Published in final edited form as:
Chembiochem. 2008 November 24; 9(17): 2796–2799. doi:10.1002/cbic.200800577.

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

Yoshiyuki Hari[=],
Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037 (USA)

Gil Tae Hwang[=],
Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037 (USA)

Aaron M. Leconte [Prof.],
Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037 (USA)

Nicolas Joubert [Prof.],
Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Gilead Sciences & IOCB Research Center, CZ-16610, Prague 6 (Czech Republic)

Michal Hocek, and
Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Gilead Sciences & IOCB Research Center, CZ-16610, Prague 6 (Czech Republic)

Floyd E. Romesberg*
Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037 (USA)

Keywords

DNA; DNA replication; Genetic alphabet; Unnatural nucleobases; Pyridine

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 (*i.e.* by insertion of the unnatural dNTP opposite its partner in the template) by the exonuclease-deficient Klenow fragment of *E. coli* DNA polymerase I (Kf), none are efficiently extended (*i.e.* by continued primer elongation), most likely due to interstrand nucleobase intercalation and distortion of the primer terminus.[10] Thus, a range of nucleotides bearing smaller phenyl-based nucleobases that should be 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

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

*Fax: (+1) 858-784-7472, E-mail: floyd@scripps.edu.
[=]These authors contributed equally to this work.

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 analog might optimize heteropair replication.

Here, we report the synthesis and characterization of a series of substituted 2-pyridine derivatives (Figure 1) designed to systematically examine the effect of nucleobase shape, size, and hydrophobicity as well as structural and electronic modifications within the interbase interface. The analogs 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.

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 $6.2 \times 10^3\ M^{-1}min^{-1}$),[3] demonstrating 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 $4.1 \times 10^4\ M^{-1}min^{-1}$, 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 analogs 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 efficiently 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 5- and 2-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 2-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 analogs 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 with a 6- to 12-fold increase in the rate of self pair extension with methyl substitution. The effects are roughly additive, with extension of the d**45DMPy** self pair 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 identified to date. This data suggests 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** analogs 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 10-fold slower than that for insertion of d**MMO2**TP. Likewise, d**5SICS**TP is inserted opposite either 2-pyridyl analog in the template with rates that are ~10-fold less efficient than insertion opposite d**MMO2**. Thus, at least with these analogs, 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 10- to 100-fold slower than the 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 6-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 analogs 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 11. 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 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 is 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 analogs. Judicious placement of methyl groups

alone, yielding d**45DMPy**, results in a self pair that is extended with a natural-like rate and can be used to synthesize site-specifically modified DNA in good yield. Moreover, while none of the pyridyl analogs 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 analogs and further optimization should result in self pairs or heteropairs for *in vitro*, and possibly even *in vivo*, applications.[13,14]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Henry AA, Olsen AG, Matsuda S, Yu C, Geierstanger BH, Romesberg FE. J. Am. Chem. Soc 2004;126:6923–6931. [PubMed: 15174862]
2. Leconte AM, Matsuda S, Hwang GT, Romesberg FE. Angew. Chem. Int. Ed 2006;45:4326–4329.
3. Kim Y, Leconte AM, Hari Y, Romesberg FE. Angew. Chem. Int. Ed 2006;45:7809–7812.
4. Matsuda S, Henry AA, Romesberg FE. J. Am. Chem. Soc 2006;128:6369–6375. [PubMed: 16683801]
5. Leconte AM, Hwang GT, Matsuda S, Capek P, Hari Y, Romesberg FE. J. Am. Chem. Soc 2008;130:2336–2343. [PubMed: 18217762]
6. Yang Z, Sismour AM, Sheng P, Puskar NL, Benner SA. Nucleic Acids Res 2007;35:4238–4249. [PubMed: 17576683]
7. Krueger AT, Lu H, Lee AHF, Kool ET. Acc. Chem. Res 2007;40:141–150. [PubMed: 17309194]
8. Hirao I, Mitsui T, Kimoto M, Yokoyama S. J. Am. Chem. Soc 2007;129:15549–15555. [PubMed: 18027940]
9. a) Chairamonte M, Moore CL, Kincaid K, Kuchta RD. Biochemistry 2003;42:10472–10481. [PubMed: 12950174] b) Kincaid K, Beckman J, Zivkovic A, Halcomb RL, Engels JW, Kuchta RD. Nucleic Acids Res 2005;33:2620–2628. [PubMed: 15879351]
10. Matsuda S, Fillo JD, Henry AA, Rai P, Wilkens SJ, Dwyer TJ, Geierstanger BH, Wemmer DE, Schultz PG, Spraggon G, Romesberg FE. J. Am. Chem. Soc 2007;129:10466–10473. [PubMed: 17685517]
11. Ogawa A, Wu Y, McMinn DL, Liu J, Schultz PG, Romesberg FE. J. Am. Chem. Soc 2000;122:3274–3287.
12. Urban M, Pohl R, Klepetarova B, Hocek M. J. Org. Chem 2006;71:7322–7328. [PubMed: 16958526]
13. a) Ieven MJ. Clin. Virol 2007;40:259–276. b) Cropp TA, Chin JW. Curr. Opin. Chem. Biol 2006;10:601–606. [PubMed: 17055771] c) Bittker JA, Phillips KJ, Liu DR. Curr. Opin. Chem. Biol 2002;6:367–374. [PubMed: 12023118]
14. a) Wang L, Xie J, Schultz PG. Annu. Rev. Biophys. Biomol. Struct 2006;35:225–249. [PubMed: 16689635] b) Xie J, Schultz PG. Curr. Opin. Chem. Biol 2005;9:548–554. [PubMed: 16260173]

