Supplementary Information (SI) for Nanoscale Horizons.

Supporting Information

RNA nanostructures of three-letter coding with non-canonical base pairs

Jianqiu Zhao[1,2], Yan Qin[1,2], Qiancheng Xiong[3], Fang Fang[4,*] and Bryan Wei[1,2,*]

[1] School of Life Sciences, Tsinghua University, Beijing, 100084, China.

[2] Center for Synthetic and Systems Biology, Tsinghua University, Beijing, 100084, China.

[3] Bioprocessing Technology Institute, Agency for Science, Technology and Research (A*STAR), Singapore 138668, Singapore.

[4] Reproductive and Genetic Center, The First Affiliated Hospital of USTC, Center for Advanced Interdisciplinary Science and Biomedicine of IHM, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230026, China.

[*] Corresponding Author. Email: ffang24@ustc.edu.cn (F.F.) and bw@tsinghua.edu.cn (B.W.).

Methods

DNA templets and sequences design

Double strand DNA templets were obtained by multiple rounds of polymerase chain reactions. The sequences of DNA templets are shown in **Table S1**.

All the kissing loops (KLs) of RNA nanostructures, including branched KLs[1] and 180° KLs[2], were changed into exclusive A-U or G-C base pairs, with reverse complementary of corresponding KL sequences. For example, sequence of branched KLs from GGAGGCA to GGCGGCA and from GCGAGCA to GCGCGCA, sequence of 180° KLs from AAGGAGGCA to AAGGCGGCA. For other designs, 90° kink motifs with the sequence of AACUA[3], U was replaced by C for both of them are pyrimidine bases with more similar chemical structures. For triangle design, the space sequence of open three way junction (O3WJ) motifs[4] were changed from UUUU to AAAA, and GAAA which belongs to the GNRA family of tetraloops[5] with stems were designed to stabilize the three angles. ~5% or ~10% A·C base pairs were inserted in the duplex region of three-letter RNA sequences to replace G-C Watson-Crick base pairs. A·C base pairs were placed at least 3-bp apart from loops and 5'/ 3' ends and at least 2-bp apart from another A·C base pair with basically symmetric positions. The four-letter sequence of z-shaped motif and S derived from the designs in the literature.[1,2] All the RNA sequences were checked by Mfold[6] software to make sure that the secondary structures were correct.

Polymerase chain reactions

DNA templates were amplified by polymerase chain reactions using PrimeSTAR Max DNA Polymerase (TAKARA). For the templates were amplified step by step, 2 μM of forward and reverse primers were dissolved in 100 μL 1×PrimeSTAR Max Premix. The solution was initially denatured for 3 min at 98°C, and started 35 cycles at 98°C for 10 s, 55°C for 10 s, and 72°C for 1 min. Then kept the final extension at 72°C for another 5min. The first step PCR products worked as the templets of next-round PCR. The final DNA templates were purified by V-ELUTE Gel Mini Purification Kit (Beijing Zoman Biotechnology) and were stored at −20°C.

In vitro transcription

RNA nanostructures were synthesized by in vitro transcription. For 10 μL system, 200 ng DNA templates, 2 mM of each NTPs, 5 mM DTT, 26 mM $MgCl_2$, 2 mM spermidine, with RNase inhibitor (TAKARA), pyrophosphatase (New England Biolabs) and T7 RNA polymerase (New England Biolabs) were contained in the solution. The solution was incubated at 37°C for 2-3 hours, followed by DNase I (New England Biolabs) digestion of DNA templates. The RNA products were stored at -20°C or -80°C.

RNA nanostructures formation

For assembly of z-shaped motif of asymmetric arm lengths with KLs connected in exclusive A-U base pairs or G-C base pairs, here goes three different annealing procedures. Annealing procedure 1: 70°C/5 min, 50°C/30 min, 37°C/30 min, 22°C/30 min, and 4°C/30 min; annealing procedure 2: 70°C/5 min, 70°C to 50°C over 10 min, 50°C to 37°C over 25 min, 37°C to 4°C over 2 h, and 4°C/30 min; annealing procedure 3: 70 °C/5 min, 70°C to 40°C over 30 min, 40°C to 4°C over 15 h, and 4°C/30 min.

Different type of RNA nanostructures were annealed in different procedure, annealing procedure 2 for z-shaped motif with asymmetric or symmetric arm; annealing procedure 1 for triangle and square. Four-letter S and three-letter Z mixed nanostructures were annealed in procedure 2. Annealing buffer contained 40 mM Tris base, 20 mM acetic acid, 2 mM EDTA (1*TAE buffer) and 12.5 mM $MgCl_2$.

Atomic force microscopy (AFM) imaging

Add 2 µL RNA sample into freshly cleaved mica, with 50 µL scanning buffer (1*TAE buffer, 12.5 mM $MgCl_2$, 2 mM NiCl) and then incubated for 2 min. Samples were imaged under liquid ScanAsyst mode AFM (Bruker Multimode 8 with Nanoscope V controller) with the SNL-10 silicon nitride cantilever chip.

Urea PAGE

10% Urea PAGE gel was prepared with 29:1 acrylamide/bisacrylamide gel and was run at 50°C. The gel was stained by 1:10,000 dilute Sybr green II (Thermo Fisher Scientific) and was detected by Amersham Typhoon scanner equipped with Cy2 filter.

