Organic & Biomolecular Chemistry

Dynamic Article Links

Cite this: *Org. Biomol. Chem.*, 2012, **10**, 1007

www.rsc.org/obc

PAPER

DNA duplexes and triplex-forming oligodeoxynucleotides incorporating modified nucleosides forming stable and selective triplexes†

Takashi Kanamori, Yoshiaki Masaki, Masahiro Mizuta, Hirosuke Tsunoda, Akihiro Ohkubo, Mitsuo Sekine* and Kohji Seio*

Received 18th August 2011, Accepted 19th October 2011
DOI: 10.1039/c1ob06411h

We have previously reported DNA triplexes containing the unnatural base triad G-PPI·C3, in which PPI is an indole-fused cytosine derivative incorporated into DNA duplexes and C3 is an abasic site in triplex-forming oligonucleotides (TFOs) introduced by a propylene linker. In this study, we developed a new unnatural base triad A-ψ·C^{R1} where ψ and C^{R1} are base moieties 2′-deoxypseudouridine and 5-substituted deoxycytidine, respectively. We examined several electron-withdrawing substituents for R1 and found that 5-bromocytosine (C^{Br}) could selectively recognize ψ. In addition, we developed a new PPI derivative, PPI^{Me}, having a methyl group on the indole ring in order to achieve selective triplex formation between DNA duplexes incorporating various Watson–Crick base pairs, such as T-A, C-G, A-ψ, and G-PPI^{Me}, and TFOs containing T, C, C^{Br}, and C3. We studied the selective triplex formation between these duplexes and TFOs using UV-melting and gel mobility shift assays.

Introduction

DNA triplexes are supramolecular structures formed by sequence-specific recognition of DNA duplexes by triplex-forming oligonucleotides (TFOs).[1] Many studies have been reported on the application of DNA triplex formation to gene regulation,[2] gene correction,[3] gene detection,[4] and DNA nanotechnology.[5]

TFOs can bind to DNA duplexes in either of the two orientations: parallel and anti-parallel modes. In a parallel triplex, TFOs recognize DNA duplexes that form Hoogsteen-type base pairs[1] where thymine (T) and protonated cytosine (C+) recognize adenine (A) and guanine (G) in the major groove of the duplex, respectively (Fig. 1A, B). In this paper, the Watson–Crick base pairs in the duplex are depicted with hyphens (*e.g.*, T-A, C-G), while the Hoogsteen base pairs and other interactions between a DNA duplex and TFO are represented using dots (*e.g.*, A·T and G·C+). Note that in natural-type DNAs, there are no Hoogsteen-type base pair structures that allow binding with the pyrimidine bases in duplexes. Thus, when natural-type DNA duplexes and TFOs are used, duplexes that can bind to TFOs are restricted to those with homopyrimidine–homopurine sequences, and the TFO should be composed of pyrimidine bases.

A number of studies for developing unnatural nucleic acids to overcome this sequence limitation have been reported.[6–8] For example, Fox *et al.* reported that all four base pairs (A-T, G-C, C-G, and T-A) could be recognized by TFOs containing four different unnatural bases including BAU [2′-aminoethoxy-5-(3-aminoprop-1-ynyl)uridine],[6]

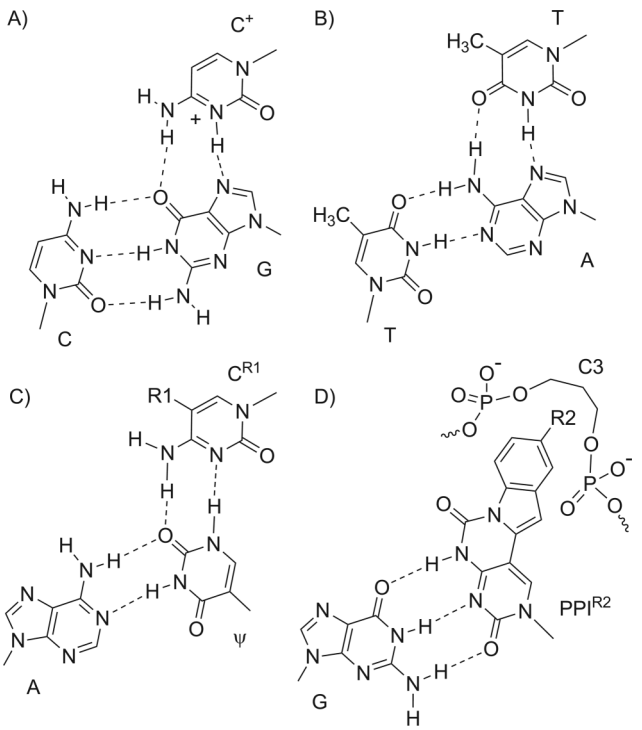


Fig. 1 The four base triads introduced in this study. A) Canonical C-G·C+, B) T-A·T, C) A-ψ·C^{R1} (R1 = F, Cl, Br, CN), and D) G-PPI^{R2}·C3 (R2 = H, Me, OMe).

4259, Nagatsuta, Midori-ku, Yokohama, Japan. E-mail: kseio@bio.titech.ac.jp, msekine@bio.titech.ac.jp; Fax: +81-45-924-5144; Tel: +81-45-924-5136

† Electronic supplementary information (ESI) available: Synthetic schemes, procedures, ^{1}H, ^{13}C and ^{31}P NMR spectra of newly synthesized compounds.. See DOI: 10.1039/c1ob06411h

