

Identification of structural elements critical for inter-domain interactions in a group II self-splicing intron

Jean-Luc Jestin¹, Elise Dème and Alain Jacquier²

Laboratoire du Métabolisme des ARN, CNRS (URA 13CO)
Département des Biotechnologies, Institut Pasteur, 28 rue du Dr Roux,
F-75724 Paris Cedex 15, France

¹Present address: Laboratory of Molecular Biology and Centre for Protein Engineering, Medical Research Council, Hills Road, Cambridge CB2 2QH, UK

²Corresponding author

Thus far, conventional biophysical techniques, such as NMR spectroscopy or X-ray crystallography, allow the determination, at atomic resolution, of only structural domains of large RNA molecules such as group I introns. Determination of their overall spatial organization thus still relies on modeling. This requires that a relatively high number of tertiary interactions are defined in order to get sufficient topological constraints. Here, we report the use of a modification interference assay to identify structural elements involved in inter-domain interactions. We used this technique, in a group II intron, to identify the elements involved in the interactions between domain V and the rest of the molecule. Domain V contains many of the active site components of these ribozymes. In addition to a previously identified 11 nucleotide motif involved in the binding of the domain V terminal GAAA tetraloop, a small number of elements were shown to be essential for domain V binding. In particular, we show that domain III is specifically required for the interaction with sequences encompassing the conserved 2 nucleotide bulge of domain V.

Keywords: catalytic RNA/group II introns/modification interference/ribozyme/RNA structure

Introduction

Group II introns are found in bacteria as well as in fungi, algae and plant organelles. They are characterized by complex secondary structures that are conserved despite a high degree of sequence divergence (Michel *et al.*, 1989). Some of them have been shown to be ribozymes able to catalyze, *in vitro*, their own excision by a two transesterification pathway that resembles the nuclear pre-mRNA splicing pathway. Group II intron ribozymes are typically large molecules: efficient catalysis of the specific reactions that lead to their excision as lariat molecules requires a set of structures that, altogether, represents ~700 nucleotides (nt).

Modeling of the three-dimensional architecture of such a complex structured RNA requires the identification of a sufficient number of inter-domain tertiary interactions in order to get enough structural constraints. Modeling of

the overall group I intron tertiary structure relied for a large part on constraints obtained from phylogenetic sequence analysis (Michel and Westhof, 1990). Such sequence comparisons have enabled the determination of a reliable model of group II intron secondary structures (Michel *et al.*, 1989). It is organized in six domains radiating away from a conserved wheel. Thus far, however, only three interdomain interactions have been reported (see Figure 1): ζ/ζ' is an interaction between the terminal GAAA tetraloop of the highly conserved domain V and its 11 nt receptor motif within domain I (Costa and Michel, 1995), θ/θ' consists of the interaction between the GNRA tetraloop capping the IC1 subdomain and its receptor site at the base of domain II (Costa *et al.*, 1997) and η/η' is yet another GNRA tetraloop/receptor interaction between domain II and domain VI. Interestingly, this interaction, in which the location of the tetraloop and its receptor can be exchanged between domains II and VI, is involved in a conformational change occurring between the two chemical steps of the splicing reaction (Chanfreau and Jacquier, 1996; Costa *et al.*, 1997). It is thus striking that all interdomain interactions known so far in group II introns involve terminal tetraloops of the GNRA type. This results, in part, from the fact that at least some of the rules governing the association of a given GNRA tetraloop sequence with the different possible receptor motifs have recently been elucidated, making this type of interaction particularly amenable to phylogenetic analysis (Michel and Westhof, 1990; Jaeger *et al.*, 1994; Costa and Michel, 1995, 1997).

One limitation of the co-variation approach resides in the fact that it is restricted to the analysis of the bases, while the interactions between structural domains are expected to rely in part on tertiary contacts involving the sugar phosphate backbone. For these reasons, we have been looking for an approach allowing the experimental determination of structural elements involved in inter-domain interactions. Such an approach could make use of the ability, which has been reported in several instances (see, for examples, Doudna *et al.*, 1987; Jarrell *et al.*, 1988; van der Horst *et al.*, 1991; Doudna and Cech, 1995), to divide a structured RNA (here a ribozyme) into separate structural domains which can reassemble *in vitro*, via tertiary interactions only, to reconstitute a functional molecule. Such a reconstitution of an active ribozyme from complementary parts of a group II intron has been achieved with a large RNA molecule composed of exon 1 and domains I–II and III (E1–DI,II,III transcripts) on the one hand and the small isolated domain V (DV) added *in trans* on the other hand (Jarrell *et al.*, 1988). Domain V is of particular interest because it is the most conserved element of group II introns, both in primary and secondary structure, and it is believed to carry at least some of the catalytic components of the molecule (Jarrell *et al.*, 1988;

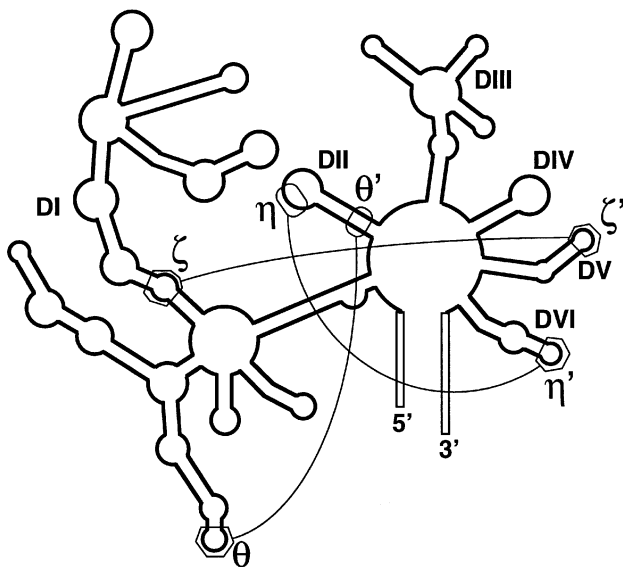


Fig. 1. Schematic representation of the ai5γ intron RNA secondary structures. The thick line represents the intron sequences. The boxes represent the exons. The thin lines point to the three known interdomain interactions: ζ/ζ', η/η' and θ/θ' (Costa and Michel, 1995; Chanfreau and Jacquier, 1996; Costa *et al.*, 1997). DI–DVI label domains I–VI, respectively.

Chanfreau and Jacquier, 1994; Peebles *et al.*, 1995; Abramovitz *et al.*, 1996). This *trans*-complementation assay has been previously used to distinguish kinetically between modifications or substitutions introduced within domain V, those that specifically affect the binding of DV to the E1–DI,II,III molecule (Franzen *et al.*, 1993; Peebles *et al.*, 1995; Abramovitz *et al.*, 1996). Furthermore, direct analysis of the interaction, without relying on ribozyme activity, was made possible by the development of a binding assay based on gel filtration chromatography (Pyle and Green, 1994). Use of this assay in a modification interference analysis allowed the determination of domain V elements required for interaction with the E1–DI,II,III molecule (Chanfreau and Jacquier, 1994) (see also below). The gel filtration assay could be used in that analysis because the small domain V (43 nt) could easily be separated from domain V complexed to the large E1–DI,II,III molecule (~700 nt). In contrast, because free E1–DI,II,III molecules and E1–DI,II,III molecules complexed to domain V have molecular weights in the same range, this assay could not be used for the reciprocal identification of E1–DI,II,III elements required for domain V binding.