Supporting Figures

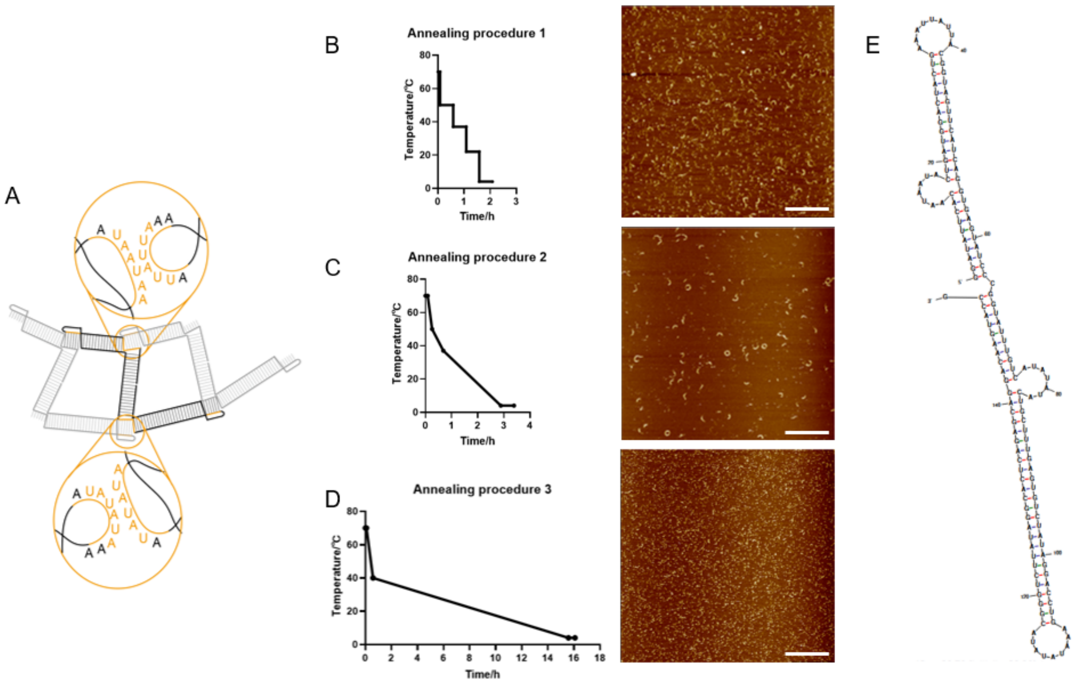

Figure S1. Z-shaped motif of asymmetric arm lengths with KLs connected in exclusive A-U base pairs. (A) Strand diagram, with KL interactions are depicted in orange. (B) Atomic force microscopy (AFM) results following annealing procedure 1, a step cooling from 70°C to 4°C. (C) AFM results following annealing procedure 2, a short ramp cooling from 70°C to 4°C (fast cooling from 70°C to 37°C and slow cooling from 37°C to 4°C). (D) AFM results following annealing procedure 3, a long ramp cooling from 70°C to 4°C (fast cooling from 70°C to 40°C and overnight cooling from 40°C to 4°C). For B-D, left: annealing procedures; right: AFM images of A-U base-paired-KLs. Scale bars: 400 nm. (E) The target secondary structure predicted by Mfold.

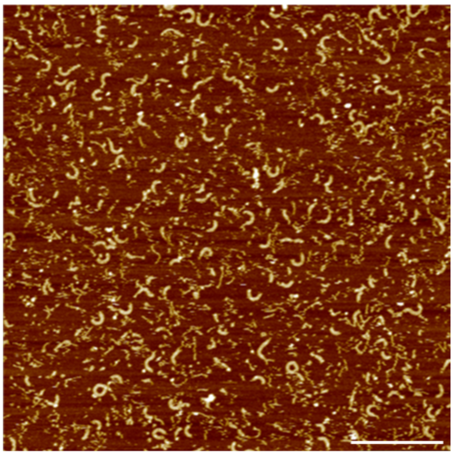

Figure S2. Additional AFM results of z-shaped motif of asymmetric arm lengths with KLs connected in exclusive A-U base pairs in annealing procedure 1. Extremely rare ring appearance in this large-scale image. Scale bar: 400 nm.

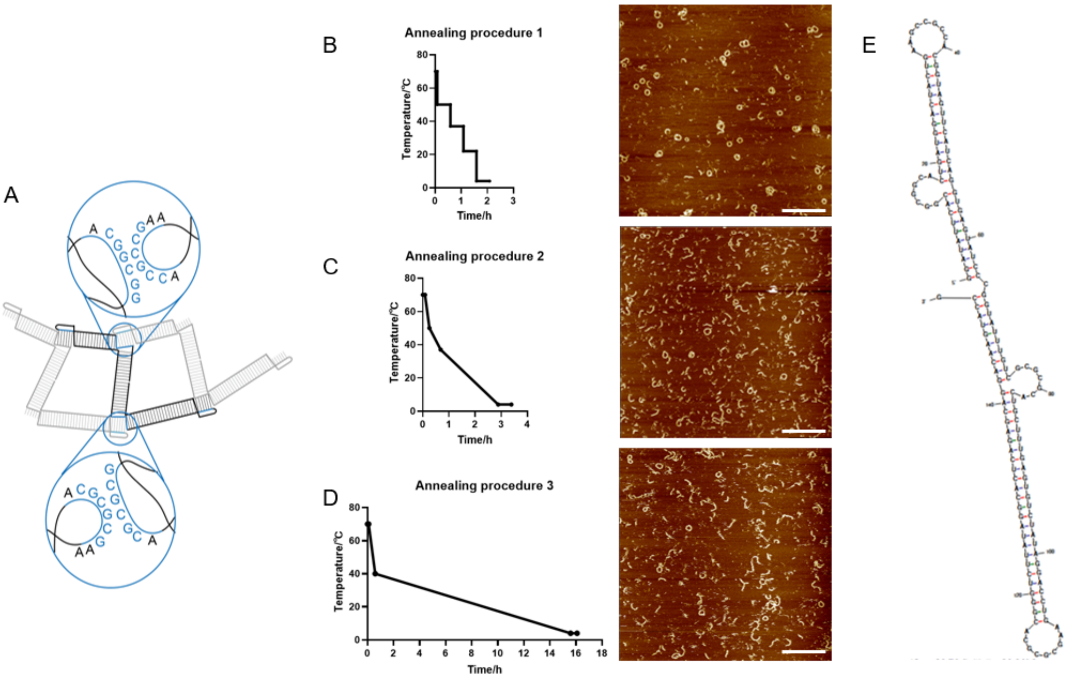

Figure S3. Z-shaped motif of asymmetric arm lengths with KLs connected in exclusive G-C base pairs. (A) Strand diagram, with KL interactions are depicted in blue. (B) AFM results following annealing procedure 1, a step cooling from 70°C to 4°C. (C) AFM results following annealing procedure 2, a short ramp cooling from 70°C to 4°C (fast cooling from 70°C to 37°C and slow cooling from 37°C to 4°C). (D) AFM results following annealing procedure 3, a long ramp cooling from 70°C to 4°C (fast cooling from 70°C to 40°C and overnight cooling from 40°C to 4°C). For B-D, left: annealing procedures; right: AFM images of G-C base-paired-KLs. Scale bars: 400 nm. (E) The target secondary structure predicted by Mfold.

