A multi-axial RNA joint with a large range of motion promotes sampling of an active ribozyme conformation

Nicolas Girard, Pierre Dagenais, Julie Lacroix-Labonté and Pascale Legault*

Département de biochimie et médecine moléculaire, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, QC, H3C 3J7, Canada

ABSTRACT

Investigating the dynamics of structural elements in functional RNAs is important to better understand their mechanism and for engineering RNAs with novel functions. Previously, we performed rational engineering studies with the Varkud satellite (VS) ribozyme and switched its specificity toward non-natural hairpin substrates through modification of a critical kissing-loop interaction (KLI). We identified functional VS ribozyme variants with surrogate KLIs (ribosomal RNA L88/L22 and human immunodeficiency virus-1 TAR/TAR*), but they displayed ∼100-fold lower cleavage activity. Here, we characterized the dynamics of KLIs to correlate dynamic properties with function and improve the activity of designer ribozymes. Using temperature replica exchange molecular dynamics, we determined that the natural KLI in the VS ribozyme supports conformational sampling of its closed and active state, whereas the surrogate KLIs display more restricted motions. Based on *in vitro* selection, the cleavage activity of a VS ribozyme variant with the TAR/TAR* KLI could be markedly improved by partly destabilizing the KLI but increasing conformation sampling. We formulated a mechanistic model for substrate binding in which the KLI dynamics contribute to formation of the active site. Our model supports the modular nature of RNA in which subdomain structure and dynamics contribute to define the thermodynamics and kinetics relevant to RNA function.

INTRODUCTION

RNA molecules adopt complex 3D structures to fulfill essential function in diverse cellular processes. RNA architecture is often recognized as being highly modular because RNA molecules contain a diverse but limited set of recurrent structural motifs within their secondary and tertiary structural elements, such as hairpin loops, multi-way junctions and kissing-loop interactions (KLIs) (1). RNA folding typically follows a hierarchical process, where secondary structure elements form first followed by compaction through tertiary contacts (2,3). As part of the RNA secondary structure, paired regions form helical structures and unpaired regions adopt specific structural motifs, which impart orientations to the associated helical domains and contribute to forming tertiary contacts (4). It is becoming clear that local structural elements have inherent flexibility that allows them to explore conformational space. Consequently, each structural element should not simply be viewed as a static structure but rather as a sub-ensemble of structures with specific dynamic properties (4–10). However, despite our growing knowledge of how structural motifs define the 3D structure of RNA and increased appreciation for the importance of RNA dynamics, much remains to be known regarding the dynamics that are inherent to specific structural motifs and how they contribute to RNA structure and function.

The modular and hierarchical organization of RNA structure has been made evident through successful efforts in RNA engineering, an emerging field of research that holds tremendous potential for applications in synthetic biology, nanotechnology and molecular medicine (11,12). Novel RNA function can be created by *de novo* design of self-assembled nanostructures or incorporation of structural elements into known RNA architectures (11–14). Our laboratory is exploring the latter approach to modify the substrate specificity of the *Neurospora* Varkud satellite (VS) ribozyme (15,16). The wild-type VS ribozyme specifically recognizes and cleaves its folded hairpin substrate, and we are exploring the potential of this ribozyme to cleave other RNA hairpins, since such structural element regulates a wide variety of cellular and viral RNAs.

The VS ribozyme is a member of the small self-cleaving ribozyme family, which also includes the *glmS*, hammerhead, hairpin, HDV, hatchet, pistol, twister and twister sister ribozymes (17–19). Like many other self-cleaving ribozymes,

---
*To whom correspondence should be addressed. Tel: +514 343 7326; Fax: +514 343 2210; Email: pascale.legault@umontreal.ca

it represents a good model system for structure, function and engineering studies because of its relatively small size and its ability to cleave its substrates either in *cis* or in *trans* (20,21). The minimal catalytic sequence is comprised of ∼160 nt that form six paired domains. The *trans* ribozyme (domains II-VI) recognizes its hairpin substrate (stem-loop I or SLI) by way of a KLI, which forms between loop I and loop V (Figure 1A), and this KLI is crucial for efficient cleavage activity (22). The I/V KLI is highly stable and allows formation of the open state of the ribozyme, which is likely the predominant conformation in solution (Figure 1B) (23–26). Subsequently, the G_{638} loop, known as the cleavage site internal loop of the SLI substrate, intimately associate with the A_{756} loop to form the active site within the closed state of the ribozyme (Figure 1A and B) (27) The cleavage mechanism has been thoroughly investigated and involves the N1 atoms of guanine (G_{638}) and adenine (A_{756}) nucleobases as general base and acid, respectively (18). However, the dynamic processes associated with formation of the active site and involving the transition from the open state to the closed state are yet to be examined.

In recent rational engineering studies (16), we substituted the I/V KLI by two other KLIs with similar thermodynamic and structural properties (Figure 1C): the L88/L22 KLI from the large subunit of the *Deinococcus radiodurans* ribosomal RNA and the synthetic TAR/TAR* KLI related to the human immunodeficiency virus (HIV-1). The KLI substitutions were successful since the overall structure of the ribozyme remained intact as judged from SHAPE analysis and substantial levels of cleavage activity were obtained for variant substrate/ribozyme (S/R) complexes that incorporate either the L88/L22 KLI (S_{L88}/R_{L22}) or TAR/TAR* KLI (S_{TAR}/R_{TAR*}). However, the k_{obs} values for the most active variant S/R complexes (S_{L88+1}/R_{L22-1} and S_{TAR-1}/R_{TAR*-0}) were 50- to 160-fold lower than that of the reference S_0/R_0 system, even though the catalytic activities of the variant ribozymes were optimized by fine-tuning the lengths of stems Ib and V (Figure 1C). These kinetic results suggest that the transition from the open state to the closed state is rate-limiting in these KLI variants of the VS ribozyme.

In this work, we investigated the dynamic properties of three kissing-loops (KL) complexes, the I/V, L88/L22 and TAR/TAR* KL complexes, to correlate the dynamics of KLIs with ribozyme function. We ran 50-ns unrestrained Temperature-Replica Exchange Molecular Dynamics (T-REMD), an enhanced atomistic sampling method that has been shown to better explore conformational space than standard molecular dynamics (MD) and that is well suited to investigate ensembles of RNA populations (28). We analyzed the resulting trajectories to characterize the intrinsic dynamics of these KLIs and establish how the dynamics may contribute to the formation of a closed state for substrate cleavage in the VS ribozyme. The T-REMD simulations support the hypothesis that the reduced cleavage activity observed with our KLI variants of the VS ribozyme correlates with restricted motion of their KLI. Furthermore, to improve the cleavage activity of one of the S_{TAR}/R_{TAR*}

ribozyme variants, we performed *in vitro* selection and identified an S_{TAR}/R_{TAR*} variant with both increased dynamics at the KLI and markedly improved cleavage activity. Based on our results, we formulated a mechanistic model of substrate binding in which the KLI dynamics contributes to the probability of aligning the two key internal loops to form the active site.

MATERIALS AND METHODS

Computational tools and hardware

All system minimization, equilibration and MD simulations were carried out in GROMACS version 2016.3 (29,30) on Compute Canada's Graham GPU server. For the equilibration and MD simulations, the Amber ff99SB (31) force field was used with the ParmBSC0 nucleic acid parameters (32,33) and chi_ol corrections for glycosidic torsion angle chi (34) using explicit aqueous solvent. The Particle Mesh Ewald method was used to handle electrostatic interactions (35,36). GROMACS tools were employed for processing and analysis of T-REMD trajectories as described in Supplementary Materials and Methods. MC-Sym 3.3.2 (37) and Pymol (The PyMOL Molecular Graphics System, Version 1.9.0 Schrödinger, LLC.) were used for system building and/or structural analysis.

