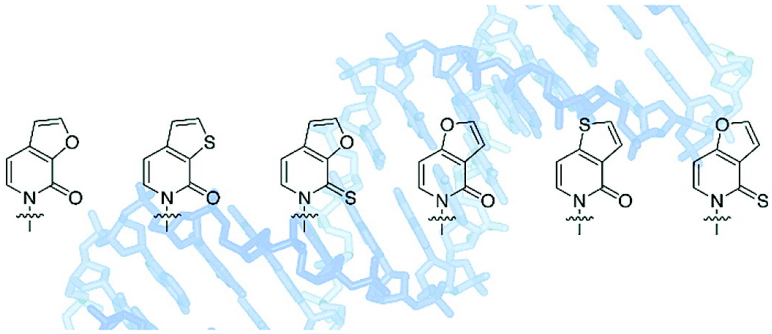
Article

Determinants of Unnatural Nucleobase Stability and Polymerase Recognition

Allison A. Henry, Chengzhi Yu, and Floyd E. Romesberg

J. Am. Chem. Soc., **2003**, 125 (32), 9638-9646• DOI: 10.1021/ja035398o • Publication Date (Web): 18 July 2003



J|A|C|S
ARTICLES
Published on Web 07/18/2003

Determinants of Unnatural Nucleobase Stability and Polymerase Recognition

Allison A. Henry, Chengzhi Yu, 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 31, 2003; E-mail: floyd@scripps.edu

Abstract: Six new unnatural nucleobases have been synthesized and characterized in terms of stability and selectivity of self-pairing in duplex DNA and efficiency and fidelity of self-pairing during polymerase-mediated replication. Each nucleobase has a conserved ring structure but differs from the others in its specific pattern of substitution with oxygen and sulfur atoms. Heteroatom derivatization within the conserved scaffold is shown to have only moderate effects on unnatural self-pair synthesis by the polymerase; larger effects were observed on the thermal stability and polymerase-mediated extension of the self-pairs. The largest effects of heteroatom substitution were on the stability and synthesis of mispairs between the unnatural and natural bases. Certain heteroatom substitutions were found to have a general effect while others were found to have effects that were specific for a particular unnatural or natural base. The data are useful for designing stable and replicable third base pairs and for understanding the contributions of nucleobase shape, polarity, and polarizability to the stability and replication of DNA.

1. Introduction

The storage and replication of biological information is based on complementary interactions between the natural nucleobases dA:dT and dG:dC. The identification of enzymatically replicable unnatural base pairs with thermal stabilities comparable to those of dA:dT and dG:dC would significantly expand the biological and chemical potential of DNA. While the hydrogen-bonding (H-bonding) patterns of the natural base pairs are an obvious source of stability and specificity, efforts to design unnatural bases with different H-bonding patterns have proven difficult.[1−5] An alternative unnatural nucleobase design strategy is based on the observation that H-bonds are not an absolute requirement for the stability or replication of DNA.[6−10] Thus, we have explored the utility of hydrophobic and van der Waals forces in developing additional base pairs to store and replicate increased genetic information.[11−17]

In characterizing the stability and replication of these non-H-bonding unnatural nucleobases, we identified several promising base pairs formed between two of the same bases. In principle, an unnatural self-pair, as opposed to a heteropair (like the natural base pairs), is not a limitation on expansion of the genetic alphabet as the addition of a single self-pair would result in 64 new codons. In practice, a self-pair is advantageous because the potential for mispairing with dA, dC, dG, and dT is reduced. Previously, the most interesting unnatural nucleobases were based on the azaindole, isocarbostyril, or naphthyl scaffolds: **7AI**:**7AI**,[13,15] **ICS**:**ICS**,[12] **PICS**:**PICS**,[11] and **3MN**:**3MN**[14] (Figure 1). These self-pairs are stable in duplex DNA and enzymatically synthesized with reasonable efficiency and selectivity; however, continued primer extension downstream of the unnatural self-pair is inefficient in each case.

There are at least two possible reasons for the observed poor unnatural self-pair extension rates. One obvious feature of natural DNA is the presence of heteroatoms in both the major and minor grooves. These atoms are not only involved in interbase H-bonding but also contribute to the physical properties of the nucleobases, including dipole moment and polarizability.[18−22]

(1) Bain, J. D.; Switzer, C.; Chamberlin, A. R.; Benner, S. A. *Nature* **1992**, *356*, 537−539.
(2) Horlacher, J.; Hottiger, M.; Podust, V. N.; Hübscher, U.; Benner, S. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6329−6333.
(3) Lutz, M. J.; Held, H. A.; Hottiger, M.; Hübscher, U.; Benner, S. A. *Nucleic Acids Res.* **1996**, *24*, 1308−1313.
(4) Lutz, M. J.; Horlacher, J.; Benner, S. A. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1149−1152.
(5) Lutz, S.; Burgstaller, P.; Benner, S. A. *Nucleic Acids Res.* **1999**, *27*, 2792−2798.
(6) Moran, S.; Ren, R. X.-F.; Rumney, S. I.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056−2057.
(7) Moran, S.; Ren, R. X.-F.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10506−10511.
(8) Morales, J. C.; Kool, E. T. *Nat. Struct. Biol.* **1998**, *5*, 950−954.
(9) Kool, E. T. *Biopolymers* **1998**, *48*, 3−17.
(10) Kool, E. T.; Morales, J. C.; Guckian, K. M. *Angew. Chem., Int. Ed.* **2000**, *39*, 990−1009.
(11) McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **1999**, *121*, 11585−11586.
(12) Ogawa, A. K.; Wu, Y.; McMinn, D. L.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 3274−3287.
(13) Wu, Y.; Ogawa, A. K.; Berger, M.; McMinn, D. L.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 7621−7632.
(14) Ogawa, A. K.; Wu, Y.; Berger, M.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 8803−8804.
(15) Tae, E. L.; Wu, Y. Q.; Xia, G.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2001**, *123*, 7439−7440.
(16) Berger, M.; Luzzi, S. D.; Henry, A. A.; Romesberg, F. E. *J. Am. Chem. Soc.* **2002**, *124*, 1222−1226.
(17) Yu, C.; Henry, A. A.; Romesberg, F. E.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2002**, *41*, 3841−3844.
(18) 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.
(19) Lister, J. H. *The Purines. Supplement 1*; John Wiley and Sons: New York, 1996; Vol. 54.

10.1021/ja035398o CCC: $25.00

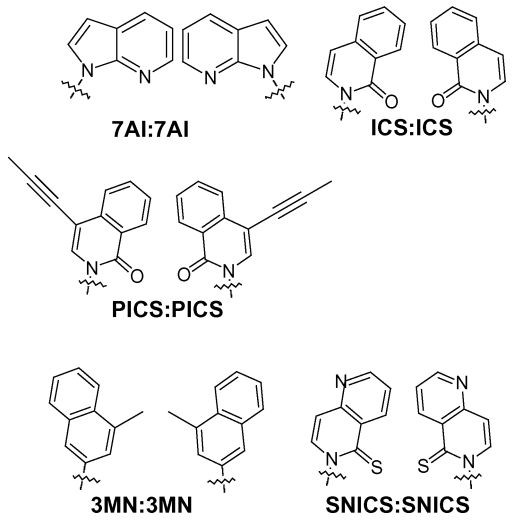


Figure 1. 7AI, ICS, PICS, 3MN, and SNICS self-pairs.

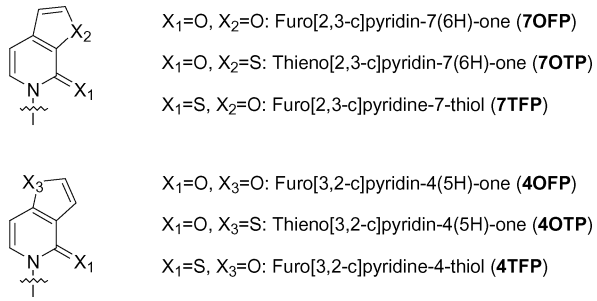


Figure 2. Unnatural nucleobases. Derivatization with an oxygen or sulfur heteroatom is indicated by the letter X and differentiated by a numerical subscript corresponding to positions 1, 2, and 3, assumed to be minor groove, self-pair interface, and major groove, respectively (see text).

It is not simple to evaluate how these properties affect stability and replication, but it seems possible that the general absence of these heteratoms from the previously reported unnatural bases could contribute to their poor extension efficiency.[23] For example, **ICS** lacks heteroatoms except for the carbonyl that is expected to be oriented toward the minor groove. By comparison, its analogue, **SNICS** (Figure 1), with a nitrogen and a sulfur atom that may be positioned in the major and minor grooves, respectively, forms self-pairs that are extended with improved efficiency.[17] The second possible origin of the poor extension rates may be the large aromatic surface area of the unnatural nucleobases. This could cause steric repulsion or interstrand intercalation between the unnatural bases that results in distortion of the primer terminus to a geometry that is inappropriate for continued primer extension.

