Organic & Biomolecular Chemistry

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/obc

Journal Name

RSCPublishing

ARTICLE

A new modified cytosine base capable of base pairing with guanine using four hydrogen bonds

Cite this: DOI: 10.1039/xoxxoooooox

Ken Yamada, Yoshiaki Masaki, Hirosuke Tsunoda, Akihiro Ohkubo, Kohji Seio, and Mitsuo Sekine*

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/xoxxoooooox

www.rsc.org/

Oligonucleotides, containing 4-*N*-(1*H*-pyrrol-2-ylcarbonyl)deoxycytidine (dC^{Pyc}) and related derivatives, were synthesized *via* deprotection using 1.5 M NaOMe/MeOH. Among them, oligodeoxynucleotides containing dC^{Pyc} exhibited a higher hybridization affinity for DNA and RNA than the unmodified oligodeoxynucleotides. Comparative analysis between dC^{Pyc} and its derivatives by molecular dynamic simulation indicated that the C^{Pyc} residue could form four hydrogen bonds with the opposite G nucleobase keeping a more planar structure than the C^{Inc} residue where the Pyc group was replaced with a 1*H*-indol-2-ylcarbonyl group.

Introduction

The structure-based design of modified nucleobases without disturbing duplex formation is a keystone to functionalize oligonucleotides. In current nucleic acid chemistry, various types of modified cytosine derivatives have been developed. Particularly, bi- and tri-cyclic cytosine-based scaffolds, both of which contain completely planar geometry, have been extensively utilized for acquiring biochemical and biophysical properties of oligonucleotides.[1]

Among the modified cytosine bases, Matteucci reported a tricyclic aminoethyl-phenoxazine deoxycytidine, the so called "G-clamp," which could tightly bind to a guanine (G) base through the formation of four hydrogen bonds in addition to a strong π-stacking effect of the tricyclic phenoxazine ring.[1b] Egli also reported a "Guanidino G-clamp" capable of forming a base pair with G using two additional hydrogen bonds.[1c] Hudson developed [bis-*o*-(aminoethoxy)phenyl]pyrrolocytosine (boPhpC) that was designed to acquire tight binding affinities to the guanine base and sensitive fluorescent properties.[1h] These analogues achieved their goals in terms of a significant increase of the target-binding affinities of oligonucleotides and the development of sensitive fluorophore oligonucleotides. However, the synthesis of bi- or tri-cyclic cytosine scaffold analogues requires a long synthetic process.

We previously reported on the hybridization property of oligonucleotides incorporating 4-*N*-acetylcytosine bases.[2] This study disclosed that the acetyl group did not interfere with the formation of the base pair with G, keeping a slightly enhanced hybridization affinity for the complementary strands since the acetyl oxygen of 4-*N*-acetyldeoxycytidine was oriented to the 5-vinyl hydrogen *via* a unique hydrogen bond with the 5-proton in a planar structure.[3] This unique property of the 4-*N*-acetylcytosine base has recently been applied to the development of a new type of "protected" oligonucleotide probes in DNA chips in our laboratory.[4] However, we also reported that 4-*N*-aroylation of the cytosine base with benzoyl, furoyl, picolyl, and nicotinoyl groups resulted in significant destabilization of the DNA duplexes.[2b] This result was explained in terms of a lack of planarity. These *N*-acylated oligonucleotides were synthesized using a (4,5-dichloro)phthaloyl (DCP) linker that can be cleaved with DBU, since the acyl groups on the cytosine base were sensitive to aqueous basic conditions, such as concentrated aqueous ammonia required for the final deprotection step in the usual DNA synthesis.[2c] We have been searching for new acyl groups that can not only be stable as nucleobase-modifiers, even under strongly basic conditions, but can also significantly enhance the hydrogen bonding ability with G base. Particularly, we came up with an idea that the 6-carbonyl oxygen of G might simultaneously accept two donor protons if an additional NH-group could be allocated at a suitable site of the acyl moiety. In this paper, we report a synthetically accessible new cytosine base modified with a 1*H*-pyrrol-2-ylcarbonyl (Pyc) group that

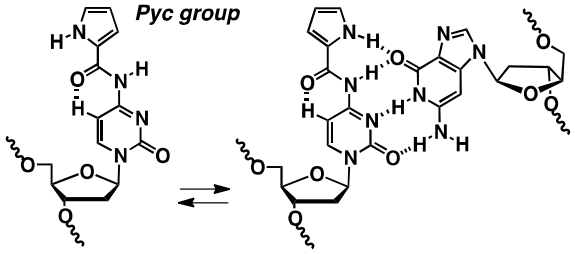


Fig. 1 A base pair between dC^{Pyc} and dG using an additional hydrogen bond.

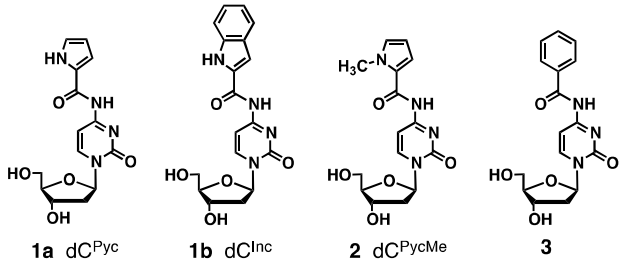
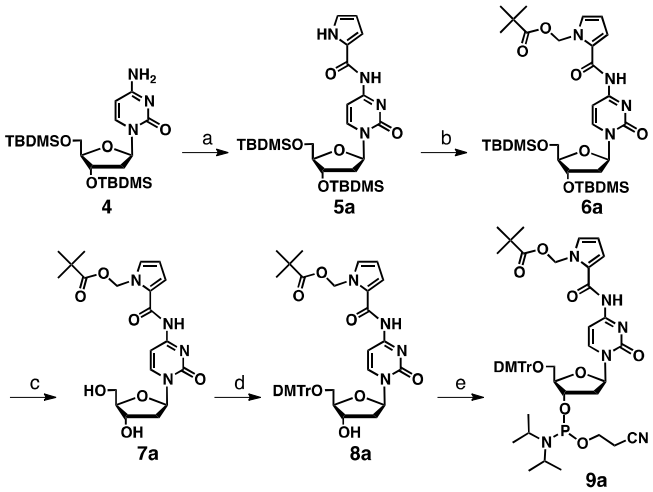


Fig. 2 4-*N*-acylated deoxycytidine derivatives **1–3**.



Scheme 1. Synthesis of phosphoramidite **9a**. *Reagents and conditions*: (a) pyrrole-2-carbonyl chloride (1.2 equiv.), pyridine-CH_3CN (1: 1, v/v), rt, 2 h, 90%; (b) POM-Cl (1.5 equiv.), NaH (3.0 equiv.), THF, 0 °C to rt, 3 h, 81%; (c) TBAF (2.4 equiv.), THF, rt, 2 h, 84%; (d) DMTr-Cl (1.2 equiv.), pyridine, rt, 2 h, 98%; (e) CEO-P(N*i*-Pr_2)$_2$ (1.5 equiv.), 1*H*-tetrazole (0.6 equiv.), *i*-Pr_2NH (0.6 equiv.), CH_3CN, rt, 2 h, 83%.

proved to be extraordinarily stable under basic conditions and capable of strengthened hydrogen bonding with G in a completely planar geometry (Fig. 1).

