Supporting Information

Single-molecule study reveals novel rod-like structures formed by the thrombin aptamer repeat sequence

Jianyu Liu[a], Wei Feng[b], Wenke Zhang*[a]

[a]State Key Laboratory of Supramolecular Structure and Materials, College of Chemistry, Jilin University, Changchun 130012, People's Republic of China

[b]Institute of Atomic and Molecular Physics, Jilin University, Changchun 130012, People's Republic of China

*Tel: +86 431 85159203; Fax: +0431-85159203; Email: zhangwk@jlu.edu.cn

Table of Contents

Sequences of oligonucleotides:

TBA forming ssDNA for single-molecule stretching:

Flank one: 5'-Phosphate-GGAACGCAACAATCCTACGAC-3'

Flank two: 5'-TCACTCCACTCTACCG-3'

TBA: 5'-GGTTGGTGTGGTTGG-3'

Monomeric sequence: 5'-Phosphate-GTAGCCGGTAGAGTGGAGTGATTT(TBA)TTTGTCGTAGGATTGTTGC-3'

Dimeric sequence: 5'-Phosphate-GTAGCCGGTAGAGTGGAGTGATTT(TBA)T_{20}(TBA)TTTGTCGTAGGATTGTTGC-3'

Control sequence for agarose gel electrophoresis: 5'-Phosphate-GTAGCCGGTAGAGTGGAGTGATTT(TBA)T_{20}AGTCCGTGGTAGGGCAGGTTGGGGTGACTTTTGTCGTAGGATTGTTGC-3'

N20 template for PC-RCA: 5'-GATCCTAAAACCAACCACACCAACCA$_{20}$CCAACCACACCAACCAAAACCACAC-3'

N30 template for PC-RCA: 5'-GATCCTAAAAAAAAACCAACCACACCAACCA$_{30}$CCAACCACACCAACCAAAAAAAAACCACAC-3'

PC-RCA ssDNA primer: 5'-NH_2-TTAGGATCGTGTGGTT-3'

Thymine sequence template for RCA: 5'-Phosphate-AAAAAAAAAAAAAAAAAAAAAAAAAAAA-3'

Thymine sequence primer for RCA: 5'-TTTTTTTTTTTTTTT-3'

ssDNA for CD:

Monomer: TBA

Dimer: TBA-T_{20}-TBA

Trimer: TBA-T_{20}-TBA-T_{20}-TBA

Materials

DNA oligonucleotides used in this study were purchased from Sangon Biotech (Shanghai) Co., Ltd. Enzymes were purchased from New England Biolabs (NEB, Ipswich, MA), and all of the chemicals (>99% in purity) were purchased from Sigma-Aldrich. The surface-functionalized beads for the single-molecule experiments were obtained from Thermo Fisher Scientific Inc. The amino-functionalized reference beads (PS-NH_2) were purchased from Polysciences, Inc.

Circular dichroism spectroscopy experiments

CD spectroscopy has been proven to be a sensitive technique for determining the conformation of G4 DNA. We used CD to determine the overall conformations of TBA sequences with different copy numbers at room temperature. Spectra were recorded in experiment buffer (100 mM KCl, 10 mM Tris-HCl, pH 7.4). The concentration per copy number of TBA was 10 μM. CD measurements were carried out using a 0.2-cm path cell. Data were obtained with a 1-nm slit width from 330 to 220 nm. The CD spectrum of the buffer solution was subtracted from that of the sample containing DNA, and the CD spectra were averaged over three scans.

Quartz crystal microbalance with dissipation (QCM-D) experiment

Quartz crystal microbalance with dissipation monitoring (QCM-D) has been proven to be a powerful research tool to investigate in situ interactions between nanoparticles and different functionalized surfaces in liquids. QCM-D can also be used to quantitatively determine adsorption kinetics of polymers, DNA and proteins from solutions on various substrate surfaces while providing insights into conformations of adsorbed molecules. We use QCM to prove the biotin were modified on ssDNA. As shown in Figure S8, the substrates were blocked by T sequence ssDNA and streptavidin were added until frequency change reaches equilibrium that means streptavidin is well adsorbed on the substrate. Subsequently, NH_2-ssDNA (produced by RCA with phi29 DNA polymerase) and NH_2-ssDNA-biotin (produced by PC-RCA) were introduced into the system until frequency change equilibrium is reached. The frequency is named A when NH_2-ssDNA and NH_2-ssDNA-biotin were added. After rinsing with

water, the equilibrated frequency is named B. The frequency changes were compared between A and B. For the NH_2-ssDNA, the frequency A is less than the frequency B, see Figure S8a, but the opposite in Figure S8b. This indicates that biotin has been successfully introduced to NH_2-ssDNA-biotin.

