Linking the group II intron catalytic domains: tertiary contacts and structural features of domain 3

Olga Fedorova and Anna Marie Pyle*

Howard Hughes Medical Institute, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA

Despite its importance for group II intron catalytic activity, structural information on conserved domain 3 (D3) is extremely limited. This domain is known to specifically stimulate the chemical rate of catalysis and to function as a ‘catalytic effector’. Of all the long-range tertiary contacts that have been identified within group II introns, none has included D3 residues. Furthermore, little is known about the atoms and functional groups in D3 that contribute to catalysis. Using a nucleotide analog interference mapping assay with an extended repertoire of nucleotide analogs, we have identified functional groups in D3 that are critical for ribozyme activity. These data, together with mutational analysis, suggest the formation of noncanonical base pairs within the phylogenetically conserved internal loop at the base of D3. Finally, a related nucleotide analog interference suppression study resulted in the identification of a direct tertiary interaction between D3 and catalytic domain 5, which sheds new light on D3 function in the group II intron structure and mechanism.

The EMBO Journal (2005) **24**, 3906–3916. doi:10.1038/sj.emboj.7600852; Published online 27 October 2005

Subject Categories: RNA

Keywords: enzymology; ribozyme; RNA folding; splicing; tertiary structure

Introduction

Group II introns are large catalytic RNAs that self-splice by a mechanism similar to spliceosomal processing of pre-mRNA. In addition to catalyzing the two-step self-splicing reaction, group II introns are also mobile genetic elements that are capable of inserting themselves into an RNA or double-stranded DNA. Group II introns are generally found in bacteria and lower eukaryotes such as plants, fungi and algae (Lehmann and Schmidt, 2003; Pyle and Lambowitz, 2006).

Surprisingly for such multifunctional ribozymes, the primary sequence of group II introns is not highly phylogenetically conserved. However, there is strong conservation of the intron secondary structure, which is subdivided into six domains (Michel *et al*, 1989) (Figure 1). Each intronic domain has a distinct functional role in splicing and/or

mobility. Of the six domains, only domains 1 (D1) and 5 (D5) are indispensable for chemical catalysis (Koch *et al*, 1992; Michels and Pyle, 1995). D1 is the scaffold for the assembly of the catalytic core, while the highly conserved D5 is a central component of the ribozyme active site (Lehmann and Schmidt, 2003; Pyle and Lambowitz, 2006). Domain 3 (D3) has long been a subject of interest because, while not strictly required for catalysis, it markedly stimulates the chemical rate of ribozyme reactions and is actually required for *cis*-splicing (Griffin *et al*, 1995; Podar *et al*, 1995; Konforti *et al*, 1998; Fedorova *et al*, 2003).

Due to its apparent role as a ‘catalytic effector’ in group II intron function, we have focused recent attention on the structure and function of D3 (Fedorova *et al*, 2003). We have observed that D3 does not facilitate the binding of catalytic domain D5 (Konforti *et al*, 1998), which suggests that D3 may have its own role to play in catalysis. Although previous results indicate that the basal stem of D2 recruits D3 into the ribozyme active site (Fedorova *et al*, 2003), the modes of interaction between D3 and the rest of the intron have remained a mystery.

The lack of phylogenetic covariation between D3 and other intronic domains suggests that D3 does not form any long-range tertiary contacts through simple Watson–Crick base-pairing. Nonetheless, D3 binds the rest of the intron with relatively high affinity: the K_d of D3 is approximately 500 nM (Podar *et al*, 1995), which is comparable to that of D5 (Pyle and Green, 1994). These data indicate that D3 forms an extensive network of tertiary contacts with other domains, which is consistent with the fact that sections of D3 are highly protected during hydroxyl radical protection assays (Swisher *et al*, 2001). DEPC modification interference studies suggest that D3 interacts with D5 (Jestin *et al*, 1997) and that the internal bulge, basal stem and the pentaloop of D3 are important functional regions.

By using nucleotide analog interference mapping (NAIM) analysis with an extended set of nucleotide analogs, we have identified atoms and functional groups in D3 that are critical for group II intron catalysis. Analysis of the interference patterns and phylogenetic data allows us to predict a likely conformation for the most conserved substructure in D3—the internal bulge within the basal stem. In addition, by combining forward and reverse nucleotide analog interference suppression (NAIS) results, we have identified the first direct tertiary contact between D3 and the rest of the intron. This interaction joins two bases of the D3 pentaloop with the backbone of G844 in D5. Taken together, the data provide new insight for the role of D3 in group II intron catalysis.

