Design Principles for Single-Stranded RNA Origami Structures

Cody W. Geary and Ebbe Sloth Andersen*

Interdisciplinary Nanoscience Center, Aarhus University,
Gustav Wieds Vej 14, 8000 Aarhus, Denmark
esa@inano.au.dk

Abstract. We have recently introduced an experimental method for the design and production of RNA-origami nanostructures that fold up from a single strand while the RNA is being enzymatically produced, commonly referred to as cotranscriptional folding. To realize a general and scalable architecture we have developed a theoretical framework for determining RNA crossover geometries, long-distance interactions, and strand paths that are topologically compatible with cotranscriptional folding. Here, we introduce a simple parameterized model for the A-form helix and use it to determine the geometry and base-pair spacing for the five types of RNA double-crossover molecules and the curvature resulting from crossovers between multiple helices. We further define a set of paranemic loop-loop and end-to-end interactions compatible with the design of folding paths for RNA structures with arbitrary shape and programmable curvature. Finally, we take inspiration from space-filling curves in mathematics to design strand paths that have high-locality, programmed folding kinetics to avoid topological traps, and structural repeat units that might be used to create infinite RNA ribbons and squares by rolling circle transcription.

Keywords: RNA, structure, folding, kinetics, space-filling curves.

1 Introduction

RNA molecules have a greater structural diversity than DNA and are thus of interest as design substrate for biomolecular engineering. The field of RNA nanotechnology has taken inspiration from the modular nature of structurally characterized RNA molecules to develop a design paradigm where structural modules can be composed to achieve complex geometric shapes and lattices [13]. However, RNA modular design is currently limited to rather small RNAs, suggesting that there is still a need to develop a more standardized design approach such as that used in DNA nanotechnology [2] and best exemplified by the DNA origami method [22]. Particularly, in DNA nanotechnology the double-crossover (DX) motif plays a central role in organizing DNA helices into large arrays, but the RNA field has yet to leverage RNA DX motifs to build similarly large structures.

* Corresponding author.

S. Murata and S. Kobayashi (Eds.): DNA 2014, LNCS 8727, pp. 1–19, 2014.

DX molecules have been described extensively for B-form DNA helices [9], but to our knowledge not in great detail for A-form RNA helices, which we will now describe. Several papers have revealed that using a simplified model considering only the helical twist of 11 base pairs (bps) per turn for RNA is not sufficiently accurate for fully realizing design in RNA. For example, for the RNA/DNA hybrid design of Mao and colleagues it was revealed that the inclination of the bases in RNA influences the tiling behavior of the RNA/DNA nanostructures [17]. Likewise, in Delebecque et al. RNA assemblies designed considering only the helical twist required the inclusion of unpaired bases within the tile [5], which we here propose generate a required flexibility to counteract the distance offset caused by the base-pair inclination. Furthermore, studies of RNA assemblies based on the paranemic crossover motif (PX) have found that the ideal tile designs differ significantly from the DNA versions, due to the difference in width of the major and minor grooves of RNA compared to DNA [1]. Recently, we initiated the design and testing of RNA double-crossover molecules with crossover spacings based on calculated distances and three-dimensional (3D) modeling, and demonstrated very robust formation of RNA DX motifs [12]. Here, we extend upon this work and describe the DX motifs for RNA in detail, and propose a new method for the systematic design of RNA nanostructures.

A useful feature of RNA structure is that it can be produced directly from the RNA polymerase enzyme by the cotranscriptional self-assembly process. In this process the RNA folds locally as it is transcribed and the structure gradually builds up as more sequence is produced by the enzyme. The autonomous enzymatic process furthermore allows engineered RNA structures to be genetically encoded and expressed in cells, much like natural RNAs. We have recently demonstrated the design of single-stranded RNA origami structures and their production by both heat-annealing and/or cotranscriptional folding [12]. To develop this method further we consider the cotranscriptional folding process as it relates to requirements for the order of folding and assembly events along the RNA strand, and propose geometrical and topological principles for single-stranded RNA nanostructures.

2 Simple Model of a Double Helix

In DNA nanotechnology simplified helices are often used when designing larger constructs to decrease the computational requirements for handling large 3D objects, since only a few parameters are needed to determine crossover positions [24]. By contrast, many artificial RNA nanostructures are still designed using fully-atomistic models [23,14], in a tedious process requiring specialized experience, or by using over-reduced models that do not capture important details of the RNA helix [5]. Thus, there is a need for an appropriately simplified RNA model that can still be easily manipulated. Here, we propose such a model that is a parameterization describing only the positions of phosphate atoms on the helix, which we call the "P-stick model". Our model generates the atomic positions of phosphates relative to the central axis of the helix, based on five parameters: base inclination relative to the helix-axis, rise between bases, helix

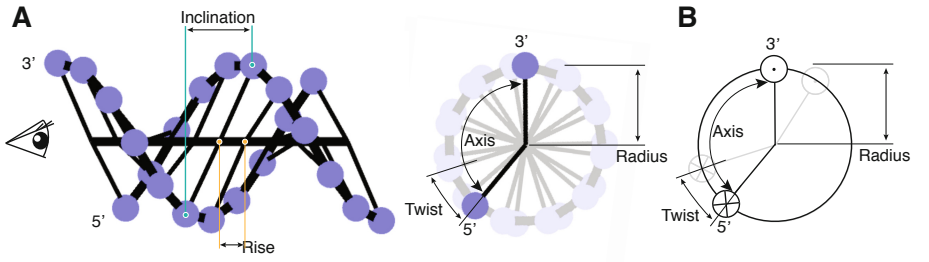

Fig. 1. P-stick model for the A-form helix. (A) Left: Model of A-form helix shown in side view with indications of inclination and rise distances along helical axis. Right: End view of the same A-form helix model as seen along the helical axis from the left as indicated by the eye symbol. The radius, axis and twist angles are indicated. Lines connecting the phosphates to the central axis depict the connectivity of the helix and do not represent base-pairs. **(B)** Schematics of P-stick model with 3' end indicated by circle with a dot (tip of arrow) and 5' end indicated by a circle with a cross (end of arrow). The second bp is shown in grey to indicate twist angle and the right handed rotation of the helix.

Table 1. Parameters for P-stick helix model. Parameters are measured by the authors based on data from Arnot et. al. 1973 [3] and Dickerson et al. 1982 [6]. We also note that the inclination variable measures the inclination of paired-phosphates relative to the helical axis, and should not be confused with base-pair inclination.

Parameters	Variables	A-form	B-form
Radius	R	8.7 Å	9.3 Å
Rise	D	2.81 Å	3.4 Å
Inclination	I	-7.45 Å	3.75 Å
Axis	A	139.9°	170.4°
Twist	T	32.73°	34.48°
Helicity	H	11 bp	10.44 bp

radius, axis angle across the minor groove, and helicity in bps per turn (described in Fig 1A).

