The group II intron ribonucleoprotein precursor is a large, loosely packed structure

Tao Huang[1], Tanvir R. Shaikh[1], Kushol Gupta[2], Lydia M. Contreras-Martin[1], Robert A. Grassucci[3], Gregory D. Van Duyne[2], Joachim Frank[3,4] and Marlene Belfort[1,*]

[1]Wadsworth Center, New York State Department of Health, Center for Medical Sciences, 150 New Scotland Avenue, Albany, NY 12201-2002, [2]Department of Biochemistry & Biophysics, University of Pennsylvania School of Medicine and Howard Hughes Medical Institute, 422 Curie Boulevard, Philadelphia, PA 19104, [3]Department of Biochemistry and Molecular Biophysics, Howard Hughes Medical Institute, Columbia University, 650 West 168th Street, New York, NY 10032, USA and [4]Department of Biological Sciences, Columbia University, 600 Fairchild Center, 1212 Amsterdam Avenue New York, NY 10027, USA

ABSTRACT

Group II self-splicing introns are phylogenetically diverse retroelements that are widely held to be the ancestors of spliceosomal introns and retrotransposons that insert into DNA. Folding of group II intron RNA is often guided by an intron-encoded protein to form a catalytically active ribonucleoprotein (RNP) complex that plays a key role in the activity of the intron. To date, possible structural differences between the intron RNP in its precursor and spliced forms remain unexplored. In this work, we have trapped the native *Lactococcus lactis* group II intron RNP complex in its precursor form, by deleting the adenosine nucleophile that initiates splicing. Sedimentation velocity, size-exclusion chromatography and cryo-electron microscopy provide the first glimpse of the intron RNP precursor as a large, loosely packed structure. The dimensions contrast with those of compact spliced introns, implying that the RNP undergoes a dramatic conformational change to achieve the catalytically active state.

INTRODUCTION

Group II introns are catalytic RNAs found in all three domains of life (1–3). Some of the interest in this class of introns derives from their suspected ancestral relationship to nuclear spliceosomal introns (4–9). Additionally, many group II introns are mobile genetic elements that are related to retrotransposons (10). Both mobile group II introns and most retrotransposons encode reverse transcriptase, the agent of their mobility. The intron-encoded protein (IEP) also functions as a maturase, to facilitate the proper folding of the intron RNA, for both splicing and intron mobility. Finally, the IEP can have DNA endonuclease activity, that plays a role in mobility (11,12).

The *Lactococcus lactis* group II intron LtrB (Ll.LtrB), comprises an RNA of 2492 nt, which forms a ribonucleoprotein (RNP) complex with two molecules of the IEP of 599 amino acids (70 163 Da) (13–15). This intron has been well characterized both genetically and biochemically, but biophysical and structural studies are more limited, with a lack of information on such parameters as size, shape and molecular flexibility. Although there is a crystal structure available for a bacterial group IIC intron RNA (16), there are only models of the class IIA Ll.LtrB intron, one for the IEP (15) and another for the complete RNP (17). Additionally, a low-resolution cryo-EM structure of the Ll.LtrB intron, which was artificially fused to a ribosomal RNA submit, is available (18).

The group II intron RNP is a dynamic particle involved not only in intron excision from the precursor RNA, but also in targeting and invasion of double-stranded DNA to initiate the retromobility reaction, and subsequently in completing cDNA synthesis (19,20). We first developed a purification scheme for the intron RNP precursor from the natural *L. lactis* host. We isolated the RNP with a branch-point mutation (ΔA) in the intron such that the first step of splicing was blocked and therefore the group II intron remained attached to short exons. The intron active site was otherwise intact. These precursor RNPs were examined by analytical centrifugation, gel filtration and cryo-electron microscopy (cryo-EM) to determine their size and shape, revealing a large loosely packed pre-splicing RNP structure. Comparison with the smaller

---

catalytically competent spliced intron suggests that the precursor RNP undergoes a major conformational change, compacting into its active structure.

MATERIALS AND METHODS

Plasmid constructs

A *L. lactis/Escherichia coli* shuttle vector, pLNRK, a pLE2-based plasmid containing a nisin promoter (21,22), was used for plasmid construction. The LtrB (ΔORF+A) fragment was amplified by PCR from pLNRK-23S-LtrB(ΔORF)+LtrA (18), whereas the LtrB (ΔORFΔA) fragment was amplified from pLNRK–23s-LtrB(ΔORF)+LtrA (18). For details see Supplementary Methods. The nisin-LtrA fragment (nisin promoter followed by LtrA ORF sequence) was amplified by PCR from pLE-nisin-LtrA and the intein CBD fragment from pImp-1P (23) and ligated to generate nisin-LtrA-intein-CBD, which was further ligated with linearized pLNRK-nisLtrB(ΔORF+A) or pLNRK-nisLtrB(ΔORFΔA). The resulting plasmids were called pLNRK-nisLtrB(ΔORF+A)+nisLtrA-intein-CBD abbreviated in the text to LtrBΔORF-LIC or '+A' and pLNRK-nisLtrB(ΔORFΔA)+nisLtrA-intein-CBD abbreviated in the text to LtrBΔORFΔA-LIC or 'ΔA', respectively. These two plasmids were then used as templates for cloning the other various plasmids used in this study. Plasmid pLNRK-nisLtrB(ΔORF+A)+nisLtrA-intein-CBD was cleaved by XbaI so that the part of nisin-LtrB(ΔORF+A) was removed and the rest of the vector was gel-purified and ligated to obtain the plasmid of pLNRK-nisLtrA-intein-CBD.

RNP purification

Lactococcus lactis IL1403 containing the aforementioned plasmids was grown in GM17 media at 30°C to an OD$_{600}$ of 0.5–0.6. Nisin was added at 10 ng/ml to induce the cells for 2–3 h. The culture was then centrifuged and the pellet was washed in a buffer containing 10 mM Tris–HCl, pH 7.5, 150 mM NaCl and 1 mM EDTA. The washed cell pellet was re-suspended in 20 mM Tris–HCl, pH 8.0, 500 mM NaCl, 0.1 mM EDTA and 1 mM PMSF. This cell suspension was frozen at −80°C before purification. The cell lysate was loaded onto a chitin column, which was washed extensively, and then immersed in 40 mM DTT to activate the intein of the C-terminal tag to release the LtrA protein in complex with LtrB group II intron RNA (for details see Supplementary Methods). The RNP was further purified away from other contaminants by sucrose gradient sedimentation. Each sucrose gradient fraction was measured by its UV absorbance. Usually, an OD$_{260}$ peak could be identified in the middle of the sucrose gradient.

Primer extension analysis

Primer extension was performed as described (24), using primer IDT1073 (5'-GTA CCT TAA ACT ACT TGA CTT AAC ACC-3'), which was complementary to the 5'-end of the LtrB group II intron RNA. Primer extension products were separated in 10% polyacrylamide/8 M urea gels, visualized by autoradiography, and quantified by phosphorimaging and ImageQuant 5.2.