Results and Discussion

After extensive screening of such possible 4-*N*-acylated cytosine structures, 1*H*-pyrrol-2-ylcarbonyl (Pyc), and 1*H*-indol-2-ylcarbonyl (Inc) were selected, as seen in the corresponding deoxynucleosides **1a** and **1b** of Fig. 2. The *ab initio* calculations of cytosine derivatives acylated with these Pyc or Inc groups indicated that, upon their Watson–Crick base pairing with G, the NH proton of their rotamers can approach to the G's 6-carbonyl oxygen from the Hoogsteen face of the complementary base paired guanine so that, expectedly, an additional hydrogen bond could be formed (Fig. 1 and Fig. S1–2). The hydrogen bond energy estimated by this additional hydrogen bond was −27.7 kcal mol^{-1} that was 2.2 kcal mol^{-1} higher than that of a G–C base pair (−25.5 kcal mol^{-1}) (Fig. S2).

Table 1. Modified oligodeoxynucleotides containing dC^{Pyc} or dC^{Inc}

DNA	Sequence
ODN**0**	5′-d(TTCTTCCCTTCTT)-3′
ODN**1**	5′-d(TTCTTC**C**^{Pyc}CTTCTT)-3′
ODN**2**	5′-d(TT**C**^{Pyc}TTC**C**^{Pyc}CTT**C**^{Pyc}TT)-3′
ODN**3**	5′-d(TTCTT**C**^{Pyc}**C**^{Pyc}**C**^{Pyc}TTCTT)-3′
ODN**4**	5′-d(TT**C**^{Pyc}TT**C**^{Pyc}**C**^{Pyc}**C**^{Pyc}TT**C**^{Pyc}TT)-3′
ODN**5**	5′-d(TTCTTC**C**^{Inc}CTTCTT)-3′
ODN**6**	5′-d(TT**C**^{Inc}TTC**C**^{Inc}CTT**C**^{Inc}TT)-3′
cODN	3′-d(AAGAAG**X**GAAGAA)-5′ (X = G, T, A, or C)

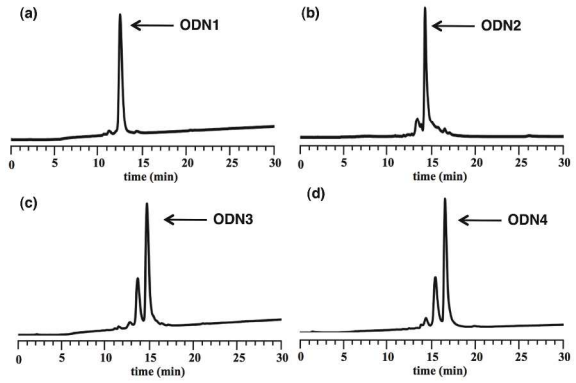


Fig. 3 Reverse-phase HPLC profiles of the crude mixtures of (a) ODN**1**, (b) ODN**2**, (c) ODN**3**, and (d) ODN**4** obtained after 1.5 M NaOMe-MeOH treatment.

Before synthesizing oligonucleotides incorporating dC^{Pyc} and related compounds, we examined the chemical stability of the newly designed acyl groups substituted at the 4-amino position of deoxycytidine. To evaluate the stability of the newly designed acyl group, we measured the completion time (T_{comp}) of deacylation under 1.5M NaOMe/MeOH at room temperature. Surprisingly, compound **1a** ($T_{comp.}$ = 6 h) showed a significantly longer $T_{comp.}$ value than 4-*N*-benzoyldeoxycytidine (**3**) [5] ($T_{comp.}$ = 1 h), whereas the *N*-methylated species (**2**) of **1a** showed a somewhat longer $T_{comp.}$ (8 h) value than compound **1a**. Khorana *et al.* and Köster *et al.* reported that the resonance effect of dissociated *N*-acyl groups affected the stability of the amide bond under basic conditions of NaOH/MeOH.[6] Our results suggested that the significantly longer $T_{comp.}$ of compounds **1a** and **2** than **3** can be explained by the strong electron donating properties of the pyrrolyl moiety of the acyl groups.[7] It was also found that the pivaloyloxymethyl (POM) group, which was introduced as a protecting group into the pyrrolyl NH group of compound **1a**, could be cleaved very

Table 2. T_m values of DNA/DNA duplexes containing dC^{Pyc} or dC^{Inc}

	T_m[a] (ΔT_m[b]) (°C)						
X	ODN**0**	ODN**1**	ODN**2**	ODN**3**	ODN**4**	ODN**5**	ODN**6**
G	55.2	56.5	59.6	59.4	63.1	54.5	51.0
T	36.8 (18.4)	42.0 (14.5)	45.4 (14.2)	44.8 (14.6)	47.0 (16.1)	45.8 (8.7)	40.6 (10.4)
A	35.0 (20.2)	39.2 (17.3)	35.4 (24.2)	41.1 (18.3)	36.2 (26.9)	41.8 (12.7)	34.6 (16.4)
C	30.9 (24.3)	40.8 (15.7)	34.5 (25.1)	42.4 (17.0)	35.7 (27.4)	46.4 (8.1)	41.5 (9.5)

[a] The T_m values are accurate within ±0.5 °C. T_m measurements were carried out in a buffer containing 10 mM sodium phosphate (pH 7.0), 1 M NaCl, 0.1 mM EDTA, and 2 μM duplex.
[b] ΔT_m is the difference in the T_m value between the duplexes of ODNs **0**–**6** with cODN (X = G) and cODN (X = T, A, or C).

quickly ($T_{comp.}$ = 15 min) in 1.5 M NaOMe/MeOH without significant cleavage of the 4-*N*-amide bond. Therefore, we considered that oligonucleotides having dC^{Pyc} could be synthesized by use of this POM protection.[8]

To examine if oligodeoxynucleotides having dC^{Pyc} could be synthesized by the usual phosphoramidite approach, we tried to obtain ODN**1** (Table 1) using the phosphoramidite unit **9a**, which was synthesized, as shown in Scheme 1. In this oligonucleotide synthesis, 5-(*bis*-3,5-trifluoromethylphenyl)-1*H*-tetrazole (Activator 42)[9] gave better results than 5-benzylthio-1*H*-tetrazole[10] as an activator for the coupling of dC^{Pyc} phosphoramidite. After the final chain elongation, the 5′-terminal DMTr group was removed and the ODN was rapidly cleaved from the solid support (CPG) by treatment with 1.5 M NaOMe in MeOH at room temperature for 30 min. Under these conditions, the POM group at the pyrrole residue, the protecting groups at the nucleobase and the internucleotide phosphate triester were simultaneously removed. The solution was neutralized by Dowex 50W X8. After the usual workup, ODN**1** was obtained as a major product, as shown in Fig. 3a. In a similar manner, dC^{Pyc} residues containing oligonucleotides, ODN**2** (discretely 3 points)**,** ODN**3** (consecutively 3 points) and ODN**4** (5 points), were synthesized (Table 1). In these syntheses, minor byproducts lacking one pyrrol-2-ylcarbonyl group were observed (Fig. 3b-3d). These deacylated byproducts dominantly were observed in consecutively modified ODNs (ODN**3** and ODN**4**). As mentioned above, the stabilization of the amide bond toward nucleophilic attack is due to the delocalization of the negative charge on the amide bond.[6] Thus, our results suggested that the negative charges on both of the neighboring dC^{Pyc} residues were generated less likely so that the solvolysis might predominantly occur. This effect should be taken into account for the synthesis of multi-modified longer ODNs.

We also synthesized two oligonucleotides having 4-*N*-(indol-2-ylcarbonyl)deoxycytidines, *i.e.*, ODN**5** and ODN**6** (For the details, see Scheme S1 and experimental section), in a manner similar to that described in the synthesis of oligonucleotides having dC^{Pyc}. However, this Inc group was found to be less stable than the Pyc group (data not shown) and the removal of the POM group on the nitrogen required longer (1 h) than that of the Pyc group when 1.5 M NaOMe/MeOH was used. The T_m values of the duplexes formed between ODNs **1–6** and the complementary strand 5′-d(AAGAAGXGAAGAA)-3′ (cODN: X = G, T, A, or C) are summarized in Table 2.

