

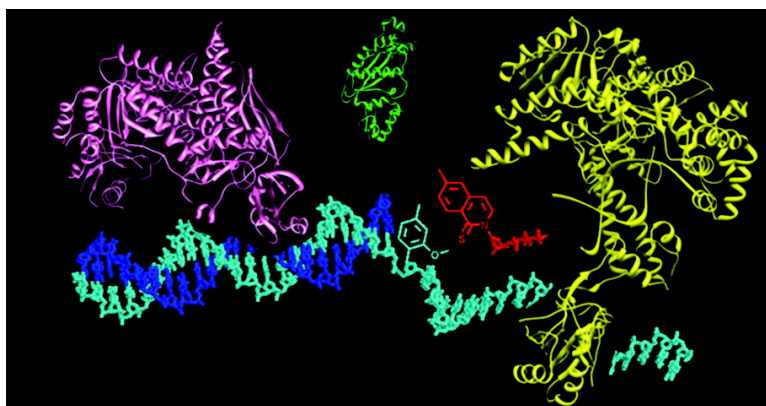
Article

Unnatural Substrate Repertoire of A, B, and X Family DNA Polymerases

Gil Tae Hwang, and Floyd E. Romesberg

J. Am. Chem. Soc., **2008**, 130 (44), 14872-14882 • DOI: 10.1021/ja803833h • Publication Date (Web): 11 October 2008

Downloaded from <http://pubs.acs.org> on February 8, 2009



More About This Article

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

- Supporting Information
- 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](#)

Unnatural Substrate Repertoire of A, B, and X Family DNA Polymerases

Gil Tae Hwang and Floyd E. Romesberg*

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

Received May 22, 2008; E-mail: floyd@scripps.edu

Abstract: As part of an effort to develop unnatural base pairs that are stable and replicable in DNA, we examined the ability of five different polymerases to replicate DNA containing four different unnatural nucleotides bearing predominantly hydrophobic nucleobase analogs. The unnatural pairs were developed based on intensive studies using the Klenow fragment of DNA polymerase I from *E. coli* (Kf) and found to be recognized to varying degrees. The five additional polymerases characterized here include family A polymerases from bacteriophage T7 and *Thermus aquaticus*, family B polymerases from *Thermococcus litoralis* and *Thermococcus* 9°N-7, and the family X polymerase, human polymerase β . While we find that some aspects of unnatural base pair recognition are conserved among the polymerases, for example, the pair formed between two d3FB nucleotides is typically well recognized, the detailed recognition of most of the unnatural base pairs is generally polymerase dependent. In contrast, we find that the pair formed between d5SICS and dMMO2 is generally well recognized by all of the polymerases examined, suggesting that the determinants of efficient and general recognition are contained within the geometric and electronic structure of these unnatural nucleobases themselves. The data suggest that while the d3FB:d3FB pair is sufficiently well recognized by several of the polymerases for in vitro applications, the d5SICS:dMMO2 heteropair is likely uniquely promising for in vivo use. T7-mediated replication is especially noteworthy due to strong mismatch discrimination.

1. Introduction

The ability to replicate and amplify DNA in vitro has revolutionized biotechnology by enabling a wide range of techniques, including PCR, cloning, and DNA sequencing. Much interest is also focused on the use of oligonucleotides for the development of novel materials,^{1,2} diagnostics,³ and therapeutics.^{4–7} All of these efforts require, or are at least facilitated by, sequence-specific amplification of DNA by DNA polymerases, which is currently limited by the natural substrate repertoires of the polymerases. Identification of novel nucleotides that selectively pair in duplex DNA as well as during replication, i.e., an unnatural base pair, would allow for the expansion of these in vitro techniques to include oligonucleotides that are site-specifically modified with interesting functionalities (catalytic groups, fluorophores, etc.), thereby expanding their scope and potential applications. Additionally, an unnatural base pair would lay the foundation for construction of a synthetic organism with increased potential for information storage that

might be used, for example, to site-specifically label RNA in vivo or encode proteins with unnatural amino acids.^{8–10}

Toward these goals, we synthesized and evaluated a large number of nucleotides bearing predominantly hydrophobic nucleobase analogs^{11–15} and identified several that are recognized with at least moderate efficiency and selectivity by the exonuclease deficient Klenow fragment of DNA polymerase I from *E. coli* (Kf). These studies have elucidated many of the determinants of efficient unnatural base pair replication and were punctuated by the discovery of four base pairs of particular significance (Figure 1). First, while examining a variety of isocarbotiryl-based nucleoside analogs, we showed that Kf synthesizes the dPICS self pair (i.e., by insertion of the unnatural nucleotide opposite itself in the template) with moderate efficiency and fidelity; however, the resulting primer is not extended (i.e., by continued primer elongation resulting from the insertion of the next correct dNTP).¹³ Structural studies

- (1) Seeman, N. C. *Mol. Biotechnol.* **2007**, *37*, 246–257.
- (2) Seeman, N. C. *Trends Biochem. Sci.* **2005**, *30*, 119–235.
- (3) Ieven, M. J. *Clin. Virol.* **2007**, *40*, 259–276.
- (4) Sinha, N. D.; Michaud, D. P. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 807–818.
- (5) Lee, J. F.; Stovall, G. M. *Curr. Opin. Chem. Biol.* **2006**, *10*, 282–289.
- (6) Ng, E. W. M.; Shima, D. T.; Calias, P.; Cunningham, E. T., Jr.; Guyer, D. R.; Adams, A. P. *Nat. Rev. Drug Discovery* **2006**, *5*, 123–132.
- (7) Nimjee, S. M.; Rusconi, C. P.; Sullenger, B. A. *Annu. Rev. Med.* **2005**, *56*, 555–583.

(8) Wang, L.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249.

(9) Xie, J.; Schultz, P. G. *Curr. Opin. Chem. Biol.* **2005**, *9*, 548–554.

(10) Xie, J.; Schultz, P. G. *Methods* **2005**, *36*, 227–238.

(11) Henry, A. A.; Olsen, A. G.; Matsuda, S.; Yu, C.; Geierstanger, B. H.; Romesberg, F. E. *J. Am. Chem. Soc.* **2004**, *126*, 6923–6931.

(12) Leconte, A. M.; Hwang, G. T.; Matsuda, S.; Capek, P.; Hari, Y.; Romesberg, F. E. *J. Am. Chem. Soc.* **2008**, *130*, 2336–2343.

(13) McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **1999**, *121*, 11585–11586.

(14) Ogawa, A. K.; Wu, Y.; McMinn, D. L.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 3274–3287.

(15) Tae, E. L.; Wu, Y. Q.; Xia, G.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2001**, *123*, 7439–7440.

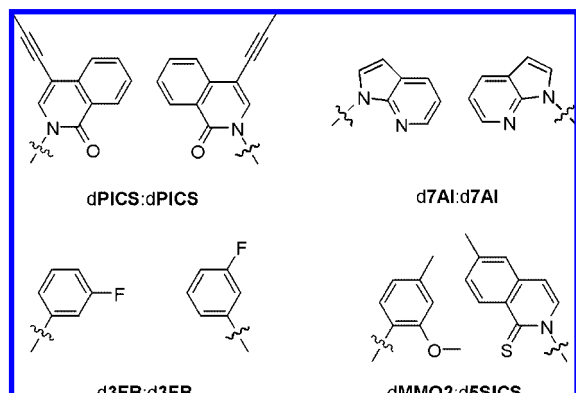


Figure 1. Unnatural base pairs used in this study. Only the paired nucleobase analogs are shown for clarity.

revealed that in duplex DNA the **dPICS** nucleobases interact by intercalation,¹⁶ which we speculate facilitates self pair synthesis but results in a primer terminus structure that is refractory to extension, at least with Kf. Second, after examining a variety of indole-based scaffolds, we found that Kf synthesizes the **d7AI** self pair with moderate efficiency but cannot extend it,¹⁴ while rat pol β cannot synthesize the self pair but extends it with a rate comparable to that of a natural base pair.¹⁵ Thus, full-length synthesis of DNA containing the **d7AI** self pair is possible when both Kf and pol β are present.

The results of these early studies prompted us to examine nucleotides bearing small phenyl-ring nucleobase analogs. These analogs are expected to be incapable of interstrand intercalation, and we hypothesized that they might be derivatized to facilitate synthesis and extension. After extensive investigations, we found that inclusion of a single fluorine substituent at the position meta to the *C*-glycosidic bond significantly facilitates Kf-mediated self pair synthesis and extension (**d3FB**, Figure 1).¹¹ Structural and biochemical studies have shown that the **d3FB** self pair forms a more natural-like ‘in plane’ primer terminus and also suggested that the specific dipole alignment within the self pair may be important for both its efficient synthesis and extension.^{11,16}

Finally, we recently conducted two separate screens of 3600 possible base pair candidates for self pairs or heteropairs that are efficiently synthesized and extended by Kf.¹² After optimization of the initial hits, these screening efforts yielded the **d5SICS:dMMO2** heteropair (Figure 1). Steady-state kinetic analysis revealed that this heteropair is synthesized and extended by Kf with remarkable efficiency and fidelity in either strand context. Further characterization suggested that, at least with Kf, efficient replication relies on the substituents at the position ortho to the glycosidic linkage, which are likely disposed in the developing minor groove at the primer terminus.

Our efforts to design predominantly hydrophobic unnatural base pairs that are efficiently replicated have been based mainly on their recognition by Kf. However, Kf appears to recognize nucleoside analogs differently than other polymerases, such as the thermostable polymerases Vent and 9^oN-7^{17–19} or the mesophilic polymerase pol α .^{20,21} Moreover, work from the

Benner laboratory with nucleobase analogs that pair via orthogonal H-bonding topologies indicates that recognition can be polymerase dependent.^{22–26} In addition, Kool et al. have shown that efficient extension of H-bonding-deficient nucleobase isosteres by different polymerases relies to different extents on polymerase-nucleobase H-bonds²⁷ and that different polymerases accommodate changes in nucleobase size differently.^{28–31} Finally, our own data showing the significant difference in recognition of the **d7AI** self pair by Kf and pol β ¹⁵ suggests that there may be significant variability in how different polymerases recognize the predominantly hydrophobic base pairs.