Calculation of the change in contour length (ΔLc)

The change in extension (Δz) of the unfolding of monomeric and dimeric TBA (Figure 1) at a certain force (F) was converted to the change in contour length (ΔLc) using the worm-like-chain model[1,2] given below:

$$\frac{\Delta z}{\Delta Lc} = 1 - \frac{1}{2}\left(\frac{k_B T}{F L_P}\right)^{\frac{1}{2}} + \frac{F}{K} \tag{1}$$

Force extension curves of ssDNA (Figure 2d) were fitted freely-jointed chain (FJC) model[3]:

$$z = Lc\left[\coth\left(\frac{2FL_P}{k_B T}\right) - \frac{k_B T}{2FL_P}\right]\left(1 + \frac{F}{K}\right) \tag{2}$$

The change in extension (Δz) of the unfolding event in ssDNA extension curves (Figure 2c) at a certain force (F) was converted to the change in contour length (ΔLc) using the freely-jointed chain model given below[4]:

$$\Delta z = \Delta Lc\left[\coth\left(\frac{2FL_P}{k_B T}\right) - \frac{k_B T}{2FL_P}\right]\left(1 + \frac{F}{K}\right) \tag{3}$$

WLC, Worm-like chain; FJC, freely-jointed chain; F, force; z, extension (end-to-end distance); Δz, the change in extension; Lp, persistence length; Lc, contour length; ΔLc, the change in contour length; K, elastic modulus; k_BT, Boltzmann's constant times absolute temperature. The following parameters Lp = 1 nm, K ~800 pN[5] were used for ssDNA. And a contour length value of 0.56 nm for an individual nucleotide[6,7] is used during the calculation.

Calculation of the unfolding rate constants k0 u

The unfolding force distribution was analyzed by Bell's model which describes the force-dependent unfolding rate[7] as

$$k_u(f) = k_u^0 \, exp(\Delta x_u f / k_B T) \quad (1)$$

where k_B is the Boltzmann constant, T is absolute temperature, k0 u the unfolding rate constants, and Δx_u is the transition distance to the transition state. It predicts an unfolding force distribution with a single force peak[7] pBell unfold(f) as

$$p_{unfold}^{Bell}(f) = \frac{k_u^0}{r} exp\left\{\frac{\Delta x_u f}{k_B T} + \frac{k_B T k_u^0}{\Delta x_u r}\left[1 - exp\left(\frac{\Delta x_u f}{k_B T}\right)\right]\right\} \quad (2)$$

where r is the loading rate. The histograms of unfolding forces (Figure 3) were fitted by equation (2), which produced corresponding fitting parameters, k0 u which was summarized in Table S4. For the monomeric TBA with TB, the unfolding force shows two peaks. We determined k0 u Ⅱ= (1.0 ± 0.3) × 10^{-3} s^{-1} (average ± standard error), and k0 u Ⅲ= (0.9 ± 0.2) × 10^{-3} s^{-1}. Uncertainties represent the standard error. Our results have shown that the TBA sequences can form multiple G4's, and the forms (Ⅰ, Ⅱ and Ⅲ) have slow unfolding rate constants in the order of 10^{-3} s^{-1} at zero force. The k0 u value for monomeric TBAs have the same order of magnitude as previous researcher by a nanopore encapsulating single-molecule method[8]. In addition, the unfolding rate constants of rod-like TBA structures are basically the same order of magnitude as that of monomer. This result indicates that the presence of rod-like TBA does not significantly change TBA unfolding rate constants in the different copy numbers.

Preparation of dsDNA handles

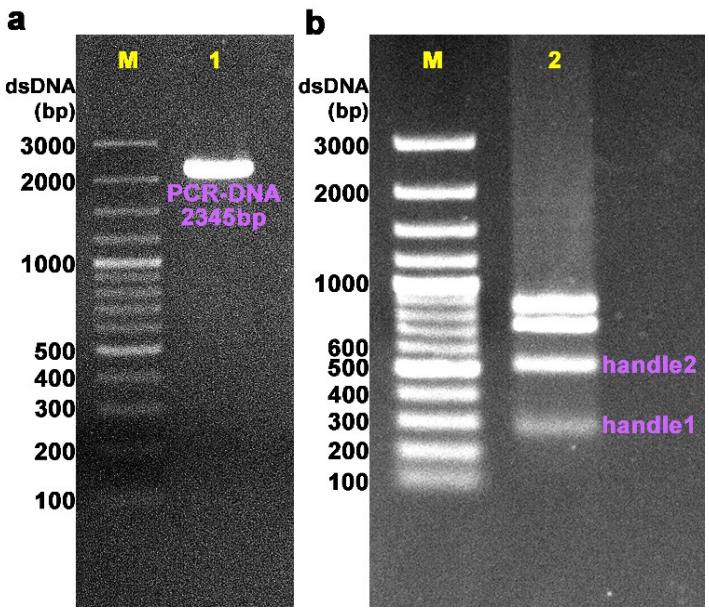

Figure S1. Agarose gel electrophoresis of 2345 bp dsDNA and the 262 bp and 509 bp dsDNA handles by digestion with HgaI restriction enzymes. The lane M represents DNA marker and the lane 1 represents the PCR production. The lane 2 represents the digestion product by HgaI restriction enzyme. The stripes of dsDNA handles (262 bp handle 1 and 509 bp handle 2) can be clearly identified on the DNA gel.

