CHEMBIOCHEM

Supporting Information

CHEMBIOCHEM

Supporting Information

for

Optimization of the Pyridyl Nucleobase-Scaffold for Polymerase Recognition and Unnatural Base Pair Replication

Yoshiyuki Hari, Gil Tae Hwang, Aaron M. Leconte, Nicolas Joubert, Michal Hocek, and Floyd E. Romesberg*

Experimental Section

Materials. 2-Bromo-6-methylpyridine **1a**, 2-bromo-5-methylpyridine **1b** and 2-bromo-4-methylpyridine **1c** were purchased from Aldrich. 5,6-Dimethylpyridin-2-ylamine for the synthesis of 2-bromo-5,6-dimethylpyridine **1d** was purchased from Oakwood Product Inc (Scheme S1). 1,1´-Azobis(*N*,*N*-dimethylforamide) was purchased from Tokyo Chemical Industry Co., Ltd. Other reagents and solvents were purchased from Aldrich or Acros and used without further purification. Free nucleosides **7a**-**d**,[1] 2-bromo-4,6-dimethylpyridine **1e**,[2] 2-bromo-4,5-dimethylpyridine **1f**[3] and 2-bromoquinoline **1g**[4] were synthesized according to reported procedures. All reactions were carried out with dry glassware under argon atmospheres. Analytical TLC was carried out on Merck 60F_{254} silica gel plate and column chromatography was performed on silica gel 60 (Geduran, 40-63 µm, Merck). ^{1}H, ^{13}C and ^{31}P NMR spectra were taken on a Bruker NMR spectrometer (DRX-500 or AMX-400). The ^{1}H and ^{13}C chemical shifts are referenced relative to TMS, and the ^{31}P chemical shifts are referenced relative to 85% phosphoric acid in D_2O. High resolution mass spectroscopic data were obtained on an Agilent ESI-TOF mass spectrometer at The Scripps Research Institute Center for Mass Spectrometry

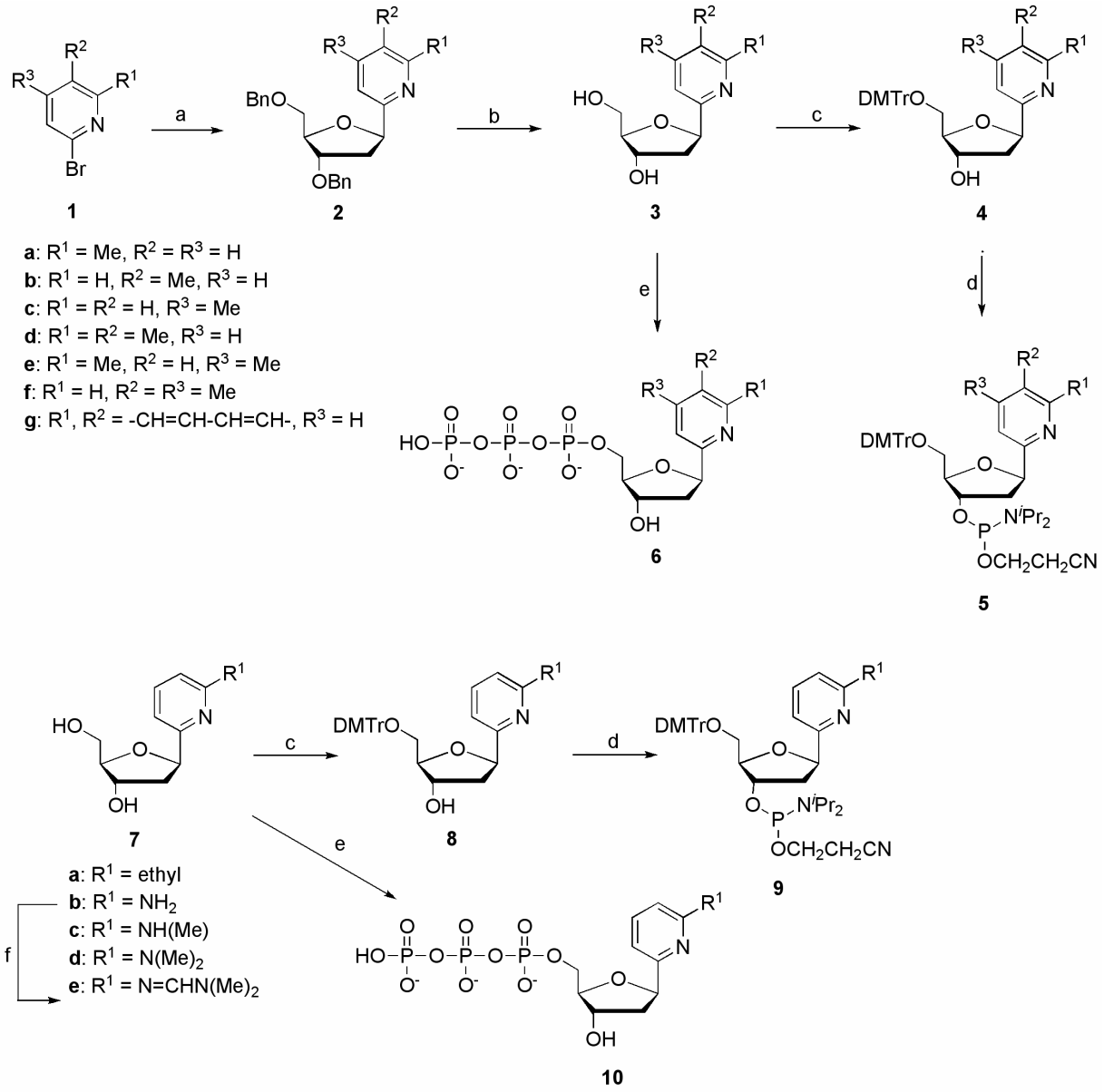
Steady-state kinetics. Primer was 5' radiolabeled with [γ-^{33}P]-ATP (GE Healthcare) and T4 polynucleotide kinase (New England Biolabs). Primer-template duplexes were annealed in the reaction buffer by heating to 90 °C and slow cooling to room temperature. Assay conditions include: 40 nM template-primer duplex, 0.30-1.2 nM enzyme (Klenow fragment exo$^-$, GE Healthcare), 50 mM Tris buffer (pH 7.5), 10 mM $MgCl_2$, 1 mM DTT and 50 µg/mL BSA. The reactions were initiated by adding the DNA-enzyme mixture to an equal volume (5 µL) of a 2× triphosphate stock solution, incubated at 25°C for 3-12 min and quenched with 20 µL of loading buffer (95% formamide, 20 mM EDTA). The reaction mixture (8 µL) was then analyzed by 15% polyacrylamide gel electrophoresis. Radioactivity was quantified using a Phosphorimager (Molecular Dynamics) with overnight exposures and the ImageQuant program. The Michaelis-Menten equation was fit to the data using the program Kaleidagraph (Synergy software). The data presented are averages of triplicates.

Heteropair synthesis/extension screen. The primer was 5' radiolabeled with [γ-^{33}P]-ATP and T4 polynucleotide kinase. Primer- template duplexes were annealed in the reaction buffer by heating to 90 ºC and slow cooling to room temperature. Assay conditions were as follows: 40 nM template-primer duplex, 0.6 nM enzyme, 50 mM Tris buffer (pH 7.5), 10 mM $MgCl_2$, 1 mM DTT and 50 µg/mL BSA. The reactions were initiated by adding the DNA-enzyme mixture to an equal volume (5 µL) of a 2 x triphosphate stock solution resulting in a final concentration of 5, 20, 100 and 500 µM d**X**TP and dCTP, incubated at 25 ºC for 5 min, and quenched with 20 µL of loading buffer (95% formamide, 20 mM EDTA). The reaction mixture (8 µL) was then analyzed by 15% polyacrylamide gel electrophoresis. Radioactivity was quantified using a Phosphorimager (Molecular Dynamics) with overnight exposures and the ImageQuant program.

General full-length assay protocol. Primer was 5' radiolabeled as described above. Primer-template duplexes were annealed in the reaction buffer by heating to 90 °C and slow cooling to room temperature. Assay conditions include: 40 nM template-primer duplex, 6 nM enzyme, 50 mM Tris buffer (pH 7.5), 10 mM $MgCl_2$, 1 mM DTT and 50 µg/mL BSA. The reactions were initiated by adding the DNA-enzyme mixture to an equal volume (5 µL) of a 2× triphosphate stock solution containing either 20 µM of the four natural dNTPs or 20 µM of dNTP with 500 µM of d**45DMPy** triphosphate. Reactions were incubated at 25 °C for 3 min, and quenched with 20 µL of loading buffer (95% formamide, 20 mM EDTA). The reaction mixture (8 µL) was then ana-

lyzed by 15% polyacrylamide gel electrophoresis. Radioactivity was quantified using a Phosphorimager (Molecular Dynamics) with overnight exposures and the Image-Quant program.

Compound Preparation



Scheme S1. Synthesis of 2-pyridine analogs a) *i.* ^{n}BuLi, 3,5-di-*O*-benzyl-2-deoxy-D-erythro-pentofuranose, THF, -78 ºC; *ii,* 1,1´-azobis(*N*,*N*-dimethylformamide), $^{n}Bu_3P$, benzene, RT; b) 20 % $Pd(OH)_2$/C, cyclohexene, EtOH, reflux; c) DMTrCl, DMAP, pyridine, RT; d) 2-cyano-ethyl *N*,*N*-diisopropylchlorophosphoramidite, DIPEA, CH_2Cl_2, RT; e) Proton Sponge, $POCl_3$, $^{n}Bu_3N$, $^{n}Bu_3NPPi$, $(MeO)_3P$, DMF, -10 ºC; f) *N*,*N*-Dimethylformamide dimethyl acetal, MeOH, reflux.