No comprehensive comparisons of substrate recognition have been reported for DNA-dependent DNA polymerases; however, they have been grouped into six families, A, B, C, D, X, and Y, based on sequence homology.^{32–34} Of these, the two largest and most thoroughly studied are the A family (or type I), exemplified by DNA polymerase I from *E. coli*, and the B (or pol α) family, exemplified by eukaryotic DNA polymerase α . In addition, the family X polymerases, especially pol β , have also been well characterized and are structurally and functionally diverged from the A and B family polymerases. To begin to explore the potential substrate repertoires of natural DNA polymerases, we analyzed the replication of the **dPICS**, **d7AI**, and **d3FB** self pairs and the **d5SICS:dMMO2** heteropair by two family A polymerases (T7 and Taq) and two family B polymerases (Vent and Thermanator). These polymerases are among the most useful for biotechnology applications, and they have been characterized extensively.^{35–41} In addition, each is a replicative polymerase that is responsible for both leading and lagging strand replication in vivo. Due to its structural and functional divergence, we also analyzed unnatural DNA replication by the human family X polymerase pol β , which is a repair

- (16) Matsuda, S.; Leconte, A. M.; Romesberg, F. E. *J. Am. Chem. Soc.* **2007**, *129*, 5551–5558.
 (17) Gardner, A. F.; Joyce, C. M.; Jack, W. E. *J. Biol. Chem.* **2004**, *279*, 11834–11842.
 (18) Gardner, A. F.; Jack, W. E. *Nucleic Acids Res.* **2002**, *30*, 605–613.
 (19) Gardner, A. F.; Jack, W. E. *Nucleic Acids Res.* **1999**, *12*, 2545–2553.
 (20) Chiamonte, M.; Moore, C. L.; Kincaid, K.; Kuchta, R. D. *Biochemistry* **2003**, *42*, 10472–10481.

- (21) Kincaid, K.; Beckman, J.; Zivkovik, A.; Halcomb, R. L.; Engels, J. W.; Kuchta, R. D. *Nucleic Acids Res.* **2005**, *33*, 2620–2628.
 (22) Horlacher, J.; Hottiger, M.; Podust, V. N.; Hübscher, U.; Benner, S. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6329–6333.
 (23) Lutz, M. J.; Held, H. A.; Hottiger, M.; Hübscher, U.; Benner, S. A. *Nucleic Acids Res.* **1996**, *24*, 1308–1313.
 (24) Lutz, M. J.; Horlacher, J.; Benner, S. A. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1149–1152.
 (25) Lutz, M. J.; Horlacher, J.; Benner, S. A. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 499–504.
 (26) Yang, Z.; Sismour, A. M.; Sheng, P.; Puskar, N. L.; Benner, S. A. *Nucleic Acids Res.* **2007**, *35*, 4238–4249.
 (27) Morales, J. C.; Kool, E. T. *Biochemistry* **2000**, *39*, 12979–12988.
 (28) Kim, T. W.; Briebe, L. G.; Ellenberger, T.; Kool, E. T. *J. Biol. Chem.* **2006**, *281*, 2289–2295.
 (29) Sintim, H. O.; Kool, E. T. *Angew. Chem., Int. Ed.* **2006**, *45*, 1974–1979.
 (30) Kim, T. W.; Delaney, J. C.; Essigmann, J. M.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15803–15808.
 (31) Kool, E. T. *Annu. Rev. Biochem.* **2002**, *71*, 191–219.
 (32) Hübscher, U.; Maga, G.; Spadari, S. *Annu. Rev. Biochem.* **2002**, *71*, 133–163.
 (33) Filé, J.; Forterre, P.; Sen-Lin, T.; Laurent, J. *J. Mol. Evol.* **2002**, *54*, 763–773.
 (34) Braithwaite, D. K.; Ito, J. *Nucleic Acids Res.* **1993**, *21*, 787–802.
 (35) Kiefer, J. R.; Mao, C.; Braman, J. C.; Beese, L. S. *Nature* **1998**, *391*, 304–307.
 (36) Doublé, S.; Tabor, S.; Long, A. M.; Richardson, C. C.; Ellenberger, T. *Nature* **1998**, *391*, 251–258.
 (37) Li, Y.; Waksman, G. *Protein Sci.* **2001**, *10*, 1225–1233.
 (38) Li, Y.; Korolev, S.; Waksman, G. *EMBO J.* **1998**, *17*, 7514–7525.
 (39) Li, Y.; Kong, Y.; Korolev, S.; Waksman, G. *Protein Sci.* **1998**, *7*, 1116–1123.
 (40) Franklin, M. C.; Wang, J.; Steitz, T. A. *Cell* **2001**, *105*, 657–667.
 (41) Rodriguez, A. C.; Park, H. W.; Mao, C.; Beese, L. S. *J. Mol. Biol.* **2000**, *299*, 447–462.

polymerase involved in gap filling.^{42–47} The results, along with available structural data, allow us to begin to define the general and polymerase-specific determinants of natural and unnatural base pair recognition. In addition, these studies should help in the design of unnatural base pairs, not only for in vitro applications but also for in vivo applications, which will require general recognition by the multiple polymerases involved in genome replication and maintenance.

2. Results

Two A family polymerases were selected for characterization to augment our data with Kf, the mesophilic polymerase from bacteriophage T7 (T7 polymerase) and the thermophilic polymerase from *Thermus aquaticus* (Taq polymerase), both of which have high homology to Kf and have been particularly well characterized. For our initial characterization of B family polymerases we selected two thermophilic polymerases, Vent, from the archaea *Thermococcus litoralis*,⁴⁸ and Terminator, a mutant of the native DNA polymerase from the *Thermococcus species* 9°N-7.^{18,19} In addition, based on our previous characterization of pol β with d7AI (vide supra) and its diverged role as a DNA repair enzyme, we included the human variant of this X family polymerase. All of the polymerases used in this study are either naturally exonuclease deficient or rendered so by point mutation. The mesophilic polymerases were characterized at 25 °C with the exception of pol β , which due to generally reduced activity with both natural and unnatural substrates was characterized at its optimal temperature of 37 °C. The thermophilic polymerases were characterized at 50 °C to allow the use of the same primer-template substrates employed in our previous studies (significant melting of the duplex would occur at higher temperatures). While this is approximately 20 °C below their optimal temperatures, each polymerase retains the majority of its activity under these conditions.^{49–51}

The recognition of the different unnatural base pairs was evaluated using steady-state kinetics, which provides a convenient means to assay the overall rate at which product is produced. The most relevant and interpretable data available from this approach is the ratio k_{cat}/K_M (or efficiency), which is a second-order rate constant relating the duplex-bound polymerase and free dNTP to the rate-limiting transition state for multiple turnover dNTP insertion. The lower limit of detection of the assay is $k_{cat}/K_M \approx 1.0 \times 10^3 \text{ M}^{-1} \text{ min}^{-1}$. The individual values of k_{cat} and K_M are reported in the Supporting Information but are generally not discussed as their interpretation depends on the rate-limiting step of the reaction, which may vary among

the different polymerases and unnatural base pairs. The absolute efficiencies of synthesis (correct pair and all possible mismatches) and extension (correct pair and select mismatches) are listed in Tables 1 and 2, respectively, for the A family polymerases and Tables 4 and 5 for the B and X family polymerases. While the synthesis of all possible mismatches was characterized, we only characterized the extension of the most efficiently synthesized mismatches in order to approximate the overall fidelity of replication. In cases where no mismatch is efficiently synthesized, the mismatch with dA was characterized as it is commonly one of the most problematic.^{11,16,52–54} The overall fidelity is calculated as the product of the synthesis and extension fidelities. It is important to note that unless otherwise noted, the calculated fidelities are minimum fidelities, meaning the rate of correct unnatural base pair synthesis and/or extension, relative to the most competitively synthesized and/or extended mismatch. Tables 3 and 6 summarize the most relevant information, including the efficiencies of synthesis and extension, and the overall fidelity for the replication of each unnatural base pair by the A family polymerases (Table 3) or the B and X family polymerases (Table 6). To facilitate comparison of the mesophilic and thermophilic polymerases in the following sections, it is convenient to compare the second-order rate constants after normalization by the efficiencies for natural synthesis under identical conditions. Throughout, heteropairs (natural or unnatural) are denoted as dX:dY, with dX indicating the nucleotide in the primer strand and dY indicating the nucleotide in the template strand.

2.1. Kf Polymerase.

2.1.1. Kf-Mediated Synthesis and Extension of Natural Base Pairs. To provide a reference for the unnatural base pairs, we first measured the steady-state rates at which dATP or dGTP is inserted opposite a template dT. Kf inserts the correct and incorrect triphosphate with second-order rate constants of $3.2 \times 10^8 \text{ M}^{-1} \text{ min}^{-1}$ ⁵⁵ and $5.7 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$, respectively, resulting in a fidelity under these conditions of 5.6×10^3 . Preferential extension of a correct primer terminus is also critical for the high fidelity replication of DNA, thus we determined the steady-state rates of dA:dT pair and dG:dT mismatch extension by the insertion of dCTP opposite a template dG. Kf extends the dA:dT pair and dG:dT mismatch with second-order rate constants of $1.7 \times 10^8 \text{ M}^{-1} \text{ min}^{-1}$ ⁵⁵ and $4.8 \times 10^5 \text{ M}^{-1} \text{ min}^{-1}$, respectively, resulting in a fidelity in this context of 350. The overall fidelity, which is the selectivity for the synthesis and extension of dA:dT, relative to the dG:dT mismatch in this strand context is 2×10^6 .

2.1.2. Recognition of Unnatural Base Pairs by Kf. Synthesis of the unnatural base pairs and to some extent their extension by Kf have been reported previously^{11–14} and are included here to facilitate comparison. At 25 °C, Kf synthesizes the dPICS self pair with a second-order rate constant of $2.4 \times 10^5 \text{ M}^{-1} \text{ min}^{-1}$, but no extension of the nascent self pair is detected. The only mismatch synthesized with a significant rate results from insertion of dTTP, which results in a minimal synthesis fidelity of 20.¹³ Similarly, as mentioned above, Kf synthesizes the d7AI self pair with reasonable efficiency ($k_{cat}/K_M = 2.2 \times 10^5 \text{ M}^{-1}$

(42) Sawaya, M. R.; Prasad, R.; Wilson, S. H.; Kraut, J.; Pelletier, H. *Biochemistry* **1997**, *36*, 11205–11215.

(43) Pelletier, H.; Sawaya, M. R.; Wolffe, W.; Wilson, S.; Kraut, J. *Biochemistry* **1996**, *35*, 12742–12761.

(44) Sawaya, M. R.; Pelletier, H.; Kumar, A.; Wilson, S. H.; Kraut, J. *Science* **1994**, *264*, 1930–1935.

