The EMBO Journal Vol.17 No.23 pp.7091–7104, 1998

Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis

Marc Boudvillain and Anna Marie Pyle[1]

The Howard Hughes Medical Institute and Department of Biochemistry and Molecular Biophysics, 701 W. 168th Street, Room 616, Hammer Health Sciences Center, Columbia University, New York, NY 10032, USA

[1]Corresponding author
e-mail: amp11@columbia.edu

Group II introns are self-splicing RNA molecules that are of considerable interest as ribozymes, mobile genetic elements and examples of folded RNA. Although these introns are among the most common ribozymes, little is known about the chemical and structural determinants for their reactivity. By using nucleotide analog interference mapping (NAIM), it has been possible to identify the nucleotide functional groups (Rp phosphoryls, 2′-hydroxyls, guanosine exocyclic amines, adenosine N7 and N6) that are most important for composing the catalytic core of the intron. The majority of interference effects occur in clusters located within the two catalytically essential Domains 1 and 5 (D1 and D5). Collectively, the NAIM results indicate that key tetraloop–receptor interactions display a specific chemical signature, that the ε–ε′ interaction includes an elaborate array of additional features and that one of the most important core structures is an uncharacterized three-way junction in D1. By combining NAIM with site-directed mutagenesis, a new tertiary interaction, κ–κ′, was identified between this region and the most catalytically important section of D5, adjacent to the AGC triad in stem 1. Together with the known ζ–ζ′ interaction, κ–κ′ anchors D5 firmly into the D1 scaffold, thereby presenting chemically essential D5 functionalities for participation in catalysis.

Keywords: catalysis/interference/ribozyme/RNA/structure

Introduction

Splicing of group II introns is essential for the expression of organellar genes in organisms such as plants, yeast, protists and fungi (Michel and Ferat, 1995). Some group II introns are autocatalytic, promoting their own excision from primary transcripts *in vitro* (Peebles *et al*., 1986; Schmelzer and Schweyen, 1986; Michel *et al*., 1989) and therefore containing all of the molecular constituents required for chemical catalysis. As expected for a self-splicing RNA (Cech, 1993), group II introns can also function as ribozymes that are capable of binding and cleaving other RNA molecules (Mörl and Schmelzer, 1990; Michels and Pyle, 1995). An intrinsically high sequence specificity (Xiang *et al*., 1998) and cleavage-site fidelity (L.H.Su and A.M.Pyle, in preparation) make group II introns attractive candidates for biomedical applications. Group II introns are also mobile genetic elements that can invade RNA (Augustin *et al*., 1990; Mörl and Schmelzer, 1990) and double-stranded DNA (Zimmerly *et al*., 1995). All of these biological functions are intimately dependent on the ability of group II intron RNA to properly fold into a catalytically competent three-dimensional structure (or set of structures). RNA tertiary structure is stabilized by networks of interactions involving a variety of RNA functional groups and, in many cases, metal ions bound to specific sites. The complex three-dimensional shape imposed by RNA tertiary interactions can lead to the formation of a ribozyme active site that insures proper positioning of catalytic core components (Pyle and Green, 1995; Westhof *et al*., 1996). The identification and characterization of such interactions is therefore a fundamental step towards elucidating the strategies used by RNA enzymes to regulate and catalyze chemical reactions.

Group II intron secondary structure is divided into six domains, of which Domains 1 and 5 (D1 and D5) are known to be critical for all reactions catalyzed by the intron (Pyle, 1996). Other regions are necessary for particular steps of splicing (Jacquier and Michel, 1990; Chanfreau and Jacquier, 1993), enhancing catalytic activity (Podar *et al*., 1995a; Xiang *et al*., 1998), adding core stabilization (Costa *et al*., 1997) or promoting partial intron rearrangement required between the two steps of splicing (Chanfreau and Jacquier, 1996). Group II introns have a relatively conserved secondary structure; however, their low level of base conservation and a paucity of phylogenetic covariations have permitted only a partial identification of tertiary interactions (Figure 1). These include Watson–Crick pairing between the intron and 5′-exon binding sequences (IBS–EBS), pairings within D1 (α–α′, ε–ε′, β–β′) and several inter-domain tertiary contacts (γ–γ′, ζ–ζ′, η–η′, θ–θ′) (Michel and Ferat, 1995; Qin and Pyle, 1998). This lack of genetically evident interactions suggests that additional tertiary contacts will involve components of the sugar–phosphate backbone and unusual base–base interactions.

At the present time, the mechanism for group II intron folding is not understood and a three-dimensional model is not available. In addition, only a few functional groups (all in D5) have been demonstrated to participate in chemical catalysis although additional moieties (potentially in other domains) are likely to participate in transition-state stabilization. Although it is possible to extend our understanding of critical functionalities by surveying the reactivity of group II intron domains with single-atom mutations or deletions (Abramovitz *et al*., 1996; Konforti *et al*., 1998a), this strategy is not always appropriate and it is generally time-consuming since each chimeric molecule must be synthesized and tested individually. For

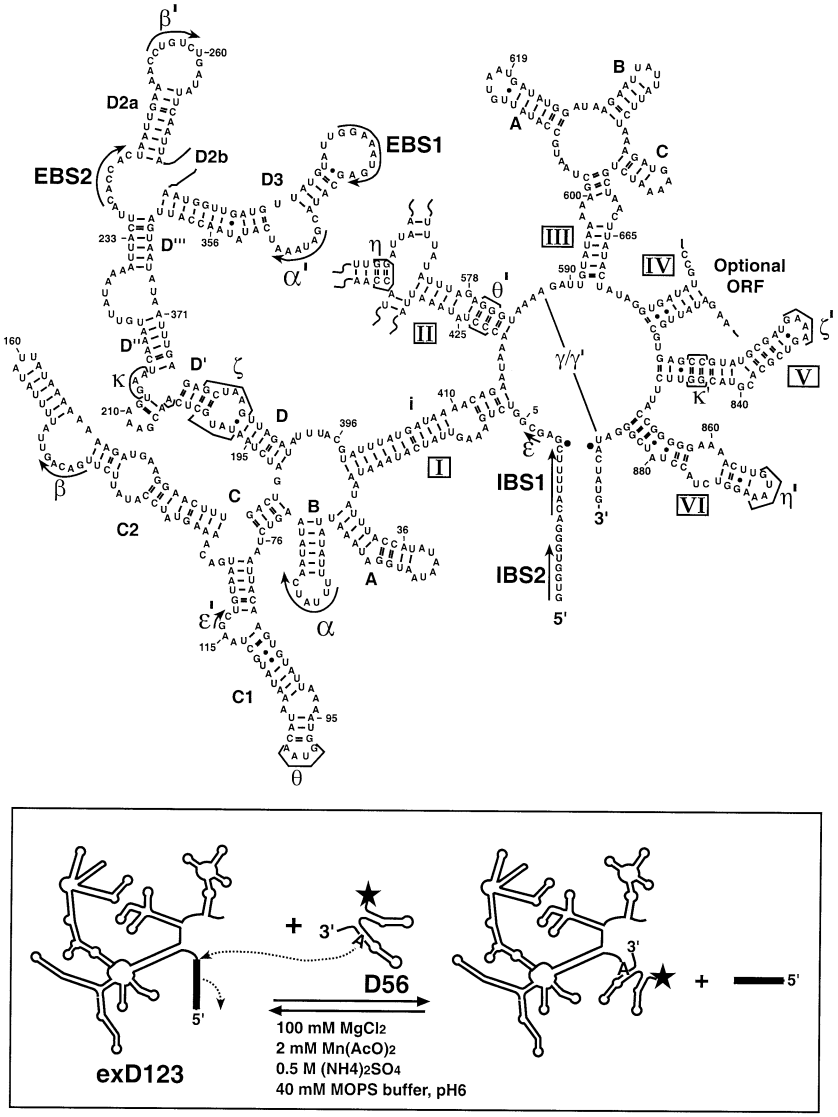

Fig. 1. A two-piece branching system derived from the ai5γ group II intron. The secondary structure of the intron is shown, together with known tertiary interactions (represented by Greek letters) and the IBS and EBS intron/exon recognition sequences. The six intron domains, which radiate from a central wheel, are numbered within rectangles. Helix D2b of D1, as well as sections of D2, are not shown, as no interferences were observed in these areas. Intron/exon boundaries are identified by black dots. Inset: the *trans*-branching assay involves reaction of 5′-^{32}P (★) end-labeled D56 with 5′-exon/Domain 1–3 molecules (exD123 or 17D123, depending on the assay) under the conditions shown. For mapping the 3′-end of D123, an alternative labeling approach was used (see Materials and methods).

that reason, it was of interest to exploit the faster and broader approach known as nucleotide analog interference mapping (NAIM) (Conrad *et al*., 1995; Strobel and Shetty, 1997). NAIM uses T7 RNA polymerase to randomly incorporate NTPαS analogs into RNA transcripts. The NTPαS can contain natural or modified nucleosides. After a subset of the transcripts is selected from an RNA pool (based on their ability to catalyze a certain reaction, participate in complex formation, and so on), iodine cleavage of the phosphorothioate linkages (Gish and Eckstein, 1988) reveals the positions of interference. This strategy has been used to identify a set of non-bridging Rp phosphate oxygens contributing to group II intron catalysis (Chanfreau and Jacquier, 1994; Jestin *et al*., 1997). The phosphorothioate group can also be used as a chemical tag to probe the importance of functional groups on base or sugar moieties (Conrad *et al*., 1995; Strobel and Shetty, 1997; Ortoleva-Donnelly *et al*., 1998). In more elaborate strategies, NAIM can be combined with site-directed mutagenesis and modification to identify specific pairs of functional groups that engage in a particular tertiary interaction (Strobel and Shetty, 1997; Strobel *et al*., 1998). Known as nucleotide analog interference suppression (NAIS), this methodology applies a type of genetic analysis to identify a broad spectrum of tertiary interactions.

In this study, we have focused on the identification of backbone and base functional groups that are critical for the first step of group II intron self-splicing. This reaction is highly distinctive and markedly differentiates group II introns from other ribozymes (Cech, 1993): the 2′-OH of a bulged adenosine located within Domain 6 (D6) attacks the phosphodiester bond at the 5′-splice junction, leading to the release of both the 5′-exon and a branched lariat intermediate, as observed in nuclear pre-mRNA splicing (Michel and Ferat, 1995). Because branching results in

covalent attachment of D6 to the first nucleotide of the intron, the reaction provides an ideal selection method for conducting interference studies such as NAIM. Here, a *trans*-branching construct derived from the ai5γ group II intron (Dib-Hajj *et al*., 1993; Chin and Pyle, 1995) is found to react under experimental conditions that enhance interference effects and facilitate elucidation of important functional groups (Figure 1). The experiments reveal clusters of essential groups located almost exclusively in D1 and D5, many of which appear to be involved in tertiary contacts. NAIM is then combined with site-directed mutagenesis to reveal a specific new tertiary interaction (κ–κ′) that, together with the ζ–ζ′ interaction, precisely anchors D5 within D1 (Figure 1). The results provide a further demonstration that group II intron cores are built from inter-dependent functional modules that assemble through a complex network of long-range interactions.