^{Me}P (3-methyl-2-aminopyridine),[9] ^{A}PP (6-(3-aminopropyl)-7-methyl-3*H*-pyrrolo[2,3-*d*]pyrimidin-2(7*H*)-one),[6] and S [*N*-(4-(3-acetamidophenyl)thiazol-2-yl-acetamide)].[10] They designed ^{A}PP and S as unnatural nucleobases that recognize C and T in the duplexes, respectively. In addition to this approach, the recognition of pyrimidine bases in duplexes by modified bases in TFOs can be theoretically achieved by incorporating unnatural bases into the DNA duplexes. This concept was first proposed by Trapane and Ts'o *et al.* who reported the design of various base triads[11] and later reported experimental evidence of the formation of the triad adenine (A)-pseudouracil (uracil-5-yl, ψ)·cytosine (C) (Fig. 3C).[12] Similarly, Switzer *et al.* reported another A-ψ·A triad.[13] In addition, Ganesh *et al.* reported DNA duplexes incorporating 5-aminouracil (U#),[14,15] which is recognized by adenine in TFOs.[12] We have reported a nucleotide triad with PPI, an indole-fused cytosine derivative, incorporated into DNA duplexes as shown in Fig. 1D (R2 = H) and Fig. 2.[16] PPI can form a Watson–Crick base pair with guanosine in DNA duplexes, and its indole ring, which extrudes into the major groove, can selectively interact with an abasic site in TFOs containing a propylene linker (C3). This triad is stabilized by stacking interactions between PPI and the upstream and downstream nucleobases of C3 in TFOs without forming hydrogen bonds. These results clearly suggest that the combined use of unnatural pyrimidine residues such as ψ, U#, and PPI incorporated into DNA duplexes and appropriately designed TFOs is a promising strategy for expanding the base triad codes of DNA triplexes.

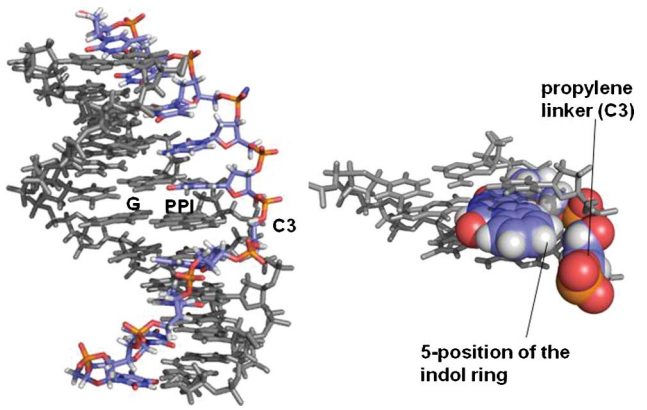


Fig. 2 3D structure of the DNA triplex incorporating a G-PPI·C3 triad and the extracted structure of the G-PPI·C3 triad obtained by the MD simulation. Here, PPI and C3 are shown in CPK style.

As the next step, we aimed to develop a set of four base triads including T-A·T, C-G·C+, A-ψ·C^{R1}, and G-PPI^{R2}·C3, where C^{R1} and PPI^{R2} are a 5-substituted cytosine derivative and an indole-fused cytosine derivative, respectively (Fig. 1C, D). In theory, different 4^{n} *n*-mer TFOs and 4^{n} DNA duplexes that form DNA triplexes in a sequence-selective manner can be designed using the four base triads.

In this paper, we report the triplex formation between TFOs containing various modified cytosine bases C^{R1} (R1 = Br, F, Cl, and CN) and DNA duplexes incorporating ψ. In addition, we describe the interaction of new PPI derivatives, PPI^{R2} (R2 = CH_3 and OCH_3), in DNA duplexes with C3 in TFOs. Finally, we combined these unnatural base triads with canonical Hoogsteen-type base triads and developed a set of four base triads capable of sequence-selective triplex formation. As a result, we found that the combined use of C^{Br} and PPI^{Me} gave the best result in terms of stability and selectivity upon DNA triplex formation.

Results and discussion

Design of C^{R1} and properties of the corresponding nucleosides

We designed the 5-substituted cytosine analog C^{R1}, which can be incorporated into TFOs by standard phosphoramidite chemistry and can recognize the A-ψ base pair in the duplexes. Ts'o *et al.*[12] previously reported that the cytosine base (C) can bind to A-ψ to give the triad A-ψ·C. However, in our case, C itself could not be used in TFOs because our system was designed to use its protonated form (C+) for recognizing G in the duplex, as shown in Fig. 1A. This combination forms a very stable G·C+ Hoogsteen base pair. Thus, we decided to develop less basic analogs of C, which could form stable base pairs with ψ, as shown in Fig. 1C and 3D, but could not form stable Hoogsteen-type base pairs with G because they would remain in an unprotonated form under weakly acidic conditions (Fig. 3B).

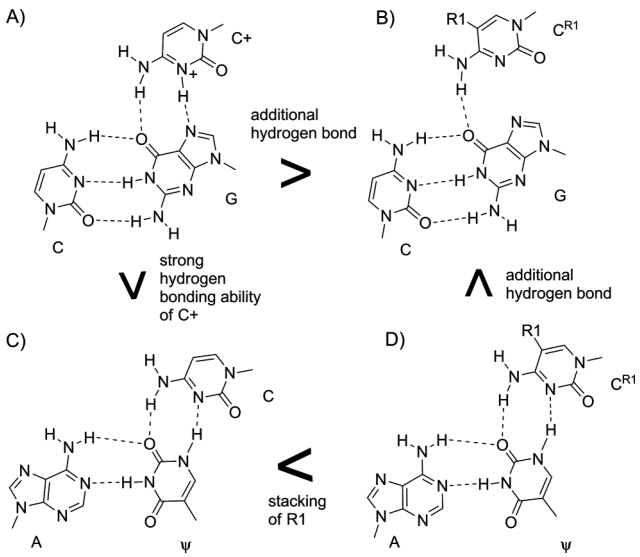


Fig. 3 Relative stabilities among A) C-G·C+ matched triad, B) C-G·C^{R1} mismatched triad, C) A-ψ·C mismatched triad, and D) A-ψ·C^{R1} (R1 = Br, Cl, F, or CN) matched triad indicated by the inequality signs.

