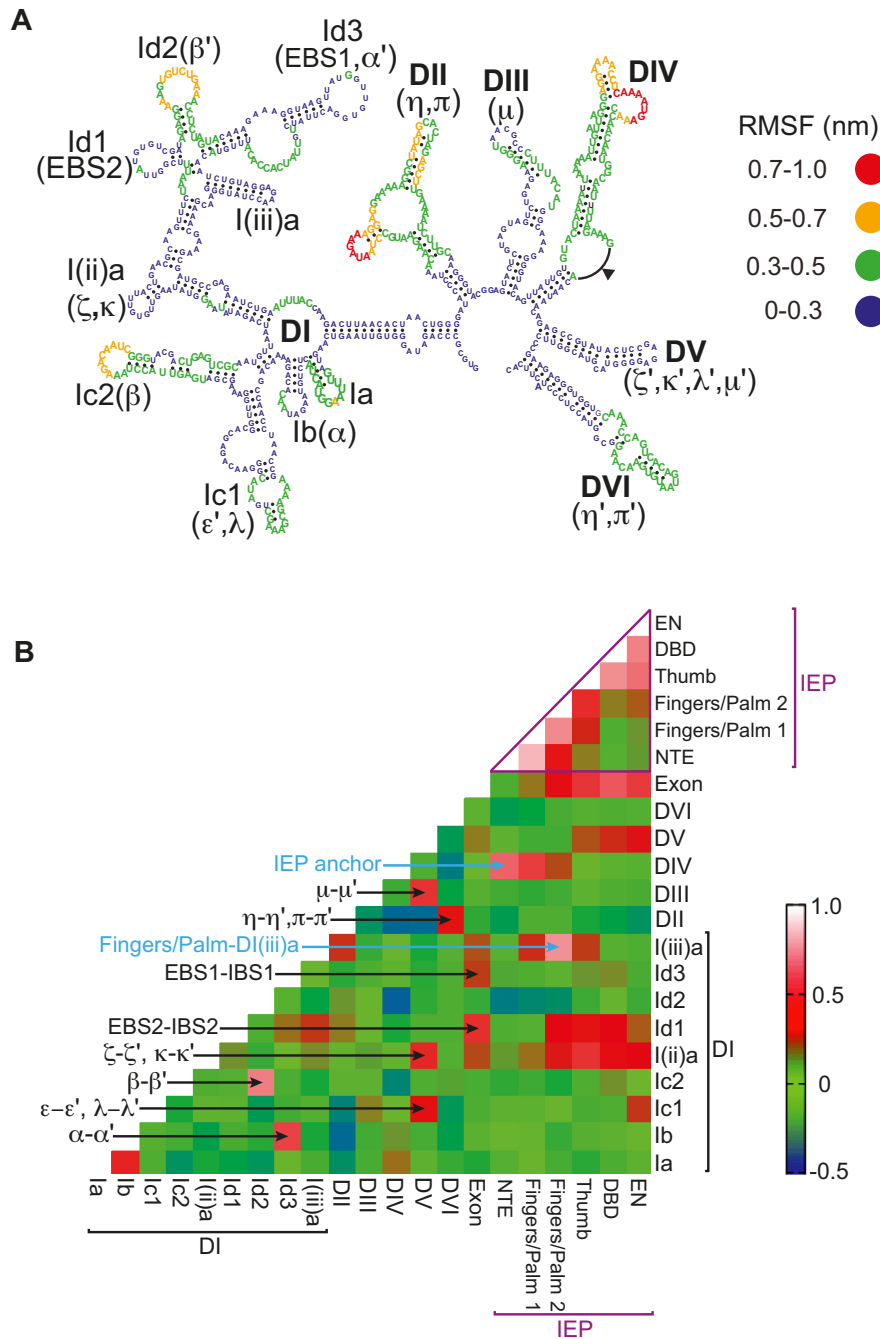
Figure 6. Molecular dynamic simulations of the +A RNP. (**A**) Domain fluctuations of the intron RNA highlighting nucleotides with moderate (green), high (orange) and very high (red) fluctuations. Nucleotides with RMSF below 0.3 are shown in dark blue. Tertiary interaction motifs and the domains are noted. (**B**) Fluctuation correlations among domains in the intron RNA, the exons and the IEP. Pale-red (1.0) and blue-green (–0.5) colors indicate high and low ends of correlation values, respectively. Fingers/Palm domain is divided at the insertion 4a truncation (Y252-L301 as shown in Supplementary Figure S8A), with Fingers/Palm 1 adjacent to the NTE and Fingers/Palm 2 close to the Thumb domain at the action interface. The black and blue arrows indicate representative tertiary and protein interactions, respectively.

