This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal’s standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

www.rsc.org/chemcomm
Identification of selective G-quadruplex DNA binder by a multistep virtual screening approach

Jin-Qiang Hou, a,b 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*

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. 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.

Thus far, a number of small molecules targeting biologically relevant G-quadruplexes have been reported. Their anticancer activities and mechanism of actions were carefully studied to direct medicinal chemistry research targeting G-quadruplexes. 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 and SYUIQ-5 bind to duplex DNA via intercalation, while some unfused aromatic G-quadruplex ligands, such as distamycin A, bind to the groove of duplex DNA. 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.

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. 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. Secondly, the remaining 28,530 small molecules were docked to the NMR structure of c-MYC G-quadruplex (PBD entry: 1XAV) 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 1K2Z)\(^1\) 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 1Z3F)\(^1\) 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†). 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 \(\mu\)M, while the compound has no duplex binding with a concentration of up to 10.0 \(\mu\)M, 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 (4FID) assay (Fig. S4, ESI†) were carried out, and the \(K_D\) value of 6.6\( \times \)10\(^5\) \(\text{M}^{-1}\) and DC50 value of 1.8 \(\mu\)M were determined respectively. The result from the 4FID 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\(^\circ\)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 \(\mu\)M competitor, while 20 \(\mu\)M 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 \(\mu\)M 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.

**Table 1 Hit list table from virtual ligand screening.**

<table>
<thead>
<tr>
<th>Compd. ID</th>
<th>Score_G4</th>
<th>Score_intercalator</th>
<th>Score_groovebinder</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>VS1</td>
<td>9.72</td>
<td>7.29</td>
<td>5.99</td>
<td>1.33</td>
<td>1.62</td>
</tr>
<tr>
<td>VS2</td>
<td>9.35</td>
<td>8.51</td>
<td>4.90</td>
<td>1.10</td>
<td>1.91</td>
</tr>
<tr>
<td>VS3</td>
<td>9.21</td>
<td>8.11</td>
<td>6.20</td>
<td>1.14</td>
<td>1.49</td>
</tr>
<tr>
<td>VS4</td>
<td>8.07</td>
<td>7.85</td>
<td>5.44</td>
<td>1.03</td>
<td>1.48</td>
</tr>
<tr>
<td>VS5</td>
<td>8.99</td>
<td>7.62</td>
<td>4.59</td>
<td>1.18</td>
<td>1.96</td>
</tr>
<tr>
<td>VS6</td>
<td>8.75</td>
<td>7.51</td>
<td>3.66</td>
<td>1.53</td>
<td>2.39</td>
</tr>
<tr>
<td>VS7</td>
<td>7.86</td>
<td>6.37</td>
<td>5.17</td>
<td>1.23</td>
<td>1.52</td>
</tr>
<tr>
<td>VS8</td>
<td>7.91</td>
<td>7.55</td>
<td>5.82</td>
<td>1.29</td>
<td>1.67</td>
</tr>
<tr>
<td>VS9</td>
<td>10.65</td>
<td>7.33</td>
<td>6.20</td>
<td>1.45</td>
<td>1.72</td>
</tr>
<tr>
<td>VS10</td>
<td>9.09</td>
<td>7.94</td>
<td>7.97</td>
<td>1.14</td>
<td>1.14</td>
</tr>
<tr>
<td>VS11</td>
<td>8.28</td>
<td>5.36</td>
<td>4.36</td>
<td>1.54</td>
<td>1.90</td>
</tr>
<tr>
<td>VS12</td>
<td>6.60</td>
<td>5.67</td>
<td>3.83</td>
<td>1.16</td>
<td>1.72</td>
</tr>
<tr>
<td>VS13</td>
<td>6.64</td>
<td>4.86</td>
<td>4.45</td>
<td>1.37</td>
<td>1.49</td>
</tr>
</tbody>
</table>

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.

**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\(^\prime\) end of the G-quartet and another two docking poses representing ligand stacking over the 3\(^\prime\) 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†). 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.

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. 5B, 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 III 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. 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 III 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.

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 fluorocenter 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, 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.

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 different concentrations for 24 h on c-MYC transcription.

In addition, 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.
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.bH. Tan) Tel: +86 20 39943053, E-mail: tanjiah@mail.sysu.edu.cn; *(Z.bS. Huang) Tel: +86 20 39933056, E-mail: ceshbz@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 5K7 Canada.

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


