Published online 8 August 2018

Nucleic Acids Research, 2018, Vol. 46, No. 17 ***8679–8688***
doi: 10.1093/nar/gky704

Modification of the aminopyridine unit of 2′-deoxyaminopyridinyl-pseudocytidine allowing triplex formation at CG interruptions in homopurine sequences

Lei Wang, Yosuke Taniguchi*, Hidenori Okamura and Shigeki Sasaki*

Graduate School of Pharmaceutical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812–8582, Japan

Received June 01, 2018; Revised July 19, 2018; Editorial Decision July 23, 2018; Accepted July 23, 2018

ABSTRACT

The antigene strategy based on site-specific recognition of duplex DNA by triplex DNA formation has been exploited in a wide range of biological activities. However, specific triplex formation is mostly restricted to homo-purine strands within the target duplex DNA, due to the destabilizing effect of CG and TA inversion sites where there is an absence of natural nucleotides that can recognize the CG and TA base pairs. Hence, the design of artificial nucleosides, which can selectively recognize these inversion sites with high affinity, should be of great significance. Recently, we determined that 2-amino-3-methylpyridinyl pseudo-dC (^{3Me}AP-ΨdC) possessed significant affinity and selectivity toward a CG inversion site and showed effective inhibition of gene expression. We now describe the design and synthesis of new modified aminopyridine derivatives by focusing on small chemical modification of the aminopyridine unit to tune and enhance the selectivity and affinity toward CG inversion sites. Remarkably, we have newly found that 2-amino-4-methoxypyridinyl pseudo-dC (^{4OMe}AP-ΨdC) could selectively recognize the CG base pair in all four adjacent base pairs and form a stable triplex structure against the promoter sequence of the human gene including multiple CG inversion sites.

INTRODUCTION

In the past decade, an increasing number of genetic studies demonstrated that RNA regulation by oligonucleotides is essential for the therapeutic application and biochemical technology. Although oligonucleotides for direct interaction with a single-stranded RNA to regulate the amount of RNA are actively investigated, those with a high affinity to duplex DNA can also provide an interesting method to control the amount of RNA. The antigene strategy is based on the sequence-specific recognition of duplex DNA by triplex-forming oligonucleotides (TFOs) at the major groove side, which can modulate gene expression at the transcriptional level. TFOs have been exploited in a wide range of biological activities (1–4), including gene expression regulation (5–7), site-specific DNA damage (8–11), DNA repair, recombination (12,13) and mutagenesis (14,15).

Triplex DNA is formed by the interaction of TFOs with the homo-purine strand consisting of 2′-deoxy guanosine (dG) and 2′-deoxy adenosine (dA) in the target duplex DNA. According to the binding orientation of the phosphate backbone of the TFOs, triplex DNA formations are divided into two types. In a parallel type triplex DNA, a pyrimidine-rich TFO binds to the homo-purine strand with the same direction of the phosphate backbone to form T/AT and C+/GC triplets through two Hoogsteen hydrogen bonds under acidic conditions. In an antiparallel type triplex DNA, a purine-rich TFO binds to the homo-purine strand in the reverse direction to the phosphate backbone to form the A/AT and G/GC triplets through two reverse Hoogsteen hydrogen bonds under neutral conditions (Figure 1A). However, there is no natural nucleoside can that form stable hydrogen bonds with thymidine (T) and 2′-deoxy cytidine (dC) in TA and CG base pairs from the major groove side, which are known as inversion sites (Figure 1B). Their presence in the target purine DNA sequences drastically restricts the formation of stable triplex DNA. Hence, the design of an artificial nucleoside, which can selectively recognize these inversion sites with a high affinity, should be of great significance. A variety of artificial nucleosides have already been developed to selectively recognize inversion sites (16–24). Nevertheless, the achievement of potential analogues, which can selectively recognize inversion

*To whom correspondence should be addressed. Tel: +81 92 642 6615; Fax: +81 92 642 6615; Email: taniguch@phar.kyushu-u.ac.jp
Correspondence may also be addressed to Shigeki Sasaki. Tel: +81 92 642 6615; Fax: +81 92 642 6615; Email: sasaki@phar.kyushu-u.ac.jp

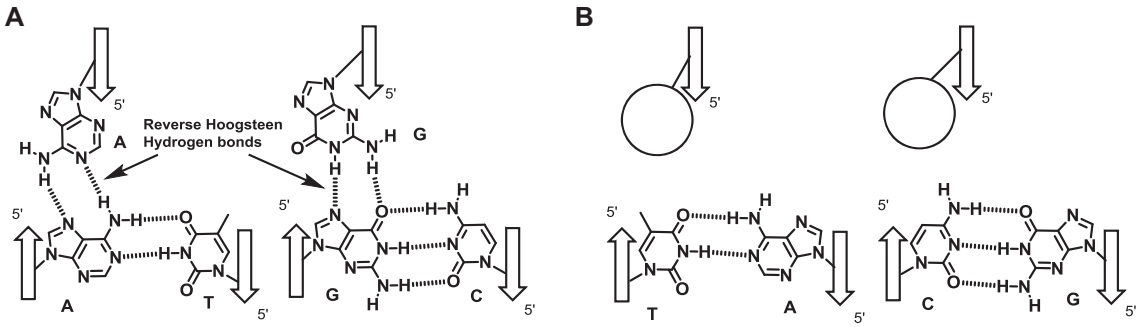

Figure 1. The schematic base triplet structure of antiparallel type triplex DNA (**A**) canonical A/AT and G/GC triplets and (**B**) TA and CG base inversion sites. Arrow shows the direction of the phosphate backbone.

sites with sufficient affinity, remains much more challenging.

Recently, we synthesized 2-amino-3-methylpyridinyl pseudo-dC (^{3Me}AP-ΨdC) and demonstrated that an antigene TFOs incorporating ^{3Me}AP-ΨdC possessed significant affinity and selectivity toward the CG inversion site and effectively inhibited intracellular transcription of the human telomerase reverse transcriptase (hTERT) gene (25). It was anticipated that the 1-*N* position of 2-aminopyridine within ^{3Me}AP-ΨdC was protonated to form three hydrogen bonds with the CG base pair, as schematically shown in Figure 2A. This hypothesis was affirmed by a AP-ΨdC derivative in which the 3-methyl group of the aminopyridine unit was replaced by a halogen atom (Cl, Br or I) that caused a reduction of the pK_a value of the 1-*N* position (pK_a value is ∼4.5) and also decreased the triplex stability and selectivity to the CG inversion site (26). These results also supported the hypothesis of the contribution of the protonated 1-*N* position to triplex formation with the AP-ΨdC derivatives. Therefore, this study was initiated to improve the hydrogen binding affinity, by increasing the pK_a of the 1-*N* position of aminopyridine within AP-ΨdC derivatives. According to the molecular design concept, we devised two types of new AP-ΨdC derivatives to confirm the hydrogen bonding effect of the 1-*N* position and substituent effect of the aminopyridine unit. One series includes bicyclic compounds; i.e. 2,3-dihydro-7-azaindole (DHAID-ΨdC), 1,2,3,4-tetrahydro-1,8-naphthyridine (THNTD-ΨdC) and 7-azaindole (AID-ΨdC) (Figure 2B, C and D, respectively). In another series, the methyl or methoxy group was introduced at the 4-position of the 2-aminopyridine unit, 2-amino-4-methylpyridinyl pseudo-dC (^{4Me}AP-ΨdC) (Figure 2E) and 2-amino-4-methoxypyridinyl pseudo-dC (^{4OMe}AP-ΨdC) (Figure 2F), respectively. We now describe the design, synthesis and evaluation of the triplex-forming ability of the TFOs incorporating the newly designed artificial AP-ΨdC derivatives.