In order to circumvent this problem, we have adapted a technique initially developed by von Ahsen and Noller (1995) to identify 16S rRNA elements essential for tRNA binding. This technique makes use of streptavidin-coated magnetic beads to select chemically modified rRNA complexed to 3' end-biotinylated tRNA. In our assay, we use 3' end-biotinylated DV molecules, bound to streptavidin-coated magnetic beads, to select chemically modified E1–DI,II,III transcripts. This specific modification interference assay allowed us to identify elements within the E1–DI,II,III molecule that are critical for domain V binding. Some of these elements fall within the domain I 11 nt motif (ζ) previously shown to bind the terminal GAAA tetraloop of domain V (Costa and Michel, 1995). In addition, several other structures were counterselected,

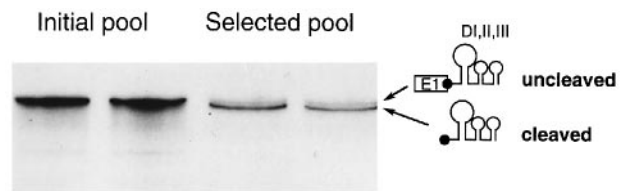


Fig. 2. Preparative acrylamide/bis-acrylamide/urea gel (4%/0.2%/50%) from which input (initial pool) or bound (selected pool) modified, end-labeled E1–DI,II,III transcripts were purified prior to mapping of the modified positions. The doublet, visible on the selected pool, represents the uncleaved (top band) and the 5' splice site cleaved transcripts. The cleaved and uncleaved transcripts are schematized on the right of the gel. E1, exon I; DI,II,III, intron domains I, II and III; the black dot represents the reactive phosphate. The migration was deliberately very short in order to purify the cleaved and uncleaved transcripts in a single gel slice in order to ensure that selection is performed only on binding and not on activity. In this particular gel, the E1–DI,II,III multiple subunit transcripts were modified by incorporation of few phosphorothioates at Gs (left samples) or Us (right samples) in the KNX subunit which was 5' end labeled (see Materials and methods).

identifying new potential sites of interaction. In particular, domain III appears as an essential element for binding to domain V. Moreover, we could show that domain III is required for the interaction with only a subset of domain V elements involved in binding. Finally, we note that most of the sites identified correspond to highly conserved sequences, an observation in agreement with the fact that domain V itself constitutes one of the most conserved structures of group II introns.

Results

Development of the selection procedure

In our first attempts to apply to group II introns the approach described by von Ahsen and Noller, we tried to select specifically on the streptavidin-coated beads the E1–DI,II,III/DV RNA complexes previously formed in solution. However, the same low amount of labeled E1–DI,II,III molecules was recovered by this selection scheme whether domain V was biotinylated or not, indicating that we were not recovering specific complexes. Several hypotheses could explain this observation; for example, the RNA-linked biotin could be shielded in the complex, making it inaccessible to streptavidin. We thus tested another approach in which the streptavidin-coated magnetic beads were previously saturated with the biotinylated domain V. The E1–DI,II,III molecules were then selected on solid phase. The selections were performed under splicing conditions (see Materials and methods). Using this approach, the ratio of specific over unspecific retention of the E1–DI,II,III molecules was typically of the order of 20–50 when non-domain V-coated beads were used in controls.

In a typical experiment, 5% of the E1–DI,II,III labeled molecules were selected for their interaction with domain V. Interestingly, when analyzed by gel electrophoresis, two types of selected molecules were observed: ~10% of the molecules had the size of the starting E1–DI,II,III transcript, while ~90% were found as a discrete product migrating slightly below (Figure 2). Precise size determination showed that this smaller product corresponds to the 5' splice junction cleaved transcript DI,II,III (data

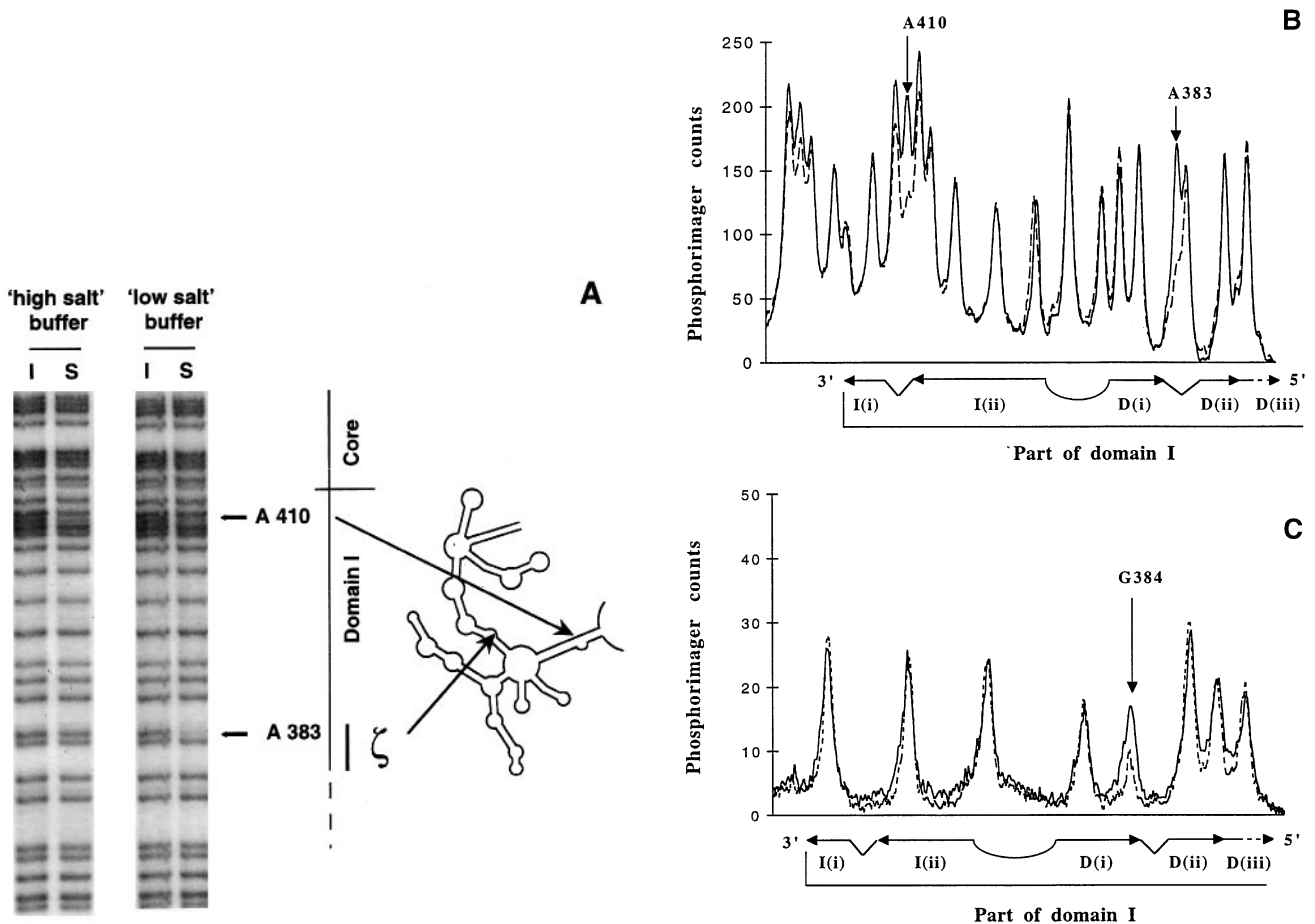


Fig. 3. Examples of phosphorothioate modification interference. (A) The gels (acrylamide/bis-acrylamide/urea gel: 5%/0.25%/50%) show the pattern of phosphorothioate modification at adenosines for the input (lanes I) or selected (lanes S) E1–DI,II,III multiple subunit transcripts in which phosphorothioate adenosines were incorporated in the KNX subunit, labeled at its 5' end (see Materials and methods). The two positions, A410 and A383, at which phosphorothioate incorporation interferes with binding to domain V are labeled, and their respective positions in the secondary structure model of domain I (shown at right) indicated by arrows (see also Figure 5). (B) Phosphorimager profiles of part of the gel shown in (A) ('low salts'). The continuous line corresponds to the input RNA, while the dashed line corresponds to the selected RNA. The secondary structure model of the corresponding part of domain I is schematized below the profiles and labeled according to Michel *et al.* (1989). (C) As in (B), except that the profiles correspond to transcripts in which thiophosphates were incorporated 5' to guanosines. The arrow points to position G384 at which a significant modification interference can be seen after selection in the low-salt buffer.

not shown). This indicates that the selected complexes correspond to catalytically competent structures.

