Role of helical constraints of the EBS1–IBS1 duplex of a group II intron on demarcation of the 5′ splice site

MILENA POPOVIĆ[1,2] and NANCY L. GREENBAUM[1,3,4]

[1]Department of Chemistry and Biochemistry, Hunter College of the City University of New York, New York, New York 10065, USA
[2]Department of Chemistry and Biochemistry, The Florida State University, Tallahassee, Florida 32306-4390, USA
[3]The Graduate Center of the City University of New York, New York, New York 10016, USA

ABSTRACT

Recognition of the 5′ splice site by group II introns involves pairing between an exon binding sequence (EBS) 1 within the ID3 stem–loop of domain 1 and a complementary sequence at the 3′ end of exon 1 (IBS1). To identify the molecular basis for splice site definition of a group IIB ai5γ intron, we probed the solution structure of the ID3 stem–loop alone and upon binding of its IBS1 target by solution NMR. The ID3 stem was structured. The base of the ID3 loop was stacked but displayed a highly flexible EBS1 region. The flexibility of EBS1 appears to be a general feature of the ai5γ and the smaller *Oceanobacillus iheyensis* (*O.i.*) intron and may help in effective search of conformational space and prevent errors in splicing as a result of fortuitous base-pairing. Binding of IBS1 results in formation of a structured seven base pair duplex that terminates at the 5′ splice site in spite of the potential for additional A-U and G•U pairs. Comparison of these data with conformational features of EBS1–IBS1 duplexes extracted from published structures suggests that termination of the duplex and definition of the splice site are governed by constraints of the helical geometry within the ID3 loop. This feature and flexibility of the uncomplexed ID3 loop appear to be common for both the ai5γ and *O.i.* introns and may help to fine-tune elements of recognition in group II introns.

Keywords: RNA; group II intron; ai5γ; ID3; EBS1; IBS1; EBS1–IBS1 interaction; NMR

INTRODUCTION

Group II introns are notable for their ability to catalyze their removal from precursor RNA of organellar genes of plants, fungi, and lower eukaryotes, as well as the mobile DNA elements of bacteria (Dai and Zimmerly 2002). This autocatalytic activity enables them to reinsert into canonical (Meunier et al. 1990), and ectopic DNA sites (Yang et al. 1998); as such, group II introns have been successfully used for gene targeting (Guo et al. 2000; Karberg et al. 2001; Perutka et al. 2004; Mastroianni et al. 2008; Zhuang et al. 2009). Group II introns display mechanistic similarities with the nuclear (spliceosomal) splicing mechanism and have been used as model systems for the study of the splicing of precursor messenger RNA (pre-mRNA). Because of the importance of group II introns for the study of genetic expression and evolution, as well as practical applications in biotechnology, it is important to understand their structural features and the impact of those on recognition of their substrates.

Group II introns (400–800 nt) comprise six secondary structure domains (D1–D6) (Fig. 1A), the largest of which, D1, helps to position the active elements of the ribozyme and thereby contributes to reaction specificity (Xiang et al. 1998; Qin and Pyle 1999) and efficiency (Hetzer et al. 1997; Boudvillain et al. 2000). Sequence-specific target recognition in the first step of splicing is achieved through formation of a number of base pairs between the ubiquitous exon binding sequence (EBS) 1 and the complementary segment of the exon, the intron binding sequence (IBS) 1 that ends at the 5′ splice site. EBS1–IBS1 interaction in the ai5γ introns is a required step for splicing via the transesterification pathway (Jacquier and Michel 1987). All group II introns except the small IIC introns also have an EBS2 sequence that pairs with the IBS2 segment of the exon. In IIB introns, the most commonly studied family and the family to which the ai5γ intron belongs, the EBS1 and EBS2 segments are found within a stem–loop (ID3) and a large internal loop of D1, respectively, whereas the IBS1 and IBS2 form a contiguous segment immediately upstream of the 5′ splice site. Little is known about the precise timing of the recognition events involved

[4]**Corresponding author**
E-mail nancy.greenbaum@hunter.cuny.edu
Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.039701.113.

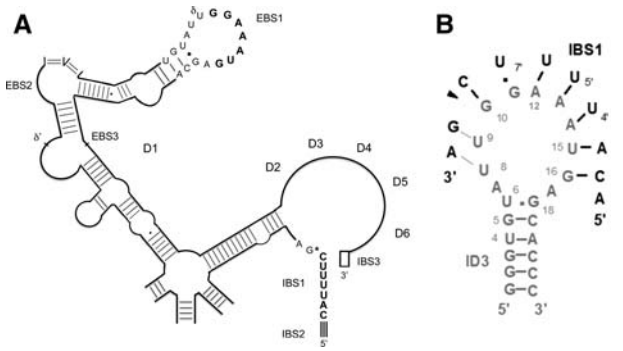

FIGURE 1. (*A*) Schematic depiction of the secondary structure of the ai5γ group II intron from the *Saccharomyces cerevisiae* mitochondrial COX I gene. Domains 1–6 extend from the central "wheel." The segments involved in tertiary interactions contributing to recognition of the 5′ and 3′ splice sites: Exon binding sequences (EBS) 1–3 and intron binding sequences (IBS) 1–3 and δ–δ′ are noted. The EBS1 guide sequence (bold script) is contained within the apical loop of the ID3 stem–loop, the interaction of which with the IBS1 of exon 1 helps define the 5′ splice site. (*B*) The proposed secondary structure of the EBS1–IBS1 interaction involved in recognition of the 5′ splice site of the ai5γ group II intron. Watson-Crick base pairs identified in this study are marked with dashes and G•U wobble pairs with dots in black. The potential base pairs that were investigated here are denoted by gray dashes. Residues of the 5′ exon are shown in black. No pairing is formed between bases downstream from the 5′ splice site (noted by the wedge).

in splicing, which raises the question of which elements of structure are recognized before the first step of splicing.

The sequences of EBS1 and IBS1 vary widely, and the length of the EBS1–IBS1 duplex of IIB introns comprises an average of 6 bp (Toor et al. 2001), although it ranges from four in the *Oceanobacillus iheyensis* (*O.i.*) IIC intron to seven in the ai5γ and bacterial class D introns (Candales et al. 2012). Also, the size of the ID3 loops and the relative position of the EBS1 sequence within the loop of ID3 vary among different families of group II introns (Toor et al. 2001). As an example, the EBS1–IBS1 interaction in the ai5γ intron of the mitochondrial cytochrome *c* oxidase gene from *Saccharomyces cerevisiae*, typical of IIB introns, the subject of study in this report, involves seven of the 11 nucleotides (nt) of the ID3 loop and is positioned at the 3′ end of the ID3 loop (Fig. 1B; Jacquier and Michel 1987).

It was originally shown that cleavage at the 5′ splice site during the forward reaction of ai5γ introns invariably occurs at the 3′ phosphate opposite the 5′-terminal G of EBS1 (Jacquier and Jacquesson-Breuleux 1991). In contrast, subsequent studies of the ai5γ intron suggest that the 5′ splice site is defined by the junction between a single- and double-stranded region of the most thermodynamically stable EBS1–IBS1 pairing (Su et al. 2001). A duplex of seven base pairs was found for all the mutants tested, albeit with a shifted pairing register in some, suggesting that the size of the loop of ID3 impacts on the extent of base-pairing.

In an intact group II intron, the EBS1–IBS1 interaction occurs in the context of conserved tertiary interactions, such as the δ–δ′ base pair that helps to tether the 5′ and 3′ exons during splicing, and contacts with D5 (Toor et al. 2010). The δ–δ′ base pair is formed between a nucleobase of the ID3 loop (δ) located immediately 5′ to EBS1 and a base within an asymmetric loop of D1 (δ′) (Fig. 1A). There appears to be no agreement, however, regarding the precise order of formation of these tertiary interactions. The absence of a consensus model for the order of formation of interactions among members of the IIB family of group II introns is intriguing. The Pl.LSU/2 intron from the algae *Pylaiella littoralis* appears to require formation of the δ–δ′ tertiary interaction and the presence of D5 for high affinity binding of the exon substrate (Costa and Michel 1999; Costa et al. 2000). Conversely, D1 of the ai5γ intron has been shown to be an independently folding unit that recognizes and binds the exon substrate with affinity comparable to that of the full intron, without the requirement for recognition by D5 (Qin and Pyle 1997, 1999; Xiang et al. 1998). These findings suggest that individual secondary structural elements of ai5γ D1 represent valid systems for the study of 5′ splice site recognition.

We have determined a solution structure of the native ai5γ ID3 stem–loop and investigated the changes in its structure upon interaction with its target IBS1 sequence. We found that the EBS1 sequence is highly flexible and shows evidence of conformational exchange in the absence of IBS1. We suggest that the flexibility of EBS1 may help to prevent fortuitous pairing, presenting an additional layer of control in the recognition of the 5′ splice site that is general for both the ai5γ and *O.i.* group II introns. Upon pairing with IBS1, the loop of ID3 becomes structured and forms a helix that terminates at the splice site in spite of the potential to form two additional pairs. In the complete and folded intron, the ID3 residue immediately 5′ of the EBS1 sequence is involved in the tertiary δ–δ′ interaction; however, this minimal system does not include the δ′ nucleotide; thus the fact that no additional base pairs between EBS1 and IBS1 form beyond the splice site in spite of the presence of additional base-pairing partners suggests that definition of the 5′ splice site is at least partly determined by the limits on the length of the EBS1–IBS1 duplex imposed by helical constraints of the ID3 loop. We propose this to be a general feature of 5′ splice site definition in group II introns.

