

# SCIENTIFIC REPORTS



OPEN

## Triplex-quadruplex structural scaffold: a new binding structure of aptamer

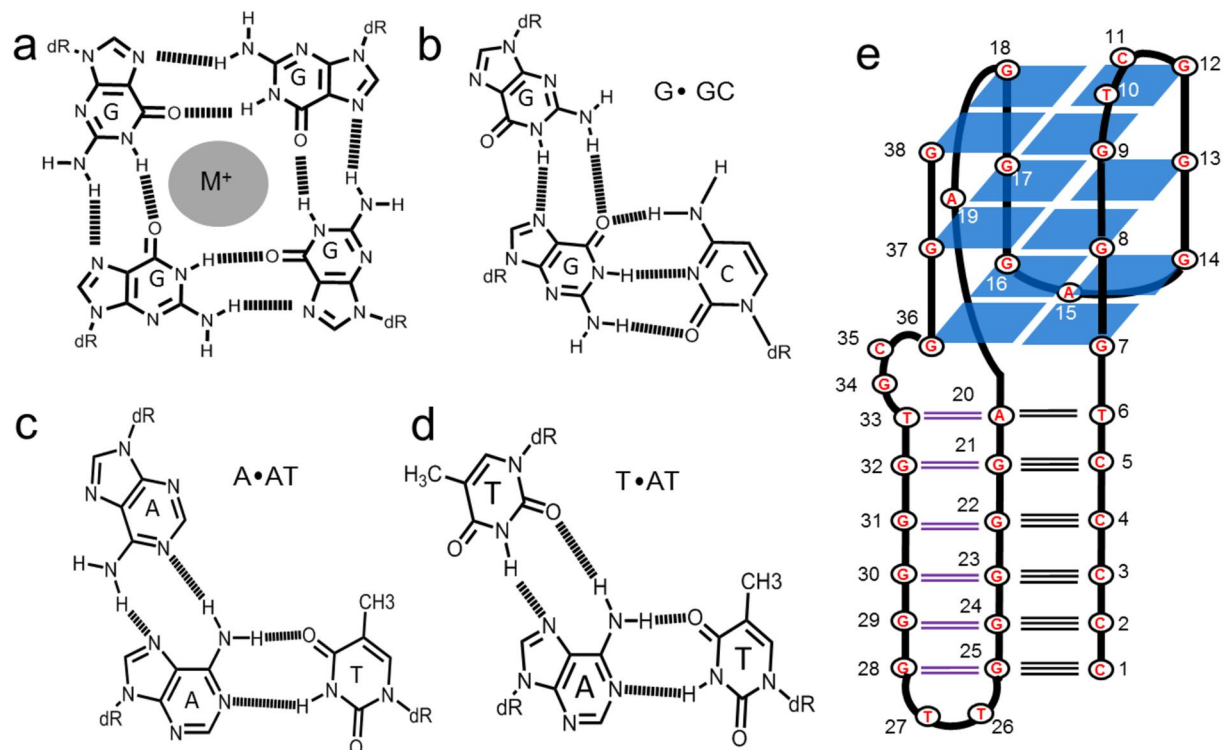
Tao Bing<sup>1,2</sup>, Wei Zheng<sup>1</sup>, Xin Zhang<sup>1,2</sup>, Luyao Shen<sup>1,2</sup>, Xiangjun Liu<sup>1,2</sup>, Fuyi Wang<sup>1,2</sup>, Jie Cui<sup>1</sup>, Zehui Cao<sup>1</sup> & Dihua Shangguan<sup>1,2</sup>

Apart from the canonical Watson-Crick duplex, nucleic acids can often form other structures, e.g. G-quadruplex and triplex. These structures give nucleic acid additional functions besides coding for genetic information. Aptamers are one type of functional nucleic acids that bind to specific targets with high selectivity and affinity by folding into special tertiary structures. Despite the fact that numerous aptamers have been reported, only a few different types of aptamer structures are identified. Here we report a novel triplex-quadruplex hybrid scaffold formed by a codeine binding aptamer (CBA). CBA and its derivatives are G-rich DNA sequences. Codeine binding can induce the formation of a complex structure for this aptamer containing a G-quadruplex and a G-GC triplex, while codeine is located at the junction of the triplex and quadruplex. When split CBA into two moieties, codeine does not bind either moieties individually, but can bind them together by inducing the formation of the triplex-quadruplex scaffold. This structure formation induced by codeine binding is shown to inhibit polymerase reaction, which shows a potential application of the aptamer sequence in gene regulations.

Nucleic acid sequences can form many non-canonical structures other than the Watson-Crick duplex structure (e.g. pseudoknot, quadruplex, triplex and i-motif) through base-pairing interactions, such as Watson-Crick base pairing, wobble base pairing and Hoogsteen base pairing. These structures are found in various functional nucleic acids, such as transfer RNA, riboswitches, DNazymes, ribozymes, aptamers, and many nanodevices<sup>1-5</sup>. Among these structures, G-quadruplexes are four-stranded structures formed by repetitive G-rich nucleic acid sequences. They have attracted considerable interest because of their important biological functions and applications in biosensors and nanotechnology<sup>6</sup>. G-quadruplexes are composed of planar arrangements of four guanine-bases stabilized by eight Hoogsteen hydrogen bonds (known as G-quartet) (Fig. 1a)<sup>7</sup>. These structures are widely considered to play important roles in regulation of many cellular functions, such as telomere maintenance<sup>8</sup>, transcription<sup>9,10</sup> and translation<sup>11,12</sup>. Triplexes are triple helical structures, in which the third strand binds to a homopurine strand of the Watson-Crick duplex within the major groove<sup>13</sup>. There are two different triplex motifs, purine motif (R) and pyrimidine motif (Y). In the R motif, the third strand binds to the duplex purine strand in an antiparallel orientation via reverse-Hoogsteen hydrogen bonds and forms A·AT, T·AT, and G·GC base-triplets (Fig. 1b-d)<sup>14,15</sup>. In the Y motif the third strand binds to the duplex purine strand in a parallel orientation via Hoogsteen hydrogen bonds and forms T·AT and C·GC base-triplets<sup>14</sup>. Potential triplex-forming oligonucleotides are found throughout genomic DNA<sup>14,15</sup> and RNA<sup>16,17</sup>, and are also considered to play important roles in gene transcription and expression<sup>18</sup>.

Aptamers are single strand oligonucleotides (DNA or RNA), generated *in vitro* using the SELEX technique. They can bind to various targets with high affinity and specificity, with potential applications in affinity separation, biosensors, drug discovery, and diagnosis<sup>1,19-21</sup>. The specific binding of aptamers results from their well-defined tertiary structures composed of simple secondary structures, such as hairpins, bulges, pseudoknots, T-junctions, G-quadruplexes, and their combination<sup>22,23</sup>. These complex tertiary structures create well-defined binding shapes, facilitating the formation of electrostatic interactions, hydrogen bonds and  $\pi$ - $\pi$  stacking interactions between aptamers and their targets. The structural elucidation of aptamers can provide insights into the principles, patterns and diversity associated with nucleic acid architecture, molecular recognition and the adaptive binding<sup>24</sup>.

<sup>1</sup>Beijing National Laboratory for Molecular Sciences, Key Laboratory of Analytical Chemistry for Living Biosystems, CAS Research/Education Center for Excellence in Molecular Sciences, Institute of Chemistry, Chinese Academy of Sciences, Beijing, 100190, China. <sup>2</sup>University of Chinese Academy of Sciences, Beijing, 100049, China. Correspondence and requests for materials should be addressed to D.S. (email: [sgdh@iccas.ac.cn](mailto:sgdh@iccas.ac.cn))



**Figure 1.** (a) Structure of G-quartet; (b) Structure of G-GC triad; (c) Structure of A-AT triad; (d) Structure of T-AT triad; (e) The schematic depiction of predicted scaffold 1 of CBA-1.

