THE UNIVERSITY OF CHICAGO

SYNTHETIC NUCLEIC ACID TOPOLOGY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY
DI LIU

CHICAGO, ILLINOIS
DECEMBER 2016

This thesis is dedicated to my parents.

TABLE OF CONTENTS

List of Figures

List of Tables

ACKNOWLEDGEMENTS

First, I would like to express my gratitude to my advisor, Prof. Yossi Weizmann, for the continuous support through my Ph.D. study. Yossi has allowed me the extraordinarily high levels of freedom to conceive ideas and design experiments according to my own interest. He also provides me the precious opportunities to attend conferences, to learn new techniques, to establish collaborations, and to practice negotiating with editors and reviewers. Without such freedom and opportunities, which are very rare for the graduate students in other labs, it would be impossible for me to fully exercise my creativity and to conduct the work presented here.

I am grateful to Prof. Joseph A Piccirilli and Prof. Bozhi Tian for serving on my thesis committee. Prof. Piccirilli served on my Ph.D. candidacy examining committee before and provided me with great suggestions and encouragement. Later, we collaborated with his lab on the RNA work, including the RNA topology and crystallization of RNA nanostructures. Prof. Piccirilli has been an outstanding scientist in RNA research with keen insights. Talking with him is always joyous and stimulating. I also admire his wonderful skills of clear and effective scientific communication, which I hope to obtain one day. Prof. Tian has provided me great guidance in the past years. During the submission of our DNA topology paper, he taught me a lot in writing response to reviewers. He is also very generous to give me advices in my career plan.

I must thank all the researchers in the DNA nanotechnology community. We are proud to build such a harmonious and prosperous community, where collaborations are greatly facilitated and ideas are freely expressed. Presenting unpublished work in conference, which is unimaginable in most other fields, is normal in our field. I am glad to be one member in this community! I want to thank Prof. Chengde Mao (Purdue University) for his kind suggestion and support over the years. I also want to thank Prof. Hao Yan (Arizona State University) and the members in his lab for great discussion and encouragement. Our RNA work (which accounts the largest portion of my research) would be impossible without Dr. Cody Geary (Caltech). I am very luck to make this friend in the field. Cody was very generous to teach me how to play with RNA and has been willing to share idea with me. He also built the first model of the bKL motif, which has greatly facilitated my research.

Of course, my research is not possible without the help from the Weizmann group members, collaborators and staff members. In particular, I want to thank Dr. Gang Chen, Dr. Leilei Tian, Timothy Cronin, Usman Ahkter, Ted Gannett, Di Wen (Ram Mahato group in University of Nebraska Medical Center), Dr. Chunbai He (Wenbin Lin group). Working with them has been very impressive and memorable experiences. I am very glad and fortunate to work with Dr. Yaming Shao (Piccirilli group) on our unprecedented and wonderful efforts to crystallize artificially designed RNA nanostructures and I have learned a lot of knowledge on biological experiment and crystallography from him. I am grateful to Prof. Yuk-Ching Tse-Dinh (Florida International University) for the enzymes and advices. Thanks to Jonathan Remis (Northwestern University) for teaching me the CryoEM technique, and Dr. Qiti Guo, Dr. Justin Jureller, and Dr. Elena Solomaha for the facility training.

Special thanks go to Prof. Luping Yu, Prof. Jun Yin and Prof. Viresh H. Rawal for serving on my Ph.D. candidacy examining committee. I appreciated their patience and encouragement. I am honored to be supported by the Martha Ann and Joseph A. Chenicek Graduate Research Fund from the department in the third year and by the HHMI International Student Research Fellowship in the fourth and fifth years.

I must also thank all the friends I have made in Chicago. Out of study and research, we have much fun together. Without them, my days would be tedious and colorless in the past few years.

Last, but most importantly, I want to thank my parents – my father Yimin Liu and my mother Xinhua Yu – for their unconditional support and endless love all the time. I also appreciate the support and care provided by our relatives to my family.

ABSTRACT

Knots and links are ubiquitous and constantly provide inspiration in all spheres of human endeavor. For chemists, the synthesis of entangled and interlocked molecules represents a significant challenge, but also an excitingly grand pursuit. We have been working on construction of topological nucleic acid structures using the methodologies and principles developed in the field of structural DNA nanotechnology. According to the rigidity of the systems, different aspects should be considered. Specifically, nucleic acid topological structures have been created at three different levels with different nucleic acid structures. Level 1: Single-stranded (ss) structures. Topological construction with the highly flexible ss nucleic acids basically deals with the problem of how to generate nodes with the desired handedness and in the correct location. To achieve this, we have developed a more general and controllable strategy based on the stacked X-structure of nucleic acid four-way junctions (4WJs) by realizing that the two helical strands within a 4WJ form a node. The handedness of nodes can be readily controlled by manipulating the 4WJs within geometrical constraints, such as forming the tensegrity triangles. Level 2: Double-stranded (ds) structures. At this level, aside from the node formation, embodiment of topology also necessitates the consideration of 3D geometry. Therefore, curvature and torsion of the ds structures should be carefully designed when ss structures are converted to the double-stranded version. Level 3: Self-assemble structures. Self-assembled nucleic acid nanostructures can be rigid enough for enabling the topological construction solely via geometrical control. We have designed a T-shaped RNA branched kissing loop (bKL) structural motif, in which a hairpin loop and a bulge form programmable Watson-Crick base-pairings. Tiles based on this bKL motif permit the control over curvature and torsion of the

assemblies. This ultimately allows us to design tiles which can assemble into topologically complex nanostructures.

DNA topology is a prominent and fundamental theme in modern biology, and largely defines the structural, biological and functional principles of DNA and most DNA-processing enzymes. Synthetic DNA topological structures can provide invaluable tools to reveal how the problems of DNA topology are tackled in living cells. Meanwhile, the existence of naturally occurring RNA structures with nontrivial topologies remains an enigma. Synthetic RNA topological structures are significant for understanding the physical and biological properties pertaining to RNA topology, and these structures in turn could facilitate identifying naturally occurring knotted or interlocked RNAs. Indeed, RNA topology is an important but unfortunately neglected subject in current research of RNA biology.

While constructing self-assembled topological structures (level 3), we have also attempted to study some artificially designed RNA nanostructures with single-crystal X-ray diffraction. The initial goal of these efforts was to obtain high-resolution 3D structural information of the new motif bKL, but eventually we realize that our approach may be developed into a geometry-promoted strategy to overcome the challenging RNA crystallization problem.

All in all, our studies on synthetic nucleic acid topological structures have yielded not only aesthetically appealing entities, but also practical tools and strategies for solving fundamental questions in biology.

Chapter 1 – Introduction

The elegant double helical structure of DNA[1] is one of the most significant scientific discoveries of modern science, and also becomes, without any exaggeration, one of the icons of human civilization. The impact of double helical DNA structure has been immense. There is no doubt that it represented the turning point for a new era of modern biology by suggesting the mechanisms of the storage and transmission of genetic information. We should note that there are two key words in the term “double helical structure” – “double” and “helical”. More attention has been paid to the word “double” than the word “helical”. This is understandable due to the fact that DNA being “double” seems to bear more biological relevance. If the double-stranded DNA molecule adopted a “non-helical” or “side-by-side” structure (which was indeed suggested and supported by some scientists years ago[2]), most of the fundamentals of molecular biology would not have been much different. In fact, DNA has been always illustrated (understandably, for simplicity) as two antiparallel side-by-side strands in textbooks and literature. Nonetheless, the DNA structure being “helical” is as important as, if not more than, being “double” for other areas of research. My graduate research has been mostly focused on two of these areas – structural DNA nanotechnology[3] and DNA topology[4].

1.1 From structural DNA nanotechnology to digital biotechnology

In the field of structural DNA nanotechnology[3], DNA is utilized as building blocks to create complex materials. The extraordinarily precise spatial control of matter with DNA is due to its well-defined geometric features (helical parameters) and highly predictable recognition properties (Watson-Crick base pairings or complementarity), both of which together are regarded as the programmability of DNA. The field was initiated by Nadrian Seeman’s theoretical paper in 1982[5], where he suggested constructing crystalline lattices with DNA multiway junctions to host guest macromolecules (such as proteins) for structural determination by single-crystal X-ray crystallography. During the beginning years, Seeman and his coworkers were almost alone in the field, but laid the very foundation for the field. Their proposed principles regarding sequence and motif design are still the gospels in the field today.

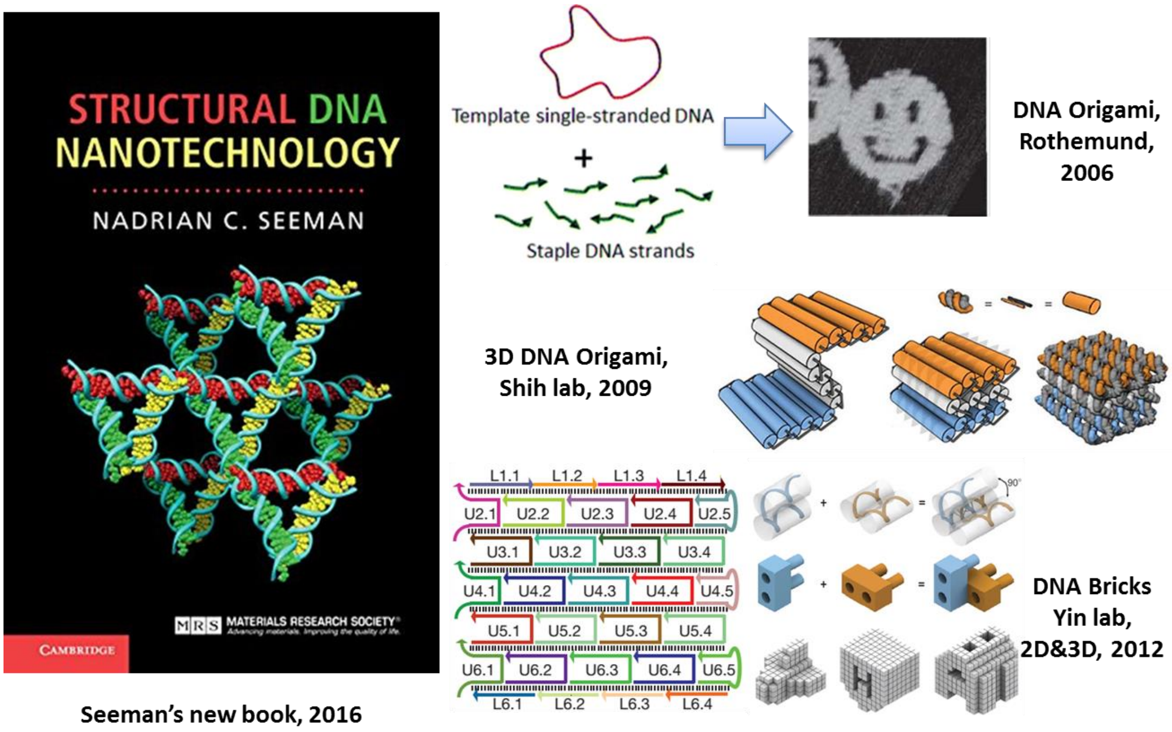

Figure 1.1 Structural DNA nanotechnology. Seeman's book *Structural DNA Nanotechnology*[3], which provides an overview and development of the field, is shown on the left. The techniques of DNA origami[6,7] and DNA bricks[8,9] are shown on the right.

The most important boost in the field is the invention of DNA origami by Paul Rothemund in his 2006's Nature paper[6], where he show that a long single-stranded (ss) viral DNA (~7000 nt long for M13mp18 phage), serving as a "scaffold", can be folded into arbitrary shapes and patterns by hundreds of designed short synthetic oligonucleotides (called "staples"). The invention of DNA origami technique has two major far-reaching outcomes. On the one hand, it revealed the extraordinarily high complexity that can be enabled by DNA self-assembly. This eventually led to the advent of a kaleidoscope of new techniques derived or at least inspired by DNA origami to create nanostructures with ever increasing complexity. Variants of DNA origami, include, but are not limited to, hollow 3D structures with 2D walls[10-12], solid 3D structures with helices arranged in honeycomb-[7], square-[13] and hexagonal[14]-lattices, and various gridiron[15] and wireframe[16-18] structures. Another scaffold-free approach is the “DNA brick”, where 2D[8] or 3D[9] shapes are assembled solely with short ssDNA tiles. Though seemingly different (regarding the using of scaffolds), the design principles of DNA origami and DNA bricks are essentially similar. On the other hand, professionals from other disciplines began to pay attention to this field. Structural DNA nanotechnology enables the nanometer-precise spatial control of other molecules or components, and a lot of otherwise unimaginable feats have been achieved as a result. Catalyzing the process is the fact that there are a lot of free and easy-to-use software, such as caDNAno[19], to aid the design process. Therefore, even “non-experts” nowadays can design complex DNA origami structures at their disposal and obtain the designed structures within a few days.

Besides providing the methods of designing and constructing complex nanoarchitectures, structural DNA nanotechnology also represents the prototype of the concept of digital biotechnology. Again, programmability is the key word for this concept. Like the real-world computer program, techniques belonging to digital biotechnology should be able to generate foreseeable results with high success rate based on encoded commands and information. Another important component for the concept of digital biotechnology is the relatively small dependence on professional expertise and specialized (and expensive in most cases) instruments. These can facilitate the wide-spreading and the concomitant further development of the techniques. In retrospect, several previously developed techniques also meet the standards, and these include PCR, RNA interference (RNAi), and the more recent CRISPR-Cas9 genome editing. Interestingly, all of these are based on the programmability of DNA or its chemical cousin RNA. In addition, the field of structural DNA nanotechnology has also evolved into the stage when we are beginning to witness new techniques evolving from the field, such as toehold exchange hybridization probes[20], DNA-PAINT[21], and toehold riboswitch[22]. I wish the research that will be presented in this thesis can also contribute to developing the broad concept of techniques of digital biotechnology.

1.2 Basics of topology

Topological problem is one important consequence of the double helical structure of DNA and also the keynote of this thesis. Therefore, it is necessary to introduce here some basic concepts and terminologies in topology (knot theory in particular) in advance. It should be noted that topology is a vast subject that is mostly out of the scope of this thesis. This section only scratches the surface of this subject in order to provide enough background information for the discussion in the following chapters.

Topology is a mathematical discipline that studies the invariant properties of shapes under continuous deformations, such as stretching and bending. There is a famous joke about topologist – someone who cannot distinguish between a donut and a coffee cup. Most biomacromolecules, such as nucleic acids and proteins, can be abstracted into a string, and the topological problems of them naturally fall into the category of knot theory.

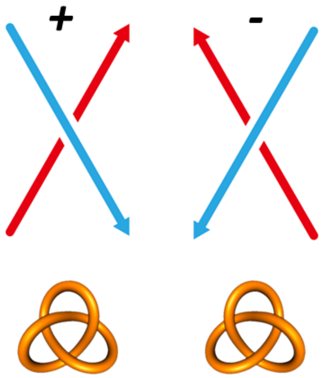

Figure 1.2 Chiralities of nodes and trefoil knots.

The simplest nontrivial knot is a trefoil knot, which contains three nodes. Actually, by deforming a trefoil knot, one can obtain more nodes, but three is the minimum number of nodes, which is an invariant for a knot. Therefore, a 2D representation of the knot containing the minimum number of nodes is called “minimum plane projection”. In fact, there are two distinct trefoil knots, which are not mutually interconverted by continuous deformations. Close inspection reveals that they are of different chiralities and the nodes within are of different handednesses (Figure 1.2). The node convention also applies for links, which are composed of multiple closed components. Links are always called “catenane” by chemists.

The most commonly used nomenclature of knots and links is the Alexander-Briggs notation n_i or n_i^C, where n is the minimal number of nodes, C is the number of components (for links), and i distinguishes different topologies with the same n and C. Unlike knots and links with small number of nodes, there can be multiple distinct knots with the same number of nodes for more complex knots and links. Readers are referred to *Knots and Links*[23] by Dale Rolfsen for a more complete table of knots and links. Here, I only show the ones that we will encounter through the thesis (Figure 1.3). Some very famous topologies also possess more common names (or names dubbed by chemists), which are denoted as well. Some topologies can be considered as the connection sum of two or more nontrivial topologies, and are called composite knots or links. A hash (#) is put between the component topologies in their Alexander-Briggs notations.

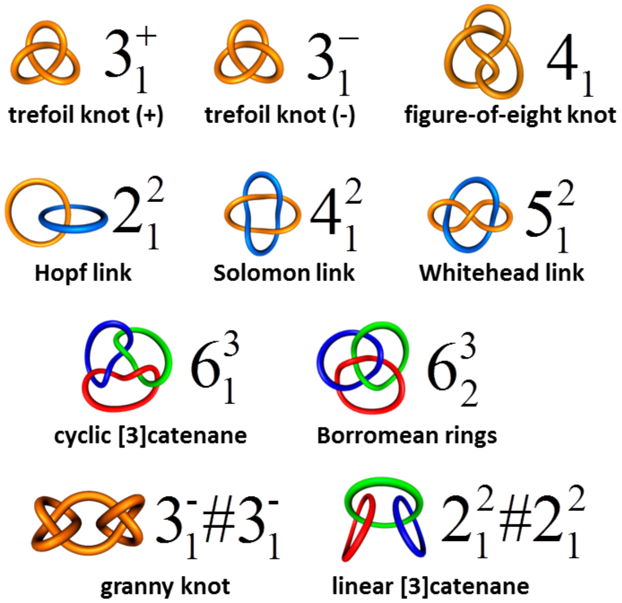

Figure 1.3 Representative topologies and their Alexander-Briggs notations.

1.3 Scope of thesis

My graduate research mainly focused on the construction of topological structures with nucleic acids. Specifically, I have been working on developing different strategies for topological construction based on the rigidity of three different systems – single-stranded, double -stranded, or self-assembled structures. During constructing these unusual structures or during studying them, there are a lot of unexplored aspects of nucleic acids discovered, as will be discussed in the following chapters.

Chapter 2 presents the junction-based method to create ssDNA topological structures. A total of nine distinct nontrivial topologies and four series of topoisomers are constructed with ssDNA.

The ssDNA topological structures, serving as templates, are also converted to the corresponding dsDNA versions. We further investigate two important enzymatic processes *in vitro* using a constructed ssDNA knot – the topological conversion catalyzed by DNA topoisomerase and the DNA replication under topological constraints.

In Chapter 3, topological structures containing single-stranded RNA (ssRNA) are created either by configuring RNA-DNA hybrid four-way junctions or via DNA-templated synthesis. We then demonstrate the use of a constructed ssRNA knot as a model to probe the RNA topoisomerase activity and to obtain the topology-dependent RT-PCR patterns.

Chapter 4 describes our control over the curvature and torsion of RNA assemblies via an artificially designed motif formed by the kissing interactions between a loop and a bulge. This is part of our attempts to create self-assembled topological structures. Our current efforts to crystallize self-assembled RNA nanostructures are included as well.

The thesis concludes with Chapter 5, where the importance of topology in nucleic acid research and in the broader scientific research is discussed. Several future opportunities of structural DNA nanotechnology (relating to my research but not discussed in chapters 2-4) are tentatively provided.

1.4 References

1. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 1953, **171**(4356)**:** 737-738.
2. Rodley, G. A., Scobie, R. S., Bates, R. H. & Lewitt, R. M. A possible conformation for double-stranded polynucleotides. *Proc. Natl. Acad. Sci. USA* 1976, **73**(9)**:** 2959-2963.
3. Seeman, N. C. *Structural DNA Nanotechnology*. Cambridge University Press, 2016.
4. Bates, A. D. & Maxwell, A. *DNA topology*, 2 edn. Oxford Univ. Press: Oxford, 2005.
5. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* 1982, **99**(2)**:** 237-247.
6. Rothemund, P. W. Folding DNA to create nanoscale shapes and patterns. *Nature* 2006, **440**(7082)**:** 297-302.

7. Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* 2009, **459**(7245)**:** 414-418.
8. Wei, B., Dai, M. & Yin, P. Complex shapes self-assembled from single-stranded DNA tiles. *Nature* 2012, **485**(7400)**:** 623-626.
9. Ke, Y., Ong, L. L., Shih, W. M. & Yin, P. Three-dimensional structures self-assembled from DNA bricks. *Science* 2012, **338**(6111)**:** 1177-1183.
10. Andersen, E. S. *et al.* Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* 2009, **459**(7243)**:** 73-76.
11. Ke, Y. *et al.* Scaffolded DNA origami of a DNA tetrahedron molecular container. *Nano Lett* 2009, **9**(6)**:** 2445-2447.
12. Han, D. *et al.* DNA origami with complex curvatures in three-dimensional space. *Science* 2011, **332**(6027)**:** 342-346.
13. Ke, Y. *et al.* Multilayer DNA origami packed on a square lattice. *J. Am. Chem. Soc.* 2009, **131**(43)**:** 15903-15908.
14. Ke, Y., Voigt, N. V., Gothelf, K. V. & Shih, W. M. Multilayer DNA origami packed on hexagonal and hybrid lattices. *J. Am. Chem. Soc.* 2012, **134**(3)**:** 1770-1774.
15. Han, D. *et al.* DNA gridiron nanostructures based on four-arm junctions. *Science* 2013, **339**(6126)**:** 1412-1415.
16. Benson, E. *et al.* DNA rendering of polyhedral meshes at the nanoscale. *Nature* 2015, **523**(7561)**:** 441-444.
17. Zhang, F. *et al.* Complex wireframe DNA origami nanostructures with multi-arm junction vertices. *Nature Nanotechnol.* 2015, **10**(9)**:** 779-784.
18. Veneziano, R. *et al.* Designer nanoscale DNA assemblies programmed from the top down. *Science* 2016, **352**(6293)**:** 1534.
19. Douglas, S. M. *et al.* Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res.* 2009, **37**(15)**:** 5001-5006.
20. Zhang, D. Y., Chen, S. X. & Yin, P. Optimizing the specificity of nucleic acid hybridization. *Nature Chem.* 2012, **4**(3)**:** 208-214.
21. Jungmann, R. *et al.* Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nat Methods* 2014, **11**(3)**:** 313-318.
22. Green, A. A., Silver, P. A., Collins, J. J. & Yin, P. Toehold switches: de-novo-designed regulators of gene expression. *Cell* 2014, **159**(4)**:** 925-939.
23. Rolfsen, D. *Knots and Links*. Publish or Perish, Inc: Berkeley, CA, 1976.

Chapter 2 – Synthesizing DNA topological structures by configuring DNA four-way junctions

2.1 Introduction

The synthesis of topologically interesting molecules is very challenging both intellectually and technically[1-3]. In order to successfully construct topological structures at nanometer-scale dimensions, an extraordinarily high level of the manipulation of molecular recognitions is required for the controllable generation of nodes (or crossovers) with the correct handedness (positive or negative) and in the precise position (how they are linked). As has been introduced in Chapter 1, DNA is a very promising molecule for this purpose[4]. Meanwhile, DNA topological structures are also very desirable because DNA topology plays a profound role in major cellular processes in the living cells, such as DNA replication, transcription and recombination[5]. Synthetic DNA topology may help elucidate these processes.

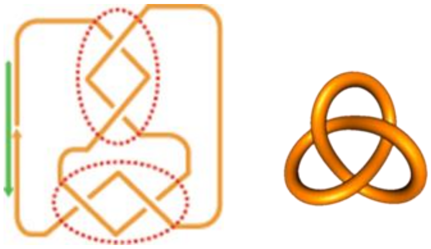

Figure 2.1 Helix-based method to construct a DNA trefoil knot[6]. The two helical regions are highlighted within red dashed ovals.

In 1991, Seeman produced the first synthetic DNA topological structures – a ssDNA trefoil knot, by ligating a ssDNA containing complementary pairing regions interspersed between poly(dT) tracts produced (Figure 2.1)[6]. This method is referred to as helix-based method in the thesis, because it utilizes the topological property of the dsDNA helix, a half turn (5 or 6 bp) of which produces to a node. The most common right-handed B-form helix readily generates negative nodes (Figure 2.2a)[6]. However, the access to positive nodes is not as convenient because the formation of left-handed DNA helix would require the less common Z-form DNA (Figure 2.2b)[7-9] or the unnatural L-DNA (containing L-nucleotides)[10]. This helix-based method has tremendously expanded the spectra of accessible molecular topologies[11]. Its adaptability is demonstrated by several elegant examples, including the realization of three knotted topologies using the same DNA sequence[8] and the construction of DNA Borromean rings[9], the latter of which predated by seven years the small molecule version with the metal-templation method by Stoddart group[12].

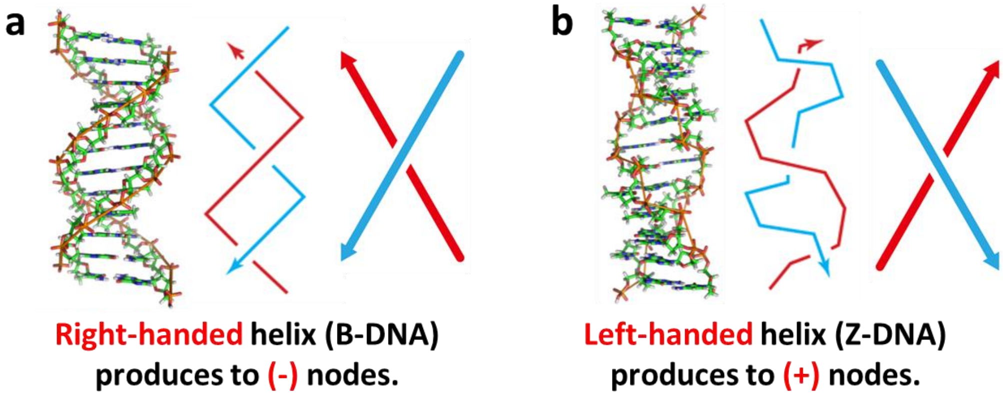

Figure 2.2 Nodes are produced by the DNA helices[7,8]. Right-handed B-DNA (**a**) and left-handed Z-DNA (**b**) helices produce the negative and positive nodes respectively.

However, this helix-based method has three obvious limitations. First, the resulting topological constructs have restricted applications due to the intrinsic strong base-pairings. Second, it is not convenient to generate positive nodes since the most common DNA duplex is B-form. Finally, due to the undesired braiding of the ssDNA linkers (even with the relatively unstructured sequence of poly(dT)) connecting the complementary pairing regions, a surplus of nodes is often produced. “Topological protection”[13] should be used in these cases, but it inevitably complicates the synthesis. Although the helix-based method has these drawbacks, almost all the synthetic DNA topological structures until now were constructed based on the principle of using helix to generate nodes, with the exception of a dsDNA catenane utilizing DNA-binding polyamides[14] and a reconfigurable Mobius strip assembled via the scaffolded DNA origami[15]. Consequently, the direction of DNA topological structures is relatively less developed than the geometrical structures. To solve these problems, we herein present a versatile design strategy based on the DNA four-way junction (4WJ)[16] with the goal of creating complex DNA topological structures. (The contents of this chapter is partially adapted from a published paper[17].)

2.2 Results and discussion

2.2.1 Structural properties of DNA 4WJ

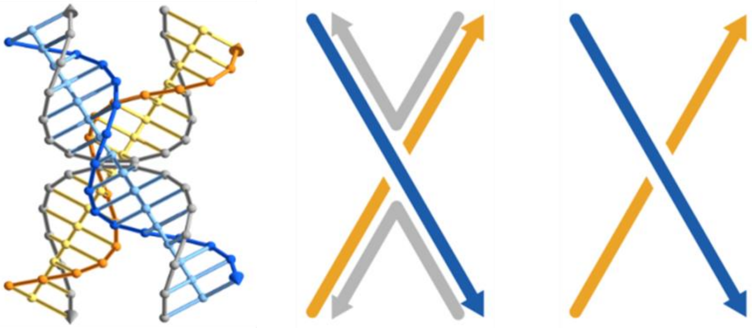

Figure 2.3 Junction-based strategy for topological construction. Helical (left) and schematic (middle) representations of a DNA four-way junction (4WJ) are shown with a node generated by the two helical strands (right).

The DNA 4WJ adopts an X-structure composed of four ssDNA strands that form two pairs of stacked helices (Figure 2.3)[18]. These four strands are grouped into two classes according to their continuity throughout the stacked helices: two helical strands are continuous and the other two crossover strands are exchanged. The fact that the two helical strands within the 4WJ form a node is the fundamental design principle for our junction-based method for topological construction. Individual free-standing 4WJ has been shown to have the two helical domains oriented antiparallel, producing a right-handed torsion angle of approximately 60 °. Therefore, a positive node will be obtained from a free-standing 4WJ.

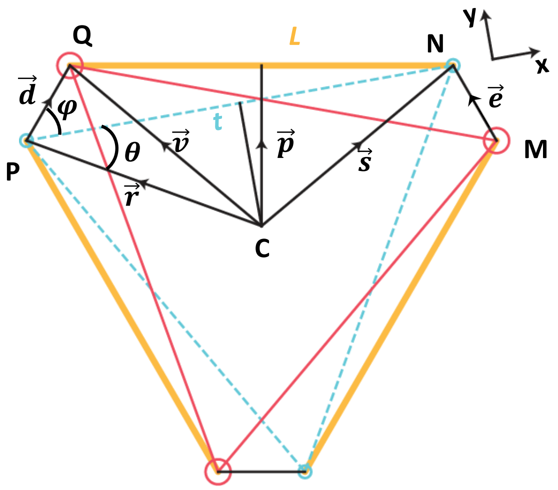

Figure 2.4 Geometry analysis of the tensegrity triangle[19]. The orange segments represent the axes of DNA duplexes forming the triangle. Note that the duplexes are not parallel to the paper, but with the larger red circles marking the points in the front and the smaller blue circles marking the points in the rear. Each black segment connecting a front point from one orange segment and a rear point from another represents the distance between these two orange segments in the 4WJ and thus is the common perpendicular of them.

It is important to realize that the torsion angle of the 4WJ is malleable and can be manipulated when geometrical constraints are applied. When 4WJs are connected into rigid structural motifs, the torsion angles are forced to accommodate the geometrical requirement for the maximization of base-pairing interactions. In the antiparallel[20] and parallel[21] double-crossover, the 4WJs have a torsion angle of 0 ° and 180 ° (approximately -60 ° and +120 ° distortion from the free-standing 4WJ). While the distortion is relative small for the 4WJs in the various right-handed tensegrity triangles[19,22,23], it is substantially large for the 14-bp-edged left-handed tensegrity triangles with a distortion of approximately -133 °[19]. In the “DNA Gridiron” structures[24], 4WJs of more variations of torsion angles are created. Due to malleability of their torsion angles, nodes with different signs can be readily obtained from the DNA 4WJs. Another important fact is that multiple nodes can be generated within one structural motif constructed by 4WJs. In this study, the structural motif tensegrity triangle is well exploited.

Table 2.1 Strain analysis of the tensegrity triangles with edges ranging from 13 to 18 bp.

Edge Length	13 bp	14 bp	15 bp	16 bp	17 bp	18 bp
Helix turning	445.7°	480.0°	514.3°	548.6°	582.9°	617.1°
2α	111.3°	113.9°	116.5°	118.7°	120.9°	123.0°
Geometric turning (-)	471.3°	473.9°	476.4°	478.7°	480.9°	483.0°
Strain (-)	0.057	-0.013	-0.074	-0.127	-0.175	-0.217
Geometric turning (+)	608.7°	606.1°	603.6°	601.3°	599.1°	597.0°
Strain (+)	0.366	0.263	0.174	0.096	0.028	-0.033
Handedness	-	-	-	+	+	+

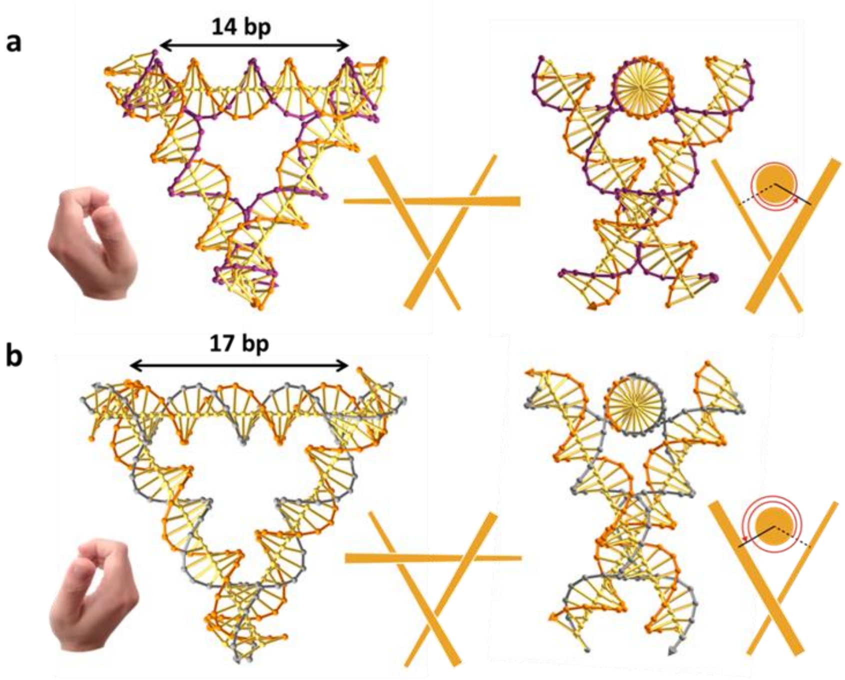

Figure 2.5 Schematic illustration of left- and right-handed tensegrity triangles. Staples are colored in magenta or gray in the left- (**a**) or right-handed (**b**) tensegrity triangle. Insets represent the 3-dimensional relationship of the dsDNA helices.

