OPEN

Existence of G-quadruplex structures in promoter region of oncogenes confirmed by G-quadruplex DNA cross-linking strategy

Libo Yuan[1]*, Tian Tian[1]*, Yuqi Chen[1], Shengyong Yan[1], Xiwen Xing[1], Zhengan Zhang[1], Qianqian Zhai[1], Liang Xu[1], Shaoru Wang[1], Xiaocheng Weng[1], Bifeng Yuan[1], Yuqi Feng[1] & Xiang Zhou[1,2]

[1]College of Chemistry and Molecular Sciences, Key Laboratory of Biomedical Polymers of Ministry of Education, Wuhan University, Wuhan, Hubei, 430072, P. R. China, [2]State Key Laboratory of Natural and Biomimetic Drugs, Peking University, Beijing, 100191, P. R. China.

Existence of G-quadruplex DNA *in vivo* always attract widespread interest in the field of biology and biological chemistry. We reported our findings for the existence of G-quadruplex structures in promoter region of oncogenes confirmed by G-quadruplex DNA cross-linking strategy. Probes for selective G-quadruplex cross-linking was designed and synthesized that show high selectivity for G-quadruplex cross-linking. Further biological studies demonstrated its good inhibition activity against murine melanoma cells. To further investigate if G-quadruplex DNA was formed *in vivo* and as the target, a derivative was synthesized and pull-down process toward chromosome DNAs combined with circular dichroism and high throughput deep sequencing were performed. Several simulated intracellular conditions, including *X. laevis* oocytes, Ficoll 70 and PEG, was used to investigate the compound's pure cross-linking ability upon preformed G-quadruplex. Thus, as a potent G-quadruplex cross-linking agent, our strategy provided both valuable evidence of G-quadruplex structures *in vivo* and intense potential in anti-cancer therapy.

T he G-quadruplex, stabilized by Hoogsteen hydrogen bonds, is one of the most important secondary structures of nucleic acids, which always forms in G-rich sequences under some monovalent cations[1–3]. In recent years, the G-quadruplex has attracted intense interests because of its potential biological functions, such as gene regulation, gene expression and antitumor potential[4–7]. Of note, G-rich sequences are unevenly distributed on some regions of the human genome, including telomeric ends, immunoglobulin switch regions and regulatory elements in some gene promoters, such as *c-myc*, *c-kit*, and so on[8–10]. During these decades, studies indicated that formation of G-quadruplexes in these regions may play important regulatory roles[11–14]. Therefore, the G-quadruplex is considered to be a promising target for antitumor drug design[15,16]. However, since G-quadruplex structure's polymorphism and variability under complex intracellular conditions, its biological function in cells remains unclear, and even its existance *in vivo* is always in controversy[17,18]. Although plentiful small molecules had develped as G-quadruplex stablizer, not much could really target G-quadruplex *in vivo*, since that they always function through noncovalent binding, resulting in declined efficiency. Evidence of G-quadruplex structure's existence *in vivo* was only proved in *Stylonychia lemnae*[19]. There are not many direct evidence presented in mammalian living systems except one article published by S. Balasubramanian[20], while our article is under reviewing.

Cross-linking, a kind of covalent interaction, has been suggested to be an efficient strategy for probing specific DNA structures. Thus, cross-linking of G-quadruplex is representing a quite promising strategy for its functional and structural study, and development of such G-quadruplex cross-linking agents have aroused wide interests. Previous reported agents were generally confined to telomeric DNAs[21–24], besides telomeric sequences, other G-rich sequences are also considerably significant, such as ones in oncogenic promoter regions which have close relationship with gene expression. To best of our knowledge, G-quadruplex cross-linking agents targeting G-rich sequences in oncogenic promoter region through endogenous induction remained to be developed. Due to frangibility, reactions of nucleic acids were required to be very mild, selective and readily controllable. *o*-Quinone intermediate,

an actively potent DNA cross-linking unit resulted from a catechol precursor upon oxidation, has been proved to possess good cross-linking activity toward duplex DNAs without resulting in G-quadruplex cross-linked[25]. A superiority of the *o*-quinone mediated DNA cross-linking was that it could be readily stimulated by tyrosinase, which is highly expressed in some malignant tumors[26–29], thus bringing in good cell selectivity. Taken together, the catechol moiety was an ideal scaffold that could fit multifaceted requirements. We previously reported N,N′-(3,4-Dihydroxylbenzyl)-N,N,N′,N′–tetramethyl-4,4′-biphenyldiamine dibromide as double strand DNA cross-linking agent[25]. In these studies, for purpose of demonstration of G-quadruplex's existence *in vivo*, we designed a new series of Schiff-base catechol derivatives as G-quadruplex cross-linking candidates.

Results

Preliminary CD study of the interaction between compounds 1-3 and some typical G-rich single strands. Besides cross-linking unit of catechol, structural component with preference for G-quadruplex was required. Consequently, compound **1** was constructed containing an aromatic plane, which could interact with G-quadruplex through π-π stacking and result in a further cross-linking effect. To further study the structure and activity relationship, N,N′-bis(3, 4-dihydroxybenzylidene)-diaminobenzene derivatives[30] were also designed and synthesized (Figure 1a, SI Schemes S1-3). Their interactions with several G-quadruplex DNAs were firstly studied by Circular Dichroism (CD). For preliminary screen, two typical guanine rich sequences respectively in gene promoters and telomere ends were used. The Pu27 sequence of 5′-TGGGGAGGGTGGGG-AGGGTGGGGAAGG-3′ was firstly used, which was known as a 27-nt guanine-rich segment in *c-myc* gene promoter region that controlled approximately 90% of the total *c-myc* transcription[12]. All three synthesized bis(catechol) compounds were expected to be converted to *o*-quinone active intermediate under oxidation by tyrosinase. CD was used to characterize the formation of G-quadruplex and corresponding thermodynamic stability ΔT_m. It was found that compounds **1–3** could individually induce and enhance stability of parallel G-quadruplex with an increased T_m. Compound **1** possess a best ΔT_m (29.6°C) over the other two ones (Figure 1b, c and SI Figures S1 and S3). Herein, it needs to be pointed
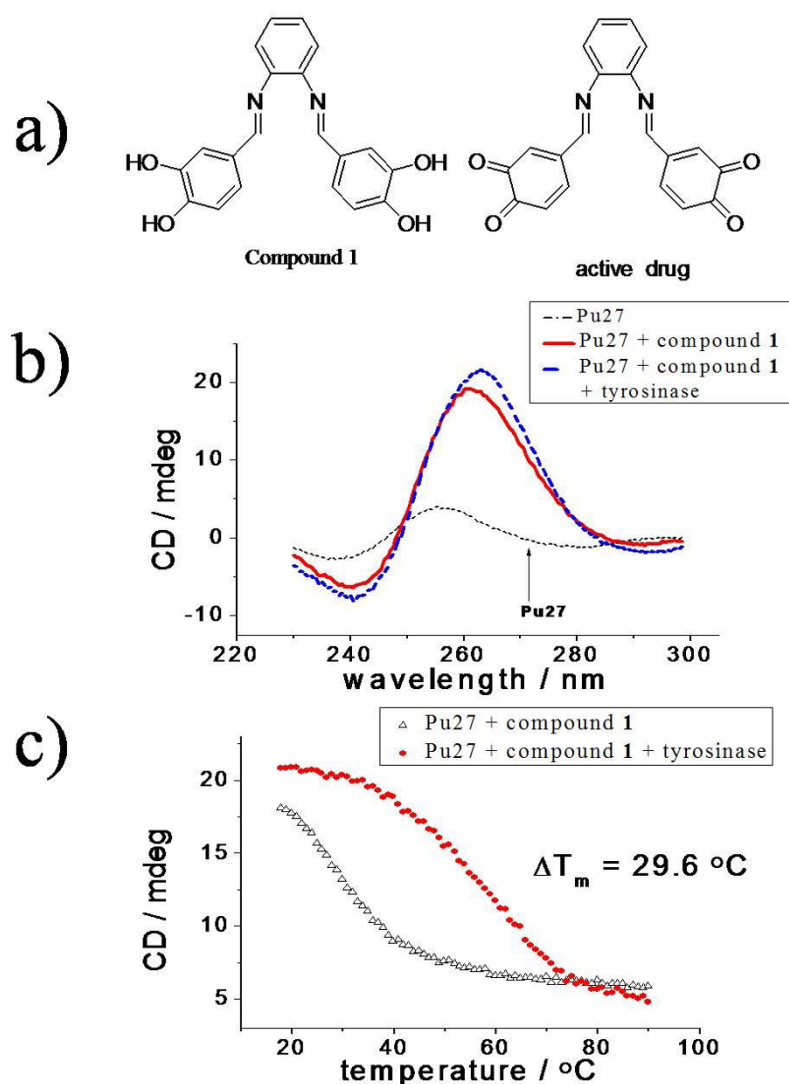