(45) Beard, W. A.; Wilson, S. H. *Structure* **2003**, *11*, 489–496.

(46) Werneburg, B. G.; Ahn, J.; Zhong, X.; Hondal, R. J.; Kravynov, V. S.; Tsai, M. D. *Biochemistry* **1996**, *35*, 7041–7050.

(47) Osheroff, W. P.; Jung, H. K.; Beard, W. A.; Wilson, S. H.; Kunkel, T. A. *J. Biol. Chem.* **1999**, *274*, 3642–3650.

(48) Perler, F. B.; Comb, D. G.; Jack, W. E.; Moran, L. S.; Qiang, B.; Kucera, R. B.; Benner, J.; Slatko, B. E.; Nwankwo, D. O.; Hempstead, S. K.; Carlow, C. K. S.; Jannasch, H. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5577–5581.

(49) Datta, K.; LiCata, V. J. *Nucleic Acids Res.* **2003**, *31*, 5590–5597.

(50) Kong, H.; Kucera, R. B.; Jack, W. E. *J. Biol. Chem.* **1993**, *268*, 1965–1975.

(51) Ichida, J. K.; Horhota, A.; Zou, K.; McLaughlin, L. W.; Szostak, J. W. *Nucleic Acids Res.* **2005**, *33*, 5219–5225.

(52) Kim, Y.; Leconte, A. M.; Hari, Y.; Romesberg, F. E. *Angew. Chem., Int. Ed.* **2006**, *45*, 7809–7812.

(53) Matsuda, S.; Henry, A. A.; Romesberg, F. E. *J. Am. Chem. Soc.* **2006**, *128*, 6369–75.

(54) Hwang, G. T.; Romesberg, F. E. *Nucleic Acids Res.* **2006**, *34*, 2037–45.

(55) Hwang, G. T.; Leconte, A. M.; Romesberg, F. E. *ChemBioChem* **2007**, *8*, 1606–1611.

Table 1. Family A Polymerase-Mediated Incorporation (k_{cat}/K_M , $\text{M}^{-1} \text{min}^{-1}$) of dXTP Opposite Unnatural Nucleotides in the Template^a

5' -dTAAATACGACTCACTATAGGGAGA
3' -dATTATGCTGAGTGATATCCCTCT (**Y**) GCTAGGTTACGGCAGGATCGC

X	Y	Kf ^b	T7	Taq
dA	dT	$(3.2 \pm 0.6) \times 10^8$	$(1.5 \pm 0.4) \times 10^7$	$(4.0 \pm 0.5) \times 10^7$
dG	dT	$(5.7 \pm 0.5) \times 10^4$	$(9.8 \pm 0.7) \times 10^2$	$(2.5 \pm 0.5) \times 10^4$
dPICS	dPICS	$(2.4 \pm 0.3) \times 10^5$	$<1.0 \times 10^3$	$(2.6 \pm 0.5) \times 10^4$
dA	dPICS	$(1.8 \pm 0.4) \times 10^3$	$<1.0 \times 10^3$	$<1.0 \times 10^3$
dC	dPICS	$(1.8 \pm 0.5) \times 10^3$	$<1.0 \times 10^3$	$<1.0 \times 10^3$
dG	dPICS	$(1.7 \pm 0.4) \times 10^3$	$<1.0 \times 10^3$	$<1.0 \times 10^3$
dT	dPICS	$(1.2 \pm 0.2) \times 10^4$	$<1.0 \times 10^3$	$(1.2 \pm 0.3) \times 10^3$
d7AI	d7AI	$(2.2 \pm 0.5) \times 10^5$	$(2.2 \pm 0.3) \times 10^3$	$(1.2 \pm 0.1) \times 10^4$
dA	d7AI	$(5.9 \pm 1.0) \times 10^3$	$(1.2 \pm 0.5) \times 10^3$	$(4.1 \pm 0.1) \times 10^3$
dC	d7AI	$(5.5 \pm 0.9) \times 10^3$	$(1.1 \pm 0.4) \times 10^3$	$<1.0 \times 10^3$
dG	d7AI	$(1.6 \pm 0.4) \times 10^3$	$<1.0 \times 10^3$	$<1.0 \times 10^3$
dT	d7AI	$(1.6 \pm 0.6) \times 10^3$	$<1.0 \times 10^3$	$(2.3 \pm 0.6) \times 10^3$
d3FB	d3FB	$(2.1 \pm 0.8) \times 10^6$	$(5.3 \pm 1.6) \times 10^4$	$(1.5 \pm 0.1) \times 10^4$
dA	d3FB	$(5.4 \pm 1.0) \times 10^5$	$(1.1 \pm 0.1) \times 10^5$	$(9.1 \pm 2.0) \times 10^4$
dC	d3FB	$(2.1 \pm 1.2) \times 10^3$	$(1.8 \pm 0.2) \times 10^3$	$(2.1 \pm 2.9) \times 10^3$
dG	d3FB	$(1.8 \pm 0.6) \times 10^3$	$<1.0 \times 10^3$	$<1.0 \times 10^3$
dT	d3FB	$(9.5 \pm 1.1) \times 10^4$	$(8.1 \pm 0.3) \times 10^4$	$(2.8 \pm 0.7) \times 10^4$
dMMO2	d5SICS	$(3.6 \pm 0.7) \times 10^5$	$(3.7 \pm 0.1) \times 10^4$	$(8.6 \pm 0.9) \times 10^4$
d5SICS	d5SICS	$(2.7 \pm 0.9) \times 10^4$	$<1.0 \times 10^3$	$(3.2 \pm 0.7) \times 10^3$
dA	d5SICS	$(2.2 \pm 0.3) \times 10^4$	$(1.7 \pm 0.2) \times 10^3$	$(2.7 \pm 0.7) \times 10^3$
dC	d5SICS	$<1.0 \times 10^3$	$<1.0 \times 10^3$	$<1.0 \times 10^3$
dG	d5SICS	$(1.3 \pm 0.8) \times 10^5$	$<1.0 \times 10^3$	$(3.7 \pm 0.5) \times 10^4$
dT	d5SICS	$(1.3 \pm 0.4) \times 10^4$	$<1.0 \times 10^3$	$(3.7 \pm 0.9) \times 10^3$
d5SICS	dMMO2	$(4.7 \pm 0.4) \times 10^7$	$(1.1 \pm 0.2) \times 10^6$	$(3.5 \pm 0.6) \times 10^6$
dMMO2	dMMO2	$(1.2 \pm 0.1) \times 10^5$	$(1.5 \pm 0.3) \times 10^4$	$(2.1 \pm 0.6) \times 10^4$
dA	dMMO2	$(1.0 \pm 0.1) \times 10^5$	$(5.8 \pm 1.7) \times 10^4$	$(1.1 \pm 0.3) \times 10^5$
dC	dMMO2	$<1.0 \times 10^3$	$<1.0 \times 10^3$	$<1.0 \times 10^3$
dG	dMMO2	$<1.0 \times 10^3$	$<1.0 \times 10^3$	$<1.0 \times 10^3$
dT	dMMO2	$<1.0 \times 10^3$	$(1.9 \pm 0.1) \times 10^3$	$(8.2 \pm 0.8) \times 10^3$

^a See text and Supporting Information for experimental details. ^b References 11–14. Error reported is standard deviation calculated from three independent experiments.

Table 2. Family A Polymerase-Mediated Extension (k_{cat}/K_M , $\text{M}^{-1} \text{min}^{-1}$) of dX:dY Pairs^a

5' -dTAAATACGACTCACTATAGGGAGA (**X**)
3' -dATTATGCTGAGTGATATCCCTCT (**Y**) GCTAGGTTACGGCAGGATCGC

X	Y	Kf ^b	T7	Taq
dA	dT	$(1.7 \pm 0.2) \times 10^8$	$(1.3 \pm 0.3) \times 10^7$	$(1.9 \pm 0.2) \times 10^7$
dG	dT	$(4.8 \pm 0.5) \times 10^5$	$(1.8 \pm 0.1) \times 10^4$	$(7.4 \pm 1.2) \times 10^3$
dPICS	dPICS	$<1.0 \times 10^3$	$(2.8 \pm 0.4) \times 10^3$	$<1.0 \times 10^3$
dA	dPICS	nd	$<1.0 \times 10^3$	nd
dT	dPICS	$(2.0 \pm 1.1) \times 10^5$	nd	$(1.6 \pm 0.1) \times 10^3$
d7AI	d7AI	$<1.0 \times 10^3$	$<1.0 \times 10^3$	$<1.0 \times 10^3$
dA	d7AI	$(1.9 \pm 0.2) \times 10^3$	$<1.0 \times 10^3$	$<1.0 \times 10^3$
dC	d7AI	$(4.5 \pm 0.7) \times 10^3$	$<1.0 \times 10^3$	nd
d3FB	d3FB	$(3.3 \pm 0.6) \times 10^5$	$(1.5 \pm 0.4) \times 10^5$	$(1.6 \pm 0.4) \times 10^4$
dA	d3FB	$(6.3 \pm 1.7) \times 10^4$	$<1.0 \times 10^3$	$<1.0 \times 10^3$
dC	d3FB	$(1.4 \pm 0.5) \times 10^4$	nd	nd
dG	d3FB	$<1.0 \times 10^3$	nd	nd
dT	d3FB	$(1.9 \pm 0.4) \times 10^5$	$(3.4 \pm 0.1) \times 10^3$	$(1.1 \pm 0.2) \times 10^3$
dMMO2	d5SICS	$(1.9 \pm 0.2) \times 10^6$	$(5.4 \pm 1.2) \times 10^5$	$(2.2 \pm 0.2) \times 10^4$
d5SICS	d5SICS	$<1.0 \times 10^3$	$<1.0 \times 10^3$	$<1.0 \times 10^3$
dA	d5SICS	nd	$<1.0 \times 10^3$	$<1.0 \times 10^3$
dG	d5SICS	$(4.9 \pm 0.5) \times 10^3$	$<1.0 \times 10^3$	$<1.0 \times 10^3$
dT	d5SICS	$(4.0 \pm 0.2) \times 10^5$	nd	$(4.7 \pm 0.7) \times 10^3$
d5SICS	dMMO2	$(6.7 \pm 1.1) \times 10^5$	$(1.1 \pm 0.2) \times 10^6$	$(6.4 \pm 1.6) \times 10^4$
dMMO2	dMMO2	$(5.3 \pm 0.9) \times 10^3$	$(5.9 \pm 0.4) \times 10^3$	$<1.0 \times 10^3$
dA	dMMO2	$(4.6 \pm 0.2) \times 10^4$	$<1.0 \times 10^3$	$(1.7 \pm 0.1) \times 10^3$
dT	dMMO2	nd	nd	$(1.1 \pm 0.1) \times 10^4$

