Cyclic mismatch binding ligand CMBL4 binds to the 5'-T-3'/5'-GG-3' site by inducing the flipping out of thymine base

Sanjukta Mukherjee¹, Chikara Dohno¹, Kaori Asano² and Kazuhiko Nakatani¹,*

¹Department of Regulatory Bioorganic Chemistry, The Institute of Scientific and Industrial Research (ISIR), Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan and ²Comprehensive Analysis Center, The Institute of Scientific and Industrial Research (ISIR), Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Received May 15, 2016; Revised July 15, 2016; Accepted July 19, 2016

ABSTRACT

A newly designed cyclic bis-naphthyridine carbamate dimer CMBL4 with a limited conformational flexibility was synthesized and characterized. Absorption spectra revealed that two naphthyridines in CMBL4 were stacked on each other in aqueous solutions. The most efficient binding of CMBL4 to DNA was observed for the sequence 5'-T-3'/5'-GG-3' (T/GG) with the formation of a 1:1 complex, which is one of possible structural elements involved in the higher order structures of (TGG)ₙ repeat DNA triggering the genome microdeletion. Surface plasmon resonance assay also showed the binding of CMBL4 with TGG repeat DNA. Potassium permanganate oxidation studies of CMBL4-bound duplex containing the T/GG site showed that the CMBL4-binding accelerated the oxidation of thymine at that site, which suggests the flipping out of the thymine base from a π-stack. Preferential binding was observed for CMBL4 compared with its acyclic variants, which suggests the marked significance of the macrocyclic structure for the recognition of the T/GG site.

INTRODUCTION

DNA encodes the genetic information necessary for life (1). DNA structures and sequences are allied intrinsically with many vital biological processes (2,3). Errors in polymerase reactions, chemical degradations of nucleotide bases, and several physical factors (UV or ionizing radiations) are responsible for the formation of imperfect DNA structures with damaged nucleotide bases, abasic sites, bulge bases and mismatched base pairs (4-9). Furthermore, the secondary structures of trinucleotide repeat DNA causing neurological diseases involve a number of mismatched base pairs and are known to be responsible for repeat expansion (10-15). Alteration of gene expression due to changes in DNA sequences can be lethal if they produce a nonfunctional variant of an essential protein. Sequence-specific recognition of these damaged DNA structures by small molecules is a promising tool and is an active research field in molecular biology providing a basis for novel chemotherapeutic agents (16-27).

So far our group has succeeded in synthesizing a series of compounds named mismatch-binding ligands (MBLs) containing two heterocycles, typically 1,8-naphthyridine moieties linked by a linker either at (a) C-2 or (b) C-7 positions (Figure 1A). These molecules were selectively bound to the characteristic DNA sequences involving bulge or mismatched base pairs (G-G, G-A, A-A or C-C) through the formation of selective hydrogen bonding with nitrogenous bases and stacking of the naphthyridine units with the neighboring base pairs (23,28-39).

Linkers connecting two heterocycles play an important role in offering a conformation that is suitable for the simultaneous binding of two heterocycles to the target nucleotide bases (36). Due to the increased conformational freedom of two heterocycles connected by a long linker, the necessary conformational changes for the binding should be accompanied by the penalty of entropy loss. To minimize the extra energy required for the conformational change of MBLs for the complex formation with DNA, we decided to design cyclic MBLs (CMBLs), where two heterocycles are connected by linkers at both (a) C-2 and (b) C-7 positions, to provide conformational restrictions to the dynamic motion of the two heterocycles.

The class of macrocyclic compounds where a suitable bridging unit and spacer group connects nonadjacent positions of aromatic units is known as cyclophanes (40,41). Several cyclophane derivatives have been effectively utilized for host–guest complexation and undergo selective interactions with nucleotides, amino acids, nucleic acids and proteins (21,41–54), among which the water-soluble functional cyclophanes are especially interesting due to their better biomimetic scope. Cyclobisintercalands (CBIs)
are the most extensively studied water-soluble macrocycles belonging to the class cyclophane; they comprise homo- or heterodimeric planar aromatic units (such as acridine, naphthalene, phenanthridine, phenazine, tetracationic porphyrin, anthracene, imidazolium, bipyridine and/or biphenyl) (42–54). Two interactive aromatic units of CBIs stack between two separate base pairs for the recognition of nucleic bases and mispaired DNA sites, such as mismatches, bulges, and an abasic site.

In the course of our studies on hydrogen-bonding-mediated sequence-specific recognition of nucleic acids by synthetic ligands and to obtain deeper insights into the interaction of the bis-naphthyridine compound with nucleic acids, we herein reported the synthesis, photophysical properties and DNA-binding activity of CMBL4. We found that CMBL4 was selectively bound to the sequence 5′-T-3′/5′-GG-3′ (T/GG) with 1:1 stoichiometry. Potassium permanganate oxidation showed that thymine (T) in the segment was flipped out from the T/GG stack. Since the T/GG site is an important structural element of (TGG)n repeat causing microdeletion (55–58), CMBL4 could be a useful probe for the studies on (TGG)n repeat and related disorders. We investigated the binding of CMBL4 with TGG repeat DNA by using a surface plasmon resonance (SPR) assay, which clearly showed the binding of CMBL4 to TGG repeat DNA.

MATERIALS AND METHODS

Synthesis and conformation analysis of CMBL4

CMBL4 was synthesized from tert-butoxycarbonyl (Boc) protected methylcarbamoylnaphthyridine dimer (MCND).

