Published online 28 October 2021

Nucleic Acids Research, 2021, Vol. 49, No. 20 ***11959–11973***
https://doi.org/10.1093/nar/gkab963

An integrative NMR-SAXS approach for structural determination of large RNAs defines the substrate-free state of a *trans*-cleaving *Neurospora* Varkud Satellite ribozyme

Pierre Dagenais, Geneviève Desjardins and Pascale Legault*

Department of Biochemistry and Molecular Medicine, Université de Montréal, Box 6128, Downtown Station, Montreal, QC H3C 3J7, Quebec, Canada

Received February 04, 2021; Revised September 01, 2021; Editorial Decision September 29, 2021; Accepted October 26, 2021

ABSTRACT

The divide-and-conquer strategy is commonly used for protein structure determination, but its applications to high-resolution structure determination of RNAs have been limited. Here, we introduce an integrative approach based on the divide-and-conquer strategy that was undertaken to determine the solution structure of an RNA model system, the *Neurospora* VS ribozyme. NMR and SAXS studies were conducted on a minimal *trans* VS ribozyme as well as several isolated subdomains. A multi-step procedure was used for structure determination that first involved pairing refined NMR structures with SAXS data to obtain structural subensembles of the various subdomains. These subdomain structures were then assembled to build a large set of structural models of the ribozyme, which was subsequently filtered using SAXS data. The resulting NMR-SAXS structural ensemble shares several similarities with the reported crystal structures of the VS ribozyme. However, a local structural difference is observed that affects the global fold by shifting the relative orientation of the two three-way junctions. Thus, this finding highlights a global conformational change associated with substrate binding in the VS ribozyme that is likely critical for its enzymatic activity. Structural studies of other large RNAs should benefit from similar integrative approaches that allow conformational sampling of assembled fragments.

INTRODUCTION

The concept of divide and conquer consists in approaching a complex problem by breaking it down into smaller, simpler parts that can be solved independently. It represents a modern problem-solving strategy that is commonly applied in a variety of scientific disciplines, such as engineering, computer science, biology and medicine. It is perfectly suited to structural biology when determining the high-resolution structure of a complete biological macromolecule is too challenging. In such cases, a divide-and-conquer approach can provide significant advances in understanding the biomolecule of interest through determination of high-resolution structures of individual domains in isolation and integration of the resulting information into a global structural model. Such an approach appears perfectly suited to investigate RNA molecules given their modular nature and the fact that high-resolution structural investigation of many functional RNAs in their entirety is not manageable due to their large size and highly dynamic nature (1–3). However, few structural studies have focused on reconstructing the structure of large RNA systems based on high-resolution structures of small subdomains (4–8). Given the widespread importance of RNA molecules in biological processes, there is a critical need for methodological advances in this area.

Despite significant improvements in RNA structure determination, there is a limited number of high-resolution structures of functional RNAs available in the database, especially for large RNAs (>60 nucleotides). This is due in part to the inherent nature of RNA molecules, since many are dynamic and adopt several conformational states that are crucial to their function (1–3,9,10). The limited number of RNA structures also results from technical limitations of current structural biology methods. Obtaining diffracting crystals of RNA for high-resolution structure determination by X-ray crystallography (11) can be a challenge and high-resolution structural investigations of large RNAs by NMR spectroscopy is generally limited to small domains (<50 nucleotides) due to signal degeneracy and faster relaxation in larger molecules (12,13). Although recent

*To whom correspondence should be addressed. Tel: +1 514 343 7326; Email: pascale.legault@umontreal.ca
Present address: Geneviève Desjardins, Zymeworks Inc., 1385 West 8th Avenue, Suite 540, Vancouver, BC V6H 3V9, Canada.

advances in cryo-EM methods are extremely promising, cryo-EM maps do not always display the necessary resolution to allow high-resolution structure determination (11,14,15). Thus, there is a great need to develop hybrid approaches that combine low-resolution and high-resolution methods in order to define the structure and dynamics of functional RNAs at the atomic level (16). Notably, it has been shown that small angle X-ray scattering (SAXS) data can provide a complementary source of structural information when used in combination with solution NMR spectroscopy (17). For example, residual dipolar coupling (RDC) measurements have been combined with SAXS data to probe the structure of several large RNAs and to generate models displaying their global architecture in solution (18–23). SAXS data have also been used to validate the global folding of NMR-derived RNA structures (4,6) or to filter structural ensembles presenting multiple structural models (24–26). Since SAXS data provide global structural information and NMR data give structural details at the atomic level, combining SAXS studies with high-resolution NMR studies has great potential to allow both high-resolution structure determination and dynamic characterisation of RNAs in solution. Moreover, given the general importance of Mg^{2+} ions in RNA folding and the fact that only a few solution structures of large RNAs have been determined in the presence of Mg^{2+} (7,15,18,21), there is a critical need to examine the role of Mg^{2+} ions as part of these studies.

The *Neurospora* Varkud satellite (VS) ribozyme is a small nucleolytic ribozyme (27,28) that has been extensively characterized over the years and has proven to be a versatile model system to study RNA structure and function. Importantly, it serves as an excellent system to illustrate how dynamics of structural subdomains contribute to define multiple conformational states that are important for RNA function (29–31). The VS ribozyme was originally found in the mitochondria of natural isolates of *Neurospora* and catalyzes both its self-cleavage and self-ligation at a specific phosphodiester bond. *In vitro*, the minimal functional VS ribozyme is composed of six helical segments (I–VI; Figure 1A) that fold into two separate domains: stem–loop I is the substrate, which contains the cleavage site and can be separated from the rest of the ribozyme, whereas helical domains II–VI form the catalytic domain, or *trans* ribozyme, and are organized around two three-way junctions. Several experimental approaches, including biochemical studies (32–36), FRET (30,37,38), SAXS (39) and NMR methods (40,41), have been previously used to define the general organization of helical domains within the global fold of the VS ribozyme [reviewed in (42)]. More recently, crystal structures of an extended VS ribozyme were reported (43,44) in which the ribozyme was trapped as a symmetric dimer, where the substrate domain of each protomer interacts with the catalytic domain of the other protomer. These structures capture the two tertiary contacts by which the stem–loop I substrate interacts with the catalytic domain, namely its highly stable kissing-loop interaction with stem–loop V and its loop-loop interaction with the A_{730} loop of stem VI that forms the active site. Additionally, in an effort to structurally characterize the *trans* ribozyme in solution, an NMR-based divide-and-conquer approach was undertaken in our laboratory. High-resolution structures were determined by NMR spectroscopy for several important subdomains of the *trans* ribozyme (Figure 1B), including its two three-way junctions (40,41,45,46). This divide-and-conquer approach initially allowed for the reconstruction of a three-dimensional model of the VS ribozyme that provided useful information on the global folding of the RNA (40,42), but lacked the long-range structural information necessary for refinement of a high-resolution structure.

In this study, we present an integrative approach based on the divide-and-conquer strategy for determining high-resolution structures of large RNAs that combines NMR and SAXS data of both the full-length RNA and its various subdomains. This NMR-SAXS approach, which allows conformational sampling by fragment assembly, was applied to determine the high-resolution structure of a minimal *trans* VS ribozyme in solution and to investigate the role of Mg^{2+} in its folding. It is a multi-step procedure that first involves pairing high-resolution NMR structures with SAXS data to obtain structural subensembles of the various subdomains. These subdomain structures are then combined to build a large ensemble of structures for the *trans* ribozyme and, finally, structures from this ensemble are filtered based on SAXS data to derive a final high-resolution structural ensemble. The resulting hybrid NMR-SAXS structures of the minimal *trans* VS ribozyme in solution are similar to the previous crystal structures but display an important structural difference and, thus, define a distinct conformational state that is likely critical for ribozyme function.

MATERIALS AND METHODS

New plasmid DNA template

For synthesis of the TR4P ribozyme, a plasmid template was created by inserting a double-stranded PCR fragment encoding TR4 (TR4P ribozyme precursor with a 19-nt extension at the 3′-end: AAG GGC GUC GUC GCC CCG A-3′) flanked by a T7 promoter into a pTZ19R-derived vector using unique HindIII and EcoRI sites. The purified plasmids were fully linearized by Ava I digestion (New England Biolabs) prior to transcription.

Preparation of $SLI_{20}d1$

The non-cleavable substrate $SLI_{20}d1$ was obtained from Dharmacon. It was deprotected, purified by denaturing gel electrophoresis and anion exchange chromatography (47), concentrated, exchanged in NMRG buffer (10 mM sodium cacodylate pH 6.5, 50 mM KCl, 0.05 mM NaN_3 and 10% D_2O) using a 3-kDa-cutoff ultrafiltration device (Millipore) and stored at –80°C.

Preparation of VS ribozyme subdomains

The SL6, J345 and J236 RNAs were transcribed *in vitro* and purified by denaturing gel electrophoresis, as previously described (40,41,48).

Preparation of the TR4P ribozyme

TR4P was transcribed *in vitro* using T7 RNA polymerase (prepared in house) in 40 mM Tris–HCl (pH 7.6), 1 mM

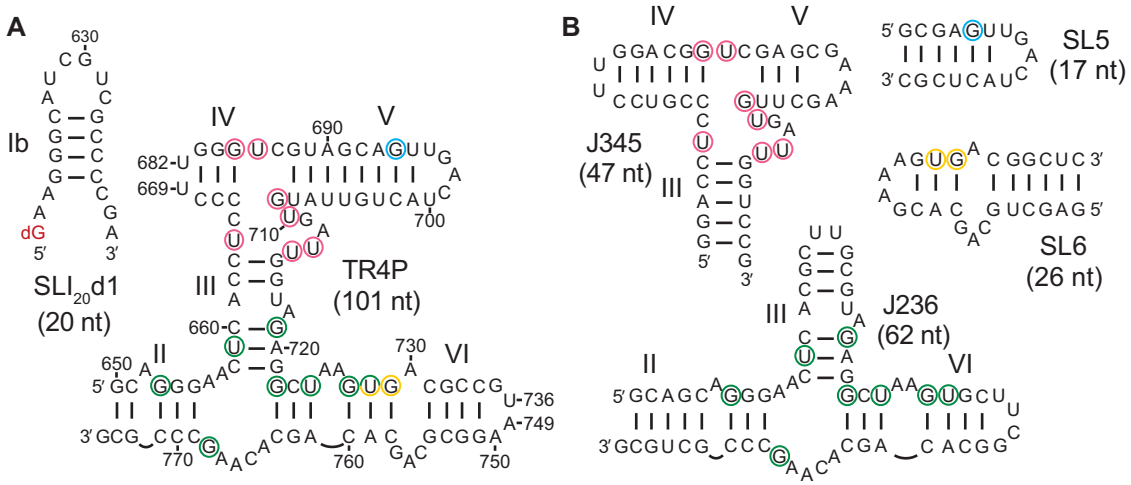


Figure 1. Sequence and secondary structures of VS ribozyme subdomains for structural studies. (**A**) Non-cleavable substrate ($SLI_{20}d1$) and functional *trans* ribozyme (TR4P) investigated in this study. (**B**) Four selected VS ribozyme subdomains for which high-resolution NMR structures were previously determined in the presence of Mg^{2+} ions, J345 (41), J236 (40), SL5 (45) and SL6 (46). Residues that display signals with characteristic chemical shifts in the ^{1}H–^{15}N HSQC spectra of those subdomains are circled using the same color code for the subdomains (B) and the *trans* ribozyme (A). The number of nucleotides (nt) for each RNA is indicated in parenthesis.

spermidine, 0.01% Triton X-100, 25 mM $MgCl_2$, 25 mM DTT, 4 mM of each NTP (ATP, CTP, UTP and GTP), 160 μg/mL of linearized plasmid and incubating at 37°C for 5–6 h. Similar transcription reaction conditions were used for synthesis of ^{15}N-labeled TR4P, except that unlabeled NTPs were replaced by 8 mM ^{15}N-labeled NTPs prepared in house (49). The TR4P RNA was purified by electrophoresis using a 5% denaturing polyacrylamide gel, visualized by UV shadowing, excised and extracted from the gel by crush and soak using TEN buffer (10 mM Tris, 1 mM EDTA, 0.3 M NaCl, pH 7.6). To remove polyacrylamide contaminants, the RNA was further purified using ion exchange FPLC chromatography on a semi-preparative Dionex™ DNAPac PA100 column (9 × 250 mm) heated at 65°C, as previously described (47). The eluted RNA was collected on ice and pooled, after which it was concentrated and exchanged in NMRG buffer using 3-kDa-cutoff ultrafiltration devices (Millipore) and then stored at -80°C.