Preparation and characterization of dsDNA and TBA construct

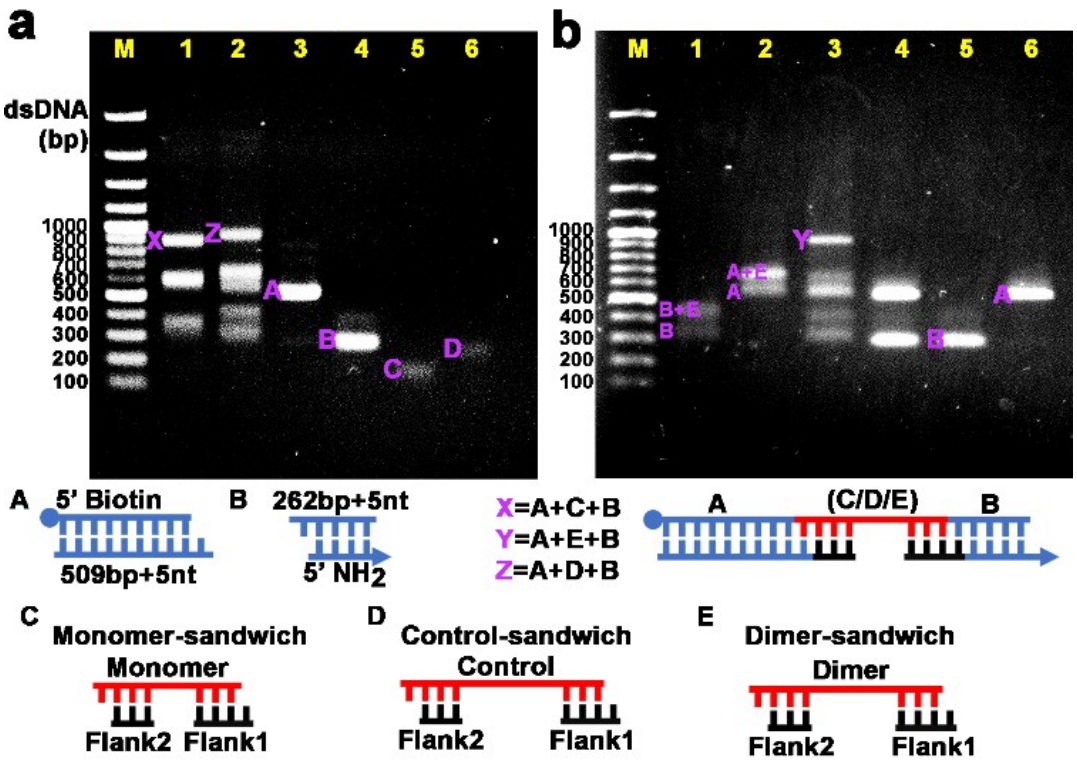

Figure S2. Agarose gel electrophoresis of preparation of Monomeric and Dimeric TBA construct for SMFS. (a) Lane M: dsDNA marker; Lane 1: A+B+C; Lane 2: A+B+D; Lane 3: A; Lane 4: B; Lane 5: C; Lane 6: D; (b) Lane M: dsDNA marker; Lane 1: B+E; Lane 2: A+E; Lane 3: A+B+E; Lane 4: A+B+E without T4 ligase; Lane 5: B; Lane 6: A;

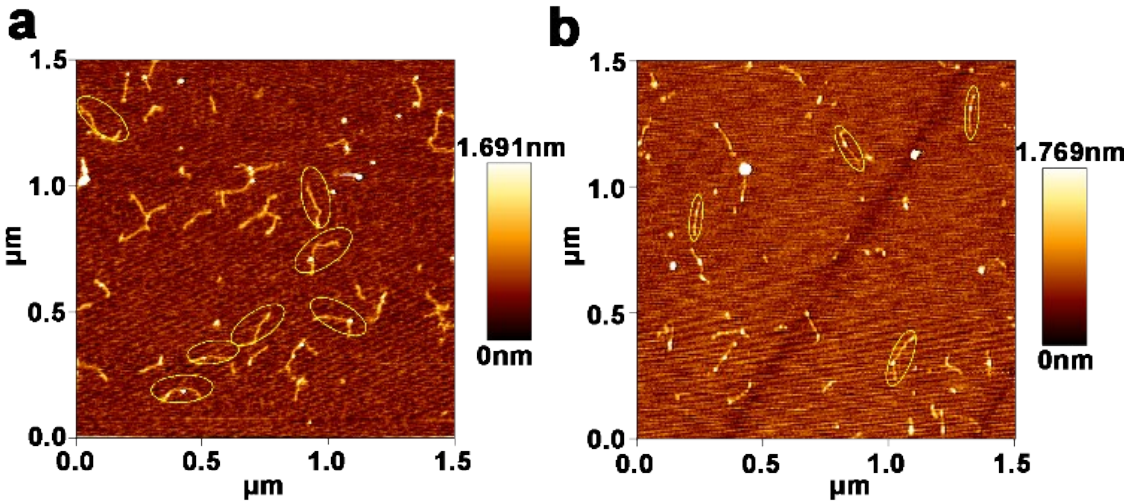

Figure S3. AFM imaging of (a) Monomeric and (b) Dimeric TBA-dsDNA construct. The G4 structures in dsDNA-ssDNA-dsDNA sandwich structures are higher than the DNA handles.

