Published online 6 November 2013 Nucleic Acids Research, 2014, Vol. 42, No. 3 1947–1958
doi:10.1093/nar/gkt1051

Visualizing the ai5γ group IIB intron

Srinivas Somarowthu[1], Michal Legiewicz[1], Kevin S. Keating[1] and Anna Marie Pyle[1,2,3,*]

[1]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06511, USA, [2]Department of Chemistry, Yale University, New Haven, CT 06511, USA and [3]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

Received August 2, 2013; Revised October 2, 2013; Accepted October 12, 2013

ABSTRACT

It has become apparent that much of cellular metabolism is controlled by large well-folded noncoding RNA molecules. In addition to crystallographic approaches, computational methods are needed for visualizing the 3D structure of large RNAs. Here, we modeled the molecular structure of the ai5γ group IIB intron from yeast using the crystal structure of a bacterial group IIC homolog. This was accomplished by adapting strategies for homology and *de novo* modeling, and creating a new computational tool for RNA refinement. The resulting model was validated experimentally using a combination of structure-guided mutagenesis and RNA structure probing. The model provides major insights into the mechanism and regulation of splicing, such as the position of the branch-site before and after the second step of splicing, and the location of subdomains that control target specificity, underscoring the feasibility of modeling large functional RNA molecules.

INTRODUCTION

Group II introns are self-splicing ribozymes and retroelements that are found in bacteria, archaea and eukaryotes (1,2). Understanding the structure and behavior of group II introns is of great biological significance owing to their impact on gene expression and genomic organization in modern organisms and their evolutionary relationship with the eukaryotic spliceosome (3). Group II introns are also significant because they show great potential as gene targeting agents that may ultimately be applicable for gene therapy (4–6). From a biophysical perspective, they serve as excellent model systems for exploring basic principles of RNA folding, structure and catalysis (7,8).

Group II introns are structurally classified into three distinct groupings (Group IIA, IIB and IIC) (9,10). The group IIB introns are of particular interest because they are the most structurally complex, they are exceptionally sequence-specific and they are believed to represent a lineage most closely related to the spliceosome (11). While there is no available structure of a group IIB intron, crystal structures of a group IIC intron have become available (12–14). Group IIC introns lack many of the structural domains and enzymatic capabilities of the IIB introns (9,11,15). However, they represent a useful starting point for visualizing the basic architectural plan of all group II introns (12). Much has been learned from previous attempts to model group IIA (16) and IIB (17,18) introns, but the availability of group IIC intron crystal structures now makes it possible to use homology modeling methods. In the field of protein structural biology, a growing trend is to solve one crystal structure of a protein family, and then obtain structures of the relatives by homology modeling (19). This trend is supported by increasingly powerful tools, which yield accurate models for novel proteins that would be impossible to express or crystallize (20). Inspired by this growing trend in protein homology modeling, we have used a similar approach to build an atomic structural model of the ai5γ group IIB intron (ai5γIIB) from yeast mitochondria (21). To make this possible on such a large multidomain RNA (~800 nt), we adapted existing computational approaches and combined them with a new tool we developed for structure building and refinement of exceptionally large RNAs.

A diversity of excellent computational tools has made it increasingly possible to predict RNA structures from sequence and prior knowledge of architectural organization [see recent reviews (22,23)]. The ability to model large RNAs is of critical importance given the explosion of information on large, highly structured noncoding RNAs that direct everything from chromatin remodeling to translational regulation (24,25). Homology modeling tools such as ModeRNA (26) use sequence alignment and template structures to predict RNA structures. Fragment assembly based methods, such as FARNA (27) and MC-Sym (28), build RNA by assembling nucleotide fragments extracted from crystal structures. NAST

*To whom correspondence should be addressed. Tel: +1 203 436 4047; Fax: +1 203 432 5316; Email: anna.pyle@yale.edu
Present address:
Kevin S. Keating, Schrödinger, Inc, NY 10036, USA.

(29) and iFoldRNA (30) predict structures by performing coarse-grained molecular dynamics simulations. These methods use different input criteria, and there are various advantages and disadvantages associated with each method. For example, homology modeling tools are relatively fast and accurate, but they require template structures with good homology. On the other hand, *de novo* methods do not require template structures but are computationally expensive and therefore limited to small RNAs. Methods for backbone optimization and refinement are nearly nonexistent. Also, current tools are not fully automated and require manual manipulation at various levels, especially to model large RNAs.

To facilitate structural modeling of large RNAs, we expanded the capabilities of RCrane (31,32), which is a program that we originally developed for automated modeling of RNA into electron density, and which is now used for building RNA crystal structures (32). Here, we adapted RCrane for correcting backbone conformations within large RNA 3D models in the absence of electron density. Using this new tool, together with existing homology and *de novo* building methods, we built an all-atom 3D model of a group IIB intron. The resulting structure is consistent with known genetic and biochemical data on IIB architecture, and it reveals fundamentally important new insights into the mechanism of self-splicing.

MATERIALS AND METHODS

Computational methods

A homology model of the core structure was constructed with ModeRNA (26) using the OiIIC crystal structure (PDBID: 3IGI) (12) as a template. Sequence alignment was performed manually with BioEdit (33). Additional subdomains specific to the IIB intron were modeled using MC-Sym (28) (see Supplementary Materials for an example of MC-Sym script) and manually docked onto the core structure with PyMol (Schrodinger, LLC) and Discovery Studio (Accelrys Software Inc). All helices were rebuilt using MC-Sym, replaced using PyMol (Schrödinger, LLC) and Discovery Studio (Accelrys Software Inc) and finally, the backbone was refined with RCrane. The new application of RCrane is included in the current version 1.1, which is available at http://www.pylelab.org/software/index.html, and also with Coot 0.7, which is available from http://www.biop.ox.ac.uk/coot/. The final model was subjected to energy minimization using the AMBER10 force field and generalized Born implicit solvent model available in the AMBER package (34). Minimization was started with 500 steps of steepest descent followed by conjugate gradient minimization for at least 500 steps. The quality of resultant structure was assessed using Molprobity (35) and the number of minimization cycles was increased in cases of high clash score. See supplementary Table S1 for complete list of software and additional details.

RNA constructs and purification

Plasmid pSS01 encoding the β-β′ variant (stem loop #252–264 replaced with UUCG tetraloop) was constructed from pQL71 using the QuickChange site-directed mutagenesis kit. Both plasmids pQL71 (wild type) and pSS01 (β-β′ variant) were linearized with HindIII before transcription and purified as described before (18).

Hydroxyl radical footprinting

RNA samples (4 pmoles) were incubated in a buffer containing 25 mM potassium cacodylate, pH 7.0, 500 mM KCl and 0.2 mM EDTA at 90°C for 2 min and then cooled to room temperature. Folding was initiated by incubating RNA (100 µl of final volume) at 42°C for 30 min in a buffer containing 25 mM potassium cacodylate, pH 7.0, 500 mM KCl and 100 mM $MgCl_2$ (or water instead of $MgCl_2$ for cleavage in the absence of Mg^{2+} ions). Footprinting reactions were initiated by applying 2 µl of each 2.5 mM $(NH_4)Fe(SO4)_2$, 2.75 mM EDTA (pH 8.0), 1.5% H_2O_2 and 50 mM sodium ascorbate to the inside wall of the tube followed by vortexing the sample. After 15 s, the reactions were quenched with 20 µl of stop mix (containing 100 mM thiourea and 200 mM EDTA). RNA was precipitated with 2.5 volumes ethanol (100%), 10% 3 M sodium acetate (pH 4.8) and 0.5 µl of glycogen (Roche).

SHAPE

For selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) experiments, 4 pmoles of RNA (in 40 µl of buffer containing 500 mM KCl, 50 mM HEPES pH 7.4, 0.1 mM EDTA) were denatured at 90°C for 2 min and then cooled to room temperature. Folding was initiated by incubating RNA (150 µl of final volume) at 42°C for 30 min in a buffer containing 500 mM KCl, 50 mM HEPES pH 7.4, 0.1 mM EDTA and 100 mM $MgCl_2$ (or water instead of $MgCl_2$ for probing in absence of Mg^{2+} ions). RNA was divided equally (72 µl) between two tubes, and 8 µl of Dimethyl sulfoxide (DMSO) (for control) or 10 mM 1M7 (1-methyl-7-nitroisatoic anhydride) (36) were added. The modification reaction was allowed to proceed for 5 min at 37°C, and samples were precipitated with 2.5 volumes ethanol (100%), 10% 3 M sodium acetate (pH 4.8) and 0.5 µl of glycogen (Roche).