Parameters for RNA A-form helices were measured from standard helices generated by Westhof's Assemble2 program [16], which are based on classical A-form parameters [3]. Parameters for DNA B-form helices were derived from various literature sources: The helicity was chosen to be 10.44 bps/turn based on NMR measurements in solution [26] and parameters found to be the optimal for producing twist-corrected DNA origami structures [27]. The inclination between phosphates was found by aligning the B-DNA crystal structure (PDB-ID 1BNA) [7] and averaging for all phosphate positions giving a value of 3.748 Å. Parameters for both A- and B-form helices are provided in Table 1.

The P-stick model consists of two equations that generate xyz coordinates for the phosphate atoms along each strand of a helix, with the axis centered on the origin pointing along the x-axis:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} p \cdot D \\ R \cdot \cos\left(p \cdot \left(\frac{2\pi}{H}\right)\right) \\ R \cdot \sin\left(p \cdot \left(\frac{2\pi}{H}\right)\right) \end{bmatrix}, \; p|p \in \mathbb{Z} \tag{1}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} p \cdot D + I \\ R \cdot \cos\left(p \cdot \left(\frac{2\pi}{H}\right) + \left(\frac{A \cdot \pi}{180^\circ}\right)\right) \\ R \cdot \sin\left(p \cdot \left(\frac{2\pi}{H}\right) + \left(\frac{A \cdot \pi}{180^\circ}\right)\right) \end{bmatrix}, \; p|p \in \mathbb{Z} \tag{2}$$

Eq(1) describes the strand running from 3' to 5' along the x-axis and Eq(2) describes the complement strand running 5' to 3'.

3 RNA Crossover Motifs

The alignment of phosphates between helices is taken as an indicator of positions where a crossover junction can join the helices [24]. Here we use the P-stick model to predict optimal crossover spacings for A- and B-form helices. To construct a crossover between two helices we first align one pair of phosphates between the helices, then break the backbone bond on the same side of each phosphate and rejoin them with the opposite strand. Likewise, to construct a DX between two helices we orient the helices in parallel and rotate the helices to align two pairs of phosphates in the plane, and perform two crossover operations on the backbone.

3.1 The RNA DX Types

DXs can either be formed between anti-parallel (A) or parallel (P) strands, and the spacing between the two crossovers (measured by the number of half turns) can either be even (E) or odd (O). Even-spaced crossovers result in both crossovers bridging the same strand, while odd-spacing results in the two crossovers formed between opposite strands (Fig. 2). Thus, DX molecules can take the form of AE, AO, PE and PO. Two possibilities exist for PO since the DX can bridge either the major or minor groove. For B-form helices the major groove is wide (W) and the minor groove is narrow (N), and the DXs are thus called PON and POW [9]. For A-form helices the major groove is narrow and deep (D) and the minor groove is wide and shallow (S), and thus we propose to name the A-form PO-DX motifs POD and POS. PO-DXs with more than one groove spanning the crossovers are named depending on which groove is predominant. PO-DXs that span a single groove width have the property of being paranemic, and have been described and studied for both DNA and RNA, although with non-ideal spacings for RNA[1]. Interestingly, the DX molecules all have different pseudosymmetry axes (shown in Fig. 2) that are important for their use as building blocks, which we will discuss later.

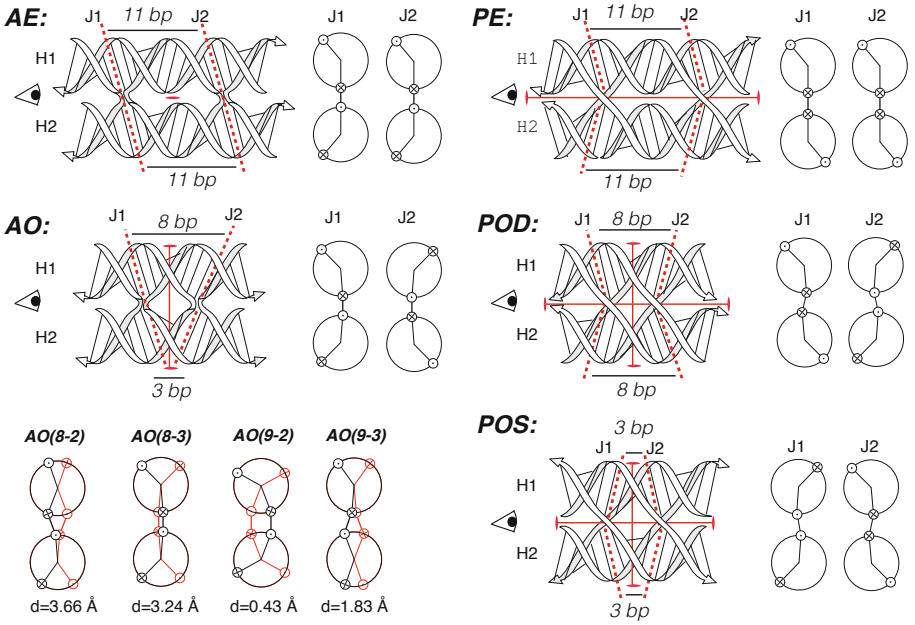

Fig. 2. Minimal A-form DX molecules. A-form DX of type AE, AO, PE, POD and POS (see section 3.1 for naming conventions) that can be expanded by $11 * n$ bps. **Left:** Ribbon diagrams show how the helix inclination affects the base-pair spacing between the two crossovers. Red dashed lines show the junctions and their inclinations. Base-pair spacing is indicated above and below, with horizontal lines. Red lines and ellipse shapes show C2 pseudosymmetry axes. Ribbon diagrams are based on P-stick models and drawn using positions of P and C3 atoms. Helices (H) and junctions (J) are numbered from top to bottom and left to right, respectively. **Right:** The two junctions are shown with schematics similar to Fig. 1B as seen along the helical axis from the eye symbol. The DX adjusted rotation of the helices are shown. **Bottom left:** Four AO DXs with different spacing are shown in side view, where black diagram is J1 and red diagram is J2. The distance between P atoms at crossovers is written below each side view.

3.2 Spacing between A-Form DXs

We use the P-stick model to calculate optimal crossover positions, which is done by measuring the distance between pairs of phosphates on two strands and finding the pairs for which the distances are at a minimum. For each crossover pattern that we analyze, we provide distance and angular measurements to qualify how well-aligned the crossover pattern is (Table 2). The distance between two crossovers (DX-$dist_x$) is measured from the midpoint between the phosphates of one crossover to the midpoint of phosphates of the second crossover. The angular gap (ΔAngle) of the crossovers is calculated by perfectly aligning

the phosphates of the first crossover, and then measuring the angular offset of the phosphates from the center-line of the second crossover. We calculate the bp spacing, derived from the number of phosphates between the crossovers, and the direction of the strands as they pass through the crossovers. Because of the different strand directionality in odd- versus even-spaced junctions, the number of bps in the junction have the following relations:

$$\begin{aligned} bp &= P - 1 \quad \text{for AE and PE} \\ bp &= P \quad\quad\ \ \text{for AOD and POD} \\ bp &= P - 2 \quad \text{for AOS and POS.} \end{aligned}$$

Thus, for AO junctions where the minor groove of the first helix (H1) aligns with the major groove of the second helix (H2), the two strands must be considered differently. In Table 2, bp values (b_2,b_1) for each set of phosphate spacings (p_2,p_1) are reported.

