VOLUME 122, NUMBER 32
AUGUST 16, 2000

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Efforts toward Expansion of the Genetic Alphabet: Optimization of Interbase Hydrophobic Interactions

Yiqin Wu, Anthony K. Ogawa, Markus Berger, Dustin L. McMinn, Peter G. Schultz,* and Floyd E. Romesberg*

Contribution from the Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received March 20, 2000

Abstract: Six novel unnatural nucleobases have been characterized that form stable base pairs in duplex DNA, relying not on hydrogen bonds, but rather on interbase hydrophobic interactions. These nucleobases are derivatives of the hydrophobic base pair between 7-azaindole (7AI) and isocarbostyril (ICS). Derivatives of 7AI and ICS were examined that have increased hydrophobic surface area, as well as increased polarizability. As observed with 7AI and ICS, these derivatives are recognized as substrates by Klenow fragment of *Escherichia coli* DNA polymerase I. The unnatural base pair between pyrrolopyrizine (PP) and C3-methylisocarbostyril (MICS) is enzymatically incorporated into DNA with high efficiency ($k_{cat}/K_M = 10^6\ M^{-1}\ min^{-1}$) and moderate selectivity. These studies represent a significant step toward the generation of a stable, orthogonal base pair that can be enzymatically incorporated into DNA with good fidelity.

1. Introduction

The structure of duplex DNA is based on the complementary Watson−Crick hydrogen-bonding patterns of adenine with thymine (dA:dT base pair) and guanine with cytosine (dG:dC base pair).[1] Previous efforts to expand the number of bases available for information storage in DNA have used N- or C-glycosidic nucleosides with purine- or pyrimidine-like nucleobases that have unique patterns of hydrogen bond donors and acceptors.[2−11] For example, the hydrogen-bonding base pairs *κ*:X and d*iso-C*:d*iso-G* have been studied extensively. However, an approach based solely on unique H-bonding patterns limits the available unnatural nucleobases to fewer than those that can be constructed by synthetic chemists. Furthermore, this approach may be limited by the existence of relatively stable base tautomers that could mispair with native bases in duplex DNA or during polymerase-catalyzed DNA synthesis.[12−15] Consistent with this notion, it has not been possible to identify a DNA

* To whom correspondence should be addressed. Tel.: (858) 784-7290. Fax: (858) 784-7472. E-mail: floyd@scripps.edu.

(1) Kornberg, A.; Baker, T. A. *DNA Replication*, 2nd ed.; W. H. Freeman and Co.: New York, 1992.
(2) Lutz, M. J.; Held, H. A.; Hottiger, M.; Hübscher, U.; Benner, S. A. *Nucleic Acids Res.* **1996**, *24*, 1308−1313.
(3) Horlacher, J.; Hottiger, M.; Podust, V. N.; Hübscher, U.; Benner, S. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6329−6333.
(4) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33−37.
(5) Bain, J. D.; Switzer, C.; Chamberlin, A. R.; Benner, S. A. *Nature* **1992**, *356*, 537−539.
(6) Kovacs, T.; Otvos, L. *Tetrahedron Lett.* **1988**, *29*.
(7) Sugiyama, H.; Ikeda, S.; Saito, I. *J. Am. Chem. Soc.* **1996**, *118*, 9994−9995.
(8) Tor, Y.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 4461−4467.
(9) Lutz, M. J.; Horlacher, J.; Benner, S. A. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 499−504.
(10) Switzer, C.; Moroney, S. E.; Benner, S. A. *J. Am. Chem. Soc.* **1989**, *111*, 8322−8323.
(11) Switzer, C. Y.; Moroney, S. E.; Benner, S. A. *Biochemistry* **1993**, *32*, 10489−10496.
(12) Beaussire, J.-J.; Pochet, S. *Nucleosides Nucleotides* **1999**, *18*, 403−410.
(13) Roberts, C.; Chaput, J. C.; Switzer, C. *Chem., Biol.* **1997**, *4*, 899−908.

10.1021/ja0009931 CCC: $19.00
Published on Web 07/28/2000

polymerase that can replicate DNA containing these unnatural H-bonding base pairs.[2−4,9,11,16]

In contrast, the design of the novel base pairs described herein is based on hydrophobic packing and thus is not susceptible to effects of tautomerization. By analogy to protein biochemistry, the hydrophobic interactions between the bases are predicted to be strong and selective for hydrophobic pair formation.[17] Furthermore, mispairs between the unnatural and native bases should be disfavored due to the energetic cost of desolvation of the native base. In addition to stable and selective pairing, the unnatural hydrophobic bases must be enzymatically replicable. A DNA polymerase must be capable of recognizing the hydrophobic substrates, both in template and as a triphosphate, and efficiently incorporating them with high selectivity into DNA. Finally, the polymerase must efficiently continue primer extension after synthesis of the nascent unnatural base pair.

We have recently shown that the Watson−Crick H-bonds in duplex DNA can be replaced with intrabase hydrophobic interactions.[18,19] A variety of hydrophobic nucleobases have been shown to form base pairs in duplex DNA with a stability and selectivity equivalent to or greater than those of native base pairs.[18−20] Moreover, several hydrophobic bases have been identified which are efficiently incorporated into DNA with rates approaching those of natural base pairs. Among these, the base pair formed between the 7-azaindole (7AI) and isocarbostyril (ICS) nucleosides possessed thermodynamic and kinetic properties that made it attractive for further investigation. The 7AI:ICS base pair was only slightly less stable than an dA:dT base pair (T_m = 57.2 °C and 59.2 °C for 7AI:ICS and dA:dT, respectively). The base pair was also thermally selective against mispairing with natural bases; the most stable mispairs were 2.1 and 3.8 °C less stable for 7AI:ICS and dA:dT, respectively. However, while the 7AI:ICS pair was 1.7 °C thermally selective against the 7AI:7AI self-pair, which melted at 55.5 °C, it was not selective against the ICS:ICS self-pair which melted at 59.3 °C. Therefore, it seemed that the interbase interactions between 7AI and ICS were not yet optimal for selective pairing in duplex DNA.

In the context of enzymatic DNA replication, the 7AI:ICS base pair also exhibited properties that made it attractive for further optimization.[19] The rates of exonuclease-deficient Klenow fragment (KF)-mediated insertion of dICSTP opposite 7AI in the template strand, and d7AITP opposite ICS in the template strand, were both reasonably efficient, with the value of $k_{cat}/K_M = 2.7 \times 10^5$ and 1.8×10^5 M^{-1} min^{-1}, respectively. While these rates are roughly 100-fold slower than those typical of native pair synthesis, the rates for mispairing with natural dNTPs were significantly slower (less than 6×10^3 M^{-1} min^{-1}), resulting in reasonable fidelity. Preliminary results show that the 7AI:ICS base pair is extendable with moderate efficiency (10^4 M^{-1} min^{-1}). However, as was the case with thermal selectivity, the self-pairs constituted the major fidelity problem for the 7AI:ICS base pair. The 7AI:7AI and ICS:ICS self-pairs were synthesized nearly as efficiently as the 7AI:ICS pair.

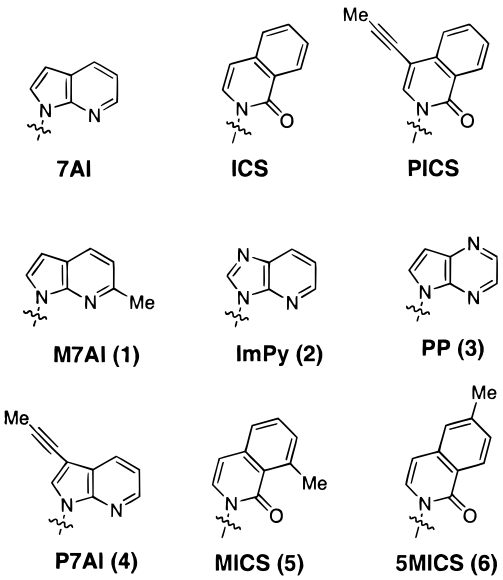
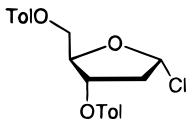


Figure 1. Hydrophobic nucleosides.



7

Figure 2.

However, unlike the 7AI:ICS pair, neither of the self-pairs was extended under any conditions examined. To optimize the unnatural base pair, we have synthesized and characterized a variety of derivatives of 7AI and ICS. The resulting base pairs between the 7AI derivatives and the ICS derivatives have been characterized thermodynamically in duplex DNA, and as substrates for KF.

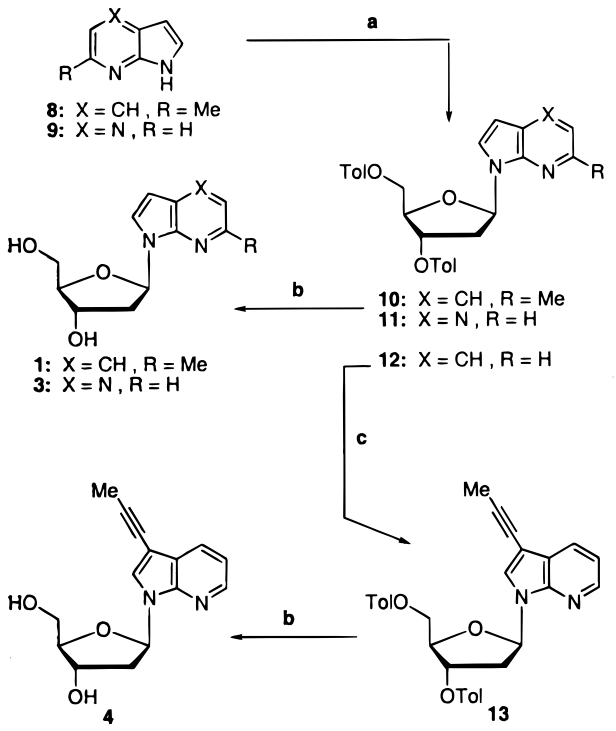
2. Results

2.1. Base Pair Design and Synthesis. In an effort to design 7AI and ICS analogues with improved thermodynamic and kinetic properties, the 7AI and ICS derivatives shown in Figure 1 were synthesized. The methyl-substituted derivatives were designed to probe the effects of minor groove (M7AI), major groove (5MICS), and interbase (MICS) packing. Propynyl-substituted derivatives (P7AI and PICS) were designed to increase nucleobase hydrophobic surface area, and nitrogen-substituted derivatives (ImPy and PP) were designed to modify the electronic nature of the nucleobase without significant structural perturbations.