Synthesis of 1d. Bromine (1.68 mL, 32.8 mmol) was added dropwise to a solution of 5,6-dimethylpyridin-2-ylamine (2.00 g, 16.4 mmol) in 48 % HBr aq. (9 mL) at -5 °C. After being stirred for 10 min at -5 °C, a solution of $NaNO_2$ (2.03 g, 29.4 mmol) in H_2O (3 mL) was added while the temperature was kept below 5 °C and furthermore H_2O (10 mL) was added. After being stirred for additional 30 min, the reaction solution was made alkaline by addition of a solution of NaOH (7 g) in H_2O (10 mL). The mixture was extracted with Et_2O and the organic layer was washed with H_2O and brine, dried over Na_2SO_4 and concentrated in vacuo. The residue was purified by silica gel column chromatography (6.3% EtOAc in hexane) as an eluent to give **1d** (2.25 g, 74 %). ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.25 (d, *J* = 7.5 Hz, 1H; ArH), 7.21 (d, *J* = 7.5 Hz, 1H; ArH), 2.47 (s, 3H; $ArCH_3$), 2.23 (s, 3H; $ArCH_3$); ^{13}C NMR (125 MHz, $CDCl_3$): d 158.5, 139.6, 138.0, 130.6, 125.2, 22.3, 18.4; HRMS (*m/z*): $[M + H]^+$ calcd for $C_7H_9Br_1N_1$, 185.9913; found, 185.9910.

Synthesis of 7e. To a stirred solution of **7b** (97.4 mg, 0.463 mmol) in MeOH (7.7 mL) was added dropwise *N*,*N*-dimethylformamide dimethyl acetal (0.22 mL, 1.66 mmol) at 0 °C. The solution was refluxed overnight. After the reaction was complete (15 h), The mixture was evaporated and purified by chromatography on a silica gel column (15% MeOH in EtOAc) afforded protected **7e** (120 mg, 98%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.19 (s, 1H; imine-H), 7.47 (t, *J* = 7.8 Hz, 1H; ArH), 6.85 (d, *J* = 8.0 Hz, 1H; ArH), 6.79 (d, *J* = 7.2 Hz, 1H; ArH), 5.18 (dd, *J* = 6.8, 9.2 Hz, 1H; H-1'), 4.55 (d, *J* = 4.8 Hz; H-3'), 4.14 (br s, 1H; H-4'), 3.89 (dd, *J* = 2.8, 12.0 Hz, 1H; H-5'), 3.66 (dd, *J* = 2.0, 12.0 Hz, 1H; H-5'), 3.07 and 3.04 (2s, 6H; CH_3), 2.46–2.39 (m, 1H; H-2'), 2.25–2.21 (m, 1H; H-2'); ^{13}C NMR (101 MHz, $CDCl_3$): δ 162.5, 159.4, 155.7, 138.4, 117.9, 115.9, 88.7, 80.8, 75.2, 64.0, 40.7, 34.7; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{13}H_{20}N_3O_3$, 266.1499; found, 266.1502.

General Procedure for benzyl-protected nucleoside synthesis. A solution of 3,5-di-*O*-benzyl-2-deoxy-D-erythro-pentofuranose[5] (1 equiv) in THF (0.25 M) was added to the solution prepared from **1** (3.0 equiv) and ^{n}BuLi (1.6 M in hexane, 3.0 equiv) in THF (0.38 M), at -78 °C. After being stirred for 1 h at 0 °C, the reaction mixture was quenched by addition of H_2O and extracted with EtOAc. The organic layer was washed with H_2O and brine, dried over Na_2SO_4 and concentrated in vacuo. The residue was purified by chromatography on a silica gel column (50% EtOAc in hexane) to give appropriate compounds, which were dissolved in benzene (42 mM) and 1,1´-azobis(*N*,*N*-dimethylformamide) (1.2 equiv) and ^{n}Bu$_3$P (1.2 equiv) were added at RT.

After being stirred for 5 h at RT, the mixture was filtered through a pad of Celite and the filtrate was concentrated in vacuo. The residue was purified by silica gel column chromatography (20% EtOAc in hexane) to give β-anomer **2** and a-anomer.

Compound 2a. 36% yield. ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.51 (t, *J* = 7.6 Hz, 1H; ArH), 7.33-7.24 (m, 11H; ArH), 7.00 (d, *J* = 7.6 Hz, 1H; ArH), 5.23 (dd, *J* = 5.9, 9.8 Hz, 1H; H-1'), 4.59 (d, *J* = 11.9 Hz, 1H; $ArCH_2$), 4.58 (s, 2H; $ArCH_2$), 4.51 (d, *J* = 11.9 Hz, 1H; $ArCH_2$), 4.34-4.32 (m, 1H H-3'), 4.16-4.15 (m, 1H; H-4'), 3.67 (dd, *J* = 4.7, 10.2 Hz, 1H; H-5'), 3.61 (dd, *J* = 5.2, 10.2 Hz, 1H; H-5'), 2.56 (ddd, *J* = 1.1, 5.9, 13.2 Hz, 1H; H-2'), 2.52 (s, 3H; $ArCH_3$), 2.02 (ddd, *J* = 5.9, 9.8, 13.2 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, $CDCl_3$): δ 161.1, 157.4, 138.2, 138.2, 136.8, 128.3, 128.3, 127.6, 127.5, 121.8, 117.0, 84.0, 81.3, 81.0, 73.4, 71.0, 39.2, 24.4; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{25}H_{28}N_1O_3$, 390.2064; found, 390.2075.

Compound 2b. 39% yield. ^{1}H NMR (500 MHz, $CDCl_3$): δ 8.35 (s, 1H; ArH), 7.42 (d, *J* = 8.0 Hz, 1H; ArH), 7.39 (d, *J* = 8.0 Hz, 1H; ArH), 7.33-7.26 (m, 10H; ArH), 5.24 (dd, *J* = 5.7, 9.8 Hz, 1H; H-1'), 4.58 (d, *J* = 11.9 Hz, 1H; $ArCH_2$), 4.57 (s, 2H; $ArCH_2$), 4.52 (d, *J* = 11.9 Hz, 1H; $ArCH_2$), 4.34-4.33 (m, 1H; H-3'), 4.16-4.15 (m, 1H; H-4'), 3.65 (dd, *J* = 4.6, 10.0 Hz, 1H; H-5'), 3.59 (dd, *J* = 5.2, 10.0 Hz, 1H; H-5'), 2.53 (dd, *J* = 5.7, 13.1 Hz, 1H; H-2'), 2.28 (s, 3H; $ArCH_3$), 2.06 (ddd, *J* = 6.0, 9.8, 13.1 Hz, 1H; H-2'); ^{13}C NMR (101 MHz, $CDCl_3$): δ 158.4, 149.1, 138.1, 138.0, 137.0, 131.6, 128.3, 128.2, 127.5, 127.5, 119.8, 83.9, 81.0, 80.9, 73.3, 70.9, 38.9, 18.0; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{25}H_{28}N_1O_3$, 390.2064; found, 390.2063.

Compound 2c. 37% yield. ^{1}H NMR (500 MHz, $CDCl_3$): δ 8.37 (d, *J* = 4.9 Hz, 1H; ArH), 7.34-7.27 (m, 11H; ArH), 6.95 (d, *J* = 4.9 Hz, 1H; ArH), 5.24 (dd, *J* = 5.9, 9.9 Hz, 1H; H-1'), 4.59 (d, *J* = 11.9 Hz, 1H; $ArCH_2$), 4.58 (s, 2H; $ArCH_2$), 4.51 (d, *J* = 11.9 Hz, 1H; $ArCH_2$), 4.34-4.33 (m, 1H; H-3'), 4.16-4.15 (m, 1H; H-4'), 3.67 (dd, *J* = 4.7, 10.1 Hz, 1H; H-5'), 3.61 (dd, *J* = 5.2, 10.1 Hz, 1H; H-5'), 2.55 (dd, *J* = 5.9, 13.2 Hz, 1H; H-2'), 2.28 (s, 3H; $ArCH_3$), 2.05 (ddd, *J* = 6.0, 9.9, 13.2 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, $CDCl_3$): δ 161.1, 148.5, 147.7, 138.1, 138.0, 128.3, 128.2, 127.5, 127.5, 127.5, 127.4, 123.2, 121.0, 84.0, 81.0, 80.9, 73.3, 70.9, 70.9, 39.0, 21.0; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{25}H_{28}N_1O_3$, 390.2064; found, 390.2063.

