Supplementary materials

Involvement of G-triplex and G-hairpin in the multi-pathway folding of human telomeric G-quadruplex

Xi-Miao Hou[1,§,*], Yi-Ben Fu[2,§], Wen-Qiang Wu[1], Lei Wang[1], Fang-Yuan Teng[1], Ping Xie[2], Peng-Ye Wang[2] and Xu-Guang Xi[1,3,*]

[1]College of Life Sciences, Northwest A&F University, Yangling, Shaanxi 712100, China;

[2]Beijing National Laboratory for Condensed Matter Physics and CAS Key Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China;

[3]Laboratoire de Biologie et Pharmacologie Appliquée, Ecole Normale Supérieure de Cachan, Centre National de la Recherche Scientifique, 61 Avenue du Président Wilson, 94235 Cachan, France

[*]To whom correspondence should be addressed. Tel: +86 29 8708 1664; Fax: +86 29 8708 1664; Email: houximiao@nwsuaf.edu.cn

Correspondence may also be addressed to Xu-Guang Xi. Tel: +33 1 4740 7754; Fax: +33 1 4740 7754; Email: xxi01@ens-cachan.fr

[§]Xi-Miao Hou and Yi-Ben Fu contributed equally to this work.

Supplementary Table S1. Sequences of substrates used in the experiments.

Substrates for single-molecule FRET (5′-3′)	
	G4 at the 3′ end of duplex DNA
dG4	GCGTGGCACCGGTAATAGGAAATAGGAGAGGGTTAGGGTTAGGGTTAGGG (Cy3)
dG3	GCGTGGCACCGGTAATAGGAAATAGGAGAGGGTTAGGGTTAGGGTTTTTT (Cy3)
dG3*	GCGTGGCACCGGTAATAGGAAATAGGAGATTTTTTGGGTTAGGGTTAGGG (Cy3)
dG2	GCGTGGCACCGGTAATAGGAAATAGGAGATTTTTTTTTTTTTGGGTTAGGG (Cy3)
d4G2	GCGTGGCACCGGTAATAGGAAATAGGAGATTTTTTTTTTTTTGGGGTTAGGG G(Cy3)
stem	TCTCCT(iCy5)ATTTCCTATTACCGGTGCCACGC-biotin
	Poly T at the 3′ end of duplex DNA
dT_{20}	GCGTGGCACCGGTAATAGGAAATAGGAGATTTTTTTTTTTTTTTTTTT (Cy3)
stem	TCTCCT(iCy5)ATTTCCTATTACCGGTGCCACGC-biotin
	G4 at 5′ end of duplex DNA
G4d	(Cy3)GGGTTAGGGTTAGGGTTAGGGATGTATGACAAGGAAGG
G3d	(Cy3)GGGTTAGGGTTAGGGTTTTTTATGTATGACAAGGAAGG
G2d	(Cy3)GGGTTAGGGTTTTTTTTTTTTATGTATGACAAGGAAGG
stem	biotin-CCTTCCTTGTCAT(iCy5)ACAT
	G4 at 3′ end of partial duplex DNA
dsG4	GCGTGGCACCGGTAATAGGAAATAGGAGAT$_{12}$GGGTTAGGGTTAGGGTTAGGG (Cy3)
stem	(Cy5)TCTCCTATTTCCTATTACCGGTGCCACGC-biotin
	G4 at 5′ end of partial duplex DNA
G4sd	(Cy3)GGGTTAGGGTTAGGGTTAGGGT$_{12}$ ATGTATGACAAGGAAGG
stem	biotin-CCTTCCTTGTCATACAT(Cy5)
	Substrates for CD measurement (5′-3′)
3G4	GGGTTAGGGTTAGGGTTAGGG
3G3	GGGTTAGGGTTAGGG
3G2	GGGTTAGGG
s_{12}	CTCTGCTCGACG
	Substrates for gel assay
T_{12}-FAM	TTTTTTTTTTTT(FAM)
3G2	GGGTTAGGG
4G2	GGGGTTAGGGG
$s_{12}s_6$-FAM	CGTCGAGCAGAGTTTTTT(FAM)
s_{12}	CTCTGCTCGACG

