Ligand-mediated G-quadruplex Induction in a Double-stranded DNA Context by Cyclic Imidazole/Lysine Polyamide

Sefan Asamitsu,[a] Yue Li,[a] Toshikazu Bando,*[a] and Hiroshi Sugiyama*[a,b]

Abstract: G-quadruplex (G4) DNA is often observed as a DNA secondary structure in guanine-rich sequences and is thought to be relevant to pharmacological and biological events. Therefore, G4 ligands have attracted great attention regarding potential anticancer therapies or molecular probe applications. In the present study, we designed cyclic imidazole/lysine polyamide (cIKP) as a new class of G4 ligand, which was readily synthesized without time-consuming column chromatography. cIKP enables the selective recognition of particular G4 structures with low nanomolar affinity. Moreover, cIKP exhibited the ability to induce G4 formation on the promoter of G4-containing DNAs in the context of stable double-stranded DNA (dsDNA) under molecular crowding conditions. This cIKP may be applicable as a molecular probe for the detection of potential G4-forming sequences in dsDNA.

Introduction

G-quadruplex (G4) DNA is the higher-order structure of four-stranded nucleic acids including several G-quartets comprising four planar guanines that stabilize each other via Hoogsteen hydrogen bonding. G4 structures can be formed in G-rich sequences, and monovalent metal cations (K^+ or Na^+) reinforce the stability of such structures.[1] It has been shown that G4-forming sequences are widely present in human genomes, notably in the promoter and 5′-UTR regions of genes that are involved in cellular proliferation or are relevant to several diseases, as well as in telomere sequences.[2] Hence, G4 structures have attracted great attention regarding biological events, including gene alteration,[3a,b] epigenetic regulation,[3c] genome stability,[3d] and targets for these therapies.[3e–g] To date, several synthetic molecules have been reported as G4 ligands, such as flat aromatic, macrocyclic, and crescent-shaped compounds.[4] In particular, the fact that a naturally occurring macrocycle, telomestatin, exhibits telomerase-inhibition activity by binding to telomeric G4 structures has accelerated attempts to study synthetic macrocycles for the direct targeting of these structures, aiming to develop potential anticancer therapies or molecular probe applications.[3d,5] For example, furan-based cyclic oligopeptides were shown to be highly selective for G4 structures, and some cyclic oligopeptides (24-membered rings) suppressed the mRNA expression of the *c-Myc* oncogene, the promoter of which contains G4-forming sequences.[6] A cationic telomestatin derivative shows strong binding affinity for telomeric G4 over other G4 structures.[7] In addition to these compounds, several macrocyclic compounds have been reported as G4 ligands, and, in some cases, G4 topology selectivity was successfully achieved.[3b,8] Therefore, such macrocyclic structures are attractive scaffolds to achieve both binding affinity and selectivity for particular G4 structures.

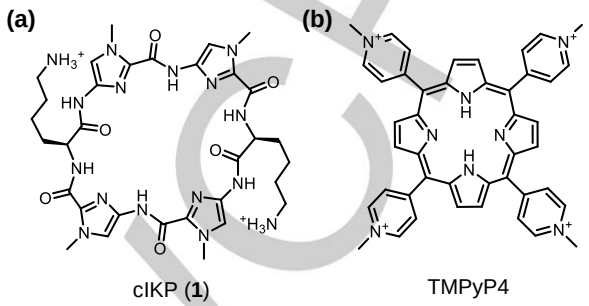


Figure 1. Chemical structures of cIKP (**1**) (a) and TMPyP4 (b).

For years, we have studied a synthetic DNA minor groove binder named pyrrole–imidazole polyamide (PIP). PIP was developed by Prof. Dervan and colleagues as a programmable DNA-binding molecule and distinctly recognized A/T base parings and G/C base parings with modest affinity and specificity.[9] In the course of such PIP studies, we have reported that consecutive imidazole rings held together via amide bonds adopt a highly planar conformation, as estimated by density functional theory.[10] In the present study, we designed cyclic Imidazole/Lysine Polyamide, cIKP (**1**) as a new class of G4 ligand, which was readily synthesized and which enabled selective recognition of the particular G4s with higher binding affinity compared to TMPyP4, a well-studied G4-interactive compound (Figure 1). Furthermore, cIKP exhibited the ability to induce G4 formation on the promoter G4 DNAs in the context of stable dsDNA under molecularly crowded conditions.

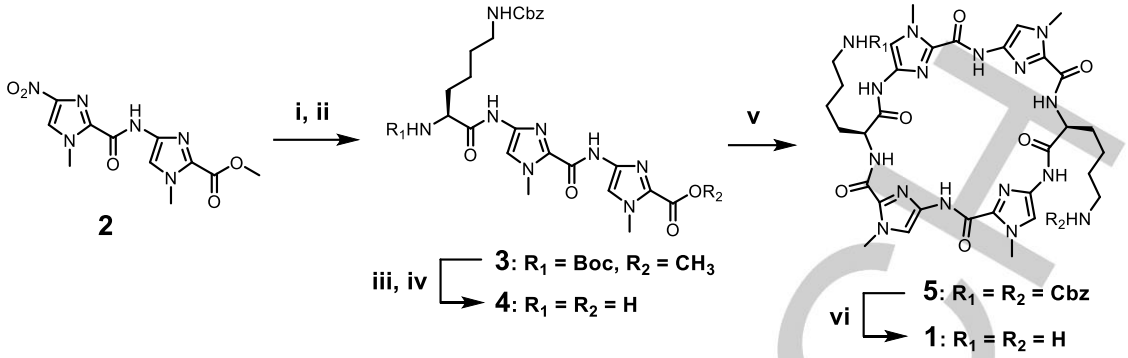
Results and Discussion

Molecular design and synthesis

Based on the above mentioned report, the imidazole moiety was selected as a key component for the molecular planarity of the structure, although other heteroaromatic rings were widely used to construct the planar backbones of most of the other macrocyclic G4 ligands.[4a,11] The lysine residue was adopted as another core mainly because its positively charged side chains

[a] S. Asamitsu, Y. Li, Dr. T.Bando, Prof. H. Suigiyama
Department of Chemistry, Graduate School of Science
Kyoto University Kitashirakawa-Oiwakecho
Sakyo-ku, Kyoto 606-8502 (Japan)
E-mail: hs@kuchem.kyoto-u.ac.jp (H. Sugiyama),
Bando@kuchem.kyoto-u.ac.jp (T. Bando).
[b] Prof. H. Sugiyama
Institute for Integrated Cell-Material Sciences (WPI-iCeMS)
Kyoto University Yoshida-Ushinomiyacho
Sakyo-ku, Kyoto 606-8501 (Japan).

Supporting information for this article is given via a link at the end of the document.



Scheme 1. Synthetic route of cIKP (**1**) in solution-phase. Compound **2** was prepared from NO_2-I-$COCCl_3$. Reagents and conditions; (i) H_2, 10 wt% Pd/C, AcOEt, MeOH, rt. (ii) BocNH-Lys(Cbz)-OH, HCTU, DIEA, DMF, rt, 64% over two steps. (iii) NaOH, MeOH, H_2O, 45 °C. (iv) TFA, DCM, rt. (v) FDPP, DIEA, DMF, rt, 52% over 3 steps. (vi) TfOH, TFA, rt, 98%.

can electrostatically interact with the negatively charged phosphate backbone of the G4s.[12] Although the incorporation of the two lysine moieties seems likely to sacrifice the high planarity of the molecule, the relatively flexible compound could mold into the surface of the G-quartet.[12] cIKP (**1**) was synthesized as shown in Scheme 1. We successfully synthesized the target compound using a total of nine steps without any time-consuming chromatographic purification, with the exception of that of the final product. Moreover, although, generally, 2 or 3 days are necessary to finish the ring-closing step upon the construction of the macrocyclic backbone in low yield,[6,7,8b] the dimerization of the linear imidazole/lysine polyamide gave a better yield and reduced the reaction time. This may be attributed to internal H-bond formation between the N of imidazole and the H of the adjacent amide.[10] Three commonly used condensation agents, PyBOP, HCTU, and DPPA, were applicable to this dimerization reaction, with comparable yields achieved within 4 h (see Supporting Information).