Compound 2d. 28% yield. ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.33-7.22 (m, 12H; ArH), 5.23 (dd, *J* = 5.8, 9.9 Hz, 1H; H-1'), 4.57 (d, *J* = 12.1 Hz, 1H; $ArCH_2$), 4.56 (d, *J* = 11.9 Hz, 1H; $ArCH_2$), 4.55 (d, *J* = 12.1 Hz, 1H; $ArCH_2$), 4.49 (d, *J* = 11.9 Hz, 1H;

$ArCH_2$), 4.33-4.31 (m, 1H; H-3'), 4.16-4.14 (m, 1H; H-4'), 3.64 (dd, *J* = 4.7, 10.2 Hz, 1H; H-5'), 3.59 (dd, *J* = 5.3, 10.2 Hz, 1H; H-5'), 2.53 (ddd, *J* = 1.8, 5.8, 13.2 Hz, 1H; H-2'), 2.44 (s, 3H; $ArCH_3$), 2.19 (s, 3H; $ArCH_3$), 2.03 (ddd, *J* = 6.0, 9.9, 13.2 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, $CDCl_3$): δ 158.0, 155.8, 138.0, 137.4, 129.7, 128.1, 128.1, 127.4, 127.3, 127.3, 117.4, 83.8, 81.0, 80.9, 73.2, 70.8, 70.7, 39.0, 22.3, 18.66; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{26}H_{30}N_1O_3$, 404.2220; found, 404.2227.

Compound 2e. 33% yield. ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.31-7.20 (m, 10H; ArH), 7.16 (s, 1H; ArH), 6.76 (s, 1H; ArH), 5.23 (dd, *J* = 5.9, 9.9 Hz; H-1'), 4.56-4.54 (m, 3H; $ArCH_2$), 4.47 (d, *J* = 11.9 Hz, 1H; $ArCH_2$), 4.34-4.31 (m, 1H; H-3'), 4.15-4.14 (m, 1H; H-4'), 3.65 (dd, *J* = 4.5, 10.1 Hz, 1H; H-5'), 3.60 (dd, *J* = 5.1, 10.1 Hz, 1H; H-5'), 2.55 (ddd, *J* = 1.5, 5.9, 13.2 Hz, 1H; H-2'), 2.45 (s, 3H; $ArCH_3$), 2.18 (s, 3H; $ArCH_3$), 2.01 (ddd, *J* = 6.0, 9.9, 13.2 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, $CDCl_3$): δ 160.5, 156.7, 147.4, 137.9, 137.8, 127.9, 127.1, 127.1, 127.1, 122.4, 117.5, 83.7, 80.9, 80.6, 72.9, 70.6, 70.5, 38.9, 23.8, 20.5; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{26}H_{30}N_1O_3$: 404.2220; found, 404.2217.

Compound 2f. 36% yield. ^{1}H NMR (500 MHz, $CDCl_3$): δ 8.23 (s, 1H; ArH), 7.34-7.23 (m, 11H; ArH), 5.21 (dd, *J* = 5.8, 10.0 Hz, 1H; H-1'), 4.59 (d, *J* = 11.9 Hz, 1H; $ArCH_2$), 4.58 (s, 2H; $ArCH_2$), 4.51 (d, *J* = 11.9 Hz, 1H; $ArCH_2$), 4.36- 4.33 (m, 1H; H-3'), 4.17-4.15 (m, 1H; H-4'), 3.67 (dd, *J* = 4.7, 10.2 Hz, 1H; H-5'), 3.61 (dd, *J* = 5.2, 10.2 Hz, 1H; H-5'), 2.52 (ddd, *J* = 1.8, 5.8, 13.2 Hz, 1H; H-2'), 2.19 (s, 3H; $ArCH_3$), 2.19 (s, 3H; $ArCH_3$), 2.06 (ddd, *J* = 6.1, 10.0, 13.2 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, $CDCl_3$): δ 158.7 148.9, 146.2, 138.1, 138.0, 130.7, 128.2, 127.5, 127.5, 127.4, 127.4, 121.1, 83.9, 81.0, 80.9, 73.2, 70.9, 38.9, 19.1, 16.0; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{26}H_{30}NO_3$, 404.2220; found, 404.2229.

Compound 2g. 18% yield. ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.97-7.94 (m, 2H; ArH), 7.63 (d, *J* = 7.9 Hz, 1H; ArH), 7.58 (d, *J* = 8.6 Hz, 1H; ArH), 7.56-7.52 (m, 1H; ArH), 7.35 (t, *J* = 7.2 Hz, 1H; ArH), 7.24-7.14 (m, 10H; ArH), 5.35 (dd, *J* = 5.9, 10.0 Hz, 1H; H-1'), 4.49 (d, *J* = 11.9 Hz, 1H; $ArCH_2$), 4.48 (d, *J* = 12.1 Hz, 1H; $ArCH_2$), 4.46 (d, *J* = 12.1 Hz, 1H; $ArCH_2$), 4.41 (d, *J* = 11.9 Hz, 1H; $ArCH_2$), 4.31-4.28 (m, 1H; H-3'), 4.10-4.08 (m, 1H; H-4'), 3.57 (dd, *J* = 4.5, 10.2 Hz, 1H; H-5'), 3.54 (dd, *J* = 4.9, 10.2 Hz, 1H; H-5'), 2.53 (ddd, *J* = 1.5, 5.9, 13.3 Hz, 1H; H-2'), 2.04 (ddd, *J* = 5.9, 10.0, 13.3 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, $CDCl_3$): δ 162.1, 147.2, 138.0, 138.0, 136.6,

129.3, 128.9, 128.3, 128.2, 127.5, 126.0, 118.3, 84.3, 81.7, 81.0, 73.3, 70.9, 70.9, 39.26; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{28}H_{28}N_1O_3$, 426.2064; found, 426.2065.

Generral procedure for debenzylation. A solution of **2** (1 equiv), 20 % $Pd(OH)_2/C$ and cyclohexene (27 equiv) in EtOH (0.15 M) was refluxed for 2 h and after cooling to RT, the mixture was filtered through a filter paper. The filtrate was concentrated in vacuo and the residue was purified by silica gel column chromatography (4% MeOH in EtOAc) to give free nucleoside **3**.

Compound 3a. 51% yield. ^{1}H NMR (500 MHz, CD_3OD): δ 7.68 (t, *J* = 7.7 Hz, 1H; ArH), 7.33 (d, *J* = 7.7 Hz, 1H; ArH), 7.17 (d, *J* = 7.7 Hz, 1H; ArH), 5.18 (dd, *J* = 6.3, 9.5 Hz, 1H; H-1'), 4.39-4.38 (m, 1H; H-3'), 4.04-4.02 (m, 1H; H-4'), 3.76 (dd, *J* = 3.9, 11.9 Hz, 1H; H-5'), 3.69 (dd, *J* = 4.0, 11.9 Hz, 1H; H-5'), 2.50 (s, 3H; $ArCH_3$), 2.31 (dd, *J* = 6.3, 13.1 Hz, 1H; H-2'), 2.08 (ddd, *J* = 5.9, 9.5, 13.1 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, CD_3OD): δ 162.4, 159.0, 139.1, 123.9, 119.3, 89.7, 81.8, 74.5, 64.2, 44.1, 23.8; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{11}H_{16}N_1O_3$, 210.1125; found, 210.1124.

Compound 3b. 89% yield. ^{1}H NMR (500 MHz, CD_3OD): δ 8.29 (s, 1H; ArH), 7.61 (d, *J* = 8.0 Hz, 1H; ArH), 7.44 (d, *J* = 8.0 Hz, 1H; ArH), 5.18 (dd, *J* = 6.0, 9.8 Hz, 1H; H-1'), 4.38-4.37 (m, 1H; H-3'), 4.03-4.01 (m, 1H; H-4'), 3.73 (dd, *J* = 4.1, 11.9 Hz, 1H; H-5'), 3.68 (dd, *J* = 4.5, 11.9 Hz, 1H; H-5'), 2.30 (s, 3H; $ArCH_3$), 2.30-2.27 (m, 1H; H-2'), 2.04 (ddd, *J* = 5.9, 9.8, 13.1 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, CD_3OD): δ 159.7, 149.7, 139.3, 134.1, 121.9, 89.6, 81.5, 74.2, 64.0, 43.9, 18.1; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{11}H_{16}N_1O_3$, 210.1125; found, 210.1124.

Compound 3c. 83% yield. ^{1}H NMR (500 MHz, CD_3OD): δ 8.31 (d, *J* = 5.0 Hz, 1H; ArH), 7.46 (s, 1H; ArH), 7.14 (d, *J* = 5.0 Hz, 1H; ArH), 5.17 (dd, *J* = 6.0, 9.9 Hz, 1H; H-1'), 4.37-4.36 (m, 1H; H-3'), 4.02-4.00 (m, 1H; H-4'), 3.74 (dd, *J* = 4.1, 11.9 Hz, 1H; H-5'), 3.68 (dd, *J* = 4.5, 11.9 Hz, 1H; H-5'), 2.37 (s, 3H; $ArCH_3$), 2.30 (ddd, *J* = 1.2, 6.0, 13.1 Hz, 1H; H-2'), 2.03 (ddd, *J* = 5.9, 9.9, 13.1 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, CD_3OD): δ 162.5, 150.8, 149.2, 125.0, 123.1, 89.7, 81.7, 74.2, 64.0, 44.0, 21.2; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{11}H_{16}N_1O_3$, 210.1125; found, 210.1121.

Compound 3d. 96% yield. ^{1}H NMR (500 MHz, CD_3OD): δ 7.52 (d, *J* = 7.7 Hz, 1H; ArH), 7.24 (d, *J* = 7.7 Hz, 1H; ArH), 5.16 (dd, *J* = 6.3, 9.6 Hz, 1H; H-1'), 4.40-4.38 (m, 1H; H-3'), 4.03-4.01 (m, 1H; H-4'), 3.77 (dd, *J* = 3.9, 11.9 Hz, 1H; H-5'), 3.68 (dd, *J* = 3.9, 11.9 Hz, 1H; H-5'), 2.46 (s, 3H; $ArCH_3$), 2.31-2.26 (m, 1H; H-2'), 2.29 (s, 3H;

$ArCH_3$), 2.10 (ddd, *J* = 5.8, 9.6, 13.2 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, CD_3OD): δ 159.5, 157.6, 139.7, 132.5, 119.9, 89.7, 81.7, 74.7, 64.4, 44.2, 21.8, 18.8; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{12}H_{17}N_1O_3$, 224.1281; found, 224.1279.

