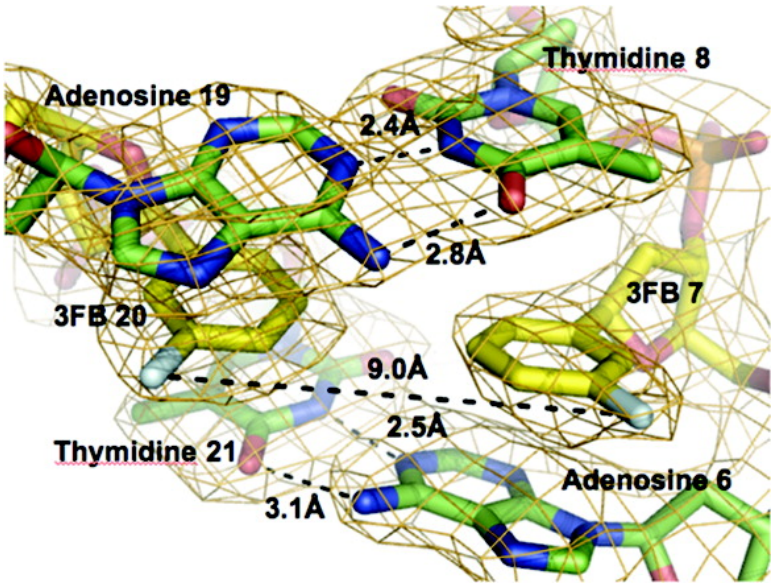
Article

Efforts toward Expansion of the Genetic Alphabet: Structure and Replication of Unnatural Base Pairs

Shigeo Matsuda, Jeremiah D. Fillo, Allison A. Henry, Priyamrada Rai, Steven J. Wilkens, Tammy J. Dwyer, Bernhard H. Geierstanger, David E. Wemmer, Peter G. Schultz, Glen Spraggon, and Floyd E. Romesberg

J. Am. Chem. Soc., **2007**, 129 (34), 10466-10473• DOI: 10.1021/ja072276d • Publication Date (Web): 08 August 2007



More About This Article

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

- Supporting Information
- Links to the 18 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 08/08/2007

Efforts toward Expansion of the Genetic Alphabet: Structure and Replication of Unnatural Base Pairs

Shigeo Matsuda,[†] Jeremiah D. Fillo,[‡,§] Allison A. Henry,[†] Priyamrada Rai,[||] Steven J. Wilkens,[‡] Tammy J. Dwyer,[§] Bernhard H. Geierstanger,[‡] David E. Wemmer,[||] Peter G. Schultz,[†,‡] Glen Spraggon,[‡] and Floyd E. Romesberg*[,†]

Contribution from the Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, Genomics Institute of the Novartis Research Foundation, 10675 John J. Hopkins Drive, San Diego, California 92121, Department of Chemistry, University of San Diego, 5998 Alcala Park, San Diego, California 92110, and Department of Chemistry, University of California, Berkeley, California 94720

Received April 1, 2007; E-mail: floyd@scripps.edu

Abstract: Expansion of the genetic alphabet has been a long-time goal of chemical biology. A third DNA base pair that is stable and replicable would have a great number of practical applications and would also lay the foundation for a semisynthetic organism. We have reported that DNA base pairs formed between deoxyribonucleotides with large aromatic, predominantly hydrophobic nucleobase analogues, such as propynylisocarbostyril (d**PICS**), are stable and efficiently synthesized by DNA polymerases. However, once incorporated into the primer, these analogues inhibit continued primer elongation. More recently, we have found that DNA base pairs formed between nucleobase analogues that have minimal aromatic surface area in addition to little or no hydrogen-bonding potential, such as 3-fluorobenzene (d**3FB**), are synthesized and extended by DNA polymerases with greatly increased efficiency. Here we show that the rate of synthesis and extension of the self-pair formed between two d**3FB** analogues is sufficient for in vitro DNA replication. To better understand the origins of efficient replication, we examined the structure of DNA duplexes containing either the d**3FB** or d**PICS** self-pairs. We find that the large aromatic rings of d**PICS** pair in an intercalative manner within duplex DNA, while the d**3FB** nucleobases interact in an edge-on manner, much closer in structure to natural base pairs. We also synthesized duplexes containing the 5-methyl-substituted derivatives of d**3FB** (d**5Me3FB**) paired opposite d**3FB** or the unsubstituted analogue (d**BEN**). In all, the data suggest that the structure, electrostatics, and dynamics can all contribute to the extension of unnatural primer termini. The results also help explain the replication properties of many previously examined unnatural base pairs and should help design unnatural base pairs that are better replicated.

Expansion of the genetic alphabet to include a third base pair would be a fundamental accomplishment that would not only have immediate utility for a number of applications, such as site-specific oligonucleotide labeling, but also lay the foundation for an organism with an expanded genetic code. Efforts toward this goal were first reported by Benner and co-workers,[1] who designed nucleobase analogues to pair on the basis of hydrogen-bonding (H-bonding) patterns that are complementary to each other, but not to any of the natural nucleobases. While these analogues have found practical applications and improvements continue to be reported, work from the Kool group has shown that H-bonds are not absolutely essential for polymerase-mediated base pair synthesis.[2−6] This work demonstrated that forces other than H-bonding could control polymerase-mediated base pair synthesis, and it has inspired a variety of novel nucleobase design strategies.

We,[7−18] and others,[19−30] have examined a large number of

[†] The Scripps Research Institute.
[‡] Genomics Institute of the Novartis Research Foundation.
[§] University of San Diego.
[||] University of California.

(1) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33−37.
(2) Morales, J. C.; Kool, E. T. *Nat. Struct. Biol.* **1998**, *5*, 950−954.
(3) Moran, S.; Ren, R. X.-F.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10506−10511.
(4) Moran, S.; Ren, R. X.-F.; Rumney, S. I.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056−2057.
(5) Kool, E. T. *Biopolymers* **1998**, *48*, 3−17.
(6) Kool, E. T. *Curr. Opin. Chem. Biol.* **2000**, *4*, 602−608.
(7) Tae, E. L.; Wu, Y. Q.; Xia, G.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2001**, *123*, 7439−7440.
(8) Ogawa, A. K.; Wu, Y.; McMinn, D. L.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 3274−3287.
(9) Ogawa, A. K.; Wu, Y.; Berger, M.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 8803−8804.
(10) McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **1999**, *121*, 11585−11586.
(11) Matsuda, S.; Romesberg, F. E. *J. Am. Chem. Soc.* **2004**, *126*, 14419−14427.
(12) Matsuda, S.; Henry, A. A.; Romesberg, F. E. *J. Am. Chem. Soc.* **2006**, *128*, 6369−6375.
(13) Leconte, A. M.; Matsuda, S.; Romesberg, F. E. *J. Am. Chem. Soc.* **2006**, *128*, 6780−6781.
(14) Leconte, A. M.; Matsuda, S.; Hwang, G. T.; Romesberg, F. E. *Angew. Chem., Int. Ed.* **2006**, *45*, 4326−4329.

