Supplementary Information

Reverse Watson-Crick G-G base pair in G-Quadruplex formation

Soma Mondal[a], Jyotsna Bhat[a], Jagannath Jana[a] , Meghomukta Mukherjee[a] and Subhrangsu Chatterjee[a]*

[§]Department of Biophysics, Bose Institute, P-1/12 CIT Scheme VIIM, Kankurgachi, Kolkata, WB, India.

Methods

Materials: The sequences were purchased from Eurofins Genomics India Pvt Ltd and the matrix materials for MALDI-TOF that is 3-Hydroxypicolinic acid, Ammonium citrate dibasic (DAC) were purchased from SIGMA-ALDRICH. The stock solution for CD and NMR were prepared by dissolving the oligonucleotides in 20 mM Tris-HCl buffer containing 1mM EDTA and was annealed at 90˚C for 5 minutes. The stock solution was stored at 4˚C and 100 mM KCl was added. The matrix for mass spectrometry was prepared by making a saturated solution 3-hydroxypicolinic acid in H2O:ACN 50:50 (v:v), containing 10mg/ml DAC.

Circular Dichroism: CD spectroscopy was used to determine the secondary structures adopted by GG4 sequences using Jasco 815 spectrometer. All the samples were diluted with 20 mM Tris-HCl buffer containing 100 mM KCl or NH_4Cl and 1 mM EDTA at pH 7.4 to 20 uM strand concentration. Spectra were recorded over a range of 220 to 310 nm with accumulation of three scans at a speed of 100 nm/min having data interval of 1nm. Cuvette having 0.1 cm path length were used. Spectra were taken at room temperature and were base line corrected.

MALDI-TOF spectrometry: MALDI-TOF spectrometry were performed to elucidate the strand stoichiometry leading to various intermolecular structures using Bruker Daltonics autoflex TOF/TOF model.The samples were dissolved in 100 mM Ammonium citrate dibasic and were heated at 85˚C for 5 minutes and kept at 4˚C after it was allowed to cool to room temperature. The final strand concentration was adjusted to 10 µM. 100 µM cisplatin was added to GG4(T) and incubated for two days. Ground steel plate (MTP 384) were used for sample spotting. The spectra were recorded in negative ion mode with the accelerating voltage set at 20 KV and 100 laser shots were given for each spectrum.

NMR Spectroscopy: All NMR spectra were recorded using Bruker AVANCE III 500 MHz NMR spectrometer equipped with a 5 mm SMART probe at 283 K. Data processing and acquisition were performed with TopspinTM*v3.1* software. NMR samples were prepared in 20 mM Tris-HCl ,100 mM KCl ,1mM EDTA at pH 7.4 containing 10% D2O. DSS (2, 2- dimethyl-2-silapentane 5-sulfonate sodium salt) was used as an internal standard (0.0 ppm), the strand concentration of the samples were 300 µM. One dimensional proton spectra for all oligonucleotide sequences were recorded using Bruker Pulse programme “zgesgp” with a spectral width of 20 ppm, number of scan 512 with calibrated pulse length.

Native Polyacrylamide gel electrophoresis: Generally for a small size nucleic acid separation, native polyacrylamide gel electrophoresis (PAGE) has been done. The size of our samples was very small, so we also performed a non-denaturing PAGE for its analysis. The samples were run in 16% polyacrylamide gel, where the ratio of acrylamide and bis-acrylamide was 29:1 and the voltage was 5 mili-amps. The samples that are incubated for two months at 4˚C were used. The samples concentrations were 500 ng/ µL in each lane. To perform the native PAGE experiment Tris-buffer was used in all cases. The gels were stained by ethidium bromide and imaged in gel-dock (Bio-rad gel dock xr+).

Model building and Molecular Dynamics Simulations: We needed to build models of **‘XXXGGGGXXX’** intermolecular dimer where G-G intermolecular pairing is with Hoogsteen hydrogen bonds and terminal bases are X=A, T and mixture where one strand X=A and another strand X=T. A racemic crystal structure of tetra-molecular G-Quadruplex has been released recently (PDB-ID:4R44)[1]; each crystal is composed of four chains of ‘TGGGGT’. Chain A and chain B of the crystal were extracted and taken as a starting template. To make $(AAAGGGGAAA)_2$i.e GG4(A) terminal thymines were edited and adenines were built in Maestro 9.0[2]. The edited structure was energy minimized and taken for further simulation purposes. Similarly,starting models were built for $(TTTGGGGTTT)_2$i.e GG4(T) and (TTTGGGGTTT)+ (AAAGGGGAAA) i.e GG4(AT).
For all the three models simulation systems were built in tleap module of AMBER14[3] and all the parameters and simulation steps are maintained identical for each system. ff99SB and PARMBSC0 force fields of AMBER14 are applied over nucleic acids whereas, General Amber Force Field (GAFF) parameters were applied for inorganic entities such as ions[4, 5]. All systems were

neutralized with the addition of counter ions and further solvated with 8ÅTIP3PBOX water model[6]. Systems were further minimized in two steps; **Step1**: only ions and water molecules were minimized and nucleic acid regions were kept fixed, for first 500 cycles steepest descent method was applied and for next 2000 cycles conjugate gradient method was applied, **Step2**: in this step all atoms were allowed to move, for first 1000 cycles steepest descent method was applied while for next 2500 cycles were performed with conjugate gradient method. Minimized systems were further heated to reach the temperature of 300 K. Heating was carried out with NVT ensemble for 50 ps with restrained force constant of 2 kcal/mol/Å^2 over nucleic acid region. Equilibration was conducted for 1 ns with a NPT ensemble. Final production phase was performed for 50 ns simulation period, with a NPT ensemble at 300 K temperature and 1 atm pressure; with the step size of 2 fs. A Langevin thermostat and barostat was used for temperature and pressure coupling[7, 8]. A SHAKE algorithm was applied to constrain all bonds containing hydrogen atoms[9]. The non-bonded cut-off was kept at 12 Å and long range electrostatic interactions were treated by the Particle Mesh Ewald (PME) method with fast Fourier transform grid having approximately 0.1 nm space[10]. Trajectory snapshots were taken at each 10 ps, which were finally used for analysis.Further analysis and visualization was conducted with cpptraj, VMD, PYMOL and Chimera[11-14].

