Selective recognition of c-MYC Pu22 G-quadruplex by a fluorescent probe

Qianqian Zhai1,2,†, Chao Gao3,†, Jieqin Ding2,†, Yashu Zhang2, Barira Islam4, Wenxian Lan5, Haitao Hou2, Hua Deng2, Jun Li2, Zhe Hu1, Hany I. Mohamed2,6, Shengzhen Xu2, Chunyang Cao5, Shozeb M. Haider4 and Dengguo Wei1,2,∗

1State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan 430070, PR China, 2College of Science, Huazhong Agricultural University, Wuhan 430070, China, 3College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, PR China, 4UCL School of Pharmacy, University College London, 29–39 Brunswick Square, London WC1N 1AX, UK, 5State Key Laboratory of Bioorganic and Natural Products Chemistry and Collaborative Innovation Center of Chemistry for Life Sciences, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 345 Lingling Road, Shanghai, 200032, China and 6Chemistry Department, Faculty of Science, Benha University, Benha 13518, Egypt

Received October 26, 2018; Revised January 22, 2019; Editorial Decision January 23, 2019; Accepted January 25, 2019

ABSTRACT

Nucleic acid mimics of fluorescent proteins can be valuable tools to locate and image functional biomolecules in cells. Stacking between the internal G-quartet, formed in the mimics, and the exogenous fluorophore probes constitutes the basis for fluorescence emission. The precision of recognition depends upon probes selectively targeting the specific G-quadruplex in the mimics. However, the design of probes recognizing a G-quadruplex with high selectivity in vitro and in vivo remains a challenge. Through structure-based screening and optimization, we identified a light-up fluorescent probe, 9CI that selectively recognizes c-MYC Pu22 G-quadruplex both in vitro and ex vivo. Upon binding, the biocompatible probe emits both blue and green fluorescence with the excitation at 405 nm. With 9CI and c-MYC Pu22 G-quadruplex complex as the fluorescent response core, a DNA mimic of fluorescent proteins was constructed, which succeeded in locating a functional aptamer on the cellular periphery. The recognition mechanism analysis suggested the high selectivity and strong fluorescence response was attributed to the entire recognition process consisting of the kinetic match, dynamic interaction, and the final stacking. This study implies both the single stacking state and the dynamic recognition process are crucial for designing fluorescent probes or ligands with high selectivity for a specific G-quadruplex structure.

INTRODUCTION

G-quadruplex structures have been demonstrated to play important roles in mediating biological processes including functioning as diagnostic tools (1–7). However, the lack of ligands or fluorescent probes selectively targeting a specific G-quadruplex topology limits their unambiguous identification in live cells (8–10). Endogenic 4-hydroxybenzimidazolinone (HBI) formed from three residues in the nascent green fluorescent proteins (GFP), work as essential visualizing tools that locates proteins and monitors biological process in living cells and organisms (11–13). As intrinsically fluorescent RNA is not known, an RNA aptamer ‘Spinach’ was identified as mimic of GFP with HBI derivatives as the exogenous fluorophores (14–16). Crystal structures demonstrated that it was the two-layer non-canonical G-quadruplex and an adjacent base-triplet in the RNA aptamer that served as a pocket to accommodate HBI derivatives, and restrain their vibration (17,18). Similar restraint was assumed to occur in DNA mimics of red fluorescent proteins and other RNA fluogenic light-up aptamers, such as Mango and Corn (19,20). The stacking between the G-quartet surface and the fluorophores constitutes the basis for the fluorescent property of these RNA\DNA mimics of FPs, which provide tools for genetic encoding of fluorescent nucleic acids and for tracking bio-molecules in cells (17,18,21,22). However, precision of these nucleic acid mimics of FPs on locating target biological molecules in cells is affected by multiple G-quadruplexes that can form physically (23,24). They share a common G-quartet surface,
which has a potential to interact with the exogenous analogs of HBI (25–30). To improve precision, it is essential to develop methods for designing probes that recognize a specific G-quadruplex forming oligonucleotide with high selectivity and specificity in vitro and in vivo (9,10,31–34).

G-quadruplexes (G4) are four-stranded structures formed by guanine-rich DNA or RNA sequences, in which four guanines are assembled in a square co-planar arrangement by Hoogsteen hydrogen bonding to form a G-quartet (26–30). The G-quartets stack on top of one another to form a G-quadruplex. G-quadruplex-forming sequences are widely distributed in eukaryotic telomeres (35,36), and the promoter regions of genes such as c-MYC (37–39), c-KIT (40,41), k-RAS (42,43), BCL2 (44,45) and VEGF (46,47). G-quadruplex structures can be classified into various groups according to the orientation of the nucleic acid strands, such as parallel, antiparallel or hybrids thereof, and these structures share a common G-quartet. In the reported G-quadruplex–ligand complex structures in PDB, most of ligands interact with G-quadruplex via G-quartet stacking (30,48). Some compounds display light-up fluorescence property when interacting with G-quadruplex, and part of them exhibit selectivity to distinguish G-quadruplexes from the DNA duplex structure, which is derived from the π–π stacking between the planar aromatic core of the compounds and the common G-quartet surface (49,50). However, generic G-quartet stacking results in poor selectivity between G-quadruplex structures. Interaction with the flanking loops around the G-quartet enhances the binding of the stacked ligands (51). The adjacent grooves extend the binding pockets, and dynamic movement between G-quadruplex and the ligands explore more intermediate states (52,53). These could provide more diversified space for designing ligands with selectivity for a specific G-quadruplex structure.

