Supporting Information for:

The Hairpin Form of r(G_4C_2)^{exp} in c9ALS/FTD is RAN Translated and a Target for Bioactive Small Molecules

Zi-Fu Wang[1], Andrei Ursu[1], Jessica L. Childs-Disney[1], Rea Guertler[1], Wang Yong-Yang[1], Viachaslau Bernat[1], Suzanne G. Rzuczek[1], Rita Fuerst[1], Yong-Jie Zhang[2], Tania F. Gendron[2], Ilyas Yildirim[3] , Brendan Dwyer[1], Joseph E. Rice[4], Leonard Petrucelli[2], and Matthew D. Disney[1]*

[1]Departments of Chemistry and Neuroscience, The Scripps Research Institute, 130 Scripps Way, Jupiter, FL 33458 USA

[2]Department of Neuroscience, Mayo Clinic, 4500 San Pablo Rd., Jacksonville, FL 32224 USA

[3]Department of Chemistry and Biochemistry, Florida Atlantic University, Jupiter, FL 33458 USA

[4] Department of Medicinal Chemistry, Ernest Mario School of Pharmacy, Rutgers, The State University of New Jersey, 160 Frelinghuysen Road Piscataway, NJ 08854 USA

*Author to whom correspondence is addressed; Email: Disney@scripps.ed

SUPPLEMENTARY TABLES & FIGURES

Table S1. Related to Figure 1 and Table 2. Dissociation constant (K_d) of known G-quadruplex ligands by biolayer interferometry.

Compound	K_d for r$(G_4C_2)_4$ (μM)	K_d for r(UAGGG(UUAGGG)$_3$) (μM)
BMVC	1.3(±0.6)	1.4(±0.6)
BMVC2	0.6(±0.06)	0.14(±0.08)
BMVC4	0.7(±0.07)	0.15(±0.08)
Braco19	0.7(±0.1)	0.21(±0.08)
PDS	0.07(±0.02)	0.09(±0.06)
cPDS	0.2(±0.07)	0.13(±0.07)
TMPyP4	0.05(±0.02)	0.11(±0.07)

Table S2. Related to Table 2. Binding parameters of **4** and r(G_4C_2)$_x$ (x = 2,4,6) repeats of various length in K^+-containing buffer determined by biolayer interferometry.[a]						
RNA	$K_{d,1}$ **(M)**	$k_{on,1}$ **(1/M×s)**	$k_{off,1}$ **(1/s)**	$K_{d,2}$ **(M)**	$k_{on,2}$ **(1/M×s)**	$k_{off,2}$ **(1/s)**
r(G_4C_2)$_2$	$2.4(\pm0.3)\times10^{-7}$	$1.8(\pm0.2)\times10^{5}$	$4.3(\pm0.1)\times10^{-2}$	$3.2(\pm0.02)\times10^{-6}$	$2.7(\pm0.1)\times10^{4}$	$8.6(\pm0.3)\times10^{-2}$
r(G_4C_2)$_4$	$1.9(\pm0.9)\times10^{-7}$	$1.4(\pm0.7)\times10^{5}$	$2.6(\pm0.05)\times10^{-2}$	$2.0(\pm0.09)\times10^{-6}$	$2.5(\pm0.1)\times10^{4}$	$5.1(\pm0.1)\times10^{-2}$
r(G_4C_2)$_6$	$3.0(\pm0.3)\times10^{-7}$	$1.0(\pm0.1)\times10^{5}$	$3.0(\pm0.1)\times10^{-2}$	$2.6(\pm0.6)\times10^{-6}$	$2.2(\pm0.5)\times10^{4}$	$5.6(\pm0.4)\times10^{-2}$

[a] k_{on} and k_{off} values were generated by fitting the obtained data using ForteBio's Data Analysis 7.1 software to global 2:1 (heterogeneous) binding equation using entire time range.

Table S3. (**STAR Methods Sections**, " *Effect of compound treatment on* $(G_4C_2)_{66}$*–No ATG–GFP transcript levels* " and "*Polysome profiling*") Sequences of qPCR primers used in these studies. All primers were purchased from Integrated DNA Technologies (IDT) except for 18S rRNA primers, which were obtained from Eurofins.

RNA	Primer sequence
GFP (fwd)	5'-GCACGACTTCTTCAAGTCCGCCATGCC-3'
GFP (rev)	5'-GCGGATCTTGAAGTTCACCTTGATGCC-3'
18S (fwd)	5'-GTAACCCGTTGAACCCCATT-3'
18S (rev)	5'-CCATCCAATCGGTAGTAGCG-3'
β-actin (fwd)	5'-GATTACTGCTCTGGCTCCTAGCA-3'
β-actin (rev)	5'-GCTCAGGAGGAGCAATGATCTT-3'

