

Optimization of Unnatural Base Pair Packing for Polymerase Recognition

Shigeo Matsuda, Allison A. Henry, and Floyd E. Romesberg*

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

Received November 6, 2005; E-mail: floyd@scripps.edu

Abstract: As part of an effort to expand the genetic alphabet, we have been examining the ability of predominately hydrophobic nucleobase analogues to pair in duplex DNA and during polymerase-mediated replication. We previously reported the synthesis and thermal stability of unnatural base pairs formed between nucleotides bearing simple methyl-substituted phenyl ring nucleobase analogues. Several of these pairs are virtually as stable and selective as natural base pairs in the same sequence context. Here, we report the characterization of polymerase-mediated replication of the same unnatural base pairs. We find that every facet of replication, including correct and incorrect base pair synthesis, as well as continued primer extension beyond the unnatural base pair, is sensitive to the specific methyl substitution pattern of the nucleobase analogue. The results demonstrate that neither hydrogen bonding nor large aromatic surface area is required for polymerase recognition, and that interstrand interactions between small aromatic rings may be optimized for replication. Combined with our previous results, these studies suggest that appropriately derivatized phenyl nucleobase analogues represent a promising approach toward developing a third base pair and expanding the genetic alphabet.

1. Introduction

An expanded genetic alphabet, which includes a third base pair to supplement the natural base pairs formed between guanosine and cytosine and adenine and thymine, would allow for a wide range of biotechnology applications, such as site-directed oligonucleotide labeling and *in vitro* selections with oligonucleotides bearing increased chemical diversity.¹ Additionally, a third base pair would lay the foundation for an organism with an expanded genetic code.² Efforts toward developing a third base pair have focused on nucleobase analogues designed to pair via orthogonal hydrogen bonding (H-bonding), based on work of the Benner group,³ and more recently on predominantly non-H-bonding analogues that pair via hydrophobic interactions, based on work of the Kool group.⁴ Pursuing the latter strategy, we⁵ and the Yokoyama and Hirao groups⁶ have shown that a wide variety of unnatural base pairs formed between identical nucleobase analogues (self-pairs) or

between different analogues (heteropairs) may be formed between hydrophobic nucleobase analogues that lack H-bonding potential and have little structural similarity to the natural purines or pyrimidines. In comparison to the stability of a natural base pair, the relative thermal stability of at least some of the predominantly hydrophobic unnatural base pairs was shown to result from a more favorable entropy change, implying that the classical hydrophobic effect contributes to their pairing.^{5k} As with proteins, hydrophobicity appears to be a suitable force to control molecular recognition within duplex DNA.

While hydrophobic interactions may be sufficient for base pair stability, it is unclear to what extent they can stabilize the transition states corresponding to polymerase-mediated DNA replication. In particular, the determinants of efficient and selective synthesis of an unnatural base pair, by insertion of the triphosphate opposite its cognate base in the template, and

- (1) Bittker, J.; Phillips, K. J.; Liu, D. R. *Curr. Opin. Chem. Biol.* **2002**, *6*, 367–374.
- (2) (a) Wang, L.; Magliery, T. J.; Liu, D. R.; Schultz, P. G. *J. Am. Chem. Soc.* **2000**, *122*, 5010–5011. (b) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498–500.
- (3) (a) Switzer, C. Y.; Moroney, S. E.; Benner, S. A. *J. Am. Chem. Soc.* **1989**, *111*, 8322–8323. (b) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33–37. (c) Piccirilli, J. A.; Moroney, S. E.; Benner, S. A. *Biochemistry* **1991**, *30*, 10350–10356. (d) Horlacher, J.; Hottiger, M.; Podust, V. N.; Huebscher, U.; Benner, S. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6329–6333.
- (4) (a) Moran, S.; Ren, R. X.-F.; Rumney, S.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056–2057. (b) Moran, S.; Ren, R. X.-F.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10506–10511. (c) Morales, J. C.; Kool, E. T. *Nat. Struct. Biol.* **1998**, *5*, 950–954. (d) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **1999**, *121*, 2323–2324. (e) Kool, E. T. *Curr. Opin. Chem. Biol.* **2000**, *4*, 602–608. (f) Kool, E. T. *Annu. Rev. Biochem.* **2002**, *71*, 191–219.

- (5) (a) McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **1999**, *121*, 11585–11586. (b) Ogawa, A. K.; Wu, Y.; McMinn, D. L.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 3274–3287. (c) Wu, Y.; Ogawa, A. K.; Berger, M.; McMinn, D. L.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 7621–7632. (d) Ogawa, A. K.; Wu, Y.; Berger, M.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 8803–8804. (e) Tae, E. L.; Wu, Y.; Xia, G.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2001**, *123*, 7439–7440. (f) Yu, C.; Henry, A. A.; Romesberg, F. E.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2002**, *41*, 3841–3844. (g) Berger, M.; Luzzi, S. D.; Henry, A. A.; Romesberg, F. E. *J. Am. Chem. Soc.* **2002**, *124*, 1222–1226. (h) Matsuda, S.; Henry, A. A.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2003**, *125*, 6134–6139. (i) Henry, A. A.; Yu, C.; Romesberg, F. E. *J. Am. Chem. Soc.* **2003**, *125*, 9638–9646. (j) Henry, A. A.; Olsen, A. G.; Matsuda, S.; Yu, C.; Geierstanger, B. H.; Romesberg, F. E. *J. Am. Chem. Soc.* **2004**, *126*, 6923–6931. (k) Matsuda, S.; Romesberg, F. E. *J. Am. Chem. Soc.* **2004**, *126*, 14419–14427.
- (6) (a) Mitsui, T.; Kimoto, M.; Sato, A.; Yokoyama, S.; Hirao, I. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4515–4518. (b) Mitsui, T.; Kimoto, M.; Harada, Y.; Sato, A.; Kitamura, A.; To, T.; Hirao, I.; Yokoyama, S. *Nucleic Acids Res. (supplement)* **2002**, *2*, 219–220.