Observations of simulation results:

Backbone atom RMSD represents the overall flexibility of the structure and thus the stability of the structure. Here, RMSD analyses of three systems (Figure S8 (top))suggested that the overall structures of all the three systems are undergoing significant changes in comparison to their starting conformation. Though they acquired high RMSD (average values) such as GG4(A):6.21 Å, GG4(AT):6.30 Å and GG4(T):7.60 Å but the RMSD changes are steady after 20 ns of simulation run. Therefore RMSD calculations were again performed by considering coordinates after 20ns simulation run as a starting frame, here the RMSD is decreasing drastically such as GG4(A):3.41 Å, GG4(AT):3.16 Å and GG4(T):2.89 Å (average values) (Figure S8(bottom)).This indicates that the resulting structures are different from the starting conformations however they are achieving conformational stability. Water and ion grids are calculated over the average structure of GG4(A),GG4(T) and GG4(AT). As seen in Figure S7 water and ion density is thicker in case GG4(A) and thinner around GG4(T) and GG4(AT).Denser grid indicates that overall structure is same as a starting structure i.e. less flexible and thinner density indicates higher changes in overall geometry of the structure. Thus conformational changes are significant more in case of GG4(T) and GG4(AT).A B-factor analysis further supports the overall conformational stability, it indicates that the G-region of all the three systems are stable most compared to the terminal bases (Figure 2B).When the whole trajectories were visually observed it was found that the starting structures of each system is undergoing significant change and rearranging to new conformation which is stable throughout the simulation period (simulation movie of GG4(AT) is provided in the supplementary data). The rearrangement is occurring majorly by the means of changes in hydrogen bond pattern it is discussed in details in the main manuscript. In brief initial Hoogsteen hydrogen bonding is interchanged with reverse Watson Crick hydrogen bonding in G-G base pairing. In summary the simulation observations are suggesting that, despite the fact that we started with Hoogsteen G-G pairing we acquired new conformation with dissimilar hydrogen bonding pattern i.e. reverse Watson Crick hydrogen bonding among G-G pairing. The new conformations are energetically stable, less flexible and they contain higher and stable hydrogen bonds.

Table S 1.Sequences taken, Abbreviation and Molecular Weight

Sequences taken	Abbreviation	Molecular Weight
$A_3G_4A_3$	GG4(A)	3134.0
$T_3G_4T_3$	GG4(T)	3080.0
$C_3G_4C_3$	GG4(C)	2989.0
G_{10}	GG4(G)	3230.0
($A_3G_4A_3$+$T_3G_4T_3$) mixture	GG4(AT)	-
($C_3G_4C_3$+G_{10}) mixture	GG4(GC)	-

Table S2.Calculation of theoretical mass values (X = ammonium ion).

	monomer	dimer	trimer	tetramer	1X	2X	3X
GG4(A)	3134.0	6268.0	9402.0	12536.0	12554.0	12572.0	12590.0
GG4(T)	3080.0	6160.0	9240.0	12320.0	12338.0	12356.0	12374.0
GG4(C)	2989.0	5978.0	8967.0	11956.0	11974.0	11992.0	12010.0
GG4(G)	3230.0	6460.0	9690.0	12920.0	12938.0	12956.0	12974.0

Table S3.Calculation of observed mass values as obtained after overnight incubation at 4 ˚C (X = ammonium ion).

	monomer	dimer	trimer	tetramer	1X	2X	3X
GG4(A)	3124.7	6260.0	-	-	-	-	-
GG4(T)	3076.0	6160.8	-	-	-	-	-
GG4(C)	2983.6	5978.0	8967.2	-	-	-	-
GG4(G)	3221.6	-	-	-	-	-	-
GG4(AT)	3130.0,3075.1	6153.8,6214.8,6267.8	-	-	-	-	-
GG4(GC)	2980.2,3219.3	5977.1	8966.6	-	-	-	-

Table S4.Calculation of observed mass values as obtained after one month incubation at 4˚C (X = ammonium ion).

	monomer	dimer	trimer	tetramer	1X	2X	3X
GG4(A)	3132.8	6271.8	9404.1	-	-	-	-
GG4(T)	3074.6	6160.8	9234.2	-	-	-	-
GG4(C)	2985.6	5978.0	8965.7	-	-	-	-
GG4(G)	3230.5	-	-	-	-	-	-
GG4(AT)	3076.5,3132.1	6160.0,6216.3,6269.8	9234.0,9403.5	-			
GG4(GC)	2983.7,3226.5	5976.6	8965.1	-	-	-	-

Table S5.Calculation of observed mass values as obtained after two month incubation at 4˚C (X = ammonium ion).

	monomer	dimer	Trimer	tetramer	1X	2X	3X
GG4(A)	3126.8	6261.7	9386.8	12538.0	12552.7	12570.0	-
GG4(T)	3071.0	6152.1	9227.0	12298.8	12333.6	-	12372.6
GG4(C)	2984.0	5974.0	8963.0	-	-	-	-
GG4(G)	3231.0	-	-	-	-	-	-
GG4(AT)	3071.5, 3126.0	6152.3,6209.2, 6262.5	9239.5, 9387.4	12538.0, 12312.8	-	12575.5, 12346.6	12371.9
GG4(GC)	2987.3,3231.2	5980.7	8962.2	-	-	-	-