Elements of domains I, II and III essential for binding to domain V

In our modification interference approach, the analyzed transcript is end labeled and randomly modified at a ratio of less than one modification per molecule. This modified population of transcripts is submitted to the selection. Transcripts carrying modifications interfering with the binding to domain V are counterselected. Specific cleavage at the modified positions followed by gel electrophoresis on a polyacrylamide denaturing gel allows identification of the positions whose modification interferes with binding (Conway and Wickens, 1989).

Because the E1–DI,II,III transcripts are large (~750 nt), 3' end labeling does not allow the mapping of the entire sequence. Moreover, 5' end labeling cannot be used because it is removed by the 5' splice site hydrolysis which occurs during the selection procedure. We thus used a multiple subunit transcript. In this construct, E1–DI,II,III was split within a non-essential structure (the

D2b stem-loop within domain I) which was replaced by two long GC-rich complementary sequences ('GC clamp') (Chanfreau and Jacquier, 1996). A functional E1–D,I,II,III transcript can be reconstituted from these two separate transcripts by annealing of the GC clamp. By using 5' or 3' end labeling and modification of either of the two subunits, mapping of the entire E1–DI,II,III transcript could be performed (see Materials and methods).

Two types of modifications were introduced. First, the pro-R phosphodiester oxygen atoms were substituted by a sulfur atom by incorporation of phosphorothioates during *in vitro* transcription (Eckstein, 1985) (see Figure 3). Second, modifications of the bases were analyzed. Purines were modified by diethylpyrocarbonate (DEPC) (disruption of the five-membered aromatic ring) (see Figure 4), while pyrimidines were modified by hydrazine (deletion of the base) (Conway and Wickens, 1989).

Two different splicing buffers were used in the selection procedure. First, a 'high-salt' buffer, containing 0.5 M KCl (see Materials and methods), has previously been reported to be highly permissive to defects affecting the structural stability of the intron (Koch *et al.*, 1992). This

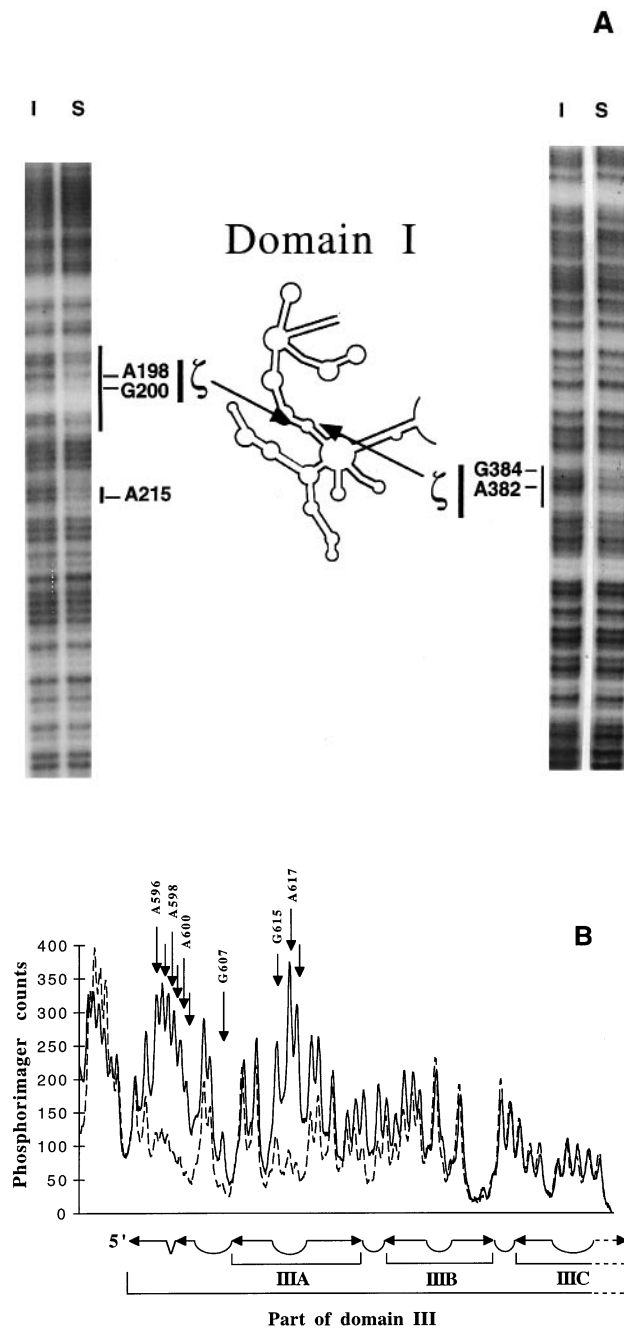


Fig. 4. Examples of DEPC modification interference. (A) Examples of DEPC modification interference gels comparing the input (lanes I) or selected (lanes S) E1-DI,II,III transcripts modified with DEPC. On the left and right sides of the figure, portions of an acrylamide/bis-acrylamide/urea gel (5%/0.25%/50%) encompassing, respectively, the 5' and 3' side of the ζ motif (Costa and Michel, 1995) are shown. On the left-side gel, the $\Delta 52$ -NK subunit of the multiple subunit E1-DI,II,III transcript (see Materials and methods) was modified and 3' end labeled, while on the right-side gel, the KNX subunit was modified and 5' end labeled. Vertical bars point to regions showing a strong modification interference effect. The secondary structure model of domain I is schematized in the middle, with the arrows pointing to the positions of the 5' and 3' sides of the ζ motif, respectively. (B) The figure shows the Phosphorimager profile of the input lane (continuous line) or the selected lane (dashed line) for a region of a gel encompassing part of domain III. The KNX subunit was modified and 3' end labeled. The secondary structure model of the corresponding part of domain III is schematized below the profiles and labeled as in Figure 5. The vertical arrows point to positions showing the strongest modification interference effects, some of which are labeled for identification on Figure 5.

buffer was used for the base modification analysis in order to minimize indirect binding defects due to overall destabilization of the secondary and tertiary structures. Phosphorothioate incorporation, which results in a single atom substitution per molecule, was shown not to affect greatly the RNA secondary structure (Eckstein, 1985). Thus, in addition to the high-salt buffer, a 'low-salt' buffer (see Materials and methods), was used in order to increase the sensitivity of the selection when using this type of modification.

Figure 3A shows examples of gels on which 5' pro-R phosphodiester oxygen atoms of adenosines important for binding to domain V can be seen [compare the initial pools (lanes I) with the selected pools (lanes S)]. Phosphorimager profiles obtained by scanning gels analyzing the same region of the molecule after phosphorothioate incorporation 5' to adenosines or guanosines are shown in Figure 3B and C, respectively. Phosphorothioate interference is seen at the A383, G384 and A410 positions. Incorporation 5' to A383 or G384 inhibits binding only in the most sensitive conditions (low-salt buffer). No other phosphorothioate interference was observed in the DI,II,III molecule. A383 and G384 are located within the ζ motif, while A410 belongs to a highly conserved sequence at the bottom of domain I.

Examples of base modification interference obtained after DEPC treatments are shown in Figure 4 for the region encompassing the ζ motif (Figure 4A) or domain III (Figure 4B).