Electron microscopy and image processing

Cryo-EM specimens were prepared on Quantifoil R 2/4 grids (Quantifoil Micro Tools GmbH, Jena, Germany), onto which thin carbon was floated. Grids were glow-discharged for 25 sec using a plasma cleaner (Harrick PKC-3XG, Harrick Scientific Corp., Ossining, NY, USA) to make them hydrophilic (25). Samples of 4 μl were pipetted onto a grid. Grids were blotted in ∼85% humidity for 10 s and plunge-frozen into liquid ethane at liquid-nitrogen temperature using a Vitrobot (FEI, Hillsboro, OR, USA). We also made grids in parallel with negative staining using uranyl acetate as previously described (26,27).

Cryo-EM was performed on an FEI (Eindhoven, The Netherlands) Tecnai F20 at a nominal magnification of 50 000× under low-dose conditions (20 e$^-$/Å^2). Tilt pairs were collected, first at nominally 60° and then at 0°, onto SO-163 film (Eastman Kodak, Rochester, NY, USA). Micrographs were digitized on a Zeiss Imaging scanner (Z/I Imaging Corporation, Huntsville, AL, USA) at a step size of 14 μm, corresponding to a calibrated pixel size of 2.76 Å on the object scale.

Image processing was performed using SPIDER (28). Batch files for reconstruction by random-conical tilt (RCT) (29) were adapted from Shaikh *et al.* (30). First, reference-free alignment was implemented based on the 'pyramidal system for pre-alignment reconstruction' (31). The corresponding SPIDER batch file *pairwise.spi* is located at http://www.wadsworth.org/spiderdoc/spider/docs/techs/MSA/index.htm. Second, classification-based particle-verification (32) was incorporated into the image selection procedure. Then, the RCT reconstruction was used as a reference for projection matching (33) against the tilted-specimen images.

Size-exclusion chromatography

For determination of the Stokes radius (R_s), size-exclusion chromatography (SEC) experiments were performed with a Superdex 200 10/300 GL column (GE Healthcare) at 0.5 ml/min at room temperature in buffer containing 20 mM Tris 7.5, 200 mM KCL and 5 mM MgCl$_2$. The column was calibrated using the following proteins (Bio-Rad): thyroglobulin (670 kDa, R_s = 85 Å), γ-globulin (158 kDa, R_s = 52.2 Å), ovalbumin (44 kDa, R_s = 30.5 Å), myoglobin (17 kDa, R_s = 20.8 Å) and Vitamin B12 (1350 Da). Blue-Dextran (Sigma) was used to define the void volume of the column.

Sedimentation velocity

Purified +A and ΔA particles were diluted to ∼10 ng/μl concentrations in buffer that contained 20 mM Tris 7.5, 200 mM KCl and 5 mM MgCl$_2$. Sedimentation velocity (SV) ultracentrifugation experiments were performed at 4°C with an XL-A analytical ultracentrifuge (Beckman) and a TiAn60 rotor with two-channel charcoal-filled epon centerpieces and quartz windows. Complete sedimentation velocity profiles were collected every 30 s for

200 boundaries at 45 000 r.p.m. Data were fit using the c(s) distribution model of the Lamm equation as implemented in the program SEDFIT (34). After optimizing meniscus position and fitting limits, the sedimentation coefficient (s) and best-fit frictional ratio (f/f_0) was determined by iterative least squares analysis. Sedimentation coefficients were corrected to $s_{20,w}$ based on the calculated solvent density (ρ) and viscosity (η) derived from chemical composition by the program SEDNTERP (35). To derive the partial specific volume (υ_{bar}) of the composite protein–RNA particles, component densities were derived using the program NUCPROT (36) and mass-averaged as described previously (37). For +A, this value for υ_{bar} is 0.63 cm^3/g; for ΔA, 0.62 cm^3/g. These hydrodynamic parameters were combined in SEDFIT to calculate the weight average molecular weights for integrated peaks.

RESULTS

Purification of the native LtrB group II intron precursor RNP

We designed an affinity purification scheme to facilitate the isolation of the LtrB group II intron RNP in its pre-splicing state from its native *L. lactis* host. The intron and the IEP, called LtrA, are expressed from independent promoters (Figure 1A). LtrA was fused to an affinity tag, chitin binding domain (CBD), via an intein at the C-terminus of the LtrA (LIC: LtrA-Intein-CBD) (Figure 1B, construct 1). A truncated version of the intron, deleted for most of the 1797-nt LtrA open reading frame (ΔORF) (deletion of 1590 nt) was expressed upstream of the LIC cassette (Figure 1B, construct 2).

The size of the model intron used in this study (902 nt) is average relative to the known group II introns in eubacteria (2), but this intron is significantly larger than that for which a crystal structure has been determined (16,17). The size of the ΔA precursor, including the attached exons, is 983 nt, corresponding to a 901-nt intron (with the ΔA mutation) with a 27-nt 5′exon and a 54-nt 3′exon. The LtrA protein, which binds as a dimer to the intron RNA, has 599 amino acids, corresponding to a molecular weight of 70 163 Da (for each monomer).

Expression of both the LtrB ΔORF RNA and the LIC fusion protein were driven by two independent nisin A promoters, which were induced by addition of 10 ng/ml of nisin. The construct designed for purifying the group II intron RNP in its precursor form had a deletion of the nucleophile for the first step of splicing, the branch-site adenosine, to keep the LtrB group II intron in its precursor form (Figure 1B, construct 3). We designated this construct as LtrBΔORFΔA-LIC and abbreviate its name to 'ΔA', in contrast to its splicing-competent '+A' counterpart (Figure 1B, construct 2). Upon activation by thiol reagents, the intein cleaves at its N-terminus, to release LtrA and its associated LtrB group II intron from the CBD column, as an RNP complex in its native form (Figure 1A). Purification of the +A RNPs was similar, and will be reported elsewhere.

We next conducted a functional analysis of the ΔA RNP precursor and compared it to the +A RNP splice
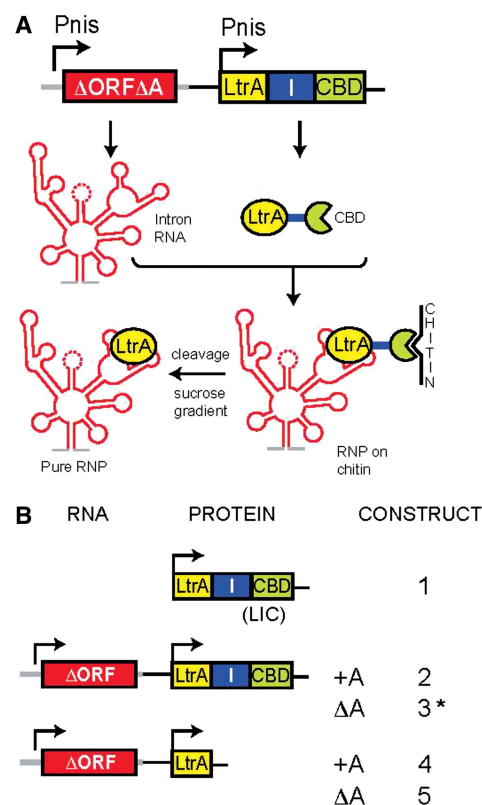

