JOC Article

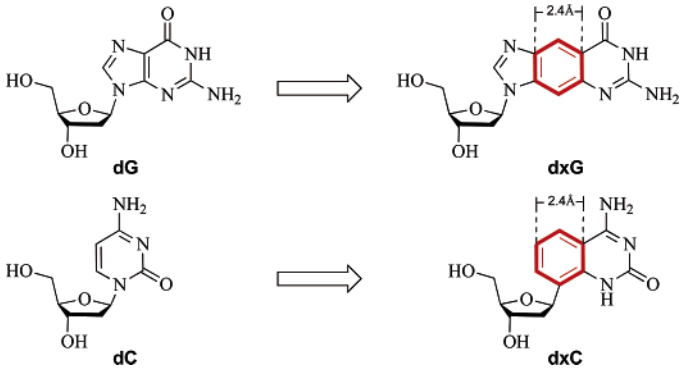
Size-Expanded Analogues of dG and dC: Synthesis and Pairing Properties in DNA

Haibo Liu, Jianmin Gao, and Eric T. Kool*

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

kool@stanford.edu

Received September 16, 2004



We describe the completion of the set of four benzo-fused expanded DNA (xDNA) nucleoside analogues. We previously reported the development of benzo-fused analogues of dA and dT and their inclusion in an exceptionally stable new four-base genetic system, termed xDNA, in which the base pairs were expanded in size. Here we describe the preparation and properties of the second half of this nucleotide set: namely, the previously unknown dxC and dxG nucleosides. The dxC analogue was prepared from the previously reported dxT nucleoside in three steps and 57% yield. The large-sized deoxyguanosine analogue was prepared from an intermediate in the synthesis of dxA, yielding dxG in 14 steps overall (2.4%). Suitably protected versions of the deoxynucleosides were prepared for oligonucleotide synthesis following standard procedures, and they were readily incorporated into DNA by automated synthesizer. "Dangling-end" measurements revealed that the benzo-fused homologues stack considerably more strongly on neighboring DNA sequences than do their natural counterparts. Base pairing experiments with xC or xG bases showed that they pair selectively with their Watson−Crick partners, but with mild destabilization, due apparently to their larger size. Overall, the data suggest that the fluorescent xG and xC bases may be useful probes of steric effects in the study of biological nucleotide recognition mechanisms. In addition, the completion of the xDNA nucleoside set makes it possible in the future to construct full four-base xDNA strands that can target any sequence of natural DNA and RNA.

Introduction

The biological recognition and processing of nucleotides requires efficient and selective binding of nucleobases. Differential recognition of GTP and ATP, for example, regulates metabolic energy usage in cells and switching and signaling in biochemical pathways.[1] Similarly, the four natural deoxynucleotides—dATP, dCTP, dTTP, and dGTP—must be distinguished from one another with high selectivity during replication of the cellular genome.[2] Of course, recognition of chemical differences in the bases themselves is paramount in such processes, and thus, steric effects, solvation effects, and hydrogen-bonding effects could all possibly play roles in this selectivity.[3] We have previously explored the importance of hydrogen bonding and solvation effects by the molecular strategy

(1) Alaimo, P. J.; Shogren-Knaak, M. A.; Shokat, K. M. *Curr. Opin. Chem. Biol.* **2001**, *5*, 360−367.

(2) (a) Goodman, M. F.; Creighton, S.; Bloom, L. B.; Petruska, J. *Crit Rev. Biochem. Mol. Biol.* **1993**, *28*, 83−126. (b) Kunkel, T. A.; Bebenek, K. *Annu. Rev. Biochem.* **2000**, *69*, 497−529. (c) Kool, E. T. *Annu. Rev. Biochem.* **2002**, *71*, 191−219.

(3) Kool, E. T. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 1−22.

10.1021/jo048357z CCC: $30.25
Published on Web 12/23/2004

of using nonpolar isosteres of natural nucleosides,[4] and with them have gained useful insights into mechanisms of DNA replication and DNA repair enzymes.[5,6] However, a complementary approach to study of such biomolecular recognition would be to directly test steric effects by maintaining hydrogen bonding groups, but changing the sizes of the nucleobases. Here, we describe the completion of a full set of nucleoside analogues for such an approach.

We recently reported the synthesis of two nucleosides, dxT and dxA, in which the nucleobases of dT and dA are expanded in size by benzo-homologation.[7] This is a strategy adapted from early work by Leonard, who described ATP analogues having *lin*-benzoadenine as a base.[8] Analogous to the addition of a benzene ring to the framework of the purine base adenine, we designed a size-expanded analogue of the pyrimidine T by use of the bicyclic quinazolinedione skeleton. Studies demonstrated that the two analogues could be successfully incorporated into DNA oligonucleotides, and their pairing properties in natural DNA were explored.[9] Finally, we showed that entire duplexes consisting of size-expanded xA−T and xT−A pairs could be formed and that these large-diameter "xDNA" duplexes were more stable than natural DNAs of the analogous sequence.[10]

However, that early work left a number of problems unsolved. First, for universal application of expanded nucleotide analogues as biochemical tools, one would need variants of all four bases, while the initial studies described only two of them, leaving C and G analogues unexplored. Second, the high thermodynamic stability of xDNA duplexes suggests the possibility of four-base xDNA oligomers as efficient targeting agents for natural sequences of DNA and RNA. However, in previous studies only two of the four xDNA nucleosides were reported, making general sequence recognition impossible. Finally, the development of xDNA analogues of C and G would make it conceptually possible to form a novel eight-base genetic system using combinations of four natural and four expanded bases. Thus, the current study addresses these problems by the preparation and study of dxC and dxG nucleosides.

Experimental Section

Preparation of dxG Nucleoside.

6-Amino-3,7-dihydroimidazo[4,5-*g*]quinazolin-8-one (12). A mixture of 3-acetyl-3,5-dihydroimidazo[4,5-*g*]quinazoline-6,8-dione and 1-acetyl-1,5-dihydroimidazo[4,5-*g*]quinazoline-6,8-dione[11] (6.974 g) and Na_2CO_3 (6.59 g) was suspended in 100 mL of CH_3CN and 25 mL of water and heated at reflux 3 h. After being cooled to room temperature, the reaction mixture was evaporated and the residue was suspended in 100 mL of water. After the pH was adjusted to 7.0, the mixture was stored at 4 °C overnight. After filtration, washing with water, and drying, 3.093 g of product (54%) was obtained: ^{1}H NMR (DMSO-d_6, 500 MHz) δ 8.20 (s, 1H), 8.16 (s 1H), 7.31 (s, 1H); HRMS calcd for $C_9H_8N_5O$ (M + H) 202.0729, found 202.0737.

***N*^{6}-Isobutyryl-6-amino-3,7-dihydroimidazo[4,5-*g*]-quinazolin-8-one (13).** To a suspension of 2.732 g of 6-amino-3,7-dihydroimidazo[4,5-*g*]quinazolin-8-one **12** in 45 mL of anhydrous pyridine was added 5.8 mL of isobutyryl chloride. After being heated at reflux for 4.5 h, the reaction mixture was cooled to room temperature and evaporated. Fifty milliliters of ethanol was added and the solution heated at reflux for 1.5 h. After ethanol was removed by rotary evaporation, 100 mL of water was added and the pH was adjusted to 7.0. After storage at 4 °C overnight, 3.014 g of product (82%) was obtained after filtration and drying: ^{1}H NMR (DMSO-d_6, 500 MHz) δ 8.35 (s, 1H), 8.29 (s, 1H), 7.61 (s, 1H), 2.82−2.75 (m, 1H) (d, 1H, Hz); HRMS calcd for $C_{13}H_{14}N_5O2$ (M + H) 272.1148, found 272.1155.