10.1021/ja072276d CCC: $37.00

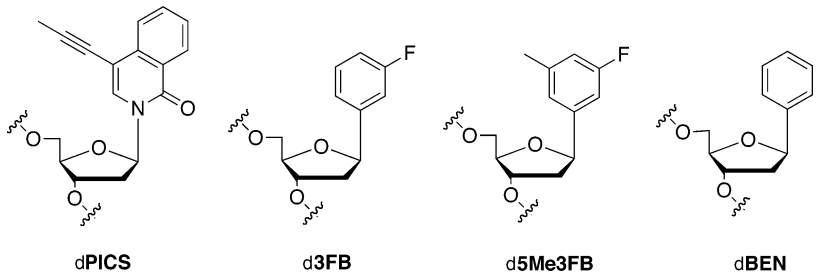


Figure 1. Unnatural nucleotides used in this study.

unnatural nucleotides bearing nucleobase analogues that pair on the basis of packing and hydrophobic interactions rather than H-bonding. (While many of the analogues are not actually basic, we refer to them as nucleobases for simplicity.) Packing and hydrophobic interactions have a well-documented role in protein folding, structure, and stability and should be inherently orthogonal to the H-bonding forces that mediate pairing of the natural base pairs. Indeed, DNA containing simple hydrophobic nucleobase analogues, such as benzene rings,[31] has been studied for years, and more recent studies have shown that hydrophobic nucleobase analogues may be incorporated into duplex DNA without significant structural distortions.[32,33] Our original efforts focused primarily on nucleobase analogues derived from relatively large bicyclic scaffolds with an extended aromatic surface area. These analogues were designed to preserve duplex stability in the absence of interstrand H-bonding, by increasing intrastrand packing. We identified several heteropairs[11−14,34] (i.e., formed by pairing two different analogues) and self-pairs[10−12,17] (i.e., formed by pairing two identical analogues) that are stable in duplex DNA and efficiently synthesized by polymerase-catalyzed insertion of the unnatural triphosphate opposite its cognate base in a DNA template. For example, the propynyl-isocarbostyril (d**PICS**) self-pair (Figure 1) is stable in duplex DNA and also synthesized (by insertion of the triphosphate opposite the analogue in the template) by the exonuclease-deficient Klenow fragment of *Escherichia coli* DNA polymerase I (Kf) with reasonable efficiency and selectivity.[10] These results demonstrated that packing and hydrophobicity are sufficient to mediate duplex stability and unnatural base pair synthesis.

While many of the unnatural base pairs formed between the relatively large nucleobase analogues, such as d**PICS**, are stable and efficiently synthesized, continued primer extension is very inefficient in all cases examined, regardless of the DNA polymerase employed. Thus, the determinants of base pair stability and synthesis are different from those of extension. Rate-limiting extension has also been observed during the replication of other unnatural base pairs. We have speculated that while a large aromatic surface area may stabilize base pairing, as well as the transition state for unnatural nucleotide insertion, it may also result in a structure at the primer terminus that is poorly recognized by DNA polymerases. To probe the role of the nucleobase aromatic surface area, we designed and evaluated nucleoside analogues bearing a wide variety of simple, derivatized phenyl rings.[11,12,15−17] The shape, hydrophobicity, and electronic properties of the nucleobase analogues were systematically varied by derivatization with fluoro, bromo, cyano, and/or methyl substituents. Surprisingly, we found that a large aromatic surface area is not required for stable and selective pairing within the duplex nor for the relatively efficient synthesis of the unnatural base pairs. More importantly, while we observed that extension remained inefficient for the majority of the pairs, several were extended with a significantly higher efficiency. Most notably, 3-fluorobenzene (d**3FB**) (Figure 1) is incorporated opposite itself and then correctly extended by Kf with a rate that is within 100-fold of a natural base pair.[17] This result is remarkable considering that the d**3FB** self-pair lacks both H-bonds and an extended aromatic surface area, which are thought to underlie the stability and replication of natural DNA.

In this work, we first demonstrate that the synthesis and extension efficiencies of d**3FB** are sufficient for Kf to synthesize long strands of DNA containing the self-pair. Then, to better understand why the d**3FB** self-pair is efficiently synthesized and extended, while similar pairs, or pairs formed between the larger nucleobase analogues are not, we examine the structure of DNA duplexes containing either the d**3FB** or d**PICS** self-pair. In addition, the role of hydrophobic packing at the interface between the nucleobase analogues was further examined by comparing the replication and structure of the d**3FB** self-pair with that of a heteropair formed between d**3FB** and its 5-methyl-3-fluoro derivative (d**5Me3FB**) or between d**5Me3FB** and a simple phenyl nucleotide (d**BEN**). The data suggest that the structure, electrostatics, and dynamics each contribute to efficient unnatural base pair replication.

