

Topological constraints of RNA pseudoknotted and loop-kissing motifs: applications to three-dimensional structure prediction

Xiaojun Xu ^{1,*} and Shi-Jie Chen ^{2,*}

¹Institute of Bioinformatics and Medical Engineering, Jiangsu University of Technology, Changzhou, Jiangsu 213001, China and ²Department of Physics, Department of Biochemistry, and Informatics Institute, University of Missouri, Columbia, MO 65211, USA

Received April 04, 2020; Editorial Decision May 14, 2020; Accepted May 19, 2020

ABSTRACT

An RNA global fold can be described at the level of helix orientations and relatively flexible loop conformations that connect the helices. The linkage between the helices plays an essential role in determining the structural topology, which restricts RNA local and global folds, especially for RNA tertiary structures involving cross-linked base pairs. We quantitatively analyze the topological constraints on RNA 3D conformational space, in particular, on the distribution of helix orientations, for pseudoknots and loop-loop kissing structures. The result shows that a viable conformational space is predominantly determined by the motif type, helix size, and loop size, indicating a strong topological coupling between helices and loops in RNA tertiary motifs. Moreover, the analysis indicates that (cross-linked) tertiary contacts can cause much stronger topological constraints on RNA global fold than non-cross-linked base pairs. Furthermore, based on the topological constraints encoded in the 2D structure and the 3D templates, we develop a 3D structure prediction approach. This approach can be further combined with structure probing methods to expand the capability of computational prediction for large RNA folds.

INTRODUCTION

Most RNAs fold in a hierarchical pathway, with the folding of secondary structures typically preceding the formation of tertiary interactions (1–3). Chain (bond) connectivity, excluded volume, and the linkage between the different helices, play an essential role in shaping the RNA conformational space and the global topology of the native structure (4–16). Different RNA 2D structural motifs show different linkages between helices. Understanding the degree

to which RNA structural topology restricts RNA 3D conformational space (i.e. topological constraints) can improve the accuracy of structure prediction as well as our understanding for RNA folding principles.

Compared with the protein backbone, RNA contains more rotatable bonds per residue, contributing to the substantial conformational flexibility. Muthy et al. performed a grid search for all the potential conformers for a dinucleotide, and found that hard sphere steric exclusion and bond connectivity can restrict torsion angles in nucleic acids to <5% of all the possible conformations (4). For the helix-junction-helix (HJH) motif, Chu et al. studied two helices joined by flexible single- or double-stranded polyethylene glycol tethers (5). Stochastic dynamics simulations and small-angle X-ray scattering experiments showed that the HJH junction topology can further significantly reduce the conformational space and influence the preferred location and orientation of the adjoining helices. Conformational analysis based on TOPRNA (6) and MC-sym (7,8) showed that RNA global conformation is largely defined by topological constraints of RNA secondary structure while the electrostatics, intra- and inter-loop and other interactions select specific conformations from the accessible conformational ensemble.

Different approaches have been developed to investigate the impact of RNA 2D structural constraint on 3D conformations. For example, guided by the knowledge-based statistical potentials for bending and torsional degrees of freedom of the internal loop and radius of gyration, graph theory-based tool (RAG) (14,15) can efficiently sample the global helical topologies (as represented by graphs) in 3D space. Erwin (16), a coarse-grained helix-centered model as another example, explores global arrangements of helices and loops within RNA structures. Combined with a novel energy function for the positioning of stems and loops, the model can predict RNA 3D structures.

RNA 3D structure prediction (17–39), on the other hand, has been developed for decades to help in understanding

*To whom correspondence should be addressed. Tel: +1 573 882 3335; Email: chenshi@missouri.edu
Correspondence may also be addressed to Xiaojun Xu. Email: xuxiaojun@jsut.edu.cn

RNA structure-function relationships, as well as to promote the rational design of RNA-based molecular systems. The prediction methods result from various coarse-grained levels (23–30), such as iFoldRNA (27), SimRNA (28,29), and IsRNA (30), have the ability to predict RNA structures and folding kinetics, but the accuracy of the predictions can be limited by the large conformational space for some RNAs. Template-assembly approaches build RNA 3D structures based on the known structural modules ranging from piecewise fragments to whole structural motifs (8,31–39). For example, FARNA/FARFAR (34,35) predicts RNA structures through the assembly of 1–3 nucleotides from known structures via a Monte Carlo procedure. It can reach atomic resolution for most short RNAs (<30 nt). MC-sym (8), RNA-composer (36) and Vfold3D (37,38), furthermore, take the advantage of strong topological constraints of secondary structural (bulge, internal loops and multi-branched junctions, shown in Supplementary Figure S1) motifs, and select templates from known structures with homologous sequence information to predict RNA 3D structures.

RNA structures often involve cross-linked tertiary base pairs (as indicated in the circular representations in Supplementary Figure S1), such as pseudoknots (PK) and hairpin-hairpin loop kissing (KISS) motifs. Most motif template-based RNA structure prediction methods use the motif type and loop/helix size-dependent scoring scheme to search for appropriate templates to build structures. However, the success rate to find a proper template is usually low for large RNAs with tertiary contacts. A possible approach to alleviate the problem of limited availability of templates is to search for the topology-based templates, i.e. templates with the same motif type but may have different sizes and sequences for the loops and helices. Here, we present a comprehensive analysis for the topological constraints of four types of RNA structural motifs with tertiary (cross-linked) contacts. Specifically, we investigate the helix and loop size dependence of the conformational space, particularly the topological constraints induced by the cross-linked base pairs in the PK and KISS motifs. We use the fractional change in the number of viable conformations to quantify the topological constraint. Compared with the RNA motifs without cross-linked base pairs, the tertiary contacts impose much stronger topological constraints on RNA global folds. Therefore, the use of cross-linked contacts can significantly improve the conformational sampling quality for 3D structure prediction. Furthermore, we here also develop and illustrate a multi-stage method, which incorporates Vfold3D, VfoldLA (39) and tertiary topologies, in RNA 3D structure prediction. By taking the advantage of the strong topological constraints imposed by the tertiary contacts, as shown in the benchmark tests, our multi-stage model can provide improved predictions for RNA 3D structures with cross-linked base pairs.

ALGORITHM AND METHODS

Transformation matrix for helix orientation

The different configurations for a pair of helices can be generated through the translation-rotation transformation between the helices. Mathematically, the transformation of a

rigid body can be described by a 4×4 matrix for the translation and rotation, as shown in Figure 1(A). Vector (dx, dy, dz) denotes the translational displacement between the helices, and the nine-element 3×3 submatrix \mathbf{R} represents the rotation with three independent parameters: polar angle θ , azimuthal angle φ for the rotation axis, and rotation angle α , as shown in Figure 1(B). Therefore, the spatial configuration of a two-helix system can be described and generated by six parameters $(\theta, \varphi, \alpha, dx, dy, dz)$. Correspondingly, the parameters for any given helix configuration can be uniquely extracted from the relative displacement and orientation between the first base pairs of the helices (see the red oval in Figure 1B).

Vfold for allowed/disallowed configurations

To quantify the degree to which structural topology restricts the sampling of 3D conformations, we use the coarse-grained representation of RNA structures (Vfold model) (40) and sample the coarse-grained loop structures with given motif type, helix sizes and loop sizes.

- For a given helix–loop–helix (HLH) 2D motif, to determine the minimal bounding box (MBB) for the 3D conformations, we use Vfold to sample the coarse-grained 3D conformations; See the P–C4′–P virtual-bond loop structures in Figure 1C. We choose the geometric center of the terminal base pair in one of the helices as the center of the bounding box. In the conformational sampling, the (geometric center of) terminal base pair in the other helix sweeps out a 3D region. The minimum box that contains such a region defines the MBB. See the Supplementary Figure S2 for detailed descriptions about the Vfold model and the MBB calculations. By definition of the MBB, we only need to sample HLH configurations within the minimal bounding box.

