Frequent use of the same tertiary motif by self-folding RNAs.

Maria Costa, François Michel

The EMBO Journal vol.14 no.6 pp.1276–1285, 1995

Frequent use of the same tertiary motif by self-folding RNAs

Maria Costa and François Michel

Centre de Génétique Moléculaire du CNRS,
91190 Gif-sur-Yvette, France

Communicated by I.W.Mattaj

We have identified an 11 nucleotide RNA motif, [CCUAAG...UAUGG], that is extraordinarily abundant in group I and group II self-splicing introns at sites known, or suspected from co-variation analysis, to interact with hairpin terminal loops with a GNRA consensus sequence. Base substitution experiments using a ribozyme–substrate system derived from a group I intron reveal that this motif interacts preferentially with GAAA terminal loops and binds them with remarkable affinity, compared with other known combinations of GNRA loops and matched targets. A copy of the [CCUAAG...UAUGG] motif which is present in domain I of many group II introns is shown to interact with the GAAA terminal loop that caps domain V. This is the first interaction to be identified between these two domains, whose mutual recognition is known to be necessary and sufficient for group II ribozymic activity. We conclude that interaction of [CCUAAG...UAUGG] with GAAA loops is one of the most common solutions used by nature to solve the problem of compacting and bringing together RNA structural domains.

Key words: GAAA hairpin loops/RNA structure/RNA tertiary motif/ribozymes/self-splicing introns

Introduction

Some large catalytic RNAs, including many group I and some group II self-splicing introns, as well as the RNA component of bacterial RNase P, can function alone in salt solutions (reviewed in Cech, 1993). These molecules have a compact active three-dimensional structure (Latham and Cech, 1989), which ought to include numerous tertiary interactions in order for electrostatic repulsion between backbone phosphates to be overcome. Yet, these RNAs have only four monomeric units at their disposal (modified bases have not been reported from self-splicing introns or bacterial RNase P) and with such a restricted pool to draw from, natural selection is unlikely to have come up with a vast number of useful structural solutions. In other words, it should not come as a surprise if self-folding RNA molecules were to make intensive use of only a relatively small set of tertiary motifs. Identifying these motifs would greatly aid modelling enterprises, which will remain essential as long as the crystallization of large RNAs remains a difficult task.

One illustration of the fact that nature preferentially retains some RNA motifs over others is provided by the distribution of terminal loops in large structured RNAs. Sequence analyses of ribosomal RNAs (Woese *et al.*, 1990), group I and group II self-splicing introns (Michel *et al.*, 1989; Michel and Westhof, 1990) and the catalytic RNA component of bacterial RNase P (James *et al.*, 1988) have revealed an unexpectedly high frequency of some four-residue loops, most notably the ones with a GNRA consensus sequence (R stands for purine and N for any nucleotide). In group I introns, two cases of co-variation involving terminal GNRA loops were interpreted as evidence that these loops contact the minor groove of specific helices in the group I structure and, by extension, it was proposed that the ability of GNRA loops to participate in RNA–RNA tertiary interactions could be a major reason for their selective advantage (Michel and Westhof, 1990). Since then, two lines of evidence have supported this suggestion. In the first case, chemical modification of the *Tetrahymena thermophila* group I intron demonstrated a structural linkage between the GAAA terminal loop of stem P5b and the terminal base pair of helix P6a (Murphy and Cech, 1994; Figure 1). In the second case, introduction of base substitutions in the L9 loop of the *sunY* intron of bacteriophage T4 and its putative P5 helix target (Figure 1) showed that the preferred partner of a CU:AG helix is a GUGA loop, whereas a CC:GG helix prefers a GUAA loop: this was regarded as evidence that residue L9-3 interacts with the base pair that was changed (Jaeger *et al.*, 1994).

In nature, GUAA and GUGA sequences are used with roughly the same frequency in the L9 loops of group I introns. In contrast, other sites in large structured RNAs show strong preference for only one member of the GNRA family (Woese *et al.*, 1990). For instance, most group II introns seem to require a GAAA sequence to cap hairpin V (this stem–loop structure is one of the six constituent domains of the ribozyme component of group II introns; Figure 4; Michel *et al.*, 1989). Such biases might have been thought to result from selection pressure for thermodynamic stabilization of the underlying helix, but in contrast to UNCG loops (Tuerk *et al.*, 1988), GCAA or GAAA terminal loops do not seem to contribute markedly to the stability of hairpin structures (SantaLucia *et al.*, 1992). Strong conservation of the sequence of a loop could also reflect interaction with a protein macromolecule. In fact, recognition of specific members of the GNRA family by proteins is known to occur in the case of the 7S RNA of the signal recognition particle (Siegal and Walter, 1988) and the sarcin/ricin loop of the large subunit ribosomal RNA (Szweczak *et al.*, 1993). In group II introns, however, protein recognition cannot be the sole explanation for the presence of a GAAA loop at the tip of domain V, since *in vitro* chemical modification has shown that this loop is involved in binding of domain V

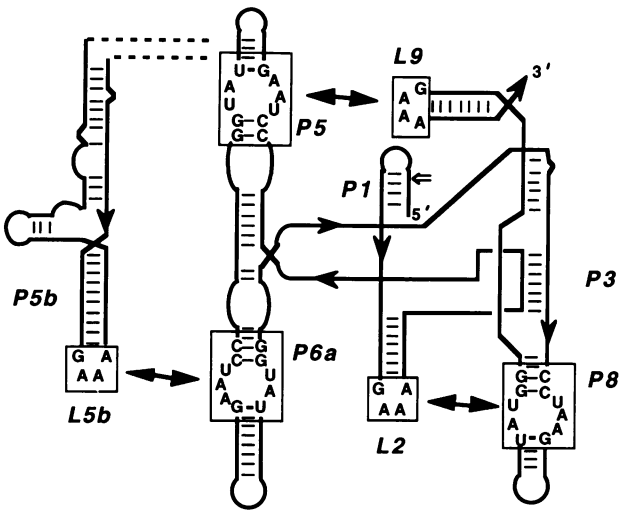

Fig. 1. Location of [CCUAAG...UAUGG] motifs and interacting GAAA loops in group I introns. The secondary structure representation is redrawn from Cech *et al.* (1994) (dashed lines are intended as a reminder that only a fraction of group I introns possess a peripheral domain distal of P5). (i) Twenty two out of 46 currently known introns belonging to the IC1 and IC2 subgroups (Michel and Westhof, 1990; Table II) possess a [CCUAAG...UAUGG] motif or a closely related sequence at the distal end of the P6a pairing and in 21 of these introns, the sequence of the L5b terminal loop is GAAA. (ii) The [CCUAAG...UAUGG] motif or a closely related sequence is present within stem P5 in many introns belonging to the IA, IB and IC subgroups (Michel and Westhof, 1990) and most of these introns have a GAAA L9 sequence (Tables I and II). (iii) The intron in the $tRNA^{Ile}$ gene of *Azoarcus* spp. has a [CCUAAG...UAUGG] sequence starting at P8 bp 3 and a related sequence (CCUAAA...CACGG) exists at the same site in the $tRNA^{Arg}$ intron of *A.tumefaciens* (Reinhold-Hurek and Shub, 1992). In both these introns the sequence of the L2 loop is GAAA.

by domain I (Chanfreau and Jacquier, 1994) (mutual recognition of domains I and V, by some so far unidentified tertiary interactions, has been known for some time to be both necessary and sufficient for hydrolysis of the 5′ intron–exon junction; Jarrell *et al.*, 1988; Koch *et al.*, 1992; W.J.Michels and A.M.Pyle, personal communication). In summary, there seemed to exist a specific RNA partner for interaction with the GAAA loop of domain V.

We have now identified an 11 nt RNA motif that binds GAAA loops with remarkable efficiency. The extraordinary abundance of this motif in group I and group II introns at sites of interaction with GAAA loops testifies to the fact that it has been selected repeatedly by nature as part of one of the optimal solutions to the problem of ensuring efficient mutual recognition of separate RNA domains in large RNA molecules.