Measurements of absorption and fluorescence spectra

UV and fluorescence spectra were measured in 10 mM sodium cacodylate buffer (pH 7.0) and 100 mM NaCl using BECMAN COULTER DU800 UV/Visible spectrophotometer and JASCO FP-8500 spectrofluorometer respectively. Fluorescence spectra were measured at absorption maxima.

Measurements of thermal denaturation profiles of duplexes

Thermal denaturation profiles of the duplexes 5′-d(GTC CAG X GCA ACG)-3′/5′-d(GTG TGC YZ CTG GAC)-3′ and 5′-d(GT CAG WX GCA ACG)-3′/5′-d(GT TGC YZ CTG GAC)-3′ with general formula d(5′-X-3′/5′-YZ-3′) and d(5′-WX-3′/5′-YZ-3′) respectively (5 μM each strand) were measured in sodium cacodylate buffer (10 mM, pH 7) containing sodium chloride (100 mM) using SHIMADZU UV-2700 spectrometer equipped with SHIMADZU TMSPC-8 temperature controller using a 1 cm path length cell. The absorbance of the samples was monitored at 260 nm from 2°C to 95°C with heating rate of 1°C/min. Tm values were calculated by using the median or differential method.
Circular dichroism (CD) measurements

Circular dichroism (CD) measurements of the duplex containing 5′-T-3′/5′-GG-3′ (d1/d2) were carried out on a J-725 CD spectropolarimeter (JASCO) using a 1.0 cm path length cell. CD titration spectra of DNAs (7.5 μM) were measured while titrating with ligand (0, 4, 6, 8, 10, 12, 14, 16 and 18 μM) at ambient temperature in 10 mM sodium cacodylate (pH 7.0) containing 100 mM sodium chloride.

Cold-spray ionization time-of-flight mass (CSI-TOF-MS) measurements of ligand–DNA complex

Samples were prepared by mixing d1/d2 DNA duplex (20 μM) and ligand (40, 60 or 80 μM) in 50% MeOH in water containing 0.1 M ammonium acetate. Mass spectra were obtained using BRUKER microTOF II mass spectrometer. Spray temperature was cooled at −10°C during the injection with a sample flow rate of 0.18 ml/h. Nitrogen gas was used as a desolvation gas as well as a nebulizer.

Matrix-assisted laser desorption/ionization time-of-flight mass (MALDI-TOF-MS) measurements

MALDI-TOF-MS spectra of DNA oligomers were measured using BRUKER ultraflex III. DNA oligomers were mixed with HPA (3-Hydroxy picolinic acid) (0.5 μM; 10 mg/ml)/DAC (Diammonium citrate) (0.5 μM; 1 mg/ml) matrix, which were applied to anchor chip followed by ionization using 60–70% laser irradiation.

Surface plasmon resonance (SPR) assay

Surface plasmon resonance (SPR) single cycle kinetics assay was performed using Biacore T200 platform (GE Healthcare, Life Science). Immobilization of the oligomers on the Series S sensor chip SA surface was carried out using avidin-biotin coupling in HEPES -NaCl running buffer (HBS-N) (0.01 M, HEPES, 0.15 M NaCl, pH 7.4). 5′-Biotinylated oligonucleotides (10 μM aqueous stock) were diluted to 0.2 μM in 10 mM HEPES−500 mM NaCl, was injected to reach the response of around 500 RU. Blank immobilization was performed in the flow cell 1 to permit reference subtraction. Ligand solution (1 mM aqueous) was diluted using HBSEP+ buffer (0.01 M HEPES, 0.15 M NaCl, 3.0 mM EDTA, pH 7.4, 0.005%(v/v) Surfactant P20). Sensorgrams were obtained in the concentration range of 0.25–4.0 μM. All sensorgrams were corrected by reference subtraction of blank flow cell response and buffer injection response.

Reactions and analysis methods used for the Investigation of (T/GG) binding site

Potassium permanganate (KMnO4) was prepared as a 2 mM stock solution. KMnO4 oxidation reaction of d1/d2 DNA duplexes (5 μM each) were carried out with KMnO4 (0.1 mM) in sodium cacodylate buffer (10 mM, pH 7) containing sodium chloride (100 mM) in presence and absence of ligand at 0°C for a duration of 180 min. HPLC profiles were recorded with injection of 50 μl of reaction mixture after every 30 min. Isolated oxidized product was treated with piperidine at 90°C for 30 min. Phosphorylated termini of the cleaved DNAs were removed by the alkaline phosphatase treatment at 37°C for 30 min. HPLC profiles were recorded using CHEMCOBOND 5-ODS-H column and acetoniitrile/0.1M triethylammonium acetate (TEAA) as mobile phase. Adenosine was added as an internal standard (adenosine is eluted at retention time of 5.9 min).

RESULTS

Molecular design, synthesis and characterization of CMBL4

CMBL4 consisted of two 2-amino-1,8-naphthyridine heterocycles connected by two linkers at both (a) C-2 and (b) C-7 positions to provide conformational restrictions to the dynamic motion of two heterocycles (Figure 1A and B). The heterocycle had a hydrogen-bonding surface fully complementary with that of guanine (Figure 1D). Both linkers contained a secondary amino group, which functions as a positively charged site upon protonation at neutral pH to gain water solubility and ensure electrostatic attraction with the negatively charged phosphate groups of DNA. Arrangement of two heterocycles in CMBL4 could be either parallel or antiparallel in the hydrogen-bonding direction, and stacked or unstacked conformation (Figure 1C and Supplementary Figure S1). In parallel arrangement, with either stacked or unstacked conformations, hydrogen-bonding surfaces of two N-acyl-2-amino-1,8-naphthyridine units in CMBL4 point in the same direction, which provides the possibility to bind to two consecutive guanines (GG) if these Gs were involved in the structure with a dynamic motion and were therefore kinetically accessible to CMBL4 (Figure 1E). Conformational analysis of CMBL4 using an Amber* force field with mixed torsional/low mode sampling method (Maestro version 10.0.0.13; Schrödinger Inc.) showed that the most stable conformation and subsequent nine conformations possessed parallel-stacked structures (Supplementary Figure S2; Tables S1 and S2). Energy minimization of the lowest energy conformation of CMBL4 by DFT calculation (wb97xd/6-311+(G)) showed a parallel-stacked structure (Supplementary Figure S3), which suggests the possibility of binding to two consecutive guanines.