Results

Group II introns are readily divided into separate pieces that associate and react to recapitulate specific steps of splicing. These multi-component reactions provide for good kinetic and thermodynamic control, since the concentration and structure of the individual components can be easily varied. To catalyze branching *in-trans*, an RNA containing Domains 5–6 (D56) was reacted with an RNA consisting of the 5′-exon joined to Domains 1–3 (exD123; Figure 1) in a reaction that proceeds only through the first step of splicing (Chin and Pyle, 1995). Either component can be transcribed with NTPαS analogs to probe the importance of functional groups in each domain.

Calibrating the method

For optimization of trans-branching reaction conditions, 5′-end labeled D56 transcripts were randomly modified with either a NTPαS or dNTPαS and reacted with an excess of unlabeled exD123 under a variety of conditions. The *trans*-branching reaction was found to require high [Mg^{2+}] and it was almost undetectable in the absence of monovalent cations (data not shown). The lowest possible stimulatory concentration of monovalent salts was used in order to maximize sensitivity of the reaction. To limit the extent of incubation time, to control RNA degradation and to ensure that conditions were not overly permissive, final salt and pH conditions were chosen to be 100 mM $MgCl_2$, 500 mM $(NH_4)_2SO_4$, 2 mM $Mn(AcO)_2$ and 40 mM MOPS pH 6. The low pH is necessary for favoring branching over the competing hydrolysis reaction at the 5′-splice site (Chin and Pyle, 1995). The exD123 molecule is present at a non-saturating concentration (1.5 μM) only slightly above its K_m (0.5 μM) (Chin and Pyle, 1995) in order to ensure that the assay would be sensitive to effects on both binding and catalysis (k_{cat}/K_m effects). During this initial study, the assay was designed to reveal functional groups that are critical for any aspect of group II intron tertiary folding, inter-domain binding or catalysis.

Strong phosphorothioate effects preclude NAIM at positions 1, 4, 7, 117, 816, 817 and 839

Before investigating ribose and base functionalities that were essential for catalysis, it was necessary to locate intron positions where a strong phosphorothioate interference precluded identification of effects induced by nucleoside modifications. In particular, it was of interest to identify interferences that could not be relieved by thiophilic metal ions (2 mM Mn^{2+} or Cd^{2+}) provided in the reaction buffer. Phosphorothioate effects were clustered in two areas of D5 (Figure 2), containing three positions of strong interference that could not be relieved, even partially, by the presence of thiophilic metal ions (816, 817 and 839). A similar observation was made for interferences at positions 1, 4, 7 and 117 of D1 (see below). These effects are likely to reflect steric hindrance by the larger S atom, electronic effects of the P–S bond, important H-bonds to phosphoryl oxygens (Lecuyer *et al.*, 1996; Loverix *et al.*, 1998), or the loss of critical outer- or even inner-sphere metal–phosphate interactions (Brautigam and Steiz, 1998). The effects precluded probing of other nucleotide modifications at these seven positions. Nevertheless, NAIM analysis remained possible for positions of milder phosphorothioate interference, provided that the effect was taken into account during quantitation.

Branching in trans is more stringent than cis splicing

Analysis of the phosphorothioate effects in D5 (Figure 2) reveals positions not identified in previous work (Chanfreau and Jacquier, 1994) where a *cis*-splicing system was used under non-permissive salt conditions (5 mM $MgCl_2$, 2 mM spermidine and 40 mM Tris–Cl pH 7.5). Three novel positions (at nucleotides 831, 832 and 835) were identified recently in a study of D5 binding (Jestin *et al*., 1997). Highly reproducible effects observed at positions 819, 820 and 827 were not identified in either of the previous studies. Therefore, despite the higher ionic strength used in the assays, the *trans*-branching system appears to be particularly sensitive for identifying functionally important atoms.

5′-splice site hydrolysis and branching recruit the same set of D5 2′-hydroxyl group

A 5′-splice site hydrolysis reaction was previously used to identify a set of specific 2′-hydroxyl groups that are essential for D5 catalysis (Abramovitz *et al*., 1996). Although that reaction is different from the branching reaction studied herein, D5 is of central importance to both pathways for 5′-exon cleavage, both of which lead to splicing *in vitro* and *in vivo* (Podar *et al*., 1998b). It was therefore of interest to compare previous deoxynucleotide effects with those obtained through NAIM to reveal whether D5 uses similar functional groups in both reactions. NAIM studies using dNTPαS analogs reveal a set of 2′-hydroxyl groups that play an important role in the branching reaction (Figure 2). These include all the hydroxyl groups previously shown to make strong contributions to D5 binding or catalysis during the hydrolysis reaction (Abramovitz *et al*., 1996). Moreover, the deoxynucleotide interferences observed at C825, G829 and A830 each correspond to hydroxyl moieties previously observed to have smaller effects on hydrolysis (D.Abramovitz and A.M.Pyle, unpublished results). Only deoxynucleotide interference observed at two additional positions, G815 and C848, may reflect the requirement of the branching reaction for supplementary D5 contacts. The majority of

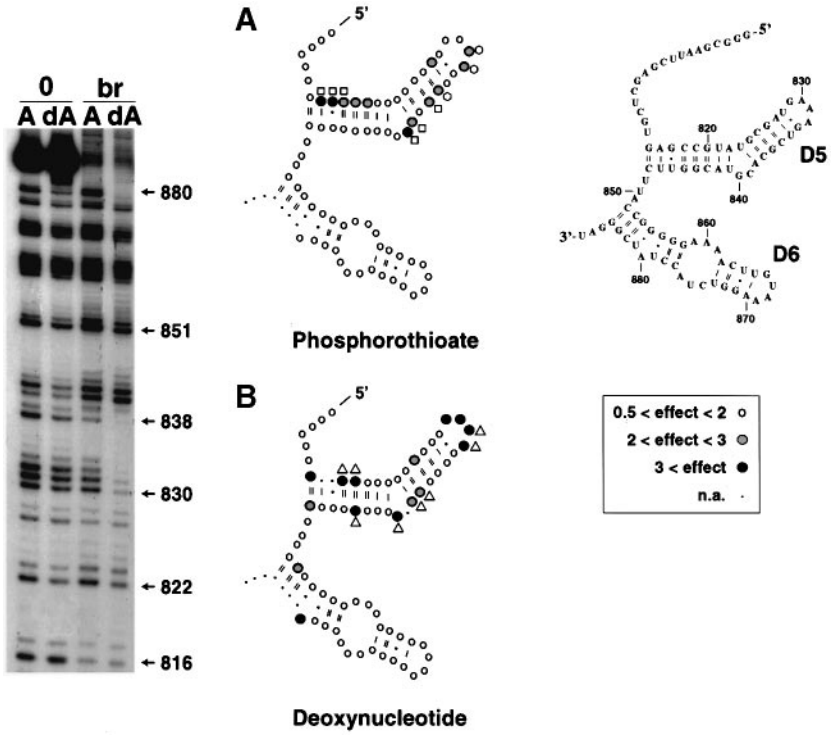

Fig. 2. Interference effects within D56: backbone and sugar modifications. A representative polyacrylamide gel shows phosphorothioate (A lane) and deoxynucleotide (dA lanes) interference effects after iodine cleavage of unreacted (0) and branched (br) transcripts modified with either ATPαs or dATPαS. Phosphoroathioate (**A**) and deoxynucleotide (**B**) interferences are summarized on secondary structure diagrams, with the sequence of D56 shown at the right. Note that strong phosphorothioate effects at A816, G817 and C839 preclude probing of nucleotide modifications at these residues [indicated by the n.a. symbol (●) in (**B**)]. Circles are used to indicate the effects and intensities observed in the present study. Also indicated are positions of phosphorothioate effects identified previously through *cis*-splicing (□) and binding assays (hexagons) and deoxynucleotide effects observed previously by direct binding and catalysis assays (△).

strong 2′-hydroxyl effects are observed at the same positions for both reactions, which indicates that the active site is very similar in both cases and that the new NAIM assay is sensitive to known functional group effects.

Identification of important base functionalities in D5 and D6

Three of the most important base functionalities known to play a role in tertiary interactions are the exocyclic amine of G, the N7 group of A and the N6 group of A. NAIM readily probes the role of these groups by transcriptional incorporation of the analogs inosine-αS (ITPαS), 7-deaza-adenosine-αS (C7ATPαS) and 2-amino-purine-αS ([2-AP]TPαS), respectively. Transcription of the latter results in the selective incorporation of the modified nucleotide at A and not G positions (Ortoleva-Donnelly *et al*., 1998). The first NAIM studies of base functional groups in group II introns focused on the atoms of D5 and D6, which were studied by reacting exD123 with D56 molecules containing NTPαS analogs (Figure 3).

The D5 internal loop and surrounding nucleotides contain critical base functionalities. Previous work on the base atoms of D5 has focused on substituents of G817, which is the invariant nucleotide that is part of the chemically critical AGC triad at the base of the D5 stem (Figure 1). Through synthetic inosine incorporation, previous work has shown that the exocyclic amine of G817 is very important for D5 docking, contributing 2.6 kcal/mol of interaction energy (Konforti *et al*., 1998). To rapidly identify other important guanosine amino groups in this region, D56 was transcribed with ITPαS and examined by NAIM. Strong interferences were observed at positions 824 and 840, and a moderate effect was observed at position 845 (Figure 3). The prominent inosine effect previously observed at position 817 could not be seen in this case because of a strong phosphorothioate effect at that position (Figure 2). Position 824 corresponds to a guanine involved in a highly conserved G–C pair in which the cytosine 2′-OH group is known to contribute to chemical catalysis (Abramovitz *et al*., 1996). Position 840 corresponds to the guanine of the D5 internal loop that separates the two stems of D5 and whose critical role in group II intron function has been inferred recently (Schmidt *et al*., 1996). The internal loop region is known to be a site of many important 2′-OH and phosphoryl atoms as well (Chanfreau and Jacquier, 1994; Abramovitz *et al*., 1996). In fact, the 2′-OH of G840 has already been shown to participate in catalysis (Abramovitz *et al*., 1996), implicating a role for both base and ribose substituents in the function of G840.

