SUPPLEMENTARY INFORMATION

Targeting a c-MYC G-quadruplex DNA with a fragment library

Hamid R. Nasiri,[†] Neil M. Bell,[†] Keith I. E. McLuckie,[‡] Jarmila Husby,[§] Chris Abell,[†] Stephen Neidle[§*] and Shankar Balasubramanian[†,‡,*]

[†]Department of Chemistry, The University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK
[‡]Cancer Research UK Cambridge Institute, Li Ka Shing Centre, Cambridge, CB2 0RE, UK
[§]UCL School of Pharmacy, 29-39 Brunswick Square, London WC1N 1AX, UK

SUPPLEMENTARY MATERIALS AND METHODS

SUPPLEMENTARY FIGURES

SUPPLEMENTARY TABLES

SUPPLEMENTARY REFERENCES

EXPERIMENTAL SECTION:

Reagents

The c-MYC DNA oligonucleotides were synthesized and purified by IBA GmbH, and stock solutions (100 μM) were made by resuspending the DNA in molecular biology grade water and quantified by A_{260} at 95 °C, using ε_{260} values as provided by the manufacturers, before the DNAs were aliquoted and stored at −80 °C. All samples were freshly prepared prior to each experiment, the Myc DNA was dissolved in an Tris/KCl, annealing buffer (Tris 10 mM Tris- HCl pH 7.4 & 100 mM KCl) heated at 90 °C for 5 min, and slowly cooled down to room temperature overnight. All other reagents were purchased from Sigma-Aldrich unless otherwise stated, fragments determined as hit from the initial intercalator-displacment assay (IDA) were reordered for the biological assays (Table S1).

Intercalator-Displacement Assay (IDA)

Preliminary experiments determined the K_d for TO binding to c-MYC DNA to be 3.5 μM using the conditions 0.25 μM DNA, 0.5 μM Thiazole Orange, 2.5% DMSO, 20 mM Na caco, 140 mM KCl, pH 7 (25 μL/well) (Figure S3). For assay optimisation 32 negative and 32 positive controls were used, DMSO only (2.5% v/v) wells, which contained no small molecule, were used as a negative control, while positive control wells consisted target DNA and the intercalator TO. By using both negative and positive controls, an excellent Z'-score was calculated (mean Z'=0.89 for five tested plates), indicating a statistically robust assay.[1]

Fragment Screen

The IDA was performed using a 1377 fragment molecule library, comprised of structurally and chemically diverse fragments, with each member obeying the 'rule of three', where; MW< 300 Da, cLogP < 3, with ≤ 3 H bond donors and acceptors).[2] All fragment molecules were ≥ 95% purity and had >1 mM aqueous solubility. They were obtained from commercial sources, or synthesised internally. For screening, 1.25 μL of each fragment from its original 100 mM DMSO stock plate was transferred to a 384 well assay plate (low volume flat bottom black NBS treated, Corning 3820) with each 384 well plate containing 320 fragments 32 negative and 32 positive controls. To the fragments were added 23.75 μL of the annealed MYC oligo containing 0.25 μM DNA, 0.5 μM Thiazole Orange, 20 mM Na caco, 140 mM KCl, pH 7 and the plate incubated for 45 min at room temperature. The fluorescent measurements were taken at 25 °C using a Pherastar^{+} platereader (BMG LabTech) with an excitation filter of 510 nm and an emission filter of 540 nm. Dilution, transfer and mixing of all solutions were carried out using a Biomek NX liquid handling robot (Beckman Coulter). On each plate, the controls were used to calculate the Z' score and any plate which failed to gain a Z' greater than

0.5 was rejected and the plate was rescreened.[1] The fragments were ranked according to their TO displacement effect and those fragments showing ≥ 95% displacement were subjected to a dose response, under the original screening conditions. The 50% displacement value (DP_{50}) and subsequent K_d, were calculated using the Prusoff-Cheng equation[3] (Table S3). The top 10 fragments hits were used in molecular modelling and docking studies as well as the cellular c-MYC expression assessed in human HT1080 fibrosarcoma cells.

Data analysis: DP_{50} were calculated by nonlinear regression with sigmoidal dose response curve fitting using Prism version 4.02 (GraphPad Software).

Molecular modeling; System setup and molecular docking.