NMR spectroscopy studies

1D NMR data were collected on Varian ^{Unity}Inova 500 MHz and 600 MHz NMR spectrometers equipped with 5-mm inverse $^{1}H\{^{13}C/^{15}N\}$ triple resonance probes, whereas 2D and 3D NMR data were collected on a Bruker AVANCE NEO 700 MHz NMR spectrometer equipped with a 5-mm inverse $^{1}H\{^{13}C/^{15}N\}$ triple resonance (TXI) probe. The temperature calibration of NMR probes was performed using an external standard of methanol. The ^{1}H and ^{15}N chemical shifts were referenced to an external standard of 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) at 0.00 ppm (50). All NMR data were collected at 15°C unless otherwise specified. For 1D NMR studies, 1D ^{1}H WATERGATE spectra (51) were generally collected with pre-saturation of the cacodylate methyl resonance and, if needed, using ^{15}N decoupling. For the Mg^{2+} and substrate titrations monitored by 1D ^{1}H NMR, 0.5 mM samples of TR4P and $SLI_{20}d1$ were initially prepared in NMRG buffer. TR4P was heated at 37°C for 5 min and $SLI_{20}d1$ was heated at 95°C for 1 min, and both RNAs were then transferred to ice for at least 5 min. Mg^{2+}-ion titrations of TR4P and $SLI_{20}d1$ were performed by adding small volumes (2.5–7.5 μl) of concentrated solutions of $MgCl_2$ (99.995% trace metals from Sigma-Aldrich, MO) directly to the concentrated RNA samples to achieve final concentrations of 0.5, 1, 2, 5, 10 and 20 mM $MgCl_2$. Following the Mg^{2+} additions, the $SLI_{20}d1$ sample (∼300 μl) was titrated into the TR4P sample (∼300 μl) to achieve a 1:1 complex of $SLI_{20}d1$:TR4P. For 1D NMR studies of temperature stability and 2D/3D NMR data collection for resonance assignments, 1.7 mM ^{15}N-labeled TR4P samples in NMRG-5 buffer (NMRG buffer with 5 mM $MgCl_2$) were prepared by first exchanging in NMRG buffer, heating at 37°C for 5 min and snap cooling on ice, and then exchanging in NMRG-5 buffer by ultrafiltration. For imino $^{1}H/^{15}N$ resonance assignment, 2D imino-optimized flip-back WATERGATE ^{1}H–^{15}N HSQC (51–54), 2D ^{15}N-decoupled/cacodylate-presaturated WATERGATE ^{1}H–^{1}H NOESY (mixing times of 75 and 150 ms) (55) and 3D ^{15}N-edited WATERGATE NOESY-HSQC (mixing time of 100 ms) (51) spectra were collected. NMR spectra were processed with the NMRPipe/NMRDraw package (56) and analyzed with the CCPNMR suite (57).

SAXS data collection and analysis

For SAXS studies, each RNA sample (SL6, J345, J236 and TR4P) was first exchanged in SAXS buffer (50 mM MES pH 6.5 and 50 mM KCl) by ultrafiltration. The RNA concentration was adjusted to 20 μM, and the RNA was refolded by heating at 95°C (SL6, J345, J236) or 65°C (TR4P) for 2 min, followed by snap cooling on ice. RNAs were then exchanged in SAXS-5 buffer (SAXS buffer with 5 mM $MgCl_2$) by ultrafiltration, and, for each RNA, samples were prepared at the different concentrations used for SAXS data

acquisition, namely 0.5, 1.0, 2.0, 3.0, 5.0, 6.0 and 8.0 mg/ml. Ultrafiltration flowthroughs (exact buffer match) were kept for buffer subtraction as part of SAXS data collection and analysis.

All the parameters for SAXS data collection as well as the software used for data processing and analysis are detailed in Supplementary Table S1. Briefly, all SAXS data collection was performed at 20°C, either on the G-1 beamline at the Cornell High Energy Synchrotron Source (MacCHESS; for SL6, J345 and J236) (58) or on a local Xenocs BioXolver L from SAXSLab equipped with an Excillum MetalJet D2 + 70 kV using a 95% (weight) liquid gallium anode alloy (for TR4P). Initial image processing was achieved using BioXTAS RAW v.1.6.3 (59). For each sample, expositions of the SAXS dataset were analysed individually for radiation damage, averaged and underwent buffer subtraction. Guinier analysis was performed using ATSAS v.2.8.4. (60), and data from samples showing signs of aggregation or interparticle interference were discarded. For each RNA, datasets from the lowest and highest concentrations showing no concentration dependence were merged and averaged into a single scattering curve for further analysis. Pair distance (P(r)) distribution functions were calculated using GNOM (61). Maximum dimension (D_{max}) values were determined such that the $P(r)$ distribution profile is non-negative, displays zero values at P(0) and ends smoothly at D_{max}. Real-space R_g and $I_{(0)}$ values were validated to be within 10% of their Guinier-approximated values. Volume of correlation (V_c) values were calculated to determine and validate the molecular weight of each sample (62). Theoretical scattering curves of all models were calculated and fitted to the experimental scattering curve using CRYSOL (63) with default parameters and a q_{max} of 0.3 Å^{-1}. The ensemble optimization method (EOM) was performed with GAJOE 2.0 (64) on external pools of 10 000 molecules, using default parameters. The χ^2_{free} values were obtained by first calculating the number of Nyquist–Shannon points (n_s) required to completely reconstitute the SAXS signal using $n_s = (q_{max} \times D_{max})/\pi$ with q_{max} values of 0.3 Å^{-1} and D_{max} values inferred by GNOM. The SAXS data was then divided into n_s equal bins, and one random data point was selected from each bin. The resulting SAXS profiles were analysed by CRYSOL to calculate χ^2 values. The χ^2_{free} values were determined by taking the median over 2500 rounds (62).

DENSS *ab initio* electron density maps

DENSS (65) was used to calculate the *ab initio* electron density map directly from each GNOM outputs. For each molecule, 20 reconstructions were performed in slow mode using default parameters and averaged.

Fine-grained refinement of structural models

Structures of isolated VS ribozyme subdomains (SL5, SL6, J345 and J236) as well as final TR4P ribozyme models were refined using QRNAS v.0.3 (66) with default parameters and no additional restraint. QRNAS uses Born electrostatics and a modified version of the AMBER force field with additional energy terms, and those associated with explicit hydrogen bonds, ideal base-pair planarity and regularization of the backbone conformation were selected. The final conformer, generated after 100 000 refinement steps, was considered the refined structure.

Fragment assembly of the minimal VS ribozyme

The fragment assembly of TR4P was initiated by selecting from the NMR structures of isolated VS ribozyme subdomains (SL6, J345 and J236 with pdb entry codes 2L5Z, 2MTJ and 2N3Q, respectively) each individual state of the NMR ensemble and refining them using QRNAS. For each subdomain, the resulting ensemble of refined NMR structures was used to select individual conformers that best fit the SAXS data based on CRYSOL and GAJOE analyses. The conformer showing the best CRYSOL fit (smallest χ^2_{free}) and the conformers forming the best GAJOE ensemble (smallest χ^2) were used to generate subensembles of hybrid NMR-SAXS structures. The conformer pool used as an input for the genetic algorithm of GAJOE was generated by copying each of the individual states of a given refined NMR structure an equal number of times to create an ensemble of 10,000 conformers. Given that SAXS data were not collected for the SL5 subdomain, each SL5 structure from the NMR ensemble was minimized using QRNAS and three representative conformers were selected to create a subensemble of SL5 structures.

The assembly of the TR4P structure was performed by aligning overlapping helical segments present in the NMR structures of isolated subdomains and by merging the structures using PyMOL-Python scripts. The process was carried out in a pairwise fashion, aligning J236 onto J345, then onto SL5 and SL6 conformers to generate all possible conformer combinations of different TR4P ribozyme models. Point mutations were performed to match the exact TR4P ribozyme sequence by using the PyMOL mutagenesis wizard. The final 108 NMR-based models were then refined using QRNAS. A crystal-based model of TR4P was also created in PyMOL from the crystal structure of the VS ribozyme (pdb entry 4R4V) by deletion of helical regions and point mutations to match the TR4P ribozyme sequence. QRNAS was used to refine the final structure.

Hybrid NMR-SAXS structure determination of the TR4P ribozyme

The ensemble of 108 refined NMR-based models of the TR4P ribozyme was filtered to first discard models with steric clashes by using a modified version of the pdbcheck script from the pdbremix library (https://github.com/boscoh/pdbremix). χ^2_{free} values for each selected model were then calculated using CRYSOL and TR4P SAXS data collected in 5 mM $MgCl_2$. The final hybrid NMR-SAXS structural ensemble contains the 10 models exhibiting the best χ^2_{free} values. A detailed structural analysis of these models was performed using PyMOL-Python scripts to confirm the presence of important features in the NMR structures of the J236 and J345 subdomains (Supplementary Table S2).

RESULTS

TR4P: a minimal *trans* VS ribozyme for structural studies in solution

The TR4P RNA (101 nucleotides or 33 kDa; Figure 1A) was selected for these studies because it represents the smallest known RNA sequence derived from the natural VS ribozyme that retains cleavage activity of an isolated hairpin substrate *in vitro* (67). TR4P is an efficient *trans* cleaving ribozyme that has been successfully used in the preparation of several RNAs, including the J345 and J236 subdomains, by processing RNA transcripts containing a VS ribozyme substrate at their 3′-end (40,41). Accordingly, TR4P was efficiently produced by self-cleavage from its precursor, TR4, directly in the transcription reaction (Supplementary Figure S1A) (67). The sequence of TR4P is well defined at both its 5′- and 3′-ends; the 5′-sequence homogeneity is imparted by the identity of the first three nucleotides (5′-GCA) (68), whereas the 3′-sequence homogeneity results from precise and efficient self-cleavage. As a result, the purification of TR4P by denaturing gel electrophoresis produced a highly pure RNA with good yield (Supplementary Figure S1A). Subsequent purification by anion exchange chromatography under denaturing conditions removed polyacrylamide contaminants (69) and significantly improved detection of the imino signals in the 1D ^{1}H NMR spectrum of the TR4P RNA (Supplementary Figure S1B).

