Rational Design of an Unnatural Base Pair with Increased Kinetic Selectivity

Anthony K. Ogawa,† Yiqin Wu,† Markus Berger, Peter G. Schultz,* and Floyd E. Romesberg*

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

Received April 26, 2000

In an effort to expand the genetic code we have examined a variety of unnatural hydrophobic base pairs which pair selectively in duplex DNA and are also synthesized efficiently by a DNA polymerase.[1−6] In some cases, the observed rates for the synthesis of DNA containing unnatural base pairs approach those of natural base pairs.[2,4] However, the kinetic selectivity of the unnatural base pairs has generally been limited by the fast enzymatic misincorporation of dATP opposite the hydrophobic bases in the template. For example, as previously reported[2] **2MN**, **DMN**, and **TM** (Figure 1) each direct the exonuclease deficient Klenow fragment of *E. coli* DNA polymerase I (KF) to incorporate dATP with an efficiency (apparent k_{cat}/K_M) in excess of 10^5 M^{-1} min^{-1}. The "A-rule" proposes that dATP is preferentially inserted opposite noninstructive sites in the template.[7−9] This may imply that as a result of the unnatural bases having none of the expected H-bonding patterns, KF reads them in the template as "noninstuctive". An alternative to the "A-rule" for explaining the efficient insertion of dATP is based on specific interactions between the unnatural bases and the natural triphosphate. It is plausible that the hydrophobic C2-methine of adenine (unique among the natural bases) could interact favorably with the hydrophobic *o*-methyl substituents of **2MN**, **DMN**, and **TM** in the developing minor groove. If this model is correct, deletion of the *o*-methyl substituent could increase selectivity against dATP insertion and result in more selective synthesis of the unnatural base pair.

To examine this model, **3MN** and **2Np** (Figure 1) were synthesized and characterized. The **3MN** and **2Np** nucleosides were synthesized following a previously reported strategy (see Supporting Information) and converted to the corresponding triphosphates and phosphoramidites by literature methods.[10,11] The unnatural C-nucleosides were incorporated at position X of oligonucleotide 5′-dATTATGCTGAGTGATATCCCTCTXGT-CA to evaluate template-directed polymerase synthesis. We report below that **3MN** selectively enhances self-pair formation relative to misincorporation of dATP. This leads to an unnatural base pair that is enzymatically synthesized with an efficiency and fidelity approaching those characteristic of native base pair synthesis.

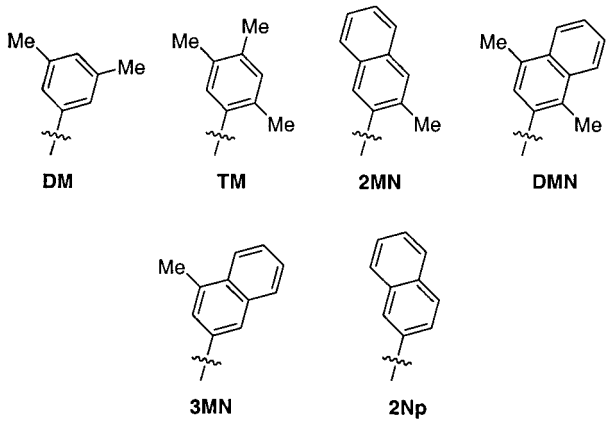

Figure 1. Unnatural hydrophobic bases.

Steady-state kinetic data corresponding to single nucleotide incorporation are shown in Table 1.[12] There is an 17-fold reduction in incorporation efficiency for dATP opposite **3MN** relative to its methylated analogue **DMN**, while the k_{cat}/K_M for incorporation of other native triphosphates opposite **3MN** was largely unaffected.[2] The **3MN**:**3MN** self-pair is efficiently synthesized ($k_{cat}/K_M = 2.4 \times 10^6$ M^{-1} min^{-1}) resulting in a 62-fold selectivity for correct self-pair synthesis. Although DNA synthesis beyond the **3MN:3MN** self-pair proceeded inefficiently, its highly selective formation represents an important development in our efforts to design a third base pair.

To deconvolute the impact of the *m*-methyl group, **2Np** was synthesized (Figure 1). Surprisingly, KF efficiently inserts dATP opposite **2Np** in the template. The differing behaviors of the regioisomeric *o*- and *m*-methyl substitution may be rationalized based on the specific interactions between the hydrophobic template and dATP. This model hypothesizes that ortho substituents favor the incorporation of dATP by providing hydrophobic contacts with the incoming adenine nucleobase, while meta substituents disfavor the KF mediated insertion of dATP due to forced desolvation of the adenine hydrophilic endocyclic amide moiety.[2,4,13] The kinetic data support this model if rotational isomerization about the C-glycosidic linkage is considered (Figure 2). **TM** is known to reside in the anti configuration about the C-glycosidic linkage,[14] and the rotation to the syn conformation is disfavored due to the resulting eclipsing interactions between the *o*-methyl group and the ribosyl oxygen lone pairs. The same is likely true for **2MN** and **DMN**. Therefore, **2MN**, **DMN**, and **TM** each present the incoming triphosphate nucleobase with an *o*-methyl group disposed toward the developing minor groove.