The design of probes that recognizes a G-quadruplex topology with high selectivity and specificity in vitro and in vivo remains a challenge. In this study, we approach this issue by identifying a fluorescent probe that selectively recognizes a G-quadruplex highlighted that the entire recognition process might consist of a kinetic match, dynamic interaction, and the final stacking. This recognition mechanism rationalizes the high selectivity of 9CI, and has been further supported by the relationship between the structures of analogous compounds of 9CI and their measured fluorescent intensity.

### MATERIALS AND METHODS

#### Materials

All oligonucleotides used in this study were purchased from Sangon (Shanghai, China). G-quadruplex forming oligonucleotides were annealed in the buffer (10 mM K2HPO4/KH2PO4 pH 7.0, 100 mM KCl) by heating to 95°C for 5 min, followed by gradual cooling to room temperature. Most compounds in this study were purchased from the Chinese national compound library of Peking University (PKU-CNCL), and compounds 5 (9CI) and 6 were synthesized in-house (Scheme S1). During the synthesis, all chemical reagents were commercially available and immobilizes the coplanar chromophore and ensures the dominance of fluorescence emission in excited state energy release (56). In this study, virtual screening based on the pockets was carried out to identify fluorescent probes of c-MYC Pu22 G-quadruplex. Commercially available compounds with fluorophores were docked into the ligand binding cavities in the structures to judge their fitness, and in vitro fluorescence assay was employed to measure their fluorescent response. Using identified hits as templates for further optimization, compound 9CI (6-(2-(anthracen-9-ylmethylene) hydrazinyl)-N,N,N'-diphenyl-1,3,5-triazine-2,4-diamine (9CI) and its division into sub-structures. The triazine ring is labeled as A, and B represents the anthracene ring. C and D represent the phenyl rings. Nitrogen atoms connecting the triazine ring and the phenyl rings are labeled as 1 and 2, respectively.

Scheme 1. Structure of 6-(2-(anthracen-9-ylmethylene) hydrazinyl)-N,N,N'-diphenyl-1,3,5-triazine-2,4-diamine (9CI) and its division into sub-structures. The triazine ring is labeled as A, and B represents the anthracene ring. C and D represent the phenyl rings. Nitrogen atoms connecting the triazine ring and the phenyl rings are labeled as 1 and 2, respectively.
treated with standard methods before use unless otherwise noted. Silica gel column chromatography (CC) (silica gel 200–300 mesh) was ordered from Qingdao Makall Group Co., Ltd, Qingdao, China. 1H NMR and 13C NMR spectra were recorded in DMSO-d6 on a 600 MHz Varian VNMR Spectrometer and resonances were given in ppm relative to tetramethylsilane (TMS). The following abbreviations were used to designate chemical shift multiplicities: s = singlet, d = doublet, t = triplet, m = multiplet, br = broad. High-resolution mass spectra (HRMS) were acquired in positive mode on a MALDI SYNAPT G2 high-definition mass spectrometer (Waters, Milford, MA, USA).

Synthesis of 9CI (compound 5) and compound 6
Compounds I (1.0 equivalent) and compound II (1.0 equivalent) (Scheme S1) were synthesized according to the literature (59,60). Both of them were dissolved in 10 ml anhydrous EtOH, and two drops of acetic acid were added to the mixture (59,60). Both of them were dissolved in 10 ml anhydrous EtOH, and two drops of acetic acid were added to the mixture. The reaction was kept at 80°C for 7 h; and then, the solvent was removed under reduced pressure, and the residue was purified by silica column chromatography to get 9CI (compound 5) and 6 as yellow solid.

For compound 5 (9CI): 73% yields. 1H NMR (600 MHz, DMSO-d6): 11.22 (s, 1H), 9.44 (s, 3H), 8.80 (s, 1H), 8.71 (d, J = 6.0 Hz, 2H), 8.17 (d, J = 6.0 Hz, 2H), 7.95–7.90 (m, 4H), 7.65–7.59 (m, 4H), 7.28 (s, 4H), 7.00 (s, 2H). 13C NMR (150 MHz, DMSO-d6): 164.87, 164.71, 142.27, 140.60, 131.46, 129.96, 129.41, 128.85, 125.98, 125.57, 122.38, 122.46, 122.41, 120.68. HRMS cacld for [M+H]+: 482.2088, found: 482.2146.

For compound 6: 80% yields. 1H NMR (600 MHz, DMSO-d6): 11.17 (s, 1H), 9.47 (s, 2H), 8.62 (s, 1H), 8.58 (s, 1H), 8.45 (s, 1H), 8.18–8.10 (m, 6H), 7.90 (s, 3H), 7.56–7.54 (m, 2H), 7.34 (s, 4H), 7.03 (t, J = 7.2 Hz, 2H).