[*N*^{6}-Isobutyryl-6-amino-3,7-dihydroimidazo[4,5-*g*]-quinazolin-8-one]-1′-β-D-ribofuranoside 2′,3′,5′-Triacetate (14a and N7 Regioisomer 14b). The isobutyryl-protected base **13** (1.051 g, 3.88 mmol) was dried in vacuo overnight and was then suspended in 16.0 mL of anhydrous CH_3CN. An 11.6 mL portion of *N*,*O*-bistrimethylsilyl acetamide was then added to the reaction mixture. The starting material dissolved under heating at reflux for 15 min. Then 23.3 mL of a solution of 0.20 mM β-D-ribofuranose 1,2,3,4-tetraacetate in anhydrous CH_3CN was added, followed by addition of trimethylsilyl triflate (5.8 mL). After the reaction was heated at reflux for 5 h, it was cooled to room temperature, concentrated, and redissolved in 140 mL ethyl acetate. The organic phase was washed with water, 5% sodium bicarbonate, and brine (70 mL each). The organic layer was then dried over sodium sulfate, concentrated, and purified by silica gel column chromatography (0−3−5% MeOH in CH_2Cl_2). An inseparable ~1:1 mixture of N1 and N3 coupling products was obtained in 70% yield (1.436 g): ^{1}H NMR ($CDCl_3$, 500 MHz) δ 8.69 (s, 0.5H), 8.45 (s, 0.5H), 8.39 (s, 0.5H), 8.32 (s, 0.5H), 7.91 (s, 0.5H), 7.65 (s, 0.5H), 6.20 (d, 0.5H, J = 6.5 Hz), 6.13 (d, 0.5H, J = 5.0 Hz), 5.62−5.60 (m, 1H), 5.47−5.41 (m, 1H), 4.51−4.50 (m, 1H), 4.42−4.389 (m, 2H), 2.77 (m, 1H), 2.18−2.07 (m, 9H), 1.25 (d, 6H, J = 7.0 Hz); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 179.0, 178.7, 170.7, 170.7, 169.9, 169.9, 169.7, 169.6, 162.1, 161.8, 149.6, 145.5, 145.1, 144.9, 144.1, 143.9, 142.7, 138.0, 131.7, 119.6, 117.5, 117.1, 116.5, 108.8, 106.5, 87.6, 87.0, 81.0, 80.5, 73.6, 73.1, 70.8, 70.3, 63.2, 63.0, 36.9, 36.9, 21.2, 21.1, 20.9, 20.8, 20.7, 20.7, 19.3, 19.3; HRMS calcd for $C_{24}H_{28}N_5O_9$ (M + H) 530.1887, found 530.1878.

[*N*^{6}-Isobutyryl-6-amino-3,7-dihydro-imidazo[4,5-g]-quinazolin-8-one]-1′-β-D-ribofuranoside (regioisomers 15a and 15b). The mixture of isomers **14a** and **14b** (1.420 g, 2.68 mmol) was dissolved in 30.0 mL anhydrous MeOH, to which was added 10.0 mL of 2.0 M ammonia in MeOH. After stirring at room temperature for 7.5 h, the reaction was evaporated. The residue was suspended in ethyl acetate, filtered, washed with EtOAc, and dried in vacuo, which furhished 0.846 g (95%) free dxG (**2**). This crude product was essentially pure by ^{1}H NMR. Without further purification, this intermediate was

(4) Schweitzer, B. A.; Kool, E. T. *J. Org. Chem.* **1994**, *59*, 7238−7242.

(5) (a) Moran, S.; Ren, R. X.-F.; Rumney, S.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056−2057. (b) Moran, S.; Ren, R. X.-F.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10506−10511. (c) Morales, J. C.; Kool, E. T. *Nature Struct. Biol.* **1998**, *5*, 950−954. (d) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **1999**, *121*, 2723−2724. (e) Matray, T. J.; Kool, E. T. *Nature* **1999**, *399*, 704−708. (f) Dzantiev, L.; Alekseyev, Y.; Morales, J. C.; Kool, E. T.; Romano, L. J. *Biochemistry* **2001**, *40*, 3215−3221. (g) Delaney, J. C.; Henderson, P. T.; Helquist, S. A.; Morales, J. C.; Essigmann, J. M.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4469−4473.

(6) (a) Schofield, M. J.; Brownewell, F. E.; Nayak, S.; Du, C.; Kool, E. T.; Hsieh, P. *J. Biol. Chem.* **2001**, *276*, 45505−8. (b) Drotschmann, K.; Yang, W.; Brownewell, F. E.; Kool, E. T.; Kunkel, T. A. *J. Biol. Chem.* **2001**, *276*, 46625−9. (c) Begley, T. J.; Haas, B. J.; Morales, J. C.; Kool, E. T.; Cunningham, R. P. *DNA Repair* **2002**, *2*, 107−120. (d) Washington, M. T.; Helquist, S. A.; Kool, E. T.; Prakash, L.; Prakash, S. *Mol. Cell. Biol.* **2003**, *23*, 5107−5112. (e) Francis, A. W.; Helquist, S. A.; Kool, E. T.; David, S. S. *J. Am. Chem. Soc.* **2003**, *125*, 16235−16242.

(7) Liu, H.; Gao, J.; Maynard, L.; Saito, Y. D.; Kool, E. T. *J. Am. Chem. Soc.* **2004**, *126*, 1102−1109.

(8) Bommarito, S.; Peyret, N.; SantaLucia, J., Jr. *Nucleic Acids Res.* **2000**, *28*, 1929−1934.

(9) Gao, J.; Liu, H.; Kool, E. T. *J. Am. Chem. Soc.* **2004**, *126*, 11826−11831.

(10) (a) Liu, H.; Gao, J.; Lynch, S. R.; Saito, Y. D.; Maynard, L.; Kool, E. T. *Science* **2003**, *302*, 868−871. (b) Liu, H.; Gao, J.; Kool, E. T. *J. Am. Chem. Soc.* In press.

(11) Leonard, N. J.; Sprecker, M. A.; Morrice, A. G. *J. Am. Chem. Soc.* **1976**, *98*, 3987−3994.

subjected to N6 protection. An 840 mg (2.54 mmol) portion of this intermediate was coevaporated with anhydrous pyridine (3 × 5 mL) and then was suspended in 30.0 mL of anhydrous pyridine, followed by addition of TMSCl (4.2 mL, 32.6 mmol). After 30 min of stirring, the reaction solution was treated with isobutyric anhydride (0.83 mL, 5.27 mmol). After being stirred at room temperature for 24 h, the reaction was cooled to 0 °C and treated with 5.0 mL of water. After 5 min, 5.0 mL of 25% ammonium hydroxide was added. After an additional 15 min stirring at 0 °C, volatiles were removed by evaporation. The product was purified in nearly quantitative yield by silica gel column chromatography (10−20% MeOH in CH_2Cl_2): ^{1}H NMR (CD_3OD, 500 MHz) δ 8.78 (s, 0.4H), 8.70 (s, 0.6H), 8.54 (s, 0.4H), 8.49 (s, 0.6H), 7.83 (s, 0.6H), 7.81 (s, 0.4H), 6.06 (d, 0.4H, J = 7.0 Hz), 6.02 (d, 0.6H, J = 7.5 Hz), 4.51−4.48 (m, 1H), 4.32−4.29 (m, 1H), 4.18−4.15 (m, 1H), 3.93−3.87 (m, 1H), 3.84−3.79 (m, 1H), 2.75−2.72 (m, 1H), 1.24 (d, 1H, J = 8.5 Hz); ^{13}C NMR (DMSO-d_6, 125 MHz) δ 181.0, 180.9, 149.8, 147.7, 146.7, 142.6, 138.9, 132.1, 117.5, 116.6, 116.3, 109.3, 89.6, 86.48, 86.2, 84.1, 74.7, 74.4, 70.8, 70.7, 61.9, 61.7, 35.5, 19.7; HRMS calcd for $C_{18}H_{22}N_5O_6$ (M + H) 404.1570, found 404.1575.