The geometry derivation and the strain analysis have been reported previously[19], we provide here its brief derivation for the convenience of our further discussion (Figure 2.4). The strain analysis is based on the difference between the geometric turning of the tensegrity triangle and the helical turning of the double helix. Because of the intrinsic chirality of helix of the B-form DNA, the geometric turning in this case is different for left-handed (-) and right-handed (+) tensegrity triangle. It is $n \times 360° + 2\alpha$ for left-handed and $n \times 360° - 2\alpha$ for right-handed, where n is the integer that makes the geometric turning most close to the helix turning. The strain analysis results for equilateral tensegrity triangles with edge length ranging from 13 bp to 18 bp are listed in Table 2.1. In this study, we have used the left-handed tensegrity triangle with 14-bp edges for the generation of negative nodes and the right-handed with 16 or 17-bp edges for positive nodes (Figure 2.5).

2.2.2 Three representative topological structures with the junction-based method

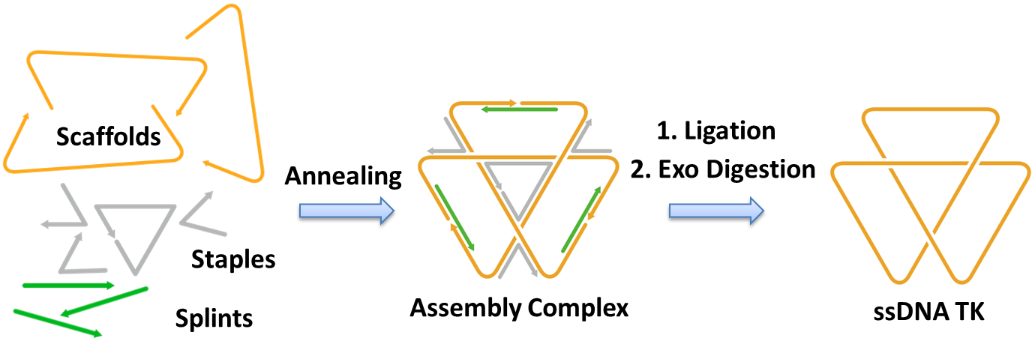

Figure 2.6 All-in-one protocol for the preparation of TK. All the strands are annealed in a one-pot fashion. The ssDNA knot **TK** is formed after ligation and exonuclease (Exo) digestion.

We first illustrate our concept in the construction of the simplest knot, a trefoil knot, **TK**. After an annealing process, the assembled complex is formed with three kinds of strands: scaffolds, staples and splints (Figure 2.6). Similar to the scaffolded DNA origami[25], the scaffolds are held together by the staples to form 4WJs and are connected end-to-end by the splints. T4 DNA ligase then fixes the topology by sealing the nicks in the scaffolds. After removal of all linear species by exonuclease digestion, the relaxed ssDNA topological constructs are generated without intended intramolecular base-pairings. The formation of **TK** is shown by denaturing polyacrylamide gel electrophoresis (dPAGE) analyses of the ligation reactions (Figure 2.7). In the presence of staples (lanes 1 and 2), a dominant closed species (with 40% gel yield) is formed with greater electrophoretic mobility than the sample without staples (lanes 3 and 4). This indicates the formation of **TK**, which migrates faster than the monomeric unknot, **Circle**[6].

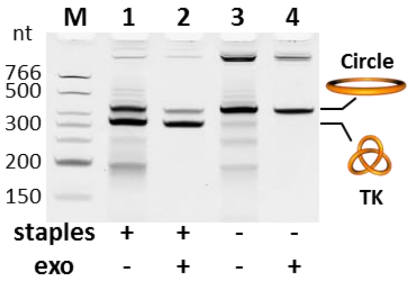

Figure 2.7 Denaturing polyacrylamide gel electrophoresis (dPAGE) showing the construction of TK.

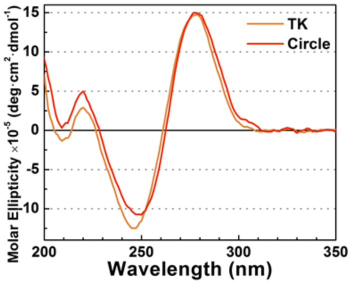

Figure 2.8 CD spectra of TK and Circle.

Circular dichroism (CD) spectroscopy is used to reveal the three-dimensional (3D) structural differences. As shown in F#, the CD spectrum of the **TK** fundamentally resembles that of the **Circle**, but some noticeable distinctions are observed below 250 nm. Compared to the **Circle**,

TK has a more negative peak at 245 nm with a blueshift of 5 nm and a less positive peak at 220 nm. The CD spectrum of ssDNA is largely contributed by the base stacking and unavoidable secondary structures[26]. These differences, therefore, reflect perturbations in base stacking and secondary structure formation due to knotting, which changes the writhe of the ssDNA molecule.

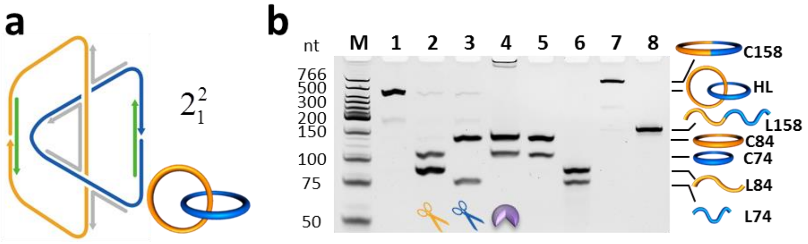

Figure 2.9 Construction for the Hopf link, HL. (**a**) Assembly complex for **HL**. (**b**) Topological analysis of **HL**.

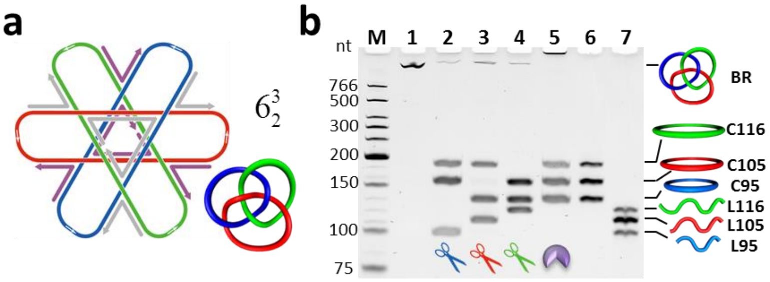

Figure 2.10 Construction for the Borromean rings, BR. (**a**) Assembly complex for **BR** (splints and nicks in scaffolds are omitted for clarity). (**b**) Topological analysis of **BR**.

Similarly, the topology of a Hopf link, **HL** is also constructed by folding two scaffolds (Figure 2.9a), which are the precursor for the two rings in this topology. Its topology is confirmed by the specific cleavage with two different nickases, whose restriction sites are installed in the two component rings of different sizes (Figure 2.9b). Treatment with either nickase disassembles the **HL** into one circular and one linear strand (lanes 2 and 3). **HL** is also resolved by the *E. coli* topoisomerase I (Topo I) into the component rings (lane 4).

We further constructed a more challenging topology, 6_2^3 Borromean rings[27], **BR** to demonstrate our capability of producing both negative and positive nodes. The defining feature of this famous topology is that cleaving one ring makes the whole structure fall apart. This requires an equal number of positive and negative nodes and, consequently, two tensegrity triangles with different handedness are designed in the assembly complex (Figure 2.10a). It should be noted that the topology of **BR** is the exact topology of 6_2^3, whereas the DNA Borromean rings constructed previously with the helix-based method contain 18 nodes[9]. The three component rings are designed to be different sizes, and each is installed with a unique restriction site, so that the topology of **BR** is unambiguously proved by the cleavage with nickases and relaxation with Topo I (Figure 2.10b).

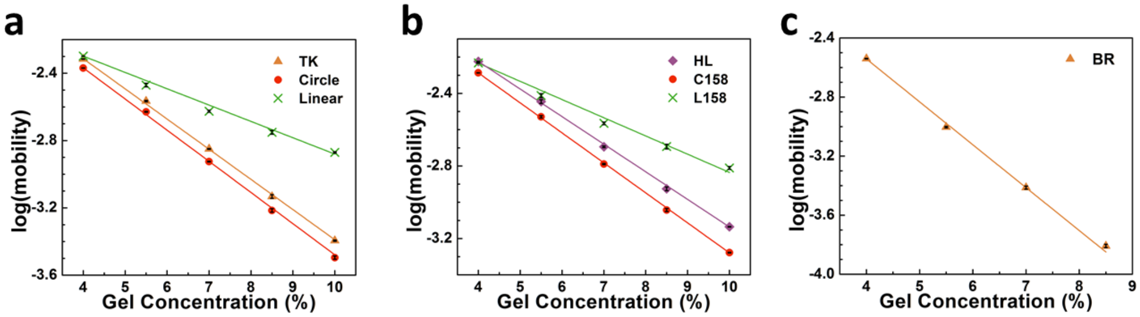

Figure 2.11 Ferguson plot analyses. Ferguson plot analyses of **TK** (**a**) and **HL** (**b**) with their respective circular and linear counterparts. (**c**) Ferguson plot for **BR**.

The electrophoretic behaviors of these three topological structures with their related species were studied by Ferguson plots (Figure 2.11), in which the logarithm of the electrophoretic mobility (M) is plotted as a function of the total acrylamide concentration (T): $\log(M) = \log(M_0) - K_R T$[28]. The extrapolated intercept M_0 is the free electrophoretic mobility, which is related to the net surface charge density. The slope K_R is the retardation coefficient, which is related to the apparent molecular size. Similar to previous studies[6-8], the linear species have very different parameters from the knotted or interlocked species and the circular species, featuring a more gradual slope and a smaller intercept (presumably due to the worm-like motion). Compared with their respective circular species, **TK** (Figure 2.11a) and **HL** (Figure 2.11b) have a smaller slope (absolute value) and a larger intercept. This indicates that, as expected, the knotted or interlocked topology makes these molecules more compact, exhibiting a particle-like electrophoretic behavior. **BR** features a large intercept and a steep slope, reflecting its compactness (Figure 2.11c).

2.2.3 Left- and right-handed trefoil knots with the same sequence

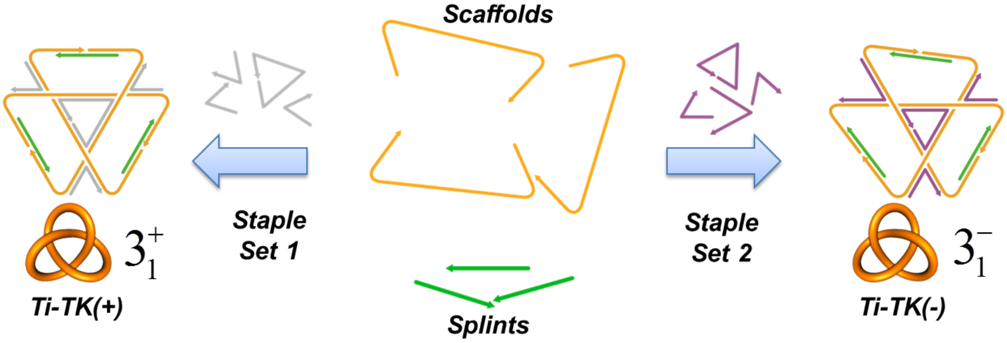

Figure 2.12 Constructing trefoil knots of both handednesses.

The construction of topoisomers can be achieved by designing the sequences of the scaffolds so that different tensegrity triangles are formed with different sets of staples. Accordingly, trefoil knots of both handedness[8], 3_1^- **Ti-TK(-)** and 3_1^+ **Ti-TK(+)**, are constructed (Figure 2.12). This enables us to investigate topological properties independent of composition. The electrophoretic properties of both trefoil knots are almost identical under the denaturing condition (Figure 2.13a). This reflects that the mobility of the DNA topoisomers in dPAGE is only dependent on topology, and that knots with different handedness cannot be differentiated. However, **Ti-TK(-)** migrates slightly faster than **Ti-TK(+)** in the native gel (Figure 2.13b). Unlike the denaturing condition, some unavoidable weak base-pairing interactions are formed under native condition. Promotion of these interactions makes the molecule more compact, which is reflected by higher mobility. Our results indicate that the topology of a negative knot promotes the formation of some weak base-pairings whereas a positive knot inhibits their formation. This is understandable because most base-pairings form right-handed anti-parallel helices, which are negative-node-producing. The addition of Mg^{2+} to the electrophoresis buffer can further promote the formation of base-pairings (Figure 2.13c). Under this condition, the mobility differences among these species becomes less obvious and the unknot **Ti-C** almost catches up with **Ti-TK(+)**.

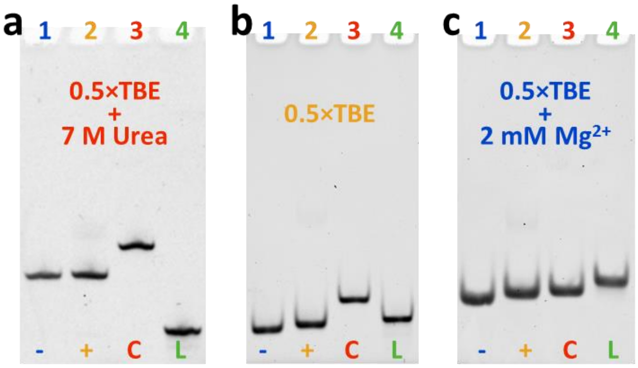

Figure 2.13 PAGE analyses of the two trefoil knots with their linear and circular counterparts under different conditions. (**a**) dPAGE analyses. (**b**) Native PAGE without the addition of Mg^{2+}. (**c**) Native PAGE in the presence of 2 mM Mg^{2+}.

The CD spectra (Figure 2.14) of **Ti-TK(+)** and **Ti-C** have noticeable distinctions, which are consistent with the previous case of **TK** and **Circle**. However, the spectrum of **Ti-TK(-)** closely resembles that of **Ti-C**. This fact also indicates that knotting with different handedness has very different effects on the higher-order structures of ssDNA, which are likely due to the weak base-pairing interactions (as has been indicated by the native gel results).

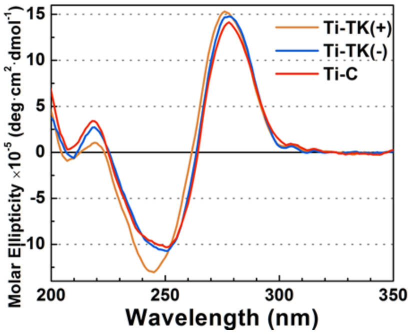

Figure 2.14 CD spectra of Ti-TK(+), Ti-TK(-) and the unknot Ti-C.

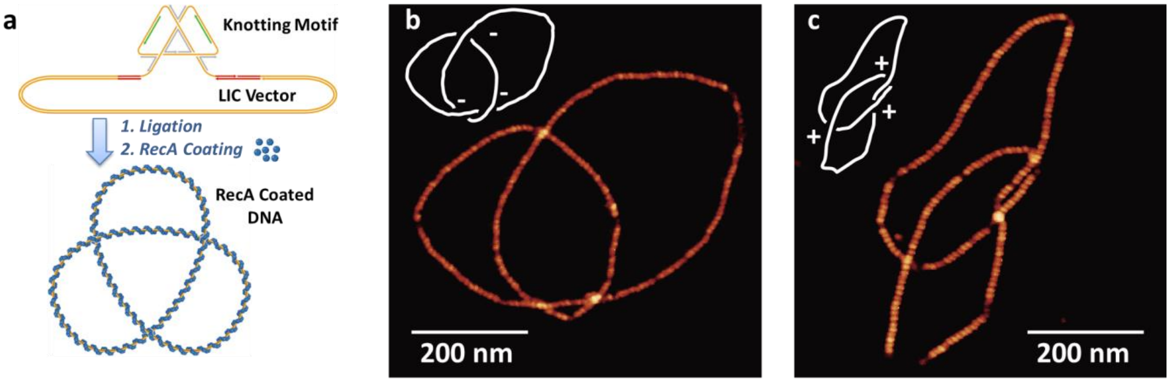

Figure 2.15 “Topological grafting” for direct visualization of topologies. (**a**) Knots are enlarged by ligating the knotting motifs (derived from the assembly complexes) to a 4.5-kb LIC vector (pLATE11 by Thermal Scientific) for direct visualization after RecA-coating. (**b** and **c**) Representative AFM images of the RecA-coated enlarged DNA knots with negative (**b**) and positive (**c**) nodes. Inserts show the outlined paths of the RecA-coated DNA knots and the handedness of each node.

The RecA-coating method is routinely utilized for direct visualization of the DNA topology[29]. However, our knots thus far have been too small (a contour length of ~120 nm) to apply this technique. Therefore, we devised a new method, coined “topology grafting”, to enlarge the knots for RecA-coating (Figure 2.15). This grafting is achieved by ligating the knotting motif, which is derived from the assembly complex, to a ligase independent cloning (LIC) vector. Therefore, the knotted topology of the vector has been conferred by the knotting motif, but not through the extremely low probability of random knotting of the closure of linear dsDNA of this size[30]. During the subsequent incubation with RecA, which can form a helical coating along DNA, the staples no longer hold the junctions due to the high incubation temperature and stiffening of the structure. The RecA-coated DNA knots are then subjected to atomic force microscopy (AFM) imaging for topology identification (Figure 2.15b and c). As a result, the formation of nodes with designed handedness can be directly visualized.

2.2.4 Topoisomers of higher-order complexity

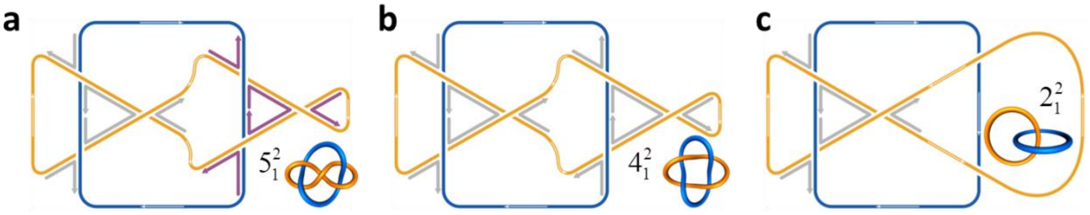

Figure 2.16 Topoisomers of two-component links. The Whitehead link **Ti-WL** (**a**), the Solomon link **Ti-SL** (**b**) and the Hopf link **Ti-HL** (**c**) are formed by configuring two tensegrity triangles.

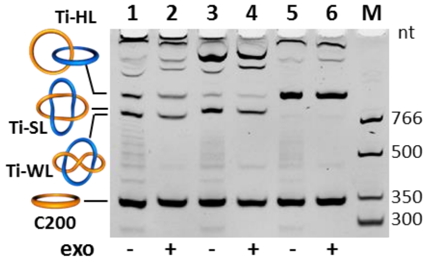

Figure 2.17 Formation of the two-component links analyzed by denaturing PAGE. Bands with different mobility are formed corresponding to **Ti-WL** (lanes 1 and 2), **Ti-SL** (lanes 3 and 4) and **Ti-HL** (lanes 5 and 6).

Even more complex topoisomers can be created by configuring two tensegrity triangles. Figure 2.16 shows our strategy for creating a series of two-component topoisomers: 5_1^2 Whitehead link, **Ti-WL** (Figure 2.16a), 4_1^2 Solomon link, **Ti-SL** (Figure 2.16b), and 2_1^2 Hopf link, **Ti-HL** (Figure 2.16c). The crude ligation reaction mixtures were analyzed by dPAGE (Figure 2.17). The order of electrophoretic mobility is **Ti-HL** < **Ti-SL** < **Ti-WL**, consistent with the order of compactness of these topoisomers; and this is also the order of the number of nodes in the minimum projection. Although there is **Ti-HL** byproduct formed in the reaction of **Ti-WL** and **Ti-SL**, likely due to the formation failure of one triangle in the assembly complexes, no **Ti-SL** is formed in the reaction of **Ti-WL** and *vice versa*. This validates that connecting junctions within geometrical constraint can precisely control the handedness of nodes. The topologies of these three topoisomeric links are further supported by the treatment of nickases and Topo I (Figure 2.18).

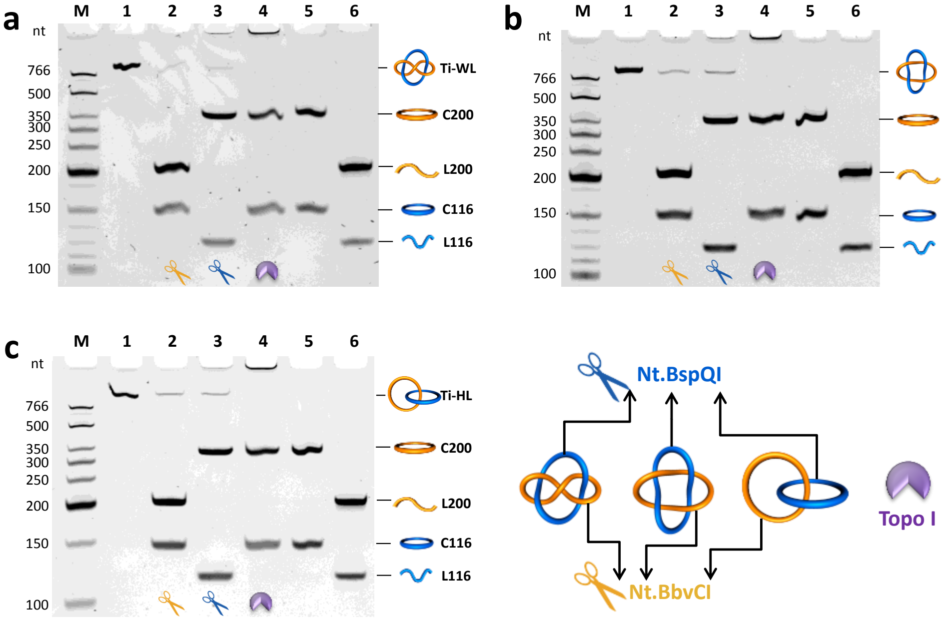

Figure 2.18 Confirmation of the two-component topologies by enzyme digestions. dPAGE analyses of the reaction of **Ti-WL** (**a**), **Ti-SL** (**b**) and **Ti-HL** (**c**) with various enzyme digestions. Each gel was loaded with the same order: lane 1, purified **Ti-WL** (**a**), **Ti-SL** (**b**) or **Ti-HL** (**c**); lane 2, digestion by Nt.BbvCI (which cleaves the orange strand); lane 3, digestion by Nt.BspQI (which cleaves the blue strand); lane 4, digestion by *E. coli* topoisomerase I; lane 5, the two circular references; lane 6, two linear references.

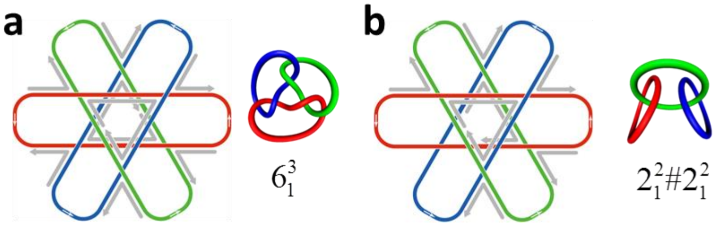

Figure 2.19 Topoisomers of three-component links. Two topoisomers of the Borromean rings **Ti-BR**, which are cyclic [3]catenane **Ti-C3C** (**a**) and the linear [3]catenane **Ti-L3C** (**b**), are designed by keeping one triangle fixed as 17-bp-edged and the other variable.

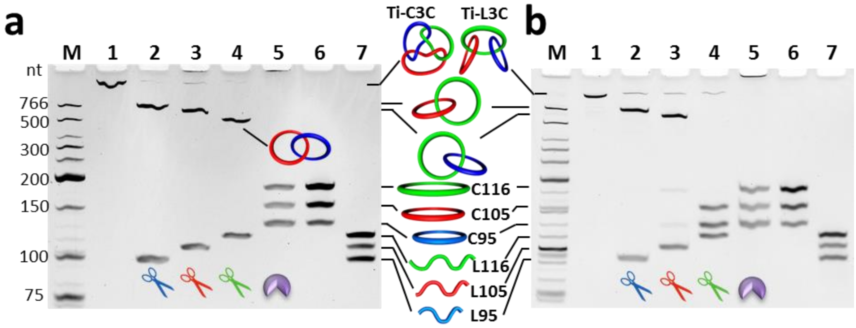

Figure 2.20 Confirmation of the three-component topologies by enzyme digestions. Topological analysis of **Ti-C3C** (**a**) and **Ti-L3C** (**b**). Lane 1, purified **Ti-C3C** or **Ti-L3C**; lanes 2 to 4, samples cleaved by three different nickases (Nt.BbvCI, Nt.AlwI and Nt.BspQI); lane 5, samples resolved by Topo I; lanes 6 and 7, circular and linear references of the three individual components.

Likewise, three topoisomers of three-component links can also be generated. Figure 2.19 shows the designed assembly complexes for two topoisomers of the Borromean rings (**Ti-BR**, which has a similar design with **BR** as shown in Figure 2.10, but of different sequence), the torus link cyclic [3]catenane (6_1^3), **Ti-C3C** (Figure 2.19b), and the composite link linear [3]catenane ($2_1^2\#2_1^2$), **Ti-L3C** (Figure 2.19c). The topologies of these molecules can be readily identified by nickase cleavage. As for **Ti-C3C** (Figure 2.20a), each pair of rings is mutually interlocked and therefore a Hopf link remains when any ring is cleaved. The difference between the **Ti-L3C** and

Ti-C3C is that cleavage of the middle green ring in the **Ti-L3C** takes apart the other two rings (lane 4 in Figure 2.20b).

2.2.5 Validation of new structural motif by topology

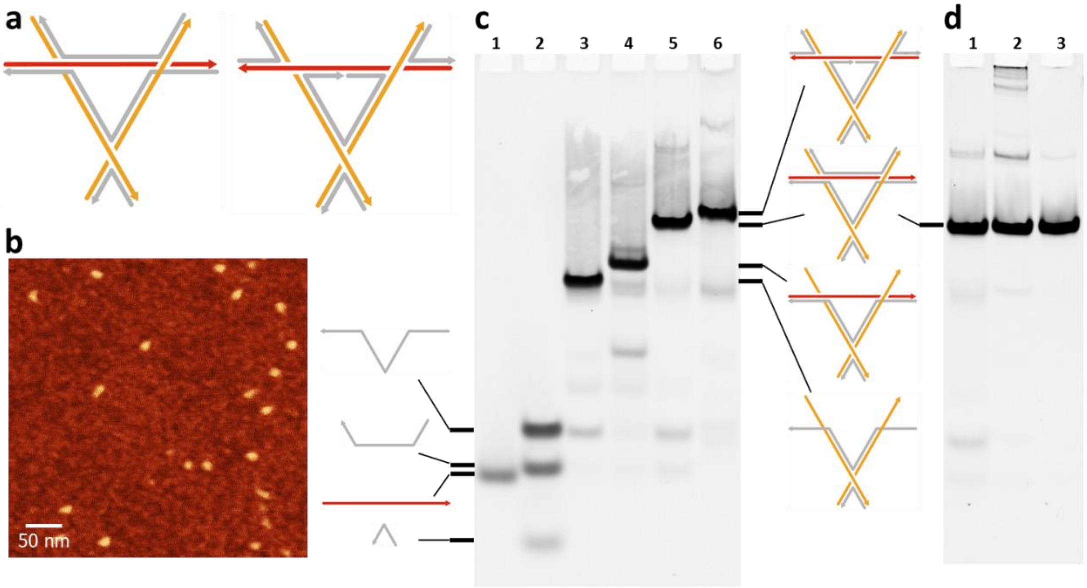

Figure 2.21 Design the 17-bp-edged reverse tensegrity triangle. (**a**) Schematics comparing the reverse tensegrity triangle (left) and the normal tensegrity triangle (right). (**b**) AFM image showing the formation of the reverse tensegrity triangle. (**c**) Native PAGE analysis showing the formation of the reverse tensegrity triangle. Lanes 1 to 4, annealing reactions of assembling partial components of the reverse tensegrity triangle; lane 5, reverse tensegrity triangle (annealed with concentration of 2 µM, gel yield 94%); lane 6, normal tensegrity triangle. The reverse tensegrity triangle shows slightly higher mobility than the normal tensegrity triangle. (**d**) For the sake of accurate stoichiometry, the formed reverse tensegrity triangle was purified from the gel and reanalyzed by native PAGE. Lane 1, the triangle assembled directly from component strands; lane 2, the triangle purified from the gel showing formation of some oligomers, probably formed during the gel extraction process; lane 3, the purified product is re-annealed (1 µM concentration) and assembles with accurate stoichiometry to the reverse tensegrity triangle with a yield of 98%.

Structural motifs, such as tensegrity triangles, have been used for topological construction so far; the successful construction of target topology can, in turn, validate the formation of the designed motifs. To exhibit this, we constructed a modified tensegrity triangle by reversing the polarity of helix of one edge in the normal tensegrity triangle (Figure 2.21). This new motif, which we dubbed a “reverse tensegrity triangle”, was computationally predicted by Sherman[31], but has not yet been shown experimentally. This reverse tensegrity triangle is a very desirable motif for topological construction because nodes with different handedness are produced within one triangle. Therefore, we incorporated it into the design of a 4_1 figure-of-eight knot, **Ti4-FE**, (Figure 2.22). The formation of **Ti4-FE** provides the topological detail of the threading of strands within the reverse tensegrity triangle, which otherwise could be very difficult to obtain. By using different sets of staples, we also designed the negative and positive trefoil knots with the same sequence (Figure 2.22). As a result, we also achieved the synthesis of four topoisomers, three of which are knotted[8].

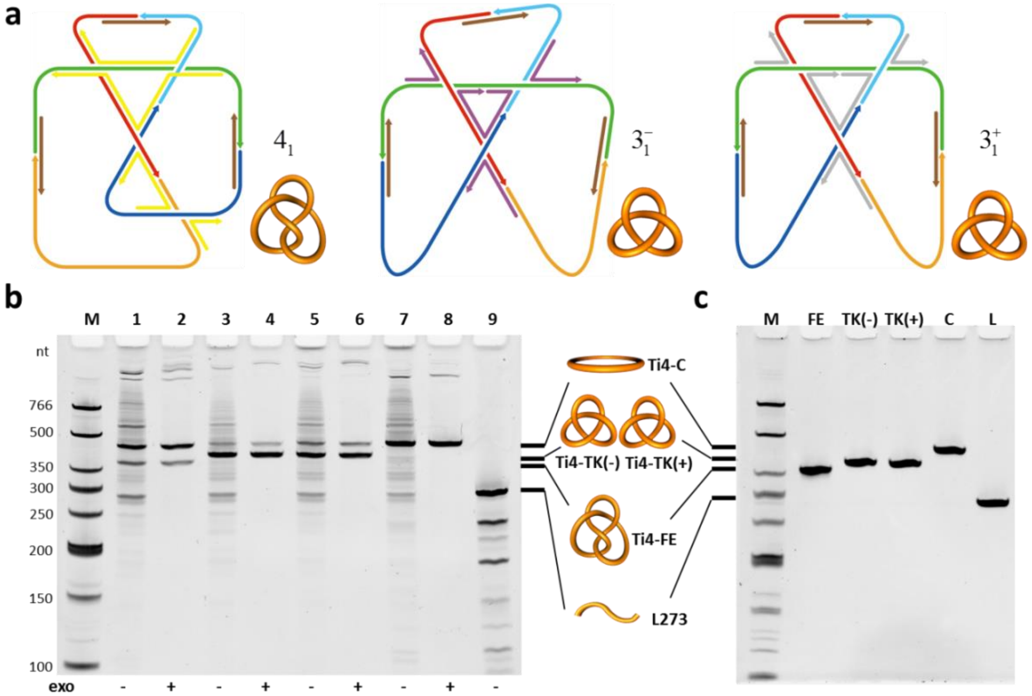

Figure 2.22 Topological validation of the designed 17-bp-edged reverse tensegrity triangle. (**a**) Schematics of the assembly complexes for the figure-of-eight knot **Ti4-FE** (left), negative trefoil knot **Ti4-TK(-)** (middle) and positive trefoil knot **Ti4-TK(+)** (right). (**b**) dPAGE showing the crude products for the construction of **Ti4-FE** (lanes 1 and 2), **Ti4-TK(-)** (lanes 3 and 4), **Ti4-TK(+)** (lanes 5 and 6) and corresponding circle **Ti4-C** and linear species **L273**. Lanes 2, 4, 6 and 8 is the crude products further treated by a mix of exo I and exo III. (**c**) dPAGE analysis of the purified product of **Ti4-FE**, **Ti4-TK(-)**, **Ti4-TK(+)**,**Ti4-C** and **L273**.

