SUPPORTING INFORMATION

Size-expanded Analogues of dG and dC: Synthesis and Pairing Properties in DNA

Haibo Liu, Jianmin Gao, Eric T. Kool*

Department of Chemistry, Stanford University, Stanford CA  94305-5080

*to whom correspondence should be addressed:  kool@stanford.edu

CONTENTS:

General synthesis procedures…………………………………………………… p. S2
Nucleoside phosphoramidite synthesis and characterization……………… p. S2
UV-vis absorption and fluorescence………………………………………… p. S8
Heats of formation preliminary calculations…………………………… p. S8
Melting curves………………………………………………………………… p. S9
Oligonucleotide characterization by NMR………………………………… p. S10
NMR spectra………………………………………………………………… p. S12
General Synthesis Procedures

Reagents were purchased and were used without further purification. All water sensitive reactions were carried out in oven-dried glassware with a Teflon-coated stirring bar under a nitrogen or argon atmosphere. Anhydrous solvents pyridine, CH₃CN, CH₂Cl₂, and THF, were distilled under a nitrogen atmosphere. THF was dried over Na₂O₂, and the other three solvents were dried over CaH₂. All other anhydrous solvents were purchased and used directly. All ¹H, ¹³C and ³¹P NMR spectra were recorded on 400 MHz or 500 MHz instruments as noted. ³¹P NMR spectra were referenced to 85% H₃PO₄ in water as external standard (0.0 ppm). 2D-COSY and 2D-ROESY spectra were acquired on a 500 MHz instrument.

Nucleoside Synthesis and Characterization for isobutyryl dxG (19). Details are given in the main text. See below for ¹H-NMR spectra of intermediates. Subsequent steps for preparation of 5'-DMT-3’-O-phosphoramidite derivative are described here:

[N⁶-isobutyryl-6-amino-3,7-dihydro-imidazo[4,5-g]quinazolin-8-one]-1’-β-D-2’-deoxyribofuranoside-5’-O-(4,4’-dimethoxytrityl)ether (20). Starting material 19 (102 mg, 0.26 mmol) was dissolved in 7.0 mL anhydrous pyridine and 0.5 mL DIPEA. DMT-Cl (611 mg) was added and the solution was stirred overnight. After quenching with MeOH, the reaction mixture was concentrated, redissolved in EtOAc (80 mL) and washed with saturated NaHCO₃ and brine (40 mL each). The organic layer was dried over NaSO₄, concentrated and purified by silica gel column chromatography (0-5% MeOH in CH₂Cl₂). A light yellow foamy solid was obtained as product in 93% yield (170 mg). ¹H NMR (CDCl₃ 500MHz): δ 8.74(s, 1H) 8.21(s, 1H) 7.52(s, 1H) 7.38-7.36(m, 2H) 7.27-7.17(m, 7H) 6.78-6.75(m, 4H) 6.36(t, 1H, J=8.5Hz) 4.64-4.62(m, 1H) 4.21-4.20(m, 1H) 3.73(s, 6H) 3.41-3.38(m, 1H) 3.33-3.30(m, 1H) 2.67-2.62(m, 1H) 2.58-2.57(m, 1H) 1.27(d, 6H, J=6.0Hz) ¹³C NMR (CDCl₃ 125MHz): δ 179.3 170.4 162.5 158.75 145.5 144.7 144.1 142.4 138.8 137.8 135.7 130.3 128.3 128.2 127.2 119.2 116.6 113.5 106.3 86.8 86.6 85.6 72.7 64.1 55.5 40.7 36.7 34.8 30.0 23.6 19.3 19.3 15.1; HRMS calcd for C₃⁹H₄₀N₅O₇ (M+H) 690.2922; found 690.2900.
Starting DMT ether 20 (174 mg, 0.25mmol) was dissolved in 15.0 mL anhydrous CH$_2$Cl$_2$ and 1.5 mL DIPEA, to which was added 303 µL 2-cyanoethyl diisopropylchlorophosphoramidite. After the reaction was stirred at room temperature for 72 h, it was purified by silica gel column chromatography (50-100% EtOAc in hexanes). 173 mg (77%) of the product was obtained as a white foamy solid. $^1$H NMR (CDCl$_3$ 500MHz): δ 8.86(s, 1H) 8.71(s, 1H) 8.22(s, 1H) 7.56(s, 1H) 7.38(m, 2H) 7.28-7.15(m, 7H) 6.79-6.75(m, 4H) 6.33(t, 1H, J=8.5Hz) 4.72(br, 1H) 4.33(d, br, 1H) 4.14-4.06(m, 1H) 3.75(s, 6H) 3.65-3.57(m, 2H) 3.38-3.33(m, 2H) 2.70-2.59(m, 4H) 2.47(t, 1H, J=8.5Hz) 1.28-1.13(m, 18H); $^{13}$C NMR (CDCl$_3$ 125MHz): δ 178.1 162.4 158.8 145.3 144.7 144.0 144.0 142.77 138.7 135.8 135.7 135.7 130.4 130.3 128.4 128.2 128.2 127.2 127.2 119.3 117.9 117.8 116.6 113.5 106.4 86.8 86.0 85.6 74.3 73.8 73.7 63.7 63.5 58.6 58.6 58.5 58.5 54.5 43.5 40.0 36.8 24.8 20.7 20.5 19.3 $^{31}$P NMR (CDCl3 202MHz) δ 146.5; HRMS calcd for C$_{48}$H$_{57}$N$_7$O$_8$P (M+H) 890.4001; found 890.3994.

