Supplementary Information

Minidumbbell Structures Formed by ATTCT Pentanucleotide Repeats in Spinocerebellar Ataxia Type 10

Pei Guo[1,*] and Sik Lok Lam[2†,*]

[1]School of Biology and Biological Engineering, South China University of Technology, Guangzhou, Guangdong 510006, China, [2]Department of Chemistry, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China. [†]In memory of Professor Sik Lok Lam.

*To whom correspondence should be addressed. Pei Guo. Email: peiguo@scut.edu.cn

Correspondence may also be addressed to Sik Lok Lam. Email: lams@cuhk.edu.hk

Table S1. A summary of the NMR restraints used for structure calculations of *(ATTCT)*$_2$.

Distribution of NOE-derived distance restraints (total = 284; inter/intra-residue = 169/115)

Residue	A1	T2	T3	C4	T5	A6	T7	T8	C9	T10
A1	14	21	1				1	2	5	3
T2		11	11	4		1				
T3			9	14		2		2		
C4				10	8	9				
T5					7	8				
A6						13	23	2		
T7							14	14	8	
T8								7	19	
C9									16	11
T10										14

Hydrogen bond restraints

Atom pair	Distance (Å)
A1 N1-T7 N3, A6 N1-T2 N3	2.72-2.92
A1 N6-T7 O4, A6 N6-T2 O4	2.85-3.05

Sugar, backbone and glycosidic torsion angle restraints

Residue	$^3J_{H1'-H2'}$ (Hz)	H1'-C1'-C2'-H2' torsion angle (°)	$^3J_{H4'-H5'}$ (Hz)	$^3J_{H4'-H5''}$ (Hz)	γ (°)	χ (°)
A1	[a]	-	2.8	3.3	30-90 (*gauche*$^+$)	90-330 (high *anti*)
T2	[a]	-	5.7	[b]	30-90 (*gauche*$^+$)	90-270 (*anti*)
T3	8.8	153 ± 15	[d]	[d]	-	90-270 (*anti*)
C4	[a]	-	[c]	[c]	-	90-270 (*anti*)
T5	[a]	-	[d]	[d]	-	90-270 (*anti*)
A6	8.4	150 ± 15	[d]	[d]	-	90-330 (high *anti*)
T7	[a]	-	[d]	[d]	-	90-270 (*anti*)
T8	9.0	154 ± 15	5.1	[b]	30-90 (*gauche*$^+$)	90-270 (*anti*)
C9	[a]	-	[c]	[c]	-	90-270 (*anti*)
T10	[a]	-	[d]	[d]	-	90-270 (*anti*)

[a] The coupling constant was not determined due to peak overlap.
[b] Due to weak coupling, the correlation peaks were not observed.
[c] The coupling constant was not determined due to peak broadening or weak coupling.
[d] The coupling constant was not determined due to peak overlap/broadening or weak coupling.

Chirality restraints for all residues[a]

C2', O4', N1, H1'	O3', C2', C4', H3'	C3', C5', O4', H4'
60-80°	60-80°	60-80°

[a] The chirality restraints were generated by AMBERTools. A total of 30 chirality restraints were obtained for *(ATTCT)*$_2$.

Table S2. Proton NMR chemical shifts (ppm) of *(ATTCT)$_2$*.[a]

Residue	H3	NH_2[b]	H6/ H8	H2/H5/ H7	H1'	H2'	H2''	H3'	H4'	H5'	H5''
A1	-	[c]	8.34	8.12	6.24	2.84	2.86	4.79	4.23	4.00	3.89
T2	13.36	-	7.47	1.39	6.19	2.14	2.38	4.93	4.41	4.23	4.11
T3	[d]	-	7.76	1.98	6.31	2.06	2.52	4.79	4.40	4.24	4.11
C4	-	7.25/6.53	7.63	5.69	5.69	1.83	2.28	4.61	3.48	4.03	4.01
T5	[d]	-	7.56	1.85	6.19	2.16	2.44	4.82	4.27	3.70	3.70
A6	-	[c]	8.49	8.18	6.37	2.88	2.96	4.86	4.47	4.19	4.19
T7	13.00	-	7.34	1.37	6.19	2.09	2.42	4.91	4.40	4.33	4.12
T8	[d]	-	7.74	1.96	6.32	2.17	2.58	4.81	4.38	4.22	4.14
C9	-	7.29/6.58	7.68	5.78	5.69	2.01	2.20	4.65	3.66	4.15	4.03
T10	[d]	-	7.52	1.73	6.20	2.20	2.28	4.48	4.07	3.89	3.89

[a] The chemical shifts of labile and non-labile protons were measured at 0 and 5 °C, respectively.
[b] The IUPAC nomenclature for the amino protons of A and C are H61/H62 and H41/H42, respectively.
[c] The chemical shifts could not be determined as these labile protons were too broad to be observed.
[d] The chemical shifts could not be determined as these H3 signals were overlapped.

Table S3. DNA sequences used for *in vitro* primer extension assays.

	Sequence
Template in *P1-T* and *P2-T*	5'-AGTCTG AGAAT AGAAT AGAAT AGAAT AGAAT AGAAT AGAAT AGAAT AGAAT AGAAT GCACTG C3s-3'
Primer in *P1-T*	5'-ATTCT ATTCT ATTCT ATTCT ATTCT-3'
Primer in *P2-T*	5'-GC ATTCT ATTCT ATTCT ATTCT ATTCT-3'
Ref-56nt in *P1-T*	5'-ATTCT ATTCT ATTCT ATTCT ATTCT ATTCT ATTCT ATTCT ATTCT ATTCT CAGACT-3'
Ref-58nt in *P2-T*	5'-GC ATTCT ATTCT ATTCT ATTCT ATTCT ATTCT ATTCT ATTCT ATTCT ATTCT CAGACT-3'
Template in *MDB+6bp*, *MDB+5bp*, *MDB+4bp* and *MDB+3bp*	5'-CAGTGCTGCGTGTCTGCATG-3'
Primer in *MDB+6bp*	5'-CATGCAGACACG ATTCTATTCT CAGCAC-3'
Primer in *MDB+5bp*	5'-CATGCAGACACG ATTCTATTCT CAGCA-3'
Primer in *MDB+4bp*	5'-CATGCAGACACG ATTCTATTCT CAGC-3'
Primer in *MDB+3bp*	5'-CATGCAGACACG ATTCTATTCT CAG-3'
Ref-30nt in *MDB+6bp*, *MDB+5bp*, *MDB+4bp* and *MDB+3bp*	5'-CATGCAGACACG ATTCTATTCT CAGCACTG-3'

Note: C3s is the three-carbon spacer (Glen Research) to avoid template extension or cleavage by KF.

Figure S1. Schematics of the TTTA MDB (left) (1), CCTG MDB (middle) (2) and CTTG MDB (right) (3). In each of the MDBs, the first (L1/L1') and the fourth loop residues (L4/L4') form loop-closing base pairs, the second loop residues (L2/L2') sit in the minor groove, and the third loop residues (L3/L3') stack on the loop-closing base pairs. In the TTTA MDB, the bases of L2 and L2' partially stack with each other. In the CCTG and CTTG MDBs, L2 and L2' form a C·C and T·T base pairs, respectively.

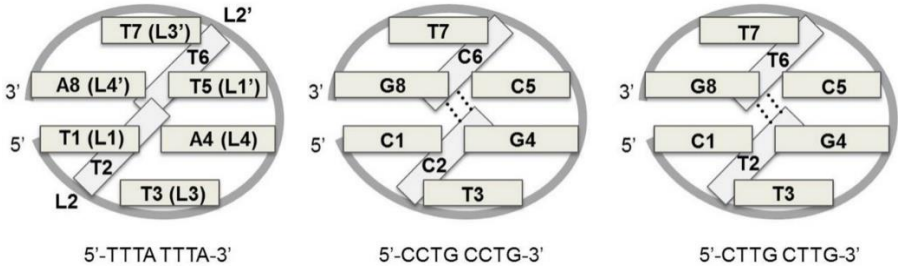


Figure S2. 1D ^{1}H NMR spectra show the aromatic and methyl proton resonance assignments of (A) *(ATTCT)$_2$*, *(ATTCT)$_3$*, *(ATTCT)$_4$*, and *(ATTCT)$_5$*, and (B) *(TTCTA)$_2$*. The spectra were acquired at 5 °C.

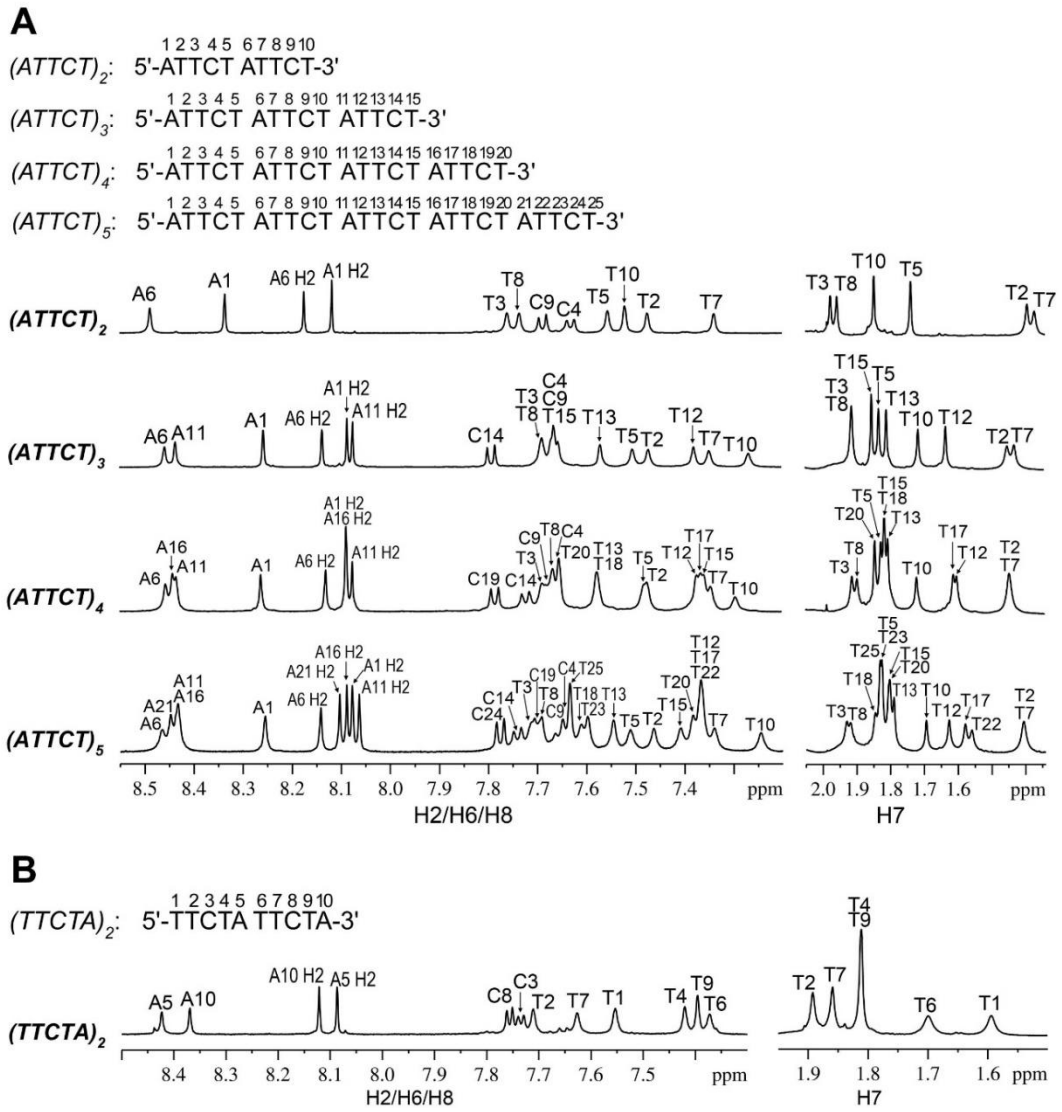


Figure S3. 1D ^{1}H NMR spectra show the aromatic and methyl proton resonance assignments of (A) *(ATTCT)$_2$A*, *(ATTCT)$_2$AT*, *(ATTCT)$_2$ATT*, and *(ATTCT)$_2$ATTC*, (B) *T(ATTCT)$_2$*, *CT(ATTCT)$_2$*, *TCT(ATTCT)$_2$*, *TTCT(ATTCT)$_2$*, and (C) *T(ATTCT)$_2$A*. The spectra were acquired at 5 °C.

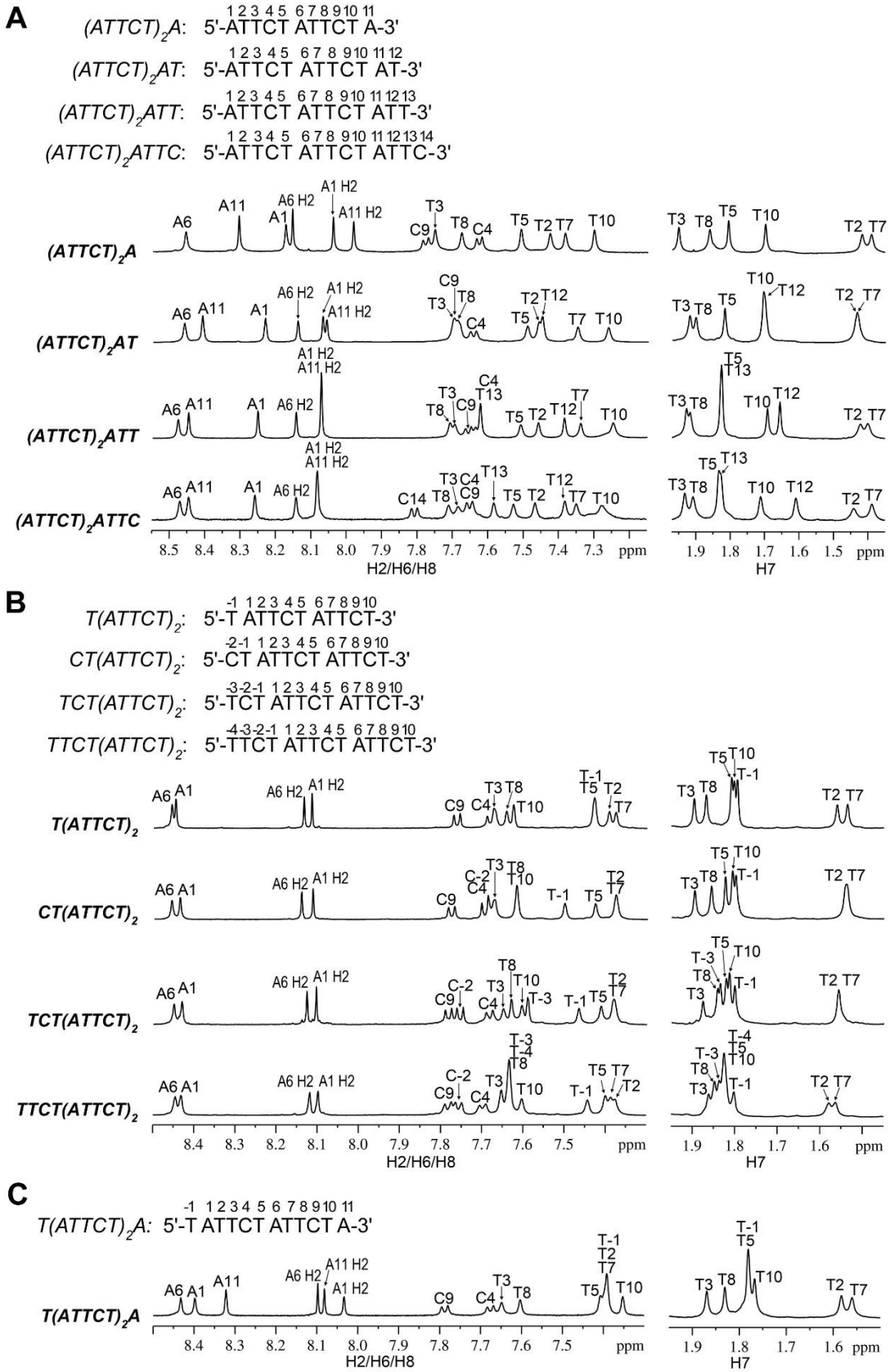


Figure S4. (A) The H6/H8-H1' sequential resonance assignment of *(ATTCT)$_2$* under 10 mM NaPi. The NOESY NMR spectrum was acquired at 5 °C with a mixing time of 300 ms. (B) 1D ^{1}H NMR spectra show the aromatic and methyl proton resonances of *(ATTCT)$_2$* under 10 mM NaPi (top), and addition of 150 mM NaCl (bottom) at 5 °C.