3′,5′-Tetraisopropyldisiloxane-Protected [*N*^{6}-Isobutyryl-6-amino-3,7-dihydroimidazo[4,5-*g*]quinazolin-8-one]-1′-*β*-D-ribofuranoside (Regioisomers 16a and 16b). The amine-protected riboside mixture **15a**,**b** (496 mg, 1.23 mmol) and imidazole (300 mg) were dissolved in 20.0 mL of anhydrous DMF, to which was added 481 μL (1.48 mmol) of 1,3-dichloro-1,1,3,3-tetraisopropylsiloxane. After being stirred at room temperature for 24 h, the solvent was removed by evaporation. The residue was dissolved in 250 mL of ether. After being washed with 125 mL of water and 125 mL of aqueous saturated $NaHCO_3$, the organic phase was dried over $NaSO_4$, filtered, concentrated, and purified by silica gel column chromatography (50−100% EtOAc in hexanes). A white solid was obtained as product in 71% yield (560 mg): ^{1}H NMR ($CDCl_3$, 500 MHz) δ 12.00 (s, 0.5H), 11.93 (s, 0.5H), 8.75 (s, 0.5H), 8.47 (s, 0.5H), 8.42 (s, 0.5H), 8.39 (s, 0.5H), 7.95 (s, 0.5H), 7.54 (s, 0.5H), 6.03 (s, 0.5H), 5.99 (s, 0.5H), 4.60−4.55 (m, 1H), 4.34−4.20 (m, 3H), 4.16−4.12 (m, 1H), 3.32 (s, 1H), 3.129 (s, 1H), 2.72−2.67 (m, 1H), 1.34−1.27 (m, 6H), 1.14−0.99 (m, 28H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 170.7, 170.6, 170.0, 169.9, 169.8, 169.7, 149.3, 145.3, 144.0, 142.5, 137.9, 131.5, 119.4, 117.2, 116.7, 108.9, 87.5, 87.1, 80.9, 80.4, 73.5, 73.2, 70.7, 70.2, 63.2, 62.9, 36.6, 21.1, 21.0, 20.9, 20.8, 20.7, 20.7, 19.3, 19.3; HRMS calcd for $C_{30}H_{47}N_5O_7Si_2$ (M + H) 646.3087, found 646.3082.

3′,5′-Tetraisopropyldisiloxane-Protected [*N*^{6}-Isobutyryl-6-amino-3,7-dihydroimidazo[4,5-*g*]quinazolin-8-one]-1′-*β*-D-Ribofuranoside-2′-*O*-thiocarbonate Phenyl Ester (17). Starting material **16a**,**b** (859 mg, 1.33 mmol) and DMAP (974 mg, 7.98 mmol) were dissolved in 40 mL of anhydrous CH_3CN, to which was added phenyl chlorothionoformate (368 μL, 2.66 mmol). After being stirred at room temperature 24 h, the reaction mixture was evaporated to dryness. The residue was dissolved in 120 mL of EtOAc, and the organic phase was washed with saturated $NaHCO_3$ (60 mL) and brine (60 mL). After being dried over $NaSO_4$, the organic layer was concentrated and purified by silica gel column chromatography (25−50−67% EtOAc in hexanes). Separation of imidazole regioisomers was successful in this step. The desired product (**17**) was obtained as a pale yellow foamy solid in 50% yield (516 mg): ^{1}H NMR ($CDCl_3$, 500 MHz) δ 11.99 (s, 1H), 8.67 (s, 1H), 8.72 (b, 1H), 8.51 (s, 1H), 7.59 (s, 1H), 7.463 (t, 2H, J = 7.5 Hz), 7.35 (t, 1H, J = 7.0 Hz), 7.179 (d, 2H, J = 8.0 Hz), 6.21 (s, 1H), 6.03 (d, 1H, 7.0 Hz), 4.84−4.81 (m, 1H), 4.36−4.33 (m, 1H), 4.28−4.26 (m, 1H), 4.26−4.12 (m, 1H), 2.67−2.65 (m, 1H), 1.28 (d, 6H, J = 7.0 Hz), 1.16−1.00 (m, 28H); ^{13}C NMR ($CDCl_3$ 125 MHz) δ 194.1, 178.6, 162.2, 153.6, 145.38, 144.9, 143.6, 142.8, 137.8, 129.9, 127.2, 121.9, 121.2, 119.6, 117.0, 105.9, 87.8, 84.2, 82.7, 68.6, 60.7, 60.2, 36.9, 19.3, 17.7, 17.6, 17.6, 17.6, 17.3, 17.20, 17.2, 17.0, 14.5, 13.64, 13.2, 13.0; HRMS calcd for $C_{37}H_{52}N_5O_8Si_2S$ (M + H) 782.3070, found 782.3058.

The other regioisomer was eluted from the silica column with 100% EtOAc in 21% yield (214 mg).

3′,5′-Tetraisopropyldisiloxane-Protected [*N*^{6}-Isobutyryl-6-amino-3,7-dihydroimidazo[4,5-*g*]quinazolin-8-one]-1′-*β*-D-2′-deoxyribofuranoside (18). Starting material **17** (480 mg, 0.61 mmol) and AIBN (45 mg) were dissolved in 18.0 mL of toluene, to which was added 2.0 mL of n-Bu_3SnH. After being heated at reflux for 2 h, the reaction mixture was cooled. Volatiles were removed by evaporation, and compound **18** was purified by silica gel column chromatography (50−75% EtOAc in hexanes). The product was obtained as a white foamy solid in 71% yield (276 mg): ^{1}H NMR ($CDCl_3$ 500 MHz) δ 12.04 (s, 1H), 8.75 (s, 1H), 8.35 (s, 1H), 7.47 (s, 1H), 6.24 (dd, 1H, J = 3.5, 7.0 Hz), 4.73−4.72 (m, 1H), 4.12−3.97 (m, 3H), 2.70−2.65 (m, 2H), 2.61−2.59 (m, 1H), 1.32 (d, 6H, J = 6.5 Hz), 1.12−0.99 (m, 28H); ^{13}C NMR ($CDCl_3$ 125 MHz) δ 178.5, 162.2, 145.3, 144.5, 143.8, 142.9, 138.1, 119.5, 116.6, 85.6, 84.1, 69.9, 61.8, 40.2, 36.9, 19.3, 17.8, 17.7, 17.6, 17.4, 17.3, 17.1, 13.7, 13.4, 13.2, 12.7; HRMS cald. for $C_{30}H_{48}N_5O_6Si_2$ (M + H) 630.3138, found 630.3182.

[*N*^{6}-Isobutyryl-6-amino-3,7-dihydroimidazo[4,5-*g*]-quinazolin-8-one]-1′-*β*-D-2′-deoxyribofuranoside (19). Starting material **18** (276 mg, 0.44 mmol) was dissolved in 12.0 mL of anhydrous THF and 4.0 mL of anhydrous pyridine, to which was added 70% HF in pyridine (350 μL). After the reaction mixture was stirred at room temperature for 80 h, it was concentrated and purified by silica gel column chromatography (10−20% MeOH in CH_2Cl_2). The product was obtained as a white solid in essentially quantitative yield: ^{1}H NMR (CD_3OD, 500 MHz) δ 11.98 (s, 1H), 11.51 (s, 1H), 8.68 (s, 1H), 8.33 (s, 1H), 7.71 (s, 1H), 6.45 (t, 1H, J = 7.0 Hz), 5.41 (d, 1H, J = 4.0 Hz), 5.01 (t, 1H, J = 5.0 Hz), 4.44−4.42 (m, 1H), 3.89 (q, 1H, J = 4.0 Hz), 3.61−3.54 (m, 2H), 2.81−2.78 (m, 1H), 2.67−2.62 (m, 1H), 2.37−2.33 (m, 1H), 1.13 (d, 6H, J = 6.5 Hz); ^{13}C NMR (DMSO-d_6, 125 MHz) δ 181.0, 161.6, 146.5, 146.2, 144.9, 142.6, 139.0, 117.5, 116.2, 107.1, 88.3, 85.4, 71.1, 62.1, 35.5, 19.7; HRMS calcd for $C_{18}H_{22}N_5O_5$ (M + H) 388.1621, found 388.1631.

Base-Protected Phosphoramidite Derivative of 19. Synthetic details and characterization are given in the Supporting Information.

Synthesis of dxC Nucleoside.

