

Artificial Genetic Systems: Exploiting the “Aromaticity” Formalism To Improve the Tautomeric Ratio for Isoguanosine Derivatives

Theodore A. Martinot and Steven A. Benner*

Department of Chemistry, University of Florida,
P.O. Box 117200, Gainesville, Florida 32611-7200

benner@chem.ufl.edu

Received February 4, 2004

Abstract: The tautomerism of 2'-deoxy-7-deaza-isoguanosine (**2**) was studied and compared to that of 2'-deoxy-isoguanosine (**1**). The fixed 1*N*-methyl (**8**) and *O*-methyl (**4**) derivatives were synthesized to represent the pure extremes of each tautomer. The replacement of the imidazole ring in **1** with a pyrrole ring in **2** makes the keto form in the latter more favored by 2 orders of magnitude (K_{TAUT} for **2** $\approx 10^3$, as opposed to K_{TAUT} for **1** ≈ 10).

It has been a decade since it was shown that the geometry of the Watson–Crick nucleobase pair can accommodate as many as 16 nucleobases forming eight mutually exclusive pairs, to yield an artificially expanded genetic information system (AEGIS).¹ By providing molecular recognition on demand in aqueous solution, similar to nucleic acids but with a coding system that is orthogonal to the system in DNA and RNA, AEGIS today enables clinical assays such as Bayer’s branched DNA diagnostics tool that monitors the load of viruses in patients infected with HIV and hepatitis C viruses.² Both assays are FDA-approved and are widely used to provide personalized patient care in the clinic. Further, AEGIS components enable an assay for the early detection of the SARS virus.³ With the emergence of the first six-letter PCR,⁴ AEGIS now supports the development of a synthetic biology where higher order processes of living systems, including reproduction and Darwinian evolution, are duplicated by artificial, designed chemical systems.

One issue remaining before a synthetic biology based on AEGIS is fully implemented arises from the tautomeric form displayed by isoguanosine (and 2'-deoxy-isoguanosine, disoG, **1**). In its major keto form (Figure 1), disoG implements a hydrogen bond donor–donor–acceptor pattern (proceeding from the major to the minor groove) on a purine heterocycle (Figure 1). A minor enolic tautomer of disoG is known to be present in aqueous solution to perhaps 10%,⁵ however. The enolic tautomer

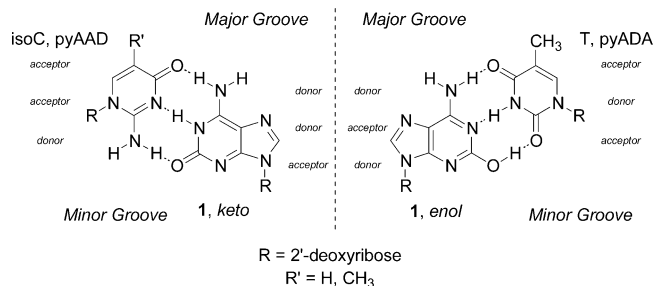


FIGURE 1. Isoguanosine (**1**), a purine (pu), can generate two possible tautomers, an *enol* tautomer (giving a puDAD form, which will pair with T or U) or a *keto* tautomer (giving a puDDA form, which will pair with isoC). The extent of this tautomerism has been reported to be 10:1 in favor of the keto form.

presents the donor–acceptor–donor hydrogen bonding pattern that is complementary to the thymidine and uridine heterocycles.⁶ Although the presence of the minor tautomer does not adversely affect the use of isoguanosine in clinical diagnostics, it does inconvenience some polymerases that prefer to place thymidine (T) and/or uridine (U), rather than isocytidine (isoC), opposite isoguanosine in a template.⁷

Given that the keto forms of pyrimidinones normally dominate over enolic forms, we viewed the enolic tautomer of disoG as an exception in need of explanation. We reasoned that the enolic form was present in appreciable concentration in disoG because it restores formal aromaticity to the five-membered imidazole ring, which must be cross-conjugated in the keto tautomer (Figure 1). It is well-known that pyrrole is less “aromatic” than imidazole, with the former being 59% as aromatic as benzene, the latter being 64% (compared to pyridine at 86%, thiophene at 66%, and furan at 43%).⁸ This suggested that replacement of the imidazole ring of disoG by a pyrrole ring to give 7-deaza-isoguanine might generate a purine analogue that would also implement the puDDA hydrogen bonding pattern (like disoG), but with less of the enolic tautomer present at equilibrium.