The duplex resulting from ODN**1** with a single dC^{Pyc} was found to show slightly enhanced thermostability (+1.3 °C) than the unmodified duplex. In ODNs **2** and **3** having three dC^{Pyc}s arranged in discontinuous and consecutive manners, their

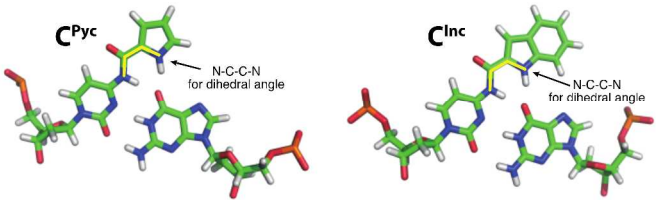


Fig. 4 C^{Pyc} and C^{Inc} base pairing during MD simulations.

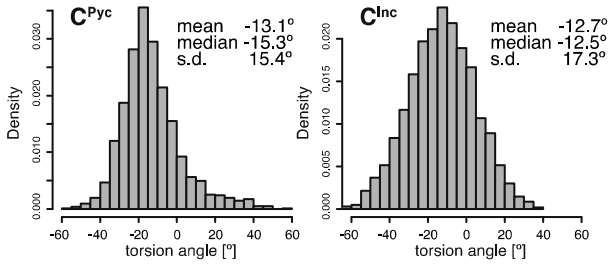


Fig. 5 Histograms of torsion angle (N-C-C-N) during MD simulations. The mean, median, and s.d. denote mean torsion angle, median torsion angle, and standard deviation, respectively.

hybridization affinities were higher by +4.4 °C and +4.2 °C, respectively, than that of unmodified ODN**0**, as shown in Table 2. The hybridization affinity of ODN**4** having five dC^{Pyc}s showed a further increase of the T_m value, which was higher by +7.9 °C than that of ODN**0**. The ODNs**5** and **6** containing one and three discontinuous dC^{Inc}s showed lower binding affinity for the matched-sequence template by −0.7 °C and −4.2 °C, respectively, than that of ODN**0**. As far as the base recognition is concerned, ODNs**1**–**4** showed sufficient ΔT_m values of more than 14 °C that are the difference in the T_m value between matched and the most stable mismatched duplexes. In sharp contrast, ODNs**5** and **6** showed a significant decrease of base recognition ability. For further evaluations, dC^{Pyc} was incorporated into the Dickerson–Drew self complementary mixed-sequence of 5′-d(CGCGAATTC^{Pyc}GCG) (ODN**7**). In this sequence, the thermostability (T_m = 58.1 °C) of the oligonucleotide (ODN**7**) having dC^{Pyc} was also higher by +7.3 °C than that of unmodified ODN (T_m = 50.8 °C).

For duplexes of dC^{Pyc}-modified ODNs**1**–**4** and complementary RNA strands 3′-r(AAGAAGYGAAGAA)-5′ (Y = G, A, U, or C), the steady increase of the T_m values (Y = G) was also observed except for ODN**1** (Table 3).

To understand the results of the T_m experiments, we performed molecular dynamic (MD) simulations using

Table 3. T_m values of DNA/RNA duplexes containing dC^{Pyc} or dC^{Inc}

	$T_m{}^a$ ($\Delta T_m{}^b$) (°C)						
X	ODN**0**	ODN**1**	ODN**2**	ODN**3**	ODN**4**	ODN**5**	ODN**6**
G	67.1	64.3	69.1	69.7	72.7	61.9	59.7
U	45.6 (21.5)	46.0 (18.3)	48.8 (21.3)	48.0 (21.7)	50.4 (22.3)	48.4 (13.5)	42.6 (17.1)
A	46.0 (21.1)	46.0 (18.3)	50.0 (19.1)	49.8 (19.9)	53.8 (18.9)	48.0 (13.9)	42.4 (17.3)
C	42.4 (24.7)	46.8 (17.5)	48.6 (20.5)	49.0 (20.7)	51.6 (21.1)	47.8 (14.1)	42.4 (17.3)

[a]The T_m values are accurate within ±0.5 °C. T_m measurements were carried out in a buffer containing 10 mM sodium phosphate (pH 7.0), 1 M NaCl, 0.1 mM EDTA, and 2 μM duplex.
[b]ΔT_m is the difference in the T_m value between the duplexes of ODNs **0–6** with 3′-r(AAGAAG**G**GAAGAA)-5′ and 3′-r(AAGAAG**X**GAAGAA)-5′ (X = U, A, or C).

AMBER 12.0 and analyzed the torsion angles between the pyrrole or indole ring and the *N*-carbonyl moiety. As described in Fig. 1, the dynamic behavior of the torsion angles was expected to reflect the stability of the additional hydrogen bond. For the MD simulation, the energy profile around the carbonyl group and the pyrolyl or indolyl residue of the 1-methylated species of C^{Pyc}, C^{Inc}, and (1-*N*-methyl-pyrrol-2-yl)carbonylcytosine (C^{PycMe}, the base part of compound **2** in Fig. 2) were obtained by using the quantum calculation at the level of B3LYP/6-31+G(d). To perform the MD simulations, we fitted the force field parameters of the torsion angle of each compound to the corresponding energy profile (Fig. S7). The charges for the noncanonical residues were determined by RESP charge fitting.[11] The sequences for MD simulations are same as ODN1 (C^{Pyc}) and ODN5 (C^{Inc}). For C^{PycMe}, C^{Pyc} residue in ODN1 was substituted to C^{PycMe} residue (Table1). The duplexes of these ODNs with cODN (X = G) were simulated. The minimization and equilibration processes were run according to the protocol used successfully in the previous studies.[12] The torsion angles around the heteroaromatic ring and the carbonyl group of the C^{Pyc} and C^{Inc} residues were rather kept with the mean of *ca.* −13° during the unrestrained MD simulations (Fig. 5). However, the *N*-methylpyrrol ring of the C^{PycMe} residue, which cannot form an additional hydrogen bond, was completely rotated to the reverse side (Fig. S9). These results suggested the formation of the additional hydrogen bond in the case of C^{Pyc} and C^{Inc}. Interestingly, the distribution pattern of the C^{Pyc} torsion angles was different from those of C^{Inc} (Fig. 5). The latter showed a wider distribution in the torsion angle axis than the former. It seemed that the wider fluctuation results from a weakness of the additional hydrogen bond. In addition, the hydrophobic benzene ring of the indole moiety is oriented toward the major groove without any hydrophobic interaction partners at the upstream and downstream sites (Fig. S10). Since the phenyl ring of the indole moiety extrudes into the outer hydrophilic space surrounding the water network, this type of hydrophobic moiety could be expected to be entropically unfavorable in such a rigid duplex state.[13] The stabilization effect by the additional hydrogen bond with the C^{Inc} residue might be weakened by the hydrophobic indole moiety.

