Published in final edited form as:

J Mol Biol. 2009 October 30; 393(3): 717–734. doi:10.1016/j.jmb.2009.08.001.

A Method for Helical RNA Global Structure Determination in Solution Using Small Angle X-ray Scattering and NMR Measurements

Jinbu Wang^{1,*}, Xiaobing Zuo^{1,*}, Ping Yu^{1,2}, Huan Xu^{1,&}, Mary R. Starich³, David M. Tiede⁴, Bruce A. Shapiro⁵, Charles D. Schwieters⁶, and Yun-Xing Wang¹

¹Protein Nucleic Acid Interaction Section, Structural Biophysics Laboratory, National Cancer Institute at Frederick, National Institutes of Health, Frederick, MD 21702, USA

²SAIC-Frederick, Inc., National Cancer Institute at Frederick, National Institutes of Health, Frederick, MD 21702, USA

³Office of Chief, Structural Biophysics Laboratory, National Cancer Institute at Frederick, National Institutes of Health, Frederick, MD 21702, USA

⁴The Chemical Sciences and Engineering Division, Argonne National Laboratory, Argonne, Illinois 60439, USA

⁵Center for Cancer Research, Nanobiology Program, National Cancer Institute at Frederick, National Institutes of Health, Frederick, MD 21702, USA

⁶Division of Computational Bioscience, Center for Information Technology, National Institutes of Health, Bethesda, MD 20892, USA

Abstract

We report a “top-down” method that uses mainly duplexes' global orientations and overall molecular dimension and shape restraints, which were extracted from experimental NMR and small angle X-ray scattering (SAXS) data respectively, to determine global architectures of RNA molecules consisting of mostly A-form like duplexes. The method is implemented in the G2G (from Global measurement to Global Structure) toolkit of programs. We demonstrate the efficiency and accuracy of the method by determining the global structure of a 71-nucleotide RNA using experimental data. The backbone *root-mean-square-deviation (RMSD)* of the ensemble of the calculated global structures relative to the X-ray crystal structure using the experimental data is $3.0 \pm 0.3 \text{ \AA}$, and the *RMSD* is only $2.5 \pm 0.2 \text{ \AA}$ for the three duplexes that were orientation-restrained during the calculation. The global structure simplifies interpretation of multi-dimensional nuclear Overhauser spectra for high resolution structure determination. The potential general application of the method for RNA structure determination is discussed.

To whom correspondence should be addressed: Y-X Wang: wangyu@ncifcrf.gov, (phone) 301-846-5985.

*These authors contributed equally to this work

&Current address: Gynecology and Obstetric Hospital, Fudan University, 419 Fangxie Road, Shanghai 200011, People's Republic of China

Supporting Information Available: A more detailed description of the methods and materials, including a detailed description of the G2G toolkit is provided in the Supporting Information. The calculation protocols, and the G2G toolkit package, all data files together with the coordinates and restraint files can also be downloaded from the authors' web sites <http://xxx>.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

RNA structure; RDC; SAXS; NMR; duplex orientation; RDC waves

Introduction

Among the greatest advances in biology today are the discoveries of the various roles played by RNA in biological functions. RNA is an active participant in regulation of gene expression by interference¹ or by riboswitches,² in the processing of RNA introns,³ in the maintenance of chromosome ends by telomerase,⁴ in protein synthesis by ribosome.⁵ RNA function is encoded in its dynamics and structures^{6,7} and determination of RNA structures remains a major goal in contemporary biology. Currently, despite significant advances in X-ray crystallography and solution NMR, structure determination of any given mid- to large-size RNA molecule with a complex fold remains a daunting task. This is because of the great difficulty in growing crystals and/or obtaining phase information in X-ray crystallography, and severe size constraints on structure determination by solution NMR spectroscopy.

The prevailing approach for structure determination of RNA in solution is a “bottom-up” approach, similar to the approach used for determining protein structures,⁸ despite vast differences in both structural features and chemical compositions between these two types of biomacromolecules. In RNA, very high structural similarities among its basic building units leads to very similar chemical shift environments, thus, a very narrow chemical shift dispersion and severe NMR signal overlaps, making it extremely difficult, if not impossible, to extract sufficient local and global structural information to construct global structures of mid- and large-size RNAs.^{9,10} Consequently, the current “bottom-up” approach runs into a size barrier.

A survey of RNA X-ray crystal structures with resolution better than 3.0 Å in the Protein Data Bank reveals that A-form like duplexes are the most predominant building blocks in RNA structures and the A-form conformation in term of the sugar-phosphate backbones is very conserved (Tables 1 and Table S1 in Supporting Information, SI). The other A-form structural parameters also vary in narrow ranges, except for the basepair tilt, which heavily depend on the basepair types (Table 1). Thus, RNA duplexes (stems) depicted in a secondary structure can be treated as A-form like and can be generated from RNA databases as approximate initial structures with an acceptable accuracy. Therefore, in determining a global architecture of a RNA molecule that mostly consists of duplexes, there are essentially two problems: (1) orientations and phases (the rotation of duplex around the helical axis) of duplexes, and (2) the relative positions of the duplexes in a global sense. Once the global relative orientations, phases and relative positions of these duplexes are determined, so is the approximate global structure of the RNA. We use RDC-structure periodicity correlation to derive the discrete relative orientations (DRO) of duplexes in terms of the polar angles of a duplex axis, Θ and Φ , and phase angle ρ_0 (Figure 1a, 1b & 1c), and utilize the SAXS-derived molecular envelope to identify the correct combination of duplex orientations and approximate relative positions of the duplexes. The programs ORIENT, BLOCK and PACK in the G2G toolkit calculates DROs, generates duplex coordinates from a RNA structure database library, and packs duplexes together, respectively. In the next step, the starting structures that are generated by G2G are subjected to a rigid-body simulated annealing (SA) refinement, where the orientations and phases of duplexes and the overall shape of the RNA are restrained. We demonstrated the accuracy and efficiency of the method by determining the global structure of the adenine-riboswitch (riboA) (71 nt) using experimental RDC and SAXS data. The structure of riboA with a different sequence (see Materials and Methods section) has previously been determined using X-ray crystallography¹² was used as a benchmark for the G2G method. It is noteworthy

that combined use of RDCs and SAXS data to refine a known structure of tRNA has been reported in literature.¹³

Results

The global measurements to global structure (G2G) method

The four DROs of a chiral molecule (a domain or a whole protein or RNA) are usually derived using the singular value decomposition (SVD) method if the coordinates are known¹⁴ or using periodicity-RDC correlation¹¹ if the structure is periodic structural elements, as described briefly in the following. Duplexes belong to a unique class of chiral molecules that consist of repetitive structural units that are arranged in certain periodic patterns. This structural periodicity is reflected in and correlated with geometrical measurements, such as dipolar coupling or residual dipolar coupling of particular types of spin pairs.^{11,15} DROs are derived using an efficient nonlinear least-squares optimization routine to fit the RDC-structure periodicity correlation to experimental imino RDCs of duplexes.¹¹ The general RDC equation in terms of the spherical coordinates (θ , φ) of the bond vector connecting the two nuclei A and B

$$D_{AB} = D_a \left\{ (3 \cos^2 \theta - 1) \frac{3}{2} R \sin^2 \theta \cos 2\varphi \right\} \quad (1)$$

was rewritten to explicitly express the RDC as a function of a duplex orientation relative to an alignment tensor:

$$D_{AB} = C_1(\Theta, \Phi, \delta_n) \cos 2\rho_n + C_2(\Theta, \Phi, \delta_n) \sin 2\rho_n + C_3(\Theta, \Phi, \delta_n) \cos \rho_n + C_4(\Theta, \Phi, \delta_n) \sin \rho_n + C_5(\Theta, \Phi, \delta_n) \quad (2)$$

