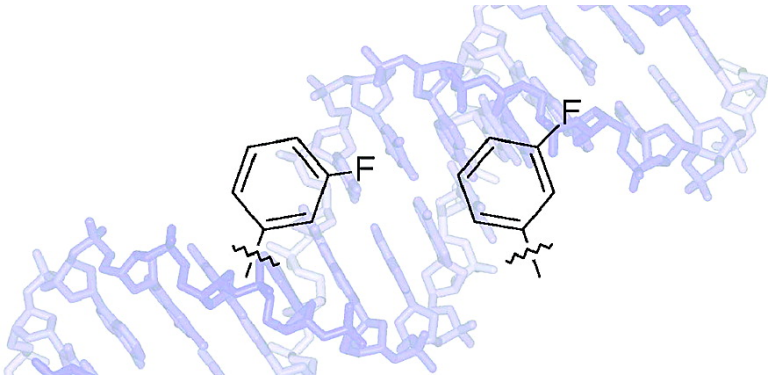
Article

Efforts To Expand the Genetic Alphabet: Identification of a Replicable Unnatural DNA Self-Pair

Allison A. Henry, Anne Goldbech Olsen, Shigeo Matsuda, Chengzhi Yu, Bernhard H. Geierstanger, and Floyd E. Romesberg

J. Am. Chem. Soc., **2004**, 126 (22), 6923-6931• DOI: 10.1021/ja049961u • Publication Date (Web): 18 May 2004



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 10 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 05/18/2004

Efforts To Expand the Genetic Alphabet: Identification of a Replicable Unnatural DNA Self-Pair

Allison A. Henry,† Anne Goldbech Olsen,† Shigeo Matsuda,† Chengzhi Yu,† Bernhard H. Geierstanger,‡ and Floyd E. Romesberg*,†

Contribution from the Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California, 92037, and Genomics Institute of the Novartis Research Foundation, 10675 John-Jay-Hopkins Drive, San Diego, California 92121-1125

Received January 3, 2004; E-mail: floyd@scripps.edu

Abstract: Six unnatural nucleotides featuring fluorine-substituted phenyl nucleobase analogues have been synthesized, incorporated into DNA, and characterized in terms of the structure and replication properties of the self-pairs they form. Each unnatural self-pair is accommodated in B-form DNA without detectable structural perturbation, and all are thermally stable and selective to roughly the same degree. Furthermore, the efficiency of polymerase-mediated mispair synthesis is similar for each unnatural nucleotide in the template. In contrast, the efficiency of polymerase-mediated self-pair extension is highly dependent on the specific fluorine substitution pattern. The most promising unnatural base pair candidate of this series is the 3-fluorobenzene self-pair, which is replicated with reasonable efficiency and selectivity.

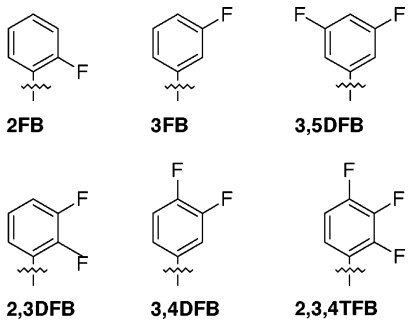
1. Introduction

Genetic information is encoded by a four-letter alphabet. Faithful replication of this information is ensured by the selective interactions between natural nucleobases: adenine with thymine and guanine with cytosine. The identification of unnatural base pairs that are enzymatically replicated and that have thermal stability comparable to the natural pairs would allow for the introduction of additional information, biological or abiological, within a cell. Ongoing efforts to design unnatural base pairs that interact via nonnatural hydrogen-bonding (H-bonding) patterns have yielded some success,[1−7] but have generally proven difficult, possibly due in part to facile tautomerization which converts the unnatural nucleobase to a form that readily pairs with a natural nucleobase.[8−14] In addition, this approach limits the number of available unnatural nucleobases to fewer than can be constructed by synthetic chemists. Furthermore, it is unclear at present that H-bonding is the only intermolecular force suitable for the storage of genetic information. In fact, it has been known for nearly a decade that unnatural nucleobases without H-bonding functionality can be substrates for DNA polymerases.[15−19] Among the alternative intermolecular forces that could mediate interbase interactions to a degree sufficient for information storage is the hydrophobic force, whose contribution to protein structure and folding has been appreciated for decades.[20] Thus, we have explored the utility of hydrophobic and van der Waals forces in developing additional base pairs for the storage and replication of increased genetic information.[21−27]

† The Scripps Research Institute.
‡ Genomics Institute of the Novartis Research Foundation.

(1) Switzer, C.; Moroney, S. E.; Benner, S. A. *J. Am. Chem. Soc.* **1989**, *111*, 8322−8323.
(2) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33−37.
(3) Switzer, C. Y.; Moroney, S. E.; Benner, S. A. *Biochemistry* **1993**, *32*, 10489−10496.
(4) Piccirilli, J. A.; Moroney, S. E.; Benner, S. A. *Biochemistry* **1991**, *30*, 10350−10356.
(5) Moser, M. J.; Prudent, J. R. *Nucleic Acids Res.* **2003**, *31*, 5048−5053.
(6) Moser, M. J.; Marshall, D. J.; Greiner, J. K.; Kieffer, C. D.; Killeen, A. A.; Ptacin, J. L.; Richmond, C. S.; Roesch, E. B.; Scherrer, C. W.; Sherrill, C. B.; Van Hout, C. V.; Zanton, S. J.; Prudent, J. R. *Clin. Chem.* **2003**, *49*, 407−414.
(7) Tor, Y.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 4461−4467.
(8) Bain, J. D.; Switzer, C.; Chamberlin, A. R.; Benner, S. A. *Nature* **1992**, *356*, 537−539.
(9) Horlacher, J.; Hottiger, M.; Podust, V. N.; Hübscher, U.; Benner, S. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6329−6333.
(10) Lutz, M. J.; Held, H. A.; Hottiger, M.; Hübscher, U.; Benner, S. A. *Nucleic Acids Res.* **1996**, *24*, 1308−1313.
(11) Lutz, M. J.; Horlacher, J.; Benner, S. A. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1149−1152.
(12) Lutz, S.; Burgstaller, P.; Benner, S. A. *Nucleic Acids Res.* **1999**, *27*, 2972−2798.
(13) Lutz, M. J.; Horlacher, J.; Benner, S. A. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 499−504.
(14) Roberts, C.; Bandaru, R.; Switzer, C. *Tetrahedron Lett.* **1995**, *36*, 3601−3604.
(15) Moran, S.; Ren, R. X.-F.; Rumney, S. I.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056−2057.
(16) Moran, S.; Ren, R. X.-F.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10506−10511.
(17) Morales, J. C.; Kool, E. T. *Nat. Struct. Biol.* **1998**, *5*, 950−954.
(18) Kool, E. T. *Biopolymers* **1998**, *48*, 3−17.
(19) Kool, E. T. *Curr. Opin. Chem. Biol.* **2000**, *4*, 602−608.
(20) Rose, G. D.; Wolfenden, R. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 381−415.
(21) McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **1999**, *121*, 11585−11586.
(22) Ogawa, A. K.; Wu, Y.; McMinn, D. L.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 3274−3287.
(23) Wu, Y.; Ogawa, A. K.; Berger, M.; McMinn, D. L.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 7621−7632.
(24) Ogawa, A. K.; Wu, Y.; Berger, M.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 8803−8804.
(25) Tae, E. L.; Wu, Y. Q.; Xia, G.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2001**, *123*, 7439−7440.
(26) Berger, M.; Luzzi, S. D.; Henry, A. A.; Romesberg, F. E. *J. Am. Chem. Soc.* **2002**, *124*, 1222−1226.
(27) Yu, C.; Henry, A. A.; Romesberg, F. E.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2002**, *41*, 3841−3844.

Chart 1



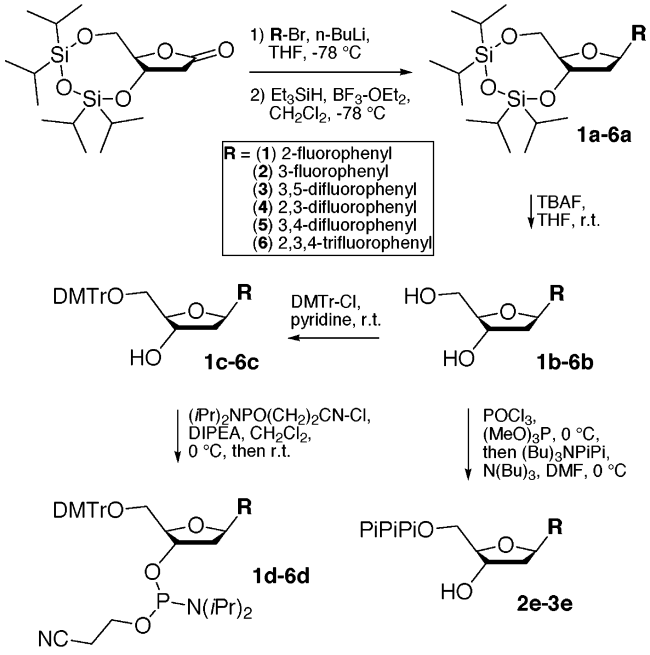
In the course of characterizing these non-H-bonding unnatural nucleobases, we identified several promising pairs formed between two of the same analogues. The use of such "self-pairs" is not a limitation since the addition of a single self-pair to the genetic alphabet would create 61 new codons. Moreover, a self-pair is advantageous because the potential for mispairing with the natural nucleobases is significantly reduced. Unfortunately, though several self-pairs have been identified that are thermally selective in duplex DNA and selectively synthesized by the Klenow fragment of *Escherichia coli* DNA polymerase I (Kf), the nascent self-pairs have been in general poor substrates for Kf, and synthesis continues very inefficiently. Apparently, the physiochemical properties of the nucleobase analogues necessary for stability and enzymatic base pair synthesis are not sufficient for extension.