Important base functionalities in the first stem and tetraloop of D5. A moderate inosine interference highlights the role of the G845 exocyclic amine, which has an unpaired proton that falls on the binding face of D5 (Konforti *et al*., 1998a) and is likely to be involved in tertiary contacts with D1 (see below). Also in the first stem of D5, the previously identified G817 base functionalities lie immediately opposite position G845, on the other strand (Peebles *et al*., 1995; Konforti *et al*., 1998a). In the tetraloop of D5, the N2 group of the guanosine does

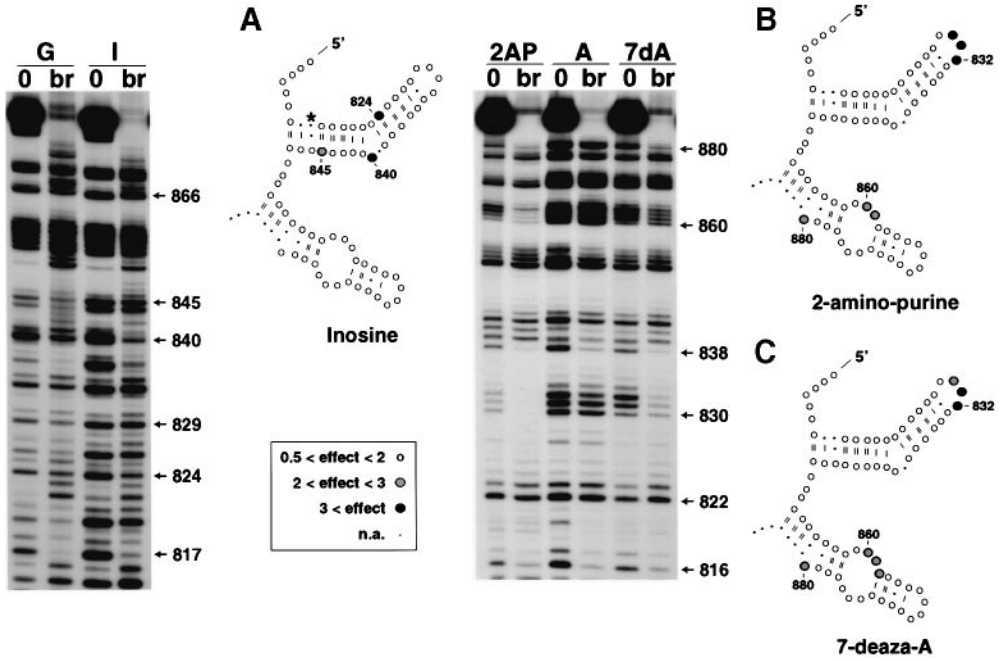

Fig. 3. Interference effects within D56: base modifications. High-resolution sequencing gels and secondary-structure maps summarize the effects of inosine [I lanes, (**A**)], 7-deaza adenosine [7dA lanes, (**B**)] and 2-aminopurine [2AP lanes, (**C**)] substitution in D56. Transcripts were modified specifically with the NTPαs analog identified above each lane. The asterisk indicates the position of G817, which has been shown to result in a strong effect when synthetically substituted with inosine (Konforti *et al.*, 1998a). Gels are labeled as shown in Figure 2.

not appear to play a role, despite its presumed importance in formation of a sheared G–A pair important for intra-tetraloop stabilization (Jucker *et al*., 1996). In fact, inosine effects are not observed for any of the GNRA tetraloops in the group II intron (Figures 3 and 5). In contrast, there are moderate and strong effects of 2-AP and 7-deazaA at positions A830, A831 and A832 (Figure 3). D5 is known to dock into a receptor motif in D1 (ζ–ζ′) (Costa and Michel, 1995) and therefore many interferences in D5 may be explained by analogy to a crystallographically-characterized tetraloop–receptor interaction from a group I intron (Cate *et al*., 1996a). The 2-AP effects observed at positions A830 and A831 are consistent with the loss of H-bonds between N6 groups of these residues and the A198 N1 and 2′-OH group of the first paired guanosine in the receptor motif (G200). However, docking studies on the tetraloop–receptor crystal structure reveal that 2-AP interferences at positions 830–832 could also be attributed to steric interference by the 2-amino group. Although one cannot rule out steric and electronic effects of the C7 substituent at these positions, the 7-deazaA effects in the tetraloop are consistent with contacts between N7 of A831 and the 2′-OH of G829, while N7 of A832 is expected to participate in the sheared G–A pair of the tetraloop.

Few D6 functional groups contribute to its role in branching. Despite its important role in bearing the branch-site, only a few interferences are observed within D6, which is consistent with its low level of phylogenetic conservation. In particular, there are no ribose or base modification effects within the GUAA tetraloop that caps D6 (Figures 2 and 3). This suggests that the tetraloop is not involved in any essential tetraloop–receptor interaction during branching. Such an observation supports the hypothesis that the D6 tetraloop interacts with its receptor (located within D2) either after a conformational rearrangement that follows the first step of splicing (Chanfreau and Jacquier, 1996) or to favor 5′-splice site hydrolysis over branching (Costa *et al*., 1997). Despite the absence of effects in the tetraloop, 7-deaza-A and 2-AP interferences are observed in other regions of D6, particularly at positions 860, 861 and 862 (Figure 3) which are located in the D6 internal loop. Interestingly, suppressor mutations of branching-deficient D6 variants have been found in the same loop (V.T.Chu, P.S.Perlman and A.M.Pyle, in preparation), suggesting that this poorly conserved region plays a role in branching. Moderate effects of base substituents are seen at the branch point itself, which has been the subject of other functional group analyses (Gaur *et al*., 1997; Liu *et al*., 1997). Given this collection of relatively minor effects in D6, it is likely that determinants such as correct docking of D5 into the active site, the length of the D5–D6 junction (Boulanger *et al*., 1996), or simply the D6 shape are critical for correct D6 positioning. The results are consistent with the suggestion that D5 is the active site anchor, dragging D6 and the branch point into place (Dib-Hajj *et al*., 1993; Chin and Pyle, 1995).

Identification of important nucleoside functionalities in D1–3

Development of customized NAIM approaches for mapping a large RNA. In order to identify critical functionalities located within Domains 1–3, exD123 RNA was transcribed with a battery of NTPαS analogs, 3′-end labeled and then reacted with an excess of D56 RNA. Although this method was useful for analyzing functionalities in D3, D2 and the 3′ portion of D1 (Figure 4A), the large size of D123 (683 nts) prevented the resulting branched molecules from being completely sequenced to the 5′-end. To examine nucleotides near the 5′-end of the intron (such as residues

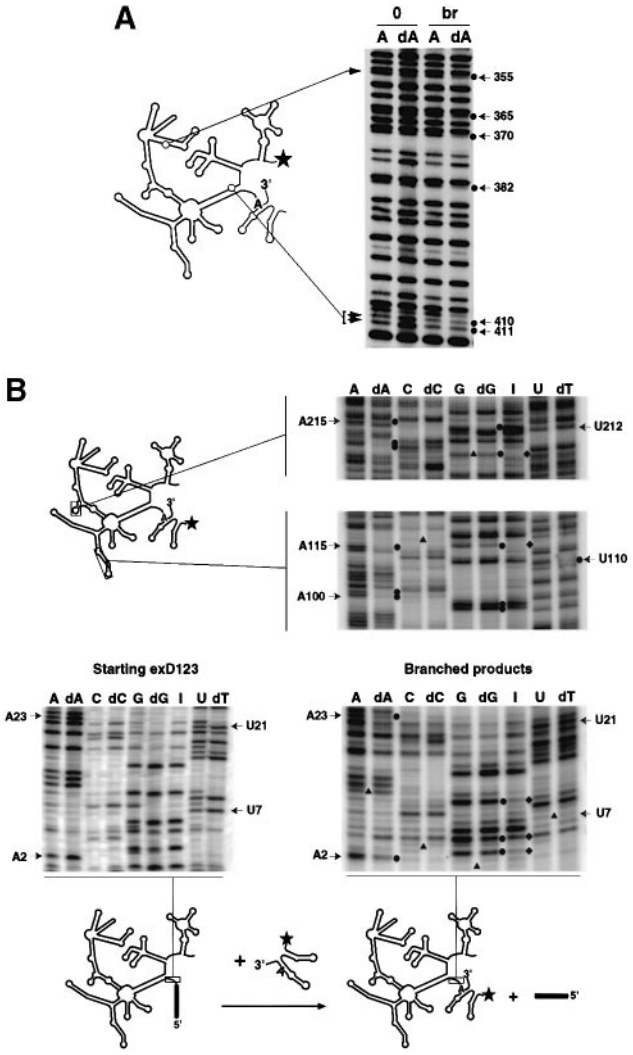

Fig. 4. Interference mapping of D123. (**A**) Autoradiograph of interferences at the 3′-end of D123, assayed by reaction of D56 with 3′-end-labeled exD123 modified with either ATPαs (lanes A) or dATPαS (lanes dA). Lanes 0 and br refer to the unreacted and branched products, respectively. (**B**) Interferences assayed by tagging D123 with [5′-^{32}P]D56. Results are shown for specific regions indicated on the adjacent secondary structure diagrams. A representative comparative assay between starting material and branched products is shown (bottom) In all cases, the positions of phosphorothioate (▲), deoxynucleotide (●), or inosine (◆) interference are indicated by symbols to the right of the appropriate lane.

in D1), simple 5′-end labeling of exD123 could not be used because the 5′-exon is removed during the selection step. Instead, a complementary approach was developed in which 5′-^{32}P end-labeled D56 transcripts were reacted with 17D123 RNA containing modified nucleotides of interest (see Figure 1, inset and 4B, bottom). Shortening of the 5′-exon from 293 to 17 nucleotides had only minor effects on *trans*-branching in our hands (Kevin Chin and A.M.Pyle, unpublished results), in contrast with what has been reported previously for *cis*-splicing (Jacquier and Michel, 1987; Nolte *et al.*, 1998). After branching, this procedure effectively tags D123 at its 5′-end by attachment to a [^{32}P]D56 molecule. The method is particularly powerful because only the selected molecules acquire a 5′-^{32}P label, making detection of interferences easier. Notably, the [^{32}P]D56 tagging method was highly sensitive to 3′-end heterogeneity of D56 resulting from run-on transcription by T7 RNA polymerase. This effect was found to obscure the sequencing pattern of branched products (data not shown). To overcome this problem, [^{32}P]D56 RNA was prepared with a homogeneous 3′-end by treating 5′-^{32}P end-labeled D56-Exon2 transcripts with a DNAzyme (Santoro and Joyce, 1997), which cleaves between nucleotides A886 and U887 under single-turnover conditions. This treatment leads to a readable sequencing pattern (see branched-products gel, Figure 4B). By using both approaches to tag D123 molecules (either by ^{32}P 3′-end labeling of precursor exD123, as shown in Figure 4A, or [^{32}P]D56-tagging of 17D123, as shown in Figure 4B, bottom), it was possible to visualize interference effects for all regions within D1–3.

Critical functional groups are found primarily in D1, clustered into distinctive groups. A striking feature of the interferences in D1–3 is that almost all ribose, phosphate and base effects correspond to nucleotides in D1 (Figure 5). Only three positions within D3 exhibit moderate deoxynucleotide effects (A619, A662 and C663), while no modifications within D2 affect branching. Similarly, no interferences were found within inter-domain linkers such as the section of J2/3 that contains highly conserved nucleotides (A589 and G588) that crosslink with D5 (Podar *et al.*, 1998a) and participate in the γ–γ pairing important for the second step of splicing (Jacquier and Michel, 1990).