Where C_1 to C_5 are functions of duplex orientation (Θ, Φ), D_a and R .¹¹ The latter two are the axial and rhombic components of an alignment tensor, respectively. The bond vector of the n th residue in the duplex in spherical coordinates is given by bond length d_{AB} , angle δ , to the duplex axis (defined in Figures 1d&1e), and angle ρ_n with the x-axis (in the x-y plane perpendicular to the helix axis). The angle phase $\rho_n = (\alpha_n + \rho_0)$ where ρ_0 is the initial phase offset, $\alpha_n = 2\pi (n-1)/T$, $n = 1, 2, 3$, and T is the period of the duplex (A-RNA: $T = 11$).

Eq. 2 has five unknown variables, D_a , R , Θ , Φ , ρ_0 , and these are extracted using the nonlinear least squares fitting¹¹ method in the program ORIENT and the best fits are selected based on the RDC root mean square deviation (*RMSD*):

$$RMSD = \frac{1}{N} \sqrt{\sum_i^N (RDC_i^{\text{exp}} - RDC_i^{\text{fit}})^2} \quad (3)$$

In theory, the D_a and R values obtained by fitting RDCs in each RNA duplex should be consistent with each other. In practice, the quality of fits depends on quality of the data, number of RDCs per duplex, closeness of the duplexes to idealized A-form RNA and the rigidity of structures. The values of D_a , R and orientations and phases, (Θ , Φ , ρ_0), can be extracted by fitting each duplex individually with independent sets of tensors, or using a simultaneous fit where the same D_a and R are assumed for all duplexes, assuming the system is rigid. In the

individual fitting, each duplex is fit independently and produces its own parameter set of D_a , R and (Θ, Φ, ρ_0) . The individual fit works well only when the duplexes are long. In the case where the amount of available RDC data is not sufficient to obtain a reliable parameter set for every duplex, the initial D_a and R values are obtained from the longest duplex, and are assumed as initial values for all duplexes. The alternative method performs a simultaneous RDC fit for all duplexes in a molecule, under the rigid-body assumption, that produces a common D_a and R for all duplexes, and orientations and phases for individual duplexes.

In principle, the unique orientation of each duplex can unambiguously be resolved by applying a second independent alignment tensor.¹⁶ However, the dual alignment tensors may not be the most practical approach, partly because of the lack of a truly independent second alignment medium for RNA.¹⁷ Therefore, we applied a combined approach by using both the shape information of a RNA molecule and the DROs to determine the unique orientation for each duplex in a molecular frame. This is done by restraining angles between duplexes to be consistent with the overall molecular envelope, which is calculated from SAXS data, in the program ORIENT. An example of an input to the program is shown in the Supporting Information (SI).

RNAs consisting of mostly duplexes have the following unique information and features, which must be considered in determining global structures: (1) Approximate secondary structures of RNAs are usually known prior to structure determination, based on either predictions,¹⁸ experimental verification,¹⁹ or both; (2) RNA bases tend to be stacked either sequentially or in tertiary interactions (more than 95% of bases in any given RNA structure are stacked in a folded state, as shown by our survey of the RNA database). Bases that are not stacked tend to be at the tip of loops or in bulges. RNA duplexes and well-determined tetraloop hairpins can be considered, to a first degree of approximation, as rigid building blocks for delineating the global architecture of a large RNA molecule.

We generate coordinates of these duplex building blocks and linker residues from MOSAICLIB in a database using the BLOCK program in the G2G toolkit. MOSAICLIB was built based on the accepted geometric nomenclature, classification and standards.^{20,21} These duplex building blocks are then connected by linkers with arbitrary initial conformations. Similar structures can also be generated from non-experimental predictions using the software^{22,23} that appeared in the literature during this study. A more detailed description of the program BLOCK is given in SI. Packing the discretely oriented building blocks is accomplished by the program PACK. In principle, when RDCs are measured from only one alignment medium, each duplex can be determined to four possible discrete orientations. In the case of three-duplex containing RNA such as riboA, there would be $4 \times 4 \times 4$ possible combinations of these orientations, or four equivalent sets of 16 possible combinations. The actual number of viable combinations of orientations is lower, because of steric clashes between duplexes and linkers, as shown in the riboA case. A more detailed description about PACK is given in SI.

SAXS data contains information about the overall dimension and shape of a biomacromolecule in solution.²⁴ This is particularly true for nucleic acids because a large electron density of the phosphate-sugar backbones dominates the SAXS intensity (Fig. S2 in SI).²⁵ Thus, the shape of a molecular envelope reflects the outline shape of an overall structure of a RNA backbone. This information complements NMR spectroscopy, which is not particularly useful in extracting the phosphate-sugar backbone conformation of large RNA molecules. The molecular envelope, determined from SAXS data using the program DAMMIN,²⁶ was utilized to define the approximate dimensions, which are indicative of packing of the RNA duplexes. In the case of riboA, the top width of the front view of the topological envelope is about 36 Å (Fig. S3, SI), a little less than the sum of the width of the two duplexes but much greater than

that of a single duplex. This width suggests that the two duplexes are packed side by side and “intercalated”. The width of the side view of the envelope, about 22 Å, simply suggests that the two duplexes are packed in a parallel or an anti-parallel arrangement (see the analysis in the next section).

Determining the global structure of the 71-nt riboA RNA in solution

In order to evaluate thus demonstrate the utility of the G2G method, we applied it to solve the global structure of the 71-nt riboA RNA whose secondary structure is shown in Figure 2a. The NMR spectra used for assigning imino signals, identifying basepairs in both duplexes and non-duplex regions, and measuring RDC are shown in Figures 2b & 2c and 3, respectively. The dipolar waves of the three duplexes of riboA are shown in Figure 4.

Four possible discrete orientations were derived from the one set of RDC data for each duplex by fitting eq 2. Assuming duplex H1 takes any one of four orientations, there are four possible relative orientations of duplexes H2 (Figures 5a–d), and the same is true for duplexes H1 and H3 (Figure 5e–h). Thus there are eight possible combinations for the orientations of duplexes H1 and H2, and duplexes H1 and H3. Covalent bond linkages and steric hindrance between linkers and duplexes, detected by the PACK program, rule out combinations b and d, and f and g as non-viable conformations. The correct combination of the DROs of all three duplexes was identified by shape-aided orientation analyses (Figure 6), which led to the conclusion that the three duplexes in riboA are packed in either approximately a parallel or an anti-parallel arrangement (ae in Figure 6).