3′,5′-Diacetyl-1′-*β*-[8-(6-methylquinazoline-2, 4-dione)]-D-2′-deoxyribofuranoside (4). 1′-β-[8-(6-Methylquinazoline-2, 4-dione)]-2′-D-deoxyribofuranose **3**[7] (200 mg, 0.69 mmol) was dissolved in 5 mL of dry pyridine. To the solution was added 200 μL of acetic anhydride. The mixture was stirred at room temperature for 8 h. Volatiles were removed under vacuum, and crude product was purified by silica column chromatography (hexanes/ethyl acetate 1:1 followed by hexanes/ethyl acetate 1:2) to give the product as a white foam (195 mg, 76%): ^{1}H NMR ($CDCl_3$, 500 MHz), δ= 9.74 (S, 1H), 9.39 (s, 1H), 7.93 (s, 1H), 7.28 (s, 1H), 5.26 (m, 2H), 4.52 (m, 1H), 4.33 (m, 2H), 2.38 (s, 3H), 2.40−2.29 (m, 2H), 2.21 (s, 3H), 2.15 (s, 3H); ^{13}C NMR ($CDCl_3$, 125.7 MHz) δ 171.3, 170.9, 163.3, 150.2, 136.5.06, 133.0, 128.3, 123.9, 116.0, 83.9, 80.7, 75.6, 64.0, 39.7, 21.3, 21.2, 20.8; HRMS (EI+) calcd for $C_{18}H_{21}N_2O_7$ $[M + H]^+$ m/z = 377.1349, found 377.1349.

3′,5′-Diacetyl-1′-*β*-[8-(6-methyl-4-thio-2-quinazolone)]-D-2′-deoxyribofuranoside (5). Compound **4** (70 mg, 0.19 mmol) and phosphorus pentasulfide (211 mg, 0.95 mmol) were charged into a round-bottom flask equipped with condenser. Four milliliters of dry pyridine was added to form a suspension, which turned into a clear solution upon heating to reflux. TLC showed the reaction was complete after 4 h. Solvent was removed under vacuum, and the residue was suspended in 15 mL of warm water. The aqueous suspension was extracted with 3 × 20 mL dichloromethane. The organic layer was collected, dried, and concentrated. Silica column chromatography (hexanes/ethyl acetate 1:1) gave product **5** as yellow foam (56 mg, 77%): ^{1}H NMR ($CDCl_3$, 500 MHz) δ 10.74 (s, 1H), 9.62 (s, 1H), 8.33 (s, 1H), 7.29 (s, 1H), 5.25 (m, 2H), 4.52 (m, 1H), 4.34 (m, 2H), 2.40 (s, 3H), 2.36 (m, 1H), 2.30 (m, 1H),

2.21 (s, 3H), 2.16 (s, 3H); ^{13}C NMR ($CDCl_3$, 125.7 MHz) δ 191.9, 171.2, 170.8, 147.0, 135.5, 133.9, 133.6, 131.8, 123.9, 122.0, 84.0, 80.7, 75.5, 63.9, 39.7, 21.3, 21.2, 20.96; HRMS (EI+) calcd for $C_{18}H_{21}N_2O_6S$ $[M + H]^+$ *m/z* 393.1120, found 393.1112.

1′-β-[8-(4-Amino-6-methyl-2-quinazolone)]-D-2′-deoxyribofuranoside (1). In a high-pressure reaction tube, compound **5** (106 mg, 0.27 mmol) was dissolved in 10 mL of anhydrous methanol and chilled to 0 °C with an ice−water bath. Ammonia gas was bubbled into the solution for 1 min. The tube was sealed and heated to 100 °C in an oil bath. The reaction mixture was stirred at 100 °C for 8 h and then was allowed to cool to room temperature. Methanol was evaporated under vacuum, and residue was suspended in 10 mL of dichloromethane. The solid was collected by filtration to give product **1** as pale solid (77 mg, 97%): ^{1}H NMR (DMSO-d_6, 500 MHz) δ 9.80 (s, broad, 1H), 7.79 (s, 1H), 7.9−7.5 (broad peak, 2H), 7.45 (s, 1H), 5.27 (dd, 1H, J = 11 Hz, 5 Hz), 5.13 (s, broad, 1H), 4.25 (m, 1H), 3.86 (m, 1H), 3.58 (m, 2H), 2.31 (s, 3H), 2.09 (m, 1H), 1.96 (m, 1H); ^{13}C NMR (DMSO-d_6, 125.7 MHz) δ 164.6, 156.4, 138.7, 133.8, 130.2, 126.6, 124.2, 109.3, 88.73, 78.6, 72.9, 62.2, 42.1, 21.1; HRMS (EI+) calcd for $C_{14}H_{18}N_3O_4$ $[M + H]^+$ *m/z* 292.1297, found 292.1295.

Base-Protected Dimethoxytrityl Phosphoramidite Derivative of 1. Synthetic details and characterization are given in the Supporting Information.

Oligonucleotide Synthesis. 5′-*O*-Dimethoxytritylated phosphoramidite analogues of nucleosides were synthesized as described above and in the Supporting Information. Oligodeoxynucleotides were synthesized on a 1.0 μmol scale on a DNA synthesizer using standard β-cyanoethyl phosphoramidite chemistry for DNA synthesis, but with extended (120 s) coupling times for modified nucleosides. Stepwise coupling yields for nonnatural nucleosides were all greater than 95% as determined by trityl cation monitoring. Oligomers were deprotected in concentrated ammonium hydroxide (55 °C, 16 h), purified by preparative 20% denaturing polyacrylamide gel electrophoresis, and isolated by excision and extraction from the gel, followed by dialysis against water. The recovered material was subsequently quantified by absorbance at 260 nm with molar extinction coefficients determined by the nearest neighbor method. Values for oligomers containing modified nucleotides were estimated by measuring the molar extinction coefficients of modified nucleosides at 260 nm and adding these values to the calculated values of natural DNA fragments. Extinction coefficients of dxG and dxC at 260 nm were measured at 8060 and 5800 $cm^{-1}\cdot M^{-1}$, respectively. Intact incorporation of modified nucleotides was confirmed by characterization of short unpurified oligomers (sequence TXT) by ^{1}H NMR (see the spectra in the Supporting Information). These were also characterized by ESI-MS: calcd for TxGT ($C_{34}H_{42}N_9O_{18}P_2$) [M + H] 926.2, found 926.1; calcd for TxCT ($C_{34}H_{42}N_7O_{18}P_2$) $[(M - H)^-]$ *m/z* = 898.2, found 898.2.

Thermal Denaturation Studies. Solutions for the thermal denaturation studies contained either a 1:1 ratio of two complementary oligomers or only a self-complementary oligomer. Total oligomer concentrations ranged from 2.0 to 30 μM. The buffer contained NaCl (100 mM), $MgCl_2$ (10 mM), and Na·PIPES(10 mM), at pH = 7.0, except where noted otherwise. After solutions were prepared, they were heated to 95 °C for 10 min and allowed to cool to room temperature over at least 2 h and then cooled at 5 °C for at least 4 h. Melting studies were carried out in Teflon-stoppered 1 cm path length quartz cells (under nitrogen atmosphere when temperature was below 20 °C) on a UV−vis spectrophotometer equipped with a thermoprogrammer and temperature control. Absorbance was monitored while the temperature was changed at a rate of 0.5 °C/min. Melting data were essentially the same for heating or cooling at this rate. Experiments were monitored at 260 nm. In most cases, the complexes displayed apparent two-state transitions, with all-or-none melting curves from bound duplex to free single strands. Computer fitting of the melting data using Meltwin 3.0b provided both melting temperatures T_m and free energy values for the complexes. The free energies were also calculated from van't Hoff plots by plotting $1/T_m$ vs ln(C/4); in all cases, close agreement was observed, indicating that the two-state approximation is reasonable for these specific sequences.

Results

Design Principles. We chose a 4-amino-6-methyl-2-quinazolone C-nucleoside as a size-expanded analogue of deoxycytidine. This compound was previously unknown. A 4-amino-2-quinazolone nucleoside with the glycosidic attachment on N1 has been reported.[12] However, it was not incorporated into oligonucleotides, and in any case, with attachment at N1 of the heterocycle, the reported nucleoside would not have an expanded pairing dimension. Space-filling models of 4-amino-6-methyl-2-quinazolone with electrostatic potentials showed (Figure 1) that it is expected to have similar electrostatic charges along the Watson−Crick analogous hydrogen-bonding edge as cytosine, suggesting its capability for forming hydrogen bonds with guanine.