Most of the interferences in D1 are clustered within several distinct regions (Figure 5). Not surprisingly, a significant number of these correspond to nucleotides that are conserved phylogenetically, partially or completely (Michel *et al.*, 1989). The following clusters are notable:

(i) The first six nucleotides of the intron. This area corresponds to the single-stranded region between the 5′-splice site and helix I of D1 (Figures 1 and 5). As expected, a particularly strong phosphorothioate effect is observed at the first nucleotide of the intron, reflecting the inhibition of branching caused by Rp phosphorothioate incorporation at the 5′-splice site (Padgett *et al.*, 1994; Podar *et al.*, 1995b). Weak deoxynucleotide and 7-deaza interferences were found at nucleotide A2, which has not been implicated previously in function. In contrast, almost total inosine and phosphorothioate blocks are observed at positions 3 and 4, respectively, highlighting the importance of the two nucleotides involved in the ε–ε′ pairing (Figures 4B and 5). The loss of a H-bonding amino group (inosine effect) can be rationalized as a strong destabilizing factor for Watson–Crick pairing (especially if only two base pairs are involved in ε–ε′); however I–C pairs still form (Mirau and Kearns, 1984; Green *et al.*, 1991) and may not necessarily disrupt adjacent regions of secondary structure. This suggests an additional role for the G3N2 group. The phosphorothioate effect at cytosine 4 is particularly intriguing since the phosphate moiety has not been shown to contribute to the ε–ε′ interaction. In addition, the universally conserved guanosine (G5) that is located directly adjacent to the ε dinucleotide shows both moderate deoxy and strong inosine effects.

(ii) Helix I of the first D1 stem. Strong interferences extend into the first helix of D1, where a strong phos-

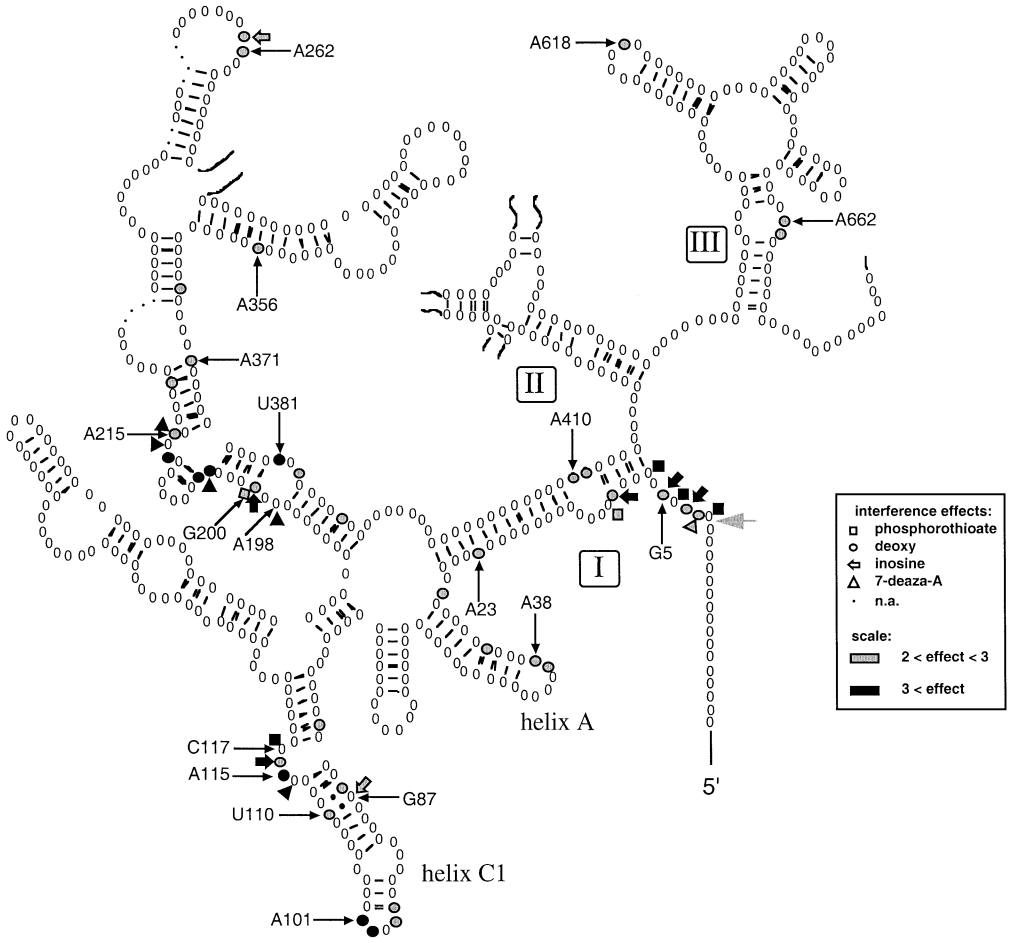

Fig. 5. Summary of interferences within D123: schematic representation of positions within the D1–3 portion of the ai5γ secondary structure, showing sites of interference by phosphorothioate, deoxynucleotide, inosine and 7-deaza adenosine modification (key is inset). The 5′-splice site (gray arrow) as well as a portion of the 5′-exon are also shown. Weak interferences with intensities between 0.5- and 2-fold greater than control bands are not shown. The interference levels shown herein represent a conservative estimate for the experimental window appropriate for NAIM experiments on large molecules (Ortoleva-Donnely *et al*., 1998). Note that effects induced by 2-aminopurine-αS were not determined for D123. The 7-deaza-A interferences were analyzed only between residues A2 and A262.

phorothioate effect at U7 is accompanied by a moderate phosphorothioate effect at A11 as well as deoxynucleotide and strong inosine effects at G10. Phosphorothioate effects on the opposite strand, at nucleotides A410, A411 and C412, have also been reported for branching under low-salt conditions (Jestin *et al*., 1997). These effects were not detected under the experimental conditions used in this study; however, deoxynucleotide interferences at A410 and A411 were detected (Figures 4A and 5). The collective results suggest that internal loop of helix I(i) probably has an important function, consistent with the multiplicity of interference effects and the presence of several conserved nucleotides.

(iii) Helix A. Although this region is not highly conserved, it is organized structurally in a similar fashion in many group II introns. Therefore this helical stem capped by an apical loop might constitute an element whose global organization is important for intron function, as inferred by modest but reproducible deoxynucleotide interferences at C34, A38 and U39.

(iv) Helix C. This subdomain shows numerous, intense interference effects (Figure 5), most of which cluster in or around the helix central bulge containing the GC dinucleotide involved in the ε–ε′ interaction. The results indicate that ε–ε′ is much more than a simple pairing, as its constituent nucleotides show strong interferences at many functional groups, including a strong phosphorothioate effect at C117, deoxynucleotide and inosine effects at G116. At the adjacent tandem G–U wobble motif, deoxynucleotide (U86 and U110) and inosine interferences (G87) suggest an important role for this motif, which has been proposed to be a major-groove metal binding domain in other systems (Cate and Doudna, 1996). Moderate and strong deoxynucleotide effects are observed also in the area of the GUAA tetraloop that caps helix C (positions 97, 98, 100 and 101).

(v) The tetraloop–receptor region of the D1D stem. This section of D1, which has been proposed to dock with the tetraloop at the D5 terminus (the ζ–ζ′ interaction; Costa and Michel, 1995), contains strong interferences both within the internal loop and in the adjacent set of inverted tandem G–C pairs. By comparing analogous tetraloop and receptor positions with those from a crystallograph-determined structure of this motif (Cate *et al*., 1996a), the following conclusions can be drawn. Strong 7-deaza interference at A198 is consistent with a hydrogen bond to U381 (forming an unusual A–U pair within the receptor). Deoxynucleotide effects observed at positions

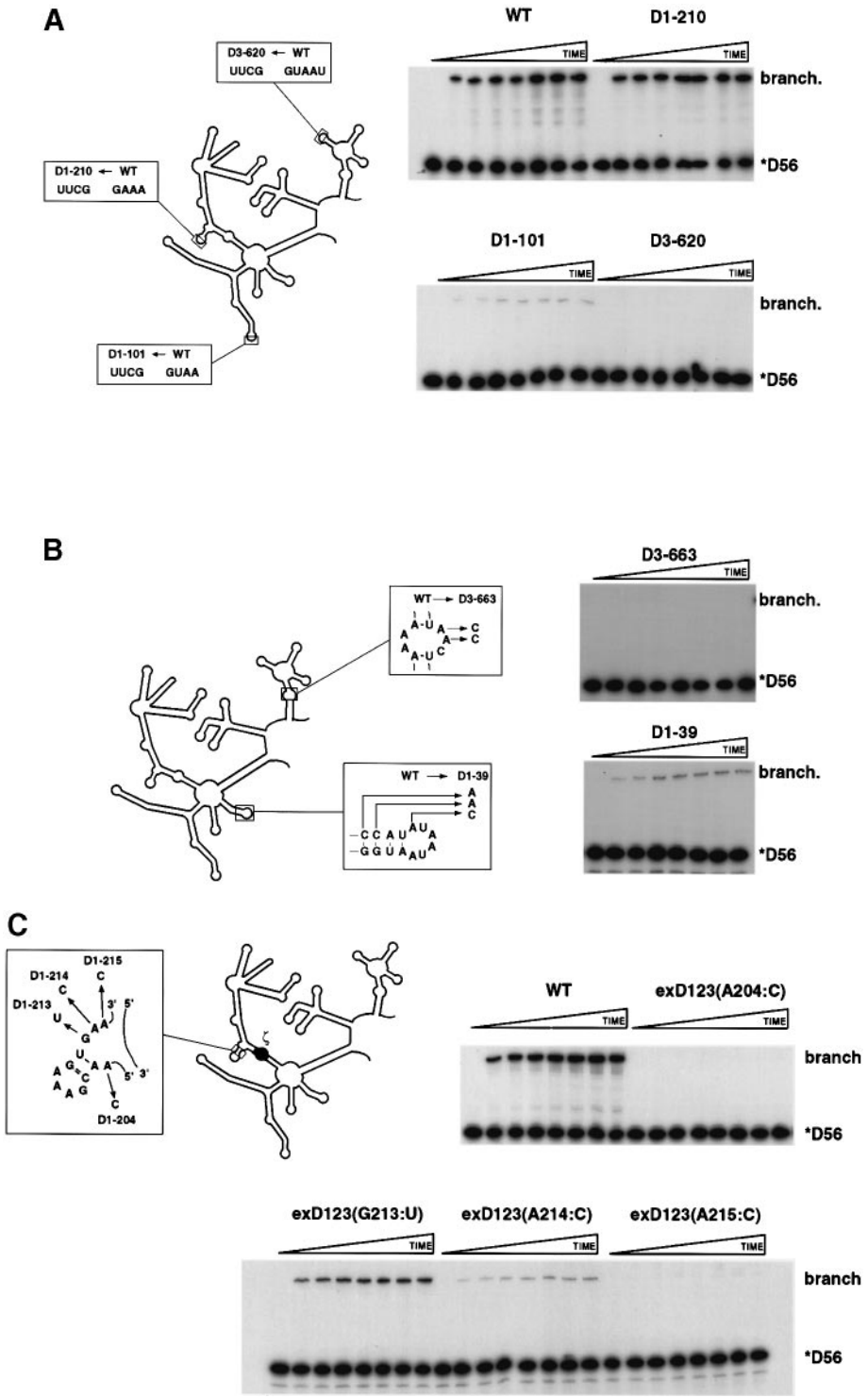

Fig. 6. Effect of intronic mutations on branching kinetics: autoradiographs of stacked 4:20% (top:bottom) denaturing polyacrylamide gels are shown, together with the sequences of mutants subjected to kinetic analysis (left). For each gel, the lanes correspond to timepoints for aliquots removed at 0, 20, 40, 60, 90, 120, 150 and 180 min. (**A**) Kinetic effects induced by replacement of GNRA loops by UUCG loops at three places in the intron. (**B**) Kinetic effects of mutations in the D1A helix and D3 central loop, which share homology with tetraloop receptors. (**C**) Branching defects induced by single point mutations of the D1 three-way junction adjacent to the ζ receptor. Under the conditions described in Materials and methods and assuming a 38% endpoint in each case k_{obs} was $(175 \pm 4)\times10^{-4}$ min^{-1} for the wild-type construct; $(4.6 \pm 0.3)\times10^{-4}$, $(1.8 \pm 0.2)\times10^{-4}$ and $(270 \pm 4)\times10^{-4}$ min^{-1} for D1–39, D1–101 and D1–210, respectively; $(24 \pm 2)\times10^{-4}$ and $(3.7 \pm 0.3)\times10^{-4}$ min^{-1} for exD123(G213:U) and exD123(A214:C), respectively. Values for k_{obs} could not be measured for the other mutants, because they were inactive.