7-Deaza-isoguanine (C7isoG) and its 2'-deoxynucleoside analogue (dC7isoG, **2**) are both known.^{9–11} The crystal structure of dC7isoG¹² and its base-pairing ability have both been studied.^{13–15} The tautomerism of both C7isoG and dC7isoG remain undetermined, however.

(6) Robinson, H.; Gao, Y.-G.; Bauer, C.; Roberts, C.; Switzer, C.; Wang, A. H.-J. *Biochemistry* **1998**, *37*, 10897–10905.

(7) Switzer, C. Y.; Moroney, S. E.; Benner, S. A. *Biochemistry* **1993**, *32*, 10489–10496.

(8) Katritzky, A. R.; Pozharskii, A. F. *Handbook of Heterocyclic Chemistry*, 2nd ed.; Pergamon Press: Oxford, 2000.

(9) Seela, F.; Menkhoff, S.; Behrendt, S. *J. Chem. Soc., Perkin Trans. 2* **1986**, 525–530.

(10) Kazimierczuk, Z.; Mertens, R.; Kawczynski, W.; Seela, F. *Helv. Chim. Acta* **1991**, *74*, 1742–1748.

(11) Seela, F.; Muth, H.-P.; Kaiser, K.; Bourgeois, W.; Muehlegger, K.; Von der Eltz, H.; Batz, H.-G. U.S. Patent 6,211,158 B1, 2001.

(12) Seela, F.; Wei, C.; Reuter, H.; Kastner, G. *Acta Crystallogr.* **1999**, *C55*, 1335–1337.

(13) Seela, F.; Wei, C. *Chem. Commun.* **1997**, 1869–1870.

(14) Seela, F.; Wei, C. *Helv. Chim. Acta* **1999**, *83*, 726–745.

(15) Seela, F.; Wei, C.; Becher, G.; Zulauf, M.; Leonard, P. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 289–292.

(1) Geyer, C. R.; Battersby, T. R.; Benner, S. A. *Structure* **2003**, *11*, 1485–1498.

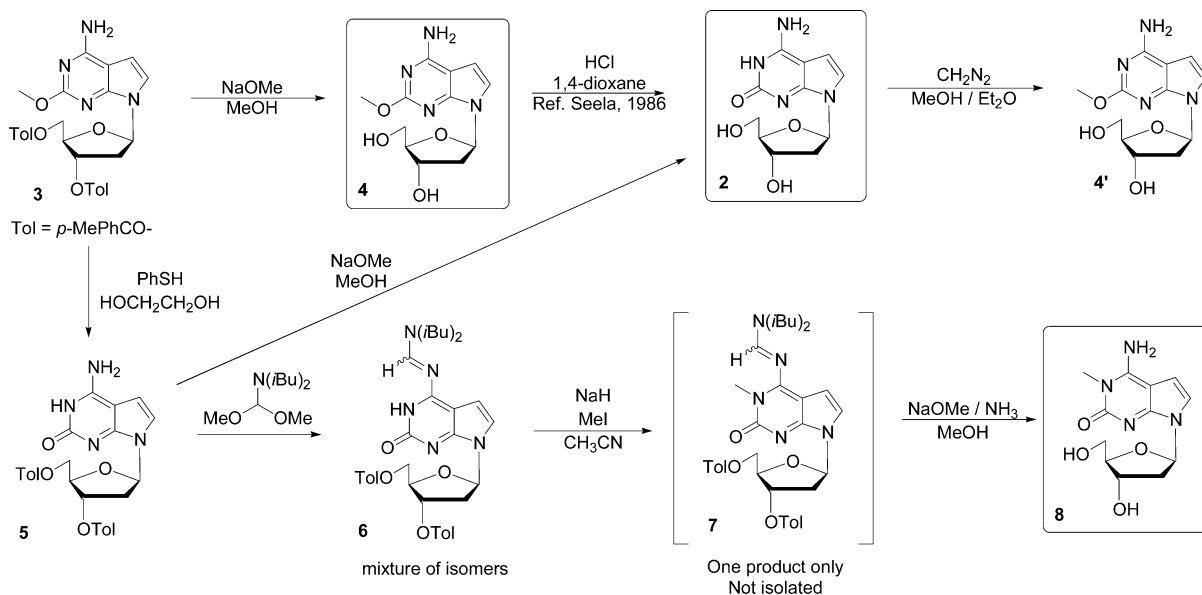
(2) Collins, M. L.; Irvine, B.; Tyner, D.; Fine, E.; Zayati, C.; Chang, C. A.; Horn, T.; Ahle, D.; Detmer, J.; Shen, L. P.; Kolberg, J.; Bushnell, S.; Urdea, M. S.; Ho, D. D. *Nucleic Acids Res.* **1997**, *25*, 2979–2984.

