Non-natural Base Pairs

DOI: 10.1002/ange.200601272

# Efforts towards Expansion of the Genetic Alphabet: Pyridone and Methyl Pyridone Nucleobases**

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

A non-natural base pair, which is stable in duplex DNA and enzymatically synthesized with high efficiency and selectivity, would greatly expand the utility of nucleic acids, both in terms of their genetic-coding capacity and chemical functionality. Unlike the natural base pairs, a third base pair does not necessarily require H-bonding,[1] and we,[2] and others,[3] have demonstrated that hydrophobic and van der Waals forces can mediate the selective interbase interactions required for base-pair stability and synthesis.

Our early efforts focused on nucleotides bearing nucleobase analogues with large extended aromatic surfaces. A variety of self-pairs (formed between the same analogue) and heteropairs (formed between different analogues) were identified. These pairs were both stable in duplex DNA and efficiently synthesized by the exonuclease deficient Klenow fragment of DNA polymerase I (Kf).[2a,4] However, replication was consistently limited by poor extension (i.e. continued primer elongation after synthesis of the non-natural pair), which we reasoned could be the result of interbase intercalation of the large aromatic scaffolds that distort the newly formed primer terminus and prevent its efficient extension. Our second-generation non-natural base pairs were formed between nucleotides that bear small, aromatic nucleobase analogues that were designed to be too small to intercalate and therefore more likely to interact edge on. Although several base pairs have been found that are more efficiently extended by Kf, the rates remain insufficient.[2b,5] Further-

[*] A. M. Leconte, S. Matsuda, G. T. Hwang, 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

[**] Acknowledgement for financial support is given to the National Institutes of Health (GM60005 to F.E.R.).

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

more, many of the smaller analogues bearing small hydrophobic nucleobases are also limited by facile and stable mispairing with dA.[2b,11] Thus, although small aromatic nucleobase analogues are promising, further optimization is required.

One factor that may contribute to the reduced Kf-mediated extension of the small aromatic pairs is their inability to engage the polymerase at a primer terminus. The natural base pairs have a similar geometric shape with a hydrogen bond acceptor located in an analogous position in the minor groove, N3 in purines or O2 in pyrimidines. When at the primer terminus, these H-bond acceptors have been shown to be important for polymerase-mediated extension.[6,7] Herein, we report the synthesis and characterization of nucleotides bearing pyridone-nucleobase analogues that are expected to position their carbonyl group into the minor groove (Figure 1a).

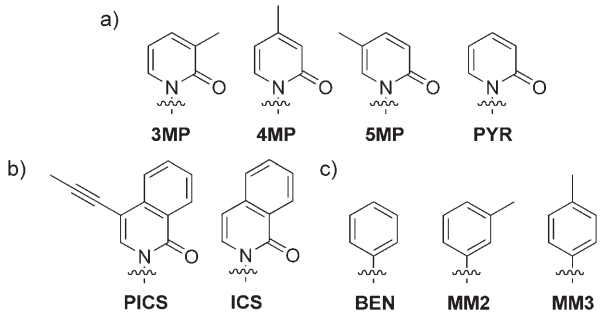


***Figure 1.*** Non-natural nucleobases. a) Pyridone analogues, b) previously reported ICS analogues,[2a,4] and c) previously reported carbocyclic analogues.[9,11] 3MP = 3-methylpyridone, 4MP = 4-methylpyridone, 5MP = 5-methylpyridone, PYR = pyridone.

The β-nucleotides, triphosphates, and phosphoramidites of the pyridone analogues were synthesized by a modification of a literature procedure[8] as described in the Supporting Information. To evaluate the thermodynamic stability of the non-natural base pairs in duplex DNA, each non-natural nucleoside was incorporated into complementary oligonucleotides and the melting temperature ($T_m$) of the resulting duplexes was determined by thermal denaturation experiments (Table 1). The non-natural self-pairs are generally more stable than mispairs formed with natural nucleotides, with $T_m$ values that range from 49.3 to 53.4 °C. With the exception of **3MP**, which is 3 °C more stable than the other self-pairs, the $T_m$ values are only slightly dependent on methyl substitution. The relative stability of the **3MP** self-pair likely reflects optimized packing of the 3-position methyl groups in the interbase interface. Also, with the exception of **3MP**, these pyridone self-pairs are less stable than the self-pairs of the corresponding carbocyclic analogues (Figure 1c).[9] This relative destabilization may result from reduced hydrophobicity and/or from electrostatic repulsion between the carbonyl groups in the minor groove.