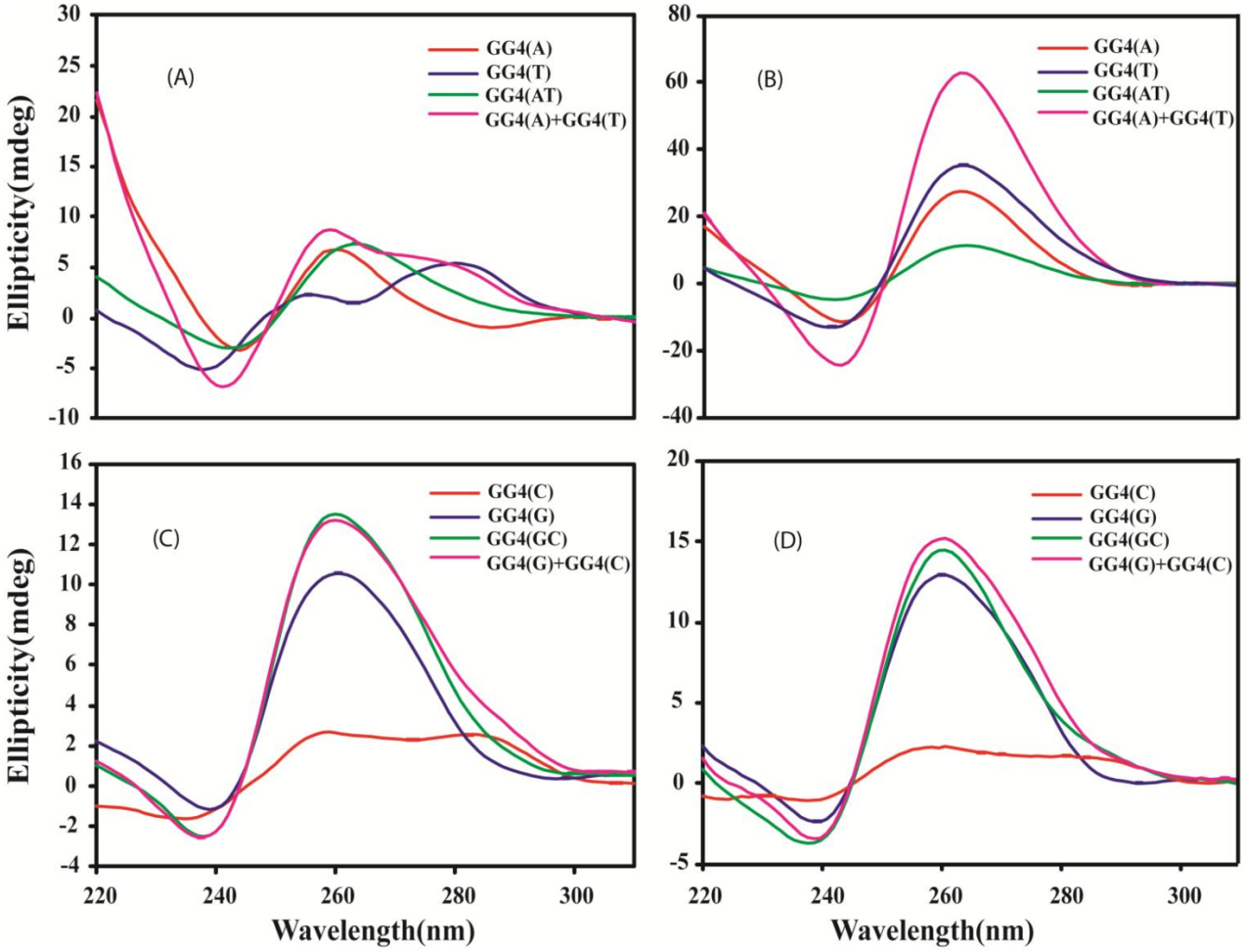


Figure S1.Circular Dichroism spectra of GG4 sequences (A) GG4(A), GG4(T), GG4(AT) after overnight incubation (B)GG4(A), GG4(T), GG4(AT) after two months incubation (C)GG4(C), GG4(G), GG4(GC) after overnight incubation (D)GG4(C), GG4(G), GG4(GC) after two months incubation. All samples were dissolved in 20 mMTris-HCl, pH 7.4, containing 1 mM EDTA and 100mMKCl, annealed at 90°C and incubated at 4°C.

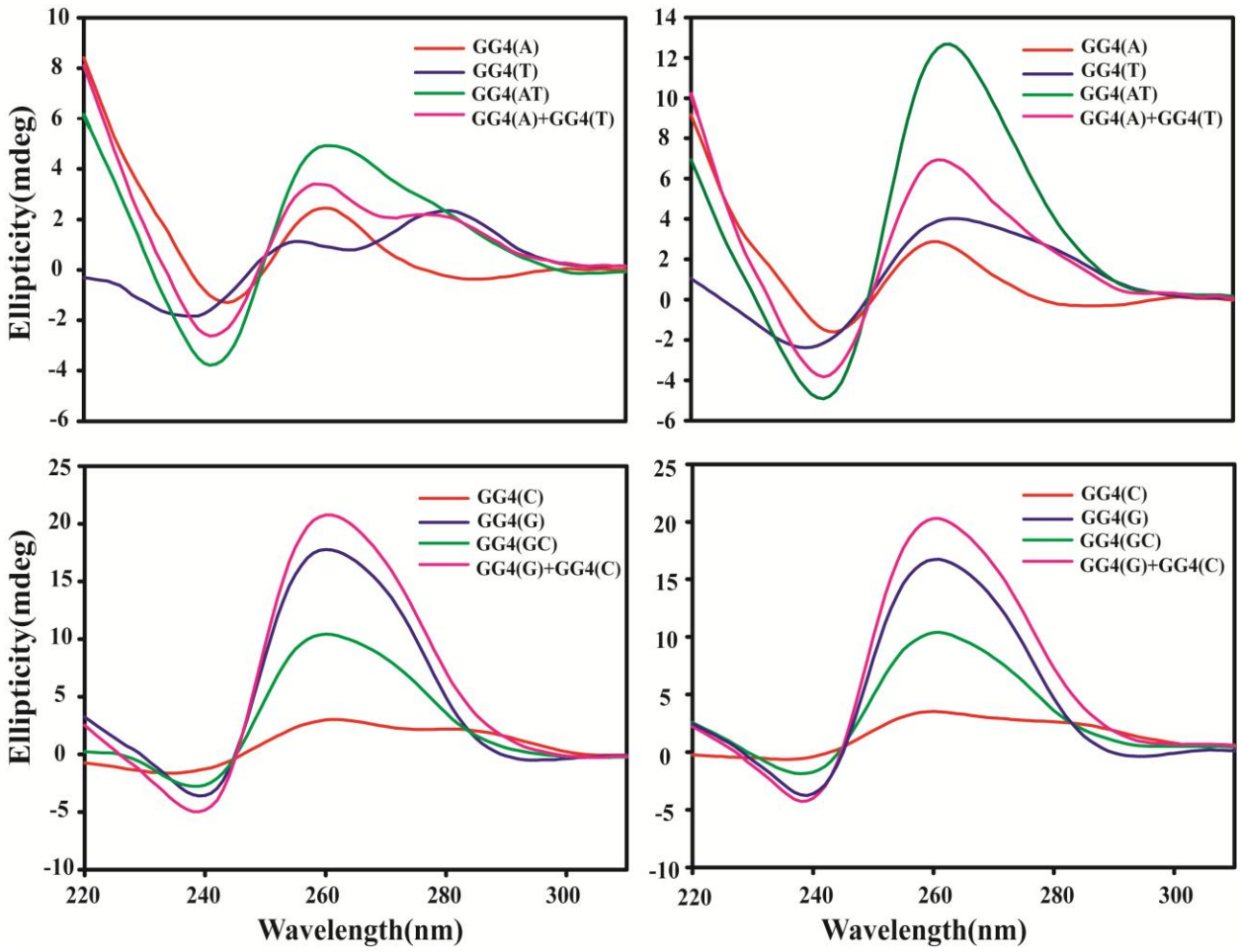


Figure S2.Circular Dichroism spectra of GG4 sequences (A)GG4(A), GG4(T), GG4(AT) after overnight incubation (B) GG4(A), GG4(T), GG4(AT) after two months incubation (C)GG4(C), GG4(G), GG4(GC) after overnight incubation (D) GG4(C), GG4(G), GG4(GC) after two months incubation. All samples were dissolved in 20 mMTris-HCl, pH 7.4, containing 1mM EDTA and 100 mM DAC, annealed and incubated at 4˚C