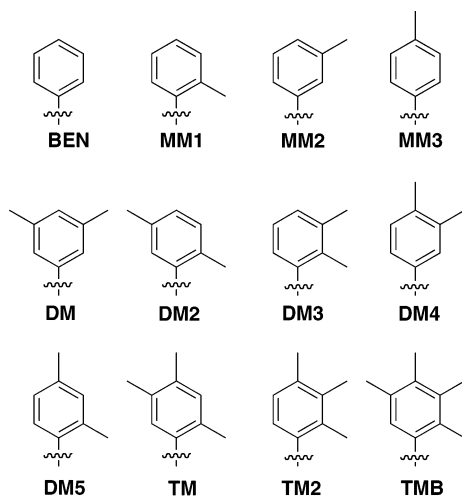


Figure 1. Unnatural nucleobases used in this study.

efficient continued primer extension of the nascent unnatural primer terminus are largely unknown. While it is now apparent that hydrophobic base pairs may be designed that are efficiently synthesized by DNA polymerases, replication is commonly limited by the competitive insertion of dATP or dTTP opposite the unnatural base in the template and by the efficient continued primer extension after unnatural base pair synthesis. However, the potential contribution of interbase hydrophobic interactions to fidelity and extension has not yet been systematically explored.

Within duplex DNA, aromatic nucleobase analogues may interact either via face packing, where one nucleobase interstrand intercalates between its partner and a flanking nucleobase, or via edge-to-edge packing, in a manner topologically similar to a natural base pair. In general, pairs formed between large aromatic nucleobase analogues appear to pair via intercalative face packing (D. Wemmer and F. Romesberg, unpublished results).⁷ After polymerase-mediated synthesis of the unnatural base pair, this mode of pairing at the nascent primer terminus is likely to induce structural distortions that might result in the observed poor extension rates. Thus, we hypothesized that nucleobase analogues with an aromatic surface area that is insufficient to cause interstrand intercalation, but that are optimized to interact edge-on, might form unnatural base pairs that are both efficiently synthesized and extended.

As a first step toward testing this hypothesis, we recently reported the systematic thermodynamic analysis of base pairs formed between 12 novel nucleotides bearing simple phenyl rings derivatized with methyl groups^{5k} (Figure 1). Surprisingly, we found that these rather simple unnatural nucleotides can form base pairs that are virtually as stable and selective as the natural pairs, despite lacking both H-bonding functionality and large aromatic surface area. We now report the complete kinetic analysis of these analogues as substrates for the exonuclease-deficient Klenow fragment of *E. coli* DNA polymerase I (Kf). We observed that methyl group substitution has a large effect on all steps of DNA synthesis, with the different substitution patterns resulting in pairs that are synthesized and extended with rate constants varying over 1–3 orders of magnitude. In

Table 1. Rates of Unnatural Self-Pair Synthesis^a

5'-d(TAATCGACTCACTATAGGGAGA) 3'-d(ATTATGCTGAGTGATATCCCTCTXGCTAGTTACGGCAGGATCGC)				
X	triphosphate	k_{cat} (min ⁻¹)	K_M (μM)	k_{cat}/K_M (min ⁻¹ M ⁻¹)
BEN	BEN	nd ^b	nd ^b	<1.0 × 10 ³
DM	DM	1.0 ± 0.1	359 ± 88	2.8 × 10 ³ ^c
DM2	DM2	0.85 ± 0.2	47 ± 11	1.8 × 10 ⁴
DM3	DM3	5.9 ± 0.6	54 ± 13	1.1 × 10 ⁵
DM4	DM4	2.8 ± 0.6	25 ± 3	1.1 × 10 ⁵
DM5	DM5	50 ± 4.6	25 ± 6	2.0 × 10 ⁶
TM	TM	31 ± 1.8	14 ± 3	2.2 × 10 ⁶ ^c
dT	dATP	163 ± 7 ^d	3.5 ± 1 ^d	4.7 × 10 ⁷ ^d

^a See Experimental Section for details. ^b Reaction was too inefficient for k_{cat} and K_M to be determined independently. ^c See ref 5b. ^d See ref 5a.

addition, polymerase-mediated mispair synthesis by incorporation of a natural triphosphate opposite an unnatural nucleobase analogue in the template reveals interesting trends that contribute to our understanding of unnatural base pair replication fidelity, as well as polymerase fidelity in general. Along with our previous studies, these data convincingly demonstrate that unnatural base pairs require neither H-bonds nor large aromatic surface for thermal stability and replication.

2. Results

2.1. Efficiency of Self-Pair Synthesis. We have chosen to examine the effect of methyl group substitution in the context of self-pairs. We have reported data on the synthesis of a wide variety of self-pairs which serve as a reference point for the current studies.⁵ It should also be noted that self-pairs do not limit the effort to expand the genetic alphabet, as a functional self-pair would more than double the number of codons available for encoding proteins, and in fact, the use of self-pair limits the potential mispairings with natural bases by a factor of 2. The second-order rate constants (i.e., k_{cat}/K_M), or efficiency, for Kf-mediated **DM** and **TM** self-pair synthesis were previously reported as 2.8×10^3 and $2.2 \times 10^6 \text{ M}^{-1} \text{ min}^{-1}$, respectively.^{5b} To further explore the effect of methyl group substitution, we examined the steady-state rates for **BEN**, **DM2**, **DM3**, **DM4**, and **DM5** self-pair synthesis (Table 1), which along with the **DM** and **TM** provide a systematic analysis of the effect of double and triple methyl group substitution. Interestingly, the rate constants varied by over 3 orders of magnitude, from below the limit of detection (<10³ M⁻¹ min⁻¹) for the **BEN** self-pair to greater than 10⁶ M⁻¹ min⁻¹ for the **DM5** and **TM** self-pairs. In general, there is no correlation between the rates and the extent of substitution. This is best illustrated by comparing the disubstituted self-pairs, **DM**, **DM2**, **DM3**, **DM4**, and **DM5**, which are synthesized with rates that vary over 3 orders of magnitude. Remarkably, the **DM5** and **TM** self-pairs are synthesized only about 20-fold less efficiently than a natural base pair in the same sequence context.^{5b} The efficiency of **DM5** and **TM** self-pair synthesis results predominantly from a large k_{cat} , which is 5–60-fold larger than that for the other self-pairs, and only 3–5-fold smaller than that for a natural base pair in the same sequence context.^{5b} The similar rates with which the **DM5** and **TM** self-pairs are synthesized suggest that substitution at the 2- and 4-positions is sufficient for efficient synthesis, while substitution at the 5-position is less important, highlighting the importance of the specific substitution pattern as opposed to the extent of substitution.