The quantitative determination of orientations and phases of duplexes was then carried out using program ORIENT. In the calculation, the angles between duplexes were restrained to either $0 \pm 30^\circ$ or $180 \pm 30^\circ$ for all three pairs of duplexes to account for parallel or anti-parallel arrangements. The best simultaneous fit yields axial component $D_a = -26.4$ Hz, and rhombic component $R = 0.35$, with orientations and phases, (Θ, Φ, ρ_0) of $(151^\circ, 281^\circ, 101^\circ)$, $(22^\circ, 97^\circ, 259^\circ)$ and $(40^\circ, 101^\circ, 63^\circ)$ for duplexes H1, H2 and H3, respectively, which represent angles of 173° , 169° and 18° between H1-H2, H1-H3, and H2-H3, respectively. The top fits with an overall *RMSD* no greater than 1.5 Hz in RDCs yield average values of $D_a = -25.5 \pm 2.0$, $R = 0.28 \pm 0.03$, with average values of (Θ, Φ, ρ_0) of $(160 \pm 9^\circ, 288 \pm 18^\circ, 104 \pm 4^\circ)$, $(13 \pm 7^\circ, 94 \pm 6^\circ, 264 \pm 6^\circ)$ and $(20 \pm 20^\circ, 100 \pm 4^\circ, 64 \pm 3^\circ)$ for duplexes H1, H2 and H3, respectively, where the error ranges are standard deviations.

We then derived the structural topology (Figure 7a, left), using the duplex orientation and phase in terms of (Θ, Φ, ρ_0) , and the overall envelope, and built a three-dimensional starting structure using the BLOCK and PACK programs of G2G (Figure 7a, right). In addition to the restraints imposed by the orientations, phases and the overall shape of the envelope, the degrees of freedom for the relative translational positions of the three duplexes are also constrained in part by linkers of short stretches of nucleotides, which were generated with arbitrary conformations using the MOSAIC library.

This structure was then used as a starting structure for regularization and a hybrid rigid-body simulated annealing (SA) refinement protocol similar to that used for refining the global structure of a dimeric RNA:RNA complex²⁷ using Xplor-NIH.²⁸ The regularization procedure in Xplor-NIH connects linkers with duplex building blocks and removes any gross covalent geometry distortions before SA refinement. During the refinement, the duplex orientations and phases were held fixed in space, but duplex translations and arbitrary linker motions were allowed using the IVM facility²⁹ of the Xplor-NIH package (version 2.22 or newer). In addition to the loose-distance and torsion-angle restraints that were applied to maintain the conformation of approximate A-form duplexes, we added the following restraints to the calculation: (1) the experimental SAXS data and an explicit restraint on the radius of gyration that was extracted

from the SAXS data; (2) uniform distance restraints to maintain neighboring base-stacking throughout the whole chain; (3) restraints for hydrogen bonds, extracted from imino HNN-COSY^{30,31} and NOE experiments, in addition to those in the duplex regions; (4) imino RDC restraints for residues in both the duplex and non-duplex regions; and (5) approximate dimension restraints derived from the envelope of riboA in the form of approximate phosphorus-phosphorus distances. To avoid gross close contacts, we also added a minimum distance repulsive restraint, 6 Å that is the sequential phosphorus distance in an A-form duplex and is generally shortest possible distance separating any given two phosphate groups in RNA structures. The backbone RMSD of the average structure of the top 10% of the lowest energy structures for overall structure excluding the flexible loops from the rigid-body SA calculation is about 3.3 Å comparing to that of the X-ray crystal structure (Figure 7b) (Table S3 in SI for the calculation statistics).

The global structure, shown in Figure 7b, greatly facilitates assigning cross peaks in two-dimensional Nuclear Overhauser Effect (NOE) and ¹⁵N-HNN-COSY spectra and led to identification of tertiary hydrogen bonding interactions/close contacts suggested by experiments (Figure 2). For example, the initial structure puts the two loops in L2 and L3 facing each other in space (Figure 7b). This “closeness” in space between the two loops led to assignments of two unassigned sequential GC pairs: G38:C60 and G37:C61 (Figure 2b & 2c), because they immediately follow the G59:C67 pair in H3 in the imino NOE walk path. In the second example, an imino signal of a Watson-Crick G:C pair, revealed by the HNN-COSY spectrum, has a cross peak to that of U25 in the imino NOE walk path from H2 throughout to H1 (Figure 2b & 2c) and was assigned as G46:C53 for the following reasons: G46 is in close vicinity of U25 according to the arrangements of the duplexes in the topology. Because of the arrangements among H1, H2 and H3, C53 is only unpaired cytosine in the close vicinity of G46 and is likely to be the one forming this Watson-Crick base pair with G46. In addition, the imino of G46 has cross peaks with those of two other Us, which are U22 and U47 because they are the only two Us in the close vicinity of G46 in the initial global structure. The third example is about assigning signals related to ligand adenine. A previous biochemical study indicated that U74 interacts with the ligand most likely via a Watson-Crick pairing.³² We assigned the imino proton signal H9 of the ligand adenine base and identified its position using a combination of the two spectra and the initial topology structure of riboA as follows. U74 is located in the junction formed with linkers between H1-H2, H2-H3 and H1-H3. The U74 imino was assigned by its strong cross peak to that of U75 that is at the end of H1 (Figure 2b & 2c). The H9 signal of the ligand adenine base was identified by the peak position at 11.43 ppm in the ¹H-NOESY spectrum (Figure 2c) but without corresponding signal at the same position in the ¹⁵N-HNN-COSY spectrum (Figure 2b), suggesting that the signal is from the non-isotope labeled N9H9 imino of the adenine ligand. Furthermore, the U74 imino forms a hydrogen bond with the non-labeled ligand adenine as suggested by the absence of the acceptor signal (Figure 2b and 2c). The ligand also forms a hydrogen bond with another U imino 13.24 ppm (Figure 2c), suggested by the absence of the acceptor signal (Figure 2b). This U imino has also strong cross peaks with the U74 and U75 imino protons (Figure 2c). Based on the initial structure, U51 is the only U in the close vicinity of the ligand, U75 and U74. We therefore assigned the imino at 13.24 ppm to that of U51. These long range distance restraints were then applied to restrain the structure in the rigid-body SA calculation, where the duplex orientations and phases were initially fixed to maintain the global folding, and then later during refinement the each duplex as rigid-body was allowed to rotate to achieve the best overall fit to the RDC data.