Figure 1 | a) Structures of compound **1** and *o*-quinone intermediate induced by tyrosinase. b) CD spectra of Pu27 DNA (20 µM) in 10 mM Tris/HCl buffer at pH 7.0 in the absence or presence of compound **1** ([compound **1**]/[DNA strand] = 5) without or with incubation with tyrosinase (400 Units) at 37°C for 1 h. c) CD melting of Pu27 DNA curves in the presence of compound **1** ([compound **1**]/[DNA strand] = 5) without or with incubation with tyrosinase (400 Units) at 37°C for 1 h, as monitored by the CD intensity at 265 nm. T_m values of the two G-quadruplexes are 30.3°C and 59.9°C, respectively.

out that the CD and ΔT_m test were performed to measure the pure cross-linking and/or stabilizing effect of the compound towards G-quadruplex, so that no cations were added in order to display our compound's pure effect without interference of cations. Meanwhile, we also tested the cases with addition of K$^+$ from 5 mM, 50 mM to 100 mM (SI Figure S4a–d). After screening the series of three compounds, compound **1** was used for further investigation. Beside Pu27 sequence, other typical G-rich sequences, such as c-kit-20, bcl-2–23, and human telomeric sequence were also screened to test sequence selectivity of compound **1**. It was indicated that compound **1** can induce G-quadruplex formation of c-kit-20 and bcl-2-23 sequences other than human telomeric sequence(SI Figure S11). Stability of G-quadruplex induced by compound **1** in c-kit-20 and bcl-2-23 were weaker than that in Pu27 (SI Figures S7–S10). Based on the above findings, compound **1** has a preference for Pu27 sequence, which has a maximum guanine density in the studied sequences. Thus, Pu27 sequence was used for further study.

Verification of G-quadruplex cross-linking *in vitro* through gel and MS analysis. Based on preliminary test, compound **1** was implied to be a good inducing and stabilizing agent of G-quadruplexes toward single strand G-rich sequences. However, for cross-linking, a covalent interaction, ΔT_m was supposed to be larger than 30°C. We suggested it could be resulted from incomplete cross-linking efficiency, since the unreacted strands in the system could lower the ΔT_m. Thus, we need to verify that compound **1** interact with DNA through cross-linking other than the common noncovalent manner. Denaturing gel analysis, which could exclude noncovalent interaction, was used to further prove our hypothesis. It was shown that with increasing concentrations of compound **1** upon oxidation by tyrosinase, new band behind original band appeared and its intensity gradually increased. That could be inferred to be the alkylated DNA (Figure 2a). Similarly, compounds **2** and **3** also interact with G-quadruplex sequences through cross-linking, just not that strong as compound **1** (SI Figures S2 and S6), which was consistent with previous CD and T_m analysis.

Since large amount of double strand DNA predominate in genome, one practical G-quadruplex cross-linking agent should have enough selectivity. To evaluate whether compound **1** exhibits selectivity for G-quadruplex over duplex DNA, a competition experiment was performed. A fixed amount of compound **1** was incubated with a fixed amount of 5′-end TAMRA-labeled Pu27 DNA in the presence of increasing concentrations of unlabeled Pu27 DNA or unlabeled ds-Pu27 DNA. As shown in Figure 2b, unlabeled Pu27 DNA was able to effectively compete for adduct formation (adduct amount decreased to less than 1% of the total labeled DNA in the presence of a 10-fold excess of the unlabeled Pu27 DNA competitor) (lane 7, Figure 2b). On the contrary, the alkylated adduct was only modestly affected with unlabeled ds-Pu27 competitor DNA (lane 8-12, Figure 2b). We quantified the alkylating ability of compound **1** in these competition conditions (SI Figure S12), in which with addition of 10 folds of double strands, the cross-linking percentage of G-quadruplex still maintained no change. This indicated the obvious preference of the compound **1** towards G-quadruplex over double strands. These data clearly demonstrated that compound **1** could specifically alkylate G-quadruplex DNA even in presence of large amount of duplex DNA.

Of note, in most common conditions, double strand DNA is the stable conformation, while single strand, G-quadruplex or some other secondary structures were supposed to be transient and in dynamic equilibrium with double stranded state when corresponding complementary sequence exist. Even in complex intracellular environment, most G-rich segments are in double strand region and the dissociation of duplexes is believed to happen only under the co-effect of helicase and/or some other proteins, thus most non-covalent-type G-quadruplex inducers or probes were not strong enough to function at those segments and maintain fixed G-quadruplex in common condition.[31] Taken the above into account, to affirm and evaluate the covalent cross-linking effect between our compound and G-quadruplex structure, we examined whether compound **1** could trap Pu27 sequence in the presence of its complementary strand (Pu27-c). As shown in Figure 3a, the G-quadruplex formed when both compound **1** and tyrosinase were present, and the quadruplex structure was stable enough even with addition of a 2-fold excess of Pu27-c. To further confirm the cross-linking of G-quadruplex, we extracted the DNA from the gel and subsequently tested its T_m. Figure 3b shows a parallel G-quadruplex structure and to our delight, its melting curve look like a flat line and did not drop, indicating the extremely high thermodynamic stability derived from the cross-linking G-quadruplex (Figure 3c). MALDI-TOF mass
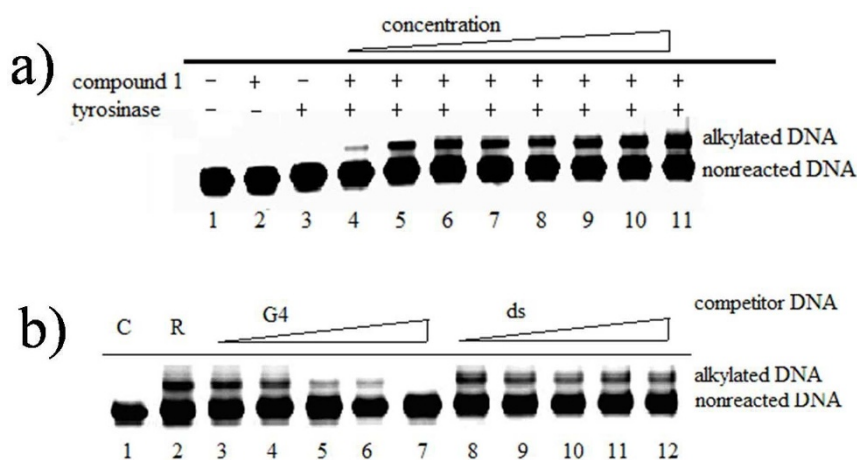

