

Recognition of 5-methyl-CG and CG base pairs in duplex DNA with high stability using antiparallel-type triplex-forming oligonucleotides with 2-guanidinoethyl-2'-deoxynebularine

Ryotaro Notomi¹, Shigeki Sasaki² and Yosuke Taniguchi^{1,*}

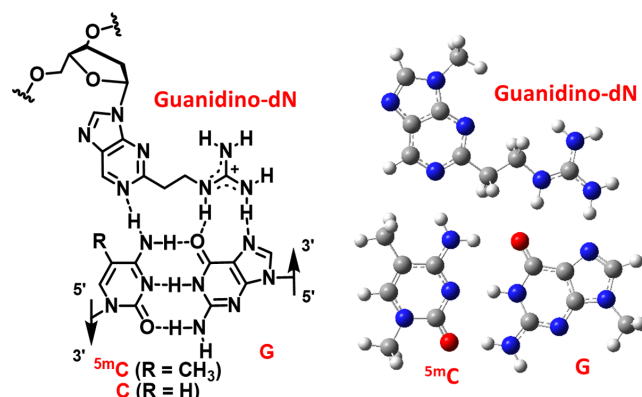
¹Graduate School of Pharmaceutical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan and ²Graduate School of Pharmaceutical Sciences, Nagasaki International University, 2825-7 Huis Ten Bosch Machi, Sasebo city, Nagasaki 859-3298, Japan

Received July 08, 2022; Revised October 26, 2022; Editorial Decision October 27, 2022; Accepted November 05, 2022

ABSTRACT

The formation of triplex DNA is a site-specific recognition method that directly targets duplex DNA. However, triplex DNA formation is generally formed for the GC and AT base pairs of duplex DNA, and there are no natural nucleotides that recognize the CG and TA base pairs, or even the 5-methyl-CG (^{5m}CG) base pair. Moreover, duplex DNA, including ^{5m}CG base pairs, epigenetically regulates gene expression *in vivo*, and thus targeting strategies are of biological importance. Therefore, the development of triplex-forming oligonucleotides (TFOs) with artificial nucleosides that selectively recognize these base pairs with high affinity is needed. We recently reported that 2'-deoxy-2-aminonebularine derivatives exhibited the ability to recognize ^{5m}CG and CG base pairs in triplex formation; however, this ability was dependent on sequences. Therefore, we designed and synthesized new nucleoside derivatives based on the 2'-deoxy-nebularine (dN) skeleton to shorten the linker length connecting to the hydrogen-bonding unit in formation of the antiparallel motif triplex. We successfully demonstrated that TFOs with 2-guanidinoethyl-2'-deoxynebularine (guanidino-dN) recognized ^{5m}CG and CG base pairs with very high affinity in all four DNA sequences with different adjacent nucleobases of guanidino-dN as well as in the promoter sequences of human genes containing ^{5m}CG base pairs with a high DNA methylation frequency.

GRAPHICAL ABSTRACT



INTRODUCTION

The formation of triplex DNA has a great potential to become a biological or medicinal tool. In order to form a stable triplex DNA, triplex-forming oligonucleotides (TFOs) interact with a homopurine strand of target duplex DNA from the major groove. In formation of the antiparallel motif triplex, the guanine (G) and adenine (A) nucleobases of TFOs form two reverse Hoogsteen hydrogen bonds to the GC and AT base pairs of target duplex DNA, thereby forming stable G/GC and A/AT triplets (Figure 1A). However, natural nucleosides do not form two hydrogen bonds with CG, TA and ^{5m}CG base pairs (Figure 1B) (1,2). These base pairs are called mismatch sites in the formation of triplex DNA, and their presence prevents further development of biological tools, etc. (3–6).

5-Methylcytosine (^{5m}C) of the ^{5m}CG base pair, which is one of the mismatch sites, is generated by the methylation of the 5-carbon atom of the cytosine (C) nucleobase and regulates gene expression *in vivo* (7–10). This small chemical modification is mainly induced by DNA methyltransferase

*To whom correspondence should be addressed. Tel: +81 92 642 6569; Fax: +81 92 642 6876; Email: taniguch@phar.kyushu-u.ac.jp

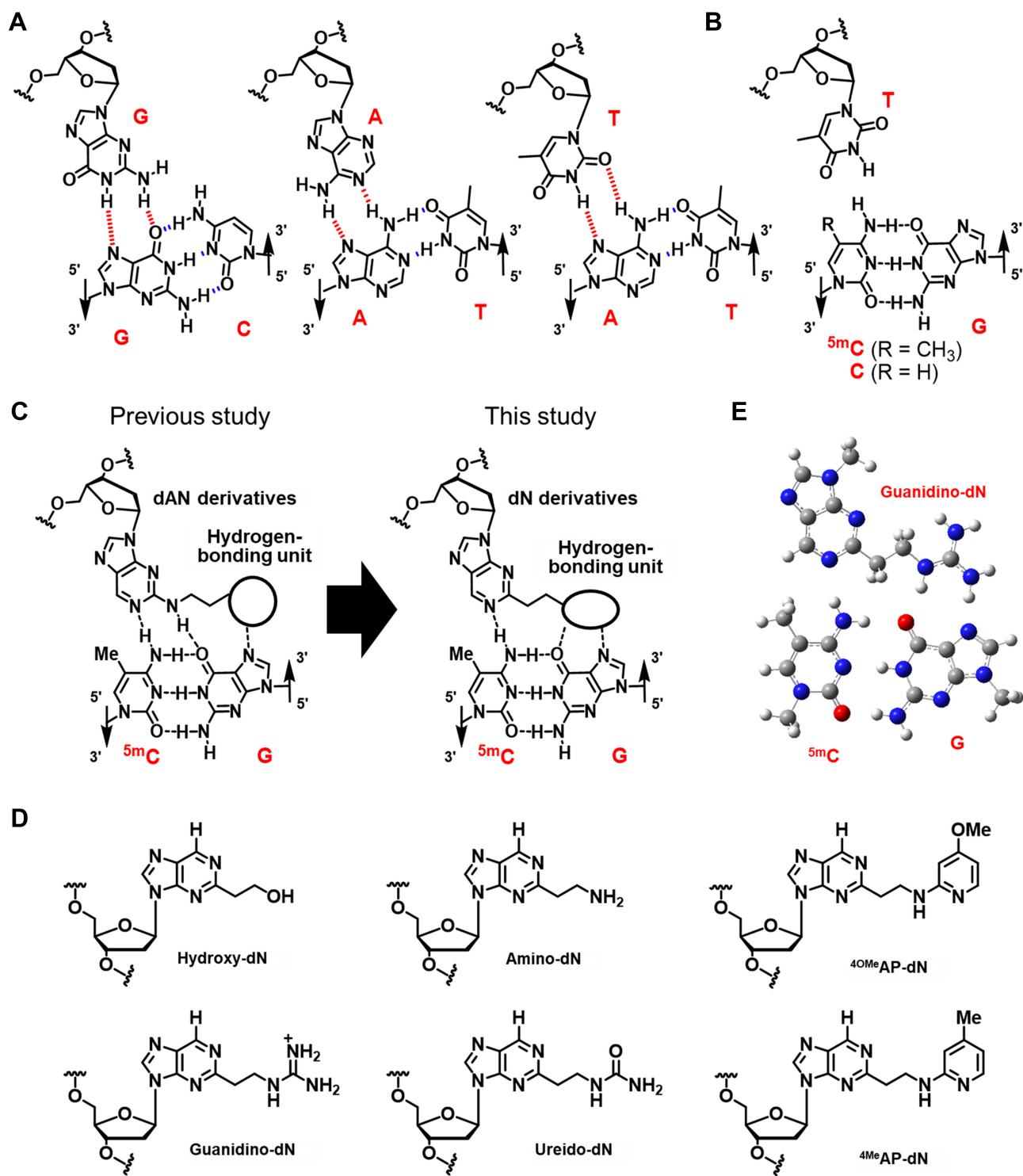


Figure 1. Molecular structures of dN derivatives in the present study. (A) G/GC, A/AT and T/AT triplets. (B) ^{5m}C/CG and CG mismatch sites. (C) Molecular design with a shortened linker portion. (D) Structures of dN derivatives. (E) Predicted recognition structure of the guanidino-dN/^{5m}C triplet.

