Supplementary Information

Specific suppression of D-RNA G-quadruplex-protein interaction with an L-RNA aptamer

Mubarak I. Umar[1], Chun Kit Kwok[1,2*]

[1]Department of Chemistry, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Hong Kong SAR, China

[2] Shenzhen Research Institute of City University of Hong Kong, Shenzhen, China

* To whom correspondence should be addressed. Email: ckkwok42@cityu.edu.hk

SI Tables and Figures

References

Table S1. Oligonucleotides used for the *in vitro* selection.

Oligos	Sequences (5' - 3')	Type
N_{40} Library	TTCTAATACGACTCACTATAGGTTACCAGCCTTCACTGC **(N_{40})** GCACCACGGTCGGTCACAC	D-DNA
Forward Primer	TTCTAATACGACTCACTATAGGTTACCAGCCTTCACTGC	
Reverse Primer	GTGTGACCGACCGTGGTGC	
5'Biotin L-*hTERC* rG4	Biotin – GGGUUGCGGAGGGUGGGCCU	L-RNA

Note: The 39 and 19 nucleotides before and after the N40 regions are the fixed forward and reverse primers respectively. The nucleotides highlighted in red are the T7 promoter regions.

Table S2. Conditions used for the *in vitro* selection process.

Mining rounds	**1**	**2**	**3**	**4**	**5**	**6**	**7**
RNA library conc. (uM)	3.3	2	1	0.1	0.1	0.05	0.02
Target conc. (uM)	0.65	0.65	0.65	0.33	0.1	0.05	0.02
Negative selection	2 hrs	2 hrs	2 hrs	1 hr	1 hr	1 hr	1 hr
Positive selection (Incubation)	30 mins	30 mins	30 mins	30 mins	30 mins	30 mins	30 mins
Positive selection (Washing)	1min, pipette mix	1min, pipette mix	1min, pipette mix	10 mins, 300 rpm	10 mins, 300 rpm	10 mins, 300 rpm	10 mins, 300 rpm
Temperature (^{o}C)	25	25	37	37	37	37	37
MgCl2 conc. (mM)	5	5	5	1	1	1	1
PCR cycles	12	8	6	6	6	8	10 Dream Taq

Note: The selection stringency increased as the *in vitro* selection progresses. For example, each selection round had lowered D-RNA library and L-*hTERC* rG4 concentrations. The $MgCl_2$ concentration also decreased from 5 mM in the early rounds to 1 mM from round four. Washing time increased from 1 to 10 minutes from round four. During the entire selection rounds, the concentrations of 150 mM KCl and 25 mM Tris-HCl (pH 7.5) were maintained during the selection rounds.

Table S3. Representative sequences obtained from Sanger sequencing.

Colony	Sequence (5' – 3')
1	GGTTACCAGCCTTCACTGC **GTGCCTGCTCTCCTAGGGTGGTGGTGGATGGAGGGCAGCG** GTGTGACCGACCGTGGTGC
2	GGTTACCAGCCTTCACTGC **GCGCTGATGCGTCCGATGGTGGTGGTGGATGGGCGCAGC** GTGTGACCGACCGTGGTGC
3	GGTTACCAGCCTTCACTGC **CTGGGGTTGCGGATGCGGTGGTGGTGGATTCCGCAGCCCA** GTGTGACCGACCGTGGTGC
4	GGTTACCAGCCTTCACTGC **TCGCTGCGCCCTATTGGTGGTGGTGGGAGGGCCAGCGCGG** GTGTGACCGACCGTGGTGC
5	GGTTACCAGCCTTCACTGC **GGCGTCTCTCTATGGTGGTGGTGGGAGGGAGGCGCCGATA** GTGTGACCGACCGTGGTGC

Note: Bold sequences are the N40 region, the 19 and 19 nucleotides before and after the N40 regions are the fixed region of RNA aptamer after T7 *in vitro* RNA transcription respectively.

Table S4. Sequences of aptamers, targets and peptide.

Name	Sequences (5' - 3')	Nucleotides /Amino acid residues
D-Apt.4	GGUUACCAGCCUUCACUGCUCGCUGCGCCCUAUUGGU GGUGGUGGGAGGGCCAGCGCGGGCACCACGGUCGGUC ACAC	78
D-Apt.4-1	GCCCUAUUGGUGGUGGUGGGAGGGC	25
D-Apt.4-1a	GCCCUAUUGAUGAUGAUGAGAGGGC	25
D-Apt.4-1b	GCCCUAUUGGCGGCGGCGGGAGGGC	25
D-Apt.4-1c	GCCCUAAAGGUGGUGGUGGGAGGGC	25
D-Apt.4-1d	GCCCUAGGUGGUGGUGGGAGGGC	23
D-Apt.4-1c(G9->A9)	GCCCUAAAAGUGGUGGUGGGAGGGC	25
D-Apt.4-1c(G10->A10)	GCCCUAAAGAUGGUGGUGGGAGGGC	25
D-Apt.4-1c(G12->A12)	GCCCUAAAGGUAGUGGUGGGAGGGC	25
D-Apt.4-1c(G13->A13)	GCCCUAAAGGUGAUGGUGGGAGGGC	25
D-Apt.4-1c(G15->A15)	GCCCUAAAGGUGGUAGUGGGAGGGC	25
D-Apt.4-1c(G16->A16)	GCCCUAAAGGUGGUGAUGGGAGGGC	25
D-Apt.4-1c(G18->A18)	GCCCUAAAGGUGGUGGUAGGAGGGC	25
D-Apt.4-1c(G20->A20)	GCCCUAAAGGUGGUGGUGGAAGGGC	25
D-Apt.4-1c(G18-20->A18-20)	GCCCUAAAGGUGGUGGUAAAAGGGC	25
D-Apt.4-1c(U11->C11)	GCCCUAAAGGCGGUGGUGGGAGGGC	25
D-Apt.4-1c(U14->C14)	GCCCUAAAGGUGGCGGUGGGAGGGC	25
D-Apt.4-1c(U17->C17)	GCCCUAAAGGUGGUGGCGGGAGGGC	25
D/L *hTERC* rG4	*FAM*-GGGUUGCGGAGGGUGGGCCU	20
D-*hTERC* rG4 mutant1	*FAM*-GGGAAGCGGAGGGUGGGCCU	20
D-*hTERC* rG4 mutant2	*FAM*-GGGUUGGGAGGGUGGGCCU	19
D-*hTERC* rG4 mutant3	*FAM*-GGGUUGCGGUGGGUGGGCCU	20
D-*hTERC* rG4 mutant4	*FAM*-GGGUUGCGGAGGGAGGGCCU	20
5'ext. D-*hTERC* rG4	*FAM*-AAAGGGUUGCGGAGGGUGGGCCU	23