Figure 5 summarizes the data obtained for the complete DI,II,III molecule. Three regions appear critical: a region encompassing the ζ motif, a bulge and a terminal loop of domain III, and a single pro-Rp oxygen within the bulge at the bottom of domain I.

Comparison between the interference effects observed with the binding assay and an activity assay

It is important at this point to evaluate to what extent the elements identified by this binding assay are indeed specifically involved in domain V binding (possibly through direct interactions) or whether they are rather involved in the global folding of the E1-DI,II,III molecule, influencing domain V binding only secondarily. Comparison with results obtained previously with other types of assays is, in this regard, most revealing. Modification interference on group II introns has been reported using different assays based on activity of the ribozyme. The simplest assay monitors lariat formation in a normal transcript (with all sequences *in cis*). Because, with this type of transcript, the enzyme and substrates are all on the same molecule, one cannot uncouple RNA folding from catalysis. Thus, the assay not only monitors catalysis of branch formation, but also RNA folding and tertiary structure stability. A subset of the positions identified by this assay should thus be involved in inter-domain tertiary interactions and should thus include positions identified by our more specific binding assay.

The complete DEPC modification interference map has been reported for the *Sc.cox1/5* intron (Chanfreau and Jacquier, 1996). A very small subset of the domain I and II positions shown in this activity map is indeed found by our binding assay. In contrast, virtually all domain III

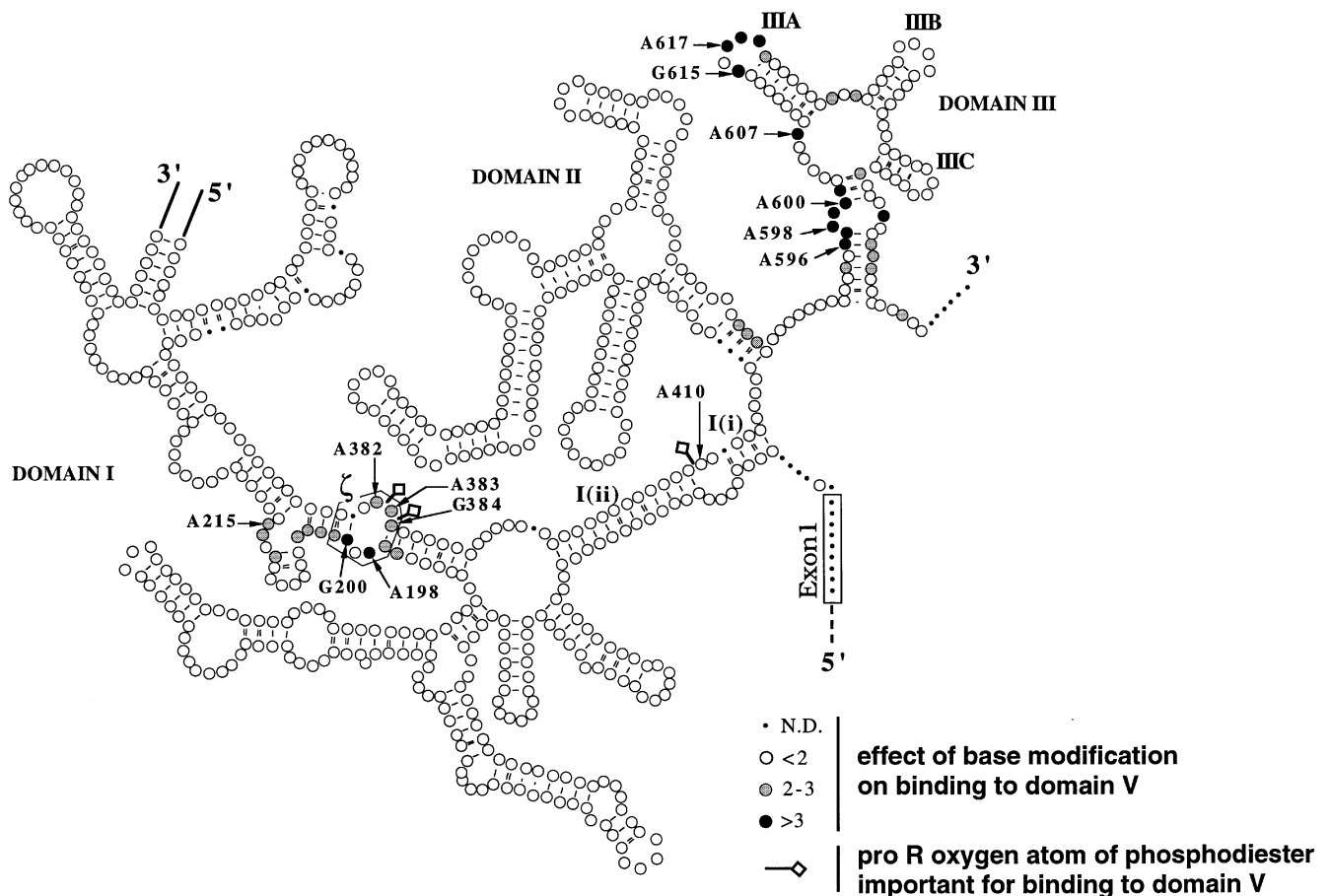


Fig. 5. Summary of the modification interference effects observed on the DI,II,III transcripts for binding to domain V. Each circle represents a single nucleotide on the secondary structure model which is as shown in Chanfreau and Jacquier (1996). White, gray and black circles refer to different levels (as indicated) of inhibition, upon base modification, of binding to domain V. The values are the ratios of band intensities in the input transcript over band intensities in the bound transcripts. N.D., not determined (essentially cytosines which cannot be analyzed with hydrazine from 5' end-labeled transcripts). The diamond pinheads point to positions at which incorporation of a phosphorothioate inhibits binding to domain V. The nucleotides labeled in Figures 3 and 4 are summarized in the figure and indicated by arrows.

positions important in the activity assay are identified by our binding assay (see below and Discussion).

The complete activity map of phosphorothioate interference for branch formation has now also been completed (E.Dème and A.Jacquier, in preparation). Out of several domain I phosphates found to be important for first-step activity, only one, A410, is found to be important for domain V binding. A410 belongs to a sequence, well conserved among group II introns (Michel *et al.*, 1989), in which the incorporation of a phosphorothioate at any of three consecutive positions, A410, A411 or C412, inhibits ribozyme activity (E.Dème and A.Jacquier, in preparation). This observation, which strikingly illustrates the specificity of our binding assay, is pictured in Figure 6 which compares, for this region of the molecule, the phosphorothioate interference profiles for adenosines either with the domain V binding assay or with the first-step activity assay.

In addition, two phosphates, 5' to nucleotides A383 and G384, were shown (in the low-salt buffer only) to be involved in domain V binding, while they were not observed in the 'cis' activity assay (see Discussion).

Domain III is required for the interaction with only a subset of the binding elements of domain V

As mentioned above, the interference pattern previously observed for first-step activity in domain III is very similar to the interference pattern we observe for domain V binding. This raises the possibility that domain III is important for the stability of the global tertiary structure of domain I, affecting domain V binding only indirectly. A prediction of this hypothesis is that the deletion of domain III should affect all elements of domain V involved in binding to the EI-DI,II,III molecule and not specifically a subset of these elements.

