J|A|C|S
ARTICLES
Published on Web 05/25/2005

An Efficient Unnatural Base Pair for a Base-Pair-Expanded Transcription System

Tsuneo Mitsui,[†] Michiko Kimoto,[‡] Yoko Harada,[‡] Shigeyuki Yokoyama,*[,‡,§,||] and Ichiro Hirao*[,†,‡]

Contribution from the Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan, Protein Research Group, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan, Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan, and RIKEN Harima Institute at SPring-8, 1-1-1 Kohto, Mikazuki-cho, Sayo, Hyogo 679-5148, Japan

Received December 13, 2004; E-mail: ihirao@riken.jp; yokoyama@biochem.s.u-tokyo.ac.jp

Abstract: For the site-specific incorporation of artificial components into RNA by transcription, an efficient, unnatural base pair between 2-amino-6-(2-thiazolyl)purine (denoted as **v**) and 2-oxo(1H)pyridine (denoted as **y**) was developed. The substrates of **y** and 5-substituted **y** were site-specifically incorporated into RNA by T7 RNA polymerase opposite **v** in templates. The efficiency and fidelity of the **v**−**y** pairing in transcription were as high as those of the natural A−T(U) and G−C pairings. Furthermore, RNAs containing two adjacent **y** bases were also transcribed from DNA templates containing two **v** bases. This specific transcription allows the large-scale preparation of artificial RNAs and can be combined with other systems to simultaneously incorporate several different components into a transcript.

Transcription mediated by unnatural base pairs enables the site-specific incorporation of artificial components into RNA for generating novel RNA molecules with increased functionality.[1,2] In addition, the combination of this specific transcription with conventional translation systems can expand the genetic code toward the synthesis of proteins containing amino acid analogues.[3,4] For these purposes, the creation of unnatural base pairs that are recognized and replicated by polymerases has been studied on the basis of several different concepts, relying on nonstandard hydrogen-bonding patterns,[5,6] shape complementarity,[7−9] and hydrophobic interactions.[10−12] However, a few unnatural base pairs, such as isoguanine (isoG)−isocytosine (isoC)[5,13] and xanthosine−diaminopyrimidine,[6] have been tested in transcription. Only one example of the incorporation of a functional component into RNA, using N^6-(6-aminohexyl)isoG and 5-methyl-isoC, has been reported, and the yield of the transcript containing N^6-(6-aminohexyl)isoG was approximately 50% of that for transcription using natural templates and substrates.[14] There was no other extra base pair for the efficient, site-specific incorporation of a variety of artificial components into RNA.

Recently, we developed unnatural base pairs of 2-amino-6-(*N*,*N*-dimethylamino)purine (denoted by **x**) and 2-oxo(1H)-pyridine (denoted by **y**)[15] and of 2-amino-6-(2-thienyl)purine (denoted by **s**) and **y**[16] (Figure 1). These unnatural base pairs were designed to employ different shape fitting and hydrogen-bonding patterns from those of the natural base pairs, and they exhibit high specificity in transcription. In particular, the fidelity of the **y** incorporation into RNA opposite **s** in templates by T7 RNA polymerase is as high as that of the natural base incorporations opposite the complementary bases. Using the **s**−**y** pair, we have created an extra codon−anticodon interaction and achieved the site-specific incorporation of an amino acid analogue, 3-chlorotyrosine, into a protein.[16] In addition, chemically synthesized 5-modified **y**, such as 5-iodo-**y** (I-**y**)[17] and 5-phenylethynyl-**y** (Ph-**y**),[18] can also be incorporated into

[†] Research Center for Advanced Science and Technology, The University of Tokyo.
[‡] RIKEN Genomic Sciences Center.
[§] Graduate School of Science, The University of Tokyo.
[||] RIKEN Harima Institute at SPring-8.

(1) Benner, S. A.; Burgstaller, P.; Battersby, T. R.; Jurczyk, S. In *The RNA World*, 2nd ed.; Gesteland, R. F., Cech, T. R., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1999; p 163.
(2) Bergstrom, D. E. *Chem. Biol.* **2004**, *11*, 18.
(3) Wang, L.; Schultz, P. G. *Chem. Commun.* **2002**, 1.
(4) Hendrickson, T. L.; de Crécy-Lagard, V.; Schimmel, P. *Annu. Rev. Biochem.* **2004**, *73*, 147.
(5) Switzer, C.; Moroney, S. E.; Benner, S. A. *J. Am. Chem. Soc.* **1989**, *111*, 8322.
(6) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33.
(7) Morales, J. C.; Kool, E. T. *Nat. Struct. Biol.* **1998**, *5*, 950.
(8) Matray, T. J.; Kool, E. T. *Nature* **1999**, *399*, 704.
(9) Rappaport, H. P. *Biochemistry* **1993**, *32*, 3047.
(10) McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **1999**, *121*, 11585.
(11) Wu, Y.; Ogawa, A. K.; Berger, M.; McMinn, D. L.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 7621.
(12) Mitsui, T.; Kitamura, A.; Kimoto, M.; To, T.; Sato, A.; Hirao, I.; Yokoyama, S. *J. Am. Chem. Soc.* **2003**, *125*, 5298.
(13) Switzer, C. Y.; Moroney, S. E.; Benner, S. A. *Biochemistry* **1993**, *32*, 10489.
(14) Tor, Y.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 4461.
(15) Ohtsuki, T.; Kimoto, M.; Ishikawa, M.; Mitsui, T.; Hirao, I.; Yokoyama, S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4922.
(16) 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.
(17) Kimoto, M.; Endo, M.; Mitsui, T.; Okuni, T.; Hirao, I.; Yokoyama, S. *Chem. Biol.* **2004**, *11*, 47.

