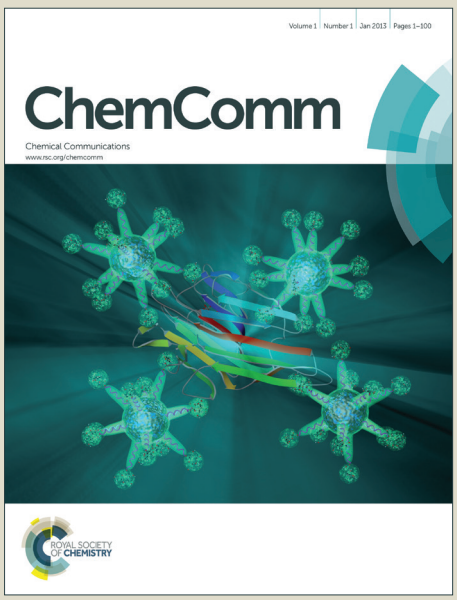
# ChemComm

Accepted Manuscript





ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

RSCPublishing

COMMUNICATION

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

Cite this: DOI: 10.1039/x0xx00000x

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

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

**To efficiently identify small molecules binding to G-quadruplex structure while avoiding binding to duplex DNA, we performed a multistep structure-based virtual screening by simultaneously taking into account G-quadruplex DNA and duplex DNA. Among the 13 compounds selected, one outstanding ligand shows significant selectivity for G-quadruplex binding as determined by SPR, FRET-based competition and luciferase activity assay.**

In addition to the Watson–Crick duplex, DNA and RNA are able to form higher order structures called G-quadruplexes from guanine-rich sequences. It has been shown that G-quadruplex formation in regions of biological significance, such as human telomeres, promoter regions of oncogene and ribosomal DNA, plays important roles in cellular aging and cancer.[1] Recently, an elegant study reported that DNA G-quadruplex structures can be visualized quantitatively in human cells, which confirms the existence of G-quadruplex structures in the human genome and reinforces the idea that DNA G-quadruplexes might be a promising class of molecular targets for anticancer drugs.[2]

Thus far, a number of small molecules targeting biologically relevant G-quadruplexes have been reported.[3] Their anticancer activities and mechanism of actions were carefully studied to direct medicinal chemistry research targeting G-quadruplexes.[4] However, there is only one small molecule (quarfloxin) that has entered into clinical trials for human cancer and none has made it to market. The reasons are complex, including inherent difficulties in drug discovery process, difficulties in translation from *in vitro* data to *in vivo* data and the side-toxicity of the small molecules because of their poor selectivity, or nonspecific interactions with the duplex DNA. Typically, fused aromatic G-quadruplex ligands such as BRACO-19,[5] and SYUIQ-5 bind to duplex DNA via intercalation,[6] while some unfused aromatic G-quadruplex ligands, such as distamycin A, bind to the groove of duplex DNA.[7] Importantly, duplex DNAs are more abundant than G-quadruplex DNAs in cells. Therefore, the selectivity for G-quadruplex over duplex DNA is crucial for small molecules to be promising G-quadruplex binders. A number of studies have been devoted to the development of novel G-quadruplex binders or discovery of new chemical platforms for G-quadruplex through virtual screening protocols.[8] However, to the best of our knowledge, none of them includes the aspect of selectivity screening for G-quadruplex over duplex DNA. Herein, we report on the development of a sequential computational approach to enable the identification of prospective ligands that bind to G-quadruplex selectively without duplex binding. The multistep strategy combines docking-based virtual screening for G-quadruplex binders, duplex DNA intercalators and duplex DNA minor groove binders.