3'ext. D-*hTERC* rG4	*FAM*-GGGUUGCGGAGGGUGGGCCUAAA	23
RNA hairpin	*FAM*-CAGUACAGAUCUGUACUG	18
DNA hairpin	*FAM*-CAGTACAGATCTGTACTG	18
Poly rA	*FAM*-AAAAAAAAAAAAAAAAAA	18
Poly rU	*FAM*-UUUUUUUUUUUUUUUUUU	18
Poly rC	*FAM*-CCCCCCCCCCCCCCCCCC	18
D-*TERRA* rG4	*FAM*-UUAGGGUUAGGGUUAGGGUUAGGG	24
D-*NRAS* rG4	*FAM*-GGGAGGGGCGGGUCUGGG	18
D-*MT3* rG4	*FAM*-GAGGGAGGGAGGGAGAGGGA	20
D-*Bcl2* rG4	*FAM*-GGGGGCCGUGGGGUGGGAGCUGGGG	25
D-*TRF2* rG4	*FAM*-CGGGAGGGCGGGGAGGGC	18
D-*ZIC1* rG4	*FAM*-GGGUGGGGGGGGCGGGGGAGGCCGGGG	27
D-*Adam10* rG4	*FAM*-GGGGGACGGGUAGGGGCGGGAGGUAGGGG	29
D-*hTERC* dG4	*FAM*-GGGTTGCGGAGGGTGGGCCT	20
D-*hTELO* dG4	*FAM*-TTAGGGTTAGGGTTAGGGTTAGGG	24
D-*VEGF* dG4	*FAM*-CGGGGCGGGCCGGGGGCGGGGTC	23
D-*Bcl2Mid* dG4	*FAM*-GGGCGCGGGAGGAAGGGGGCGGG	23
D-*c-Kit* dG4	*FAM*-AGGGAGGGCGCTGGGAGGAGGG	22
D-*hTERC* dG4 for CD	GGGTTGCGGAGGGTGGGCCT	20
D-*hTELO* dG4 for CD	TTAGGGTTAGGGTTAGGGTTAGGG	24
D-*VEGF* dG4 for CD	CGGGGCGGGCCGGGGGCGGGGTC	23
Bcl2Mid dG4 for CD	GGGCGCGGGAGGAAGGGGGCGGG	23
D-*c-Kit* dG4 for CD	AGGGAGGGCGCTGGGAGGAGGG	22
Nucleolin GAR domain	GFGGRGGGRGGFGGRGGGRGGRGGFGGRGRGGFGGRG GFRGGRGGGGDHKPQGKKTKFE	59

Note: The red nucleotide(s) indicates mutated nucleotide(s).

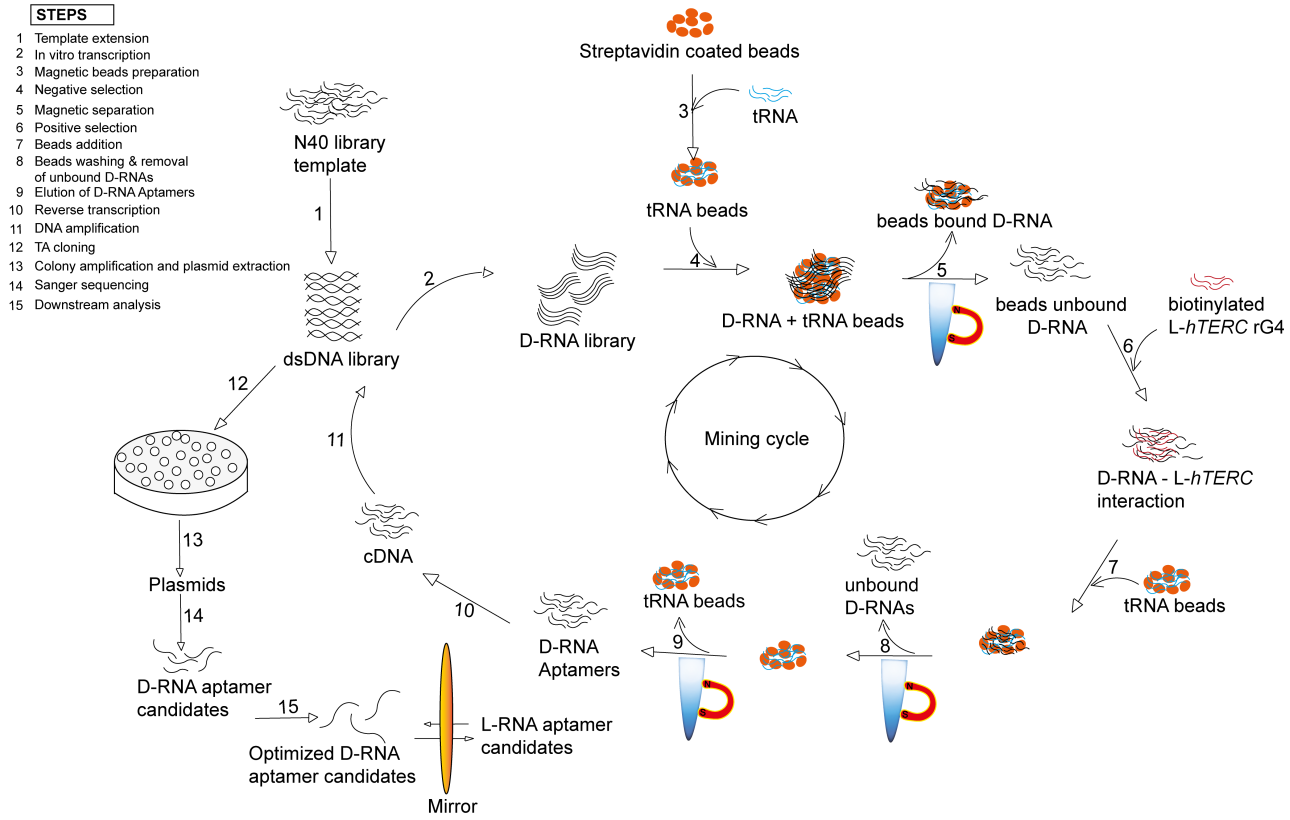

Figure S1. Schematic representation of the *in vitro* selection method used in this study. See detail descriptions for each step in the method section.

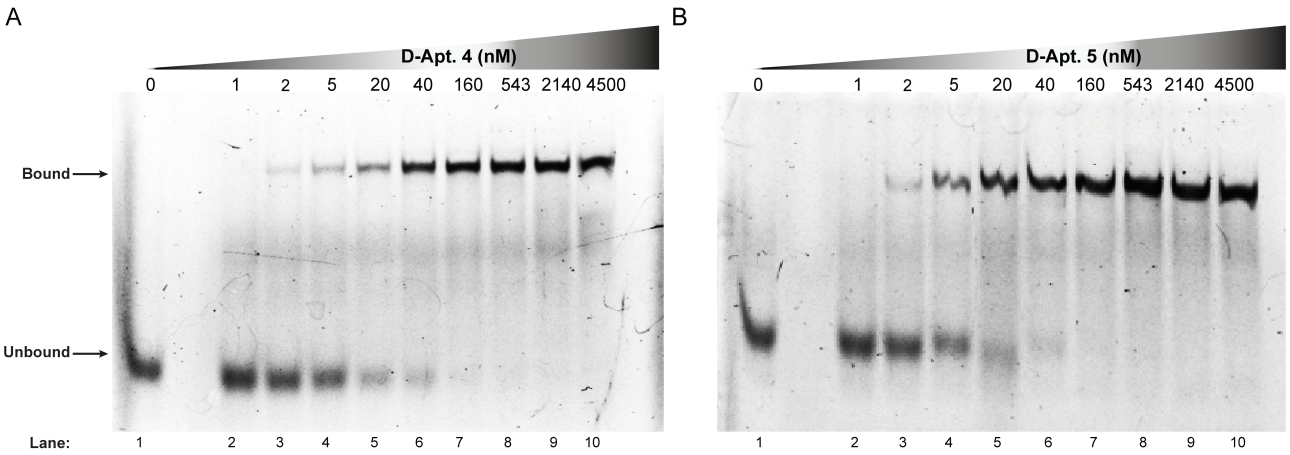

Figure S2. Binding of D-Apt. 4 and D-Apt. 5 with L-*hTERC* rG4. A) The binding between D-Apt.4 and FAM-L-*hTERC* rG4 was monitored by EMSA. 12% native gel was used to resolve the bound and unbound fraction (see methods). Each lane contained 1 nM FAM-L-*hTERC*, and D-Apt.4 was added from 0-4500 nM. With increasing concentration of D-Apt.4 (from left to right on the gel), the bound band (upper band) became darker, whereas the unbound band (lower band) became weaker. This suggests direct interaction between D-Apt.4 and L-*hTERC* rG4. **B)** Same as A), except that D-Apt.5 was used.