Figure 1. Purification scheme of intron RNP. (**A**) Purification strategy. Lysates from induced cells are passed over a CBD column, which traps the LtrA-intein-CBD (LIC) protein fusion and associated RNA. Cleavage of LtrA from the intein with DTT releases the RNP from the chitin column. (**B**) Constructs. The purification construct (3, marked with asterisk) (ΔA) and control constructs are illustrated.

product (Figure 1B, constructs 2 and 3), to verify that the protein and RNA components of the RNPs were present and that intact particles of appropriate activity were isolated. We compared these ΔA and +A particles with those from constructs lacking the group II intron (Figure 1B, construct 1) and missing the LIC purification module (Figure 1B, constructs 4 and 5). First, these constructs were analyzed with respect to LtrA expression on a 10% SDS–polyacrylamide gel (Figure 2A). A nisin-inducible band of ~64 kD for the ΔA and +A constructs lacking the LIC module (lanes 4 and 5) was replaced by a >100 kDa nisin-inducible band for all the LIC-containing constructs (lanes 1–3), including the ΔA purification construct (lane 3). The bands in the +A and ΔA constructs were of similar intensity (lanes 2 and 3), indicating that the LtrA intein fusion was being properly expressed, and highly soluble.

We then examined the RNA from the same constructs separated on a 1.2% agarose–formaldehyde gel (Figure 2B). Besides the 23S and 16S rRNA bands, intron RNA was clearly visible as a ~1-kb band, again in nisin-dependent fashion, for all constructs containing the group II intron (lanes 2–5). The ΔA-introns had a band of ~1 kb (lanes 3+ and 5+), corresponding well to the size of the 983-nt precursor. In contrast, the +A intron had only a faint band of this size. Rather, the +A constructs had a slightly smaller band, which we identified as
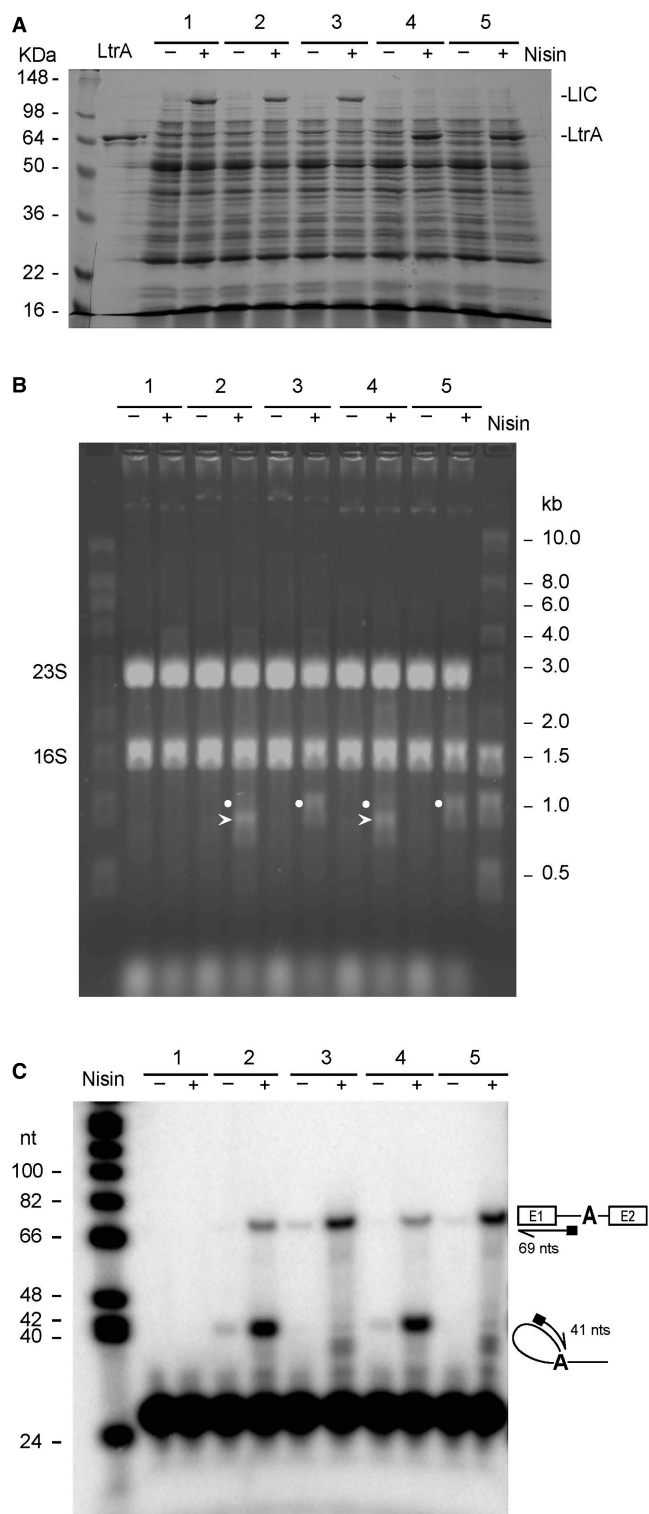
Figure 2. LtrA protein and the LtrB intron RNA were well induced in active form. (**A**) LtrA expression. Total protein lysate corresponding to constructs 1–5 in Figure 1B was prepared from each strain and separated by 10% SDS–PAGE. Protein bands were visualized after Coomassie blue staining. All the LIC constructs showed a distinct protein band of ∼130 Kd, corresponding in size to LtrA-intein-CBD upon induction with nisin, whereas the constructs without the intein-CBD fusion showed a protein band of ∼70 Kd, corresponding to the size of LtrA. (**B**) Intron RNA expression. Total RNA was prepared from the same set of constructs as above, and 5 μg of total

the excised 902-nt intron (lanes 2+ and 4+). To verify that these band assignments were correct, we conducted northern blots (data not shown) and a primer-extension termination assay, in which we separated cDNA products of the RNAs on a 10% denaturing acrylamide gel (Figure 2C). Again whereas cDNAs corresponding only to precursor species accumulated in both ΔA constructs (lanes 3+ and 5+), an abundance of cDNA corresponding to excised intron was observed in the +A constructs, representing an 80% splicing efficiency (intron/total) (lanes 2+ and 4+). We therefore conclude that the RNP in our purification construct is catalytically active, and that splicing is effectively inhibited by the ΔA branch-point mutation, allowing us to purify the RNP precursor prior to splicing. We further demonstrated that the +A but not the ΔA construct was active in DNA targeting in *L. lactis* (data not shown).