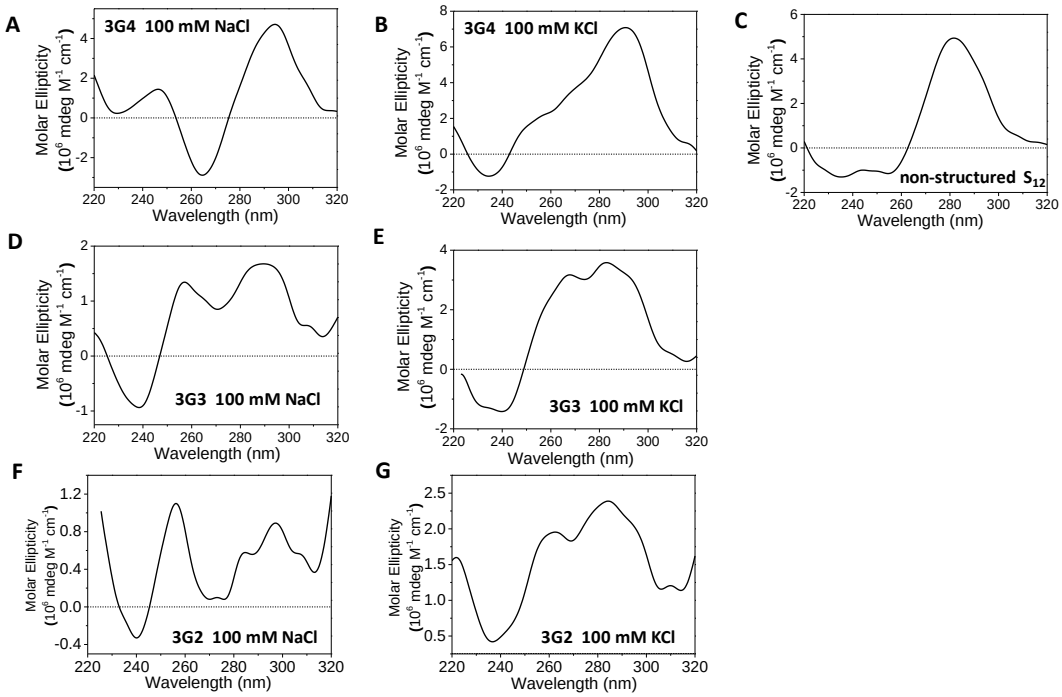
Supplementary Table S2. Dye parameters

Dye	Radius1 (Å)	Radius2 (Å)	Radius3 (Å)	Linker length (Å)	Linker width (Å)
Cy3	6.8	3.0	1.5	20	4.5
Cy5	11	3.0	1.5	22	4.5

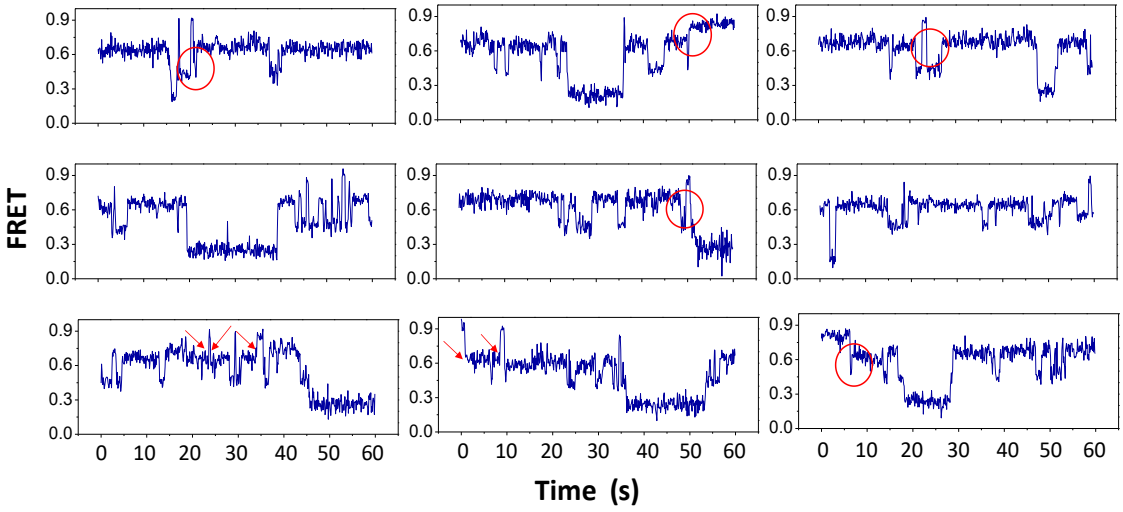
Supplementary Table S3. The interconversion rate between different states (sec).