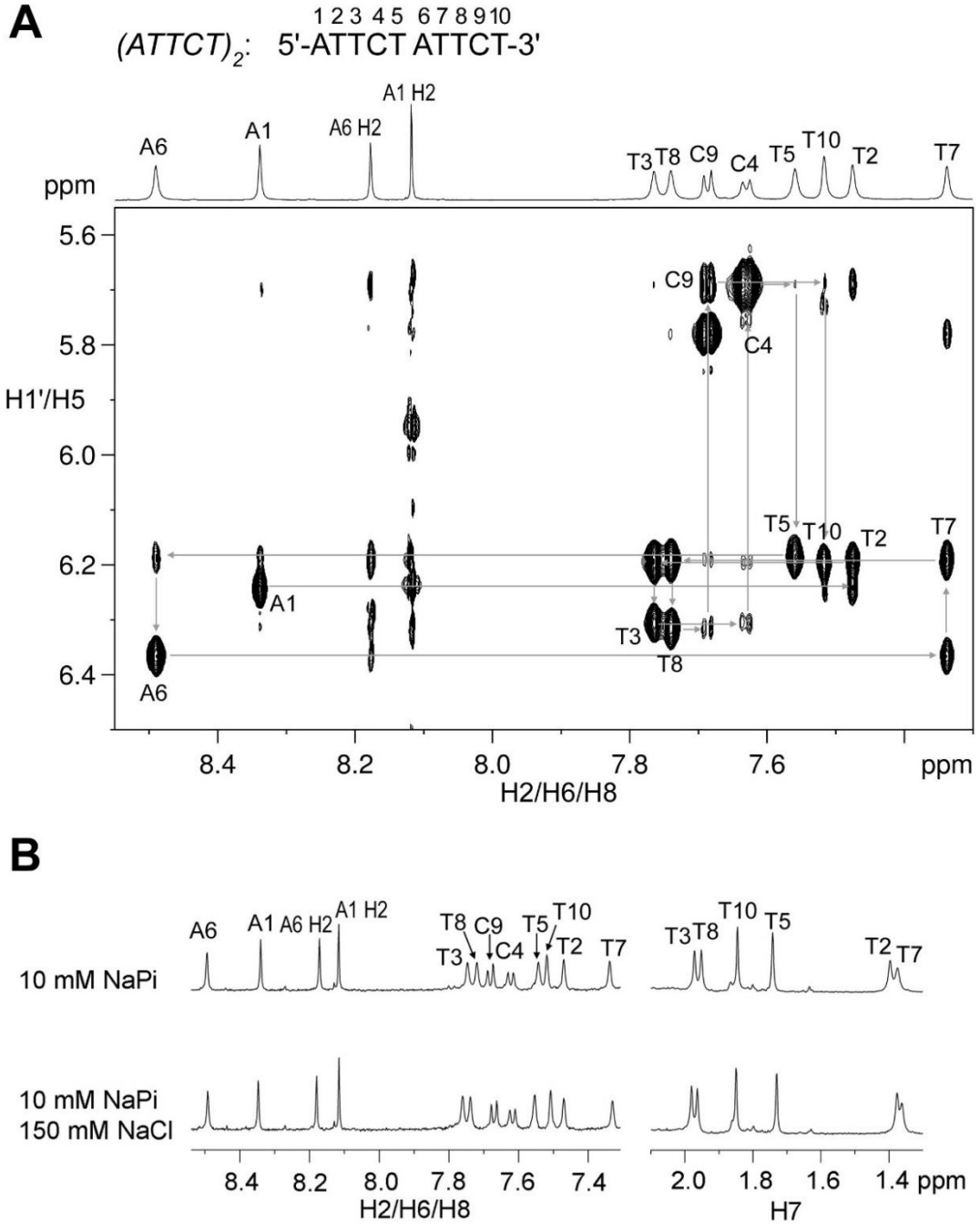


Figure S5. The H6/H8-H1' sequential resonance assignment of *(TTCTA)$_2$*. The NOESY spectrum was acquired at 0 °C with a mixing time of 600 ms.

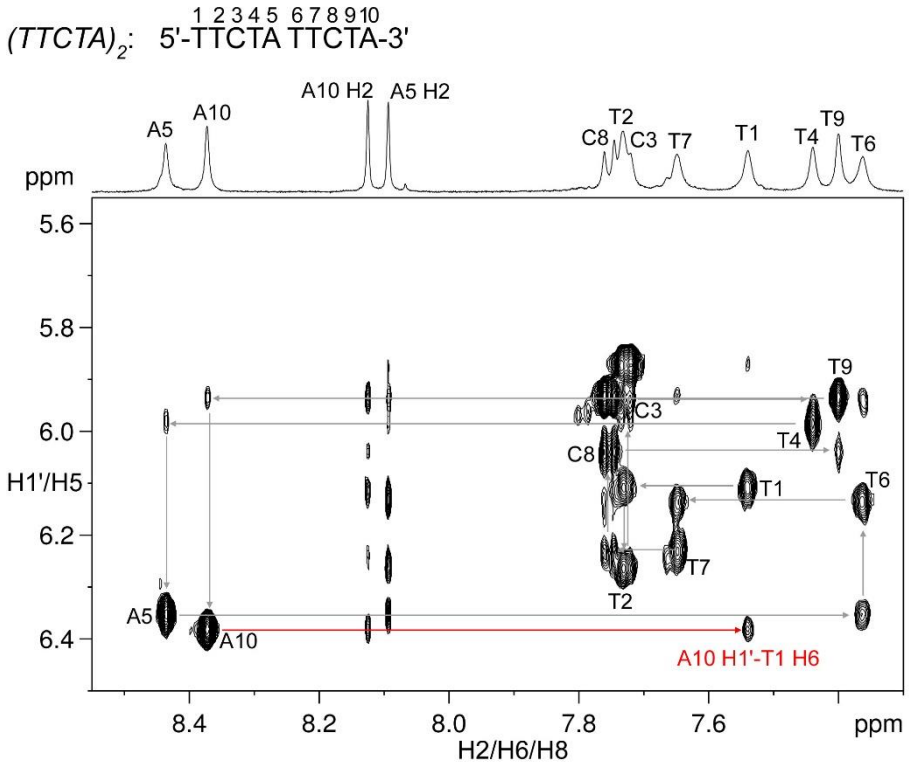


Figure S6. The H6/H8-H1' sequential resonance assignments of (A) *(ATTCT)$_3$* and (B) *(ATTCT)$_4$*. The NOESY spectra were acquired at 10 °C with a mixing time of 600 ms.

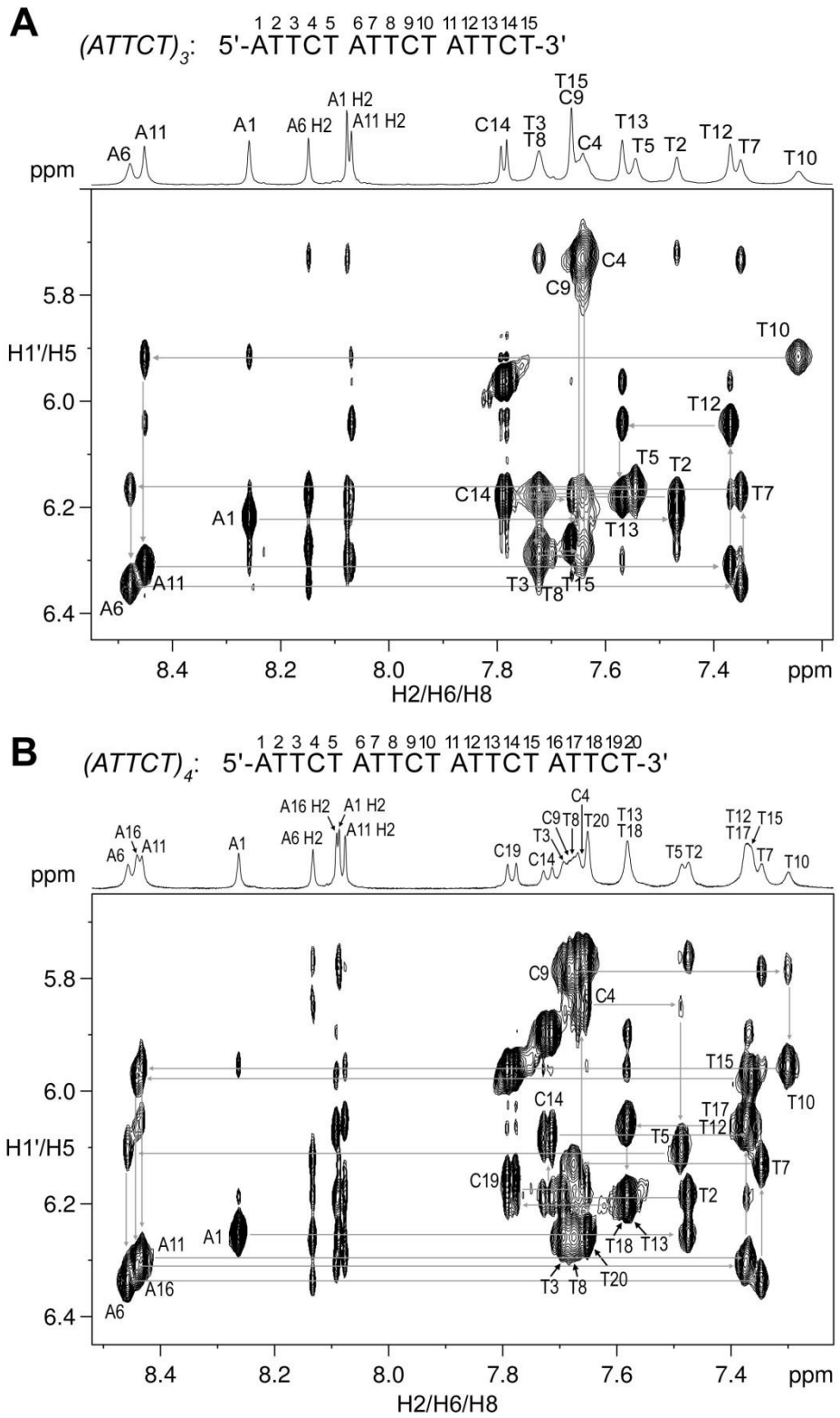


Figure S7. The H6/H8-H1' sequential resonance assignments of *(ATTCT)$_5$*. The NOESY spectrum was acquired at 15 °C with a mixing time of 600 ms.

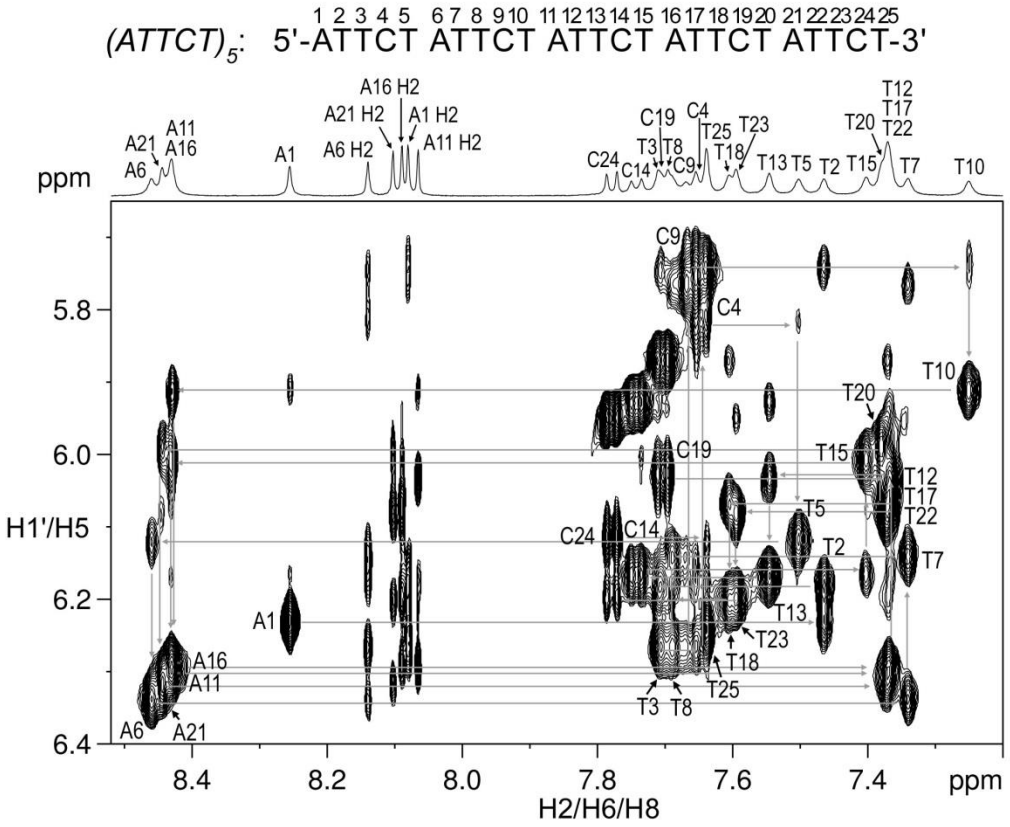


Figure S8. The H6/H8-H1' sequential resonance assignments of (A) *(ATTCT)$_2$A* and (B) *(ATTCT)$_2$AT*. The NOESY spectra were acquired at 0 °C with a mixing time of 300 ms for *(ATTCT)$_2$A*, and at 5 °C with a mixing time of 600 ms for *(ATTCT)$_2$AT*.