The unnatural nucleobase properties that may be important for polymerase recognition and efficient replication include aromatic surface area and heteroatom substitution. For example, the large surface area of many previously reported unnatural nucleobases may cause the nascent base pair to adopt a thermally stable non-Watson−Crick-like structure, favoring synthesis but also leading to a primer terminus with a geometry that is not recognized by the polymerase. Heteroatom substitution may be critical for continued synthesis by affecting the dipole moment and polarizability of the nucleobase. Recently, we found that heteroatom-substituted nucleobases of intermediate size may form thermally stable pairs in duplex DNA, despite their reduced aromatic surface area, and that these unnatural base pairs may be synthesized and extended by DNA polymerases with improved efficiency.[28,29] Here we further explore this approach using a simpler scaffold and report the synthesis and characterization of six novel fluorine-substituted benzene nucleobase analogues (Chart 1). The aromatic carbon−fluorine bonds are expected to impart a significant dipole moment to each nucleobase without introducing H-bonding functionality that would be complementary to the natural bases. One of the self-pairs, formed between two 3-fluorobenzene nucleobase analogues, is found to be thermally selective and efficiently synthesized and extended by Kf, making it the most promising unnatural base pair identified to date.

2. Results

2.1. Nucleobase Analogue Synthesis. Tetraisopropyl disiloxane-protected nucleosides were synthesized by aryllithium

Scheme 1



coupling to the disiloxane-protected lactone (Scheme 1), based on a procedure in the literature for simple aromates not containing fluorine atoms.[30,31] In the case of the *o*-fluorobromobenzene derivatives, it was essential to add the haloaromate to the *n*-butyllithium, as the conventional reverse manner gave very low yields. Treatment with triethylsilane and strong Lewis acid resulted in reduction to give the protected aryl *C*-nucleoside in moderate yield; the kinetic preference for hydride attack on the planar C1′ carbocation from below the ring[30,31] was manifested in the β-anomer being the major product of the reaction. The minor α-anomer was readily removed by flash column chromatography. Deprotection was achieved by treatment with TBAF. In all cases, the 1-D ^{1}H NMR splitting pattern for H1′ conformed to literature precedents in which β-stereochemistry is indicated by a doublet of doublets for *C*-nucleosides.[30] In each case, the conversion of free nucleoside to phosphoramidite or triphosphate was accomplished according to standard literature procedures.[32,33]

2.2. Preliminary Structural Characterization of DNA Containing Unnatural Self-Pairs. To determine whether the presence of any unnatural self-pair in a short DNA duplex perturbed its overall structure, each unnatural nucleotide was incorporated, at position **X**, into the complementary oligonucleotides 5′-d(GCGATG**X**GTAGCG) and 5′-d(CGCTAC**X**CATCGC). Circular dichroism (CD) spectra were collected for each unnatural duplex, as well as for the duplexes containing the correct pairs dA:dT or dG:dC or the mispair dT:dT. The spectra of the unnatural duplexes are virtually identical and are within the extrema set by the fully natural duplexes, indicating that the unnatural base pairs do not significantly distort the B-form duplex (Figure 1).

More detailed structural characterization of a short DNA duplex containing the **3FB** self-pair has also been initiated. This particular base pair was chosen because of its favorable

(28) Matsuda, S.; Henry, A. A.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2003**, *125*, 6134−6139.
(29) Henry, A. A.; Yu, C.; Romesberg, F. E. *J. Am. Chem. Soc.* **2003**, *125*, 9638−9646.
(30) Wichai, U.; Woski, S. A. *Org. Lett.* **1999**, *1*, 1173−1175.
(31) Wichai, U.; Woski, S. A. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 3465−3468.
(32) Schweitzer, B. A.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 1863−1872.
(33) Kovacs, T.; Otvos, L. *Tetrahedron Lett.* **1988**, *29*, 4525−4528.

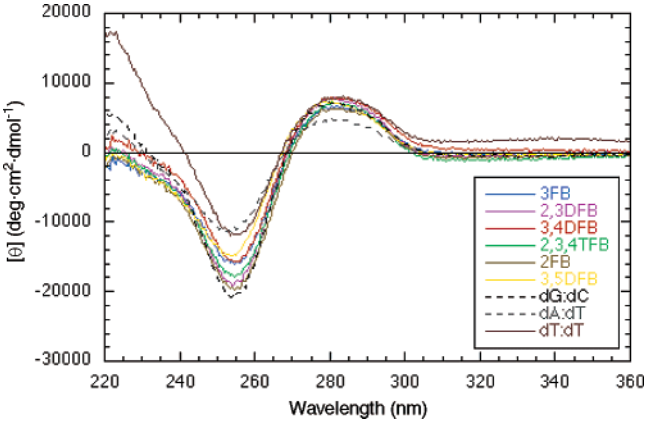


Figure 1. CD spectra of duplex DNA [5′-d(GCGATG**X**GTAGCG):3′-d(CGCTAC**X**CATCGC)] containing each unnatural self-pair, each natural correct pair, or a dT:dT mismatch at the position labeled **X**. See Experimental Section for details.

replication properties (see below). The oligonucleotides 5′-d($G_1C_2A_3G_4A_5$**3FB**$_6C_7G_8A_9T_{10}C_{11}C_{12}$) and 5′-d($G_{13}G_{14}A_{15}$-$T_{16}C_{17}G_{18}$**3FB**$_{19}T_{20}C_{21}T_{22}G_{23}C_{24}$) were synthesized, and the resulting duplex was characterized by NMR spectroscopy. This sequence context was chosen for its similarity to that used in the kinetic assays. All proton resonances associated with the base moieties, as well as H_1' and H_2'/H_2'' of the deoxyribose rings, were assigned by 2D NOESY and DQF-COSY spectra recorded in D_2O and H_2O at several temperatures and buffer conditions. Assignments of the resonances of the two **3FB** bases were further confirmed by $^1H-^{13}C$ HSQC spectra.

In general, all observations are consistent with an overall standard B-form DNA structure.[34,35] This conclusion is supported by characteristic aromatic to H_1' (Figure 2), aromatic to H_2'/H_2'', and aromatic to dT methyl NOE connectivities, as well as the intensities of the H_1' to H_2'/H_2'' NOE cross-peaks (data not shown). Imino to imino NOE connectivities and imino to adenine H_2 connectivites are observed along the center of the DNA helix from base pair dC_2-dG_{23} to dA_5-dT_{20}, and from dG_8-dC_{17} to dC_{11}-dG_{14} (data not shown). The amino protons of all dC residues were assigned, and NOE cross-peaks to the imino protons (dG residues) confirm Watson−Crick base pairing along the entire length of the DNA duplex, including the dA_5-dT_{20} base pair adjacent to the d**3FB**$_6$-d**3FB**$_{19}$ base pair. The imino resonances of the other adjacent base pair, dC_7-dG_{18}, are broad presumably because of enhanced solvent exchange. This may be the result of a base pair stacking defect, or it may simply be due to the two relatively small **3FB** nucleobases incompletely occupying the space that is normally occupied by a purine−pyrimidine base pair. The **3FB** nucleobases do not interrupt the aromatic to H_1' (Figure 2) and aromatic to H_2'/H_2'' (data not shown) walk along the B-DNA helix. These observations and the NOE cross-peaks observed between the dT_{20} CH_3 protons and the d**3FB**$_{19}$ H_6 or H_2 proton (data not shown; H_6 and H_2 resonances are degenerate in chemical shift), between dA_5 H_2 and d**3FB**$_6$ H_6 and H_2 (Figure 2, boxes), and between dA_5 H_2 and d**3FB**$_6$ H_5 (Figure 2, boxes), suggest that both fluorobenzene bases are positioned within the DNA helix. However, NOE cross-peaks from d**3FB**$_6$ H_2 and H_6 to d**3FB**$_6$ H_2'/H_2'' and to

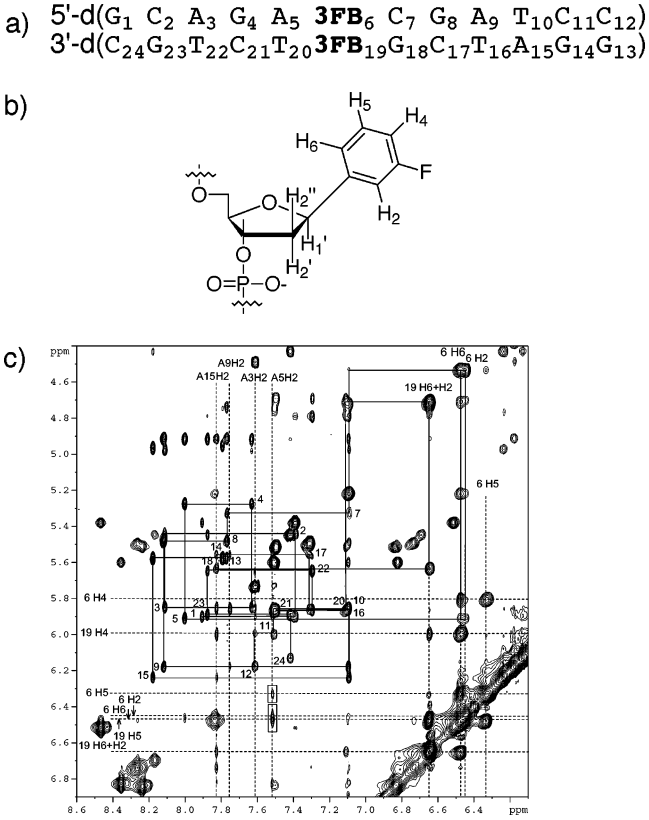


Figure 2. NMR characterization of duplex DNA containing the **3FB** self-pair. (a) DNA duplex sequence with residue numbering, (b) unnatural nucleotide proton numbering guide, and (c) aromatic proton to H_1' region of a 2D NOESY of 0.6 mM duplex in 8.5 mM sodium phosphate, pH 7.0, 134 mM NaCl, 0.25 mM EDTA, and 90%/10% H_2O/D_2O recorded at 295 K with a mixing time of 200 ms. Boxes represent NOE cross-peaks between H_2 of dA_5 and H_6, H_2 and H_5 of d**3FB**$_6$. Intranucleotide connectivities are labeled by the residue number. d**3FB** and dA H_2 resonances are shown as broken lines. See Experimental Section for details.

dA_5 H_2'/H_2'' suggest that the **3FB**$_6$ nucleobase base exists within the duplex in two orientations that are related by rotation about the glycosidic bond.