Results

NAIM identifies critical nucleobase atoms in D3

Experimental design. As in previous NAIM and NAIS studies of the group II intron (Boudvillain and Pyle, 1998; Boudvillain

*Corresponding author. Howard Hughes Medical Institute, Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, Box 208114, New Haven, CT 06520, USA.
Tel.: +1 203 432 5733; Fax: +1 203 432 5316;
E-mail: anna.pyle@yale.edu

Received: 8 August 2005; accepted: 6 October 2005; published online: 27 October 2005

Figure 1 A secondary structure of the ai5γ group II intron. The new tertiary interaction μ-μ' is shown in dark gray. Nucleotides involved in the μ-μ' interaction (G844, A617 and A618) are highlighted by gray shading. The tandem sheared A·A pairs in the internal bulge of D3 are denoted in light gray by respective symbols according to base-pairing classification by Leontis and Westhof (1998); Leontis *et al.*, 2002).

et al., 2000), the present study employed a two-piece construct in which branching occurs *in trans*, thereby providing a system for *in-vitro* selection. The branch site of D56 RNA attacks the 5'-splice site of exD123 RNA (Figure 2) and becomes covalently attached. While early NAIM studies utilized this approach to identify several important atoms in D3, the analysis was necessarily limited (Boudvillain and Pyle, 1998; Boudvillain *et al.*, 2000). The present study employed an expanded set of nucleotide analogs, particularly for probing adenosine positions, which are predominant in D3 (Figure 1). To specifically examine nucleobase functional groups in D3, we transcribed exD123 RNA in the presence of 7-deaza adenosine (7-deaza A), 2,6-diaminopurine (2,6-DAP), N6-methyl adenosine (N6-MeA) and inosine (I) α-thiotriphosphates, which allowed us to probe the Hoogsteen face, major groove and minor groove of A residues and the minor groove of G residues for involvement in important tertiary contacts.

It was previously demonstrated that there are no phosphorothioate interferences in D3 (Boudvillain and Pyle, 1998), and therefore any observed interference can be attributed to an analog effect. Inosine interferences are also not observed in D3 (Fedorova *et al.*, 2003), suggesting that guanosine residues in D3 do not form tertiary contacts involving the exocyclic amino group, or that any such contacts are dispensable for catalysis. However, analogs of adenosine were highly informative, resulting in clusters of residues that exhibit multiple interference effects (Figure 3).

Interferences in the internal bulge. Some of the strongest nucleobase NAIM effects are clustered in the A-rich internal bulge of D3, which is the most phylogenetically conserved part of the domain (Michel *et al.*, 1989; Michel and Ferat, 1995; Toor *et al.*, 2001). Three adenosines in this region showed a combination of 7-deaza A, N6-MeA and 2,6-DAP

Figure 2 A two-piece branching construct used in NAIM and NAIS studies. An asterisk denotes the ^{32}P label at the 3'-end of exD123 for NAIM and reverse NAIS analysis of D3. For the forward NAIS assay exD123 is not radiolabeled and D56 is 5'-end labeled (see Materials and methods).

effects (Figure 3), suggesting that they are involved in non-canonical base pairing and/or formation of base-triples from both the minor groove edge and the Hoogsteen edge. The N6-MeA effects may also be interpreted as an indication of tight packing of the internal loop, whereby the methyl group on amine group causes a steric clash.

Interference effects in the hairpin-loop regions. Of the three hairpin-loop regions that extend from D3, NAIM effects are particularly pronounced in the pentaloop G615–U619, which is consistent with its important role in catalysis (Boudvillain and Pyle, 1998). Both adenosine residues A617 and A618 exhibit 2,6-DAP interferences (Figure 3). The interference pattern at A618 also includes a 7-deaza A effect (Figure 3) as well as 2'-analog effects (see below). The other two stem-loop regions contain few catalytically critical functionalities. No interference effects were observed in the tetraloop UAUU (residues 635–638), suggesting that it is not important for catalysis. This is consistent with this substructure being the most variable region in D3 according to the phylogenetic analysis (Michel and Ferat, 1995). In the GAAA tetraloop (residues 650–653), the last A (A653) exhibits N6-methyl A and 2,6-DAP interferences and A652 does not show any interferences at all (Figure 3), which are inconsistent with the pattern expected for a GNRA-like fold.

Interferences in the linker regions. The present study reveals, for the first time, that the D3 linker regions (nucleotides 603–607, 627–629, 644–645; Figure 3) contain functional groups that are critical for group II intron catalysis. Both A604 and A605 exhibit 7-deaza A and 2,6-DAP interference effects (Figure 3). A605 also shows interference with the N6-MeA analog. One of the strongest 2,6-DAP effects in D3 is observed at A627 (Figure 3). A629 exhibits a weak 7-deaza A interference, and there is a 2,6-DAP effect at A630 (Figure 3). These data suggest that the linker nucleotides may be involved in noncanonical pairings with each other, as observed for other large internal RNA loops, and/or they are involved in long-range tertiary contacts with other intronic domains. Importantly, the same experiments reveal a combination of 7-deaza A, 2,6-DAP and N6-MeA effects at A589, which is located in the adjacent J2/3 region (the single-stranded linker

between domains 2 and 3). This nucleotide was previously shown to be critical for the second step of splicing (Mikheeva *et al*, 2000); however, the interference results indicate that this nucleotide is important for the first step as well.

NAIM elucidates the location and role of important 2'-hydroxyl (OH) groups in D3

Early NAIM work implicated three 2'-OH groups in the function of D3 (at A618, A661 and A662 (Boudvillain and Pyle, 1998)). To determine the mechanistic role of these 2'-OH groups and to determine if ribose functionalities at other positions are important, we conducted NAIM with exD123 RNAs transcribed in the presence of 2'-fluoro- and 2'-O-methyl adenosine thiotriphosphates (2'-fluoro adenosine (2'-FA) and 2'-OMe adenosine (2'-OMe A), respectively). These analogs display NAIM effects at positions where the 2'-OH is a catalytically important hydrogen bond donor. In addition, if interference is observed with 2'-fluoro, but not 2'-deoxy, analogs, it indicates a sugar pucker in the unusual C2'-endo conformation (Ortoleva-Donnelly *et al*, 1998).

