Published in final edited form as:

Nat Struct Mol Biol. 2008 November ; 15(11): 1221–1222. doi:10.1038/nsmb.1509.

Structural basis for exon recognition by a group II intron

Navtej Toor^{1,2}, Kanagalaghatta Rajashankar³, Kevin Keating⁴, and Anna Marie Pyle^{1,2}

¹Department of Molecular Biophysics and Biochemistry, Yale University

²Howard Hughes Medical Institute, Yale University

³The Northeastern Collaborative Access Team (NE-CAT) at the Advanced Photon Source at Argonne National Laboratory, Yale University

⁴Program in Computational Biology and Bioinformatics, Yale University.

Abstract

Free group II introns are infectious retroelements that can bind and insert themselves into RNA and DNA molecules via reverse-splicing. Here we report the 3.4Å crystal structure of a complex between an oligonucleotide target substrate and a group IIC intron, as well as the refined free intron structure. The structure of the complex reveals the conformation of motifs involved in exon recognition by group II introns.

Group II introns are retroelements that have invaded and populated the genomes of many prokaryotes and eukaryotes, resulting in high levels of terrestrial genomic complexity.¹ One hypothesis for the prevalence of regulated splicing in eukaryotes is that pre-mRNA introns (along with the eukaryotic splicing apparatus, or spliceosome) evolved from ancient group II introns that effectively divided early genes into pieces, thereby giving rise to multiple exons and multiple protein products from single genes.¹ The group IIC class of introns is considered to be the most basal, or primitive lineage of group II introns² and, like members of the IIA and IIB classes, the IIC introns use RNA pairing interactions to recognize target sequences for nucleophilic attack and subsequent retrotransposition. Domain I of Group IIC introns contains two exon binding sequences, EBS1 and EBS3, which recognize the 5'-exon and 3'-exon sites of target sequences, respectively.³ Consistent with their relatively low level of sequence specificity, group IIC introns form only 3 to 4 base pairs with the 5'-exon (via EBS1) and they engage the 3'-exon through a single pairing interaction between EBS3 and the first nucleotide of the 3'-exon.⁴ Indeed, the molecular basis for stabilization of the 3'-exon has long been a mystery. The thermodynamics and specificity of substrate recognition by group II introns has been well studied, as the topic is of intense importance for the development of group II introns as therapeutics.⁵ Nonetheless, the structural basis for substrate recognition and catalytic activation has remained unknown until now.

Recently we reported the 3.1Å crystal structure of a self-spliced group IIC intron from *Oceanobacillus iheyensis*,⁶ which revealed a complex molecular architecture that is entirely consistent with decades of biochemical work on group II intron structure and function. In addition, the structure revealed two divalent metal ions in the catalytic core coordinating to the

Correspondence should be addressed to Anna Marie Pyle or Kanagalaghatta Rajashankar Anna Marie Pyle, Room 334A Bass Building, MB&B Yale University, 266 Whitney Avenue, New Haven, CT 06520; 203-432-5633, 203-432-5316; anna.pyle@yale.edu, Kanagalaghatta Rajashankar, Bldg. 436E, APS/ANL, 9700 S. Cass Av. Argonne, IL 60504; 630-252-0694, 630-252-0687, rajashankar@anl.gov.

Accession codes. Protein Data Bank: Coordinates and structure factors for the oligonucleotide co-crystallization and refined intron structure have been deposited with accession codes of 3EOG and 3EOH, respectively.

precise phosphoryl oxygens in domain V that had been biochemically implicated in chemical catalysis. The presence and location of these ions is consistent with the hypothesis that group II introns utilize a two-metal-ion mechanism of catalysis, analogous to that of many protein enzymes and group I introns. The previous model did not contain an exon substrate and exon binding motifs were not well ordered, so a role for these metals in the reaction mechanism could not be determined. However, upon further refinement and modeling of this data ($R_{\text{work}} = 21.3\%$ and $R_{\text{free}} = 24.1\%$; supplementary table 1, PDB accession code 3EOH), along with data obtained from intron co-crystallized with a 6-mer complementary oligonucleotide target RNA ($R_{\text{work}} = 20.7\%$ and $R_{\text{free}} = 25.6\%$; see supplementary table 1, PDB accession code 3EOG), we have produced a highly refined structure of the intron in complex with ligated exons. This has allowed us to visualize regions of the intron which previously lacked electron density and it has also revealed the structural basis for exon recognition by free group IIC introns.

In this structure, electron density has greatly improved in specific regions of domains I and III. For example, it is now possible to visualize the coordination loop, the κ - κ' interaction (supplementary Figure 1), and additional regions of stems Id1 and Id2, as well as additional nucleotides in domain III that were not evident in the previous model. There is also clear density for the 6-nt RNA target oligonucleotide (consisting of a four-nucleotide 5' exon and a 2-nucleotide 3' exon) (Fig. 1, Fig. 2A, supplementary Figure 2), which forms a continuous helix with the complementary EBS1 and EBS3 sequences and places the scissile phosphate over the divalent metal ion binding site that is localized in the bulge of domain V. Crystallization with the RNA oligonucleotide improves electron density for the conserved adenosines that flank EBS3 and positions them toward the inside of the coordination loop, which is a substructure known to be essential for organizing reactants within a group II intron active site. Some electron density corresponding to bound oligonucleotide can be observed in the refined form of the original free intron data, and it is presumably the result of bound RNA and/or DNA fragments that are left over from our native purification protocol. However, this density is weaker than that observed for the co-crystallization data.