RNA system building and preparation

Three RNA systems were initially built for molecular dynamic studies: the I/V, L88/L22 and TAR/TAR* KL complexes. The starting conformation of the L88/L22 KL complex was obtained from PDB ID: 4IOA [chain A; (38)]. For the other two KL complexes, the starting conformation was obtained by extracting the lowest-energy structure of their representative structural ensembles, namely the second model of PDB ID: 2MI0 (24) and 1KIS (39). The hairpin stems within these three KL complexes were substituted by standard A-form stems generated with MC-Sym 3.3.2 (37) through heavy-atom superposition of the first Watson–Crick base pair following the loop-closing base pair. The resulting hybrid structures (Supplementary Figure S1a) retained the hairpin loop residues and closing base-pair [2MI0: residues 6–16 (I) and 108-114 (V); 1KIS: residues 5–12 (TAR) and 21–28 (TAR*) and 4IOA: residues 423–430 (L22) and 2383–2392 (L88)]. Each RNA complex was solvated with ∼20 000 TIP3P water molecules (40) in a periodic rhombic dodecahedral box and neutralized by addition of K^+ placed at randomized positions. In addition, K^+ and Cl^- ions equivalent to a concentration of 0.1 M KCl were added to the solvent (41). For simplicity, the addition of Mg^{2+} ions was not considered here. In addition, it was previously shown to have no significant effect on the global dynamics of the I/V KLI (42). For T-REMD of the TAR/TAR-S1 KL complex derived from *in vitro* selection (Figure 7), the RNA system was built by taking the hybrid structure of the TAR/TAR* KL complex, replacing residues of TAR* such that they match the TAR-S1 sequence (Figure 7D) and modeling the four potential WC base pairs between the TAR and TAR-S1 loops. For RNA
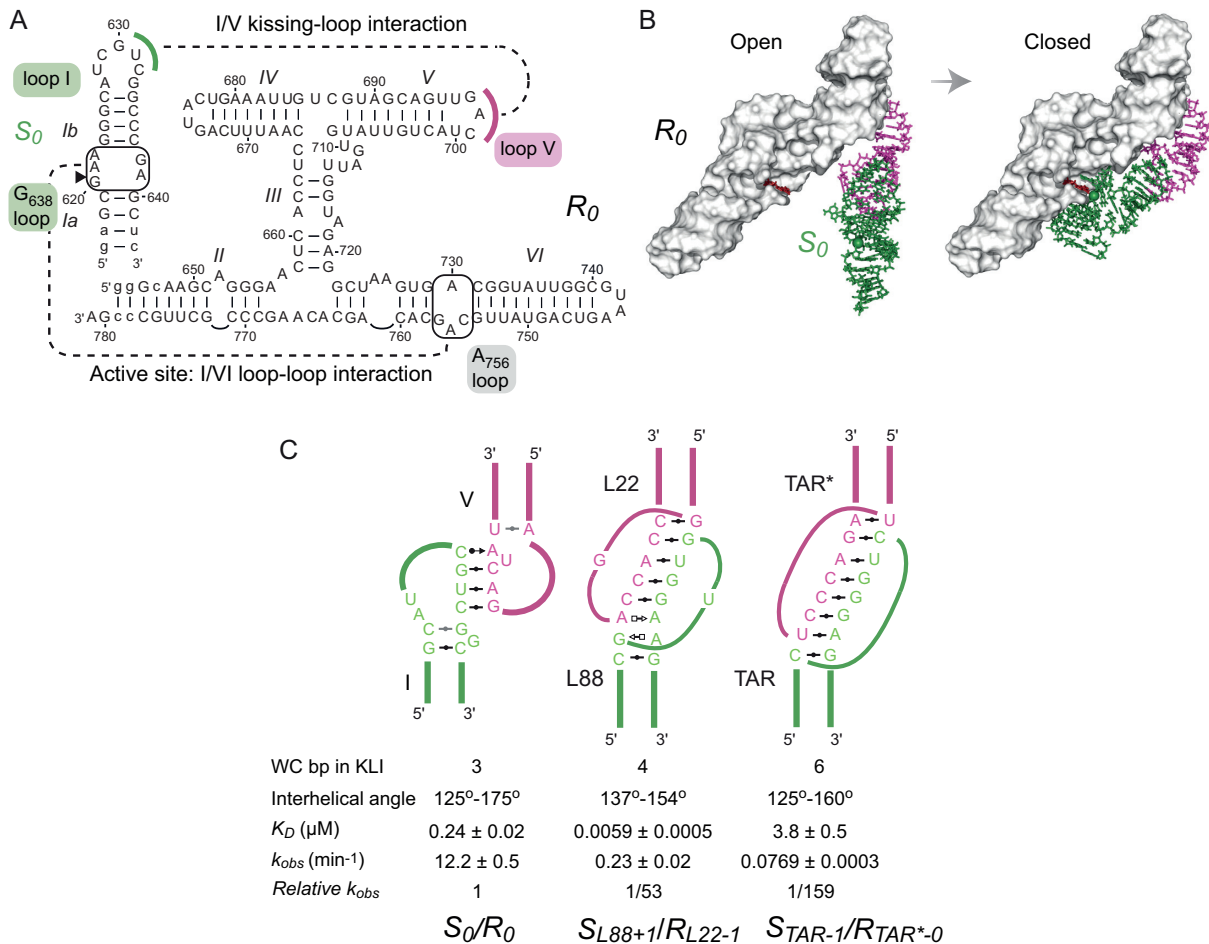
Figure 1. Substitution of the I/V KLI in the *Neurospora* VS ribozyme. (**A**) Primary and secondary structures of the parental *trans* VS ribozyme (R_0) and substrate (S_0). Substrate recognition involves a KLI between loop I and loop V as well as a loop/loop interaction between the G638 loop and the A756 loop, which defines the active site. (**B**) Representation of the open and closed states of the VS ribozyme. The closed state (right panel) is taken from the crystal structure of the *Neurospora* VS ribozyme substrate/ribozyme complex with SLI (green) and the SLV (pink) in sticks, helical domains II, III, IV and VI as a white surface, A_{756} in red sticks and the scissile phosphate as a green sphere. A single representative of the many conformations that likely populate the open state (left panel) is similarly illustrated, but with the I/V KLI from the crystal structure replaced by the I/V KLI from the NMR structure of the SLI/SLV complex (24). (**C**) The I/V KLI of the S_0/R_0 ribozyme complex was previously substituted by the L88/L22 KLI from the rRNA large subunit of *Deinococcus radiodurans* and by the HIV-1 TAR/TAR* KLI (16). Compared to the VS ribozyme I/V KLI, these other KLIs also display a small number of WC base pairs at their interface (WC bp in KLI), have similar interhelical angles and are thermodynamically stable as judged from their dissociation constant [K_D; (16)]. The k_{obs} values of the resulting VS ribozyme variants are given for the parental S_0/R_0 pair as well as for the S_{L88+1}/R_{L22-1} and the S_{TAR-1}/R_{TAR*-0} variants (16). Residues at the KLI, either free or involved in KL base pairs, are shown together with the loop closing base pairs.

system preparation, each RNA system was then copied several times, and each copy was energy minimized and equilibrated at the target temperature and pressure using standard methods (see Supplementary Materials and Methods).

Production T-REMD simulations

The T-REMD simulations were run in several replicas at evenly distributed temperatures from 300 to 375 K (inclusive bounds), to allow enhanced sampling of the conformational space. The number and temperature of the replicas were automatically generated for each KL complex to ensure an exchange rate of ∼20% between adjacent replicas within the temperature scale (43). This resulted in 58, 60 and 60 replicas for the I/V, L88/L22 and TAR/TAR* KL complexes, respectively. The T-REMD simulations were run for 50 ns with 2-fs time steps, attempting exchange between

replicas at each 500 steps and recording a frame for each replica every 200 fs.

Convergence assessment on demultiplexed trajectories

To assess the convergence of the MD simulations (44), the T-REMD trajectories were first demultiplexed to reconstruct continuous trajectories, each one following a single starting structure through the temperature space. An RMSD profile was then built for each demultiplexed trajectory by calculating the pairwise heavy-atom RMSD of each frame with respect to the starting structure.

Computation of Euler angles between the two stems of the KL complexes

Each frame of the T-REMD trajectories at 300 K was analyzed to derive the three Euler angles α, β and γ between

the two stems of the KL complexes. The Euler angles were calculated by adapting a method previously developed for RNA interhelical junctions (45). The interhelical angle θ was derived by taking the supplementary angle of the β Euler angle. For comparative analysis, the α, β, γ and θ angles were also calculated for the VS ribozyme crystal structure [PDB ID: 4R4P (27)].

Analysis of the T-REMD trajectories in the context of the crystal structure of the VS ribozyme

To investigate the T-REMD trajectory at 300 K of a given KL complex in the context of the crystal structure of the VS ribozyme, composite models were built that merged the KL interaction of an individual T-REMD frame with a substrate/ribozyme complex, termed S_X/R_X, taken from the crystal structure of the VS ribozyme $A_{756}G$ variant [PDB ID: 4R4P (27)], in which S_X is the SLI substrate of one protomer and R_X the helical domains II–VI of the other protomer. First, each frame of the T-REMD trajectories was aligned on the crystal structure by superposing the four core base pairs of stem V (residues 685–688 and 698–701 in the crystal structure; residues 5–8 and 18–21 in the I/V KL complex; residues 5–8 and 19–22 in the L88/L22 KL complex and residues 5–8 and 19–22 in the TAR/TAR* KL complex). Subsequently, the internal loop and stem Ia of SLI from the crystal structure was added to the model, superposing stem Ib from the crystal structure (residues 623–625 and 635–637) on the corresponding residues of the I/V, L88/L22 and TAR/TAR* KL complexes (residues 4–6 and 18–20, 4–6 and 17–19 and 4–6 and 17–19, respectively). From these superpositions, d_{PP} values were calculated by measuring the distance between phosphorus atoms of the scissile phosphate from the crystal structure and the composite model derived from a given frame of the T-REMD trajectory. Based on the same superpositions, heavy atom RMSD values were calculated between the G_{638} loop and closing base pairs (residues 619–623 and 637–640) of the crystal structure and the frame-derived model.

Principal component analysis of the T-REMD trajectories to characterize the molecular mechanics of KL complexes

A principal component (PC) analysis was performed on the T-REMD trajectories at 300 K to extract the PCs of motions and evaluate their contributions to the overall dynamics of the KL complexes. For each PC, the two extreme projections along the trajectory were reported on the average structure and interpolating frames were calculated to obtain a pseudo-trajectory. The interpolated frames of the pseudo-trajectory were represented in the context of the VS ribozyme crystal structure by building composite models as described above for frames of the T-REMD trajectories.

Root mean square fluctuation analysis of the T-REMD trajectories and identification of the residues responsible for the dynamics of the KL complexes

To capture the dynamics of individual nucleotides with respect to their immediate neighbors in the sequence, the heavy-atom Root-Mean Square Fluctuation (RMSF) was calculated for each pair of consecutive nucleotides. From the two RMSF values calculated for each non-terminal nucleotide, the largest one was retained.