(15) Kim, Y.; Leconte, A. M.; Hari, Y.; Romesberg, F. E. *Angew. Chem., Int. Ed.* **2006**, *45*, 7809−7812.
(16) Hwang, G. T.; Romesberg, F. E. *Nucleic Acids Res.* **2006**, *34*, 2037−2045.
(17) Henry, A. A.; Olsen, A. G.; Matsuda, S.; Yu, C.; Geierstanger, B. H.; Romesberg, F. E. *J. Am. Chem. Soc.* **2004**, *126*, 6923−6931.
(18) Henry, A. A.; Romesberg, F. E. *Curr. Opin. Chem. Biol.* **2003**, *7*, 727−733.
(19) Ishikawa, M.; Hirao, I.; Yokoyama, S. *Tetrahedron Lett.* **2000**, *41*, 3931−3934.
(20) Hirao, I.; Harada, Y.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Yokoyama, S. *J. Am. Chem. Soc.* **2004**, *126*, 13298−13305.
(21) Mitsui, T.; Kimoto, M.; Harada, Y.; Yokoyama, S.; Hirao, I. *J. Am. Chem. Soc.* **2005**, *127*, 8652−8658.
(22) Mitsui, T.; Kitamura, A.; Kimoto, M.; To, T.; Sato, A.; Hirao, I.; Yokoyama, S. *J. Am. Chem. Soc.* **2003**, *125*, 5298−5307.
(23) Hirao, I.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Kawai, R.; Sato, A.; Harada, Y.; Yokoyama, S. *Nat. Methods* **2006**, *3*, 729−735.
(24) Kincaid, K.; Beckman, J.; Zivkovic, A.; Halcomb, R. L.; Engels, J. W.; Kuchta, R. D. *Nucleic Acids Res.* **2005**, *33*, 2620−2628.
(25) Chiaramonte, M.; Moore, C. L.; Kincaid, K.; Kuchta, R. D. *Biochemistry* **2003**, *42*, 10472−10481.
(26) Hirao, I. *Curr. Opin. Chem. Biol.* **2006**, *10*, 622−627.
(27) Matray, T. J.; Kool, E. T. *Nature* **1999**, *399*, 704−708.
(28) Zhang, X.; Lee, I.; Berdis, A. J. *Biochemistry* **2005**, *44*, 13101−13110.
(29) Zhang, X.; Lee, I.; Zhou, X.; Berdis, A. J. *J. Am. Chem. Soc.* **2006**, *128*, 143−149.
(30) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **1999**, *121*, 2323−2324.
(31) Millican, T. A.; Mock, G. A.; Chauncey, M. A.; Patel, T. P.; Eaton, M. A. W.; Gunning, J.; Cutbush, S. D.; Neidle, S.; Mann, J. *Nucleic Acids Res.* **1984**, *12*, 7435−7453.
(32) Guckian, K. M.; Krugh, T. R.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 6841−6847.
(33) Smirnov, S.; Matray, T. J.; Kool, E. T.; de los Santos, C. *Nucleic Acids Res.* **2002**, *30*, 5561−5569.
(34) Wu, Y. Q.; Ogawa, A. K.; Berger, M.; McMinn, D. L.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 7621−7632.

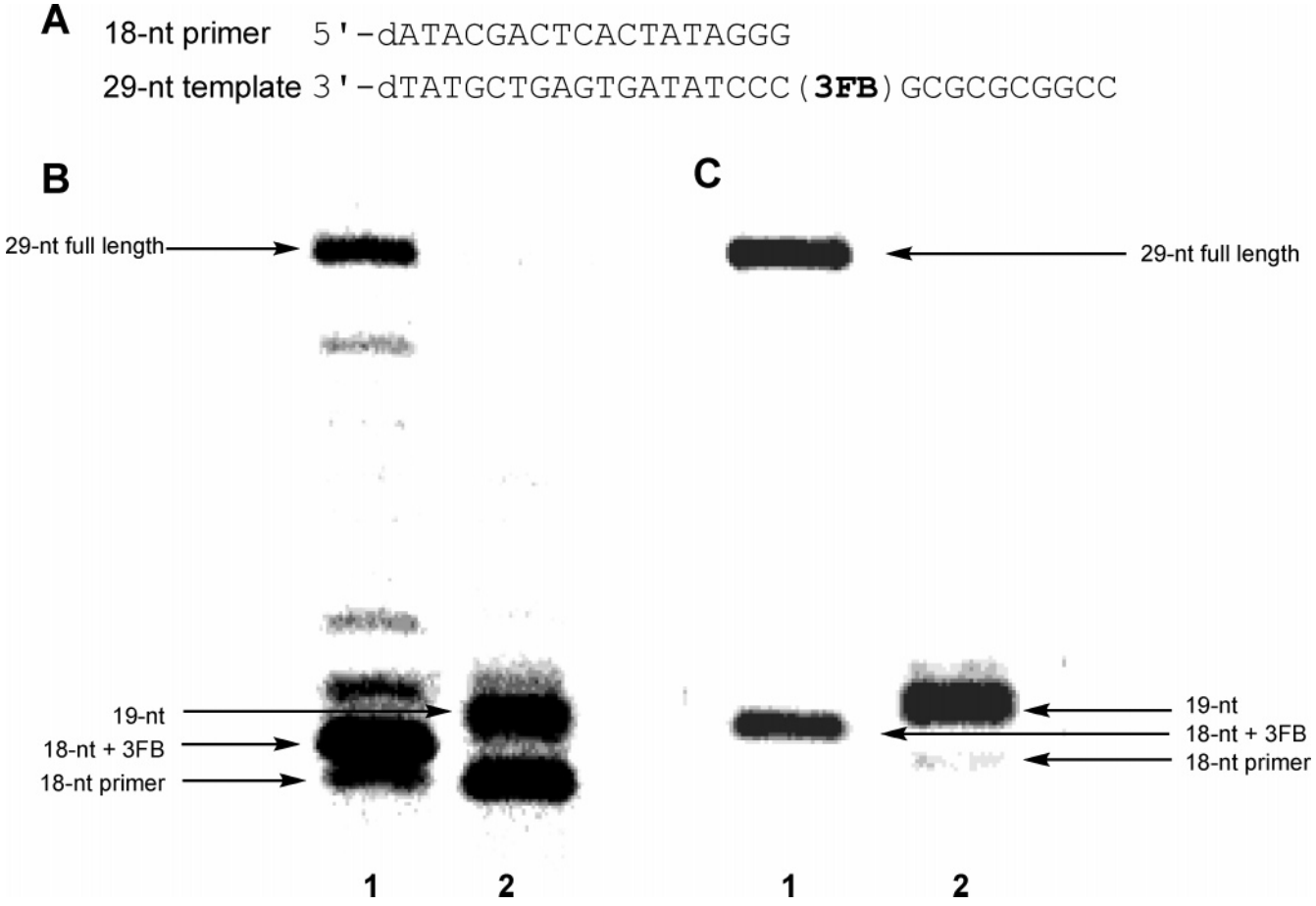


Figure 2. (A) Sequence of the 18-nt primer and 29-nt template used for full-length synthesis. (B) Full-length synthesis by KF exo$^-$: 40 nM primer−template, 1.23 nM enzyme, 50 mM Tris−HCl, pH 7.5, 1 mM DTT, 50 μg/mL acetylated BSA, 2 h at 25 °C, 5 mM $MgCl_2$, 1 mM $MnCl_2$, and 50 μM concentration each of dCTP and dGTP, with (lane 1) or without (lane 2) 200 μM d(**3FB**)TP. (C) Full-length synthesis by KF exo$^+$: 40 nM primer−template, 1.23 nM enzyme, 50 mM Tris−HCl, pH 7.5, 1 mM DTT, 50 μg/mL acetylated BSA, 2 h at 25 °C, 10 mM $MgCl_2$, 1 mM $MnCl_2$, and 200 μM dCTP/dGTP, with (lane 1) or without (lane 2) 200 μM d(**3FB**)TP.

