Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters
Bioorganic & Medicinal Chemistry Letters 14 (2004) 4887–4890

# Unnatural base pairs between 2- and 6-substituted purines and 2-oxo(1H)pyridine for expansion of the genetic alphabet

Ichiro Hirao,[a,b,*] Tsuyoshi Fujiwara,[b] Michiko Kimoto[b] and Shigeyuki Yokoyama[b,c,d,*]

[a]*Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan*
[b]*Protein Research Group, RIKEN Genomic Sciences Center (GSC), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan*
[c]*Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[d]*RIKEN Harima Institute at SPring-8, 1-1-1 Kouto, Mikazuki-cho, Sayo, Hyogo 679-5148, Japan*

Received 18 June 2004; revised 15 July 2004; accepted 17 July 2004
Available online 6 August 2004

**Abstract**—An unnatural base pair between 2-amino-6-(2-thienyl)purine (denoted by **s**) and 2-oxo(1H)pyridine (denoted by **y**) shows high selectivity in transcription and translation. Toward the further development of unnatural base pairs that also have exclusive selectivity in replication, we examined the roles of the 2-amino and 6-thienyl groups of **s** using base pairs between **y** and purine-analogs, 6-thienylpurine and 2-amino-6-furanylpurine, as well as **s**. The results obtained from the thermal stability and DNA polymerase single-nucleotide insertion experiments suggest that the 2-amino group of **s** contributes toward the shape complementarity of the pairing with **y**, rather than the hydrogen bonding with the 2-keto group of **y**. In addition, the bulkiness of positions 2 and 6 of the unnatural purines cooperatively determines the selectivity of the noncanonical pairing with **y** or the natural pyrimidines in replication. This information is useful not only for the development of unnatural, orthogonal base pairs, but also for understanding the mechanisms of base pair formation in replication.


Unnatural base pair studies provide valuable information for understanding the mechanisms of base pair selectivity in replication and transcription.[1–6] This information is also useful for the further development of unnatural base pairs that can function in replication, transcription, and translation.[7,8] These unnatural base pairs enable the site-specific incorporation of extra components into nucleic acids and proteins, and eventually will permit the expansion of the genetic alphabet and code.[9–12]

Recently, we developed several types of unnatural bases and base pairs. Among them, the base pair between 2-amino-6-(2-thienyl)purine (denoted by **s**) and 2-oxo(1H)pyridine (denoted by **y**) was designed, based on the concepts of hydrogen-bonding patterns and shape complementarity (Fig. 1a).[13,14] The bulky 6-thienyl group of **s** prevents noncognate pairing with the natural bases, but the relatively small hydrogen at position 6 of **y** maintains the shape complementarity of the **s**–**y** pairing. The **s**–**y** pair works with the natural A–T(U) and G–C base pairs, especially in transcription, and the nucleoside 5′-triphosphates of **y** and 5-substituted **y** bases can be site-specifically incorporated into RNA opposite **s** in DNA templates by T7 RNA polymerase.[14–16] Thus, the **s**–**y** pair is useful for the development of RNA molecules with increased functionality. In addition, in a coupled transcription–translation system, the **s**–**y** pair can expand the genetic code and enable the site-specific incorporation of amino acid analogs into proteins.[14]

The **s**–**y** pair also shows specificity in replication,[13] but the selectivity is not yet sufficient for practical uses, such as PCR and in vivo experiments. In DNA amplification, even if a small mis-incorporation occurs by incorrect base pairing, the error significantly accumulates in the DNA products. In addition, the complementarity of the **s**–**y** pair is not so high that the **s** incorporation opposite **y** into nucleic acids is lower than the **y** incorporation

*Keywords*: Unnatural base pairs; Transcription; Replication.
* Corresponding authors. Tel.: +81-45-503-9644; fax: +81-45-503-9645 (I.H.); tel.: +81-35-841-4413; fax: +81-35-841-8057 (S.Y.); e-mail addresses: ihirao@postman.riken.jp; yokoyama@biochem.s.u-tokyo.ac.jp