Fragment analysis

Cleavage products from both hydroxyl radical footprinting and SHAPE were analyzed by primer extension using fluorescently labeled primers. The fluorescent dye coding and instrument precalibration was applied as described (37). Briefly, primers with dyes (Anaspec) 6-JOE (reagent present) and 5-FAM (reagent absent) were used for reverse transcription, and those with 6-TAMRA (ddT) and 5-ROX (ddC) were used for cycle sequencing (Affymetrix cycle sequencing kit). Each primer sequence was corrected for mobility shift in ShapeFinder. Primers for reverse transcription were annealed by incubating 2 pmol RNA with 1 µl of 2 µM primer (12 µl of total volume) at 95°C for 3 min, 4°C for 5 min and 42°C for

2 min. Reverse transcription was initiated by adding 8 µl of mixture containing 4 µl of 5× First strand buffer, 1 µl of 100 mM DTT, 1 µl of 10 mM dNTPs, 0.5 µl of Superscript III (U/µl) and water at 48°C for 45 min. The (+) and (−) reagent reactions were quenched and combined followed by ethanol precipitation. Recovered cDNAs were resuspended in deionized formamide and mixed with sequencing ladders and were analyzed on ABI 3730xl DNA analyzer. Data were analyzed with ShapeFinder (38) as described before (39).

RESULTS

Three-dimensional modeling of a group IIB intron

Homology modeling

The core region of the ai5γIIB (Figure 1) was modeled with ModeRNA (26), using a published crystal structure of the *Oceanobacillus iheyensis* IIC intron (PDBID: 3IGI) (12) as the template (OiIIC). Sequence alignment was performed manually using BioEDIT (33). Before performing the alignment, regions of the IIB intron that are not found in the OiIIC template were removed, including D1c2, D1d2a, D1d2b, D3a, D3b, D3c and D6. The lengths of several helices and loops vary between the two introns, and these regions were left as gaps within the sequence alignment. All the helices were later rebuilt as idealized A-form helices using MC-Sym (28). Among the various helices generated by MC-Sym, the helix that fits best with the existing helix in the core homology model was selected and replaced using PyMol and Discovery Studio (DS) Visualizer. Long-range interactions with different structural forms were adjusted manually: For example, the ζ-ζ′ interaction (GAAA tetraloop-receptor) (40,41) was modeled from a similar motif taken from the PDB (PDBID: 2JYF). The α-α′ kissing-loop interaction was modeled as a short helix and replaced using PyMol and DS Visualizer. The resulting model consists of a full-length D5, the basal stem of D3, a truncated D2 and D4 and all elements of D1 except D1c2a, D1d2a and D1d2b. The model coordinates are available in the Supplementary Material.

De novo modeling

D6 and various subdomains that are only found within the ai5γIIB intron, such as D3a, D3b, D3c, D1c2a, D1d2a

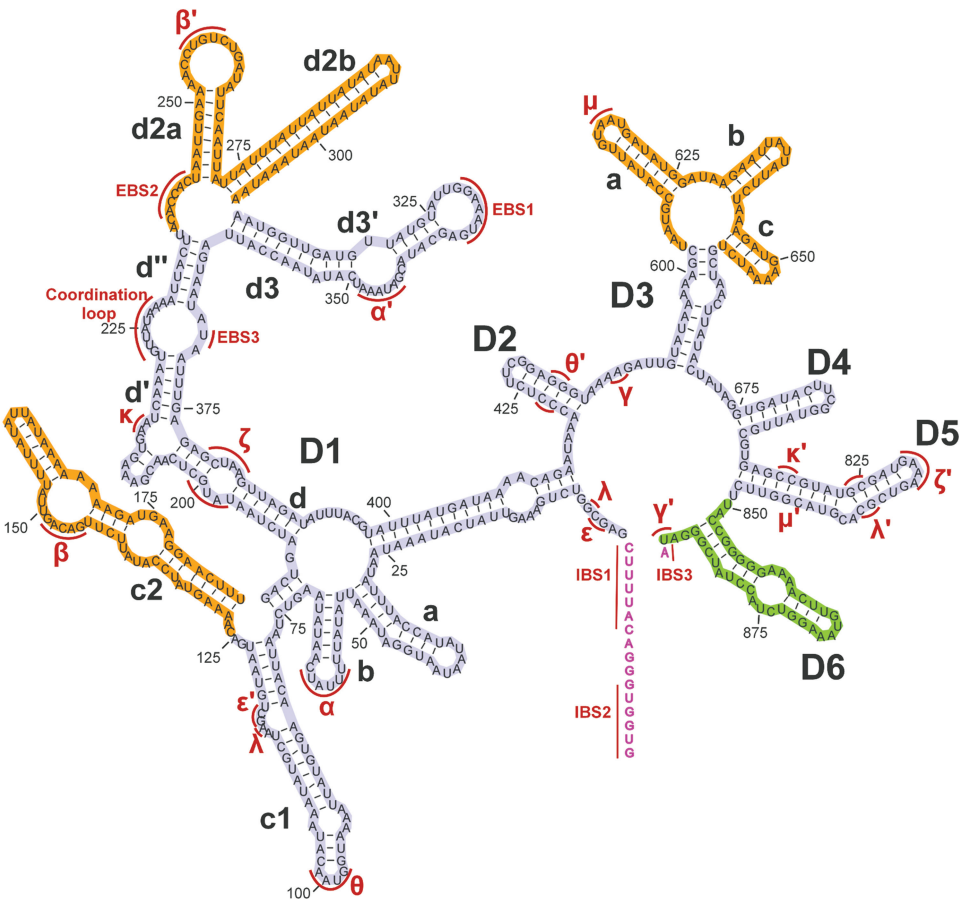

Figure 1. Schematic secondary structure of ai5γIIB intron. Domains 1–6 are labeled in capital letters, subdomains are labeled in small letters. Exons are colored in magenta. Exon and intron binding sites are labeled as EBS and IBS, respectively. Regions highlighted in light blue are common in ai5γIIB and OiIIC introns (core region); they were built with homology modeling using OiIIC crystal structure as the template. Additional insertions specific to ai5γIIB (highlighted in orange) were built *de novo* and docked on to the homology model. D6 (highlighted in green), which is missing in all available crystal structures of OiIIC, was also modeled *de novo*. Long-range tertiary interactions are labeled in red.

and D1d2b, were modeled *de novo* using MC-Sym (28) and then docked manually onto the core model using PyMol and DS visualizer (Figure 2). For a given secondary structure, MC-Sym generates an ensemble of decoy structures using fragment-based libraries. As the input for MC-Sym, the domain to be modeled was appended to a known region that was already part of the core model, and which served as an anchor for later docking steps. The ensemble of structures predicted by MC-Sym were clustered, docked onto the core model and finally filtered based on known biochemical data.

To model D6 in the context of the intron core, our strategy involved sampling several possible conformations of D6 that are sterically allowed, but which do not significantly alter the structure of the intron core. D6 is covalently linked to D5 via a 3-nt linker, and both domains are part of a higher-order six-way junction. Five helices of this six-way junction are already present in the homology model. In the absence of the other helices, D5 and D6 can obviously sample many conformations; however, relatively few of these are sterically allowed when they are part of a six-way junction. We predicted conformations of D5 and D6 as a two-way junction using MC-Sym. The resulting conformations were clustered and docked by aligning D5 with the corresponding region of the core model, and then filtered to remove conformations with steric clashes. From the remaining conformations, we selected one conformation where the bulged adenosine of D6 is proximal to the active site. We then manually adjusted D6 such that the 2′-OH group of bulged adenosine occupies the position of a nucleophilic water that had been visualized crystallographically in structures of the OiIIC intron (42). Additional subdomains D3a, D3b, D3c, D1c2a, D1d2a and D1d2b were modeled in a similar fashion (see Supplementary Results).