To examine the thermal orthogonality of the non-natural base pairs, we determined the duplex melting temperatures with the non-natural nucleotides paired opposite each natural nucleotide (Table 1). These mispairs are significantly less stable than even the least-stable self-pair, with the exception

***Table 1:*** $T_m$ values of DNA duplexes.[a]

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

| X | Y | $T_m$ [°C] | X | Y | $T_m$ [°C] |
|---|---|---|---|---|---|
| **3MP** | **3MP** | 53.4 | **5MP** | **5MP** | 49.3 |
| **3MP** | A | 49.4 | **5MP** | A | 48.3 |
| **3MP** | C | 46.4 | **5MP** | C | 43.3 |
| **3MP** | G | 51.4 | **5MP** | G | 52.3 |
| **3MP** | T | 49.4 | **5MP** | T | 47.3 |
| **4MP** | **4MP** | 50.4 | **PYR** | **PYR** | 50.2 |
| **4MP** | A | 48.4 | **PYR** | A | 49.2 |
| **4MP** | C | 44.4 | **4MP** | C | 43.2 |
| **4MP** | G | 52.4 | **PYR** | G | 52.2 |
| **4MP** | T | 47.2 | **PYR** | T | 47.2 |
| A | T | 59.2[10] | A | G | 55.4[10] |

[a] For experimental details, see the Supporting Information.

of the mispairs with dG. The relative stability of the mispairs with dG is likely due to the increased aromatic surface area of guanine, as well as its exocyclic minor-groove amine, which may H-bond with the pyridone-nucleobase analogues. In contrast, the fully carbocyclic analogues form their most stable mispairs with dA,[9] likely due to the increased hydrophobicity of adenine relative to the other natural nucleobases. Relative to substitution at the 4- or 5-position, substitution at the 3-position has a more variable effect on mispair stability: it has little effect on the mispair with dA, stabilizes the mispairs with dC and dT, and destabilizes the mispair with dG. The stability of the **3MP** self-pair and the destabilization of the **3MP**:dG mispair results in a self-pair thermal selectivity of 2–7 °C, which approaches that of a natural base pair.[10]

The polymerase-mediated synthesis of these pyridone self-pairs was characterized by incorporating the analogues into template DNA and measuring the rate at which Kf incorporates the non-natural triphosphates (Table 2). Self-pair synthesis was inefficient, with second-order rate constants ($k_{cat}/K_M$) varying between $< 10^3$ and $4 \times 10^3\,M^{-1}\,min^{-1}$ (for comparison, a natural base pair is synthesized in a similar sequence context with a rate constant of $5 \times 10^7\,M^{-1}\,min^{-1}$ [4]). The **BEN** self-pair (Figure 1) was synthesized with a similar rate,[11] suggesting that the N-glycoside moiety does not facilitate non-natural base pair synthesis, at least not within the simple pyridone scaffold. The rates for pyridone self-pair synthesis are inefficient in contrast with self-pairs formed between other N-glycoside analogues bearing larger aromatic surface areas, such as **ICS** and **PICS** (Figure 1b), which are synthesized with second-order rate constants of approximately $10^5\,M^{-1}\,min^{-1}$.[2a,4] Perhaps the intercalation of the large

***Table 2:*** Steady-state rate constants of Kf-mediated incorporation of dXTP with dX in the template.[a]

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

| X | $k_{cat}$ [$min^{-1}$] | $K_M$ [μM] | $k_{cat}/K_M$ [$M^{-1}\,min^{-1}$] |
|---|---|---|---|
| **3MP** | $0.15 \pm 0.01$ | $76 \pm 23$ | $2.0 \times 10^3$ |
| **4MP** | $1.1 \pm 0.2$ | $389 \pm 96$ | $2.8 \times 10^3$ |
| **5MP** | $0.53 \pm 0.18$ | $135 \pm 65$ | $3.9 \times 10^3$ |
| **PYR** | n.d.[b] | n.d.[b] | $< 10^3$ |

[a] For experimental details, see the Supporting Information. [b] Rate too slow to determine $k_{cat}$ and $K_M$ independently.