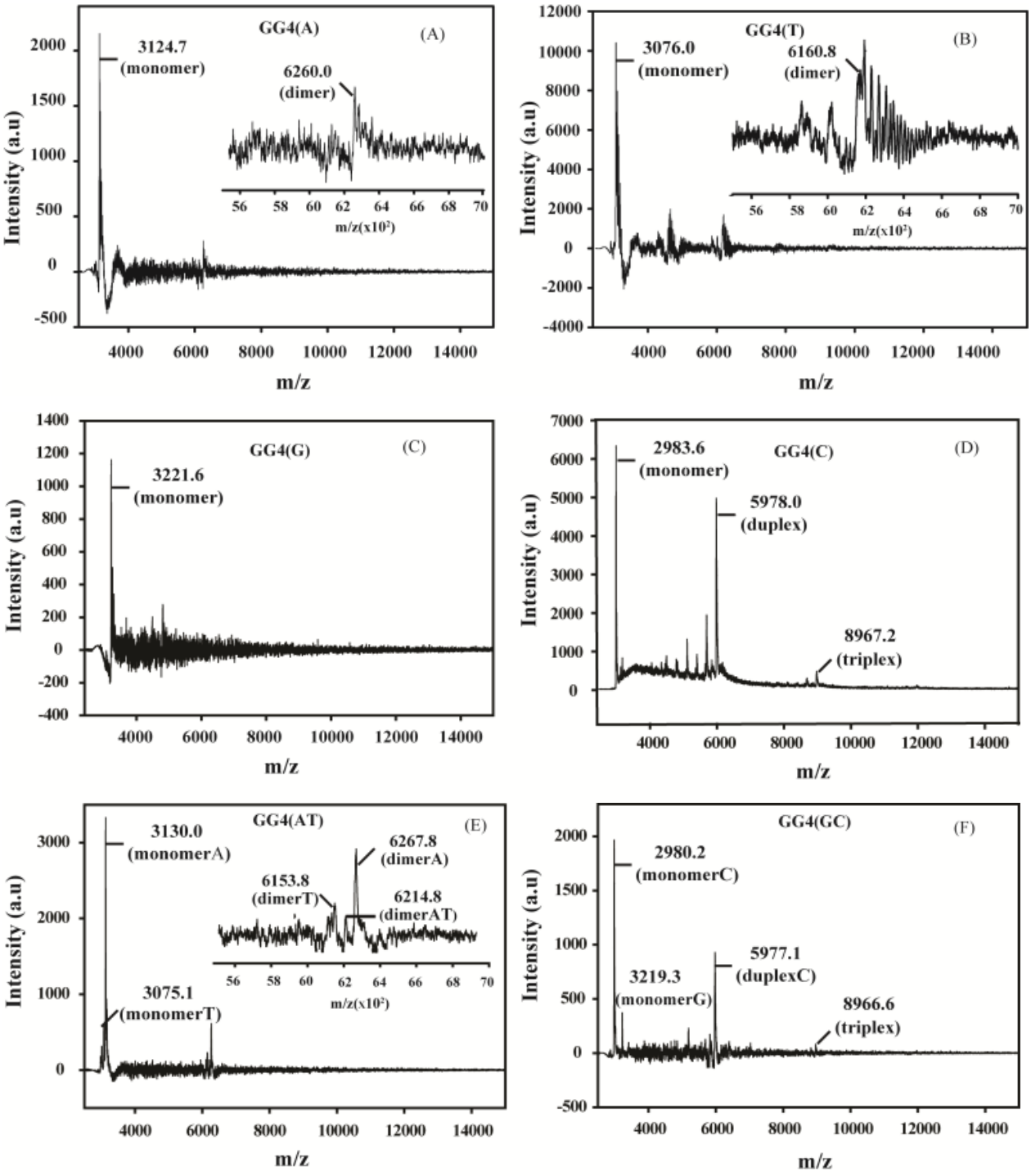


Figure S3.MALDI-TOF spectra of GG4 sequences after overnight incubation at 4˚C using 3-HPA as matrix (A)GG4(A) , inset [zoomed region of GG4(A) dimer] (B)GG4(T) , inset [zoomed region of GG4(T) dimer] (C)GG4(G) (D)GG4(C) (E)GG4(AT), inset [zoomed region of GG4(AT) dimer] (F)GG4(GC).

