Supplementary Information

Identification of selective G-quadruplex DNA binder by a multistep virtual screening approach

Jin-Qiang Hou,[ab] Shuo-Bin Chen,[a] Li-Peng Zan,[a] Tian-Miao Ou,[a] Jia-Heng Tan,*[a] Leonard G. Luyt[b,c] and Zhi-Shu Huang*[a]

[a] School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou 510006, China.

[b] London Regional Cancer Program, 790 Commissioners Road East, London, Ontario N6A 4L6 Canada.

[c] Depts. Oncology, Chemistry, Medical Imaging, The University of Western Ontario, London, ON N6A 3K7 Canada.

Table of Contents:

1. Experimental Section

1.1 Materials

All oligonucleotides used in this study were purchased from Invitrogen (Shanghai, China). Calf thymus DNA (CT DNA) and thiazole orange (TO) were purchased from Sigma-Aldrich (Singapore). Compounds used in the experimental validation were purchased from ChemBridge (USA). The pGL4.10[*luc2*] Vector was purchased from Promega (USA). Tumor cell lines were obtained from the American Type Culture Collection. All the oligonucleotides were dissolved in relevant buffer. Their concentrations were determined from the absorbance at 260 nm on the basis of respective molar extinction coefficients using NanoDrop 1000 Spectrophotometer (Thermo Scientific). To obtain G-quadruplex formation, highly concentrated solutions of oligonucleotides were annealed in relevant buffer containing KCl by heating to 90 °C for 5 min, followed by gradual cooling to room temperature and incubation at 4 °C overnight. Stock solutions of compounds (10 mM) used in the experimental validation were dissolved in DMSO and stored at -80 °C. Further dilutions of compounds to working concentrations were made with relevant buffer immediately prior to use.

1.2 Virtual Screening Studies

Molecular docking was carried out with Surflex-dock (Tripos, Inc. St. Louis, MO). The protomol represents a set of molecular fragments that characterizes the binding site of the receptor (Fig. S1). For G-quadruplex docking, the protomol was generated using a receptor-based approach. While for duplex docking, it was generated using ligand-based approach. The "proto_thresh" and "proto_bloat" are two important factors which determine how far the protomol extends into and outside of the concavity of the binding site.[1,2] Both of them significantly affect the size and extent of the protomol. For docking with G-quadruplex, "proto_thresh" and "proto_bloat" were set to 0.01 and 0 in order to obtain a protomol which encompasses the G-quartets at both ends as well the groove of G-quadruplex. For docking with duplex, proto_thresh" and "proto_bloat" were set to 0.5 and 1. "Additional starting conformations per Molecule" was set to 20. "Pre-Dock Minimization" and "Post-Dock Minimization" were switched on. All other parameters accepted default settings. The NMR structure of *c-MYC* G-quadruplex DNA 5′-d(TGAGGGTGGGTAGGGTGGGTAA)-3′ (PDB id 1XAV)[3] was used as the G-quadruplex target. The crystal structure of duplex DNA 5′-d(CGATCG)-3′ complexed with ellipticine (PDB id: 1Z3F)[4] that is a representative structure for a ligand intercalating into duplex DNA was used for the modeling of intercalation interaction. While the crystal structure of duplex DNA 5′-d(GGCCAATTGG)-3′ complexed with distamycin (PDB id: 1K2Z)[5] which is a representative structure for a ligand binding to the groove region of duplex DNA was used for modeling of the grooving binding.

1.3 Surface Plasmon Resonance (SPR) Studies

SPR measurements were performed on a ProteOn XPR36 Protein Interaction Array system (Bio-Rad Laboratories, CA) using a Streptavidin-coated GLH sensor chip. Biotinylated oligonucleotides (*c-MYC* G-quadruplex DNA: 5′-biotin-d(ACGTACGTGGGGAGGGTGGGGAGGGTGGGGAAGGTGGGG)-3′; duplex DNA: 5′-biotin-d(TTCGCGCGCGTTTTCGCGCGCG)-3′) were attached to the chip. In a typical experiment, biotinylated DNA was folded in filtered and degassed running buffer (50 mM Tris-HCl, 100 mM KCl, pH 7.4). The DNA samples were then captured (~1000 RU) in two flow cell, leaving one flow cell as a blank. Solutions of tested compounds were prepared with running buffer through serial dilutions of stock solution. Six concentrations were injected simultaneously at a flow rate of 100 μL/min for 150 s of association phase, followed with 250 s of dissociation phase at 25 °C. The GLH sensor chip was regenerated with short injection of 50 mM NaOH between consecutive measurements. The final graphs were obtained by subtracting blank sensorgrams from different DNA sensorgrams. Data were analyzed with ProteOn manager software, using the Langmuir model for fitting data.

1.4 UV Titration Experiment

UV titration experiments were performed on a UV-2450 spectrophotometer (Shimadzu, Japan) using 1 cm path length quartz cuvette. The *c-MYC* G-quadruplex DNA 5′-d(TGAGGGTGGGTAGGGTGGG TAA)-3′ was employed in the assay. Small aliquots of a stock solution of oligonucleotide were added into the solution containing **VS10** at fixed concentration (10 μM) in Tris-HCl buffer (10 mM, pH 7.2) containing 100 mM KCl. The final concentration of oligonucleotide was varied from 0 to 20.0 μM. After each addition of oligonucleotide, the reaction was stirred and allowed to equilibrate for at least 1 min and absorbance measurement was taken. The binding affinity of **VS10** for *c-MYC* G-quadruplex DNA was determined by fitting the absorption change at 405 nm to the Scatchard binding model.

1.5 G-Quadruplex Fluorescent Intercalator Displacement (G4-FID) Assay

G4-FID assays were performed on a LS-55 luminescence spectrophotometer (Perkin-Elmer, USA). A quartz cuvette with 2 mm × 10 mm path length was used for the spectra recorded at 10 nm excitation and emission slit widths. The *c-MYC* G-quadruplex DNA 5′-d(TGAGGGTGGGTAGGGTGGGTAA)-3′ was employed in the assay. The binding affinity of TO for *c-MYC* G-quadruplex DNA was first determined by fitting the emission change at 535 nm to the Scatchard binding model to validate the usability of such DNA. To a mixture of *c-MYC* G-quadruplex DNA (0.25 μM) and TO (0.50 μM) in Tris-HCl buffer (10 mM, pH 7.4) containing 100 mM KCl, an increasing amount of compound sample was then added. After an equilibration time of 3 min the emission spectrum was recorded between 510 and 650 nm with an excitation wavelength of 500 nm. The percentage of displacement is calculated from the fluorescence intensity using the formula: percentage of displacement (TO displacement %) = 100-[(fluorescence

intensity of sample/fluorescence intensity of standard)×100]. The fluorescence intensity of standard is the fluorescence of TO binding to DNA without addition of compound sample. The percentage of displacement is plotted as a function of the concentration of added compound. The DNA affinity was evaluated by the concentration of compound required to decrease the fluorescence of the probe by 50%, noted DC_{50}, and determined after non-linear fitting of the displacement curve.