An interesting feature of all AO- and PO-DXs is that the crossovers occur between opposite strands which, due to the incline of the phosphates, results in an asymmetry. The asymmetry is much more apparent for A-form RNA helices than for B-form DNA helices, and results in a substantial difference in the number of bps on each side of the crossover. Our calculations for the optimal AO-DX (Fig. 2) find that there should be 9 bps in H1 and only 2 bps in H2 (Table 2) and that the phosphates do not perfectly align, resulting in a large 42° angular gap between the crossovers, but a short distance of 0.4 Å between the two crossover phosphates along the helical axis.

Likewise, PO-DXs in RNA have different spacings depending on whether they cross the deep groove (POD) or the shallow groove (POS) of the RNA, where POD-DX have a spacing of 8 bp on each helix and a distance of 12.2 Å between crossovers and POS-DX have only 3 bp on each helix and a distance of 18.7 Å between crossovers. All spacings for RNA crossovers can be extended by $n * 11$ bps, for positive integer n. Spacings for DNA crossovers can be extended by roughly $n*21$ bps, but because of the 10.44bp/turn helicity they are not perfectly equivalent.

Using the previously defined variables for Eq(1), we define the alignment of two helices with parameters that translate and rotate H2 relative to H1, such that the two helices are aligned with phosphate-1 of each strand superimposed. The primary strand is thus defined similarly to Eq(1), where p_1 represents the nth phosphate in strand 1 on H1:

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} (p_1 - 1) \cdot D \\ R \cdot \cos\left(\frac{2\pi(p_1-1)}{H}\right) \\ R \cdot \sin\left(\frac{2\pi(p_1-1)}{H}\right) \end{bmatrix} \tag{3}$$

Likewise, the complement strand 2 is defined by a shift and rotation about the helical axis by the axis and inclination variables, similar to Eq(2):

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} (p_1 - 1) \cdot D + I \\ R \cdot \cos\left(\frac{2\pi(p_1-1)}{H} + \left(\frac{A \cdot \pi}{180^\circ}\right)\right) \\ R \cdot \sin\left(\frac{2\pi(p_1-1)}{H} + \left(\frac{A \cdot \pi}{180^\circ}\right)\right) \end{bmatrix} \tag{4}$$

For the purpose of calculating the crossover spacing distances, H2 is aligned such that its phosphate-1 is superimposed on the phosphate-1 of H1, where variables for the translation and rotation of H2 are defined as follows:

$$Tr_x = I \tag{5}$$

$$Tr_y = R \cdot \cos\left(A \cdot \left(\frac{\pi}{180^\circ}\right)\right) - R \cdot \cos(\theta) \tag{6}$$

$$Tr_z = R \cdot \sin\left(A \cdot \left(\frac{\pi}{180^\circ}\right)\right) - R \cdot \sin(\theta) \tag{7}$$

where $\theta = \pi + A \cdot \left(\frac{\pi}{180^\circ}\right)$. The directionality of H2 is changed by $\delta = 1$ for antiparallel helices and $\delta = -1$ for parallel helices giving the following equations, where p_2 is the nth phosphate:

$$\begin{bmatrix} x_3 \\ y_3 \\ z_3 \end{bmatrix} = \begin{bmatrix} (p_2 - 1) \cdot D + Tr_x \\ R \cdot \cos\left(\frac{2\pi \cdot (p_2-1)}{H} + \theta\right) + Tr_y \\ R \cdot \sin\left(\frac{2\pi \cdot (p_2-1)}{H} + \theta\right) + Tr_z \end{bmatrix} \tag{8}$$

$$\begin{bmatrix} x_4 \\ y_4 \\ z_4 \end{bmatrix} = \begin{bmatrix} (p_2 - 1) \cdot D + \delta \cdot I + Tr_x \\ R \cdot \cos\left(\frac{2\pi \cdot (p_2-1)}{H} + \frac{\delta \cdot A \cdot \pi}{180} + \theta\right) + Tr_y \\ R \cdot \sin\left(\frac{2\pi \cdot (p_2-1)}{H} + \frac{\delta \cdot A \cdot \pi}{180} + \theta\right) + Tr_z \end{bmatrix} \tag{9}$$

The distance between phosphates is then calculated by measuring the euclidian distance between p_1 and p_2 on the pair of strands forming the crossover:

$$dist_{AO} = \sqrt{(x_1 - x_4)^2 + (y_1 - y_4)^2 + (z_1 - z_4)^2} \tag{10}$$

$$dist_{AE} = \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2 + (z_2 - z_3)^2} \tag{11}$$

Thus, for any crossover domain of length p_1 and p_2, we need only evaluate the equations to get the table of distances (P-$dist_{xyz}$ in Table 2).

Using the P-stick model we have calculated parameters for what we propose are the ideal spacings for all four DX types in both A-form and B-form helices.

Table 2. Base-pair spacing for A- and B-form DX molecules. Numbers are given for phosphate and base-pair spacings on helix 1 and 2. The spacings are minimum spacings and can be extended by full helical turns (+11 bps for RNA, and +21 bps for DNA). DX-$dist_x$ is the distance between P-P midpoints of the two crossovers. P-$dist$ is the P-P distance at the J2 crossover. *Negative numbers for p_1 and p_2 are converted to phosphate spacings by taking: -p+2. **Has symmetry equivalent: (p_2,p_1) = (p_1,p_2).

Helix	DX	(p_2,p_1)	(b_2,b_1)	DX-$dist_x$ (Å)	ΔAngle (°)	P-$dist_x$ (Å)	P-$dist_{xyz}$ (Å)
A	AE/PE	(12,12)	(11,11)	30.9	0	0	0
	AO	(8,4)	(8,2)	14.1	25.4	1.83	6.20
	AO	(8,5)	(8,3)	15.5	9.0	3.24	6.48
	AO	(9,4)	(9,2)	15.5	41.7	0.43	4.49
	AO	(9,5)	(9,3)	16.9	25.4	1.83	6.20
	AO	(18,20)	(18,18)	50.6	72.8	10.26	25.17
	POD	(-6,-6)*	(8,8)	12.2	9.0	0	2.73
	POS	(5,5)	(3,3)	18.7	9.0	0	2.73
B	AE/PE	(22,22)	(21,21)	71.4	4.1	0	1.34
	AE/PE	(11,11)	(10,10)	34.0	15.2	0	4.91
	AE/PE	(11,12)**	(10,11)	35.7	17.2	1.7	3.56
	AO	(5,7)	(5,5)	17.0	44.1	0.3	5.68
	AO	(16,17)	(16,15)	52.7	26.8	2.0	4.86
	AO	(16,18)	(16,16)	54.4	44.1	0.3	6.19
	AO	(26,28)	(26,26)	88.4	44.1	0.3	5.37
	POW	(-4,-5)*,**	(6,7)	22.4	17.2	1.7	3.50
	POW	(-15,-15)*	(17,17)	58.1	2.1	0	0.69
	PON	(6,6)	(4,4)	13.3	2.0	0	0.66
	PON	(17,16)	(15,14)	49.0	17.2	1.7	3.73
	PON	(16,16)	(14,14)	47.3	13.2	0	4.26