	buffer	G4-G3	G3-G4	G4-G2	G2-G4	G3-G2	G2-G3	G2-ssDNA	ssDNA-G2
dG4	NaCl	4.53	12.59	3.77	2.52	7.00	1.52	1.55	8.16
dG4	KCl	13.32	17.24	15.54	2.38	9.30	3.60	1.86	4.57
dG3	NaCl					13.29	16.1	15.90	11.98
dG3	KCl					15.16	18.58	17.68	11.08
dG2	NaCl							15.17	7.43
dG2	KCl							22.44	8.08

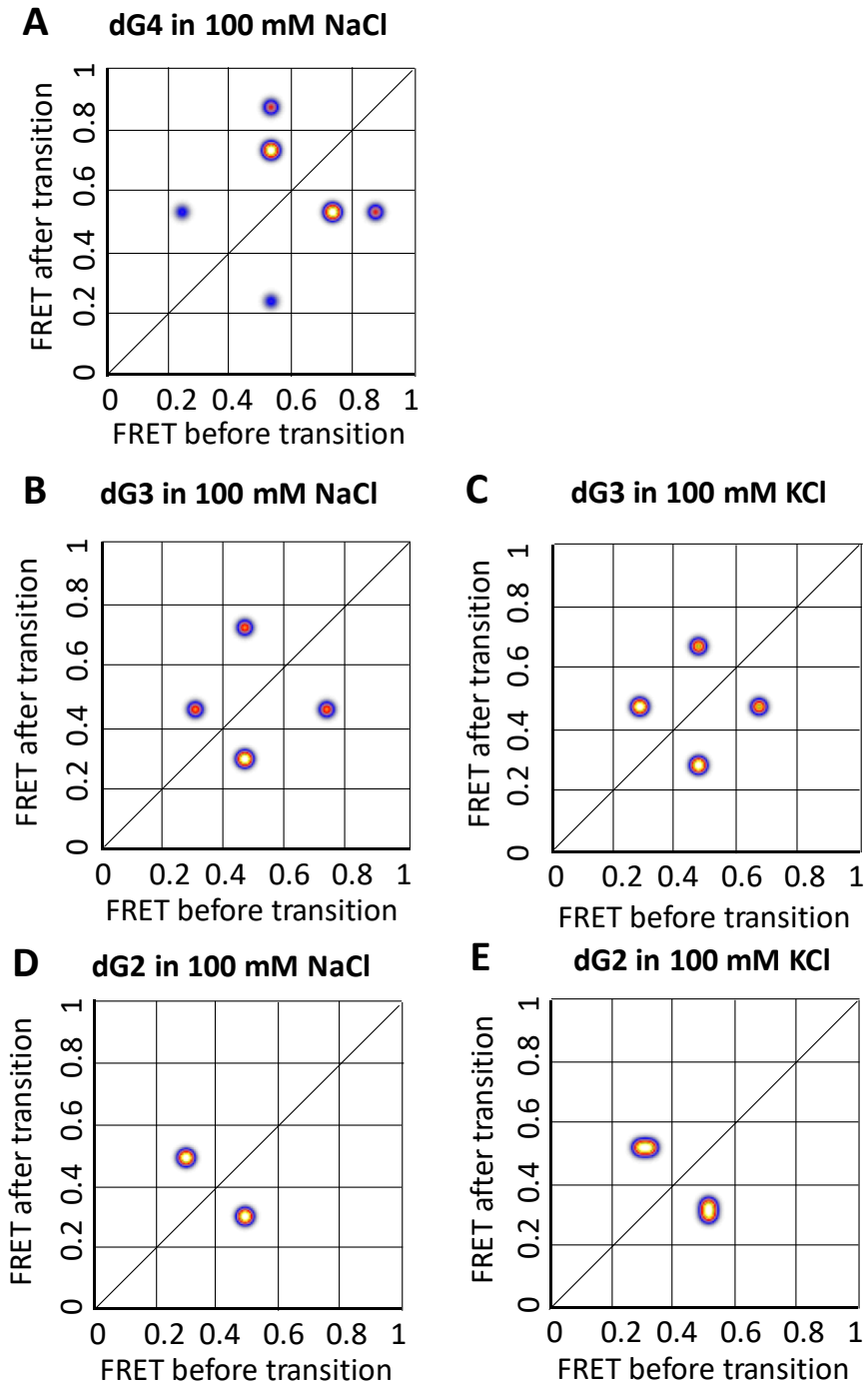
G3, G2 denote G-triplex and G-hairpin respectively. The time was obtained from dG4, dG3 and dG2 traces in 100 mM NaCl or KCl showing dynamic transitions.