Conclusions

In this study, we designed a Pyc group capable of not only forming a planar structure but also forming additional hydrogen bonds with the 6-carbonyl oxygen. This choice allowed the planarity of the resulting base pair with G since the 5-membered pyrrole ring gave less steric hindrance compared the previously reported aroyl groups.[2b] Actually, oligonucleotides modified with dC^{Pyc} showed a significant increase of hybridization affinity for the complementary strand compared with the oligonucleotides modified with an aroyl group.[2b] It should be noted that oligonucleotides incorporating dC^{Pyc} bases were easily synthesized by the conventional phosphoramidite approach and the ultrafast deprotection/cleavage procedure using 1.5 M NaOMe/MeOH instead of the canonical *conc.* ammonia. This synthetic advantage is essentially arising from the electronically inherent property of the pyrrole group that showed its strong electron-donating effect on the carbonyl group. Probably, the Pyc group would be used as a new type of scaffold to introduce a variety of functional moieties into oligonucleotides *via* ester and amide bonds since further modification on the Pyc group is possible.

Acknowledgements

This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and in part by Health Sciences Research Grants for Research on Psychiatric and Neurological Diseases and Mental Health from the Ministry of Health, Labor and Welfare of Japan.

Experimental

General remarks

^1H, ^{13}C, and ^{31}P NMR spectra were recorded at 500, 120, and 203 MHz, respectively. The chemical shifts were measured from tetramethylsilane (0 ppm), $CDCl_3$ (7.26 ppm) or DMSO-d_6 (2.49 ppm) for ^1H NMR spectra, $CDCl_3$ (77.0 ppm) or DMSO-d_6 (39.7 ppm) for ^{13}C NMR spectra, and 85% H_3PO_4 as an external standard for ^{31}P NMR spectra. Column chromatography was performed with silica gel C-200 purchased from Wako Co. Ltd, and a mini-pump for a goldfish bowl was conveniently used to attain sufficient pressure for rapid chromatographic separation. The synthesis of modified oligonucleotides was performed using a DNA/RNA synthesizer 392 (Applied Biosystem). The unmodified oligonucleotides were purchased from Sigma-Aldrich Japan. HPLC was performed using the following systems: reversed-exchange HPLC was done on Waters Alliance systems (2690 and 996) with a Waters 3D UV detector and a Waters XTerra MS C18 column (4.6 × 150 mm). A linear gradient (0%–30%) of solvent I (0.1 M ammonium acetate buffer (pH 7.0)) in solvent II (CH_3CN) was used at 50 °C at a flow rate of 1.0 mL min^{-1} for 30 min. Anion-exchange HPLC was performed on a Waters Alliance system with a Waters 3D UV detector HPLC for analysis of the purity of the products (on a Shimadzu SCL-10A system with Shimadzu CTO-10A and SPD-M10A for isolation of the pure product) and a Gen-Pak FAX column (Waters, 4.6 × 100 mm). A linear gradient (0%–60%) starting from 25 mM sodium phosphate buffer (pH 7.0) and applying 25 mM sodium phosphate buffer (pH 7.0) containing 1 M NaCl was used at a flow rate of 1 mL min^{-1} for 30 min at 50 °C. High resolution ESI mass spectrometry was performed by use of a Mariner^{TM} (PerSeptive Biosystems Inc.). MALDI-TOF mass was performed by use of Brüker Daltonics. The alphabets (a, b, and c) after compound number, *i.e.,* **a** for compound **5a**, denote the C^{Pyc}, C^{Inc} and C^{PyMe} derivatives, respectively.

Synthesis of 4-*N*-(1*H*-Pyrrol-2-ylcarbonyl)deoxycytidine (1)

Compound **5a** (2.46 g, 4.4 mmol) was rendered anhydrous by repeated co-evaporation with CH_3CN and finally dissolved in anhydrous THF (46 mL). Et_3N-3HF (4.0 mL, 22.0 mmol) was added to the solution and the resulting mixture was stirred at room temperature for 12 h. The mixture was partitioned between $CHCl_3$ and H_2O. The organic layer was collected, dried over Na_2SO_4, filtered, and evaporated under reduced pressure. The crude product was purified by C-200 silica gel column chromatography with $CHCl_3$–MeOH to yield **1** (690 mg, 48%). ^1H NMR (DMSO-d_6) δ: 2.01–2.06 (1H, m), 2.26–2.31 (1H, m), 3.55–3.64 (1H, m), 3.84–3.86 (1H, m), 4.22–4.24

(1H, m), 5.05 (1H, t, J = 5.1 Hz), 5.26 (1H, d, J = 4.2 Hz), 6.13 (1H, t, J = 6.3 Hz), 6.16–6.17 (1H, m), 7.05 (1H, s), 7.34 (1H, s), 7.37 (1H, d, J = 7.3 Hz), 8.32 (1H, d, J = 7.6 Hz), 10.80 (1H, br), and 11.90 (1H, br); ^{13}C NMR (DMSO-d_6) δ: 41.4, 61.5, 70.5, 86.5, 88.4, 96.4, 110.2, 115.1, 125.3, 125.4, 145.0, 154.9, 160.4, 163.4; HRMS (ESI): m/z calcd for $C_{14}H_{17}N_4O_5^+$ [M+H]: 321.1199, found 321.1193.

Synthesis of 4-*N*-(1-methyl-1*H*-pyrrol-2-yl-carbonyl) deoxycytidine (2)

Compound **5c** (2.45 g, 4.4 mmol) was rendered anhydrous by repeated coevaporation with CH_3CN and finally dissolved in anhydrous THF (120 mL). TBAF (3.14 g, 12.0 mmol) was added to the solution and the resulting mixture was stirred at room temperature for 4 h. The mixture was partitioned between $CHCl_3$ and H_2O. The organic layer was collected, dried over Na_2SO_4, filtered, and evaporated under reduced pressure. The crude product was purified by C-200 silica gel column chromatography with $CHCl_3$–MeOH to yield **2** (1.84 g, 75 %). ^{1}H NMR (DMSO-d_6) δ: 2.00–2.06 (1H, m), 2.26–2.30 (1H, m), 3.56–3.64 (2H, m), 3.96 (4H, s), 4.23 (1H, br), 5.06 (1H, t, J = 4.7 Hz), 5.27 (1H, d, J = 3.9 Hz), 6.08–6.09 (1H, m), 6.13 (1H, t, J = 6.3 Hz), 7.12 (1H, s), 7.30 (1H, d, J = 7.1 Hz), 7.34 (1H, br), 8.30 (1H, d, J = 7.3 Hz), and 10.7 (1H, br); ^{13}C NMR (DMSO-d_6) δ: 37.8, 41.8, 61.9, 70.9, 87.0, 88.8, 96.8, 108.5, 117.9, 124.8, 132.2, 145.3, 155.5, 161.4, and 164.0; HRMS (ESI): calcd for $C_{15}H_{19}N_4O_5^+$ m/z [M+H] $^+$: 335.1355, found 335.1318.