0960-894X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2004.07.038

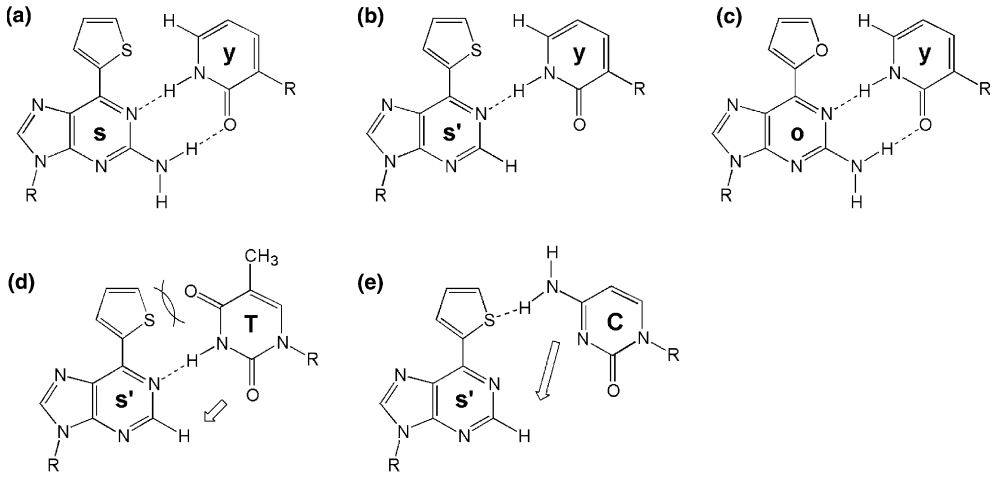


**Figure 1.** Unnatural and noncognate base pairs. In these structures, R denotes ribose. The arrows in the structures of the **s**′–C (d) and **s**′–T (e) pairs show the possible movement of each pyrimidine.

opposite **s**. Thus, unnatural base pairs that have exclusive selectivity and complementarity in replication are required.

Toward the further development of unnatural base pairs with exclusive selectivity and complementarity in replication, we synthesized two **s** analogs, 6-thienylpurine (denoted by **s**′) and 2-amino-6-furanylpurine (denoted by **o**),[13] and investigated the roles of the 2- and 6-groups of **s** by comparisons with these analogs. Here, we report the characterizations of the **s**–**y**, **s**′–**y**, and **o**–**y** pairs (Fig. 1a–c) through investigations of the thermal stability of DNA duplexes and the enzymatic insertion of the 2′-deoxyribonucleoside 5′-triphosphate of **y** (d**y**TP) into DNA opposite **s**, **s**′, or **o** in DNA templates by the Klenow fragment of *Escherichia coli* DNA polymerase I.

The syntheses of the 2′-deoxyribonucleosides of **s**′ and **o** were carried out by the same procedure as that for **s**,[13,17] and these nucleosides were converted to the amidites for DNA synthesis. The 2-amino group of **o** was protected with a phenoxyacetyl group, because the isobutyryl and acetyl groups were too stable as protecting groups. We also synthesized the 2′-deoxyribonucleoside of 6-furanylpurine (**o**′).[13] However, this nucleoside is unstable under the alkaline conditions used for DNA synthesis, and thus we did not further examine the nucleoside of **o**′. The DNA fragments (12-mer and 35-mer) containing **s**, **s**′, or **o** were synthesized with an Applied Biosystems Model 392 DNA synthesizer.

Assessments of the thermal stabilities of DNA duplexes containing the **s**–**y**, **s**′–**y**, or **o**–**y** pair were performed using the DNA fragments, 5′-GGTAAC**N**$_1$ATGCG (**N**$_1$ = **s**, **s**′, or **o**), and 5′-CGCATN$_2$GTTACC (**N**$_2$ = **y**, T, or C) (each 5 μM), in 10 mM sodium phosphate (pH 7.0) containing 100 mM NaCl and 0.1 mM EDTA. The melting temperature ($T_m$) of each DNA duplex is shown in Table 1. Interestingly, the thermal stability of the DNA duplex containing the **s**′–**y** pair was as high as that containing the **s**–**y** pair, and in comparison to the noncognate **s**–T and **s**–C pairs, the **s**′-T, and **s**′–C pairs exhibited slightly enhanced duplex stabilities. These results suggest that the 2-amino group of **s** is less functional as a proton donor group for hydrogen bonding with the 2-keto groups of **y** and the natural pyrimidines. This finding agrees with the fact that the 2-amino group of **s** is less chemically reactive than that of G.[13]

The thermal stability of the DNA duplex containing the **o**–**y** pair was higher than that of the duplex containing the **s**–**y** pair. In comparison to the noncognate **s**–T and **s**–C pairs, the **o**–T pair decreased the duplex stability, but the **o**–C pair increased it. The oxygen atom in the 6-furanyl group of **o** is sterically smaller than the sulfur atom in the 6-thienyl group of **s**, and the hydrogen bonding ability of the oxygen atom is higher than that of the sulfur atom. Thus, **y** sterically fits better with **o** than with **s**. In addition, the oxygen atom of **o** electrostatically clashes with the 4-keto group of T, but attracts the 4-amino group of C.