Both 2'-FA and 2'-OMe A interferences were observed at all three positions that had previously been implicated (A618, A661 and A662; Figure 4), suggesting that 2'-OH groups at these positions serve as hydrogen bond donors rather than as acceptors. Five additional positions were identified where interference was observed only with 2'-OMe A (A596–A598 and A605; Figure 4). In these cases, substitution of a hydroxyl with the bulkier methyl group is likely to create a steric clash with some functionality located in close proximity to the OH group (Ortoleva-Donnelly *et al*, 1998).

Mutational analysis for characterizing the D3 pentaloop

Consistent with previous studies (Jestin *et al*, 1997; Boudvillain and Pyle, 1998), the GUAU pentaloop in D3 is observed to play an important role in D3 function. Given the sequence of the loop, it might fold like a GNRA tetraloop, which could dock with a cognate receptor (Abramovitz and Pyle, 1997; Legault *et al*, 1998). Alternatively, it might fold into a different set of possible interaction motifs. To differentiate between these possibilities and to better characterize the role of constituent nucleotides, we conducted a mutational and kinetic analysis of the D3 pentaloop. Mutant

Figure 3 (A) An autoradiograph of representative high-resolution sequencing gels showing base analog interference effects after iodine cleavage of unreacted exD123 RNAs and branched products. Lines corresponding to precursor (unreacted exD123) and branched RNAs are labeled 'pre' and 'br', respectively. Numbers on the left indicate adenosine positions in D3. (B) Summary of reproducible base analog interference effects in D3. (C) Representative average interference values with different base analogs.

exD123 RNAs were constructed and their kinetic effects were tested using the *trans*-branching reaction with D56 RNA (Figure 2 and Table I). Branching was most severely affected when the D3 pentaloop was replaced with the tetraloop sequence UUCG (30-fold inhibition; Table I and Figure 5). Deletion of U619, which replaces the pentaloop with a GUAA tetraloop (Δ 619 mutant), decreases the rate of branching by 11-fold (Table I and Figure 5). Single-base mutations in the pentaloop were similarly revealing. Most notably, mutation of

G615 to an A or a U did not affect the reaction (Table I and Figure 5C), which is in good agreement with the NAIM results showing lack of interference at this position. When U619 is changed to an A, the mutant is as active as the wild type (Table I and Figure 5), suggesting that the adverse effect of Δ 619 is caused by changes in distance between critical functional groups in the pentaloop. Taken together, the mutational data are clearly inconsistent with a GNRA-like fold for the D3 pentaloop.

Figure 4 (A) An autoradiograph of representative high-resolution sequencing gels showing 2'-analog interference effects after iodine cleavage of unreacted exD123 RNAs and branched products. Lines corresponding to precursor (unreacted exD123) and branched RNAs are labeled 'pre' and 'br', respectively. Numbers on the left indicate adenosine positions in D3. (B) Summary of reproducible 2'-analog interference effects in D3. (C) Representative average interference values with different 2'-analogs.

Mutational analysis of positions 617 and 618 indicates that one or both adenosine residues of the pentaloop are important for the ribozyme function. The mutation of either A617 or A618 to a pyrimidine or a purine residue results in a 3–5-fold decrease in branching activity (Figure 5). However, when both adenosine residues are mutated to cytidines at the same time, the double mutant is as defective as the UUCG mutant (Table I and Figure 5). This indicates that both adenosine residues of the D3 pentaloop are important for catalysis, and that they may function together.

NAIS analysis reveals an important tertiary interaction between D3 and D5

Having confirmed that specific nucleotides in the pentaloop are important for catalytic function, cognate mutants in this region were used for NAIS analyses designed to identify interaction partners between the pentaloop and the rest of the intron. For these studies, D56 RNAs were transcribed in the presence of various nucleotide analog α -thiotriphosphates. The resultant pools of modified D56 molecules were then reacted with exD123 molecules that contain muta-

tions in the D3 pentaloop. After cleavage and electrophoresis, interference and suppression patterns in D5 and D6 were then examined using standard methods (Boudvillain *et al.*, 2000). Pentaloop mutations at positions A617 and/or 618 resulted in specific suppressions of dG interference at G844 in D5 (Figure 6). Importantly, mutations of the D3 internal loop (i.e. A600G), which also cause strong branching defects (Table I and Figure 5), do not suppress the dG interference at G844, indicating that the pentaloop suppressions are highly specific (Figure 6). Taken together, these data are consistent with a tertiary contact between the 2'-OH at G844 and functional groups on A617 and/or A618.

