Nonnatural Base Pairs

DOI: 10.1002/ange.200602579

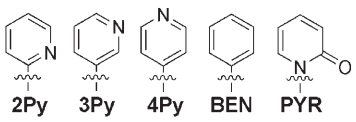
Stability and Polymerase Recognition of Pyridine Nucleobase Analogues: Role of Minor-Groove H-Bond Acceptors**

*Yoonkyung Kim, Aaron M. Leconte, Yoshiyuki Hari, and Floyd E. Romesberg**

The information potential of a genome is limited by the four-letter, two-base-pair genetic alphabet. Significant efforts have been directed towards developing a third, unnatural base pair, which would expand the genetic alphabet and lay the foundations of an expanded genetic code. Studies from the Benner and Kool groups have elegantly demonstrated that the interbase hydrogen bonds (H-bonds) may be rearranged,[1] or removed altogether,[2] without complete loss of stability or polymerase recognition; this information has inspired design strategies based on the use of predominantly hydrophobic nucleobase analogues.[3]

While base-pair stability and efficient replication are required for any unnatural base-pair candidate, predominantly hydrophobic analogues are consistently limited by extension, that is, efficient primer elongation after synthesis of the unnatural base pair at a primer terminus.[3a,c] One possible explanation for inefficient extension is that, while properly positioned H-bonds are not required to mediate interbase recognition, they are required at the primer terminus to appropriately engage the polymerase for efficient DNA synthesis.[4,5] We have previously optimized nucleobase analogues by the inclusion of natural-nucleobase-like carbonyl groups in a position β to an *N*-glycosidic linkage.[6] However, this natural-like H-bond acceptor also reintroduces interbase H-bonds and facilitates mispairing. By contrast, H-bond acceptors at the α position of C-nucleosides should be less recognized by the natural nucleobases but may still be able to engage the polymerase appropriately for DNA synthesis.

Here, we report the synthesis and analysis of three pyridyl nucleotide analogues that are designed to explore the effect of positioning a nitrogen atom in the minor groove (**2Py**), the interbase interface (**3Py**), or the major groove (**4Py**) (Scheme 1). We focus on the evaluation of the corresponding self pairs (that is, pairs formed between two identical analogues). Each nucleoside was synthesized, converted into the corresponding phosphoramidite, and incorporated into DNA as described in the Supporting Information.



Scheme 1. Unnatural nucleobases. **BEN**[3c,7b] and **PYR**[6] have been previously reported and are included for comparison (see the text).

To explore the effect of aza substitution on thermal stability and selectivity, the melting temperatures (T_m) of duplexes containing the unnatural base pairs or mispairs were determined (Table 1). The **2Py** self pair was the most stable ($T_m = 52.2\,°C$), followed by the **3Py** and **4Py** self pairs ($T_m =$ 50.2 and 48.1 °C, respectively). Thus, the introduction of a single nitrogen substituent can alter self-pair stability by more than 4 °C. This is in contrast to the much smaller variations associated with fluoro, bromo, cyano, or methyl derivatization, which were each less than 2 °C.[3b,7] For all but **2Py**, the self pairs are significantly less stable than the parental, unsubstituted **BEN** self pair ($T_m = 52.8\,°C$; Scheme 1). We also examined the stability of heteropairs formed between the pyridyl nucleotides; heteropairs containing **2Py** were again

[*] Y. Kim, A. M. Leconte, Y. Hari, Prof. Dr. F. E. Romesberg
Department of Chemistry
The Scripps Research Institute
10550 North Torrey Pines Road, La Jolla, CA 92037 (USA)
Fax: (+1) 858-784-7472
E-mail: floyd@scripps.edu

[**] We acknowledge the National Institutes of Health (GM60005 to F.E.R.) and the Uehara Memorial Foundation (Y.H.) for financial support.

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

Table 1: Denaturation temperatures for duplex DNA containing **2Py**, **3Py**, and **4Py**.[a]

5′-dGCGTAC**X**CATGCG
3′-dCGCATG**Y**GTACGC

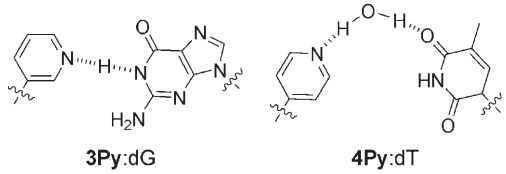
X	Y	T_m [°C]	X	Y	T_m [°C]	X	Y	T_m [°C]
2Py	**2Py**	52.2	**3Py**	**2Py**	51.1	**4Py**	**2Py**	50.1
2Py	**3Py**	51.1	**3Py**	**3Py**	50.2	**4Py**	**3Py**	49.6
2Py	**4Py**	49.6	**3Py**	**4Py**	49.1	**4Py**	**4Py**	48.1
2Py	A	50.2	**3Py**	A	50.2	**4Py**	A	49.1
2Py	C	44.2	**3Py**	C	44.7	**4Py**	C	45.1
2Py	G	49.2	**3Py**	G	52.7	**4Py**	G	49.6
2Py	T	47.7	**3Py**	T	49.2	**4Py**	T	50.1

