Helix-length compensation studies reveal the adaptability of the VS ribozyme architecture

Julie Lacroix-Labonté[1], Nicolas Girard[1], Sébastien Lemieux[2] and Pascale Legault[1],*

[1]Département de Biochimie, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, QC, Canada, H3C 3J7 and [2]Département d'Informatique et de Recherche Opérationnelle et Institut de Recherche en Immunologie et en Cancérologie, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, QC, Canada, H3C 3J7

ABSTRACT

Compensatory mutations in RNA are generally regarded as those that maintain base pairing, and their identification forms the basis of phylogenetic predictions of RNA secondary structure. However, other types of compensatory mutations can provide higher-order structural and evolutionary information. Here, we present a helix-length compensation study for investigating structure–function relationships in RNA. The approach is demonstrated for stem-loop I and stem-loop V of the *Neurospora* VS ribozyme, which form a kissing–loop interaction important for substrate recognition. To rapidly characterize the substrate specificity (k_{cat}/K_M) of several substrate/ribozyme pairs, a procedure was established for simultaneous kinetic characterization of multiple substrates. Several active substrate/ribozyme pairs were identified, indicating the presence of limited substrate promiscuity for stem Ib variants and helix-length compensation between stems Ib and V. 3D models of the I/V interaction were generated that are compatible with the kinetic data. These models further illustrate the adaptability of the VS ribozyme architecture for substrate cleavage and provide global structural information on the I/V kissing–loop interaction. By exploring higher-order compensatory mutations in RNA our approach brings a deeper understanding of the adaptability of RNA structure, while opening new avenues for RNA research.

INTRODUCTION

Compensatory mutations play an important role in the evolution of macromolecular systems. Their identification has proven to be important for the characterization of macromolecular structure and for revealing important aspects of function. Studies of compensatory mutations in functional RNA molecules are particularly informative, because of the simple base-pairing rules that govern the formation of RNA helical structures. Compensatory mutations in RNA are therefore generally seen as those that maintain base pairing interactions, and their identification has been central for comparative sequence analyses in the prediction of RNA secondary structures and for phylogenetic analyses.

Compensatory mutations in RNA are not strictly limited to base-pairing interactions. Other types of structural compensations, such as helix-length compensation, can help provide higher-order structural and evolutionary information. The concept of helix-length compensation was first introduced in the prediction of coaxially stacked helices for rRNAs (1). It was suggested that since coaxially stacked helices are expected to maintain a constant overall length in a functional RNA architecture, comparative sequence analysis could support coaxially stacked helices when one helix in a group of organisms is shorter, while its partner helix in that group is longer by the same number of base pairs (1). In support of this theory, comparative sequence analysis previously predicted the existence of coaxially stacked helices in 23S rRNA [helices 2646–2652/2668–2674 and 2675–2680/2727–2732; (2)], which was subsequently confirmed in the X-ray structure of the 50S rRNA subunit (3). Coaxial stacking of the group I intron P1 and P2 helices were also previously inferred from helix-length variations (4) and later confirmed from crystal structures (5,6). A more elaborate investigation of tertiary structure compensation has been previously reported as part of a comparative sequence analysis of mitochondrial tRNA sequences that do not fit the standard cloverleaf consensus (7). 3D modeling helped describe several types of structural compensations, including helix-length compensations, that

can help maintain the global L-shape in these atypical tRNAs (7).

In this work, we extend the concept of helix-length compensation for structure–function studies of RNA using the *Neurospora* Varkud satellite (VS) ribozyme as a model system [for reviews, see (8–10); Figure 1A]. Like other nucleolytic ribozymes, the VS ribozyme adopts a simple RNA architecture that allows phosphodiester cleavage activity. Notably, the VS ribozyme has the unique ability to specifically recognize its stem-loop I (SLI) substrate through formation of a kissing-loop interaction with stem-loop V [SLV; (11–13)]. This I/V kissing-loop interaction of the VS ribozyme is ideally suited for performing helix-length compensation studies since variant

substrates and ribozymes can be synthesized separately and then mixed in various combinations for kinetic characterization (Figure 1B). Furthermore, in the 3D models of the VS ribozyme, helices I and V are extended away from each other in a similar manner as coaxially stacked helices (14–18). It is expected that adding (or removing) base pairs within stem Ib would disrupt productive docking of the SLI cleavage loop with the A730 loop of stem-loop VI to form the proposed active site (10,19–27) and significantly reduce cleavage activity, as previously observed for stem V length variants (16,28). We hypothesize that cleavage activity of stem Ib length variants could be restored by adjusting the number of base pairs within stem V of the ribozyme, such that the total number of base pairs within stems Ib and V is preserved (Figure 1B). To explore this possibility, we combined kinetic characterization of variants and 3D modeling. Our helix-compensation study with the VS ribozyme provides global structural information and reveals the adaptability of RNA architecture.

MATERIALS AND METHODS

Plasmids

The pAvapl plasmid used for transcription of Avapl (Figure 1A) was constructed by PCR from the A-3 clone (18,29). The plasmids for transcription of Avapl ribozyme variants (Figure 1B) were prepared by QuikChangeII site-directed mutagenesis from pAvapl (Stratagene, CA). All plasmids were validated by DNA sequencing.

Preparation of RNA for kinetic studies

The R_0 ribozyme and variant sequences (Figure 1) were prepared by *in vitro* T7 RNA polymerase transcription from the pAvapl plasmid (or derivatives) linearized at the AvaI site. All ribozymes were purified as previously described (18). The S_0 substrate and variant sequences (Figure 1) were synthesized *in vitro* using T7 RNA polymerase and a single-stranded synthetic oligonucleotide template (Integrated DNA Technologies, IA). They were purified by denaturing gel electrophoresis, dephosphorylated with calf alkaline phosphatase and 5′-end-labeled with γ-(^{32}P) ATP (PerkinElmer, Ont.) using established procedures (18).