To systematically evaluate the effects of surface area and heteroatom derivatization, we synthesized and characterized six nucleobases that share a scaffold based on the fusion of a pyridone or thiopyridone with a furan or thiophene (Figure 2). These base analogues can be coupled with 2′-deoxyribofuranose by an *N*-glycosidic linkage while retaining aromaticity. Each has reduced surface area, compared to the aforementioned bases, and may be derivatized with heteroatoms likely to be located in the minor groove (position 1), at the nucleobase interface (position 2) or in the major groove (position 3). This provides an opportunity to vary the heteroatom substitution pattern in a controlled manner and is expected to help deconvolute the contributions of nucleobase shape, hydrophobicity, dipole moment, and polarizability to the stability and replication of the unnatural self-pairs, as well as the mispairs between unnatural and natural bases. Herein, we show that heteroatom substitution has variable effects on the stability and polymerase-catalyzed synthesis of the unnatural self-pairs and mispairs. Studies of this nature not only aid in the design of unnatural nucleobases to expand the genetic alphabet but also contribute to a general understanding of the determinants of DNA stability and replication.

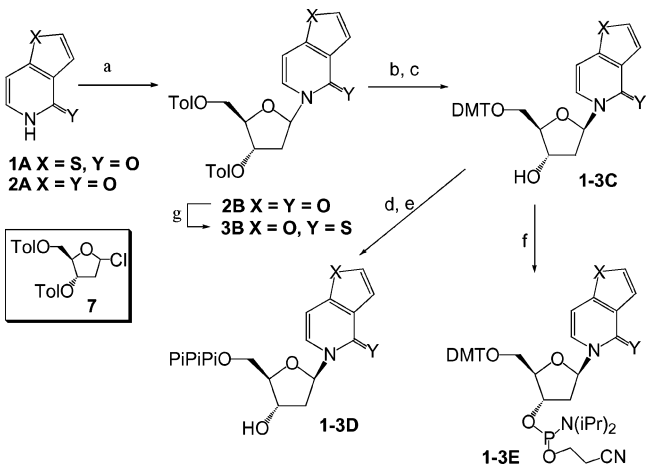
2. Results

2.1. Base Pair Design and Synthesis. Using the modeling program Insight II (Biosym/MSI), we predicted that self-pairs formed between unnatural nucleobases comprised of six- and five-membered ring fusions would be well accommodated in B-form DNA. For the six-membered ring component, pyridone and thiopyridone were chosen to give an *N*-glycoside with either an oxygen or a sulfur atom positioned in the minor groove. For the five-membered ring component, furan and thiophene were chosen in order to allow manipulation of the physical properties of the unnatural nucleobases, such as their dipole moment and polarizability. These elements lead to the six unnatural nucleobase analogues (see Figure 2): furo[3,2-c]pyridin-4(5H)-one (**4OFP**), thieno[3,2-c]pyridin-4(5H)-one (**4OTP**), furo[3,2-c]-pyridine-4-thiol (**4TFP**), furo[2,3-c]pyridin-7(6H)-one (**7OFP**), thieno[2,3-c]pyridin-7(6H)-one (**7OTP**), and furo[2,3-c]pyridine-7-thiol (**7TFP**). To facilitate the following discussion, we refer to heteroatom substitution at the positions labeled 1, 2, and 3 (Figure 2). Substitution at position 1 is expected to orient the heteroatom in the minor groove of duplex DNA (or in the developing minor groove during replication); substitution at position 2 is expected to bury the heteroatom within the duplex; and substitution at position 3 is expected to position the heteroatom in the major groove. These structural predictions are based on an assumption that the unnatural nucleobases will adopt a pyrimidine-like conformation in duplex DNA.

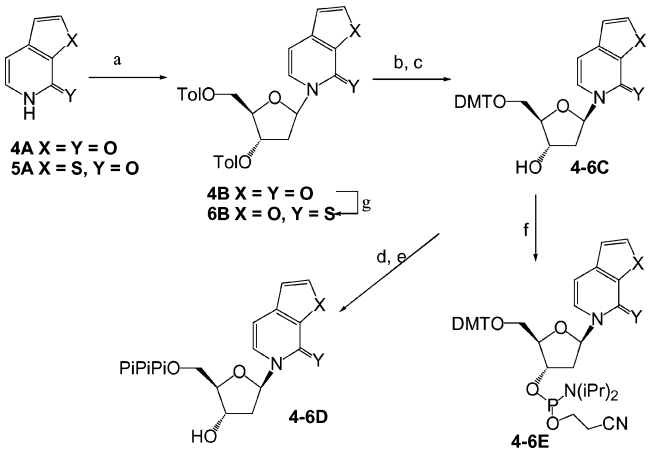
Unnatural bases **1**−**3** and **4**−**6** were synthesized as shown in Schemes 1 and 2, respectively. Starting materials **1A**, **2A**, **4A**, and **5A** were prepared using a literature protocol.[24] Analogues **3B** and **6B**, which contain a thiocarbonyl group, were synthesized from the corresponding lactams using Lawesson's chemistry, as previously reported.[17] Briefly, formation of the *N*-glycosidic bond was effected by silyl−Hilbert−Johnson methodology under Vorbruggen conditions.[25] After removal of the toluoyl protecting groups and 5′-tritylation of the free nucleoside, the desired β-anomer was easily separated from the α-anomer by silica gel column chromatography. The assignment of β-stereochemistry at C1′ for each dimethoxytrityl (DMTr)-protected nucleoside was based on COSY and NOESY spectra, which showed cross-peaks between H1′ and both H4′ and α-H2′. The free nucleosides to be used in the syntheses of the corresponding

(20) Hurst, D. T. *An Introduction to the Chemistry and Biochemistry of Pyrimidines, Purines and Pteridines*; Page Bros.: Norwich, U. K., 1980.
(21) Gilchrist, T. L. *Heterocyclic Chemistry*; Longman Scientific & Technical: Essex, 1985.
(22) 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.
(23) Guo, M.-J.; Hildbrand, S.; Leumann, C. J.; McLaughlin, L. W.; Waring, M. J. *Nucleic Acids Res.* **1998**, *26*, 1863−1869.
(24) Shiotani, S.; Morita, H. *J. Heterocyclic Chem.* **1982**, *19*, 1207−1209.
(25) Niedballa, U.; Vorbruggen, H. *J. Org. Chem.* **1974**, *39*, 3654−3660.

Scheme 1 [a]



Scheme 2 [a]



[a] General conditions: (a) Bis-TMS acetamide, rt, **7**, then $SnCl_4$, 0 °C; (b) 0.5 M NaOMe, MeOH, rt; (c) DMTrCl, pyridine, rt; (d) $LiBF_4/CH_3CN$, MeOH−CH_2Cl_2 (1:9), rt; (e) Proton Sponge, $POCl_3$, Bu_3N, Bu_3NPPi, $(MeO)_3P$, DMF, 0 °C; (f) 2-Cyanoethyl diisopropylchlorophosphoramidite, DIEA, CH_2Cl_2, 0 °C; (g) Lawesson's reagent, toluene, reflux.

triphosphates were obtained by treatment of the DMTr-protected nucleoside intermediates with $LiBF_4$.[26] In all cases, preparation of phosphoramidites from the corresponding DMTr-protected nucleosides followed reported procedures.[11]

2.2. Unnatural Self-Pair Stability and Selectivity in Duplex DNA. The stabilities of the unnatural self-pairs and the mispairs with natural bases were determined by thermal denaturation of duplex DNA. The duplex melting temperature (T_m) was measured using the complementary oligonucleotides 5′-GC-GATGXGTAGCG-3′ and 5′-CGCTACYCATCGC-3′, with either an unnatural or a natural base incorporated at positions X and Y (Table 1).

Heteroatom substitution was found to have significant effects on self-pair stability in duplex DNA. For example, comparison of **4OFP** to **4TFP** and **7OFP** to **7TFP** shows that in both cases an amide conferred greater self-pair stability than a thioamide, by 1.3 °C. Heteroatom substitution in the five-membered ring had less general effects on duplex stability, though in all cases placement of the heteroatom at position 3 resulted in more stable self-pairs than placement of the heteroatom at position 2. This effect was less pronounced with oxygen as the heteroatom,

(26) Chen, A.; Zheng, Y.; Zhou, X. *Synth. Commun.* **1999**, *29*, 3421−3423.