RESULTS

Folding and secondary structure of the ID3 stem–loop in solution

We verified that ID3 exists in a monomeric stem–loop conformation by comparison of the electrophoretic mobility of the oligomer sequence with that of other sequences already shown to form stem–loop on nondenaturing gels (data not shown). A single major band of the ID3 sample was taken as evidence for a single conformation of the stem–loop of ID3.

Circular dichroism (CD) measurements were recorded using samples with either 50 or 100 mM KCl or NaCl, as well as low ionic strength solution containing 10 mM KCl. CD spectra of the ID3 stem–loop at medium and high ionic strength (50 and 100 mM KCl or NaCl, respectively) displayed a large negative ellipticity at 210 nm and a positive peak at 265 nm, typical of an A-form RNA helix (Bloomfield et al. 2000), with a positive shoulder at 280 nm, suggesting the presence of a single-stranded or unstructured region (Fig. 2; Supplemental Fig. S2a). Addition of Mg^{2+} (20:1 ratio Mg^{2+}:RNA) resulted in an slight increase in positive ellipticity, suggesting a small increase in helicity associated with the presence of Mg^{2+} ions. Spectra recorded in the presence of either KCl or NaCl are virtually identical, suggesting that these monovalent metal ions have a similar effect on the secondary structure of ID3. Although only marginal, the slightly increased ellipticity at 265 nm is consistent with improved stacking in the presence of Na^+ compared with K^+.

Lower ellipticity at 265 nm in the presence of 10 mM KCl (Supplemental Fig. S1b) suggests poorly stacked bases at low ionic strength. We detected maximum ellipticity of ID3 upon addition of 20 eq (400 μM) $MgCl_2$, comparable to the values observed in the presence of either 100 mM NaCl or KCl, suggesting improved stacking in the presence of Mg^{2+}. Similar to data on a modified ID3 loop published by Kruschel and Sigel (2008), our CD spectra of the mixture of ID3 and IBS1 in the presence of 10 mM KCl suggest absence of full binding. Interestingly, however, the melting effect upon titration of Mg^{2+} to ID3 under low ionic conditions (10 mM KCl) observed by Kruschel and Sigel (2008) was not detected in our studies. To the contrary, we observed increased stacking in the presence of Mg^{2+}. We attribute this to the fact that the native ID3 loop, unlike the modified loop used by these investigators, is single-stranded and flexible prior to the interaction with the target strand.

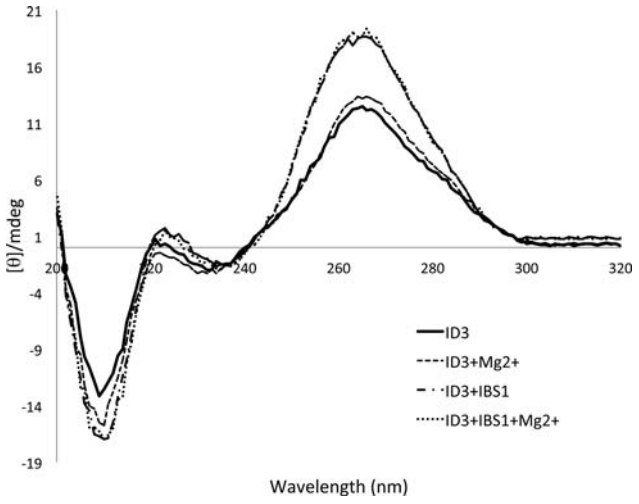

FIGURE 2. Circular dichroism (CD) spectra of the ID3 stem–loop at 100 mM NaCl. Ellipticity was shown in mdeg for all spectra. Spectral features of the ID3 stem–loop, e.g., a minimum at 210 nm and maximum 265 nm, were typical of an A-form RNA helix; a shoulder at 280 nm suggested a single-stranded region. Addition of $MgCl_2$ resulted in a slight increase in the intensity of the band at 265 nm, consistent with an enhancement of stacking. Addition of the IBS1 target induced an increase in the positive ellipticity at 265 nm, consistent with additional helix formation. Addition of Mg^{2+} to the ID3–IBS1 complex did not result in further spectral changes.

Spectral and structural features of the ID3 stem–loop

In order to probe structural features of the ID3 stem–loop, we collected solution NMR spectra. Resonance assignments were made using standard homo- and heteronuclear NMR spectroscopic methods (Dingley and Grzesiek 1998). Base-pairing in the stem was identified by assignment of four sharp resonances of imino protons (G2, G3, U4, and G5) attributed to Watson-Crick base pairs and upfield shifted imino protons attributed to a G•U wobble pair between U6 and G18. The assignments of base pairs in the stem of a fully labeled ID3 were confirmed by detecting $^{2h}J_{NN}$ through-bond coherence transfer in an HNN-COSY spectrum (data not shown). Six direct correlations were observed at the chemical shifts of imino ^{1}Hs and ^{15}Ns of G2, G3, U4, G5, U6, and G18 residues, as well as cross-H-bond correlations of imino ^{1}Hs to the imino ^{15}N of the residues C22, C21, A20, and C21, to which the imino protons were hydrogen-bonded, confirming the initial assignments of stem imino ^{1}Hs. The absence of imino ^{1}H resonances for the residues of the loop is taken as evidence for a lack of stable pairing interactions among loop residues.

Well-resolved NOEs between aromatic H6/8 and anomeric H1′ resonances of the stem–loop residues (Fig. 3) implied A-type helical conformation for the stem and stacking interactions through U9 (i.e., the first three residues of the loop), as well as for A17 (the last residue of the loop). However, the downfield chemical shifts for A7 and A17 H2 resonances (7.76 and 8.02 ppm, respectively) correlate with increased flexibility, as detected previously (Newby and Greenbaum 2001, 2002). Specifically, the relatively downfield chemical shift of 8.02 ppm for the A17 H2 resonance suggests that the A17 residue is highly deshielded and/or spends a significant fraction of time unstacked. Sequential base-to-ribose NOEs associated with A-type helical parameters were weak or absent in the rest of the loop, consistent with absence of rigid structure involving those nucleotides, as was initially suggested by the spectra of exchangeable protons. The apparent lack of rigid structure is also evident from the downfield chemical shifts of A12–14 residues of the loop that range from 7.93–8.01 ppm, similar or even more downfield than those in the fully exposed A residues, such as the A H2 resonance of the extrahelical branch site A at 7.85 ppm (Newby and Greenbaum 2001) or that of the flipped-out apical A residue of the GAAA loop at 7.82 ppm (Hoogstraten et al. 1998). Broad H2 resonances are consistent with increased conformational flexibility of the A12–14 residues compared with stable A-form helices. C8 resonances of G10 and G11 exhibited intense intraresidue H8-H1′ NOEs, as well as an absence

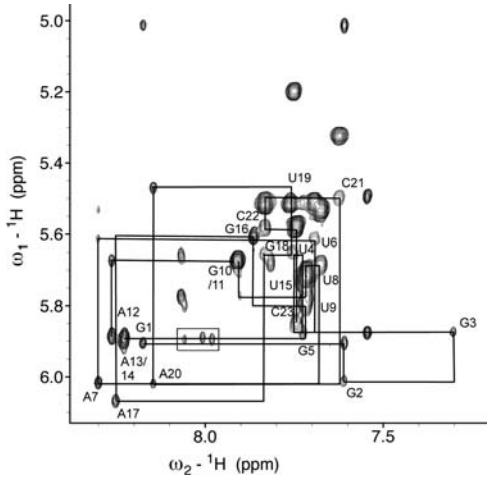

FIGURE 3. Base-H1′ region of a 2D NOESY spectrum of nonexchangeable ^{1}Hs of the ID3 stem–loop. Sequential aromatic–anomeric (H8/6-H1′) connectivities are marked by lines. Sequential NOE connectivities were interrupted at residues A13/A14 and at 14 H1′-U15 H6. NOEs involving H2 resonances of A12, A13, and A14, which were broadened in the uncomplexed ID3, are boxed (from *left* to *right*): A14 H2-A14 H1′, A12 H2-A13 H1′, and A12 H2-A12 H1′. The spectrum was acquired with 72 scans and 576 increments and a mixing time of 350 msec.

of intraresidue H8-H3′ NOEs, in a NOESY spectrum collected with a mixing time of 50 msec, consistent with a *syn* glycosidic angle conformation for both. Moreover, C8 resonances of G10 and G11 displayed marked downfield shifts relative to the mean value for C8 resonances of the stem residues, confirming the *syn* conformation (Greene et al. 1995; Nikolova et al. 2011).