Single-substrate kinetic studies

All reactions were carried out at 30°C in reaction buffer (50 mM Tris pH 7.5, 25 mM KCl, 2 mM spermidine and 25 mM MgCl$_2$). Prior to starting the cleavage reaction, the 5′-^{32}P-labeled substrate was heated to 95°C for 1 min in storage buffer (10 mM Tris pH 7.6 and 1 mM EDTA) and snap-cooled to promote hairpin formation. Cleavage reactions were initiated by addition of equal volumes (20 µl) of the 5′-^{32}P-labeled S and R, either the wild-type R (R_0) or variants (Figure 1B), after a 5-min pre-incubation of S and R at 30°C in reaction buffer. Typically, about 10 aliquots of 3 µl were removed from the 40-µl reaction mixture at specific times, and the reaction was stopped by addition of 12 µl of stop mix (50 mM EDTA pH 8.0,
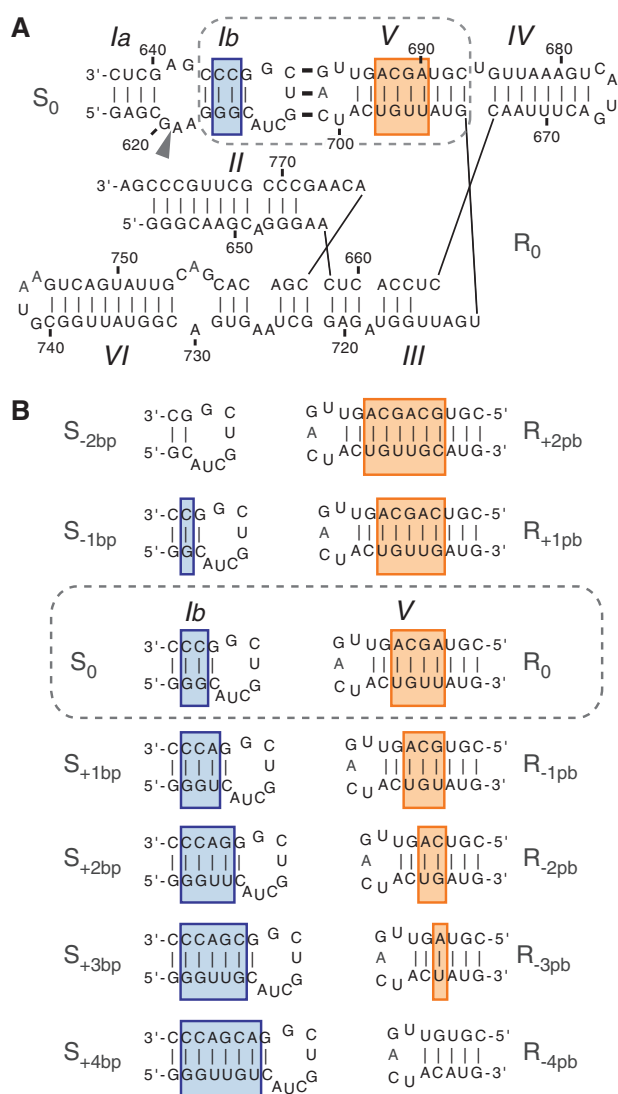
Figure 1. Primary and secondary structures of the substrate/ribozyme (S/R) pairs characterized here. (**A**) Parental VS ribozyme (R_0) and substrate (S_0). Substrate recognition involves a kissing-loop interaction between residues 630-632 of S_0 and 697–699 of R_0. The arrowhead points to the cleavage site and the dashed box encloses the region under investigation. (**B**) Variant S/R pairs. Only the mutated regions are shown along with the nomenclature of the mutated substrates and ribozymes.

0.04% xylene cyanol, 0.04% bromophenol blue and $\geq 85\%$ formamide) and transfer to 4°C. Samples were analyzed by denaturing gel electrophoresis [20% acrylamide:bisacrylamide (19:1) / 7 M urea gels]. For each time point, the amounts of substrate and product were quantified using a Molecular Imager FX and the Quantity One 1D analysis software version 4.6.2 (Bio-Rad Laboratories, Ont.).

Single-turnover kinetic experiments were typically performed with $[S] = 250$ pM and excess R (0.005–5 µM). The percentage of remaining substrate $[F = S_t/(S_t+P_t)*100\%$; where S_t and P_t are the amount of substrate and product at time t] was plotted against time. The resulting time courses were fitted to the equation $F = Ae^{-(k_{obs}*t)} + F_\infty$, where A is the amplitude, k_{obs} is the rate of the reaction and F_∞ is a variable that represents the amount of substrate at the end point of the cleavage reaction. The value of k_{obs}, the standard error for k_{obs}, and F_∞ were estimated by non-linear regression analysis within the OriginPro 8.1 software (OriginLab, MA). The F_∞ values are <10% for all S/R combinations.

The dependence of k_{obs} on $[R]$ was analyzed by linear regression to derive the second order rate constant of the reaction, k_{cat}/K_M. The linear dependence of k_{obs} on $[R]$ was observed from 0.005 to 1 µM for S_0/R_0 and from 0.005 to 3 µM for both $S_{+1\,bp}/R_0$ and $S_{-1\,bp}/R_0$. The quality of the fit for the linear dependence of the k_{obs} on $[R]$ was obtained from the square of the correlation coefficient (R^2), and in all cases R^2 was ≥ 0.99. Single-substrate kinetic experiments were all performed from at least three independent experiments, and reported k_{cat}/K_M and their errors are average values and standard deviations, respectively, from these multiple measurements.

Multiple-substrate kinetic studies

Multiple-substrate kinetic studies were carried out as single-turnover reactions under the same conditions as the single-substrate kinetic studies. The main difference is that the cleavage reactions were initiated by addition of equal volumes (20 µl) of R (0.005–5 µM of either R_0 or a variant) and a mixture of 5′-^{32}P-labeled nucleic acids containing all substrates ($S_{-2\,bp}$, $S_{-1\,bp}$, S_0, $S_{+1\,bp}$, $S_{+2\,bp}$, S_{+3bp} and $S_{+4\,bp}$; 35 pM of each) and, in certain cases, a non-cleavable DNA hairpin control (5′- CGG CGA AGC ATC GTC GGC TTC GCC G -3′; 35 pM). In addition, the percentage of remaining substrate was expressed as $F = (S_t/S_0)*(C_0/C_t)*100\%$ and plotted against time, where S_0 and C_0 are respectively the amount of substrate and non-cleavable $S_{-2\,bp}$ control at time $t = 0$ and S_t and C_t represent the corresponding values at time $t \neq 0$. Values for the reaction rate k_{obs}, amount of uncleaved substrate F_∞ and specificity constant k_{cat}/K_M were derived as for the single-substrate kinetic studies. All multiple-substrate kinetic experiments were performed at least three times, and reported k_{cat}/K_M and their errors are average values and standard deviations, respectively, from these multiple measurements.

Localization of the substrate cleavage site

The site of cleavage was determined for all S/R combinations where cleavage was observed. For this purpose, the substrates were subjected to cleavage in single-substrate kinetic conditions chosen to yield ~50% cleavage. Cleavage reactions were analyzed by denaturing gel electrophoresis [20% acrylamide:bisacrylamide (19:1) / 7 M urea gels] and by comparison with partial T1 nuclease digest and alkaline hydrolysis of S_0.

Native gel studies of RNA substrates

Non-denaturing polyacrylamide gel electrophoresis was carried out on a 7.5% polyacrylamide bisacrylamide (37.5:1) gel in TB buffer (50 mM Tris borate pH 8.0 and 5 mM magnesium acetate). Prior to loading, the RNA substrates were heated at 95°C for 2 min in storage buffer and transferred on ice for 5 min. The gel was pre-run for 30 min at 200 V, run for 2 h at 250 V and 4°C and then stained with Stains All (Sigma-Aldrich, Ont.).