Fluorescence measurement
Fluorescence spectra were scanned by Spectrofluorophotometer RF-5301PC (Shimadzu, Japan) at room temperature. For fluorescence intensity measurement, 0.5 μM annealed DNA samples were mixed with compounds at 0.5 μM in the buffer (10 mM K2HPO4/KH2PO4 pH 7.0, 100 mM KCl), and the mixture was incubated for 10 min at 30°C before the spectra collection. Full UV absorbance profiles were scanned for compounds and mixture with DNA and compounds. The peaks appeared in the profile were used as the initial excitation wavelengths to obtain the emission spectra, in which, the maximum emission wavelength was determined. With the fixed maximum emission wavelengths, the peak wavelength in the excitation spectra was chosen as the final excitation wavelengths. For the fluorescence change of 9CI with the titration of DNA samples, small aliquots of annealed DNA oligonucleotide were added into the solution with 0.5 μM 9CI in the buffer (10 mM K2HPO4/KH2PO4 pH 7.0, 100 mM KCl). The final concentration of DNA samples varied from 0 to 2 μM. The mixture with DNA and 9CI was stirred and allowed to equilibrate for 10 min at 30°C before the measurement. The emission between 420 and 580 nm was recorded at the excitation wavelength of 405 nm with 5 nm excitation slit width and 5 nm emission slit width.

The fluorescence variation of 9CI with the titration of c-MYC Pu 22 G-quadruplex was used to derive Ka value according to the method of Benesi-Hildebrand (61), using the following equation: 1/(ΔF) = 1/(b[DNA][H]0Kb) + 1/[b[H]0]. ΔF represents the change in the fluorescence intensity (ΔF = F − F0); F denotes the fluorescence intensity in the presence of different concentration of c-MYC DNA; F0 is the maximum fluorescence intensity of the free compound 9CI. [DNA] is the total added c-MYC Pu 22 G-quadruplex forming oligonucleotide concentration, and [H]0 is the total concentration of compound 9CI.

Gel electrophoresis
The native gel electrophoresis was performed on a 20% polyacrylamide gel in 1× TBE running buffer containing 100 mM KCl at room temperature. 8 μl annealed DNA (10 μM) in 1× loading buffer were loaded into the different wells except for c-MYC Pu22 (20 μM) in lane 3. The electrophoresis ran at 100 V. SYBR Gold was used to stain one gel, after being diluted ten thousand times with TBE (Tris–borate–EDTA) buffer. The second gel was obtained with DNA samples premixed with 20 μM 9CI before the electrophoresis. Bands were detected by Tanon-1600 trans-UV light. The UV wavelength used for image acquisition is 302 nm.

Cell cytotoxicity assay
The cell growth effect of 9CI against A549 (human lung carcinoma cell) were tested via measuring the quantity of living cells using the Cell Counting Kit-8 (CCK8) (DOJINDO, Japan) according to the instruction. A549 cells were cultured in Dulbecco’s Modified Eagle Medium (Gibco) with 1% glutamine and 10% fetal bovine serum (FBS) for 24 h at 37°C. These cells were treated with 9CI at the indicated concentrations (2.5, 5, 10, 20 μM) for 24 h. Then, 100 μl of 10% CCK-8 solution was added into wells, and the cells were incubated for another 1 h at 37°C. The cytotoxicity was evaluated basing on the percentage of cell survival in a dose-dependent manner regard to the control, and the optical density (OD) detected at 450 nm by automated microplate reader (Bio-Rad) was used to calculate the survival rate. All the experiments were performed in parallel and in triplicate.

Confocal imaging of co-localization
A549 cells were cultured in Dulbecco’s modified Eagle’s medium (Gibco) with 1% glutamine and 10% fetal bovine serum (FBS) for 24 h at 37°C. 40 pmol annealed c-MYC Pu22, TTA and c-KIT1 G-quadruplex forming oligonucleotides labelled with 5’-Cy5 was transfected into cells by using lipofectamine 2000 (Thermo Fisher) and incubated at 37°C for 5 h. Then 5 μM 9CI was added and incubated
Table 1. Sequences of the oligonucleotides used in imaging SL1

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence(5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cy5-c-MYC Pu22</td>
<td>5'-Cy5-TGAGGGTGAGGTGGCTGGGTTGAA</td>
</tr>
<tr>
<td>Cy5-TTA</td>
<td>5'-Cy5-GGGTTAGGGTTAGGGTTAGGG</td>
</tr>
<tr>
<td>Cy5-c-KIT 1</td>
<td>5'-Cy5-TAATGGAGGAGGGCGGTTGGGAGGGAGGGG</td>
</tr>
<tr>
<td>c-MYC Pu22-MFP</td>
<td>TGAGGGTGGGTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTGAAGTCTGACTGCAAGGCA</td>
</tr>
</tbody>
</table>
| TTA2-MFP            | GAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG
Figure 1. Structures of compounds 1–9 and their fluorescence response to eighteen DNA oligonucleotides. The excitation and emission wavelengths of these compounds are listed in Supplementary Table S2.

charges were assigned to the selected representative compounds. A search box centered at the quality center of compound quindoline in the complex structure was set up with size_x = 30, size_y = 25, size_z = 23 to encompass the cleft and three grooves in the c-MYC Pu22 G-quadruplex structure. Ten separate docking runs were performed for each selected compound by AutoDock Vina, using the default search parameters. Compounds with docking scores from AutoDock Vina in the top 30% were re-docked using Autodock4.0 for more accurate docking. Grid points of 80 x 66 x 60 with a 0.375 Å spacing, centered on quindoline structure were calculated for all the ligand atom types using AutoGrid 4.2. Each docking calculation with Autodock 4.0 consisted of 25 million energy evaluations and a maximum of 270 000 generations using the Lamarckian genetic algorithm local search method. A population size of 300 and 300 rounds of Solis and Wets local search were applied with a probability of 0.06. A mutation rate of 0.02 and a crossover rate of 0.8 were used to generate new docking trials for subsequent generations, and the best individual from each generation was propagated to the next generation. One hundred binding poses were predicted for each compound by Autodock 4.0, and were clustered with root-mean-square deviation (RMSD) cutoff value 2 Å. The predicted binding poses with the frequency of occurrence higher than 10/100 were visually inspected for their fitness with the binding pocket.