SPR-binding assays

To evaluate the binding properties of **1** for G4s and double-stranded DNAs (dsDNA) quantitatively, we performed an SPR-binding assay (Figure 2).[13] We selected five DNAs harboring G4-forming sequences located in a human telomere, and the promoter regions of the *c-Myc*, *c-kit-1*, *c-kit-2*, and *BCL2* genes, which were studied well via quantitative analysis using SPR-binding assays (Table S1).[6,14] The sensorgrams obtained for *c-Myc* or telomeric/*c-kit-1*/*c-kit-2*/*BCL2* G4s were better fitted by a single-site or a two-site model, respectively, suggesting a 1:1 or 2:1 stoichiometry for the binding of **1** to *c-Myc* or telomeric/*c-kit-1*/*c-kit-2*/*BCL2* G4s within optimized concentration ranges. These two fitting models include equations that reflect mass transfer limitation effect. The association rate (k_a), dissociation rate (k_d), and dissociation constant (K_D) for the interaction of **1** with the G4 DNAs are shown in Table 1. For telomeric/*c-kit-1*/*c-kit-2*/*BCL2* G4s, **1** exhibited a preference for one site over the other. The kinetic parameters and dissociation constants for the weaker

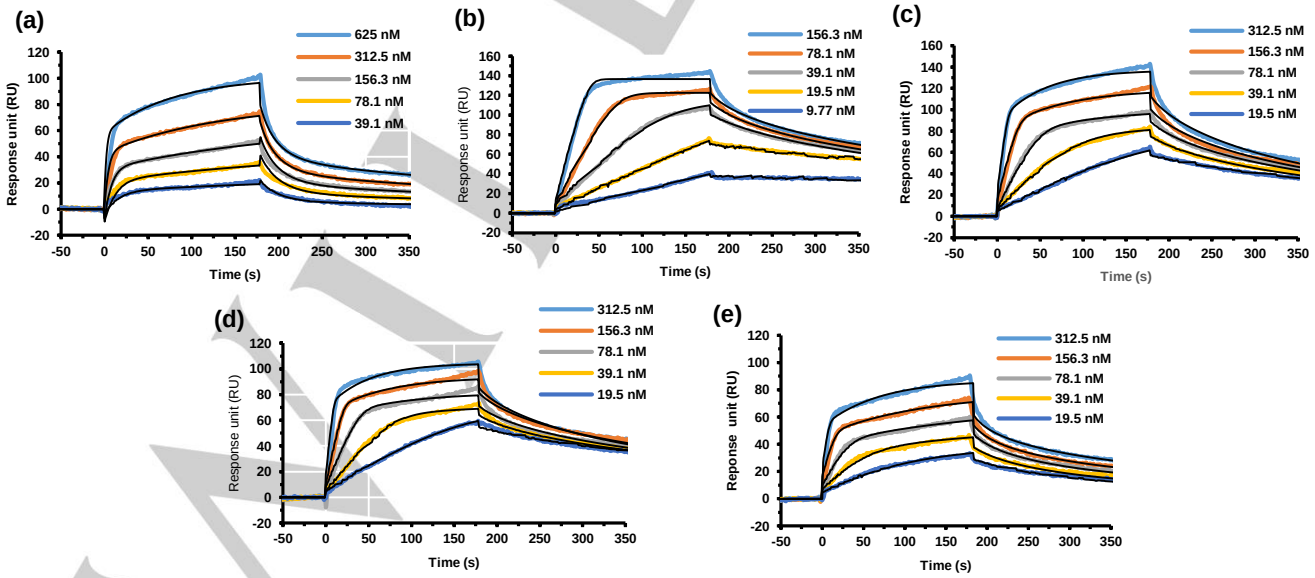


Figure 2. SPR-binding assays to evaluate the binding properties of cIKP (**1**). Each SPR sensorgram is fitted with an appropriate fitting model (black line). (a) SPR sensorgrams for interactions with telomere G4. (b) SPR sensorgrams for interactions with *c-Myc* G4. (c) SPR sensorgrams for interactions with *c-kit-1* G4. (d) SPR sensorgrams for interactions with *c-kit-2* G4. (e) SPR sensorgrams for interactions with *BCL2* G4.

Table 1. Values of the association rates (k_a) and dissociation rates (k_d) obtained from curve fittings of the sensorgrams, and dissociation constants (K_D).

	cIKP (**1**)		
DNA	k_a [$M^{-1}s^{-1}$]	k_d [s^{-1}]	K_D [nM]
Telomere[a]	7.5×10^5	6.7×10^{-2}	90
c-Myc[b]	2.0×10^6	1.3×10^{-2}	6.2
c-kit-1[a]	2.2×10^6	1.6×10^{-2}	7.4
c-kit-2[a]	3.3×10^6	1.2×10^{-2}	3.8
BCL2[a]	2.4×10^6	4.1×10^{-2}	17
dsDNA[a]	Few response	Few response	Few response

[a] Determined by fitting with a modified heterogeneous ligand-binding model (two-site binding model). [b] Determined by fitting with a 1:1 binding model with mass transfer.

binding sites were described in Table S2. Compound **1** showed higher affinity for *c-Myc* (6.2 nM), *c-kit-1* (7.4 nM), *c-kit-2* (3.8 nM), and *BCL2* (17 nM) G4s, whereas it had modest affinity for the telomeric G4 (90 nM). This indicates that **1** has a 24-fold selectivity toward particular G4 structures (*c-kit-2* vs telomeric G4s). Furthermore, a low response of **1** for duplex DNA (both GC-rich and AT-rich sequences) was observed even at higher concentrations, indicating a significantly high selectivity toward G4 structures (Figure S1). Compared with the K_D values of TMPyP4 (Figure 1b), which were measured for some G4s using the SPR method,[8h] **1** exhibited approximately 2-fold and 5–10-fold higher binding affinities for telomeric G4s and *c-Myc*/*c-kit-1* G4s, respectively. Collectively, **1** showed high selectivity and low nanomolar affinity toward G4 compared with dsDNA and exhibited some modest selectivity toward particular G4s.

CD spectra analysis

Next, we performed a CD spectra analysis to investigate the binding behavior of **1** to G4 structures using *c-Myc*, *c-kit-1*, *c-kit-2*, and *BCL2* G4s (Table S1). CD titration using several concentrations of **1** revealed that **1** can recognize and bind to G4 structures and does not cause conformational changes upon binding (Figure S2). Moreover, the ability of **1** to form G4 structures was evaluated in the absence of K^+ (Figure S3).[15] Regarding *c-Myc*, *c-kit-2*, and *BCL2* ssDNA, no significant differences were observed compared with the case of the presence of K^+, suggesting that **1** forms the same G4 structures according to CD spectra (Figure S3a,c,d).[16] Conversely, **1** exhibited a different behavior for *c-kit-1* ssDNA, including the amplitude of the positive peak detected at ~295 nm (Figure S3b). This suggests that **1** drives *c-kit-1* ssDNA to form hybrid or antiparallel/parallel mixed G4s instead of the parallel structure.[16] Those results demonstrate that **1** has the ability to form G4 structures from ssDNA.

Induction of G4 formation in dsDNA context

In living cells, genomic DNA is usually present as a double helix. We examined the binding properties of **1** for four G4s located in the promoter regions in the presence of their complementary strands. A previous pioneer work clearly reveled induction of G4 formation by a small molecule using the core G4-forming sequence and its complementary strand by gel electrophoresis.[17]

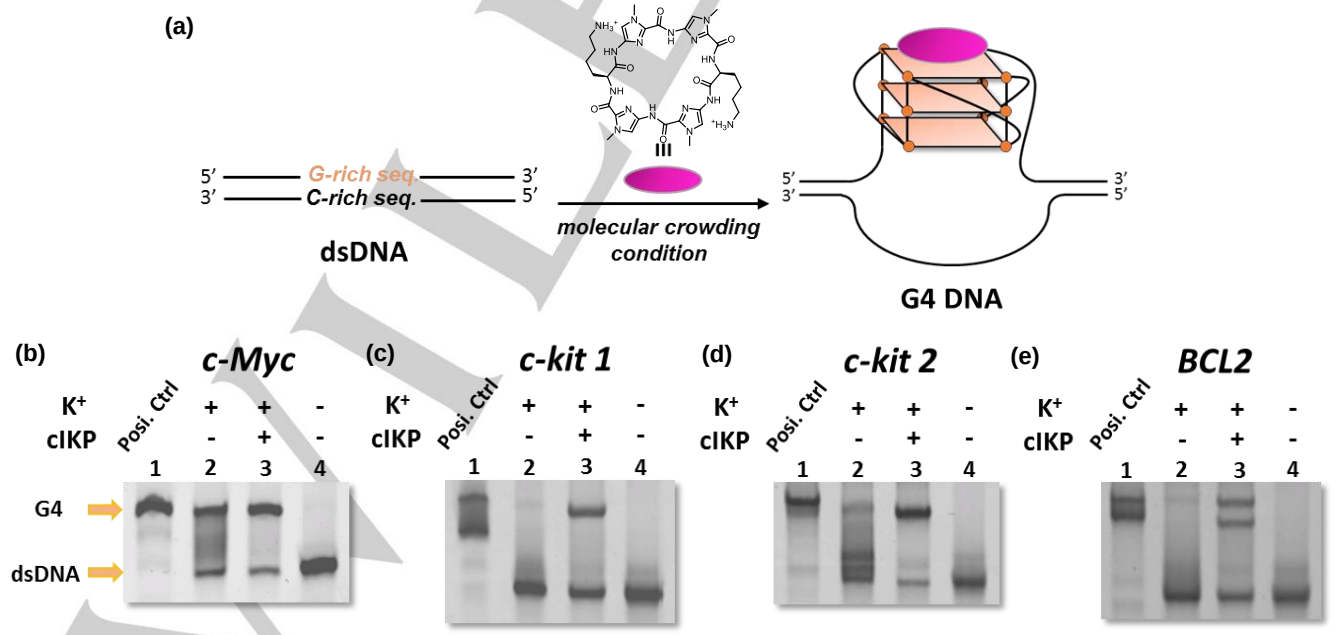


Figure 3. Induction of G4 formation by **1** in the presence of KCl, as assessed by native gel electrophoresis. (a) Schematic representation of G4 induction by the ligand. (b) *c-Myc* dsDNA, (c) *c-kit-1* dsDNA, (d) *c-kit-2* dsDNA, and (e) *BCL2* dsDNA. Lane 1 is a positive control for G4 formation bands. Lane 4 shows dsDNA formation bands. Note that the two bands showing G4 formation for *c-kit-1* and *BCL2* are attributable to several potential G4 conformations on each sequence.[19]