Results

Comparative analyses

In *Tetrahymena* and related group I introns (subgroups IC1 and IC2 in Michel and Westhof, 1990), the P5b helix (Figure 1) often ends with a GAAA loop. It was recently suggested (Murphy and Cech, 1994), based on chemical probing of molecules carrying nucleotide substitutions, that the L5b loop interacts with the C:G base pair at the end of stem P6a. We have noticed that this base pair is part of an 11 nt motif, [CCUAAG...UAUGG], that is

Table I. Sequences of loop L9 and stem P5 in 42 subgroup IA introns

P5 stem	L9 loop			
	GAAA	GYAA	GNGA	Other
[CCUAAG...UAUGG] and related motifs	12	0	0	0
CC:GG helix	0	8	0	1
CU:AG helix	4	2	7	2
Other	3	0	0	3

Sequences are from Michel and Westhof (1990) and our own unpublished compilations. P5 stems are scored as containing a [CCUAAG...UAUGG] or related motif when they include a sequence which differs from that consensus by at most three substitutions (3 unpaired nt, 2 nt and 1 bp, or 1 nt and 2 bp, but bp 2 must always be C:G) and such that the second, constant C:G pair is P5 bp 2 (counting as in Michel and Westhof, 1990). CC:GG and CU:AG refer to bp 2 and 3 of uninterrupted P5 helices. Underlined numbers correspond to optimal L2×P8 combinations in *in vitro* kinetic tests (see Table III and Figure 9A).

present in identical or slightly altered form at several other locations in many group I and group II introns (Figures 1 and 4). Remarkably, one of these locations, in group I stem P5, coincides with a site that had already been shown to be involved in recognition of terminal loops of the GNRA family. As seen in Table I, subgroup IA introns with a GUAA or GCAA L9 loop tend to have two consecutive C:G pairs in stem P5, whereas those with a GNGA loop apparently require that the second one of these pairs be U:A; these preferences have already been noticed by Michel and Westhof (1990) and confirmed by base substitution experiments (Jaeger *et al.*, 1994). Introns with a GAAA L9 loop are now seen to form a third category, since they show strong preference for the [CCUAAG...UAUGG] consensus motif, instead of an uninterrupted P5 helix. Actually, significant correlation between occurence of a GAAA L9 loop and that of the [CCUAAG...UAUGG] motif in P5 is not confined to subgroup IA introns (compare Tables I and II). Moreover, the same type of statistical bias is also apparent when loop L5b and stem P6a are considered (Table II).

The [CCUAAG...UAUGG] motif (or a closely related sequence) is also present within the P8 stem of two group I introns (see Figure 1). This is again a site which has been proposed to interact with loops of the GNRA family (Michel and Westhof, 1990) and which could, through these interactions, contribute to the proper positioning of the 5′ splice site in the catalytic centre. In group I introns, 5′ cleavage occurs 3′ of a U:G pair, which is part of a helix (P1) whose backbone must be bound by the ribozyme core of the intron for cleavage to occur (reviewed by Michel and Westhof, 1994). Recognition of the conserved U:G pair by the core may not suffice for correct positioning of the P1 helix, for most group I introns appear to have evolved additional devices to this end. In particular, many introns have a second hairpin structure (P2) next to P1, with a GNRA terminal loop. This loop is located at a fixed number of base pairs from the U:G pair in P1 (which is consistent of course with the notion that P1 and P2 are stacked on top of one another) and its presence imposes specific constraints on the sequence of helix P8 in the ribozyme core (Michel and Westhof, 1990).

Table II. Combined occurences of GAAA loops and [CCUAAG...UAUGG] motifs in group I and group II introns

Combinations (loop×stem)	GAAA loop [CCUAAG... UAUGG] stem	Other loop [CCUAAG... UAUGG] stem	GAAA loop Other stem	Other loop Other stem
L5b×P6a (IC1, IC2)	21	1	12	12
L9×P5 (IC1, IC2)	28	6	2	10
dV×ID$^{(i)}$–ID$^{(ii)}$ (group II)	38	18	11	18

For scoring of motifs see legend to Table I; in P5 stems of subgroup IC1 and IC2 introns with an extended P5 domain (see text and Figure 1), the second C:G pair must coincide with P5 bp 3 (rather than P5 bp 2 in subgroup IA introns). Sequences are from Michel and Westhof (1990), Michel *et al.* (1989) and our own unpublished compilations.

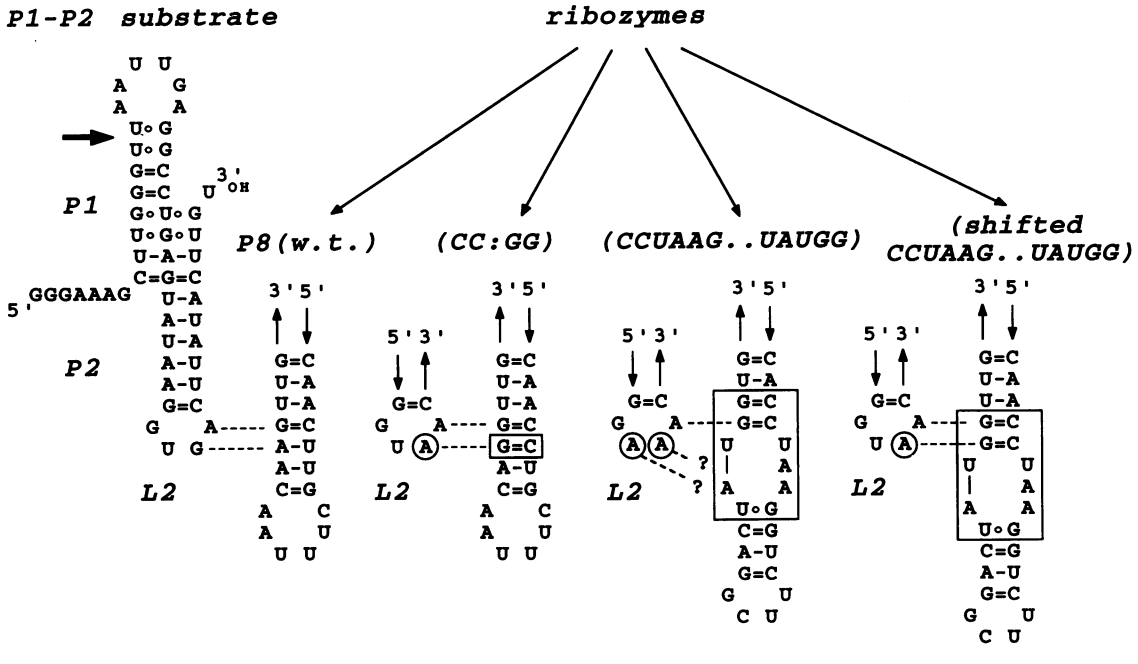

Fig. 2. Base changes introduced in the P8 stems of *td*-derived ribozymes and L2 loops of P1–P2 substrates. The unmutated P1–P2 substrate (far left; residues 8–49 correspond to the wild-type *td* sequence) is shown interacting with the wild-type P8 sequence (see text). The horizontal arrow points to the site of cleavage in P1. Note that the UGAG sequence at the P1–P2 junction can pair with either the sequence to its left (5′ of P1) or the one to its right (3′ of P2). The L2 loop shown facing a mutated P8 stem is that among the three sequences tested (GUGA, GUAA and GAAA) which led to the fastest reaction (see Table III). Relevant base substitutions are circled (GNRA loops) or boxed (P8 stems). Putative base–base tertiary interactions between GNRA loops and target sequences (see Jaeger *et al.*, 1994; Murphy and Cech, 1994) are indicated by dashed lines.

Measuring the strength of binding of GNRA loops by complementary motifs: the L2×P8 interaction in group I introns

We have now tested the possibility of a direct interaction between the L2 loop and stem P8 and also the ability of the [CCUAAG...UAUGG] motif to substitute for a continuous P8 helix, by first splitting the *td* intron of bacteriophage T4 (Shub *et al.*, 1988) into a substrate, consisting of helices P1 and P2 and a short 5′ extension (Figure 2), and a ribozyme (the rest of the intron, minus its last five nucleotides). In the presence of the guanosine cofactor of group I splicing, the former piece is specifically cleaved at the 5′ splice site by the latter, which acts as a catalyst (Figure 2; Materials and methods; M.Costa and F.Michel, unpublished data).