Typical force-extension curves obtained on monomeric, dimeric and polymeric TBAs

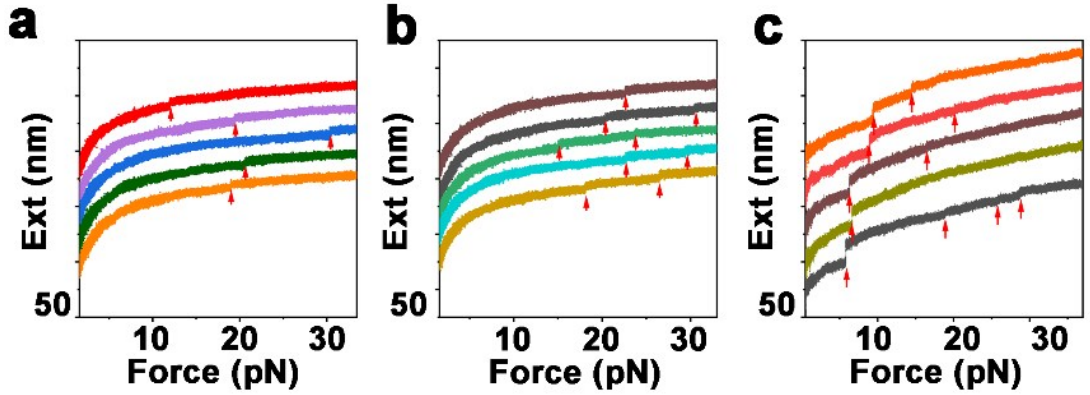

Figure S4. Typical force-extension curves for: (A) monomer, (B) dimer, (C) polymer. In each Figure panel, 5 consecutive force-extension curves recorded in force-increase scan (loading rate = ~0.01 pN/sec) taken from repeating stretch-relax cycles are shown. The curves are shifted along the

extension axis for visual clarity. TBA unfolding is indicated by stepwise extension increases (red arrows).

The calculation of the expected change in contour length during force-induced unzipping.

The expected ΔL during the unfolding of the structure was calculated using the following equation

$$\Delta L = (N \times L) - x \quad (1)$$

where N is the number of nucleotides, L is the contour length per nucleotide in ssDNA (0.56 nm)[6,7], and x=1 is the end-to-end distance of a folded G4 structure as has been used by other researchers[9] for the G4 structure. This calculation gave ΔL expected values of monomer TBA, separated dimer TBA, rod-like TBA, respectively.

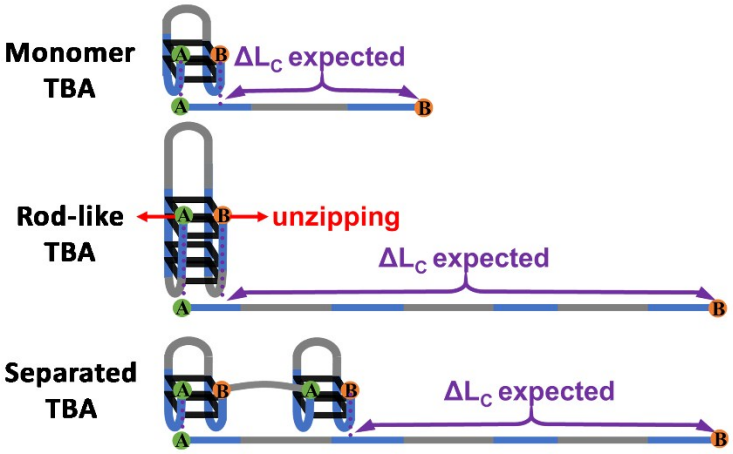

Figure S5. The calculation schematic of the expected change in contour Length.

CD Spectra of TBA monomer, dimer, trimer and polymer

TBA, d(GGTTGGTGTGGTTGG), adopts a chair-type anti-parallel structure in solution. As shown in Figure S6, CD profiles of TBA monomers (black line) show maxima at about 248 and 295 nm and a minimum at 270 nm, which is consistent with an anti-parallel G4 architecture characterized by a positive band near 248 nm, a maximum near 295 nm, and a minimum near 270 nm[10]. In contrast, the CD spectroscopy of a parallel quadruplex presents a maximum near 265 nm[11,12]. For TBA dimers and polymers, the overall shape of the CD spectra is similar, but the magnitude of the peaks at 248 nm and 298 nm decreases gradually. In addition, the maxima and minima gradually move to shorter wavelengths. These results indicate that the multiple repeated TBA aptamers form a structure that is

progressively different from monomeric structures. However, all copy numbers in our system show a maximum centered at a wavelength of 280–295 nm and a minimum around 270 nm, indicating the formation of anti-parallel G-quadruplex structures[11-13]. Almost all CD spectra of the TBA in the 3' or 5' extended based (without antisense sequence) show peaks at wavelengths very close to those of TBA monomers, as previously reported[13]. This means the spacer has little effect on the TBA CD peak wavelengths. The overall decrease in CD peak wavelength with the increase in copy number may be attributed to the change/disruption of standard anti-parallel G-quadruplex TBA structures caused by the rod-like TBA structures.

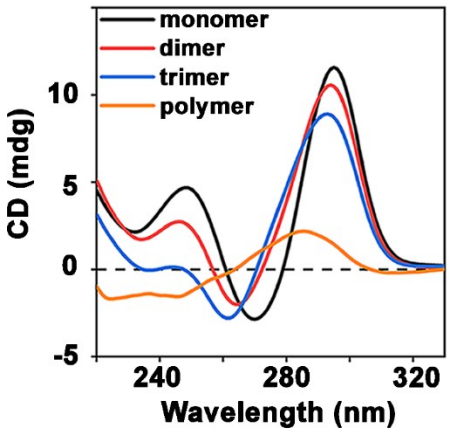

Figure S6. Circular dichroism (CD) spectra of TBA repeat sequences of different copy numbers (monomer, black trace; dimer, red trace; trimer, blue trace; polymer, orange trace).