  

aromatic rings reorients the pyridone carboxy groups or provides sufficient stabilization in the transition state to overcome their forced desolvation.

To explore synthesis as a function of methyl group substitution, we examined the synthesis of the 12 possible heteropairs (see the Supporting Information). Although most heteropairs are synthesized only inefficiently (with $k_{cat}/K_M$ between $< 10^3$ and $6 \times 10^3\,M^{-1}\,min^{-1}$), the triphosphate of **4MP** is inserted opposite **3MP** with an efficiency of $2 \times 10^4\,M^{-1}\,min^{-1}$, which is at least fivefold more efficient than either self-pair. The increased rate results largely from an increased $k_{cat}$, suggesting that the interbase interface is stabilized by packing interactions between the methyl groups in the developing transition state, and, importantly, that the pyridone scaffold may be optimized for synthesis.

To examine the kinetic orthogonality against pairing with the natural bases, we characterized the ability of the pyridone nucleobases to direct Kf to incorporate the natural deoxyribonucleotide triphosphates (dNTPs; see the Supporting Information). All of the mispairs are formed with a second-order rate constant of less than $5 \times 10^4\,M^{-1}\,min^{-1}$. The most efficiently synthesized mispairs resulted from insertion of deoxyguanosine triphosphate (GTP), which was inserted against the pyridone analogues significantly faster than against the corresponding carbocyclic analogues. For the carbocyclic analogues, each templates dGTP misincorporation at rates less than $2 \times 10^3\,M^{-1}\,min^{-1}$.[11] Methyl group substitution appears to have little effect on dGTP misincorporation, thus, it seems likely that the relatively efficient synthesis of the mispair is driven by a H-bond formed between the pyridone keto group and the dGTP amine.

After dGTP, the next most efficiently inserted natural triphosphate was generally deoxyadenosine triphosphate (dATP), which was inserted with a second-order rate constant ranging from $4 \times 10^3$–$2 \times 10^4\,M^{-1}\,min^{-1}$, with the exception of **3MP**, which did not template dATP insertion at a detectable rate ($k_{cat}/K_M < 10^3\,M^{-1}\,min^{-1}$). In contrast, the fidelity of the carbocyclic analogues is typically limited by facile insertion of dATP (often with rates in excess of $10^5\,M^{-1}\,min^{-1}$ [11]). This data further supports the previous suggestion that hydrophobic minor-groove functional groups increase the efficiency of dATP incorporation, possibly through a favorable packing interaction with the methyne group of dATP. Furthermore, the pyridones are more pyrimidine-like than their carbocyclic counterparts, and their mispairs with dA may be better recognized by the polymerase which evolved to discriminate against natural mispairs.[12]

The misincorporation of deoxythymidine triphosphate (dTTP) was inefficient ($k_{cat}/K_M < 5 \times 10^3\,M^{-1}\,min^{-1}$). Relative to the carbocyclic analogues,[11] inefficient incorporation of dTTP opposite the pyridone analogues generally appears to be due to weaker apparent binding of the natural triphosphate, again possibly reflecting forced desolvation of the minor-groove carbonyl groups in the transition state. dCTP was not incorporated opposite any pyridone analogue at a detectable rate ($k_{cat}/K_M < 10^3\,M^{-1}\,min^{-1}$).