To truly reflect and better understand G4 induction effect by **1** in the genomic DNA context, we employed DNAs that consisted of a single G4-forming sequence and two flanking duplex regions and their complementary DNAs, and conducted these experiments under molecularly crowded condition.[18] DNAs including a single G4-forming sequence and their complementary DNAs were annealed in 40% (w/v) PEG 200 solution. The DNA solution was incubated at 37 °C for 30 min in the presence or absence of **1** and was then subjected to native gel electrophoresis. As positive controls, poly-T sequences, with which the complementary sequences of G4-forming ones were replaced, were used to confirm the band showing the DNA containing a G4 on the gel (Table S1). Such G4-containing DNAs migrated behind the same sequences remaining in the dsDNA.[18] The migration behavior of DNAs upon the addition of **1** are shown in Figure 3. Regarding the *c-kit-1*, *c-kit-2*, and *BCL2* dsDNAs, G4s were almost not formed in the absence of **1**, but migrated bands indicating G4 formation appeared after incubation with **1**, with a mobility on the gel that corresponded to that of a positive control (Figure 3c–e). Note that two bands showing G4 for *c-kit-1* and *BCL2* can be attributed to several potential G4 conformations on each sequence.[19] Conversely, TMPyP4 did not exhibit the G4 induction effect under the same conditions containing the PEG solution (Figure S4). Given that such G4 induction did not occur in the absence of PEG solution (Figure S5), **1** has the property of inducing G4 formation from the stable dsDNA under a molecular crowding condition. Regarding the *c-Myc* dsDNA, G4 had already been formed in the absence of **1** and, therefore, it is difficult to observe the induction effect by **1** in the dsDNA context (Figure 3b). To circumvent this problem, we performed the same experiment in the absence of KCl. Interestingly, even in the absence of KCl, such induction of G4 formation by **1** was observed (Figure S6). This result also demonstrates the induction ability of **1** to form G4 structures in a stable dsDNA context. The dose-dependency observed in Figure S6 was not significant within the measured concentration ranges. We performed additional experiments using the lower concentration ranges of **1**. The results are shown in Figure S7. Even in the lower concentration ranges, the dose-dependency was not significant. We are unable to provide a definitive reason for the observations, but we propose two possible reasons for this. First, a ligand may remain on the G4 structure that was induced by the ligand and, therefore, is not recruited for the next continuous G4 induction. Second, once G4 structures are generated, the ligands may bind nonspecifically to the G4s through other weaker binding sites or additional ligands may aggregate G4-ligand complexes; these changes would be observed as greater SPR responses at higher concentrations. These factors might suppress the ability to induce G4 formation in the presence of increased concentrations of G4s.

Conclusions

We designed and synthesized cyclic imidazole/lysine polyamide, cIKP (**1**), as a new class of G4 ligand based on the high planarity of consecutive imidazole scaffolds, as hinted by structural studies of PIP. cIKP had high selectivity and affinity toward G4 structures over dsDNA, and modest selectivity for the particular G4 structures (24-fold). Interestingly, cIKP displayed the ability to induce G4 formation, even in the presence of the complementary strand. In recent studies, researchers have reported several new motifs that formed stable intramolecular G4s *in vitro*, such as motifs with bulges, long loops, and (4n–1) guanines.[20] Considering the high affinity and induction ability of cIKP toward G4, we anticipate that it might be a promising molecular probe for the detection of such new motifs or relatively unstable G4 structures that have never been discovered before in genome-wide analysis.

Experimental Section

General

^{1}H NMR spectra were recorded on JEOL JNM ECA-600 spectrometer (600 MHz for ^{1}H and 150 MHz for ^{13}C), with chemical shifts reported in parts per million relative to residual solvent and coupling constants in hertz. The following abbreviations were applied to spin multiplicity: s (singlet), d (doublet), dd (doublet of doublets), t (triplet), m (multiplet), br (broad). Regular column chromatography was performed using Silica Gel 60 (70–230 mesh, Merck Chemicals). Analytical HPLC was performed on a Jasco Engineering PU-2080 plus series system using a 150 X 4.6 mm X-Terra MS C18 reversed-phase column in 0.1% TFA in water with acetonitrile as the eluent at a flow rate of 1.0 mL/min and a linear gradient elution of 0–100% acetonitrile in 20 or 40 min with detection at 254 nm. Collected fractions were analyzed by ESI-TOF-MS (Bruker). Reversed-phase flash column chromatography was performed on CombiFlash Rf (Teledyne Isco, Inc.) using a 4.3 g reversed-phase flash column (C18 RediSep Rf) in 0.1% TFA in water with acetonitrile as the eluent at a flow rate of 18.0 mL/min and a linear gradient elution of 0–35% acetonitrile in 5–40 min with detection at 254 nm. 10wt% Pd/C and Pentafluorophenyl diphenylphosphinate (FDPP) were purchased from Aldrich. *O*-(1*H*-6-chlorobenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HCTU) were purchased from Peptide International. Diphenylphosphoryl azide (DPPA), O_2N-I-COCl$_3$, trifluoromethansulfonic acid (TfOH), and *N,N*-dimethylformamide (DMF) were purchased from Wako, and (Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyBOP) was purchased from Novabiochem. BocHN-Lys(Cbz)-CO_2H was purchased from Watanabe Chemical Industries, LTC. Diisopropylethylamine (DIEA) was purchased from Nacalai Tesque, Inc. Trifluoroacetic acid (TFA) was purchased from Kanto Chemical Co., Inc. Dichloromethane (DCM) was purchased from Sasaki chemical co., Ltd. The other reagents and solvents were purchased from standard suppliers and used without further purification.

Synthesis of cIKP (1)

Synthesis of compound 3

To a solution of compound **2** (1.01 g, 3.28 mmol), which had been synthesized in 3 steps from NO_2-I-COCCl$_3$ in overall 83 % yield,[21] in AcOEt (40 mL) and MeOH (20 mL) was added 10 wt% Pd/C (90 mg). The solution was stirred at rt under H_2 gas atmosphere (P = 0.21 MPa) for 2 h. After the filtration of Pd/C, the filtrate was concentrated on an evaporator to give the amine compound (803 mg, 2.88 mmol) as a light brown powder, which was used the next step without further purification. The amine compound (121 mg, 0.43 mmol), BocHN-Lys(Cbz)-CO_2H (182 mg, 0.49 mmol), HCTU (198 mg, 0.48 mmol) was dissolved in dry DMF (8 mL) and to the mixture was added DIEA (83 µL, 0.48 mmol). The solution was

stirred at rt for 2.5 h, then evaporated the solvent and washed with water to produce the oil crude, which are used in the next reaction step without further purification. For spectral data, it was purified by column chromatography (silica gel, Hexane:AcOEt 1:10) to give compound **3** as a white powder (203 mg, 73 % yield). ^{1}H NMR (600 MHz, DMSO-d_6) δ 10.32 (s, 1H), 9.80 (s, 1H), 7.70 (s, 1H), 7.50 (s, 1H), 7.35-7.30 (m, 5H) 7.23 (t, *J* = 5.6 Hz, 1H), 6.97 (d, *J* = 7.6 Hz, 1H), 4.99 (s, 2H), 4.10 (dd, *J* = 13.7, 8.9 Hz, 1H), 3.96 (s, 3H) 3.95 (s, 3H), 3.82 (s, 3H), 2.97 (t, *J* = 17 Hz, 2H), 1.56 (m, 2H), 1.38 (s, 9H), 1.31 (m, 2H), 1.24 (m, 2H). ^{13}C NMR (DMSO-d_6) δ 158.8, 156.1, 155.7, 155.4, 137.3, 136.2, 133.0, 131.3, 128.3, 127.7, 115.4, 114.5, 99.5, 78.0, 65.1, 59.7, 54.2, 51.8, 35.6, 35.0, 31.5, 29.1, 28.2, 22.8. ESI-MS m/z calcd for $C_{30}H_{40}N_8O_8$ [M+H] 641.3042, found 641.3112.

Synthesis of compound 5

To a solution of compound **3** (203 mg, 0.32 mmol) in MeOH (6 mL) and H_2O (6 mL) was added a solid NaOH (460 mg), then the mixture was heated at 45 ºC for 1 h. After the evaporation of MeOH, the resulting solution was acidized with 6N HCl aq to produce a white precipitation. It was collected by filtration and dried in vacuo to give the carboxylic acid (159 mg) as a white powder, which was used the next step without further purification. To a solution of the carboxylic acid (32.6 mg) in DCM (1.5 mL) was added TFA (1.5 mL). The solution was stirred at rt for 1 h, then evaporated the solvent and dried in vacuo to give the TFA salt of compound **4** as a light brown solid, which was used the next step without further purification. The salt of compound **4** and FDPP (79 mg, 0.21 mmol) was dissolved in dry DMF (8.7 mL) and to the mixture was added DIEA (54 µL, 0.31 mmol). The solution was stirred at rt for 4 h, then evaporated the solvent. The crude was dissolved in a minimum amount of DCM and recrystallized with Et_2O to give compound **7** as a white powder (27.4 mg, 52 % for 3 steps). Three ordinary condensation agents, PyBOP, HCTU, and DPPA, were also used for this dimerization reaction with comparable yield (40%, 36%, and 32%, respectively). ^{1}H NMR (600 MHz, DMSO-d_6) δ 10.19 (s, 2H), 9.41 (s, 2H), 8.82 (d, *J* = 9.7 Hz , 2H), 7.52 (s, 2H), 7.46 (s, 2H) 7.36-7.29 (m, 10H), 7.25 (t, *J* = 5.7 Hz, 2H), 4.99 (s, 4H), 4.59 (dd, *J* = 15.8, 9.6 Hz, 2H), 4.00 (s, 6H) 3.94 (s, 6H), 2.99 (t, *J* = 13.1 Hz, 4H), 1.85 (m, 2H), 1.75 (m, 2H), 1.49-1.42 (m, 4H), 1.38-1.31 (m, 4H). ^{13}C NMR (DMSO-d_6) δ 168.8, 157.9, 156.1, 154.8, 137.3, 135.6, 134.7, 134.3, 133.2, 128.3, 127.7, 114.9, 113.0, 65.1, 53.7, 34.9, 32.2, 28.9, 22.8. ESI-MS m/z calcd for $C_{48}H_{56}N_{16}O_{16}$ $[M+H]^+$ 1017.4438, found 1017.4467.