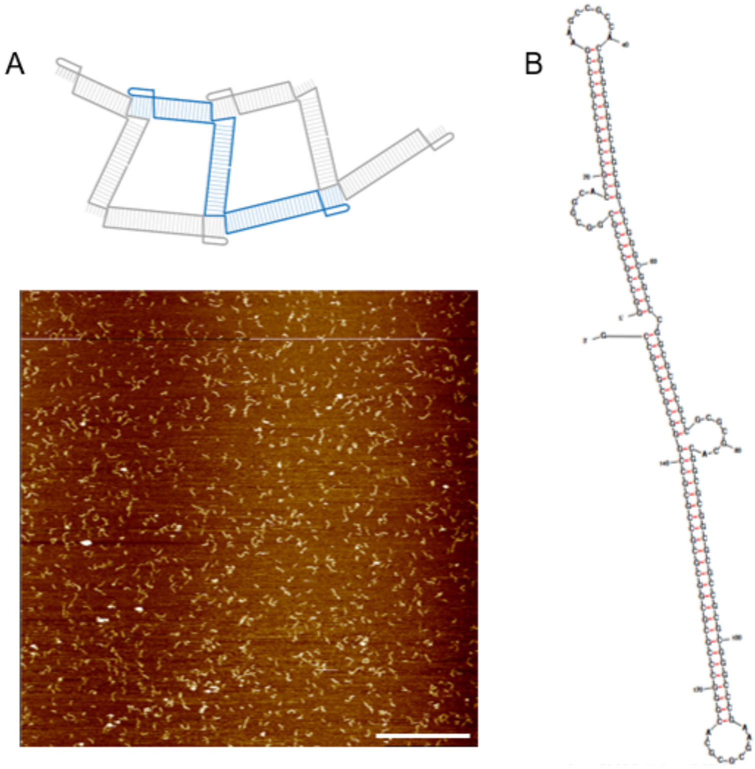

Figure S4. Z-shaped motif of asymmetric arm lengths with no A·C base pair inserted. (A) AFM results. Top: strand diagram (G-C base pairs depicted in blue); bottom: AFM images. Scale bar: 400 nm. (B) The target secondary structure predicted by Mfold.

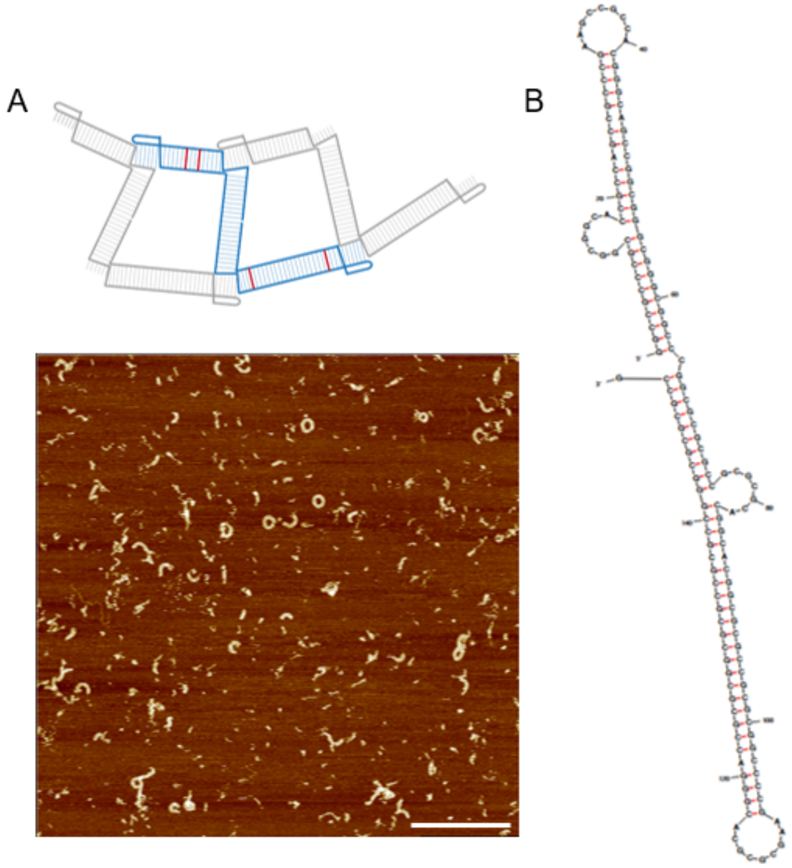

Figure S5. Z-shaped motif of asymmetric arm lengths with ~5% A·C base pairs. (A) AFM results. Top: strand diagram (A·C base pairs depicted in red and G-C base pairs in blue); bottom: AFM images. Scale bar: 400 nm. (B) The target secondary structure predicted by Mfold.