Model refinement

Backbone refinement: a novel application of RCrane

The RNA backbone is highly flexible, and given that seven torsion angles describe each individual nucleotide, computational modeling can be a daunting task (43). Previously, we developed an alternative conformational description that reduces the dimensionality of RNA structure. Using this approach, each nucleotide is represented by two pseudotorsions η and θ, where η represents torsion between $C4'_{i-1}$, P_i, $C4'_i$ and P_{i+1}, and θ represents torsion between P_i, $C4'_i$, P_{i+1} and $C4'_{i+1}$ (44). Further, using both the RNA pseudotorsions and the RNA backbone conformers (45) we developed the program RCrane, which is a computational tool for semiautomated model building of RNA into electron-density maps (31,32). Here, we developed a novel application of RCrane, using it to automatically correct the backbone configurations in the models even before an electron-density map is available. In the absence of an electron-density map, RCrane first predicts the appropriate backbone conformer based on the positions of phosphate and base, and then subsequently improves the placement of both the phosphate and the backbone sugar. This diversification of RCrane's capabilities demonstrates that the approach has additional

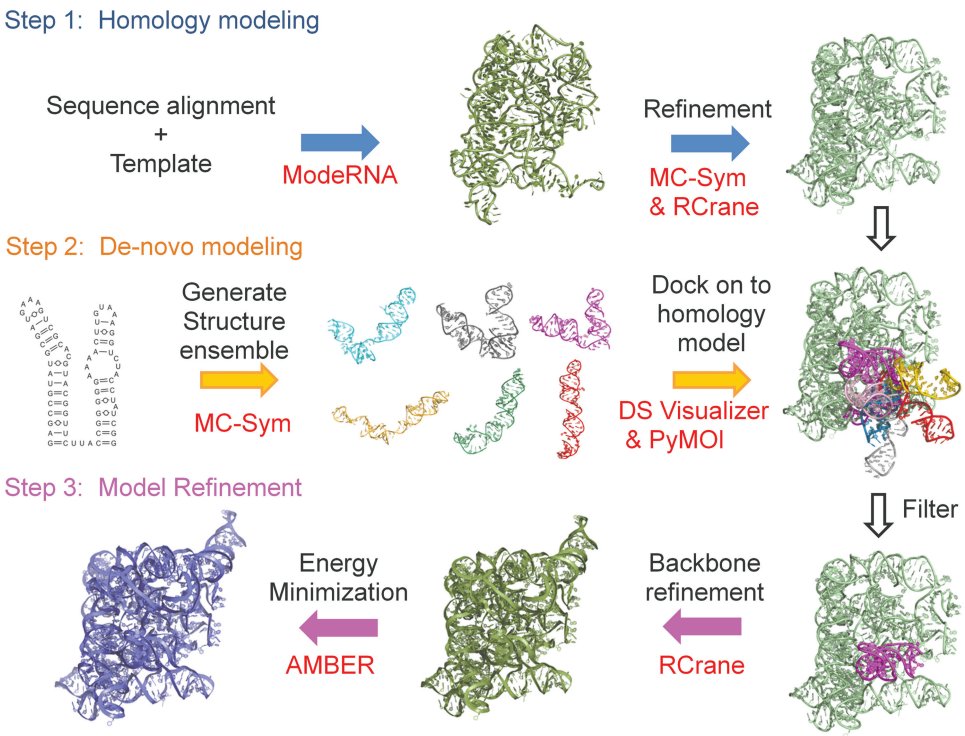

Figure 2. Overview of model building. Modeling was performed in three steps, in the first step the core structure of the intron was generated with homology modeling using crystal structure from IIC as the template. Next, all additional regions specific to the IIB intron were modeled using MC-Sym and docked onto the core structure. Finally, the model was refined using RCrane (see text) and AMBER.

applications in building, modeling and analyzing RNA structures. This new application is included in the current version of RCrane 1.1 available at http://www.pylelab.org/software/index.html and also with Coot 0.7 available from http://www.biop.ox.ac.uk/coot/.

We used RCrane to refine the backbone of the ai5γIIB model at various stages of the modeling process (see Figure 2). The resulting model was subjected to energy minimization with AMBER (34) to remove steric clashes, resulting in an overall clash score that is well within the acceptable range (0.2, 99th percentile) (35).

Model validation

The resulting computational model was validated experimentally using hydroxyl radical footprinting. Hydroxyl radical footprinting probes the solvent-accessible regions of a molecular structure (46). Recent work has shown that there can be good correlation between hydroxyl radical reactivity and number of neighboring nucleotides (47) in the crystal structures of various RNAs. In this work, we calculated the number of contacts of each nucleotide in the model and correlated with hydroxyl radical reactivity.

First, we performed hydroxyl radical footprinting followed by primer extension and fragment analysis with capillary electrophoresis on D135, which is an ai5γIIB construct that contains full length D1, D3 and D5, a truncated form of D2 and D4, while it lacks D6 and both exons. To be consistent with the experiment, the exons and D6 were removed from the model before calculating the number of contacts. Numbers of contacts for each nucleotide in the model were calculated as described previously (47), using 20 Å distance cutoff. There is good correlation between numbers of contacts in the model and hydroxyl radical reactivities (Figure 3A). The observed correlation coefficient ($r = -0.61$) is as good as that observed for crystal structures of other RNAs (47) and it is also statistically significant with $P < 10^{-24}$ (Figure 3A).

The locations of D1c2 and D1d2a (stems containing the loops that form the β-β′ interaction) were also validated with mutation and hydroxyl radical footprinting. Previous results have shown that both D1c2 and D1d2a are exposed to solvent, and they are believed to lie at the periphery of D1 (18). However, disrupting the kissing-loop interaction between these stems should disengage them and expose regions that lie protected beneath these stems. To test this hypothesis, and thereby test the model, we disrupted the β-β′ kissing-loop by replacing the D1d2a stem loop (nucleotides 252–264) with a UUCG tetraloop (the β-β′ mutant).

Hydroxyl radical footprinting was then performed on both the WT D135 and the β-β′ mutant. In the case of the β-β′ mutant, both the D1c2 and the D1d2a stems showed significant decreases in hydroxyl radical protection (Figure 3B). Disruption of the β-β′ kissing-loop therefore causes the two stems to become more dynamic in solution and more solvent exposed relative to their position in WT. Significant decreases in hydroxyl radical protection were also observed in regions of D1b (57–59), D1d'' (232–236), D1d3' (326–328), D3b (632–634, 638–644), D3c (655–657) and D4 (808–813). Among these regions, D1b (57–59), D1d'' (232–236) and D1d3' (326-328) lie beneath the interacting D1c2 and D1d2a stems in the model, indicating that mutation has exposed them and that the model is in good agreement with the experimental data (Figure 3C).

Interestingly, there are a few regions that are more protected in the mutant construct: D1d' (216–218), D1d2b (302–306), D1d3 (357–365), D1d (386–391, 394–398) and D1 (i) (404–408). Most of these regions are in the vicinity of D1c2 or D1d2a, suggesting that when the β-β′ kissing-loop is disrupted, D1c2 and D1d2a do not simply move around randomly. Rather, they may form new types of discrete stabilizing interactions.

To ensure that the observed increases in reactivity for the β-β′ mutant were not simply a result of localized RNA unfolding, we performed SHAPE experiments to evaluate whether secondary structural elements are still intact in this mutant. As expected, nucleotides (145–149) involved in the β-β′ kissing-loop interaction showed a significant increase in SHAPE reactivities, as the other half of this kissing-loop pair was replaced with a tetraloop (Supplementary Figure S1). This local reactivity increase was consistent and comparable with the SHAPE reactivity differences in WT when probed in the presence of 100 mM Mg^{2+} (folded, β-β′ interactions formed) and without Mg^{2+} (not folded, β-β′ interaction not formed). For the remaining regions of the mutant, there were no significant changes in SHAPE reactivity when compared with WT, suggesting that the secondary structure has not been affected by the mutations (Supplementary Figure S1). Taken together, these results support the location of D1c2 and D1d2a in our model.

Finally, to validate the location of D3, we compared previously published experimental footprinting data on the ai5γIIB D135 construct with data on a D15 construct that lacks D3 (48). On deletion of D3, regions in contact with D3 should become solvent exposed and are expected to be less protected in D15 construct compared with D135 construct. Again, there is good agreement between the model and experimental data, as most of the regions that are less protected in the D15 construct are located beneath D3 in the model (Supplementary Figure S2).