(3) EraGen: SARS Diagnostic: <http://www.eragen.com/diagnostics/sars.html>.

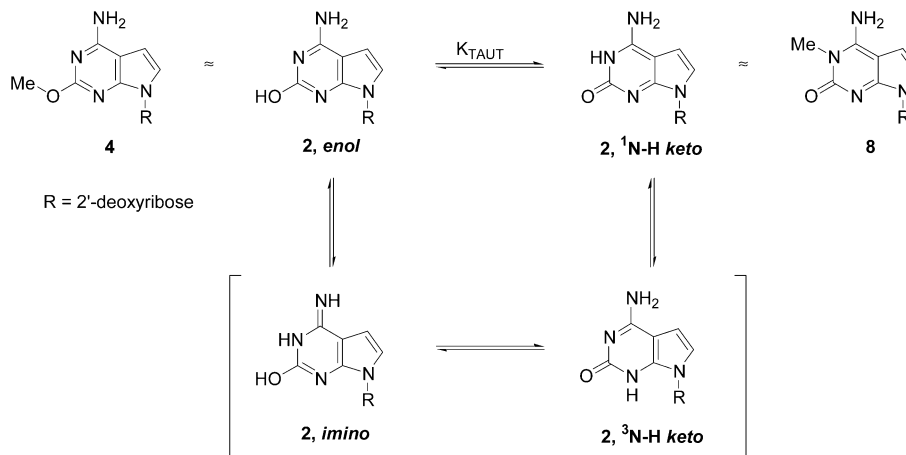
(4) Sismour, A. M.; Lutz, S.; Park, J.-H.; Lutz, M. J.; Boyer, P. L.; Hughes, S. H.; Benner, S. A. *Nucleic Acids Res.* **2004**, *32*, 728–735.

(5) Sepiol, J.; Kazimierczuk, Z.; Shugar, D. *Z. Naturforsch. C* **1976**, *31*, 361–370.

SCHEME 1



SCHEME 2



To test our hypothesis, the *O*-methyl-7-deaza-isoguanine derivative was prepared by the procedure of Davoll.¹⁶ Reaction with 1-(α -chloro-3,5-di-*O*-(*p*-toluoyl)-2-deoxy-D-ribose¹⁷ gave the 3',5'-ditoluoyl-protected 2'-deoxyriboside of 2-methoxy-6-amino-7-deaza-purine (**3**). Seela's procedure⁹ for the removal of the 2-methoxy group (HCl, 1,4-dioxane, 100 °C) did not work well in our hands (Scheme 1). We therefore used thiophenol in ethylene glycol¹⁸ to generate the di-*O*-(*p*-toluoyl)-derivative of **2** (**5**). This compound was converted to the 2'-deoxynucleoside using NaOMe in MeOH to afford **2**, which had spectroscopic properties identical to those reported for this compound by Seela et al.⁹ The structure of this compound was further confirmed by treatment with diazomethane, which generated **4'**. The structure of **4'** was identical in all aspects to that of **4**. The *N*-methyl derivative was prepared by protection of the exocyclic amino group as the diisobutylformamide^{19,20} (generating **6**), reaction with CH₃I (to give **7**), and immediate deprotection by