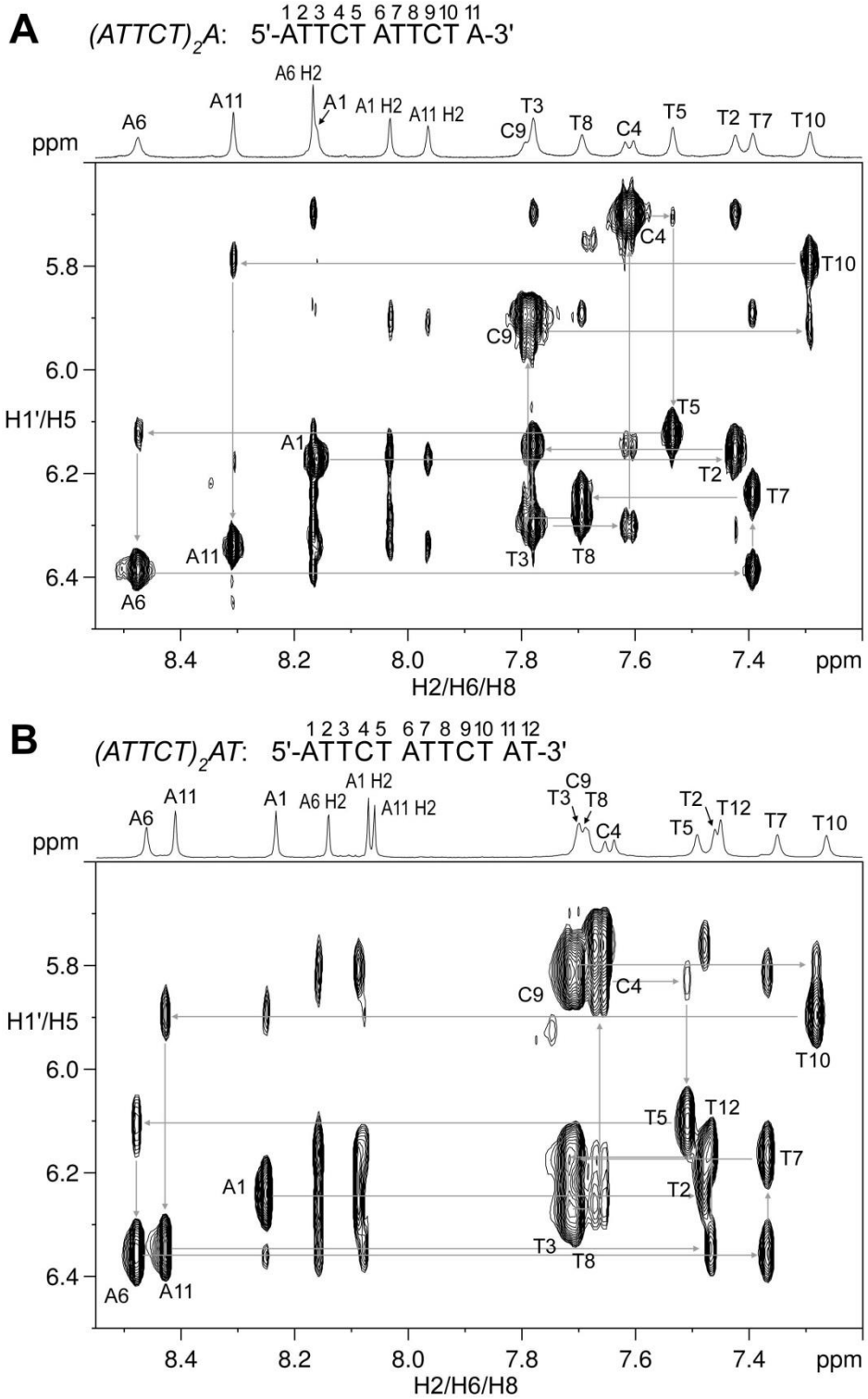


Figure S9. The H6/H8-H1' sequential resonance assignments of (A) *(ATTCT)$_2$ATT* and (B) *(ATTCT)$_2$ATTC*. The NOESY spectra were acquired at 0 °C with a mixing time of 300 ms.

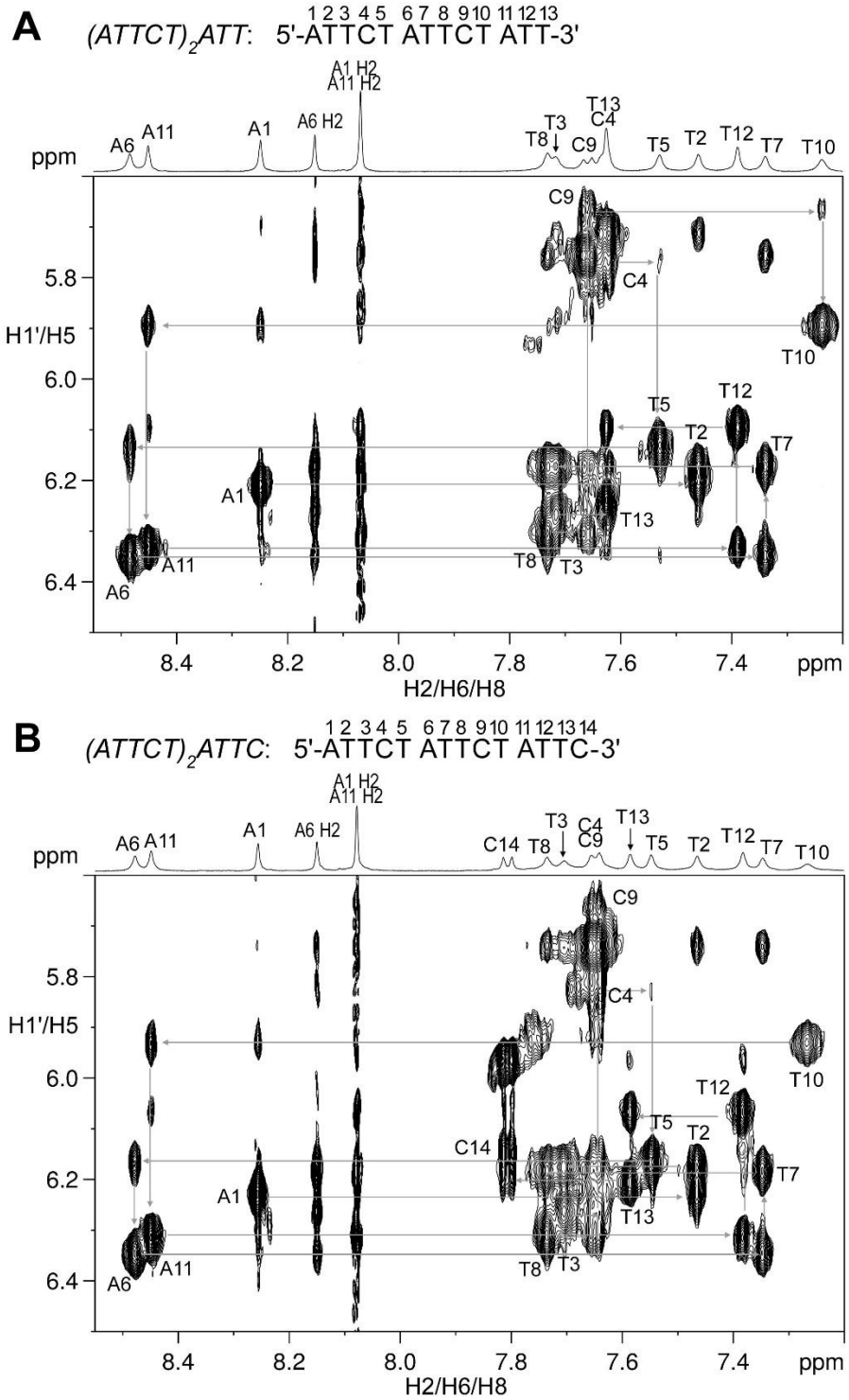


Figure S10. The H6/H8-H1' sequential resonance assignments of (A) *T(ATTCT)$_2$* and (B) *CT(ATTCT)$_2$*. The NOESY spectra were acquired at 0 °C with a mixing time of 600 ms.

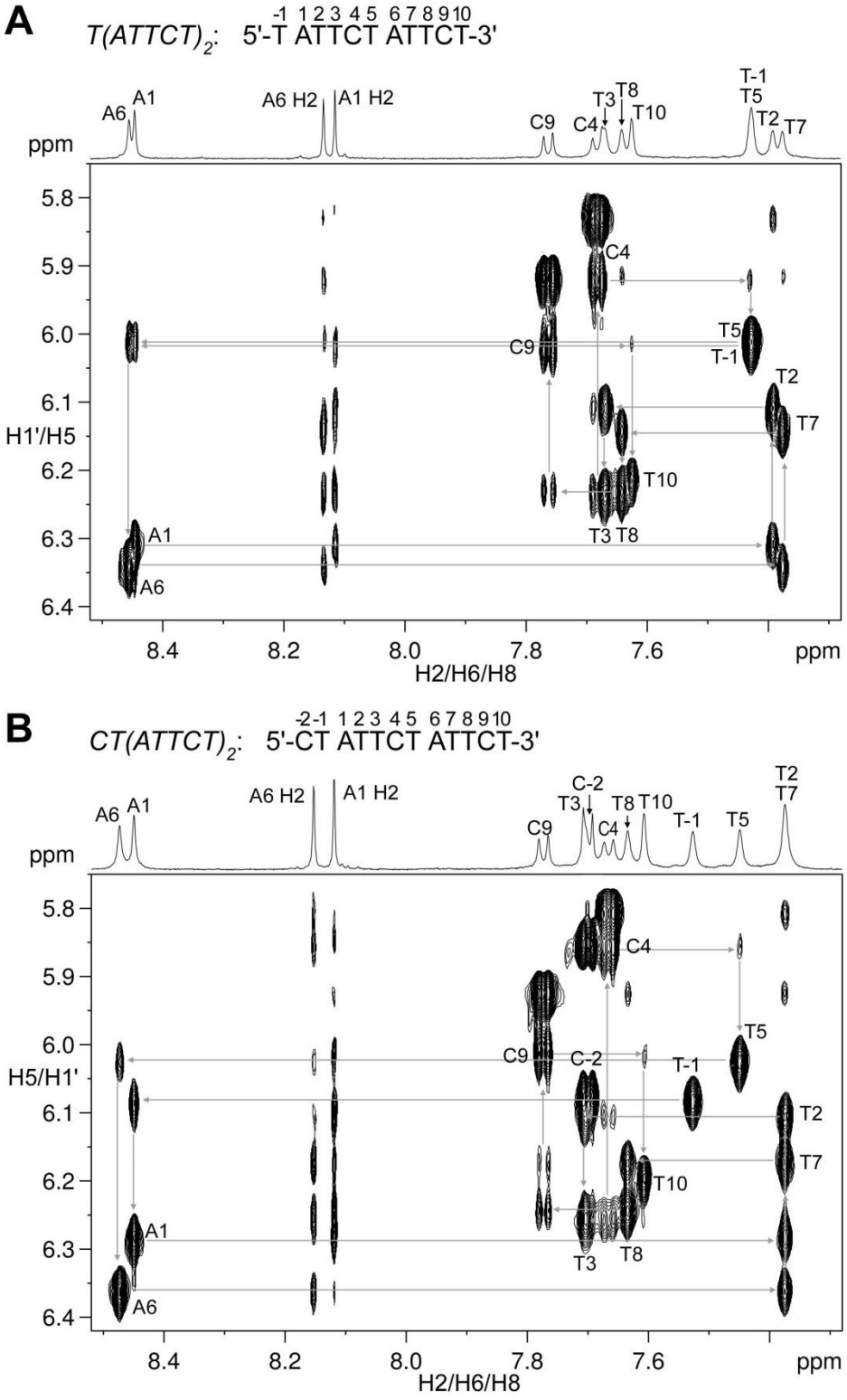


Figure S11. The H6/H8-H1' sequential resonance assignments of (A) *TCT(ATTCT)$_2$* and (B) *TTCT(ATTCT)$_2$*. The NOESY spectra were acquired at 0 °C with a mixing time of 600 ms.

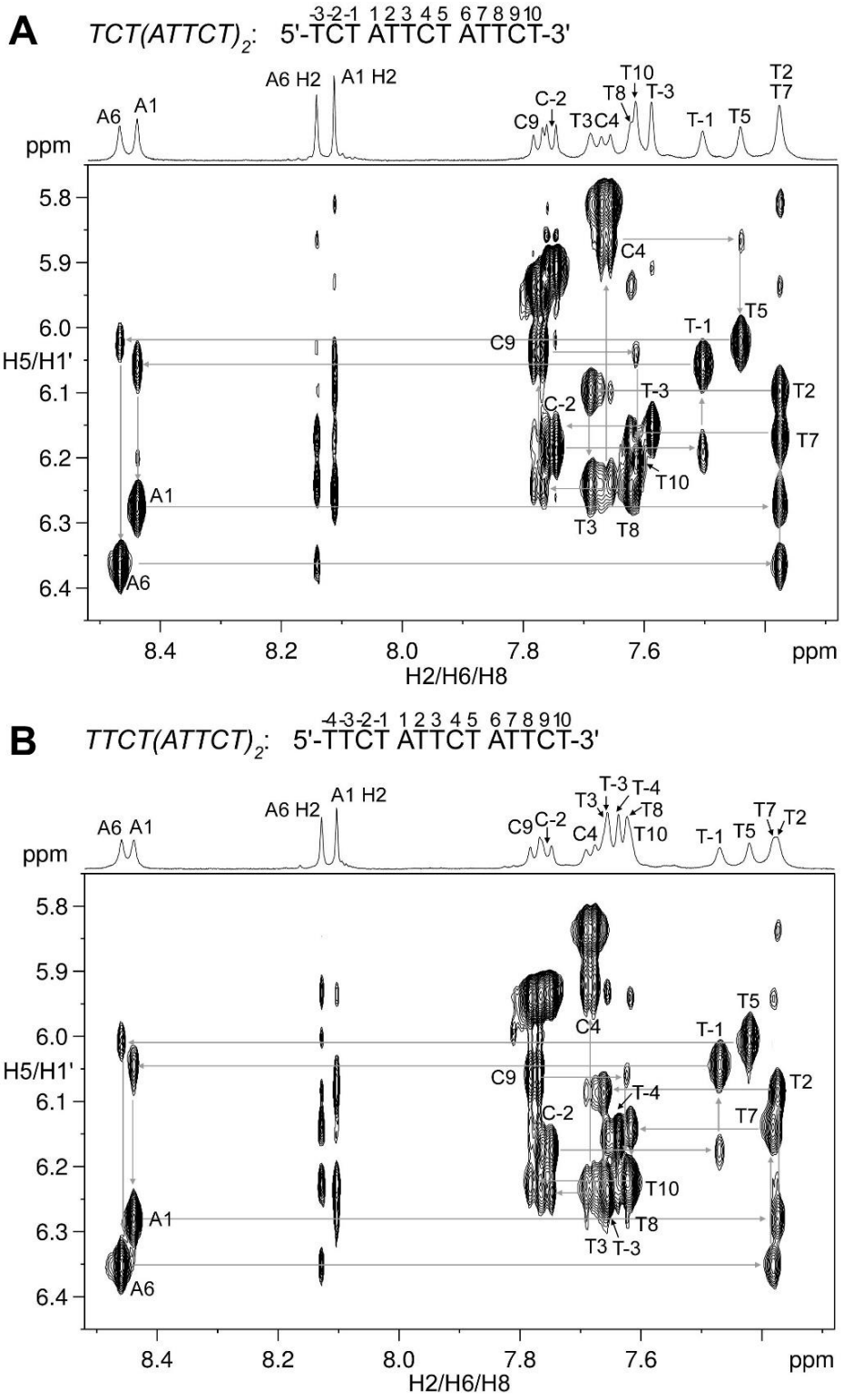


Figure S12. The H6/H8-H1' sequential resonance assignment of *T(ATTCT)$_2$A*. The NOESY spectrum was acquired at 0 °C with a mixing time of 600 ms.