G200 and U381 are consistent with hydrogen bonds to the amine of A831 and the N3 and 2′-OH of A831, respectively. The inosine effect observed at G200 is consistent with weakened G–C pairing in the receptor and a hydrogen bond to N3 of A832. The moderate deoxynucleotide and phosphorothioate effects observed at positions A383 and G200 are not explained by the P456 structure, as the 2′-OH group and Rp-phosphoryl oxygen

at these positions are located on its exterior and are likely to contact other elements in a larger RNA (as in the intact group I intron or the group II intron studied here).

(vi) The three-way junction in the D1D stem. Adjacent to the tetraloop receptor of the ζ–ζ′ interaction, there is an unusual substructure that is commonly observed in secondary structural maps of group II introns. Although it has not been previously implicated in group II intron function, this three-way junction (often represented by a large open loop in other introns) contains many of the strongest interference effects observed in D123. Deoxynucleotide effects are observed at A204, A205, G213 and A215, as well as 7-deaza interferences at A204, A214 and A215. This pattern of uniformly strong interferences implicates the region in interaction with D56 and/or participation in catalysis (see below).

(vii) Interference effects at isolated positions within D1. Moderate NAIM effects are observed at several isolated spots in D1. Deoxynucleotide interference is observed at A220, A371 and G388, which are sites that are near the tetraloop receptor and three-way junction. In the D1D2a stem thought to contain the β–β′ interaction (Figure 1), deoxynucleotide and inosine effects are observed at G261 and a deoxynucleotide effect was found at A262. This suggests that the β—β′ pairing may promote the formation of other tertiary contacts involving nearby nucleotides.

(viii) A collection of interferences in D3. It is remarkable that there were no strong interferences observed in D3, despite the stimulatory role it has shown in catalysis (Griffin *et al*., 1995; Podar *et al*., 1995a; Xiang *et al*., 1998). Moderate interference effects were observed at position A618, which is contained in a form of GNRA pentaloop (Abramovitz and Pyle, 1997; Massire *et al*., 1998) and at A662 and C663, which are located in a region with the appearance (Costa and Michel, 1997) and DMS reactivity (Konforti *et al*., 1998b) of a tetraloop receptor. Note that NAIM effects were not observed in D2, consistent with its low level of phylogenetic conservation and its role in second step efficiency (Chanfreau and Jacquier, 1996).

Mutations confirm the importance of specific D123 subdomains

Many of the interferences observed in this study belong to sub-structures identical to or resembling motifs commonly involved in RNA tertiary interactions. To further assess the role of these sub-structures, a set of exD123 mutants was prepared and subjected to kinetic analysis. In the first set of mutations, GNRA loops were replaced by UUCG loops, which are not believed to be capable of participation in long-range interaction. The mutations introduced in region D1C1 (101) and D3 (620) had marked effects on the rate of branching, while alteration of the GAAA tetraloop in the D1D three-way junction (210) did not (Figure 6A). Significantly, strong effects on branching tend to correlate closely with the presence of deoxynucleotide interferences in a loop (compare Figure 6A with 5), suggesting that deoxynucleotide effects in a GNRA loop (particularly at the loop 3 and 4 positions; see also Abramovitz *et al*., 1996) are a good indication that the loop is involved in a long-range contact to a receptor. Other mutations were introduced into the internal loop of D3 (663) and helix A of D1 (39). The gross mutational changes made at these positions had radical effects on the kinetics of branching (Figure 6B), underscoring their importance in structure or mechanism.

In this study, many interferences were observed at positions or structures known to be phylogenetically conserved. However, one of the strongest regions of interference lies in an area that has not been recognized previously as a conserved feature: the three-way junction adjacent to ζ receptor (nts 204–215, Figures 1 and 5). To understand better the role of this substructure, single-base mutations were made throughout and analyzed kinetically. Remarkably, every single mutation, exD123(A204:C), exD123(G213:U), exD123(A214:C) and exD123(A215:C), is highly detrimental to the branching reaction (Figure 6C). This contrasts sharply with the lack of effects from the GAAA→UUCG mutation, introduced a few nucleotides away (nts 207–210; Figure 6A).

Identification of a critical long-range tertiary interaction between the D5 catalytic region and the D1D stem

Because two of the D1 bulge mutants, exD123(G213:U) and exD123(A214:C), exhibited residual branching activity, it was possible to exploit them in the search for long-range interactions with D5 using the technique of NAIS, (Strobel and Shetty, 1997; Strobel *et al*., 1998). This experiment rests on the fact that an important mutation or single-atom change in one molecule (such as exD123) will result in the loss of nucleotide analog interference at positions of tertiary interaction on a partner molecule (such as D56). In the mutant context, the energetic coupling between both partners has already been disrupted and no additional penalty is expected to result from a second modification. After reacting wild-type and mutant exD123 transcripts with [^{32}P]D56 molecules containing NTPαS analogs, branched products were isolated and sequenced, and interference effects were compared. No significant differences were observed between patterns obtained with wild-type exD123 and the exD123(G213:U) mutant (not shown). In contrast, deoxynucleotide and inosine interferences previously observed at C819 and G845, respectively, disappeared upon branching with the exD123 (A214:C) mutant (Figure 7). These interference suppressions are likely to reflect the loss of tertiary contact between the D1 bulge region and stem 1 of D5. Designated κ–κ′, this is a novel interaction, of a type and position not reported previously.

Discussion

Inherent advantages of the trans-branching experimental system

Previous studies on mutations and modifications of group II introns have utilized effectively *cis*-splicing assays for examination of effects. In contrast, the present study applied a *trans*-branching approach in order to isolate the first step of splicing and sensitize the reaction to perturbations of tertiary structure. This approach is so effective because the stabilization energy resulting from essential tertiary interactions must counterbalance the entropic penalty inherent to the assembly of a multi-component system. Therefore, disruption of interactions

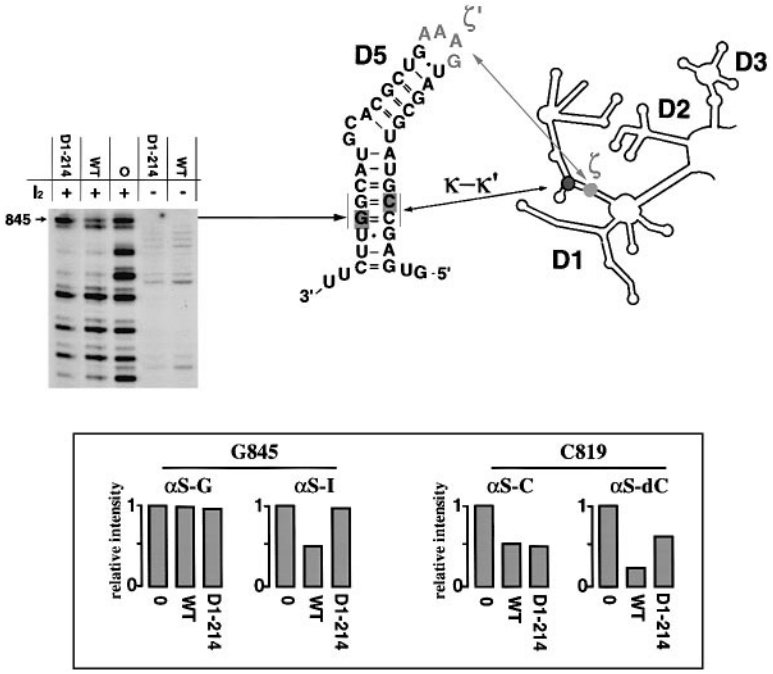

Fig. 7. NAIS within D5: mutants exD123(G213:U) and exD123(A214:C) were tested for potential changes in D5 interference pattern. The wild-type and mutant exD123 transcripts were reacted with 5′-^{32}P-labeled D56 that was transcribed with an array of NTPαS analogs. The autoradiograph corresponds to iodine-cleaved (+) or uncleaved (–) transcripts modified with inosine-αS and shows that the exD123(A214:C) mutant (D1–214 lane) induces suppression of the inosine interference observed at position 845 for the wild-type exD123 (WT lanes). Lane 0 refers to unreacted D56. A similar observation was made for the deoxynucleotide effect at position C819 (data not shown), although it was partially obscured by concomitant phosphorothioate interference (see inset). Inset: the normalized intensities of bands corresponding to positions 845 (left) and C819 (right) are compared for unreacted exD123 (0 bars), wild-type exD123 (WT bars) and the exD123(A214:C) mutant (D1–214 bars). The intensity observed for starting material was arbitrarily fixed to 1. For branched products (right and middle bars in each case), the size of the bars is related to 1/(interference effect), i.e. the smaller the bar, the stronger the interference.

during trans-branching by D56 (which interacts with exD123 solely through formation of tertiary contacts) is likely to have more dramatic effects than in unimolecular systems. Furthermore, the NAIM approach inherently limits the number of false positives resulting from steric clashes that are induced by bulky chemical modifications. For example, the lesion resulting from DEPC modification (Chanfreau and Jacquier, 1996) is much larger than the C–H group substituted for N7 of A in NAIM studies. For these reasons, the non-permissive character of this assay has permitted the identification of more significant functional groups, with higher precision, than reported previously.

The D56 *trans*-branching system was particularly useful for focusing on tertiary interactions between D1 and D5. This is because D6 does not make many tertiary contacts with the intron core and D5 serves as the anchor for branching. Therefore, any assay that depends on strong D5 docking prior to reaction will be highly sensitive to defects in tertiary interactions between D5 and the rest of the intron.

A chemical signature for GNRA tetraloop–receptor interactions

Tetraloop–receptor interactions are important building blocks for RNA tertiary structure. However, it is not always straightforward to predict whether a GNRA loop is engaged in contact to a receptor and, even if it is, which receptor is its partner. This is because GNRA loops and their receptors are often poorly conserved substructures. Furthermore, GNRA loops in certain positions have evolved to play other roles, such as protein binding (Legault *et al*., 1998). In that case, the second base of a GNRA loop (which stacks with receptor bases during tertiary contact formation) has been observed to stack with phenylalanine in RNA–protein interactions. Because of the multifunctional nature of GNRA loops, it would be desirable to have a rapid screen for detecting those that function by interacting specifically with receptor motifs.