The efficiency of the unnatural base pairings in replication was assessed by single-nucleotide insertion experiments, using d**y**TP, DNA templates containing **s**, **s**′, or **o**, and the exonuclease-deficient Klenow fragment.[1,18] The steady-state kinetic parameters (Table 2) were determined by using the automated ABI 377 DNA sequencer with the *GeneScan* software and a primer labeled with 6-carboxyfluorescein.[19] The **y** incorporation opposite **s**′ was 3.1-fold less efficient than that opposite **s** (Table 2, entries 6 and 1). On the contrary, the **y** incorporation opposite **o** was 3.7-fold more efficient than that opposite **s** (Table 2, entries 11 and 1). In the context of the noncognate pairing, the T incorporation opposite **s**′ was 2.5-fold more efficient than that opposite **s** (Table 2, entries 7 and 2), although the C incorporation opposite **s**′ was 23-fold less than that opposite **s** (Table 2, entries 8 and 3). The noncognate pairings between **o** and the natural pyrimidines showed relatively high efficiency; the T incorporation opposite **o** was 3.9-fold more efficient than that opposite **s** (Table 2, entries 12 and 2), and the efficiency of the C incorporation opposite **o** was as high as that opposite **s** (Table 2, entries 13 and 3).

**Table 1.** $T_m$ values for DNA duplexes containing the unnatural base pairs

5′-d(GGTAAC**N**$_1$ATGCG)
3′-d(CCATTG**N**$_2$TACGC)

| $N_1$–$N_2$ | $T_m$ (°C)[a] | $N_1$–$N_2$ | $T_m$ (°C) | $N_1$–$N_2$ | $T_m$ (°C) |
|---|---|---|---|---|---|
| **s**–**y** | 43.4 | **s**′–**y** | 43.4 | **o**–**y** | 44.9 |
| **s**–T | 40.3 | **s**′–T | 41.0 | **o**–T | 39.5 |
| **s**–C | 41.1 | **s**′–C | 41.6 | **o**–C | 43.7 |

[a] The absorbance at 260 nm of the DNA fragments (5 μM), in 10 mM sodium phosphate (pH 7.0), 100 mM NaCl, and 0.1 mM EDTA, was monitored on a Beckman model DU650 spectrometer.

**Table 2.** Steady-state kinetic parameters for insertion of single nucleotides into a template-primer duplex by the exonuclease-deficient Klenow fragment[a]

Primer 5′-ACTCACTATAGGGAGGAAGA
Template 1 3′-TATTATGCTGAGTGATATCCCTCCTTCT**N**TCTCGA
Template 2 3′-TATTATGCTGAGTGATATCCCTCCTTCT**N**TCTCTT

| Entry | Template base **N** (template) | Nucleoside triphosphate | $K_M$ (μM) | $V_{max}$ (%min$^{-1}$)[d] | Efficiency, $V_{max}/K_M$ (%min$^{-1}$M$^{-1}$) |
|---|---|---|---|---|---|
| 1 | **s** (template 1) | **y** | 170 (40)[b] | 11 (2) | $6.5\times10^4$ |
| 2 | **s** (template 1) | T | 270 (40) | 3.1 (0.2) | $1.1\times10^4$ |
| 3 | **s** (template 1) | C | 430 (180) | 18 (9) | $4.2\times10^4$ |
| 4 | **s** (template 1) | A | 87 (27) | 0.31 (0.07) | $3.6\times10^3$ |
| 5 | **s** (template 1) | G | Nd[c] | Nd | — |
| 6 | **s**′ (template 2) | **y** | 180 (70) | 3.8 (0.7) | $2.1\times10^4$ |
| 7 | **s**′ (template 2) | T | 180 (70) | 4.9 (0.6) | $2.7\times10^4$ |
| 8 | **s**′ (template 2) | C | 210 (80) | 0.37 (0.04) | $1.8\times10^3$ |
| 9 | **s**′ (template 2) | A | 150 (50) | 3.1 (0.4) | $2.1\times10^4$ |
| 10 | **s**′ (template 2) | G | Nd | Nd | — |
| 11 | **o** (template 2) | **y** | 190 (90) | 46 (22) | $2.4\times10^5$ |
| 12 | **o** (template 2) | T | 280 (90) | 12 (3) | $4.3\times10^4$ |
| 13 | **o** (template 2) | C | 520 (290) | 27 (1) | $5.2\times10^4$ |
| 14 | **o** (template 2) | A | 84 (28) | 0.31 (0.05) | $3.7\times10^3$ |
| 15 | **o** (template 2) | G | Nd | Nd | — |
| 16 | **A** (template 2) | T | 1.2 (0.3) | 3.5 (1.3) | $2.9\times10^6$ |
| 17 | **A** (template 2) | C | 1600 (600) | 2.5 (1.3) | $1.6\times10^3$ |