For this purpose, we designed several cytosine analogs (C^{R1}) having electron-withdrawing groups such as cyano, fluoro, chloro, and bromo at the 5-position of the cytosine ring. By this modification, the basicity at N3 of the cytosine ring would be decreased, thereby avoiding protonation at this position. We expected that this unprotonated C^{R1} would bind more tightly to ψ than to G because ψ·C^{R1} formed two hydrogen bonds (Fig. 3D), while G·C^{R1} had only one (Fig. 3B). In addition, the triplex incorporating ψ·C^{R1} would be more stable than that incorporating ψ·C because of the stacking interactions between the R1 group and the nearby bases (Fig. 3C and 3D).

To confirm these expectations, we first checked the basicity of the deoxynucleosides dC^{Br} (**1a**),[17,18] dC^{Cl} (**1b**),[18,19,20] dC^{CN} (**1c**),[21,22] and dC^{F} (**1d**)[23] having C^{R1} as the aglycon (Fig. 4). The reported pK_a

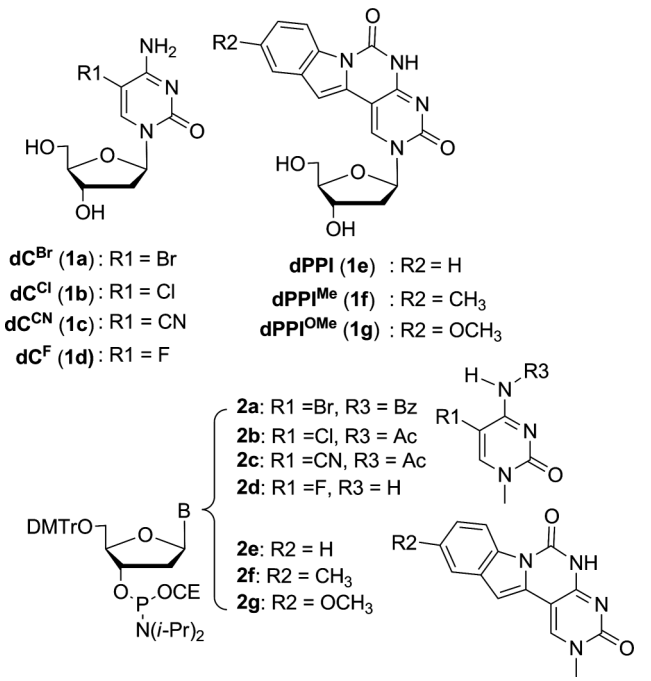


Fig. 4 Structures of the nucleosides and their phosphoramidites having C^{R1} and PPI^{R2} bases.

values of dC^F and dC^{Br} are 2.7[24] and 2.8,[25] respectively. The pK_a of dC^{CN} was determined to be 1.7 by a UV-titration experiment (see Supporting Information†). Considering the previous result[26] that pK_a of deoxycytidine derivatives is proportional to the Hammett constants of the substituents at the 5-position, dC^{Cl} was also expected to be less basic than dC. These pK_a values indicated that when incorporated in TFOs, these modified deoxycytidines would favor the unprotonated forms under the conditions used for DNA triplex formation (pH 5.4) and form a weak hydrogen bond with the guanine residue in the duplex, as shown in Fig. 3B. To confirm this hypothesis, we studied the stability and selectivity of A-ψ·C^{R1} base triads, where C^{R1} is C^F, C^{Cl}, C^{Br}, or C^{CN}, by measuring the T_m values of the DNA triplexes incorporating them.

Synthesis of phosphoramidite units and oligodeoxynucleotides for the study of A-ψ·C^{R1} base triads

To measure the thermal stability of the DNA triplexes, we synthesized hairpin-type DNA duplexes, **HP-cg** and **HP-aψ** (Fig. 5), incorporating C-G and A-ψ base pairs, respectively, at the X-Y positions. These hairpin DNAs were synthesized on an automated synthesizer using commercially available phosphoramidite units. We also synthesized **TFO-F**, **TFO-Cl**, **TFO-Br**, **TFC-CN**, and **TFO-C**, where the residues at the Z position are C^F, C^{Cl}, C^{Br}, C^{CN}, and C, respectively. In these HP and TFO combinations, nucleobases at positions X, Y, and Z are arranged to form the base triad X-Y·Z.

For the synthesis of these TFOs, the phosphoramidites of dC^{Br} (**2a**), dC^{Cl} (**2b**), and dC^{CN} (**2c**) were synthesized from nucleosides **1a–c**, as described in the Supporting Information†. The phosphoramidite of dC^F (**2d**) was synthesized according to the known procedure.[23]

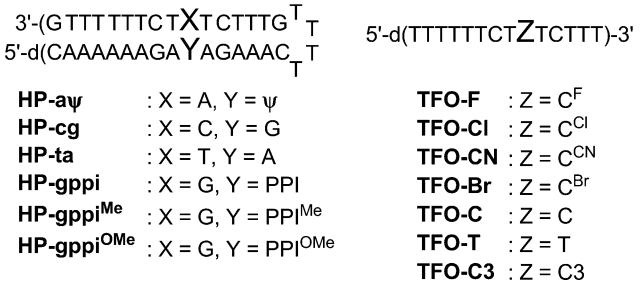


Fig. 5 Sequences of hairpin DNAs incorporating ψ and PPI derivatives (left) and TFOs incorporating C^{R1} derivatives (right).

Stability and selectivity of triplexes of hairpin oligodeoxynucleotides incorporating ψ and TFOs incorporating C^{R1}

Using these oligonucleotides, UV-melting temperatures of various DNA triplexes were measured. The data are shown in Box-1 in Table 1, where the underlined values are those of the matched base triads C-G·C+ and A-ψ·C^{R1}. Initially, we checked the stability of matched **HP-aψ/TFO-F** and mismatched **HP-aψ/TFO-C** in order to clarify the effect of the electron-withdrawing group at the 5-position of cytosine. As shown in Box-1 in Table 1, the triplex with **TFO-C** (C at Z position) bound to **HP-aψ** (A-ψ base pair at the X-Y positions) had a T_m of 42 °C, and the triplex with **TFO-F** bound to **HP-aψ** demonstrated the same affinity (T_m = 42 °C). These results suggested that the electron-withdrawing fluoro group at the 5-position of cytosine did not decrease the hydrogen bond energy between ψ and C^F.