Mg^{2+}-dependent folding and substrate binding of a minimal VS ribozyme by NMR spectroscopy

To investigate the effect of Mg^{2+} ions on the structure of the minimal *trans* ribozyme, 1D imino ^{1}H spectra of TR4P were collected following addition of increasing concentration of Mg^{2+} ions. Formation of stable hydrogen bonds involving the imino protons of G and U residues in RNA slows down their exchange rates with the solvent such that they are readily observable by ^{1}H NMR spectroscopy. As a result, observation of imino ^{1}H signals in RNA can be interpreted in terms of stable base-pairing interactions, thus providing valuable information on secondary and tertiary structures. Several imino signals are observed for TR4P in the absence of Mg^{2+} ions, which is compatible with formation of many helical segments defining its secondary structure (Figure 2A). However, the addition of Mg^{2+} ions (5–20 mM) leads to significant changes in the imino ^{1}H region (Figure 2A), including an overall better signal dispersion and two additional downfield-shifted signals (between 14.0 and 15.5 ppm), as previously observed upon Mg^{2+}-dependent folding of the J236 subdomain (40). These 1D imino ^{1}H NMR data confirm that Mg^{2+} ions are required for formation of a stable tertiary structure for the TR4P ribozyme.

To investigate the effect of substrate binding on the structure of the minimal *trans* ribozyme, a non-cleavable SLI substrate was titrated into the TR4P sample. For these studies, a minimal substrate carrying a 2′-deoxy modification at the cleavage site was used (SLI_{20}d1; Figure 1A). TR4P efficiently cleaves the non-modified form of this substrate (SLI_{20}; data not shown). As shown in a previous study, SLI_{20}d1 adopts a stable hairpin structure in the presence of Mg^{2+} ions with four well-resolved signals in the 1D imino ^{1}H spectrum (Figure 2B) (70). Titration of the Mg^{2+}-folded SLI_{20}d1 to TR4P leads to significant changes in the 1D imino ^{1}H spectrum (Figure 2B), including an overall broadening of the signals, as expected due to the increase in molecular weight, as well as the appearance of two new signals at 9.8 and 15 ppm that were not previously observed in the stable complex formed between SLI_{20} and an isolated stem–loop V (SLV) (70). Specific changes in imino ^{1}H signals associated with formation of the kissing-loop interaction previously observed with the SLI_{20}/SLV complex could not be detected for the SLI_{20}d1/TR4P complex due to signal overlap in the 11–14.5 ppm region (70). However, the imino ^{1}H spectrum of the SLI_{20}d1/TR4P complex is quite different from the mathematical addition of the individual spectrum of its two components (Figure 2B), which provides strong evidence that complex formation is accompanied by structural changes not previously observed in the SLI_{20}/SLV complex. Taken together, the 1D ^{1}H imino NMR data are compatible with formation of a 40-kDa SLI_{20}d1/TR4P complex in the presence of Mg^{2+} ions.

Validation of the divide-and-conquer approach by NMR spectroscopy

To validate the divide-and-conquer approach for structure determination of a minimal VS ribozyme, NMR spectra of TR4P were collected under optimized conditions, namely with a free Mg^{2+} concentration of 5 mM at 15°C. To obtain a free Mg^{2+} concentration of 5 mM, the RNA sample was prepared by repeated exchanges in a buffer containing 5 mM $MgCl_2$, instead of adding Mg^{2+} ions directly to the concentrated RNA sample as in our initial experiments (Figure 2). This resulted in narrower imino ^{1}H linewidths and, thus, higher quality NMR data (Supplementary Figure S2A). In addition, the RNA sample was kept at low temperatures (<20°C) to ensure data quality, since incubation at higher temperatures led to irreversible NMR line broadening and decreased signal intensity (Supplementary Figure S2B). The 2D ^{1}H–^{15}N HSQC spectrum of Mg^{2+}-folded TR4P (Figure 3A) is very well-resolved and displays the expected number of signals (∼35). This provides evidence for formation of a single stable conformation in solution for this minimal VS ribozyme. For NMR signal assignment, a 2D ^{1}H–^{1}H NOESY spectrum was collected, which allowed for detection of several imino-to-imino crosspeaks (Figure 3B and Supplementary Figure S3). The 2D NOESY data combined with known chemical shifts for characteristic imino signals of the various VS ribozyme subdomains allowed for unambiguous assignment of all imino signals in the 2D ^{1}H–^{15}N HSQC spectrum (Figure 3 and Supplementary Table S3). Spin-diffusion imino-to-imino NOESY crosspeaks were particularly important to validate assignments of imino signals in complex structural elements (Supplementary Figure S3). For example, the spin-diffusion crosspeaks between the G_{719} and G_{721} imino signals and between the U_{659} and G_{768} imino signals confirm their assignment as well as the continuous stacking between several base pairs at the heart of the II-III-VI junction (the C_{660}–G_{719}, U_{659}–A_{720}, C_{658}–G_{721} and G_{768}–A_{656} base pairs). In addition, the unambiguous assignment of the two most downfield-shifted imino signals

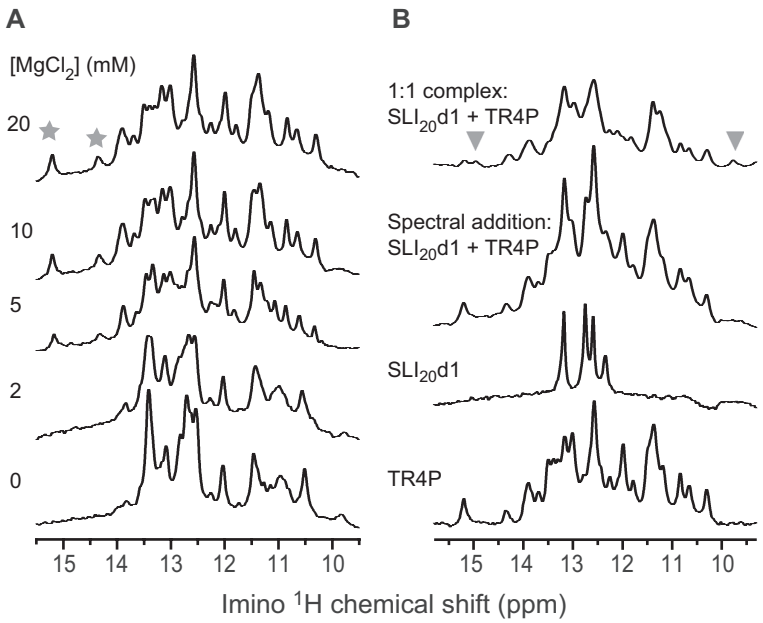


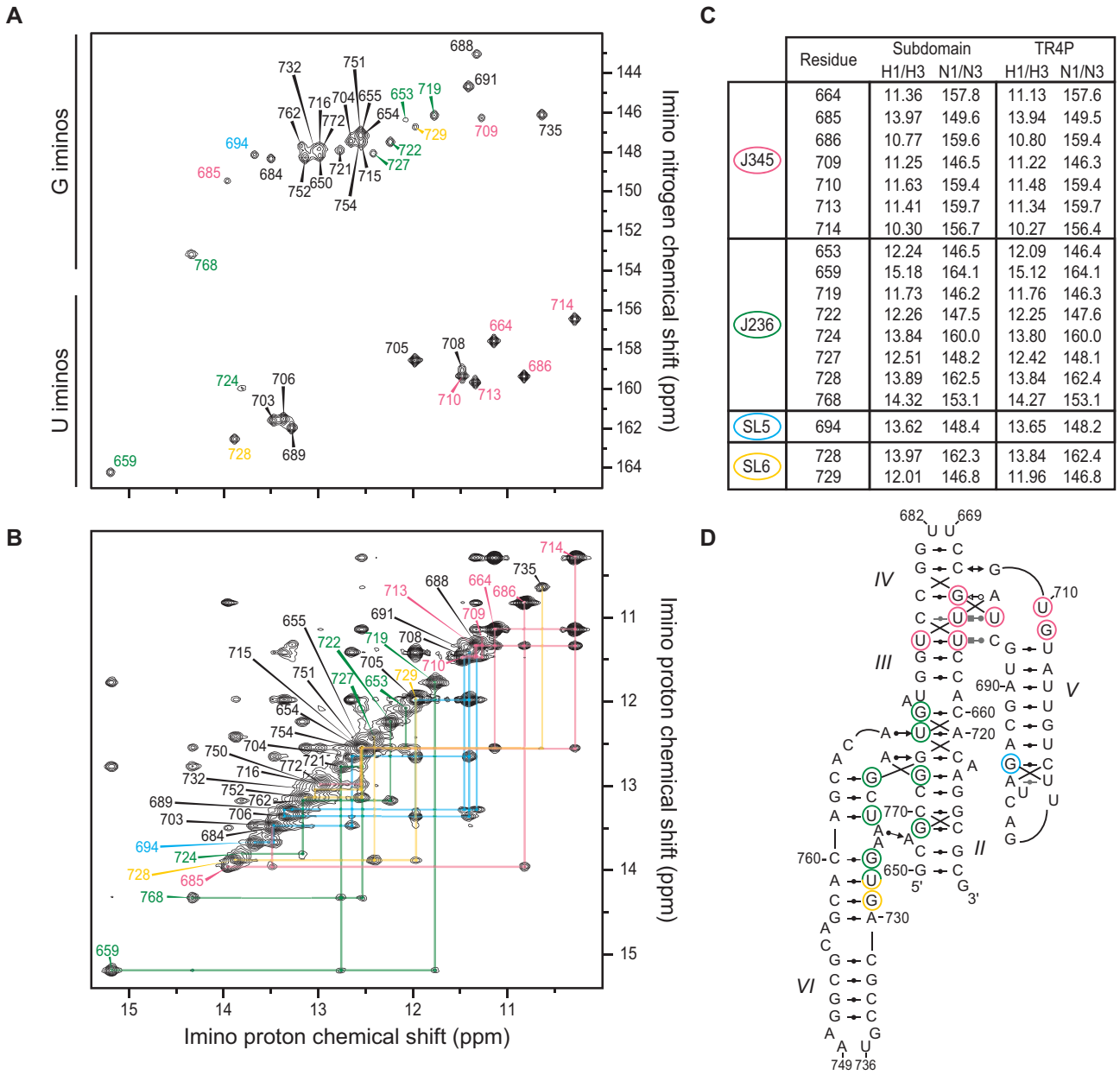
Figure 2. Mg^{2+}-dependent folding and substrate binding of the TR4P ribozyme by 1D imino NMR studies. (**A**) Effect of Mg^{2+} ions on the 1D imino ^{1}H NMR spectrum of 0.5 mM TR4P in NMRG buffer. Small aliquots (~5 μl) of concentrated $MgCl_2$ were added directly to the NMR sample. The two signals annotated with a grey star at the highest Mg^{2+} concentration are characteristic signals of a Mg^{2+}-folded II–III–VI junction. (**B**) Formation of a 1:1 SLI_{20}d1:TR4P complex followed by 1D imino ^{1}H NMR spectra starting with 0.5 mM TR4P and titrating 0.5 mM SLI_{20}d1, both pre-folded with Mg^{2+} ions. Shown here is a summary of the substrate titration data, with the 1D ^{1}H spectra of the free TR4P and SLI_{20}d1, the simulated 1D ^{1}H spectrum resulting from the mathematical addition and line broadening of these spectra, and the observed 1D ^{1}H spectrum of the 1:1 complex. Two imino ^{1}H signals that appear upon complex formation are indicated with grey arrows.

to U_{659} and G_{768} of the II–III–VI junction confirms that stable folding of this junction within TR4P depends on the presence of Mg^{2+} ions. Of note, all imino-to-imino NOEs were assigned for TR4P based on its secondary structure and the expected signals observed for the isolated subdomains, and, thus, no additional NOEs were observed that could have provided evidence for changes in the structure of subdomains or unexpected interactions between subdomains.