MATERIALS AND METHODS

All starting materials, reagents and solvents are purchased from Sigma-Aldrich, Inc. and Tokyo Chemical Industry Co. Ltd., and analytical reagents used without further purification. The ^{1}H-NMR (400 MHz, 500 MHz) and ^{13}C-NMR (125 MHz) spectra were recorded by Varian UNITY-400 and Bruker Ascend-500 spectrometers. The ^{31}P-NMR (202 MHz) spectrum was recorded using 10% phosphoric acid in D_2O as the internal standard at 0 ppm. Infrared spectra were obtained using a PerkinElmer FTIR-SpectrumOne. The high-resolution mass spectra were recorded by an Applied Biosystems Mariner System 5299 spectrometer using bradykinin, angiotensin and neurotensin as the internal standard. The Matrix Assisted Laser Desorption Ionization-Time of Flight-MS (MALDI-TOF-MS) spectra were recorded by a BRUKER DALTONICS Microflex. The ultraviolet (UV)–vis spectra were measured by a BECKMAN COULTER DU-800.

The pK_a determination by ^{1}H-NMR

All the diol compounds were dissolved or suspended in 0.1 M Na_2HPO_4 in D_2O and titrated with 0.5 M citric acid in D_2O. The ^{1}H-NMR spectra were collected at each point of the titration. The chemical shift of the proton on the pyridine ring was plotted versus the p*D* value. The half-equivalence point was determined, and the pK_a in H_2O was re-calculated according to the equation.

Preparation of the TFOs incorporating ΨdC derivatives

The 18-mer TFOs (3′-GGAAGGNZN'GAGGAGGGA-5′, Z = DHAID-ΨdC, THNTD-ΨdC, AID-ΨdC, ^{4Me}AP-ΨdC and ^{4OMe}AP-ΨdC), the hTERT TFOs (5′-GGAAAGGZGZZGGGGZGGGAGAGGAG-$_{C3}$-NH_2, Z = T, ^{3Me}AP-ΨdC, THNTD-ΨdC and ^{4OMe}AP-ΨdC) and Cyclin D1 TFOs (5′-GZGGGZGGGGGZGGGGGG-$_{C3}$-NH_2, Z = T, ^{3Me}AP-ΨdC and ^{4OMe}AP-ΨdC) were synthesized on a 1 μmol scale using a DNA automated synthesizer (Nihon Techno Service Co., Ltd.) by the standard phosphoramidite chemistry. Cleavage from the resin was accomplished by treatment overnight with 28% ammonium hydroxide at 55°C, followed by reverse-phase high pressure liquid chromatography (HPLC) purification (column: Nacalai tesque COSMOSIL 5C18-MS-II, 10 × 250 mm, solvents: A: 0.1 M Triethylamine Acetate (TEAA) buffer (pH = 7.0), B: CH_3CN, linear gradient: B for 10–40%/20 min, flow rate: 3.0 ml/min, UV: 254 nm, column oven: 35°C). The DMTr group was removed in 5% aqueous acetic acid in an aqueous solution at room temperature for 30 min, followed by reverse-phase HPLC

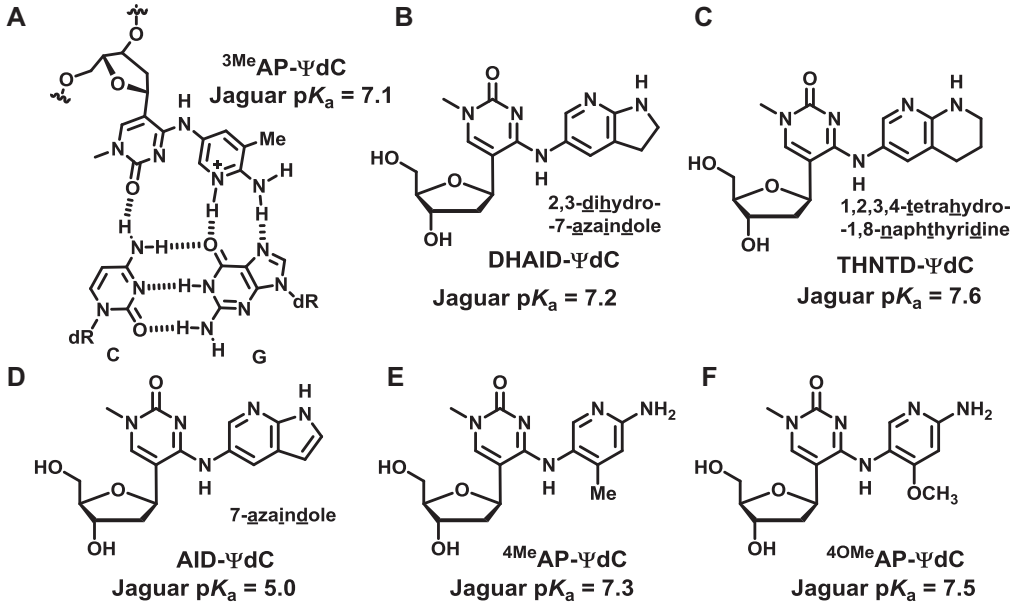

Figure 2. (**A**) The basic molecular design of ^{3Me}AP-ΨdC, and (**B**–**F**) structures of the chemically modified ΨdC derivatives in this study. Jaguar pK_a showed the pK_a values of the *N* position on the pyridine ring.

purification (column: SHISEIDO MG-I C18, 4.6 × 250 mm, solvents: A: 0.1 M TEAA buffer (pH 7.0), B: CH_3CN, B: 5% to 15%/20 min, flow rate: 1.0 ml/min, UV: 254 nm, column oven: 50°C). The structural integrity of the TFOs was confirmed by MALDI-TOF-MS measurements.

General procedure for the evaluation of the triplex-forming ability of TFOs

The FAM-labeled duplex DNA (100 nM) was incubated with increasing concentrations of the TFO (0–1000 nM) in buffer containing 20 mM Tris–HCl and 20 mM $MgCl_2$ for 12 h at 37°C and pH 7.5. Electrophoresis was performed at 4°C using 10% non-denatured polyacrylamide gel. The gel was visualized using a Luminoimage analyzer LAS-4000 (FUJIFILM), and the fluorescence intensity of each band was quantified for the calculation of the association constants. K_s (10^6 M^{-1}) = [Triplex]/([TFO][Duplex]). The standard deviations (± S.D.) for each K_s value were calculated from three independent experiments.

Evaluation of DNase I footprinting digestion

5′-FAM-labeled duplex DNA (100 nM) was incubated with TFO (0–250 nM) in the buffer containing 20 mM Tris–HCl, 20 mM $MgCl_2$ at 37°C and pH 7.5. The mixture was subsequently treated with DNase I (0.05 U) for 1 min at 37°C. The reaction was terminated by adding formamide containing 20 mM ethylenediaminetetraacetic acid (EDTA) and heating at 90°C. The products were separated by 20% denatured polyacrylamide gel containing 8 M urea and visualized by LAS-4000.

Computational calculation of the base triplet structure

Geometry optimization for the base triplet structures was performed using the B3LYP density functional with 6–31G* basis for the complex between artificial nucleoside and the CG base pair *in vacuo* as implemented in the Gaussian09 program. These optimized complex structures were inserted into the triplex DNA with the TFO sequence of the 3′GZA5′ context. The neighboring base triplet structures of the XAP-ΨdC/CG triplet (X = 3Me, 4Me and 4OMe) were then optimized under the amber force field in HyperChem.