In the MBB calculations shown in the Supplementary Figure S2, we sample the orientation of the rotation axis (θ, φ) with 100 quasi-uniformly distributed points on the surface of a sphere (see Supplementary Figure S3). We use 18 uniformly spaced rotation angles α between 0 and 2π to rotate the helices for each given rotation axis, which leads to $R = 1800$ total rotations. To achieve the optimal balance between the computational efficiency and the sampling quality, we use loop-size dependent adaptive grid size $((L + 1) \text{ \AA})$ for an L -nt loop) in MBB for translational displacement of rotated helices. Specifically, the total grids in MBB of HLH with $L = 0, 1, 2,$ and 3 nt are $T = 23548, 8820, 4913$ and 3840 , respectively. The total possible helical orientations is equal to $N_{\text{tot}}^{\text{MBB}} = R \times T$. For example, $N_{\text{tot}}^{\text{MBB}} = 1800 \times 8820$ for an HLH motif with $L = 1$ nt (grid size of 2 \AA in MBB).

- We generate the possible helix configurations based on $(\theta, \varphi, \alpha, dx, dy, dz)$. For the PK motif with two helices and three loops, denoted by the size of H1–L1–L2–L3–H2 (see Supplementary Figure S4), we use the HLH motif of H1–L2–H2 to exhaustively sample the Vfold-allowed two-helix configurations within its MBB. $N_{\text{tot}}^{\text{MBB}}$ of H1–L2–H2 gives the total number of the possible two-helix configurations described by $(\theta, \varphi, \alpha, dx, dy, dz)$ for a PK motif.

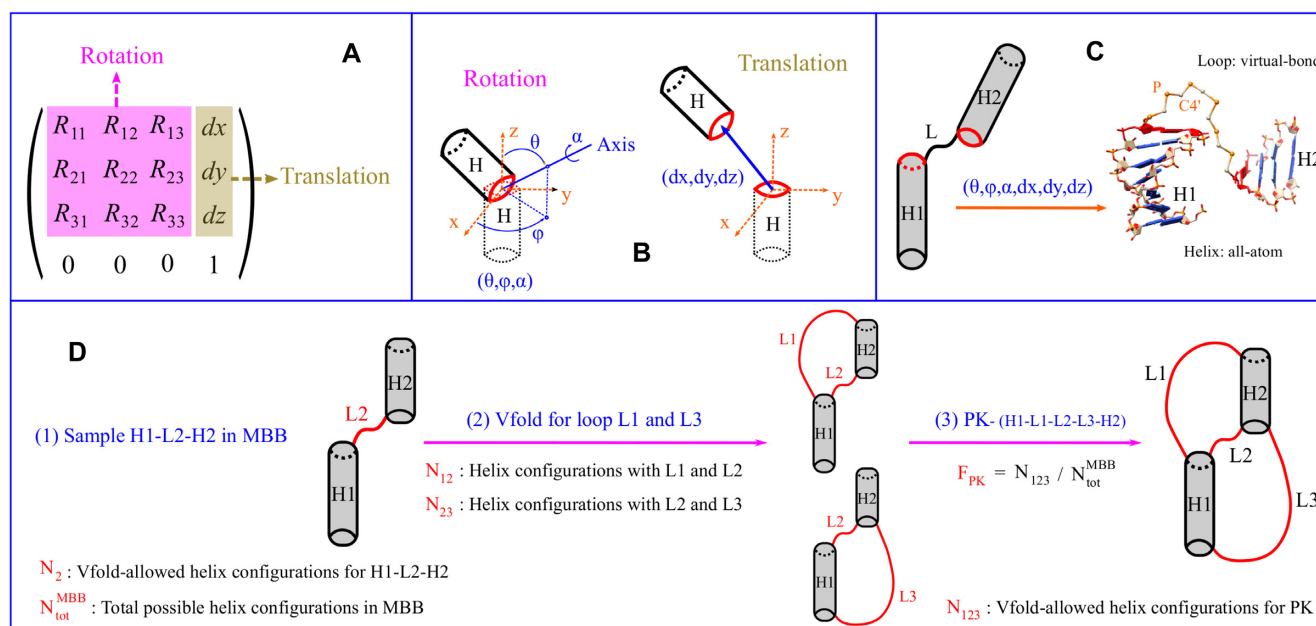


Figure 1. (A) Transformation matrix for the rigid-body translational and rotational transformations. (B) The nine-element rotation matrix in (A) is characterized by three independent parameters for the rotation axis (θ, φ) passing through the geometric center of the first base pair (red oval), with the rotation angle α . Vector (dx, dy, dz) denotes the translational displacement between the helices. (C) The definition of an HLH motif, and the sampling of the 3D helix configurations defined by $(\theta, \varphi, \alpha, dx, dy, dz)$ with Vfold. Helices are modeled with all atom A-form structures. Loops are generated by the Vfold (P-C4'-P) backbone structures. (D) Topological constraints for PK, which contains two helices and three loops. The HLH of H1-L2-H2 is used to sample the helix configurations in MBB. Loops are modeled by Vfold to determine the allowed and disallowed helix configurations. N_x is the number of allowed helix configurations with the respective loop connections. F_{PK} is the fraction of Vfold-allowed helix configurations of a PK motif.

The KISS motif, as illustrated in Supplementary Figure S4, contains three helices and six loops, denoted by the size of (H1-H2-H3)-(L1-L2-L3-L4-L5-L6). In order to sample the conformational ensemble with viable computational feasibility, we randomly sample the three-helix configurations with the HLHs of H1-L2-H2, and H2-L4-H3 in their corresponding MBBs. As shown in Supplementary Figure S5, N_{tot}^{24} gives the total randomly sampled (three-helix) helical configurations of a KISS motif, characterized by two sets of $(\theta, \varphi, \alpha, dx, dy, dz)$.

- For each given helix configuration, we generate Vfold loop conformations with the consideration of helix-loop excluded volume interactions. If at least one viable conformation can be realized for the loops, the given helix configuration is considered as being acceptable (allowed), otherwise, it is considered topologically unacceptable (disallowed). As shown in Figure 1D and Supplementary Figure S5, we take count of the allowed helix configurations for the two-helix (PK) and three-helix (KISS) systems with the different loop linkages. Our purpose is to assess the impact of topological constraints. We use N_x to denote the number of the Vfold-allowed helical configurations of the given helix system. For example, N_{124} in Supplementary Figure S5 gives the number of the viable helical configurations of the three helices connected by the three viable loops of L1, L2 and L4. Therefore, the decrease of the number of the viable helical configurations from N_{124} to N_{1245} indicates the impact of the topological constraints from loop L5 in the presence of loops L1, L2 and L4.

- To calculate the fraction of forming viable helical configurations. From the ratio between the numbers of the allowed and the total possible helical conformations, we evaluate the degree to which RNA structural topology restricts the conformations for the different motif types, helix sizes and loop sizes. We use the (topological constraint) fraction $F_x = N_x / N_{\text{tot}}$ ($N_{\text{tot}}^{\text{MBB}}$ for PK and N_{tot}^{24} for KISS) to measure the topological constraints of the helix system with the respective loop connections. Therefore, as shown in Figure 1D and Supplementary Figure S5, $F_{\text{PK}} (= N_{123} / N_{\text{tot}}^{\text{MBB}})$ and $F_{\text{KISS}} (= N_{123456} / N_{\text{tot}}^{24})$ gives the topological constraints of PK and KISS motifs, respectively.

Topology-based 3D structure prediction

To take the advantage of structural topological constraints, in particular, from tertiary contacts, we develop a multi-stage RNA 3D structure prediction method. Specifically, as shown in Figure 2, we integrate Vfold3D, VfoldLA and tertiary topologies to predict RNA structures containing cross-linked base pairs.