Table I Kinetic parameters for branching reactions of exD123 RNAs that contain mutations in D3

exD123	$k_{\text{obs}} \times 100 \text{ (min}^{-1}\text{)}$	$k_{\text{obs}} \text{ (rel)}$
WT	3.2 ± 0.2	1
A600G	0.35 ± 0.07	0.1
A617G	0.6 ± 0.1	0.19
A617U	1.0 ± 0.1	0.31
A618G	0.97 ± 0.03	0.3
G615A	2.0 ± 0.2	0.6
G615U	2.4 ± 0.3	0.75
U619A	2.1 ± 0.2	0.66
UUCG	0.13 ± 0.03	0.04
$\Delta 619$	0.265 ± 0.005	0.08
A617C, A618C	0.06 ± 0.01	0.018

Mutation of the entire pentaloop to a UUCG tetraloop caused a full suppression, whereas single-base mutations at A617 and A618 caused partial suppression of dG interference at this position (Figure 6). These data suggested that not just one, but both A617 and A618 may be involved in the interaction with D5. When both A617 and A618 were mutated to C residues at the same time, the double mutant also caused a full suppression of the dG interference at G844 (Figure 6), which supports this assumption. Notably, single-base mutations of U619 did not suppress dG844 interference even partially (Figure 6), although deletion of U619 ($\Delta 619$ mutant) results in full suppression of the above interference (Figure 6), much like the UUCG substitution. The loss of the D3–D5 interaction in the $\Delta 619$ case is possibly caused by promotion of an undesirable GNRA-fold. Take together, the NAIS results and mutational data support a critical tertiary interaction between the D3 pentaloop and the first stem of D5 (the μ – μ' interaction; Figure 1).

Interestingly, some mutations in the pentaloop of D3 caused an exacerbation of existing dG interference effects at G836 and G840 (Figure 6). This increased sensitivity to pentaloop mutations was not specific to this region of D5; it was also observed for the adenosine residues in the tetraloop of D5 when other analogs were used (data not shown). This indicates that, in order to compensate for the loss of the μ – μ' interaction, the intron puts more emphasis on tertiary

Figure 5 Mutations in the pentaloop of D3 reduce branching efficiency. (A) A schematic of mutations in the pentaloop of D3 used for kinetic studies. (B) Representative 5% denaturing gels showing a time-course of branching for wild-type and mutant exD123 RNAs. (C) Kinetic analysis of the influence of mutations in D3 on the rate of branching. Values of k_{obs} (Table I) were determined as described previously (Boudvillain and Pyle, 1998).

Figure 6 NAIS analysis demonstrates suppression of 2'-deoxy interference at G844 in D5 in the presence of exD123 RNAs with mutations in the pentaloop of D3. **(A)** The autoradiograph of representative high-resolution sequencing gels after iodine cleavage of unreacted D56 RNAs and branched products. The selection step (branching reaction) was carried out in the presence of wild-type and mutant unlabeled exD123 RNAs and 5'-end labeled wild-type D56 transcribed in the presence of dGTP α S analog. Numbers on the left indicate guanosine positions in D5. The position of G844 is indicated by an arrow. **(B)** A diagram comparing the relative intensity of the band corresponding to G844 in precursor (0) and branched products (WT and pentaloop mutants).

contacts involving other D5 residues, thus increasing energetic penalty for their disruption.

Reverse NAIS confirms the tertiary contacts between D3 and D5

Further evidence of a direct specific interaction between two adenosines (A617 and A618) of the D3 pentaloop and a guanosine residue (G844) of D5 was provided by a 'reverse' NAIS experiment. In this case, a D56 RNA was synthesized with a single 2'-deoxy substitution at position G844. This RNA was reacted with exD123 RNAs that were transcribed in the presence of nucleotide analog α -thiotriphosphates. The resultant iodine cleavage patterns of exD123 were then examined for interference suppressions in D3. Strikingly, 2,6-DAP interference at both A617 and A618 is suppressed upon reaction with the dG844 D56 molecule, suggesting that interaction with the pentaloop occurs on the minor groove edge of both adenosine residues (Figure 7). The effect was specific to the pentaloop adenosines: strong interference at A627 and A630 remained unchanged when the dG844 mutant was used instead of the wild-type D56 RNA (Figure 7). No suppression of 2,6-DAP interference was observed in the D3 pentaloop upon reaction with a D56 RNA containing a single atom modification on a minor groove edge of a neighboring

G845 (D56 I845) (Figure 7). Since the exo-cyclic amine of G845 was previously shown to be a part of a κ - κ' interaction, this suggests an absence of energetic coupling between κ - κ' and the μ - μ' interaction.

Discussion

Identification of μ - μ' : a critical tertiary contact between catalytic intron domains

We have successfully applied NAIS and mutational analysis to identify the first functional tertiary contact between D3 and another region of the intron. Consistent with a role for D3 in catalysis, specific interactions are observed between D3 and catalytic D5. The μ - μ' contact involves interactions between the ribose of G844 in D5 and the minor groove edge of two adenosine residues in the pentaloop of D3: A617 and A618 (Figures 1 and 6). Specifically, the data are consistent with interaction between the 2'-OH group of G844 and the N3 or H2 atoms of A617 and A618. The 2'-OH group of G844 was previously shown to be one of the few 2'-OH groups in D5 that directly affects the chemical rate of catalysis, rather than binding of D5 to the rest of the intron (Abramovitz *et al*, 1996). Since D3 specifically increases the chemical rate constant in group II intron ribozymes (Griffin *et al*, 1995;

Figure 7 Reverse NAIMS analysis confirms a direct interaction between 2'-OH of G844 in D5 and the pentaloop of D3. (A) The autoradiograph of representative high-resolution sequencing gels after iodine cleavage of unreacted exD123 RNAs and branched products. Selection step (branching reaction) was carried out in the presence of wild-type and mutant unlabeled D56 RNAs and 3'-end labeled wild-type exD123 transcribed in the presence of 2,6-DAP. Numbers on the left indicate adenosine positions in exD123. (B) A diagram comparing relative intensity of bands corresponding to A617 and A618 in precursor (0) and branched products.