Molecular dynamics simulations

The stability of 9CI interacting with c-MYC Pu22 G-quadruplex was further assessed by MD simulations. The docked 9CI - c-MYC Pu22 G-quadruplex complex was set up using the xleap module in AMBER16 molecular dynamics package (http://ambermd.org/). The ligand parameter was generated using the GAFF2 force field implemented in the Antechamber program. The simulation was run using parmbsc1 (65) version of Cornell et al. force field with OL15 modifications (66–68). Two potassium ions were placed at the center of the G-tetrads. The complex was further neutralized by adding K+ ions. The system was solvated using Amber-adapted Joung and Chetham (69) parameters specific for TIP3P waters. The solvated box whose boundary extended at least 10 Å from any solute atom was used as the starting configuration for our simulation. The complex was first minimized using 1000 steps of steepest descent, followed by 1000 steps of conjugate gradient minimization. A 300 ps MD equilibration was run in which the quadruplex was restrained, while the ions and water were allowed to equilibrate. The system was gently heated from 0 to 300 K with a time step of 0.5 ps followed by another round at constant pressure at 300 K for 1 ns. The constraints were gradually relaxed until no constraints were applied on the systems. The final production MD run was carried out for 1000 ns using ACEMD (70) molecular dynamics engine. The default parameters were used with 10 Å non-bonded
Figure 2. (A) Cartoon representation of superimposed c-MYC Pu22 G-quadruplex structures in apo state (brown, model 1 in the NMR structure, PDB 1xav) and with quindoline (purple, model 1 in the NMR structure, PDB 2l7v). The two quindoline molecules are shown in cyan. (B) Superimposed structures without ligands. (C) Connolly surface of c-MYC Pu22 G-quadruplex structure from PDB 2l7v. The red spheres highlight the central platform on top of 3' end G-quartet and at the bottom of 5' end G-quartet, while green, yellow and blue spheres represent the locations of grooves connecting to the central platform, respectively.

Figure 3. (A) Fluorescence emission spectra of 0.5 μM 9CI in the presence of various oligonucleotides (0.5 μM), λex = 405 nm. Fluorescence emission of compound 9CI alone is shown in black. 0.5 μM oligonucleotide samples were in 10 mM K2HPO4/KH2PO4 buffer solution with 100 mM KCl at pH 7.0. (B) Fluorescence intensity enhancement of 0.5 μM 9CI against concentration of different DNA oligonucleotides. (C) Fluorescence intensity of 9CI in the presence of c-MYC Pu18, c-MYC Pu22, c-MYC Pu24 and c-MYC Pu27. Both 9CI and DNA are 0.5 μM. λex = 405 nm, λem = 472 nm.

Figure 4. Electrophoresis of DNA oligonucleotides on 20% non-denaturing acrylamide gel. (A) Gel stained by SYBR Gold after electrophoresis; (B) DNA samples were premixed with 20 μM 9CI before the electrophoresis. SYBR Gold used for staining was diluted ten thousand times from original stock solution with TBE (Tris-borate-EDTA) buffer. 80 pmol DNAs were loaded into the different wells except for c-MYC Pu22 (160 pmol) in lane 3. Lane 1: DNA marker with 22 nucleotides; lane 2: c-MYC Pu22 (80 pmol); lane 3: c-MYC Pu22 (160 pmol); lane 4: TTA; lane 5: c-KIT1; lane 6: c-KIT2; lane 7: VEGF; lane 8: TBA; lane 9: CT4; lane 10: 19AT.

RESULTS

Structure-based virtual screening for identifying fluorescent probes

In the quindoline-c-MYC Pu22 G-quadruplex complex structure (PDB code 2l7v), both 3' and 5' flanking sequences exhibit ligand-induced reorientation in the binding pockets. Pockets in the complex structure and in apo state (Figure 2) were used as potential binding sites. Consensus docking using a combination of Autodock Vina (62)
Fluorescence intensity measurement in solution and structure optimization

Eighteen DNA oligonucleotides (Supplementary Table S1) with the potential to form single strand, duplex, triplex, and G-quadruplexes were employed to check the fluorescence response of the selected compounds. Compound 1 and 2 exhibited enhanced fluorescence intensity at different exciting and emitting wavelengths (Figure 1, Supplementary Figure S3 and Table S2). In view of the selectivity of compound 2 for c-MYC Pu22 G-quadruplex, it was chosen as the template to further search for compounds with higher selectivity from the database. New ligands with groups common in compound 2 or with bioisosteres to them were purchased and synthesized (Figure 1, Supplementary Figure S1, Schemes S1 and Figures S4 and S5), and their fluorescence response to the eighteen oligonucleotides was measured (Figure 1 and Supplementary Figure S1).