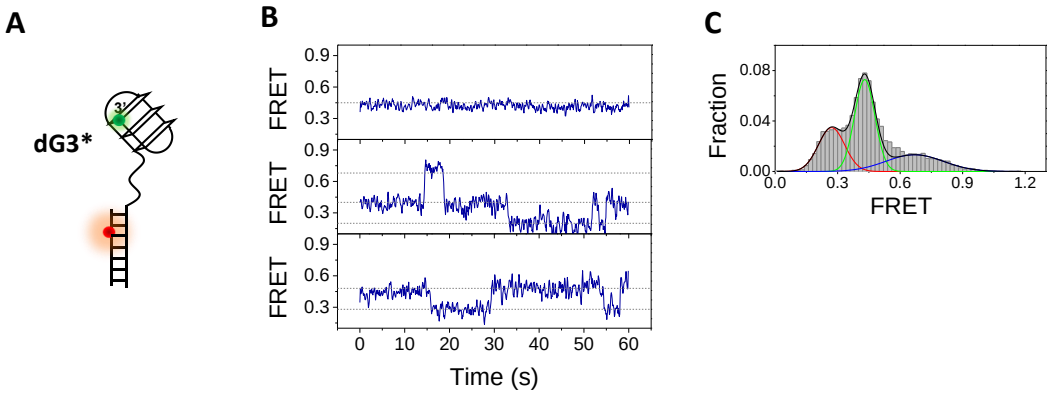
Supplementary Figure S1. CD spectrum. All DNA concentrations were 2 μM. **(A)** In 100 mM NaCl, CD spectrum of 3G4 has one positive peak at ~295 nm and one negative peak at ~265 nm. **(B)** In 100 mM KCl, CD spectrum of 3G4 has one dominant positive peak at ~290 nm. **(C)** For the non-structured ssDNA, in 100 mM KCl, CD spectrum has one peak at ~280 nm. **(D)** In 100 mM NaCl, CD spectrum of 3G3 has two major positive peaks at ~ 255 nm and 290 nm. **(E)** In 100 mM KCl, CD spectrum of 3G3 has two major positive peaks at ~ 265 nm and 285 nm. **(F)** In 100 mM NaCl, CD spectrum of 3G2 has two major positive peaks at ~ 255 and 295 nm, and two minor peaks at ~ 280 nm and 310 nm. **(G)** In 100 mM KCl, CD spectrum of 3G2 has two major positive peaks at ~ 260 nm and 285 nm.



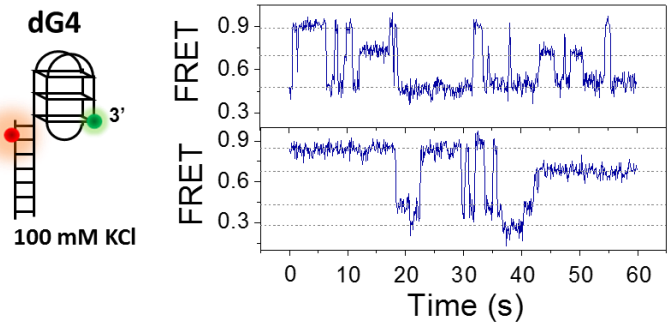
Supplementary Figure S2. Selected smFRET traces of dG4 in 100 mM NaCl showing 4 states. Representative direct transitions between $E_{0.9}$-$E_{0.7}$ are marked by red arrows, and indirect transitions are marked by red circles.