1 (DNMT1), which functions to maintain methylation patterns after DNA replication (11). Methylated binding domain proteins recognize and bind ^{5m}C, resulting in the aggregation of the chromatin structure, which inhibits transcription factor functions and suppresses gene expression (12,13). Therefore, DNA methylation is one of the important epigenetic modifications in living cells. These chemical modifications are frequently located in CpG islands, which are methylated to 60–80% (14–16). Furthermore, the abnormal methylation of CpG that accumulates in promoter regions has been implicated in many genetic diseases (17–21). Therefore, the ^{5m}CG base pair is an important genomic target, and molecules that directly target the ^{5m}CG base pairs of duplex DNA are attracting increasing attention for their potential as chemical tools to prevent intracellular demethylation and DNA maintenance methylation processes. ^{5m}C has been identified in genomic DNA by bisulfite sequencing or nanopore sequencing methods (22,23). A photo-cross-linking reaction using a psoralen-conjugated oligonucleotide and a cationic comb-type co-polymer has been employed to detect ^{5m}C in duplex DNA (24). However, these methods are difficult to apply to living cells. Therefore, the formation of triplex DNA may be used to directly target and recognize ^{5m}C using TFOs, which specifically form triplex DNA sequences in the major groove of duplex DNA (25–28).

In the process of developing artificial nucleoside analogs that recognize these mismatch sites (29–34), we designed TFOs containing 2'-deoxy-2-aminonebularine (dAN), aminoethyl-dAN, which recognized ^{5m}CG and CG base pairs in a target duplex DNA (Figure 1C) (35). This finding prompted us to develop further derivatives. Therefore, we focused on the 2'-deoxynebularine (dN) skeleton using a shorter linker length and closer hydrogen bonding unit to the G nucleobase within the ^{5m}CG or CG base pair (Figure 1C). An ethylene linker was directly introduced from the 2-position of the dN skeleton, and a hydrogen-bonding unit was placed at the end of the linker. As hydrogen-bonding units, we adopted a hydroxyl group (hydroxyl-dN) or amino group (amino-dN), which can form one additional hydrogen bond, or a guanidino group (guanidino-dN) or ureido group (ureido-dN), which are expected to form two hydrogen bonds, as well as a para-substituted aromatic pyridine ring (^{4OMe}AP-dN and ^{4Me}AP-dN) (Figure 1D). Furthermore, the predicted base triplet structure of guanidino-dN and the ^{5m}CG base pair confirmed that the N¹-position of nebularine and the amino group at the 4-position of ^{5m}C, the guanidino group and carbonyl group at the 6-position, and the nitrogen atom at the 7-position of G were each in positions at which hydrogen bonding was possible (Figure 1E). We herein described the synthesis of artificial dN derivatives and TFOs with these, as well as the evaluation of their triplex-forming ability.

MATERIALS AND METHODS

General

All starting materials, reagents and solvents were purchased from Sigma-Aldrich, Inc., Tokyo Chemical Industry Co., Ltd and Nacalai Tesque, Inc., and analytical reagents were

used without further purification. ¹H-nuclear magnetic resonance (NMR) (500 MHz), ¹³C-NMR (125 MHz) and ³¹P-NMR (202 MHz) spectra were recorded by Varian UNITY-500 and Bruker Ascend-500 spectrometers. High-resolution mass spectra were recorded by Bruker microTOF II. Spectra obtained by matrix-assisted laser desorption/ionization-time of flight-mass spectrometry (MALDI-TOF-MS) were recorded by Bruker Autoflex II. DNA concentrations were measured by NANODROP ONE. The gel was visualized using the FUJIFILM Luminoimage analyzer LAS-4000.

Assessment of pKa values of the nitrogen atom on aminopyridine units

The pKa assessment of the aminopyridine units of ^{4OMe}AP-dN (11a) and ^{4Me}AP-dN (11b) was performed using ¹H-NMR (Supplementary Figure S1). Corresponding diol compounds were dissolved in 0.1 M Na₂HPO₄ in D₂O and titrated with 0.5 M citric acid in D₂O. ¹H-NMR spectra were collected at each point of the titration. The chemical shift in the proton on the pyridine ring was plotted versus the pD value. The half equivalence point was evaluated, and pKa in H₂O was re-calculated according to the following equation: pH = pD + 0.40 (32)

Prediction of pKa values of the nitrogen atom on recognition units

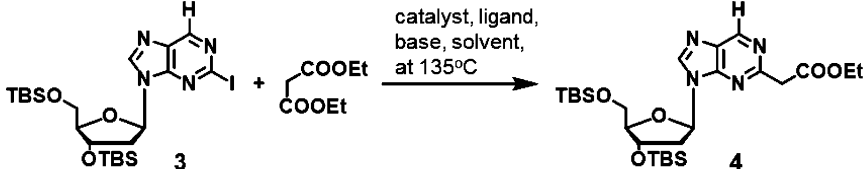
The free energy of each hydrogen bond state was calculated by the Hartree–Fock calculation using Gaussian 16. Geometry optimization and frequency were performed with the 6-31 + G (d,p) basis set using the B3LYP/IEFPCM model. Calculated results were applied to the following equation to calculate pKa in water: pKa = ΔG_{soln}(HA)/2.303RT. The free energy of H⁺ was –270.26 kcal/mol (36).

Optimization of base triplet structures of dN derivatives and ^{5m}CG base pairs

Predicted base triplet structures were optimized by the DFT calculation using Gaussian 16. Geometry optimization was performed with the 6-31 + G (d,p) basis set using B3LYP for the complex structure between the artificial nucleobase and the ^{5m}CG base pair *in vacuo*.

Preparation of TFOs incorporating dN derivatives

The 18-mer TFOs (3'-GGAAGGGZGGAGGAGGGA-5', 3'-GGAAGGGZAGAGGAGGGA-5', 3'-GGAAGGAZGGAGGAGGGA-5' and 3'-GGAAGGAZAGAGGAGGGA-5' Z = hydroxy-dN, amino-dN, guanidino-dN, ureido-dN, ^{4OMe}AP-dN or ^{4Me}AP-dN) and RASSF1A TFOs (5'-GGAGGAAGZAGGGGAGGAG-3' and 5'-GGGGGGZGGAAZGGGAAGG-3', Z = guanidino-dN) were synthesized on an automated DNA synthesizer (Nihon Techno Service Co., Ltd) using standard phosphoramidite chemistry. Cleavage from Controlled Pore Glass (CPG) was achieved by a treatment with 28% ammonia

Table 1. Optimization of coupling reaction conditions


entry	catalyst	ligand	solvent	base (equiv.)	yield (%)
1	CuI	2-picolinic acid	DMF	3.0	n.d.
2	Pd(OAc) ₂	Xantphos	diethyl malonate	3.0	16
3	Pd(OAc) ₂	Xantphos	diethyl malonate	2.0	49
4	Pd(OAc) ₂	Xantphos	diethyl malonate	1.5	87

Conditions: catalyst (0.1 eq.), ligand (0.2 eq.), 135°C, overnight. n.d. = not detected.

aqueous solution at 55°C overnight, followed by reverse-phase high-performance liquid chromatography (HPLC) purification [column, Shiseido CAPCELL PAK C18, type MG, 4.6 × 250 mm; solvents, A: 0.1 M triethylamine acetate buffer (pH 7.0), B: CH₃CN, linear gradient: B for 10–40%/20 min, flow rate: 1.0 ml/min, UV: 254 nm, column oven: 35°C]. The DMTr group was removed in 5% acetic acid aqueous solution at room temperature for 15 min. The structure and purity of TFOs were confirmed by MALDI-TOF-MS (Supplementary Table S1; Supplementary Figures S2–S7).