Podar *et al.*, 1995; Xiang *et al.*, 1998; Fedorova *et al.*, 2003), but has no effect on D5 binding (Konforti *et al.*, 1998), our results suggest that the 2'-OH of G844 contributes to catalysis by anchoring D3 into the ribozyme active site.

Despite a growing database of correlations between NAIM patterns and the corresponding high-resolution crystal structures, it was not possible to assign μ - μ' as part of a known RNA structural motif. The four types of known A-minor motifs can be ruled out for the following reasons: A-minor type I and II motifs would require that one of the pentaloop adenosines (A617 or A618) interacts with the G844-C819 base pair while the other adenosine participates in an adjacent, energetically coupled A-minor interaction. However, the data are inconsistent with interactions that extend beyond G844. Furthermore, C819 has been unambiguously demonstrated to be involved in the very well characterized κ - κ' interaction, and its 2'-OH is fully engaged with interactions to its counterpart in D1 (Boudvillain and Pyle, 1998). A type III motif is fully inconsistent with the NAIMS data. We cannot completely exclude the possibility of the type 0 A-minor motif, although it has been predicted to be much more

tolerant to the C2 substitution at the adenosine position (Strobel, 2002), which is inconsistent with our data. Despite these issues, A617 and A618 may be involved in a form of A-minor motif that has not yet been reported.

The D3 pentaloop does not participate in a GNRA tetraloop-receptor interaction

At the first glance, it seemed likely that the GUAAU pentaloop (residues 615–619) might function as a GNRA tetraloop and that there is a receptor for it somewhere in the intron. In this case, the pentaloop would exhibit typical NAIM signatures such as strong 2'-deoxy and 7-deaza A interferences for third and fourth adenosine residues and a 2'-deoxy interference for the G (Boudvillain and Pyle, 1998). However, the 2,6-DAP interference effect at A618 is inconsistent with GNRA-like fold (Figure 3), as are the absence of 2'-deoxy and 7-deaza A interferences at A617. Since the G residue of a GNRA tetraloop is invariant, and its pairing with the last A is one of the signature features of the motif, mutation of the G is expected to severely affect catalytic function. However, the pentaloop is completely tolerant to G615 mutations, as this residue can be changed to an A or to a U without significant catalytic defect (Figure 5C). Importantly, these mutations have no effect on the μ - μ' interaction (Figure 6), which indicates that pentaloop tertiary contacts do not involve a GNRA-like fold. Finally, deletion of the last U in the pentaloop is detrimental for catalysis (Figure 5) and it disrupts the μ - μ' interaction (Figure 6), while a simple mutation of this U to an A (Figure 5) or a C (data not shown) does not affect the rate of branching at all, indicating that D3 function is disrupted when the pentaloop is forced into a GNRA-like conformation.

NAIM and phylogenetic analysis define the functional components of D3

NAIM results indicate that D3 contains a defined set of nucleotides that play a major role in D3 function. Consistent with previous studies (Jestin *et al.*, 1997; Boudvillain and Pyle, 1998; Konforti *et al.*, 1998) and with analysis of phylogenetic conservation, we observe that the major functional regions of D3 are the internal loop of the basal stem, the GUAAU pentaloop and in the linker regions that connect the four helical stems of the domain (Figures 3 and 4). Residues A599 and A662 in the internal loop, which exhibit strong interferences with a diversity of analogs, are highly conserved in almost all group II introns identified thus far (Michel *et al.*, 1989; Michel and Ferat, 1995; Toor *et al.*, 2001). A major functional role for these residues is supported by their high levels of protection against hydroxyl radicals (Swisher *et al.*, 2001).

Linker residues A604, A605, A627 and A629, which exhibited strong interference effects with different base analogs, are conserved in group II introns of chloroplast-like class 1 and 2 and semiconserved in group II introns of bacterial classes B and D (Toor *et al.*, 2001), suggesting that this region is important for D3 function in these classes of introns. However, D3 of mitochondrial class group II introns, which do not have single-stranded junctions, function in a different manner. The pentaloop of D3 is conserved in all group II introns from chloroplast-like classes and in most bacterial introns except for bacterial class C (Toor *et al.*, 2001).

Figure 8 A correlation of the experimental NAIM data for D3 bulge nucleotides with the predicted NAIM pattern for the AA platform (A) and a *trans*-Hoogsteen–sugar edge A · A base pair (B). Predicted NAIM effects are shown in red, experimentally identified NAIM effects are shown in green. The structure that best fits the experimental data is indicated by light gray shading.

Structural organization of the D3 internal loop

One of the most important elements in D3 is an A-rich internal loop in the basal stem that resembles a classic 11-nucleotide tetraloop receptor (see the ζ' region of D1 for comparison; Figure 1). Based on this similarity, we set out to determine if the D3 internal loop functions as a tetraloop receptor and, if not, what type of structure it is likely to adopt. NAIM analysis has been particularly successful at correctly predicting the conformational state of various RNA motifs (Ortoleva-Donnelly *et al.*, 1998; Strobel, 1999).

