Pharmacophore-based discovery of triaryl-substituted imidazole as new telomeric G-quadruplex ligand

Shuo-Bin Chen, Jia-Heng Tan *, Tian-Miao Ou, Shi-Liang Huang, Lin-Kun An, Hai-Bin Luo, Ding Li, Lian-Quan Gu, Zhi-Shu Huang *

School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou 510006, People's Republic of China

ABSTRACT

Discovery of potent and selective ligands for telomeric G-quadruplex DNA is a challenging work. Through a combination approach of pharmacophore model construction, model validation, database virtual screening, chemical synthesis and interaction evaluation, we discovered and confirmed triaryl-substituted imidazole **TSIZ01** to be a new telomeric G-quadruplex ligand with potent binding and stabilizing activity to G-quadruplex DNA, as well as a 8.7-fold selectivity towards telomeric G-quadruplex DNA over duplex DNA.

Human telomeric DNA, located at the very end of chromosomes, consists of tandem repeats of sequence d[(TTAGGG)$_n$] which ends as the single strand.[1] This guanine-rich single strand can adopt higher-order and functionally useful structures called the G-quadruplex DNA. The building blocks of G-quadruplexes are the G-quartets, which stack up with one on top of another to form such secondary DNA structures (Fig. 1).[2] Induction and stabilization of the telomeric G-quadruplexes by small molecules have been shown to interfere with telomere biological functions, inhibit telomerase activity and eventually alter telomere maintenance.[3] Telomere maintenance is crucial for the unlimited proliferative potential of cancer cells,[4] thus, the design of drugs targeting the telomeric G-quadruplex DNA is a rational and promising approach for cancer chemotherapy.[5–7]

Several classes of ligands have been identified to be G-quadruplex interactive ligands, including those based on an acridine scaffold.[8] Neidle and co-workers previously reported a number of such acridine derivatives with well-characterized activity.[9–11] The best compound BRACO-19 (Table 1, compound **1**) exhibits strong G-quadruplex binding and stabilizing activity, effective telomerase inhibitory activity, and accompanying significant in vivo anticancer activity in tumor xenografts associated with telomere maintenance inhibition.[12–15] These findings accompanied with recent 3D-QSAR studies have provided very useful information for further

drug design.[16] However, the acridine derivatives might face stability problems during the preparation of dosage forms and their storage due to the hydrolysis of the amide bonds on the acridine ring.[17] Accordingly, discovery of potent, selective and stable ligands for telomeric G-quadruplex DNA is still a crucial and challenging work.

In view of above researches, here we report the development of new ligand-based pharmacophore models on the basis of the acridine derivatives in the present study. The best quantitative pharmacophore model generated was subsequently used as a 3D query to screen our in-house compound database to find novel structurally diverse G-quadruplex ligands. Finally, their interactions with telomeric G-quadruplex DNA were examined by an array of biophysical assays.