between the α–α′ and DId3. Similarly, DII and DVI, both highly flexible, demonstrated high motion correlation with each other by joining via π–π′ and η–η′ interactions (Figure 6). In addition, DIc2 and DId2, which sandwich the exons from the side opposite to the IEP, showed high fluctuations, but exhibited coordinated motions via the β–β′ interaction (Figure 6). At the core of the RNP complex, DV showed

correlated motions with DI(ii)a via κ–κ′ and ζ–ζ′ interactions, with DIc1 via the ε–ε′ and λ–λ′ interactions and with DIII via μ–μ′ interaction (Figure 6B). With the persistent network of interactions, these locations exhibited uniformly low fluctuations (Figure 6A). With respect to RNA–protein interaction sites, the NTE and DIVa showed high fluctuation as well as high correlations, suggesting that the IEP an-

chor site is structurally flexible, in spite of the high-affinity binding (Figure 6 and Supplementary Figure S8B). Overall, the simulation results were in agreement with the SHAPE analyses, suggesting that the group II intron RNA is able to maintain its scaffold through stabilization from tertiary interactions and IEP associations while still exhibiting considerable dynamics, which are likely important for its different functionalities.

DISCUSSION

The use of SHAPE data in combination with the available cryo-EM structure not only provided cross-validation, but also added essential information, including interactions and dynamics, to the structure (Figure 5). The oligonucleotide probing data provided additional information regarding gross accessibility of the nucleotides, including those in base-paired RNA regions. The results showed that the group II intron Ll.LtrB RNP has a well-folded, compact and generally stable RNA architecture in the context of functional transitions as well as in the cellular environment. The IEP anchor site and various tertiary interactions remain fixed throughout the splicing transition and there is no indication for large conformational changes. In contrast, DId in the RNA, which is interacting with the IEP at the action interface, undergoes structural rearrangements accompanying splicing.

Persistent protein-RNA and tertiary RNA–RNA interactions maintain the RNP scaffold

The high-affinity nature of the DIVa–LtrA interaction was demonstrated by a K_d of 0.25 pM (45). Deletion of DIVa not only significantly decreased the binding to the IEP and greatly hindered the splicing of the intron (45), but also strongly inhibited retromobility (47). Consistent with these findings, our data show strong and stable RNA–protein contacts at the IEP anchor site in both catalytic states. The tight binding at the anchor site is not only important for maintaining a stable RNP structure, but also for clamping the IEP to the RNA as it functions at the action interface.

Furthermore, stability of DI of the intron RNA is strengthened by numerous tertiary interactions such as α–α′ and β–β′. Other tertiary interactions that remain unchanged in the two states include ε–ε′, θ–θ′, π–π′ and DIa containing the conserved T-loop, which are also likely to maintain the RNA scaffold. It is worth noting that while those long-range pairings appear to be stable during splicing, SHAPE profile changes were observed in adjacent regions. Accordingly, MD simulations indicated that domains joined by tertiary interactions, while maintaining the strong bonds, exhibited highly correlated structural versatility (Figure 6B).