Compound 3e. 80% yield. ^1H NMR (500 MHz, CD_3OD): δ 7.18 (s, 1H; ArH), 7.01 (s, 1H; ArH), 5.14 (dd, *J* = 6.3, 9.6 Hz, 1H; H-1'), 4.39-4.37 (m, 1H; H-3'), 4.03-4.01 (m, 1H; H-4'), 3.77 (dd, *J* = 3.8, 12.0 Hz, 1H; H-5'), 3.69 (dd, *J* = 4.0, 12.0 Hz, 1H; H-5'), 2.45 (s, 3H; $ArCH_3$), 2.33 (s, 3H; $ArCH_3$), 2.29 (ddd, *J* = 2.0, 13.2, 6.3, 13.2 Hz, 1H; H-2'), 2.07 (ddd, *J* = 5.9, 9.6, 13.2 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, CD_3OD): δ 162.1, 158.7, 150.9, 124.6, 120.2, 89.7, 81.7, 74.5, 64.2, 44.2, 23.5, 21.0; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{12}H_{18}N_1O_3$, 224.1281; found, 224.1282.

Compound 3f. 82% yield. ^1H NMR (500 MHz, CD_3OD): δ 8.18 (s, 1H; ArH), 7.35 (s, 1H; ArH), 5.13 (dd, *J* = 6.0, 9.9 Hz, 1H; H-1'), 4.36-4.34 (m, 1H; H-3'), 4.00-3.97 (m, 1H; H-4'), 3.74 (dd, *J* = 4.0, 11.9 Hz, 1H; H-5'), 3.67 (dd, *J* = 4.5, 11.9 Hz, 1H; H-5'), 2.32 (s, 3H; $ArCH_3$), 2.28-2.24 (m, 1H; H-2'), 2.27 (s, 3H; $ArCH_3$), 2.03 (ddd, *J* = 6.0, 9.9, 13.1 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, CD_3OD): δ 160.1, 149.5, 149.2, 133.2, 123.2, 89.7, 81.6, 74.3, 64.1, 44.0, 19.3, 16.1; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{12}H_{18}N_1O_3$, 224.1281; found, 224.1280.

Compound 3g. 60% yield. ^1H NMR (500 MHz, CD_3OD): δ 8.30 (d, *J* = 8.5 Hz, 1H; ArH), 7.98 (d, *J* = 8.5 Hz, 1H; ArH), 7.89 (d, *J* = 7.9 Hz, 1H; ArH), 7.75-7.72 (m, 1H; ArH), 7.71 (d, *J* = 8.5 Hz, 1H; ArH), 7.57-7.54 (m, 1H; ArH), 5.39 (dd, *J* = 6.2, 9.6 Hz, 1H; H-1'), 4.44-4.42 (m, 1H; H-3'), 4.10-4.08 (m, 1H; H-4'), 3.81 (dd, *J* = 4.2, 11.8 Hz, 1H; H-5'), 3.75 (dd, *J* = 1.0, 4.5, 11.8 Hz, 1H; H-5'), 2.41 (ddd, *J* = 2.2, 6.2, 13.1 Hz, 1H; H-2'), 2.16 (ddd, *J* = 5.9, 9.6, 13.1 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, CD_3OD): δ 163.9, 148.2, 139.0, 131.2, 129.2, 129.1, 128.8, 127.8, 120.0, 89.8, 82.3, 74.1, 64.0, 44.1; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{14}H_{16}NO_3$, 246.1125; found, 246.1124.

General procedure for DMTr-protection. To a solution of free nucleoside **3** or **7** (1 equiv) in anhydrous pyridine was added 4,4'-dimethoxytrityl chloride (1.5 equiv) and 4-(dimethylamino)pyridine (0.5 equiv) and the mixture was stirred overnight under argon at room temperature. The reaction was quenched by addition of MeOH. The mixture was evaporated, the crude product was purified by chromatography on a silica gel column (30-40% EtOAc in hexane).

Compound 4a. 81% yield. ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.51 (t, *J* = 7.6 Hz, 1H; ArH), 7.46 (d, *J* = 7.4 Hz, 2H; ArH), 7.36-7.33 (m, 5H; ArH), 7.26 (dd, *J* = 7.3, 7.4 Hz, 2H; ArH), 7.19 (t, *J* = 7.3 Hz, 1H; ArH), 7.01 (d, *J* = 7.6 Hz, 1H; ArH), 6.81 (d, *J* = 8.7 Hz, 4H; ArH), 5.27 (dd, *J* = 6.2, 9.1 Hz, 1H; H-1'), 4.41-4.40 (m, 1H; H-3'), 4.14-4.12 (m, 1H; H-4'), 3.77 (s, 6H; OCH_3), 3.34 (dd, *J* = 4.6, 9.8 Hz, 1H; H-5'), 3.31 (dd, *J* = 5.2, 9.8 Hz, 1H; H-5'), 2.60 (br s, 1H; OH), 2.51 (s, 3H; $ArCH_3$), 2.41 (ddd, *J* = 2.7, 6.2, 13.0 Hz, 1H; H-2'), 2.15 (ddd, *J* = 5.9, 9.1, 13.0 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, $CDCl_3$): δ 161.3, 158.4, 157.4, 144.9, 136.9, 136.1, 130.1, 128.2, 127.8, 126.7, 121.8, 117.0, 113.1, 86.2, 80.7, 74.1, 64.4, 55.2, 42.4, 24.3; HRMS (*m/z*): $[M + Na]^+$ calcd for $C_{32}H_{33}N_1O_5Na_1$, 534.2251; found, 534.2256.

Compound 4b. 62% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.33 (d, *J* = 0.9 Hz, 1H; ArH), 7.46- 7.17 (m, 11H; ArH), 6.81 (d, *J* = 8.9 Hz, 4H; ArH), 5.28 (dd, *J* = 6.2, 9.3 Hz, 1H; H-1'), 4.42-4.39 (m, 1H; H-3'), 4.14-4.11 (m, 1H; H-4'), 3.77 (s, 6H; OCH_3), 3.33 (dd, *J* = 4.6, 9.7 Hz, 1H; H-5'), 3.28 (dd, *J* = 5.2, 9.7 Hz, 1H; H-5'), 2.73 (br s, 1H; OH), 2.39 (ddd, *J* = 2.6, 6.2, 13.0 Hz, 1H; H-2'), 2.30 (s, 3H; $ArCH_3$), 2.16 (ddd, *J* = 6.1, 9.3, 13.0 Hz, 1H; H-2'); ^{13}C NMR (101 MHz, $CDCl_3$): δ 158.8, 158.4, 149.0, 144.8, 137.2, 136.0, 131.7, 130.1, 128.2, 127.9, 126.7, 119.7, 113.1, 86.2, 86.1, 80.4, 74.1, 64.4, 55.2, 42.2, 18.1; HRMS (*m/z*): $[M + Na]^+$ calcd for $C_{32}H_{33}N_1O_5Na_1$, 534.2251; found, 534.2249.

Compound 4c. 67% yield. ^{1}H NMR (500 MHz, $CDCl_3$): δ 8.34 (d, *J* = 5.0 Hz, 1H; ArH), 7.45 (d, *J* = 7.4 Hz, 2H; ArH), 7.48 (s, 1H; ArH), 7.35 (d, *J* = 8.8 Hz, 2H; ArH), 7.34 (d, *J* = 8.8 Hz, 2H; ArH), 7.25 (dd, *J* = 7.3, 7.4 Hz, 2H; ArH), 7.18 (t, *J* = 7.3 Hz, 1H; ArH), 6.90 (d, *J* = 5.0 Hz, 1H; ArH), 6.80 (d, *J* = 8.8 Hz, 2H; ArH), 6.80 (d, *J* = 8.8 Hz, 2H; ArH), 5.30 (dd, *J* = 6.1, 9.5 Hz, 1H; H-1'), 4.43-4.41 (m, 1H; H-3'), 4.17-4.14 (m, 1H; H-4'), 3.76 (s, 6H; OCH_3), 3.40 (br s, 1H; OH), 3.32 (d, *J* = 4.6 Hz, 2H; H-5'), 2.43 (ddd, *J* = 2.3, 6.1, 13.0 Hz, 1H; H-2'), 2.26 (s, 3H; $ArCH_3$), 2.18 (ddd, *J* = 5.9, 9.5, 13.0 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, $CDCl_3$): δ 161.6, 158.4, 148.3, 148.0, 144.9, 136.0, 136.0, 130.0, 128.1, 127.7, 126.7, 123.2, 121.0, 113.0, 86.4, 86.1, 80.5, 74.0, 64.4, 55.1, 42.5, 21.1; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{32}H_{34}N_1O_5$, 512.2431; found, 512.2421.