To ascertain the structural similarity of VS ribozyme subdomains between their isolated form and in the context of TR4P, ^{1}H and ^{15}N chemical shift values of characteristic signals were compared, and it was found that they vary by less than 0.23 ppm and 0.3 ppm, respectively (Figure 3C). Superposition of the 2D ^{1}H–^{15}N HSQC spectrum of TR4P on the spectrum of either the J345 or the J236 subdomain further illustrates the chemical shift similarity for characteristic signals (Supplementary Figure S4). Together, these observations help validate the chemical shift assignments and provide evidence of structural similarity for several subdomains in isolation and within TR4P (Figure 3D), thereby validating the divide-and-conquer approach used here for the structure determination of TR4P.

SAXS studies of TR4P and VS ribozyme subdomains

To complement the NMR results, SAXS studies were conducted on TR4P as well as on the individual VS ribozyme subdomains. Scattering data were collected for TR4P in the presence of 0, 5 and 20 mM $MgCl_2$ to investigate the effect of the free Mg^{2+} ion concentration on its global folding (Supplementary Figure S5A). The values for the radius of gyration (R_g) derived from Guinier analysis are significantly different at 0 mM (35.8 Å) and 5 mM (29.4 Å) $MgCl_2$, which confirms that the tertiary folding of TR4P is Mg^{2+} dependent (Table 1). A similar conclusion is reached when comparing the R_g and maximal dimension (D_{max}) values derived from the pair distance distribution function [$P(r)$] (Table 1). Moreover, the dimensionless Kratky plot is compatible with TR4P adopting an extended state in the absence of Mg^{2+} ions and a more compact shape in the presence of Mg^{2+} ions (Supplementary Figure S5B). Intriguingly, the distinct Kratky profiles observed at 5 mM and 20 mM $MgCl_2$ indicate that TR4P adopts different conformations under these two conditions. Since the Porod–Debye plots at 5 and 20 mM $MgCl_2$ present similar profiles with a clear plateau, both distinct from the plateau-less profile observed in the absence of Mg^{2+} ions (Supplementary Figure S5C), this suggests greater flexibility of the Mg^{2+}-free state, while ruling out the possibility of increased flexibility at the highest Mg^{2+} ion concentration (71). Taken together with the increase in R_g values from 29.4 to 34.4 Å when going from 5 to 20 mM $MgCl_2$, this suggests that TR4P forms a monomer at 5 mM Mg^{2+} but partially multimerizes at higher Mg^{2+} ion concentrations. Complementary native gel analyses of TR4P are consistent with its main conformation being monomeric (Supplementary Figure S6). However, when TR4P is folded at 20 mM $MgCl_2$, about 10–20% of the RNA adopts a multimeric form (Supplementary Figure S6B), which is barely detectable after folding the RNA under the same conditions but with 5 mM $MgCl_2$ (Supplementary Figure S6A, lane 3). This partial RNA multimer-



	Residue	Subdomain H1/H3	Subdomain N1/N3	TR4P H1/H3	TR4P N1/N3
J345	664	11.36	157.8	11.13	157.6
	685	13.97	149.6	13.94	149.5
	686	10.77	159.6	10.80	159.4
	709	11.25	146.5	11.22	146.3
	710	11.63	159.4	11.48	159.4
	713	11.41	159.7	11.34	159.7
	714	10.30	156.7	10.27	156.4
J236	653	12.24	146.5	12.09	146.4
	659	15.18	164.1	15.12	164.1
	719	11.73	146.2	11.76	146.3
	722	12.26	147.5	12.25	147.6
	724	13.84	160.0	13.80	160.0
	727	12.51	148.2	12.42	148.1
	728	13.89	162.5	13.84	162.4
	768	14.32	153.1	14.27	153.1
SL5	694	13.62	148.4	13.65	148.2
SL6	728	13.97	162.3	13.84	162.4
	729	12.01	146.8	11.96	146.8

Figure 3. Validation of the divide-and-conquer approach by 2D imino NMR studies of the Mg^{2+}-folded TR4P ribozyme. (**A**) Assigned 2D ^{1}H–^{15}N HSQC spectrum of 1.7 mM ^{15}N-labeled TR4P in NMRG-5 buffer containing 5 mM $MgCl_2$ at 15°C. (**B**) Assigned 2D ^{15}N-decoupled ^{1}H–^{1}H NOESY spectrum of the same sample collected at 15°C with a mixing time of 150 ms. (**C**) Chemical-shift comparison of characteristic imino signals observed in the isolated VS ribozyme subdomains and TR4P. The chemical shift values were measured either at 25°C (J345, J236, SL5, TR4P) or 15°C (SL6). (**D**) Tertiary structure schematics of TR4P based on high-resolution structure of individual subdomains. Residues with characteristic imino signals are circled in (D) and annotated in (A), (B) and (C) using the same color code. Base pairs are depicted using the Leontis–Westhof notation (80), either in black (≥2 hydrogen bonds) or grey (one hydrogen bond).

ization is consistent not only with the increased R_g values at 20 mM $MgCl_2$, but also with line broadening of imino NMR signals under non-ideal conditions (Supplementary Figure S2) and the natural tendency of some VS ribozyme sequences to form dimers, as the observed in crystal structures (43,44). Thus, a free $MgCl_2$ concentration of 5 mM was used for the remainder of the SAXS investigations.

SAXS data collected on TR4P and isolated VS subdomains (J345, J236 and SL6; Figure 4A) result in a linear fit in their respective Guinier region (Figure 4B), demonstrating that all samples were monodisperse as well as being free of contaminants and interparticle interference that would significantly contribute to the scattering. Dimensionless Kratky profiles (Figure 4C) are consistent with well-folded RNA molecules in solution adopting a flattened prolate spheroid shape (72), resembling the shape of a flat American football. $P(r)$ distributions (Figure 4D) of each molecule are also well behaved, displaying zero values at $P(0)$ and ending smoothly at D_{max}. In addition, molecular weight values estimated from the volume-of-correlation

Table 1. SAXS results and parameters for selected subdomains of the VS ribozyme and the TR4P ribozyme

	SL6	J345	J236	TR4P		
Mg^{2+} ion concentration	5 mM	5 mM	5 mM	0 mM	5 mM	20 mM
MW chemical composition (Da)	8 934	16 010	21 184	34 455	34 455	34 455
MW estimates from V_c (Da)	8.10E + 03	1.80E + 04	2.20E + 04	3.65E + 04	3.30E + 04	3.67E + 04
R_g from Guinier analysis (Å)	16.4	20.3	21.6	35.8	29.4	34.4
R_g from $P(r)$ analysis (Å)	17.1	20.9	22.3	37.6	30.7	34.9
D_{max} (Å)	61	75	77	124	110	114

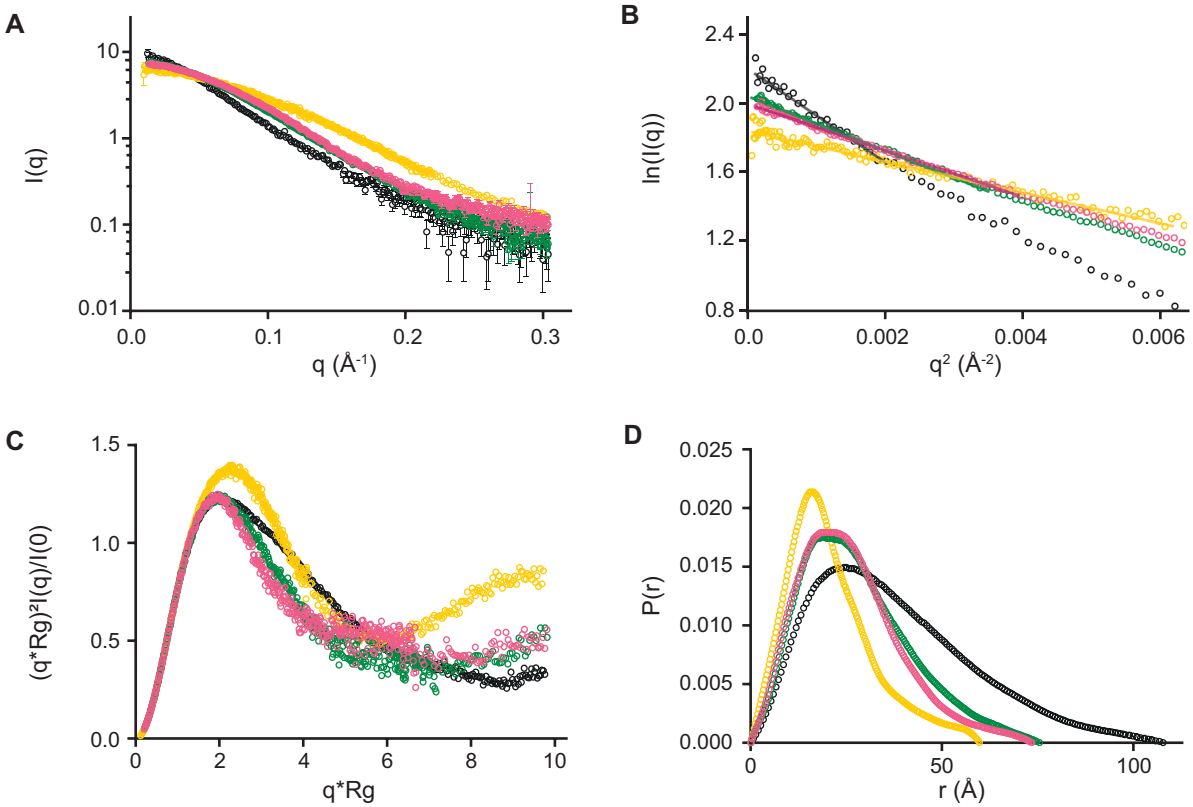


Figure 4. SAXS results of TR4P and isolated VS ribozyme subdomains collected in NMRG-5 buffer containing 5 mM $MgCl_2$ at 20°C. (**A**) $I(q)$ versus q plot. (**B**) Guinier plot with linear regression for $q^*R_g < 1.3$ ($R^2 \geq 0.98$). (**C**) Dimensionless Kratky plot and (**D**) Pr distribution. TR4P (dark grey), SL6 (gold), J345 (magenta) and J236 (green) SAXS data are displayed.

