DOI: 10.1002/cbic.200800577

Optimization of the Pyridyl Nucleobase Scaffold for Polymerase Recognition and Unnatural Base Pair Replication

Yoshiyuki Hari,[a] Gil Tae Hwang,[a] Aaron M. Leconte,[a] Nicolas Joubert,[b] Michal Hocek,[b] and Floyd E. Romesberg*[a]

As part of an effort to increase both the biological and biotechnological applications of DNA, we[1–5] and others[6–9] have explored the DNA polymerase-mediated replication of a wide range of unnatural base pairs. In our initial efforts we examined large, aromatic, unnatural nucleotides, both as self pairs of two identical nucleotides and heteropairs of different nucleotides.[1–5,10,11] While several of these unnatural base pairs are efficiently synthesized (through insertion of the unnatural dNTP opposite its partner in the template) by the exonuclease-deficient Klenow fragment of *Escherichia coli* DNA polymerase I (Kf), none are efficiently extended (by continued primer elongation); this is most likely due to interstrand nucleobase intercalation and distortion of the primer terminus.[10] Thus, a range of nucleotides bearing smaller phenyl-based nucleobases incapable of intercalation were explored, and several modifications that facilitate extension were identified.[1–4] Of these, aza-substitution at the 2-position (**2Py**, Figure 1) appears to be the only modification that facilitates self pair extension without significantly facilitating mispairing.[3]

We have also recently found that another of the phenyl-based nucleotides, d**MMO2**, forms a heteropair with d**5SICS** that is synthesized and extended with relatively high efficiency and fidelity by a variety of different DNA polymerases (Figure 1).[5] However, it is unclear whether d**MMO2** is the best phenyl-based nucleotide for heteropair formation with d**5SICS**, or if a derivatized pyridyl analogue might optimize heteropair replication.

Here, we report the synthesis and characterization of a series of substituted 2-pyridine derivatives, (Figure 1) which were designed to systematically examine the effect of nucleobase shape, size, and hydrophobicity, as well as structural and electronic modifications within the interbase interface. The analogues are examined both as self pairs and as part of a heteropair with d**5SICS**. All nucleosides, phosphoramidites, oligonucleotides, and triphosphates were synthesized as described in the Supporting Information or in ref. [12].

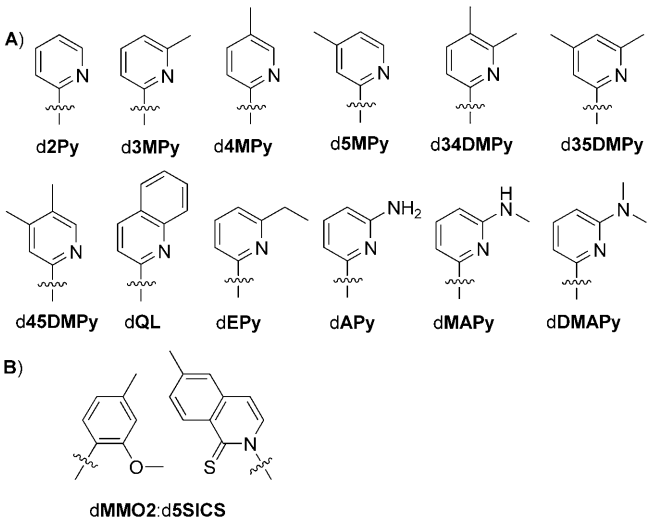

Figure 1. A) 2-Pyridyl nucleotides synthesized and characterized in this study. B) d**MMO2**:d**5SICS** base pair. Only the nucleobase analogue is shown. The wavy line indicates connection to the sugar and phosphate backbone, which have been omitted for clarity.