The syntheses of nucleosides **1**−**6** involved two strategies for stereoselective glycosidic bond formation. In all cases, the bis-toluoyl-protected chloroglycoside **7**[21] (Figure 2) served the electrophile for condensation with each nucleobase.[22] Deprotection under basic methanolic conditions[22] afforded the respective nucleosides, which were converted to both nucleoside triphosphates[6] and 5′-trityl phosphoramidites for use in automated DNA synthesis (Scheme 3).[23] As previously reported, C1′-stereochemical assignment of **1**−**6** was based on NOESY data.[19]

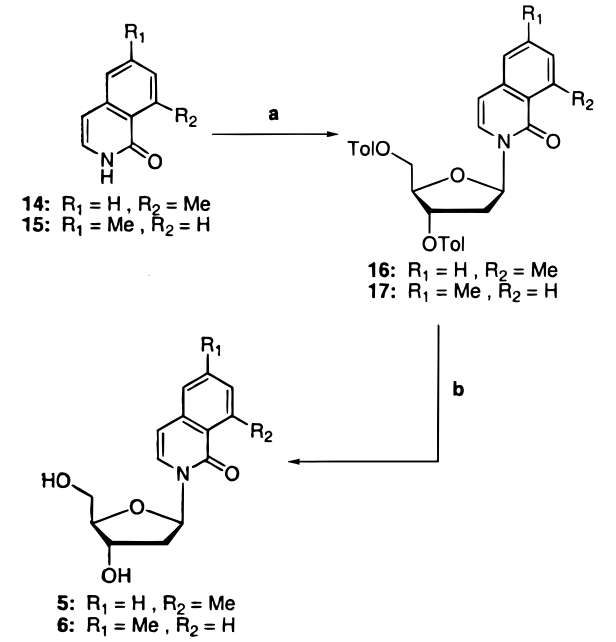
(14) Roberts, C.; Bandaru, R.; Switzer, C. *J. Am. Chem. Soc.* **1997**, *119*, 4640−4649.
(15) Robinson, H.; Gao, Y.-G.; Bauer, C.; Roberts, C.; Switzer, C.; Wang, A. H.-J. *Biochemistry* **1998**, *37*, 10897−10905.
(16) Lutz, M. J.; Horlacher, J.; Benner, S. A. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1149−1152.
(17) Dill, K. A. *Biochemistry* **1990**, *29*, 7133−7155.
(18) McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **1999**, *121*, 11585−11586.
(19) Ogawa, A. K.; Wu, Y.; McMinn, D. L.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 3274−3287.
(20) Berger, M.; Ogawa, A. K.; McMinn, D. L.; Wu, Y.; Schultz, P. G.; Romesberg, F. E. *Angew. Chem., Int. Ed.*, in press.
(21) Takeshita, M.; Chang, C.-N.; Johnson, F.; Will, S.; Grollman, A. P. *J. Biol. Chem.* **1987**, *262*, 10171−10179.
(22) Schweitzer, B. A.; Kool, E. T. *J. Org. Chem.* **1994**, *59*, 7238−7242.
(23) Schweitzer, B. A.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 1863−1872.

Scheme 1[a]



[a] Conditions: (a) NaH, then **7**; (b) NaOMe; (c) ICl, CuI, $Cl_2Pd(Ph_3)_2$, NEt_3, propyne, −78 °C to room temperature.

Scheme 2[a]

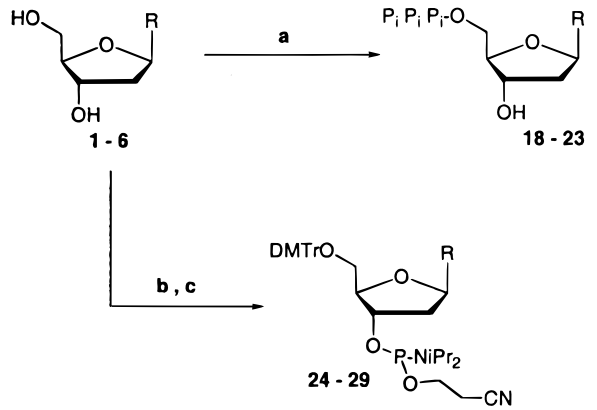


[a] Conditions: (a) bis-TMS-acetamide, then **7**, $SnCl_4$; (b) NaOMe.

Purine analogues **1**−**4** followed from condensing the nucleobase sodium salts onto **7** and deprotection using the aforementioned conditions (Scheme 1).[24] Specific to the synthesis of nucleoside **4**, 7-azaindole nucleoside **12** was first converted to propynyl-substituted **13** prior to deprotection.[18] Isocarbostyril analogues **5** and **6** were synthesized via Vorbruggen glycosylation methodology (Scheme 2) and deprotected as above.[25]

2.2. Thermodynamic and Kinetic Properties of 7AI:ICS Base Pair Derivatives. So that we could evaluate the stability of the hydrophobic bases, they were incorporated into the complementary oligonucleotides, 5′-GCGATGXGTAGCG-3′ and 5′-CGCTACYCATCGC-3′, where X and Y are either unnatural or natural bases. The relative stability of the base pairs, as well as the mispairs with natural bases, was determined in this sequence context by temperature-dependent absorption changes as measured with a Cary 300 Bio UV/vis spectrophotometer. The absorption at 260 nm was monitored while the samples were heated at a rate of 0.5 °C/min. We use the terms "thermal stability" or "thermodynamic stability" of the X:Y base pair to refer to the stability of the duplex as measured by the duplex melting temperature (T_m) with oligonucleotides containing X and Y at the indicated positions. The stability difference between a given base pair and the most stable mispair with another base is referred to as thermal or thermodynamic selectivity.

Scheme 3[a]



[a] Conditions: (a) $POCl_3$, 0 °C, then Bu_3N-PP_i; (b) DMTr-Cl; (c) $ClP(NiPr_2)(OCH_2CH_2CN)$, NEt_3.

To facilitate the presentation of the data, the following results are divided into two sections. The 7AI derivatives paired only with ICS are discussed first, followed by an analysis of ICS derivatives paired with 7AI. The base pairs, in which both 7AI and ICS have been modified, are discussed only where important. In each case, the thermal stability and selectivity of the base pairs are first presented, followed by a discussion of base pair recognition by KF.

2.2.1. Thermodynamic Properties of 7AI Derivatives. As mentioned above, the 7AI:ICS unnatural base pair is 2.0 °C less stable than a dA:dT base pair in the examined sequence context.[19] The thermal selectivity of the unnatural base pair (2.1 °C, relative to the most stable mispair) is also somewhat compromised relative to that of dA:dT (3.8 °C).[19] In an effort to improve the interbase hydrophobic packing, the nucleoside derivative, 6-methyl-7-azaindole (M7AI), was synthesized and characterized. The methyl group is expected to occupy space in the minor groove of the DNA duplex and may pack efficiently opposite another hydrophobic base. When the unnatural base is incorporated opposite ICS, PICS, MICS, or 5MICS, C6-methyl substitution is found to increase duplex stability by 0.8−2.7 °C (Table 1). Thus, the stability of the pairs formed between M7AI and ICS derivatives is comparable with that of a native dA:dT base pair in the same sequence context. However, the mispairs between M7AI and the native bases are also stabilized relative to the mispairs with 7AI by approximately the same amount. This results in little change in thermal selectivity of M7AI:ICS, as compared to that of 7AI:ICS. Furthermore, the M7AI:M7AI self-pair is 2.5 °C more stable than the 7AI:7AI

(24) Seela, F.; Gumbiowski, R. *Heterocycyles* **1989**, *29*, 795−805.
(25) Niedballa, U.; Vorbruggen, H. *J. Org. Chem.* **1974**, *39*, 3654−3660.

Table 1. T_m Values for Duplex-Containing 7AI Derivatives[a]

5′-dGCGTACXCATGCG
3′-dCGCATGYGTACGC

X	Y	T_m (°C)	X	Y	T_m (°C)
A	T	59.2	G	C	61.8
7AI	7AI	55.5	ImPy	ImPy	52.7
	ICS	57.2		ICS	54.7
	PICS	57.6		PICS	57.3
	MICS	59.7		MICS	59.7
	PIM	62.5		PIM	60.0
	5MICS	57.3		5MICS	56.6
	A	52.5		A	52.3
	T	50.5		T	50.0
	G	51.5		G	51.0
	C	48.5		C	46.3
M7AI	M7AI	58.0	PP	PP	51.3
	ICS	58.7		ICS	56.3
	PICS	58.6		PICS	56.3
	MICS	60.5		MICS	58.5
	PIM	59.7		PIM	57.2
	5MICS	60.0		5MICS	56.3
	A	54.2		A	50.0
	T	52.2		T	50.8
	G	52.5		G	50.5
	C	49.5		C	46.3
P7AI	P7AI	56.3			
	ICS	56.8			
	PICS	58.6			
	MICS	59.3			
	PIM	61.2			
	5MICS	57.9			
	A	51.2			
	T	50.3			
	G	50.2			
	C	45.7			

[a] See text for experimental details.

Table 2. Steady-State Kinetic Parameters for KF Exo-Mediated Synthesis of DNA with 7AI, M7AI, and P7AI in the Template[a]

5′-dTAATACGACTCACTATAGGGAGA
3′-dATTATGCTGAGTGATATCCCTCTXGTCA

template (X)	nucleoside triphosphate	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} min^{-1})
7AI	7AI	7.6 ± 0.4	34 ± 5	2.2×10^5
	M7AI	1.5 ± 0.1	124 ± 23	1.2×10^4
	ICS	7.4 ± 0.3	27 ± 3	2.7×10^5
	PICS	2.9 ± 0.2	7.5 ± 2.0	3.9×10^5
	MICS	22 ± 1	15 ± 1	1.5×10^6
	5MICS	3.3 ± 0.2	26 ± 2.0	1.3×10^5
	A	0.32 ± 0.02	54 ± 10	5.9×10^3
	G	0.08 ± 0.01	49 ± 13	1.6×10^3
	C	0.17 ± 0.02	31 ± 11	5.5×10^3
	T	0.23 ± 0.05	146 ± 40	1.6×10^3
M7AI	M7AI	0.42 ± 0.03	67 ± 12	6.3×10^3
	7AI	2.2 ± 0.2	29 ± 8	7.6×10^4
	ICS	3.4 ± 0.1	29 ± 2	1.2×10^5
	PICS	0.9 ± 0.1	5.1 ± 1.2	1.8×10^5
	MICS	20 ± 1	19 ± 2	1.1×10^6
	5MICS	2.0 ± 0.1	32 ± 2	6.3×10^4
	A	0.11 ± 0.01	28 ± 11	3.9×10^3
	G	0.06 ± 0.01	38 ± 10	1.5×10^3
	C	0.03 ± 0.01	30 ± 10	1.0×10^3
	T	0.17 ± 0.02	76 ± 24	2.2×10^3
P7AI	P7AI	11.2 ± 0.1	8.1 ± 0.3	1.4×10^6
	ICS	2.9 ± 0.1	26 ± 3	1.1×10^5
	PICS	1.8 ± 0.1	5.3 ± 0.9	3.4×10^5
	MICS	8.0 ± 0.3	18 ± 2	4.4×10^5
	5MICS	0.9 ± 0.2	31 ± 15	2.9×10^4
	A	0.41 ± 0.02	36 ± 8	1.1×10^4
	G	1.1 ± 0.1	23 ± 4	4.8×10^4
	C	*2.3 ± 0.1*	*0.5 ± 0.3*	*4.6×10^6*
	T	1.1 ± 0.1	56 ± 20	2.0×10^4

[a] See text for experimental details.

self-pair, which further limits the thermal orthogonality of the M7AI nucleobase.

The effect of C3-propynyl substitution of the azaindole ring was examined with propynyl 7-azaindole (P7AI). The duplex containing the P7AI:ICS base pair is destabilized 0.4 °C relative to the 7AI:ICS base pair. All of the mispairs with native bases are also destabilized (by 0.2−2.8 °C). In the case of 7AI, the most stable mispairs are 7AI:dA and 7AI:dG. With propynyl substitution, both of these mispairs are more destabilized than the pairs with ICS or ICS derivatives. Therefore, compared to 7AI, P7AI is more thermally orthogonal relative to mispairing with native bases. However, the P7AI nucleobase is less orthogonal relative to self-pairing, as the P7AI:P7AI self-pair is stabilized by 0.8 °C relative to the 7AI self-pair.