2.3. Stability of Unnatural Self-Pairs. To evaluate the thermodynamic stability of DNA duplexes containing an unnatural self-pair or mispair, each unnatural and natural nucleotide was incorporated into the complementary oligonucleotides 5′-d(GCGATG**X**GTAGCG) and 5′-d(CGCTAC**Y**CATCGC) at positions **X** and **Y**. The melting temperature of each duplex was determined by thermal denaturation experiments. As used in this text, the term "stability" is intended to refer only to the duplex melting temperature (T_m). Each self-pair-containing duplex is about as stable as a pyrimidine−pyrimidine mispair (Table 1). There is little dependence of the T_m value on the number of fluorine substituents or on the substitution pattern: all self-pair-containing duplexes melt between 52 and 54 °C. For reference, the stability of this duplex containing dA:dT is 59.2 °C, and the stability of this duplex containing dT:dT is 53.3 °C (Table 1, footnote). The stabilities of all possible mispairs are less than any self-pair, resulting in a range of thermal selectivities from 1.9 to 6.7 °C for the **3FB** self-pair and from 4.1 to 10.0 °C for the **3,5DFB** self-pair (Table 1). Mispairs with dA are consistently the most stable, followed closely by those with dT (except **3,4DFB** for which the dT mispair is slightly more stable). Mispairs involving dG and dC

(34) Hare, D. R.; Wemmer, D. E.; Chou, S.-H.; Drobny, G.; Reid, B. R. *J. Mol. Biol.* **1983**, *171*, 319−3368
(35) Wuethrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, 1986.

Table 1. T_m Values for Duplexes Containing Unnatural Base Pairs[a]

5′-d(GCGTAC**X**CATGCG)					
3′-d(CGCATG**Y**GTACGC)					
X	**Y**	T_m (°C)	**X**	**Y**	T_m (°C)
2FB	**2FB**	54.0	**2,3DFB**	**2,3DFB**	53.0
	A	52.0		A	49.1
	C	46.0		C	45.2
	G	46.0		G	45.1
	T	49.0		T	48.3
3FB	**3FB**	52.0	**3,4DFB**	**3,4DFB**	53.0
	A	50.2		A	50.1
	C	45.4		C	47.2
	G	45.3		G	45.3
	T	48.4		T	51.1
3,5DFB	**3,5DFB**	53.2	**2,3,4TFB**	**2,3,4TFB**	53.0
	A	49.0		A	49.1
	C	43.2		C	46.2
	G	44.3		G	45.2
	T	47.1		T	48.3

[a] Uncertainty in values is less than 0.1 °C. The duplex containing a dA:dT correct pair has a T_m value of 59.2 °C, and the one containing a dT:dT mismatch has a T_m value of 53.3 °C. See Experimental Section for details.

Table 2. Rates of Correct Extension of Unnatural Self-Pairs and **3FB** Mispairs[a]

5′-d(TAATACGACTCACTATAGGGAGA**X**)				
3′-d(ATTATGCTGAGTGATATCCCTCT**Y**GCTAGGTTACGGCAGGATCGC)				
X	**Y**	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}\,M^{-1}$)
3FB	**3FB**	28 ± 4	85 ± 24	3.3×10^5
3,5DFB	**3,5DFB**	25 ± 6	125 ± 29	2.0×10^5
2,3,4TFB	**2,3,4TFB**	3.5 ± 0.7	289 ± 59	1.2×10^4
2,3DFB	**2,3DFB**	4.1 ± 1.5	127 ± 31	3.2×10^4
2FB	**2FB**	5 ± 1	254 ± 55	2.0×10^4
3,4DFB	**3,4DFB**	0.5 ± 0.3	242 ± 75	2.1×10^3
A	**3FB**	17 ± 4	268 ± 60	6.3×10^4
C	**3FB**	4.2 ± 0.7	310 ± 80	1.4×10^4
G	**3FB**	nd[b]	nd[b]	$<1.0 \times 10^3$
T	**3FB**	25 ± 4	132 ± 60	1.9×10^5
3FB	A	2.5 ± 0.4	237 ± 48	1.1×10^4
3FB	C	8 ± 1	219 ± 54	3.7×10^4
3FB	G	nd[b]	nd[b]	$<1.0 \times 10^3$
3FB	T	42 ± 1	40 ± 14	1.1×10^6

[a] See Experimental Section for details. [b] Reaction was too inefficient for k_{cat} and K_M to be determined independently.

result in duplexes of similar stability that have significantly reduced T_m values relative to duplexes containing mispairs involving dA and dT.

2.4. Polymerase-Mediated Replication of the Unnatural Self-Pairs. As mentioned above, continued synthesis after incorporation of the unnatural base pair has consistently limited the replication of unnatural DNA. Here we refer to this step as self-pair extension. Thus, to determine the effect of fluorine substitution on self-pair extension step, each unnatural nucleotide was first incorporated at positions **X** and **Y**, respectively, into the primer oligonucleotide 5′-d(TAATACGACTCACTATAGGGAGA**X**) and the template oligonucleotide 5′-d(CGCTAGGACGGCATTGGATCG**Y**TCTCCCTATAGTGAGTCGTATTA). Each primer-template assembly was used as a substrate in a single nucleotide (dCTP) incorporation assay with Kf polymerase. Unlike base pair stability, fluorine substitution had a pronounced effect on the self-pair extension rate, which varies by more than 2 orders of magnitude (Table 2). The **3,4DFB** self-pair is extended with the lowest efficiency, $2.1 \times 10^3\ M^{-1}\ min^{-1}$, due to both a low k_{cat} and a high K_M. The three nucleobases having a fluorine substitutent ortho to the glycosidic bond (**2FB**, **2,3DFB**, and **2,3,4TFB**) form self-pairs that are extended with intermediate efficiency, approximately $10^4\ M^{-1}\ min^{-1}$, due largely to an increase in k_{cat}, relative to the **3,4DFB** self-pair. The nucleobases having only meta-fluorine substituents (**3FB** and **3,5FB**) are extended very efficiently because of both an increase in k_{cat} and a decrease in K_M. The **3FB** and **3,5DFB** self-pairs are extended with an efficiency ($>10^5\ M^{-1}\ min^{-1}$) that is only approximately 100-fold reduced relative to natural synthesis in the same sequence context.

Table 3. Incorporation of Natural and Unnatural Triphosphates Opposite Unnatural Bases in the Template[a]

5′-d(TAATACGACTCACTATAGGGAGA)				
3′-d(ATTATGCTGAGTGATATCCCTCT**X**GCTAGGTTACGGCAGGATCGC)				
X	triphosphate	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}M^{-1}$)
3FB	**3FB**	25 ± 4	12 ± 2	2.1×10^6
	A	26 ± 4	48 ± 10	5.4×10^5
	C	0.4 ± 0.1	195 ± 35	2.1×10^3
	G	0.3 ± 0.1	168 ± 80	1.8×10^3
	T	15 ± 8	158 ± 23	9.5×10^4
3,5DFB	**3,5DFB**	25 ± 5	218 ± 8	1.1×10^5
	A	5 ± 2	31 ± 16	1.6×10^5
	C	0.5 ± 0.2	4 ± 3	1.3×10^5
	G	0.34 ± 0.31	106 ± 33	3.2×10^3
	T	16 ± 7	187 ± 94	8.6×10^4
2,3,4TFB	A	8 ± 2	41 ± 13	2.0×10^5
	C	0.5 ± 0.2	334 ± 153	1.5×10^3
	G	0.7 ± 0.3	123 ± 43	5.7×10^3
	T	17 ± 8	79 ± 13	2.2×10^5
2,3DFB	A	7 ± 2	51 ± 10	1.4×10^5
	C	0.39 ± 0.02	6 ± 4	6.5×10^4
	G	1.6 ± 0.5	131 ± 28	1.2×10^4
	T	9 ± 3	49 ± 17	1.8×10^5
2FB	A	25 ± 9	24 ± 5	1.0×10^6
	C	nd[b]	nd[b]	$<1.0 \times 10^3$
	G	0.16 ± 0.01	77 ± 22	2.1×10^3
	T	1.4 ± 0.9	163 ± 39	8.6×10^3
3,4DFB	A	30 ± 9	63 ± 23	4.8×10^5
	C	0.8 ± 0.2	255 ± 93	3.1×10^3
	G	0.27 ± 0.20	184 ± 95	1.5×10^3
	T	20 ± 14	164 ± 52	1.2×10^5

[a] See Experimental Section for details. [b] Reaction was too inefficient for k_{cat} and K_M to be determined independently.

Because mispair extension rates also contribute to overall replication fidelity, we determined the rate at which each possible mispair between **3FB** and a natural base is extended, using the same assay and the appropriate primer-template combinations (Table 2). The most efficiently extended mispairs involve dT. With dT at the primer terminus, paired opposite **3FB** in the template, dCTP is incorporated opposite dG with an efficiency of $1.1 \times 10^6\ M^{-1}\ min^{-1}$; the mispair formed with **3FB** at the primer terminus and dT in the template is extended with an efficiency of $1.9 \times 10^5\ M^{-1}\ min^{-1}$. All other mispairs in either context are extended with efficiencies less than $6.4 \times 10^4\ M^{-1}\ min^{-1}$ (Table 2).