(V_c) (62) of each molecule are comparable to the molecular weights calculated based on their respective chemical composition, confirming their monomeric states (Table 1). The R_g values estimated from each Guinier fit are also within 10% of the real-space values derived from their respective $P(r)$ distributions. Overall, the SAXS data presented here are of high quality and can be combined with high-resolution NMR data to derive high-resolution structural ensembles of a functional VS ribozyme in solution.

Strategy to determine hybrid NMR-SAXS structural ensembles

To obtain high-resolution structural ensembles of TR4P in solution, we established a general strategy combining high-resolution NMR structures, structural refinement and SAXS data (Figure 5). As a starting point, NMR-derived structural ensembles of each VS ribozyme subdomain (J345, J236, SL6 and SL5) were refined using QRNAS (66), a convenient software for fine-grained refinement of nucleic acids. Then, the resulting ensemble of high-quality NMR structures of the subdomains that contain an internal loop and/or a helical junction (J345, J236 and SL6) were paired with the SAXS data to generate a subensemble that best fit the data. To allow for a high degree of conformational heterogeneity within the structural subensembles of each subdomain while still limiting their size, two criteria were used to select conformers. First, we selected the conformer presenting the best χ^2_{free} value when comparing the calculated scattering curve to the experimental data using CRYSOL (62). Second, we selected the ensemble of conformers that best describe the experimental data using the GAJOE ensemble optimization method (EOM) (64). Overall, these hybrid NMR-SAXS structural ensem-

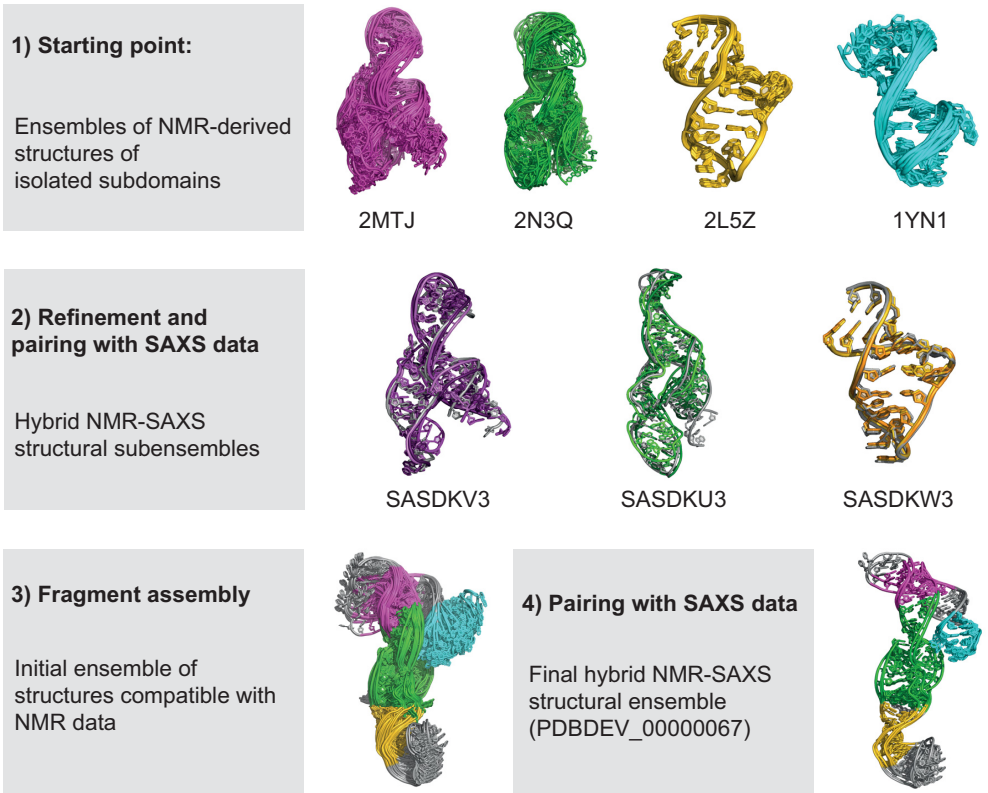


Figure 5. Fragment assembly strategy. Ensembles of NMR-derived structures of isolated subdomains are refined and paired with SAXS data to form hybrid NMR-SAXS subensembles. Fragment assembly is performed by combining each conformer present in the subensembles in a pairwise fashion. The final TR4P ribozyme models are then paired with SAXS data. The 10 models exhibiting the lowest χ^2_{free} values are selected to form the final hybrid NMR-SAXS structural ensemble of TR4P.

bles are in close agreement with experimental SAXS data, as demonstrated by both the χ^2 values and the visual fit of the calculated scattering curves to the SAXS data as well as the overall disposition of the structures within DENSS *ab initio* electron density maps (Figure 6) (65). Subsequently, structural models of TR4P were built by fragment assembly allowing all possible combinations of conformers from the hybrid NMR-SAXS structural ensembles of the J345 (four conformers), J236 (three conformers) and SL6 (three conformers) subdomains as well as from the refined NMR structural ensemble of SL5 (3 representative conformers). The resulting 108 structural models of TR4P were refined with QRNAS and then filtered to remove models with unresolved steric clashes. Finally, the remaining 71 structures were paired with the experimental SAXS data of TR4P, and since these structures did not fit the SAXS data equally well given the observed large range of CRYSOL χ^2_{free} values (1.08–5.96), only the 10 structures presenting the lowest χ^2_{free} values were selected to form the final hybrid NMR-SAXS structural ensemble. Thus, applying a steric clash filter and SAXS data fits allowed us to discriminate between the different RNA conformers obtained from fragment assembly. Our approach is similar to previous implementation of χ^2 fits to help discriminate between different conformational states (4,10,25), except that we added conformational sampling via fragment assembly.

Description of the NMR-SAXS structural ensemble

The NMR-SAXS structural ensemble of TR4P is well defined as illustrated by the superposition of the 10 best conformers and the pairwise heavy-atom RMSD of 2.7 ± 1.6 Å (Figure 7A). Moreover, these 10 best conformers are highly compatible with the experimental SAXS data, given the low CRYSOL χ^2_{free} values of 1.08–1.41. The overall structure of the *trans* VS ribozyme is well represented by the single best-fitting conformer (χ^2_{free} of 1.08), since the 10 best NMR-SAXS structures display only subtle differences in the relative orientations of their helical segments, their topologies being strikingly similar. The overall shape is best described by a flat, elongated ovoid, within which RNA helices combine in a single layer to create a double herringbone pattern. The main architectural frame of the ribozyme is formed by an extended helical segment resulting from the continuous stacking of stems II, III and IV. The two three-way junctions each form a V-shaped pattern, in which stems V and VI extend alongside stems III and II, respectively. As expected, a detailed structural analysis of the III–IV–V and II–III–VI junctions confirms that key structural features previously observed in the NMR structures of the J345 and J236 subdomains and in the crystal structures of the VS ribozyme are preserved in the 10 best TR4P structures (Supplementary Table S2). This includes the U-turn fold, four stacked base triples and a ribose zipper within the III–IV–V junc-

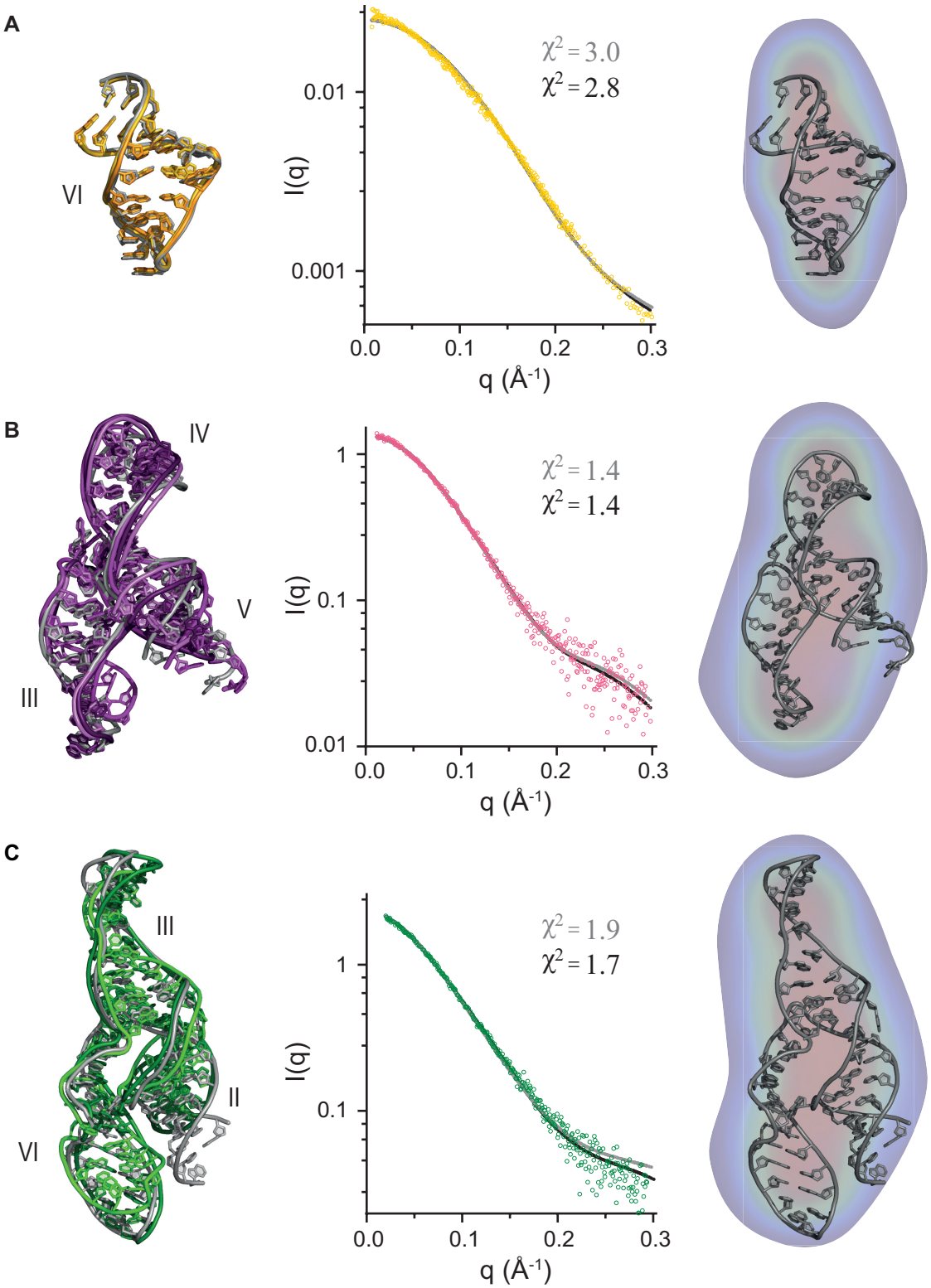


Figure 6. Hybrid NMR-SAXS structures of Mg^{2+}-folded subdomains of the VS ribozyme. Structures of the SL6 (**A**), J345 (**B**) and J236 (**C**) subdomains are shown with quality controls. Left panels: superposition of the selected NMR-SAXS subensembles of models. Conformers selected by the SAXS-oriented ensemble optimization method are shown in different colored shades, whereas the best-fit conformer selected by the χ^2_{free} method is in grey. Middle panels: experimental SAXS data (colored circles) along with the fit to the scattering curves calculated using the ensemble optimization method (black) and the scattering curves of the best-fit conformer selected by the χ^2_{free} method (grey). Right panels: experimental DENSS *ab initio* electron density map and its visual fit to the best-fit NMR-SAXS structure are presented. Contours of the density maps are as follows: 15σ (red), 10σ (yellow), 7.5σ (green), 5σ (cyan) and 0.8σ (blue).