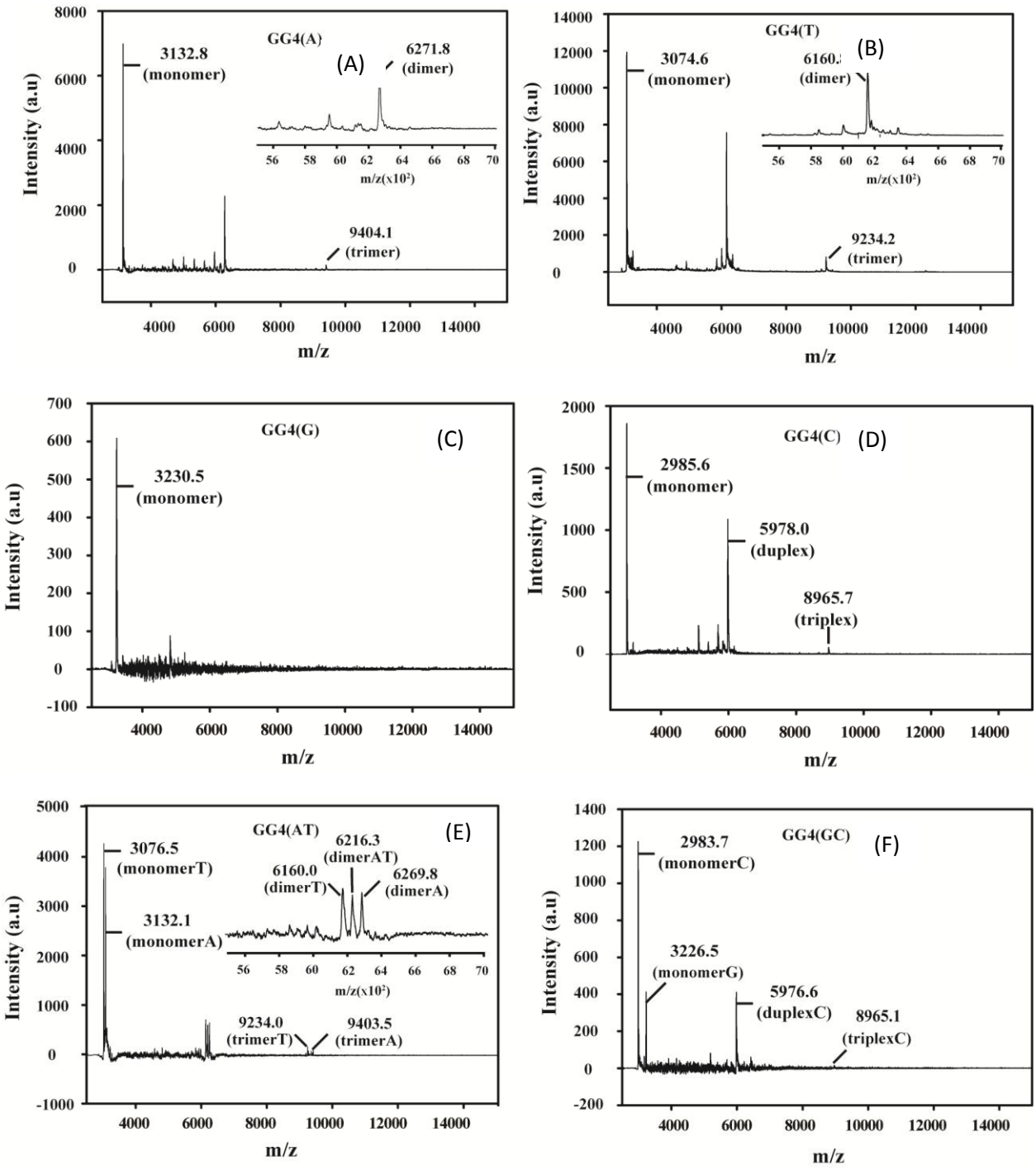


Figure S4.MALDI-TOF spectra of GG4 sequences after one month incubation at 4˚C using 3-HPA as matrix (A)GG4(A) , inset [zoomed region of GG4(A) dimer] (B)GG4(T) , inset[zoomed region of GG4(T) dimer] (C)GG4(G) (D)GG4(C) (E) GG4(AT), inset [zoom region of GG4(AT) dimer] (F)GG4(GC)

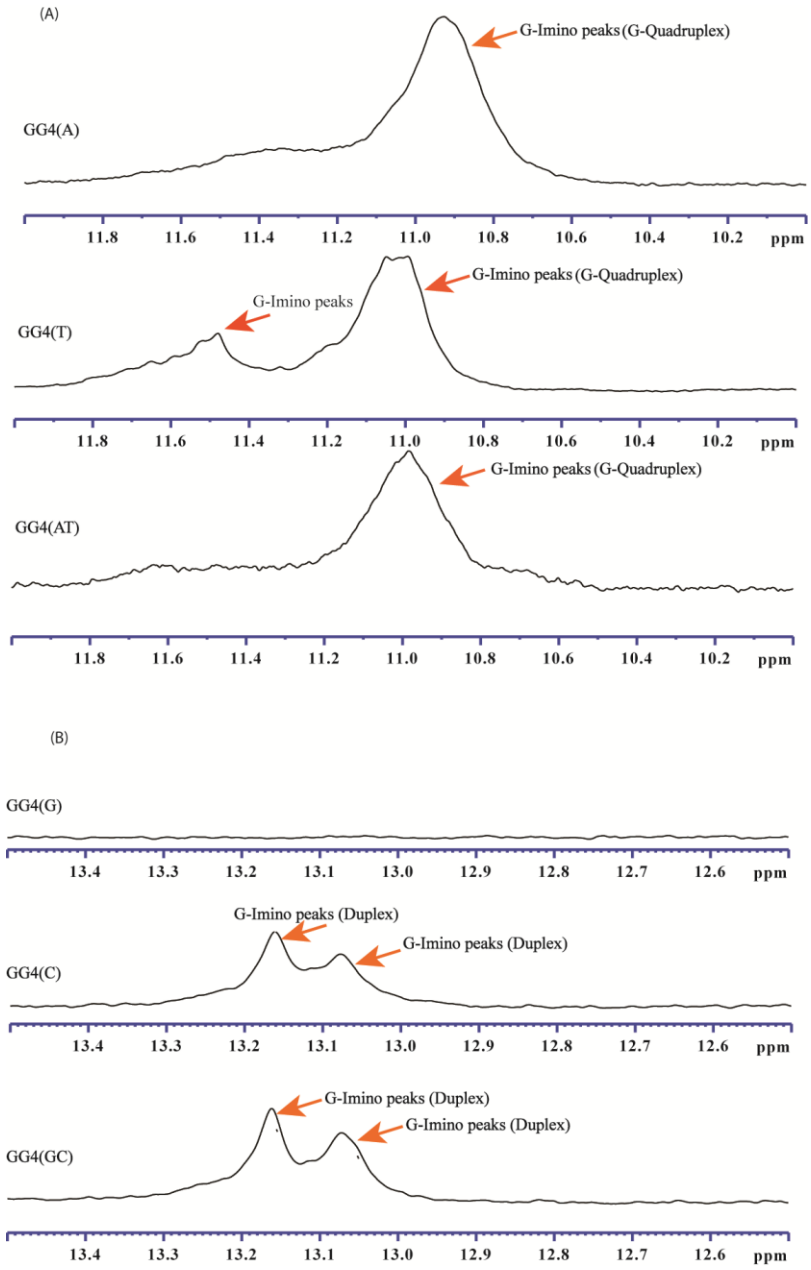


Figure S5.One dimensional proton NMR spectra showing Iminoregion (A) GG4(A/T/AT) (B) GG4(G/C/GC). Samples were dissolved in 20 mM Tris-HCl, pH 7.4 containing 1 mM EDTA and 100 mMKCl.