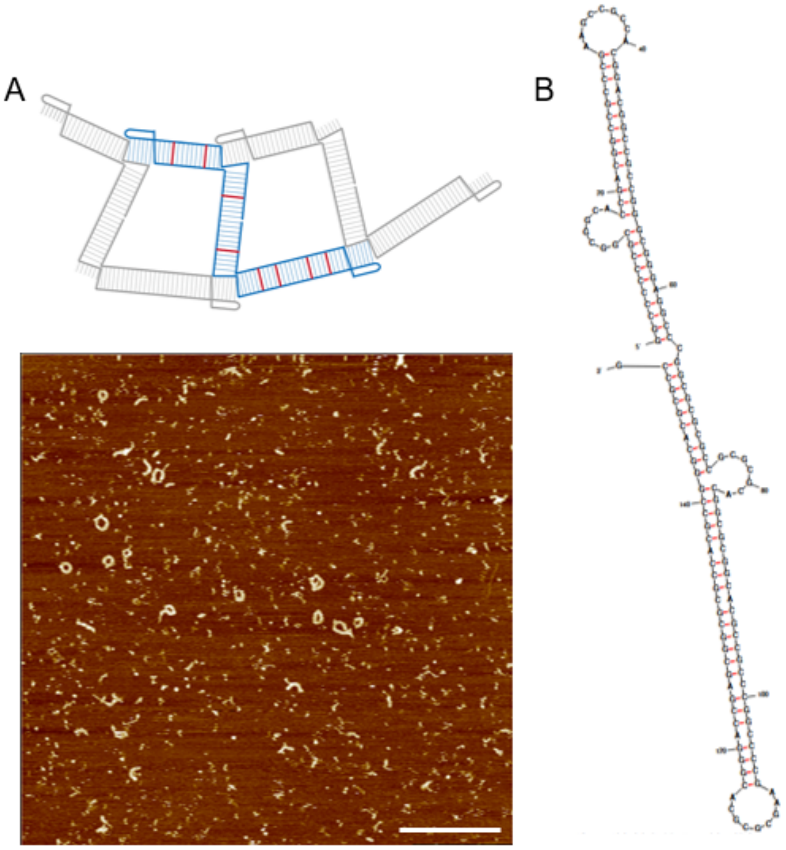

Figure S6. Z-shaped motif of asymmetric arm lengths with ~10% A·C base pairs. (A) AFM results. Top: strand diagram (A·C base pairs depicted in red and G-C base pairs in blue); bottom: AFM images. Scale bar: 400 nm. (B) The target secondary structure predicted by Mfold.

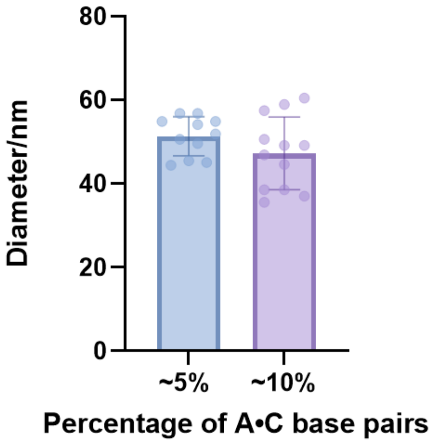

Figure S7. Statistics of diameter distribution for z-shaped motif with ~5% and ~10% A·C inserted. Data are presented as mean ± SD (N=11 for~5% A·C, N=12 for ~10% A·C).

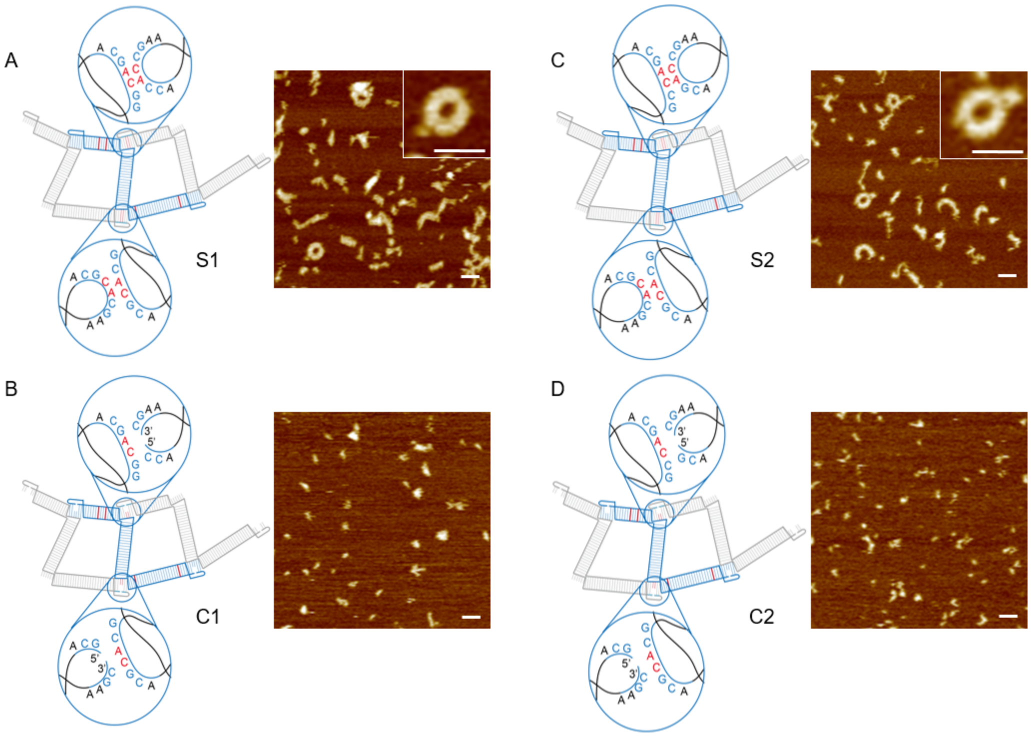

Figure S8. Z-shaped motif of asymmetric arm lengths with A·C base pairs designed in KLs. (A) Strand diagram of three-letter z-shaped motif of asymmetric arm lengths with KLs of sequence design S1, and had a 0.5% yield. (B) The negative control (C1) of (A) with A·C base pairs omitted. (C) Strand diagram of three-letter z-shaped motif of asymmetric arm lengths with KLs sequence design S2, and had a 3.3% yield. (D) The negative control (C2) of (C) with A·C base pairs omitted. A·C interaction was highlight in red. Scale bars: 50 nm.