In addition, we report measurements for commonly used spacings in the literature, for comparison (Table 2). Interestingly, we find that some of the commonly used DX spacings are not the best possible arrangements. However, we also note that non-ideal phosphate spacings may still be preferential for symmetry reasons, or out of a need to have domains of the same length, and thus might represent necessary compromises for structural or thermodynamic reasons that are not considered in our model. Of particular interest, we find that for narrow DNA-AE crossovers a 10/11 bps ((b_2/b_1), respectively) arrangement results in shorter P-$dist_{xyz}$ than a 10/10 bp spacing. Likewise, we find that DNA-AO crossovers have an asymmetry of 16/15 bps which is better than the regularly used 16/16 bp spacing, consistent with early modeling predictions for DNA [9]. For DNA crossovers, we find that the shortest crossover gaps are a PON crossover with 4/4 bps spacing and a POW crossover with 17/17 bps spacing. By contrast in RNA, the 18/18 bps spacing used in Delebecque et al. [5] to form RNA-AO crossovers has a large P-$dist_{xyz}$, likely explaining why the experimenters needed several unpaired bases to provide the necessary flexibility to form these crossovers.

3.3 Multiple Crossovers and Curvature

The DX motifs can be further combined into multiple crossover (MX) molecules. We define the distance between DX molecules in a MX molecule as follows: position 1 (n_1) is set at the first crossover between H1 and H2, where the 5' end of H1 enters the crossover from the left. Position 2 (n_2) is the position of the second crossover between H2 and H3. The spacing (s) is given as the difference between the crossover positions:

$$s = n_2 - n_1$$

and relates to the number of basepairs between the two crossovers. Only certain spacings are possible since some spacings will bring the helices to overlap and sterically clash.

Depending on the DX type, only certain spacings are allowed for producing MX arrays without steric clashes. For RNA-AE, spacings of

$$s = -5, -4, -3, -2, -1, 0, 1, 2$$

are allowed, as well as spacings of $s - 11 * n$. For RNA-PE, spacings of

$$s = -2, -1, 0, 1, 2, 3, 4, 5$$

are allowed, plus spacings of $s - 11 * n$. For both AE and PE, the curvature of multiple helices can be controlled in this fashion. AO has very strict spacing restrictions when the crossovers are placed in close proximity, since the two junctions bend in opposite directions (Fig. 3). This is because AO-DX has an even/odd number of bases in the shallow/deep groove, respectively, and it is thus not possible to have symmetrically opposed curvature on both junctions with this spacing. Thus, to obtain a symmetric molecule we choose a less optimal DX spacing of 9-3 (see Fig. 2 and Table 2), which results in a MX spacing of $s = -3$ between the three helices. The curvature is expected to cancel out by pushing the crossovers in opposite directions (see 3H-AO in Fig. 3) and this molecule has been tested and found to assemble [12]. Similarly, the POS and POD crossovers for RNA also require consideration of symmetry if they are to be used in MX arrays.

Interestingly, another consequence of symmetry is that if one forms MX arrays by repeating a PE crossover pattern then the bending of the array evens out because the structure becomes corrugated with alternating ridges and valleys. Likewise, for a continuous sheet of POS and POD crossovers, the bending also evens out internally in the sheet but results in a strained although flat conformation of the sheet.

The most flat RNA-AE spacing is $s = -1*(n*11)$, which bends only 7 degrees into the plane. However, as the spacing can be variegated row-by-row, another way to make planar sheets of AE or PE DXs is to alternate between two or more different spacings e.g. $s = 0, -2, 0, -3$ for AE as will be used later to construct large planar AE strand paths.

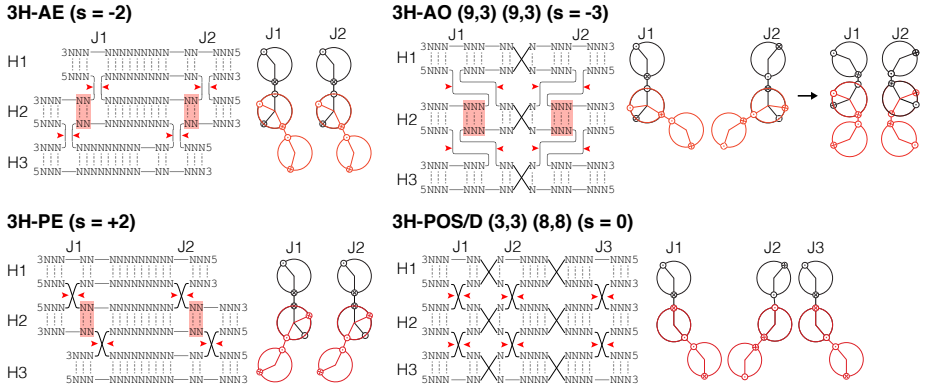

Fig. 3. Crossovers between three helices. RNA 3-helix double-crossover molecules shown as strand diagrams and in side-view. Helices (H) and junctions (J) are numbered from bottom to top and left to right, respectively. Spacing (s) between the sets of crossovers joining the three helices are marked in parenthesis behind the name of the motif, and shown in red on the strand diagrams. Red arrows indicate the positions of the crossovers. H1 and H2 are fixed in their common plane. H3 (marked in red) is rotated depending on the crossovers spacing. For AO crossovers, the curvature induced by crossovers on each side are in opposition and are expected to cancel out by deforming the geometry of the crossover junction, as indicated at the top right.

4 Single-Stranded RNA Origami

We have recently introduced a general architecture called single-stranded RNA origami, where we use the crossover motifs introduced above to fold a single RNA strand into complex topologies [12]. To force the otherwise multi-stranded architecture into a single-stranded form we use kissing-loop interactions to bridge the crossover junctions (Fig. 4), and use minimal spacings at helix junctions (named dovetail seams). The art of folding RNA structures from a single strand without getting into topological problems will be discussed below.

4.1 Paranemic Connectors

The structural diversity of RNA provides numerous long-range RNA-RNA interactions which can be used to form programmable contacts between specific elements in the RNA structure. Long-range interactions can be especially useful when considering crossover patterns that would otherwise require many separate strands. Many different connectors can be found from natural RNA structures, some of which are listed in Fig. 5. Loop-loop interactions are a class of pseudoknot that form between preformed secondary structure elements, bringing distant elements in a 2D structure together in a sequence-programmable fashion. The kissing-loops are typically in the range of 2-7 bps, and are limited to be less than one helical turn in order to avoid being topologically entangled.