Synthesis of 3′,5′-*O*-bis-(*tert*-butyldimethylsilyl)-4-*N*-(1*H*-pyrrol-2-ylcarbonyl)deoxycytidine (5a)

A solution of 1*H*-pyrrole-2-carbonyl chloride (712 mg, 5.5 mmol) in dry CH_2Cl_2 (3.4 mL) was added to a solution of 3′,5′-*O*-bis(*tert*-butyldimethylsilyl)deoxycytidine (**4**) (2.28 g, 5.0 mmol) in dry pyridine (3.4 mL). The mixture was stirred at room temperature for 2 h and the mixture was diluted with $CHCl_3$. The $CHCl_3$ solution was washed successively with H_2O and saturated $NaHCO_3$. The organic layer was collected, dried over Na_2SO_4, filtered, and evaporated under reduced pressure. The crude product was purified by C-200 silica gel column chromatography with hexane–chloroform to yield **5a** (2.46 g, 90 %). ^{1}H NMR (DMSO-d_6) δ: 0.06–0.08 (12H, m), 0.85–0.88 (18H, m), 2.14–2.19 (1H, m), 2.28–2.33 (1H, m), 3.72–3.75 (1H, m), 3.82–3.85 (1H, m), 3.86–3.88 (1H, m), 4.36–4.39 (1H, dd, J = 9.9 Hz, 5.0 Hz), 6.10–6.13 (1H, t, J = 5.9 Hz), 6.15–6.16 (1H, m), 7.05 (1H, br), 7.35 (1H, br), 7.41–7.42 (1H, d, J = 7.5 Hz), 8.21–8.22 (1H, d, J = 7.5 Hz), 10.8 (1H, s), and 11.9 (1H, s); ^{13}C NMR (DMSO-d_6) δ: −5.2, −5.1, −4.6, −4.3, 18.1, 18.4, 26.1, 26.2, 41.4, 62.3, 71.1, 86.2, 87.5, 96.2, 110.2, 115.1, 125.2, 125.4, 144.4, 154.8, and 160.3, 163.5; HRMS (ESI): m/z calcd for $C_{26}H_{44}N_4NaO_5Si_2^+$ [M+Na]$^+$: 571.2748, found 571.2751.

Synthesis of 3′,5′-*O*-bis-(*tert*-butyldimethylsilyl)-4-*N*-(1*H*-indol-2-ylcarbonyl)deoxycytidine (5b)

A solution of 1*H*-indole-2-carbonyl chloride (2.2 g, 12 mmol) in dry CH_2Cl_2 (20 mL) was added to a solution of 3′,5′-*o*-bis(*tert*-butyldimethylsilyl)deoxycytidine (**4**) (4.56 g, 10 mmol) in dry pyridine (20 mL). The mixture was stirred at room temperature for 1 h and the mixture was diluted with $CHCl_3$. The $CHCl_3$ solution was washed successively with H_2O and saturated $NaHCO_3$. The organic layer was collected, dried over Na_2SO_4, filtered, and evaporated under reduced pressure. The crude product was purified by C-200 silica gel column chromatography with hexane–chloroform to yield **5b** (5.4 g, 90%). ^{1}H NMR (DMSO-d_6) δ: 0.08–0.10 (12H, m), 0.87–0.94 (18H, m), 2.17–2.22 (1H, m), 2.31–2.36 (1H, m), 3.74–3.77 (1H, m), 3.83–3.86 (1H, m), 3.89–3.90 (1H, m), 4.38–4.41 (1H, dd, J = 4.9 Hz), 6.11–6.14 (1H, t, J = 5.9 Hz), 7.05–7.08 (1H, t, J = 7.4 Hz), 7.23–7.26 (1H, t, J = 7.6 Hz), 7.44–7.45 (1H, d, J = 4.8 Hz), 7.64–7.66 (1H, d, J = 8.1 Hz), 11.30 (1H, br), and 11.90 (1H, br); ^{13}C NMR ($CDCl_3$) δ: −4.65, −4.61, −4.1, −3.8, 18.6, 18.9, 26.6, 26.7, 41.1, 62.8, 71.6, 86.9, 88.1, 96.8, 107.9, 113.4, 121.6, 123.3, 125.8, 127.8, 130.8, 138.4, 145.3, 155.3, 162.0, 163.8; HRMS (ESI): m/z calcd for $C_{30}H_{47}N_4O_5Si_2^+$ [M+H]$^+$: 599.3085, found 599.3033.

Synthesis of 3′,5′-*O*-bis-(*tert*-butyldimethylsilyl)-4-*N*-(1-methyl-1*H*-pyrrol-2-ylcarbonyl)deoxycytidine (5c)

A solution of 1-methy-1*H*-pyrrole-2-carbonyl chloride (790 mg, 5.5 mmol) in dry CH_2Cl_2 (10 mL) was added to a solution of 3′,5′-*O*-bis(*tert*-butyldimethylsilyl)deoxycytidine (**4**) (2.28 g, 5.0 mmol) in dry pyridine (10 mL). The mixture was stirred at room temperature for 2 h and the mixture was diluted with $CHCl_3$. The $CHCl_3$ solution was washed successively with H_2O and saturated $NaHCO_3$. The organic layer was collected, dried over Na_2SO_4, filtered, and evaporated under reduced pressure. The crude product was purified by C-200 silica gel column chromatography with hexane–chloroform to yield **5c** (2.45 g, 87 %). ^{1}H NMR ($CDCl_3$) δ: 0.03–0.04 (6H, m), 0.10–0.11 (6H, m), 0.86 (9H, s), 0.92 (9H, s), 2.09–2.14 (1H, m), 2.48–2.53 (1H, m), 3.75–3.78 (1H, m), 3.92–3.93 (1H, d, J = 3.2 Hz), 3.95 (4H, s), 4.36–4.39 (1H, dd, J = 10.5 Hz, J = 5.3 Hz), 6.14 (1H, t, J = 3.3 Hz), 6.26 (1H, t, J = 5.7 Hz), 6.77 (1H, br), 6.83 (1H, br), 7.39 (1H, br), 8.32 (1H, d, J = 7.1 Hz), and 8.37 (1H, br); ^{13}C NMR ($CDCl_3$) δ: −5.6, −5.5, −5.0, −4.7, 17.9, 18.3, 25.6, 25.9, 37.1, 42.3, 61.8, 70.1, 86.7, 87.7, 95.6, 108.3, 114.9, 124.1, 130.8, 144.2, 155.1, 159.4, and 162.0; HRMS (ESI): m/z calcd for $C_{27}H_{47}N_4O_5Si_2^+$ [M+H]$^+$: 563.3085, found 563.3094.

Synthesis of 3′,5′-*O*-bis-(*tert*-butyldimethylsilyl)-4-*N*-(*1*-pivaloyloxymethyl-1*H*-pyrrol-2-ylcarbonyl)deoxycytidine (6a)

Compound **5a** (274 mg, 0.5 mmol) was rendered anhydrous by repeated co-evaporation with pyridine, toluene, and CH_2Cl_2 and finally dissolved in anhydrous THF (1.5 mL). NaH (60 mg, 1.5 mmol, containing 40% oil) was added to the solution, and the resulting mixture was stirred at 0 °C for 1 h. The mixture was added to pivaloyloxymethyl chloride (0.11 μL, 0.75 mmol) and stirred at room temperature for 2 h. The mixture was diluted with ethyl acetate. The ethyl acetate solution was washed successively with H_2O and saturated $NaHCO_3$. The organic layer was collected, dried over Na_2SO_4, filtered, and evaporated under reduced pressure. The crude product was purified by C-200 silica gel column chromatography with hexane–chloroform to yield **6a** (270 mg, 81 %). ^{1}H NMR (DMSO-d_6) δ: 0.07 (12H, m), 0.86–0.87 (18H, m), 1.07 (9H, s), 2.14–2.19 (1H, m), 2.29–2.34 (1H, m), 3.77–3.75 (1H, dd, J = 11.5 Hz), 3.82–3.85 (1H, dd, J = 11.5 Hz), 3.88–3.89 (1H, m), 4.36–4.39 (1H, q, J = 4.9 Hz), 6.09–6.11 (1H, t, J = 6.0 Hz), 6.16–6.22 (3H, m), 7.27–7.29 (1H, d, J = 7.1 Hz), 7.47 (1H, s), 8.19–8.20 (1H, d, J = 7.1 Hz), and 10.86 (1H, br); ^{13}C NMR (DMSO-d_6) δ: −4.8, −4.7, −4.2, −3.9, 18.5, 18.8, 26.5, 26.6, 27.5, 39.1, 41.8, 62.8, 71.6, 72.2, 86.7, 88.0, 96.7, 109.6, 119.5, 124.6, 132.2, 144.8, 155.2, 160.8, 163.8, and 177.4. HRMS (ESI): m/z calcd for $C_{32}H_{55}N_4O_7Si_2^+$ [M+H]$^+$: 663.3609, found 663.3615.