A more detailed analysis of how the substituents impact base pairing during replication requires an assumption about how

(7) (a) Henry, A. A.; Romesberg, F. E. *Curr. Opin. Biotechnol.* **2005**, *16*, 370–377. (b) Brotschi, C.; Mathis, G.; Leumann, C. J. *Chem.—Eur. J.* **2005**, *11*, 1911–1923.

Table 2. Rates of Correct Extension of Unnatural Self-Pairs^a

5'-d(TAATACGACTCACTATAGGGAGAX) 3'-d(ATTATGCTGAGTGATATCCCTCTYGCTAGGTTACGGCAGGATCGC)				
X	Y	k_{cat} (min ⁻¹)	K_M (μM)	k_{cat}/K_M (min ⁻¹ M ⁻¹)
BEN	BEN	0.18 ± 0.02	112 ± 20	1.6 × 10 ³
MM1	MM1	0.25 ± 0.07	173 ± 100	1.4 × 10 ³
MM2	MM2	nd ^b	nd ^b	<1.0 × 10 ³
MM3	MM3	0.67 ± 0.38	110 ± 56	6.1 × 10 ³
DM	DM	nd ^b	nd ^b	<1.0 × 10 ³
DM2	DM2	0.99 ± 0.08	154 ± 3	6.4 × 10 ³
DM3	DM3	nd ^b	nd ^b	<1.0 × 10 ³
DM4	DM4	1.8 ± 0.02	92 ± 28	2.0 × 10 ⁴
DM5	DM5	6.5 ± 1.1	161 ± 17	4.0 × 10 ⁴
TM	TM	7.9 ± 1.4	152 ± 32	5.2 × 10 ⁴
TM2	TM2	0.10 ± 0.002	25 ± 0.7	4.0 × 10 ³
TMB	TMB	nd ^b	nd ^b	<1.0 × 10 ³

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

the nucleobase analogues are orientated with respect to the C-glycosidic linkage. In principle, the analogues may exist in either a *syn*- or *anti*-orientation, defined by the 2-position relative to the C-glycosidic linkage. However, when there is a methyl group at the 2-position, the *syn*-orientation is unlikely to be stable due to eclipsing interactions between the methyl group and the ribosyl oxygen lone pair of electrons. To avoid this potential structural ambiguity, it is instructive to compare the rates for **DM2**, **DM3**, and **DM5** self-pair synthesis. These analogues are all locked into the *anti*-orientation by the methyl group at the 2-position and bear an additional substituent at one of the remaining three unique positions. Clearly, substitution at the 4-position is most favorable, followed by substitution at the 3-position, and then at the 5-position. The effects are substantial, resulting in a 2 order of magnitude variation in the rate of self-pair synthesis.

2.2. Efficiency of Self-Pair Extension. The step that consistently limits the synthesis of DNA containing unnatural self-pairs (or heteropairs) is continued primer elongation after incorporation of the unnatural triphosphate opposite its partner in the template (i.e., extension). We thus examined the efficiency of extension of all 12 self-pairs and found that methyl group substitution has a significant effect, with the second-order rate constants varying over 2 orders of magnitude (Table 2). The analogues can be roughly grouped into four categories. No extension could be detected with the **MM2**, **DM**, **DM3**, and **TMB** self-pairs ($k_{\text{cat}}/K_M < 10^3 \text{ M}^{-1} \text{ min}^{-1}$). These nucleobase analogues all have a methyl substituent at the 3-position. For the **BEN** and **MM1** self-pairs, extension proceeded with a k_{cat}/K_M of $\sim 1.5 \times 10^3 \text{ M}^{-1} \text{ min}^{-1}$. The similar extension rate of the **BEN** and **MM1** self-pairs suggests that a single methyl substituent at the 2-position has little effect on extension. The **MM3**, **DM2**, and **TM2** self-pairs were extended with second-order rate constants of approximately $5 \times 10^3 \text{ M}^{-1} \text{ min}^{-1}$. Finally, the **DM4**, **DM5**, and **TM** self-pairs were extended more efficiently than the other self-pairs (k_{cat}/K_M of $2\text{--}5 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$). **DM4**, **DM5**, and **TM** all have a methyl substituent at the 4-position in addition to at least one additional substituent at the 2- or 5-position. In comparison to the extension of a natural base pair, the extension of **DM4**, **DM5**, and **TM** is limited both by binding the next correct dNTP (K_M) and by reduced turnover of the complex (k_{cat}).