TOCSY spectra revealed strong H1′-H2′ and H1′-H3′ correlations for U8, U9, and U15 residues, indicative of C2′-*endo* conformation (Wijmenga et al. 1994). The A7, G10, G11, A12, A13, A14, G16, and A17 residues of the ID3 loop exhibited H1′-H2′ cross-peaks of intermediate intensity, consistent with ribose groups in alternative conformations or undergoing exchange between C3′- and C2′-*endo* conformations (data not shown). The ^{13}C chemical shift of C1′ residues of U8, U9, and U15 (89.95–90.40 ppm) were also consistent with a C2′-*endo* conformation of these riboses. Conversely, residues of the stem displayed C1′ chemical shifts ranging from 92.18–94.23 ppm, associated with a C3′-*endo* conformation of those riboses (Varani and Tinoco 1991).

Structure calculations of the ID3 stem–loop

An initial set of structures was calculated from a total of 428 structural restraints (Table 1). Two hundred eighty-seven NOE-derived distance restraints were used for structure calculation as well as 32 individually imposed hydrogen bond restraints for base pairs identified by HNN-COSY and NOESY of exchangeable ^{1}Hs. One hundred forty-one dihedral angle restraints were applied. Among them, the backbone angles for the paired stem residues were constrained to A-form helical values. The initial structures were subjected to a torsion angle molecular dynamics (TAMD) protocol (Rice and Brunger 1994; Stein et al. 1997) applied in XPLOR-NIH (Schwieters et al. 2003), described in Materials and Methods. Ten of the 20 lowest energy structures were refined further using the DELPHIC database potentials, applied upon the stem residues only (Clore and Kuszewski 2003). The 10 lowest energy structures had no NOE violations >0.5 Å and no dihedral violations >5°.

Structural features of the ID3 stem–loop

The stem adopts a defined A-form structure comprising five Watson-Crick base pairs and a closing G18•U6 wobble pair (Fig. 4). The base of A7 stacks onto the base of the U6 residue. The A17 base intercalates between the U6 and A7 bases. U8 and U9 bases are also stacked facing the interior of the loop. The Watson-Crick face of A17 is within hydrogen-bonding distance to both U8 and U9 and is nearly coplanar with the base U8 residue. In contrast, the bases of EBS1 are solvent exposed and display no stacking. The atomic position RMSD among the loop residues 10–16 (EBS1) is high, consistent with an unstructured and/or dynamic loop. Residues comprising the EBS1 (10–16) are poorly defined in the family of structures, as has previously been seen in large RNA loops (Bouvet et al. 2001). Although we cannot definitively exclude the possibility that the loop nucleotides are underconstrained,

TABLE 1. Summary of input data and NMR structure statistics

No. of distance restraints	287
Intranucleotide	130
Internucleotide	99
Long-range	26
Hydrogen bonds	32
Planarity restraints	12
Dihedral angle restraints	141
Backbone	88
Ribose pucker	36
Glycosydic	17
NOEs per residue	12.5
NOEs and dihedrals per residue	18.4
Mean RMSD from ideal experimental restraints	
Distance restraints (Å)	0.045 ± 0.008
Dihedral restraints (°)	0.419 ± 0.096
Mean RMSD from ideal covalent geometry	
Bond length (Å)	0.00595 ± 0.00032
Angle (°)	0.894 ± 0.014
Impropers (°)	0.525 ± 0.012
No. of NOE violations (>0.5 Å)	0
No. of dihedral violations (>5°)	0
Mean nonhydrogen atom pairwise RMSD (1–9, 17–23)	1.26 ± 0.33
Mean all atom pairwise RMSD (1–9, 17–23)	1.29 ± 0.28
Mean nonhydrogen atom pairwise RMSD (10–16)	3.41 ± 1.86
Mean all atom pairwise RMSD (10–16)	3.55 ± 1.89

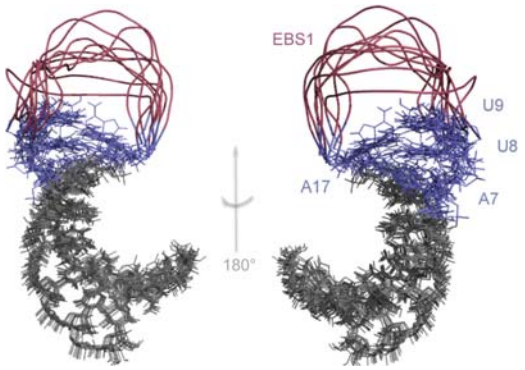

FIGURE 4. Structure of the ID3 stem–loop showing the 5′ end of the ID3 loop on the *left* side; the figure on the *right* is rotated by 180°. The stem of ID3 adopts an A-type helix. Stacking involving bases A7, U8, U9, and A17 continues into the loop, but the riboses of these residues display C2′-*endo* and mixed conformations. The residues of the EBS1 recognition sequence are not well defined in the family of structures, consistent with a high degree of disorder.

we argue that data from chemical shifts are the result of a high degree of flexibility and conformational exchange.

Evidence for flexibility within the ID3 loop

Chemical shifts (δ) of resonances in highly disordered regions have been shown to be similar to those of nucleotide monophosphates (NMPs) (Eichhorn et al. 2012), whereas those in the structured regions have upfield-shifted positions (lower ppm values) (Fares et al. 2007). Carbon chemical shifts for aromatic nuclei in the stem of ID3 (Fig. 5A) resonated upfield relative to those of NMPs, suggesting a well-ordered region with stacking interactions (Fares et al. 2007). Comparable ^{13}C resonances in the loop, however, displayed similar shifts to those of NMPs, suggesting that loop residues are disordered (Fig. 5B).

It has been shown that highly flexible residues display intense resonances compared with those of the well-ordered residues in a ^{13}C-HSQC spectrum (Zhang et al. 2006; Shajani et al. 2007; Eichhorn et al. 2012). We compared cross-peak intensities to that of the G3 stem residue, which was normalized to a value of one. Resonance intensities of the stem were low (Fig. 5C). In contrast, the resonances in the loop of ID3: U9, G10, G11, and U15 displayed a marked increase of intensity (92 ± 42% increase), suggesting increased flexibility on a pico- to nanosecond time scale, consistent with increased disorder. Interestingly, C8-H8 cross-peaks of U9 and U15 display the highest intensity, suggesting increased fluctuations, whereas the residues 10–14 show a higher degree of stacking, commensurate with behavior previously observed in single-stranded regions of RNA (Eichhorn et al. 2012). C6/8-H6/8 cross-peaks of U8, G16, A17, and G18 were broadened beyond detection and were therefore not appropriate for such analysis, consistent with a degree of conformational exchange, as previously observed (Johnson and Hoogstraten 2008).

Interaction of the ID3 stem–loop with IBS1

We assayed binding between the ID3 stem–loop and the exon strand by nondenaturing gel electrophoresis, CD, and NMR spectroscopy. We first measured changes in migration of individual oligomers representing the ID3 stem–loop and the IBS1 strand upon interaction on nondenaturing gels (data not shown). A 1:1 stoichiometric combination of IBS1 and the ID3 stem–loop at 4°C resulted in appearance of one band that migrated more slowly than either of the two individual components, suggesting a single conformation of the complex. A faint band (<3% the intensity of the major band) migrating more slowly in each of the ID3 and ID3 +IBS1 lanes suggested a very small amount of dimer formation (data not shown). A parallel investigation of complex formation between the ID3 stem–loop and a longer target strand (14 nt) displayed similar results (data not shown).

CD spectra of the ID3–IBS1 complex in the presence of sodium exhibited an increase in the positive ellipticity at 265 nm, as well as a decrease of the shoulder at 280 nm, compared with the spectra of the ID3 alone in the same buffer. This change is consistent with an increase in helical character and a decrease of a single-stranded region (Fig. 2). As was the case for the ID3 loop alone, substitution of KCl for NaCl resulted in slightly lesser ellipticity of the ID3–IBS1 complex at 265 nm, suggesting that K^+ is less effective at fostering stacking than Na^+. Addition of Mg^{2+} (20:1 Mg^{2+}:RNA) to the ID3–IBS1 complex in the presence of Na^+ did not result in any significant spectral changes, suggesting that Mg^{2+} does not cause major changes in helical content of the ID3–IBS1 complex in NaCl solution. However, addition of Mg^{2+} to RNA in KCl resulted in a spectrum identical to that in NaCl alone.

NMR assignment and structural features of the ID3–IBS1 complex

NOESY spectra of exchangeable protons provided evidence for formation of five additional base pairs as compared with ID3 alone. When temperature was increased to 25°C, all imino proton resonances associated with the intermolecular base pairs broadened beyond detection. Similarly, at temperatures above 30°C, NOESY spectra of nonexchangeable protons of the complex were identical to the spectra of the ID3 stem–loop alone, consistent with an absence of complex formation. To prevent thermal melting of the complex, all spectra were acquired at ≤15°C.