Figure 2 | a) Concentration dependence of compound **1** for tyrosinase oxidation DNA alkylating were incubated with 5′-end TAMRA labeled Pu27 DNA in 10 mM Tris/HCl buffer at pH 7.0 for 1 h at 37°C. The amounts of labeled DNA and tyrosinase were fixed as 10 pmol and 40 Units, respectively. Lane 1–3 are control lanes, lane 4–11 are increasing concentrations of compound **1** (10 μM, 20 μM, 50 μM, 0.1 mM, 0.2 mM, 0.5 mM, 1 mM, 2 mM) incubation with labeled Pu27 DNA and tyrosinase (40 Units). The alkylated oligo was separated from the nonreacted DNA by 20% denaturing polyacrylamide gel. b) Compound **1** (1 mM) was incubated with labeled Pu27 DNA (10 pmol) and tyrosinase (40 Units) in the presence of increasing molar ratios (0.5, 1, 2, 5, 10) of unlabeled Pu27 DNA(G4) or unlabeled ds-Pu27 DNA(double-stranded, ds) for 1 h at 37°C . The alkylated oligo was separated from the nonreacted DNA by 20% denaturing polyacrylamide gel. "C" refers to labeled Pu27 DNA (10 pmol) treated at 37°C for 1 h, and "R" refers to the Pu27-compound **1** reaction product in the absence of any competitor DNA.
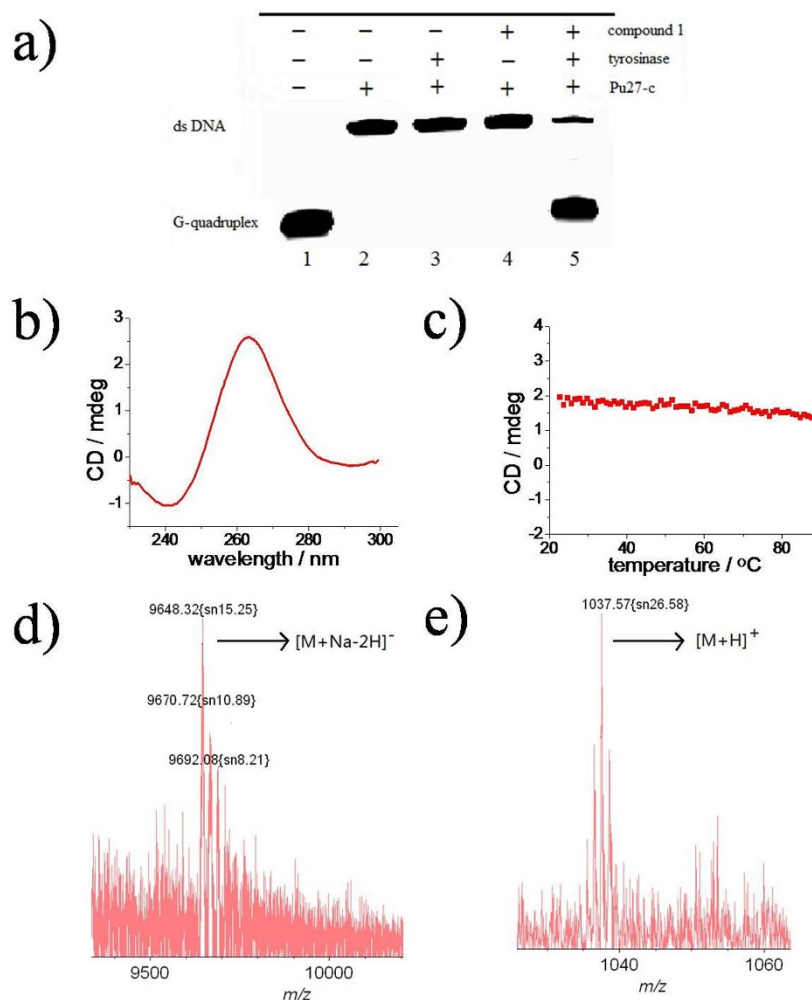
Figure 3 | a) Analysis of cross-linking G-quadruplex in the presence of complementary strand of Pu27 DNA (Pu27-c) in 10 mM Tris/HCl buffer at pH 7.0 and 10 mM KCl by 20% polyacrylamide gel. Lane 1: labeled Pu27 DNA (10 pmol) treated at 37°C for 1 h in 10 mM KCl; lane 2: control lane without tyrosinase and compound **1**; lane 3: control lane only with tyrosinase and without the compound **1**; lane 4: control lane only with compound **1** and without the tyrosinase; lane 5: compound **1** (2 mM) incubation with labeled Pu27 DNA (10 pmol) and tyrosinase (40 Uints) in the presence of excess of Pu27-c (20 pmol). b) CD spectrum for examining the structural features of the cross-linked complex which was extracted from gel. c) CD melting curves of the cross-linked complex as monitored by the CD intensity at 265 nm. d) MALDI-TOF MS spectrum of the cross-linked complex, calcd. $[M + Na-2H]^-$ 9649.77, found: 9648.32. e) MALDI-TOF MS spectrum of the cross-linked complex degradation products by DNase I, calcd. $[M + H]^+$ 1037.52, found: 1037.57.

spectrometry was also used to provide confirmation of quadruplex cross-linking. Molecular ion peak of the Pu27-compound **1** adduct was detected in MS spectrum upon analysis of gel extracted sample. These data indicated a ratio of compound **1**: DNA to be 1:1 (Figure 3d). Based on all the above evidences, it was pretty solid that compound **1** interact with Pu27 sequence through cross-linking and the cross-linking effect made the G-quadruplex fixed even under competition of its complementary sequence. It need to be pointed out that if complementary sequence added first and formed double strand with Pu27 precedingly, our compound could not induce G-quadruplex in this situation. This was consistent with previous results.

After confirmation of cross-linking, a more detailed study was conducted, such as the sites of cross-linking. Of note, the *o*-quinone intermediate tends to form covalent bonds with guanines[32]. To test whether alkylation sites of compound **1** were guanines, we employed DNase I to randomly digest the DNA[33] and then used MS to detect whether guanine-compound existed, since adduct of guanosine and compound **1** would remained intact upon digestion. As expected, the molecular ion peak of G-compound-G adduct was detected by MALDI-TOF MS (Figure 3e). These results clearly showed that the interaction between compound **1** and G-quadruplex is through *o*-quinone-mediated guanine cross-linking. Additionally, we performed the MBTH(3-methyl-2-benzothiazolinone hydrazone) color test to characterize the *o*-quinone-mediated oxidation mechanism in the study[29] (SI Figure S13). The dihydroxyphenyl groups were oxidized by tyrosinase into *o*-quinones, which could react with MBTH through a Michael reaction to form a dark-pink pigment. The formed *o*-quinone intermediate could further induce and cross-link G-quadruplex structure (SI Figure S14). These indicated that our compound **1** function through the active intermediate of *o*-quinone.

***In vivo* studies of compound 1 N,N′-bis(3,4-dihydroxybenzylidene)-1,2-diaminobenzene toward tumor cells.** In consideration of its good specificity toward G-quadruplex cross-linking even in presence of excess duplex DNA, we were encouraged to investigate its activities *in vivo*. Due to inducible oxidation by tyrosinase, B16-F1, the murine melanoma cell line with high levels of tyrosinase could serve as an excellent host in our study[34]. First we carried out the proliferation assay upon drug exposure (SI Figure S15). A

gradual dosage-dependent inhibition of cell proliferation was observed upon 6 days treatment with compound **1**. Then DNA damage by compound **1** in the B16-F1 cells was characterized through alkaline comet assay[35,36]. For negative control experiment, HeLa cells were used and nearly no DNA damage was observed in the even treated with higher concentrations of compound **1** (SI Figure S16a–f), which supported the tyrosinase-stimulated cross-linking mechanism *in vivo*.

To better understand the correlation between the demonstrated cytoxicity and the G-quadruplex cross-linking activity, we extracted the genomic DNA from 40 μM compound **1** treated B16-F1 cells. After treatment with DNAse I, a peak corresponding to the guanosine-compound-guanosine was observed by MALDI-TOF MS (SI Figure S19c). We were confident that the proliferation inhibition and DNA damage of B16-F1 cells were resulted from tyrosinase-induced G-quadruplex cross-linking by compound **1**. Then we studied *c-myc* gene expression upon incubation with compound **1**. Using the software of ChemDoc XRS+ imaging system (BioRad, USA), quantitative analysis was done. Detailedly, the strength ratio of c-Myc and β-actin level at each compound concentration was calculated, and then using the value of control sample as 100%, other inhibition percentage could be obtained as figure 4 showed. It was indicated that when compound **1** reached 50 μM, 3 day treatment could provide almost 80% *c-myc* gene expression inhibition, which was considered to be strong inhibitive effect. Based on the above results, we hypothesized that the selective inhibition of melanoma cell proliferation is due to G-quadruplex cross-linking within the oncogene promoters.

