JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

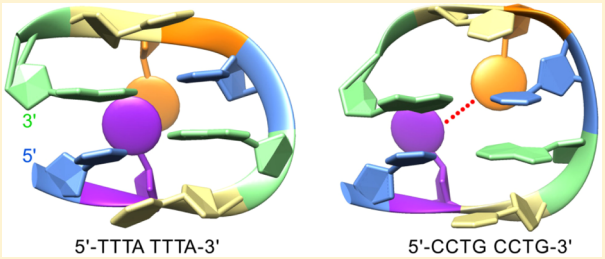
# Minidumbbell: A New Form of Native DNA Structure

Pei Guo and Sik Lok Lam*

Department of Chemistry, The Chinese University of Hon Kon , Shatin, New Territories, Hon Kon

Supporting Information

**ABSTRACT:** The non-B DNA structures formed by short tandem repeats on the nascent strand durin DNA replication have been proposed to be the structural intermediates that lead to repeat expansion mutations. Tetranucleotide TTTA and CCTG repeat expansions have been known to cause reduction in biofilm formation in *Staphylococcus aureus* and myotonic dystrophy type 2 in human, respectively. In this study, we report the first three-dimensional minidumbbell (MDB) structure formed by natural DNA sequences containin two TTTA or CCTG repeats. The formation of MDB provides possible pathways for strand slippa e to occur, which ultimately leads to repair escape and thus expansion mutations. Our result here shows that MDB is a hi hly compact structure composed of two type II loops. In addition to the typical stabilizin interactions in type II loops, MDB shows extensive stabilizin forces between the two loops, includin two distinctive modes of interactions between the minor roove residues. The formation of MDB enriches the structural diversity of natural DNA sequences, reveals the importance of loop−loop interactions in unusual DNA structures, and provides insi hts into novel mechanistic pathways of DNA repeat expansion mutations.




## INTRODUCTION

DNA is well-known to adopt the ri ht-handed double-helical B-form structure.[1] Over the past five decades, DNA has also been shown to be capable of adoptin non-B structures such as A- and Z-form DNA,[2,3] bul e, hairpin,[4,5] dumbbell,[6] three-way junction,[7] triplex,[8] sticky DNA,[9] quadruplex,[10] i-motif,[11,12] and cruciform.[13] These non-B structures have been demonstrated to participate in various biolo ical processes such as ene re ulation, DNA replication, transcription, dama e, and repair. In particular, the non-B structures adopted by short tandem repeats in the nascent strand durin DNA replication, e. ., CTG hairpin,[5] GAA triplex,[14] CGG quadruplex,[15] GCC i-motif,[12] and CCTG dumbbell,[6] have been proposed to be the culprits leadin to repeat expansion mutations[16−18] which brin about nearly 30 human enetic disorders.[17] Meanwhile, non-B structures have also become fascinatin buildin blocks in DNA nanotechnolo y and material science owin to their unique structural features.[19,20]

Recently, we showed that two TTTA[21] or CCTG repeats[22] are capable to fold into a minidumbbell (MDB) structure, which not only provides possible pathways for the occurrence of TTTA and CCTG repeat expansions in *Staphylococcus aureus* and myotonic dystrophy type 2 patients, respectively, but also enriches the structural diversity of natural DNA sequences. The MDB structure comprises two tetranucleotide type II loops with 3′-5′ terminal stackin . In a type II loop,[23−25] the first and fourth loop residues form a loop-closin base pair whereas the second and third residues fold into the minor roove and stack on the base pair, respectively. Yet how two adjacent type II loops in a sin le DNA strand lead to the formation of MDB remains elusive. Therefore, we have determined the three-dimensional solution structures of TTTA and CCTG MDBs in this study. Our results reveal these MDBs are hi hly compact with extensive stabilizin interactions between the two loops. We have also identified two distinctive modes of stabilization between the minor roove residues.

## MATERIALS AND METHODS

This section only provides a brief description of key materials and experimental methods. The detailed experimental procedures are described in Supportin Information (SI) Materials and Methods.

**DNA Samples.** The two DNA samples used in this study contain the sequence 5′-TTTA TTTA-3′ and 5′-CCTG CCTG-3′, respectively. For simplicity, these two DNA samples were named as "$(TTTA)_2$" and "$(CCTG)_2$". NMR samples were prepared by dissolvin 0.5 $\mu$mol of purified DNA into 500 $\mu$L buffer solutions containin 10 mM sodium phosphate (pH 7.0), and 0.1 mM 2,2-dimethyl-2-silapentane-5-sulfonic acid.[21,22]

**NMR Spectroscopy.** The details of NMR experiments and resonance assi nments are described in Supportin Information. To extract the nonlabile proton NOEs, the samples were prepared in a 99.96% $D_2O$ buffer solution, and 2D NOESY spectra were acquired with mixin times of 100, 300, and 600 ms at 5 °C unless otherwise specified. To study the labile protons, the solvent was exchan ed with a 90% $H_2O$/10% $D_2O$ buffer solution. The 2D NOESY and 1D NOE difference spectra were acquired usin the excitation sculptin water suppression method.[26] For the measurements of the $^3J_{H1'H2'}$, $^3J_{H4'H5'}$, and $^3J_{H4'H5''}$ couplin constants, DQF-COSY spectra were acquired.