In vitro fluorescence selectivity of 9CI

9CI (compound 5), with anthracene ring replacing 4-ethenyl-N,N-dimethylamino in compound 2, exhibited high fluorescence response and selectivity against c-MYC Pu22 G-quadruplex (Figure 3). The absorbance spectrum of 9CI displayed two vibronic bands centered at 263 and 405 nm (Supplementary Figure S6A). To avoid overlap with the nucleic acid absorbance, the fluorescence spectra were scanned with the excitation wavelength at 405 nm. Compound 9CI alone in buffer displayed extremely negligible emission at 472 nm. With gradual titration of c-MYC Pu22 G-quadruplex DNA (up to 2 μM), the fluorescence emission at 472 and 503 nm exhibited a dramatic increase (Figure 3B), even without a distinct shift in absorption (Supplementary Figure S6B). In contrast, non-significant fluorescence enhancements were observed when 9CI was mixed or titrated with other DNA sequences, including antiparallel G-quadruplex (Bom17, Asc20, CT4, TBA, 21CTA), parallel G-quadruplex (c-KIT 1, c-KIT 2, Kras, VEGF, AKTI), mixed G-quadruplex (TATA, TTA 45, Pias24), trinucleotide (CGG12), duplex (19AT), single strand and triplex (Figure 1, Supplementary Table S1, Figure 3). 9CI did not yield significant fluorescent response towards the wild-type c-MYC G-quadruplex forming DNA oligonucleotides (Pu18, Pu24 and Pu27) (Supplementary Table S1, Figure S2).

Selective recognition of 9CI with c-MYC Pu22 G-quadruplex

Co-localization of c-MYC Pu22 G-quadruplex and 9CI is an intuitive way to determine whether 9CI is able to recognize c-MYC Pu22 G-quadruplex in live cells. Negligible cytotoxic effect was observed for 9CI over 24 h incubation and Autodock 4.0 (63) was carried out to screen for fluorescent probes. Compounds with fluorophores from the Chinese National Compound Library of Peking University (PKU-CNCL) were clustered through structural similarity analysis using Open Babel (64). The representative compounds were selected for virtual screening. The binding modes of the representative compounds in the clefts were first predicted by AutoDock Vina, which achieves an approximately two orders of magnitude speed-up compared with Autodock 4.0. Compounds with docking scores ranked in the top 30% from Autodock Vina were selected and re-docked using Autodock 4.0 to predict their binding poses with greater accuracy. Binding poses with a cluster population size > 10 (out of 100) were visually inspected for their fitness within the binding pockets. Twenty representative compounds were purchased and tested for their fluorescence response to DNA oligonucleotides that adopt different secondary structures.

Figure 5. Visualization of 9CI mixed with DNA oligonucleotides under UV light. (A) 5 μM 9CI and 5 μM DNA samples in 10 mM KH₂PO₄/K₂HPO₄ buffer with 100 mM KCl at pH 7.0. (B) Lane 1: 200 μM c-MYC Pu22 DNA; lane 2: 50 μM 9CI; lane 3: 25 μM c-MYC Pu22 DNA + 25 μM 9CI; lane 4: 50 μM c-MYC Pu22 DNA + 50 μM 9CI; lane 5: 100 μM c-MYC Pu22 DNA + 100 μM 9CI; lane 6: 200 μM c-MYC Pu22 DNA + 200 μM 9CI.
Compound 9CI displayed highly selective fluorescent recognition for c-MYC Pu22 G-quadruplex in vitro and in live cells. This encouraged us to exploit the 9CI - c-MYC Pu22 G-quadruplex complex as the fluorescent response core in a DNA mimic of fluorescent proteins to locate bio-functional molecules in cells, such as aptamers, and biomarkers. Aptamers are single-stranded DNA or RNA oligonucleotides with unique molecular recognition and binding capabilities specific to various targets, including small molecules, proteins, and live cells (73–75). The identification of their targeted cells and targets in live cell benefits the mechanism exploration and further optimization. Aptamer SL1 was found to affect the phosphorylation level in A549 and suppress cancer cell motility (76). To locate aptamer SL1 in cells, a fluorescent probe (c-MYC Pu22-MFP; Table 1) was constructed with two segments: a reporter strand bearing two repeated c-MYC Pu22 G-quadruplex sequences binding with 9CI and a matched recognition strand with the extended aptamer SL1-M (Table 1). The constructed fluorescent probe (c-MYC Pu22-MFP and 9CI) showed high fluorescent recognition for the extended SL1 (SL1-M) in vitro through hybridization (Supplementary Figure S13). At first, SL1-M was incubated with A549 cells in DPBS for 30 min at room temperature, following which c-MYC Pu22-MFP and 9CI were incubated to the cells sequentially (Figure 7). After the cells were washed with buffer three times, the bright blue ring and the green ring on the cellular periphery of A549 cells at 405 nm excitation suggested the aptamer SL1 might target the membrane receptors of A549 cells (Figure 8), whereas only weak fluorescence was observed in the control HepG2 cells. This is consistent with the reported mechanism that SL1 might target c-Met, a cell surface receptor for the hepatocyte growth factor in A549 cells (76). However, no signal was observed for SL1-M in control experiment carried out with a fluorescent probe (TAA2-MFP in Table 1) constructed with the TTA G-quadruplex forming oligonucleotide sequence replacing c-MYC Pu22 G-quadruplex forming oligonucleotide sequence (Figure 8).