General procedure for evaluating the triplex-forming ability of TFOs

FAM-labeled duplex DNA (100 nM) was incubated with increasing concentrations of TFOs (0–1000 nM) in buffer containing 20 mM Tris–HCl and 5 mM MgCl₂ at 37°C for 16 h and pH 7.5. Electrophoresis was performed at 4°C for 4 h using a 10% non-denaturing polyacrylamide gel. Faster mobility bands were observed as unreacted duplex DNA, while slower mobility bands were detected as triplex DNA, which was a complex between duplex DNA and TFOs. The gel was visualized using the Luminoimage analyzer LAS-4000 (FUJIFILM) (Supplementary Figure S8), and the fluorescence intensity of each band was quantified for the calculation of association constants: K_s (10⁶/M) = [Triplex]/([TFO][Duplex]). All K_s values were similarly calculated using a narrow range of TFO concentrations corresponding to each TFO (Supplementary Figure S9). The standard deviation (± SD) of each K_s value was calculated from three independent experiments.

TET oxidation of DNA containing 5^mdC

Duplex DNA containing 5^mdC (24 bp; 5.0 μM) was incubated with TET1 (1.0 μg) in a buffer containing 100 mM HEPES, 1.0 mM Fe(NH₄)₂(SO₄)₂, 1.0 mM 2-oxoglutaric acid, 2.0 mM ascorbic acid, 10 mM MgCl₂ at 37°C for 3 h and pH 7.7. The reaction was stopped by heating at 90°C for 5 min.

Inhibition of TET oxidation by triplex formation

Duplex DNA containing 5^mdC (24 bp; 5.0 μM) and TFO (18-mer; 25.0 μM) having guanidino-dN were incubated in a buffer containing 100 mM HEPES, 1.0 mM Fe(NH₄)₂(SO₄)₂, 1.0 mM 2-oxoglutaric acid, 2.0 mM ascorbic acid, 10 mM MgCl₂ at 37°C for 16 h and pH 7.7. Then the mixture was treated with TET1 (1.0 μg) for 3 h.

Enzymatic digestion of DNA

Bacterial alkaline phosphatase (BAP; 0.4 U), snake venom phosphodiesterase (VPDE; 0.1 U), nuclease P1 (0.4 U) and 10× BAP buffer (1.0 μl) were added to the DNA solution and incubated at 37°C for 1 h.

RESULTS AND DISCUSSION

Synthesis of TFOs incorporating dN derivatives

The synthesis of the amidite unit of new dN derivatives is shown in Scheme S1 and was prepared using dG as the starting material. The carbonyl group was converted to the chlorine group at the 6-position of dG to obtain compound **1**. The acetyl group was then removed under dehalogenation reaction conditions, and hydroxyl groups were protected with *tert*-butyldimethylsilyl (TBS) groups to provide compound **2** in 83% yield in two steps. The exocyclic amino group of the nucleobase was converted to an iodine group, and compound **3** was coupled with diethyl malonate and decarboxylated to give compound **4** in 87% yield. Since this coupling reaction did not proceed with a copper catalyst, a palladium catalyst was used (Table 1) (37). In the case of a large number of base equivalents, a compound was obtained in which the decarboxylation reaction did not proceed; therefore, we successfully obtained a good yield by reducing base equivalents to 1.5 equivalents. The ethyl ester group was converted to a primary alcohol to give compound **5** as the key intermediate (38). This hydroxyl group was protected by a benzoyl group, and the deprotection of TBS groups was performed to obtain compound **6**. On the other hand, the hydroxyl group

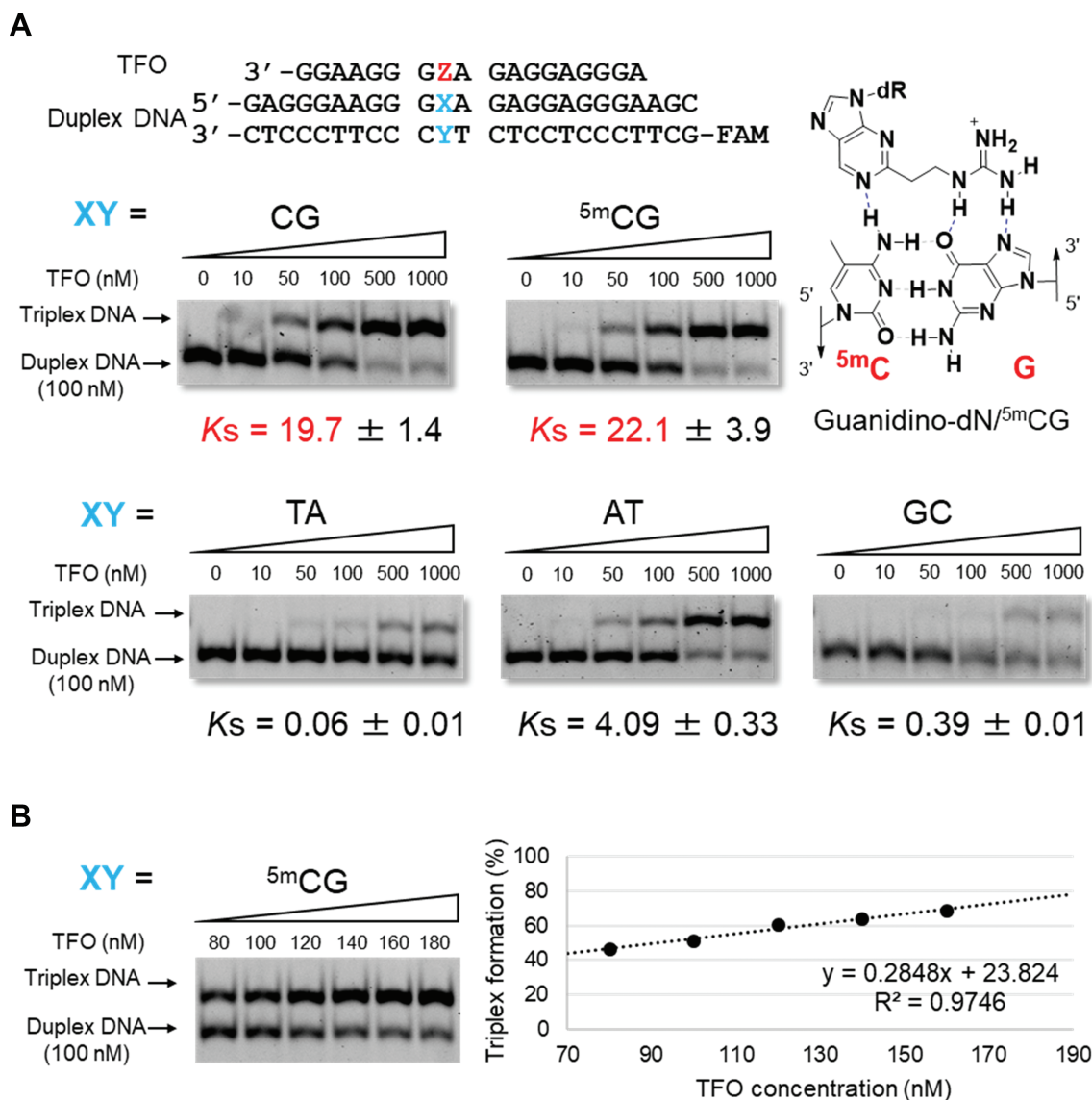


Figure 2. Gel results and determination of the triplex-forming ability of TFOs with guanidino-dN in sequences with 3'-GZA-5' (Z = guanidino-dN). (A) Conditions: FAM-labeled duplex DNA (24 bp; 100 nM) was incubated with increasing concentrations of TFOs (18-mer; 0, 10, 50, 100, 500 and 1000 nM) in buffer containing 20 mM Tris-HCl and 5 mM MgCl₂ at 37°C for 16 h and pH 7.5. Electrophoresis was performed with a 10% non-denaturing polyacrylamide gel at 4°C. (B) Conditions were the same except for TFO concentration (80, 100, 120, 140, 160 and 180 nM) and plot of the yield of formed triplex bands against the added concentration of TFO. K_s (10⁶/M) = [Triplex]/([TFO][Duplex]).