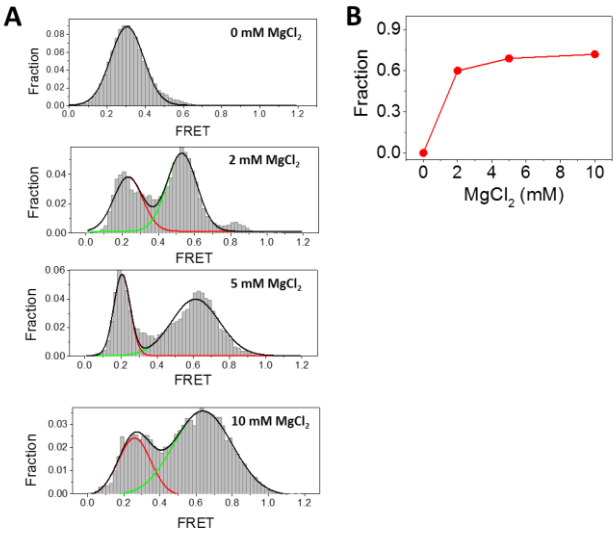
Supplementary Figure S3. Transition density plots for dG4, dG3 and dG2 in 100 mM NaCl or KCl. The color intensity corresponds to transition probability. The x-axis is the initial FRET prior to transition, and the y-axis is the final FRET after transition. (**A**) For dG4 in 100 mM NaCl, the most frequent transitions are between $E_{0.7}$-$E_{0.5}$, transitions between $E_{0.9}$-$E_{0.5}$ are less frequent. There are also a small amount of transitions between $E_{0.5}$-$E_{0.3}$. (**B-C**) For dG3 in 100 mM NaCl or 100 mM KCl, the most frequent transitions are between $E_{0.5}$-$E_{0.3}$, transitions between $E_{0.5}$-$E_{0.7}$ are less frequent. (**D-E**) For dG2 in 100 mM NaCl or 100 mM KCl, the transitions are between $E_{0.5}$-$E_{0.3}$. Transition density plot for dG4 in 100 mM KCl was not shown because the fraction of traces showing dynamic transitions is very low.



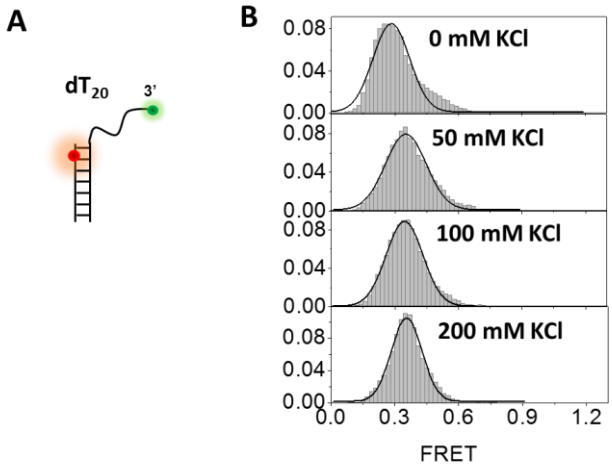
Supplementary Figure S4. Conformational dynamics of G-triplex at the 3′ end of ss/dsDNA. **(A)** Design of dG3* substrate. **(B)** Representative traces of dG3* in 100 mM NaCl. (**C**) FRET histograms of dG3* in 100 mM NaCl. Multi-peak Gaussian distributions were used to fit the histograms.