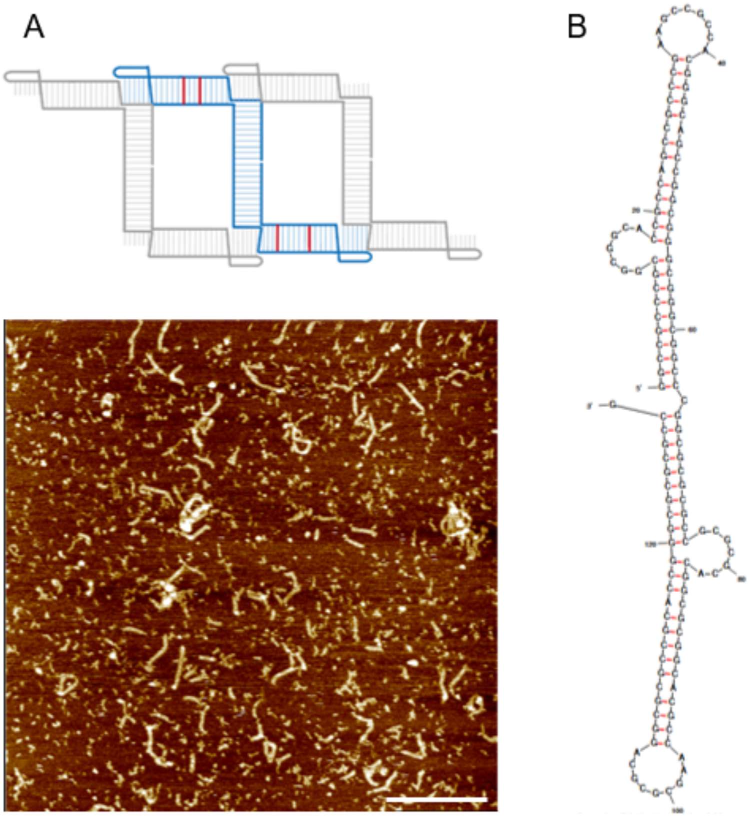

Figure S9. Assembly of z-shaped motif of symmetric arm lengths in three-letter coding. (A) AFM results. The longest contour of ribbons was ~228 nm. Top: strand diagram (A·C base pairs depicted in red and G-C base pairs in blue); bottom: AFM images. Scale bar: 400 nm. (B) The target secondary structure predicted by Mfold.

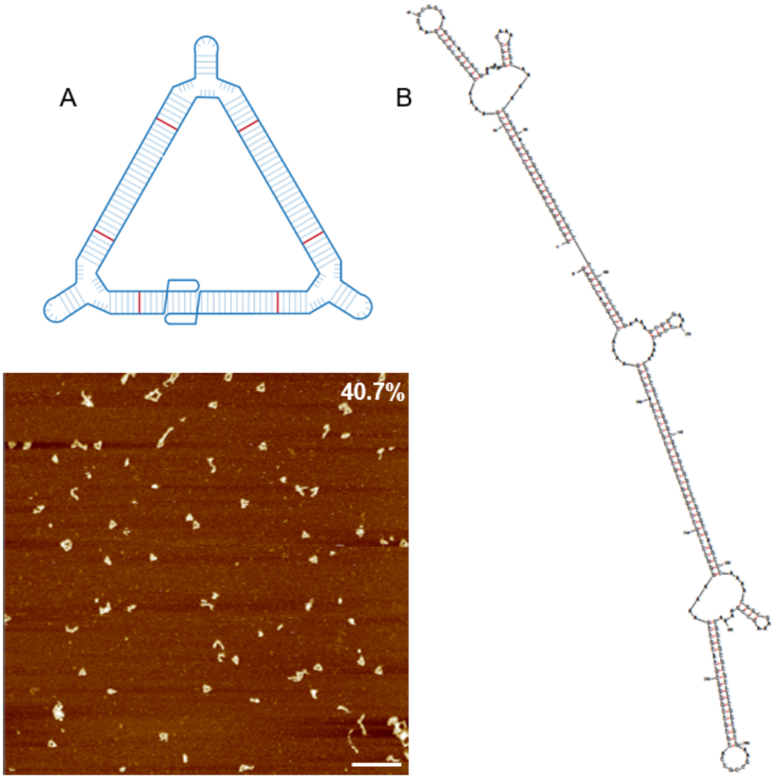

Figure S10. Assembly of triangle in three-letter coding. (A) AFM results with the yield of 40.7% (N=64). Top: strand diagram (A·C base pairs depicted in red and G-C base pairs in blue); bottom: AFM images. Scale bar: 100 nm. (B) The target secondary structure predicted by Mfold.