The 5' truncated version (5'-dT removed) of the full-length 22-mer sequence d(TGAGGG TGGGTAGGGTGGGTAA) in the NMR structure (PDB id 1XAV) of the biologically relevant G-quadruplex element in the human c-MYC promoter was used as a target starting point for in silico modeling. The 3D structures of the fragments were built by means of the ChemBio Office suite (www.cambridgesoft.com), and their conformations were optimized by a short cycle (500 steps) of the MM2 energy minimisation procedure. With the exception of fragment 2F2 (net charge +2) their overall net charges were kept neutral. Suggested initial conformations of those fragments, with functional groups attached to their substituted heterocyclic ring, such as 4H11 and 7A3, were manually adjusted by means of the Discovery Studio Visualizer program (www.accelrys.com). The equatorial position for their functional groups was verified, as it is sterically more plausible than the axial position. The 10 fragments were then docked with the energy-minimized G4-MYC 5' truncated 21-mer using the DOCK v 6.4 program (www.dock.compbio.ucsf.edu/DOCK_6).[4] The entire surface of the G-quadruplexes was defined as a "binding site" (all the spheres generated by the sphgen program of DOCK 6, representing the binding site, to allow all possible binding poses of the small molecule ligands to be examined. The anchor-and-growth strategy for incremental ligand construction, allowing for the ligand's flexibility, was employed. Grid-based (primary) and the Hawkins GBSA (secondary) scoring functions were subsequently used to score the three best ligand orientations, with the highest-scoring binding pose of each fragment being further examined.

Molecular Dynamics simulations.

5 ns Molecular dynamics simulations were performed for the 21-mer alone (as a reference), and for the ten 21-mer/fragment complexes with the best binding poses for the ligands suggested by molecular docking. All full-atom simulations were performed with the GROMACS v 4.5.3 program (www.gromacs.org), employing the parmbsc025 force field previously ported into GROMACS. The topologies and other parameters for the small-molecule fragments were obtained via the ACPYPE tool, employing the ANTECHAMBER module of the AMBER11 program with the GAFF force field.[6-9] All

molecular dynamics protocols were kept identical for consistency of the results. Explicit solvent simulations were performed at T=300 K with a time constant for coupling of 0.1 ps under the control of a velocity rescaling thermostat, and isotropic constant-pressure boundary conditions controlled by the Parinello-Rahman algorithm of pressure coupling.[10-11] Long-range electrostatics were calculated using the PME algorithm[12] with grid spacing of 1.17 Å, and the LINCS algorithm[13] was employed to constrain all bonds. Non-bonded van der Waals interactions were treated with the Lennard-Jones 12-6 potential with a 10.0 Å cut-off. The solute was soaked in a triclinic box of TIP3P water molecules with a minimal clearance of 20.0 Å between periodic images for the starting configurations. Additionally, positively-charged K^+ counter-ions were included in the systems to neutralize the negative net charge on the DNA backbone. In each of the MD runs, there were two temperature-coupling groups; DNA with the structural K^+ ions (and fragment, when present), and water with counter-ions. Subsequently, the systems were subjected to 10,000 steps of potential energy minimisation, followed by 300 ps of molecular dynamics at 200 K while keeping the solutes constrained, and further 100 ps of molecular dynamics during which the systems were slowly heated to 300 K and further equilibrated prior to unconstrained 5 ns production-level molecular dynamics trajectory calculations. The time-step applied was 2.0 fs with coordinates saved every 5.0 ps. The initial 500 ps were then rejected for subsequent MM/PB(GB)SA calculations, that were carried out over 450 frames representing the last 4.5 ns of the 5 ns production runs of the 21-mer fragment complexes.

Molecular Mechanics/Poisson-Boltzmann/Generalized-Born calculations.

The MM/PB(GB)SA method[14,15] computes the relative free energies of binding, employing the thermodynamic cycle that combines molecular mechanics (MM) energies with implicit solvent methods. This method takes advantage of multiple snapshots from a trajectory, to provide an average of energies. The change of free-energy of the molecules upon complex formation was calculated (for each of the snapshots) as a difference of free energy between their bound and unbound states. The corresponding MM/PB(GB)SA protocol as described in Husby et al[16] was employed here. The entropy term ($T\Delta S$) was not included in these simulations. All the calculations were performed employing the single-trajectory approach.