* To whom correspondence should be addressed: Telephone (858) 784-7290, Fax (858) 784 7472, E-mail floyd@scripps.edu.
† These authors contributed equally to this work.

(1) McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **1999**, *121*, 11585−11586.
(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) Berger, M.; Ogawa, A. K.; McMinn, D. L.; Wu, Y.; Schultz, P. G.; Romesberg, F. E. *Angew. Chem., Int. Ed.* **2000**, *39*, 2940−2942.
(4) Wu, Y.; Ogawa, A. K.; Berger, M.; McMinn, D. L.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 7621−7632.
(5) For alternative strategies for expanding the genetic code, see: (a) Lutz, S.; Burgstaller, P.; Benner, S. A. *Nucleic Acids Res.* **1999**, *27*, 2792−2798. (b) Lutz, M. J.; Horlacher, J.; Benner, S. A. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 499−504. (c) Lutz, M. J.; Horlacher, J.; Benner, S. A. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1149−1152. (d) Lutz, M. J.; Held, H. A.; Hottiger, M.; Hübscher, U.; Benner, S. A. *Nucleic Acids Res.* **1996**, *24*, 1308−1313. (e) Horlacher, J.; Hottiger, M.; Podust, V. N.; Hübscher, U.; Benner, S. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6329−6333. (f) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33−37. (g) Bain, J. D.; Switzer, C.; Chamberlin, A. R.; Benner, S. A. *Nature* **1992**, *356*, 537−539.
(6) For studies of hydrophobic nucleobases see: (a) Barsky, D.; Kool, E. T.; Colvin, M. E. *J. Biomol. Struct. Dyn.* **1999**, *16*, 1119−1134. (b) Matray, T. J.; Kool, E. T. *Nature* **1999**, *399*, 704−708. (c) Kool, E. T. *Biopolymers* 1998, *48*, 3−17. (d) Guckian, K. M.; Scheweitzer, B. A.; Sheils, C. J.; Tahmassebi, D. C.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 2213−2222. (e) Kool, E. T.; Morales, J. C.; Guckian, K. M. *Angew. Chem., Int. Ed.* **2000**, *39*, 991−1009. (f) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 1001−1007. (g) Morales, J. C.; Kool, E. T. *Nature Struct. Biol.* **1998**, *5*, 950−954. (h) Kool, E. T. *Chem. Rev.* **1997**, *97*, 1473−1487.
(7) Sagher, D.; Strauss, B. *Biochemistry* **1983**, *22*, 4518−4526.
(8) Randall, S. K.; Eritja, R.; Kalplan, B. E.; Petruska, J.; Goodman, M. F. *J. Biol. Chem.* **1987**, *262*, 6864−6870.
(9) Takeshita, M.; Chang, C.-N.; Johnson, F.; Will, S.; Grollman, A. P. *J. Biol. Chem.* **1987**, *262*, 10171−10179.
(10) Kovacs, T.; Otvos, L. *Tetrahedron Lett.* **1988**, *29*.
(11) Schweitzer, B. A.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 1863−1872.
(12) Goodman, M. F.; Creighton, S.; Bloom, L. B.; Petruska, J. *Crit. Rev. Biochem. Mol. Biol.* **1993**, *28*, 83−126.
(13) Shih, P.; Pedersen, L. G.; Gibbs, P. R.; Wolfenden, R. *J. Mol. Biol.* **1998**, *280*, 421−430.
(14) Jarso, S.; Wemmer, D. Unpublished results.

10.1021/ja001450u CCC: $19.00 © 2000 American Chemical Society
Published on Web 08/26/2000

Table 1. Steady-State Kinetic Constants for KF Mediated Synthesis of DNA with **3MN** and **2Np** in the Template[a]

5′-dTAATACGACTCACTATAGGGAGA
3′-dATTATGCTGAGTGATATCCCTCTX̲GTCA

template (*X*)	nucleoside triphosphate	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} min^{-1})
3MN	3MN	32 ± 2	14 ± 3	2.4×10^6
	A	0.37 ± 0.02	49 ± 7	7.6×10^3
	G	n.d.[b]	n.d.[b]	$\leq 1.0 \times 10^3$
	C	0.86 ± 0.05	108 ± 15	8.0×10^3
	T	2.4 ± 0.1	61 ± 4	3.9×10^4
2Np	2Np	22 ± 1	8 ± 1	2.8×10^6
	A	2.5 ± 0.2	22 ± 7	1.1×10^5
	G	n.d.[b]	n.d.[b]	$\leq 1.0 \times 10^3$
	C	0.72 ± 0.03	85 ± 12	8.5×10^3
	T	1.0 ± 0.1	86 ± 16	1.2×10^4
dT	A	260 ± 50	4.1 ± 1.0	6.3×10^7