Generation of the pharmacophore model: In this approach, telomeric G-quadruplex ligands were firstly collected from the same literature for the training set in order to assure their comparability (Table 1).[11] These compounds included a series of acridine derivates with telomerase inhibition activity (EC$_{50}$ values) ranged from 0.067 to 6.9 µM. 3D-QSAR Pharmacophore Generation module/Discovery Studio (version 2.5, Accelrys Inc., San Diego, CA) was used to construct pharmacophore model. Since the aromatic planar feature can be recognized as the hydrophobic moiety in the software, we used hydrophobic moiety here instead of the former so as to discover more structurally diverse compounds.[18] Therefore, a minimum number of features involving hydrogen bond acceptor and donor, positive ionizable and hydrophobic moieties were specified while generating the pharmacophore model.
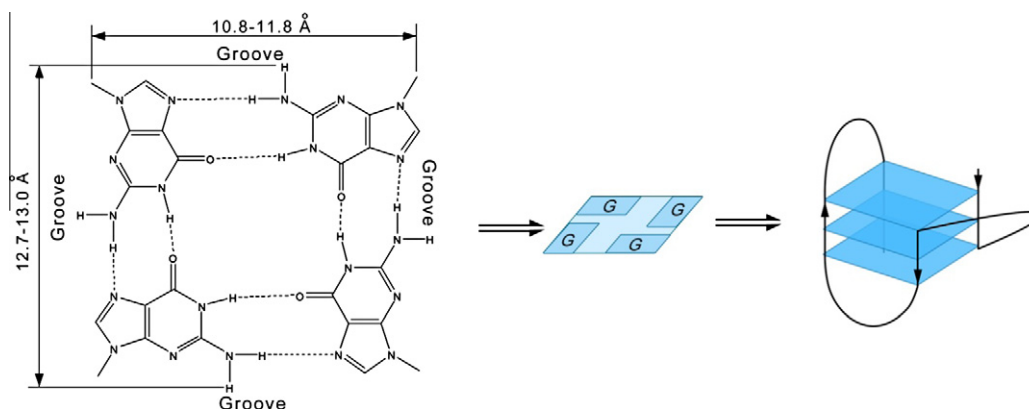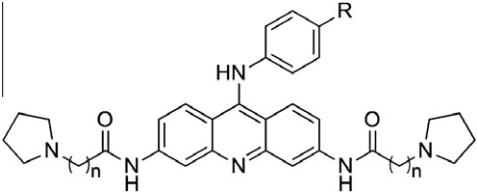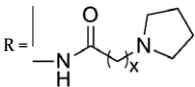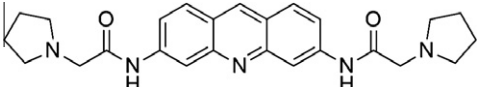

Figure 1. Basic structure of human telomeric G-quadruplex DNA.

Table 1
Chemical structures of the 11 training set compounds together with their experimental inhibitory activities on telomerase (EC_{50} values)

				R = $-N(CH_3)_2$				R = (pyrrolidine amide)	
1	$n = 2$	$EC_{50} = 113$ nM	**5**			$n = 2$	$x = 1$	$EC_{50} = 167$ nM	
2	$n = 3$	$EC_{50} = 99$ nM	**6**			$n = 2$	$x = 2$	$EC_{50} = 67$ nM	
3	$n = 4$	$EC_{50} = 1930$ nM	**7**			$n = 2$	$x = 3$	$EC_{50} = 117$ nM	
4	$n = 5$	$EC_{50} = 6910$ nM	**8**			$n = 3$	$x = 2$	$EC_{50} = 326$ nM	
			9			$n = 4$	$x = 2$	$EC_{50} = 255$ nM	
			10			$n = 5$	$x = 2$	$EC_{50} = 146$ nM	
			11			$EC_{50} = 5200$ nM			

HypoGen hypothesis generation process produced 10 top-scored hypotheses based on the activity values of the training set molecules. Three statistical parameters of such hypotheses, including null cost, fixed cost and total cost values, were used to signify the reliability of a pharmacophore model. The null cost represented the highest cost of a pharmacophore without features, which estimated activity to be the average of the activity data of the training set molecules. The fixed cost represented the simplest model that fits all data perfectly, while the total cost was the cost analysis of each hypothesis. A reliable pharmacophore model should have a high correlation co-efficiency, low total cost, low RMS (root mean square) and error values. The difference between the null and fixed cost should be greater than 70 bits, and the difference between null and total cost (cost difference) should be greater than 60 bits. In this work, the null cost was 5943.6, and the fixed cost was 29.0. The difference between null and fixed cost was found to be greater than 70 bits. Among the 10 hypotheses, Hypo1 showed the highest correlation co-efficiency values of 0.998, demonstrating its good prediction ability. Meanwhile, it also showed the lowest total cost values of 73.9. Moreover, higher cost difference and correlation value with low RMS and error values have been observed for Hypo1 when compared with other hypotheses.