 10.1021/ja0425280 CCC: $30.25

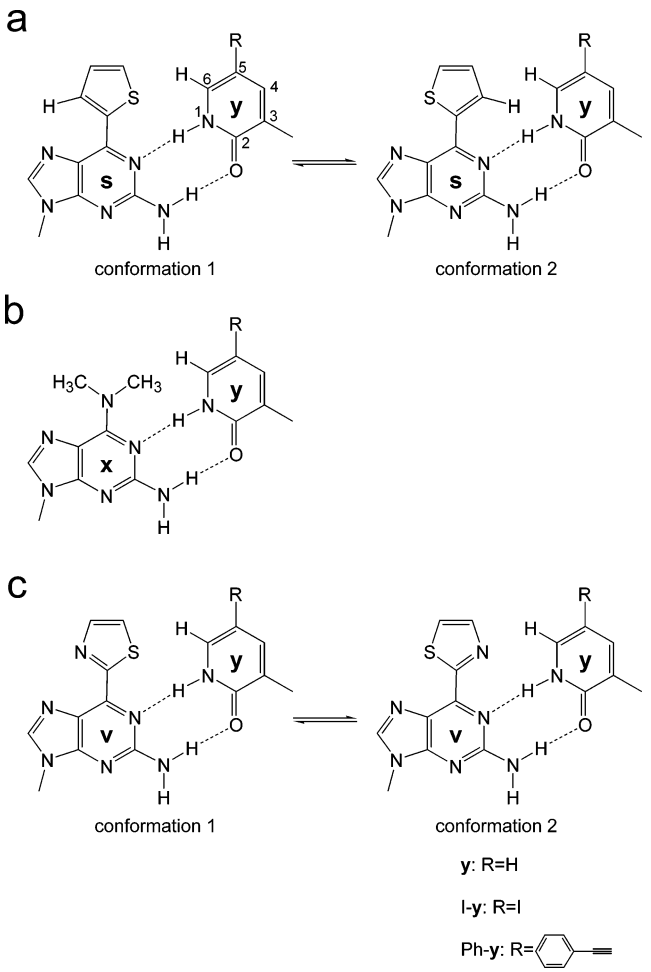


Figure 1. Unnatural base pair structures. Two conformations of the **s**−**y** pair (a), the **x**−**y** pair (b), and two conformations of the **v**−**y** pair (c). The substrates of 5-modified **y** bases can be synthesized chemically.

RNA molecules; I-**y** functions as a photocrosslinking component or a rational phasing tool for X-ray crystallography, and Ph-**y** stabilizes higher-ordered RNA structures. In addition, various functional substrates can also be derived chemically from the nucleoside of I-**y**. Thus, the method for the site-specific incorporation of **y** and modified **y** bases into RNA would be a powerful tool for the creation of novel RNA molecules with increased functionality.

Although the **s**−**y** pairing exhibits high selectivity with T7 RNA polymerase, the transcription efficiency involving one **s**−**y** pair is 50−60% lower than that of native transcription mediated by only the natural base pairs. In addition, this transcription cannot synthesize RNA molecules containing two adjacent **y** bases. Thus, to develop more efficient base-pair-expanded transcription systems, we have continued to improve the unnatural base pair. One of the problems of the **s**−**y** pair is the two possible conformations of **s**, caused by the free rotation of the thienyl group at position 6 of **s**, as shown in Figure 1a. Either the electronegative sulfur atom of conformation 1 or the C−H moiety of conformation 2 is located on the pairing surface with **y**, but both groups sterically prevent the pairing with the natural bases. The sulfur atom of conformation 1 may be more efficient for the specific pairing with **y** than the C−H moiety of conformation 2, because the function of the C−H moiety of **s** is similar to that of the methyl group in the 6-dimethylamino group of **x** (Figure 1b), and the efficiency and selectivity of the **s**−**y** pairing in transcription and replication are higher than those of the **x**−**y** pairing.[16,19] Thus, the replacement of the C−H moiety of **s** with other electronegative atoms could improve the incorporation efficiency of **y** into RNA by transcription.

