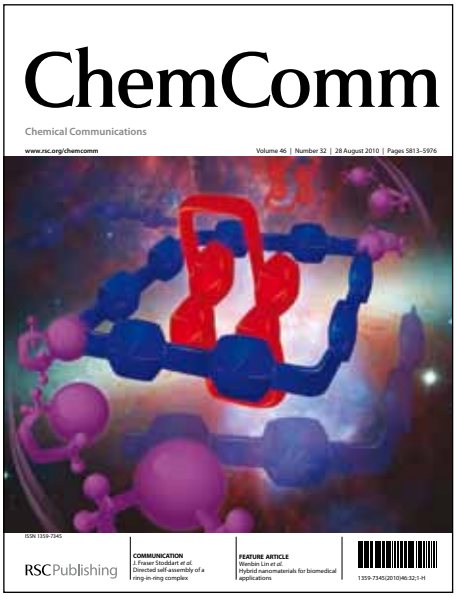
ChemComm

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the RSC Publishing peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, which is prior to technical editing, formatting and proof reading. This free service from RSC Publishing allows authors to make their results available to the community, in citable form, before publication of the edited article. This *Accepted Manuscript* will be replaced by the edited and formatted *Advance Article* as soon as this is available.

To cite this manuscript please use its permanent Digital Object Identifier (DOI®), which is identical for all formats of publication.

More information about *Accepted Manuscripts* can be found in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics contained in the manuscript submitted by the author(s) which may alter content, and that the standard **Terms & Conditions** and the **ethical guidelines** that apply to the journal are still applicable. In no event shall the RSC be held responsible for any errors or omissions in these *Accepted Manuscript* manuscripts or any consequences arising from the use of any information contained in them.

Journal Name

RSCPublishing

COMMUNICATION

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

Cite this: DOI: 10.1039/xoxxoooooox

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

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/xoxxoooooox

www.rsc.org/

We report here on the screening of a fragment library against a G-quadruplex element in the human c-MYC promoter. The ten fragment hits had significant concordance between a biophysical assay, in silico modelling and c-MYC expression inhibition, highlighting the feasibility of applying a fragment-based approach to the targeting of a quadruplex nucleic acid.

Guanine-rich nucleic acid sequences can self-associate via Hoogsteen hydrogen bonds to form planar G-quartets, which π-π stack to form four-stranded G-quadruplexes (G4s) in the presence of stabilizing alkali metal cations.[1,2] G4 motifs occur in tandem repeat regions of telomeres[1,3,4] and in other G-rich regions of the genome[5] including gene promoters[6], untranslated regions[7] and gene bodies.[8] They are prevalent in the promoters of a number of functionally important oncogenes such as MYC,[9] KRAS,[10] and KIT.[11,12] Small molecules that bind and stabilize G4s can attenuate the transcriptional activity of such proto-oncogenes in human cancer cells[13-16] and are thus novel potential therapeutic agents. A well-established example is quarfloxin, the first G4 compound to enter clinical trials in human cancer.[17]

A number of novel drug-like small molecules that target biologically relevant G4s have been discovered through a combination of rational design and molecular modeling based on knowledge of the secondary and tertiary structure of the target G4. An alternative approach is to assemble smaller fragments around a target, as exemplified by the identification of a potent telomere-targeting small molecule by using in-situ click chemistry.[18] Fragment-based screening methods have been used with considerable success to discover high-affinity ligands for protein active sites, as well as targeting protein-protein-interfaces.[19] The application of this approach to nucleic acid targeting has been largely unexplored for DNA, although a recent study has demonstrated its feasibility with RNA riboswitches.[20] In principle fragment-based methods can lead to the design of compounds with enhanced drug-like features compared to most current G4-binding ligands.

We present here an initial study using fragment-based targeting of a DNA G4 structure derived from the c-MYC promoter. At the outset it was not clear whether low-molecular weight fragments could bind to G4 nucleic acids with sufficient affinity and selectivity for high-throughput detection, and thus whether they could produce effects on promoter function.