Compound 4d. 72% yield. ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.37 (d, *J* = 7.4 Hz, 2H; ArH), 7.27-7.24 (m, 5H; ArH), 7.19 (d, *J* = 7.8 Hz, 1H; ArH), 7.14 (dd, *J* = 7.3, 7.4 Hz, 2H; ArH), 7.07 (t, *J* = 7.3 Hz, 1H; ArH), 6.69 (d, *J* = 8.6 Hz, 4H; ArH), 5.19 (dd, *J* =

6.0, 9.4 Hz, 1H; H-1'), 4.32-4.31 (m, 1H; H-3'), 4.07-4.04 (m, 1H; H-4'), 3.64 (s, 6H; OCH_3), 3.36 (br s, 1H; OH), 3.21 (d, *J* = 4.2 Hz, 2H; H-5'), 2.33 (s, 3H; $ArCH_3$), 2.33-2.29 (m, 1H; H-2'), 2.12 (s, 3H; $ArCH_3$), 2.00 (ddd, *J* = 5.8, 9.4, 13.0 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, $CDCl_3$): δ 158.5, 158.3, 155.8, 144.8, 137.8, 136.0, 130.0, 129.8, 128.1, 127.6, 126.6, 117.5, 113.0, 86.2, 86.0, 80.4, 73.8, 64.4, 55.1, 42.6, 22.1, 18.74; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{33}H_{36}N_1O_5$, 526.2588; found, 526.2586.

Compound 4e. 74% yield. ^1H NMR (500 MHz, $CDCl_3$): δ 7.46 (d, *J* = 7.4 Hz, 2H; ArH), 7.36-7.34 (m, 4H; ArH), 7.27-7.18 (m, 4H; ArH), 6.84 (s, 1H; ArH), 6.81 (d, *J* = 7.6 Hz, 4H; ArH), 5.26 (dd, *J* = 6.1, 9.4 Hz, 1H; H-1'), 4.42-4.41 (m, 1H; H-3'), 4.15-4.12 (m, 1H; H-4'), 3.77 (s, 6H; OCH_3), 3.33 (d, *J* = 4.6 Hz, 2H; H-5'), 2.67 (br s, 1H; OH), 2.47 (s, 3H; $ArCH_3$), 2.42 (ddd, *J* = 2.4, 6.1, 13.1 Hz, 1H; H-2'), 2.22 (s, 3H; $ArCH_3$), 2.15 (ddd, *J* = 5.9, 9.4, 13.1 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, $CDCl_3$): δ 161.2, 158.4, 157.0, 148.1, 144.9, 136.1, 136.1, 130.1, 128.2, 127.7, 126.7, 122.8, 117.9, 113.0, 86.3, 86.1, 80.7, 74.0, 64.5, 55.1, 42.8, 24.0, 21.0; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{33}H_{36}N_1O_5$, 526.2588; found, 526.2585.

Compound 4f. 89% yield. ^1H NMR (500 MHz, $CDCl_3$): δ 8.15 (s, 1H; ArH), 7.38 (d, *J* = 7.5 Hz, 2H; ArH), 7.28-7.17 (m, 7H; ArH), 7.12 (t, *J* = 7.2 Hz, 2H; ArH), 6.73 (d, *J* = 8.7 Hz, 2H; ArH), 6.73 (d, *J* = 8.7 Hz, 2H; ArH), 5.17 (dd, *J* = 6.3, 9.2 Hz, 1H; H-1'), 4.35-4.33 (m, 1H; H-3'), 4.05-4.02 (m, 1H; H-4'), 3.70 (s, 6H; OCH_3), 3.27 (dd, *J* = 4.4, 9.8 Hz, 1H; H-5'), 3.23 (dd, *J* = 5.2, 9.8 Hz, 1H; H-5'), 2.31 (ddd, *J* = 2.5, 6.3, 13.1 Hz, 1H; H-2'), 2.16-2.10 (m, 1H; H-2'), 2.14 (s, 3H; $ArCH_3$), 2.11 (s, 3H; $ArCH_3$); ^{13}C NMR (126 MHz, $CDCl_3$): δ 159.1, 158.4, 148.9, 146.4, 144.9, 136.1, 130.8, 130.1, 128.2, 127.8, 126.7, 121.1, 113.1, 86.2, 86.2, 80.5, 74.3, 64.5, 55.2, 42.3, 19.3, 16.1; HRMS (*m/z*): $[M + Na]^+$ calcd for $C_{33}H_{35}N_1O_5Na_1$, 548.2407; found, 548.2401.

Compound 4g. 83% yield. ^1H NMR (500 MHz, $CDCl_3$): δ 8.11 (d, *J* = 8.5 Hz, 1H; ArH), 8.04 (d, *J* = 8.5 Hz, 1H; ArH), 7.79 (d, *J* = 8.0 Hz, 1H; ArH), 7.73-7.67 (m, 2H; ArH), 7.53-7.50 (m, 1H; ArH), 7.46 (d, *J* = 7.3 Hz, 2H; ArH), 7.35 (d, *J* = 8.9 Hz, 2H; ArH), 7.34 (d, *J* = 8.9 Hz, 2H; ArH), 7.25 (t, *J* = 7.3 Hz, 2H; ArH), 7.19 (t, *J* = 7.3 Hz, 1H; ArH), 6.80 (d, *J* = 8.6 Hz, 1H; ArH), 6.79 (d, *J* = 8.5 Hz, 1H; ArH), 5.46 (dd, *J* = 6.3, 9.3 Hz, 1H; H-1'), 4.49-4.47 (m, 1H; H-3'), 4.22-4.19 (m, 1H; H-4'), 3.76 (s, 6H; OCH_3), 3.38 (s, 1H; H-5'), 3.37 (s, 1H), 2.51 (ddd, *J* = 2.6, 6.3, 13.1 Hz, 1H; H-2'), 2.34 (ddd, *J* = 6.0, 9.3, 13.1 Hz, 1H; H-2'); ^{13}C NMR (126 MHz, $CDCl_3$): δ 162.3, 158.4, 147.3, 144.8, 136.8, 136.0, 136.0, 130.1, 129.5, 128.9, 128.2, 127.8, 127.6,

127.5, 126.8, 126.2, 118.4, 113.1, 86.5, 86.2, 81.3, 74.2, 64.3, 55.2, 42.5; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{35}H_{34}N_1O_5$, 548.2431; found, 548.2424.

Compound 8a. 83% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.56 (t, *J* = 8.0 Hz, 1H; ArH), 7.47 (d, *J* = 8.0 Hz, 2H; ArH), 7.36–7.33 (m, 5H; ArH), 7.30–7.25 (m, 2H; ArH), 7.23–7.18 (m, 1H; ArH), 7.04 (d, *J* = 8.0 Hz, 1H; ArH), 6.82 (d, *J* = 8.0 Hz, 4H; ArH), 5.31 (dd, *J* = 5.8, 9.0 Hz, 1H, H-1'), 4.45–4.41 (m, 1H; H-3'), 4.17–4.14 (m, 1H; H-4'), 3.78 (s, 6H; OCH_3), 3.37–3.30 (m, 2H; H-3'), 2.78 (q, *J* = 5.0 Hz, 2H; CH_2CH_3), 2.46–2.41 (m, 1H; H-2'), 2.19–2.13 (m, 1H; H-2'), 2.02 (t, *J* = 6.0 Hz, 3H; CH_2CH_3); ^{13}C NMR (101 MHz, $CDCl_3$): δ 162.6, 161.2, 158.4, 144.8, 137.0, 136.0, 130.1, 128.2, 127.7, 126.7, 120.5, 117.3, 113.0, 86.2, 86.1, 80.7, 74.0, 64.4, 55.1, 42.4, 31.1, 14.0; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{33}H_{36}N_1O_5$, 526.2588; found, 526.2592.

Compound 8c. 82% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.48 (d, *J* = 7.6 Hz, 2H; ArH), 7.43–7.34 (m, 5H; ArH), 7.29–7.21 (m, 2H; ArH), 7.20–7.15 (m, 1H; ArH), 6.83–6.76 (m, 5H; ArH), 6.26 (d, *J* = 8.0 Hz, 1H; ArH), 5.10 (t, *J* = 7.6 Hz, 1H; H-1'), 4.42–4.40 (m, 1H; H-3'), 4.12–4.07 (m, 1H: H-4'), 3.78 (s, 6H; OCH_3), 3.36–3.26 (m, 2H; H-5'), 2.83 (s, 3H; CH_3), 2.34–2.29 (m, 1H; H-2'), 2.23–2.17 (m, 1H; H-2'); ^{13}C NMR (101 MHz, $CDCl_3$): δ 159.9, 159.1, 158.4, 144.9, 138.2, 136.1, 130.1, 129.2, 127.8, 126.7, 113.1, 109.1, 86.1, 86.1, 80.6, 74.2, 64.6, 55.2, 41.7, 29.2; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{32}H_{35}N_2O_5$, 527.2540; found, 527.2545.

Compound 8d. 69% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.48 (d, *J* = 8.4 Hz, 2H; ArH), 7.43–7.36 (m, 5H; ArH), 7.29–7.25 (m, 2H; ArH), 7.20–7.16 (m, 1H; ArH), 6.86–6.81 (m, 4H; ArH), 6.71 (d, *J* = 7.2 Hz, 1H; ArH), 6.39 (d, *J* = 8.4 Hz, 1H; ArH), 5.13 (t, *J* = 7.4 Hz, 1H; H-1'), 4.48–4.46 (m, 1H; H-3'), 4.15–4.11 (m, 1H: H-4'), 3.78 (s, 6H; OCH_3), 3.40–3.36 (m, 1H; H-5'), 3.28–3.24 (m, 1H; H-5'), 3.00 (s, 6H; CH_3), 2.35–2.28 (m, 2H; H-2'); ^{13}C NMR (101 MHz, $CDCl_3$): δ 159.2, 158.9, 158.4, 144.9, 137.6, 136.2, 130.1, 128.2, 127.8, 126.7, 113.1, 108.2, 104.5, 86.1, 86.0, 81.0, 74.5, 64.7, 55.2, 41.0, 38.9, 37.8; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{33}H_{37}N_2O_5$, 541.2697; found, 541.2715.