^a See text and Supporting Information for experimental details. ^b References 11–14. Error reported is standard deviation calculated from three independent experiments. nd, not determined.

min^{-1}), while mismatches are synthesized only inefficiently, resulting in a synthesis fidelity of 37; however, neither the dPICS nor the d7AI self pair is extended at a detectable rate. In contrast, Kf not only synthesizes the d3FB self pair with a high efficiency ($2.1 \times 10^6 \text{ M}^{-1} \text{min}^{-1}$) that is only 150-fold

reduced relative to the efficiency for natural base pair synthesis but also extends the self pair with a reasonable efficiency ($3.3 \times 10^5 \text{ M}^{-1} \text{min}^{-1}$).¹¹ Mismatch synthesis with dGTP or dCTP is inefficient, although dATP and dTTP are inserted opposite d3FB slightly more efficiently, reducing the fidelity of Kf-mediated

Table 3. Efficiencies and Overall Fidelities of Family A Polymerase-Mediated Replication

5' -dTAAATACGACTCACTATAGGGAGA (**X**)
3' -dATTATGCTGAGTGATATCCCTCT (**Y**) GCTAGGTTACGGCAGGATCGC

X:Y	Kf			T7			Taq		
	synthesis efficiency	extension efficiency	overall fidelity ^a	synthesis efficiency	extension efficiency	overall fidelity ^a	synthesis efficiency	extension efficiency	overall fidelity ^a
dA:dT	3.2×10^8	1.7×10^8	2.0×10^6	1.5×10^7	1.3×10^7	1.1×10^7	4.0×10^7	1.9×10^7	4.2×10^6
dPICS:dPICS	2.4×10^5	$<1.0 \times 10^3$	nd ^b	$<1.0 \times 10^3$	2.8×10^3	nd ^b	2.6×10^4	$<1.0 \times 10^3$	nd ^b
d7AI:d7AI	2.2×10^5	$<1.0 \times 10^3$	nd ^b	2.2×10^3	$<1.0 \times 10^3$	nd ^b	1.2×10^4	$<1.0 \times 10^3$	nd ^b
d3FB:d3FB	2.1×10^6	3.3×10^5	20	5.3×10^4	1.5×10^5	29	1.5×10^4	1.6×10^4	>2.6
dMMO2:d5SICS	3.6×10^5	1.9×10^6	130	3.7×10^4	5.4×10^5	$>1.2 \times 10^4$	8.6×10^4	2.2×10^4	>51
d5SICS:dMMO2	4.7×10^7	6.7×10^5	7100	1.1×10^6	1.1×10^6	1.4×10^4	3.5×10^6	6.4×10^4	1200

^a Overall fidelity refers to the product of synthesis and extension fidelities; see text for details. ^b Overall fidelity not determined due to low efficiency of unnatural base pair synthesis and/or extension.

Table 4. Family B and Family X Polymerase-Mediated Incorporation (k_{cat}/K_M , $\text{M}^{-1} \text{min}^{-1}$) of dXTP Opposite Unnatural Nucleotides in the Template^a

5' -dTAAATACGACTCACTATAGGGAGA
3' -dATTATGCTGAGTGATATCCCTCT (**Y**) GCTAGGTTACGGCAGGATCGC

X	Y	Vent	Terminator	Pol β
dA	dT	$(1.2 \pm 0.1) \times 10^7$	$(1.5 \pm 0.4) \times 10^8$	$(9.8 \pm 0.5) \times 10^4$
dG	dT	$(1.7 \pm 0.4) \times 10^4$	$(1.9 \pm 0.1) \times 10^6$	$<1.0 \times 10^3$
dPICS	dPICS	$(6.4 \pm 2.1) \times 10^4$	$(2.7 \pm 0.2) \times 10^6$	$<1.0 \times 10^3$
dA	dPICS	$(1.6 \pm 0.7) \times 10^4$	$(1.1 \pm 0.2) \times 10^6$	$<1.0 \times 10^3$
dC	dPICS	$<1.0 \times 10^3$	$(1.8 \pm 0.6) \times 10^5$	$<1.0 \times 10^3$
dG	dPICS	$(5.5 \pm 2.8) \times 10^3$	$(2.6 \pm 0.4) \times 10^5$	$<1.0 \times 10^3$
dT	dPICS	$(8.3 \pm 2.0) \times 10^3$	$(4.3 \pm 0.4) \times 10^5$	$<1.0 \times 10^3$
d7AI	d7AI	$(2.2 \pm 0.6) \times 10^4$	$(4.2 \pm 1.1) \times 10^6$	$<1.0 \times 10^3$
dA	d7AI	$(1.5 \pm 0.1) \times 10^4$	$(9.2 \pm 1.8) \times 10^5$	$<1.0 \times 10^3$
dC	d7AI	$(3.0 \pm 1.1) \times 10^3$	$(1.2 \pm 0.3) \times 10^5$	$<1.0 \times 10^3$
dG	d7AI	$(2.0 \pm 0.4) \times 10^3$	$(1.1 \pm 0.2) \times 10^5$	$<1.0 \times 10^3$
dT	d7AI	$(1.7 \pm 0.4) \times 10^4$	$(5.3 \pm 0.4) \times 10^5$	$<1.0 \times 10^3$
d3FB	d3FB	$(4.6 \pm 1.3) \times 10^4$	$(2.3 \pm 0.3) \times 10^6$	$<1.0 \times 10^3$
dA	d3FB	$(6.4 \pm 0.1) \times 10^4$	$(4.0 \pm 0.6) \times 10^6$	$<1.0 \times 10^3$
dC	d3FB	$(3.0 \pm 0.1) \times 10^3$	$(2.3 \pm 0.4) \times 10^5$	$<1.0 \times 10^3$
dG	d3FB	$(5.1 \pm 1.1) \times 10^3$	$(4.1 \pm 0.3) \times 10^5$	$<1.0 \times 10^3$
dT	d3FB	$(2.6 \pm 0.7) \times 10^4$	$(4.6 \pm 0.6) \times 10^5$	$<1.0 \times 10^3$
dMMO2	d5SICS	$(1.3 \pm 0.4) \times 10^6$	$(4.5 \pm 0.5) \times 10^7$	$<1.0 \times 10^3$
d5SICS	d5SICS	$(5.0 \pm 0.3) \times 10^5$	$(1.4 \pm 0.3) \times 10^7$	$<1.0 \times 10^3$
dA	d5SICS	$(5.0 \pm 0.2) \times 10^4$	$(9.5 \pm 2.4) \times 10^5$	$<1.0 \times 10^3$
dC	d5SICS	$<1.0 \times 10^3$	$(6.0 \pm 4.2) \times 10^4$	$(7.8 \pm 2.2) \times 10^3$
dG	d5SICS	$(3.8 \pm 1.6) \times 10^4$	$(8.1 \pm 0.9) \times 10^5$	$<1.0 \times 10^3$
dT	d5SICS	$(7.0 \pm 0.5) \times 10^4$	$(4.5 \pm 0.4) \times 10^5$	$<1.0 \times 10^3$
d5SICS	dMMO2	$(9.9 \pm 1.0) \times 10^6$	$(4.0 \pm 0.7) \times 10^7$	$(1.6 \pm 0.8) \times 10^3$
dMMO2	dMMO2	$(1.2 \pm 0.1) \times 10^5$	$(7.3 \pm 2.8) \times 10^6$	$<1.0 \times 10^3$
dA	dMMO2	$(3.2 \pm 0.6) \times 10^4$	$(2.6 \pm 0.5) \times 10^6$	$<1.0 \times 10^3$
dC	dMMO2	$(4.4 \pm 0.4) \times 10^4$	$(5.1 \pm 0.4) \times 10^5$	$(6.1 \pm 1.6) \times 10^3$
dG	dMMO2	$(2.9 \pm 0.4) \times 10^3$	$(1.2 \pm 0.3) \times 10^5$	$<1.0 \times 10^3$
dT	dMMO2	$(1.2 \pm 0.3) \times 10^5$	$(1.2 \pm 0.3) \times 10^6$	$<1.0 \times 10^3$

^a See text and Supporting Information for experimental details. Error reported is standard deviation calculated from three independent experiments.

d3FB self pair synthesis to only ~ 4 . The mismatch with dA is extended with a k_{cat}/K_M of $6.3 \times 10^4 \text{ M}^{-1} \text{min}^{-1}$, resulting in an extension fidelity of ~ 5 . Combined, these single-step fidelities result in an overall fidelity of 20 for Kf-mediated d3FB self pair replication.

The d5SICS:dMMO2 pair is the only predominantly hydrophobic *heteropair* we identified to date that is efficiently synthesized and extended by a polymerase in both possible strand contexts.¹² With d5SICS in the template, Kf inserts dMMO2TP with a second-order rate constant of $3.6 \times 10^5 \text{ M}^{-1} \text{min}^{-1}$ and inserts dGTP, d5SICSTP, dATP, and dTTP with low to moderate efficiency, resulting in a synthesis fidelity of ~ 3 . In the opposite context, Kf inserts d5SICSTP opposite dMMO2 with the remarkable efficiency of $4.7 \times 10^7 \text{ M}^{-1} \text{min}^{-1}$, which is only 7-fold slower than natural synthesis under identical conditions. The only mismatches synthesized with any

efficiency are those with dA and dMMO2, but the correct heteropair is still synthesized with a fidelity of 390.

Extension of dMMO2:d5SICS and d5SICS:dMMO2 (referring to primer:template) by Kf is also efficient with a k_{cat}/K_M of 1.9×10^6 and $6.7 \times 10^5 \text{ M}^{-1} \text{min}^{-1}$, respectively. Extension of the most efficiently synthesized mismatch, dG:d5SICS, is not efficient; however, dTTP is inserted with reasonable efficiency, and the resulting dT:d5SICS mismatch is also extended with moderate efficiency, which limits the extension fidelity to only 4.8.¹² For dMMO2 in the template, dATP and dMMO2TP insertion results in the most efficiently synthesized mismatches, but extension of the self pair is barely detectable and extension of the dA:dMMO2 mismatch proceeds inefficiently, resulting in an extension fidelity of 15. Taken together the single-step fidelities result in overall minimum fidelities from 130 (for d5SICS in the template) to 7100 (for dMMO2 in the template).