**Experimental Restraints.** Proton−proton distance restraints were obtained from NOESY spectra based on the intensities of NOE cross peaks. A total of 242 and 274 distance restraints were obtained for

**Received:** July 4, 2016
**Published:** September 2, 2016





DOI: 10.1021/jacs.6b06897
J. Am. Chem. Soc. 2016, 138, 12534−12540

$(TTTA)_2$ and $(CCTG)_2$, respectively. Besides, distance restraints based on crystallo raphic data for hydro en bonds in Watson−Crick T-A and C-G base pairs[27] were used. The H1′−C1′−C2′−H2′ su ar torsion an les were determined by the $^3J_{H1'H2'}$ couplin constants measured from the DQF-COSY spectra and the Karplus equation.[28] Glycosidic torsion an les $\chi$ were obtained based on the intranucleotide H6/H8−H1′ NOE intensities. Restraints for backbone torsion an les $\gamma$ were determined based on the analysis of $^3J_{H4'H5'/H5''}$ couplin constants.[29] A summary of the restraints used in calculatin the structures of $(TTTA)_2$ and $(CCTG)_2$ is shown in Tables S1 and S2.

**Structure Calculations.** Restrained molecular dynamics (rMD) calculations were performed usin AMBER[30] with the ff12SB force field.[31] See the protocol in Supportin Information.

**Data Analysis.** The pseudorotation phase an les ($P$) of deoxyribose puckers were measured usin the CPPTRAJ module of AMBER 12.[32] The criteria of hydro en bond and hydrophobic interaction are stated in Supportin Information. All fi ures of the calculated structures were enerated usin UCSF Chimera.[33]

## RESULTS AND DISCUSSION

**Overview of the TTTA MDB Solution Structure.** For $(TTTA)_2$, 20 refined structures with lowest restraint violation ener ies were selected in the final representative ensemble ensemble (PDB ID: 5GWQ). Superimposition of them shows the TTTA MDB structure was well-defined with reasonable precision (Fi ure 1A). Specifically, the first and fourth loop residues, i.e., T1 and A4, and T5 and A8, form the loop-closin base pairs. The second loop residues, namely T2 and T6, fold into the minor roove and partially stack with each other whereas the third loop residues T3 and T7 stack on T1-A4 and T5-A8 base pairs, respectively. Amon the 20 structures, the avera e pairwise RMSD was found to be 0.87 ± 0.20 Å and the RMSD from the mean structure was 0.60 ± 0.14 Å for all residues (Table 1). All these structures show (i) satisfactory a reement with experimental restraints with no lar e distance and torsion an le violations and (ii) ood covalent eometries with no si nificant bond and an le violations (Table 1). The backbone and lycosidic torsion an les and pseudorotation phase an les are summarized in Fi ure S1.

**The Core Scaffold of TTTA MDB Was Constructed by Two Watson−Crick Loop-Closing Base Pairs with 3′-5′ Terminal Stacking.** The core scaffold was constructed by two well-defined loop-closin T-A base pairs with an avera e pairwise RMSD of 0.78 ± 0.20 Å (Table 1). The T1-A4 and T5-A8 base pairs adopt Watson−Crick pairin eometry with an extensive stack between the A8 and T1 termini (Fi ure 1B). As supported by the 1D NOEs of A4/A8 H2 but not H8 by saturatin the T1/T5 imino si nals at ∼13.5 ppm (Fi ure S2A,B), Watson−Crick hydro en bond restraints were added in the structural refinement process to increase the chance of obtainin the structures with lowest restraint violation ener y. To verify the Watson−Crick pairin modes, we also repeated the structural refinement process by removin these hydro en bond restraints. Amon 100 rMD trials with random startin velocities, the structure with lowest restraint violation remains an MDB with two T-A Watson−Crick base pairs (Fi ure S2C).

As the observed Watson−Crick base pairin s differ from the Hoo steen pairin s found in the TTTA loop of hairpin[23,34] and TTCA loop of cyclic DNA,[35] we a ain repeated the structural refinement by incorporatin Hoo steen hydro en bond restraints to avoid underestimatin the possibility of formin Hoo steen loop-closin base pairs. Amon 100 trials, the lowest restraint violation ener y was found to be about 6-fold hi her than that with Watson−Crick hydro en bond restraints.

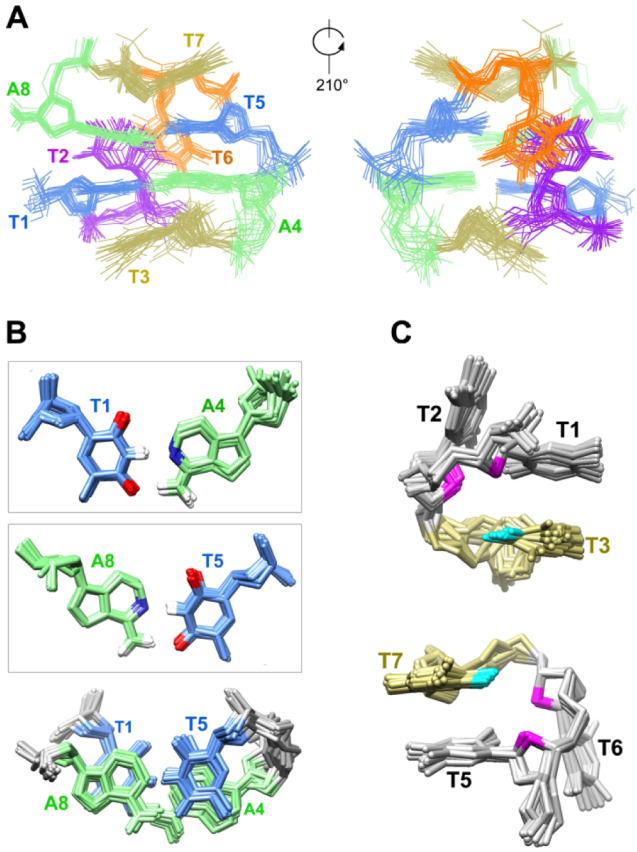



**Figure 1.** MDB structure of $(TTTA)_2$. (A) The major and minor roove views of 20 superimposed structures of $(TTTA)_2$. The third loop residues T3 and T7 stack on the T1-A4 and T5-A8 loop-closin base pairs while the second loop residues T2 and T6 fold into the minor roove and partially stack with each other. (B) T1-A4 and T5-A8 form the Watson−Crick loop-closin base pairs (top) with extensive base−base stackin (bottom). (C) Hydrophobic interactions were observed between T3/T7 methyl (cyan) and the 2′-methylene roups (ma enta) of its two precedin residues.