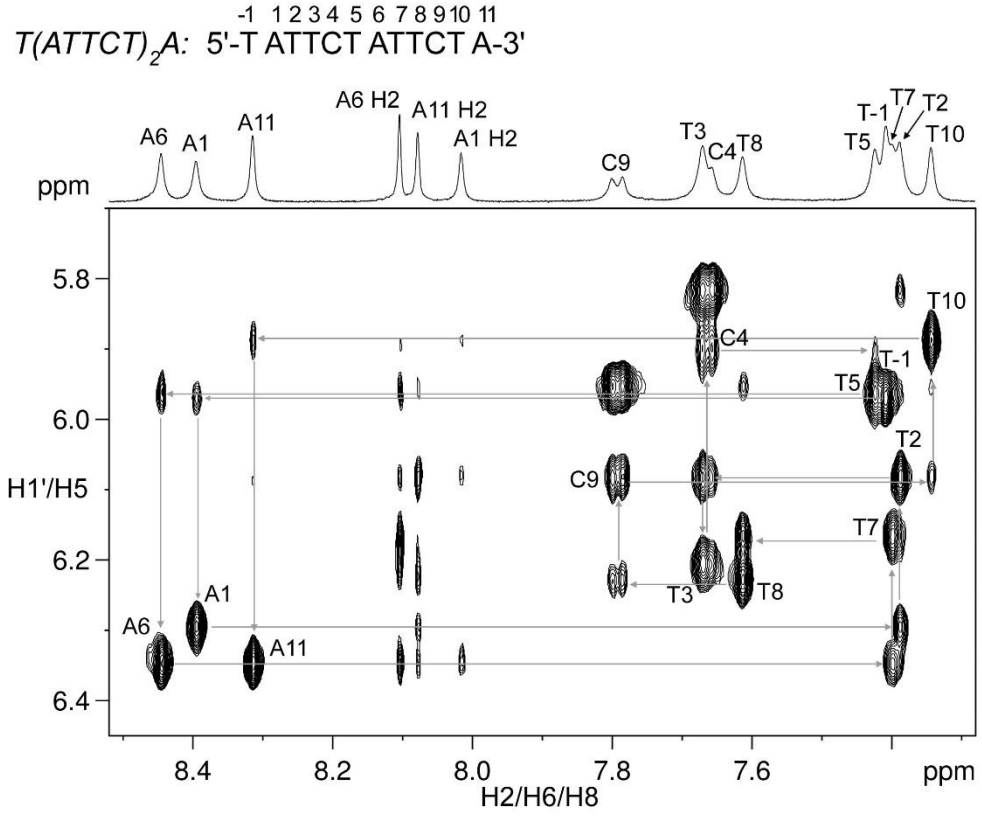


Figure S13. The ^{1}H-^{13}C HMBC spectra show the adenine H2 resonance assignments of (A) *(ATTCT)*$_2$ and (B) *(ATTCT)*$_3$. The spectra were acquired at 10 °C.

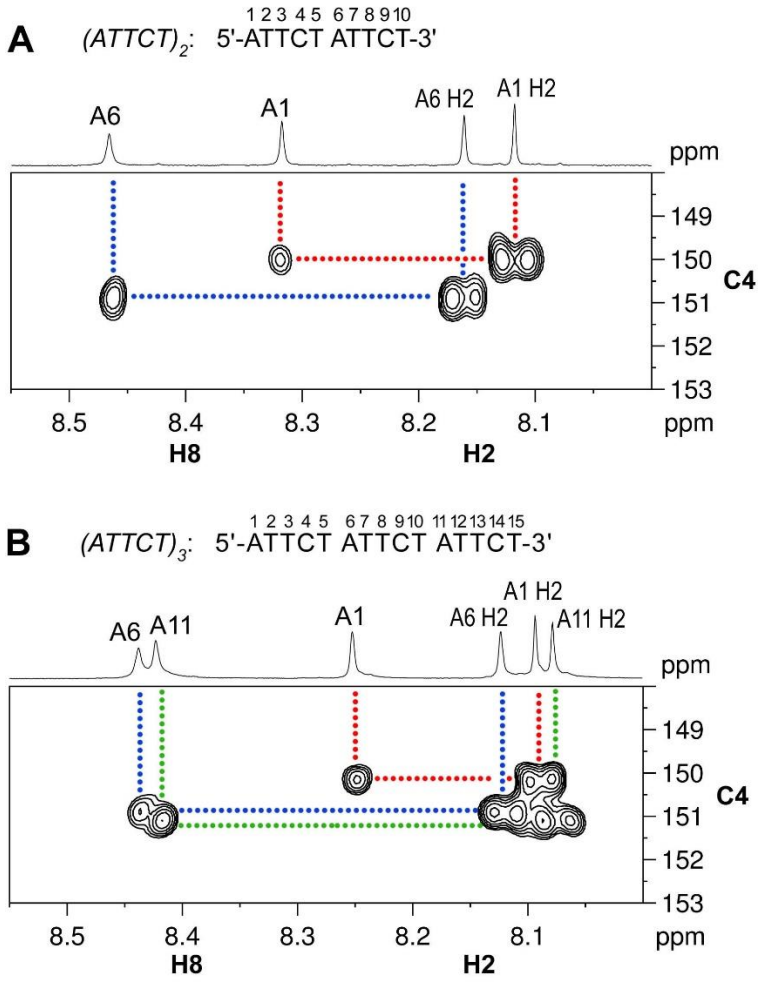


Figure S14. The ^{1}H-^{13}C HMBC spectra of (A) *(ATTCT)$_4$* and (B) *(ATTCT)$_5$* show that except for A1, the C4 signals of other adenine residues were overlapped, thus their H2 resonances could not be assigned using these HMBC spectra. (C-D) Alternatively, adenine H2 resonances were assigned based on intranucleotide H1'/H2'/H2''-H2 NOEs from the NOESY spectra. The spectra in (A), (C) and (D) were acquired at 10 °C, and the spectrum in (B) was acquired at 15 °C.

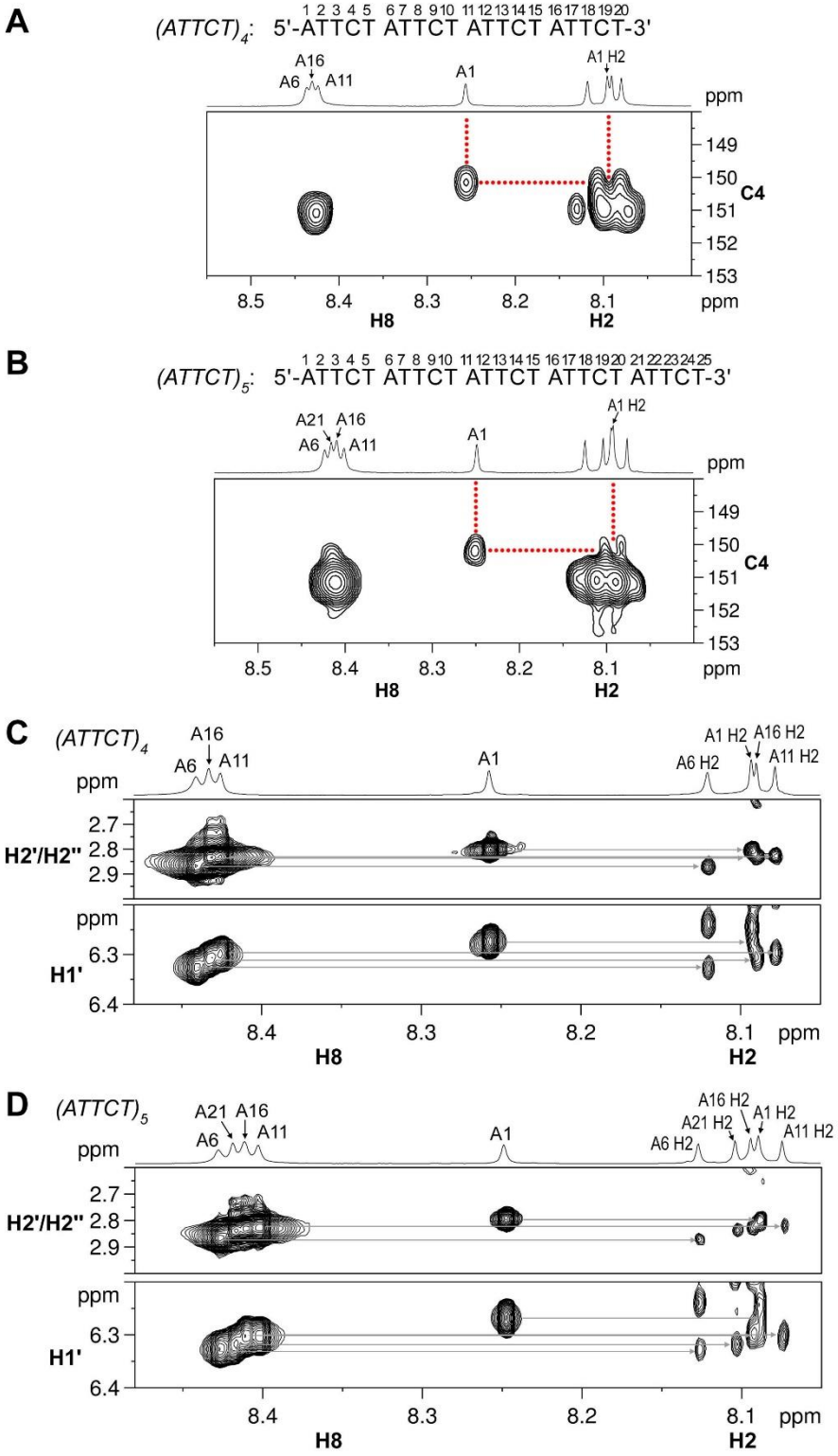


Figure S15. The ^{1}H-^{13}C HMBC spectra show the adenine H2 resonance assignments of (A) *(ATTCT)$_2$A*, (B) *(ATTCT)$_2$AT*, (C) *(ATTCT)$_2$ATT* and (D) *(ATTCT)$_2$ATTC*. The spectra in (A), (B), (C) and (D) were acquired at 10, 0, 15 and 15 °C, respectively.

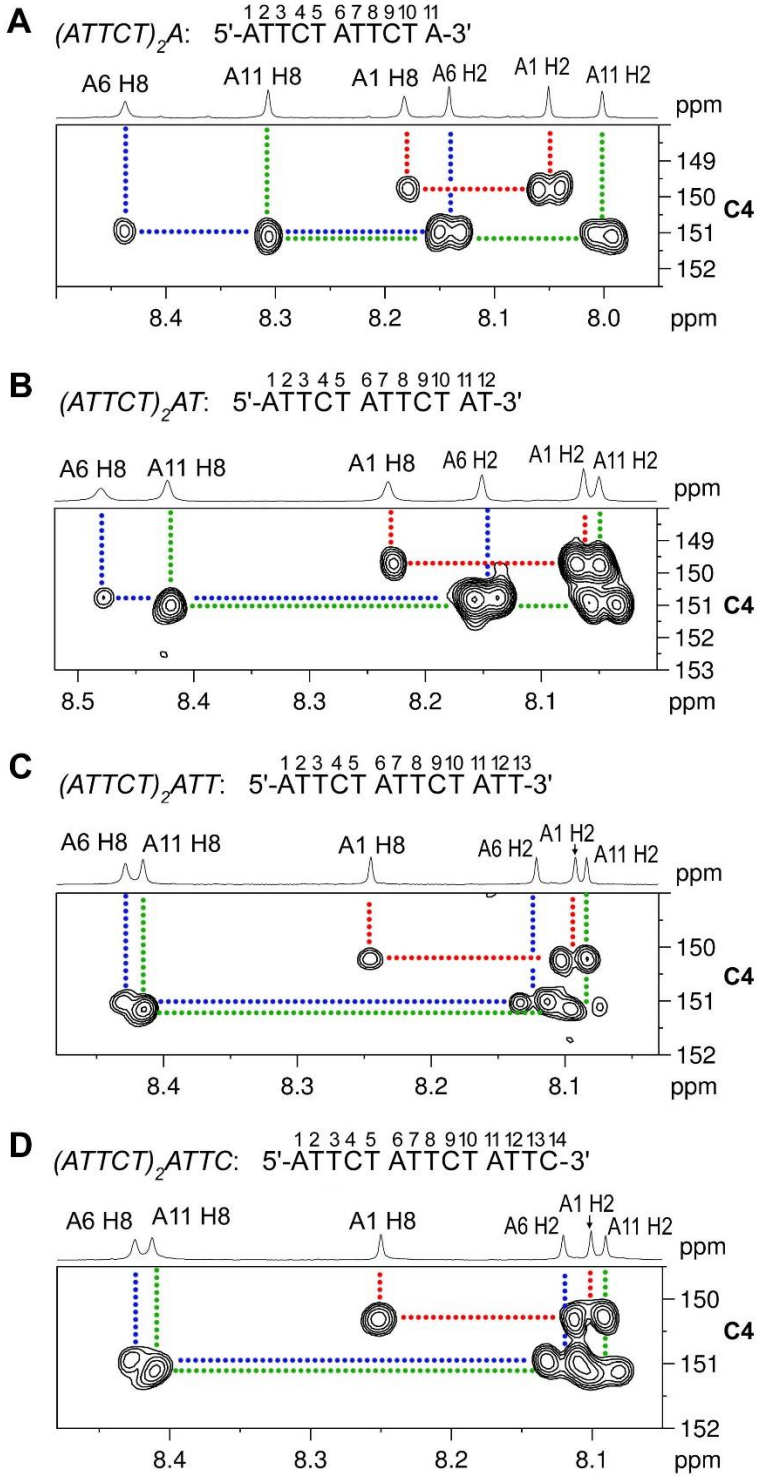


Figure S16. The ^{1}H-^{13}C HMBC spectra show the adenine H2 resonance assignments of (A) *T(ATTCT)*$_2$, (B) *CT(ATTCT)*$_2$, (C) *TCT(ATTCT)*$_2$, (D) *TTCT(ATTCT)*$_2$ at 5 °C and (E) *T(ATTCT)*$_2$*A* at 0 °C.