2.2.6 Design and construction of dsDNA knots

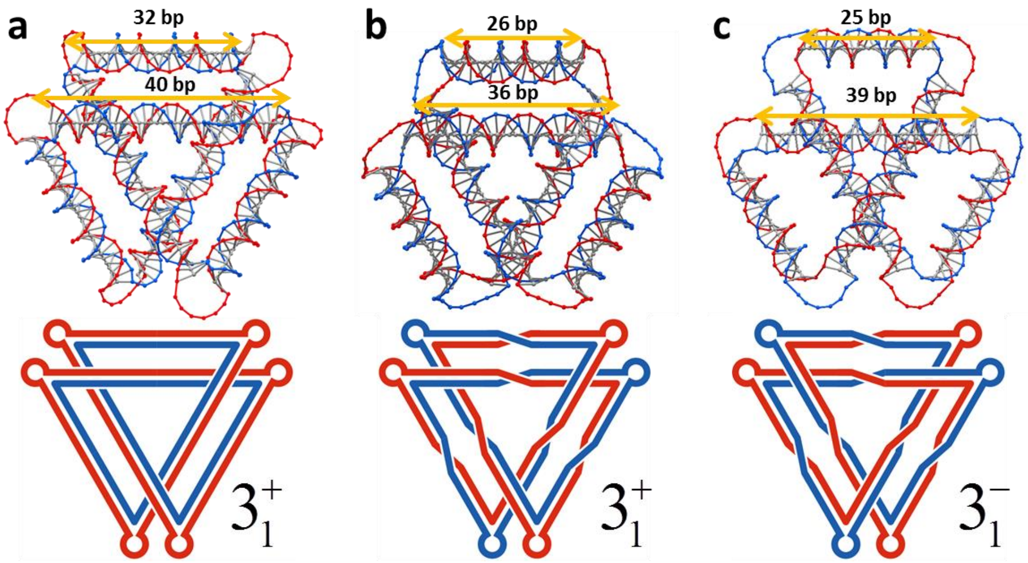

Figure 2.23 Schematics of three dsDNA trefoil knots (dsKnots) constructed.

Table 2.2 Characteristic parameters of the dsDNA trefoil knots.

dsknot	symmetry	handedness	outer helix	inner helix
dsTKCP	*C3*	+	32 bp	40 bp
dsTKDP	*D3*	+	26 bp	36 bp
dsTKDN	*D3*	-	25 bp	39 bp

Due to the more rigid and well-defined 3D structures, dsDNA topological structures are, in some cases, more desirable for nanofabrication compared with the ssDNA topological structures constructed thus far. Although a few dsDNA topological nanostructures have been constructed previously[14,32], these structures are only composed of dsDNA nanocircles such as catenanes or rotaxanes. In fact, it is more difficult to synthesize small dsDNA knots. Unlike ssDNA or sufficiently long dsDNA, short fragments of dsDNA behave like a rigid rod, the embodiment of a knotted topology in which necessitates the consideration of 3D geometry. Specifically, two main aspects have to be considered. The first is curvature. According to the Fary-Milnor theorem, the total curvature of a knot is larger than 4π, which is more than two times that required for a circle. The second is torsion, because a knot is by no means a planar structure. To achieve these, we designed the dsDNA knots to contain six poly(dT)$_6$ bulges (for curvature), which are separated by dsDNA helices with a specific number of base pairs (for torsion). Accordingly, we designed three different dsDNA knots (Figure 2.23 and Table 2.2) with either *C3*- or *D3*-symmetry (neglecting the sequence details). For the *C3*-symmetry dsDNA knot **dsTKCP** (Figure 2.23a), all of the six bulges are from one component strand and are separated by dsDNA helices of close to an integer number of turns. In contrast, in the two *D3*-symmetry dsDNA knots **dsTKDP** (Figure 2.23b) and **dsTKDN** (Figure 2.23c), each component strand contributes three bulges, which are separated by dsDNA helices of close to a half-integer number of turns.

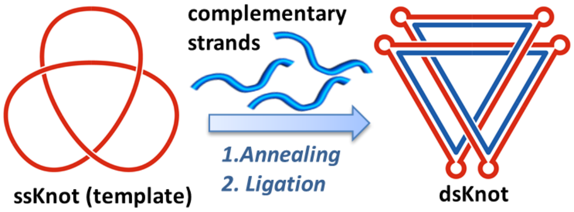

Figure 2.24 Schematic showing the preparation of dsKnots.

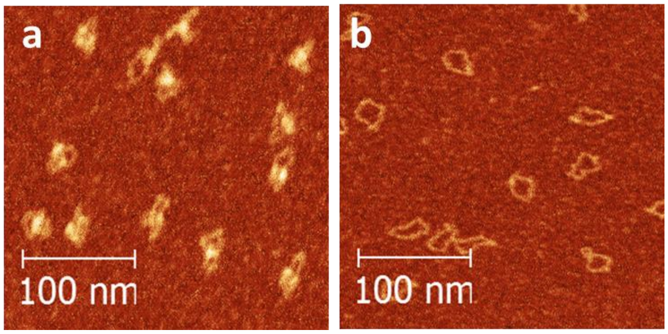

Figure 2.25 AFM images for dsTKCP (a) and its corresponding dsDNA circle (b).

Here, I only show the construction of **dsTKCP** as an example to demonstrate our capability to synthesis dsDNA knot. The dsDNA knots were successfully constructed by annealing and ligating complementary strands to the pre-prepared ssDNA knots using our junction-based method (Figure 2.24). An AFM image of the positive *C3*-symmetry dsDNA knot **dsTKCP** is shown (Figure 2.25a), and compared with its dsDNA circle counterpart (Figure 2.25b), it is more compact with higher bumps indicating helix crossings. To aid in proving the topology and geometry of **dsTKCP**, we installed two different restriction sites in one inner and one outer helix for cleavage, and analyzed the purified dsDNA knot and circle along with their digested products using dPAGE (Figure 2.26). Due to its highly compact structure, **dsTKCP** features very high electrophoretic mobility (lane 1): much higher than the corresponding dsDNA circle (lane 5) and even slightly higher than the precursor ssDNA knot template (lane 9). As shown in lane 2, Nt.BspQI nicks the template strand in one outer helix of the dsDNA knot and releases the knotted complementary strand. The designed geometry of the dsDNA knot is reflected by cleavage protection against the nickases (Nt.BbvCI and Nb.BbvCI) with the restriction site buried in the center of the inner helix (lanes 3 and 4), which is not observed for the corresponding dsDNA circle (lanes 7 and 8).

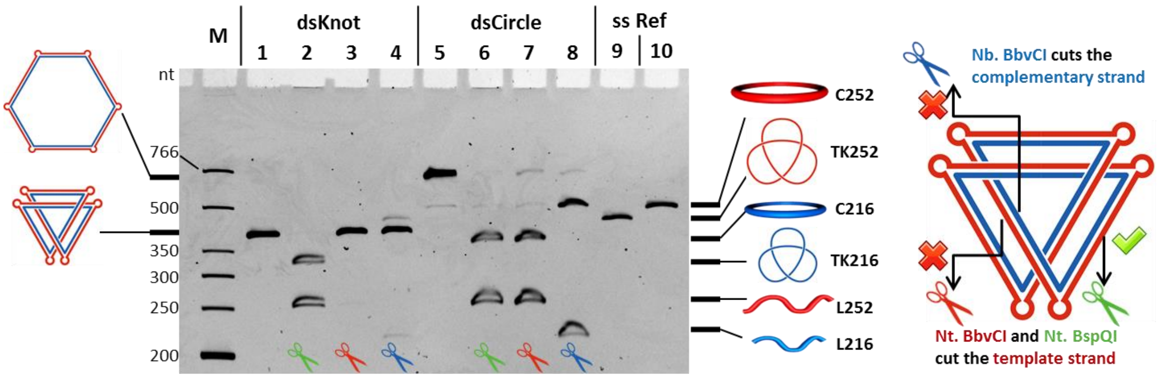

Figure 2.26 Nickase digestion confirming the formation of the dsKnot dsTKCP.

2.2.7 Topological conversion by DNA topoisomerase

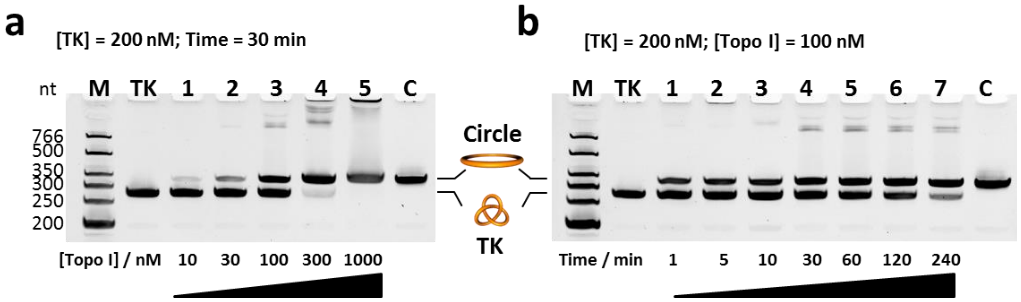

Figure 2.27 TK is converted to its topoisomer unknot Circle catalyzed by *E. coli.* Topo I. Concentration-dependent (**a**) and time-dependent (**b**) assays.

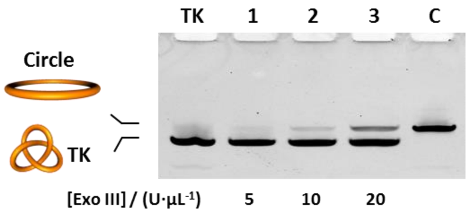

Figure 2.28 Detecting the topoisomerase contamination in commercial *E. coli.* Exo III.

DNA topoisomerase regulates the topology of DNA in both prokaryotic and eukaryotic cells, and represents an important target for anti-bacterial and anti-cancer drugs[33]. We show here that the ssDNA knot, as a topological probe, permits a robust assay for topoisomerase activity. Figure 2.27a shows the increasing conversion of **TK** to **Circle** catalyzed by increasing concentrations of *E. coli* Topo I. In the presence of sufficient Topo I, the **TK** can be converted completely to **Circle** (lane 5). This is strikingly different than the case with knots synthesized by the helix-based method[34], where the conversion can only reach an equilibrium due to the intrastrand base-pairing interaction. The time course of the topological conversion (Figure 2.27b) reveals that the action of Topo I is very fast and the stoichiometric conversion is complete in approximately 30 minutes. The reaction can proceed further over a longer incubation period, although very slowly.

This is likely due to the enzyme's ssDNA-binding activity causing its slow dissociation from the converted product. Based on these facts, we detected the topoisomerase contamination in the commercial product of *E. coli* Exo III (Figure 2.28)[34]. It is estimated that the contaminating topoisomerase is equivalent to the activity of 5 fmole of Topo I per unit of Exo III (NEB). We believe that our DNA topological constructs, in combination with other powerful tools developed by structural DNA nanotechnology, will definitely allow us a better understanding of the mechanism of this important type of enzyme. A good example is the single-molecule analysis enabled by the DNA origami frame structure[35], which has been employed to reveal a secondary binding site in human Topo IB[36].

2.2.8 DNA replication under topological constraints

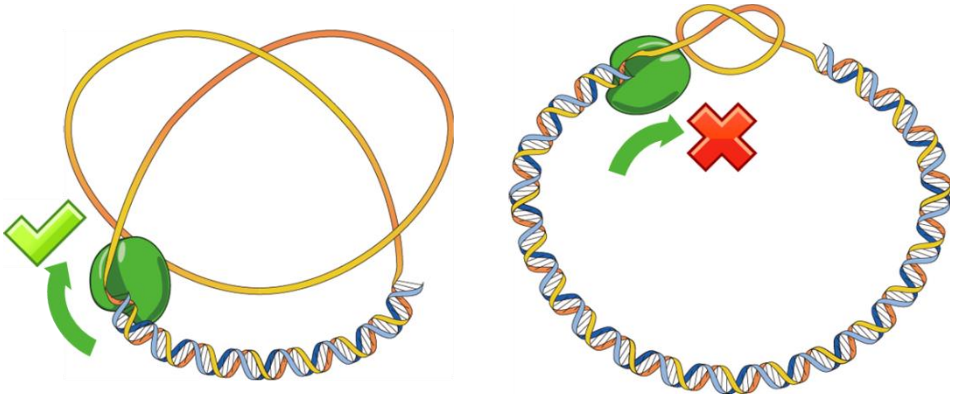

Figure 2.29 Schematics of the topological blockage of DNA polymerase procession.

The ssDNA knot also allows for the study of DNA polymerases (Pols) under topological constraints, which may help reveal the mechanics of DNA replication. As previously mentioned, the total curvature of a knot is larger than 4π. Consequently, the bending energy of a dsDNA knot is no less than four times that of a dsDNA circle. This means that it would be energetically unfavorable for the Pol to complete the replication on the small knotted template. We hypothesize that in the beginning stage, the Pol can extend the primer on the knotted template in a normal fashion (Figure 2.29 left). As the replication proceeds, a larger region of dsDNA forms along the knotted template. Since ssDNA (persistence length L_p =3 nm and length per base L_d =0.63 nm) is less rigid than dsDNA (L_p =50 nm and L_d =0.34 nm), the procession of Pol would tighten the diminishing ssDNA region, which bears the largest portion of the curvature. At a certain point, the high free energy cost associated with the tightening of the ssDNA region and bending of the dsDNA region would cause the procession of Pol to stall (Figure 2.29 right). To verify this hypothesis, DNA replication with the non-strand-displacing T4 DNA Pol was assayed using templates with different topologies (Figure 2.30). Only incomplete products (about 155-165 nt) were synthesized from the knotted template **TK**, (lane 1), while the full-length product (189 nt) was synthesized from Topo I-treated **TK** (lane 2) and unknotted template **Circle** (lane 3). Additional assays with two other non-strand-displacing DNA Pols (Q5® Hot Start and T7 DNA Pol) showed similar results of topological blockage (data not shown).

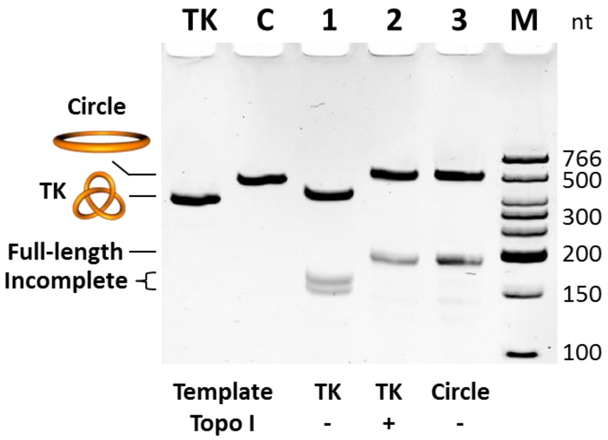

Figure 2.30 T4 DNA Pol is inhibited on the knotted template.

For the strand-displacing DNA Pols, rolling-circle amplification (RCA) was used to probe the topological effects. We measured the RCA curves of the phi29 DNA Pol under different topological conditions by monitoring the real-time fluorescence in the presence of SYBR® Green II (Figure 2.31). Unlike the non-strand-displacing Pols, phi29 DNA Pol was not completely blocked and could synthesize long ssDNA products containing tandem repeats of the full-length monomeric complementary strands from the **TK** template. The lack of complete blockage possibly reflects a larger stall force for this enzyme, presumably related to its strand-displacing activity. Nevertheless, the RCA rate with the **TK** template is only 2% of that with the **Circle** template. As expected, Topo I can recover the RCA by unknotting the **TK** template. This feature will allow for the development of a DNA-amplification-based and electrophoresis-free detection method of topoisomerase activity, which is promising for high-throughput screening of topoisomerase inhibitors (Figure 2.32).

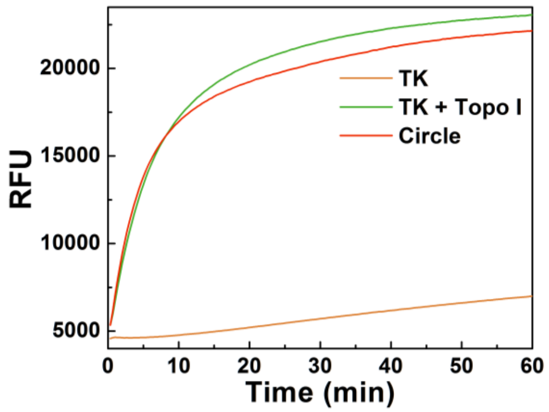

Figure 2.31 The RCA of phi29 DNA Pol is inhibited by the knotted DNA template.

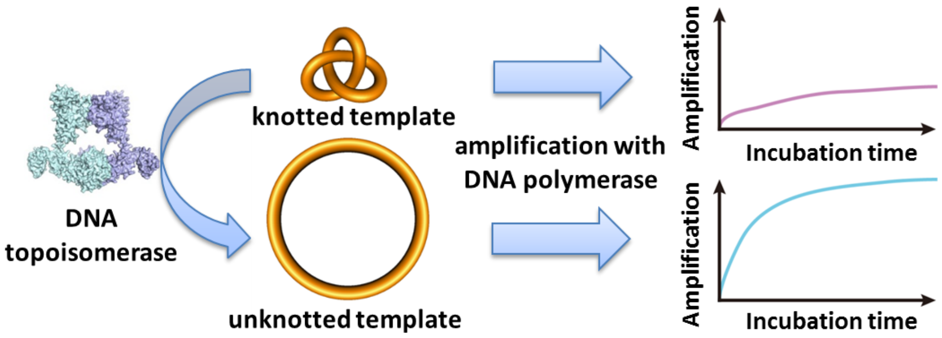

Figure 2.32 DNA-amplification-based for detecting topoisomerase activity.

2.2.9 Conclusions

Our junction-based strategy has substantially expanded the richness of molecular topology. This approach has enabled us to be the first to realize the molecular topology of the 5_1^2 Whitehead link and to create the first DNA Borromean rings with the exact 6_2^3 topology. The resulting ssDNA topological constructs has no intended intramolecular base-pairing interactions. This has also facilitated the construction of dsDNA knots. In each of our constructed dsDNA knot, the knotting of the complementary strand is guided by the topology of the knotted template strand. This opens up the possibility of using our ssDNA topological structures as universal templates for the synthesis of other materials[37]. Furthermore, the access to complex topological molecules from ssDNA or dsDNA should result in new opportunities for the study of elastic properties of DNA[38], or in the discovery of unknown enzymatic actions associated with DNA topology[39]. Of course, this junction-based method can be expanded to construct synthetic RNA topologies, which will be presented in the next chapter.

2.3 Experimental Section

2.3.1 Topological design and construction

The sequence design followed the rule of sequence symmetry minimization[40] and was assisted by CANADA[41]. All oligonucleotides were ordered from Integrated DNA Technologies. Generally, the ssDNA topological constructs were prepared with the all-in-one protocol, which involves three steps. (1) Annealing to get the assembly complex. Equimolecular quantities of all strands (scaffolds, staples and splints) were mixed in 1×T4 DNA Ligase Buffer (NEB, 1×buffer: 50 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, pH 7.5 @ 25 ℃) and annealed from 90 ℃ to 16 ℃ over 4 hours in a water bath (PolyScience). (2) Ligation to seal the nicks. To each 100 μL of reaction mixture, 2.5 μL of T4 DNA ligase (NEB, 400 U/μL), 2 μL of 100 mM fresh DTT (NEB) and 2 μL of 100 mM fresh ATP (NEB) were added and incubated at 16 ℃ for at least 16 hours. (3) Exonuclease digestion to eliminate linear species. To each 100 μL of reaction mixture, 2.5 μL of exonuclease I (NEB, 20 U/μL) and 2.5 μL of exonuclease III (NEB, 100 U/μL) were added and incubated at 37 ℃ for 45 min. For some specific topologies, small modifications were made to increase the yield. The dsDNA knots were prepared using the corresponding ssDNA knots as templates. The complementary strands were designed as 3 substrands, which were annealed to the ssDNA knot templates with a ratio of complementary:template = 1.2:1 and T4 DNA ligase was then used to seal the nicks.

2.3.2 Denaturing polyacrylamide gel electrophoresis (dPAGE)

Gels of different concentrations (which were carefully chosen to ensure the proper separations between different topologies as well as the references) were prepared using 30% acrylamide and bis-acrylamide solution (Bio-Rad, 19:1) with 7 M urea in 0.5×TBE buffer (Bio-Rad) and run on a PROTEAN® II xi cell (Bio-Rad) or a Mini-PROTEAN® Tetra cell (Bio-Rad). Samples were mixed 1:1 with TBE-Urea Sample Buffer (Bio-Rad) and heated at 85 ℃ for 1 min before they were loaded.. For imaging, gels were stained with GelRed^{TM} (Biotium) and images were taken by Gel Doc^{TM} XR+ (Bio-Rad) system and processed by Image Lab^{TM} (v.4.0.1, Bio-Rad). For Ferguson plot analysis, absolute mobilities (cm/min/V) were measured and their base-10 logarithms were taken. For PAGE purification, gels (without staining) were visualized by UV shadowing against a fluorescent TLC plate and bands of interest were cut. Products were then eluted using the crush-and-soak method and the eluent was collected on 3K Nanosep® filters (Pall). DNA concentration was determined by measuring the OD_{260}.

2.3.2 Enzymatic treatment

Four nickases (or nicking restriction endonuclease) – Nt.BbvCI (NEB, 10 U/µL), Nb.BbvCI (NEB, 10 U/µL), Nt.AlwI (NEB, 10 U/µL) and Nt.BspQI (NEB, 10 U/µL) – were used to aid confirming the topologies. An additional short strand (most likely a splint strand) binding to each ssDNA construct has to be added with 2 equivalents to form the restriction site for each nickase. Restriction enzyme EcoRV (NEB, 20 U/µL) was used to cleave the hairpin in one design (Supplementary Fig. 5). We used the reaction condition (including buffer condition and temperature) for each enzyme recommended by the data sheet provided by NEB. Special attention, however, should be paid to the concentration of Nt.AlwI. Though 10 units were added to each 10 µL of reaction (containing 200 nM of ssDNA substrate) for Nt.BbvCI, Nb.BbvCI and Nt.BspQI, only 2 units were added for the reaction of Nt.AlwI because it exhibited very pronounced “star activity” at higher concentration (probably due to topoisomerase contamination). The use of the restriction version of each nickase (cutting both strands of dsDNA), though not tested, should in principle gives similar results. Experimental details of the reaction conditions for *E. coli* topoisomerase I and various polymerases are also according to the protocols recommended by the enzyme providers.

2.3.3 Circular dichroism (CD) spectrum

DNA samples were prepared with a concentration of 1 μM in 10 mM sodium phosphate buffer (pH 7.4) supplemented with 0.1 mM EDTA. CD was obtained on an AVIV Model 420 CD spectrometer at constant temperature of 25 ℃.

2.3.4 AFM imaging

For the RecA-coated enlarged knots, the coating mixture was diluted 10 times with 1×RecA Buffer (NEB, 1×buffer: 70 mM Tris-HCl, 10 mM $MgCl_2$, 5 mM DTT, pH 7.6 @ 25 ℃). 5 μL of the diluted sample was deposited onto a freshly cleaved mica surface (Ted Pella) and left to adsorb for 1 min. Mica was rinsed with 20 μL of water and then dried with compressed air. For the dsDNA structures and the reverse tensegrity triangle, 30 uL of 0.1 mg/mL polyornithine (Sigma) solution was added to freshly cleaved mica and stand for 3 min to increase the binding. The mica was rinsed with 1 mL of water and dried with compressed air. 5 uL of each sample (about 5 nM) in 1×TAE-Mg buffer (11 mM $MgCl_2$, 40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.0) was applied to the treated mica and stand for 1 min. Then the mica was rinsed with 1 mL water and dried with compressed air. AFM imaging was performed on a Veeco 8 AFM in the ScanAsyst® in Air mode using the scanayst-air tips (Veeco). The images were processed with Gwyddion.

2.4 References.

1. Forgan, R. S., Sauvage, J. P. & Stoddart, J. F. Chemical topology: complex molecular knots, links, and entanglements. *Chem. Rev.* 2011, **111**(9)**:** 5434-5464.
2. Ayme, J. F., Beves, J. E., Campbell, C. J. & Leigh, D. A. Template synthesis of molecular knots. *Chem. Soc. Rev.* 2013, **42**(4)**:** 1700-1712.

3. Gil-Ramirez, G., Leigh, D. A. & Stephens, A. J. Catenanes: fifty years of molecular links. *Angew. Chem. Int. Ed.* 2015, **54**(21)**:** 6110-6150.
4. Seeman, N. C. Nanomaterials based on DNA. *Annu. Rev. Biochem.* 2010, **79:** 65-87.
5. Bates, A. D. & Maxwell, A. *DNA topology*, 2 edn. Oxford Univ. Press: Oxford, 2005.
6. Mueller, J. E., Du, S. M. & Seeman, N. C. Design and synthesis of a knot from single-stranded DNA. *J. Am. Chem. Soc.* 1991, **113**(16)**:** 6306-6308.
7. Du, S. M. & Seeman, N. C. Synthesis of a DNA knot containing both positive and negative nodes. *J. Am. Chem. Soc.* 1992, **114**(24)**:** 9652-9655.
8. Du, S. M., Stollar, B. D. & Seeman, N. C. A synthetic DNA molecule in three knotted topologies. *J. Am. Chem. Soc.* 1995, **117**(4)**:** 1194-1200.
9. Mao, C., Sun, W. & Seeman, N. C. Assembly of Borromean rings from DNA. *Nature* 1997, **386**(6621)**:** 137-138.
10. Ciengshin, T., Sha, R. & Seeman, N. C. Automatic molecular weaving prototyped by using single-stranded DNA. *Angew. Chem. Int. Ed.* 2011, **50**(19)**:** 4419-4422.
11. Seeman, N. C. Synthetic DNA Topology. In: Sauvage JP, Dietrich-Buchecker C (eds). *Molecular Catenanes, Rotaxanes and Knots*. Wiley: Weinheim, 2007, pp 323-356.
12. Chichak, K. S. *et al.* Molecular borromean rings. *Science* 2004, **304**(5675)**:** 1308-1312.
13. Chen, J. H. & Seeman, N. C. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* 1991, **350**(6319)**:** 631-633.
14. Schmidt, T. L. & Heckel, A. Construction of a structurally defined double-stranded DNA catenane. *Nano Lett.* 2011, **11**(4)**:** 1739-1742.
15. Han, D., Pal, S., Liu, Y. & Yan, H. Folding and cutting DNA into reconfigurable topological nanostructures. *Nature Nanotechnol.* 2010, **5**(10)**:** 712-717.
16. Kallenbach, N. R., Ma, R.-I. & Seeman, N. C. An immobile nucleic acid junction constructed from oligonucleotides. *Nature* 1983, **305**(5937)**:** 829-831.
17. Liu, D., Chen, G., Akhter, U., Cronin, T. M. & Weizmann, Y. Creating complex molecular topologies by configuring DNA four-way junctions. *Nature Chem.* 2016.
18. Lilley, D. M. Structures of helical junctions in nucleic acids. *Q. Rev. Biophys.* 2000, **33**(2)**:** 109-159.
19. Birac, J. J., Sherman, W. B., Kopatsch, J., Constantinou, P. E. & Seeman, N. C. Architecture with GIDEON, a program for design in structural DNA nanotechnology. *J. Mol. Graph. Model.* 2006, **25**(4)**:** 470-480.
20. Fu, T. J. & Seeman, N. C. DNA double-crossover molecules. *Biochemistry* 1993, **32**(13)**:** 3211-3220.
21. Kumara, M. T., Nykypanchuk, D. & Sherman, W. B. Assembly pathway analysis of DNA nanostructures and the construction of parallel motifs. *Nano Lett.* 2008, **8**(7)**:** 1971-1977.
22. Liu, D., Wang, M., Deng, Z., Walulu, R. & Mao, C. Tensegrity: construction of rigid DNA triangles with flexible four-arm DNA junctions. *J. Am. Chem. Soc.* 2004, **126**(8)**:** 2324-2325.
23. Zheng, J. *et al.* From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature* 2009, **461**(7260)**:** 74-77.
24. Han, D. *et al.* DNA gridiron nanostructures based on four-arm junctions. *Science* 2013, **339**(6126)**:** 1412-1415.
25. Rothemund, P. W. Folding DNA to create nanoscale shapes and patterns. *Nature* 2006, **440**(7082)**:** 297-302.

26. Bloomfield, V. A., Crothers, D. M. & Tinoco, I. *Nucleic Acids: Structures, Properties, and Functions*. Univ. Science Books: Sausalito, 2000.
27. Cantrill, S. J., Chichak, K. S., Peters, A. J. & Stoddart, J. F. Nanoscale borromean rings. *Acc. Chem. Res.* 2005, **38**(1)**:** 1-9.
28. Rodbard, D. & Chrambach, A. Estimation of molecular radius, free mobility, and valence using polyacrylamide gel electrophoresis. *Anal. Biochem.* 1971, **40**(1)**:** 95-134.
29. Krasnow, M. A. *et al.* Determination of the absolute handedness of knots and catenanes of DNA. *Nature* 1983, **304**(5926)**:** 559-560.
30. Shaw, S. & Wang, J. Knotting of a DNA chain during ring closure. *Science* 1993, **260**(5107)**:** 533-536.
31. Sherman, W. B. HolT Hunter: software for identifying and characterizing low-strain DNA Holliday Triangles. *J. Comput. Chem.* 2012, **33**(15)**:** 1393-1405.
32. Ackermann, D. *et al.* A double-stranded DNA rotaxane. *Nature Nanotechnol.* 2010, **5**(6)**:** 436-442.
33. Chen, S. H., Chan, N. L. & Hsieh, T. S. New mechanistic and functional insights into DNA topoisomerases. *Annu. Rev. Biochem.* 2013, **82:** 139-170.
34. Du, S. M., Wang, H., Tse-Dinh, Y.-C. & Seeman, N. C. Topological transformations of synthetic DNA knots. *Biochemistry* 1995, **34**(2)**:** 673-682.
35. Rajendran, A., Endo, M. & Sugiyama, H. Single-molecule analysis using DNA origami. *Angew. Chem. Int. Ed.* 2012, **51**(4)**:** 874-890.
36. Subramani, R. *et al.* A novel secondary DNA binding site in human topoisomerase I unravelled by using a 2D DNA origami platform. *ACS nano* 2010, **4**(10)**:** 5969-5977.
37. Niu, J., Hili, R. & Liu, D. R. Enzyme-free translation of DNA into sequence-defined synthetic polymers structurally unrelated to nucleic acids. *Nature Chem.* 2013, **5**(4)**:** 282-292.
38. Marko, J. F. & Siggia, E. D. Bending and twisting elasticity of DNA. *Macromolecules* 1994, **27**(4)**:** 981-988.
39. Wang, H., Di Gate, R. J. & Seeman, N. C. An RNA topoisomerase. *Proc. Natl. Acad. Sci. USA* 1996, **93**(18)**:** 9477-9482.
40. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* 1982, **99**(2)**:** 237-247.
41. Feldkamp, U. CANADA: Designing Nucleic Acid Sequences for Nanobiotechnology Applications. *J. Comput. Chem.* 2010, **31**(3)**:** 660-663.

Chapter 3 – Synthesizing RNA-containing topological structures

3.1 Introduction

In Chapter 2, we discussed our efforts to construct synthetic DNA topological structures and the biological importance of DNA topology was also introduced. An interesting question naturally arises regarding the importance of RNA topology. Although the recent systematic screening of the Protein Data Bank (PDB) indicated the absence of genuine linear knots in known RNA structures[1,2], it is yet premature to disclaim the existence of naturally occurring knotted (linear or closed) RNA structures. On the one hand, the set of solved RNA structures are not representative of all RNA molecules. There are many more RNAs with unknown structures, and RNAs that remain to be identified. It is likely that knotted RNA structures can be discovered as more RNA structures are solved. On the other hand, RNA pseudoknots with two sufficiently long (at least around 11 bp) helices can adopt knotted conformation and several likely knot-forming candidates have been suggested[2] based on their sequences from an RNA pseudoknot database[3]. Instead of passively waiting for RNA topological structures to emerge through the accumulation of structural data, RNA topologies can be created and investigated through synthesis. Importantly, synthetic RNA topological structures can provide reagents possessing the properties necessary for developing new tools and methods to identify the naturally occurring RNA topological structures.