The results of the above two tests support the Watson−Crick pairin eometry in the two loop-closin base pairs in TTTA MDB. To ether with the extensive 3′-5′ terminal stack (Fi ure 1B), as supported by the base−base NOEs between A8 and T1 (Fi ure S3A), these loop-closin base pairs provide substantial stabilization in constructin the core scaffold of TTTA MDB.

**The Third Loop Residues Show Stacking and Hydrophobic Interactions.** In type II TTTA loops, the third thymine residue stabilizes the loop throu h stackin with the loop-closin base pair.[23,34] In TTTA MDB, T3 and T7 were also found to stack on the loop-closin base pairs (Fi ure 1A). These stackin interactions are supported by the base−base NOEs includin T3 H7-T1 H6, T3 H6-A4 H2, T7 H7-T5 H6, and T7 H6-A8 H2 (Fi ure S3A). It has been su ested that the formation of a Hoo steen instead of Watson−Crick loop-closin base pair would provide better stackin for the third residue in type II loop due to the shorter C1′−C1′ distance.[23,24,36] However, it is apparent that the 3′-5′ terminal stack between the two Watson−Crick loop-closin base pairs is more crucial toward the formation of TTTA MDB. As a result, the stackin interactions between T1-A4 and T5-A8 Watson−Crick base pairs outwei h the enhanced stabilizin effects of T3 on T1-A4 and T7 on T5-A8 Hoo steen base pairs, makin T-A


DOI: 10.1021/jacs.6b06897
J. Am. Chem. Soc. 2016, 138, 12534−12540

**Table 1. NMR and Refinement Statistics**

| | $(TTTA)_2$ | $(CCTG)_2$ |
|---|---|---|
| Structural Restraints | | |
| number of distance restraints | | |
| inter-residue | 103 | 82 |
| intraresidue | 139 | 192 |
| hydro en bond | 4 | 6 |
| subtotal | 246 | 280 |
| number of torsion an le restraints | | |
| lycosidic ($\chi$) | 8 | 8 |
| su ar (H1′−C1′−C2′−H2′) | 2 | 4 |
| backbone ($\gamma$) | 3 | 5 |
| subtotal | 13 | 17 |
| Restraint Satisfaction | | |
| distance restraints (Å) | | |
| number of violations >0.2 Å | 1.6 ± 0.9 | 0.6 ± 0.7 |
| maximum violation | 0.26 | 0.25 |
| avera e violation | 0.09 ± 0.06 | 0.08 ± 0.05 |
| torsion an le restraints (de ) | | |
| number of violations >5° | 0 | 0 |
| maximum violation | 0.7 | 0 |
| avera e violation | 0.6 ± 0.1 | 0 |
| Deviations from Covalent Geometry | | |
| bonds (Å) | 0.0086 ± 0.0002 | 0.0100 ± 0.0003 |
| an les (de ) | 2.5 ± 0.1 | 2.5 ± 0.1 |
| Heavy Atomic RMSD (Å)[a] | | |
| all residues | | |
| avera e pairwise RMSD | 0.87 ± 0.20 | 1.06 ± 0.28 |
| RMSD from mean structure | 0.60 ± 0.14 | 0.73 ± 0.19 |
| loop-closin base pairs | | |
| avera e pairwise RMSD | 0.78 ± 0.20 | 0.77 ± 0.24 |
| RMSD from mean structure | 0.54 ± 0.10 | 0.54 ± 0.15 |

[a]Pairwise RMSD was calculated amon 20 refined structures.

Watson−Crick base pairs more favorable than T-A Hoo steen base pairs in TTTA MDB.

Apart from base−base stackin , the su ar rin s of the third loop residues T3 and T7 also directly stack on A4 and A8 of the loop-closin base pairs, respectively (Fi ure 1A), as supported by the more upfield chemical shifts of the H1′, H2′, and H2″ su ar protons than those of other residues (Table S3). In addition, hydrophobic interactions were also observed between the methyl roup of T3 or T7 and the 2′-methylene roups of its two precedin residues (Fi ure 1C). In these hydrophobic cores, the distances from T3 C7 to T1 C2′/T2 C2′ and from T7 C7 to T5 C2′/T6 C2′ were found to be 3.4 ± 0.1 Å/4.7 ± 0.2 Å, and 3.6 ± 0.1 Å/5.4 ± 0.2 Å, respectively. In eneral, the hydrophobic interaction involvin T3 or T7 with its second precedin residue was found to be stron er than that with its first precedin residue.