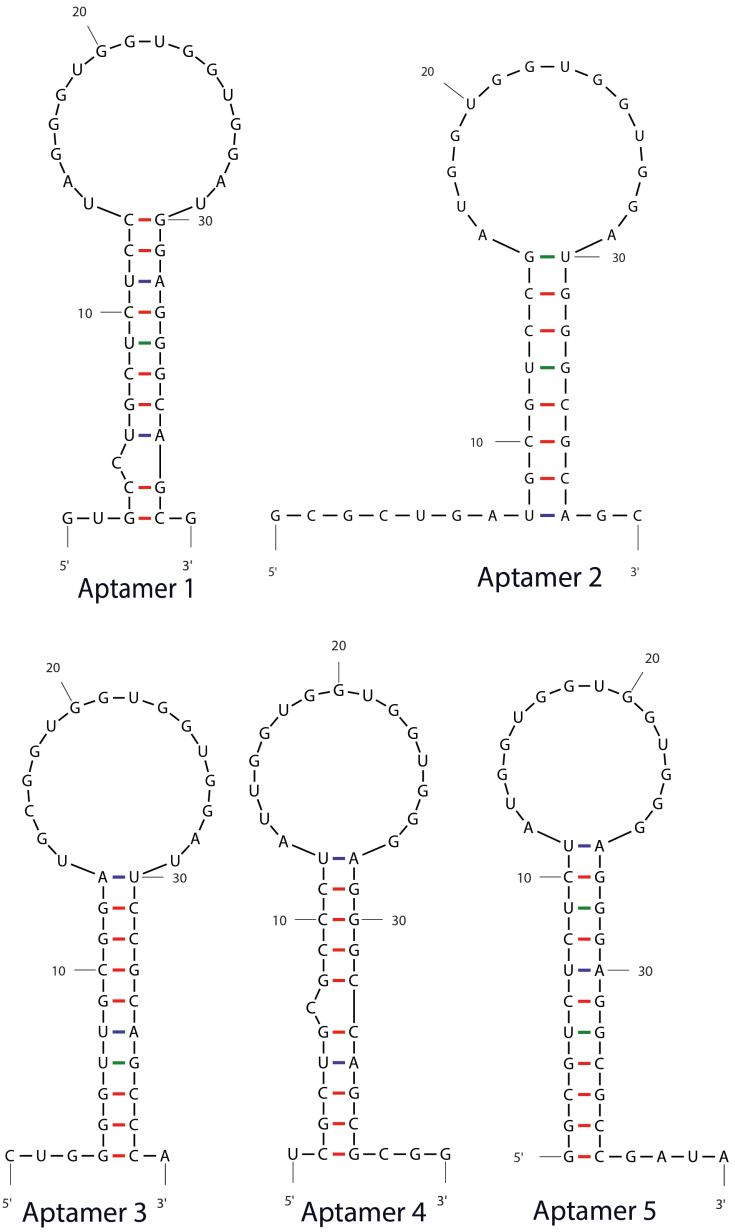

Figure S3. RNA secondary structure prediction of aptamer candidates. Mfold [1] was used to predict the secondary structure of sequences from representative clones (Table S3, and see methods). The fixed primer region was removed and only the N40 region was shown for each case. Comparative analysis of the sequences/structures were performed in Fig. 1D. Note that Aptamer 2 only has 39 nucleotides only.

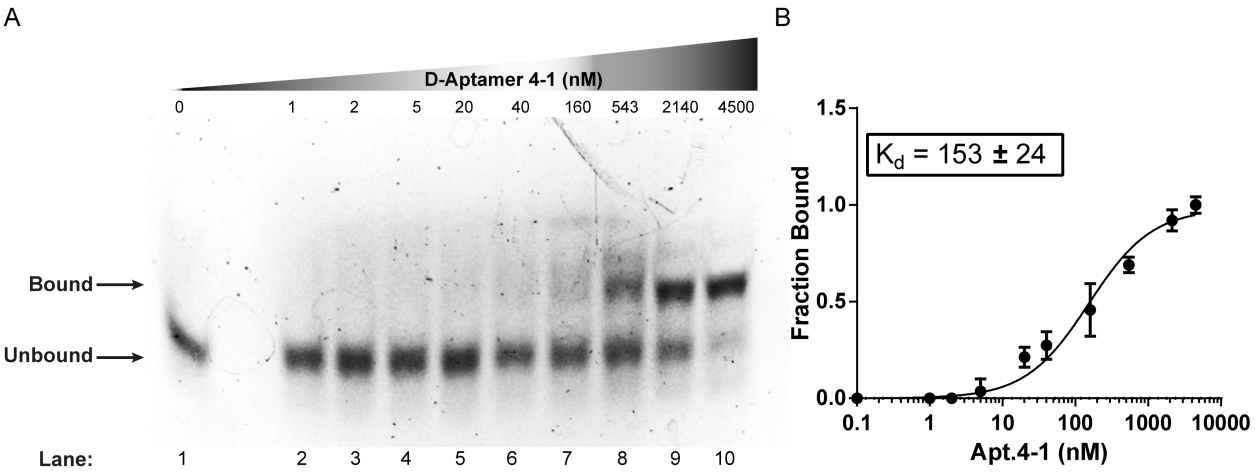

Figure S4. Binding of D-Apt. 4-1 with L-*hTERC* rG4. A) 12% EMSA gel used to analyze the binding of D-Apt. 4-1 to FAM-L-*hTERC* rG4. 1 nM FAM-L-*hTERC* added to all lanes, and the bound bands (upper band) increases with an increase in D-Apt.4-1 concentration from 0 – 4500 nM, while the lower bands (unbound bands) weakens. **B)** Saturation plot for the binding of D-Apt. 4-1 with FAM-L-*hTERC* rG4. The *Kd* was determined to be 153 ± 24 nM from 3 independent replicates. The error bars represent the standard deviation.