An Intercaltor-displacement assay (IDA) employing thiazole orange (TO) was used to screen for fragment interactions with the c-MYC G4. TO is a well-validated probe for screening G4-binding ligands.[21,22] It is highly fluorescent when bound and quenched after displacement (λ_{Ex}= 501 nm, λ_{Em}= 539 nm). The screening library, which has previously been used for targeting RNA riboswitches,[20] is comprised of structurally and chemically diverse fragments. The top 10 confirmed and profiled hits selected from the initial IDA screen were used in molecular modelling and docking studies with the NMR structure for the c-MYC G4 (PBD entry 1XAV). Their effects on the down-regulation of cellular c-MYC expression were assessed in human HT1080 fibrosarcoma cells.

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).[23] All fragment molecules were ≥ 95% purity and had >1 mM aqueous solubility. They were obtained from commercial sources, or were synthesized in-house. For screening, a 384-well assay plate format was used; each plate contained 320 fragments together with 32 negative and 32 positive controls.

The top 10 fragments from the initial IDA screen can be subdivided into three groups based on their structural similarity: 2G5, 9B4, 11D6 and 14H8 have fused 5- and 6-membered

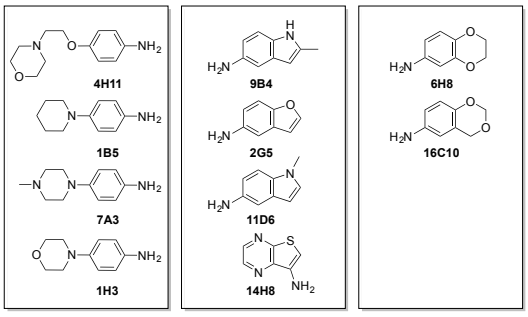


Figure 1. The top 10 fragment hits obtained from the thiazole orange (TO) intercalator displacement assay (IDA).

Compound	2G5	9B4	11D6	14H8	16C10	6H8	1B5	1H3	7A3	4H11
MW	133	146	146	151	151	151	176	178	191	222
IDA[(a)]	2998.6 (10)	133.0 (6)	21.8 (1)	129.5 (5)	326.3 (7)	1128.7 (9)	330.7 (8)	75.2 (3)	59.5 (2)	125.1 (4)
In Silico[(b)]	29;18 (10)	15;15 (6)	29;20 (9)	17:7 (5)	10;2 (2)	5;9 (1)	19;19 (8)	20;18 (7)	12;15 (3)	12;15 (4)
Cellular[(c)]	37 (5)	42 (6)	21 (1)	32 (4)	31 (2)	32 (3)	51 (8)	42 (7)	84 (10)	52 (9)

Table 1. Fragment molecules scored according to their binding free energies (kcal/mole) and stability plots, over the ten 5 ns MD simulations. Rank positions are in parentheses. Free energy values are in the Supplementary Data. (a) TO Displacement Assay (IDA)-hit profiling data with the corresponding inhibition constants (Ki_{50}). (b) *In silico* data quoted as overall best score. (c) Cellular data quoted as % c-MYC expression compared to control (100% corresponds to the most effective dose).

heterocyclic rings while 16C10 and 6H8 have two fused 6-membered heterocyclic rings. 1B5, 1H3, 7A3 and 4H11 are all 4-substituted aniline derivatives. Interestingly all fragment derivatives contain an amino functionality, which may aid in the electrostatic binding of the fragments to G4 DNA.