Compound 8e. 78% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.26 (s, 1H; imine-H), 7.51–7.48 (m, 3H; ArH), 7.38–7.35 (m, 4H; ArH), 7.26–7.16 (m, 3H; ArH), 7.07 (d, *J* = 7.8 Hz, 1H; ArH), 6.84–6.79 (m, 5H; ArH), 5.18 (t, *J* = 7.6 Hz, 1H; H-1'), 4.42–4.40 (m, 1H; H-3'), 4.11–4.06 (m, 1H: H-4'), 3.77 (s, 6H; OCH_3), 3.34–3.27 (m, 2H; H-5'), 3.02 and 2.76 (2s, 6H; CH_3), 2.33–2.28 (m, 2H; H-2'); ^{13}C NMR (101 MHz, $CDCl_3$): δ 161.4,

159.7, 158.4, 155.6, 145.0, 138.2, 136.1, 130.1, 128.2, 127.8, 126.7, 116.2, 114.5, 113.1, 86.3, 86.1, 81.0, 74.2, 64.5, 55.2, 40.4, 34.7; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{34}H_{38}N_3O_5$, 568.2806; found, 568.2807.

General procedure for phosphoramidation. 2-Cyanoethyldiisopropyl chlorophosphoramidite (1.5 equiv) was added dropwise to a solution of **4** or **8** (1 equiv) and diisopropyethylamine (4 equiv) in CH_2Cl_2 at room temperature. After the reaction had reached completion (30 min), the mixture was concentrated in vacuo and purified by chromatography through a short column of SiO_2 (13-25% EtOAc in CH_2Cl_2) to yield **5** or **9**.

Compound 5a. 88% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.54-7.19 (m, 11H; ArH), 7.02 (d, *J* = 7.2 Hz, 1H; ArH), 6.82-6.79 (m, 4H; ArH), 5.24 (dd, 1H, *J* = 6.4, 9.2 Hz, 1H; H-1'), 4.49 (br s, 1H; H-3'), 4.27 (br s, 1H; H-4'), 3.82-3.78 (m, 1H; OCH_2), 3.78 (s, 6H; OCH_3), 3.66-3.54 (m, 3H; NCH, OCH_2, and H-5'), 3.38-3.23 (m, 2H; NCH and H-5'), 2.56 (t, *J* = 6.4 Hz, 1H; CH_2CN), 2.58-2.45 (m, 1H; H-2'), 2.52 (s, 3H; $ArCH_3$), 2.41 (t, *J* = 6.4 Hz, 1H; CH_2CN), 2.19-2.13 (m, 1H; H-2'), 1.18-1.06 (m, 12H; $CHCH_3$); ^{31}P NMR (162 MHz, $CDCl_3$): δ 148.5, 148.1; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{41}H_{51}N_3O_6P_1$, 712.3510; found, 712.3511.

Compound 5b. 88% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.36 (s, 1H; ArH), 7.48-7.19 (m, 11H; ArH), 6.82-6.79 (m, 4H; ArH), 5.25 (dd, *J* = 5.8, 9.8 Hz, 1H; H-1'), 4.53-4.47 (m, 1H; H-3'), 4.27 (br s, 1H; H-4'), 3.83-3.77 (m, 1H; OCH_2), 3.78 (s, 6H; OCH_3), 3.68-3.55 (m, 3H; NCH, OCH_2, and H-5'), 3.35-3.23 (m, 2H; NCH and H-5'), 2.59 (t, *J* = 6.8 Hz, 1H; CH_2CN), 2.58-2.45 (m, 1H; H-2'), 2.42 (t, *J* = 6.8 Hz, 1H; CH_2CN), 2.31 (s, 3H; $ArCH_3$), 2.20-2.07 (m, 1H; H-2'), 1.18-1.07 (m, 12H; $CHCH_3$); ^{31}P NMR (162 MHz, $CDCl_3$): δ 148.7, 148.2; HRMS (*m/z*): $[M + Na]^+$ calcd for $C_{41}H_{50}N_3O_6P_1Na_1$, 734.3329; found, 734.3334.

Compound 5c. 88% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.39 (d, *J* = 4.8 Hz, 1H; ArH), 7.47-7.19 (m, 10H; ArH), 6.99 (d, *J* = 4.8 Hz, 1H; ArH), 6.82-6.79 (m, 4H; ArH), 5.26 (dd, *J* = 6.0, 9.6 Hz, 1H; H-1'), 4.53-4.48 (m, 1H; H-3'), 4.28 (br s, 1H; H-4'), 3.83-3.77 (m, 1H; OCH_2), 3.78 (s, 3H; OCH_3), 3.77 (s, 3H; OCH_3), 3.68-3.56 (m, 3H; NCH, OCH_2, and H-5'), 3.39-3.24 (m, 2H; NCH and H-5'), 2.61-2.47 (m, 1H; H-2'), 2.59 (t, *J* = 6.4 Hz, 1H; CH_2CN), 2.42 (t, *J* = 6.4 Hz, 1H; CH_2CN), 2.28 (s, 3H; $ArCH_3$), 2.28-2.11 (m, 1H; H-2'), 1.18-1.07 (m, 12H; $CHCH_3$); ^{31}P NMR (162 MHz, $CDCl_3$): δ

148.7, 148.1; HRMS (*m/z*): $[M + Na]^+$ calcd for $C_{41}H_{50}N_3O_6P_1Na_1$, 734.3329; found, 734.3333.

Compound 5d. 91% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.49-7.46 (m, 2H; ArH), 7.38-7.18 (m, 9H; ArH), 6.82-6.78 (m, 4H; ArH), 5.22 (dd, *J* = 5.6, 9.6 Hz, 1H; H-1'), 4.51-4.49 (m, 1H; H-3'), 4.26-4.24 (m, 1H; H-4'), 3.83-3.77 (m, 1H; OCH_2), 3.78 (s, 3H; OCH_3), 3.77 (s, 3H; OCH_3), 3.69-3.54 (m, 3H; NCH, OCH_2, and H-5'), 3.37-3.22 (m, 2H; NCH and H-5'), 2.60-2.52 (m, 1H; H-2'), 2.59 (t, *J* = 6.6 Hz, 1H; CH_2CN), 2.46 (s, 3H; $ArCH_3$), 2.41 (t, *J* = 6.6 Hz, 1H; CH_2CN), 2.25 (s, 3H; $ArCH_3$), 2.22-2.10 (m, 1H; H-2'), 1.18-1.06 (m, 12H; $CHCH_3$); ^{31}P NMR (162 MHz, $CDCl_3$): δ 148.5, 148.1; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{42}H_{53}N_3O_6P_1$, 726.3666; found, 726.3671.

Compound 5e. 86% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.50-7.47 (m, 2H; ArH), 7.39-7.34 (m, 4H; ArH), 7.27-7.19 (m, 4H; ArH), 6.89 (s, 1H; ArH), 6.82-6.79 (m, 4H; ArH), 5.22 (dd, *J* = 5.8, 9.8 Hz, 1H; H-1'), 4.53-4.48 (m, 1H; H-3'), 4.26-4.25 (m, 1H; H-4'), 3.85-3.75 (m, 1H; OCH_2), 3.78 (s, 3H; OCH_3), 3.77 (s, 3H; OCH_3), 3.68-3.54 (m, 3H; NCH, OCH_2, and H-5'), 3.40-3.22 (m, 2H; NCH and H-5'), 2.60 (t, *J* = 6.4 Hz, 1H; CH_2CN), 2.60-2.55 (m, 1H; H-2'), 2.48 (s, 3H; $ArCH_3$), 2.40 (t, *J* = 6.6 Hz, 1H; CH_2CN), 2.22 (s, 3H; $ArCH_3$), 2.22-2.11 (m, 1H; H-2'), 1.18-1.06 (m, 12H; $CHCH_3$); ^{31}P NMR (162 MHz, $CDCl_3$): δ 148.5, 148.0; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{42}H_{53}N_3O_6P_1$, 726.3666; found, 726.3663.

Compound 5f. 90% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.25 (s, 1H; ArH), 7.48-7.20 (m, 10H; ArH), 6.82-6.79 (m, 4H; ArH), 5.22 (dd, *J* = 5.8, 9.8 Hz, 1H; H-1'), 4.54-4.47 (m, 1H; H-3'), 4.26-4.25 (m, 1H; H-4'), 3.83-3.54 (m, 3H; NCH, OCH_2, and H-5'), 3.79 (s, 3H; OCH_3), 3.78 (s, 3H; OCH_3), 3.44-3.21 (m, 2H; NCH and H-5'), 2.61-2.40 (m, 3H; CH_2CN and ; H-2'), 2.22 (s, 3H; $ArCH_3$), 2.22-2.10 (m, 1H; H-2'), 2.19 (s, 3H; $ArCH_3$), 1.18-1.06 (m, 12H; $CHCH_3$); ^{31}P NMR (162 MHz, $CDCl_3$): δ 148.7, 148.1; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{42}H_{53}N_3O_6P_1$, 726.3666; found, 726.3663.