Affinity Enrichment of DNAs by compound 1a from Genomic DNA fragments of B16-F1 cells and Further Deep-sequencing Study. All the above demonstrated our compound could exactly reach the genome region in mammalian system and cross-link on guanosine. Then for purpose of characterizing corroborative G-quadruplex cross-linking mechanism toward whole genmome in chromosome of living cell systems was explored in depth. Biotin
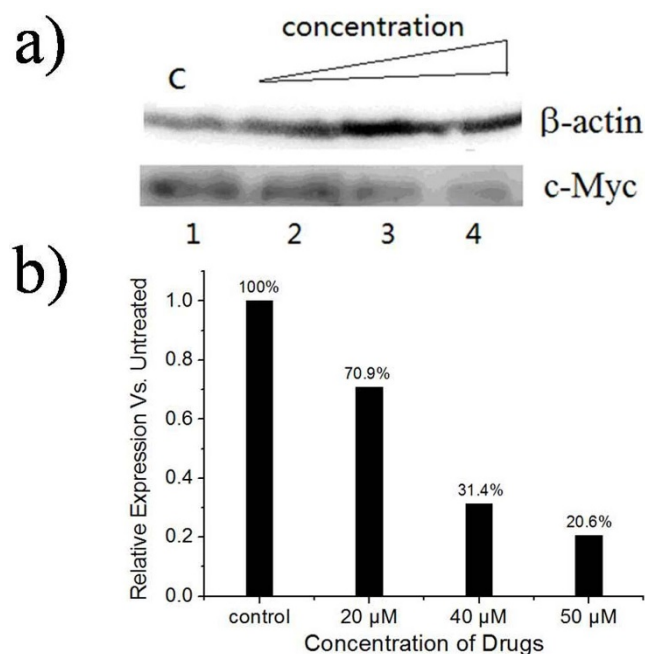
tag was widely used for affinity capture and/or enrichment of targets in a complicated environment for further detection[37,38]. Herein, to further study the explicit interaction between compound **1** and chromosome DNA of melanoma cells and confirm whether the DNA damage and downstream protein regulation were exactly referred to the compound's G-quadruplex-cross-linking, a biotin group was introduced on compound **1** through amide bonds to get compound **1a** (structure shown in Figure 5a), which could be used as hosts for isolation of guests DNA from cells (Figure 5a).

First we need to demonstrate the influence of the chemical modification on compound **1** for its recognition property, and compound **1a** need to retain a good cross-linking activity toward G-quadruplex DNAs. As results shown in Supplementary Information Figures S17 and S18, compound **1a** still reserve good activities in induction and cross-linking of Pu27 G-quadruplex DNA. Thus it was used in the following pull-down experiments. Compound **1a** in 40 μM was directly added to B16-F1 cells and treated for 3 days, then genomic DNA was extracted, with control experiment simultaneously. A sonication was used to fragment the chromosome DNA and the gel assay showed an average fragment length of 100 bp, which was suitable for ChIP-Seq assay (Figure 5b). Then hydrophilic streptavidin magnetic beads were used to isolate compound **1a** probed target DNAs, which was further detected by CD experiment. An evident peak representing the existing of a large quantity of parallel G-quadruplex structure could be observed (Figure 5c), while almost no DNA was captured by streptavidin-coated magnetic beads in the control sample without interference of compound **1a**. A huge difference has been demonstrated from the circular dichroism spectra.

To further illustrate the exact sequences targeted by compound **1a** and their locations on chromosome, we employed a high-throughput

Figure 4 | Western Blot was used to determine the expression of *c-myc* gene in the B16-F1 cells treated with compound 1. The cells were treated with medium (Line 1) and increasing concentrations of compound **1**(20, 40 and 50 μM) for 3 days, and the total protein was extracted and subjected to Western Blot for c-Myc, and β-actin (control).
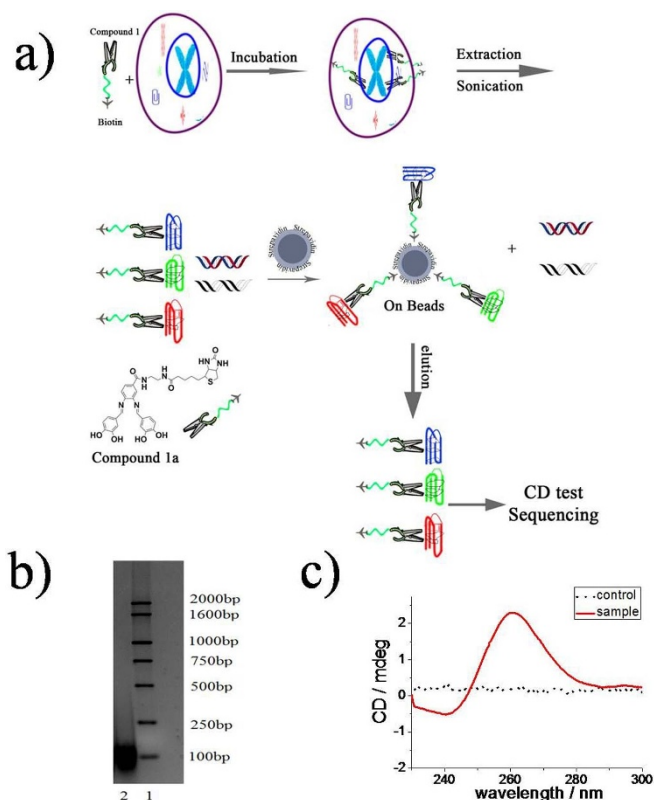
Figure 5 | a) Experimental design for detecting formation of G-quadruplexes and sequencing in oncogene promoters in B16-F1 cells. b)1.5% agarose gel. Lane 1: DNA ladder and lane 2: approximately 100 bp sonicated in genomic DNA. c) CD spectrum of genomic DNA short fragments which was extracted from B16-F1 cells in the absence or presence of compound **1a** (40 μM) for 3d.

deep sequencing technique to investigate small-molecule–induced DNA damage identifies alternative DNA structures in human genes[7]. The affinity purified DNA fragments by compound **1a** from B16-F1 cells was detached from the strepavidin beads, and they were submitted for high throughput sequencing. As expected, 120 putative G-quadruplex–forming sequences (PQSs) were identified from the 1294 sequences which located in the promoter regions(SI Table S1), although not all sequences from Chip-Seq are quadruplex–forming sequences. From the sequencing studies, it was shown that our compound functioned at G-rich sequences in the gene promoter regions, and by cross linking, the sequences could be pulled down. To demonstrate that quadruplex–forming sequences from Chip-Seq were undoubtedly G-quadruplex forming sequences, other than some ones containing random guanosine sites, we synthesized the 120 putative G-quadruplex–forming sequences (PQSs)[39,40] and analyzed them with CD. Among the 120 PQSs, we observed that all the sequences adopted stable folded G-quadruplex structures. Each individual PQS which was incubated with potassium chloride buffer displays signals at 265 nm and/or 295 nm characteristic of parallel G-quadruplex conformations and/or antiparallel conformations, respectively. The parallel structure occupy the vast majority of G-quadruplex conformations, several spectra exhibit both signals, demonstrating the presence of a mixture of conformers or the presence of mixed-type parallel/antiparallel conformation (SI Figure S20). After incubated with compound **1a** and tyrosinase in potassium chloride containing buffer, the majority of the sequence's G-quadruplex structure has not changed except for sequences 4, 32, 42, 50, 53, 58, 74. For these G-rich sequences, after addition with compound **1a** and tyrosinase, the conversion from antiparallel to parallel G-quadruplex structure can be observed (SI Figure S20–4, 32, 42, 50, 53, 58, 74), which showed that compound tend to induce parallel G-quadruplex structures for these G-rich sequences. The cases in which that when potassium absent, only under effect of compound and tyrosinase were also tested, giving no typical G-quadruplex peaks. This indicated that the compound could not induce G-quadruplex alone, so that it made us to deduce the G-quadruplex structure enriched and pulled down from genome was preformed and existed originally, just be captured by our compound.

Beside CD test, UV thermal difference spectrum (TDS)[41] and NMR were further used to affirm the quadruplex structure. The UV thermal difference spectrum (TDS) was considered as the finger print of different DNA secondary structure, in which G-quadruplex structure's typical TDS factor was defined to above 4 or below 2. Since too many sequences figured out by Chip-Seq, we tried to choose some typical ones (Sequence 1, 3, 19, 65, 97, 119) for save of expense and time (Sequences in Table S1). The results were consistent with literature report, in which the parallel structure with TDS above 4 was the dominant. (Figure S21).

On the other hand, the NMR test also confirmed G-quadruplex structure. The NMR was firstly done in 100 mM K$^+$ containing tris-buffer condition, respectively with and without compound + tyrosinaes system.^{1}H spectrum were collected and typical peaks representing imino hydrogen around the region of 10–12ppm were visualized, which could be a solid evidence of G-quadruplex structure (Figure S22). The NMR was also performed in *X. laevis* oocytes extracts, a condition very close to the real in-cell environment[42]. To our success, imino region of hydrogen representing G-quadruplex was also visualized (Figure S23). These results indicated that the PQSs figured out through Chip-Seq could indeedly have the strong potential to form G-quadruplex structure and be further stabilized as parallel structure by our compound both *in vitro* and *in vivo*.