Structural modeling of SLI/SLV complexes

3D models of the SLI/SLV complex were generated using a stepwise procedure (Supplementary Figure S3). As a first step, RNA fragments were modeled using the constraint-satisfaction software MC-Sym 3.3.2 (30). Several models of the I/V kissing-loop junction core (residues 626–634 and 695–701) were first obtained separately using the NMR structure of magnesium-bound SLV (31) and constraints derived from biochemical studies, as previously described (18). A unique structural model was obtained for each type of SLI stems (S_{-1bp} to S_{+4bp}; residues 619–625 and 635–640 for S_0), in accordance with the secondary structure model (Figure 1) and the NMR structure of the active conformation of the cleavage site internal loop (32). A unique structural model was also obtained for each type of SLV stems (R_{-4bp} to R_{+1bp}; residues 689–694 and 702–707 for R_0), in accordance with the secondary structure model (Figure 1). Before being grafted together, the core and stem models were energy minimized using the molecular dynamics software Sander (33) based on the Amber force field and including covalent, electrostatic, Van der Waals energies, as well as implicit solvation in the form of a Generalized Born model (34). Several rounds of minimization (100 000) were first applied including restraints to maintain Watson-Crick base pairing and followed by fewer rounds of minimization (1000) without these restraints. The minimized core and stem models were then filtered using a custom C++ program based on the freely accessible C++ library MC-Core (30). The filter-selected models that preserved global structural integrity (heavy-atom rmsd between the pre- and post-minimization model <2.0 Å) and key structural features (covalent integrity of backbone, Watson-Crick base pairs at the I/V kissing-loop junction, and hydrogen bonds typical of U-turn structure in SLI between U628 2′-O and G630 N7 and between U628 N3 and U631). Each remaining core model was then assembled using MC-Sym with all possible combinations

of stem I (S_{-1bp} to S_{+4bp}) and stem V (R_{-4bp} to R_{+1bp}), energy minimized and filtered to create 36 sub-models that form a common-core group.

For all sub-models, helical axes were computed using Curves+ for stem Ib and V (35). The interhelical angle (θ_{I-V}) and axial displacement (d_{I-V}) between these two axes were computed by vector methods within Pymol using Python scripts. For calculation of d_{I-V}, a reference point O was first set midway between atoms corresponding to A698 N3 and U631 N1 of the kissing-loop interaction. Two segments OA and OB were calculated that start at this reference point to reach the two helical axes at point A and B in a perpendicular orientation. The distance of the AB segment was taken to represent the axial displacement. The reported values of θ_{I-V} and d_{I-V} are average values of all sub-models from a common-core group.

### Selection of SLI/SLV models compatible with kinetic data

In order to select common-core groups that are compatible with the kinetic data, the sub-models of each common-core group were first compared to their reference S_0/R_0 sub-model using an rmsd criteria. The sub-models were first superposed to the S_0/R_0 sub-model based on alignment of the first base pair of SLV (residues 689 and 707 in S_0/R_0). Heavy atoms pairwise rmsds between the SLI cleavage site region of the S_0/R_0 sub-model (residues 619–623 and 637–640) and that of each sub-models were then computed. For each common-core group, the rmsd distributions for the high-activity [$(k_{cat}/K_M)/(k_{cat}/K_M) \geq 0.25$)] and low-activity [$(k_{cat}/K_M)/(k_{cat}/K_M) \leq 0.01$)] classes of S/R pairs were compared using a Welch's t-test (36). The normality assumption was verified using the Shapiro-Wilks test (37) on each distribution, and no significant violation was found ($\alpha = 0.05$). Common-core groups compatible with kinetic data were selected when statistically significant differences (at $\alpha = 0.02$) were observed between the mean rmsd values of high-activity and low-activity classes of S/R pairs.

RESULTS

For helix-length compensation studies, we performed kinetic studies of trans cleavage with the Avapl VS ribozyme and an SLI substrate, herein termed R_0 and S_0, respectively (Figure 1A). The S_0 substrate adopts a pre-shifted active state (Figure 1A) that is rapidly cleaved by the VS ribozyme (18). Helix-length mutations were introduced within stem Ib of S_0 (Figure 1B) that remove up to 2 bp (S_{-1bp} and S_{-2bp}) or add up to 4 bp (S_{+1bp}, S_{+2bp}, S_{+3bp} and S_{+4bp}). Helix-length mutations were also introduced within stem V of R_0 (Figure 1B) that add up to 2 bp (R_{+1bp} and R_{+2bp}) or remove up to 4 bp (R_{-1bp}, R_{-2bp}, R_{-3bp} and R_{-4bp}). Compensatory substrate/ribozyme (S/R) variant pairs are those for which the total number of base pairs in stems Ib and V is identical to the S_0/R_0 reference pair (Figure 1B). We hypothesize that, in the simplest scenario, compensatory S/R variant pairs would retain cleavage activity, whereas other S/R variant pairs would not.

Single-substrate kinetic studies

We first examined the specificity of R_0 for each substrate individually by determining the second-order rate constant (k_{cat}/K_M) under single-turnover conditions (18), as illustrated for S_0 in Figure 2. Cleavage reactions were performed with ^{32}P-labeled S_0 and excess R_0 and monitored by denaturing gel electrophoresis (Figure 2A). The percentage of remaining substrate was plotted against time to derive the value of the first-order rate constant k_{obs} (Figure 2B). For S_{-1bp}, S_0 and S_{+1bp} a linear relationship is observed when k_{obs} is plotted as a function of R_0 concentration, and the values of catalytic efficiency (k_{cat}/K_M) were obtained from the slope of this graph (Figure 2C and Table 1). For all other substrates (S_{-2bp}, S_{+2bp}, S_{+3bp} and S_{+4bp}), the value of k_{obs} is below the detection limit ($\leq 0.001 \, min^{-1}$) at R_0 concentrations up to 5 μM, and therefore the value of k_{cat}/K_M cannot be
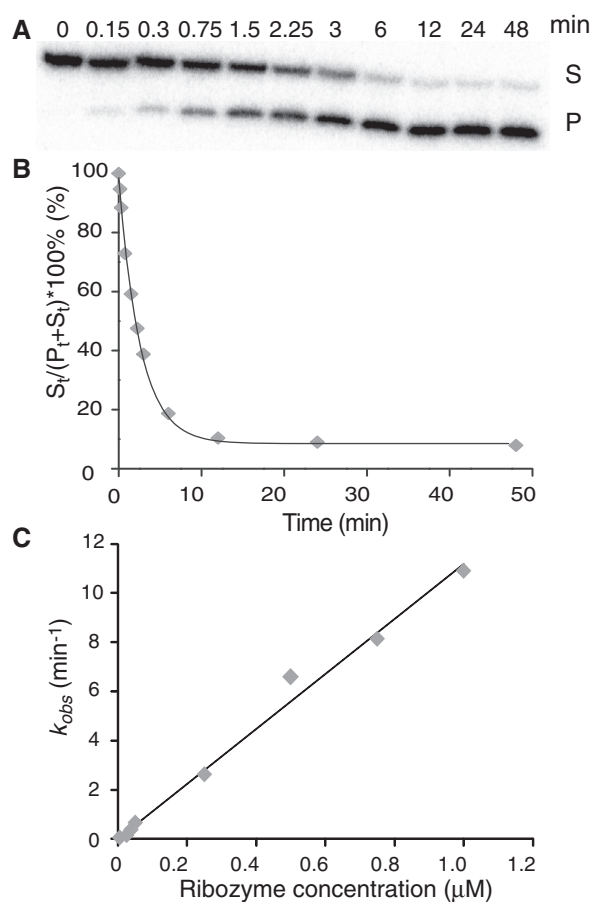
Figure 2. Single-substrate kinetic analysis of R_0. (**A**) Cleavage of $5'$-^{32}P-labeled S_0 by R_0 (at 37.5 nM) monitored by denaturing gel electrophoresis. (**B**) The percentage of remaining substrate was plotted against time. The data were fitted to the non-linear equation $F = Ae^{-(k_{obs} * t)} + F_\infty$, to extract the value of k_{obs} ($k_{obs} = 0.332 \pm 0.008 \, min^{-1}$; $R^2 = 1.00$). (**C**) Cleavage reactions were performed at various concentrations of R_0, and the value of k_{obs} was plotted against ribozyme concentration. For all k_{obs} values, error bars are smaller than the data point on the graph. The data were fitted to a single exponential equation, and the k_{cat}/K_M is extracted by linear regression (k_{cat}/K_M = value of the slope = $11.2 \, min^{-1} \mu M^{-1}$ and $R^2 = 0.99$).