In contrast, SHAPE reactivity for nucleotides at the κ–κ′ and ζ–ζ′ interactions decrease in the post-catalytic state (Table 1), with implications for altered tertiary interactions as well as adjustments of DV at the active site. It is interesting to note that the π–π′ interaction between DII and DVI, which had previously been proposed to mediate the transition between the two splicing steps by an on/off mechanism (8), stayed unchanged after splicing. While we do not rule out the possibility that π temporarily disengages from π′ between the two transesterification steps, our data suggest that it may be more likely that the π–π′ interaction persists throughout the splicing process and the toggling of DVI is achieved by the dynamic movements of DII as shown in the MD simulations.

Importantly, *in vivo* SHAPE showed structural consistency of the RNPs in the living cell with purified RNPs, suggesting stable folding. Nevertheless, quantitative differences in SHAPE profiles were identified (Supplementary Figure S7), possibly due to effects of the cellular environment and/or, in the case of the +A cells, heterogeneity of spliced and unspliced species in the cell (Figure 2B).

Although the fluctuations of the nucleotides in the MD simulation were generally consistent with their SHAPE profiles, there were some differences between the two datasets, especially at tertiary and protein interaction sites. This likely reflects the technical hurdle of predicting interactions at regions entangled in higher-order structures upon which further development of computational methods may shed light.

Structural changes accompanying splicing

A 1:1 protein-RNA ratio was demonstrated in the cryo-EM structure of the spliced Ll.LtrB intron (9). This is also the case for group II intron GsI-IIC RT bound to template/primer substrate and for telomerase RTs (TERTs) (28,48,49). However, there have also been biochemical and biophysical reports of IEP dimers, in many cases with a ratio of two proteins to one intron RNA, although the underlying biological significance is not known (29,50–52). The possibility of transient dimerization of the IEP or the RNP has been raised as well (53). Whereas the spliced intron always behaved as a monomer, we indeed observed a temperature-dependent change in monomer (1:1) versus dimer (2:2) formation for both of our precursor RNPs (Figure 2D and Supplementary Figure S1A). Such inconsistency with the previously reported 2:1 ratio (two proteins to one intron RNA) may reflect the preference of the RNP to form dimers that dissociate in the splicing process. Interestingly, group II intron RNP in the monomer and dimer forms shared similar SHAPE profiles. It is thus likely that the stoichiometry change is not dependent upon the structure of the RNA. It is also unlikely that the monomers join via exon interactions, given the observation that the SHAPE profiles of the exons in the precursor RNPs were unchanged at different temperatures (Supplementary Figure S2), as well as that exon competition did not disrupt dimerization (Supplementary Figure S1B). Thus it is more likely that dimers form mainly via the IEPs. Similar temperature-dependent stoichiometry changes of proteins had been reported (54,55). Such dimerization in the precursor state may affect the configuration of the IEP and hinder splicing so that its initiation can be down-regulated beyond the rate-limiting first step of splicing.