Isostructural modifications of the electronic nature of the nucleobases were examined by substitution of CH groups with N atoms (aza-substitution). Specifically, the imidizopyridine (ImPy) and pyrrolopyrizine (PP) nucleosides were examined. ImPy:ICS and PP:ICS are both less stable than 7AI:ICS, by 2.5 and 0.9 °C, respectively (Table 1). The mispairs are also destabilized with ImPy, by 0.2−0.5 °C for dA, dT, and dG, and by 2.2 °C for dC. This decrease in mispair stability is perhaps surprising, at least in the case of ImPy:dC and ImPy:dT, considering the more purine-like ring structure of ImPy relative to that of 7AI. For PP, the mispair with dT is slightly stabilized, while those with dC, dA, and dG are destabilized by 1.0−2.5 °C, again relative to the same mispairs with 7AI. The low stability of these mispairs, and the reasonable stability of the PP:ICS and PP:MICS base pairs, give this azaindole analogue increased thermodynamic orthogonality relative to mispairing with native bases. Moreover, the PP:PP self-pair is significantly destabilized, by 4.2 °C relative to 7AI:7AI, providing good thermal orthogonality for PP against all possible mispairs.

2.2.2. Kinetic Properties of 7AI Derivatives. The 7AI derivative nucleobases were also evaluated as substrates for KF. Initial velocities were determined by literature methods using [γ-^{32}P]primer extension reactions with varying concentrations of nucleoside triphosphates.[26] The reactions were analyzed by polyacrylamide gel electrophoresis and subsequent Phosphorimager (Molecular Dynamics) analysis to quantify gel band intensities corresponding to the extended primer. The measured velocities were plotted against the concentration of dNTP and fit to the Michaelis−Menten equation. Unnatural nucleobases were assayed both in template DNA and as incoming nucleoside triphosphates. Steady-state kinetic parameters for single nucleotide incorporation (apparent k_{cat} and apparent K_M) are reported in Tables 2 and 3.

As mentioned above, the7AI:ICS base pair was synthesized by KF with reasonable efficiency, while mispairs with native bases were synthesized significantly slower.[19] However, the synthesis of 7AI:ICS was significantly less efficient than that for native base pairs. Therefore, it is desirable to further optimize the rate at which KF synthesizes the unnatural base pair. Increased fidelity both against mispairing with native bases, and against self-pairing, is also necessary to make DNA replication practical.

The effect of C6-methyl substitution on enzymatic synthesis was examined first. With M7AI in template, the efficiency of dM7AITP incorporation is low ($k_{cat}/K_M = 6.3 \times 10^3$ M^{-1} min^{-1},

(26) Goodman, M. F.; Creighton, S.; Bloom, L. B.; Petruska, J. *Crit. Rev. Biochem. Mol. Biol.* **1993**, *28*, 83−126.

Table 3. Steady-State Kinetic Parameters for KF Exo-Mediated Synthesis of DNA with ImPy and PP in the Template[a]

5′-dTAATACGACTCACTATAGGGAGA
3′-dATTATGCTGAGTGATATCCCTCTXGTCA

template (X)	nucleoside triphosphate	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} min^{-1})
ImPy	ImPy	2.0 ± 0.1	109 ± 11	1.8 × 10^4
	ICS	3.3 ± 0.5	19 ± 6	1.7 × 10^5
	PICS	2.4 ± 0.3	2.7 ± 1.4	8.9 × 10^5
	MICS	30 ± 1	14 ± 1	2.1 × 10^6
	5MICS	5.5 ± 0.2	28 ± 3	2.0 × 10^5
	A	1.6 ± 0.1	50 ± 7	3.2 × 10^4
	G	0.12 ± 0.01	134 ± 27	9.0 × 10^2
	C	2.6 ± 0.2	190 ± 50	1.4 × 10^4
	T	0.89 ± 0.05	157 ± 19	5.7 × 10^3
PP	PP	2.5 ± 0.1	15 ± 5	1.7 × 10^5
	ICS	3.5 ± 0.1	25 ± 3	1.4 × 10^5
	PICS	2.0 ± 0.1	3.5 ± 0.5	5.7 × 10^5
	MICS	20 ± 1	10 ± 1	2.0 × 10^6
	5MICS	2.5 ± 0.1	28 ± 3	8.9 × 10^4
	A	1.1 ± 0.1	49 ± 7	2.2 × 10^4
	G	0.54 ± 0.04	41 ± 6	1.3 × 10^4
	C	2.5 ± 0.2	27 ± 12	9.3 × 10^4
	T	0.85 ± 0.04	98 ± 11	8.7 × 10^3

[a] See text for experimental details.

Table 4. T_m Values for Duplex-Containing ICS Derivatives[a]

5′-dGCGTACXCATGCG
3′-dCGCATGYGTACGC

X	Y	T_m (°C)	X	Y	T_m (°C)
A	T	59.2	PICS	PICS	62.6
G	C	61.8		7AI	56.8
ICS	ICS	59.3		M7AI	59.7
	7AI	56.5		P7AI	58.8
	M7AI	58.5		PP	56.0
	A	55.1		A	55.5
	T	53.0		T	53.7
	G	51.0		G	54.5
	C	52.2		C	51.4
MICS	MICS	62.3	5MICS	5MICS	61.2
	A	53.6		A	55.7
	T	55.1		T	55.5
	G	54.8		G	55.5
	C	54.8		C	54.3

[a] See text for experimental details.

Table 2). Therefore, despite the high thermal stability of the M7AI:M7AI self-pair, it is not efficiently synthesized by KF, making M7AI a promising nucleobase to pair with ICS derivatives. The incorporation efficiency of dICSTP opposite M7AI is within a factor of 2.5 of that with 7AI in the template (k_{cat}/K_M = 2.3 × 10^5 and 7.6 × 10^4 M^{-1} min^{-1} for 7AI and M7AI in template, respectively). Native triphosphates are incorporated opposite M7AI in the template with rates of approximately 2 × 10^3 M^{-1} min^{-1}. Therefore, KF selectively incorporates dICSTP opposite M7AI in the template by a factor of 20−80-fold relative to incorporation of native triphosphates. This selectivity is 2-fold reduced relative to that observed with 7AI in the template (dICSTP is inserted opposite 7AI 40−140-fold faster than any native triphosphate).

Differences between M7AI and 7AI are more pronounced when the unnatural base is present as a triphosphate. The triphosphate dM7AITP is incorporated 9−23-fold less efficiently than d7AITP opposite ICS and its derivatives in the template (Tables 5 and 6). However, dM7AITP is inserted even less efficiently opposite native bases relative to d7AITP (Table 7). For example, the incorporation efficiency of dM7AITP opposite dA is reduced 50-fold relative to d7AITP opposite dA (from 1.6 × 10^5 to 2.8 × 10^3 M^{-1} min^{-1}). The decreased rates for insertion result from both k_{cat} and K_M effects.

Table 5. Steady-State Kinetic Parameters for KF Exo-Mediated Synthesis of DNA with ICS and PICS in the Template[a]

5′-dTAATACGACTCACTATAGGGAGA
3′-dATTATGCTGAGTGATATCCCTCTXGTCA

template (X)	nucleoside triphosphate	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} min^{-1})
ICS	ICS	3.6 ± 0.1	57 ± 6	6.3 × 10^4
	PICS	2.6 ± 0.4	33 ± 11	7.8 × 10^4
	7AI	10.6 ± 0.5	60 ± 15	1.8 × 10^5
	M7AI	1.6 ± 0.2	204 ± 60	7.8 × 10^3
	P7AI	28.8 ± 1.6	15 ± 2	1.9 × 10^6
	ImPy	1.3 ± 0.1	32 ± 15	4.1 × 10^4
	PP	4.9 ± 0.2	12 ± 2	4.1 × 10^5
	A	0.15 ± 0.01	73 ± 10	2.1 × 10^3
	G	0.28 ± 0.02	83 ± 12	3.4 × 10^3
	C	0.03 ± 0.01	34 ± 19	8.8 × 10^2
	T	1.2 ± 0.1	234 ± 83	5.1 × 10^3
PICS	ICS	1.2 ± 0.1	25 ± 3	4.8 × 10^4
	PICS	0.90 ± 0.04	3.7 ± 0.7	2.4 × 10^5
	7AI	4.5 ± 0.1	21 ± 2	2.1 × 10^5
	M7AI	0.88 ± 0.04	56 ± 9	1.6 × 10^4
	P7AI	43 ± 1	4.1 ± 0.1	1.0 × 10^7
	ImPy	1.3 ± 0.1	25 ± 5	5.2 × 10^4
	PP	2.8 ± 0.2	6 ± 2	4.7 × 10^5
	A	0.17 ± 0.07	92 ± 30	1.8 × 10^3
	G	0.12 ± 0.04	69 ± 22	1.7 × 10^3
	C	0.22 ± 0.12	123 ± 81	1.8 × 10^3
	T	0.89 ± 0.15	76 ± 36	1.2 × 10^{xt}

[a] See text for experimental details.

Table 6. Steady-State Kinetic Parameters for KF Exo-Mediated Synthesis of DNA with MICS and 5MICS in the Template[a]

5′-dTAATACGACTCACTATAGGGAGA
3′-dATTATGCTGAGTGATATCCCTCTXGTCA

template (X)	nucleoside triphosphate	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} min^{-1})
MICS	MICS	4.6 ± 0.1	14 ± 1	3.3 × 10^5
	ICS	2.2 ± 0.2	37 ± 14	5.9 × 10^4
	7AI	7.0 ± 0.3	44 ± 5	1.8 × 10^5
	M7AI	1.0 ± 0.1	57 ± 15	1.8 × 10^4
	P7AI	19 ± 1	6.2 ± 1.2	3.1 × 10^6
	ImPy	1.5 ± 0.1	18 ± 4	8.3 × 10^4
	PP	8.4 ± 0.7	5.9 ± 1.6	1.4 × 10^6
	A	0.36 ± 0.02	15 ± 2	2.4 × 10^4
	G	0.28 ± 0.02	48 ± 6	5.8 × 10^3
	C	0.27 ± 0.03	43 ± 15	6.3 × 10^3
	T	1.6 ± 0.1	80 ± 7	2.0 × 10^4
5MICS	5MICS	nd[b]	nd[b]	≤1.0 × 10^3
	ICS	1.6 ± 0.2	112 ± 42	1.4 × 10^4
	7AI	6.9 ± 0.1	100 ± 5	6.9 × 10^4
	M7AI	0.9 ± 0.1	124 ± 27	7.3 × 10^3
	P7AI	23 ± 2	16 ± 4	1.4 × 10^6
	ImPy	0.7 ± 0.1	26 ± 10	2.7 × 10^4
	PP	2.3 ± 0.1	7.3 ± 1.0	3.2 × 10^5
	A	nd[b]	nd[b]	≤1.0 × 10^3
	G	nd[b]	nd[b]	≤1.0 × 10^3
	C	nd[b]	nd[b]	≤1.0 × 10^3
	T	nd[b]	nd[b]	≤1.0 × 10^3

[a] See text for experimental details. [b] Rates too slow for determination of k_{cat} and K_M independently.

