

# Single-Molecule Analysis of the Improved Variants of the G-Quadruplex Recognition Protein G4P

Department of Biochemistry and Molecular Biology, Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA; masayoshi-honda@uiowa.edu (M.H.)

DYKDDDDKSSGHHHHHHHSSGLVPRGSHMQDSPEKTTLAAKLAKKQASIYRGTGSGAGTGSAG  
TGSGAQDSPEKTTLAAKLAKKQASIYRGTGSGAGTGSAGTGSAGTGSAGTGSAGTGSAGTGSAGQD  
SPEKTTLAAKLAKKQASIYRGTGSGAGTGSAGTGSAGTGSAGQDSPEKTTLAAKLAKKQASIYRGTGS  
GA\*

**Supplementary Table S1. Sequences of DNA used in this study.**

Name and purpose	Sequence	Structure	Topology
Biotin base strand (for smTIRFM surface tethering)	5'-GCCTCGCTGCCGTCGCCA-bio-3'	ssDNA	NA
<i>c</i> -MYCG4 (for smTIRFM surface tethering)	5'-TGGCGACGGCAGCGAGGCTGGGTGGGTAGGGTGGGTTT -3'	G4	Parallel
hTelG4 (for smTIRFM surface tethering)	5'- TGGCGACGGCAGCGAGGCTTAGGGTTAGGGTTAGGGTTTAGG GTTAGGG-3'	G4	Mixed
2hTelG4 (for smTIRFM surface tethering)	5'- TGGCGACGGCAGCGAGGCTTAGGGTTAGGGTTAGGGTTTAGG GTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG -3'	G4	Mixed
Cy5- <i>c</i> -MYCG4	5'-/5Cy5/TGGGTGGGTAGGGTGGGTTT-3'	G4	Parallel
Cy5-hTelG4	5'-/5Cy5/TTAGGGTTAGGGTTAGGGTTTAGGGTTAGGG-3'	G4	Mixed
Cy5-2hTelG4	5'- /5Cy5/TTAGGGTTAGGGTTAGGGTTTAGGGTTAGGGTTAGGGT TAGGGTTAGGGTTAGGGTTAGGGTTAGGG-3'	G4	Mixed
hTelG4	5'-/5Fluor/TTAGGGTTAGGGTTAGGGTTTAGGGTTAGGG -3'	G4	Mixed
<i>c</i> -MYCG4	5'-/TGGGTGGGTAGGGTGGGTTT3' 6-FAM-3'	G4	Parallel
Cy5 Biotin base strand (smFRET)	5'-Cy5-GCCTCGCTGCCGTCGCCA-bio-3	ssDNA	NA
hTel (smFRET)	5'- TGGCGACGGCAGCGAGGCGGGTTAGGGTTAGGGTTAGGGTT A-Cy3-3'	G4	Parallel

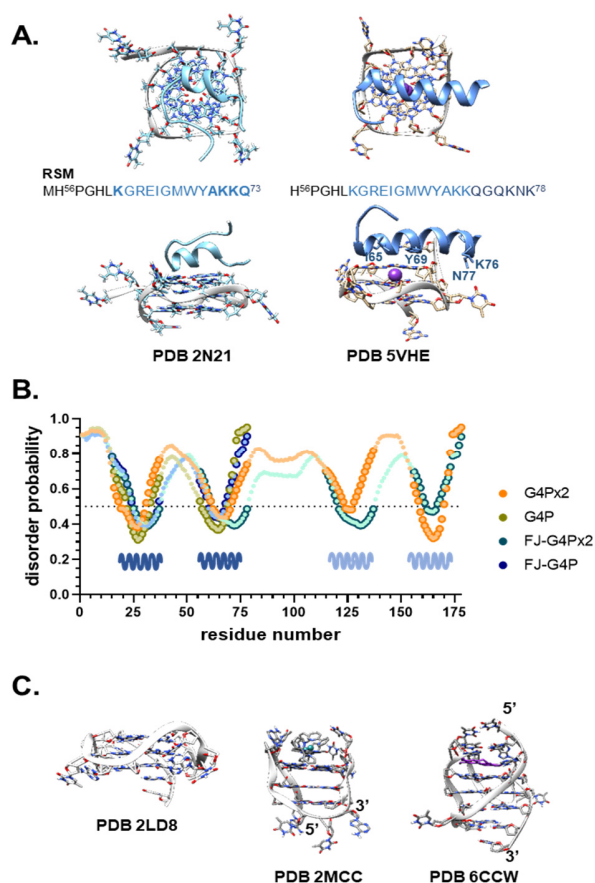
The G4-forming sequences within each oligo are underlined, and sequences complementary to the biotinylated oligo are **in blue**.

**Supplementary Table S2. Comparison of dissociation constant (Kd) values obtained by different methods and from previous studies.** Note: The numbers obtained from single molecule experiments were adjusted based on the Cy3 labeling efficiency of each respective protein.

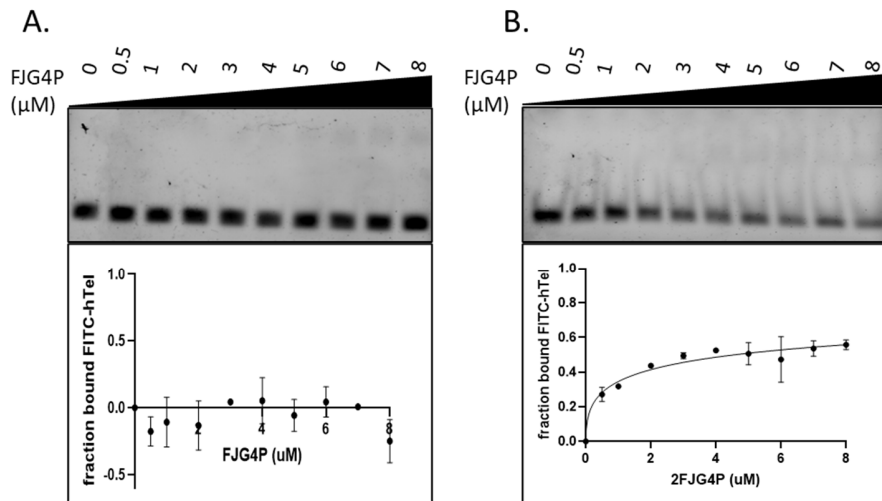
S. No.	Complex (Protein-DNA)	smTIRFM experiments				EMSA Kd (nM)	EMSA Kd Previously published [1] (nM)
		DNA	Protein	Cy3 labeling Efficiency of Protein %	smTIRFM Kd (nM)		
1.	<b>G4P-<i>c</i>-MYCG4</b>	100 pM	150 pM	55%	0.22 ± 0.02	84.14 ± 8 nM	2.7 nM
2.	<b>2G4P-<i>c</i>-MYCG4</b>	100 pM	100 pM	71%	0.16 ± 0.008	27.66 ± 4 nM	NA
3.	<b>G4P-hTelG4</b>	100 pM	200 pM	55%	1.6 ± 0.13	558 ± 156.8 nM	14 nM
4.	<b>2G4P-hTelG4</b>	100 pM	200 pM	71%	3.4 ± 0.33	86.19 ± 6 nM	NA
5.	<b>FJG4P-<i>c</i>-MYCG4</b>	100 pM	200 pM	42%	0.27 ± 0.03	110.2 ± 58.9 nM	NA
6.	<b>2FJG4P-<i>c</i>-MYCG4</b>	100 pM	200 pM	28.5%	0.2 ± 0.01	130.5 ± 63 nM	NA
7.	<b>FJG4P-hTelG4</b>	100 pM	5000 pM	42%	32.7 ± 4.8	NA	NA
8.	<b>2FJG4P-hTelG4</b>	100 pM	2000 pM	28.5%	10.7 ± 1.7	4526 ± 553 nM	NA
9.	<b>G4P-2hTelG4</b>	100 pM	200 pM	55%	1.9 ± 1.4	NA	NA
10.	<b>2G4P-2hTelG4* (3 states)</b>	100 pM	200 pM	71%	NA	NA	NA
11.	<b>FJG4P-2hTelG4</b>	100 pM	2000 pM	42%	14.5 ± 2.2	NA	NA
12.	<b>2FJG4P-2hTelG4</b>	100 pM	1000 pM	28.5%	4.6 ± 0.55	NA	NA