Synthesis of 3′,5′-*O*-bis-(*tert*-butyldimethylsilyl)-4-*N*-(1-pivaloyloxymethyl-1*H*-indol-2-ylcarbonyl)deoxycytidine (6b)

Sodium hydride (900 mg, 15 mmol, containing 40% oil) was added to a solution of **5b** (3.0 g, 5 mmol) in dry THF (15 mL) and stirred for 2 h at room temperature. Pivaloyloxymethyl chloride (1.1 mL, 7.5 mmol) was added and stirred at room temperature for 1 h. The mixture was quenched by H_2O and then evaporated. The crude solution was extracted with ethyl acetate and H_2O and then washed successively with H_2O. The organic layer was collected, dried over Na_2SO_4, filtered and evaporated under reduced pressure. The crude product was purified by C-200 silica gel column chromatography with hexane–chloroform to yield **6b** (3.6 g, 99%). ^{1}H NMR (DMSO-d_6) δ: 0.07–0.08 (12H, m), 0.86 (9H, s), 0.88 (9H, s), 1.03 (9H, s), 2.16–2.21 (1H, m), 2.31–2.36 (1H, m), 3.73–3.76 (1H, m), 3.83–3.86 (1H, m), 3.89–3.90 (1H, m), 4.37–4.40 (1H, dd, J = 5.0 Hz, 10.0 Hz), 6.11–6.13 (1H, t, J = 5.9 Hz), 6.50–6.56 (2H, m), 7.18–7.21 (1H, t, J = 7.4 Hz), 7.31 (1H, br), 7.38–7.41 (1H, t, J = 7.6 Hz), 7.68–7.71 (3H, m), 8.24–8.25 (1H, d, J = 5.6 Hz), and 11.30 (1H, br); ^{13}C NMR ($CDCl_3$) δ: −5.7, −5.6, −5.1, −4.8, 17.7, 18.1, 25.5, 25.7, 26.7, 38.7, 42.0, 61.6, 67.1, 69.8, 86.5, 87.5, 96.5, 110.3, 110.5, 121.8, 122.5, 125.9, 126.0, 130.2, 139.3, 144.0, 154.1, 161.6, and 177.7; HRMS (ESI): *m*/*z* calcd for $C_{36}H_{57}N_4O_7Si_2^+$ $[M+H]^+$: 713.3766, found 713.3797.

Synthesis of 4-*N*-(1-pivaloyloxymethyl-1*H*-pyrrol-2-ylcarbonyl)deoxycytidine (7a)

Compound **6a** (3.26 g, 5.0 mmol) was rendered anhydrous by repeated co-evaporation with pyridine, toluene, and CH_2Cl_2 and finally dissolved in anhydrous THF (46 mL). TEA·3HF (4.0 mL, 25 mmol) was added to the solution, and the resulting mixture was stirred at room temperature for 1 h. The mixture was partitioned between $CHCl_3$ and H_2O. The organic layer was collected, dried over Na_2SO_4, filtered, and evaporated under reduced pressure. The crude product was purified by C-200 silica gel column chromatography with $CHCl_3$–MeOH to yield **7a** (1.9 g, 94%). ^{1}H NMR (DMSO-d_6) δ: 1.06 (9H, s), 2.02–2.07 (1H, m), 2.27–2.32 (1H, m), 3.55–3.60 (1H, m), 3.61–3.65 (1H, m), 3.86–3.87 (1H, m), 4.22–4.24 (1H, m), 5.02–5.04 (1H, t, J = 5.0 Hz), 5.25–5.26 (1H, d, J = 4.5 Hz), 6.12–6.14 (1H, t, J = 6.3 Hz), 6.18–6.20 (3H, m), 7.25 (1H, br), 7.30 (1H, s), 7.44 (1H, br), 8.31–8.32 (1H, d, J = 7.0 Hz), and 10.81 (1H, br); ^{13}C NMR ($CDCl_3$) δ: 27.7, 39.2, 41.9, 62.0, 71.0, 72.1, 87.1, 88.9, 96.9, 109.7, 119.6, 124.8, 132.3, 145.6, 155.4, 160.9, 163.8, and 177.6; HRMS (ESI): *m*/*z* calcd for $C_{20}H_{27}N_4O_7^+$ $[M+H]^+$: 435.1880, found 435.1820.

Synthesis of 4-*N*-(1-pivaloyloxymethyl-1*H*-indol-2-ylcarbonyl)deoxycytidine (7b)

Compound **6b** (2.85 g, 4.0 mmol) was rendered anhydrous by repeated co-evaporation with CH_3CN and finally dissolved in anhydrous THF (96 mL). TBAF (2.51 g, 9.6 mmol) was added to the solution and the resulting mixture was stirred at room temperature for 2 h. The mixture was partitioned between $CHCl_3$ and H_2O. The organic layer was collected, dried over Na_2SO_4, filtered, and evaporated under reduced pressure. The crude product was purified by C-200 silica gel column chromatography with $CHCl_3$–MeOH to yield **7b** (1.15 g, 59 %). ^{1}H NMR (DMSO-d_6) δ: 1.01–1.03 (9H, m), 2.03–2.08 (1H, m), 2.29–2.32 (1H, m), 3.55–3.64 (2H, m), 3.86–3.87 (1H, m), 4.23 (1H, br), 5.06 (1H, t, J = 5.3 Hz), 5.27 (1H, d, J = 4.0 Hz), 6.13 (1H, t, J = 6.0 Hz), 6.54 (2H, s), 7.20 (1H, t, J = 7.5 Hz), 7.30 (1H, br), 7.40 (1H, t, J = 7.5 Hz), 7.68–7.72 (3H, m), 8.38 (1H, br), and 11.3 (1H, s); ^{13}C NMR ($CDCl_3$) δ: −27.6, 39.3, 41.9, 61.9, 68.7, 70.9, 87.2, 88.9, 96.9, 112.0, 122.5, 122.6, 123.6, 126.7, 126.8, 140.0, and 177.8; HRMS (ESI): *m*/*z* calcd for $C_{24}H_{29}N_4O_7^+$ $[M+H]^+$: 485.2036, found 485.2034.

Synthesis of 5′-*O*-DMTr-4-*N*-(1-pivaloyloxymethyl-1*H*-pyrrol-2-ylcarbonyl)deoxycytidine (8a)

Compound **7a** (434 mg, 1.0 mmol) was rendered anhydrous by repeated co-evaporation with pyridine, toluene, and CH_2Cl_2 and finally dissolved in anhydrous pyridine (5.0 mL). To the solution was added 4,4′-dimethoxytrityl chloride (407 mg, 1.2 mmol). After being stirred at room temperature for 2 h, the mixture was quenched by addition of water and evaporated *in vacuo*. The residue was dissolved with $CHCl_3$. The solution was washed with brine and aqueous $NaHCO_3$. The organic layer was dried over Na_2SO_4 and concentrated *in vacuo*. The residue was chromatographed on a column of silica gel with $CHCl_3$–MeOH (95: 5, v/v) containing 0.5% triethylamine to yield **8a** (721 mg, 98%). ^{1}H NMR (DMSO-d_6) δ: 1.05 (9H, s), 2.11–2.16 (1H, m), 2.31–2.36 (1H, m), 3.73 (6H, s), 3.95–3.97 (1H, m), 4.25–4.29 (1H, m), 5.24–5.35 (1H, d, J = 4.9 Hz), 6.11–6.13 (1H, t, J = 5.9 Hz), 6.18–6.22 (3H, m), 6.88–6.90 (4H, m), 7.22–7.26 (5H, m), 7.29–7.32 (3H, m), 7.36–7.38 (2H, m), 7.46 (1H, br), 8.10–8.12 (1H, d, J = 6.6 Hz), and 10.9 (1H, br); ^{13}C NMR ($CDCl_3$) δ: 26.8, 38.8, 42.0, 55.2, 62.7, 69.9, 70.9, 76.8, 77.0, 77.3, 86.4, 86.8, 87.1, 96.2, 109.5, 113.2, 116.7, 123.9, 125.2, 127.0, 128.0, 128.1, 128.2, 129.0, 129.9, 130.0, 131.0, 135.3, 135.4, 144.2, 144.3, 155.3, 158.6, 158.7, 162.1, and 177.9; HRMS (ESI): *m*/*z* calcd for $C_{41}H_{44}N_4NaO_9^+$ $[M+Na]^+$: 759.3006, found 759.3090.