The **2Py** self pair is largely limited by inefficient self pair synthesis; thus, we first characterized the steady-state rates for Kf-mediated synthesis of the 2-pyridyl-based nucleotide self pairs (Table 1). As with the fully carbocyclic scaffold,[4] methyl group substitution has a significant effect on self pair synthesis. For example, the d**4MPy** and d**45DMPy** self pairs are synthesized far more efficiently than the parent d**2Py** self pair (which is synthesized with an efficiency of $6.2\times10^3\,\text{M}^{-1}\,\text{min}^{-1}$);[3] this demonstrates that simple methyl substitution can substantially increase the rates of synthesis. The similar rates with which the d**4MPy** and d**45DMPy** self pairs are synthesized (1.6×10^4 and $4.1\times10^4\,\text{M}^{-1}\,\text{min}^{-1}$, respectively) suggest that substitution at the 4-position is sufficient for the observed increase in efficiency. Additionally, self pairs of d**QL**, which combines both 3- and 4-position substitution with increased aromatic surface area, are synthesized with an efficiency very similar to those of d**4MPy** and d**45DMPy** (Table 1), suggesting that the packing interactions mediated by the 4-position methyl group stabilize the dNTP insertion transition state as much as the intercalative interactions mediated by the larger aromatic group. The rates of synthesis for the self pairs of the remaining analogues are all less than that for d**2Py**, indicating that substitution at the 3- or 5-positions does not facilitate self pair synthesis.

We also characterized the efficiency with which Kf inserts the natural dNTPs opposite a pyridyl nucleotide in the template to gauge the fidelity of unnatural base pair synthesis (Table S1). For reference, opposite d**2Py**, dATP is the most effi-

[a] Dr. Y. Hari,[+] Dr. G. T. Hwang,[+] A. M. Leconte, Prof. Dr. F. E. Romesberg
Department of Chemistry, The Scripps Research Institute
La Jolla, CA 92037 (USA)
Fax: (+1) 858-784-7472
E-mail: floyd@scripps.edu

[b] N. Joubert, Prof. Dr. M. Hocek
Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic
Gilead Sciences and IOCB Research Center
16610, Prague 6 (Czech Republic)

[+] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under http://www.chembiochem.org or from the author.

Table 1. Incorporation rates of unnatural triphosphates d**X**TP.[a]

5′-dTAATACGACTCACTATAGGGAGA
3′-dATTATGCTGAGTGATATCCCTCT**X**GCTAGGTTACGGCAGGATCGC

X	k_{cat} [min^{-1}]	K_M [μM]	k_{cat}/K_M [M^{-1} min^{-1}]
d**3MPy**	0.42 ± 0.14	163 ± 23	2.6×10^3
d**4MPy**	3.2 ± 0.6	201 ± 27	1.6×10^4
d**5MPy**	0.34 ± 0.02	117 ± 33	2.9×10^3
d**34DMPy**	1.2 ± 0.2	267 ± 22	4.7×10^3
d**35DMPy**	0.10 ± 0.02	89 ± 19	1.2×10^3
d**45DMPy**	2.0 ± 0.2	49 ± 7	4.1×10^4
d**QL**	0.74 ± 0.05	20 ± 3	3.7×10^4
d**EPy**	nd[b]	nd[b]	$<1.0\times10^3$
d**APy**	nd[b]	nd[b]	$<1.0\times10^3$
d**MAPy**	nd[b]	nd[b]	$<1.0\times10^3$
d**DMAPy**	0.26 ± 0.02	162 ± 42	1.6×10^3

[a] For details see the Supporting Information. [b] Reaction was too inefficient for k_{cat} and K_M to be determined independently.

ciently inserted natural dNTP, and it is actually inserted 32-fold faster than d**2Py**TP.[3] While substitution at the 5-position has no significant effect, the rate of incorporation of dATP decreases as the steric bulk at the 3- or 4-position increases. In each case, dATP remains the most efficiently inserted natural dNTP, followed by dGTP, dTTP, and last by dCTP. The increased rate of self pair synthesis and the reduced rates of mispairing with dA combine so that the d**4MPy** and d**45DMPy** self pairs are synthesized only five- and two-fold (respectively) slower than dATP is inserted.