1.6 Circular Dichroism (CD) Studies

CD studies were performed on a Chirascan circular dichroism spectrophotometer (Applied Photophysics, UK). The *c-MYC* G-quadruplex DNA 5′-d(TGAGGGTGGGTAGGGTGGGTAA)-3′ was employed in the assay. A quartz cuvette with a 4 mm path length was used for the recording of spectra over a wavelength range of 230–330 nm with a 1 nm bandwidth, 1 nm step size and time of 0.5 s per point. CD melting was performed at a fixed concentration of G-quadruplex (5.0 μM), either with or without a fixed concentration (20.0 μM) of **VS10** in Tris-HCl buffer (10 mM, pH 7.2) containing 10 mM KCl. The data was recorded at intervals of 5 °C over a range of 10–95 °C, with a heating rate of 0.5 °C/min. A buffer baseline was collected in the same cuvette and was subtracted from the sample spectra. Final analysis of the data was conducted using Origin 7.0 (OriginLab Corp.).

1.7 Fluorescence Resonance Energy Transfer (FRET) Melting Experiment

FRET melting experiments were performed on a Roche LightCycler 2 real-time PCR machine. The dual labeled oligonucleotide (*c-MYC* G-quadruplex DNA: 5′-*FAM*-d(TGAGGGTGGGTAGGGTGGG TAA)-*TAMRA*-3′; donor fluorophore *FAM* is 6-carboxyfluorescein; acceptor fluorophore *TAMRA* is 6-carboxytetramethylrhodamine) was employed in the assay. Fluorescence melting curves were determined using a total reaction volume of 24 μL, with 0.6 μM of labeled oligonucleotide in Tris-HCl buffer (10 mM, pH 7.2) containing 5 mM KCl. Fluorescence readings with excitation at 470 nm and detection at 530 nm were taken at intervals of 1 °C over the range 37-99 °C, with a constant temperature being maintained for 30 s prior to each reading to ensure a stable value. The melting of the *c-MYC* G-quadruplex DNA was monitored alone or in the presence of 10 μM of compounds without and with double-stranded competitor CT DNA. Final analysis of the data was carried out using Origin 7.0 (OriginLab Corp.).

1.8 Molecular Dynamics (MD) Simulation

The NMR structure of *c-MYC* G-quadruplex DNA 5′-d(TGAGGGTGGGTAGGGTGGGTAA)-3′ (PDB id 1XAV)[3] was used as template to study the interactions between **VS10** and G-quadruplex. Ligand structures were constructed with SYBYL7.3 (Tripos Inc., St Louis, MO, USA) and optimized with GAUSSIAN 03 at HF/6-31G* level.[6] MD simulations were performed using the AMBER 12 program suite.[7] Based on the electrostatic potential (ESP) calculations at the ab initio HF/6-31G* level, the partial atomic charges for ligands were refined by using RESP calculation in antechamber module. Other

parameters for ligands were taken from the Generalized Amber Force field (GAFF).[8] Two K^+ ions were included in the central core of the complexes. The ligand-receptor system was then solvated in a truncated octahedron box of TIP3P water molecules with a 10.0 Å buffer along each dimension. Additional positively-charged counter ions were added in the system to neutralize the negative charge on the DNA backbone. The AMBER ff99 force field was applied for G-quadruplexes, ions, and water molecules.[9] Periodic boundary conditions were applied to avoid the edge effect. The Particle Mesh Ewald (PME) method was used to calculate long-range electrostatic interactions with a 10 Å residue-based cutoff.[10] The hydrogen bonds were constrained using SHAKE algorithm. The final systems were subjected to initial minimization to equilibrate the solvent and counter cations, and were then heated from 0 to 300 K in a 200 ps simulation, and followed with a 1 ns simulation to equilibrate the density. Afterwards, 50 ns MD simulation was performed in an NPT ensemble at 1 atm and 300 K. The calculation was carried out with the PMEMD module. The MM/PBSA method and NMODE module were used to calculate the binding free energy and the entropic contribution.[11] All the waters and counter-ions were stripped off. A total of 100 snapshots were taken from the last 2 ns trajectory with an interval of 20 ps. Calculations were performed on high-performance computing (HPC) facilities within the Compute Canada network.

1.9 NMR Spectroscopy Studies

NMR spectroscopy experiments were performed on an AVANCE AV 600 MHz spectrometer (Bruker, USA). The *c-MYC* G-quadruplex forming sequence Pu24I (5′-d(TGAGGGTGGIGAGGGTGGGGAA GG)-3′) was employed in the assay. All of the titration experiments were carried out at 25 °C in a 90% H_2O/10% D_2O solution containing 150 mM KCl and 25 mM potassium phosphate buffer (pH 7.2). The concentration of the Pu24I was 0.5 mM for NMR measurements.

1.10 Plasmid Construction

A DNA fragment of the human *c-MYC* gene promoter region, containing wild-type *c-MYC* G-quadruplex forming sequence, was extracted from Hela cell total DNA by PCR. This fragment was then inserted into pGL4.10[*luc2*] Vector, named *c-MYC* native construct. Then the *c-MYC* native construct was site-mutagenesis as *c-MYC* mutated construct as showed below:

c-MYC native construct: AGGGGCGCTTATGGGGAGGGTGGGGAGGGTGGGGAAGG

c-MYC mutated construct: AGGAGCGCTTATGGAGAGGGTGGAGAGGGTGGAGAAGG

The conformations of wild-type and mutant *c-MYC* G-quadruplex forming sequence were further identified by CD spectroscopy. As showed in Fig. S6, a positive peak at 262 nm was observed for native sequence, indicating the formation of parallel G-quadruplex structure. While for mutated sequence, the intensity of peak was reduced and a 2 nm shift of the peak was observed, suggesting that the formation of G-quadruplex was significantly suppressed.