Table 1. Denaturation Temperatures for Duplex DNA Containing Unnatural Base Pairs[a]

5′-dGCGTAC*X*CATGCG
3′-dCGCATG*Y*GTACGC

X	Y	T_m, °C	X	Y	T_m, °C
4OTP	**4OTP**	57.9 ± 0.04	**4OFP**	**4OFP**	55.0 ± 0.07
	dA	51.5 ± 0.10		dA	51.7 ± 0.27
	dT	50.6 ± 0.02		dT	47.2 ± 0.19
	dC	49.0 ± 0.03		dC	48.9 ± 0.04
	dG	49.8 ± 0.06		dG	50.7 ± 0.33
4TFP	**4TFP**	53.7 ± 0.14	**7OFP**	**7OFP**	53.3 ± 0.10
	dA	50.1 ± 0.03		dA	50.3 ± 0.11
	dT	50.2 ± 0.10		dT	46.6 ± 0.14
	dC	46.2 ± 0.13		dC	48.6 ± 0.31
	dG	47.0 ± 0.11		dG	49.5 ± 0.16
7OTP	**7OTP**	52.6 ± 0.07	**7TFP**	**7TFP**	52.0 ± 0.01
	dA	51.2 ± 0.21		dA	48.2 ± 0.02
	dT	50.7 ± 0.24		dT	47.4 ± 0.02
	dC	45.7 ± 0.13		dC	44.4 ± 0.03
	dG	47.5 ± 0.30		dG	46.3 ± 0.04

[a] See text for experimental details.

where **4OFP** was found to be 1.7 °C more stable than **7OFP** and more pronounced with sulfur as the heteroatom, where **4OTP** was found to be 5.3 °C more stable than **7OTP**. In fact, the **4OTP** self-pair was the most stable of all, only 1.3 °C less stable than a dA:dT pair in the same sequence context (T_m = 57.9 and 59.2 °C for **4OTP**:**4OTP** and dA:dT, respectively). In contrast, in the case of heteroatom placement at position 2, the duplex containing an unnatural self-pair was slightly more stabilized by oxygen than sulfur (compare **7OFP** to **7OTP**).

Each of the unnatural nucleobases was strongly selective for self-pairing (Table 1). This thermal selectivity was greatest for **4OTP**, a self-pair that was 6.4 °C to 8.9 °C more stable than any mispair, comparable to the thermal selectivity observed among the natural base pairs. Stabilities of the mispairs between **4OTP** and the natural bases varied over a range of only 2.5 °C but generally increased with the hydrophobicity of the natural base, dA being the most hydrophobic and forming the most stable mispair. The **4OFP** self-pair was the next most selective, with mispairs destabilized by 3.3 °C to 7.8 °C, and again, the relative stabilities paralleled the hydrophobicity of the natural base. It is apparent that the slightly reduced selectivity of the **4OFP** self-pair, relative to **4OTP**, resulted from a decrease in the stability of the self-pair while the absolute stabilities of the mispairs remained the same, except in the case of dT. In this case, substitution of a sulfur atom (**4OTP**:dT) with an oxygen atom (**4OFP**:dT) resulted in a 3.4 °C decrease in duplex T_m. A comparison of **4OFP** with **4TFP** reveals that substitution of the minor groove oxygen atom with sulfur resulted in decreased stability of the **4TFP** self-pair and its mispairs to give an overall selectivity that was virtually identical to that observed with **4OFP**; however, the mispair with dT was again an exception as the presence of sulfur at position 1, like at position 3, stabilized the mispair with dT, by 3.0 °C compared to **4OFP**:dT. The **7OFP**, **7OTP**, and **7TFP** self-pairs, with oxygen or sulfur at positions 1 and 2, were 1.4 to 6.9 °C more stable than all mispairs with natural bases. The thermal selectivities of **7OFP**, **7OTP**, and **7TFP** were the result of decreases in self-pair stability as well as decreases in the stabilities of the mispairs.

2.3. Replication of DNA Containing the Unnatural Base Pairs. In general, the two components of DNA replication are

Table 2. Incorporation of Natural and Unnatural Triphosphates Opposite Unnatural Templating Bases[a]

template base	triphosphate	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}\ M^{-1}$)
4OFP	d(**4OFP**)TP	nd[b]	nd[b]	$<1.0 \times 10^3$
	dATP	0.09 ± 0.01	28 ± 20	3.4×10^3
	dCTP	nd[b]	nd[b]	$<1.0 \times 10^3$
	dGTP	1.4 ± 0.8	157 ± 63	8.8×10^3
	dTTP	2.0 ± 0.8	178 ± 76	1.2×10^4
4OTP	d(**4OTP**)TP	0.19 ± 0.09	32 ± 6	5.9×10^3
	dATP	nd[b]	nd[b]	$<1.0 \times 10^3$
	dCTP	nd[b]	nd[b]	$<1.0 \times 10^3$
	dGTP	0.6 ± 0.4	130 ± 23	4.8×10^3
	dTTP	1.8 ± 1.4	181 ± 122	1.0×10^4
7OFP	d(**7OFP**)TP	0.16 ± 0.03	32 ± 4	5.2×10^3
	dATP	0.10 ± 0.04	42 ± 29	2.5×10^3
	dCTP	nd[b]	nd[b]	$<1.0 \times 10^3$
	dGTP	53 ± 24	35 ± 6	1.5×10^6
	dTTP	1.7 ± 0.3	188 ± 14	9.1×10^3
7TFP	d(**7TFP**)TP	0.06 ± 0.01	5.3 ± 0.2	1.1×10^4
	dATP	nd[b]	nd[b]	$<1.0 \times 10^3$
	dCTP	nd[b]	nd[b]	$<1.0 \times 10^3$
	dGTP	nd[b]	nd[b]	$<1.0 \times 10^3$
	dTTP	nd[b]	nd[b]	$<1.0 \times 10^3$
7OTP	d(**7OTP**)TP	0.28 ± 0.08	17 ± 4	1.6×10^4
	dATP	0.20 ± 0.06	41 ± 12	4.9×10^3
	dCTP	nd[b]	nd[b]	$<1.0 \times 10^3$
	dGTP	2.8 ± 0.8	146 ± 63	1.9×10^4
	dTTP	14 ± 5	118 ± 48	1.2×10^5
4TFP	d(**4TFP**)TP	0.56 ± 0.07	6.0 ± 3.5	9.3×10^4
	dATP	1.4 ± 0.6	3.4 ± 0.9	4.2×10^5
	dCTP	nd[b]	nd[b]	$<1.0 \times 10^3$
	dGTP	1.8 ± 0.6	0.7 ± 0.5	2.7×10^6
	dTTP	2.4 ± 1.4	51 ± 27	4.7×10^4

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

base pair synthesis (i.e., selective incorporation of the correct triphosphate opposite a template base) and extension (synthesis of the next base pair). Each unnatural and natural base was incorporated into the primer-template substrates shown in Tables 2−4. The efficiency and selectivity of unnatural DNA synthesis by KF exo$^-$ were examined both in the context of self-pair synthesis (Tables 2 and 3) and self-pair extension (Table 4).

2.3.1. Efficiency of Unnatural Self-Pair Synthesis. Self-pair synthesis efficiencies for the six unnatural nucleobases vary over 2 orders of magnitude (Table 2). The most efficiently formed is the **4TFP** self-pair, with a specificity constant (k_{cat}/K_M) of $9.3 \times 10^4\ min^{-1}\ M^{-1}$, which is 500-fold reduced relative to a dA:dT in the same sequence context ($4.7 \times 10^7\ min^{-1}\ M^{-1}$). The next most efficient are the **7OTP** and **7TFP** self-pairs, which are synthesized 5.8- and 8.4-fold slower than the **4TFP** self-pair ($1.6 \times 10^4\ min^{-1}\ M^{-1}$ and $1.1 \times 10^4\ min^{-1}\ M^{-1}$, respectively). Following **7OTP** and **7TFP**, the next most efficiently synthesized self-pairs are those of **4OTP** and **7OFP**, with rates of $5.9 \times 10^3\ min^{-1}\ M^{-1}$ and $5.2 \times 10^3\ min^{-1}\ M^{-1}$, respectively (each approximately 16-fold reduced relative to the **4TFP** self-pair). Finally, the **4OFP** self-pair is the least efficiently synthesized of the group with an almost undetectable synthesis rate, at least 2 orders of magnitude reduced compared to **4TFP**.