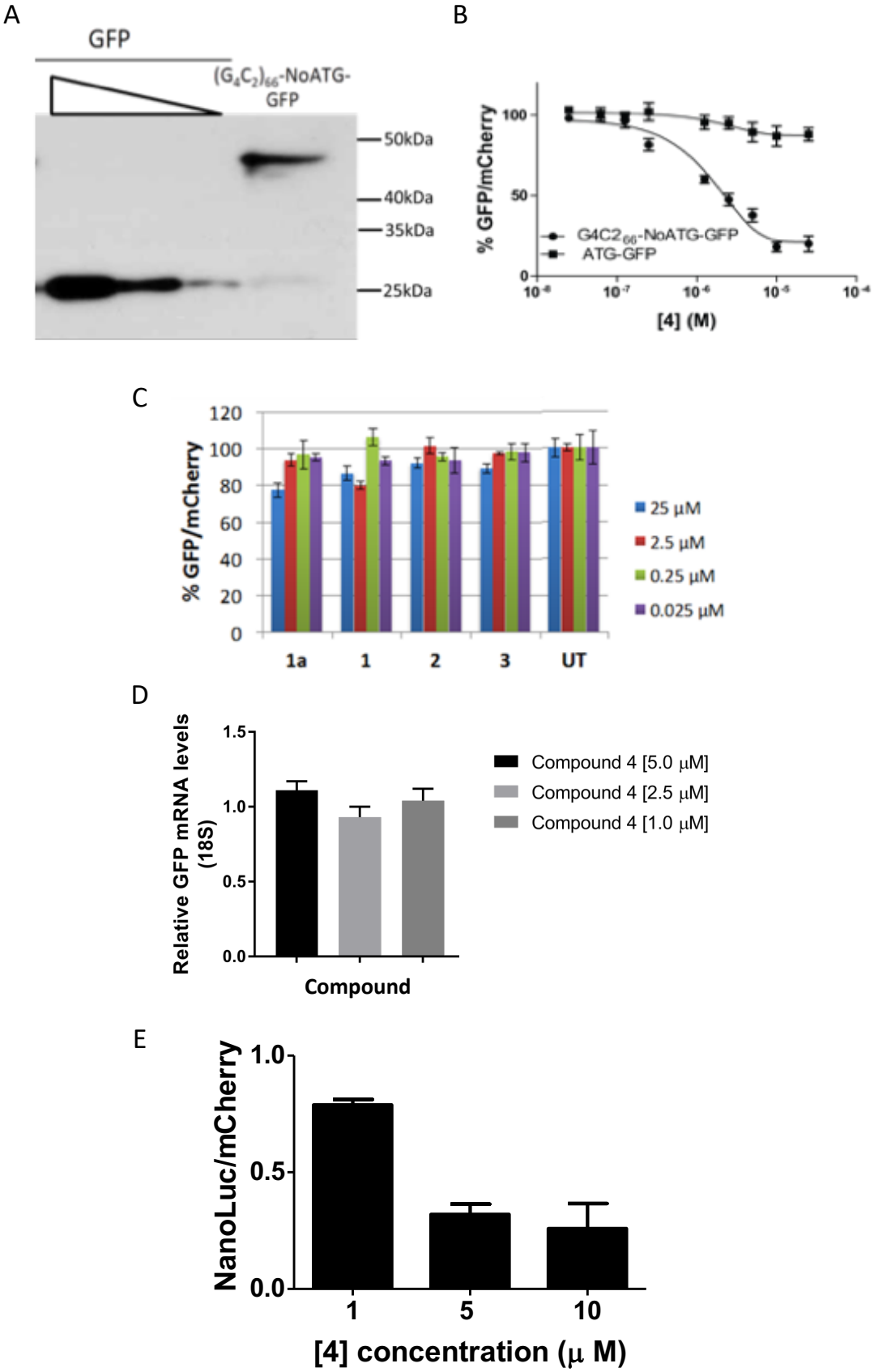

A
GFP
(G4C2)66-NoATG-GFP
50kDa
40kDa
35kDa
25kDa
B
% GFP/mCherry
100
50
0
G4C266-NoATG-GFP
ATG-GFP
10-8
10-7
10-6
10-5
10-4
[4] (M)
C
% GFP/mCherry
120
100
80
60
40
20
0
1a
1
2
3
UT
25 µM
2.5 µM
0.25 µM
0.025 µM
D
Relative GFP mRNA levels (18S)
1.5
1.0
0.5
0.0
Compound
Compound 4 [5.0 µM]
Compound 4 [2.5 µM]
Compound 4 [1.0 µM]
E
NanoLuc/mCherry
1.0
0.5
0.0
1
5
10
[4] concentration (µ M)

Figure S1. **Related to Figure 1**. **A cell-based assay to monitor RAN translation inhibition upon addition of 4 that is amendable to high-throughput screening**. (**A**) Western blotting of the proteins generated from *$(G_4C_2)_{66}$-No ATG-GFP* using an anti-GFP antibody. (**B**). Dose-response curve of **4** in *$(G_4C_2)_{66}$-No ATG-GFP* or *ATG-GFP* transfected cells to demonstrate that **4** selectively inhibits RAN translation but not traditional translation as evidenced by the decrease in GFP/mCherry signal for *$(G_4C_2)_{66}$-No ATG-GFP* (**circle**) but no change in *ATG-GFP* signal (**square**). (**C**) Effect of **1a** and the four most promising **1a**-like small molecules on canonical GFP translation using a plasmid that encodes GFP with a canonical ATG start codon. "UT" indicated untreated. The effect of **4** on canonical GFP translation is shown in panel B. (**D**) Effect of **4** on *$(G_4C_2)_{66}$-No ATG-GFP* mRNA transcript levels as determined by RT-qPCR. (**E**) Dose dependent effect of **4** in *$(G_4C_2)_{66}$-NOATG-NanoLuc* transfected HEK293T cell.