Results

Previously we demonstrated that the d**3FB** self-pair is efficiently synthesized by Kf and then extended by the addition of a single dCTP opposite dG in the template.[17] As discussed above, the efficient extension of the d**3FB** self-pair, relative to others that have been examined, including the d**PICS** self-pair, suggests that the d**3FB** self-pair forms a more natural-like primer terminus. To determine whether the d**3FB** self-pair is compatible with efficient full-length synthesis, we examined the ability of Kf to synthesize a full-length strand of DNA that contained d**3FB**, as well as an additional 10 nucleotides (Figure 2). d**3FB** has high selectivity against misincorporation of guanine and cytosine, but only modest selectivity against adenine and thymine; thus, the template contained only dG and dC nucleotides, and only the triphosphates dCTP, dGTP, and d(**3FB**)TP were added to the reaction mixture. Under these conditions, full-length product is efficiently produced, demonstrating that the d**3FB** self-pair does not interfere with the addition of downstream nucleotides to the growing primer strand. By comparison, attempts to replicate the d**PICS** self-pair in the same sequence context resulted in complete termination of primer elongation after insertion of the unnatural triphosphate.

To identify possible determinants of efficient replication and, in particular, why the d**3FB** self-pair is both efficiently synthesized and extended, while the d**PICS** self-pair is efficiently synthesized but not extended, we examined the structure of duplex DNA containing these unnatural base pairs. The structure of DNA containing a d**PICS** self-pair was determined in the duplex d($C_1G_2T_3T_4T_5C_6$**PICS**$_7T_8T_9C_{10}T_{11}C_{12}$): d($G_{13}A_{14}G_{15}A_{16}A_{17}$**PICS**$_{18}G_{19}A_{20}A_{21}A_{22}C_{23}G_{24}$) using NMR and restrained molecular modeling (Figure 3). Overall, the duplex assumes a canonical B-form structure and is not substantially distorted relative to natural DNA[35] except at the site of the d**PICS** self-pair. The d**PICS** propynyl groups are oriented into the major groove, while the carbonyl groups are positioned in the minor groove. This conformation, which is referred to as *anti* by analogy to natural nucleotides, likely minimizes repulsive electrostatic interactions between the nucleobase carbonyl oxygen and the oxygen and/or 5′-carbon of the ribose ring. The most notable feature of the structure is that the d**PICS** nucleotides interact through interstrand stacking: each nucleobase analogue intercalates between the other and its flanking base pair, rather than pairing in an edge-on manner as observed with natural Watson−Crick base pairs. The positions of the bases were verified by NOE contacts between the edge of both d**PICS** nucleobases and the sugar of the opposite strand and demonstrate that only one of the two possible intercalated structures is formed. The preference for

(35) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

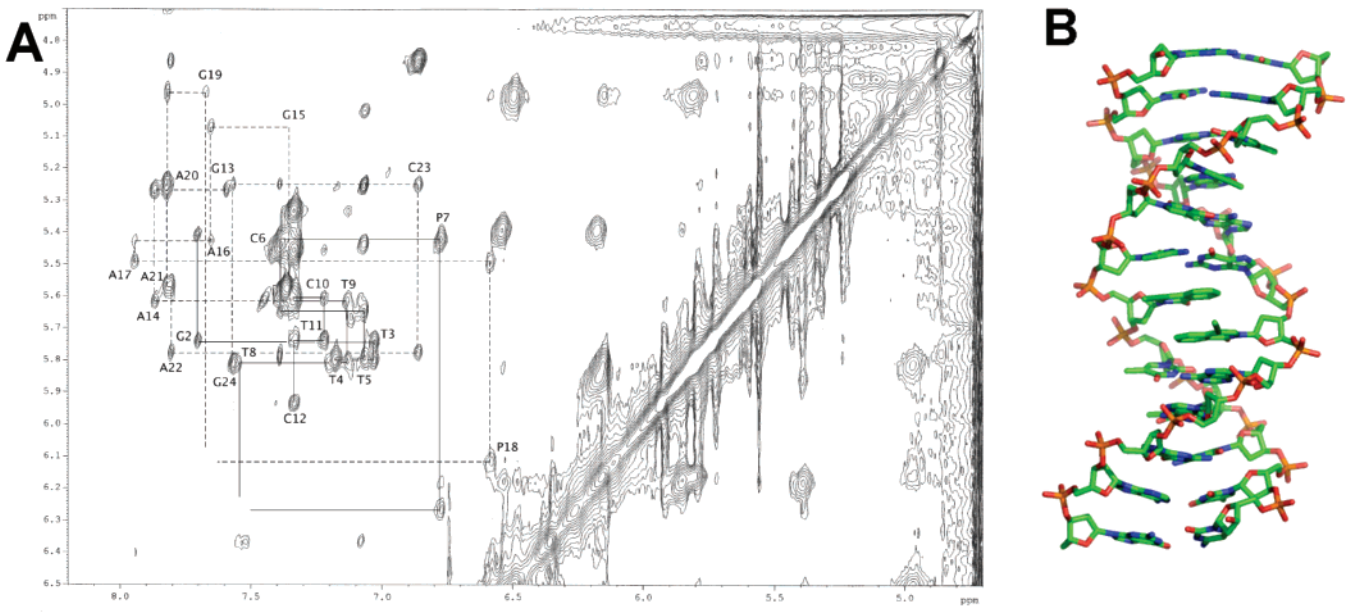


Figure 3. Structural characterization of the d**PICS** self-pair in the duplex d($C_1G_2T_3T_4T_5C_6$**PICS**$_7T_8T_9C_{10}T_{11}C_{12}$):d($G_{13}A_{14}G_{15}A_{16}A_{17}$**PICS**$_{18}G_{19}A_{20}A_{21}A_{22}C_{23}G_{24}$). (A) Aromatic to H1′ region of a 200 ms NOESY spectrum. (B) Model of the duplex. The d**PICS** self-pair is shown in the center of the duplex.

the observed structure likely results from optimized packing interactions between the d**PICS** nucleobase and the flanking natural nucleobases. While we observed breaks in the normal NOE connectivities between the sugar and the flanking natural nucleobase at both d**PICS**$_{18}$dG_{19} and d**PICS**$_7$dT_8, the imino proton resonances were observed for all Watson−Crick pairs, except the terminal dG:dC nucleotides, indicating that the hydrogen bonds of the flanking DNA remain intact. The intrastrand phosphate distances are approximately 7.0 Å, which is typical of B-form DNA. The O4′−C1′−N1−C2 dihedral angles of the d**PICS** nucleotides are approximately −135°, also in the range for B-form DNA. The torsion angles about the exocyclic C4′−C5′ bonds, which position the 5′-phosphate group relative to the sugar and the nucleobase, are 60−70° for d**PICS**$_{18}$ and the flanking natural nucleotides in the purine-rich strand. The same torsion angle is approximately 45° for d**PICS**$_7$ in the pyrimidine-rich strand and approximately 70° for its flanking nucleotides, similar to that found with natural B-DNA. However, the neighboring O5′−C5′ torsion angles for d**PICS**$_{18}$ and d**PICS**$_7$ are approximately 100°, while normal B-DNA values are approximately 170°. These structural readjustments appear to help accommodate unnatural base pair intercalation. In the model, intrastrand packing appears to be better optimized within the purine-rich strand than the pyrimidine-rich strand, which is distorted by the buckling at the self-pair.