1.11 Transfection and Luciferase Activity Assay

DNA transfections were performed with Hela cells in log phase. The *c-MYC* native construct or mutated construct and pRL-SV40 were cotransfected into MCF-7 cell with Lipofectamine 2000 transfection reagent (Invitrogen, USA) as manufacturer's instructions. Then after 6 h of transfection, different concentrations of the compounds were added into medium. Cells were incubated for another 36 h at 37 °C in 5% CO_2, then the luciferase activity was evaluated by Dual-Glo Luciferase Assay System (Promega, USA). The efficiency of transfection was normalized by pRL-SV40.

1.12 RT-PCR Studies

For RNA extraction, about 1×10^6 CA46 or RAji cells were seeded in 6-well plates with appropriate concentrations of compounds and incubated for 24 h. Cell pellets from each well of the culture plates were harvested and lysed in TRIpure solution. Total RNA was extracted with RNAiso Plus Kit (Takara, Japan) according to manufacturer's protocol and eluted in distilled, deionized water with 0.1% diethyl pyrocarbonate (DEPC) to a final volume of 20 μL. RNA was quantitated spectrophotometrically and stored at -80 °C.

Total RNA was used as a template for reverse transcription using Reverse Transcriptase M-MLV (RNase H-, Takara, Japan) according to the manufacturer's protocol. For each cDNA sample, a 20 μL reaction, containing 1 × PCR buffer, 500 μM dNTPs, 0.15 μM *β-actin* primers, 0.15 μM *c-MYC* primers, 1 U of Taq polymerase, and 2 μL of the cDNA template, was performed. Both *β-actin* and *c-MYC* were amplified by using PCR. The PCR products were resolved in 1% agarose gel, and photographs were taken on an AlphaImager (ProteinSimple, USA).