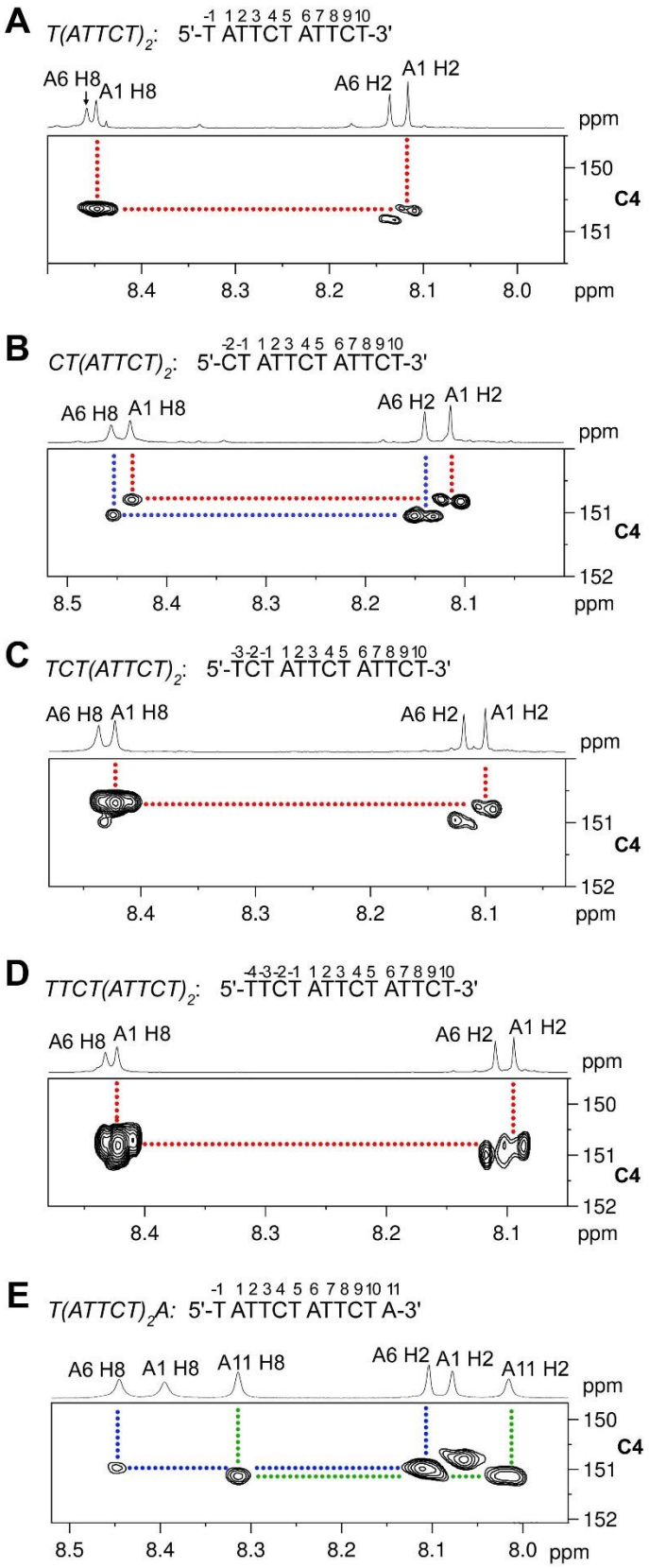


Figure S17. The TOCSY (top and middle) and ^{1}H-^{31}P HSQC (bottom) spectra show the ^{31}P resonance assignments of (A) $(ATTCT)_2$ and (B) $(ATTCT)_3$. The spectra for $(ATTCT)_2$ and $(ATTCT)_3$ were acquired at 5 °C and 10 °C, respectively.

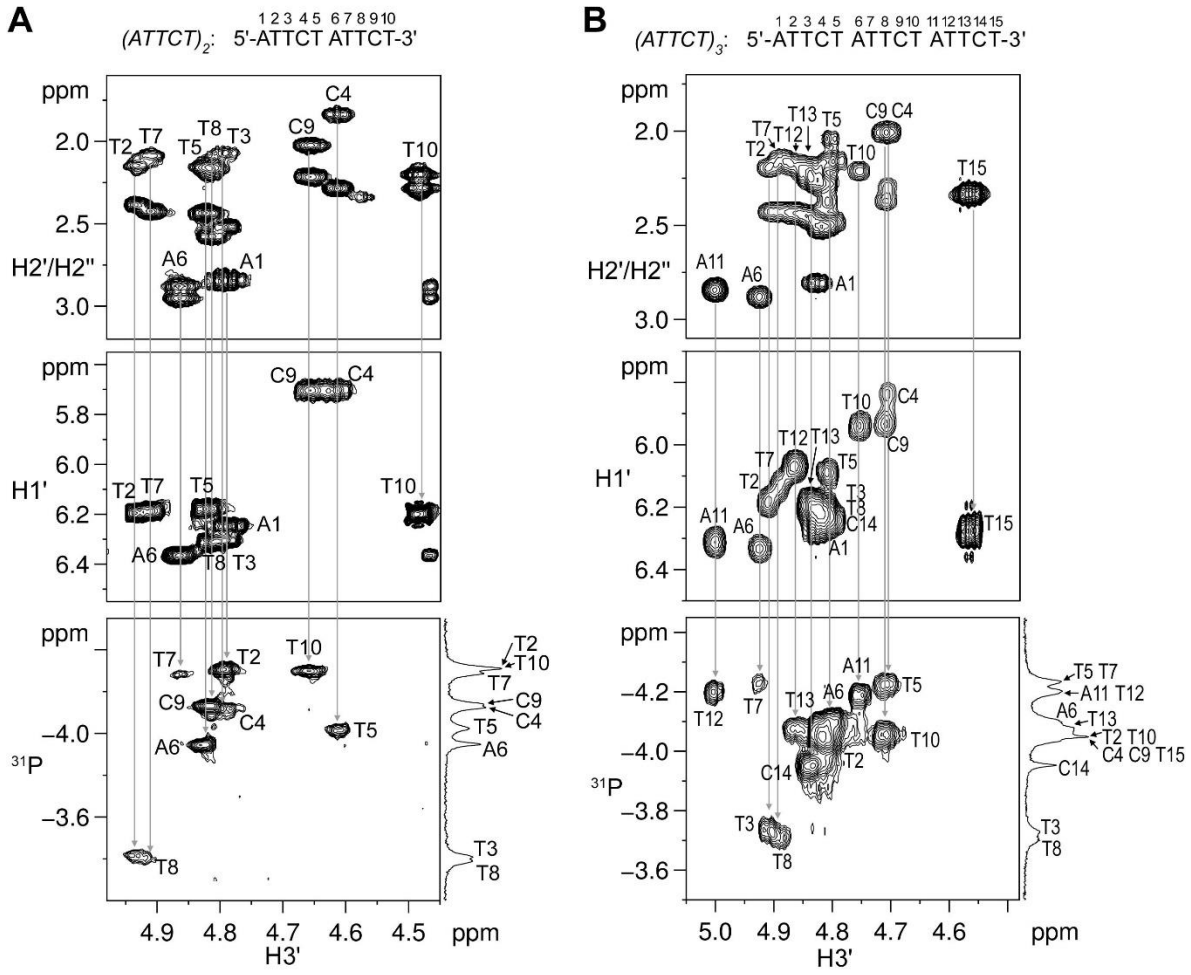


Figure S18. The TOCSY (top and middle) and ^{1}H-^{31}P HSQC (bottom) spectra show the ^{31}P resonance assignments of (A) *(ATTCT)$_4$* and (B) *(ATTCT)$_5$*. Due to peak overlaps, the ^{31}P resonances could only be partially assigned. The spectra for *(ATTCT)$_4$* and *(ATTCT)$_5$* were acquired at 10 °C and 15 °C, respectively.

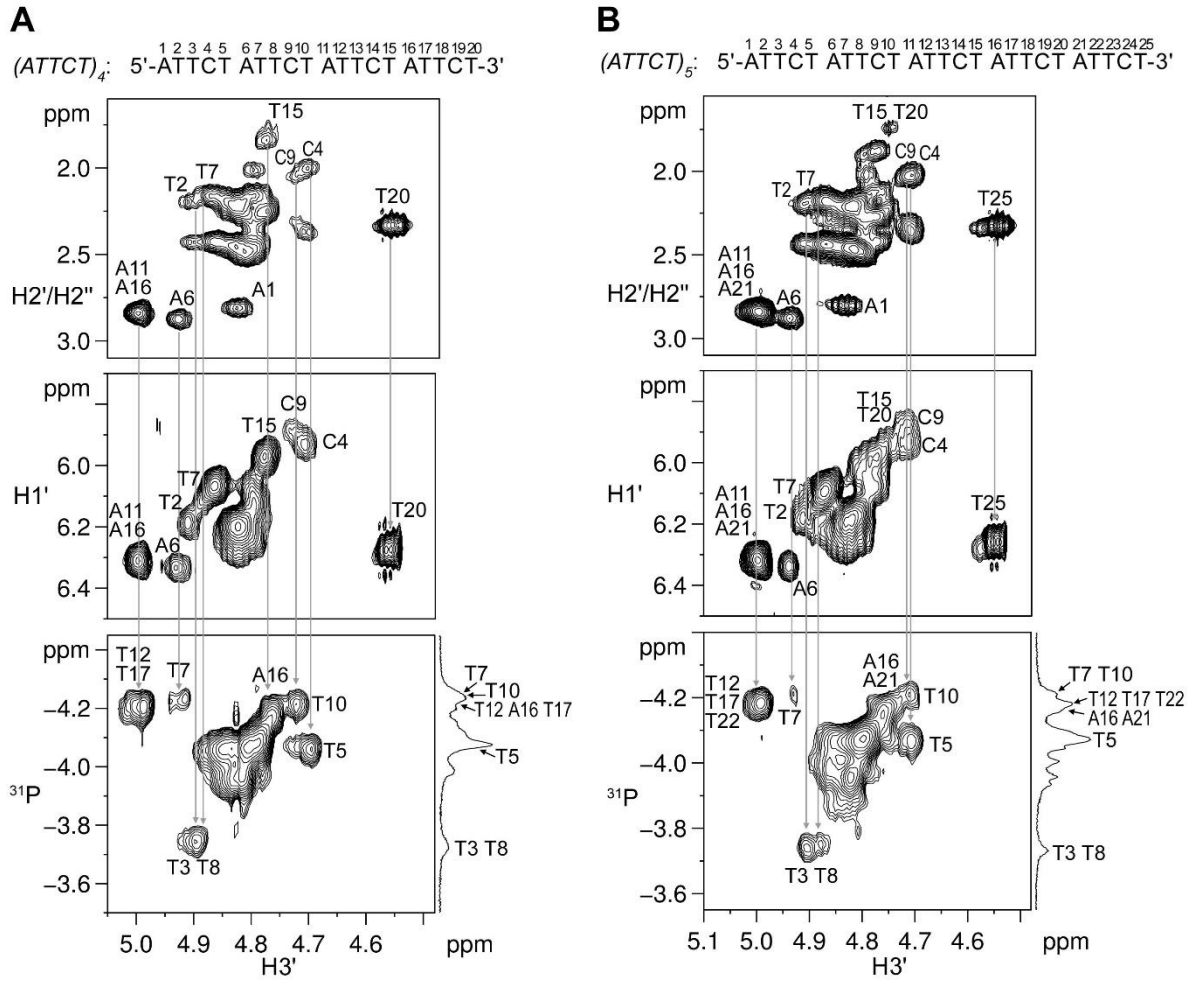


Figure S19. The TOCSY (top and middle) and ^{1}H-^{31}P HSQC (bottom) spectra show the ^{31}P resonance assignments of (A) *(ATTCT)$_2$A* and (B) *(ATTCT)$_2$AT*. The spectra of *(ATTCT)$_2$A* and *(ATTCT)$_2$AT* were acquired at 10 °C and 0 °C, respectively.

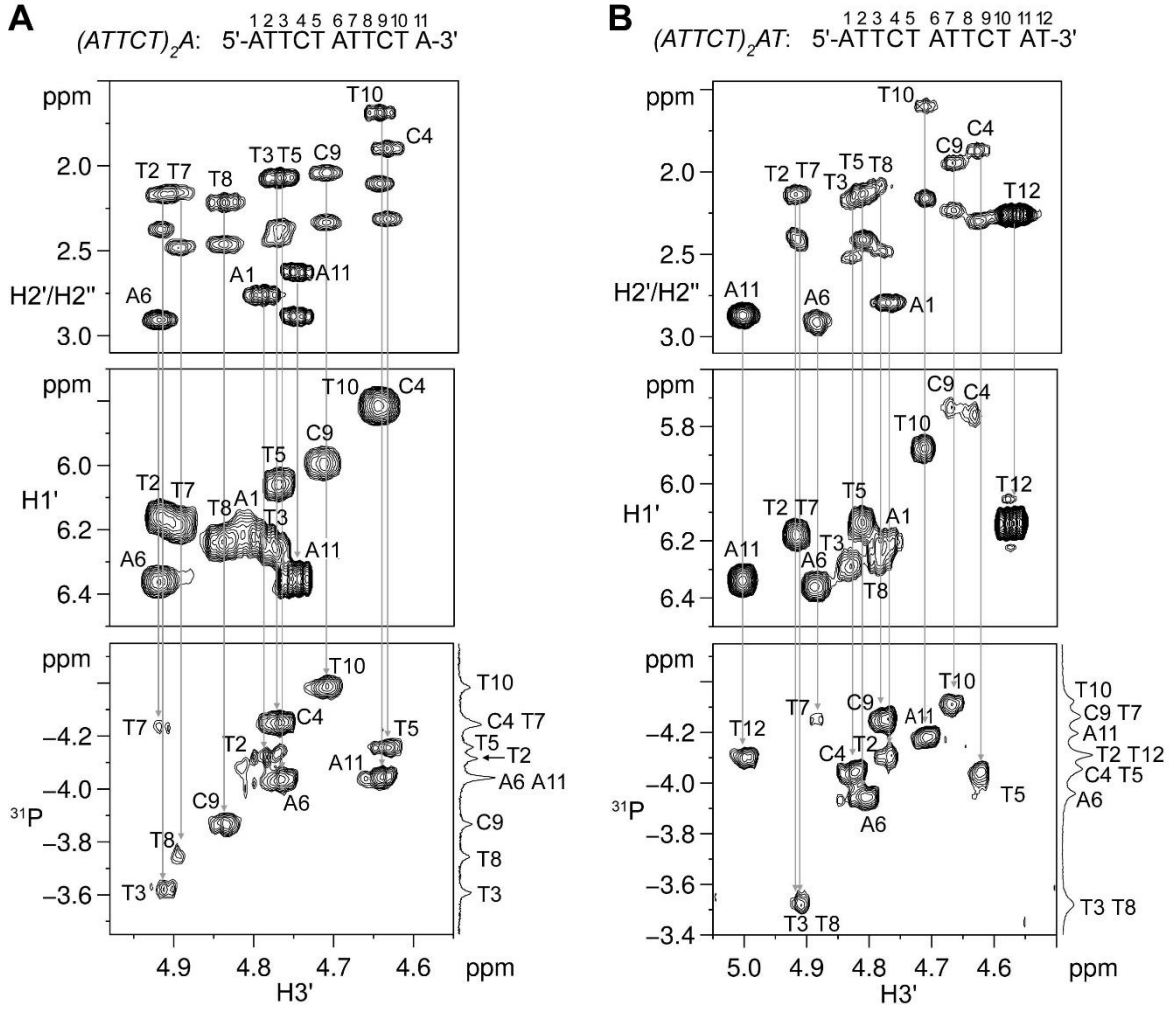


Figure S20. The TOCSY (top and middle) and ^{1}H-^{31}P HSQC (bottom) spectra show the ^{31}P resonance assignments of (A) *(ATTCT)$_2$ATT* and (B) *(ATTCT)$_2$ATTC*. The spectra were acquired at 0 °C. Due to peak overlaps, the ^{31}P resonances of *(ATTCT)$_2$ATTC* were partially assigned.