Next, we checked the stability of the mismatched triplex **HP-cg/TFO-F**. **TFO-F** had a binding affinity for **HP-cg** giving T_m = 42 °C, which was 10 °C lower than that of the **HP-cg/TFO-C** triplex (T_m = 52 °C). This result is consistent with the above mentioned hypothesis that the less basic C^F, which does not favor the protonated form under mild acidic conditions, cannot form a stable Hoogsteen-type base pair with G (Fig. 3B).

Subsequently, we incorporated C^{CN}, C^{Cl}, and C^{Br} into TFOs at the Z position. These modified residues were expected to bind more securely than C^F to **HP-aψ** because of the stacking effect of their larger substituents (CN, Cl, Br > F). The **HP-aψ/TFO-CN** triplex had a T_m value of 45 °C, which was higher than that of the **HP-aψ/TFO-F** triplex (T_m = 42 °C). In addition, **TFO-CN** showed a stronger binding affinity for **HP-aψ** than for **HP-cg**, (ΔT_m = +7 °C, Table 1). Similarly, **TFO-Cl** and **TFO-Br** also showed higher affinities for **HP-aψ** (T_m = 48 °C, both TFOs) than **HP-cg** (T_m = 45 °C, both TFOs).

Among the 5-substituted deoxycytidine, C^{CN} had the highest selectivity for the A-ψ pair over the C-G pair (ΔT_m = 7 °C), while both C^{Cl} and C^{Br} were less selective (ΔT_m = 3 °C). However, considering the affinity for the A-ψ base pair independently, both C^{Cl} and C^{Br} triplexes showed a higher T_m (48 °C) than that of the C^{CN} triplex (45 °C). Because **TFO-Cl** and **TFO-Br** showed identical T_m profiles, we used only C^{Br} for the subsequent studies.

Studies using C^{CN} and C^{Br} in combination with the G-PPI·C3 triad

Using the above A-ψ·C^{R1} base triad where C^{R1} is C^{CN} or C^{Br}, we compared the stability of the four matched base triads

Table 1 T_m values [°C] of various triplexes[a]

	TFO-F Z = C^F	-Cl = C^{Cl}	-CN = C^{CN}	-Br = C^{Br}	-C = C	-T = T	-C3 = C3
	Box-1					Box-2	
HP-aψ (X-Y = A-ψ)	42	48	45	**48**	42	36	21
HP-cg (X-Y = C-G)	42	45	38	45	**52**	21	12
HP-ta (X-Y = T-A)	n.t.	n.t.	20	21	18	**48**	14
HP-gppi (X-Y = G-PPI)	n.t.	n.t.	49	49	49	45	56
HP-gppi^{Me} (X-Y = G-PPI^{Me})	n.t.	n.t.	n.t.	45	45	45	**53**
HP-gppi^{OMe} (X-Y = G-PPI^{OMe})	n.t.	n.t.	n.t.	46	46	45	52
				Box-3			

[a] n.t. not tested

(T-A·T, C-G·C+, A-ψ·C^{R1} and G-PPI·C3) and all possible mismatched triads. To this end, we synthesized the hairpin duplexes **HP-ta** and **HP-gppi** and the TFOs **TFO-T** and **TFO-C3** as shown in Fig. 5. The PPI residue was incorporated using the previously reported phosphoramidite unit **2e**[16] (see the Supporting Information†) and C3 was introduced using commercially available 3-(4, 4′-dimethoxytrityloxy)propyl 2-cyanoethyl *N*,*N*-diisopropylphosphoramidite.

Using the hairpin duplexes and TFOs as mentioned above, we measured the T_m values of all possible triplexes. The data are shown in Box-2 in Table 1.

In Box-2 in Table 1, the underlined values are those of the matched base triads C-G·C+, T-A·T, G-PPI·C3, and A-ψ·C^{R1}. The T_m values were compared to evaluate the sequence selectivity of each hairpin DNA for the individual TFOs and of each TFO for the different hairpin DNAs. Considering the former selectivities (see T_m values in Box-2 in Table 1 horizontally), each matched T_m value is higher than the mismatched ones. In contrast, in the case of selectivities of TFOs for hairpin DNAs (see T_m values in Box-2 in Table 1 vertically), it was found that **TFO-Br** and **TFO-CN** did not selectively bind to **HP-aψ** (T_m = 48 and 45 °C, respectively) but instead bound more tightly to **HP-gppi** (T_m = 49 °C, both TFOs). This is because PPI-related mismatches are thermally stable (T_m is from 45 °C to 49 °C). Therefore, we next attempted to modify the structure of PPI to PPI^{R2} by the introduction of a substituent (R2) to suppress the T_m of these mismatched triads so that all the matched triads became more stable than the corresponding mismatched ones. For the A-ψ pair, we selected C^{Br}, which bound more tightly to A-ψ than C^{CN}, expecting that **TFO-Br** can bind more tightly to **HP-aψ** than to the modified **HP-gppi^{R2}** which incorporates PPI^{R2} in place of PPI.

We anticipated that the mismatched triplexes incorporating modified PPI^{R2} would have a T_m lower than 48 °C so that sequence-selective triplex formation would be achieved in all possible patterns.