With P7AI in the template, the incorporation efficiency of dICSTP is reduced by approximately 2-fold relative to that with 7AI in the template (Table 2). Conversely, native triphosphates are incorporated more efficiently opposite P7AI than opposite 7AI, by 2−30-fold for dATP, dTTP, and dGTP. However, the incorporation of dCTP opposite P7AI is approximately 1000-fold more efficient than that opposite 7AI in the template. This remarkable, and selective, enhancement in the efficiency of

Table 7. Steady-State Kinetic Parameters for KF Exo-Mediated Synthesis of DNA with A, G, C, and T in the Template[a]

5′-dTAATACGACTCACTATAGGGAGA
3′-dATTATGCTGAGTGATATCCCTCTXGTCA

template (X)	nucleoside triphosphate	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} thrmin^{-1})
A	T	163 ± 7	3.5 ± 1.0	4.7×10^7
	7AI	7.5 ± 0.8	48 ± 10	1.6×10^5
	M7AI	0.28 ± 0.02	100 ± 20	2.8×10^3
	P7AI	21 ± 1	8.8 ± 1.7	2.4×10^6
	ImPy	1.6 ± 0.1	70 ± 10	2.3×10^4
	PP	5.6 ± 0.2	11 ± 2	5.1×10^5
	ICS	3.0 ± 0.3	50 ± 16	6.0×10^4
	PICS	0.40 ± 0.10	20 ± 10	2.0×10^4
	MICS	9.3 ± 0.2	36 ± 2	2.6×10^5
	5MICS	0.89 ± 0.02	69 ± 5	1.3×10^4
G	7AI	0.63 ± 0.02	30 ± 3	2.1×10^4
	M7AI	nd[b]	nd[b]	$\leq 1.0 \times 10^3$
	P7AI	7.6 ± 0.2	26 ± 2	2.9×10^5
	ImPy	0.27 ± 0.01	60 ± 9	4.5×10^3
	PP	0.71 ± 0.06	9.3 ± 2.6	7.6×10^4
	ICS	0.48 ± 0.02	27 ± 3	1.8×10^4
	PICS	0.09 ± 0.01	9.3 ± 3.6	9.7×10^3
	MICS	2.25 ± 0.03	29 ± 1	7.8×10^4
	5MICS	0.16 ± 0.01	51 ± 11	3.1×10^3
C	7AI	1.7 ± 0.1	88 ± 4	1.9×10^4
	M7AI	nd[b]	nd[b]	$\leq 1.0 \times 10^3$
	P7AI	15.9 ± 0.2	22 ± 1	7.2×10^5
	ImPy	0.73 ± 0.04	40 ± 6	1.8×10^4
	PP	5.0 ± 0.2	9.9 ± 1.0	5.1×10^5
	ICS	1.7 ± 0.2	72 ± 10	2.4×10^4
	PICS	1.05 ± 0.04	2.6 ± 0.7	4.0×10^5
	MICS	4.7 ± 0.2	35 ± 3	1.3×10^5
	5MICS	0.54 ± 0.04	140 ± 23	3.9×10^3
T	7AI	4.6 ± 0.1	113 ± 7	4.1×10^4
	M7AI	0.28 ± 0.02	100 ± 20	2.8×10^3
	P7AI	21.4 ± 0.7	24 ± 2	8.9×10^5
	ImPy	1.2 ± 0.1	75 ± 12	1.6×10^4
	PP	4.9 ± 0.2	19 ± 2	2.6×10^5
	ICS	4.9 ± 0.2	103 ± 13	4.8×10^4
	PICS	2.80 ± 0.10	9.4 ± 2.0	3.0×10^5
	MICS	15 ± 1	54 ± 5	2.8×10^5
	5MICS	2.1 ± 0.10	126 ± 15	1.7×10^4

[a] See text for experimental details. [b] Rates too slow for determination of k_{cat} and K_M independently.

dCTP incorporation upon propynyl substitution results from a 14-fold increase in apparent k_{cat} and a 62-fold decrease in apparent K_M(dCTP).

The triphosphate, dP7AITP is inserted opposite ICS with high efficiency (1.9×10^6 M^{-1} min^{-1}, Table 5). Relative to the incorporation of d7AITP opposite ICS, dP7AITP is inserted 10-fold more efficiently due to both an increase in k_{cat} (3-fold) and a decrease in K_M (4-fold). However, the triphosphate dP7AITP is also inserted with increased efficiency opposite all of the native templates, relative to d7AITP. Insertion opposite purines (dA and dG) is approximately 15-fold more efficient, while insertion opposite pyrimidines (dT and dC) is 20−40-fold more efficient. The self-pair P7AI:P7AI is also synthesized with increased efficiency (by 6-fold), relative to 7AI:7AI, again limiting the usefulness of this nucleobase for expansion of the genetic alphabet.

Isostructural alterations of the electronic nature of the nucleobase were also examined with respect to enzymatic DNA synthesis. The effect of N3 substitution was determined by comparing the rate for dICSTP incorporation opposite 7AI and ImPy. Insertion of dICSTP opposite ImPy in the template proceeds with the same efficiency (1.7×10^7 M^{-1} min^{-1}) as that of dICSTP opposite 7AI. Native triphosphates are inserted opposite ImPy with efficiencies that are all less than 3.2×10^4 M^{-1} min^{-1} (2−5-fold faster than insertion opposite 7AI for dATP, dCTP, and dTTP, and 2-fold slower for dGTP). A 15-fold increase in k_{cat} for dCTP insertion is offset by a 6-fold increase in K_M(dCTP). The insertion of dATP is only 6-fold slower than the insertion of dICSTP, limiting the orthogonality of ImPy. Insertion of dImPyTP opposite ICS proceeds with an efficiency of 4.1×10^4 M^{-1} min^{-1}, which is 4-fold reduced relative to the insertion of d7AITP (1.8×10^5 M^{-1} min^{-1}). The insertion efficiency of dImPyTP opposite dC is equal to that of d7AITP opposite dC, while insertion of dImPyTP opposite dG, dA, and dT is 2−7-fold slower by comparison. The ImPy:ImPy self-pair is synthesized by KF 12-fold less efficiently than the 7AI:7AI self-pair (Table 7).

The rate of KF-mediated insertion of dICSTP opposite PP in the template is reasonably efficient ($k_{cat}/K_M = 1.4 \times 10^5$ M^{-1} min^{-1}), but approximately 2-fold slower than the rate of insertion of dICSTP opposite 7AI. The insertions of triphosphates dGTP, dATP, and dTTP are slightly more efficient opposite PP than opposite 7AI (by 4−8-fold). The insertion of dCTP opposite PP is 17-fold more efficient than the insertion of dCTP opposite 7AI (Table 3).

The triphosphate dPPTP is inserted opposite ICS in the template approximately 2-fold faster than d7AITP is inserted opposite ICS. Similar to the behavior of PP in the template, insertion of dPPTP opposite native templates dA, dG, and dT is slightly faster than the insertion of d7AITP (3−6-fold), while the insertion of dPPTP opposite dC is 27-fold faster than that of d7AITP (Table 7). Therefore, inclusion of the nitrogen at position 4 of the azaindole ring selectively favors the synthesis of the mispair between the unnatural base and natural bases (especially cytosine) in both template−triphosphate combinations. Despite the increased rates for the incorporation of dPPTP opposite dC, the insertion of the PP triphosphate is still 2 orders of magnitude slower than the insertion of dGTP. The self-pair PP:PP is synthesized slightly slower than the 7AI:7AI self-pair, with an efficiency of 1.7×10^5 M^{-1} min^{-1}.

2.2.3. Thermodynamic Properties of ICS Derivatives. To further optimize the unnatural base pair we examined the stability of base pairs formed with the ICS derivatives shown in Figure 1. As reported previously, C7-propynyl substitution marginally stabilizes the unnatural pair 7AI:PICS relative to 7AI:ICS (T_m = 57.6 and 57.2 °C, respectively, Table 4).[19] The mispairs between PICS and dA or dT are both slightly more stable than the mispairs between ICS and dA or dT (by 0.4−0.7 °C). The PICS mispair with dG is 1.8 °C more stable, while the mispair with dC is 0.8 °C less stable, relative to the same mispairs with ICS. Relative to ICS, PICS stabilizes the unnatural pairs with each 7AI derivative slightly more than it stabilizes the most stable of the mispairs (PICS:dA). Therefore, PICS is slightly more thermodynamically orthogonal to the native bases than is ICS. The remarkable stability of the PICS:PICS self-pair (T_m = 62.6 °C) limits the thermal orthogonality of the 7AI:PICS unnatural base pair.

The effect of C3-methyl substitution was examined with the isocarbostyril derivative MICS. The added methyl group of MICS has a pronounced effect on base pair stability, with 7AI:MICS stabilized by 2.5 °C relative to the 7AI:ICS base pair (T_m = 59.7 °C relative to 57.2 °C, Table 4). The mispairs formed between MICS and dT, dG, and dC are also stabilized by 2.1−3.8 °C, while the mispair with dA is destabilized by 1.5 °C. The stabilities of the most stable mispair formed between ICS or MICS and a native base are identical (T_m = 55.1 °C for ICS:dA and MICS:dT). However, 7AI:MICS is more stable than 7AI:ICS, and thus MICS is more thermodynamically orthogonal than ICS relative to the native bases. In fact, the 7AI:MICS

base pair is both more stable than a dA:dT base pair (T_m = 59.2 °C for dA:dT) and more selective against mispairing with native bases (thermal selectivity, relative to the most stable mispair, is 4.6 and 3.8 °C for 7AI:MICS and dA:dT, respectively). 7AI:MICS is more stable than the 7AI:7AI self-pair (by 4.2 °C) but less stable than the MICS:MICS self-pair (by 2.6 °C).

In addition to the effects of methyl substitution at C3, the effects of methyl substitution at C5 of the isocarbostyril core (5MICS) were examined. Methyl substitution at the C5 position is significantly less stabilizing than that at the C3 position (Table 4). The stability of the 7AI:5MICS base pair is equal to that of the 7AI:ICS base pair. Surprisingly, however, the mispairs between 5MICS and the native bases are all stabilized by the inclusion of the C5-methyl group. The mispair with dA is stabilized by 0.6 °C, relative to the ICS mispairs, while the mispairs with dC, dT, and dG are stabilized by 2.1−4.5 °C. The C5-methyl substitution also stabilizes the self-pair by 1.9 °C. The increased stabilities of the mispairs and the self-pair make 5MICS less thermally orthogonal than ICS.

2.2.4. Kinetic Properties of ICS Derivatives. As reported previously, PICS has been examined as a substrate for KF.[18,19] With PICS in the template, there was little difference in efficiency of d7AITP incorporation relative to that with ICS in the template (Table 5). The rates for insertion of native triphosphates opposite PICS are also relatively unchanged, as compared to those with ICS in the template. There is a slight increase in insertion efficiency for dCTP and dTTP, and a slight decrease in efficiency for insertion of dGTP and dATP opposite PICS in the template. When present as the triphosphate, the dPICSTP propynyl group again had little effect on incorporation efficiencies opposite 7AI (Table 2). The propynyl group had larger effects on the efficiency of incorporation of the unnatural triphosphate opposite native templates (Table 7). In each case, the added propynyl group resulted in a decrease in k_{cat} (1.6−8-fold) and a relatively larger decrease in K_M (2−30-fold). Opposite dA and dG, dPICSTP was incorporated slower than dICSTP (by 2- and 3-fold, respectively), and opposite dT and dC, dPICSTP was incorporated 6- and 17-fold faster than dICSTP. The PICS:PICS self-pair is synthesized almost 4-fold more efficiently than the ICS:ICS self-pair, limiting the kinetic orthogonality of PICS.

The effects of the C3-methyl group on KF-mediated DNA synthesis were examined with MICS in the template and d7AITP (Table 6). Small differences are found upon modification of unnatural base in the template. However, opposite MICS, the insertion of both dImPyTP (8.3×10^4 M^{-1} min^{-1}) and dPPTP (1.4×10^6 M^{-1} min^{-1}) is more efficient than that opposite other ICS analogues in the template. For example, relative to d7AITP, dPPTP is inserted 8-fold more efficiently opposite MICS. The triphosphate d7AITP is inserted with the same efficiency opposite both ICS and MICS in the template (1.8×10^5 M^{-1} min^{-1}). The rates for insertion of the native triphosphates opposite MICS are 2−11-fold more efficient than those for insertion opposite ICS. The largest increase observed is in the incorporation of dATP, which proceeds with an efficiency of 2.4×10^4 M^{-1} min^{-1}.