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27																		
A.	T	G	G	G	G	A	G	G	G	T	G	G	G	G	A	G	G	G	T	G	G	G	G	A	A	G	G																		
B.						A	G	G	G	T	G	G	G	G	G	A	G	G	G	T	G	G	G	G																					
C.					T	G	A	G	G	G	T	G	G	G	G	A	G	G	G	T	G	G	G	G	A	A																			
D.						A	G	G	G	T	G	G	G	G	G	A	G	G	G	T	G	G	G	G	A																				
E.					T	G	A	G	G	G	T	G	G	G	T	A	G	G	G	T	G	G	G	T	A	A																			
F.						T	G	G	G	T	G	G	G	G	T	A	G	G	G	T	G	G	G	T	T																				
G.					T	G	A	G	G	G	T	G	G	G	T	A	G	G	G	T	G	G	G	T	G	C	G	C	T	G	C	G	G	C	T	G	G	C	T	C	G	A	G	G	C

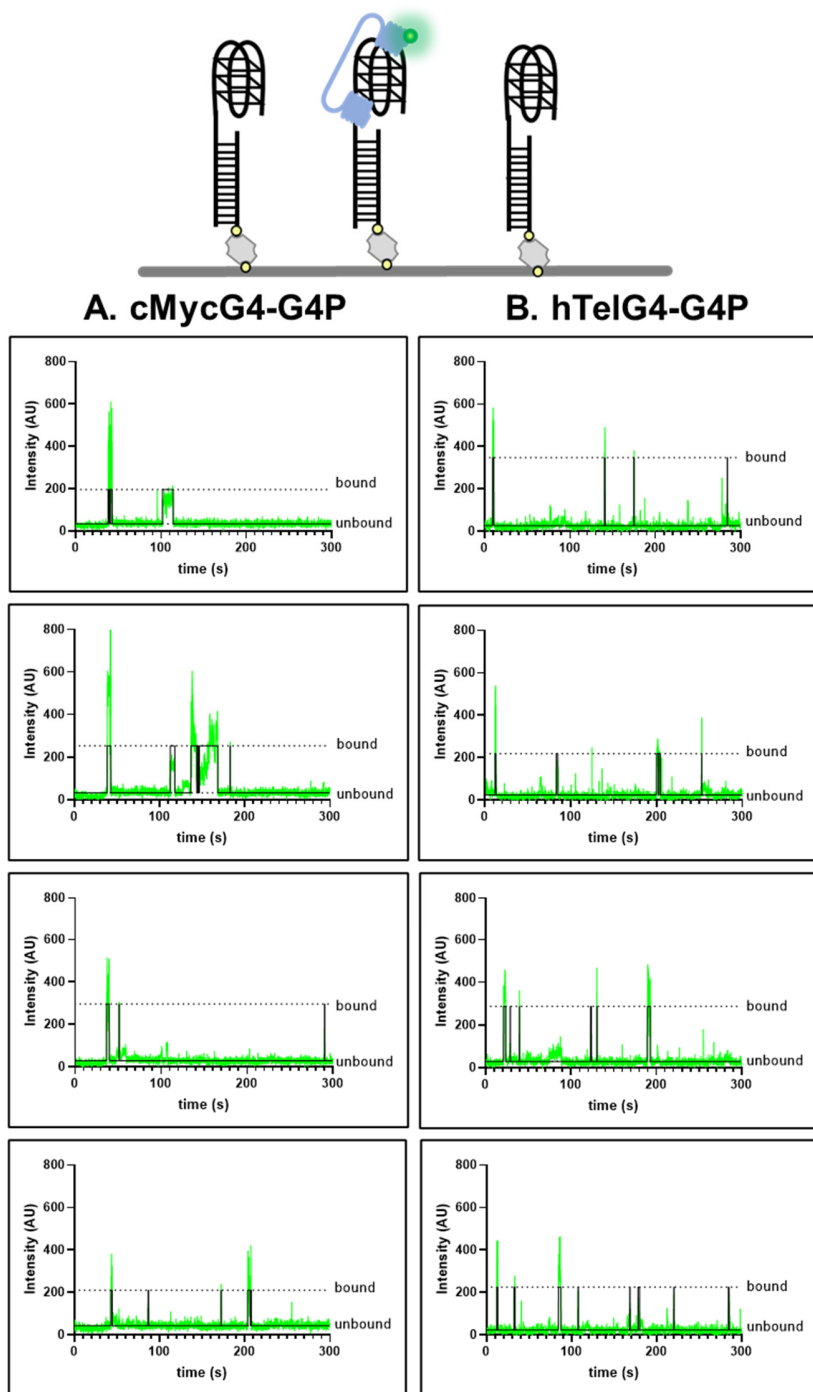
**Supplementary Figure S1.** The promoter sequence of the NHE III1 element of the *c-MYC* gene and its modified versions. A. *c-MYCG4*-Pu27: This is the wild-type 27-mer G-rich sequence of the c-Myc NHE III1. B. *c-MYCG4*-Pu18: an 18-nucleotide sequence derived from Pu27, which excludes the G2-G5 guanine nucleotides that are not involved in quadruplex formation. C. Myc2345: This is a sequence that can form four major loop-isomers, which can be distinguished by introducing dual G-T mutations at positions 14 and 23, 11 and 23, 14 and 20, and 11 and 20, respectively. These are denoted as 14/23, 11/23, 14/20, and 11/20. D. *c-MYCG4*: This is a shorter version of Myc2345 that was used by Zhen et al. (NAR)[1]. E. *c-MYCG4* 14/23: This is a modified version of Myc2345 that forms only one major loop-isomer with the 14/23 mutation. F. *c-MYCG4* 14/23\*: This is a further modified version of *c-MYCG4* 14/23 that was used in this study. G. *c-MYCG4*\*: This is another modified version of *c-MYCG4* that was used in a previous study published in FASEB[2].



**Supplementary Figure S2. In all G4P constructs, the RSM elements fold into  $\alpha$ -helixes connected by disordered regions.** **A.** NMR structure of the RHAU RSM peptide (PDB 2N21) shows different contacts between the RSM  $\alpha$ -helix compared to the longer  $\alpha$ -helix containing RSM when a part of RHAU protein (part of the crystal structure from PDB 5VHE). Structures were visualized in UCSF Chimera 1.14. **B.** Disorder prediction for the four constructs using PrDOS server. **C.** Structures of human telomeric G4 in different configurations: 2LD8 – parallel, 2MCC – anti-parallel, 2CCW – hybrid-2.

