

An Efficiently Extended Class of Unnatural Base Pairs

Aaron M. Leconte, Shigeo Matsuda, and Floyd E. Romesberg*

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

Received February 4, 2006; E-mail: floyd@scripps.edu

The four natural nucleotides of DNA form two base pairs that are capable of encoding complex genetic information as well as functional nucleic acids. Expansion of the genetic alphabet to include a third base pair, formed between two identical or different unnatural nucleotides, referred to as self-pairs and heteropairs, respectively, would expand the informational and functional potential of DNA. While the pairing of the natural nucleobases is mediated by hydrogen-bonding (H-bonding) complementarity, it has become clear that, like proteins, hydrophobic forces are also capable of mediating these interactions.¹ Indeed, a variety of nucleotides bearing predominantly hydrophobic nucleobase analogues form pairs that are both stable in duplex DNA and synthesized by the exonuclease-deficient Klenow fragment of DNA Pol I (Kf), by incorporation of an unnatural dNTP opposite the unnatural base in the template, with reasonable efficiency and selectivity.^{1b,2} However, extension of the newly synthesized unnatural base pair, by incorporation of the next correct dNTP, has consistently limited the enzymatic synthesis of unnatural DNA. Only recently have efforts begun to focus on the determinants of efficient extension;^{1b,3} however, they remain largely unknown. This has limited the rational design of viable unnatural base pairs.

We have synthesized over 80 unnatural nucleotides bearing predominantly hydrophobic nucleobase analogues and systematically evaluated the synthesis and extension of the corresponding self-pairs. However, because these nucleotides may be combined to form more than 6400 heteropairs, their synthetic evaluation requires a more efficient screening strategy. In an attempt to identify more efficiently extended heteropairs, as well as to begin to understand the determinants of extension, we developed and performed a simple screen of the unnatural nucleotides to identify those that formed heteropairs which Kf efficiently extends.

Each nucleobase was incorporated at the 3'-end of a 24-mer primer oligonucleotide, as well as at position 24 of a 45-mer template oligonucleotide. Hybridization of these two oligonucleotides forms a polymerase substrate with a primer that terminates at the X:Y heteropair, where X and Y are the unnatural nucleotides in the primer and template, respectively. In all cases, the next nucleobase in the template is dG; incorporation of dCTP results in correct extension of the unnatural heteropair. Sixty primers were organized into 10 interrelated groups based on their shape and functionality. In each experiment, a single group, consisting of a pool of 4–10 radiolabeled primers, was annealed to a common template strand, producing heteropairs differentiated by the nucleotide X at the primer terminus. Kf (0.3 nM) and dCTP (400 μ M) were added in reaction buffer and allowed to incubate with the unnatural DNA for 5 min before quenching with EDTA. The amount of the primer pool extended by the addition of dCTP was quantified by PAGE and phosphorimaging. While the detailed analysis of the results will be published elsewhere, it was immediately apparent that one group of primers, the bicyclic 5/6

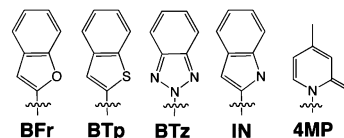


Figure 1. Unnatural nucleobases.

Table 1. Steady-State Rate Constants of Unnatural Terminus Extension by Insertion of dCTP^a

5' - dTAATACGACTCACTATAGGGAGA (X)				
3' - dATTATGCTGAGTGATATCCCTCT (Y) GCTAGGTTACGGCAGGATCGC				
X	Y	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M ($\text{M}^{-1}\text{min}^{-1}$)
BFr	4MP	5.4 ± 1.2	6.1 ± 2.1	8.8×10^5
BTp	4MP	3.3 ± 1.0	11.5 ± 1.3	2.9×10^5
BTz	4MP	5.4 ± 1.7	7.6 ± 2.2	7.1×10^5
IN	4MP	5.3 ± 1.2	3.9 ± 0.9	1.4×10^6
4MP	BFr	30.9 ± 6.5	226 ± 16	1.4×10^5
4MP	BTp	19.8 ± 5.1	205 ± 39	9.7×10^4
4MP	BTz	14.9 ± 3.3	89 ± 23	1.7×10^5
4MP	IN	15.8 ± 4.1	402 ± 108	3.9×10^4

^a See Supporting Information for experimental details.

series,⁴ including **BFr**, **BTp**, **BTz**, and **IN**, was extended efficiently when paired with 4-methyl pyridone (**4MP**) in the template (Figure 1). We were thus interested in the more detailed characterization of this class of unnatural heteropair.