Agarose gel electrophoresis of two-end modified ssDNA prepared by PC-RCA

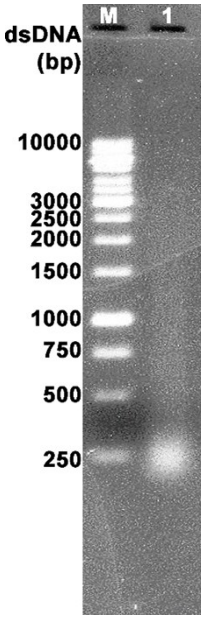

Figure S7. Agarose gel electrophoresis of two-end modified ssDNA synthetized by polymerase change-rolling circle amplification (PC-RCA). lane 1 represents agarose gel electrophoresis of the products using GenElute PCR Clean-Up Kit, and the stripes were diffused from 250 bp to 2000 bp.

Characterization of biotin-labeled ssDNA by QCM-D

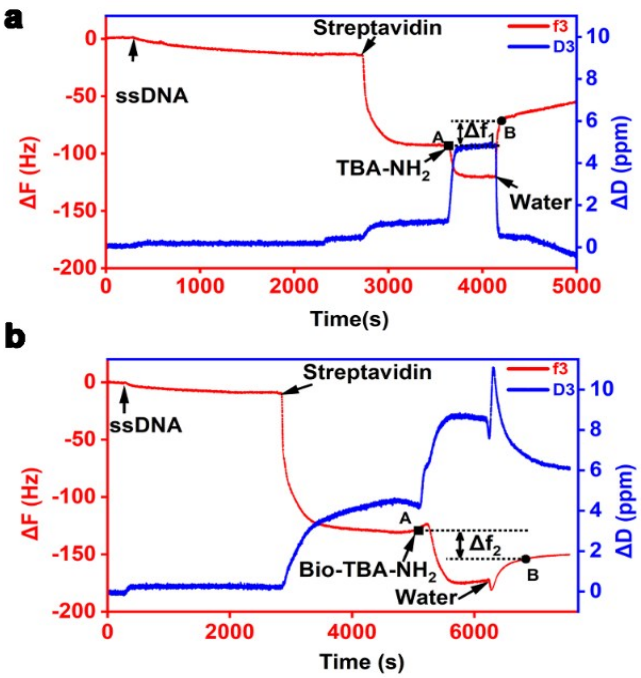

Figure S8. Characterization of preparation of biotin-modified ssDNA by QCM-D. The substrates were blocked by T sequence ssDNA and streptavidin were added until frequency change equilibrium is reached that means streptavidin is well adsorbed on the substrate. Subsequently, (a) NH_2-ssDNA prepared by RCA with phi29 DNA polymerase and (b) NH_2-ssDNA-biotin prepared were introduced. The frequency is named A when NH_2-ssDNA and NH_2-ssDNA-biotin were added. After rinsing with water, the equilibrated frequency is named B. The frequency changes were compared between A and B. The reduction of frequency after the introduction of NH_2-ssDNA-biotin indicates the successful modification of biotin group to the ssDNA.

Characterization of two-end labeled ssDNA by optical microscope

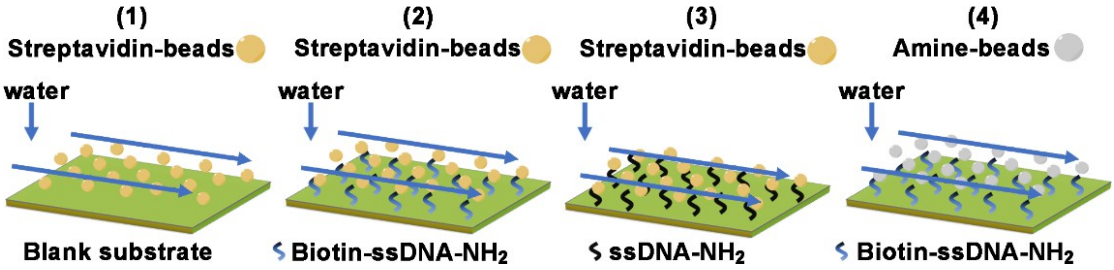

Figure S9. Characterization of the two-end modified ssDNA (PC-RCA product). (1) The streptavidin-beads were added on the amino-silanized silicon substrates, (2) Biotin-ssDNA-NH_2 modified substrates and (3) ssDNA-NH_2 modified substrates in flow chamber. (4) The amine-beads were added on the Biotin-ssDNA-NH_2 modified substrates in flow chamber. After incubation, 100 µL of water was vertically dropped on one side of the flow chamber using a pipettor to remove loosely bond beads. The change of bead number was estimated by CCD camera. The number of fixed beads represents the number of beads successfully tethered to the substrates. The results are summarized in Table S2.

AFM imaging of structures in TBA repeat sequences

We directly observed the rod-like TBA of different lengths along the direction of ssDNA in TBA repeat sequences using AFM imaging. The long intramolecular stacking structures of TBA repeat sequences were directly observed. It can be clearly observed that there are discontinuous bead structures with different lengths along the direction of the ssDNA.