[a] Determined in 10 mM 1,4-piperazinediethanesulfonic acid (PIPES), 10 mM $MgCl_2$, and 100 mM NaCl (pH 7). The error in the T_m values is ±0.1 °C

the most stable, although they were not as stable as the **2Py** self pair (Table 1).

The stability of the pyridyl base pairs probably results from a compromise between stabilizing hydrophobic forces, due to transfer of the hydrophobic analogues into the less polar duplex interior, and destabilizing desolvation effects. Forced desolvation will destabilize base-pair formation as the pyridyl analogues are expected to be more favorably solvated in single-stranded DNA.[8] Forced desolvation also explains the relative stability of the pairs involving **2Py**, as only its nitrogen atom will be oriented into the duplex minor groove where stabilizing H-bonds with water molecules from the spine of solvation[9] can offset the cost of desolvation.

Interestingly, as described above, **2Py** formed more stable unnatural base pairs than **3Py** and **4Py**; however, it did not form more stable mispairs (Table 1). Thus, the **2Py** self pair is thermally selective by 2.0–8.0 °C. This compares favorably to the thermal selectivity observed among natural base pairs.[10] The thermal selectivity of the **3Py** and **4Py** self pairs is limited by stable mispairing with dG or dT, respectively. In each case, the stability of the mispair may result from H-bonding (Scheme 2).[9b,11] While only speculative, these arguments



Scheme 2. H-bonding interactions that may underlie the relative stability of the **3Py**:dG and **4Py**:dT mispairs.

are consistent with the idea that desolvation is important for both mispair and unnatural base-pair stability, as H-bonds within the duplex environment would again compensate for the H-bonds with water that are lost upon duplex formation. Regardless of the interpretation, it appears that the minor-groove H-bond acceptor at the α position of **2Py** is not appropriately positioned to interact with the H-bond donors of the natural nucleobases.

The ability of the exonuclease-deficient Klenow fragment of *Escherichia coli* DNA polymerase I (Kf) to extend each self pair was examined by using single-nucleotide-incorporation steady-state kinetics (Table 2). As reported previously, extension of the fully carbocyclic **BEN** self pair, by incorporation of deoxycytidine triphosphate (dCTP) opposite deoxyguanosine (dG) in the template, occurs with a second-order rate constant of $1.6\times10^3\,\text{M}^{-1}\,\text{min}^{-1}$.[3c] The **3Py** and **4Py** self pairs are extended with rates that are too low to detect ($k_{cat}/K_M < 10^3\,\text{M}^{-1}\,\text{min}^{-1}$). However, the **2Py** self pair is extended by Kf with a k_{cat}/K_M value of $3.4\times10^4\,\text{M}^{-1}\,\text{min}^{-1}$, at least 34-fold more efficiently than the other pyridyl analogues. Thus, as with base-pair stability, aza substitution has a large effect

Table 2: Correct extension of unnatural self pairs.[a]

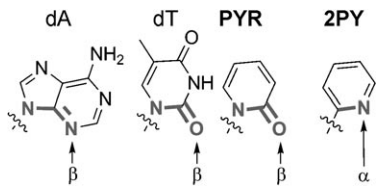
5′-dTAATACGACTCACTATAGGGAGA**X**
3′-dATTATGCTGAGTGATATCCCTCT**X**GCTAGGTTACGGCAGGATCGC
↓dCTP
5′-dTAATACGACTCACTATAGGGAGA**X**C
3′-dATTATGCTGAGTGATATCCCTCT**X**GCTAGGTTACGGCAGGATCGC

Unnatural self pair	k_{cat}[b] [min^{-1}]	K_M[b] [μM]	k_{cat}/K_M[b] [M^{-1} min^{-1}]
2Py	2.1 ± 0.8	61 ± 17	3.4×10^4
3Py	n.d.[c]	n.d.[c]	$<1.0\times10^3$
4Py	n.d.[c]	n.d.[c]	$<1.0\times10^3$

[a] See the Supporting Information for experimental details. [b] k_{cat} = rate of catalysis, K_M = Michaelis constant. [c] n.d. = not determined. Rates were too slow for the determination of k_{cat} and K_M independently.

on self-pair extension, with substitution at the *ortho* position substantially increasing the rate of extension and substitution at the *meta* or *para* positions decreasing the rate of extension. The fidelity of **2Py** self-pair extension was also high; the rates of self-pair extension by deoxyguanosine triphosphate (dGTP), deoxyadenosine triphosphate (dATP), or deoxythymidine triphosphate (dTTP) insertion opposite dG were all too low to detect ($k_{cat}/K_M < 10^3\,\text{M}^{-1}\,\text{min}^{-1}$).