Cells and cell culture.

Human HT1080 fibrosarcoma cells (ATCC# CCL-121) were grown in Dulbecco's Minimum Es-sential Media (supplemented with 10% fetal calf serum) at 37 °C in 5% CO2 in air.

In-cell Western blot assay.

Cells were seeded in 96-well clear bottomed black plates (10,000/well; Corning) and incubated at 37 °C in 5% CO2 in air for 18 h to attach. Cells were treated with fragments

(125 or 250 μM) and incubated at 37 °C in 5% CO2 in air for 24 h. Each well had a final volume of 0.1% DMSO. Media was removed and cells were fixed with neutral buffered formalin (CellPath, Newtown, UK) for 20 min, washed with 0.1% Triton-X (in PBS; 5 x 5 min with gentle shaking) and incubated with Odyssey blocking buffer (Li-Cor, Lincoln, NE, US) for 90 min. Rabbit anti-MYC (1:200 dilution; sc-764, Santa Cruz) and mouse anti ACTB1 (1:200 dilution; ab6276, Abcam, Cambridge, UK) were diluted with Odyssey blocking buffer and incubated at 4 °C overnight. Unbound antibodies were removed with 0.1% Tween (in PBS; 5 x 5 min). Secondary antibodies (IRDye® 800CW goat anti-rabbit and 680LT goat anti-mouse, Li-Cor) were diluted with Odyssey blocking buffer (1:500) and incubated for 1 h at room temperature. Unbound antibodies were removed with 0.1% tween and plates allowed to dry before imaging using a Li-Cor Odyssey system. Background expression was removed (control containing no primary antibody) for each channel. For each well MYC expression was normalized to ACTB1 and then finally all wells were normalized to non-treated control. The assay was repeated in at least triplicate. Statistical significance was calculated by ANOVA (Kruskall-Wallis test; $P<0.0001$) or student's t-test using Graphpad Prism version 4.02.

Further details of the modelling

Structural stability was observed throughout the eleven 5 ns MD runs (comprising ten fragment complexes and the reference native 21-mer structure), the G-quadruplex structures remaining entirely intact for all fragment-bound 21-mers and the reference (21-mer alone) model, with the structural K^+ ions remaining present in the channel. RMSD values for the all-atom 21-mer, and the fragment-bound 21-mers, respectively, as a function of simulation time, were used as a measure of stabilization of the ten complexes. The simulation for the native structure stabilized at ~ 2.4 Å, and ~ 1.4 Å respectively for the time-averaged RMSD plots (Figure S1 a). However, whereas the majority of the fragments remained at their initial binding site (the T14-A15 loop) or in its vicinity throughout the MD runs, fragments 2G5, 11D6 and eventually 9B4 left the binding site completely and 'escaped'. The fragment, 1B5 was found to relocate from their initial binding site on top of the 3rd G4-tetrad formed by G9-G13-G18-G22, with the latter stacking with G18 (Figure 3).

Binding poses of fragments 6H8 and 16C10, which obtained the highest overall score, are shown both upon docking, and at the end of each 5 ns MD runs (Figure 3). The two fragments with the highest overall score are structurally very similar (Figure 1) with oxygen atoms in the di-substituted heterocyclic ring being either in para (6H8) or meta (16C10) positions; these were then found to be involved in specific hydrogen bonds with N2 of G12. Also fragment 14H8 scored very well when the individual scorings of the (1) relative free energy of binding, (2) stability throughout the MD run, (3) and formation of hydrogen bonds were considered, and combined into one value, a overall score (Table S3). Similarly, 2G5 and 11D6 at the other end of the overall score are structurally similar and together with fragment 9B4 they all left their binding site through the course of the MD run. Structurally similar fragments 1H3 and 1B5 scored towards the lower end of the group, however, fragment 7A3 performed well in terms of stability and binding energy.

The difference may suggest an extra methyl group at the para position of the disubstituted heterocyclic ring, which may contribute towards its preferred binding properties in a G-quadruplex groove.