treatment with methanolic NH₃/NaOMe, affording **8**. *N*-Methylation at the ¹N position was confirmed by long-range (HMBC²¹) and short-range (HMQC²²) ¹H–¹³C NMR correlation spectroscopy.

The p*K*_a of the nucleoside analogue (**2**) was determined by spectrophotometric titration (pH 2.8 to 13.7)²³ at 240–350 nm to give values for p*K*₁ and p*K*₂ as 4.3 (±0.1) and 9.9 (±0.2), respectively (Supporting Information). These are comparable to the p*K*_a values of natural nucleobases and ensure that the C7isoG heterocycle is largely uncharged in neutral water.

Four tautomeric forms are worthy of consideration for 7-deaza-isoguanine (Scheme 2). On the basis of literature precedent, as well as precedent in other heterocycles, the imino form was considered to be the least likely.^{5,24–26}

(19) Zemlicka, J.; Holy, A. *Collect. Czech. Chem. Commun.* **1967**, *32*, 3159–3168.

(20) Froehler, B. C.; Matteucci, M. D. *Nucleic Acids Res.* **1983**, *11*, 8031–8036.

(21) Bax, A.; Summers, M. F. *J. Am. Chem. Soc.* **1986**, *108*, 2093–2094.

(22) Hurd, R. E.; John, B. K. *J. Magn. Reson.* **1991**, *91*, 648–653.

(23) Kazimierzuk, Z.; Shugar, D. *Acta Biochim. Pol.* **1974**, *21*, 455–463.

(16) Davoll, J. *J. Chem. Soc.* **1960**, 131–138.

(17) Hoffer, M. *Chem. Ber./Recl.* **1960**, *93*, 2777–2781.

(18) Wildes, J. W.; Martin, N. H.; Pitt, C. G.; Wall, M. E. *J. Org. Chem.* **1971**, *36*, 721–723.