**Confirmation of anomeric geometry for dxG derivative 17.** The isolated regioisomers of 17 were identified through extensive NMR experiments to determine their structures. First, HMBC was used to identify the Ha and Hb in the two isomers (Figure S1). Ha has three carbons that are three bonds away from it (labeled “#”), whereas Hb only has two (labeled “*”). This difference led to the easy identification of Ha and Hb as seen in Figure S2. The coupling between the carbonyl carbon (163 ppm) and the proton at 8.70 ppm is clearly visible and therefore, Ha was assigned at 8.70 ppm and Hb at 7.42 ppm. With this information in hand, we were then able to identify the two regioisomers based upon NOE measurements (below).
**Figure S1.** The structure of desired regioisomer of dxG and illustration of an expected NOE.

**Figure S2.** HMBC spectrum of 17.
A 2D-NOESY experiment was performed to see whether Ha or Hb exhibited NOE with H1’. As seen in Figure S3, there is a stronger NOE between Hb (δ 7.42 ppm) and H1’ than between Ha (δ 8.70 ppm) and H1’. This observation allowed for the conclusive assignment of 17 as the structure of the isolated regioisomer. Further confirmation was obtained by the observation of an NOE between Hb and H2’. NOEs between the imidazole proton (8.29 ppm) and sugar 1’, 2’, 3’, and 5’ (weakest) were also observed. Due to the distance between Ha and the sugar, no NOEs were observed. A 2D-NOESY experiment was also carried out for the other regioisomer of 17, and an NOE between Ha and H1’ was observed, while no NOE between Hb and H1’ was seen.

Figure S3. 2D-NOESY spectrum of one regioisomer of 17.
Nucleoside Synthesis and Characterization for dxC. Details are given in the main text. See below for 1H-NMR spectra of intermediates. Subsequent steps for preparation of 5'-O-DMT-3'-O-phosphoramidite derivative are described here.

1'-β-[8-(4-N^4-(N,N-dimethylformamidino)-amino-6-methyl-2-quinazolone)]-2'-D-deoxyribofuranose (6). The free xC-nucleoside 1 (76 mg, 0.26 mmol) was dissolved in 5 mL dry pyridine. To the solution was added N,N-dimethylacetamide dimethyl acetal (500 μL, 3.75 mmol) in one portion. The reaction mixture was stirred at room temperature for 5 h. Volatiles were removed in vacuo and the residue was purified by silica column chromatography (CH$_2$Cl$_2$:CH$_3$OH 10:1 followed by CH$_2$Cl$_2$:CH$_3$OH 5:1) to give the product 6 as white solid (74 mg, 82%). $^1$H-NMR (CD$_3$OD, 400 MHz), δ 8.74 (s, 1H), 8.06 (s, 1H), 7.40 (s, 1H), 5.31 (dd, 1H, J=11 Hz, 5 Hz), 4.46 (m, 1H), 4.05 (m, 1H), 3.83 (m, 2H), 3.28 (s, 6H), 2.36 (s, 3H), 2.22 (m, 1H), 2.14 (m, 1H); $^{13}$C-NMR (CD$_3$OD, 100.6 MHz), δ 170.9, 158.7, 138.1, 134.7, 131.7, 126.7, 125.2, 115.7, 88.8, 80.8, 73.3, 61.9, 42.2, 40.7, 34.5, 19.7; HRMS (EI+) calcd for C$_{17}$H$_{23}$N$_4$O$_4$ [M+H]$^+$ m/z = 347.1719, found 347.1704.