As showed in Figure 2A, Hypo1 hypothesis had four features, which included one hydrogen bond donor, one hydrophobic and two positive ionizable sites. The distance constraints of these pharmacophore features were labeled. It showed that the positively charged atoms in acridine derivatives contributed positively to the activity. They might interact with the phosphate backbone of telomeric G-quadruplex DNA via electrostatic association. Meanwhile, the hydrogen bond interactions between ligands and bases in the DNA grooves or loops could also enhance their activity. Their middle hydrophobic features might further serve as a linker of other features and fill in the active sites, which could increase the van der Waals interactions between the ligands and the G-quadruplex DNA.[19]

On the other hand, although all training set molecules contain the acridine structure, the activities of these compounds were greatly different. Such differences of the activities might be derived from their distinct substituent groups. Therefore, the pharmacophore model did not consider the tricyclic aromatic plane as an important feature in the HypoGen hypothesis generation process.[18] As showed in Figure 2B, all features of Hypo1 hypothesis were nicely mapped with the corresponding chemical functional groups of highly active compound **1**. In comparison, compound
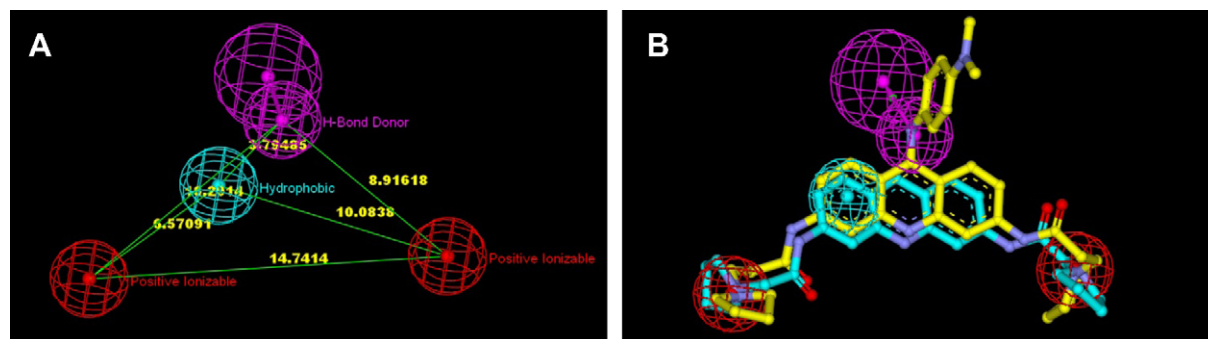
Figure 2. Pharmacophore model of telomeric G-quadruplex ligand generated by HypoGen. (A) The best pharmacophore model Hypo1. (B) Mapping of the representative compounds **1** (yellow) and **11** (cyan) onto Hypo1 hypothesis.
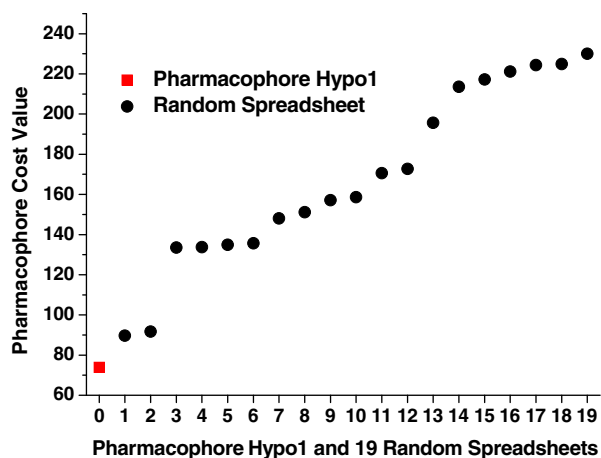

Figure 3. The difference in cost values between Hypo1 runs and the 19 lowest cost values scrambled runs.

Figure 4. ROC curve analysis.

Table 2
Decoy test results for Hypo1 hypothesis

Parameter	Values
D, total number of molecules in database	4000
A, total number of actives in database	74
Ht, total number of hit molecules from the database	111
Ha, total number of active molecules in hit list	69
% Yield of actives [(**Ha/Ht**) × 100]	62.16
% Ratio of actives [(**Ha/A**) × 100]	93.24
EF, enrichment factor, [(**Ha/Ht**)/(**A/D**)]	33.60
False negatives [**A** − **Ha**]	5
False Positives [**Ht** − **Ha**]	42
AUC of receiver operating characteristic	0.96553

generated, and some of the results were listed in Figure 3. All 99 resulting hypotheses after randomization exhibited worse performance than Hypo1. As a result, Hypo1 hypothesis was not generated by chance.