Synthesis of compound 1

To a solution of compound **5** (11.6 mg, 11.4 µmol) in TFA (800 µL) was slowly added TfOH (80 µL). The solution was stirred at rt for 5 min, then poured into cold Et_2O (5 mL) to produce a light brown precipitate, which was collected by a centrifuge. The pellet was purified by reverse-phase flash chromatography to give compound **1** as a white powder (10.9 mg, 98 % yield). ^{1}H NMR (600 MHz, DMSO-d_6) δ 10.27 (s, 2H), 9.42 (s, 2H), 8.86 (d, *J* = 9.6 Hz, 2H), 7.72 (br s, 4H), 7.53 (s, 2H) 7.46 (s, 2H), 4.61 (dd, *J* = 15.8, 9.6 Hz, 2H), 4.01 (s, 6H) 3.95 (s, 6H), 2.80 (t, *J* = 13.1 Hz, 4H), 1.87 (m, 2H), 1.76 (m, 2H), 1.64-1.55 (m, 4H), 1.45-1.35 (m, 4H). ^{13}C NMR (DMSO-d_6) δ 169.3, 158.5, 155.3, 136.1, 135.2, 134.8, 133.7, 115.5, 113.6, 54.0, 39.2, 35.4, 32.6, 27.1, 23.0. ESI-MS m/z calcd for $C_{32}H_{44}N_{16}O_6$ $[M+H]^+$ 749.3702, found 749.3704.

SPR-binding experiments

SPR experiments were performed on a Biacore X instrument (GE healthcare) acording to previous reports with some modifications. 5´ Biotinylated DNAs were purchased from JbioS (Table S1). The biotinylated DNAs were immobilized to streptavidin-functionalized SA sensor chips to obtain the desired immobilization level. SPR measurements were carried out using degassed and filtered HBS buffer (10 mM HPES pH 7.4, 150 mM NaCl, 3mM EDTA, and 0.005% Surfactant P20) with 0.1% DMSO and 100 mM KCl at 25°C. A series of sample solutions with a wide range of concentrations were prepared in the buffer with 0.1% DMSO and 100 mM KCl, and injected at a flow rate of 20 µl/min. After each cycle, the samples remainig on the DNA was detatched with 50 mM NaOH/1M NaCl buffer until the baseline of the sensorgrams was recovered. Optimized concentration range was then selected for the subsequent quantitative anlalysis. The resulting sensorgrams were fitted with 1:1 (single site) or modified heterogenous ligand (two site) binding model with mass tranfer using BIAevaluation 4.1 program in order to obtain the values of kinetics parameters (k_a and k_d) and binding affinites (K_D). These two fitting models include equations that reflect mass transfer limitation effect. The best fitted curves were shown in Figure 2.

CD spactra measurements

DNA samples for CD spectra were prepared in 10 mM Tris-HCl (pH 7.5) buffer. Annealing was performed by heating to 95 °C for 5 min and gradually cooling down to room temperature in the presence or absence of **1**. For CD titration assays, to the solution of the annealed DNA (5 µM) in the buffer was added various concentrations of **1**, the solution was incubated for 30 min. CD spectra were measured in 0.5-nm steps from 340 to 220 nm using JASCO J-805LST Spectrometer in a 1-cm quartz cuvette, and shown in Figure S2 and S3.

Native gel electrophoresis analysis

DNA samples for gel electrophoresis were prepared in 10 mM Tris-HCl (pH 7.5) buffer containing 40% (w/v) PEG 200 with or without 50 mM KCl. The DNA samples was heated to 95 °C for 5 min and gradually cooled down to room temperature. **1** was added to the resulting DNA solution (200 nM) and incubated for 30 min at 37 ºC. Those samples were loaded on 12% polyacrylamide gel containing 50mM KCl, and electrophoresed at 4 °C, 100 V, in 1×TBE buffer containing 50 mM KCl for 90 min. The gel was stained with SYBR gold for 15 min and then imaged by FUJIFILM FLA-3000 (FUJIFILM).

Acknowledgements

This work was partially supported by JSPS KAKENHI (grant number 24225005 to H. S and 24310155 to T. B.).

Keywords: cyclic polyamide • G-quadruplexes • G4 induction • heterocycles • molecular crowding condition

[1] a) M. Gellert, M. N. Lipsett, D. R. Davies, *Proc .Nat.l Acad. Sci. U. S. A.* **1962**, *48*, 2013–2018. b) D. Sen, W. Gilbert, *Nature* **1988**, *334*, 364–366. c) M. W. da silva, *Chem. Eur.–J.* **2007**, *13*, 9738–9745. d) Sannohe, Y.; Sugiyama, H. *Curr. Protoc. Nucleic Acid Chem*. **2010**, *17.2*, 1–17. e) D. J. Patel, A. T. Phan, V. Kuryavyi, *Nucleic Acids Res.* **2007**, *35*, 7429–7455.

[2] a) J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res*. **2005**, *33*. 2908–2916. b) J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res.*, **2007**, *35*, 406–413. c) A. M. Zahler, J. R. Williamson, T. R. Cech, D. M. Prescott, *Nature* **1991**, *350*, 718–720. d) E. Y. N. Lam, D. Beraldi, D. Tannahill, S. Balasubramanian, *Nat. Commun*. **2013**, *4*, 1796–1811. e) S. Muller, S. Kumari, R. Rodriguez, S. Balasubramanian, *Nat. Chem.* **2010**, *2*, 1095–1098.

[3] a) R. Rodriguez, K. M. Miller, J. V. Forment, C. R. Bradshaw, M. Nikan, S. Britton, T. Oelschlaegel, B. Xhemalce, S. Balasubramanian, S. P.

Jackson, *Nat. Chem. Biol.* **2012**, *8*, 301–310. b) C. L. Grand, H. Han, R. M. Muñoz, S. Weitman, D. D. Von Hoff, L. H. Hurley, D. J. Bearss, *Mol. Cancer Ther.* **2002**, *1*, 565–573. c) M .J. Law, K. M. Lower, H. P. Voon, J. R. Hughes, D. Garrick, V. Viprakasit, M. Mitson, M. De Gobbi, M. Marra, A. Morris, A. Abbott, S. P. Wilder, S. Taylor, G.M. Santos, J. Cross, H. Ayyub, S. Jones, J. Ragoussis, D. Rhodes, I. Dunham, D. R. Higgs, R. J. Gibbons, *Cell* **2010**, *143*, 367–378. d) M.−Y. Kim, H. Vankayalapati, K. Shin-ya, K. Wierzba, L. H. Hurley *J. Am. Chem. Soc.* **2002**, *124*, 2098–2099. e) Y. Qin, L. H. Hurley, *Biochemie* **2008**, *90*, 1149-1171. f) T. A. Brooks, L. H. Hurley, *Genes & Cancer* **2010**, *1*, 641–649. g) S. Balasubramanian, L. H. Hurley, S. Neidle, *Nat. Rev. Drug Discovery* **2011**, *10*, 261–175.

[4] a) D. Monchaud, M.–P. Teulade–Fichou, *Org. Biomol. Chem.*, **2008**, *6*, 627–636. b) N. W. Luedtke, *Chimia* **2009**, *63*, 134–139.

[5] a) K. Shin–ya, K. Wierzba, K.–I. Matsuo, T. Ohtani, Y. Yamada, K. Furihata, Y. Hayakawa, H. Seto, *J. Am. Chem. Soc.* **2001**, *123*, 1262–1263. b) T. Doi, M. Yoshida, K. Shin–ya, T. Takahashi, *Org. Lett.* **2006**, *8*, 4165–4167.

[6] T. Agarwal, S. Roy, T. K. Chakraborty, S. Maiti, *Biochemistry* **2010**, *49*, 8388–8397.

[7] M. Tera, H. Ishizuka, M. Takagi, M. Suganuma, K. Shin-ya, K. Nagasawa, *Angew. Chem. Int. Ed.* **2008**, *47*, 5557–5560.

[8] a) D. P. N. Goncalves, R. Rodriguez, S. Balasubramanian, J. K. M. Sanders, *Chem. Commun.* **2006**, *45*, 4685–4687. b) P. S. Shirude, E. R. Gillies, S. Ladame, F. Godde, K. Shin-ya, I. Huc, S. Balasubramanian, *J. Am. Chem. Soc*. **2007**, *129*, 11890–11891. c) K. Shinohara, Y. Sannohe, S. Kaieda, K. Tanaka, H. Osuga, Y. Xu, T. Bando, H. Sugiyama, *J. Am. Chem Soc.* **2010**, *132*, 3778-3782. d) K. Jantos, R. Rodriguez, S. Ladame, P. S. Shirude and S. Balasubramanian, *J. Am. Chem. Soc.*, **2006**, *128*, 13662–13663. e) Q. Zhang, X. Cui, S. Lin, J. Zhou, G. Yuan, *Org. Lett.* **2012**, *14*, 6126–6129. f) J. M. Nicoludis, S. P. Barrett, J. L. Mergny, L.A. Yatsunyk, *Nucleic Acids Res.*, **2012**, *40*, 5432–5447. g) D. M. Kong, Y. E. Ma, , J. H. Guo, W. Yang, H. X. Shen, *Anal. Chem.*, **2009**, *81*, 2678–2684. h) A. Arora, S. Maiti, *J. Phys. Chem. B* **2008**, *112,* 8151–8159. i) G. S. Minhas, D. S. Pilch, J. E. Kerrigan, E. J. LaVoie, J. E. Rice, *Bioorg. Med. Chem. Lett.*, **2006**, *16*, 3891–3895.