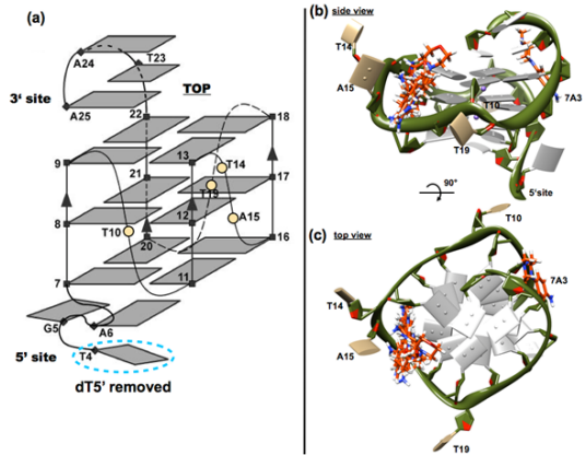


Figure 2. Schematic views of (a) the c-MYC 21- and 22-mer G4s. Strand directionality is indicated by arrows and the loop nucleotides are shown as yellow circles. (b,c) Two views of the 21-mer with the binding poses of the fragments found by DOCK, in ribbon representation (green) with guanine bases coloured grey, and loop nucleotides shown in yellow. The fragments, in stick representation, are coloured red.

The 5' truncated version (5'-dT removed) 21-mer G4 element d(TGAGGGTGGGTAGGGTGGGTAA) in the sequence of the human c-MYC promoter, was used for molecular docking studies with the top ten fragment hits (Figure 2). 5 ns molecular dynamics (MD) simulations were performed for the 21-mer alone (as a reference) and for the ten 21-mer/fragment complexes with the best binding poses of the fragments suggested by docking. The overall score for each fragment was assessed by scoring each according to its averaged predicted binding energy (MM/PB(GB)SA) with stability indicated from 1 (best) to 10 (worst). H bonds formed between the G4-fragment complexes were also included within the overall score (Table 1 and Supplementary Information).

Structural stability was observed throughout the eleven 5 ns molecular dynamics runs (comprising ten fragment complexes and the reference native 21-mer structure). The G4 structures remained entirely intact for all fragment-bound 21-mers and the reference (21-mer alone) model, with the structural K^+ ions. The small loop fluctuations reflect the stable nature of the two single (T10 and T19) and double-nucleotide (T14-A15) loops. The 3'-flanking T23-A24-A25 sequence is stacked over the top of the 3rd G-quartet in all sixteen models, showing some structural flexibility, suggesting that the 3´-end maintains a stable conformation over the course of the 5 ns simulations.

The majority of the ligands remained at their initial binding site (the T14-A15 loop) or in its vicinity throughout the MD runs. The poorly-scoring fragments 2G5, 11D6 and 9B4 left the binding site completely during the MD. Fragment 1B5 was found to relocate from its initial binding site on top of the 3rd G-quartet formed by G9-G13-G18-G22, with the latter stacking with G18. All ten fragment molecules stabilized the T14-A14 loop region where they were initially docked, compared to the native 21-mer. The binding of fragments 6H8 and 4H11 resulted in significantly reduced loop flexibility. In contrast, the 5´-flanking G5-A6 sequence show increased flexibility in the majority of the 21-mer fragment complexes (except with 14H8) compared to the native 21-mer.

Overall, fragments 6H8 and 16C10 performed the best with consistent predicted binding energy and complex stability throughout the MD trajectories. These gave improved G4 stabilization, H bond formation and strongly favourable binding energies. Fragments 14H8, 7A3 and 4H11 also scored towards the high end of the group. At the other side of the scale, 2G5 and 11D6 did not perform well since their intermolecular binding interactions were significantly less favourable than any of the other fragments

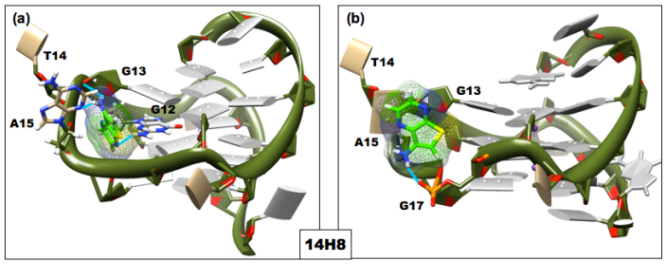


Figure 3. The consensus best predicted fragment 14H8 shown bound to the c-MYC G4 21-mer. The initial binding pose of the fragment upon docking is shown in panel (a) and at the end of the 5 ns MD run in panel (b). The fragment is in stick representation, and is coloured green. H-bonds formed between the fragment and sites on the 21-mer are coloured cyan.

