New insights into the fundamental role of topological constraints as a determinant of two-way junction conformation

Anthony M. Mustoe, Maximillian H. Bailor, Robert M. Teixeira, Charles L. Brooks III and Hashim M. Al-Hashimi*

Departments of Chemistry & Biophysics, The University of Michigan, 930 North University Avenue, Ann Arbor, MI 48109-1055, USA

ABSTRACT

Recent studies have shown that topological constraints encoded at the RNA secondary structure level involving basic steric and stereochemical forces can significantly restrict the orientations sampled by helices across two-way RNA junctions. Here, we formulate these topological constraints in greater quantitative detail and use this topological framework to rationalize long-standing but poorly understood observations regarding the basic behavior of RNA two-way junctions. Notably, we show that the asymmetric nature of the A-form helix and the finite length of a bulge provide a physical basis for the experimentally observed directionality and bulge-length amplitude dependence of bulge induced inter-helical bends. We also find that the topologically allowed space can be modulated by variations in sequence, particularly with the addition of non-canonical GU base pairs at the junction, and, surprisingly, by the length of the 5′ and 3′ helices. A survey of two-way RNA junctions in the protein data bank confirms that junction residues have a strong preference to adopt looped-in, non-canonically base-paired conformations, providing a route for extending our bulge-directed framework to internal loop motifs and implying a simplified link between secondary and tertiary structure. Finally, our results uncover a new simple mechanism for coupling junction-induced topological constraints with tertiary interactions.

INTRODUCTION

A fundamental goal of structural biology and biophysics is to understand the underlying forces that govern the folding and conformational behavior of macromolecules. Stacking interactions and base-pairing have been shown to be the primary forces that drive the formation of RNA secondary structure, and thermodynamic frameworks that treat these interactions are available and can be used to predict RNA secondary structure with >70% success (1,2). Despite great advances in understanding the contributions of tertiary interactions in RNA folding and indeed its coupling to secondary structure and metal association (3–5), we are far from being able to predict an RNA fold based on a first principle understanding of RNA folding (6–8). The problem is even more complicated when one considers that RNA structures are typically highly flexible, and often adopt not one, but an ensemble of native conformations that can vary in response to cellular signals and/or changes in the physiochemical conditions (9,10). Much less is known about the principles that guide such dynamic aspects of RNA conformation (11).

RNA secondary structure is dominated by A-form helical domains that carry various receptors involved in tertiary interactions and are linked together by a variety of junctions, including bulges, internal loops, three-way and other higher order junctions. Lilley (12–14), Griffith (15), Draper (16) and Hagerman (17,18) showed more than two decades ago that bulges and internal loops induce directional bends in DNA and RNA duplexes by amounts dependent on the length, asymmetry and sequence of the junction as well as on ionic strength. Surveys of higher order junction atomic structures have also suggested that there is strong dependence of junction conformation on secondary structure (19–21). While it has long been recognized that junctions can direct specific inter-helical conformations, the degree to which secondary structure alone defines RNA global structure has only recently been fully appreciated and placed on a quantitative footing (22). In particular, studies by us (23) and the Herschlag group (24) have shown that two-way junctions impose steric and connectivity constraints that restrict the

---

*To whom correspondence should be addressed. Tel: +1 734 615 3361; Fax: +1 734 647 4865; Email: hashimi@umich.edu

relative orientation of A-form helices. These constraints include inter-helix steric clashes and connectivity constraints due to the finite length of the junction nucleotides linking the helices. Using a simple heuristic model that assumed a canonical A-form helix geometry and maximum inter-nucleotide distance of 4.9 Å, we showed that these constraints alone restrict the range of accessible inter-helical conformations to between 4 and 20%, and on average 7%, of the total space (23). Despite this narrow range the model accommodates ∼85% of all inter-helical orientations observed in all two-way junctions in the PDB, including bulges and internal loops (23).

Here, we explore how various parameters of interest influence the topologically allowed space with the goal of developing a deeper and more predictive understanding of how two-way junctions encode 3D RNA conformation and dynamics. In particular, we examine how the allowed conformational space varies with local variations in A-form helix geometry, the presence/absence of non-canonical base pairs at the junction, relative lengths of the 5′ and 3′ helices and overall junction topology (bulge, symmetric and asymmetric internal loops). We also put on a firm quantitative footing the well-known yet poorly understood directionality of junction-induced bends as well as explain the origins of recently reported correlated variations in the inter-helical twist and bend angles observed across two-way junctions (23,25). We also survey the structures of symmetric and asymmetric internal loops and show that the resulting allowed inter-helical conformational space is accurately modeled by assuming that internal loops will maximize non-canonical base pairing, marking a significant simplification in the link between RNA secondary and tertiary structure. Our results also uncover a new simple mechanism for coupling junction-induced topological constraints with tertiary interactions.

METHODS

Computing topologically allowed inter-helical orientations

We utilize the $H_iS_XH_jS_Y$ notation (23,26) to describe the topology of a two-helix junction where i and j specify the length of the 5′ and 3′ helices, respectively, and X and Y ($X \geq Y$) denote the number of single-stranded nucleotides in the 5′ and 3′ strands with the respective strands assigned in accordance to our prior conventions (23). For simplicity we refer to a junction topology in the body of the text as S_XS_Y, with the values of i and j implied by context; junctions from the PDB possess $i,j \geq 3$, and in reference to our model, unless otherwise stated, $i = j = 4$. For a structure of a given two-way junction topology, the inter-helical conformation can be defined through the use of the three Euler angles needed to describe the orientation of the chiral helical objects with respect to one another. As described previously (23,27) these angles, denoted α_h β_h γ_h, are obtained from the rotation matrix needed to transform the 3′ helix from its observed orientation to one where it is coaxially stacked upon its 5′ partner, and where α_h and γ_h are interpreted as twists around the 3′ and 5′ helices and β_h as an inter-helical bend angle (Figure 1A).

As described previously (23), this representation of inter-helical conformation can readily be inverted to compute the sterically and connectivity allowed conformations. Briefly, an idealized A-form helix with helical axis oriented along the positive z-axis from 5′ to 3′ direction was divided into two sub-helices, denoted the 5′ and 3′ helices, and with the angle between the 5′ helix closing base pair y-axis and the molecular y-axis, T_h, set to 51.1° (27). The entire system was translated such that the first phosphorous atom belonging to the 5′ helix on the strand oriented along the negative z-direction from the 5′ to 3′ end was located at the origin, serving as the rotation pivot point. A given α_h β_h γ_h inter-helical orientation was generated by applying the Euler rotation $R_Z(-\gamma_h) \times R_Y(-\beta_h) \times R_Z(-\alpha_h)$ to the 3′ helix while keeping the 5′ helix fixed, where R_Z and R_Y refer to Euler rotations of an object about the Z-axis and Y–axis, respectively (27). Rotations were performed for all non-degenerate (27) permutations of α_h β_h γ_h on a 5° grid. Conformations that resulted in inter-helix collisions, taken as an inter-helical atomic distance of <1.4 Å excluding hydrogen atoms and O1P and O2P atoms connected to the pivot P atom, were classified as sterically disallowed. If the distance between the last O3′ atom of the 5′ helix and first P atom of the 3′helix at the division between the two helices on the strand opposite the pivot was greater than the linker cutoff of 7.26 Å per bulged nucleotide then the conformation was classified as connectivity disallowed. This corresponds to a cutoff of 7.26 Å, 14.52 Å, 21.78 Å and 29.04 Å when calculating connectivity constraints for S_1S_0, S_2S_0, S_3S_0 and S_4S_0 bulges, respectively. The fraction of space allowed for a given junction is computed by dividing the number of allowed conformations by the number of unique α_h β_h γ_h permutations, 191 736 (27).

Computing topologically allowed space with translational relaxation

We computed the topologically allowed conformations when allowing for some degree of inter-helix translation that may relieve inter-helix steric collisions. Translational degrees of freedom were defined as deviations in the x, y and z coordinates of a 3′ helix relative to the position of an idealized reference 3′ helix that has the same α_h β_h γ_h orientation. Analysis of crystal structures of S_1S_0, S_2S_0, S_3S_0 and S_4S_0 junctions in our database (see PDB survey of two-way junctions) revealed that these translational deviations are roughly isotropic in x, y and z, centered near 0, and on average $\sqrt{x^2+y^2+z^2} \leq 2$ Å. We therefore allowed 1–2 Å of translational freedom in each of the x, y and z directions to be representative of that found in the PDB. To simulate the effects that these extra degrees of freedom may have to relieve inter-helix steric collisions, we modified our rotation protocol. As in the translation-free procedure, a rotation $R_Z(-\gamma_h) \times R_Y(-\beta_h) \times R_Z(-\alpha_h)$ is applied to the 3′ idealized helix, iHII, while keeping the 5′ helix, iHI, fixed. Following each rotation, iHII was translated by all combinations of −1, 0 and 1 Å in the x, y and z directions for a 1-Å cutoff, or by all combinations of −2, −1, 0, 1 and 2 Å for a 2-Å cutoff. This procedure captures >80% of the inter-helix translations observed in

the PDB. A given inter-helical angle was accepted provided one of the translations alleviated any existing steric collisions. Connectivity-allowed conformations were taken to be the same those for the translation free case.