We next examined the rate at which each derivatized self pair is extended by insertion of dCTP opposite a dG in the template (Table 2). Substitution at the 3-position has widely varying effects. While methyl substitution has no effect, ethyl substitution (d**EPy**) increases the rate of extension by two-fold relative to the d**2Py** self pair, and the amino substituents (d**APy**, d**MAPy**, and d**DMAPy**) decrease efficiency of extension to an extent that is correlated with substituent size. Thus, it appears that while the interface between these pyridyl nucleobase analogues may be optimized by increased packing, it is not tolerant of altered electrostatics. The effects at the 4- and 5-position are much more promising, and increase display a six- to twelve-fold increase in the rate of self pair extension with methyl substitution. The effects are roughly additive; extension of the d**45DMPy** self pair is increased 70-fold (relative to the unmodified self pair) to a rate that is only 70-fold slower than that for a natural base pair. In fact, the d**45DMPy** self pair is the most efficiently extended self pair that has been identified to date. These data suggest that increased steric bulk in the nucleobase interface induces a structure that is less efficiently extended, perhaps by causing a widening or distortion of the base pair, while modification at the 4- and 5-positions apparently favors extension, perhaps due to better interbase packing within the major groove or optimized packing with flanking nucleobases. Considering all steps, methyl substitution at the 4-position is the most beneficial as it increases both the efficiency and fidelity of both unnatural base pair synthesis and extension. Substitution at the 5-position, which does not significantly affect synthesis, but does favor extension, also facilitates replication.

We next examined the potential of the pyridyl nucleotides as d**MMO2** analogues by screening them for their ability to pair with d**5SICS**. Examination of primer extension by gel electrophoresis revealed that d**5MPy** and d**34DMPy** were most efficiently paired with d**5SICS** (Figure S1). Thus, we characterized the d**5MPy**:d**5SICS** and d**34DMPy**:d**5SICS** heteropairs in greater detail (Table 3). The triphosphates of d**5MPy** and d**34DMPy** are inserted opposite d**5SICS** with second order rate constants that are approximately ten times slower than the rate constant for insertion of d**MMO2**TP. Likewise, d**5SICS**TP is inserted opposite either 2-pyridyl analogue in the template with rates that are approximately ten times less efficient than insertion opposite d**MMO2**. Thus, at least with these analogues, methyl substitution at the 3-, 4-, and 5-positions appears to have similar effects on synthesis, and the aza nitrogen atom appears to be slightly less beneficial than the methoxy substituent at the 2-position.

Interestingly, the extension of the two pyridyl heteropairs is very different. While the d**34DMPy**:d**5SICS** (primer:template) and d**5SICS**:d**34DMPy** heteropairs are extended ten- to 100-fold slower than the

Table 2. Extension rates of unnatural self pairs.[a]

5′-dTAATACGACTCACTATAGGGAGA**X**
3′-dATTATGCTGAGTGATATCCCTCT**X**GCTAGGTTACGGCAGGATCGC

X	k_{cat} [min^{-1}]	K_M [μM]	k_{cat}/K_M [M^{-1} min^{-1}]
d**3MPy**	2.2 ± 0.1	100 ± 3	2.2×10^4
d**4MPy**	7.5 ± 1.6	36 ± 6	2.1×10^5
d**5MPy**	6.5 ± 1.1	16 ± 2	4.1×10^5
d**34DMPy**	1.3 ± 0.08	132 ± 23	9.9×10^3
d**35DMPy**	6.8 ± 0.9	68 ± 15	1.0×10^5
d**45DMPy**	15 ± 2	6.2 ± 0.3	2.4×10^6
d**QL**	0.073 ± 0.007	57 ± 5	1.3×10^3
d**EPy**	8.6 ± 1.6	102 ± 36	8.4×10^4
d**APy**	1.0 ± 0.1	41 ± 2	2.5×10^4
d**MAPy**	0.77 ± 0.04	123 ± 17	6.3×10^3
d**DMAPy**	nd[b]	nd[b]	$<1.0\times10^3$

[a] For details see the Supporting Information. [b] Reaction was too inefficient for k_{cat} and K_M to be determined independently.

Table 3. Steady-state rate constants of d**5MPy**:d**5SICS** and d**34DMPy**:d**5SICS** heteropairs.[a]

5′-dTAATACGACTCACTATAGGGAGA
3′-dATTATGCTGAGTGATATCCCTCT**Y**GCTAGGTTACGGCAGGATCGC

d**X**TP	**Y**	k_{cat} [min^{-1}]	K_M [μM]	k_{cat}/K_M [M^{-1} min^{-1}]
d**5MPy**	d**5SICS**	4.1 ± 0.1	152 ± 12	2.7×10^4
d**34DMPy**	d**5SICS**	5.3 ± 0.1	132 ± 4	4.0×10^4
d**5SICS**	d**5MPy**	4.5 ± 0.5	0.69 ± 0.03	6.6×10^6
d**5SICS**	d**34DMPy**	4.1 ± 0.7	2.5 ± 0.2	1.6×10^6