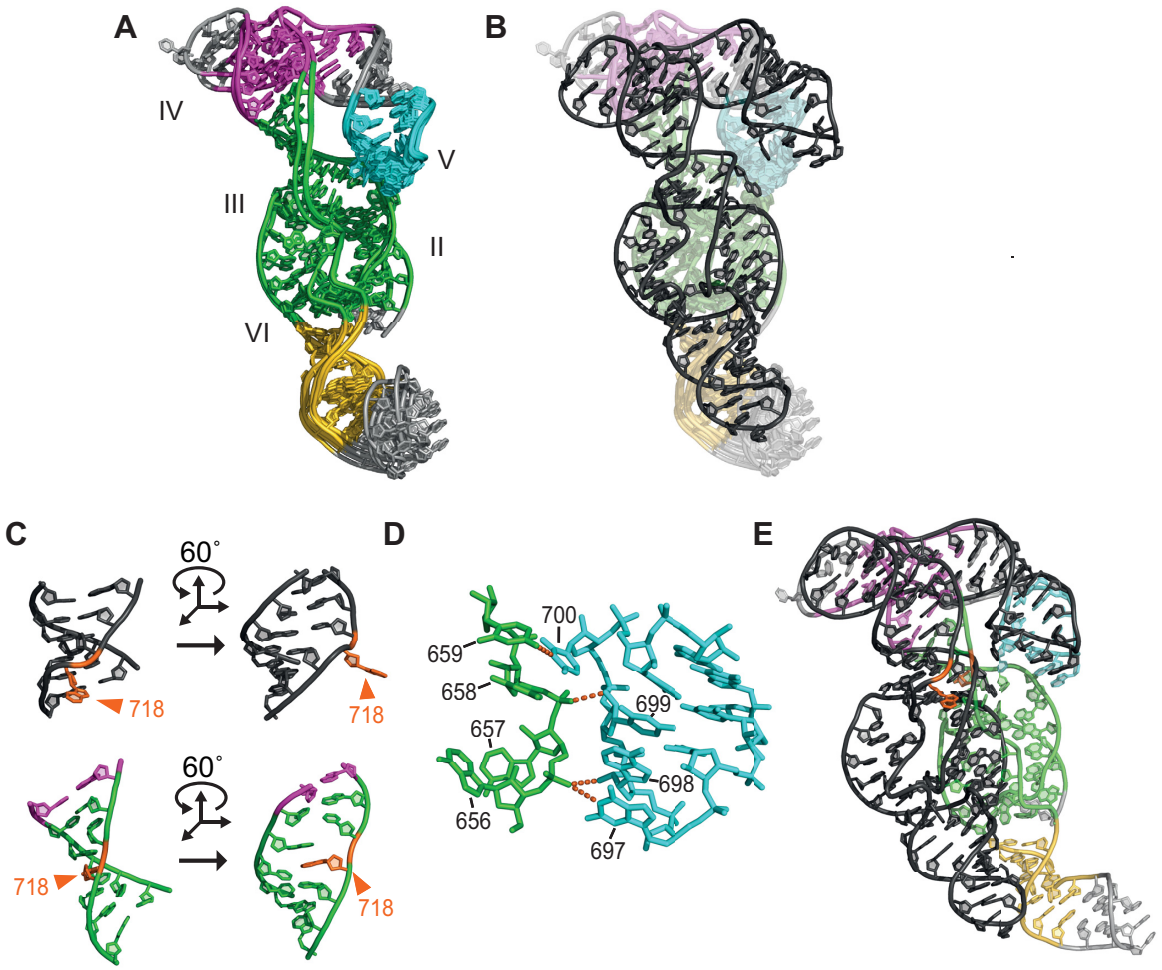


Figure 7. Hybrid NMR-SAXS structural ensemble of the Mg^{2+}-folded TR4P ribozyme and comparison with the crystal structure. (**A**) Superposition of the 10 conformers from the hybrid NMR-SAXS structural ensemble of TR4P exhibiting the best χ^2_{free} values (1.08–1.41). (**B**) Structural ensemble shown in (A) superposed to the crystal-based model of TR4P (dark grey; $\chi^2_{free} = 1.8$). This crystal-based model is derived from the crystal structure of the VS ribozyme (pdb entry 4R4V) that was modified to fit the TR4P sequence (see Materials and Methods). (**C**) Local conformational difference in stem III between a representative NMR-SAXS structure (bottom) and the crystal structure (top). (**D**) Interactions between the SLV loop and the II–III–VI junction observed for the best-fit conformer of the final NMR-SAXS structural ensemble of TR4P. Potential hydrogen bonds are indicated by dash lines (G697 N1 to A657 OP1, A698 2′-O to A657 OP1, C699 2′-O to C658 OP2, and U700 N3 to U659 O4). (**E**) Effect of the local conformational difference shown in (C) on the relative orientation of the III–IV–V and II–III–VI junctions illustrated by superposing only the III–IV–V moiety of the two molecules described in (C).

tion as well as the ACAA turn, two base-triples, a WC/WC A_{656}–G_{768} base pair and a long-range bulge-bulge interaction within the II–III–VI junction. Of note, in the NMR-SAXS structure of the TR4P ribozyme, the individual helical subdomains of the VS ribozyme are held rigidly with respect to one another. This fairly rigid structure results from both the well-defined NMR structures of the isolated subdomains and the selection of ribozyme conformations that are compatible with the SAXS data of the TR4P ribozyme.

To compare the NMR-SAXS structures with crystal structures derived from X-ray diffraction data, the crystal structure of an extended *cis* VS ribozyme (44) was modified to match the sequence of TR4P and then aligned onto the NMR-SAXS ensemble (Figure 7B). This crystal-based model of TR4P is compatible with the experimental SAXS data (χ^2_{free} values of 1.8), although the fit is not as good as for the NMR-SAXS structures (χ^2_{free} values of 1.08–1.41). The crystal-based model adopts the same general topology but a more compact fold than the NMR-SAXS structures, in which the double herringbone pattern is flatter and thus more clearly visible. This more compact fold largely originates from a local structural difference in stem III, which connects the two three-way junctions. Residue A_{718} of stem III is completely bulged out in the crystal structure (numbered A_{713} in pdb entry 4R4V) instead of stacked within the helix as in the NMR-SAXS structure of TR4P (Figure 7C) and the NMR structure of the J236 subdomain (40,42). In addition, we noted that the SLV loop interacts with the II–III–VI junction in 3 out of the 10 final structures (Figure 7D). This potential inter-domain interaction is likely transient since it is not supported by our imino NMR data.

Given the important structural difference in stem III between NMR-derived structures and the crystal structure, we reexamined the evidence for the stacked in conforma-

tion in the structures of both the isolated J236 junction and TR4P. For the isolated J236 junction, several NOE signals between base protons of A_{718} and nearby protons support the stacked in conformation observed in the NMR structure, including an NOE between the amino protons of A_{718} and the imino proton of G_{719} as well as cross-strand NOEs between A_{718} H2 and A_{661} H1' (40). Given that the NMR-SAXS structure of TR4P was built by fragment assembly using the structure of the isolated J236 subdomain, it is not surprising that this stacked in conformation is preserved in TR4P. However, a potential pitfall of divide-and-conquer approaches is that the subdomains may adopt different structures in their isolated form and in the context of the full molecule. Even though the structural context of the A_{718} bulge in stem III is identical in the isolated J236 subdomain and TR4P with the same four base pairs surrounding the bulge, there is still a possibility for unanticipated non-nearest neighbor effects that could affect the structure of the A_{718} bulge in TR4P. We have done our best to reduce this possibility by careful selection of subdomains and validation by NMR spectroscopy. For the A_{718} bulge specifically, three base pairs on each side of the bulge in TR4P (U_{659}-A_{720}, C_{660}-G_{719} and A_{661}-U_{717}) present the same imino NMR data signature in the isolated J236 domain: U_{717} is not observable, whereas G_{719} and U_{659} have very similar chemical shifts and imino-to-imino NOE cross-peak patterns (Figure 3). Moreover, when examining additional NOEs involving the imino proton of G_{719}, we found a striking resemblance with those previously observed for the isolated J236 subdomain, providing strong evidence for a similar chemical environment for A_{718} in the subdomain and the full RNA (Supplementary Figure S7). Thus, taken together, the imino NMR data support the stacked in conformation for the A_{718} bulge in TR4P and do not provide evidence for the existence of an alternate conformation that would be significantly populated (>15%).

The local structural difference within the A_{718} bulge between the crystal structure and NMR-SAXS structure affects the helical twist of stem III and has a large effect on the relative positioning of the helical domains from the two three-way junctions (Figure 7E). Based on this observation, we suggest that the VS ribozyme undergoes a conformational change in stem III that is associated with substrate binding since the NMR-SAXS structural ensemble of TR4P describes a substrate-free *trans* ribozyme and the crystal structure captures a substrate-bound ribozyme in a pre-catalytic state. More importantly, this conformational change is likely required for formation of the active site and subsequent substrate cleavage. Thus, the NMR-SAXS approach presented here allowed us to define the high-resolution solution structure of a minimal *trans* ribozyme in the absence of substrate and brings important insights into a structural change that is likely critical for VS ribozyme function.

DISCUSSION

Previous structural studies on the *Neurospora* VS ribozyme have determined the crystal structure of a domain-swapped dimer by X-ray crystallography (43,44) and solution structures of several isolated VS ribozyme subdomains by NMR spectroscopy (40,41,45,46,70,73). The present study combines the high-resolution NMR structures of subdomains to determine the solution structure of a minimal *trans* VS ribozyme in its functional monomeric state. To achieve this, we have developed an integrative NMR-SAXS approach based on the divide-and-conquer strategy for high-resolution structure determination of large RNAs. First, both NMR and SAXS studies were used to define conditions under which the minimal *trans* VS ribozyme adopts a stable and compact monomeric fold in solution and binds to a non-cleavable substrate. Then, imino NMR studies were conducted to validate that the subdomains for which high-resolution structures are available adopt similar structures within the minimal *trans* ribozyme. For structure determination, NMR and SAXS data were combined with fine-grained structural refinement to derive a high-resolution structural ensemble of this large RNA. The method presented here allows for a good exploration of conformational space within a relatively short computational time by combinatorial sampling of a diverse set of conformations for the individual subdomains. Although more extensive molecular dynamics simulations could increase the sampling of conformational space, it was not deemed necessary to generate models that adequately fit both the NMR and the SAXS data.