of compound **5** was converted to an azide group to give compound **7**. This azide group was reduced, and the resulting amino group was protected by an Fmoc group to give compound **8**, which was then converted to the corresponding diol compound **9**. The amino group formed by the reduction of the azide group of compound **7** was used to connect with 4-methoxypyridine or 4-methylpyridine by the Buchwald-Hartwig cross-coupling reaction to give compounds **10a** or **10b**, respectively. These compounds were converted to the corresponding diol compound **11a** or **11b**. All diol compounds (**6**, **9**, **11a** or **11b**) were protected by the dimethoxytrityl (DMTr) group of the hydroxyl group at the 5'-position, and converted to the corresponding phosphoramidite compounds (**13**, **15**, **17a** or **17b**, respectively).

The phosphoramidite units of dN derivatives were incorporated into the middle of 18-mer oligodeoxynucleotides (ODNs) using an automated DNA synthesizer. Synthesized ODNs were treated with 28% ammonia aqueous solution and purified by reverse-phase HPLC. The DMTr group was finally deprotected with 5% acetic acid aqueous solution to obtain TFOs containing hydroxy-dN, amino-dN, ⁴⁰MeAP-dN and ^{4Me}AP-dN. On the other hand, TFOs containing guanidino-dN or ureido-dN were obtained by the post-modification method using CPG-amino-dN according to a previously reported method (39). The structures and purities of all synthesized TFOs were confirmed by MALDI-TOF-MS (Supplementary Table S1; Supplementary Figures S2-S7).

Table 2. *K_s* values of each TFO in four different sequences

TFO Target Duplex		3'-GGAAGG N _Z N GAGGAGGGA 5'-GAGGGAAGG N _X N GAGGAGGGAAGC 3'-CTCCCTTC M _Y M CTCCTCCCTTCG-FAM				
3' NZN ^{5'}	Z	<i>K_s</i> (10 ⁶ /M) for XY				
		TA	AT	CG	GC	^{5m} CG
3' AZA ^{5'}	T	n.d.	8.59 ± 1.30	n.d.	n.d.	n.d.
	dN	n.d.	N.d.	n.d.	n.d.	n.d.
	Hydroxy-dN	n.d.	n.d.	n.d.	n.d.	n.d.
	Amino-dN	n.d.	n.d.	n.d.	n.d.	n.d.
	Guanidino-dN	n.d.	n.d.	5.87 ± 0.05	n.d.	4.30 ± 0.60
	Ureido-dN	n.d.	n.d.	n.d.	n.d.	n.d.
	^{40Me} AP-dN	0.21 ± 0.08	n.d.	3.43 ± 0.23	n.d.	3.51 ± 0.13
	^{4Me} AP-dN	n.d.	n.d.	6.55 ± 0.04	n.d.	2.65 ± 0.01
3' AZG ^{5'}	T	n.d.	27.8 ± 1.1	n.d.	n.d.	0.14 ± 0.03
	dN	0.09 ± 0.01	0.32 ± 0.17	0.07 ± 0.02	n.d.	n.d.
	Hydroxy-dN	n.d.	n.d.	0.89 ± 0.34	0.63 ± 0.03	0.5 ± 0.17
	Amino-dN	n.d.	n.d.	n.d.	n.d.	n.d.
	Guanidino-dN	0.22 ± 0.03	n.d.	10.6 ± 0.71	n.d.	9.43 ± 1.05
	Ureido-dN	n.d.	n.d.	0.08 ± 0.04	0.3 ± 0.07	0.82 ± 0.16
	^{40Me} AP-dN	4.10 ± 0.10	2.16 ± 0.06	7.07 ± 0.01	3.76 ± 0.15	6.94 ± 0.31
	^{4Me} AP-dN	4.54 ± 0.08	0.80 ± 0.07	9.33 ± 0.04	4.64 ± 0.03	9.86 ± 0.12
3' GZA ^{5'}	T	n.d.	46.9 ± 5.0	0.13 ± 0.05	n.d.	0.21 ± 0.04
	dN	n.d.	n.d.	0.42 ± 0.04	n.d.	n.d.
	Hydroxy-dN	n.d.	n.d.	0.16 ± 0.06	n.d.	n.d.
	Amino-dN	n.d.	0.89 ± 0.05	5.05 ± 0.39	n.d.	4.29 ± 0.06
	Guanidino-dN	0.06 ± 0.01	4.09 ± 0.33	19.7 ± 1.4	0.39 ± 0.01	22.1 ± 3.9
	Ureido-dN	n.d.	n.d.	2.11 ± 0.15	n.d.	2.07 ± 0.02
	^{40Me} AP-dN	2.62 ± 0.20	n.d.	11.7 ± 0.1	4.37 ± 0.38	14.5 ± 0.6
	^{4Me} AP-dN	1.39 ± 0.14	n.d.	15.2 ± 0.1	11.5 ± 0.3	15.1 ± 0.6
3' GZG ^{5'}	T	n.d.	1.00 ± 0.36	0.26 ± 0.01	n.d.	0.32 ± 0.12
	dN	3.54 ± 0.30	5.86 ± 0.96	12.6 ± 0.4	n.d.	3.17 ± 0.29
	Hydroxy-dN	0.36 ± 0.03	2.25 ± 0.14	4.13 ± 0.37	1.62 ± 0.04	2.71 ± 0.22
	Amino-dN	0.14 ± 0.03	0.75 ± 0.02	0.17 ± 0.02	0.45 ± 0.08	0.21 ± 0.01
	Guanidino-dN	3.29 ± 0.24	3.14 ± 0.17	20.6 ± 1.5	3.74 ± 0.30	21.4 ± 2.62
	Ureido-dN	1.31 ± 0.17	6.15 ± 0.37	5.20 ± 0.05	2.05 ± 0.03	6.11 ± 0.84
	^{40Me} AP-dN	6.88 ± 0.24	7.61 ± 0.35	9.35 ± 1.02	9.71 ± 0.89	8.91 ± 0.02
	^{4Me} AP-dN	13.3 ± 1.0	11.0 ± 0.6	20.6 ± 1.5	20.9 ± 3.1	13.4 ± 0.1

Conditions: FAM-labeled duplex DNA (24 bp; 100 nM) was incubated with increasing concentrations of TFOs (18-mer; 0–1000 nM) in buffer containing 20 mM Tris-HCl and 5 mM MgCl₂ at 37°C for 16 h and pH 7.5. Electrophoresis was performed with a 10% non-denaturing polyacrylamide gel. *K_s* (10⁶/M) = [Triplex]/([TFO][Duplex]); n.d. = not detected (*K_s* < 0.01).