determined (Table 1). To insure that the lack of reactivity of certain substrates is not due to major folding defects, they were analyzed by native gel electrophoresis (Supplementary Figure S1). All substrates, with exception of S_{-2bp}, migrate as a single band on the gel, at a position typical of small RNA hairpins, in agreement with the predicted secondary structure (38). Most of the S_{-2bp} substrate migrates slower, indicating that it adopts a duplex conformation that is incompatible with cleavage by R_0.

For R_0, the highest value of k_{cat}/K_M is obtained with S_0 ($11 \pm 1\,min^{-1}\mu M^{-1}$). Surprisingly, the k_{cat}/K_M values for S_{-1} ($5.2 \pm 0.8\,min^{-1}\mu M^{-1}$) and S_{+1} ($2.7 \pm 0.3\,min^{-1}\mu M^{-1}$) are reduced by only 2- and 4-fold, respectively, with respect to S_0 (Table 1). The similar substrate specificities of R_0 for S_{-1}, S_0 and S_{+1} point to the adaptability of the VS ribozyme architecture to cleave a range of substrates with 3–5 bp in stem Ib, indicating a certain level of promiscuity in substrate selection.

Multiple-substrate kinetic studies

Single-substrate kinetic studies performed with R_0 are fairly tedious and time-consuming, especially given that several time-courses are needed to derive k_{cat}/K_M values for an individual substrate, each k_{cat}/K_M are determined at least three times, and multiple substrates are being investigated. To accelerate the kinetic characterization of the remaining S/R variant pairs (Figure 1B), we set up a multiple-substrate kinetic procedure in which one enzyme is simultaneously incubated with a mixture of substrates (29). Single-turnover conditions were chosen for our studies because the enzyme is in excess with respect to all substrates, and there should be no competition between the various substrates. Given the different sizes of substrates investigated (Figure 1B), the disappearance of any given substrate is easily monitored by denaturing gel electrophoresis (Figure 3). The percentage of remaining substrate can be reliably estimated from the ratio of substrate remaining at time t to that at time zero provided that normalization of input RNA in each lane can be achieved. For normalization, we initially added a non-cleavable DNA control in the cleavage reaction, but later found that quantification is more reliable using the non-cleavable S_{-2bp} RNA that gives a better-defined band on the gel (Figure 3).

The multiple-substrate kinetic procedure was first performed with R_0, revealing essentially identical k_{cat}/K_M values to those obtained with the single-substrate kinetic method (Table 1). These results confirm the reliability of the multiple-substrate procedure and, thus, similar studies were conducted with the variant ribozymes, except for R_{+2bp} because its helix-length complementary substrate S_{-2bp} is not cleavable. Interestingly, substrate cleavage is observed for several S/R pairs, with many having similar k_{cat}/K_M values as the S_0/R_0 pair (Table 2). To verify that substrate cleavage for all functional S/R pairs yields the same product, the reaction products were analyzed by

Table 1. Summary of kinetic data for R_0[a]

	Values of k_{cat}/K_M ($min^{-1}\mu M^{-1}$)	
	Single-substrate kinetics	Multiple-substrate kinetics
S_{-2bp}	n.d.[a]	n.d.
S_{-1bp}	5.2 ± 0.8	5.9 ± 0.7
S_0	11 ± 1	11 ± 2
S_{+1bp}	2.7 ± 0.3	2.7 ± 0.4
S_{+2bp}	n.d.	n.d.
S_{+3bp}	n.d.	n.d.
S_{+4bp}	n.d.	n.d.

[a]The k_{cat}/K_M and their errors (in $min^{-1}\mu M^{-1}$) are average values and standard deviations, respectively, from at least three independent experiments. For several S/R pairs cleavage activity was not detected (n.d.; $k_{obs} \leq 0.001\,min^{-1}$) at $[R] \leq 5\,\mu M$.
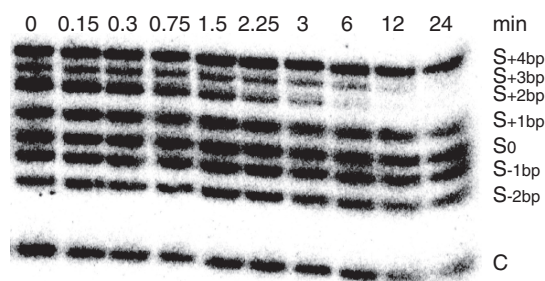

Figure 3. Simultaneous kinetic characterization of R_{-3bp} with multiple substrates. The cleavage reactions of $5'$-^{32}P-labeled substrates (S) by R_{-3bp} (at 100 nM) were monitored by denaturing gel electrophoresis. A DNA hairpin was used as a normalization control (C).

Table 2. Summary of multiple-substrate kinetic data[a]

	R_{-4bp}		R_{-3bp}		R_{-2bp}		R_{-1bp}		R_0		R_{+1bp}	
	k_{cat}/K_M	Ratio	k_{cat}/K_M	Ratio	k_{cat}/K_M	Ratio	k_{cat}/K_M	Ratio	k_{cat}/K_M	Ratio	k_{cat}/K_M	Ratio
S_{-1bp}	n.d.		n.d.		n.d.		0.11 ± 0.02	0.0095	5.9 ± 0.7	0.54	0.13 ± 0.02	0.012
S_0	n.d.		n.d.		0.11 ± 0.02	0.01	12.0 ± 0.9	1.1	11 ± 2	1	0.019 ± 0.002	0.0017
S_{+1bp}	n.d.		0.051 ± 0.006	0.0046	10 ± 1	0.87	14 ± 2	1.3	2.7 ± 0.4	0.25	0.013 ± 0.002	0.0012
S_{+2bp}	0.07 ± 0.01	0.0064	3.1 ± 0.2	0.28	7.4 ± 0.8	0.67	1.9 ± 0.3	0.17	n.d.		n.d.	
S_{+3bp}	0.35 ± 0.04	0.032	0.66 ± 0.09	0.064	0.20 ± 0.05	0.018	n.d.		n.d.		n.d.	
S_{+4bp}	0.040 ± 0.004	0.0036	0.020 ± 0.003	0.0018	n.d.		n.d.		n.d.		n.d.	