Because of the efficient and selective extension of the **3FB** self-pair, we characterized the efficiency and selectivity of its Kf-mediated synthesis. For comparison, we also characterized the efficiency and fidelity of **3,5FB** self-pair synthesis, which is also efficiently extended (Table 3). Both self-pairs are synthesized with a k_{cat} value of $25\ min^{-1}$. However, the d**3FB**TP substrate is bound 20-fold more tightly by the polymerase−DNA binary complex, resulting in more efficient self-pair

Table 4. Incorporation of Unnatural Triphosphates Opposite Natural Bases in the Template[a]

5′-d(TAATACGACTCACTATAGGGAGA)

3′-d(ATTATGCTGAGTGATATCCCTCTN**G**CTAGGTTACGGCAGGATCGC)

N	triphosphate	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($min^{-1}M^{-1}$)
A	**3FB**	6 ± 1	14 ± 3	4.3×10^5
	3,5DFB	5 ± 2	215 ± 4	2.3×10^4
C	**3FB**	nd[b]	nd[b]	$<1.0 \times 10^3$
	3,5DFB	0.06 ± 0.01	11 ± 3	5.5×10^3
G	**3FB**	14 ± 5	19 ± 11	7.4×10^5
	3,5DFB	10 ± 5	45 ± 27	2.2×10^5
T	**3FB**	6 ± 2	33 ± 24	1.8×10^5
	3,5DFB	12 ± 3	241 ± 106	5.0×10^4

[a] See Experimental Section for details. [b] Reaction was too inefficient for k_{cat} and K_M to be determined independently.

synthesis: 2.1×10^6 M^{-1} min^{-1}, compared to 1.1×10^5 M^{-1} min^{-1} for the **3,5DFB** self-pair. Remarkably, the **3FB** self-pair is synthesized only 20-fold less efficiently than is a natural base pair in the same sequence context. The fidelity of self-pair synthesis was determined by measuring the rate of incorporation of each natural dNTP opposite the unnatural base in the template, and comparing these with the value for self-pair synthesis efficiency. In the template, **3,5DFB** is somewhat promiscuous, directing Kf to insert dATP, dCTP, and dTTP with an efficiency of roughly 10^5 M^{-1} min^{-1}. In contrast, **3FB** in the template is less prone to mispair synthesis, directing Kf to insert dATP and dTTP with efficiencies of 5.4×10^5 M^{-1} min^{-1} and 9.5×10^4 M^{-1} min^{-1}, respectively, and dCTP and dGTP with efficiencies of only 2.1×10^3 M^{-1} min^{-1} and 1.8×10^3 M^{-1} min^{-1}, respectively.

Neither d**3FB**TP nor d**3,5DFB**TP are inserted opposite a natural base in the template at a rate competitive with that of natural Watson−Crick pair synthesis (Table 4). d**3FB**TP is inserted opposite the natural bases in the template with rates between 1.8×10^5 and 7.4×10^5 M^{-1} min^{-1}, except for dC in which case insertion is not detectable ($<10^3$ M^{-1} min^{-1}). Insertion of d**3,5DFB**TP is also least efficient opposite dC (5.5×10^3 M^{-1} min^{-1}) and otherwise varies between 2.3×10^4 M^{-1} min^{-1} and 2.2×10^5 M^{-1} min^{-1}. Thus, each unnatural nucleotide is selective for self-pair synthesis both in the template and as a triphosphate.

3. Discussion

Efforts to expand the genetic alphabet are based on developing a thermally stable and replicable unnatural base pair that is orthogonal to the natural dG:dC and dA:dT base pairs. There is no reason why such an unnatural base pair must resemble a natural base pair in any way. The only constraints are functional: the unnatural base pair must not significantly destabilize duplex DNA, and it must be efficiently and selectively replicated by a DNA polymerase. With this in mind, we have synthesized and characterized a wide variety of the self-pairs and heteropairs formed between nucleobase analogues with different structures and substituents.[22,23,26−29]

Several promising self-pairs based on the azaindole, isocarbostyril, or naphthyl scaffolds have been identified.[21,24,25] These self-pairs are thermally selective in duplex DNA and are enzymatically synthesized with reasonable efficiency and selectivity. However, after insertion of the unnatural triphosphate opposite its partner in the template, continued primer extension is inefficient in each case. There are at least two possible reasons for the poor self-pair extension rates: one is interstrand intercalation due to large aromatic surface area, and the other is the absence of suitably positioned heteroatoms to provide the nucleobases with optimal physiochemical properties such as H-bond acceptors, dipole moment, or polarizability. For example, heteroatom derivatization of the isocarbostyril scaffold changed the dipole moment of the nucleobase and lead to an improved self-pair extension rate.[27]

To explore the potential of heteroatom derivatization of a small nucleobase scaffold, we have begun a systematic analysis of the stability and polymerase recognition of self-pairs and heteropairs formed between two derivatized phenyl rings. Importantly, this scaffold is expected to abrogate interstrand intercalation by virtue of its reduced size. In this article, we report the synthesis and characterization of six unnatural self-pairs formed between nucleotides bearing fluorine-substituted benzene nucleobase analogues.

3.1. Stability and Thermal Selectivity of Unnatural Self-Pairs. Each self-pair is accommodated in B-form duplex DNA with reasonable stability; however, there is a surprising lack of dependence on the extent and pattern of fluorine substitution. All of the self-pair-containing duplexes denature between 52 and 54 °C, approximately 6 °C lower than a duplex with dA:dT in the same sequence context, and similar to a duplex containing a dT:dT mismatch. The stability of the parent phenyl self-pair is 52.8 °C,[36] demonstrating that the addition of fluorine substituents results in little or no stabilization of the self-pairs.

Despite their only moderate stability, the self-pairs are formed with good selectivity because of the destabilization of all possible mispairs with natural bases. In this respect, the most discriminating unnatural nucleobase analogue is **2FB** as the T_m values range from 46 to 52 °C. The least discriminating nucleobase analogue is **3FB**, which nonetheless retains thermal selectivity of 1.9 to 6.7 °C, typical of that observed among the natural pairs.

It is interesting to speculate about the nucleobase characteristics that may contribute to the self-pair and mispair stabilities, including structure, permanent electrostatic moments, polarizability, and hydrophobicity. Neither the specific fluorine substitution pattern nor the extent of fluorine substitution appears to affect the stability of either the self-pairs or the mispairs. Thus, it seems unlikely that the unnatural bases engage in specific electronic or structural interbase interactions within duplex DNA or that the polarizability of π-electron density within the unnatural nucleobase analogue is important. However, the mispair stabilities do roughly follow the hydrophobicity of the natural nucleobases as thymine is the most hydrophobic, followed closely by adenine, and next cytosine, then guanine. Thus, forced desolvation of the natural nucleobases appears to be the dominant factor in the thermal selectivity of the unnatural self-pairs, which do not require desolvation.

3.2. Replication Properties of Unnatural Self-Pairs. Kinetic Orthogonality of Unnatural Bases in Template DNA. Replication fidelity requires that the unnatural base in the template not direct the efficient misinsertion of any natural substrate (dNTP). The mispair synthesis rates for each unnatural base in the template reveal that not only are the unnatural mispairs with dA and dT the most thermally stable (see above), but they are also the most efficiently synthesized. In each case, efficient

(36) Matsuda, S.; Romesberg, F. E. Unpublished results.

synthesis is the result of a large k_{cat} value. The correlation between mispair stability and k_{cat} implies that the energetic cost of desolvation, argued above to be the most important determinant of stability in this series, is also a significant component of the activation barrier for dNTP insertion.

The only unnatural nucleotide that does not direct the preferential insertion of both dATP and dTTP is **2FB**. In the template, **2FB** directs Kf to insert dTTP with a k_{cat}/K_M of only 8.6×10^3 M^{-1} min^{-1}. While this is significantly more efficient than the insertion of dGTP and dCTP, it is more than 100-fold less efficient than the insertion of dATP. The reduced insertion efficiency of dTTP results from both a decrease in k_{cat} and an increase in K_M. The preferential insertion of dATP opposite **2FB** is reminiscent of the efficiency with which other unnatural nucleobase scaffolds bearing methyl groups alpha to the glycosidic bond direct Kf to insert dATP.[24] Previously, it was speculated that the methyl group was oriented in the developing minor groove, where it participated in stabilizing van der Waals interactions with the adenine methine. Perhaps the *ortho*-fluorine atom of **2FB** is also well accommodated both in the ternary complex with dATP and its corresponding transition state.

A second exception to the preferential synthesis of mispairs involving dA and dT is the relatively efficient insertion of dCTP opposite **2,3DFB** (6.5×10^4 M^{-1} min^{-1}), and especially **3,5DFB** (1.3×10^5 M^{-1} min^{-1}). The latter is as efficient as dATP incorporation opposite **3,5DFB** (1.6×10^5 M^{-1} min^{-1}) and more efficient than dTTP incorporation opposite **3,5DFB** (8.6×10^4 M^{-1} min^{-1}). Efficient insertion of dCTP in these cases results largely from a decrease in K_M which, at approximately 5 μM, is similar to that of natural base pair synthesis in the same sequence context. **2,3DFB** and **3,5DFB** each have *meta*-fluorine substitution that may facilitate dCTP binding. However, **3FB**, **3,4DFB**, and **2,3,4DFB**, which also possess *meta*-fluorine substitutents, direct dCTP insertion very inefficiently ($\sim 10^3$ M^{-1} min^{-1}), indicating that there is a complex role for the substitution pattern in dCTP insertion.

Overall, the unnatural nucleobases in the template resist insertion of dGTP and dCTP. Each member of the series directs the insertion of dATP and dTTP more efficiently. Nonetheless, all of the mispair synthesis rates are slow relative to natural synthesis and would not be expected to compromise the replication fidelity of an efficiently synthesized and extended unnatural self-pair.

3.3. Extension of Unnatural Self-Pairs. As discussed above, continued enzymatic polymerization after unnatural self-pair synthesis has limited replication of the previously described unnatural base pairs. The K_M values for extension of the fluorobenzene self-pairs vary by only 3.4-fold. dCTP is bound most tightly (85 μM) by the enzyme−DNA complex containing the **3FB** self-pair. This value indicates only 20-fold weaker binding than in natural DNA synthesis. The addition of a second *meta*-fluorine substitutent results in a slight decrease in dCTP binding (125 μM). Both *ortho*- and *para*-fluorine substitution within the self-pair also reduced the affinity of the polymerase−duplex complex for dCTP. There appears to be no correlation between these K_M values and self-pair stability.

In contrast to its effect on substrate binding, the nature of the unnatural base has a pronounced effect on the turnover rate associated with self-pair extension: the k_{cat} values vary by more than 50-fold. When a *meta*-fluorine substitutent is accompanied by an *ortho*- and/or *para*-fluorine, the turnover rate is modest (0.5−4.1 min^{-1}). However, when not accompanied by an *ortho*- or *para*-fluorine, as in **3FB** and **3,5DFB**, the turnover rate is remarkably high ($\sim$25 min^{-1}). In fact, these k_{cat} values are only approximately 6-fold reduced relative to natural synthesis (approximately 160 min^{-1}). These data support two possible hypotheses. First, that there exist strong anisotropic interactions between the incoming dNTP and the polymerase−DNA complex that develop upon proceeding from the polymerase−duplex−dNTP ternary complex to its transition state. Second, that there exist subtle yet functionally important differences in the structure of the primer terminus for the different self-pairs.