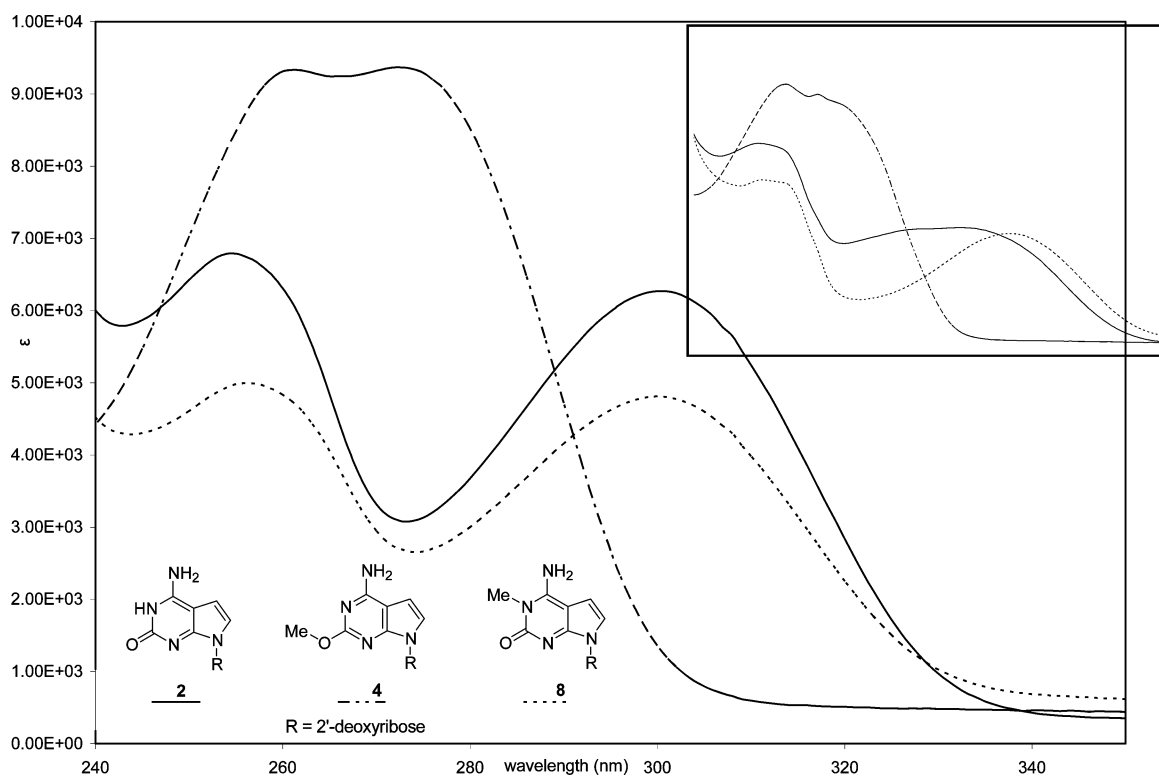


FIGURE 2. UV spectra of compounds **2**, **4**, and **8** in water. (Inset) UV spectra of the same compounds, in 99:1 1,4-dioxane/water.

This was confirmed by NMR: the -NH_2 group was easily identified in the NMR by integration and D_2O exchange.

The UV spectra of 2'-deoxy-7-deaza-isoguanosine (**2**) in aqueous solution resembles closely that of the N-methylated derivative (**8**) and not the O-methylated derivative (**4**) (Figure 2). This similarity suggests but does not prove that the keto tautomer predominates, following the same logic used by Sepiol et al. to infer that the keto tautomer predominates for isoguanosine.⁵

The multiwavelength analysis described by Dewar and Urch²⁷ was used to suggest that, at most, one additional tautomer was present as a major contributor to the mixture (Supporting Information).

The procedure of Voegel et al.²⁸ was then used to estimate a value of the keto:enol tautomers in pure water for dC7isoG. This procedure exploits the fact that in nonpolar solvents the UV spectra for both disoG and C7isoG shift from their form in pure water (which resembles the N-methylated derivatives) to a form that resembles the UV spectra of the O-methylated derivatives. This is interpreted as evidence that the tautomeric equilibrium shifts from one favoring the keto tautomer in pure water to one favoring the enol tautomer in pure dioxane.

To obtain quantitative data, the ratio of the extinction coefficients at two wavelengths ($\lambda = 296$ and 255 nm for disoG and $\lambda = 305$ and 254 nm for dC7isoG; Supporting

Information), chosen to maximize the difference between the keto and enol forms, was chosen as a metric for the tautomeric equilibrium constant in dioxane/water mixtures in varying proportions. This was plotted against the Dimroth $E_T(30)$ value, which provides a measure of the local dielectric constant.^{29–33}

Even qualitatively, disoG and dC7isoG behave differently in these experiments. The UV spectrum of disoG changes well before the water is completely removed. In contrast, the UV spectrum of dC7isoG is not identical to that of the O-methylated derivative even at the highest concentrations of dioxane tested. Further, the UV spectrum of disoG continues to change even in mixtures approaching pure water; this suggests that the conversion of the enolic tautomer to the keto tautomer of disoG is not complete even in pure water (the conclusion of Sepiol et al.⁵). With dC7isoG, however, the UV spectrum ceases to be solvent-dependent when the fraction of water is greater than 50%. These results suggest that our hypothesis at outset had manipulative value: the keto tautomer of dC7isoG is more stable relative to its enolic tautomer than the keto tautomer of disoG is relative to its enolic tautomer.