Binding poses of the best-ranked fragments 6H8, 16C10 and 14H8 are shown in Figure 3. All fragments, with the exception of 7A3, are docked within the T14-A15 loop region. The two best-ranked fragments are structurally very similar (Figure 1) with oxygen atoms in the dioxan rings being either in *para* (6H8) or *meta* (16C10) positions; these are involved in specific H bonds with N2 of G12. Also fragment 14H11 scored very well when the individual scorings of (1) relative free energy of binding, (2) stability throughout the molecular dynamics runs, and (3) formation of H bonds, were considered, and combined into a single total score (Table S3). Similarly, 2G5 and 11D6 at the other end of the ranking list are structurally similar and together with fragment 9B4 they all left their binding site through the course of the simulations. Structurally similar fragments 1H3 and 1B5 scored towards the lower end of the group, however, fragment 7A3 performed better in terms of stability and binding energy. This suggests that an extra methyl group at the

para position of the dioxin ring is beneficial, contributing towards its preferred binding properties in a G4 groove.

For the cellular evaluation the top ten IDA hit fragments were investigated using a 96-well in-cell Western blot immunoassay to allow concurrent screening of all fragments at 125 μM and 250 μM. These doses were chosen after preliminary cell-growth inhibition studies to ensure conditions did not lead to short or long-term cytotoxicity (data not shown). Human HT1080 cells were treated for 24 h with fragments or vehicle-only control. For each well the measured MYC protein level was normalised against an actin beta1 control and background (no primary antibody) and non-treated control (Figure 4). All fragments, except 7A3 (both doses) and 2G5 (at 125 μM), showed significant changes in MYC protein expression relative to control. Four of the fragments produced a robust reduction in MYC levels in all repeat experiments, of which 11D6 was the best inhibitor (Table 1). These four fragments were used to carry out further inhibitory studies in pair-wise combinations (Supplementary Information) that reveal every combination tested (125 μM of each component) induced a significant reduction in c-MYC protein compared to control (ANOVA). The best binary mixtures were 6H8*11D6 and 11D6*16C10, which were both slightly more effective than 11D6 as a single treatment, although not statistically significant (c.f. 11D6; students t-test, 2-tail).

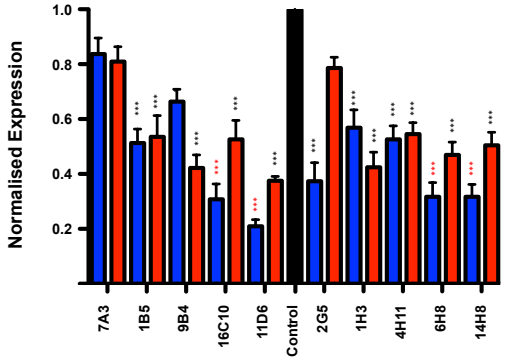


Figure 4. Cellular effect of hit fragments showing quantification of MYC protein expression using an in-cell Western assay 250 (blue) and 125 (red) μM fragment concentration. These levels are necessary to compensate for their relative low affinity. Combined data from five plates is shown, with at least 18 data points per bar. Statistical significance was calculated by ANOVA (Kruskall-Wallis test; * P<0.05, *** P<0.0001).