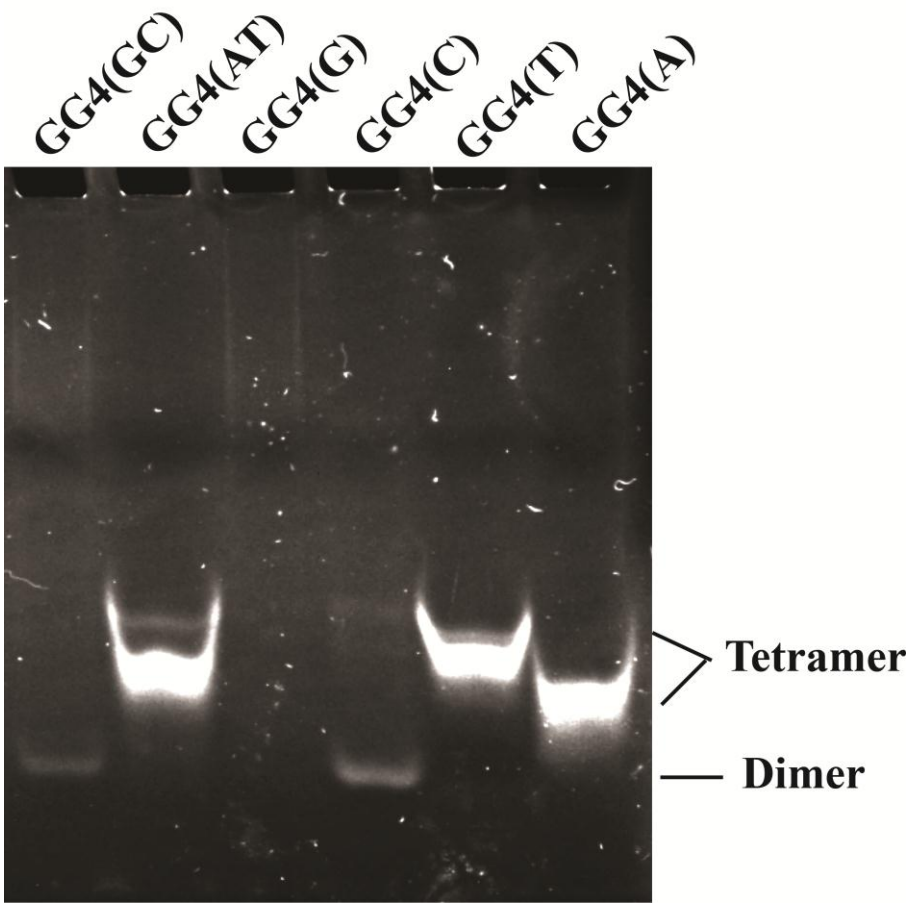


Figure S6. Gel electrophoresis of GG4 sequences

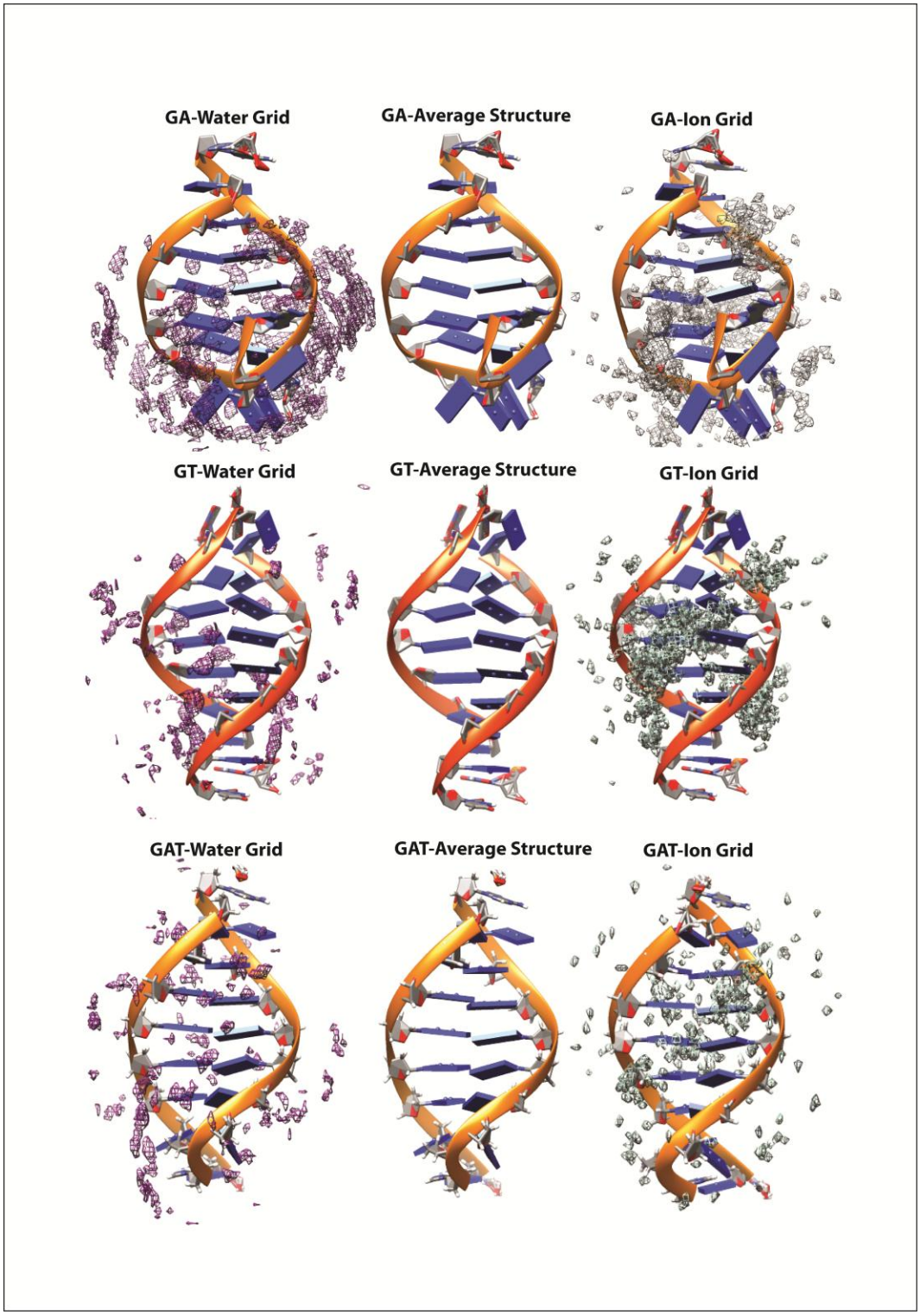


Figure S7. Water molecule grid and sodium ion grid around the average structure of GG4(A), GG4(T) and GG4(AT)