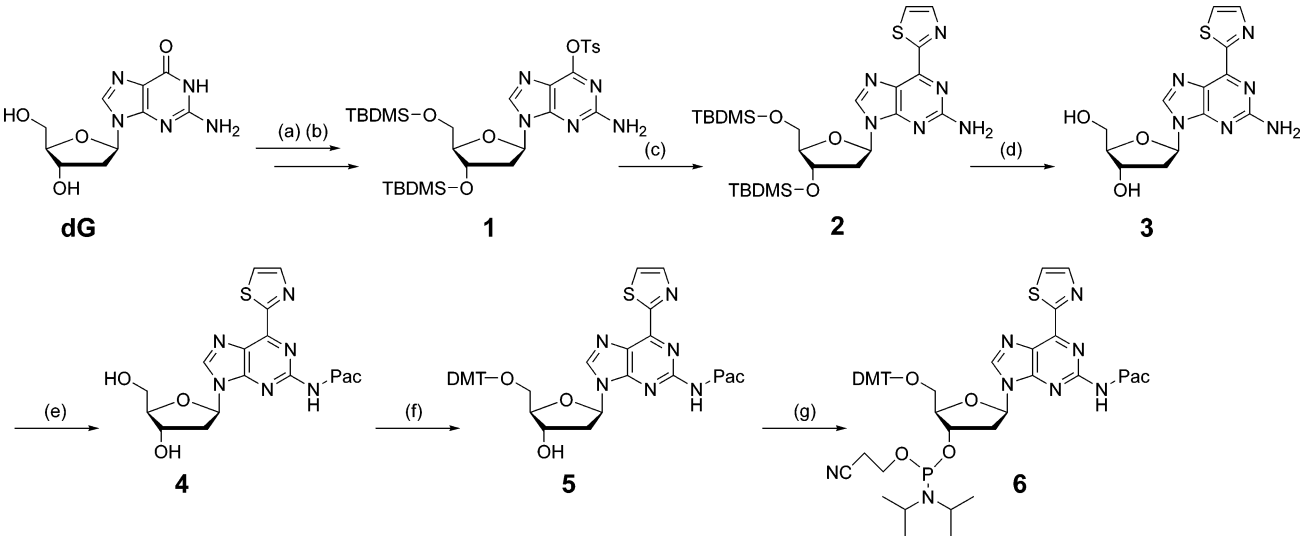
Here, we have designed an unnatural base, 2-amino-6-(2-thiazolyl)purine (denoted by **v**) (Figure 1c), instead of **s**, as a pairing partner of **y**. The two conformations of the **v** base, resulting from the free rotation of the 6-thiazolyl group, exhibit similar electronegative effects on the pairing partners, thus enhancing the transcription efficiency of the **v**−**y** pairing over that of the **s**−**y** pairing.

The 2′-deoxyribonucleoside of **v** (**3**) was synthesized by coupling the 2-amino-6-toluenesulfonylpurine nucleoside derivative (**1**) and 2-tributylstannylthiazole[20,21] (47% overall yield from 2′-deoxyguanosine (**dG**) in four steps, Scheme 1). It is noteworthy that the 2′-deoxyribonucleoside **3** exhibited a fluorescence emission centered at 457 nm, characterized by two major excitation maxima (295 and 364 nm), and its fluorescence quantum yield (excitation at 363 nm) was 0.41 in ethanol. The fluorescence emission of **v** was stronger than that of **s** (the quantum yield was 0.23).[22] For the amidite synthesis, the 2-amino group was protected by a phenoxyacetyl group, and the 5′- and N^2-protected nucleoside **5** was converted to the phosphoramidite **6**. The 2′-deoxyribonucleoside of **v** was slightly decomposed by treatments with concentrated ammonia for 6 h at 55 °C. Thus, for DNA synthesis, base-labile protection, using phenoxyacetyl for A, *p*-isopropylphenoxyacetyl for G, and acetyl groups for C, was used,[23] and deprotection of DNA fragments was carried out for 2−3 h at 55 °C with concentrated ammonia. The **v** nucleotide in DNA fragments was more stable than that expected from the stability of the 2′-deoxyribonucleoside, and no decomposition of the DNA fragments was observed under the deprotection conditions. The coupling efficiency of the amidite of **v** was more than 98% on a DNA synthesizer (Applied Biosystems, CA).