As discussed above, the replication of predominantly hydrophobic non-natural base pairs is consistently limited by their extension.[5,11] Thus, understanding determinants of efficient extension is of great importance. To determine how the pyridone and methyl modifications impact this critical step, we examined the efficiencies with which Kf extended primers (by incorporation of the next correct triphosphate, deoxycytidine triphosphate; dCTP) that terminated with the self-pairs (Table 3). Both the **5MP** and **PYR** self-pairs were

***Table 3:*** Selected rate constants for Kf-mediated extension of non-natural termini by incorporation of dCTP.[a]

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

| X | Y | $k_{cat}$ [$min^{-1}$] | $K_M$ [μM] | $k_{cat}/K_M$ [$M^{-1}\,min^{-1}$] |
|---|---|---|---|---|
| **3MP** | **3MP** | $1.1 \pm 0.1$ | $270 \pm 107$ | $3.9 \times 10^3$ |
| **4MP** | **4MP** | $2.9 \pm 0.9$ | $202 \pm 56$ | $1.5 \times 10^4$ |
| **5MP** | **5MP** | $7.3 \pm 0.7$ | $107 \pm 12$ | $6.8 \times 10^4$ |
| **PYR** | **PYR** | $9.6 \pm 1.0$ | $117 \pm 32$ | $8.2 \times 10^4$ |
| **5MP** | **3MP** | $13 \pm 4$ | $198 \pm 21$ | $6.4 \times 10^4$ |
| **5MP** | **4MP** | $16 \pm 3$ | $204 \pm 56$ | $7.8 \times 10^4$ |
| **PYR** | **3MP** | $14 \pm 2$ | $322 \pm 76$ | $4.3 \times 10^4$ |
| **PYR** | **4MP** | $16 \pm 4$ | $146 \pm 56$ | $1.1 \times 10^5$ |

[a] For experimental details and additional heteropair extension rate constants, see the Supporting Information.

extended relatively efficiently at rates of $7 \times 10^4$ and $8 \times 10^4\,M^{-1}\,min^{-1}$, respectively, which is approximately 50-fold faster than the **BEN** self-pair.[11] The **4MP** and **3MP** self-pairs were extended less efficiently, at rates of $1 \times 10^4$ and $4 \times 10^3\,M^{-1}\,min^{-1}$. All of the pyridone self-pairs are extended significantly more efficiently than their carbocyclic analogues.[11] This is largely due to increased $k_{cat}$ values. The large difference in $k_{cat}$ between **PYR** and **BEN** self-pair extension suggests that the minor-groove H-bond is important for mediating interactions in the developing transition state. This agrees with previous studies that show that removing the minor groove H-bond acceptor from dG at a primer terminus results in a specific decrease in the extension $k_{pol}$.[13] As was suggested for the extension of a natural base pair,[6b] it seems likely that the pyridone minor groove H-bond acceptor stabilizes the rate-limiting transition state by engaging in a H-bonding network with Arg 668 of Kf and the ribosyl oxygen of the incoming dNTP.

To further explore extension as a function of methyl group substitution, we characterized the rates at which Kf extended a primer terminating at a heteropair (**X**:**Y**, in which **X** is the primer nucleobase and **Y** is the template nucleobase, Table 3 and the Supporting Information). Heteropairs formed between either **3MP** and **5MP** or **3MP** and **PYR** are extended at rates ranging from $4 \times 10^4$ to $9 \times 10^4\,M^{-1}\,min^{-1}$. These rates are 10- to 30-fold more efficient than those for the corresponding self-pairs and are similar to the extension of the **5MP** and **PYR** self-pairs. Likewise, heteropairs formed between **4MP** and either **5MP** or **PYR** were extended significantly faster than the **4MP** self-pair with rates ranging from $8 \times 10^4$ to $2 \times 10^5\,M^{-1}\,min^{-1}$. This suggests that steric interactions are responsible for the inefficient extension of the **3MP** and **4MP** self-pairs, and that shape complementarity is important for extension and possibly for the correct positioning of the minor groove H-bond acceptor to productively engage the polymerase. It is also interesting to note that the **4MP**:**5MP**, **PYR**:**4MP**, and **PYR**:**5MP** heteropairs are

  

extended within twofold of the most efficiently extended non-natural base pairs reported to date.