The first synthetic RNA topological structure was a single-stranded RNA (ssRNA) trefoil knot described by Seeman and coworkers in 1996[4]. As a sequel to a series of their remarkable research on synthetic single-stranded DNA (ssDNA) topologies[5-8], the ssRNA knot was constructed based on the fact that a half turn (5 or 6 bp) of RNA duplex generates a node (Fig. 1a and b). The realization of synthetic RNA topology essentially led to the discovery of the first RNA topoisomerase (RNA Topo) – *E. coli* DNA Topo III[4]. Based on this, RNA Topo activity has also been recently found in other Type IA DNA Topos[9,10], including the human DNA Topo 3β[9], which is crucial to neurodevelopment. Similar to the case of DNA, RNA topology and RNA Topos may transform our understanding of fundamental RNA biology.

However, synthetic RNA topology has received far less attention the its DNA counterpart[11-13]. Until now, Seeman's helix-based method is still the only way to access it[4]. In Chapter 2, we have introduced a versatile method to construct synthetic DNA topologies based on the stacked X structure of DNA four-way junction (4WJ), in which the two helical strands (continuous along the stacked helices) are held by the two crossover strands (exchanged at the junction) to form a node for topological construction[14]. We herein demonstrate that this method can be further developed for synthetic RNA topologies by using the RNA-DNA hybrid 4WJ, which contains RNA helical strands and DNA crossover strands. Thus, RNA topological structures can be generated by folding RNA scaffolds into hybrid 4WJs with DNA staples. In addition, we also show an alternative way for DNA-templated construction of an RNA topology, i.e. a pre-prepared ssDNA topological structure can be used as a template to synthesize the corresponding RNA topological structure.

3.2 Results and discussion

3.2.1 Producing RNA molecules suitable for ligation

In order to produce closed RNA topological structures, the linear RNA strands can be ligated enzymatically. However, RNA strands synthesized by *in vitro* run-off transcription, which is the most convenient and economic method to produce long RNA strands, contain both 5'- and 3'-

heterogeneities. Therefore, their enzymatic ligation poses a significant challenge. To solve these problem, a 5’-end hammerhead (HH) ribozyme and a 3’-end hepatitis delta virus (HDV) ribozyme[15] were designed in each RNA transcript to undergo self-cleavage to produce uniform ends (Fig. 2a). In the presence of ATP, T4 polynucleotide kinase can add a phosphate group to the 5’-hydroxyl and remove the 2’,3’-cyclic phosphate[16] of self-cleaved RNA transcripts. The resulting RNA molecules contain both ends proper for ligation and can serve as the scaffolds for topological construction.

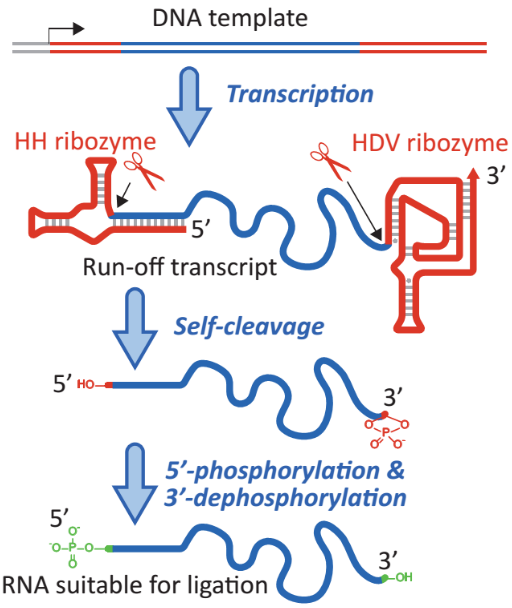

Figure3.1 Preparation of the ssRNA strand with uniform ends and proper end groups for the DNA-splinted RNA ligation.

3.2.2 Right- and left-handed RNA trefoil knots with the same sequence

The RNA-DNA hybrid 4WJs presumably have structural properties similar to that of DNA 4WJs due to the fact that the former have also been utilized in various RNA-DNA hybrid nanostructures[17,18], within which the 4WJs are composed of helical strands of RNA and crossover strands of DNA. Therefore, the design principles of DNA nanostructures based on

DNA 4WJs also hold for those containing the RNA-DNA hybrid 4WJs. The tensegrity triangle[19], a structural motif containing three 4WJs, is a case in point and it is utilized extensively in this work. The number of base pairs between the 4WJs dictates the tensegrity triangle to be either right- or left-handed[20], which results in positive or negative nodes, respectively. After geometrical analysis according to previous work[14,20] (Table 3.1), we used a 17-bp-edged tensegrity triangle for the right-handed RNA trefoil knot (3_1^+), **TK$_j$(+)**, and 14-bp-edged for the left-handed trefoil knot (3_1^-), **TK$_j$(-)**.

Table 3.1 Strain analysis of the RNA-DNA hybrid tensegrity triangles.

Edge Length	13 bp	14 bp	15 bp	16 bp	17 bp	18 bp	19 bp
Helix turning	425.5°	458.2°	490.9°	523.6°	556.4°	589.1°	621.8°
2α	107.7°	110.3°	112.7°	115.0°	117.2°	119.3°	121.2°
Geometric turning (-)	467.7°	470.3°	472.7°	475.0°	477.2°	479.3°	481.2°
Strain (-)	0.099	0.026	-0.037	-0.093	-0.142	-0.186	-0.226
Geometric turning (+)	612.4°	609.7°	607.3°	605.0°	602.8°	600.7°	598.8°
Strain (+)	0.439	0.331	0.237	0.155	0.083	0.020	-0.037
Handedness	-	-	-	-	+	+	+

The construction of topoisomers of both right- and left-handed trefoil knots can be achieved by folding the same RNA scaffolds into different tensegrity triangles with different sets of DNA staples (Figure 3.2). Notably, for the helix-based method, generating positive nodes would demand the formation of left-handed Z-form RNA, which requires harsher conditions (higher salt concentration and higher temperature)[21] than formation of Z-form DNA requires. In contrast, our junction-based method provides a more convenient way to generate the positive nodes because the handedness of the nodes is controlled by applying geometric constraints to 4WJs, but not by the formation of nonstandard duplex structures.

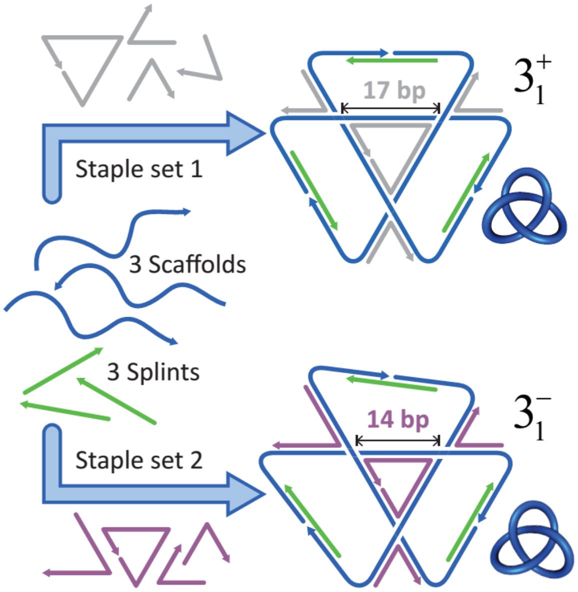

Figure 3.2 Preparing RNA trefoil knot topoisomers.

Experimentally, three 76-nt ssRNA scaffolds were annealed with either DNA staple set and splints to form each assembled complex containing three 4WJs (to form the tensegrity triangle) joined end-to-end by the splints (Figure 3.2). Subsequent ligation seals the nicks in between the scaffolds and thereby fixes the topology. T4 RNA ligase 2 was used for the DNA-splinted RNA ligation because lower enzyme concentration is needed compared to T4 DNA ligase (which ligates RNA less efficiently and only catalyzes approximately stoichiometric ligation). All the assisting DNA strands (staples and splints) are dissociated upon purification with denaturing polyacrylamide gel electrophoresis (dPAGE), and consequently relaxed 228-nt ssRNA topological constructs are generated without strong intrinsic intramolecular base-pairings. **TK_j(+)** and **TK_j(-)** were obtained in yields of 13% and 28%, respectively, comparable to the previous synthesis of DNA trefoil knots of similar size[14]. Figure 3.3 shows the purified RNA trefoil knots analyzed by dPAGE: lane 1 for **TK_j(+)**, lane 3 for **TK_j(-)**, along with the circular (**C_j**, lane 5) and linear (**L_j**, lane 7) references. Similar to the previous results with DNA topological structures[14],

TK_j(+) and **TK_j(-)** have almost identical electrophoretic mobility, which is higher than that of **C_j**. The resistance to digestion with RNase R proves the closed structure for knots (lanes 2 and 4) and circle (lane 6) but not for the linear species (lane 8).

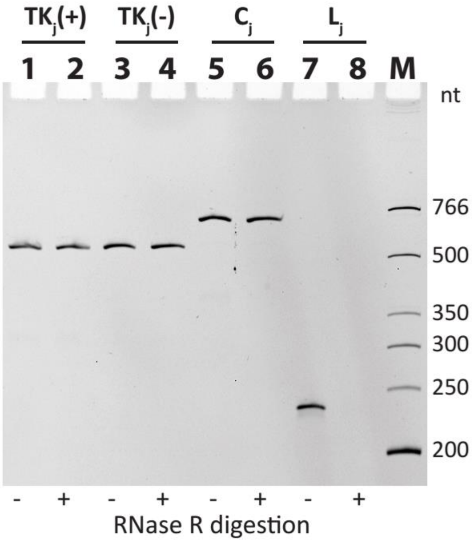

Figure 3.3 dPAGE analysis of RNA trefoil knots and their circular and linear counterparts.

3.2.3 Hybrid Borromean rings (BR) molecule containing both ssRNA and ssDNA rings

To demonstrate the high-level complexity of topological structures we can achieve with RNA, we then create a hybrid BR (6_2^3) structure containing one ssRNA and two ssDNA rings. The defining feature of this famous topology is that the whole assembly of three rings falls apart upon the scission of any one ring. This requires an equal number of positive and negative nodes to ensure that no two rings are interlocked. In chapter 2's construction of ssDNA BR[14], two tensegrity triangles with different handednesses were designed in the assembly complex to meet this requirement. Similarly, in the assembly complex for this hybrid BR, a 95-nt circular RNA, a 105-nt circular DNA and two linear DNA (57- and 59-nt, respectively) scaffolds are folded into a 17-edged right-handed tensegrity triangle for the three positive nodes and a 14-edged left-handed one for the three negative nodes (Figure 3.4). The two linear DNA scaffolds (precursors for the 116-nt DNA ring) are joined by two splints and the hybrid BR structure is formed after ligation by T4 DNA ligase. To conclusively prove the topology of this hybrid BR, each DNA ring is installed with a unique restriction site. As shown in Fig. 2e, the hybrid BR is disassembled by the cleavage of the ssRNA ring by RNase H (lane 2), or either ssDNA ring by the corresponding nickase (lanes 3 and 4).

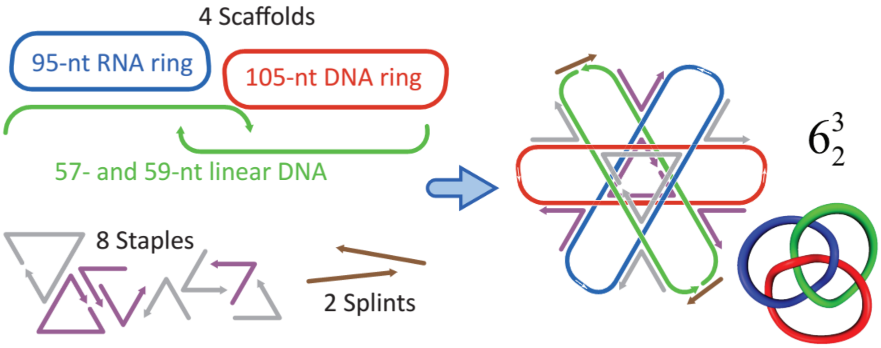

Figure 3.4 The assembly complex for the hybrid Borromean rings (BR).

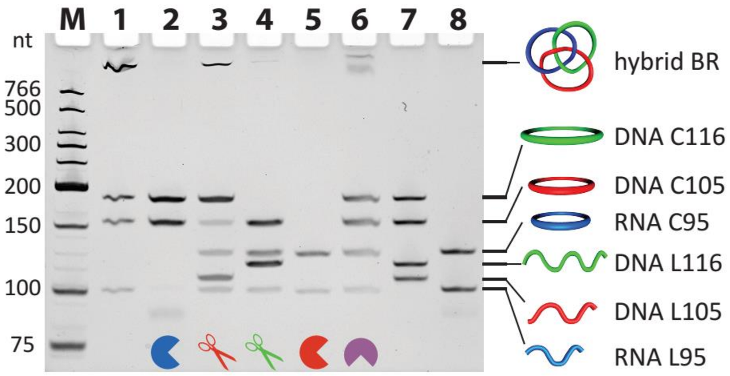

Figure 3.5 Topological analyses of the hybrid BR.

In the field of chemical topology, molecular BR is considered an Everest[22]. Like the real-world Mount Everest, the fact that it has been conquered once and again[8,14,23-26] by no means makes it less challenging, but only manifests its attractiveness to the committed synthetic adventurers. Here, we further extend the chemical diversity of this topological target by creating this hybrid BR structure, which is the first BR molecule to contain component rings of different materials. To the best of our knowledge, it is also the first topological structure composed of both DNA and RNA. Additionally, the successful construction of a topological target as complex as BR reasonably implies that using our junction-based method we ought to be able to access the ssRNA or ssRNA-ssDNA-hybrid versions of any of our previous topological targets realized with ssDNA[14].

3.2.4 Accessing ssRNA topological structure via DNA-templated synthesis

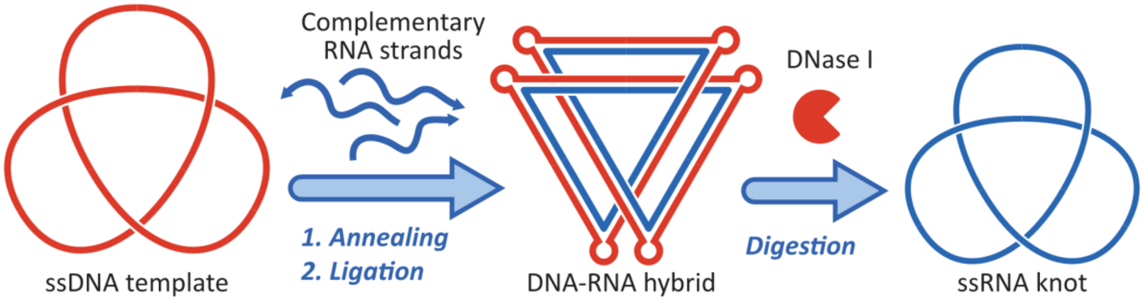

Figure 3.6 Using the ssDNA trefoil knot as a template for synthesizing ssRNA trefoil knot.

Synthetic ssDNA topological structures constructed previously, in principle, can direct the synthesis of ssRNA structures of the same topology (Figure 3.6). The intermediates would be the double-stranded (ds) RNA-DNA hybrid structures; however, the conversion of the structures from ssDNA to ds version is not always straightforward. This is especially true for small ds knots, which are more difficult to synthesize than other ds topological structures that contain only rings (such as rotaxanes[5] or catenanes[6]). Because ds nucleic acid structure is more rigid and adopts a better-defined geometry compared to ssDNA or ssRNA, constructing ds knot necessitates the careful design of curvature and torsion in 3D space[14]: (1) the total curvature of a knot should be larger than 4π (more than two times that required for a circle) according to the Fary-Milnor theorem; (2) the torsion of a knotted space curve should not be zero everywhere because a knot cannot be flattened onto the plane.

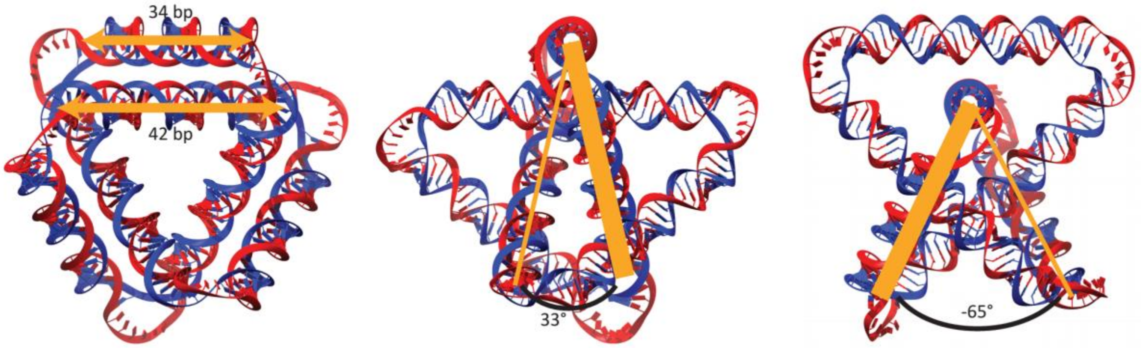

Figure 3.7 Three views of 3D helical model of the ds DNA-RNA hybrid knot.

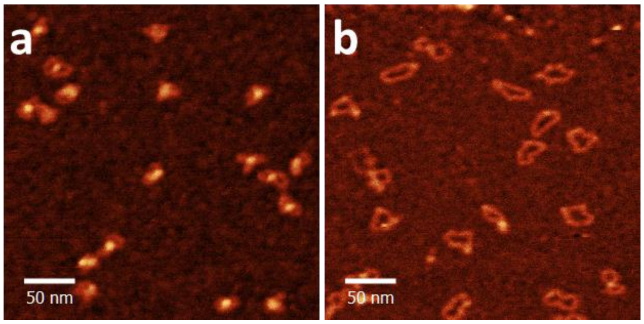

Figure 3.8 AFM images for the ds hybrid knot (a) and the corresponding circle (b).

To construct the ds RNA-DNA hybrid knot, a 264-nt ssDNA positive trefoil knot was pre-prepared (using the junction-based method) to serve as the template, and three 76-nt complementary RNA strands were then annealed onto the DNA template and ligated to form a 228-nt RNA strand. Six poly(dT)$_6$ bulges are designed in the DNA strand of the hybrid to provide six curving points, and they are alternatively placed at the ends of three outer 34-bp helices for a 33° torsion angle and three inner 42-bp ones for a -65° torsion angle (Figure 3.7). If the sequence details are neglected, this ds hybrid knot is a *C3*-symmetry molecule. Using ssDNA circle as the template, we also prepared a ds hybrid circle as a topoisomeric reference. The AFM images of the ds hybrid knot (Figure 3.8a) and circle (Figure 3.8b) reveal the obvious structural differences between them, and the ds hybrid knot adopts a more compact structure with strand crossings (reflected by the higher bumps in the AFM image). Furthermore, these ds hybrid structures were digested by nucleases and subsequently analyzed by dPAGE (Figure 3.9). The highly compact structure of the hybrid knot is again reflected by the very high electrophoretic mobility (lane 1), which exceeds the hybrid circle (lane 4) and the ssDNA template (lane 7). DNase I digests the DNA strand of either hybrid structure and, as expected, the ssRNA knot (lane 2) or circle (lane 5) is released. The ssDNA templates are recovered by RNase H digestion of the hybrid structures (lanes 3 and 6).

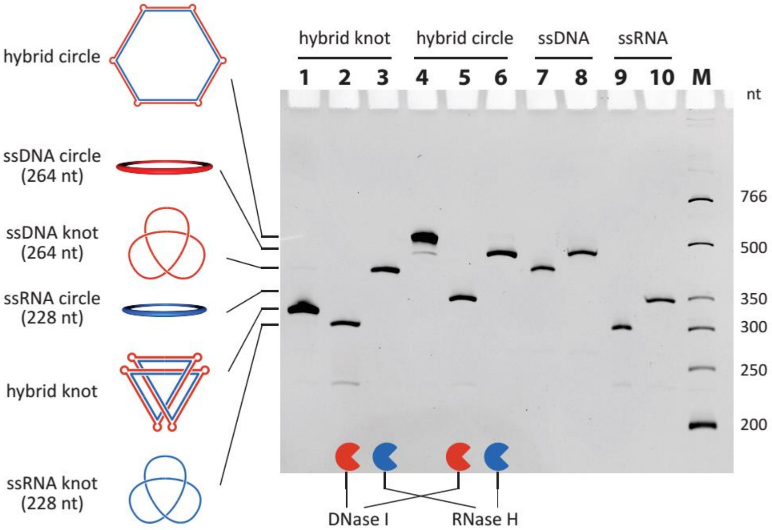

Figure 3.9 Nuclease digestion confirming the formation of the ds hybrid knot.

Not only does our approach provide the alternative route of DNA-templated synthesis for the construction of ssRNA topological structures, but also demonstrates the first realization of a ds RNA-DNA hybrid topological structure. This kind of ds hybrid structure, with more rigid and better-defined 3D structure, may find other potential applications in nanobiotechnology and nanofabrication. Additionally, the principle of using a knotted DNA template to guide the topology of RNA in the current research reveal the possibility of synthesizing topological structures of other non-nucleic acid materials with the more general DNA-templated synthesis[27].

3.2.5 Preparing several helix-based RNA topological structures

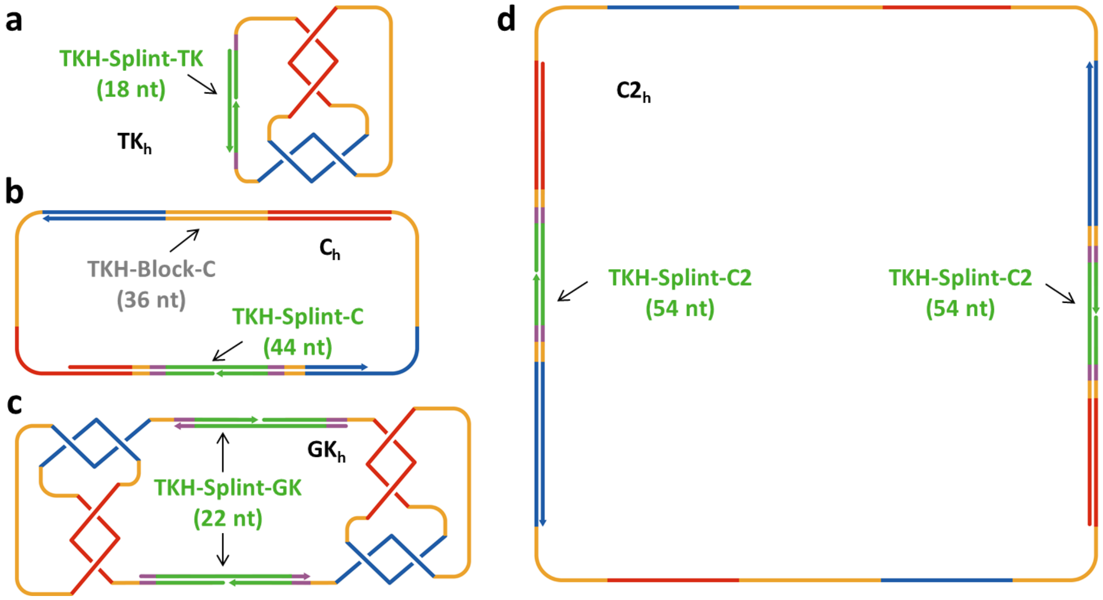

Figure 3.10 Schematics illustrating the preparation of four different helix-based topological structures using the same linear RNA precursor. (**a**) Preparation of the monomeric trefoil knot (**TK_h**) via Seeman's helix-based method[4]. (**b**) Preparation of the monomeric circle (**C_h**). **C_h** is the trickiest structure to construct in all these four structures. On the one hand, the intrastrand base-pairing should be inhibited. On the other hand, the single-stranded linker should be sufficiently long for the cyclization of the monomer. Therefore, a 36-nt block strand (TKH-Block-C) and a 44-nt splint (TKH-Splint-C) are used to inhibit the formation of the two helical region in the RNA. (**c**) Preparation of the dimeric granny knot (**GK_h**). A 22-nt splint (TKH-Splint-GK), without affecting the formation of helices in the RNA, decreased the single-stranded linker regions and therefore the dimeric knot is formed. (**d**) Preparation of the dimeric circle (**$C2_h$**). The 54-nt splint (TKH-Splint-C2) is long enough to inhibit the formation of helices in the RNA and to discourage the cyclization of the monomer.

One of the most significant discoveries resulting from synthetic RNA topology is the identification of *E. coli* DNA Topo III as the first RNA Topo[4]. More recently, several other DNA

Topos were also reported to have RNA Topo activity[9,10]. Unlike DNA Topos, the research on RNA Topos is belated and rare, mainly due to the lack of proper RNA Topo probes. The ssRNA knot prepared with our junction-based method does not contain strong base-pairings, and here we show that it serve as a more sensitive probe for RNA Topo activity compared with the previous helix-based probe[4]. Additionally, four RNA structures containing strong base-pairings were constructed from the same 114-nt linear RNA strand, $\mathbf{L_h}$, according to Seeman's helix-based method[4] with minor modifications. By adding different sets of assisting DNA strands in the synthesis, the monomeric trefoil knot, $\mathbf{TK_h}$, and circle, $\mathbf{C_h}$, and the dimeric Granny knot ($3_1^- \# 3_1^-$, a complex knot), $\mathbf{GK_h}$, and circle, $\mathbf{C2_h}$, were prepared (Figure 3.10). Figure 3.11 shows dPAGE analyses of these structures and the electrophoretic mobility is a function of both size and topology.

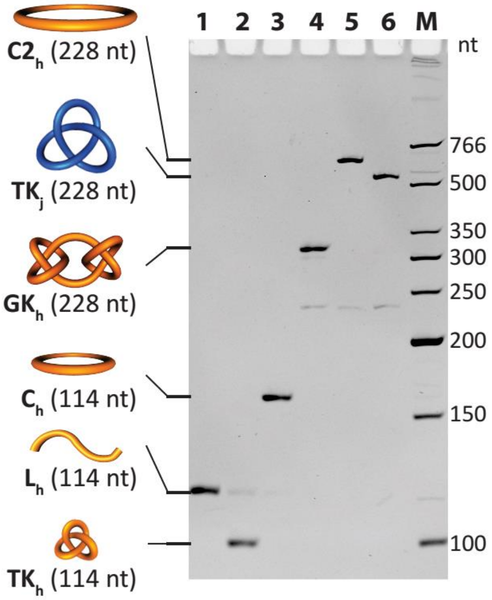

Figure 3.11 dPAGE analyses of various RNA topological structures.

3.2.6 Probing RNA topoisomerase activity

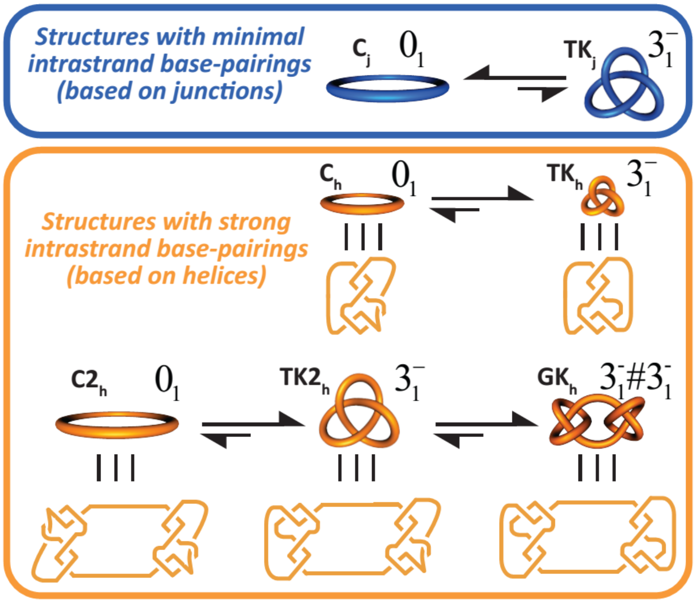

Figure 3.12 Hypothetical conversions of various topological structures under "ideal" RNA Topo conditions.

We tested *E. coli* DNA Topo I (denoted as Topo I afterwards for clarity) for the RNA Topo assay, and in contrast to previous report[4], we find that Topo I indeed possesses RNA Topo activity. (While we were preparing the paper, the RNA Topo activity of Topo I was independently reported[10].) We used our ssRNA negative trefoil knot, **TK_j(-)** (denoted as **TK_j** afterwards for clarity), as the topological probe, which would be converted to the ssRNA circle, **C_j**, in the presence of an RNA Topo activity. Figure 3.12 illustrates the topological conversion of these structures under "ideal" RNA Topo conditions, i.e. when the RNA strand-passage events occur freely and lead to the most thermodynamically stable topoisomers. The most stable topoisomer for the structures free of strong base-pairings (with junction-based method) is that of a simpler topology (Figure 3.12 top). In contrast, the most stable topoisomers of the helix-based structures are those favoring maximum base-pairings and consequently are those of more complex topology (Figure 3.12 middle and bottom). Therefore, **C$_h$**, and **C2$_h$** were used as the helix-based RNA Topo probes to be compared with the junction-based **TK$_j$**.

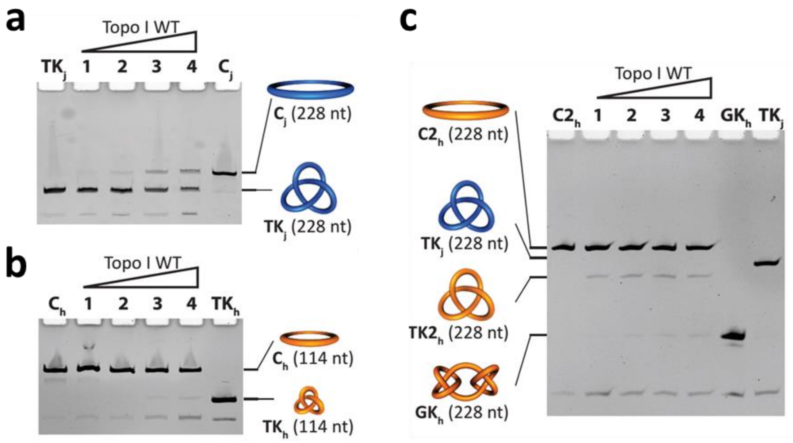

Figure 3.13 Topological relaxations of TK$_j$ (a), C$_h$ (b) and C2$_h$ (c) catalyzed by increasing concentrations of wild-type *E. coli.* DNA Topo I. In **a** and **c**, the RNA probe substrates (**TK$_j$** or **C2$_h$**) were 80 nM and Topo I in lanes 1 to 4 was 40, 80, 160 and 320 nM. In **b**, the RNA probe substrate (**C$_h$**) was 160 nM and Topo I in lanes 1 to 4 was 80, 160, 320 and 640 nM. In e, **C2$_h$** is relaxed to the trefoil knot **TK2$_h$** and to **GK$_h$** after one and two strand passage events respectively. All the reactions were incubated at 37 ℃ for 30 min.

Figure 3.13a shows the increasing conversion of **TK$_j$** to **C$_j$** catalyzed by increasing concentrations of wild-type (WT) Topo I. Though Topo I has RNA Topo activity, this activity is relatively low. Based on our previous result that Topo I catalyzes the fast approximately stoichiometric unknotting of ssDNA knot within 30 minutes[14], the RNA Topo activity of Topo I is estimated to be 1/15~1/12 of the DNA Topo activity. Comparing with **TK$_j$**, **C$_h$** is a much less sensitive RNA Topo probe (Figure 3.13b). Only 3% conversion was observed when the molar ratio of Topo I to RNA is 4:1 (lane 4). Figure 3.13c shows the RNA Topo assay using probe **C2$_h$**. Though **C2$_h$** is more sensitive than **C$_h$**, probably due to the more severe topological stress, it is still not as sensitive as **TK$_j$**. As expected, there are two products of the topological conversion of **C2$_h$**, the larger-amount intermediate trefoil knot **TK2$_h$** (after one strand-passage event), and the smaller-amount final **GK$_h$** (after two strand-passage events). We also find that **TK2$_h$** migrates slightly faster than **TK$_j$**, though both are of the same size and topology, probably due to the different sequences or the formation of transient base-pairings within **TK2$_h$** even during migration in the dPAGE.