After validating the constructs, ΔA RNP precursor from the LtrA-intein-CBD LIC fusion, (Figure 1B, construct 3), was purified as described in 'Materials and methods' section (Figure 1A), and monitored for protein on polyacrylamide gels (Figure 3A) and for RNA on agarose gels (Figure 3B). Briefly, the RNP was trapped on a chitin column (Figures 1A and 3A, lanes 1–4; Figure 3B, lane 10), concentrated through a sucrose cushion (Figure 3A, lane 5; Figure 3B, lane 11) and then the pellet was fractionated by sucrose gradient sedimentation to purify away contaminants (Figure 3B, lanes 1–8). The 1-kb intron RNA co-purified with 16S ribosomal RNA, but there were fractions early in the gradient that were essentially free of 16S RNA (Figure 3B, lane 6, pooled clean fractions). The purified RNP contained a single 64-kD protein on a Coomassie-stained SDS–PAGE gel (Figure 3A, lane 6), and a single intron RNA band of ∼1 kb on an agarose gel (Figure 3B, lane 12). LtrA was confirmed by Western blot with anti-LtrA antibody and LtrB RNA by northern blot using corresponding probes (data not shown).

The interaction between the group II intron RNP and 16S rRNA seemed to be specific. When the LtrB group II intron expression plasmid with untagged LtrA, designed to flow through the chitin column, was used as the negative control for the purification, no 16S rRNA co-purification was observed (data not shown). Likewise 23S rRNA co-purification, which we observed to much lower levels, was specific. The association of the group II intron RNP with the 30S ribosomal subunit is being investigated in an independent study.

RNA from each, with (+) or without (−) nisin induction, was separated on 1.2% agarose gel containing 2.2 M formaldehyde. All the strains expressing LtrB intron RNA showed an RNA band of ∼1 kb with nisin induction (lanes 2, 3, 4 and 5, dots). The slight difference in migration of the spliced intron was due to the size difference between the intron precursor and the intron lariat (lanes 2 and 4, arrows). (**C**) Primer extension assay. A primer was designed to bind to the 5′-end of the LtrB intron. A 69-nt primer extension product was generated from LtrB intron precursor (lanes 2–5); and a 41 nt primer extension product was generated from LtrB intron lariat (lanes 2 and 4). Primer extension products were separated in 10% polyacrylamide/ 8 M urea gels, visualized by autoradiography.
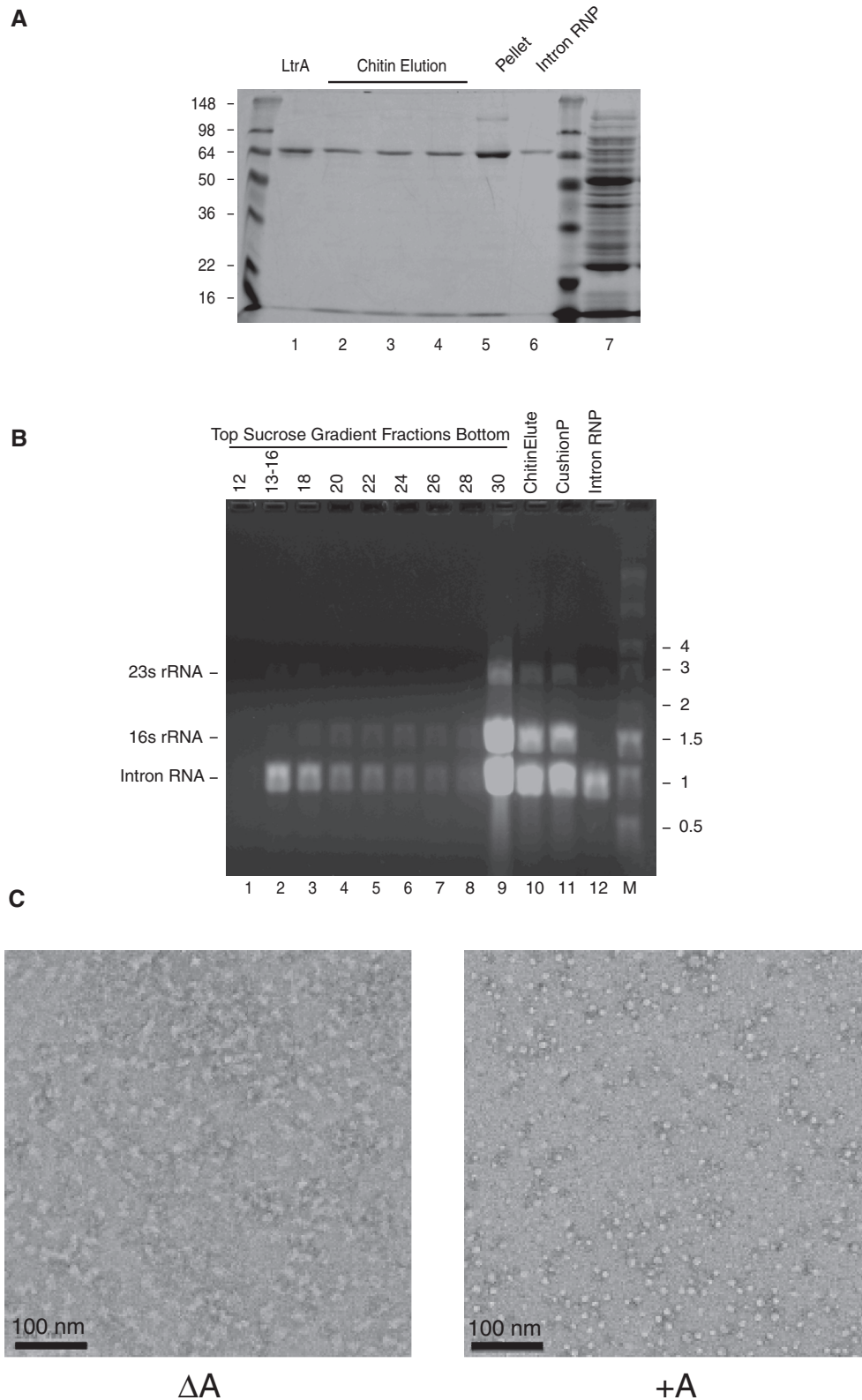
Figure 3. Intron RNP precursor purification. (**A**) Purified intron RNP showed a single protein band, whose size corresponded to the intron-encoded protein, LtrA. Lane 1: purified LtrA; lanes 2–4: chitin elution fractions; lane 5: re-suspension of the pellet from sucrose cushion centrifugation; lane 6: purified intron RNP; lane 7: total cell lysate. (**B**) RNA gel of the purified intron RNP. A band of ~1 kb corresponds to the size of LtrB intron precursor. The RNA content of each even-numbered sucrose gradient fraction was analyzed by 1.2% agarose gel containing 2.2 M formaldehyde. Fractions 13–16 were pooled and concentrated as the purified intron RNP. The RNA content of the pooled chitin elution and the re-suspension of the sucrose cushion pellet were also analyzed. (**C**) Negative-stain micrographs of ΔA and +A RNPs. Specimens were prepared in uranyl acetate.