A comparison of the ensemble of the top 10% of the lowest energy G2G structures and the X-ray crystal structure is shown in Figure 8a. The structure calculation statistics are listed in Table 2. The backbone RMSD of the ensemble of the calculated global structures relative to the X-ray crystal structure is 3.0 ± 0.3 Å for the whole molecule and 2.5 ± 0.2 Å for the three orientation and phase restrained duplexes. The backbone RMSD between the X-ray crystal and

the average structure of the three duplexes is 2.5 Å. The agreement between the G2G structure and the experimental RDC and SAXS data was also examined. The correlation coefficient r between the back-calculated RDCs based on the ensemble of the top 10% lowest energy structures from the rigid-body calculation and the experimental data is about 0.83 (Figure 8b). The correlation coefficient between the back-calculated based on the regularized average structure of the ensemble and the experimental RDCs is about 0.95 (Figure 8c). For the comparison, the correlation between the RDCs calculated based on the X-ray crystal structure and the experimental data is about 0.77 (Figure 8d). The relatively low correlation for the X-ray crystal structure may in part be due to differences between the structures in solution and in crystalline states as well as the sequence difference in duplex H1 that might result in direct or indirect changes in the structure nearby or even in the entire structure (see the Materials and Methods). The quality of the G2G structure is also demonstrated by the comparison of the back-calculated SAXS curves with the experimental one and the *RMSD* between the two is about 0.29 ± 0.04 (Figure 8e). The PDDF comparison is shown in Figure 8f. With the current data, the approximate position of the adenine ligand was also determined (Figure 8g).

Discussion

The “top-down” approach of the G2G method presents a new strategy for determining global RNA structure in solution. In general, the *RMSD* of the G2G structure relative to the “true” structure may approximately be estimated using the following empirical formula:

$$\text{RMSD} = [\alpha^2 \cdot p^{\text{duplex}} + \beta^2 \cdot (1 - p^{\text{duplex}})]^{1/2} \quad (7)$$

Where α is the possible *RMSD* between the “true” and the database-derived duplex structures in the context of the structure, β is the possible *RMSD* between of the “true” and the G2G structures of non-duplex regions, such as long linkers and under-determined loops; p^{duplex} is the percentage of duplex residues in the RNA. For A-form like duplexes, α of individual duplex is well below 2.0 Å, based on *RMSDs* from the database (Table 1). The value of β can vary significantly, depending on the length of non-duplex regions such as linkers and loops. In the riboA case, duplexes make up more than 60% of the total residues and the linkers between H1 & H2, and H2 & H3 are relatively short, the overall *RMSD* between the G2G and the “true” structure is estimated to be about 3.3 Å or better, assuming α and β are about 2.5 and 4.0 Å for duplexes and the long linker/loops, respectively.

The global structures may imply potential intricate networks of interactions, such as multi-base pairing³³ at junctions and among residues that are next to each other in a three-dimensional fold but otherwise are far apart from each other in sequence, as seen in the riboA RNA structure (an illustration identifying such tertiary basepairings are presented in the Results section). Incorporating those tertiary basepairings in the rigid-body calculation results in an improvement of the structure, as indicated by a small decrease in overall backbone *RMSD*, about 0.4 Å, between the ensemble vs the crystal (Figures 7b and 8a). Thus, the G2G global structures of RNAs may serve as structural scaffolds for rapid determination of high-resolution structures in a way that is similar to using a structural model driven approach to determine the high-resolution structures of homologous proteins.³⁴ Furthermore, since there is virtually no size limitation in using SAXS data to derive the approximate molecular envelope, this method can potentially be used to study the structures of large RNAs in solution, provided that imino signals can be assigned and enough RDCs can be measured by using partial and/or segmental labeling schemes¹⁰ in conjunction with a SAXS-aided divide-and-conquer strategy. The G2G method may also be valuable in deriving more accurate structures with a combined use of the MC-fold and MC-Sym pipeline.^{22,23}

The efficacy and accuracy of the G2G method heavily depends on how well orientations and phases of duplexes can be determined, which depends in part on the quality of RDCs and number of available RDCs per duplex. When orientations of the X-ray structure was taken in place of the experimentally determined ones, the calculation with the same number of restraints yielded an ensemble of structures with the backbone RMSDs of 0.9 or 1.7 Å for the three duplexes or all residues, respectively (Figure S5, SI). In principle, the longer a duplex is, the better its orientation can be determined, because more RDCs per duplex can possibly be extracted. Furthermore, in general, dynamics with large amplitude may complicate interpretations of both SAXS and RDC data. Therefore, the G2G method is limited to rigid systems. In practice, segments and residues of a RNA may undergo motions at various time scales^{6,7}. Both the SAXS and RDCs are time-averaged measurements on fraction of ms-sec time scales. We examined the validity of the rigid-body assumption for riboA by comparing the duplex orientations calculated with and without the rigid-body assumption (Table 3). The angles between duplexes computed from the best simultaneous fit (under the rigid-body assumption) are 173° for H1 & H2, 169° for H1 & H3 and 18° for H2 & H3. Those angles agree remarkably well with those calculated from the best individual fits (without the rigid-body assumption) with 172°, 162°, 23° for H1&H2, H1&H3, and H2&H3, respectively. This agreement indicates that riboA is well packed with no large amplitude internal motions among duplexes and the rigid body assumption approximately applicable for riboA. The rigid-body assumption for riboA is also consistent with the linearity of the Guinier plot (indicating uniform molecular size), the bell-shaped Kratky plot (well folded conformation,²⁴ Figure S6), the convergence and compactness of the molecular envelope derived from SAXS data.

Materials and Methods

RNA sample preparation

A plasmid containing template coding for the version of riboA used in this study was a gift from Professor David Draper at the Johns Hopkins University. This riboA molecule differs from that of the X-ray crystal structure³⁵ in four basepairs: G14-U82 => G14-C82, C15G81 => G15-C81, U16-A80 => A16-U80 and U17-A79 => A17-U79. All these mutated basepairs are located in duplex H1 (Figure 2a) and have no tertiary interactions with other parts of the RNA. These changes made it much more stable (personal communication with David Draper) for the structural study. The riboA RNA samples were prepared using in vitro transcriptions with T7 RNA polymerase, followed by a polyacrylamide gel electrophoresis (PAGE) purification. ¹⁵N-isotope-labeled RNA samples were prepared from ¹⁵N-labeled NTPs (Sigma-Aldrich, St. Louis, MO). NMR samples were extensively dialyzed against pH = 6.8 buffer containing 50 mM potassium acetate, 2 mM MgCl₂ and 2 mM adenine. For scattering experiments, the Tris-HCl buffer was used in place of potassium acetate. The concentrations of riboA were 1-5 mg/ml for the X-ray scattering and 0.5–1.0 mM for all NMR experiments.

NMR experiments

For assigning imino signals, we recorded two-dimensional ¹H-¹H NOESY⁸ and HN-N COSY spectra^{30,31} at 15, 25 °C on a Varian Innova spectrometer operating at proton frequency of 800 MHz equipped with a triple-resonance cryo-probe. Mixing times of 100, 120 and 150 ms, depending on temperatures, were used for recording the NOESY experiments. All spectra were processed and analyzed using nmrPipe, nmrDraw³⁶ and NMRViewJ (One Moon Scientific, New Jersey).

Assigning imino proton and hydrogen bond signals

The G2G methodology requires assigning imino signals and identifying hydrogen bonds involved in canonical and noncanonical basepairs in duplexes. The assignments of imino signals of the riboA RNA were accomplished by an NOE-walk of the two-dimensional NOESY

spectrum of the imino region, aided by the two-dimensional [$^{15}\text{N}\cdots\text{H}-^{15}\text{N}$]HNN-COSY spectrum.^{30,31} The assignments of the imino signals in the regular duplex regions were straightforward (Figure 2). The initial fold, which was generated based on the information on the orientations and packing of duplexes (Figure 7), provided a visible model of the approximate relative locations of residues in a global sense and made it relatively easy to identify close contact and potential hydrogen bonds involved in tertiary interactions in the junction and loops (see the previous section).