SUPPLEMENTARY FIGURES

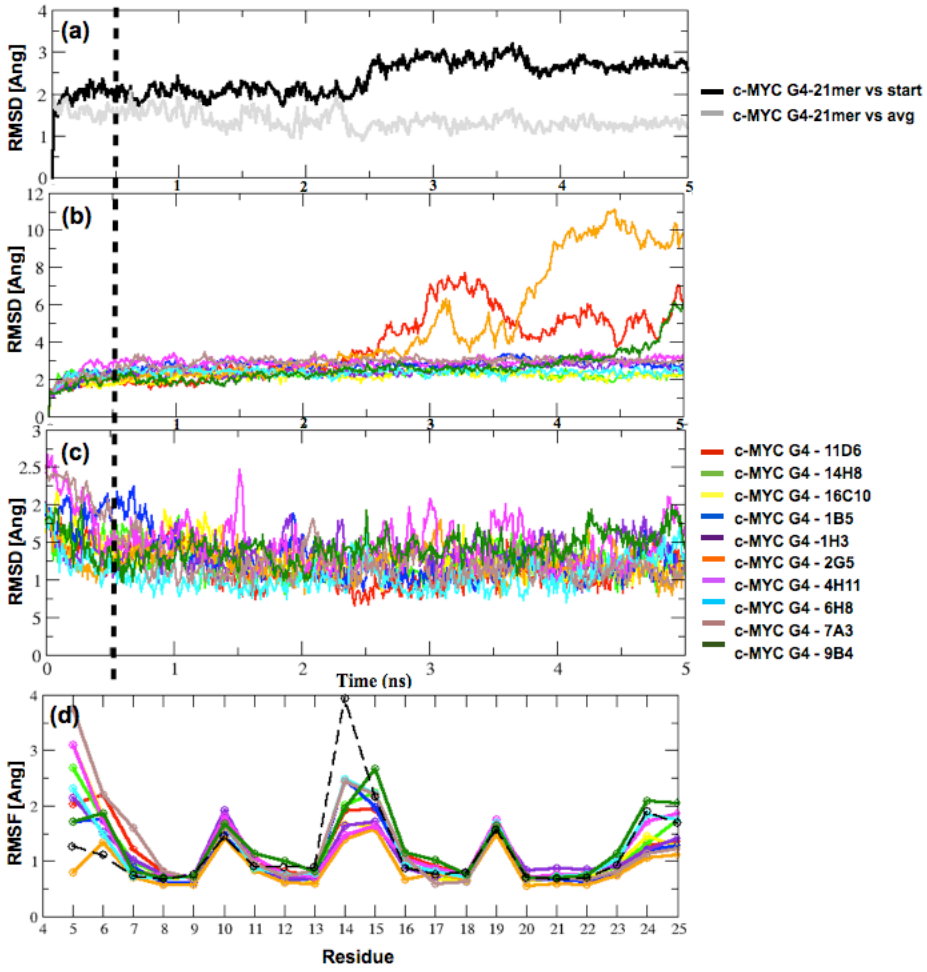

Figure S1. RMSD and RMSF plots showing the stability of the simulated systems during the molecular dynamics simulations of the c-MYC 21-mer/fragment complexes. All-atom RMSDs of the truncated 21-mer are shown (a) with respect to the initial (black) and the time-averaged (grey) structure; (b) all-atom RMSD plots of the G-quadruplex/fragment complexes with respect to their starting structures, and (c) all-atom RMSD of the complex-bound 21-mer with respect to the time-averaged structures over the 5 ns molecular dynamics runs. Fragments 2G5, 11D6 and 9B4 'escape' the binding site. (d) RMSF per residue plots of the complex-bound 21-mers with respect to the initial structure. The native truncated 21-mer data is shown as a black dotted line, to demonstrate the stabilizing effect of the fragments.