In a completely analogous way to the design of dxT nucleoside,[7] attachment of 2′-deoxyribose to the C8 of 4-amino-6-methyl-2-quinazolone would give the expanded cytidine analogue dxC (Figure 1). Compared with the structure of cytidine, dxC has an extra methyl group attached to the C6 of 2-quinazolone. Thus, rigorously speaking, the xC base is an expanded analogue of 5-methylcytosine, rather than cytosine. We decided to retain the methyl group in our molecular design for ease of synthesis from an intermediate in the xT preparation (see below).

Analogous to xT base, the xC base has as number of possible tautomers as well. Preliminary semiempirical calculations of heats of formation suggested that the correct tautomer for pairing with guanine, at least in the gas phase, was the most stable one (Supporting Information). Notably, a second tautomer, an imino variant of the exocyclic amine, was predicted to be less stable by a small margin of 1.3 kcal/mol, raising a possible issue of selectivity in pairing. In fact, further study showed that xC is highly selective in pairing with G (see below), indicating that this imino tautomer, if it does exist in aqueous solution, does not interfere with the pairing properties of xC.

The design of the xG base was a straightforward adaptation of the structure of the previously known xA base. The deoxyriboside derivative of xG (dxG, Figure 1) was not previously known in the literature, although the ribose derivative was reported by Leonard 25 years ago.[13] As with the xA base, the xG base is larger than G by benzo-homologation and is longer by approximately 2.4 Å. A space-filling model is shown in Figure 1, and the calculated electrostatic potentials along the Watson−Crick-analogous edge are essentially the same as those on guanine.

The xG base has a number of possible tautomeric forms. Preliminary heat-of-formation calculations suggested, however (see Supporting Information), that the desired guanine-like tautomer should be the energetically

(12) Stout, R. G.; Robins, R. K. *J. Org. Chem.* **1968**, *33*, 1219−1225.

(13) Leonard, N. J.; Keyser, G. E. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 4262−4264.

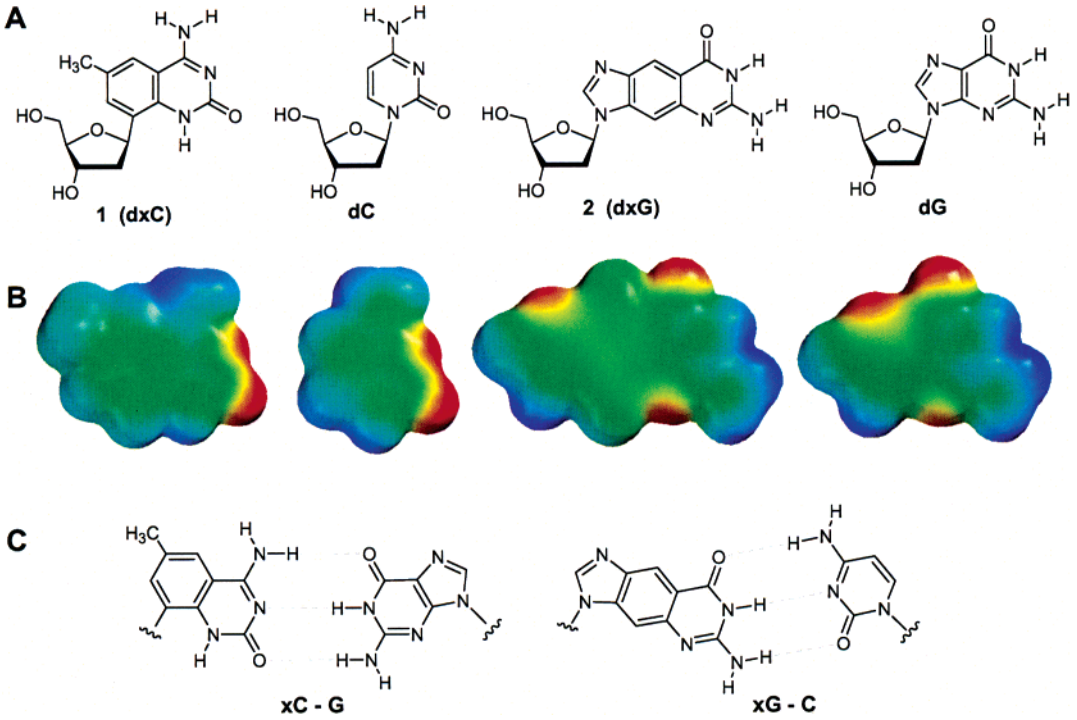


FIGURE 1. Structures of size-expanded (xDNA) homologues of cytosine and guanine. (A) Structures of the benzo-homologated deoxyribosides, with natural dC and dG nucleosides shown for comparison. (B) Space-filling models of the free bases xC, C, xG, and G, respectively, with electrostatic potentials mapped on the surfaces. Geometry optimization and electrostatics calculated by AM1 using Spartan (Wavefunction, Inc). Colors depict an electrostatic scale of −70 to +50 (red = negative potential, blue = positive). (C) Proposed structures of expanded base pairs formed between xC or xG with the Watson−Crick complements.

most stable one relative to two other tautomers, by relatively large margins of 7−10 kcal/mol.

Synthesis of dxC and dxG Nucleoside Homologues. Synthesis of dxC proved to be quite straightforward (see Scheme 1) and proceeded from the previously reported quinazolinedione intermediate **3**.[7] The 5′- and 3′-hydroxyls were acylated with acetic anhydride in pyridine to give the protected nucleoside **4**. Regioselective conversion of C4 carbonyl to thiocarbonyl was accomplished by treating compound **4** with phosphorus pentasulfide in refluxing pyridine.[12] The reaction was closely monitored to prevent overreaction to give the 2,4-dithioquinazolone. The regioselectivity of the sulfurization reaction was confirmed by heteronuclear multibond coherence (HMBC) experiments with compounds **4** and **5** (see details in the Supporting Information).

Methanolic ammonia treatment of compound **5** converted the 4-thio group to 4-amino and removed the diacetyl protecting groups on the sugar, yielding the dxC free nucleoside **1** in high yield. In preparation for oligonucleotide synthesis, the 4-amino group on quinazolone was protected with a formamidine group to give compound **6**, which was subsequently 5′-dimethoxytritylated and phosphitylated to afford the desired dxC-phosphoramidite **8** in very good yield.

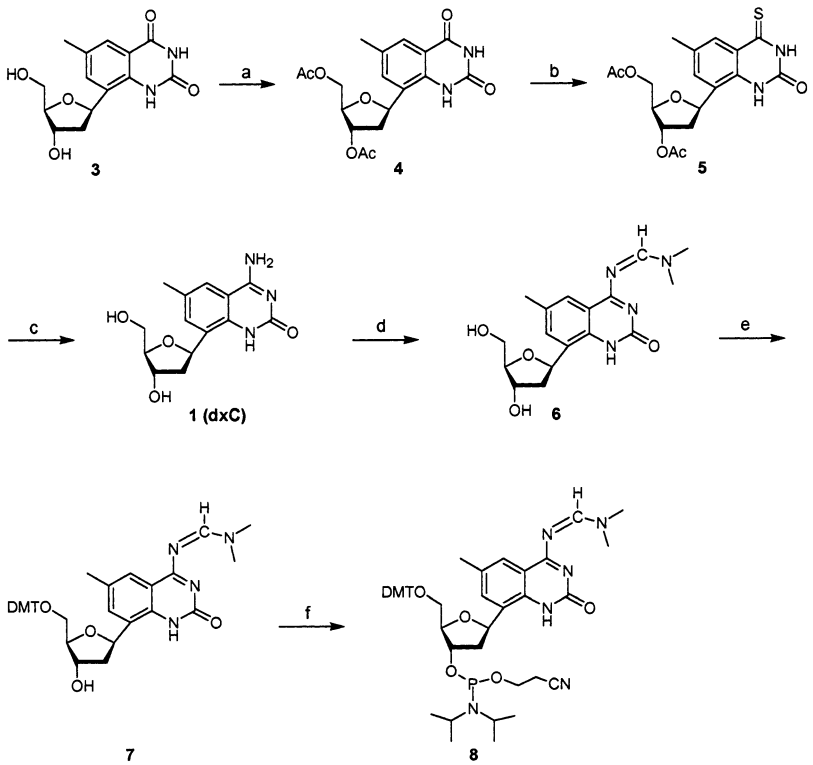
Experiments showed that dxC phosphoramidite **8** was compatible with standard oligodeoxynucleotide synthesis protocols. A short test oligonucleotide synthesis (sequence d(T−xC−T)) showed that good coupling yield (93%) can be obtained. The product was characterized by NMR and mass spectrometry for its chemical integrity and purity (Supporting Information).