Synthesis of 5′-*O*-DMTr-4-*N*-(1-pivaloyloxymethyl-1*H*-indol-2-ylcarbonyl)deoxycytidine (8b)

Compound **7b** (969 mg, 2.0 mmol) was rendered anhydrous by repeated co-evaporation with pyridine and finally dissolved in anhydrous pyridine (10 mL). Dimethoxytrityl chloride (0.81 mg, 2.4 mmol) was added to the solution and the resulting mixture was stirred at room temperature for 2 h. The mixture was quenched by saturated $NaHCO_3$ and partitioned between $CHCl_3$ and H_2O. The organic layer was collected, dried over Na_2SO_4, filtered and evaporated under reduced pressure. The crude product was purified by C-200 silica gel column chromatography with $CHCl_3$–MeOH to yield **8b** (1.65 g, 99%). ^{1}H NMR (DMSO-d_6) δ: 1.03 (9H, s), 2.03–2.08 (1H, m), 2.29–2.32 (1H, m), 3.56–3.64 (2H, m), 3.87 (1H, s), 4.23 (1H, s), 5.05–5.07 (1H, t, J = 5.3 Hz), 5.27–5.28 (1H, d, J = 4.0 Hz), 6.12–6.14 (1H, t, J = 6.0 Hz), 6.54 (2H, s), 7.19–7.22 (1H, t, J = 7.5 Hz), 7.30 (1H, br), 7.38–7.41 (1H, t, J = 7.5 Hz), 7.68–7.72 (3H, m), 8.38 (1H, br), and 11.30 (1H, s); ^{13}C NMR (DMSO-d_6) δ: −9.5, 27.5, 39.2, 41.0, 46.5, 56.0, 58.8, 63.9, 68.7, 70.3, 86.7, 86.9, 87.0, 112.0, 114.2, 122.6, 123.6, 126.3, 126.7, 126.8, 127.7, 128.7, 128.9, 129.1, 129.9, 130.7, 136.2, 136.3, 140.0, 145.6, 159.1, and 177.8; HRMS (ESI): *m*/*z* calcd for $C_{45}H_{47}N_4O_9^+$ $[M+H]^+$: 787.3343, found 787.3304.

Synthesis of 5′-*O*-DMTr-4-*N*-(1-pivaloyloxymethyl-1*H*-pyrrol-2-ylcarbonyl)deoxycytidine-3′-(2-cyanoethyl *N*,*N*-diisopropylphosphoramidite) (9a)

Compound **8a** (338 mg, 0.5 mmol) was rendered anhydrous by repeated co-evaporation with pyridine, toluene, and CH_2Cl_2 and finally dissolved in anhydrous CH_2Cl_2 (5.0 mL) under an argon atmosphere. To the solution, diisopropylamine (42.2 μL, 0.6 mmol), 1*H*-tetrazole (21 mg, 0.6 mmol), and a solution of 2-cyanoethoxy-bis(*N*, *N*-diisopropylamino)phosphine (193 μL, 0.6 mmol) were added. After being stirred at room temperature for 2 h, the mixture was diluted with $CHCl_3$. The solution was

washed with brine and aqueous $NaHCO_3$. The organic layer was dried over Na_2SO_4 and filtered. The solution was evaporated *in vacuo*. The residue was chromatographed on a column of silica gel with $CHCl_3$–MeOH (98: 2, v/v) containing 0.5% triethylamine to give compound **9a** (389 mg, 83%). ^{1}H NMR ($CDCl_3$) δ: 1.07–1.20 (21H, m), 2.29–2.30 (1H, m), 2.43–2.46 (1H, t, *J* = 6.5 Hz), 2.61–2.64 (1H, t, *J* = 6.5 Hz), 2.71–2.80 (1H, m), 3.36–3.42 (1H, m), 3.48–3.62 (5H, m), 3.78–3.81 (6H, m), 4.21–4.22 (1H, m), 4.58–4.66 (1H, m), 6.22–6.23 (1H, t, *J* = 3.4 Hz), 6.27–6.30 (3H, m), 6.84–6.88 (4H, m), 7.22 (1H, br), 7.24–7.34 (9H, m), 7.37–7.43 (2H, m), 8.20–8.30 (1H, br), and 8.40 (1H, br); ^{13}C NMR ($CDCl_3$) δ: 20.1, 20.2, 20.2, 20.3, 24.5, 26.8, 26.8, 38.8, 40.8, 41.2, 43.1, 43.1, 43.2, 43.3, 55.1, 55.2, 58.1, 58.3, 61.7, 62.4, 69.9, 71.6, 72.7, 85.6, 86.8, 95.9, 109.4, 109.5, 113.1, 113.2, 116.4, 117.3, 117.5, 124.0, 125.2, 127.1, 127.9, 128.1, 128.2, 129.0, 130.0, 130.1, 130.9, 131.0, 135.1, 135.2, 135.3, 135.3, 137.8, 144.1, 158.6, 161.9, and 177.8; ^{31}P NMR ($CDCl_3$) δ: 149.8, and 150.4; HRMS (ESI): *m*/*z* calcd for $C_{50}H_{62}N_6O_{10}P^+$ $[M+H]^+$: 937.4265, found 937.4220.

Synthesis of 4-*N*-(1-pivaloyloxymethyl-1*H*-indole-2-ylcarbonyl)-5′-*O*-(4,4′-dimethoxytrityl)deoxycytidine-3′-(2-cyanoethyl *N,N*-diisopropylphosphoramidite) (9b)