For a triphosphate, incorporation of dMICSTP by KF opposite the 7AI analogue bases in the template is 4−14-fold more efficient than that of dICSTP opposite the same templates (Tables 2 and 3). The greatest increase is observed for the incorporation of dMICSTP opposite PP, which proceeds with an efficiency of 2×10^6 M^{-1} min^{-1}. With templates containing native bases, dMICSTP is also incorporated by KF approximately 5-fold more efficiently than dICSTP. This results primarily from a 3−5-fold increase in k_{cat}, along with an up to 2-fold decrease in K_M for dMICSTP relative to that for dICSTP. Insertions of dMICSTP opposite dA, dC, and dT were catalyzed with large values of k_{cat}/K_M, approaching 3×10^5 M^{-1} min^{-1} (Table 7). However, this rate is 2 orders of magnitude lower than the rate at which KF pairs the same templates with their native partners. Therefore, incorporation of the unnatural base at these positions does not compete with native synthesis.

Enzymatic synthesis of base pairs containing 5MICS was also examined. The 5MICS:5MICS self-pair is synthesized with low efficiency (less than 1×10^3 M^{-1} min^{-1}, Table 5). However, the generally slow rates of unnatural base pair synthesis with 5MICS in the template or as the triphosphate (6.9×10^4 and 1.3×10^5 M^{-1} min^{-1}, respectively with 7AI) limit the practical utility of 5MICS.

3. Discussion

The practical replication of DNA containing three base pairs requires that the rate for unnatural pair synthesis approach those rates characteristic of native pair synthesis. Fidelity of replication must also be high, especially with respect to self-pairing. Any improvements in enzymatic replication must be accomplished without significantly compromising base pair stability or thermal selectivity in duplex DNA.

Hydrophobic bases, such as derivatives of 7AI:ICS that are efficiently recognized by KF, and recognized with good fidelity, are potentially realistic candidates for expansion of the genetic alphabet. Modifications of the 7AI:ICS base pair were found which improved both the efficiency and the fidelity of synthesis. Examination of the data also reveals much about the general stability of hydrophobic bases in duplex DNA, as well as the recognition of these bases by KF during DNA synthesis.

3.1. Alkyl and Aza Substituent Effects. 3.1.1. Effect of Methyl Substitution. The optimization of interstrand hydrophobic packing was examined by generating methyl-substituted derivatives of 7AI and ICS. The 7-azaindole analogue M7AI has a C6-methyl group that is expected to reside in the minor groove. One ICS analogue, MICS, has a C3-methyl group that is expected to be positioned in the interface with the pairing base. Another ICS analogue, 5MICS, has a C5-methyl group that is expected to be positioned in the major groove.

C6-Methyl substitution of 7AI (M7AI) and C3-methyl substitution of ICS (MICS) strongly stabilized both pairing with other unnatural bases and mispairing with native bases. A single exception was the mispair between MICS:dA, which was destabilized by 1.5 °C relative to ICS:dA. Methyl substitution at the C5 position of ICS (5MICS) weakly stabilized pairing with other hydrophobic bases but stabilized mispairing with native bases more strongly. All of the methyl-substituted derivatives were found to form more stable self-pairs than 7AI or ICS.

The data imply that inclusion of a methyl group in the minor groove, major groove, or interbase interface methyl groups leads to specific interbase interactions in addition to increased buried hydrophobic surface area. This follows from the observation that although methyl substitution in general increased the stability of all base pairs, the extent of stabilization was not uniform. For example, the greatest increase in stabilization afforded by the addition of the C3-methyl group of MICS was observed when the unnatural base was incorporated opposite less hydrophobic dG and dC.[27] In contrast, there was a

(27) Shih, P.; Pedersen, L. G.; Gibbs, P. R.; Wolfenden, R. *J. Mol. Biol.* **1998**, *280*, 421−430.

significant destabilization opposite dA, the most hydrophobic of the native bases.[27] Selective destabilization of the mispair with dA is particularly advantageous for the orthogonality of MICS and shows a clear improvement over the highly stable ICS:dA mispair.

The discovery that H-bonds are not an absolute requirement for enzymatic DNA synthesis[28−30] has strengthened the argument that the fidelity of DNA synthesis derives, at least in part, from a polymerase-mediated steric matching of the pairing bases.[31−33] Therefore, optimization of interstrand hydrophobic packing, by generating methyl-substituted derivatives of 7AI and ICS, might allow for the selective optimization of correct unnatural pair synthesis relative to the synthesis of mispairs. The major fidelity problem of the parental 7AI:ICS base pair resulted from the 7AI:7AI and ICS:ICS self-pairs.[19] The addition of either a major groove (5MICS) or a minor groove (M7AI) methyl group decreased the rate of self-pair synthesis. The M7AI:M7AI and 5MICS:5MICS self-pairs were synthesized 37- to >50-fold less efficiently than the 7AI:M7AI and ICS:ICS self-pairs, respectively. Despite the slow rates of self-pair synthesis, M7AI and 5MICS are limited in their practical utility because KF does not efficiently incorporate d5MICSTP opposite any of the hydrophobic bases examined (see below). In contrast, the C3-methyl substitution of ICS increased the rate of self-pair formation (MICS:MICS) by a factor of 5, relative to that of the ICS:ICS self-pair. However, the relatively efficient formation of the MICS:MICS self-pair does not preclude the use of MICS for heteropair formation with 7AI derivatives, including PP (see below), due to the high efficiency with which the heteropairs are formed.

The shape of the unnatural base also affected the synthesis of heteropairs. However, the nature of the base in the template, natural or unnatural, had only small effects on the rates of insertion of a given triphosphate. Addition of a methyl group typically changed the rate for the insertion of a natural or an unnatural triphosphate by less than 5-fold. The rates of unnatural pair synthesis were more sensitive to modification of the triphosphate. The largest effects were observed when the isocarbostyril ring was modified at the C3 position (dMICSTP). Relative to dICSTP, the rates for dMICSTP insertion opposite 7AI analogue templates increased by 6−15-fold, resulting in very efficient rates of 10^6 M^{-1} min^{-1}. The rates for incorporation of dMICSTP opposite native templates were also increased (by 4−6-fold), but not to the extent that they were competitive with insertion of a correct, native triphosphate. The C5-methyl group of d5MICSTP resulted in a small decrease in the rate of enzymatic pairing with 7AI derivatives (by 2−4-fold) or native templates (by 3−6-fold). A significant decrease in efficiency was observed for the incorporation of dM7AITP opposite the ICS analogues by 10−20-fold relative to the incorporation of d7AITP. The effect of the C6-methyl group was even larger for the incorporation of the unnatural triphosphate opposite native templates, where none of the rates exceeded 10^4 M^{-1} min^{-1} (Table 7).

3.1.2. Effect of Propynyl Substitution. Previously, with isocarbostyril base pairs, it was observed that unsymmetrical propynyl substitution (where only one base of the pair has a propynyl group, i.e., ICS:PICS) was destabilizing.[19] However, symmetrical propynyl substitution (i.e., PICS:PICS) was strongly stabilizing.[19] This trend was also apparent in the data discussed above. The base pair between 7AI and ICS was stabilized by −0.4 to 0.4 °C upon addition of a single propynyl group to either unnatural base (i.e., 7AI:PICS or P7AI:ICS). However, with symmetrical propynyl substitution (P7AI:PICS, P7AI:P7AI, or PICS:PICS), the base pairs were stabilized by 0.8−3.3 °C.

Similar to the effects of methyl group modification, the addition of a propynyl group to the unnatural base in the template resulted in only modest effects on base pair recognition by KF. Exceptions were the efficient incorporation of dTTP, dGTP, and especially dCTP opposite P7AI, relative to 7AI in the template. In these cases, the added propynyl group increased binding of the native triphosphates (by 2−60-fold) and catalytic turnover (by 5−14-fold), resulting in a large increase in the incorporation efficiency of these dNTPs opposite P7AI, as compared to 7AI (13−800-fold). In fact, the rate of dCTP incorporation opposite P7AI (4.6×10^6 M^{-1} min^{-1}) was only 10-fold slower than the rate of dG:dC synthesis in the same sequence context. In the context of KF-mediated DNA synthesis, P7AI proved not to be orthogonal to native templates, but rather to be an efficient dG mimic. As this effect is not reflected in the thermodynamic data (in fact, dC:P7AI is 2.8 °C less stable than dC:7AI), it seems that the efficient insertion of dCTP opposite P7AI is mediated by interactions specific to the polymerase or the insertion transition state.

The addition of a propynyl group to the isocarbostyril triphosphate (dPICSTP) also had only modest effects, with efficiencies that were within a factor of 5 relative to dICSTP, resulting primarily from tighter binding of the triphosphate with a propynyl group. However, addition of a propynyl group to 7AI had larger effects. Opposite both unnatural and natural templates, dP7AITP is inserted 10−50-fold more efficiently than is d7AITP. This increased efficiency resulted from a combination of k_{cat} and K_M effects. The largest increase was for the incorporation of dP7AITP opposite PICS, the only unnatural template to also have a major groove propynyl group. It is tempting to speculate that increased kinetic efficiency might be related to the thermal stability characteristic of base pairs that are symmetrically substituted with propynyl groups, as discussed above. This would imply that the interactions which stabilize the symmetrically substituted base in duplex DNA are significantly developed in the transition state for the incorporation of an unnatural triphosphate bearing a propynyl group. Unlike P7AI in the template, dP7AITP is not a specific mimic of dGTP, being incorporated less efficiently opposite dC than opposite either dA or dT.

3.1.3. Electronic Effects. Electronic effects may also play an important role in the stability of unnatural base pairs in DNA.[34−36] The 7AI ring has a single nitrogen atom (N7) and is otherwise hydrophobic. Aza-substitution is expected to affect the physical properties of the nucleobases. First, the presence of the extra nitrogen atoms in the polyaza heterocycles decreases the basicity at N7, relative to that of 7AI, due to the inductive electron-withdrawing effect of the added nitrogen atoms and to unfavorable charge localization in the protonated forms

(28) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 1001−1007.
(29) Moran, S.; Ren, R. X.-F.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10506−10511.
(30) Moran, S.; Ren, R. X.-F.; Rumney, S. I.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056−2057.
(31) Goodman, M. F. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10493−10495.
(32) Morales, J. C.; Kool, E. T. *Nature Struct. Biol.* **1998**, *5*, 950−954.
(33) Kool, E. T. *Biopolymers* **1998**, *48*, 3−17.
(34) Bergman, E. D.; Weiler-Feilchenfeld. In *The Dipole Moments of Purines*; Bergman, E. D., Pullman, B., Eds.; The Israel Academy of Sciences and Humanities: Jerusalem, 1971; pp 21−28.
(35) Lister, J. H. *The Purines. Supplement 1*; John Wiley and Sons: New York, 1996; Vol. 54.
(36) Hurst, D. T. *An Introduction to the Chemistry and Biochemistry of Pyrimidines, Purines and Pteridines*; Page Bros.: Norwich, Great Britain, 1980.