CMBL4 was synthesized from tert-butoxycarbonyl Boc-protected MCND (36) in the following three steps: deprotection of the carbomethoxy group, amide coupling and finally deprotection of the Boc group. Details of the synthesis (Supplementary Figure S4) and the characterization of products are given in the Supplementary Data.

Conformational studies of CMBL4 by absorption and fluorescence spectra

NCD and MCND contained two 1,8-naphthyridine units linked in an acyclic framework at either C-2 or C-7 positions, respectively, and monomer, which contained a 1,8-naphthyridine scaffold common to all three compounds (CMBL4, NCD and MCND) (Figure 2A). The absorption spectra of CMBL4 in aqueous buffer solution (pH 7) containing 100 mM sodium chloride was compared with those of parent molecules NCD, MCND, and monomer (Figure 2B). Intense absorption band at 318 nm with a shoulder peak at 332 nm was observed for CMBL4, whereas dual
Figure 2. (A) Structures of NCD, MCND, and a monomer unit. (B) Absorption and (C) fluorescence spectra excited at absorption maxima (318 nm for CMBL4 and NCD, 320 nm for MCND and 332 nm for monomer) of CMBL4 (blue), NCD (red), MCND (green) and monomer (black) (10 μM) in 10 mM sodium cacodylate buffer (pH 7.0) and 100 mM NaCl. Fluorescence of monomer is reduced by a factor 10 in the intensity for clarity. (D) Absorption spectra of CMBL4 with variable concentrations (10, 25, 50 and 100 μM). Vertical arrow indicates the increased ligand concentration. The inset shows the linear correlation of absorption at 318 nm against CMBL4 concentration with correlation factor of >0.99. (E) Absorption spectra of CMBL4 with increasing percentage of methanol (0, 10, 25, 50, 79 and 100% (dotted line)). Vertical arrow indicates the increased percentage of methanol. (F) Tₘ increase of the duplex d₁/d₂ (5 μM) in the presence of CMBL4, NCD, MCND and monomer (10 μM) in 10 mM sodium cacodylate buffer (pH 7.0) and 100 mM NaCl.

Fluorescence spectra of these compounds reflected the differences in the absorption spectra (Figure 2C). While monomer emitted broad and intense fluorescence that peaked at 364 nm, CMBL4 showed emissions not only at 364 nm, but also at 438 nm. This dual emission was almost similar to that of MCND with a higher I₄₃₈/I₃₆₄ ratio of 0.49 compared to 0.36 for MCND. The fluorescence spectrum of NCD was observed neither at 364 nm nor 438 nm, but at 390 nm. Since the absorption spectra of NCD was similar to that of monomer, the broad and structureless emission of NCD that peaked at 390 nm is likely due to excimer fluorescence. On the other hand, the emission at 438 nm observed for MCND and CMBL4 is due to the excimer fluorescence, but is due to the characteristic ground electronic state of the molecule, which is likely due to the intramolecular stacking of two naphthyridines in an aqueous buffer solution. The absorption at 318 nm of CMBL4 did not shift regardless to the concentration and fulfilled the Lambert–Beer law (Figure 2D and inset), whereas the absorption spectra markedly changed upon increasing the solution hydrophobicity from 0% to 100% methanol. The absorption spectrum of CMBL4 in 100% methanol resembled that of NCD in 100% aqueous buffer solution with dual absorption bands at 320 and 332 nm (Figure 2E). In fact, the fluorescence spectrum of CMBL4 in methanol resembled that of NCD in aqueous buffer with a broad peak at 400 nm (Supplementary Figure S5B). In addition, ¹H NMR spectra of CMBL4 showed a significant up-field shift of signals in D₂O compared with those in CD₃OD (Figure 3). These spectroscopic analyses indicated that naphthyridine moieties in CMBL4 are in intramolecularly stacked conformations in aqueous solutions.

Sequence dependency of CMBL4 binding evaluated by thermal melting temperature (Tₘ) analyses

As CMBL4 consists of two structural units of N-acyl-2-amino-1,8-naphthyridine with three hydrogen-bonding groups fully complementary to that of G (Figure 1D), we anticipated the possibility of CMBL4 binding to two consecutive Gs, if these Gs were kinetically accessible for
Complex formation between CMBL4 and T/GG site

The formation of the CMBL4-bound d1/d2 duplex was confirmed by CSI-TOF-MS. In the presence of 40 μM CMBL4 (2 molar equivalent to d1/d2 duplex) (Figure 4 and Supplementary Figure S9A and B) ions corresponding to the 1:1 complex between CMBL4 and d1/d2 duplex were observed as a predominant complex ([d1] / [d2]-CMBL4)\(^{5-}\) (m/z: found 1752.592; calcd. 1752.591) and [d1] / [d2]-CMBL4\(^{6-}\) (m/z: found 1460.320; calcd. 1460.325) (Figure 4). Formation of 1:2 [d1] / [d2]-2CMBL4\(^{5-}\) (m/z: found 1855.91; calcd. 1855.90) complex was detectable with increased concentrations of CMBL4 (60 and 80 μM) (Supplementary Figure 9C and D), but its intensity was much smaller than that of the 1:1 complex, which was still the dominant species. When the ligand concentration was increased, a concomitant increase of the ion peak corresponding to the unbound ligand was observed with the almost disappearance of the ion peak corresponding to d1 / d2. These data clearly indicated that CMBL4 bound to d1 / d2 duplex preferentially with 1:1 stoichiometry.