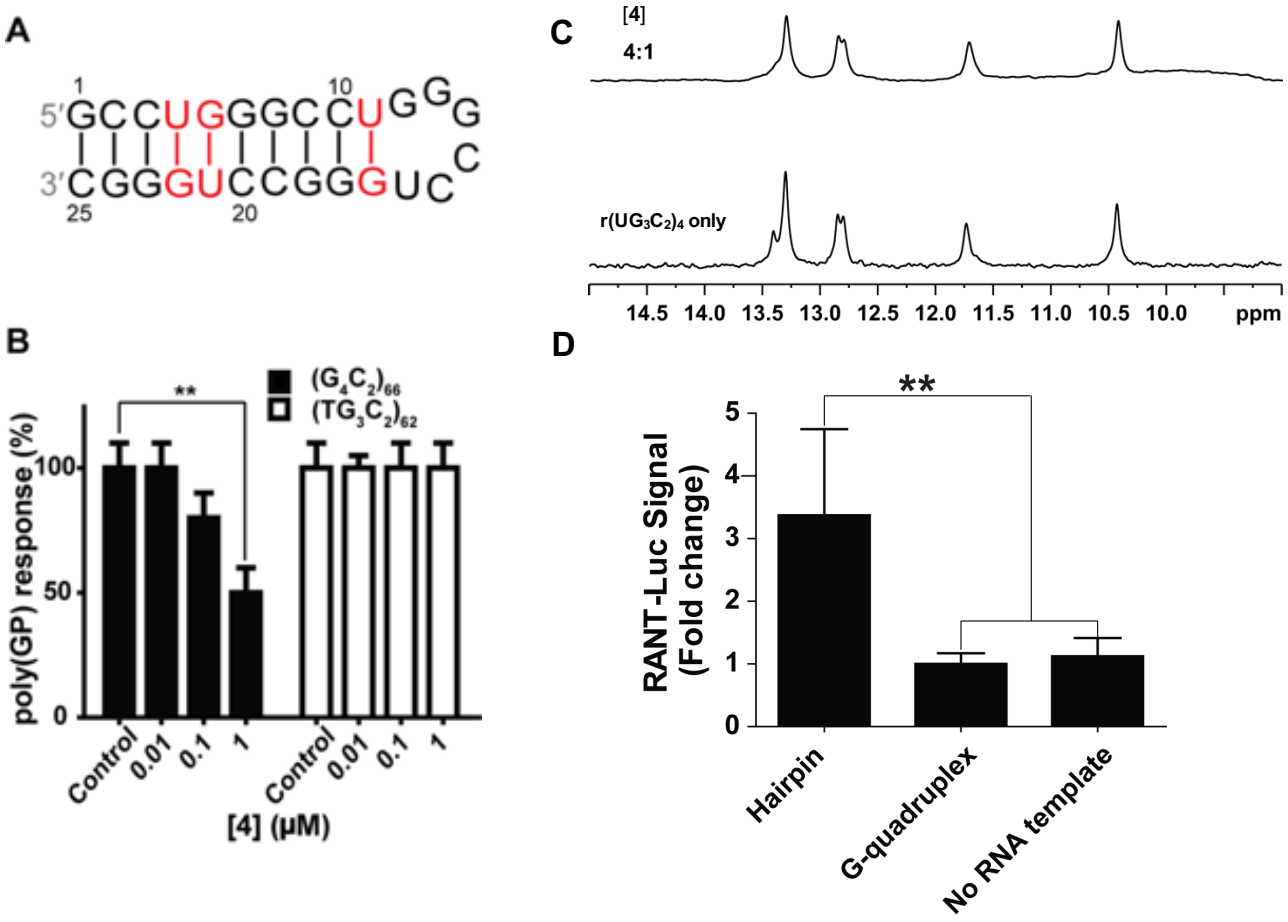
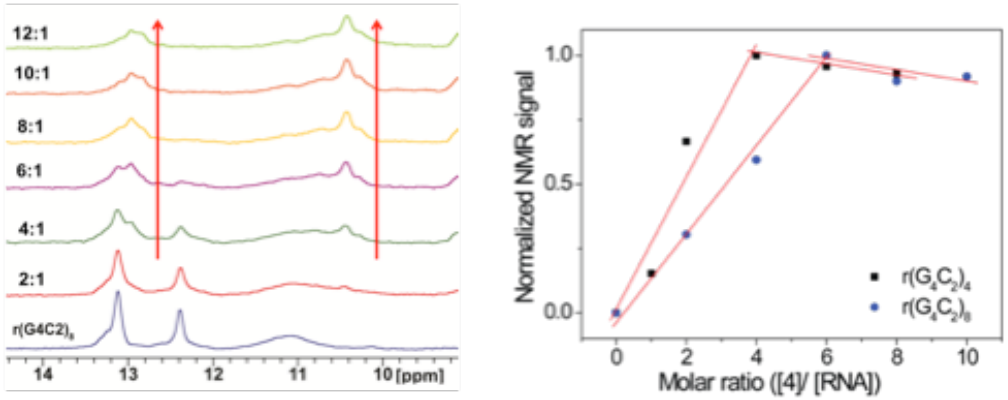
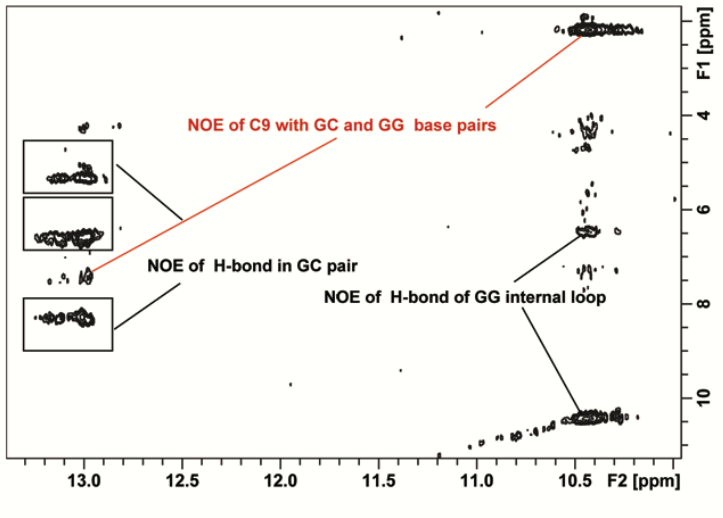


Figure S2. Related to Figures 1 and 2. Poly(GP) activity assay to explore single base modification of $r(G_4C_2)^{exp}$ versus $r(UG_3C_2)^{exp}$. (A) Secondary structure of $r(UGGGCC)_4$. **(B)** Poly(GP) activity assay in $(G_4C_2)_{66}$ or $(TG_3C_2)_{62}$-plasmid transfected HEK293T cell. ** indicates $p < 0.01$ as determined by a two-tailed Student *t*-test. **(C)** Imino ^1H NMR spectra of $r(UG_3C_2)_4$ (**bottom**) and upon addition of **4** equivalents of **4** (**top**) in a buffer containing 10 mM sodium phosphate. **(D) Related to Figure 3.** An in vitro translation assay shows that the hairpin structure, not the G-quadruplex, undergoes RAN translation. The RNA template of a *r(G4C2)66-luciferase* construct was pre-folded under the conditions that dictate hairpin (no K^+ present) or G-quadruplex structure (K^+ present). ** indicates $p < 0.01$ as determined by a two-tailed Student *t*-test.