Compound **8b** (787 mg, 1.0 mmol) was rendered anhydrous by repeated co-evaporation with pyridine, toluene, and CH_2Cl_2 and finally dissolved in dry CH_2Cl_2 (10 mL) under an argon atmosphere. To the solution, 1*H*-tetrazole (42 mg, 0.6 mmol) ethyldiisopropylamine (84.4 μL, 0.6 mmol) and a solution of 2-(cyanoethoxy)bis(*N*,*N*′-diisopropylamino)phosphine (386 μL, 1.2 mmol) were added. After being stirred at room temperature for 2 h, the mixture was diluted with $CHCl_3$. The solution was washed with brine and aqueous $NaHCO_3$. The solution was evaporated *in vacuo*. The mixture was diluted by Et_2O, washed with 0.5 M NaOH aqueous solution and the organic layer was dried over Na_2SO_4 and filtered. The solution was evaporated *in vacuo*. The residue was chromatographed on a column of silica gel with hexane–$CHCl_3$ containing 0.5% triethylamine to yield **9b** (795 mg, 80%). ^{1}H NMR ($CDCl_3$) δ: 1.08–1.11 (12H, m), 1.17–1.19 (9H, m), 2.28–2.36 (1H, m), 2.45 (1H, t, *J* = 6.5 Hz), 2.62 (1H, t, d, *J* = 6.3 Hz), 2.70–2.80 (1H, m), 3.38–3.44 (1H, m), 3.50–3.68 (5H, m), 3.72–3.87 (6H, m), 4.23 (1H, br), 4.62–4.68 (1H, m), 6.28–6.32 (1H, m), 6.59 (2H, br), 6.85–6.88 (4H, m), 7.13–7.17 (1H, m), 7.22–7.28 (4H, m), 7.30–7.34 (5H, m), 7.40–7.43 (3H, m), 7.55 (1H, d, *J* = 8.0 Hz), 7.69 (1H, d, *J* = 7.5 Hz), 8.23–8.33 (1H, m), and 8.74 (1H, br); ^{13}C NMR ($CDCl_3$) δ: 20.1, 20.2, 20.2, 20.3, 24.5, 24.6, 26.9, 38.8, 40.8, 43.1, 43.1, 43.2, 43.3, 55.1, 55.2, 58.1, 58.3, 61.9, 62.3, 67.2, 85.7, 86.8, 110.7, 113.2, 117.3, 117.5, 122.1, 122.7, 126.1, 126.3, 127.1, 128.0, 128.1, 128.2, 130.0, 130.0, 130.1, 135.1, 135.3, 139.4, 144.1, 158.6, and 177.9; ^{31}P NMR ($CDCl_3$) δ: 149.6, and 150.4; HRMS (ESI): *m*/*z* calcd for $C_{54}H_{64}N_6O_{10}P^+$ $[M+H]^+$: 987.4422, found 987.4440.

Notes and references

Department of Life Science, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa, Japan
E-mail: msekine@bio.titech.ac.jp

† Electronic Supplementary Information (ESI) available: NMR spectra of all new compounds, *ab initio* calculations, computational methods for MD simulations, RESP charges, force field parameters, and representative structure of dC^{Inc} during MD simulation. See DOI: 10.1039/b000000x/

1 (a) K. Y. Lin, R. J. Jones and M. D. Matteucci, *J. Am. Chem. Soc*. 1995, **117**, 3873; (b) K. Y. Lin and M. D. Matteucci, *J. Am. Chem. Soc*. 1998, **120**, 8531; (c) C. J. Wilds, M. A. Maier, V. Tereshko, M. Manoharan and M. Egli, *Angew. Chem. Int. Ed*. 2002, **41**, 115; (d) O. Nakagawa, S. Ono, Z. Li, A. Tsujimoto and S. Sasaki, *Angew. Chem. Int. Ed*. 2007, **46**, 4500; (e) L. M. Wilhelmsson, A. Holmén, P. Lincoln, P. E. Nielsen and B. Nordén, *J. Am. Chem. Soc*. 2001, **123**, 2434; (f) D. A. Berry, K. Y. Jung, D. S. Wise, A. D. Sercel, W. H. Pearson, H. Mackie, J. B. Randolph and R. L. Somers, *Tetrahedron Lett*. 2004, **45**, 2457; (g) P. Sandin, K. Börjesson, H. Li, J. Mårtensson, T. Brown, L. M. Wilhelmsson and B. Albinsson, *Nucleic Acids Res*. 2008, **36**, 157; (h) F. Wojciechowski and R. H. E. Hudson, *J. Am. Chem. Soc*. 2008, **130**, 12574; (i) A. S. Wahba, A. Esmaeili, M. J. Damha and R. H. E. Hudson, *Nucleic Acids Res*. 2010, **38**, 1048; (j) S. Preus, K. Kilså, L. M. Wilhelmsson and B. Albinsson, *Phys. Chem. Chem. Phys*. 2010, **12**, 8881; (k) A. Ohkubo, T. Sakaue, H. Tsunoda, K. Seio and M. Sekine, *Chem. Lett*. 2010, **39**, 726; (l) R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chem. Rev*. 2010, **110**, 2579; (m) K. G. Rajeev, M. A. Maier, E. A. Lesnik and M. Manoharan, *Org. Lett*. 2002, **4**, 4395; (n) J. A. Ortega, J. R. Blas, M. Orozco, A. Grandas, E. Pedroso and J. Robles, *Org. Lett*. 2007, **9**, 4503.

2 (a) T. Wada, A. Kobori, S. Kawahara and M. Sekine, *Tetrahedron Lett*. 1998, **39**, 6907; (b) T. Wada, A. Kobori, S. Kawahara and M. Sekine, *Eur. J. Org. Chem*. 2001, 4583; (c) T. Wada, A. Kobori, S. Kawahara and M. Sekine, *Tetrahedron Lett*. 1998, **39**, 6907; (d) A. Ohkubo, K. Seio and M. Sekine, *J. Am. Chem. Soc*. 2004, **126**, 10884.

3 (a) R. Parthasarathy, S. L. Ginell, N. C. De and G. B. Chheda, *Biochem. Biophys. Res. Commun*. 1978, **83**, 657; (b) W. Saenger, in *Principles of Nucleic Acid Structure*, Springer Verlag, New York, 1984.

4 (a) A. Ohkubo, R. Kasuya, K. Sakamoto, K. Miyata, H. Taguchi, H. Nagasawa, T. Tsukahara, T. Watanobe, Y. Masaki, K. Seio and M. Sekine, *Nucleic Acids Res*. 2008, **36**, 1952; (b) S. K. Mzumdar and W. Saenger, *J. Mol. Biol*. 1974, **85**, 213.

5 H. Schaller, G. Wimann, B. Lerch and J. G. Khorana, *J. Am. Chem. Soc*. 1963, **85**, 3821.

6 R. K. Ralph and H. G. Khorana *J. Am. Chem. Soc*. 1961, **83**, 2926; (b) H. Köster, K. Kulikowski, T. Liese, W. Heikens, V. Kohli, *Tetrahedron* 1981, **37**, 363.

7 (a) S. Clementi and G. Marino, *Tetrahedron* 1969, **25**, 4599; (b) A. D. Allen, J. M. Kwong-Chip, W. C. Lin, P. Nguyen and T. T. Tidwell, *Can. J. Chem*. 1990, **68**, 1709.

8 L. Araki, K. Morita, M. Yamaguchi, Z. Zhao, T. J. Wilson, D. M. J. Lilley and S. Harusawa, *J. Org. Chem*. 2009, **74**, 2350; (b) S. Harusawa, H. Yoneyama, D. Fujisue, M. Nishiura, M. Fujitake, Y. Usami, Z. Zhao, S. A. McPhee, T. J. Wilson and D. M. J. Lilley, *Tetrahedron Lett*. 2012, **53**, 5891.

9 M. Leuck and A. Wolter, PCT Int. WO 2006116476, 2006.

10 R. Welz and S. Müller, *Tetrahedron Lett*. 2002, **43**, 795.

11 C. I. Bayly, P. Cieplak, W. Cornell, P.A. Kollman, *J. Phys. Chem*. **1993**, *97*, 10269.

12 (a) Y. Masaki, R. Miyasaka, A. Ohkubo, K. Seio, M. Sekine, *J. Phys. Chem. B* 2010, **114**, 2517. (b) Y. Masaki, R. Miyasaka, K. Hirai, H. Tsunoda, A. Ohkubo, K. Seio, M. Sekine, *Chem. Commun*. 2012, **48**, 7313.

13 P. Auffinger and E. Westhof, *Angew. Chem. Int. Ed*. 2001, **40**, 4648.