The structures of two different duplexes containing the d**3FB** self-pair were determined, one using X-ray crystallography and the other using NMR spectroscopy (Figure 4). Figure 4A shows the central section of the DNA duplex d($C_1G_2{}^{Br}C_3G_4A_5A_6$-**3FB**$_7T_8T_9C_{10}G_{11}C_{12}G_{13}$)$_2$ containing a single d**3FB** self-pair as determined by X-ray crystallography with a resolution of 2.8 Å (Table 1 and the Supporting Information). Six copies of the duplex are present in the crystallographic asymmetric unit. Four copies are well ordered (chains A−H) and well defined by the electron density. The remaining two copies are less well ordered (chains I−L) and characterized by diffuse electron density; however, the density was successfully fit using the bromine atoms in the Patterson maps (Supporting Information). Analysis of the four ordered duplexes using the 3DNA package[36] revealed a right-handed B-form DNA conformation with a mean helix diameter of 19.9 Å, consistent with standard Watson−Crick base pairing (except at the 5′ and 3′ ends of the duplex, where the nucleobases intercalate into neighboring duplexes in the crystal to form two sets of semicontinuous helices). The root-mean-square deviation between the duplexes is 1.50 Å for backbone atoms and 0.84 Å for nucleobase atoms. The root-mean-square deviations between the average duplex and an ideal B-form duplex are 1.26 and 0.58 Å for sugar−phosphate backbone and nucleobase atoms, respectively. The d**3FB** nucleobases are oriented so that their fluorine atoms are positioned in the major groove of the duplex, separated by 9.8 Å (Figure 4B). At their closest approach, the nucleobase analogues are separated by an average carbon to carbon distance of 3.75 Å. This is slightly greater than the sum of the van der Waals radii (3.4 Å), which suggests that the nucleobase analogues are not optimally edge-to-edge packed. The unnatural base pairs adopt an average propeller twist of −12°, which is virtually identical to that of canonical B-form DNA. The mean distance between the d**3FB** nucleobase analogue and the flanking natural nucleobases is 3.2 Å, which suggests that the analogues pack favorably with their flanking natural nucleobases. In fact, the only significant deviation from an ideal duplex geometry appears to be due to these stacking interactions, as the flanking natural nucleobases tilt to achieve optimal coplanarity with the unnatural nucleobases (Figure 4B).

Characterization of d($C_1G_2C_3$**3FB**$_4A_5A_6T_7T_8$**3FB**$_9G_{10}C_{11}G_{12}$)$_2$ by NMR spectroscopy and NOE-restrained MD simulations (Supporting Information) also indicates a canonical B-form DNA duplex as demonstrated by characteristic NOE connectivities and intensities (Figure 4C). The base−sugar connectivities along each strand are not interrupted at the d**3FB** self-pair (Figure 4C), unlike the connectivities observed for the d**PICS** self-pair which show clear breaks between the d**PICS** sugars and their 3′ neighbor base protons. In addition, imino to

(36) Lu, X.-J.; Olson, W. K. *Nucleic Acids Res.* **2003**, *31*, 5108−5121.