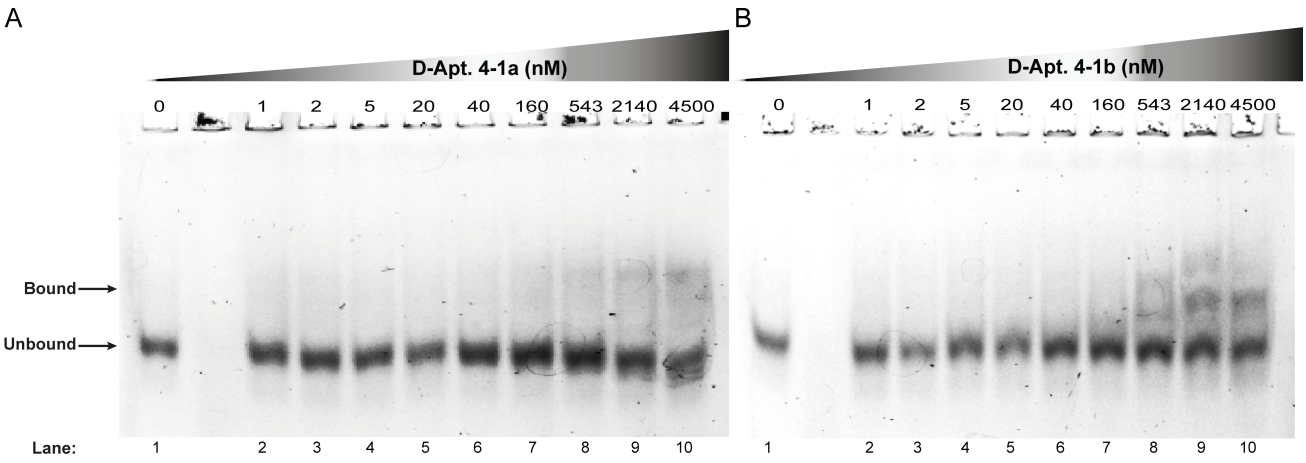

Figure S5. Binding of D-Apt. 4-1a & D-Apt. 4-1b with L-*hTERC* rG4. A) The binding of D-Apt. 4-1a to FAM-L-*hTERC* rG4 was analyzed by 12% EMSA gel. All lanes contained 1 nM FAM-L-*hTERC*, and increasing concentrations of D-Apt. 4-1a from 0 – 4500 nM. **B)** Same as A) above except the D-Apt. 4-1b was used instead. Both A & B showed no/very weak upper band (bound bands), suggested D-Apt. 4-1a and D-Apt. 4-1b do not interact/interact very weakly with L-*hTERC* rG4. The *Kd* was unable to be determined due to the no/very weak binding.

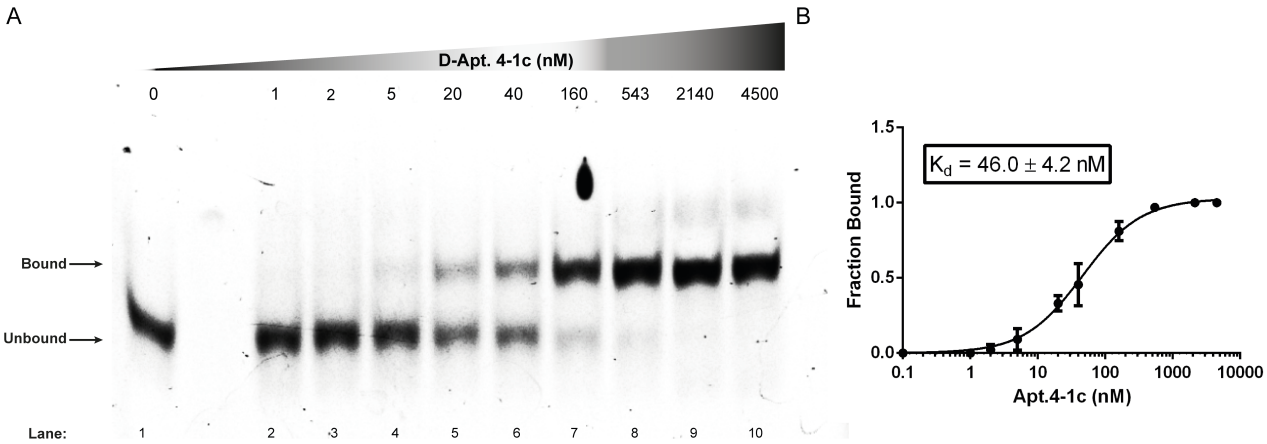

Figure S6. Binding of D-Apt. 4-1c with L-*hTERC* rG4. A) The binding of D-Apt. 4-1c to FAM-L-*hTERC* rG4 was examined by EMSA. 12% native gel was used to observe the bound band (upper bands) and unbound band (lower bands). All lanes contained 1 nM FAM-L-*hTERC* rG4 while D-Apt. 4-1c concentration varied from 0 – 4500 nM. The intensity of the bound band (upper bands) increases when the concentration of Apt. 4-1c increases, while the unbound band (lower bands) diminishes. **B)** Saturation plot for binding of D-Apt. 4-1c with FAM -L-*hTERC* rG4. The *Kd* value was found to be 46.0 ± 4.2 nM.

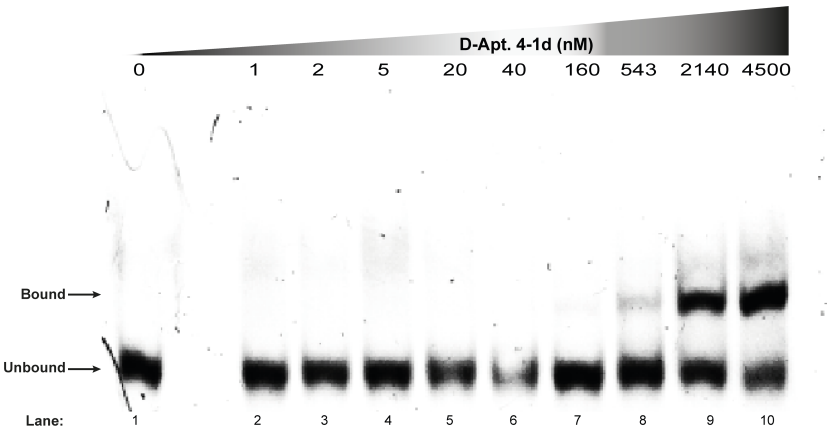

Figure S7. Binding of D-Apt. 4-1d with L-*hTERC* rG4. The binding between D-Apt. 4-1d and FAM-L-*hTERC* was monitored by 12% EMSA gel. All lanes contained 1 nM FAM-L-*hTERC* rG4, whereby D-Apt. 4-1d concentration increases from 0 – 4500 nM. The upper bands (bound band) was found to increase slightly at 543 – 4500 nM aptamer concentrations respectively, while the lower bands (unbound bands) slightly weakens. This result indicates D-Apt. 4-1d binds to L-*hTERC* rG4 weakly.

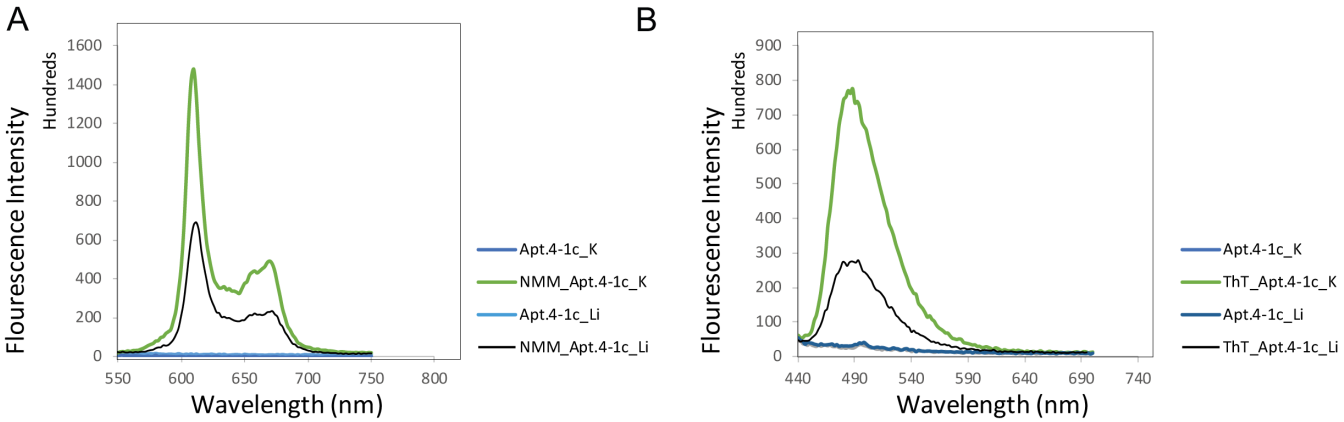

Figure S8. G4 ligand-enhanced fluorescence assays on D-Apt. 4-1c. A) NMM ligand-enhanced fluorescence spectra of D-Apt. 4-1c. 1:1 concentration of D-Apt. 4-1c and NMM was used. The result showed an enhanced fluorescence intensity (~2.5-fold at 610 nm) under 150 mM KCl as compared to 150 mM LiCl, indicating the formation of G-quadruplex. **B)** Same as A) except another G4 ligand ThT was used and the fold intensity was found to be ~3-fold at 488 nm.