Both the purified ΔA precursor RNP and the spliced +A particle were visualized in the electron microscope, which showed many particles of similar size in each case (Figure 3C). Strikingly, the precursor RNP particles appeared larger than the +A spliced intron RNPs (characterization to be described elsewhere) (Figure 3C). These homogeneous images of the precursor RNP demonstrated the high quality of the purification and paved the way towards biophysical characterization of the LtrB group II intron RNP in its pre-splicing form.

Solution analyses suggest that the precursor RNP forms a loosely packed structure

The large apparent volume of the ΔA precursor particle prompted us to perform biophysical analyses of the ΔA RNP, with the +A particle as a control. We first examined the hydrodynamic properties of purified +A and ΔA RNPs using SEC and SV. Because of the strong absorbance properties of these particles, we were able to conduct our analyses at relatively low concentrations (10–200 ng/μl). At these very low molar concentrations, the complexes were tractable and remained intact, implying a very high affinity between the LtrA protein and its binding site in both particles.

The +A particle eluted as a single peak (+A$_1$) by SEC (Figure 4A), with a determined Stokes radius (R_s) of $70.6 \pm 6.8\,\text{Å}$ ($n = 4$). In contrast, injections of the ΔA sample yielded a species with a much larger hydrodynamic radius [ΔA$_1$, $R_s = 98.5 \pm 7.1\,\text{Å}$ ($n = 8$)], as evidenced by a peak with a shorter retention time, appearing immediately following the void volume of the column (Figure 4A). This difference in R_s (defined as the radius of a hard sphere in solution) would imply a large difference in volume between the two particles. Additionally, a second species evolved over the course of the first hour after thawing of the ΔA precursor particle. This second species [ΔA$_2$, $R_s = 76.1 \pm 5.9\,\text{Å}$ ($n = 4$)] is largely consistent with the retention time and R_s of the +A particle. Isolation, concentration and reinjection of the ΔA$_1$ peak resulted in both peaks being observed, suggesting that these species are in equilibrium in solution (data not shown). For both precursor and +A preparations, we suspect that the species with a small apparent molecular mass (ΔA$_3$, $R_s = 21.4\,\text{Å}$; +A$_2$, $R_s = 22.1\,\text{Å}$) may reflect a low degree of particle degradation or a buffer artifact; these species had no measureable light scattering properties (data not shown).

In SV experiments, the absorbance data collected for the +A particle were readily fit to the Lamm equation across all data time boundaries (starting at time 0; Figure 4B). Consistent with SEC, SV analysis of the +A particle showed a single peak within the concentration distribution of S-values, c(S), with an $S_{20,w}$ value of 12.9, and a frictional coefficient (f/f_o) of 2.3. This latter value is consistent with an elongated shape (Figure 4D). Using these determined parameters and a mass-averaged partial specific volume derived from our understanding of the composition of the particle (see 'Materials and methods' section), a mass-averaged molecular weight of 432 729 Da was calculated, in concordance with the
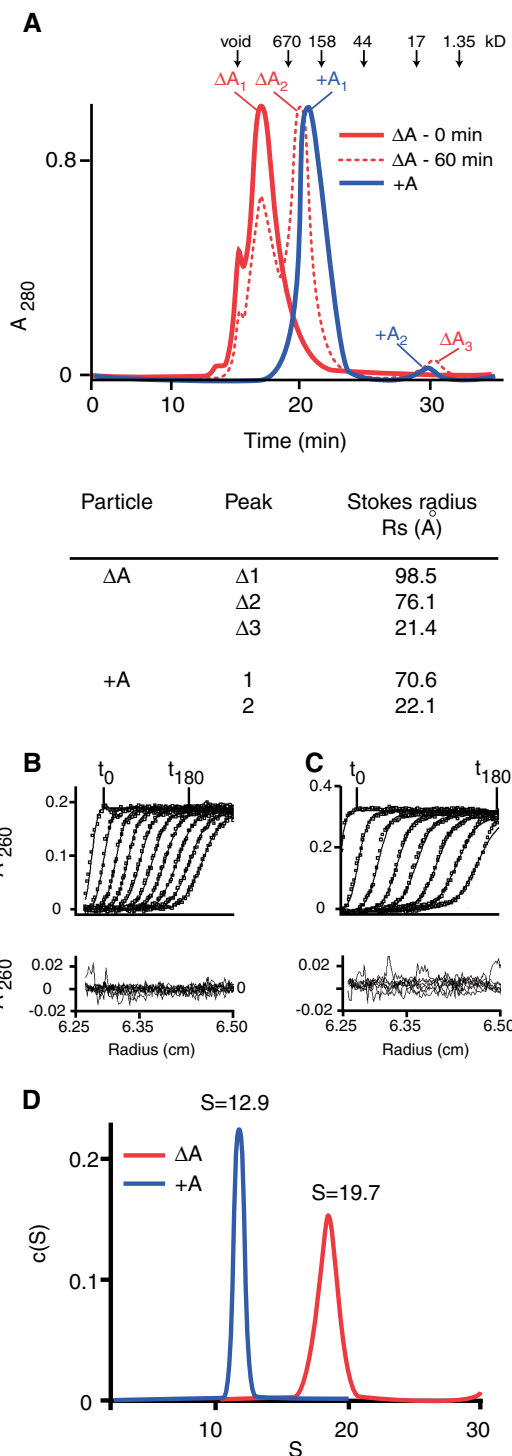

Particle	Peak	Stokes radius Rs (Å)
ΔA	Δ1	98.5
	Δ2	76.1
	Δ3	21.4
+A	1	70.6
	2	22.1

Figure 4. Hydrodynamic comparison of precursor and product introns indicates a major conformational change. (**A**) SEC. The size-exclusion peaks for the +A and ΔA RNP particles are compared. The identity of the small 17-kDa peak with both samples is not known. (**B** and **C**) SV analysis. Shown in (B and C) are the experimental data for +A and ΔA RNP, respectively, rendered as black points on solid black lines that are the fits to the Lamm equation. Each boundary shown corresponds to a 30-s time interval, starting at time = 0 (data are shown only for the initial boundaries). Residuals, showing the agreement between the absorbance data collected and the theoretical fit to the Lamm equation, are shown below each panel as a function of the radius of the experimental cell. (**D**) RNP particles are compared on a concentration distribution of S values, c(S).

theoretical molecular weight of ~432 000 Da (corresponding to an LtrA dimer bound to a 902-nt intron RNA).

In contrast, for the ∆A RNP, only data through the first several boundaries of the experiment could be modeled (Figure 4C), with fits becoming increasingly discrepant as a function of time. This phenomenon is consistent with a particle undergoing a change in shape in solution during the course of the experiment, and mirrors the observations made by SEC. A comparison of the residuals for both particles (Figure 4B and C, lower panels) indicates the quality of the fitting through the first several initial boundaries for both particles. For the ∆A particle, this analysis yields a determined $S_{20,w}$ value of 19.7 and a frictional coefficient (f/f_o) of 1.63 (Figure 4D), suggesting a more globular shape than the +A particle. A molecular weight of 473 167 Da was derived, consistent with the predicted mass of ~458 000 Da. In agreement with these derived masses, the RNA and protein compositions of the ∆A and the +A particles do not suggest any major differences that could account for the dramatic disparities in hydrodynamic properties observed between these two particles. Collectively, our measurements raise the possibility that ∆A forms a loosely packed structure that undergoes significant compaction with time to achieve a similar conformation to the +A RNP particle.