To test this hypothesis, we compared the elements, within domain V, which are important for binding to the E1-DI,II,III or E1-DI transcripts (in which domain III was deleted, along with domain II in which no elements were found to be to be important for domain V binding). To do so, we used an adaptation of the gel filtration column chromatography method previously used for selecting domain V molecules bound to E1-DI,II,III transcripts (Chanfreau and Jacquier, 1994; Pyle and Green, 1994) (see Materials and methods). After pre-incubation

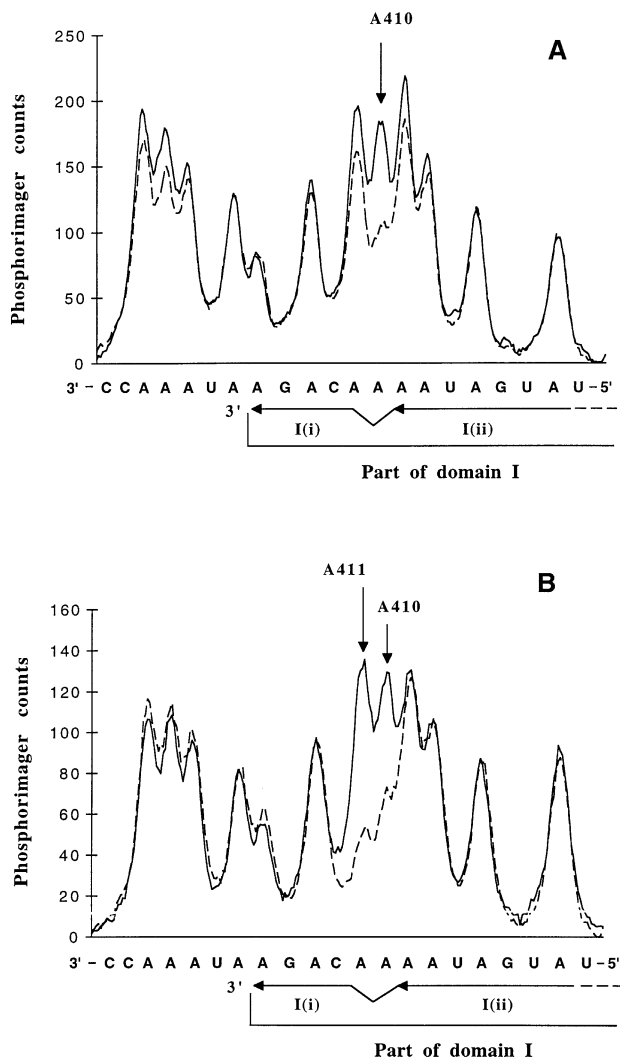


Fig. 6. Comparison of phosphorothioate interference profiles obtained with the binding versus the first-step activity assays. The figure compares, for a sequence overlapping nucleotide A410, the interference profiles, obtained after incorporation of phosphorothioates at adenosines, for the binding assay (A) or the first-step activity assay (B). In (A), the Phosphorimager profile obtained with the input RNA is shown as a continuous line, while the profile obtained for the bound fraction is shown as a dashed line. In (B), the continuous line represents the profile of the unreacted precursor population, while the dashed line represents the lariat population. Both the binding and the activity assays were performed in the low-salt buffer. The arrows and labels point to the positions at which a significant interference effect is visible. The sequence and the corresponding secondary structures are shown below the profiles and labeled as in Figure 5.

with 1 μ M of the E1-DI,II,III molecules in the high-salt buffer, ~48% of the labeled DV transcripts were found complexed (see Materials and methods), consistent with the K_d (0.8 μ M) previously reported for this interaction (Pyle and Green, 1994). In contrast, when the E1-DI molecule was used at 1 μ M, <1% of the DV transcripts were bound. At 5 and 25 μ M, 1.8 and 10% of DV were complexed, respectively. This is likely to correspond to a K_d for the DV/E1-DI interaction at least 50 times higher than that observed in these conditions for the DV/E1-DI,II,III interaction. At low monovalent ion concentration (low-salt buffer, see Materials and methods), 10% of DV was found complexed in the presence of 1 μ M E1-

DI,II,III, while, in the presence of 5 or 25 μ M E1-DI, 1.2 and 4.7% were found complexed, respectively.

In Figure 7, the modification patterns of the selected and input DV transcripts are compared. Essentially, two regions of DV appear most important for binding to E1-DI,II,III: the GAAA terminal tetraloop and a region encompassing the 2 nt bulge. The interference pattern for DV binding after phosphorothioate incorporation has previously been analyzed in the high-salt buffer only (Chanfreau and Jacquier, 1994). No thiophosphate modification was found to inhibit significantly DV binding to E1-DI,II,III in these conditions. In the low-salt buffer, the incorporation of a phosphorothioate at several positions of the DV transcript was found to interfere with binding to the E1-DI,II,III molecules (Figure 7A and C). Most significantly, substitution of the phosphate 5' to either of the last two As of the GAAA tetraloop and to the A preceding the 2 nt bulge (A838) significantly inhibited formation of the complex, while substitution of the phosphate 5' to the C within this bulge (C839) reproducibly enhanced DV binding. Binding to E1-DI leads to very similar modification patterns, except for the region (nucleotides G836-C839) encompassing the 2 nt bulge. The DEPC modification interference of DV binding to E1-DI,II,III was previously reported and correlates with the one observed in this experiment (both experiments were performed in the high-salt buffer). Three positions (G840, A842, G844; see Figure 6B and C), whose modification induced some of the strongest inhibition for E1-DI,II,III binding, are now virtually neutral when analyzed for E1-DI binding.

In conclusion, domain III is specifically required for the interaction between the E1-DI,II,III transcript and structural elements surrounding the 2 nt bulge of domain V, but not with other domain V regions (especially the GAAA tetraloop).

Discussion

To date, the determination of the complete detailed structure of very large RNA molecules, such as group II introns, by conventional approaches like NMR spectroscopy or X-ray crystallography has not been possible. One of the possible routes to solve this problem is to determine the structure of smaller domains and to reassemble these domains, by modeling, in order to get a view of the entire molecule. For example, the latest achievement in the field reports the analysis, by X-ray crystallography, of a structural domain (P4-P6) of a group I intron (Cate *et al.*, 1996a). This is, however, only possible if one has sufficient information relative to the overall spatial organization of these individual structural domains relative to each other.

In addition to phylogenetic/mutational studies, several experimental approaches have been used to gain data on tertiary structures, such as cross-linking (e.g. Harris *et al.*, 1994) or footprinting with structural chemical probes (e.g. Murphy and Cech, 1994). These techniques, however, require that the vast majority of the transcript population is folded in the native conformation, a criterion which can be difficult to meet.

One of the aims of this work was to evaluate whether an alternative approach, based on a modification interference technique, could be successful in identifying sequence