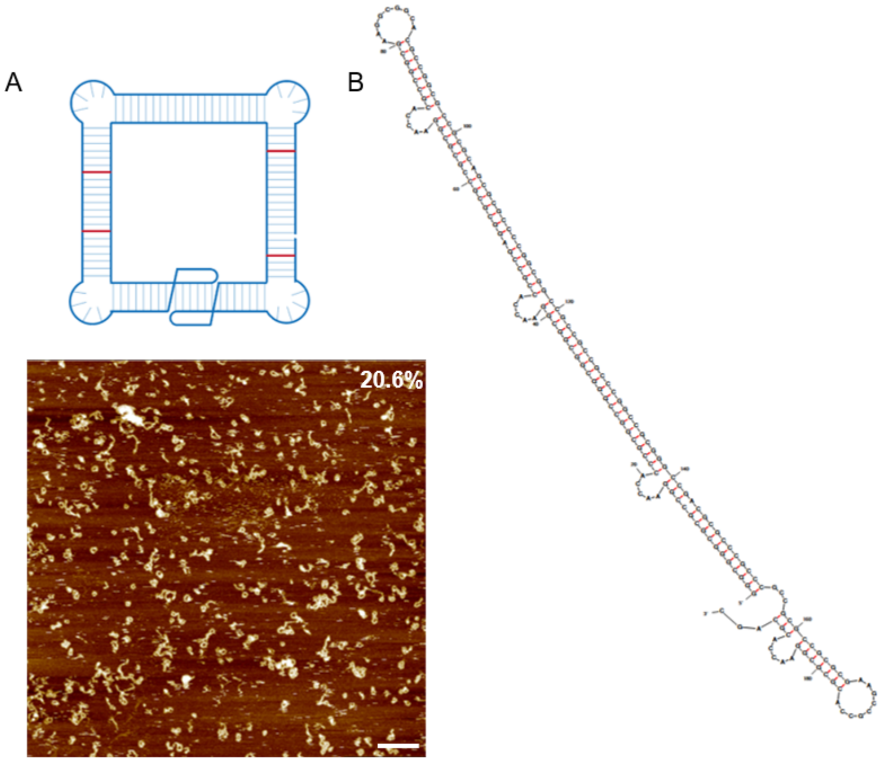

Figure S11. Assembly of square in three-letter coding. (A) AFM results with the yield of 20.6% (N=398). Top: strand diagram (A·C base pairs depicted in red and G-C base pairs in blue); bottom: AFM images. Scale bar: 100 nm. (B) The target secondary structure predicted by Mfold.

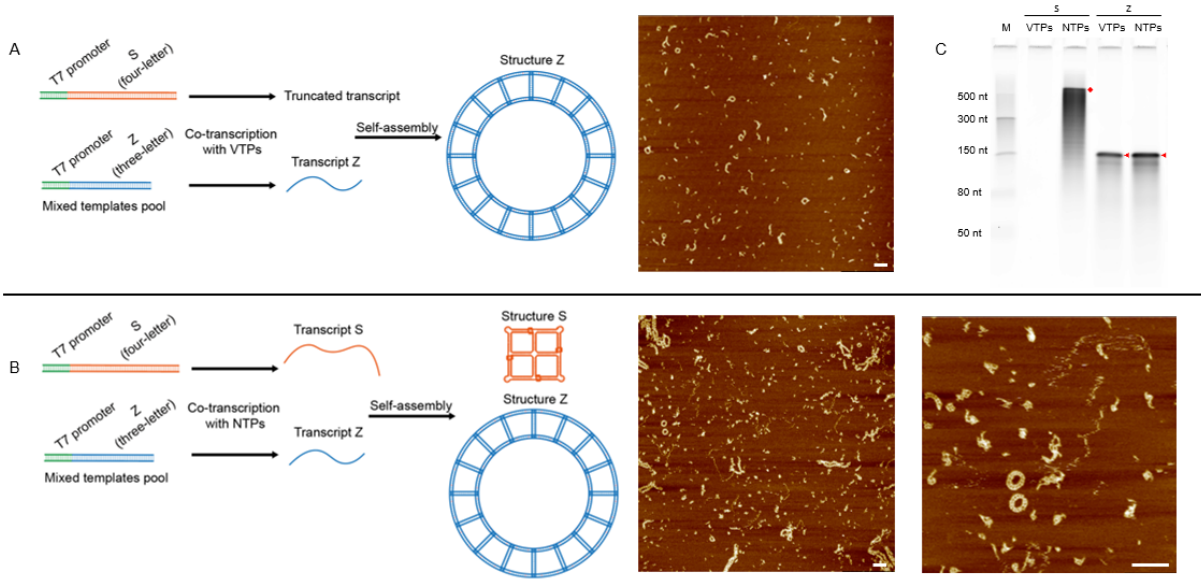

Figure S12. Co-transcription system with mixed three-letter Z and four-letter S templates. (A) Co-transcription with VTPs (ATP, CTP and GTP). Left: system diagram. Right: AFM images. (B) Co-transcription with NTPs (ATP, CTP, GTP and UTP). Left: system diagram. Middle: AFM images. Right: zoom in AFM images. (C) 10% Urea PAGE results of *in vitro* transcription of three-letter Z and four-letter S templates individually. Four-letter S templates cannot be transcribed by VTPs. Scale bars: 100 nm.

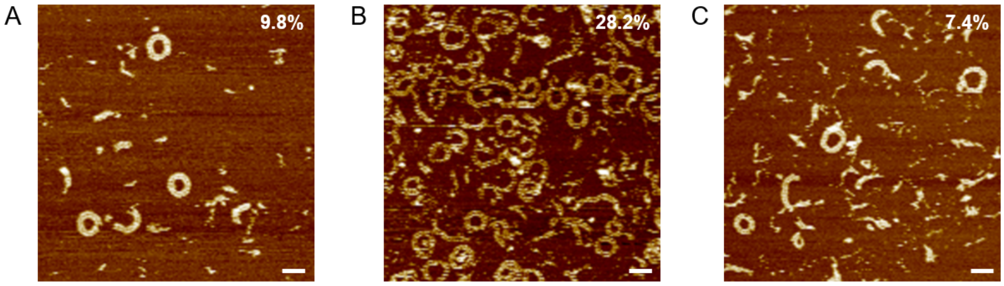

Figure S13. Z-shaped motifs in three-letter coding and four-letter coding. (A) Three-letter system with a yield of 9.8%. (B) Four-letter system with a yield of 28.2%. (C) Z-shaped motifs with the same backbone sequences of four-letter system but KLs connected in exclusive G-C base pairs with a yield of 7.4%. All of the AFM results are collected following annealing procedure 2. Scale bars: 50 nm.