Compound 5g. 83% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.13-8.04 (m, 2H; ArH), 7.81-7.68 (m, 3H; ArH), 7.54-7.19 (m, 10H; ArH), 6.81-6.78 (m, 4H; ArH), 5.46 (dd, *J* = 6.0, 10.0 Hz, 1H; H-1'), 4.58-4.53 (m, 1H; H-3'), 4.34-4.32 (m, 1H; H-4'), 3.85-3.77 (m, 1H; OCH_2), 3.77 (s, 3H; OCH_3), 3.77 (s, 3H; OCH_3), 3.69-3.58 (m, 3H; NCH, OCH_2, and H-5'), 3.44-3.26 (m, 2H; NCH and H-5'), 2.72-2.54 (m, 1H; H-2'), 2.62 (t, *J* = 6.6 Hz, 1H; CH_2CN), 2.45 (t, *J* = 6.6 Hz, 1H; CH_2CN), 2.39-2.26 (m, 1H; H-2'),

1.19-1.08 (m, 12H; CHCH_3); ^{31}P NMR (162 MHz, $CDCl_3$): δ 148.7, 148.3; HRMS (*m/z*): [*M* + H]$^+$ calcd for $C_{44}H_{51}N_3O_6P_1$, 748.3510; found, 748.3502.

Compound 9a. 92% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.59–7.55 (m, 1H; ArH), 7.50–7.47 (m, 2H; ArH), 7.39–7.35 (m, 5H; ArH), 7.30–7.19 (m, 3H; ArH), 7.05 (d, *J* = 7.6 Hz, 1H; ArH), 6.83–6.80 (m, 4H; ArH), 5.27 (dd, *J* = 6.0, 9.6 Hz, 1H; H-1'), 4.54–4.50 (m, 1H; H-3'), 4.29–4.27 (m, 1H: H-4'), 3.80 and 3.79 (2s, 6H; OCH_3), 3.89–3.77 (m, 1H; OCH_2), 3.71–3.55 (m, 3H; OCH_2, NCH, and H-5'), 3.32–3.24 (m, 2H; NCH, and H-5'), 2.71 (q, *J* = 5.0 Hz, 2H; CH_2CH$_3$), 2.62 (t, *J* = 6.6 Hz, 1H; CH_2CN), 2.58–2.42 (m, 2H; CH_2CN and H-2'), 2.25–2.14 (m, 1H; H-2'), 1.30–1.08 (m, 15H; CH_2CH_3 and CHCH_3); ^{31}P NMR (162 MHz, $CDCl_3$): δ 148.4, 148.0; HRMS (*m/z*): [*M* + H]$^+$ calcd for $C_{42}H_{53}N_3O_6P_1$, 726.3666; found, 726.3665.

Compound 9c. 75% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.50–7.17 (m, 10H; ArH), 6.85–6.79 (m, 5H; ArH), 6.26 (d, *J* = 8.0 Hz, 1H; ArH), 5.08 (dd, *J* = 5.8, 9.8 Hz, 1H; H-1'), 4.52–4.50 (m, 1H; H-3'), 4.26–4.24 (m, 1H: H-4'), 3.79 (2s, 6H; OCH_3), 3.86–3.77 (m, 1H; OCH_2), 3.72–3.49 (m, 3H; OCH_2, NCH, and H-5'), 3.32–3.16 (m, 2H; NCH, and H-5'), 2.85 (s, 3H; NCH_3), 2.70 (t, *J* = 6.2 Hz, 1H; CH_2CN), 2.61 (t, *J* = 6.6 Hz, 1H; CH_2CN), 2.48–2.44 (m, 1H; H-2'), 2.27–2.18 (m, 1H; H-2'), 1.37–1.09 (m, 12H; CHCH_3); ^{31}P NMR (162 MHz, $CDCl_3$): δ 148.4, 148.0; HRMS (*m/z*): [*M* + H]$^+$ calcd for $C_{41}H_{52}N_4O_6P_1$, 727.3619; found, 727.3620.

Compound 9d. 89% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.53–7.21 (m, 10H; ArH), 6.87–6.76 (m, 5H; ArH), 6.41 (d, *J* = 8.4 Hz, 1H; ArH), 5.13 (dd, *J* = 5.8, 9.4 Hz, 1H; H-1'), 4.59–4.56 (m, 1H; H-3'), 4.28–4.27 (m, 1H: H-4'), 3.80 and 3.79 (2s, 6H; OCH_3), 3.90–3.28 (m, 6H; OCH_2, NCH, and H-5'), 3.02 (2s, 6H; NCH_3), 2.61 (t, *J* = 6.4 Hz, 1H; CH_2CN), 2.50–2.31 (m, 3H; CH_2CN and H-2'), 1.21–1.01 (m, 12H; CHCH_3); ^{31}P NMR (162 MHz, $CDCl_3$): δ 148.4, 148.0; HRMS (*m/z*): [*M* + H]$^+$ calcd for $C_{42}H_{54}N_4O_6P_1$, 741.3775; found, 741.3769.

Compound 9e. 71% yield. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.27 (s, 1H; imine-H), 7.52–7.49 (m, 3H; ArH), 7.41–7.36 (m, 4H; ArH), 7.25–7.18 (m, 3H; ArH), 7.11 (d, *J* = 8.0 Hz, 1H; ArH), 6.85–6.78 (m, 5H; ArH), 5.18–5.16 (m, 1H; H-1'), 4.53–4.51 (m, 1H; H-3'), 4.20–4.12 (m, 1H: H-4'), 3.77 (s, 6H; OCH_3), 3.83–3.77 (m, 1H; OCH_2), 3.69–3.47 (m, 3H; OCH_2, NCH, and H-5'), 3.40–3.16 (m, 2H; NCH, and H-5'), 3.02 and 2.72 (2s, 6H; NCH_3), 2.77–2.23 (m, 4H; CH_2CN and H-2'), 1.36–1.05 (m, 12H; CHCH_3); ^{31}P

NMR (162 MHz, $CDCl_3$): δ 148.2, 147.9; HRMS (*m/z*): $[M + H]^+$ calcd for $C_{43}H_{55}N_5O_6P_1$, 768.3884; found, 768.3892.

General procedure for triphosphate synthesis. Proton sponge (1.5 equiv) and the free nucleoside **3** or **7** (1 equiv) were dissolved in trimethyl phosphate (0.3 M) and cooled to -10 ºC. $POCl_3$ (1.5 equiv) was added dropwise, and the purple slurry was stirred at -10 ºC for 2 h. Tri-*n*-butylamine (6.2 equiv) was added, followed by a solution of tributylammonium pyrophosphate (5.0 equiv) in DMF (0.5 M). After 5 min, the reaction was quenched by addition of 0.5 M aqueous $Et_3NH_2CO_3$ (20 vol. equiv). The resulting solution lyophilized. Purification by reverse-phase (C18) HPLC (4-35% CH_3CN in 0.1 M $Et_3NH_2CO_3$, pH 7.5) followed by lyophilization afforded **6** or **10** as a white solid. The amidine-protected triphosphates **10e** was dissolved in 1.5 mL of NH_4OH and stirred at room temperature for 15 h. The solution was diluted by addition of 3 mL of aqueous $Et_3NH_2CO_3$ and purified by reverse-phase (C18) HPLC.

Compond 6a. ^{31}P NMR (162 MHz, D_2O): δ -8.84 (d, *J* = 20.6 Hz; γ-P), -10.71 (d, *J* = 19.9 Hz; α-P), -22.63 (dd, *J* = 20.6, 19.9 Hz; β-P).

Compond 6b. ^{31}P NMR (162 MHz, D_2O): δ -5.90 (d, *J* = 21.1 Hz; γ-P), -10.56 (d, *J* = 19.8 Hz; α-P), -22.10 (dd, *J* = 21.1, 19.8 Hz; β-P).

Compond 6c. ^{31}P NMR (162 MHz, D_2O): δ -9.15 (d, *J* = 20.3 Hz; γ-P), -10.65 (d, *J* = 20.3 Hz; α-P), -22.61 (t, *J* = 20.3 Hz; β-P).

Compond 6d. ^{31}P NMR (162 MHz, D_2O): δ -7.98 (d, *J* = 20.6 Hz; γ-P), -10.65 (d, *J* = 19.9 Hz; α-P), -22.47 (dd, *J* = 20.6, 19.9 Hz; β-P).

Compond 6e. ^{31}P NMR (162 MHz, D_2O): δ -8.84 (d, *J* = 18.8 Hz; γ-P), -10.65 (d, *J* = 20.9 Hz; α-P), -22.61 (dd, *J* = 20.9, 18.8 Hz; β-P).

Compond 6f. ^{31}P NMR (162 MHz, D_2O): δ -6.10 (d, *J* = 20.1 Hz; γ-P), -10.58 (d, *J* = 19.8 Hz; α-P), -22.12 (dd, *J* = 20.1, 19.8 Hz; β-P).

Compond 6g. ^{31}P NMR (162 MHz, D_2O): δ -6.80 (d, *J* = 20.0 Hz; γ-P), -10.61 (d, *J* = 20.0 Hz; α-P), -22.24 (t, *J* = 20.0 Hz; β-P).

Compound 10a. ^{31}P NMR (162 MHz, D_2O): δ -6.25 – -6.37 (m; γ-P), -12.60 (d, *J* = 18.0 Hz; α-P), -24.18 – -24.57 (m; β-P).

Compound 10b. ^{31}P NMR (162 MHz, D_2O): δ -6.19 – -6.30 (m; γ-P), -11.76 – -11.91 (m; α-P), -22.31 – -22.49 (m; β-P).