Structural architecture of ai5γIIB

As expected, the active-site region and the overall architecture of the ai5γIIB model is similar to the known structure of OiIIC. However, IIB introns possess additional subdomains and tertiary interactions (such as β-β′, μ-μ′ and EBS2-IBS2) that are important for both folding and catalysis (11,49). Unlike IIC introns, IIB introns react with high sequence-specificity and undergo branching efficiently. The ai5γIIB model allowed us to visualize the additional subdomains and thereby provided valuable hints on functional roles of structural features specific to IIB introns. For example, the β-β′and μ-μ′ tertiary interactions form on opposite sides of D5 (Figure 4 and Supplementary Movie S1). D1c2 and D1d2a (the stems supporting the β-β′ interaction) are located on the 'catalytic face', while D3 falls on the 'binding face' (50) of D5. D1c2 and D1d2a protrude like two extra arms from

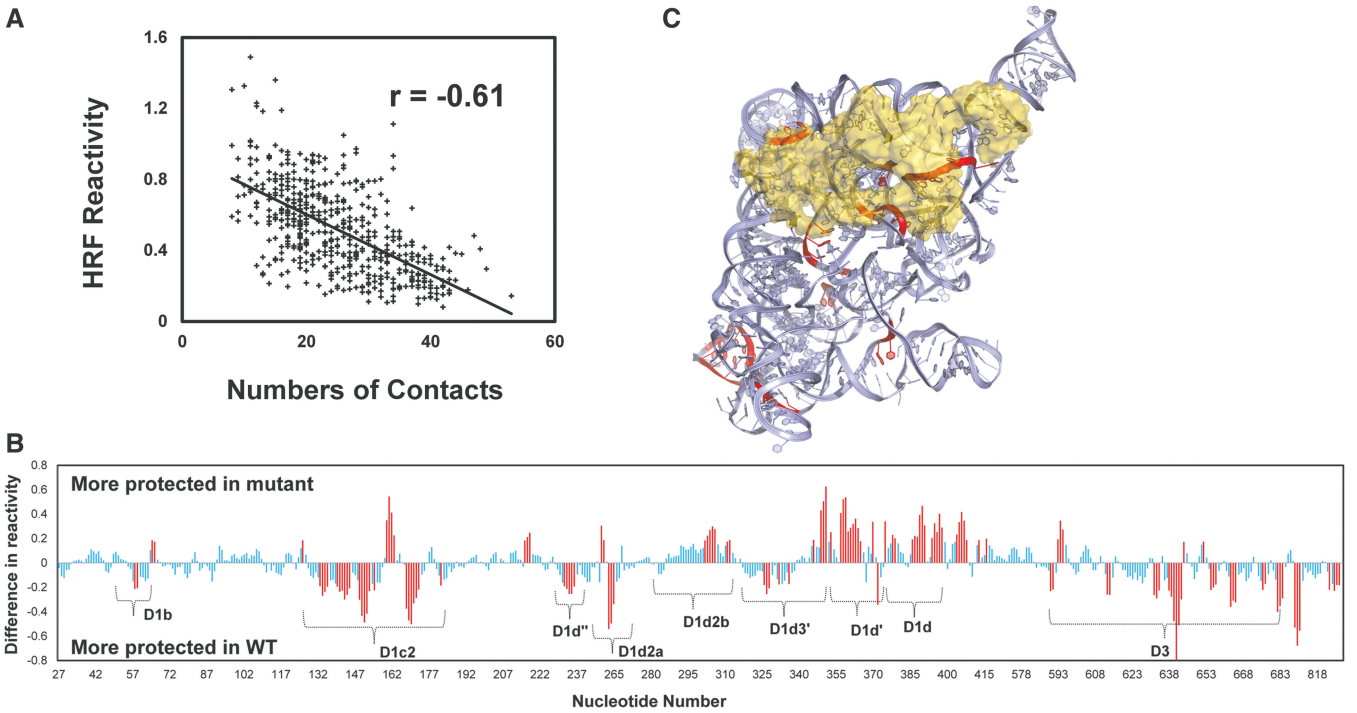

Figure 3. Model Validation. (**A**) Correlation plot between numbers of contacts of each nucleotide in the model with hydroxyl radical reactivities. The correlation coefficient (r = −0.61) is as good as that observed for crystal structures of other RNAs such as group I introns, RNase P and riboswitches (47). (**B**) Difference plot of hydroxyl radical reactivities between WT D135 and β-β′ mutant. Nucleotides with significant increases or decreases (greater than one unit of standard deviation) in reactivities are colored in red. (**C**) D1c2 and D1d2a stems are shown as a transparent surface representation colored in yellow. Regions with significant increase in hydroxyl reactivities are colored in red. The rest of the intron is colored in light blue. Most of the regions with significant increases in reactivities lie adjacent to D1c2 and D1d2a, indicating that mutation of the kissing-loop interaction between D1c2 and D1d2a has exposed the regions beneath them and that the model is in good agreement with the experimental data.

opposite sides of D1, and the β-β′ kissing-loop interaction between these hairpin loops interlocks like a bridge that is likely to rigidify the scaffold of D1 (Figure 4A). On the other side of D5, D3a and D3b form a wall along the 'binding face' that may mediate interactions between D5 and the rest of the intron (Figure 4B and C).

The second exon binding site, or EBS2, is a major distinguishing feature of the highly derived IIB and IIA introns. The model shows that EBS2 is far more solvent exposed than EBS1, which is consistent with footprinting data on ai5γIIB intron (18). The model also shows that the EBS1-IBS1 and EBS2-IBS2 helices, which form on target or exon binding, are not coaxial with each other, consistent with previous spectroscopic and mutational analysis of the ai5γIIB intron (51) and the model of *Pylaiella littoralis* intron (Pl.LSU/2) (17). We see that the EBS2-IBS2 helix is located immediately adjacent to the β-β′ interaction (Figure 5), suggesting that function of the two motifs may be linked. Importantly, the modeling process revealed that, when no exon is bound, the EBS2 region and adjacent helices are flexible. This indicates that exon binding, and formation of the EBS2-IBS2 helix, rigidifies EBS2 and the surrounding substructures. Thus, formation of the EBS2-IBS2 helix induces conformational rearrangements that are likely to considerably reduce overall entropy, with important implications for the known role of EBS2-IBS2 in substrate specificity and reaction chemistry (51,52).

Among the subdomains that are specific to IIB introns, D1d2b is the only substructure that appears to be pointed away from the intron core and has no contacts with the rest of the intron. However, D1d2b is coaxial with D1d2a, and it may therefore play an indirect role in the formation of β-β′ kissing-loop between D1d2a and D1c2.

The active and silent conformers of D6

D6 can exist in two different conformations (Figure 6A). In the conformation capable of branching, (the active form) the 2′-OH group of the branch-site adenosine is positioned precisely to initiate the first step of splicing within the active site (Figure 6B). After the first step, the branch-point nucleotide flips out of the active site, resulting in an alternate conformation (the silent form) (11). The toggling of D6 between active and silent forms is likely to represent an important regulator of splicing in group II introns, and potentially, in the spliceosome as well, for cognate domains (53). To prepare the active site for second step of splicing, and to position the 3′-exon appropriately, D6 will need to exit from the active site. However, cross-linking studies (54,55) have shown that nucleotides in D5 and D6 are located approximately in the same position through the entire cycle of splicing, which is consistent with studies showing that the same active-site elements play a role in both steps of group II intron splicing (56).