RESULTS AND DISCUSSION

Synthesis of modified aminopyridinyl-pseudo deoxycytidine derivatives

The syntheses of the corresponding aromatic amine monomers are shown in Supplementary Scheme S1 in the Supporting Data. The synthesis of the DHAID recognition unit **16a** was started using compound **1**. The reaction of **1** with 69% nitric acid and concentrated sulfuric acid at low temperature produced the 1-nitro compound **2** in 95% yield. The 1-nitro compound **2** in concentrated sulfuric acid was converted to the 5-nitro isomer **3** in 35% yield (27). The selective reduction reaction of the nitro group of **3** was then carried out in the presence of 10% Pd on carbon under a hydrogen atmosphere in order to obtain the intermediate **16a**. The synthesis of the THNTD recognition unit **16b** was started using compound **4**. 1,8-Naphthyridine **4** was selectively reduced using 10% Pd on carbon in an ethanol solution to obtain compound **5** in 88% yield. The 1,6-dinitro compound **6** was obtained by the treatment of compound **5** with 69% nitric acid and concentrated sulfuric acid at low temperature in 72% yield. The 1-nitro group selective cleavage occurred using 10% Pd on carbon under a hydrogen atmosphere in the presence of acetic acid for 4 days in 65% yield. Finally, the nitro group was converted to the corresponding amino group to obtain compound **16b** using the same procedure as already described. The syn-

thesis of the 7-azaindole recognition unit **16c** was started from compound **8** (28,29). The 2-amino-5-nitropyridine **8** was iodinated at the 3-position using the Batkowski procedure to obtain 2-amino-3-iodo-5-nitropyridine **9** in 91% yield. After the iodination reaction, compound **10** was synthesized under Sonogashira coupling conditions in 94% yield. The cycloisomerization reaction was carried out to obtain 5-nitro-7-azaindole in 57% yield. Finally, the reduction product of **16c** was obtained using the general procedure as already described. The 2-amino-4-methylpyridine recognition unit (^{4Me}AP) **16d** was obtained from compound **12** in the presence of 10% Pd on carbon under a hydrogen atmosphere in a good yield. The synthesis of the 2-amino-4-methoxylpyridine recognition unit (^{4OMe}AP) **16e** was started from compound **13**. Compound **13** was treated with fuming nitric acid and concentrated sulfuric acid at low temperature to provide the 2-nitraminopyridine intermediate, followed by rearrangement to introduce a nitro group at the 5-position in 25% yield over two steps (30). Compound **15** was then obtained by condensation with sodium methoxide in methanol in 87% yield. Finally, the reduction product **16e** was obtained in the presence of 10% Pd on carbon under a hydrogen atmosphere in a good yield.

The syntheses of the corresponding diol and phosphoramidite monomers of the modified AP-ΨdC derivatives are shown in Supplementary Scheme S2. The substitution reaction that began with compound **17** was carried out with a series of newly designed aromatic amines **16a–e** in the present of DIPEA in anhydrous acetonitrile to produce compound **18a–e**. Afterward, these compounds (**18a–e**) were converted to the corresponding diol compounds (**19a–e**) by treatment with 0.5 M HCl in a methanol solution in good yields. All the diol compounds were used to determine the pK_a values of the pyridine unit by ^{1}H-NMR measurement in a D_2O solution. On the other hand, compounds **18a–e** were treated with phenoxyacetic anhydride to protect the amine groups, followed by desilylation of the hydroxyl group with $3HF/Et_3N$ at the 3′-terminal of the sugar part in two steps to produce compounds **20a–e**. Finally, after conversion to the corresponding phosphoramidite compounds **21a–e**, they were incorporated into the middle of the 18-mer oligodeoxynucleotides (ODNs) using an automated DNA synthesizer. The synthesized ODNs were cleaved from the CPG and the protecting groups of the nucleobases were removed by heating in a 28% ammonium hydroxide solution at 55°C overnight, then purified by reverse-phase HPLC. The DMTr group at the 5′ end of ODN was deprotected by treatment with 5% acetic acid in an aqueous solution at room temperature for 30 min, then the final target TFOs sequences were again purified by HPLC. The structural integrity and purity of the TFOs were confirmed by MALDI-TOF-MS measurements (Supplementary Table S1 and Supplementary Figure S1).

Determination of pK_a values of the nitrogen atom on the pyridine recognition units

In order to determine the pK_a values of the nitrogen atom on the pyridine ring, the ^{1}H-NMR spectra of all the designed diol compounds (**19a–e**) in D_2O were measured at different pD conditions. The chemical shifts of the hydrogen atom on the aminopyridine ring were then simultaneously plotted (Supplementary Figure S2). The half-equivalence points were determined from these plots, and the pK_a values in H_2O were re-calculated according to the following equation, pH = pD + 0.40 (Table 1) (31). To compare these results, we simulated these values by the Jaguar pK_a prediction method using compounds in which the ΨdC part was replaced with a methyl group (32). Based on the results of these values, there existed a nice correlation between the measured pK_a values and calculated pK_a values of all the modified diol compounds. These values demonstrated that the nitrogen atom on the pyridine ring of the newly designed compounds except for AID-ΨdC (pK_a = 4.1, Table 1) should be protonated under neutral conditions.

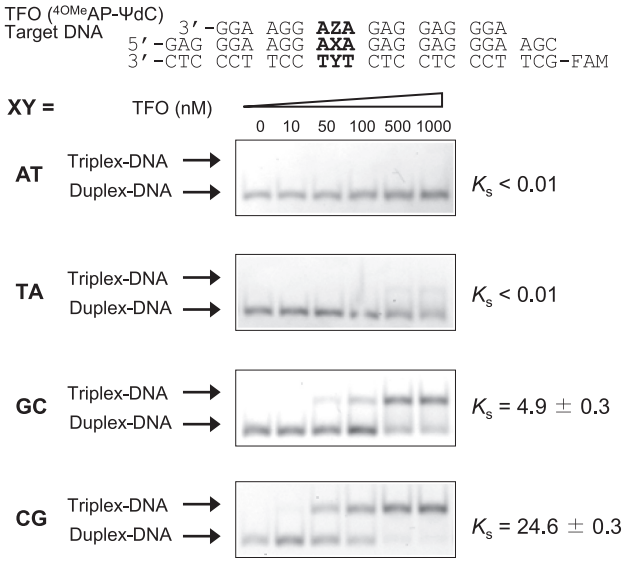

Figure 3. The gel result of triplex-forming ability of TFO containing ^{4OMe}AP-ΨdC in sequences in the 3′AZA5′ context. Conditions: The FAM-labeled duplex DNA (24 bp; 100 nM) was incubated with increasing concentrations of TFO (18 mer; 0–1000 nM) in a buffer containing 20 mM Tris–HCl and 20 mM $MgCl_2$ at pH 7.5 and 37°C. Electrophoresis was performed with 10% non-denatured polyacrylamide gel at 4°C. K_s (10^6 M^{-1}) = [Triplex]/([TFO][Duplex]).

The evaluation of the triplex-forming ability of the TFOs containing the modified AP-ΨdC derivatives against a single CG base pair

The triplex-forming ability of the synthesized TFOs was evaluated by gel-electrophoretic mobility shift assays using FAM-labeled duplex DNAs of an antiparallel DNA type. An example of the gel result of the 3′AZA5′ (Z = ^{4OMe}AP-ΨdC) sequence is shown in Figure 3. The duplex DNAs and triplex DNAs were detected by a fluorescence image analyzer, then the equilibrium association constants (K_s) were calculated by quantification of the fluorescent intensities of these bands. According to this gel result and K_s values for all four target base pairs, the TFO containing ^{4OMe}AP-ΨdC apparently underwent triplex formation with the target duplex DNA containing a CG inversion site with high selectivity.