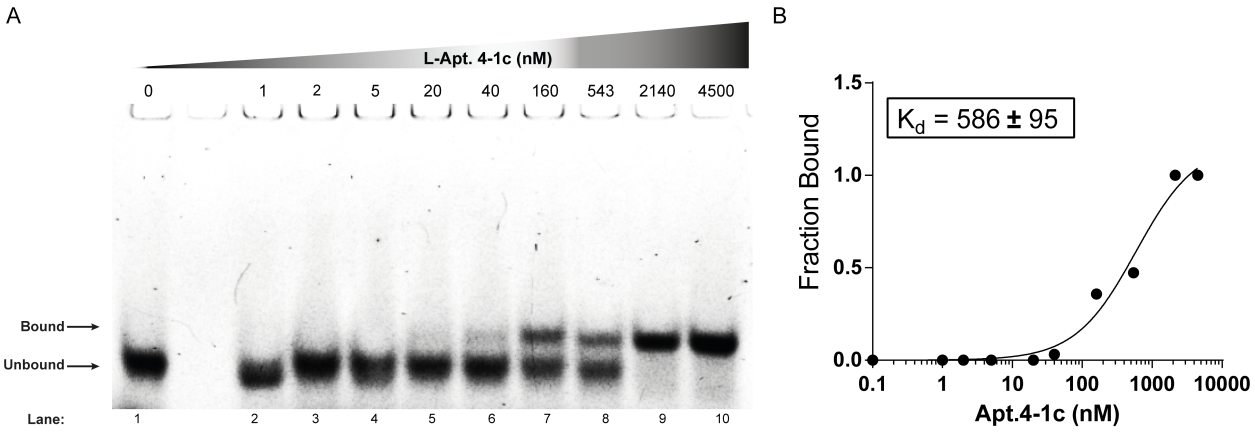

Figure S9. Binding of L-Apt. 4-1c with D-*hTERC* rG4 in the absence of Mg^{2+}. A) The binding of L-Apt. 4-1c to FAM-L-*hTERC* rG4 was tested without Mg^{2+}. 12% native gel was used to identify the upper bands (bound band) and lower bands (unbound band). All lanes contained 1 nM FAM-L-*hTERC* rG4 while L-Apt. 4-1c concentration varied from 0 – 4500 nM. As the concentration of Apt. 4-1c increases the intensity of the bound band (upper bands) increases, while the unbound band (lower bands) decreases. These results showed a decrease affinity in absence of Mg^{2+}, an indication that Mg^{2+} enhances the binding affinity for Apt. 4-1c (Fig. 3A). **B)** Saturation plot for binding of L-Apt. 4-1c with FAM -L-*hTERC* rG4 in absence of Mg^{2+}. The *Kd* value was found to be 586 ± 95 nM.

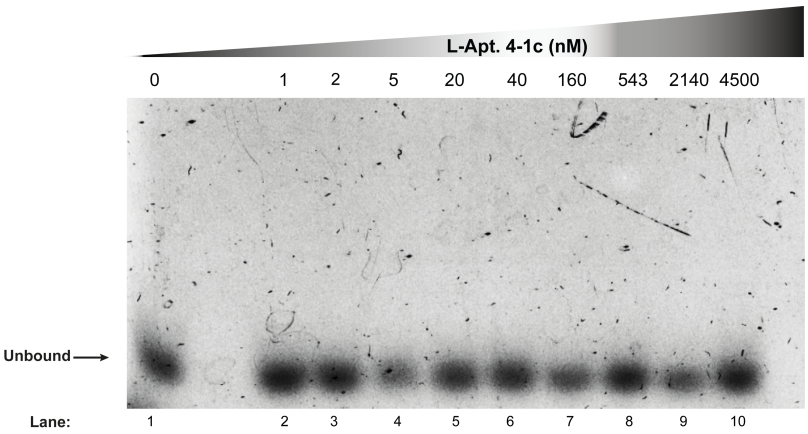

Figure S10. Binding of L-Apt. 4-1c with D-*hTERC* rG4 in the presence of Li^+ instead of K^+. The binding of L-Apt.4-1c to FAM-D-*hTERC* rG4 was tested in the presence of Li^+ instead of K^+. EMSA gel was used to resolved bound (upper bands) and unbound bands (lower bands). All lanes contained 1 nM FAM-D-*hTERC* rG4 and increasing L-Apt.4-1c concentrations (0-4500 nM). EMSA showed no binding (upper bands) as the concentration of L-Apt.4-1c increases (0-4500 nM). This result indicated the presence of Li^+ instead of K^+ abolished the binding, likely by unfolding the rG4 in both the aptamer and target.