As shown in Figure 3 and Table III, the ability of *td*-derived ribozymes to recognize and cleave a given P1–P2 substrate depends critically on which combination of sequences was chosen for the L2 loop and P8 stem. When rates of cleavage in the presence of excess enzyme ('single turnover' conditions) are compared, ribozymes with the wild-type *td* sequence are seen to prefer the wild-type GUGA L2 sequence over GUAA, but this preference is reversed when P8 bp 5 is changed from U:A to C:G (Figure 3A). This effect, which is just that predicted by phylogenetic analyses of natural group I intron sequences (Michel and Westhof, 1990), has already been reported for molecules carrying homologous substitutions in the L9 loop and P5 helix (Jaeger *et al.*, 1994).

We then explored the consequences of introducing the [CCUAAG...UAUGG] motif at a location within the P8 stem (Figure 2) that should be appropriate, since it is the one at which this motif is found in the *Azoarcus* tRNA^{Ile} intron. Strikingly, whereas ribozymes with a continuous P8 helix poorly catalyze cleavage of a P1–P2 substrate with a GAAA L2 loop (Figure 3A), introduction of the [CCUAAG...UAUGG] motif results in a rate of reaction much higher than for any other combination (Figure 3B). At the same time, ability to cleave a substrate with the wild-type GUGA loop is severely reduced. In contrast,

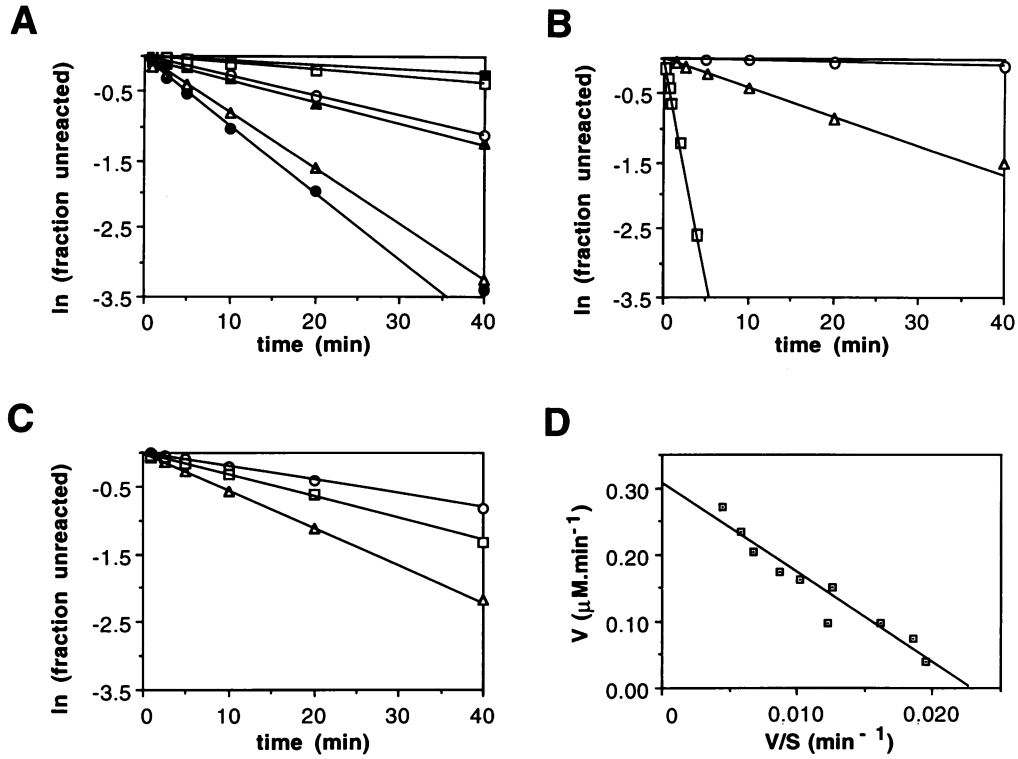

Fig. 3. Kinetics of cleavage of P1–P2 substrates by *td*-derived ribozymes. (**A**) Time course of reaction under single turnover conditions (500 nM ribozyme, 50 nM substrate). The ordinate is the natural logarithm of the fraction unreacted. Filled symbols, wild-type ribozyme; empty symbols, P8 (CC:GG) ribozyme (see Figure 2). Circles, wild-type (GUGA) substrate; triangles, GUAA substate; squares, GAAA substrate. (**B**) As panel (A) but with a P8 [CCUAAG...UAUGG] ribozyme. For reaction with a GAAA substrate, the ribozyme and substrate concentrations were lowered to 100 and 10 nM respectively. (**C**) As panel (A) but with a P8 ('shifted CCUAAG...UAUGG') ribozyme. (**D**) Eadie–Hofstee plot of initial reaction rates with wild-type substrate in excess of wild-type ribozyme (see Materials and methods). Each point is the average of three replicates.

mispositioning of the same motif by just 1 bp (Figure 2) allows only limited discrimination between the three loops (Figure 3C; note that as expected from bp 5 being C:G, GUAA is preferred over GUGA).

Kinetic parameters were determined for the wild-type L2×P8 combination (GUGA L2 loop with CU:AG P8 helix) with substrate in excess of ribozyme (the concentration of the latter was kept at 160 nM). Values of 13.5 μM for K_m and 1.88/min for k_{cat} were obtained by varying the concentration of the P1–P2 substrate (Figure 3D). The resulting value of k_{cat}/K_m (1.4×10^5/min/M) is not too different from the reaction rate constant (2.1×10^5/min/M) determined under single turnover conditions at a ribozyme concentration of 500 nM (Materials and methods and Table III). Since reaction rates for other ribozyme–substrate combinations are generally lower than for the wild-type, it is reasonable to assume that this concentration of ribozyme is far below saturation, not only for wild-type molecules, but in other cases as well, so that the single turnover rates in Table III are likely to correspond also to 'k_{cat}/K_m' conditions. The same should also be true of multiple turnover rates (at 1 μM substrate), most of which do not differ significantly from single turnover rates when expressed relative to the wild-type (see Table III). The one possible exception is the combination of a GAAA loop with the [CCUAAG...UAUGG] motif, which leads to much faster kinetics than even the wild-type combination. Determination of kinetic parameters for this ribozyme–substrate pair (again with substrate in excess of ribozyme; data not shown) yielded K_m = 0.54 μM and k_{cat} = 2.8/min.

Table III. Values of k_{cat}/K_m relative to the wild-type L2×P8 combination

Ribozymes	Substrate (L2 loop)		
	GUGA	GUAA	GAAA
P8 (wild-type)	1 (1)	0.31 (0.26)	0.068 (0.065)
P8 (CC:GG)	0.27 (0.30)	0.78 (0.73)	0.094 (0.120)
P8 (CCUAAG...UAUGG)	0.023 (0.036)	0.39 (0.39)	31.3 (37.3[a])
P8 (shifted CCUAAG...UAUGG)	0.20 (0.15)	0.55 (0.37)	0.32 (0.25)

Single turnover rates (upper values) were determined as described in Materials and methods (see also Figure 3) and divided by the value (2.1×10^5/M/min) obtained for the combination of a wild-type ribozyme and a P1–P2 substrate with a GUGA L2 loop. Values in parentheses (lower values) correspond to multiple turnover rates (see Materials and methods) divided by the correponding wild-type value (1.4×10^5/M/min).

[a]For the combination of a P8 [CCUAAG...UAUGG] ribozyme with a GAAA substrate, k_{cat} (2.8/min) and K_m (0.54 μM) were separately determined by varying the substrate concentration (see Materials and methods).

Comparison with wild-type values reveals that by far the major effect is on K_m, which is consistent of course with the notion that changing partners at L2 and P8 should affect primarily the affinity of the P1–P2 substrate for the ribozyme.

More generally, we believe the values of k_{cat}/K_m we measured directly reflect the strength of the interaction between L2 loops and complementary motifs in the P8 stem: (i) changing the sequence of the P8 helix does not seem to destabilize the *td* intron, as its melting temperature, determined by measuring the fraction of active molecules (Jaeger *et al.*, 1993), remains unchanged (F.Michel, unpublished data); (ii) all our P1–P2 substrates appear to bind in the correct register, since we did not observe miscleavage for any of the combinations tested (compare with Doudna *et al.*, 1989; Herschlag, 1992; Strobel and Cech, 1994); (iii) assuming differences in K_m (and, at constant k_{cat}, in k_{cat}/K_m) indeed reflect the distinct affinities of P8 sequences for L2 loops, the average gap (3.1-fold in rate constant or −0.7 kcal in binding energy; see Table III) between combinations matched and mismatched for loop base 3 and helix bp 5 is about the same as the average ΔΔG (−0.6 kcal) measured from equilibrium melting curves for the corresponding combinations of L9 loops and P5 helices (Jaeger *et al.*, 1994).