We observed slight downfield shifts for the imino proton resonances of the closing G18•U6 wobble pair of the ID3 stem in the complex compared with those in ID3 alone; no other chemical shift changes within the stem were detected. Accuracy of assignment of imino and amino protons was facilitated by use of 2D NOESY spectra obtained using samples of a uniformly ^{15}N/^{13}C-labeled ID3 stem–loop paired with an unlabeled IBS1 target and an unlabeled ID3 stem–loop paired

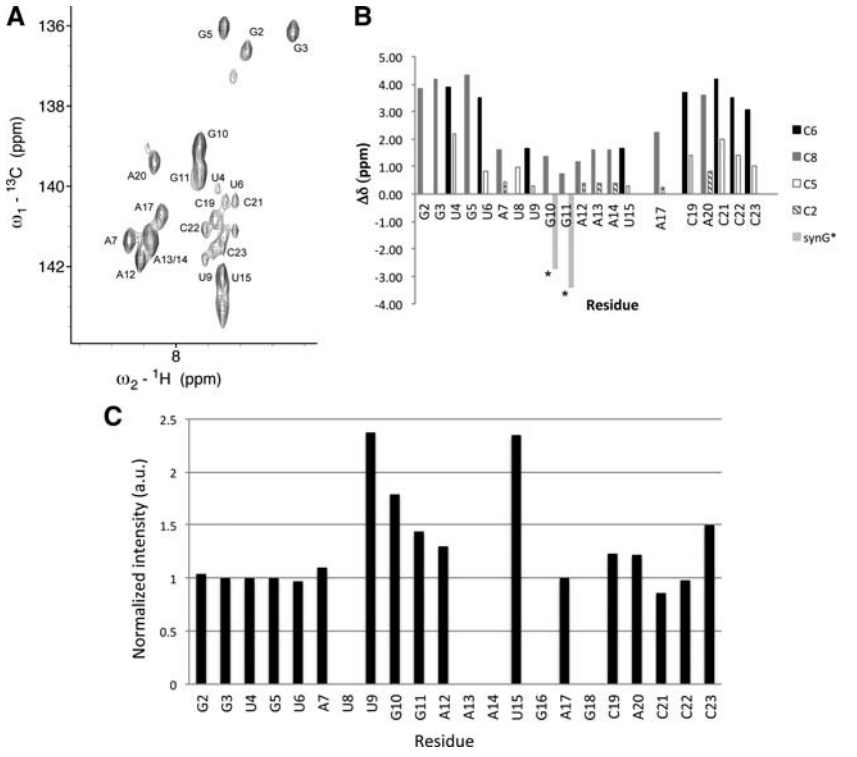

FIGURE 5. (*A*) Section of a ^{13}C-HSQC spectrum of the ID3 stem–loop showing C6/8-H6/8 correlations. In this spectrum, pyrimidine cross-peaks appear as doublets as a result of the unrefocused $^{1}J_{C5\text{-}C6}$ scalar coupling, with an unmarked downfield partner in the ^{13}C dimension. The downfield-shifted chemical shifts of C8 resonances for the residues of the loop of ID3 are consistent with increased flexibility. Concentration of RNA in the sample was ~0.3 mM; the spectrum acquired with 64 scans and 256 increments. (*B*) Plot of chemical shift values for aromatic carbon resonances were compared to analogous positions of nucleotide monophosphates (NMPs). The chemical shift difference is defined as δ (NMP)-δ (ID3) to obtain positive values. Resonances in the loop of ID3 display similar chemical shift values to those of NMPs, suggesting a disordered loop. Bars marked by *syn* G* identify residues that displayed a large downfield chemical shift relative to the value for C8 resonances of the residues in the stem. The large negative value for δ (G_{stem}) − δ (G_{loop}) is consistent with a *syn* glycosidic angle of those residues. (*C*) Plot of normalized intensity of C6/8-H6/8 cross-peaks from a nonconstant time ^{13}C-HSQC spectrum shows relative flexibility of residues. The intensity of aromatic cross-peaks was compared with the intensity of the C8 resonance of the G3 residue in the stem (value set at 1.0). Increased intensity is associated with higher pico- to nanosecond fluctuations than in the residues of the ID3 stem.

with a ^{15}N/^{13}C-cytosine-labeled IBS1 target in which ^{1}H-^{15}N scalar couplings were not refocused. From analysis of splittings resulting from $^{1}J_{HN}$ scalar couplings, we attributed imino ^{1}H resonances to the ID3 stem–loop or the IBS1 target unambiguously (Supplemental Fig. S2).

Through this method, we identified five base pairs resulting from complex formation, specifically G10-C8′, G11•U7′, A12-U6′, A13-U5′, and A14-U4′ (where n′ denotes residues of the IBS1 target strand). In addition, inter- and cross-strand $AH2_{(n)}$-$H1'_{(n+1)}$ and AH2-AH2 NOEs confirmed stacking of residues and implied formation of the U15-A3′ base pair (Supplemental Fig. S3a). The identity of A-U base pairs was confirmed by A C2-H2 correlations in the ^{13}C HSQC spectra of a sample in which a labeled ID3 stem–loop was paired with an unlabeled IBS1 strand (Supplemental Fig. S3b).

Identification of G16-C2′, for which no imino proton was observed, was achieved indirectly by taking advantage of different line-width patterns for paired versus unpaired resonances. The resonance linewidth of an A3′ H2 proton (IBS1 target) of 12 Hz was comparable to the linewidths of the other two A H2s of the EBS1 (also 12 Hz), which are base-paired and flanked by additional base pairs, suggesting that A3′ is not a terminal base pair and, indirectly, that the G16-C2′ base pair forms (Fig. 1B). Overall, we collected direct and indirect evidence for formation of 7 bp between EBS1 and IBS1, six Watson-Crick and one G•U wobble, so that the ID3–IBS1 pairing begins at G10-C8′ and ends at the G16-C2′ base pair. The IBS1 segment displays evidence of stacking interactions throughout, whereas we do not observe similar signature for the EBS1 sequence, suggesting somewhat altered helical parameters from A-type helix in the loop of ID3. We evaluated the possibility of formation of additional base pairs 3′ to the splice site involving interaction between U8 and U9 of EBS1 with A10′ and G9′ of IBS1, respectively. However, none of the spectral features associated with base pair formation were detected. Specifically, there was no NOE detectable between imino protons of U8 or U9 in 2D spectra or of the individual resonances in 1D spectra; in particular, a NOE between U9 NH3 and G9′ NH1 would have been anticipated in an uncrowded region of the spectrum. Moreover, the A10′ H2 resonance exhibited a downfield shift and displayed no cross-strand NOEs, confirming the unpaired status of A10′.

The absence of TOCSY correlations suggested C3′-*endo* conformation for all riboses of the complex with the exception of the terminal base pair of the ID3 stem. ^{31}P chemical shifts well within the A-form envelope were consistent with no apparent backbone perturbation in the ID3–IBS1 complex.

Based upon these observations and constraints, these features allow us to propose a structural model of the complex (Fig. 6). The ID3 stem is unchanged by complex formation. Nucleotides at the base of the ID3 loop (A7, U8, U9, and A17) are unpaired. Seven base pairs form between the ID3 and IBS1 in the loop, although the 2 bp at the 3′ side of the ID3 stem–loop may undergo a more rapid exchange rate. There is clear evidence that no additional pairs form beyond the EBS1–IBS1 pairing. In contrast with the

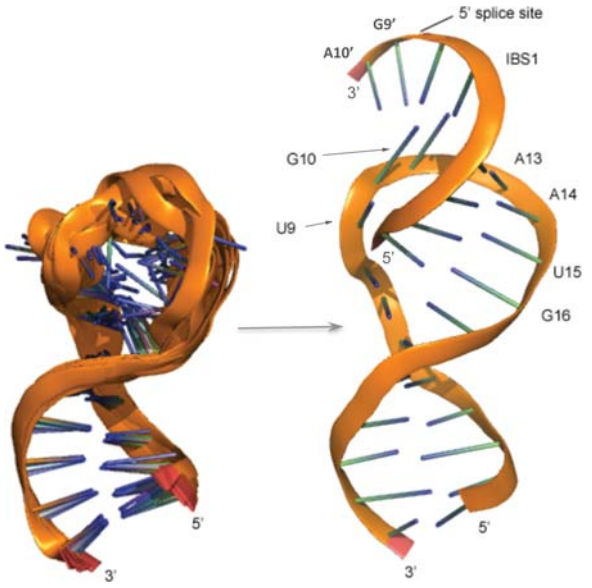

FIGURE 6. A comparison of the lowest energy ensemble of structures of the ID3 stem–loop and a cartoon of the ID3–IBS1 complex. Starting with the structure of the ID3–IBS1 complex of Kruschel and Sigel (PDB: 2K64), we modified conformation of the backbone of ID3 to accommodate our NMR data using the modeling features of Pymol (The PyMOL Molecular Graphics System, version 1.2r3pre, Schrödinger) and added two residues 3′ to IBS1 that were used in our studies. Geometry was based on A-form backbone parameters and stacking information from base-base NOEs within the complex. Helical constraints of the EBS1–IBS1 pairing, limited by the length of the ID3 loop, result in a large distance between U8 and A10′ and U9 and G9′, preventing formation of additional base pairs.

uncomplexed ID3 loop, in which several of the bases are splayed out in most of the structures, pairing with IBS1 results in a well-structured duplex.