Three out of the four top fragments 16C10, 6H8 and 14H8 (ranked 2-4) in the expression inhibition screen are also in the top four from the *in silico* screen (Table 1), although these three are not in the top half on the basis of the IDA screen alone. On the other hand the IDA and the expression screens concur in indicating that fragment 11D6 is the top-scoring one; this is low in the *in silico* ranking. In terms of consensus fragment scoring, 14H8 and 9B4 are moderately well indicated by all three methods, suggesting that an amino-thieno-pyrazine or amino-indole motif may be a good starting-point for fragment expansion (fragment growing and linking) to more potent compounds. The amino-benzodioxan skeleton found to be favoured in both the *in silico* and expression screens would also be appropriate.

Conclusions

This study demonstrates the feasibility of applying a fragment-based approach to the targeting of a G4 nucleic acid. It has led to fragment molecules that exhibit inhibition of c-MYC protein production. These may well provide suitable future starting-points for the design of drug-like molecules with high selectivity for the c-MYC quadruplex target. There is a degree of correspondence between the predicted and experimental ranking order of the fragments, although the best predicted fragment is not consistently the best experimentally. These differences may reflect (i) limitations of the initial docking which maintains the c-MYC quadruplex in a rigid conformation, as well as limitations in the energy calculation methodology, (ii) potential differences in binding site preferences between the TO-displaced IDA site(s) and that indicated by the docking and simulations. The possibility that the fragments cause c-MYC expression changes by non-promoter G4 pathways cannot be discounted at this stage[24], although the good correspondence between the expression and *in silico* data suggest otherwise. We suggest that the use of consensus scores from experimental and *in silico* studies may minimise the first two factors.

It is notable that the presence of two single-nucleotide loops and a two-nucleotide loop in the c-MYC G4 ensures conformational stability during the docking and subsequent MD simulations. The human telomeric G4 with three three-nucleotide loops would have significantly increased flexibility[25]. Thus *in silico* screening of G4s with fragment and other small-molecule libraries, is most likely to be useful when the G4s have some short loops, as in the case of the c-MYC and many other promoter G4s.[5,6]

Notes and references

[a] Department of Chemistry, The University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK. sb10031@cam.ac.uk [b] Cancer Research UK Cambridge Institute, Li Ka Shing Centre, Cambridge, CB2 0RE, UK. [c] UCL School of Pharmacy, 29-39 Brunswick Square, London WC1N 1AX, UK.

S. B and S. N. are grateful to CRUK for programme grant support to their laboratories.

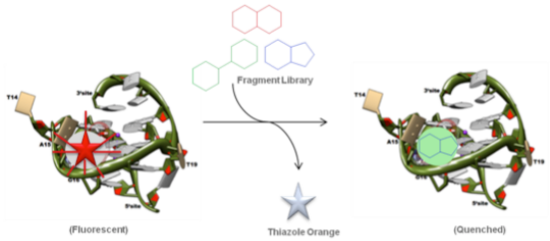
Electronic Supplementary Information (ESI) available: See DOI: 10.1039/c000000x/