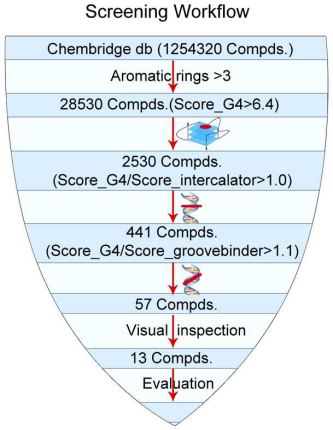



**Fig. 1** Workflow diagram of the virtual screening cascade protocol.

The *c-MYC* gene promoter G-quadruplex structure was selected as the target in view of its important role linked to human cancers, with the additional advantage that it adopts only one conformation, a propeller-type parallel-stranded G-quadruplex.[9] A database containing 1.2 million compounds from ChemBridge was used for screening. All the small molecules are drug-like and purchasable. Firstly, only small molecules with more than 3 aromatic rings were selected, as suggested by previously identified G-quadruplex ligands being mostly comprised of more than 3 aromatic rings, which may help the ligands stack over the G-quartet (Fig. 1). Secondly, the remaining 28,530 small molecules were docked to the NMR structure of *c-MYC* G-quadruplex (PBD entry: 1XAV)[10] using Surflex-DOCK (Fig. S1, ESI†). Compared to the crystal structure of

*c-MYC* G-quadruplex, this NMR structure may be more physiologically relevant and has flanking sequences that interact with the G-quartets, thus contributing to the conformational complexity of the structure, which might participate in ligand binding. The docking score was calculated and referred as Score_G4. Compounds were selected if satisfying the criteria of Score_G4 > 6.4. Thirdly, the resulting 2,530 compounds were docked to duplex DNA structure (PDB entry 1Z3F)[11] via intercalation interaction. The score was calculated and referred as Score_intercalator. Compounds were selected if satisfying the criteria of Score_G4/Score_intercalator > 1.0. Fourthly, the resulting 441 compounds were docked to duplex DNA structure (PDB entry 1K2Z)[12] via groove binding interaction. The score was calculated and referred as Score_groovebinder. The compounds were selected if satisfying the criteria of Score_G4/Score_groovebinder > 1.1. The resulting 57 compounds were subjected to visual inspection, which has been found to be one of the most critical steps in virtual screening, and undesirable compounds were discarded. The final step yielded 13 candidates that proceeded to experimental validation (Table 1, Table S1, ESI†).

Table 1 Hit list table from virtual ligand screening.

| Compd. ID | Score_G4 | Score_intercalator | Score_groovebinder | A | B |
|---|---|---|---|---|---|
| **VS1** | 9.72 | 7.29 | 5.99 | 1.33 | 1.62 |
| **VS2** | 9.35 | 8.51 | 4.90 | 1.10 | 1.91 |
| **VS3** | 9.21 | 8.11 | 6.20 | 1.14 | 1.49 |
| **VS4** | 8.07 | 7.85 | 5.44 | 1.03 | 1.48 |
| **VS5** | 8.99 | 7.62 | 4.59 | 1.18 | 1.96 |
| **VS6** | 8.75 | 5.71 | 3.66 | 1.53 | 2.39 |
| **VS7** | 7.86 | 6.37 | 5.17 | 1.23 | 1.52 |
| **VS8** | 9.71 | 7.55 | 5.82 | 1.29 | 1.67 |
| **VS9** | 10.65 | 7.33 | 6.20 | 1.45 | 1.72 |
| **VS10** | 9.09 | 7.94 | 7.97 | 1.14 | 1.14 |
| **VS11** | 8.28 | 5.36 | 4.36 | 1.54 | 1.90 |
| **VS12** | 6.60 | 5.67 | 3.83 | 1.16 | 1.72 |
| **VS13** | 6.64 | 4.86 | 4.45 | 1.37 | 1.49 |

A= Score_G4/Score_intercalator; B= Score_G4/Score_groovebinder.