DISCUSSION

In order to identify the molecular basis for splice site definition of a group IIB ai5γ intron, we probed the solution structure of the ID3 stem–loop and changes that occur upon binding of its IBS1 target by solution NMR spectroscopy. The lowest energy ensemble of structures of the ID3 stem–loop of the mitochondrial ai5γ intron of *S. cerevisiae,* described here, shows a structured stem and a flexible base of the loop, as well as a highly flexible EBS1 region of the ID3 loop (residues G10–G16). The Watson-Crick faces of all residues in the EBS1 region are extruded from the interior of the ID3 loop and exposed to the solvent.

Biophysical studies have previously suggested that EBS1 is exposed and flexible to interact with its target. Qin and Pyle (1999) have shown that the EBS1 sequence interacts with its exon with high affinity even in the absence of D5. Folding of D1 is the rate-limiting step for folding of the full ai5γ group II intron (Waldsich and Pyle 2007); the free energy of binding between D1 (as well as the D13 segment) and the IBS1 sequence is similar to that of base-pairing calculated for a RNA duplex of the same length and sequence, suggesting that EBS1 and IBS1 are unstructured and exposed prior to interaction with the target in the ai5γ group II intron.

Flexibility of the EBS1 sequence in these minimal systems from a IIB intron is in full agreement with the disorder observed for the EBS1 sequence in crystal structures of a small bacterial IIC intron from *O.i.* The use of crystallographic thermal (B)-factors to represent the effects of thermal motion and positional disorder, although not without limitations (DePristo et al. 2004), is a widely utilized method for assessment of flexibility in biomolecules. The reported B-factors of a number of the crystallographic models of *O.i.* introns (Marcia and Pyle 2012; Marcia et al. 2013) indicate that the EBS1 segment is markedly more flexible than the surrounding residues in the substrate-free ID3 loop and becomes more ordered when in complex with the target strand. Although the 7-nt EBS1 in the ai5γ intron may have very different structural features from the 4-nt EBS1 strand of the *O.i.* intron, these findings show that flexibility of the EBS1 sequence is a general feature for these members of group IIB and IIC introns.

It is important to note that the flexibility of EBS1 in the intact *O.i.* intron is fully recapitulated in our minimal system, suggesting that the use of a minimal system was suitable for this study. Moreover, we propose that the flexibility of EBS1 is a general characteristic of group II introns that may impact on splice site recognition.

The major biological implication for a flexible ID3 loop is that it may assist in positioning of the bases of the EBS1 sequence by making them easily accessible for base-pairing with the exon, thus facilitating the 5′ splice site formation. The flexibility of the EBS1 sequence could aid in the rapid search of conformational space. It is tempting to speculate on the benefits of a flexible target binding sequence, including that weak EBS1–IBS1 binding and flexibility are likely to prevent fortuitous pairing and splicing errors and therefore act as a control mechanism for the proper 5′ splice site recognition.

The identity of monovalent metal ions has an impact on the catalytic pathway of in vitro splicing activity of group II introns. Although sodium ions have been used for in vitro splicing assays (Wank et al. 1999), a high concentration of sodium has a slight inhibitory effect on the forward splicing reaction (Costa et al. 1997), which brings to question the utility of sodium ions for structural studies of group II introns.

The CD studies we performed in the presence of either KCl or NaCl are indicative of a similar secondary structure of the ID3 stem–loop and slightly better stacking in the presence of Na^+. Previous studies have shown that stability of a loop–loop interaction involving formation of six Watson-Crick pairs and a substantial deformation of the RNA backbone is inversely proportional to the cation size. Specifically, it was shown that Na^+ cations were preferred over K^+ in stabilization of the Tar–Tar* complex (Chen et al. 2009; Lambert et al. 2009), whereas the stabilization of the Tar* hairpin, as is the case for ID3, shows only minor dependence on the cation

type (Lambert et al. 2009). Our findings are in full agreement with these previous observations.

Kruschel and Sigel also investigated the structural features of a modified ai5γ ID3 stem–loop and the ID3–IBS1 complex by solution NMR (PDB: 2K63 and 2K64). In particular, they replaced two adenosine residues with cytosines and made compensatory changes in the IBS1 residues in order to improve thermal stability of the ID3–IBS1 pairing. In spite of the sequence difference and the use of different monovalent K^+ ions (for review, see Sigel 2005), both groups identified similar structural properties in the ID3 stem–loop.

A number of crystal structures of the *O.i.* intron in which potassium ions were replaced by various monovalent cations, including Na^+, display a remarkably well folded intron scaffold (Marcia and Pyle 2012), suggesting that a wide range of monovalent metal ions is effective for screening of charge that is required for complex tertiary interactions. Importantly, no specifically bound K^+ ions in the crystal structure of the *O.i.* intron, which may account for the observed K^+ requirement for the forward-splicing reaction, are in the vicinity of the EBS1–IBS1 pairing. Taken together, these findings suggest that the use of Na^+ as the monovalent counter ion is advantageous for the system studied here.

Ribozymes that cleave RNA typically recognize and position their substrates by base-pairing between guide sequences and their targets (Zaug et al. 1986; Jacquier and Michel 1987; Uhlenbeck 1987). Recognition of the 5′ splice site in group II introns is characterized by formation of base pairs between the EBS1 sequence (within the loop of ID3) and IBS1 (Jacquier and Michel 1987; Qin and Pyle 1998; Boudvillain et al. 2000; Costa et al. 2000; Su et al. 2001). Our study indicates that the EBS1 region of ID3 undergoes a significant structural change, resulting in formation of 7 bp with the IBS1 target, whereby both the IBS1 and EBS1 strands become more structured upon interaction via an induced-fit or conformational capture mechanisms. These findings are consistent with those for the modified ID3–IBS1 complex deposited by Kruschel and Sigel (PDB 2K64); however, the IBS1 in their studies terminates at the 5′ splice site and therefore yields no information about the conformation of the splice site and the potential for base-pairing beyond the splice site.

We evaluated the impact of the looped conformation of EBS1 on the length of the EBS1–IBS1 duplex by examination of coordinates of deposited structures. We found that the EBS1–IBS1 pairing terminates at the 5′ splice site despite the presence of additional complementary bases (A10′-U8 and G9′•U9 in Fig. 1B) in the ID3 alone. The distance between C1′ residues in the ideal A-form RNA does not differ between strands, although small deviations have been observed in solution structures of RNA stems. Interestingly, the variation in distance between terminal C1′ residues for the strands of the EBS1–IBS1 pairing of the modified ID3 loop of an ai5γ intron are very small for the first 5 and 6 bp (PDB: 2K64; 2.4% and 3.3%, respectively). In contrast, the distance between the terminal C1′ residues seven bases apart in the EBS1 sequence is 18.8 Å in the structure of this modified ID3 of the ai5γ intron, while the same distance in IBS1 is 20.9 Å, an average increase of 11.2%, suggesting an increasing deviation from A-form parameters toward the end of the pairing and a significant deviation for the last base pair. Although less pronounced, a similar trend is observed for the short EBS1–IBS1 pairing of the *O.i.* intron in the crystal structure of the precatalytic state (PDB: 4FAQ), where the representative distances between termini (C1′–C1′) are 18.8 and 20.0 Å, a difference of 6%. The perturbation of the helical parameters at the 5′ splice site suggests an increase in strain upon the last base pair that may create a limit on the length of the EBS1–IBS1 pairing. We thus propose that the location of the splice site is at least partly determined by the helical constraints of the ID3 loop in group II introns.

Interestingly, an increasing deviation is observed for EBS1–IBS1 duplexes of both the ai5γ and the *O.i.* introns, in solution NMR and crystal structures, respectively. We consider it significant that an increase in deviation of the EBS1 orientation toward the end of the EBS1–IBS1 pairing occurs regardless of whether the ID3 is alone in solution or in the context of an intact group II intron, suggesting that the minimal system behaves similarly to the full-length intron. Moreover, the effect is similar for the IIB and smaller IIC introns, suggesting a general mechanism for regulation of the 5′ splice site recognition.

Our two-component sample comprising the ID3 loop and a 10-nt oligomer representing the IBS1 target sequence does not include tertiary interactions found in the complete group II intron prior to splicing (Costa and Michel 1999; Costa et al. 2000). In particular, the residue of the ID3 loop 5′ to the terminal base in the EBS1–IBS1 pairing is involved in a tertiary interaction in all group II introns, forming either a long range δ–δ′ interaction by pairing with the δ′ residue located in a remote part of D1 (IIB and IIC introns) or the EBS3–IBS3 pairing with the first residue of exon 2 (IIA introns) (Lambowitz and Zimmerly 2011). However, we have found that the EBS1–IBS1 pairing terminates at the proposed 5′ splice site, even though there is the potential for formation of two additional pairs. Our findings suggest that the tertiary interactions involving the δ residue are not the cause for the absence of additional base pairs. In fact, the possibility for formation of additional base pairs has been observed for a number of introns for which sequences and secondary structures have been determined (Dai et al. 2003). We propose that both the geometry of the EBS1–IBS1 pairing in the context of the ID3 loop and the formation of the δ–δ′ interaction are involved in the 5′ splice site recognition. This raises the question of whether the δ–δ′ interaction is facilitated by the limit in the length of the EBS1–IBS1 pairing. It appears that the termination of the EBS1–IBS1 pairing by the constraints of the ID3 loop is consistent with accessibility of the δ residue for the δ–δ′ interaction, rather than competing with it for the extension of the pairing.