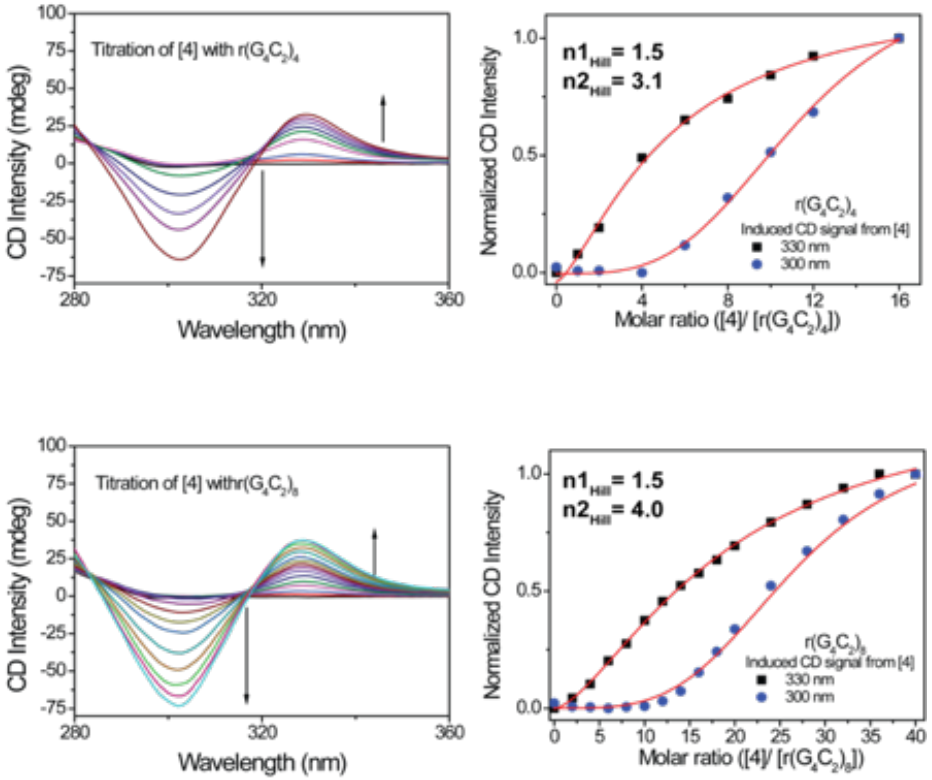
A


12:1
10:1
8:1
6:1
4:1
2:1
r(G4C2)4
14
13
12
11
10 [ppm]
Normalized NMR signal
1.0
0.5
0.0
0
2
4
6
8
10
Molar ratio ([4]/ [RNA])
r(G4C2)4
r(G4C2)8

B


NOE of C9 with GC and GG base pairs
NOE of H-bond in GC pair
NOE of H-bond of GG internal loop
F1 [ppm]
F2 [ppm]
13.0
12.5
12.0
11.5
11.0
10.5

C


Titration of [4] with r(G4C2)4
CD Intensity (mdeg)
Wavelength (nm)
n1Hill= 1.5
n2Hill= 3.1
Normalized CD Intensity
r(G4C2)4
Induced CD signal from [4]
330 nm
300 nm
Molar ratio ([4]/ [r(G4C2)4])
Titration of [4] withr(G4C2)8
n1Hill= 1.5
n2Hill= 4.0
r(G4C2)8
Molar ratio ([4]/ [r(G4C2)8])

D

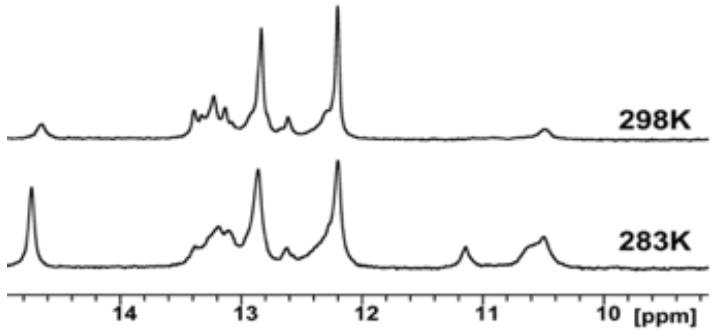

298K
283K
14
13
12
11
10 [ppm]

Figure S3. Related to Figure 2. (**A**) **Left,** Imino ^{1}H NMR spectra of r$(G_4C_2)_8$ upon increasing equivalents of **4** in a buffer containing 10 mM sodium phosphate (pH 7.0). The ratio is the concentrations of **4** : r$(G_4C_2)_8$; **Right**, stoichiometry of **4** and r$(G_4C_2)_4$ or r$(G_4C_2)_8$ as determined by NMR spectroscopy. (**B**) Expanded NOESY spectrum of r$(G_4C_2)_4$ in complex with four equivalents of **4** recorded with a mixing time of 400 ms in 10 mM sodium phosphate buffer (pH 7.0). (**C**) The cooperative binding of 4 with r$(G_4C_2)_4$ and r$(G_4C_2)_8$ as determined by CD spectroscopy**.** **Top left**, CD spectra upon addition of **4** to r$(G_4C_2)_4$. **Top right**, binding curve of **4** and r$(G_4C_2)_4$ generated by plotting the CD signal at 330 nm and 300 nm as a function of the molar ratio of [**4**]/ [r$(G_4C_2)_4$]. **Bottom left,** CD spectra upon addition of **4** to r$(G_4C_2)_8$. **Bottom right,** binding curve of **4** and r$(G_4C_2)_8$ generated by plotting the CD signal at 330 nm and 300 nm as a function of the molar ratio of [**4**]/ [r$(G_4C_2)_8$]. Cooperative binding curves were fitted to equation 2. (**D**) Imino ^{1}H NMR spectra of r$(GUCGCCGGGGCCGGGCGAC)_2$ at 298 K and 283 K in a buffer containing 10 mM sodium phosphate, pH 7.0 and 100 mM LiCl.