Table 5. Family B and Family X Polymerase-Mediated Extension (k_{cat}/K_M , $\text{M}^{-1} \text{min}^{-1}$) of dX:dY Pairs^a

5' -dTAATACGACTCACTATAGGGAGA (**X**)
3' -dATTATGCTGAGTGATATCCCTCT (**Y**) GCTAGGTTACGGCAGGATCGC

X	Y	Vent	Terminator	Pol β
dA	dT	$(2.9 \pm 0.6) \times 10^6$	$(3.4 \pm 1.1) \times 10^6$	$(2.1 \pm 0.2) \times 10^5$
dG	dT	$(1.5 \pm 0.1) \times 10^4$	$(7.1 \pm 1.6) \times 10^5$	$(2.8 \pm 0.4) \times 10^3$
dPICS	dPICS	$<1.0 \times 10^3$	$(1.1 \pm 0.2) \times 10^4$	$(9.7 \pm 1.0) \times 10^3$
dA	dPICS	$<1.0 \times 10^3$	$(1.4 \pm 0.1) \times 10^4$	$<1.0 \times 10^3$
dG	dPICS	$<1.0 \times 10^3$	nd	nd
dT	dPICS	$(1.1 \pm 0.1) \times 10^3$	nd	$(1.3 \pm 0.1) \times 10^3$
d7AI	d7AI	$<1.0 \times 10^3$	$(1.2 \pm 0.1) \times 10^5$	$(4.4 \pm 0.5) \times 10^4$
dA	d7AI	$<1.0 \times 10^3$	$(8.7 \pm 1.7) \times 10^3$	$(7.6 \pm 0.3) \times 10^3$
dT	d7AI	$(1.0 \pm 0.1) \times 10^3$	$(2.8 \pm 0.4) \times 10^4$	nd
d3FB	d3FB	$(5.7 \pm 0.2) \times 10^3$	$(7.8 \pm 0.8) \times 10^5$	$(8.3 \pm 0.5) \times 10^4$
dA	d3FB	$(1.1 \pm 0.4) \times 10^3$	$(3.0 \pm 0.4) \times 10^3$	$(6.7 \pm 0.2) \times 10^3$
dT	d3FB	$<1.0 \times 10^3$	nd	nd
dMMO2	d5SICS	$(6.2 \pm 1.6) \times 10^3$	$(4.6 \pm 0.3) \times 10^5$	$(4.4 \pm 0.5) \times 10^4$
d5SICS	d5SICS	$<1.0 \times 10^3$	$(4.5 \pm 1.1) \times 10^3$	$(2.2 \pm 1.0) \times 10^4$
dA	d5SICS	$<1.0 \times 10^3$	$(8.8 \pm 1.5) \times 10^4$	$(3.9 \pm 0.3) \times 10^3$
dC	d5SICS	nd	nd	$(5.3 \pm 0.2) \times 10^3$
dG	d5SICS	$<1.0 \times 10^3$	$(1.5 \pm 0.2) \times 10^5$	$(4.1 \pm 0.1) \times 10^3$
dT	d5SICS	$<1.0 \times 10^3$	$(2.3 \pm 0.5) \times 10^5$	$(4.4 \pm 0.1) \times 10^3$
d5SICS	dMMO2	$(7.9 \pm 0.4) \times 10^3$	$(2.5 \pm 0.6) \times 10^5$	$(4.0 \pm 2.1) \times 10^5$
dMMO2	dMMO2	$(1.2 \pm 0.1) \times 10^3$	$(6.8 \pm 2.0) \times 10^4$	$(1.2 \pm 0.4) \times 10^4$
dA	dMMO2	nd	$(1.5 \pm 0.1) \times 10^3$	$(2.2 \pm 0.1) \times 10^4$
dC	dMMO2	nd	$(3.5 \pm 0.2) \times 10^4$	$(2.1 \pm 0.3) \times 10^4$
dG	dMMO2	nd	$(1.2 \pm 0.1) \times 10^3$	nd
dT	dMMO2	$(3.8 \pm 0.8) \times 10^3$	$(4.4 \pm 1.9) \times 10^4$	nd

^a See text and Supporting Information for experimental details. Error reported is standard deviation calculated from three independent experiments. nd, not determined.

Table 6. Efficiencies and Overall Fidelities of Family B and Family X Polymerase-Mediated Replication

5' -dTAATACGACTCACTATAGGGAGA (**X**)
3' -dATTATGCTGAGTGATATCCCTCT (**Y**) GCTAGGTTACGGCAGGATCGC

X:Y	Vent			Terminator			Pol β		
	synthesis efficiency	extension efficiency	overall fidelity ^a	synthesis efficiency	extension efficiency	overall fidelity ^a	synthesis efficiency	extension efficiency	overall fidelity ^a
dA:dT	1.2×10^7	2.9×10^6	1.3×10^5	1.5×10^8	3.4×10^6	380	9.8×10^4	2.1×10^5	>7400
dPICS:dPICS	6.4×10^4	$<1.0 \times 10^3$	nd ^b	2.7×10^6	1.1×10^4	2.0	$<1.0 \times 10^3$	9.7×10^3	nd ^a
d7AI:d7AI	2.2×10^4	$<1.0 \times 10^3$	nd ^b	4.2×10^6	1.2×10^5	34	$<1.0 \times 10^3$	4.4×10^4	nd ^b
d3FB:d3FB	4.6×10^4	5.7×10^3	3.7	2.3×10^6	7.8×10^5	150	$<1.0 \times 10^3$	8.3×10^4	nd ^b
dMMO2:d5SICS	1.3×10^6	6.2×10^3	>16	4.5×10^7	4.6×10^5	170	$<1.0 \times 10^3$	4.4×10^4	nd ^b
d5SICS:dMMO2	9.9×10^6	7.9×10^3	170	4.0×10^7	2.5×10^5	20	1.6×10^3	4.0×10^5	>4.9

^a Overall fidelity refers to the product of synthesis and extension fidelities; see text for details. ^b Overall fidelity not determined due to low efficiency of unnatural base pair synthesis or extension.

2.2. T7 Polymerase.

2.2.1. T7-Mediated Synthesis and Extension of Natural Base Pairs. T7 inserts dATP and dGTP opposite dT in the template with second-order rate constants of 1.5×10^7 and $9.8 \times 10^2 \text{ M}^{-1} \text{min}^{-1}$, respectively, resulting in a fidelity under these conditions of 1.5×10^4 , which is approximately 3-fold higher than that observed for Kf. T7 extends the dA:dT and dG:dT pairs with second-order rate constants of 1.3×10^7 and $1.8 \times 10^4 \text{ M}^{-1} \text{min}^{-1}$, resulting in a fidelity in this context of 720, 2-fold higher than that observed for Kf. Overall, the selectivity for the synthesis and extension of the dA:dT pair relative to the dG:dT mismatch in this strand context is 1.1×10^7 .

2.2.2. Recognition of Unnatural Base Pairs by T7. Despite its homology with Kf, T7 does not synthesize or extend the dPICS or d7AI self pairs; however, it does synthesize and extend the d3FB self pair with second-order rate constants that are only 40- and 2-fold reduced relative to Kf. Also, similar to Kf, T7 competitively inserts dATP and dTTP opposite d3FB, with the rates of both actually exceeding that of d3FBTP.

Despite their reasonably efficient rates of synthesis, neither the dA:d3FB or dT:d3FB mispairs is efficiently extended, resulting in an overall fidelity of 29.

Generally, T7 recognizes the d5SICS:dMMO2 heteropair better than it recognizes any of the self pairs. T7 inserts dMMO2TP opposite d5SICS with a second-order rate constant of $3.7 \times 10^4 \text{ M}^{-1} \text{min}^{-1}$, which is virtually identical to that for insertion of d3FBTP opposite d3FB. Insertion of the natural dNTPs or d5SICSTP is barely detectable, resulting in a fidelity in this context of 22. With dMMO2 in the template, T7 inserts d5SICSTP with an efficiency of $1.1 \times 10^6 \text{ M}^{-1} \text{min}^{-1}$, which remarkably is only 14-fold less efficient than that for a natural base pair. T7 does not synthesize any mispairs except those with dA and dMMO2, and these are only synthesized inefficiently, resulting in a synthesis fidelity of 19.

Extension of d5SICS:dMMO2 and dMMO2:d5SICS by T7 proceeds with high efficiency ($k_{\text{cat}}/K_M = 1.1 \times 10^6$ and $5.4 \times 10^5 \text{ M}^{-1} \text{min}^{-1}$, respectively). The most efficiently synthesized mispair with either d5SICS or dMMO2 in template is that with dA; however, neither is extended at a detectable level. Although

the dMMO2 self pair is synthesized less efficiently than the dA mismatch, it is extended more efficiently. These rates result in a selectivity for correct pair extension of at least 540 with d5SICS in template and 190 with dMMO2 in template. Remarkably, the overall fidelities for T7-mediated unnatural heteropair synthesis are at least 1.2×10^4 and 1.4×10^4 in its two strand contexts.

2.3. Taq Polymerase.

2.3.1. Taq-Mediated Synthesis and Extension of Natural Base Pairs. At 50 °C, Taq polymerase inserts dATP and dGTP opposite dT in the template with second-order rate constants of 4.0×10^7 and $2.5 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$, resulting in a fidelity of 1.6×10^3 . Taq extends the dA:dT and dG:dT pairs with second-order rate constants of 1.9×10^7 and $7.4 \times 10^3 \text{ M}^{-1} \text{ min}^{-1}$, resulting in a fidelity in this context of 2.6×10^3 . Overall, the selectivity for the synthesis and extension of dA:dT relative to the dG:dT mismatch in this strand context is 4.2×10^6 (It is not appropriate to compare the fidelities of this thermophilic polymerase at 50 °C with those of the mesophilic polymerases at 25 °C due to the temperature differences and because the temperature at which Taq is assayed is further from its optimal temperature of $\sim 75 \text{ }^\circ\text{C}^{50}$).

2.3.2. Recognition of Unnatural Base Pairs by Taq. Despite the temperature differences, self pair recognition by Taq is similar to that by Kf with second-order rate constants that only vary between 1.2×10^4 and $2.6 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$ and extension observed only in the case of the d3FB self pair ($k_{\text{cat}}/K_{\text{M}} = 1.6 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$). Surprisingly, Taq extends the d3FB self pair with the same efficiency that it synthesizes it. Taq does not efficiently synthesize any mismatches with either dPICS or d7AI in the template, resulting in fidelities of 22 and 2.9, respectively, but with d3FB in the template, Taq actually inserts dATP and dTTP opposite d3FB faster than it inserts d3FBTP. Nonetheless, as also observed with T7, Taq does not extend either the dA:d3FB or dT:d3FB mismatch, resulting in a minimum overall selectivity for d3FB self pair synthesis of 2.6.