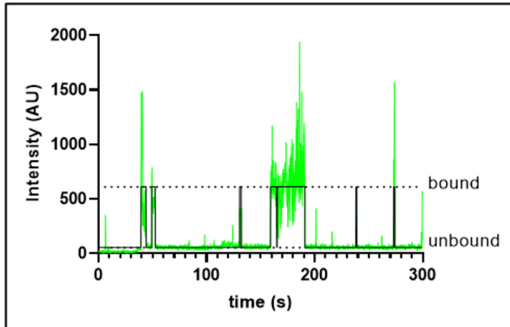
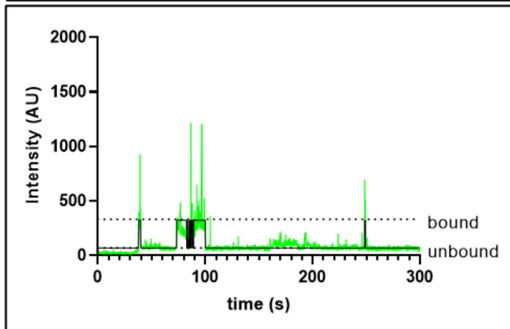
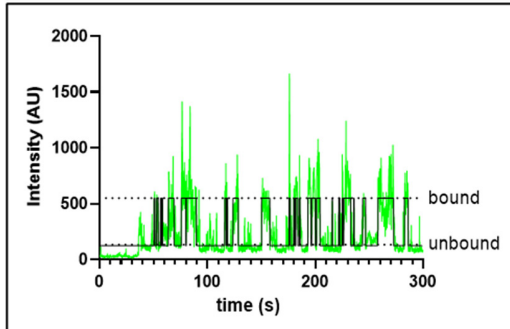
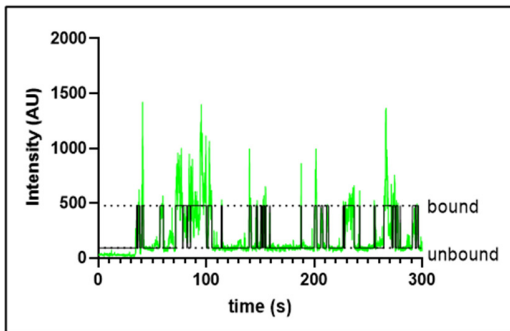


**Supplementary Figure S3. Electrophoretic Mobility Shift Assay (EMSA). A and B.** Indicated concentrations of purified FJG4P and 2FJG4P were incubated with fluorescently-labeled hTelG4 DNA for 10 minutes at room temperature. The samples were resolved on an 8% native polyacrylamide gel and imaged using a ChemiDoc MP imaging system. The bands were quantified using ImageJ software and the data were plotted using GraphPad Prism (n=2).

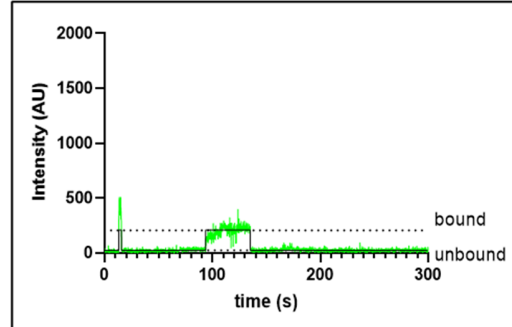
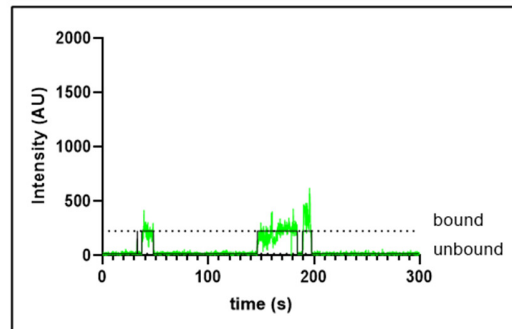
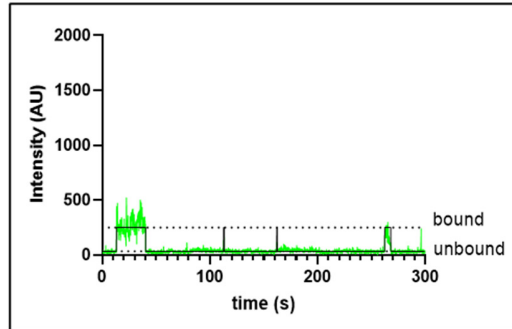
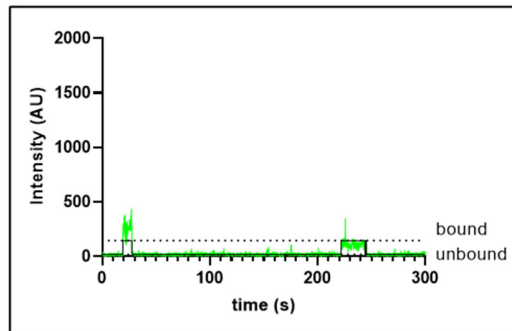


**Supplementary Figure S4. Representative fluorescence trajectories of Cy3-labeled proteins binding to immobilized DNA structures.** Unlabeled DNA was immobilized in the TIRFM reaction chamber, and Cy3-labeled proteins were flown into the chamber. The time-dependent changes in Cy3 fluorescence were recorded, with raw data shown in green and an idealized trajectory shown in black. Panel **A** and **B** show the 4 representative fluorescence trajectories for *c*-MYCG4-G4P, and hTelG4-G4P each, respectively.

### A. cMycG4-2G4P



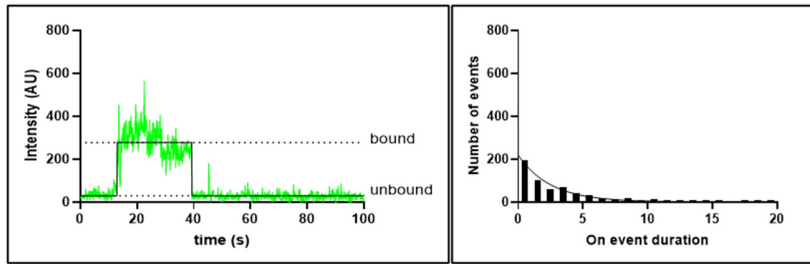
### B. hTelG4-2G4P



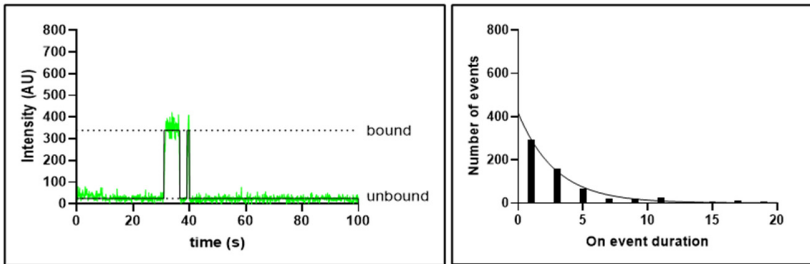
**Supplementary Figure S5. Representative fluorescence trajectories of Cy3-labeled proteins binding to immobilized DNA structures.** Unlabeled DNA was immobilized in the TIRFM reaction chamber, and Cy3-labeled proteins were flown into the chamber. The time-dependent changes in Cy3 fluorescence were recorded, with raw data shown in green and an idealized trajectory shown in black. Panel **A** and **B** show the 4 representative fluorescence trajectories for *c*-MYCG4-2G4P, and hTelG4-2G4P each, respectively.