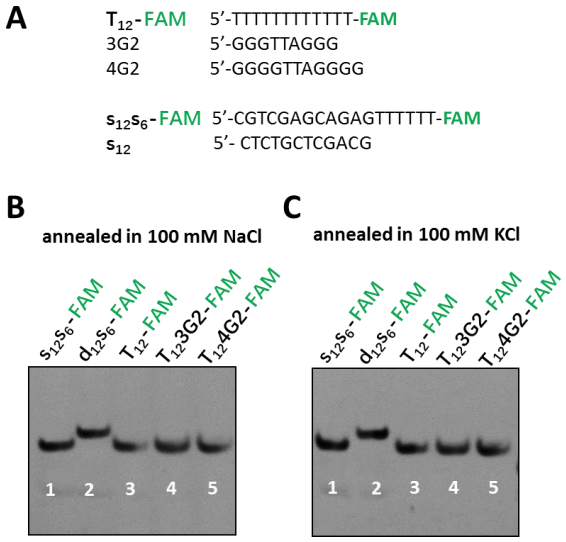
Supplementary Figure S5. In 100 mM KCl, transitions between 3-4 states can be occasionally observed for dG4.



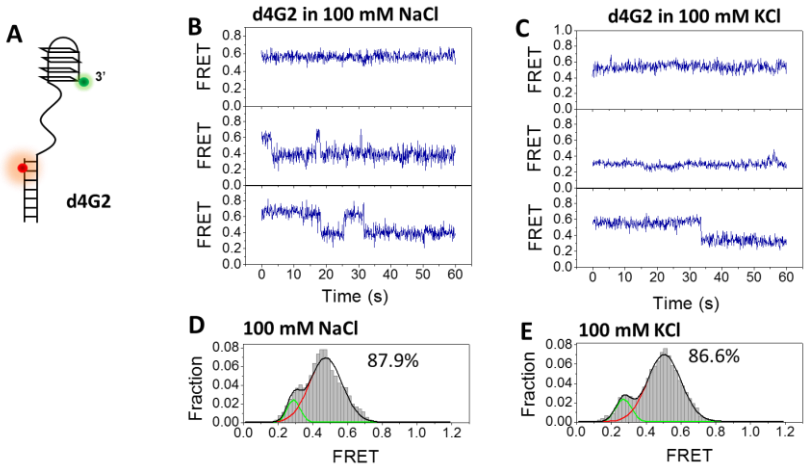
Supplementary Figure S6. Effects of $MgCl_2$ on the folding of G-hairpin. (A) FRET distributions of dG2 in 0-10 mM $MgCl_2$. Single-peak and two-peak Gaussian distributions were used to fit those histograms. **(B)** Fractions of G-hairpin in different concentrations of $MgCl_2$.



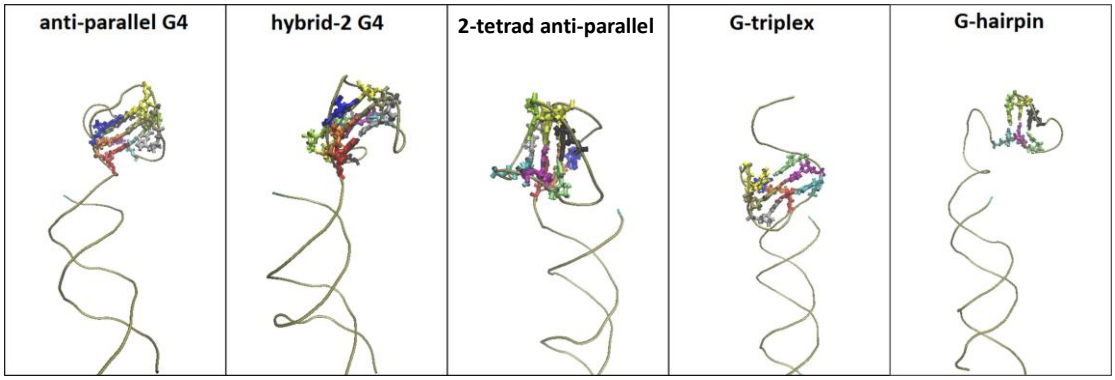
Supplementary Figure S7. Poly-T sequence in different concentration of KCl. (**A**) Schematic representation of experimental design for dT_{20}. **(B)** FRET distributions of dT_{20} in 0, 50, 100 and 200 mM KCl respectively. With the increasing of KCl concentration, there is no additional band at higher FRET value, suggesting no higher order structure has been formed.