Evaluation of the triplex-forming ability of TFOs containing dN derivatives

The triplex-forming abilities of synthesized TFOs were assessed by electrophoretic mobility shift assays using a non-denaturing polyacrylamide gel with FAM-labeled duplex DNAs. An example of the gel results of a TFO containing guanidino-dN with dG on its 3' side and dA on the 5' side (3'-GZA-5') is shown in Figure 2. According to the gel results, triplex DNA formation was observed following the addition of a low concentration of TFOs to a constant concentration of target duplex DNA containing a CG or ^{5m}CG base pair. Regarding other target base pairs, triplex DNA formation was detected only in the presence of a high concentration of TFOs, compared with the CG or ^{5m}CG base pairs (Figure 2A). To estimate the *K_s* value, a narrow range of TFO concentrations was evaluated and a binding curve was created by plotting the yield of the triplex bands against the added concentration of TFO (Figure 2B; Supplementary Figure S9). From this formula, the concentration of TFO that forms 50% of triplex DNA was calculated

and the *K_s* value was estimated by using the free TFO concentration (Supplementary Figure S9). This experiment was performed in triplicate, and the average value is shown as the *K_s* value below each gel (Figure 2A). To confirm the sequence dependency of the synthesized TFOs containing dN derivatives, four different types of TFOs were examined with a combination of G or A at adjacent nucleobases of the dN derivatives (3'-GZG-5', 3'-GZA-5', 3'-AZG-5' and 3'-AZA-5'; Z = dN derivatives) (Supplementary Figure S8). All *K_s* values were similarly calculated using a narrow range of TFO concentrations corresponding to each TFO, whose values are summarized in Table 2. According to these *K_s* values, the natural nucleoside of thymidine (T) can form two hydrogen bonds with adenine of the AT base pair, and thus selective triplex DNA formation was observed, although there were some differences in the stability of adjacent nucleobases (Figure 1A). Except for the TFO sequence of 3'-AZA-5', the interaction was also observed for CG or ^{5m}CG base pairs with low affinity (Table 2). The T/^{5m}CG triplet was slightly more stable than the T/CG triplet, suggesting that the methyl group may contribute to the stabiliza-

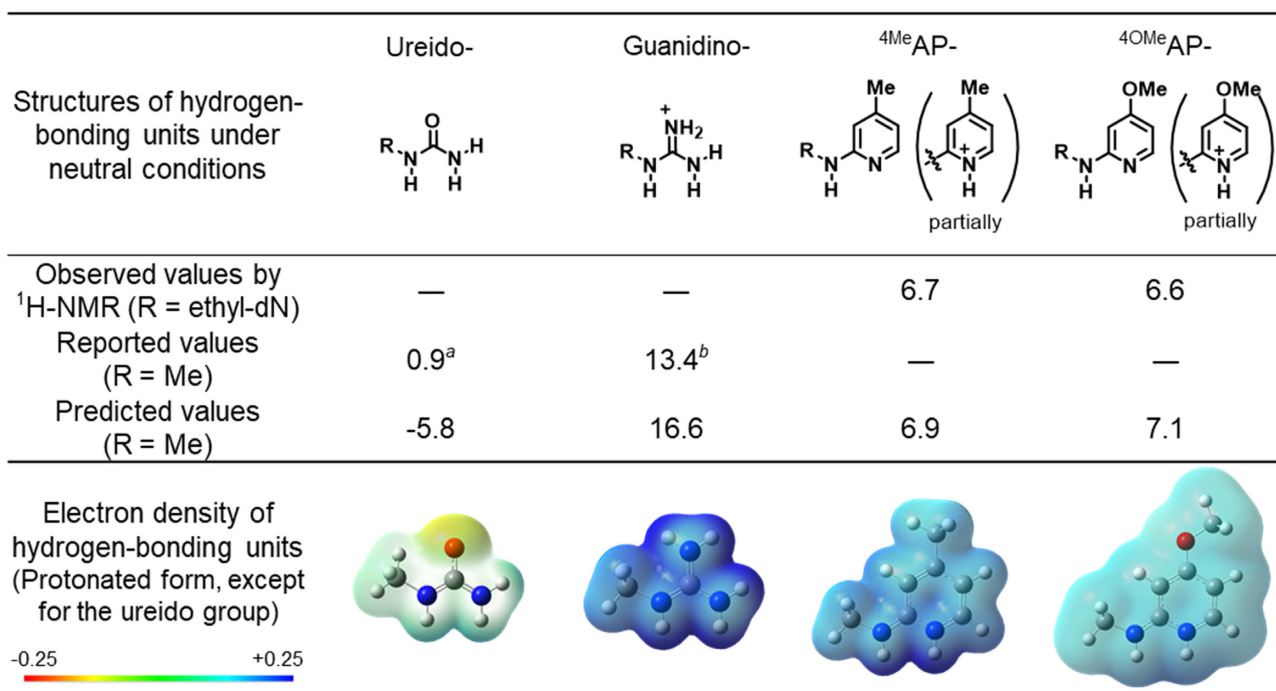


Figure 3. pK_a values and electron density maps of hydrogen-bonding units of dN derivatives. The electron density maps are shown at -0.25 (red) and +0.25 (blue). ^a pK_a value from ref 40. ^b pK_a value from ref 41.

tion effect for the base triplet structure (Figure 1B). In the case of TFOs with hydroxyl-dN or amino-dN, triplex DNA formation was observed for duplex DNA containing ^{5m}CG base pairs, although only slightly more frequently than for TFOs containing dN. This result may be due to the formation of two hydrogen bonds between the artificial nucleoside and the ^{5m}CG base pair; however, these interactions may be very weak. On the other hand, TFOs having guanidino-dN showed triplex DNA formation for target duplex DNA containing CG or ^{5m}CG base pairs in all four sequences. Although their stability was lower than that of the natural-type T/AT triplet, they showed relatively stable triplex DNA formation with duplex DNA containing ^{5m}CG or CG base pairs that cannot form stable triplexes with TFOs containing natural nucleobases. TFOs with ureido-dN showed only a slight affinity for duplex DNA containing CG and ^{5m}CG base pairs, except for the 3'-AZA-5' sequence. These results indicate that the hydrogen bonding ability of the guanidino group was stronger even with the formation of similar hydrogen bonds with the ureido group. This was the reason for the high pK_a value of the guanidino group, which was protonated under neutral conditions (Figure 3) (40,41). On the other hand, the ureido group had a very low pK_a value and existed as a neutral species under neutral conditions. Furthermore, protonation occurred at the oxygen atom of the ureido group. In a comparison of the electron densities of these hydrogen-bonding units, the guanidino group had a lower electron density, suggesting its superior hydrogen-bonding donor properties (Figure 3). These differences may alter the affinity for the ^{5m}CG base pair. TFOs incorporating ⁴OMe AP-dN and ⁴Me AP-dN, an aminopyridine ring introduced into the hydrogen-bonding site of artificial nucleosides, showed high affinity for CG and

^{5m}CG base pairs in all DNA sequences. However, with these TFOs, the formation of triplex DNA was confirmed for duplex DNA containing TA, AT and GC base pairs in the sequences 3'-GZG-5' and 3'-GZA-5'. These results can be attributed to a change in the protonation state of the molecule by the moderate pK_a of the pyrimidine ring N¹-position (Figure 3). Next, we examined the effect of monovalent cations. In general, it is known that the presence of potassium ions prevents the formation of triplex DNA due to self-aggregation of G-rich TFOs. Indeed, the newly developed TFOs containing dN derivatives were found to lose affinity in the presence of potassium ions using TFO as shown for guanidino-dN with the 3'-GZA-5' sequence (Supplementary Figure S10A). However, this reduced affinity has not affected the base selectivity (Supplementary Figure S10B). Consequentially, TFOs incorporating guanidino-dN were able to recognize CG or ^{5m}CG base pairs without sequence dependency, and triplex DNA formation was successfully extended to include methylated C.