[a] Assays were carried out at 37 °C for 1–20 min using 5 μM template-primer duplex, 3–50 nM enzyme, and 0.6–2100 μM nucleoside triphosphate in a solution (10 μL) containing 50 mM Tris · HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM DTT, and 0.05 mg/mL bovine serum albumin.
[b] Standard deviations are given in parentheses.
[c] Nd: No inserted products were detected after an incubation for 20 min with 1500 μM nucleoside triphosphate and 50 nM enzyme.
[d] The values were normalized to the enzymatic concentration (20 nM) for the various enzyme concentrations used.

A comparison between **s** and **s**′ suggests that the protruding 2-amino group of **s** is important for fixing the locations of the pairing partners. The removal of the 2-amino group of **s** may allow the T to have more flexibility in the **s**′–T pairing, to relieve the steric hindrance between the thienyl group of **s**′ and the 4-keto group of T (Fig. 1d). This distortion of the **s**′–T pairing is permitted by the polymerase recognition. Thus, **s**′ loses the recognition between **y** and T, and the $K_M$ and $V_{max}$ values of the **s**′–**y** pair are similar to those of the **s**′–T pair. Similarly, the C in the **s**′–C pairing also moves to relieve the steric hindrance between the thienyl group of **s**′ and the 4-amino group of C (Fig. 1e). Although the sulfur atom in the 6-thienyl group may electrostatically interact with the 4-amino group of C, the nitrogen at position 1 of **s**′ is more attractive to the 4-amino group. Thus, **s**′ promotes the wobble pairing with C to form the hydrogen bond between the 1-nitrogen atom of **s**′ and the 4-amino group of C. However, the large movement of the C in the **s**–C pair is prevented by the 2-amino group of **s**. Thus, the distortion of the **s**′–C pair might be greater than that of the **s**–C pairing. The kinetic parameters of the **s**′–C pairing show that this large distortion disadvantageously affects the $V_{max}$ value (Table 2, entry 8). In contrast, changing the template base **s** to **o** reduces the steric hindrance of the purine analog, and thereby increases the incorporation efficiencies of all of the pyrimidines, T and C, and the **y** bases.

The $T_m$ values and the incorporation efficiencies of these base pairings provide clues about the roles of the 2-amino group and the 6-bulky heterocycles of the purine analogs in replication. In the **s**′–**y**, **s**′–C, **o**–T, and **o**–C pairs, there are low correlations between the incorporation efficiencies and the $T_m$ values of the DNA duplexes. This suggests that the stable structures of the noncanonical base pairs in the DNA duplexes are not always the same as those of the base pairing between the template base and the incoming base in the polymerase complex. Nevertheless, the $T_m$ values indicate that the 6-thienyl and 6-furanyl groups function sterically and electrostatically, but the 2-amino group functions sterically, rather than electrostatically, for the base pair formation. In addition, the single-nucleotide insertion experiments suggest

that the cooperation between the 2-amino group and the 6-bulky heterocycles determines the base pair selectivity in replication. The combination of the 2-amino and 6-thienyl groups of **s** excludes the noncognate pairing with T. In contrast, in the absence of the 2-amino group, the 6-thienyl group effectively excludes another noncognate pairing with C. In addition, for efficient pairing with **y**, the 6-thienyl group of **s** might be sterically large.[20] Thus, the 6-furanyl group of **o** improves the shape fitting with **y**, although the efficiency of the noncognate pairing with C or T also increases.

In some cases, certain noncanonical base pairs, which show high efficiency in the single-nucleotide insertion experiments, have low processibility after the base pair insertion by DNA polymerases.[21] As shown in Table 2 (entries 4, 9, and 14), the A incorporation showed relatively high efficiency, especially opposite **s**′. However, after the A incorporation opposite the purine analogs, the primer elongations were prevented (data not shown). In contrast, the incorporation efficiencies of the base pairings of the purine analogs with **y**, T, and C correlated with their elongation efficiencies after the incorporation (data not shown). Thus, the kinetic parameters obtained by the single-nucleotide insertion experiments for the base pairs between purine analogs and **y** or pyrimidines reflect the selectivity of these pairings in replication.