Computing the topologically allowed space with non-idealized helices

Eighteen 8-bp continuous helices of different sequences were chosen from junctions returned from a search of the RNA FRABASE (28) and were divided into 2×4-bp helices, named the 5′ and 3′ helices, respectively (Supplementary Table S1). The helices were chosen such that all combinations of WC pairs and two GU wobble pairs were represented at the 'junction' between the two 4-bp helices. To remove any inter-helical bends or twists the 5′ and 3′ helices of each 2×4-bp system were individually superimposed on top of their corresponding helices in the idealized A-form system used in the idealized topologically allowed computations. In cases where the pivot P atom was not located at the origin after the superpositions, a translation was applied to both 5′ and 3′ helices such that the pivot was at the origin as desired. After these 'non-idealized' systems were prepared, the same rotation procedure used for the idealized helices was used.

Cylinder rotations

An in-house Python script was used to generate two cylinders of diameter 17.5 Å, height 24 Å and surfaces represented by a ~1.5-Å square mesh (distance between points was closer on cylinder ends). The chosen diameter of 17.5 Å corresponds to the distance between P(i) and O3′ (j) across the idealized A-form helix in Watson–Crick base-paired residues i and j. The bottom cylinder axes were vectors originating at $(8.75, 0, -0.75)$ and $(0, 0, 0)$ and running along the negative z direction for the bulge and center connected systems, respectively. The top cylinder axes began at $(8.75, 0, 0.75)$ for the bulge and $(0, 0, 14.52)$ for the center connected systems and ran along the positive z direction. Orientations that resulted in mesh points from opposing cylinders being <1.5 Å apart were determined to be sterically disallowed. Connectivity constraints for the bulge-like system were implemented by requiring that the distance between the linker termini, $(17.5, 0, -0.75)$ and $R(-\gamma_h, -\beta_h, -\alpha_h) \times (17.5, 0, 0.75)$, were <14.52 Å. The linker length of the center-connected cylinder system is invariant across all rotations and is thus free of connectivity constraints.

PDB survey of two-way junctions

A search of RNA two-way junctions was performed using the RNA FRABASE (28) on 25 June 2010 for $H_3S_1H_3S_0$, $H_3S_2H_3S_0$, $H_3S_3H_3S_0$ and $H_3S_4H_3S_0$ bulge motifs; on 27 July 2010 for $H_3S_1H_3S_1$, $H_3S_2H_3S_1$, $H_3S_3H_3S_1$, $H_3S_4H_3S_1$, $H_3S_2H_3S_2$, $H_3S_3H_3S_2$, $H_3S_4H_3S_2$, $H_3S_3H_3S_3$, $H_3S_4H_3S_3$ and $H_3S_4H_3S_4$ internal loop motifs; and 22 September 2010 for all H_6H_6 helices (or 6-bp continuous helices) as described previously (23). Inter-helical orientations for each identified junction were measured using in-house Perl scripts and the α_h β_h γ_h convention of Bailor *et al.* (23,27,29). Unless otherwise noted, angles

for nuclear magnetic resonance (NMR) structures were measured on the first model reported in the ensemble. The $<2\%$ of junctions identified by the RNA FRABASE whose flanking helices were >2 Å in backbone RMSD from idealized A-form, had a strand possessing all deoxyribose sugars, or had non-continuous chain connectivity at the junction were excluded from our analysis. Measured orientations are designated as falling within the topologically allowed space if the shortest distance between the $(\alpha_h^i \beta_h^i \gamma_h^i)$ measured for conformation i and $(\alpha_h^T \beta_h^T \gamma_h^T)$ for topologically allowed point T, $d = \sqrt{(\alpha_h^i - \alpha_h^T)^2 + (\beta_h^i - \beta_h^T)^2 + (\gamma_h^i - \gamma_h^T)^2}$, is $<5°$. Bulge-linker distances were calculated along the bulge strand between the O3′ atom of the 5′ helix closing base pair and the P atom of the 3′ helix's first base pair with an in-house Perl script (23).

In-house Perl scripts were also used for the determination of the inter-helical translations of each junction by adapting the procedure used to determine inter-helical angles. In obtaining the α_h β_h γ_h for a junction a reference helix, denoted iH2′, is superimposed on top of the 3′ helix and then this iH2′ is transformed to an orientation that is perfectly coaxially stacked upon the 5′ helix (27). This transformation is done by first translating the center of mass of iH2′ to the origin, performing a rotation, and then translating the rotated iH2′ from the origin to its coaxially stacked position. The difference between these first and second translations defines a net translation of the center of mass of iH2′ from its coaxially stacked position. This measured net translation is composed of two terms, the movement of the center of mass due the rotation of the helix and the deviation of the helix away from this 'pure rotation' position. We calculated the 'pure rotation' position of a given α_h β_h γ_h by applying a rotation of $-\gamma_h$ $-\beta_h$ $-\alpha_h$ to the 3′ helix of an idealized, coaxially stacked helical system, and then determined the net translation incurred when transforming this 'pure rotation' conformation back to its original coaxially stacked state. The difference between the net translation measured for the PDB conformation and net translation of the 'pure rotation' conformation was taken as the translational deviation of the junction.

FR3D analysis of internal loop motifs

FR3D (Find RNA 3D) (30) pairwise interaction files that were available for PDBs within our database were downloaded from the FR3D website at http://rna.bgsu.edu/FR3D/AnalyzedStructures/PDB/PDB_interactions.html on 23 August 2010, where PDB is replaced with the name of the PDB structure of interest. The interaction files were parsed with in-house Python scripts to search for base pairing and stacking interactions. Bases were assigned as 'looped-in' if FR3D indicated that they possessed a stacking interaction with residues of either strand of the junction or with the bases comprising the closing base pairs of the flanking helices. The $<2\%$ of junctions without available FR3D files were excluded from the analysis.

Identification of tertiary interactions

Intra-RNA tertiary interactions were identified by searching the FR3D interaction files for pairing or stacking interactions between junction or closing base pair nucleotides and non-junction nucleotides with an in-house Python script. Junctions without available FR3D files were by default classified as having tertiary contacts. A separate in-house python script was used to search for protein, ligand and RNA tertiary interactions that were not identified by FR3D. PDB files corresponding to each junction in our database were searched exhaustively for any C, N, S, O or P atoms within 4 Å of junction or closing base pair atoms while excluding atoms from the chain(s) containing the junction and all 'HOH' O atoms.

Manual annotation of $H_3S_{\geq 2}H_3S_2$ junctions

Crystal structures of $H_3S_{\geq 2}H_3S_2$ junctions within our database were sorted according to sequence and inter-helical angles. Junctions with identical sequences and within an 11.5° distance in α_h β_h γ_h space were clustered together and assumed to have common features. One junction from each group was then manually inspected and each loop base was qualitatively classified as to whether it was non-canonically base paired, 'near non-canonically paired' or neither, and whether it was stacked into the junction. Base paired residues were defined as two bases from opposing strands having at least one hydrogen bond donor to acceptor distance of ≤ 3.6 Å and angle $\geq 125°$. Near non-canonical base pairs were defined as two bases from opposing strands that were roughly coplanar and looped into the junction but did not possess clear hydrogen bonds. Tertiary contacts, defined as protein, ligand, crystal or other non-junction RNA hydrogen bonding contacts with any junction or closing base pair base (excluding backbone contacts) were also recorded.

RESULTS AND DISCUSSION

Bulges

Sterics. We performed an exhaustive search of the inter-helical conformational space accessible to a bulge by carrying out a series of rigid body rotations to two, 4-bp-long, idealized A-form helices. In Figure 1B, we show the inter-helical orientations allowed after excluding all conformations that possess inter-helix steric clashes. The angles α_h and γ_h, denoting the twists of the 3′ and 5′ helices about their respective helical axes, and β_h, the inter-helical bend angle, are plotted as $3 \times 2D$ projections of the 3D α_h β_h γ_h space of inter-helical conformations (27). Of the theoretically possible α_h β_h γ_h orientations (27), 53% are excluded by steric constraints alone. Visual inspection of the projections also reveals that the allowed orientations are asymmetric in α_h and γ_h. Calculations using perfectly symmetric cylinders (Supplementary Figure S1) reveal that the asymmetry of the sterically allowed space arises from the inherent asymmetry and chirality of the helices.

A notable consequence of the anisotropy of the sterically allowed space is that it encodes a specific directionality to any bending that occurs between the two helices (Figure 1C). Steric collisions between the closing base pairs of the two helices make it such that the density of accessible states is much greater for bends that correspond to movement of the 3′ helix into the space above the plane of the 5′ closing base pair and 'away' from the 5′ helix, rather than into the closing base pair plane and 'towards' the 5′ helix. This result provides a quantitative and simple explanation for the directionality of RNA and DNA bulge-induced bends (12,14,16,31,32).

Interestingly, the sterically allowed orientations are also dependent on helix length. While 53% of conformations are excluded with 4-bp helices, the fraction excluded increases to 62% and 69% for 6-bp and 8-bp helices, respectively, and decreases to 49% with 3-bp helices. The steric constraints only depend on the length of the 5′ helix when the 3′ helix has >3 bp (Figure 1D). This can be attributed to the fact that the helices are asymmetric about the rotation origin (pivot P atom), which is located closer to atoms in the 5′ helix as compared to the 3′ helix. As a result, elongation of the 5′ helix results in steric occlusion, whereas the new volume occupied by elongation of the 3′ helix falls outside of the sphere swept by the 3′ helix. Thus, varying the length of the 5′ helix can, in principle, modulate the allowed inter-helical space.