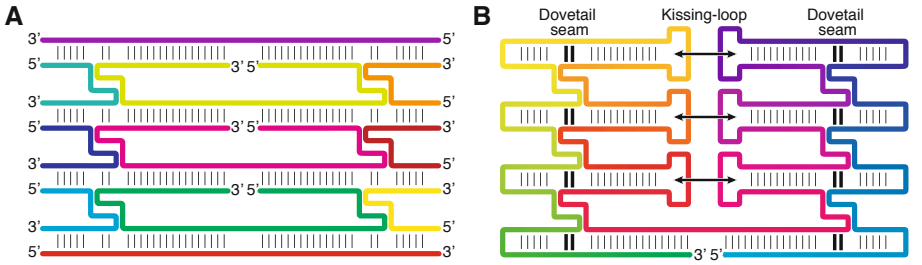

Fig. 4. Multi-stranded to single-stranded conversion. (A) Eleven strands in a multi-crossover motif between four helices. **(B)** Loops are inserted to convert the motif into a single-stranded architecture. Loops on the edges cap the end of the helix. Loop-loop interactions are inserted inside the structure to connect the helices across the crossover junctions.

Loop-receptor interactions (Fig. 5C) represent an entirely different class of programmable long-range interaction that have been extensively studied [4,15,10]. Unlike kissing-loop interactions, these tertiary contacts do not form any Watson-Crick bps in their assembly, but are still highly sequence specific. Because they do not form bps with each other, loop-receptor interactions do not form pseudo-knotted structures, and also have a higher salt requirement compared to kissing-loop interactions. These properties make loop-receptor interactions of interest for use as paranemic connectors because they likely fold in a different time scale compared to kissing loops and are a completely orthogonal class of interaction from kissing loops.

Lastly, RNA-PX crossovers represent a third class of interaction that can be implemented sequence-specifically [1]. Of all the paranemic connectors we have presented here, PX crossovers are the only variety that have been implemented in DNA [25,28].

4.2 Dovetail Seams

Aside from paranemic type connectors, helices can also interact by stacking at their helix ends. In the context of single-stranded RNA origami we call these end-end interactions "dovetail seams" because they can click rigidly together without creating topological problems. The dovetail bps are only possible when the structure is at least three helices wide, and to our knowledge, there are no natural RNA molecules that adopt this architecture. The interesting feature of dovetail seams is that they can be used to program a specific stacking conformation of a junction, as well as define the curvature of the single-stranded RNA origami tile. The small stems formed by the dovetail seams are in themselves a secondary structure motif and do not increase the topological complexity of the structure. Additionally, as discussed in section 3.3, only certain spacings are allowed. For AE crossovers, the dovetail seams can have spacings of

$$s = -5, -4, -3, -2, -1, 0, 1, 2.$$

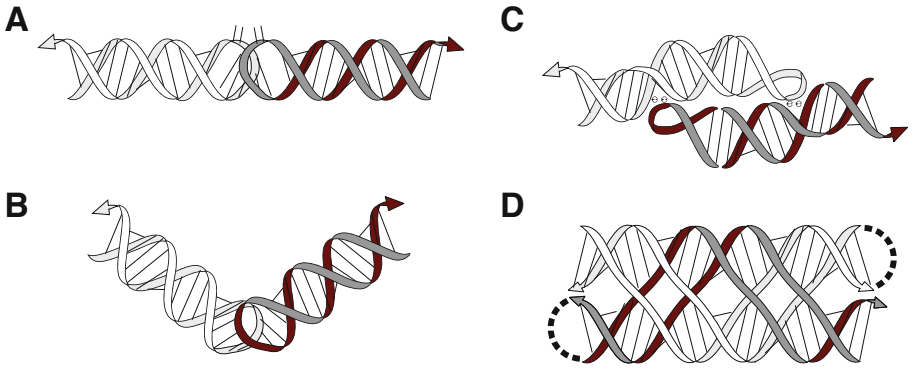

Fig. 5. Examples of paranemic interactions. (A) 180° kissing-loop interaction from HIV-1 [8]. **(B)** 120° kissing-loop motif [19]. **(C)** Loop-receptor interaction [4,15,10]. **(D)** Paranemic crossover motif PX [1].

The other DX types have similar allowable dovetail seam spacings, although we will focus on AE crossovers here as it is the simplest example.

4.3 Folding Path Design

The different types of DX lattices can all be used as the basis of designing single-stranded folding paths, simply by inserting loop-loop interactions as shown in Fig. 4. The AE crossover pattern is especially easy to design since it has a regular folding path and the curvature can be precisely controlled by the choice of dovetail spacings. As illustrated in Fig. 6, an 8 helix tall structure can be expanded in two-dimensions by utilizing multiple dovetail seams, and through the choice of the dovetail spacings on each row can be programmed to be relatively flat. Long helices can potentially cause topological problems, as they require a complement strand that wraps around the helix by more than a full turn. This problem can be circumvented by careful strand-path design to minimize the number of long helices, placing the long helices where they will be less likely to generate a trap, and by choosing the location of the 3' end so that it wraps less than one helical turn at the end of the folding proces. As a general rule we find that it is important that long helices are completely formed before pseudoknotted long-range interactions form connections that block access to the helix. In analogy to tying a knot, we wish to avoid cases where the end of the rope needs to be threaded back through the structure, and thus knots should be formed only in the bight of the rope. Even more complicated strand paths can be designed (Fig. 6), and in these cases the choice of strand path and position of the 5' and 3' ends becomes an even more important factor.