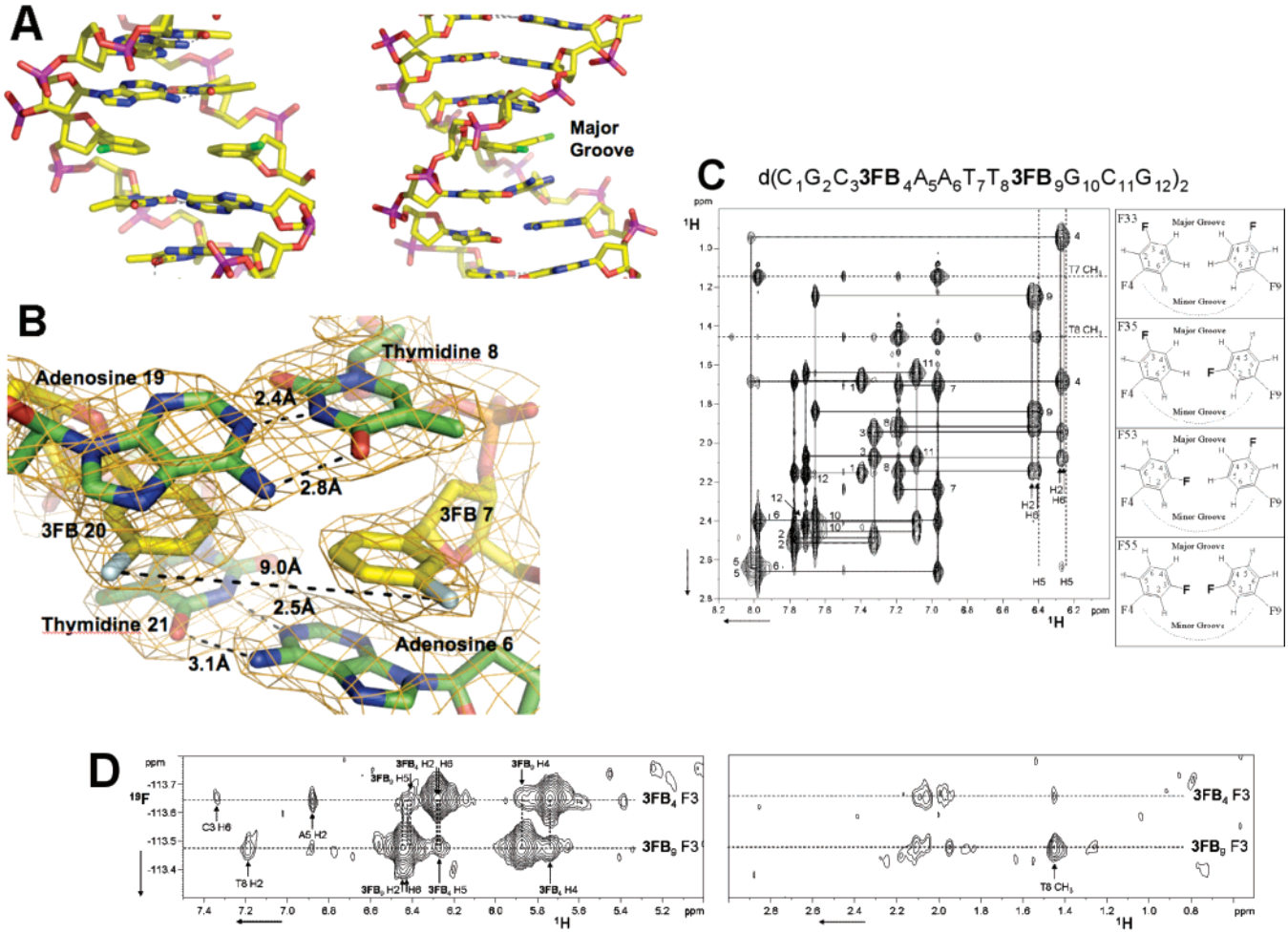


Figure 4. Structural characterization of the d**3FB** self-pair. (A) X-ray structure of the DNA duplex d($C_1G_2{}^{Br}C_3G_4A_5A_6$**3FB**$_7T_8T_9C_{10}G_{11}C_{12}G_{13}$)$_2$ at 2.8 Å resolution (PDB ID 2PIS). The d**3FB** self-pair is shown with flanking base pairs. Carbon atoms are shown in yellow, nitrogen in blue, oxygen in orange, phosphorus in magenta, and fluorine in green. (B) $2F_o - F_c$ electron density map contoured at 1.2σ around the model of the d**3FB** self-pair and its adjacent residues. The fluorine atoms are oriented into the major groove and are colored light cyan. The figure was produced by Pymol (www.pymol.org). (C) NMR characterization of the DNA duplex d($C_1G_2C_3$**3FB**$_4A_5A_6T_7T_8$**3FB**$_9G_{10}C_{11}G_{12}$)$_2$. Shown is the aromatic to H2′/H2″ region of a 2D NOESY experiment with the NOE walk indicated by solid lines. Numbers indicate the aromatic to H2′ and H2″ intranucleotide peaks. Both the H2 and H6 protons of d**3FB**$_9$ and of d**3FB**$_4$ show cross-peaks with the H2′ and H2″ protons of the preceding residue, indicating rapid flipping of the d**3FB** rings. The four possible ring orientations of d**3FB** used to model the NOE data are shown on the right. (D) ^{1}H−^{19}F NOESY spectrum of d($C_1G_2C_3$**3FB**$_4A_5A_6T_7T_8$**3FB**$_9G_{10}C_{11}G_{12}$)$_2$ acquired as described previously.[46] The fluorine of d**3FB**$_9$ resonates at −113.48 ppm, while the chemical shift of the fluorine of d**3FB**$_4$ is −113.64 ppm. ^{19}F−^{1}H NOE peaks are labeled according to (C). Experimental details and spectra are available in the Supporting Information.

imino and imino to adenine H2 NOE connectivities are observed throughout the DNA helix except at the terminal base pair (data not shown). The similarity of all proton chemical shifts to those reported for the fully natural DNA duplex (containing dG_4 and dC_9, Table S1 and Figure S1 in the Supporting Information)[37] further demonstrates that both d**3FB** self-pairs are accommodated within the double helix without substantial structural distortions.

Interestingly, the NMR data indicate that the d**3FB** self-pair adopts multiple conformations that are related by simple ring flips about each C-glycosidic linkage (Figure 4C,D). This heterogeneity is demonstrated by NOE cross-peaks from d**3FB**$_4$ H2 and H6 to dC_3 H2′/H2″ and from d**3FB**$_9$ H2 and H6 to T_8 H2′/H2″ (Figure 4C), as well as by heteronuclear NOEs between the fluorine of d**3FB**$_4$ and both dA_5 H2 and dC_3 H6, which are mutually exclusive for a single ring orientation (Figure 4D). An NOE between the d**3FB**$_9$ fluorine and both dT_8 CH_3 and dT_8 H6, as well as between the H4 and H5 protons of d**3FB**$_4$, further demonstrates that both nucleobase analogues undergo rapid ring flipping. Since single resonance lines are observed for the two fluorine atoms and for each aromatic proton, the rate of base flipping must be fast on the chemical shift time scale for these resonances, i.e., exchange lifetimes on the submillisecond time scale. This ring flipping model is also supported by NOE-restrained MD simulations. When all 662 NOE restraints are applied, the d**3FB** self-pair adopts conformations with large propeller twists, causing distortions of the neighboring base pairs and high restraint violation energies (Table 2, Figures S2 and S3 in the Supporting Information). Grouping NOEs into four separate sets, corresponding to each of the four possible self-pair conformations, eliminated these violations (Tables 3 and S3 in the Supporting Information) and distortions (Figure S4 in the Supporting Information), suggesting that they arise from a superposition of signals corresponding to multiple conformations present in solution. The predicted energy of the conformation with both fluorine atoms oriented into the

(37) Hare, D. R.; Wemmer, D. E.; Chou, S. H.; Drobny, G.; Reid, B. R. *J. Mol. Biol.* **1983**, *171*, 319−336.

Table 1. Data Collection, Phasing, and Refinement Statistics

	remote	*f*′	*f*″
space group	$I4_122$		
unit cell params (Å)	$a = b = 146.00$, $c = 93.21$		
wavelength (Å)	0.9050	0.92017	0.92030
resolution range (Å)	50.0−2.8	50.0−2.8	50.0−2.8
R_{sym} (in highest resolution shell)	0.105(0.747)	0.089 (0.59)	0.099 (0.706)
no. of unique reflns (obsd)	12331 (186023)	12382 (190046)	12271 (92360)
completeness (%) (highest shell)	97.7(87.5)	97.8(88.5)	97.1 (84.1)
highest resolution shell (Å)	2.9−2.8		
mean $I/\sigma(I)$	25.2 (2.9)	19.6(1.5)	12.0(3.1)
	Phasing Statistics		
no. of Br sites	10		
mean figure of merit	0.42 (0.19)		
	Model and Refinement Statistics		
no. of reflns (total)	11082		
no. of reflns (test)	565		
R_{cryst} (R_{free})[a,b]	23.1,(30.8)		
no. of nucleic acid atoms	2474		
no. of heteroatoms	678		
	Stereochemical Parameters		
rmsd(bonds) (Å)	0.014		
rmsd(angles) (deg)	1.742		
av isotropic *B* value (Å^2)	56.75		
ESU[c] based on R_{free} (Å)	0.483		