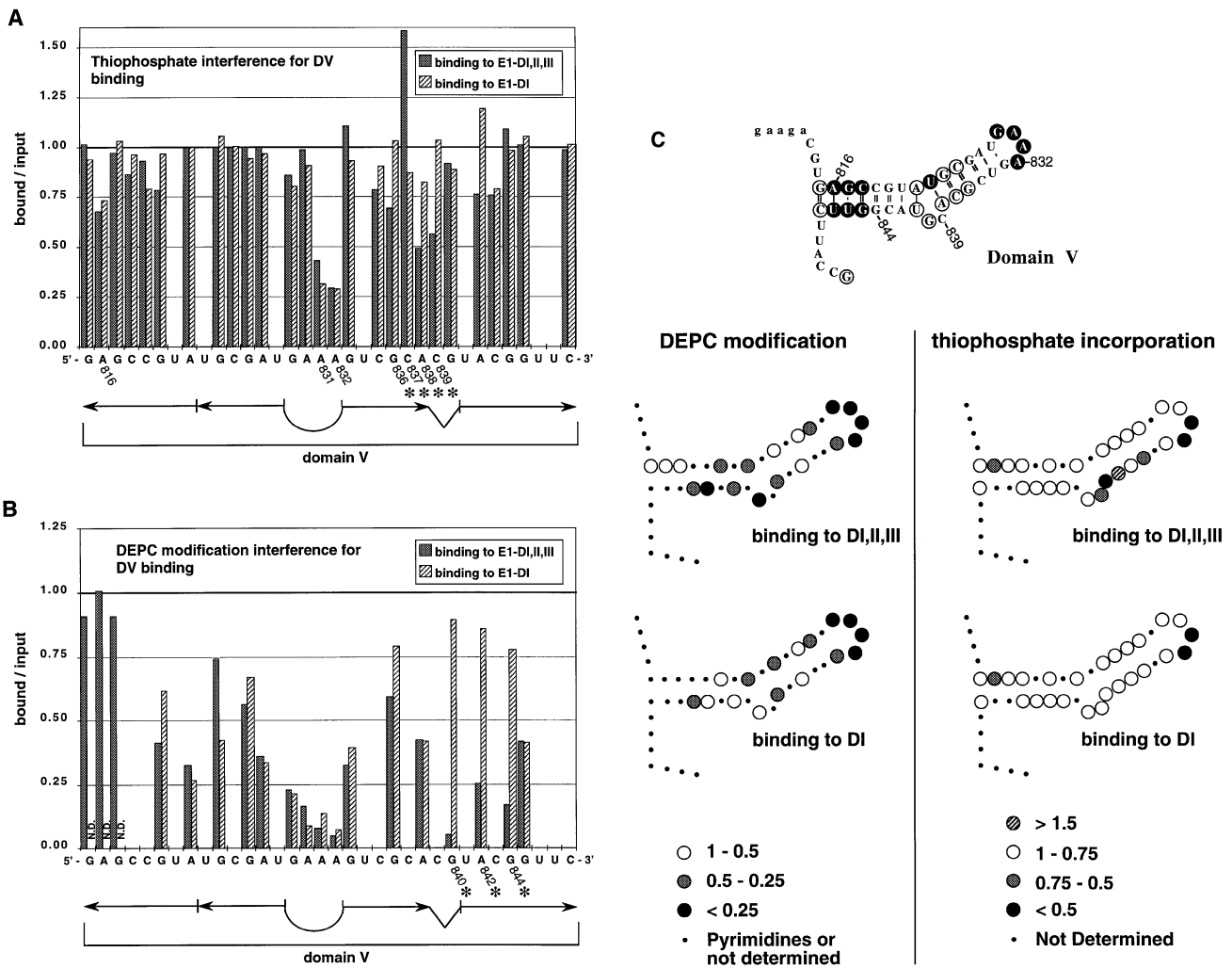


Fig. 7. Domain III is required for the interaction with only a subset of the binding elements of Domain V. (A) Interference experiment for domain V binding (gel filtration assay) to E1-DI,II,III (shaded rectangles) or E1-DI (hatched rectangles) after phosphorothioate incorporation at A, G or Cs. The histogram shows the ratio, for a given domain V nucleotide, of band intensities in the bound fraction over the band intensities in the input RNA. Band intensities were calculated from a Phosphorimager profile of an acrylamide/bis-acrylamide/urea gel (17.5%/0.875%/50%). The sequence of domain V and its secondary structure is indicated below the histogram. Numbered nucleotides refer to positions at which phosphorothioate incorporation significantly affected domain V binding (value inferior to 0.75 or superior to 1.25). When these positions show a differential effect when analyzed for binding to E1-DI,II,III or E1-DI, they are labeled by an asterisk. (B) Same as (A), but for the DEPC modification. Only the nucleotides showing a differential effect when analyzed for binding to E1-DI,II,III or E1-DI are summarized. Not Determined: the effect of DEPC modification at positions G815, A816 and G817 on binding to E1-DI was not analyzable on the gels used, but was not further investigated because no significant effect on binding to E1-DI,II,III was observed with our binding assay. (C) The sequence of the domain V-containing transcript used in the gel filtration column assay is shown at the top of the figure, with the most conserved nucleotides (according to Michel and Ferat, 1995) circled (the most highly conserved nucleotides are circled in black). Lower-case letters refer to a nucleotide present at the 5' end of the transcript, but not encoded within the intron. On the right and left panels, the effects of phosphorothioate incorporation or DEPC modification (as defined in Figure 6A and B, respectively) on binding to E1-DI,II,III or to E1-DI are summarized. Not Determined: the effect of DEPC modification at positions G815, A816 and G817 on binding to E1-DI was not analyzable on the gels used, but was not further investigated because no significant effect on binding to E1-DI,II,III was observed with our binding assay.

elements involved in interdomain interactions. Several observations indicate that this is indeed the case.

The technique was used in such a way as to obtain structural information on an active form of the ribozyme. The RNA selection step has been achieved under splicing conditions, buffers and temperature. Most importantly, >90% of the bound E1-DI,II,III RNAs were found to be cleaved by hydrolysis at their 5' splice site, yielding a DI,II,III molecule (Figure 2), indicating that the selection occurs on a properly folded fraction of the transcript population. Note that the hydrolysis reaction occurring is thought to be relevant with respect to the self-splicing reaction because it has the same regioselectivity (cleavage occurs at the 5'-splice junction) (Jarrell *et al.*, 1988) and

stereospecificities (reaction occurs only with Sp- and not with Rp-phosphorothioates located at the 5'-splice junction) (Podar *et al.*, 1995b) as branch formation.

In general, the chemical modification of a catalytic RNA can affect activity at different levels. It can affect the formation of the secondary structure, of local tertiary structures (folding of structural domains), or of interdomain interactions. In addition, it can directly affect catalysis by modifying catalytic site components. In the normal *cis* reaction, all these effects are expected to take place. In the binding assay, the effects on the formation of secondary structures and local tertiary structures are expected to be minimized because, the assay being *in trans*, the molecules can be separately pre-incubated to

ensure proper folding prior to the selection. Likewise, because the assay does not rely on activity, the effects on catalysis should not be observed. In contrast, when affecting the formation of the complex between the two separately transcribed molecules, the effects on interdomain interactions should be specifically enhanced in the binding assay when compared with the *cis* activity assay.

Comparison of the modification interference map obtained by our binding assay with the previously reported DEPC modification interference map for the first splicing step (Chanfreau and Jacquier, 1996) shows that most of the positions we identify within domain I happen to satisfy these criteria. Strikingly, many of these positions, located in the region between helices D(i) and D(iii), belong to the ζ helix–bulge–helix 11 nt RNA motif previously shown to bind the GAAA terminal tetraloop of domain V (noted ζ') (Costa and Michel, 1995). Conversely, we show that this terminal tetraloop of domain V (ζ') is clearly the most important element for binding to domain I alone. The consistency of the results obtained independently by us and by Costa and Michel thus provides a nice internal control for our binding assay. Interestingly, the two phosphates revealed within the ζ motif by the binding assay (only in the most sensitive conditions) are invisible in the *cis* activity assay (E.Dème and A.Jacquier, in preparation). This most likely results from the configuration *in trans* of the binding assay which makes it not only more specific, but also more sensitive than the *cis* assay for this type of interaction.

In addition to the ζ 11 nt motif, the effects of modifications of several bases, close to ζ , and of the phosphate 5' to nucleotide A410 are also specifically enhanced in the binding versus the first-step activity assay. This identifies these elements as critical for domain V binding.

In contrast to what is observed for domain I, the interference pattern for domain III is very similar for the binding and the first-step assays. However, we show that domain III is involved in the binding of only a subset of the domain V elements involved in the interdomain interaction. This excludes the possibility that domain III is only required for the stability of the global tertiary structure of domain I. Otherwise, the absence of domain III would equally affect the interaction with all domain V elements involved in binding, and not only those encompassing the 2 nt bulge. We must thus conclude that domain III is part of a structural domain specifically involved in binding those specific domain V elements. The fact that the same pattern of modification is observed with the binding and the first-step activity assay indicates that the principal role of domain III is to stabilize the interaction between domain I and domain V. This conclusion is in agreement with previous reports suggesting that domain III is involved in the structure stability of the ribozyme and not in catalysis (Koch *et al.*, 1992; Michels and Pyle, 1995). It is unlikely, however, that most of the elements important for domain V binding identified within domain III directly contact domain V. Domain III was indeed shown, by UV cross-linking, to be a self-structured RNA domain able to adopt its active conformation alone *in vitro* (Podar *et al.*, 1995a). Modifications introduced within domain III might alter its structural stability, affecting in turn the ability of some of its components to

directly, or indirectly, contact the sequence encompassing the 2 nt bulge of domain V.