Mimic of intracellular conditions using long chain DNA for G-quadruplex cross-linking studies. So far as we know, few report had trapped G-quadruplex structure *in vivo*, herein, using our G-quadruplex cross-linking agent, we successfully captured this structure, which could be a more direct evidence of existence of G-quadruplex. To further affirm our proclaimation and exclude the possibility of compound's induction effect, we used PEG-containing system and long DNA duplex to simulate the genome in molecular crowd effect in cellular condition, then the compound's behaviour was examined through gel analysis. Of note, PEG and KCl solution was a commonly used condition that could simulate molecular crowd effect of cellular environment. In this condition stable double strand could open and G-quadruplex forming sequence could form stable quadruplex structure[43]. Herein, we used this condition as literature reported to run gel for characterization of the G-quadruplex cross-linking behaviour. First, in 40% PEG containing gel and 100 mM KCl the 124 bp long double strand DNA *c-myc* and its complementary sequence with TAMRA labelled *c-myc*-c converted to G-quadruplex which shows a much looser structure (lane 2, Figure 6a), which was consistent with literature. In the absence of PEG, the long dsDNA formed a well-matched duplex structure (lane 1, Figure 6a). After incubation with compound **1a** and tyrosinase, the bands became a little bit more retard, which could be attributed to compound's covalently binding to DNA. Then, we tried our compound alone. Without PEG and KCl, *c-myc* and *c-myc*-c formed a well-matched duplex structure (lane 2, Figure 6b), in the presence of PEG but without KCl (lane 3, Figure 6b), there is no looser structure existence. It was also found that no G-quadruplex formed even our compound was at high concentration up to 2 mM. That was to say, our compound alone was not potent enough to convert G-rich sequence in stable long DNA duplex, even in molecular crowd situation, to G-quadruplex structure. Meanwhile, Ficol 70 was also used for purpose of avoiding PEG condition's potential one-sideness[42]. We first anneal to prepare double strands, then gel was ran. Results were similar to that before (Figure S24).
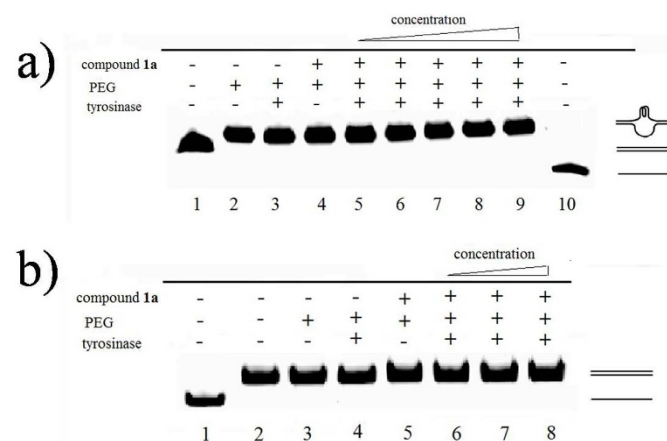

Figure 6 | G-quadruplex formation in long dsDNA carrying G-quadruplex-forming sequence from the *c-myc* gene examined by native gel electrophoresis. a) DNA samples were loaded on 8% polyacrylamide gel containing 150 mM KCl, 40% (w/v) PEG 200 and electrophoresed at in 1× TBE buffer containing 150 mM KCl. Lane 1:*c-myc* and *c-myc*-c without PEG; Lane 2: *c-myc* and *c-myc*-c with PEG; Lane 3 and 4 are control lanes, *c-myc* and *c-myc*-c incubation with compound **1a** (2 mM) or tyrosinase (40 Units); lane 5-10 are increasing concentrations of compound **1a** (100 μM, 200 μM, 500 μM, 1 mM, 2 mM) incubation with *c-myc* and *c-myc*-c in the presence of tyrosinase (40 Units); Lane 10: *c-myc*-c only. All samples contain 150 mM KCl. b) DNA samples were loaded on 8% polyacrylamide gel containing 40% (w/v) PEG 200 and electrophoresed at in 1× TBE buffer. Lane 1: *c-myc*-c only; Lane 2: *c-myc* and *c-myc*-c without PEG; Lane 3: *c-myc* and *c-myc*-c with PEG; Lane 4 and 5 are control lanes, *c-myc* and *c-myc*-c incubation with compound **1a** (2 mM) or tyrosinase (40 Units); lane 6-8 are increasing concentrations of compound **1a** (500 μM, 1 mM, 2 mM) incubation with *c-myc* and *c-myc*-c in the presence of tyrosinase (40 Units), All samples were free off KCl.

Besides gel analysis, NMR was also use to monitor the G-quadruplex in mimetic cellular condition. Since the 124 bp long chain was too expensive to afford for NMR's amount, here, two sequences from the 120 G-rich sequences (Sequence 78, 119)same as that chosen in the above were tested in absence of compound **1a** (here, the sequences were as pre-annealed double strands with their complementary sequences). To our success, imino region of hydrogen representing G-quadruplex was visualized around 10–12 ppm. These NMR results indicated that under the *X. laevis* oocytes mimetic macromolecular and ionic conditions, without compound's function, the double strands could be detached and the G-rich segment could form quadruplex structure. When adding compound **1a** and tyrosinase, the peaks representing imino hydrogens maintained (Figure S25). Although 124 bp sequence did not have NMR, we suggest their behavior and mechanism in the intracellular conditions should be consistent, the main difference between genome in real condition and synthetic long DNA in mimetic condition could be that the copy number of natural G-rich sequences in genome was much less than the synthetic ones, that's why it was very difficult to be directly visualized that through NMR or some other methods at in situ level.

For further exclusion of compound's function, For more precise quantitative analysis, we measured the cellular uptake of the compound through UV absorption. We first did work curve in cell culture medium and then calculated the concentration of compound treated medium. As shown in Figure S27 the concentration of compound in medium after 1 day treatment was 28 μM, after 2 days and 3 days, it was generally the same. That was to say, the compound uptake by cells was quite a small amount about only 10 μM level, which exactly could not induce G-quadruplex alone. So we suggest that cells re-exposure to compound **1a** would not induce quadruplex. Taking all the above into account, we suggest the captured G-quadruplex structure in preceding experiments was the preformed structure other than induced by compound, just that it's transient and difficult to be visualized in common condition.

Discussion

Known G-quadruplex inducers or stabilizers, such as porphyrins and so on, commonly function through noncovalent binding. Due to modest stability, the noncovalently induced G-quadruplex was always vulnerable, especially in real complex living system, which largely limited their practicality. In our work, a novel strategy based on tyrosinase oxidation mediated G-quadruplex cross-linking has been developed. Three catechol derivatives have been designed and synthesized, and compound **1** was found to be best in G-quadruplex cross-linking. Based on the according study, we suggested compound **1**′s conformation fit best with G-quadruplex structure compared to the other two. What we want to explain was that the reason why we did those in absence of potassium before was that the Pu27 sequence in potassium solution had very high T_m, even at 5 mM K^+, the T_m was more than 70°C. Thus, if at physiological potassium concentration about 100 mM, the ΔT_m between our compound treated cases and control one would absolutely be overwhelmed, and the effect of cross-linking could not be measured. The preliminary test provided about 30°C ΔT_m under compound **1** with tyrosinase compared to the control one, which in the mass showed good cross-linking and stabilizing ability and that is the premise encouraged us to do more.

The exact cross-linking mode was confirmed by double strand-competition and complementary single-strand competition expreriments. Gel analysis showed exclusive results that compound **1** could cross-link G-quadruplex DNAs even in presence of excess single complementary strands with a T_m larger than 90°C. MS assay also demonstrated molecular ion peaks corresponding to complex of compound-Pu27 or dimmer of G-compound-G. These undoubtedly proved the exact cross-linking interaction and the binding sites of guanines. MBTH color test characterized the *o*-quinone mediated

oxidation, which demonstrated existance of *o*-quinone intermediate in the cross-linking process.

The strong cross-linking potency and considerable G-rich sequence preference over double strands and complementary strands provided potential for further applications in antitumor drug development. Since its tyrosinase-mediated oxidation mechanism, melanoma cell with tyrosinase overexpressed, was an ideal host. Long term cell preliferation inhibition effect other than acute toxicity indicated that cross-linking mediated DNA damage could be major pathway of our compound. Comet assays showed obvious DNA damage with increasing dosage of compound. MS assay afforded advantageous imply, which trapped molecular ion peak of dimer of G-compound-G from genomic DNA of the compound treated cells. Gene expression regulation was also examined, since that it's an important criterion for evaluation of an antitumor candidate. Because of our compound's preference of Pu27 sequence over other G-rich sequences, it was chosen as the primary target in corresponding downstream protein expression assay. Through Western blot, obvious c-Myc down-regulation (about 80% *c-myc* gene expression inhibition)was observed. These assays together demonstrated that our compound could enter the cell nuclei and interact with DNA through cross-linking to cause transcriptional inhibition of downstream oncogenes.