The above steric model assumes an idealized A-form helix geometry. However, RNA helices can deviate from the canonical geometry and this in turn can affect the sterically allowed conformations. To examine these contributions, we computed the sterically allowed conformations when using 18 non-idealized 2×4-bp helices taken from the PDB (Supplementary Table S1). Each sequence had different closing base pairs at the site of the junction, one representative for each of the 16 unique canonical base pairs, and two sequences possessing a GU wobble pair at the junction. The fraction of sterically disallowed conformations varied between 35 and 50% and on average 44% of the total space, in good agreement with 53% when using idealized A-form helices. On average, 97% of the orientations that are allowed when using idealized A-form helices are also allowed when using non-idealized helices, whereas an average of 82% of the orientations sampled by the non-idealized helices are accessible to idealized A-from helices. Thus, local distortions in the helix geometry can slightly increase the space that can be sampled and allow access to otherwise sterically disallowed inter-helical conformations.

Our model also assumes that the two helices are fixed at the pivot point. However, some of the steric constraints can, in principle, be relaxed through translations of the helices relative to one another. Analysis of the bulge motifs within our structural database revealed that the average net translation from the origin is ≤ 2 Å. If the sterically allowed space is recalculated with an allowance for ~1 Å translational relaxation the fraction of sterically allowed conformations increases from 47% to 68% and with ~2 Å to 80%. Studies that apply a more sophisticated treatment of the coupling between rotational and translational degrees of freedom are needed to clarify the significance of inter-helix translations in determining junction conformation.
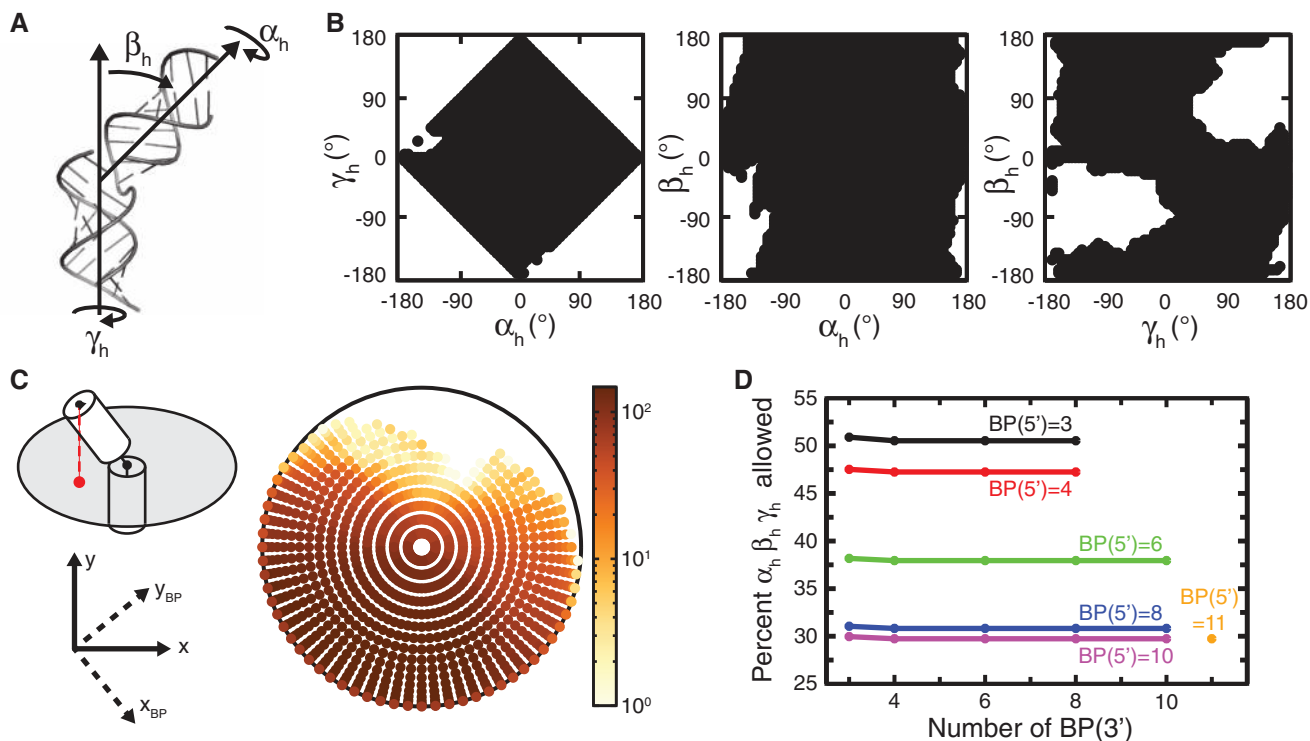
Figure 1. Sterically allowed inter-helical conformations. (**A**) α_h β_h γ_h convention used to describe inter-helical orientations and (**B**) α_h–γ_h, α_h–β_h and γ_h–β_h projections of the conformations allowed (black) given the steric constraints. (**C**) Density of allowed conformations of the upper (3′) helical axis in relation to the fixed lower (5′) helical axis projected in the plane perpendicular to the lower helical axis. A cartoon of the relationship between a helix conformation and a projected point as well as the coordinate frame of the closing bp of the 5′ helix (67) in relation to the coordinate frame of the projection is shown for reference. (**D**) Fraction of α_h β_h γ_h space sterically accessible as a function of length of the 5′ and 3′ helices.

As a whole, these results suggest that steric constraints exclude 20–53% of the possible inter-helical orientations. However, we emphasize that these values represent a lower bound to the degree of steric confinement in RNA junctions. The heavy atom van der Waals diameter of 1.4 Å used in our models is well below commonly accepted values of 3–4 Å. Inclusion of hydrogens and the solvation shell will also increase the effective van der Waals radii. Finally, it is likely that electrostatic repulsion may serve to further narrow the space (33).

Connectivity. The set of conformations that are accessible to a bulge junction is also subject to connectivity constraints arising from the finite length of the bulged nucleotides linking the helices. We approximate these constraints by requiring that the end-to-end distance between the last O3′ atom of the lower (5′) helix and the first P atom of the upper (3′) helix to be less than a given cutoff length. We previously used a bulge length cutoff equal to 4.9 Å per nucleotide, which was assumed to be the maximum length of a given residue (23). Here, we compute the average per bulge residue distance length for all $H_3S_{4 \geq X \geq 1}H_3S_0$ junctions within our database. The computed distribution is roughly Gaussian with a mean of 5.78 Å and a standard deviation of 1.48 Å (Supplementary Figures S2 and S3). In the new calculations, we use a cutoff of 7.26 Å per bulged nucleotide, or the mean plus one standard deviation, which accommodates 84% of the per-nucleotide bulge lengths found within the PDB. With the distance cutoff,

the connectivity constraints alone exclude 95, 83, 62 and 31% of the possible orientations for S_1S_0, S_2S_0, S_3S_0 and S_4S_0 bulges, respectively (Figure 2A).

Our analysis does not account for intra-linker and linker–helix steric clashes. As a first estimate of the impact of such steric effects, we identified orientations where the line segment drawn between the O3′ and P atoms (the linker termini) was <2 Å from either helix and within 2 Å of the cutoff distance, meaning a bulge connecting the two termini would have no alternative path. This analysis suggests that 0%, 1%, 2% and 4% of the connectivity allowed conformations may be inaccessible due to linker-helix steric overlaps for S_1S_0, S_2S_0, S_3S_0 and S_4S_0 motifs, respectively. When considering both steric and connectivity constraints, as discussed in greater detail in following sections, 0%, 2%, 8% and 11% of the S_1S_0, S_2S_0, S_3S_0 and S_4S_0 allowed conformations are estimated to contain linker-helix steric overlaps.

The connectivity constraints, and their variation with bulge length, depend critically on the ratio of the per-nucleotide length of the bulge, which is determined by the backbone structure, and to the helix diameter, as determined by base-pair structure. In Figure 2B, we plot the results of a derived mathematical relationship (Supplementary Figure S4) that relates the set of accessible α_h β_h γ_h angles to the number of bulged nucleotides and the 'helix' diameter. At the effective diameter of the A-form helix, increasing the number of bulged nucleotides results in incremental increases in the size of the allowed

space up to 5 nt, above which there are no connectivity constraints. However, for a hypothetically smaller helix diameter, there are fewer, more coarsely spaced putative levels. Conversely, there are greater, more finely spaced levels for larger helix diameters, underlining the unique tuneability of RNA junction conformation that is endowed by RNA's chemical structure.

A strong prediction of Figure 2B as well as the computed topologically allowed space is that while the accessible space increases with bulge length, it plateaus at 5 nt. This prediction is in general agreement with the FRET experiments of Gohlke *et al.* (34) where the FRET efficiency between probes located at the ends of the two helices increased as the number bulged nucleotides increases from one to five but then changes very little for bulges that are 7 and 9 nt long. Additionally, gel electrophoresis experiments also indicate a convergence to a common electrophoretic mobility for bulges that are longer than 6 nt (14,18,35). Interestingly, a search of bulge motifs in PDB with the RNA FRABASE (28) reveals a marked scarcity of bulges with >5 nt (Figure 2C). This suggests that bulges primarily play the role of tuning the inter-helical orientation and there is rarely reason to have bulges >5 nt long.