Apart from stems, loops, and bulges, G-quadruplexes are often found in aptamer, and act as the binding motif of many aptamers<sup>25–30</sup>. In contrast, triplexes have been rarely reported in aptamers.

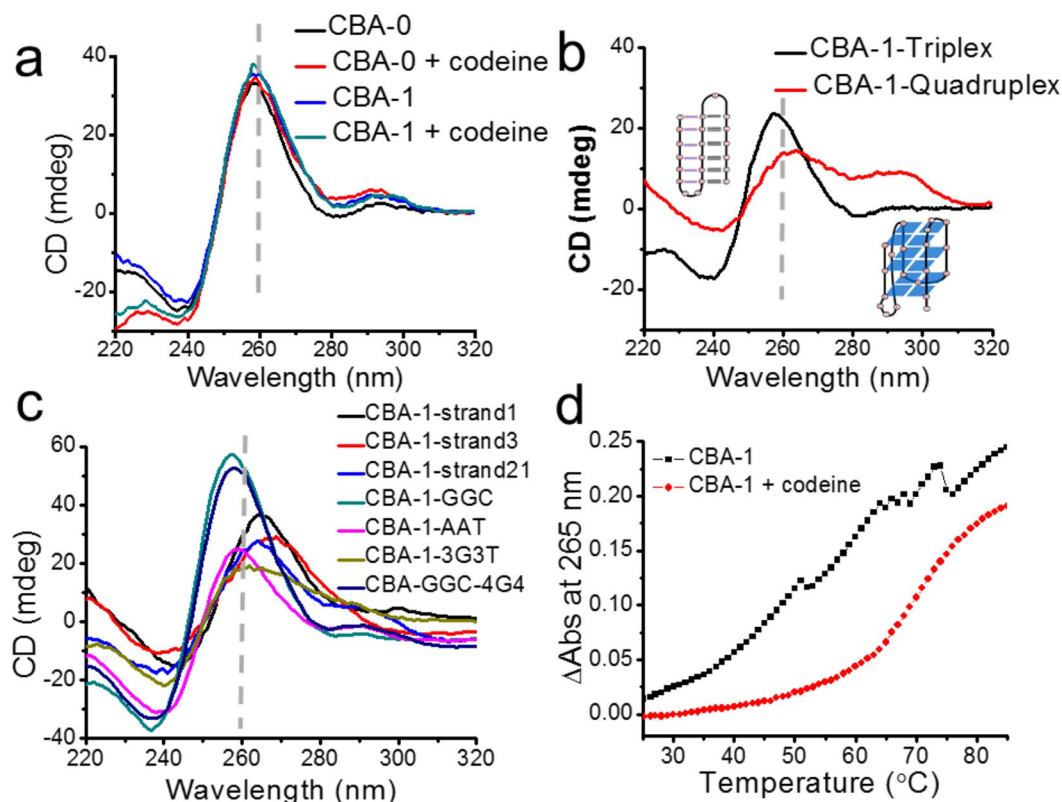
Herein we report a novel aptamer structure composed of a G-quadruplex and a triplex. The binding features of this codeine binding aptamer (CBA) to codeine<sup>31</sup> were investigated. We also show that the formation of the triplex-quadruplex scaffold induced by codeine can lead to inhibition of DNA polymerase activity.

## Results and Discussion

**The potential structure of codeine binding aptamer.** The original CBA (CBA-0) is a highly G-rich sequence generated in our lab previously, which contains six guanine (G) tracts and a cytosine (C) tract (5C-T-3G-TC-3G-A-3G-AA-5G-TT-5G-TGC-2G) (Table S1). G-rich sequences often fold into G-quadruplexes. The formation of a G-quadruplex motif only needs four G-tracts, the extra G-tracts in a G-quadruplex-forming sequence act as loops<sup>32,33</sup>. Usually, a sequence containing more than four G-tracts may form multiple G-quadruplexes in solution. In the case of CBA-0, there are a 5C-tract and two 5G-tracts, which have the potential to fold into a perfect G-GC triplex structure. In addition, previously, we have found that removal of the 2G-tract at 3'-end of CBA-0 causes the loss of its binding ability to codeine<sup>31</sup>, suggesting that the 2G-tract may be involved in the formation of a G-quadruplex together with the other three 3G-tracts. As a result, CBA-0 may adopt a combination of triplex and G-quadruplex. Based on this hypothesis, the G-quadruplex moiety of CBA-0 likely forms an imperfect G-quadruplex structure because of the three 3G-tracts and one 2G-tract. A new sequence, CBA-1, was therefore designed by adding a G-base on the 3'-end of CBA-0 so that it can form a perfect G-quadruplex with four 3G-tracts. Because our previously study found that CBA-0 tightly bound codeine in buffers with high concentration of Na<sup>+</sup> (> 50 mM) and low concentration of K<sup>+</sup> (≤ 50 mM)<sup>31</sup>, unless otherwise indicated, all the experiment in this paper were performed in phosphate buffer saline (PBS, 140 mM NaCl + 2.5 mM KCl + 1.6 mM KH<sub>2</sub>PO<sub>4</sub> + 15 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.4) with 2.5 mM MgCl<sub>2</sub>. CBA-1 was measured to have a much higher binding affinity ( $K_d = 0.07 \pm 0.01 \mu\text{M}$ ) to codeine than that of CBA-0 ( $K_d = 0.25 \pm 0.08 \mu\text{M}$ ) by isothermal titration calorimetry (ITC) (Figure S1), which suggests that a stable G-quadruplex moiety is critical for codeine binding.

Because the orientation of the three strands in G-GC type triplex is fixed, the G<sub>21–25</sub> tract of CBA-1 inclines to form a Watson-Crick duplex with the 5C tract, and the G<sub>28–32</sub> tract acts as the third strand in an antiparallel orientation so that G<sub>36–38</sub> and G<sub>7–9</sub> can locate on the same side for the formation of G-quadruplex. G-quadruplexes can be classified into parallel, anti-parallel and hybrid structures based on the arrangement and orientation of the strands, therefore, CBA-1 may adopt five possible structures with hybrid G-quadruplex (scaffold 1 and 2), parallel G-quadruplex (scaffold 3) or antiparallel G-quadruplex (scaffold 4 and 5) (Figs 1e and S2).

**Circular dichroism spectra and binding affinity.** Circular dichroism (CD) spectra are commonly used to provide a signature profile of the secondary structure of a given DNA. The intramolecular triplex DNA displays a negative peak around 240 nm, a positive peak around 257 nm and a second negative peak around 280 nm in