5′-dTAATACGACTCACTATAGGGAGA**X**
3′-dATTATGCTGAGTGATATCCCTCT**Y**GCTAGGTTACGGCAGGATCGC

X	**Y**	k_{cat} [min^{-1}]	K_M [μM]	k_{cat}/K_M [M^{-1} min^{-1}]
d**5MPy**	d**5SICS**	1.5 ± 0.1	0.056 ± 0.015	2.7×10^7
d**34DMPy**	d**5SICS**	2.1 ± 0.1	7.9 ± 0.5	2.6×10^5
d**5SICS**	d**5MPy**	3.2 ± 0.4	13 ± 1	2.4×10^5
d**5SICS**	d**34DMPy**	0.47 ± 0.06	106 ± 3	4.4×10^3

[a] For details see the Supporting Information.

d**MMO2** heteropair, the d**5MPy**:d**5SICS** heteropair is extended with an efficiency of 2.7×10^7 M^{-1} min^{-1}, which is actually more efficient than the corresponding d**MMO2** heteropair, and remarkably, only six-fold slower than extension of a natural base pair in the same sequence context. Moreover, the d**5SICS**:**d5MPy** heteropair is extended with an efficiency that is only marginally reduced relative to the heteropair with d**MMO2**. Thus, derivatization, particularly methyl substitution at the 5-position, has a significant and beneficial effect on the pairing of the 2-pyridyl analogues with d**5SICS** and, at least for extension, can actually optimize the heteropair so that it is better recognized than d**MMO2**:d**5SICS**.

While these data will be helpful for the design of optimized base pairs, it is apparent that the d**45DMPy** self pair is the most promising of the unnatural base pairs examined in the current study. To further explore the utility of this self pair, we determined the rates at which all possible mispairs are extended (Table S2), which along with the rates at which the mispairs are synthesized (see above), allow for a determination of the overall fidelity. The most efficiently synthesized mispair, dA:d**45DMPy**, is extended approximately 20-fold less efficiently than the correct pair. Thus, the overall fidelity for self pair replication (synthesis and extension) relative to the mispair with dA is eleven. The most efficiently extended mispair is that with dT, which is extended with a rate of 6.4×10^5 M^{-1} min^{-1}; however, due to the mispair's inefficient synthesis, the overall fidelity of the self pair relative to the mispair with dT is 33. The mispairs with dC and dG are extended with rates of 3.9×10^3 and 2.1×10^3 M^{-1} min^{-1}, respectively, resulting in an overall fidelity of 1.1×10^4 for dC and 6.0×10^3 for dG.

To explore the potential elimination of mispairs that are efficiently synthesized but not extended, we characterized full length DNA synthesis with Kf or exonuclease-proficient Kf (Kf exo$^+$) (Figure 2). As expected, full-length synthesis was observed with a natural template (X = dT) with both Kf and Kf exo$^+$. With d**45DMPy** in the template, Kf exo$^+$, and only natural triphosphates present, an equilibrium was observed between the primer and the n+1 extension product, presumably resulting from dATP insertion and excision. However, when d**45DMPy**TP was added to the reaction, full length product was observed in good yield. Thus, while further optimization is clearly desired, the d**45DMPy** self pair might have immediate use in a variety of different in vitro applications.[13]

Cumulatively, these results demonstrate that every step of replication may be optimized by derivatization of the 2-pyridyl nucleobase analogues. Judicious placement of methyl groups alone, yielding d**45DMPy**, results in a self pair that is extended with a rate similar to the natural one; this self pair can be used to synthesize site-specifically modified DNA in good yield. Moreover, while none of the pyridyl analogues is actually a better heteropair partner for d**5SICS** than d**MMO2** itself, the results demonstrate that at least in some contexts, heteropair extension is better facilitated by a 2-pyridyl nitrogen atom than a methoxy group. Thus, the 2-pyridyl scaffold remains among the most promising nucleobase analogues and further optimization should result in self pairs or heteropairs for in vitro, and possibly even in vivo, applications.[13,14]