[a]The k_{cat}/K_M and their errors (in $min^{-1}\mu M^{-1}$) are average values and standard deviations, respectively, from at least three independent experiments. The ratios correspond to $(k_{cat}/K_M)/(k_{cat}/K_M)_0$, where $(k_{cat}/K_M)_0$ is the value obtained using R_0 and S_0. Data for high-activity S/R pairs with ratios ≥ 0.25 are highlighted in grey. For several S/R pairs cleavage activity was not detected (n.d.; $k_{obs} \leq 0.001\,min^{-1}$) at $[R] \leq 5\,\mu M$.

comparing their migration with alkaline and T1 digests on denaturing gels. The results clearly demonstrate that cleavage occurs between G620 and A621 for all substrates cleaved by either R_0 or variant ribozymes (Supplementary Figure S2). Thus, the helix-length mutations investigated here have no effect on cleavage site selection.

As expected, helix-length compensation is observed for all variant ribozymes, since the highest values of k_{cat}/K_M are observed with the compensatory substrate or one that differs by a single base pair. More precisely, the most active S/R combinations for each ribozyme are: $S_{-1\,bp}/R_{+1\,bp}$, S_0/R_0, $S_{+1\,bp}/R_{-1\,bp}$, $S_{+1\,bp}/R_{-2\,bp}$, $S_{+2\,bp}/R_{-3bp}$ and $S_{+3\,bp}/R_{-4\,bp}$. However, k_{cat}/K_M values similar to that of S_0/R_0 (within 4-fold; shaded in Table 2) are observed only for four of the six ribozymes (R_0, $R_{-1\,bp}$, R_{-2bp} and $R_{-3\,bp}$). For the R_{+1bp} and R_{-4bp} ribozymes, the maximal k_{cat}/K_M values are reduced significantly, ~80- and ~30-fold, respectively, compared to S_0/R_0.

## Structural modeling of SLI/SLV complexes compatible with kinetic data

Catalysis generally requires the precise positioning of chemically active groups within the active site, which in the VS ribozyme results from the precise docking of the cleavage site and A730 internal loops (10,19–27). Thus, it is likely that the cleavage site loop is presented in a favorable position for productive docking in the most active S/R combinations, but not in the inactive or poorly active S/R combinations. To explore how the relative positioning of the cleavage site loop is affected by helix-length mutations, a structural modeling approach was used. We chose to model an SLI/SLV complex as opposed to models of complete S/R pairs to simplify the modeling procedure and limit the structural consequences of helix-length mutations. Furthermore, the SLI/SLV models allow us to examine the effect of helix-length variations on the positioning of the scissile phosphate in SLI relative to a common base pair in SLV. This approach is consistent with our kinetic studies of S/R pairs in which ribozyme variants retain a common core sequence and interact with the free substrates primarily through the I/V kissing-loop interaction (Bouchard, P., Delfosse, V. and Legault, P., unpublished data). We assumed that realistic 3D models would be compatible with the kinetic data, such that the relative positioning of the cleavage site loop in active S/R combinations would be distinct from that found with inactive S/R combinations.

3D models of the SLI/SLV complexes were generated using a stepwise procedure (Supplementary Figure S3). First, several models of the I/V kissing-loop junction core were obtained using previously described constraints (18). Of these core models, 56 were then selected because each could be used to build a common-core group of 36 sub-models that contain all combinations of SLI (S_{-1bp} to $S_{+4\,bp}$) and SLV stems (R_{-4bp} to $R_{+1\,bp}$) and remain consistent with input structural constraints after energy minimization. For all selected sub-models (56×36), the interhelical angle and axial displacement between stems Ib and V were computed and average values were determined for each common-core group. A wide range

of axial displacements ($d_{I-V} = 12$–24 Å), and interhelical angles ($\theta_{I-V} = 108°$ to $175°$; Figure 4) are observed among the common-core groups, pointing to the extensive conformational sampling of the modeling procedure.

We hypothesized that for common-core groups compatible with the kinetic data, the sub-models of the high-activity class $[(k_{cat}/K_M)/(k_{cat}/K_M)_0 \geq 0.25$; Table 2] should be significantly different than those of the low-activity class $[(k_{cat}/K_M)/(k_{cat}/K_M)_0 \leq 0.01$; Table 2] of S/R pairs. In order to compare the structures of the sub-models, pairwise rmsd's were calculated between the sub-models of each common-core group and their reference S_0/R_0 sub-model. For each common-core group, the rmsd distributions for the high-activity and low-activity classes of S/R pairs were compared using a Welch's t-test (36). Interestingly, we observe a good correlation between P-values and interhelical angles (Figure 4), where common-core groups with average interhelical angles (θ_{I-V}) closer to $180°$ show smaller P-values and are thus more likely to be consistent with the kinetic data. Group 13 ($\theta_{I-V} = 166.0° \pm 3.6$) was selected for structural analysis, but similar conclusions are obtained for other common-core groups with P-values smaller than the α threshold value of 0.02 (group 12: $\theta_{I-V} = 165.4° \pm 3.5°$ and group 20: $\theta_{I-V} = 171.4° \pm 4.5°$).