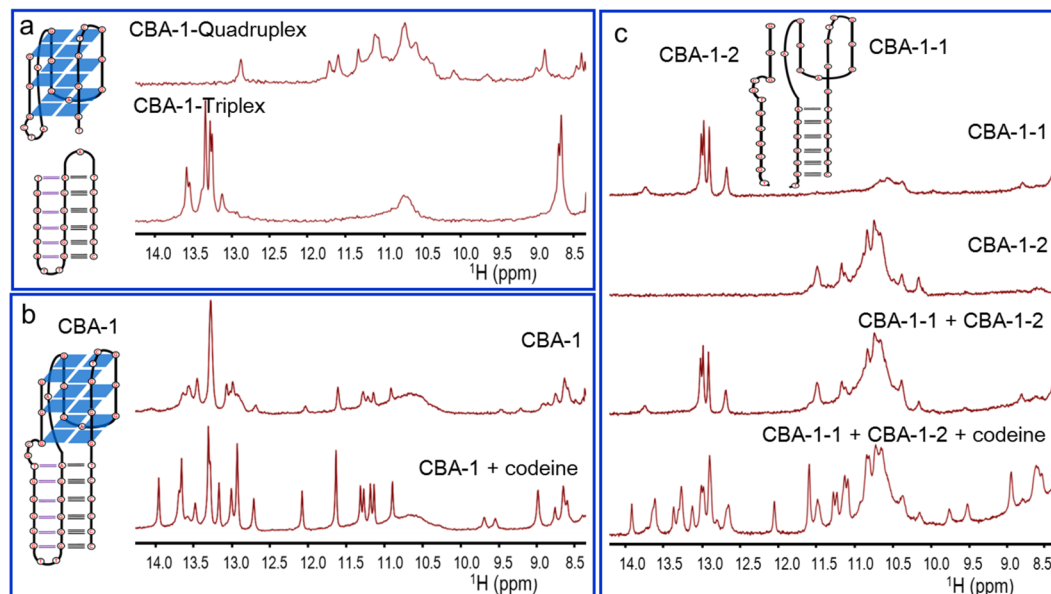


**Figure 2.** (a) CD spectra of CBA-0 and CBA-1 (4 μM) in the absence or presence of codeine (4 μM). (b) CD spectra of CBA-1-triplex and CBA-1-quadruplex (4 μM). (c) CD spectra of derivative sequences of CBA-1 (4 μM). (d) The melting curve of CBA-1 (4 μM) in the absence or presence of codeine (8 μM) (absorbance at 265 nm).

CD spectra<sup>15</sup>. For G-quadruplexes, the antiparallel structure has a positive CD peak around 295 nm and a negative peak around 260 nm; the parallel structure has a positive peak around 264 nm and a negative peak around 240 nm; and hybrid structures show positive signals around 295 and 265 nm and a negative peak around 240 nm<sup>7</sup>. The CD spectra of CBA-0 and CBA-1 (Fig. 2a) showed a positive main peak near 260 nm, and a small positive peak near 290 nm (combined signal of triplex and G-quadruplex); the addition of codeine caused minor change of their CD spectra. Due to the spectral overlap of triplex and parallel/hybrid G-quadruplexes in CD spectra, structural information of triplex and quadruplex was difficult to be elucidated from the CD spectra of the CBAs. To solve this problem, CBA-1 was split into two moieties, G-quadruplex and triplex (Table S1), and their CD spectra were collected. As shown in Fig. 2b, the CBA-1-triplex showed the typical CD signals of a triplex. And CBA-1-quadruplex showed the CD signals of a hybrid structure (3 + 1) of G-quadruplex that contained three parallel and one antiparallel strands (a positive main peak at 265 nm, a shoulder peak near 290 nm and a negative at 240 nm)<sup>34</sup>.

In order to demonstrate whether the triplex and G-quadruplex structures were essential for the binding of CBA, mutation assay was performed. The triplex moiety of CBA-1 was mutated by replacing the 5C-tract with an 5A-tract (CBA-1-strand1), replacing a 5G-tract ( $G_{28-32}$ , the third strand of the triplex) with TATAT (CBA-1-strand3), or interchanging the 5C-tract and a 5G-tract ( $G_{21-25}$ ) (CBA-1-strand21). The CD spectra of these sequences showed that their main CD peaks shifted from 257 nm to around 265 nm, while the negative CD peak at 280 nm disappeared, suggesting the loss of triplex structure (Fig. 2c). These mutations also caused the sequences to lose binding ability to codeine (Figure S3). As expected, extending the triplex of CBA-1 with three G-GC triads (CBA-1-GGC) or three A-AT triads (CBA-1-AAT) did not cause any CD signal shift (Fig. 2c), and kept the binding to codeine ( $K_d = 0.05 \pm 0.02$  and  $0.07 \pm 0.03 \mu\text{M}$ , respectively) (Figure S4). These results suggest that the triplex moiety is essential for the target binding of CBA.

When replacing all the 3G-tracts of CBA-1 with 3T-tracts, the resulted sequence CBA-1-3G3T lost binding ability to codeine (Figure S3d), and showed a broad positive CD signal with the maximum at 257 nm (combined signal of triplex and single strand) (Fig. 2c). In addition, extending the G-quadruplex moiety of CBA-1-GGC to form a four-layer quadruplex (CBA-1-GGC-4G4) only caused a very small shift of the main positive CD signal but led to reduced binding ability ( $K_d = 0.30 \pm 0.17 \mu\text{M}$ ) (Figs 2c and S4c), suggesting that the added quartet did not destroy the conformation of the aptamer but might have slightly affected the binding pocket. These additional results confirmed that G-quadruplex plays an important role for codeine binding. Since no significant peak at 295 nm was found in CD spectra of all the CBA sequences, their G-quadruplex moiety is not likely to adopt an antiparallel structure, i.e. scaffold 4 and 5 may not be the structure of CBA-1.

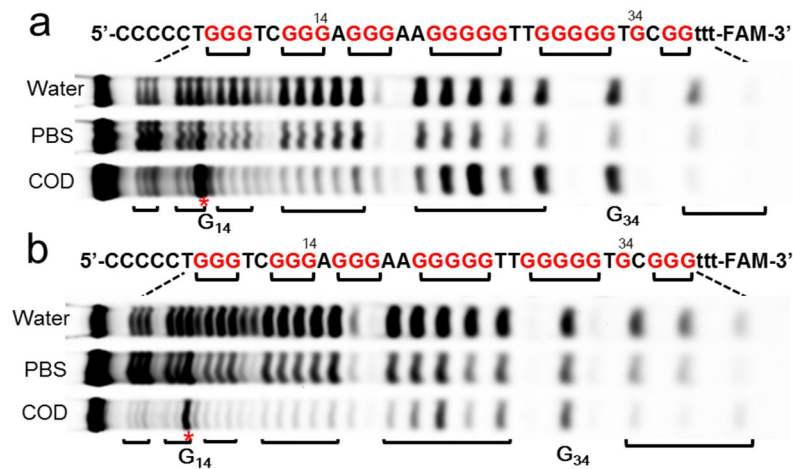


**Figure 3.** DNA sequences related to CBA-1 and their 1D  $^1\text{H}$  NMR spectra in the range of 8.4–14.5 ppm.

The influence of codeine on the structural stability of CBA-1 was investigated by thermal denaturation. The melting curve of CBA-1 showed several transitions in the absence of codeine, but only one transition in the presence of codeine and with a significant increase of melting temperature (Fig. 2d). This finding suggests that codeine induced the formation of a unique and more stable structure.

**G-quadruplex probe competition assay.** In order to further confirm the G-quadruplex in CBA-1, a fluorescent G-quadruplex probe, BPBC, that binds to G-quadruplexes through end-stacking on G-quartet surface was used to compete with codeine<sup>35</sup>. BPBC emit a very weak fluorescence in the absence of G-quadruplex. After mixing with CBA-1 or CBA-1-strand1 (non-binding sequence without the predicted triplex moiety), the fluorescence of BPBC increased approximately 200 fold (DNA:BPBC = 1:1) and 400 fold (DNA:BPBC = 1:3) (Figure S5), indicating the presence of G-quadruplex in both sequences. The addition of equivalent amount of codeine and BPBC to CBA-1 (DNA:BPBC:codeine = 1:3:3) resulted in a fluorescence intensity about half of that from the mixture of DNA/BPBC at 1:3 ratio. Interestingly, further addition of codeine (DNA:BPBC:codeine = 1:3:10) did not continue to cut down BPBC fluorescence. One possible explanation is that codeine only bound at one end of G-quadruplex moiety of CBA-1 while BPBC could bind at both ends through end-stacking. Compared to CBA-1, the addition of codeine did not affect the fluorescence of CBA-1-strand1/BPBC solution (DNA:BPBC:codeine = 1:3:3), indicating codeine binding did not happen without the triplex structure. These results suggest that codeine most likely binds to only one end of the G-quadruplex that is adjacent to the triplex structure, which agrees with the ITC results that CBA only binds one codeine.