The data have several important implications for future efforts to design non-natural base pairs. Generally, at least for the pyridone scaffold, shape complementarity is less important for mispair stability and synthesis than is H-bonding. However, shape complementarity between the nucleobase analogues does appear to be important for base pair synthesis and extension. It may contribute to efficient synthesis by forming a well-packed interbase interface that stabilizes the developing transition state. It may contribute to efficient extension by contributing to the formation of a primer terminus that properly orients the minor-groove H-bond acceptor for polymerase recognition. The effects of shape complementarity were generally less pronounced with pairs formed between larger nucleobase analogues, likely owing to their intercalative mode of interaction.[2a,4] The data also show that, compared to the fully carbocyclic analogues, the pyridone base pairs are generally extended with dramatically increased rates and are also more orthogonal to the natural nucleotides, especially dA.[11] In contrast with other classes of predominantly hydrophobic nucleobase analogues, non-natural base pairs formed between the pyridone analogues appear to be limited by their synthesis efficiencies. However, the relative efficiency with which the **4MP**:**3MP** pair is synthesized suggests that these rates might be increased by nucleobase optimization, as has been observed for other scaffolds.[2b,14] In addition, the synthesis might be optimized through the directed evolution of polymerases to more efficiently catalyze this step as it has for the **PICS** self-pair.[15] Thus, further exploring these non-natural base-pair scaffolds and evolving polymerases that more efficiently synthesize them represents a promising strategy for the expansion of the genetic alphabet.

Received: March 31, 2006
Published online: May 30, 2006

**Keywords:** DNA replication · genetic alphabet · hydrophobic effect · nucleobases · pyridones

[1] S. Moran, R. X.-F. Ren, S. Rumney, E. T. Kool, *J. Am. Chem. Soc.* **1997**, *119*, 2056 – 2057.
[2] 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.
[3] I. Hirao, Y. Harada, M. Kimoto, T. Mitsui, T. Fujiwara, S. Yokoyama, *J. Am. Chem. Soc.* **2004**, *126*, 13298 – 13305.
[4] A. K. Ogawa, Y. Wu, D. L. McMinn, J. Liu, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **2000**, *122*, 3274 – 3287.
[5] G. T. Hwang, F. E. Romesberg, *Nucleic Acids Res.* **2006**, *34*, 2037 – 2045.
[6] a) J. C. Morales, E. T. Kool, *J. Am. Chem. Soc.* **1999**, *121*, 2323 – 2324; b) A. S. Meyer, M. Blandino, T. E. Spratt, *J. Biol. Chem.* **2004**, *279*, 33043 – 33046.
[7] a) Y. Li, G. Waksman, *Protein Sci.* **2001**, *10*, 1225 – 1233; b) S. J. Johnson, L. S. Beese, *Cell* **2004**, *116*, 803 – 816.
[8] F. E. Romesberg, C. Yu, S. Matsuda, A. A. Henry in *Current Protocols in Nucleic Acids Chemistry* (Eds.: S. L. Beaucage, D. E. Bergstrom, G. D. Glick, R. A. Jones), Wiley, New York, **2002**, pp. 1.5.1 – 1.5.36.
[9] S. Matsuda, F. E. Romesberg, *J. Am. Chem. Soc.* **2004**, *126*, 14419 – 14427.
[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] S. Matsuda, A. A. Henry, F. E. Romesberg, *J. Am. Chem. Soc.* **2006**, *128*, 6369 – 6375.
[12] K. Kincaid, J. Beckman, A. Zivkovic, R. L. Halcomb, J. W. Engels, R. D. Kuchta, *Nucleic Acids Res.* **2005**, *33*, 2620 – 2628.
[13] T. E. Spratt, *Biochemistry* **2001**, *40*, 2647 – 2652.
[14] a) A. K. Ogawa, Y. Wu, M. Berger, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **2000**, *122*, 8803 – 8804; b) C. Yu, A. A. Henry, F. E. Romesberg, P. G. Schultz, *Angew. Chem.* **2002**, *114*, 3997 – 4000; *Angew. Chem. Int. Ed.* **2002**, *41*, 3841 – 3844.
[15] A. M. Leconte, L. Chen, F. E. Romesberg, *J. Am. Chem. Soc.* **2005**, *127*, 12470 – 12471.

  