[9] P. B. Dervan, B. S. Edelson, *Curr. Opin. Struct. Biol.* **2003**, *13*, 284−299.

[10] Y.-W. Han, T. Matsumoto, H. Yokota, G. Kashiwazaki, H. Morinaga, K. Hashiya, T. Bando, Y. Harada, H. Sugiyama, *Nucleic Acids Res.* **2012**, *40*. 11510–11517.

[11] M. J. Cocco, L. A. Hanakahil, M. D. Huber, N. Maizels, *Nucleic Acids Res.*, **2003**, *31*, 2944–2951.

[12] W. J. Chung, B. Heddi, M. Tera, K. Iida, K. Nagasawa, A. T. Phan, *J. Am. Chem. Soc.* **2013**, *135*, 13495−13501.

[13] a) E. R. Lacy, N. M. Le, C. A. Price, M. Lee, M. W. D. Wilson, *J .Am. Chem. Soc.* **2002**, *124*, 2153−2163. b) S. Asamitsu, S. Y. Kawamoto, F. Hashiya, K. Hashiya, M. Yamamoto, S. Kizaki, T. Bando, H. Sugiyama, *Bioorg. Med. Chem.*, **2014**, *22*, 4646–4657.

[14] a) J. Seenisamy, S. Bashyam, V. Gokhale, H. Vankayalapati, D. Sun, A. Siddiqui−Jain, N. Streiner, K. Shin−ya, E. White, W. D. Wilson, L. H. Hurley, *J. Am. Chem. Soc.* **2005**, *127*, 2944–2959. b) E. M. Rezler, J. Seenisamy, S. Bashyam, M.−Y. Kim, E. White, W. D. Wilson, L. H. Hurley, *J. Am. Chem. Soc.* **2005**, *127*, 9439–9447. c) Bejugam, S. Sewitz, P. S. Shirude, R. Rodriguez, R. Shahid, S. Balasubramanian, *J. Am. Chem. Soc.* **2007**, *129*, 12926–12927. d) X.−D. Wang, T.−M. Ou, Y.−J. Lu, Z. Li, Z. Xu, C. Xi, J.−H. Tan, S.−L. Huang, L.−K. An, D. Li, L.−Q. Gu, Z.−S. Huang, *J. Med. Chem.* **2010**, *53*, 4390–4398.

[15] R. Rodriguez, G. D. Pantos, D. P. N. Goncalves, J. K. M. Sanders, S. Balasubramanian, *Angew. Chem. Int. Ed.* **2007***, 46,* 5405–5407.

[16] A. I. Karsisiotis, N. M. Hessari, E. Novellino, G. P. Spada, A. Randazzo, M. W. da Silva, *Angew. Chem. Int. Ed.* **2011**, *50*, 10645–10648.

[17] A. Rangan, O. Y. Fedoroff, L. H. Hurley, *J. Biol. Chem.* **2001**, *276*, 4640–4646

[18] a) K. Zheng, Z. Chen, Y. Hao, Z. Tan, *Nucleic Acids Res*. **2010**, *38*, 327–338. b) C. Zhang, H. Liu, K. Zheng, Y. Hao, Zheng Tan, *Nucleic Acids Res.* **2013**, *41*, 7144–7152.

[19] a) S. Rankin, A. P. Reszka, J. Huppert, M. Zloh, G. N. Parkinson, A. K. Todd, S. Ladame, S. Balasubramanian, S. Neidle, *J. Am. Chem. Soc.* **2005**, *127*, 10584−10589. b) A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle, D. J. Patel, *J. Am. Chem. Soc.* **2007**, *129*, 4386−4392. c) T. S. Dexheimer, D. Sun, L. H. Hurley, *J. Am. Chem. Soc.* **2006**, *128*, 5404−5415.

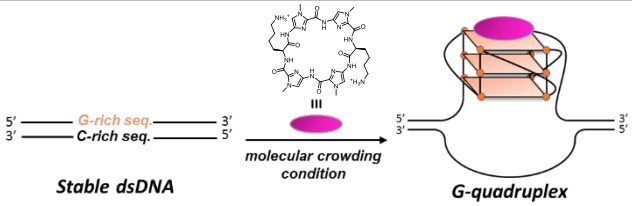
[20] a) V. S. Chambers, G. Marsico, J. M. Boutell, M. D. Antonio, G. P. Smith, S. Balasubramanian *Nat. Biotechnol.*, **2015**, *33*, 877–881. b) V. T. Mukundan, A. T. Phan, *J. Am. Chem. Soc.,* **2013**, *135*, 5017–5028. c) B. Heddi, N. Martìn-Pintado, Z. Serimbetov, T. M. A. Kari, A. T. Phan, *Nucleic Acids Res*. **2016**, *44*, 910–916.

[21] J. Xiao, G. Yuan, W. Huang, *J. Org. Chem.* **2000,** *65,* 5506–5513.

Entry for the Table of Contents (Please choose one layout)

Layout 2:

FULL PAPER



G-quadruplex inducer: cyclic imidazole/lysine polyamide (cIKP) as a new class of G-quadruplex ligand was designed and synthesized. It exhibited the ability to induce G4 formation, even in the presence of the complementary strand. It may be applicable as a molecular probe for the detection of potential novel G4s.

S. Asamitsu, Y. Li, T. Bando, H. Sugiyama**

Page No. – Page No.

Ligand-mediated G-quadruplex Induction in a Double-stranded DNA Context by Cyclic Imidazole/Lysine Polyamide

Supporting Information

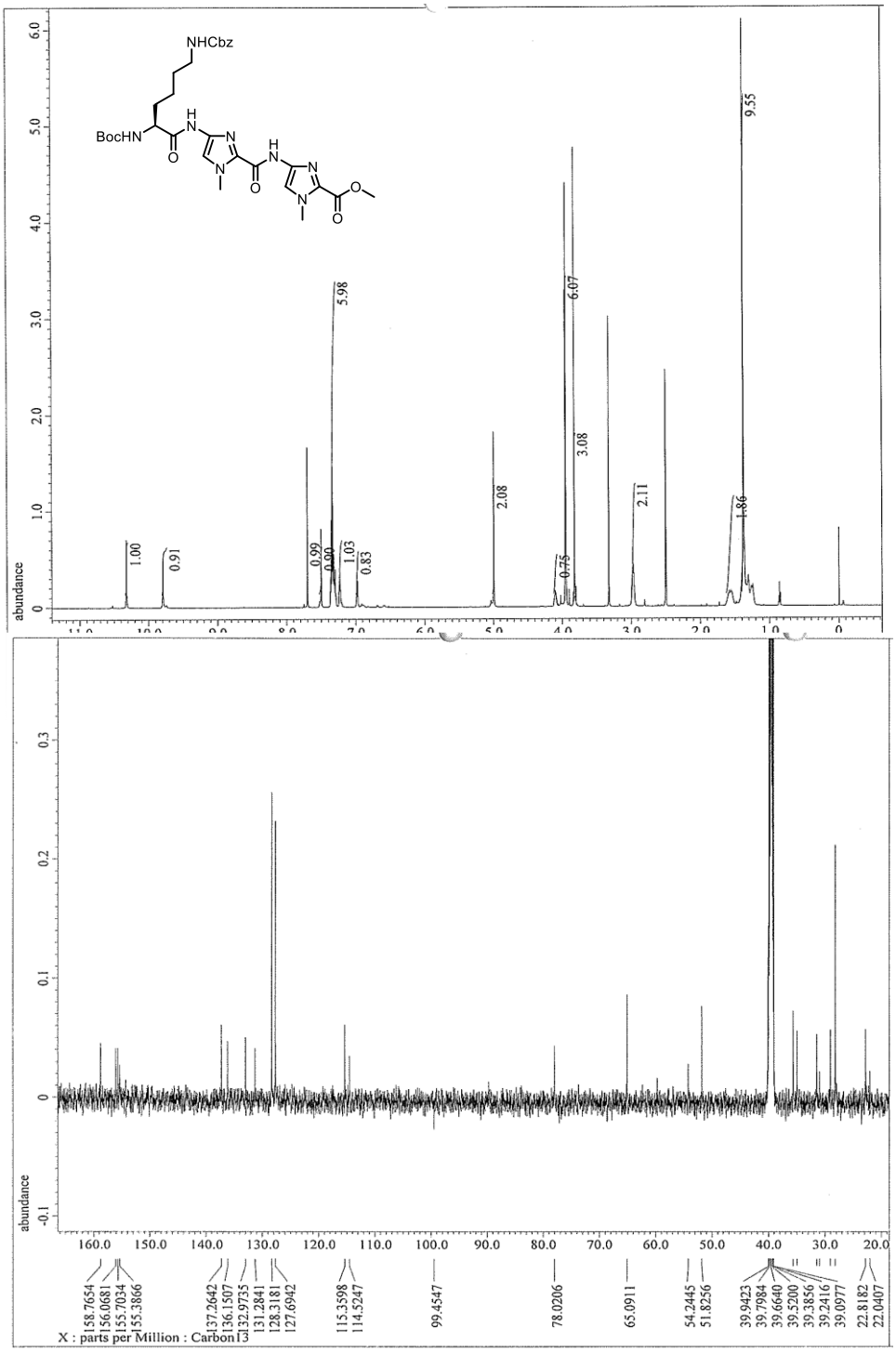
Ligand-mediated G-quadruplex Induction in a Double-stranded DNA Context by Cyclic Imidazole/Lysine Polyamide