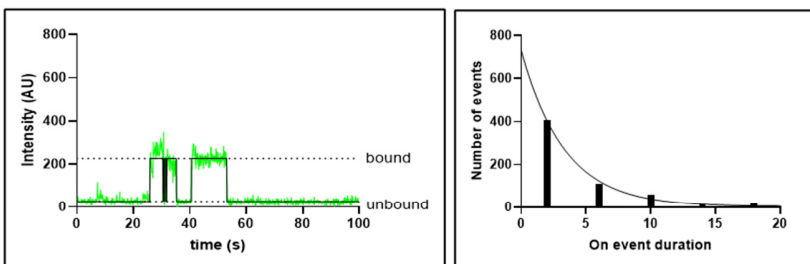
### A. 2hTelG4-G4P



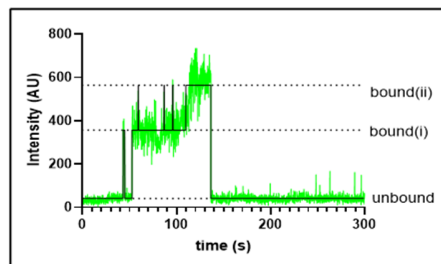
### B. 2hTelG4-FJG4P



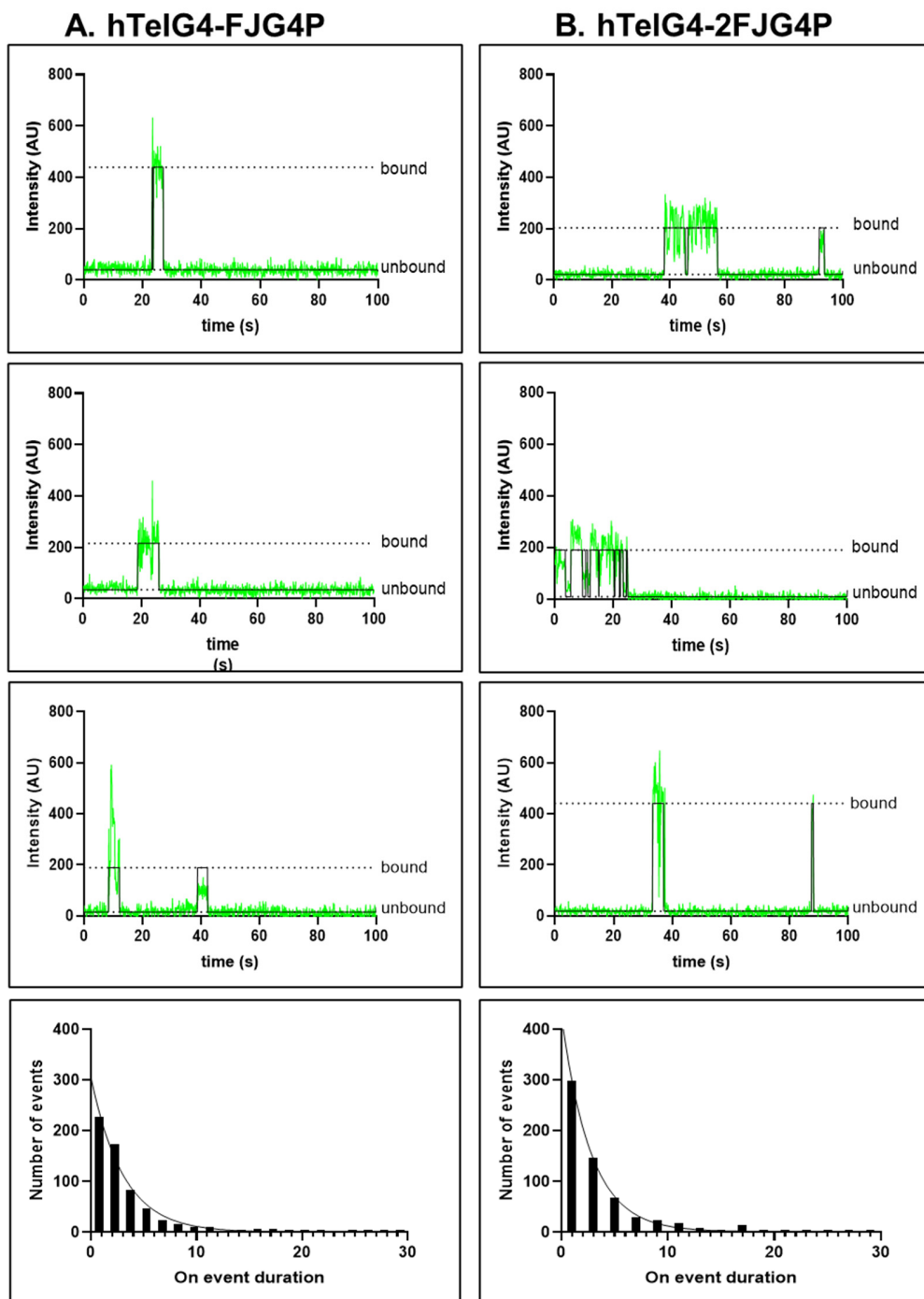
### C. 2hTelG4-2FJG4P



### D. 2hTelG4-2G4P

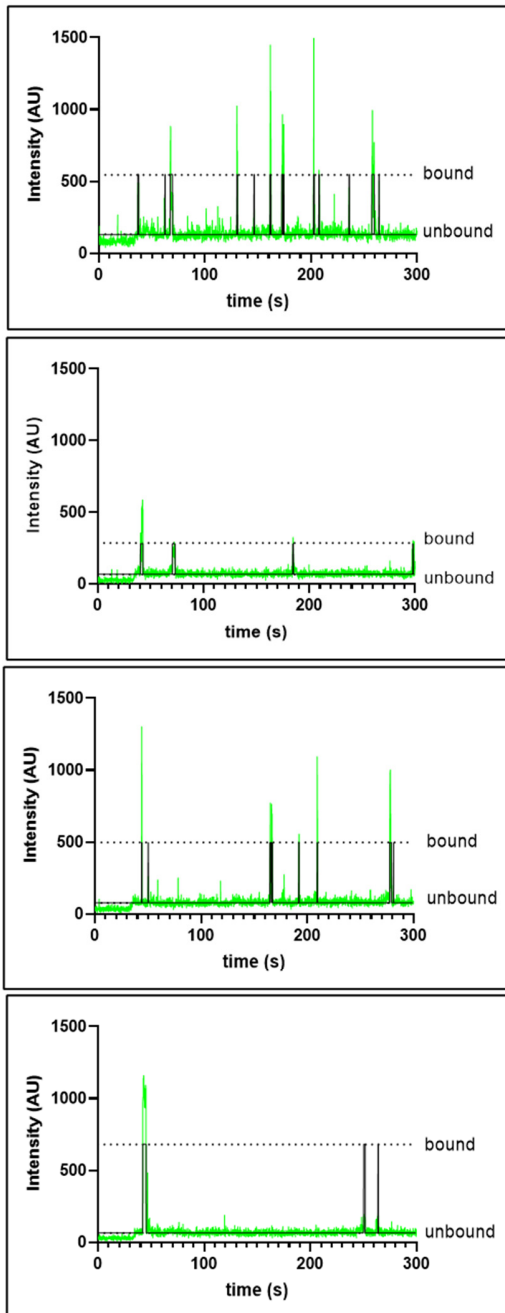


**Supplementary Figure S6. Representative fluorescence trajectories of Cy3-labeled proteins binding to immobilized DNA structures and the dwell-time distribution of all "ON" events.** Unlabeled DNA was immobilized in the TIRFM reaction chamber, and Cy3-labeled proteins were flown into the chamber. The time-dependent changes in Cy3 fluorescence were recorded, with raw data shown in green and an idealized trajectory shown in black. Panel **A**. 2hTel-G4P, **B**. 2hTel-FJG4P, **C**. 2hTel-2FJG4P and **D**. 2hTel-2G4P showing 3 state binding model.