The high-resolution structures of the minimal *trans* ribozyme represent one of the first hybrid NMR-SAXS structural ensemble of a relatively large RNA determined in the presence of Mg^{2+} ions (18,21). Although Mg^{2+} ions are required for the enzymatic activity and tertiary folding of the self-cleaving VS ribozyme (74,75), as well as the SL6 (46) and J236 (37) isolated subdomains, it was still unclear whether Mg^{2+} ions are required for tertiary folding of the *trans* ribozyme. We demonstrated here that the tertiary folding of the monomeric form of a minimal *trans* ribozyme is strictly dependent on the presence of Mg^{2+} ions. More specifically, imino NMR data provide evidence that Mg^{2+} ions are required for folding of a stable II-III-VI junction and, concomitantly, the transition from a highly base-paired secondary structure to a stable tertiary fold for TR4P. SAXS data further supports this model in which the minimal VS ribozyme adopts a rather flexible, loose fold in the absence of Mg^{2+} ions that becomes more compact and rigid in the presence of 5 mM Mg^{2+}. Both NMR and SAXS studies provide evidence that higher concentrations of Mg^{2+} ions (20 mM) are not desirable for structural investigations, as they lead to increases in signal linewidths in the 1D imino NMR spectrum and unexpected changes in SAXS parameters, possibly due to partial dimer formation.

Our structural investigations of the *trans* VS ribozyme in the presence of Mg^{2+} ions posed specific experimental challenges, and several guidelines can be derived from this work to reduce aggregation or the formation of higher order multimers for structural investigation of large RNAs. First, working at the lowest Mg^{2+} ion concentration possible, in this case 5 mM, was found to be extremely important. Second, as judged from the narrower imino ^1H linewidths, more homogeneous RNA samples can be obtained when preparing the Mg^{2+}-folded RNA by exchanging the RNA by ultrafiltration from a Mg^{2+}-free buffer to a Mg^{2+}-containing buffer, instead of adding Mg^{2+} ions

directly to a concentrated RNA sample. Finally, refraining from heating the Mg^{2+}-folded and concentrated RNA to high temperatures (here more than 20°C) resulted in a more homogeneous RNA population, since such treatment can lead to irreversible RNA misfolding. Although great care in sample preparation made it possible to maintain the TR4P ribozyme in a monomeric state for the present study, we encountered difficulties handling this RNA that prevented us from obtaining complementary RDC data in a liquid crystalline alignment medium and further investigating the complex between the SLI_{20}d1 substrate and TR4P ribozyme in solution.

The high-resolution NMR-SAXS structural ensemble of the VS *trans* ribozyme determined here displays a global fold generally compatible with previous low- and high-resolution models of the VS ribozyme (30,38–40,43,44,76). However, when compared to the catalytic domain of the extended VS ribozyme found in the crystal structure of the pre-catalytic state, important differences are observed. Firstly, the minimal *trans* VS ribozyme adopts a monomeric state in solution, in contrast to the domain-swapped dimer found in the crystal structure. Secondly, local structural differences are observed that are associated with substrate binding, namely subtle changes in the terminal loop of SLV to form the kissing-loop interactions with SLI and remodeling of the internal loop of SLVI to form the active site with the internal loop of SLI. These local structural differences have been previously described in detail (31,42,70). Finally, but importantly, the global shape of the catalytic domain of the ribozyme appears more compact in the crystal structure. This was found to result mainly from a local structural change in stem III, which connects the two three-way junctions. In the crystal structure of the extended ribozyme, stem III exhibits a twisted conformation, where A_{718} is extruded from the helix. However, in the NMR-SAXS solution structure of the *trans* ribozyme, stem III adopts an untwisted conformation, in which residue A_{718} is stacked within the helix (42).

A comparison of the solution and crystal structures clearly captures how this local structural change leads to two different conformational states of the VS ribozyme. In the *trans* ribozyme, the stacked in conformation of the stem III bulge is stabilized by nearest neighbor interactions and, possibly, by transient interactions between the terminal loop of SLV and the II–III–VI junction. In the crystal structure, the bound substrate is stabilized by a kissing-loop interaction with SLV and a loop-loop interaction with the A_{730} loop of stem VI. Since substrate binding involves contacts between two separate parts of the substrate rigidly held together and two different domains of the ribozyme, having a flexible bulge in the central stem III likely facilitates proper positioning of remote ribozyme domains for bipartite substrate binding. This concept is generally consistent with previous mutational studies in which both the specific location of the bulge and the length of stem III were found to be critical for VS ribozyme activity (38). Notably, although A_{718} could be replaced by a base pair with retention of 85% activity and could be removed completely or replaced by two base pairs with only an order of magnitude loss of activity, further shortening and lengthening of stem III profoundly affected activity (38). Together, these results lead us to propose that the dynamic equilibrium between the two states of the A_{718} bulge is important for substrate binding, formation of the active site and subsequent cleavage.

From our NMR investigations, we demonstrated the formation of a stable complex between the minimal *trans* VS ribozyme and a minimal non-cleavable substrate. Upon titration of the substrate to the *trans* ribozyme, two imino signals emerged that were not previously observed for the SLI_{20}/SLV complex, supporting the idea that the substrate makes additional contact with the ribozyme and/or that the ribozyme undergoes structural changes upon substrate binding. Future structural investigations are needed to bring additional information on structural changes associated with substrate binding in solution. The NMR-SAXS approach introduced here should be very valuable for these investigations as well as for future studies aimed at better understanding the high-resolution structures and intricate dynamics that allow other RNAs to fulfill their biological function. Derivatives of this integrative structural biology approach could incorporate other sources of experimental data such as residual dipolar couplings, cryo-EM maps or chemical probing. Given the modular and dynamic nature of RNA, we can also envision that divide-and-conquer approaches like the one presented here, which allow conformation sampling of assembled fragments, will become more common and straightforward in the future as more structural and dynamic information becomes available for a wide variety of small RNA domains.

DATA AVAILABILITY

SAXS data were deposited in the SAS experimental data and model data bank (SASBDB: https://www.sasbdb.org) (77) with the following accession codes: SASDKU3, SASDKV3, SASDKW3, SASDKX3. SASDKY3 and SASDKZ3 (Supplementary Table S1). NMR data were deposited in the biological magnetic resonance bank (BMRB: https://bmrb.io/) (78) under BMRB ID 50637. The structural coordinates of the hybrid NMR-SAXS structure of TR4P were deposited in the Protein Data Bank PDB-Dev archives (79) under accession code 00000067. Residues in the PDB files are numbered in agreement with Figure 1.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Calvin Byrd, Philipe Lampron, Élyse Benoit and Geneviève Di Tomasso for sample preparation and preliminary studies, Normand Cyr for assistance with collection and analysis of both SAXS and NMR data, Jurgen Sygusch for assistance with SAXS data collection and useful discussions, Nicolas Girard and Ryan Richter for computer support. We also thank Normand Cyr and James G. Omichinski for critical reading of the manuscript.

FUNDING

Canadian Institutes of Health Research (CIHR) [PJT-148801 to P.L.]; Bristol-Myers-Squibb Research Chair in

Molecular Biology at the Université de Montréal (to P.L.); Ph.D. scholarships from the Fonds de Recherche du Québec – Nature et Technologies (FRQ-NT); Université de Montréal (UdeM) (to P.D.); Master's scholarships from FRQ-NT, Luigi-Liberatore Foundation and UdeM (to G.D.); SAXS data collected at the Cornell High Energy Synchrotron Source (CHESS) using MacCHESS resources were supported by NSF & NIH/NIGMS [DMR-1829070, GM-124166]; The 700 MHz NMR data, SAXS-Lab SAXS data and computational research were performed at the Structural Biology Platform of the Department of Biochemistry and Molecular Medicine at the Université de Montréal, funded and currently supported by the Canada Foundation for Innovation [30574]. Funding for open access charge: Canadian Institutes of Health Research (CIHR) [PJT-148801 to P.L.].
Conflict of interest statement. None declared.