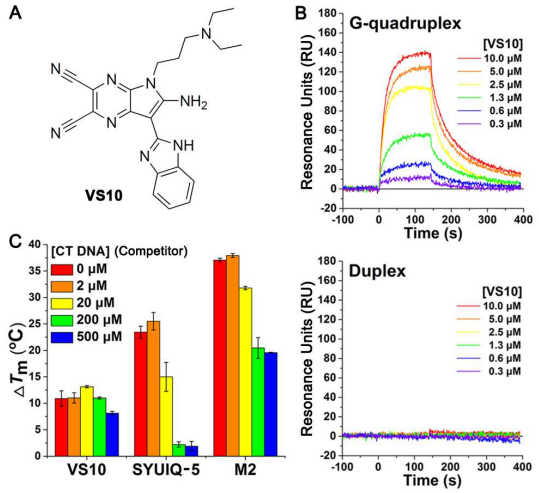



**Fig. 2** (A) Structure of **VS10**. (B) SPR sensorgrams overlay for binding of **VS10** to G-quadruplex and duplex DNA. (C) Competitive FRET results for **VS10** without and with excess of duplex DNA competitor (CT DNA). G-quadruplex ligand SYUIQ-5 and M2 were used as reference compound.

SPR assay was first performed to rapidly screen candidates by evaluating the G-quadruplex binding affinity as well as selectivity over duplex DNA. Biotinylated *c-MYC* G-quadruplex and duplex DNA were used. Five compounds were identified to bind to the G-quadruplex DNA with no binding seen to the duplex DNA (Fig. S2, ESI†). The compound **VS10** shows strong binding affinity with *c-MYC* G-quadruplex DNA with a $K_D$ value of 2.0 μM, while the compound has no duplex binding with a concentration of up to 10.0 μM, as indicated by Fig. 2B. $K_D$ values of the other compounds were not determined as they did not bind to the G-quadruplex strongly enough. To verify the effective binding of **VS10** towards *c-MYC* G-quadruplex DNA, UV titration experiment (Fig. S3, ESI†) and G-quadruplex fluorescent intercalator displacement (G4-FID) assay (Fig. S4, ESI†) were carried out, and the $K_a$ value of $6.6\times10^5$ $M^{-1}$ and $DC_{50}$ value of 1.8 μM were determined respectively. The result from the G4-FID assay also showed that binding of **VS10** towards *c-MYC* G-quadruplex DNA was comparable to that of reference compound SYUIQ-5 and bisquinolinium G-quadruplex ligand M2.[13]

In addition, the stabilization of **VS10** to the G-quadruplex was studied by CD spectroscopy through measuring the thermal stability profile of the *c-MYC* G-quadruplex DNA incubated with the compound. CD studies showed that **VS10** increased the melting temperature of *c-MYC* G-quadruplex by 5°C (Fig. S5, ESI†). To further access the G-quadruplex binding selectivity of **VS10** under competitive conditions, FRET-based competition assays were performed (Fig. 2C), where the ability of ligand to retain G-quadruplex stabilizing ability was challenged by nonfluorescent duplex DNA (CT DNA). In the presence of various amounts of duplex DNA, the thermal stabilization of *c-MYC* G-quadruplex DNA enhanced by **VS10** was only slightly affected until the addition of 500 μM competitor, while 20 μM competitor sharply disrupted the binding of SYUIQ-5 to the G-quadruplex. Besides, the selectivity of **VS10** for G-quadruplex was also found to be better than that of bisquinolinium G-quadruplex ligand M2 whose binding to the G-quadruplex was obviously disrupted upon the addition of 200 μM competitor. These results were in agreement with those found by the SPR assay, suggesting the multistep virtual screening protocol is effective in identifying selective G-quadruplex ligands.