The signature feature of a tetraloop-receptor motif is the AA platform, in which two adjacent adenosine residues lie side by side and adopt characteristic sugar pucker configurations (Figure 8A, left) (Cate *et al.*, 1996a,b). If internal loop residues A661 and A662 form an AA platform, the sugar pucker of A661 will be C2'-endo. This will result in a strong NAIM effect with a 2'-fluoro analog, which forces a C3'-endo conformation of the sugar. Conversely, one will not observe interference with a 2'-deoxy analog (Figure 8A, middle panel) (Ortoleva-Donnelly *et al.*, 1998). In addition, 2,6-DAP and possibly 2'-OMe analog interferences will also be expected at this position (Ortoleva-Donnelly *et al.*, 1998). However, a strong 2'-deoxy interference (Boudvillain and Pyle, 1998)

and the absence of a 2,6-DAP interference at A661 contraindicate involvement in a platform. Furthermore, the strong 7-deaza, 2,6-diaminourine and 2'-analog interferences at A662 (Figures 3 and 4) are also inconsistent with the AA-platform NAIM signature (Figure 8A, middle panel). It is therefore highly unlikely that the D3 internal loop forms a classic 11-nucleotide tetraloop receptor motif.

Given the abundance of functional data on the loop residues, we set out to build a plausible model for the conformational state of the D3 internal loop. The interference patterns at A600, A599 and A662 indicate that these residues are engaged in interactions from both the major and minor grooves (Figures 3 and 4). One explanation for the observed NAIM is that these nucleotides form noncanonical base pairs across the loop, thereby creating an interface for interaction with another intronic substructure. The interference pattern at A600 is consistent with involvement in a reverse-Hoogsteen A600:U660 pair that is involved in additional interactions from the minor groove. Similarly, the data are consistent with A598:A662 and A599:A661 pairings. Having analyzed different classes of A · A pairs, we conclude that the bulge nucleotides are probably forming tandem sheared *trans*-Hoogsteen–sugar edge pairs (Figure 8B) (Leontis and

Westhof, 1998; Leontis *et al.*, 2002). This would explain the 7-deaza A and N6-MeA interferences observed for A599 and A662 and the strong 2'-deoxy effect at A661 (Boudvillain and Pyle, 1998). Importantly, the minor groove of this hypothetical structure is free to interact with other intronic substructures (Figure 8B), as predicted by the NAIM data.

This noncanonical pairing of the D3 internal loop is completely consistent with available phylogenetic data, which show that A598A599:A661A662 is frequently replaced with UA:AA (group II A1, bacterial class D), AA:GA (group II B1); GA:GA (group II B2) or AG:AA (bacterial class C) (Toor *et al.*, 2001). The proposed sheared *trans*-Hoogsteen-sugar edge pairs readily tolerate the above co-variations, as A·G and A·A are isosteric and interchangeable in many observed structural motifs (Leontis and Westhof, 1998), and the tandem $5'RA^{3'}/3'AR^{5'}$ can interchange with $5'UA^{3'}/3'AR^{5'}$, as observed for the J4/5 region in group I introns (Cate *et al.*, 1996a; Leontis and Westhof, 1998).

Conclusions

Having used NAIM, NAIS, mutational and phylogenetic analysis to examine the function of intron D3, we conclude that this important catalytic domain exerts its function in a variety of ways. The unusual D3 pentaloop is engaged in a docking interaction (μ - μ') that joins D3 with other intronic substructures that play an important role in catalysis. The highly conserved internal loop at the base of D3 is closed by network of specific noncanonical base pairings that create an accessible minor groove platform for critical interactions with the rest of the intron. Finally, we show that the linker regions in D3 play an important, and as yet undefined, role in the function of D3. Taken together, the data underscore the functional importance of D3 and they lay a foundation for understanding its role in the catalytic mechanism of group II introns.

Materials and methods

DNA templates, RNA transcription and synthesis

The exD123 plasmid (pJDI3'-673) (Jarrell *et al.*, 1988) was kindly provided by Dr PS Perlman. All exD123 plasmids with mutations in D3 (i.e. pQL 38 (G615UAAU619:UUCG), pQL 55 (A617:U), pQL 57 (A662:U), pQL 58 (A617:G), pQL 104 (G615:A), pQL 108 (G615:U), pQL 114 (A618:G), pQL 119 (A600:G), pEL 01 (Δ 619), pEL 40 (A617:C, A618:C), pEL 41 (U619:A), pEL 42 (U619:C) were created by site-directed mutagenesis (Kunkel *et al.*, 1991) of the wild-type exD123 construct. Design of the pT7-D56 plasmid is described elsewhere (Chin and Pyle, 1995).

Prior to transcription, wild-type and mutant exD123 plasmids were linearized with *Bam*HI, and pT7-D56 construct was digested with *Eco*RV. All RNA transcriptions were carried out essentially as described previously (Pyle and Green, 1994; Chin and Pyle, 1995). For NAIM and NAIS experiments, RNAs with randomly incorporated phosphorothioate analogs were prepared by adding NTP α S to the transcription mixture (Gish and Eckstein, 1988; Ortoleva-Donnelly *et al.*, 1998). All NTP α S analogs were purchased from Glen Research (Sterling, VA).