Supporting Tables

Structure	**DNA template sequences**
Z-shaped motif of asymmetric arm lengths with KLs connected in exclusive A-U base pairs.	**GGGTAATACGACTCACTATA**GGATATTCACAATAATACTGATGGACTACTGAAATTATTACGGTAGTTCATCAGGTGAGTATCCCGGTATTTGTCATATATACTGCTTTGAGTGTCTATAGGACCTGAAATATATACGGGTCTTATAGGCACTCAGAGCAGGACAAGTACCG
Z-shaped motif of asymmetric arm lengths with KLs connected in exclusive G-C base pairs.	**GGGTAATACGACTCACTATA**GGATATTCACGGCGGCACTGATGGACTACTGAAGCCGCCACGGTAGTTCATCAGGTGAGTATCCCGGTATTTGTCGCGCGCACTGCTTTGAGTGTCTATAGGACCTGAAGCGCGCACGGGTCTTATAGGCACTCAGAGCAGGACAAGTACCG
Z-shaped motif of asymmetric arm lengths with no A·C base pair inserted.	**GGGTAATACGACTCACTATA**GGCCGCCCGCGGCGGCACCGCCGGCCGCCCGAAGCCGCCACGGGCGGCCGGCGGGCGGGCGGCCCGGCGCGCGCCGCGCGCACGGCGCGGCGCGCCGCGCGGGCCCGAAGCGCGCACGGGCCCGCGCGGCGCGCCGCGCCGGGCGCGCGCCG
Z-shaped motif of asymmetric arm lengths with ~5% A·C base pairs.	**GGGTAATACGACTCACTATA**GGCCGCCCGCGGCGGCACCGCCAGCCGCCCGAAGCCGCCACGGGCAGCCGGCGGGCGGGCGGCCCGGCGCGCGCCGCGCGCACGGCACGGCGCGCCGCGCGGCCCCGAAGCGCGCACGGGACCGCGCGGCGCGCCGCGCCGGGCGCGCGCCG
Z-shaped motif of asymmetric arm lengths with ~10% A·C base pairs.	**GGGTAATACGACTCACTATA**GGCCCCCCGCGGCGGCACCGACGGCCGCCCGAAGCCGCCACGGACGGCCGCCGGGCGGGAGGCCCGGCGCGCGCCGCGCGCACGGCGCGGCACGCCGCCCGGCCCCGAAGCGCGCACGGGACCGAGCGGCGCGCCACGCCGGGCACGCGCCG
Z-shaped motif of symmetric arm lengths.	**GGGTAATACGACTCACTATA**GGCCGCCCGCGGCGGCACCGCCAGCCGCCCGAAGCCGCCACGGGCAGCCGGCGGGCGGGCGGCCCGGCGCGCGCCGCGCGCACGGCGCGGCACGCCAAGCGCGCAGGCGCGCCGCACCGGGCGCGCGCCG
Triangle.	**TTCTAATACGACTCACTATA**GGCGGGCGGGCGCCCGCGCCCGAAAACGCGGCGCCGAAGGCGGCACGGCACCGCGAAAAGCGCGAAAGCGCAAAACGGGCACGGGCGCCCGCCCGCCCCCGGCCCGGGAAAAGCGGGAAACCGCAAAACCGCGCCGGGGCGCGGCGGCCCGGGCGAGGCCCAAAACGGCGAAAGCCGAAAACGGCGCGCGCCCCGGCGCGAAGCCGCCACGCGCCGGGGCGCACGCCGAAAAGGGCCCCGCCCGGGCCGCCGCGCCCCGACGCGGAAAACCCGGACCGGG
Square.	**TTCTAATACGACTCACTATA**GGGCGGGCGCGCCGGAACCACCCGCGGCCGGGCGGCGGCGGAACCACCGCCGAGGCGCGCCGCGCGGAACCACGCCGGCGAAGGCGGCACGCCGGCGCCGCGCAGCGCGCCCCGGCGGCCGCCGCCGCCCGGCCGCGGGCCGAC

	GCGCCCGCCCGCCGCGCCGCGCGAAGCCGCCACGCGCGGAACCACGCAGC
Four-letter S.	**TTCTAATACGACTCACTATA**GGTCTAGAGGTGAGTGTGTAACTACGGTCTAGTCGTGGAGGAACTGGTCGTAGTTCCTCCGCGACTAGGCCGACACACTCACCCATCACTCCAGCCACTTCCAGGGAAACTCCGGCTTCGTGTGATCTCGACATTCTGGGGATACGACCTGGTTACCCTCCTAGACTATCAACTAGGTGGGAGTGAATGCTGTTCTGGGAGTAGAACAGCGTTCACTCTCACCGATAGTCTGGGAGGGTAATCAGCCCAGAGTGTCGAGATTACACGTTCGTACAGACACGTCTGAGTGGGGATACTCCCTGAACTCAGGTGCTTGAGCTCAACTAGGAGTTGAGTCTTCTCAACGTGTTGCTACGTTGAGAAGATTCAACTCCGAGTTCAAGCACCTGGGTTCAGCCCACTCAGGCGTGTCTGTGCGTTGTGAGTATCGACAGGATAGAGGGATAGCAACGTGCAGTGGAGAACATTACACAACTAGTGTCAGAGCTAGCAGGTCGTGAGCGTACGACCTGCTGGCTCTGACACGTGTGATGTTCTCCATTGCACGCCTCTGTCCTGTCGATGCTCACTTGCTGGAGTTTCTCTGGAAGTGGGATACGCTCTGGAGTGATG
Three-letter Z (the same as z-shaped motif of asymmetric arm lengths with ~5% A·C base pairs).	**GGGTAATACGACTCACTATA**GGCCGCCCGCGGCGGCACCGCCAGCCGCCCGAAGCCGCCACGGGCAGCCGGCGGGCGGGCGGCCCGGCGCGCGCCGCGCGCACGGCACGGCGCGCCGCGCGGCCCCGAAGCGCGCACGGGACCGCGCGGCGCGCCGCGCCGGGCGCGCGCCG
Z-shaped motif of asymmetric arm lengths with A-C base pairs designed in KLs (S1).	**GGGTAATACGACTCACTATA**GGCCGCCCGCGGCAGCACCGCCAGCCGCCCGAAGCCACCACGGGCAGCCGGCGGGCGGGCGGCCCGGCGCGCGCCGCACGCACGGCACGGCGCGCCGCGCGGCCCCGAAGCACGCACGGGACCGCGCGGCGCGCCGCGCCGGGCGCGCGCCG
The negative control of S1 (C1).	**GGGTAATACGACTCACTATA**CCACGGGCAGCCGGCGGGCGGGCGGCCCGGCGCGCGCCGCACGCACGGCACGGCGCGCCGCGCGGCCCCGAAGC; **GGGTAATACGACTCACTATA**GCACGGGACCGCGCGGCGCGCCGCGCCGGGCGCGCGCCGGGCCGCCCGCGGCAGCACCGCCAGCCGCCCGAAGC
Z-shaped motif of asymmetric arm lengths with A-C base pairs designed in KLs (S2).	**GGGTAATACGACTCACTATA**GGCCGCCCGCGCCAGCACCGCCAGCCGCCCGAAGCCAGCACGGGCAGCCGGCGGGCGGGCGGCCCGGCGCGCGCCGCACGCACGGCACGGCGCGCCGCGCGGCCCCGAAGCACGCACGGGACCGCGCGGCGCGCCGCGCCGGGCGCGCGCCG
The negative control of S2 (C2).	**GGGTAATACGACTCACTATA**GCACGGGCAGCCGGCGGGCGGGCGGCCCGGCGCGCGCCGCACGCACGGCACGGCGCGCCGCGCGGCCCCGAAGC; **GGGTAATACGACTCACTATA**GCACGGGACCGCGCGGCGCGCCGCGCCGGGCGCGCGCCGGGCCGCCCGCGCCAGCACCGCCAGCCGCCCGAAGC