Sefan Asamitsu, Yue Li, Toshikazu Bando, and Hiroshi Sugiyama**

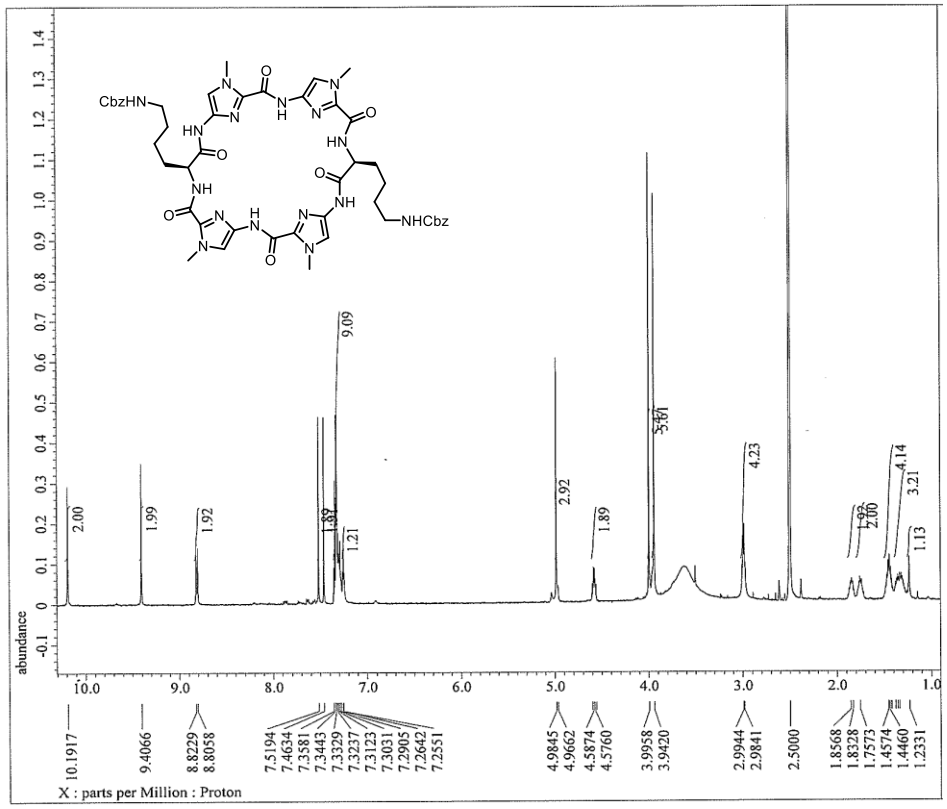
Table of contents

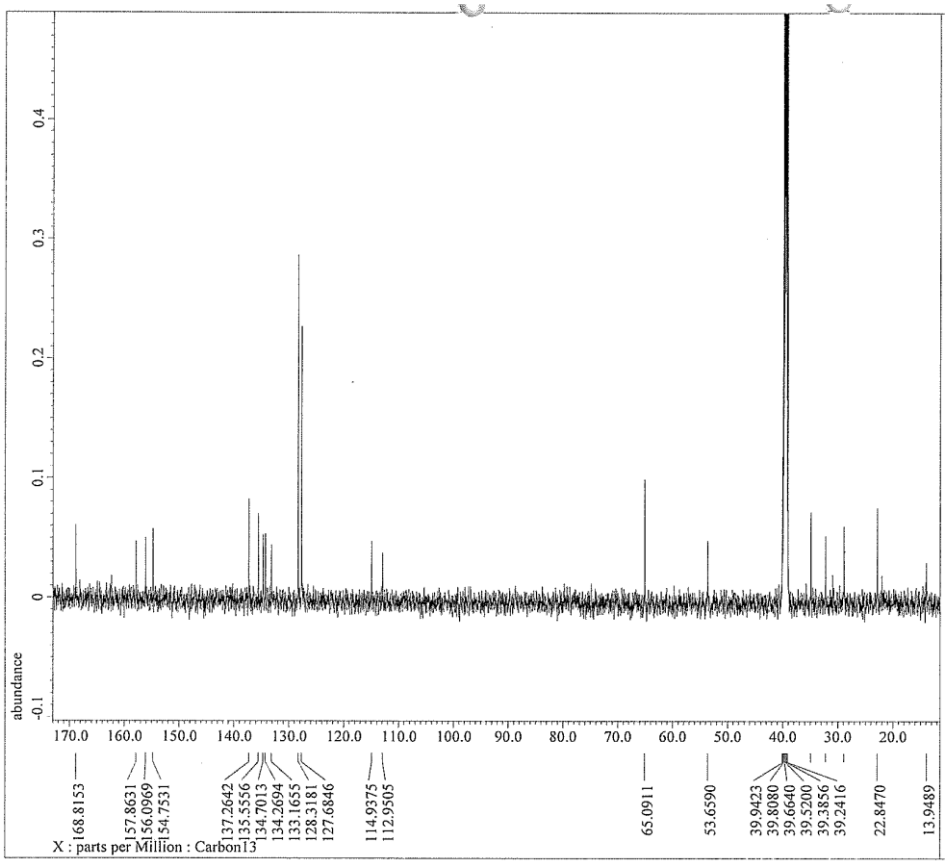
Spectral data

^{1}H NMR and ^{13}C NMR spectra of compound 3

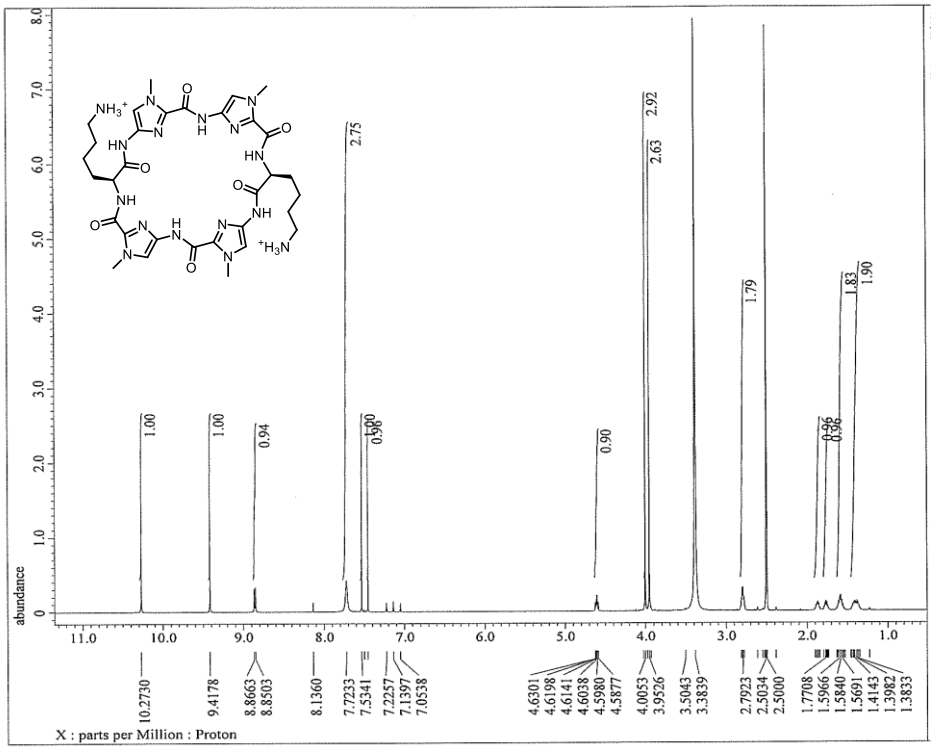


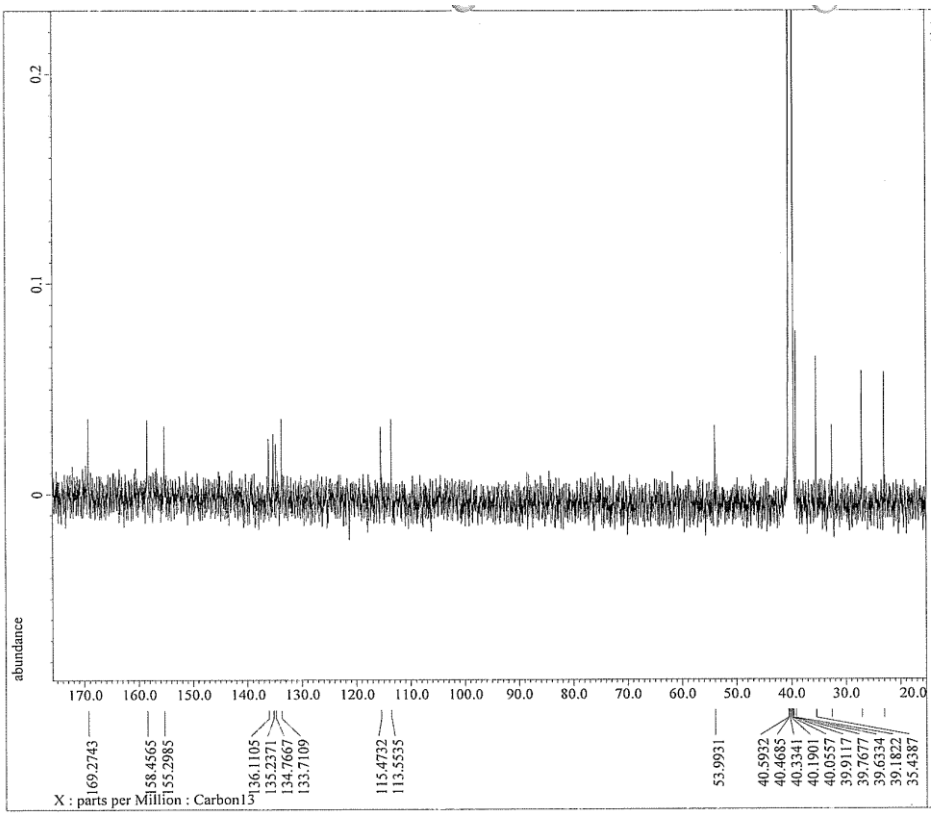
^{1}H NMR and ^{13}C NMR spectra of compound 5





^{1}H NMR and ^{13}C NMR spectra of compound 1





Supplemental data

Table S1. DNA oligomers used in this study.