The observation that **C$_h$** and **C2$_h$** (helix-based) are not as sensitive probes as **TK$_j$** (junction-based) can be explained by both the Topo I binding to ssRNA and its low RNA Topo activity. Topo I binding is expected to inhibit the formation of base-pairings, countering the thermodynamic driving force for the topological conversion illustrated in Figure 3.12 for **C$_h$** and **C2$_h$**. This problem is further exaggerated due to the low RNA Topo activity, which necessitates a higher concentration of Topo I and ultimately leads to even more severe protein binding. The low sensitivity of the helix-based probe may account for the previous failure of finding RNA Topo activity of Topo I[4]. Besides, better sensitivity would circumvent the inconvenience associated with the use of autoradiograph with ^{32}P-labeled RNA as was in the recent study[10].

3.2.7 Substrate-specificity study of Topo I mutants

In the crystal structure of the Topo I-ssDNA covalent complex[28], domain IV of the enzyme provides several contacts with the DNA substrate and is important for the binding and recognition of the substrate (Figure 3.14a). A key residue is R173, which interacts with the -4 position cytosine base via hydrogen bonding. Previous studies demonstrated that the R173A point mutation displays an approximately 100-fold decrease in the relaxation activity of supercoiled (sc) plasmid DNA[29] and completely loses the ability to relax the helix-based ssRNA probe[10]. With our junction-based probes of both ssDNA (a previously prepared ssDNA trefoil knot, **dTK$_j$**) and ssRNA (**TK$_j$**), the R173A mutant was investigated regarding its relaxation activities of ssDNA and ssRNA. Interestingly, we could still detect the unknotting activity of

R173A mutant for the ssDNA probe (Figure 3.14b), but not for ssRNA (Figure 3.14c). The Y319F mutant was assayed as the negative control, unable to unknot either **dTK$_j$** or **TK$_j$** due to the loss of the active-site tyrosine residue.

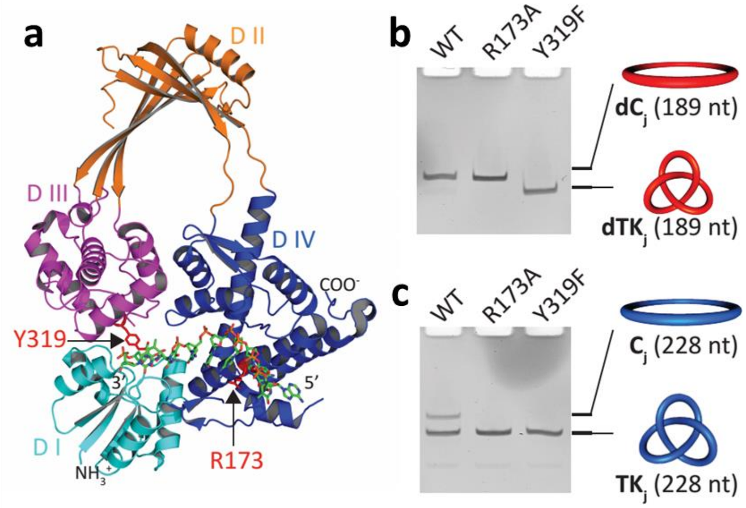

Figure 3.14 Substrate specificity of two Topo I mutants. (**a**) Crystal structure of the Topo I-ssDNA covalent (PDB ID: 3PX7)[28]. The two mutated residues in mutant study, R173 and Y319, highlighted in red. (**b** and **c**) Investigating the ssDNA- and ssRNA-unknotting activity of WT Topo I and two mutants using ssDNA knot, **dTK$_j$** (**b**) and ssRNA knot, **TK$_j$** (**c**) as probes. Probe substrates were 80 nM, treated by 320 nM of proteins (30 min incubation at 37 ℃).

To further determine the ssDNA-unknotting efficiency of the R173A mutant, the concentration-dependent topological conversion assay was conducted with ssDNA probe **dTK$_j$** (Figure 3.15). It turns out that the R173A mutant has almost identical ssDNA unknotting activity as the WT enzyme[14], both catalyzing the approximately stoichiometric topological conversion of ssDNA knot within 30 minutes. The different activities of the WT enzyme and the two mutants are summarized in Table 3.2Table 3.2 Summary of activities of the WT Topo I and two mutants tested in this work.. The supercoiled (sc) plasmid relaxation activities were taken from previous work[29]. This is the first report of an amino acid substitution in topoisomerases that affect DNA and RNA Topo activity differently. Our results imply that the region containing R173 in the domain IV, which was suggested to be important for the sequence selectivity of different Type IA DNA Topo[28], plays an important role in the enzyme's specificity to different nucleic acid substrates, for example, ssDNA, ssRNA and partially unwound dsDNA to different extents (as in sc plasmid). It is also possible that the substrate specificity can be tuned by engineering this region, and more efficient RNA Topo or RNA-specific Topo is envisioned, which can serve as a promising tool for the identification and the consequent concomitant studies of RNA topology.

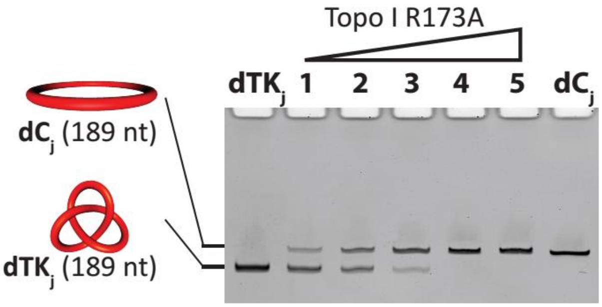

Figure 3.15 Topo I R173A mutant does not show decrease in ssDNA unknotting activity. 80 nM of **dTK$_j$** was incubated with increasing concentrations of Topo I R173A mutant (lanes 1 to 5 was 20, 40, 80, 160 and 320 nM) for 30 min at 37 ℃.

Table 3.2 Summary of activities of the WT Topo I and two mutants tested in this work.

Topo I	WT	R173A	Y319F
ssDNA deknotting	Yes	Yes	No
ssRNA deknotting	Low	No	No
sc plasmid relaxing	Yes	Low	No

3.2.8 Topological inhibition of reverse transcription (RT)

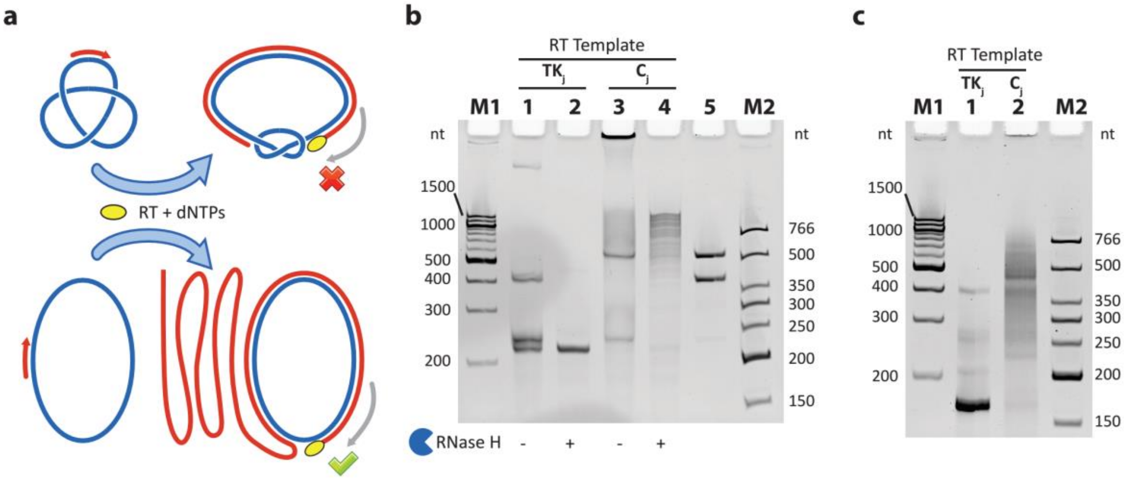

Figure 3.16 Different RT-PCR patterns resulting from RNA of different topologies. (**a**) Schematics illustrating topological blockage of reverse transcription (RT). (**b**) RT reactions of **TK$_j$** (lanes 1 and 2) and **C$_j$** (lanes 3 and 4) analyzed by dPAGE. Samples in lanes 2 and 4 were treated by RNase H. (**c**) PCR reactions of the cDNA products from **TK$_j$** (lane 1) and **C$_j$** (lane 2) analyzed by dPAGE.

DNA topology affects the DNA-templated processes in the living cell, such as DNA replication and transcription[30]. In Chapter 2, we showed that *in vitro* the procession of various DNA polymerases can be blocked on a knotted ssDNA template[14]. Analogously, we expect that RT would also be affected by RNA topology. If a knotted ssRNA is used as the RT template, the RT enzyme can extend the primer only to a certain point, when the increased free energy associated with the tightening of the diminishing ssRNA region and bending of the ds RNA-DNA hybrid region causes the enzyme to stall. As a result, only incomplete product of complementary DNA (cDNA) is generated (Figure 3.16a top). In contrast, on the circular ssRNA template, the RT enzyme, as a strand-displacing polymerase, could synthesize long ssDNA products containing tandem repeats of the cDNA with a rolling-circle fashion (Figure 3.16a bottom). Therefore, we conducted the cDNA synthesis with the ProtoScript® II reverse transcriptase (NEB) on both knotted (**TK$_j$**) and circular (**C$_j$**) ssRNA templates (228 nt) and the reaction mixtures were analyzed by dPAGE (Figure 3.16b). As expected, only incomplete cDNA product (approximately 210 nt) was generated from $\mathbf{TK_j}$, (lanes 1 and 2), and the rolling-circle product (up to approximately 1500 nt) was generated from $\mathbf{C_j}$ (lanes 3 and 4). The faint top band in lane 1 and that sticking to the well in lane 3 are probably due to the strong association of the RT enzyme with the RNA-DNA complex, even after heat denaturation and during dPAGE. After treatment with RNase H, these bands disappear (lanes 2 and 4).

The cDNA products obtained were subsequently subjected to PCR with a pair of convergent primers to amplify a 167-nt region. Figure 3.16c shows that the different RT-PCR patterns from RNA of different topologies. A single major band corresponding to the correct target PCR product was observed for $\mathbf{TK_j}$ (lane 1). However, a smear containing multiple bands was observed with $\mathbf{C_j}$ (lane 2) due to the tandem-repeat sequence of its cDNA. As a result, different RT-PCR patterns have been obtained for ssRNA knot and circle with the same sequence. This topology-dependent RT-PCR feature can be utilized to identify closed RNA knots from the naturally occurring circular RNAs[31]. Assays of screening other proteins for RNA Topo activity are also envisioned.

3.2.9 Conclusions

Today, the exciting field of RNA nanotechnology is beginning to emerge[32,33]. However, it is unfortunate that there was virtually no progress on RNA topological structures after Seeman's first construction of RNA knot. Synthetic RNA topology will certainly catalyze the further developments of this field. In fact, considering RNA's structural and functional diversity, future work is likely to yield a plethora of design strategies and practical applications of synthetic RNA topology. Our current work, serving as a starting point, has greatly expanded the richness of

RNA topological structures, including the first realization of RNA topoisomers of both positive and negative trefoil knots, and two different forms of RNA-DNA hybrid structures.

Furthermore, the tools and methods that have been demonstrated or suggested in our work can help solve several unexplored problems associated with RNA topology. Aside from searching for naturally occurring RNA topological structures, whether or not there exists RNA-specific Topo is also a fundamental question worth pursuing. Though RNA Topo activity has been spotted in some proteins, there is to date no evidence supporting the cellular functions of these proteins as being directly related to the RNA Topo activity. Hypothetically, RNA Topos may correct the misfolded RNA structures or resolve RNA entanglements that could inhibit the normal functions. If so, this will open new opportunities for fundamental RNA biochemical and biophysical research, as well as novel therapeutic inventions.

3.3 Methods

3.3.1 RNA preparation

The design of the sequences followed the rule of sequence symmetry minimization[34] and was assisted by the program CANADA[35]. All RNA molecules were synthesized by *in vitro* transcription using the HiScribe™ T7 High Yield RNA Synthesis Kit from the New England Biolabs (NEB). The corresponding DNA templates were generated by the PCR amplification of the gBlocks® gene fragments from the Integrated DNA Technologies (IDT) using the Q5® Hot Start High-Fidelity DNA Polymerase (NEB). To enhance the ribozyme cleavage, five thermal cycles were performed after transcription, with each cycle containing three steps: 70 ℃ for 10 seconds, 50 ℃ for 1 minute, and 37 ℃ for 10 minutes. The target RNA molecules were then purified by the dPAGE. Each purified RNA molecule was treated with T4 polynucleotide kinase

(NEB) in 1×T4 DNA Ligase Buffer (NEB, 1×buffer: 50 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, pH 7.5 @ 25 ℃) at 37 ℃ for 6 hours to remove the 2',3'-cyclic phosphate[16] and to add a phosphate to 5'-hydroxyl end. After the treatment, these RNA molecules were directly used as RNA scaffolds for the topological construction.

3.3.2 Topological construction

The ssRNA topological constructs were prepared with the all-in-one protocol. It involves two steps: (1) Annealing to get the assembly complex. Equimolecular quantities (normally with a final concentration of 1 μM each) of all strands (RNA scaffolds, DNA staples and splints) were mixed in a buffer with the ultimate concentration of 1×T4 DNA Ligase Buffer (by adding 10×T4 DNA Ligase Buffer because the kinase treated RNA scaffolds were in 1×Buffer) and annealed from 70 ℃ to 16 ℃ over 4 hours. (2) Ligation to seal the nicks. To each 100 μL of reaction mixture, 4 μL of T4 RNA ligase 2 (NEB, 10 U/μL), 1.5 μL of 100 mM fresh DTT (NEB) and 1.5 μL of 100 mM fresh ATP (NEB) were added and incubated at 16 ℃ at least 16 hours for the ligation. The ds RNA-DNA hybrid structures were prepared using the corresponding ssDNA knots as templates, which were prepared according to previous work[14]. The complementary RNA strands were designed as 3 substrands, which were annealed to the ssDNA knot templates with a ratio of complementary:template = 1.2:1. After the annealing, T4 RNA ligase 2 was used to seal the nicks in complementary RNA strands.

3.3.3 dPAGE

Gels of different concentrations were prepared using 30% acrylamide and bis-acrylamide solution (Bio-Rad, 29:1) with 7 M urea in 0.5×TBE buffer (Bio-Rad) and run on a PROTEAN® II xi cell (Bio-Rad) or a Mini-PROTEAN® Tetra cell (Bio-Rad). Samples were mixed 1:1 with TBE-Urea Sample Buffer (Bio-Rad) and heated at 70 ℃ for 5 min before they were loaded in the wells. Gel concentrations were carefully chosen to ensure the proper separations between different topologies as well as the references. For imaging, gels were stained with GelRed™ (Biotium), and images were taken by Gel Doc™ XR+ (Bio-Rad) imaging system and processed by software Image Lab™ (v.4.0.1, Bio-Rad). For purification, gels (without staining) were visualized by UV shadowing against a fluorescent TLC plate and bands of interest were cut. The bands were then eluted using the crush-and-soak method and the eluent was purified and concentrated on 3K Nanosep® filters (Pall). The concentration of product was determined by measuring the OD_{260}. Optionally, the ssRNA knots and circles can be digested by RNase R (Epicentre) after the gel extraction to remove the unavoidable cleaved linear RNA during the purification. However, RNase R digestion is not useful for the hybrid BR.

3.3.4 Digestion with various nucleases

Various nucleases were used in this work, including Nt.AlwI (NEB), Nt.BspQI (NEB), RNase H (NEB), DNase I (NEB), RNase R (Epicentre). We used the reaction conditions for these enzymes as recommended by the providers.

3.3.5 Topoisomerase assay

The *E. coli* Topo I WT and mutant proteins were expressed and purified as described in previous study[36]. Commercial product of *E. coli* Topo I (NEB) was also tested and RNA Topo activity was found to be slightly higher than the in-house prepared WT enzyme, probably due to the contamination of *E. coli* Topo III (Supplementary Fig. 7). The reaction buffer contained 1×NEBuffer 4 (NEB, 1×buffer: 50 mM KOAc, 20 mM Tris-acetate, 10 mM $Mg(OAc)_2$, 1 mM DTT, pH 7.9 @ 25 ℃) and 100 µg/mL BSA. The concentration of substrates and proteins were described in the text. Reactions were quenched by phenol-chloroform extraction followed by ethanol precipitation. The reactions were then analyzed by dPAGE.

3.3.6 AFM imaging

For AFM imaging of the ds RNA-DNA hybrid structures, 30 µL of 0.1 mg/mL polyornithine (Sigma) solution was added to freshly cleaved mica and stand for 3 min to increase the binding to the structures before applying the samples. Then the mica was rinsed with 1 mL water and dried with air. 5 uL of each sample (about 5 nM) in 1×TAE-Mg buffer (11 mM $MgCl_2$, 40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.0) was applied to the treated mica and stand for 1 min. Then the mica was rinsed with 1 mL water and dried with compressed air. AFM imaging was performed on a Veeco 8 AFM in the ScanAsyst® in Air mode using the scanayst-air tips (Veeco). The AFM images were processed with the software Gwyddion.

3.3.7 RT-PCR

The ProtoScript® II First Strand cDNA Synthesis Kit (NEB) was used for the RT. Each reaction (20 µL) contained 100 nM ssRNA template ($\mathbf{TK_j}$ or $\mathbf{C_j}$) and 200 nM primer (sequence is shown in Supplementary Information) and followed the recommended protocol. After RT, the enzyme was inactivated at 80 ℃ for 5 min. Then the reactions were treated by RNase H. For the subsequent PCR, HotStart-IT® FideliTaq™ DNA Polymerase (Affymetrix) was used and 1 µL of each RT mixture was added to each 50 µL of PCR reaction for amplification.

3.4 References

1. Micheletti, C., Di Stefano, M. & Orland, H. Absence of knots in known RNA structures. *Proc. Natl. Acad. Sci. USA* 2015, **112**(7)**:** 2052-2057.
2. Burton, A. S., Di Stefano, M., Lehman, N., Orland, H. & Micheletti, C. The elusive quest for RNA knots. *RNA Biol.* 2016, **13**(2)**:** 134-139.
3. van Batenburg, F. H. D. PseudoBase: a database with RNA pseudoknots. *Nucleic Acids Res.* 2000, **28**(1)**:** 201-204.
4. Wang, H., Di Gate, R. J. & Seeman, N. C. An RNA topoisomerase. *Proc. Natl. Acad. Sci. USA* 1996, **93**(18)**:** 9477-9482.

5. Mueller, J. E., Du, S. M. & Seeman, N. C. Design and synthesis of a knot from single-stranded DNA. *J. Am. Chem. Soc.* 1991, **113**(16)**:** 6306-6308.

6. Du, S. M. & Seeman, N. C. Synthesis of a DNA knot containing both positive and negative nodes. *J. Am. Chem. Soc.* 1992, **114**(24)**:** 9652-9655.

7. Du, S. M., Stollar, B. D. & Seeman, N. C. A synthetic DNA molecule in three knotted topologies. *J. Am. Chem. Soc.* 1995, **117**(4)**:** 1194-1200.

8. Mao, C., Sun, W. & Seeman, N. C. Assembly of Borromean rings from DNA. *Nature* 1997, **386**(6621)**:** 137-138.

9. Xu, D. *et al.* Top3β is an RNA topoisomerase that works with fragile X syndrome protein to promote synapse formation. *Nature Neurosci.* 2013, **16**(9)**:** 1238-1247.

10. Ahmad, M. *et al.* RNA topoisomerase is prevalent in all domains of life and associates with polyribosomes in animals. *Nucleic Acids Res.* 2016.

11. Seeman, N. C. Synthetic DNA Topology. In: Sauvage JP, Dietrich-Buchecker C (eds). *Molecular Catenanes, Rotaxanes and Knots*. Wiley: Weinheim, 2007, pp 323-356.

12. Jester, S. S. & Famulok, M. Mechanically interlocked DNA nanostructures for functional devices. *Acc. Chem. Res.* 2014, **47**(6)**:** 1700-1709.

13. Lu, C. H., Cecconello, A. & Willner, I. Recent Advances in the Synthesis and Functions of Reconfigurable Interlocked DNA Nanostructures. *J. Am. Chem. Soc.* 2016, **138**(16)**:** 5172-5185.

14. Liu, D., Chen, G., Akhter, U., Cronin, T. M. & Weizmann, Y. Creating complex molecular topologies by configuring DNA four-way junctions. *Nature Chem.* 2016.

15. Avis, J. M., Conn, G. L. & Walker, S. C. Cis-acting ribozymes for the production of RNA in vitro transcripts with defined 5' and 3' ends. *Methods. Mol. Biol.* 2012, **941:** 83-98.

16. Wichlacz, A., Legiewicz, M. & Ciesiolka, J. Generating in vitro transcripts with homogenous 3' ends using trans-acting antigenomic delta ribozyme. *Nucleic Acids Res.* 2004, **32**(3)**:** e39.

17. Ko, S. H. *et al.* Synergistic self-assembly of RNA and DNA molecules. *Nature Chem.* 2010, **2**(12)**:** 1050-1055.

18. Wang, P., Ko, S. H., Tian, C., Hao, C. & Mao, C. RNA-DNA hybrid origami: folding of a long RNA single strand into complex nanostructures using short DNA helper strands. *Chem. Commun.* 2013, **49**(48)**:** 5462-5464.

19. Liu, D., Wang, M., Deng, Z., Walulu, R. & Mao, C. Tensegrity: construction of rigid DNA triangles with flexible four-arm DNA junctions. *J. Am. Chem. Soc.* 2004, **126**(8)**:** 2324-2325.

20. Birac, J. J., Sherman, W. B., Kopatsch, J., Constantinou, P. E. & Seeman, N. C. Architecture with GIDEON, a program for design in structural DNA nanotechnology. *J. Mol. Graph. Model.* 2006, **25**(4)**:** 470-480.

21. Hall, K., Cruz, P., Tinoco, I., Jovin, T. M. & van de Sande, J. H. 'Z-RNA'—a left-handed RNA double helix. *Nature* 1984, **311**(5986)**:** 584-586.

22. Cantrill, S. J., Chichak, K. S., Peters, A. J. & Stoddart, J. F. Nanoscale borromean rings. *Acc. Chem. Res.* 2005, **38**(1)**:** 1-9.

23. Chichak, K. S. *et al.* Molecular borromean rings. *Science* 2004, **304**(5675)**:** 1308-1312.

24. Huang, S. L., Lin, Y. J., Hor, T. S. & Jin, G. X. Cp*Rh-based heterometallic metallarectangles: size-dependent Borromean link structures and catalytic acyl transfer. *J. Am. Chem. Soc.* 2013, **135**(22)**:** 8125-8128.

25. Thorp-Greenwood, F. L., Kulak, A. N. & Hardie, M. J. An infinite chainmail of M6L6 metallacycles featuring multiple Borromean links. *Nature Chem.* 2015, **7**(6)**:** 526-531.

26. Kim, T. *et al.* Selective Synthesis of Molecular Borromean Rings: Engineering of Supramolecular Topology via Coordination-Driven Self-Assembly. *J. Am. Chem. Soc.* 2016.

27. Niu, J., Hili, R. & Liu, D. R. Enzyme-free translation of DNA into sequence-defined synthetic polymers structurally unrelated to nucleic acids. *Nature Chem.* 2013, **5**(4)**:** 282-292.

28. Zhang, Z., Cheng, B. & Tse-Dinh, Y. C. Crystal structure of a covalent intermediate in DNA cleavage and rejoining by Escherichia coli DNA topoisomerase I. *Proc. Natl. Acad. Sci. USA* 2011, **108**(17)**:** 6939-6944.

29. Narula, G. & Tse-Dinh, Y. C. Residues of E. coli topoisomerase I conserved for interaction with a specific cytosine base to facilitate DNA cleavage. *Nucleic Acids Res.* 2012, **40**(18)**:** 9233-9243.

30. Bates, A. D. & Maxwell, A. *DNA topology*, 2 edn. Oxford Univ. Press: Oxford, 2005.

31. Frank-Kamenetskii, M. D. RNA topology. *Artificial DNA: PNA & XNA* 2014, **4**(2)**:** 35-36.

32. Guo, P. The emerging field of RNA nanotechnology. *Nature Nanotechnol.* 2010, **5**(12)**:** 833-842.

33. Grabow, W. W. & Jaeger, L. RNA self-assembly and RNA nanotechnology. *Acc Chem Res* 2014, **47**(6)**:** 1871-1880.

34. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* 1982, **99**(2)**:** 237-247.

35. Feldkamp, U. CANADA: Designing Nucleic Acid Sequences for Nanobiotechnology Applications. *J. Comput. Chem.* 2010, **31**(3)**:** 660-663.

36. Narula, G. *et al.* The strictly conserved Arg-321 residue in the active site of Escherichia coli topoisomerase I plays a critical role in DNA rejoining. *J. Biol. Chem.* 2011, **286**(21)**:** 18673-18680.

Chapter 4 – Controlling the curvature and torsion of RNA assemblies via artificially designed branched kissing loop

4.1 Introduction

Two geometric parameters– curvature and torsion – at each point on a space curve determine the overall shape of the curve. Curvature measures the bending of the curve; and torsion measures the twisting of the curve from the plane of curvature (i.e. the osculating plane). Obviously, arbitrary 1D string-like nanostructures (such as tubes, ribbons or rods) can be created if the curvature and torsion can be conveniently engineered at the nanoscale dimensions. The programmable self-assembly of nucleic acids is one of the very few promising approaches to the control of curvature and torsion. For example, Shih and coworkers have demonstrated this concept using 3D DNA origami structures, where bundles of DNA helices are arranged in the honeycomb lattice[1]. In their approach, the number of base pairs (bps) is adjusted in selected regions so that the structures have to be bent or twisted in order to release the strain due to the deviation from 10.5 bps per turn of the ideal B-form helix. However, the technique of DNA origami, though elegant, is still too complicated and costly. A simpler system needs to be developed for this purpose.

Here we address this problem by designing RNA tiles that can self-assemble into various ladder-like structures. The cohesion between individual tiles is through a de novo designed RNA structural motif. This motif is dubbed branched kissing loop (bKL) because it is branched by being formed via the Watson-Crick base-pairing interactions between a hairpin loop and a bulge. The design of this new motif is inspired by the artificially designed DNA T-junction[2] and the naturally occurring HIV-1 DIS RNA kissing loop (KL)[3], both of which are also successfully employed in the construction of various nanostructures[2,4-8]. Therefore, the bKL combines the advantages of both T-junction and KL. On the one hand, it is branched and therefore suitable for the design of complex nanostructures. On the other hand, the cohesion is paranemic, which eliminates sticky ends and thereby decreases the number of unique strands necessary. Furthermore, the control of curvature and torsion is readily realized by adjusting the number of bps in the helical regions of each tile based on the bKL.

4.2 Results and discussion

4.2.1 Design of the bKL motif

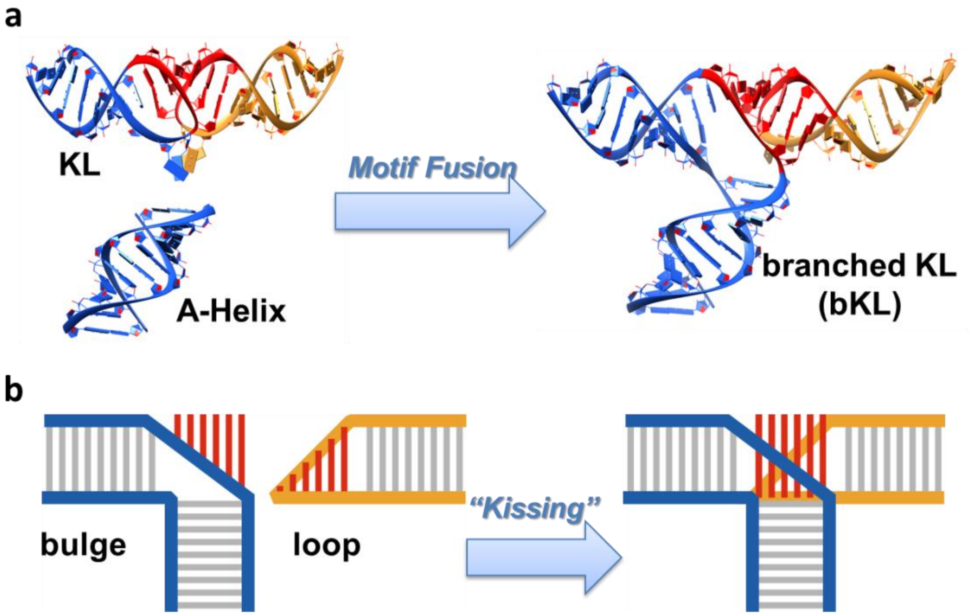

Figure 4.1 Design of the bKL. (**a**) The bKL is designed by fusing an A-helix to a KL. (**b**) Schematic illustrating the formation of bKL.

In the structure of the HIV-1 DIS KL[3], two purines are ahead of the 5' side of 6-nt complementary region in each hairpin loop and span a distance of ~15 Å. This distance is very close to the width of an A-form RNA helix (~18 Å). The bKL is designed by an operation of motif fusion (Figure 4.1a). The two purines in one hairpin loop of the original KL are replaced with an A-form RNA helix and the other hairpin loop is kept unchanged. Consequently, the paranemic kissing interaction is formed by the 6-bp complementary regions between the bulge (or internal loop) and the hairpin loop (or external loop) (Figure 4.1b). This bKL motif is expected to have a T-shaped structure with the stacking between the stems and kissing-loop region in the original KL preserved. The kissing interaction between the bulge and loop is via Watson-Crick base-pairings and therefore has the advantage of specificity and programmability.

4.2.2 Control of torsion

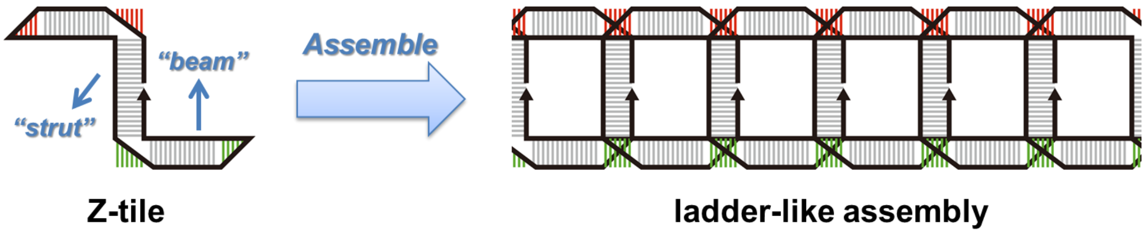

Figure 4.2 Z-tile based on the bKL self-assembles into ladder-like nanostructure.

Tiles folded from a single strand of RNA can be designed to be able to self-assemble via the cohesion of the bKLs. A designed Z-shaped tile (Z-tile) containing two loop-bulge pairs is shown in Figure 4.2, which can self-assemble into a ladder-shaped structure. In this tile, the helices belonging to the stem-loops are dubbed the "beams" and the helix connecting the two bulges is dubbed the "strut". By adjusting the length of both beams simultaneously, the dihedral angle between adjacent tiles can be tuned and therefore the torsion of the resulting assembly can be controlled. Each bKL motif contributes to the helical twist equivalent to an 8-bp A-from RNA helix. Therefore, helical ladder structures with either left- or right-handed torsion can be assembled from the Z-shaped tiles containing the beam helices of 13 or 15 bps, respectively (Figure 4.3). From CryoEM images, the left-handed helical ladder contains 16-18 tiles per helix turn and the right-handed ladder contains 17-21 tiles per helix turn. Furthermore, the right-handed ladder appears to have more flexibility than the left-handed ladder.