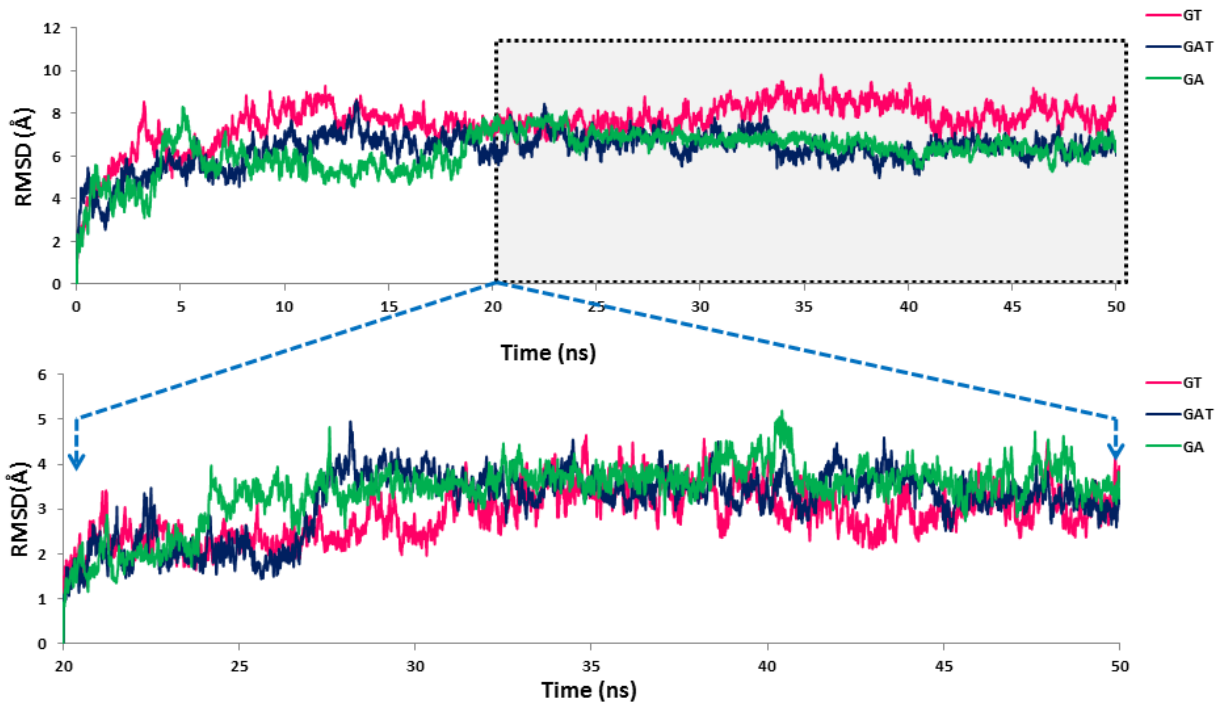


Figure S8.Backbone RMSD of GG4(T), GG4(AT)and GG4(A) is evaluated with respect to time, in top figure, RMSD is calculated for the entire simulation run and reference frame is the structure taken after the equilibration phase. I the bottom figure, RMSD is calculated for last 30 ns simulation run and reference frame is the structure taken after 20 ns simulation run.

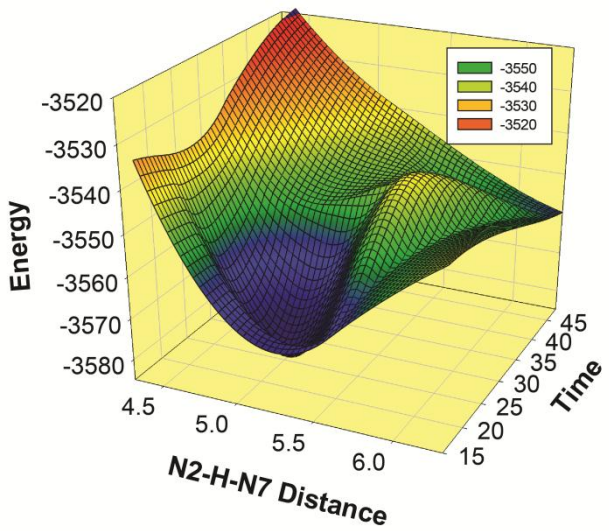
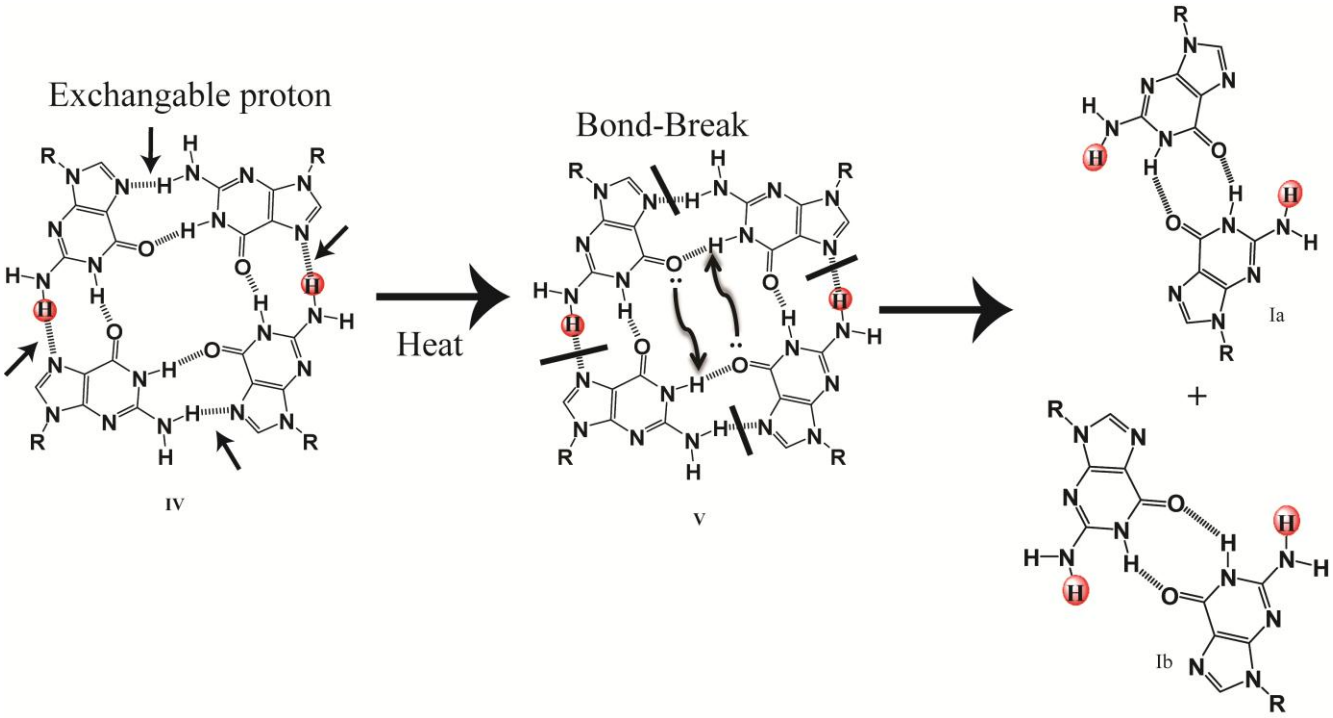
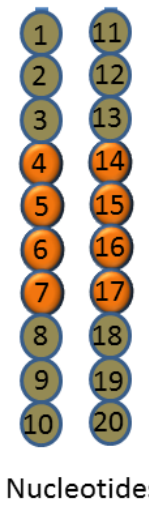