Secondly, decoy set was generated to evaluate the efficiency of Hypo1 hypothesis by computing enrichment factor (EF) value.[21] The screening was performed using the Ligand Pharmacophore Mapping protocol in DS, and the results were shown in Table 2. EF value was found to be 33.6, thus showing very good indications of the high efficiency of the screening. Furthermore, the ROC curve of the test set screened by Hypo1 hypothesis was showed in Figure 4 and the area under the ROC curve (AUC) was 0.96, which meant that a selected randomly active compound had a higher score than a randomly selected inactive compound for 9.6 times out of 10.[22] All of the above validation results suggested that Hypo1 hypothesis would be valuable and reliable in identifying the compounds for further screening.

Virtual and experimental screening: After the validation procedure, Hypo1 hypothesis was used as a 3D structural query for retrieving potential telomeric G-quadruplex ligands from our in-house compound library, which consisted of more than 5000 natural products and their derivatives. A total of 31 molecules were identified with Ligand Pharmacophore Mapping protocol, and their fit values were ranged from 3.8 to 14.7. Compounds with molecular weight over 850 were removed, and 21 molecules were left for the second step. Furthermore, only those compounds that could be easily obtained from the compound library or re-synthesized by ourselves were retained. At this stage, the number of hits decreased to 10. Among these compounds, six molecules that belong

11 with low activity just mapped hydrophobic and positive ionizable features, while the other feature of hydrogen bond donor was not mapped. In view of the above results, Hypo1 was selected as a best hypothesis for further study.

Validation of the pharmacophore model: Validating the hypothesis is one of the significant approaches in pharmacophore generation. In this approach, the reliability of Hypo1 hypothesis was validated by means of Fischer's randomization method, enrichment factor (EF) analysis and receiver-operating characteristic (ROC) curve analysis. Fischer's randomization method was firstly used to evaluate the statistical significance of Hypo1 hypothesis.[20] Validation was carried out by randomizing the activity data among the training set compounds and generating pharmacophore model using the same features and parameters originated for Hypo1. To achieve the confidence level of 99%, 99 random spreadsheets were
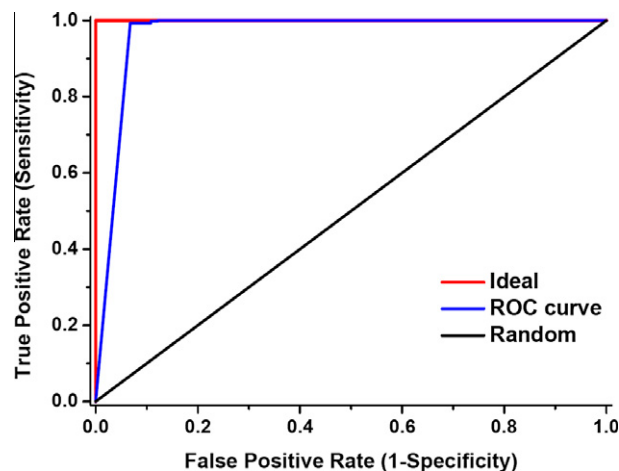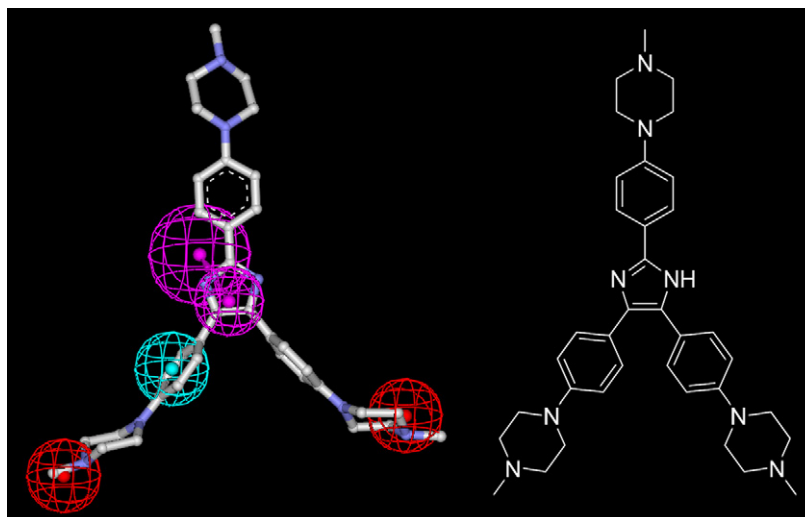
Figure 5. Chemical structure of **TSIZ01** and mapping of **TSIZ01** onto Hypo1 hypothesis.