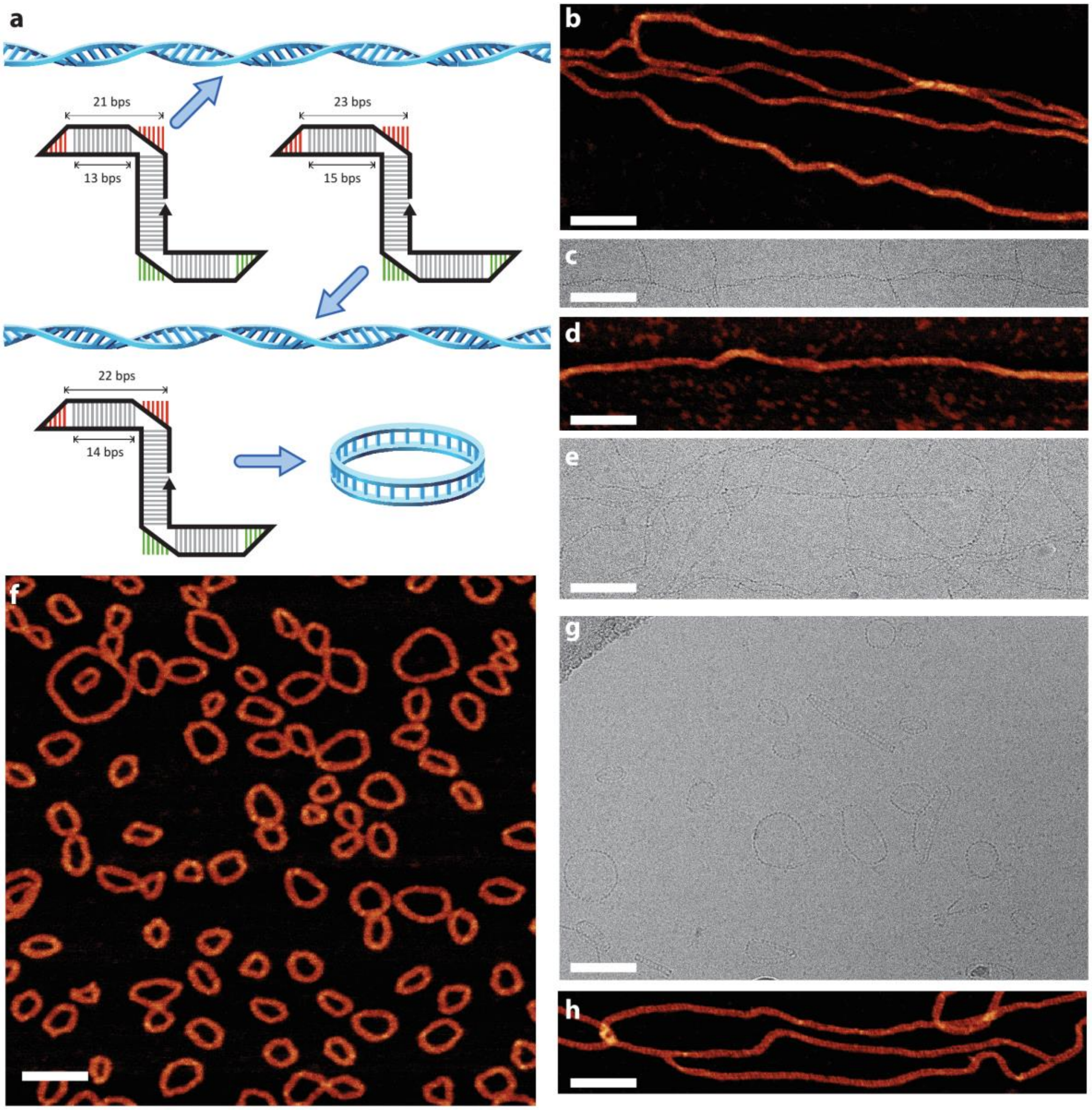

Figure 4.3 Various structures assembled from different Z-tiles. (**a**) Torsion can be controlled by changing the length of beams. (**b** and **c**) Left-handed helical ladder is formed with the 13-bp-beamed Z-tile. (**d** and **e**) Right-handed helical ladder is formed with the 15-bp-beamed Z-tile. (**f** and **g**) In the absence of torsion, ring structures are formed with the 14-bp-beamed Z-tile. (**h**) Long assembled structures occasionally observed by AFM show the minimal torsion of the assembly. Scale bar: 100 nm.

By setting both beams in the tile to be 14 bps, the dihedral angle between adjacent tiles is zero because the helical twist in between is two turns of helix (22 bps). We expected to observe the assembled ladder structure with minimal torsion. However, the majority of the structures are closed rings (Figure 4.3f and g). This fact indicates that the bKL motif has intrinsic bending and thereby introduces the out-of-plane curvature between the adjacent assembling tiles. In the absence of torsion, the curvature can be accumulated and the rings are formed as a result. The minimal torsion of the design is also reflected by the occasionally observed longer assembled structures (Figure 4.3h), which contain no crossover point over a very long tract of the assembly.

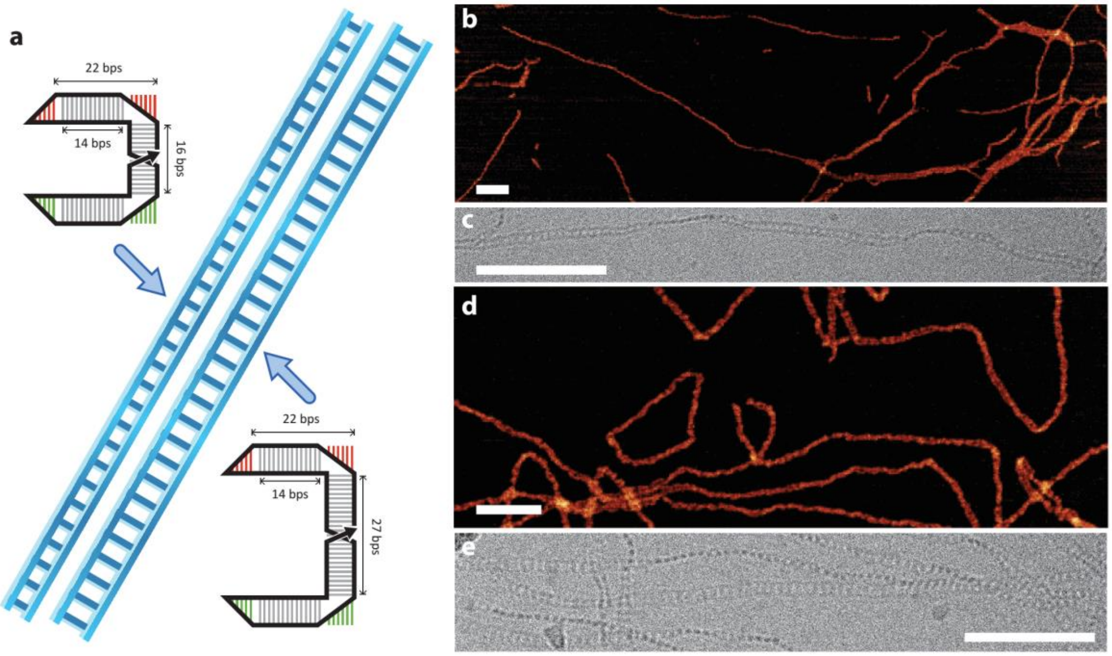

Figure 4.4 Assemblies of the C-tiles. (**a**) Ladders with minimal torsion and curvature are formed by the C-tiles. (**b** and **c**) Ladder formed by the 16-bp-strutted C-tile. (**d** and **e**) Ladder formed by the 27-bp-strutted C-tile. Scale bar: 100 nm.

To cancel out the accumulated bending in the absence of torsion, C-shaped tiles (C-tiles) can be designed with the length of the strut helix close to an odd number of helical half-turns. Accordingly, two different C-tiles are designed with the strut helix to be 16 or 27 bps respectively (Figure 4.4 a). With the use of C-tiles, the formation of rings can be greatly discouraged because the out-of-plane curvatures along the two beam helices are of the opposite directions. Therefore, the C-tiles assemble into ladder structures with minimal torsion and curvature (Figure 4.4 b to d). Though very rarely rings are formed, they are very large, which is unlike the case of Z-tile. Interestingly, bundles were observed in the AFM image of the structures assembled from the 16-bp-strutted C-tile (Figure 4.4 b), but they were not observed in by the CryoEM (Figure 4.4 c). These bundles are likely due to the stacking of the struts of individual ladders during the sample preparation. This fact also reflects that the assembled structures are almost free of torsion and curvature.

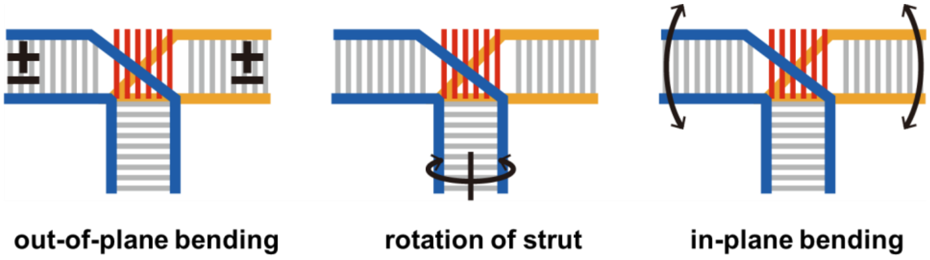

Figure 4.5 Flexibility of the bKL motif.

The structures created thus far indicate that the bKL motif is intrinsically flexible from three aspects (Figure 4.5). Firstly, its flexibility of the out-of-plane bending along the beams is reflected by the various sizes of the closed rings assembled from the 14-bp-beamed Z-tile. Secondly, the strut has some freedom of rotation along its axis. In the designed 13-, 14- and 15-bp-beamed Z-tiles, the struts are of the same length (21 bps) and rotational flexibility of the strut is necessary to accommodate the different torsions in the resulting assembled structures. Finally, the bKL also has in-plane bending flexibility along the beams. The in-plane bending is necessary in the formation of the left- and right-handed helical ladders for the curvature of beams to follow the trajectory of a circular helix.

4.2.3 DNA version of the bKL motif

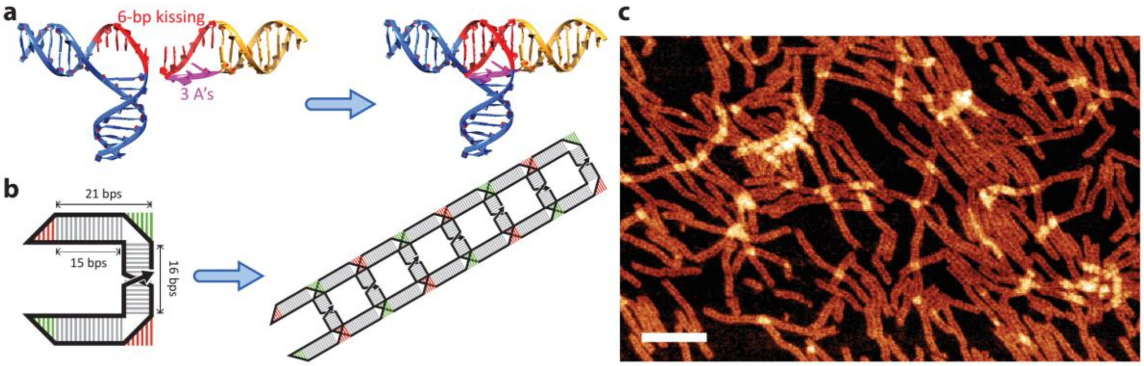

Figure 4.6 DNA bKL motif. (**a**) Design of the DNA bKL. (**b**) Ladder formed by the 16-bp-strutted C-tile. (**d** and **e**) Ladder formed by the 27-bp-strutted C-tile. Scale bar: 100 nm.

The design of the bKL motif can further be extended to DNA, but with a slight variation (Figure 4.6a). The DNA bKL is formed by the Watson-Crick base-pairings between a 6-nt bulge with a 6-nt region in a 9-nt loop. The other three nucleotides in the loop are three unpaired dA on the 5’ side. Because each nucleotide spans ~6 Å, these three nucleotides serve to bridge the distance of ~18 Å between the 3’ end of the stem and the 5’ end of 6-nt kissing region. Each DNA bKL motif contributes to the helical twist equivalent to a 6-bp B-from DNA helix. Based on this DNA bKL, a 15-beamed C-tile is designed with a single strand of DNA (Figure 4.6b). The ladder structure with minimal twist is assembled from this DNA C-tile (Figure 4.6c).

4.2.4 Control of the in-plane curvature

The capability of in-plane bending of the bKL can be further utilized to design closed annulus structures by using Z-tiles containing the two beams with different lengths. Figure 4.7a shows a Z-tile with 14/25-bp beams. Accordingly, each tile contributes 2 helical turns (22 bps) to the inner circle and 3 helical turns (33 bps) to the outer circle of the assembly. This 11-bp (corresponding to one helical turn) difference between the two beams causes the overall bending of assembled structures toward the side of the shorter beam. Geometrical analysis shows that an ideal planar annulus should contain about 15 tiles. However, due to the intrinsic flexibility of the bKL and the kinetics of the assembly, annuli of different sizes were observed in our experiment (Figure 4.7b and c). From the CryoEM image, the assembled structures containing more or fewer tiles were observed to be distorted from the plane in order to release the strain: the assemblies containing fewer tiles have the geometry of a conical frustum, and those containing more tiles are curved into boat-like structures.

To further construct annuli with even larger in-plane curvature, a 14/36-bp-beamed Z-tile has been designed. It was expected that smaller annuli would be assembled from about 8 tiles due to the larger length difference between the two beams. Interestingly, two distinct structures are assembled with the same tile (Figure 4.7). One is the expected smaller annulus; the other is a long linear structure, which appears to be thicker than the annulus in the AFM image. The latter can be explained by the formation of a two-layered structure from the 14/36-bp-beamed Z-tile. One side (the single-layered side) of this structure is formed by the 14-bp beam of each tile connected by the bKLs and each tile contributes two helical turns on this side. The other side (the double-layered side) contains two layers, both of which are formed by the connection of the 36-bp beam of every other tile via the bKLs (each tile contributes four helical turns, which are compensated by two tiles in the single-layered side). Passing the annealing mixture through the filter with a pore size of 0.22 µm can remove most of the two-layered structures, leaving the annuli as the majority in the filtrate (Figure 4.7f and g), though with a small amount of short fragments and circularized species of the two-layered structures.

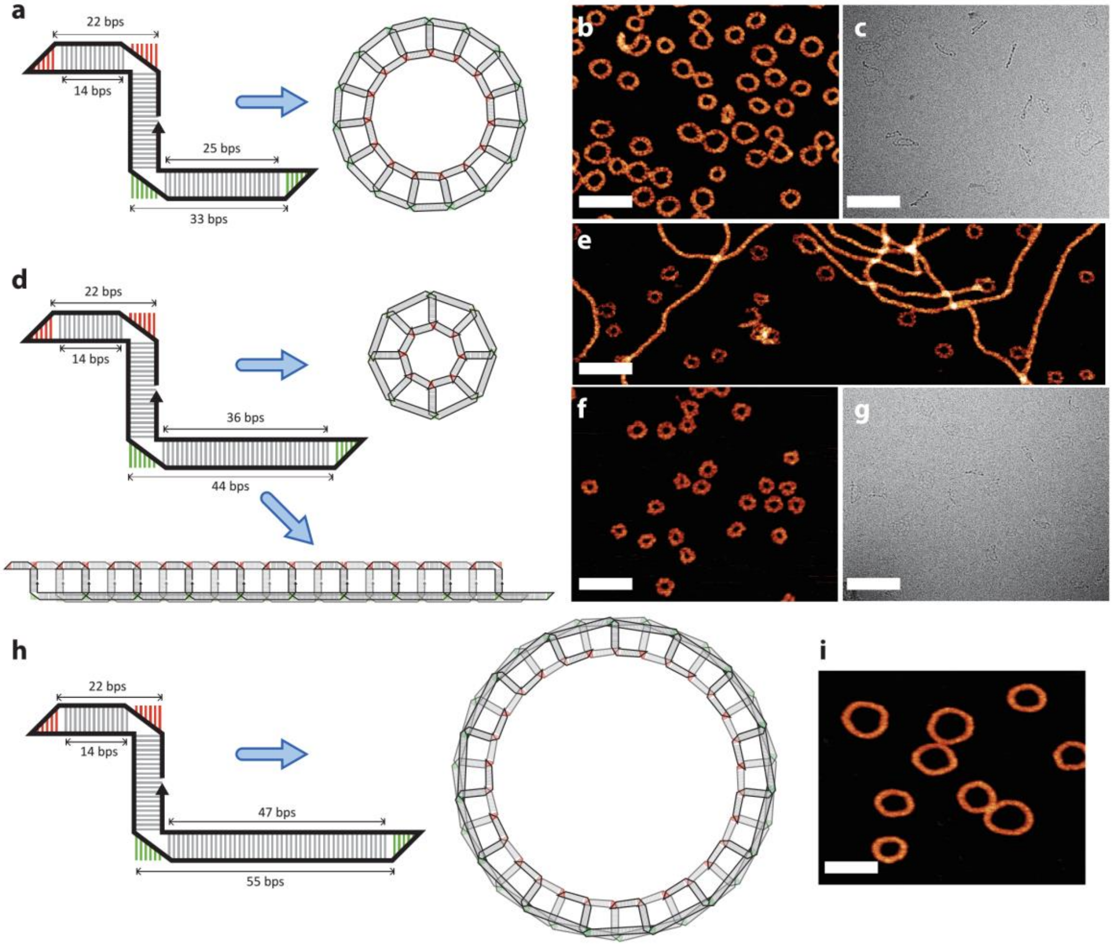

Figure 4.7 In-plane curvature generated by tiles with beams of different length. (**a**) 14/25-bp-beamed Z-tile. (**b** and **c**) 2D annulus structures are assembled from the14/25-bp-beamed Z-tile. (**d**) 14/36-bp-beamed Z-tile. (**e**) Two different kinds of structures are assembled from the14/36-bp-beamed Z-tile. (**f** and **g**) The smaller 2D annulus structures are the major species after passing through the filter with a pore size of 0.22 µm. (**h**) 14/47-bp-beamed Z-tile. (**i**) The two-layered larger annulus structures assembled from the14/47-bp-beamed Z-tile. Scale bar: 100 nm.

The formation of two different structures from the 14/36-bp-beamed Z-tile reflects the reluctance of the bKL to accommodate so large in-plane curvature. In fact, this might be near the maximum in-plane curvature that the bKL can achieve. As for the 14/47-bp-beamed Z-tile (Figure 4.7h), only the two-layered annuli are formed, instead of the even smaller and more curved single-layered annuli. In each two-layered annulus, the structure is curved to the single-layered side because one tile contributed 5 helical turns to the double-layered side, one-turn longer than the 4 turns contributed by two tiles to the single-layered side.

4.2.5 Combining torsion and curvature

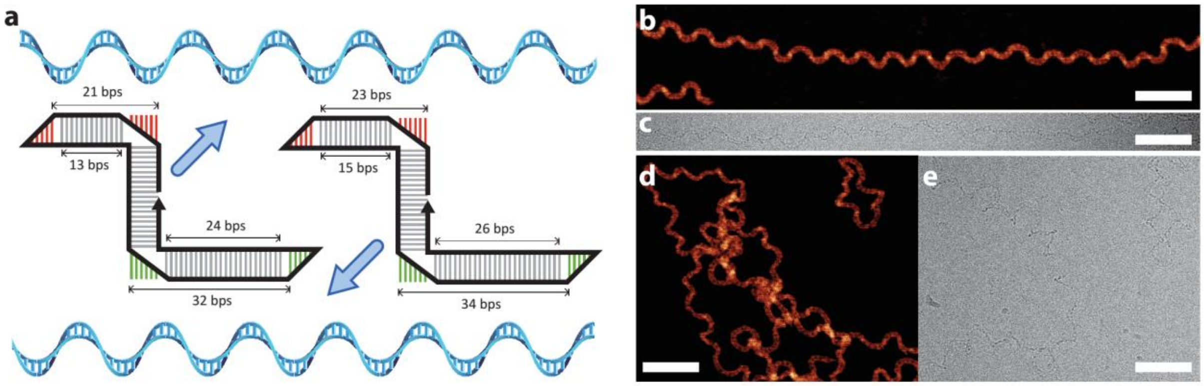

Figure 4.8 Combining torsion and curvature to generate ramen-like structures. (**a**) 13/24-bp-beamed and 15/26-bp-beamed Z-tiles. (**b** and **c**) Left-handed nano-ramen assembled from the 13/24-bp-beamed Z-tile. (**d** and **e**) Right-handed nano-ramen assembled from the 15/26-bp-beamed Z-tile. Scale bar: 100 nm.

More complex structures can be created by incorporating both torsion and curvature. Based on the annulus structure assembled from the 14/25-bp-beamed Z-tile, left- or right-handed torsion can be added to the structure by removing or adding one bp from both beams of the tile (Figure 4.8a). The introduced torsions prevent the tiles assembling into closed structure. As a result, ramen-like spiral structures (nano-ramen) are created. In each nano-ramen, the two beams approximately make a pair of Bertrand curves. Compared to the left-handed nano-ramen assembled from the 13/24-bp-beamed Z-tile (Figure 4.8b and c), the right-handed nano-ramen assembled from the 15/26-bp-beamed Z-tile is more flexible as is shown by the larger chance of formation of smaller circularized structures (Figure 4.8d and e).

4.2.6 Promoting out-of-plane curvature for the formation of nanocage

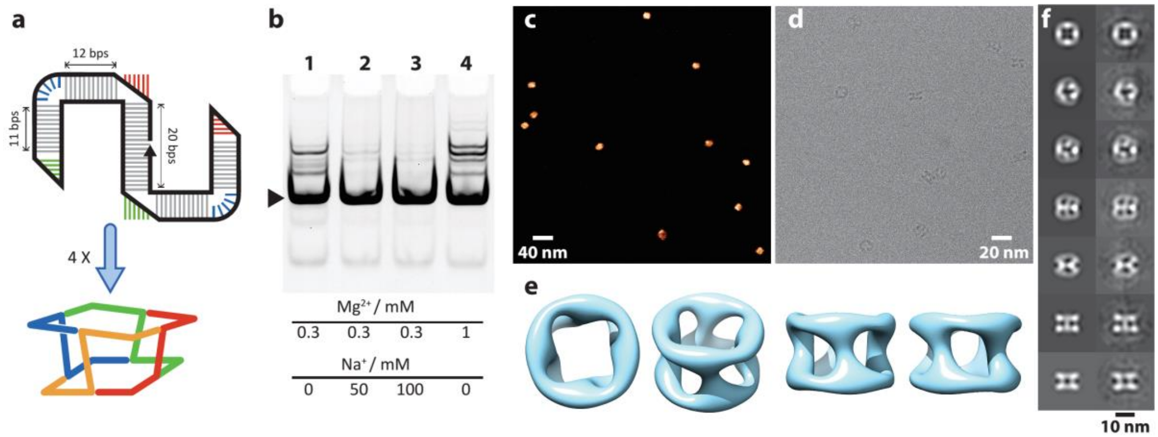

Figure 4.9 A self-assembled tetrameric nanocage. (**a**) A 90°-angle bulge is designed in the beams of the Z-tile to promote the out-of-plane curvature for the formation of a tetrameric cage. (**b**) Native PAGE showing the optimization of the assembly by changing the salt concentration in the annealing buffer. The band corresponding to the target tetrameric nanocage is marked by a black triangle. (**c** and **d**) Characterization of the tetrameric cage by AFM (**c**) and CryoEM (**d**). (**e**) Different views of the structural model of the nanocage generated by CryoEM single-particle reconstruction. (**f**) Pair-wise comparison between the representative 2D projections of the reconstructed 3D model (left) and the corresponding class-averages of the individual particles (right).

By incorporating curved structural motifs in the beams of the tile, the out-of-plane curvature can be further increased and enclosed nanocages can be formed. The 5-nt (AACUA) bulge in the domain IIa of the internal ribosome entry site (IRES) of the hepatitis C virus (HCV) RNA bends with a 90° angle[9]. This bulge motif has been utilized previously to construct various RNA nanostructures[10,11]. Here, we have inserted this 90°-angle bulge into each beam of a designed Z-tile (Figure 4.9a). Consequently, four copies of this kind of Z-tile are expected to assemble into a tetrameric nanocage. However, the intrinsic flexibility of the bKL motif poses a considerable obstacle for the production of nanocages with uniform size. In our experiment, we found that the composition of buffer in which the RNA is annealed plays a very important role in obtaining uniform assembly products (Figure 4.9b). In the annealing buffer containing 100 mM of Na^+ and

0.3 mM of Mg^{2+} (lane 3), more than 95% yield for the target tetrameric nanocage can be achieved. As far as we know, our experiment is the first demonstration of the importance of the buffer components in optimizing the self-assembly of RNA nanostructures, which has been largely overlooked in most previous research. Because both thermodynamic and kinetic factors are important in the assembly of the tiles, tuning the concentrations of different cation species in the buffer can finely orchestrate the intramolecular folding and intermolecular assembling processes during the annealing. This argument may also be extended into the tile-based DNA nanocages.

The assembled tetrameric nanocage is visualized by the AFM image, which shows square-like particles with uniform size (Figure 4.9c). Its successful formation is further proved by CryoEM image (Figure 4.9d). Figure 4.9e shows the structural model of the nanocage obtained with the single-particle reconstruction technique. The pair-wise comparison between the representative 2D projections of the reconstructed 3D model and the corresponding class-averages of the individual particles is shown in Figure 4.9f. A more complete comparison is shown in Figure 4.10. Though a handful of RNA nanocages have been previously constructed using multiple tiles or using a single tile formed by multiple RNA strands, to the best of our knowledge, this is the first example of obtaining highly uniform oligomeric nanocage assembled with a single strand of RNA.

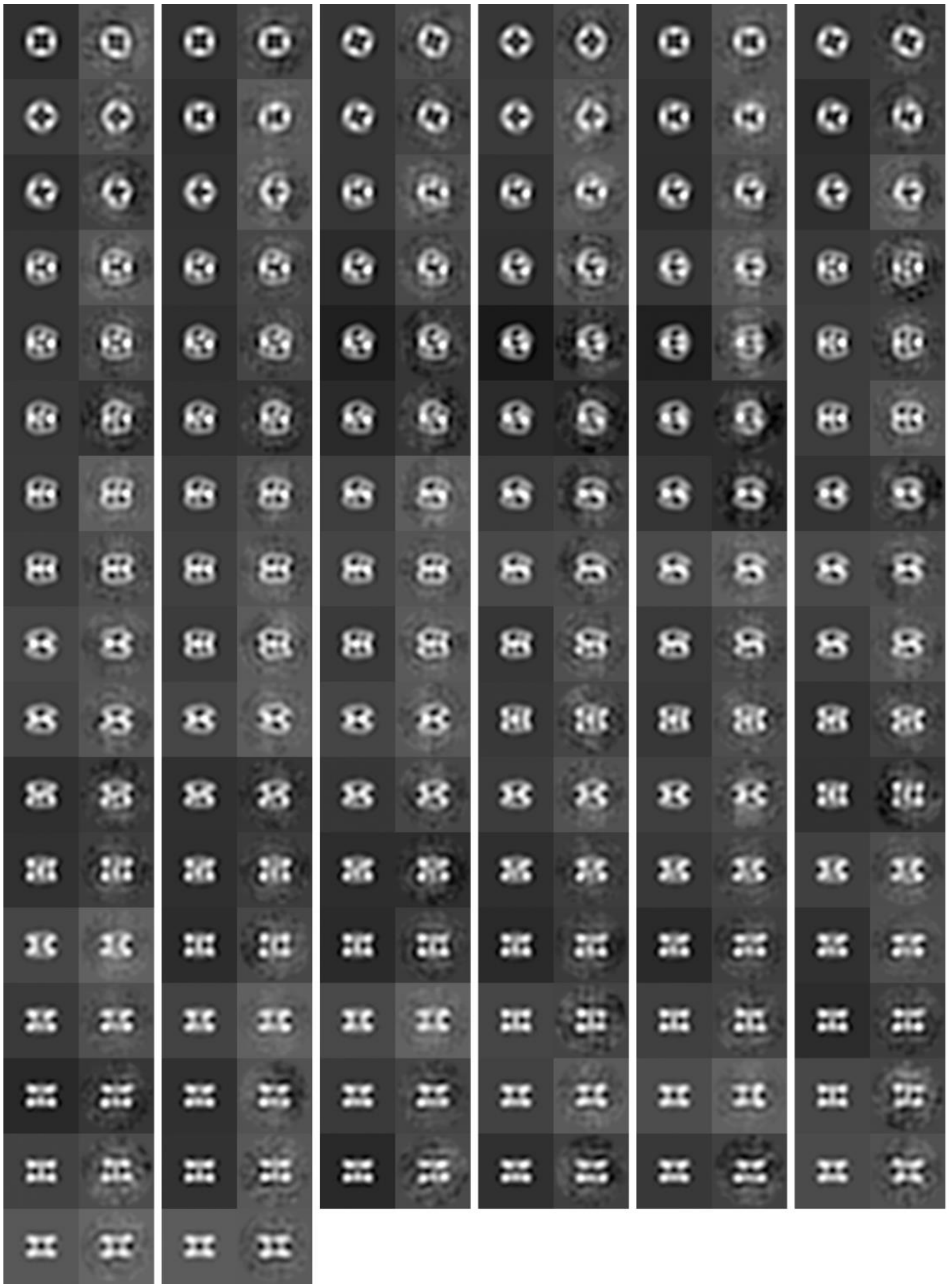

Figure 4.10 A more complete list of the comparison between the 2D projections of the reconstructed 3D model (left) and the corresponding class-averages of the individual particles (right).

4.2.7 Multi-helix bundle structures

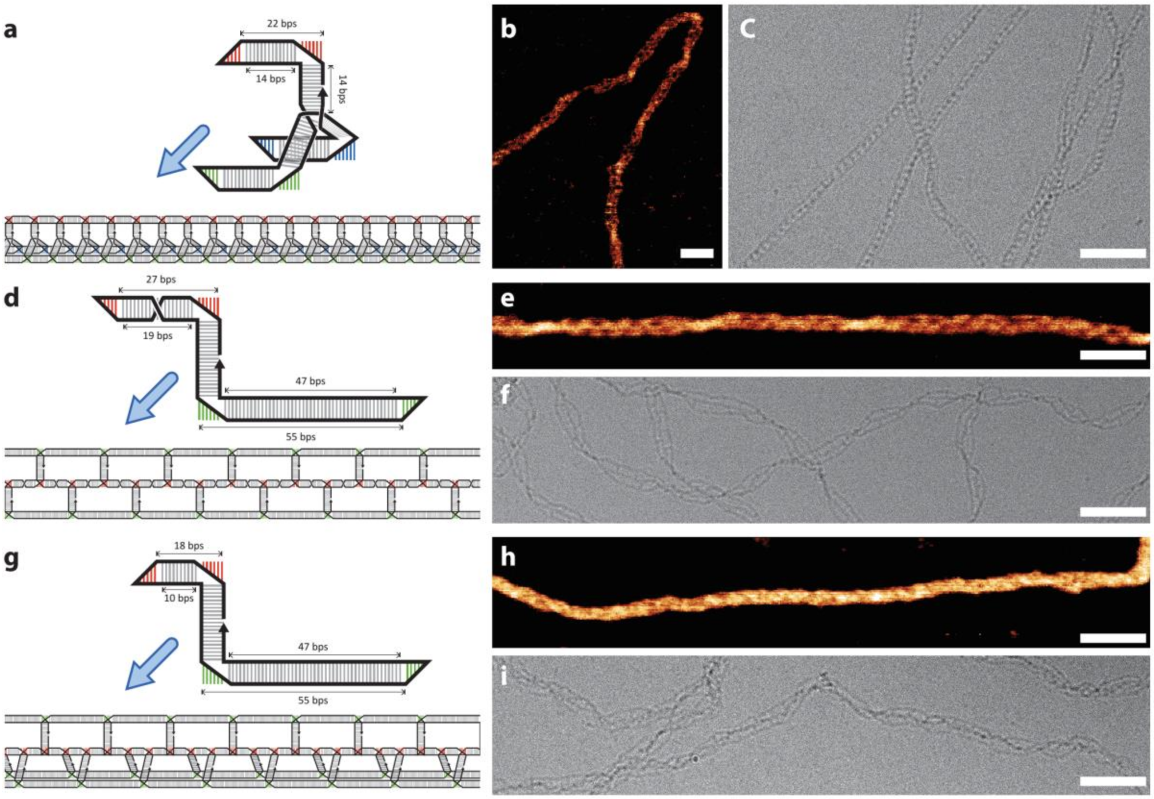

Figure 4.11 Multi-helix bundle structures. (**a**) A claw-like tile containing a 3WJ is designed to self-assemble into a three-helix bundle structure. (**b** and **c**) The three-helix bundle structure imaged by AFM (**b**) and CryoEM (**c**). (**d**) A 19/47-bp-beamed Z-tile is designed to self-assemble into a three-helix bundle structure. (**e** and **f**) The three-helix bundle structure imaged by AFM (**e**) and CryoEM (**f**). (**g**) A 10/47-bp-beamed Z-tile is designed to self-assemble into a four-helix bundle structure. The band corresponding to the target tetrameric nanocage is marked by a black triangle. (**h** and **i**) The four-helix bundle structure imaged by AFM (**h**) and CryoEM (**i**). Scale bar: 50 nM.

Two approaches can be used for the design of the RNA tiles that can self-assemble into structures containing more than two helices in the longitudinal direction. The first approach is to use a claw-like tile that has multiple beams. As shown in Figure 4.11a, a branched tile with 3 beams stemming from a three-way junction (3WJ) has been designed that can be self-assembled into a three-helix bundle structure (Figure 4.11b and c). The other approach is to utilize the tiles of uneven beams, with the longer beam (plus the bKL) spanning full helical turns and being approximately divided evenly by the shorter beam (plus the bKL). The previously shown two-layered structure assembled from the 14/36-bp-beamed Z-tile, though unexpected, is also a case in point. Of course, the shorter beam (plus the bKL) does not necessarily span full helical turns. As for the 19/47-bp-beamed Z-tile (Figure 4.11d), a three-helix bundle structure can be assembled (Figure 4.11e and f). In this structure, the shorter beam (plus the bKL) of each tile generates an approximate 180 ° torsion along central helix and, consequently, each peripheral helix is formed by joining the longer beam of every other tile. Analogously, a 10/47-bp- Z-tile can self-assemble into a four-helix bundle structure (Figure 4.11g-i).