**Structure study by  $^1\text{H}$ -NMR.** In order to further confirm the triplex-quadruplex scaffold of CBA,  $^1\text{H}$ -NMR spectra were collected. Guanines in the tetrads of G-quadruplexes exhibit characteristic chemical shifts around 10–12 ppm, which are assigned to the imino protons in Hoogsteen hydrogen bonding<sup>36–38</sup>. In triplex, the hydrogen bonded imino protons of thymine and guanine exhibit characteristic chemical shifts in the range of 12.5–14.5 ppm in addition to those expected from the Watson-Crick duplex, and the cytosine amino protons display a set of downfield shifted resonances around 9 ppm<sup>15,38,39</sup>. The  $^1\text{H}$ -NMR spectra of the split triplex and quadruplex moieties of CBA-1 (CBA-1-Quadruplex and CBA-1-Triplex) showed a group of signals in the range of 10–12 ppm and 13–14 ppm respectively (Fig. 3a), which correspond to the imino proton signals of quadruplex and triplex. CBA-1-Triplex also showed strong  $^1\text{H}$ -NMR signals around 8.7 ppm corresponding to cytosine amino protons in triplex. In the absence of codeine, CBA-1 displayed similar  $^1\text{H}$ -NMR signals with low resolution in the range of 13–14 ppm and 8.6–8.8 (Fig. 3b), suggesting the presence of triplex. It also showed broad envelopes and some small peaks in the range of 10–12 ppm, which suggests that CBA-1 might form multiple quadruplex species. In contrast, in the presence of codeine, CBA-1 showed well-resolved NMR signals of triplex and quadruplex, suggesting that codeine caused CBA-1 to form a well-defined and more stable unique triplex-quadruplex hybrid scaffold. Similar results were also obtained from other codeine binding sequences (CBA-0, CBA-1-GGC and CBA-1-AAT) (Figures S6 and S7). The non-binding sequences (CBA-1-strand1, CBA-1-strand3, CBA-1-strand21 and CBA-1-3G3T) did not show well-resolved NMR signals of triplex or quadruplex whether in the presence or absence of codeine (Figure S8). These results further confirm that a well-defined unique triplex-quadruplex scaffold is essential for the aptamer to bind codeine. In addition, compared to CBA-1-Quadruplex, the  $^1\text{H}$ -NMR peaks of all codeine-binding sequences in the range of 10–10.8 disappeared, and only 8 imino well-dissolved peaks of G-quadruplex were observed in the presence of codeine, suggesting that the quadruplex moiety of these



**Figure 4.** Cleavage fragments of CBA-0 (a) and CBA-1 (b) resolved by denaturing gel electrophoresis; Lane water: DNA cleaved in water; Lane PBS: DNA cleaved in binding buffer; Lane COD: DNA cleaved in binding buffer containing codeine. The full-length gel is presented in Supplementary Figure S19.

sequences only contains two layers of G-tetrads, i.e. the binding of codeine might preclude the formation of G-quadruplex with three layers of G-tetrads.

Since high concentration of  $K^+$  was found unfavourable to the binding of CBA and codeine<sup>31</sup>, we further investigated the effect of 500 mM  $K^+$  on the structure of CBA-1 by  $^1H$ -NMR spectra. In the buffer only containing 500 mM  $K^+$  (without  $Na^+$ ), CBA-1 showed broad NMR signals of G-quadruplex (10–12 ppm), but did not show any signals of triplex in the range of 13–14 ppm whether in the absence or presence of codeine (Figure S9). In the buffer without  $K^+$  (only  $Na^+$ ), the  $^1H$ -NMR spectra of CBA-1 were very similar with those in PBS buffer (containing 4.1 mM  $K^+$ ) in the absence or presence of codeine. After adding 500 mM KCl in the buffer containing  $Na^+$  and codeine, the well-resolved signals of triplex (13–14 ppm) disappeared and the signals of G-quadruplex became broad (Figure S10). These results suggest that high concentration of  $K^+$  led CBA-1 to form multiple G-quadruplexes and disturbed the formation of G-GC triplex structure.

G-rich sequences can simultaneously fold into many quadruplex species in solution, and the populations of the various structures are present in a dynamic equilibrium under any given conditions<sup>40</sup>. In the absence of codeine, CBA sequences formed multiple intermolecular and intramolecular structures, which could have resulted in the low-resolution NMR signals and the multiple transitions of the melting curve (Fig. 2d). The binding of codeine stabilized one of the structures (binding structure), and moved the equilibrium toward the formation of a unique binding structure, resulting in the well-defined NMR signals, as well as the single transition of the melting curve and a much higher  $T_m$  value.

Based on the predicted structures (Figs 1e and S2), the 12-nt long sequence at the 3'-end of CBA-1 caught our attention. This part contains the third strand of triplex and the fourth strand of G-quadruplex. The melting of triplexes has shown that the third strand dissociates at a much lower temperature than that of the dissociation of the core duplex<sup>41–43</sup>. The G-triplex structure, a stable folding intermediate of the G-quadruplex formed by G:G:G triad planes has also been proved<sup>44</sup>. Therefore, it is possible that the 3' end of CBA sequences may open and close in the absence of codeine. The assumption was further proved by DNA-sequence splitting. CBA-1 was split into a 12-nt sequence (CBA-1-2) and a 26-nt sequence (CBA-1-1), and their binding of codeine was measured with codeine-modified beads (Figure S11). The split sequences did not bind codeine separately, but bound it together. Similar result was also obtained in split CBA-1-GGC. The melting assay showed that codeine greatly increased the  $T_m$  value ( $\Delta T_m \approx 4^\circ C$ ) of the mixture of CBA-1-1/CBA-1-2, indicating the formation of a stable complex (Figure S12). In addition, the  $^1H$ -NMR spectrum of CBA-1-1 showed strong signals around 13 ppm (Fig. 3c), which were expected from the Watson-Crick duplex. And CBA-1-2 showed a group of  $^1H$ -NMR signals in the range of 10–12 ppm, which might be due to the formation of intermolecular quadruplexes. In the absence of codeine, the  $^1H$ -NMR spectrum of the CBA-1-1/CBA-1-2 mixture was simply the assembly of both individual spectra, implying no new structure was formed. Interestingly, after addition of codeine, well-resolved NMR signals similar to those of CBA-1/codeine complex appeared in the  $^1H$ -NMR spectrum of CBA-1-1/CBA-1-2 mixture (Fig. 3c), suggesting that codeine brought both sequences together to form the triplex-quadruplex scaffold as with CBA-1.