2. Other Supporting Tables, Spectra and Graphs

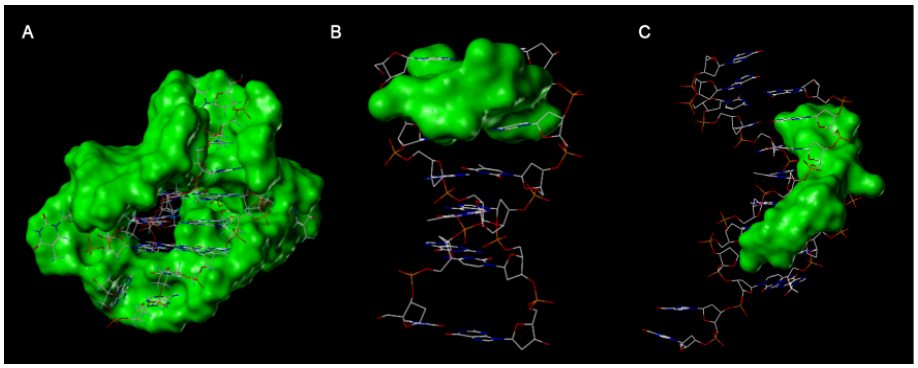
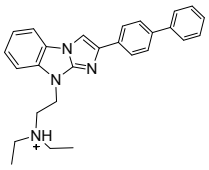
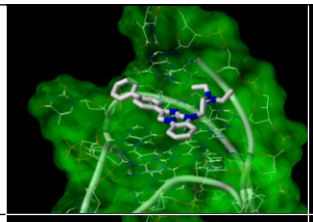
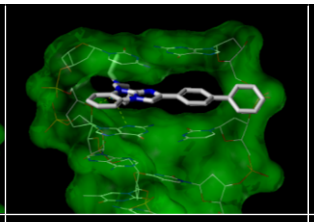
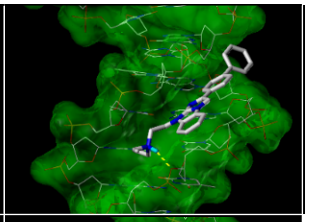
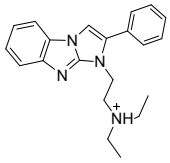
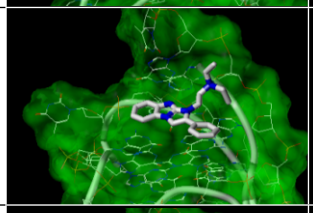
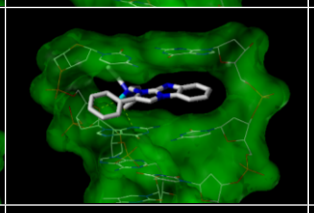
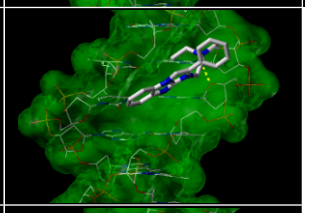
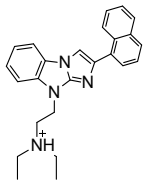
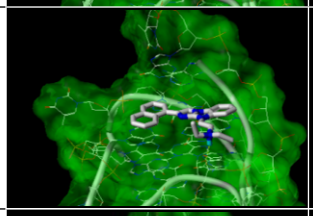
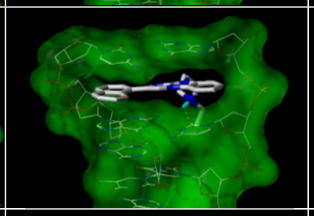
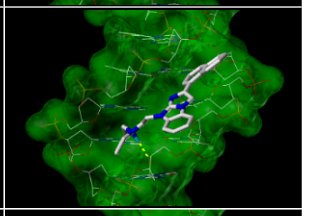
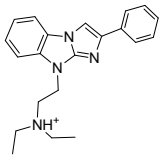
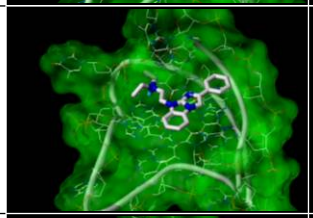
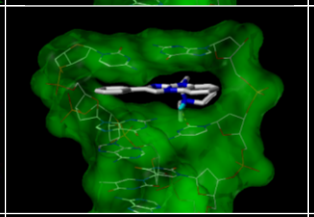
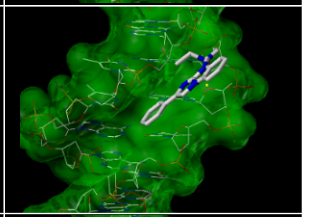
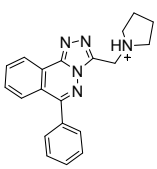
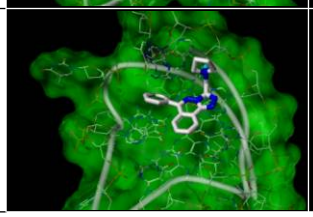
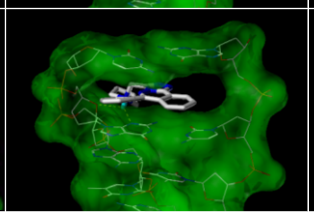
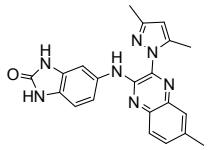
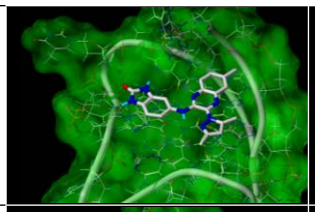
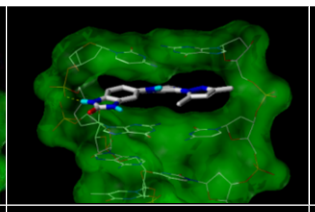
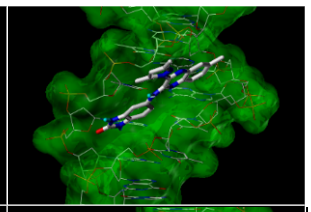
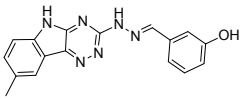
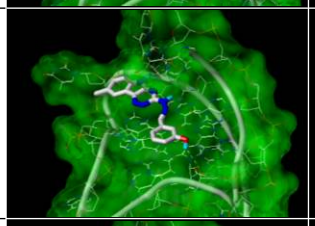
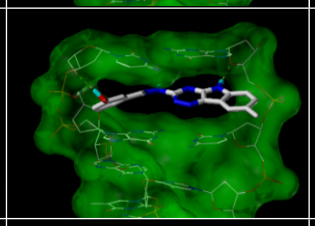
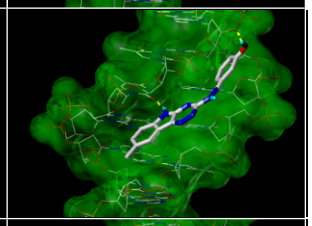
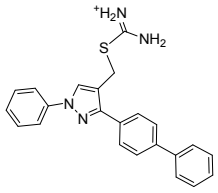
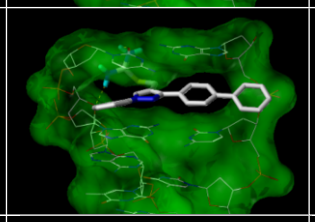
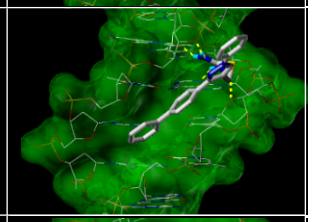
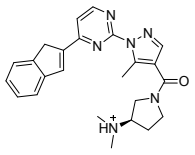
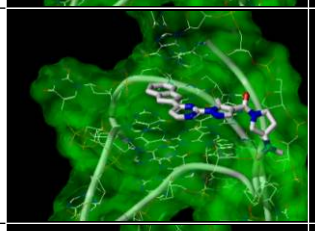
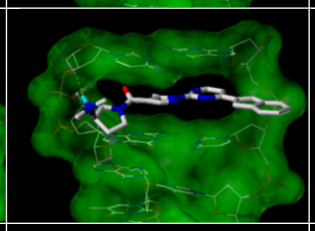
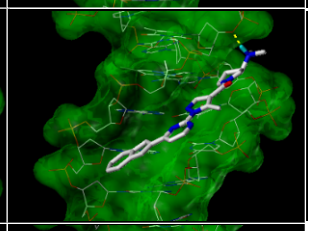
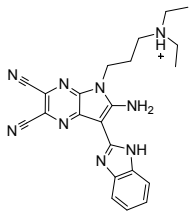
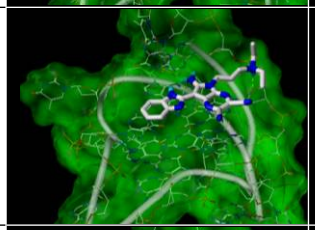
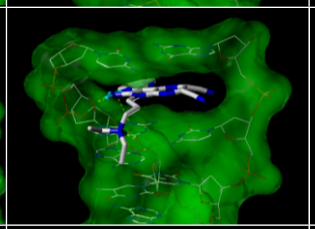
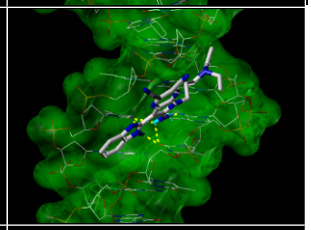
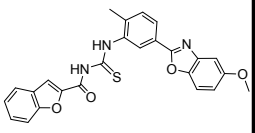
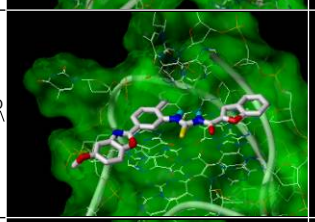
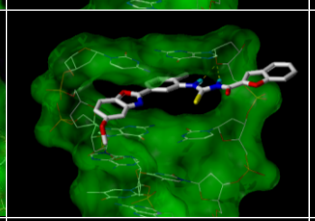
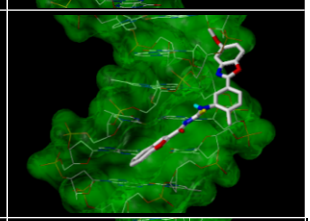
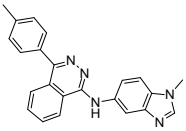
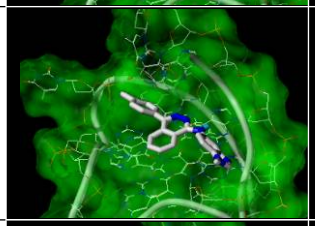
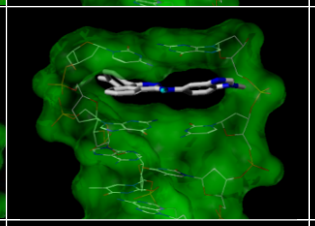
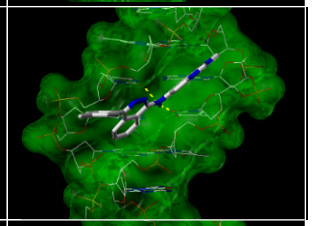
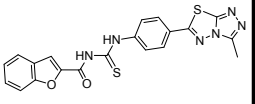
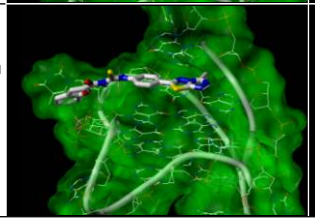
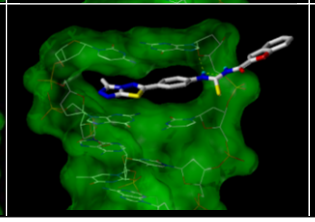
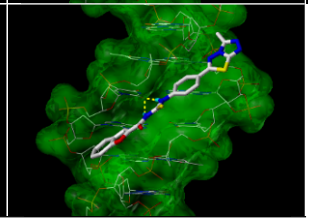