Table 1. pK_a values by ^{1}H-NMR measurement and prediction[a]

ΨdC derivatives	^{3Me}AP	DHAID	THNTD	AID	^{4Me}AP	^{4OMe}AP
pK_a value by NMR	6.3[b]	6.8	7.4	4.1	6.8	7.3
Jaguar pK_a prediction	7.1	7.2	7.6	5.0	7.3	7.5

[a]The pK_a values were determined from the titration plots in Supplementary Figure S2. The pK_a of the 1-*N* of aminopyridine was calculated using a compound in which the ΨdC part was replaced by a methyl group using Jaguar (ver. 8.1) (32).
[b]Data from (25).

To investigate the effect of adjacent base pairs to the modified AP-ΨdC derivatives for triplex-forming ability, four groups of TFOs bearing different combinations of flanking bases for each derivative (3′AZA5′, 3′AZG5′, 3′GZA5′ and 3′GZG5′; Z = new modified AP-ΨdC derivatives) and the corresponding target duplex DNAs were prepared. The K_s values are summarized in Table 2. In the case of TFOs incorporating the DHAID-ΨdC, THNTD-ΨdC and AID-ΨdC derivatives, which were designed as bicyclic structures, THNTD-ΨdC showed the highest affinity for the CG base pair in all four 3′NZN'5′ combinations. In particular, THNTD-ΨdC exhibited a relatively high affinity for the CG base pair in the 3′AZA5′ and 3′GZG5′ sequences compared to the original ^{3Me}AP-ΨdC, whereas DHAID-ΨdC showed only a weak interaction with the CG base pair. In the 3′AZG5′ or 3′GZA5′ sequence, in which ^{3Me}AP-ΨdC selectively stabilized triplexes with the CG base pair, both THNTD-ΨdC and DHAID-ΨdC could not recognize the CG base pair with a sufficient stability compared to the ^{3Me}AP-ΨdC derivative.

Although the introduction of a bicyclic ring structure (THNTD-ΨdC and DHAID-ΨdC) increased the pK_a values and protonation ability of 1-*N* of the pyridine ring, a negative impact also occurred. For instance, the presence of steric hindrance, reduction of hydrophilicity triggered by the restricted structure and conformational change in the tautomerization between two close *N*-atoms in the recognition units could make a great contribution to the triplex destabilizing effect. Under these circumstances, the TFOs incorporating THNTD-ΨdC derivative showed preferable triplex-forming ability with CG base pairs compared with that of DHAID-ΨdC derivative. The relatively higher pK_a value and suitable position of 1-*N* to form a hydrogen bond, both in orientation and close proximity with a guanine nucleobase of a CG base pair for THNTD-ΨdC derivative, were thought to be responsible for this occurrence. In contrast, the TFOs incorporating AID-ΨdC showed a weak affinity toward the CG base pair in all sequences, except for the 3′AZG5′ sequence. Presumably, the protonation of 1-*N* of the pyridine ring did not occur due to its lower pK_a value, and the corresponding hydrogen bond to a guanine nucleobase of a CG base pair was not formed.

The ^{4Me}AP-ΨdC and ^{4OMe}AP-ΨdC derivatives showed higher pK_a values compared to the parent ^{3Me}AP-ΨdC (Table 1). The ^{4Me}AP-ΨdC-TFO formed a stable triplex only when it was incorporated into the 3′GZA5′ sequence (Table 2). In remarkable contrast, TFOs containing the ^{4OMe}AP-ΨdC derivative demonstrated triplex-forming ability and selectivity with the CG base pair in all sequences (Figure 3 and Table 2). Although a weak interaction, unlike the ^{4Me}AP-ΨdC-TFO, was observed for the GC base pair, this result is presumably due to a change in the protonation state of the pyridine ring that allows its interaction with the carbonyl group at the 6 position of dG (Supplementary Figure S4). It should be noted that the ^{4OMe}AP-ΨdC-TFO has overcome the sequence dependence problem with a higher affinity compared to the parent ^{3Me}AP-ΨdC-TFO. Consequently, these results supported our hypothesis that the protonated form of the pyridine ring at the 1-*N* position and the exocyclic primary amino group of the aminopyridinyl-ΨdC nucleoside analogues contributed to recognition of the guanine base of a CG base pair in duplex DNA.

Evaluation of DNase I footprinting digestion

To corroborate the selective and stable recognition ability of a TFO containing the ^{4OMe}AP-ΨdC derivative for a CG base pair, a DNase I footprinting assay was performed using two long duplex DNAs (5′-FAM-labeled) including a CG base pair as the target site and an AT base pair as the negative control (Figure 4).

It was obviously found that the bands corresponding to the CG target region disappeared in the presence of 50–250 nM TFO containing ^{4OMe}AP-ΨdC, indicating that this TFO formed triplexes with the CG inversion site and protected them from the DNase I enzymatic digestion. On the other hand, the bands corresponding to the AT target region were present and mostly the same as the result in the absence of TFO containing ^{4OMe}AP-ΨdC, indicating that this TFO could not form stable triplexes with the AT base pair site for concentrations of 50–250 nM and DNase I enzymatic digestion occurred. These results indicated that a TFO containing ^{4OMe}AP-ΨdC selectively recognized the CG target site and inhibited the interaction of the enzyme protein with the target duplex DNA.

Verification of distinguishability of 5-methylcytosine and cytosine

Although genomic information is generally described by four types of nucleosides, A, T, G and C, it is known that a wide variety of modified nucleosides exist in intracellular genomic DNA (33). Among them, 5-methylcytosine (^{m}C) is one of the most important nucleosides (34,35). Therefore, we applied the newly developed AP-ΨdC derivatives for the recognition of a ^{m}CG base pair. The gel results for the ability to form triplex DNA between a TFO incorporating ^{4OMe}AP-ΨdC and target DNA including the ^{m}CG base pair with different adjacent bases are shown in Supplementary Figure S5. Unfortunately, when the target base pair was ^{m}CG, hardly any triplex DNA was formed. It was thought that insufficient triplex DNA formation might be due to the

Table 2. The equilibrium association constants (K_s values) of each TFO in four different sequence contexts[a]