- To classify motifs with given 2D structure. As shown in Supplementary Figure S1, we define three types of motifs: (a) Tertiary structural motifs, such as PK, and KISS; (b) Secondary structural motifs, such as junctions, internal and bulge loops; (c) Single-stranded loops, such as hairpin, helix2 and tails. For the purpose to determine the irreducible tertiary ‘topology’ of a complicated structure, we replace the isolated substructures, such as helices,

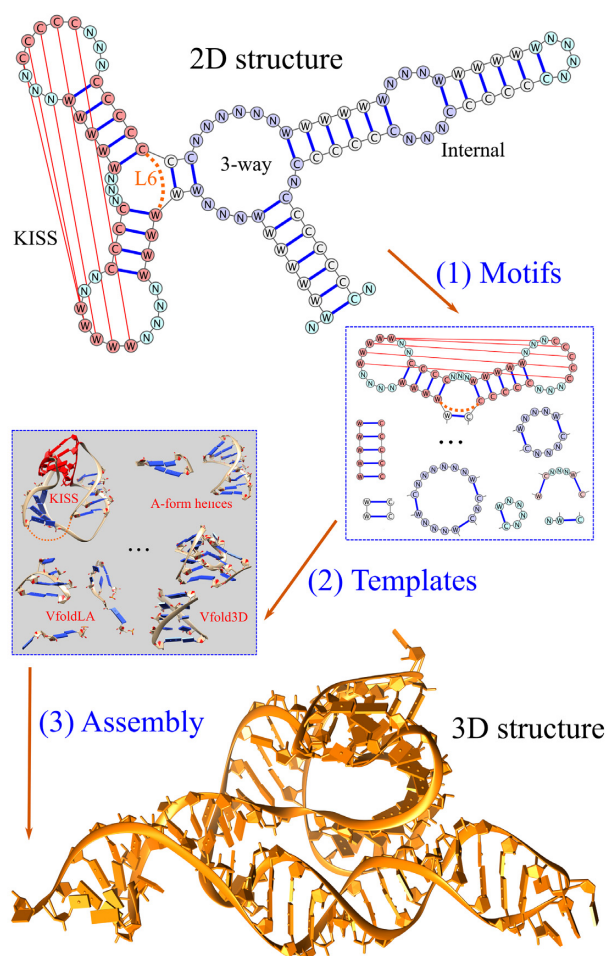


Figure 2. The multi-stage approach, which incorporates Vfold3D, VfoldLA and tertiary topologies, for RNA 3D structure prediction. The method involves three steps: (1) extract motifs from a given 2D structure; (2) select templates from known RNA structures; (3) generate all-atom models to predict the 3D structures.

stem-loop, and junctions, with the terminal base pairs of the substructures. For example, the given 2D structure in Figure 2 is broken down into the structural motifs of one KISS, one three-way junction, one internal loop, and several helices and single-stranded loops. The size of the KISS motif is (4–5–5)-(4–2–3–3–3–2), with the size of 2-nt (one terminal base pair) for loop L6.

- To build helices with all-atom A-form helical structures.
- To construct 3D structures for tertiary structural motifs. We select the optimal templates from the structural database for the same motif type. It should be noted that the current template database for the PK and KISS motifs is quite limited, compared with the database of the secondary structural motifs. If no templates with the exact helix and loop sizes are found, We choose the templates with the similar helix and loop sizes. We select the optimal templates according to the size-dependent scoring scheme, $S^{\text{PK/KISS}} = a \sum_{\text{helix}} \Delta H + b \sum_{\text{loop}} \Delta L$, where ΔH and ΔL are the size differences between the corresponding helix and loop, respectively. Here the size of a helix is defined as the number of base pairs. We set the

prefactor $a = 2$ for the (helix) ΔH -contribution. Since the loops L2 for PK and L2 and L4 for KISS are usually shorter (<4 nts) with stronger topological constraints than other loops. We set $b = 5$ for these loops and 1 for others. For example, the KISS motif extracted from the given 2D structure in Figure 2 is (4–5–5)-(4–2–3–3–3–2). The score for the query template with the size of (5–5–4)-(3–1–4–4–5–2) is 18.

- To build all-atom structures for secondary structural motifs using Vfold3D (37,38). Vfold3D predicts RNA 3D structures through the assembly of the 3D templates for secondary structural motifs (bulge, internal and multi-branched loops, shown in Supplementary Figure S1) and A-form helices. The criteria for template searching in Vfold3D is motif type, loop size and sequence dependent. If no templates can be found for the given secondary structural motifs, we break the motifs into single-stranded loops.
- To build all-atom structures for single-stranded loops using VfoldLA (39). VfoldLA builds RNA 3D structures through the assembly of 3D templates for single-stranded loops (hairpin, helix2 and tails, shown in Supplementary Figure S1) and A-form helices. The criteria for the template search in VfoldLA is loop type, size, and sequence dependent.
- To assemble structures from helices, motifs, and loops to build all-atom structures of the whole RNA, and refine the assembled structures using energy minimization, such as AMBER and NAMD, to remove the steric clash. Manual local structure adjustment may be applied to disentangle the strands (for loop strand penetrating helices) as a pre-simulation treatment. Here, we use the 5000-step NAMD energy minimization to refine the predicted all-atom structures, embedded in a water box, neutralized by Na^+ .

RESULTS AND DISCUSSION

We use the 20 PK motifs and 7 KISS motifs listed in Supplementary Tables S1 and S2 to elucidate the topological constraints induced by the cross-linked base pairs, as compared with the constraints encoded in secondary structural motifs, and its potential applications to RNA 3D structure prediction.

Pseudoknotted motifs

As shown in Supplementary Figure S4(A), the 20 pseudoknotted structures with different helix and loop sizes (H1–L1–L2–L3–H2, as listed in Table 1) show similar global folds, with the two helices (nearly) coaxially stacked. We use the Vfold coarse-grained conformational model to sample all the two-helix configurations using the discrete transformation matrix of $(\theta, \varphi, \alpha, dx, dy, dz)$.

As shown in Figure 3A and B and Supplementary Figures S6–S10, the pseudoknot topology imposes significant constraints on otherwise free helix orientations. For example, the 2tpk pseudoknot (5–1–0–7–7) in Figure 3(A) has ($N_2 =$) 181 249 viable Vfold-allowed helical configurations of H1–L2–H2 in its MBB. The topological constraints of H1–L2–H2 (from loop L2) can be quantified, as listed in Table 1,