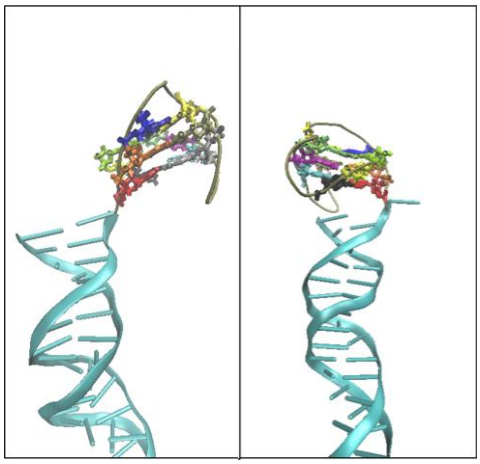
Supplementary Figure S8. In dG2 construct, there are no stable base pairs between the poly-T linker and G-rich part. (A) DNA sequences used in the gel assay. Poly-T sequence is labeled by FAM at 3' end for visualization. If poly-T and G-rich sequences can form stable base pairs, T_{12}3G2-FAM and T_{12}4G2-FAM (the annealed products of T_{12}-FAM and 3G2 or 4G2) will migrate slower in native PAGE gel. Difference between $s_{12}s_6$-FAM and $d_{12}s_6$-FAM (the annealed products of $s_{12}s_6$-FAM and s_{12}) is a positive control to show the reliability of gel analysis.



Supplementary Figure S9. G-hairpin with longer G-rich sequence at the 3′ end of ss/dsDNA in the presence of 100 mM NaCl or KCl. (**A**) Design of d4G2 substrate. **(B-C)** In 100 mM NaCl or 100 mM KCl, d4G2 is either keeping static or going through dynamic transitions between two FRET states at ~0.5 and ~0.3. (**D**) smFRET histograms of d4G2 in 100 mM NaCl display two peaks at $E_{0.29}$ and $E_{0.47}$. (**E**) smFRET histograms of d4G2 in 100 mM KCl display two peaks at $E_{0.27}$ and $E_{0.51}$. The fractions of folded G-hairpin are 87.9% and 86.6% in NaCl and KCl.



Supplementary Figure S10. Representative structures of different G4 folding states obtained from MD simulations. The positions of both anti-parallel and hybrid G4 tend to be at one side of duplex DNA, and the direction of G-tetrad plane is almost perpendicular to the long axis of duplex. 2-tetrad anti-parallel G4 is predominantly tilted relative to duplex, and G-tetrad plane is almost parallel to duplex. The orientation of G-triad plane in G-triplex tends to be almost perpendicular to long axis of duplex.

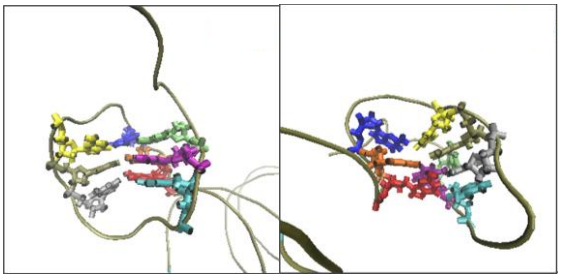


Supplementary Figure S11. Selected hybrid-2 G4 structures at different time during simulation. In left panel, G4 is at one side of duplex. In right panel, G4 is on the top of duplex. The first base pair of duplex is disrupted.