Two other elements of the ∆A precursor construct might account for the larger apparent hydrodynamic state observed relative to the active +A RNP particle. First, the precursor contains the ∆A deletion in the intron RNA, introducing a change in RNA composition relative to the +A RNA. Second, the ∆A precursor RNA contains ~80 additional nucleotides corresponding to the exons, which are not found in the active +A RNP particle. To evaluate the first possibility, we used a bioinformatic analysis that predicts secondary structure changes introduced by defined mutations (38). The correlation coefficient was >0.99 for the ∆A mutation, suggesting no effect in the base-pairing energy landscape that would be indicative of a structural change introduced by the ∆A mutation. To evaluate the second possibility, we performed theoretical calculations to predict the size and volumes of the RNP particles and their component parts (Supplementary Table S1). Assuming that the binding sites of the LtrA protein are maintained, our calculations suggest that both RNPs, as compact and well-folded macromolecules, should yield similar hydrodynamic properties. The deviation from theoretical predictions observed, despite a similar measured mass between the two particles, (Supplementary Table S1), suggests that the discrepancy in apparent volume observed stems from loose packing of the ∆A RNP.

Electron microscopy and three-dimensional reconstruction confirm a loosely packed structure for the ∆A precursor

We proceeded to collect images of the precursor particle by electron microscopy to analyze its overall shape. In addition to preparation of cryo-EM grids, we also prepared grids in parallel with negative staining, using uranyl acetate and ammonium molybdate. The preservation

of the precursor RNP appeared best in the cryo-specimen. While the yield of tilted images unaffected by beam-induced movement would probably have been better on the negative stain images, we pursued the cryo-specimens for the precursor RNP. A *de novo* reconstruction was computed using the random-conical tilt (RCT) method (29).

Ten out of 28 tilt pairs of micrographs were found to be of high quality and were digitized. From these 10 pairs, 1586 particles were selected. The particle images were subjected to reference-free alignment and classification (see 'Materials and Methods' section). Class averages of common views were selected and used for 2D multi-reference alignment and classification. Multi-reference alignment and classification were iterated one more time.

Two common views were seen (Figure 5A), which we termed 'face' view and a 'chain link' view. After the second iteration of 2D alignment and classification, 298 particles were chosen from the face view and 93 from the chain-link view. Reconstructions were computed for each (Figure 5B, a1–a3 and b1–b3). In addition, the face-view reconstruction was used as a reference for projection-matching of all particles, with the hope to include particles that adopted other, less-common orientations. In this merged reconstruction, a total of 647 particles were included (Figure 5B, c1–c3). The chainlink-view reconstruction (Figure 5B, b1–b3) is less defined and quite noisy compared to the face-view and merged reconstructions, on account of the small number of particles included. In addition, it must be noted that as a rule, RCT reconstructions appear best defined when viewed in the direction of the view they were created from. Thus, Figure 5B, a1 displays the best resolution for the face-view reconstruction, while Figure 5B, b3 shows the best resolution for the chainlink-view reconstruction, in a view which is equivalent to a side view of the face-view reconstruction. Although the possibility cannot be excluded that the view preference is caused by conformational differences, we find it reasonable to assume that the chainlink-view reconstruction depicts the molecule in the same conformation as the reconstruction from face-view particles. Therefore, we believe that the merged reconstruction (Figure 5B, c1–c3) is the best overall depiction of the molecule because of the inclusion of particles in multiple views.

The precursor RNP shows a loosely packed conformation, with dimensions of about 250 Å (the longest dimension, along the "forehead" of the face) × 200Å × 140 Å. In the face view (Figure 5B, a1–a3), the 'eyes' do not correspond to holes penetrating the entire structure, as might be assumed from the appearance of the 2D averages, but the cavity corresponding to the left eye is much deeper than that of the right eye. The 'mouth' visible in the face views is a large gaping hole in the reconstruction. The side view (Figure 5B, a3, b3 and c3) is characterized by the appearance of three lobes. Overall, the structure can be characterized as being porous and unexpectedly large according to mass, consistent with the data we obtained from the solution studies.
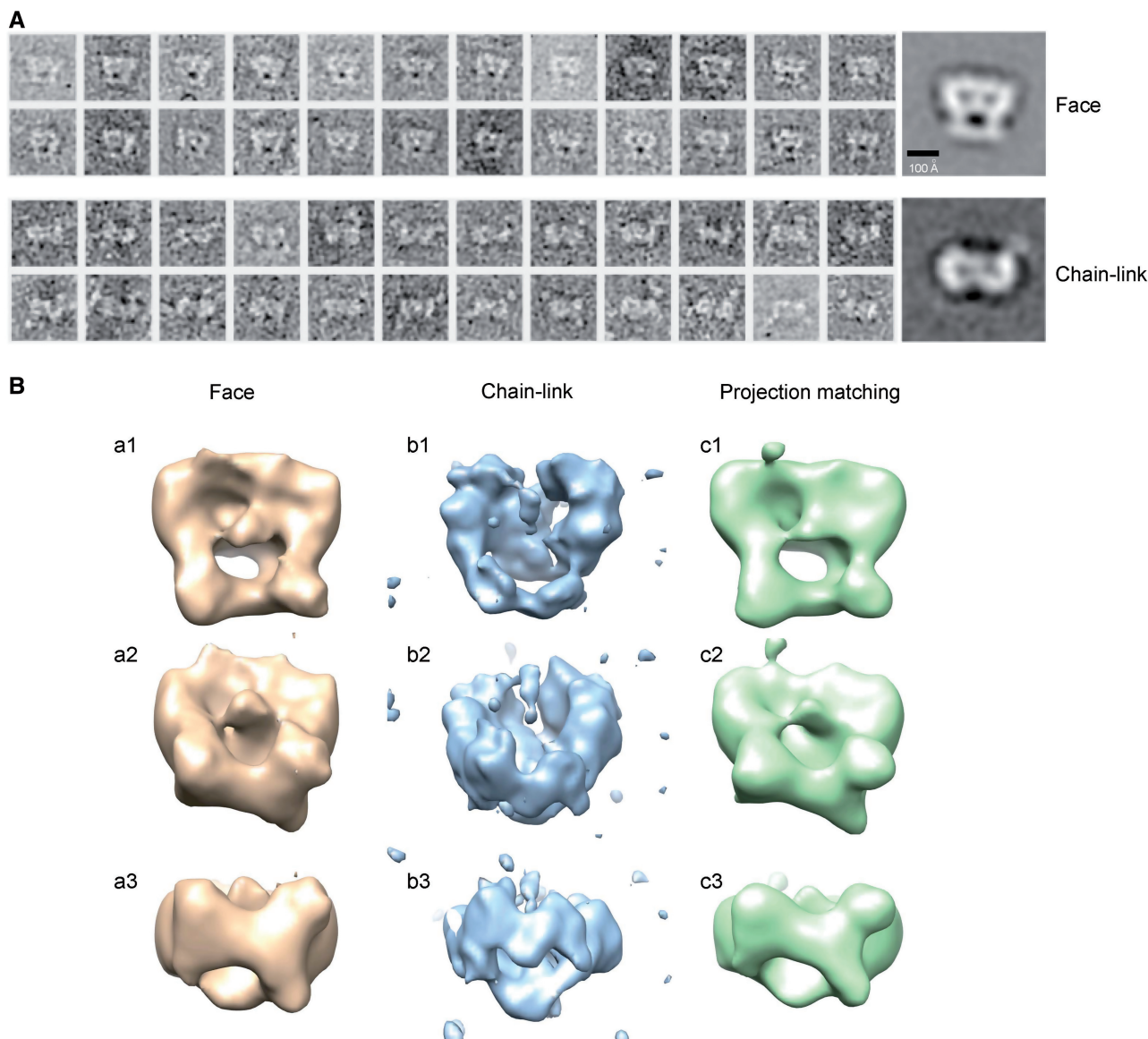
Figure 5. (A) Cryo-EM images of the group II intron RNP precursor. (A) Top two rows are montages of individual, low-pass filtered particles corresponding to the dominant 'face' view. At the right is an average of 298 particles. Scale bar represents 100 Å. Bottom two rows are montages of individual, low-pass filtered particles corresponding to the 'chain-link' view. At the right is an average of 93 particles. (B) RCT reconstructions and merged reconstruction shown in front (first row), intermediate (second row) and side views (third row). (a_1, a_2, a_3) RCT reconstruction from 298 particles in face view, some of which are depicted in (A). (b_1, b_2, b_3) RCT reconstruction from 93 particles in side view, some of which are depicted in (A). (c_1, c_2, c_3) Reconstruction of 647 particle images using as a reference the face-view reconstruction shown in panels (a_1, a_2, a_3).