Fig. S1 Protomols were depicted with Connolly surfaces in Green. (A) *c-MYC* G-quadruplex DNA (1XAV); (B) Duplex DNA (1Z3F); (C) Duplex DNA (1K2Z).

Table S1 The binding modes of the selected compounds with *c-MYC* G-quaduplex and duplex DNA.

ID	Compounds	*c-MYC* G-quadruplex	Duplex	Duplex
VS1				
VS2				
VS3				
VS4				
VS5				

VS6				
VS7				
VS8				
VS9				
VS10				
VS11				
VS12				
VS13				

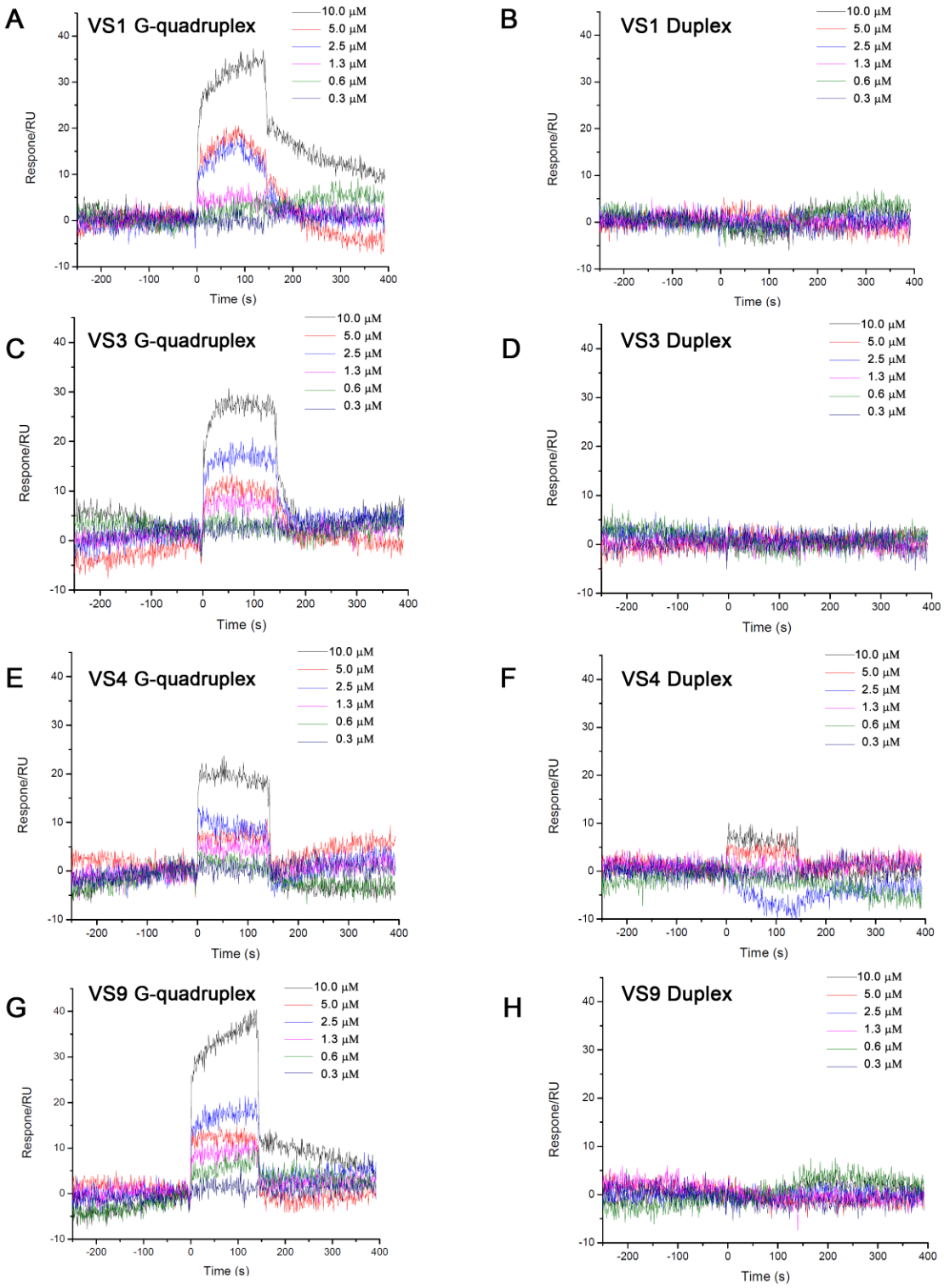


Fig. S2 SPR sensorgrams for binding of the selected compounds to the immobilized *c-MYC* G-quadruplex (A:**VS1**; C:**VS3**; E:**VS4**; G:**VS9**) and duplex DNA ((B:**VS1**; D:**VS3**; F:**VS4**; H:**VS9**). All of the compounds exhibited selectivity for *c-MYC* G-quadruplex over duplex DNA.