The efficiency and fidelity of **2Py** self-pair extension suggests that the *ortho*-pyridyl nitrogen atom is positioned in the developing minor groove and productively engages the polymerase. Interestingly, extension of the **2Py** self pair is only about twofold less efficient than the pyridone scaffold (**PYR**, Scheme 1; extended at $8.2\times10^4\,\text{M}^{-1}\,\text{min}^{-1}$).[6] This small difference is somewhat surprising when it is considered that the minor-groove H-bond acceptor of **2Py** is located at the α position relative to the glycosidic bond, whereas the H-bond acceptor of **PYR** is located at a position β to the glycosidic bond, which is more analogous to the natural nucleobases (Scheme 3).



Scheme 3. Minor-groove H-bond acceptors within natural and unnatural scaffolds.

Based on the efficiency and fidelity of **2Py** self-pair extension, we synthesized the corresponding nucleoside triphosphate (see the Supporting Information) and examined the efficiency of Kf-mediated self-pair synthesis (Table 3).

Table 3: Triphosphate insertion opposite pyridyl analogues in the template.[a]

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

X	Y	k_{cat} [min^{-1}]	K_M [μM]	k_{cat}/K_M [$M^{-1}min^{-1}$]
2Py	**2Py**	0.92 ± 0.16	149 ± 12	6.2×10^3
	A	3.6 ± 1.2	18 ± 2	2.0×10^5
	C	0.16 ± 0.02	98 ± 30	3.2×10^3
	G	0.8 ± 0.3	41 ± 9	2.0×10^4
	T	0.16 ± 0.06	39 ± 9	4.1×10^3
3Py	A	6.8 ± 0.9	25 ± 0.5	2.7×10^5
	C	n.d.[b]	n.d.[b]	$<1.0\times10^3$
	G	11 ± 3	58 ± 13	1.9×10^5
	T	0.54 ± 0.05	83 ± 15	6.5×10^3
4Py	A	11 ± 3	20 ± 6	5.5×10^5
	C	n.d.[b]	n.d.[b]	$<1.0\times10^3$
	G	n.d.[b]	n.d.[b]	$<1.0\times10^3$
	T	1.6 ± 0.3	49 ± 16	3.3×10^4

[a] See the Supporting Information for experimental details. [b] Rates were too slow for the determination of k_{cat} and K_M independently.

The self pair is synthesized with a second-order rate constant of $6.2\times10^3\,M^{-1}min^{-1}$. While this synthesis is less efficient than that of self pairs comprised of nucleobase analogues with a large, extended aromatic surface area[3a] or more highly substituted phenyl rings,[3b,c] it is actually high compared to other minimally substituted scaffolds. For example, the **BEN** and **PYR** self pairs are both synthesized at rates less than $10^3\,M^{-1}min^{-1}$.[3c,6]

We also examined the rates at which each natural triphosphate is inserted opposite each pyridyl analogue (Table 3). As with most small hydrophobic analogues, dATP was the most efficiently incorporated deoxynucleoside triphosphate (dNTP) opposite the pyridyl analogues. The rate of dATP insertion showed little dependence on aza substitution and ranged from only 2×10^5–$5.5\times10^5\,M^{-1}min^{-1}$. The magnitude and template independence of the rate of dATP insertion suggests that it results from nonspecific hydrophobic interactions. dGTP was the next most efficiently inserted natural triphosphate opposite **2Py** and **3Py**, but its insertion was not detectable opposite **4Py**. The rates are at least roughly correlated with the expected ability of the pyridine moiety to H-bond with guanosine (Scheme 2). The absolute rate at which the **2Py**:dG, **PYR**:dG ($1.2\times10^4\,M^{-1}min^{-1}$), and **BEN**:dG ($1.6\times10^3\,M^{-1}min^{-1}$) mispairs are formed suggests that while the *ortho* nitrogen atom does not form stable H-bonds with dG in duplex DNA, it facilitates dGTP insertion, although less efficiently than the *meta* nitrogen substituent of **3Py**. dTTP misincorporation opposite **2Py** and **3Py** proceeded with rates of $\approx5\times10^3\,M^{-1}min^{-1}$; however, its insertion was an order of magnitude more efficient opposite **4Py**. Perhaps the same factors that stabilize the **4Py**:dT mispair also contribute to its efficient synthesis. dCTP was not efficiently inserted opposite any of the pyridyl analogues. Overall, the rates at which **2Py** directs the synthesis of mispairs are similar to those of **BEN**,[3c] a result suggesting that the minor-groove H-bond acceptor has only a minor effect on mispair synthesis.