Figure S9.Change in total energy in GG4AT with respect to time and the corresponding change in N2H-N7 distance



Scheme S2.Proposed mechanism of G-quartet breakage to building blocks G-G reverse Watson Crick pairs

Table S6.Percentoccupancy of Hydrogen Bond with respect to the paired bases



H-Bond Pair in GT		%Occupancy
T1	T11	5.86
T1	T12	6.63
T2	T12	12.76
T3	T12	4.45
T3	T11	3.44
T3	T13	5.11
G4	T12	10.31
G4	T13	28.69
G5	T13	9.09
G4	G14	12.02
G5	G14	36.97
G5	G15	20.2
G6	G15	43.42
G6	G16	11.94
G7	G16	45.97
G7	G17	7.7
T8	G17	53.06
T9	T18	47.01
T10	T19	58.72

H-Bond Pair in GA		%Occupancy
A1	G17	10.42
A2	G16	15.02
A3	G14	11.46
G4	A12	21.98
G5	A11	21.56
G4	G14	19.24
G4	G15	9.35
G5	G14	11.96
G5	G15	48.82
G5	G16	11.96
G6	G16	41.61
G7	G17	47.22
A8	A18	13.17
A8	A19	3.23
A9	A18	6.67
A9	A19	2.02
A10	A20	5.05

H-Bond Pair in GAT		%Occupancy
A2	T13	80.34
A3	T13	13.02
A3	T12	11.95
G4	T13	5.85
G4	G14	57.77
G5	G14	6.29
G5	G15	63.31
G6	G15	4.93
G6	G16	68.4
G7	G16	2.42
G7	G17	73.11
A8	T18	68
A9	T19	19.96
A10	T20	7.79

Table S7.Total energy content of GA, GAT and GT estimated through MMGBSA calculations

Energy Component (kcal/mol)	GA	GAT	GT
E_{BOND}	168.56 ± 11.50	167.70 ± 8.11	171.08 ± 16.69
E_{ANGLE}	350.83 ± 14.42	344.32 ± 14.30	337.46 ± 6.35
E_{DIHED}	403.72 ± 10.43	437.98 ± 6.10	438.98 ± 8.37
$E_{VDWAALS}$	-385.16 ± 7.82	-351.59 ± 7.21	-356.43 ± 7.24
E_{EEL}	2268.25 ± 36.17	1896.42 ± 45.61	1836.59 ± 45.72
$E_{1\text{-}4\ VDW}$	170.02 ± 3.36	164.56 ± 3.92	160.55 ± 5.26
$E_{1\text{-}4\ EEL}$	-1845.64 ± 16.54	-1493.89 ± 14.02	-1150.39 ± 11.69
E_{GB}	-4621.60 ± 34.96	-4408.94 ± 38.91	-4451.68 ± 41.32
E_{SURF}	24.84 ± 0.44	26.97 ± 0.43	27.83 ± 0.36
G_{gas}	1130.57 ± 38.90	1165.50 ± 43.24	1437.86 ± 54.00
G_{solv}	-4596.76 ± 34.81	-4381.97 ± 38.99	-4423.84 ± 41.31
E_{Total}	**-3466.19 ± 20.77**	**-3216.47 ± 11.77**	**-2985.98 ± 22.21**

E_{BOND}	Bond energy	E_{GB}	Polar solvation energy
E_{ANGLE}	Angle energy	E_{SURF}	Non-polar solvation energy
E_{DIHED}	Dihedral Energy	G_{gas}	Total gas phase free energy
$E_{VDWAALS}$	van der Waals energy	G_{solv}	Total solvation free energy
E_{EEL}	Electrostatic energy	E_{Total}	Total Energy
$E_{1\text{-}4\ VDW}$	1-4 van der Waals energy		
$E_{1\text{-}4\ EEL}$	1-4 Electrostatic energy		

References:

1. P. K. Mandal, G. W. Collie, B. Kauffmann and I. Huc, Angewandte Chemie International Edition, 53, 14424-14427.
2. Schrödinger Release 2013-3: Maestro, version 9.6, Schrödinger, LLC, New York, NY, 2013.
3. D. A. Case, V. Babin, J. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham Iii, T. A. Darden, R. E. Duke and H. Gohlke.
4. J. Wang, W. Wang, P. A. Kollman and D. A. Case, Journal of molecular graphics and modelling, 2006, 25, 247-260.
5. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, Journal of computational chemistry, 2004, 25, 1157-1174.
6. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, The Journal of chemical physics, 1983, 79, 926-935.
7. R. J. Loncharich, B. R. Brooks and R. W. Pastor, Biopolymers, 1992, 32, 523-535.
8. J. A. Izaguirre, D. P. Catarello, J. M. Wozniak and R. D. Skeel, The Journal of chemical physics, 2001, 114, 2090-2098.
9. V. Kräutler, W. F. van Gunsteren and P. H. Hünenberger, Journal of computational chemistry, 2001, 22, 501-508.
10. T. Darden, D. York and L. Pedersen, The Journal of chemical physics, 1993, 98, 10089-10092.
11. D. R. Roe and T. E. Cheatham Iii, Journal of Chemical Theory and Computation, 9, 3084-3095.
12. W. Humphrey, A. Dalke and K. Schulten, Journal of molecular graphics, 1996, 14, 33-38.
13. W. L. DeLano, 2002.
14. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, Journal of computational chemistry, 2004, 25, 1605-1612.