2.3.2. Fidelity of Unnatural Self-Pair Synthesis. We determined the fidelity of self-pair synthesis in the context of single incorporation of an unnatural or a natural triphosphate opposite the unnatural base in the template (Table 2). Despite an only modest rate of self-pair synthesis, the **7TFP** self-pair is replicated with the greatest selectivity. In fact, no natural

Table 3. Incorporation of Unnatural Triphosphates Opposite Natural Templating Bases[a]

template base	triphosphate	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}\ M^{-1}$)
A	d(**4OFP**)TP	1.25 ± 0.03	176 ± 47	7.1×10^3
	d(**4OTP**)TP	2.5 ± 0.5	48 ± 12	5.3×10^4
	d(**7OFP**)TP	2.7 ± 1.0	40 ± 12	6.7×10^4
	d(**7TFP**)TP	2.7 ± 1.6	16 ± 7	1.7×10^5
	d(**7OTP**)TP	0.21 ± 0.02	12 ± 4	1.8×10^4
	d(**4TFP**)TP	0.4 ± 0.2	14 ± 5	3.2×10^4
C	d(**4OFP**)TP	nd[b]	nd[b]	$<1.0 \times 10^3$
	d(**4OTP**)TP	0.13 ± 0.04	30 ± 10	4.2×10^3
	d(**7OFP**)TP	0.11 ± 0.04	23 ± 6	4.7×10^3
	d(**7TFP**)TP	0.08 ± 0.05	7.1 ± 4.3	1.2×10^4
	d(**7OTP**)TP	0.11 ± 0.01	15 ± 9	7.4×10^3
	d(**4TFP**)TP	0.27 ± 0.08	8.3 ± 3.5	3.3×10^4
G	d(**4OFP**)TP	nd[b]	nd[b]	$<1.0 \times 10^3$
	d(**4OTP**)TP	41 ± 7	15 ± 1	2.8×10^6
	d(**7OFP**)TP	37 ± 11	17 ± 5	2.2×10^6
	d(**7TFP**)TP	24 ± 8	12 ± 5	2.0×10^6
	d(**7OTP**)TP	10 ± 2	2.6 ± 2.0	3.8×10^6
	d(**4TFP**)TP	13 ± 2	8.4 ± 2.6	1.5×10^6
T	d(**4OFP**)TP	1.5 ± 0.4	25 ± 13	6.2×10^4
	d(**4OTP**)TP	4.2 ± 0.2	42 ± 25	1.0×10^5
	d(**7OFP**)TP	5.5 ± 1.3	31 ± 12	1.8×10^5
	d(**7TFP**)TP	3.8 ± 1.3	17 ± 9	2.3×10^5
	d(**7OTP**)TP	0.54 ± 0.04	15 ± 8	3.5×10^4
	d(**4TFP**)TP	4.5 ± 1.3	25 ± 4	1.8×10^5

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

Table 4. Correct Single Nucleotide Extension of Unnatural Self-Pairs[a]

unnatural self-pair	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}\ M^{-1}$)
4OFP	7.4 ± 0.1	183 ± 12	4.0×10^4
4OTP	2.4 ± 0.1	181 ± 23	1.3×10^4
7OFP	8.4 ± 0.4	80 ± 3	1.1×10^5
7TFP	nd[b]	nd[b]	$<1.0 \times 10^3$
7OTP	1.6 ± 1.2	171 ± 21	9.4×10^3
4TFP	1.0 ± 0.6	4.5 ± 2.0	2.3×10^5

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

triphosphate is found to be inserted opposite **7TFP** with a detectable rate. Thus, it is not possible to measure the selectivity of **7TFP** self-pair synthesis but is only possible to estimate that it is greater than 11-fold. In all other cases, the unnatural self-pair is synthesized slower than its most efficiently synthesized mispair. Despite a somewhat reduced rate of self-pair synthesis compared to **7TFP**, **4OTP** is the next most selective unnatural base in the template. The **4OTP** self-pair is synthesized more efficiently than all its mispairs except that with dT. dTTP is inserted opposite **4OTP** 1.7-fold more efficiently than d(**4OTP**)-TP, despite a high K_M (181 μM) and because of a relatively large k_{cat} (1.8 min^{-1}). Relative to dTTP and d(**4OTP**)TP, dGTP is inserted opposite **4OTP** less efficiently, due to a decreased k_{cat} relative to dTTP and an increased K_M relative to d(**4OTP**)-TP. dATP and dCTP are inserted opposite **4OTP** at undetectable rates. The third most selective self-pair is **7OTP**, which is synthesized from 3.4-fold to at least 16-fold more efficiently than its mispairs with dA and dC, respectively; however, **7OTP** mispairs with dTTP 7.3-fold more efficiently than it correctly pairs with d(**7OTP**)TP. This is the result of an increased k_{cat} (14 min^{-1}) that overcomes the reduced binding of dTTP relative to the unnatural triphosphate. **7OTP** also directs KF to insert dGTP with a rate approximately equal to that for insertion of

the unnatural triphosphate. Despite a comparatively efficient rate of self-pair synthesis (9.3×10^4 min^{-1} M^{-1}), the **4TFP** self-pair is not selectively formed due to the remarkably efficient incorporation of both dGTP (2.7×10^6 min^{-1} M^{-1}) and dATP (4.2×10^5 min^{-1} M^{-1}). In this case, the efficiency of dGTP insertion results largely from very tight binding of the triphosphate ($K_M = 0.7$ μM). The pyrimidine triphosphates dTTP and dCTP are inserted less efficiently opposite **4TFP** than is d(**4TFP**)TP, with rates of 4.7×10^4 min^{-1}M^{-1} and $<1.0 \times 10^3$ min^{-1}M^{-1}, respectively. **4OFP** is not selective for self-pair synthesis primarily because of inefficient self-pair synthesis; dTTP and dGTP are inserted against **4OFP** 9-fold and 12-fold faster than d(**4OFP**)TP, respectively. Finally, the least selective unnatural base, in the template, is **7OFP** against which dGTP is inserted 290-fold more efficiently than d(**7OFP**)TP. Remarkably, the k_{cat} for insertion of dGTP against **7OFP** (53 min^{-1}) is only about 3-fold reduced compared to that for natural base pair synthesis.[12] The **7OFP** self-pair is synthesized 1.8-fold less efficiently than the **7OFP**:dT mispair, but 2- and 5-fold more efficiently than the mispairs with dA and dC, respectively.

We also characterized the efficiency of insertion of the unnatural triphosphates against each natural base in the template (Table 3). The rates were found to depend most strongly on which natural base was in the template. The greatest rates were consistently seen with dG in the template, followed by dT, dA, and finally dC. The decrease in rates was largely due to k_{cat} effects. Each unnatural triphosphate, with the exception of d(**4OFP**)TP, is efficiently inserted opposite dG with rates that vary only between 1.5×10^6 min^{-1} M^{-1} and 3.8×10^6 min^{-1} M^{-1}. The specificity constants are only about 20-fold less than that for natural DNA synthesis. While there was some variation in the rates of unnatural triphosphate insertion opposite dT, dA, and dC, this variation was less than 10-fold in all cases. The only exception was the reasonably efficient insertion of d(**4OFP**)TP opposite dT; d(**4OFP**)TP was inserted opposite dG, dA, and dC with efficiencies that were reduced by up to 3 orders of magnitude compared to the other five unnatural bases.

2.3.3. Extension of Self-Pair by One Natural Nucleotide. There is no obvious correlation between self-pair synthesis efficiency and self-pair extension efficiency except that the most efficiently synthesized self-pair of this family, **4TFP**, is also the most efficiently extended self-pair, with a specificity constant of 2.3×10^5 min^{-1} M^{-1} (Table 4). Thus, **4TFP** exhibits one of the best rates of single nucleotide extension of an unnatural self-pair that we have observed, and it is only 200-fold reduced relative to the rate of extension of a dA:dT in the same sequence context.[12] The second most efficiently extended self-pair of the group is **7OFP**; with a rate of 1.1×10^5 min^{-1} M^{-1}, the **7OFP** self-pair is extended only 2-fold less efficiently than the **4TFP** self-pair. The **4OFP** self-pair is the third most efficiently extended with a specificity constant of 4.0×10^4 min^{-1} M^{-1}, 5.8-fold less than **4TFP**. With a rate of 1.3×10^4 min^{-1} M^{-1}, the fourth most efficiently extended self-pair is that of **4OTP**. The least efficiently extended self-pairs are those of **7OTP** and **7TFP**, with specificity constants of 9.4×10^3 min^{-1} M^{-1} and $<1.0 \times 10^3$ min^{-1} M^{-1}, respectively. These rates represent a more than 25-fold reduction compared to the extension rate of **4TFP**. With the exception of **7TFP**, all of the self-pairs in this family are extended at least 5-fold more efficiently than any of the previous first generation unnatural base pairs, such as the **ICS** self-pair.

3. Discussion

Despite sharing the same nucleobase scaffold, the six unnatural bases were found to have very different properties with respect to self-pair stability and recognition by the polymerase. This demonstrates the importance of heteroatom substitution, which affects nucleobase hydrophobicity, polarity, and polarizability.[18−22] Several trends in the thermodynamic and kinetic data are now discussed.

Stability. In all cases, the unnatural self-pairs are more stable than all possible corresponding mispairs with natural bases. This thermodynamic selectivity illustrates the strength of the hydrophobic effect in duplex DNA, as has been observed with other unnatural base pairs.[27] The most stable self-pair was that of **4OTP**, which was only 1.3 °C less stable than a dA:dT pair in the same sequence context. The **4OTP** self-pair was also as selective for "correct" pairing as the natural bases, with each mispair between **4OTP** and a natural base conferring a T_m of 6.4 °C to 8.9 °C below that of the self-pair.