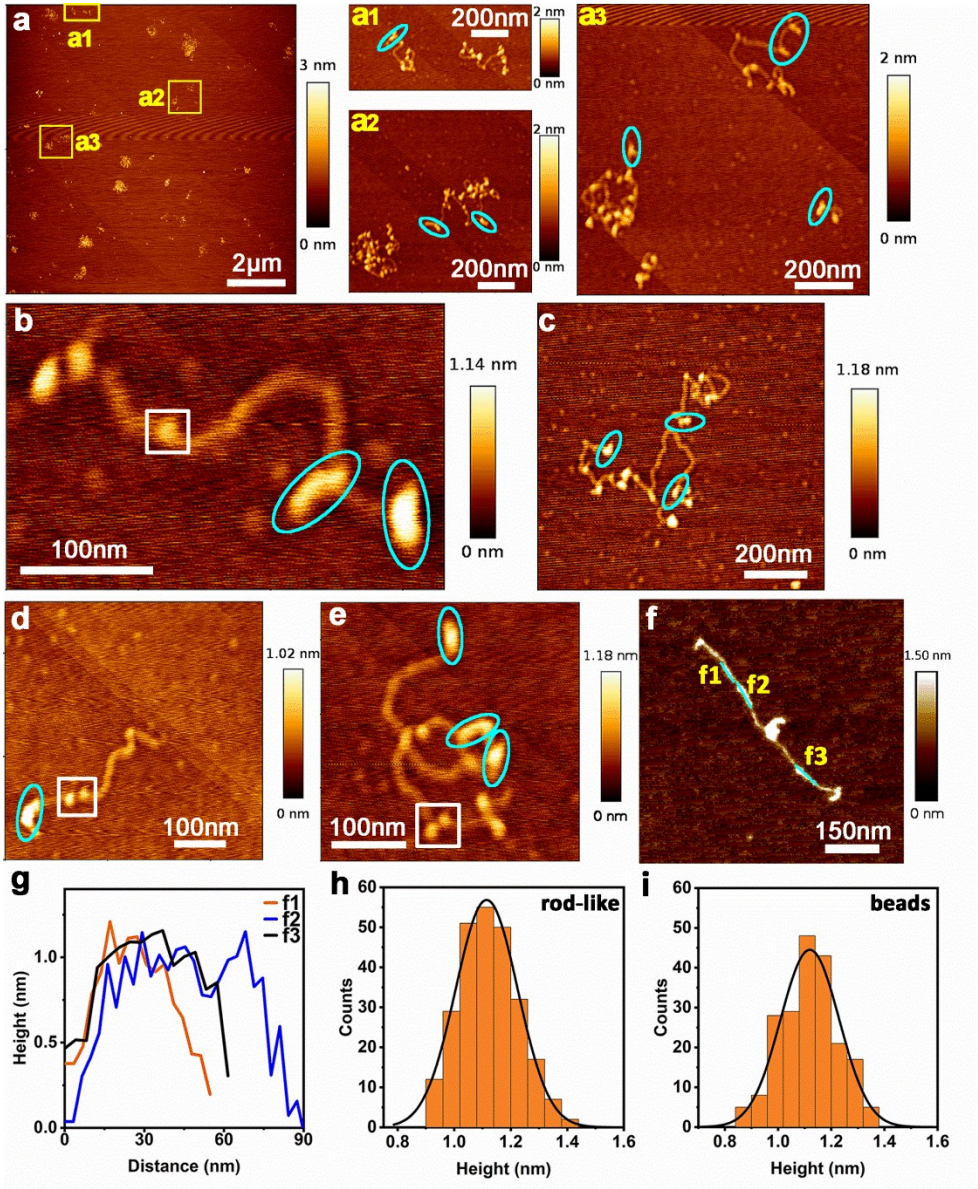

Figure S10. (a-f) AFM images of long-chain ssDNA containing rod-like G4 structures (sky-blue circle) and normal G4 beads (white rectangle in b), d) and e)). a1-a3 correspond to zoom of yellow rectangle in a. (g) The height profile of the sky-blue lines in the f. (h) The height histogram of rod-like G4. The statistic number is 255. The height histogram was fitted by using the Gaussian function to obtain the most probable value at 1.11 ± 0.11 nm. (i) The height histogram of normal G4 beads (marked by white rectangle). The statistic number is 204. The height histogram was fitted by using the Gaussian function to obtain the most probable value at 1.12 ± 0.11 nm.

AFM imaging of poly (thymine) ssDNA

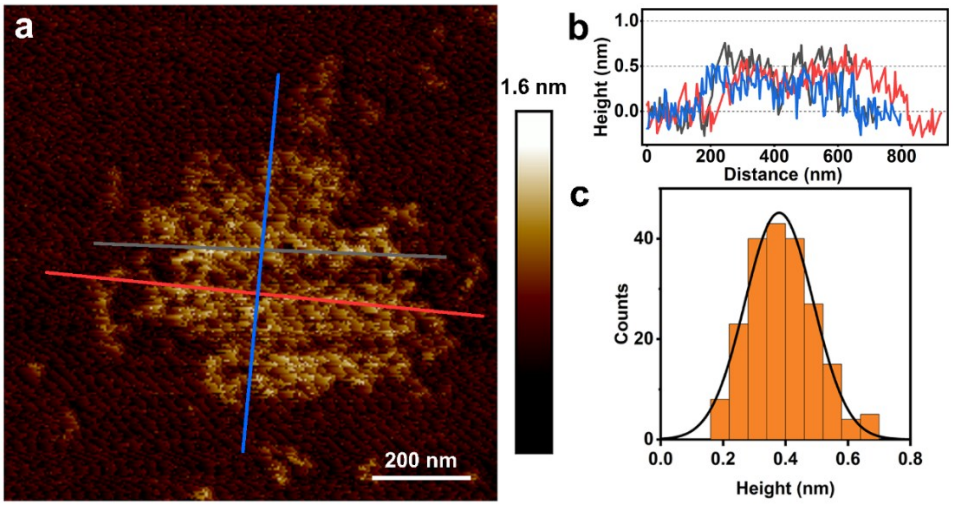

Figure S11. AFM Imaging of poly(thymine) ssDNA. (a) The long-chain thymine sequences exhibit random coil conformation (b) The height profile of poly(T) ssDNA along the lines (red, blue and grey) (c) The height histogram of poly(thymine) ssDNA. The statistic number is 205. The height histogram was fitted by using the gaussian function to obtain the most probable value at 0.38 ± 0.11 nm.