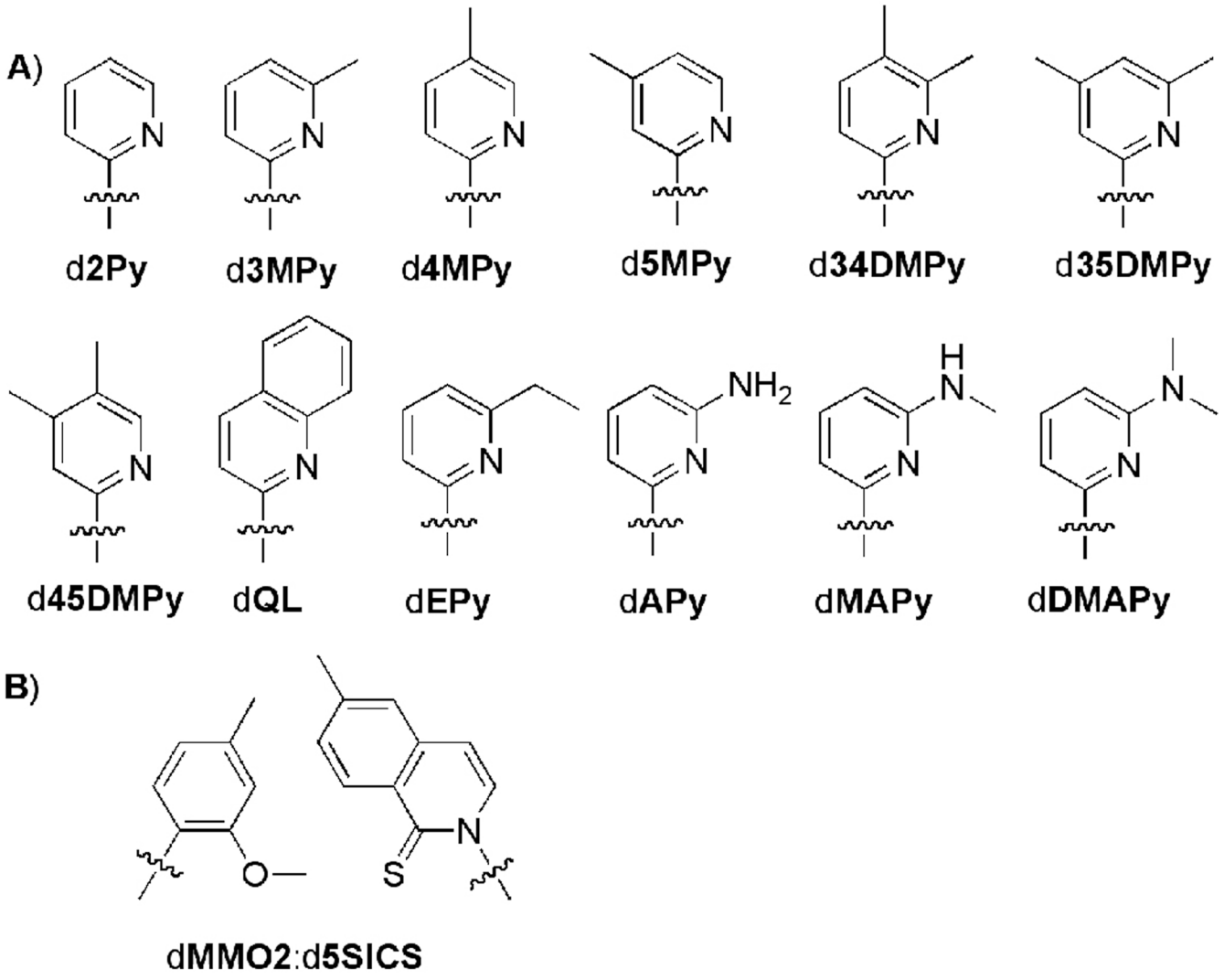

Figure 1.

a) 2-Pyridyl nucleotides synthesized and characterized in this study. b) d**MMO2**:d**5SICS** base pair. Only the nucleobase analog is shown with the wavy line indicating connection to the sugar and phosphate backbone, which have been omitted for clarity.