More generally, the identification of elements critical for an interdomain interaction does not imply that these elements take part directly in this interaction. For example, the effect observed after modification of phosphates could be mediated by a magnesium ion (unfortunately a manganese rescue experiment could not yet be performed with this assay because the addition of manganese considerably enhanced the amount of non-specific binding of the RNA to the beads). Indeed, we note that, for the 11 nt motif found in the crystal structure of the P4–P6 domain of the *Tetrahymena thermophila* group I intron (Cate *et al.*, 1996a,b), Cate *et al.* report the observation of a strong density, in the electron density map, which is consistent with a metal ion or water molecule in the major groove coordinated to the phosphate of the 3'A of the platform (A383 in our group II intron). The incorporation of a phosphorothioate 5' to A383 could thus inhibit the binding of a magnesium ion involved in the stabilization of the local tertiary structure of the 11 nt ζ motif. This would, in turn, indirectly inhibit docking of the domain V GAAA tetraloop in this motif. This hypothesis is in good accordance with our results (see Figure 5). Cate *et al.* also report, in the X-ray structure, a potential water or magnesium ion coordinated to the N7 and O6 of G in the G·U wobble in the motif (the equivalent of G384 in our group II intron). This could also be compatible with our observation of a weak inhibition of domain V binding upon incorporation of a phosphorothioate at this position (see Figure 6C). Indeed, the pro-R oxygen atom is close to N7 and could thus potentially interact with the same ion [in another motif (L5c loop) involving an adenosine platform found within the crystal, the G·U wobble is replaced by a A·U non-canonical pair in which a putative magnesium ion is seen coordinated to the phosphate of the A].

Some phosphates of domain V had previously been assigned a structural role because their substitution by a phosphorothioate affected activity only at low monovalent ion concentrations; these elements were found to be neutral for domain V binding when analyzed by a gel filtration column technique analogous to the one used here, but performed in a buffer with a high concentration of monovalent ions (high-salt buffer) (Chanfreau and Jacquier, 1994). It is thus interesting to compare these data with the results obtained here with the binding assay performed in low monovalent ion concentration (low-salt buffer). From this previous work, it was proposed that phosphorothioate incorporation at four positions inhibits RNA folding: positions A816, C818, G836 and A838. In this work, we indeed found that the phosphates at positions A816, G836 and A838 affect domain V binding to the E1–DI,II,III transcript (while only A816 affects binding to E1–DI). The effect observed in the previous work on activity at position C818 was very weak and the fact that we do not detect it in binding probably reflects the limit of sensitivity of our assay. In addition, the phosphates 5' to the last two As of the GAAA terminal tetraloop are shown here to affect the affinity of domain V for domain I. So, in this region of the molecule, the binding assay appears more sensitive than the activity assay. Note that in the crystal structure of the P4–P6 domain of the *Tetrahymena* group I intron (Cate *et al.*, 1996a), the pro-R

oxygen atom of the phosphate 5' to the last A of the GAAA tetraloop was found to interact directly with the N7 of its own adenosine and with the G of the tetraloop. Substitution of this phosphate by a phosphorothioate is thus likely to perturb the local tertiary structure of the tetraloop, thus interfering indirectly with the interaction with the 11 nt motif. The positive effect that we observe on binding upon thiophosphate incorporation at position C837 could also be observed on activity in the low-salt buffer (see Figure 2D in Chanfreau and Jacquier, 1994). Phosphorothioate substitutions at positions G817 and C839 were previously shown to abolish totally activity in all conditions tested, including the high-salt buffer in which these substitutions exhibited no phenotype in the binding assay. For this reason, it was deduced that these modifications primarily affect catalysis. Our observations strongly support this conclusion. Indeed, phosphorothioate incorporation at position G817 is totally neutral for binding in the low-salt buffer as well. This result is very significant because the neighboring nucleotide A816, which was assigned a structural role because it inhibited activity in low-salt but not in the high-salt buffer, is indeed detected in our binding assay (see Figure 7). The phosphate 5' to C839 now appears to have a dual role, structural and catalytic. Indeed, incorporation of a phosphorothioate at this position affects domain V binding to E1-DI,II,III (but not to E1-DI) in the low-salt buffer, indicating a role in RNA folding. However, this cannot be its unique role because, while its substitution by a phosphorothioate is neutral for binding in the high-salt buffer, it totally inhibits activity in all conditions tested (including the high-salt buffer), indicating an effect on catalysis (Chanfreau and Jacquier, 1994). This contrasts, for example, with the phosphorothioate substitution at the neighboring position (A838) which strongly inhibits binding in our assay, but is neutral for activity at high monovalent ion concentrations (Chanfreau and Jacquier, 1994). This particular phosphorothioate substitution (A838) is also of interest because it affects activity in low salt much more strongly in a first-step assay than in a second-step specific assay (Chanfreau and Jacquier, 1994). In fact, the first-step assay was done 'in cis', precluding pre-incubation of the precursor molecules in splicing conditions, while in the second-step specific assay, an intron-exon 2 molecule was pre-incubated in splicing conditions and the reaction initiated by the addition of exon 1 *in trans*. We now interpret this apparent first-step specific effect by the fact that, in these conditions, RNA folding is much more limiting in the first-step assay than in the second-step assay which includes a pre-incubation step. All these observations are thus fully compatible with the hypothesis that the role of this phosphate is purely structural.

If we cannot conclude that the elements identified by our binding assay are directly contacting domain V, it must be noted, however, that only a very small number of structural elements have been defined as critical for domain V binding. Indeed, aside from the ζ motif, which is known to interact with the GAAA tetraloop of domain V, we are left with only a few elements to interact with the rest of domain V: a few nucleotides close to ζ , the pro-R oxygen atom at position A410, and domain III. It is thus safe to predict that at least some of these elements are at least 'closely associated' with domain V, information

invaluable to orient modeling as well as future experiments aiming at the identification of specific contacts between domain V and the rest of the molecule.

Materials and methods

Transcript preparation

For the modification interference analysis of the E1-DI,II,III RNA, this transcript was reconstituted by annealing, via long GC clamps, of two separate transcripts, $\Delta 52$ -NK and KNX. Transcript $\Delta 52$ -NK was synthesized by transcription *in vitro*, with SP6 RNA polymerase, of *Bam*HI-digested plasmid $\Delta 52$ NK (Chanfreau and Jacquier, 1996). This transcript contains the last 52 nt of the *Sc.cox1/5c* 5' exon, followed by the first 278 nt of the *Sc.cox1/5c* intron and the 5' side of the GC clamp K (Chanfreau and Jacquier, 1996). The KNX transcript was synthesized by transcription *in vitro*, with SP6 RNA polymerase, of the *Xba*I-digested SK+KNXC plasmid (Chanfreau and Jacquier, 1996). It contains the 3' side of the G-C clamp K followed by nucleotides 303–681 of the *Sc.cox1/5c* intron. For the modification interference analysis of the DV molecule bound to E1-DI,II,III, the latter molecule was synthesized by transcription *in vitro*, with SP6 RNA polymerase, of *Xba*I-digested plasmid $\Delta 52$ -XBA which consists of plasmid $\Delta 52$ (Jacquier and Michel, 1987) in which peripheral structures of domain IV (intron nt 682–803) were deleted and replaced by the *Xba*I restriction site. This produces a transcript containing the last 52 nt of the *Sc.cox1/5c* 5' exon, followed by the first 681 nt of the *Sc.cox1/5c* intron. For analysis of domain V bound to E1-DI, the latter molecule was synthesized by transcription *in vitro*, with the SP6 RNA polymerase, of *Sall*-digested plasmid $\Delta 52$ - Δ II which consists of plasmid $\Delta 52$ in which peripheral structures of domain II (intron nt 430–573) were deleted and replaced by the *Sall* restriction site (Chanfreau and Jacquier, 1996). This yields a transcript containing the last 52 nt of the *Sc.cox1/5c* 5' exon, followed by the first 429 nt of the *Sc.cox1/5c* intron. The DV transcript was synthesized from a PCR DNA template as described in Chanfreau and Jacquier (1994). Conditions for RNA transcription, modification, 32 P-end-labeling, annealing and purification are as described previously (Chanfreau and Jacquier, 1994, 1996).