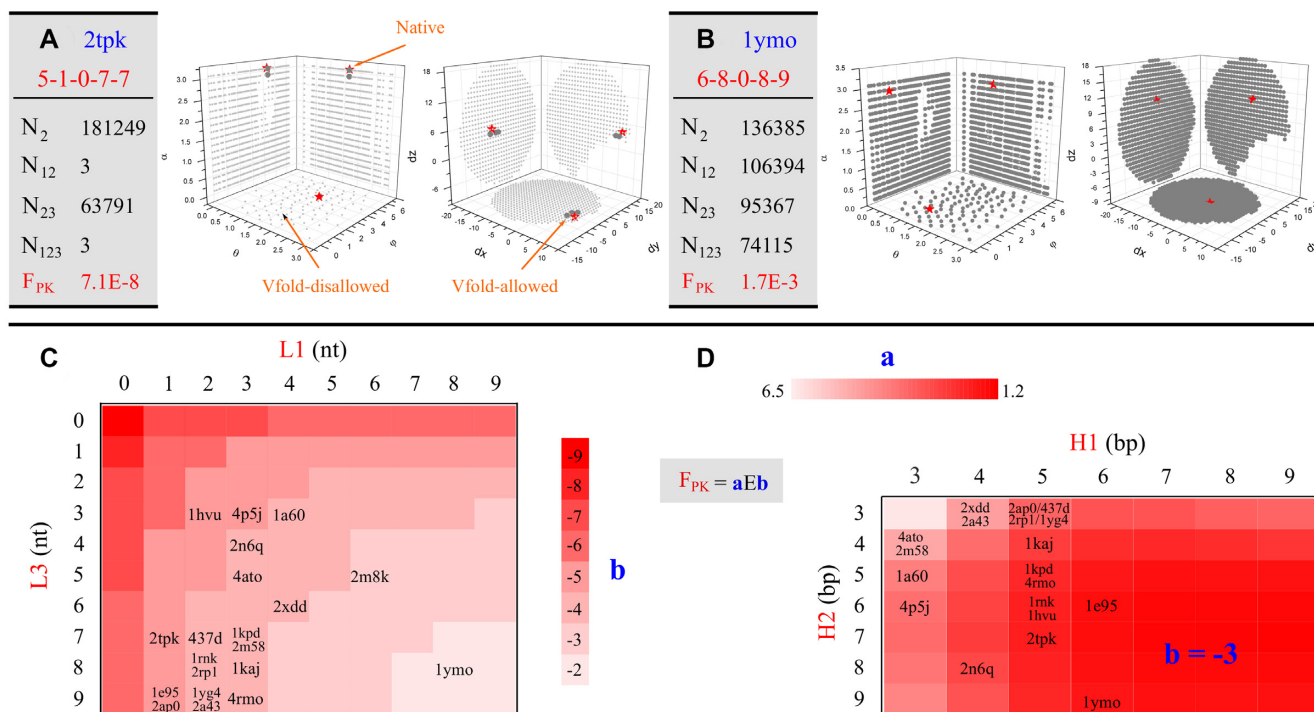


Figure 3. The quantification of the conformational sampling and the projections of the allowed (large dark gray) and disallowed (small light gray) helix configurations on six planes (three for translation, and three for rotation) for (A) 2tpk, and (B) lymo, respectively. The red stars are from the corresponding native structures. N_x is the number of the Vfold-allowed helix configurations with the respective loop connections. F_{PK} is the fraction of the Vfold-allowed helix configurations of a PK motif. (C) Loop L1 and L3 dependence of the topological constraints for PK, $F_{PK}(L1, L3)$, averaged over helices H1 and H2 (from 3 to 9 bp) and loop L2 (from 0 to 3 nt). The color bar is based on the value b of F_{PK} . PDB IDs are placed according to (L1, L3). (D) Helix H1 and H2 dependence of the topological constraints for PK, $F_{PK}(H1, H2)$, averaged over loops L1 and L3 (from 0 to 9 nt) and loop L2 (from 0 to 3 nt). The color bar is based on the value a of F_{PK} (with $b = -3$). PDB IDs are placed according to (H1, H2).

Table 1. Topological constraints of the PK motif, defined by the size of H1–L1–L2–L3–H2. We use loop L2 to exhaustively sample all the possible two-helix configurations in the MBB of H1–L2–H2. F_x is the fraction of the Vfold-allowed helical configurations with the corresponding loops. $F_{PK}(F_{123})$ denotes the total topological constraints of a PK structure

PDB	Size	F_2	F_{12}	F_{23}	F_{PK}
1a60	3–4–0–3–5	5.2E–3	1.8E–4	2.0E–4	4.6E–5
1e95	6–1–0–9–6	3.2E–3	1.8E–6	2.9E–3	1.8E–6
1hvu	5–2–0–3–6	4.3E–3	6.5E–6	3.8E–5	1.2E–7
lymo	6–8–0–8–9	3.2E–3	2.5E–3	2.2E–3	1.7E–3
2n6q	4–3–0–4–8	4.9E–3	3.9E–5	2.8E–4	9.2E–6
2tpk	5–1–0–7–7	4.3E–3	7.1E–8	1.5E–3	7.1E–8
4p5j	3–3–0–3–6	5.2E–3	4.2E–5	2.0E–4	2.1E–5
2m8k	5–6–0–5–12	4.3E–3	5.8E–4	2.4E–4	2.6E–5
2ap0	5–1–1–9–3	2.8E–2	1.0E–4	2.6E–2	1.0E–4
1kpd	5–3–1–7–5	2.7E–2	3.8E–4	9.6E–3	3.1E–4
1rnk	5–2–1–8–6	2.7E–2	6.3E–5	1.7E–2	6.3E–5
1yg4	5–2–1–9–3	2.8E–2	4.6E–4	2.6E–2	4.5E–4
2a43	4–2–1–9–3	3.1E–2	4.6E–4	3.0E–2	4.6E–4
2rp1	5–2–1–8–3	2.8E–2	4.6E–4	2.0E–2	3.8E–4
2xdd	4–4–1–6–3	3.1E–2	3.6E–3	1.0E–2	1.6E–3
4ato	3–3–1–5–4	3.2E–2	8.6E–4	6.9E–3	5.4E–4
437d	5–2–1–7–3	2.8E–2	4.6E–4	1.2E–2	2.3E–4
4rmo	5–3–2–9–5	5.6E–2	8.3E–4	4.2E–2	8.3E–4
1kaj	5–3–2–8–4	5.7E–2	1.6E–3	3.1E–2	1.4E–3
2m58	3–3–3–7–4	9.7E–2	3.3E–3	6.1E–2	3.0E–3
Average		2.5E–2	7.9E–4	1.5E–2	5.6E–4

by the fraction of $F_2 = N_2/N_{\text{tot}}^{\text{MBB}} = 4.3E-3$. Here, the total possible helix configurations in MBB $N_{\text{tot}}^{\text{MBB}} = R \times T$, with

$R = 1800$ for the rotations, and $T = 23\,548$ ($L2 = 0$ nt) of the total grids in MBB for the translations. For each Vfold-allowed helical configurations of H1–L2–H2, we assess the topological constraints with additional loops by counting for the number of viable helical configurations. N_{12} ($= 3$) and N_{23} ($= 63791$) gives the number of viable configurations for the H1–L2–H2 motif with loop L1 and L3, respectively. Since loop L1 ($= 1$ nt) is much shorter than L3 ($= 7$ nt), the topological constraint from L1 (from $4.3E-3$ (of F_2) to $7.1E-8$ (of F_{12})) is much stronger than that from L3 ($F_{23} = 1.5E-3$). Furthermore, N_{123} denotes the total number of viable helical configurations for the PK motif. The total topological constraints on 2tpk is $F_{PK} = N_{123}/N_{\text{tot}}^{\text{MBB}} = 7.1E-8$ with the major contributions from L1 and L2 loops.

For the lymo pseudoknot (6–8–0–8–9) shown in Figure 3(B), because of the relatively large size of loops L1 ($= 8$ nt) and L3 ($= 8$ nt), we find that the total topological constraints is $F_{PK} = 1.7E-3$, with the major contributions from loop L2 ($= 0$ nt) with $F_2 = 3.2E-3$. On the other hand, 2m8k (5–6–0–5–12), similar to lymo, has large size of loops L1 and L3. However, 2m8k involves a much stronger constraint than lymo ($F_{PK} = 2.6E-5$ for 2m8k versus $1.7E-3$ for lymo), suggesting the helix contribution to the overall topological constraints. From Table 1 of the 20 cases with loop L2 ranging from 0 to 3 nt in length, we find that the constraints from the L2 loop, F_2 , of H1–L2–H2 decreases with the length of L2 (the average of F_2 for $L2 = 0, 1, 2$ and 3 nt are $4.3E-3, 2.9E-2, 5.6E-2$ and $9.7E-2$, respectively).

Shorter L2 leads to stronger excluded volume effects between helices. Moreover, the asymmetrical A-form structure of helices may also contribute to the significant loop- and helix-size dependent of the helical configurations of a pseudoknot. The total topological constraints, F_{PK} , ranges from E-3 to E-8 in magnitude, with 5.6E-4 in average, for the 20 cases.