TFO 3′-GGAAGG NZN' GAGGAGGGA Duplex DNA 5′-GAGGGAAGG NXN' GAGGAGGGAAGC 3′-CTCCCTTCC MYM' CTCCTCCCTTCG-FAM					
		K_s (10^6 M^{-1}) for XY			
3′ NZN 5′	Z (-ΨdC)	AT	TA	GC	CG
3′ AZA 5′	^{3Me}AP[b]	<0.01	<0.01	0.2 ± 0.1	20.8 ± 0.9
	DHAID	<0.01	<0.01	0.1 ± 0.1	2.1 ± 0.3
	THNTD	<0.01	<0.01	0.2 ± 0.1	22.8 ± 0.8
	AID	<0.01	<0.01	<0.01	<0.01
	^{4Me}AP	< 0.01	<0.01	<0.01	6.1 ± 0.3
	^{4OMe}AP	<0.01	<0.01	4.9 ± 0.3	24.6 ± 0.3
3′ AZG 5′	^{3Me}AP[b]	<0.01	0.2 ± 0.1	1.8 ± 0.6	19.4 ± 1.8
	DHAID	<0.01	0.2 ± 0.1	2.1 ± 0.4	2.8 ± 0.4
	THNTD	<0.01	<0.01	1.0 ± 0.1	5.9 ± 0.1
	AID	0.3 ± 0.1	3.6 ± 1.1	1.1 ± 0.7	17.0 ± 2.5
	^{4Me}AP	<0.01	0.7 ± 0.1	2.1 ± 0.4	3.0 ± 0.1
	^{4OMe}AP	<0.01	<0.01	7.9 ± 1.2	22.5 ± 1.7
3′ GZA 5′	^{3Me}AP[b]	<0.01	<0.01	1.8 ± 0.5	32.6 ± 0.5
	DHAID	<0.01	<0.01	0.3 ± 0.1	13.4 ± 0.9
	THNTD	<0.01	<0.01	0.1 ± 0.1	13.7 ± 0.6
	AID	<0.01	0.3 ± 0.1	3.0 ± 0.3	<0.01
	^{4Me}AP	<0.01	1.5 ± 0.1	3.0 ±0.6	28.5 ± 0.3
	^{4OMe}AP	<0.01	1.0 ± 0.2	7.8 ± 0.4	34.7 ± 1.1
3′ GZG 5′	^{3Me}AP[b]	0.8 ± 0.1	2.6 ± 0.5	5.3 ± 0.4	16.6 ± 0.5
	DHAID	0.7 ± 0.1	3.2 ± 0.2	4.9 ± 0.1	10.6 ± 1.0
	THNTD	0.9 ± 0.2	3.5 ± 0.4	6.6 ± 0.3	19.1 ± 0.6
	AID	1.1 ± 0.6	5.7 ± 0.1	3.6 ± 0.6	8.5 ± 0.5
	^{4Me}AP	1.2 ± 0.2	6.1 ± 0.6	6.1 ± 0.7	8.7 ± 1.0
	^{4OMe}AP	1.0 ± 0.3	4.5 ± 1.5	6.9 ± 0.9	18.9 ± 0.5

[a]Conditions: FAM-labeled duplex DNA (24 bp; 100 nM) was incubated with increasing concentrations of TFO (18 mer; 0–1000 nM) in the buffer containing 20 mM Tris–HCl and 20 mM $MgCl_2$ at pH 7.5 and 37°C. Electrophoresis was performed with 10% non-denatured polyacrylamide gel at 4°C. K_s (10^6 M^{-1}) = [Triplex]/([TFO][Duplex]).
[b]Data from (25).

steric hindrance of the methyl group of 5-methylcytosine in the ^{m}CG base pair.

Verification of recognition ability for double strand

We next tested selectivity for double-stranded DNA (ds-DNA), RNA (dsRNA) and DNA/RNA heteroduplexes. The gel results are shown in Supplementary Figure S6. According to these results, the TFOs containing ^{4Me}AP-ΨdC and ^{4OMe}AP-ΨdC derivatives can selectively recognize CG inversion site only within DNA duplexes. It is likely that the major groove structure of the duplex affects the stability of the triplex formation, i.e. B-form dsDNA and A-form dsRNA have differences in helical pitch and stacking structure, which are probably affecting triplex formation (Supplementary Figure S7) (36).

The evaluation of triplex-forming ability of TFOs containing modified AP-ΨdC derivatives against multiple CG base pairs

Taking into consideration that the CG inversion sites in the promoter regions of the human gene are multiple and consecutive, the recognition ability of TFOs containing AP-ΨdC derivatives for multiple CG base pairs was further demonstrated by triplex formation against human gene sequences. It is reported that overexpression of gene transcription of the hTERT (or HTR) was selective in tumor cells, but not in normal cells, and played a critical role in human tumorigenesis (37–39). There is an available triplex-forming region in the promoter sequence containing four CG inversion sites with two of them consecutive (Figure 5A). On the other hand, Cyclin D1 (Cy) is an important regulator of the cell cycle progression and can function as a transcriptional co-regulator (40). At the same time, overexpression of cyclin D1 has been linked to the development and progression of cancer (41,42). There also exists an available triplex-forming region in the promoter sequence containing three CG inversion sites (Figure 5B).

The promoter regions of the hTERT and Cyclin D1 genes were confirmed as target genes to investigate the recognition ability of AP-ΨdC derivatives by triplex formation. The antigene TFOs incorporating THNTD-ΨdC and ^{4OMe}AP-ΨdC, which showed good recognition affinity for CG base pair in the adjacent 3′GZG5′ sequence, were designed to bind to the target CG inversion sites in an antiparallel type triplex formation. The corresponding antigene TFOs were synthesized and purified by reverse-phase HPLC. The structural integrity and purity of the TFOs were confirmed by MALDI-TOF-MS measurements (Supplementary Table S2 and Supplementary Figure S8).

The triplex-forming ability was evaluated by a gel shift assay, and the gel results and K_s values of the targeting for the promoter region of the hTERT and Cyclin D1 gene are depicted in Figure 5C and D, respectively. In the case of tar-

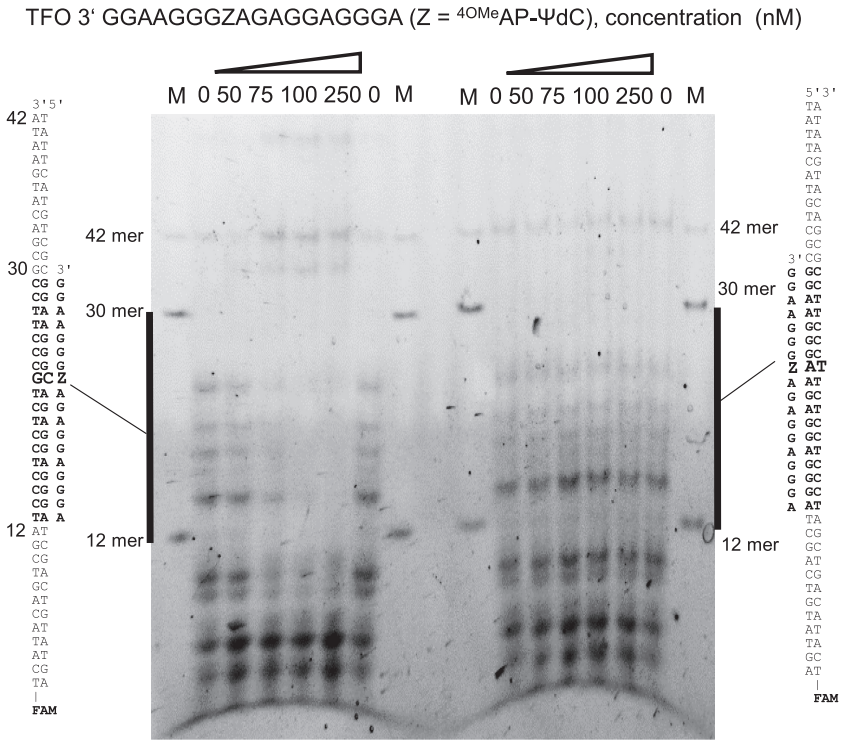

Figure 4. DNase I footprinting assay for confirmation of triplex formation. 5′-FAM-labeled duplex DNA (100 nM) was incubated with TFO (0–250 nM) in buffer containing 20 mM Tris–HCl, 20 mM $MgCl_2$ at 37°C and pH 7.5. The mixture was subsequently treated with DNase I (0.05 U) for 1 min at 37°C. The reaction was terminated by adding formamide containing 20 mM EDTA and heating at 90°C. M is the mixture of 5′-FAM-labeled 12-, 30- and 42-mer ODNs as makers. The products were separated by 20% denatured polyacrylamide gel containing 8 M urea and visualized by LAS-4000.