Compound 10c. ^{31}P NMR (162 MHz, D_2O): δ -10.21 – -10.52 (m; γ-P), -11.00 (d, *J* = 18.5 Hz; α-P), -22.86 (t, *J* = 19.8 Hz; β-P).

Compound 10d. ^{31}P NMR (162 MHz, D_2O): δ -6.37 (d, *J* = 20.9 Hz; γ-P), -10.57 (d, *J* = 19.8 Hz; α-P), -22.11 (t, *J* = 20.7 Hz; β-P).

Synthesis of Oligonucleotides. Oligonucleotides were prepared by the β-cyanoethylphosphoramidite method on controlled pore glass supports (1 µmol) using an Applied Biosystems Inc. 392 DNA/RNA synthesizer as standard method. After automated synthesis, the oligonucleotides were cleaved from the support by concentrated aqueous ammonia for 1 h at room temperature, deprotected by heating at 56 °C for 10 h, and purified by denaturing polyacrylamide gel electrophoresis (12-15%, 8 M urea). The primer oligonucleotides containing unnatural bases at the 3'-end were obtained using Universal Support II or 3'-phosphate CPG, which was treated with alkaline phosphatase after deprotection. The oligonucleotides were purified by PAGE, visualized by UV shadowing and recovered by electroelution. After ethanol precipitation, the concentration of oligonucleotides was determined by UV/vis absorption.

Table S1: Misincorporation rates of natural triphosphates dNTP.[a]

5'-dTAATACGACTCACTATAGGGAGA 3'-dATTATGCTGAGTGATATCCCTCT (**Y**) GCTAGGTTACGGCAGGATCGC				
N	**Y**	k_{cat} [min^{-1}]	K_M [μM]	k_{cat}/K_M [M^{-1} min^{-1}]
dA	d**3MPy**	2.9 ± 0.4	37 ± 5	7.9×10^4
dC	d**3MPy**	nd[b]	nd[b]	$<1.0 \times 10^3$
dG	d**3MPy**	1.7 ± 0.3	125 ± 13	1.4×10^4
dT	d**3MPy**	0.75 ± 0.20	133 ± 13	5.7×10^3
dA	d**4MPy**	2.6 ± 0.4	32 ± 1	8.1×10^4
dC	d**4MPy**	nd[b]	nd[b]	$<1.0 \times 10^3$
dG	d**4MPy**	0.43 ± 0.07	64 ± 4	6.8×10^3
dT	d**4MPy**	0.39 ± 0.12	118 ± 15	3.3×10^3
dA	d**5MPy**	9.0 ± 1.5	36 ± 7	2.5×10^5
dC	d**5MPy**	nd[b]	nd[b]	$<1.0 \times 10^3$
dG	d**5MPy**	0.47 ± 0.04	68 ± 6	6.9×10^3
dT	d**5MPy**	0.39 ± 0.08	142 ± 19	2.8×10^3
dA	d**34DMPy**	0.93 ± 0.17	22 ± 0.3	4.2×10^4
dC	d**34DMPy**	0.36 ± 0.10	217 ± 18	1.6×10^3
dG	d**34DMPy**	0.88 ± 0.03	65 ± 0.4	1.4×10^4
dT	d**34DMPy**	1.7 ± 0.5	124 ± 21	1.4×10^4
dA	d**35DMPy**	1.2 ± 0.1	21 ± 3	5.7×10^4
dC	d**35DMPy**	nd[b]	nd[b]	$<1.0 \times 10^3$
dG	d**35DMPy**	1.3 ± 0.2	87 ± 4	1.5×10^4
dT	d**35DMPy**	0.96 ± 0.21	122 ± 3	7.9×10^3
dA	d**45DMPy**	2.3 ± 0.2	31 ± 2	7.4×10^4
dC	d**45DMPy**	0.63 ± 0.14	281 ± 16	2.2×10^3
dG	d**45DMPy**	0.35 ± 0.03	45 ± 8	7.8×10^3
dT	d**45DMPy**	0.79 ± 0.08	167 ± 14	4.7×10^3
dA	d**QL**	0.47 ± 0.04	25 ± 2	1.9×10^4
dC	d**QL**	0.27 ± 0.09	133 ± 10	2.0×10^3
dG	d**QL**	0.31 ± 0.07	69 ± 13	4.5×10^3
dT	d**QL**	0.31 ± 0.03	119 ± 7	2.6×10^3
dA	d**EPy**	1.2 ± 0.3	57 ± 8	2.0×10^4
dC	d**EPy**	nd[b]	nd[b]	$<1.0 \times 10^3$
dG	d**EPy**	1.3 ± 0.1	174 ± 40	7.5×10^3
dT	d**EPy**	1.0 ± 0.2	266 ± 46	3.8×10^3
dA	d**APy**	2.5 ± 0.4	36 ± 7	6.9×10^4
dC	d**APy**	nd[b]	nd[b]	$<1.0 \times 10^3$
dG	d**APy**	0.44 ± 0.06	103 ± 15	4.3×10^3
dT	d**APy**	0.13 ± 0.03	110 ± 24	1.2×10^3
dA	d**MAPy**	2.1 ± 0.9	85 ± 7	2.5×10^4
dC	d**MAPy**	nd[b]	nd[b]	$<1.0 \times 10^3$
dG	d**MAPy**	0.10 ± 0.02	86 ± 18	1.1×10^3
dT	d**MAPy**	0.25 ± 0.07	223 ± 48	1.1×10^3
dA	d**DMAPy**	0.92 ± 0.07	62 ± 19	1.5×10^4
dC	d**DMAPy**	nd[b]	nd[b]	$<1.0 \times 10^3$
dG	d**DMAPy**	nd[b]	nd[b]	$<1.0 \times 10^3$
dT	d**DMAPy**	1.2 ± 0.2	264 ± 56	4.7×10^3

[a] See 'Steady-state kinetics' for details. [b] Reaction was too inefficient for k_{cat} and K_M to be determined independently.

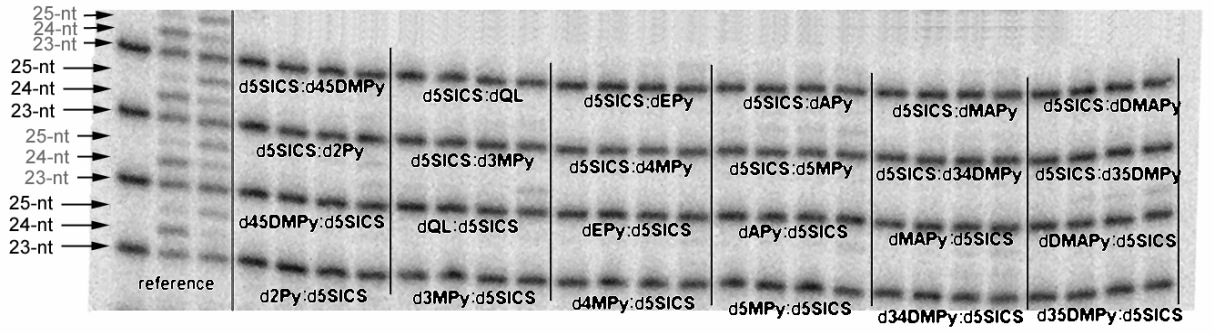


Figure S1. Heteropair synthesis and extension screen between pyridyl analogs and d**5SICS** (d**X**:d**Y**, primer:template). Assay conditions were as follows: 40 nM DNA duplex, 0.6 nM Kf, and 5, 20, 100, and 500 µM d**X**TP + dCTP (from left to right). Reference is natural dA:dT pair with none, 100 µM dATP, and 100 µM dATP + dCTP (from left to right). All Reactions were incubated at 25 ºC for 5 min. See heteropair synthesis/extension described above for details.

Table S2: Rates of mispair extension.[a]

5'-dTAATACGACTCACTATAGGGAGA (**X**) 3'-dATTATGCTGAGTGATATCCCTCT (**Y**) GCTAGGTTACGGCAGGATCGC				
X	**Y**	k_{cat} [min^{-1}]	K_M [µM]	k_{cat}/K_M [M^{-1} min^{-1}]
dA	d**45DMPy**	4.9 ± 0.4	42 ± 2	1.2×10^5
dC	d**45DMPy**	0.34 ± 0.05	89 ± 9	3.9×10^3
dG	d**45DMPy**	0.26 ± 0.03	127 ± 5	2.1×10^3
dT	d**45DMPy**	11 ± 1	17 ± 2	6.4×10^5

[a] See description of steady-state kinetics, above for details.

References.

[1] M. Urban, R. Pohl, B. Klepetárová, M. Hocek, *J. Org. Chem.* **2006**, *71*, 7322- 7328.

[2] G. J. Bridger, R. T. Skerlj, S. Padmanabhan, S. A. Martellucci, G. W. Henson, M. J. Abrams, H. C. Joao, M. Witvrouw, K. De Vreese, R. Pauwels, E. De Clercq, *J. Med. Chem.* **1996**, *39*, 109- 119.

[3] T. Kaminski, P. Gros, Y. Fort, *Eur. J. Org, Chem.* **2003**, 3855- 3860.

[4] O. Sugimoto, M. Mori, K. Moriya, K. Tanji, *Helv. Chim. Acta* **2001**, *84*, 1112- 1118.

[5] W. Wierenga, H. I. Skulnick, *Carbohydrate Res.* **1981**, *90*, 41- 52.