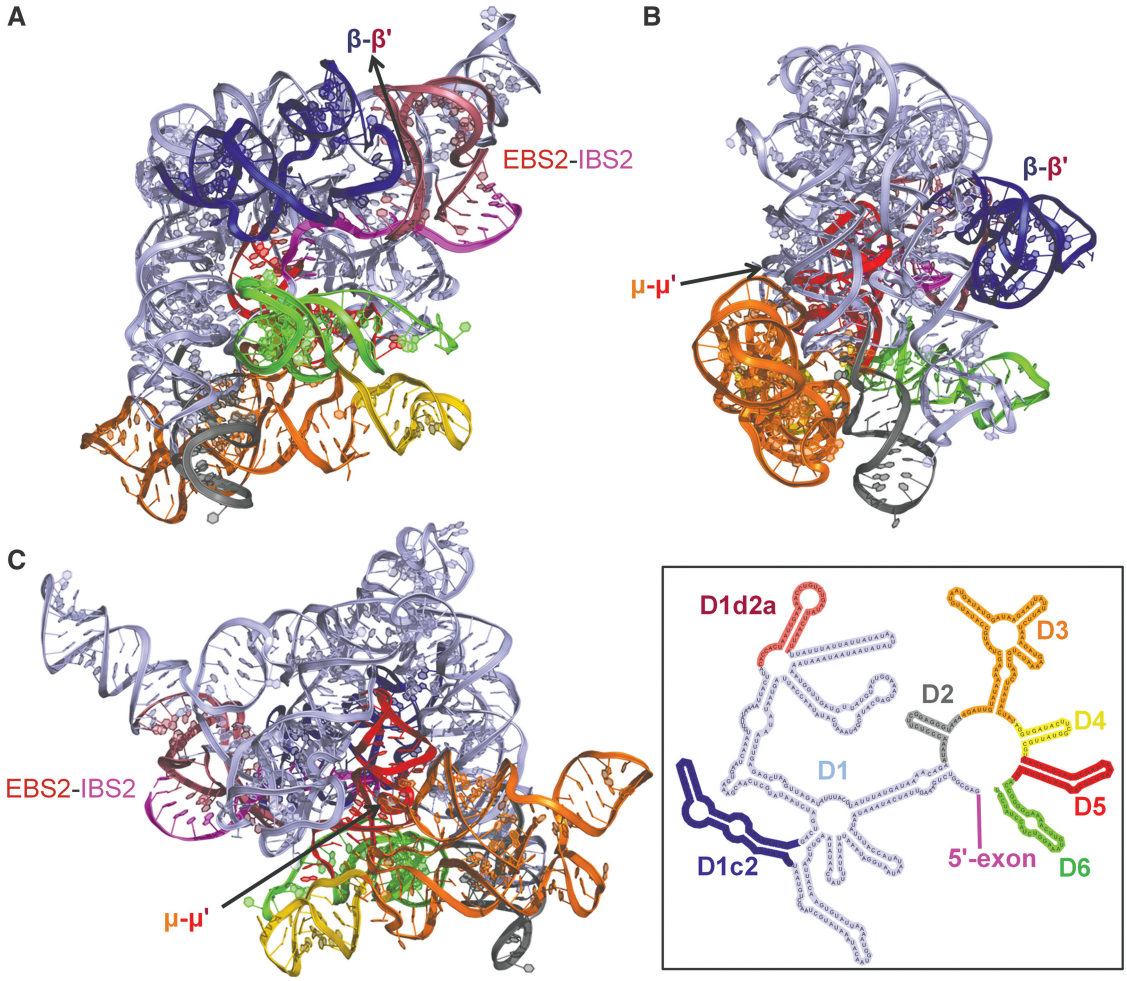

Figure 4. Structural architecture of ai5γIIB. (**A**) Cartoon representation of the model depicting the overall architecture (also see Supplementary Movie S1 and Supplementary Figure S4). All domains and subdomains are color-coded as shown in the insert. Locations of three additional interactions, β-β′, μ-μ′ and EBS2-IBS2 are labeled and colored the same as domains involved in that particular interaction. D1c2 (shown in blue) and D1d2a (shown in raspberry) are the two additional subdomains involved in the β-β′ kissing-loop interaction. (**B**) 90° rotated view of the image shown in (A). (**C**) 180° rotated view of the image shown in (A).

Understanding this paradox is of utmost importance, and various studies already attempted to model D6 in active form (16,57). Here, we modeled both active and silent forms of D6 and sought to visualize its conformational toggling. The location of D6 in active form is consistent with recent studies on the *P. littoralis* intron (Pl.LSU/2) (57). The silent conformation was modeled such that the 3′-exon was positioned within the active site, where it forms the EBS3-IBS3 interaction and the γ-γ′ interaction between the last nucleotide of the intron (U887) and residue A587 (Figure 6C). Surprisingly, the overall spatial location of D6 is similar in the active and silent forms (Figure 6A). Close comparison of the two states reveals that D6 undergoes more of a rotation than a translation along its helical axis on toggling, so that the global position of D6 with respect to the intron remains similar, but individual nucleotides within D6 shift their position. This is consistent with the fact that, after the first step of splicing, D6 cannot move too far from the intron core, as the branched adenosine is now covalently linked (2′–5′ lariat bond) with 5′-end of the intron. Further, we noticed that this conformational change is sufficient to support formation of the η-η′ interaction between D6 and D2 (Supplementary Figure S3). The crystal structure of OiIIC contains only the basal stem of D2, while the D2 stem that participates in η-η′ is longer, and part of a four-way junction in ai5γIIB intron. Surprisingly, we were able to model this junction in a way that allows formation of η-η′ by moving D6 only slightly from the 'active' conformation and without disturbing the position of D2 basal stem and the θ-θ′ interaction. It is interesting to point out that the tip of D6 in 'active' form is not too far from the receptor of η-η′ interaction in D2, suggesting that D2 may play a role in positioning D6. This is consistent with previous studies showing that mutations of η-η′ affect exon ligation (58).

DISCUSSION

Here, we report an all-atom model of an intact ai5γIIB, which was built using a combination of homology and *de novo* methods. This approach has enabled us to visualize

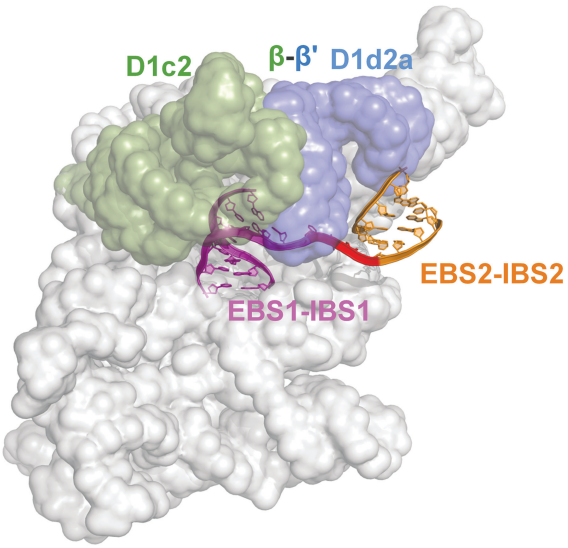

Figure 5. Location of the EBS2 binding site. The EBS2-IBS2 duplex is shown in orange, the EBS1-IBS1 duplex in shown in magenta. The linker between IBS1 and IBS2 is shown in red. The rest of the intron is shown as a transparent surface representation. The EBS2-IBS2 interaction is located next to the kissing-loop (β-β′) interaction that joins D1c2 (green) and D1d2a (blue), suggesting that function of the two motifs may be linked.

key group II intron structural features, such as the two conformations of intron D6, and the EBS-IBS2 specificity helix. The relative locations and orientations of these subdomains, along with structural innovations such as β-β′, explain many of the functional attributes unique to group IIB introns, such as their extreme sequence-specificity and their propensity to splice through branching.

One of the features most readily apparent from the model is that D5 is surrounded by a larger and more protective shell in the ai5γIIB intron relative to OiIIC (Figure 7). This well-packed IIB intron scaffold, along with added interactions such as β-β′ and μ-μ′, is likely to affect the overall dynamics of the intron, influence the electrostatic potential at the active site and provide additional stability to the active site as it assembles on the surface of D5. Previous studies of ai5γIIB have shown that D5 binds so tightly to rest of the intron (59,60) that it can be added as a separate domain to truncated variants of the intron (59–61).

Perhaps most importantly, the model provides insights into the branching mechanism of group II introns. Splicing through branching is one of the most notable features of group II introns, but it has remained difficult to visualize because the existing IIC crystal structures lack electron density for D6, and IIC introns do not readily