To better reveal the fundamental principles of topological constraints encoded in PK, we compute the fraction F_{PK} for all PKs with H1 and H2 ranging from 3 to 9-bp, L1 and L3 from 0 to 9-nt, and L2 from 0 to 3-nt to calculate the averaged fraction of $F_{PK}(L1, L3)$, as shown in Figure 3(C). The magnitude of F_{PK} ranges from 1E-2 to 1E-9. Smaller L1 and L3 in size would cause stronger topological constraints (decreasing b in F_{PK}). The overall shape of the 'landscape' might be funnel-like, with quite good symmetry with respect to L1 and L3. According to the sizes of 20 native PK structures, as listed in Table 1, we find that (i) most cases have the sizes of (L1, L3) with moderate averaged topological constraints (ranging from 1E-4 to 1E-6); (ii) larger size of L3 (compared with L1) is preferable. As listed in Table 1, the average F_{12} ($= 7.9E-4$) is much smaller than F_{23} ($= 1.5E-2$). Therefore, loops L1 and L2 play the major role in the topological constraints of the wild-type PKs.

Energetically, RNA topological constraint is related to the entropic contribution to RNA stability. Stronger topological constraints with shorter loops lead to a higher entropic folding free energy (and lower stability). At the 2D level, the enthalpic contribution of RNA stability is mainly from the helices (base pairs). Larger size of helices (depending on the sequences) can result in a lower enthalpic free energy (higher stability). Therefore, there is a balance between the sizes of loops and helices in RNA folding. As a compromise, the wild-type native structures in general involve moderate topological constraints. On the other hand, the tertiary interactions (depending on the sequences) at the 3D level, such as the coaxial stacking between H1 and H2, base triples between L3 and H2, and non-canonical interactions within loops, can further stabilize the structures while reducing the corresponding conformational space (strengthening the topological constraints). For instance, the strong topological constraints of 2tpk ($F_{PK} = 7.1E-8$ with very limited number of allowed helix configurations $N_{123} = 3$), as shown in Figure 3(A), may be counteracted by the stability from two helices and rich tertiary interactions.

Moreover, as shown in Figure 3D, we also calculate the averaged topological constraint fraction $F_{PK}(H1, H2)$. Compared with $F_{PK}(L1, L3)$ shown in Figure 3C, the 'landscape' becomes much flattened (all F_{PK} have the same magnitude of 1E-3). Larger H1 and H2 in size result in stronger topological constraints (decreasing a in F_{PK}). However, for a given H1 (H2), there is a trend of decreasing then increasing in a with the elongation of H2 (H1). For example, $6.5 \rightarrow 3.5 \rightarrow 4.2$ for the a value for H2 = 3, 7 and 9 bp, respectively, with H1 = 3 bp. This factor should be related to the helical structure of an A-form RNA helix. According to the sizes of native PKs, as listed in Table 1, we find that the wild-type PK structures prefer small sizes of H1 (<7-bp) with a weaker preference for H2.

Although the impact of the helix contribution on the topological fraction (averaged over different loops) is

weaker than that of the loop contribution (Figure 3 C), the two helices and three loops in a PK motif are strongly coupled structurally by the loop-helix connectivity and loop-helix excluded volume effect in 3D. Such a coupling results in loop thermodynamic parameters dependent on both loop and helix sizes (41-44). In contrast, the loop parameter for a secondary structural motif, such as an internal loop, is mainly determined by loop itself. The structural coupling effects above become stronger for PKs with large helices and short loops, such as 2n6q and 2tpk.

Hairpin-hairpin kissing motifs

As shown in Supplementary Figure S4(B-1), the hairpin-hairpin kissing motif involves at least three (H1, H2 and H3) helices and four (L1, L2, L4 and L5) loops, with the inter-strand hairpin-hairpin kissing base pairs of H2. Loop L3 would convert a two-strand KISS motif into a one-strand KISS motif with enhanced structural topological constraints. Furthermore, several RNA structures involve the kissing base pairs between the two hairpins within a multi-branched junction, as shown in Supplementary Figure S4(D-1) of the hairpin-hairpin kissing in a three-way junction. The loops La and Lb and the helix Hab may contribute weakly to the preference of the orientation of Hab, however, their constraints for the three-helix (H1-H2-H3) orientation may be much stronger. Here, we use the virtual loop L6, shown as the dotted orange line in Supplementary Figure S4(B-1), to denote the constraints from loops La and Lb and the helix Hab. In the following calculations, we substitute the helix Hab with two nucleotides, resulting in an effective loop length of L6: $L6 = La + Lb + 2$. The effective hairpin-hairpin kissing motif, described by the size of (H1-H2-H3)-(L1-L2-L3-L4-L5-L6), involves three helices and six loops.

As shown in Supplementary Figure S4(B-2), the superposition of the seven crystal KISS structures in Table 2 with different helix and loop sizes suggests that these structures share the similar global topology (helical configurations). Since the L2 and L4 loops are likely shorter than others, for a fixed H2 helix, we use the HLH motifs of H1-L2-H2 and H2-L4-H3 to efficiently sample the three-helix configurations in their corresponding MBB. For each three-helix configuration, as shown in Supplementary Figure S5, we use the Vfold model to sample the loop conformations and determine the topological constraints. We note that unlike a PK motif, a KISS motif needs two sets of ($\theta, \varphi, \alpha, dx, dy, dz$) to describe each configurations.

As listed in Table 2, the contributions from the L2 and L4 loops, F_{24} , ranges from E-3 to E-5 in magnitude. Shorter loop size, as expected, creates stronger constraints. For example, 4wfl and 4uyk (with longer L2 and L4) involve weaker constraints. The contributions from other four loops, i.e. F_{124} , F_{234} , F_{245} and F_{246} , also show the loop size and helix size dependence of the topological constraints. F_{245} ($= 4.8E-7$) of 3sk1, with L5 = 2 nt, shows a stronger constraint than that from other three loops. Because of the longer loop sizes of 3ds7, 3ivn, and 4lx5, the contribution from all four loops (F_{24x}/F_{24}) are E-1 (much weaker constraint). On the other hand, a longer kissing helix, such as the 5-bp kissing helix in 4wfl and 4uyk, results in a stronger

Table 2. Topological constraints of the KISS motif, defined by the size of (H1–H2–H3)-(L1–L2–L3–L4–L5–L6). We use loops L2 and L4 to randomly sample the three-helix configurations in the MBBs of H1–L2–H2 and H2–L4–H3. F_x is the fraction of the Vfold-allowed helical configurations with the corresponding loops. F_{KISS} (F_{123456}) denotes the total topological constraints of a KISS structure

PDB	Size	F_{24}	F_{124}	F_{234}	F_{245}	F_{246}	F_{KISS}
1e8o	(7–2–4)–(3–1–4–1–5–2)	8.5E–4	6.0E–5	3.9E–5	3.6E–5	2.7E–5	1.4E–7
3skl	(6–2–7)–(5–0–8–0–2–7)	2.3E–5	9.4E–6	6.7E–6	4.8E–7	1.2E–5	4.4E–8
3ds7	(7–2–6)–(5–0–8–0–5–7)	2.0E–5	8.8E–6	6.8E–6	9.4E–6	1.5E–5	1.0E–6
3ivn	(6–2–5)–(6–1–8–1–6–7)	9.4E–4	6.5E–4	2.8E–4	7.2E–4	7.1E–4	1.3E–4
4lx5	(6–2–6)–(6–1–8–0–5–7)	1.5E–4	9.9E–5	4.2E–5	7.5E–5	9.8E–5	1.1E–5
4wfl	(4–5–5)–(4–2–3–3–3–2)	5.2E–3	1.7E–4	4.0E–5	9.4E–5	3.7E–5	9.9E–11
4uyk	(5–5–4)–(3–1–4–4–5–2)	3.4E–3	4.1E–5	8.2E–5	3.3E–4	4.1E–5	2.5E–9

topological restriction (F_{24x}/F_{24} ranges from E–2 to E–3 for all four loops). This is because the formation of the kissing helix H2 in a KISS motif tends to rigidify the two originally flexible hairpin loops. The total topological constraints, F_{KISS} , ranges from 1E–4 to 1E–11 for the seven cases.