AFM imaging of long-chain repeated TBA-thrombin complexes

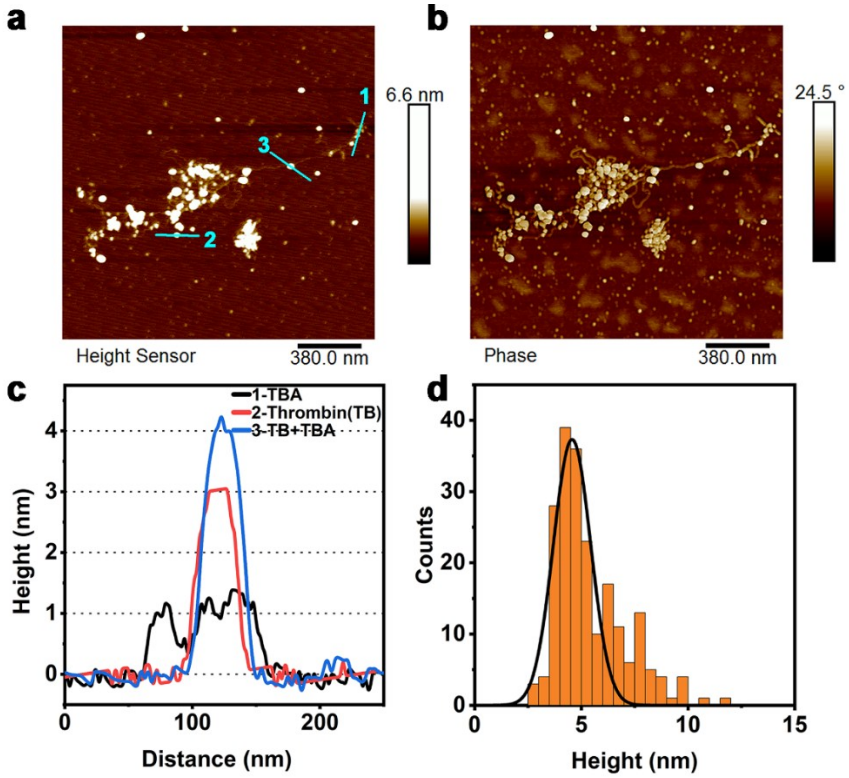

Figure S12. Representative AFM (a) height image and (b) phase image of TBA repeat sequences in complexation with thrombin (TB). (c) the heights profiles in selected areas marked by the sky-blue line in the Figure S13a. The structures with a height of ~1.16 nm correspond to G4 structures (The statistic height is 1.11 ± 0.11 and statistic number is 255.) as marked by the number 1 sky-blue line. The structures with a height of ~3.06 nm correspond to thrombin as marked by the number 2 sky-blue line. The structures with a height of ~4.34 nm can be ascribed to thrombin binding to TBA in the sitting-on mode as marked by the number 3 sky-blue line. (d) The height histogram of TBA-thrombin complexes. The statistic number is 206. The height histogram was fitted by using the Gaussian function to obtain the most probable value at 4.56 ± 0.87 nm.

AFM imaging of thrombin

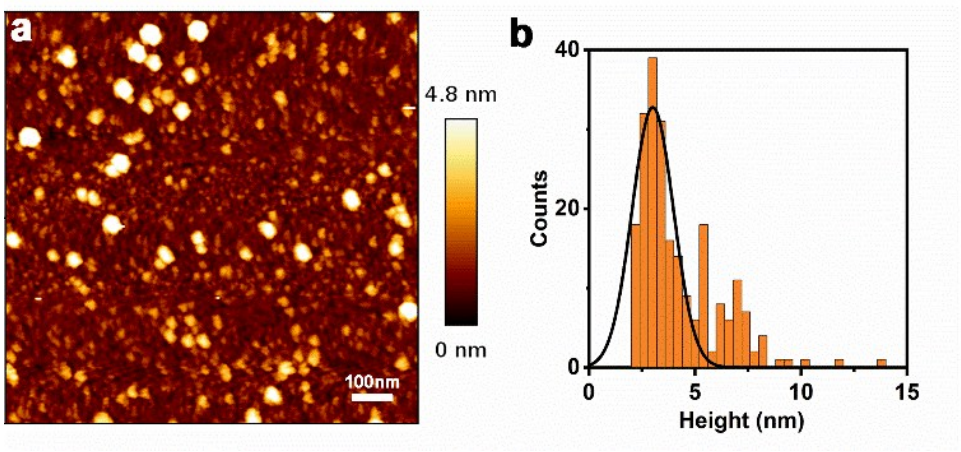

Figure S13. (a) AFM images of thrombin. (b) The height histogram of thrombin. The statistics about the height of thrombin is 3.01 ± 0.98 nm and the statistic number is 229.