The structure of the complex reveals that EBS1 and 3 are interdigitated motifs that form a continuous binding interface for simultaneous recognition of both the 5' and 3' exon sequences. They are joined together by the δ - δ' interaction⁴ (nucleotides 155 and 180) and there is a magnesium ion that bridges the backbone of both motifs (at positions 181 and 223), binding through inner shell coordination and stabilizing the EBS1/EBS3 substructure (Fig. 2A). As a result of this arrangement, the EBS1 nucleotides stack on top of EBS3 as if they were components of one continuous strand. This allows EBS1 and EBS3 to bind across the putative exon junction site and base pair with the 5' and 3' exons simultaneously. Consistent with phylogenetic prediction,³ EBS1 forms four base pairs with the 5'-exon portion of the target sequence and EBS3 pairs with the first nucleotide of the 3'-exon sequence. This arrangement presents the scissile phosphate to the active site within Domain V.

The scissile phosphate at the exon junction is coordinated by two divalent metal ions and, in agreement with the reported diastereomeric preferences of group II introns,⁷ the metals are within coordination distance to both the Rp and Sp phosphate oxygens (Fig 2). Again, consistent with biochemical work,⁸ a divalent metal ion (M_2) directly coordinates the O3' of the scissile linkage and a second metal ion (M_1) appears proximal to O5' of the linkage (Fig. 2B), thereby establishing the type of polarity and ionic positions that would be expected for a two-metal ion mechanism of catalysis. That said, this structure does not attempt to mimic the transition-state: it represents the ground-state form of the intron that immediately follows exon ligation and precedes product release. It does not contain a reaction nucleophile, the substrate is uncleaved in the crystal and group IIC introns are known to require their cognate maturase proteins to catalyze reverse-splicing.⁹ Therefore, the precise role for M_1 and M_2 in catalysis

is not yet known and it remains likely that additional catalytic strategies are required for chemical reactivity by group II introns. However, the juxtaposition of EBS1 and EBS3, together with their arrangement precisely over the domain V bulge explains the molecular basis for intron substrate specificity and indicates that both steps of splicing are tightly coordinated within a highly organized intron active-site.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the staff of the NE-CAT beamline 24-ID-C at the Advanced Photon Source of Argonne National Laboratory. We also thank Dr. Olga Fedorova for advice and support. This work was supported by the Howard Hughes Medical Institute (HHMI) and NIH grant GM50313 (to A.M.P.). N.T. is a postdoctoral associate and A.M.P. is an investigator of the HHMI.

References

1. Martin W, Koonin EV. Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 2006;440:41–45. [PubMed: 16511485]
2. Rest JS, Mindell DP. Retroids in archaea: phylogeny and lateral origins. *Mol Biol Evol* 2003;20:1134–1142. [PubMed: 12777534]
3. Toor N, Hausner G, Zimmerly S. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* 2001;7:1142–1152. [PubMed: 11497432]
4. Costa M, Michel F, Westhof E. A three-dimensional perspective on exon binding by a group II self-splicing intron. *Embo J* 2000;19:5007–5018. [PubMed: 10990464]
5. Pyle, A.; Lambowitz, A. *The RNA World*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2006.
6. Toor N, Keating KS, Taylor SD, Pyle AM. Crystal structure of a self-spliced group II intron. *Science* 2008;320:77–82. [PubMed: 18388288]
7. Podar M, Perlman PS, Padgett RA. The two steps of group II intron self-splicing are mechanistically distinguishable. *Rna* 1998;4:890–900. [PubMed: 9701281]
8. Gordon PM, Fong R, Piccirilli JA. A second divalent metal ion in the group II intron reaction center. *Chem Biol* 2007;14:607–612. [PubMed: 17584608]
9. Robart AR, Seo W, Zimmerly S. Insertion of group II intron retroelements after intrinsic transcriptional terminators. *Proc Natl Acad Sci U S A* 2007;104:6620–6625. [PubMed: 17420455]

Figure 1.

The refined structure of the group II intron co-crystallized with exon substrate reveals additional structural features such as the coordination loop, additional regions of stems Id1 and Id2, κ - κ' (all highlighted by background space-filling models, see also supplementary Figure 1), and the co-crystallized exon junction target (pink ribbon). A total of 391 nucleotides were modeled, including the exons.

Figure 2.

A. Close-up view of the active site with bound exons. EBS1 and EBS3 are held together by the $\delta-\delta'$ interaction and a magnesium ion that forms inner shell contacts to the backbones of both regions (shown as an orange sphere with yellow, dashed lines). Electron density for exon oligonucleotide is clearly visible in the Fo-Fc omit map shown as a mesh contoured at 3.3σ . The second nucleotide of the 3'-exon does not participate in base pairing and, as expected, is less ordered than upstream portions of the oligonucleotide. A nucleotide has been modeled at this position as observed at a contour level of 1.5σ . B. Diagram of the phosphodiester linkage at the exon junction and its coordination with metal ions. Red and blue dashed lines indicate distances (in angstroms) observed for M_1 and M_2 , respectively. M_1 would presumably coordinate an attacking hydroxyl nucleophile (NUC-OH) during a reverse splicing reaction. The distance between M_1 and M_2 is 3.9 Å.