4.2.8 Co-transcriptional assembly and cellular production

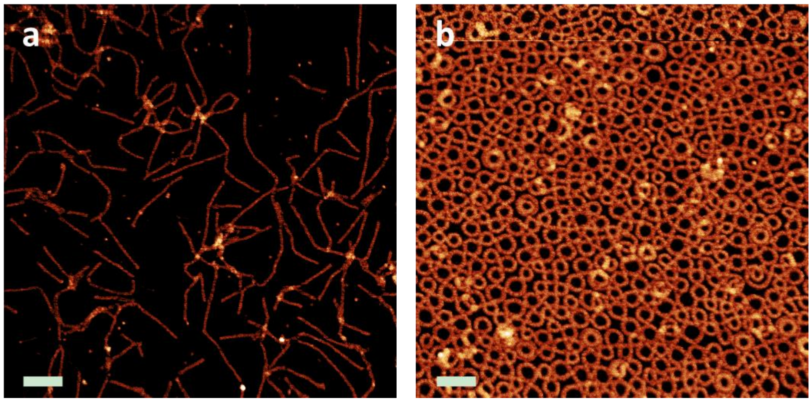

Figure 4.12 Co-transcriptional assembly. (**a**) The helical ladder from the 13-bp-beamed Z-tile and (**b**) the annulus from the 14/25-bp-beamed Z-tile. Scale bar: 100 nm.

In all the previous structures, the RNA molecules were purified from the *in vitro* transcription mixture first and then were annealed for the assembly. We expect that each RNA molecule can fold itself into the corresponding tile during the transcription, and the produced tiles then self-assemble into the target structures. In other words, these structures can also be assembled co-transcriptionally[8]. Two representative structures were tested with the co-transcriptional protocol. The first is the helical ladder structure assembled from the 13-bp-beamed Z-tile; the second is the annulus assembled from the 14/25-bp-beamed Z-tile. Figure 4.12 shows the AFM images of the diluted transcription mixtures of the two structures and both of them are, as expected, assembled co-transcriptionally. The successful co-transcriptional assembly of these structures reasonably indicates the possibility of assembling the structures from the RNA expressed in the cells.

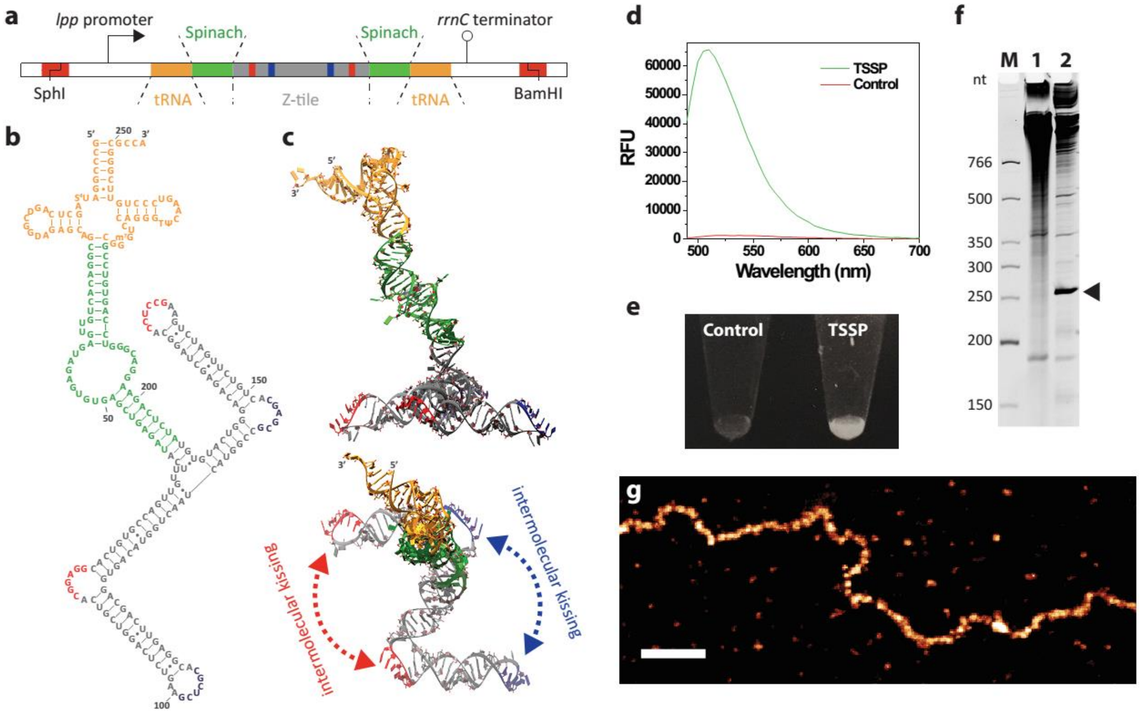

Figure 4.13 Cellular production of RNA nanostructures. (**a**) Design of gene coding the RNA construct, **TSSP**. (**b** and **c**) The secondary structure (**b**) and 3D model (**c**) of **TSSP**. The secondary structure presentation has been prepared with the assistance of VARNA[12]. (**d**) Emission spectra of the bacterial cells with and without plasmid expressing **TSSP** in the presence of 20 µM of 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI). (**e**) A photograph of the same samples in (**d**). (**f**) The total RNA extracted from the bacterial cells without (lane 1) and with (lane 2) plasmid expressing **TSSP** analyzed by dPAGE. The band corresponding to the target RNA construct, **TSSP** (254 nt), is marked by a black triangle. (**g**) AFM image of the assembled structure from cell-produced and gel-purified **TSSP**. Scale bar: 100 nm.

To generate the RNA nanostructure in the bacterial cells, we constructed an RNA encoding gene which can be inserted between the SphI and BamHI restriction sites of the pUC19 plasmid (Figure 4.13a). The design is based on the *in vivo* transcription using the tRNA scaffold[13,14]. Due to the unique fold of tRNA, it can be precisely processed by cellular enzymes and escape cellular RNase, leading to the cellular accumulation. As a result, various RNA molecules have been successfully expressed in *E. coli* by fusing to the tRNA scaffold[13,14]. Between the *lpp* promoter and the *rrnC* terminator is the DNA sequence encoding a chimera RNA construct named **TSSP**. **TSSP** is a fusion of the tRNA scaffold, the Spinach (as the fluorescence tag)[15] and the 13-bp-beamed Z-tile (Figure 4.13b and c). The Spinach, adopting an elongated conformation[16], is inserted between the tRNA scaffold and the Z-tile. One end of the Spinach is fused to anticodon stem of the tRNA scaffold, and the other end of it is connected to the strut of the Z-tile through the well-structured 3WJ from the phi29 prohead RNA[17]. To avoid the possible steric hindrance of the sticking out Spinach with the tRNA scaffold during the assembly, each pair of kissing bulge and loop are designed at the ends of different beams of the Z-tile so that each tile has to rotate 180 ° relative to its neighboring tiles in the assembled structure.

After the transformation of the vector expressing **TSSP**, *E. coli* cells become fluorescent after the addition of 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI) (Figure 4.13d and e), indicating the successful expression of **TSSP** as well as its correct folding. Furthermore, the total cell RNA was analyzed by dPAGE and band corresponding to **TSSP** (254 nt) was observed for the cells transformed with the RNA-expressing vector (Figure 4.13f). The gel-purified target RNA can successfully assemble *in vitro* as is shown by the AFM (Figure 4.13g).

4.2.9 Geometry-based assembly of topological structures

With the capability to control both curvature and torsion of the assembled structures, a handful of complex space curves can be generated. Topological structures, as a special case, should also be constructed. Obviously, in order to tie the ladder-like structure into a knot, a set of several different tiles generating different torsions and curvatures should be used because the curvature and torsion of the trajectory of a knot are not everywhere constant. Nevertheless, we may be able to configure the beams of the assembled structure into complex topology with the one-tile system.

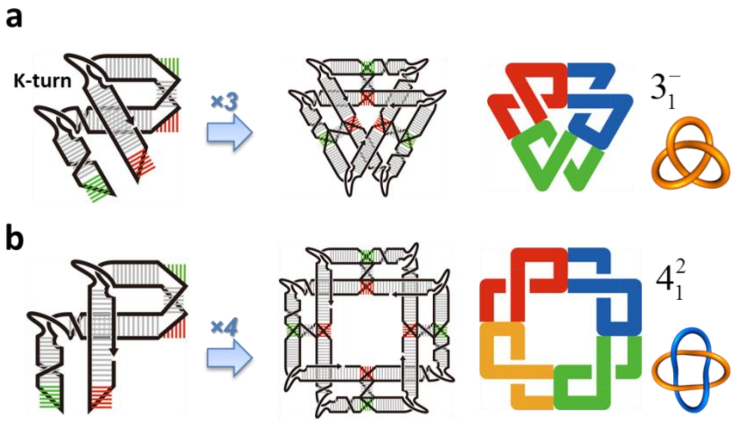

Figure 4.14 Design of the tile that can self-assemble into topological structures. (**a**) The tile is designed to self-assemble into a trimer with the beams wrapping in the topology of a trefoil knot. (**b**) The same tile can also assemble into a tetramer with the beams wrapping in the topology of a Solomon link.

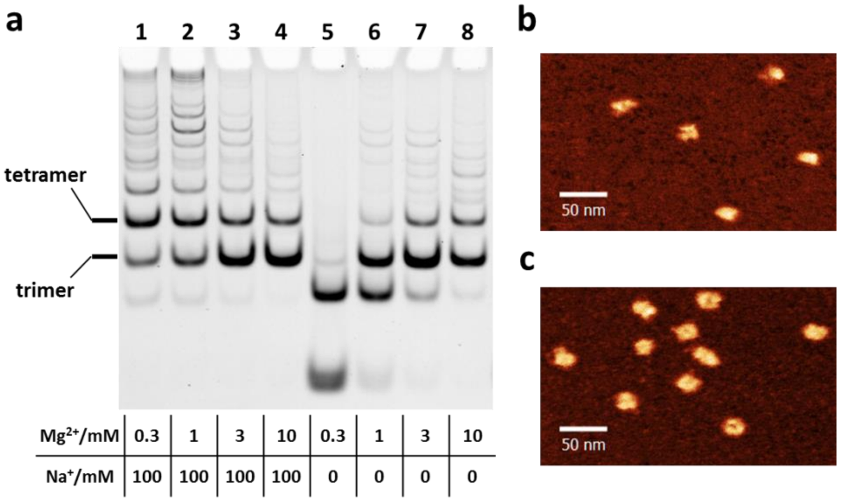

Figure 4.15 Optimizing the annealing condition for the pure assembly. (**a**) Native PAGE analyses of the annealed products of the tile in different buffers. (**b** and **c**) AFM images for the gel-purified trimer (**b**) and tetramer (**c**).

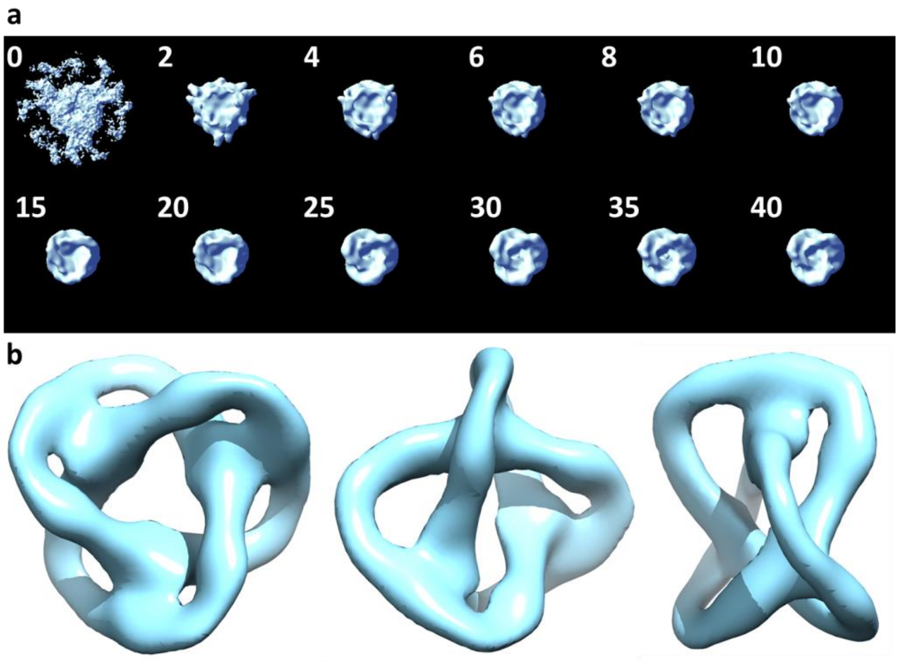

Figure 4.16 CryoEM reconstruction of the trimeric assembly with the topology of a trefoil knot. (**a**) Evolution of the 3D structural model through 40 rounds of refinement iteration starting with a "bad" initial model (#0). (**b**) Three views of the final model (#40 in **a**) after changing the threshold level of the map.

Figure 4.14a shows a designed tile for connecting the beams into a trefoil knot. We chose the K-turn motif[18], which bends about 120 ° (external angle), for the curvature. As for the desired torsion, the length of each fragment of helix is finely designed. Though the annealing buffer was optimized, we failed to obtain satisfactorily neat assembly from the designed tile as shown in the native PAGE (Figure 4.15a). There are two major products formed after annealing in certain buffer conditions (such as in lane 4, the two products accounts for 73% and 21% respectively). They are suggested as the trimeric and tetrameric products based on the mobility in the native PAGE. Though the trimer is the initially designed structure with the beams of the tiles connected into the topology of a trefoil knot, the tetramer can also be formed due to the flexibility of the tile (Figure 4.14b). The beams of the tiles in the tetramer are connected into the topology of a Solomon link.

The bands corresponding to the products of trimer and tetramer were subsequently purified from the gel. AFM images show that the trimeric product has an approximate triangular shape (Figure 4.15b) and tetrameric product has an approximate square shape (Figure 4.15c), implying the formation of the suggested topologies. Furthermore, we attempted to characterize the gel-purified trimer by CryoEM. Even though we failed to generate a reasonable initial model from the reference-free class-averages, we could still obtain a model revealing the topology of a trefoil knot after 40 iterations of refinement using a “bad” initial model (Figure 4.16a). After changing the threshold of the map of the final model, the topology of the connected beams can be clearly seen (Figure 4.16b). Admittedly, this reconstructed model need more validation, but still all the evidence shown above should be sufficient to prove the successful construction of the trefoil knot. The current design of the tile still has pretty large room for optimization and the balance between the structural flexibility and strain should be better considered. After all, the successful design of such delicate structures inevitably necessitates the very fine and precise design of the geometry of the tile. Consequently, we feel it very necessary to better understand the structural detail of the bKL motif.

4.2.10 Crystal structure of an RNA nanobracelet

X-ray crystallography is one of the most widely used and reliable techniques in structural biology for obtaining high-resolution structural information of biological macromolecules. We estimated that it would be very challenging to directly crystallize the bKL due to well-known difficulty associated with RNA crystallization. Its branched structure further increases the difficulty of crystal packing. Therefore, we attempted to crystallize a nanostructure assembled from a tile based on the bKL. The tile was originally designed to assemble into a tetrameric nanocage (Figure 4.17a). However, optimizing the annealing buffer conditions failed to generate the target tetrameric nanocage with good yield (Figure 4.17c). Interestingly, this tile can exclusively assemble into a dimer under a very low Mg^{2+} concentration (0.3 mM). This dimer structure turns out to be a nanobracelet structure as is determined by X-ray crystallography at 2.96 Å (Figure 4.17b, d and e).

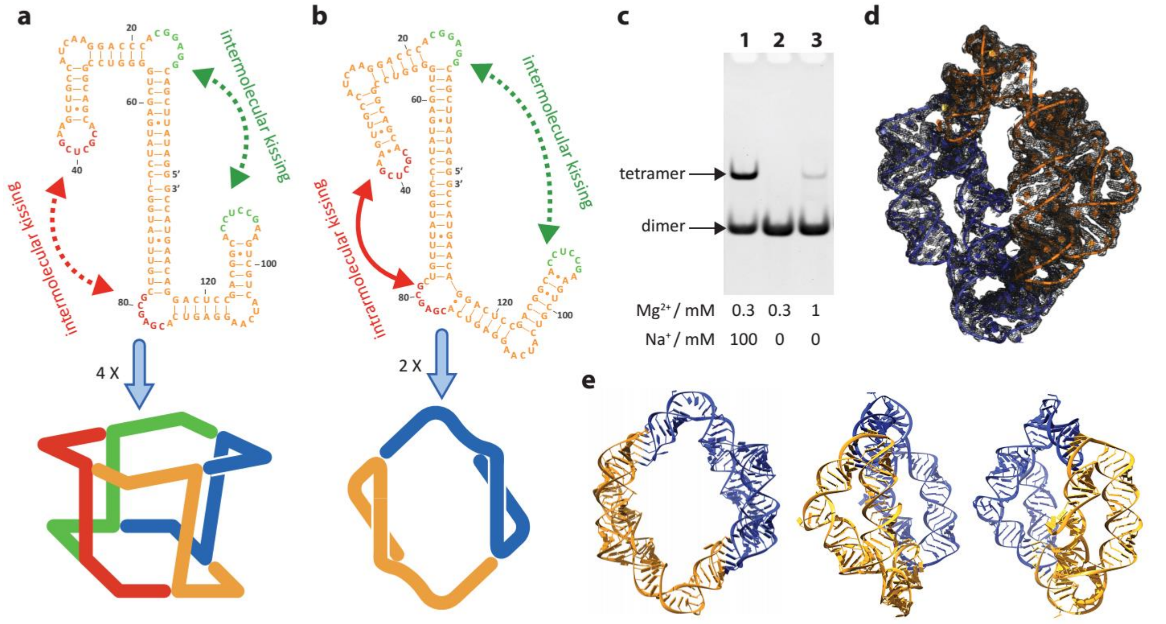

Figure 4.17 RNA nanobracelet. (**a**) Predicted secondary structure of the tile to assemble into tetrameric cage via the formation of two intermolecular bKLs per tile. (**b**) The determined secondary structure of the tile that assembles into the dimeric nanobracelet. Based on the crystal structure, there is an intramolecular bKL formed within each tile and two copies of the tiles associate via the formation of two bKLs. (**c**) Native PAGE analyses of the assembly products of the RNA tile under different annealing buffers. (**d**) Crystal structure of the nanobracelet represented with the electron density map ($2F_o{-}F_c$ map, dark grey mesh) contoured at $\sigma = 1.0$. (**e**) Cartoon representations of the nanobracelet. Three views of a series of 60 °rotation are shown.

There is one intramolecular bKL formed within each tile and two copies of the tiles assemble into the nanobracelet through two intermolecular bKLs. The crystal structure provides high-resolution structural insights of both kinds of bKLs (Figure 4.18). Though all the designed six Watson-Crick base-pairings are formed in both kinds of bKLs, there are substantial structural differences between them, reflecting the intrinsic structural flexibility of the bKL motif. The intermolecular bKL is essentially in accordance with our initial design of the bKL (Figure 4.18a). However, the conformation of the stem-loop part of the bKL is slightly different from the KL in HIV-1 DIS[3], from which the bKL is designed. In the intermolecular bKL, there are two extra A's (A103 and A111) forming a non-canonical base-pair and extending the stem. This A ·A base-pair has the pattern of the bifurcated pair[19]. The last unpaired A (A104) in the loop stacks on A103. Compared with the intermolecular bKL, the intramolecular one deviates more from the initial design (Figure 4.18b). The largest deviation is that the expected G-C base-pair (G124 and C78) in the strut adjacent to the bulge does not form, but G124 pairs with the 3' A (A85) in the bulge. In the stem-loop part of the intramolecular bKL, the two A's (A36 and A44) adjacent to the stem also form base-pair, but with the pattern of *cis* Watson-Crick edge/Watson-Crick edge[19]. The last unpaired A (A37) in this loop inserts itself between C78 and A125 in the strut. Due to this structural deviation, the angle between the two helices adjacent to the bulge is enlarged and therefore accommodates the geometry required by the formation of intramolecular bKL.

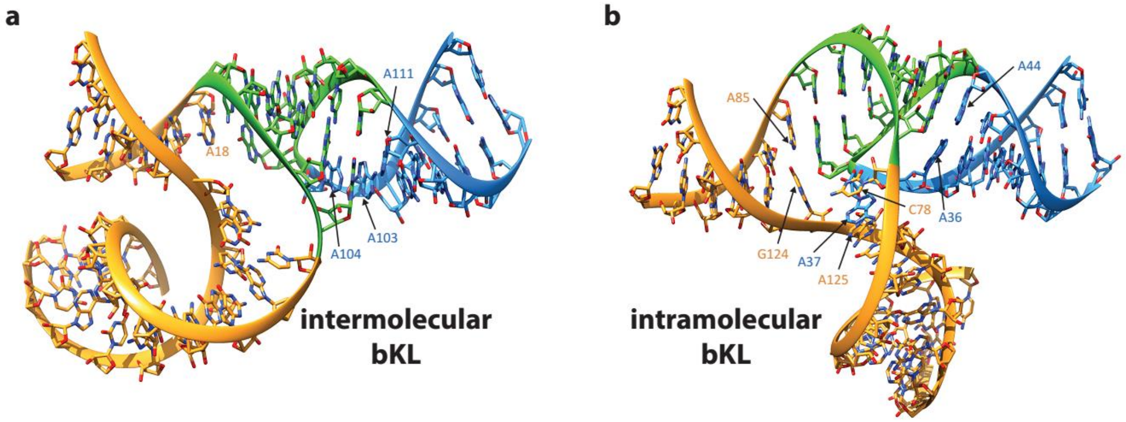

Figure 4.18 Structural details of the intermolecular (a) and intramolecular (b) bKLs.

Each asymmetric unit in the crystal is the dimer, with the two monomeric tiles related with the the non-crystallographic symmetry. The formation of the intramolecular bKLs creates two double-sided “walls” in the nanobracelet and consequently a cleft is embraced in between. The crystal-packing contact is maintained mainly by the shape complementarity. As is shown in Figure 4.19a, the two-sided walls of the nanobracelets can fit into the cleft of the nanobracelets in the adjacent layers along the screw axis (y axis). The shape complementarity also prevents the formation of channels through the hole of the nanobracelets (Figure 4.19b). Additionally, there is no flipped base providing interparticle base-paring, nor is there stacking interaction. This is very rarely observed in nucleic acid crystals.

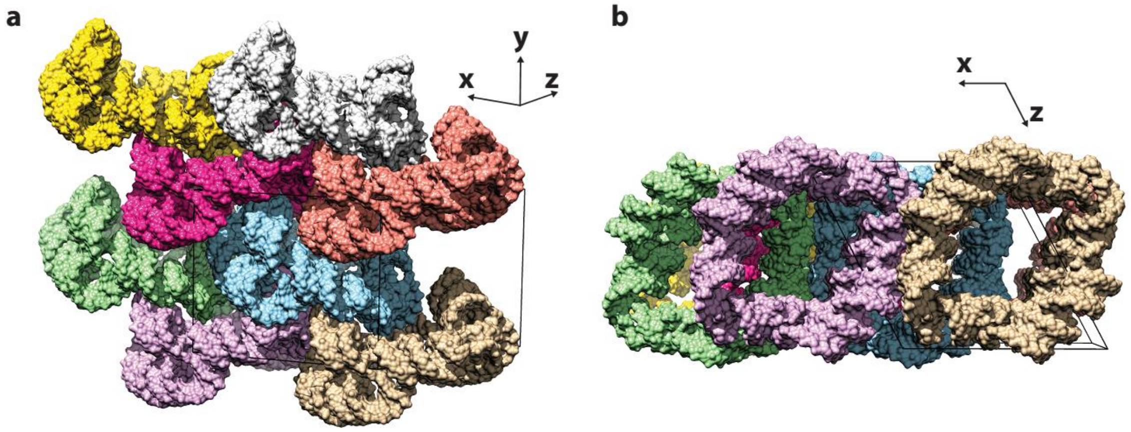

Figure 4.19 Crystal packing for the RNA nanobracelet.

4.2.11 Crystallizing artificially designed RNA nanostructures

Inspired by the success of obtaining the crystal structure of the above RNA nanobracelet, we further extend our efforts in the crystallization of several other artificially designed RNA nanostructures. The design strategy is RNA architectonics, which involves fusing RNA structural motifs to generate closed geometry[20,21]. Experimentally, two steps are critical. The first step is optimizing the folding and assembling of the designed structures. In most cases, the high homogeneity (95%) of resulting assembled structure is desired. Though we also purified several structures by native PAGE after the annealing, this purification step inevitably complicates the experimental procedure. The second step is the crystallization of the well-folded/assembled or native PAGE purified structures. In our experiment, we obtained crystals from all of the constructs. However, only five crystals (including the RNA nanobracelet) have diffraction better than 4 Å.

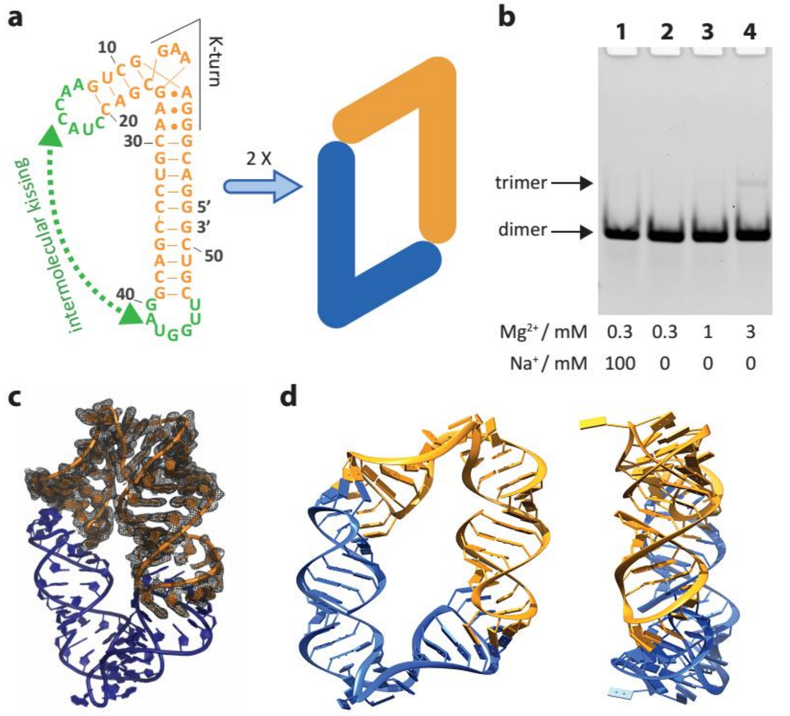

Figure 4.20 RNA dimeric parallelogram. (**a**) Secondary structure of the tile that dimerizes into the parallelogram. (**b**) Native PAGE analyses of the assembly products of the RNA tile under different annealing buffers. (**c**) Crystal structure of the parallelogram with one monomer represented with the electron density map ($2F_o{-}F_c$ map, dark grey mesh) contoured at $\sigma = 1.0$. (**d**) Cartoon representations of the nanobracelet. Two views related by a 90 ° rotation are shown.

The best diffracting crystal we have obtained thus far is that of a 2D homodimeric parallelogram structure. We designed a single strand of RNA that can fold itself into a tile with the secondary structure like a bent dumbbell (Figure 4.20a). The bending is rendered by the K-turn motif with a 120 ° bend angle (60 ° internal angle)[18]. Two copies of the tiles self-assemble into the parallelogram through the formation of two intermolecular KLs. The KLs are adopted from the 7-bp kissing complex formed by the palindrome I of RNA I and the palindrome I RNA II transcribed by *ColE1* plasmid of *E. coli*[22], and have been determined by NMR to have an approximate 80 ° bend angle (100 ° internal angle)[23]. As shown by the native PAGE (Figure 4.20b), the dimer is the exclusive assembled structure after annealing in a range of different buffers (though trimer emerges as the Mg^{2+} concentration increases).

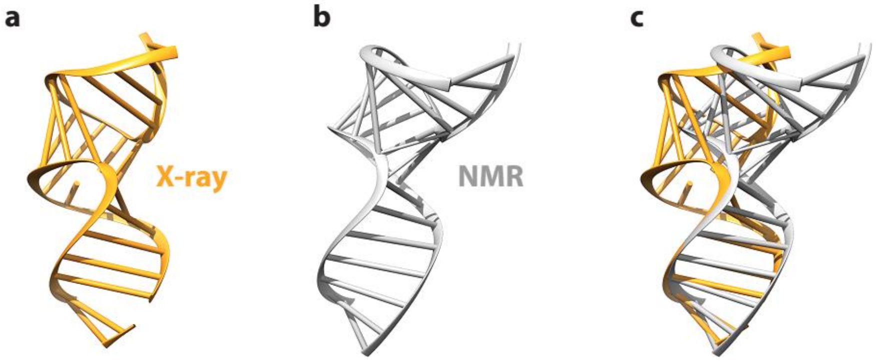

Figure 4.21 Comparing the 7-bp KL structure determined by X-ray crystallography and by NMR. (**a**) The X-ray structure of the 7-bp KL determined in the present study. (**b**) The NMR structure of the 7-bp KL determined previously (PDB ID: 2BJ2)[23]. (**c**) Superimposition of the X-ray and NMR structures.

We obtain the crystal of the dimeric parallelogram that diffracts to 2.09 Å. The overall structural feature is in good accordance with the initial design and the closed structure is formed due to the bending collectively contributed by the two K-turns and the two 7-bp KLs in the dimer (Figure 4.20c and d). Furthermore, obtaining the high-resolution crystal structure of the dimeric parallelogram allows us to compare the 7-bp KL structure determined by X-ray crystallography (Figure 4.21a) with that determined previously by NMR[23] (Figure 4.21b). It turns out that the X-ray structure is slightly different from the NMR structure. Whereas this KL structure was shown to adopt a wide range of bend angle with an average internal angle of approximately 100 ° as determined by NMR, the X-ray structure exhibit an internal angle of approximately 120 °(Figure

4.21c). This smaller inter-helix bending of the KL in our X-ray structure is likely a result of the geometrical constraints of being incorporated within the closed nanostructure. Still, the X-ray structure presented here is in the possible conformation that the KL can adopt and can provide valuable structural insights. We note that the crystal structure of this KL would be very difficult to obtain without configuring it into a well-formed nanostructure.

Based on the observations in our experiment, we tentatively propose the following two rules that may help in designing the RNA nanostructures that can diffract to high resolution. Firstly, smaller structures tend to produce better crystals. This is likely due to the smaller cavities and higher rigidity of the smaller structures. Generating nanocages with larger cavities may not be necessarily impossible as long as to increase the framework and linkage points more rigid. Secondly, pseudosymmetry of the nanostructure is a not good thing. The very tiny structural differences of different orientations of the pseudosymmetrical nanostructure are very hard to be differentiated during the crystal packing and this can limit the resolution that the crystal can achieve. Even though we have obtained a quite high-resolution crystal of a pseudosymmetrical nanostructure, this structure is very difficult to refine due to the suspected static disorder stemming from pseudosymmetry.

4.2.12 Conclusions

RNA nanotechnology is an excitingly emerging field[24]. Compared to its DNA cousin, RNA nanotechnology has many additional advantages associated with the properties pertaining to RNA, which include the more diverse functionalities, a cornucopia of structural motifs and the ease of production by *in vitro* or *in vivo* transcription and etc. Despite the plethora of structural motifs in naturally occurring RNA folds, most of them suffer from the lack of modularity – their formation may be context-dependent or they may be difficult for generalized reconfiguration, which limits their usage in the design of artificial RNA nanostructures. Our artificially designed bKL is a very desirable motif as is shown in this chapter. Opportunities of new nanodesign strategies are always emerging with the invention of new structural motifs. We reasonably anticipate that this bKL motif can be incorporated in a vast number of new structures in the near future. Moreover, there exist very similar loop-bulge kissing interactions as essential structural elements in various naturally occurring RNA structures, such as the group II intron ribozyme[25] (an intramolecular 7-bp bKL is formed between the loop of stem IB and the bulge of stem ID2, i.e. the α-α' interaction), the RNase P RNA[26,27] (the P6 in the archaeal and A-type bacterial RNase P RNA is formed by the intramolecular bKL interaction), and the phi29 prohead RNA[28] (an intermolecular 4-bp bKL is formed between the loop L_D and the bulge L_{CE}). Our artificially designed bKL may be useful in re-engineering such functional RNAs.