Taq inserts dMMO2TP opposite d5SICS with a second-order rate constant of only $8.6 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$, but it does not efficiently insert d5SICSTP or any natural dNTP, resulting in a heteropair synthesis fidelity in this context of 2.3. As with the other polymerases, Taq synthesizes the heteropair more efficiently in the opposite strand context, inserting d5SICSTP opposite dMMO2 with an efficiency of $3.5 \times 10^6 \text{ M}^{-1} \text{ min}^{-1}$, which is only 11-fold lower than that for a natural base pair. While Taq does not insert dCTP or dGTP opposite dMMO2, it does insert dATP, dTTP, and dMMO2TP but does so at least 32-fold slower than it inserts the correct unnatural triphosphate.

Taq extends dMMO2:d5SICS and d5SICS:dMMO2 with moderate efficiencies ($k_{\text{cat}}/K_{\text{M}} = 2.2 \times 10^4$ and $6.4 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$). The most efficiently synthesized mismatch with d5SICS in the template, dG:d5SICS, is not extended at a detectable level nor is the self pair or the mismatches with dA, while the mismatch with dT is extended only very inefficiently, resulting in an extension fidelity in this context of 4.7. The most efficiently synthesized mismatch with dMMO2 in the template, dA:dMMO2, is extended at a barely detectable rate, resulting in an extension fidelity of 38. These efficiencies of correct pair and mismatch synthesis and extension result in a minimum overall fidelity for dMMO2:d5SICS and d5SICS:dMMO2 replication of at least 51 and 1200, respectively.

2.4. Vent Polymerase.

2.4.1. Vent-Mediated Synthesis and Extension of Natural Base Pairs. At 50 °C, the family B polymerase Vent inserts

dATP and dGTP opposite dT in the template with second-order rate constants of 1.2×10^7 and $1.7 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$, resulting in a fidelity of 710. Vent extends the dA:dT and dG:dT pairs with second-order rate constants of 2.9×10^6 and $1.5 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$, resulting in a fidelity in this context of 190. Overall, the selectivity for the synthesis and extension of the dA:dT relative to the dG:dT mismatch in this strand context is 1.3×10^5 .

2.4.2. Recognition of Unnatural Base Pairs by Vent. Vent synthesizes each self pair with reasonable efficiency (between 2.2×10^4 and $6.4 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$). While the absolute efficiencies are similar to those of the A family polymerases, relative to the synthesis of a natural base pair, synthesis of the dPICS and d7AI self pairs is significantly better with Vent than with the A family enzymes. Vent does not insert dCTP opposite dPICS but does insert the other natural dNTPs with moderate efficiencies, resulting in a synthesis fidelity of 4. Insertion of dGTP and dCTP opposite d7AI is barely detectable, but dATP and dTTP insertion is slightly more efficient. Opposite d3FB, Vent does not insert dGTP or dCTP efficiently but does insert dTTP and dATP with moderate efficiencies. Nonetheless, none of the synthesized mismatches are efficiently extended, resulting in overall fidelity of 3.7 for Vent-mediated d3FB self pair synthesis.

Vent synthesizes the unnatural heteropair much more efficiently than it synthesizes the self pairs. Insertion of dMMO2TP opposite d5SICS proceeds with a second-order rate constant of $1.3 \times 10^6 \text{ M}^{-1} \text{ min}^{-1}$, an efficiency only 9-fold lower than that of a natural base pair. dATP, dGTP, and dTTP are inserted with only moderate efficiencies, but the self pair of d5SICS is synthesized with a slightly greater efficiency, which limits fidelity in this context to 2.6. As with the other polymerases, d5SICSTP insertion opposite dMMO2 is even more efficient; in fact, it is as efficient as insertion of dATP opposite dT in the same sequence context. Unlike the A family polymerases, which preferentially insert dATP opposite dMMO2, Vent most efficiently inserts dTTP and dMMO2 opposite dMMO2 in the template, while dCTP and dATP are inserted less competitively and dGTP the least competitively. This results in a fidelity of 83. Also unlike the other polymerases, Vent cannot efficiently extend the heteropair in either strand context ($k_{\text{cat}}/K_{\text{M}} \approx 7 \times 10^3 \text{ M}^{-1} \text{ min}^{-1}$) nor can it extend any of the mismatches of these two nucleotide analogs with the natural nucleotides.

2.5. Terminator Polymerase.

2.5.1. Terminator-Mediated Synthesis and Extension of Natural Base Pairs. At 50 °C, the family B polymerase Terminator inserts dATP and dGTP opposite dT in the template with second-order rate constants of 1.5×10^8 and $1.9 \times 10^6 \text{ M}^{-1} \text{ min}^{-1}$, resulting in a fidelity of 79, which is significantly lower than that of the other polymerases. Terminator extends both the dA:dT and dG:dT pairs efficiently with second-order rate constants of 3.4×10^6 and $7.1 \times 10^5 \text{ M}^{-1} \text{ min}^{-1}$, resulting in a fidelity of only 4.8. The low fidelities at each step result in an overall fidelity for natural DNA synthesis of only 380.

2.5.2. Recognition of Unnatural Base Pairs by Terminator. At 50 °C, Terminator synthesizes each of the self pairs with remarkable efficiency, varying only between 2.3×10^6 and $4.2 \times 10^6 \text{ M}^{-1} \text{ min}^{-1}$. The efficient synthesis results in large part from a small apparent K_{M} for triphosphate binding (Supporting Information). However, Terminator also efficiently synthesizes a wide range of mismatches. Opposite both dPICS and d7AI in the template, Terminator inserts the natural dNTPs with reasonable efficiency, limiting fidelity to 2.5 and 4.6,

respectively. Opposite **d3FB**, Terminator inserts dCTP, dGTP, and dTTP with moderate efficiencies, but as with all of the other polymerases except Kf, it inserts dATP more efficiently than it inserts **d3FBTP**.

The **d5SICS:dMMO2** heteropair is recognized by Terminator at 50 °C in both strand contexts with remarkable efficiencies that are only 3- to 4-fold reduced relative to natural synthesis. This is again due to a remarkably small apparent K_M value (Supporting Information), as also observed for self pair synthesis by Terminator. However, as with the self pairs, Terminator also efficiently synthesizes mismatches with both **d5SICS** and **dMMO2**. With **d5SICS** in the template, dCTP is inserted with moderate efficiency, while the remaining dNTPs are inserted with slightly higher efficiencies. The most efficiently synthesized mismatch with **d5SICS** in the template is its self pair, which limits heteropair synthesis fidelity to 3.2. Opposite **dMMO2**, Terminator inserts **dMMO2**, dATP, and dTTP very efficiently and dCTP and dGTP only slightly less efficiently. This results in a fidelity of heteropair synthesis of 5.5.

Like Vent, Terminator at 50 °C catalyzes the extension of **d5SICS:dMMO2** and **dMMO2:d5SICS** less efficiently than it catalyzes their synthesis. However, with second-order rate constants between 2 and $5 \times 10^5 \text{ M}^{-1} \text{ min}^{-1}$, the absolute efficiency of Terminator-mediated extension is significantly greater than that of Vent-mediated extension. With **d5SICS** in the template, the most efficiently synthesized mismatches are those with **d5SICS** and dG (although the mismatches with dA and dT are synthesized with nearly identical efficiencies). While extension of the self pair is inefficient, extension of the mismatches with dT and dG is reasonably efficient and extension of the mismatch with dA is only slightly less efficient. This results in a minimum selectivity for correct pair extension of 2. Mismatches with **dMMO2** in the template are extended only inefficiently, resulting in an extension fidelity of 3.7. Synthesis and extension fidelities combine to yield an overall fidelity for the heteropair of 20 and 170 in its two strand contexts. Remarkably, this is only 2- to 19-fold less than the fidelity observed for natural synthesis.

2.6. Polymerase β .

2.6.1. Polymerase β -Mediated Synthesis and Extension of Natural Base Pairs. Due to low activity, the family X polymerase pol β was assayed at its optimal temperature of 37 °C. At this temperature, pol β inserts dATP and dGTP opposite dT in the template with second-order rate constants of 9.8×10^4 and less than $1.0 \times 10^3 \text{ M}^{-1} \text{ min}^{-1}$, resulting in a fidelity of at least 98. Pol β extends the dA:dT and dG:dT pairs with second-order rate constants of 2.1×10^5 and $2.8 \times 10^3 \text{ M}^{-1} \text{ min}^{-1}$, resulting in a fidelity in this context of 75. Overall, the selectivity for the synthesis and extension of the dA:dT pair relative to the dG:dT mismatch in this strand context is at least 7.4×10^3 . Relatively low-fidelity pol β -mediated extension of a simple primer template (i.e., not a gapped substrate) has been reported previously,^{47,56,57} as has the high frequency of the dG:dT mismatch.⁵⁸

2.6.2. Recognition of Unnatural Base Pairs by Polymerase β . Pol β does not synthesize any of the unnatural pairs examined,

except for the heteropair, and in this case only by insertion of **d5SICSTP** opposite **dMMO2** and even then only at a barely detectable rate. Likewise, pol β also does not synthesize any mismatches with the exception of both mismatches between dC and the nucleotides of the heteropair. In dramatic contrast, pol β extends all of the unnatural pairs examined with second-order rate constants that vary from as efficient as that for natural base pair extension to only 20-fold less efficient. However, mismatch extension is also efficient, resulting in fidelities of 5.8–12 for the self pairs and 2–18 for the heteropair (in its two strand contexts).