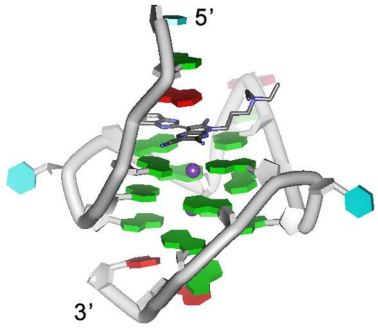



**Fig. 3** Modeling structure of **VS10** and *c-MYC* G-quadruplex complex from MD simulation.

Molecular docking and molecular dynamics simulations were then carried out to investigate the binding mode of **VS10** to *c-MYC* G-quadruplex DNA. The NMR determined propeller-type parallel-stranded G-quadruplex structure used in above screening process was also employed as the template. Given the possibility that ligand molecules may utilize one or the other side of its aromatic ring system to make stacking interactions with the G-quartet, we selected two docking poses representing ligand stacking over the 5′ end of the G-quartet and another two docking poses representing ligand stacking over the 3′ end of the G-quartet. A loop binding mode was

also observed, with the ligand binding to the G-quadruplex propeller loop. In total, five docking poses representing all possible binding modes were subjected to molecular dynamics simulation (Fig. S6, ESI†). The binding energy estimated by MM/PBSA approach shows that ligand stacking over the 5′ end of the G-quartet has the most negative binding free energy (-14.1 kcal/mol, Table S2, ESI†), suggesting that **VS10** might prefer to bind to the 5′ end of the G-quartet (Fig. 3).

To clarify the binding mode between **VS10** and *c-MYC* G-quadruplex structure, a NMR spectroscopy study was carried out using G-quadruplex Pu24I derived from *c-MYC* (PDB entry 2A5P) instead of 1XAV structure, as Pu24I has clearly assigned and more easily distinguished G-quartet imino proton signals (Fig. S7, ESI†).[14] As shown in Fig. 4, upon addition of **VS10** to a solution of Pu24I, the imino proton resonances of the 5' terminal G-quartet residues (G4, G8 and G13) shift upfield and G17 shifts downfield. While no shift of the imino proton resonances of the 3' terminal G-quartet residues (G6, G15, G19 and G24) was observed. The data suggests that **VS10** interacts with G-quadruplex structure by stacking on the 5' terminal G-quartet. This result is consistent with the molecular modeling study. Particularly, the folding of Pu24I is slightly different in the loop region as compared to that of 1XAV. However, such a difference is not expected to significantly affect **VS10** binding to the G-quartet.

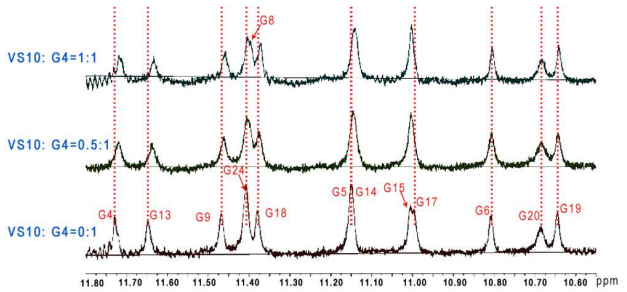



**Fig. 4** NMR titration of *c-MYC* G-quadruplex Pu24I with **VS10** at various ratios of [**VS10**]/[Pu24I]. The imino proton resonances of the residues in the G-quartet were assigned based on the data from literature.[14]

In addition to the above studies, it is also important to evaluate the cellular effects of **VS10** and see whether it could also bind to the *c-MYC* G-quadruplex DNA in cellular conditions and accordingly reduce the gene transcriptional and expression level. Thus, a luciferase activity assay was employed to further explore the cellular effect of **VS10** on *c-MYC* promoter activity. Two luciferase constructs were designed and used. One contained a full-length wild-type promoter of *c-MYC* with a native G-quadruplex forming sequence, while the other had a mutated sequence that suppresses the G-quadruplex formation as suggested by CD studies (Fig. S8, ESI†). The effects of **VS10** on *c-MYC* promoter activity are shown in Fig. 5A, and a compound SYUIQ-5 with low G-quadruplex selectivity over duplex DNA was used as reference. Dose-dependent decreased luciferase activities on the wild promoter construct were observed for both compounds (Fig. S9, ESI†). The addition of **VS10** at 2.0 μM resulted in 33% reduction of luciferase activity for wild promoter construct, which is 2.3 times stronger than its reduction of mutant promoter construct (14%). While at the same concentration, SYUIQ-5 exhibited close inhibitory effects on luciferase activity of wild and mutant constructs (51% vs 52%). The results suggest that **VS10** could inhibit the activity of *c-MYC* promoter through interacting with the promoter G-quadruplex structure and **VS10** has better selectivity than SYUIQ-5 for wild promoter construct over the mutant construct.