The absolute stabilities of the mispairs were more varied than those of the self-pairs, indicating that the unnatural−natural interbase interactions are more specific than simple hydrophobicity. Comparison of **4OFP** to **4TFP** and **7OFP** to **7TFP** reveals the effect of the atom at position 1, namely oxygen or sulfur, which is presumably disposed in the minor groove. For each of these four unnatural bases, the self-pair was most stable, followed by the mispairs with dA, dG, and dC, in that order. The greater stability of the purine mispairs indicates that packing interactions may be important. Both the self-pairs and mispairs (except the mispair with dT) were destabilized by the substitution of oxygen with sulfur. Since sulfur is expected to be more polarizable than oxygen, a property favorable for packing interactions, sulfur's destabilizing effect relative to oxygen likely results from unfavorable steric repulsion caused by its greater size. The four mispairs with dT behaved very differently from the other mispairs in that sulfur stabilized the mispair. In fact, the **4TFP**:dT and **7TFP**:dT mispairs were virtually as stable as the corresponding mispairs with dA. It is interesting that the dT mispair behaved so differently from the dC mispair, considering the similar size and minor groove functionality of the two pyrimidines (see below).

The effect of heteroatom substitution is also informative in the series **4OFP**, **4OTP**, **7OFP**, and **7OTP**. In this case, each unnatural nucleobase has an amide carbonyl group oriented toward the minor groove, with either an oxygen or a sulfur atom at position 2 or 3. Substitution at position 3 is expected to orient the heteroatom toward the major groove, which is available for solvation by water, while substitution at position 2 results in a different overall dipole and a heteroatom that is expected to be packed within the duplex. Placement of a heteroatom at position 3 was more stabilizing for the unnatural self-pairs than placement of a heteroatom at position 2. Compared to an oxygen atom, a sulfur atom at position 3 significantly stabilized the self-pair, by about 3 °C, while a sulfur atom at position 2 slightly destabilized the self-pair. The stabilities of mispairs between the unnatural bases and the natural bases were also sensitive to

(27) Berger, M.; Ogawa, A. K.; McMinn, D. L.; Wu, Y.; Schultz, P. G.; Romesberg, F. E. *Angew. Chem., Int. Ed.* **2000**, *39*, 2940−2942.

the identity of the heteroatom at these two positions. Surprisingly, specific stabilization of the dT mispair by sulfur substitution, mentioned above for position 1, was also observed at both positions 2 and 3 (**7OTP**:dT and **4OTP**:dT). The remaining mispairs, with dA, dG, and dC, were less sensitive to the identity of the heteroatom at position 3 and were affected differently by the identity of the heteroatom at position 2 (sulfur destabilized the mispairs with dC and dG by 2 °C to 3 °C and stabilized the mispair with dA by 0.9 °C).

It is remarkable that sulfur substitution at each of the three positions examined results in selective stabilization of the mispair between the unnatural base and dT. This is unlikely to be the result of a nonspecific increase in the hydrophobic force, as dT is not the most hydrophobic of the natural bases. The only other strong isotropic force that could be increased by sulfur substitution is polarizability. Thus, it may be that sulfur substitution results in increased polarizability that favors intrastrand and interstrand packing interactions in the unnatural mispair with dT.[18−22] As discussed below, this effect is not apparent in the kinetic data, implying that different forces control stability and replication of DNA.[28]

Replication. Self-pair synthesis efficiency was not strongly sensitive to the specific heteroatom substitution pattern of the unnatural base; the Michaelis constants of the unnatural triphosphates varied by only 6-fold, and the value of k_{cat} varied by only 10-fold. In general, the specificity constants for self-pair synthesis are very different from those reported for other unnatural self-pairs indicating that nucleobase "shape" may be important for efficiency.

Compared to self-pair synthesis, the self-pair extension rates are more sensitive to heteroatom substitution. With the exception of the **4TFP** self-pair, the presence of any of the unnatural self-pairs at the primer terminus appears to interfere with the binding of dCTP. In the case of **4TFP**, dCTP seems to pack on the unnatural self-pair, opposite dG in the template, well enough to give a Michaelis constant typical of natural synthesis. However, as the catalyzed chemistry proceeds with a reduced k_{cat}, it may be inferred that the ternary complex with an unnatural primer-template is structurally or dynamically different from the fully natural complex. In this case, tight binding may actually inhibit replication by stabilizing the complex in a form that requires rearrangement prior to the chemical step. Nonetheless, the **4TFP** self-pair is extended with an efficiency only 200-fold reduced relative to natural synthesis.[12] This indicates that unnatural nucleobases with a conserved shape may be optimized for extension efficiency by variation of the heteroatom substitution pattern.

Heteroatom substitution has its greatest effect on the polymerase-catalyzed synthesis of mispairs between the unnatural and natural bases. In general, the unnatural triphosphates described herein are tightly bound opposite all natural bases in the template. Therefore, the unnatural triphosphates must interact strongly, but nonspecifically, with the polymerase and the natural primer-template. Based on the various crystal structures available for KF and homologous polymerases (type I DNA polymerases from *B. stearothermophilus* and *T. aquaticus*[29,30]), the base of the incoming triphosphate packs between the natural base at the primer terminus and the side chain of a tyrosine in the polymerase (Tyr766 in KF). Packing between the terminal base of the primer and Tyr766 may be particularly favorable for the base of the unnatural triphosphate due to its aromatic surface area and heteroatoms, which may act to increase its polarizability.[13] This tight binding is most pronounced for the **7OTP** triphosphate, which is bound opposite dG with an affinity similar to that of dCTP. Opposite each of the other three natural bases, d(**7OTP**)TP is bound with a Michaelis constant only 4-fold lower than that for the correct natural triphosphate.

As discussed previously with respect to the unnatural self-pair extension, tight binding of the unnatural triphosphates does not produce ternary complexes that lead to efficient catalytic turnover, as evidenced by the small k_{cat} values for insertion opposite the natural bases in the template, except in the case of dG. Opposite dG, the unnatural triphosphates are apparently tightly bound in a geometry that is suitable for efficient catalysis; the observed k_{cat} values are only 3-fold to 10-fold smaller than those typical for correct insertion of a natural dNTP.[12] With the exception of d(**4OFP**)TP, the specificity constants for unnatural triphosphate insertion opposite dG vary by less than 2.5-fold, from 1.5×10^6 min^{-1} M^{-1} to 3.8×10^6 min^{-1} M^{-1}, within 10-fold to 20-fold of the rate for dCTP insertion.

The specific dipole moment and heteroatom substitution pattern of the unnatural base had little effect on its overall efficiency of insertion against dG, with the sole exception of d(**4OFP**)TP, which was not inserted at a rate sufficient for detection ($< 1 \times 10^3$ min^{-1} M^{-1}). The inefficiency of d(**4OFP**)-TP incorporation with dG in the template must not be based on the shape of the nucleobase but rather on its specific heteroatom substitution pattern. Interestingly, of all the unnatural bases evaluated, the heteroatom substitution pattern of **4OFP** is most similar to a natural base, specifically thymidine. Perhaps KF accepts or rejects unnatural triphosphate substrates based not on their engagement of favorable interactions but rather based on their avoidance of unfavorable interactions. This would explain the specific rejection of d(**4OFP**)TP, which may be recognized by the polymerase as dTTP whose insertion opposite dG the polymerase evolved to avoid.

The unnatural mispairs with dG were synthesized in the opposite sense by insertion of dGTP opposite the unnatural bases in the template, with a much broader distribution of rates, ranging from $< 10^3$ min^{-1} M^{-1} for insertion opposite **7TFP** to $> 10^6$ min^{-1} M^{-1} for insertion opposite **7OFP** and **4TFP**. The insertion of dGTP opposite **7OFP** and **4TFP** is efficient for different reasons. Against **7OFP** in the template, dGTP is inserted with a remarkably high k_{cat} (53 min^{-1}, only about 3-fold reduced relative to natural synthesis) and a moderate K_M (35 μM). In contrast, against **4TFP** in the template, dGTP is inserted with a relatively low k_{cat} (1.8 min^{-1}) and the remarkably low K_M of 0.7 μM, which is similar to that for insertion of dGTP opposite dC. This tight binding in the polymerase active site is not correlated with greater thermal stability as **4TFP**:dG was one of the least stable mispairs in duplex DNA.

Conclusion. In duplex DNA, all six unnatural self-pairs are stable and selective, demonstrating the suitability of non-H-bonding forces for the storage of genetic information. Sulfur substitution at each position examined stabilized mispairs

(28) Petruska, J.; Goodman, M. F.; Boosalis, M. S.; Sowers, L. C.; Cheong, C.; Ignacio Tinoco, J. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6252−6256.

(29) Kiefer, J. R.; Mao, C.; Braman, J. C.; Beese, L. S. *Nature* **1998**, *391*, 304−307.

(30) Kim, Y.; Eom, S. H.; Wang, J.; Lee, D.-S.; Suh, S. W.; Steitz, T. A. *Nature* **1995**, *376*, 612−616.

between the unnatural bases and dT. These stabilizing interactions were not apparent in the K_M values determined in the gel-based single nucleotide incorporation assay. In fact, as has been observed with natural DNA,[28] there is no correlation between thermal stability and the efficiency and fidelity of polymerase-catalyzed synthesis, in this case of the unnatural self-pairs. This implies either that the rate-limiting transition state is relatively early or, more likely, that interactions between the bases of the nascent pair at a primer terminus in a polymerase active site are significantly different from those of a base pair in duplex DNA. Heteroatom substitution of the unnatural nucleobases had a small effect on self-pair synthesis efficiency, a larger effect on self-pair extension efficiency, and the largest effect on polymerase-mediated pairing of an unnatural base with a natural base. This information is expected to aid in the future design of unnatural bases that are efficiently and selectively replicated. Moreover, the data imply that dipole and other electrostatic effects play an important role in the fidelity of replication of natural DNA.