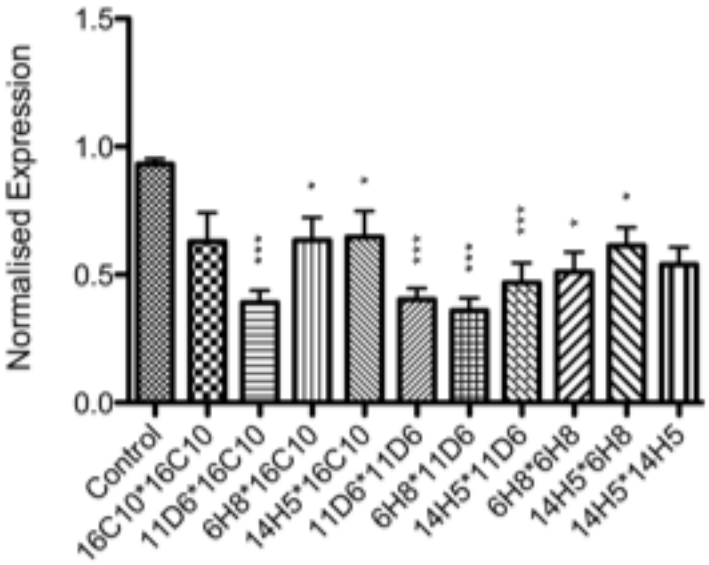

Figure S2. The cellular effects of binary fragments, quantifying c-MYC protein expression using an in-cell Western blot assay. Combined data from three plates is shown, with at least eight data points per bar. Statistical significances were calculated by ANOVA (Kruskall-Wallis test; * P<0.05, ***P<0.0001).

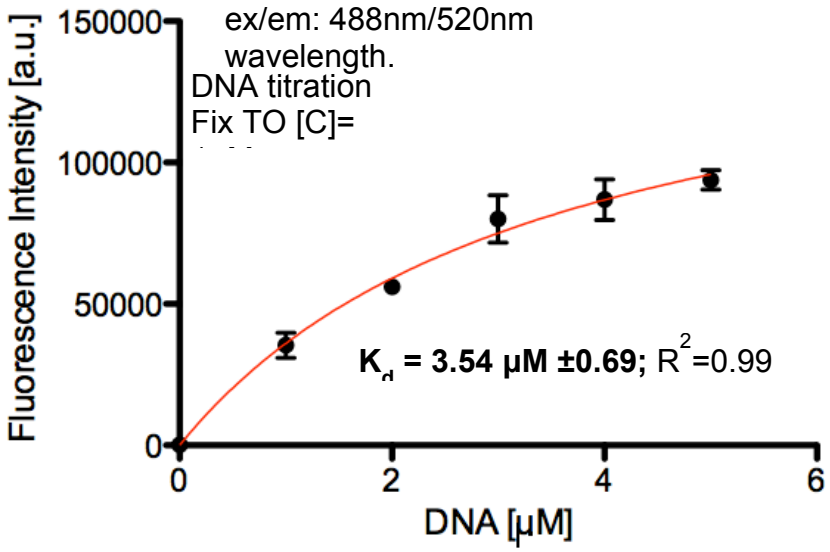

Figure S3. K_d determination of TO for the c-MYC G-quadruplex.