In the present study and in other published work (Abramovitz *et al*., 1996), deoxynucleotide interferences are consistently observed in GNRA tetraloops that interact with receptors (namely GAAA of D5 and GUAA of the C helix; Figures 2 and 5). The patterns of interferences are consistent with crystallographic studies showing that 2′-OH groups of the loop 3 and 4 nucleotides are involved in important hydrogen bonds to cognate receptors (Pley *et al*., 1994; Cate *et al*., 1996a), contributing ~2 kcal/mol per 2′-OH group (Abramovitz *et al*., 1996). Branching defects induced by mutation parallel the presence of deoxynucleotide interferences at specific GNRA loops (Figure 6A). Furthermore, deoxynucleotide effects were not observed for the GUAA tetraloop of D6, which is consistent with the absence of an interaction with its putative D2-receptor during branching (Chanfreau and Jacquier, 1996; Costa *et al*., 1997). We suggest therefore that the pattern of NAIM deoxynucleotide interferences within GNRA loops is a chemical signature for their participation in a receptor interaction.

It is nonetheless important to note that the number and intensity of GNRA deoxynucleotide effects may vary depending on sequence and context (Figures 2 and 5). In

some cases, these effects may be accompanied by 7-deaza and phosphorothioate interferences (see Results). However, the absence of deoxynucleotide effects might simply indicate that the tetraloop is involved in a dispensable tertiary interaction, as recently shown in a group I intron (Ortoleva-Donnelly *et al*., 1998). By comparing patterns of interferences, one may be able to deduce a chemical signature for interacting receptor sequences. For example, a distinctive pattern was observed for the ζ receptor in D1. However, interferences were not observed for an alternative form of receptor in D2 (θ′, Figure 1). Determinants such as the tetraloop–receptor interaction strength, sequence variation, context, or relative contribution to folding and mechanism may account for these differences.

The variety of 2′-OH effects implicates a variety of tertiary interaction motifs in group II introns

There is growing sentiment that long-range base-pairing interactions and tetraloop–receptor interactions constitute the major tertiary structural features that stabilize large folded RNAs. However, strong deoxynucleotide effects are evident in many other types of motifs (Pyle *et al*., 1992; Strobel and Cech, 1993; Hardt *et al*., 1996; Ortoleva-Donnelly *et al*., 1998). Because they lack abundant base–base interactions, group II introns may rely heavily on tertiary interactions involving backbone functionalities such as the 2′-OH. This dependence may be characteristic of RNA systems that favor non-classical interactions to control RNA spatial organization and reactivity.

In the case of the *Tetrahymena* group I intron, a collection of 2′-OH groups on the P1 helix (which contains the 5′-exon/intron boundary) contribute to the docking of this substructure with single-stranded nucleotides in J8/7 of the catalytic core (Pyle *et al*., 1992; Strobel and Cech, 1993). By analogy, deoxynucleotide interferences at positions involved in the ε–ε′ pairing (G3 and G116), may reflect the involvement of these 2′-OH groups with single stranded nucleotides near the 5′-splice site. Alternatively, the ε–ε′ hydroxyl groups may interact with D5, thereby comprising part of the intron active site. This is supported by DMS footprinting experiments showing strong protections in the same region upon D5 binding to D1 (Konforti *et al.*, 1998b).

ε–ε′: a more elaborate, extended interaction

Deletion and mutational experiments indicate that the ε–ε′ interaction is critical for the catalytic activity of group II introns (Suchy and Schmelzer, 1991; Podar *et al*., 1995b; Costa *et al*., 1997). However, the high level of ε–ε′ conservation suggests that Watson–Crick pairing may not be the only important feature of this interaction. Indeed, base mutation experiments indicate that the two wild-type base pairs correspond to the pairing combination with the highest catalytic efficiency (Jacquier and Michel, 1990). Strong inosine interferences at G3 and G116 further underscore a requirement for wild-type base identity (Figure 5).

The ε–ε′ interaction is problematic because a two base-pair helix is not expected to be sufficiently stable (Freier *et al*., 1986), suggesting that other interactions afford supplementary stabilization energy. As suggested above, such a contribution might be provided by interactions between 2′-OH groups of the ε′ nucleotides and single-stranded residues near the ε nucleotides. Alternatively, NAIM results suggest that additional stabilization could be provided by adjacent base pairings (Figure 5). The highly conserved G5 and A115 nucleotides are located next to those comprising the ε–ε′ interaction. Phylogenetic conservation is particularly apparent for G5, which is invariant, while A115 is sometimes replaced by a guanine. This suggests that an adjacent non-canonical G–A or G–G (Limmer, 1997) base pair may reinforce the ε–ε′ interaction. Note that among 67 group II intron sequences belonging to both IIA and IIB subgroups (appendix in Michel *et al*., 1989), the occurrence of the G–A combination at the positions corresponding to G5 and A115 (91%) is comparable to the level of conservation of the two base pairs constituting the ε–ε′ interaction (98.5 and 62.7%, respectively). Although interference results argue against the formation of a typical sheared G–A pair (the strong inosine effect at G5 is not accompanied by a 7-deaza effect at A115), other published arrangements might accommodate either a G–A or G–G pair (Battiste *et al*., 1996) without extensive variation of the backbone geometry or functional group pattern. It is noteworthy that G–A base pairing often occurs at the end of RNA helices where they stabilize structural elements such as RNA loops (Gautheret *et al*., 1994; Gutell *et al*., 1994; Heus *et al*., 1997) and serve as anchor points for tertiary contacts (Pley *et al*., 1994). Further NAIM experiments are being designed to differentiate the G5–A115 pair from alternative models, such as a triple interaction between G5 and a potential AA platform (Cate *et al*., 1996b) that might be formed by A114 and A115.

A strong phosphorothioate block is observed at nucleotides C4 and C117, indicating that contacts involving the two non-bridging oxygens may also contribute to additional stabilization of the ε–ε′ interaction. The absence of thiophilic ion rescue at these positions suggests that the Rp phosphoryl oxygens are not involved in direct Mg^{2+}-mediated interactions (Pecoraro *et al*., 1984) although disruption of a metal ion binding site by phosphorothioate substitution cannot be excluded (Brautigam and Steitz, 1998). The intensity of the effects suggests that Rp oxygens at C4 and C117 might be part of or close to the catalytic center.

The significance of D2 and D3

The most critical functionalities identified fall within D1 and D5. However, weak deoxynucleotide effects (A618, A662 and C663) were also observed in two regions of D3. The results are consistent with DMS footprints (Konforti *et al.*, 1998b) and base modification interferences (Chanfreau and Jacquier, 1996; Jestin *et al*., 1997) that were also found in the same regions of D3. The sequences, morphologies and DMS sensitivities of the loop-capping D3A and of the region encompassing A662 and C663 suggest that these substructures may act as a GNRA tetraloop and a tetraloop receptor, respectively, which may or may not interact with each other. Complementary experiments (DMS footprinting and direct binding measurements) have shown recently that D1–D5 interdomain binding is independent of the presence of D3 (Konforti *et al.*, 1998b), although D3 is known to enhance chemical catalysis during splicing and by group II intron

ribozymes (Griffin *et al*., 1995; Podar *et al*., 1995a; Xiang *et al*., 1998) and constitutes an independent folding domain (Podar *et al*., 1995a). These results suggest therefore that D3 is an important catalytic effector that, like D5, may dock into a region of the D1 scaffold where it serves to augment features of the active site.

That no interferences were found in D2 is in accord with previous results showing that this non-conserved substructure is dispensable for intron activity (Kwakman *et al*., 1990; Koch *et al*., 1992). However, deletion of D2 in the trans-branching reaction (replacing nucleotides 424–579 by a UUCG tetraloop) significantly slows reaction rate (>10×; data not shown). It has been proposed that the GUAA-tetraloop capping region D1C (~A101) interacts with a small tandem G–C pair receptor located at the base of the D2 stem (θ–θ′; Figure 1; Costa *et al*., 1997). That the D1C tetraloop is involved in a receptor interaction of some kind is consistent with the NAIM results. However, cognate interferences in D2 are not observed. Thus, the θ–θ′ interaction cannot be confirmed by the data herein.

κ–κ′: a critical new tertiary interaction between D1 and D5

The results presented herein show that an abundance of motifs, many of which have been previously identified, are likely to be involved in tertiary interactions or catalysis. These include nucleotides such as G5 and other residues that flank the ε–ε′ interaction (Figures 1 and 5); the terminus of the D1A stem; the internal loop at the base of D3, which has the appearance of a tetraloop receptor; the GA dinucleotide (A262) that flanks β′ (Figure 1 and 5); and, most strikingly, nucleotides in the three-way junction that lies at the center of the D1D stem (~nt A204). Interferences in the latter are among the strongest observed for the entire intron, suggesting that this region plays a critical role. Using a form of NAIS, it has been possible to implicate this three-way junction in a tertiary interaction with at least two nucleotides in the lower stem of D5 (Figure 7). The latter are immediately adjacent to the AGC triad of D5 containing the nucleotides most critical for catalysis by the intron (Peebles *et al*., 1995). Thus, the κ–κ′ interaction is likely to be essential for orienting D5 properly in the active site.

It is important to note that the type of NAIS experiments conducted here exploited mutants of D1 to search for interference suppressors in D5. In a large RNA such as the ai5γ intron, mutational approaches to NAIS are a good way to identify partners for suspected tertiary interactions. Now that the regions of contact have been identified in D1, it will be possible to conduct single-atom NAIS experiments (Strobel *et al*., 1998) to explore the molecular nature of the tertiary contact. Although it is difficult to predict which type of tertiary interaction is involved (and conceding that it may be a new form of contact), there is reason to suggest that the κ–κ′ interaction may represent a form of tetraloop receptor. Recent studies have shown that GNRA loops can interact with receptors even once expanded by additional nucleotides or split and interrupted by additional motifs (Abramovitz and Pyle, 1997; Massire *et al*., 1998). GNRA-like motifs can be imbedded in more complex structures without losing their characteristic ability to fold and bind receptor sequences. It is possible therefore that G213, A214, A215 and A204 form a GAAA tetraloop (Figure 6C). The region of contact is a pair of tandem G–C pairs (Figure 7), which are known to be a form of receptor for GNRA loops (Pley *et al*., 1994; Costa *et al*., 1997). While an exact assignment of the motif awaits further studies, a modified tetraloop–receptor interaction remains a likely option for the κ–κ′ contact. If this is the case, it means that ζ–ζ′ and κ–κ′ represent a 'yin-yang' type of association; in one case D5 provides the tetraloop and in the other case it provides the receptor. Both contacts occur on the binding face (the same side) of D5 (Abramovitz *et al*., 1996), in the minor groove, where they anchor D5 and present its opposite chemical face for participation in the catalysis of phosphodiester cleavage and ligation.