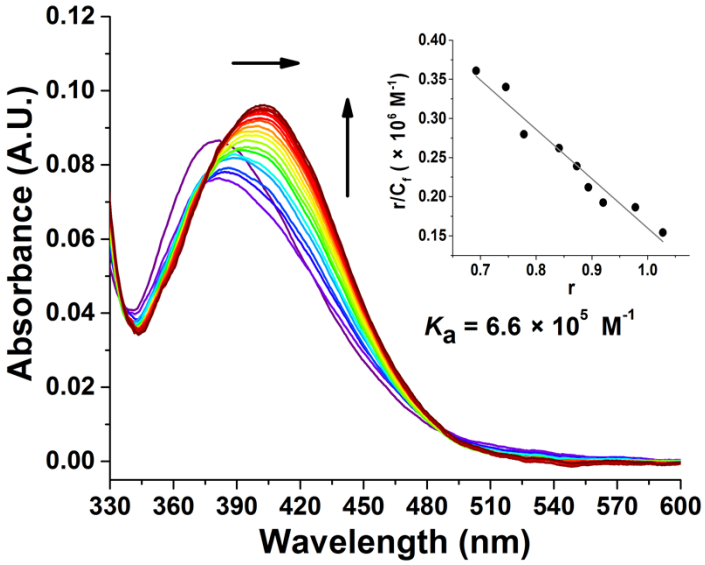


Fig. S3 UV titrations of 10 μM **VS10** with stepwise addition of *c-MYC* G-quadruplex DNA in 10 mM Tris-HCl buffer, 100 mM KCl, pH 7.2. The final concentration of oligonucleotide was varied from 0 to 20.0 μM. The binding affinity of **VS10** for *c-MYC* G-quadruplex DNA was determined by fitting the absorption change at 405 nm to the Scatchard binding model that were plotted in the inner panel.

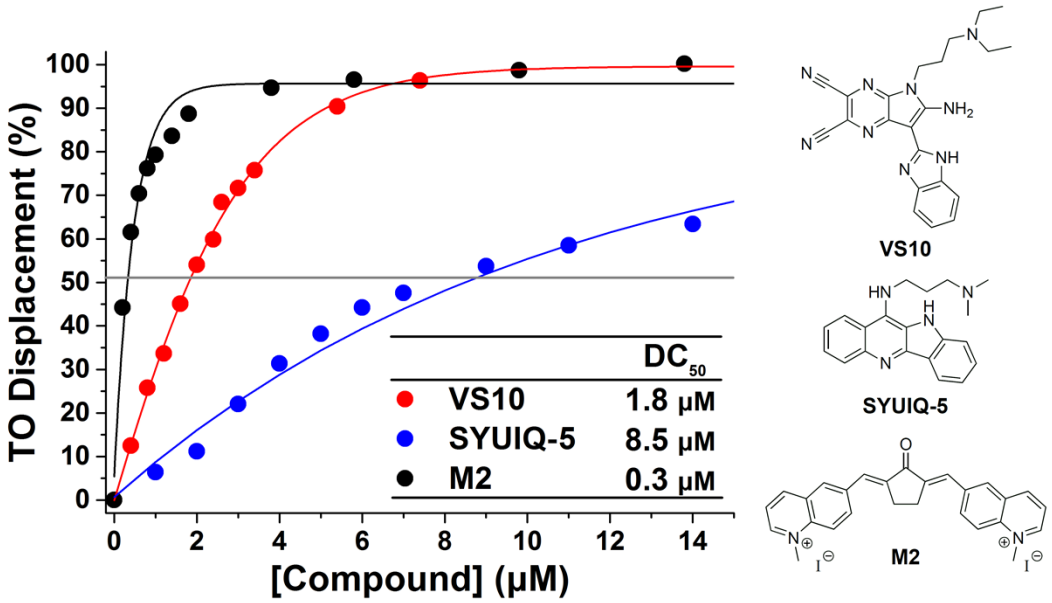


Fig. S4 G4-FID results of **VS10**, SYUIQ-5 and M2 with *c-MYC* G-quadruplex DNA in 10 mM Tris-HCl buffer, 100 mM KCl, pH 7.4. The binding affinity of TO for *c-MYC* G-quadruplex DNA was 3.2×10^6 M^{-1}, suggesting this DNA could be used in the assay.