Table S1. The contour length change during the unfolding of monomeric and dimeric TBA. The expected and measured contour length change are listed in the presence and absence of thrombin. Uncertainties represent the standard error. The measured contour length change agrees well with the calculated/expected contour length change for unfolding a rod-like dimer, which includes two copy numbers of TBA and a spacer), as shown in Table S1. This result can prove the existence of rod-like TBA structure.

Transition	Monomer (15 nt)	Rod-like Dimer (50 nt)
ΔLc expected	7.4 nm	27.0 nm
ΔLc measured (no TB)	7.7 ± 1.3 nm	23.3 ± 5.8 nm
ΔLc measured (with TB)	8.2 ± 1.4 nm	24.6 ± 5.8 nm

Table S2. The statistical table of number of beads while rinsing.

NO.	Substrates	Beads	Number of beads while rinsing
(1)	Blank	Streptavidin-beads	236 to 81 (remaining 34.3%)
(2)	Biotin-ssDNA-NH_2	Streptavidin-beads	277 to 272 (remaining 98.2%)
(3)	ssDNA-NH_2	Streptavidin-beads	201 to 20 (remaining 10%)
(4)	Biotin-ssDNA-NH_2	NH_2-beads	264 to 50 (remaining 18.9%)

Table S3. The contour length change during the unfolding of long-chain TBA with two different spacer lengths (N20 represent TBA repeat sequence with 20-nt spacer and N30 represent TBA repeat sequence with 30-nt spacer). The probability for forming rod-like TBAs that contains different copy numbers of the aptamer are listed.

Number of TBA in structures	N20 ΔLc Expected (nm)	N20 Probability (%)	N20+TB Probability (%)
1	7.4	88.4	64.2
2	27	8.1	22.1
3	46.6	1.5	5.3
4	66.2	1.0	1.1
5	85.8	1.0	2.1
6	105.4	0.0	1.1
7	125	0.0	3.2
8	144.6	0.0	0.0
9	164.2	0.0	0.0
10	183.8	0.0	1.1

Number of TBA	N30 ΔLc Expected	N30 Probability	N30+TB Probability

in structures	(nm)	(%)	(%)
1	7.4	88.6	79.2
2	32.6	7.3	9.2
3	57.8	2.4	7.7
4	83	0.0	0.0
5	108.2	1.6	0.8
6	133.4	0.0	0.8
7	158.6	0.0	0.8
8	183.8	0.0	0.8
9	209	0.0	0.0
10	234.2	0.0	0.8

Table S4. The unfolding rate constants ko u of TBA were obtained from analysis of unfolding force distribution in the presence and absence of thrombin. The I, Ⅱ and Ⅲ represent three forms in the different copy numbers.

ko u	monomer+TB	dimer	dimer+TB	N20	N20+TB	N30	N30+TB
I (10^{-3} s^{-1})		0.9±0.1	1.6±0.1	0.9±0.3	1.0±0.3	2.0±0.1	1.9±0.5
Ⅱ (10^{-3} s^{-1})	1.0±0.3	1.2±0.1	1.4±0.1	1.0±0.4	1.0±0.3	0.9±0.2	0.7±0.3
Ⅲ (10^{-3} s^{-1})	0.9±0.2	0.8±0.01	0.7±0.1	0.8±0.09	0.7±0.1	0.5±0.1	0.7±0.3

References

1. C. G. Baumann, S. B. Smith, V. A. Bloomfield and C. Bustamante, *Proc. Natl. Acad. Sci. USA.* 1997, **94**, 6185-6190.

2. Z. Yu and H. Mao, *Chem. Rec.* 2013, **13**, 102-116.

3. M. D. Wang, H. Yin, R. Landick, J. Gelles and S. M. Block, *Biophys. J.* 1997, **72**, 1335-1346.

4. S. Haldar, R. Tapia-Rojo, E. C. Eckels, J. Valle-Orero and J. M. Fernandez, *Nat. Commun.* 2017, **8**, 1-8.

5. Y. Seol, G. M. Skinner and K. Visscher, *Phys. Rev. Lett.* 2004, **93**, 118102:1-4.

6. W. Saenger, Principles of Nucleic Acid Structure, Springer, New York, 1984.

7. H. You, J. Wu, F. Shao and J. Yan, *J. Am. Chem. Soc.* 2015, **137**, 2424-2427.

8. J. W. Shim, Q. Tan and L. Q. Gu, *Nucleic Acids Res.* 2009, **37**, 972-982.

9. S. Selvam, Z. Yu and H. Mao, *Nucleic Acids Res.*, 2015, **44**, 45-55.

10. D. Zhao, X. Dong, N. Jiang, D. Zhang and C. Liu, *Nucleic Acids Res.* 2014, **42**, 11612-11621.

11. H. Saneyoshi, S. Mazzini, A. Avino, G. Portella, C. Gonzalez, M. Orozco, V. E. Marquez and R. Eritja, *Nucleic Acids Res.* 2009, **37**, 5589-5601.

12. I. N. Rujan, J. C. Meleney and P. H. Bolton, *Nucleic Acids Res.* 2005, **33**, 2022-2031.

13. M. C. Buff, F. Schäfer, B. Wulffen, J. Müller, B. Pötzsch, A. Heckel and G. Mayer, *Nucleic Acids Res.* 2010, **38**, 2111-2118.