**Dimethyl sulfate (DMS) footprinting assay.** DMS footprinting assay can probe the accessibility of N7 of guanine in G-quadruplex and triplex structure<sup>45,46</sup>. N7 of guanine in G-quartet and N7 of the second guanine of G-GC triplet are involved in hydrogen bonding, which were generally evenly protected from methylation. In order to further illustrate the structure of CBA sequences, DMS experiments were performed. As shown in Fig. 4, CBA sequences did not form stable triplex-quadruplex structure in water, and almost all G bases in CBA-0 and CBA-1 were not protected as a result. Compared with in water, the  $G_{12-13}$ ,  $G_{16-18}$ ,  $G_{21-25}$ ,  $G_{28-32}$ ,  $G_{34}$  and  $G_{36}$  of CBA-0 were partly protected from methylation in PBS buffer, suggesting these G-bases were involved in the formation of triplex or G-quadruplex. Upon binding with codeine,  $G_{7-9}$ ,  $G_{12-13}$ ,  $G_{16-18}$ ,  $G_{21-25}$  and  $G_{35}$  of CBA-0

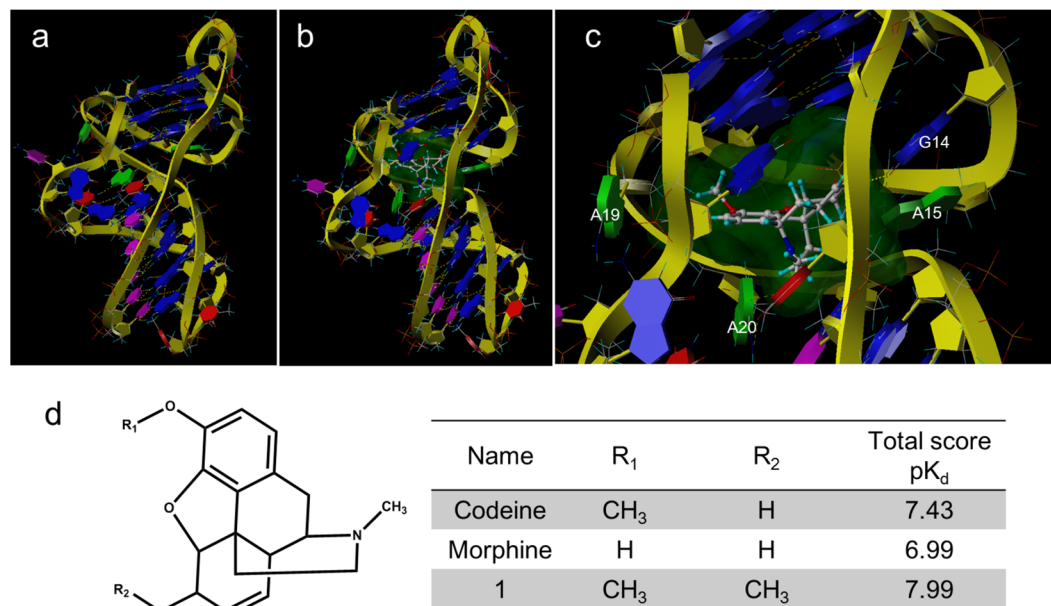
were further protected, but G<sub>28–32</sub> and G<sub>34</sub> became less protected, which agreed well with the formation of the triplex-quadruplex structure. Similar result was obtained in the DMS experiment of CBA-1 (Fig. 4b). In the absence of codeine (PBS buffer), the G-bases in quadruplex and in the second G-strand (G<sub>21–25</sub>) of triplex of CBA-0 and CBA-1 were less protected than those in the presence of codeine, which may be due to the structural diversity of both sequences in the absence of codeine, e.g. the opening and closing of the 3' end part as discussed earlier. Interestingly, G<sub>14</sub> in both sequences was not protected and even more hypersensitive to DMS methylation when binding to codeine, indicating that this G-base was exposed to DMS methylation and not involved in G-quadruplex formation. This is also in agreement with the NMR result that the G-quadruplex moiety contains two layers of G-tetrads after binding codeine.

**The binding site study.** The binding site in CBA was investigated by mutation assay. As G<sub>14</sub> and G<sub>34</sub> of CBA-1 were not protected from methylation in DMS assay, G<sub>14</sub> of CBA-1 was replaced by A (CBA-1-G14A). The mutated sequence still showed high binding ability to codeine (Figure S13). The <sup>1</sup>H-NMR spectra of CBA-1-G14A in the presence of codeine also showed well-resolved and sharp signals of triplex and quadruplex (Figure S14). These results further confirmed that G<sub>14</sub> of CBA-1 was not involved in the formation of G-quadruplex after codeine binding, and might not interact with codeine directly. Mutating G<sub>34</sub> of CBA-1 with T (CBA-1-G34T) greatly decreased its binding ability. Removing G<sub>34</sub> (CBA-1-G34R) or G<sub>34</sub>C<sub>35</sub> (CBA-1-G34C35R) caused the loss of binding ability totally, but did not affect the CD spectra of these sequences (Figures S13, S15), suggesting that G<sub>34</sub>C<sub>35</sub> played a key role in codeine binding. Since G<sub>34</sub>C<sub>35</sub> is located at the junction between G-quadruplex and triplex, codeine is likely bound in the pocket formed by G<sub>34</sub>C<sub>35</sub>, G-quadruplex and triplex.

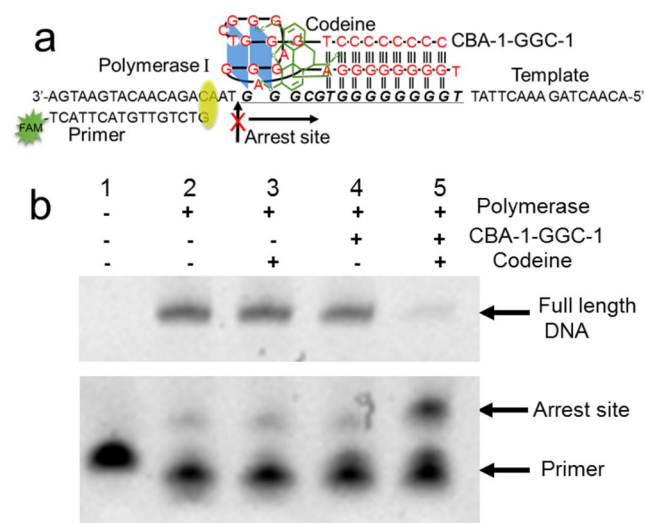
2-aminopurine (2-AP), an adenine analog with high intrinsic fluorescence, usually serves as a versatile, site-specific conformational reporter of nucleic acids<sup>47</sup>. The fluorescence of 2-AP is sensitive to its environment, e.g. quenched by stacking with other bases or molecule<sup>48</sup>. In order to further study the conformation, G<sub>14</sub>, A<sub>15</sub>, A<sub>19</sub> and A<sub>20</sub> of CBA-1 were replaced with 2-AP respectively (CBA-1-A14AP, CBA-1-A15AP, CBA-1-A19AP and CBA-1-A20AP). The mutated sequences remained their binding ability to codeine except for CBA-1-A19AP whose binding ability was greatly decreased (Figure S16). The fluorescence intensity of CBA-1-A14AP and CBA-1-A15AP in the absence of codeine was much higher than that of CBA-1-A19AP and CBA-1-A20AP (Figure S17), suggesting that AP<sub>14</sub> and AP<sub>15</sub> were free of interaction with other bases. Codeine binding to CBA-1-A14AP, CBA-1-A15AP and CBA-1-A20AP partly quenched their fluorescence, suggesting that codeine binding might have influenced the local environment of AP<sub>14</sub>, AP<sub>15</sub> and AP<sub>20</sub>, i.e. these bases are near the binding pocket. According to these results, the predicted scaffold 2-3 may not be the binding structure of CBA. Since replacing A<sub>19</sub> of CBA-1 with 2-AP (CBA-1-A19AP) greatly decreased its binding ability, this A<sub>19</sub> was further replaced with T or G (CBA-1-A19T and CBA-1-A19G), which also caused great drop of the binding ability to codeine for both sequences (Figure S16). The CD spectra of CBA-1-A19AP, CBA-1-A19G and CBA-1-A19T showed that their main CD peaks shifted from 261 nm to 257 nm (Figure S18), suggesting that A<sub>19</sub> is essential for maintaining the quadruplex structure. All the above results suggest that the predicted scaffold 1 is the most likely binding structure of CBA-1.