It is also important to realize that this bKL motif has a pretty large room to be reengineered. For example, the unpaired A's in the bulge and the loop may be altered. The number of base-pairings between the bulge and the loop is another variant that can be engineered. The flexibility and the bending propensity of the bKL are therefore might be fine-tuned, making it better suitable in the geometry control. Indeed, we have incorporated some modified versions of the bKLs in the some nanocages for crystallization (data not shown). We expect to see these structures and consequently learn the structural details of the bKLs within very soon.

Obtaining well-ordered crystals remains one of the most challenging bottlenecks in X-ray crystallography. This issue is more severe for the crystallization of DNA and RNA, and this is evident by looking at the very small amount of nucleic acid crystal structures in the PDB. The difficulties associated with nucleic acid structure crystallization include the elongated shapes (in contrast to the globular shape of most proteins) and the indistinguishable negatively charged phosphate backbones. As a result, the crystallization of even a very small fragment of nucleic acid (< 40 nt) can be fairly challenging. For example, only recently was reported the first solved crystal structure of a DNAzyme[29]. Stacking normally plays an important role in the intermolecular contact or packing within the DNA or RNA crystals. Therefore, end- or loop-engineering is routinely employed and extensively screening of different constructs is always unavoidable. However, it heavily depends on luck to have the exact and proper stacking that leads to the formation of ordered crystals. Here, we want to raise the argument that nucleic acid structures with closed and more regular geometry can be better crystallizable. This is because the closed geometry can create geometrically differentiated surface and can also restrict the swing of the otherwise flexible RNA molecules. Furthermore, the self-assembled structures are rotationally symmetrical. The symmetry is another beneficial factor that can promote the crystallization (due to the fewer distinct crystal contacts required to form a connected network in 3D space) as was shown by Yeates and coworkers[30,31].

Two strategies may be potentially developed based on this work for nucleic acid crystallography. (1) Our crystallized nanostructures contain interior cavities and can serve as nanoscaffolds to include other RNA or DNA inside. Because the crystal packing is dominated solely by the contact of the nanoscaffolds, this strategy can greatly streamline the conventional crystal screening process. Another benefit of this strategy is that the nanoscaffolds can also provide phase information for solving the structure. (2) A lot of nucleic acid structures (such as aptamers or DNAzyme) or structural motifs have the extended and elongated structures, they can be conveniently engineered to form closed structures with regular and/or symmetric geometry, which are more likely to form well-ordered crystals. Even though some structures do not have large bending, they can be fused to well-studied bended motifs (such as bulges or K-turns) to form the closed geometry.

4.3 Methods

4.3.1 RNA preparation

The design of the sequences followed the rule of sequence symmetry minimization[32] and was assisted by the program CANADA[33]. The secondary structure of each RNA molecule was further predicted by Mfold[34] to ensure the correct folding of the RNA into the designed tile. All RNA molecules were synthesized by *in vitro* transcription using the HiScribe™ T7 High Yield RNA Synthesis Kit from the New England Biolabs (NEB). The corresponding DNA templates were generated by the PCR amplification of the gBlocks® gene fragments from the Integrated DNA Technologies (IDT) using the Q5® Hot Start High-Fidelity DNA Polymerase (NEB). To reduce transcriptional heterogeneity at the 3' end, the first two nucleotides of the reverse primer are modified with 2'-OMe[35]. The RNA molecules were then purified by the dPAGE, ethanol precipitated and suspended in pure water.

4.3.2 Assembly of the nanostructure

The RNAs were denatured at 90 ℃ for 1 min and snap-cooled on ice before assembly. For most structures, RNAs were diluted to 600 nM in 1×TAE-Mg buffer (11 mM $MgCl_2$, 40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.0) and annealed from 70 ℃ to 4 ℃ with the following assembly protocol: 70 ℃ to 50 ℃ over 6 min, 50 ℃ to 37 ℃ over 20 min, 37 ℃ to 4 ℃ over 2 hr. For the assembly of several nanocage structures, the buffer composition was optimized using 10×TAE-Mg buffer and 1 M NaCl to obtain the desired concentrations of Mg^{2+} and Na^+.

4.3.3 AFM imaging

The annealing mixture was diluted 10 times with 1×TAE-Mg buffer. 5 μL of the diluted sample was deposited onto a freshly cleaved mica surface (Ted Pella) and left to adsorb for 1 min and dried with compressed air. The mica was then rinsed with 20 μL of 2 mM $Mg(OAc)_2$ and dried with compressed air. AFM imaging was performed on a Veeco 8 AFM in the ScanAsyst® in Air mode using the scanayst-air tips (Veeco). The images were processed with Gwyddion.

4.3.4 CryoEM imaging and single-particle reconstruction

For the long structures, the annealing mixture containing 600 nM of RNA tiles was directly used for sample preparation. For closed structures, the annealing mixture was concentrated 10 times with an Amicon Ultra centrifugal filter (MWCO 30 kDa) before the sample preparation. 3 μL of solution was applied onto a glow-discharged C-flat holey carbon grid (CF-1.2/1.3-4C), blotted for 5.5 s and immediately flash frozen by the liquid nitrogen-cooled liquid ethane with Cryoplunge 3 System (GATAN). Images were collected on a JEOL 3200FS TEM operated at 300 kV with a K2 Summit direct electron counting camera (GATAN) under low-dose mode. For the structures without single-particle reconstruction, images were recorded at 12000× or 25000× microscope magnification with the defocus ranging from about -3.0 μm to -5.0 μm. For the structures for single-particle reconstruction, images were recorded at 25000× microscope magnification with the defocus ranging from about -1.0 μm to -4.0 μm.

Single-particle reconstruction was performed using EMAN2[36]. 598 particles were used to generate the reference-free class averages, which were used to build the initial model. The refinement of the model was conducted with the full set of 1440 particles. Resolution of the resulted structural density map were estimated to be 18.1 Å using gold-standard FSC = 0.143

criterion without applying any mask. The reconstructed model was visualized by UCSF Chimera[37].

4.3.5 Cellular production of TSSP

DH5α single colony was picked from the agar plate and inoculated into fresh LB medium. After overnight growth, 500 μL of cells were inoculated into 50 mL of fresh LB medium and grow at 37 ℃ with shaking at 220 rpm until OD_{600} reaches 0.4. Cool down the cells on ice for 5min. The cells were spun down at 4000 ×g for 15min and then re-suspended with 5 mL of pre-chilled 0.1 M $CaCl_2$ solution. The cells were spun at 4000 ×g for 15min again and finally re-suspended with 1 mL of pre-chilled 0.1 M $CaCl_2$ solution.

The PCR product containing the gene for expressing **TSSP** was digested with BamH I and Sph I and ligated with pUC19 vector digested with same restriction enzyme using T4 DNA ligase (NEB). 1 μL of ligation mixture was added into 150 μL of DH5α competent cells and the mixture was left on ice for 30 min. The mixture was quickly incubated at 42 ℃ for exact 60 s and was then quickly put back on ice for 2 min. 850 μL of SOC medium was added and the mixture was incubated at 37 ℃ for 45 min with shaking at 220 rpm. 200 μL of the culture was spread onto an agar plate with ampicillin antibiotic. The plate was incubated at 37 ℃ overnight. Single colony was picked and grown in 5 mL of fresh LB medium at 37 ℃ overnight. The plasmid was extracted by plasmid miniprep kit (Qiagen) and the sequence of plasmid was confirmed by sequencing.

The DNA plasmid with correct sequence was transformed into DH5α again. Single colony was picked and grown overnight in fresh LB medium. The total RNA was prepared by PCI extraction and ethanol precipitation. The target RNA **TSSP** was purified by dPAGE.

4.3.6 Crystallization and structure determination

The folded/assembled nanostructures were concentrated to ~5 μg/μL using Amicon Ultra centrifugal filters (MWCO 10 kDa). The initial crystallization trails were carried out using the high-throughput hanging-drop vapor-diffusion method in 96-well plates set up by a Mosquito liquid handling robot (TTP Labtech). Each drop contains 0.1 μL of RNA and 0.1 μL of reservoir solution. Four commercially premade screening kits (Natrix (Hampton), Nucleix (Qiagen), Nuc-Pro (Jena Biosciences), Crystallization Kit for RNA (Sigma)) were used. Initial hits were identified and the drops were enlarged to contain 1.2 μL of RNA and 1.2 μL of reservoir solution on siliconized glass slides. The final crystallization condition for the nanobracelet contains 80 mM NaCl, 12 mM KCl, 20 mM $MgCl_2$, 40 mM Sodium cacodylate (pH = 6.0), 30 % (v/v) MPD, 12 mM spermine tetrahydrochloride. The final crystallization condition for the parallelogram contains 50 mM HEPES (pH = 7.0), 20 mM KCl, 5 mM $MnCl_2$, 35 % (v/v) MPD.

Data sets were collected at NE-CAT beamline 24-ID at the Advanced Photon Source (APS). For solving the structure of nanobracelet, the bulge AACUA motif (PDB ID: 2NOK)[9] flanked by two standard 6-bp A-form helices was built. Molecular replacement was performed by using two copies of bulge structure and two copies of standard 10-bp A-form helices as the search models. An initial solution with LLG of 458 was obtained. Relying on the initial map from this solution, more nucleotides were built into the map. The complete model was obtained by iteratively building more nucleotides and improving the map. For solving the structure of parallelogram, molecular replacement was performed using the structure of K-turn (PDB ID: 4CS1)[18] with a flanking 9-bp A-form helix and a 6-bp A-form helixes as the search models. An initial solution with LLG of 258 and an initial map were obtained. The kissing-loop was built by placing six base pair A-form helix into its supposed position in the electron density map followed by rigid body refinement. The model was completed by ligating all fragments. All models were built in COOT[38]. All molecular replacement and refinement were performed using PHENIX[39]. All of the figures were prepared with Pymol (DeLanoScientific LLC) or UCSF Chimera[37]. The statistics of data collection and refinement were tabulated in Table 4.1.

Table 4.1 Statistics for data collection and refinement

Crystal	nanobracelet	parallelogram
Data collection		
Space group	$P2_1$	$P4_32_12$
Resolution (Å)	70.00-2.96 (3.01-2.96)	40.22-2.09 (2.15-2.09)
Cell dimensions		
a, b, c (Å)	86.44, 71.99, 89.44	40.22, 40.22, 202.43
α, β, γ (°)	90, 117.61, 90	90, 90, 90
R_{merge} (%)	10.7 (105.9)	4.5 (110.6)
$I/\sigma I$	36 (2.15)	23.3 (0.9)
Completeness (%)	99.2 (98.2)	96.8 (77.7)
Redundancy	15.9 (12.9)	7.1 (2.4)
Refinement		
No. of reflections	20,356	9,931
R_{work}/R_{free}	0.1935/0.2327	0.2303/0.2604
r.m.s deviations		
Bond angles (°)	1.254	1.227
Bond length (Å)	0.006	0.005
B-factors	89.22	58.38

* Values in the parentheses are for the highest resolution shell.

4.4 References

1. Dietz, H., Douglas, S. M. & Shih, W. M. Folding DNA into twisted and curved nanoscale shapes. *Science* 2009, **325**(5941)**:** 725-730.

2. Hamada, S. & Murata, S. Substrate-assisted assembly of interconnected single-duplex DNA nanostructures. *Angew. Chem. Int. Ed.* 2009, **48**(37)**:** 6820-6823.

3. 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. *Nature Struct. Biol.* 2001, **8**(12)**:** 1064-1068.

4. Li, X. *et al.* DNA polyhedra with T-linkage. *ACS nano* 2012, **6**(6)**:** 5138-5142.

5. Tian, C. *et al.* Approaching the limit: can one DNA strand assemble into defined nanostructures? *Langmuir* 2014, **30**(20)**:** 5859-5862.

6. Severcan, I., Geary, C., Verzemnieks, E., Chworos, A. & Jaeger, L. Square-shaped RNA particles from different RNA folds. *Nano Lett* 2009, **9**(3)**:** 1270-1277.

7. Severcan, I. *et al.* A polyhedron made of tRNAs. *Nat Chem* 2010, **2**(9)**:** 772-779.

8. Geary, C., Rothemund, P. W. & Andersen, E. S. RNA nanostructures. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science* 2014, **345**(6198)**:** 799-804.

9. Dibrov, S. M., Johnston-Cox, H., Weng, Y. H. & Hermann, T. Functional architecture of HCV IRES domain II stabilized by divalent metal ions in the crystal and in solution. *Angew. Chem. Int. Ed.* 2007, **46**(1-2)**:** 226-229.

10. Dibrov, S. M., McLean, J., Parsons, J. & Hermann, T. Self-assembling RNA square. *Proc. Natl. Acad. Sci. USA* 2011, **108**(16)**:** 6405-6408.

11. Yu, J., Liu, Z., Jiang, W., Wang, G. & Mao, C. De novo design of an RNA tile that self-assembles into a homo-octameric nanoprism. *Nat Commun* 2015, **6:** 5724.

12. Darty, K., Denise, A. & Ponty, Y. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 2009, **25**(15)**:** 1974-1975.

13. Ponchon, L. & Dardel, F. Recombinant RNA technology: the tRNA scaffold. *Nat Methods* 2007, **4**(7)**:** 571-576.

14. Ponchon, L., Beauvais, G., Nonin-Lecomte, S. & Dardel, F. A generic protocol for the expression and purification of recombinant RNA in Escherichia coli using a tRNA scaffold. *Nat Protoc* 2009, **4**(6)**:** 947-959.

15. Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA mimics of green fluorescent protein. *Science* 2011, **333**(6042)**:** 642-646.

16. Huang, H. *et al.* A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nat Chem Biol* 2014, **10**(8)**:** 686-691.

17. Zhang, H. *et al.* Crystal structure of 3WJ core revealing divalent ion-promoted thermostability and assembly of the Phi29 hexameric motor pRNA. *RNA* 2013, **19**(9)**:** 1226-1237.

18. McPhee, S. A., Huang, L. & Lilley, D. M. A critical base pair in k-turns that confers folding characteristics and correlates with biological function. *Nature Commun.* 2014, **5:** 5127.

19. Leontis, N. B., Stombaugh, J. & Westhof, E. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* 2002, **30**(16)**:** 3497-3531.

20. Jaeger, L. & Chworos, A. The architectonics of programmable RNA and DNA nanostructures. *Curr Opin Struct Biol* 2006, **16**(4)**:** 531-543.

21. Grabow, W. W. & Jaeger, L. RNA self-assembly and RNA nanotechnology. *Acc Chem Res* 2014, **47**(6)**:** 1871-1880.

22. Tomizawa, J.-i. Control of cole 1 plasmid replication: The process of binding of RNA I to the primer transcript. *Cell* 1984, **38**(3)**:** 861-870.

23. Lee, A. J. & Crothers, D. M. The solution structure of an RNA loop–loop complex: the ColE1 inverted loop sequence. *Structure* 1998, **6**(8)**:** 993-1007.

24. Guo, P. The emerging field of RNA nanotechnology. *Nature Nanotechnol.* 2010, **5**(12)**:** 833-842.

25. Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* 2008, **320**(5872)**:** 77-82.

26. Evans, D., Marquez, S. M. & Pace, N. R. RNase P: interface of the RNA and protein worlds. *Trends Biochem Sci* 2006, **31**(6)**:** 333-341.

27. Torres-Larios, A., Swinger, K. K., Krasilnikov, A. S., Pan, T. & Mondragon, A. Crystal structure of the RNA component of bacterial ribonuclease P. *Nature* 2005, **437**(7058)**:** 584-587.

28. Ding, F. *et al.* Structure and assembly of the essential RNA ring component of a viral DNA packaging motor. *Proc. Natl. Acad. Sci. USA* 2011, **108**(18)**:** 7357-7362.

29. Ponce-Salvatierra, A., Wawrzyniak-Turek, K., Steuerwald, U., Hobartner, C. & Pena, V. Crystal structure of a DNA catalyst. *Nature* 2016, **529**(7585)**:** 231-234.

30. Wukovitz, S. W. & Yeates, T. O. Why protein crystals favour some space-groups over others. *Nature Structural Biology* 1995, **2**(12)**:** 1062-1067.

31. Banatao, D. R. *et al.* An approach to crystallizing proteins by synthetic symmetrization. *Proc. Natl. Acad. Sci. USA* 2006, **103**(44)**:** 16230-16235.

32. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* 1982, **99**(2)**:** 237-247.

33. Feldkamp, U. CANADA: Designing Nucleic Acid Sequences for Nanobiotechnology Applications. *J. Comput. Chem.* 2010, **31**(3)**:** 660-663.

34. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 2003, **31**(13)**:** 3406-3415.

35. Kao, C., Zheng, M. & RÜDisser, S. A simple and efficient method to reduce nontemplated nucleotide addition at the 3′ terminus of RNAs transcribed by T7 RNA polymerase. *RNA* 1999, **5**(9)**:** 1268-1272.

36. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* 2007, **157**(1)**:** 38-46.

37. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 2004, **25**(13)**:** 1605-1612.

38. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004, **60**(Pt 12 Pt 1)**:** 2126-2132.

39. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 2010, **66**(Pt 2)**:** 213-221.

Chapter 5 – Looking back and looking/moving forward

In summary, we have developed various ways to create topological structures with nucleic acids. Since DNA topology is a real puzzle for the living cells and, highly likely, so is RNA topology, the constructed structures should be expected to hold promising applications towards the practical ends. Nonetheless, designing and constructing these exotic structures by itself is an exciting and rewarding adventure.

5.1 Respecting topology as a scientific philosophy

Topology is indeed one of the very few topics that have been constantly fascinating researchers from almost all scientific disciplines. Looking into the history of science, we can find that a number of significant scientific breakthroughs were inspired by thinking topologically. As for DNA (or biology more generally), two of the most prominent discoveries are DNA supercoiling[1] and DNA topoisomerase[2], both belonging to the study of DNA topology. "Unfortunately, these discoveries have not received the credit they merit outside of the community of scientists interested in the DNA structure and DNA biological roles," wrote Patrick Forterre in the book *DNA Topoisomerase and Cancer*[3], "The fact that James Wang has not (yet) been rewarded by the Nobel Prize is astonishing, considering the importance of DNA topoisomerases in both fundamental chemistry and medicine (as testified by this book). More generally, the study of DNA topology and DNA topoisomerases is not given appropriate credit in life science and these topics are still missing from too many biological degree courses at universities. To play down the importance of DNA topology is highly damaging for someone whose aim is to understand how modern living organisms thrive on our planet."

When Watson and Crick proposed the double helical model of DNA in 1953[4], it was not very readily accepted by all biologists. This was mainly due to the intrinsic topological problem stemming from the helical nature of their model, which necessitates the two strands being wrapped around each other. Realizing this topological puzzle stimulated scientists to pursue the answers and the abovementioned two discoveries were made as a result. During the pursuit to the correct answers, some scientists proposed the non-helical model, or side-by-side model, of DNA[5], trying to provide a theoretical solution to the topological puzzle of DNA. Though this model has not been supported by the later experimental observations, their efforts should be remembered and respected. After all, offering an imperfect answer with all-around thinking that was allowed by the time is better than a total ignorance.

Unfortunately, researchers (out of mathematics) nowadays seem to be more lacking in topological thinking than their counterparts in the old times. Two sad facts reinforce my statement. The first is that unprofessional, incorrect presentations of DNA are too frequent. DNA models shown to be "left-handed", or nucleosome models with DNA wrapping around in a "right-handed" manner, appear too often in papers (or cover images) published (even in high-profile journals), in popular science readings, in presentation slides and posters, in the websites of research groups, and in the commercial advertisements. The second sad fact reflecting the ignorance of topology is that the word "topology" is often misused by chemists and biologists. In most cases when the term "topology" is used in chemistry or biology papers, it should be better to be replaced by "geometry", "shape", or "pattern". The reason why "topology" is more favored than the other may be that "chemists are not immune to the temptation to sound, if not wise, at least mathematically sophisticated," wrote by Michelle Francl in an essay pointing out that the meaning of topology has been "stretched" by chemists[6].

Nonetheless, there are still a line of researchers outside of mathematics nowadays who esteem topology with more serious respect. For some of them (including myself), the pursuit of topology from their respective disciplines is purely out of rather eccentric personal interest. In fact, topology plays a prominent role in the development of several research areas. Structural DNA nanotechnology is a case in point[7]. In the very beginning of the field, most of the earliest DNA nanostructures[8,9] were actually topological targets and the design principles were gained by building and characterizing them on the topological level. Another example is the field of supramolecular chemistry[10]. Synthesis of molecules with nontrivial topologies has been one of the most important impetuses for the exploration of various strategies in molecular recognition. The formation and realization of the concept of molecular machines or molecular motors would not be possible without synthetic molecular topologies.

5.2 Several related points worth exploring in the future

This is the last section of my thesis. Here I would like to mention several points that can be explored. Because the scope of the field of DNA structural nanotechnology is very vast and, at the same time, is replete with collaborative opportunities with other disciplines of research, I limit these points to be only related to my graduate research presented or not presented in the previous chapters. Meanwhile, I refrain from reiterating the points that have been already mentioned in the previous chapters.

5.2.1 Further extending the complexity of molecular topological structures with nucleic acids

Though we have greatly enriched the accessible topologies with ssDNA or ssRNA, it is important to realize the intrinsic limitations with single-stranded structures –they are unbranched.

Within the more generic supramolecular chemistry, some unusual topological structures besides knots and links have been synthesized, such as various interlocked nanocages[11] and suit[3]ane[12]. Branching can be achieved with double-stranded structures, which can be used to explore these unusual topological entities. In fact, it is at least theoretically possible that these double-stranded topological structures can be created with a multistep fashion, just like Seeman's early approach to synthesize the dsDNA nanocages[8,9]. However, there would be more pain than gain in doing so, especially when considering the questions of practical and theoretical values. Moreover, it is important to realize that in all the previous construction of double-stranded topological structures, the nodes were actually produced by the single-stranded nucleic acids, either with the helix-based method[13,14] or our newly invented junction-based method[15], and that at least a certain double-stranded region in each structure was finally formed by hybridizing the complementary strand(s) to the corresponding single-stranded region. This method, though doable, would unavoidably complicate the synthesis and limit the complexity that can ultimately be achieved. Therefore, it is desirable to create nodes directly with double-stranded nucleic acids. One possibility is proposed recently by Seeman that dsDNA nodes can be directly produced by the paranemic crossover DNA motif, where two dsDNA helices appear to wrap around each other[16]. Another possibility is that the helix-helix interaction (mediated either by unpaired bases or by nucleic-acid binding proteins) can be explored to form nodes of double-stranded structures. Though the practical applications of these double-stranded topological structures are vague at the moment, exploring them would definitely enrich the current design principles in structural DNA nanotechnology.

Self-assembled nucleic acid structures have the advantage to achieve higher structural complexity. However, their capability to construct topologically interesting structures was rarely explored. In chapter 4, we proposed the possibility for topological construction by controlling the two geometrical parameters – curvature and torsion. In principle, this strategy can be readily realized by 3D DNA origami or DNA bricks. Here, I would like to further suggest another possibility. Instead of assembling, the structures can be carved out of a larger precursor, like carving the sculpture out of the marble. Very recently, Yin's group has created the giant DNA bricks with dimensions up to 100 nm (personal communication). These large bricks can possibly be served as a sufficiently large "marble" to carve various topological structures by removing a certain set single-stranded tiles using their corresponding complementary strands. This technique may be called "DNA sculpture". However, the expense of making this giant DNA "marble" and the "carver" strands is immense, which cannot be withstood by many research groups.

5.2.2 Synthetic topological probes for other kinds of DNA topoisomerases besides Type IA topoisomerases

In our research, the ssDNA knot has been used to probe the activity of *E. coli* DNA topoisomerase I (a Type IA topoisomerase). Other type of DNA topoisomerases were also tested against the ssDNA knot in our experiment, but no detectable conversion was observed. This is understandable because Type IA topoisomerases target ssDNA and catalyze the topological conversion with the strand-passage mechanism. The other two most important types of topoisomerases are the Type IB and Type IIA. Type IB topoisomerases, acting like a swivel, catalyze the rotation of one strand in the dsDNA around the other strand. Type IIA topoisomerases use a mechanism similar to Type IA topoisomerases but with dsDNA substrates. A candidate synthetic probe for Type IB topoisomerases can be a synthetic supercoiled "miniplasmid" with the helix-based method. In our research, we also tested the dsDNA knots for probing Type IIA topoisomerases, but got negative results. It is likely that the dsDNA knots are of a too compact geometry for the enzyme actions. A dsDNA catenane, which has not been attempted in my research but should be easily constructed, might be a likely probe for Type IIA topoisomerases.

A desirable synthetic probe for topoisomerase should have the capability of translating the topological conversion to other signals that can be conveniently monitored, just as shown in our research that the unknotting of ssDNA knot can be detected by DNA amplification. This is beneficial for the development of high-throughput methods for screening topoisomerase inhibitors or engineering the topoisomerases for other applications. Therefore, strategies of signal conversion and, more favorably, signal amplification should be incorporated in the design of abovementioned two probes for Type IB and Type IIA topoisomerases, respectively.

5.2.3 Exploring new design strategies for RNA nanostructures

Comparing to its chemical cousin of DNA, RNA has two additional advantages making it more desirable for designing nanostructures. The first advantage of RNA is its higher structural diversity. The available DNA motifs can be used for designing nanostructure is very limited, and almost all the current DNA nanodesign strategies are based on the motif of DNA four-way junction[17] or its derivative. All the structures that have been realized by DNA should be theoretically recast by RNA. Additionally, considering the plethora of structural motifs of RNA, there is much room for improvement for the design of RNA nanostructures. It is very important to keep in mind that a versatile structural motif should be modular and with well-studied structural properties. This is why we took so many efforts in studying our newly designed bKL motif (presented in Chapter 4) before we can confidently put it into the RNA nanodesign toolbox.

The second advantage of RNA is that *in vitro* transcription can conveniently produce very long RNA strands with very few sequence constraints. The use of structural motifs of paranemic characteristic can further make it possible to create highly complex nanostructure using only one strand of RNA, which is dubbed single-stranded RNA origami. This concept was first realized in 2014 by Geary and Andersen by exploiting the RNA kissing loop (which is paranemic) to construct 2D origami structures[18]. It has been recently further developed by Andersen group to enable the construction of various 3D origami structures (personal communication). Yan group recently achieved single-stranded RNA origami structures with the motif of paranemic crossover (personal communication). Though larger structures with more structural complexity have been created with Yan's approach, these single-stranded RNA origami structures, unfortunately, cannot be folded cotranscriptionally. We reasonably expect that our newly designed bKL motif can also be incorporated in the design of more complex single-stranded RNA origami structures.

5.2.4 Engineering nucleic acid crystallization

The initial motivation for structural DNA nanotechnology is the construction of crystalline lattice with an infinite network of DNA multiway junctions and sticky end cohesions[19]. The position of every nucleotide, the crystal contacts, and the symmetry of the resulting lattice would be exactly predicted based on the designed model. Therefore, a crystal obtained this way is called a "rationally designed crystal". The first high-resolution rationally designed crystal was reported by Seeman's and Mao's groups collectively in 2009 based on the rigid structural motif of tensegrity triangle[20]. The development of rationally designed crystal is relatively slower comparing to other directions in the field of structural DNA nanotechnology. As far as I know, only three research groups are making efforts towards this direction. People's reluctance to be involved in this branch of research in turn reflects the daunting challenge associated with it.

However, as is always the case, this challenging pursuit will definitely prove itself to be rewarding in the end.

Herein, I present two suggestions that might be helpful for the future research of rationally designed crystal. Firstly, the structural motifs (multiway junctions) should be as simple as possible without compromising the rigidity. In Seeman and Mao's crystal, the multiway junction motif is the tensegrity triangle (as a six-way junction), which is formed with 7 strands (though with 3 distinct sequences). Though this motif fulfills the requirement for rigidity, the high complexity is unavoidably detrimental for the percentage of correctly formed motif. Considering the difficulty associated with designing sufficiently rigid structural motifs with a small number of strands, branching could be offered by synthetic organic cores, to which multiple strands of DNA can be connected to generate the artificial DNA junctions. As a result, a new category of materials, dubbed by myself as "NAOF" (Nucleic Acid-Organic Framework), could be envisaged. Secondly, the strength of sticky end cohesion should be minimized. Though the crystal is "rationally designed", it is important to note that the progress of crystal formation is still essentially crystallization. The progress of conventional DNA programmed self-assembly, which often comprises molecules no more than several thousand, exploits stronger cohesions provided by longer sticky ends (enthalpy-driven). In contrast, the process of crystallization for biomacromolecules, which requires millions of molecules to be in a regular register, is primarily entropically driven by the release of ordered water molecules from the site of packing contacts in the crystal. The forces maintaining the crystal contact, though important, are always very small. If these forces are too large, any packing error during the crystallization could not readily be corrected and poorly diffracted, if any, crystals are resulted. Therefore, 1-bp sticky end should be sufficient to guide the designed motifs to grow into the designed lattice. In fact, Seeman recently improved the resolution from 4.0 Å to 2.6 Å by changing the sticky end from 2 to 1 bp (personal communication).

It should be noted that the strategy of promoting RNA crystallization by creating closed geometry presented in our research cannot be regarded as rationally designed crystal because the crystal contact cannot be predicted. Notwithstanding, even the abovementioned rationally designed crystal is not purely rationally designed. There are still some numerous uncontrollable factors governing the real process of crystallization and the diffracting resolution that can be finally achieved. Admittedly, the power of contemporary human beings has substantially outreached our ancestors; the forces of Nature should still be respected and well utilized for now as well as in the foreseeable future.

5.3 References

1. Vinograd, J., Lebowitz, J., Radloff, R., Watson, R. & Laipis, P. The twisted circular form of polyoma viral DNA. *Proc. Natl. Acad. Sci. USA* 1965, **53**(5)**:** 1104-1111.
2. Wang, J. C. Interaction between DNA and an Escherichia-Coli Protein Omega. *J. Mol. Biol.* 1971, **55**(3)**:** 523-&.
3. Pommier, Y. *DNA Topoisomerases and Cancer*, 1 edn. Springer-Verlag New York, 2012.
4. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 1953, **171**(4356)**:** 737-738.
5. Rodley, G. A., Scobie, R. S., Bates, R. H. & Lewitt, R. M. A possible conformation for double-stranded polynucleotides. *Proc. Natl. Acad. Sci. USA* 1976, **73**(9)**:** 2959-2963.
6. Francl, M. Stretching topology. *Nat Chem* 2009, **1**(5)**:** 334-335.
7. Seeman, N. C. Synthetic DNA Topology. In: Sauvage JP, Dietrich-Buchecker C (eds). *Molecular Catenanes, Rotaxanes and Knots*. Wiley: Weinheim, 2007, pp 323-356.
8. Chen, J. H. & Seeman, N. C. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* 1991, **350**(6319)**:** 631-633.
9. Zhang, Y. & Seeman, N. C. A solid-support methodology for the construction of geometrical objects from DNA. *J. Am. Chem. Soc.* 1992, **114**(7)**:** 2656-2663.
10. Forgan, R. S., Sauvage, J. P. & Stoddart, J. F. Chemical topology: complex molecular knots, links, and entanglements. *Chem. Rev.* 2011, **111**(9)**:** 5434-5464.
11. Fujita, M., Fujita, N., Ogura, K. & Yamaguchi, K. *Nature* 1999, **400**(6739)**:** 52-55.

12. Stoddart, J. F. The chemistry of the mechanical bond. *Chem Soc Rev* 2009, **38**(6)**:** 1802-1820.
13. Ackermann, D. *et al.* A double-stranded DNA rotaxane. *Nature Nanotechnol.* 2010, **5**(6)**:** 436-442.
14. Schmidt, T. L. & Heckel, A. Construction of a structurally defined double-stranded DNA catenane. *Nano Lett.* 2011, **11**(4)**:** 1739-1742.
15. Liu, D., Chen, G., Akhter, U., Cronin, T. M. & Weizmann, Y. Creating complex molecular topologies by configuring DNA four-way junctions. *Nature Chem.* 2016.
16. Baas, N. A., Seeman, N. C. & Stacey, A. Synthesising Topological Links. *J. Math. Chem.* 2015, **53**(1)**:** 183-199.
17. Kallenbach, N. R., Ma, R.-I. & Seeman, N. C. An immobile nucleic acid junction constructed from oligonucleotides. *Nature* 1983, **305**(5937)**:** 829-831.
18. Geary, C., Rothemund, P. W. & Andersen, E. S. RNA nanostructures. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science* 2014, **345**(6198)**:** 799-804.
19. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* 1982, **99**(2)**:** 237-247.
20. Zheng, J. *et al.* From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature* 2009, **461**(7260)**:** 74-77.