Good biological activities of compound **1** encouraged us to further elucidate its acting mechanism. To determinate distribution of function sites and affirm G-quadruplex targeting and cross-linking ability, the most direct way was to modify the compound with an isolating tag for use of enrichment of compound-DNA complex from complex cellular system. We combined a biotin group with compound **1** to form compound **1a**. After verification of compound **1a**'s G-quadruplex cross-linking ability *in vitro*, we replace compound **1** with compound **1a** for following cell treatment. Extracted genomic DNA from compound **1a** treated living cells was sonicated and purified through biotin-strepavidin affinity. CD test showed a direct evidence of G-quadruplex involving interaction between compound **1a** and chromosomal DNAs. It need to be emphasized that the CD spectra in Figure 5c was the real sample from compound treated cells, before extraction, extra compound in medium was washed away, and after extraction, no extra K^+ was added, the sample was directly poured to do CD test. This indicated that the G-quadruplex was enriched from the genomic DNA, maintained and fixed by compound, rather than induced by K^+ or re-exposure under compound afterwards, otherwise, the quadruplex structure enriched from real sample would break down because we did not add extra cations or compounds afterwards. We considered that compound **1a** could exactly target G-quadruplex formation sequences and completely block further biological processing, then subsequently capture the G-quadruplex structures from living cells. Thus, we were encouraged to pursue more corresponding information. The sequencing analysis showed about 120 putative quadruplex-forming sequences (PQS) from the biologically enriched DNA sample. It was shown that the sequences figured out in Chip-Seq generally had very high amount of guanine. Of note, under physiological condition, parallel G-quadruplex structure was the most potential structure of such G-rich sequences. Thus, we investigated G-quadruplex structure of them *in vitro* from the opposite side. We used CD, UV and NMR for structural characterization, which all provided affirmative evidence of that the sequences enriched had strong propensity to form the quadruplex secondary structure. Detailedly, under physiological condition, most sequences among the 120 PQSs formed parallel conformation and could be further enhanced after addition of our compound **1a** and tyrosinase. When in absence of K^+, for most G-rich sequences, the compound **1a** with tyrosinase could not induce or stabilize G-quadruplex even at 100 μM. This also further verified our viewpoint that at micro molar level, for most G-rich sequences, our compound could only block and trap the preformed G-quadruplexs,

instead of directly inducing them. That was to say, the enriched sequences and visualized quadruplex structures were the preformed and existed ones *in vivo*. All the results strongly demonstrated that compound **1a** could act on G-rich sequences in gene promoter regions through the whole genome and further fix parallel G-quadruplex through cross-linking. Another point we would like to explain was that about the percentage of TSS from deep sequencing. The percentage 10% was calculated on the statistically level, our focus herein was on G-quadruplex's existence *in vivo* and our compound's essential role in capture of the originally existed but transient G-quadruplex structure. Through deep sequencing, what we could claim was that our compound could target and capture the existed G-quadruplex *in vivo*.

To further ascertain G-quadruplex's preformation *in vivo* and exclude compound's effect, we chose 124 bp long chain in several commonly used mimetic conditions, including Ficoll 70, PEG and *X. laevis* oocytes, to simulate genome in real living cells, because absolute real in-cell process was very hard to access. Through NMR and gel assay in the mimetic biomacromolecular environment, G-quadruplex structure in synthetic DNA sequences could be observed without any compound's function, which proved that the genomic DNA could have the propensity to be unlink under effect of various enzyme and biomacromolecules in intracellular environment and some G-rich segments could form transient quadruplexs due to physiological high ionic condition.

Our synthesized bis(catechol) Schiff base derivative has the ability to capture the G-quadruplex structures from living cells unlike other previously published molecular probes. Previous G-quadruplex inducers or stabilizers only showed good potency in simple *in vitro* systems, however, in complex intracellular conditions where there are large amounts of double-stranded DNA and the histones protect the chromosome, the molecule probes' interactions toward certain short G-rich segments were rendered ineffective. Because of these problems, transient G-quadruplexes could not be fixed. However, our cross-linking agent is efficient and traps transient quadruplex structures through cross-linking.

In conclusion, a bis(catechol) Schiff-base derivative that functions as a highly selective and inducible G-quadruplex cross-linking agent was designed and synthesized. Through an integrated approach that included CD, MALDI-TOF MS, enzymatic assay, gel and sequencing affinity-enriched fragments from genomic DNA, this derivative was found to cross-link G-quadruplexes in promoter region of oncogenes. Importantly, the biotinylated derivative compound **1a** was able to capture G-quadruplex DNA *in vivo* through a cross-linking way. Mimic of cellular condition by PEG also gave us supportive information that G-quadruplex structure exists in cells. To the best of our knowledge, this was the first example of G-quadruplex cross-linking agent that selectively targeted G-rich sequences in the promoter regions of the oncogenes *in vivo* and captured the G-quadruplex DNA structures from *in vivo* systems.

Methods

General methods. Oligomers labelled with carboxytetramethylrhodamine (TAMRA) were purchased from TaKaRa Biotech (China). Other oligomers used in this study were purchased from Invitrogen (China). CD relating assays were all carried out on a Chirascan CD spectrometer (Applied Photophysics, UK) equipped with a Peltier temperature controller. Polyacrylamide gel electrophoresis products were finally scanned with Pharos FX Molecular imager (Bio-Rad, USA) operated in the fluorescence mode. MALDI-TOF-MS spectra were collected with an Axima-TOF² mass spectrometry (Shimadzu, Japan). Mass spectra were performed in negative ion mode or positive ion reflector mode, respectively, with 3-hydroxypicolinic acid as matrix.

DNase I degradation. DNase I was purchased from Fermentas (Canada). DNA alkylated complex was incubated with DNase I (10 Units/mL) in 10 mM Tris-HCl reaction buffer (pH 7.5) and 2.5 mM $MgCl_2$ at 37°C for 1 h, and then 5 mM EDTA was added and incubated at 65°C for 10 min. Reaction products were monitored by MALDI-TOF-MS.

Alkaline comet assay for B16-F1 and Hela cells. Base comet assay was performed following standard method. Drug-exposed samples together with control one were collected and suspended at 10^6/mL after 72 h treating. The first layer of normal agarose gel was performed by adding 100 μL 0.5% agarose on to slide, and then a cellslip was covered. After it became solid, remove the cellslip, 100 μL low melting agarose gel containing 10 μL cell solution above was add to the first layer gel. Another cellslip was covered. After solidifying, remove cellslip, then lysis of cells was followed by immerging the cell-containing gel in lysis buffer for 4°C 2 h. After washing, use basic buffer to denature. The last step was electrophoresis, which was at 5 V/cm for 20 min. Confocal microscopy was used to visualize the tailing after stained with PI. One-photon confocal images were taken by NOL-LSM 710 (Carl Zeiss, Germany).

Western Blot assay. Cells were treated same as the above for 72 h with increasing concentrations of compound **1** (20, 40 and 50 μM), and total proteins were extracted by lysis buffer RIPA. Then use BCA protein assay to measure concentrations of each sample. After electrophoresis on 10% PAGE-SDS gel and blotting, anti-*c-myc* antibody and corresponding secondary antibody was used to detect target protein through standard washing and incubating steps.

Cell uptake of compound 1a. Cells were grown in T75 tissue culture flask at 10^5/mL. Compound **1a** in DMEM with a concentration of 40 μM was added 24 h after seeding. Concentration of compound remained in medium was measured by UV every day for 3 days overall, according to standard job curve done in the medium of the control sample.