In vitro selection procedure

All DNA sequences (Supplementary Materials and Methods) were obtained from IDT (Integrated DNA Technologies, IA) and gel purified. Double-stranded DNA libraries were generated by polymerase chain reaction (PCR) using 5′- and 3′-primers and the Pfu DNA polymerase, extracted with phenol/chloroform and precipitated with 1/10 volume NaAc 3 M pH 5.3 and 3 volumes ethanol 95%. RNA libraries were prepared by *in vitro* transcription using T7 RNA polymerase in Transcription buffer (40 mM Tris pH 7.6, 50 mM dithiothreitol, 1 mM spermidine, 0.1% triton, 7.5 mM $MgCl_2$ and 4 mM of each adenosine triphosphate, cytidine triphosphate, uridine triphosphate and guanosine triphosphate) and incubation for 90 min at 37°C. To improve ribozyme cleavage of the active population, the reaction was incubated for an additional 1–20 min (20, 10, 10, 5, 2 and 1 min for cycles 1, 2, 3, 4, 5 and 6, respectively) after addition of 17.5 mM $MgCl_2$. The large cleavage product was purified by denaturing gel electrophoresis [10% acrylamide:bisacrylamide (19:1) and 7 M urea]. The RNA product was then reverse transcribed using Superscript III reverse transcriptase with the RT-primer, according to the manufacturer's protocol (Invitrogen, CA). The next rounds of selection were performed by following the same protocol, starting by regenerating the double-stranded DNA library, as described above.

Next-generation sequencing and sequence analysis of selected ribozymes

PCR products from the sixth round of selection were sequenced by Illumina MiSeq Spike (McGill University and Génome Québec Innovation Centre, Montréal, Canada) and analyzed using tools from the Galaxy Project (46). Briefly, the raw fastq output was first filtered using a Phred quality score of 30, such that 141 511 reads were kept. The dataset was first analyzed to verify that there were no mutations within the ribozyme outside the randomized region, which was found to be the case for all sequences. The ribozyme sequences were then truncated, keeping the eight bases flanking both sides of the randomized region, and the data was processed to obtain each unique sequence and its number of reads. Sequences with 10+ reads were kept and analyzed in terms of their potential for contiguous Watson–Crick (WC) base pairing with loop I. Then, the sequences that could formed at least four base pairs with loop I (96.4% of sequences) were analyzed to identify the positions of loop V residues that could form the four base pairs, and from this analysis two main classes were identified (Figure 7). For each class, position weight matrices were generated by counting the frequency of each base at every position within the randomized region and sequence logo representations were generated using a modified version of SeqLogo (47).

Single-turnover kinetic studies of the S_{TAR}/R_{TAR-S1} complex

The S_{TAR}/R_{TAR-S1} complex is a derivative of the S_0/R_0 complex previously characterized kinetically (15,16), modified in stem Ia and stem-loop Ib to match the HIV-1 TAR RNA sequence (S_{TAR}: 5'-GCA GAG AAG AGC CUG GGA GCU CGA UCU G-3') and in stem-loop V to match the selected R_{TAR-S1} sequence (Figure 7D). The 5'-end-^{32}P-labeled S_{TAR} and unlabeled *trans* ribozymes *(R_0, R_{TAR*-0}* and *R_{TAR-S1})* were prepared using established procedures (16). Single-turnover kinetic studies with the S_{TAR}/R_{TAR-S1}, S_{TAR}/R_0 and S_{TAR}/R_{TAR*-0} complexes were carried out at 37°C in Kinetic Reaction buffer (50 mM Tris–HCl pH 7.6, 25 mM KCl, 2 mM spermidine and 25 mM MgCl$_2$) to determine the values of the observed rate constant (k_{obs}) using a substrate concentration ([S]) of 250 pM and a ribozyme concentration ([R]) of 0.5 μM, as previously described (16).

Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) studies, including RNA preparation, data collection and analysis, were performed as previously described (23) with relevant experimental details provided in Supplementary Figure S8.

RESULTS AND DISCUSSION

T-REMD simulations of kissing-loop complexes

We ran 50-ns T-REMD atomistic simulations to characterize the structural dynamics of three KL complexes in explicit solvent (Supplementary Figure S1). The T-REMD approach was used to enhance conformational sampling; MD simulations were run in parallel for ∼60 replicas at evenly-distributed temperatures between 300 and 375 K, with periodic attempts to exchange RNA coordinates between replicas of adjacent temperatures. Thus, the conformational space was sampled for a total MD simulation time of ∼3 μs for each KL complex.

To ensure that the T-REMD protocol allowed for a thorough exploration of the conformational space, the convergence of the demultiplexed trajectories was assessed (44). The demultiplexed trajectories were first obtained by following the continuous trajectory of each starting structure through the temperature space, and for each demultiplexed trajectory an RMSD frequency profile was generated based on heavy atom RMSDs between each frame and the T-REMD starting structure. The resulting RMSD profiles are all very similar with only a few outliers for the I/V KL complex (Supplementary Figure S2a), as well as for the L88/L22 and TAR/TAR* KL complexes (Supplementary Figure S2bc). Moreover, the detection of a single peak in the overlaid profiles indicates that all of the conformations are very similar in terms of RMSD, and that the T-REMD protocol did not induce large structural changes. Thus, the T-REMD trajectories of the I/V, L88/L22 and TAR/TAR* KL complexes converge reasonably well, and this validates the analysis of their structural dynamics.

Basic structural analysis of T-REMD trajectories

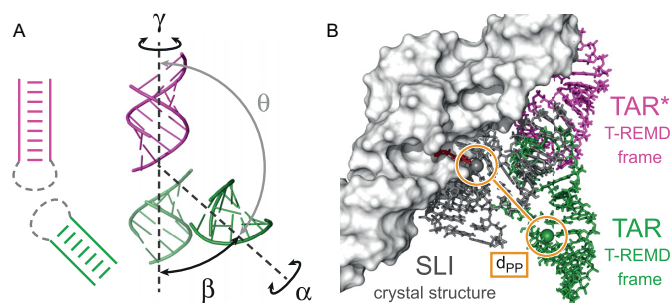For each KL complex, the 250 000 frames of the lowest-temperature T-REMD trajectory (300 K) were structurally

Figure 2. Schematic representation of the parameters used to analyze the T-REMD trajectories. (**A**) The Euler angles (α, β and γ) were defined from the KL complexes as shown and the interhelical angle θ is calculated as $180° - |\beta|$. (**B**) Each T-REMD frame of the trajectories (SLV in pink sticks and SLI in green sticks) was analyzed in the context of the VS ribozyme crystal structure (PDB ID: 4R4P; *trans* ribozyme in white surface with A$_{756}$ in red stick and SLI in gray sticks), superposing the composite model of the frame on stem V as described in 'Materials and Methods' section. In this context, the P-P distance (d_{PP}) corresponds to the distance between the phosphorus atom of the scissile phosphates within the crystal structure (gray sphere) and the selected frame model (green sphere).

characterized using metrics that are either intrinsic to the KL complex or evaluated in the context of the VS ribozyme crystal structure (Figure 2). First, the distribution of Euler angles (α, β and γ) between the two helical stems of the KL complexes were calculated (45), which helped derive the more intuitive interhelical angle θ (Figure 2A). Then, composite models were built starting from the KL complex of each simulated frame and then attaching the internal loop and stem Ia of SLI to one stem (I, L88 or TAR) and the *trans* VS ribozyme to the other stem (V, L22 or TAR*) via superposition of common helical elements. These composite models allowed for the analysis of individual frames in the context of the VS ribozyme crystal structure. From heavy atom superposition of central base pairs of stem V within the composite model and the crystal structure, we measured the distance between the phosphorus atoms at the scissile phosphate of the two SLI internal loops (d_{PP}; Figure 2B) as well as the heavy-atom RMSD between these two loops.

The 2D heat maps showing distributions of Euler angles demonstrate that the I/V KL complex explores a wider region of the Euler angle space compared to the L88/L22 and TAR/TAR* KL complexes (Figure 3). The I/V KL complex access interhelical angles that the surrogate KL complexes never adopt during the simulation, notably the set of Euler angles measured for the I/V KL complex within the crystal structure (marked by a star in Figure 3). Thus, the analysis of interhelical Euler angles indicate that the intrinsic global dynamics of the I/V KL complex are less constrained than those of the TAR/TAR* and L88/L22 KL complexes. The restricted global mobility of the surrogate KL complexes is particularly important for the TAR/TAR* KL complex, which is confined to a much narrower range of α, β and γ angles.