As with self-pair synthesis, the second-order rate constants for extension do not generally reflect hydrophobicity, as no

extension could be detected with the most substituted analogue (**TMB**). Analysis of the data reveals that substitution at the 4-position (compare **BEN** to **MM3**, **MM1** to **DM5**, **MM2** to **DM4**, **DM2** to **TM**, and **DM3** to **TM2**) is the most favorable, while substitution at the 3-position (compare **BEN** to **MM2**, **MM1** to **DM3**, **MM3** to **DM4**, and **DM5** to **TM2**) is the most unfavorable. Substituents at the 4-position may pack favorably with one another, resulting in a self-pair structure that is efficiently recognized and extended by Kf. Substituents at the 3-position may be oriented toward one another within the self-pair interface and introduce a steric clash that results in the self-pair adopting a geometry that is not well recognized by the polymerase. Generally, these effects result predominantly from changes in k_{cat} , suggesting that the interactions are manifest at the interbase interface of the developing transition state. An exception is the **DM4** self-pair. **DM4** differs from **MM3** by the presence of a methyl group at the 3-position, but forms a self-pair that is extended 3-fold faster than the self-pair formed by **MM3**. This may be due to the **DM4** analogue adopting an orientation that positions the methyl substituent at the 3-position away from the interface with the pairing analogue.

A methyl group at the 2-position had a more variable effect, depending on the specific substitution pattern of the nucleobase analogue. Addition of a substituent at the 2-position of a ring that already possessed a methyl substituent at the 4- or 5-position increased the efficiency of extension (compare **MM2** to **DM2**, **MM3** to **DM5**, and **DM4** to **TM**). **DM4** again provided an exception. Addition of a methyl group at the 2-position of **DM4**, resulting in **TM2**, slightly decreased the efficiency of extension. Again, this may be due to the **DM4** analogues adopting an orientation that positions the methyl group at the 3-position away from the interbase interface, while the 2-methyl group of **TM2** forces the analogues into an *anti*-orientation that positions the substituent at 3-position into the interbase interface and results in a primer terminus geometry that is less well recognized by Kf. However, addition of a substituent at the 2-position of the parent phenyl ring (compare **BEN** to **MM1**) or a ring that already possessed methyl substituent at the 3-position (compare **MM2** to **DM3** and **DM4** to **TM2**) resulted in no detectable change or a decrease in the extension rate.

2.3. Efficiency of Polymerase-Mediated Mispairing of Natural dNTPs with Nucleobase Analogues in the Template. The efficiency of Kf-mediated mispair synthesis was examined by characterizing the rates at which the natural triphosphates were inserted opposite each of the 12 analogues in the template (Table 3). In general, dCTP and dGTP are the least efficiently inserted, with second-order rate constants less than $5 \times 10^3 \text{ M}^{-1} \text{ min}^{-1}$. The second most efficiently inserted natural triphosphate was dTTP. Interestingly, in this case, the efficiencies are at least roughly correlated with the number of methyl groups that are expected to be presented to the incoming thymidine. **BEN**, **MM1**, **MM2**, **MM3**, **DM**, **DM2**, and **DM5** can each present zero or one substituent to the incoming thymidine. Insertion of dTTP opposite each of these analogues proceeds with k_{cat}/K_M of 3.6×10^3 to $1.6 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$. In the case of **DM3**, **DM4**, or **TM**, each analogue may present two methyl groups to the incoming dTTP, and insertion proceeded with a k_{cat}/K_M of 3.0×10^4 to $9.0 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$. Finally, the most substituted interfaces, presented by **TM2** and **TMB**, template addition of dTTP with a k_{cat}/K_M of 1.5×10^5 and 4.1×10^5

Table 3. Incorporation of Natural Triphosphates Opposite Unnatural Bases in the Template^a

		5'-d(TAATACGACTCACTATAGGGAGA) 3'-d(ATTATGCTGAGTGATATCCCTCTAGGTTACGGCAGGATCGC)		
X	triphosphate	k_{cat} (min ⁻¹)	K_M (μM)	k_{cat}/K_M (min ⁻¹ M ⁻¹)
BEN	A	21 ± 4.5	61 ± 11	3.4 × 10 ⁵
	C	0.34 ± 0.01	328 ± 41	1.0 × 10 ³
	G	0.09 ± 0.01	49 ± 13	1.8 × 10 ³
	T	0.70 ± 0.10	192 ± 2	3.6 × 10 ³
MM1	A	72.8 ± 18.1	25.4 ± 3.3	2.9 × 10 ⁶
	C	0.18 ± 0.02	81 ± 69	2.2 × 10 ³
	G	nd ^b	nd ^b	<1.0 × 10 ³
	T	0.56 ± 0.15	135 ± 21	4.1 × 10 ³
MM2	A	27.2 ± 12.4	56.5 ± 8.7	4.8 × 10 ⁵
	C	nd ^b	nd ^b	<1.0 × 10 ³
	G	nd ^b	nd ^b	<1.0 × 10 ³
	T	1.3 ± 0.47	203 ± 34	6.4 × 10 ³
MM3	A	14.2 ± 5.79	36.2 ± 2.98	3.9 × 10 ⁵
	C	0.36 ± 0.04	303 ± 69	1.2 × 10 ³
	G	nd ^b	nd ^b	<1.0 × 10 ³
	T	1.28 ± 0.45	123 ± 14	1.0 × 10 ⁴
DM^c	A	1.1 ± 0.1	75 ± 15	1.5 × 10 ⁴
	C	0.68 ± 0.03	307 ± 28	2.2 × 10 ³
	G	nd ^b	nd ^b	<1.0 × 10 ³
	T	2.9 ± 0.5	182 ± 30	1.6 × 10 ⁴
DM2	A	42 ± 10	10 ± 2	4.2 × 10 ⁶
	C	0.37 ± 0.17	83 ± 33	4.5 × 10 ³
	G	nd ^b	nd ^b	<1.0 × 10 ³
	T	1.5 ± 0.34	143 ± 6	1.0 × 10 ⁴
DM3	A	9.9 ± 1.8	28 ± 5	3.5 × 10 ⁵
	C	0.43 ± 0.08	297 ± 36	1.4 × 10 ³
	G	nd ^b	nd ^b	<1.0 × 10 ³
	T	8.4 ± 4.8	93 ± 2	9.0 × 10 ⁴
DM4	A	9.1 ± 2.3	34 ± 8	2.7 × 10 ⁵
	C	1.7 ± 0.36	258 ± 12	6.6 × 10 ³
	G	0.06 ± 0.02	51 ± 7	1.2 × 10 ³
	T	7.2 ± 1.7	113 ± 20	6.4 × 10 ⁴
DM5	A	30 ± 1.8	20 ± 3.9	1.5 × 10 ⁶
	C	nd ^b	nd ^b	<1.0 × 10 ³
	G	0.22 ± 0.08	208 ± 7	1.1 × 10 ³
	T	1.7 ± 0.39	169 ± 49	1.0 × 10 ⁴
TM^c	A	6.6 ± 0.2	26 ± 5	7.6 × 10 ⁵
	C	0.18 ± 0.04	381 ± 35	4.7 × 10 ²
	G	0.07 ± 0.01	140 ± 25	5.0 × 10 ²
	T	6.9 ± 0.6	227 ± 42	3.0 × 10 ⁴
TM2	A	16 ± 4.7	38 ± 14	4.2 × 10 ⁵
	C	2.3 ± 0.19	300 ± 19	7.7 × 10 ³
	G	0.13 ± 0.03	117 ± 49	1.1 × 10 ³
	T	24 ± 0.61	156 ± 14	1.5 × 10 ⁵
TMB	A	83 ± 11	29 ± 4.8	2.9 × 10 ⁶
	C	nd ^b	nd ^b	<1.0 × 10 ^{3d}
	G	0.08 ± 0.01	54 ± 22	1.5 × 10 ³
	T	42 ± 3.1	103 ± 20	4.1 × 10 ⁵