Residual dipolar coupling (RDC)

The RDC data were derived from the IPAP-HSQC³⁷ spectra, which were recorded using ^{15}N -isotopic RNA samples in isotropic and anisotropic conditions. The anisotropic riboA sample was prepared by adding about 9.7 mg/ml pf1 phage (ASLA Biotech, Burlington, NC), which resulted in a split of 9.8 Hz in deuterium signal. RDC values range from 2.2 to 22.7 Hz for the imino ^{15}N - ^1H of riboA. The IPAP spectra of the imino resonance splitting of the riboA in the alignment medium are shown in Figure 3 and the RDC-structural periodicity correlation curves of the three duplexes are shown in Figure 4.

SAXS experiments, data analysis and molecular envelope calculation

Both small-angle and wide-angle X-ray scattering (SAXS and WAXS) were performed at beamlines 12-ID and 18-ID of Advanced Photon Sources (APS) at Argonne National Laboratory. The wavelength, λ , of X-ray radiation was set as 1.033 Å and the scattered X-ray photons were recorded with a charge-coupled device X-ray detector (Mar). An X-ray flow cell made of a cylindrical quartz capillary with a diameter of 1.5 mm and a wall of 10 µm was used. The X-ray beam with size of $0.1 \times 0.2 \text{ mm}^2$, was adjusted to pass through the center of the cell. The exposure time was set to 0.5-1.0 seconds to avoid detector saturation and radiation damage. Potential radiation damage was further reduced by flowing the samples.

Twenty images were taken for each sample or buffer solution to get good statistics. The two-dimensional images were reduced to one-dimensional scattering profiles using the program MarDetector (Tiede, unpublished). In this program, the center of beam and the sample-to-detector distance, and thus q values, of individual detector pixels, were first calibrated using scattering data of silver behenate powder. The 2-D scattering images of buffers and samples were azimuthally averaged after solid angle correction and then normalized with the intensity of the incident X-ray beam. The resulting 1-D scattering data sets were averaged before buffer background subtraction. The scattering profile of a sample solute was calculated with the following equation:

$$I^{\text{solute}}(q) = I^{\text{sample}}(q) - \alpha I^{\text{buffer}}(q), \quad (4)$$

where $I(q)$ is the scattering intensity at q , and α is the scaling factor that denotes the relative contribution from the buffer. For a quick quality check on SAXS data, α was estimated as $\alpha \approx 1 - c_{\text{mass}}\beta/1000$, where c_{mass} is the concentration of sample in mg/ml, and β is the partial specific volume of a solute, 0.54 for nucleic acids.

The range of momentum transfer q [$=4\pi \sin\theta/\lambda$, where 2θ is the scattering angle] of SAXS experiments was 0.006-0.250 Å⁻¹, and that of WAXS was 0.1 – 2.6 Å⁻¹. The SAXS data quality was evaluated by the linearity of Guinier plot (eq. 5). The more accurate background subtractions were performed in the following way. First, the WAXS profile was obtained by using eq. 4 and tuning the value of α to totally get rid of the scattering from buffer, indicating by the disappearance of the solvent peak at 2.0 Å⁻¹. Then, the resulting WAXS profile was

used as a guide for the SAXS background subtraction, by tuning the value of α in SAXS subtraction and overlaying the resulting SAXS profile with WAXS profile at the overlapping q range, i.e., $0.1 - 0.25 \text{ \AA}^{-1}$ in our experiments. The final scattering data were obtained by piecing the resulting SAXS and WAXS data together in the range of $0.006 - 2.5 \text{ \AA}^{-1}$.

The radius of gyration (R_g) was calculated from data at low q values in the range of $qR_g < 1.2$, using the Guinier approximation (eq. 5),

$$\ln I(q) = \ln(I(0)) - R_g^2 q^2 / 3 \quad (5)$$

The R_g values using this reciprocal space method were $19.2 \pm 0.2 \text{ \AA}$ for riboA. The scattering intensities at and near $q = 0$ were extrapolated with the Guinier equation, when needed.

The PDDF, $p(r)$, in real space was calculated using GNOM,³⁸ which is an indirect Fourier transform program, using real space perceptual criteria based on a solid sphere. To avoid under-estimation of the molecular dimension and consequent distortion in low resolution structural reconstruction, the parameter $R_{\max}$, the upper end of distance r , was chosen such that the resulting PDDF has a short, near zero-values tail at large r . The maximum distances ($D_{\max}$) were estimated as $62 \pm 5 \text{ \AA}$ from the PDDF for the riboA molecule. The R_g value of riboA calculated from $p(r)$ using GNOM was $19.8 \text{ \AA}$, in good agreement with that calculated directly from Guinier approximation. We used the program DAMMIN²⁶ to obtain an approximate molecular envelope, which outlines the phosphate-sugar backbone outline of an RNA structure. In DAMMIN, a spherical space with a radius of $R_{\max}$, read from the PDDF result, is built and initially filled with multiple-phase small beads or dummy atoms. To avoid distortion caused by possible under-estimation of $D_{\max}$, the $R_{\max}$ was usually set to be about 10 -20 $\text{\AA}$ greater than $D_{\max}$. At each step, the envelope evolves by randomly phasing a dummy-atom in (as a part of the molecule) or out (as a part of the solvent). A simulated annealing algorithm is employed in DAMMIN for driving the envelope evolution, by reducing the discrepancy between the experimental and calculated scattering curves during the annealing process. All reconstructions for riboA were run in the “jagged” mode, which works well for middle size molecules, yielding dummy atoms sufficient for resolving RNA duplexes, but is less computationally demanding than “slow” mode. The resulting structural models were subjected to averaging using the program package DAMAVER.³⁹ In this program, the normalized spatial discrepancy (NSD) values between each pair of models were computed. The model with lowest average NSD with respect to the rest of models was chosen as the reference model. The remaining models were superimposed onto the reference model using SUPCOMB⁴⁰ except that possible outliers identified by NSD criteria were discarded. The dummy atoms of these superimposed models were remapped onto a densely packed grid of atoms with each grid point marked with its occupancy factor. The grids with non-zero occupancies were chosen to generate a final consensus model with the volume equal to the average excluded volume of all the models. Scattering data in a q range of $0.006-0.33 \text{ \AA}^{-1}$, which reflect the global shape without significant undesired influence from the internal structure, were used in DAMMIN calculations. 16 independent DAMMIN runs were performed and the resulting bead models were subjected to averaging by DAMAVER. The R_f values, as defined in the following:

$$(R_f^2 = \sum_q [sc * I_{\text{mod}}(q) - I_{\text{exp}}(q)]^2 / \sum_q [I_{\text{exp}}(q) * q^2]^2) \quad (6)$$

were in the range of 0.008 – 0.009, indicating good match between the experimental scattering data and the calculated ones for individual models, and the average NSD for those samples was 0.61 ± 0.01 , which indicate the excellent convergences in both individual DAMMIN fits and overall bead model ensembles for each sample.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Professor D. E. Draper for providing us with the plasmid template for in vitro transcription of the riboA RNA for helpful discussions and reviewing the manuscript. This research was supported [in part] by the Intramural Research Program of the NIH, National Cancer Institute, Y-X.W. and B.A.S., by the Intramural Research Program of the NIH, the CIT Intramural Research Program to C.D.S. Work at Argonne National Laboratory (DMT) and the Advanced Photon Source was supported by the Office of Basic Energy Sciences, Department of Energy under contract DE-AC02-06CH11357. We thank Drs. L. Guo (BioCAT, sector 18-ID) and S. Seifert (BESSRC, sector 12-ID) at Argonne National Laboratory for their support for synchrotron experiments. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract No. W-31-109-ENG-38. BioCAT is a National Institutes of Health-supported Research Center RR-08630. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Center for Research Resources or the National Institutes of Health.