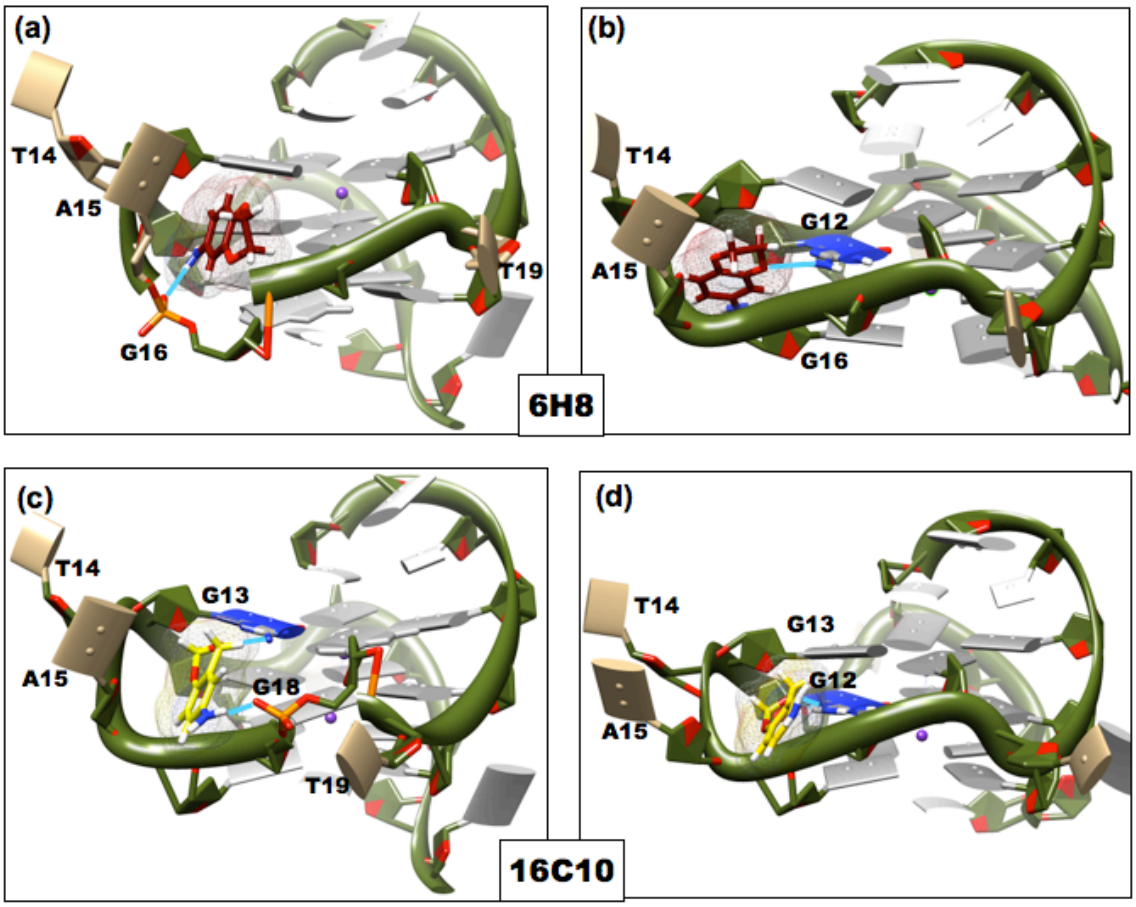

Figure S4. Further consensus predicted fragments shown bound to the c-MYC G4 21-mer. Initial binding poses of the fragments upon docking are shown in panels (a, c) and at the end of the 5 ns MD run in panels (b, d). The fragment is in stick representation, and coloured brown (6H8) and yellow (16C10).

SUPPLEMENTARY TABLES

Hit	UPAC name	Order number	Supplier
7A3	4-(4-Methylpiperazino)aniline 97%	SS00001DA	Maybridge
11D6	1-Methyl-1H-Indol-5-Amine 97+%	CC41414DA	Maybridge
6H8	1,4-Benzodioxan-6-Amine 99%	AC18578-2500	Across Organics
4H11	4-[2-(4-Morpholinyl)ethoxy]aniline 96%	CC42114CB	Maybridge
9B4	2-Methyl-1H-Indol-5-Amine 97%	AC34875	Maybridge
1H3	4-Morpholinoaniline 98+%	197157-5G	Sigma-Aldrich
2G5	1-Benzofuran-5-Amine 97+%	CDS017825-25MG	Sigma-Aldrich
1B5	4-(1-Piperidino)aniline, 97+%	556629-1G	Sigma-Aldrich
14H8	Thieno[2,3-B] Pyrazin-7-Amine 95+%	MO07822,	Acros Organics
16C10	4,5-Dihydro-1,3-Benzodioxine-6-Amine 97+%	SEW03598CB	Maybridge

Table S1. Name, Order number and supplier of the fragments identified in the original TO-displacement assay (IDA).

Cmpd	HB (docked)	HB (trajectory)	Binding site	RMSD [Å]	MM/PB(GB)SA [kcal/mol]		E(bin) AVG
			(MD 5-ns)	G4 21mer	GB	PB	
2G5	G16_OP	---	escapes	2.26	-2.75	-1.47	-2.11± 0.91
9B4	G13_O4'	G12; G13	escapes	2.33	-9.51	-11.37	-10.44 ± 1.32
11D6	G16_OP	---	escapes	2.32	-2.64	-1.98	-2.31 ± 0.47
14H8	G12_N2; G13_O3'; A15_N3	G13; G17	T14-A15 loop	2.35	-11.74	-9.06	-7.15 ± 1.90
16C10	G13_N2; G18_OP	G12; G13	T14-A15 loop	2.42	-11.21	-11.76	-11.49 ± 0.39
6H8	G16_OP	G12	T14-A15 loop	2.07	-11.85	-12.7	-12.28 ± 0.60
1B5	G13_O4'	---	T14-A15 loop → 3'site	2.27	-10.58	-9.71	-10.15 ± 0.62
1H3	G16_OP	G17	T14-A15 loop	2.18	-8.46	-10.89	-9.68 ± 1.72
7A3	G8_OP	G8; G22	groove	2.19	-9.28	-13.87	-11.58 ± 3.25
4H11	G16_OP; G12_N2	G13; A15; G17	T14-A15 loop	2.26	-9.97	-12.53	-11.25 ± 1.81