to quindoline and berberine derivatives have been previously proved to be potent telomeric G-quadruplex ligands (Supplementary data), and four new triaryl-substituted imidazole derivatives were then subjected to fast experimental screening with FRET assay. Of the tested compounds, only one molecule, **TSIZ01**, with the fit value of 7.071 (Supplementary data), was found to bind and stabilize telomeric G-quadruplex DNA significantly. As shown in Figure 5, **TSIZ01** was nicely mapped with the pharmacophore model. It appeared that the terminal nitrogen atoms in methylpiperazine groups of **TSIZ01** mapped to the two positive ionizable features, one of its phenyl ring mapped to the hydrophobic feature, and the N–H moiety of imidazole ring mapped to the hydrogen bond donor feature. Meanwhile, it should be noted that other three triaryl-substituted imidazole derivatives could also be mapped with the pharmacophore model and had better fit values (Supplementary data). However, their aromatic amines were hard to be protonated at physiological pH, thus showing undetectable activity in the FRET assay.

Detailed biophysical investigation of TSIZ01: To gain insight into the interaction between **TSIZ01** and telomeric G-quadruplex DNA, CD, FRET and SPR experiments were carried out. CD spectroscopy was firstly used to investigate the induction and binding property of **TSIZ01** to telomeric G-quadruplex DNA.[23] In the presence of K[+], HTG21 oligonucleotide (5′-d(GGG[TTAGGG]$_3$)-3′) formed the hybrid-type G-quadruplex structure.[24] The CD spectrum of such preformed structure showed a strong positive band at 290 nm, a positive shoulder band at 270 nm, a minor positive band at 252 nm, and a negative band at 235 nm (Fig. 6, black line).[25] Upon titration of **TSIZ01** into this DNA solution, the CD spectra changed with enhancement of both 290 and 270 nm band. Meanwhile, an induced CD signal was observed between 310 and 380 nm. The induced CD signal was an additional evidence for the interaction between the G-quadruplex and **TSIZ01**, because its appearance in the CD spectra was indicative of an achiral ligand binding tightly to a chiral host.[26] The above CD experimental results indicated that **TSIZ01** could not induce major changes in telomeric quadruplex topology upon binding.[27]

Secondly, FRET melting experiments were carried out to study the stabilization effect and selectivity of **TSIZ01** for telomeric G-quadruplex DNA.[28] The effect of 1 μM **TSIZ01** on the melting temperature enhancement (ΔT_m) of two labeled oligonucleotides in K[+] solution was shown in Table 3, and its concentration dependent melting curves were shown in Figure 7A. F21T (5′-*FAM*-d(GGG[TTAGGG]$_3$)-*TAMRA*-3′) represented the human telomeric DNA sequence, while F10T (5′-*FAM*-dTATAGCTATA-HEG-TATAGC TATA-*TAMRA*-3′) is a hairpin duplex DNA. The FRET-melting data demonstrated that **TSIZ01** effectively stabilized the telomeric G-quadruplex DNA. At 1 μM of ligand, the ΔT_m value was 23.5 °C. This value would further increase to 32.0 °C when the concentration of **TSIZ01** was up to 3 μM. On the other hand, **TSIZ01** exhibited good preference for binding to telomeric G-quadruplex rather than duplex DNA. ΔT_m value of 1 μM **TSIZ01** binding to the duplex DNA was just 2.7 °C, which was much lower than that of G-quadruplex DNA. Moreover, the G-quadruplex selectivity of **TSIZ01** was further assessed with a FRET-based competition assay, and the activity of the ligand to retain the stability of G-quadruplex was challenged by nonfluorescent duplex DNA (ds26). In the presence of 3 μM of competitor ds26, the thermal stabilization of F21T enhanced by **TSIZ01** was not significantly affected. Even with the addition of 10 μM competitor ds26, **TSIZ01** (2 μM) still retained its activity of stabilizing G-quadruplex (Fig. 7B). The combined results of these FRET assays demonstrated that **TSIZ01** had high levels of stabilization activity and good selectivity for the telomeric G-quadruplex DNA.[27,29,30]