Affinity enrichment of G-quadruplex fragments from cells. Cells were grown in T25 tissue culture flask at 10^5/mL. Compound **1a** in DMEM with a concentration of 40 μM was added 24 h after seeding. After 3 days' incubation, genomic DNA was exacted by TIANamp Genomic DNA kit (TIANGEN biotech, China) and isolated genomic DNA from B16-F1 cells was sheared with the SB-5200 DTD sonicator (300 W, Ningbo scientz biotechnology, China) for 60 min at high power with a pulse of 30 sec on/30 sec off, to yield the fragment sizes of approximately 100 bp (Figure 5b). 100 μL of sonicated DNA was incubated with 10 μL hydrophilic streptavidin magnetic beads (New England biolabs, UK) for 1 h at 37°C. The magnetic beads was then incubated with 10 mM EDTA and 95% formamide (2.5 μL 0.2 M EDTA pH = 8.0 and 47.5 μL formamide) at 90°C for 10 min. Collected DNA fractions were purified by loading onto a illustra™ NAP™-5 Columns Sephadex™ G-25 DNA Grand (GE Healthcare, UK) equilibrated with double distilled water. The purified DNA fragments were directly used for CD and sequencing experiments. Sequencing was collaborated with Invitrogen in China, and data were collected using the Ion Torrent Genome Analyzer (Invitrogen, China) according to ChIP-Seq protocol. Putative quadruplex sequences (PQS) were proposed the following Folding Rule: A sequence of the form $d(G_3 + N_{1-7}G_3 + N_{1-7}G_3 + N_{1-7}G_3 +)$ will fold into a quadruplex under near-physiological conditions.

Long chain DNA for G-quadruplex cross-linking studies by native gel electrophoresis in PEG and Ficoll 70 condition. DNA samples were made in 10 mM Tris–HCl (pH 7.0) buffer containing the indicated concentration of KCl and PEG 200 or Ficoll 70, heated at 95°C for 5 min and then cooled down to room temperature at a rate of 0.02°C per second. Then incubated with indicated concentration of compound and tyrosinase at 37°C for 1 h. The DNA samples were loaded with the indicated concentration of PEG 200 or Ficoll 70 and electrophoresed at 4°C with 8 V/cm in 1× TBE buffer with or without KCl.

Preparation of X. laevis oocytes extracts. *X. laevis* were cultured and the oocytes were collected by injection of progesterone to the *X. laevis* 1 day in advance. The oocytes were transferred to a petri dish on iced and washed with Ori buffer [5 mM HEPES (pH = 7.6), 110 mM NaCl, 5 mM KCl, 2 mM $CaCl_2$, 1 mM $MgCl_2$], using 2% solution of cystine, the membrane of ooctyes could be removed. Then they were transferred into another petri dish prefilled with 10 mL ice cold buffer mimicking the oocyte salt environment (25 mM HEPES pH 7.5, 10.5 mM NaCl, 110 mM KCl, 130 nM $CaCl_2$, 1 mM $MgCl_2$,). Afterward oocytes were transferred into an eppendorf tube and mechanically crushed, then insoluble fractions were removed by centrifugation at 16 000 g for 20 min. The supernatant was transferred into an eppendorf tube and heated to 95°C for 10 min. Precipitated proteins were removed by centrifugation at 16 000 g for 10 min. The supernatant (250 μL) was used for NMR measurements.

1. Zahler, A. M., Williamson, J. R., Cech, T. R. & Prescott, D. M. Inhibition of telomerase by G-quartet DMA structures. *Nature.* **350**, 718–720 (1991).
2. Rodriguez, R., Pantos, G. D., Goncalves, D. P. N., Sanders, J. K. M. & Balasubramanian, S. Ligand-driven G-quadruplex conformational switching by using an unusual mode of interaction. *Angew. Chem. Int. Ed.* **46**, 5405–5407 (2007).
3. Wang, X. L., Huang, J., Zhou, Y. Y., Yan, S. Y., Weng, X. C., Wu, X. J., Deng, M. G. & Zhou, X. Conformational switching of G-quadruplex DNA by photoregulation. *Angew. Chem. Int. Ed.* **49**, 5305–5309 (2010).
4. Yang, P., Cian, A. D., Teulade-Fichou, M. P., Mergny, J. L. & Monchaud, D. Engineering bisquinolinium/thiazole orange conjugates for fluorescent sensing of G-quadruplex DNA. *Angew. Chem. Int. Ed.* **48**, 2188–2191 (2009).

5. Yue, X., Schwartz, J. C., Chu, Y., Younger, S. T., Gagnon, K. T., Elbashir, S., Janowski, B. A. & Corey, D. R. Transcriptional regulation by small RNAs at sequences downstream from 3′ gene termini. *Nat. Chem. Biol.* **6**, 621–629 (2010).

6. Ito, K., Go, S., Komiyama, M. & Xu, Y. Inhibition of translation by small RNA-stabilized mRNA structures in human cells. *J. Am. Chem. Soc.* **133**, 19153–19159 (2011).

7. Rodriguez, R., Miller, K. M., Forment, J. V., Bradshaw, C. R., Nikan, M., Britton, S., Oelschlaege, T., Xhemalce, B., Balasubramanian, S. & Jackson, S. P. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **8**, 301–310 (2012).

8. Ou, T. M., Lu, Y. J., Tan, J. H., Huang, Z. S., Wong, K. Y. & Gu, L. Q. G-quadruplexes: targets in anticancer drug design. *ChemMedChem.* **3**, 690–713 (2008).

9. Bugaut, A., Jantos, K., Wietor, J. L., Rodriguez, R., Sanders, J. K. M. & Balasubramanian, S. Exploring the differential recognition of DNA G-quadruplex targets by small molecules using dynamic combinatorial chemistry. *Angew. Chem. Int. Ed.* **47**, 2677–2680 (2008).

10. Balasubramanian, S., Hurley, L. H. & Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nature Rev. Drug Discovery.* **10**, 261–275 (2011).

11. Slamon, D. J., deKernion, J. B., Verma, I. M. & Cline, M. J. Expression of cellular oncogenes in human malignancies. *Science.* **224**, 256–262 (1984).

12. Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J. & Hogan, M. E. Site-specific oligonucleotide binding represses transcription of the human c-myc gene in vitro. *Science.* **241**, 456–459 (1988).

13. Siebenlist, U., Hennighausen, L., Battey, J. & Leder, P. Chromatin structure and protein binding in the putative regulatory region of the c-myc gene in burkitt lymphoma. *Cell.* **37**, 381–391 (1984).

14. Davis, T. L., Firulli, A. B. & Kinniburgh, A. J. Ribonucleoprotein and protein factors bind to an H-DNA-forming c-myc DNA element: possible regulators of the c-myc gene. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9682–9686 (1989).

15. Phan, A. T., Modi, Y. S. & Patel, D. J. Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.* **126**, 8710–8716 (2004);

16. Ou, T. M., Lin, J., Lu, Y. J., Hou, J. Q., Tan, J. H., Chen, S. H., Li, Z., Li, Y. P., Li, D., Gu, L. Q. & Huang, Z. S. Inhibition of cell proliferation by quindoline derivative (SYUIQ-05) through its preferential interaction with *c-myc* promoter G-quadruplex. *J. Med. Chem.* **54**, 5671–5679 (2011).

17. Grand, C. L., Han, H., Munoz, R. M., Weitman, S., Von Hoff, D. D., Hurley, L. H. & Bearss, D. J. The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Mol. Cancer Ther.* **1**, 565–573 (2002).

18. Seenisamy, J., Bashyam, S., Gokhale, V., Vankayalapati, H., Sun, D., Siddiqui-Jain, A., Streiner, N., Shin-ya, K., White, E., Wilson, W. D. & Hurley, L. H. Design and Synthesis of an Expanded Porphyrin That Has Selectivity for the c-MYC G-Quadruplex Structure. *J. Am. Chem. Soc.* **127**, 2944–2959 (2005).

19. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H. J. & Plückthun, A. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8572–8577 (2001).

20. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **5**, 182–186 (2013).

21. Xu, Y., Suzuki, Y. & M. Komiyama. Click chemistry for the identification of G-quadruplex structures: discovery of a DNA-RNA G-quadruplex. *Angew. Chem. Int. Ed.* **48**, 3281–3284 (2009).

22. Antonio, M. D., Doria, F., Richter, S. N., Bertipaglia, C., Mella, M., Sissi, C., Palumbo, M. & Freccero, M. Quinone methides tethered to naphthalene diimides as selective G-quadruplex alkylating agents. *J. Am. Chem. Soc.* **131**, 13132–13141 (2009).

23. Nadai, M., Doria, F., Antonio, M. D., Sattin, G., Germani, L., Percivalle, C., Palumbo, M., Richter, S. N. & Freccero, M. Naphthalene diimide scaffolds with dual reversible and covalent interaction properties towards G-quadruplex. *Biochimie* **93**, 1328–1340 (2011).

24. Doria, F., Nadai, M., Folini, M., Antonio, M. D., Germani, L., Percivalle, C., Sissi, C., Zaffaroni, N., Alcaro, S., Artese, A., Richter, S. N. & Freccero, M. Hybrid ligand-alkylating agents targeting telomeric G-quadruplex strucyure. *Org. Biomol. Chem.* **10**, 2798–2806 (2012).