In conclusion, the kinetically derived values in Table III probably constitute generally valid estimates of the relative affinities of GNRA loops and complementary motifs. If this is correct, it then means that those loop sequences most often selected by nature to match a given motif (GUAA for CC:GG, GUGA for CU:AG and GAAA for [CCUAAG...UAUGG]; Michel and Westhof, 1990; this work) are precisely the ones which give rise to tightest binding of the interacting partners.

Group II introns: interaction between the terminal loop of domain V and domain I

The [CCUAAG...UAUGG] motif, or a closely related sequence, is present at the junction of stems $ID^{(i)}$ and $ID^{(ii)}$ (Figure 4) in many group II introns belonging to both the major subdivisions of group II (IIA and IIB; Michel *et al.*, 1989). When searching for a potential partner for this motif, we noted that although terminal GNRA loops are commonly used by group II introns (e.g. the secondary structure models in Jacquier and Michel, 1987), only one of these GNRA loops, that which caps domain V (dV), is shared by a majority of group II sequences in each subgroup. The sequence of the dV loop is most often GAAA and, in fact, possession by a group II intron of a dV GAAA loop correlates with the presence of the [CCUAAG...UAUGG] motif, or closely related sequences, at the junction of helices $ID^{(i)}$ and $ID^{(ii)}$ (Table II). Such phylogenetic co-variation strongly suggests that these two components interact in the three-dimensional structure of group II introns. We have proven the reality of this tertiary interaction, the first one to be reported between domains I and V, by introducing compensatory base substitutions at the two potentially interacting sites.

To study the potential interaction between domains I and V, we split the ai5γ intron, an *in vitro* self-splicing group II intron (Peebles *et al.*, 1986; Van der Veen *et al.*, 1986), into a substrate and a ribozyme (see Materials and methods and Figures 4 and 5; note that in intron ai5γ, the sequence of the $ID^{(i)}$–$ID^{(ii)}$ junction differs by 1 bp from the [CCUAAG...UAUGG] consensus). The substrate piece, called a5ΔID, is the ai5γ intron with subdomain ID deleted and flanked by 5′ and 3′ exons of 66 and 184 nt respectively. The other piece, called ID, is a free-standing,

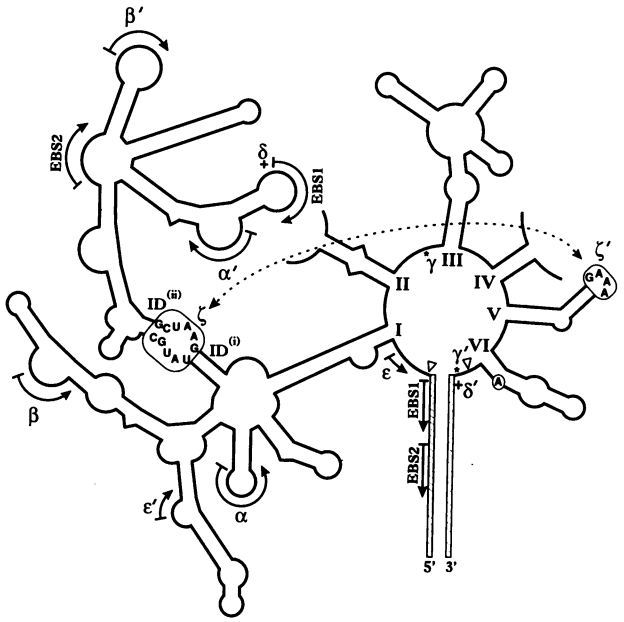

Fig. 4. Schematic representation of the secondary structure of intron ai5γ of *S.cerevisiae* according to Jacquier and Michel (1987). The sequence of the intron is shown as a continuous line and exon sequences as hollow lines. I–VI designate the six secondary structure domains (only the basal part of domains II and IV is shown). $ID^{(i)}$ and $ID^{(ii)}$ are the first and second helices respectively of subdomain D, which constitutes part of domain I. The circled A in domain VI is the site of lariat formation. Those currently known tertiary base pairings that consist of classical base pairs (EBS1–IBS1, EBS2–IBS2, α–α′, β–β′, γ–γ′, δ–δ′ and ε–ε′) are indicated (Jacquier and Michel, 1987, 1990; Michel and Jacquier, 1987; Jacquier and Jacquesson-Breuleux, 1991). Except for γ–γ′ and δ–δ′, arrows indicate extent of pairing and orientation of helix strands. Asterisks mark the two nucleotides involved in the γ–γ′ pairing. The δ–δ′ interaction consists of a base pair between the nucleotide immediately upstream of the EBS1 sequence and the first nucleotide of the 3′ exon (Jacquier and Jacquesson-Breuleux, 1991); these nucleotides are designated by + symbols. Empty triangles indicate an interaction between the first and penultimate nucleotides of the intron (Chanfreau and Jacquier, 1993). The dotted curve with two arrowheads indicates the novel tertiary interaction described in this work. This interaction, called ζ–ζ′, involves the two boxed sequences.

circularly permuted version of the ai5γ ID subdomain, which may be regarded as a ribozyme, since it remains unchanged in a splicing reaction. This ID fragment differs from the wild-type ai5γ ID element in only two respects (Figures 4 and 5): (i) in our construct the ID2b element is interrupted 8 bp from the base of the stem; and (ii) the $ID^{(i)}$ helix was replaced by a 3 bp stem (5′-GGC:GCU-3′) which was capped by the very stable UUCG loop (Tuerk *et al.*, 1988) in order to favour formation of the new, shortened $ID^{(i)}$ helix. The reason we chose this particular partition of the ai5γ intron was that we feared that replacement of the wild-type $ID^{(i)}$–$ID^{(ii)}$ internal loop by different structures would cause changes in the geometry of domain ID, thus preventing one or several tertiary contacts necessary for the splicing reaction. By setting subdomain ID free from the rest of the intron, we have introduced a degree of freedom in the relative positioning of the two ai5γ fragments that should make geometrical changes in the $ID^{(i)}$–$ID^{(ii)}$ section no longer prohibitive.

We have tested this bimolecular system for its ability to react and generate the same final splicing products, i.e.