**Molecular modeling and docking.** Because we failed to obtain the single crystal of CBA-1/codeine complex for X-ray diffraction analysis; and failed to obtain the detail structure of CBA-1/codeine complex through 2D NMR<sup>34</sup>, we tried to understand the binding between codeine and CBA-1 through molecular docking using Sybyl X 2.0. Based on all of the results above, the binding structure of CBA-1 is most likely a triplex-quadruplex hybrid structure, in which the G-quadruplex moiety of the predicted scaffold 1 is similar to the (3 + 1) G-quadruplex form 2 of human telomere sequence<sup>37,49</sup>. The starting structure of CBA-1 was therefore generated by combining the reported NMR structures of (3 + 1) G-quadruplex (PDB entries 2JPZ)<sup>50</sup> and the G-GC triplex structures (PDB entries 134D)<sup>39</sup>. This molecular model was then energy-minimized using the Tripos force field with Gasteiger-Marsili charges (Fig. 5a–c). The binding site is located in the junction between the triplex and G-quadruplex, and G<sub>14</sub>, A<sub>15</sub>, A<sub>19</sub> and A<sub>20</sub> are near the binding pocket. Molecular docking was performed using Surflex-Dock program, with the docking scores expressed in  $-\lg K_d$  units, where  $K_d$  is the dissociation constant. The docking result showed that codeine bound to CBA-1 with a score of 7.34. According to the equation  $\Delta G = -RT \ln K$ ,  $\Delta G$  was calculated to be  $-41.87$  kJ/mol, which was very close to that measured by ITC ( $-40.9 \pm 0.5$  kJ/mol), suggesting that the binding was very stable. In the presence of codeine, the G-quartet plane facing the binding pocket was destroyed by the binding of codeine, which agreed with the NMR and DMS results that the G-quadruplex moiety contains two layers of G-tetrads. Furthermore, two codeine analogues, morphine and methyl substituted codeine (compound 1) were also docked into the binding pocket (Fig. 5d). Morphine showed a slightly weaker binding affinity than codeine (score: 6.99 vs 7.43), which is consistent with our previous results<sup>31</sup>. As a comparison, Methyl substituted codeine displayed higher affinity (score: 7.99) to CBA-1 than codeine, suggesting the substitution at the hydroxyl group of codeine did not affect the aptamer binding. This result is expected since CBA was selected using codeine immobilized beads by SELEX technique, in which codeine was linked to beads through this hydroxyl group.

**Polymerase Stop Assays.** Both triplex and G-quadruplex are reported to form in cells and repress gene transcription<sup>51,52</sup>. They are widely utilized as tools to regulate gene expression and study the molecular mechanisms of DNA repair, recombination, and mutagenesis<sup>53,54</sup>. However, a few factors may limit their application. For example, these structures form spontaneously in the presence of a physiological amount of monovalent cations<sup>55,56</sup>; G-rich sequences often form multiple inter/intra-molecular structures; and the triplex structures are much less stable than duplex and G-quadruplex<sup>57</sup>. In comparison, the CBA/codeine system may have some notable advantages because of the unique and stable structure induced by codeine. Since we have shown that codeine could induce the formation of triplex-quadruplex scaffold of the split CBA sequences, we have tested



**Figure 5.** (a) The predicted molecular model of CBA-1. (b) The predicted binding model of codeine to CBA-1 by molecular docking. (c) The detailed binding pocket of CBA-1 with codeine. The predicted binding pocket was shown as green. (d) The docking total score of codeine analogues.



**Figure 6.** (a) The principle of the polymerization arrest caused by triplex-quadruplex hybrid scaffold formation. (b) The gel electrophoresis of the extension product by polymerase I. DNA polymerization was arrest at the triplex-quadruplex scaffold induced by codeine. The full-length gel is presented in Supplementary Figure S20.

the feasibility of using codeine to control gene express by performing a DNA polymerase I stop assay. The DNA template contained the 3' end sequence of CBA-1-GGC (CBA-1-GGC-2, T-8G-TGC-3G); DNA polymerase I (Klenow Fragment) was used to extend primer sequence; sequence CBA-1-GGC-1 (the other part of CBA-1-GGC) was used to form the triplex-quadruplex structure with CBA-1-GGC-2 in the template (Fig. 6a). As shown in Fig. 6b, codeine or CBA-1-GGC-1 alone did not affect the extension of primer, but codeine and CBA-1-GGC-1 together almost totally halted the extension, indicating that codeine likely induced the formation of the triplex-quadruplex structure that blocked polymerase progression. Unlike many organic ligands that bind to most G-quadruplexes, codeine is highly specific for CBA<sup>31</sup>. These results may indicate great potentials for precise regulation of gene functions by using CBA sequences and codeine.

In summary, the binding structure of CBA has been investigated and a novel triplex-quadruplex scaffold has been proposed and confirmed. CBA and its derivatives are G-rich DNA sequences containing six G-tracts and a C-tract, which prefer to form structures with G-quadruplex and G-GC based triplex. The binding of codeine

induces CBA to form a well-defined triplex-quadruplex scaffold, and greatly enhances their thermal stability. The codeine binding pocket is located in the cavity of the junction formed by the triplex and quadruplex. When split in two, the two individual parts of CBA cannot bind codeine on their own, and can only bind codeine when both are present. A polymerase stop experiment showed that the split sequences could form the triplex-quadruplex scaffold induced by codeine, resulting in blocking of the polymerase reaction. This novel scaffold not only has implication in rational design of aptameric sensors, but also has the potential to function as regulators of gene expression. Although the detailed structure of CBA cannot be elucidated currently, the triplex-quadruplex hybrid scaffold offers new insights in DNA architecture and molecular recognition.

## Materials and Methods

**Materials.** Synthetic oligodeoxynucleotides were obtained from Sangon Biotech Co. Ltd. (Shanghai, China) or Zixi Biotech Co. Ltd. (Beijing, China). Codeine was purchased from the Material Evidence Identification Center of Ministry of Public Security (Beijing, China). The binding buffer was phosphate buffer saline (PBS, 140 mM NaCl, 2.5 mM KCl, 1.6 mM  $\text{KH}_2\text{PO}_4$ , 15 mM  $\text{Na}_2\text{HPO}_4$ ) with 2.5 mM  $\text{MgCl}_2$ . Unless otherwise stated, all experimental samples were annealed prior to use at 95 °C for 5 min followed by quenching on ice for 10 min.

**Isothermal titration calorimetry assay.** Isothermal titration calorimetry (ITC) assay was performed on a Nano ITC System (TA Instruments Inc., New Castle, DE, USA). Injections of 10  $\mu\text{L}$  of 250  $\mu\text{M}$  codeine were added from a computer-controlled microsyringe at an interval of 400 s into the DNA aptamer solution (about 8–15  $\mu\text{M}$ , cell volume 1.0 mL) with 300 rpm stirring at 25 °C. Heat produced by codeine dilution was evaluated by performing a control experiment, titrating codeine into the buffer alone. The experimental data were fitted to a theoretical titration curve using software supplied by TA, with  $\Delta G$  (Gibbs free energy),  $K_d$  (association constant), and  $n$  (number of binding sites per monomer), as adjustable parameters.

**Fluorescence spectra experiments.** Fluorescence spectra were collected with a Hitachi F-4600 fluorescence fluorometer (Kyoto, Japan) at 25 °C. For the competition experiments with G-quadruplex probe BPBC, 1  $\mu\text{M}$  DNA sequence and 3  $\mu\text{M}$  BPBC were incubated in the absence or presence of codeine (3  $\mu\text{M}$ ) for 30 min at 25 °C. Subsequently, emission spectra were measured from 420 nm to 600 nm with excitation of 408 nm, where excitation and emission slits were set to 5 nm. Single 2-aminopurine substituted codeine binding aptamers were excited at 305 nm and emission spectra were measured from 330 nm to 460 nm, where excitation and emission slits were set to 5 nm.

**Circular dichroism analysis.** Circular dichroism analysis were performed on a JASCO spectrometer in 10 mm path length quartz cuvettes. The spectra were recorded with 100 nm/min, at 0.5 nm bandwidth, at 25 °C and normalized by subtraction of the background scan of the buffer.