This disparity in conformational space between the I/V KL complex and the surrogate KL complexes was also examined from the 2D heat maps showing the distribution of interhelical angle θ as a function of d_{PP} (Figure 4). For the I/V KL complex, several frames of the T-REMD trajectory
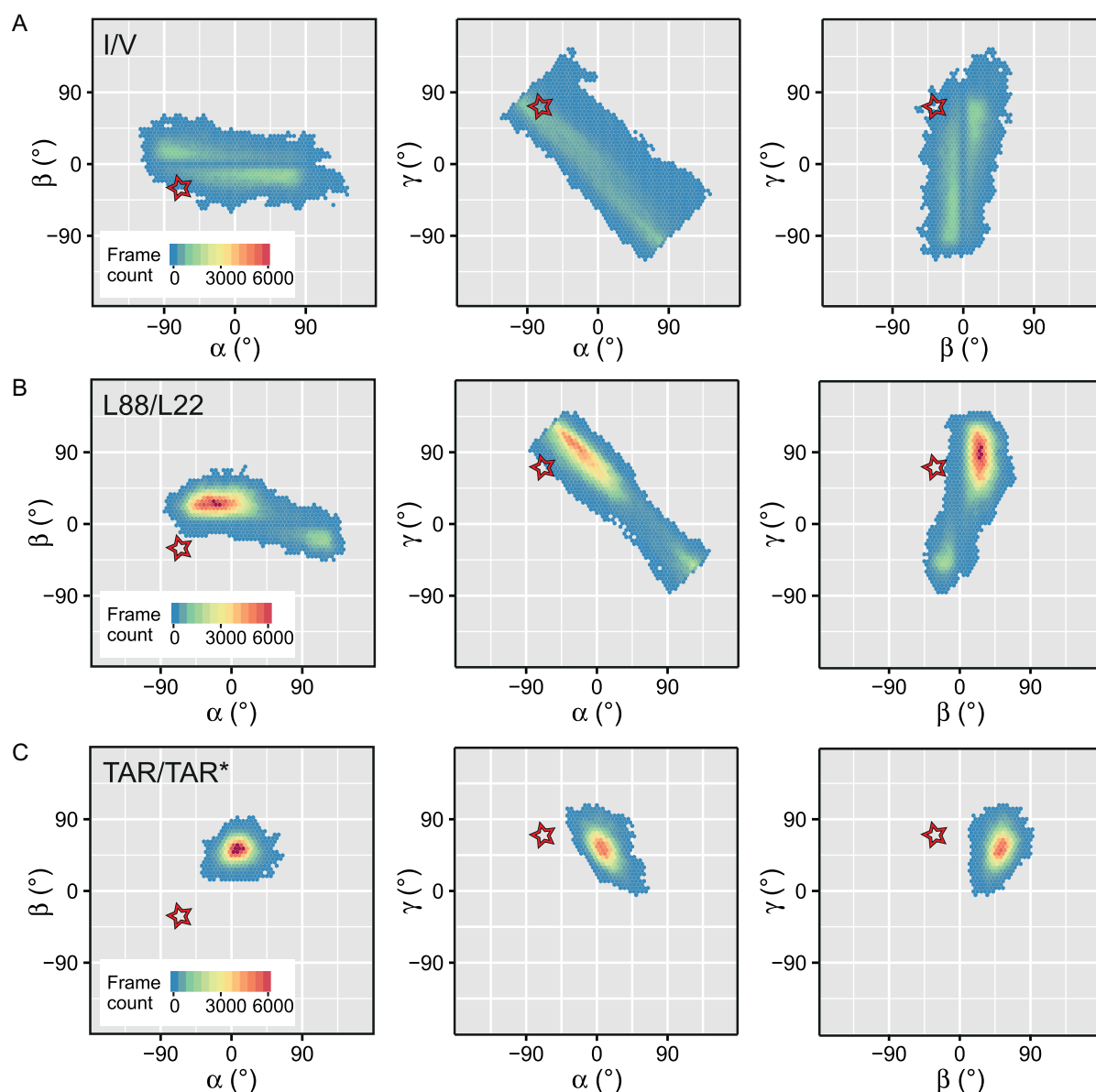
Figure 3. Difference in conformational space between the I/V KL complex and surrogate KL complexes from Euler angle analysis. The 2D heat maps of Euler angle combinations representing all frames of the T-REMD trajectories at 300 K are shown for the (**A**) I/V, (**B**) L88/L22 and (**C**) TAR/TAR* KL complexes. In the 2D maps, the star points to the Euler angle values measured for the I/V KLI within the reference crystal structure (PDB ID: 4R4P; $\alpha = -69°$; $\beta = -31°$; and $\gamma = 71°$).

adopt d_{PP} values approaching 0 Å, and small d_{PP} values are observed for a wide range of interhelical angles ($d_{PP} < 2$ Å for $\theta = 136°–178°$), matching the range observed in the ensemble of 20 lowest-energy NMR structures of the I/V KL complex [$\theta = 127°–175°$; (24)]. This analysis clearly indicates that the trajectory of the I/V KL complex brings the internal loop of the substrate close to its position in the crystal structure (marked by a star in Figure 4), which is compatible with formation of the active site. In contrast, for the L88/L22 and TAR/TAR* KL complexes, d_{PP} values were all greater than 6 and 11.5 Å, respectively, and these d_{PP} values directly correlate with the interhelical angle θ. Similar observations were made by examining the 2D heat maps showing the distribution of interhelical angle θ as a

function of RMSD (Supplementary Figure S3), where the smallest RMSD value attained by the I/V KLI (2.6 Å), is much smaller than those for both the L88/L22 KL complex (8.7 Å) and the TAR/TAR* KL complexes (14.9 Å). Thus, the I/V KLI allows for a more complex range of motion and enables a large subpopulation of structures to approach the closed state in which the active site is formed.

Principal component analysis of T-REMD trajectories

To better understand the dynamics associated with the three different KL complexes, a PC analysis was performed on the atomic coordinates of each T-REMD trajectory. This analysis extracts non-overlapping concerted atomic movements, or PCs, from the seemingly random vibrations of
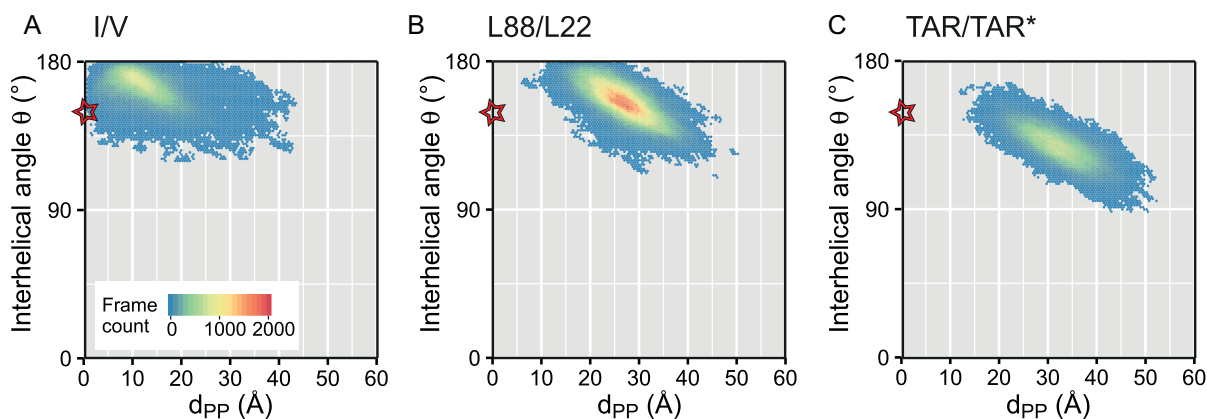
Figure 4. Correlation between the interhelical angle θ and the d_{PP} value for the I/V KL complex and surrogate KL complexes. The 2D heat maps of the interhelical angle θ versus d_{PP} representing all frames of the T-REMD trajectories at 300 K are shown for the (**A**) I/V, (**B**) L88/L22 and (**C**) TAR/TAR* KL complexes. In the 2D maps, the star points to values measured for the I/V KLI within the reference crystal structure (PDB ID: 4R4P; $\theta = 149°$ and $d_{PP} = 0$).

atomic positions observed in the T-REMD trajectories. For the three KL complexes, the first five PCs together represent more than 70% of the cumulative percentage of variance (Supplementary Figure S4 and Table 1). Thus, they provide the main contributions of the global motion. Pseudo-trajectories for each of these first five PCs were produced, in which atomic movements are projected along a single PC. Although these pseudo-trajectories can give rise to some partially unrealistic local structures, they are very useful in illustrating the direction and relative amplitude of concerted motions. They are displayed here in the context of the VS ribozyme crystal structure (Figure 5 and Supplementary Movies S1-S15) to provide insights into how specific motions may contribute to active-site formation.

For the I/V KL complex, the first five PCs help to bring the scissile phosphate of the SLI substrate in proximity of the A_{756} internal loop within the ribozyme, with the first two PCs having the greatest impact (Figure 5 and Supplementary Movies S1–S5). The first PC, which accounts for 34% of the variance (Table 1) is characterized by a large-amplitude twist motion of the SLI substrate around an axis that is close to its own principal axis (Figure 5A). Interestingly, this twist motion appears to be correlated with an important structural transition in loop I between compact and loose states (Movie S1). The second PC, with 16% of the variance, is also of large amplitude, but is better described as a side-to-side bend motion that swings the substrate perpendicularly to the II-VI plane formed by helical domains II and VI (Figure 5B). The third PC describes a twist motion that is smaller in amplitude than the first PC, and it accounts for 9.5% of the variance. Finally, the fourth and fifth PCs both describe back-to-front bend motions, during which the substrate travels parallel to the II-VI plane (Figure 5D and E). Taken together these PCs of motion indicate that the I/V KLI can be viewed as a multi-axial 'RNA joint' that resembles a ball-and-socket shoulder joint, since it allows circular movements in three orthogonal directions.