**Folding of T2 and T6 into the Minor Groove.** In TTTA MDB, both of the second loop residues T2 and T6 were folded into the minor roove, with T2 bein closer to the loop-closin base pairs than T6 (Fi ure 1A). Their relative positions are supported by (i) the presence of NOEs between T2 and A4/A8 and (ii) the absence of NOEs between T6 and A4/A8 (Fi ure S3B). Amon the 20 refined structures, five show a T2 H3-T5 O2 hydro en bond (Fi ure 2A), nine show a T2 H3-T7 O5′ hydro en bond (Fi ure 2B), and six show a T2 H3-T7 OP1 hydro en bond (Fi ure 2C). These indicate that the foldin of T2 into the minor roove was driven by the formation of hydro en bond between T2 imino and the loop-closin base pair T5 O2 or the phosphodiester backbone T7 O5′/OP1 of the second TTTA repeat. Owin to the stabilization between T2 and T5/T7, the loop formed by the second repeat was more well-defined than the loop formed by the first repeat (Fi ure 2A−C), as evidenced by a much smaller avera e pairwise RMSD of 0.61 ± 0.16 Å for the second repeat than that of 0.89 ± 0.23 Å for the first repeat.

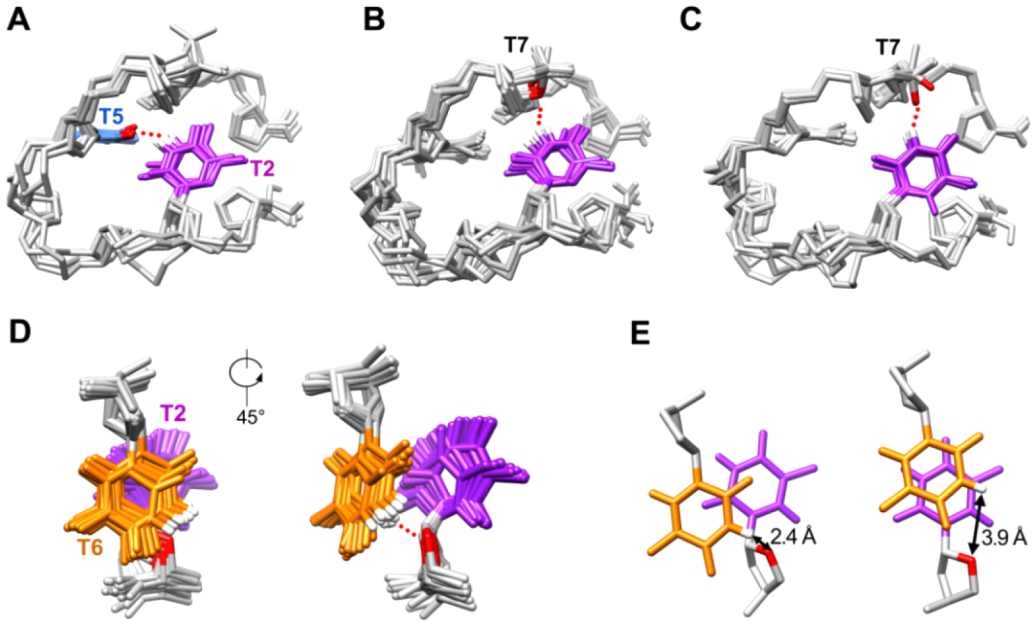



**Figure 2.** Stabilizin interactions involvin the minor roove T2 and T6 residues. (A) Five structures show a T2 H3-T5 O2 hydro en bond with a bond len th of 2.2 ± 0.4 Å. (B) Nine structures show a T2 H3-T7 O5′ hydro en bond with a bond len th of 2.9 ± 0.3 Å. (C) Six structures show a T2 H3-T7 OP1 hydro en bond with a bond len th of 2.1 ± 0.6 Å. (D) T6 stacks with T2 (left). An additional a T6 H3-T2 O4′ hydro en bond was observed with a bond len th of 2.8 ± 0.4 Å (ri ht). (E) A stron er T6 H3-T2 O4′ hydro en bond was observed in the case that T2 and T6 were not well-stacked (left). With better stackin , the T6 H3-T2 O4′ hydro en bond was found to be weaker (ri ht).



DOI: 10.1021/jacs.6b06897
J. Am. Chem. Soc. 2016, 138, 12534−12540

Instead of a hydro en bond, the foldin of T6 into the minor roove was predominantly driven by base–base stackin with T2 (Fi ure 2D). This stackin eometry is supported by the NOEs of T2 H7-T6 H1′ and T2 H1′-T6 H7 (Fi ure S3B). In addition to this minor roove residues' stack which has also been observed in cyclic DNA,[35] an unprecedented T6 H3-T2 O4′ hydro en bond was found to complement the base–base stackin in the minor roove (Fi ure 2D). Pendin the de ree of stackin overlap between T2 and T6, the len th of this hydro en bond was found to vary from 2.4 to 3.9 Å in the 20 refined structures (Fi ure 2E).

**Overview of the CCTG MDB Solution Structure.** Similarly, 20 refined structures of $(CCTG)_2$ with lowest restraint violation ener ies were selected in the final representative ensemble (PDB ID: 5GWL). The core scaffold of CCTG MDB was also well-defined by the two loop-closin base pairs (Fi ure 3A) with an avera e pairwise RMSD of 0.77 ± 0.24 Å (Table 1). CCTG MDB shows some structural similarities to TTTA MDB, includin (i) two Watson–Crick loop-closin base pairs (Fi ure 3B) as supported by the NOEs between G4/G8 imino and C1/C5 amino protons (Fi ure S4A), (ii) the 3′-5′ terminal stack (Fi ure 3B) as supported by the base–base NOEs of C1 H6-G8 H8 and C1 H5-G8 H8 (Fi ure S4B), (iii) the stackin of the third loop residues on the base pairs (Fi ure 3A) as supported by the base–base NOEs between T3 and C1, and between T7 and C5 (Fi ure S4B) and the more upfield T3/T7 H1′, H2′, and H2″ su ar proton chemical shifts (Table S3). Stabilizin hydrophobic interactions were also found between the methyl roup of T3 or T7 and the 2′-methylene roups of its two precedin residues (Fi ure 3C). The backbone and lycosidic torsion an les and pseudorotation phase an les are summarized in Fi ure S5.