[a] $R_{cryst} = \sum|F_o - F_c|/\sum|F_c|$, where F_o and F_c are observed and calculated structure factors, respectively. R_{free} was calculated from a test set (5%) omitted from the refinement. [b] $R_{factor} = \sum|I_i - \langle I_i\rangle|/\sum|I_i|$, where I_i is the scaled intensity of the *i*th measurement and $\langle I_i\rangle$ is the mean intensity for that reflection. [c] Estimated overall coordinate error.[44,45]

Table 2. NOE-Restrained MD Calculations with d($C_1G_2C_3$**3FB**$_4A_5A_6T_7T_8$**3FB**$_9G_{10}C_{11}G_{12}$)$_2$ Using All Restraints[a]

self-pair conformation[b]	no. of structures	rmsd	rmsd from mean structure	av *E*(viol)	av *E*(AMBER)	percentage of *E*(AMBER) due to *E*(viol)	total no. of restraints	no. of restraints w/av viol >0.2 Å
F33	22	1.80	1.24	156.0	−5184	3.0	662	26
F35	17	2.06	1.41	197.9	−5146	3.8	662	21
F53	22	1.88	1.30	170.2	−5160	3.3	662	21
F55	21	1.93	1.34	183.3	−5122	3.6	662	16

[a] Identical restraints were applied to the four different families of d**3FB** conformers (Figure 4C). [b] For a definition of d**3FB** self-pair conformations, see Figure 4C.

Table 3. NOE-Restrained MD Calculations with d($C_1G_2C_3$**3FB**$_4A_5A_6T_7T_8$**3FB**$_9G_{10}C_{11}G_{12}$)$_2$ after Organization of the Restraints into Four Sets Consistent with the Four Different Starting Conformations (Figure 4C)

self-pair conformation[a]	no. of structures	rmsd	rmsd from mean structure	av *E*(viol)	av *E*(AMBER)	percentage of *E*(AMBER) due to *E*(viol)	total no. of restraints	no. of restraints w/av viol >0.2 Å
F33	18	1.97	1.35	19.87	−5307	0.37	608	0
F35	22	2.04	1.41	28.19	−5261	0.54	610	0
F53	20	1.93	1.33	21.79	−5275	0.41	618	0
F55	17	1.92	1.32	17.87	−5242	0.34	616	0

[a] For a definition of d**3FB** self-pair conformations, see Figure 4C.

duplex, as well as the absence of an observable fluorine−fluorine NOE cross-peak (Figure S5 in the Supporting Information), suggests this conformation is not significantly populated. However, the NMR data suggest that all three other possible self-pair conformations are populated (Figure 4C).

Spectral congestion precludes an accurate assessment of the population of the three self-pair conformations; however, the d**3FB** self-pair conformation, in which both fluorine atoms are positioned in the major groove, as seen in the X-ray structure, yields the lowest energy structure in the NOE-restrained MD simulations (Table 3), suggesting that it may also be the most populated in solution. To approximate the actual populations of each self-pair conformation, we examined the NMR spectra of the d**3FB** in the d**5Me3FB** heteropair (see below). In contrast to the self-pair, the signals of the d**3FB**:d**5Me3FB** heteropair are more resolved and the d**3FB** nucleotide appears to adopt the same conformations. The intensity of the intranucleotide aromatic to H2′/H2″ NOE peaks of the d**3FB** analogue of the heteropair suggests that the fluorine atom is oriented into the major groove approximately 70% of the time, while the remainder of the time it is oriented into the interior of the duplex. If the same ratio is assumed for both nucleobases of the self-pair, then one can conclude that the conformation with both fluorine atoms disposed in the major groove is populated approximately 40% of the time, while the two conformations where one fluorine atom is positioned in the major groove and the other within the duplex are each populated approximately 30% of the time (Figure 4C). These populations are consistent with the observed X-ray structure. In addition to rapid ring flipping, additional self-pair motion is suggested by line broadening of intranucleotide NOE cross-peaks of the flanking nucleobases dA_5 and dG_{10} and dA_5 imino resonance broadening, which may result from enhanced solvent exchange, possibly

due to increased solvent accessibility or base pair fluctuations.[38] Such fluctuations may be coupled to the observed ring flipping of the d**3FB** analogues.

Taken together, the X-ray and NMR data indicate that the d**3FB** self-pair does not perturb the structure of a B-form DNA duplex and that the nucleobase analogues are coplanar and well packed with their flanking natural nucleobases. However, while the d**3FB** self-pair is significantly more stable than the mispairs of the natural bases,[17] it is somewhat less stable than pairs formed between more highly substituted analogues, and the structures suggest that d**3FB** is not optimized for edge-on packing within the self-pair. Less than optimal packing may underlie the self-pair dynamics observed in the NMR experiments. To explore the effect of increased interbase packing, we synthesized and examined an analogue with a methyl group at position 5 of the benzene ring, d**5Me3FB** (Figure 1). This base analogue was incorporated into the duplex d($C_1G_2C_3$-**5Me3FB**$_4$$A_5A_6T_7T_8$**3FB**$_9$$G_{10}C_{11}G_{12}$)$_2$, which pairs d**5Me3FB** opposite d**3FB**. NOE analysis unambiguously demonstrates that the d**5Me3FB** analogue does not undergo base flipping and that its methyl group is positioned in the duplex where it packs against the pairing d**3FB** (Figure S6 in the Supporting Information). Consistent with a more optimally packed interface, the UV melting temperature of the duplex with the d**3FB**:d**5Me3FB** heteropair was ~2 °C higher than that of the analogous duplex containing the d**3FB** self-pair (Supporting Information). However, we cannot exclude the possibility that packing interactions between the methyl group and flanking nucleobases also contribute to the increased stability. While the d**5Me3FB** nucleobase is less dynamic, rotation of the d**3FB** nucleobase of the heteropair remains fast on the chemical shift time scale, and line broadening of the flanking nucleotides dA_5 and dG_{10} is observed, as in the d**3FB** self-pair. Similar observations were made when d**5Me3FB** was paired opposite **dBEN** (Figure 1) in the same sequence context (Figure S7 in the Supporting Information). While d**5Me3FB** is locked into one orientation, the phenyl ring undergoes rapid flipping as indicated by the observation of only three aromatic resonances rather than the five expected for a static nucleobase analogue in the magnetically nonsymmetric environment of a DNA duplex. These data indicate that the addition of the methyl group restrains the modified nucleobase to a single, well-defined conformation but that the pairing nucleobase remains dynamic. The data also confirm that sufficient space is available for at least a single fluorine within the interface between the nucleobase analogues and suggests that the positioning of the fluorine atoms of the d**3FB** self-pair must result from other interactions, such as electrostatic or packing interactions with the flanking nucleobases.