Erroneously spliced longer exon products are retrimmed to the correct length by repairing with the EBS1 sequence (Su et al. 2001). This mechanism has been shown to be a part of proof reading done by the ai5γ group II introns (Wang and Silverman 2006). The geometric component of the 5′ splice site control may therefore have a direct role in a general proofreading mechanism of group II introns.

MATERIALS AND METHODS

Design of RNA samples

The RNA oligomers used for spectroscopic studies were designed to represent the ID3 stem–loop and its oligonucleotide target IBS1 of the ai5γ group II intron from *S. cerevisiae*. The RNA oligomer that represents the ID3 stem–loop comprises 23 nt, which form a 6-bp stem (three terminal G-C pairs were added to increase thermal stability) and an 11-nt loop. The 10-nt oligomer designed to represent the target contains the native IBS1 sequence, including 2 nt of the native intron sequence adjacent to the 5′ splice site. Slightly longer IBS1 target oligomers (14 nt) were designed for use in heteronuclear NMR experiments to optimize yield of in vitro transcriptions.

Preparation of the RNA samples

All RNA oligomers representing the ID3 stem–loop (5′-GGGUGUA UUGGAAAUGAGCACCC-3′) were transcribed in vitro by T7 RNA polymerase (expressed and purified in the Greenbaum laboratory) (He et al. 1997) from a double-stranded DNA template. RNA was purified using a 20% denaturing polyacrylamide gel, and the desired band was eluted from gel slices by electroelution or crushed and soaked in TBE buffer. RNA was precipitated by ethanol precipitation, and the buffer was exchanged by successive washes in 1000 MWCO Omega concentrators (Pall). Samples were dried and resuspended in autoclaved ultrapure water and the indicated salt concentration at the indicated pH. For the heteronuclear NMR study, a uniformly labeled ID3 molecule was transcribed using $^{13}C/^{15}N$ labeled NTPs (Cambridge Isotopes Laboratories [CIL]) and purified as described above.

Unlabeled IBS1 oligomers (5′-ACAUUUUCGA-3′) were purchased form Dharmacon, deprotected according to the supplied protocol, and purified as described above. Two samples of IBS1 oligomers were transcribed in vitro with labeled nucleotides, with $^{13}C/^{15}N$-labeled CTP (*C) (5′-GGA*CAUUUU*CGAG*C-3′) and labeled GTP (*G) (5′-*G*GACAUUUU C*GA*GC-3′), using a DNA template that included a hammerhead ribozyme sequence for cleavage at the desired site during transcription. RNA oligomers were purified as described above.

Concentration was determined using Varian Cary 50 UV-Vis spectrophotometer and extinction coefficients of 233400 (L/mol/cm) for ID3, 128900 (L/mol/cm) for IBS1 (14 nt), and 104100 (L/mol/cm) for IBS1 (10 nt).

Gel mobility shift assay of the ID3–IBS1 complex

Samples used for the nondenaturing PAGE comprised the ID3 stem–loop and the IBS1 target oligomers (10 nt). RNA oligomers in a solution containing 100 mM NaCl were mixed in stoichiometric ratios and annealed by heating to 72°C and snap cooled. The gels used for nondenaturing PAGE contained 20% polyacrylamide (19:1 acrylamide:bisacrylamide), Tris-HEPES buffer, and 20 mM NaCl (pH 8.0) with and without 10 mM $MgCl_2$. Electrophoresis was performed at 4°C. Gels were stained with SybrGold or Nuclistain.

CD of ID3 stem–loop and ID3–IBS1 complex

Samples used for CD comprised 22 μM ID3 stem–loop and a mixture of 1.0:1.05 ratio of ID3–IBS1 strands (the slight excess of the target strand was included to minimize the amount of free ID3). The samples were dissolved in 10, 50, and 100 mM KCl and 50 and 100 mM NaCl with or without 400 μM $MgCl_2$. CD spectra were recorded on an Aviv Biomedical CD spectrophotometer (model 202-01) equipped with a thermoelectric temperature controller. A quartz cuvette with path length of 0.1 cm and a volume of 400 μL was used (Starna Cells). All spectra were collected at 10°C. The background signal was subtracted, and spectra were plotted using Microsoft Excel. Ellipticity was shown in millidegrees (mdeg).

NMR data collection for structure calculation of the ID3 stem–loop

The final concentration of RNA oligomers used for NMR experiments was ~1 mM, unless otherwise noted. NMR samples of ID3 and ID3–IBS1 complex were prepared. For the study of exchangeable protons, the ID3 stem–loop and ID3–IBS1 complex were dissolved in 50–100 mM sodium chloride and 0.1 mM EDTA, in 90% H_2O/7% D_2O (99.96% at pH 6.4; Cambridge Isotope Laboratories). Microvolume NMR tubes (Shigemi) were used for all NMR data collection. For the observation of nonexchangeable protons, samples were lyophilized and resuspended in 99.96% D_2O (DSS; Cambridge Isotope Laboratories) twice and dried a final time and resuspended in 99.96% D_2O, 0.01 mg/mL DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid). The absolute standard frequencies for chemical shifts of ^{13}C were indirectly referenced by multiplying the absolute frequency for ^{1}H signal (using the methyl resonance of the DSS reference) by 0.251449530 (Wishart et al. 1995; Markley et al. 1998).

Some experiments were carried out in the presence of 5 mM $MgCl_2$ in order to evaluate the possibility of magnesium ion-dependent structural changes. Magnesium was removed by addition of 10 mM EDTA and multiple washes of the sample in 1000 MWCO filter (Pall).

All NMR experiments used in structural studies of the ID3 stem–loop samples were collected on a 600-MHz Bruker Avance NMR spectrometer equipped with a TCI cryoprobe (Hunter College). 2D experiments were processed using Bruker TopSpin 2.1 software and nmrPipe (Delaglio et al. 1995). 3D experiments were processed using nmrPipe (Delaglio et al. 1995). Spectra were apodized with a cosine-squared window function and zero-filled in all the direct and indirect dimensions. 2D Spectra were analyzed and assigned using Sparky (T.D. Goddard and D.G. Kneller, University of California, San Francisco) and CCPN (Vranken et al. 2005).

2D NOESY spectra of the exchangeable protons were collected using excitation sculpting techniques (Callihan et al. 1996) in order to assess the base-pairing patterns of the ID3 stem–loop and the

ID3–IBS1 complex. Mixing times of 120 and 200 msec were used in collection of NOESY spectra, at 0–4°C to slow the rate of exchange of RNA protons with the bulk water ^{1}Hs. The methyl proton resonance of DSS was referenced to 0 ppm. Nonexchangeable proton spectra were recorded by using a phase-sensitive pulse program, which employs presaturation for suppression of the HDO signal. Spectra of the ID3 stem–loop were collected at 25°C, while those of the ID3–IBS1 complex were collected at 15°C. Pyrimidine H5-H6 cross-peaks were observed in TOCSY spectra collected with 35- and 70-msec mixing times.

Initial assignments of ribose protons and carbons was achieved with the use of constant time heteronuclear single quantum correlation spectroscopy (^{13}C-ctHSQC) (Vuister and Bax 1992). Aromatic protons and carbons of bases were assigned through the use of nonconstant time gradient sensitivity enhanced (Palmer et al. 1991) HSQC. Adenine H2 assignments of A20, A7, and A17 residues were made by correlating H2 and H8 resonances by using a TROSY relayed HCCH-COSY pulse sequence (Simon et al. 2001). Chemical-shift assignments of ribose spin systems were made by using a gradient enhanced triple-resonance HCCH-TOCSY (Kay et al. 1993) and HCCH-COSY experiments. Hydrogen bonds between Watson-Crick pairing partners were observed directly through the use of an HNN-COSY experiment (Dingley and Grzesiek 1998) in the fully labeled ID3 molecule.

1D proton-decoupled ^{31}P spectra of the ID3 stem–loop and ID3–IBS1 complex were acquired in order to assess the backbone dihedral angles. A 500-MHz Varian INOVA spectrometer with an HCX probe equipped with *z*-axis gradients was used for collection of ^{31}P spectra. The X-channel was tuned to 202.3 MHz in order to detect phosphorous. Eighty-five percent phosphoric acid (H_3PO_4) was used as an external referencing standard at 0.00 ppm (3.46 ppm upfield of trimethyl phosphate). Spectra were collected at 15°C –25°C.