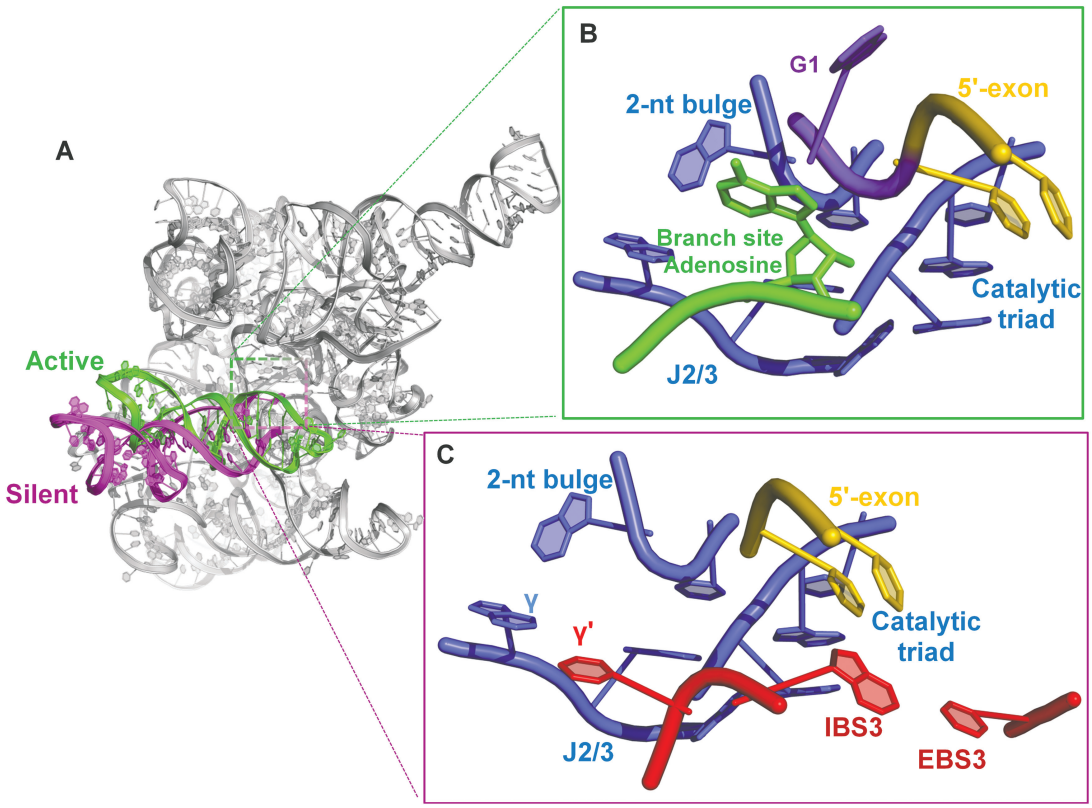

Figure 6. Active versus silent form of D6: (**A**) D6 can exist in two different conformations: the active (green) and silent forms (magenta). (**B**) In the conformation needed for the first step of splicing (the active form) the 2′-OH group of branch-site adenosine (green) is positioned in the active site to initiate the first step of splicing. Active-site elements (catalytic triad, 2-nt bulge and J2\3 junction) are shown in blue, 5′-exon is shown in yellow, first nucleotide of the intron G1 is shown in purple. (**C**) In the conformation that forms after the first step of splicing, and which is required for the second step (the silent-form), the branch-site adenosine flips out and positions the 3′-exon in the active site and forms two new interactions (shown in red) IBS3-EBS3 and γ-γ′. Based on the model, toggling between the 'active' and 'silent' conformations of D6 involves a simple rotation of the domain rather than a large translational movement away from the active site.

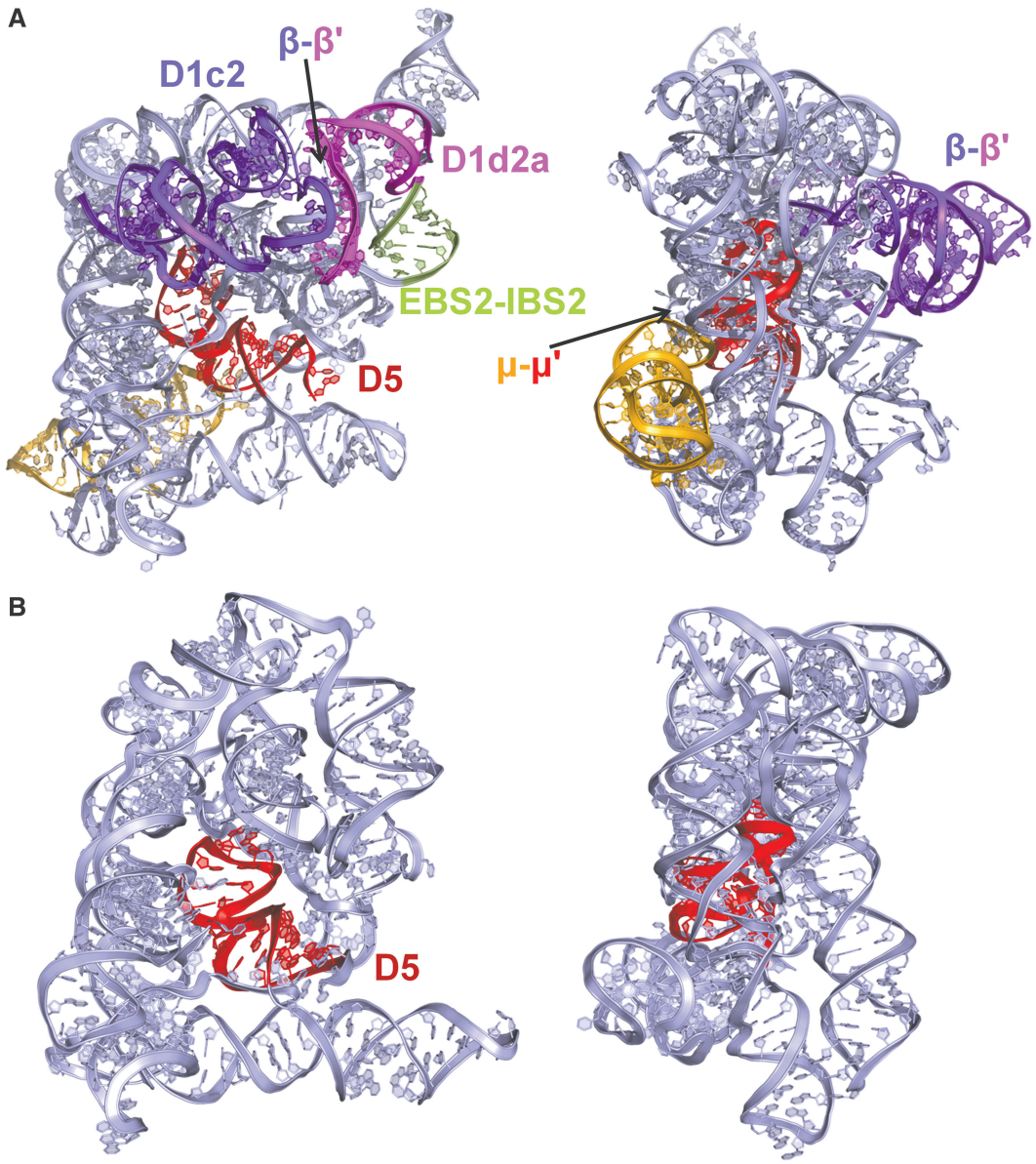

Figure 7. Comparison of the ai5γIIB model with the crystal structure of OiIIC. (**A**) Front and side views of the ai5γIIB model. Regions common to both introns are colored in light blue; D5 is colored in red. Long-range interactions that are not present in OiIIC are labeled in Greek letters. D1c2 is shown in purple, D1d2a is shown in magenta, D3a and D3b are shown in orange. EBS2-IBS2 helix is shown in green. For comparison with OiIIC, D6 is not shown in both views. (**B**) Front and side views of OiIIC crystal structure.

branch (they splice only through hydrolysis *in vitro*) (62,63). Previous studies have proposed two candidate receptors of D6 for branching: D1c1 (57) and the coordination loop (55). In the model provided here, D6 (in the active form) is located close to D1c1; however, the coordination loop is also in the vicinity of D6, and it is possible that D6 interacts with the coordination loop before the branch-site adenosine is recruited into the active site. During the modeling process, we observed that the presence of additional subdomains (D1c1 and D1d2a) restrict the possible conformations for D6 in ai5γIIB relative to D6 in OiIIC, which is likely to enhance branching relative to hydrolytic attack at the 5′-splice site. In addition, we observe that toggling between the 'active' and 'silent' conformations of D6 involve a simple rotation of the domain rather than a large translational movement away from the active site as proposed previously (11), thereby explaining the many biochemical studies, showing that intron domains do not radically reorganize between the two steps of splicing (54) (Figure 6).