In the presence of NCD, formation of a much less intense peak corresponding to the ions [d1] / [d2]-NCD\(^{5-}\) (m/z: found: 1750.001; calcd: 1749.991) (Supplementary Figure S9E) indicated the reduced binding affinity of acyclic compound NCD to d1 / d2 duplex compared to the cyclic variant CMBL4 (Supplementary Figure S9B). Another acyclic compound, MCND, did not produce any ion peaks corre-
Table 2. ΔTm values of the different bulge/mismatch duplexes d(5′-X-3′/5′-YZ-3′) with CMBL4

| Entry | X/YZ       | Tm (–) | Tm (+) | ΔTm  |
|-------|------------|--------|--------|------|
| 1     | X/GG       | 43.8 (0.7) | 52.8 (0.8) | 9.0 (0.5) |
| 2     | T/GG       | 45.3 (0.2) | 56.7 (0.3) | 0.1 (0.1) |
| 3     | A/GG       | 48.0 (0.2) | 58.4 (0.2) | 1.9 (0.1) |
| 4     | G/GG       | 48.2 (0.2) | 58.2 (0.2) | 3.2 (0.2) |
| 5*    | T or C/YY  | 49.1 (0.2) | 59.1 (0.5) | 1.3 (0.4) |
| 6*    | T/TT       | 41.0 (0.3) | 51.0 (0.5) | 2.0 (0.3) |
| 7**   | C/CC       | 43.8 (0.2) | 54.8 (0.3) | 1.8 (0.2) |
| 8*    | T/II       | 43.8 (0.1) | 53.8 (0.1) | 2.0 (0.1) |
| 9*    | C/AA       | 45.0 (0.2) | 56.0 (0.2) | 1.8 (0.2) |
| 10    | C/TT       | 44.0 (0.1) | 54.0 (0.1) | 2.0 (0.1) |
| 11    | C/CC       | 43.2 (0.5) | 53.2 (0.5) | 3.2 (0.5) |
| 12*   | T/YG or GZ | 45.3 (0.2) | 55.3 (0.5) | 1.3 (0.4) |
| 13*   | T/AG       | 46.0 (0.4) | 56.0 (0.8) | 2.0 (0.5) |
| 14*   | T/TG       | 47.0 (0.1) | 57.0 (0.1) | 2.0 (0.1) |
| 15*   | T/GA       | 48.3 (0.2) | 58.3 (0.2) | 1.0 (0.2) |
| 16*   | T/GT       | 41.0 (0.1) | 51.0 (0.1) | 3.0 (0.5) |
| 17*   | T/GC       | 43.5 (0.5) | 53.5 (0.5) | 2.5 (0.5) |

*Thermal denaturation profile of the duplex 5′-d(GTC CAG X GCA ACG)-3′/5′-d(CGT TGC YZ CTG GAC)-3′ (5 μM) was measured in sodium cacodylate buffer (10 mM, pH 7) containing sodium chloride (100 mM). The temperature was increased at a rate of 1°C/min. Tm values (°C) were calculated by median method. All measurements were made three times and standard deviations are shown in parentheses.

bTm values of the oligomers.
cTm values of oligomers in the presence of ligands (10 μM).
dΔTm was calculated as the difference between Tm(+) and Tm(–).

**Unable to determine the Tm value due to the biphasic nature of the Tm curve.

Table 3. ΔTm values of the different bulge/mismatch duplexes d(5′-WX-3′/5′-YZ-3′) with CMBL4

| Entry | WX/YZ       | Tm (–) | Tm (+) | ΔTm  |
|-------|------------|--------|--------|------|
| 1     | CC/GG      | 68.0 (0.5) | 66.6 (0.1) | –1.4 (0.7) |
| 2     | TT/GG      | 52.5 (0.1) | 53.1 (0.1) | 0.6 (0.1) |
| 3     | TG/TG      | 48.4 (0.1) | 48.2 (0.1) | –0.2 (0.4) |
| 4     | TG/G       | 46.0 (0.4) | 46.9 (0.1) | 0.9 (0.4) |
| 5     | G/TG       | 48.5 (0.2) | 48.9 (0.1) | 0.4 (0.4) |
| 6     | GT/GT      | 47.8 (0.2) | 47.9 (0.1) | 0.1 (0.3) |
| 7     | GT/G       | 46.4 (0.2) | 47.9 (0.1) | 1.5 (0.3) |
| 8     | G/GT       | 46.3 (0.3) | 47.6 (0.3) | 1.3 (0.3) |
| 9     | G/G        | 54.2 (0.4) | 53.3 (0.1) | –0.9 (0.4) |

*Thermal denaturation profile of the duplex 5′-d(GTC CAG WX GCA ACG)-3′/5′-d(CGT TGC YZ CTG GAC)-3′ (5 μM) was measured in sodium cacodylate buffer (10 mM, pH 7) containing sodium chloride (100 mM). The temperature was increased at a rate of 1°C/min. Tm values (°C) were calculated by median method. All measurements were made three times and standard deviations are shown in parentheses.

bTm values of the oligomers.
cTm values of oligomers in the presence of ligands (10 μM).
dΔTm was calculated as the difference between Tm(+) and Tm(–).