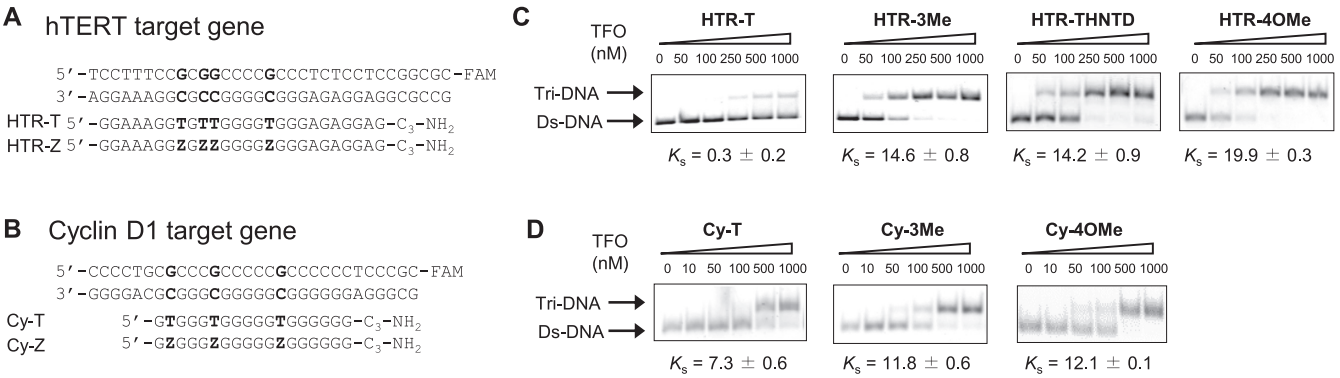

Figure 5. Evaluation of triplex-forming ability of TFOs containing modified AP-ΨdC derivatives against multiple CG base pairs. (**A**) The hTERT promoter sequence (hTR) containing four CG inversion sites and sequences of the corresponding TFOs (HTR-T and HTR-*Z*). (**B**) The Cyclin D1 promoter sequence (Cy) containing three CG inversion sites and sequences of the corresponding TFOs (Cy-T and Cy-*Z*). (**C**) FAM-labeled hTR (32 bp; 100 nM) was incubated with increasing concentrations of each TFO (26 mer; 0–1000 nM) in the buffer containing 20 mM Tris–HCl, 2.5 mM $MgCl_2$ and 2.5 mM spermidine at pH 7.5 and 37°C. Electrophoresis was performed with a 10% non-denaturing polyacrylamide gel at 4°C. K_s ($10^6 M^{-1}$) = [Triplex]/([TFO][Duplex]). (**D**) FAM-labeled Cy (30 bp; 100 nM) was incubated with increasing concentrations of each TFO (18 mer; 0–1000 nM) in the buffer containing 89 mM Tris–HCl, 10 mM $MgCl_2$ at pH 7.4 and 37°C. Electrophoresis was performed with a 10% non-denaturing polyacrylamide gel at 4°C. K_s ($10^6 M^{-1}$) = [Triplex]/([TFO][Duplex]).

geting the hTERT sequence, the TFO containing THNTD-ΨdC showed a moderate triplex-forming ability compared to ^{3Me}AP-ΨdC, due to the structural disturbance resulting from the sequence dependency. In contrast, the TFO containing ^{4OMe}AP-ΨdC could form a stable triplex DNA with high affinity even at a low concentration compared to the original ^{3Me}AP-ΨdC. On the other hand, for the target Cyclin D1 sequence with CG inversion sites, the TFO containing ^{4OMe}AP-ΨdC derivative showed a moderate triplex-forming ability compared to ^{3Me}AP-ΨdC. The affinity of T/CG base triplet using a natural TFO is highly influenced by the sequence contexts (25). For the Cyclin D1 sequence, the natural TFO(Cy-T) showed moderate affinity to the target sequences. Therefore, the difference between the K_s values of natural TFO(Cy-T) and TFO(Cy-Z) was relatively small compared to the results of the targeting for

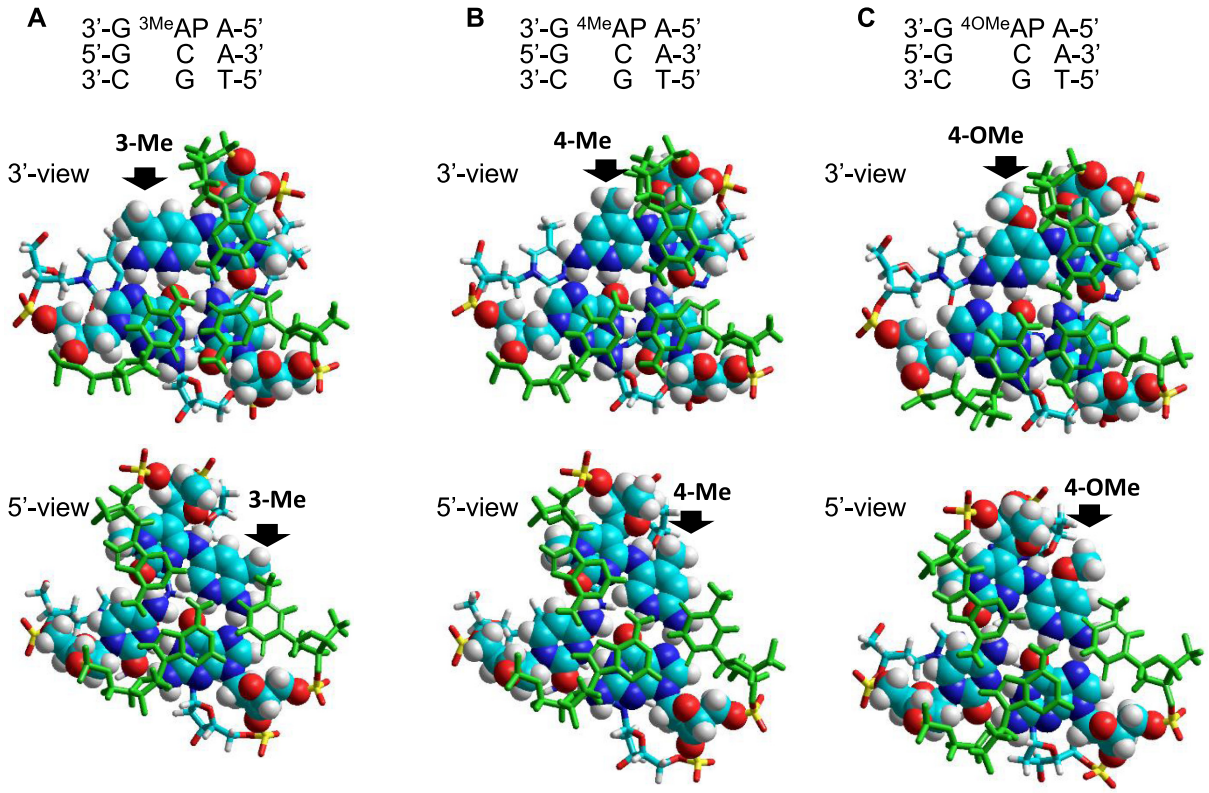

Figure 6. Optimize structure, CPK model, of AP-ΨdC derivatives-CG complex using Gaussian09 B3LYP/6–31G* *in vacuo*. The neighboring base triplets were optimized using HyperChem under the Amber force field. Parts of the triplex structure were depicted in 3′- and 5′-view of the TFOs. The front base triplet is shown in green color. (**A**) ^{3Me}AP-ΨdC/CG, (**B**) ^{4Me}AP-ΨdC/CG and (**C**) ^{4OMe}AP-ΨdC/CG complexes in the sequence of 3′G and 5′ A in the TFO.

the hTERT sequence. These results indicated that ^{4OMe}AP-ΨdC derivative could be used for the formation of a stable triplex as well as ^{3Me}AP-ΨdC even in the presence of multiple and consecutive CG base pairs.