^a See Experimental Section for details. ^b Reaction was too inefficient for k_{cat} and K_M to be determined independently. ^c See ref 5b. ^d Due to rapid mispair extension, the rate of dCTP insertion opposite **TMB** was measured with a template modified to contain a dA as the next 5' base in the template.

M⁻¹ min⁻¹, respectively. This dependence of the second-order rate constant on the number of methyl groups results predominantly from an increased k_{cat} . In fact, the most substituted analogues show k_{cat} values that are only 4–6-fold reduced relative to that for a natural base pair in the same sequence context.^{5b}

The most efficiently inserted triphosphate opposite each unnatural base in the template is dATP. Interestingly, the insertion efficiencies generally fall into one of two groups, those in the range of 2.7 × 10⁵ to 4.8 × 10⁵ M⁻¹ min⁻¹ (**BEN**, **MM2**, **MM3**, **DM3**, **DM4**, **TM**, and **TM2**) or 1.5 × 10⁶ to 4.2 × 10⁶ M⁻¹ min⁻¹ (**MM1**, **DM2**, **DM5**, and **TMB**). An exception is **DM**, opposite which dATP is inserted with a second-order rate constant of 1.5 × 10⁴ M⁻¹ min⁻¹. Comparing **BEN** to **MM1**,

MM2, and **MM3**, it is apparent that the only single substitution that significantly affects dATP insertion is at the 2-position, which increases the second-order rate constant by about an order of magnitude. This results from a large increase in k_{cat} and a small decrease in K_M .

Comparing **BEN** to **DM** indicates that a methyl group at the 3-position decreases the efficiency of dATP insertion. **MM2** and **DM4** are presumably oriented such that these methyl groups are directed away from the incoming nucleobase, and they likely template dATP in a manner similar to **BEN**. The effect of a methyl group at the 2- and 3-positions appears to approximately cancel with **DM3**. The similar rates at which dATP is inserted opposite **MM1**, **DM2**, and **DM5** further support the suggestion that substituents at the 2- and 3-positions are generally the only ones that significantly impact dATP insertion within this scaffold. An exception is the insertion of dATP opposite **TMB**, where the additional two methyl groups (relative to **DM3**) increase the efficiency of dATP insertion by an order of magnitude. Remarkably, dATP is inserted opposite **MM1**, **DM2**, **DM5**, or **TMB** with a second-order rate constant that is only approximately an order of magnitude reduced relative to the rate at which dATP is inserted opposite dT in the template. As with **MM1** relative to **BEN**, the increased rates of dATP insertion opposite **DM2**, **DM5**, and **TMB**, relative to the other analogues, result mostly from an increase in k_{cat} and a small decrease in K_M . Remarkably, with each of these four analogues, the observed k_{cat} 's are within 3–9-fold of that for natural synthesis, and the K_M 's are within 2–7-fold.^{5b} When these rates are compared to those for the insertion of the other natural dNTPs, it is apparent that **MM1**, **DM2**, and **DM5** (and to a lesser extent, **TMB**) are functional mimics of dT, directing the efficient and selective insertion of dATP.

3. Discussion

Efforts to develop an orthogonal third base pair are expected to be facilitated by a general understanding of the determinants that underlie DNA duplex stability and recognition by DNA polymerases. With simple methyl-derivatized phenyl nucleoside analogues, we previously reported a systematic evaluation of the contribution of nucleobase shape and hydrophobicity to pairing stability.^{5k} Surprisingly, we found that despite a lack of H-bonding capacity and a significantly reduced aromatic surface area relative to a natural base pair, these small nucleobase analogues can form unnatural base pairs with high stability and selectivity. Using these same analogues, we now report the systematic evaluation of nucleobase shape and hydrophobicity to polymerase recognition. We examined base pair synthesis in the context of self-pairs, which are formed between two identical nucleobases. While this allows us to systematically address the effects of nucleobase modification, it is also of practical utility as self-pairs offer the simplest route to the expansion of the genetic code, reducing the number of potential mispairs with the natural bases by a factor of 2.