To quantitate the efficiency with which Kf extends the **BFr:4MP**, **BTp:4MP**, **BTz:4MP**, and **IN:4MP** heteropairs, we employed steady-state kinetics (Table 1). In this context, with the bicyclic nucleobase analogue in the primer and **4MP** in the template, the second-order rate constants range between 3×10^5 and $1 \times 10^6 \text{ M}^{-1} \text{ min}^{-1}$. The most efficiently extended pairs, **BFr:4MP**, **BTz:4MP**, and **IN:4MP**, are extended at rates that are 2–4-fold higher than those of the most efficiently extended unnatural base pair reported to date, and the **BTp:4MP** heteropair was extended only slightly less efficiently. The efficient extension rates are predominantly due to low K_M values, which range from 4 to 12 μM and, remarkably, are all within a factor of 10 of that typical for the extension of a natural base pair (3.5 μM).² The K_M values for the corresponding self-pairs are all significantly higher, ranging between 85 and 294 μM , indicating that the tight binding of dCTP is specific to the heteropairs (Supporting Information). The data suggest that, in this sequence context, the heteropair forms a specific primer terminus that tightly binds the incoming dNTP.

Viable hetero base pair candidates must be efficiently extended in both sequence contexts. With **4MP** in the primer and the bicyclic analogue in the template, each heteropair is also extended efficiently, with second-order rate constants ranging from 4×10^4 to $2 \times 10^5 \text{ M}^{-1} \text{ min}^{-1}$ (Table 1). In stark contrast to the data discussed above, efficient extension of the unnatural heteropairs in this context results primarily from large k_{cat} values, which range from 15 to 31 min^{-1} . Remarkably, these k_{cat} 's are all within 10-fold of that for the

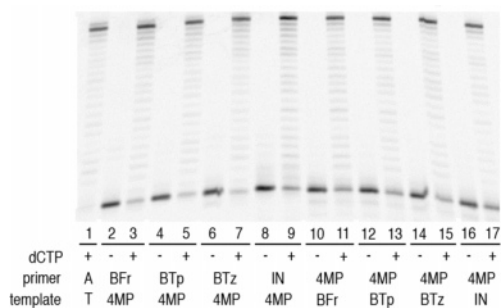


Figure 2. Full-length extension of DNA termini. All reactions contain 50 μM dATP, dGTP, and dTTP. Odd lanes also contain 50 μM dCTP. Sequences are identical to Table 1. See Supporting Information for full experimental details.

extension of a natural base pair (163 min^{-1}).² With the exception of the **BTz** self-pair, which is extended at a rate of $2 \times 10^5 \text{ M}^{-1} \text{ min}^{-1}$, the rates in either context are significantly faster than any of the rates of extension of the corresponding self-pairs, which range from 7×10^2 to $1 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$. This again indicates that efficient extension is generally specific for the heteropairs. In addition, the data suggest that, in this sequence context, the heteropair forms a primer terminus that interacts favorably with the incoming dCTP in the developing transition state.

The efficient extension of the heteropairs must not come at the expense of fidelity. To characterize the fidelity with which each heteropair is extended, we examined extension of all eight unnatural termini with dATP, dGTP, and dTTP. Importantly, the rates of misincorporation of these nucleobases opposite dGTP are all below detectable rates ($k_{\text{cat}}/K_{\text{M}} < 10^3 \text{ M}^{-1} \text{ min}^{-1}$). Thus, the unnatural heteropairs are extended with high efficiency and fidelity.

Replicative DNA polymerases, such as Kf, interact with up to five base pairs upstream of the primer terminus, and it has been demonstrated that modified nucleobases at these upstream positions may have an effect on replication.⁵ To determine if the heteropairs compromise full length synthesis, we examined synthesis in the presence of all dNTPs (Figure 2). Kf (3.7 nM), dNTPs (50 μM), and primer/template (40 nM) were incubated for 5 min in the presence or absence of dCTP. After 5 min of incubation, full length synthesis was the predominant product in the presence of dCTP, but only unextended primer was apparent in the absence of dCTP. While the presence of a small amount of unextended unnatural termini indicates that the reaction is less efficient than wild type, the data clearly demonstrate that full length synthesis is both efficient and selective.