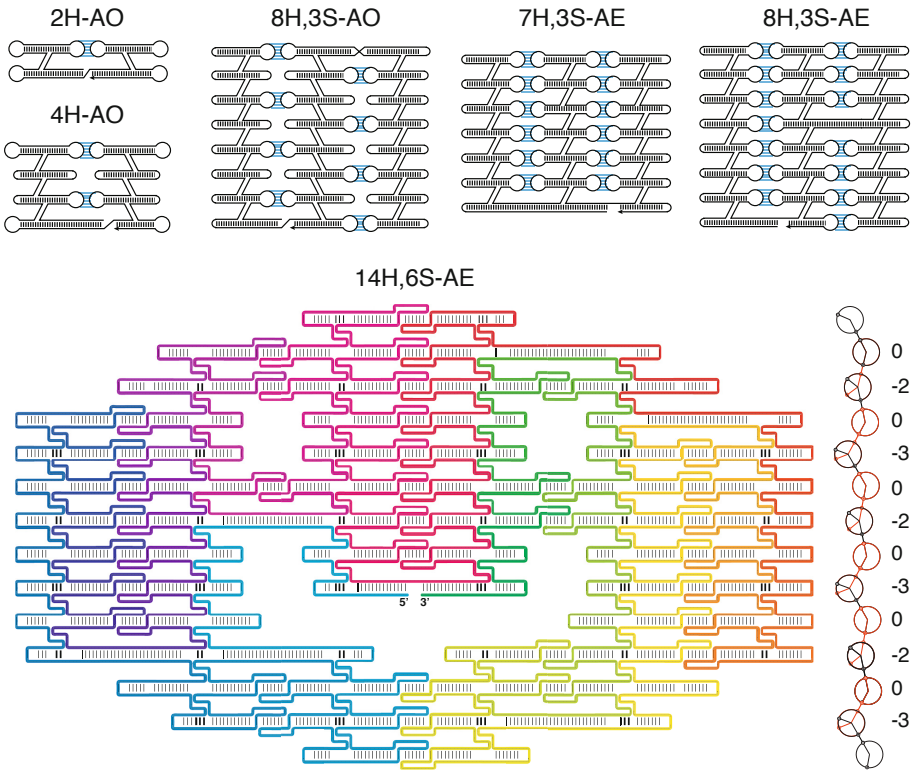

Fig. 6. Folding paths of single-stranded RNA origami structures. Top: Various strand paths for RNA origami in different size ranges, to illustrate how AO and AE junctions might be incorporated into arbitrary designs. The designs are named by the number of helices tall (H), the number of dovetail seams wide (S), and the type of DX used (AO or AE).AO strand paths are more intertwined than AE strand paths, and thus for larger designs AE patterns might be preferable. **Bottom:** An example of a single-stranded RNA origami with a complicated arbitrary strand path, here inspired from the DNA origami smiley face [22]. The dovetail junctions must have the same spacing within any given row to minimize strain in the structure. A particularly flat combination of spacings for AE is $s = 0, -2, 0, -3$, shown in side-profile at the right.5' and 3' ends are found near the middle of the structure (on the upper lip). The strand path has been colorized to highlight the order of synthesis.

Depending on the choice of the loop-loop placement within the strand path, the locality of interactions can either be decreased or increased. We choose to design structures by maximizing the locality of long-range interactions, with the goal of helping the kinetic folding of loops and their cognate partners. Following this logic, the shorter time between the expression of a loop and the expression of its partner loop, the less likely that these loops are to bind in the wrong place, especially if degenerate loop sequences are used. In terms of co-transcriptional folding where the 5' end is made first by the RNA polymerase and has time to fold before downstream sequence is produced, we believe this design principle can help to narrow choices in strand path design. In particular, locality seems to be beneficial for two main reasons: 1) the structural core forms faster and new elements can be added gradually, 2) loop-loop interaction sequences can be reused if the former loop is already base paired in the structure. The timing of formation of loop-loop interactions can be further programmed by choosing loop-loop interactions with different association-dissociation constants. Additionally, it may be possible to even further tune the folding pathway by implementing pausing sites for the RNA polymerase [18], by which a short 16 nucleotide sequence can be programmed into the DNA that causes the polymerase to pause for long enough to allow time for one folding event to complete before the next structure is synthesized.

4.4 Space-Filling Folding Paths

We have found inspiration in space-filling curves from mathematics for designing RNA folding paths with high locality. Some space-filling curves have similar strand paths to the structures shown in Fig. 6. In particular, the Peano curve has a pattern (Fig. 7B, top) that can be interpreted as RNA helices with dovetail seams and kissing-loop interactions (Fig. 7B, bottom). This strand path differs from natural RNAs in that sequences are left unpaired and that the 5' and 3' ends do not meet, where most natural RNAs fold into structures resulting in the two ends co-localized [29]. The Peano-inspired RNA structure also has some long stems internal to the structure that could form potential topological problems. However, these topological barriers can be bypassed with the ability to control the speed of transcription, pausing sites, and strength of kissing-loop interactions. We have highlighted all of the barrier-forming kissing loop interactions in orange and indicated the point in the sequence (shown as orange dots) in which the 2D structure is required to fold up to before the kissing loops form. Additionally, variations of the repeating pattern in the Peano-curve with either fewer or more kissing-loop interactions are expected to have an effect on the fold (Fig. 7A). We present a second Peano-inspired curve (Fig. 7C, top) and its translation into RNA secondary structure (Fig. 7C, bottom), to further illustrate the large variety of possible space-filling patterns that can be translated into RNA strand path patterns.

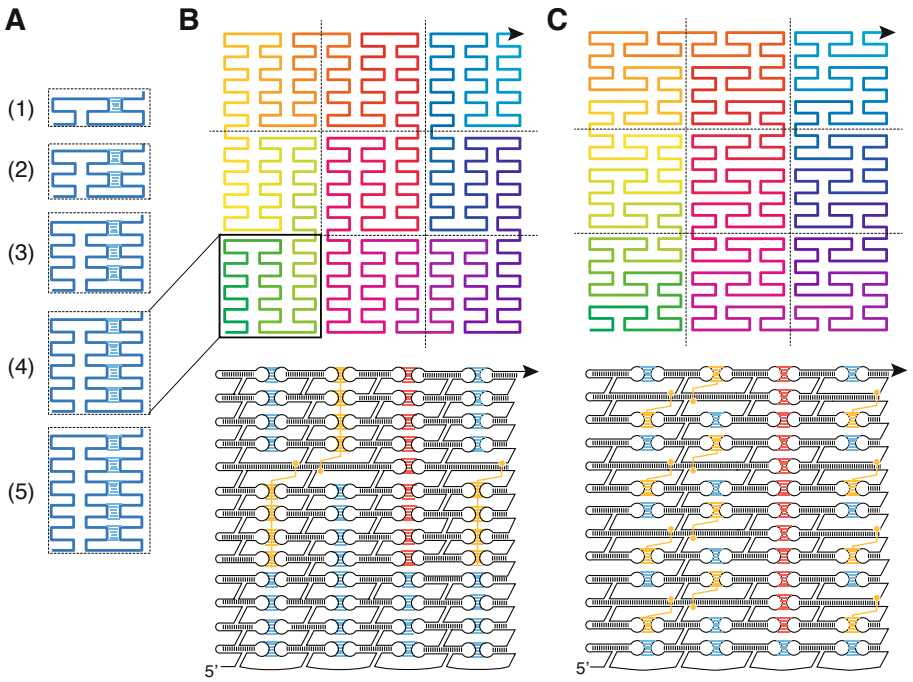

Fig. 7. Space-filling curves and their translation into RNA folding paths. (A) A set of simple repeating strand path patterns that can be tiled to form a Peano-like curve. **(B, top)** RNA Peano folding path based on pattern (4). The strand is colored to illustrate the strand directionality. Dotted lines indicate repetitions of the tile pattern. **(B, bottom)** The translation of the Peano curve into an RNA secondary structure. Colored lines indicate kissing loop interactions and their order of assembly, where blue forms first, orange second, and red last. Orange lines indicate kissing-loop interactions that have a requirement of slow folding, such that they must fold only after the 2D structure has formed up to the indicated orange dot. Red lines indicate a long seam of kissing-loop interactions. **(C)** A second RNA Peano-inspired folding path.