Materials and methods

Transcriptions

The following plasmids were constructed by site-directed mutagenesis (Kunkel *et al*., 1991) of plasmid pJDI3′-673 (Podar *et al*., 1998a): plasmid pKC02 encoding 17D123 RNA in which the 5′-exon is shortened to the last 17 nucleotides (comprising the two IBS sequences); plasmids pQL42, pQL40, pQL41 and pQL39 encoding single-point mutant RNAs exD123(A204:C), exD123(G213:U), exD123(A214:C) and exD123-(A215:C), respectively; plasmids pQL36, pQL37, pQL35, pQL38 and pQL43 encoding the D1–39, D1–101, D1–210, D3–620 and D3–662 mutants of exD123 (Figure 6A and B), respectively. All plasmids were linearized with *Bam*HI, except pT7-D56 (Chin and Pyle, 1995) which was linearized with *Eco*RV. Transcriptions of pJDI3′-673 and pT7-D56 (Chin and Pyle, 1995) plasmids to generate exD123 and D56 RNA, respectively, have been described (Pyle and Green, 1994; Chin and Pyle, 1995). Other plasmid transcriptions were carried out in the same manner. The pT7-D56 plasmid was linearized with *Dra*I to prepare D56-Exon2 transcripts containing D5 and D6 as well as the 79 first nucleotides of the 3′-exon. The D56-Exon2 transcripts were subsequently cleaved with a 36 nucleotide DNAzyme (Santoro and Joyce, 1997) producing D56 RNA with a homogeneous 3′-end. The 17D123, exD123 and D56 RNAs containing approximately one phosphorothioate per molecule were prepared as above with the addition of the desired NTPαS to the transcription mixture (Eckstein and Gish, 1989; Christian and Yarus, 1992). Dr Scott Strobel kindly provided αS-ITP. Other non-commercial NTPαS analogs were prepared from nucleoside precursors (Sigma) following published procedures (Arabshahi and Frey, 1994; Strobel and Shetty, 1997; Ortoleva-Donnelly *et al*., 1998). Nucleotide analogs were randomly incorporated into D56, 17D123, or exD123, at a level of about one modification per molecule using either wild-type or the Y639F mutant of T7 RNA polymerase (Sousa and Padilla, 1995) and adjusted NTPαS/NTP ratios as described (Strobel and Shetty, 1997; Ortoleva-Donnelly *et al*., 1998). The 2′-OH of uridine residues were probed by incorporating dTTPαS into transcripts. Therefore, interference effects at these positions could not be unambiguously attributed to deoxynucleotide modification. All RNA was purified by PAGE and stored at 20°C in 10 mM MOPS, 1 mM EDTA pH 6.

Branching reactions

In order to examine interferences within D5 and D6, 5′-^{32}P end-labeled, NαS-modified D56 transcripts (final concentration 1–10 nM) were mixed with exD123 (final concentration 1.5 μM) and heated at 95°C for 1 min. The mixture was cooled to 42°C and salts were added (100 mM $MgCl_2$, 2 mM $Mn(AcO)_2$, 0.5 M $(NH_4)_2SO_4$, 40 mM MOPS, pH 6) before further incubation at 42°C. Mapping of the 5′-portion of D123 was done by reacting, under the conditions described above, 5′-^{32}P end-labeled D56 transcripts (final concentration 1–10 nM) with 17D123 transcripts (final concentration 1.5 μM) that were modified with NTPαS analogs. To identify interferences in the 3′-portion of D123, NαS-modified exD123 transcripts were 3′-end labeled (Huang and Szostak, 1996) and reacted with an excess of D56 (final concentration 1.5 μM) under the above reaction conditions. The reaction time was adjusted in each case so that branched molecules represent ~20% of the total labeled material. Two volumes of formamide were then added to each sample, which was then loaded on a 4 or 8% polyacrylamide denaturing gel. Branched and,

when relevant, unreacted RNAs were recovered from the gel, eluted and ethanol precipitated prior to sequencing. Despite different experimental setups for the three assays described above, they are expected to be equally sensitive to modification based on the following evidence: salt, temperature, refolding procedure and reaction extent were the same for all three; each system revealed both strong and moderate interferences; the intensity of deoxynucleotide and inosine interferences at G261 is similar when using either assay for D123 mapping.

Branching kinetics with exD123 mutants

Every exD123 mutant (final concentration 2 μM) was reacted with 5′-end labeled D56 RNA (10 nM) under the branching conditions described above. At selected time points (Figure 6), aliquots were withdrawn and combined with 4 volumes of quench buffer (95% formamide, 5 mM EDTA pH 6, 0.1% of xylene cyanol and bromophenol blue dyes) before placing on ice. Samples were then subjected to electrophoresis on denaturing 4:20% (top:bottom) polyacrylamide gels which were quantitated on a Packard Instant Imager using methods described previously (Pyle and Green, 1994; Chin and Pyle, 1995).

NAIS experiments

Wild-type exD123 (final concentration 0.5 μM) and mutant exD123 (A214:C, final concentration 3 μM) were mixed with 5′-^{32}P end-labeled, NαS-modified, D56 transcripts (final concentration 10 nM). The mixture was preincubated and reacted as described above. After 3h of incubation at 42°C, reaction products were resolved by PAGE. Both unreacted and branched D56 transcripts were recovered from the gel and cleaved with iodine to reveal the sequence (see below). In order to insure that interference rescue did not result from experimental artifacts, variations in RNA concentration, reaction time and reaction extent were tested [for each assay two out of the three parameters were identical for both wild-type and the ExD123(A214:C) mutant], with similar results (not shown).

Sequencing of transcripts containing nucleotide analogs

Branched and unreacted (or precursor) RNA molecules were re-suspended in 10 μl of a 1:1 formamide: (10 mM MOPS, 1 mM EDTA pH 6) mixture, heated 1 min at 95°C and chilled on ice. Cleavage reactions were initiated by the addition of 1 μl of 1mM iodine dissolved in ethanol. After 1 min of reaction at room temperature, 240 μl of 0.3 M sodium acetate was added and the mixtures were ethanol precipitated in the presence of 1 μg of tRNA carrier. Iodine-cleaved products were resolved by PAGE. Gel bands were quantitated using a PhosphorImager (Molecular Dynamics). Interference quantitation was performed as described previously (Ortoleva-Donnelly *et al*., 1998). Only normalized interferences more than two were considered significant. Experimental error in interference magnitude did not exceed 20%, based on independent experiments.

Acknowledgements

We are grateful to Dr Scott Strobel for helpful suggestions and the gift of ITPαS. We also thank Qiaolian Liu for mutagenesis, Dr Rui Sousa for the clone of the Y639F RNA polymerase and Dr Philip S.Perlman for the gift of plasmid pJDI3′-673. M.B. is a Postdoctoral Research Associate and A.M.P. is an Assistant Investigator with the Howard Hughes Medical Institute, which we thank for financial support of this work.

References

Abramovitz,D.L. and Pyle,A.M. (1997) Remarkable morphological variability of a common RNA folding motif: the GNRA tetraloop-receptor interaction. *J. Mol. Biol*., **266**, 493–506.

Abramovitz,D.L., Friedman,R.A. and Pyle,A.M. (1996) Catalytic role of 2′-hydroxyl groups within a group II intron active site. *Science*, **271**, 1410–1413.

Arabshahi,A. and Frey,P.A. (1994) A simplified procedure for synthesizing nucleoside 1-thiotriphosphates. *Biochem. Biophys. Res. Commun*., **204**, 150–155.

Augustin,S., Müller,M.W. and Schweyen,R.J. (1990) Reverse self-splicing of Group II intron RNAs *in vitro*. *Nature*, **343**, 383–386.

Battiste,J.L., Mao,H., Rao,N.S., Tan,R., Mohandiram,D.R., Kay,L.E., Frankel,A. and Williamson,J.R. (1996) Alpha helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science*, **273**, 1547–1551.

Boulanger,S.C., Faix,P.H., Yang,H., Zhou,J., Franzen,J.S., Peebles,C.L. and Perlman,P.S. (1996) Length changes in the joining segment between domain 5 and 6 of a group II intron inhibit self-splicing and alter 3′ splice site selection. *Mol. Cell. Biol*., **16**, 5896–5904.

Brautigam,C.A. and Steitz,T.A. (1998) Structural principles for the inhibition of the 3′-5′ exonuclease activity of *E.coli* DNA polymerase I by phosphorothioates. *J. Mol. Biol*., **277**, 363–377.

Cate,J.H. and Doudna,J.A. (1996) Metal-binding sites in the major groove of a large ribozyme domain. *Structure*, **4**, 1221–1229.

Cate,J.H., Gooding,A.R., Podell,E., Zhou,K., Golden,B.L., Kundrot,C.E., Cech,T.R. and Doudna,J.A. (1996a) Crystal structure of a group I ribozyme domain reveals principles of higher order RNA folding. *Science*, **273**, 1678–1685.

Cate,J.H., Gooding,A.R., Podell,E., Zhou,K., Golden,B.L., Szewczak,A.A., Kundrot,C.E., Cech,T.R. and Doudna,J.A. (1996b) RNA tertiary structure mediation by adenosine platforms. *Science*, **273**, 1696–1699.

Cech,T.R. (1993) Structure and mechanism of the large catalytic RNAs: group I and group II introns and ribonuclease P. In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 239–270.

Chanfreau,G. and Jacquier,A. (1993) Interaction of intronic boundaries is required for the second splicing step efficiency of a group II intron. *EMBO J*., **12**, 5173–5180.

Chanfreau,G. and Jacquier,A. (1994) Catalytic site components common to both splicing steps of a group II intron. *Science*, **266**, 1383–1387.

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

Chin,K. and Pyle,A.M. (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.

Christian,E.C. and Yarus,M. (1992) Analysis of the role of phosphate oxygens on the Group I intron from *Tetrahymena*. *J. Mol. Biol*., **228**, 743–758.

Conrad,F., Hanne,A., Gaur,R.K. and Krupp,G. (1995) Enzymatic synthesis of 2′-modified nucleic acids: identification of important phosphate and ribose moieties in RNase P substrates. *Nucleic Acids Res*., **23**, 1845–1853.

Costa,M. and Michel,F. (1995) Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J*., **14**, 1276–1285.

Costa,M. and Michel,F. (1997) Rules for RNA recognition of GNRA tetraloops deduced by *in vitro* selection: comparison with *in vivo* evolution. *EMBO J*., **97**, 3289–3302.

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

Dib-Hajj,S.D., Boulanger,S.C., Hebbar,S.K., Peebles,C.L., Franzen,J.S. and Perlman,P.S. (1993) Domain 5 interacts with Domain 6 and influences the second transesterification reaction of group II intron self-splicing. *Nucleic Acids Res*., **21**, 1797–1804.

Eckstein,F. and Gish,G. (1989) Phosphorothioates in molecular biology. *Trends Biochem. Sci*., **14**, 97–100.

Freier,S.M., Kierzek,R., Jaeger,J.A., Sugimoto,N., Caruthers,M.A., Neilson,T. and Turner,D.H. (1986) Improved free energy parameters for predictions of RNA duplex stability. *Proc. Natl Acad. Sci. USA*, **83**, 9373–9377.