Inter-helical correlations. A notable feature of Figure 2A is that the inter-helical angles do not vary independently of one another; rather we observe two linear correlations, one positive and one anti, in the α_h–γ_h projection. Thus, each helix can only twist about its axis a certain amount, generally determined by the length of the linker, before the other helix must compensate with a correlated twist or bend to maintain linker connectivity (Supplementary Figure S5). This codependence can be attributed to the unique topology of the nucleic acid two-way junction where the helices are linked at two positions around the helix circumference. By comparison, an artificial system in which helices are connected through a single, centrally located linker exhibits no correlations between α_h and γ_h (Supplementary Figure S1D).

The source of this correlation can be found in the mathematical relationship linking the linker length, L, to the fixed helix diameter, D, and the applied $\alpha_h \, \beta_h \, \gamma_h$ rotation transformation (Supplementary Figure S4, Supplementary Information):

$$\frac{L^2}{D^2} = 2 - 2\cos(\alpha_h)\cos(\beta_h)\cos(\gamma_h) + 2\sin(\alpha_h)\sin(\gamma_h) \qquad (1)$$
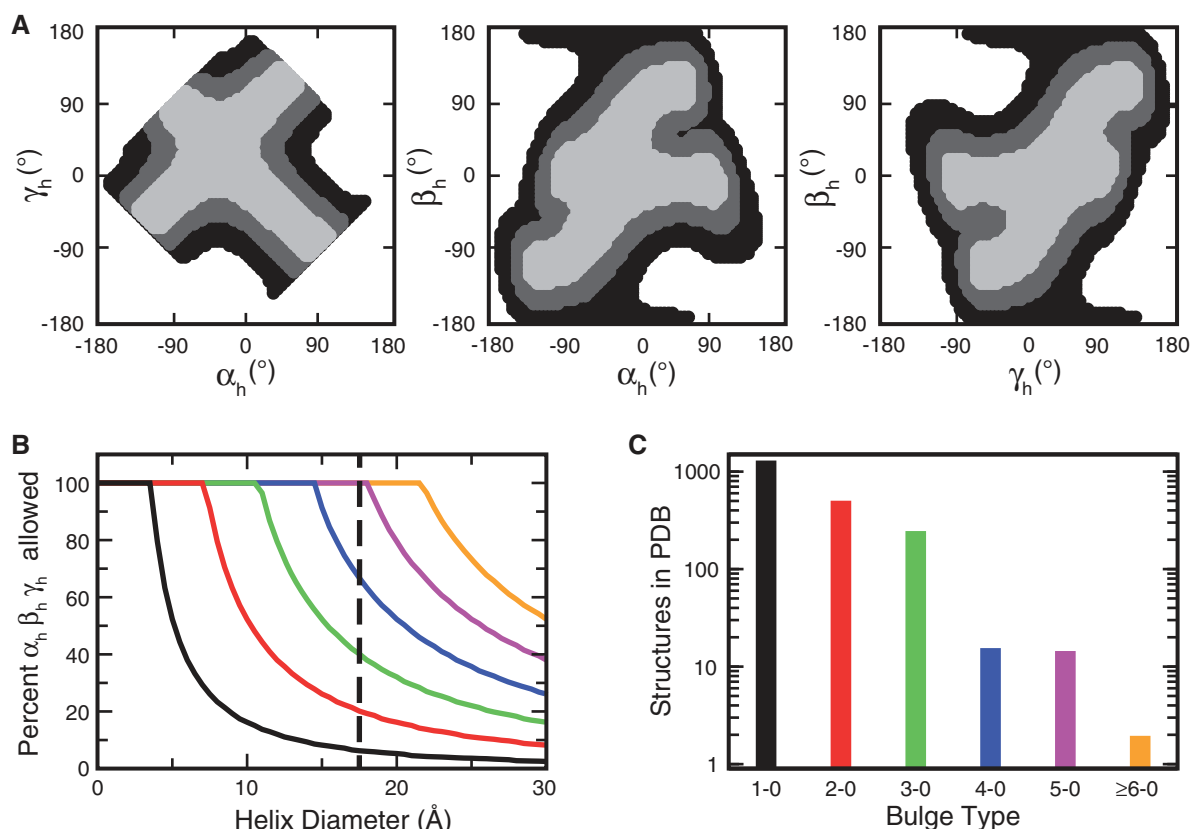
Figure 2. Connectivity allowed inter-helical conformations. (**A**) The set of conformations allowed given the connectivity constraints of 1 (light gray), 2 (dark gray) and 3 (black) bulged nucleotides, where the allowed conformations for *i* number of bulged nucleotides include those for *i−1*, *i−2*, . . .1 bulged nucleotides. (**B**) The percentage of $\alpha_h \, \beta_h \, \gamma_h$ orientations that are connectivity allowed as a function of effective helix diameter for a linker cutoff of 1 (black), 2 (red), 3 (green), 4 (blue), 5 (magenta) or 6 (orange) bulged nucleotides. The effective linker diameter of an A-form helix, 17.5 Å, is marked with a vertical dashed line. (**C**) The number of S_1S_0 (black), S_2S_0 (red), S_3S_0 (green), S_4S_0 (blue), S_5S_0 (magenta) and $S_{\geq 6}S_0$ (orange) structures in PDB.

A derivative of Equation (1) with respect to an arbitrary time, t, while maintaining a constant inter-helical bend, β_h, and constant linker length, shows that changes in α_h must be offset by corresponding changes in γ_h through the equality

$$\frac{d\alpha_h}{dt} = \frac{\sin(\alpha_h)\cos(\gamma_h) + \cos(\alpha_h)\cos(\beta_h)\sin(\gamma_h)}{-\cos(\alpha_h)\sin(\gamma_h) - \sin(\alpha_h)\cos(\beta_h)\cos(\gamma_h)} \frac{d\gamma_h}{dt} \quad (2)$$

highlighting the interdependence of these quantities. In the proper limits, Equation (2) reduces to

$$\frac{d\alpha_h}{dt} = \pm \frac{d\gamma_h}{dt} \quad (3)$$

and changes in α_h, which denotes a twist of the 3′ helix, must be exactly offset corresponding changes in γ_h, or twist of the 5′ helix, and explaining the positive and negative α_h–γ_h correlations in Figure 2A.

Generally, we do not expect β_h and linker length to remain fixed; rather all variables will be sampling values throughout their allowed range. Under this scenario, α_h and γ_h are no longer directly correlated as a change in α_h or γ_h could be offset by a change in the linker length or β_h. However, the limited range of the linker length and a limited β_h range due to steric constraints ensure that the system will experience the α_h–γ_h constraints, giving rise to on-average correlated movements of the two helices as has been previously observed from experiment (23,25). These correlations will be particularly strong in junctions with smaller bulges where the range of accessible linker lengths and β_h's is more significantly restricted.

Union of connectivity and sterics. The full range of accessible inter-helical conformations is restrained by both connectivity and steric constraints. In Figure 3A, we show the intersection of both of the spaces. Visual comparison between Figures 2A and 3A reveals that sterics and connectivity exclude distinct parts of conformational space. For the case of S_1S_0 and S_2S_0 bulges, 95% and 83% of the space is disallowed when considering connectivity alone, yet when combined with the steric constraints, 98% and 95%, respectively, of the space is excluded. While the connectivity constraints are symmetric, this symmetry is broken through the imposition of the steric constraints. Conspicuously, the negative α_h–γ_h correlation for the S_1S_0 and S_2S_0 bulges disappears and the range of accessible β_h angles decreases; this implies that bends that maintain or shorten the distance between the two closing base pairs, and therefore preserve a short linker length, are precluded due to steric collisions, whereas bends that result in an opening of the space between the pairs, and are accessible to bulges with longer linker lengths, are not.

Nevertheless, this picture of the topological constraints is incomplete—A-form helices themselves possess intrinsic orientational degrees freedom that are independent of the constraints imposed by the bulge. Analysis of all coaxially stacked 2×3-bp helices, or all continuous 6-bp helices, in the PDB revealed that the $\alpha_h \beta_h \gamma_h$ values for these systems varied with $-20 \le (\alpha_h + \gamma_h) \le 20$ and $-25 \le \beta_h \le 35$ (Supplementary Figure S6). To obtain a complete topologically allowed space we take the union of those conformations that satisfy either these constraints or those calculated for the bulge topology of interest with an
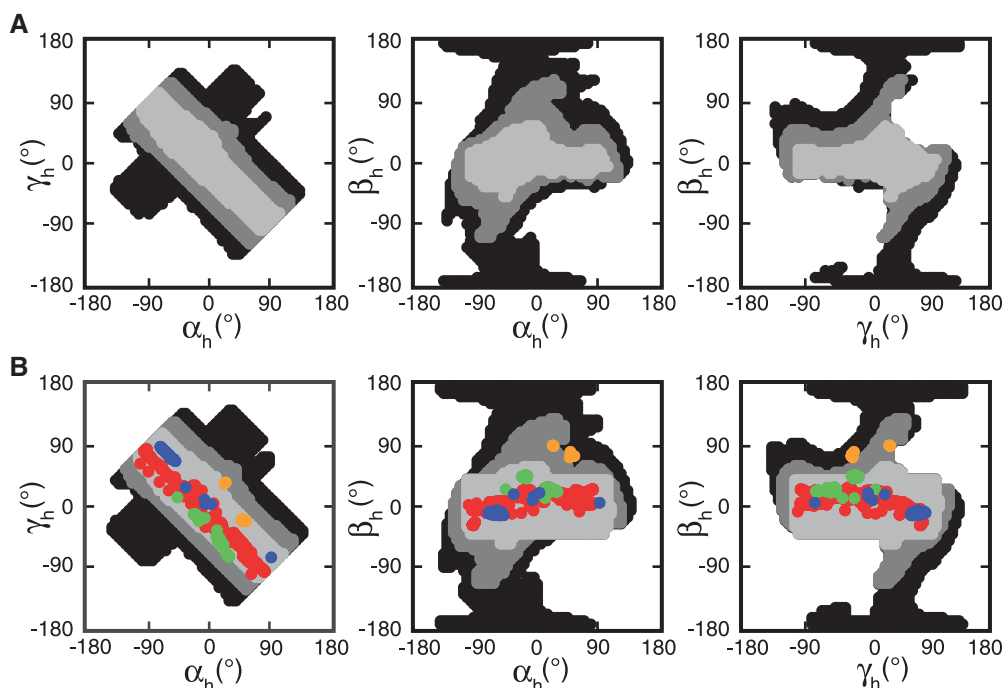
Figure 3. Union of sterics and connectivity constraints. (**A**) Plane projections of the set of conformations satisfying both the steric and connectivity constraints for 1 (light gray), 2 (dark gray) or 3 (black) bulged nucleotides, where allowed conformations for i number of bulged nucleotides are inclusive of those for $i-1$, $i-2$, . . .1 bulged nucleotides. (**B**) Projections of the total topologically allowed space, obtained by taking the union of that in (A) with those conformations accessible to continuous helices and an additional 5° error. Inter-helical orientations measured for 728 $H_3S_1H_3S_0$, 160 $H_3S_2H_3S_0$, 65 $H_3S_3H_3S_0$ and 10 $H_3S_4H_3S_0$ crystal structure junctions in the PDB are plotted in red, green, blue and orange, respectively.