In order to highlight the application of locating aptamers in live cell, additional DNA oligonucleotides, based on AS1411, were constructed (Supplementary Table S3). Using the same protocol, a blue ring and a green ring was observed on the cellular periphery of A549 cells at 405 nm excitation (Supplementary Figure S14), suggesting aptamer AS1411 might target protein at the cell surface. This is consistent with the report that aptamer AS1411 could bind with nucleolin on the external surface of the tumor cells (77,78).

Figure 6. Fluorescence images of A549 cells transfected with 40 pmol c-MYC Pu22 DNA G-quadruplex forming oligonucleotide labelled with 5′-Cy5 using lipofectamine 2000 and then incubated with 5 μM 9CI for 2 hours successively. (A) Fluorescence signal collected between 655–755 nm at λex = 635 nm, (B) fluorescence signal collected between 425–470 nm at λex = 405 nm, (C) fluorescence signal collected between 500 and 545 nm at λex = 405 nm, (D) merged image of (A) and (C), (E) merged images of (A), (B), (C) and the bright field. Scale bar, 20 μm.
Figure 7. Schematic illustration of locating aptamers in cells by DNA mimic of fluorescent proteins with the 9CI–c-MYC Pu22 G-quadruplex complex as the fluorescent response core.

Table 2. c-MYC Pu22 G-quadruplex sequence and the mutants in the binding site analysis

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5′-3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c-MYC Pu22</td>
<td>TGAGGGTGGGTAGGGTGGGTAA</td>
</tr>
<tr>
<td>c-MYC Pu22-G2&gt;C2</td>
<td>TCAGGTTGGTAGGTTGGGTAA</td>
</tr>
<tr>
<td>c-MYC Pu22-A3&gt;T3</td>
<td>TGAGGTTGGTAGGTTGGGTAA</td>
</tr>
<tr>
<td>c-MYC Pu22-T11&gt;A11</td>
<td>TGAGGTTGGTAAGGTTGGGTAA</td>
</tr>
<tr>
<td>c-MYC Pu22-T20&gt;A20</td>
<td>TGAGGTTGGTAGGTTGGGA</td>
</tr>
<tr>
<td>c-MYC Pu22-A22&gt;T22</td>
<td>TGAGGTTGGGTAAGGTTGGGTAT</td>
</tr>
<tr>
<td>c-MYC Pu22-G10&gt;T10</td>
<td>TGAGGTTGGGTAGGTTGGGTAA</td>
</tr>
</tbody>
</table>

Binding site analysis

Compound 9CI exhibited high selectivity for c-MYC Pu22 G-quadruplex in \textit{in vitro} and \textit{ex vivo} and the constructed probe succeeded in locating aptamer SL1 at the periphery of A549 cells. These implied its potential on improving the precision of nucleic acid mimics of fluorescent proteins to track biomolecules in cells. Dissection of the recognition mechanism would provide important clues for designing ligands or fluorescent probes for a specific G-quadruplex structure. Job plot analysis showed that 9CI bound with c-MYC Pu22 G-quadruplex with 1:1 stoichiometry (Supplementary Figure S15). ITC (Isothermal titration calorimetry) assay and fluorimetric titration showed that the dissociation constants were on the order of magnitude of $10^{-3}$ M$^{-1}$ (Supplementary Figures S16 and S17), suggesting a relative weak interaction between 9CI and c-MYC Pu22 DNA G-quadruplex. CD (Circular Dichroism) spectra showed that the addition of 9CI did not change the topology of the c-MYC Pu22 DNA G-quadruplex and other DNA structures (Supplementary Figure S18).

Fluorescence alterations upon recognition between 9CI and the mutant c-MYC Pu22 G-quadruplex sequences (Table 2) were analyzed to investigate the binding site of 9CI. The oligonucleotides mutated at the 5′ end loop (G2 > C2, A3 > T3), 3′ end loop (T20 > A20, A22 > T22), middle loop (T11 > A11), and 3′ end G-quartet (G10 > T10) were purchased from Sangon (Shanghai, China). The CD spectra showed that these mutated DNA oligonucleotides adopted the same topology as c-MYC Pu22 DNA G-quadruplex,
for G10 in the 3′ end G-quartet did not change the G-quadruplex structure (Supplementary Figure S19). Instead, the weak fluorescence response (Figure 9) indicated that the mutation on G10 might prevent the interaction between 9CI and 3′ end G-quartet. The mutation from T20 to A20 resulted in negligible fluorescence response, and the oligonucleotide with mutation A22 > T22 exhibited the highest fluorescence enhancement to 9CI. Non-significant fluorescence change with mutation in the 5′ end loop and significant fluorescence variation with the mutation in the 3′ end G-quartet and the T20-A22 loop indicated that compound 9CI might be located between the T20-A22 loop and 3′ end G-quartet.

Recognition mechanism explored by automated docking and molecular dynamics simulations.