Gaur,R.K., Mclaughlin,L.W. and Green,M.R. (1997) Functional group substitutions of the branchpoint adenosine in a nuclear pre-mRNA and a group II intron. *RNA*, **3**, 861–869.

Gautheret,D., Konings,D. and Gutell,R.R. (1994) A major family of motifs involving G–A mismatches in ribosomal RNA. *J. Mol. Biol*., **242**, 1–8.

Gish,G. and Eckstein,F. (1988) DNA and RNA sequence determination based on phosphorothioate chemistry. *Science*, **240**, 1520–1522.

Green,R., Szostak,J.W., Benner,S., Rich,A. and Usman,N. (1991) Synthesis of RNA containing inosine: analysis of the sequence requirements for the 5′ splice site of the *Tetrahymena* group I intron. *Nucleic Acids Res*., **15**, 4161–4166.

Griffin,E.A., Qin,Z.-F., Michels,W.A. and Pyle,A.M. (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.

Gutell,R.R., Larsen,N. and Woese,C.R. (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol. Rev.*, **58**, 10–26.

Hardt,W.-D., Erdmann,V.A. and Hartmann,R.K. (1996) Rp-deoxy-phosphorothioate modification interference experiments identify 2′-OH groups in RNase P RNA that are crucial to tRNA binding. *RNA*, **2**, 1189–1198.

Heus,H.A., Wijmenga,S.S., Hoppe,H. and Hilbers,C.W. (1997) The detailed structure of tandem G.A mismatched base-pair motifs in RNA duplexes is context dependent. *J. Mol. Biol.*, **271**, 147–158.

Huang,Z. and Szostak,J.W. (1996) A simple method for 3′-labeling of RNA. *Nucleic Acids Res.*, **24**, 4360–4361.

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

Jacquier,A. and Michel,F. (1990) Base-pairing interactions involving the 5′- and 3′-terminal nucleotides of group II self-splicing introns. *J. Mol. Biol.*, **213**, 437–447.

Jestin,J.-L., Deme,E. and Jacquier,A. (1997) Identification of structural elements critical for inter-domain interactions in a group II self-splicing intron. *EMBO J.*, **16**, 2945–54.

Jucker,F.M., Heus,H.A., Yip,P.F., Moors,E.H.M. and Pardi,A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.*, **264**, 968–980.

Koch,J.L., Boulanger,S.C., Dib-Hajj,S.D., Hebbar,S.K. and Perlman,P.S. (1992) Group II Introns deleted for multiple substructures retain self-splicing activity. *Mol. Cell. Biol.*, **12**, 1950–1958.

Konforti,B.B., Abramovitz,D.L., Duarte,C.M., Karpeisky,A., Beigelman,L. and Pyle,A.M. (1998) Ribozyme catalysis from the major groove of group II intron domain 5. *Mol. Cell*, **1**, 1–20.

Kunkel,T.A., Bebenek,K. and McClary,J. (1991) Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.*, **204**, 125–139.

Kwakman,J.H., Konings,D.A., Hogeweg,P., Pel,H.J. and Grivell,L.A. (1990) Structural analysis of a group II intron by chemical modifications and minimal energy calculations. *J. Biomol. Struct. Dyn.*, **8**, 413–430.

Lecuyer,K.A., Behlen,L.S. and Uhlenbeck,O.C. (1996) Mutagenesis of a stacking contact in the MS2 coat protein RNA complex. *EMBO J.*, **15**, 6847–6853.

Legault,P., Mogridge,J., Kay,L.E. and Greenblatt,J. (1998) NMR structure of the bacteriophage λ N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell*, **93**, 289–299.

Limmer,S. (1997) Mismatch base pairs in RNA. *Prog. Nucleic Acid Res. Mol. Biol.*, **57**, 1–39.

Liu,Q., Chu,V.T. and Pyle,A.M. (1997) Mechanistic characterization of branch-site mutations and other alterations of Domain 6: Implications for group II intron reactivity. *J. Mol. Biol.*, **267**, 163–171.

Loverix,S., Winquist,A., Stromberg,R. and Steyaert,J. (1998) An engineered ribonuclease preferring phosphorothioate RNA. *Nature Struct. Biol.*, **5**, 365–368.

Massire,C., Jaeger,L. and Westhof,E. (1998) Derivation of the three-dimensional architecture of bacterial Ribonuclease P RNAs from comparative sequence analysis. *J. Mol. Biol.*, **279**, 773–793.

Michel,F. and Ferat,J.-L. (1995) Structure and activities of group II introns. *Annu. Rev. Biochem.*, **64**, 435–461.

Michel,F., Umesono,K. and Ozeki,H. (1989) Comparative and functional anatomy of group II catalytic introns-a review. *Gene*, **82**, 5–30.

Michels,W.J. and Pyle,A.M. (1995) Conversion of a group II intron into a new multiple-turnover ribozyme that selectively cleaves oligonucleotides: elucidation of reaction mechanism and structure/function relationships. *Biochemistry*, **34**, 2965–2977.

Mirau,P.A. and Kearns,D.R. (1984) Effect of environment, conformation, sequence and base substituents on the imino proton exchange rates in guanine and inosine-containing DNA, RNA and DNA–RNA duplexes. *J. Mol. Biol.*, **177**, 207–227.

Mörl,M. and Schmelzer,C. (1990) Integration of group II intron bI1 into a foreign RNA by reversal of the self-splicing reaction *in vitro*. *Cell*, **60**, 629–636.

Nolte,A., Chanfreau,G. and Jacquier,A. (1998) Influence of substrate structure on *in vitro* ribozyme activity of a group II intron. *RNA*, **4**, 694–708.

Ortoleva-Donnelly,L., Szewczak,A.A., Gutell,R.R. and Strobel,S.A. (1998) The chemical basis of adenosine conservation throughout the *Tetrahymena* ribozyme. *RNA*, **4**, 498–519.

Padgett,R.A., Podar,M., Boulanger,S.C. and Perlman,P.S. (1994) The stereochemical course of group II intron self-splicing. *Science*, **266**, 1685–1688.

Pecoraro,V.L., Hermes,J.D. and Cleland,W.W. (1984) Stability constants of Mg^{2+} and Cd^{2+} complexes of adenine nucleotides and thionucleotides and rate constants for formation and dissociation of MgATP and MgADP. *Biochemistry*, **23**, 5262–5271.

Peebles,C.L., Perlman,P.S., Mecklenburg,K.L., Petrillo,M.L., Tabor,J.H., Jarrell,K.A. and Cheng,H.-L. (1986) A self-splicing RNA excises an intron lariat. *Cell*, **44**, 213–223.

Peebles,C.L., Zhang,M., Perlman,P.S. and Franzen,J.F. (1995) Identification of a catalytically critical trinucleotide in Domain 5 of a group II intron. *Proc. Natl Acad. Sci. USA*, **92**, 4422–4426.

Pley,H.M., Flaherty,K.M. and McKay,D.B. (1994) Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature*, **372**, 111–113.

Podar,M., Dib-Hajj,S. and Perlman,P.S. (1995a) A UV-induced Mg^{2+}-dependent cross-link traps an active form of Domain 3 of a self-splicing group II intron. *RNA*, **1**, 828–840.

Podar,M., Perlman,P.S. and Padgett,R.A. (1995b) Stereochemical selectivity of group II intron splicing, reverse-splicing and hydrolysis reactions. *Mol. Cell. Biol.*, **15**, 4466–4478.

Podar,M., Zhou,J., Zhang,M., Franzen,J.S., Perlman,P.S. and Peebles,C.L. (1998a) Domain 5 binds near a highly-conserved dinucleotide in the joiner linking domains 2 and 3 of a group II intron. *RNA*, **4**, 151–166.

Podar,M., Chu,V.T., Pyle,A.M. and Perlman,P.S. (1998b) Group II intron splicing in-vivo by first step hydrolysis. *Nature*, **391**, 915–918.

Pyle,A.M. (1996) Catalytic reaction mechanisms and structural features of group II intron ribozymes. In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology* (*Catalytic RNA*). Springer Verlag, New York, Vol. 10, pp. 75–107.

Pyle,A.M. and Green,J.B. (1994) Building a kinetic framework for group II intron ribozyme activity: quantitation of interdomain binding and reaction rate. *Biochemistry*, **33**, 2716–2725.

Pyle,A.M. and Green,J.B. (1995) RNA Folding. *Curr. Opin. Struct. Biol.*, **5**, 303–310.

Pyle,A.M., Murphy,F.L. and Cech,T.R. (1992) Ribozyme substrate binding site in the catalytic core of the *Tetrahymena* ribozyme. *Nature*, **358**, 123–128.

Qin,P.Z. and Pyle,A.M. (1998) The architectural organization and mechanistic function of group II intron structural elements. *Curr. Opin. Struct. Biol.*, **8**, 301–308.

Santoro,S.W. and Joyce,G.F. (1997) A general purpose RNA-cleaving DNA enzyme. *Proc. Natl Acad. Sci. USA*, **94**, 4262–4266.

Schmelzer,C. and Schweyen,R.J. (1986) Self-splicing of group II introns *in vitro*: mapping of the branch point and mutational inhibition of lariat formation. *Cell*, **46**, 557–565.

Schmidt,U., Podar,M., Stahl,U. and Perlman,P.S. (1996) Mutations of the two-nucleotide bulge of D5 of a group II intron block splicing *in vitro* and *in vivo*: phenotypes and suppressor mutations. *RNA*, **2**, 1161–1172.

Sousa,R. and Padilla,R. (1995) A mutant T7 RNA polymerase as a DNA polymerase. *EMBO J.*, **14**, 4609–4621.

Strobel,S.A. and Cech,T.R. (1993) Tertiary interactions with the internal guide sequence mediate docking of the P1 helix into the catalytic core of the Tetrahymena ribozyme. *Biochemistry*, **32**, 13593–13604.

Strobel,S.A. and Shetty,K. (1997) Defining the chemical groups essential for Tetrahymena group I intron function by nucleotide analog interference mapping (NAIM). *Proc. Natl Acad. Sci. USA*, **94**, 2903–2908.

Strobel,S.A., Ortoleva-Donnelly,L., Ryder,S.P., Cate,J.H. and Moncoeur,E. (1998) Complementary sets of noncanonical base pairs mediate RNA helix packing in the group I intron active site. *Nature Struct. Biol.*, **5**, 60–65.

Suchy,M. and Schmelzer,C. (1991) Restoration of the self-splicing activity of a defective group II intron by a small *trans*-acting RNA. *J. Mol. Biol.*, **222**, 179–187.

Westhof,E., Masquida,B. and Jaeger,L. (1996) RNA tectonics: towards RNA design. *Folding Design*, **1**, 78–88.

Xiang,Q., Qin,P.Z., Michels,W.J., Freeland,K. and Pyle,A.M. (1998) The sequence-specificity of a group II intron ribozyme: Multiple mechanisms for promoting unusually high discrimination against mismatched targets. *Biochemistry*, **37**, 3839–3849.

Zimmerly,S., Guo,H., Eskes,R., Yang,J., Perlman,P.S. and Lambowitz,A.M. (1995) A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell*, **83**, 529–538.

Received August 18, 1998; revised September 29, 1998; accepted October 2, 1998