The 3D structure of the S_0/R_0 sub-model from the selected group reveals that, although the helical axes of stems Ib and V are more or less colinear, the two stems are not coaxial (Figure 5A). Rather, there is a significant perpendicular displacement ($d_{I-V} \sim 20$ Å) between these axes that results from the structure of the kissing-loop junction. The spatial relation between these two axes is reflected in the structure of variant S/R pairs, as illustrated
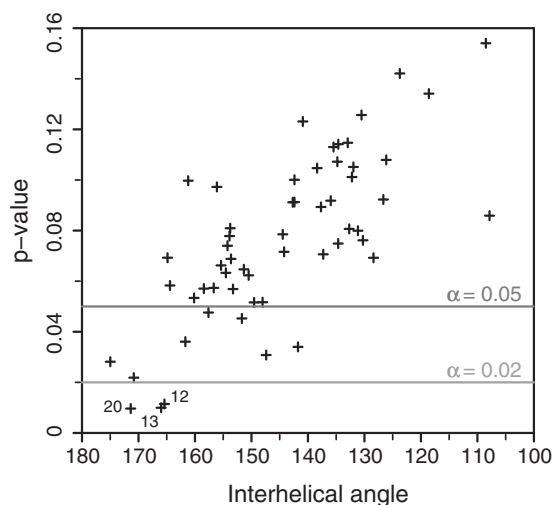
Figure 4. Identification of SLI/SLV models compatible with the kinetic data. Plot of the P-value (Welch's t-test) against the mean interhelical angle. The P-value was obtained from a Welch's t-test that compared the rmsd distributions of the high-activity and low-activity classes of S/R pairs for a given common core group. The mean interhelical angle was obtained by computing the average interhelical angles of the 36 S/R pairs for each common-core group. All common-core groups showing a P-value < 0.02 are numbered. Standard deviations on the interhelical angles vary between 2° and 7°, but were omitted from the graph for clarity.
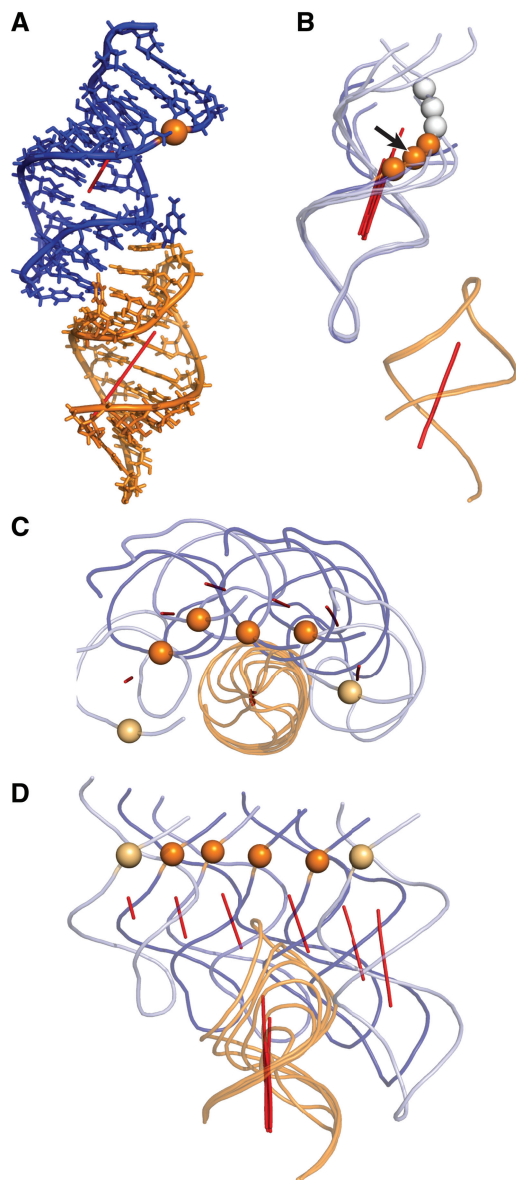
Figure 5. Model of an SLI/SLV complex compatible the kinetic data (common-core group 13). (**A**) Stick representation of the S_0/R_0 sub-model. (**B**) Cartoon representation of a superposition of the five variant S/R_0 sub-models on the S_0/R_0 sub-model. The black arrow points at the scissile phosphate of S_0/R_0. The scissile phosphates of the following substrates are shown from bottom to top: S_{-1}, S_0, S_{+1}, S_{+2}, S_{+3}, S_{+4}. (**C**) Top-down and (**D**) front views of the superposition of the most catalytically efficient sub-models of each ribozyme. The following S/R pairs are shown from left to right: S_{-1}/R_{+1}, S_0/R_0, S_{+1}/R_{-1}, S_{+1}/R_{-2}, S_{+2}/R_{-1}, S_{+3}/R_{-4}. In (B), (C) and (D), the sub-models were superposed to the S_0/R_0 sub-model based on heavy atom alignment of the first base pair of SLV (residues 689 and 707 in S_0/R_0). SLI and SLV are represented in deep blue and orange, respectively. Helical axes are represented as red rods and the phosphorus atom at the cleavage site is shown as a sphere. The phosphorus atoms are color-coded according to the catalytic efficiency of the sub-models $\{[(k_{cat}/K_M)/(k_{cat}/K_M) \geq 0.25]$: deep orange; $[0.25 > (k_{cat}/K_M)/(k_{cat}/K_M) > 0.01]$: light orange and $[(k_{cat}/K_M)/(k_{cat}/K_M) \leq 0.01]$: white}.

in Figure 5B–D. The superposition of all S/R pairs containing R_0 shows that the scissile phosphates of the high-activity substrates (deep orange spheres in Figure 5B) cover a $\sim$7.5 Å distance on one side of stem Ib, whereas those of low-activity substrates (white spheres in Figure 5B) rotate around the helical axis towards a different side of the helix. Thus, this superposition indicates that limited helical displacements of the scissile phosphate are compatible with productive docking of the substrate internal loop with the catalytic domain.

To better understand the helix-length compensation observed between selected S/R pairs, we superposed the S/R pairs that produce maximal cleavage activity for each ribozyme (Figure 5C and D). It is clear from both the top-down (Figure 5C) and front views (Figure 5D) of this superposition that the scissile phosphates are distributed along the arc of a circle. Those of the high-activity S/R pairs (deep orange in Figure 5C and D) span a chord length of $\sim$22 Å. This superposition suggests that large, but restricted displacements of the scissile phosphate along this arc are compatible with productive docking of the substrate internal loop with the catalytic domain.

DISCUSSION

In this study, we explored an unusual type of compensatory mutations in RNA, termed helix-length compensations, using the VS ribozyme as a model system. VS substrate and ribozyme variants were prepared in which the number of base pairs in stems Ib and V were either reduced or increased compared to the reference S_0/R_0 pair. Several S/R combinations were found to be nearly as active as the reference pair, with precise and efficient cleavage obtained for substrate derivatives that contain 3–6 bp in stem Ib. Thus, this work expands the substrate specificity of the VS ribozyme architecture, while revealing its adaptability for efficient substrate cleavage. Two interesting aspects of this adaptability are limited substrate promiscuity with respect to stem Ib length and helix-length compensation between stems Ib and V.

Substrate promiscuity in the VS ribozyme

Substrate cleavage by R_0 occurs most efficiently with a stem Ib containing 3–5 potential base pairs, whereas the catalytic efficiency for substrates with 6–8 bp in stem Ib is reduced at least 50 000-fold compared to S_0. Thus, there is limited substrate promiscuity with respect to stem Ib length for cleavage by R_0, and comparable results are obtained with R_{-1} and R_{-2}. A similar phenomenon has previously been observed for the 5′-splicing reaction of the group I intron ribozyme, where the cleavage site coincides with a G–U wobble pair in the first stem-loop (P1) of the intron (39). Cleavage at the correct site was observed when the G–U wobble pair was inserted within a 3-bp window of the P1 stem (39). Although the cleavage mechanisms of the group I intron and VS ribozymes differ significantly, both require productive docking of a hairpin substrate with the catalytic domain. Thus, for both ribozymes, helical displacement of the scissile phosphate within a

window equivalent to 3 bp is compatible with productive docking and cleavage at the correct site.