There are three facets to the synthesis of DNA containing an unnatural self-pair: first, the pair must be efficiently synthesized by insertion of the unnatural triphosphate opposite the unnatural base in the template; second, insertion of a natural dNTP must not be competitive; and finally, the self-pair must be efficiently extended. In this work, we examined all of these facets in detail for each analogue. We showed that methyl substitution has a

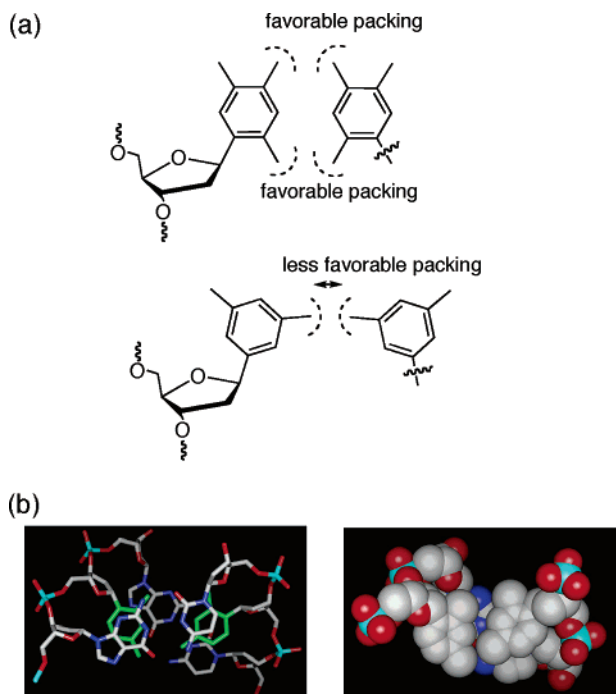


Figure 2. (a) Model of pairing with hydrophobic base pairs. (b) Predicted structure of **DM5** self-pair.^{5k}

pronounced effect on the rates of self-pair synthesis. Overall, the specific methyl group substitution pattern appears to be more important than hydrophobicity. This is most evident from an examination of the rates at which the disubstituted analogues (**DM**, **DM2**, **DM3**, **DM4**, and **DM5**) form self-pairs. Despite having similar surface area and hydrophobicity, the self-pair synthesis rates vary over 3 orders of magnitude. Substitution at the 2- and 4-positions appears to be sufficient for efficient synthesis. Substitution at the other positions has little or a deleterious effect. Synthesis of the **DM5** and **TM** self-pairs is only approximately an order of magnitude less efficient than that of a natural base pair in the same sequence context.^{5b} This is remarkable considering that these pairs do not form H-bonds, either between the nucleobase analogues or to the polymerase, and that they have significantly reduced aromatic surface area relative to a natural purine–pyrimidine base pair. In addition to being most efficiently synthesized, the **DM5** self-pair is also the most stable of this series of analogues in duplex DNA.^{5k} However, the rates for self-pair synthesis do not generally reflect base pair stability as measured by the duplex melting temperature (see Supporting Information). Rather, the efficiencies result predominantly from large k_{cat} 's. Apparently, the methyl groups at the 2- and 4-positions form a stable interface between the analogues in the rate-determining transition state, methyl groups at the 3-position are less stabilizing, presumably due to increased eclipsing interactions, and substituents at the 5-position are oriented away and do not contribute to the interface (Figure 2).

Evaluation of the rates with which polymerases insert natural dNTPs opposite an unnatural base in the template allows for an assessment of unnatural base pair synthesis fidelity. Additionally, it allows for the systematic evaluation of the interactions that mediate natural base pair synthesis. Interestingly, the methyl group substitution pattern has a very different effect for each natural dNTP. Generally, dGTP and dCTP, the

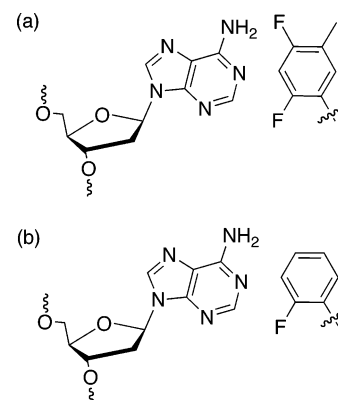


Figure 3. (a) dA:F base pair and (b) dA:2FB base pair.

most hydrophilic of the natural dNTPs,⁸ are not inserted well opposite any of the 12 unnatural bases in the template. This suggests that desolvation may contribute to mispair synthesis. dTTP is inserted with a rate that at least roughly parallels the extent of methyl group derivatization, with less dependence on the specific substitution pattern. While the K_M 's vary by approximately a factor of 2, the k_{cat} 's vary by 75-fold, suggesting that the increased hydrophobicity stabilizes the transition state rather than a Michaelis-like complex, perhaps by improving the orientation of the scissile phosphate bond relative to the nucleophilic 3'OH at the primer terminus.

Of the four natural triphosphates, dATP is consistently the most efficiently inserted triphosphate opposite the unnatural analogues in the template. However, unlike insertion of dTTP, these rates do not appear to be dependent on hydrophobic surface area, but instead are dependent on the specific substitution pattern. Generally, addition of a methyl group at the 2-position of the unnatural nucleobase favors insertion of dATP, while substitution at the 3-position disfavors it. The efficient insertions appear to result predominantly from an elevated k_{cat} and, to a lesser extent, from a reduced K_M . Remarkably, dATP is inserted opposite **MM1**, **DM2**, **DM5**, or **TMB** with a second-order rate constant that is only approximately an order of magnitude reduced relative to the rate at which it is inserted opposite dT in the template.^{5b} These results suggest that overall shape mimicry of dT (i.e., shape complementarity) is less important than the presence of a single appropriately placed substituent that is capable of packing with adenine in the developing transition state, likely at the hydrophobic methine moiety. This hypothesis is supported by data from the literature. For example, the efficient incorporation of dATP opposite the dT shape mimic, **F**, in the template is one of the central observations supporting the importance of shape complementarity (Figure 3).^{4a,b} However, **2FB** templates dATP with a similar efficiency,^{5j} suggesting that the fluorine substituent at the 2-position is sufficient for recognition of dATP. The importance of specific electrostatic and packing interactions, as opposed to shape complementarity, is also supported by recent studies that have shown that the efficiency of polymerase-mediated replication is better correlated with electronic properties of the nucleobase than with the shape.^{5c,h-j,9}