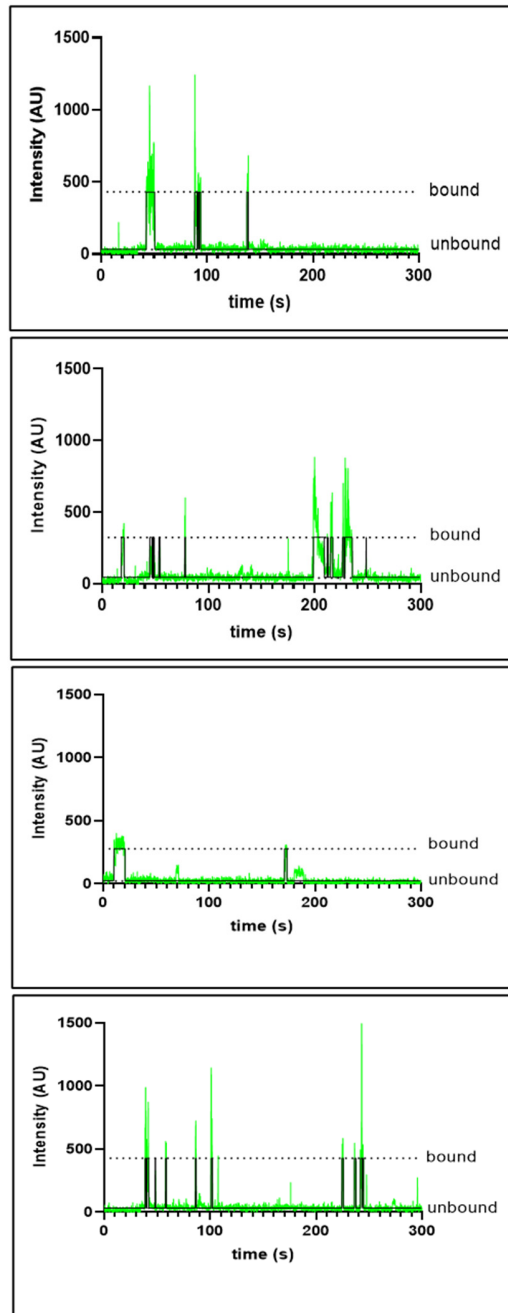


**Supplementary Figure S7. Representative fluorescence trajectories of Cy3-labeled proteins binding to immobilized DNA structures.** Unlabeled DNA was immobilized in the TIRFM reaction chamber, and Cy3-labeled proteins were flown into the chamber. The time-dependent changes in Cy3 fluorescence were recorded, with raw data shown in green and an idealized trajectory shown in black. Panel **A.** and **B.** represent three represented fluorescence trajectories and the dwell-time distribution of all "ON" events for hTelG4-FJG4P and hTelG4-2FJG4P, respectively.

### A. cMycG4-FJG4P

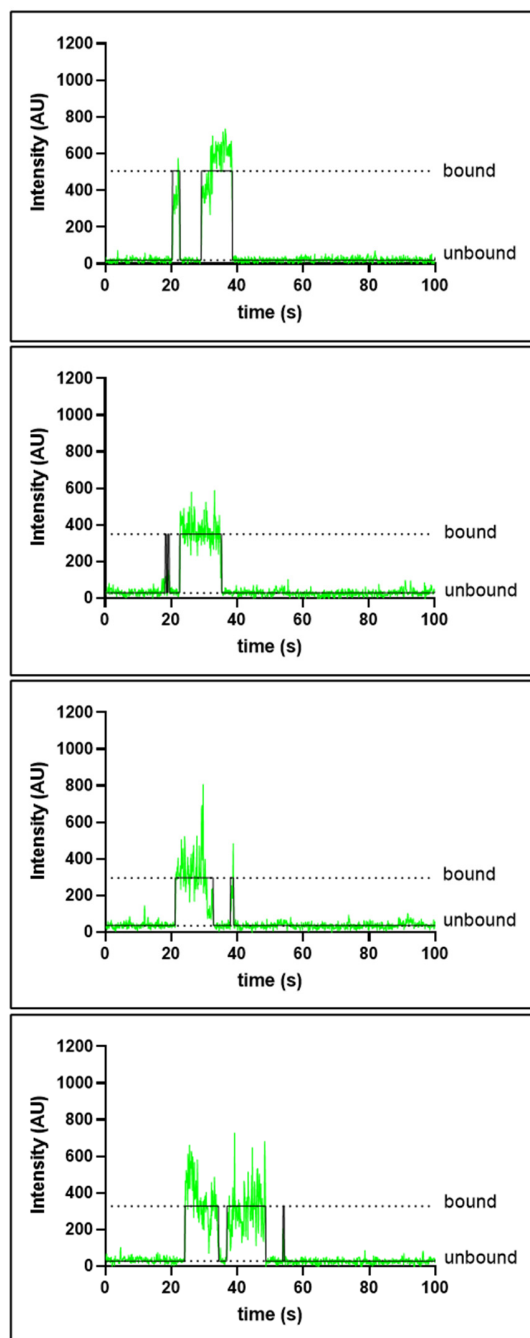


### B. cMycG4-2FJG4P

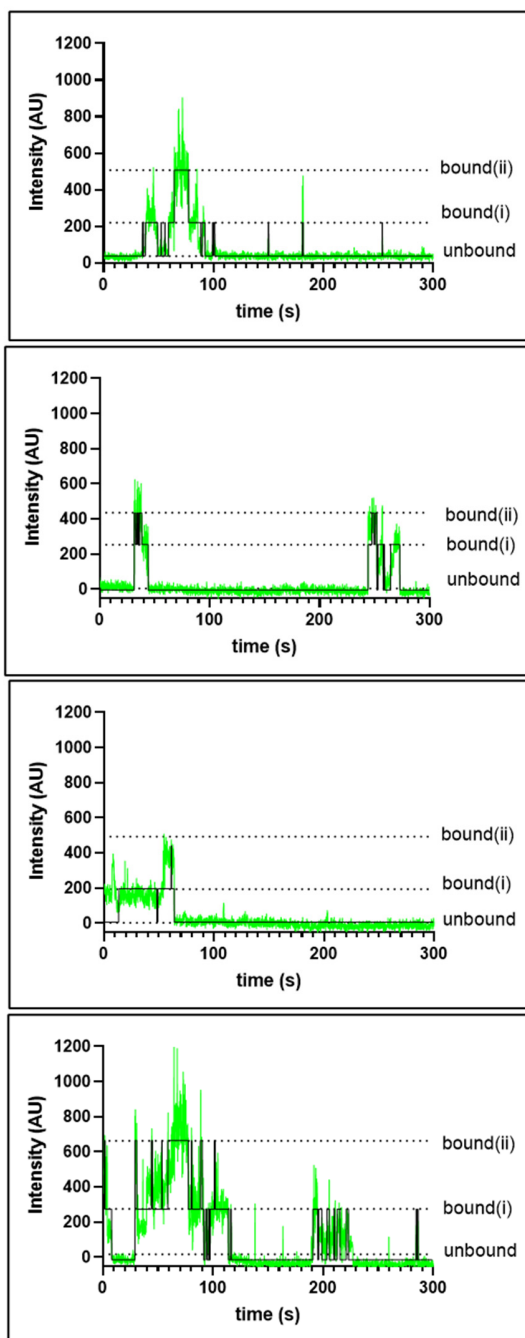


**Supplementary Figure S8. Representative fluorescence trajectories of Cy3-labeled proteins binding to immobilized DNA structures.** Unlabeled DNA was immobilized in the TIRFM reaction chamber, and Cy3-labeled proteins were flown into the chamber. The time-dependent changes in Cy3 fluorescence were recorded, with raw data shown in green and an idealized trajectory shown in black. Panel A and B show the 4 representative fluorescence trajectories for *c*-MYCG4-FJG4P, and *c*-MYCG4-2FJG4P each, respectively.