Interestingly, most of the post catalytic SHAPE increases were located at the action interface of the RNA–protein interaction, including nucleotides at EBS1 and EBS2, at DI(iii)a and adjacent to α′. These results indicate that both RNA–RNA and RNA–protein interactions at the action
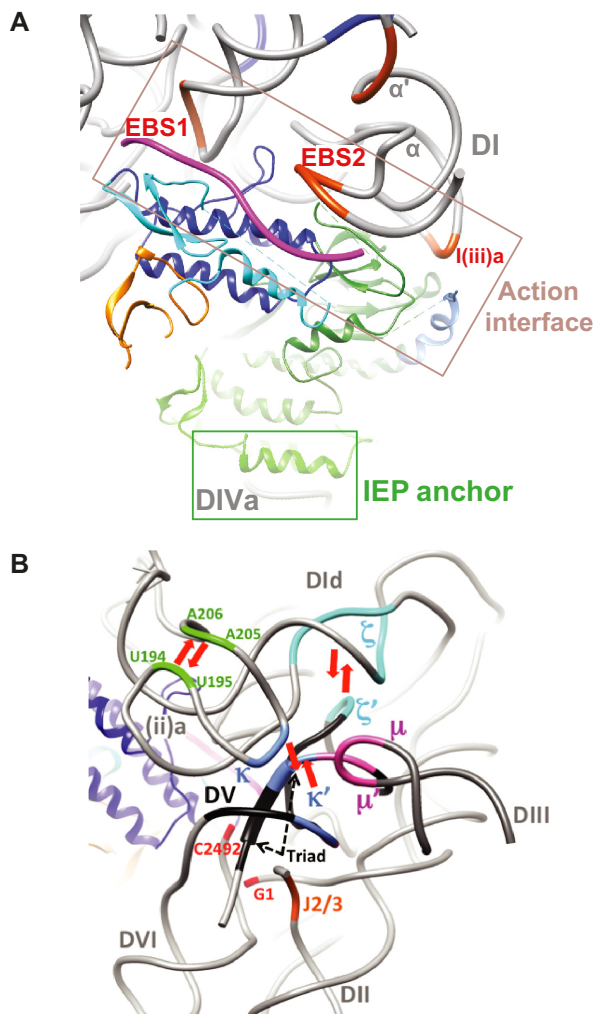
Figure 7. Putative structural rearrangements in transition between the two catalytic states. (**A**) Key regions with gained SHAPE reactivity in the +A RNP are located at the action interface of the RNP. Retained mRNA is shown in magenta. Light and dark green represent the NTE and the Fingers/Palm domain, respectively. Blue denotes the Thumb domain, cyan denotes the DNA binding domain and orange denotes the DNA endonuclease domain. The boxes indicate the action interface (merlot) and the IEP anchor (green). (**B**) Spatial arrangements of κ–(ii)a–ζ region and the adjacent DVI at the active site. In the post-catalytic RNP the κ loop has strengthened tertiary interactions with κ′, indicated by red arrows. ζ–ζ′ also has strengthened interactions as do A205-A206 and U195-U194 (green). DV is shown in black. Light blue, cyan and dark pink represent the κ–κ′, ζ–ζ′ and μ–μ′ interactions, respectively. The catalytic triad nucleotides are shown as a bold tube. 5′ and 3′ ends are shown in red and J2/3 is shown in orange.

interface undergo structural rearrangements accompanying splicing (Figure 7A). EBS-IBS interactions are essential for group II intron splicing and mobility (3). After splicing, the ligated exons, no longer a part of the intron RNA, are retained in the RNP rather than being promptly removed (9) (Figure 7A). However, the SHAPE profiles suggest weakening of the EBS-IBS base pairings. These base pairings likely become disengaged with time, concomitant with structural changes of the protein that alter the local architecture of the action interface. Interestingly, DNA oligonucleotide probing suggested that EBS1 dramatically gains accessibility af-

ter splicing (Figure 3), consistent with a previously reported role of a similar EBS as an exon sensor in a spliced IIC intron modified from *Oceanobacillus iheyensis* (27). Throughout these transitions, the intron RNP remains stabilized by IEP binding at DIVa and by multiple long-range tertiary interactions of the intron RNA. Pre- and post-catalytic states have also been reported for crystal structures of the *O. iheyensis* intron (24,26). Although direct comparison to the group IIA Ll.LtrB intron RNP is difficult given the differences between the two intron complexes, large conformational changes are also unlikely between the two states of the *O. iheyensis* intron.