The EBS2-IBS2 interaction is another distinguishing feature of group IIB and IIA introns, and the model helps to show why this structural innovation has come to play such an important role in intron function. There are introns, such as *Clostridium beijerinckii* I2 (64) and *Pelotomaculum thermopropionicum* strain SI I2 (64), which contain an EBS2-IBS2 interaction with no apparent role in self-splicing (64,65). When we compared

the secondary structures of those introns with ai5γIIB, we noticed that the β-β′ interaction is absent in *C. beijerincki* I2 (64), *P. thermopropionicum* strain SI I2 (64) and *Sinorhizobium meliloti* RmInt1 (65). In the model, the β-β′ interaction lies in a cleft between the EBS1-IBS1 and EBS2-IBS2 duplexes (Figure 5), consistent with a known cross-link between IBS1 and D1c2 (66) (one of the stems involved in β-β′). Thus, a functional EBS2-IBS2 duplex may require the presence of β-β′ and the two may act synergistically to rigidify the overall intron structure and promote specificity. Another important observation is that EBS2 is more solvent-exposed than EBS1. This suggests that EBS2 may interact with its cognate IBS2 site first, triggering a conformational change that makes EBS1 more accessible, and then leading to EBS1-IBS1 formation. Such a mechanism might be important during specific retro-homing events, whereby the intron protects EBS1 from binding to nonspecific sites, but allows binding at EBS1 to occur only after a specific EBS2-IBS2 duplex is formed. Finally, the lack of coaxial stacking between the EBS-IBS duplexes explains the extraordinary sequence-specificity of group IIB introns. Because each duplex is fully independent, and the overall exon interaction energy is not supplemented by coaxial stacking, the recognition helices are less tolerant of mismatches (51), leading to the large specificity index for target site recognition that is observed for group IIB introns (51). The specificity index may also be enhanced by the insertion of β-β′ into the cleft between the EBS-IBS duplexes (Figure 5), as the resultant steric clashing may give rise to the energetic penalty that is known to accompany binding of an oligonucleotide to both EBS1 and EBS2 simultaneously. In fact, it is known that binding of long oligonucleotide targets (which bind both EBS1 and EBS2) is ∼8 kcal/mol less favorable than binding of short oligos to EBS1 and EBS2 individually (52).

Although the RNA computation field has significantly progressed in the past decade, it was necessary for us to create new tools to complete and refine the model of the IIB intron. Here, we developed a broadly applicable computational tool for automatically correcting the backbone configurations in the models even before an electron-density map is available. This new tool, which is part of the RCrane program (32), will be helpful in modeling large RNA structures in conjunction with existing RNA modeling methods. Several features of RCrane as a 3D modeling tool are notable: First, existing, homology modeling methods (including those used for protein structures) derive models by copying coordinates from a template structure, which is a process that often leads to improper backbone configurations and steric clashes, especially in regions with low sequence similarity. RCrane rebuilds the entire backbone while keeping the position of the bases constant, and thereby provides the unique advantage of preserving secondary structure throughout refinement. In addition, large RNAs like group II introns may use different structural motifs but still share similar global structures, and RCrane preserves this information content. For example, the tetraloop and receptor interaction of D5 in IIC introns has a different structural form compared with other group II introns (40) and various long-range interactions also display slight differences, such as the number of base pairs involved in a kissing-loop. Currently, no methods exist to automatically correct for such differences, and users would have to manually adjust or rebuild these regions. By contrast, RCrane semiautomates this process, allowing the user to adjust or rebuild the nucleotide base, while RCrane rebuilds the backbone accordingly and ensures that it conforms to allowed rotameric states. RCrane can also be used to enhance *de novo* modeling methods, such as MC-Sym, which build structures by assembling fragments from nucleotide libraries. This process often results in structures with poor backbone connectivity. Again, RCrane can be used to fix backbone inconsistencies in these models without disrupting the secondary structure. In this work, we used RCrane at various stages to successfully build the model of ai5γIIB intron.

The ability to create a detailed homology model of a large multidomain RNA is of particular interest at this time, as it is becoming apparent that large RNAs are central to most aspects of gene expression in biology (25). There is also increasing evidence that large noncoding RNAs are, in many cases, highly structured (24). Since it is not currently feasible to experimentally determine structures for a majority of these RNAs, there will be an increasing demand for computational methods to predict RNA structures. The modeling strategies used here can be applied to other group II introns and to other multidomain RNA molecules for which empirical structural data are available.

In conclusion, we have shown that it is now possible to model large RNAs even from remote homologs. We report and test a 3D model of the ai5γIIB intron that provides structural insights into the mechanistic behavior of group II introns and explains functional differences between group IIC introns and more evolutionarily derived group II introns such as ai5γIIB.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online, including [9,11,26,28,32–35,48,49, 67].

ACKNOWLEDGEMENTS

The authors thank all members of Pyle lab, in particular Dr Isabel Chillón Gázquez, Dr Marco Marcia, Dr Olga Fedorova, Dr Maximilian Bailor and Dr Laura Murray for many helpful discussions. They also thank David Rawling, Dr Isabel Chillón Gázquez and Dr Marco Marcia for critical reading of this manuscript. A.M.P. is an investigator of Howard Hughes Medical Institute.

FUNDING

National Institutes of Health [RO1 GM50313]. Funding for open access charge: Howard Hughes Medical Institute.

Conflict of interest statement. None declared.