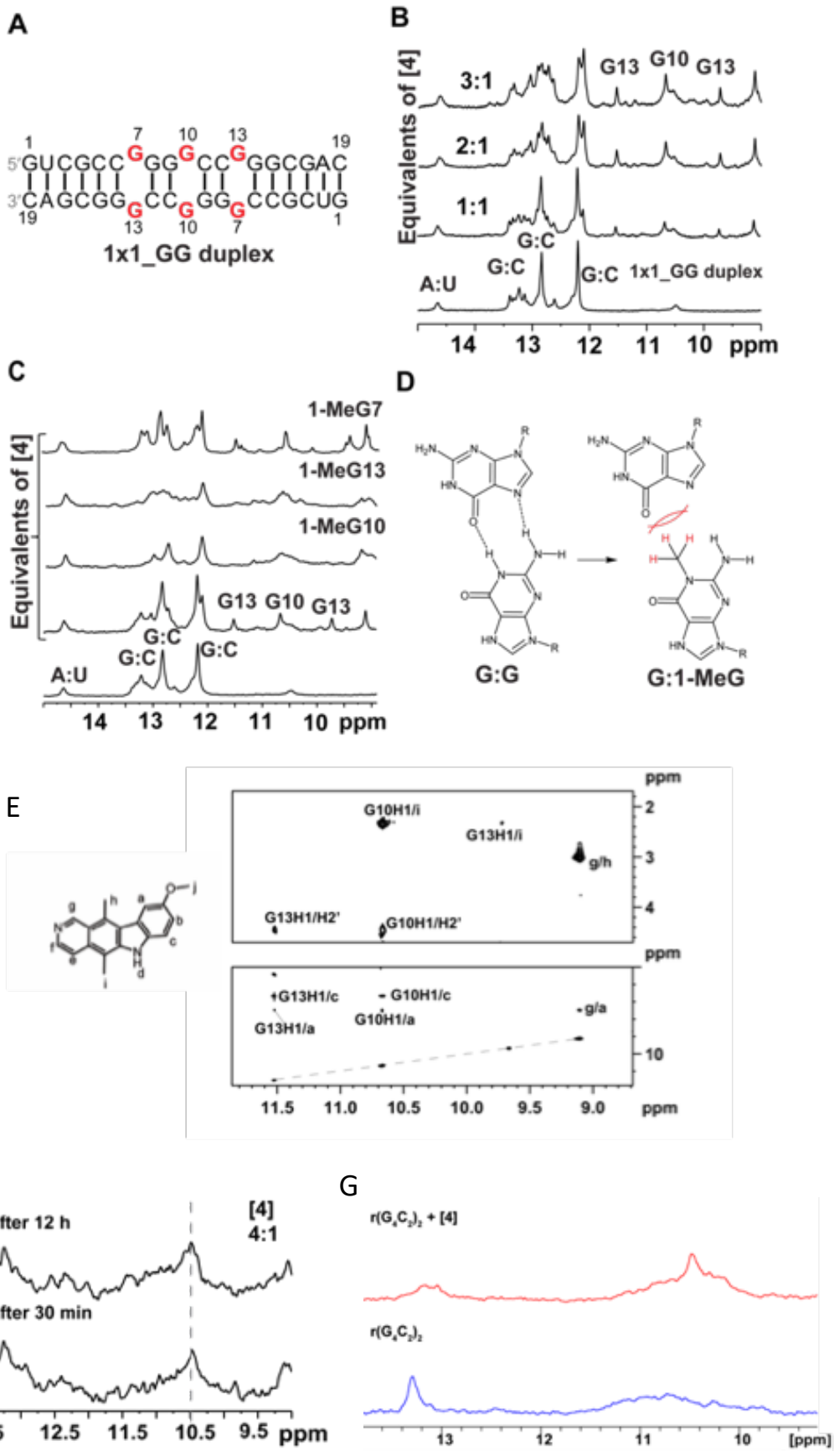

A
1x1_GG duplex
B
Equivalents of [4]
3:1
2:1
1:1
G13 G10 G13
G:C
A:U
1x1_GG duplex
14 13 12 11 10 ppm
C
Equivalents of [4]
1-MeG7
1-MeG13
1-MeG10
G13 G10 G13
G:C
A:U
14 13 12 11 10 ppm
D
G:G
G:1-MeG
E
G10H1/i
G13H1/i
g/h
G13H1/H2'
G10H1/H2'
+G13H1/c
G10H1/c
G13H1/a
G10H1/a
g/a
11.5 11.0 10.5 10.0 9.5 9.0 ppm
F
+K+ after 12 h
+K+ after 30 min
[4]
4:1
13.5 12.5 11.5 10.5 9.5 ppm
G
r(G4C2)2 + [4]
r(G4C2)2
13 12 11 10 [ppm]