Our results also indicated rearrangements at the active site. The region encompassing κ–(ii)a–ζ had been reported to be crucial for the proper folding of DI (56). In addition, κ–κ′ and ζ–ζ′ interactions link this region with DV containing the catalytic triad (Figures 1B and 7B). Interestingly, our SHAPE data suggested that, compared to the pre-catalytic RNP, the spliced RNP possessed the characteristic of formation of a loop encompassing κ, and strengthening of κ–κ′ and ζ–ζ′ interactions, where nucleotides became less SHAPE-sensitive in the post-catalytic state (Figure 4B). Our results suggest the possibility that this region acts as a structural modulator that rearranges during splicing to change the conformation of the catalytic triad. In addition, such rearrangement of DV may be important in repositioning the active site for DNA integration.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Lambowitz,A.M. and Belfort,M. (2015), *Mobile DNA III.* Microbiol Spectr. ASM Press, Vol. **3**, pp. 1209–1236.
2. Novikova,O. and Belfort,M. (2017) Mobile group II introns as ancestral eukaryotic elements. *Trends Genet.: TIG*, **33**, 773–783.
3. Lambowitz,A.M. and Zimmerly,S. (2011) Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.*, **3**, a003616.

4. 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.

5. Pyle,A.M. (2010) The tertiary structure of group II introns: implications for biological function and evolution. *Crit. Rev. Biochem. Mol. Biol.*, **45**, 215–232.

6. Harris-Kerr,C.L., Zhang,M. and Peebles,C.L. (1993) The phylogenetically predicted base-pairing interaction between alpha and alpha' is required for group II splicing in vitro. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 10658–10662.

7. Michel,F. and Ferat,J.L. (1995) Structure and activities of group II introns. *Ann. Rev. Biochem.*, **64**, 435–461.

8. Robart,A.R., Chan,R.T., Peters,J.K., Rajashankar,K.R. and Toor,N. (2014) Crystal structure of a eukaryotic group II intron lariat. *Nature*, **514**, 193–197.

9. Qu,G., Kaushal,P.S., Wang,J., Shigematsu,H., Piazza,C.L., Agrawal,R.K., Belfort,M. and Wang,H.W. (2016) Structure of a group II intron in complex with its reverse transcriptase. *Nat. Struct. Mol. Biol.*, **23**, 549–559.

10. Zimmerly,S., Hausner,G. and Wu,X. (2001) Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Res.*, **29**, 1238–1250.

11. Blocker,F.J., Mohr,G., Conlan,L.H., Qi,L., Belfort,M. and Lambowitz,A.M. (2005) Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. *RNA*, **11**, 14–28.

12. Mohr,G., Perlman,P.S. and Lambowitz,A.M. (1993) Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Res.*, **21**, 4991–4997.

13. Singh,R.N., Saldanha,R.J., D'Souza,L.M. and Lambowitz,A.M. (2002) Binding of a group II intron-encoded reverse transcriptase/maturase to its high affinity intron RNA binding site involves sequence-specific recognition and autoregulates translation. *J. Mol. Biol.*, **318**, 287–303.

14. Zoschke,R., Nakamura,M., Liere,K., Sugiura,M., Borner,T. and Schmitz-Linneweber,C. (2010) An organellar maturase associates with multiple group II introns. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 3245–3250.

15. Mohr,G. and Lambowitz,A.M. (2003) Putative proteins related to group II intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants. *Nucleic Acids Res.*, **31**, 647–652.

16. Schmitz-Linneweber,C., Lampe,M.K., Sultan,L.D. and Ostersetzer-Biran,O. (2015) Organellar maturases: A window into the evolution of the spliceosome. *Biochim. Biophys. Acta*, **1847**, 798–808.

17. Sultan,L.D., Mileshina,D., Grewe,F., Rolle,K., Abudraham,S., Glodowicz,P., Niazi,A.K., Keren,I., Shevtsov,S., Klipcan,L. *et al.* (2016) The reverse Transcriptase/RNA maturase protein MatR is required for the splicing of various group II introns in brassicaceae mitochondria. *Plant cell*, **28**, 2805–2829.

18. Cui,X., Matsuura,M., Wang,Q., Ma,H. and Lambowitz,A.M. (2004) A group II intron-encoded maturase functions preferentially in cis and requires both the reverse transcriptase and X domains to promote RNA splicing. *J. Mol. Biol.*, **340**, 211–231.