Z-shaped motifs in four-letter coding.	**GGGTAATACGACTCACTATA**GGATATTCACGGAGGCACTGATGGACTACTGAAGCCTCCACGGTAGTTCATCAGGTGAGTATCCCGGTATTTGTCGCGAGCACTGCTTTGAGTGTCTATAGGACCTGAAGCTCGCACGGGTCTTATAGGCACTCAGAGCAGGACAAGTACCG

Table S1. Sequences of DNA templates. All the T7 promoters are marked as bold. All the KLs are in red. The open three way junctions (O3WJ) are in green. 'GAAA' tetraloops are in purple. 90° kink sequences are in blue. The backbone sequences of z-shaped motif of asymmetric arm lengths with KLs connected in exclusive A-U or G-C base pairs together with four-letter coded construct derived from the original designs in the literature,[1] the sequence S also derived from original literature.[2]

	Tiles$_{\text{min}}$	**Tiles$_{\text{max}}$**	**Diameter$_{\text{min}}$/nm**	**Diameter$_{\text{max}}$/nm**	**N**
Z-shaped motif of asymmetric arm lengths with KLs connected in exclusive A-U base pairs (annealing procedure 2).	13	16	~50	~60	7
Z-shaped motif of asymmetric arm lengths with KLs connected in exclusive G-C base pairs (annealing procedure 1).	13	18	~42	~70	20
Z-shaped motif of asymmetric arm lengths with KLs connected in exclusive G-C base pairs (annealing procedure 2).	11	20	~48	~75	6
Z-shaped motif of asymmetric arm lengths with KLs connected in exclusive G-C base pairs (annealing procedure 3).	11	16	~46	~68	7
Z-shaped motif of asymmetric arm lengths with ~5% A·C base pairs.	12	16	~44	~56	11
Z-shaped motif of asymmetric arm lengths with ~10% A·C base pairs.	7	16	~36	~60	12

Table S2. Stastics of tiles and diameter for one closed ring based on z-shaped motif of asymmetric arm lengths. N represents the number of rings.

	Yield	**N**
Z-shaped motif of asymmetric arm lengths with KLs connected in exclusive A-U base pairs (annealing procedure 2).	6.6%	6
Z-shaped motif of asymmetric arm lengths with KLs connected in exclusive G-C base pairs (annealing procedure 1).	27.8%	20
Z-shaped motif of asymmetric arm lengths with KLs connected in exclusive G-C base pairs (annealing procedure 2).	7.4%	10
Z-shaped motif of asymmetric arm lengths with KLs connected in exclusive G-C base pairs (annealing procedure 3).	8.5%	9
Z-shaped motif of asymmetric arm lengths with ~5% A·C base pairs.	9.8%	8
Z-shaped motif of asymmetric arm lengths with ~10% A·C base pairs.	14.3%	18
Z-shaped motif of asymmetric arm lengths with A-C base pairs designed in KLs (S1).	0.5%	2
Z-shaped motif of asymmetric arm lengths with A-C base pairs designed in KLs (S2).	3.3%	9
Z-shaped motifs in four-letter coding.	28.2%	64

Table S3. Stastics of yield for rings based on z-shaped motif of asymmetric arm lengths. Yield analysis is based on the ratio of ring structures and all ribbon-like structures identifiable. N represents the number of rings.

References

[1] D. Liu, C. W. Geary, G. Chen, Y. Shao, M. Li, C. Mao, E. S. Andersen, J. A. Piccirilli, P. W. K. Rothemund and Y. Weizmann, *Nat. Chem.*, 2020, **12**, 249-259.

[2] M. Li, M. Zheng, S. Wu, C. Tian, D. Liu, Y. Weizmann, W. Jiang, G. Wang and C. Mao, *Nat. Commun.*, 2018, **9**, 2196.

[3] S. M. Dibrov, H. Johnston-Cox, Y. Weng and T. Hermann, *Angew. Chem. Int. Ed.*, 2007, **46**, 226-229.

[4] A. Lescoute and E. Westhof, *RNA*, 2006, **12**, 83-93.

[5] H. A. Heus and A. Pardi, *Science*, 1991, **253**, 191-194.

[6] M. Zuker, *Nucleic Acids Res.*, 2003, **31**, 3406–3415.