Fluorescence changes of 9CI with the mutant c-MYC Pu22 G-quadruplex oligonucleotides showed 3′ end cleft could be the recognition site. Exploring their recognition mode at the atomic level could provide detailed information for further structural modification. Automated docking coupled with molecular dynamics simulations were carried out to explore its recognition mechanism in the 3′ end cleft. 9CI was docked into the cleft at the 3′ end of c-MYC Pu22 G4-quadruplex structure (PDB ID: 2l7v) with 200 repeat runs using Autodock 4.0. The predicted binding conformation with the highest frequency of occurrence among 200 conformations was considered as the initial recognition mode. In this orientation (Supplementary Figure S20), the anthracene ring was found to stack on the top of G10 and G15 in 3′ end G-quartet surface; three nitrogen atoms in the triazine ring and nitrogen atoms N1 or N2 (Scheme 1) interacted with the phosphate oxygen atoms of T11 via electrostatic interaction and the phenyl ring stacked with the thymine of T11. In this case, the initial interaction between 9CI and the G-quadruplex structure constituted the basis for the dynamic recognition model.

One μsec molecular dynamics simulation (Supplementary Figure S21) was carried out to check the dynamic recognition between 9CI and c-MYC Pu22 G-quadruplex. In the initial contact, the triazine ring was attracted to the phosphate oxygen atoms through electrostatic interaction, while the anthracene ring of 9CI stacked on two guanines of the 3′ terminal end G-quartet (Figure 10A). As the anthracene ring approached the T20–A21–A22 loop, A22 moved away from the approaching ligand (Figure 10B). Further insertion of 9CI disrupted T20–A22 base pair, and forced the bases in loop T20–A21–A22 to stack (Figure 10C). In the following structural perturbations, 9CI and the T20–A21–A22 loop translated and rotated to obtain an optimal stacking mode. Finally, T20–A22 base pair reformed and the anthracene ring was sandwiched between T20–A22 base pair and G15–G19 on the 3′ end G-quartet (Figure 10F, Supplementary Figure S22). The triazine and the phenyl ring stacked on top of G6 (Figure 10F), while the bridge nitrogen atom interacted with the phosphate oxygen atom in T7 via electrostatic interaction (Supplementary Figure S22C). These stable interactions prevented intramolecular motions of 9CI, and ensured the dominance of emission pathway in excited state energy release. It has been
Figure 10. Dynamic recognition between 9CI and c-MYC Pu22 G-quadruplex explored by molecular dynamic simulations.

Figure 11. Schematic illustration for the induced fitting between 9CI and c-MYC Pu22 G-quadruplex.
confirmed by the remarkable fluorescence enhancement of 9CI in the glycerol at different concentrations (Supplementary Figure S23).

DISCUSSION

Selectivity for the cleft at the 3' end

As a functional structural unit, G-quadruplex structures have been assumed to be anti-tumour or anti-viral targets (79,80). However, the design of ligands selectively targeting a specific G-quadruplex structure in cells still remains a challenge. Surprisingly, G-quadruplex structure was found to be a key component for the fluorescent response of exogenous fluorophores (HBI analogues) in some RNA/DNA mimics of fluorescent proteins. Due to the lack of selectivity for G-quadruplex structures in the mimics of FP's, the addition of the exogenous HBI analogues would result in enormous fluorescence noise by encountering hundreds of G-quadruplex in the genome. The noise can be reduced by identifying self-illuminating nucleic acid aptamers that generate endogenous fluorophores or an alternative approach is to discover probes that selectively recognize the specific G-quadruplex topology in the aptamer. However, majority of G-quadruplex ligands described thus far show little selectivity among different G-quadruplex structures. 9CI displays high selectivity for c-MYC Pu22 G-quadruplex structure over other G-quadruplex topologies. No obvious fluorescence enhancement with three wide-type c-MYC G-quadruplex forming DNA oligonucleotides (Pu18, Pu24 and Pu27) (Figure 3C) further suggested the potential of 9CI and c-MYC Pu22 G-quadruplex complex as the fluorescence response core on precisely locating target biological molecules in cells. CD spectra showed that addition of 9CI did not change the topologies of c-MYC Pu22 DNA G-quadruplex (Supplementary Figure S18). Fluorescence variance upon the recognition between 9CI and c-MYC Pu22 G-quadruplex forming sequences mutated in the loops at 3' end, 5' end or in the middle suggested that compound 9CI bound close to the T20-A22 loop and at the 3'-end G-quartet instead of 5' end. The preference for only one terminal end G-quartet was further supported by Job plot analysis, which showed a 1:1 stoichiometry between 9CI and c-MYC Pu22 G-quadruplex.

Induced fit recognition mechanism supported by the relationship between the fluorescence response and the structures of 9CI analogues

Generally, the fluorescence response of G-quadruplex probes arises from the π stacking stabilization with G-quartet (32,81). The optimization of analogues was derived from the chemical skeleton of compound 2. Twenty three compounds with similar chemical groups in compound 2 or with bioisostere to them were measured for their fluorescence response to eighteen DNA oligonucleotides. The complicated relationship between these structures and their fluorescence response to c-MYC Pu22 G-quadruplex could not be explained by any single stacking state. However, some structure-fluorescence relationship could be rationalized.

(1) Except for c-MYC Pu22 G-quadruplex, 9CI does not exhibit obvious fluorescence to other G-quadruplex structures (Figure 1). This suggests that stacking with the common G-quartet is not determinant for the fluorescence of 9CI.