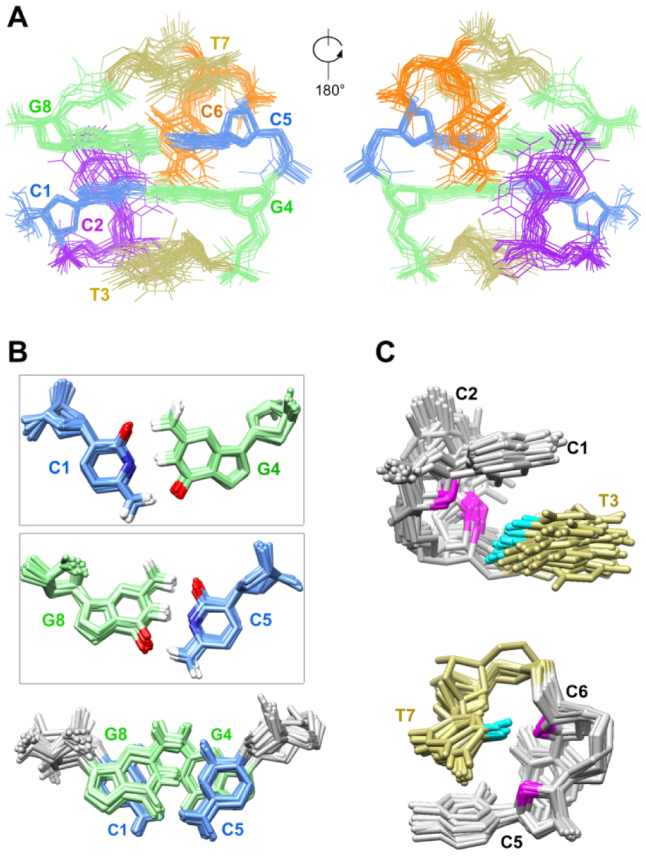



**Figure 3.** MDB structure of $(CCTG)_2$. (A) The major and minor roove views of 20 superimposed structures of $(CCTG)_2$. The third loop residues T3 and T7 stack on the C1-G4 and C5-G8 loop-closin base pairs, while the second loop residues C2 and C6 fold into the minor roove and pair up with multiple eometries. (B) C1-G4 and C5-G8 form the Watson–Crick loop-closin base pairs (top) and stack extensively with each other (bottom). (C) Hydrophobic interactions were observed between T3/T7 methyl (cyan) and the 2′-methylene roups (ma enta) of its two precedin residues.

**Exchangeable Pairing Geometries of C·C Mismatch in the Minor Groove of CCTG MDB.** In CCTG MDB, the second loop residues C2 and C6 were also folded into the minor roove. However, instead of stackin with each other, C2 and C6 were found to pair up with six different eometries via hydro en bond(s) and/or $Na^+$-mediated electrostatic interaction(s) (Fi ure 4A). These include (i) 13 cases showin a C2 O2-C6 H41 hydro en bond with C2 O2/N3···$Na^+$···C6 O2/N3 electrostatic interactions, (ii) two showin a C2 O2-C6 H41 hydro en bond without $Na^+$-mediated electrostatic interaction, (iii) two showin only C2 O2/N3···$Na^+$···C6 O2/N3 electrostatic interactions, (iv) one showin a C2 H41-C6 O2 hydro en bond, (v) one showin a C2 H41-C6 N3 hydro en bond, and (vi) one showin C2 N3-C6 H41 and C2 H41-C6 N3 hydro en bonds. As su ested by the seriously broadened C2 and C6 H6 si nals (Fi ure S4C), there is conformational exchan e amon these pairin eometries. Owin to these multiple C2·C6 pairin eometries, a relatively lar er avera e pairwise RMSD of 1.06 ± 0.28 Å for all residues was obtained (Table 1).

The exchan e between different pairin eometries can occur via hydro en bond(s) breakin /formin and/or the addition/removal of $Na^+$-mediated electrostatic interaction(s). It is reasonable that we observed more sin le hydro en bond pairin eometries in the refined structures, as they are involved in the conformational exchan e pathways between the one containin two hydro en bonds and the one containin no hydro en bonds. For the pairin eometry with a C2 O2-C6 H41 hydro en bond and C2 O2/N3···$Na^+$···C6 O2/N3 electrostatic interactions (Fi ure 4A, i), it occurs more frequently than the others probably because it is involved in more conformational exchan e pathways. This pairin eometry can be formed from the one with only a C2 O2-C6 H41 hydro en bond (Fi ure 4A, ii) by simply ainin $Na^+$-mediated electrostatic interactions or from the one with C2 O2/N3···$Na^+$···C6 O2/N3 electrostatic interactions (Fi ure 4A, iii) by formin a C2 O2-C6 H41 hydro en bond.