Analysis of the individual PCs for the surrogate KL complexes indicate that their global dynamics are similarly composed of a mixture of movements, which includes a twist motion of the substrate around its main axis and different
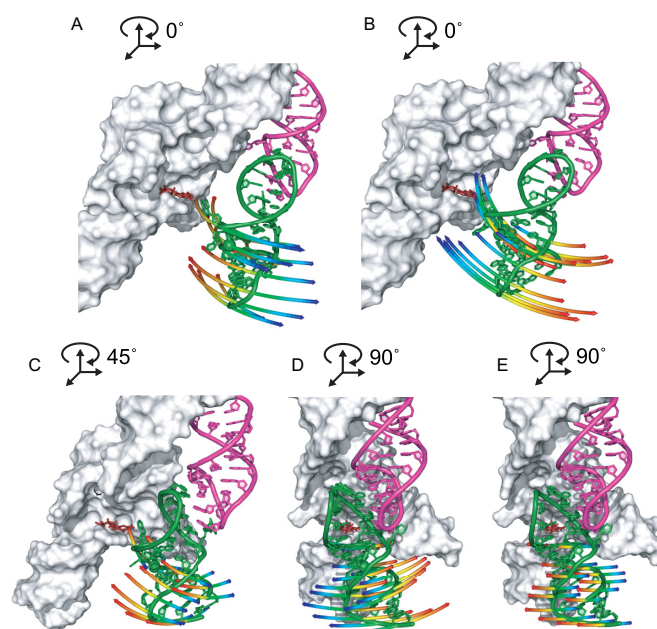

Figure 5. Cartoon representations of the first five PCs of motion for the I/V KL complex. (**A**) The large twist of the first PC, (**B**) the large side-to-side bend of the second PC, (**C**) the small twist of the third PC and (**D** and **E**) the two small back-to-front bends of the fourth and fifth PCs. The PCs of motion were extracted from the T-REMD trajectory at 300 K for the I/V KL complex and are represented in the context of the reference VS ribozyme X-ray structure (PDB ID: 4R4P; in white surface with A_{756} in red stick), by superposing interpolated frames from the pseudo-trajectory (SLV in pink and SLI in green) as described in 'Materials and Methods'. The movements of the backbone phosphorus atoms from SLI residues 616-623 and 637-643 are materialized as rainbow-colored arrows along with a single interpolated frame taken at the halfway point of the pseudo-trajectory.

bend motions (Supplementary Figures S5 and S6; Supplementary Movies S6-S15). However, the contribution and amplitude of these movements differ between the three KL complexes (Table 1). Only the PCs for the I/V KL complex have sufficient amplitudes to bring the G_{638} loop in proximity of the A_{756} loop, in agreement with the short d_{PP} values

Table 1. Characteristics of the first five PCs of motion for the I/V, L88/L22 and TAR/TAR* KL complexes

	Principal components				
	1	2	3	4	5
I/V					
% of variance	34.4	16.3	9.5	7.0	5.2
Type of motion	Twist	Side-to-side bend	Twist	Back-to-front bend	Back-to-front bend
Amplitude of motion (Å)[a]	44	31	13	21	19
L88/L22					
% of variance	24.4	20.8	13.5	8.9	5.3
Type of motion	Bend[b]	Bend[b]	Twist	Back-to-front bend	Side-to-sidebend
Amplitude of motion (Å)	38	32	6	11	17
TAR/TAR*					
% of variance	31.5	21.7	9.6	7.6	5.3
Type of motion	Bend[b]	Side-to-side bend	Bend[b]	Twist	Twist
Amplitude of motion (Å)	34	35	19	6	11

[a]The amplitude of motion was calculated as the distance travelled by the scissile phosphate in the pseudo-trajectory.
[b]It contains almost equivalent back-to-front and side-to-side components.

observed for this complex. Moreover, the I/V KL complex is the only one in which the preeminent motion is a twist of the substrate around an axis that is close to its own principal axis. Whereas twist motions are observed for the L88/L22 and TAR/TAR* KL complexes, their contribution and amplitude are far less important.

Dynamics of KL complexes versus cleavage activity of *S/R* complexes

In order for the VS ribozyme to cleave its substrate, the G_{638} and A_{756} loops must intimately associate so that the reactive atoms are precisely positioned in the active site. However, in solution, the VS ribozyme is best represented by an ensemble of open conformations in which these two loops are most often apart from each other (25,26,48). Our analysis of T-REMD trajectories suggest that the higher activity of the S_0/R_0 complex compared to the S_{L88}/R_{L22} and S_{TAR}/R_{TAR*} complexes (Figure 1C) is linked to the dynamics of the I/V KL complex, which is better suited to allow spatial closeness and proper positioning of the G_{638} loop relative to the A_{756} loop. The twist and side-to-side motions of large amplitude described by the first two PCs can bring the two internal loops in proximity, while other twist and bend motions of smaller amplitude can fine tune the relative position of the two internal loops.

Given that the S_{L88}/R_{L22} and S_{TAR}/R_{TAR*} variant complexes display cleavage activity (Figure 1C), it is intriguing that the dynamics of the corresponding KL complexes characterized here does not allow the G_{638} and A_{756} loops to come in proximity. This is clearly shown both from the large d_{PP} values of the T-REMD frames (Figure 4B and C) and visualization of the main PCs of motions (Supplementary Figures S5 and S6; Supplementary Movies S6-S15). The lack of frames with smaller d_{PP} values may be due to the limited sampling of the T-REMD simulations. MD/T-REMD simulations that extend to longer timescales or explore higher temperatures might capture closed conformations, possibly by exploring less stable KL structures with fewer base pairs. Indeed, we observed surrogate KL complexes with disrupted KL base pairs in frames with the shortest d_{PP} values. Alternatively, other dynamic hotspots in the VS ribozyme, potentially the stem III bulge and the two-
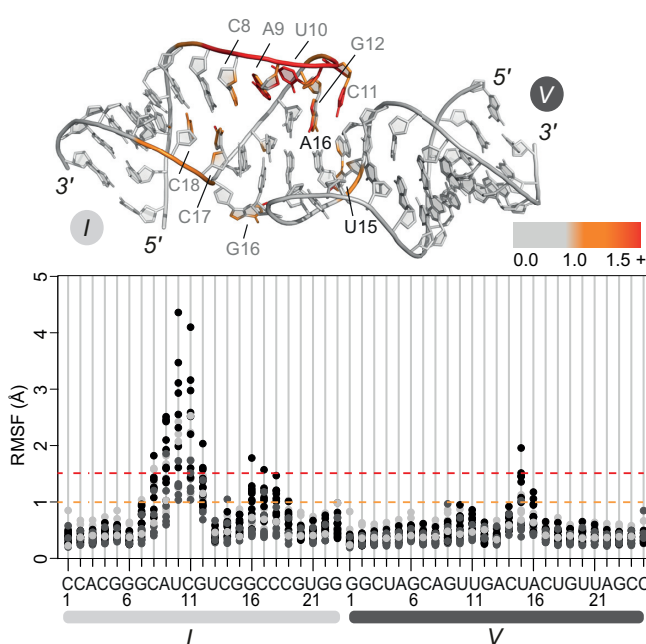

Figure 6. Identification of mobile and rigid residues in the I/V KL complex. RMSF values were calculated for phosphate atoms (pale gray), ribose ring atoms (dark gray) and base heavy atoms (black) of the T-REMD trajectories at 300 K. The RMSF values are mapped according to the displayed color scheme on the 3D structure of the I/V KL complex (top panel) that was used to initiate the T-REMD simulation.

three way junctions (26), could help to bring the G_{638} and A_{756} loops in sufficient proximity to allow substrate cleavage.

Identification of the local hotspots of motion responsible for the KLI dynamics

To identify local hotspots of motion, we calculated the RMSF of every consecutive pair of residues along the stem-loop sequences for each KL complex (Figure 6). Remarkably, the I/V KL complex exhibits several clusters of residues with high RMSF values (>1.5 Å), defining three local hotspots of motion (residues 8–12 of loop I; residues

16–17 of loop I and residue 15 of loop V), in agreement with previous NMR studies (24). In contrast, fewer residues with high RMSF values were identified for the L88/L22 and TAR/TAR* KL complexes (Supplementary Figure S7), in agreement with more restricted motions associated with these surrogate KL complexes. It is not surprising that for all three KL complexes, the most dynamic residues are found at the junction between the hairpin stem and the KL stem. Unlike the surrogate KL complexes, the I/V KL complex contains several unpaired residues (I: A_9, U_{10} and G_{16}; V: U_{11} and U_{15}) and unstable base pairs (I: C_8-G_{15}; V: U_{10}-A_{16}) at this junction (Figure 1 and Supplementary Figure S1) (24).

Through their association with dynamic hotspots, the unpaired residues and unstable base pair in loop I largely contribute to the conformational dynamics of the I/V KL complex. Although there is a diversity of loop I conformations in the T-REMD trajectory, we noted that several structures adopt a compact loop I conformation in which the two base pairs closing the loops are formed (G_7-C_{17} and C_8-G_{15}), whereas several other structures adopt a loose loop I conformation in which these base pairs are clearly disrupted. The loose and compact conformations of loop I can both populate the open and closed states when placed in the context of the VS ribozyme (Figure 1B). Interestingly, the large-amplitude twist motion defining the first PC in the I/V KL complex involves a structural transition between compact and loose states of loop I that is coupled to helical winding and unwinding (Movie S1). Such structural transition between loose and compact loop I conformations is not evident for the surrogate KL complexes, and this could explain why the preeminent twist motion observed for the I/V KL complex is not observed for these KL complexes.