Table 5. Overall Fidelity of 3FB Self-Pair Replication

correct pair/mispair (primer:template)	synthesis fidelity[a,c]	extension fidelity[b,c]	total fidelity[d]
dA:dT/d**3FB**:dT	222	36	7992
dT:dA/d**3FB**:dA	93	3636	338 148
dC:dG/d**3FB**:dG	54	>40 000	>2 160 000
dG:dC/d**3FB**:dC	>40 000	1081	>43 240 000
d**3FB**:d**3FB**/dA:d**3FB**	4	5	20
d**3FB**:d**3FB**/dT:d**3FB**	22	2	44
d**3FB**:d**3FB**/dC:d**3FB**	1000	24	24 000
d**3FB**:d**3FB**/dG:d**3FB**	1167	>330	>385 110

[a] Ratio of synthesis efficiencies (Tables 3 and 4) for correct pair and mispair shown in first column of current table. [b] Ratio of extension efficiencies (Table 2) for correct pair and mispair shown in first column of current table. [c] Using a specificity constant for natural synthesis of 4.7×10^7 min^{-1} M^{-1}.[21] [d] Product of values in second and third columns of current table.

3.4. The 3FB Self-Pair and the Effort To Expand the Genetic Alphabet. On the basis of thermal stability data and CD spectra, the unnatural **3FB** nucleobase appears to selectively form a self-pair in an undistorted B-form DNA duplex of a given sequence. This interpretation is further supported by the NMR characterization of a different **3FB** self-pair-containing DNA duplex, one that closely resembles the sequence used in the kinetic assays. Preliminary data indicate a B-form conformation with Watson−Crick base pairing along the entire length of the DNA duplex, including the dA:dT base pair adjacent to the **3FB** self-pair, although the other flanking base pair may be somewhat more dynamic. The molecular details of the effect of incorporating **3FB** self-pairs into additional sequences is currently being studied by both NMR and X-ray crystallography. These studies will be of paramount importance in defining structure−function relationships that would facilitate optimization of the unnatural base pairs. For now, it is interesting to speculate that efficient single nucleotide extension of the **3FB** self-pair is related to the fact that it does not perturb a DNA duplex or, by extension, the primer-template terminus where it is recognized by the DNA polymerase.

The **3FB** self-pair is well-replicated by Kf polymerase, including efficient and selective synthesis, and most notably efficient and selective extension. In fact, the efficiency of self-pair synthesis and extension are only approximately 20-fold and 100-fold reduced, respectively, relative to natural DNA synthesis. The most significant fidelity problems result from the insertion of dATP and dTTP opposite **3FB** in the template and extension of these mispairs. This occurs with overall efficiencies that are 20- and 44-fold reduced, respectively, relative to self-pair replication (Table 5). Compared to mispairs involving dG and dC, the fidelity of **3FB** self-pair replication is in excess of 24 000-fold (Table 5). We also point out that enzymatic

proofreading, absent in the Kf mutant used in these experiments, is predicted to contribute to fidelity because of the lower thermal stability of the mispairs with dA and dT relative to the unnatural self-pair.[25]

Despite its lack of H-bonding and shape complementarity, the **3FB** self-pair appears to have all of the properties required of a third base pair, including those associated with structure, stability, and replication. This argues that shape complementarity and H-bonding are not unique in their ability to control the specific interbase interactions required for DNA stability and replication. It seems likely that through continued derivatization, optimal hydrophobic and van der Waals forces will result in a third base pair with stability and polymerase recognition sufficient to expand the genetic alphabet.

4. Experimental Section

General Synthetic Methods. All reactions were carried out in oven-dried glassware under inert atmosphere unless otherwise stated. All solvents were dried over 4 Å molecular sieves except dicloromethane (distilled from CaH_2), tetrahydrofuran (distilled from sodium and potassium metal), and diethyl ether (distilled from $LiAlH_4$). All other reagents were purchased from Aldrich. High-resolution mass spectroscopic data were obtained from the facilities at The Scripps Research Institute. All small molecule NMR spectra were collected by means of a Varian Mercury 200 MHz spectrometer (^{13}C 50.3 MHz; ^{19}F 188.2 MHz; ^{31}P 81 MHz). The ^{1}H and ^{13}C chemical shifts are referenced relative to TMS, and the ^{31}P and ^{19}F chemical shifts are referenced relative to 85% phosphoric acid in D_2O and neat trichlorofluoromethane, respectively.

Representative Procedure for Synthesis of Tetraisopropyldisiloxanediyl-Protected Nucleosides. To a solution of *n*-butyllithium (1.6 M in hexane, 2.8 mL) at −78 °C was added dropwise a solution of 1-bromo-2-fluorobenzene (0.467 mL, 4.27 mmol) in THF (7.5 mL). After being stirred at this temperature for 40 min, a solution of 3′,5′-*O*-((1,1,3,3-tetraisopropyl)disiloxanediyl)-2′-deoxy-D-ribono-1,4-lactone (1.0 g, 2.67 mmol) in THF (7.5 mL) was added via cannula. After being stirred for 1 h, the reaction was quenched with saturated aqueous NH_4Cl, extracted with diethyl ether, washed with saturated aqueous NH_4Cl and brine, and dried over $MgSO_4$. After concentration in vacuo, the resulting crude oily product (0.8 g, 1.69 mmol) was dissolved in CH_2Cl_2 (6 mL) and cooled to −78 °C after which was added Et_3SiH (0.81 mL, 5.06 mmol) followed by $BF_3{-}OEt_2$ (0.64 mL, 5.06 mmol), added dropwise. After being stirred at −78 °C for 7 h, the reaction was quenched by the addition of aqueous $NaHCO_3$ (10 mL). Additional water was added, the solution was extracted with diethyl ether, and after washing with aqueous $NaHCO_3$, water, and brine, it was dried over $MgSO_4$ and evaporated to dryness. Purification by column chromatography on silica gel (0−20% ethyl acetate in hexane) afforded the β-anomer of tetraisopropyl disiloxane-protected nucleoside **1a** in 16% yield over two steps.

Compound 1a. ^{1}H NMR ($CDCl_3$): δ 8.69 (1H, d, J = 3.5 Hz), 7.70−6.99 (3H, m), 5.38 (1H, t, J = 6.8 Hz), 4.60−4.54 (1H, m), 4.18 (1H, br d, J = 10.5 Hz), 4.03−3.94 (2H, m), 2.53−2.47 (1H, m), 2.15−2.09 (1H, m), 1.13−1.06 (28H, m). ^{19}F NMR ($CDCl_3$): δ −119.1. HRMS (MALDI−FTMS): calcd for $C_{23}H_{39}FO_4Si_2Na$ (MNa^+), 477.2263; found, 477.2261.

Compound 2a. ^{1}H NMR ($CDCl_3$): δ 7.21 (1H, dd, J = 6.9, 1.3 Hz), 7.06−6.89 (1H, m), 6.84 (1H, t, J = 7.6 Hz), 4.98 (1H, t, J = 7.2 Hz), 4.47−4.40 (1H, m), 4.03 (1H, t, J = 7.4 Hz), 3.82 (2H, br s), 2.31−2.20 (1H, m), 2.01−1.86 (1H, m), 1.13−1.06 (28H, m). ^{19}F NMR ($CDCl_3$): δ −113.3 (1F, t, J = 3.9 Hz). ^{13}C NMR ($CDCl_3$): δ 165.6, 160.7, 145.3, 130.1, 121.5, 114.6, 113.1, 86.6, 73.1, 63.7, 43.4, 18.2, 17.8, 17.7, 17.6, 17.5, 17.3, 17.2, 13.8, 13.7, 13.3, 13.2, 12.8. HRMS (MALDI−FTMS): calcd for $C_{23}H_{39}FO_4Si_2Na$ (MNa^+), 477.2263; found, 477.2260.

Compound 3a. ^{1}H NMR ($CDCl_3$): δ 7.99 (2H, d, J = 6.6 Hz), 7.70 (2H, d, J = 6.6 Hz), 7.27−7.16 (4H, m), 7.01−6.96 (2H, m), 6.77−6.65 (1H, m), 6.45−6.40 (1H, m), 5.61−5.58 (1H, m), 5.35 (1H, dd, J = 7.8, 5.2 Hz), 4.75−4.62 (1H, m), 4.59 (2H, s), 3.06−2.82 (1H, m), 2.41 (3H, s), 2.39 (3H, s), 2.40−2.30 (1H, m). ^{13}C NMR ($CDCl_3$): δ 166.5, 166.3, 165.9, 165.6, 160.9, 160.7, 145.3, 144.5, 144.1, 129.9, 129.4, 127.2, 127.1, 109.1, 108.5, 103.2, 83.5, 79.9, 64.7, 41.8, 21.9. HRMS (MALDI−FTMS): calcd for $C_{27}H_{25}F_2O_5$ (MH^+), 466.1593; found, 466.1588.

Compound 4a. ^{1}H NMR ($CDCl_3$): δ 7.22−7.16 (1H, m), 7.01−6.92 (2H, m), 5.25 (1H, t, J = 7.0 Hz), 4.47−4.37 (1H, m), 4.05 (1H, dd, J = 11.5, 2.9 Hz), 3.91−3.74 (2H, m), 2.43−2.33 (1H, m), 2.07−1.97 (1H, m), 1.07−0.94 (28H, m). ^{19}F NMR ($CDCl_3$): δ −139.8 (1F, m), −144.5 (1F, m). ^{13}C NMR ($CDCl_3$): δ 124.3, 124.1, 122.0, 122.0, 116.3, 116.0, 85.9, 73.0, 72.4, 63.1, 41.9, 17.8, 17.7, 17.6, 17.5, 17.3, 17.2, 13.7, 13.6, 13.2, 12.8. HRMS (MALDI−FTMS): calcd for $C_{23}H_{38}F_2O_4Si_2Na$ (MNa^+), 495.2169; found, 495.2169.