On the basis of results from the luciferase activity assay, a study on the effects of **VS10** on *c-MYC* transcription was carried out using two Burkitt's lymphoma cell lines (RAji and CA46). The NHE III1 element of *c-MYC* gene containing G-quadruplex forming sequence is removed together with P1 and P2 promoter in the CA46 cell line, while the RAji cell line still retains this element after translocation.[15] Regulation of the transcription was evaluated by quantitation of mRNA using RT-PCR. As shown in Fig. 5B, upon the addition of **VS10**, the amount of *c-MYC* PCR product did not change much in CA46 cell line. However, in the RAji cell line with the NHE III1 element deleted, the transcription of *c-MYC* was inhibited in a dose-dependent fashion. Upon the treatment of **VS10** at different concentrations for 24 h, transcription of *c-MYC* was reduced by 6%, 18%, 26% and 34%, respectively, related to a control gene β-actin. These results provide additional evidence suggesting that **VS10** could bind to the *c-MYC* G-quadruplex DNA in cellular conditions and accordingly reduce the gene transcriptional level.

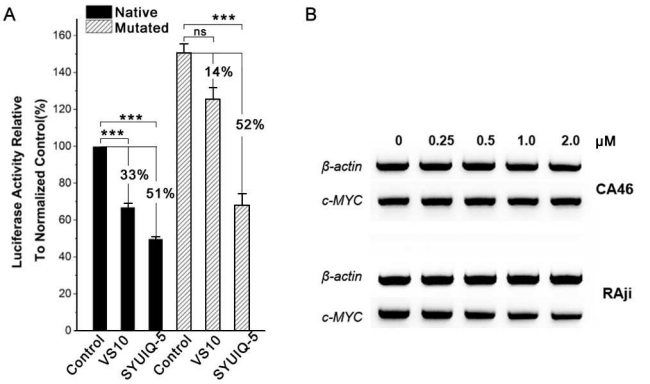



**Fig. 5** (A) Effects of **VS10** and SYUIQ-5 at the concentration of 2.0 μM on *c-MYC* promoter activity. The error bars represent the standard error from triplicates of three independent experiments. ***:significantly different ($P < 0.001$). ns: no significant difference ($P > 0.05$). (B) Effect of **VS10** at the concentration of 0.25, 0.5, 1.0 and 2.0 μM on *c-MYC* transcription.

In summary, a multistep structure-based virtual screening was carried out, incorporating docking with G-quadruplex and duplex DNA in order to identify selective G-quadruplex binders. A new ligand, **VS10**, was identified, which exhibited high selectivity for *c-MYC* G-quadruplex versus duplex DNA as determined by SPR, FRET-based competition and luciferase activity assays. The results demonstrate the feasibility of applying a multiple step structure-based virtual screening approach to selective targeting of G-quadruplex. This work may shed light on the search for selective binders for G-quadruplexes against duplex DNA. Furthermore, it represents an important first step that can be used in the process of discovering selective binders that target a specific G-quadruplex structure.

The interaction between **VS10** and *c-MYC* G-quadruplex DNA was further studied by UV titration, G-quadruplex fluorescent intercalator displacement, molecular modeling and NMR spectroscopy methods. Notably, there is still room for improvement of the binding of **VS10** to G-quadruplex. Structural modification of **VS10** to further improve the binding affinity is currently underway. It's also worth mentioning that an identical compound emerged recently as a multi-target antimalarial agent,[16] which provides added evidence that this compound is a potential drug lead that requires further comprehensive structural modification and understanding of its mechanism of action.