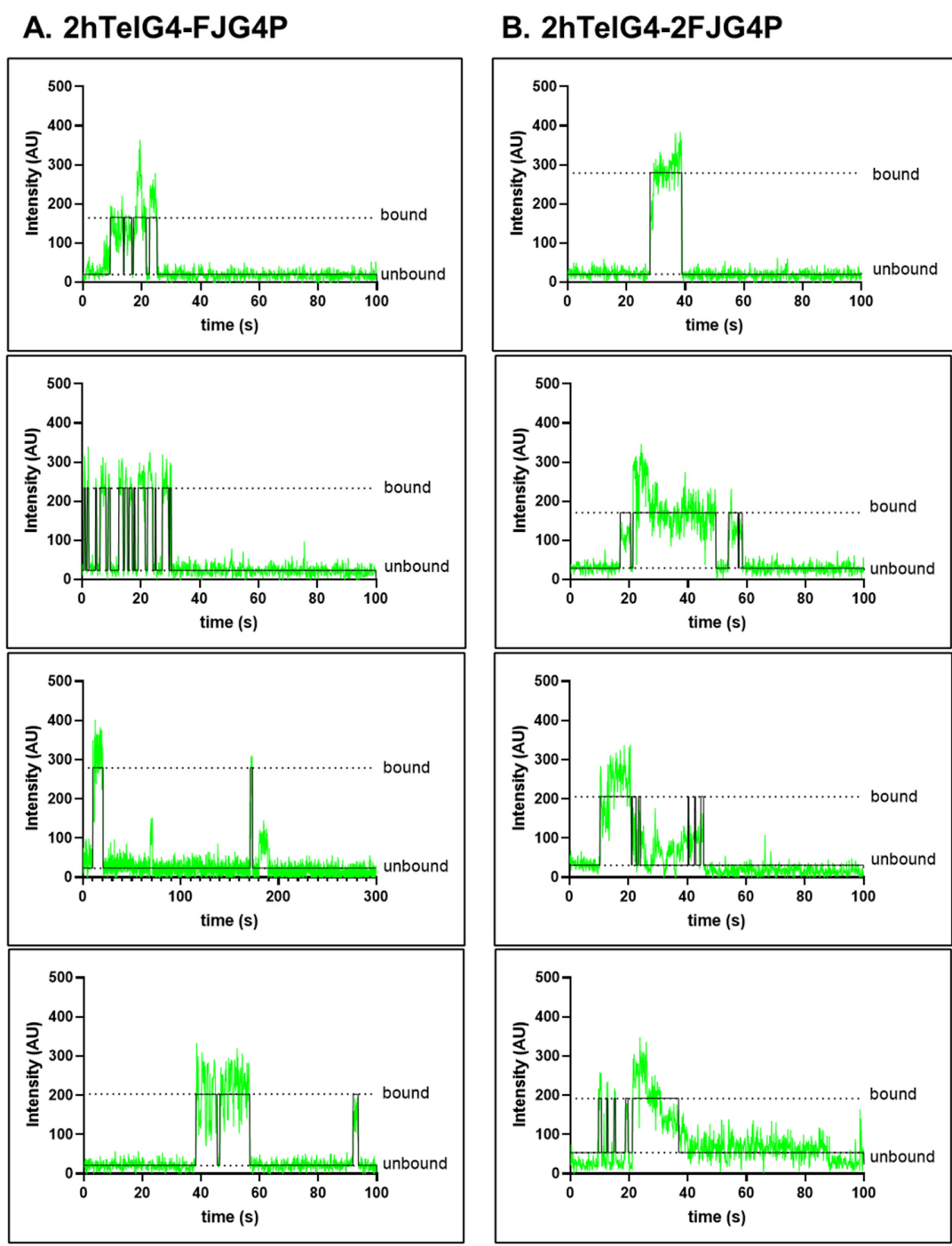
### A. 2hTelG4-G4P



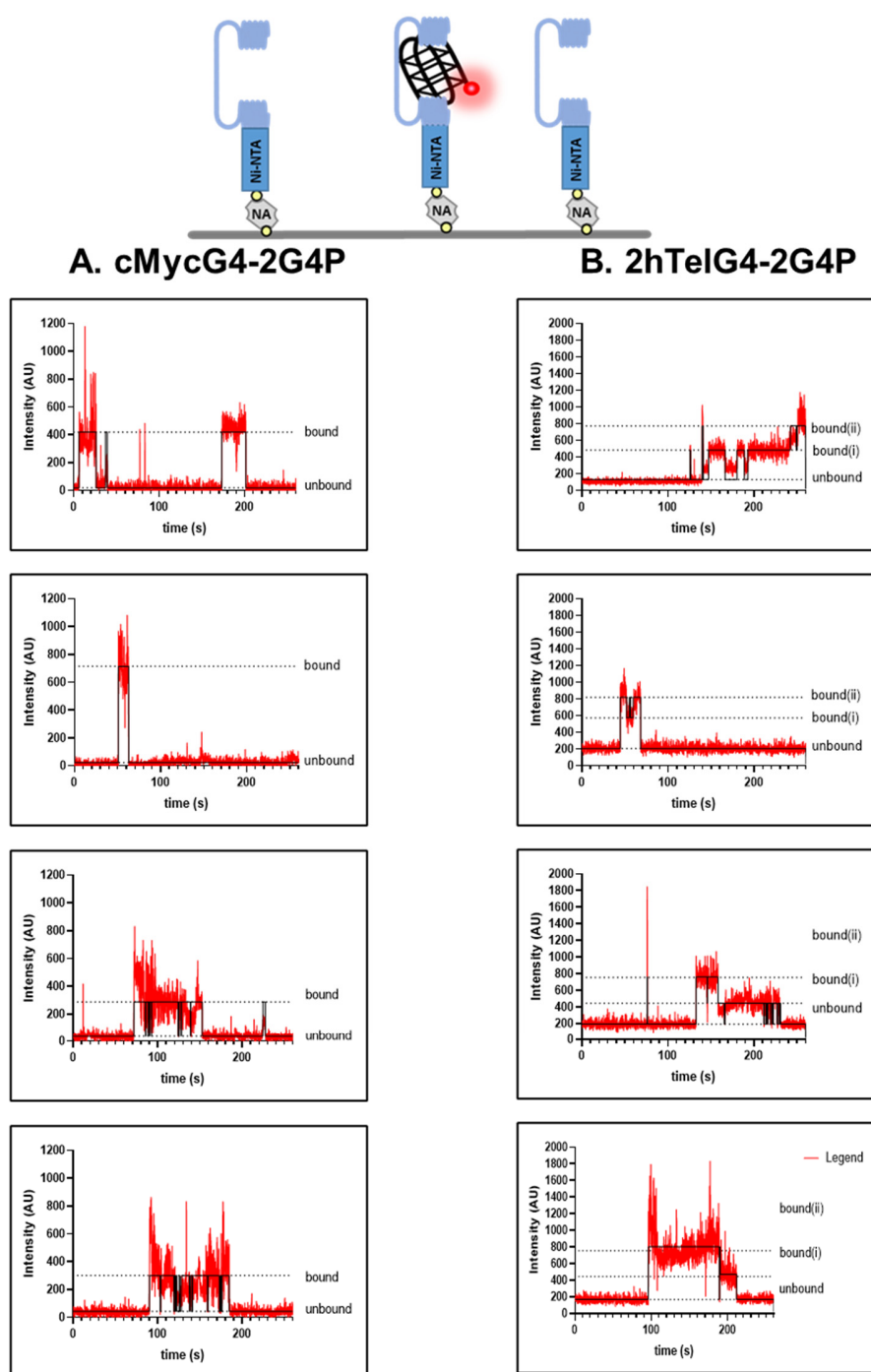
### B. 2hTelG4-2G4P



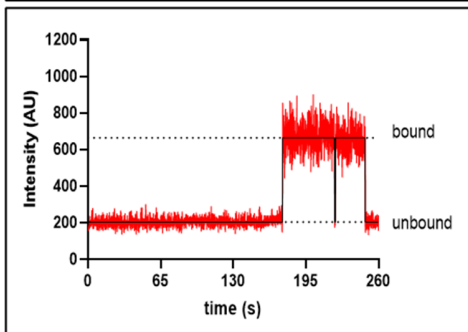
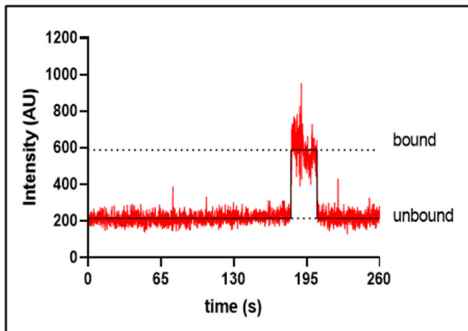
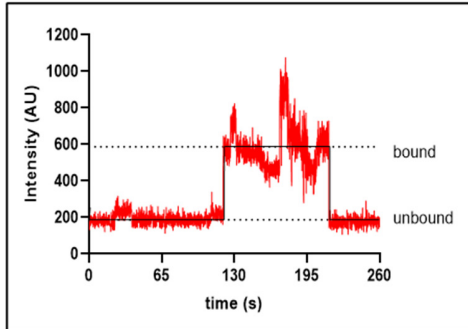
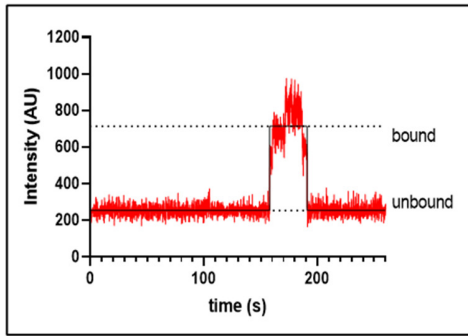
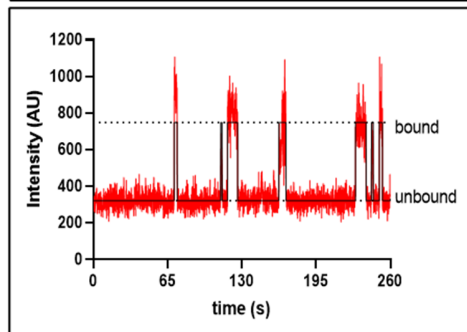
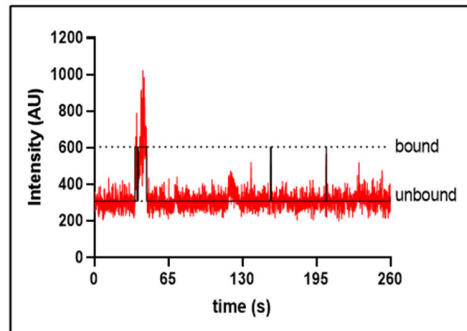
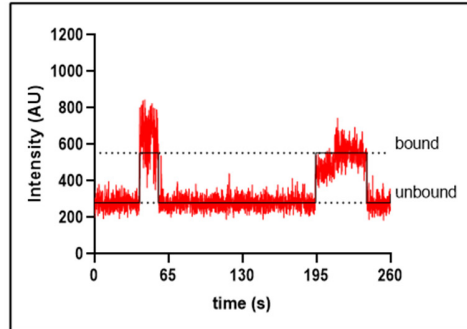
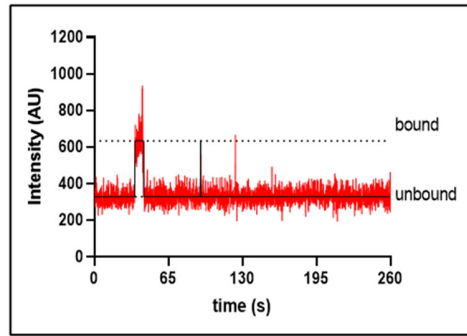
**Supplementary Figure S9. Representative fluorescence trajectories of Cy3-labeled proteins binding to immobilized DNA structures.** Unlabeled DNA was immobilized in the TIRFM reaction chamber, and Cy3-labeled proteins were flown into the chamber. The time-dependent changes in Cy3 fluorescence were recorded, with raw data shown in green and an idealized trajectory shown in black. Panel **A.