Evaluation of the inhibition of the demethylation reaction by triplex DNA formation

Ten eleven translocation (TET) enzymes exhibit methylcytosine dioxygenase activity and contribute to intracellular demethylation reactions (42). We investigated whether the activity of TET enzymes may be inhibited by triplex DNA formation against target duplex DNA containing ^{5m}CG base pairs (Figure 4). Duplex DNA containing a ^{5m}CG base pair was treated with nuclease and analyzed by HPLC. dG and dA were detected as large peaks, dC and T as moderate peaks and ^{5m}dC as a small peak (Figure 4A, black line). Figure 4A shows an expanded region where ^{5m}dC,

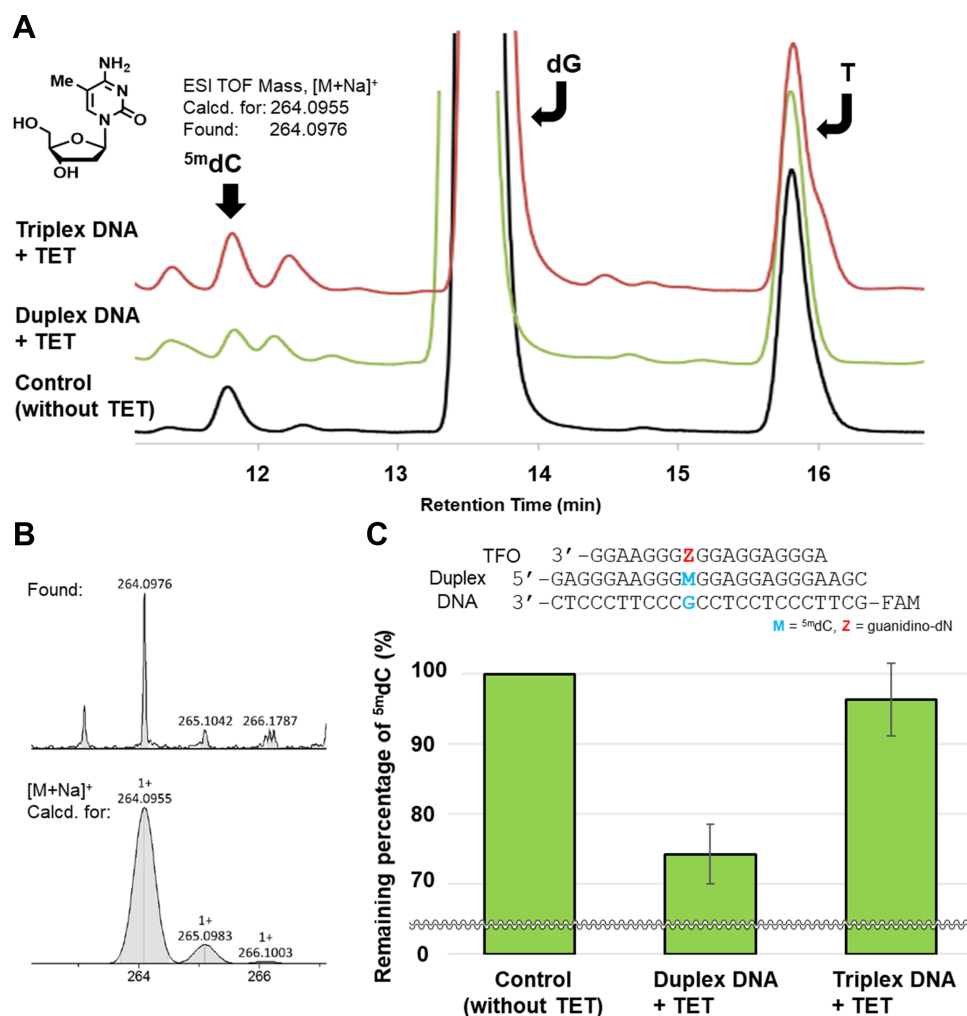


Figure 4. Inhibition of TET by triplex DNA formation. (A) HPLC chart with expanded 5^mdC , dG and T region. (B) ESI-TOF-MS result of the separated product of 5^mdC . (C) Bar graph showing the remaining percentage of 5^mdC . Control: duplex DNA containing 5^mdC (24 bp; 5.0 μM) and TFO (18-mer; 25.0 μM) incorporating guanidino-dN were incubated in buffer containing 100 mM HEPES, 1.0 mM $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$, 1.0 mM 2-oxoglutaric acid, 2.0 mM ascorbic acid and 10 mM MgCl_2 at 37°C and pH 7.7. Duplex DNA + TET: duplex DNA was incubated with TET1 in buffer at 37°C for 3 h and pH 7.7. Triplex DNA + TET: duplex DNA and TFO were incubated at 37°C for 16 h and pH 7.7, and then treated with TET1 in buffer for 3 h.

dG and T are observed. The 5^mdC peak was confirmed to be a peak by overlapping it with the standard compound. After treating duplex DNA with TET1, duplex DNA was enzymatically hydrolyzed, and analyzed by HPLC (Figure 4A, green line). The peak at 5^mdC clearly decreased, which suggested a reduction in the amount of 5^mdC as a result of the oxidation reaction. Under these conditions, when TFO containing guanidino-dN was added to form triplex DNA, a peak of similar intensity to that of the control was observed at the position of 5^mdC (Figure 4A, red line). This peak was isolated and confirmed to be 5^mdC by electrospray ionization-MS (ESI-MS; Figure 4B). The results of the quantification of 5^mdC from these peak areas are summarized in Figure 4C as a bar graph. The results obtained showed that the amount of 5^mdC was reduced by $\sim 25\%$ following the treatment with TET1 (Figure 4C). On the other hand, the treatment with TET1 after triplex DNA formation did not markedly affect the amount of 5^mdC (Figure 4C). These results demonstrate the utility of the enzyme

inhibition approach through triplex DNA formation and suggest its development as a new DNA demethylation inhibitor.

Evaluation of the ability of artificial TFOs to form triplex DNA for regions containing methylated DNA in gene sequences

To demonstrate the expansion of triplex formation using artificial nucleosides, triplex-forming ability was evaluated by targeting the *RASSF1A* gene sequence (Figure 5). The *RASSF1A* gene is a tumor suppressor gene, the expression of which is reportedly suppressed by hypermethylation of the CpG region in the promoter sequence (18). There is a purine-rich sequence between base pairs 17 842 and 17 885 of this promoter region, and the underlined C position in Figure 5A is methylated (43). Therefore, we designed TFO1 and TFO2 targeting this region and synthesized TFOs containing T as a control and guanidino-dN