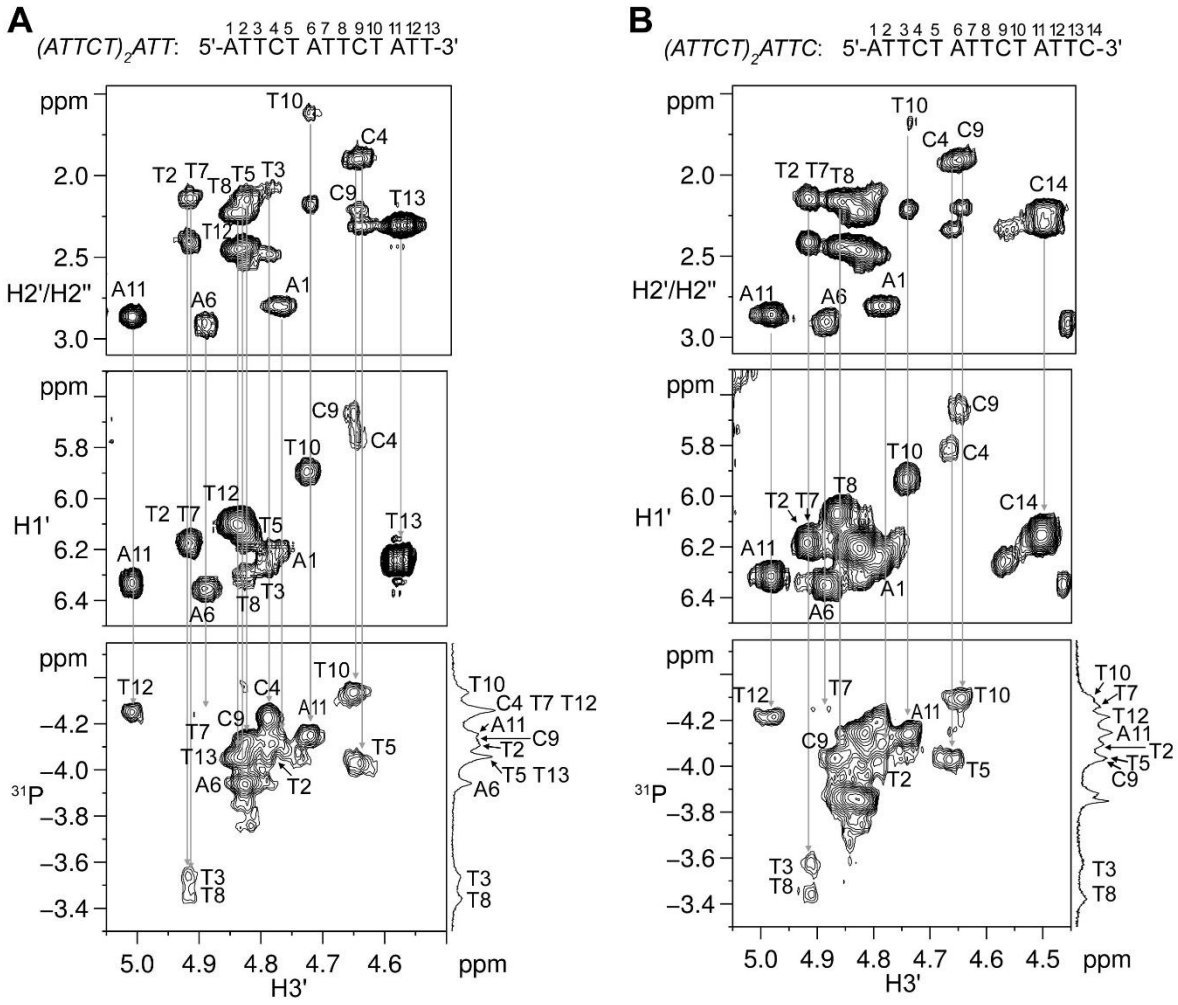


Figure S21. The TOCSY (top and middle) and ^{1}H-^{31}P HSQC (bottom) spectra show the ^{31}P resonance assignments of (A) *T(ATTCT)$_2$* and (B) *CT(ATTCT)$_2$*. The spectra in (A) and (B) were acquired at 5 and 0 °C, respectively.

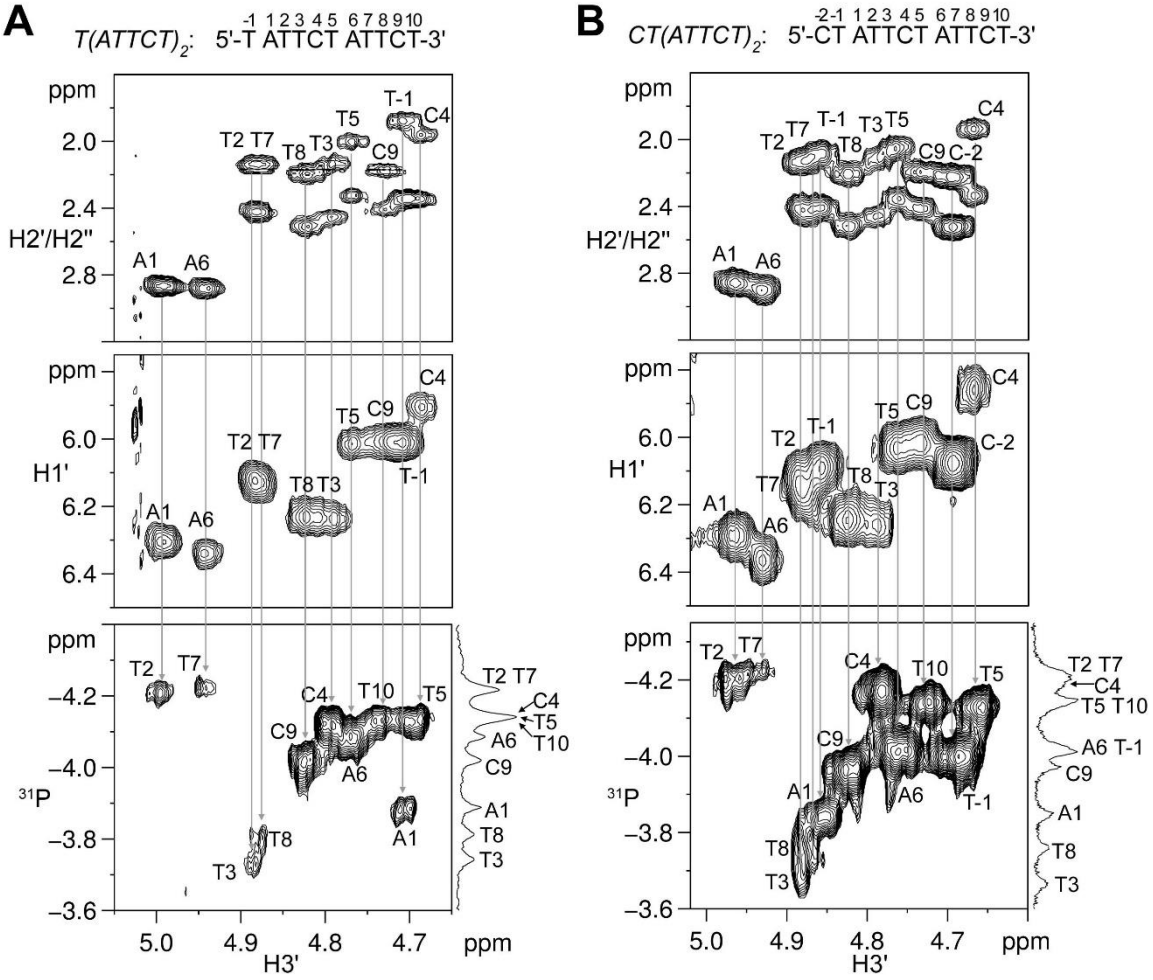


Figure S22. The TOCSY (top and middle) and ^{1}H-^{31}P HSQC (bottom) spectra show the ^{31}P resonance assignments of (A) *TCT(ATTCT)$_2$* and (B) *TTCT(ATTCT)$_2$*. The spectra were acquired at 0 °C. Due to peak overlaps, the ^{31}P resonances of *TTCT(ATTCT)$_2$* were partially assigned.

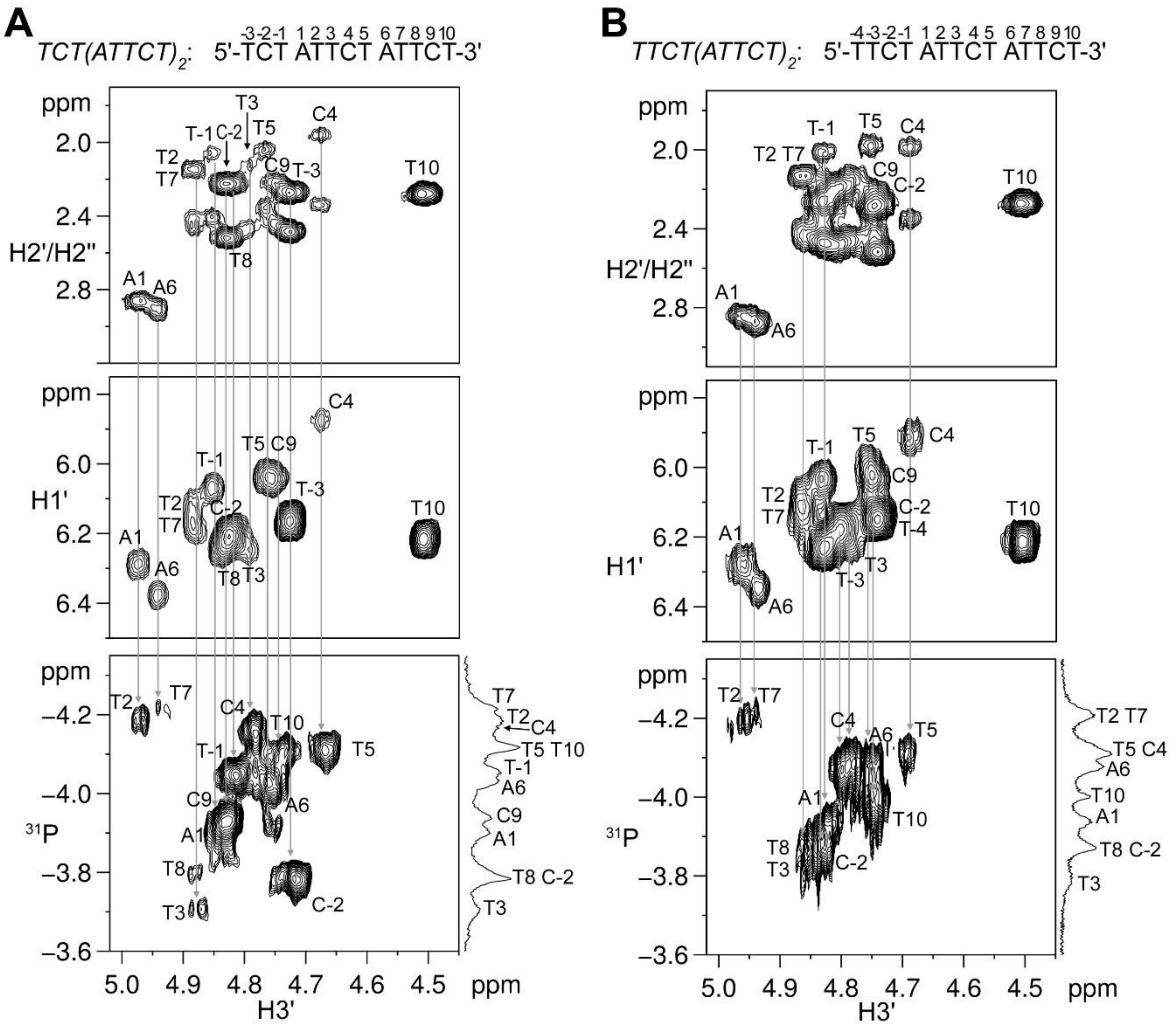


Figure S23. The TOCSY (top and middle) and ^{1}H-^{31}P HSQC (bottom) spectra show the ^{31}P resonance assignments of *T(ATTCT)$_2$A*. The spectrum was acquired at 5 °C.

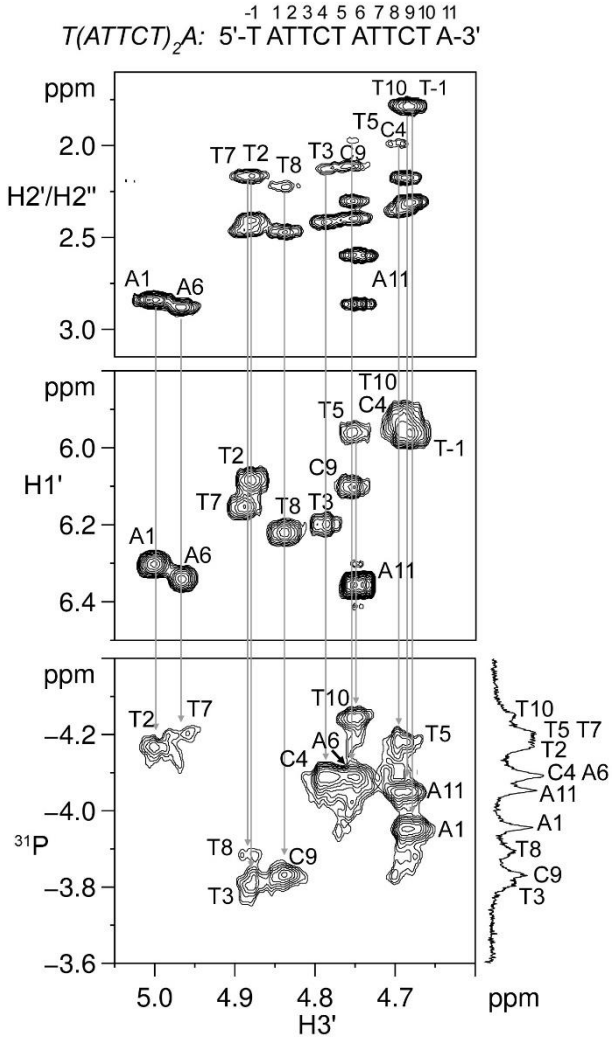


Figure S24. The native PAGE results of (A) *(ATTCT)$_2$*, *(TTCTA)$_2$*, *(ATTCT)$_3$*, *(ATTCT)$_4$*, *(ATTCT)$_5$*, and (B) *(ATTCT)$_2$A*, *(ATTCT)$_2$AT*, *(ATTCT)$_2$ATT*, *(ATTCT)$_2$ATTC*, *T(ATTCT)$_2$*, *CT(ATTCT)$_2$*, *TCT(ATTCT)$_2$*, and *TTCT(ATTCT)$_2$*. All these sequences adopted monomeric conformations by a single strand, as their mobilities were similar to those corresponding single-strand DNA ladders. The gel assays were conducted at ~5 °C in a fridge. DNA sample condition: ~1 mM DNA in 10 mM NaPi (pH 7). DNA ladder information: sequences containing two, three, four, five, six, seven, and eight TTTA repeats which correspond to 8, 12, 16, 20, 24, 28 and 32 nt in 10 mM NaPi (pH 7) and 8 M urea. The monomeric states of these TTTA repeating sequences have been demonstrated in our previous study (4).