Helix-length compensation between stems Ib and V

We initially hypothesized that the reduced cleavage activity of defective stem Ib variants by R_0 could be restored by adjusting the number of base pairs within stem V, such that the total number of base pairs within the two stems would be preserved. Examination of the kinetic data indicates that there is, as predicted, helix-length compensation between stems Ib and V. Indeed, substrates for which cleavage is not detected with R_0 ($S_{+2\,bp}$, $S_{+3\,bp}$, $S_{+4\,bp}$) are cleaved by a ribozyme variant. However, the cleavage activities of the compensatory S/R pairs are not all restored to the reference level. In agreement with these results, 3D modeling of the I/V kissing-loop complex indicates that the global orientation of stems Ib and V significantly affects the position of the scissile phosphate in the variant S/R pairs.

An unexpected aspect of this helix-length compensation study is that it helps provide global structural information on the I/V kissing-loop interaction. From an ensemble of SLI/SLV models that display a wide range of interhelical angles ($108° \leq \theta_{I-V} \leq 175°$) and axial displacements ($d_{I-V} = 12–24\,\text{Å}$), three models were selected that best fit the kinetic data. These selected models display a much smaller range of interhelical angles and axial displacements ($\theta_{I-V} = 165°–171°$ and $d_{I-V} = 18–21\,\text{Å}$), with near colinearity of their helical axes. By comparison, our previously reported model of the SLI/SLV complex (18) exhibits interhelical parameters ($\theta_{I-V} = 106°$ and $d_{I-V} = 16\,\text{Å}$) that are somewhat compatible with the ensemble of models, but not with our selected models. In contrast, the approximate colinearity of stems Ib and V is consistent with a 3D model of the VS ribozyme derived from biochemical and mutational data (14,28), however sufficient constraints were not available to precisely define the interhelical angle between these two stems nor to rule out alternative models (Collins, R.A., personal communication). Similarly, the global model of the VS ribozyme based on fluorescence resonance energy transfer data does not incorporate sufficient constraints, particularly for the SLI/SLV interaction, to allow a fair comparison with our selected models (15,16). Interestingly, the model of the complete VS ribozyme derived from small-angle X-ray scattering (SAXS) shows an interhelical angle close to that of our selected models [$\theta_{I-V} = 154°$ and $d_{I-V} = 25\,\text{Å}$; (17)]. This similarity between our selected models and the SAXS model is particularly surprising given that the folding of the complete ribozyme was not accounted for in our modeling studies. In addition, the SAXS model includes helix VII that is not present in the VS ribozyme used for our kinetic study. The small difference in axial displacement between our selected models and the SAXS model likely reflects structural differences at the kissing-loop junction. Nevertheless, this comparison with the SAXS model suggests that the combination of kinetic data from helix-length variants with 3D modeling represents a potentially powerful approach to extract interhelical parameters.

Adaptability of the VS ribozyme architecture for substrate cleavage

Given the intrinsic adaptability of the VS ribozymes architecture for substrate cleavage, the modeling results give an appreciation of the level of structural adaptation that is likely at stake for productive docking of the SLI cleavage site internal loop with the A730 loop of SLVI. However, these models do not explain how high-activity S/R pairs move relative to the wild-type S/R pair to allow formation of the active site and efficient cleavage.

An emerging view in RNA structure formation is that tertiary interactions form as a result of conformational sampling of helical domains within a distribution of conformations that is limited, in part, by the inherent flexibility of helical connectivities [for recent reviews see (40,41)]. It is likely that in the VS ribozyme, flexible elements such as its bulges, internal loops, three-way junctions and even its I/V kissing-loop junction (42,43) allow conformational sampling of helical domains, including SLI and SLVI, through which a well-defined active site can emerge. According to this view, the ability to sample the active conformation may be conserved in high-activity S/R pairs but not in those with lower activity. Future investigations are needed to help identify the dynamic features responsible for the conformational adaptability of the VS ribozyme and determine if a similar adaptability exists in other functional RNAs.

Prospects for RNA research

Helix-length compensation represents an alternative approach for investigating the sequence and structural repertoire of functional RNAs that could be applied to other helix pairs. In principle, helix-length compensation could be observed for helices that are more or less coaxially stacked (1–6) or those that are not coaxial but nearly colinear as observed for the VS ribozyme. Although helix-length compensations have been previously identified by comparative sequence analysis (1–6), such analysis may not identify all possible helix-length mutants. Indeed, the genetic mechanism through which such complex mutation pattern evolves likely requires multiple mutational events with an overall lower probability than simple covariation of base-pairing residues. Moreover, sequence searches may not always take into account the possibility for helix-length compensation. Systematic helix-length compensation studies could help improve the consensus sequence and structural model for a given functional RNA. Subsequently, revised RNA consensus sequence could be used for bioinformatics database search of additional representatives of known functional RNAs (44–47). Helix-length compensation studies could also help reveal the inherent adaptability of other complex RNA architectures and be useful for the design of RNAs with novel functions.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures S1–S3.