References

1 J. R. Williamson, M. K. Raghuraman and T. R. Cech. *Cell*, 1989, **59**, 871.
2 D. Sen and W. A. Gilbert. *Nature*, 1990, **344**, 410.
3 E. Henderson, C. C. Hardin, S. K. Walk, I. Tinoco, Jr. and E. H. Blackburn. *Cell*, 1987, **51**, 899.
4 Y. Wang and D. J. Patel. *J. Mol. Biol*., 1993, **234**, 1171.
5 J. Huppert and S. Balasubramanian, *Nucleic Acids Res.* 2005, **33**, 2908; A. Todd, M. Johnston and S. Neidle. *Nucleic Acids Res.* 2005, **33**, 2901.
6 J. L. Huppert and S. Balasubramanian, *Nucleic Acids Res.* 2007, **35**, 406.
7 J. L. Huppert, A. Bugaut, S. Kumari and S. Balasubramanian. *Nucleic Acids Res*., 2008, **36**, 6260
8 R. Rodriguez, K. M. Miller, J. V. Forment, C. R. Bradshaw, M. Nikan, S. Britton, T. Oelschlaegel, B. Xhemalce, S. Balasubramanian and S. P. Jackson, *Nat. Chem. Biol*., 2012, 8, 301.
9 A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L. H. Hurley, *Proc. Natl. Acad. Sci. USA* 2002**, 99**, 11593
10 S. Cogoi, and L. E. Xodo. *Nucleic Acids Res*., 2006, **34**, 2536.
11 S. Rankin, A. P. Reszka, J. Huppert, M. Zloh, G. N. Parkinson, A. K. Todd, S. Ladame, S. Balasubramanian and S. Neidle. *J. Amer. Chem. Soc*., 2005, **127**, 10584; H. Fernando, A. P. Reszka, J. Huppert, S. Ladame, S. Rankin, A. R. Venkitaraman, S. Neidle and S. Balasubramanian. *Biochemistry*, 2006, **45**, 7854.
12 L. H. Hurley, D. D. Von Hoff, A. Siddiqui-Jain and D. Yang. *Semin. Oncol*., 2006, **33**, 498.
13 J. N. Liu, R. Deng, J. F Guo, J. M. Zhou, G. K. Feng, Z. S. Huang, L. Q. Gu, Y. X. Zeng and X. F. Zhu. *Leukemia*, 2007, **21**, 1300.
14 K. I. McLuckie, Z. A. Waller, D. A. Sanders, D. Alves, R. Rodriguez, J. Dash, G. J. McKenzie, A. R. Venkitaraman and S. Balasubramanian. *J. Amer. Chem. Soc*., 2011, **133**, 2658.
15 Z. A. Waller, S. A. Sewitz, S. T. Hsu and S. Balasubramanian. *J. Amer. Chem. Soc*., 2009, **131**, 12628.
16 S. Balasubramanian and S. Neidle, *Curr. Opin. Chem. Biol*., 2009, **13**, 345.
17 D. Drygin, A. Siddiqui-Jain, S. O'Brien, M. Schwaebe, A. Lin, J. Bliesath, C. B. Ho, C. Proffitt, K. Trent, J. P. Whitten, J. K. Lim, D. Von Hoff D, K. Anderes and W. G. Rice. *Cancer Res*., 2009, **69**, 7653.
18 M. Di Antonio, G. Biffi, A. Mariani, E. A. Raiber, R. Rodriguez and S. Balasubramanian. *Angew. Chem. Int. Ed. Engl*., 2012, **51**, 11073.
19 A. Ciulli, T. L. Blundell and C. Abell. *Drug Future*, 2007, **32**, 13.
20 L. Chen, E. Cressina, F. J. Leeper, A. G. Smith and C. Abell. *ACS Chem. Biol*., 2010, **5**, 355.
21 D. L. Boger and W. C. Tse. *Bioorg. Med. Chem*., 2001, **9**, 2511.
22 D. Monchaud, C. Allain and M.-P. Teulade-Fichou. *Bioorg. Med. Chem. Lett*., 2006, 16, 4842-5.
23 M. Congreve, R. Carr, C. Murray and H. A. Jhoti. *Drug Discov. Today*, 2003, **8**, 876.
24 P. V. Boddupally, S. Hahn, C. Beman, B. De, T. A. Brooks, V. Gokhale and L. H. Hurley. *J. Med. Chem*., 2012, **55**, 6076.
25 J. Husby, A. K. Todd, J. A. Platts and S. Neidle. *Biopolymers*, 2013, **99**, 989.

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



We report here on the screening of a fragment library against a G-quadruplex element in the human c-MYC promoter. The ten fragment hits had significant concordance between a biophysical assay, *in silico* modelling and c-MYC expression inhibition, highlighting the feasibility of applying a fragment-based approach to the targeting of a quadruplex nucleic acid.