DV-RNA was linked to biotin at its 3' end using an amidocaproyl hydrazide group as described previously (von Ahsen and Noller, 1995). The precipitated DV-RNA (3 nmol) was oxidized specifically at its 3' end with an aqueous solution of KIO_4 (40 mM, 100 μ l) during 30 min at room temperature. To quench the reaction, ethylene glycol (50%, 100 μ l) was added and reacted for 5 min at room temperature. After ethanol precipitation, the DV-RNA reacted with biotin amidocaproyl hydrazide (Sigma, 10 mM, 100 μ l) at 37°C for 90 min. Addition of NaBH_4 (0.2 M, 100 μ l) and Tris-HCl (1 M, pH 8) at 4°C for 30 min resulted in the reduction reaction. The product was precipitated with a 7.5 M ammonium acetate/ethanol mixture (v:v = 1:6), washed with 80% ethanol, dissolved in formamide (20 μ l), denatured for 3 min at 37°C and loaded on a preparative gel (acrylamide/bisacrylamide 6%/0.3%, urea 50%).

Binding assays

Streptavidin-coated magnetic beads (Dynal, M-280, 30 μ l, 10 mg/ml) were washed twice with a 0.1 M NaOH, 0.05 M NaCl solution, twice with a [0.1 M NaCl, 1% bovine serum albumin (BSA)] solution before incubation with biotinylated DV-RNA (75 pmol) in 40 μ l of 1 M NaCl at room temperature for 10 min. Once the supernatant had been discarded, the magnetic beads were resuspended in 40 μ l of the appropriate splicing buffer [high-salt buffer: 100 mM MgCl_2 , 0.5 M KCl, 40 mM Tris-Cl (pH 7.5); low-salt buffer: 100 mM MgCl_2 , 40 mM Tris-Cl (pH 7.5)] containing 10 pmol of the labeled and modified E1-DI,II,III RNAs. The mixture was incubated for 10 min at 45°C followed by 15 min at room temperature. The magnetic beads, which can be captured with a magnet, were washed twice with high-salt splicing buffer. The RNAs bound specifically to the biotinylated DV-RNA linked to the streptavidin-coated magnetic beads were eluted with a formamide solution (10 mM NaOH, 0.5% SDS, 20 mM EDTA). These RNAs were purified on a gel (acrylamide/bisacrylamide 4%, urea 50%) after denaturation for 5 min at 37°C. Reaction tubes were siliconized.

For the modification interference analysis of domain V bound to the E1-DI,II,III or E1-DI molecules, we used a modification of the gel filtration column technique previously described (Chanfreau and Jacquier, 1994; Pyle and Green, 1994). The main modification was the use of small spun columns in place of the normal column for gel filtration.

Among other advantages, this technique minimizes the time between the incubation of the samples under splicing conditions and the actual purification of the complexed molecules. In brief, 3' end-labeled, modified DV transcripts (1 pmol), pre-incubated for 5 min at 45°C in 20 µl splicing buffer (high- or low-salt buffer) were mixed with 20 µl of E1–DI,II,III transcripts (2 µM) or E1–DI transcripts (10 or 50 µM) pre-incubated for 5 min at 45°C in splicing buffer. The 40 µl of mixed transcripts were incubated for 10 min at 45°C and then immediately loaded onto a CHROMA SPIN+TE-100 column (Clontech), equilibrated in splicing buffer according to the procedure described by the manufacturer. The column was then spun for 5 min at 700 g at room temperature. Forty microliters of formamide loading buffer were added to the samples (recovered in a siliconized tube) which were then loaded on a preparative acrylamide/bis-acrylamide/urea gel (8%/0.4%/50%).

Modification interference

Gel purifications and analysis of the modified positions were carried out as described previously (Chanfreau and Jacquier, 1994, 1996). The gels were quantified using a PhosphorImager (Molecular Dynamics). Modification interference on first-step activity was performed as described previously (Chanfreau and Jacquier, 1996).

Acknowledgements

We thank P.Legrain, M.Costa and F.Michel for stimulating discussions and critical reading of the manuscript, and U.von Ahsen for advice. This work was funded in part by a grant from the French Ministry of Higher Education and Research (ACC-SV5). E.D. was supported by a predoctoral fellowship from the French Ministries of Research and Technology and of Higher Education and Research.

References

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.

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

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

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.

Conway,L. and Wickens,M. (1989) Modification interference analysis of reactions using RNA substrates. *Methods Enzymol.*, **180**, 369–379.

Costa,M. and Michel,F. (1995) Frequent use of the same tertiary motif by self-splicing 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.*, in press.

Costa,M., Dème,E., Jacquier,A. and Michel,F. (1997) Multiple tertiary interactions involving domain II of group II self-splicing introns. *J. Mol. Biol.*, in press.

Doudna,J.A. and Cech,T.R. (1995) Self-assembly of a group I intron active site from its component tertiary structural domains. *RNA*, **1**, 36–45.

Doudna,J.A., Gerber,A.S., Cherry,J.M. and Szostak,J.W. (1987) Genetic dissection of an RNA enzyme. *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 173–180.

Eckstein,F. (1985) Nucleoside phosphorothioates. *Annu. Rev. Biochem.*, **54**, 367–402.

Franzen,J.S., Zhang,M.C. and Peebles,C.L. (1993) Kinetic analysis of the 5' splice junction hydrolysis of a group-II intron promoted by domain-5. *Nucleic Acids Res.*, **21**, 627–634.

Harris,M.E., Nolan,J.M., Malhotra,A., Brown,J.W., Harvey,S.C. and Pace,N.R. (1994) Use of photoaffinity crosslinking and molecular modeling to analyze the global architecture of ribonuclease P RNA. *EMBO J.*, **13**, 3953–3963.

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

Jaeger,L., Michel,F. and Westhof,E. (1994) Involvement of GNRA

tetraloop in long-range RNA tertiary interactions. *J. Mol. Biol.*, **236**, 1271–1276.

Jarrell,K.A., Dietrich,R.C. and Perlman,P.S. (1988) Group II intron domain 5 facilitates a *trans*-splicing reaction. *Mol. Cell. Biol.*, **8**, 2361–2366.

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.

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

Michel,F. and Westhof,E. (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.*, **216**, 585–610.

Michel,F., Umeson,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.

Murphy,F.L. and Cech,T.R. (1994) GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. *J. Mol. Biol.*, **236**, 49–63.

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

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

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

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

van der Horst,G., Christian,A. and Inoue,T. (1991) Reconstitution of a group I intron self-splicing reaction with an activator RNA. *Proc. Natl Acad. Sci. USA*, **88**, 184–188.

von Ahsen,U. and Noller,H.F. (1995) Identification of bases in 16S rRNA essential for tRNA binding at the 30S ribosomal P site. *Science*, **267**, 234–237.

Received on December 3, 1996; revised on January 23, 1997