relative to the unprotonated forms of these heterocycles.[36] For example, the pK_a of pyrazine (pK_a = 0.7) is significantly lower than that of pyridine (pK_a = 5.2), as is the N3 pK_a of 2′-deoxyxanthosine (pK_a = 5.7) relative to that of 2′-deoxy-7-deazaxanthosine (pK_a = 7.2).[2,37] Second, aza-substitution is also expected to affect the dipole moment of the base.[34,36] In analogy to native purine rings, the effects of aza-substitution on the overall electron distribution in the 7AI analogues may be expected to result from the sharing of electron density between nitrogen atoms of the fused five- and six-membered rings.[34,36] The N4-aza-substitution of PP should result in a more π-electron deficient pyrazine ring and should contribute to a dipole moment oriented toward the six-membered ring, more similar to the native purine ring dipole moment. The N3-aza-substitution of ImPy is expected to result in a more π-electron deficient pyrrole ring and should contribute to a dipole oriented toward the five-membered ring, opposite to the case with a purine ring system. Finally, aza-substitution is also expected to affect the polarizability of the base. A significant component of base stacking results from polarization forces, which manifest themselves structurally as the partial overlap of the polar bonds of one nucleobase with the polarizable ring system of an adjacent base.[38] The addition of heteroatoms to otherwise nonpolar hydrophobic rings, such as benzene and naphthalene, has a pronounced effect on the structure of these molecules.[38] The nonpolar molecules form structures with the characteristic herringbone pattern with no ring stacking, while those bearing heteroatom substituents crystallize in forms dominated by ring-stacking interactions with the polar heteroatoms packing with the polarizable ring system of an adjacent molecule.[38]

The T_m data in Table 1 reveal that both N3 of ImPy and N4 of PP result in a consistent decrease in duplex stability. The destabilization is unlikely to result from direct interactions between pairing bases due to the large backbone deformations that would be necessary to position the nitrogen atoms in a region of the duplex accessible to the pairing base. This is consistent with preliminary NMR data which indicate that the 7AI:ICS packs in DNA without significant backbone deformations or intercalation. Since the extent of destabilization is roughly independent of the pairing base, it may derive from less than optimal stacking interactions. However, PP and ImPy are expected to stack more favorably with adjacent bases due to the predicted overlap with polarizable regions of the adjacent bases. Moreover, poor stacking would also be surprising, at least for ImPy, considering its more purine-like imidizo ring structure. It seems more likely that the decrease in duplex stability may result from a decrease in the stability of H-bonding interactions with the ordered water molecules along the minor groove[39−43] due to the decreased ability of the minor groove N7 of 7AI to accept a H-bond relative to the corresponding minor groove nitrogen atom of guanine or adenine. The H-bond-accepting ability of the minor groove N7 of PP and ImPy is further compromised by the inductive electron-withdrawing effects of aza-substitution. Despite these effects, the overall reduction in duplex stability is not a significant obstacle to the use of these unnatural bases. The stability of the unnatural base pairs comprised of PP and ImPy is comparable, or only slightly reduced, relative to that of native base pairs.

The effects of nitrogen substitution were also examined in the context of KF-mediated DNA synthesis. The effect on unnatural base pair synthesis was negligible for modification of the templating base with either N3 or N4 substitution. In the template, 7AI, PP, and ImPy directed the incorporation of each ICS analogue triphosphate with virtually identical rates. However, unnatural base pair synthesis was more affected when these substitutions were made in the triphosphate. Compared to the insertion of d7AITP opposite the ICS analogues, dImPyTP was inserted less efficiently (by 2−7-fold), resulting predominately from a decrease in apparent k_{cat}, while dPPTP was inserted more efficiently (by 2−8-fold), resulting primarily from a decrease in apparent K_M(triphosphate). The largest difference observed was in the insertion of dPPTP opposite 5MICS (14-fold smaller apparent K_M) or MICS (7-fold smaller apparent K_M, and 8-fold increased k_{cat}/K_M). It seems unlikely that the differing behavior of the two aza-substituted triphosphates results from differing bascisity of N7, as each analogue is less basic. The opposite effects that N3- and N4-aza-substitution have on KF recognition imply a role for the dipole moment of the bases, as the addition of the electron-withdrawing heteroatoms should contribute to dipole moments oriented in opposite directions. Unnatural nucleobases that are functionalized with heteroatoms at a variety of positions are currently being further examined.

Unlike hydrophobic pair synthesis, the rates with which KF inserted native bases opposite the template were sensitive to aza-substitution. Generally, the native triphosphates were inserted by KF more efficiently opposite both ImPy and PP, as opposed to 7AI, predominately due to an increase in apparent k_{cat}. The effect is most dramatic for the insertion of dCTP opposite both ImPy and PP templates, where the apparent k_{cat} for insertion of dCTP was increased 15-fold, relative to the insertion of dCTP opposite 7AI. When present as the triphosphate, the two N-substitutions had different effects on KF recognition. Both dImPyTP and dPPTP bound to the enzyme−template complex most tightly opposite dC. However, the only consistent and significant effect was tighter binding of dPPTP opposite native templates, with K_M(dPPTP) 3−9-fold smaller than K_M(d7AITP). The effect was greatest opposite dC in the template, which when combined with a 3-fold increase in insertion efficiency results in a greater than 20-fold increase in the pairing of dPPTP opposite dC, relative to that of d7AITP. The electronic effects resulting from C3- or C4-aza-substitution selectively favored pairing with deoxycytosine. Like the efficient insertion of dCTP opposite P7AI discussed above, the efficient pairing of PP and ImPy with deoxycytosine is not reflected in the thermal stability of the base pair, as dC:PP and dC:ImPy are both 2.2 °C less stable than dC:7AI, again implying that the factors governing efficient replication are decoupled from those governing pair stability in duplex DNA. The specific increase in efficiency must result from either polymerase-mediated interactions or an early insertion transition state bearing little resemblance to products.

3.2. Optimization of Interbase Hydrophobic Interactions for Efficient Replication. The efficiencies of unnatural base insertion opposite another unnatural base in the template show a surprising trend: while modifications of the incoming triphosphate base may have large effects on the rate of base pair synthesis, the base in the template is of less importance. Each triphosphate is inserted with little variation in rate opposite each

(37) Gilchrist, T. L. *Heterocyclic Chemistry*; Longman Scientific & Technical: Essex, 1985.
(38) Bugg, C. E. In *Solid-state packing patterns of purine bases*; Bergmann, E. D., Pullman, B., Eds.; The Israel Academy of Sciences and Humanities: Jerusalem, 1971; pp 178−204.
(39) Drew, H. R.; Dickerson, R. E. *J. Mol. Biol.* **1981**, *151*, 535−56.
(40) Drew, H. R.; Travers, A. A. *Cell* **1988**, *37*.
(41) Kopka, M. L.; Fratini, A. V.; Drew, H. R.; Dickerson, R. E. *J. Mol. Biol.* **1983**, *163*, 129−46.
(42) Liepinsh, E.; Otting, G.; Wüthrich, K. *Nucleic Acids Res.* **1992**, *20*, 6549−53.
(43) Kubinec, M. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 8739−8740.

unnatural base in the template. Opposite each 7AI analogue in the template dMICSTP is inserted the most efficiently, followed by dPICSTP and dICSTP, and finally d5MICSTP. Opposite each ICS analogue in the template, dP7AITP is inserted the most efficiently, followed by dPPTP, d7AITP, dImPyTP, and dM7AITP. The effects were the greatest for the insertion of the 7AI analogue triphosphates opposite the ICS analogues in the template. The insertion of a given 7AI analogue triphophate varied by 2−7-fold, depending on the unnatural base in the template, while the insertion of the 7AI analogue triphosphates opposite a given template varied by 200−600-fold. Relative to the unnatural base in the template, the unnatural base of the triphosphate was 2−3 orders of magnitude more sensitive to the addition of methyl groups in the developing major or minor groove, propynyl substitution, or isostructural aza-substitution.

This asymmetry is not easily explained by any simple model which relies on interbase interactions for correct base-pairing. For example, improvements in shape complementarity due to methyl group substitution should stabilize the nascant base pair, regardless of whether the modified base is in the template or the triphospahte. This asymmetry also has important implications for our efforts to develop a third orthogonal base pair for expansion of the genetic alphabet. While rates for unnatural base pair synthesis are sensitive to base modification, and may approach those characteristic of native DNA synthesis, special care must be taken to prevent the efficient mispairing of the hydrophobic bases.

3.3. An Unnatural Base Pair with Improved Kinetic Properties. Efforts to expand the genetic code are reliant on development of a third base pair to supplement the naturally existing base pairs, dA:dT and dG:dC. The nucleobases that comprise the third base pair must be thermodynamically and kinetically orthogonal to the natural bases. Base pairs reported above between 7AI analogues and ICS analogues are significant improvements over those previously reported.[18,19] For example, the reasonable stability of the M7AI:5MICS pair, as well as the improved selectivity against self-pair synthesis, make it a potentially interesting target for further optimization. However, the low efficiency for M7AI:5MICS synthesis by KF currently limits its practical utility.

The PP:MICS unnatural base pair is more promising, even though it is not thermodynamically selective against the MICS:MICS self-pair. In the template, PP directs KF to incorporate dMICSTP very efficiently ($k_{cat}/K_M = 2.0 \times 10^6$ M^{-1} min^{-1}), while dPPTP is inserted opposite MICS with an equally efficient k_{cat}/K_M value of 1.4×10^6 M^{-1} min^{-1}. These rates are both only approximately 20-fold reduced relative to those typical of native synthesis. The PP:MICS unnatural pair is also synthesized 4- and 12-fold faster than the self-pairs (1.7×10^5 and 3.3×10^5 M^{-1} min^{-1} for PP:PP and MICS:MICS self-pairs, respectively). The most efficiently synthesized mispairs between unnatural and native bases are dCTP insertion opposite PP (9.3×10^4 M^{-1} min^{-1}) and dATP insertion opposite MICS (2.4×10^4 M^{-1} min^{-1}), resulting in selectivities of 21- and 58-fold favoring correct unnatural pair synthesis. Moreover, preliminary results show that the PP:MICS pair is also extended by KF with reasonable efficiency.

While these rates are impressive, the selectivity is several orders of magnitude reduced relative to what will be required for information storage. However, we emphasize that the PP:MICS base pair is the first unnatural base pair to be incorporated into DNA with a measurable selectivity against all possible mispairs.

4. Conclusion and Future Prospects

Hydrophobicity is a strong and selective force in biological systems. The hybridization and substrate properties of hydrophobic bases may be tuned by judicious electronic and substituent modification. We have derivatized the earlier reported 7AI:ICS base pair and characterized the resulting PP:MICS base pair. PP:MICS is the first unnatural base pair reported that is efficiently accepted by a DNA polymerase with kinetic selectivity against all possible mispairs. We are further examining derivatives of these hydrophobic ring structures, as well as others, in a continuing effort to optimize the interbase hydrophobic interactions required for information storage and retrieval.

5. Experimental Section

General Experimental Methods. NMR spectra were obtained from Bruker ARX400, ARX500, and ARX600 spectrometers. High-resolution mass spectra were obtained by the Mass Spectroscopy Facilities at TSRI. Solvents and reagents were used as supplied from commercial sources with the following exceptions: all moisture-sensitive reagents were either distilled prior to use or stored over 4-Å molecular sieves for greater than 12 h. Tetrahydrofuran was distilled from sodium benzophenone ketyl immediately prior to use. Dichloromethane was distilled from calcium hydride. All reactions involving moisture-sensitive reagents were performed under an argon atmosphere.

General Toluoyl Deprotection Procedure. To a stirred solution of nucleoside (1 equiv) in CH_3OH (0.1 M) at room temperature was added NaOMe (>2.2 equiv). After the reaction was complete, the reaction was quenched via addition of solid NH_4Cl (excess) and concentrated. Purification via column chromatography on silica gel (gradient of CH_3OH in CH_2Cl_2) afforded the nucleoside.