Figure S4. Related to Figure 2. NMR analysis of 4 binding to GG internal loops. (A) Site-specific methylation of duplex model. To further explore our hypothesis that **4** binds the GG internal loop in the hairpin structure, the following duplex stem [r(GUCGCC**GGGGCC**GGGCGAC)$_2$] was designed to contain two r(G_4C_2) repeats and adopt three 1x1 GG internal loops, which provide three distinct GG internal loops for unambiguous characterization (red). **(B)** Imino ^{1}H NMR spectra of r(GUCGCC**GGGGCC**GGGCGAC)$_2$ upon adding equivalents of **4** to the 1×1 GG duplex. The imino proton NMR spectrum of the 1×1 GG internal loop duplex exhibited three peaks in the 10 to 12 ppm region only after the addition of **4**, which were assigned to G7:G13 and G10:G10. **(C)** Overlaid imino proton NMR spectra of the 1×1 GG duplex from bottom to top: (1) duplex without **4**, (2) duplex containing **4**, (3-5) duplexes containing **4** with 1-methylguanine modifications at G10 (3), at G13 (4) and at G7 (5). Samples containing **4** all have three equivalents of **4.** These data suggest that **4** stabilized hydrogen bonding between the GG pairs. When G7 was methylated (1-MeG7 in, all three imino proton peaks from G10 and G13 were present, indicating that methylation did not affect hydrogen bonding between the GG mismatched pairs. However, when G10 and G13 were methylated (1-MeG10 and 1-MeG13), all three imino proton peaks were absent. **(D)** Chemical structure of GG pair hydrogen bonding disrupted by 1-methylguanine modification. Since methylation prevents the hydrogen bond donor from donating, G13 is a hydrogen bond donor, G7 is an acceptor, and G10 is both a donor and an acceptor. (**E**) Expanded NOESY spectrum of 1x1 GG duplex, r(GUCGCCGGGGCCGGGCGAC)$_2$ with two equivalents of **4** recorded with a mixing time of 400 ms in 100 mM LiCl and 10 mM sodium phosphate buffer (pH 7.0). The NOESY spectrum of **4** with the 1×1 GG duplex showed several NOE signals of **4** with the imino protons of G10 and G13 in the GG internal loop, providing additional evidence that **4** interacts with the 1×1 GG internal loop. (**F**) Compound **4** stabilizes the hairpin structure and inhibits G-quadruplex formation in r(G_4C_2)$_4$. Imino proton NMR spectra of r(G_4C_2)$_4$ with four equivalents of **4** at 30 min (bottom) and 12 h (top) after addition of K^+. The dashed line at 10.5 ppm is the imino protons from the GG internal loop pairs. To compare the binding of the G-quadruplex ligands and **4**, we tested whether adding K^+ after **4** binds to r(G_4C_2)^{exp} prevents conversion to G-quadruplex form. We added four equivalents of **4** before the addition of K^+. After 12 h of incubation with K^+ incubation, the imino proton signal due to GG pairs 10.5 ppm) did not shift or change. (**G**) Moreover, we characterized the interaction of **4** with r(G_4C_2)$_2$ by NMR spectroscopy, which cannot form

an intramolecular G-quadruplex as it only has two G_4 repeats. Imino proton NMR spectra of $r(G_4C_2)_2$ with four equivalents of **4** (top) or without **4** (bottom). A new peak appears at ~10.5 ppm upon addition of **4**, which is consistent with **4** binding and stabilizing the hydrogen bonds formed between GG pairs. Indeed, **4** induced the same imino proton signal (10.5 ppm) as observed in the presence of $r(G_4C_2)_4$, suggesting that **4** still recognizes the GG mismatch pairs in short repeats and not the G-quadruplex structure. Taken together, these findings indicate that **4** stabilizes the hydrogen bonding of the GG internal loops displayed in the stem of the hairpin structure and inhibits conversion into the G-quadruplex.

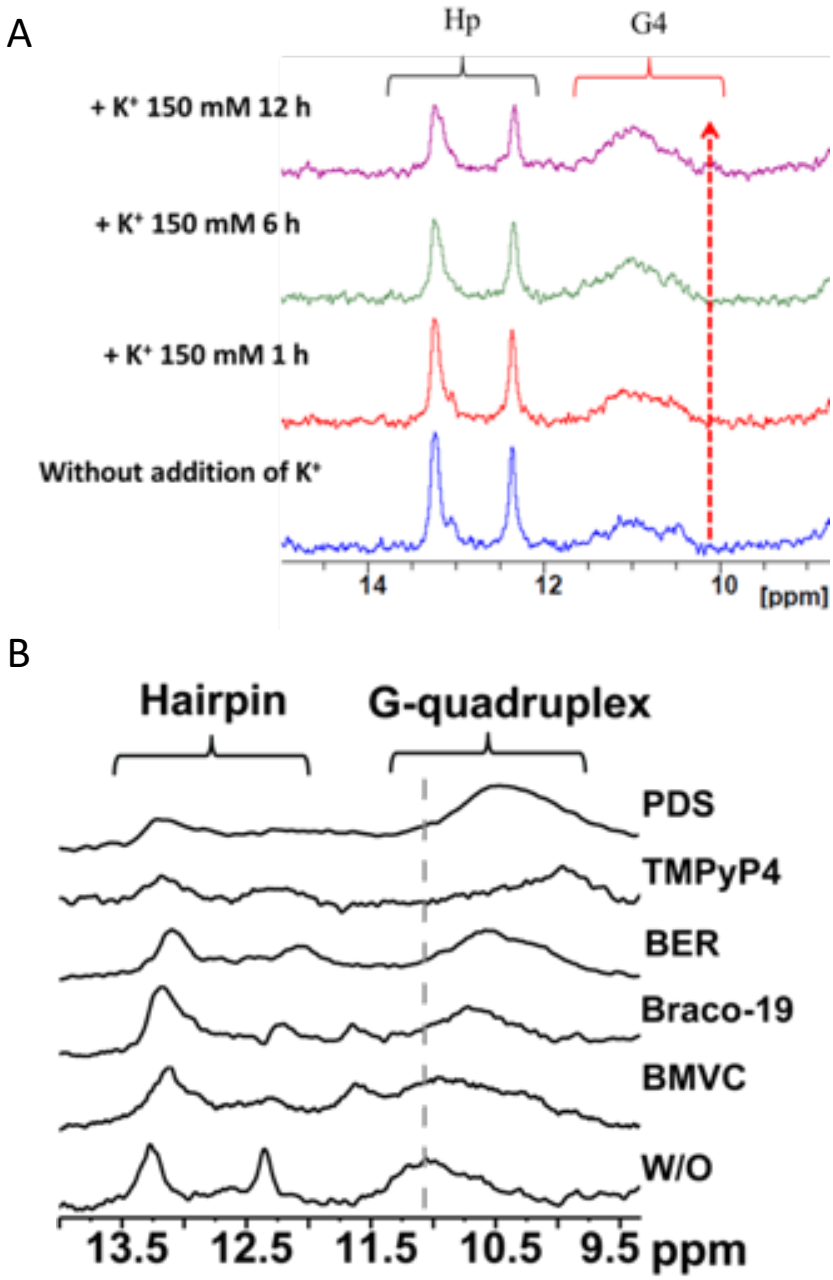


Figure S5. Related to Figure 2. Imino proton NMR spectra of $r(G_4C_2)_4$ in the absence and presence of 150 mM K^+ and $r(G_4C_2)_4$ with various G-quadruplex ligands. (**A**) Having established that **4** binds the GG internal loop in the hairpin structure, we hypothesized that targeting the G-quadruplex form results in poor inhibition of RAN translation. Therefore, we proceeded to characterize the binding of the G-quadruplex ligands to confirm that they do in fact bind to the G-quadruplex structure of $r(G_4C_2)^{exp}$ using imino proton NMR, which can distinguish between the hairpin and G-quadruplex structures formed by $r(G_4C_2)^{exp}$.(Bugaut et al., 2012; Kuo et al., 2015; Su et al., 2014) The imino