The origins of the efficient extension of the bicyclic pyridone heteropairs likely lie in specific interactions between the unnatural base pair, the polymerase, and the incoming dNTP. For the efficient extension of a natural Watson–Crick base pair, Kf appears to require the precise assembly of an arginine fork—formed between the primer terminus minor groove H-bond acceptor, residue Arg668 of Kf, and the incoming dNTP ribosyl oxygen.⁶ With the bicyclic analogues in the primer, the larger aromatic surface area may facilitate dNTP binding via favorable stacking interactions. However, this strong interaction may induce the formation of a slightly distorted primer terminus–Kf–dNTP complex, resulting in the slightly compromised k_{cat} observed. With **4MP** at the primer

terminus, its pyrimidine-like structure may result in a more natural-like primer terminus structure, including the correct alignment of the nucleophilic 3'-OH and minor groove H-bond acceptor. However, compared to the bicyclic analogues at the primer terminus, the relatively reduced surface area of **4MP** may underlie its weaker dCTP binding. Nonetheless, the base pairs are reasonably well extended in either context, while the self-pairs are not, suggesting that the corresponding nucleobase analogues are complementary and truly form a base pair.

Recent systematic efforts to understand the role of hydrophobicity, aromatic surface area, and electronic contributions have begun to shed light on the determinants of polymerase-mediated insertion of unnatural dNTPs.^{1b,7} However, the rate-limiting extension of the resulting base pairs has remained relatively less studied and, thus, less understood, limiting the rational design of viable unnatural base pairs. From a pool of thousands of novel nucleobase combinations, we identified one class of unnatural heteropairs that are extended by Kf with second-order rate constants that are orders of magnitude higher than those typically observed for unnatural base pairs. In fact, they represent the most efficiently extended unnatural base pairs identified to date. Thus, it is apparent that interbase interactions mediated by predominantly hydrophobic interactions are sufficient to support both base pair synthesis and extension. The fact that each of the four members of the bicyclic analogue family formed efficiently extended heteropairs when paired opposite **4MP** suggests that their recognition is general to this class of heteropair. Experiments designed at further understanding the efficient extension of this family of heteropairs should prove invaluable for future base pair design efforts and may also yield derivatives that are even better recognized by DNA polymerases. Moreover, further progress will be stimulated by more sophisticated screens which might sort through libraries of heretopairs, based not only on their efficient extension but also on their efficient and high fidelity synthesis. Such screens are currently under development.

Acknowledgment. This work was supported by the National Institutes of Health (NIGMS 60005).

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

References

- (1) (a) Moran, S.; Ren, R. X.-F.; Sumney, S., IV; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056–2057. (b) Henry, A. A.; Olsen, A. G.; Matsuda, S.; Yu, C.; Geierstanger, B. H.; Romesberg, F. E. *J. Am. Chem. Soc.* **2004**, *126*, 6923–6931. (c) Hirao, I.; Harada, Y.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Yokoyama, S. *J. Am. Chem. Soc.* **2004**, *126*, 13298–13305.
- (2) Ogawa, A. K.; Wu, Y.; McMinn, D. L.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 3274–3287.
- (3) (a) Adelfinskaya, O.; Nashine, V. C.; Bergstrom, D. E.; Davisson, V. J. *J. Am. Chem. Soc.* **2005**, *127*, 16000–16001. (b) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **1999**, *121*, 2323–2324.
- (4) Matsuda, S.; Henry, A. A.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2003**, *125*, 6134–6139.
- (5) Miller, H.; Grollman, A. P. *Biochemistry* **1997**, *36*, 15336–15342.
- (6) Meyer, A. S.; Blandino, M.; Spratt, T. E. *J. Biol. Chem.* **2004**, *279*, 33043–33046.
- (7) (a) Kincaid, K.; Beckman, J.; Zivkovic, A.; Halcomb, R. L.; Engels, J.; Kuchta, R. D. *Nucleic Acids Res.* **2005**, *33*, 2620–2628. (b) Zhang, X.; Lee, I.; Zhou, X.; Berdis, A. J. *J. Am. Chem. Soc.* **2006**, *128*, 143–149. (c) Kim, T. W.; Delaney, J. C.; Essigmann, J. M.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15803–15808.

JA060853C