We then estimated the $K_{\text{TAUT}} (= [\text{keto}]/[\text{enol}])$ for dC7isoG using the method of Voegel et al.²⁸ Here, the fraction of enol form was estimated over a range of $E_T(30)$ values where the $K_{\text{TAUT}} \approx 1$, regions where an

(24) Katritzky, A. R.; Lagowski, J. M. *Adv. Het. Chem.* **1963**, *1*, 339–437.

(25) Cox, R. H.; Bothner-By, A. A. *J. Phys. Chem.* **1968**, *72*, 1642–1645.

(26) Cox, R. H.; Bothner-By, A. A. *J. Phys. Chem.* **1968**, *72*, 1646–1649.

(27) Dewar, M. J. S.; Urch, D. S. *J. Chem. Soc.* **1957**, 345–347.

(28) Voegel, J. J.; von Krosigk, U.; Benner, S. A. *J. Org. Chem.* **1993**, *58*, 7542–7547.

(29) Kosower, E. M. *J. Am. Chem. Soc.* **1958**, *80*, 3253–3260.

(30) von Dimroth, K.; Reichardt, C.; Siepmann, T.; Bohlmann, F. *Liebigs Ann. Chem.* **1963**, *661*, 1–37.

(31) Reichardt, C.; Harbusch-Goernert, E. *Liebigs Ann. Chem.* **1983**, *721*–743.

(32) Gordon, A.; Katritzky, A. R. *Tetrahedron Lett.* **1968**, *23*, 2767–2770.

(33) von Jouanne, J.; Palmer, D. A.; Kelm, H. *Bull. Chem. Soc. Jpn.* **1978**, *51*, 463–465.