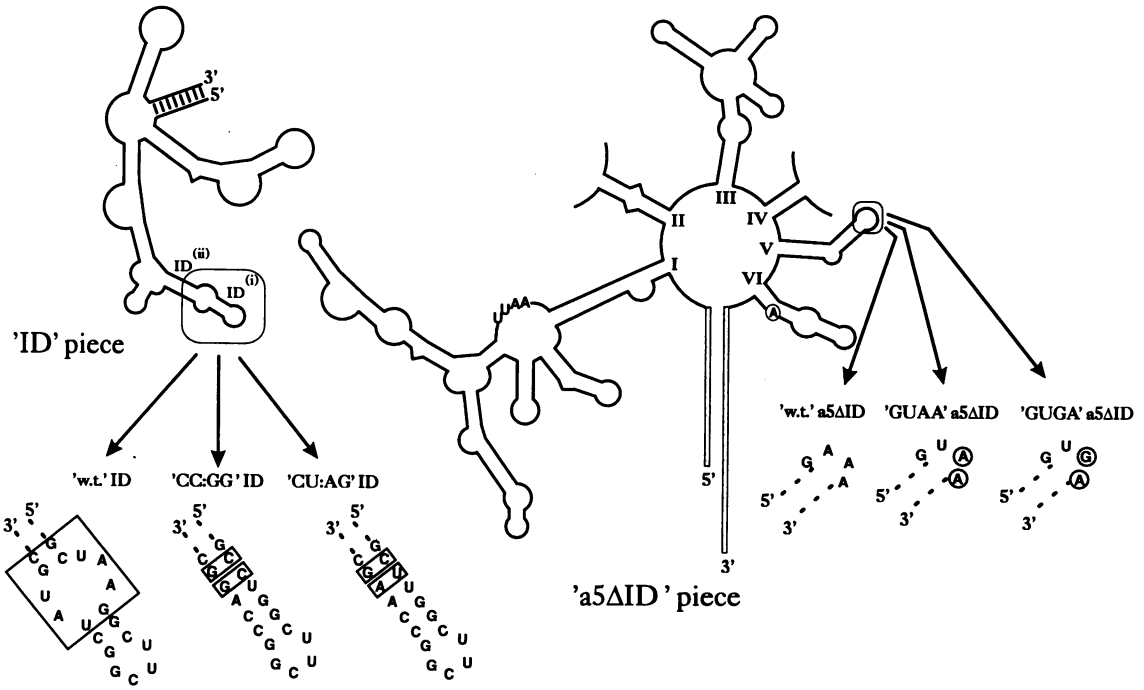

Fig. 5. Schematic representation of the wild-type and mutant versions of the two RNA fragments that constitute the ai5γ bimolecular splicing system. Construction of the ID piece, which is a circulary permuted version of the wild-type ID subdomain with minor changes, is described in Materials and methods. The boxed structure in that piece corresponds to the internal loop lying between the $ID^{(i)}$ and $ID^{(ii)}$ helices in the wild-type ai5γ intron, together with adjacent base pairs. Sequences shown are that of the original (wild-type) structure, as well as those of mutant versions of it (CC:GG ID and CU:AG ID) that were used in this work. The a5ΔID piece corresponds to the ai5γ intron and exon sequences minus subdomain ID, which was replaced by the UUAA sequence shown in domain I. See Materials and methods for the strategy used to obtain this RNA fragment. The boxed structure in this construct corresponds to the terminal loop at the tip of domain V. Wild-type and mutated versions of this loop are shown. Circled nucleotides within the GUAA and GUGA loops are the ones that are believed to interact with the boxed base pairs in the CC:GG ID and CU:AG ID molecules (see text).

the intron excised as a lariat and the ligated exons, as the normal ai5γ reaction. As shown in Figure 6, our (a5ΔID + ID) system is reactive and most probably proceeds through the same reaction pathway as the normal, monomolecular self-splicing reaction, since reaction products with the expected electrophoretic mobilities are obtained. We have also compared rates of reaction for the two splicing systems (Figure 7). Comparison of bimolecular reaction kinetics in the presence of excess ID molecule (single turnover conditions) with the kinetics of the monomolecular reaction show that: (i) except for the first 10 min, the two reactions obey approximately linear kinetics in semi-logarithmic coordinates; and (ii) the (a5ΔID + ID) bimolecular reaction is only 30% slower than the monomolecular one (we checked that the concentration of the ID piece used in this experiment was a saturating one; data not shown). One possible reason for this 30% difference could be that association of the two pieces fails to reconstitute as efficient a catalyst as the complete ai5γ intron, possibly because the $ID^{(i)}$ element does not have the same length and structure in the monomolecular and bimolecular systems. Alternatively, it could have been that not all substrate and/or enzyme molecules are properly folded to react, but this seems unlikely, since we checked that the apparent rate of reaction is insensitive to the duration (up to 15 min) of separate pre-incubation of the a5ΔID and ID pieces at 45°C in splicing buffer (data not shown). Finally, these experiments reveal a striking delay of ~10 min before the reaction reaches steady-state in both splicing systems. This lag had not been reported previously for the ai5γ intron. It must reflect magnesium-induced rearrangements of the tertiary structure of the ai5γ intron prior to reaction in the splicing buffer used in these experiments (pre-incubation of samples in the same buffer without magnesium does not suppress the lag). Since the same delay is observed in the bimolecular system, it seems probable that after association of the two fragments, the same kind of tertiary rearrangements also take place in that system.

In order to test for the existence of an interaction between the terminal loop of domain V and the internal loop at the junction of helices $ID^{(i)}$ and $ID^{(ii)}$, we have introduced into these two structural components base substitutions that we believed could replace the wild-type motifs structurally and functionally (Figure 5). Indeed, based on phylogeny and experimental data (see previous sections), we expected that replacement of the separate $ID^{(i)}$ and $ID^{(ii)}$ helices by a single continuous helix would be feasible, provided the U:U mismatch in the $ID^{(i)}-ID^{(ii)}$ internal loop was replaced either by a C:G pair (in the case of a GUAA dV loop) or an U:A pair (for a GUGA loop). In contrast, all other possible combinations between the three dV terminal loops and three $ID^{(i)}-ID^{(ii)}$ motifs should be unfit to some extent. Figure 8 shows reaction time courses (remaining substrate versus time) under single turnover conditions, with 1 μM enzyme and 100 nM substrate, for all possible combinations. We believe this enzyme concentration to be saturating in all cases, since even for the less efficient pair [wild-type $ID^{(i)}-ID^{(ii)}$ with a GUGA dV loop], the rate of reaction did not vary

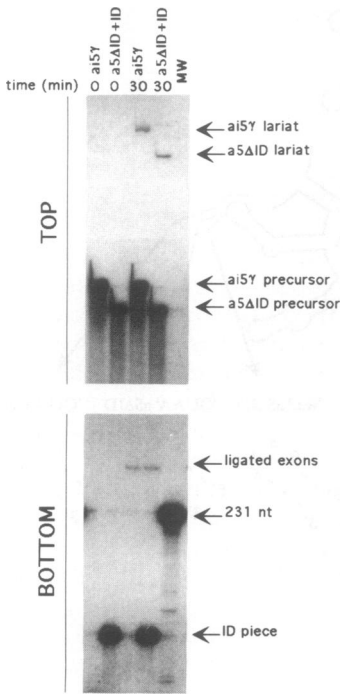

Fig. 6. Comparison of reaction products for the monomolecular and bimolecular ai5γ splicing systems. Only the top and botton of the gel are shown (no splicing products were detected in that part of the gel that is missing). Reaction products are the ones generated when molecules were incubated for 30 min under the conditions described in the legend to Figure 7. Products with electrophoretic mobilities consistent with their expected size and structure were readily identified (lengths of RNA molecules: ai5γ precursor, 1138 nt; a5ΔID precursor, 938 nt; ai5γ lariat, 888 nt; a5ΔID lariat, 688 nt; ligated exons, 250 nt; ID piece, 173 nt). MW lane, RNA molecular weight marker (231 nt).

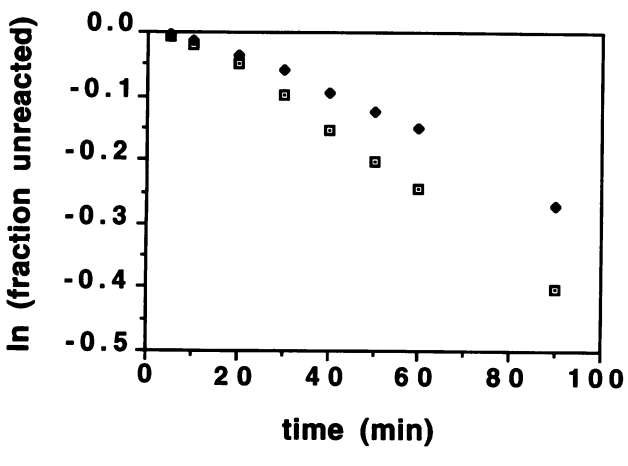

Fig. 7. Comparison of reaction time courses for the monomolecular and bimolecular ai5γ splicing systems. Experimental points for the complete ai5γ intron (⊡) and for the combination of wild-type a5ΔID and ID molecules (◆) were generated by plotting the natural log of remaining precursor (1 − fraction reacted) versus time. The fraction of reacted precursor was determined by dividing the lariat counts (after background correction and molecular size normalization) by the sum of lariat plus precursor counts. In the monomolecular splicing reaction the ai5γ concentration was 500 nM; for the bimolecular splicing system the a5ΔID concentration was 50 nM and the ID concentration was 500 nM.

significantly when the concentration of subdomain ID was varied from 500 nM to 7 μM (with 50 nM substrate; data not shown). Figure 9B summarizes bimolecular reaction