The computational molecular modeling of the complex between the AP-ΨdC derivatives and the CG base pair

To further understand the triplex-forming ability, for a CG base pair, of the AP-ΨdC derivatives, the base triplet structures were optimized by Density Functional Theory (DFT) calculation using the protonated AP-ΨdC derivatives and the CG base pair. These optimized complex structures were inserted into triplex DNA with the TFO sequence of a 3′GZA5′ context, which formed the stable and selective triplex DNA (Table 2). The neighboring base triplet structures of the XAP-ΨdC/CG triplet (X = 3Me, 4Me and 4OMe) were then optimized under the amber force field. The regions of interest are shown in Figure 6, which illustrate the 3′-side and 5′-side views of triplex DNA with a 3′GZA5′ sequence. According to the predicted triplex DNA structures, the 3- and 4-substituents of the aminopyridine ring may not suffer from steric repulsions with bases or sugar backbones and may be allowed in the triplex DNA (Figure 6A, B and C). Also, it was found that the guanine base on the 3′ side of the TFO is nicely stacked with the peudocytosine unit as shown in the 3′-view pictures. Furthermore, the adenine base on the 5′ side of the TFO is also stacked with the pseudocytidine unit as shown in the 5′-view pictures. It is reasonable to interpret that these stacking interactions together with the H-bond formation with the protonated 1-*N* position of the pyridine ring stabilize triplex formation. These results suggest that the recognition ability of the artificial nucleoside for CG inversion sites may be adjusted by small chemical modification in the proper position of the aminopyridine part of the pseudocytidine nucleoside analogue (Table 2 and Figure 6)

CONCLUSIONS

In summary, with the aim of tuning and improving the protonation ability of the pyridine ring unit, we designed, synthesized and evaluated triplex-forming ability of bicyclic derivatives (THNTD-ΨdC, DHAID-ΨdC and AID-ΨdC) and 4-substituted derivatives (^{4Me}AP-ΨdC and ^{4OMe}AP-ΨdC) toward the CG base pair. Based on the pK_a studies of their monomers, it was suggested that the 1-*N* of aminopyridine in the AP-ΨdC derivatives are protonated under neutral conditions except for AID-ΨdC. For the bicyclic derivatives (THNTD-ΨdC, DHAID-ΨdC and AID-ΨdC), their stability is lower than the original ^{3Me}AP-ΨdC with a sequence dependency. On the other hand, in the case of the 4-substituted derivatives, ^{4OMe}AP-ΨdC showed a remarkable affinity and selectivity toward the CG base pair regardless of adjacent base pairs, which was also confirmed by DNase I footprinting assays. It was found that ^{4OMe}AP-

ΨdC distinguishes between cytosine and 5-methylcytosine and recognizes the CG base pair in dsDNA compared to dsRNA and DNA/RNA heteroduplexes. Remarkably, ^{4OMe}AP-ΨdC formed a stable triplex DNA with the sequence from the promoter region of the hTERT and the Cyclin D1 gene including multiple CG inversion sites. There is another problem when developing antigene method using an antiparallel type triplex DNA formation *in vivo*. The purine-rich TFOs have potential to form G-quadruplex structures in the presence of potassium ion and compete with triplex DNA formations. However, the existence of artificial nucleic acid in TFO may prevent the formation of such G-quadruplex structures, and a detailed study will be conducted in the future (43,44). We are now going to investigate the antigene effect on cultured cells using a large number of antigene TFOs containing ^{4OMe}AP-ΨdC and their derivatives for targeting sequences including multiple CG base pairs.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Grant-in-Aid for Scientific Research (B) [16H05100 to Y.T., 15H04633 to S.S.]; Japan Society for the Promotion of Science (JSPS) for Challenging Exploratory Research [26670056 to Y.T.]; China Scholarship Council [201506220182 to L.W.]. Funding for open access charge: Grant-in-Aid for Scientific Research (B) [16H05100 to Y.T.].

Conflict of interest statement. None declared.