References

1. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 1998;391:806–811. [PubMed: 9486653]
2. Tucker BJ, Breaker RR. Riboswitches as versatile gene control elements. *Curr Opin Struct Biol* 2005;15:342–348. [PubMed: 15919195]
3. Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 1982;31:147–157. [PubMed: 6297745]
4. Blackburn EH. Telomerases. *Annu Rev Biochem* 1992;61:113–129. [PubMed: 1497307]
5. Moore PB, Steitz TA. The involvement of RNA in ribosome function. *Nature* 2002;418:229–235. [PubMed: 12110899]
6. Zhang Q, Stelzer AC, Fisher CK, Al-Hashimi HM. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* 2007;450:1263–1267. [PubMed: 18097416]
7. Zhang Q, Sun X, Watt ED, Al-Hashimi HM. Resolving the motional modes that code for RNA adaptation. *Science* 2006;311:653–656. [PubMed: 16456078]
8. Wüthrich, K. *NMR of Proteins and Nucleic Acids*. John Wiley & Sons; New York: 1986.
9. Lukavsky PJ, Kim I, Otto GA, Puglisi JD. Structure of HCV IRES domain II determined by NMR. *Nat Struct Biol* 2003;10:1033–1038. [PubMed: 14578934]
10. D'Souza V, Dey A, Habib D, Summers MF. NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J Mol Biol* 2004;337:427–442. [PubMed: 15003457]
11. Walsh JD, Cabello-Villegas J, Wang YX. Periodicity in residual dipolar couplings and nucleic acid structures. *J Am Chem Soc* 2004;126:1938–1939. [PubMed: 14971918]
12. Serganov A, Yuan YR, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, Patel DJ. Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol* 2004;11:1729–1741. [PubMed: 15610857]
13. Grishaev A, Ying J, Canny MD, Pardi A, Bax A. Solution structure of tRNA^{Val} from refinement of homology model against residual dipolar coupling and SAXS data. *J Biomol NMR* 2008;42:99–109. [PubMed: 18787959]
14. Losonczi JA, Andrec M, Fischer MW, Prestegard JH. Order matrix analysis of residual dipolar couplings using singular value decomposition. *J Magn Reson* 1999;138:334–342. [PubMed: 10341140]

15. Mesleh MF, Veglia G, DeSilva TM, Marassi FM, Opella SJ. Dipolar waves as NMR maps of protein structure. *J Am Chem Soc* 2002;124:4206–4207. [PubMed: 11960438]
16. Fischer MW, Losonczi JA, Weaver JL, Prestegard JH. Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. *Biochemistry* 1999;38:9013–9022. [PubMed: 10413474]
17. Latham MP, Hanson P, Brown DJ, Pardi A. Comparison of alignment tensors generated for native tRNA(Val) using magnetic fields and liquid crystalline media. *J Biomol NMR* 2008;40:83–94. [PubMed: 18026844]
18. Mathews, DH.; Turner, DH.; Zuker, M. RNA secondary structure prediction. In: Beaucage, Serge L., et al., editors. *Current protocols in nucleic acid chemistry*. 2007. Chapter 11, Unit 11 12
19. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* 2005;127:4223–4231. [PubMed: 15783204]
20. Leontis NB, Westhof E. Geometric nomenclature and classification of RNA base pairs. *RNA* 2001;7:499–512. [PubMed: 11345429]
21. Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, Harvey SC, Heinemann U, Lu XJ, Neidle S, Shakked Z, Sklenar H, Suzuki M, Tung CS, Westhof E, Wolberger C, Berman HM. A standard reference frame for the description of nucleic acid base-pair geometry. *J Mol Biol* 2001;313:229–237. [PubMed: 11601858]
22. Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* 2008;452:51–55. [PubMed: 18322526]
23. Martinez HM, Maizel JV Jr, Shapiro BA. RNA2D3D: a program for generating, viewing, and comparing 3-dimensional models of RNA. *J Biomol Struct Dyn* 2008;25:669–683. [PubMed: 18399701]
24. Glatter, O.; Kratky, O. *Small Angle X-ray Scattering*. Academic Press; New York: 1982.
25. Zuo X, Cui G, Merz KM Jr, Zhang L, Lewis FD, Tiede DM. X-ray diffraction “fingerprinting” of DNA structure in solution for quantitative evaluation of molecular dynamics simulation. *Proc Natl Acad Sci U S A* 2006;103:3534–3539. [PubMed: 16505363]
26. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing (vol 76, pg 2879, 1999). *Biophys J* 1999;77:2896–2896.
27. Zuo XB, Wang JB, Foster TR, Schwieters CD, Tiede DM, Butcher SE, Wang YX. Global molecular structure and interfaces: Refining an RNA : RNA complex structure using solution X-ray scattering data. *J Am Chem Soc* 2008;130:3292–+. [PubMed: 18302388]
28. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 2003;160:65–73. [PubMed: 12565051]
29. Schwieters CD, Clore GM. Internal coordinates for molecular dynamics and minimization in structure determination and refinement. *J Magn Reson* 2001;152:288–302. [PubMed: 11567582]
30. Dingley AJ, Masse JE, Feigon J, Grzesiek S. Characterization of the hydrogen bond network in guanosine quartets by internucleotide 3hJ(NC)' and 2hJ(NN) scalar couplings. *J Biomol NMR* 2000;16:279–289. [PubMed: 10826880]
31. Wohnert J, Dingley AJ, Stoldt M, Grolach M, Grzesiek S, Brown LR. Direct identification of NH... N hydrogen bonds in non-canonical base pairs of RNA by NMR spectroscopy. *Nucleic Acids Res* 1999;27:3104–3110. [PubMed: 10454606]
32. Mandal M, Breaker RR. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature structural & molecular biology* 2004;11:29–35.
33. Bloomfield, VA.; Crothers, DM.; Tinoco, I. *Nucleic Acids, Structures, properties, and Functions*. University Science Books; Sausalito, CA: 2000.
34. Duggan BM, Legge GB, Dyson HJ, Wright PE. SANE (Structure Assisted NOE Evaluation): an automated model-based approach for NOE assignment. *J Biomol NMR* 2001;19:321–329. [PubMed: 11370778]
35. Serganov A, Patel DJ. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nature reviews* 2007;8:776–790.
36. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 1995;6:277–293. [PubMed: 8520220]

37. Ottiger M, Delaglio F, Bax A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson* 1998;131:373–378. [PubMed: 9571116]
38. Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Cryst* 1992;25:495–503.
39. Petoukhov MV, Konarev PV, Kikhney AG, Svergun DI. ATSAS 2.1 – towards automated and websupported small-angle scattering data analysis. *J Appl Cryst* 2007;40:s223–s228.
40. Kozin MB, Svergun DI. Automated matching of high- and low-resolution structural models. *J Appl Cryst* 2001;34:33–41.
41. Clore GM, Gronenborn AM, Tjandra N. Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J Magn Reson* 1998;131:159–162. [PubMed: 9533920]

Figure 1.