proton signals of the GC base pairs in $r(G_4C_2)_4$ without K^+ present were consistent with previous studies that report the hairpin form exists as the native predominant structure of G-rich DNA and RNA sequences as K^+ is essential for the stabilization of G-quadruplexes.(Bugaut et al., 2012; Kuo et al., 2015) Approximately 12 h post-addition of 150 mM KCl, $r(G_4C_2)_4$ reaches an equilibrium between the hairpin and G-quadruplex. "Hp" indicates peaks derived from the hairpin form of $r(G_4C_2)_4$ (GG internal loops) while "G4" indicates peaks derived from the G-quadruplex form. (**B**) G-quadruplex-binding ligands alters the equilibrium between G-quadruplex and hairpin structures of $r(G_4C_2)_4$ to varying extents. Imino proton NMR spectra of $r(G_4C_2)_4$ after addition of K^+ (12 h) without (W/O; bottom) and with two equivalents of **BMVC**, **Braco-19**, **BER**, **TMPyP4**, or **PDS**. Upon addition of the G-quadruplex ligands **BMVC**, **Braco-19**, **BER**, **TMPyP4** and **PDS**, the imino proton signal in the G-quadruplex region shifted upfield, which indicates binding to the G-quadruplex structure. Interestingly, **PDS** shifts the equilibrium towards the G-quadruplex structure, as evidenced by the decrease in the signal of the hairpin imino proton. Furthermore, **PDS** induced the largest upfield shift of the imino proton signal of G-quadruplex as expected given PDS had the lowest K_d in the BLI studies (Supporting information Table S1). Therefore, these well-known G-quadruplex ligands do in fact bind to the G-quadruplex structure of $r(G_4C_2)^{exp}$ as expected.

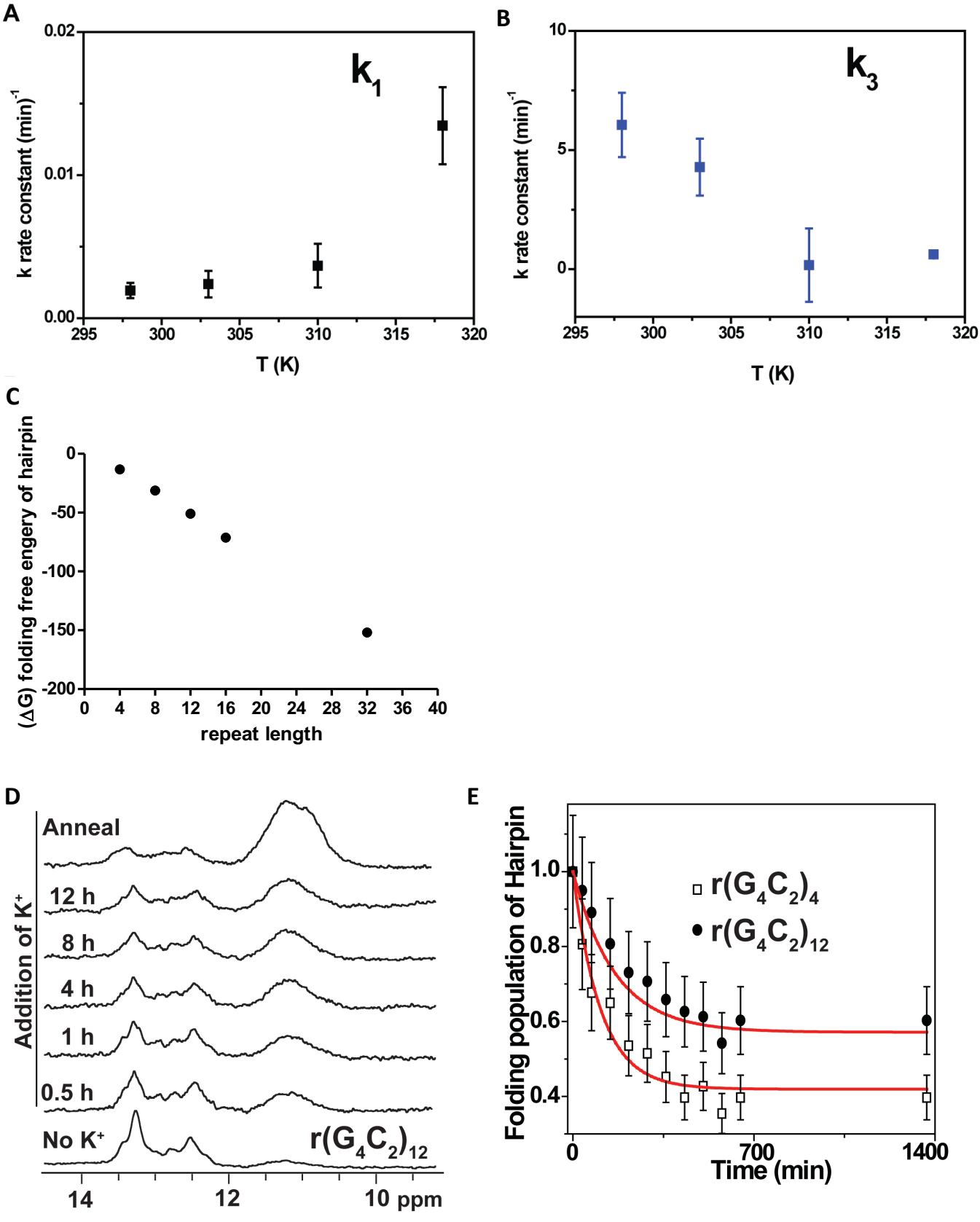

A
k rate constant (min)-1
k1
T (K)
B
k rate constant (min)-1
k3
T (K)
C
(ΔG) folding free engery of hairpin
repeat length
D
Addition of K+
Anneal
12 h
8 h
4 h
1 h
0.5 h
No K+
r(G4C2)12
14
12
10 ppm
E
Folding population of Hairpin
r(G4C2)4
r(G4C2)12
Time (min)