General Triphosphate Synthesis Procedure. Proton sponge (1.5 equiv) and nucleoside (1 equiv) were dissolved in trimethyl phosphate (final concentration ~0.3 M) and cooled to 0 °C. $POCl_3$ (1.05 equiv) was added dropwise, and the lavender slurry was stirred at 0 °C for an additional 2 h. Tributylamine (4 equiv) was added, followed by a solution of tributylammonium pyrophosphate (>2.5 equiv) in DMF (to final concentration ~0.15 M). After 1 min, the reaction was quenched by addition of 1 M aqueous triethylammonium bicarbonate (20 vol equiv). The resulting crude solution was lyophilized and purified by reverse-phase (C18) HPLC (4−35% CH_3CN in 0.1 M NEt_3−HCO_3, pH 7.5) to afford the triphosphate as a white solid.

General Phosphoramidite Synthesis Procedure. To a solution of nucleoside (1 equiv) in pyridine (0.2 M) was added triethylamine (3 equiv), followed by DMTr-Cl (1.5 equiv). After being stirred for 30 min at room temperature, the reaction mixture was concentrated. Purification via column chromatography on silica gel (50−100% ethyl acetate in hexane) afforded the tritylated product, which was dissolved in CH_2Cl_2 (0.1 M) and cooled to 0 °C. A catalytic amount of DMAP was added, followed by triethylamine (5 equiv) and 2-cyanoethyl diisopropylaminochlorophosphoramidite (2 equiv). After 15 min, the reaction mixture was partitioned between CH_2Cl_2 and saturated aqueous $NaHCO_3$. The layers were separated, and the aqueous layer was extracted twice with CH_2Cl_2. The combined organics were dried over Na_2SO_4, filtered, and concentrated. Purification via column chromatography on silica gel (15−50% ethyl acetate in 5% triethylamine/hexane) afforded the desired phosphoramidite as a mixture of two diastereomers.

General Procedure for Oligonucleotide Synthesis and Purification. Oligonucleotides containing unnatural bases were synthesized with an ABI 392 DNA/RNA synthesizer using standard β-cyanoethyl phosphoramidite chemistry. Oligonucleotides were cleaved from solid support and deprotected by 12 h of incubation with concentrated ammonia at 55 °C. Oligonucleotides were then purified by preparative 15% denaturing polyacrylamide gel electrophoresis, isolated from the gel, and quantitated by absorbance at 260 nm.

General Polymerase Kinetic Assay Protocol. Primer was 5′-end ^{32}P-labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP (10 nCi/10 μL reaction). Primer−template duplexes were annealed by mixing in

the reaction buffer, heating to 90 °C, and slowly cooling to room temperature. Assay conditions: 40 nM template−primer duplex, 0.11−1.34 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× dNTP stock solution, incubated at room temperature for 1−10 min, and quenched by the addition of 20 μL of loading buffer (95% formamide, 20 mM EDTA). The reaction mixture (5 μL) was then analyzed by 15% polyacrylamide gel electrophoresis. Radioactivity was quantified using a PhosphorImager (Molecular Dynamics) and the ImageQuant program. The kinetic data were fit to the Michaelis−Menten equation using the program Kaleidograph (Synergy Software). Data presented are averages of duplicates or triplicates.

Compound 1. To a stirred solution of 6-methyl-7-azaindole (56 mg, 0.427 mmol) in DMF (2 mL) at 0 °C was added sodium hydride (12 mg, 0.512 mmol). The resulting dark brown mixture was stirred at 0 °C for 15 min, at which time it was added slowly dropwise to a solution of chloroglycoside (199 mg, 0.512 mmol) in DMF (3 mL). After being stirred at 0 °C for 1 h, the reaction mixture was partitioned between saturated aqueous $NaHCO_3$ (15 mL) and ethyl acetate (20 mL). The layers were separated, and the aqueous layer was extracted with 2 × 20 mL of ethyl acetate. The combined organics were dried over Na_2SO_4, filtered, and concentrated. Purification via column chromatography on silica gel (25−70% ethyl acetate in hexane) afforded bis-tolyl nucleoside **10** (105 mg, 51% yield), which was dissolved in CH_3OH (4 mL) and deprotected using the general procedure outlined above. Purification by column chromatography on silica gel (1−4% CH_3OH in CH_2Cl_2) afforded nucleoside **1** (31 mg, 58%): ^{1}H NMR (600 MHz, $CDCl_3$) δ 7.82 (1H, d, J = 8.0 Hz), 7.13 (1H, d, J = 3.6 Hz), 6.96 (1H, d, J = 8.0 Hz), 6.37 (1H, m), 6.29 (1H, dd, J = 9.8, 5.6 Hz), 4.79 (1H, d, J = 5.2 Hz), 4.21 (1H, s), 4.00 (1H, dd, J = 12.5, 1.7 Hz), 3.84 (1H, dd, J = 12.5, 1.6 Hz), 3.25 (1H, ddd, J = 13.6, 9.8, 5.2 Hz), 2.61 (3H, s), 2.21 (1H, dd, J = 13.5, 5.6 Hz); ^{13}C NMR (150 MHz, $CDCl_3$) δ 151.4, 145.6, 130.4, 130.0, 129.1, 127.9, 121.2, 116.8, 100.2, 90.2, 88.7, 73.9, 63.7, 39.9, 23.4; HRMS calcd for $C_{13}H_{17}N_2O_3$ (MH^+) 249.1239, found 249.1232.

Compound 2. Compound 2 was synthesized according to a literature procedure:[44] ^{1}H NMR (400 MHz, CD_3OD) δ 8.61 (1H, s), 8.36 (1H, dd, J = 4.9, 1.4 Hz), 8.10 (1H, dd, J = 8.1, 1.4 Hz), 7.36 (1H, dd, J = 8.1, 4.8 Hz), 6.57 (1H, dd, J = 7.9, 6.0 Hz), 4.61 (1H, m), 4.09 (1H, dd, J = 5.9, 3.1 Hz), 3.85 (1H, dd, J = 12.2, 3.1 Hz), 3.75 (1H, dd, J = 12.2, 3.4 Hz), 2.89 (1H, ddd, J = 13.6, 7.9, 5.9 Hz), 2.43 (1H, ddd, J = 13.4, 6.1, 2.8 Hz); ^{13}C NMR (100 MHz, CD_3OD) δ 147.0, 145.3, 145.1, 137.1, 129.2, 120.2, 89.8, 87.1, 73.1, 63.7, 41.3.

Compound 3: ^{1}H NMR (400 MHz, CD3OD) δ 8.37 (1H, d, J = 2.7 Hz), 8.24 (1H, d, J = 2.7 Hz), 8.05 (1H, d, J = 3.8 Hz), 6.68 (1H, dd, J = 8.1, 6.0 Hz), 6.67 (1H, d, J = 3.8 Hz), 4.53 (1H, m), 3.99 (1H, dd, J = 6.6, 4.1 Hz), 3.77 (1H, dd, J = 12.0, 3.5 Hz), 3.70 (1H, dd, J = 12.0, 4.1 Hz), 2.72 (1H, ddd, J = 13.8, 7.7, 6.0 Hz), 2.35 (1H, ddd, J = 13.5, 6.2, 3.0 Hz); HRMS calcd for $C_{11}H_{14}N_3O_3$ (MH^+) 236.1035, found 236.1034.

Compound 4. To a stirred solution of **12** (450 mg, 0.96 mmol) in CH_2Cl_2 (5 mL) was added 1 M ICl in CH_2Cl_2 (1.3 mL, 1.3 mmol), and the mixture was heated to 50 °C for 1 min. After cooling to room temperature over 30 min, the mixture was diluted with CH_2Cl_2 (20 mL), and the excess ICl was quenched with saturated aqueous $Na_2S_2O_3$ (15 mL). The layers were separated, and the aqueous layer was extracted with 3 × 20 mL of CH_2Cl_2. The combined organic layers were washed with brine, dried over Na_2SO_4, filtered, and concentrated in vacuo. Purification by column chromatography on silica gel (20% ethyl acetate in hexane) afforded 540 mg of white crystals, which were suspended in NEt_3 (20 mL) in a pressure tube. Argon was bubbled through the mixture for 15 min, and $(Ph_3P)_2PdCl_2$ (63 mg, 0.09 mmol) and CuI (34 mg, 0.18 mmol) were added. The mixture was cooled to −78 °C, and propyne (∼1 mL) was condensed in the tube. The reaction was sealed and allowed to warm to room temperature over 4 h. Prior to quenching, the reaction mixture was vented at −78 °C and then warmed to room temperature and concentrated. The crude residue was dissolved in $CHCl_3$ (50 mL), and the organic layer was washed two times with aqueous 5% EDTA and brine, dried over Na_2SO_4, filtered, and concentrated. Purification by column chromatography on silica gel (20% ethyl acetate in hexane) afforded **13** (330 mg). To a stirred solution of **13** (330 mg, 0.65 mmol) in CH_3OH (10 mL) was added 1 M NaOMe (2 mL). After 45 min, the excess NaOMe was quenched with NH_4Cl (∼100 mg). The resulting slurry was concentrated and purified by column chromatography on silica gel (hexane/ethyl acetate/MeOH 4:4:1), which afforded **4** (160 mg, 59% over three steps): ^{1}H NMR (400 MHz, CD_3OD) δ 8.25 (1H, d, J = 4.3 Hz), 8.04 (1H, dd, J = 8.1, 1.5 Hz), 7.35 (1H, s), 7.15 (1H, dd, J = 8.1, 4.8 Hz), 6.28 (1H, dd, J = 9.5, 5.7 Hz), 4.75 (1H, d, J = 5.1 Hz), 4.21 (1H, m), 3.96 (1H, m), 3.80 (1H, m), 3.18 (1H, ddd), 2.23 (1H, ddd, J = 13.5, 5.9, 3.7 Hz), 2.10 (3H, s); HRMS calcd for $C_{15}H_{17}N_2O_5$ (MH^+) 273.1234, found 273.1235.

Compound 5. NaH (60%, 70 mg, 1.7 mmol) was suspended in acetonitile (3 mL) and cooled to 0 °C. Compound **14** (265 mg, 1.65 mmol) was added in acetonitile (2 mL) over 15 min. The cooling bath was removed, and the mixture was stirred for 1 h. Chloroglycoside **7** (622 mg, 1.6 mmol) was added in four portions, and the mixture was stirred for 2 h. The slurry was filtered through Celite, and the residue was washed with Et_2O (∼20 mL). The filtrate was concentrated, and purification by column chromatography on silica gel afforded 560 mg of a 1:1 mixture of the *N*-glycosidic and the *O*-glycosidic compounds. To a solution of the mixture in CH_3OH (10 mL) and THF (2 mL) was added 1 M NaOMe (3 mL). After 45 min, the excess NaOMe was quenched with solid NH_4Cl (∼200 mg). The resulting slurry was concentrated and purified by column chromatography on silica gel (hexane/ethyl acetate/MeOH 4:4:1), which afforded 160 mg of **5** (34% over two steps): ^{1}H NMR (250 MHz, $CDCl_3$) δ 7.44 (1H, d, J = 7.6, 7.3 Hz), 7.31 (1H, d, J = 7.3 Hz), 7.28 (1H, d, J = 7.6 Hz), 7.19 (1H, d, J = 7.5 Hz), 6.50 (1H, dd, J = 7.2, 6.3 Hz), 6.43 (1H, d, J = 7.5 Hz), 4.60 (1H, ddd, J = 6.5, 4.0, 1.0 Hz), 4.10 (1H, ddd, J = 6.5, 4.0, 1.0 Hz), 3.94 (1H, dd, J = 12.1, 3.3 Hz), 3.85 (1H, dd, J = 12.1, 4.0 Hz), 2.92 (3H, s), 2.48 (1H, ddd, J = 13.5, 6.3, 4.0 Hz), 2.37 (1H, ddd, J = 13.5, 7.2, 1.0 Hz); HRMS calcd for $C_{15}H_{17}NO_4Na$ (MNa^+) 298.1050, found 298.1050.