The definition of duplex orientation (Φ, Θ) and phase ρ_0 (a & b), simulated RDC-structural periodicity correlation curves (c) that are color-coded the same as in (a) and the definition of duplex axis (d and e). The shapes of the RDC-structural periodicity correlation curves (RDC waves) depend on the orientation and phase of duplexes. The shapes of the RDC-structural periodicity correlation curves depend on orientation (Θ, Φ) and the phase ρ_0 . The orientations and phases of duplexes colored in red, blue and cyan are (Θ, Φ, ρ_0), are ($45^\circ, 45^\circ, 0^\circ$), ($60^\circ, 300^\circ, 120^\circ$) and ($75^\circ, 128^\circ, 232^\circ$), respectively. The RDC waves are simulated using $D_a = -15.6$ Hz and $R = 0.22$. The A-form duplexes were generated using Discovery Studio 2.0 (Accelrys Software, Inc.). (d) Each ball represents a phosphate atom (P) and a basepair PP vector stands for the vector from 3' P to 5' P in the same base pair. Big bold black arrow shows the duplex orientation. Big bold black arrow indicates the duplex orientation. (e) The vector $\overline{\Delta P_i} = \overline{p_{i+1} p_{j+1}} - \overline{p_i p_j}$ belongs to the plane perpendicular to the duplex axis. The cross product $\vec{n}_i = \overline{\Delta P_i} \otimes \overline{\Delta P_{i+1}}$ is the normal of the plane perpendicular to the duplex axis, which is parallel to the duplex axis. The duplex orientation is calculated as an average of the normals.

Figure 2.

Schematic drawing of the secondary structure of the riboA RNA (a) and its 2D NMR spectra, (b) HNN-COSY spectrum of ^{15}N -labeled riboA sample, and (c) 2D NOESY spectrum of non-labeled riboA in the imino and partial amino proton region. In (a), the ligand adenine is labeled as A99 in the diagram and is in a close contact with U74. In (b), AU pairs are linked with magenta lines, and GC with green lines. Dashed lines denote missing acceptor because the ligand adenine is the acceptor but not isotope-labeled. In (c), two NOE-walking paths are shown in red and blue.

Figure 3. Imino ^{15}N - ^1H IPAP 9 spectra of the ^{15}N -labeled riboA, recorded in an anisotropic solution containing about 0.5 mM RNA and 9.7 mg/ml pf1 alignment medium. The spectra were recorded at 25 °C on a VARIAN Inova spectrometer operating at proton frequency 800 MHz and equipped with a cryogenic triple resonance probe with a z-gradient. The imino peaks are generally well resolved.

Figure 4.

The dipolar waves for H1, H2 and H3 in riboA from the simultaneous fits. Four DROs simultaneously satisfy each wave. The solid curves are back-calculated waves using experimental D_a , R , orientation (Θ, Φ) and phase ρ_0 from simultaneous fits using the program ORIENT (see the text and SI) and structural parameters of a standard A-form duplex. The experimental RDCs are drawn as circles. The deviation of experimental RDCs from the curves at several residues may be due in part to deviation from the ideal A-form geometry. RDC of residue G78 was not used due to large error caused by peak overlapping.

Figure 5.

Combinations of possible orientations for duplexes H1 and H2 (a-d), and duplexes H1 and H3 (e-h) for riboA. Duplex H1, H2 and H3 are shown in green, red and in blue, and linkers in cyan. In general, two conformations in each combination are not viable because steric clashes between a linker and duplexes, as in conformations b, d, f and g. The coordinates of the possible conformers were generated by using the programs BLOCK and PACK and dawn using Pymol (DeLano, W.L. DeLano Scientific, San Carlos, CA, USA. <http://www.pymol.org>).

Figure 6.

The shape-assisted identification of the correct combination of DRO. The comparisons of the riboA molecular envelope and the initial structures built from viable orientation combinations leads to the conclusion that either a parallel or an anti-parallel arrangement of the three duplexes, conformer ae, can possibly fit into the envelope.

Figure 7.

(a) A two dimensional topological drawing (left) and the three dimensional starting structure (right) that was generated using the G2G toolkit. The orientations and phases (Θ, Φ, ρ_0) of the duplexes are given alongside the duplexes. The broken lines represent linker residues. Close contacts between H2 and H3 involving imino protons were unambiguously detected in the NOESY spectrum (NMR experiments section). The three duplexes have correct orientations and phases which determine the terminal positions of linkers and loops. The initial relative positions of the duplexes were arranged to be approximately consistent with the dimensions of the envelope. The hairpin loops and linkers, derived from the MOSAIC library, were added with arbitrary conformations so long as they can be connected to the ends of the three duplexes via 3'O to a phosphate atom (SI) without a sharp twist. The deviations from covalent geometry, such as abnormal bond lengths and angles at the joint junctions, were corrected in the regularization process in the SA calculation. (b) The initial fold of the riboA structure that was calculated using the rigid-body SA refinement protocol. The backbone RMSDs between the top 10% initial fold structures and the X-ray structure are 2.5 ± 0.2 and 3.4 ± 0.3 Å for the three duplexes and the whole molecule, respectively.

Figure 8.

The ensemble of global structures of the riboA RNA determined using the G2G “top-down” method, a SAXS curve comparison and the adenine ligand position. (a) The front (center) and side (left and right) views of the superimposed riboA backbone structures (top 10% lowest energy) superimposed within the envelope. The X-ray crystal structure is in cyan, the G2G structures in red, and the average structure of the ensemble in blue. The local structures of the hairpin loops and linker between H2 and H3 are relatively poorly defined. The backbone *RMSDs* of the ensemble (top 10% of 100 structures calculated) relative to the mean is about 0.8 $\text{\AA}$. (b) The superimposed correlation plots between the imino RDCs that were back-calculated from the ensemble of the top 10% lowest energy structures using the Powell-Grid-search method⁴¹ and the experimental RDCs. (c) The correlation plot between the imino RDCs that were back-calculated from the regularized average structure of the ensemble of the top 10% and the experimental RDCs. (d) The correlation plot between the imino RDCs that were back-calculated from the X-ray crystal structure (1Y26) and the experimental RDCs. (e) The comparison of experimental (circle), back-calculated SAXS curves based on the ensemble (red), average (blue) and the X-ray crystal structure (cyan) and the *RMSD* between the first and the third is about 0.29 ± 0.04 , which is calculated based on the logarithms of the normalized (i.e., $I(q=0)=1.0$) SAXS intensities. (f) The comparison of experimental PDDF (black), and those for the ensemble (red), the average (blue) and the X-ray crystal structure (cyan). (g) The comparison of the adenine ligand positions in the X-ray crystal structure (cyan) and the average structure of the top 10% lowest energy G2G structures. The adenine ligands are shown in sphere mode. The adenine position was defined by five close contacts that were detected in NOESY spectra (Figures 2c and S4 in SI).