Before testing the base pairs in transcription, to assess the selectivity and efficiency of the **v**−**y** pairing in polymerase reactions, we determined the kinetic parameters of the single-nucleotide insertion of **y** and natural substrates into DNA opposite **v** or **s** in templates, using the exonuclease-deficient Klenow fragment of *Escherichia coli* DNA polymerase I. This is because the assessment of unnatural base pairings by the kinetic parameters in replication is much easier than that in transcription.[24,25] The single-nucleotide insertion experiments were carried out using the nucleoside triphosphate of **y** (d**y**TP) or the natural bases and a partially double-stranded template (35-mer, Table 1) containing **v** or **s**, with a primer labeled with 6-carboxyfluorescein (20-mer), in which various bases in the template were adjacent to the 3′ end of the primer. The insertion

(18) Endo, M.; Mitsui, T.; Okuni, T.; Kimoto, M.; Hirao, I.; Yokoyama, S. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2593.
(19) Fujiwara, T.; Kimoto, M.; Sugiyama, H.; Hirao, I.; Yokoyama, S. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2221.
(20) Peters, D.; Hörnfeldt, A.-B.; Gronowitz, S. *J. Heterocycl. Chem.* **1990**, *27*, 2165.
(21) Gutierrez, A. J.; Terborst, T. J.; Matteucci, M. D.; Froehler, B. C. *J. Am. Chem. Soc.* **1994**, *116*, 5540.
(22) Hirao, I.; Harada, Y.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Yokoyama, S. *J. Am. Chem. Soc.* **2004**, *126*, 13298.
(23) Schulhof, J. C.; Molko, D.; Teoule, R. *Tetrahedron Lett.* **1987**, *28*, 51.
(24) Vaught, J. D.; Dewey, T.; Eaton, B. E. *J. Am. Chem. Soc.* **2004**, *126*, 11231.
(25) Goodman, M. F.; Creighton, S.; Bloom, L. B.; Petruska, J. *Crit. Rev. Biochem. Mol. Biol.* **1993**, *28*, 83.

Scheme 1. Synthesis of the Nucleoside and the Phosphoramidite of **v**[a]



[a] Conditions: (a) *tert*-butyldimethylsilyl chloride, imidazole, DMF, 90%; (b) *p*-toluenesulfonyl chloride, Et_3N, 4-(dimethylamino)pyridine, CH_2Cl_2, 94%; (c) 2-tributylstannylthiazole, $Pd(PPh_3)_4$, LiCl, dioxane, 70%; (d), TBAF, THF, 80%; (e) (i) trimethylsilyl chloride, pyridine, (ii) 1-hydroxybenzotriazole, phenoxyacetyl chloride, pyridine, CH_3CN, 90%; (f) 4,4′-dimethoxytrityl chloride, pyridine, 99%; (g) $ClP(N{-}iPr_2)(OCH_2CH_2CN)$, iPr_2EtN, THF, 83%. Abbreviations: TBDMS, *tert*-butyldimethylsilyl; Ts, *p*-toluenesulfonyl; Pac, phenoxyacetyl; DMT, 4,4′-dimethoxytrityl.

Table 1. Steady-State Kinetic Parameters for Insertion of Single Nucleotides into a Template-Primer Duplex by the Exonuclease-Deficient Klenow Fragment[a]

primer 1	5′-ACTCACTATAGGGAGGAAGA
template 1	3′-TATTATGCTGAGTGATATCCCTCCTTCT***N***TCTCTT
template 2	3′-TATTATGCTGAGTGATATCCCTCCTTCT***N***TCTCGA

entry	template (**N**)	nucleoside triphosphate	K_M (μM)	V_{max} (% min^{-1})	efficiency (V_{max}/K_M)[d]
1	template 1 (**v**)	**y**	210 (140)[b]	40 (20)	1.9×10^5
2	template 1 (**v**)	T	280 (90)	1.9 (0.7)	6.8×10^3
3	template 1 (**v**)	C	370 (70)	21 (5)	5.7×10^4
4	template 1 (**v**)	A	89 (25)	0.37 (0.08)	4.2×10^3
5	template 1 (**v**)	G	n.d.[c]	n.d.[c]	
6	template 2 (**s**)	**y**	170 (40)	11 (2)	6.5×10^4
7	template 2 (**s**)	T	270 (40)	3.1 (0.2)	1.1×10^4
8	template 2 (**s**)	C	430 (180)	18 (9)	4.2×10^4
9	template 2 (**s**)	A	87 (27)	0.31 (0.07)	3.6×10^3
10	template 2 (**s**)	G	n.d.[c]	n.d.[c]	
11	template 1 (A)	T	1.2 (0.3)	3.5 (1.3)	2.9×10^6
12	template 1 (A)	C	1600 (600)	2.5 (1.3)	1.6×10^3

[a] Assays were carried out at 37 °C for 1−20 min using 5 μM template-primer duplex, 3−66 nM enzyme, and 0.6−3000 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] No inserted products were detected after an incubation for 20 min with 1500 μM nucleoside triphosphate and 50 nM enzyme. [d] The units of this term are % min^{-1} M^{-1}.

of each substrate opposite each base in the template was analyzed with an automated ABI 377 DNA sequencer with the *GeneScan* software[26] (Table 1).