A KISS motif contains four looping circuits: L1–H2 (one chain)-L2–H1 (terminal base pair), L4–H2 (one chain)-L5–H3 (terminal base pair), L2–H1 (one chain)-L3–H3 (one chain)-L4–H2 (terminal base pair), and L3–H3 (terminal base pair)-L6–H1 (terminal base pair), as shown the red dotted lines in Supplementary Figure S1. It is the simultaneous effects of the four ‘loops’, in addition to the helix formation and the resultant excluded volume and chain rigidity effects, that contribute to the total topological constraints of a KISS motif. For a PK motif in Supplementary Figure S1, there are only two circuits: L1–H2 (one chain)-L2–H1 (terminal base pair) and L2–H1 (one chain)-L3–H2 (terminal base pair), thus, the topological constraints on average are weaker than the KISS motif. On the other hand, a secondary structural motif (hairpin, internal, and bulge loops and multi-branched junctions) involves only one circuit, and shows much weaker topological constraints than a tertiary structural motif.

Extended kissing motifs

The generalized hairpin-hairpin kissing motif can be extended to the following more complicated kissing motifs, involving multiple helices and loops. Supplementary Figure S4(C-1) gives a hairpin-internal loop kissing motif. Supplementary Figure S4(D-1) denotes a hairpin loop kissed by a stem-loop terminal loop. Due to the extremely long computational time, we do not perform the exhaustive conformational sampling and comprehensive analysis for the topology constraints for these two extended kissing motifs. It is important to note that as shown in Supplementary Figure S4(C-2 and D-2), the crystal structures of the corresponding RNAs are similar to each other, despite the different helix and loop sizes and different sequence (listed in Supplementary Tables S3 and S4).

Here we use the RNApdbee (45) to extract the 2D structures for all the cases of the two extended kissing motifs listed in Supplementary Tables S3 and S4. Although some base pairs are not treated as standard canonical base pairs in RNApdbee, according to the global structures shown in Supplementary Figures S4(C-2 and D-2), we manually categorize them into the above two kissing motifs. The small perturbation in the sequence leads to slight changes in 2D structures, such as the different sizes of loops and helices.

Such changes would subsequently cause changes in the conformational space. However, the structure similarity of the above two extended KISS motifs, as well as the PK and KISS motifs indicates that RNAs with same tertiary structural motifs may conserve their global structural folds with the tolerance of slight changes in the sizes of loops and helices, which raises a potential application for RNA 3D structure prediction.

Application to RNA structure prediction

Despite the rapid development of various computational models for RNA structures (17–39), accurate and efficient predictions for RNA tertiary folds remain a significant challenge and impact our mechanistic understanding of RNA functions. Knowledge-based RNA 3D structure prediction methods, such as MC-sym, Vfold3D, RNAcomposer, are severely limited by the completeness of the motif templates database. Due to the limited numbers of known RNA 3D structures in PDB database, currently, the low success rate of finding a proper template for a given motif, especially for the motifs with cross-linked contacts, hampers predictions of large RNAs with tertiary base pairs. From our analysis of the structural topological constraints for the PK and KISS motifs, which concludes that the tertiary base pairs can significantly reduce the conformational space, we develop a topology-directed RNA 3D structure prediction approach to overcome the limitation of the motif template-based prediction methods.

As a benchmark test, we compare the new multi-stage model with RNAcomposer and VfoldLA servers for 22 RNAs as shown in Figure 4(A). The comparisons are based on the same 2D structures provided as input. The corresponding 2D structures are extracted from the native 3D structures (see Supplementary Table S5 for details). All the tested RNAs contain cross-linked base pairs of different structural complexity, with the sizes ranging from 44 to 107 nt. To avoid using templates from the native structures, we exclude all the native structures from the template database. For RNAcomposer, we upload the native structure to exclude native 3D structure elements and set the maximum number of 3D models for each prediction to be ten. For VfoldLA, we use the five clusters given by the default output. The model with the minimum heavy-atom RMSD is selected as the predicted structures for each test case.

As shown in Figure 4A, our model gives the low-resolution predictions for most tested RNAs, with 14 (out of 22) cases having RMSDs in the range of (5.0, 10.0) Å. One RNA (PDB: 5kh8) achieves the near-atomic prediction

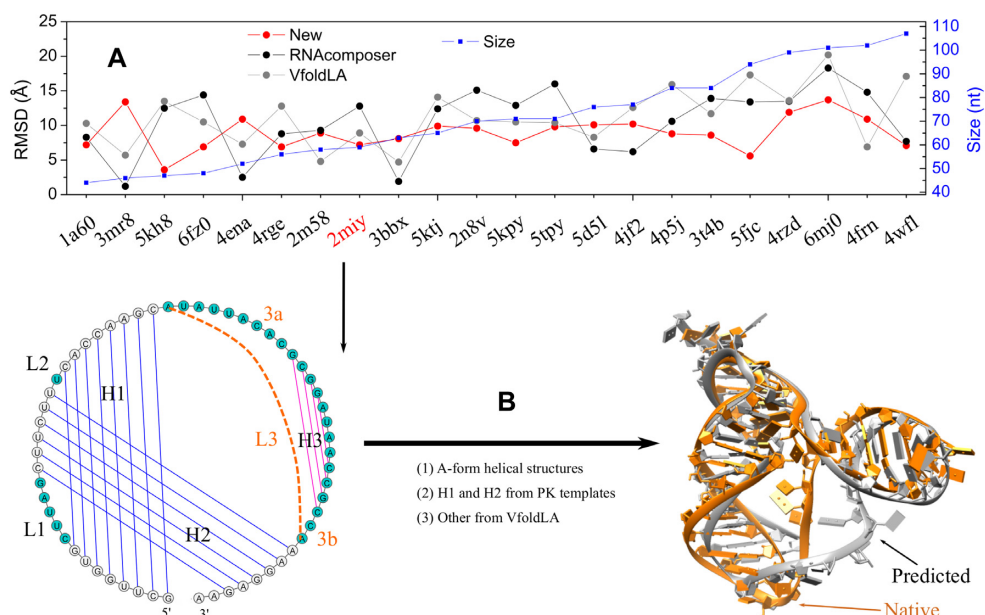


Figure 4. (A) Comparison of the RMSDs between RNAcomposer, VfoldLA, and the current new model. The sequences and 2D structures of all the test RNAs are listed in supplementary Table S5. (B) Topology-based RNA 3D structure prediction for the case of 2miy, a PK with substructure. The method integrates tertiary structural motif templates for helix orientation, Vfold3D for secondary structural motifs, and VfoldLA for single-stranded loop structures.

with the RMSD < 5.0 Å, while seven of them have large RMSDs (> 10.0 Å). The average RMSD is 8.9 Å for the 22 tested cases. RNAcomposer and VfoldLA have the similar performance, with the average RMSDs of 10.6 Å and 11.2 Å, respectively. Although the native 3D structure elements are excluded from predictions, RNAcomposer can still reach the atomic resolution (< 3.0 Å) for three cases. However, there are 13 (15) tested cases with large RMSDs (> 10.0 Å) from RNAcomposer (VfoldLA) predictions. As shown in Figure 2, our model builds the helix configurations of tertiary motifs (involving cross-linked base pairs) based on the templates of known structures. The better treatment of the topological constraints induced by the cross-linked base pairs, such as PK and KISS motifs, may contribute to the improved accuracy.

As shown in Figure 4B, the preQ1 Class II riboswitch from *Streptococcus pneumoniae* (PDB: 2miy) (46), is a pseudoknotted structure with a hairpin substructure in loop L3. By substituting the hairpin substructure with two nucleotides in loop L3, we convert its 2D structure into a PK motif attached by a 2-nt 3' tail and a hairpin structure (helix H3 and a 4-nt hairpin loop) with flanking 9-nt and 2-nt tails on two sides. We build the two-helix (H1 and H2) orientation based on a PK motif template (1yg4), which has the similar loop and helix sizes. We then predict the structures for the hairpin and all the loops using VfoldLA. As shown the 3D structures in Figure 4B, the predicted structure preserves the coaxial stacking interactions between H1 and H2. Compared with the VfoldLA predictions, the improved accuracy (7.2 Å from the current new model versus 8.9 Å from VfoldLA) stems from the better prediction of the helix configuration (H1 and H2) for the PK motif. In fact, the structure of the hairpin substructure (with weak topological constraints) is determined mainly by the interactions

within loops (3a and 3b shown in Figure 4B) and between loops and helices. The prediction based on single-stranded loop sequences (VfoldLA) may not completely preserve the required interactions, leading to the (slightly) misplaced helix H3. Therefore, additional information, such as the tertiary interactions (contacts), may be needed to further improve the prediction.