Compound 5a. ^{1}H NMR ($CDCl_3$): δ 7.26−7.07 (3H, m), 5.04 (1H, t, J = 7.1 Hz), 4.54−4.46 (1H, m), 4.09 (1H, t, J = 4.4 Hz), 3.89 (1H, dd, J = 14.6, 7.0 Hz), 2.44−2.31 (1H, m), 2.17−1.92 (1H, m), 1.08−0.87 (28H, m). ^{13}C NMR ($CDCl_3$): δ 121.9, 117.4, 117.1, 115.2, 114.8, 86.6, 63.4, 43.3, 17.9, 17.7, 17.6, 17.5, 17.3, 17.2, 13.8, 13.6, 13.3, 12.8. ^{19}F NMR ($CDCl_3$): δ −138.11 (1F, ddd, J = 20.7, 11.6, 6.9 Hz), −140.28 (1F, ddd, J = 17.1, 7.9, 5.3 Hz). HRMS (MALDI−FTMS): calcd for $C_{23}H_{38}F_2NaO_4Si_2$ (MNa^+), 495.2174; found, 495.2169.

Compound 6a. ^{1}H NMR ($CDCl_3$): δ 7.19−7.11 (1H, m), 6.90−6.76 (1H, m), 5.29 (1H, t, J = 6.8 Hz), 4.46−4.36 (1H, m), 4.03 (1H, dd, J = 11.7, 3.1 Hz), 3.91−3.75 (2H, m), 2.45−2.32 (1H, m), 2.04−1.91 (1H, m), 1.00−0.94 (28H, m). ^{19}F NMR ($CDCl_3$): δ −137.6 (1F, m), −140.2 (1F, m), −161.7 (1F, m). ^{13}C NMR ($CDCl_3$): δ 120.6, 112.2, 111.7, 85.7, 72.6, 72.0, 62.9, 41.7, 31.9, 22.9, 17.8, 17.6, 17.6, 17.4, 17.3, 17.2, 13.7, 13.6, 13.2, 12.8. HRMS (MALDI−FTMS): calcd for $C_{23}H_{37}F_3O_4Si_2Na$ (MNa^+), 513.2080; found, 513.2077.

Representative Procedure for Deprotection of Tetraisopropyldisiloxanediyl-Protected Nucleosides. To a stirred solution of **4a** (0.134 g, 0.283 mmol) in THF (6 mL) was added dropwise TBAF (1 M in THF, 0.848 mL). After being stirred at room temperature for 45 min, the reaction was quenched with saturated aqueous NH_4Cl (5 mL) and extracted with ether/THF (1:1). NaCl was added to the aqueous phase, which was further extracted, dried over $MgSO_4$, and evaporated to dryness. Purification by column chromatography on silica gel (50−100% ethyl acetate in hexane) afforded free nucleoside **4b** in 92% yield.

Compound 1b. ^{1}H NMR ($CDCl_3$): δ 7.58 (1H, dt, J = 7.6, 1.8 Hz), 7.26−6.98 (3H, m), 5.38 (1H, dd, J = 10.3, 5.2 Hz), 4.33 (1H, t, J = 3.5 Hz), 3.99−3.74 (1H, m), 3.74 (2H, br s), 2.33−2.22 (1H, m), 1.99−1.82 (1H, m). ^{19}F NMR ($CDCl_3$): δ −121.64 (1F, t, J = 5.3 Hz). ^{13}C NMR ($CDCl_3$): δ 129.0, 127.3, 124.2, 114.9, 114.5, 87.7, 74.2, 74.1, 73.2, 62.8, 42.4. HRMS (MALDI−FTMS): calcd for $C_{11}H_{14}FO_3$ (MH^+), 213.0927; found, 213.0924.

Compound 2b. ^{1}H NMR ($CDCl_3$): δ 7.32 (1H, dd, J = 7.9, 5.9 Hz), 7.15 (2H, d, J = 7.4 Hz), 7.01−6.89 (1H, m), 5.12 (1H, dd, J = 10.2, 5.4 Hz), 4.32 (1H, dd, J = 4.2, 1.8 Hz), 4.01−3.93 (1H, m), 3.66 (2H, d, J = 5.0 Hz), 2.22 (1H, ddd, J = 13.0, 5.6, 1.6 Hz), 2.00−1.82 (1H, m). ^{13}C NMR ($CDCl_3$): δ 165.6, 145.3, 130.0, 121.7, 114.2, 112.7, 88.1, 79.7, 73.2, 62.9, 43.8. ^{19}F NMR ($CDCl_3$): δ −115.54 (1F, t, J = 5.3 Hz). HRMS (MALDI−FTMS): calcd for $C_{11}H_{13}FNaO_3$ (MNa^+), 235.0746; found, 235.0738.

Compound 3b. ^{1}H NMR ($CDCl_3$): δ 7.01−6.80 (2H, m), 6.78−6.73 (1H, m), 5.11 (1H, t, J = 7.4 Hz), 4.40 (1H, dd, J = 6.8, 5.2 Hz), 4.02 (1H, dd, J = 5.2, 3.4 Hz), 3.70−3.55 (2H, m), 2.82−2.61 (1H, m), 1.95−1.78 (1H, m). ^{13}C NMR ($CDCl_3$): δ 165.9, 148.1, 136.2, 113.5, 128.9, 102.7, 86.7, 74.8, 64.8, 55.5, 43.3. HRMS (MALDI−FTMS): calcd for $C_{11}H_{13}F_2O_3$ (MH^+), 231.0755; found, 231.0750.

Compound 4b. ^{1}H NMR (CD_3OD): δ 7.40−7.34 (1H, m), 7.20−7.08 (2H, m), 5.37 (1H, dd, J = 10.2, 5.5 Hz), 4.34−4.31 (1H, m), 3.99−3.92 (1H, m), 3.73−3.64 (2H, m), 2.34−2.25 (1H, m), 2.00−

1.84 (1H, m). ^{19}F NMR (CD_3OD): δ −138.5 (1F, m), −143.6 (1F, m). ^{13}C NMR (CD_3OD): δ 131.9, 124.3, 122.1, 115.9, 115.6, 87.9, 73.8, 73.1, 72.1, 62.7, 42.4. HRMS (MALDI−FTMS): calcd for $C_{11}H_{12}F_2O_3$-Na (MNa^+), 253.0647; found, 253.0644.

Compound 5b. ^{1}H NMR ($CDCl_3$): δ 7.31 (1H, dd, J = 12.0, 8.6 Hz), 7.22−7.11 (2H, m), 5.08 (1H, dd, J = 10.6, 5.4 Hz), 4.82 (1H, br s), 4.32 (1H, d, J = 6.0 Hz), 4.03−3.81 (1H, m), 3.67 (2H, d, J = 4.6 Hz), 2.25−2.15 (1H, m), 1.95−1.80 (1H, m). ^{13}C NMR ($CDCl_3$): δ 139.9, 122.3, 117.0, 116.7, 115.0, 114.6, 88.1, 79.2, 73.2, 62.8, 43.9. ^{19}F NMR ($CDCl_3$): δ −140.53 (1F, ddd, J = 23.8, 14.0, 5.3 Hz), −142.36 (1F, dd, J = 13.3, 6.6 Hz). HRMS (MALDI−FTMS): calcd for $C_{11}H_{12}F_2NaO_3$ (MNa^+), 253.0647; found, 253.0648.

Compound 6b. ^{1}H NMR (CD_3OD): δ 7.40−7.32 (1H, m), 7.13−6.99 (1H, m), 5.32 (1H, dd, J = 10.2, 5.9 Hz), 4.34−4.30 (1H, m), 3.97−3.91 (1H, m), 3.71−3.64 (2H, m), 2.33−2.22 (1H, m), 2.00−1.83 (1H, m). ^{19}F NMR (CD_3OD): δ -139.4 (1F, m), −143.2 (1F, m), −161.1 (1F, m). ^{13}C NMR (CD_3OD): δ 121.2, 121.1, 112.2, 112.1, 111.8, 111.7, 87.9, 73.5, 73.0, 62.7, 42.3. HRMS (MALDI−FTMS): calcd for $C_{11}H_{11}F_3O_3Na$ (MNa^+), 271.0552; found, 271.0559.

General Procedure for Phosphoramidite Synthesis. To a solution of free nucleoside (0.6 mmol) coevaporated with toluene (10 mL) was added pyridine (2.5 mL), followed by DMTr-Cl (1.5 equiv) in pyridine (1.2 mL) over 10 min. After being stirred for an additional 30 min at room temperature, ethyl acetate was added (50 mL), and the organic phase was washed with saturated $NaHCO_3$ (20 mL) and brine (20 mL) and dried over Na_2SO_4. The solvent was removed in vacuo, and flash column chromatography on silica gel (20−50% ethyl acetate in hexane) yielded a light yellow foam. The tritylated nucleoside (0.2 mmol) was dissolved in CH_2Cl_2 (2 mL), diisopropylethylamine (4 equiv) was added, and the solution was cooled to 0 °C. 2-Cyanoethyl diisopropyl-aminochloro phosphoramidite (1.3 equiv) was added dropwise, and the reaction was allowed to reach room temperature over 15 min. The solution was transferred to ethyl acetate (50 mL), and the organic phase was washed with saturated $NaHCO_3$ (20 mL) and brine (20 mL) and dried over Na_2SO_4. The solvent was removed in vacuo and flash column chromatography on silica gel washed with 2% triethylamine in hexane (20% ethyl acetate in hexane with 2% triethylamine) afforded a white foam. The phosphoramidite was generally obtained in 70% yield over two steps, as a mixture of two diastereoisomers.

Compound 1c. ^{1}H NMR ($CDCl_3$): δ 7.64−7.02 (13H, m), 6.85 (4H, d, J = 8.6 Hz), 5.48 (1H, dd, J = 9.4, 6.0 Hz), 4.44 (1H, t, J = 2.8 Hz), 4.13 (1H, dd, J = 7.4, 5.0 Hz), 3.79 (6H, s), 3.34 (2H, t, J = 5.0 Hz), 2.82 (1H, br s), 2.38 (1H, dd, J = 13.2, 5.8 Hz), 2.18−1.95 (1H, m). ^{19}F NMR ($CDCl_3$): δ −118.93 (1F, t, J = 5.3 Hz). ^{13}C NMR ($CDCl_3$): δ 162.5, 158.7, 157.6, 149.7, 145.2, 136.6, 136.3, 130.8, 129.7, 129.5, 129.0, 128.9, 128.5, 128.1, 127.5, 127.4, 127.1, 124.4, 124.3, 115.5, 115.1, 113.4, 110.0, 86.5, 86.3, 85.0, 75.0, 74.5, 74.4, 64.7, 55.5, 42.9. HRMS (MALDI−FTMS): calcd for $C_{32}H_{31}FNaO_5$ (MNa^+), 537.2053; found, 537.2051.