In addition to the pairin interactions between C2 and C6, hydro en bondin interactions were also found between C2/C6 and the loop-closin base pairs or the phosphodiester backbone to assist the foldin of C2 and C6 into the minor roove in CCTG MDB. For the most frequently observed C2·C6 pairin eometry which shows a C2 O2-C6 H41 hydro en bond and $Na^+$-mediated electrostatic interactions (Fi ure 4A, i), two to three hydro en bonds were usually formed with the loop-closin base pairs via C2 O2-G4/G8 H22, C2 N3-G8 H22 (Fi ure 4B, left), and/or C6 H42-G4 N3 (Fi ure 4C, left), and two to three hydro en bonds with the phosphodiester backbone via C2 H41/H42-G8/OP1/OP2/O4′/O5′ (Fi ure 4B, ri ht) and/or C6 H42-T3 OP1 (Fi ure 4C, ri ht), indicatin both of the minor roove residues are capable of formin hydro en bonds with the loop-closin base pairs and backbone. For the C2·C6 pairin eometry with two symmetric hydro en bonds (Fi ure 4A, vi), three additional hydro en


DOI: 10.1021/jacs.6b06897
J. Am. Chem. Soc. 2016, 138, 12534–12540

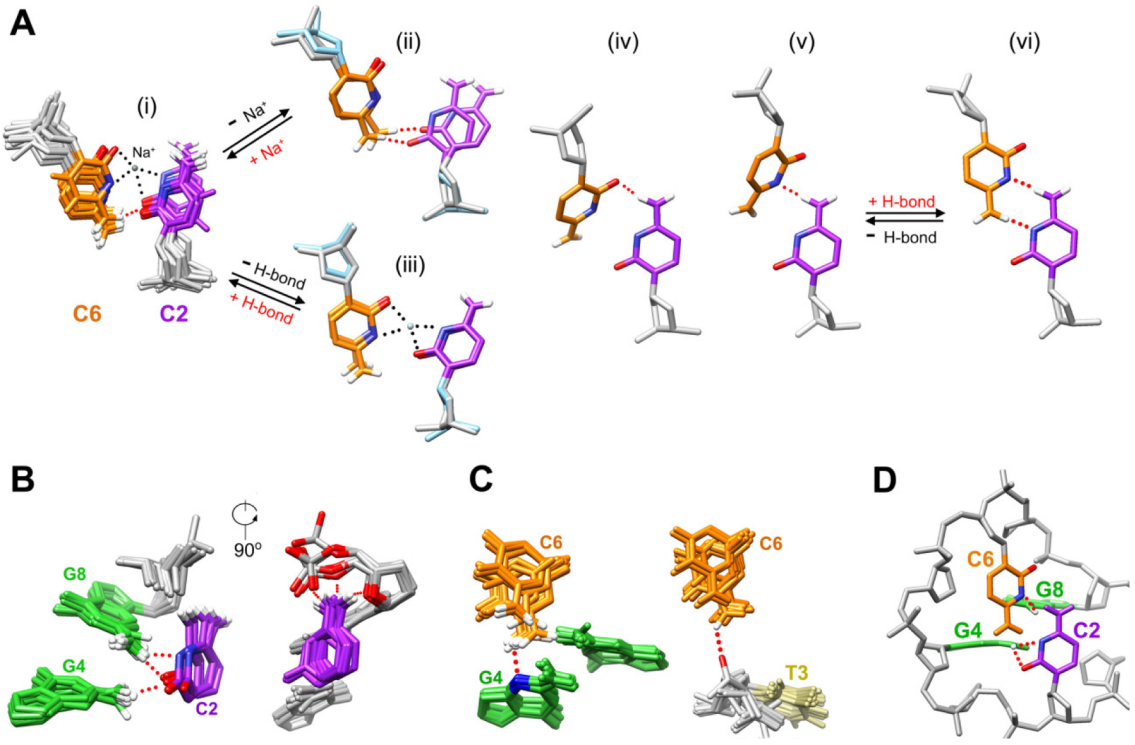



**Figure 4.** Multiple C2·C6 pairin eometries in CCTG MDB. (A) Six C2·C6 pairin eometries were observed, includin (i) a C2 O2-C6 H41 hydro en bond and $Na^+$-mediated C2 O2/N3···$Na^+$···C6 O2/N3 electrostatic interactions, (ii) C2 O2-C6 H41 hydro en bond, (iii) $Na^+$-mediated C2 O2/N3···$Na^+$···C6 O2/N3 electrostatic interactions, (iv) C2 H41-C6 O2 hydro en bond, (v) C2 H41-C6 N3 hydro en bond, and (vi) two symmetric C2 N3-C6 H41 and C2 H41-C6 N3 hydro en bonds. These pairin eometries are interconvertible by simply breakin /formin hydro en bond(s) and/or losin / ainin $Na^+$-mediated electrostatic interactions. Arrows were added between the pairin eometries that differed by one hydro en bond or the presence/absence of $Na^+$-mediated electrostatic interactions. In the pairin eometry with C2 O2-C6 H41 hydro en bond and $Na^+$-mediated electrostatic interactions, (B) C2 and (C) C6 form additional hydro en bonds with the loop-closin base pair and the phosphodiester backbone. (D) In the pairin eometry with two symmetric hydro en bonds, C2 and C6 can only form hydro en bonds with the loop-closin base pairs.

bonds were formed with the loop-closin base pairs via C2 O2/N3-G4 H22 and C6 N3-G8 H22 but not the backbone (Fi ure 4D).