Finally, to provide insight into the binding affinity and selectivity of **TSIZ01** for telomeric G-quadruplex DNA, SPR experiments were carried out to quantitatively determine the kinetic constants of
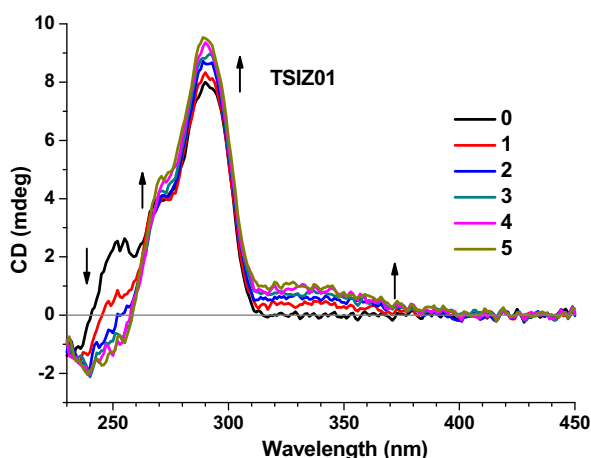

Figure 6. CD titration spectra of HTG21 (5 μM) at increasing concentrations (arrows: 0–5 mol equiv) of **TSIZ01** in 10 mM Tris–HCl buffer with 100 mM KCl, pH 7.2.

Table 3
Stabilization temperatures (ΔT_m) determined with FRET melting experiment, and equilibrium binding constants (K_A) with SPR

	$\Delta T_m{}^a$ (°C) at 1 μM		K_A (M^{-1})		Specificity G4:duplex
	F21T	F10T	Telomeric G4	Duplex DNA	
TSIZ01	23.5 ± 1.6	2.7 ± 0.6	2.19×10^6	2.51×10^5	8.7:1

^{a} $\Delta T_m = T_m$ (DNA + ligand) − T_m (DNA). The concentrations of F21T and F10T were both 0.2 μM. In the absence of ligand, T_m values of annealed F21T and F10T were 62.5 and 63.5 °C, respectively.
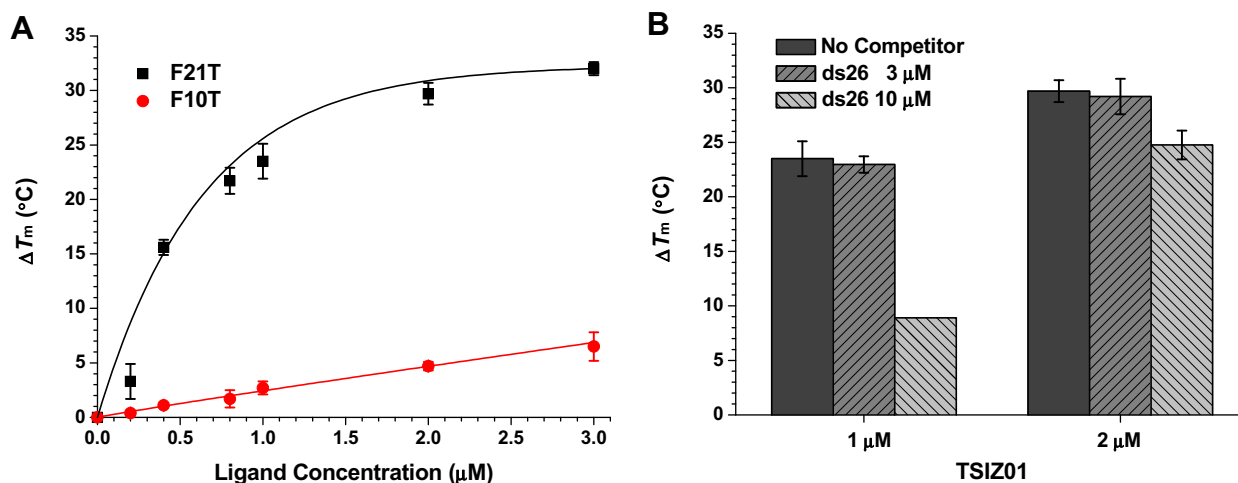