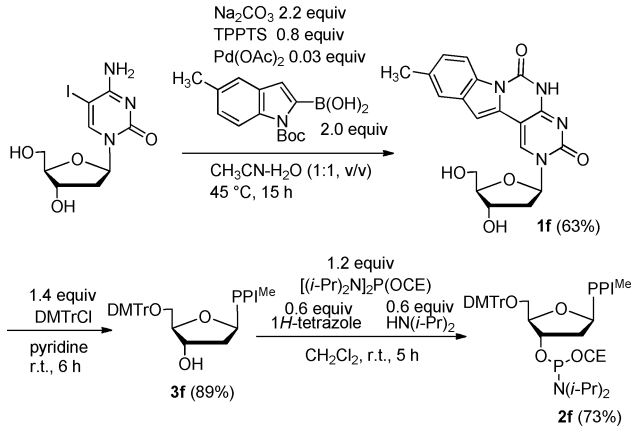
Stability and base recognition of triplexes incorporating PPI^{R2} and C3

We attempted to tune the base pairing property of PPI by introducing a substituent into its indole ring. Before the synthesis, we used a molecular dynamics simulation of a DNA triplex incorporating a G-PPI·C3 triad to identify the positions where the introduction of a substituent reduces the stability of the mismatched triads by steric repulsion, but does not significantly reduce the stability of the matched G-PPI^{R2}·C3 triad. The 3D structure obtained by the MD simulation[16] is shown in Fig. 2.

As is evident, the indole ring of PPI fits the space formed by the C3 linker and the upstream and downstream bases in the TFO strand. According to the 3D structure, the 5-position (shown by a line in Fig. 2) could be the potential site for modification because substitution at this position does not seem to interfere with binding to the TFO containing C3. Therefore, we designed two deoxynucleosides having PPI derivatives as the aglycon, including $dPPI^{Me}$ (**1f**) and $dPPI^{OMe}$ (**1g**) incorporating a methyl and methoxy group, respectively, at the 5-position (Fig. 4). The synthetic protocols for $dPPI^{OMe}$ (**1g**) and its phosphoramidite have been previously described[27] and $dPPI^{Me}$ was newly synthesized according to a procedure similar to that in our previous papers,[16,27,28] as shown in Scheme 1.

Briefly, 5-iododeoxycytidine[17] was converted into the indole nucleoside **1f** by treatment with 5-methylindole-boronic acid in the presence of a palladium catalyst and triphenylphosphine trisulfonate (TPPTS). DMTr protection followed by phosphitylation gave the phosphoramidite **2f** in 73% yield.

The hairpin duplexes **HP-gppi^{Me}** and **HP-gppi^{OMe}** were synthesized using the phosphoramidites **2f** and **2g** so that the T_m values of their triplexes with various TFOs could be measured. The hairpin DNA sequences are shown in Fig. 5 and the T_m data are listed in Box-3 in Table 1. As expected, the results show that **HP-gppi^{Me}** incorporating PPI^{Me} demonstrated lower affinities for the mismatched TFOs such as **TFO-C** and **TFO-Br** than did **HP-gppi**



Scheme 1 Synthesis **of dPPI^{Me} (1f) and its phosphoramidite 2f**.

incorporating unsubstituted PPI. T_m decreased from 49 °C (**HP-gppi/TFO-C** and **HP-gppi/TFO-Br**) to 45 °C (**HP-gppi^{Me}/TFO-C** and **HP-gppi^{Me}/TFO-Br**), which is a decrease of 4 °C. In the case of another mismatched TFO (**TFO-T**), the **HP-gppi** and **HP-gppi^{Me}** triplexes gave identical T_m values of 45 °C. Although its affinity for matched **TFO-C3** was also decreased by the incorporation of the methyl group from 56 °C (**HP-gppi/TFO-C3**) to 53 °C (**HP-gppi^{Me}/TFO-C3**), **HP-gppi^{Me}** maintained the largest T_m value with matched **TFO-C3** compared to mismatched **TFO-C**, **TFO-Br**, and **TFO-T** strands (compare second-row values horizontally). Thus, we concluded that PPI^{Me} had a higher affinity for the matched TFO incorporating C3 and lower affinity for the mismatched TFOs.

In the case of **HP-gppi^{OMe}**, the T_m profiles are similar to those of **HP-gppi^{Me}**. Thus, we selected PPI^{Me} for further studies.

Selection of four base triads capable of selective recognition

On the basis of their T_m properties, we selected C-G·C+, T-A·T, and A-ψ·C^{Br} (Box-2 in Table 1) and G-PPI^{Me}·C3 (Box-3 in Table 1) as the set of selective base triads for further studies. The T_m data are extracted from Tables 1 and listed again in Table 2.

As shown in Table 2, the T_m of the matched triplexes **HP-cg/TFO-C**, **HP-ta/TFO-T**, **HP-gppi^{Me}/TFO-C3**, and **HP-aψ/TFO-Br** shown in the diagonal cells are larger than those for the mismatched TFOs shown in the non-diagonal cells. Thus, it is expected that DNA duplexes with mixed sequences comprised of C-G, T-A, G-PPI^{Me}, and A-ψ can be selectively recognized by TFOs incorporating C, T, C3, and C^{Br}, respectively, in the appropriate sequences. To test this expectation, we synthesized the hairpin oligodeoxynucleotides **HP-1** and **HP-2** incorporating all four base pairs (T-A, C-G, A-ψ, and G-PPI^{Me}) and the corresponding **TFO-1** and **TFO-2** incorporating T, C, C^{Br}, and C3, which recognize **HP-1** and **HP-2**, respectively (Fig. 6).