For NAIS experiments, D56 RNAs (wild-type and containing single-atom substitutions dG844 or I845) were prepared by enzymatic ligation of two chemically synthesized oligonucleotides as described previously (Fedorova *et al.*, 2005).

Branching reactions for NAIM

Reactions were carried out as described previously (Boudvillain and Pyle, 1998; Fedorova *et al.*, 2003). All exD123 transcripts containing

NTP α S analogs were 3'-end labeled (Huang and Szostak, 1996), gel-purified and re-suspended in 10 mM MOPS, pH 6.0, 1 mM EDTA. The 3'-end-labeled exD123 RNA (10 nM final concentration) was then incubated with unlabeled wild type D56 RNA transcript (final concentration 1.5 μ M) at 42°C under splicing conditions (40 mM MOPS, pH 6.8, 100 mM MgCl₂, 0.5 M (NH₄)₂SO₄), until ~20% of the branched product was formed. The unreacted exD123 and reaction products (branching and/or hydrolysis) were separated on a 5% polyacrylamide denaturing gel, eluted from the gel, ethanol precipitated and subjected to iodine cleavage as described below.

Branching kinetics

Wild type and mutant exD123 RNAs (2 μ M final concentration) and 5'-end labeled D56 RNA (10 nM) were separately denatured in MOPS pH 7.5 (40 mM final concentration), for 1 min at 95°C and then cooled to 42°C. The exD123 and D56 RNAs were then combined with the simultaneous addition of MgCl₂ and NH₄Cl (final concentrations 100 mM and 0.5 M, respectively), and reactions were carried out at 42°C as described (Chin and Pyle, 1995; Boudvillain and Pyle, 1998). Aliquots taken at specific time points were quenched with denaturing dye solution (Fedorova *et al.*, 2002), chilled on ice and analyzed on a stacking 5–8% (top-bottom) denaturing polyacrylamide gel. Bands corresponding to precursor D56 and branched product were quantified using a PhosphorImager (Molecular Dynamics), and results were analyzed as described previously (Boudvillain and Pyle, 1998).

NAIS experiments

Branching reactions for NAIS experiments were performed as described (Boudvillain and Pyle, 1998; Boudvillain *et al.*, 2000). For the 'forward' NAIS experiments, wild-type and mutant exD123 RNAs (1.5 μ M final concentration) were reacted with 5'-end labeled NTP α S-containing D56 (10 nM final concentration) under the conditions described above for NAIM experiments. The reaction time was adjusted so that only 20% of the branched product was formed in each case. The unreacted D56 and the branched product were separated on a 5% denaturing polyacrylamide gel, eluted from the gel and ethanol precipitated. Sequencing by iodine cleavage was carried out as described below. For the 'reverse' NAIS experiments, 3'-end labeled wild-type exD123 transcribed with various NTP α S analogs (10 nM final concentration) was reacted with unlabeled D56 RNA (wild type and containing dG844 or I845 substitutions, 1.5 μ M final concentration). The reactions were carried out as described above for the 'forward' NAIS assay. The unreacted exD123 and branched product were separated on a 5% denaturing gel and treated as above.

Iodine cleavage of the phosphorothioate-containing RNAs

After the NAIM or NAIS selection procedure (see above), precursor and product RNAs were subjected to sequencing by iodine in order to visualize the position of incorporated nucleotide analogs (Gish and Eckstein, 1988; Boudvillain and Pyle, 1998; Fedorova *et al.*, 2003). RNAs were resuspended in 3 μ l of water, 0.5 μ l of 10 mM ethanolic iodine solution was added and the mixture was then incubated at 37°C for 3 min. The samples were ethanol precipitated and analyzed on 8% high-resolution sequencing gels prepared using Long Ranger gel matrix (Cambrex, USA). Bands were quantified by PhosphorImager (Molecular Dynamics) and data analysis was carried out as described (Boudvillain and Pyle, 1998; Fedorova *et al.*, 2005). NAIM and NAIS experiments were repeated and quantitated at least three times. The experimental error in determining the interference magnitude did not exceed 20%.

Acknowledgements

We thank Professor Eric Westhof for helpful discussions and Dr Christina Waldsich and Dr Michael Bruno for critically evaluating the manuscript. We also thank Dr Philip S Perlman for the gift of plasmid pJDI3'-673. OF is a Research Specialist and AMP is an Investigator of the Howard Hughes Medical Institute.