The expanded deoxyguanosine nucleoside, dxG (**2**), and its protected phosphoramidite derivative **21** were prepared as shown in Scheme 2. Transformations from **9** to **11** followed reported procedures.[13] In converting **11** to the free xG base **12** we found it necessary to modify the reported reaction conditions as NH_2CN and *t*-BuOK gave unsatisfactory results in our hands. The use of *S*-methylisothiouronium sulfate resulted in higher yields. The exocyclic amine of **12** was then protected with isobutyryl chloride to give **13** in 82% yield.

The coupling reaction between **13** and tetraacetyl-*d*-ribose was carried out in the presence of *N*,*O*-bistrimethylsilyl acetamide and trimethylsilyl triflate in refluxing acenonitrile to afford compounds **14** in 70% yield as an inseparable mixture of regiosiomers. Treatment of **14** with 2.0 M ammonia in methanol followed by selective protection of the amine group on the base furnished isobutyryl-protected dxG nucleoside mixture **15** in 95% overall yield. Selective protection of the 3′- and 5′-hydroxyl groups was achieved via 1,3-dichlorotetraisopropyl siloxane to give compounds **16** in 71% yield. The free 2′-OH of **16** was then removed following a two-step dehydroxylation protocol.[14] The first step involved transformation of the free hydroxyl group of **16** to the thiocarbonate **17** in 50% yield. It was at this stage that the two regioisomers could be separated by column chromatography. The desired regioisomer was then transformed to the deoxyribose derivative **18** with tributyltin hydride in 71% yield. The structures of the isolated regioisomers

(14) Barton, D. H. R.; Ferreira, J. A.; Jaszberenyi, J. C. *Prepr. Carbohydr. Chem.* **1997**, 151−172.

SCHEME 1[a]



[a] Reagents and conditions: (a) Ac_2O, pyridine, 76%; (b) P_2S_5, pyridine, reflux, 77%; (c) methanolic ammonia, 100 °C, 97%; (d) *N*,*N*-dimethylformamide dimethyl acetal, pyridine, 82%; (e) 4,4′-dimethoxytrityl chloride, DIPEA, pyridine, 78%; (f) 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite, DIPEA, CH_2Cl_2, 86%.

were identified through NMR experiments. First, HMBC experiments were used to identify the benzo-group hydrogens proximal and distal to the ribose in the two isomers. With this information in hand, we were then able to identify the two regioisomers based upon NOE measurements (Supporting Information).

The synthesis was continued with the single regioisomer **18**. Removal of the disiloxane protecting group with pyridine·HF proceeded in nearly quantitative yield to give base-protected deoxynucleoside **19**; this could be deprotected with ammonia to yield free nucleoside **2** for characterization of fluorescence (see below). The isobutyryl protected compound **19** was then converted to the desired phosphoramidite **21** by DMT protection (93% yield) and phosphoramidite formation (77% yield).

As for dxC, the dxG nucleoside phosphoramidite derivative **21** could be readily incorporated into oligodeoxynucleotides by standard automated protocols. A short test oligomer d(T−xG−T) was prepared and characterized by NMR and mass spectrometry (Supporting Information), which confirmed intact incorporation of the homologue of deoxyguanosine.

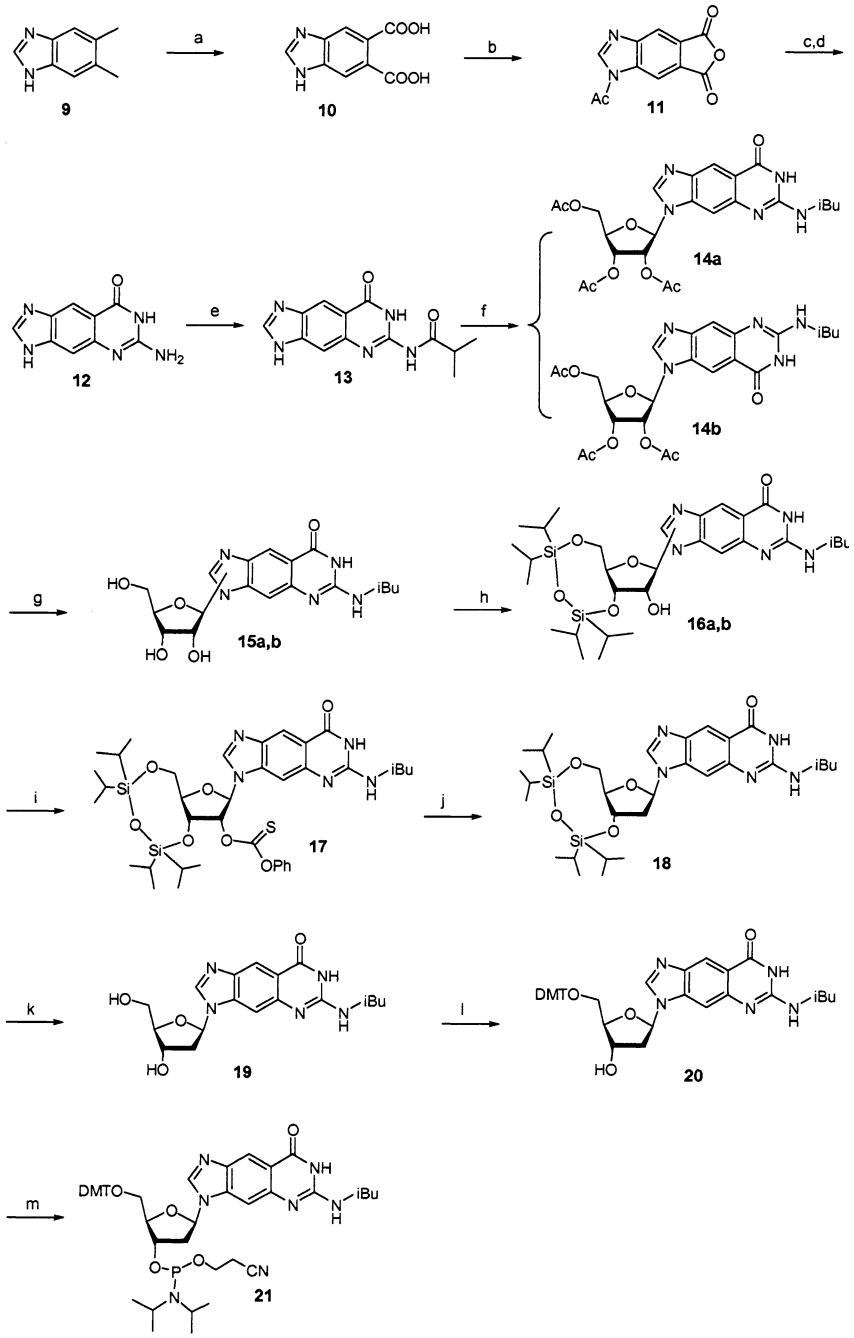
Fluorescence of dxC and dxG. Experiments showed that the free nucleoside analogues **1** and **2** were fluorescent under illumination by ultraviolet light. Excitation and emission spectra of the two compounds in methanol are shown in Figure 2. The dxC nucleoside (**1**) has, in addition, to short wavelength absorptions, a longer wavelength absorption band at 330 nm (ϵ = 4100). Excitation at this wavelength resulted in an emission band having a maximum at 388 nm. The quantum yield for fluorescence was measured to be 0.52. Similarly, the dxG free nucleoside (**2**) has a long wavelength absorption at 320 nm (ϵ = 3400) and gave fluorescence emission at 413 nm.

Pairing and Stacking Properties in the Context of Natural DNA. The initial discovery that the first size-expanded analogue, xA, pairs selectively with T in the context of DNA[9] suggests that DNA backbones are moderately flexible and that recognition between xA and T through hydrogen bonding may exist. In the current study we wished to see if this is generally true for all size-expanded base analogues, as a general test of DNA backbone flexibility. Moreover, all four expanded bases are fluorescent and might conceivably report on matched or mismatched bases at specific sites by fluorescence changes. For these reasons we tested pairing stability and selectivity of the new analogues xC and xG in the context of natural DNA duplexes.