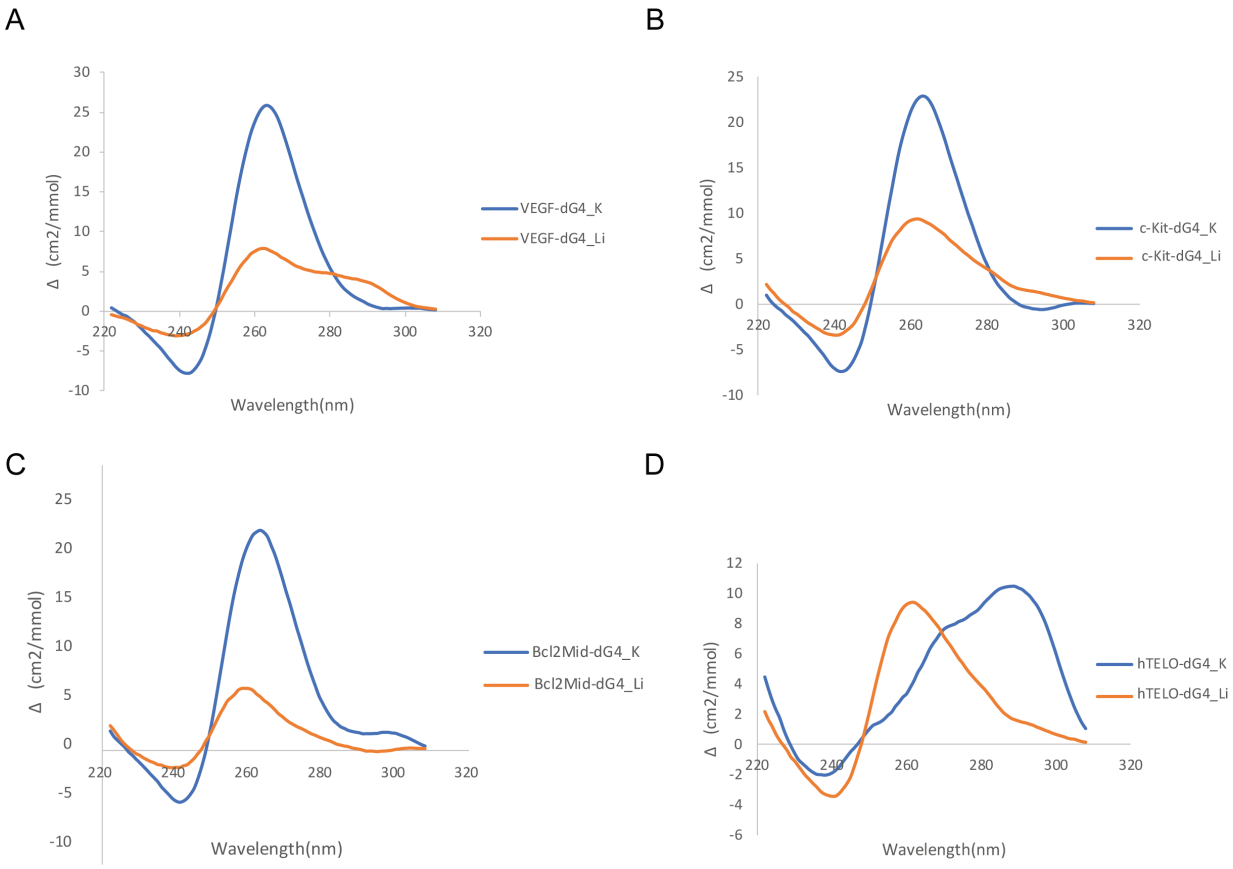

Figure S11. Biophysical characterization of dG4s excludes parallel topology as requisite for binding. CD profile of **A)** D-*VEGF*. **B)** D-*c-kit*. **C)** D-*Bcl2MidG4* and **D)** D-*hTELO* dG4s. The present of positive peak at 264 nm (except for *hTELO* dG4) and negative peak at 240 nm under 150 mM K^+ condition (compared to Li^+ condition), suggestive of parallel dG4 formation. All the dG4s tested formed parallel topology except for *hTELO* dG4 that showed CD signal for mixed topology.

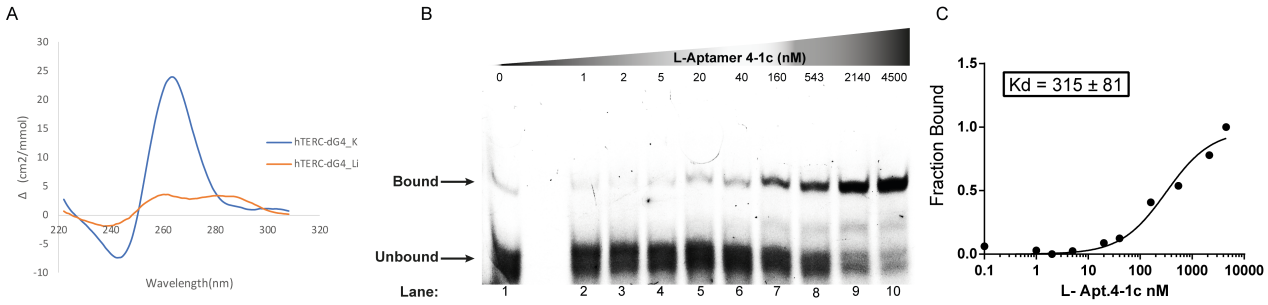

Figure S12. Biophysical characterization of D-*hTERC* dG4 and its binding with L-Apt.4-1c. **A)** CD profile of D-*hTERC* dG4 showed a parallel topology (due to the present of a positive peak at 264 nm and negative peak at 240 nm) in 150 mM K^+ condition compared to Li^+ condition. **B)** 12% EMSA native gel used to analyze the binding of L-Apt. 4-1c to FAM-D-*hTERC* dG4. 1 nM FAM-D-*hTERC* was used in all lanes, and the bound bands (upper band) increases with an increase in L-Apt.4-1c concentration from 0 – 4500 nM, while the lower bands (unbound bands) weakens. **C)** Saturation plot for the binding of L-Apt. 4-1c with FAM-D-*hTERC* dG4. The *Kd* was determined to be 315 ± 81 nM. This indicates a weaker binding affinity compared to its rG4 counterpart, *hTERC rG4*.

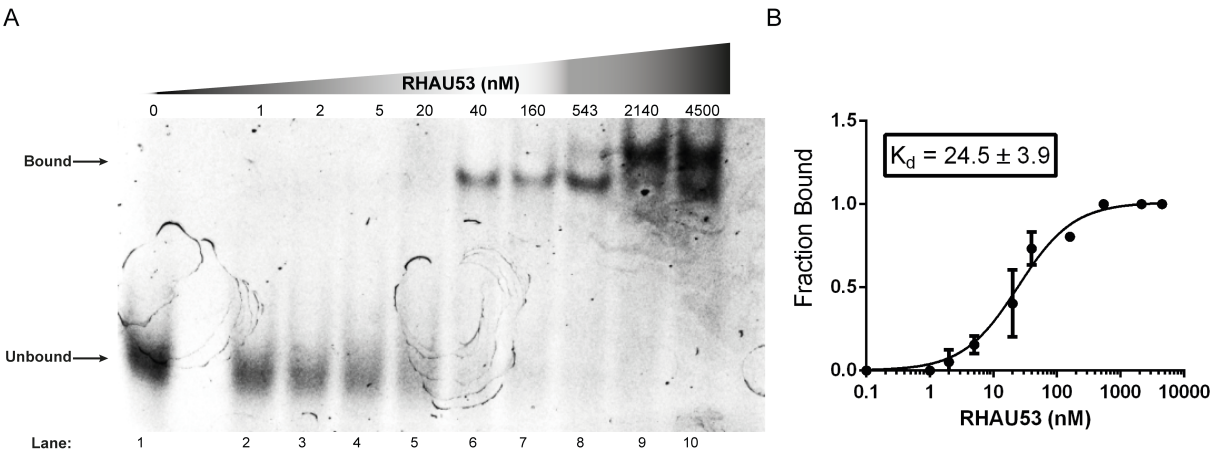

Figure S13. Binding of RHAU53 peptide with D-*hTERC* rG4. A) Binding of RHAU53 peptide with FAM-D-*hTERC* rG4 demonstrated by EMSA. 6% native gel was used to examine the bound band (upper bands) and unbound bands (lower bands). 1 nM FAM-D-*hTERC* rG4 was applied in all lanes and RHAU53 concentration ranged from 0 – 4500 nM. As RHAU53 concentration increases, the intensity of the bound bands (upper bands) increases while that of unbound band (lower bands) decreases. This indicates the RHAU53 peptide directly interact with the D-*hTERC* rG4. **B)** Saturation plot for RHAU53 peptide binding to D-*hTERC* rG4. The *Kd* was found to be 24.5 $\pm$ 3.9 nM.