Compound 6: ^{1}H NMR (400 MHz, CD_3OD) δ 8.17 (1H, d, J = 8.2 Hz), 7.72 (1H, d, J = 7.6 Hz), 7.40 (1H, s), 7.34 (1H, dd, J = 8.4, 1.5 Hz), 6.68 (1H, dd, J = 7.3, 6.2 Hz), 6.64 (1H, d, J = 7.6 Hz), 4.42 (1H, dt, J = 6.5, 3.2 Hz), 3.99 (1H, dd, J = 7.5, 3.8 Hz), 3.82 (1H, dd, J = 12.0, 3.5 Hz), 3.76 (1H, dd, J = 11.9, 4.4 Hz), 3.30 (1H, m), 2.47 (3H, s), 2.40 (1H, ddd, J = 13.5, 6.2, 3.5 Hz), 2.18 (1H, ddd, J = 13.6, 7.3, 6.5 Hz); ^{13}C NMR (150 MHz, CD_3OD) δ 158.0, 139.4, 133.2, 124.1, 122.7, 122.4, 121.4, 118.7, 102.4, 83.4, 80.8, 66.9, 57.5, 36.5, 16.2; HRMS calcd for $C_{15}H_{17}NO_4Na$ (MNa^+) 298.1050, found 298.1062.

Compound 11. To a stirred solution of **9** (100 mg, 0.840 mmol) in CH_3CN (4.5 mL) was added NaH (36 mg, 0.900 mmol, 60% dispersion in mineral oil), and the resulting mixture was stirred at room temperature for 10 min. The sodium salt was added to chloroglycoside **7** (400 mg, 1.028 mmol) and stirred at room temperature for 10 min, at which time the reaction was partitioned between ethyl acetate (25 mL) and saturated aqueous $NaHCO_3$ (20 mL). The layers were separated, and the aqueous layer was extracted with 2 × 20 mL of ethyl acetate. The combined organics were dried over Na_2SO_4, filtered, and concentrated. Purification via column chromatography on silica gel (15−30% ethyl acetate in hexane) afforded the bis-protected nucleoside **11** (161 mg, 41%): ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.38 (1H, m), 8.08 (2H, d, J = 6.4 Hz), 8.02 (2H, d, J = 6.5 Hz), 7.78 (1H, d, J = 2.1 Hz), 7.36 (2H, d, J = 6.4 Hz), 7.34 (1H, s), 7.33 (2H, d, J = 6.7 Hz), 6.91 (1H, dd, J = 6.9, 4.4 Hz), 6.83 (1H, m), 5.87 (1H, m), 4.80 (1H, dd, J = 9.7, 3.2 Hz), 4.73 (1H, dd, J = 9.4, 3.2 Hz), 4.69 (1H, m), 3.08 (1H, ddd, J = 11.6, 6.6, 5.0 Hz), 2.83 (1H, ddd, J = 11.4, 4.5, 1.8 Hz), 2.52 (3H, s), 2.50 (3H, s).

Compound 14. Data are presented for comparison to literature values:[45] ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.48 (1H, dd, J = 7.9, 7.3 Hz), 7.35 (1H, d, J = 7.9 Hz), 7.22 (1H, d, J = 7.3 Hz), 7.01 (1H, d, J = 7.0 Hz), 6.44 (1H, d, J = 7.0 Hz), 2.93 (3H, s); ^{13}C NMR (125 MHz) δ 221.7, 164.5, 141.9, 131.8, 129.8, 127.1, 124.5, 106.9, 23.7.

(44) Wenzel, T.; Seela, F. *Helv. Chim. Acta* **1996**, *79*, 169−178.

(45) Hirao, K.; Tsuchiya, R.; Yano, Y.; Tsue, H. *Heterocycles* **1996**, *42*, 415−422.

Compound 17. To a slurry of **15** (800 mg, 5.0 mmol) in CH_3CN (14.5 mL) was added bis(trimethylsilyl)acetamide (1.24 mL, 5.0 mmol). The mixture became homogeneous over 30 min, at which time it was added to chloroglycoside **7** (1.5 g, 3.86 mmol) in CH_3CN (17.4 mL). The resulting slurry was cooled to 0 °C, and $SnCl_4$ (85 μL, 0.73 mmol) was added. After 30 min, the reaction was partitioned between ethyl acetate (100 mL) and saturated aqueous $NaHCO_3$ (100 mL). The layers were separated, and the aqueous layer was extracted with 2 × 100 mL of ethyl acetate. The combined organics were dried over Na_2SO_4, filtered, and concentrated. Purification via column chromatography on silica gel (15−40% ethyl acetate in hexane) afforded protected nucleoside **17** (253 mg, 13%): ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.38 (1H, d, J = 8.2 Hz), 8.06 (2H, d, J = 8.4 Hz), 8.01 (2H, d, J = 8.4 Hz), 7.49 (1H, d, J = 7.7 Hz), 7.29−7.40 (6H, m), 6.95 (1H, dd, J = 8.3, 5.5 Hz), 6.45 (1H, d, J = 7.7 Hz), 5.72 (1H, m), 4.81 (2H, m), 4.67 (1H, dd, J = 6.1, 3.4 Hz), 2.94 (1H, ddd, J = 14.3, 5.4, 1.8 Hz), 2.55 (3H, s), 2.52 (3H, s), 2.49 (3H, s), 2.42 (1H, m); HRMS calcd for $C_{31}H_{29}NO_6Na$ (MNa^+) 534.1893, found 534.1897.

Compound 18: ^{31}P NMR (140 MHz, 50 mM Tris, 2 mM EDTA, pH 7.5 in D_2O) δ −6.03 (br), −10.41 (d, J = 15.4 Hz), −21.54 (br).

Compound 19: ^{31}P NMR (140 MHz, 50 mM Tris, 2 mM EDTA, pH 7.5 in D_2O) δ −6.35 (d, J = 16.8 Hz), −10.51 (d, J = 16.8 Hz), −21.68 (t, J = 16.8 Hz).

Compound 20: ^{31}P NMR (140 MHz, 50 mM Tris, 2 mM EDTA, pH 7.5 in D_2O) δ −5.16 (d, J = 18.0 Hz), −9.94 (d, J = 16.8 Hz), −21.27 (dd, J = 18.0, 17.4 Hz).

Compound 21: ^{31}P NMR (140 MHz, 50 mM Tris, 2 mM EDTA, pH 7.5 in D_2O) δ −5.68 (d, J = 20.9 Hz), −10.46 (d, J = 20.1 Hz), −21.84 (dd, J = 20.9, 20.1 Hz).

Compound 22: ^{31}P NMR (140 MHz, 50 mM Tris, 2 mM EDTA, pH 7.5 in D_2O) δ −5.82 (d, J = 21.4 Hz), −10.52 (d, J = 19.8 Hz), −21.90 (t, J = 20.6 Hz).

Compound 23: ^{31}P NMR (140 MHz, 50 mM Tris, 2 mM EDTA, pH 7.5 in D_2O) δ −5.79 (d, J = 18.5 Hz), −10.68 (d, J = 17.2 Hz), −21.97 (dd, J = 18.5, 17.2 Hz).

Compound 24: ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.74 (1H, d, J = 7.9 Hz), 7.43 (2H, m), 7.32 (5H, m), 7.17−7.25 (3H, m), 6.93 (1H, d, J = 7.9 Hz), 6.88 (1H, m), 6.77 (4H, m), 6.38 (1H, dd, J = 3.7, 1.9 Hz), 4.75 (1H, m), 4.21 (1H, m), 3.57−3.89 (10H, m), 3.36 (1H, m), 3.28 (1H, m), 2.71 (1H, m), 2.62 (1H, t, J = 2.4 Hz), 2.58 (3H, d, J = 1.7 Hz), 2.54 (1H, m), 2.45 (1H, m), 1.16−1.20 (9H, m), 1.09 (3H, d, J = 6.8 Hz); ^{31}P NMR (140 MHz, $CDCl_3$) δ 149.1, 148.9.

Compound 25: ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.32 (1H, dd, J = 4.8, 1.4 Hz), 8.23 (1H, d, J = 8.7 Hz), 8.05 (1H, dt, J = 8.0, 1.4 Hz), 7.39 (2H, m), 7.15−7.30 (9H, m), 6.76 (4H, m), 6.56 (1H, m) 4.77 (1H, m), 4.29 (1H, m), 3.55−3.90 (10H, m), 3.30−3.40 (2H, m), 2.92 (1H, m), 2.64 (1H, m), 2.61 (1H, t, J = 6.4 Hz), 2.46 (1H, t, J = 6.4 Hz), 1.16−1.19 (9H, m), 1.11 (3H, d, J = 6.8 Hz); ^{31}P NMR (140 MHz, $CDCl_3$) δ 149.3, 149.2.

Compound 27: ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.31 (1H, m), 7.97 (1H, d, J = 7.7 Hz), 7.56 (0.5H, s), 7.53 (0.5H, s), 7.42−7.45 (2H, m), 7.31−7.34 (4H, m), 7.18−7.29 (3H, m), 7.13 (1H, m), 6.76−6.85 (4H, m), 4.69 (1H, m), 4.18−4.25 (1H, m), 3.79 (3H, s), 3.78 (3H, s), 3.54−3.70 (2H, m), 3.36 (0.5H, dd, J = 9.9, 4.0 Hz), 3.25−3.32 (1.5H, m), 2.57−2.64 (2H, m), 2.47−2.53 (1H, m), 2.45 (1H, t, J = 6.2 Hz), 2.14 (3H, s), 1.19 (3H, d, J = 6.7 Hz), 1.18 (3H, d, J = 6.4 Hz), 1.16 (3H, d, J = 6.3 Hz),1.09 (3H, d, J = 7.0 Hz).

Compound 28: ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.69 (0.5H, d, J = 7.3 Hz), 7.60 (0.5H, d, J = 7.5 Hz), 7.43−7.48 (3H, m), 7.19−7.37 (9H, m), 6.84 (2H, d, J = 8.5 Hz), 6.82 (2H, d, J = 8. 8 Hz), 6.70 (0.5H, dd, J = 6.3, 6.2 Hz), 6.67 (0.5H, dd, J = 6.3, 6.2 Hz), 6.26 (0.5H, d, J = 7.5 Hz), 6.25 (0.5H, d, J = 7.5 Hz), 4.65 (0.5H, m), 4.60 (0.5H, m), 4.19 (1H, m), 3.79 (3H, s), 3.78 (3H, s), 3.48−3.61 (4H, m), 3.32−3.38 (1H, m), 2.92 (3H, s), 2.68 (0.5H, ddd, J = 13.5, 6.2, 3.7 Hz), 2.62 (0.5H, ddd, J = 13.5, 6.2, 3.7 Hz), 2.61 (1H, dd, J = 6.4, 6.2 Hz), 2.40 (1H, dd, J = 6.6, 6.4 Hz), 2.26 (0.5H, ddd, J = 13.5, 6.8, 6.3 Hz), 2.24 (0.5H, ddd, J = 13.5, 6.8, 6.3 Hz), 2.06 (3H, s), 1.17 (3H, d, J = 6.6 Hz), 1.16 (3H, d, J = 6.8), 1.15 (3H, d, J = 6.7 Hz), 1.05 (3H, d, J = 6.8 Hz).

Acknowledgment. Funding was provided by the National Institutes of Health (GM 60005 to F.E.R.), the Skaggs Institute for Chemical Biology (F.E.R. and P.G.S.), and a National Institutes of Health postdoctoral fellowship (F32 GM19833-01 to A.K.O.).

Supporting Information Available: Experimental procedures and characterizations (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA0009931