References

- Abramovitz DL, Friedman RA, Pyle AM (1996) Catalytic role of 2'-hydroxyl groups within a group II intron active site. *Science* **271**: 1410–1413
- Abramovitz DL, Pyle AM (1997) Remarkable morphological variability of a common RNA folding motif: the GNRA tetraloop-receptor interaction. *J Mol Biol* **266**: 493–506
- Boudvillain M, DeLencastre A, Pyle AM (2000) A new RNA tertiary interaction that links active-site domains of a group II intron and anchors them at the site of catalysis. *Nature* **406**: 315–318
- Boudvillain M, Pyle AM (1998) Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J* **17**: 7091–7104
- Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA (1996a) Crystal structure of a group I ribozyme domain reveals principles of higher order RNA folding. *Science* **273**: 1678–1685
- Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Szewczak AA, Kundrot CE, Cech TR, Doudna JA (1996b) RNA tertiary structure mediation by adenosine platforms. *Science* **273**: 1696–1699
- Chin K, Pyle AM (1995) Branch-point attack in group II introns is a highly reversible transesterification, providing a possible proof-reading mechanism for 5'-splice site selection. *RNA* **1**: 391–406
- Fedorova O, Boudvillain M, Kawaoka J, Pyle AM (2005) Nucleotide analog interference mapping and suppression: specific applications in studies of RNA tertiary structure, dynamic helicase mechanism and RNA-protein interactions. In *Handbook of RNA Biochemistry*, Hartmann RK, Bindereif A, Schoen A, Westhof E (eds), Vol. 1, pp 259–293. Weinheim: Wiley-VCH
- Fedorova O, Mitros T, Pyle AM (2003) Domains 2 and 3 interact to form critical elements of the group II intron active site. *J Mol Biol* **330**: 197–209
- Fedorova O, Su LJ, Pyle AM (2002) Group II introns: highly specific endonucleases with modular structures and diverse catalytic functions. *Methods* **28**: 323–335
- Gish G, Eckstein F (1988) DNA and RNA sequence determination based on phosphorothioate chemistry. *Science* **240**: 1520–1522
- Griffin EA, Qin ZF, Michels WA, Pyle AM (1995) Group II intron ribozymes that cleave DNA and RNA linkages with similar efficiency, and lack contacts with substrate 2'-hydroxyl groups. *Chem Biol* **2**: 761–770
- Huang Z, Szostak JW (1996) A simple method for 3'-labeling of RNA. *Nucleic Acids Res* **24**: 4360–4361
- Jarrell KA, Dietrich RC, Perlman PS (1988) Group II intron domain 5 facilitates a *trans*-splicing reaction. *Mol Cell Biol* **8**: 2361–2366
- Jestin JL, Deme E, Jacquier A (1997) Identification of structural elements critical for inter-domain interactions in a group II self-splicing intron. *EMBO J* **16**: 2945–2954
- Koch JL, Boulanger SC, Dib-Hajj SD, Hebbbar SK, Perlman PS (1992) Group II introns deleted for multiple substructures retain self-splicing activity. *Mol Cell Biol* **12**: 1950–1958
- Konforti BB, Liu Q, Pyle AM (1998) A map of the binding site for catalytic domain 5 in the core of a group II intron ribozyme. *EMBO J* **17**: 7195–7217
- Kunkel TA, Bebenek K, McClary J (1991) Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol* **204**: 125–139
- Legault P, Mogridge J, Kay LE, Greenblatt J (1998) NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell* **93**: 289–299
- Lehmann K, Schmidt U (2003) Group II introns: structure and catalytic versatility of large natural ribozymes. *Crit Rev Biochem Mol Biol* **38**: 249–303
- Leontis NB, Stombaugh J, Westhof E (2002) The non-Watson-Crick base-pairs and their associated isostericity matrices. *Nucleic Acids Res* **30**: 3497–3531
- Leontis NB, Westhof E (1998) Conserved geometrical base-pairing patterns in RNA. *Q Rev Biophys* **31**: 399–455
- Michel F, Ferat JL (1995) Structure and activities of group II introns. *Annu Rev Biochem* **64**: 435–461
- Michel F, Umesono K, Ozeki H (1989) Comparative and functional anatomy of group II catalytic introns—a review. *Gene* **82**: 5–30
- Michels WJ, Pyle AM (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
- Mikheeva S, Murray HL, Zhou H, Turczyk BM, Jarrell KA (2000) Deletion of a conserved dinucleotide inhibits the second step of group II intron splicing. *RNA* **6**: 1509–1515
- Ortoleva-Donnelly L, Sweczak AA, Gutell RR, Strobel SA (1998) The chemical basis of adenosine conservation throughout the *Tetrahymena* ribozyme. *RNA* **4**: 498–519
- Podar M, Dib-Hajj S, Perlman PS (1995) A UV-induced Mg²⁺-dependent cross-link traps an active form of domain 3 of a self-splicing group II intron. *RNA* **1**: 828–840
- Pyle AM, Green JB (1994) Building a kinetic framework for group II intron ribozyme activity: quantitation of interdomain binding and reaction rate. *Biochemistry* **33**: 2716–2725
- Pyle AM, Lambowitz AM (2006) Group II introns: ribozymes that splice RNA and invade DNA. In *The RNA World*, Gesteland RF, Cech TR, Atkins JF (eds), 3rd edn, pp 469–506. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press (in press)
- Strobel SA (1999) A chemogenetic approach to RNA function/structure analysis. *Curr Opin Struct Biol* **9**: 346–352
- Strobel SA (2002) Biochemical identification of A-minor motifs within RNA tertiary structure by interference analysis. *Biochem Soc Trans* **9**: 1126–1131
- Swisher J, Duarte C, Su L, Pyle AM (2001) Visualizing the solvent-inaccessible core of a group II intron ribozyme. *EMBO J* **20**: 2051–2061
- Toor N, Hausner G, Zimmerly S (2001) Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* **7**: 1142–1152
- Xiang Q, Qin PZ, Michels WJ, Freeland K, Pyle AM (1998) The sequence-specificity of a group II intron ribozyme: multiple mechanisms for promoting unusually high discrimination against mismatched targets. *Biochemistry* **37**: 3839–3849