The incorporation efficiency and selectivity of the **v**−**y** pairing were improved in comparison to those of the **s**−**y** pairing. The incorporation efficiency of d**y**TP opposite **v** in the template ($V_{max}/K_M = 1.9 \times 10^5$) was higher than that opposite **s** ($V_{max}/K_M = 6.5 \times 10^4$) (Table 1, entries 1 and 6). In contrast, the misincorporation efficiencies of the natural dNTPs opposite **v** were mostly similar to those opposite **s** (Table 1, entries 2−5 and 7−10). However, the efficiency of the T incorporation opposite **v** ($V_{max}/K_M = 6.8 \times 10^3$) (Table 1, entry 2) was slightly decreased, in comparison to that opposite **s** ($V_{max}/K_M = 1.1 \times 10^4$) (Table 1, entry 7). Although the efficiency of the C incorporation opposite **v** ($V_{max}/K_M = 5.7 \times 10^4$) (Table 1, entry 3) was slightly increased relative to that opposite **s** ($V_{max}/K_M = 4.2 \times 10^4$) (Table 1, entry 8), the selectivity of the cognate **v**−**y** pairing was 3.3-fold higher than that of the **v**−C pairing, and the selectivity of the **s**−**y** pairing was only 1.5-fold higher than that of the **s**−C pairing. Although the efficiency of the **y** incorporation opposite **v** was still 15-fold lower than that of the T incorporation opposite A (Table 1, entry 11), the **v**−**y** pair was more effective than the representative unnatural base pair between isoG and isoC;[5] the **y** incorporation opposite **v** ($V_{max}/K_M = 1.9 \times 10^5$) was 2.7-fold higher than the isoG incorporation opposite isoC ($V_{max}/K_M = 7.0 \times 10^4$).[12]

The different selectivity and efficiency between the **v**−**y** and **s**−**y** pairings suggest the existence of both conformations (1 and 2) of **v** and **s** in the single-stranded DNA templates. The improvement of the **v**−**y** pairing indicates that the nitrogen of conformation 2 of **v** is favorable for the C−H moiety of conformation 2 of **s**, for the selective pairing with **y**. To exclude the pairing with T, the electronegative nitrogen atom of the thiazolyl group is more effective than the C−H moiety of the thienyl group, and to sterically fit with **y**, the nitrogen of the thiazolyl group may be favorable for the C−H moiety. In addition, the relative abundances of conformations 1 and 2 of **v** might differ from those of **s**, and this difference would affect the selectivity and the efficiency of these base pairs in replication. If conformation 2 of **v** and conformation 1 of **s** are dominant in solution, then the efficiencies of the **y** incorporation opposite **v** ($V_{max}/K_M = 1.9 \times 10^5$) and **s** ($V_{max}/K_M = 6.5 \times 10^4$) would correlate with the van der Waals radii of the heteroatoms in the 6-groups of these bases opposite the H6 of **y**: the nitrogen (1.55 Å) in the thiazolyl group of **v** and the sulfur (1.80 Å) in the thienyl group of **s**. This correlation is supported by analyses of another unnatural base, 2-amino-6-(2-furanyl)purine (denoted by **o**);[27] conformation 1 of **o** has the smaller oxygen (1.52 Å) in the furanyl group opposite the H6

(26) Kimoto, M.; Yokoyama, S.; Hirao, I. *Biotechnol. Lett.* **2004**, *26*, 999.

(27) Hirao, I.; Fujiwara, T.; Kimoto, M.; Yokoyama, S. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4887.