Figure 7. (A) Concentration dependant melting curves (ΔT_m vs ligand concentration) for **TSIZ01** upon binding to F21T and F10T. (B) Competitive FRET results for **TSIZ01** (1 and 2 μM), without (dark gray) and with 15-fold (3 μM; gray) or 50-fold (10 μM; light gray) excess of duplex DNA competitor (ds26). The concentration of F21T was 0.2 μM.

the derivatives binding to either G-quadruplex or duplex DNA.[31,32] As shown in Table 3, **TSIZ01** had the higher binding affinity to telomeric G-quadruplex DNA with a K_A value of 2.19×10^6 M^{-1}. For the binding selectivity between G-quadruplex and duplex, **TSIZ01** showed an obvious selectivity of 8.7-fold (2.19×10^6 vs 2.51×10^5 M^{-1}). These results were consistent with those from the FRET studies, which showed that **TSIZ01** was a new potent and selective G-quadruplex ligand.

In summary, we developed a reliable ligand-based pharmacophore model on the basis of well-characterized telomeric G-quadruplex ligands, the acridine derivatives. By using this pharmacophore model as a 3D query to screen our in-house compound database, we successfully identified a new G-quadruplex binding ligand, namely **TSIZ01**. **TSIZ01** is a triaryl-substituted imidazole derivative with good stability. In our experimental studies, it exhibited potent binding and stabilization activity as well as good selectivity towards telomeric G-quadruplex DNA. Inspired by these promising results, our research group is currently pursuing structural optimization of this class of ligands using CADD methods in parallel with organic synthesis.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2010.12.019.

References and notes

1. Blackburn, E. H. *Nature* **1991**, *350*, 569.
2. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402.
3. De Cian, A.; Lacroix, L.; Douarre, C.; Temime-Smaali, N.; Trentesaux, C.; Riou, J.-F.; Mergny, J.-L. *Biochimie* **2008**, *90*, 131.
4. Masutomi, K.; Yu, E. Y.; Khurts, S.; Ben-Porath, I.; Currier, J. L.; Metz, G. B.; Brooks, M. W.; Kaneko, S.; Murakami, S.; DeCaprio, J. A.; Weinberg, R. A.; Stewart, S. A.; Hahn, W. C. *Cell* **2003**, *114*, 241.
5. Hurley, L. H. *Nat. Rev. Cancer* **2002**, *2*, 188.
6. Neidle, S.; Parkinson, G. *Nat. Rev. Drug Disc.* **2002**, *1*, 383.
7. Ou, T.-M.; Lu, Y.-J.; Tan, J.-H.; Huang, Z.-S.; Wong, K.-Y.; Gu, L.-Q. *ChemMedChem* **2008**, *3*, 690.
8. Tan, J.-H.; Gu, L.-Q.; Wu, J.-Y. *Mini-Rev. Med. Chem.* **2008**, *8*, 1163.
9. Read, M.; Harrison, R. J.; Romagnoli, B.; Tanious, F. A.; Gowan, S. H.; Reszka, A. P.; Wilson, W. D.; Kelland, L. R.; Neidle, S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4844.
10. Harrison, R. J.; Cuesta, J.; Chessari, G.; Read, M. A.; Basra, S. K.; Reszka, A. P.; Morrell, J.; Gowan, S. M.; Incles, C. M.; Tanious, F. A.; Wilson, W. D.; Kelland, L. R.; Neidle, S. *J. Med. Chem.* **2003**, *46*, 4463.
11. Moore, M. J. B.; Schultes, C. M.; Cuesta, J.; Cuenca, F.; Gunaratnam, M.; Tanious, F. A.; Wilson, W. D.; Neidle, S. *J. Med. Chem.* **2006**, *49*, 582.
12. Gowan, S. M.; Harrison, J. R.; Patterson, L.; Valenti, M.; Read, M. A.; Neidle, S.; Kelland, L. R. *Mol. Pharmacol.* **2002**, *61*, 1154.
13. Incles, C. M.; Schultes, C. M.; Kempski, H.; Koehler, H.; Kelland, L. R.; Neidle, S. *Mol. Cancer Ther.* **2004**, *3*, 1201.
14. Burger, A. M.; Dai, F.; Schultes, C. M.; Reszka, A. P.; Moore, M. J.; Double, J. A.; Neidle, S. *Cancer Res.* **2005**, *65*, 1489.
15. Kelland, L. R. *Eur. J. Cancer* **2005**, *41*, 971.
16. Zambre, V. P.; Murumkar, P. R.; Giridhar, R.; Yadav, M. R. *J. Chem. Inf. Model.* **2009**, *49*, 1298.
17. Taetz, S.; Muerdter, T. E.; Zapp, J.; Boettcher, S.; Baldes, C.; Kleideiter, E.; Piotrowska, K.; Schaefer, U. F.; Klotz, U.; Lehr, C. M. *Int. J. Pharm.* **2008**, *357*, 6.
18. Li, Q.; Xiang, J.; Li, X.; Chen, L.; Xu, X.; Tang, Y.; Zhou, Q.; Li, L.; Zhang, H.; Sun, H.; Guan, A.; Yang, Q.; Yang, S.; Xu, G. *Biochimie* **2009**, *91*, 811.
19. Monchaud, D.; Teulade-Fichou, M.-P. *Org. Biomol. Chem.* **2008**, *6*, 627.
20. Sakkiah, S.; Thangapandian, S.; John, S.; Kwon, Y. J.; Lee, K. W. *Eur. J. Med. Chem.* **2010**, *45*, 2132.
21. Hawkins, P. C.; Skillman, A. G.; Nicholls, A. *J. Med. Chem.* **2007**, *50*, 74.
22. Triballeau, N.; Acher, F.; Brabet, I.; Pin, J. P.; Bertrand, H. O. *J. Med. Chem.* **2005**, *48*, 2534.