First, we measured the T_m values for the fully matched triplexes **HP-1/TFO-1** and **HP-2/TFO-2** and the mismatched triplexes

Table 2 T_m summary of the set of four base triads

	TFO-C Z = C	**-T** = T	**-C3** = C3	**-Br** = C^{Br}
HP-cg (X-Y = C-G)	52	21	12	45
HP-ta (X-Y = T-A)	18	48	14	21
HP-gppi^{Me} (X-Y = G-PPI^{Me})	45	45	53	45
HP-aψ (X-Y = A-ψ)	42	36	21	48

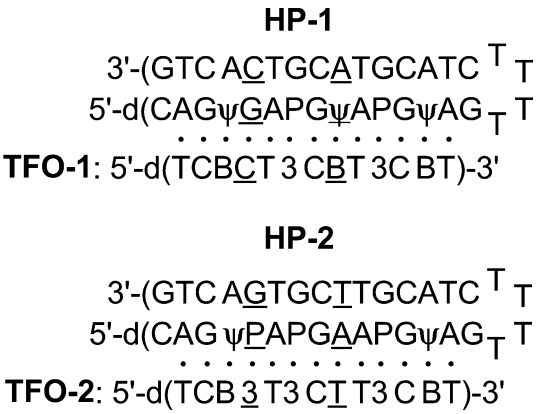


Fig. 6 Sequences of hairpin DNAs and TFOs incorporating natural and unnatural bases. B = C^{Br}; 3 = C3; P = PPI^{Me}. The underlined bases are not identical between **HP-1/TFO-1** and **HP-2/TFO-2**.

HP-1/TFO-2 and **HP-2/TFO-1**. The triplex **HP-1/TFO-2** has two C-G·C3 and A-ψ·T mismatches, whereas **HP-2/TFO-1** has G-PPI^{Me}·C and T-A C^{Br} mismatches. As a result, it was found that the matched triplexes **HP-1/TFO-1** and **HP-2/TFO-2** showed very high T_m values of 59 °C and 60 °C, respectively, while the two-base-mismatched triplexes **HP-1/TFO-2** and **HP-2/TFO-1** showed much lower T_m values of 12 °C and 28 °C, respectively. These results clearly suggest that **TFO-1** bound more strongly to the matched partner **HP-1** than to **HP-2** and, likewise, **TFO-2** substantially favored **HP-2** over **HP-1**, thereby discriminating between fully matched and two-base-mismatched hairpin duplexes.

To further evaluate the selective formation of **HP-1/TFO-1** and **HP-2/TFO-2**, the gel shift assay experiment was performed with these hairpin oligomers and TFOs. The results are shown in Fig. 7. Sequence-matched triplexes **HP-1/TFO-1** (lane 2) and **HP-2/TFO-2** (lane 6) appeared as retarded bands. In contrast, a sequence-unmatched mixture of **HP-1** and **TFO-2** (lane 3) or **HP-2** and **TFO-1** (lane 5) did not show such a retarded band. These results clearly suggested that base triads such as C-G·C+, T-A·T, G-PPI^{Me}·C3, and A-ψ·C^{Br} could selectively form base pairs as determined not only by the UV-melting conditions but also by the gel mobility shift assays.

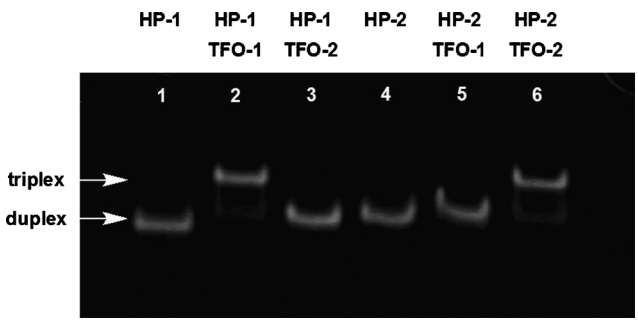


Fig. 7 Native gel shift assay using **HP-1**, **HP-2**, **TFO-1**, and **TFO-2**.

Conclusions

In this paper, we reported the development of a set of four base triads that can be incorporated into DNA duplexes and TFOs to expand the potential for sequence-selective DNA triplex

formation. We designed base triads G-PPI^{Me}·C3 and A-ψ·C^{Br} having unnatural pyrimidine bases such as PPI^{Me} and ψ, which formed Watson–Crick type base pairs with G and A in the duplexes and were selectively bound to C3 and C^{Br} in TFOs, respectively.

As shown in Table 2, the triplexes incorporating one of the matched base triads, such as G-PPI^{Me}·C3, A-ψ·C^{Br}, T-A·T, or C-G·C+, showed higher T_m than any of their mismatched counterparts. Thus, DNA duplexes having mixed base pairs such as T-A, C-G, G-PPI^{Me}, and A-ψ can be combined with TFOs that recognize them in a sequence-specific manner to generate the corresponding DNA triplexes.

In the current set of four base triads, the most stable matched triad is G-PPI^{Me}·C3, which gave a T_m of 53 °C, while the least stable matched triads are T-A·T and A-ψ·C^{Br} both of which gave a T_m of 48 °C when incorporated in the nucleotide sequences used in this study. In addition, as shown in Table 2, the T_m values of the least stable matched triads T-A·T and A-ψ·C^{Br} are only 3 °C higher than those of the most stable mismatched base triads C-G·C^{Br}, G-PPI^{Me}·C, and G-PPI^{Me}·C^{Br}, all of which have a T_m of 45 °C. Therefore, if very stringent single-base mismatch discrimination is to be achieved, the design and synthesis of new base analogs capable of further improved base discrimination might be necessary. On the other hand, these four triads formed less stable triplex under the neutral condition (pH 7.0) due to the unprotonated cytosine in the TFO (see the Supplementary Information†). Analogues of protonated cytosine such as ^{Me}P[9] might be useful for further application of our artificial triplexes under the biological conditions.

In contrast, the T_m analyses performed using **HP-1/TFO-1**, **HP-2/TFO-2**, **HP-2/TFO-1**, and **HP-1/TFO-2** and the gel electrophoretic studies indicated the ability of the current base triads to discriminate between fully matched and two-base-mismatched sequences. For example, although the triplex **HP-2/TFO-1** contains a rather stable G-PPI^{Me}·C mismatch, the formation of this triplex is substantially inhibited by the incorporation of an unstable T-A·C^{Br} mismatch as demonstrated by the gel electrophoresis.