Mispairing with dT, dC, or dG does not significantly compromise self-pair synthesis fidelity. However, dATP inser-

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

tion opposite the unnatural base pairs in the template proceeds with rates similar to or only marginally reduced relative to those for the self-pairs. Thus, mispairing with dA compromises DNA synthesis fidelity with these self-pairs. While methyl group substitution at the 2-position increases the rates at which both the self-pairs and the mispairs with dA are synthesized, substitution at the 3- and 4-position favors self-pair synthesis, but disfavors (3-position) or has little effect (4-position) on the rate at which the mispair is synthesized. Thus, orthogonality against dT, dC, and dG appears to be intrinsic to these small hydrophobic scaffolds, while orthogonality against dA may be achieved by judicious placement of methyl groups.

Perhaps most importantly from the perspective of the effort to develop unnatural base pairs, methyl group substitution was found to have a significant effect on the self-pair extension rates. The observed rates do not parallel either the previously reported thermal stabilities (Supporting Information) or the hydrophobicity of the nucleobase analogues, suggesting that the specific substitution pattern is important. The most efficiently extended self-pairs are those formed by **DM4**, **DM5**, and **TM**, all of which have a methyl group at the 4-position and at least one additional methyl group. While these self-pairs are extended significantly less efficiently than are the natural base pairs, they are extended significantly faster than the other analogues and only approximately 10-fold slower than the most efficiently extended unnatural base pairs identified to date.^{5j} These results suggest that, with continued optimization, analogues based on this nucleobase scaffold may represent attractive third base pair candidates. For example, on the basis of similar analogues, addition of a fluorine substituent to the 3-position of **DM5** or **TM** is predicted to further increase the extension rates, as already reported for the parent phenyl scaffold.^{5j}

It is interesting to speculate about how nucleobase analogues with such limited surface area and no ability to form H-bonds can be recognized by DNA polymerases. While dipole and dipole-induced dipole interactions are all known to contribute to stable intrastrand stacking in natural DNA,¹⁰ these analogues have neither large permanent dipole moments nor significant polarizability. Thus, it seems unlikely in these cases that these electrostatic interactions contribute to the differences in self-pair synthesis, fidelity, or extension. Rather, specific interbase pair packing interactions that develop in the transition state appear to underlie recognition of the analogues by Kf polymerase. This is consistent with previous studies of the **DM5** self-pair, which modeling studies predicted to be well accommodated in B-form DNA with the analogues positioned in the same plane and interacting with one another in an edge-on manner.^{5k}

It is reasonable that hydrophobic packing may replace interbase H-bonding, as both the donors and acceptors are removed and thus no desolvation is required and no hydrophilic moieties are buried in the hydrophobic core of the duplex. However, interactions between the nucleobases and the polymerase seem more problematic as compensatory changes have

not been made in the protein. These interactions might be expected to be especially important for base pair extension since structural and biochemical studies have identified important H-bonds between polymerase H-bond donors and nucleobase H-bond acceptors in the developing minor groove (N3 of purines and O2 of pyrimidines).¹¹ For example, the contribution of this H-bond has been investigated by replacing dG or dA at the primer terminus with 3-deazaguanine or 3-deazaadenine, respectively.¹² These modified primer termini were extended with a steady-state efficiency of $\sim 10^5 \text{ M}^{-1} \text{ min}^{-1}$. Remarkably, the **DM5** and **TM** self-pairs are extended with similar rates. Perhaps, these rates reflect the upper limits of Kf-mediated extension of a base pair that has suitable interbase interactions, mediated by either H-bonds or re-engineered with optimized packing interactions, but that does not engage the polymerase with an H-bond at the primer terminus. This suggests that the inclusion of a suitably positioned H-bond acceptor in the unnatural nucleobase scaffold might yield further improvements in unnatural base pair extension.^{4d}

We have previously shown that unnatural nucleotides bearing simple methyl-substituted benzene rings as nucleobase analogues may form base pairs that are virtually as stable as natural pairs, despite possessing neither H-bonds nor large aromatic surface area. We have now shown that these analogues may also be optimized for polymerase-mediated replication.⁵ These smaller nucleobases are not expected to induce distortions in duplex DNA, including those at the primer terminus that appear to limit replication of the larger unnatural base pairs. These nucleobase analogues will likely serve as scaffolds for further modification and optimization as unnatural base pairs. We are now focused on identifying combinations of methyl group substitutions and heteroatom derivatizations that will impart these smaller unnatural nucleobases with further improvements in stability and replication.

4. Experimental Section

General Methods. Chemical reagents were purchased from Sigma-Aldrich and used without further purification, unless otherwise stated. All reagents for oligonucleotide synthesis were purchased from Glen Research. ³¹P NMR spectra were recorded on a Bruker AMX-400 spectrometer. Coupling constants (*J* values) are reported in hertz. The chemical shifts are given in δ (parts per million) using 85% H₃PO₄ in D₂O for ³¹P NMR as an external standard. T4 polynucleotide kinase and Klenow fragment exo⁻ were purchased from New England Biolabs. [γ -³³P]-ATP was purchased from Amersham Biosciences.

Synthesis of Oligonucleotides. All unnatural nucleosides and nucleotides used in this study were synthesized as previously reported.^{5k} Oligonucleotides were prepared by the β -cyanoethylphosphoramidite method on controlled pore glass supports (1 mmol) using an Applied Biosystems Inc. 392 DNA/RNA synthesizer as standard method. After automated synthesis, the oligonucleotides were cleaved from the support by concentrated aqueous ammonia for 1 h at room temperature, deprotected by heating at 55 °C for 12 h, and purified by denaturing polyacrylamide gel electrophoresis (12–20%, 8 M urea). The primer oligonucleotides containing unnatural bases at the 3'-end were obtained using Universal Support, or 3'-phosphate CPG, which was treated with alkaline phosphatase after deprotection. The oligonucleotides were purified by PAGE, visualized by UV shadowing, and recovered by electroelution. After ethanol precipitation, the concentration of oligonucleotides was determined by UV/vis absorption.