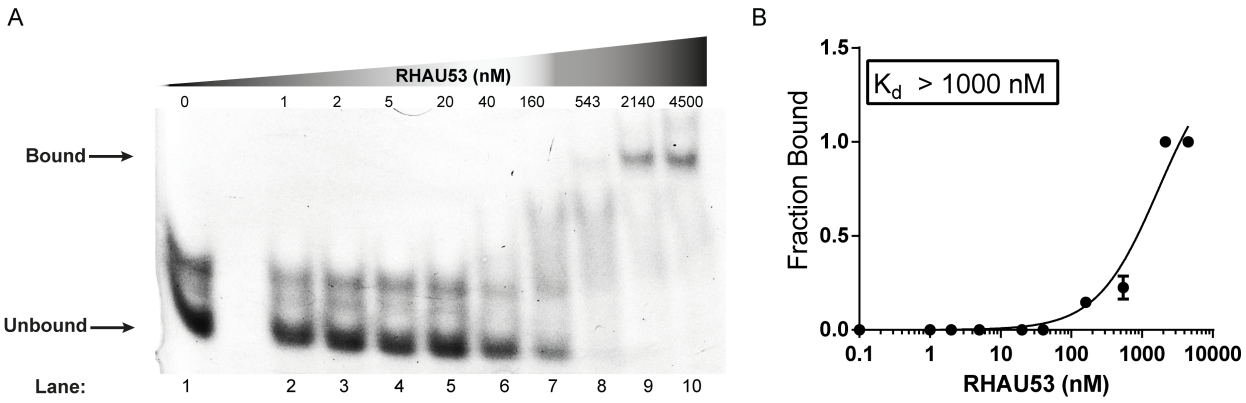

Figure S14. L-Apt.4-1c binds weakly to RHAU53 peptide at high concentration. Binding of RHAU53 peptide with FAM-L-Apt.4-1c demonstrated by EMSA. 6% native gel was used to resolve the bound band (upper bands) and unbound bands (lower bands). 1 nM FAM-L-Apt.4-1c was applied in all lanes and RHAU53 concentration ranged from 0 – 4500 nM. As RHAU53 concentration increases, the intensity of the bound bands (upper bands) increases while that of unbound band (lower bands) decreases. This indicates the RHAU53 peptide can interact with the L-Apt.4-1c much weaker compared to the D-*hTERC* rG4. **B)** Saturation plot for binding of FAM- L-Apt. 4-1c with RHAU53 peptide. The *Kd* value was found to be >1000 nM.

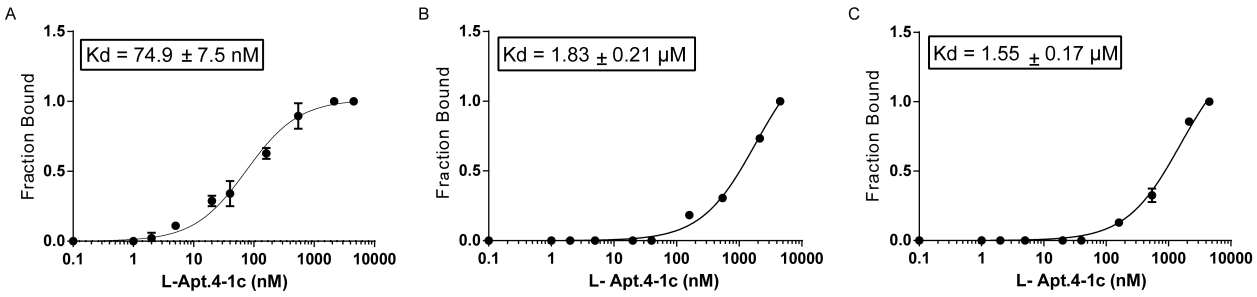

Figure S15. Binding analysis between L-Apt4.1c with D-*hTERC* rG4 and 5' or 3' extension constructs. A) Saturation plot for binding of L-Apt. 4-1c with wildtype FAM -D-*hTERC* rG4. The *Kd* value was found to be 74.9 ± 7.5 nM as demonstrated in Figure 3A. **B - C)** Saturation plot for binding of L-Apt. 4-1c with FAM 5' & 3'ext.-D-*hTERC* rG4. The *Kd* value was found to be 1.83 ± 0.21 µM and 1.55 ± 0.17 µM for 5'and 3'-ext. D-*hTERC* rG4s respectively.

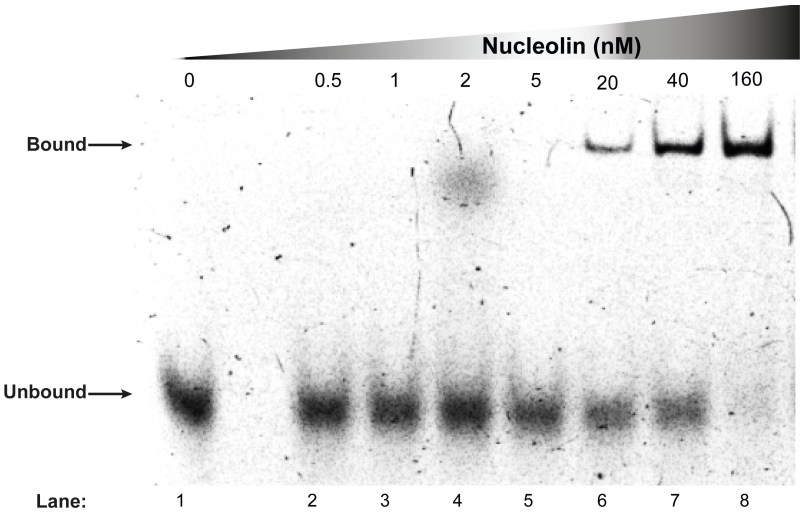

Figure S16. Binding of nucleolin protein with D-*hTERC* rG4. Same conditions as that in Fig. **S11** except nucleolin protein was used at 0-160 nM concentrations range. As nucleolin concentration increases, the intensity of the bound bands (upper bands) increases while that of unbound band (lower bands) decreases. This indicates the nucleolin protein directly interact with the D-*hTERC* rG4.

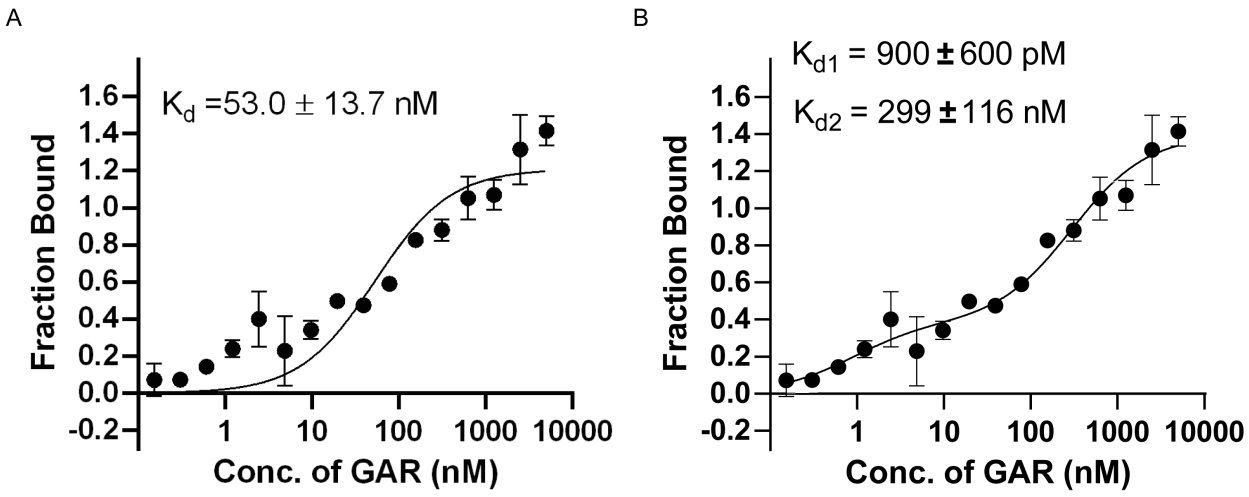

Figure S17. Binding of nucleolin GAR domain with D-*hTERC* rG4.. The binding saturation of GAR peptide with D– *hTERC* rG4 interaction monitored by MST. Reaction mixture contained 40 nM FAM *hTERC* rG4 and varying concentrations of GAR peptide (0.15 – 5000 nM). **A)** One site specific binding curve fitting. with *Kd* value of 53.0 ± 13.7 nM. B) Two sites specific binding curve fitting, with a *Kd1* value of 900 ± 600 pM and *Kd2* value of 299 ± 116 nM. All error bars represent standard deviation from 3 independent replicates.