4. Experimental Section

General Methods. All reactions were carried out in oven-dried glassware under inert atmosphere, unless otherwise noted. All solvents were dried over 4 Å molecular sieves except dichloromethane, which was freshly distilled from CaH_2. Compound **1A** was purchased from Ubichem PLC (UK). All other reagents were purchased from Sigma−Aldrich. T4 polynucleotide kinase and Klenow fragment exo- were purchased from New England Biolabs. Redivue [γ-^{33}P]-ATP was purchased from Amersham Biosciences. Oligonucleotides were synthesized using an Applied Biosystems Inc. 392 DNA/RNA synthesizer. DNA synthesis reagents were purchased from Glen Research, Stering, VA.

Oligonucleotide duplex denaturation temperature measurements were made in buffer containing 10 mM PIPES, 10 mM $MgCl_2$, and 100 mM NaCl with a final double-stranded oligonucleotide concentration of 3 μM, using a Cary 300 Bio UV−visible spectrophotometer. Measurements were taken over the range 16−80 °C at 0.5 °C/min intervals. Melting temperatures were obtained from the derivative method in the Cary WinUV thermal application software.

Oligonucleotide primers for the kinetic assays were 5′-radiolabeled with T4 polynucleotide kinase and [γ-^{33}P]-ATP. Primer-templates were annealed in the reaction buffer by heating to 95 °C followed by slow cooling to room temperature. Assay conditions included 40 nM primer-template, 0.1−1.3 nM enzyme, 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM DTT, and 50 μg/mL acetylated BSA. The reactions were carried out by combining the DNA−enzyme mixture with an equal volume (5 μL) of 2X dNTP stock solution, incubating at 25 °C for 1−10 min, and quenching by addition of 20 μL of loading dye containing 95% formamide and 20 mM EDTA. The reaction mixtures were resolved by 15% polyacrylamide gel electrophoresis, and the radioactivity was quantified by means of a PhosphorImager (Molecular Dynamics) and the ImageQuant program. A plot of k_{obs} versus [dNTP] was fit to a Michaelis−Menten equation using the program Kaleidagraph (Synergy Software). The data presented are averages of three independent determinations.

General Procedure for Bistoluoyl Nucleoside Synthesis. To a stirred solution of substrate **A** (1 equiv) in acetonitrile (0.25 M) at room temperature under argon was added bis(trimethylsilyl)acetamide (1 equiv). After stirring at ambient temperature for 40 min, 3,5-bis-(toluoyl)-2-deoxyribosyl chloride (1 equiv) was added. The reaction was brought to 0 °C. To the reaction mixture was added dropwise $SnCl_4$ (20 mmol %). After 1 h, complete dissolution of the 2′-deoxyribofuranoside had occurred. To the reaction was added EtOAc, and the resulting solution was successively extracted with saturated $NaHCO_3$ and brine. The organics were dried over anhydrous Na_2SO_4, and solvents were removed *in vacuo*. Purification via flash column chromatography on silica gel (30−50% ethyl acetate in hexanes) afforded the desired product as a mixture of two anomers.

General Toluoyl Deprotection Procedure. To a stirred solution of **B** (1 equiv) in methanol (0.2 M) was added dropwise NaOMe (2.4 equiv, 0.5 M in CH_3OH). After the reaction was complete, it was quenched by addition of saturated aqueous NH_4Cl (excess) and concentrated. Purification via flash column chromatography on silica gel (10−40% methanol in ethyl acetate) afforded free nucleoside.

General Phosphoramidite Synthesis Procedure. To a solution of free nucleoside (1 equiv), azeotroped from a small amount of toluene, in pyridine (0.2 M) was added a solution of DMTrCl (1.4 equiv) in pyridine (0.1 M) via syringe pump. After being stirred for 1 h at room temperature, the reaction mixture was quenched by saturated aqueous NH_4Cl. The reaction mixture was partitioned between ethyl acetate and brine. The organics were dried over anhydrous Na_2SO_4 and solvents were removed *in vacuo*. Purification via flash column chromatography on silica gel (30−70% hexanes in ethyl acetate) afforded the tritylated β-anomer, which was dissolved in CH_2Cl_2 (0.1 M) and cooled to 0 °C. Diisopropylethylamine (1.1 equiv) was added, followed by 2-cyanoethyl diisopropylaminochlorophosphoramidite (1 equiv). After 15 min, the reaction mixture was partitioned between ethyl acetate and brine. The layers were separated, and the aqueous layer was extracted once with ethyl acetate. The combined organic layers were dried over anhydrous Na_2SO_4, filtered, and concentrated. Purification via flash column chromatography on silica gel (15−60% ethyl acetate in 5% triethylamine/hexanes) afforded the desired phosphoramidite as a mixture of two diastereomers.

General Triphosphate Synthesis Procedure. The tritylated β-anomer nucleoside was treated with 20% trichloroacetic acid in dichloromethane (0.2 M) for 30 min. The resulting red solution was quenched by solid sodium bicarbonate (excess) and concentrated *in vacuo*. Purification via flash column chromatography afforded the desired β-free nucleoside. Proton sponge (1.5 equiv) and the free nucleoside (1 equiv) were dissolved in trimethyl phosphate (final concentration 0.3 M) and cooled to 0 °C for 5 min. $POCl_3$ (1.05 equiv) was added dropwise, and the purple 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 (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 solution stood for 40 min at 0 °C and was lyophilized and purified by reverse-phase (C18) HPLC (4−35% CH_3CN in 0.1 M Et_3N-HCO_3, pH 7.5) to afford the triphosphate as a white solid.

Compound 1B. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.90 (2H, d, J = 7.8 Hz), 7.83 (2H, d, J = 7.8 Hz), 7.56 (1H d, J = 5.6 Hz), 7.48 (1H, d, J = 7.4 Hz), 7.18−7.08 (m, 5H), 6.75 (1H, t, J = 6.6 Hz), 6.52 (1H, d, J = 7.4 Hz), 6.05 (1H, d, J = 6.0 Hz), 4.67 (s, 2H), 4.54 (1H, d, J = 2.4 Hz), 2.88 (1H, dd, J = 14.2, 4.8 Hz), 2.40−2.20 (7H, m, singlets at 2.32, 2.29, 3H each). ^{13}C NMR (100 MHz, $CDCl_3$) δ 154.1, 153.6, 146.5, 135.8, 132.2, 118.7, 117.8, 117.5, 117.4, 117.2, 116.9, 114.8, 114.7, 112.9, 112.7, 112.3, 89.4, 76.4, 73.5, 63.1, 52.3, 27.3; HRMS (MALDI−FTMS): calcd for $C_{28}H_{25}NO_6SNa$ (MNa^+), 526.1295; found, 526.1302.

Compound 1C. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.81 (1H, d, J = 7.4 Hz), 7.62 (1H, d, J = 5.6 Hz), 7.45 (1H, d, J = 6.6, 2.1 Hz), 7.36−7.22 (11H, m), 6.90−6.70 (5H, m), 6.53 (1H, d, J = 7.6 Hz), 4.56 (1H, brs), 4.45 (1H, d, J = 5.0 Hz), 4.22 (1H, d, J = 7.6 Hz), 3.76 (6H, s), 3.52−3.48 (2H, m), 2.85−2.71 (1H, m), 2.46−2.45 (1H, m). ^{13}C NMR (100 MHz, $CDCl_3$) δ 158.8, 153.3, 149.1, 148.8, 143.3, 136.0, 133.3, 130.5, 129.5, 129.3, 127., 127.2, 113.5, 107.5, 100.9, 87.5, 85.5, 72.3, 63.7, 55.5, 42.6; HRMS (MALDI−FTMS): calcd for $C_{33}H_{32}NO_6S$ (MH^+), 570.1945; found, 570.1942.

Compound 1D. ^{31}P NMR (140 MHz, 50 mM Tris, 2 mM EDTA, pH 7.5 in D_2O) δ −6.20 (d, J = 17.2 Hz), −10.71 (d, J = 15.8 Hz), −21.82 (t, J = 17.1 Hz).