RNA engineering by increasing the dynamics of an RNA junction

Given the importance of dynamics at the KLI for the high cleavage activity of the VS ribozyme complex (S_0/R_0), we postulated that the cleavage activity of the variant S_{TAR}/R_{TAR*} complex could be increased if one could add some dynamics at the TAR/TAR* KLI. In order to test this hypothesis, we first performed an *in vitro* selection to identify VS ribozyme variants that more efficiently cleave the S_{TAR} substrate (Figure 7A). For this selection, the S_{TAR} substrate was covalently bound to the ribozyme; it contains the HIV-1 TAR RNA sequence in stem Ia and stem-loop Ib but preserved the natural G_{638} loop necessary for formation of the active site. In addition, loop V (5 nt) and the adjacent two closing base pairs (4 nt) were replaced by a stretch of 10 randomized nt to form an RNA library with ~1 million (4^{10}) different loop sequences. After six rounds of selection and amplification, next-generation sequencing of the selected ribozyme pool was performed, and enriched sequences (present 10 times or more in the pool) were analyzed in terms of the potential number of consecutive WC base pairs that could form with loop I (Figure 7B). Most of the selected sequences (96%) have the potential to form at least 4 WC base pairs, confirming that the *in vitro* selection procedure allowed us to select variant VS ribozymes for which cleavage in *cis* depends on the formation of the

KLI. Formation of 4 WC base pairs at the KLI appears to be necessary, but also sufficient for cleavage activity since 64% of the sequences could form a maximum of 4 WC base pairs. Since only 2% of the ribozyme pool has the potential to form the maximum 6 WC base pairs, it appears that increasing the thermodynamic stability of the KLI is not beneficial, but instead this may in fact inhibit the activity. We found two main classes of loop V sequences with the preferred CCCA pattern starting either at position +4 or +5 of the randomized sequence (Figure 7C). In both cases, formation of 4–5 WC base pairs at the KLI could leave several residues unpaired in both KLs, which would provide more dynamics for formation of the active site and more efficient substrate cleavage.

Interestingly, when we modified the *trans* VS ribozyme R_0 to incorporate one of the selected loop V sequence predicted to form 4 consecutive WC base pairs with S_{TAR}, the resulting ribozyme (R_{TAR-S1}) cleaved S_{TAR} with a relatively high single-turnover rate (k_{obs} = 5.8 min^{-1}; Figure 7D). This rate is much higher than that previously reported for the most active rationally-engineered S_{TAR}/R_{TAR*} complex (k_{obs} = 0.077 min^{-1}; Figure 1C) and similar to the rate of the parental S_0/R_0 complex [k_{obs} = 12.2 min^{-1}; Figure 1C; (16)]. In addition, the R_0 and R_{TAR*-0} ribozymes, which can respectively form 3 and 6 WC consecutive base pairs with S_{TAR}, both cleave this substrate with a much lower rate ($\leq$200-fold; $k_{obs} \leq$0.03 min^{-1}; Figure 7D) than R_{TAR-S1}. To assess the thermodynamic stability of the KLI formed between the S_{TAR} substrate and the selected R_{TAR-S1} ribozyme with improved activity, we performed ITC studies using small stem-loops (Supplementary Figure S8). A K_D of 0.7 $\pm$ 0.2 μM was measured for this TAR/TAR-S1 KLI, indicating that it is 120-fold less stable than the TAR/TAR* KLI (0.0059 $\pm$ 0.0005 μM; Figure 1). Overall, these results demonstrate that the most active VS ribozyme variant is not the one with the most thermodynamically stable KLI.

We performed a T-REMD simulation to investigate the dynamics of the TAR/TAR-S1 KL complex (Figure 7; Supplementary Figures S9 and S10). Although this T-REMD simulation was initiated from a structural model and did not converge quite to the same extent as the other simulations reported here (Supplementary Figure S9a), broad comparisons with the original TAR/TAR* KL complex can still be drawn. We found that the selected TAR/TAR-S1 KL complex explores a much larger region of the Euler angle space (Supplementary Figure S9b) than the original TAR/TAR* KL complex (Figure 3C), and includes the values determined from the crystal structure (marked by a star in Supplementary Figure S9b). Several frames of the T-REMD trajectory adopt d_{PP} values (down to 5.6 Å; Supplementary Figure S9b) that are much smaller than the smallest values observed for the TAR/TAR* KL complex (11.5 Å; Figure 4C), and this indicates that the dynamics of TAR/TAR-S1 KL complex brings the internal loop of the substrate closer to its position in the crystal structure. Interestingly, PC analysis of the T-REMD trajectory at 300 K for the TAR/TAR-S1 KL complex reveals a dominating twist motion of the substrate around an axis close to its own principal axis (Figure 7F and Supplementary Figure S10; first PC with 38.4% variance and a 58-Å amplitude of motion), as previously observed for the I/V KL complex, but
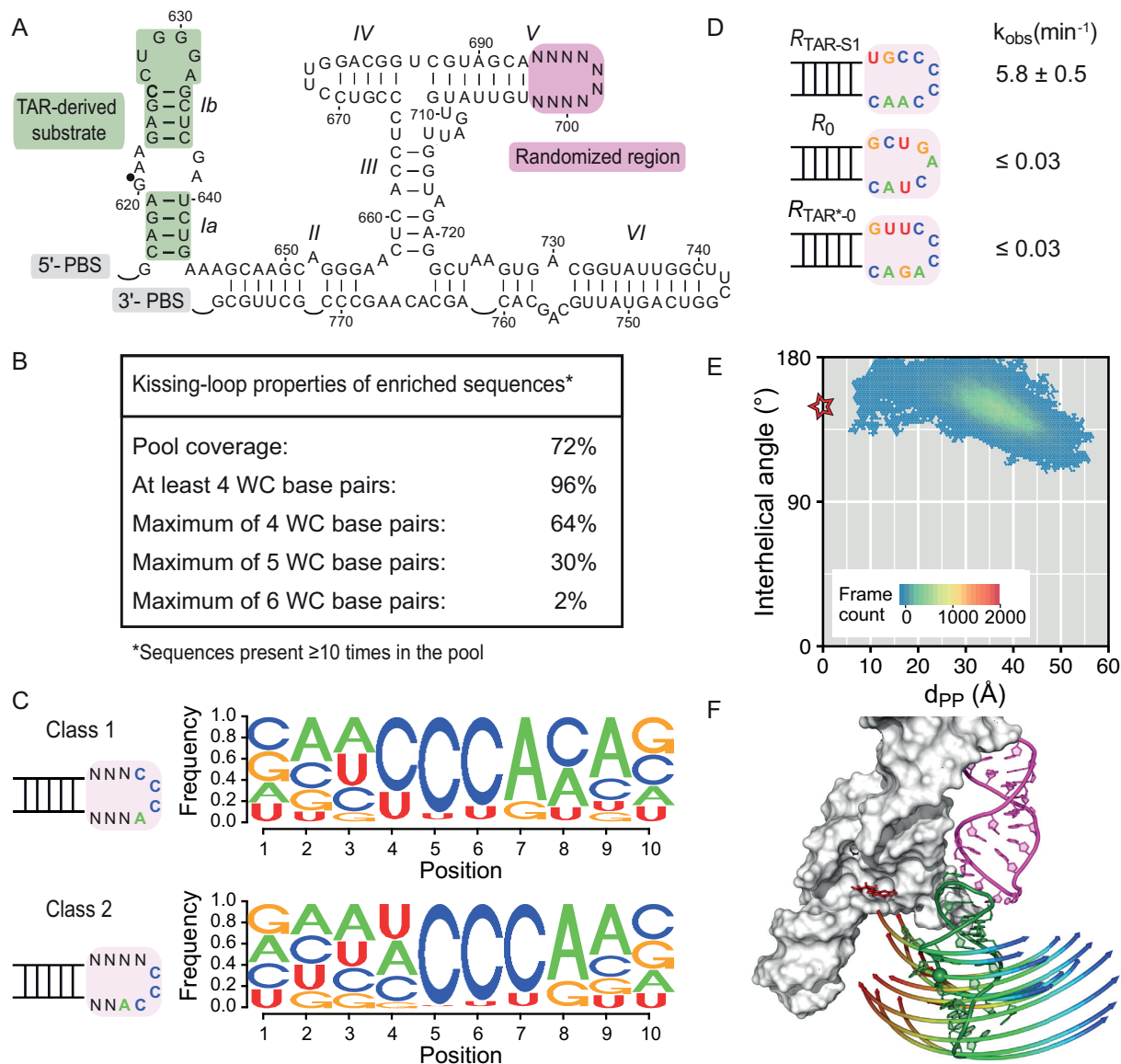
Figure 7. Improving the cleavage activity of a variant VS ribozyme for its substrate through a more dynamic KLI. (**A**) *In vitro* selection library to identify a VS ribozyme sequence that efficiently cleaves a TAR-derived substrate. (**B**) KL properties of enriched ribozyme sequences ($\geq$10 representatives in the selected pool) from the *in vitro* selection pool in terms of the number of potential Watson–Crick (WC; G-C, A-U and G-U) base-pairs between loop I and loop V. (**C**) Sequence preferences of the randomized region for the two classes of enriched ribozyme sequences. (**D**) SLV loop sequences of a selected ribozyme with high activity (R_{TAR-S1}), the reference ribozyme (R_0) and a rationally-engineered ribozyme (R_{TAR*-0}) together with their observed *trans* cleavage rates (k_{obs}) of the S_{TAR} substrate. The k_{obs} values were determined under the same conditions for results shown here and those in Figure 1C. (**E**) The 2D heat map of the interhelical angle θ versus d_{PP} representing all frames of the T-REMD trajectory at 300 K for the selected TAR/TAR-S1 KL complex. The star points to values measured for the I/V KLI within the reference crystal structure (PDB ID: 4R4P; $\theta = 149°$ and $d_{PP} = 0$). (**F**) Cartoon representation of the first PC of motion for the TAR/TAR-S1 KL complex represented in the context of the VS ribozyme X-ray structure as in Figure 5.