In addition, we use Challenge #13 of *RNA puzzles* as another example to illustrate our multi-stage method for RNA 3D structure prediction. *RNA Puzzles* is a CASP-like blind test and critical assessment of RNA structure prediction algorithms (17–19). Challenge #13 is for the structure of a 71-nt RNA sequence. To predict the 3D structure, we first predict the 2D structure using Vfold2D (38) with the information from Rfam (47). As shown in Figure 5A, the predicted 2D structure contains four helices and seven loops with a hairpin-internal loop kissing structure. From the known RNA 3D structures in the PDB database, we found that the crystal structure of the cobalamin riboswitch regulatory element (48) (PDB: 4frn) has a similar topology as the predicted 2D structure, with the two blue helices coaxial stacked and two additional helices (the two-bp helix can be considered as the intra-junction interactions). Therefore, as shown in Figure 5B, we placed four A-form helices according to the positions of the corresponding helices in the crystal structure 4frn. Furthermore, based on the MOHCA data (available at <https://rmdb.stanford.edu>), we refined the position of the helix in cyan to accommodate the tertiary contact between the nucleotides A26 and C66 (shown in red color in Figure 5B). Finally, we predicted all the loop structures using VfoldLA. The predicted structure shown in Figure 5C has the RMSD of 5.9 Å.

As indicated in the preceding section, the three helices of the hairpin-hairpin kissing motif are topologically coupled

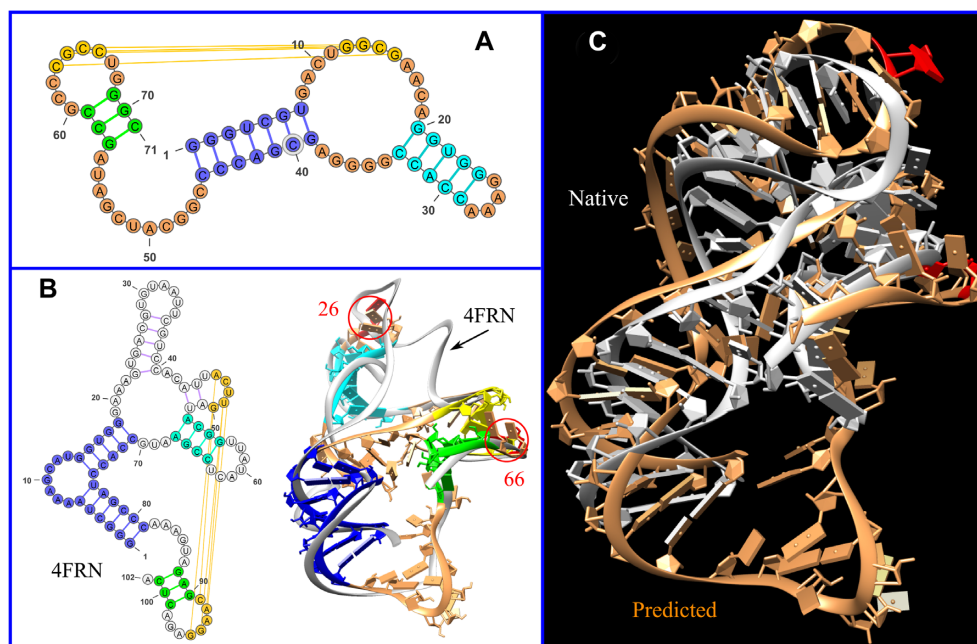


Figure 5. Topology-based RNA 3D structure prediction. (A) The 2D structure used for challenge #13 of *RNA Puzzles*, which contains four helices and seven loops. The helix in yellow is a hairpin-internal loop kissing helix. (B) The 2D and 3D (white ribbon) structures of the cobalamin riboswitch regulatory element (PDB: 4frn). We use the same color to show the alignment of the four helices in (A) to the helix parts of 4frn in (B). Since the two helices linked by a bulge loop, shown as the blue part of 4frn, are actually coaxially stacked in 3D, we consider the two helices as one effective helix. We model the four helices in (A) as A-form helices and place them according to the 3D topology of 4frn as shown in the right panel of (B). The 3D structures of the seven loops in (A) are assembled by VfoldLA. (C) The comparison between the predicted (in brown) and native (in white) structures (all-heavy-atom RMSD is 5.9 Å). We refine the position of helix in cyan according to the tertiary contact between A26 and C66 derived from the MOHCA data.

due to the loop–helix linkage. Therefore, based on the crystal structure of 4frn with the similar topology, we can predict the helix orientation (of the blue, green, and yellow helices shown in the figure). However, the additional helix (in cyan), which is one of the helices in the internal loop, has weak topological coupling with other helices and loops. The usage of the information of the tertiary interactions from the MOHCA data can improve the prediction accuracy.

CONCLUSIONS

Through a systematic analysis for the topological constraints in pseudoknot (PK) and kissing (KISS) motifs, we propose a topology-based RNA 3D structure prediction method. Using the Vfold model to construct three-dimensional maps for the helix orientations, we find that the Vfold-accessible conformational space for tertiary structural motifs, such as PK and KISS, is dominantly determined by the motif type, helix size and loop size, indicating a strong structural (topological) coupling between helices and loops in RNA tertiary motifs. The similar global folds of the native structures, as shown in Supplementary Figure S4, suggest that the native structure may be predominantly determined by the topological constraints encoded at the 2D structure level.

Based on the dominant role of tertiary contacts (cross-linked base pairs) in determining the global topology of RNA 3D structure, we develop a topology-based RNA 3D structure prediction algorithm. The approach incorporates VfoldLA for single-stranded loops, Vfold3D for secondary structural motifs and tertiary structural motifs. The key ingredient of the structure prediction method is to build

global scaffold (helix orientations) based on the homologous topology, predict component structures for loops and secondary structural motifs using previously developed models, and refine the structures with additional tertiary contacts information if available. This new method gives improved predictions for RNAs with tertiary (cross-linked) base pairs. However, similar to other template-based RNA structure prediction models, the method developed here cannot predict structures of ‘new’ topology. With the increasing number of the known RNA structures, the larger and more divergent pools of known structural topologies would lead to better predictions from the topology-based method. Further improvement of the model includes: (I) building an RNA topology database, which can accommodate non-canonical base pairs and different internal/bulge loops within helices; (II) improving the size similarity-based scoring scheme to better search for the optimal topologies in the database; (III) developing an all-atom scoring function to better select predicted structures when multiple templates are available.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Institutes of Health [R01-GM063732, R01-GM117059, R35-GM134919 to S.J.C.]. Funding for open access charge: NIH grants.

Conflict of interest statement. None declared.