Figure 4. CSI-TOF-MS of duplex d1/d2 (20 μM each) in 50% aqueous methanol and 100 mM ammonium acetate in the presence of CMBL4 (40 μM). The sample solution was cooled to –10°C during the injection with a flow rate of 0.18 ml/h.
duplex indicate the necessity of the cyclic structure for the efficient formation of the complex with duplex d₁/d₂.

The CD spectra of the duplex d₁/d₂ were measured in the presence and absence of CMBL4 to see if any structural changes could be induced upon ligand binding (Figure 5A). An induced CD band was observed at 353 nm. The molar ellipticity of the positive peak at 353 nm increased with decreasing molar ellipticity at 280 nm as the ligand concentration increased (0, 4, 6, 8, 10, 12, 14, 16 and 18 μM). The association constant (Kₐ) of CMBL4 for the duplex d₁/d₂ was calculated to be 4.3 × 10⁵ M⁻¹ (Kₐ = 2.3 μM) by curve fitting to the 1:1 binding isotherm (Figure 5B).

Chemical probing of the binding site of CMBL4

Upon ligand binding, the T/GG site possibly induced the flipping out of the T base. We investigated the reactivity of the T base in the T/GG site of the d₁/d₂ duplex toward potassium permanganate oxidation by reverse-phase HPLC (Figure 6, Supplementary Figure S10) (32, 48, 59, 60). Oxidation of the C5-C6 double bond of the T base with potassium permanganate forming 5,6-dihydroxy-5,6-dihydrothymine (thymine glycol, Tg) would be more feasible for the T base than the Tg site of the duplex (61). Treating the d₁/d₂ duplex (5 μM) with potassium permanganate (100 μM) in the presence of CMBL4 (10 μM) at 0°C for 180 min in buffer solution (pH 7.0) resulted in the formation of one major product at a retention time of 14.6 min with concomitant decrease of the d₁ strand (Figure 6D). The resultant product was isolated by HPLC, analyzed by MALDI-TOF MS, and identified as the Tg-containing d₁ (d₁(Tg)) (m/z: found 3994.67, calcd. 3994.62 [M+H]+) (Supplementary Figure S13A). The isolated d₁(Tg) was then treated with piperidine at 90°C for 30 min to give new products (Figure 6E), which suggests strand cleavage of the d₁(Tg) at the site of oxidation. MALDI-TOF MS analysis of resultant products indicated that these were the phosphorylated oligomers 5′-d(GTC CAG)-PO₃H-3′ (GTCCAGp) (m/z: found 1873.03, calcd. 1873.20 [M+H]+) and 5′-HO₃P-d(GCA ACG)-3′ (pGCAACG) (m/z: found 1882.12, calcd. 1882.21 [M+H]+) (Supplementary Figure S13B and C). Phosphorylated termini of the two identified oligomers were hydrolyzed by treatment with alkaline phosphatase to produce 5′-d(GTC CAG)-3′ and 5′-d(GCA ACG)-3′. The identities of these two oligomers were firmly confirmed by the same retention
DISCUSSION
We have designed bis-naphthyridine cyclophane CMBL4 based on the CMBL concept. The significance of two consecutive Gs opposite the single T for the CMBL4 binding was confirmed by the sequence-specific binding studies using thermal melting temperature ($T_m$) analysis. The $T_m$ of duplex $d_1/d_2$ (T/GG) in the presence of CMBL4 and its acyclic variants (NCD and MCND) suggested the importance of a cyclized structure. CSI-TOF-MS of the CMBL4-bound $d_1/d_2$ duplex confirmed the formation of a 1:1 $d_1/d_2$-CMBL4 complex. The formation of a 1:2 $d_1/d_2$-CMBL4 complex was detectable with much smaller intensity compared with the 1:1 complex. Even after increasing the concentration of CMBL4, we observed a 1:1 complex as the predominant stoichiometry. In the presence of NCD, formation of a less-intense peak corresponding to the 1:1 $d_1/d_2$-NCD complex was observed. However, the ion peak corresponding to the formation of a $d_1/d_2$-MCND complex was not detected. These data clearly support the improved binding affinity of CMBL4 compared with its acyclic variants. Flipping out of the T from the $\pi$-stack at the T/GG site upon CMBL4 binding was confirmed by the increasing rate of potassium permanganate oxidation of T in the presence of CMBL4. A much reduced rate of formation of the oxidized product $d_1$(Tg) in the presence of acyclic variants (NCD and MCND) also strongly support the increasing affinity for the binding of CMBL4 compared with its acyclic variants. The dissociation constant of CMBL4 with T/GG was calculated to be 2.3 $\mu$M. The binding affinity is higher than that of previously reported G-bulge binding molecule ($K_d = 29 \mu$M) (28), but is lower than that of the G-G mismatch-binding molecule (29). The T/GG site is one of the possible structural elements involved in the higher order structures of (TGG)$_n$ repeat DNA. The long TGG repeat has been revealed to trigger recurrent microdeletion in the long arm of paternal chromosome 14 (14q32.2) (49–52). The ligand binding to (TGG)$_n$ repeat may have potential to suppress the microdeletion.