**Loop−Loop Interactions in TTTA and CCTG MDBs.** The presence of extensive stabilizin interactions makes the foldin of an 8-nt DNA strand into a hi hly compact MDB structure feasible. In TTTA and CCTG MDBs, these stabilizin interactions include (i) 3′-5′ terminal stackin between the two loop-closin base pairs, (ii) stackin between the third loop residues with the loop-closin base pairs, (iii) hydrophobic interactions between the third loop residues with their two precedin residues, (iv) base−base stackin and/or pairin interactions between the two minor roove residues, and (v) hydro en bonds between the minor roove residue with the loop-closin base pair/phosphodiester backbone. Amon them, there are extensive loop−loop interactions overnin the MDB structure. These loop−loop interactions are absent in the lar er dumbbell structure.[6] In TTTA MDB, T1 and A4 in the first loop shows extensive base−base stackin with A8 and T5, respectively, and the minor roove T2 in the first loop stacks with T6 and forms hydro en bonds with T5, T6, and T7 (Fi ure 5A). In CCTG MDB, in addition to the stackin between the two loop-closin base pairs, C2 forms hydro en bonds with C6 and G8 in the second loop, and C6 forms hydro en bonds with T3 and G4 in the first loop (Fi ure 5B).

Loop−loop interactions have been shown to be biolo ically important in nucleic acids. In RNA, loop−loop interactions are involved in the formation of a kissin complex which serves as

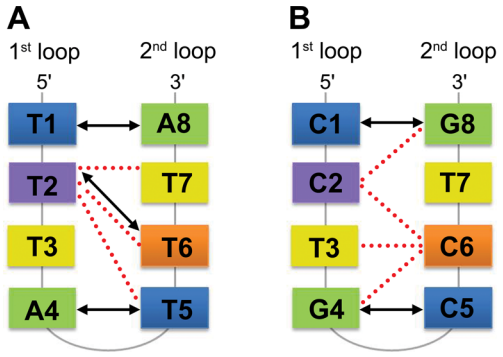



**Figure 5.** Loop−loop interactions in TTTA and CCTG MDBs. (A) The loop−loop interactions observed in TTTA MDB include (i) the stackin between T1-A4 and T5-A8 loop closin base pairs, (ii) hydro en bonds between T2 and T5/T6/T7, and (iii) base−base stackin between T2 with T6. (B) For CCTG MDB, these include (i) the stackin between C1-G4 and C5-G8 base pairs, (ii) hydro en bonds between C2 and C6/G8, and (iii) hydro en bonds between C6 and T3/G4. The stackin and hydro en bond interactions are represented by black arrows and red dotted lines, respectively.

an intermediate step in the dimerization of the RNA enomes of the human immunodeficiency virus[37] and the hepatitis C virus.[38] In DNA, loop−loop interactions have been shown to participate in CAG and CTG repeat expansion mutations.[39] More efficient mismatch repair escape was observed in the presence of adjacent CAG/CTG slip-outs than sin le slip-out,


DOI: 10.1021/jacs.6b06897
J. Am. Chem. Soc. 2016, 138, 12534−12540

revealin the si nificance of loop−loop interactions between adjacent slip-outs.[40] Recently, loop−loop interactions in DNA−DNA kissin complexes have been used in the nanotechnolo y area to construct tetrahedrons.[41] In TTTA and CCTG MDBs, these extensive loop−loop interactions play a crucial role in maintainin the structures, providin insi hts into the under-lyin chemical forces which brin about strand slippa e in TTTA and CCTG repeats durin DNA replication.

**Biological Significance of MDBs.** The formation of MDB in the nascent strand durin DNA replication can lead to variable sizes of repeat expansions.[21,22] First, MDB can be formed via a slippa e of two TTTA or CCTG repeats in the nascent strand. This provides a possible pathway for the occurrence of two-repeat expansion. Second, it has been shown that two competin MDBs can be formed in a se ment of three repeats, includin one with a 5′-overhan in repeat and one with a 3′-overhan in repeat.[21,22] Fast exchan e between these two MDBs results in the formation of a miniloop, which can lead to one-repeat expansion. Third, coexistence of multiple MDBs and/or miniloops can also occur in the nascent strand, resultin in three-repeat or lar er size expansion.

For repeat expansions to occur, the above MDBs and miniloops formed in the nascent strand must escape from DNA repair. To achieve this, conformational exchan e between the MDBs and/or miniloops can take place, which provides a potential pathway to avoid the specific reco nition by DNA repair proteins. For TTTA and CCTG MDBs, their reported meltin temperatures are 18.1 °C and 22.2 °C, respectively.[21,22] These temperatures reveal an optimized thermodynamic stability for feasible formation of MDBs and occurrence of conformational exchan e, thus brin in about TTTA and CCTG repeat expansions durin DNA replication.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supportin Information is available free of char e on the ACS Publications website at DOI: 10.1021/jacs.6b06897.

Materials and methods section; additional experimental data (Tables S1−S3 and Fi ures S1−S6) (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

*lams@cuhk.edu.hk


### Notes

The authors declare no competin financial interest.

## ■ ACKNOWLEDGMENTS

We thank Professor H. N. C. Won for his continuin support of the research activities of our roup. The work described in this paper was supported by General Research Fund (project nos. CUHK14302114 and CUHK14302915) and Special Equipment Grant (project no. SEG/CUHK09) from the Research Grants Council of the Hon Kon Special Administrative Re ion.