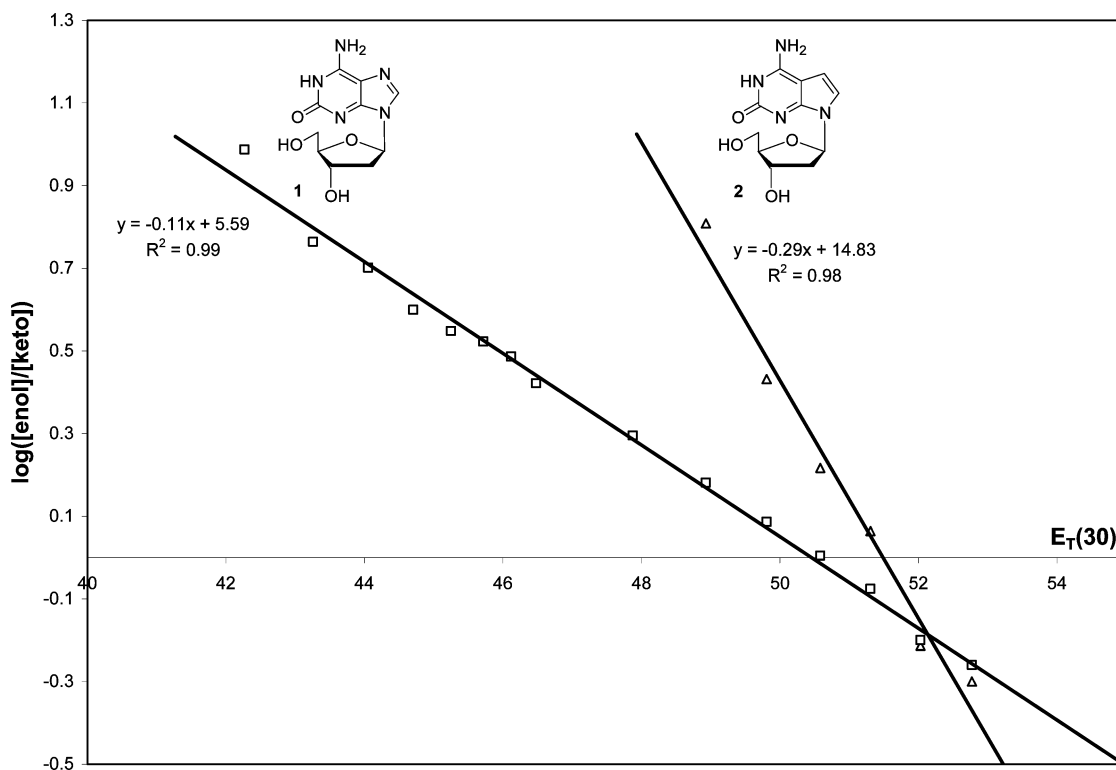


FIGURE 3. Plot of $\log([\text{enol}]/[\text{keto}])$ versus $E_T(30)$ for dC7isoG (**2**) and isoG (**1**) in mixtures of dioxane ($E_T(30) = 36.0$) and water ($E_T(30) = 63.1$).

estimate could be made with some reliability. This is the region where $E_T(30) \approx 40\text{--}50$. The log fraction of enol was plotted against $E_T(30)$ (Figure 3), and the line was extrapolated to the $E_T(30)$ of pure water (63.1).³¹ This generated a value of $K_{\text{TAUT}} \approx 10^3$. Recalculating the K_{TAUT} for isoG gave a value of ≈ 10 , consistent with that previously reported. This can be compared with the value for the natural guanosine nucleoside ($K_{\text{TAUT}} \approx 10^4\text{--}10^5$).³⁴

With this result, we have now fixed each of the problems identified in AEGIS components exploiting the full range of hydrogen bonding patterns. Adjusting the electron-withdrawing groups has generated a pyDDA derivative that is free of epimerization.³⁵ Adjusting side chains of the pyAAD system has managed deamination issues.^{36–40} Adjusting the ring nitrogens has corrected the problematic pK_a values of bases of acidic purines.^{41–44} In the work presented here, adjustment of the aromaticity

of rings has resolved a tautomerism issue.⁴⁵ Thus, with this paper, we have the ingredients for a fully functional 12-letter genetic alphabet that meets the specifications met by natural nucleic acids.

Acknowledgment. The authors are grateful to the National Institutes of Health (GM54048) and the University of Florida (fellowship for T.M.) for funding. Skillful assistance of Sonia K. Reback in the preparation of starting materials is appreciated.

Supporting Information Available: Experimental procedures, characterization data, HPLC traces, and experimental details for the pK_a measurements and tautomerization data. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JO0497959

(34) Topal, M. D.; Fresco, J. R. *Nature* **1976**, *263*, 285–289.
 (35) Hutter, D.; Benner, S. A. *J. Org. Chem.* **2003**, *68*, 9839–9842.
 (36) Kodra, J. T.; Benner, S. A. *Synlett* **1997**, 939–940.
 (37) Jurczyk, S. C.; Kodra, J. T.; Rozzell, J. D.; Benner, S. A.; Battersby, T. R. *Helv. Chim. Acta* **1998**, *81*, 793–811.
 (38) Jurczyk, S. C.; Kodra, J. T.; Park, J. H.; Benner, S. A.; Battersby, T. R. *Helv. Chim. Acta* **1999**, *82*, 1005–1015.
 (39) Switzer, C.; Moroney, S. E.; Benner, S. A. *J. Am. Chem. Soc.* **1989**, *111*, 8322–8323.
 (40) Tor, Y.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 4461–4467.

(41) Jurczyk, S. C.; Horlacher, J.; Devine, K. G.; Benner, S. A.; Battersby, T. R. *Helv. Chim. Acta* **2000**, *83*, 1517–1524.
 (42) Lutz, M. J.; Held, H. A.; Hottiger, M.; Hubscher, U.; Benner, S. A. *Nucleic Acids Res.* **1996**, *24*, 1308–1313.
 (43) Rao, P.; Benner, S. A. *J. Org. Chem.* **2001**, *66*, 5012–5015.
 (44) Piccirilli, J. A.; Moroney, S. E.; Benner, S. A. *Biochemistry* **1991**, *30*, 10350–10356.

(45) It is clear that the $E_T(30)$ value at the interior of the DNA helix is not known; our current work is looking into whether the behavior of dC7isoG in water will correlate to its behavior in DNA.