3. Discussion

We continue to explore the potential of nucleotides bearing predominantly hydrophobic nucleobase analogs as part of an effort to expand the genetic alphabet. Replication of these analogs has been extensively examined with the family A polymerase Kf. However, our work with rat pol β and the **d7AI** self pair suggests that recognition of predominantly hydrophobic base pairs may be polymerase specific. Indeed, this has been observed with other candidate unnatural base pairs,^{22–26} hydrophobic nucleobase isomers,^{27–31} and nucleotides with modified sugars or nucleobase substituents.¹⁷ However, there has been no detailed and systematic report of the kinetic rates with which different polymerases synthesize and extend unnatural base pairs. To explore the potential substrate repertoire of natural DNA polymerases with the predominantly hydrophobic unnatural base pair candidates and determine whether some aspects of replication are base pair specific or polymerase specific, we characterized the efficiency and selectivity of five different DNA polymerases with four different unnatural base pairs (Figure 1), including the **dPICS**, **d7AI**, and **d3FB** self pairs and the **d5SICS:dMMO2** heteropair. Two family A polymerases (one mesophilic (T7) and one thermophilic (Taq)), two thermophilic family B polymerases (Terminator and Vent), and one family X polymerase (pol β) were selected for characterization. Along with our previous studies of Kf, the results of this study provide a rather broad survey of natural polymerase diversity. A direct comparison of kinetic data acquired at different temperatures is not straightforward; thus, for the purposes of this discussion, we focus on relative rates determined for each polymerase by normalizing the efficiencies for unnatural base pair synthesis and extension to the corresponding values for a natural base pair under identical conditions.

3.1. Recognition of Unnatural Nucleobases.

3.1.1. Self Pairs. From a practical perspective, self pairs are particularly attractive for expanding the genetic alphabet as they require DNA polymerases that efficiently and selectively catalyze insertion of only a single unnatural dNTP and subsequent extension of only a single unnatural primer terminus. Previous structural studies¹⁶ and kinetic studies with Kf^{11,13,14} suggest that the three self pairs examined represent two structurally distinct classes. The **dPICS** nucleotides self pair within duplex DNA via intercalation, which with Kf appears to facilitate synthesis but inhibit extension. Other large aromatic nucleobase analogs have been shown to pair by intercalation,⁵⁹ and **d7AI** is also likely to pair in this manner. In contrast, and as expected based on the reduced surface area of the simple phenyl ring, structural studies have demonstrated that the **d3FB** nucleobases self pair within duplex DNA in an in-plane manner,

(56) Chagovetz, A. M.; Sweasy, J. B.; Preston, B. D. *J. Biol. Chem.* **1997**, *272*, 27501–27504.

(57) Ahn, J.; Kraynov, V. S.; Zhong, X.; Werneburg, B. G.; Tsai, M. D. *Biochem. J.* **1998**, *331*, 79–87.

(58) Beard, W. A.; Osheroff, W. P.; Prasad, R.; Sawaya, M. R.; Jaju, M.; Wood, T. G.; Kraut, J.; Kunkel, T. A.; Wilson, S. H. *J. Biol. Chem.* **1996**, *271*, 12141–12144.

(59) Johar, Z.; Zahn, A.; Leumann, C. J.; Jaun, B. *Chem. Eur. J.* **2008**, *14*, 1080–1086.

which with Kf appears to facilitate extension. Interestingly, these differences in structure appear to be accommodated very differently by the different polymerases, as we find that self pair recognition is polymerase dependent.

The intercalative self pairs **dPICS** and **d7AI** are not synthesized or extended by T7 and not synthesized by pol β . The behavior of T7 is consistent with that observed in previous studies using nucleobases of various sizes, which suggested that T7 has greater structural discrimination than Kf.²⁸ Likewise, the data with pol β is consistent with previous reports that this enzyme does not synthesize base pairs between analogs with modified H-bonding functionality.^{22,60} Despite not being able to synthesize either self pair, pol β can extend both with rates that are only 5- to 20-fold less efficient than extension of a natural base pair. However, due to the generally low rates with which this enzyme replicates natural DNA, the absolute values of the rates remain modest. Somewhat higher efficiencies for the synthesis of the **dPICS** and **d7AI** self pairs are observed with Taq, although not as high as with Kf, and like the other family A polymerases, Taq does not extend either self pair with a rate sufficient to detect. Generally, relative to natural synthesis, the family B polymerases synthesize the self pairs with the greatest efficiency. Therminator extends them with high efficiency as well, which is consistent with its ability to synthesize and extend pairs between orthogonal H-bonding nucleobases.²⁴ However, Therminator is also the only enzyme that synthesizes mispairs of either **dPICS** or **d7AI** with rates that are competitive with those for unnatural base pair synthesis. In contrast, while Vent synthesizes the self pairs only 200- to 500-fold slower than natural synthesis, it is virtually unable to extend any of the self pairs once synthesized. This behavior also is consistent with previous reports that Vent is unable to extend primers past an orthogonal H-bonding nucleotide analog in the template.²⁴

While the **d3FB** self pair is neither synthesized nor extended with even moderate efficiency by Taq and extended but not synthesized by pol β (the same behavior pol β shows with all the unnatural base pairs examined), it is synthesized and extended by all of the other polymerases with efficiencies only 4- to 500-fold reduced relative to a natural base pair. Nonetheless, when efficiencies and fidelities of both synthesis and extension are considered, it is apparent that Kf is uniquely efficient at recognizing the **d3FB** self pair. However, even though Therminator inserts dATP opposite **d3FB** more efficiently than its cognate unnatural triphosphate, the self pair is replicated with high efficiency and reasonable fidelity as the mispair is not efficiently extended. Moreover, while T7 synthesizes the self pair with only marginal efficiency, it does so with reasonable efficiency due to its very low rate of mispair synthesis and extension. Thus, both Therminator and T7 may also prove useful for replicating DNA containing the **d3FB** self pair, perhaps in the presence of a 3' to 5' exonuclease to remove any mispairs formed at the primer terminus.

3.1.2. d5SICS:dMMO2 Heteropair. The **d5SICS:dMMO2** pair is the only unnatural heteropair we identified to date that is efficiently synthesized and extended in both possible strand contexts. While no structural information is yet available, it is presumably packed in duplex DNA with the nucleobases interacting edge-on. In general, both synthesis and extension of this heteropair are efficient, but several interesting similarities and differences are apparent among the different polymerases. While Kf synthesizes the heteropair relatively efficiently in either

strand context, insertion of **d5SICSTP** opposite **dMMO2** is significantly more efficient than synthesis of the heteropair by insertion of **dMMO2TP** opposite **d5SICS**. Polar recognition of the heteropair is also observed with the other family A polymerases. The family A polymerases differ regarding insertion of dGTP opposite **d5SICS**, which is not catalyzed by T7 but proceeds only 2- to 3-fold slower than insertion of **dMMO2TP** by Kf and Taq. Interestingly, the polar recognition of the heteropair is less pronounced with Vent and not observed with Therminator; each synthesizes the heteropair in both strand contexts with rates that are within 10-fold of those for a natural base pair. As with the self pairs, virtually no synthesis of the heteropair in either strand context is observed with pol β .

Kf extends the heteropair with at least reasonable efficiency in both strand contexts, only 100- to 200-fold slower than it extends a natural base pair. The other A family enzymes show divergent behavior: T7 extends the heteropair in either strand context better than Kf, only 10- and 20-fold slower than extension of a natural base pair, while Taq extends it worse, 300- to 900-fold slower than natural synthesis. As with extension of the self pairs, the two family B polymerases showed very different behavior. Therminator efficiently extends the heteropair in both strand contexts only 10-fold slower than natural synthesis, while Vent extends it in both contexts 400- to 500-fold slower than natural synthesis. Although the absolute rates are modest, pol β extends the heteropair in both strand contexts virtually as efficiently as it extends a natural base pair. Interestingly and unlike synthesis of the heteropair, extension of the heteropair is not strongly dependent on strand context with Kf, T7, Taq, or Therminator. However, strand context is more important with pol β , which extends the heteropair with **dMMO2** in the template \sim 10-fold more efficiently than with **d5SICS** in the template. Thus, while differences are apparent, many aspects of heteropair recognition appear to be conserved among the polymerases, suggesting that unlike self pair synthesis much of the recognition of the heteropair is polymerase independent and instead encoded within the nucleobases themselves. This is likely to include a well-packed, edge-on interface between the nucleobases as well as sulfur and methoxy moieties appropriately positioned in the developing minor groove to accept H-bonds from polymerase H-bond donors (see below).

3.2. Structure–Activity Relationships. Structures of ternary complexes with primer/template duplex and triphosphate have been reported for the catalytic domain of Taq (KlenTaq),^{37–39} T7,³⁶ and pol β .^{42–44,61} As a model of the Kf active site, we use the reported ternary complex of DNA polymerase I large fragment from a thermostable strain of *Bacillus stearothermophilus* (Bf), which has high sequence and structural homology to Kf.^{35,62} As a model for Therminator and Vent we use the reported apoenzyme structure of Therminator with substrates modeled in ref 41 as well as the ternary complex of the B family polymerase from RB69.⁴⁰ It is apparent from these structures that each polymerase adopts the standard DNA polymerase fold, which consists of ‘thumb’, ‘palm’, and ‘fingers’ subdomains.

The **dPICS** and **d7AI** self pairs are unique in that their intercalative mode of pairing is likely to induce structural distortions at the primer terminus. Thus, the different recognition of the **dPICS** and **d7AI** self pairs relative to other unnatural pairs may be related to the different ability of the polymerases

(61) Pelletier, H.; Sawaya, M. R.; Kumar, A.; Wilson, S. H.; Kraut, J. *Science* **1994**, *264*, 1891–1903.

(62) Ollis, D. L.; Brick, P.; Hamlin, R.; Xuong, N. G.; Steitz, T. A. *Nature* **1985**, *313*, 762–766.

(60) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 1001–1007.

to accommodate these distortions as they develop at the primer terminus during dNTP incorporation or after they are fully manifest during extension. The polymerase structures reveal that the binding site for the base pair being synthesized is provided by a narrow slot formed by the flanking nucleobases on one side and the polymerase on the other. With the family A polymerases, the polymerase side of this binding slot is provided by an aromatic ring (F710, F667, and Y562 in BF, Taq, and T7, respectively) that tightly packs on the nucleobase of the incoming dNTP and the methylene moiety of a glycine residue (G711, G668, and G527 in BF, Taq, and T7, respectively) that packs on the nucleobase of the templating base. (The template strand bends to expose the templating base to the glycine backbone.) We refer to this packing arrangement as aromatic/aliphatic. The data suggests that this packing motif contributes to efficient catalysis of dNTP insertion when the pairing nucleobases are planar and interact in an edge-on manner, resulting in the most efficient recognition of the d3FB self pair and the unnatural heteropair in either strand context. In contrast, the polymerase side of the binding slot in pol β , when bound to a simple primer template, is provided by side chain methylene groups (D276 and K280) and in RB69 and Therminator by side chain methylene and methyl groups (N564 and L561 and L491 and I488, respectively, in RB69 and Therminator), arrangements that we refer to as aliphatic/aliphatic. The data suggest that relative to the aromatic/aliphatic motif, the aliphatic/aliphatic motifs may be better able to accommodate distortions associated with nucleobase intercalation.