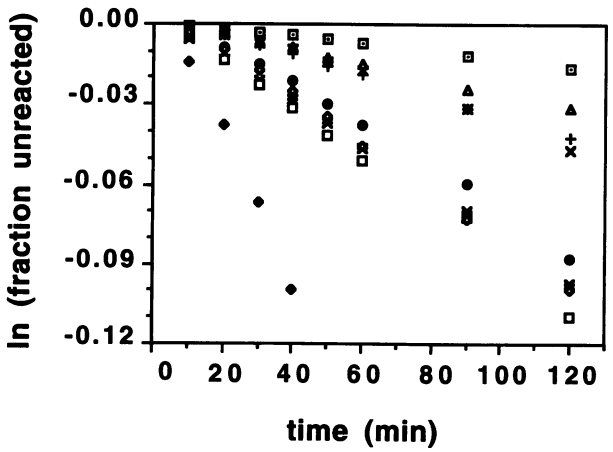

Fig. 8. Comparison of splicing reaction time courses for all possible combinations of the three different ribozymes and substrates. Ordinates as in Figure 7. All experiments were carried out under single turnover conditions with 1 μM ribozyme and 100 nM substrate. Symbols: ◆, wild-type ID + wild-type a5ΔID; ✳, wild-type ID + GUAA a5ΔID; ⊡, wild-type ID + GUGA a5ΔID; +, CC:GG ID + wild-type a5ΔID; □, CC:GG ID + GUAA a5ΔID; ●, CC:GG ID + GUGA a5ΔID; Δ, CU:AG ID + wild-type a5ΔID; ×, CU:AG ID + GUAA a5ΔID; ◇, CU:AG ID + GUGA a5ΔID.

efficiencies for all ribozyme−substrate pairs, taking the wild-type combination as reference. As can be seen: (i) the wild-type combination is the most favourable one; (ii) combinations of the wild-type GAAA loop with CC:GG or CU:AG continuous helices or of the wild-type [GCUAAG...UAUGC] motif with GUAA or GUGA dV loops are much less reactive; (iii) combining, however, a GUAA loop with a CC:GG helix or a GUGA loop with a CU:AG helix results in partial restoration of reactivity (the effect is particularly manifest when GAAA is changed into GUGA and the [GCUAAG...UAUGC] motif into a CU:AG helix).

The results described in this section are in good qualitative agreement with those we obtained for the group I L2×P8 interaction (previous section and Figure 9A; the relatively poorer performance of the GAAA loop in the group II system could reflect the lack of a canonical [CCUAAG...UAUGG] motif in intron ai5γ). These data demonstrate that mutual recognition of domains I and V in the ai5γ intron (and, presumably, in all other members of group II with similar sequences) rests in part on a long-range contact, which we now call ζ−ζ′, between the terminal loop of domain V and the motif at the junction of helices $ID^{(i)}$ and $ID^{(ii)}$. Since the discrimination between 'good' and 'bad' combinations is not abolished at saturing enzyme concentrations, there is no evidence of the ζ−ζ′ interaction being involved in the initial association of the two pieces into which we split intron ai5γ. Rather, our results are consistent with the notion that this interaction is playing a major part in a later step, which, except during the first few minutes of reaction, must be rate-limiting for the splicing process.

Discussion

We have identified an 11 nt RNA motif, [CCUAA-G...UAUGG], with remarkable affinity for GAAA terminal loops. Either this motif, or closely related sequences, are commonly found at several distinct locations in consensus

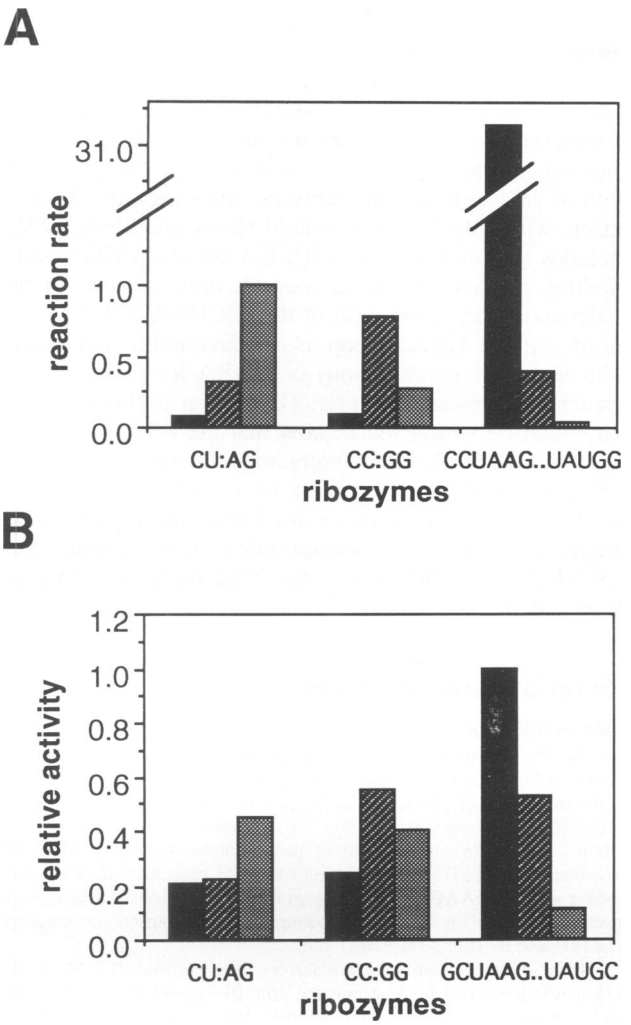

Fig. 9. Relative activities of ribozyme−substrate pairs. (**A**) Group I-derived molecules (L2×P8 interaction). The ordinates is the single turnover rate normalized to the wild-type value (see Table III). Black, GAAA loop; stippled, GUAA loop; grey, GUGA loop. (**B**) Group II-derived molecules (interaction between dV terminal loop and domain I). As (A) except that the ordinate is the inverse (relative to wild-type) of the reaction time for the same extent of reaction (fraction reacted equal to 0.017; see Figure 8).

secondary structure models of group I and group II introns (Figures 1 and 4). Two of these locations, in group I, had already been demonstrated to be sites of interaction with GNRA loops (Jaeger *et al.*, 1994; Murphy and Cech, 1994). In this work we have provided experimental evidence for the existence of two other phylogenetically suggested tertiary interactions involving GNRA loops. In the *td* intron of bacteriophage T4, the L2 terminal loop (GUGA) interacts with helix P8 and this interaction can be replaced by that of a GAAA loop with an appropriately positioned [CCUAAG...UAUGG] motif. In group II introns, the terminal loop of domain V (most often GAAA) interacts with the nucleotides at the junction of helices $ID^{(i)}$ and $ID^{(ii)}$, where a majority of introns have the [CCUAAG...UAUGG] motif or closely related sequences. This latter interaction is of particular interest, since nothing was as yet known of the mechanism by which the small dV hairpin activates the large domain I for hydrolytic cleavage at the 5′ splice site when supplied as a separate RNA piece (Jarrell *et al.*, 1988; Koch *et al.*, 1992; Pyle and Green, 1994).

Direct evidence that the [CCUAAG...UAUGG] motif and a GAAA terminal loop constitute particularly well co-adapted partners was obtained by characterizing the interaction of ribozymes derived from the *td* group I intron with substrates carrying a GNRA loop. The 25-fold smaller K_m for the combination of a GAAA loop with the [CCUAAG...UAUGG] motif compared to the combination of a GUGA loop with a CU:AG helix (as originally found in the *td* intron) translates into a 2.0 kcal energy gap [as already discussed (see Results) we believe K_m values measured in this system to reflect the strength of the L2×P8 interaction]. The fact that the GAAA× [CCUAAG...UAUGG] interaction is so advantageous in terms of binding energy could account for its prevalence whenever tertiary contacts must ensure stable association of two domains. In group II introns, for instance, its presence is likely to constitute a major contributing factor to the remarkably high affinity of the small domain V for the rest of the ribozyme: reported K_m values (190–270 nM), which probably constitute good approximations to the actual K_d (Pyle and Green, 1994), are surprisingly low given the fact that there is no evidence of classical base pairing being involved in mutual recognition (independent evidence that the bases in the dV terminal loop of intron ai5γ are critical for recognition of domain V by the rest of the intron was recently provided by chemical modification of separate dV molecules prior to a binding assay; Chanfreau and Jacquier, 1995).

Alternatively, stereochemical constraints could also be invoked to account for the striking preference for a GAAA loop at the tip of domain V: the target motif for this loop is situated within an essential domain which is presumably subject to strict geometric requirements and these might no longer be met when the separate $ID^{(i)}$ and $ID^{(ii)}$ helices are replaced by a single, continuous pairing (all natural group II sequences have at least two distinct helices at this site; see, for example, Michel *et al.*, 1989). In fact, this possibility could be tested by determining whether an uninterrupted version of intron ai5γ would be as tolerant of helix replacement as our bimolecular system [in which geometrical leeway was ensured by cleaving helix $ID^{(i)}$] and allow the same type of compensatory effects to be observed upon double substitutions.