Structural restraints used for the structure calculation of the ID3 stem–loop

Distances were calculated from peak volumes in a NOESY spectrum at 150 msec and given error bounds of ±30%. NOEs observed only at longer mixing times were estimated to be weak (2.5–5.4 Å) or very weak (3.0–6.6 Å) from their appearance in a NOESY spectrum acquired at 350 msec.

Ribose puckers of U8, U9, and U15 were estimated to be C2′-*endo*, based on strong H1′-H2′ and H1′-H3′ cross-peaks observed at 70 msec mixing time in TOCSY spectra. The υ_1, υ_2, and δ dihedral angles for riboses in C2′-*endo* conformation were set to 35 ± 5°, −35 ± 5°, and 140 ± 20°, respectively. The remaining loop residues, including the stem-closing G•U wobble, were left unconstrained. For stem residues (2–5 and 19–22), ribose pucker was constrained to C3′-*endo*, based on the absence of H1′-H2′ cross-peaks, and the υ_1, υ_2, and δ angles were set to −20 ± 10°, 35 ± 5°, and 80 ± 20°, respectively. υ_1, υ_2, and δ angles of G18 of the G•U stem closing pair were loosely constrained to −20 ± 20°, 35 ± 10°, and 80 ± 60°, respectively. Dihedral angles for the first base pair of the stem G1-C23 were left unconstrained.

The α and ζ torsion angles of the backbone were conservatively constrained to 0 ± 120° to exclude *trans* conformation of all loop residues, based on chemical shifts of the ^{31}P resonances (Gorenstein 1984). The ε torsion angle of loop residues was conservatively constrained to exclude gauche$^+$ conformation (−120 ± 120°). Backbone torsion angles α, β, ε, and ζ of stem residues (2–5 and 18–22) were constrained to A-form values: −60 ± 40°, 180 ± 50°, −160 ± 50°, and −70 ± 50°, respectively. The γ torsion angle of stem residues (2–5 and 18–22) was constrained to gauche+ conformation 60 ± 30°. The χ-angles for G10 and G11 were constrained to *syn* conformation due to the intense H6-H1′ NOE observed at a mixing time of 50 msec. The χ-angles for A12 and A13 were left unrestrained because of strong overlapping sequential H6/8-H1′ cross-peaks. The remaining bases were loosely restrained to anti-conformation (χ = −150 ± 90°).

Structure calculation and refinement

The structure of the ID3 stem–loop was calculated using the TAMD protocol (Rice and Brunger 1994; Stein et al. 1997) implemented in the XPLOR-NIH package. A total of 100 structures with proper geometry were generated. The protocol follows four stages, performed without planarity restraints. The initial high-temperature molecular dynamics stage consisted of 8000 steps with a time step of 0.008 psec, at 20,000 K. During this stage, the weight on the NOE distance restraints and repulsive (van der Waals) terms was 150 kcal/mol/Å^2 and 0.1 kcal/mol/Å^2, respectively. The energy constant for the dihedral angles was set to 5 kcal/mol/rad^2. During the second stage of simulated annealing, temperature was lowered to 1000 K, while 10,000 steps of TAMD were performed with a time step of 0.007 psec. The weight on the repulsive term was linearly increased to unity. The force constant for the dihedral term was set to 100 kcal/mol/rad^2. During the third stage, the molecule was cooled from 1000 K to 300 K, while 2000 steps of Cartesian molecular dynamics were performed, with a time step of 0.003 psec. The final minimization stage consisted of 1000 steps of Powell minimization.

Finally, structures were refined using the DELPHIC potential for relative position of close bases and dihedral angles of nucleic acids (Clore and Kuszewski 2003). The database potentials were used only for stem residues 2–5 and 19–22. Planarity restraints were used for base pairs comprising residues 2–6 and 18–22. Ten structures without violations of NOE distances >0.5 Å and dihedral angles >5° were selected among the 20 structures of lowest energy.

Comparison of ^{13}C chemical shifts between ID3 and NMPs

We compared and plotted chemical shift difference between NMPs (Jezowska-Trzebiatowska et al. 1980; Kishore et al. 2005; Eichhorn et al. 2012; Kwan 2012) and ID3. The chemical shift difference follows δ (NMP)-δ (ID3) convention, to obtain positive values for differences in δ, because NMPs display mostly downfield shifts relative to those of structured residues.

Quantification of C8-H8 resonance intensity

We quantified C6/8-H6/8 resonance intensities in nonconstant time ^{13}C HSQC spectrum acquired at 15°C. Intensities were compared to the intensity C8/H8 resonances G3 residue (value normalized to 1.0), which is in middle of the Watson-Crick paired stem and therefore structured.

DATA DEPOSITION

Coordinates of the final structures and NMR restraints have been deposited to the Protein Data Bank, and the chemical shifts were deposited to the BMRB. The PDB accession no. for the structure of ID3 is 2M12. The BMRB deposition ID is 18838.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank Dr. Manjula D. Kolipaka for assistance with structure calculations and helpful discussions, Dr. Matthew Devany for guidance with NMR experiments, Mr. Mina Farid for technical assistance with PAGE, and Mr. Ali Ergun for streamlining the precision calculations of the family of structures; and the NMR facilities at Hunter College of CUNY and at the New York Structural Biology Center. This research was supported by NSF grant MCB 0929394 to N.L.G. The project described was supported by grant RR003037 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and its contents are solely the responsibility of the authors and do not necessarily represent the official views of NCRR or NIH.

Received April 23, 2013; accepted October 9, 2013.

REFERENCES

Bloomfield VA, Crothers DM, Tinoco I Jr. 2000. Electronic and vibrational spectroscopy. In *Nucleic acids* (ed. Stiefel J), pp. 165–222. University Science Books, Sausalito, CA.

Boudvillain M, de Lencastre A, Pyle AM. 2000. A tertiary interaction that links active-site domains to the 5′ splice site of a group II intron. *Nature* **406:** 315–318.

Bouvet P, Allain FH, Finger LD, Dieckmann T, Feigon J. 2001. Recognition of pre-formed and flexible elements of an RNA stem-loop by nucleolin. *J Mol Biol* **309:** 763–775.

Callihan D, West J, Kumar S, Schweitzer BI, Logan TM. 1996. Simple, distortion-free homonuclear spectra of peptides and nucleic acids in water using excitation sculpting. *J Magn Reson B* **112:** 82–85.

Candales MA, Duong A, Hood KS, Li T, Neufeld RA, Sun R, McNeil BA, Wu L, Jarding AM, Zimmerly S. 2012. Database for bacterial group II introns. *Nucleic Acids Res* **40:** 187–190.

Chen AA, Draper DE, Pappu RV. 2009. Molecular simulation studies of monovalent counterion-mediated interactions in a model RNA kissing loop. *J Mol Biol* **390:** 805–819.

Clore GM, Kuszewski J. 2003. Improving the accuracy of NMR structures of RNA by means of conformational database potentials of mean force as assessed by complete dipolar coupling cross-validation. *J Am Chem Soc* **125:** 1518–1525.

Costa M, Michel F. 1999. Tight binding of the 5′ exon to domain I of a group II self-splicing intron requires completion of the intron active site. *EMBO J* **18:** 1025–1037.

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

Costa M, Michel F, Westhof E. 2000. A three-dimensional perspective on exon binding by a group II self-splicing intron. *EMBO J* **19:** 5007–5018.

Dai L, Zimmerly S. 2002. Compilation and analysis of group II intron insertions in bacterial genomes: Evidence for retroelement behavior. *Nucleic Acids Res* **30:** 1091–1102.

Dai L, Toor N, Olson R, Keeping A, Zimmerly S. 2003. Database for mobile group II introns. *Nucleic Acids Res* **31:** 424–426.

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: A multidimensional spectral processing system based on Unix pipes. *J Biomol NMR* **6:** 277–293.

DePristo MA, de Bakker PI, Blundell TL. 2004. Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure* **12:** 831–838.

Dingley AJ, Grzesiek S. 1998. Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^{2}J_{NN}$ couplings. *J Am Chem Soc* **120:** 8293–8297.

Eichhorn CD, Feng J, Suddala KC, Walter NG, Brooks CL 3rd, Al-Hashimi HM. 2012. Unraveling the structural complexity in a single-stranded RNA tail: Implications for efficient ligand binding in the prequeuosine riboswitch. *Nucleic Acids Res* **40:** 1345–1355.

Fares C, Amata I, Carlomagno T. 2007. ^{13}C-detection in RNA bases: Revealing structure–chemical shift relationships. *J Am Chem Soc* **129:** 15814–15823.

Gorenstein DG. 1984. *Phosphorus-31 NMR: Principles and applications*. Academic Press, New York.

Greene KL, Wang Y, Live D. 1995. Influence of the glycosidic torsion angle on 13C and 15N shifts in guanosine nucleotides: Investigations of G-tetrad models with alternating syn and anti bases. *J Biomol NMR* **5:** 333–338.

Guo H, Karberg M, Long M, Jones JP 3rd, Sullenger B, Lambowitz AM. 2000. Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science* **289:** 452–457.