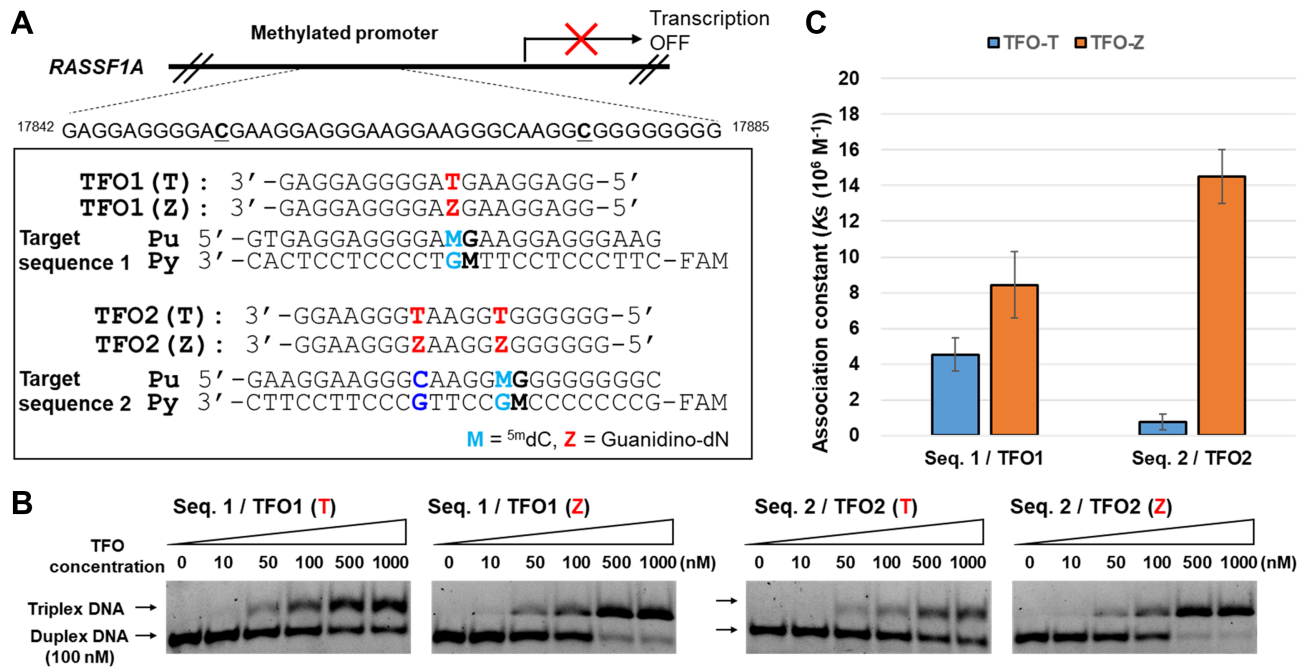


Figure 5. Gel results of the triplex-forming ability of TFOs incorporating guanidino-dN in *RASSF1A* promoter sequences containing ^{5m}dC. (A) Sequences of TFOs and target regions. (B) Gel results of the triplex formation. (C) Bar graph of K_s values. Conditions: FAM-labeled duplex DNA (25 bp; 100 nM) was incubated with increasing concentrations of TFO (19-mer; 0–1000 nM) in buffer containing 20 mM Tris-HCl and 2.5 mM MgCl₂ at 37°C for 16 h and pH 7.5. Electrophoresis was performed with a 10% non-denaturing polyacrylamide gel at 4°C. K_s (10^6 /M) = [Triplex]/([TFO][Duplex]). Blue, TFO-T/duplex DNA (control); orange, TFO-Z/duplex DNA

for each TFO (Figure 5A; Supplementary Table S2; Supplementary Figure S11). Triplex-forming ability was evaluated under a low concentration of Mg²⁺ (2.5 mM) and analyzed by a gel shift assay (Figure 5B). K_s values were obtained and are shown as bar graphs (Figure 5C; Supplementary Table S3). While the slight formation of triplex DNA was observed in TFO1(T), this was not detected in TFO2(T) because of the low affinity of T for the ^{5m}CG base pair and the presence of another CG base pair in target sequence 2. On the other hand, TFO-Z incorporating guanidino-dN showed improved stability for triplex formation in target sequences 1 and 2. Notably, TFO2(Z) incorporated two guanidino-dNs, which markedly increased the ability for triplex DNA formation by target sequence 2, which contains CG and ^{5m}CG base pairs in one target sequence (Figure 5C). Next, we evaluated the triplex-forming ability under the conditions of adding nuclear extracts of HeLa cells. Only duplex DNA was incubated in the presence of the nuclear extract. Gel electrophoresis results showed that no clear bands were observed under conditions where the reaction mixture contained >2.5% of nuclear extract (Supplementary Figure S12A). Therefore, we evaluated the ability for triplex DNA formation under the conditions in which FAM-labeled bands were observed. Although these bands were smears, triplex DNA formation was observed by the addition of a high concentration of TFO (Supplementary Figure S12B). These results indicate that TFOs with guanidino-dN form stable triplex DNA, even in methylated DNA regions, and have potential for further applications (e.g. gene expression activation or genetic recombination based on ^{5m}CG base pairs).

Conclusion

We herein designed and synthesized various dN derivatives (hydroxy-dN, amino-dN, guanidino-dN, ureido-dN, ^{40Me}AP-dN and ^{4Me}AP-dN) with a shorter linker to hydrogen-bonding units from the purine ring and evaluated their affinity for ^{5m}CG and CG base pairs in triplex DNA formation. The observed or predicted pKa of artificial nucleosides suggested that the affinity for ^{5m}CG and CG is related to the pKa and electron density of artificial nucleosides. Among synthesized dN derivatives, guanidino-dN recognized the ^{5m}CG or CG base pair with high affinity for any sequence, although its formation is inhibited by the presence of potassium ions. The formation of triplex DNA against a methylated duplex DNA sequence using this nucleoside analog may inhibit the activity of the TET enzyme that oxidizes ^{5m}CG base pairs. This is expected to sterically block the binding of TET to duplex DNA. Furthermore, to expand the triplex DNA formation region, we found that triplex DNA with an artificial nucleoside may be applied to the promoter sequence of the *RASSF1A* gene, which has a methylated CpG island. This result is very encouraging because it has the potential to recognize multiple ^{5m}CGs in future studies. Meanwhile, the properties of TFOs with multiple guanidino-dN remain unknown, and we are interested in clarifying details such as the limitations on the number of guanidino-dN to be introduced into TFOs and their effect on self-aggregation. In the future, we will examine the applicability of this artificial TFO to methylated DNA regions, and attempt to advance research on the artificial regulation of gene expression in cell studies using triplex DNA formation.

DATA AVAILABILITY

All data are available in the main article and in Supplementary Data.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

The present study was supported by a Grant-in-Aid for Scientific Research (B) [JP19H03351] from the Japan Society for the Promotion of Science (JSPS), AMED [JP21am0401026] and the JST SPRING Program [JP-MJSP2136, Japan to R.N.]. This work was also supported by the Platform Project for Supporting Drug Discovery and Life Science Research from AMED. Funding for open access charge: Grant-in-Aid for Scientific Research (B) [JP19H03351] from the Japan Society for the Promotion of Science (JSPS).

Conflict of interest statement. None declared.