Several groups have been developing small molecules binding to the characteristic sequences of DNA/RNA by applying different approaches. Zimmerman et al. focused on the rational design of small molecules comprised of melamine-acridine conjugates to target CTG and CUG repeats causing myotonic dystrophy type 1 (MD1) (24). Miller et al. focused on the development of small molecules for CUG repeats by dynamic combinatorial chemistry, which is based on the techniques of molecular design and a novel combinatorial method of small-molecule evolution (25). In contrast, Disney et al. developed small molecules based on high-throughput screening from chemical library to target selective genome sequences such as CGG repeat RNA (26,27). Our group has also reported MBLs that can stabilize the secondary structures of trinucleotide repeats. Among these MBLs, NCD binds with different trinucleotide repeat sequences with decreased efficiency along with the most strongly binding CGG repeat, probably because of the conformational freedom of the longer linker length (33). With the much shorter three-atom linkers, MCND showed improved binding to the d(GAAA)$_n$ repeats (36).

We have evaluated the binding of CMBL4 to TGG repeat DNA and UGG repeat RNA by using SPR assay. SPR single-cycle kinetics analysis using d(TGG)$_n$ and r(UGG)$_n$ immobilized sensor chip showed a concentration-dependent response for CMBL4 binding (Supplementary Figure S14a) with both repeats. Although the quantitative binding analysis is difficult for the repeat sequence because of the ambiguity of the binding stoichiometry and heterogeneity of the multiple binding sites, the SPR data clearly showed the binding of CMBL4 to both repeats. The SPR response for the r(UGG)$_n$ sequence was smaller than for the repeat DNA, which suggests that CMBL4 bound to UGG repeat RNA with lower binding affinity than TGG repeat DNA. The binding mode of CMBL4 to the RNA repeats is not necessarily the same as for the DNA repeats, which is currently unknown. These data indicated that the CMBL4 binding to T/GG site could be extended to the d(TGG) repeat DNA.

CONCLUSION
In conclusion, a new molecule CMBL4 designed based on a CMBL concept was synthesized and confirmed to bind to the T/GG site in a sequence-specific manner. The cyclophane-type structure of CMBL4 provided a characteristic electronic structure of stacked naphthyridines in aqueous solutions. The binding of CMBL4 to the T/GG site was much more efficient than that of its acyclic variants, and the binding of CMBL4 to the T/GG site induced the T to flip out from the $\pi$-stack. The 1:1 complex between CMBL4 and the T/GG site was shown by CSI-TOF-MS. These studies revealed the potential of CMBLs for the recognition of noncanonical DNA structures, which was not efficient using acyclic MBL variants. SPR single-cycle kinetics analyses suggested the binding of CMBL4 with TGG repeats, which indicated that the CMBL4 binding to T/GG site could be extended to the TGG repeat DNA. A molecular modeling simulation of CMBLs suggests that the orientation of two naphthyridines could be modulated by their linker length and structures. These advanced molecular designs may expand the possibilities of CMBLs binding to
other DNA or possibly RNA structures of biological significance to explore more complex biological systems.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

JSPS KAKENHI Grant-in-Aid for Specially Promoted Research [JP26000007 to K.N.]. Funding for open access charge: JSPS KAKENHI Grant-in-Aid for Specially Promoted Research [JP26000007].

Conflict of interest statement. None declared.