At present, the potential application of DNA triplexes in gene regulation, gene correction, gene detection, and DNA nanotechnology are attracting significant interest. However, because the formation of unnatural base triads such as G-PPI^{Me}·C3 and A-ψ·C^{Br} required the incorporation of the unnatural bases PPI^{Me} and ψ into the duplexes, these triads cannot be directly used in gene-targeting technologies such as gene regulation, gene correction, and gene detection. In contrast, our base triads can be widely used in technologies utilizing chemically synthesized DNA duplexes such as DNA nanotechnology and the development of artificial genes. These applications of the four base triads described herein will be reported elsewhere.

Experimental

Synthesis of oligodeoxynucleotides containing modified nucleosides

All oligodeoxynucleotides containing modified nucleosides were synthesized by using a DMTr-ON protocol automated DNA/RNA synthesizer on a 1 μmol scale using the standard cycle. Removal of the protecting groups and the cleavage of the linker from resins were performed for 12–24 h in 28% aqueous ammonia at room temperature. The oligonucleotides were purified on a C18-reverse phase cartridge and the DMTr group was removed by treatment with 2% CF_3COOH/H_2O. In the synthesis of **HP-aψ**, **HP-1**, and **HP-2** containing deoxypseudouridines, the deprotection conditions were slightly modified as follows to prevent cyanoethylation at the N1 position of pseudouridine.[29] After the chain elongation, the resin was flushed with a 10% solution of DBU in MeCN to remove the cyanoethyl groups on the phosphodiester residues and washed with dry acetonitrile. After the DBU treatment, the oligonucleotides were cleaved and deprotected by using 28% aqueous ammonia for 12 h. The oligonucleotides were purified by reverse-phase HPLC using a gradient of acetonitrile in 30 mM ammonium acetate. If necessary, anion-exchange HPLC was used. The purified oligonucleotides were analyzed by MALDI-TOF-MS. The MS data are as follows. **HP-aψ** [M + H]+ Calcd 11018.86, found 11020.25; **HP-gppi** [M + H]+ Calcd 11174.89, found 11174.62; **HP-gppi^{Me}** [M + H]+ Calcd 11188.91, found 11190.79; **HP-gppi^{OMe}** [M + H]+ Calcd 11204.90, found 11207.00; **HP-1** [M + H]+ Calcd 11304.88, found 11306.31; **HP-2** [M + H]+ Calcd 11473.94, found 11475.12; **TFO-Br** [M + H]+ Calcd 4228.61, found 4228.67; **TFO-Cl** [M + H]+ Calcd 4184.66, found 4183.79; **TFO-F** [M + H]+ Calcd 4168.69, found 4168.51; **TFO-CN** [M + H]+ Calcd 4175.69, found 4176.16; **TFO-C3** [M + H]+ Calcd 3999.66, found 3998.10; **TFO-1** [M + H]+ Calcd 3992.35, found 3995.74; **TFO-2** [M + H]+ Calcd 3778.40, found 3780.14.

Thermal denaturation experiments

UV melting experiments were performed with a UV spectrometer equipped with a temperature controller. All measurements were conducted in a buffer containing 10 mM $MgCl_2$, 500 mM NaCl, 10 mM sodium cacodylate at pH 5.4. Oligonucleotides were mixed in a 1 : 1 ratio at a concentration of 2.0 μM. Melting curves were recorded at 260 nm in a consecutive heating–cooling–heating cycle (0–90 °C) with a temperature gradient of 0.5 °C min^{-1}.

Gel shift assay

10 pmol of hairpin oligonucleotides (**HP-1**, **HP-2**) and 20 pmol of TFOs (**TFO-1**, **TFO-2**) were dissolved in a 20 μL cacodylate buffer, which is the same as that used in the T_m experiments. The solution was heated at 95 °C for 5 min, and then gradually cooled to ambient temperature. We used the protocol of the gel shift assay of triplex analysis previously described by Johnston *et al.*[30] All samples were analyzed by electrophoresis at 200 V on 20% native acrylamide gels in 10 mM $Mg(OAc)_2$, 40 mM NaOAc buffer (pH 5.1) at ambient temperature (26 °C) for 2.5 h and the same buffer was used as the running buffer. After the electrophoresis, DNA bands were stained with SYBR® Gold.

Acknowledgements

This study was supported by a Grants in aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan. We also thank Center for Advanced Materials Analysis (Suzukakedai), Technical Department, Tokyo Institute of Technology, for MALDI-TOF mass analysis.

Notes and references

1 M. Duca, P. Vekhoff, K. Oussedik, L. Halby and P. B. Arimondo, *Nucleic Acids Res.*, 2008, **36**, 5123; D. P. Arya, *Acc. Chem. Res.*, 2011, **44**, 134; V. Malnuit, M. Duca and R. Benhida, *Org. Biomol. Chem.*, 2011, **9**, 326.

2 C. Hélène, *Anticancer Drug Des.*, 1991, **6**, 569; D. Praseuth, A. L. Guieysse and C. Hélène, *Biochim. Biophys. Acta.*, 1999, **1489**, 181; S. Obika, *Chem. Pharm. Bull.*, 2004, **52**, 1399; R. V. Guntaka, B. R. Varma and K. T. Weber, *Int. J. Biochem. Cell Biol.*, 2003, **35**, 22; K. R. Fox, *Curr. Med. Chem.*, 2000, **7**, 17; S. Buchini and C. J. Leumann, *Curr. Opin. Chem. Biol.*, 2003, **7**, 717; B. P. Casey and P. M. Glazer, *Prog. Nucleic Acid Res. Mol. Biol.*, 2001, **67**, 163.

3 J. M. Kalish and P. M. Glazer, *Ann. N. Y. Acad. Sci.*, 2005, **1058**, 151; Y. Liu, R. S. Nairn and K. M. Vasquez, *Nucleic Acids Res.*, 2009, **37**, 6378; D. de Semir and J. M. Aran, *Curr. Gene Ther.*, 2006, **6**, 481; C.-M. Liu, D.-P. Liu and C.-C. Liang, *J. Mol. Med.*, 2002, **80**, 620; O. Igoucheva, V. Alexeev and K. Yoon, *Curr. Mol. Med.*, 2004, **4**, 445; J. Y. Kuan and P. M. Glazer, *Methods Mol. Biol.*, 2004, **262**, 173; P. P. Chan and P. M. Glazer, *J. Mol. Med.*, 1997, **75**, 267.