1'-β-[8-(4-N^4-(N,N-dimethylformamidino)-amino-6-methyl-2-quinazolone)]-5'-O-(4,4'-dimethoxytrityl)-2'-deoxy-D-ribofuranose (7). The base-protected nucleoside 6 (63 mg, 0.18 mmol) was coevaporated with 3x5 mL dry pyridine and the residue dissolved in 5 mL dry pyridine. N,N-diisopropylethylamine (188 μL, 6 equiv) and 4,4'-dimethoxytrityl chloride (244 mg, 0.72 mmol) were added to the solution. The reaction mixture was stirred at room temperature for 2 h. Volatiles were removed under vacuum and crude product was purified by silica column chromatography. Product 7 was obtained as white foam (92 mg, 78%). $^1$H-NMR (CDCl$_3$, 500 MHz), δ 11.47 (s, 1H), 8.94 (s, 1H), 8.09 (s, 1H), 7.86 (s, 1H), 7.52 (d, 2H, J=8.5 Hz), 7.40 (m, 4H), 7.29 (m, 2H), 7.22 (t, 1H, J=10 Hz), 6.84 (m, 4H), 5.87 (dd, 1H, J=11.5 Hz, 3.5 Hz), 4.49 (d, 1H, J=5 Hz), 4.40 (m, 1H), 3.81 (s, 6H), 3.44 (m, 1H), 3.34 (m, 1H), 3.31 (s, 3H), 3.28 (s, 3H), 2.86 (m, 1H), 2.40 (s, 3H), 1.94 (s, 1H); $^{13}$C-NMR (CDCl$_3$, 125.7 MHz), δ 170.5, 159.5, 158.6, 158.4, 145.0, 137.4, 136.3, 136.2, 131.7, 131.4, 130.2, 130.1, 128.3, 127.8, 127.6, 126.7, 125.5, 115.1, 113.1, 87.9, 86.0, 75.2, 64.4, 55.2, 42.5, 41.8, 35.5, 21.2; HRMS (EI+) calcd for C$_{38}$H$_{41}$N$_4$O$_6$, [M+H]$^+$ m/z = 649.3026, found 649.3012.
1’-β-[8-(4-N⁴-(N,N-dimethylformamidino)-amino-6-methyl-2-quinazolone)]-5’-O-(4,4’-dimethoxytrityl)-2’-deoxy-D-ribofuranose-3’-O-(2-cyanoethyl-diisopropylamino-phosphoramidite) (8). Compound 7 (100 mg, 0.15 mmol) was dissolved in 5 mL dry CH₂Cl₂. To the solution were added N,N-diisopropylethylamine (52 µL, 0.30 mmol) and 2-cyanoethyl tetraisopropylphosphoramidite (52 µL, 0.23 mmol). The reaction mixture was stirred at room temperature for 2 h. Volatiles were removed in vacuo and the residue was purified by silica column chromatography (ethyl acetate followed by ethyl acetate: methanol 10:1) to give product 8 (a mixture of two diastereomers) as a white foamy solid (112 mg, 86%). ¹H-NMR (CDCl₃, 400 MHz), δ 9.48 (d, 2H), 8.94 (m, 2H), 8.06 (m, 2H), 7.36-7.12 (m, 20H), 6.72 (m, 8H), 5.38 (m, 2H), 4.54 (m, 2H), 4.27 (m, 2H), 3.73 (s, 6H), 3.72 (s, 6H), 3.62-3.51 (m, 4H), 3.27-3.13 (m, 4H), 3.24 (s, 6H), 3.19 (s, 6H), 2.62 (t, 4H, J= 6 Hz), 2.48 (t, 4H, J=6 Hz), 2.36 (s, 6H), 1.25-1.10 (m, 28H); ¹³C-NMR (CDCl₃, 100.6 MHz), δ 170.1, 158.4, 158.3, 157.0, 144.6, 139.2, 139.1, 135.8, 135.8, 135.6, 132.2, 132.1, 130.6, 130.6, 129.9, 128.0, 128.0, 127.6, 126.6, 126.4, 126.30, 124.9, 117.6, 117.4, 115.1, 115.0, 112.9, 86.4, 86.2, 86.0, 77.2, 75.6, 75.4, 75.0, 74.8, 63.7, 58.3, 58.2, 58.1, 58.0, 55.1, 43.2, 43.0, 41.4, 39.6, 35.2, 24.5, 24.5, 24.4, 20.9; HRMS (EI+) calcd for C₄₇H₅₈N₆O₇P [M+H]⁺ m/z = 849.4105, found 849.4056.
**UV-vis absorption and fluorescence studies.** Absorption spectra were recorded on a UV-Vis spectrometer. Steady-state fluorescence measurements were carried out on a fluorescence spectrometer equipped with temperature controller. Quantum yields ($\Phi_f$) were calculated with Fluorescein in 0.1N NaOH solution as a reference (Demas, J. N.; Crosby, G. A. *J. Phys. Chem.* **1971**, *75*, 991-1024).