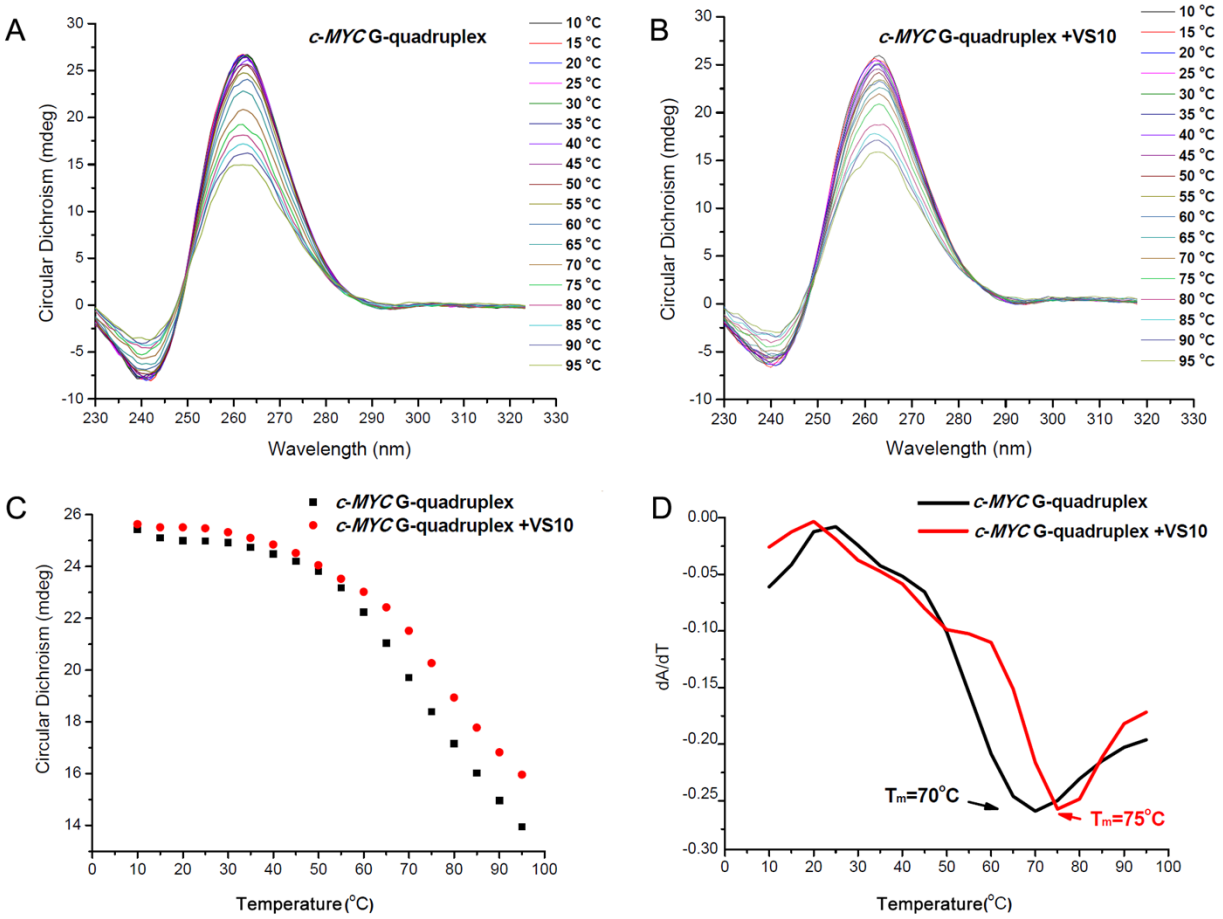


Fig. S5 CD spectra of *c-MYC* G-quadruplex DNA (A) without or (B) with **VS10** in Tris-HCl buffer (10 mM, pH 7.2) with 10 mM KCl at temperature from 10 to 95 °C. (C) CD melting curves collected at 262 nm from (A) and (B). (D) First derivative plots of the data shown in (C).

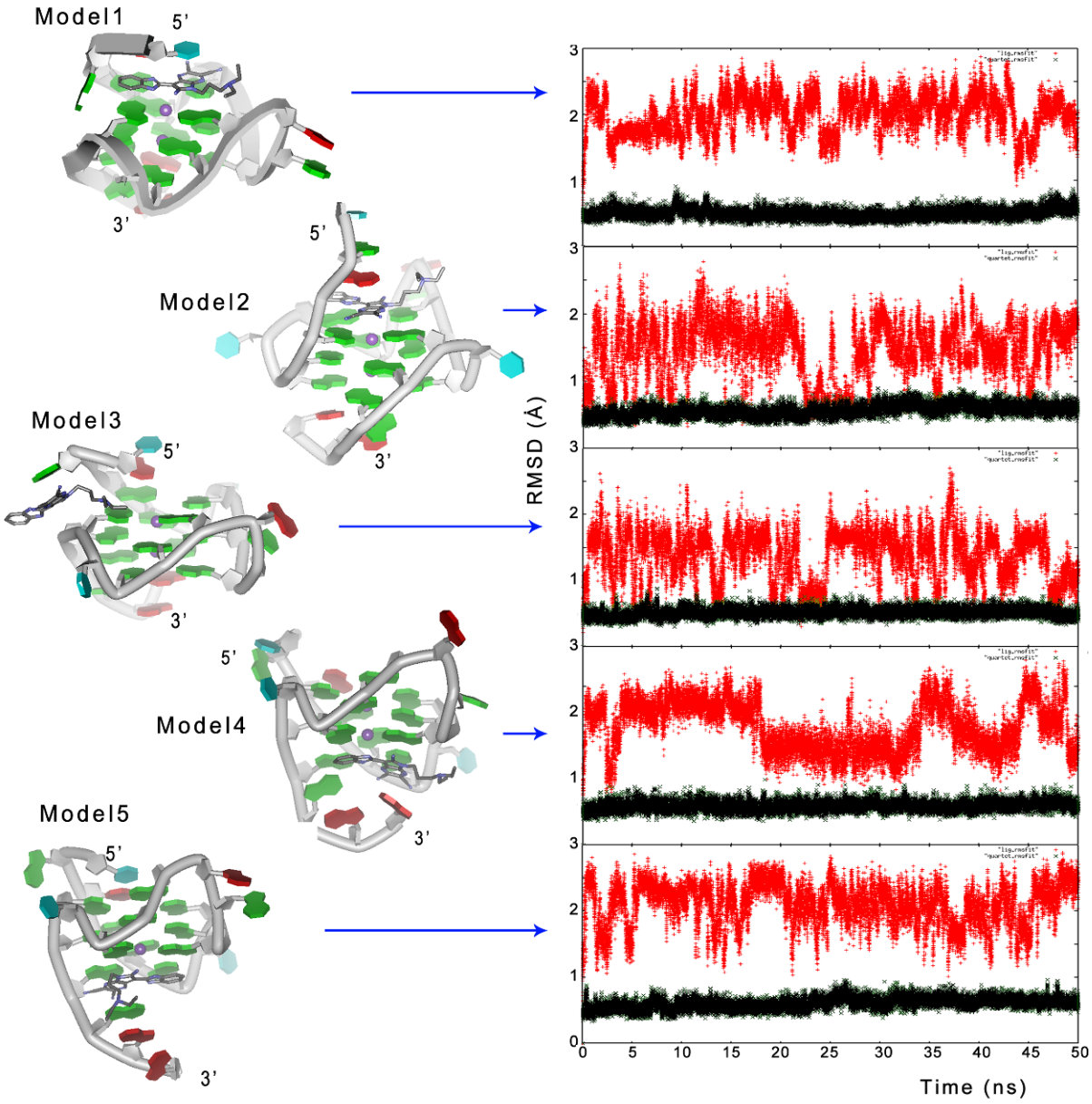


Fig. S6 Left hand side is the binding modes of **VS10** with *c-MYC* G-quadruplex DNA in five models. Right hand side is the corresponding RMSD of the G-quartets (black) and **VS10** (red) of the models during MD simulations.

Table S2 Computed binding free energies (kcal/mol) of ligand-quadruplex complexes in five models.

	ΔE_{vdw}	ΔE_{elec}	ΔE_{PB}	ΔG_{np}	ΔG_{gas}	ΔG_{solv}	ΔG_{tot}	$-T\Delta S$	ΔG_{bind}
Model1	-59.9	-498.0	545.2	-4.7	-557.9	540.5	-17.3	-23.8	6.5
Model2	-61.4	-543.2	579.6	-4.8	-604.6	574.9	-29. 7	-15.6	-14.1
Model3	-32.1	-570.1	584.8	-3.2	-602.1	581.6	-20.5	-18.6	-1.9
Model4	-50.6	-461.2	503.4	-4.1	-511.7	499.3	-12.5	-19.4	6.9
Model5	-57.1	-473.0	509.5	-4.8	-530.1	504.7	-25.4	-16.1	-9.3

ΔE_{vdw} is the van der Waals contribution from molecular mechanical (MM) force field. ΔE_{elec} is the electrostatic interaction calculated with the MM force field. ΔE_{PB} is the electrostatic contribution to the solvation energy calculated by the PB approach. ΔG_{np} is the nonpolar contribution to the solvation energy. ΔG_{gas} is the gas phase energy ($\Delta E_{vdw} + \Delta E_{elec}$). ΔG_{solv} is the total solvation energy ($\Delta E_{PB} + \Delta G_{np}$). ΔG_{tot} is the total energy without solute entropic contribution ($\Delta G_{gas} + \Delta G_{solv}$). $-T\Delta S$ is solute entropic contribution, where T = temperature and S = sum of translational, rotational, and vibrational entropies. ΔG_{bind} is the total energy with solute entropic contribution ($\Delta G_{tot} - T\Delta S$).