This work was financially supported by the National Science Foundation of China (No. 91213302, 81330077 and 21272291) and Program for Changjiang Scholars and Innovative Research Team in University of China (No. IRT1298). We also thank the Shared Hierarchical Academic Research Computing Network (SHARCNET, www.sharcnet.ca) for a generous allocation of computer resources.

## Notes and references

[a] School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou 510006, China. *(J.-H. Tan) Tel: +86 20 39943053, E-mail: tanjiah@mail.sysu.edu.cn; *(Z.-S. Huang) Tel: +86 20 39933056, E-mail: ceshzs@mail.sysu.edu.cn.

[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.

† Electronic Supplementary Information (ESI) available: [Experimental procedures, and supplemental tables, spectra and graphs]. See DOI: 10.1039/c000000x/

1 (a) T. A. Brooks and L. H. Hurley, *Nat. Rev. Cancer*, 2009, **9**, 849; (b) S. Balasubramanian, L. H. Hurley and S. Neidle, *Nat. Rev. Drug Discov.*, 2011, 1**0**, 261; (c) M. L. Bochman, K. Paeschke and V. A. Zakian, *Nat. Rev. Genet.*, 2012, **13**, 770.

2 (a) G. Biffi, D. Tannahill, J. McCafferty and S. Balasubramanian, *Nat. Chem.*, 2013, **5**, 182.

3 (a) D. Monchaud and M.-P. Teulade-Fichou, *Org. Biomol. Chem.*, 2008, **6**, 627; (b) C. Shan, J.-H. Tan, T.-M. Ou and Z.-S. Huang, *Sci. China Chem.*, 2013, **56**, 1351; (c) B. Maji and S. Bhattacharya, *Chem. Commun.*, 2014, **50**, 6422.

4 (a) S. Neidle, *Febs J.*, 2010, **277**, 1118; (b) S. A. Ohnmacht and S. Neidle, *Bioorg. Med. Chem. Lett.*, 2014, **24**, 2602.

5 R. J. Harrison, J. Cuesta, G. Chessari, M. A. Read, S. K. Basra, A. P. Reszka, J. Morrell, S. M. Gowan, C. M. Incles, F. A. Tanious, W. D. Wilson, L. R. Kelland and S. Neidle, *J. Med. Chem.*, 2003, **46**, 4463.

6 X.-D. Wang, T.-M. Ou, Y.-J. Lu, Z. Li, Z. Xu, C. Xi, J.-H. Tan, S.-L. Huang, L.-K. An, D. Li, L.-Q. Gu and Z.-S. Huang, *J. Med. Chem.*, 2010, **53**, 4390.

7 L. Martino, A. Virno, B. Pagano, A. Virgilio, S. Di Micco, A. Galeone, C. Giancola, G. Bifulco, L. Mayol and A. Randazzo, *J. Am. Chem. Soc.*, 2007, **129**, 16048.