![Chemical structures](image_url)

**Figure S4.** Stabilities of tautomers of xC and xG free bases as calculated from relative heats of formation (AM1 calculations using Spartan (Wavefunction Inc.)). Numbers were normalized to the most stable tautomer.
Figure S5. Examples of melting curves of duplexes from Table 1 (main text) containing variable base pairs/mismatches as shown in the figure. Curves were measured on a UV-vis spectrometer monitored at 260 nm in a buffer containing NaCl (100 mM), MgCl₂ (10 mM), and Na-PIPES (10 mM), at pH=7.0. Total DNA concentration was 5.0 µM for each measurement.
Figure S5. $^1$H-NMR spectrum (D$_2$O) of trimer oligodeoxynucleotide having the sequence d(T-xG-T).
Figure S6. $^1$H-NMR spectrum (D$_2$O) of trimer oligodeoxynucleotide having the sequence d(T-xC-T).
Pulse Sequence: z2pol
Solvent: CDCl3
Ambient temperature
File: 4-thio-xT-H
INOVA-500 "ui500"

Pulse 60.0 degrees
Acq. time 4.000 sec
Width 8000.0 Hz
4 repetitions

OBSERVE H1, 499.7495000 MHz
DATA PROCESSING
FT size 131072
Total time 6 min, 40 sec

AcO

NH

OAc

S

3 S

5
STANDARD PROTON PARAMETERS

Pulse Sequence: zgpol
Solvent: CDCl3
Ambient temperature
File: dmt-xC-H
INOVA-500 "ui500"

Pulse 60.0 degrees
Acq. time 4.000 sec
Width 8000.0 Hz
32 repetitions

OBSERVE 31, 499.7485607 MHz
DATA PROCESSING
FT size 131072
Total time 1 hr, 6 min, 48 sec

12 11 10 9 8 7 6 5 4 3 2 1 -0 ppm

0.84 9 0.94 2.13.01 4.21 1.39 1.03 1.11 1.06 1.05
1.01 1.01 4.15 4.21 1.39 1.03 6.38 7.58 3.24 2.44
STANDARD PROTON PARAMETERS

Pulse Sequence: s2pul
Solvent: CDC13
Ambient temperature
File: 1H_r_diSixOiBu
INOMA-500 “nmet”

Pulse 60.0 degrees
Acq. time 4.000 sec
Width 8000.0 Hz
64 repetitions
OBSERVE H1, 499.7485657 MHz
DATA PROCESSING
FT size 131072
Total time 4 min, 16 sec
STANDARD PROTON PARAMETERS

Pulse Sequence: s2pul
Solvent: CDCl3
INOVA-500 "nmri"

Pulse 60.0 degrees
Acq. time 4.000 sec
Width 8000.0 Hz
16 repetitions
OBSERVE H1, 499.7485607 MHz
DATA PROCESSING
FT size 131072
Total time 1 min
STANDARD PROTON PARAMETERS

Pulse Sequence: s2pul
Solvent: CDCl3
INOVA-500 "nrm1"

Pulse 60.0 degrees
Acq. time 4.000 sec
Width 8000.0 Hz
28 repetitions
OBSERVE H1, 499.7485607 MHz
DATA PROCESSING
FT size 131072
Total time 4 min
STANDARD PROTON PARAMETERS

Pulse Sequence: s2pul
Solvent: DMSO
INOVA-500 "nmri"

Pulse 60.0 degrees
Acq. time 4.000 sec
Width 8000.0 Hz
32 repetitions
OBSERVE H1, 499.7509370 MHz
DATA PROCESSING
FT size 131072
Total time 4 min
STANDARD 1H OBSERVE

Pulse Sequence: s2pul
Solvent: CDC13
INOVA-500 "nmr1"

Relax. delay 0.500 sec
Pulse 51.4 degrees
Acq. time 4.002 sec
Width 5997.5 Hz
16 repetitions
OBSERVE H1, 400.115375 MHz
DATA PROCESSING
FT size 65536
Total time 1 min