A

GCGTGGCACCGGTAATAGGAAATAGGAGAGGGTTAGGGTTAGGGTTTTTT
TCTCCTATTTCCTATTACCGGTGCCACGC

B

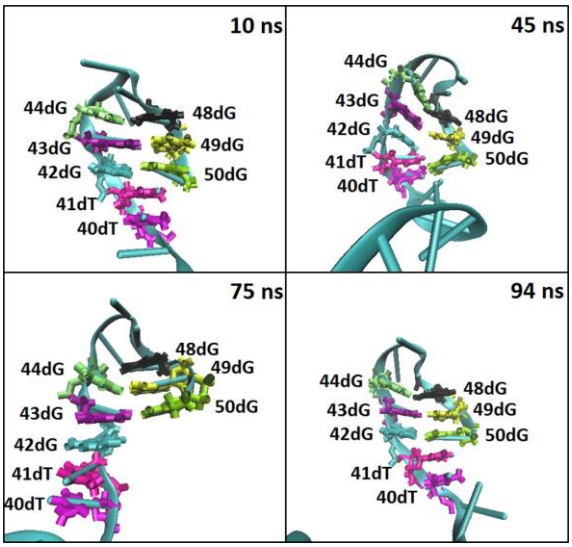


Supplementary Figure S12. MD simulations of G-triplex in dG3 construct. (A) DNA sequences for dG3. Relevant nucleotides of anti-parallel G4 were replaced with dTs, and energy-minimization was run for 100 ps before further MD simulations. (B) There are some subtle changes in the structure of G-triplex during simulation. For example, in left panel GGGTTAGGGTTAGGG sequence forms three layered triplex by well-organized G:G:G triad plane; at right panel G-triad plane seems deformed to some extent. Nevertheless, G-triplex structure keeps being folded without disruption.

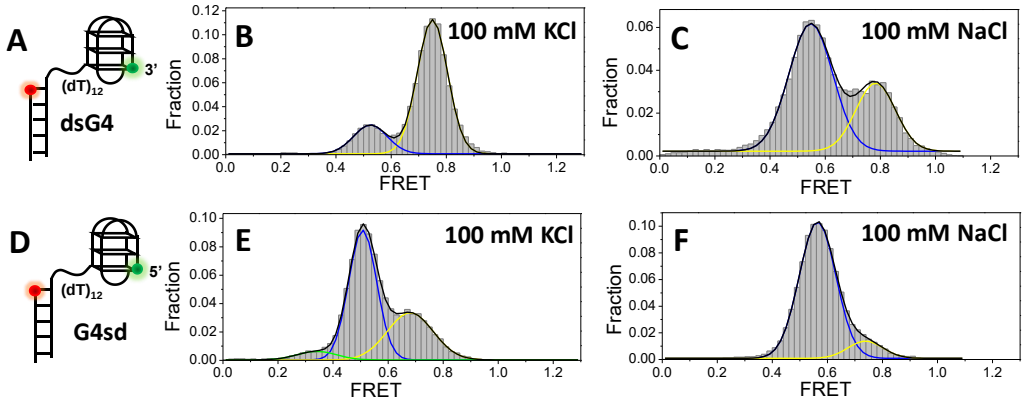
A

40 42 44 46 48 50
GCGTGGCACCGGTAATAGGAAATAGGAGATTTTTTTTTTTTGGGTTAGGG
TCTCCTATTTCCTATTACCGGTGCCACGC
41 43 45 47 49

B



Supplementary Figure S13. MD simulations of G-hairpin in dG2 construct. (A) DNA sequences for dG2. Relevant nucleotides of anti-parallel G4 were replaced with dTs, and energy-minimization was run for 100 ps before further MD simulations. (B) Selected G-hairpin structures during simulation. At 10 ns and 94 ns, G-hairpins show three G-G Hoogsteen base pairs. However, G-hairpins appear to only have two G-G base pairs at 45 ns and 75 ns, and there is even one transient T-G base pair at 45 ns. Therefore, G-hairpin may have different forms based on different base paring. However, no G-hairpins of pure G-T base pairs have been observed.



Supplementary Figure S14. The folding of G4 linked to a partial duplex DNA. (A) Design of G4 at the 3′ end of ss/dsDNA, referred as dsG4. **(B)** In 100 mM KCl, the FRET distribution of dsG4 has two peaks at $E_{0.53}$ and $E_{0.75}$. **(C)** In 100 mM NaCl, the FRET distribution of dsG4 has two peaks at $E_{0.55}$ and $E_{0.78}$. **(D)** Design of G4 at the 5′ end of ss/dsDNA, referred as G4sd. **(E)** In 100 mM KCl, the FRET distribution of G4sd has two major peaks at $E_{0.51}$ and $E_{0.68}$, and one minor peak at $E_{0.33}$. **(F)** In 100 mM NaCl, the FRET distribution of G4sd has one major peak at $E_{0.57}$ and one minor peak at $E_{0.75}$. Either in KCl or NaCl buffer, dsG4 prefers to fold into the conformation with higher FRET value, compared with G4sd.