The pairing properties of nucleosides dxC and dxG were studied in 12mer duplex DNAs by UV monitored thermal melting experiments in a pH 7.0 buffer containing 100 mM NaCl, 10 mM $MgCl_2$, and 10 mM Na·PIPES. The same 12 base−pair context used previously for dxA and dxT was employed[9] so that direct comparisons could be made. To test the effects of a change in nearest neighbor bases, we tested each expanded base in either a purine strand or a pyrimidine strand of the same duplex. The pairing data are given in Table 1.

Results showed that, when paired opposite their Watson−Crick partners, both expanded bases destabilized the duplex by a small amount as compared to the normally sized base pairs. For example, the duplex containing an xC−G pair was less stable than the control

SCHEME 2[a]



[a] Reagents and conditions: (a) $KMnO_4$, 75 °C, 55%; (b) Ac_2O, reflux, 93%; (c) $TMSN_3$, 95 °C; (d) *S*-methylisothiouronium sulfate, H_2O/THF 1:9, 54% two steps; (e) isobutyryl chloride, Py, reflux, 82%; (f) N,O−Bistrimethylsilylacetamide, trimethylsilyl triflate, ribofuranose tetraacetate, AcCN, reflux, 70%; (g) (i) 2.0 M NH_3 in MeOH, 7.5 h, 95%; (ii) TMSCl, pyridine then isobutyric anhydride, 24 h, 100%; (h) 1,3-dichlorotetraisopropylsiloxane, imidazole, 64%; (i) phenyl chlorothioformate, DMAP, 44%; (j) Bu_3SnH, AIBN, PhMe, reflux, 84%; (k) Py·HF, THF, 93%; (l) DMTrCl, Py, 93%; (m) 2-cyanoethyl diisopropylchlorophosphoramidite, DIPEA, 77%.

(containing C−G) by a small 1.9 °C in thermal melting temperature (T_m) and 0.3 kcal/mol in free energy (Table 1). Changing the sequence context (to G−xC) gave a very similar result. By comparison, the duplexes containing xG pairs were slightly more destabilized; for example, the C−xG pair destabilized the duplex by 6.3 °C in T_m and 1.3 kcal/mol. These values are quite close to those observed previously for the earlier expanded analogues xT and xA.[9]

The pairing data also revealed selectivities of the expanded bases in these two 12mer duplex contexts. The xC base showed the greatest sequence selectivity against mismatches, with drops in T_m of 11.3−18.0 °C and losses of favorable free energy of 2.7−3.8 kcal/mol, as compared with the matched pairs of xC with G (Table 1). By comparison, the selectivity of xG was lower, with drops in T_m of 7.2−11.7 °C. By this measure, the selectivities of xG are similar in magnitude to those previously measured for xT and xA,[9] while those of xC are considerably higher than the other three expanded bases. To test the relative destabilizations of the mismatched cases, we also paired the expanded bases with a tetrahydrofuran

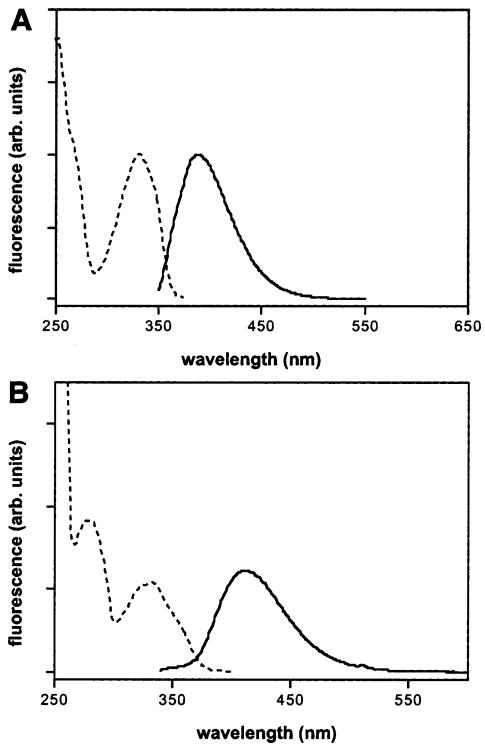


FIGURE 2. Fluorescence of the expanded-size free deoxynucleosides in methanol. (A) Excitation and emission spectra of dxC nucleoside. Excitation for the emission spectrum (solid line) was at 330 nm; the excitation spectrum (dashed line) was monitored at 388 nm. (B) Excitation and emission spectra of dxG nucleoside. Excitation for the emission spectrum (solid line) was at 330 nm; the excitation spectrum (dashed line) was monitored at 413 nm.

abasic analogue[15] (Table 1). Results showed that xC or xG paired opposite this abasic site yielded stabilities about the same as the least stable mismatches, suggesting that most of the mismatched bases contribute a small degree of stabilization by remaining stacked in the duplex. Interestingly, for both xC and xG, the least stable mismatch was with A.

Since the stability of a pair depends heavily on the stacking properties of the constituent bases,[16] we explicitly measured stacking properties for xC and xG in the context of short self-complementary sequences designed to form duplexes with a single overhanging base, putatively stacked at the end. The data are shown in Table 2, with comparison to the same duplexes lacking an overhanging base. In this system, the added stabilization is expected to arise from stacking of the overhanging base on a neighboring base pair.[16] The results show that both xC and xG in an overhanging end stabilized their core duplexes more than their nonexpanded counterparts did. The xG base stabilized the duplex by 1.2 kcal/mol per substitution, while xC stabilized it by a slightly smaller 1.0 kcal/mol. Thus, combining the data with previous measurements,[9] the relative stacking abilities of the four expanded bases are xA (2.2 kcal/mol), xT (1.6), xG (1.2), and xC (1.0). This order is similar to that for the four natural bases: A (1.0), G (0.6), T (0.5), and C (0.4).[16d] However, the four size-expanded (xDNA) bases all stack more than twice as strongly as their natural (DNA) counterparts.

TABLE 1. Thermodynamic Data for 12mer DNA Duplexes Containing a Size-expanded Base Pair or Mismatch (X−Y) in a Central Position[ab]

base pair X−Y	T_m[c] (°C)	$\phi G°_{37}$[d] (kcal/mol)
C−G	43.1 ± 0.5	−9.7 ± 0.8
G−C	45.6 ± 0.5	−10.4 ± 0.8
xC−G	41.2 ± 0.5	−9.4 ± 0.2
xC−A	25.0 ± 0.5	−5.9 ± 0.4
xC−C	29.9 ± 0.5	−6.6 ± 0.2
xC−T	28.5 ± 0.5	−6.7 ± 0.2
xC−ϕ[e]	26.5 ± 0.5	−6.0 ± 0.4
G−xC	41.1 ± 0.5	−9.4 ± 0.2
A−xC	23.1 ± 0.5	−5.6 ± 0.2
C−xC	26.4 ± 0.5	−6.0 ± 0.4
T−xC	26.0 ± 0.5	−5.9 ± 0.2
ø−xC	21.8 ± 0.5	−5.4 ± 0.2
C−G	43.1 ± 0.5	−9.7 ± 0.9
G−C	45.6 ± 0.5	−10.4 ± 1.0
xG−C	36.0 ± 0.5	−8.2 ± 0.8
xG−A	25.9 ± 0.5	−6.1 ± 0.6
xG−G	28.7 ± 0.5	−7.0 ± 0.7
xG−T	28.8 ± 0.5	−6.3 ± 0.6
xG−ϕ	28.3 ± 0.5	−6.4 ± 0.6
C−xG	40.3 ± 0.5	−9.3 ± 0.9
A−xG	28.6 ± 0.5	−6.6 ± 0.6
G−xG	32.2 ± 0.5	−7.2 ± 0.7
T−xG	31.1 ± 0.5	−7.1 ± 0.7
ø−xG	32.8 ± 0.5	−7.3 ± 0.7

[a] Conditions: 100 mM NaCl, 10 mM $MgCl_2$, 10 mM PIPES·Na (pH 7.0). [b] Sequence is 5′-AAGAAXGAAAAG·5′-CTTTTCYTTCTT. [c] T_m values are for 5.0 μM oligonucleotide. [d] Averages of values from van't Hoff and curve fitting methods. [e] "ø" is tetrahydrofuran abasic analogue.