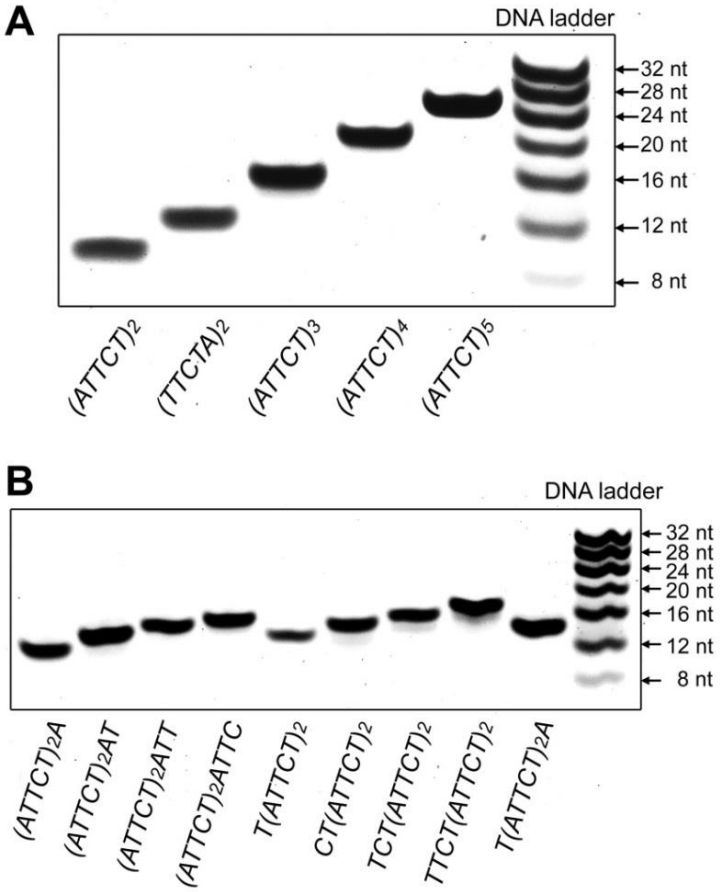


Figure S25. The fitted UV melting curves of *(ATTCT)$_2$*. The UV absorbance data at 260 nm (A_{260}) were normalized to 1.0. The T_m value of the MDB structure formed by *(ATTCT)$_2$* was determined to be 17.0 ± 0.5 °C by three replicate measurements, i.e. 16.6 °C (red diamond), 16.8 °C (blue circle), and 17.6 °C (black asterisk), respectively. The UV melting curves shown here were fitted using a two-state transition model (5).

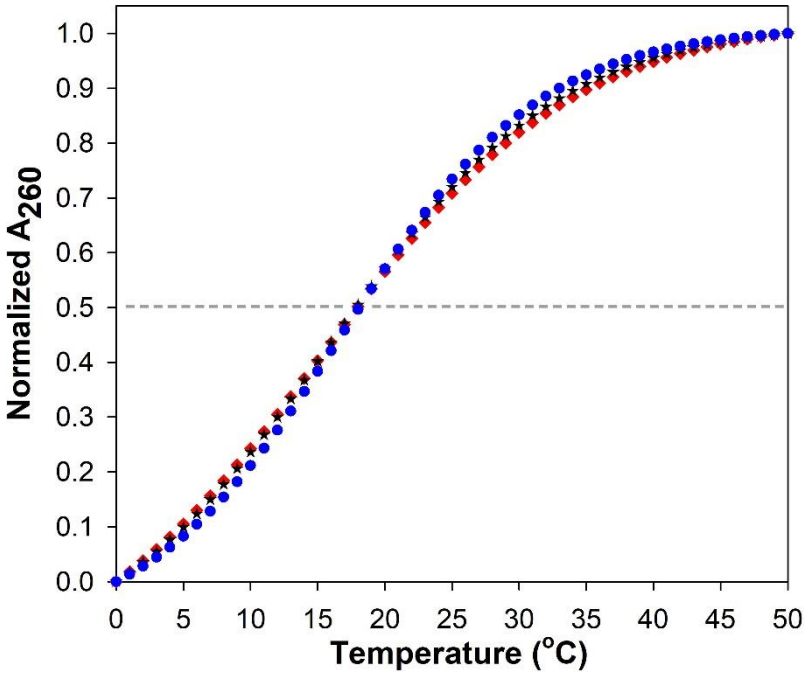


Figure S26. For *(ATTCT)*$_2$, (A) 1D NOE spectra (middle and bottom) using excitation sculpting (6) for water suppression. To enhance the H3 signals, the reference spectrum was acquired using jump-return pulse (7) for water suppression (top). (B) NOEs of T2 H6-A1 H8 and T7 H6-A6 H8 suggest base-base stackings between the two loop-closing base pairs. (C) NOEs of T3 H2'/H2''-A6 H2, T3 H7-A1 H2, T8 H2'/H2''-A1 H2 and T8 H7-A6 H2 suggest T3 and T8 located in the minor groove, as A6 and A1 H2 residues pointed to the minor groove side of the MDB. (D) The NOEs between C4 and T2-A6, and between C9 and T7-A1 suggest C4 and C9 stack on T2-A6 and T7-A1 loop-closing base pairs, respectively. (E) The 3'-5' terminal NOEs between T10 and A1. The NOESY spectra shown here were acquired at 5 °C.

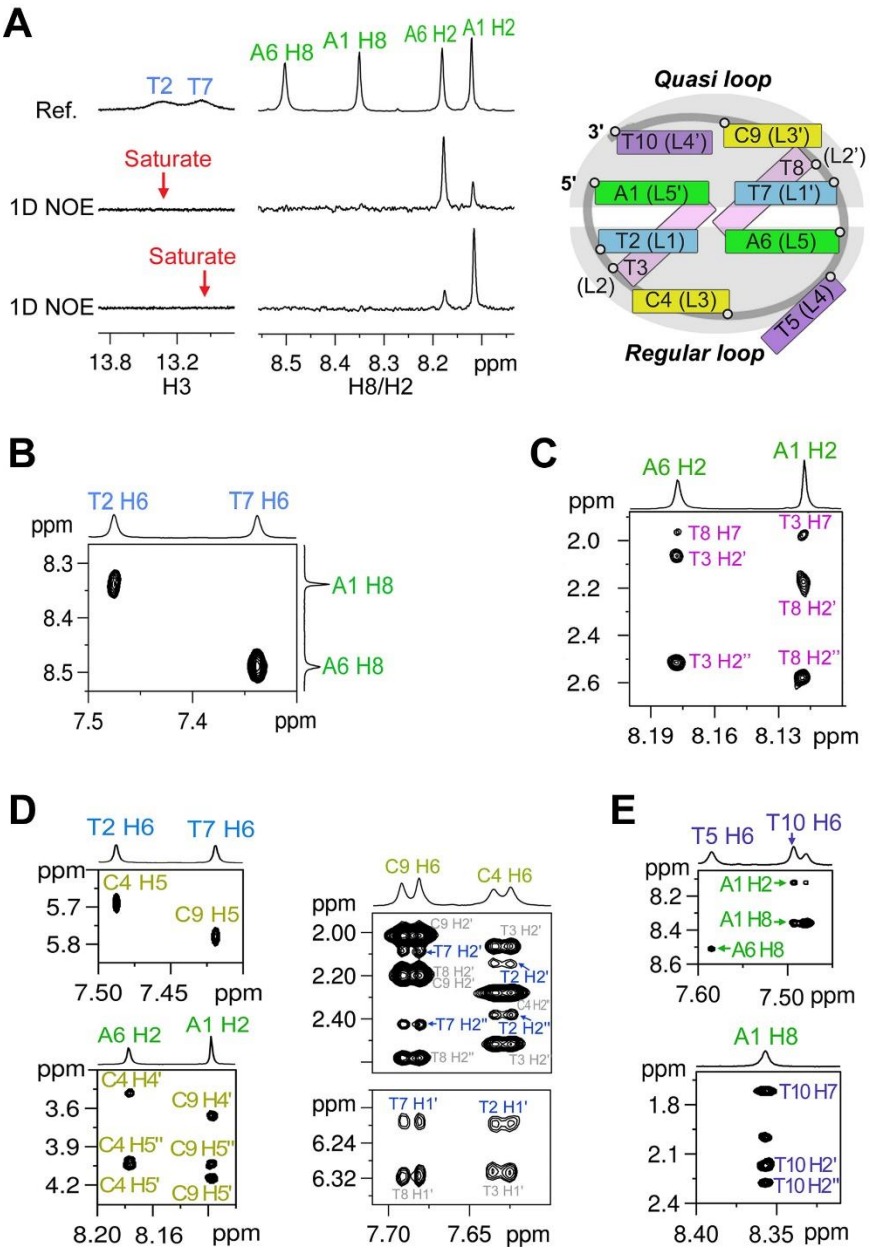


Figure S27. Among the 20 refined solution structures of the ATTCT-Q/L5' MDB formed by *(ATTCT)*$_2$, 3 structures show that T5 base is nearly perpendicular to A6 base (#4, 8 and 9), and 17 structures show that T5 base is far from A6 base, None of them show base-base stacking between T5 and A6.

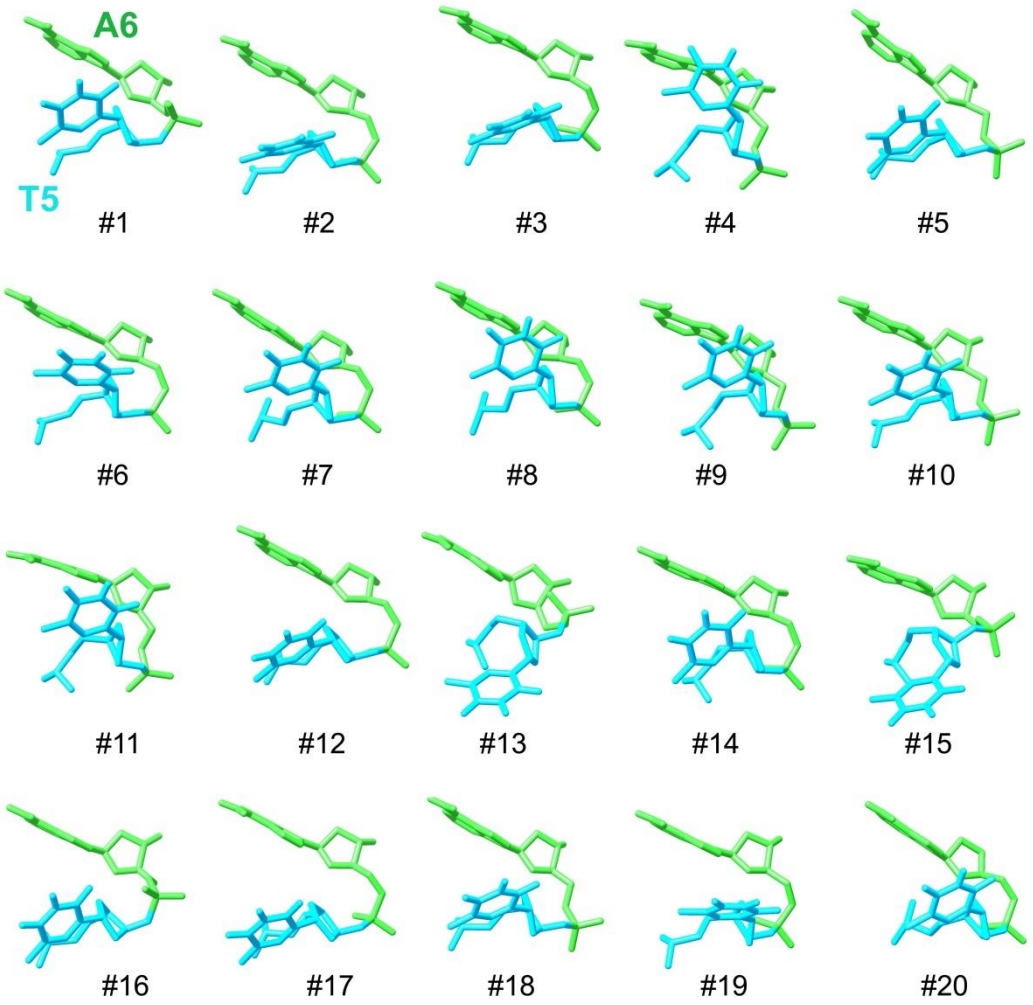


Figure S28. (*TTCTA)$_2$* formed an MDB containing two regular TTCTA pentaloops, as supported by (A) the relatively downfield shifted H7/^{31}P signals of T2 and T7, and the relatively upfield shifted H7 signals of T1 and T6, (B) the base-base NOEs of C3 H5-T1 H6 and C8 H5-T6 H6 suggesting C3 and C8 stacked on T1-A5 and T6-A10, respectively, and (C) the NOEs of T1 H6-A10 H2/H8 suggesting the existence of 3'-5' terminal stacking. The NMR spectra shown here were acquired at 0 °C.

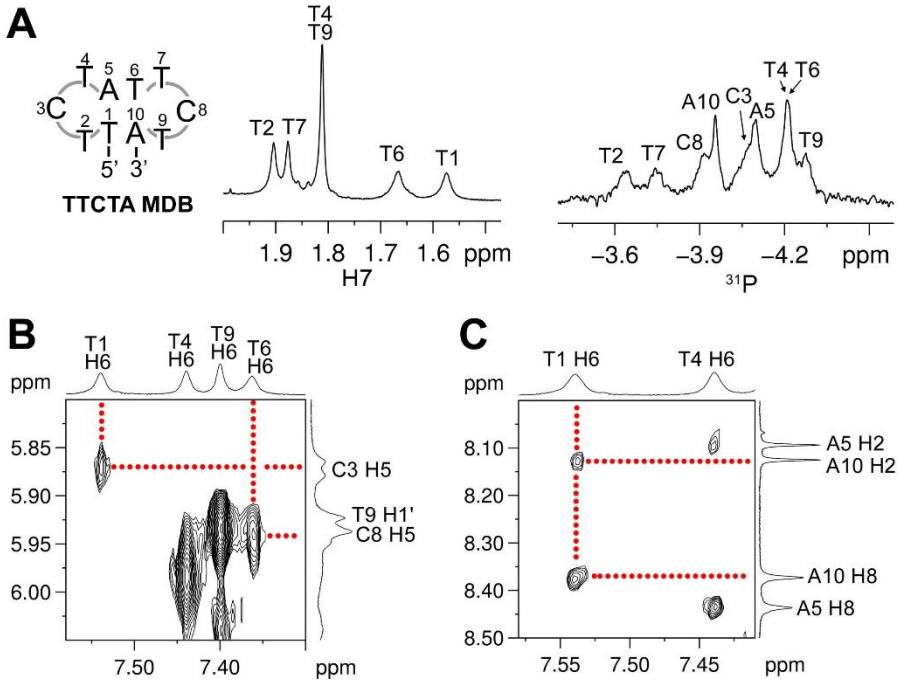


Figure S29. The NOESY NMR spectra of (A) *(ATTCT)$_4$* and (B) *(ATTCT)$_5$* show the base-base NOEs of C4 H5-T2 H6, C4 H6-A6 H2, C9 H5-T7 H6, C9 H6-A1 H2, which suggest C4 and C9 stacked on T2-A6 and T7-A1 base pairs, respectively. The spectra shown here were acquired at 5 °C.