## ■ REFERENCES


(1) Watson, J. D.; Crick, F. H. *Nature* **1953**, *171*, 737.
(2) DiMaio, F.; Yu, X.; Rensen, E.; Krupovic, M.; Pran ishvili, D.; E elman, E. H. *Science* **2015**, *348*, 914.
(3) Pohl, F. M.; Jovin, T. M. *J. Mol. Biol.* **1972**, *67*, 375.
(4) Bikard, D.; Loot, C.; Baharo lu, Z.; Mazel, D. *Microbiol. Mol. Biol. Rev.* **2010**, *74*, 570.
(5) Chi, L. M.; Lam, S. L. *Nucleic Acids Res.* **2005**, *33*, 1604.
(6) Lam, S. L.; Wu, F.; Yan , H.; Chi, L. M. *Nucleic Acids Res.* **2011**, *39*, 6260.
(7) van Buuren, B. N.; Overmars, F. J.; Ippel, J. H.; Altona, C.; Wijmen a, S. S. *J. Mol. Biol.* **2000**, *304*, 371.
(8) Asensio, J. L.; Brown, T.; Lane, A. N. *Structure* **1999**, *7*, 1.
(9) Sakamoto, N.; Chastain, P. D.; Parniewski, P.; Ohshima, K.; Pandolfo, M.; Griffith, J. D.; Wells, R. D. *Mol. Cell* **1999**, *3*, 465.
(10) Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.* **2007**, *35*, 6517.
(11) Kovanda, A.; Zalar, M.; Sket, P.; Plavec, J.; Ro elj, B. *Sci. Rep.* **2015**, *5*, 17944.
(12) Kaushik, M.; Suehl, N.; Marky, L. A. *Biophys. Chem.* **2007**, *126*, 154.
(13) Brazda, V.; Laister, R. C.; Ja elska, E. B.; Arrowsmith, C. *BMC Mol. Biol.* **2011**, *12*, 33.
(14) Gacy, A. M.; Goellner, G. M.; Spiro, C.; Chen, X.; Gupta, G.; Bradbury, E. M.; Dyer, R. B.; Mikesell, M. J.; Yao, J. Z.; Johnson, A. J.; Richter, A.; Melancon, S. B.; McMurray, C. T. *Mol. Cell* **1998**, *1*, 583.
(15) Adrian, M.; An , D. J.; Lech, C. J.; Heddi, B.; Nicolas, A.; Phan, A. T. *J. Am. Chem. Soc.* **2014**, *136*, 6297.
(16) Buard, J.; Bourdet, A.; Yardley, J.; Dubrova, Y.; Jeffreys, A. J. *EMBO J.* **1998**, *17*, 3495.
(17) Mirkin, S. M. *Nature* **2007**, *447*, 932.
(18) Zhou, K.; Aertsen, A.; Michiels, C. W. *FEMS Microbiol. Rev.* **2014**, *38*, 119.
(19) Jones, M. R.; Seeman, N. C.; Mirkin, C. A. *Science* **2015**, *347*, 1260901.
(20) Choi, J.; Majima, T. *Chem. Soc. Rev.* **2011**, *40*, 5893.
(21) Guo, P.; Lam, S. L. *FEBS Lett.* **2015**, *589*, 1296.
(22) Guo, P.; Lam, S. L. *FEBS Lett.* **2015**, *589*, 3058.
(23) Blommers, M. J.; van de Ven, F. J.; van der Marel, G. A.; van Boom, J. H.; Hilbers, C. W. *Eur. J. Biochem.* **1991**, *201*, 33.
(24) Ippel, H. H.; van den Elst, H.; van der Marel, G. A.; van Boom, J. H.; Altona, C. *Biopolymers* **1998**, *46*, 375.
(25) Cho, S. H.; Chin, K. H.; Chen, C. W. *J. Biomol. NMR* **2001**, *19*, 33.
(26) Stott, K.; Stonehouse, J.; Keeler, J.; Hwan , T. L.; Shaka, A. J. *J. Am. Chem. Soc.* **1995**, *117*, 4199.
(27) Saen er, W. *Principles of Nucleic Acid Structure*; Sprin er: New York, 1984.
(28) Altona, C.; Sundaralin am, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205.
(29) Wijmen a, S. S.; van Buuren, B. N. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 287.
(30) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerlin , C.; Wan , B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.
(31) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Lau hton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817.
(32) Roe, D. R.; Cheatham, T. E., III. *J. Chem. Theory Comput.* **2013**, *9*, 3084.
(33) Pettersen, E. F.; Goddard, T. D.; Huan , C. C.; Couch, G. S.; Greenblatt, D. M.; Men , E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605.
(34) Blommers, M. J.; Walters, J. A.; Haasnoot, C. A.; Aelen, J. M.; van der Marel, G. A.; van Boom, J. H.; Hilbers, C. W. *Biochemistry* **1989**, *28*, 7491.
(35) Escaja, N.; Gomez-Pinto, I.; Rico, M.; Pedroso, E.; Gonzalez, C. *ChemBioChem* **2003**, *4*, 623.
(36) van Don en, M. J.; Wijmen a, S. S.; van der Marel, G. A.; van Boom, J. H.; Hilbers, C. W. *J. Mol. Biol.* **1996**, *263*, 715.
(37) Paillart, J. C.; Skripkin, E.; Ehresmann, B.; Ehresmann, C.; Marquet, R. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 5572.
(38) Shetty, S.; Kim, S.; Shimakami, T.; Lemon, S. M.; Mihailescu, M. R. *RNA* **2010**, *16*, 913.
(39) Pearson, C. E.; Wan , Y. H.; Griffith, J. D.; Sinden, R. R. *Nucleic Acids Res.* **1998**, *26*, 816.
(40) Pani rahi, G. B.; Slean, M. M.; Simard, J. P.; Gileadi, O.; Pearson, C. E. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 12593.




DOI: 10.1021/jacs.6b06897
J. Am. Chem. Soc. 2016, 138, 12534−12540

(41) Barth, A.; Kobbe, D.; Focke, M. *Nucleic Acids Res.* **2016**, *44*, 1502.



DOI: 10.1021/jacs.6b06897
*J. Am. Chem. Soc.* 2016, 138, 12534−12540