- (9) (a) Chiramonte, M.; Moore, C. L.; Kincaid, K.; Kuchta, R. D. *Biochemistry* **2003**, *42*, 10472–10481. (b) Paul, N.; Nashine, V. C.; Hoops, G.; Zhang, P.; Zhou, J.; Bergstrom, D. E.; Davisson, V. J. *Chem. Biol.* **2003**, *10*, 815–825. (c) Adelfinskaya, O.; Nashine, V. C.; Bergstrom, D. E.; Davisson, V. J. *J. Am. Chem. Soc.* **2005**, *127*, 16000–16001. (d) Zhang, X.; Lee, I.; Berdis, A. *Biochemistry* **2005**, *44*, 13101–13110.
- (10) Saenger, W. *Principles of Nucleic Acid Structure*; Springer: New York, 1984; pp 105–158.

- (11) Li, Y.; Korolev, S.; Waksman, G. *EMBO J.* **1998**, *17*, 7514–7525.

General Triphosphate Synthesis Procedure. Proton sponge (1.5 equiv) and unnatural nucleoside (1 equiv) were dissolved in trimethyl phosphate (final concentration ~ 0.3 M) and cooled to 0 °C. POCl_3 (1.05 equiv) was added dropwise, and the mixture was stirred at 0 °C for 2 h. Tributylamine (5 equiv) was added, followed by a solution of tributylammonium pyrophosphate (5 equiv) in DMF (final concentration ~ 0.15 M). After 3 min, the reaction was quenched by addition of 1 M aqueous triethylammonium bicarbonate (10 vol equiv). The resulting crude solution was stirred for 30 min at 0 °C and then lyophilized. The crude material was purified by reverse phase HPLC (C18 column, 1–35% CH_3CN in 0.1 M $\text{NEt}_3\text{-HCO}_3$, pH 7.5) followed by lyophilization to afford the triphosphate as a white solid. **BEN** triphosphate¹³ and **DM** triphosphate^{5b} were synthesized as described previously. **DM2** triphosphate: ^{31}P NMR (140 MHz, D_2O) δ -5.91 (d, $J = 18.3$ Hz), -10.59 (d, $J = 17.1$ Hz), -22.16 (t, $J = 17.8$ Hz). **DM3** triphosphate: ^{31}P NMR (140 MHz, D_2O) δ -5.91 (d, $J = 17.2$ Hz), -10.45 (d, $J = 17.2$ Hz), -21.78 (t, $J = 17.8$ Hz). **DM4** triphosphate: ^{31}P NMR (140 MHz, D_2O) δ -5.91 (d, $J = 18.1$ Hz), -10.58 (d, $J = 17.2$ Hz), -22.16 (t, $J = 17.7$ Hz). **DM5** triphosphate: ^{31}P NMR (140 MHz, D_2O) δ -5.93 (d, $J = 18.5$ Hz), -10.57 (d, $J = 17.1$ Hz), -22.18 (t, $J = 17.8$ Hz).

- (12) (a) Hendrickson, L. H.; Devine, K. G.; Benner, S. A. *Nucleic Acids Res.* **2004**, *32*, 2241–2250. (b) Spratt, T. E. *Biochemistry* **2001**, *40*, 2647–2652. (c) McCain, M. D.; Meyer, A. S.; Schultz, S. S.; Glekas, A.; Spratt, T. E. *Biochemistry* **2005**, *44*, 5647–5659. (d) Guo, M. J.; Hildbrand, S.; Leumann, C. J.; McLaughlin, L. W.; Waring, M. J. *Nucleic Acids Res.* **1998**, *26*, 1863–1869.
- (13) (a) Lai, J. S.; Kool, E. T. *Chem.—Eur. J.* **2005**, *11*, 2966–2971. (b) Aketani, S.; Tanaka, K.; Yamamoto, K.; Ishihama, A.; Cao, H.; Tengeiji, A.; Hiraoka, S.; Shiro, M.; Shionoya, M. *J. Med. Chem.* **2002**, *45*, 5594–5603.

Gel-Based Kinetic Assay. Primers and templates were chosen according to which facet of replication was being examined, as shown in Tables 1 and 2. Primer oligonucleotides were 5'-radiolabeled with T4 polynucleotide kinase and $[\gamma\text{-}^{33}\text{P}]\text{-ATP}$. Primers were annealed to template oligonucleotides in the reaction buffer by heating to 90 °C followed by slow cooling to ambient temperature. Assay conditions included 40 nM primer/template, 0.1–1.3 nM enzyme, 50 mM Tris-HCl, pH 7.5, 10 mM MgCl_2 , 1 mM DTT, and 50 $\mu\text{g/mL}$ acetylated BSA. The reactions were carried out by combining the DNA–enzyme mixture with an equal volume (5 μL) of 2 \times dNTP stock solution, incubating at 25 °C for 1–10 min, and quenching by the addition of 20 μL of loading dye (95% formamide, 20 mM EDTA, and sufficient amounts of bromophenol blue and xylene cyanole). The reaction mixtures were resolved by 15% polyacrylamide and 8 M urea denaturing gel electrophoresis, and the radioactivity was quantified by means of a PhosphorImager (Molecular Dynamics) and ImageQuant software. A plot of k_{obs} versus triphosphate concentration was fit to a Michaelis–Menten equation using the program Kaleidagraph (Synergy Software). The data presented are averages of three independent determinations.

Acknowledgment. This work was supported by the National Institutes of Health (2R01 GM60005).

Supporting Information Available: Representative kinetic data and plot of kinetic data versus thermodynamic data. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA057575M