He B, Rong M, Lyakhov D, Gartenstein H, Diaz G, Castagna R, McAllister WT, Durbin RK. 1997. Rapid mutagenesis and purification of phage RNA polymerases. *Protein Expr Purif* **9:** 142–151.

Hetzer M, Wurzer G, Schweyen RJ, Mueller MW. 1997. *Trans*-activation of group II intron splicing by nuclear U5 snRNA. *Nature* **386:** 417–420.

Hoogstraten CG, Legault P, Pardi A. 1998. NMR solution structure of the lead-dependent ribozyme: Evidence for dynamics in RNA catalysis. *J Mol Biol* **284:** 337–350.

Jacquier A, Jacquesson-Breuleux N. 1991. Splice site selection and role of the lariat in a group II intron. *J Mol Biol* **219:** 415–428.

Jacquier A, Michel F. 1987. Multiple exon-binding sites in class II self-splicing introns. *Cell* **50:** 17–29.

Jezowska-Trzebiatowska B, Kozlowski H, Wolowiec S. 1980. Coordination of Gly-Tyr × Pd(II) complex to GMP nucleotide. *Acta Biochim Pol* **27:** 99–109.

Johnson JE Jr, Hoogstraten CG. 2008. Extensive backbone dynamics in the GCAA RNA tetraloop analyzed using ^{13}C NMR spin relaxation and specific isotope labeling. *J Am Chem Soc* **130:** 16757–16769.

Karberg M, Guo H, Zhong J, Coon R, Perutka J, Lambowitz AM. 2001. Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nat Biotechnol* **19:** 1162–1167.

Kay LE, Xu GY, Singer AU, Muhandiram DR, Formankay JD. 1993. A gradient-enhanced HCCH TOCSY experiment for recording side-chain H-1 and C-13 correlations in H_2O samples of proteins. *J Magn Reson B* **101:** 333–337.

Kishore AI, Mayer MR, Prestegard JH. 2005. Partial ^{13}C isotopic enrichment of nucleoside monophosphates: Useful reporters for NMR structural studies. *Nucleic Acids Res* **33:** e164.

Kruschel D, Sigel RK. 2008. Divalent metal ions promote the formation of the 5′-splice site recognition complex in a self-splicing group II intron. *J Inorg Biochem* **102:** 2147–2154.

Kwan ICM. 2012. "Structural elucidation of guanosine self-assemblies using spectroscopic and computational methods". *PhD thesis*, Queen's University, Kingston, Ontario, Canada.

Lambert D, Leipply D, Shiman R, Draper DE. 2009. The influence of monovalent cation size on the stability of RNA tertiary structures. *J Mol Biol* **390:** 791–804.

Lambowitz AM, Zimmerly S. 2011. Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol* **3:** a003616.

Marcia M, Pyle AM. 2012. Visualizing group II intron catalysis through the stages of splicing. *Cell* **151:** 497–507.

Marcia M, Somarowthu S, Pyle AM. 2013. Now on display: A gallery of group II intron structures at different stages of catalysis. *Mob DNA* **4:** 14.

Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE, Wuthrich K. 1998. Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB Inter-Union Task Group on the Standardization of Data Bases of Protein and Nucleic Acid Structures Determined by NMR Spectroscopy. *J Biomol NMR* **12:** 1–23.

Mastroianni M, Watanabe K, White TB, Zhuang F, Vernon J, Matsuura M, Wallingford J, Lambowitz AM. 2008. Group II intron-based gene targeting reactions in eukaryotes. *PLoS One* **3:** e3121.

Meunier B, Tian G-L, Macadre C, Slonimski P, Lazowska J. 1990. Group II introns transpose in yeast mitochondria. In *Structure function and biogenesis of energy transfer systems* (ed. Quagliariello E, et al.), pp. 169–174. Elsevier Scientific Publishers, Amsterdam.

Newby MI, Greenbaum NL. 2001. A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. *RNA* **7:** 833–845.

Newby MI, Greenbaum NL. 2002. Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nat Struct Biol* **9:** 958–965.

Nikolova EN, Kim E, Wise AA, O'Brien PJ, Andricioaei I, Al-Hashimi HM. 2011. Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **470:** 498–502.

Palmer AG, Cavanagh J, Wright PE, Rance M. 1991. Sensitivity improvement in proton-detected 2-dimensional heteronuclear correlation NMR-spectroscopy. *J Magn Reson* **93:** 151–170.

Perutka J, Wang W, Goerlitz D, Lambowitz AM. 2004. Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *J Mol Biol* **336:** 421–439.

Qin PZ, Pyle AM. 1997. Stopped-flow fluorescence spectroscopy of a group II intron ribozyme reveals that domain 1 is an independent folding unit with a requirement for specific Mg^{2+} ions in the tertiary structure. *Biochemistry* **36:** 4718–4730.

Qin PZ, Pyle AM. 1998. The architectural organization and mechanistic function of group II intron structural elements. *Curr Opin Struct Biol* **8:** 301–308.

Qin PZ, Pyle AM. 1999. Antagonistic substrate binding by a group II intron ribozyme. *J Mol Biol* **291:** 15–27.

Rice LM, Brunger AT. 1994. Torsion angle dynamics: Reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins* **19:** 277–290.

Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. 2003. The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* **160:** 65–73.

Shajani Z, Drobny G, Varani G. 2007. Binding of U1A protein changes RNA dynamics as observed by ^{13}C NMR relaxation studies. *Biochemistry* **46:** 5875–5883.

Sigel RKO. 2005. Group II intron ribozymes and metal ions: A delicate relationship. *Eur J Inorg Chem* **2005:** 2281–2292.

Simon B, Zanier K, Sattler M. 2001. A TROSY relayed HCCH-COSY experiment for correlating adenine H2/H8 resonances in uniformly C-13-labeled RNA molecules. *J Biomol NMR* **20:** 173–176.

Stein EG, Rice LM, Brunger AT. 1997. Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation. *J Magn Reson* **124:** 154–164.

Su LJ, Qin PZ, Michels WJ, Pyle AM. 2001. Guiding ribozyme cleavage through motif recognition: The mechanism of cleavage site selection by a group II intron ribozyme. *J Mol Biol* **306:** 655–668.

Toor N, Hausner G, Zimmerly S. 2001. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* **7:** 1142–1152.

Toor N, Keating KS, Fedorova O, Rajashankar K, Wang J, Pyle AM. 2010. Tertiary architecture of the *Oceanobacillus iheyensis* group II intron. *RNA* **16:** 57–69.

Uhlenbeck OC. 1987. A small catalytic oligoribonucleotide. *Nature* **328:** 596–600.

Varani G, Tinoco I Jr. 1991. Carbon assignments and heteronuclear couplings constants for an RNA oligonucleotide from natural abundance ^{13}C-^{1}H correlated experiments. *J Am Chem Soc* **113:** 9349–9354.

Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas P, Ulrich EL, Markley JL, Ionides J, Laue ED. 2005. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* **59:** 687–696.

Vuister GW, Bax A. 1992. Resolution enhancement and spectral editing of uniformly C-13-enriched proteins by homonuclear broad-band C-13 decoupling. *J Magn Reson* **98:** 428–435.

Waldsich C, Pyle AM. 2007. A folding control element for tertiary collapse of a group II intron ribozyme. *Nat Struct Mol Biol* **14:** 37–44.

Wang Y, Silverman SK. 2006. Experimental tests of two proofreading mechanisms for 5′-splice site selection. *ACS Chem Biol* **1:** 316–324.

Wank H, SanFilippo J, Singh RN, Matsuura M, Lambowitz AM. 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.

Wijmenga SS, Heus HA, Werten B, van der Marel GA, van Boom JH, Hilbers CW. 1994. Assignment strategies and analysis of cross-peak patterns and intensities in the three-dimensional homonuclear TOCSY-NOESY of RNA. *J Magn Reson B* **103:** 134–141.

Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD. 1995. 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR* **6:** 135–140.

Xiang Q, Qin PZ, Michels WJ, Freeland K, Pyle AM. 1998. Sequence specificity of a group II intron ribozyme: Multiple mechanisms for promoting unusually high discrimination against mismatched targets. *Biochemistry* **37:** 3839–3849.

Yang J, Mohr G, Perlman PS, Lambowitz AM. 1998. Group II intron mobility in yeast mitochondria: Target DNA-primed reverse transcription activity of aI1 and reverse splicing into DNA transposition sites in vitro. *J Mol Biol* **282:** 505–523.

Zaug AJ, Been MD, Cech TR. 1986. The *Tetrahymena* ribozyme acts like an RNA restriction endonuclease. *Nature* **324:** 429–433.

Zhang Q, Sun X, Watt ED, Al-Hashimi HM. 2006. Resolving the motional modes that code for RNA adaptation. *Science* **311:** 653–656.

Zhuang F, Karberg M, Perutka J, Lambowitz AM. 2009. EcI5, a group IIB intron with high retrohoming frequency: DNA target site recognition and use in gene targeting. *RNA* **15:** 432–449.