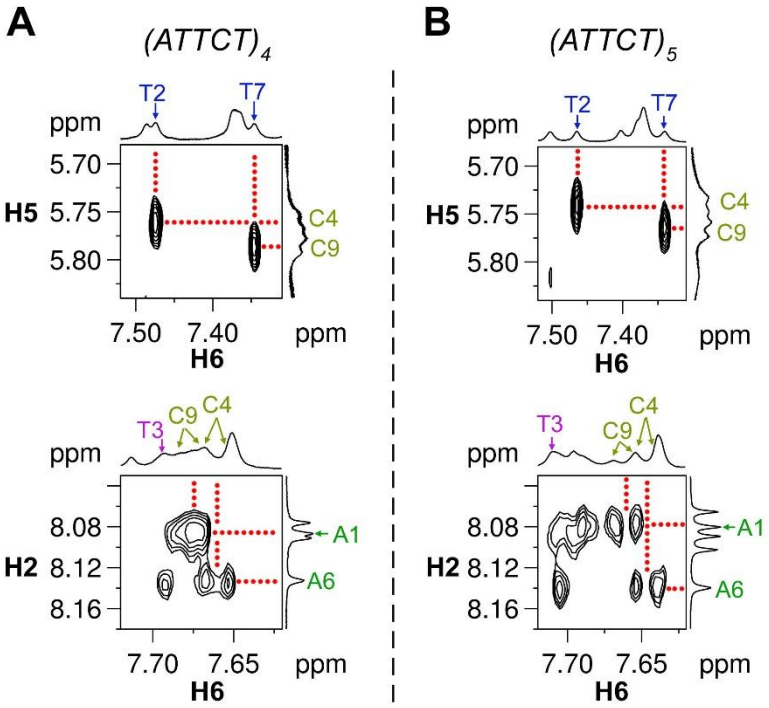


Figure S30. (A) The downfield shifted T3/T8 ^{31}P signals, and (B) the base-base NOEs of T2 H6-C4 H5 and T7 H6-C9 H5 suggest the formation of MDB in *(ATTCT)$_2$A*. (C) The NOEs of T10 H7-T8 H6, T10 H6-A11 H2/H8, and (D) the NOEs of A1 H8-A11 H1'/H2'/H2'' suggest T10 stacked on T8-A11 base pair, and A1 and A11 were close in space, respectively, in the minor dumbbell conformer. (E-F) The NMR spectral features of the MDBs formed by *(ATTCT)$_2$AT*, *(ATTCT)$_2$ATT* and *(ATTCT)$_2$ATTC*, including the relatively downfield shifted T3/T8 ^{31}P signals, and the base-base NOEs between C4/C9 H5 and T2/T7 H6. The spectra shown here were acquired at 0 °C.

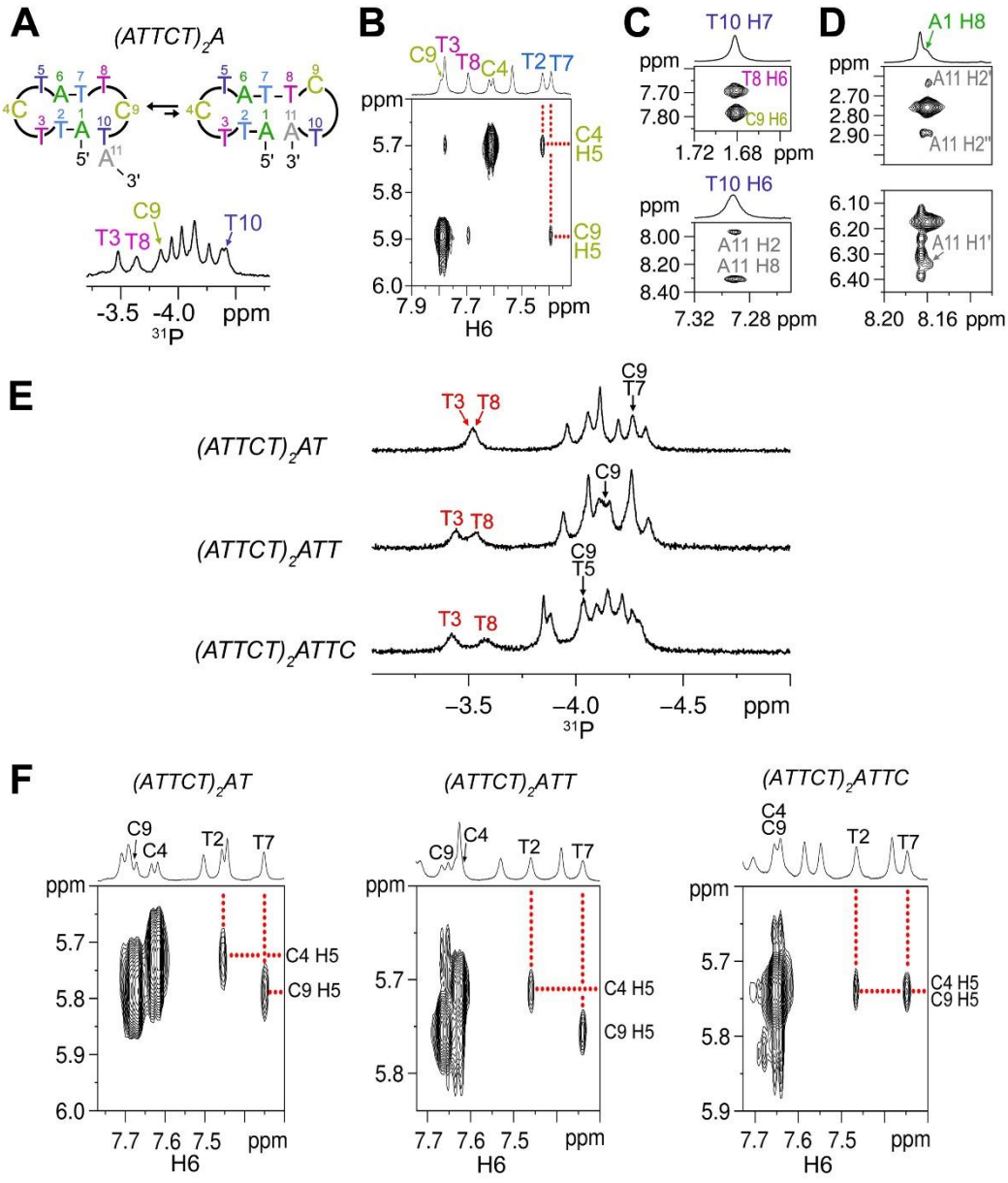


Figure S31. NMR spectral features of the MDBs formed by *T(ATTCT)*$_2$, *CT(ATTCT)*$_2$, *TCT(ATTCT)*$_2$, and *TTCT(ATTCT)*$_2$, including (A) their relatively downfield T3/T8 ^{31}P signals, and (B) the base-base NOEs between C4/C9 H5 and T2/T7 H6. The spectra shown here were acquired at 0 °C.

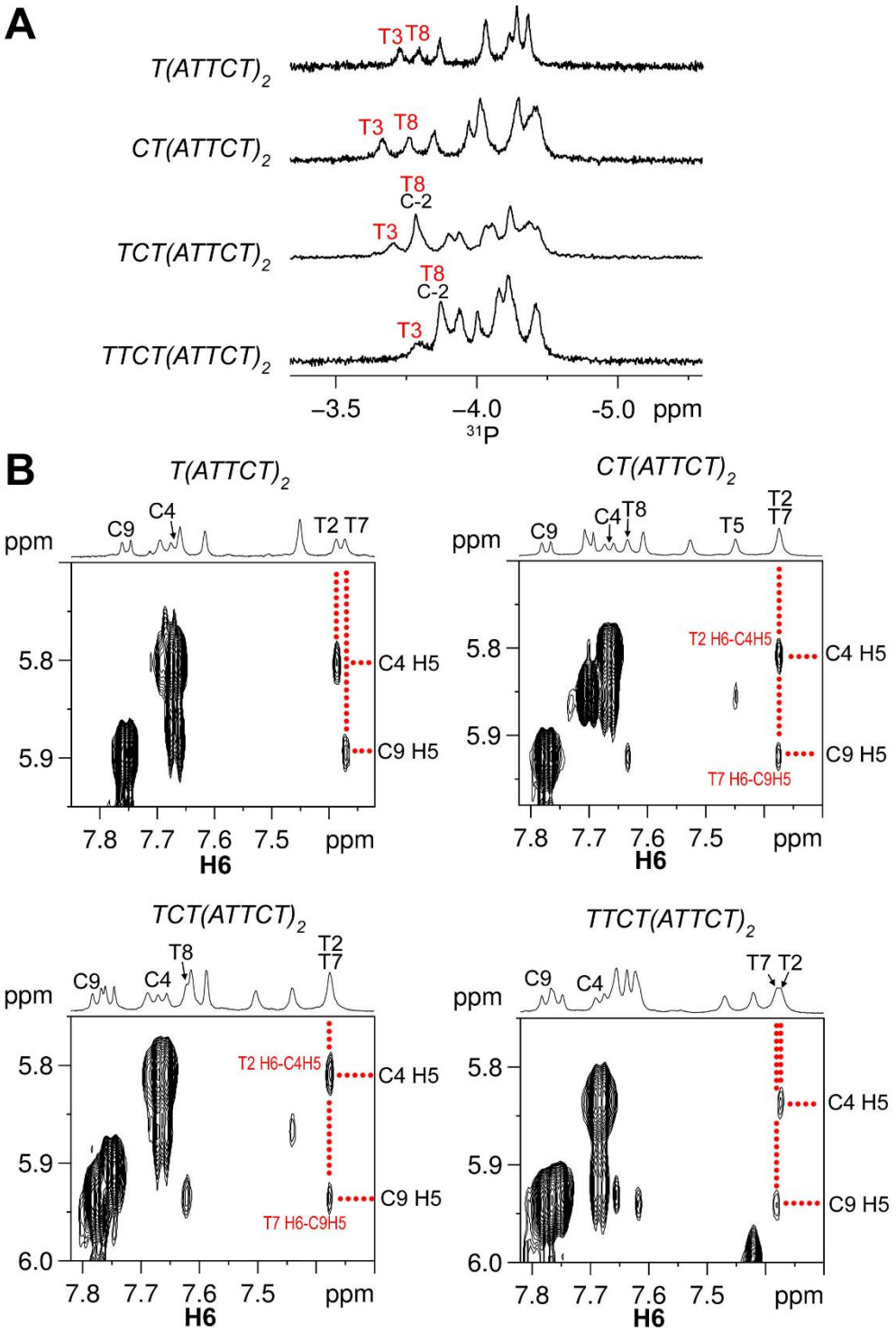


Figure S32. The NMR spectral features of the MDB formed by *T(ATTCT)$_2$A*, including (A) the downfield shifted T3/T8 ^{31}P signals, and (B) the base-base NOEs between C4/C9 H5 and T2/T7 H6. The spectra shown here were acquired at 0 °C.

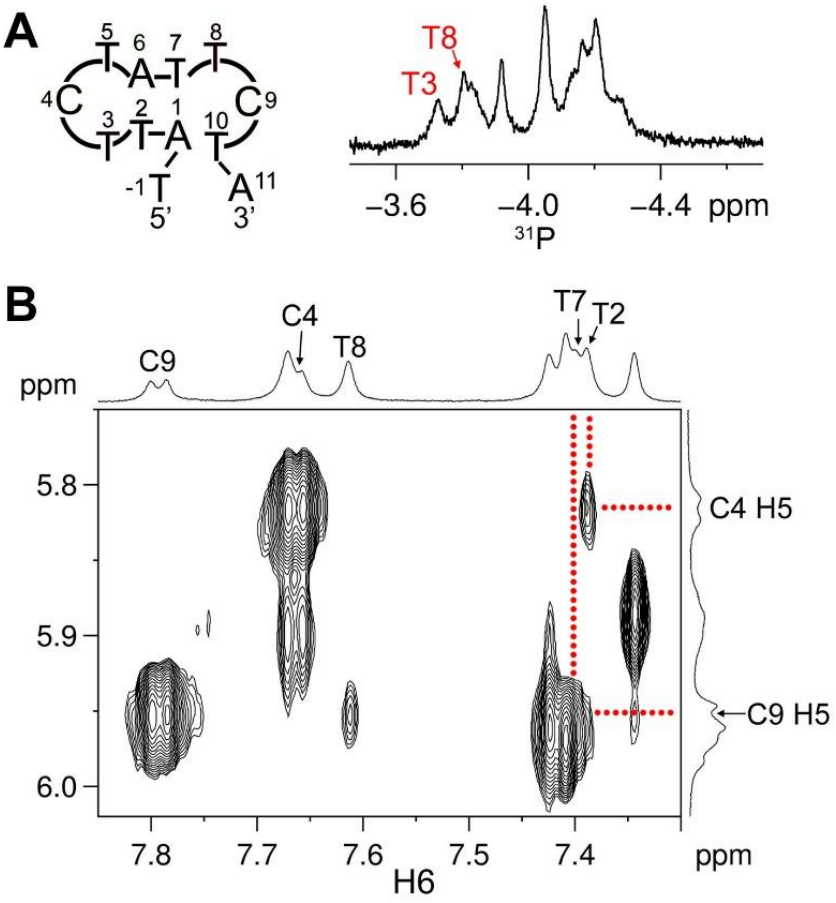
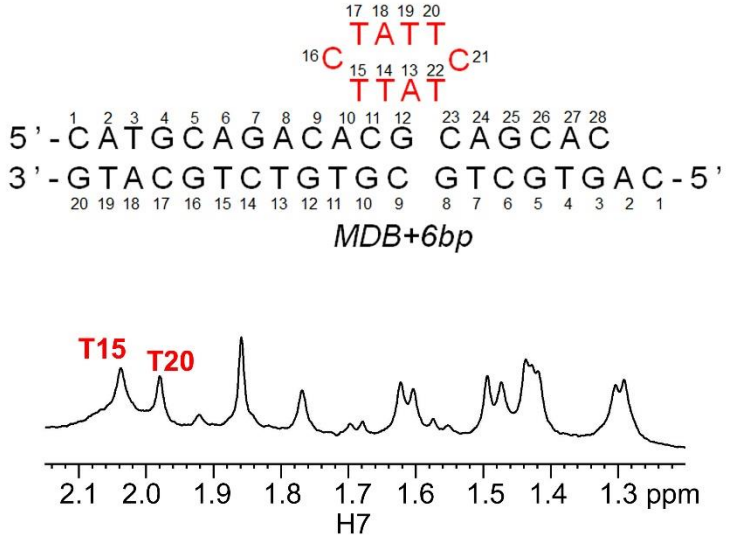


Figure S33. 1D ^{1}H NMR spectrum showing the methyl proton region of *MDB+6bp*. The unusually downfield shifted signals of T15 H7 (2.04 ppm) and T20 H7 (1.98 ppm) suggest formation of the ATTCT-Q/L5' MDB in the primer. The sample was prepared under 10 mM NaPi (pH 7), 50 mM NaCl and 10 mM $MgCl_2$. The NMR spectrum was acquired at 37 °C.



References:

1. Guo, P. and Lam, S.L. (2015) Unusual structures of TTTA repeats in *icaC gene* of Staphylococcus aureus. *FEBS Lett.*, **589**, 1296-1300.
2. Guo, P. and Lam, S.L. (2015) New insights into the genetic instability in CCTG repeats. *FEBS Lett.*, **589**, 3058-3063.
3. Liu, Y., Guo, P. and Lam, S.L. (2017) Formation of a DNA mini-dumbbell with a quasi-type II loop. *J. Phys. Chem. B.*, **121**, 2554-2560.
4. Guo, P. and Lam, S.L. (2016) The competing mini-dumbbell mechanism: new insights into CCTG repeat expansion. *Signal Transduct. Target. Ther.*, **1**, 16028.
5. Greenfield, N.J. (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.*, **1**, 2527-2535.
6. Stott, K., Stonehouse, J., Keeler, J., Hwang, T.L. and Shaka, A.J. (1995) Excitation sculpting in high-resolution nuclear magnetic resonance spectroscopy: application to selective NOE experiments. *J. Am. Chem. Soc.*, **117**, 4199-4200.
7. Plateau, P. and Gueron, M. (1982) Exchangeable proton NMR without base-line distortion, using new strong-pulse sequences. *J. Am. Chem. Soc.*, **104**, 7310-7311.