TABLE 2. Thermodynamic Data for Self-Complementary Duplexes Containing a 5′ Dangling Base[a]

dangling base	T_m[b] (°C)	$\Delta G°_{37}$[c] (kcal/mol)	$\Delta\Delta G°_{37}$[d] (kcal/mol)
core sequence 5′-dCGCGCG			
(none)	41.7 ± 0.5	−8.1 ± 0.2	
C[e]	46.2 ± 0.5	−9.0 ± 0.2	−0.5 ± 0.2
G[e]	51.5 ± 0.5	−9.6 ± 0.3	−0.8 ± 0.3
xC	53.6 ± 0.5	−10.1 ± 0.3	−1.0 ± 0.3
core sequence 5′-dACAGCTGT			
(none)	40.0 ± 0.5	−7.9 ± 0.1	
C	42.8 ± 0.5	−8.7 ± 0.1	−0.4 ± 0.2
G	44.7 ± 0.5	−8.9 ± 0.1	−0.5 ± 0.2
xG	48.9 ± 0.5	−10.2 ± 0.1	−1.2 ± 0.2

[a] Conditions: 1 M NaCl, 10 mM phosphate (pH 7.0) with 0.1 mM EDTA. [b] T_m values are for 5.0 μM oligonucleotide. [c] Averages of values from van't Hoff and curve fitting methods. [d] Values per each dangling base, relative to the core duplex. [e] Data from ref 15d.

Discussion

Our studies demonstrate that syntheses of xC and xG nucleosides can be accomplished without severe difficul-

(15) Millican, T. A.; Mock, G. A.; Chauncey, M. A.; Patel, T. P.; Eaton, M. A.; Gunning, J.; Cutbush, S. D.; Neidle, S.; Mann, J. *Nucleic Acids Res.* **1984**, *12*, 7435−7453.

(16) (a) Turner, D. H.; Sugimoto, N.; Freier, S. M. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 167−192. (b) SantaLucia, J.; Allawi, H. T.; Seneviratne, P. A. *Biochemistry* **1996**, *35*, 3555−3562. (c) Guckian, K.; Schweitzer, B. A.; Ren, R. X.-F.; Sheils, C. J.; Paris, P. L.; Tahmassebi, D. C.; Kool, E. T. *J. Am. Chem. Soc.* **1996**, *118*, 8182−8183. (d) Guckian, K. M.; Schweitzer, B. A.; Ren, R. X.-F.; Sheils, C. J.; Tahmassebi, D. C.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 2213−2222. (e) Nakano, S.; Uotani, Y.; Nakashima, S.; Anno, Y.; Fujii, M.; Sugimoto, N. *J. Am. Chem. Soc.* **2003**, *125*, 8086−8087. (g) Lai, J. S.; Qu, J.; Kool, E. T. *Angew. Chem., Int. Ed.* **2003**, *42*, 5973−5977.

ties and that the nucleotides can be incorporated intact into DNA after deprotection. The synthesis of the xC nucleoside was achieved in straightforward fashion and in relatively high yields. It conveniently makes use of the previous xT nucleoside as its starting point.[7] By contrast, the preparation of the dxG nucleoside is long and rather low-yielding. In considering possible routes to this previously unknown compound we considered direct coupling of a base precursor to a deoxyribose derivative rather than a ribose precursor, to save steps in deoxygenation of the 2′ position. However, we chose to adopt the longer procedure to take advantage of the stereoselectivity and yield in the more efficient glycosylation with tetraacetylribose. It remains to be seen whether shorter routes can be developed to increase access to this compound.

The pairing studies of xC and xG in the context of short DNAs provide new insights into understanding the flexibility of DNA backbones to accommodate larger-size base pairs. Parallel to previous results from xT and xA investigations,[9] the current studies showed that base pairs involving xC or xG generally destabilized the duplex stability by 0.3−1.3 kcal/mol. For xC in particular, the small 0.3 kcal/mol of destabilization is less than that observed for the other three stretched bases. Without further structural studies it remains unclear what the origin of this small difference is.

The observed pairing selectivities of xG for C and xC for G suggest formation of hydrogen bonded pairs within the DNA (Figure 1c), although some caution on this point should be retained until structural studies can confirm this. Presumably there is an adjustment in backbone conformation between these larger pairs and the neighboring natural ones to accommodate this transition, and this suboptimal conformation is reflected in the small degree of destabilization. We conjecture that the inherent backbone distortion cost may in fact be larger than this, but may be compensated for, to some degree, by the considerably stronger stacking of xC and xG as compared to C and G.

The completion of the xDNA nucleoside set has allowed the completion of measurement of their stacking tendencies in DNA. The data show that each expanded nucleobase stacks more favorably than its analogue by more than a factor of 2 in free energy. We hypothesize that this high affinity arises from two factors: first, the added size, which comes from the addition of the benzene ring to each molecule. One face of a benzo ring has a surface area of roughly 7.5 Å^2, increasing overall surface area of a given base by ca. 50−100%. Increased area of overlap with neighboring bases would aid stacking by increasing polarizability, thus enhancing van der Waals interactions. In addition, the addition of a hydrocarbon benzoring to a more polar heterocycle is expected to have the effect of decreasing overall polarity. This makes it more likely that solvophobic effects would also contribute favorably to stacking.[16] A simple calculation of polarities as represented by logP values for the four xDNA bases gave values of −0.02, +0.91, −0.31, and +0.69 (for xA, xT, xG, and xC, respectively), which is much less polar than the natural bases, which gave calculated values of −1.07, −0.36, −1.36, and −0.76 (for A, T, G, and C, respectively). Thus, the benzo group is expected to make a ca. 1.1 log unit difference in polarity for the purines and 1.3−1.5 log unit difference in the smaller pyrimidines.

As for the selectivity of base pairing by xC and xG bases in the present natural DNA context, the data show that both new bases do show significant selectivity. The degree of selectivity of xG is close to the magnitudes of selectivities seen previously for xT and xA, and these values are smaller than the selectivities measured for natural DNA bases. We surmise that the lower selectivity of the expanded bases in natural DNA arises from the fact that even the fully matched pairs cause some distortion (and therefore an energetic penalty) in the duplex. This is unlike natural Watson−Crick pairs, where matched pairs cause no distortion of the helix, while mismatched ones result in a distortion penalty. This idea is also supported by recent studies of the pairing selectivity of xT and xA bases in the context of *fully expanded* DNAs.[10] In that context, xT and xA should cause little distortion when correctly paired, and the data show that they do, in fact, give selectivity that is in most cases as high as, or even higher than, that of natural bases in natural DNA. Thus, we conclude that selectivity of pairing of expanded or natural bases does not arise from the sizes of the bases alone, but rather from the compatibility of pairs with their surrounding context.

Finally, the successful preparation of the second half of the xDNA nucleoside set should make possible a number of future experiments in nucleic acid recognition. One of these is the possibility that xDNA strands composed only of size-expanded nucleosides might be used to bind natural DNA or RNA sequences. xDNA strands are promising in this respect because expanded-size helices can be considerably more stable than natural duplexes.[10,17] Thus, the xDNA strands might bind even to sequences that are involved in secondary structures. The addition of the dxC and dxG building blocks to this nucleotide set should make such targeting general for all possible nucleic acid sequences. Second, xDNA bases are inherently fluorescent, and it is possible that such fluorescent bases might be useful as self-reporters of binding in mechanistic, diagnostic, or imaging applications. Third, a set of four xDNA bases might be combined with the natural four bases to make an eight-base genetic set that encodes more information than the natural genetic system. Studies are underway to test all these possibilities.

Acknowledgment. This work was supported by the National Institutes of Health (GM63587). H.L. and J.G. acknowledge support from Stanford Graduate Fellowships.

Supporting Information Available: Experimental details of nucleoside synthesis and characterization; oligonucleotide synthesis and characterization; thermal denaturation methods; fluorescence spectrometry methods. This material is available free of charge via the Internet at http://pubs.acs.org.

JO048357Z

(17) Lu, H.; Kool, E. T. *Angew. Chem., Int. Ed.* **2004**, *43*, 5834−5836.