Compound 1E. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.84 (0.5H, d, J = 7.8 Hz), 7.76 (0.5H, d, J = 7.4 Hz), 7.64 (1H, d, J = 5.0 Hz), 7.44 (2H, brs), 7.41−7.18 (8H, m), 6.85−6.79 (4H, m), 6.72 (1H, d, J = 5.8 Hz), 6.48 (1H, d, J = 7.4 Hz), 4.78−4.60 (1H, m), 4.21 (1H, brs), 3.81−3.35 [12H, m, singlet at 3.78 (6H)], 2.74−2.60 (1H, m), 2.60 (1H, t, J = 6.2 Hz), 2.42 (1H, t, J = 6.2 Hz), 2.40−2.20 (1H, m), 1.40−1.11 (9H, m), 1.02 (3H, d, J = 6.9 Hz). ^{31}P NMR (140 MHz, $CDCl_3$) δ 150.3, 149.6. HRMS (MALDI−FTMS): calcd for $C_{42}H_{48}N_3O_7$-PSNa (MNa^+), 792.2843; found, 792.2840.

Compound 2B. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.98 (2H, d, J = 6.0 Hz), 7.90 (1H, d, J = 6.0 Hz), 7.61 (1H, d, J = 5.7 Hz), 7.48 (1H, d, J = 3.0 Hz), 7.28 (2H, d, J = 6.6 Hz), 7.22 (2H, d, J = 6.0 Hz), 6.97 (1H, dd, J = 3.0, 0.6 Hz), 6.79 (1H, dd, J = 12, 3.9 Hz), 6.46 (1H, d, J = 5.7 Hz), 5.65−5.62 (1H, m), 4.78−4.60 (2H, m), 4.60 (1H, t, J = 2.1 Hz), 2.97−2.91 (1H, m), 2.44 (6H, s), 2.39−2.26 (1H, m). ^{13}C NMR (100 MHz, $CDCl_3$) δ 166.2, 166.1, 152.9, 148.6, 144.4, 144.2, 143.2, 132.8, 129.9, 129.6, 129.3, 129.2, 16.7, 126.4, 126.3, 107.3, 100.8, 85.1, 82.9, 77.4, 77.1, 76.7, 75.1, 64.3, 39.3, 21.7, 21.6. HRMS (MALDI): calcd for $C_{28}H_{25}NO_7Na$ (MNa^+), 510.1523; found, 510.1518.

Compound 2C. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.84 (1H, d, J = 6.0 Hz), 7.38 (2H, d, J = 3.8 Hz), 7.27−7.24 (6H, m), 7.19 (1H, t, J = 5.4 Hz), 6.83 (4H, d, J = 6.0 Hz), 6.26 (1H, d, J = 4.5 Hz), 5.54 (1H, d, J = 6.3 Hz), 4.49 (1H, dd, J = 8.1, 3.6 Hz), 3.97 (1H, dd, J = 5.7, 2.4 Hz), 3.35 (2H, d, J = 4.8 Hz), 2.44−2.38 (1H, m), 2.19−2.12 (1H, m). ^{13}C NMR (100 MHz, $CDCl_3$) δ 160.3, 159.7, 159.0, 145.0, 144.6, 135.7, 135.6, 130.2, 129.9, 128.2, 127.7, 126.8, 115.4, 112.9, 106.5, 96.1, 86.9, 86.6, 85.7, 70.7, 63.0, 54.5, 41.8. HRMS (MALDI): calcd for $C_{33}H_{31}NO_7Na$ (MNa^+), 576.1993; found, 576.1988.

Compound 2D. ^{31}P NMR (140 MHz, 50 mM Tris, 2 mM EDTA, pH 7.5 in D_2O) δ −6.66 (d, J = 17.2 Hz), −10.76 (d, J = 17.1 Hz), −21.98 (t, J = 17.4 Hz).

Compound 2E. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.91 (0.5H, d, J = 7.8 Hz), 7.83 (0.5H, d, J = 7.8 Hz), 7.47−7.28 (9H, m), 6.96 (1H, s), 6.86 (2H, d, J = 2.8 Hz), 6.81 (2H, d, J = 3.0 Hz), 6.70 (1H, dd, J = 25.2, 6.2 Hz), 6.30 (1H, d, J = 7.8 Hz), 4.68−4.59 (1H, m), 4.21 (1H, brs), 3.79 (6H, s), 3.78−3.35 (7H, m), 2.80−2.62 (1H, m), 2.61 (1H, t, J = 6.4 Hz), 2.43 (1H, t, J = 6.2 Hz), 2.30−2.18 (1H, m), 1.30−1.21 (9H, m), 1.19 (3H, d, J = 6.9 Hz). ^{31}P NMR (140 MHz, $CDCl_3$) δ 150.2, 149.6. HRMS (MALDI): calcd for $C_{42}H_{48}N_3O_8PNa$ (MNa^+), 776.3071; found, 776.3058.

Compound 3B. To a stirred solution of **2B** in toluene (0.1 M) was added Lawesson's reagent (1.5 equiv) at room temperature. The resulting solution was refluxed until the reaction was complete. After concentration *in vacuo*, the residue was purified via flash column chromatography on silica gel (20−40% ethyl acetate in hexanes) to afford **3B** as a yellow foam. HRMS (MALDI−FTMS): calcd for $C_{28}H_{25}NO_6SNa$ (MNa^+), 526.1300; found, 526.1272.

Compound 3C. ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.42 (d, J = 7.4 Hz, 1H), 7.44 (d, J = 2.0 Hz, 1H), 7.37−7.04 (11H, m), 6.75 (d, J = 9.0 Hz), 6.49 (d, J = 7.4 Hz, 1H), 4.53 (d, J = 5.8 Hz, 1H), 4.02 (d, J = 7.2 Hz, 1H), 3.69 (s, 6H), 3.60−3.31 (m, 2H), 2.87 (brs, 1H), 2.8−2.62 (m, 1H), 2.30−2.21 (m, 1H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 174.2, 158.8, 154.5, 145.6, 144.7, 135.7, 135.6, 133.1, 130.4, 130.3, 128.4, 128.3, 127.4, 113.5, 110.3, 101.1, 90.7, 87.2, 86.5, 78.0, 77.4, 76.7, 70.1, 62.3, 55.6, 42.2. HRMS (MALDI−FTMS): calcd for $C_{33}H_{31}$-NO_6SNa (MNa^+), 592.1764; found, 592.1733.

Compound 3D. ^{31}P NMR (140 MHz, 50 mM Tris, 2 mM EDTA, pH 7.5 in D_2O) δ −9.25 (d, J = 17.1 Hz), −10.79 (d, J = 17.0 Hz), −22.56 (t, J = 17.2 Hz).

Compound 3E. ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.50 (dd, J = 18.0, 7.4 Hz, 1H), 7.48 (d, J = 1.8 Hz, 1H), 7.40−7.05 (m, 12H), 6.79 (d, J = 3.6 Hz, 2H), 6.75 (d, J = 3.5 Hz, 2H), 6.44 (d, J = 5.8 Hz, 1H), 4.80−4.59 (m, 1H), 4.13 (brs, 1H), 3.73 (s, 6H), 3.60−3.31 (m, 6H), 2.90−2.81 (1H, m), 2.63 (1H, t, J = 6.2 Hz), 2.48−2.46 (2H, m), 1.21−1.15 (12H, m). ^{31}P NMR (140 MHz, $CDCl_3$) δ 150.6, 149.7. HRMS (MALDI−FTMS): calcd for $C_{42}H_{48}N_3O_7PSNa$ (MNa^+), 792.2848; found, 792.2832.

Compound 4B. ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.00−7.89 (3H, m), 7.77 (0.5H, d, J = 1.8 Hz), 7.74 (0.5H, d, J = 2.0 Hz), 7.63 (0.5H, d, J = 6.2 Hz), 7.52 (1H, t, J = 7.2 Hz), 7.30−7.21 (4H, m), 7.11 (1H, d, J = 7.8 Hz), 6.89 (0.5H, dd, J = 5.6, 8.2 Hz), 6.72−6.62 (1H, m), 6.53 (0.5H, d, J = 7.0 Hz), 6.40 (0.5H, d, J = 7.4 Hz), 5.64 (1H, t, J = 6.2 Hz), 4.95 (0.5H, brs), 4.73 (1H, brs), 4.60 (1H, d, J = 2.0 Hz), 4.59−4.57 (0.5H, m), 3.18−2.88 (1H, m), 2.62−2.40 (1H, m), 2.26 (3H, s), 2.17 (1.5H, s), 2.04 (1.5H, s). HRMS (MALDI−FTMS): calcd for $C_{28}H_{25}NO_7Na$ (MNa^+), 510.1523; found, 510.1509.

Compound 4C. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.89 (1H, d, J = 2.0 Hz), 7.80 (1H, d, J = 7.4 Hz), 7.43 (2H, d, J = 6.6 Hz), 7.65−7.28 (7H, m), 6.80 (4H, d, J = 8.6 Hz), 6.72 (1H, d, J = 2.0 Hz), 6.65 (1H, t, J = 6.4 Hz), 6.37 (1H, d, J = 7.4 Hz), 4.52 (1H, d, J = 5.8 Hz), 4.06 (1H, d, J = 3.6 Hz), 3.71 (6H, s), 3.48−3.42 (3H, m), 2.48−2.41 (1H, m), 2.28−2.17 (1H, m). ^{13}C NMR (100 MHz, $CDCl_3$) δ 158.8, 153.3, 149.0, 148.8, 144.8, 143.3, 136.0, 135.9, 133.3, 130.3, 128.4, 128.2, 127.8, 127.7, 127.1, 113.4, 107.4, 107.3, 100.8, 86.9, 86.7, 85.4, 85.3, 72.3, 63.7, 55.5, 55.4, 42.7. HRMS (MALDI−FTMS): calcd for $C_{33}H_{31}NO_7Na$ (MNa^+), 576.1993; found, 576.1973.