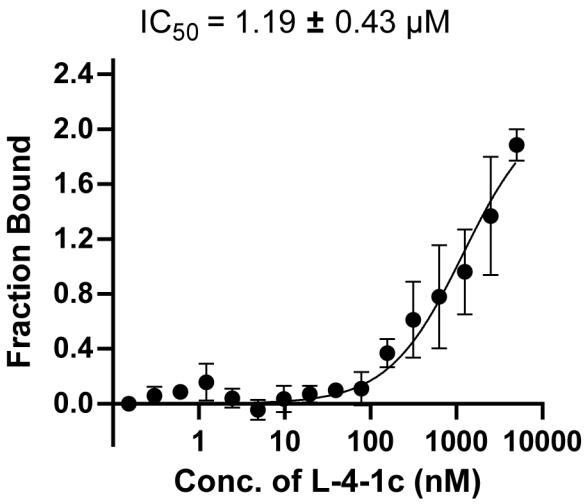

Figure S18. L-Apt.4-1c show little to no inhibitory effect towards *VEGF* dG4 – nucleolin interaction. Saturation plot of L-Apt.4-1c for its inhibition of *VEGF* dG4-nucleolin interaction. Reaction mixture contained 40 nM FAM *VEGF* dG4, 80 nM nucleolin and increasing concentrations of L-Apt.4-1c (0.15 – 5000 nM). The IC50 was found to 1.19 ± 0.43 µM, which is about 5 times lower than IC50 observed for *hTERC* rG4-nucleolin (218.9 ± 29.9 nM). All error bars represent standard deviation from 3 independent replicates.

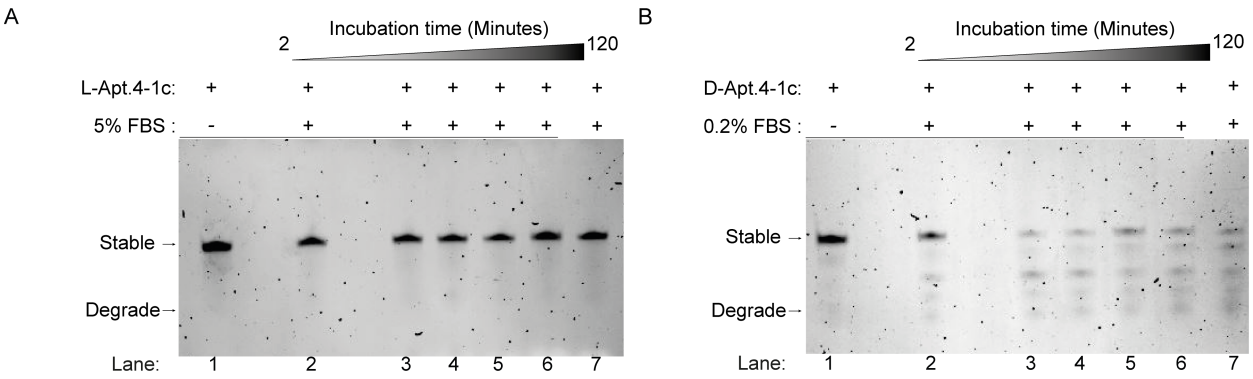

Figure S19. Biostability assay uncovers L-Apt.4-1c to be highly stable in serum-related condition than its D-aptamer counterpart. **A)** L-Apt.4-1c biostability is measured in 5% FBS over incubation periods of 2 – 120 minutes. 15% denaturing PAGE was used to resolve the stable (darker upper bands) and unstable or degraded (lower weaker bands) fragments. Each lane contained 10 nM L-Apt.4-1c dissolved in 5% FBS (except lane 1; negative control). As the incubation time increases the stability of L-Apt.4-1c (darker upper bands) remains intact with no lower weaker bands. This is suggestive of L-Apt.4-1c nuclease resistivity. **B)** Similar set-up with A) above except D-Apt.4-1c and 0.2% FBS was used instead. As the incubation time increases the stability of D-Apt.4-1c (darker upper bands) diminishes and the appearance of lower weaker bands. This is suggestive of D-Apt.4-1c succeptibility to nuclease degradation. 5% FBS was also tried, but the D-Apt.4-1c band was degraded in less than 2 minutes (data not shown).

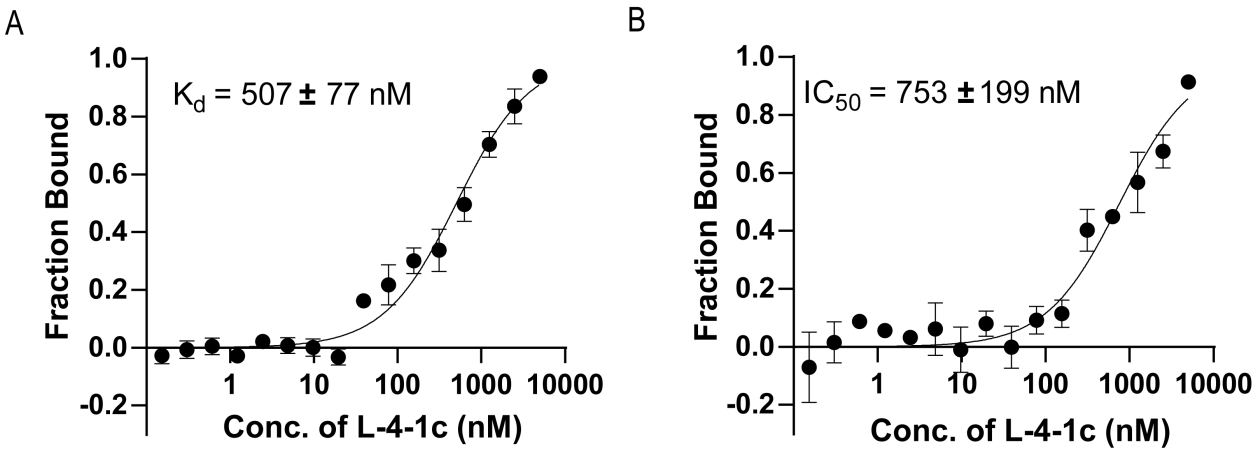

Figure S20. L-Apt.4-1c interacts with D-*hTERC* rG4 and effectively interferes with D-*hTERC* rG4 - nucleolin interaction in serum-related condition. A) Saturation plot of L-Apt.4-1c interaction with D-*hTERC* rG4. Reaction mixture contained 40 nM FAM *hTERC* rG4, 0.2% FBS, 1X ribonuclease inhibitor and increasing concentrations of L-Apt.4-1c (0.15 – 5000 nM). The *Kd* was found to be 507 ± 77 nM, which is higher than *Kd* observed without serum condition (59.1 ± 11.9 nM). **B)** Saturation plot of L-Apt.4-1c interference with D-*hTERC* rG4 – nucleolin interaction. Reaction mixture contained 40 nM FAM *hTERC* rG4, 80 nM nucleolin, 0.2% FBS, 1X ribonuclease and protease inhibitors and increasing concentrations of L-Apt.4-1c (0.15 – 5000 nM). The IC50 was found to be 753 ± 199 nM. All error bars represent standard deviation from 3 independent replicates.

References

1. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406-3415.