23. Paramasivan, S.; Rujan, I.; Bolton, P. H. *Methods* **2007**, *43*, 324.
24. Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res.* **2006**, *34*, 2723.
25. Xue, Y.; Kan, Z.-Y.; Wang, Q.; Yao, Y.; Liu, J.; Hao, Y.-H.; Tan, Z. *J. Am. Chem. Soc.* **2007**, *129*, 11185.
26. White, E. W.; Tanious, F.; Ismail, M. A.; Reszka, A. P.; Neidle, S.; Boykin, D. W.; Wilson, W. D. *Biophys. Chem.* **2007**, *126*, 140.
27. Drewe, W. C.; Nanjunda, R.; Gunaratnam, M.; Beltran, M.; Parkinson, G. N.; Reszka, A. P.; Wilson, W. D.; Neidle, S. *J. Med. Chem.* **2008**, *51*, 7751.

28. De Cian, A.; Guittat, L.; Kaiser, M.; Sacca, B.; Amrane, S.; Bourdoncle, A.; Alberti, P.; Teulade-Fichou, M.-P.; Lacroix, L.; Mergny, J.-L. *Methods* **2007**, *42*, 183.
29. Tan, J.-H.; Ou, T.-M.; Hou, J.-Q.; Lu, Y.-J.; Huang, S.-L.; Luo, H.-B.; Wu, J.-Y.; Huang, Z.-S.; Wong, K.-Y.; Gu, L.-Q. *J. Med. Chem.* **2009**, *52*, 2825.
30. De Cian, A.; Delemos, E.; Mergny, J.-L.; Teulade-Fichou, M.-P.; Monchaud, D. *J. Am. Chem. Soc.* **2007**, *129*, 1856.
31. Redman, J. E. *Methods* **2007**, *43*, 302.
32. Wang, X.-D.; Ou, T.-M.; Lu, Y.-J.; Li, Z.; Xu, Z.; Xi, C.; Tan, J.-H.; Huang, S.-L.; An, L.-K.; Li, D.; Gu, L.-Q.; Huang, Z.-S. *J. Med. Chem.* **2010**, *53*, 4390.