DNA	Sequence (5'–3')
telomere	biotin-TTTAGGGTTAGGGTTAGGGTTAGGG
c-Myc	biotin-TTTTGGGGAGGGTGGGGAGGGTGGGGAAGG
c-kit 1	biotin-TTTAGGGAGGGCGCTGGGAGGAGGG
c-kit 2	biotin-TTTCGGGCGGGCGCGAGGGAGGGG
BCL2	biotin-TTTGGGCGCGGGAGGAAGGGGGCGGG
GC-rich dsDNA	biotin-GCCGCGCGCGCTTATTTTAAGCGCGCGCGGC
AT-rich dsDNA	biotin-GCCATATATATTTATTTTTAAATATATATGGC
c-Myc ssDNA	TGGGGAGGGTGGGGAGGGTGGGGAAGG
c-kit 1 ssDNA	AGGGAGGGCGCTGGGAGGAGGG
c-kit 2 ssDNA	CGGGCGGGCGCGAGGGAGGGT
BCL2 ssDNA	GGGCGCGGGAGGAATTGGGCGGG
c-Myc dsDNA	GCGGTTCTGAACTCGATAT GGGTGGGTAGGGTGGG ATTAGTGCTAGCTACGCG CGCGTAGCTAGCACTAAT CCCACCCTACCCACCC ATATCGAGTTCAGAACCGC CGCGTAGCTAGCACTAAT TTTTTTTTTTTTTTTT ATATCGAGTTCAGAACCGC
c-kit 1 dsDNA	GCGGTTCTGAACTCGATAT AGGGAGGGCGCTGGGAGGAGGG ATTAGTGCTAGCTACGCG CGCGTAGCTAGCACTAAT CCCTCCTCCCAGCGCCCTCCCT ATATCGAGTTCAGAACCGC CGCGTAGCTAGCACTAAT TTTTTTTTTTTTTTTTTTTTT ATATCGAGTTCAGAACCGC
c-kit 2 dsDNA	GCGGTTCTGAACTCGATAT CGGGCGGGCGCGAGGGAGGGG ATTAGTGCTAGCTACGCG CGCGTAGCTAGCACTAAT CCCCTCCCTCGCGCCCGCCCG ATATCGAGTTCAGAACCGC CGCGTAGCTAGCACTAAT TTTTTTTTTTTTTTTTTTTTT ATATCGAGTTCAGAACCGC
BCL2 dsDNA	GCGGTTCTGAACTCGATAT GGGCGCGGGAGGAAGGGGGCGGG ATTAGTGCTAGCTACGCG CGCGTAGCTAGCACTAAT CCCGCCCCCTTCCTCCCGCGCCC ATATCGAGTTCAGAACCGC CGCGTAGCTAGCACTAAT TTTTTTTTTTTTTTTTTTTTT ATATCGAGTTCAGAACCGC

Table S2. Values of the association rates (k_a) and dissociation rates (k_d) obtained from curve fittings of the sensorgrams, and dissociation constants (K_D) for the weaker binding sites.

	cIKP (**1**)		
DNA	k_a [$M^{-1}s^{-1}$]	k_d [s^{-1}]	K_D [nM]
Telomere[a]	1.7×10^4	2.3×10^{-3}	131
c-kit-1[a]	3.7×10^4	9.8×10^{-3}	262
c-kit-2[a]	3.8×10^4	1.1×10^{-2}	291
BCL2[a]	4.4×10^4	2.4×10^{-3}	53

[a] Determined by fitting with a modified heterogeneous ligand-binding model (two-site binding model).

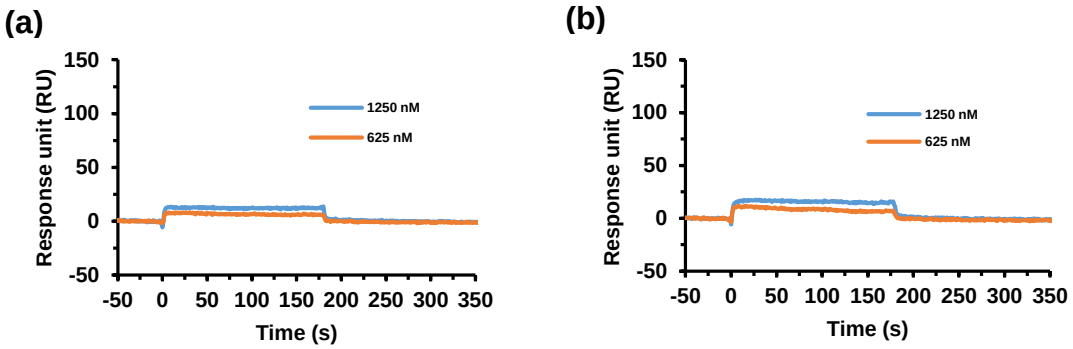


Figure S1. SPR-binding assays to evaluate the binding properties of cIKP (**1**). (a,b) SPR sensorgrams for interactions with duplex DNA including GC- or AT-rich sequences, respectively.

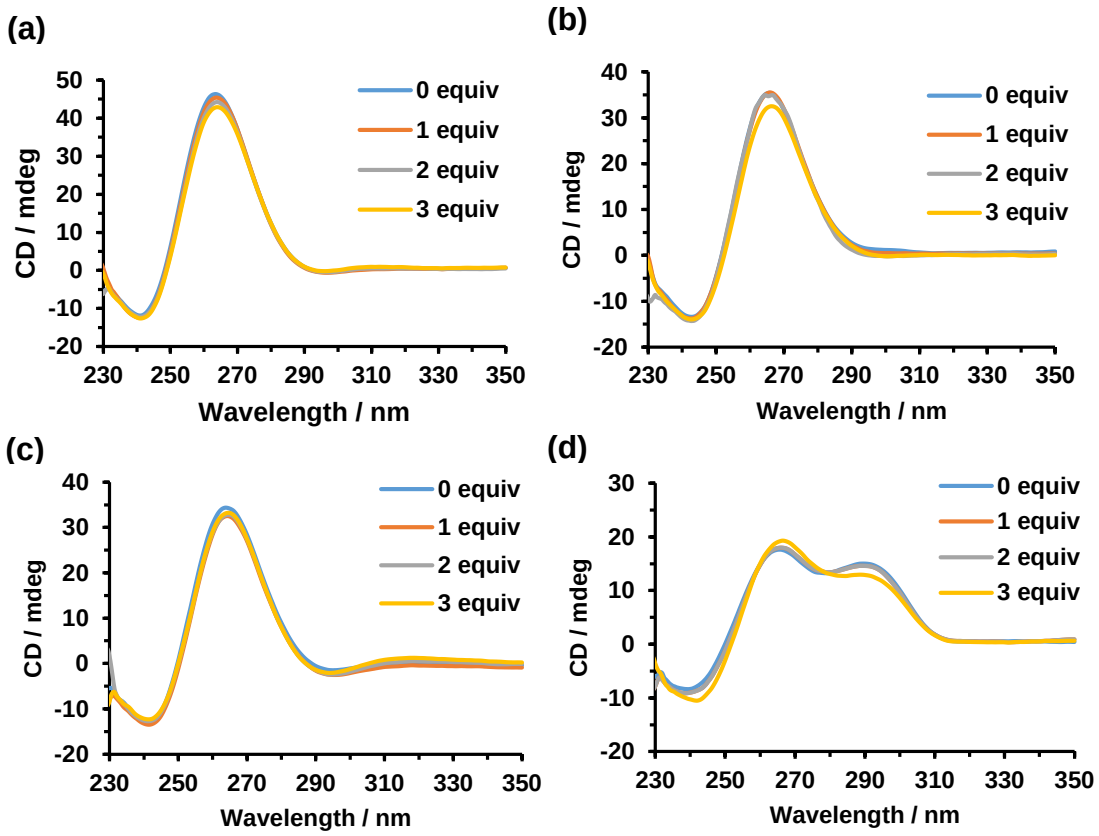


Figure S2. CD spectra of *c-Myc* (a), *c-kit 1* (b), *c-kit 2* (c), and *BCL2* (d) ssDNA (5 μM) after the addition of **1** (5 to 15 μM) under 50 mM K^+ solution. A moderate positive peak at 262 nm and a negative peak at 240 nm is characteristic of parallel conformations (a,b,c), and A positive peak around at 264 and 290 nm and a negative peak around at 240 nm is characteristic of mixed parallel/antiparallel conformations (d).