REFERENCES

1. Crick,F. (1970) Central dogma of molecular biology. Nature, 227, 561–563.
2. Latchman,D.S. (2008) Eukaryotic Transcription Factors. 5th edn. Elsevier/Academic Press.
3. Mitchell,P.J. and Tjian,R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science, 245, 371–378.
4. Kunkel,T.A. (2004) DNA replication fidelity. J. Biol. Chem., 279, 16895–16898.
5. Wang,W., Hellina,H.W. and Beele,L.S. (2011) Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. Proc. Natl. Acad. Sci. U.S.A., 108, 17644–17648.
6. Gates,K.S. (2009) An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals. Chem. Rev. Toxicol., 22, 1747–1760.
7. De Bont,R. and van Larebeke,N. (2004) Endogenous DNA damage in humans: a review of quantitative data. Mutagenesis, 19, 169–185.
8. Lighten,M., Goyon,C., Schultz,N.P., Treco,D., Szostak,J.W., Haber,J.E. and Nicolas,A. (1990) Detection of heteroduplex DNA molecules among the products of saccharomyces cerevisiae meiosis. Proc. Natl. Acad. Sci. U.S.A., 87, 7653–7657.
9. Friedberg,E.C. (2003) DNA damage and repair. Nature, 421, 436–440.
10. Wells,R.D. and Warren,S.T. (1998) Genetic Instabilities and Hereditary Neurological Diseases. Academic Press, San Diego.
11. Cummings,C.J. and Zoghbi,H.Y. (2000) Trinucleotide repeats: mechanisms and pathophysiology. Annu. Rev. Genomics Hum. Genet., 1, 281–328.
12. Pearson,C.E., Edamura,K.N. and Cleary,J.D. (2005) Repeat instability: mechanisms of dynamic mutations. Nat. Rev. Genet., 6, 729–742.
13. Orr,H.T. and Zoghbi,H.Y. (2007) Trinucleotide repeat disorder. Annu. Rev. Neurosci., 30, 575–621.
14. Lee,D.Y. and McMurray,C.T. (2014) Trinucleotide expansion in disease: why is there a length threshold? Curr. Opin. Genet. Dev., 26, 131–140.
15. Mirkin,S.M. (2007) Expandable DNA repeats and human disease. Nature, 447, 932–940.
16. Rajski,S.R. and Williams,R.M. (1998) DNA cross-linking agents as antitumor drugs. Chem. Rev., 98, 2723–2795.
17. Neidle,S. (2001) DNA minor-groove recognition by small molecules. Nat. Prod. Rep., 18, 291–309.
18. Hamilton,P.L. and Arya,D.P. (2012) Natural product DNA major groove binders. Nat. Prod. Rep., 29, 134–143.
19. Song,G. and Ren,J. (2010) Recognition and regulation of unique nucleic acid structures by small molecules. Chem. Commun., 46, 7283–7294.
20. Banerjee,S., Veale,E.B., Phelan,C.M., Murphy,S.A., Tocci,G.M., Gillespie,L.I., Frimannsson,D.O., Kelly,J.M. and Gunnaugsson,T. (2013) Recent advances in the development of 1,8-naphthalimide based DNA targeting binders, anticancer and fluorescent cellular imaging agents. Chem. Soc. Rev., 42, 1601–1618.
21. Granzhahn,A., Koteria,N. and Teulade-Fichou,M.-P. (2014) Finding needles in a basestack: recognition of mismatched base pairs in DNA by small molecules. Chem. Soc. Rev., 43, 3630–3665.
22. Gurova,K. (2009) New hopes from old drugs: revisiting DNA-binding small molecules as anticancer agents. Future Oncol., 10, 1685–1704.
23. Nakatani,K., Sando,S. and Saito,I. (2001) Scanning of guanine–guanine mismatches in DNA by synthetic ligands using surface plasmon resonance. Nat. Biotechnol., 19, 51–55.
24. Nguyen,L., Luu,L. M., Peng,S., Serrano,J. F., Chan,H. Y. E. and Zimmerman,S. C. (2015) Rationally designed small molecules that target both the DNA and RNA causing Myotonic Dystrophy Type 1. J. Am. Chem. Soc., 137, 14180–14189.
25. Gareiss,P.C., Sobezak,K., McNaughton,B. R., Palde,P. B., Thornton,C. A. and Miller,B. L. (2008) Dynamic Combinatorial selection of molecules capable of inhibiting the (CUG) repeat RNA-MBN1 interaction in vitro: discovery of lead compounds targeting Myotonic Dystrophy (DM1). J. Am. Chem. Soc., 130, 16254–16261.
26. Velagapudi,S. P., Gallo,S. M. and Disney,M. D. (2014) Sequence-based design of bioactive small molecules that target precursor microRNAs. Nat. Chem. Biol., 110, 291–297.
27. Disney,M. D., Liu,B., Yang,W.-Y., Sellier,C., Tran,T., Charlet-Berguerand,N. and Childs-Disney,J. L. (2012) A small molecule that targets r(CGG)420 and improves defects in Fragile-X-Associated Tremor Ataxia Syndrome. ACS Chem. Biol., 7, 1711–1718.
28. Nakatani,K., Sando,S. and Saito,I. (2000) Recognition of a single guanine bulge by 2-acylamino-1,8-naphthyridine. J. Am. Chem. Soc., 122, 2172–2177.
29. Nakatani,K., Sando,S., Kumasawa,H., Kikuchi,J. and Saito,I. (2001) Recognition of guanine-guanine mismatches by the dimeric form of 2-amino-1,8-naphthyridine. J. Am. Chem. Soc., 123, 12650–12657.
30. Nakatani,K., He,H., Uno,S., Yamamoto,T. and Dohno,C. (2008) Synthesis of dimeric 2-amino-1,8-naphthyridine and related DNA-binding molecules. Curr. Protoc. Nucleic Acids Chem., Unit 8.6.
31. Nakatani,K., Hagihara,S., Goto,Y., Kobori,A., Hagihara,M., Hayashi,G., Kyo,M., Nomura,M., Mishima,M. and Kojima,C. (2005) Small-molecule ligand induces nucleotide flipping in (CAG)6 trinucleotide repeats. Nat. Chem. Biol., 1, 39–43.
32. Peng,T., Dohno,C. and Nakatani,K. (2006) Mismatch-binding ligands function as a molecular glue for DNA. Angew. Chem. Int. Ed., 45, 5623–5626.
33. Peng,T. and Nakatani,K. (2005) Binding of Naphthyridine carbamate dimer to the (CAG)n repeat results in the disruption of the G-C base pairing. Angew. Chem. Int. Ed., 44, 7280–7283.
34. Hagihara,S., Kumasawa,H., Goto,Y., Hayashi,G., Kobori,A., Saito,I. and Nakatani,K. (2004) Detection of guanine-adenine mismatches by surface plasmon resonance sensor carrying naphthyridine-azaquinoalnone hybrid on the surface. Nucleic Acids Res., 32, 278–286.
35. Nakatani,K. (2009) Recognition of mismatched base pairs in DNA. Bull. Chem. Soc. Jpn., 82, 1055–1069.
36. He,H., Hagihara,M. and Nakatani,K. (2009) A small molecule affecting the replication of trinucleotide repeat d(GAA)n. Chem. Eur. J., 15, 10641–10648.
37. Hong,C., Hagihara,M. and Nakatani,K. (2011) Ligand-assisted complex formation of two DNA hairpin loops. Angew. Chem. Int. Ed., 50, 4390–4393.
38. Dohno,C. and Nakatani,K. (2011) Control of DNA hybridization by photo switchable molecular glue. Chem. Soc. Rev., 40, 5718–5729.
39. Dohno,C., Yamamoto,T. and Nakatani,K. (2009) Photoswitchable unsymmetrical ligand for DNA hetero-mismatches. Eur. J. Org. Chem., 44, 4051–4058.
40. Ferguson,J. (1986) Absorption spectroscopy of sandwich dimers and cyclophanes. Chem. Rev., 86, 957–982.
41. Ramiah,D., Neelakandan,P.P., Nair,A.K. and Avirah,R.R. (2010) Functional cyclophanes: promising hosts for optical biomolecular recognition. Chem. Soc. Rev., 39, 4158–4168.
42. Zimmerman,S.C., Lamberton,C.R., Cory,M. and Fairley,T.A. (1989) Topologically constrained bifunctional intercalators: DNA intercalation by a macrocyclic bisacridine. J. Am. Chem. Soc., 111, 6805–6809.
43. Plantanidou,J., Palm,B.S., Cudic,P., Zinic,M. and Schneider,H.-J. (2004) Interactions of acyclic and cyclic bis-phenanthridinium derivatives with ss- and ds-polynucleotides. Tetrahedron, 60, 6225–6231.
44. Fernandez-Saiz, M., Schneider, H.-J., Sartorius, J. and Wilson, W.D. (1996) A cationic cyclophane that forms a base-pair open complex with RNA duplexes. *J. Am. Chem. Soc.*, 118, 4739–4745.