**Melting assay.** For thermal denaturation measurements, folded samples were heated from 25 °C to 85 °C with a heating rate of 1 °C  $\text{min}^{-1}$  on a JASCO spectrometer in 10 mm path length quartz cuvettes. The UV signal at 265 nm was recorded every 1 °C.

**$^1\text{H}$  NMR experiments.** The DNA used for NMR was purified by RP-HPLC. If not particularly indicated, DNA samples were dissolved in 0.5 mL of PBS buffer containing 10%  $\text{D}_2\text{O}$ , annealed by heating to 95 °C for 5 min and quenched on ice for 10 min. The final concentration of DNA samples was 200  $\mu\text{M}$ . For the presence of codeine samples, the concentration of codeine was 400  $\mu\text{M}$ . NMR experiments were performed on 600 MHz Varian Bruker spectrometers with data recorded at 25 °C. 1D-proton spectra were acquired with 32 000 data points using 6k accumulated scans due to low sample concentration and processed with an exponential line broadening window function. Solvent suppression was achieved by excitation sculpting. Acquired data were processed and analyzed using Bruker Topspin and MestReNova software.

**Dimethyl sulfate (DMS) footprinting assay.** Dimethyl sulfate footprinting experiments were performed according to the protocol developed by Tan *et al.*<sup>58</sup> 10 picomoles of FAM-labeled CBA-0 or CBA-1 in 200  $\mu\text{L}$  volume were mixed with 4  $\mu\text{L}$  of 10% (v/v) dimethyl sulfate (DMS) in ethanol and incubated for 2 min at room temperature. The reaction was stopped by addition of 200  $\mu\text{L}$  of stop buffer (0.6 M NaOAc, 0.1 M  $\beta$ -mercaptoethanol, 20  $\mu\text{g}$  sperm DNA). After phenol/chloroform extraction and ethanol precipitation, the DNA was dissolved in 50  $\mu\text{L}$  of water. Thereafter, 50  $\mu\text{L}$  of 20% (v/v) piperidine in water was added and the samples were heated at 90 °C for 30 min, followed by phenol/chloroform extraction and ethanol precipitation. The precipitated DNA was dissolved in 50% (v/v) deionized formamide in water, denatured at 95 °C for 5 min and resolved on a denaturing 12% polyacrylamide gel.

**Molecular Modeling and docking.** Starting structure of (3 + 1) G-quadruplex Form II was generated from the coordinates of the reported NMR structures (PDB entries 2JPZ)<sup>50</sup>. The G-GC triplex structures was also generated from the coordinates of the reported NMR structures (PDB entries 134D)<sup>39</sup>. Molecular modeling was energy-minimized using the Tripos force field for binding energy calculation by the Powell method with a convergence criterion of 0.05 kcal/mol-Å with the maximum iterations set to 500. Molecular docking on CBA-1 was performed using Sybyl-X 2.0 program (Tripos Inc.). All of the hydrogen atoms and charges were added to define the correct configuration and tautomeric states. Docking pocket was generated manually. Then, the structure of the compound codeine built by the Sybyl program was docked into the binding pocket for docking analysis.



**Polymerase Stop Assays.** The single stand DNA template CBA-Tem was designed as the following: 5'-ACAGACTAGAACTTATTGGGGGGGGG TGCGGG TAACAGACAACA TGAATGA-3', which contained part of CBA-1-GGC (underlined sequence). The primer CBA-Pri was FAM-TCA TTC ATG TTG TCT G, with a fluorescein label at its 5' end. The other part of CBA-1-GGC, CCCCCCCTGGGTCGGGAGGG AAGGGGGGGGT (named CBA-1-GGC-1) was designed to stop the polymerase reaction. 2  $\mu$ M CBA-Tem was mixed with 4  $\mu$ M CBA-Pri and CBA-1-GGC-1 in PBS buffer and heated at 95 °C for 5 min followed by quenching on ice for 10 min. Then the mixture were diluted four times with 1X NEBuffer 2 (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT) and incubated with 5  $\mu$ M codeine at room temperature for 30 min. Subsequently, 120  $\mu$ M dNTP and 5 U DNA polymerase I (Klenow Fragment) was added and followed by polymerization for 5 min at room temperature. Denaturing PAGE was performed to analyze.

## References

- Liu, J., Cao, Z. & Lu, Y. Functional nucleic acid sensors. *Chem. Rev.* **109**, 1948–1998 (2009).
- Lau, P. S. & Li, Y. Exploration of structure-switching in the design of aptamer biosensors. *Adv. Biochem. Eng. Biot.* **140**, 69–92 (2014).
- Tang, Y. T., Ge, B. X., Sen, D. & Yu, H. Z. Functional DNA switches: rational design and electrochemical signaling. *Chem. Soc. Rev.* **43**, 518–529 (2014).
- Zhang, F., Nangreave, J., Liu, Y. & Yan, H. Structural DNA Nanotechnology: State of the Art and Future Perspective. *J. Am. Chem. Soc.* **136**, 11198–11211 (2014).
- Rich, A. & Rajbhandary, U. L. Transfer RNA: Molecular Structure, Sequence And Properties. *Annu Rev Biochem* **45**, 805–860 (1976).
- Davis, J. T. G quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed.* **43**, 668–698 (2004).
- Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **34**, 5402–5415 (2006).
- Xu, Y. Chemistry in human telomere biology: structure, function and targeting of telomere DNA/RNA. *Chem. Soc. Rev.* **40**, 2719–2740 (2011).
- Brooks, T. A., Kendrick, S. & Hurley, L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *Febs. J.* **277**, 3459–3469 (2010).
- Bugaut, A. & Balasubramanian, S. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.* **40**, 4727–4741 (2012).
- Agarwala, P., Pandey, S. & Maiti, S. Role of G-quadruplex located at 5' end of mRNAs. *BBA-Gen. Subjects* **1840**, 3503–3510 (2014).
- Wolfe, A. L. *et al.* RNA G-quadruplexes cause eIF4A-dependent oncogene translation in cancer. *Nature* **513**, 65–70 (2014).
- Felsenfeld, G., Davies, D. R. & Rich, A. Formation of a three-stranded polynucleotide molecule. *J. Am. Chem. Soc.* **79**, 2023–2024 (1957).
- Hoyne, P. R., Edwards, L. M., Viari, A. & Maher, L. J. Searching genomes for sequences with the potential to form intrastrand triple helices. *J. Mol. Biol.* **302**, 797–809 (2000).
- Holder, I. T. *et al.* Intrastrand triplex DNA repeats in bacteria: a source of genomic instability. *Nucleic Acids Res.* **43**, 10126–10142 (2015).
- Devi, G., Zhou, Y., Zhong, Z. S., Toh, D. F. K. & Chen, G. RNA triplexes: from structural principles to biological and biotech applications. *Wires Rna* **6**, 111–128 (2015).
- Conrad, N. K. The emerging role of triple helices in RNA biology. *Wires Rna* **5**, 15–29 (2014).
- Belotserkovskii, B. P. *et al.* A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. *J. Biol. Chem.* **282**, 32433–32441 (2007).
- Bruno, J. G. Predicting the Uncertain Future of Aptamer-Based Diagnostics and Therapeutics. *Molecules* **20**, 6866–6887 (2015).
- Tolle, F., Brändle, G. M., Matzner, D. & Mayer, G. A Versatile Approach Towards Nucleobase-Modified Aptamers. *Angew. Chem. Int. Ed.* **54**, 10971–10974 (2015).
- Tan, W. H., Donovan, M. J. & Jiang, J. H. Aptamers from Cell-Based Selection for Bioanalytical Applications. *Chem. Rev.* **113**, 2842–2862 (2013).
- de-los-Santos-Álvarez, N., Miranda-Ordieres, A. J. & Tuñón-Blanco, P. Aptamers as recognition elements for label-free analytical devices. *TrAC Trend. Anal. Chem.* **27**, 437–446 (2008).
- Zhang, Z., Oni, O. & Liu, J. New insights into a classic aptamer: binding sites, cooperativity and more sensitive adenosine detection. *Nucleic Acids Res* **45**, 7593–7601 (2017).
- Patel, D. J. Structural analysis of nucleic acid aptamers. *Curr. Opin. Chem. Biol.* **1**, 32–46 (1997).
- Bing, T. *et al.* G-quadruplex DNA aptamers generated for systemin. *Bioorg. Med. Chem.* **19**, 4211–4219 (2011).
- Mei, H. *et al.* Functional-group specific aptamers indirectly recognizing compounds with alkyl amino group. *Anal. Chem.* **84**, 7323–7329 (2012).
- Qi, C. *et al.* G-quadruplex DNA aptamers for zeatin recognizing. *Biosens. Bioelectron.* **41**, 157–162 (2013).
- Huang, H. *et al.* A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol.* **10**, 686–691 (2014).
- Trachman, R. J. *et al.* Structural basis for high-affinity fluorophore binding and activation by RNA Mango. *Nat. Chem. Biol.* **13**, 807–813 (2017).
- Nasiri, A. H., Wurm, J. P., Immer, C., Weickhmann, A. K. & Wohnert, J. An intermolecular G-quadruplex as the basis for GTP recognition in the class V-GTP aptamer. *RNA* **22**, 1750–1759 (2016).
- Huang, L. *et al.* A label-free electrochemical biosensor based on a DNA aptamer against codeine. *Anal. Chim. Acta.* **787**, 203–210 (2013).
- Tippana, R., Xiao, W. K. & Myong, S. G-quadruplex conformation and dynamics are determined by loop length and sequence. *Nucleic Acids Res.* **42**, 8106–8114 (2014).
- Mathad, R. I., Hatzakis, E., Dai, J. X. & Yang, D. Z. c-MYC promoter G-quadruplex formed at the 5'-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability. *Nucleic Acids Res.* **39**, 9023–9033 (2011).
- Vorlickova, M. *et al.* Circular dichroism and guanine quadruplexes. *Methods* **57**, 64–75 (2012).
- Jin, B. *et al.* Fluorescence light-up probe for parallel G-quadruplexes. *Anal. Chem.* **86**, 943–952 (2014).
- Lim, K. W. *et al.* Coexistence of Two Distinct G-Quadruplex Conformations in the hTERT Promoter. *J. Am. Chem. Soc.* **132**, 12331–12342 (2010).
- Yue, D. J. E., Lim, K. W. & Phan, A. T. Formation of (3 + 1) G-quadruplexes with a long loop by human telomeric DNA spanning five or more repeats. *J. Am. Chem. Soc.* **133**, 11462–11465 (2011).
- Feigon, J., Koshlap, K. M. & Smith, F. W. 1 H NMR spectroscopy of DNA triplexes and quadruplexes. *Method. Enzymol.* **261**, 225–255 (1995).
- Radhakrishnan, I. & Patel, D. J. Solution Structure Of a Purine Purine Pyrimidine DNA Triplex Containing G-Center-Dot-Gc And T-Center-Dot-at Triples. *Structure* **1**, 135–152 (1993).