4.5 Rolling Circle Transcription of a Space-Filling RNA Structure

Rolling circle transcription (RCT) has recently been used to produced long transcripts of repeating RNA hairpin structures that further condense into sponge-like microspheres [20]. Likewise, rolling circle replication has been used to produce well-defined nanostructures out of DNA [21], although not nearly as large as has been demonstrated for RNA. Using the RNA origami architecture introduced in this paper it might be possible to produce well-ordered RNA structures by RCT. For example, an infinite 1D ribbon of RNA might be produced from only two repeat units (shown in light and dark blue, Fig. 8A). Such a ribbon shape would consist of alternating layers of kissing-loop and dovetail connections in a continuous 1D sheet.

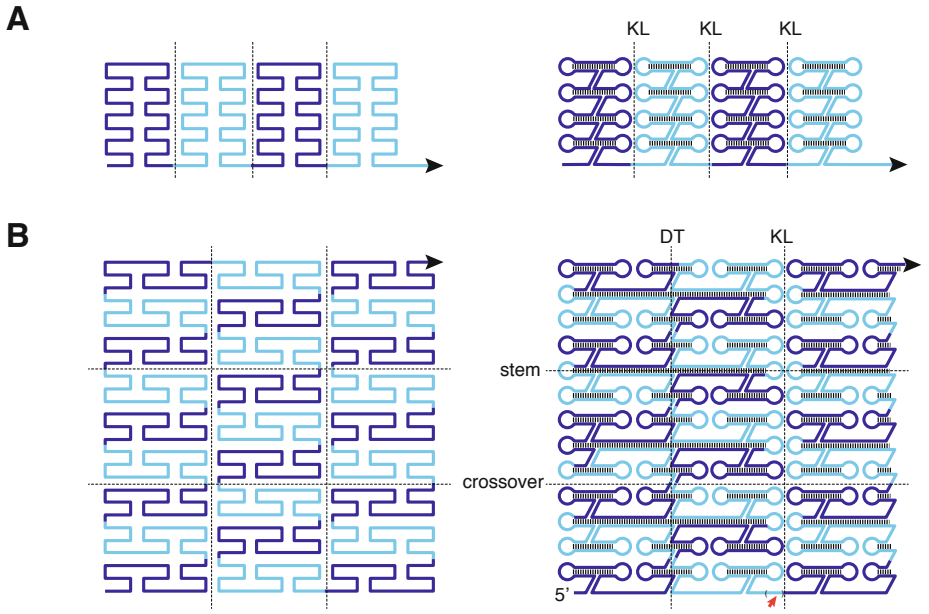

Fig. 8. Repeat units for 1D and 2D structure growth. (A) Curve and helix interpretation similar to the subpart of Fig. 7A with 4-helix repeat units shown in dark blue and cyan. Borders between repeat units are annotated as being composed of kissing-loop (KL) interactions. **(B)** Peano curve and helix interpretation similar to Fig. 7B with minimal repeating units shown in dark blue and cyan. The red arrow points to one position where a hairpin adopts a single-stranded form to fit the Peano path. Borders of the Peano curve are noted as being composed of either dovetail (DT), KL, stem or crossover interactions.

The space-filling curves of Peano inspire us to design repeating patterns of sequence and structure that could produce large and well-ordered space-filling structures by RCT. Taking the Peano curve as an example, we have made an attempt to define the sequence constraints for a tile that, when produced, will be able to adopt the infinite Peano-curve folding path while taking account of both 3D structure and topology. The basic repeating pattern is described in Fig. 8B, where just two RNA domains are required to produce the pattern. The structure is expected to raster back and forth, alternating between dovetail and kissing-loop seams. Several sequence constraints are imposed on the kissing-loop interactions and the dovetail seams, shown in more detail in Fig. 9A and 9B. Further illustrated in Fig. 9, some kissing loops form internally (denoted by different colors), while other have constraints because they are on the outside of the tile. One kissing loop and one dovetail are on the edge of the tile and are found to require base-pair palindromic symmetry in order to allow formation of the tile (marked in red).

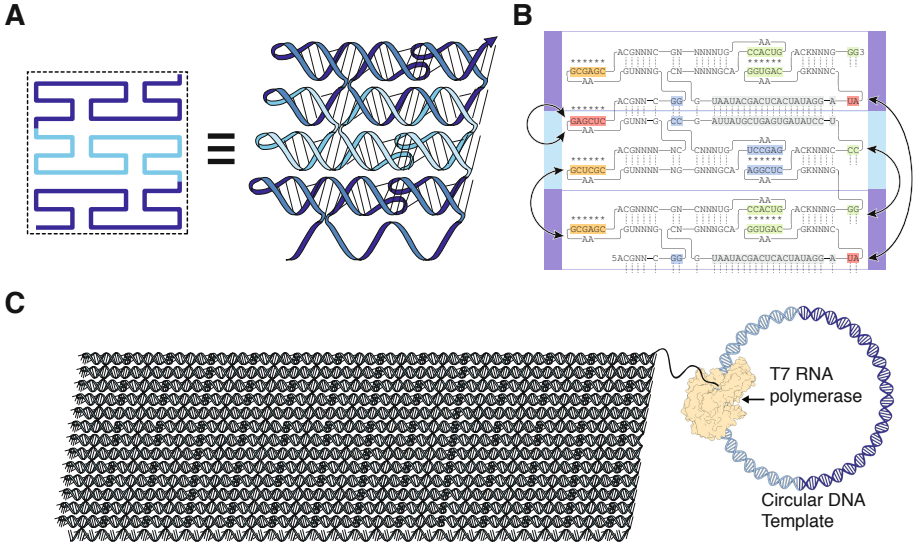

Fig. 9. Design of a circular RNA gene for producing a space-filling RNA structure. (A) The repeat units from Fig. 8B drawn as a ribbon diagram (similar to Fig. 2), with the repeating units shown in different colors. **(B)** Secondary structure with sequence constraints shown in different colors. Red indicate sequences that have to base pair with copies of itself and thus be base-pair palindromes. Sequences in orange, blue and green are orthogonal and pair according to color. Arrows indicate complementary domains. Grey sequence show the T7 promoter sequence and its complement, where GU mismatches have been introduced to disrupt the promoter signal in the opposite direction on the circular DNA template. **(C)** Ribbon diagram for an RNA structure produced from the circular DNA template gene by rolling circle transcription. T7 RNA polymerase and circular DNA template are shown to scale. Dark blue and cyan color of the circular DNA template indicate the two repeat domains.