additional 5° in each angle added to the bounds of this space. This 5° padding accounts for the errors inherent in the measurement of an arbitrary $\alpha_h\ \beta_h\ \gamma_h$ angle from the PDB as well as for capturing conformational states that may be accessible due to deviations from idealized A-form helical structure.

To test our framework we measured how well we could capture the conformations found in the PDB. In Figure 3B, we plot the predicted allowed conformations overlaid with conformations measured for the 963 non-GU closing base pair crystal structure junctions in our database. Table 1 summarizes the results for the both the fraction of PDB structures that fall within the predicted idealized allowed space and the fraction of possible $\alpha_h\ \beta_h\ \gamma_h$ states that the space is confined to. It is remarkable that our model samples as little as 5% of the allowed space, and on average 18%, and yet is able to accommodate 100% of all known bulge structures. These results compare favorably to our prior study (23) where the topologically allowed space was confined to 4%, 8%, 14% and 20% of possible states for S_1S_0, S_2S_0, S_3S_0 and S_4S_0 motifs, respectively, accommodating ~84% of 1195 cryo-EM, NMR, and crystal structure junctions. The differences in the degree of confinement for the larger bulge sizes are a function of the longer linker length used in the current model (7.26 Å versus 4.9 Å per nucleotide).

While we have focused on X-ray structures in the above analysis, 100% and 97% of the 73 cryo-EM and 58 NMR junctions without GU closing base pairs, respectively, fall within the predicted region (Supplementary Figure S7). The two NMR outliers [(1A9L,2A9L),36] are from the same set of experimental restraints and have $\alpha_h\ \beta_h\ \gamma_h$ angles 12° and 16° away from the closest predicted allowed point. A MolProbity (37) analysis of the structures revealed that there are significant steric clashes and deviations from accepted stereochemistry at the site of the S_3S_0 junction in question, suggesting that structural inaccuracies may be the cause of this disagreement. We also excluded junctions containing GU closing base pairs due to their propensity to induce significant local distortions to the base pairing geometry. Stand-alone analysis of

all 899 GU-containing junctions showed that 87% fall within our predicted ranges, where those falling outside are on average 9° from the nearest predicted point (Supplementary Figure S7). This suggests that sequence induced helix distortions can play a role in modifying the topologically allowed space. Indeed, 100% of the junctions containing GU pairs fall within at least one of two allowed distributions generated from non-idealized GU closing base pair sequences. However, it should be noted that comparison between these distributions and the idealized derived distribution share the majority of their points.

An interesting observation of Figure 3B and Supplementary Figure S7 is that certain regions of the allowed space are rarely, if ever, sampled in the PDB. This includes conformations where helices bend into each other along the major-groove face, which likely leads to unfavorable electrostatic effects. Further refinement of the model with incorporation of electrostatic interactions and larger van der Waals radii, as mentioned above, would likely reveal that such states are largely inaccessible.

Dynamic properties of the topologically allowed space. Prior studies of two-way junction motifs suggest that these systems are highly dynamic on the nanosecond to microsecond timescales (25,38), and in the case of HIV-1 TAR (23,39), roughly sample the entire topologically allowed space. If bulge systems are in fact sampling the entire topologically allowed space one would expect an ensemble average over all conformations to be similar to that observed from experiments with sensitivities less than these timescales. Remarkably, if we assume each conformation is equally weighted and calculate the ensemble average bend, or $|\beta_h|$, expected from the topologically allowed space, we obtain a result that closely replicates the values Zacharias and Hagerman (18) obtained for polyA and polyU bulges using transient electronic birefringence (Figure 4A). These values are also in agreement with estimates from FRET measurements of RNA of Gohlke *et al.* (34) and in qualitative agreement with studies performed on DNA (34,40,41). A corollary to these results is that the bent state is entropically favored for longer loops. Disregarding the conformational preferences of the loop, there is a significant entropic penalty to forming bent states ($|\beta_h| > \sim 60$) for S_1S_0 and S_2S_0 bulges, while the opposite holds true for $S_{X\geq 3}S_0$ bulges, where forming a coaxial stack (restricting $|\beta_h| < \sim 35$) is disfavored (Figure 4B). Thus, the varying conformational preferences of different junction topologies can be attributed to secondary structure encoded modulation of the topologically allowed space.

Internal Loops

Structural survey of internal loop conformations. In previous work, Bailor *et al.* (23) showed that the behavior of asymmetric internal loops can be explained by transforming a given internal loop into a reduced bulge representation. This is achieved by maximizing stacking and non-canonical base pairing between residues in the internal

Table 1. Fraction of crystal structures that fall within the predicted topologically allowed space for non-GU closing base pair crystallographic RNA junctions identified with the RNA FRABASE (28) and fraction of total unique $\alpha_h\ \beta_h\ \gamma_h$ permutations allowed for the idealized, non-ideal and idealized with translation helical systems

Bulge type	Fraction of PDBS sampled (total)	Idealized $\alpha_h\ \beta_h\ \gamma_h$ allowed	Range of non-ideal $\alpha_h\ \beta_h\ \gamma_h$ allowed (mean)	Translation $\alpha_h\ \beta_h\ \gamma_h$ allowed (w/o 5° Err)
S_1S_0	1.0 (728)	0.048	0.047–0.055 (0.049)	0.062 (0.028)
S_2S_0	1.0 (160)	0.094	0.090–0.10 (0.098)	0.14 (0.093)
S_3S_0	1.0 (65)	0.19	0.20–0.24 (0.22)	0.33 (0.23)
S_4S_0	1.0 (10)	0.37	0.38–0.45 (0.42)	0.61 (0.49)

Topologically allowed spaces represent the intersection of the sterically and connectivity allowed spaces with the addition of the $\alpha_h\ \beta_h\ \gamma_h$ accessible to a continuous helix and 5° error padding in each angle.
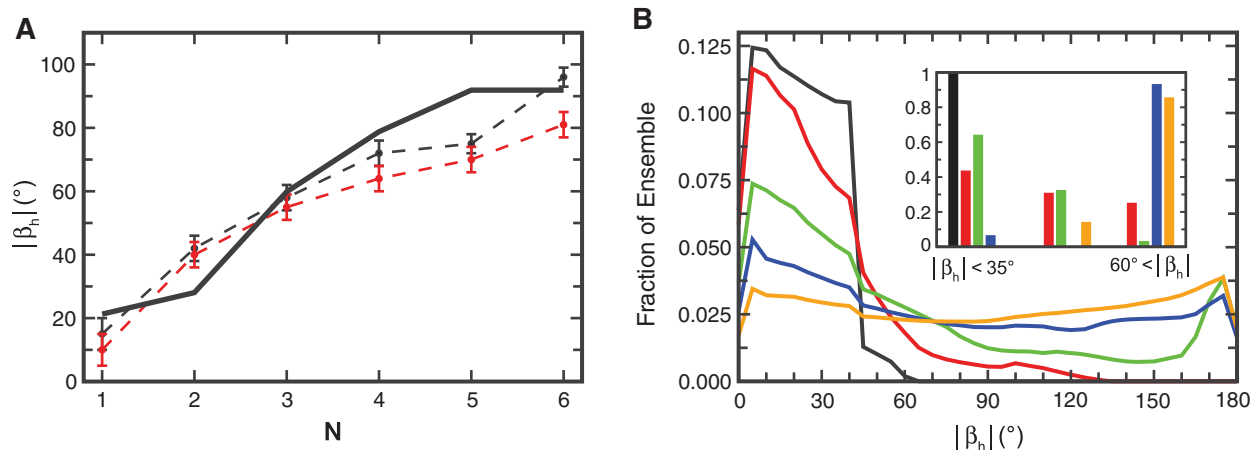
Figure 4. (A) The inter-helical bend angle for N bulged nucleotides for poly-A (dashed black) and poly-U (dashed red) RNA bulges as measured by Zacharias and Hagerman (18) in the absence of Mg^{2+}, and $|\beta_h|$ averaged over the total topologically allowed space as a function of N bulged nucleotides. (B) Populations of topologically accessible conformations as a function of $|\beta_h|$ for S_1S_0 (black), S_2S_0 (red), S_3S_0 (green), S_4S_0 (blue) and S_5S_0 (orange) bulges. The inset shows the fraction of structures from the PDB with coaxial ($|\beta_h| < 35°$), semi-bent ($35° \leq |\beta_h| \leq 60°$) or bent ($|\beta_h| > 60°$) conformations for 1253 S_1S_0 (black), 487 S_2S_0 (red), 238 S_3S_0 (green), 15 S_4S_0 (blue) and 14 S_5S_0 (orange) junctions.

loop. This 'reduced' bulge motif is subject to the characteristic topological constraints, but with the addition of Y base pairs inserted into the junction. By assuming that each inserted base-pair results in a change in $\alpha_h + \gamma_h = -34°$, the twist of the A-form helix, one could readily generate the allowed space for asymmetric internal loops by shifting the α_h and γ_h of the reduced bulge set of topologically allowed conformations by $Y*-17°$. This simple model was able to explain the observation that the α_h β_h γ_h measured for asymmetric internal loops mirrored the distribution found in bulges with systematic shifts in α_h and γ_h.