The data demonstrate that the phenyl-ring nucleobase scaffold may be optimized for both base-pair stability and replication by derivatization with an *ortho* nitrogen atom. This H-bond acceptor, which is at the α position relative to the glycosidic linkage, appears to be well positioned to mediate critical minor-groove interactions with water molecules of solvation or with polymerase-based H-bond donors. In contrast to H-bond acceptors introduced at the β position,[6] this H-bond acceptor does not stabilize mispairs in duplex DNA. While the *ortho* nitrogen atom does appear to somewhat facilitate dGTP mispair synthesis, the mispair is synthesized only inefficiently and, in fact, not significantly faster than mispairs among the natural nucleotides.[12] More problematic is the insertion of dATP opposite **2Py**, which is probably mediated by hydrophobic interactions. However, dATP insertion opposite **2Py** is less efficient than with other scaffolds that have already been successfully optimized by derivatization.[7,3b,c] Combination of these modifications with *ortho* aza substitution may result in the simultaneous optimization of interbase interactions and nucleobase–polymerase interactions and it should result in unnatural base pairs that are both more stable and efficiently replicated.

Received: June 28, 2006

Keywords: DNA replication · hydrophobic effects · nucleobases · polymerases · pyridines

[1] J. L. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* **1990**, *343*, 33–37.
[2] S. Moran, R. X.-F. Ren, S. Rumney IV, E. T. Kool, *J. Am. Chem. Soc.* **1997**, *119*, 2056–2057.
[3] a) D. L. McMinn, A. K. Ogawa, Y. Wu, J. Liu, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **1999**, *121*, 11585–11586; b) A. A. Henry, A. G. Olsen, S. Matsuda, C. Yu, B. H. Geierstanger, F. E. Romesberg, *J. Am. Chem. Soc.* **2004**, *126*, 6923–6931; c) S. Matsuda, A. A. Henry, F. E. Romesberg, *J. Am. Chem. Soc.* **2006**, *128*, 6369–6375; d) T. Mitsui, A. Kitamura, M. Kimoto, T. To, A. Sato, I. Hirao, S. Yokoyama, *J. Am. Chem. Soc.* **2003**, *125*, 5298–5307; e) K. Kinkaid, J. Beckman, A. Zivkovic, R. L. Halcomb, J. W. Engels, R. D. Kuchta, *Nucleic Acids Res.* **2005**, *33*, 2620–2628.
[4] Y. Li, G. Waksman, *Protein Sci.* **2001**, *10*, 1225–1233.
[5] A. S. Meyer, M. Blandolino, T. E. Spratt, *J. Biol. Chem.* **2004**, *279*, 33043–33046.
[6] 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.
[7] a) G. T. Hwang, F. E. Romesberg, *Nucleic Acids Res.* **2006**, *34*, 2037–2045; b) S. Matsuda, F. E. Romesberg, *J. Am. Chem. Soc.* **2004**, *126*, 14419–14427.
[8] a) I. Papai, G. Jansco, *J. Phys. Chem. A* **2000**, *104*, 2132–2137; b) L. Partay, P. Jedlovszky, G. Jansco, *J. Phys. Chem. B* **2005**, *109*, 8097–8102.
[9] a) A. N. Lane, T. C. Jenkins, T. A. Frenkiel, *Biochim. Biophys. Acta* **1997**, *1350*, 205–220; b) F. Eisenhaber, V. G. Tumanyan, F. Eisenmenger, W. Gunia, *Biopolymers* **1989**, *28*, 741–761; c) F.-Y.

Dupradeau, D. A. Case, C. Yu, R. Jimenez, F. E. Romesberg, *J. Am. Chem. Soc.* **2005**, *127*, 15612 – 15617.
[10] M. Berger, A. K. Ogawa, D. L. McMinn, Y. Wu, P. G. Schultz, F. E. Romesberg, *Angew. Chem.* **2000**, *112*, 3069 – 3071; *Angew. Chem. Int. Ed.* **2000**, *39*, 2940 – 2942.
[11] a) Y. X. Wang, S. Huang, D. E. Draper, *Nucleic Acids Res.* **1996**, *24*, 2666 – 2672; b) L. Zendlov, P. Hobza, M. Kabel, *ChemPhysChem* **2006**, *7*, 439 – 447.
[12] K. A. Eckert, T. A. Kunkel, *J. Biol. Chem.* **1993**, *268*, 13462 – 13471.