40. Dailey, M. M., Miller, M. C., Bates, P. J., Lane, A. N. & Trent, J. O. Resolution and characterization of the structural polymorphism of a single quadruplex-forming sequence. *Nucleic Acids Res.* **38**, 4877–4888 (2010).
41. Plum, G. E., Park, Y. W., Singleton, S. F., Dervan, P. B. & Breslauer, K. J. Thermodynamic Characterization of the Stability and the Melting Behavior of a DNA Triplex - a Spectroscopic and Calorimetric Study. *Proc. Natl. Acad. Sci. USA* **87**, 9436–9440 (1990).
42. Lavelle, L. & Fresco, J. R. Stabilization of nucleic acid triplexes by high concentrations of sodium and ammonium salts follows the Hofmeister series. *Biophys. Chem.* **105**, 681–699 (2003).
43. Carmona, P. & Molina, M. Binding of oligonucleotides to a viral hairpin forming RNA triplexes with parallel G\*G center dot C triplets. *Nucleic Acids Res.* **30**, 1333–1337 (2002).
44. Limongelli, V. *et al.* The G-Triplex DNA. *Angew. Chem. Int. Ed.* **52**, 2269–2273 (2013).
45. Liu, J. Q., Xiao, S., Hao, Y. H. & Tan, Z. Strand-Biased Formation of G-Quadruplexes in DNA Duplexes Transcribed with T7 RNA Polymerase. *Angew. Chem. Int. Ed.* **54**, 8992–8996 (2015).
46. Olivas, W. M. & Maher, L. J. Binding of DNA oligonucleotides to sequences in the promoter of the human bcl-2 gene. *Nucleic Acids Res.* **24**, 1758–1764 (1996).
47. Gray, R. D., Petraccone, L., Buscaglia, R. & Chaires, J. B. 2-aminopurine as a probe for quadruplex loop structures. *G-Quadruplex DNA: Methods and Protocols* 121–136 (2010).
48. Li, J., Correia, J. J., Wang, L., Trent, J. O. & Chaires, J. B. Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucleic Acids Res.* **33**, 4649–4659 (2005).
49. Phan, A. T., Kuryavyi, V., Luu, K. N. & Patel, D. J. Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K<sup>+</sup> solution. *Nucleic Acids Res.* **35**, 6517–6525 (2007).
50. Dai, J. X., Carver, M., Punchihewa, C., Jones, R. A. & Yang, D. Z. Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K<sup>+</sup> solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.* **35**, 4927–4940 (2007).
51. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. USA* **99**, 11593–11598 (2002).
52. Frankkamenetskii, M. D. & Mirkin, S. M. Triplex DNA Structures. *Annu. Rev. Biochem.* **64**, 65–95 (1995).
53. Mukherjee, A. & Vasquez, K. M. Triplex technology in studies of DNA damage, DNA repair, and mutagenesis. *Biochimie* **93**, 1197–1208 (2011).
54. Han, M. S., Lytton-Jean, A. K. R. & Mirkin, C. A. A gold nanoparticle based approach for screening triplex DNA binders. *J. Am. Chem. Soc.* **128**, 4954–4955 (2006).
55. Olivas, W. M. & Maher, L. J. Overcoming potassium-mediated triplex inhibition. *Nucleic Acids Res.* **23**, 1936–1941 (1995).
56. Hari, Y., Obika, S. & Imanishi, T. Towards the Sequence-Selective Recognition of Double-Stranded DNA Containing Pyrimidine-Purine Interruptions by Triplex-Forming Oligonucleotides. *Eur J Org Chem* **2012**, 2875–2887 (2012).
57. Fox, K. R. & Brown, T. Formation of stable DNA triplexes. *Biochem. Soc. Trans.* **39**, 629–634 (2011).
58. Zheng, K. W., Chen, Z., Hao, Y. H. & Tan, Z. Molecular crowding creates an essential environment for the formation of stable G-quadruplexes in long double-stranded DNA. *Nucleic Acids Res.* **38**, 327–338 (2010).

## Acknowledgements

This work was supported by NSF of China (21635008, 21535009, 21575147, 21375135 and 21621062).

## Author Contributions

T.B. and D.S. conceived and designed the experiments. T.B. performed the experiments with the assistance of W.Z., X.Z., L.S., J.C., T.B., D.S., W.Z., X.L. and F.W. analyzed the data. T.B., Z.C. and D.S. wrote the paper. D.S. directed the overall project.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-15797-5>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017