Compound 1d. ^{1}H NMR ($CDCl_3$): δ 7.51−7.02 (13H, m), 6.73 (4H, dd, J = 8.8, 2.6 Hz), 5.33 (1H, dd, J = 10.3, 5.3 Hz), 4.46 (1H, br s), 4.15 (1H, br s), 3.74 (6H, s), 3.72−3.41 (4H, m), 3.23 (2H, t, J = 4.6 Hz), 2.46 (1H, t, J = 6.2 Hz), 2.44−2.39 (1H, m), 2.38 (1H, t, J = 6.2 Hz), 1.97−1.81 (1H, m), 1.17−0.81 (14H, m). ^{19}F NMR ($CDCl_3$): δ −119.12 (1F, d, J = 5.3 Hz). ^{13}C NMR ($CDCl_3$): δ 158.7, 157.6, 151.1, 145.2, 136.3, 130.4, 129.2, 128.9, 128.5, 128.0, 127.5, 127.0, 124.4, 115.5, 115.1, 113.3, 110.0, 86.4, 85.8, 74.6, 64.4, 58.8, 58.4, 55.5, 43.6, 43.4, 42.2, 24.8, 24.3, 20.6. ^{31}P NMR ($CDCl_3$): δ 149.40, 148.87. HRMS (MALDI−FTMS): calcd for $C_{41}H_{48}FN_2NaO_6P$ (MNa^+), 737.3132; found, 737.3127.

Compound 2c. ^{1}H NMR ($CDCl_3$): δ 7.56−7.28 (12H, m), 6.98 (1H, t, J = 6.4 Hz), 6.87 (4H, d, J = 8.6 Hz), 5.23 (1H, dd, J = 10.0, 5.6 Hz), 4.43 (1H, br s), 4.25−4.10 (1H, m), 3.79 (6H, s), 3.36 (2H, t, J = 4.8 Hz), 2.33−2.30 (1H, m), 2.05−1.98 (1H, m). ^{19}F NMR ($CDCl_3$): δ −113.32 (1F, d, J = 5.3 Hz). ^{13}C NMR ($CDCl_3$): δ 165.6, 160.8, 158.7, 149.5, 145.2, 136.3, 130.4, 130.2, 128.5, 128.1, 127.1, 124.3, 121.9, 114.7, 114.3, 113.4, 112.9, 86.9, 86.5, 79.6, 74.5, 64.9, 55.5, 44.2. HRMS (MALDI−FTMS): calcd for $C_{32}H_{31}FO_5Na$ (MNa^+), 537.2048; found, 537.2045.

Compound 2d. ^{1}H NMR ($CDCl_3$): δ 7.51−7.02 (12H, m), 6.98 (1H, t, J = 6.2 Hz), 6.87 (4H, d, J = 8.4 Hz), 5.10 (1H, dd, J = 10.0, 5.2 Hz), 4.41 (1H, dd, J = 9.8, 5.5 Hz), 4.16 (1H, br s), 3.71 (6H, s), 3.65−3.40 (4H, m), 3.21 (2H, t, J = 4.0 Hz), 2.46 (1H, t, J = 6.2 Hz), 2.32 (1H, t, J = 6.2 Hz), 2.28−2.01 (1H, m), 2.01−1.81 (1H, m), 1.25−0.95 (14H, m). ^{19}F NMR ($CDCl_3$): δ −113.42 (1F, d, J = 5.6 Hz). ^{31}P NMR ($CDCl_3$): δ 149.16, 149.03. ^{13}C NMR ($CDCl_3$): δ 165.6, 160.8, 158.7, 145.2, 136.2, 130.4, 130.2, 130.0, 128.5, 128.1, 127.0, 121.8, 117.9, 114.7, 114.3, 113.4, 112.9, 86.4, 86.1, 86.0, 79.9, 76.2, 76.1, 75.8, 64.5, 58.8, 58.4, 55.4, 43.6, 43.4, 26.0, 24.8, 24.7, 20.7, 20.6, 20.4. HRMS (MALDI−FTMS): calcd for $C_{41}H_{49}FN_2O_6P$ (MH^+), 715.3312; found, 715.3310.

Compound 3c. ^{1}H NMR ($CDCl_3$): δ 7.51−7.20 (10H, m), 6.97−6.70 (6H, m), 6.67−6.66 (1H, m), 5.12 (1H, t, J = 7.4 Hz), 4.45 (1H, br s), 4.19 (1H, d, J = 6.4 Hz), 3.80 (6H, s), 3.40−3.20 (2H, m), 2.80−2.61 (1H, m), 2.06−1.94 (1H, m). HRMS (MALDI−FTMS): calcd for $C_{32}H_{31}F_2O_5$ (MH^+), 533.2061; found, 533.2056.

Compound 3d. ^{1}H NMR: 7.40−7.18 (12H, m), 6.89−6.62 (7H, m), 5.11 (1H, m), 4.25−4.18 (1H, m), 4.18 (1H, br s), 3.81−3.40 (12H, m), 3.20 (2H, t, J = 6.6 Hz), 2.60−2.42 (2H, m), 2.32 (2H, t, J = 6.5 Hz), 1.40 (1H, d, J = 6.2 Hz), 1.19−0.96 (6H, m). ^{31}P NMR ($CDCl_3$): δ 149.15, 149.06. HRMS (MALDI−FTMS): calcd for $C_{41}H_{48}F_2N_2O_5P$ (MH^+), 717.3191; found, 717.3187.

Compound 4c. ^{1}H NMR ($CDCl_3$): δ 7.47 (2H, d, J = 7.8 Hz), 7.40−7.18 (9H, m), 7.05 (2H, t, J = 3.5 Hz), 6.83 (4H, d, J = 8.6 Hz), 5.44 (1H, dd, J = 9.7, 4.9 Hz), 4.43 (1H, t, J = 3.0 Hz), 4.12 (1H, dd, J = 7.5, 3.0 Hz), 3.79 (6H, s), 3.36 (2H, t, J = 4.8 Hz), 2.58 (1H, br s), 2.38 (1H, dd, J = 12.8, 5.8 Hz), 2.10−1.90 (1H, m). ^{19}F NMR ($CDCl_3$): δ −144.30 (1F, t, J = 8.9 Hz), −139.77 (1F, dt, J = 10.9, 5.3 Hz). ^{13}C NMR ($CDCl_3$): δ 158.7, 145.0, 136.2, 132.2, 130.3, 128.4, 128.1, 127.0, 124.2, 122.0, 116.3, 115.9, 113.4, 86.5, 86.4, 74.4, 74.1, 64.5, 55.5, 42.8. HRMS (MALDI−FTMS): calcd for $C_{32}H_{31}F_2O_5$ (MH^+), 533.2140; found, 533.2138.

Compound 4d. ^{1}H NMR ($CDCl_3$): δ 7.38 (2H, d, J = 7.8 Hz), 7.26−7.12 (9H, m), 6.97 (2H, dd, J = 12.2, 5.6 Hz), 6.76−6.70 (4H, m), 5.34 (1H, dd, J = 12.7, 5.2 Hz), 4.44 (1H, br s), 4.15 (1H, br s), 3.70 (6H, s), 3.69−3.40 (5H, m), 3.22 (2H, dd, J = 9.8, 5.0 Hz), 2.52 (1H, t, J = 6.2 Hz), 2.37 (1H, t, J = 6.2 Hz), 2.35−2.25 (1H, m), 1.98−1.84 (1H, m), 1.25−0.96 (14H, m). ^{19}F NMR ($CDCl_3$): −144.46 (1F, t, J = 8.8 Hz), −139.88 (1F, dt, J = 10.9, 5.3 Hz). ^{13}C NMR ($CDCl_3$): δ 158.7, 145.0, 136.2, 130.4, 128.5, 128.0, 127.0, 124.3, 122.0, 116.3, 115.9, 113.3, 86.4, 85.8, 74.3, 64.1, 58.8, 58.4, 55.5, 43.7, 43.4, 42.1, 25.0, 24.8, 24.7, 20.6, 20.4. ^{31}P NMR ($CDCl_3$): δ 149.54, 149.01. HRMS (MALDI−FTMS): calcd for $C_{41}H_{47}F_2N_2NaO_6P$ (MNa^+), 755.3037; found, 755.3032.

Compound 5c. ^{1}H NMR ($CDCl_3$): δ 7.51 (2H, d, J = 8.2 Hz), 7.40 (4H, d, J = 7.8 Hz), 7.38−7.22 (4H, m), 7.18−7.07 (2H, m), 6.86 (4H, d, J = 7.4 Hz), 5.17 (1H, dd, J = 9.8, 5.4 Hz), 4.43 (1H, d, J = 2.2 Hz), 4.19 (1H, br s), 3.79 (6H, s), 3.34 (2H, t, J = 4.2 Hz), 2.27 (1H, dd, J = 13.1, 5.6 Hz), 2.06−1.93 (1H, m). ^{19}F NMR ($CDCl_3$): δ −138.15 (1F, dd, J = 17.9, 7.3 Hz), −140.21 (1F, dd, J = 14.5, 7.9 Hz). ^{13}C NMR ($CDCl_3$): δ 158.8, 149.4, 145.1, 139.6, 139.4, 136.2, 130.4, 130.2, 128.4, 128.1, 127.1, 122.3, 122.2, 122.2, 122.1, 117.4, 117.1, 115.4, 115.0, 113.4, 87.0, 86.6, 79.2, 74.4, 64.8, 55.5, 44.3. HRMS (MALDI−FTMS): calcd for $C_{32}H_{30}F_2O_5Na$ (MNa^+), 555.1953; found, 555.1948.