REFERENCES

- Frank-Kamenetskii, M.D. and Mirkin, S.M. (1995) Triplex DNA structures. *Annu. Rev. Biochem.*, **64**, 65–95.
- Fox, K.R. (2000) Targeting DNA with triplexes. *Curr. Med. Chem.*, **7**, 17–37.
- Jain, A., Wang, G. and Vasquez, K.M. (2008) DNA triple helices: biological consequences and therapeutic potential. *Biochimie*, **90**, 1117–1130.
- Seidman, M.M. and Glazer, P.M. (2003) The potential for gene repair via triple helix formation. *J. Clin. Invest.*, **112**, 487–494.
- Hu, Y., Ceconello, A., Idili, A., Ricci, F. and Willner, I. (2018) Triplex DNA nanostructures: from basic properties to applications. *Angew. Chem. Int. Ed.*, **56**, 15210–15233.
- Chandrasekaran, A.R. and Rusling, D.A. (2018) Triplex-forming oligonucleotides: a third strand for DNA nanotechnology. *Nucleic Acids Res.*, **46**, 1021–1037.
- Chen, L., MacMillan, A.M., Chang, W., Ezaz-Nikpay, K., Lane, W.S. and Verdine, G.L. (1991) Direct identification of the active-site nucleophile in a DNA (Cytosine-5)-methyltransferase. *Biochemistry*, **30**, 11018–11025.
- Santi, D.V., Garrett, C.E. and Barr, P.J. (1983) On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. *Cell*, **33**, 9–10.
- Antequera, F., Boyes, J. and Bird, A. (1990) High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. *Cell*, **62**, 503–514.
- Hsieh, C.-L. (1994) Dependence of transcriptional repression on CpG methylation density. *Mol. Cell. Biol.*, **14**, 5487–5494.
- Yoder, J.A., Soman, N.S., Verdine, G.L. and Bestor, T.H. (1997) DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. *J. Mol. Biol.*, **270**, 385–395.
- Hendrich, B. and Bird, A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.*, **18**, 6538–6547.
- Fuks, F., Hurd, P.J., Wolf, D., Nan, X., Bird, A.P. and Kouzarides, T. (2003) The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J. Biol. Chem.*, **278**, 4035–4040.
- Goyon, C., Nogueira, T.I.V. and Faugeron, G. (1994) Perpetuation of cytosine methylation in *Ascobolus immersus* implies a novel type of maintenance methylase. *J. Mol. Biol.*, **240**, 42–51.
- Ehrlich, M., Gama-Sosa, M.A., Huang, L.-H., Midgett, R.M., Kuo, K.C., Mccune, R.A. and Gehrke, C. (1982) Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Res.*, **10**, 2709–2721.
- Bird, A.P. (1986) CpG-rich islands and the function of DNA methylation. *Nature*, **321**, 209–213.
- Esteller, M. (2007) Epigenetic gene silencing in cancer. *Hum. Mol. Genet.*, **16**, 50–59.
- Burbee, D.G., Forgacs, E., Zöchbauer-Müller, S., Shivakumar, L., Fong, K., Gao, B., Randle, D., Kondo, M., Virmani, A., Bader, S. et al. (2001) Epigenetic inactivation of RASSF1A in lung and breast cancers and malignant phenotype suppression. *J. Natl Cancer Inst.*, **93**, 691–699.
- Herman, J.G., Umar, A., Polyak, K., Graff, J.R., Ahuja, N., Issa, J.J., Markowitz, S., Willson, J.K.V., Hamilton, S.R., Kinzler, K.W. et al. (1998) Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl Acad. Sci. USA*, **95**, 6870–6875.
- Ehrlich, M. (2019) DNA hypermethylation in disease: mechanisms and clinical relevance. *Epigenetics*, **14**, 1141–1163.
- Farrell, C., Doolin, K., O'Leary, N., Jairaj, C., Roddy, D., Tozzi, L., Morris, D., Harkin, A., Frodl, T., Nemoda, Z. et al. (2018) DNA methylation differences at the glucocorticoid receptor gene in depression are related to functional alterations in hypothalamic–pituitary–adrenal axis activity and to early life emotional abuse. *Psychiatry Res.*, **265**, 341–348.
- Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R. and Gunderson, K.L. (2009) Genome-wide DNA methylation profiling using infinium[®] assay. *Epigenomics*, **1**, 177–200.
- Sakamoto, Y., Zaha, S., Nagasawa, S., Miyake, S., Kojima, Y., Suzuki, A., Suzuki, Y. and Seki, M. (2021) Long-read whole-genome methylation patterning using enzymatic base conversion and nanopore sequencing. *Nucleic Acids Res.*, **49**, e81.
- Kojima, A., Nakao, J., Shimada, N., Yoshida, N., Abe, Y., Mikame, Y., Yamamoto, T., Wada, T., Maruyama, A. and Yamayoshi, A. (2022) Selective photo-crosslinking detection of methylated cytosine in DNA duplex aided by a cationic comb-type copolymer. *ACS Biomater. Sci. Eng.*, **8**, 1799–1805.
- Thuong, N.T. and Hélène, C. (1993) Sequence-specific recognition and modification of double-helical DNA by oligonucleotides. *Angew. Chem. Int. Ed.*, **32**, 666–690.
- Praseuth, D., Guieysse, A.L. and Hélène, C. (1991) Triple helix formation and the antigene strategy for sequence-specific control of gene expression. *Biochim. Biophys. Acta*, **1489**, 181–206.
- Roberts, R.W. and Crothers, D.M. (1992) Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science*, **258**, 1463–1466.
- Rajagopal, P. and Feigon, J. (1989) Triple-strand formation in the homopurine:homopyrimidine DNA oligonucleotides d(G-A)₄ and d(T-C)₄. *Nature*, **339**, 637–640.
- Taniguchi, Y. and Sasaki, S. (2012) An efficient antigenic 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.
- Akabane-Nakata, M., Obika, S. and Hari, Y. (2014) Synthesis of oligonucleotides containing N,N-disubstituted 3-deazacytosine nucleobases by post-elongation modification and their triplex-forming ability with double-stranded DNA. *Org. Biomol. Chem.*, **12**, 9011–9015.
- 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.*, **128**, 12633–12637.
- Wang, L., Taniguchi, Y., Okamura, H. and Sasaki, S. (2018) Modification of the aminopyridine unit of 2-deoxyaminopyridinyl-pseudocytidine allowing triplex formation at CG interruptions in homopurine sequences. *Nucleic Acids Res.*, **46**, 8679–8688.
- Taniguchi, Y., Magata, Y., Osuki, T., Notomi, R., Wang, L., Okamura, H. and Sasaki, S. (2020) Development of novel C-nucleoside analogues for the formation of antiparallel-type triplex DNA with duplex DNA that includes TA and dUA base pairs. *Org. Biol. Chem.*, **18**, 2845–2851.
- Ohkubo, A., Ohnishi, T., Nishizawa, S., Nishimura, Y. and Hisamatsu, S. (2020) The ability of a triplex-forming oligonucleotide to recognize T–A and C–G base pairs in a DNA duplex is enhanced by incorporating N-acetyl-2,7-diaminoquinoline. *Bioorg. Med. Chem.*, **28**, 115350–115356.
- Notomi, R., Wang, L., Sasaki, S. and Taniguchi, Y. (2021) Design, synthesis of purine nucleoside analogues for the formation of stable

- anti-parallel type triplex DNA formation with duplex DNA bearing ^{5m}CG base pair. *RSC Adv.*, **11**, 21390–21396.
36. Bahers, L.T., Adamo, C. and Ciofini, I. (2010) Photophysical properties of 8-hydroxyquinoline-5-sulfonic acid as a function of the pH: a TD-DFT investigation. *J. Phys. Chem. A*, **114**, 5932–5939.
37. Song, B., Rudolphi, F., Himmeler, T. and Gooßen, L.J. (2011) Practical synthesis of 2-arylacetic acid esters via palladium-catalyzed dealkoxycarbonylative coupling of malonates with aryl halides. *Adv. Synth. Catal.*, **353**, 1565–1574.
38. Horiba, M., Yamaguchi, T. and Obika, S. (2016) Synthesis of scpBNA-^mC, -A, and -G monomers and evaluation of the binding affinities of scpBNA-modified oligonucleotides toward complementary ssRNA and ssDNA. *J. Org. Chem.*, **81**, 11000–11008.
39. 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.
40. Penedo, J.C., Mosquera, M. and Rodríguez-Prieto, F. (2000) Role of hydrogen-bonded adducts in excited-state proton-transfer processes. *J. Phys. Chem. A*, **104**, 7429–7441.
41. Guillén Schlippe, Y.V. and Hedstrom, L. (2005) Guanidine derivatives rescue the arg418ala mutation of *Tritrichomonas foetus* IMP dehydrogenase. *Biochemistry*, **44**, 16695–16700.
42. Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. and Zhang, Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300–1303.
43. Dreijerink, K., Braga, E., Kuzmin, I., Geil, L., Duh, F., Angeloni, D., Zbar, B., Lerman, M.I., Stanbridge, E.J., Minna, J.D. *et al.* (2001) The candidate tumor suppressor gene, RASSF1A, from human chromosome 3p21.3 is involved in kidney tumorigenesis. *Proc. Natl Acad. Sci. USA*, **98**, 7504–7509.