DISCUSSION

Using biophysical methods of analysis and visualization, we have determined that the *L. lactis* group II intron RNP precursor is a particle with a loosely packed structure. In our quest to characterize the structure, we first needed to develop a purification strategy for the ΔA precursor, which included an intein-based affinity method that resulted in LtrA-intron co-purification. Yields were low (~1.5 µg/l culture) because of the association we observed between the group II intron and 16S and 23S rRNA. Yet, we were able to isolate pure group II intron RNP by careful selection of sucrose fractions, to yield RNPs that were homogeneous, well assembled and stable.

Data from SV experiments, SEC and images collected from cryo-EM collectively suggest a large, loosely packed structure for the ΔA precursor RNP, relative to its active +A counterpart. Shape reconstructions of the +A particle by small-angle X-ray scattering (SAXS) analysis supports an approximate contraction of 3–4 times the volume for the precursor RNP relative to the +A particle (Gupta K, *et al.*, manuscript in preparation). The structural aspects of the precursor RNP are consistent with previous visualization of this particle (a globule with about 200 Å in diameter; see Supplementary Figure S1), where the chimeric ΔA group II intron structure, anchored to the ribosome as a covalent fusion, appeared surprisingly large

relative to the ribosome, a particle approximately five times larger in mass (18). This chimeric ΔA group II intron structure appears even slightly larger than the one we observe in this work. This difference is likely explained by the constructs we used here, which are truncations of the intron ORF (where the RNA is only ∼900-nt long). In the previous work we used the full-length intron of ∼2500 nt. This work not only argues against the possibility that intron size in the chimera is influenced by the presence of the ribosome, but also emphasizes the common feature of the two studies, namely the surprisingly large volume of intron precursor structures.

Related to the unusual volume-to-mass ratio, another intriguing feature of the precursor RNP structure is its porous nature. The porosity is consistent with the cavity and tunnels throughout the interior of the intron that we previously hypothesized to account for the large volume of the structure (18). While studies are in progress to understand the functional significance of different features of this structure, we speculate that the cavity might accommodate the exon RNA for alignment of the splice sites within the catalytic core of the group II intron ribozyme. Alternatively or additionally, the pore might accommodate the target DNA substrate for invasion. Interestingly, cryo EM studies of the spliceosome have also revealed a cavity, which is hypothesized to provide a transient storage location for precursor RNA, not directly involved in the ongoing splicing reaction (39–41).

An interesting contrast exists between the large volume of the precursor intron from our work and the compact crystallographic structure of the RNA of the *Oceanobacillus iheyensis* intron (16). Although these are two different introns, and the *L. lactis* intron has twice the length of the *O. iheyensis* intron and has been imaged in the presence of the IEP, we assume that the size discrepancy is significant. The implication is that the precursor particle undergoes a dramatic conformational change that leads to a more tightly packed structure in preparation for splicing. This possibility is supported by the different physical properties exhibited by the precursor particle relative to the spliced +A particle. The breadth of the gel filtration profile for ΔA and its sampling of a solution state that coincides with +A is consistent with the suggestion that the precursor RNP is sampling different conformations in solution, and this is corroborated by the results of sedimentation velocity analysis.

Our hypothesis suggesting conformational changes during group II intron splicing is in accord with biochemical studies indicating long-range pairings that induce structural transitions. Specifically, formation of a tetraloop–receptor interaction (also known as the η–η' interaction) between domains II and VI mediate a conformational change of domain VI during the second step of self-splicing in both group IIA and IIB introns (42–44). These authors hypothesized that domain VI may first occupy the catalytic core to provide the nucleophile for the first step of splicing and domain VI is then displaced from the core, with the latter conformation being stabilized by the η–η' interaction. This large positional switch of domain VI between the two steps of splicing may be partly responsible for the conformational changes observed in our study. How the group II intron RNP achieves its conformational change independent of ATP and/or GTP hydrolysis and NTPases, as occurs during the assembly of a catalytically active spliceosome (45), is worthy of further exploration.