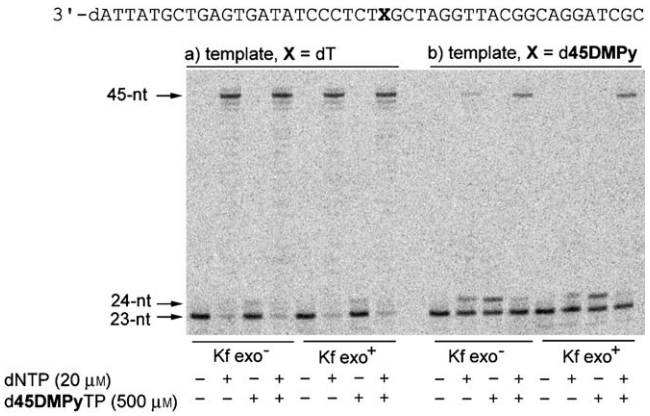

Figure 2. Extension of 23-nt primer with 45-nt templates (**X** = dT or d**45DMPy**). Reactions contain either Kf exo$^+$ or Kf exo$^-$ and the reaction time was 3 min.

Acknowledgements

We thank the National Institutes of Health (GM60005 to F.E.R.) for financial support. Support was also provided by the Uehara Memorial Foundation (Y.H.), the Czech Academy of Sciences (Z04055905 to M.H.) and MSMT (LC512 to N.J. and M.H.).

Keywords: DNA replication · DNA · genetic alphabet · pyridine · unnatural nucleobases

[1] A. A. Henry, A. G. Olsen, S. Matsuda, C. Yu, B. H. Geierstanger, F. E. Romesberg, *J. Am. Chem. Soc.* **2004**, *126*, 6923–6931.
[2] A. M. Leconte, S. Matsuda, G. T. Hwang, F. E. Romesberg, *Angew. Chem.* **2006**, *118*, 4432–4435; *Angew. Chem. Int. Ed.* **2006**, *45*, 4326–4329.
[3] Y. Kim, A. M. Leconte, Y. Hari, F. E. Romesberg, *Angew. Chem.* **2006**, *118*, 7973–7976; *Angew. Chem. Int. Ed.* **2006**, *45*, 7809–7812.
[4] S. Matsuda, A. A. Henry, F. E. Romesberg, *J. Am. Chem. Soc.* **2006**, *128*, 6369–6375.
[5] A. M. Leconte, G. T. Hwang, S. Matsuda, P. Capek, Y. Hari, F. E. Romesberg, *J. Am. Chem. Soc.* **2008**, *130*, 2336–2343.
[6] Z. Yang, A. M. Sismour, P. Sheng, N. L. Puskar, S. A. Benner, *Nucleic Acids Res.* **2007**, *35*, 4238–4249.
[7] A. T. Krueger, H. Lu, A. H. F. Lee E. T. Kool, *Acc. Chem. Res.* **2007**, *40*, 141–150.
[8] I. Hirao, T. Mitsui, M. Kimoto, S. Yokoyama, *J. Am. Chem. Soc.* **2007**, *129*, 15549–15555.
[9] a) M. Chiaramonte, C. L. Moore, K. Kincaid, R. D. Kuchta, *Biochemistry* **2003**, *42*, 10472–10481; b) K. Kincaid, J. Beckman, A. Zivkovic, R. L. Halcomb, J. W. Engels, R. D. Kuchta, *Nucleic Acids Res.* **2005**, *33*, 2620–2628.
[10] S. Matsuda, J. D. Fillo, A. A. Henry, P. Rai, S. J. Wilkens, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer, P. G. Schultz, G. Spraggon, F. E. Romesberg, *J. Am. Chem. Soc.* **2007**, *129*, 10466–10473.
[11] A. Ogawa, Y. Wu, D. L. McMinn, J. Liu, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **2000**, *122*, 3274–3287.
[12] M. Urban, R. Pohl, B. Klepetarova, M. Hocek, *J. Org. Chem.* **2006**, *71*, 7322–7328.
[13] a) M. J. Ieven, *J. Clin. Virol.* **2007**, *40*, 259–276; b) T. A. Cropp, J. W. Chin, *Curr. Opin. Chem. Biol.* **2006**, *10*, 601–606; c) J. A. Bittker, K. J. Phillips, D. R. Liu, *Curr. Opin. Chem. Biol.* **2002**, *6*, 367–374.
[14] a) L. Wang, J. Xie, P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249; b) J. Xie, P. G. Schultz, *Curr. Opin. Chem. Biol.* **2005**, *9*, 548–554.

Received: August 26, 2008