Compound 4D. ^{31}P NMR (140 MHz, 50 mM Tris, 2 mM EDTA, pH 7.5 in D_2O) δ −6.03 (d, J = 17.1 Hz), −10.73 (d, J = 17.2 Hz), −21.85 (t, J = 17.2 Hz).

Compound 4E. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.79 (0.5H, d, J = 7.4 Hz), 7.74 (0.5H, d, J = 7.4 Hz), 7.69 (1H, s), 7.46 (2H, d, J = 7.0 Hz), 7.40−7.18 (6H, m), 6.87−6.70 (5H, m), 6.59 (1H, s), 6.26 (1H, d, J = 7.4 Hz), 4.69−4.60 (1H, m), 4.20 (1H, s), 3.77 (6H, s), 3.80−3.37 (7H, m), 2.80−2.59 (1H, m), 2.63 (1H, t, J = 6.7 Hz), 2.42 (1H, t, J = 6.7 Hz), 2.40−2.21 (1H, m), 1.30−1.11 (9H, m), 1.10 (3H, d, J = 6.9 Hz). ^{31}P NMR (140 MHz, $CDCl_3$) δ 150.1, 149.6. HRMS (MALDI−FTMS): calcd for $C_{42}H_{48}N_3O_8PNa$ (MNa^+), 776.3071; found, 776.3065.

Compound 5B. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.96 (d, J = 8.0 Hz, 2H), 7.90 (d, J = 7.9 Hz, 2H), 7.67 (d, J = 5.2 Hz, 1H), 7.55 (d, J = 7.6 Hz, 1H), 7.26−7.10 (m, 5H), 6.85 (dd, J = 8.0, 5.6 Hz, 1H), 6.55 (d, J = 7.4 Hz, 1H), 5.64 (d, J = 6.2 Hz, 1H), 4.72 (brs, 2H), 4.61 (s, 1H), 2.91 (dd, J = 14.0, 5.4 Hz, 1H), 2.37 (s, 3H), 2.33 (s, 3H), 2.40−2.36 (m, 1H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 166.4, 165.9, 158.4, 145.6, 144.5, 144.4, 133.9, 130.3, 129.9, 129.6, 129.3, 127.8, 126.9, 126.5, 124.3, 102.8, 88.5, 85.6, 78.0, 77.4, 76.8, 75.3, 64.6, 39.6, 21.9. HRMS (MALDI−FTMS): calcd for $C_{28}H_{25}NO_6SNa$ (MNa^+), 526.1300; found, 526.1283.

Compound 5C. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.86 (d, J = 7.8 Hz, 1H), 7.65 (d, J = 3.0 Hz, 1H), 7.50−7.11 (m, 11H), 6.83 (d, J = 8.6 Hz, 4H), 6.50 (d, J = 7.6 Hz, 1H), 4.61 (brs, 1H), 4.38−4.20 (m, 1H), 3.77 (s, 6H), 3.54−3.41 (m, 2H), 2.78−2.72 (m, 1H), 2.36−2.26 (m, 1H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 158.7, 158.5, 145.6, 144.8, 135.9, 135.8, 134.0, 130.4, 129.8, 128.4, 128.2, 127.2, 124.4, 113.4, 103.7, 86.9, 86.8, 85.7, 72.2, 63.7, 55.5, 42.7. HRMS (MALDI−FTMS): calcd for $C_{33}H_{31}NO_6SNa$ (MNa^+), 592.1764; found, 592.1744.

Compound 5D. ^{31}P NMR (140 MHz, 50 mM Tris, 2 mM EDTA, pH 7.5 in D_2O) δ −7.64 (d, J = 17.2 Hz), −10.60 (d, J = 17.2 Hz), −22.14 (t, J = 17.1 Hz).

Compound 5E. ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.73 (dd, J = 14.0, 7.0 Hz, 1H), 7.38 (d, J = 5.2 Hz, 1H), 7.42−7.05 (m, 10H), 6.76 (d, J = 8.6 Hz, 4H), 6.72−6.60 (m, 1H), 6.37 (d, J = 7.4 Hz, 1H), 4.62−4.50 (m, 1H), 4.14 (brs, 1H), 3.70 (s, 6H), 3.65−3.23 (m, 6H), 2.70−2.53 (m, 2H), 2.38 (t, J = 6.7 Hz, 1H), 2.30−2.12 (m, 1H), 1.25−1.15 (m, 9H), 1.10 (d, J = 7.0 Hz, 3H). ^{31}P NMR (140 MHz, $CDCl_3$) δ 150.15, 149.48. HRMS (MALDI−FTMS): calcd for $C_{42}H_{48}N_3O_7PSNa$ (MNa^+), 792.2848; found, 792.2827.

Compound 6B. To a stirred solution of **4B** in toluene (0.1 M) was added Lawesson's reagent (1.2 equiv) at room temperature. The resulting solution was refluxed until the reaction was complete. After concentration *in vacuo*, the residue was purified via flash column chromatography on silica gel (10−30% ethyl acetate in hexanes) to afford **6B** as a yellow foam. ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.18−7.88 (5H, m), 7.67 (1H, d, J = 8.4 Hz), 7.60 (1H, dd, J = 7.7, 0.9 Hz), 7.40−7.28 (3H, m), 7.19 (1H, d, J = 8.1 Hz), 7.02 (1H, d, J = 7.4 Hz), 6.90−6.80 (1H, m), 5.74−5.68 (1H, m), 5.17 (0.5H, t, J = 4.3 Hz), 4.92−4.82 (0.5H, m), 4.80−4.62 (2H, m), 3.45−3.28 (1H, m), 2.50 (3H, s), 2.49 (1.5H, s), 2.43 (1.5H, s), 2.50−2.31 (1H, m). HRMS (MALDI−FTMS): calcd for $C_{28}H_{25}NO_6SNa$ (MNa^+), 526.1300; found, 526.1291.

Compound 6C. ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.47 (1H, dd, J = 11.0, 5.4 Hz), 7.96 (1H, d, J = 9.9 Hz), 7.52 (2H, dd, J = 10.7, 6.0 Hz), 7.44−7.31 (8H, m), 6.95−6.84 (4H, m), 6.78 (1H, d, J = 10.2 Hz), 6.70 (1H, dd, J = 11.1, 5.4 Hz), 4.66 (1H, t, J = 5.1 Hz), 4.18 (1H, d, J = 6.6 Hz), 3.89 (6H, s), 3.73 (2H, t, J = 8.7 Hz), 3.57 (1H, t, J = 8.1 Hz), 2.92−2.87 (1H, m), 2.42 (1H, t, J = 4.5 Hz). ^{13}C NMR (100 MHz, $CDCl_3$) δ 164.2, 158.8, 154.4, 140.9, 144.7, 135.9, 135.8, 130.9, 130.3, 129.3, 128.4, 128.3, 128.2 128.1, 127.9, 127.2, 113.5, 113.4, 107.4, 105.5, 93.2, 89.6, 86.8, 73.1, 64.3, 55.5, 42.9, 29.4. HRMS (MALDI−FTMS): calcd for $C_{33}H_{31}NO_6SNa$ (MNa^+), 592.1764; found, 592.1764.

Compound 6D. ^{31}P NMR (140 MHz, 50 mM Tris, 2 mM EDTA, pH 7.5 in D_2O) δ −7.54 (d, J = 18.5 Hz), −10.59 (d, J = 17.2 Hz), −22.04 (t, J = 17.2 Hz).

Compound 6E. ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.50 (1H, d, J = 7.8 Hz), 7.90 (1H, s), 7.47 (1H, d, J = 7.9 Hz), 7.39−7.30 (9H, m), 6.84−6.81 (4H, m), 6.69 (1H, s), 6.54 (1H, d, J = 7.0 Hz), 4.78−4.72 (1H, m), 4.22 (1H, d, J = 4.1 Hz), 3.81 (6H, s), 3.67 (1H, dd, J = 10.8, 2.6 Hz), 3.60−3.52 (4H, m), 3.47 (1H, dd, J = 10.5, 2.6 Hz), 2.90−2.85 (1H, m), 2.63 (2H, t, J = 6.2 Hz), 2.48−2.46 (1H, m), 1.21−1.15 (12H, m). ^{31}P NMR (140 MHz, $CDCl_3$) δ 149.9, 148.8. HRMS (MALDI−FTMS): calcd for $C_{42}H_{48}N_3O_7PSNa$ (MNa^+), 792.2848; found, 792.2836.

JA035398O