Figure S6. Related to Figure 3. Studying the equilibrium between hairpin and G-quadruplex forms of r(G_4C_2) repeats. The plot of rate constants (**A**) k_1(unfolding of hairpin structure), and (**B**) k_3 (folding of G-quadruplex structure) versus temperature. (**C**) The folding free energy (ΔG_f) of hairpin structure in r(G_4C_2) repeats, as predicted by *m*-fold (Zuker, 2003). (**D**) Time-resolved imino proton spectra of r$(G_4C_2)_{12}$ recorded after the addition of 150 mM K^+ at 0.5, 1, 4, 8, 12 h, and after annealing at 37 ^oC. **(E)** Kinetic trace analysis of r$(G_4C_2)_4$ (**open squares**) or r$(G_4C_2)_{12}$ (**solid circles**) as monitored by the imino proton NMR signal of G-quadruplex at 10 – 12 ppm after addition of K^+ at 37 °C. Red lines show the curve fit to a three-state kinetic model. The unfolding times of the hairpin structure in r$(G_4C_2)_4$ and r$(G_4C_2)_{12}$ are 272 to 407 min, respectively. The fraction folded for r$(G_4C_2)_4$ and r$(G_4C_2)_{12}$ is 0.4 to 0.6, respectively.

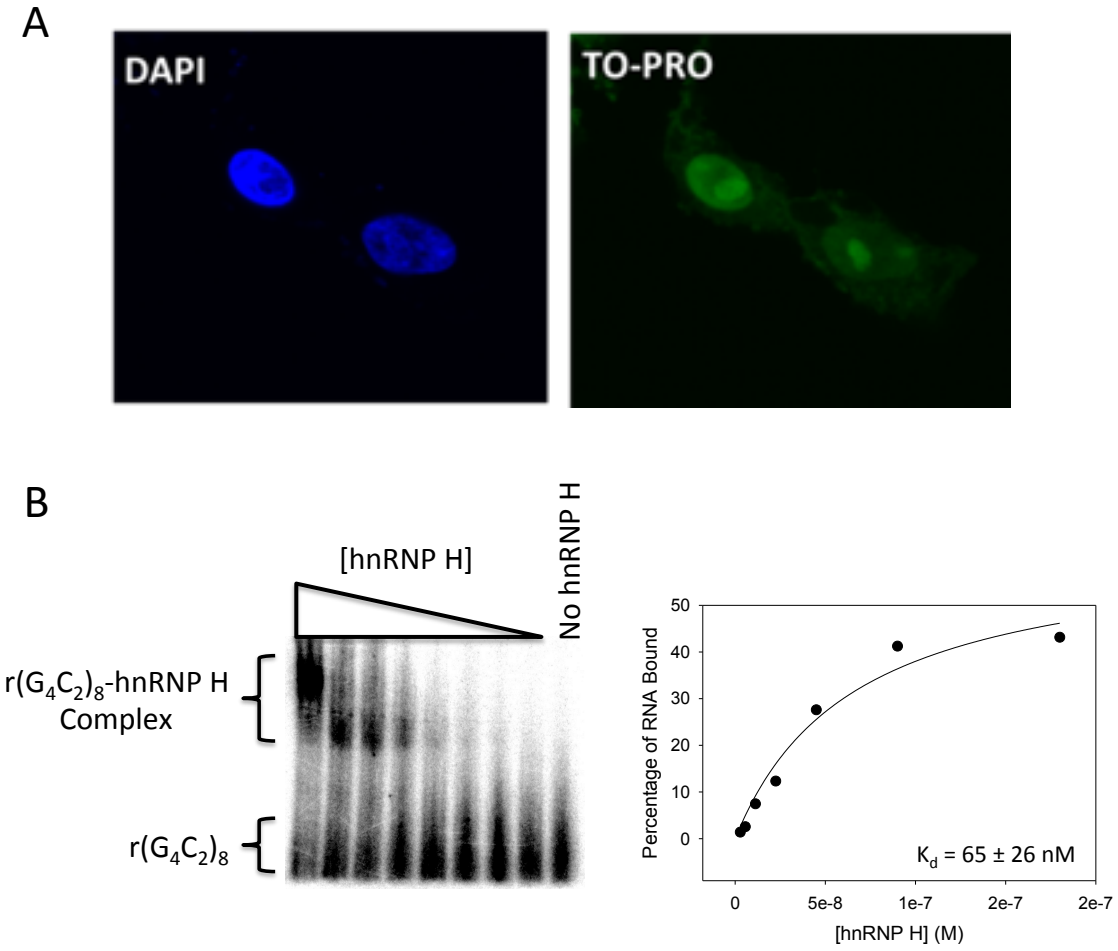


Figure S7. Related to Figures 3 & 5. (A) Fluorescence microscopy imaging of TO_Q (TO-PRO derivative). DAPI staining of the nucleus was detected in blue channel, and **TO_Q** was detected in the green channel. Representative images of **TO_Q** in *$(G_4C_2)_{66}$*-transfected cell. The cells were fixed and then post-stained by **TO_Q** 0.5 μM for 24 h. The structure of **TO_Q** is shown in the Synthetic Methods & Characterization section. (**B**) Affinity of hnRNP H for $r(G_4C_2)_8$ as determined by a gel shift assay.