19. Matsuura,M., Noah,J.W. and Lambowitz,A.M. (2001) Mechanism of maturase-promoted group II intron splicing. *EMBO J.*, **20**, 7259–7270.

20. Zimmerly,S., Guo,H., Perlman,P.S. and Lambowitz,A.M. (1995) Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell*, **82**, 545–554.

21. San Filippo,J. and Lambowitz,A.M. (2002) Characterization of the C-terminal DNA-binding/DNA endonuclease region of a group II intron-encoded protein. *J. Mol. Biol.*, **324**, 933–951.

22. Noah,J.W. and Lambowitz,A.M. (2003) Effects of maturase binding and Mg2+ concentration on group II intron RNA folding investigated by UV cross-linking. *Biochemistry*, **42**, 12466–12480.

23. 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.

24. 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.

25. Toor,N., Rajashankar,K., Keating,K.S. and Pyle,A.M. (2008) Structural basis for exon recognition by a group II intron. *Nat. Struct. Mol. Biol.*, **15**, 1221–1222.

26. Chan,R.T., Robart,A.R., Rajashankar,K.R., Pyle,A.M. and Toor,N. (2012) Crystal structure of a group II intron in the pre-catalytic state. *Nat. Struct. Mol. Biol.*, **19**, 555–557.

27. Costa,M., Walbott,H., Monachello,D., Westhof,E. and Michel,F. (2016) Crystal structures of a group II intron lariat primed for reverse splicing. *Science*, **354**, aaf9258.

28. Stamos,J.L., Lentzsch,A.M. and Lambowitz,A.M. (2017) Structure of a thermostable group II intron reverse transcriptase with template-primer and its functional and evolutionary implications. *Mol. Cell*, **68**, 926–939.

29. Gupta,K., Contreras,L.M., Smith,D., Qu,G., Huang,T., Spruce,L.A., Seeholzer,S.H., Belfort,M. and Van Duyne,G.D. (2014) Quaternary arrangement of an active, native group II intron ribonucleoprotein complex revealed by small-angle X-ray scattering. *Nucleic Acids Res.*, **42**, 5347–5360.

30. Huang,T., Shaikh,T.R., Gupta,K., Contreras-Martin,L.M., Grassucci,R.A., Van Duyne,G.D., Frank,J. and Belfort,M. (2011) The group II intron ribonucleoprotein precursor is a large, loosely packed structure. *Nucleic Acids Res.*, **39**, 2845–2854.

31. Laederach,A., Das,R., Vicens,Q., Pearlman,S.M., Brenowitz,M., Herschlag,D. and Altman,R.B. (2008) Semiautomated and rapid quantification of nucleic acid footprinting and structure mapping experiments. *Nat. Protoc.*, **3**, 1395–1401.

32. Low,J.T. and Weeks,K.M. (2010) SHAPE-directed RNA secondary structure prediction. *Methods*, **52**, 150–158.

33. Cordero,P., Kladwang,W., VanLang,C.C. and Das,R. (2012) Quantitative dimethyl sulfate mapping for automated RNA secondary structure inference. *Biochemistry*, **51**, 7037–7039.

34. Hess,B., Kutzner,C., van der Spoel,D. and Lindahl,E. (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, **4**, 435–447.

35. Parisien,M. and Major,F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51–55.

36. Bizarro,C.V., Alemany,A. and Ritort,F. (2012) Non-specific binding of Na+ and Mg2+ to RNA determined by force spectroscopy methods. *Nucleic Acids Res.*, **40**, 6922–6935.

37. Tan,Z.J. and Chen,S.J. (2007) RNA helix stability in mixed Na+/Mg2+ solution. *Biophys J.*, **92**, 3615–3632.

38. Bussi,G., Donadio,D. and Parrinello,M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.*, **126**, 014101.

39. Berendsen,H.J., Postma,J.V., Van Gunsteren,W.F., DiNola,A.R.H.J. and Haak,J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.