REFERENCES

1. Ganser,L.R., Kelly,M.L., Herschlag,D. and Al-Hashimi,H.M. (2019) The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.*, **20**, 474–489.
2. Haller,A., Souliere,M.F. and Micura,R. (2011) The dynamic nature of RNA as key to understanding riboswitch mechanisms. *Acc. Chem. Res.*, **44**, 1339–1348.
3. Mustoe,A.M., Brooks,C.L. and Al-Hashimi,H.M. (2014) Hierarchy of RNA functional dynamics. *Annu. Rev. Biochem.*, **83**, 441–466.
4. Barnwal,R.P., Loh,E., Godin,K.S., Yip,J., Lavender,H., Tang,C.M. and Varani,G. (2016) Structure and mechanism of a molecular rheostat, an RNA thermometer that modulates immune evasion by Neisseria meningitidis. *Nucleic Acids Res.*, **44**, 9426–9437.
5. Greatorex,J., Gallego,J., Varani,G. and Lever,A. (2002) Structure and stability of wild-type and mutant RNA internal loops from the SL-1 domain of the HIV-1 packaging signal. *J. Mol. Biol.*, **322**, 543–557.
6. Imai,S., Kumar,P., Hellen,C.U., D'Souza,V.M. and Wagner,G. (2016) An accurately preorganized IRES RNA structure enables eIF4G capture for initiation of viral translation. *Nat. Struct. Mol. Biol.*, **23**, 859–864.
7. Lukavsky,P.J., Kim,I., Otto,G.A. and Puglisi,J.D. (2003) Structure of HCV IRES domain II determined by NMR. *Nat. Struct. Biol.*, **10**, 1033–1038.
8. Ziegeler,M., Cevec,M., Richter,C. and Schwalbe,H. (2012) NMR studies of HAR1 RNA secondary structures reveal conformational dynamics in the human RNA. *ChemBioChem*, **13**, 2100–2112.
9. 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.
10. Plumridge,A., Katz,A.M., Calvey,G.D., Elber,R., Kirmizialtin,S. and Pollack,L. (2018) Revealing the distinct folding phases of an RNA three-helix junction. *Nucleic. Acids. Res.*, **46**, 7354–7365.
11. Zhang,H. and Keane,S.C. (2019) Advances that facilitate the study of large RNA structure and dynamics by nuclear magnetic resonance spectroscopy. *Wiley Interdiscip. Rev. RNA*, **10**, e1541.
12. Barnwal,R.P., Yang,F. and Varani,G. (2017) Applications of NMR to structure determination of RNAs large and small. *Arch. Biochem. Biophys.*, **628**, 42–56.
13. Lu,K., Miyazaki,Y. and Summers,M.F. (2010) Isotope labeling strategies for NMR studies of RNA. *J. Biomol. NMR*, **46**, 113–125.
14. Kappel,K., Zhang,K., Su,Z., Watkins,A.M., Kladwang,W., Li,S., Pintilie,G., Topkar,V.V., Rangan,R., Zheludev,I.N. *et al.* (2020) Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nat. Methods*, **17**, 699–707.
15. Zhang,K., Keane,S.C., Su,Z., Irobalieva,R.N., Chen,M., Van,V., Sciandra,C.A., Marchant,J., Heng,X., Schmid,M.F. *et al.* (2018) Structure of the 30 kDa HIV-1 RNA dimerization signal by a hybrid cryo-EM, NMR, and molecular dynamics approach. *Structure*, **26**, 490–498.
16. Shi,H., Rangadurai,A., Abou Assi,H., Roy,R., Case,D.A., Herschlag,D., Yesselman,J.D. and Al-Hashimi,H.M. (2020) Rapid and accurate determination of atomistic RNA dynamic ensemble models using NMR and structure prediction. *Nat. Commun.*, **11**, 5531.
17. Mertens,H.D.T. and Svergun,D.I. (2017) Combining NMR and small angle X-ray scattering for the study of biomolecular structure and dynamics. *Arch. Biochem. Biophys.*, **628**, 33–41.
18. Burke,J.E., Sashital,D.G., Zuo,X., Wang,Y.X. and Butcher,S.E. (2012) Structure of the yeast U2/U6 snRNA complex. *RNA*, **18**, 673–683.
19. Chen,B., Zuo,X., Wang,Y.X. and Dayie,T.K. (2012) Multiple conformations of SAM-II riboswitch detected with SAXS and NMR spectroscopy. *Nucleic Acids Res.*, **40**, 3117–3130.
20. Cornilescu,G., Didychuk,A.L., Rodgers,M.L., Michael,L.A., Burke,J.E., Montemayor,E.J., Hoskins,A.A. and Butcher,S.E. (2016) Structural analysis of multi-helical RNAs by NMR-SAXS/WAXS: application to the U4/U6 di-snRNA. *J. Mol. Biol.*, **428**, 777–789.
21. Grishaev,A., Ying,J., Canny,M.D., Pardi,A. and Bax,A. (2008) Solution structure of tRNAVal from refinement of homology model against residual dipolar coupling and SAXS data. *J. Biomol. NMR*, **42**, 99–109.
22. Wang,J., Zuo,X., Yu,P., Xu,H., Starich,M.R., Tiede,D.M., Shapiro,B.A., Schwieters,C.D. and Wang,Y.X. (2009) A method for helical RNA global structure determination in solution using small-angle x-ray scattering and NMR measurements. *J. Mol. Biol.*, **393**, 717–734.
23. Zuo,X., Wang,J., Foster,T.R., Schwieters,C.D., Tiede,D.M., Butcher,S.E. and Wang,Y.X. (2008) Global molecular structure and interfaces: refining an RNA:RNA complex structure using solution X-ray scattering data. *J. Am. Chem. Soc.*, **130**, 3292–3293.
24. Juhling,T., Duchardt-Ferner,E., Bonin,S., Wohnert,J., Putz,J., Florentz,C., Betat,H., Sauter,C. and Morl,M. (2018) Small but large enough: structural properties of armless mitochondrial tRNAs from the nematode *Romanomermis culicivorax*. *Nucleic Acids Res.*, **46**, 9170–9180.
25. Tolbert,M., Morgan,C.E., Pollum,M., Crespo-Hernandez,C.E., Li,M.L., Brewer,G. and Tolbert,B.S. (2017) HnRNP A1 alters the structure of a conserved enterovirus IRES domain to stimulate viral translation. *J. Mol. Biol.*, **429**, 2841–2858.
26. Warden,M.S., Cai,K., Cornilescu,G., Burke,J.E., Ponniah,K., Butcher,S.E. and Pascal,S.M. (2019) Conformational flexibility in the enterovirus RNA replication platform. *RNA*, **25**, 376–387.
27. Saville,B.J. and Collins,R.A. (1990) A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora* mitochondria. *Cell*, **61**, 685–696.
28. Saville,B.J. and Collins,R.A. (1991) RNA-mediated ligation of self-cleavage products of a *Neurospora* mitochondrial plasmid transcript. *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 8826–8830.
29. Zamel,R. and Collins,R.A. (2002) Rearrangement of substrate secondary structure facilitates binding to the *Neurospora* VS ribozyme. *J. Mol. Biol.*, **324**, 903–915.
30. Pereira,M.J., Nikolova,E.N., Hiley,S.L., Jaikaran,D., Collins,R.A. and Walter,N.G. (2008) Single VS ribozyme molecules reveal dynamic and hierarchical folding toward catalysis. *J. Mol. Biol.*, **382**, 496–509.
31. Girard,N., Dagenais,P., Lacroix-Labonté,J. and Legault,P. (2019) A multi-axial RNA joint with a large range of motion promotes sampling of an active ribozyme conformation. *Nucleic Acids Res.*, **47**, 3739–3751.
32. Collins,R.A. (2002) The Neurospora Varkud satellite ribozyme. *Biochem. Soc. Trans.*, **30**, 1122–1126.
33. Lilley,D.M. (2004) The Varkud satellite ribozyme. *RNA*, **10**, 151–158.
34. Lilley,D.M. (2011) Catalysis by the nucleolytic ribozymes. *Biochem. Soc. Trans.*, **39**, 641–646.
35. Lilley,D.M. (2011) Mechanisms of RNA catalysis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **366**, 2910–2917.
36. 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.
37. 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.
38. Lafontaine,D.A., Norman,D.G. and Lilley,D.M. (2002) The global structure of the VS ribozyme. *EMBO J.*, **21**, 2461–2471.

39. 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.
40. 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.
41. Bonneau,E. and Legault,P. (2014) Nuclear magnetic resonance structure of the III-IV-V three-way junction from the Varkud satellite ribozyme and identification of magnesium-binding sites using paramagnetic relaxation enhancement. *Biochemistry*, **53**, 6264–6275.
42. 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. *Wiley Interdiscip. Rev. RNA*, **8**, e1421.
43. DasGupta,S., Suslov,N.B. and Piccirilli,J.A. (2017) Structural basis for substrate helix remodeling and cleavage loop activation in the Varkud Satellite ribozymes. *J. Am. Chem. Soc.*, **139**, 9591–9597.
44. 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.
45. 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.
46. 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.
47. 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.
48. Bonneau,E. and Legault,P. (2014) NMR localization of divalent cations at the active site of the *Neurospora* VS ribozyme provides insights into RNA-metal-ion interactions. *Biochemistry*, **53**, 579–590.
49. Nikonowicz,E.P., Sirr,A., Legault,P., Jucker,F.M., Baer,L.M. and Pardi,A. (1992) Preparation of ^{13}C and ^{15}N labelled RNAs for heteronuclear multi-dimensional NMR studies. *Nucleic Acids Res.*, **20**, 4507–4513.
50. Wishart,D.S., Bigam,C.G., Yao,J., Abildgaard,F., Dyson,H.J., Oldfield,E., Markley,J.L. and Sykes,B.D. (1995) ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR*, **6**, 135–140.
51. Piotto,M., Saudek,V. and Sklenar,V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661–665.
52. Bodenhausen,G. and Ruben,D.J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.*, **69**, 185–189.
53. Mori,S., Abeygunawardana,C., Johnson,M.O. and van Zijl,P.C. (1995) Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J. Magn. Reson. B*, **108**, 94–98.
54. Sklenar,V., Piotto,M., Leppik,R. and Saudek,V. (1993) Gradient-tailored water suppression for ^{1}H-^{15}N HSQC experiments optimized to retain full sensitivity. *J. Magn. Reson. Ser. A*, **102**, 241–245.
55. Sklenar,V. (1995) Suppression of radiation damping in multidimensional NMR experiments using magnetic field gradients. *J. Magn. Reson. Ser. A*, **114**, 132–135.
56. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
57. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinas,M., Ulrich,E.L., Markley,J.L., Ionides,J. and Laue,E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
58. Acerbo,A.S., Cook,M.J. and Gillilan,R.E. (2015) Upgrade of MacCHESS facility for X-ray scattering of biological macromolecules in solution. *J. Synchrotron Radiat.*, **22**, 180–186.
59. Hopkins,J.B., Gillilan,R.E. and Skou,S. (2017) BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J. Appl. Crystallogr.*, **50**, 1545–1553.
60. Franke,D., Petoukhov,M.V., Konarev,P.V., Panjkovich,A., Tuukkanen,A., Mertens,H.D.T., Kikhney,A.G., Hajizadeh,N.R., Franklin,J.M., Jeffries,C.M. *et al.* (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.*, **50**, 1212–1225.
61. Svergun,D.I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.*, **76**, 2879–2886.
62. Rambo,R.P. and Tainer,J.A. (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*, **496**, 477–481.
63. Svergun,D., Barberato,C. and Koch,M.H.J. (1995) CRYSOL– a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.*, **28**, 768–773.
64. Tria,G., Mertens,H.D., Kachala,M. and Svergun,D.I. (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ*, **2**, 207–217.
65. Grant,T.D. (2018) Ab initio electron density determination directly from solution scattering data. *Nat. Methods*, **15**, 191–193.
66. Stasiewicz,J., Mukherjee,S., Nithin,C. and Bujnicki,J.M. (2019) QRNAS: software tool for refinement of nucleic acid structures. *BMC Struct. Biol.*, **19**, 5.
67. Rastogi,T. and Collins,R.A. (1998) Smaller, faster ribozymes reveal the catalytic core of *Neurospora* VS RNA. *J. Mol. Biol.*, **277**, 215–224.
68. Salvail-Lacoste,A., Di Tomasso,G., Piette,B.L. and Legault,P. (2013) Affinity purification of T7 RNA transcripts with homogeneous ends using ARiBo and CRISPR tags. *RNA*, **19**, 1003–1014.
69. Lukavsky,P.J. and Puglisi,J.D. (2004) Large-scale preparation and purification of polyacrylamide-free RNA oligonucleotides. *RNA*, **10**, 889–893.
70. 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.
71. Rambo,R.P. and Tainer,J.A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers*, **95**, 559–571.
72. Gopal,A., Zhou,Z.H., Knobler,C.M. and Gelbart,W.M. (2012) Visualizing large RNA molecules in solution. *RNA*, **18**, 284–299.
73. 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. U.S.A.*, **100**, 7003–7008.
74. Beattie,T.L. and Collins,R.A. (1997) Identification of functional domains in the self-cleaving *Neurospora* VS ribozyme using damage selection. *J. Mol. Biol.*, **267**, 830–840.
75. Collins,R.A. and Olive,J.E. (1993) Reaction conditions and kinetics of self-cleavage of a ribozyme derived from *Neurospora* VS RNA. *Biochemistry*, **32**, 2795–2799.
76. 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.
77. Valentini,E., Kikhney,A.G., Previtali,G., Jeffries,C.M. and Svergun,D.I. (2015) SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Res.*, **43**, D357–D363.
78. Ulrich,E.L., Akutsu,H., Doreleijers,J.F., Harano,Y., Ioannidis,Y.E., Lin,J., Livny,M., Mading,S., Maziuk,D., Miller,Z. *et al.* (2008) BioMagResBank. *Nucleic Acids Res.*, **36**, D402–D408.
79. Vallat,B., Webb,B., Westbrook,J.D., Sali,A. and Berman,H.M. (2018) Development of a prototype system for archiving integrative/hybrid structure models of biological macromolecules. *Structure*, **26**, 894–904.
80. Leontis,N.B. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.