REFERENCES

1. Mukherjee,A. and Vasquez,K.M. (2011) Triplex technology in studies of DNA damage, DNA repair, and mutagenesis. *Biochimie*, **93**, 1197–1208.
2. Bacolla,A., Wang,G. and Vasquez,K.M. (2015) New perspectives on DNA and RNA triplexes as effectors of biological activity. *PLoS Genet.*, **11**, e1005696.
3. Maldonado,R., Filarsky,M., Grummt,I. and Längst,G. (2018) Purine–and pyrimidine–triple-helix-forming oligonucleotides recognize qualitatively different target sites at the ribosomal DNA locus. *RNA*, **24**, 371–380.
4. Liu,X., Song,M. and Li,F. (2017) Triplex DNA-based bioanalytical platform for highly sensitive homogeneous electrochemical detection of melamine. *Sci. Rep.*, **7**, 4490.
5. Graham,M.K., Brown,T.R. and Miller,P.S. (2015) Targeting the human androgen receptor gene with platinated triplex-forming oligonucleotides. *Biochemistry*, **54**, 2270–2282.
6. Ye,Z., Guntaka,R.V. and Mahato,R.I. (2007) Sequence-specific triple helix formation with genomic DNA. *Biochemistry*, **46**, 11240–11252.
7. Govan,J.M., Uprety,R., Hemphill,J., Lively,M.O. and Deiters,A. (2012) Regulation of transcription through light-activation and light-deactivation of triplex-forming oligonucleotides in mammalian cells. *ACS Chem. Biol.*, **7**, 1247–1256.
8. Dahmen,V., Schmitz,S. and Kriehuber,R. (2017) Induction of the chromosomal translocation t(14;18) by targeting the BCL-2 locus with specific binding I-125-labeled triplex-forming oligonucleotides. *Mutat. Res.*, **823**, 58–64.
9. Cannata,F., Brunet,E., Perrouault,L., Roig,V., Ait-Si-Ali,S., Asseline,U., Concordet,J.-P. and Giovannangeli,C. (2008) Triplex-forming oligonucleotide–orthophenanthroline conjugates for efficient targeted genome modification. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 9576–9581.
10. Kaushik Tiwari,M., Adaku,N., Peart,N. and Rogers,F.A. (2016) Triplex structures induce DNA double strand breaks via replication fork collapse in NER deficient cells. *Nucleic Acids Res.*, **44**, 7742–7754.
11. Dahmen,V. and Kriehuber,R. (2012) Cytotoxic effects and specific gene expression alterations induced by I-125-labeled triplex-forming oligonucleotides. *Int. J. Radiat. Biol.*, **88**, 972–979.
12. Schleifman,E.B. and Glazer,P.M. (2014) *Peptide Nucleic Acids*. Springer, pp. 207–222.
13. Datta,H.J., Chan,P.P., Vasquez,K.M., Gupta,R.C. and Glazer,P.M. (2001) Triplex-induced recombination in human cell-free extracts. Dependence on XPA and HsRad51. *J. Biol. Chem.*, **276**, 18018–18023.
14. Vasquez,K.M., Narayanan,L. and Glazer,P.M. (2000) Specific mutations induced by triplex-forming oligonucleotides in mice. *Science*, **290**, 530–533.
15. Reza,F. and Glazer,P.M. (2015), *Chromosomal Mutagenesis*. Springer, pp. 39–73.
16. Semenyuk,A., Darian,E., Liu,J., Majumdar,A., Cuenoud,B., Miller,P.S., Mackerell,A.D. Jr and Seidman,M.M. (2010) Targeting of an interrupted polypurine:polypyrimidine sequence in mammalian cells by a triplex-forming oligonucleotide containing a novel base analogue. *Biochemistry*, **49**, 7867–7878.
17. Hari,Y., Akabane,M. and Obika,S. (2013) 2′,4′-BNA bearing a chiral guanidinopyrrolidine-containing nucleobase with potent ability to recognize the CG base pair in a parallel-motif DNA triplex. *Chem. Commun.*, **49**, 7421–7423.
18. Ohkubo,A., Yamada,K., Ito,Y., Yoshimura,K., Miyauchi,K., Kanamori,T., Masaki,Y., Seio,K., Yuasa,H. and Sekine,M. (2015) Synthesis and triplex-forming properties of oligonucleotides capable of recognizing corresponding DNA duplexes containing four base pairs. *Nucleic Acids Res.*, **43**, 5675–5686.
19. Kolganova,N.A., Shchyolkina,A.K., Chudinov,A.V., Zasedatelev,A.S., Florentiev,V.L. and Timofeev,E.N. (2012) Targeting duplex DNA with chimeric α,β-triplex-forming oligonucleotides. *Nucleic Acids Res.*, **40**, 8175–8185.
20. Okamura,H., Taniguchi,Y. and Sasaki,S. (2014) An isocytidine derivative with a 2-amino-6-methylpyridine unit for selective recognition of the CG interrupting site in an antiparallel triplex DNA. *ChemBioChem.*, **15**, 2374–2378.
21. Okamura,H., Taniguchi,Y. and Sasaki,S. (2013) N-(Guanidinoethyl)-2′-deoxy-5-methylisocytidine exhibits selective recognition of a CG interrupting site for the formation of anti-parallel triplexes. *Org. Biomol. Chem.*, **11**, 3918–3924.
22. Taniguchi,Y. and Sasaki,S. (2012) An efficient antigene activity and antiproliferative effect by targeting the Bcl-2 or survivin gene with triplex forming oligonucleotides containing a W-shaped nucleoside analogue (WNA-βT). *Org. Biomol. Chem.*, **10**, 8336–8341.
23. Taniguchi,Y., Nakamura,A., Senko,Y., Nagatsugi,F. and Sasaki,S. (2006) Effects of halogenated WNA derivatives on sequence dependency for expansion of recognition sequences in non-natural-type triplexes. *J. Org. Chem.*, **71**, 2115–2122.
24. Sasaki,S., Taniguchi,Y., Takahashi,R., Senko,Y., Kodama,K., Nagatsugi,F. and Maeda,M. (2004) Selective formation of stable triplexes including a TA or a CG interrupting site with new bicyclic nucleoside analogues (WNA). *J. Am. Chem. Soc.*, **126**, 516–528.
25. Okamura,H., Taniguchi,Y. and Sasaki,S. (2016) Aminopyridinyl–pseudodeoxycytidine derivatives selectively stabilize antiparallel triplex DNA with multiple CG inversion sites. *Angew. Chem. Int. Ed.*, **55**, 12445–12449.
26. Wang,L., Taniguchi,Y., Okamura,H. and Sasaki,S. (2017) Effect of the 3-halo substitution of the 2′-deoxy aminopyridinyl-pseudocytidine derivatives on the selectivity and stability of antiparallel triplex DNA with a CG inversion site. *Bioorg. Med. Chem.*, **25**, 3853–3860.
27. Robison,M.M., Robison,B.L. and Butler,F.P. (1959) 7-Azaindole. VI. Preparation of 5-and 6-substituted 7-azaindoles. *J. Am. Chem. Soc.*, **81**, 743–747.
28. Bhat,P.V., Dere,R.T., Ravikumar,S., Hindupur,R.M. and Pati,H.N. (2015) Efficient and Scalable Process for Synthesis of 5-Nitro-7-azaindole. *Org. Process Res Dev.*, **19**, 1282–1285.
29. Pearson,S.E. and Nandan,S. (2005) A practical, efficient synthesis of 5-amino-7-azaindole. *Synthesis*, **15**, 2503–2506.
30. Tzvetkov,N.T., Neumann,B., Stammler,H.-G. and Antonov,L. (2016) A simple approach to multifunctionalized N1-alkylated

7-amino-6-azaoxindole derivatives using their in situ stabilized tautomer form. *Tetrahedron*, **72**, 6455–6466.
31. Glasoe,P.K. and Long,F. (1960) Use of glass electrodes to measure acidities in deuterium oxide. *J. Phys. Chem.*, **64**, 188–190.
32. Bochevarov,A.D., Harder,E., Hughes,T.F., Greenwood,J.R., Braden,D.A., Philipp,D.M., Rinaldo,D., Halls,M.D., Zhang,J. and Friesner,R.A. (2013) Jaguar: a high-performance quantum chemistry software program with strengths in life and materials sciences, Int. *J. Quantum Chem.*, **113**, 2110–2142.
33. Carell,T., Brandmayr,C., Hienzsch,A., Müller,M., Pearson,D., Reiter,V., Thoma,I., Thumbs,P. and Wagner,M. (2012) Structure and function of noncanonical nucleobases. *Angew. Chem. Int. Ed.*, **51**, 7110–7131.
34. Song,J., Teplova,M., Ishibe-Murakami,S. and Patel,D.J. (2012) Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science*, **335**, 709–712.
35. Meissner,A., Mikkelsen,T.S., Gu,H., Wernig,M., Hanna,J., Sivachenko,A., Zhang,X., Bernstein,B.E., Nusbaum,C. and Jaffe,D.B. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, **454**, 766.
36. Semerad,C.J. and Maher,L.J. III (1994) Exclusion of RNA strands from a purine motif triple helix. *Nucleic Acids Res.*, **22**, 5321–5325.
37. Yan,T., Ooi,W.F., Qamra,A., Cheung,A., Ma,D., Sundaram,G.M., Xu,C., Xing,M., Poon,L. and Wang,J. (2018) HoxC5 and miR-615-3p target newly evolved genomic regions to repress hTERT and inhibit tumorigenesis. *Nat. Commun.*, **9**, 100.
38. Leão,R., Apolónio,J.D., Lee,D., Figueiredo,A., Tabori,U. and Castelo-Branco,P. (2018) Mechanisms of human telomerase reverse transcriptase (hTERT) regulation: clinical impacts in cancer. *J. Biomed. Sci.*, **25**, 22.
39. Zhou,J., Ding,D., Wang,M. and Cong,Y.-S. (2014) Telomerase reverse transcriptase in the regulation of gene expression. *BMB Rep.*, **47**, 8.
40. Caldon,C.E., Sutherland,R.L. and Musgrove,E.A. (2010) Cell cycle proteins in epithelial cell differentiation: implications for breast cancer. *Cell Cycle*, **9**, 1918–1928.
41. Kim,J.K. and Diehl,J.A. (2009) Nuclear cyclin D1: an oncogenic driver in human cancer. *J. Cell. Physiol.*, **220**, 292–296.
42. Tashiro,E., Tsuchiya,A. and Imoto,M. (2007) Functions of cyclin D1 as an oncogene and regulation of cyclin D1 expression. *Cancer Sci.*, **98**, 629–635.
43. Olivas,W.M. and Maher,L.J. III (1995) Overcoming potassium-mediated triplex inhibition. *Nucleic Acids Res.*, **23**, 1936–1941.
44. Olivas,W.M. and Maher,L.J. III (1995) Competitive triplex/quadruplex equilibria involving guanine-rich oligonucleotides. *Biochemistry*, **34**, 278–284.