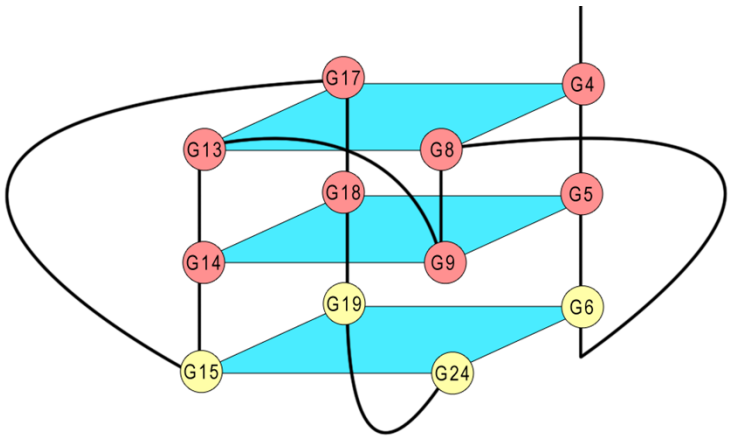


Fig. S7 Schematic diagram of G-quadruplex Pu24I with the assigned guanines. The guanines exhibiting any shifts on imino proton resonances are shown in red.

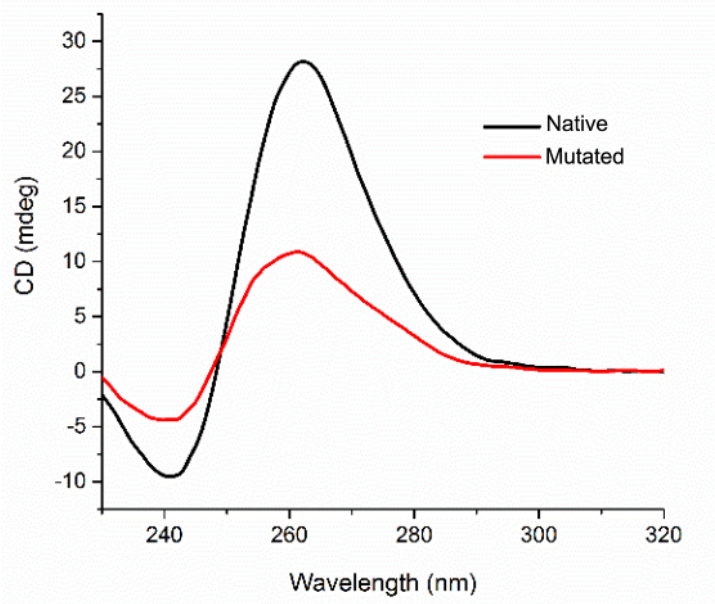


Fig. S8 CD spectra of native (black line) and mutated (red line) sequences in constructs, suggesting the formation of G-quadruplex was suppressed by sequence mutation.

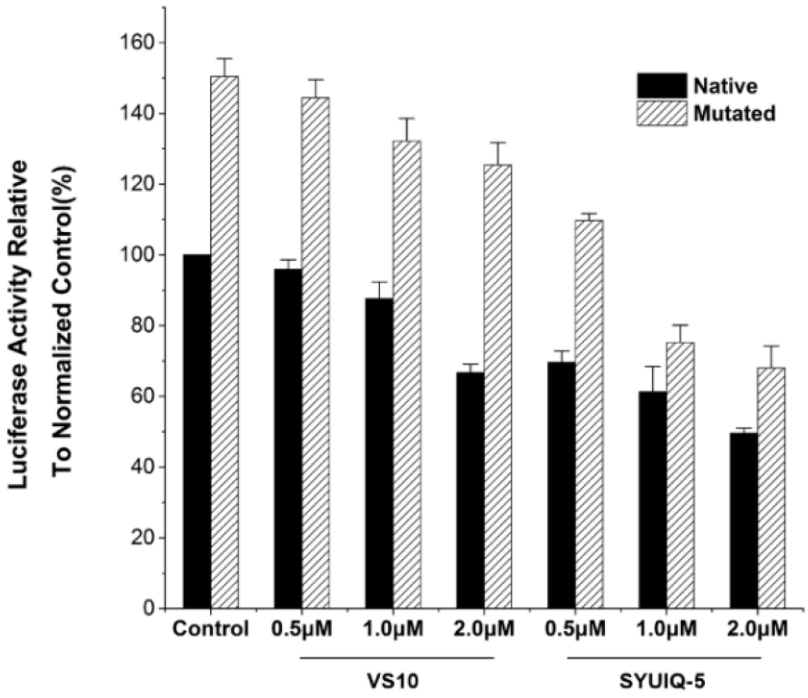


Fig. S9 Effects of **VS10** and SYUIQ-5 at the concentration of 0.5, 1.0 and 2.0 μM on *c-MYC* promoter activity.

References

1. A. N. Jain, *J. Comput. Aided. Mol. Des.*, 2007, **21**, 281.
2. A. N. Jain, *J. Med. Chem.*, 2003, **46**, 499.
3. A. Ambrus, D. Chen, J. Dai, R. A. Jones and D. Yang, *Biochemistry*, 2005, **44**, 2048.
4. A. Canals, M. Purciolas, J. Aymami and M. Coll, *Acta Crystallogr. D Biol. Crystallogr.*, 2005, **61**, 1009.
5. K. Uytterhoeven, J. Sponer and L. Van Meervelt, *Eur. J. Biochem.*, 2002, **269**, 2868.
6. Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, *Gaussian, Inc.*, Wallingford CT, 2004.
7. D.A. Case TAD, T.E. Cheatham III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q.X. Cai Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman, *AMBER 12*, University of California, San Francisco. 2012.
8. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157.
9. T. E. Cheatham, 3rd, P. Cieplak and P. A. Kollman, *J. Biomol. Struct. Dyn.*, 1999, **16**, 845.
10. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089.
11. B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *J. Chem. Theory. Comput.*, 2012, **8**, 3314.