REFERENCES

1. Ferat,J.L. and Michel,F. (1993) Group II self-splicing introns in bacteria. *Nature*, **364**, 358–361.
2. Valles,Y., Halanych,K.M. and Boore,J.L. (2008) Group II introns break new boundaries: presence in a bilaterian's genome. *PLoS One*, **3**, e1488.
3. Gordon,P.M., Sontheimer,E.J. and Piccirilli,J.A. (2000) Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns. *RNA*, **6**, 199–205.
4. Guo,H., Karberg,M., Long,M., Jones,J.P. 3rd, Sullenger,B. and Lambowitz,A.M. (2000) Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science*, **289**, 452–457.
5. Lambowitz,A.M. and Zimmerly,S. (2011) Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.*, **3**, a003616.
6. García-Rodríguez,F.M., Barrientos-Durán,A., Díaz-Prado,V., Fernández-López,M. and Toro,N. (2011) Use of RmInt1, a group IIB intron lacking the intron-encoded protein endonuclease domain, in gene targeting. *Appl. Environ. Microbiol.*, **77**, 854–861.
7. 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.
8. Donghi,D., Pechlaner,M., Finazzo,C., Knobloch,B. and Sigel,R.K. (2013) The structural stabilization of the kappa three-way junction by Mg(II) represents the first step in the folding of a group II intron. *Nucleic Acids Res.*, **41**, 2489–2504.
9. Toor,N., Hausner,G. and Zimmerly,S. (2001) Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA*, **7**, 1142–1152.
10. Michel,F., Kazuhiko,U. and Haruo,O. (1989) Comparative and functional anatomy of group II catalytic introns — a review. *Gene*, **82**, 5–30.
11. 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.
12. 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.
13. Marcia,M. and Pyle,A.M. (2012) Visualizing group II intron catalysis through the stages of splicing. *Cell*, **151**, 497–507.
14. Marcia,M., Somarowthu,S. and Pyle,A.M. (2013) Now on display: a gallery of group II intron structures at different stages of catalysis. *Mob. DNA*, **4**, 14.
15. Rest,J.S. and Mindell,D.P. (2003) Retroids in archaea: phylogeny and lateral origins. *Mol. Biol. Evol.*, **20**, 1134–1142.
16. Dai,L., Chai,D., Gu,S.-Q., Gabel,J., Noskov,S.Y., Blocker,F.J.H., 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.
17. Costa,M., Michel,F. and Westhof,E. (2000) A three-dimensional perspective on exon binding by a group II self-splicing intron. *EMBO J.*, **19**, 5007–5018.
18. Swisher,J., Duarte,D.M., Su,L.J. and Pyle,A.M. (2001) Visualizing the solvent-inaccessible core of a group II intron ribozyme. *EMBO J.*, **20**, 2051–2061.
19. Chance,M.R., Bresnick,A.R., Burley,S.K., Jiang,J.S., Lima,C.D., Sali,A., Almo,S.C., Bonanno,J.B., Buglino,J.A., Boulton,S. *et al.* (2002) Structural genomics: a pipeline for providing structures for the biologist. *Protein Sci.*, **11**, 723–738.
20. Pieper,U., Webb,B.M., Barkan,D.T., Schneidman-Duhovny,D., Schlessinger,A., Braberg,H., Yang,Z., Meng,E.C., Pettersen,E.F., Huang,C.C. *et al.* (2011) ModBase, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res.*, **39**, D465–D474.
21. Peebles,C.L., Perlman,P.S., Mecklenburg,K.L., Petrillo,M.L., Tabor,J.H., Jarrell,K.A. and Cheng,H.L. (1986) A self-splicing RNA excises an intron lariat. *Cell*, **44**, 213–223.
22. Rother,K., Rother,M., Boniecki,M., Puton,T. and Bujnicki,J.M. (2011) RNA and protein 3D structure modeling: similarities and differences. *J. Mol. Model.*, **17**, 2325–2336.
23. Laing,C. and Schlick,T. (2010) Computational approaches to 3D modeling of RNA. *J. Phys. Condens. Matter*, **22**, 283101.
24. Novikova,I.V., Hennelly,S.P., Tung,C.S. and Sanbonmatsu,K.Y. (2013) Rise of the RNA machines: exploring the structure of long non-coding RNAs. *J. Mol. Biol.*, **425**, 3731–3746.
25. Ponting,C.P., Oliver,P.L. and Reik,W. (2009) Evolution and functions of long noncoding RNAs. *Cell*, **136**, 629–641.
26. Rother,M., Rother,K., Puton,T. and Bujnicki,J.M. (2011) ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.*, **39**, 4007–4022.
27. Das,R. and Baker,D. (2007) Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl Acad. Sci. USA*, **104**, 14664–14669.
28. Parisien,M. and Major,F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51–55.
29. Jonikas,M.A., Radmer,R.J., Laederach,A., Das,R., Pearlman,S., Herschlag,D. and Altman,R.B. (2009) Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA*, **15**, 189–199.
30. Sharma,S., Ding,F. and Dokholyan,N.V. (2008) iFoldRNA: three-dimensional RNA structure prediction and folding. *Bioinformatics*, **24**, 1951–1952.
31. Keating,K.S. and Pyle,A.M. (2010) Semiautomated model building for RNA crystallography using a directed rotameric approach. *Proc. Natl Acad. Sci. USA*, **107**, 8177–8182.
32. Keating,K.S. and Pyle,A.M. (2012) RCrane: semi-automated RNA model building. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 985–995.
33. Hall,T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.*, **41**, 95–98.
34. Pearlman,D.A., Case,D.A., Caldwell,J.W., Ross,W.S., Cheatham Iii,T.E., DeBolt,S., Ferguson,D., Seibel,G. and Kollman,P. (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.*, **91**, 1–41.
35. Chen,V.B., Arendall,W.B. 3rd, Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.
36. Mortimer,S.A. and Weeks,K.M. (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.*, **129**, 4144–4145.
37. Watts,J.M., Dang,K.K., Gorelick,R.J., Leonard,C.W., Bess,J.W. Jr, Swanstrom,R., Burch,C.L. and Weeks,K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711–716.
38. Vasa,S.M., Guex,N., Wilkinson,K.A., Weeks,K.M. and Giddings,M.C. (2008) ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA*, **14**, 1979–1990.
39. McGinnis,J.L., Duncan,C.D. and Weeks,K.M. (2009) High-throughput SHAPE and hydroxyl radical analysis of RNA structure and ribonucleoprotein assembly. *Methods Enzymol.*, **468**, 67–89.
40. Keating,K.S., Toor,N. and Pyle,A.M. (2008) The GANC tetraloop: a novel motif in the group IIC intron structure. *J. Mol. Biol.*, **383**, 475–481.
41. Costa,M. and Michel,F. (1995) Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.*, **14**, 1276–1285.
42. Daniels,D.L., Michels,W.J. and Pyle,A.M. (1996) Two competing pathways for self-splicing by group II introns: a quantitative analysis of in vitro reaction rates and products. *J. Mol. Biol.*, **256**, 31–49.
43. Keating,K.S., Humphris,E.L. and Pyle,A.M. (2011) A new way to see RNA. *Q. Rev. Biophys.*, **44**, 433–466.
44. Wadley,L.M., Keating,K.S., Duarte,C.M. and Pyle,A.M. (2007) Evaluating and learning from RNA pseudotorsional space: quantitative validation of a reduced representation for RNA structure. *J. Mol. Biol.*, **372**, 942–957.

45. Richardson,J.S., Schneider,B., Murray,L.W., Kapral,G.J., Immormino,R.M., Headd,J.J., Richardson,D.C., Ham,D., Hershkovits,E., Williams,L.D. *et al.* (2008) RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA*, **14**, 465–481.
46. Tullius,T.D. and Greenbaum,J.A. (2005) Mapping nucleic acid structure by hydroxyl radical cleavage. *Curr. Opin. Chem. Biol.*, **9**, 127–134.
47. Ding,F., Lavender,C.A., Weeks,K.M. and Dokholyan,N.V. (2012) Three-dimensional RNA structure refinement by hydroxyl radical probing. *Nat. Methods*, **9**, 603–608.
48. Su,L.J., Waldsich,C. and Pyle,A.M. (2005) An obligate intermediate along the slow folding pathway of a group II intron ribozyme. *Nucleic Acids Res.*, **33**, 6674–6687.
49. Fedorova,O. and Pyle,A.M. (2005) Linking the group II intron catalytic domains: tertiary contacts and structural features of domain 3. *EMBO J.*, **24**, 3906–3916.
50. Abramovitz,D.L., Friedman,R.A. and Pyle,A.M. (1996) Catalytic role of 2'-hydroxyl groups within a group II intron active site. *Science*, **271**, 1410–1413.
51. Xiang,Q., Qin,P.Z.F., Michels,W.J., Freeland,K. and Pyle,A.M. (1998) Sequence specificity of a group II intron ribozyme: multiple mechanisms for promoting unusually high discrimination against mismatched targets. *Biochemistry*, **37**, 3839–3849.
52. Qin,P.Z. and Pyle,A.M. (1999) Antagonistic substrate binding by a group II intron ribozyme. *J. Mol. Biol.*, **291**, 15–27.
53. Abelson,J. (2013) Toggling in the spliceosome. *Nat. Struct. Mol. Biol.*, **20**, 645–647.
54. de Lencastre,A., Hamill,S. and Pyle,A.M. (2005) A single active-site region for a group II intron. *Nat. Struct. Mol. Biol.*, **12**, 626–627.
55. Hamill,S. and Pyle,A.M. (2006) The receptor for within a group II branch-site docking intron active site. *Mol. Cell*, **23**, 831–840.
56. Chanfreau,G. and Jacquier,A. (1994) Catalytic site components common to both splicing steps of a group II intron. *Science*, **266**, 1383–1387.
57. Li,C.F., Costa,M. and Michel,F. (2011) Linking the branchpoint helix to a newly found receptor allows lariat formation by a group II intron. *EMBO J.*, **30**, 3040–3051.
58. 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.
59. Pyle,A.M. and Green,J.B. (1994) Building a kinetic framework for group-ii intron ribozyme activity - quantitation of interdomain binding and reaction-rate. *Biochemistry*, **33**, 2716–2725.
60. Jarrell,K.A., Dietrich,R.C. and Perlman,P.S. (1988) Group II intron domain 5 facilitates a trans-splicing reaction. *Mol. Cell. Biol.*, **8**, 2361–2366.
61. Michels,W.J. and Pyle,A.M. (1995) Conversion of a group-ii intron into a new multiple-turnover ribozyme that selectively cleaves oligonucleotides - elucidation of reaction-mechanism and structure-function-relationships. *Biochemistry*, **34**, 2965–2977.
62. Toor,N., Robart,A.R., Christianson,J. and Zimmerly,S. (2006) Self-splicing of a group IIC intron: 5′ exon recognition and alternative 5' splicing events implicate the stem-loop motif of a transcriptional terminator. *Nucleic Acids Res.*, **34**, 6461–6471.
63. Granlund,M., Michel,F. and Norgren,M. (2001) Mutually exclusive distribution of IS1548 and GBSi1, an active group II intron identified in human isolates of group B streptococci. *J. Bacteriol.*, **183**, 2560–2569.
64. Nagy,V., Pirakitikulr,N., Zhou,K.I., Chillón,I., Luo,J. and Pyle,A.M. (2013) Predicted group II intron lineages E and F comprise catalytically active ribozymes. *RNA*, **19**, 1266–1278.
65. Barrientos-Duran,A., Chillon,I., Martinez-Abarca,F. and Toro,N. (2011) Exon sequence requirements for excision *in vivo* of the bacterial group II intron RmInt1. *BMC Mol. Biol.*, **12**, 24.
66. 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.
67. Laing,C. and Schlick,T. (2009) Analysis of four-way junctions in RNA structures. *J. Mol. Biol.*, **390**, 547–559.