Compound 5d. ^{1}H NMR ($CDCl_3$): δ 7.40−7.15 (10H, m), 7.04−6.97 (2H, m), 6.76−6.71 (4H, m), 5.04 (1H, dd, J = 10.5, 5.1 Hz), 4.47−4.39 (1H, m), 4.15−4.06 (1H, m), 3.79−3.40 (4H, m), 3.70 (3H, s), 3.23−3.18 (2H, m), 2.52 (1H, t, J = 6.6 Hz), 2.37 (1H, t, J = 6.6 Hz), 2.30−2.21 (1H, m), 1.93−1.82 (1H, m), 1.13−1.00 (12H, m). ^{19}F NMR ($CDCl_3$): δ −138.2 (1F, m), −140.3 (1F, m). ^{31}P NMR

($CDCl_3$): δ 149.2, 149.1. ESI-MS: calcd for $C_{41}H_{47}F_2N_2NaO_6P$ (MNa^+), 755; found, 755.

Compound 6c. ^{1}H NMR ($CDCl_3$): δ 7.50−7.22 (10H, m), 6.94−6.82 (5H, m), 5.38 (1H, dd, J = 9.6, 6.0 Hz), 4.44−4.41 (1H, m), 4.12−4.06 (1H, m), 3.79 (6H, s), 3.52−3.29 (2H, m), 2.40−2.31 (1H, m), 2.06−1.97 (1H, m). ^{19}F NMR ($CDCl_3$): δ −136.4 (1F, m), −140.1 (1F, m), −161.6 (1F, m). ^{13}C NMR ($CDCl_3$): δ 158.7, 145.0, 136.1, 130.3, 128.4, 128.1, 127.1, 113.4, 86.6, 86.4, 74.4, 73.8, 64.4, 55.5, 42.8. HRMS (MALDI−FTMS): calcd for $C_{32}H_{29}F_3O_5Na$ (MNa^+), 573.1859; found, 573.1844.

Compound 6d. ^{1}H NMR ($CDCl_3$): δ 7.39−7.11 (10H, m), 6.89−6.70 (5H, m), 5.27 (1H, dd, J = 10.2, 5.1 Hz), 4.47−4.41 (1H, m), 4.14−4.08 (1H, m), 3.75−3.44 (4H, m), 3.70 (6H, s), 3.25−3.13 (1H, m), 2.55−2.13 (1H, m), 2.52 (1H, t, J = 6.3 Hz), 2.49−2.29 (1H, m), 2.37 (1H, t, J = 6.3 Hz), 1.91−1.85 (1H, m), 1.20−0.92 (12H, m). ^{19}F NMR ($CDCl_3$): δ −136.6 (1F, m), −140.1 (1F, m), −161.7 (1F, m). ^{31}P NMR ($CDCl_3$): δ 149.5, 149.0. ESI-MS: calcd for $C_{41}H_{47}F_3N_2O_6P$ (MH^+), 751; found, 751.

General Procedure for Triphosphate Synthesis. Proton sponge (1.5 equiv) and the free nucleoside (1 equiv) were dissolved in trimethyl phosphate (0.3 M) and cooled to 0 °C. $POCl_3$ (1.05 equiv) was added dropwise, and the purple slurry was stirred at 0 °C for 2 h. Tributylamine (4 equiv) was added, followed by a solution of tributylammonium pyrophosphate (2.5 equiv) in DMF (0.15 M). After 1 min, the reaction was quenched by addition of 1 M aqueous Et_3N-HCO_3 (20 vol-equiv). The resulting solution stood for 40 min at 0 °C and was then lyophilized. Purification by reverse-phase (C18) HPLC (4−35% CH_3CN in 0.1 M Et_3N-HCO_3, pH 7.5) followed by lyophilization afforded the triphosphate as a white solid.

Procedure for Synthesis and Purification of Oligonucleotides Used in Thermal Denaturation, CD, and Replication Experiments. Oligonucleotides were synthesized using standard β-cyanoethylphosphoramidite chemistry by means of an Applied Biosystems Inc. 392 DNA/RNA synthesizer and reagents that were either synthesized in-house or purchased from Glen Research, Sterling, VA. Trityl-off oligonucleotides were cleaved from the CPG support and deprotected by a 12 h reaction with concentrated ammonia at 60 °C. Purification by denaturing polyacrylamide gel electrophoresis (10−20%, 8 M urea) and visualization of the ssDNA bands by UV shadowing was followed by electroelution and desalting with Sep-Pak (C18) cartridges. The pure oligonucleotides were dissolved in water, and their concentration was determined by using the value for absorbance at 260 nm in the Biopolymer Calculator (http://paris.chem.yale.edu/extinct.html); note that for oligonucleotides containing an unnatural residue, dC replaced it in the sequence.

Thermal Denaturation. Oligonucleotide duplex denaturation temperature measurements were made in buffer containing 10 mM PIPES, pH 7.0, 10 mM $MgCl_2$, and 100 mM NaCl at an ssDNA concentration of 3 μM using a Cary 200 Bio UV−visible spectrometer. Measurements of absorbance at 260 nm were taken over the temperature range 16−80 °C at 0.5 °C/min intervals after rapid annealing and cooling. Melting temperatures were obtained from one scan by means of the derivative method in the Cary Win UV software.

CD Spectroscopy. CD experiments were performed by means of an AVIV model 61 DS spectropolarimeter equipped with a Peltier thermoelectric temperature control unit. The duplex concentration was 3 μM in buffer containing 10 mM PIPES, pH 7.0, 10 mM $MgCl_2$, and 100 mM NaCl. The data were collected at 25 °C using a quartz cuvette of 1 cm path length with scanning from 360 to 220 nm, a time constant of 3 s, and a step size 0.5 nm. After subtracting the buffer reference from the sample spectra, we converted the observed signal (in units of millidegrees) to molar ellipticity (in units of degree cm^2/dmol) and plotted it against wavelength.

Gel-Based Kinetic Assay. DNA polymerase I (large fragment, exonuclease deficient) from *E. coli* was purchased from New England Biolabs. Primer oligonucleotides were 5′-radiolabeled with T4 polynucleotide kinase (New England Biolabs) and [γ-^{33}P]-ATP (Amersham Biosciences). Primers were annealed to template oligonucleotides in the reaction buffer by heating to 95 °C followed by slow cooling to ambient 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 the addition of 20 μL of loading dye (95% formamide, 20 mM EDTA, and sufficient amounts of bromophenol blue and xylene cyanole). The reaction mixtures were resolved by 15% polyacrylamide gel electrophoresis, and the radioactivity was quantified by means of a PhosphorImager (Molecular Dynamics) and ImageQuant software. A plot of k_{obs} versus triphosphate concentration was fit to a Michaelis−Menten equation using the program Kaleidagraph (Synergy Software). The data presented are averages of three independent determinations.

Procedure for Preparation of 3FB Self-Pair-Containing DNA Duplex Used in NMR Experiments. Trityl-on oligonucleotides were cleaved and deprotected by a 12 h reaction with concentrated ammonia at 60 °C, then dried with centrifugation in vacuo (Savant SpeedVac Plus SC110A). Detritylation was achieved by adding 80% acetic acid (2 mL per 1 μmol scale synthesis), vortexing, and incubating at room temperature for 20 min. The mixture was divided into two aliquots and transferred to an ice bath, and ice cold concentrated aqueous ammonium hydroxide (1 mL per one-half 1 μmol scale synthesis) was added dropwise with continuous mixing by inversion of the capped tube. The mixture was dried with centrifugation in vacuo, dissolved in 50 mM ammonium formate (0.5 mL per 1 μmol scale synthesis), and clarified with 0.65 μm centrifugal filtration devices (Ultrafree-MC, Millipore). After purification by reverse-phase (C18) HPLC (3−15% CH_3CN in 50 mM ammonium formate, pH 7.0 over 30 min) followed by lyophilization, the oligonucleotides were redissolved in water (0.5 mL per 1 μmol scale synthesis), and their concentrations were determined by using the value for absorbance at 260 nm in the Biopolymer Calculator; note that for oligonucleotides containing an unnatural residue, the position of the unnatural in the sequence was replaced by dC. Duplexes were formed by mixing equimolar amounts of each single strand, heating at 90 °C for 2 min, cooling on ice, and lyophilizing.

NMR Spectroscopic Studies of 3FB Self-Pair-Containing DNA Duplex. Lyophilized duplex DNA containing the **3FB** unnatural self-pair was dissolved in buffer containing 8.5 mM sodium phosphate, pH 7.0, 134 mM NaCl, and 0.25 mM EDTA in either 100% D_2O or in 90%/10% H_2O/D_2O at a final analyte concentration of 0.6 mM. All NMR spectra were acquired on a Bruker Avance 600 MHz spectrometer equipped with a ^{1}H/^{13}C/^{15}N-TXI CryoProbe or on an Avance 400 MHz spectrometer equipped with a ^{1}H/^{13}C/^{31}P/^{19}F-QNP probe (Bruker Biospin, Billerica, MA). Spectra were acquired at temperatures of 295, 303, and 310 K. Proton resonance assignments were made according to established procedures.[34,35,37−39] The 2D NOESY spectrum shown in Figure 2 was acquired by collecting 96 transients in each of the 700 T_1 experiments using a spectral width of 13 227 Hz, a mixing time of 200 ms, a recycle delay of 2 s, and excitation sculpting with gradients for water suppression[40] using the release pulse program noesyesgpph.

Acknowledgment. Funding was provided by the National Institutes of Health (No. GM 60005 to F.E.R.) and the JSPS Research Fellowships for Young Scientists (S.M.). We thank Peter Wright and Linda Tennant for granting access to the CD spectropolarimeter.

JA049961U

(37) Chou, S.-H.; Wemmer, D. R.; Reid, B. R. *Biochemistry* **1983**, *22*, 3037−3041.
(38) Boelens, R.; Scheek, R. M.; Dijkstra, K.; Kaptein, R. *J. Magn. Reson.* **1985**, *62*, 378−386.
(39) Sklenar, V.; Brooks, B. R.; Zon, G.; Bax, A. *FEBS Lett.* **1987**, *216*, 249−252.
(40) Hwang, T.-L.; Shaka, A. J. *J. Magn. Reson., Ser. A* **1995**, *112*, 275−279.