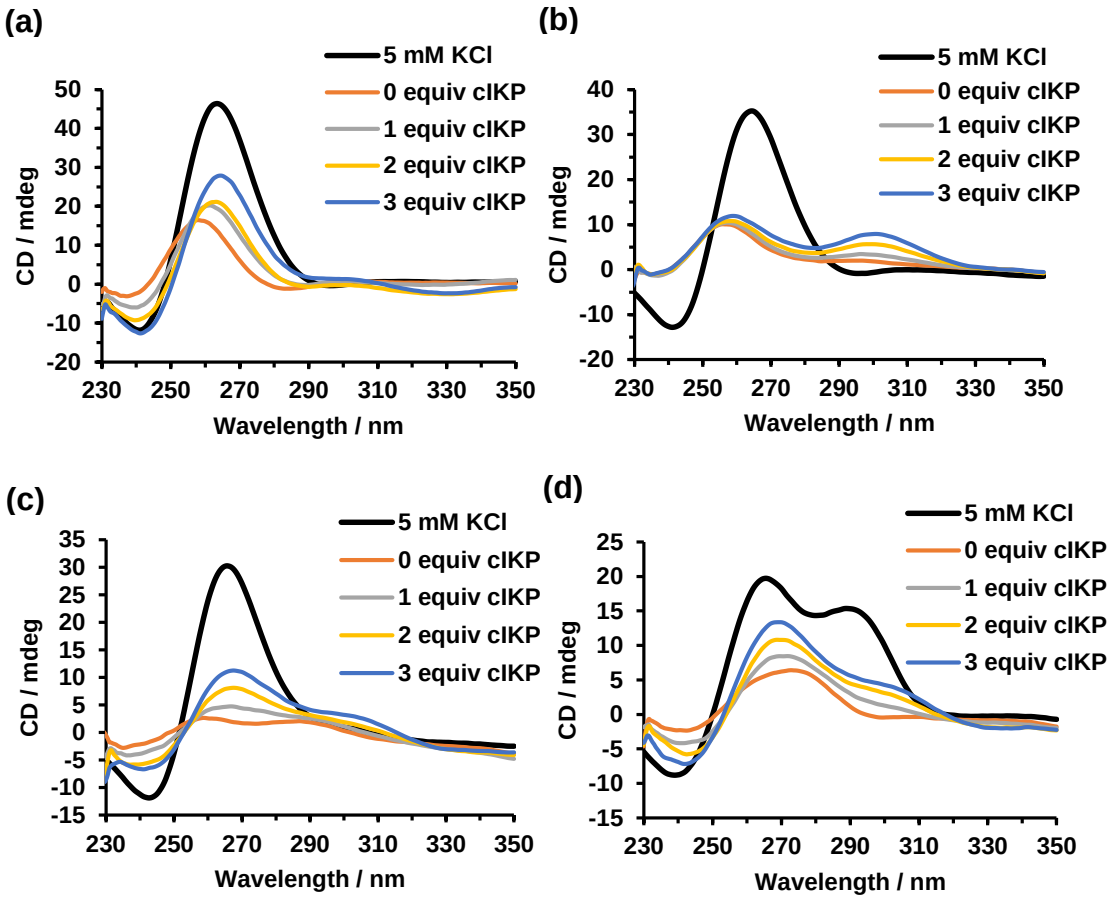


Figure S3. CD spectra of *c-Myc* (a), *c-kit 1* (b), *c-kit 2* (c), and *BCL2* (d) ssDNA (5 μM) in the presence of **1** (5 to 15 μM) under no K^+.

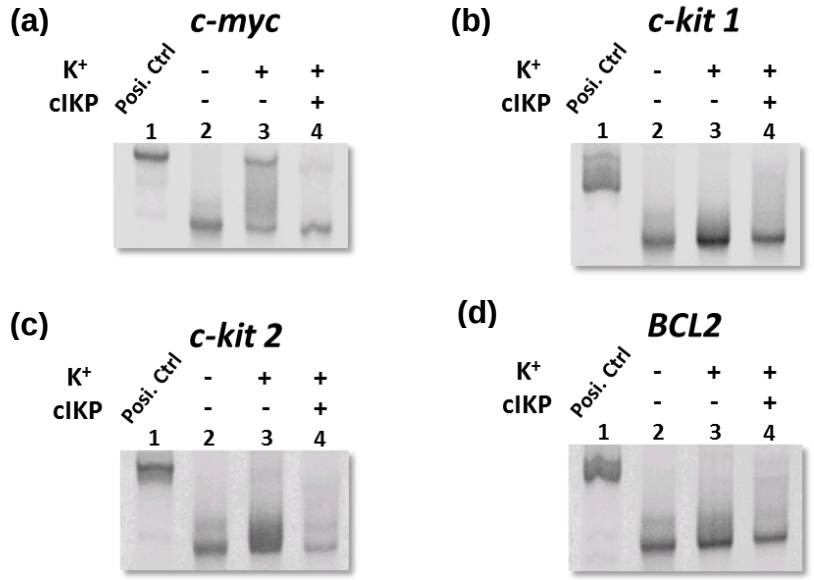


Figure S4. No induction of G4 formation by TMPyP4 (Figure 1b) by native gel electrophoresis under PEG solution; (a) *c-Myc* dsDNA, (b) *c-kit 1* dsDNA, (c) *c-kit 2* dsDNA, (d) *BCL2* dsDNA. Lane 1 is a positive control (Posi. Ctrl) for G4 formation bands. Lane 2 shows dsDNA bands as negative controls.

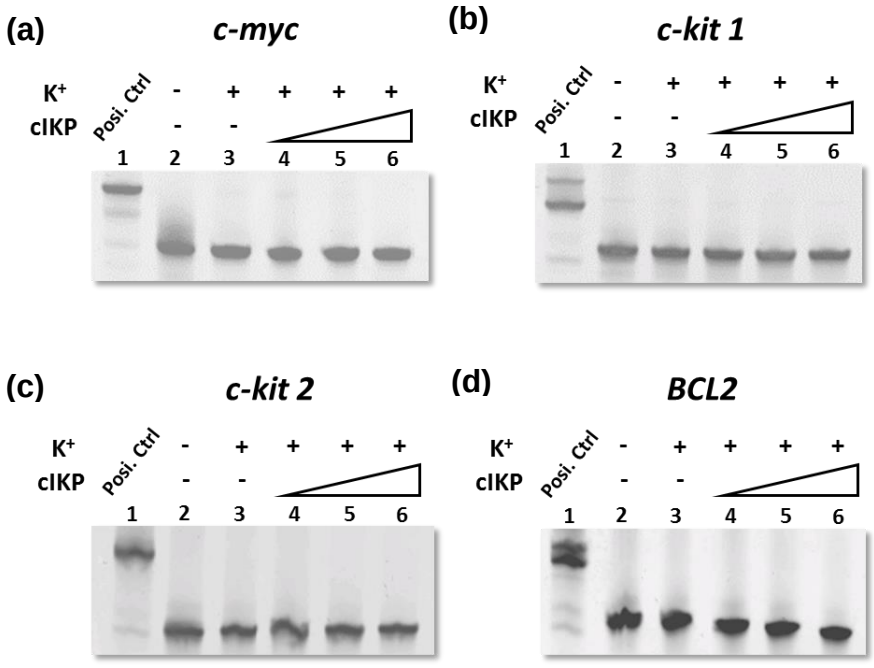


Figure S5. No induction of G4 formation by **1** under no PEG solution by native gel electrophoresis; (a) *c-Myc* dsDNA, (b) *c-kit 1* dsDNA, (c) *c-kit 2* dsDNA, (d) *BCL2* dsDNA. Lane 1 is a positive control (Posi. Ctrl) for G4 formation bands. Lane 2 shows dsDNA bands as negative controls. Concentrations of **1** are 1, 4, and 16 μM for lane 4, 5, and 6, respectively.

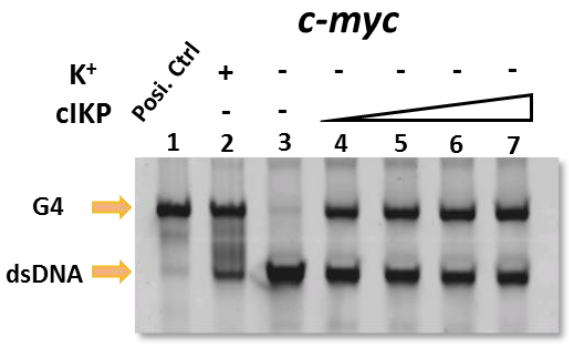


Figure S6. Induction of G4 formation on *c-Myc* dsDNA by cIKP (**1**) in the absence of K^+ under PEG solution by native gel electrophoresis. Lane 1 is a positive control (Posi. Ctrl) for G4 formation bands. Lane 3 shows dsDNA bands as negative controls. Concentrations of **1** are 1, 2, 4, and 8 μM for lane 4, 5, 6, and 7, respectively.

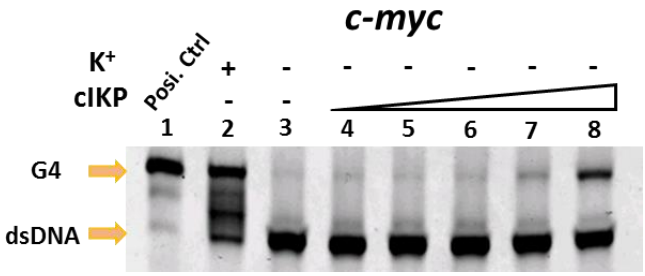


Figure S7. Induction effect of G4 formation on *c-Myc* dsDNA at the lower concentration ranges of cIKP (**1**) in the absence of K^+ under PEG solution by native gel electrophoresis. Lane 1 is a positive control (Posi. Ctrl) for G4 formation bands. Lane 3 shows dsDNA bands as negative controls. Concentrations of **1** are 62.5, 125, 250, 500, and 1000 nM for lane 4, 5, 6, 7, and 8, respectively.