Table 1
A tally of A-form duplex structures in RNA database

duplexes include basepair types: ['AU', 'UA', 'GC', 'CG', 'GU', 'UG']			
	Average	Stdev	Saenger
Dislocation	4.03	0.60	4.4 - 4.9
Tilt&	15.51	5.79	10 - 20.2
Rise	2.62	0.56	2.56 - 3.29
Rotation*	33.02	3.75	30 - 32.7
P-P dist	5.86	0.37	5.9
RNA backbone atoms: ["P", "OP1", "OP2", "O5 ", "C5 ", "C4 ", "O4 ", "C1 ", "C2 ", "O2 ", "C3 ", "O3 "].			
No.(bps)	No.(duplx)	RMSD vs ave	std
5	97	0.85	0.31
6	163	0.90	0.34
7	41	0.79	0.29
8	28	0.78	0.30
9	9	0.74	0.30
10	3	1.05	0.32
11	4	1.36	0.15
12	2	1.07	0.00
duplexes include basepair types: ['AU', 'UA', 'GC', 'CG', 'GU', 'UG', 'AC', 'GA', 'AG', 'CA', 'AA', 'CC', 'UC', 'UU', 'GG', 'CU']			
	Average	Stdev	Saenger
Dislocation	4.02	0.79	4.4 - 4.9
Tilt&	14.97	5.92	10 - 20.2
Rise	2.66	0.64	2.56 - 3.29
Rotation*	33.16	6.94	30 - 32.7
P-P dist	5.84	0.37	5.9
RNA backbone atoms: ["P", "OP1", "OP2", "O5 ", "C5 ", "C4 ", "O4 ", "CI'", "C2 ", "O2 ", "C3 ", "O3 "].			
No.(bps)	No.(duplex)	RMSD vs ave	std
5	149	0.93	0.40
6	158	1.04	0.40
7	73	0.98	0.47
8	50	1.18	0.72
9	8	1.30	0.29
10	19	1.50	0.57
11	10	1.52	0.33
12	8	1.54	0.41
Duplexes only include only basepair types: ['AU', 'UA', 'GC', 'CG']			
&: # Tilt&	16.02	5.69	10 - 20.2
: # Rotation	33.01	3.42	30 - 32.7

Note: Dislocation: dislocation of base-pairs from helical axis; tilt: base-pair tilt; rise: axial rise per nucleotide; rotation: rotation per nucleotide; P-P dist: intrastrand phosphate-phosphate distance; Saenger: structural parameters from reference. ⁴² The tally was taken based on a total of 347 duplexes of x-

ray crystal structures with resolution better than 3.0 Å and number of basepairs in duplexes ≥ 5 . *RMSD* vs ave: average values of the backbone *RMSD* of the individual structure in an ensemble vs the average structure of the ensemble. Six L-RNA structures in the database were excluded from the tally.

Table 2Structure determination statistics of riboA^a:

	Regularized average structure ^b	Top 10% ensemble
Restrains:		
Distances ^c	890	890
Hydrogen-bonded base pairs ^d	30	30
Residual dipolar couplings (RDC)	29	29
Molecular envelope size ^e	6	6
Deviations from idealized covalent geometry		
Bond, Å	0.003	0.006 ± 0.001
Angle (°)	0.75	0.88 ± 0.04
Improper (°)	0.50	0.40 ± 0.01
RMSD to experimental data		
NOE	0.15	0.19 ± 0.00
RDC ^f	1.57	1.60 ± 0.50
SAXS ^g	0.26	0.29 ± 0.04
Energy (kcal/mol)		
Total	654.3	45488.3 ± 162.0
NOE	2406.2	4246.3 ± 68.5
RDC	246.5	256.2 ± 50.1
VDW	89.2	7331.7 ± 104.5
SAXS ^g	67.0	85.6 ± 22.4
Backbone RMSD relative to the regularized average structure ^b : (Å)		0.81 ± 0.07
Backbone RMSD relative to crystal structure (Å)		
Three duplexes	2.5	2.5 ± 0.20
All except non-restrained loops	.30	3.0 ± 0.31

Note:

^aIn the Hybrid rigid-body simulated annealing calculation using Xplor-NIH-2.22. dihedral angle restraints were not used;^bexcluding loops;^c747 distance restraints to maintain A-form geometry; 13 distance restraints identified in the H1,H2 and H3 junction region after the initial fold calculation; 130 uniform generic distance restraints to maintain stacking throughout RNA non-duplex regions;^dhydrogen-bonded base pairs detected in NMR experiments;^edimension measurements (Figure S3 in SI);^frdc rmsd calculated using Powell search;^gSAXS data restraints. The RMSD of SAXS calculated in the same way as in Figure 8d. We used 20 equally q-spaced data points, up to q=0.38 Å⁻¹, in the SA calculation to speed up calculation.

Table 3

The values of D_a , R and orientations and phases (Φ , Θ , ρ_0) obtained from the individual fits and simultaneous fits for riboA.

		Duplex H1	Duplex H2	Duplex H3
individual fit	$(\Phi, \Theta, \rho_0)^a$	$(161, 297, 72)^b$	$(13, 86, 264)^b$	$(35, 107, 76)^b$
		$(161 \pm 0, 297 \pm 0, 72 \pm 0)^c$	$(14 \pm 2, 85 \pm 2, 264 \pm 1)^c$	$(24 \pm 11, 130 \pm 22, 67 \pm 11)^c$
	$(D_a, R)^a$	$(-22.6, 0.21)^b$	$(-28.8, 0.31)^b$	$(-16.7, 0.36)^b$
		$(-22.6 \pm 0.0, 0.21 \pm 0.00)^c$	$(-24.9 \pm 2.7, 0.32 \pm 0.02)^c$	$(-15.3 \pm 1.0, 0.30 \pm 0.04)^c$
simultaneous fit	$(\Phi, \Theta, \rho_0)^a$	$(151, 281, 101)^b$	$(22, 97, 259)^b$	$(40, 101, 63)^b$
		$(160 \pm 7, 288 \pm 18, 104 \pm 4)^c$	$(13 \pm 6, 94 \pm 6, 264 \pm 4)^c$	$(20 \pm 15, 100 \pm 4, 64 \pm 3)^c$
	$(D_a, R)^a$	$(-26.4, 0.35)^b$		
		$(-25.5 \pm 2.0, 0.28 \pm 0.03)^c$		

Notes:

^a values of (Φ , Θ , ρ_0) are given in degrees, and D_a in Hz;

^b obtained from the best fit;

^c the average values and the standard deviation of the top fits with a *RMSD* cutoff of 1.5 Hz.