In this report, we examined the selectivity of three purine analogs, **s**, **s**′, and **o**, pairing with **y** by $T_m$ measurements of the DNA duplexes and single-nucleotide insertion experiments using the Klenow fragment. We also characterized the 2-amino and 6-thienyl groups of **s** in the cognate and noncognate pairings in replication. Combining these results with the structure modeling[22] of these base pairs in the template–substrate–polymerase complex would make it possible to predict the geometric distortion range of the noncanonical base pairs that would be tolerated by DNA polymerases. This information will be useful for the further development of unnatural base pairs and for understanding the mechanisms of base pair formation in replication.

## Acknowledgements

This work was supported by the RIKEN Structural Genomics/Proteomics Initiative (RSGI), the National Project on Protein Structural and Functional Analyses, Ministry of Education, Culture, Sports, Science and Technology of Japan, and by a Grant-in-Aid for Scientific Research (KAKENHI 15350097) from the Ministry of Education, Culture, Sports, Science, and Technology.

## References and notes

1. Goodman, M. F.; Creighton, S.; Bloom, L. B.; Petruska, J. *Crit. Rev. Biochem. Mol. Biol.* **1993**, *28*, 83.
2. Kool, E. T.; Morales, J. C.; Gukian, K. M. *Angew. Chem., Int. Ed.* **2000**, *39*, 990.
3. Kool, E. T. *Ann. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 1.
4. Paul, N.; Nashine, V. C.; Hoops, G.; Zhang, P.; Zhou, J.; Bergstrom, D. E.; Davisson, V. J. *Chem. Biol.* **2003**, *10*, 815.
5. Chiaramonte, M.; Moore, C. L.; Kincaid, K.; Kuchta, R. D. *Biochemistry* **2003**, *42*, 10472.
6. Lee, H. J.; Lim, H. M.; Adhya, S. *J. Biol. Chem.* **2004**, *279*, 16899.
7. Kool, E. T. *Curr. Opin. Chem. Biol.* **2000**, *4*, 602.
8. Henry, A. A.; Romesberg, F. E. *Curr. Opin. Chem. Biol.* **2003**, *7*, 727.
9. Bain, J. D.; Switzer, C.; Chamberlin, A. R.; Benner, S. A. *Nature* **1992**, *356*, 537.
10. Tor, Y.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 4461.
11. Benner, S. A. *Trends Biotechnol.* **1994**, *12*, 158.
12. Ohtsuki, T.; Kimoto, M.; Ishikawa, M.; Mitsui, T.; Hirao, I.; Yokoyama, S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4922.
13. Fujiwara, T.; Kimoto, M.; Sugiyama, H.; Hirao, I.; Yokoyama, S. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2221.
14. Hirao, I.; Ohtsuki, T.; Fujiwara, T.; Mitsui, T.; Yokogawa, T.; Okuni, T.; Nakayama, H.; Takio, K.; Yabuki, T.; Kigawa, T.; Kodama, K.; Yokogawa, T.; Nishikawa, K.; Yokoyama, S. *Nat. Biotechnol.* **2002**, *20*, 177.
15. Kimoto, M.; Endo, M.; Mitsui, T.; Okuni, T.; Hirao, I.; Yokoyama, S. *Chem. Biol.* **2004**, *11*, 47.
16. Endo, M.; Mitsui, T.; Okuni, T.; Kimoto, M.; Hirao, I.; Yokoyama, S. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2593.
17. Hocek, M.; Holy, A.; Votruba, I.; Dvořáková, H. *Collect. Czech. Chem. Commun.* **2001**, *66*, 483.
18. Petruska, J.; Goodman, M. F.; Boosalis, M. S.; Sowers, L. C.; Cheong; Tinoco, I. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6252.
19. Kimoto, M.; Yokoyama, S.; Hirao, I. *Biotechnol. Lett.* **2004**, *26*, 999.
20. Bergstrom, D. E. *Chem. Biol.* **2004**, *11*, 18.
21. Tae, E. L.; Wu, Y.; Xia, G.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2001**, *123*, 7439.
22. Zhang, P.; Johnson, W. T.; Klewer, D.; Paul, N.; Hoops, G.; Davisson, V. J.; Bergstrom, D. E. *Nucleic Acids Res.* **1998**, *26*, 2208.