(2) Naphthalimides and benzoheterocycles groups stack well with the G-quartet (82,83). With these two fragments replacing the triazine ring and N-phenyl rings in compound 5 (9CI), compounds 8 and 9 were purchased and measured for their fluorescence response to different nucleic acid structures (Figure 1). Their negligible fluorescence response implies that the stacking between the anthracene ring in 9CI and the G-quartet is not crucial for the strong fluorescence response.

(3) The anthracene ring in 9CI is assumed to be the fluorophore. Compounds 3–6 with different fluorophores were measured for their fluorescence response. Compared with 9CI, compounds 3 and 4 with small aromatic rings did not display obvious fluorescent signals with or without DNA oligonucleotides. This implies that fluorophores with big stacking surface are helpful. With the same structural components, compound 5 (9CI) and compound 6 with substitutions at different positions of the anthracene ring possess conjugated system at the similar size. However, completely different fluorescence response with c-MYC Pu22 G-quadruplex suggests that only big stacking surface with the G-quartet is not enough for the strong fluorescent response.

(4) In spite of sharing > 70% similarity with 9CI, compound 7 did not display fluorescence. This suggests that the N-phenyl ring in 9CI is important in the recognition process and cannot be replaced by the piperidine ring in compound 7.

(5) Compounds 10–25 (Supplementary Figure S1) with random replacement for the N-phenyl, the triazine ring and the anthracene ring (parts A–D in Scheme 1) did not show obvious fluorescence response, suggesting that the cooperation of the rings play roles in the strong fluorescence response of 9CI.

Molecular dynamics simulations demonstrated that the recognition between 9CI and c-MYC G-quadruplex consisted of the kinetic match, dynamic recognition and the final stacking. The harmonious cooperation between the structural rearrangement of 3'-end loop and the systematic translocation of 9CI was critical for the induced fit (Figure 11). When N-phenyl is replaced by piperidine ring (compound 7), significant reduction in fluorescence could be due to the absence of the stacking between N-phenyl and T11 in the initial recognition. Weak fluorescence response of compounds 8 and 9 suggests the triazine ring and N-phenyl rings in 9CI play an important role in the initial fitting and the dynamic recognition before 9CI reaches the final stacked state. If compound 6 were able to reach the same final stacked state as 9CI, the anthracene ring could enable compound 6 to be a light-up fluorescent probe for G-quadruplex. However, insignificant fluorescence enhancement was observed for compound 6, suggesting that the relative orientation among the anthracene ring, the triazine ring and the phenyl rings might interfere with the initial contact and prevent further dynamic recognition. Finally, the stronger fluores-
cence response of 9CI for c-MYC Pu22 G-quadruplex over other G-quadruplexes highlights the importance of A20–T22 loop in the dynamic recognition and the sandwich interaction. The fluorescence intensity variance with different bioisosterisms of 9CI reflects the roles of different structural components play in the recognition process.

CONCLUSION

As functional molecular structures, G-quadruplexes have attracted wide attention from different areas. However, it has been challenging to find a ligand or a fluorescent probe targeting a specific G-quadruplex structure. In this study, we identified compound 9CI exhibiting selective and strong light-up fluorescence response to c-MYC Pu22 G-quadruplex over other G-quadruplex structures. 9CI exhibited good biocompatibility without inhibition in some tumor cells (9CI at 20 μM), fungi (9CI at 50 ppm), and insects (9CI at 1000 ppm). The fluorescent probe constructed with 9CI and c-MYC Pu22 G-quadruplex forming oligonucleotides as the fluorescent response unit succeeded in locating functional aptamer SL1 on the cellular periphery of A549 cells. Experimental and computational analysis demonstrated that: (a) the initial conformation match led to the structural interaction; (b) the high selectivity came from the dynamic rearrangement and (c) the thermodynamic stability of the complex determined the strong fluorescence response. This was further supported by the structure-fluorescence intensity relationship of the 9CI analogues. This work demonstrates that: (i) it is possible to discover fluorescence probes or ligands to recognize a specific G-quadruplex structure; (ii) besides the single recognition state, the dynamic recognition process could provide clues for designing fluorescent probes or ligands with high selectivity for a specific G-quadruplex structure; (iii) complex consisting of a specific G-quadruplex and its unique light-up fluorescent probe could work as the fluorescent response core to improve their precision on biomedical imaging and clinical diagnosis.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Prof. Stephen Neidle (UCL School of Pharmacy, University College London) for helpful comments and suggestions. We also thank Dr Likai Du (College of informatics, Huazhong Agricultural University) for the discussion about the theory of the fluorescence, Qi Wang (College of science, Huazhong Agricultural University) for the measurement of the binding affinity, and Shaoyong Ke (National Biopesticide Engineering Research Center, Hubei Academy of Agricultural Science) for the toxicity assays of 9CI.

FUNDING

National Natural Science Foundation of China [21502060, 31672558, 21732002]; Huazhong Agricultural University Scientific & Technological Self-innovation Foundation [2015RC013, 2662017PY113, 2662015PV208]; Open fund of The State Key Laboratory of Bio-organic and Natural Products Chemistry, CAS [SKLBNPC16343]; Open fund of Beijing National Laboratory For Molecular Sciences; Fundamental Research Funds for the Central Universities [2662016PY122]; Natural Science Foundation of Hubei Province of China [2016CFB230]; Funding for open access charge: Huazhong Agricultural University Scientific & Technological Self-innovation Foundation [2015RC013].

Conflict of interest statement. None declared.

REFERENCES