45. Slama-Schwok, A., Teulade-Fichou, M.-P., Vigneron, J.-P., Taillandier, E. and Lehn, J.-M. (1995) Selective binding of a macrocyclic bisacridine to DNA hairpins. *J. Am. Chem. Soc.*, 117, 6822–6830.

46. Slama-Schwok, A., Peronnet, F., Hantz-Brachet, E., Taillandier, E., Teulade-Fichou, M.-P., Vigneron, J.-P., Best-Belpomme, M. and Lehn, J.-M. (1997) A macrocyclic bis-acridine shifts the equilibrium from duplexes towards DNA hairpins. *Nucleic Acid Res.*, 25, 2574–2581.

47. Baudoin, O., Gonnet, F., Teulade-Fichou, M.-P., Vigneron, J.-P., Tabet, J.-C. and Lehn, J.-M. (1999) Molecular recognition of nucleotide pairs by a cyclo-bis-intercaland-type receptor molecule: a spectrophotometric and electrospray mass spectrometry study. *Chem. Eur. J.*, 5, 2762–2771.

48. David, A., Bleimling, N., Beuck, C., Lehn, J.-M., Weinhold, E. and Teulade-Fichou, M.-P. (2003) DNA mismatch-specific base flipping by a bisacridine macrocycle. *ChemBioChem*, 4, 1326–1331.

49. Bahri, M., Gabelica, V., Granzhahn, A., Teulade-Fichou, M.-P. and Weinhold, E. (2008) Selective recognition of pyrimidine–pyrimidine DNA mismatches by distance-constrained macrocyclic bis-intercalators. *Nucleic Acids Res.*, 36, 5000–5012.

50. Granzhahn, A. and Teulade-Fichou, M.-P. (2009) A fluorescent bisanthracene macrocycle discriminates between matched and mismatch-containing DNA. *Chem. Eur. J.*, 15, 1314–1318.

51. Granzhahn, A., Largy, E., Saettel, N. and Teulade-Fichou, M.-P. (2010) Macrocyclic DNA-mismatch-binding ligands: structural determinants of selectivity. *Chem. Eur. J.*, 16, 878–889.

52. Jourdan, M., Granzhahn, A., Guillot, R., Dumy, P. and Teulade-Fichou, M.-P. (2012) Double threading through DNA: NMR structural study of a bis-naphthalene macrocycle bound to a thymine–thymine mismatch. *Nucleic Acids Res.*, 40, 5115–5128.

53. Neelakandan, P.P. and Ramaiah, D. (2008) DNA-assisted long-lived eximer formation in a cyclophane. *Angew. Chem. Int. Ed.*, 47, 8407–8411.

54. Neelakandan, P.P., Sanju, K.S. and Ramaiah, D. (2010) Effect of bridging units on photophysical and DNA binding properties of a few cyclophanes. *Photochem. Photobiol.*, 86, 282–289.

55. Kozlowskik, P., Sobczak, K. and Krzyzosiak, W.J. (2010) Trinucleotide repeats: triggers for genomic disorders? * Genome Med.*, 2, 29, 1–5.

56. Bena, F., Gimelli, S., Migliavacca, E., Brun-Druc, N., Buiting, K., Antonarakis, S.E. and Sharp, A.J. (2010) A recurrent 14q32.2 microdeletion mediated by expanded TGG repeats. *Hum. Mol. Genet.*, 19, 1967–1973.

57. Zada, A., Mundhofir, F.E.P., Pfundt, R., Leijsten, N., Nillesen, W., Faradz, S.M.H. and de Leeuw, N. (2014) A rare, recurrent, de novo 14q32.2q32.31 microdeletion of 1.1 Mb in a 20-year-old female patient with a maternal UPD(14)-Like phenotype and intellectual disability. *Case Rep. Genet.*, Article ID 530134.

58. Nevado, J., Mergener, R., Palomares-Bralo, M., Souza, K.R., Vallespin, E., Mena, R., Martinez-Glez, V., Mori, M.A., Santos, F., Garcia-Minaur, S. et al. (2014) New microdeletion and microduplication syndromes: A comprehensive review. *Genet. Mol. Biol.*, 37, 210–219.

59. Gogos, J.A., Karayiorgou, M., Aburatani, H. and Kafatos, F.C. (1990) Detection of single base mismatches of thymine and cytosine residues by potassium permanganate and hydroxylamine in the presence of tetralkylammonium salts. *Nucleic Acids Res.*, 18, 6807–6814.

60. Rubin, C.M. and Schmid, C.W. (1980) Pyrimidine-specific chemical reactions useful for DNA sequencing. *Nucleic Acids Res.*, 8, 4613–4619.

61. Hayatsu, H. and Ukitu, T. (1967) The selective degradation of pyrimidines in nucleic acids by permanganate oxidation. *Biochem. Biophys. Res. Commun.*, 29, 556–561.