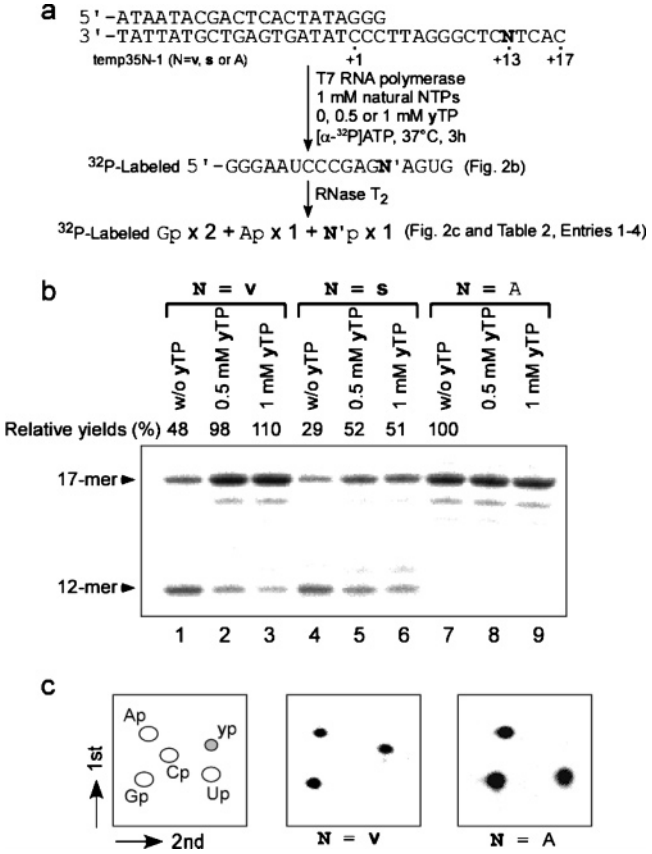


Figure 2. T7 transcription mediated by the **v**−**y** pairing. (a) Schemes of the experiments. (b) Gel electrophoresis of transcripts using the templates (N = **v**, **s**, or A) with the natural NTPs (1 mM) in the presence (0.5 or 1 mM) or absence of the substrate of **y** (**y**TP). Transcripts were internally labeled with [α-^{32}P]ATP. The relative yields of each transcript were determined by comparison to the yields of native transcripts from templates consisting of the natural bases, and each yield was averaged from three data sets. (c) 2D-TLC analysis of the labeled ribonucleoside 3′-phosphates obtained from the nuclease digestion of the transcripts (17-mer) (Figure 2b). The spots on the TLC were obtained from the 17-mer fragment transcribed in the presence of 1 mM **y**TP.

of **y**, and the efficiency of the **y** incorporation opposite **o** ($V_{max}/K_M = 2.4 \times 10^5$) is slightly higher than that opposite **v**. This suggests that the nitrogen of **v** is still somewhat large to fit well with the H6 of **y**, and thus, the efficiency of the **v**−**y** pair is also lower than that of the natural A−T pair. However, the smaller oxygen of **o** cannot effectively exclude the misincorporation of T ($V_{max}/K_M = 4.3 \times 10^4$).[27] Thus, the combination of the size and the electronegativity of the nitrogen in the thiazolyl group of **v** might be suitable for conferring high selectivity. Conformational analyses of the nucleosides of **v** and **s** are in progress.

Since the **v**−**y** pair exhibited higher selectivity and efficiency in replication than the **s**−**y** pair, we next examined the incorporation of **y**TP into RNA opposite **v** in templates by T7 RNA polymerase. Transcription was carried out by using DNA templates containing **v** or **s**, in which the unnatural base was located at a complementary site corresponding to position 13 in the transcripts (Figure 2a). The sequences of the DNA templates (temp35N-1, N = **v**, **s**, or A) were optimized to reduce the addition of one or more nontemplated nucleotides at the 3′-terminus of the nascent transcript.[28] After 3 h of transcription, the ^{32}P-labeled transcripts were analyzed on a gel (Figure 2b). The relative yields of the full-length transcripts (17-mer) containing one **y** base from the **s** template (N = **s**, temp35**s**-1) were ∼50% (Figure 2b, lanes 5 and 6). In contrast, the relative yields of the 17-mer transcripts from the **v** template (N = **v**, temp35**v**-1) (Figure 2b, lanes 2 and 3) were greatly increased (∼100%). Interestingly, even though abortive transcripts (12-mer) were observed for lanes 2 and 3 with the **v**−**y** pair, the relative yields of the full-length transcripts were as high as that obtained from the natural template (N = A, temp35A-1) with the natural NTPs (Figure 2b, lane 7).

Although the transcription of the **v** template (temp35**v**-1) without **y**TP also yielded the 17-mer transcript by the incorporation of the natural substrates, mainly CTP, opposite **v** (Figure 2b, lane 1; Table 2, entry 4), this misincorporation was completely excluded by the addition of **y**TP. The faithful **y** incorporation opposite **v** was confirmed by a nucleotide-composition analysis[15,16] of the 17-mer transcripts. For the nucleotide-composition analysis, transcripts labeled with [α-^{32}P]-ATP were fully digested to nucleoside 3′-phosphates. Afterward, the nucleotides were analyzed by 2D-TLC (Figure 2c), and the amount of each nucleoside 3′-phosphate was quantified (Table 2, entries 1−4). Since [α-^{32}P]ATP was used for labeling in transcription, the 3′-phosphates of the nucleotides that became the 5′-neighbor of A in the transcripts were labeled. Thus, the digestion of the transcripts by RNase T_2 generated one labeled Ap, two labeled Gp's, and a labeled nucleotide at position 13.

As shown in Figure 2c, among the pyrimidine substrates, only a spot corresponding to **y** was observed in the transcription of the **v** template (N = **v**), and thus no misincorporation of CTP and UTP opposite **v** occurred in the presence of **y**TP. In addition, in the transcription of the natural template (N = A) with **y**TP, no misincorporation of **y**TP opposite the natural bases was observed (Figure 2c, N = A). The quantification of each spot on the 2D-TLC also confirmed the high fidelity of the **v**−**y** pairing (Table 2, entries 1 and 2). Each nucleotide composition agreed well with the theoretical number.