40. Darden,T., York,D. and Pedersen,L. (1993) Particle mesh Ewald: an Nslog(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.

41. Horn,H.W., Swope,W.C., Pitera,J.W., Madura,J.D., Dick,T.J., Hura,G.L. and Head-Gordon,T. (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.*, **120**, 9665–9678.

42. Hess,B., Bekker,H., Berendsen,H.J.C. and Fraaije,J.G.E.M. (1997) LINCS: a linear constraint solver for molecular simulations. *J. Comp. Chem.*, **18**, 1463–1472.

43. Mauger,D.M. and Weeks,K.M. (2010) Toward global RNA structure analysis. *Nat. Biotechnol.*, **28**, 1178–1179.

44. Homan,P.J., Favorov,O.V., Lavender,C.A., Kursun,O., Ge,X., Busan,S., Dokholyan,N.V. and Weeks,K.M. (2014) Single-molecule correlated chemical probing of RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 13858–13863.

45. Wank,H., SanFilippo,J., Singh,R.N., Matsuura,M. and Lambowitz,A.M. (1999) A reverse-transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. *Mol. Cell*, **4**, 239–250.

46. Smola,M.J., Rice,G.M., Busan,S., Siegfried,N.A. and Weeks,K.M. (2015) Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.*, **10**, 1643–1669.

47. D'Souza,L.M. and Zhong,J. (2002) Mutations in the Lactococcus lactis Ll.LtrB group II intron that retain mobility in vivo. *BMC Mol. Biol.*, **3**, 17.

48. Bryan,T.M., Goodrich,K.J. and Cech,T.R. (2003) Tetrahymena telomerase is active as a monomer. *Mol. Biol. Cell.*, **14**, 4794–4804.

49. Jiang,J., Chan,H., Cash,D.D., Miracco,E.J., Ogorzalek Loo,R.R., Upton,H.E., Cascio,D., O'Brien Johnson,R., Collins,K., Loo,J.A. *et al.* (2015) Structure of Tetrahymena telomerase reveals previously unknown subunits, functions, and interactions. *Science*, **350**, aab4070.

50. Zhao,C. and Pyle,A.M. (2016) Crystal structures of a group II intron maturase reveal a missing link in spliceosome evolution. *Nat. Struct. Mol. Biol.*, **23**, 558–565.

51. Saldanha,R., Chen,B., Wank,H., Matsuura,M., Edwards,J. and Lambowitz,A.M. (1999) RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. *Biochemistry*, **38**, 9069–9083.

52. Rambo,R.P. and Doudna,J.A. (2004) Assembly of an active group II intron-maturase complex by protein dimerization. *Biochemistry*, **43**, 6486–6497.

53. Agrawal,R.K., Wang,H.-W. and Belfort,M. (2016) Forks in the tracks: Group II introns, spliceosomes, telomeres and beyond. *RNA Biol.*, **13**, 1218–1222.

54. Jaremko,M., Jaremko,L., Kim,H.Y., Cho,M.K., Schwieters,C.D., Giller,K., Becker,S. and Zweckstetter,M. (2013) Cold denaturation of a protein dimer monitored at atomic resolution. *Nat. Chem. Biol.*, **9**, 264–270.

55. Zhang,X.X., Jones,K.C., Fitzpatrick,A., Peng,C.S., Feng,C.J., Baiz,C.R. and Tokmakoff,A. (2016) Studying protein-protein binding through T-jump induced dissociation: transient 2D IR spectroscopy of insulin dimer. *J. Phys. Chem. B.*, **120**, 5134–5145.

56. Pyle,A.M., Fedorova,O. and Waldsich,C. (2007) Folding of group II introns: a model system for large, multidomain RNAs? *Trends Biochem. Sci.*, **32**, 138–145.