** and **B.** show the 4 representative fluorescence trajectories for 2hTelG4-G4P, and 2hTelG4-2G4P each, respectively.



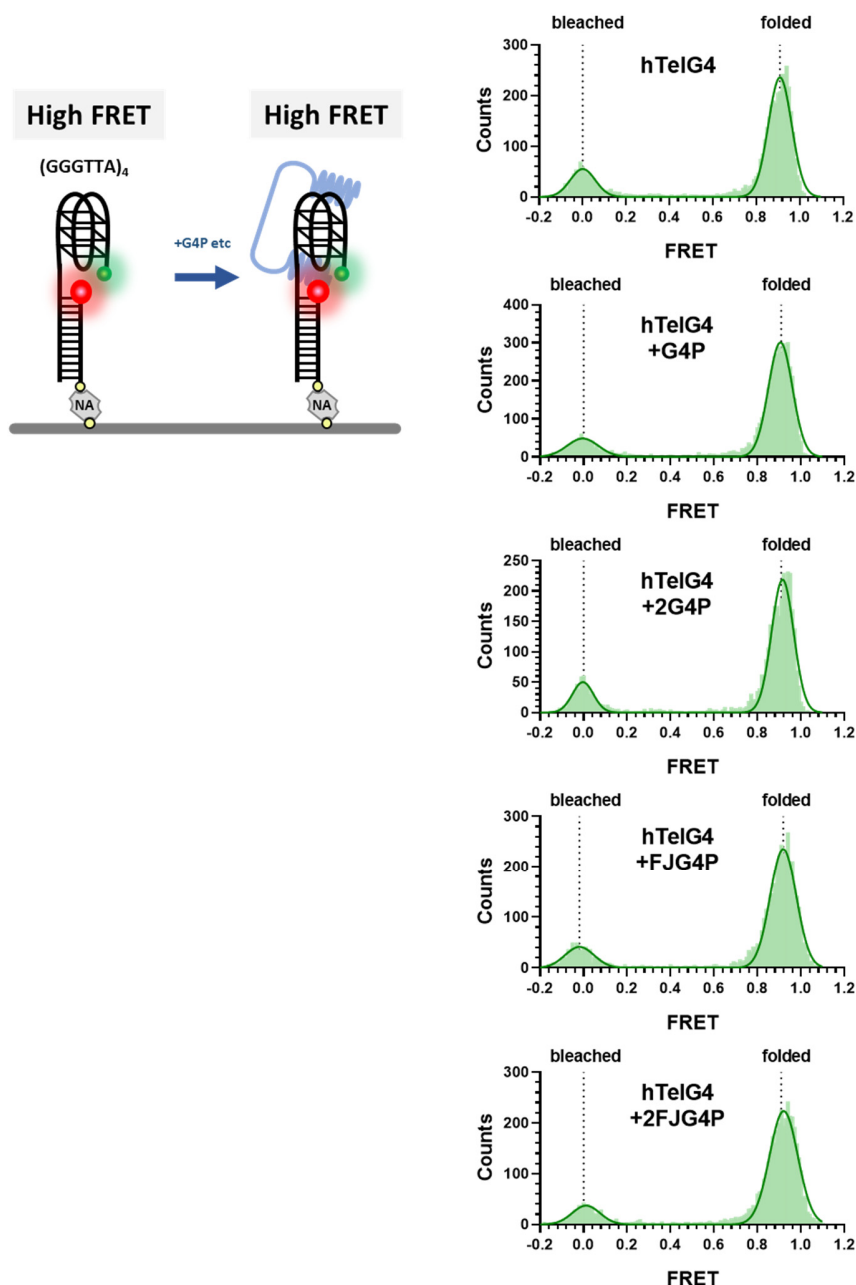
**Supplementary Figure S10. Representative fluorescence trajectories of Cy3-labeled proteins binding to immobilized DNA structures.** Unlabeled DNA was immobilized in the TIRFM reaction chamber, and Cy3-labeled proteins were flown into the chamber. The time-dependent changes in Cy3 fluorescence were recorded, with raw data shown in green and an idealized trajectory shown in black. Panel **A.** and **B.** represent the fluorescence trajectories for 2hTelG4-FJG4P and 2hTelG4-2FJG4P, respectively.