25. Bai, M. H., Huang, J., Zheng, X. L., Song, Z. B., Tang, M. R., Mao, W. X., Yuan, L. B., Wu, J., Weng, X. C. & Zhou, X. Highly selective suppression of melanoma cells by inducible DNA cross-linking agents: bis(catechol) derivatives. *J. Am. Chem. Soc.* **132**, 15321–15327 (2010).

26. Celli, C. M. & Jaiswal, A. K. Role of GRP58 in mitomycin C-induced DNA Cross-Linking. *Cancer Res.* **63**, 6016–6025 (2003).

27. Weng, X. C., Ren, L. G., Weng, L. W., Zhu, S. G., Huang, J., Zhou, X. & Weng, L. H. Synthesis and biological studies of inducible DNA cross-linking agents. *Angew. Chem. Int. Ed.* **46**, 8020–8023 (2007).

28. Ghani, U. & Ullah, N. New potent inhibitors of tyrosinase: Novel clues to binding of 1,3,4-thiadiazole-2(3H)-thiones, 1,3,4-oxadiazole-2(3H) -thiones,4-amino-1,2,4-triazole-5(4H)-thiones, and substituted hydrazidesto the dicopper active site. *Bioorg. Med. Chem.* **18**, 4042–4048 (2010).

29. Ye, Y., Chou, G. X., Mu, D. D., Wang, H., Chu, J. H., Leung, A. K. M., Fong, W. F. & Yu, Z. L. Screening of Chinese herbal medicines for antityrosinase activity in a cell free system and B16 cells. *J. Ethnopharmacol.* **129**, 387–390 (2010).

30. Salavati-Niasari, M., Davara, F. & Bazarganipourb, M. Synthesis, characterization and catalytic oxidation of *para*-xylene by a manganese(III) Schiff base complex on functionalized multi-wall carbon nanotubes (MWNTs). *Dalton Trans.* **39**, 7330–7337 (2010).

31. Xu, L., Wu, W. X., Ding, J., Feng, S., Xing, X. W., Deng, M. G. & Zhou, X. A pyridyl carboxamide molecule selectively stabilizes DNA G-quadruplex and regulates duplex–quadruplex competition. *RSC Adv.* **2**, 894–899 (2012).

32. Wang, P., Liu, R. P., Wu, X. J., Ma, H. J., Cao, X. P., Zhou, P., Wang, P., Liu, R. P., Wu, X. J., Ma, H. J., Cao, X. P., Zhou, P., J, Y., Zhang, Weng, X. C., Zhang, X. L., Qi, J., Zhou, X. & Weng, L. H. A potent, water-soluble and photoinducible DNA cross-linking agent. *J. Am. Chem. Soc.* **125**, 1116–1117 (2003).

33. Anderson, S. Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Res.* **9**, 3015–3027 (1981).

34. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11593–11598 (2002).

35. Kudugunti, S. K., Vad, N. M., Whiteside, A. J., Naik, B. U., Yusuf, M. A., Srivenugopal, K. S. & Moridani, M. Y. Biochemical mechanism of caffeic acid phenylethyl ester (CAPE) selective toxicity towards melanoma cell lines. *Chem. Biol. Interact.* **188**, 1–14 (2010).

36. Cederberg, H., Henriksson, J. & Binderup, M. L. DNA damage detected by the alkaline comet assay in the liver of mice after oral administration of tetrachloroethylene. *Mutagenesis.* **25**, 133–138 (2010).

37. Müller, S., Kumari, S., Rodriguez, R. & Balasubramanian, S. Small-molecule-mediated G-quadruplex isolation from human cells. *Nat.Chem.* **2**, 1095–1098 (2010).

38. Song, C. X., Szulwach, K. E., Fu, Y., Dai, Q., Yi, C. Q., Li, X. K., Li, Y. J., Chen, C. H., Zhang, W., Jian, X., Wang, J., Zhang, L., Looney, T. J., Zhang, B. C., Godley, L. A., Hicks, L. M., Lahn, B. T., Jin, P. & He, C. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol.* **29**, 68–72 (2011).

39. Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Research.* **33**, 2908–2916 (2005).

40. Drygin, D. Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer Res.* **69**, 7653–7661 (2009).

41. Karsisiotis, A. I., Hessari, N. M., Novellino, E., Spada, G. P., Randazzo, A. & Silva, M. W. Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angew. Chem. Int. Ed.* **50**, 10645–10648 (2011).

42. Hänsel, R., Löhr, F., Foldynová -Trantírková, S., Bamberg, E., Trantírek, L. & Dötsch,V. *Nucleic Acids Research.* **39**, 5768–5775 (2011).

43. Zheng, K. W., Chen, Z., Hao, Y. H. & Tan, Z. Molecular crowding creates an essential environment for the formation of stable G-quadruplexes in long double-stranded DNA. *Nucleic Acids Research.* **38**, 327–338 (2010).

Acknowledgments

Author contribution

L.B.Y. and T.T. conceived and designed the experiments, L.B.Y synthesized chemical compounds as well as performed gel electrophoresis, circular dichroism, enzymatic assays and ChIP-seq experiments, T.T. carried out alkaline comet assay, Western blotting and cytological experiments, Y.Q.C., S.Y.Y. and Z.G.Z. contributed some synthesis with assistance from L.B.Y., X.W.X., Q.Q.Z., B.F.Y. and Y.Q.F. contributed to the MALDI-TOF mass spectrometry, L.X., S.R.W. and X.C.W. assistant analysed ChIP-seq data. X. Z., L.B.Y. and T.T. analysed the data, discussed the results and wrote the paper.

Additional information

SCIENTIFIC REP⚙RTS

OPEN

Corrigendum: Existence of G-quadruplex structures in promoter region of oncogenes confirmed by G-quadruplex DNA cross-linking strategy

Libo Yuan, Tian Tian, Yuqi Chen, Shengyong Yan, Xiwen Xing, Zhengan Zhang, Qianqian Zhai, Liang Xu, Shaoru Wang, Xiaocheng Weng, Bifeng Yuan, Yuqi Feng & Xiang Zhou
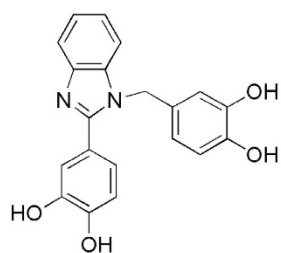
A recent study has reported that the structure of known compound 1 should be revised (revised compound 1, Fig. 1)[1]. We are aware that the structure of compound 1 reported in our Article is incorrect and should be revised. In addition, the structure of compound 1a should also be revised in the Article and Supplementary Information (Fig. 1). Revised compound 1a was the final product (HPLC purified) and was used in the pull-down experiments reported in the paper.

The structure of compound 4 (Supplementary Information) should also be revised (Fig. 1). Revised Compound 4 was not used in any of the biological studies reported in our paper, and is an intermediate in the synthesis of revised compound 1a. These revised structures do not affect the central conclusions of the paper, because the biological active groups (bis-phenol) are unchanged.
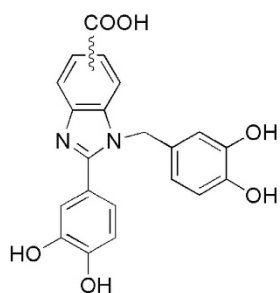
References
1. Reyes-Gutiérrez, P. E. *et al.* Structural revisions of small molecules reported to cross-link G-quadruplex DNA *in vivo* reveal a repetitive assignment error in the literature. *Sci. Rep.* **6**, 23499 (2016).
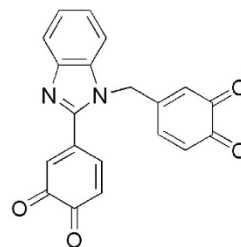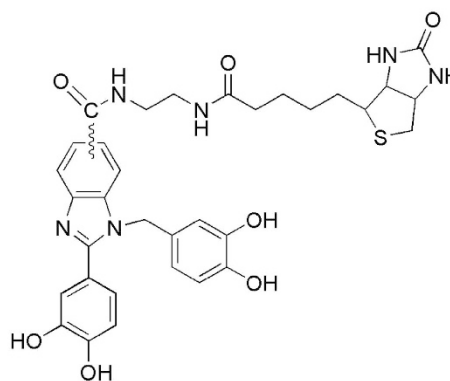
Figure 1. Revised structures.