A much larger cryo-EM data collection is currently ongoing, in the order of hundreds of thousands of particles, to obtain the necessary statistics and resolution for giving a conclusive answer on the presence and extent of heterogeneity of the ΔA particle. The reconstruction we obtained by the RCT method will be instrumental as a 3D reference in this effort. These future studies will allow us to gain further mechanistic insights into questions related to how the various pores and cavities we have observed might contribute to the catalytic activity of the RNP, and how contact with target DNA molecules might contribute to the dramatic changes in shape observed between the precursor group II intron RNP and its spliced intron counterpart.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Belfort,M., Derbyshire,V., Cousineau,B. and Lambowitz,A. (2002) Mobile introns: pathways and proteins. In Craig,N., Craigie,R., Gellert,M. and Lambowitz,A. (eds), *Mobile DNA II*. ASM Press, pp. 761–783.
2. Dai,L., Toor,N., Olson,R., Keeping,A. and Zimmerly,S. (2003) Database for mobile group II introns. *Nucleic Acids Res.*, **31**, 424–426.
3. Pyle,A.M. and Lambowitz,A.M. (2006) Group II introns: ribozymes that splice RNA and invade DNA. *The RNA World*, 3rd edn. Cold Spring Harbor Laboratory Press, pp. 469–505.
4. Cech,T.R. (1986) The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell*, **44**, 207–210.
5. Cavalier-Smith,T. (1991) Intron phylogeny: a new hypothesis. *Trends Genet.*, **7**, 145–148.
6. Sharp,P.A. (1991) Five easy pieces. *Science*, **254**, 663.
7. Hetzer,M., Wurzer,G., Schweyen,R.J. and Mueller,M.W. (1997) Trans-activation of group II intron splicing by nuclear U5 snRNA. *Nature*, **386**, 417–420.

8. Shukla,G.C. and Padgett,R.A. (2002) A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. *Mol. Cell*, **9**, 1145–1150.

9. 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.

10. Eickbush,T.H. (1999) Mobile introns: retrohoming by complete reverse splicing. *Curr. Biol.*, **9**, R11–R14.

11. 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.

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

13. Mills,D.A., McKay,L.L. and Dunny,G.M. (1996) Splicing of a group II intron involved in the conjugative transfer of pRS01 in lactococci. *J. Bacteriol.*, **178**, 3531–3538.

14. Lambowitz,A.M. and Zimmerly,S. (2004) Mobile group II introns. *Annu. Rev. Genet.*, **38**, 1–35.

15. Blocker,F.H., 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.

16. 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.

17. Dai,L., Chai,D., Gu,S.Q., Gabel,J., Noskov,S.Y., Blocker,F.J., Lambowitz,A.M. and Zimmerly,S. (2008) A three-dimensional model of a group II intron RNA and its interaction with the intron-encoded reverse transcriptase. *Mol. Cell*, **30**, 472–485.

18. Slagter-Jäger,J.G., Allen,G.S., Smith,D., Hahn,I.A., Frank,J. and Belfort,M. (2006) Visualization of a group II intron in the 23S rRNA of a stable ribosome. *Proc. Natl Acad. Sci. USA*, **103**, 9838–9843.

19. 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.

20. 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.

21. Mills,J.B., Cooper,J.P. and Hagerman,P.J. (1994) Electrophoretic evidence that single-stranded regions of one or more nucleotides dramatically increase the flexibility of DNA. *Biochemistry*, **33**, 1797–1803.

22. de Ruyter,P.G., Kuipers,O.P., Beerthuyzen,M.M., van Alen-Boerrigter,I. and de Vos,W.M. (1996) Functional analysis of promoters in the nisin gene cluster of Lactococcus lactis. *J. Bacteriol.*, **178**, 3434–3439.

23. 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.

24. Siatecka,M., Reyes,J.L. and Konarska,M.M. (1999) Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center. *Genes Dev.*, **13**, 1983–1993.

25. Grassucci,R.A., Taylor,D.J. and Frank,J. (2007) Preparation of macromolecular complexes for cryo-electron microscopy. *Nat. Protocols*, **2**, 3239–3246.

26. Stoffler,G. and Stoffler-Meilicke,M. (1983) The ultrastructure of macromolecular complexes studied with antibodies. In Tesche,H. (ed.), *Modern Methods in Protein Chemistry*. De Gruyter, Berlin, pp. 409–455.

27. Stoffler,G. and Stoffler-Meilicke,M. (1984) Immunoelectron microscopy of ribosomes. *Annu. Rev. Biophys. Bioeng.*, **13**, 303–330.

28. Frank,J., Radermacher,M., Penczek,P., Zhu,J., Li,Y., Ladjadj,M. and Leith,A. (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.*, **116**, 190–199.

29. Radermacher,M., Wagenknecht,T., Verschoor,A. and Frank,J. (1987) Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. *J. Microscopy*, **146**, 113–136.

30. Shaikh,T.R., Gao,H., Baxter,W.T., Asturias,F.J., Boisset,N., Leith,A. and Frank,J. (2008) SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat. Protocols*, **3**, 1941–1974.

31. Marco,S., Chagoyen,D., de la Fraga,L.G., Carazo,J.M. and Carrascosa,J.L. (1996) A variant to the "random approximation" of the reference-free alignment algorithm. *Ultramicroscopy*, **66**, 5–10.

32. Shaikh,T.R., Trujillo,R., LeBarron,J.S., Baxter,W.T. and Frank,J. (2008) Particle-verification for single-particle, reference-based reconstruction using multivariate data analysis and classification. *J. Struct. Biol.*, **164**, 41–48.

33. Penczek,P., Radermacher,M. and Frank,J. (1992) Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy*, **40**, 33–53.

34. Schuck,P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.*, **78**, 1606–1619.

35. Hayes,D.B., Laue,T.M. and Philo,J. (2003) SEDNTERP (Sedimentation Utility Software). Amgen Corp http://www.rasmb.bbri.org/ (6 November 2010, date last accessed).

36. Voss,N.R. and Gerstein,M. (2005) Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly. *J. Mol. Biol.*, **346**, 477–492.

37. Kar,S.R., Lebowitz,J., Blume,S., Taylor,K.B. and Hall,L.M. (2001) SmtB-DNA and protein-protein interactions in the formation of the cyanobacterial metallothionein repression complex: Zn2+ does not dissociate the protein-DNA complex in vitro. *Biochemistry*, **40**, 13378–13389.

38. Halvorsen,M., Martin,J.S., Broadaway,S. and Laederach,A. (2010) Disease-associated mutations that alter the RNA structural ensemble. *PLoS Genet.*, **6**, e1001074.

39. Azubel,M., Wolf,S.G., Sperling,J. and Sperling,R. (2004) Three-dimensional structure of the native spliceosome by cryo-electron microscopy. *Mol. Cell*, **15**, 833–839.

40. Boehringer,D., Makarov,E.M., Sander,B., Makarova,O.V., Kastner,B., Lührmann,R. and Stark,H. (2004) Three-dimensional structure of a pre-catalytic human spliceosomal complex B. *Nat. Struct. Mol. Biol.*, **11**, 463–468.

41. Sperling,J., Azubel,M. and Sperling,R. (2008) Structure and function of the Pre-mRNA splicing machine. *Structure*, **16**, 1605–1615.

42. Chanfreau,G. and Jacquier,A. (1996) An RNA conformational change between the two chemical steps of group II self-splicing. *EMBO J.*, **15**, 3466–3476.

43. Costa,M., Deme,E., Jacquier,A. and Michel,F. (1997) Multiple tertiary interactions involving domain II of group II self-splicing introns. *J. Mol. Biol.*, **267**, 520–536.

44. 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.

45. Staley,J.P. and Guthrie,C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*, **92**, 315–326.