REFERENCES

- Brion, P. and Westhof, E. (1997) Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 113–137.
- Tinoco, I. Jr and Bustamante, C. (1999) How RNA folds. *J. Mol. Biol.*, **293**, 271–281.
- Batey, R. T., Rambo, R. P. and Doudna, J. A. (1999) Tertiary Motifs in RNA Structure and Folding. *Angew. Chem. Int. Ed. Engl.*, **38**, 2326–2343.
- Murthy, V. L., Srinivasan, R., Draper, D. E. and Rose, G. D. (1999) A complete conformational map for RNA. *J. Mol. Biol.*, **291**, 313–327.
- 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.
- Mustoe, A. M., Al-Hashimi, H. M. and Brooks, C. L. III (2014) Coarse grained models reveal essential contributions of topological constraints to the conformational free energy of RNA bulges. *J. Phys. Chem. B*, **118**, 2615–2627.
- Sim, A. Y. and Levitt, M. (2011) Clustering to identify RNA conformations constrained by secondary structure. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 3590–3595.
- Parisien, M. and Major, F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51–55.
- Mustoe, A. M., Bailer, M. H., Teixeira, R. M., Brooks, C. L. III and Al-Hashimi, H. M. (2012) New insights into the fundamental role of topological constraints as a determinant of two-way junction conformation. *Nucleic Acids Res.*, **40**, 892–904.
- Bailer, M. H., Sun, X. and Al-Hashimi, H. M. (2010) Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science*, **327**, 202–206.
- Butcher, S. E. and Pyle, A. M. (2011) The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.*, **44**, 1302–1311.
- Bailer, 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.
- Behrouzi, R., Roh, J. H., Kilburn, D., Briber, R. M. and Woodson, S. A. (2012) Cooperative tertiary interaction network guides RNA folding. *Cell*, **149**, 348–357.
- Izzo, J. A., Kim, N., Elmetwaly, S. and Schlick, T. (2011) RAG: an update to the RNA-As-Graphs resource. *BMC Bioinformatics*, **12**, 219.
- Kim, N., Laing, C., Elmetwaly, S., Jung, S., Curuksu, J. and Schlick, T. (2014) Graph-based sampling for approximating global helical topologies of RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 4079–4084.
- Kerpedjiev, P., Siederdisen, C. H. and Hofacker, I. V. (2015) Predicting RNA 3D structure using a coarse-grain helix centered model. *RNA*, **21**, 1110–1121.
- Cruz, J. A., Blanchet, M. F., Boniecki, M., Bujnicki, J. M., Chen, S.-J., Cao, S., Das, R., Ding, F., Dokholyan, N. V., Flores, S. C. *et al.* (2012) RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA*, **18**, 610–625.
- Miao, Z., Adamiak, R. W., Blanchet, M. F., Boniecki, M., Bujnicki, J. M., Chen, S.-J., Cheng, C., Chojnowski, G., Chou, F. C., Cordero, P. *et al.* (2015) RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. *RNA*, **21**, 1066–1084.
- Miao, Z., Adamiak, R. W., Antczak, M., Batey, R. T., Becka, A. J., Biesiada, M., Boniecki, M. J., Bujnicki, J. M., Chen, S.-J., Cheng, C. Y. *et al.* (2017) RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA*, **23**, 655–672.
- Seetin, M. G. and Mathews, D. H. (2012) RNA structure prediction: an overview of methods. *Methods Mol. Biol.*, **905**, 99–122.
- Sun, L.-Z., Zhang, D. and Chen, S.-J. (2017) Theory and modeling of RNA structure and interactions with metal ions and small molecules. *Annu. Rev. Biol.*, **46**, 227–246.
- Miao, Z. and Westhof, E. (2017) RNA structure: advances and assessment of 3D structure prediction. *Annu. Rev. Biol.*, **46**, 483–503.
- Tan, R. K. Z., Petrov, A. S. and Harvey, S. C. (2006) YUP: A molecular simulation program for coarse-grained and multi-scaled models. *J. Chem. Theory Comput.*, **2**, 529–540.
- Xia, Z., Gardner, D. P., Gutell, R. R. and Ren, P. (2010) Coarse-grained model for simulation of RNA three dimensional structures. *J. Phys. Chem. B*, **114**, 13497–13506.
- Shi, Y. Z., Wang, F. H., Wu, Y. Y. and Tan, Z.-J. (2014) A coarse-grained model with implicit salt for RNAs: Predicting 3D structure, stability and salt effect. *J. Chem. Phys.*, **141**, 105102.
- Pasquali, S. and Derreumaux, P. (2010) HiRE-RNA: A high resolution coarse-grained energy model for RNA. *J. Phys. Chem. B*, **114**, 11957–11966.
- Sharma, S., Ding, F. and Dokholyan, N. V. (2008) iFoldRNA: three-dimensional RNA structure prediction and folding. *Bioinformatics*, **24**, 1951–1952.
- Boniecki, M. J., Lach, G., Dawson, W. K., Tomala, K., Lukasz, P., Soltysinski, T., Rother, K. M. and Bujnicki, J. M. (2016) SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.*, **44**, e63.
- Magnus, M., Boniecki, M. J., Dawson, W. and Bujnicki, J. M. (2016) SimRNAweb: a web server for RNA 3D structure modeling with optional restraints. *Nucleic Acids Res.*, **44**, W315–W319.
- Zhang, D. and Chen, S.-J. (2018) IsRNA: An iterative simulated reference state approach to modeling correlated interactions in RNA folding. *J. Chem. Theory Comput.*, **14**, 2230–2239.
- Zhang, J., Dundas, J., Lin, M., Chen, R., Wang, W. and Liang, J. (2009) Prediction of geometrically feasible three dimensional structures of pseudoknotted RNA through free energy estimation. *RNA*, **15**, 2248–2263.
- Rother, M., Rother, K., Puton, T. and Bujnicki, J. M. (2011) ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.*, **39**, 4007–4022.
- Bida, J. P. and Maher, H. J. 3rd (2012) Improved prediction of RNA tertiary structure with insights into native state dynamics. *RNA*, **18**, 385–393.
- Das, R. and Baker, D. (2007) Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 14664–14669.
- Das, R., Karanicolas, J. and Baker, D. (2010) Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods*, **7**, 291–294.
- Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K. J., Lukasiak, P., Bartol, N., Blazewicz, J. and Adamiak, R. W. (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res.*, **40**, e112.
- Cao, S. and Chen, S.-J. (2011) Physics-based de novo prediction of RNA 3D structures. *J. Phys. Chem. B*, **115**, 4216–4226.
- Xu, X., Zhao, P. and Chen, S.-J. (2014) Vfold: a web server for RNA structure and folding thermodynamics prediction. *PLoS One*, **9**, e107504.
- Xu, X. and Chen, S.-J. (2018) Hierarchical assembly of RNA three-dimensional structures based on loop templates. *J. Phys. Chem. B*, **122**, 5327–5335.
- Cao, S. and Chen, S.-J. (2005) Predicting RNA folding thermodynamics with a reduced chain representation model. *RNA*, **11**, 1884–1897.
- Cao, S. and Chen, S.-J. (2006) Predicting RNA pseudoknot folding thermodynamics. *Nucleic Acids Res.*, **34**, 2634–2652.
- Cao, S. and Chen, S.-J. (2009) Predicting structures and stabilities for H-type pseudoknots with inter-helix loop. *RNA*, **15**, 696–706.
- Cao, S. and Chen, S.-J. (2011) Structure and stability of RNA/RNA kissing complex: with application to HIV dimerization initiation signal. *RNA*, **17**, 2130–2143.
- Cao, S., Xu, X. and Chen, S.-J. (2014) Predicting structure and stability for RNA complexes with intermolecular loop-loop base pairing. *RNA*, **20**, 835–845.
- Antczak, M., Zok, T., Popenda, M., Lukasiak, P., Adamiak, R. W., Blazewicz, J. and Szachniuk, M. (2014) RNApDbee – a webserver to derive secondary structures from pdb files of knotted and unknotted RNAs. *Nucleic Acids Res.*, **42**, W368–W372.
- Kang, M., Eichhorn, C. D. and Feigon, J. (2014) Structural determinants for ligand capture by a class II preQ1 riboswitch. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E663–E671.
- Burge, S. W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E. P., Eddy, S. R., Gardner, P. P. and Bateman, A. (2013) Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.*, **41**, D226–D232.
- Johnson, J. E., Reyes, F. E., Polaski, J. T. and Batey, R. T. (2012) B12 cofactors directly stabilize an mRNA regulatory switch. *Nature*, **492**, 133–137.