not the surrogate TAR/TAR* KL complex. Furthermore, higher RMSF values were obtained for the TAR/TAR-S1 KL (Supplementary Figure S9e) compared to TAR/TAR* KL complex (Supplementary Figure S7), and in particular for residues at the junction between the hairpin and the KL stems, in agreement with increased motions associated with the selected TAR/TAR-S1 KLI. A structural analysis of the T-REMD trajectory indicates that the TAR/TAR-S1 KLI has the potential to form 6 base pairs: the 4 predicted WC base pairs involving the central residues of the TAR loop, which are found simultaneously in 75% of the

frames, and two *cis* Watson–Crick/Watson–Crick A-C base pairs involving the first C (in 72% of the frames) and last A (in 22% of the frames) of the TAR loop. As a result, the TAR/TAR-S1 KLI forms a maximum of 3, 4, 5 or 6 base pairs in 10, 23, 50 and 15% of the frames, respectively. In comparison, the TAR/TAR* KLI forms 6 WC base pairs in 76% of the frames. Taken together, our thermodynamic, kinetic and molecular dynamic results demonstrate that the cleavage activity of the variant S_{TAR}/R_{TAR*} complex can be significantly increased by changing the sequence of loop V
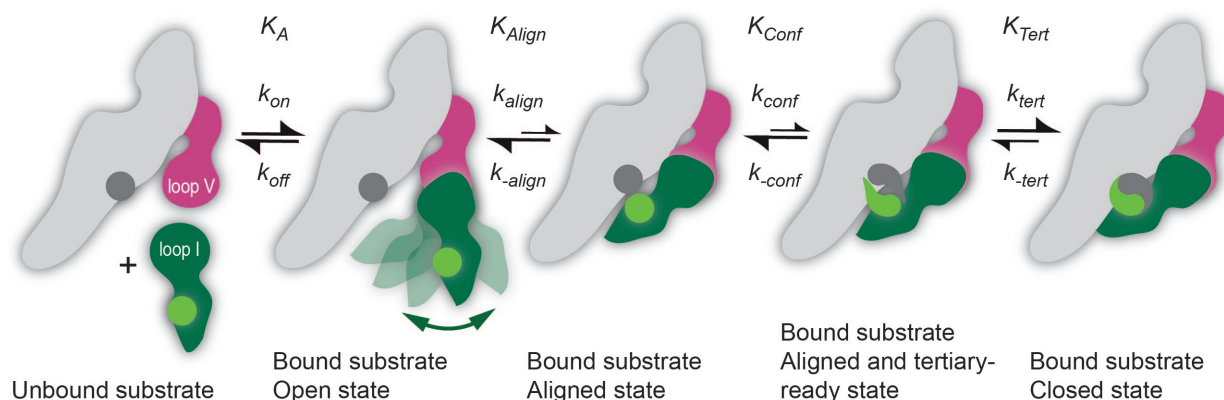
Figure 8. Simple kinetic model for binding of a pre-shifted substrate to the *trans* VS ribozyme. Formation of the catalytically-competent state involves at least four steps: binding of the SLI substrate (dark green) to the *trans* ribozyme via formation of the KLI with stem-loop V (dark pink), alignment of the tertiary contact between the G_{638} loop (light green sphere) and the A_{756} loop (dark gray sphere), adoption of a tertiary-ready state within these two loops and formation of the tertiary interaction.

to partly destabilize the KLI and thereby increasing the dynamics at the KL junction.

Mechanistic model of substrate binding in the VS ribozyme

Substrate cleavage by the *trans* VS ribozyme can be considered a two-step kinetic process that involves substrate binding and catalysis. To integrate the knowledge gained from the present studies, we propose a simple mechanistic model that breaks down substrate binding into four substeps (Figure 8). This model describes the binding of pre-shifted substrates as used here but could be expanded to the binding of shiftable substrates as found in the wild-type VS ribozyme by incorporating the helix shift associated with the formation of the KLI (23,24,26,49).

The first step is the formation of the highly-stable I/V KLI, a process that has been previously described in detail based on structural investigations of free and bound stem-loops (24,26,42,49). Formation of this KLI is the predominant thermodynamic contribution to substrate binding since the SLI substrate binds with similar affinity to an isolated SLV and the full *trans* ribozyme (23). In addition to substrate binding via the KLI, the G_{638} and A_{756} loops must associate to form the closed state, which depends on three factors: the alignment of the two internal loops (K_{Align}), the conformational change of the tertiary contact elements (K_{Conf}) and the actual formation of the tertiary interaction (K_{Tert}). The alignment of the internal loops is largely governed by the intrinsic dynamics of the KLI, and our T-REMD simulation of the isolated I/V KL complex indicates that only a small portion of the conformational space is compatible with the aligned state. However, additional factors likely contribute to define K_{Align}, including the electrostatic repulsion associated with bringing helical domains in proximity and the intrinsic dynamics of other subdomains of the VS ribozyme. Conformational changes would be necessary to bring both the G_{638} and A_{756} loops in a tertiary-ready conformation (K_{Conf}). Based on structural comparisons between the free and bound states of these loops, it was previously shown that only minor conformational changes would allow their less constrained residues

to adopt a protruded conformation compatible with the closed state (26). This suggests that the probability of forming the tertiary-ready conformation would depend on the individual probabilities of those protruded conformations within the conformational ensembles of the free G_{638} and A_{756} loops. Once the internal loops adopt an aligned and tertiary-ready state, the free energy of the tertiary interaction (ΔG_{Tert} related to K_{Tert}) would likely favor the closed state given the large number of hydrogen bonds and stacking interactions observed in the crystal structure (27,49). In summary, these forward steps, namely formation of the KLI, the aligned state, the tertiary ready state and the tertiary interaction, are all part of the on-rate kinetics for the cleavage reaction, with k_{off} being defined from the overall thermal stability of the KLI (ΔG_{KLI}), and thus all these forward steps contribute to determine if the reaction can proceed before the substrate dissociates from the ribozyme.

This mechanistic framework explains why mutations that destabilize the KLI can reduce the activity of the VS ribozyme (22,50) but also specifies that the thermodynamic stability of the KLI is not the sole determinant of substrate binding. As shown here, the more stable TAR/TAR* KLI ($K_D = 5.9$ nM) yields a less active VS ribozyme variant than the less stable and more dynamic TAR/TAR-SI KLI ($K_D = 0.7$ μM). Thus, too much stability within the KLI can significantly reduce the cleavage efficiency by not allowing sufficient sampling of the aligned state required for formation of the closed state. This relation between binding events and conformational equilibria of RNA elements is increasingly being recognized as an important hallmark of RNA. In recent pioneering studies, it was formalized and tested in a thermodynamic model of RNA folding (8) and incorporated in high-throughput virtual screening for targeting RNA with small molecules (9).

CONCLUSION

This work was conducted to test the hypothesis that the dynamic properties of the KLI in the VS ribozyme contribute to the rate of substrate cleavage. Our T-REMD simulation of the natural I/V KLI reveals that it functions like a

multi-axial RNA joint with a large range of motion, which facilitates formation of the active site through conformational sampling. In contrast, T-REMD simulations showed more restricted motion for two surrogate KLIs that can functionally substitute for the I/V KLI, but produce VS ribozymes with reduced activity, thereby supporting our initial hypothesis. We performed *in vitro* selection studies to improve the cleavage activity of the S_{TAR}/R_{TAR*} ribozyme variant by randomizing the KL of the *trans* ribozyme. Sequence analysis of the selected ribozyme variants and kinetic studies suggest that the KLI must be appropriately stable yet sufficiently dynamic for efficient substrate cleavage by the VS ribozyme. Moreover, we identified a ribozyme variant that more efficiently cleaves the S_{TAR} substrate and also forms a more dynamic KLI with the S_{TAR} substrate than the rationally-engineered R_{TAR*}. Thus, engineering an RNA junction with a larger range of motion can improve the cleavage activity of a variant ribozyme toward a non-natural substrate by promoting sampling of the active conformation. Based on these results, we propose a mechanistic model of substrate binding in the VS ribozyme that integrates conformational sampling of RNA subdomains. This model also illustrates how the modularity of RNA can be expressed in terms of defining the thermodynamics and kinetics relevant to RNA function. Such integrative view of RNA interaction and conformational dynamics will likely prove important in future applications, such as efforts to engineer RNA with novel functions and to identify specific compounds that target RNA.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEGEMENTS

FUNDING

REFERENCES

1. Parlea,L.G., Sweeney,B.A., Hosseini-Asanjan,M., Zirbel,C.L. and Leontis,N.B. (2016) The RNA 3D motif atlas: computational methods for extraction, organization and evaluation of RNA motifs. *Methods*, **103**, 99–119.

2. Brion,P. and Westhof,E. (1997) Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 113–137.

3. Tinoco,I. Jr and Bustamante,C. (1999) How RNA folds. *J. Mol. Biol.*, **293**, 271–281.

4. 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.