To investigate this assumption of maximum stacking of internal loop residues, we used the FR3D (30) nucleic acid structure analysis software to probe the conformation of nucleotides inside internal loops. Analysis of our PDB library consisting of crystal structures for all S_1S_1, S_2S_1, S_3S_1, S_4S_1, S_2S_2, S_3S_2, S_4S_2, S_3S_3, S_4S_3 and S_4S_4 junctions obtained from the RNA FRABASE (28) reveals a clear preference for bases in the Y strand to loop into the junction (Figure 5). Moreover, in >90% of the cases where all Y bases are looped in, at least an equivalent number of X strand bases are looped-in as well (filled bars of Figure 5). Interestingly, the majority of the loops that do not adopt fully looped-in conformations are directly involved in tertiary interactions (see 'Methods' section). Manual examination of the structures that were not fully looped-in and did not have a tertiary contact revealed that all were either influenced by crystal contacts or appeared to be missing coordinates where a protein would be located.

As hypothesized from the observations of Bailor *et al.* (23), we expected that not only would the junctions adopt fully looped-in conformations, but also the looped in bases should form non-canonical base pairs that propagate the flanking helices. To examine what fraction of looped in bases form non-canonical base pairs we hand annotated all S_2S_2, S_3S_2 and S_4S_2 junctions. The results revealed that >80% of the junctions exhibited a maximum

number of non-canonical base pairs or near non-canonical base pairs, which we define as cases when bases from opposing strands were looped into the junction and roughly coplanar but did not appear to form clear hydrogen bonds (Figure 5B). Results for the fraction of bases looped-in obtained from this annotation closely resemble those from FR3D (data not shown). Comparable to the results of Figure 5A, the fraction of structures with two non-canonical or near non-canonical pairs increases to 97% when we only consider structures without tertiary interactions. Of the four outliers where we do not observe maximal pairing and which do not have tertiary interactions, two show maximal pairing in other structures reported of the same junction. Of the remaining two, one is located at the surface of the ribosome and we believe, given the level of disorder of the junction nucleotides, may be experiencing the influence of tertiary interactions not reported in the coordinate file, and the other is bound in close proximity by a protein and under considerable tensile stress. We also note that when removing junctions with tertiary interactions only three $H_3S_4H_3S_2$ junctions remain, one of which is a candidate for having unreported tertiary interactions, making it impossible to draw conclusive findings for this particular motif.

Internal loop topologically allowed space. The results from our structural survey motivate the assumption that the junction residues will fully loop in and maximize non-canonical base pairs. Thus, the topologically allowed space for a S_XS_Y motif is identical to that for a $S_{X-Y}S_0$ motif shifted by $Y*-17°$ in α_h and γ_h to account for the additional twist of Y non-canonical pairs. In the case of symmetric internal loops, S_XS_X motifs reduce to S_0S_0 motifs and are subject to the same constraints on α_h β_h γ_h that we observed for continuous helices in the PDB (Supplementary Figure S5). With the correction for the additional twist of $-17°$ in both α_h and γ_h per X non-canonical base pairs, the S_XS_X allowed space contains all α_h β_h γ_h satisfying $-25 \leq \beta_h \leq 35$ and $-20-34*X \leq (\alpha_h + \gamma_h) \leq 20-34*X$ on a
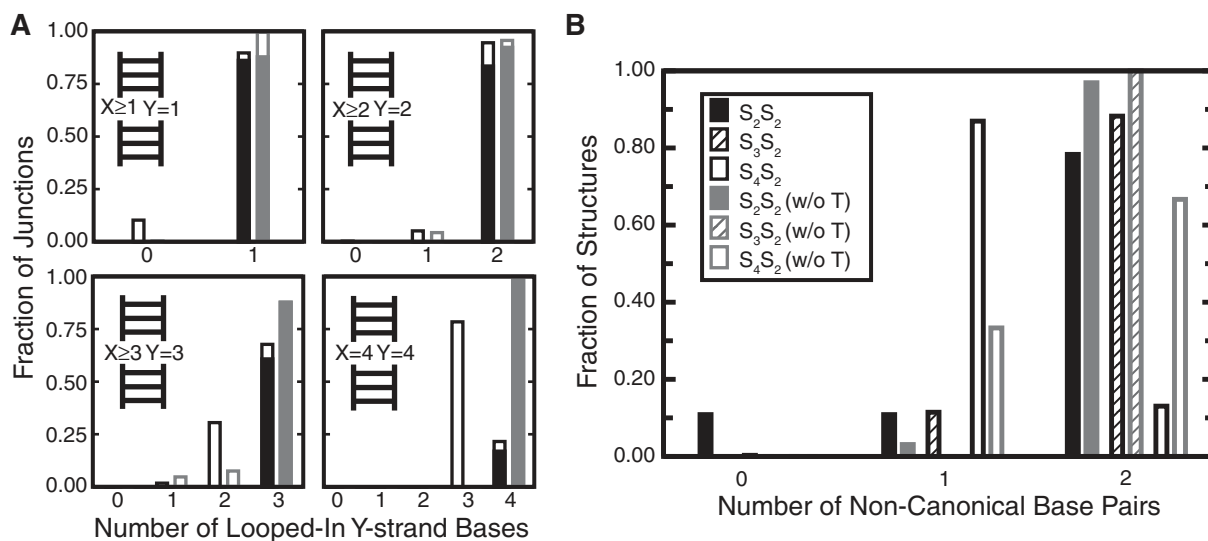
Figure 5. (A) Fraction of looped-in Y strand bases determined from FR3D (30) analysis of internal loops for $H_3S_{X \geq 1}H_sS_{Y=1}$, $H_3S_{X \geq 2}H_sS_{Y=2}$, $H_3S_{X \geq 1}H_sS_{Y=3}$, and $H_3S_{X=4}H_sS_{Y=4}$ junctions. Black bars represent all structures within our database for which FR3D analysis files were available and gray represent the subset of these structures identified as not having tertiary interactions. The degree to which the rightmost bars (fully looped-in conformations) are filled represents the fraction of those structures that have at least Y number of looped-in bases on the X strand. Results represent 474 [135] S_1S_1, 468 [97] S_2S_1, 194 [1] S_3S_1, 168 [6] S_4S_1, 315 [103] S_2S_2, 341 [77] S_3S_2, 23 [4] S_4S_2, 631 [74] S_3S_3, 165 [32] S_4S_3 and 153 [5] S_4S_4 motifs where [] denotes the number of structures without tertiary interactions. (**B**) The number of non-canonical and near non-canonical base pairs observed in all $H_3S_{X \geq 2}H_sS_2$ crystal structures obtained through a manual annotation of structures. The fraction of structures with 0, 1 or 2 pairs is plotted for S_2S_2 (solid), S_3S_2 (dashed) and S_4S_2 (open) motifs in black, with the subset of those structures free of tertiary contacts shown in gray. There are 323 [97], 349 [46] and 23 [3] S_2S_2, S_3S_2, S_4S_2 motifs, respectively, where [] denotes the number of motifs without tertiary contacts.

5°-grid, with an additional 5° in each dimension to account for alignment errors and translational degrees of freedom.

As shown in Table 2, this model is able to capture the significant majority of conformations observed within the PDB. When tertiary interactions are excluded, the fraction of PDB conformations sampled is largely 100%, where the exceptions are crystal structures that have crystal packing interactions at the junction, or are very disordered junctions that appear to be missing protein or other coordinates in close proximity to the junction site.