Indirect, but suggestive, evidence that the various interactions of GNRA loops with their targets are not equivalent in terms of geometry is provided by the fact that the relative frequencies of GNRA variants among L9 terminal loop sequences differ depending on whether or not a group I intron includes an extended domain distal of the P5 target stem. Most introns with a P5 extension have a GAAA sequence at L9 and a [CCUAA...AUGG] motif at the distal end of helix P5 (Table II; the terminal U:G pair is missing, rather, the UAA..AU loop tends to merge into a larger one, which appears to coincide with a sharp bend in the molecule; Murphy and Cech, 1993). In contrast, introns that lack such an extended domain have only a modest excess of GAAA L9 loops over GUGA and GUAA ones (Table I; Michel and Westhof, 1990). These observations, and also our own successful nucleotide substitution experiments (Jaeger *et al.*, 1994; this work), can be reconciled by assuming that for combinations of

the three GNRA loops and their co-adapted motifs to be interchangeable, the part of the molecule that lies distal to the interaction site should be spatially unconstrained (the latter condition is most probably satisfied in the case of stem P8 and, in subgroup IA introns, stem P5, since the terminal sections of these stems are highly variable from one intron to the next; see Michel and Westhof, 1990). But why then, in the latter type of situation, has only a fraction of all natural sequences adopted the GAAA and [CCUAAG...UAUGG] solution and, moreover, why should this fraction differ so much from one site to the next?

The clue to such variation is likely to lie in the notion that biologically optimal solutions need not always be thermodynamically optimal ones. For instance, completion of the group I self-splicing reaction requires the ligated exons to dissociate from their complementary sequence in the intron, which consists in part of the 3′ branch of the P1 helix (reviewed by Cech, 1990). Thus, while strong binding of the P1 helix by the ribozyme core has been shown to favour 5′ cleavage by stabilizing the interaction between the intron and the 5′ exon (Bevilacqua *et al.*, 1992; Herschlag, 1992), it might, on the other hand, interfere with release of the ligated exons (dissociation of the 5′ exon from a version of the *Tetrahymena* intron that lacked a 3′ exon was shown to be rate-limiting for 5′ cleavage under multiple turnover conditions; Herschlag and Cech, 1990). Evidence that the choice of partners at L2 and P8 could similarly reflect the need for their interaction to be a reversible one comes from the fact that, contrary to the situation at other locations where exchanges are expected to be stereochemically tolerated (e.g. at L5b and P6a or, for introns devoid of a P5 extension, L9 and P5), only one of the currently known group I introns (the one in the $tRNA^{Ile}$ gene of *Azoarcus*; Reinhold-Hurek and Shub, 1992) has a [CCUAAG...UAUGG] sequence in stem P8 (the *A.tumefaciens* $tRNA^{Arg}$ intron has a closely related sequence). Interestingly, the *Azoarcus* intron is also remarkable in that its P1 pairing is exceptionally weak, with only 3 bp (one C:G, one A:U and one U:G pair) to hold together the intron and the 5′ exon: by helping to maintain the P1 helix in the catalytic site of the intron, unusually strong binding of the L2 loop could indirectly stabilize the P1 base pairs and thus compensate for the weakness of the intron−5′ exon pairing.

Excessive stability could also be detrimental in the case of the *sunY* intron of bacteriophage T4. The catalytic core of this intron is separated from its 3′ splice site by some 800 nt. Nevertheless, uncoupling of 5′ cleavage and exon ligation and/or alternative splicing are not observed *in vivo* (or *in vitro* at low magnesium concentrations), because stable three-dimensional folding of the *sunY* ribozyme core requires it to interact with a small 3′ terminal intron domain (Michel *et al.*, 1992; Jaeger *et al.*, 1993). Replacement of tertiary interactions such as L9×P5 (in *sunY*, GUGA with CU:AG) by versions that would ensure tighter binding could be detrimental, in that it might free the core from its dependence upon complete synthesis of the intron.

More often than not, however, self-folding RNAs must be under strong pressure to achieve maximal thermodynamic stability with a minimal number of building blocks and this should be especially true of those self-splicing introns that, unlike the *Tetrahymena* intron (Van der Horst *et al.*, 1991) or subgroup IA introns (Jaeger *et al.*, 1991, 1993), cannot rely on the stabilizing effects of large peripheral domains. It is certainly no coincidence that the smallest currently known, naturally occuring group I intron with *in vitro* self-splicing abilities (the *Azoarcus* $tRNA^{Ile}$ intron, with only 205 nt; Reinhold-Hurek and Shub, 1992) includes two instances of a [CCUAAG...UAUGG] motif binding a GAAA terminal loop. It now remains to be understood why interaction of the [CCUAAG...UAUGG] motif with a GAAA loop is so favourable compared with the other combinations of GNRA loops and target sequences investigated so far. The answer to this question may soon be within reach, now that the crystallographic structure of a GAAA loop interacting with two consecutive C:G pairs in the minor groove of an A-type RNA helix has been described (Pley *et al.*, 1994) and crystals of a fragment of the *Tetrahymena* intron that includes the L5b×P6a interaction have also been obtained (Doudna *et al.*, 1993).

Materials and methods

DNA constructs

For the P1−P2 substrate, the DNA template for the RNA molecule shown in Figure 2 was inserted between the T7 promoter and *Rsa*I site of the polylinker of plasmid pTZ19U (US Biochemicals).

For the ribozyme derived from the *td* intron of bacteriophage T4 (Shub *et al.*, 1988), a fragment of the *td* intron, beginning with the sequence AATCTATC 2 nt upstream of the 5′ branch of stem P3 and ending with CTGAACA 5 nt upstream of the 3′ end of the intron, was inserted between the first G following the T7 promoter of plasmid pTZ19U and its (filled-in) *Eco*RI site.

All our group II constructs were derived from plasmid pTZ19U/Sca5, which was generated by inserting the *Hin*dIII−*Eco*RI fragment of the Δ52 construct of Jacquier and Michel (1987) into vector pTZ19U (US Biochemicals). In this pTZ19U/Sca5 plasmid, the ai5γ intron of *Saccharomyces cerevisiae* is flanked by a 5′ exon of 66 nt and a 3′ exon of 184 nt. Plasmid Sca5ΔID was derived from pTZ19U/Sca5 by replacing subdomain ID with the TTAA sequence (Figure 5; see Dichtl *et al.*, 1993); this was achieved by PCR using oligonucleotide 5′-ATCATAAATACGTAAATTTAACAGTCAAAGTTCC-3′. Construction of a free-standing, circularly permuted version of subdomain ID (Figure 5) was carried out by PCR using plasmid Sca5 DNA as a template and two pairs of primers: one pair was formed by oligonucleotides A (5′-CTCTAATACGACTCACTATAGGGAATAAAATGGTTGATGTTATGT-3′) and B (5′-TGCGATGAAGACTTCGAAGCCTTAGCTCTCAAATTATATTACT-3′) and the other by oligonucleotides C (5′-ACGCTAGAAGACTATTCGGCTATGCTCAACGAAAG-3′) and D (5′-ATACCCCCGGGAATAATAATTGAATATCAGAC-3′) [primer A, carrying the T7 promoter, and primer D define the 5′ and 3′ extremities of the circularly permuted ID piece; primers B and C define the sequence distal of helix $ID^{(i)}$; see Figure 5]. The resulting PCR products were ligated together after *Bbs*I digestion. The ligation product was eventually cloned into vector pUC19 using the *Pst*I and *Ava*I sites and the resulting construct was called Sca5ID.

Nucleotide substitutions were introduced by standard mutagenesis procedures (Sambrook *et al.*, 1989). All constructs were verified by sequencing the entire length of the insert.

RNA synthesis and purification

RNA synthesis was as described by Jaeger *et al.* (1993), except that higher concentration ratios (1.6–1.75) of magnesium over nucleotides were used during transcription by T7 RNA polymerase. DNA templates for synthesis of internally radiolabelled ([^{32}P]UTP) P1−P2 substrates were generated by *Rsa*I digestion. Templates for synthesis of *td*-derived ribozymes were generated by digestion with *Xmn*I. Templates for group II transcripts were digested either by *Eco*RI (plasmid Sca5ΔID and mutation-carrying derivatives) or *Sma*I (plasmid Sca5ID and mutation-carrying derivatives).