8 (a) D.-L. Ma, T.-S. Lai, F.-Y. Chan, W.-H. Chung, R. Abagyan, Y.-C. Leung and K.-Y. Wong, *ChemMedChem*, 2008, **3**, 881; (b) Q. Li, J. Xiang, X. Li, L. Chen, X. Xu, Y. Tang, Q. Zhou, L. Li, H. Zhang, H. Sun, A. Guan, Q. Yang, S. Yang and G. Xu, *Biochimie*, 2009, **91**, 811; (c) S. Cosconati, L. Marinelli, R. Trotta, A. Virno, L. Mayol, E. Novellino, A. J. Olson and A. Randazzo, *J. Am. Chem. Soc.*, 2009, **131**, 16336; (d) H.-M. Lee, D. S.-H. Chan, F. Yang, H.-Y. Lam, S.-C. Yan, C.-M. Che, D.-L. Ma and C.-H. Leung, *Chem. Commun.*, 2010, **46**, 4680; (e) S.-B. Chen, J.-H. Tan, T.-M. Ou, S.-L. Huang, L.-K. An, H.-B. Luo, D. Li, L.-Q. Gu and Z.-S. Huang, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 1004; (f) D. S.-H. Chan, H. Yang, M. H.-T. Kwan, Z. Cheng, P. Lee, L.-P. Bai, Z.-H. Jiang, C.-Y. Wong, W.-F. Fong, C.-H. Leung and D.-L. Ma, *Biochimie*, 2011, **93**, 1055; (g) D.-L. Ma, D. S.-H. Chan, W.-C. Fu, H.-Z. He, H. Yang, S.-C. Yan and C.-H. Leung, *PLoS One*, 2012, **7**, e43278; (h) S. Alcaro, C. Musetti, S. Distinto, M. Casatti, G. Zagotto, A. Artese, L. Parrotta, F. Moraca, G. Costa, F. Ortuso, E. Maccioni and C. Sissi, *J. Med. Chem.*, 2013, **56**, 843; (i) H. R. Nasiri, N. M. Bell, K. I. McLuckie, J. Husby, C. Abell, S. Neidle and S. Balasubramanian, *Chem. Commun.*, 2014, **50**, 1704.

9 (a) C. V. Dang, L. M. S. Resar, E. Emison, S. Kim, Q. Li, J. E. Prescott, D. Wonsey and K. Zeller, *Exp. Cell Res.*, 1999, **253**, 63; (b) T.-M. Ou, Y.-J. Lu, C. Zhang, Z.-S. Huang, X.-D. Wang, J.-H. Tan, Y. Chen, D.-L. Ma, K.-Y. Wong, J. C.-O. Tang, A. S.-C. Chan and L.-Q. Gu, *J. Med. Chem.*, 2007, **50**, 1465; (c) V. Gonzalez and L. H. Hurley, *Annu. Rev. Pharmacol. Toxicol.*, 2010, **50**, 111.

10 A. Ambrus, D. Chen, J. Dai, R. A. Jones and D. Yang, *Biochemistry*, 2005, **44**, 2048.

11 A. Canals, M. Purciolas, J. Aymami and M. Coll, *Acta Crystallogr. D Biol. Crystallogr.*, 2005, **61**, 1009.

12 K. Uytterhoeven, J. Sponer and L. Van Meervelt, *Eur. J. Biochem.*, 2002, **269**, 2868.

13 (a) D. Peng, J.-H. Tan, S.-B. Chen, T.-M. Ou, L.-Q. Gu and Z.-S. Huang, *Bioorg. Med. Chem.*, 2010, **18**, 8235; (b) S.-B. Chen, Q.-X. Shi, D. Peng, S.-Y. Huang, T.-M. Ou, D. Li, J.-H. Tan, L.-Q. Gu and Z.-S. Huang, *Biochim. Biophys. Acta, Gen. Subj.*, 2013, **1830**, 5006.

14 A. T. Phan, V. Kuryavyi, H. Y. Gaw and D. J. Patel, *Nat. Chem. Biol.*, 2005, **1**, 167

15 (a) A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L. H. Hurley, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 11593; (b) T.-M. Ou, Y.-J. Lu, C. Zhang, Z.-S. Huang, X.-D. Wang, J.-H. Tan, Y. Chen, D.-L. Ma, K.-Y. Wong, J. C.-O. Tang, A. S.-C. Chan and L.-Q. Gu, *J. Med. Chem.*, 2007, **50**, 1465.

16 D. Reker, M. Seet, M. Pillong, C. P. Koch, P. Schneider, M. C. Witschel, M. Rottmann, C. Freymond, R. Brun, B. Schweizer, B. Illarionov, A. Bacher, M. Fischer, F. Diederich and G. Schneider, *Angew. Chem., Int. Ed.*, 2014, **53**, 7079.