Examination of the tertiary contact free cryo-EM (cEM) internal loop junctions shows that 89% of 75 junctions fall within our predicted region. Of the eight outliers, seven are from a *de novo* cEM-docked model of the ribosome built before a complete crystal structure was available [(1C2W),42], and the other is a cEM-docked homology model of an rRNA segment whose full crystal structure possesses extensive protein interactions [(2GO5),43]. Analysis of 83 tertiary contact-free NMR internal loops revealed that 85% of the NMR structures agree with our predictions, where outliers were on average ~10° from the nearest allowed conformation. This decrease in performance is a result of outlier junctions adopting conformations that are not fully non-canonically or near-non-canonically paired, or, when paired, propagating with the helical twist that is significantly different than −34°. While it is certainly possible that the more flexible RNAs characterized through NMR adopt conformations not in agreement with our predictions, we note that 10 of the 12 outliers, and all of those >10°

Table 2. The fraction of internal loop motifs that fall within the predicted topologically allowed space, for all crystallographic junctions identified with the RNA FRABASE (28), the subset of those junctions without tertiary interactions, and the fraction of the total $\alpha_h \beta_h \gamma_h$ conformations predicted as allowed

Junction type	PDBs sampled (total)	PDBs w/o tert. contacts sampled (total)	$\alpha_h \beta_h \gamma_h$ allowed
S_1S_1	0.99 (482)	1.00 (135)	0.041
S_2S_2	0.79 (323)	0.99 (103)	0.041
S_3S_3	0.94 (641)	1.00 (74)	0.040
S_4S_4	0.97 (155)	0.60 (5)	0.039
S_2S_1	1.00 (482)	1.00 (97)	0.048
S_3S_2	0.99 (349)	0.96 (77)	0.047
S_4S_3	0.26 (165)	0.56 (32)	0.046
S_3S_1	0.95 (198)	1.00 (1)	0.093
S_4S_2	1.00 (23)	1.00 (4)	0.092
S_4S_1	1.00 (170)	1.00 (6)	0.19

outside, are structures solved without the use of residual dipolar coupling (RDC) restraints. As recently shown by Summers and colleagues (44) these structures are particularly reliant on the force field (and any deficiencies found therein) used in the structure refinement. That these outliers may be due to structural inaccuracies is also supported by MolProbity (37) analysis showing that each outlier suffers from steric violations and/or deviations from standard backbone angle stereochemistry around the site of the junction (data not shown). Finally, we

note that analysis of the entire structural ensemble of each outlier showed that, expect for one molecule [(1TLR),45], at least several, and on average ∼40%, of the reported structures fell within the allowed space, demonstrating that experimental constraints are not precluding agreement with our model.

We note that we previously approximated the symmetric loop case using the set of allowed conformations for a S_1S_0 bulge system that was shifted by $-17°$ in α_h and $-17°$ in γ_h per looped base (23). This approximation achieves similar results as those above, but fails to provide a physical model for its success. The reason for this agreement is because of significant overlap between the two allowed distributions—∼50% of the points are the same.

Combined, the results of our structural surveys and the success of our internal loop predicted topologically allowed space strongly argue for a 'reduced' bulge model of internal loops in the absence of external tertiary contacts. This view represents a significant simplification in the link between secondary and tertiary structure and suggests an additional level of constraints that may be incorporated in modeling RNA 3D structure from secondary structure (22). In particular, these constraints could be implemented in the *de novo* prediction of smaller RNAs (46,47) that do not possess long-range tertiary contacts, significantly reducing the conformational space that must be searched and thereby limiting computation time. The topologically allowed space might also be useful as a global orientation restraint in the refinement of NMR structures. Finally, in larger RNAs where tertiary contacts are expected to be present, such constraints could be implemented as a function of the proximity of sequence-distant RNA to the internal loop in question; on close approach, the constraints would be relaxed, allowing greater conformational freedom.

Interplay between tertiary interactions and topologically allowed conformations. Recent studies (24) have put forth the idea that topological constraints can act to preferentially select for the formation certain tertiary contacts over others, thereby providing a mechanism for secondary structure control of RNA folding and 3D structure. Intriguingly, our results suggest that there may be another dimension to the relationship between the secondary structure encoded conformational space and tertiary interactions. An implication of Table 2 is that tertiary interactions can modify the topologically allowed space by looping out junction residues from their default, maximally non-canonically paired conformation. Thus, tertiary interactions can affect the global conformation of RNA by modifying the topological constraints encoded by junctions. Several examples of tertiary contacts significantly modifying the conformation of an internal loop can be found in the literature. Most dramatically, the S_2S_2 motif of the L11 protein binding site in the 23S rRNA shifts from a fully non-canonically paired conformation to one where a base distal in sequence from the junction inserts to the middle of the junction and is stabilized by interactions with L11 (48,49), with similar behavior observed for closely related mutants (50). In the process, the junction is shifted from within the topologically

allowed space to >45° outside. Also from the ribosome, the UAA/GAN internal loop is significantly perturbed from its fully non-canonically paired conformation in the free state when bound by proteins (51–53). Examples of RNA tertiary contacts modifying internal loop conformation can be found in the GAAA receptor of *Tetrahymena* ribozyme P4-P6 domain (45,54), and stem A and stem B of the hairpin ribozyme (55–57). In some cases, downstream structure may also act to stabilize transient deviations from what would otherwise be a fully looped-in internal loop conformation, consistent with behavior of the J5/5a hinge of the *Tetrahymena* ribozyme in its folded and unfolded states (54,58–60). It is well known that internal loops serve as important sites in protein binding and intra-RNA tertiary interactions (61–64), and that their sequence-specific ability to recognize their binding partners plays a crucial role in stabilizing specific RNA 3D folds (62,65,66). However, that these tertiary interactions, once formed, may in fact be also enabling the molecule to sample new, otherwise inaccessible inter-helical conformations could offer a new paradigm to understanding RNA folding.

ACCESSIBILITY

Files containing the predicted topologically allowed spaces presented above may be obtained in plain text file format at: http://hashimi.biop.lsa.umich.edu/sites/default/files/topallowedspace.tgz.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Mathews,D.H., Moss,W.N. and Turner,D.H. (2010) Folding and finding RNA secondary structure. *Cold Spring Harb. Perspect. Biol.*, **2**, a003665.

2. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.

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

4. Chu,V.B. and Herschlag,D. (2008) Unwinding RNA's secrets: advances in the biology, physics, and modeling of complex RNAs. *Curr. Opin. Struct. Biol.*, **18**, 305–314.

5. Pyle,A.M., Fedorova,O. and Waldsich,C. (2007) Folding of group II introns: a model system for large, multidomain RNAs? *Trends Biochem. Sci.*, **32**, 138–145.

6. Shapiro,B.A., Yingling,Y.G., Kasprzak,W. and Bindewald,E. (2007) Bridging the gap in RNA structure prediction. *Curr. Opin. Struct. Biol.*, **17**, 157–165.

7. Cruz,J.A. and Westhof,E. (2009) The dynamic landscapes of RNA architecture. *Cell*, **136**, 604–609.

8. Laing,C. and Schlick,T. (2010) Computational approaches to 3D modeling of RNA. *J. Phys-Condens. Mat.*, **22**, 283101.

9. Montange,R.K. and Batey,R.T. (2008) Riboswitches: emerging themes in RNA structure and function. *Annu. Rev. Biophys.*, **37**, 117–133.

10. Solomatin,S.V., Greenfeld,M., Chu,S. and Herschlag,D. (2010) Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature*, **463**, 681–684.

11. Al-Hashimi,H.M. and Walter,N.G. (2008) RNA dynamics: it is about time. *Curr. Opin. Struct. Biol.*, **18**, 321–329.

12. Riordan,F.A., Bhattacharyya,A., McAteer,S. and Lilley,D.M. (1992) Kinking of RNA helices by bulged bases, and the structure of the human immunodeficiency virus transactivator response element. *J. Mol. Biol.*, **226**, 305–310.

13. Bhattacharyya,A. and Lilley,D.M. (1989) The contrasting structures of mismatched DNA sequences containing looped-out bases (bulges) and multiple mismatches (bubbles). *Nucleic Acids Res.*, **17**, 6821–6840.

14. Bhattacharyya,A., Murchie,A.I. and Lilley,D.M. (1990) RNA bulges and the helical periodicity of double-stranded RNA. *Nature*, **343**, 484–487.

15. Hsieh,C.H. and Griffith,J.D. (1989) Deletions of bases in one strand of duplex DNA, in contrast to single-base mismatches, produce highly kinked molecules: possible relevance to the folding of single-stranded nucleic acids. *Proc. Natl Acad. Sci. USA*, **86**, 4833–4837.

16. Tang,R.S. and Draper,D.E. (1990) Bulge loops used to measure the helical twist of RNA in solution. *Biochemistry*, **29**, 5232–5237.

17. Zacharias,M. and Hagerman,P.J. (1996) The influence of symmetric internal loops on the flexibility of RNA. *J. Mol. Biol.*, **257**, 276–289.

18. Zacharias,M. and Hagerman,P.J. (1995) Bulge-induced bends in RNA – quantification by transient electric birefringence. *J. Mol. Biol.*, **247**, 486–500.

19. Lescoute,A. and Westhof,E. (2006) Topology of three-way junctions in folded RNAs. *RNA*, **12**, 83–93.

20. Laing,C. and Schlick,T. (2009) Analysis of four-way junctions in RNA structures. *J. Mol. Biol.*, **390**, 547–559.

21. Laing,C., Jung,S., Iqbal,A. and Schlick,T. (2009) Tertiary motifs revealed in analyses of higher-order RNA junctions. *J. Mol. Biol.*, **393**, 67–82.

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

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

24. Chu,V.B., Lipfert,J., Bai,Y., Pande,V.S., Doniach,S. and Herschlag,D. (2009) Do conformational biases of simple helical junctions influence RNA folding stability and specificity? *RNA*, **15**, 2195–2205.

25. Zhang,Q., Stelzer,A.C., Fisher,C.K. and Al-Hashimi,H.M. (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature*, **450**, 1263–1267.

26. Lilley,D.M.J., Clegg,R.M., Diekmann,S., Seeman,N.C., Vonkitzing,E. and Hagerman,P.J. (1995) A nomenclature of junctions and branchpoints in nucleic-acids. *Nucleic Acids Res.*, **23**, 3363–3364.