Table S2. Overview of the results of the c-MYC 21-mer in silico study with ten small-molecule fragments. Hydrogen bonds (HB) found during the initial docking, as well as those observed through the course of the molecular dynamics runs, are listed. The RMSD values for the complex-bound 21-mers are given. Relative intermolecular binding energies were calculated by both GB and PB implicit solvent methods. (The fragments are listed according to their structural similarity, and colour coding corresponds to that in Figure S1).

ORDER	1	2	3	4	5	6	7	8	9	10
ΔG (MM/PB-SA)	7A3	6H8	4H11	16C10	9B4	1H3	1B5	14H8	11D6	2G5
ΔG (MM/GB-SA)	6H8	14H8	16C10	1B5	4H11	9B4	7A3	1H3	2G5	11D6
ΔEbin(AVG)	6H8	7A3*	16C10	4H11	9B4	1B5	1H3	14H8	11D6	2G5
RMSD (21mer)	16C10	2G5*	14H8	6H8	11D6*	9B4*	1B5	4H11	7A3	1H3
Number of H-bonds	14H8	4H11	16C10	7A3	9B4	6H8	1H3	1B5	11D6	2G5
Overall score	**16C10**	**6H8**	**4H11**	**14H8**	**7A3**	**9B4**	**1B5**	**1H3**	**11D6**	**2G5**

Table S3. Fragment molecules scored in several different ways: according to their binding free energies calculated in various ways, stability of the complex-bound G4 21-mers over ten 5 ns molecular dynamics simulations, and the formation of hydrogen bonds between the G4 and individual fragments. The overall score for each fragment was obtained by combining the scores of the relative energies of binding (ΔE_{bin}(AVG), i.e the averaged MM-PBSA and MM-GBSA), the stability plots, and hydrogen bonds formation. (Fragments that 'escaped' their binding site during the molecular dynamics runs are marked * in the RMSD column). For instance fragment 16C10's overall score was obtained as a sum of its (partial) scores: 4+3+3+1+3 = 14, which was the lowest, hence the best overall score obtained (together with fragment 6H8).

SUPPLEMENTARY REFERENCES

1. J. Zhang, T. D. Y. Chung and K. R. Oldenburg, K.R. *J. Biomol. Screen*., 1999, **4**, 67.

2. M. Congreve, R. Carr, C. Murray and H. Jhoti. *Drug Discov. Today*, 2003, **8**, 876.

3. Y.-C. Cheng and W. H. Prusoff. *Biochem. Pharmacol*., 1973, **22**, 3099.

4. Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. RNA, 2009, 15, 1219.

5. Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M. Biophys. J., 2007, 92, 3817.

6. Sousa da Silva, A. W.; Vranken, W. F. BMC Res. Notes, 2012, 5, 367.

7. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calcula-tions. J. Mol. Graph. Model., 2006, 25, 247.

8. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. J. Comput. Chem., 2004, 25, 1157.

9. Case, D.A. et al. AMBER 11, 2011 University of California, San Francisco.

10. Bussi, G.; Donadio, D.; Parrinello, M. J. Chem. Phys., 2007, 126.

11. Parrinello, M.; Rahman, A. J. Appl. Phys., 1981, 52, 7182.

12. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., J. Chem. Phys., 1995, 103, 8577.

13. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: J. Comp. Chem., 1997, 18, 1463.

14. Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. J. Am. Chem. Soc., 1998, 120, 9401.

15. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. J. Phys. Chem. A, 1997, 101, 3005.

16. Husby, J.; Todd, A. K.; Platts, J. A.; Neidle, S. Biopolymers, 2013, 99, 989.