The transcription of templates containing two adjacent **v** bases (temp35**v**-21) or two **v** bases separated by one T (temp35**v**-22) also yielded 17-mer transcripts containing two **y** bases (Figure 3a,b, lanes 2 and 3). The relative yields of the transcripts were 71% for those containing two adjacent **y** bases and 90% for those containing two **y** bases separated by one A. In contrast, the 17-mer transcripts containing two **y** bases were rarely observed in the transcription of **s** templates (Figure 3b, lanes 5 and 6). The high fidelity of this multiple **y** incorporation was also confirmed by the nucleoside-composition analysis (Figure 3c and Table 2, entries 5 and 6). In the transcription of temp35**v**-22, containing two **v** bases separated by one T, the transcripts were internally labeled with [α-^{32}P]ATP or [α-^{32}P]UTP, and thus, after RNase T_2 digestion, each nucleoside 3′-phosphate of **y** was detected on 2D-TLC; the **y** base at position 13 was labeled with [α-^{32}P]ATP, and the **y** base at position 15 was labeled with [α-^{32}P]UTP (Figure 3c). The quantification of these spots confirmed the site-specific insertion of both **y** bases (Table 2, entries 5 and 6). The multiple incorporations of **y** or modified **y** bases into RNA will be useful for the creation of RNA molecules with increased functionality and the expansion of the genetic code. For example, the multiple labeling of RNA with fluorescent-**y**TP would enhance the detection sensitivity, and

(28) Nacheva, G. A.; Berzal-Herranz, A. *Eur. J. Biochem.* **2003**, *270*, 1458.

Table 2. Nucleotide-Composition Analysis of T7 Transcripts

				composition of nucleotides incorporated as 5′ neighbor of A or U[a]				
entry	template	[α-^{32}P] NTP	**y**TP (mM)	Ap	Gp	Cp	Up	**y**p
1	temp35**v**-1	ATP	**y**TP (1)	1.01[b] [1][c] (0.01)[d]	1.97 [2] (0.06)	n.d.[e] [0] (−)	n.d. [0] (−)	1.02 [1] (0.07)
2	temp35A-1	ATP	**y**TP (1)	1.01 [1] (<0.01)	2.00 [2] (0.02)	n.d. [0] (−)	1.00 [1] (0.02)	n.d. [0] (−)
3	temp35A-1	ATP	**y**TP (0)	1.00 [1] (0.02)	1.99 [2] (0.03)	n.d. [0] (−)	1.01 [1] (0.01)	n.d. [0] (−)
4	temp35**v**-1	ATP	**y**TP (0)	1.08 [1] (0.02)	2.02 [2] (0.01)	0.79 [0] (0.01)	0.11 [0] (0.01)	n.d. [-] (−)
5	temp35**v**-22	ATP	**y**TP (1)	1.00 [1] (<0.01)	1.97 [2] (0.01)	0.01 [0] (<0.01)	0.01 [0] (<0.01)	1.02 [1] (0.01)
6	temp35**v**-22	UTP	**y**TP (1)	1.03 [1] (0.01)	n.d. [0] (−)	0.01 [0] (<0.01)	0.01 [0] (<0.01)	0.95 [1] (0.01)
7	temp35**v**-1	ATP	I-**y**TP (0.25)	1.02 [1] (0.03)	1.95 [2] (0.04)	0.01 [0] (<0.01)	0.01 [0] (0.01)	1.01 [1] (0.06)
8	temp35**v**-1	ATP	I-**y**TP (0.5)	1.00 [1] (0.01)	1.95 [2] (0.01)	0.01 [0] (<0.01)	0.01 [0] (<0.01)	1.03 [1] (0.01)
9	temp35**v**-1	ATP	I-**y**TP (1)	1.00 [1] (<0.01)	1.95 [2] (0.01)	0.01 [0] (<0.01)	0.01 [0] (0.01)	1.02 [1] (0.01)
10	temp35A-1	ATP	I-**y**TP (0.25)	1.00 [1] (0.03)	2.01 [2] (0.04)	0.01 [0] (0.01)	0.98 [1] (0.02)	0.01 [0] (<0.01)
11	temp35A-1	ATP	I-**y**TP (0.5)	1.01 [1] (0.02)	2.02 [2] (0.02)	0.01 [0] (<0.01)	0.95 [1] (0.01)	0.02 [0] (<0.01)
12	temp35A-1	ATP	I-**y**TP (1)	1.01 [1] (<0.01)	2.01 [2] (0.01)	0.01 [0] (<0.01)	0.93 [1] (0.01)	0.04 [0] (<0.01)
13	temp35**v**-1	ATP	Ph-**y**TP (0.5)	1.02 [1] (0.01)	1.98 [2] (0.03)	0.03 [0] (0.01)	0.01 [0] (0.01)	0.96 [1] (0.01)
14	temp35**v**-1	ATP	Ph-**y**TP (1)	1.02 [1] (0.01)	2.00 [2] (0.05)	0.03 [0] (0.01)	0.01 [0] (<0.01)	0.94 [1] (0.06)
15	temp35A-1	ATP	Ph-**y**TP (0.5)	1.00 [1] (0.03)	2.00 [2] (0.03)	0.01 [0] (<0.01)	0.99 [1] (0.05)	0.01 [0] (<0.01)
16	temp35A-1	ATP	Ph-**y**TP (1)	1.01 [1] (0.01)	2.01 [2] (0.01)	0.01 [0] (<0.01)	0.97 [1] (0.01)	0.01 [0] (<0.01)