[a] Assay conditions include the following: 40 nM template−primer duplex, 0.11−1.34 nM enzyme, 50 mM Tris buffer (pH 7.5), 10 mM $MgCl_2$, 1 mM DTT, 50 μg/mL BSA. The reactions were initiated by adding the DNA−enzyme mixture to an equal volume (5 μL) of a 2× dNTP stock solution, incubated at room temperature for 1−10 min, and quenched by the addition of 20 μL of loading buffer (95% formamide, 20 mM EDTA). 5 μL of the reaction mixture were then analyzed by 15% polyacrylamide gel containing 8 M urea. Radioactivity was quantified using a PhosphorImager (Molecular Dynamics), with overnight exposures and the ImageQuant program. The data were fitted to the simple Michaelis−Menten equation. Data presented here are averages of triplicates. [b] Rates too slow for determination of k_{cat} and K_M independently.

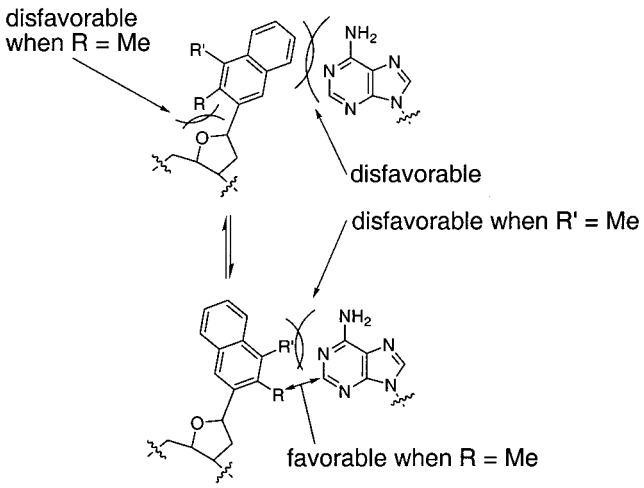

Figure 2. Rotational isomerism of naphthalene bases.

However, **DM** (Figure 1), **3MN**, and **2Np**, which lack ortho substituents, are presumably free to adopt either configuration. Despite rotation, **3MN** and **DM** each present meta substituents (methyl group or phenyl ring) to the incoming nucleobase. Rotation about the C-glycosidic bond of **2Np** presents a surface similar to an unsubstituted phenyl ring.

Table 2

template	$(k_{cat}/K_m)_{rel}$ for dATP insertion
2MN	46[a]
DMN	17[a]
2Np	14
3MN	1

[a] See ref 2.

This model allows for a consistent interpretation of the relative rates of dATP insertion opposite each unnatural base in the template (Table 2). **2MN** and **TM** direct KF to insert dATP with the greatest efficiency because each has a favorable ortho but not a disfavorable meta substituent. **DMN** and **2Np** direct the insertion of dATP with intermediate efficiencies. With **DMN** there is a cancellation of effects (favorable ortho substituent and a disfavorable meta substituent), while **2Np** possesses neither the favorable, nor the disfavorable substituent. In contrast, **3MN** and **DM** possess only disfavorable meta substituents, and thus inefficiently template dATP incorporation.

The ability to control the dATP insertion efficiency opposite hydrophobic bases indicates that these unnatural bases are not rendered "noninstructive" by the absence of H-bonds, and that the A-rule does not limit their use for information storage and replication. In this regard, these fully carbocyclic naphthyl derivatives seem more attractive as third base pair candidates than the previously reported isocarbostyril derivatives, which have similar structures, but have N-glycosidic linkages and minor groove carbonyl groups.[1−4] The incorporation of both native and unnatural triphosphates was only slightly sensitive to substitution of isocarbostyril derivatives in the template.

The **3MN**:**3MN** self-pair is synthesized by KF with an efficiency of 2.4×10^6 M^{-1}min^{-1}. This is only 20-fold reduced relative to that for the synthesis of a dA:dT base pair in the same sequence context, and may represent sufficient efficiency for biological function. The self-pair is also 62-fold selective against mispairing with native triphosphates, each of which is inserted opposite **3MN** in the template with a rate typical of that for native mispair synthesis (10^4 M^{-1} min^{-1}). The insertion of d**3MN**TP opposite a native template is also not competitive with correct triphosphate insertion. In terms of efficiency and selectivity, the **3MN**:**3MN** self-pair is the most successful third-base pair candidate yet reported. A shortcomming of the self-pair, like others that have been reported, is the inefficiency with which KF continues DNA synthesis after incorporation of the unnatural base triphosphate. We are currently addressing this issue with other design modifications, including changes to both the base and the 2′-deoxyribosyl subunits. We are also employing directed evolution techniques in an attempt to evolve variants of a DNA polymerase that can efficiently and selectively synthesize the **3MN**:**3MN** self-pair, and then efficiently continue DNA synthesis.

Supporting Information Available: Experimental procedures and characterizations (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA001450U