5. Mustoe,A.M., Brooks,C.L. and Al-Hashimi,H.M. (2014) Hierarchy of RNA functional dynamics. *Annu. Rev. Biochem.*, **83**, 441–466.

6. Herschlag,D., Allred,B.E. and Gowrishankar,S. (2015) From static to dynamic: the need for structural ensembles and a predictive model of RNA folding and function. *Curr. Opin. Struct. Biol.*, **30**, 125–133.

7. Daher,M., Mustoe,A.M., Morriss-Andrews,A., Brooks,C.L. III and Walter,N.G. (2017) Tuning RNA folding and function through rational design of junction topology. *Nucleic Acids Res.*, **45**, 9706–9715.

8. Bisaria,N., Greenfeld,M., Limouse,C., Mabuchi,H. and Herschlag,D. (2017) Quantitative tests of a reconstitution model for RNA folding thermodynamics and kinetics. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E7688–E7696.

9. Ganser,L.R., Lee,J., Rangadurai,A., Merriman,D.K., Kelly,M.L., Kansal,A.D., Sathyamoorthy,B. and Al-Hashimi,H.M. (2018) High-performance virtual screening by targeting a high-resolution RNA dynamic ensemble. *Nat. Struct. Mol. Biol.*, **25**, 425–434.

10. Denny,S.K., Bisaria,N., Yesselman,J.D., Das,R., Herschlag,D. and Greenleaf,W.J. (2018) High-throughput investigation of diverse junction elements in RNA tertiary folding. *Cell*, **174**, 377–390.

11. Geary,C., Chworos,A., Verzemnieks,E., Voss,N.R. and Jaeger,L. (2017) Composing RNA nanostructures from a syntax of RNA structural modules. *Nano Lett.*, **17**, 7095–7101.

12. Jasinski,D., Haque,F., Binzel,D.W. and Guo,P. (2017) Advancement of the emerging field of RNA nanotechnology. *ACS Nano*, **11**, 1142–1164.

13. Hallberg,Z.F., Su,Y., Kitto,R.Z. and Hammond,M.C. (2017) Engineering and in vivo applications of riboswitches. *Annu. Rev. Biochem.*, **86**, 515–539.

14. Balke,D., Hieronymus,R. and Muller,S. (2017) Challenge and perspectives in nucleic acid enzyme engineering. *Advances in Biochemical Engineering/Biotechnology*. Springer, Berlin, Heidelberg.

15. Lacroix-Labonté,J., Girard,N., Lemieux,S. and Legault,P. (2012) Helix-length compensation studies reveal the adaptability of the VS ribozyme architecture. *Nucleic Acids Res.*, **40**, 2284–2293.

16. Lacroix-Labonté,J., Girard,N., Dagenais,P. and Legault,P. (2016) Rational engineering of the *Neurospora* VS ribozyme to allow substrate recognition via different kissing-loop interactions. *Nucleic Acids Res.*, **44**, 6924–6934.

17. Jimenez,R.M., Polanco,J.A. and Luptak,A. (2015) Chemistry and biology of self-cleaving ribozymes. *Trends Biochem. Sci.*, **40**, 648–661.

18. Wilson,T.J., Liu,Y. and Lilley,D.M.J. (2016) Ribozymes and the mechanisms that underlie RNA catalysis. *Front. Chem. Sci. Eng.*, **10**, 178–185.

19. Ren,A., Micura,R. and Patel,D.J. (2017) Structure-based mechanistic insights into catalysis by small self-cleaving ribozymes. *Curr. Opin. Chem. Biol.*, **41**, 71–83.

20. Haseloff,J. and Gerlach,W.L. (1989) Sequences required for self-catalysed cleavage of the satellite RNA of tobacco ringspot virus. *Gene*, **82**, 43–52.

21. 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.

22. 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.

23. Bouchard,P. and Legault,P. (2014) A remarkably stable kissing-loop interaction defines substrate recognition by the *Neurospora* VS Ribozyme. *RNA*, **20**, 1451–1464.

24. Bouchard,P. and Legault,P. (2014) Structural insights into substrate recognition by the *Neurospora* Varkud satellite ribozyme: importance of U-turns at the kissing-loop junction. *Biochemistry*, **53**, 258–269.

25. Bonneau,E., Girard,N., Lemieux,S. and Legault,P. (2015) The NMR structure of the II-III-VI three-way junction from the *Neurospora* VS ribozyme reveals a critical tertiary interaction and provides new insights into the global ribozyme structure. *RNA*, **21**, 1621–1632.

26. Dagenais,P., Girard,N., Bonneau,E. and Legault,P. (2017) Insights into RNA structure and dynamics from recent NMR and X-ray studies of the *Neurospora* Varkud satellite ribozyme. *WIREs RNA*, **8**, e1421.

27. Suslov,N.B., DasGupta,S., Huang,H., Fuller,J.R., Lilley,D.M., Rice,P.A. and Piccirilli,J.A. (2015) Crystal structure of the Varkud satellite ribozyme. *Nat. Chem. Biol.*, **11**, 840–846.

28. Sponer,J., Bussi,G., Krepl,M., Banas,P., Bottaro,S., Cunha,R.A., Gil-Ley,A., Pinamonti,G., Poblete,S., Jurecka,P. *et al.* (2018) RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chem. Rev.*, **118**, 4177–4338.

29. Pronk,S., Pall,S., Schulz,R., Larsson,P., Bjelkmar,P., Apostolov,R., Shirts,M.R., Smith,J.C., Kasson,P.M., van der Spoel,D. *et al.* (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, **29**, 845–854.

30. Abraham,M.J., Murtola,T., Schulz,R., Páll,S., Smith,J.C., Hess,B. and Lindahl,E. (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1-2**, 19–25.

31. Hornak,V., Abel,R., Okur,A., Strockbine,B., Roitberg,A. and Simmerling,C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, **65**, 712–725.

32. Perez,A., Marchan,I., Svozil,D., Sponer,J., Cheatham,T.E. 3rd, Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.*, **92**, 3817–3829.

33. Guy,A.T., Piggot,T.J. and Khalid,S. (2012) Single-stranded DNA within nanopores: conformational dynamics and implications for sequencing; a molecular dynamics simulation study. *Biophys. J.*, **103**, 1028–1036.

34. Zgarbova,M., Otyepka,M., Sponer,J., Mladek,A., Banas,P., Cheatham,T.E. and Jurecka,P. (2011) Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.*, **7**, 2886–2902.

35. Darden,T., Perera,L., Li,L. and Pedersen,L. (1999) New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure*, **7**, R55–R60.

36. Essmann,U., Perera,L., Berkowitz,M.L., Darden,T., Lee,H. and Pedersen,L.G. (1995) A smooth particle mesh Ewald method. *J.Chem. Phys.*, **103**, 8577–8593.

37. 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.

38. Magee,T.V., Han,S., McCurdy,S.P., Nguyen,T.T., Granskog,K., Marr,E.S., Maguire,B.A., Huband,M.D., Chen,J.M., Subashi,T.A. *et al.* (2013) Novel 3-O-carbamoyl erythromycin A derivatives (carbamolides) with activity against resistant staphylococcal and streptococcal isolates. *Bioorg. Med. Chem. Lett.*, **23**, 1727–1731.

39. Chang,K.Y. and Tinoco,I. Jr (1997) The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement. *J. Mol. Biol.*, **269**, 52–66.

40. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.

41. Auffinger,P. (2012) Ions in molecular dynamics simulations of RNA systems. *RNA 3D Structure Analysis and Prediction*. Leontis,N. and Westhof,E. (eds). 27nd edn. Springer, Berlin, Heidelberg , Vol. **27**, pp. 299–318.

42. Bergonzo,C. and Cheatham,T.E. 3rd (2017) Mg(2+) binding promotes SLV as a scaffold in Varkud satellite ribozyme SLI-SLV kissing loop junction. *Biophys. J.*, **113**, 313–320.

43. Patriksson,A. and van der Spoel,D. (2008) A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.*, **10**, 2073–2077.

44. Henriksen,N.M., Roe,D.R. and Cheatham,T.E. (2013) Reliable oligonucleotide conformational ensemble generation in explicit solvent for force field assessment using reservoir replica exchange molecular dynamics simulations. *J. Phys. Chem. B.*, **117**, 4014–4027.

45. Bailor,M.H., Mustoe,A.M., Brooks,C.L. and Al-Hashimi,H.M. (2011) 3D maps of RNA interhelical junctions. *Nat. Protoc.*, **6**, 1536–1545.

46. Afgan,E., Baker,D., van den Beek,M., Blankenberg,D., Bouvier,D., Cech,M., Chilton,J., Clements,D., Coraor,N., Eberhard,C. *et al.* (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.*, **44**, W3–W10.

47. Bembon,O. (2018) seqLogo: sequence logos for DNA sequence alignments. R package version 1.46.0. https://rdrr.io/bioc/seqLogo/.

48. Desjardins,G., Bonneau,E., Girard,N., Boisbouvier,J. and Legault,P. (2011) NMR structure of the A730 loop of the *Neurospora* VS ribozyme: insights into the formation of the active site. *Nucleic Acids Res.*, **39**, 4427–4437.

49. DasGupta,S., Suslov,N.B. and Piccirilli,J.A. (2017) Structural basis for substrate helix remodeling and cleavage loop activation in the Varkud satellite ribozyme. *J. Am. Chem. Soc.*, **139**, 9591–9597.

50. 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.