**Supplementary Figure S11. Representative fluorescence trajectories of Cy5-labeled DNA binding to immobilized protein molecules.** Unlabeled protein was immobilized in the TIRFM reaction chamber, and Cy5-labeled DNA molecules were flown into the chamber. The time-dependent changes in Cy5 fluorescence were recorded, with raw data shown in red and an idealized trajectory shown in black. Panel **A.** and **B.** show the 4 representative fluorescence trajectories for *c-MYCG4-2G4P*, and *2hTelG4-2G4P* each, respectively.

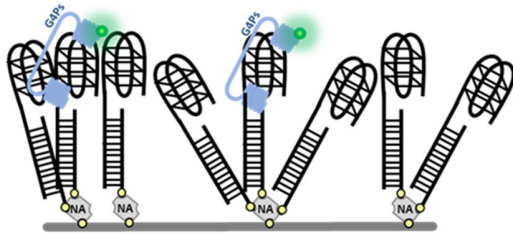
**A. cMycG4-2FJG4P****B. 2hTelG4-2FJG4P**

**Supplementary Figure S12. Representative fluorescence trajectories of Cy5-labeled DNA binding to immobilized protein molecules.** Unlabeled protein was immobilized in the TIRFM reaction chamber, and Cy5-labeled DNA molecules were flown into the chamber. The time-dependent changes in Cy5 fluorescence were recorded, with raw data shown in red and an idealized trajectory shown in black. Panel **A.** and **B.** show the 4 representative fluorescence trajectories for *c*-MYCG4-2G4P, and 2hTelG4-2G4P each, respectively.



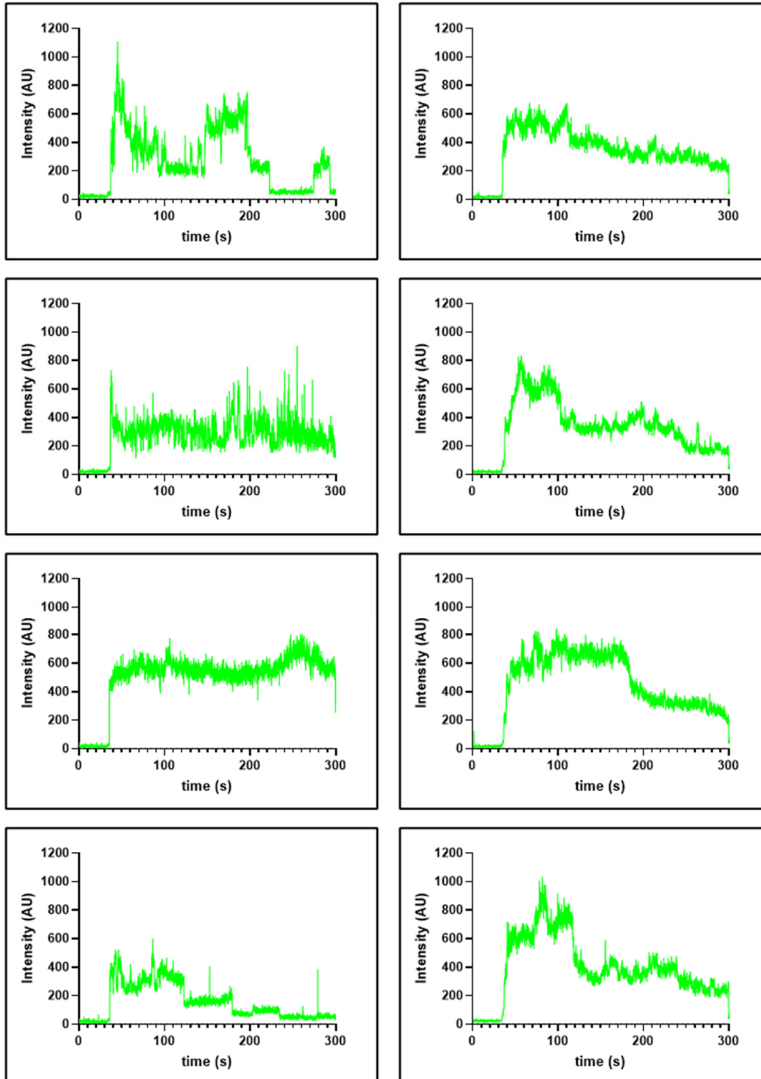
**Supplementary Figure S13. G4P and its variants do not introduce conformational changes to G4 structure.** For all experimental conditions, DNA substrates were immobilized on the surface of the TIRFM flow cell through biotin and neutravidin (NA) conjugation and incubated with saturation amount of 200 nM G4P, 2G4P and 1  $\mu$ M of FJG4P, 2FJG4P. Two-color fluorescence trajectories were extracted from short (100 frames) movies. Calculated FRET values for each molecule in the selected movies were binned with the bin size of 0.01 and plotted and fitted to double Gaussian distributions using GraphPad Prism. The resulting FRET distribution histograms are shown. Dotted lines represent means of FRET peaks of the fully folded G4 (folded) and molecules where Cy5 dye was bleached (bleached). The experimental scheme is depicted on the left.





### A. cMycG4-G4P

### B. 2hTelG4-2G4P



**Supplementary Figure S14. Representative fluorescence trajectories of Cy3-labeled proteins binding to immobilized DNA structures immobilized to the surface at high density (10 nM).** Unlabeled DNA was immobilized in the TIRFM reaction chamber, and Cy3-labeled proteins were flown into the chamber. The time-dependent changes in Cy3 fluorescence were recorded, with raw data shown in green and an idealized trajectory shown in black. Panel A. and B. represent the fluorescence trajectories for *c*-MYCG4-G4P and 2hTel-2G4P, respectively.

### Supplementary References:

1. Zheng, K.W., et al., *Detection of genomic G-quadruplexes in living cells using a small artificial protein*. Nucleic Acids Res, 2020. **48**(20): p. 11706-11720.
2. Zacheja, T., et al., *Mgs1 protein supports genome stability via recognition of G-quadruplex DNA structures*. FASEB J, 2020. **34**(9): p. 12646-12662.