REFERENCES

1. Woese,C.R., Gutell,R., Gupta,R. and Noller,H.F. (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol. Rev.*, **47**, 621–669.
2. Gutell,R.R. (1995) In: Zimmermann,R.A. and Dahlberg,A.E. (eds), *Ribosomal RNA: Structure, Evolution, Processing, and Function in Protein Biosynthesis*. CRC Press, Boca Raton, Florida, pp. 109–126.
3. Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
4. 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.
5. Adams,P.L., Stahley,M.R., Kosek,A.B., Wang,J. and Strobel,S.A. (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature*, **430**, 45–50.
6. Golden,B.L., Kim,H. and Chase,E. (2004) Crystal structure of a phage Twort group I ribozyme-product complex. *Nat. Struct. Mol. Biol.*, **12**, 82–88.
7. Steinberg,S., Leclerc,F. and Cedergren,R. (1997) Structural rules and conformational compensations in the tRNA L-form. *J. Mol. Biol.*, **266**, 269–282.
8. Collins,R.A. (2002) The *Neurospora* Varkud satellite ribozyme. *Bioch. Soc. Trans. Rev.*, **30**, 1122–1126.
9. Lilley,D.M. (2004) The Varkud satellite ribozyme. *RNA*, **10**, 151–158.
10. Wilson,T.J. and Lilley,D.M. (2011) Do the hairpin and VS ribozymes share a common catalytic mechanism based on general acid-base catalysis? A critical assessment of available experimental data. *RNA*, **17**, 213–221.
11. Rastogi,T., Beattie,T.L., Olive,J.E. and Collins,R.A. (1996) A long-range pseudoknot is required for activity of the *Neurospora* VS ribozyme. *EMBO J.*, **15**, 2820–2825.
12. Andersen,A. and Collins,R.A. (2000) Rearrangement of a stable RNA secondary structure during VS ribozyme catalysis. *Mol. Cell*, **5**, 469–478.
13. Andersen,A.A. and Collins,R.A. (2001) Intramolecular secondary structure rearrangement by the kissing interaction of the *Neurospora* VS ribozyme. *Proc. Natl Acad. Sci. USA*, **98**, 7730–7735.
14. Hiley,S.L. and Collins,R.A. (2001) Rapid formation of a solvent-inaccessible core in the *Neurospora* Varkud satellite ribozyme. *EMBO J.*, **20**, 5461–5469.
15. Lafontaine,D.A., Norman,D.G. and Lilley,D.M. (2001) Structure, folding and activity of the VS ribozyme: importance of the 2-3-6 helical junction. *EMBO J.*, **20**, 1415–1424.
16. Lafontaine,D.A., Norman,D.G. and Lilley,D.M. (2002) The global structure of the VS ribozyme. *EMBO J.*, **21**, 2461–2471.
17. Lipfert,J., Ouellet,J., Norman,D.G., Doniach,S. and Lilley,D.M. (2008) The complete VS ribozyme in solution studied by small-angle X-ray scattering. *Structure*, **16**, 1357–1367.
18. Bouchard,P., Lacroix-Labonté,J., Desjardins,G., Lampron,P., Lisi,V., Lemieux,S., Major,F. and Legault,P. (2008) Role of SLV in SLI substrate recognition by the *Neurospora* VS ribozyme. *RNA*, **14**, 736–748.
19. Lafontaine,D.A., Wilson,T.J., Norman,D.G. and Lilley,D.M. (2001) The A730 loop is an important component of the active site of the VS ribozyme. *J. Mol. Biol.*, **312**, 663–674.
20. Lafontaine,D.A., Wilson,T.J., Zhao,Z.-Y. and Lilley,D.M.J. (2002) Functional group requirements in the probable active site of the VS ribozyme. *J. Mol. Biol.*, **323**, 23–34.
21. Sood,V.D. and Collins,R.A. (2002) Identification of the catalytic subdomain of the VS ribozyme and evidence for remarkable sequence tolerance in the active site loop. *J. Mol. Biol.*, **320**, 443–454.
22. Jones,F.D. and Strobel,S.A. (2003) Ionization of a critical adenosine residue in the *Neurospora* Varkud satellite ribozyme active site. *Biochemistry*, **42**, 4265–4276.
23. Zhao,Z.Y., McLeod,A., Harusawa,S., Araki,L., Yamaguchi,M., Kurihara,T. and Lilley,D.M. (2005) Nucleobase participation in ribozyme catalysis. *J. Am. Chem. Soc.*, **127**, 5026–5027.
24. Wilson,T.J., McLeod,A.C. and Lilley,D.M. (2007) A guanine nucleobase important for catalysis by the VS ribozyme. *EMBO J.*, **26**, 2489–2500.
25. Smith,M.D. and Collins,R.A. (2007) Evidence for proton transfer in the rate-limiting step of a fast-cleaving Varkud satellite ribozyme. *Proc. Natl Acad. Sci. USA*, **104**, 5818–5823.
26. Jaikaran,D., Smith,M.D., Mehdizadeh,R., Olive,J. and Collins,R.A. (2008) An important role of G638 in the cis-cleavage reaction of the *Neurospora* VS ribozyme revealed by a novel nucleotide analog incorporation method. *RNA*, **14**, 938–949.
27. Wilson,T.J., Li,N.S., Lu,J., Frederiksen,J.K., Piccirilli,J.A. and Lilley,D.M. (2010) Nucleobase-mediated general acid-base catalysis in the Varkud satellite ribozyme. *Proc. Natl Acad. Sci. USA*, **107**, 11751–11756.
28. Rastogi,T. and Collins,R.A. (1998) Smaller, faster ribozymes reveal the catalytic core of *Neurospora* VS RNA. *J. Mol. Biol.*, **277**, 215–224.
29. Guo,H.C.T. and Collins,R.A. (1995) Efficient trans-cleavage of a stem-loop RNA substrate by a ribozyme derived from *Neurospora* VS RNA. *EMBO J.*, **14**, 368–376.
30. Major,F., Turcotte,M., Gautheret,D., Lapalme,G., Fillion,E. and Cedergren,R. (1991) The combination of symbolic and numerical computation for three-dimensional modeling of RNA. *Science*, **253**, 1255–1260.
31. Campbell,D.O., Bouchard,P., Desjardins,G. and Legault,P. (2006) NMR structure of Varkud satellite ribozyme stem-loop V in the presence of magnesium ions and localization of metal-binding sites. *Biochemistry*, **45**, 10591–10605.
32. Hoffmann,B., Mitchell,G.T., Gendron,P., Major,F., Andersen,A.A., Collins,R.A. and Legault,P. (2003) NMR structure of the active conformation of the Varkud satellite ribozyme cleavage site. *Proc. Natl Acad. Sci. USA*, **100**, 7003–7008.
33. Pearlman,D.A., Case,D.A., Caldwell,J.W., Ross,W.S., Cheatham,T.E., Debolt,S., Ferguson,D., Seibel,G. and Kollman,P. (1995) Amber, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.*, **91**, 1–41.
34. Still,W.C., Tempczyk,A., Hawley,R.C. and Hendrickson,T. (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, **112**, 6127–6129.
35. Lavery,R., Moakher,M., Maddocks,J.H., Petkeviciute,D. and Zakrzewska,K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.*, **37**, 5917–5929.
36. Satterthwaite,F.E. (1946) An approximate distribution of estimates of variance components. *Biometrics*, **2**, 110–114.
37. Shapiro,S.S. and Wilk,M.B. (1965) An analysis of variance test for normality (complete samples). *Biometrika*, **52**, 591–611.

38. Beattie,T.L., Olive,J.E. and Collins,R.A. (1995) A secondary-structure model for the self-cleaving region of *Neurospora* VS RNA. *Proc. Natl Acad. Sci. USA*, **92**, 4686–4690.

39. Doudna,J.A., Cormack,B.P. and Szostak,J.W. (1989) RNA structure, not sequence, determines the 5′ splice-site specificity of a group I intron. *Proc. Natl Acad. Sci. USA*, **86**, 7402–7406.

40. Woodson,S.A. (2010) Compact intermediates in RNA folding. *Annu. Rev. Biophys.*, **39**, 61–77.

41. Bailor,M.H., Mustoe,A.M., Brooks,C.L. 3rd and Al-Hashimi,H.M. (2011) Topological constraints: using RNA secondary structure to model 3D conformation, folding pathways, and dynamic adaptation. *Curr. Opin. Struct. Biol*, **21**, 296–305.

42. Szewczak,A.A. and Cech,T.R. (1997) An RNA internal loop acts as a hinge to facilitate ribozyme folding and catalysis. *RNA*, **3**, 838–849.

43. Bailor,M.H., Sun,X. and Al-Hashimi,H.M. (2010) Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science*, **327**, 202–206.

44. Ferbeyre,G., Smith,J.M. and Cedergren,R. (1998) Schistosome satellite DNA encodes active hammerhead ribozymes. *Mol. Cell. Biol.*, **18**, 3880–3888.

45. Gräf,S., Przybilski,R., Steger,G. and Hamman,C. (2005) A database search for hammerhead ribozyme motifs. *Biochem. Soc. Trans.*, **33**, 477–478.

46. Hammann,C. and Westhof,E. (2007) Searching genomes for ribozymes and riboswitches. *Genome Biol.*, **8**, 210.

47. McCown,P.J., Roth,A. and Breaker,R.R. (2011) An expanded collection and refined consensus model of glmS ribozymes. *RNA*, **17**, 728–736.