To examine how a better packed nucleobase interface impacts DNA synthesis, we examined the ability of Kf to extend a primer terminating with d**5Me3FB** (paired with d**3FB** in the template) or terminating with d**3FB** (paired with d**5Me3FB** in the template). With d**5Me3FB** at the primer terminus, no extension product was observed ($k_{cat}/K_M < 1 \times 10^3$ M^{-1} min^{-1}). In contrast, with d**3FB** at the primer terminus opposite d**5Me3FB**, the extension proceeded with a second-order rate constant of 1.6×10^4 M^{-1} min^{-1}, which is only 30-fold reduced relative to that of the d**3FB** self-pair. These data suggest that the most efficient extension results when d**3FB** at the primer terminus is free to adopt the conformation with its fluorine oriented into the duplex, as opposed to into the major groove.

Discussion

The expansion of the genetic alphabet requires an unnatural base pair that is efficiently replicated by DNA polymerases. The unnatural nucleoside triphosphate must be efficiently and selectively inserted opposite its partner in the template, and the resulting terminus must be efficiently extended by incorporation of the next correct nucleotide. Generally, we have found that hydrophobic packing interactions between large aromatic nucleobase analogues, such as the isocarbostyril group of d**PICS**, give rise to stable base pairs that are synthesized with reasonable efficiency, but not extended by further primer elongation. In contrast, the d**3FB** self-pair is both synthesized and extended by Kf.

To understand the basis for the differences in behavior between the d**PICS** and d**3FB** self-pairs, we characterized duplex DNA containing these unnatural base pairs. While both are accommodated within duplex DNA without significant loss of duplex stability, the large aromatic isocarbostyril rings of the d**PICS** pair in an intercalative, stacked manner. The intercalative mode of interaction within the d**PICS** self-pair is consistent with its high stability as well as its efficient synthesis, as significant hydrophobic packing is expected to be manifest in the developing transition state for d(**PICS**)TP insertion. However, it is also likely that an intercalated structure at the primer terminus results in the mispositioning of the 3′-OH in the enzyme active site, which may explain why the d**PICS** self-pair is not efficiently extended. An intercalative mode of interaction has also been predicted by computational analysis of a related self-pair[39] and has been observed with other nucleotides that have large aromatic nucleobase analogues.[40−43] Thus, stability, efficient synthesis, and poor extension are likely to be general features of unnatural base pairs formed between nucleotides with large aromatic nucleobase analogues, due to intercalative pairing.

In contrast to the d**PICS** self-pair, the nucleobase analogues of the d**3FB** self-pair are not sufficiently large to bridge the duplex and cannot intercalate. Instead, they interact in an edge-on manner similar to that of natural nucleobases. The d**3FB** self-pair stability and synthesis are likely mediated by hydrophobic forces and may have contributions from intrastrand electrostatic interactions.

While a more natural terminal base pair structure is likely to be important for the extension of the d**3FB** self-pair, other factors must also contribute, as a wide variety of similar self-pairs are not efficiently extended. We have altered the number,

(38) Guckian, K. M.; Krugh, T. R.; Kool, E. T. *Nat. Struct. Biol.* **1998**, *5*, 954−959.
(39) Reha, D.; Hocek, M.; Hobza, P. *Chem.—Eur. J.* **2006**, *12*, 3587−3595.
(40) Zahn, A.; Brotschi, C.; Leumann, C. J. *Chem.—Eur. J.* **2005**, *11*, 2125−2129.
(41) Brotschi, C.; Mathis, G.; Leumann, C. J. *Chem.—Eur. J.* **2005**, *11*, 1911−1923.
(42) Brotschi, C.; Haberli, A.; Leumann, C. J. *Angew. Chem., Int. Ed.* **2001**, *40*, 3012−3014.
(43) Brotschi, C.; Leumann, C. J. *Angew. Chem., Int. Ed.* **2003**, *42*, 1655−1658.
(44) Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307−326.
(45) Tickle, I. J.; Laskowski, R. A.; Moss, D. S. *Acta Crystallogr.* **1998**, *D54*, 243−252.
(46) Scott, L. G.; Geierstanger, B. H.; Williamson, J. R.; Hennig, M. *J. Am. Chem. Soc.* **2004**, *126*, 11776−11777.

position, and nature of the substituents of the phenyl ring scaffold and find that such modifications have a significant and generally detrimental impact on both synthesis and extension efficiencies. To date, the specific *m*-fluoro substitution pattern of d**3FB** is unique in its ability to facilitate replication. Similarly positioned methyl, bromo, and cyano substituents or alternate fluoro substituent patterns result in self-pairs with significantly reduced extension efficiencies.[12,16,17] Moreover, we have shown here that while addition of a methyl group to the *meta* position of d**3FB** stabilizes pairing, it also prevents replication. Specifically, when the analogue in the DNA template is modified with the methyl group, it has only a small effect on the efficient extension; however, when the analogue at the primer terminus is modified, extension is undetectable. Along with the NMR data, which suggest that the addition of the methyl group localizes the nucleobase to the conformation with the fluorine atom in the major groove, these data also suggest that efficient extension occurs only when the d**3FB** nucleotide at the primer terminus rotates to position its fluorine atom into the interbase interface. Thus, in addition to a natural-like primer terminus structure, base pair dynamics and dipole interactions appear to be critical for unnatural base pair extension.

Despite a complete absence of H-bonding and modest shape complementarity, the d**3FB** self-pair is reasonably stable and replicable. This argues that precise shape complementarity and H-bonding are not unique in their ability to control the specific interbase interactions required for DNA stability and replication. The data suggest that replication may be optimized within base pairs with coplanar nucleobase analogues by careful manipulation of the shape and electrostatic properties. The results also emphasize that overly stabilized unnatural base pairs might be localized to structures that are not well recognized at a primer terminus; in the absence of perfectly designed nucleobases flexibility may be advantageous for replication. These results, along with those of recent studies demonstrating that extension may also be facilitated by the addition of minor groove H-bond acceptors, help explain the replication properties of a large number of previously reported unnatural base pairs and should help design pairs that are more efficiently replicated.

Acknowledgment. Funding was provided by the National Institutes of Health (Grant GM60005 to F.E.R.).

Supporting Information Available: Kinetic, thermodynamic, NMR, and X-ray crystallography methods and supporting figures and tables. This material is available free of charge via the Internet at http://pubs.acs.org.

JA072276D