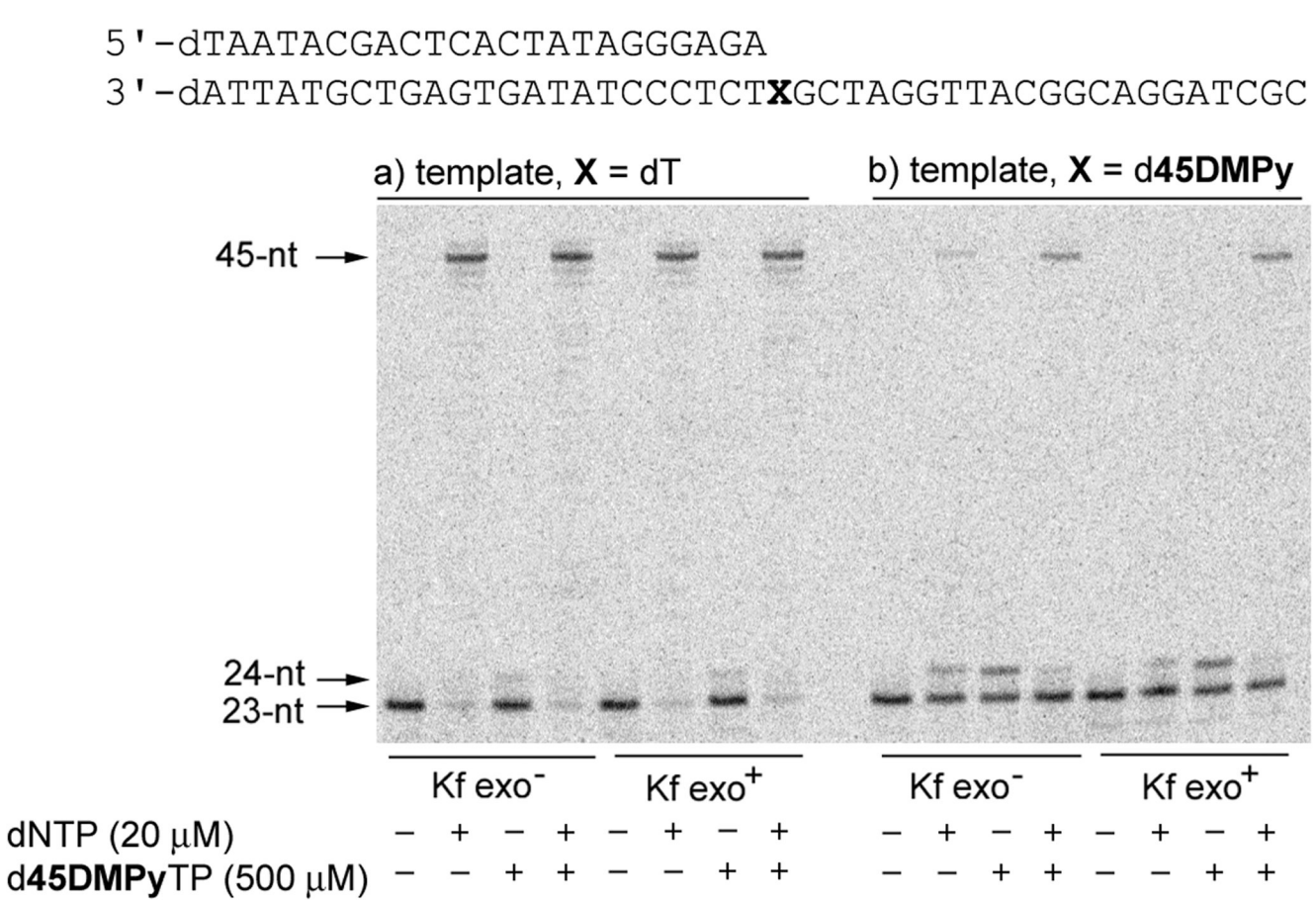

Figure 2.
Extension of 23-nt primer with 45-nt templates (**X** = dT or d**45DMPy**). Reactions contain either Kf exo$^+$ or Kf exo$^-$ and the reaction time was 3 min.

Table 1

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

5'-dTAATACGACTCACTATAGGGAGA
3'-dATTATGCTGAGTGATATCCCTCTXGCTAGGTTACGGCAGGATCGC

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]	$<1.0 \times 10^3$
d**APy**	nd[b]	nd[b]	$<1.0 \times 10^3$
d**MAPy**	nd[b]	nd[b]	$<1.0 \times 10^3$
d**DMAPy**	0.26 ± 0.02	162 ± 42	1.6×10^3

[a] See Supporting Information for details.

[b] Reaction was too inefficient for k_{cat} and K_M to be determined independently.

Table 2

Extension rates of unnatural self pairs.[a]

5'-dTAATACGACTCACTATAGGGAGAX
3'-dATTATGCTGAGTGATATCCCTCTXGCTAGGTTACGGCAGGATCGC

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]	$<1.0 \times 10^3$

[a] See Supporting Information for details.

[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'-dATTATGCTGAGTGATATCCCTCTYGCTAGGTTACGGCAGGATCGC

dXTP	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] See Supporting Information for details.