4 V. N. Potaman, *Expert Rev. Mol. Diagn.*, 2003, **3**, 481; F. McKenzie, K. Faulds and D. Graham, *Chem. Commun.*, 2008, 2367.

5 Y. Chen, S. H. Lee and C. Mao, *Angew. Chem., Int. Ed.*, 2004, **43**, 5335; J. Tumpane, R. Kumar, E. P. Lundberg, P. Sandin, N. Gale, I. S. Nandhakumar, B. Albinsson, P. Lincoln, L. M. Wilhelmsson, T. Brown and B. Nordén, *Nano Lett.*, 2007, **7**, 3832; Y. H. Jung, K. B. Lee, Y. G. Kim and I. S. Choi, *Angew. Chem., Int. Ed.*, 2006, **45**, 5960.

6 D. M. Gowers and K. R. Fox, *Nucleic Acids Res.*, 1999, **27**, 1569; D. A. Rusling, V. E. Powers, R. T. Ranasinghe, Y. Wang, S. D. Osborne, T. Brown and K. R. Fox, *Nucleic Acids Res.*, 2005, **33**, 3025.

7 D. Chen Meena, S. K. Sharma and L. W. McLaughlin, *J. Am. Chem. Soc.*, 2004, **126**, 70; H. Chen Meena and L. W. McLaughlin, *J. Am. Chem. Soc.*, 2008, **130**, 13190.

8 J.-S. Li, Y.-H. Fan, Y. Zhang, L. A. Marky and B. Gold, *J. Am. Chem. Soc.*, 2003, **125**, 2084; J.-S. Li and B. Gold, *J. Org. Chem.*, 2005, **70**, 8764; J.-S. Li, F.-X. Chen, R. Shikiya, L. A. Marky and B. Gold, *J. Am. Chem. Soc.*, 2005, **127**, 12657; D. Guianvarc'h, J.-L. Fourrey, R. Maurisse, J.-S. Sun and R. Benhida, *Org. Lett.*, 2002, **4**, 4209; D. Guianvarc'h, J.-L. Fourrey, R. Maurisse, J.-S. Sun and R. Benhida, *Bioorg. Med. Chem.*, 2003, **11**, 2751.

9 S. Hildbrand, A. Blaser, S. P. Parel and C. J. Leumann, *J. Am. Chem. Soc.*, 1997, **119**, 5499.

10 D. Guianvarc'h, R. Benhida, J. L. Fourrey, R. Maurisseb and J. S. Sun, *Chem. Commun.*, 2001, 1814.

11 T. L. Trapane and P. O. P. Ts'o, *J. Am. Chem. Soc.*, 1994, **116**, 10437.

12 T. L. Trapane, M. S. Christopherson, C. D. Roby, P. O. P. Ts'o and D. Wang, *J. Am. Chem. Soc.*, 1994, **116**, 8412.

13 R. Bandaru, H. Hashimoto and C. Switzer, *J. Org. Chem.*, 1995, **60**, 786.

14 V. S. Rana, D. Barawkar and K. N. Ganesh, *J. Org. Chem.*, 1996, **61**, 3578.

15 V. S. Rana and K. N. Ganesh, *Nucleic Acids Res.*, 2000, **28**, 1162; V. S. Rana and K. N. Ganesh, *Org. Lett.*, 1999, **1**, 631.

16 M. Mizuta, J. Banba, T. Kanamori, R. Tawarada, A. Ohkubo, M. Sekine and K. Seio, *J. Am. Chem. Soc.*, 2008, **130**, 9622.

17 D. M. Frisch and D. W. Visser, *J. Am. Chem. Soc.*, 1959, **81**, 1756.

18 V. Kumar, J. Yap, A. Muroyama and S. V. Malhotra, *Synthesis*, 2009, 3957.

19 K. Kikugawa, I. Kawada and M. Ichino, *Chem. Pharm. Bull.*, 1975, **23**, 35.

20 J. Nielse, M. Taagaard, J. E. Marugg, J. H. van Boom and O. Dahl, *Nucleic Acids Res.*, 1986, **14**, 7391.

21 G. Lahoud, V. Timoshchuk, A. Lebedev, M. de Vega, M Salas, K. Arar, Y.-M. Hou1 and H. Gamper, *Nucleic Acids Res.*, 2008, **36**, 3409.

22 Y. E. Safadi, J.-C. Paillart, G. Laumond, A.-M. Aubertin, A. Burger, R. Marquet and V. Vivet-Boudou, *J. Med. Chem.*, 2010, **53**, 1534.

23 C. J. Marasco Jr. and J. R. Sufrin, *J. Org. Chem.*, 1992, **57**, 6363.

24 A. K. Ghosh and G. B. Schuster, *J. Am. Chem. Soc.*, 2006, **128**, 4172.

25 G. B. Wisdom and B. A. Orsi, *Eur. J. Biochem.*, 1969, **7**, 223.

26 C. J. La Francois, Y. H. Jang and T. Cagin, *Chem. Res. Toxicol.*, 2000, **13**, 462.

27 M. Mizuta, K. Seio, A. Ohkubo and M. Sekine, *J. Phys. Chem. B*, 2009, **113**, 9562.

28 M. Mizuta, K. Seio, K. Miyata and M. Sekine, *J. Org. Chem.*, 2007, **72**, 5046.

29 J. Ofengand, *Biochem. Biophys. Res. Commun.*, 1965, **18**, 192.

30 B. P. Belotserkovskii and B. H. Johnston, *Electrophoresis*, 1996, **17**, 1528.