We envision such a Peano tile may be encoded on a circular DNA template, which has an internal promoter sequence that is also encoded inside the RNA structure (positioned in the stem and marked in light grey). Fig. 9C shows the Peano-tile as it grows to larger structures. Even though it conceptually is tantalizing, this structure might not form for several reasons: First, the tile unit has many outward-facing interactions that might instead form internally in the tile, resulting in a misfold. Second, as mentioned before, the long stems may produce topological problems. Third, the assembly depends on the two-part tile unit to adopt an alternative conformation at key parts of the lattice. However, the correct interactions might still be preferred because, as the structure grows large, it may try to pack into its densest form, the desired assembly shape.

5 Discussion

We have introduced the A-form DX types and shown their extension to larger multi-crossover structures. The DX types have special properties concerning their bp spacing and internal symmetry. In the multi-crossover structures between several helices the DX types also have different properties concerning bending or flattening of the structures. As such, these motifs are important to consider when building extended architectures in 2D or 3D. This is especially important when designing RNA structures to be folded from a single strand, where geometry of crossovers and topological interactions have to be considered. We highlighted several paranemic assembly motifs well suited for making non-topologically linked interactions. We have demonstrated a methodology for folding path design using kissing-loop interactions and dovetail seams, which could easily be replaced with other tertiary motifs, of which there are plenty of good examples to choose from in the literature. To realize the design of complex structures produced by cotranscriptional folding, several theoretical aspects have to be further developed: 1) Sequence design taking account of pseudoknots, 2) The development of a kinetic model of folding, 3) Coarse grained modeling of the cotranscriptional folding process, and finally, 4) A theory for tile assembly when produced on a single strand.

References

1. Afonin, K.A., Cieply, D.J., Leontis, N.B.: Specific RNA self-assembly with minimal paranemic motifs. Journal of the American Chemical Society 130, 93–102 (2008)
2. Andersen, E.S.: Prediction and design of DNA and RNA structures. New Biotechnology 27, 184–193 (2010)
3. Arnott, S., Hukins, D.W., Dover, S.D., Fuller, W., Hodgson, A.R.: Structures of synthetic polynucleotides in the A-RNA and A'-RNA conformations: x-ray diffraction analyses of the molecular conformations of polyadenylic acid–polyuridylic acid and polyinosinic acid–polycytidylic acid. Journal of Molecular Biology 81, 107–122 (1973)
4. Costa, M., Michel, F.: Rules for RNA recognition of GNRA tetraloops deduced by in vitro selection: comparison with in vivo evolution. The EMBO Journal 16, 3289–3302 (1997)
5. Delebecque, C.J., Lindner, A.B., Silver, P.A., Aldaye, F.A.: Organization of intracellular reactions with rationally designed RNA assemblies. Science 333, 470–474 (2011)
6. Dickerson, R.E., Drew, H.R., Conner, B.N., Wing, R.M., Fratini, A.V., Kopka, M.L.: The anatomy of A-, B-, and Z-DNA. Science 216, 475–485 (1982)
7. Drew, H.R., Wing, R.M., Takano, T., Broka, C., Tanaka, S., Itakura, K., Dickerson, R.E.: Structure of a B-DNA dodecamer: conformation and dynamics. Proceedings of the National Academy of Sciences of the United States of America 78, 2179–2183 (1981)
8. Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C., Dumas, P.: Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. Nat. Struct. Biol. 12, 1064–1068 (2001)

9. Fu, T.J., Seeman, N.C.: DNA double-crossover molecules. Biochemistry 32, 3211–3220 (1993)
10. Geary, C., Baudrey, S., Jaeger, L.: Comprehensive features of natural and in vitro selected GNRA tetraloop-binding receptors. Nucleic Acids Research 36, 1138–1152 (2008)
11. Geary, C., Chworos, A., Jaeger, L.: Promoting RNA helical stacking via A-minor junctions. Nucleic Acids Research 39, 1066–1080 (2011)
12. Geary, C.W., Rothemund, P.W.K., Andersen, E.S.: A single-stranded architecture for cotranscriptionally folded RNA tiles. Accepted in Science (2014)
13. Grabow, W.W., Jaeger, L.: RNA Self-Assembly and RNA Nanotechnology. Accounts of Chemical Research (2014)
14. Hao, C., Li, X., Tian, C., Jiang, W., Wang, G., Mao, C.: Construction of RNA nanocages by re-engineering the packaging RNA of Phi29 bacteriophage. Nature Communications 5, 3890 (2014)
15. Jaeger, L., Westhof, E., Leontis, N.B.: TectoRNA: modular assembly units for the construction of RNA nano-objects. Nucleic Acids Research 29, 455–463 (2001)
16. Jossinet, F., Ludwig, T.E., Westhof, E.: Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. Bioinformatics 26, 2057–2059 (2010)
17. Ko, S.H., et al.: Synergistic self-assembly of RNA and DNA molecules. Nature Chemistry 2, 1050–1055 (2010)
18. Larson, M.H., et al.: A pause sequence enriched at translation start sites drives transcription dynamics in vivo. Science 344, 1042 (2014)
19. Lee, A.J., Crothers, D.M.: The solution structure of an RNA loop-loop complex: the ColE1 inverted loop sequence. Structure 6, 993–1007 (1998)
20. Lee, J.B., Hong, J., Bonner, D.K., Poon, Z., Hammond, P.T.: Self-assembled RNA interference microsponges for efficient siRNA delivery. Nature Materials 11, 316–322 (2012)
21. Lin, C., Wang, X., Liu, Y., Seeman, N.C., Yan, H.: Rolling circle enzymatic replication of a complex multi-crossover DNA nanostructure. Journal of the American Chemical Society 129, 14475–14481 (2007)
22. Rothemund, P.W.: Folding DNA to create nanoscale shapes and patterns. Nature 440, 297–302 (2006)
23. Severcan, I., Geary, C., Chworos, A., Voss, N., Jacovetty, E., Jaeger, L.: A polyhedron made of tRNAs. Nature Chemistry 2, 772–779 (2010)
24. Sherman, W.B., Seeman, N.C.: Design of minimally strained nucleic Acid nanotubes. Biophysical Journal 90, 4546–4557 (2006)
25. Shih, W.M., Quispe, J.D., Joyce, G.F.: A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. Nature 427, 618–621 (2004)
26. Wang, J.C.: Helical repeat of DNA in solution. Proceedings of the National Academy of Sciences of the United States of America 76, 200–203 (1979)
27. Woo, S., Rothemund, P.W.: Programmable molecular recognition based on the geometry of DNA nanostructures. Nature Chemistry 3, 620–627 (2011)
28. Yan, H., Zhang, X., Shen, Z., Seeman, N.C.: A robust DNA mechanical device controlled by hybridization topology. Nature 415, 62–65 (2002)
29. Yoffe, A.M., Prinsen, P., Gelbart, W.M., Ben-Shaul, A.: The ends of a large RNA molecule are necessarily close. Nucleic Acids Research 39, 292–299 (2011)