While this potential structure–activity relationship is interesting, there are clearly exceptions to these generalizations. For example, while Taq, Kf, and T7 all employ aromatic/aliphatic packing, T7 appears to be more sensitive to nucleobase distortions (as it does not synthesize the intercalative self pairs at all) and Taq appears to be less sensitive (as it synthesizes the intercalative self pairs as efficiently as it synthesizes the d3FB self pair). Furthermore, Vent (which by homology to the other B family polymerases is likely to employ aliphatic/aliphatic packing interactions via N494 and I498) is consistently the least proficient at extending the unnatural pairs and pol β does not synthesize any of the pairs. Thus, packing at the primer terminus cannot fully account for the differences in recognition. The ‘tightness’ of the T7 active site might contribute to the unique behavior observed here as well as to the hypersensitivity of T7 to increased nucleobase size that was reported previously.²⁸ It is unclear what other factors might contribute to the unique behaviors observed with Taq and Vent, but they may be related to the generally high fidelity observed with these polymerases. With pol β , the substrate specificity during dNTP incorporation is likely to be dominated by its unique H-bonding interactions with both the dNTP and template nucleobase (see below).

While these exceptions currently render the relationship between primer–template packing motif and unnatural base pair recognition speculative, it is interesting that pol β appears to be able to utilize both packing motifs. As mentioned above, pol β appears to employ the aliphatic/aliphatic motif when bound to a simple primer–template.⁶¹ However, when bound to a gapped substrate, a conformational change results in the packing of the primer–template with an aromatic(His34)/aliphatic motif.⁴² Interestingly, the fidelity of pol β is different with the simple primer–template and gapped substrates, suggesting that the different packing arrangements may indeed contribute importantly to substrate recognition.⁴⁷

In addition to differences in packing interactions, the polymerases examined also show differences in the number of H-bonds that they make with their bound substrates. These H-bonds, between donors in the polymerase and acceptors within the developing minor groove of the substrates, are conserved in all natural nucleobases but appropriately positioned only within the context of a correct Watson–Crick base pair.^{63–66} Interestingly, the A family polymerases form more of these H-bonds than the B or X family polymerases. T7, Kf, and Taq make seven, five, and four contacts, respectively, including H-bonds with both the primer terminus nucleobase and its partner in the template, while pol β and RB69 (and by inference, Therminator⁴¹) make only two, including only a single H-bond with the primer terminus nucleobase or its partner in the template strand. Fewer H-bonds in Therminator and pol β may impart the primer terminus with increased freedom to adopt structures required for efficient unnatural base pair synthesis or extension. As with the packing interactions described above, if these H-bonding interactions do contribute to efficient extension, the data suggests that they may have diverged in Vent or that other interactions between Vent and its substrates may be more important.

Interestingly, pol β is the only one of the polymerases examined that also forms analogous H-bonds with the incoming dNTP and the templating nucleobase. We hypothesize that these unique interactions with the forming base pair make pol β hypersensitive to its structure and H-bonding topology, possibly explaining the enzyme’s general inability to synthesize unnatural base pairs. This is consistent with the inability of pol β to synthesize pairs between nucleobases with orthogonal H-bonds²² or pairs where one or both nucleobases are nonpolar isosters.⁶⁰

A primer terminus formed by a natural base pair is efficiently recognized by all DNA polymerases because it adopts an appropriate geometry, which is ensured by a combination of Watson–Crick interbase H-bonding and shape complementarity^{67–69} and because it presents appropriately positioned H-bond acceptors.^{63–65} However, the structural analysis presented above suggests that the different polymerases evolved somewhat different mechanisms to recognize these termini and that each polymerase may have a rather different potential unnatural substrate repertoire. This is likely the origin of the strongly polymerase-dependent self pair recognition. It is also possible that in the absence of suitably positioned H-bonds or shape complementarity, self pair incorporation and extension depend not on recognition per se but rather on the avoidance of the polymerase-dependent mechanisms that evolved to discriminate against natural mispairs, as suggested previously by Kuchta and co-workers.^{20,21} In contrast, while generally orthogonal to the natural nucleobases, the nucleobases of the d5SICS:dMMO2 heteropair appear to interact in a manner that better mimics a natural base pair and/or avoids the structural and/or electrostatic check points employed by polymerases to discriminate against mispairs between the natural nucleobases.

(63) Meyer, A. S.; Blandino, M.; Spratt, T. E. *J. Biol. Chem.* **2004**, *279*, 33043–33046.

(64) Polesky, A. H.; Dahlberg, M. E.; Benkovic, S. J.; Grindley, N. D.; Joyce, C. M. *J. Biol. Chem.* **1992**, *267*, 8417–8428.

(65) Seeman, N. C.; Rosenberg, J. M.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 804–808.

(66) Bebenek, A.; Dressman, H. K.; Carver, G. T.; Ng, S.-s.; Petrov, V.; Yang, G.; Konigsberg, W. H.; Karam, J. D.; Drake, J. W. *J. Biol. Chem.* **2001**, *276*, 10387–10397.

(67) Delaney, J. C.; Henderson, P. T.; Helquist, S. A.; Morales, J. C.; Essigmann, J. M.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4469–4473.

3.3. Progress toward the Expansion of the Genetic Alphabet.

An unnatural base pair that is selectively and efficiently replicated by a DNA polymerase would significantly increase the scope of biotechnologies that either employ selectively modified oligonucleotides (i.e., for novel materials development^{1,2}) or rely on the selective amplification of unnatural oligonucleotides (i.e., for *SELEX*-based selections of aptamers, ribozymes, or DNA enzymes with novel functionalities⁷⁰). Such in vitro applications are somewhat less demanding than in vivo applications in terms of efficiency and fidelity; for example, the unnatural dNTPs may be present in excess or more time may be allowed for amplification. Moreover, for in vitro applications the unnatural base pair needs only be recognized by the polymerase present. For these applications, the most promising base pair–polymerase combinations appear to be the d3FB self pair with Kf or T7 and the d5SICS:dMMO2 heteropair with Kf, T7, or Therminator (Table 3 and 6). The remarkably high fidelity with which T7 replicates DNA containing the self pair and the heteropair are particularly noteworthy. Thus, at least to help facilitate the development of such systems, these unnatural base pairs already appear to be sufficient, although continued optimization is expected to even further increase their utility.

The demands of efficiency and fidelity are much more exacting for any in vivo application. To ensure that the unnatural base pairs are stable in a genome, they should be efficiently recognized by any polymerase involved in the replication and or maintenance of the genome. It is thus critical to characterize the unnatural base pairs according to those whose recognition is polymerase specific and those recognized more generally by different polymerases. Clearly, the self pairs may be placed in the former category and the heteropair in the latter category. In fact, the d5SICS:dMMO2 heteropair is generally well recognized by all of the polymerases examined, suggesting that it might be well recognized by any replicative or repair polymerase. Our results indicate that the determinants of efficient replication are generally contained within the unnatural base pair itself. It is particularly noteworthy that the *N*-glycosidic sulfur and the *C*-glycosidic ortho methoxy substituents, which are likely to engage the polymerase from within the developing minor groove and are so central to Kf recognition,¹² also appear to be well recognized by different polymerases. The slow step with each polymerase examined is insertion of dMMO2TP opposite d5SICS in the template. Not only is this expected to result in polymerase stalling, which might be expected to be problematic in vivo, but it also results in somewhat reduced

fidelity relative to the other strand context. Current efforts are focused on the optimization of this step of unnatural base pair replication and particularly on its optimization during T7-mediated replication.

4. Materials and Methods

The unnatural triphosphates and phosphoramidites of dPICS, d7AI, d3FB, dMMO2, and d5SICS were prepared as described,^{11–15} and the phosphoramidites were incorporated into oligonucleotides using standard procedures. Kf and T7 polymerases were obtained from GE Healthcare (Piscataway, NJ), Taq, Therminator, and Vent polymerases were obtained from New England Biolabs (Ipswich, MA), and pol β was obtained from CHIMERx (Madison, WI).

Steady-State Kinetics. The unnatural nucleotides were evaluated as substrates for DNA polymerases by measuring the initial rates at which a [γ -³³P]-labeled primer–template, 5'-d(TAATAC-GACTCACTATAGGGAGAX), annealed to the 45-mer template, 5'-d(CGCTAGGACGGCATTGGATCGYTCCTCC-TATAGTGAGTCGTATTA), was extended with varying concentrations of natural or unnatural nucleoside triphosphates. Each reaction included 40 nM primer–template, 0.3–1.2 nM enzyme, and Klenow reaction buffer (50 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 1 mM DTT, and 50 μ g/mL acetylated BSA) for Kf and T7 polymerases, ThermoPol reaction buffer (20 mM Tris-HCl, 10 mM (NH₄)₂SO₄, 10 mM KCl, 2 mM MgSO₄, and 0.1% Triton X-100, pH 8.8) for Taq, Therminator, and Vent polymerases or pol β reaction buffer (50 mM Tris-HCl, pH 8.7, 10 mM MgCl₂, 100 mM KCl, 1 mM DTT, and 0.4 mg/mL acetylated BSA) for pol β . The reactions were initiated by adding a 5 μ L 2 \times dNTP solution to a 5 μ L solution containing the polymerase and primer–template and incubating at 25 (Kf and T7), 50 (Taq, Therminator, and Vent), or 37 $^{\circ}$ C (pol β) for 2–10 min and were then quenched with addition of 20 μ L of loading dye (95% formamide, 20 mM EDTA, and sufficient amounts of bromophenol blue and xylene cyanole). The reactions were analyzed by polyacrylamide gel electrophoresis, and a Phosphorimager (Molecular Dynamics) was used to quantify gel band intensities corresponding to the extended and unextended primer. The measured velocities were plotted against the concentration of dNTP and fit to the Michaelis–Menten equation to determine V_{\max} and K_M . k_{cat} was determined from V_{\max} by normalizing by the total enzyme concentration. Each reaction was run in triplicate and standard deviations determined.

Acknowledgment. Funding was provided by the National Institutes of Health (GM60005).

Supporting Information Available: Detailed steady-state kinetic data. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA803833H

(68) Goodman, M. F. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10493–10495.

(69) Kool, E. T. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 1–22.

(70) Bittker, J. A.; Phillips, K. J.; Liu, D. R. *Curr. Opin. Chem. Biol.* **2002**, *6*, 367–374.