All transcripts were purified from acrylamide–urea gels. Elution was in 250 mM NaCl, 10 mM Tris–HCl, pH 7.5, 1 mM Na_2EDTA, 0.1% sodium dodecyl sulfate (SDS). After elution, the volume of the samples was reduced to 300–500 µl with 1-butanol prior to ethanol precipitation in the presence of 3 M sodium acetate; pellets were rinsed with 70% ethanol before drying. RNA concentrations were estimated from absorbance at 260 nM.

Reaction kinetics (group I)

All RNA samples were pre-incubated for 5 min at 45°C in 50 mM Tris–HCl, pH 7.5 (at 25°C), 50 mM NH_4Cl, 100 mM $MgCl_2$, 0.02% SDS. Reactions were initiated by mixing the ribozyme solution with a solution of substrate and GTP (the final GTP concentration of 1 mM was checked to be within the optimal range) and stopped by adding an equal volume of urea loading buffer (8 M urea, 10 mM Tris–HCl, 130 mM Na_2EDTA, 0.025% w/v bromophenol blue, 0.025% w/v xylene cyanol). Extents of reaction (molar ratios of the 3′ piece of the cleaved P1–P2 substrate over cleaved plus uncleaved molecules) were determined by loading the samples on a 10% polyacrylamide–8 M urea gel and quantitating radioactivity on the unfixed, undried gel with a PhosphorImager (Molecular Dynamics).

Single turnover kinetics (Table I). The ribozyme and substrate concentrations are given in the legend to Figure 3; extents of reaction were corrected for the fraction (0.05–0.12) of substrate molecules uncleaved after 80 min incubation with an optimally adapted ribozyme.

Multiple turnover kinetics (Table 1). 'Initial' reaction rates (at 1 µM substrate, 100 nM ribozyme) were determined from the first 10% of the reaction.

Determination of kinetic parameters. For the wild-type ribozyme–substrate pair, substrate concentrations were varied from 2 to 60 µM, with 160 nM ribozyme and reaction rates were determined from the first 10% of the reaction (experiments were done in triplicate for each substrate concentration). For the combination of P8 [CCUAA-G...UAUGG] ribozyme (Figure 2) with GAAA substrate the conditions were as for the wild-type, but with 10 nM ribozyme and from 100 nM to 15 µM substrate.

Reaction kinetics (group II)

Mono- and bimolecular splicing reactions were performed at 45°C in 100 mM $MgCl_2$, 40 mM Tris–HCl, pH 7.5, 0.05% SDS. Monomolecular splicing reactions were initiated by adding buffer. For bimolecular splicing experiments, the ID and a5ΔID RNA fragments were incubated separately for 1 min at 45°C in splicing buffer in order to allow folding; reactions were initiated by combining these two RNA pieces. All reactions were stopped with one volume of urea loading buffer. Samples were heated at 80°C for 3 min before loading onto a 4% denaturing polyacrylamide gel. After electrophoresis, gels were dried and quantitated with a PhosphorImager (Molecular Dynamics).

Acknowledgements

We are grateful to Philippe Brion, Guillaume Chanfreau, Jean-Luc Ferat, Alain Jacquier and Eric Westhof for critical reading of the manuscript. This work was supported by Action 'Interface Chimie-Biologie' of the CNRS and grant Bio2CT-93-0345 of the EC.

References

Bevilacqua,P.C., Kierzek,R., Johnson,K.A. and Turner,D.H. (1992) *Science*, **258**, 1355–1358.
Cech,T.R. (1990) *Annu. Rev. Biochem.*, **59**, 543–568.
Cech,T.R. (1993) In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 239–269.
Cech,T.R., Damberger,S.H. and Gutell,R.R. (1994) *Struct. Biol.*, **1**, 273–280.
Chanfreau,G. and Jacquier,A. (1993) *EMBO J.*, **12**, 5173–5180.
Chanfreau,G. and Jacquier,A. (1994) *Science*, **266**, 1383–1387.
Dichtl,B., Pan,T., DiRenzo,A.B. and Uhlenbeck,O.C. (1993) *Nucleic Acids Res.*, **21**, 531–535.
Doudna,J.A., Cormack,B.P. and Szostak,J.W. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 7402–7406.
Doudna,J.A., Grosshans,C., Gooding,A. and Kundrot,C.E. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 7829–7833.
Herschlag,D. (1992) *Biochemistry*, **31**, 1386–1399.
Herschlag,D. and Cech,T.R. (1990) *Biochemistry*, **29**, 10159–10171.
Jacquier,A. and Jacquesson-Breuleux,N. (1991) *J. Mol. Biol.*, **219**, 415–428.
Jacquier,A. and Michel,F. (1987) *Cell*, **50**, 17–29.
Jacquier,A. and Michel,F. (1990) *J. Mol. Biol.*, **213**, 437–447.
Jaeger,L., Westhof,E. and Michel,F. (1991) *J. Mol. Biol.*, **221**, 1153–1164.
Jaeger,L., Westhof,E. and Michel,F. (1993) *J. Mol. Biol.*, **234**, 331–346.
Jaeger,L., Michel,F. and Westhof,E. (1994) *J. Mol. Biol.*, **236**, 1271–1276.
James,B.D., Olsen,G.J., Liu,J.S. and Pace,N.R. (1988) *Cell*, **52**, 19–26.
Jarrell,K.A., Dietrich,R.C. and Perlman,P.S. (1988) *Mol. Cell. Biol.*, **8**, 2361–2366.
Koch,J.L., Boulanger,S.C., Dib-Hajj,S.D., Hebbar,S.K. and Perlman,P.S. (1992) *Mol. Cell. Biol.*, **12**, 1950–1958.
Latham,J.A. and Cech,T.R. (1989) *Science*, **245**, 276–282.
Michel,F. and Jacquier,A. (1987) *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 201–212.
Michel,F. and Westhof,E. (1990) *J. Mol. Biol.*, **216**, 581–606.
Michel,F. and Westhof,E. (1994) *Nature Struct. Biol.*, **1**, 5–7.
Michel,F., Umesono,K. and Ozeki,H. (1989) *Gene*, **82**, 5–30.
Michel,F., Jaeger,L., Westhof,E., Kuras,R., Tihy,F., Xu,M.-Q. and Shub,D. (1992) *Genes Dev.*, **6**, 1373–1385.
Murphy,F.L. and Cech,T.R. (1993) *Biochemistry*, **32**, 5291–5300.
Murphy,F.L. and Cech,T.R. (1994) *J. Mol. Biol.*, 236, 49–63.
Peebles,C.L., Perlman,P.S., Mecklenburg,K.L., Petrillo,M.L., Tabor,J.H., Jarrell,K.A. and Cheng,H.-L. (1986) *Cell*, **44**, 213–223.
Pley,H.W., Flaherty,K.M. and McKay,D.B. (1994) *Nature*, **372**, 111–113.
Pyle,A.M. and Green,J.B. (1994) *Biochemistry*, **33**, 2716–2725.
Reinhold-Hurek,B. and Shub,D.A. (1992) *Nature*, **357**, 173–176.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
SantaLucia,J.,Jr, Kierzek,R. and Turner,D.H. (1992) *Science*, **256**, 217–219.
Siegal,V. and Walter,P. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 1801–1805.
Shub,D.A., Gott,J.M., Xu,M.-Q., Lang,B.F., Michel,F., Tomaschewski,J., Pedersen-Lane,J. and Belfort,M. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 1151–1155.
Strobel,S.A. and Cech,T.R. (1994) *Nature Struct. Biol.*, **1**, 13–17.
Szweczak,A.A., Moore,P.B., Chan,Y.-L. and Wool,I.G. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 9581–9585.
Tuerk,C. *et al.* (1988) *Proc. Natl Acad. Sci. USA*, **85**, 1364–1368.
Van der Horst,G., Christian,A. and Inoue,T. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 184–188.
Van der Veen,R., Arnberg,A.C., Van der Horst,G., Bonen,L., Tabak,H.F., Grivell,L.A. (1986) *Cell*, **44**, 225–234.
Woese,C.R., Winker,S. and Gutell,R.R. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 8467–8471.

Received on November 18, 1994; revised on December 14, 1994