27. Bailor,M.H., Mustoe,A.M., Brooks,C.L. III and Al-Hashimi,H.M. (2011) 3D maps of RNA inter-helical junctions. *Nat. Protoc.* (in press).

28. Popenda,M., Blazewicz,M., Szachniuk,M. and Adamiak,R.W. (2008) RNA FRABASE version 1.0: an engine with a database to search for the three-dimensional fragments within RNA structures. *Nucleic Acids Res.*, **36**, D386–D391.

29. Bailor,M.H., Musselman,C., Hansen,A.L., Gulati,K., Patel,D.J. and Al-Hashimi,H.M. (2007) Characterizing the relative orientation and dynamics of RNA A-form helices using NMR residual dipolar couplings. *Nat. Protoc.*, **2**, 1536–1546.

30. Sarver,M., Zirbel,C.L., Stombaugh,J., Mokdad,A. and Leontis,N.B. (2008) FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J. Math. Biol.*, **56**, 215–252.

31. Tang,R.S. and Draper,D.E. (1994) On the use of phasing experiments to measure helical repeat and bulge loop-associated twist in RNA. *Nucleic Acids Res.*, **22**, 835–841.

32. Lilley,D.M.J. (1995) Kinking of DNA and RNA by base bulges. *Proc. Natl Acad. Sci. USA*, **92**, 7140–7142.

33. Bai,Y., Chu,V.B., Lipfert,J., Pande,V.S., Herschlag,D. and Doniach,S. (2008) Critical assessment of nucleic acid electrostatics via experimental and computational investigation of an unfolded state ensemble. *J. Am. Chem. Soc.*, **130**, 12334–12341.

34. Gohlke,C., Murchie,A.I.H., Lilley,D.M.J. and Clegg,R.M. (1994) Kinking of DNA and RNA helices by bulged nucleotides observed by fluorescence resonance energy-transfer. *Proc. Natl Acad. Sci. USA*, **91**, 11660–11664.

35. Zacharias,M. and Hagerman,P.J. (1995) The bend in RNA created by the trans-activation response element bulge of human immunodeficiency virus is straightened by arginine and by Tat-derived peptide. *Proc. Natl Acad. Sci. USA*, **92**, 6052–6056.

36. Diener,J.L. and Moore,P.B. (1998) Solution structure of a substrate for the archaeal pre-tRNA splicing endonucleases: the bulge–helix–bulge motif. *Mol. Cell*, **1**, 883–894.

37. Davis,I.W., Leaver-Fay,A., Chen,V.B., Block,J.N., Kapral,G.J., Wang,X., Murray,L.W., Arendall,W.B. III, Snoeyink,J., Richardson,J.S. *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375–W383.

38. Olsen,G.L., Bardaro,M.F. Jr, Echodu,D.C., Drobny,G.P. and Varani,G. (2010) Intermediate rate atomic trajectories of RNA by solid-state NMR spectroscopy. *J. Am. Chem. Soc.*, **132**, 303–308.

39. Frank,A.T., Stelzer,A.C., Al-Hashimi,H.M. and Andricioaei,I. (2009) Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: new insights into RNA dynamics and adaptive ligand recognition. *Nucleic Acids Res.*, **37**, 3670–3679.

40. Stuhmeier,F., Hillisch,A., Clegg,R.M. and Diekman,S. (2000) Fluorescence energy transfer analysis of DNA structures containing several bulges and their interaction with CAP. *J. Mol. Biol.*, **302**, 1081–1100.

41. Wozniak,A.K., Schroder,G.F., Grubmuller,H., Seidel,C.A. and Oesterhelt,F. (2008) Single-molecule FRET measures bends and kinks in DNA. *Proc. Natl Acad. Sci. USA*, **105**, 18337–18342.

42. Mueller,F., Sommer,I., Baranov,P., Matadeen,R., Stoldt,M., Wohnert,J., Gorlach,M., van Heel,M. and Brimacombe,R. (2000) The 3D arrangement of the 23 S and 5 S rRNA in the *Escherichia coli* 50 S ribosomal subunit based on a cryo-electron microscopic reconstruction at 7.5 A resolution. *J. Mol. Biol.*, **298**, 35–59.

43. Halic,M., Gartmann,M., Schlenker,O., Mielke,T., Pool,M.R., Sinning,I. and Beckmann,R. (2006) Signal recognition particle receptor exposes the ribosomal translocon binding site. *Science*, **312**, 745–747.

44. Tolbert,B.S., Miyazaki,Y., Barton,S., Kinde,B., Starck,P., Singh,R., Bax,A., Case,D.A. and Summers,M.F. (2010) Major

groove width variations in RNA structures determined by NMR and impact of 13C residual chemical shift anisotropy and 1H-13C residual dipolar coupling on refinement. *J. Biomol. NMR*, **47**, 205–219.

45. Butcher,S.E., Dieckmann,T. and Feigon,J. (1997) Solution structure of a GAAA tetraloop receptor RNA. *EMBO J.*, **16**, 7490–7499.

46. Parisien,M. and Major,F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51–55.

47. Das,R., Karanicolas,J. and Baker,D. (2010) Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods*, **7**, 291–294.

48. Conn,G.L., Draper,D.E., Lattman,E.E. and Gittis,A.G. (1999) Crystal structure of a conserved ribosomal protein–RNA complex. *Science*, **284**, 1171–1174.

49. Wang,Y.X., Huang,S. and Draper,D.E. (1996) Structure of a U.U pair within a conserved ribosomal RNA hairpin. *Nucleic Acids Res.*, **24**, 2666–2672.

50. Shankar,N., Xia,T., Kennedy,S.D., Krugh,T.R., Mathews,D.H. and Turner,D.H. (2007) NMR reveals the absence of hydrogen bonding in adjacent UU and AG mismatches in an isolated internal loop from ribosomal RNA. *Biochemistry*, **46**, 12665–12678.

51. Shankar,N., Kennedy,S.D., Chen,G., Krugh,T.R. and Turner,D.H. (2006) The NMR structure of an internal loop from 23S ribosomal RNA differs from its structure in crystals of 50s ribosomal subunits. *Biochemistry*, **45**, 11776–11789.

52. Lee,J.C., Gutell,R.R. and Russell,R. (2006) The UAA/GAN internal loop motif: a new RNA structural element that forms a cross-strand AAA stack and long-range tertiary interactions. *J. Mol. Biol.*, **360**, 978–988.

53. Reblova,K., Strelcova,Z., Kulhanek,P., Besseova,I., Mathews,D.H., Nostrand,K.V., Yildirim,I., Turner,D.H. and Sponer,J. (2010) An RNA molecular switch: intrinsic flexibility of 23S rRNA Helices 40 and 68 5'-UAA/5'-GAN internal loops studied by molecular dynamics methods. *J. Chem. Theory Comput.*, **2010**, 910–929.

54. Cate,J.H., Gooding,A.R., Podell,E., Zhou,K., Golden,B.L., Kundrot,C.E., Cech,T.R. and Doudna,J.A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678–1685.

55. Cai,Z. and Tinoco,I. Jr (1996) Solution structure of loop A from the hairpin ribozyme from tobacco ringspot virus satellite. *Biochemistry*, **35**, 6026–6036.

56. Butcher,S.E., Allain,F.H. and Feigon,J. (1999) Solution structure of the loop B domain from the hairpin ribozyme. *Nat. Struct. Biol.*, **6**, 212–216.

57. Rupert,P.B. and Ferre-D'Amare,A.R. (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature*, **410**, 780–786.

58. Schlatterer,J.C., Kwok,L.W., Lamb,J.S., Park,H.Y., Andresen,K., Brenowitz,M. and Pollack,L. (2008) Hinge stiffness is a barrier to RNA folding. *J. Mol. Biol.*, **379**, 859–870.

59. Takamoto,K., Das,R., He,Q., Doniach,S., Brenowitz,M., Herschlag,D. and Chance,M.R. (2004) Principles of RNA compaction: insights from the equilibrium folding pathway of the P4-P6 RNA domain in monovalent cations. *J. Mol. Biol.*, **343**, 1195–1206.

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

61. Nissen,P., Ippolito,J.A., Ban,N., Moore,P.B. and Steitz,T.A. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl Acad. Sci. USA*, **98**, 4899–4903.

62. Costa,M. and Michel,F. (1995) Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.*, **14**, 1276–1285.

63. Draper,D.E. (1995) Protein–RNA recognition. *Annu. Rev. Biochem.*, **64**, 593–620.

64. Cate,J.H., Gooding,A.R., Podell,E., Zhou,K., Golden,B.L., Szewczak,A.A., Kundrot,C.E., Cech,T.R. and Doudna,J.A. (1996) RNA tertiary structure mediation by adenosine platforms. *Science*, **273**, 1696–1699.

65. Tan,E., Wilson,T.J., Nahas,M.K., Clegg,R.M., Lilley,D.M. and Ha,T. (2003) A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate. *Proc. Natl Acad. Sci. USA*, **100**, 9308–9313.

66. Geary,C., Chworos,A. and Jaeger,L. (2011) Promoting RNA helical stacking via A-minor junctions. *Nucleic Acids Res.*, **39**, 1066–1080.

67. Yang,H., Jossinet,F., Leontis,N., Chen,L., Westbrook,J., Berman,H. and Westhof,E. (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.*, **31**, 3450–3460.