[a] Composition of nucleotides incorporated as 5′ neighbor of A (entries 1−5 and 7−16) or U (entry 6), as shown in Figures 2, 3, and 4. [b] The values were determined using the following formula: (radioactivity of each nucleotide)/[total radioactivity of all nucleotides (3′-monophosphates)] × (total number of nucleotides at 5′ neighbor of [α-^{32}P]NTP). [c] The theoretical number of each nucleotide is shown in brackets. [d] Standard deviations are shown in parentheses. [e] Not detected.

an artificial codon containing two unnatural bases would increase the specificity of the codon−anticodon interaction, relative to codons containing one unnatural base.

The modified **y** substrates, I-**y**TP[17] and Ph-**y**TP,[18] were also incorporated into RNA opposite **v** with high efficiency (Figure 4). These transcription efficiencies mediated by the **v**−**y** pairing (Figure 4b, lanes 1 and 4) were more than 2-fold higher than those mediated by the **s**−**y** pairing (Figure 4b, lanes 2 and 5). The quantification of the nucleotide-composition analysis of the transcripts (Figure 4c and Table 2, entries 7−12) indicated that the incorporation efficiency of I-**y**TP was higher than that of **y**TP, and the slight misincorporation of I-**y**TP opposite A was observed in the transcripts by using 0.5 or 1 mM I-**y**TP. This tendency was shown in the T7 transcription using **s** as a template base and I-**y**TP as a substrate.[17] This may happen because, as shown with 5-substituted uracil derivatives,[29,30] the introduction of the electron-withdrawing iodine into **y** stabilizes the hydrogen bonding of the H−N1 position of **y** with the N1 positions of both **v** and A. Thus, for the site-specific incorporation of I-**y**TP into RNA, the combination of 0.25 mM I-**y**TP with 1 mM natural NTPs was sufficient (Table 2, entry 7). The site-specific incorporation of Ph-**y**TP was also confirmed by the nucleotide-composition analysis (Figure 4c and Table 2, entries 13−16). One of the advantages of I-**y** is that various functional substrates (5-modified **y**TPs, such as Ph-**y**TP) can be derived chemically from the I-**y** nucleoside. Actually, we succeeded in the syntheses and the site-specific incorporataion of other modified **y**TPs linked with biotin or fluorescent reagents (unpublished data by Hirao and Yokoyama), which will enable the site-specific immobilization and fluorescent labeling of RNA molecules.

In this study, we have described the specificity and efficiency of the unnatural **v**−**y** pair for a base-pair-expanded transcription system, for the site-specific incorporation of **y** or modified **y** bases into RNA. In T7 transcription, the **v**−**y** pair exhibits the highest efficiency among the unnatural base pairs developed thus far, such as the **s**−**y** and isoG−isoC pairs. The DNA templates containing the **v** bases can be prepared and amplified by PCR, using 3′-primers containing the **v** bases. Since this transcription system is quite efficient, it can be combined with other unnatural base pairs, such as isoG and isoC[13,14] or 2-amino-6-(2-thienyl)purine (**s**) and imidazolin-2-one (**z**).[22] The fluorescent **s** base can be site-specifically incorporated into RNA opposite **z** in templates. By using these unnatural base pairs, various different modified bases derived from **y**,[17,18] isoG,[14] and **s** could be simultaneously incorporated into a transcript. In addition, several different modified nucleotides of the natural bases, such as 2′-modified nucleotides,[31,32] modified bases,[24] and phosphorothioates,[33] can be applied to the specific transcription mediated by the unnatural base pairs by a T7 RNA

(29) Sanghvi, Y. S.; Hoke, G. D.; Freier, S. M.; Zounes, M. C.; Gonzalez, C.; Cummins, L.; Sasmor, H.; Cook, P. D. *Nucleic Acids Res.* **1993**, *21*, 3197.
(30) Kawahara, S.; Wada, T.; Kawauchi, S.; Uchimaru, T.; Sekine, M. *Nucleic Acids Symp. Ser.* **1998**, *39*, 123.
(31) Jellinek, D.; Green, L. S.; Bell, C.; Lynott, K.; Gill, N.; Vargeese, C.; Kirshenheuter, G.; McGee, D. P. C.; Abesinghe, P.; Pieken, W. A.; Shapiro, R.; Rifkin, D. B.; Moscatelli, D.; Janjiæ, N. *Biochemistry* **1995**, *34*, 11363.
(32) Davis, K. A.; Lin, Y.; Abrams, B.; Jayasena, S. D. *Nucleic Acids Res.* **1998**, *26*, 3915.

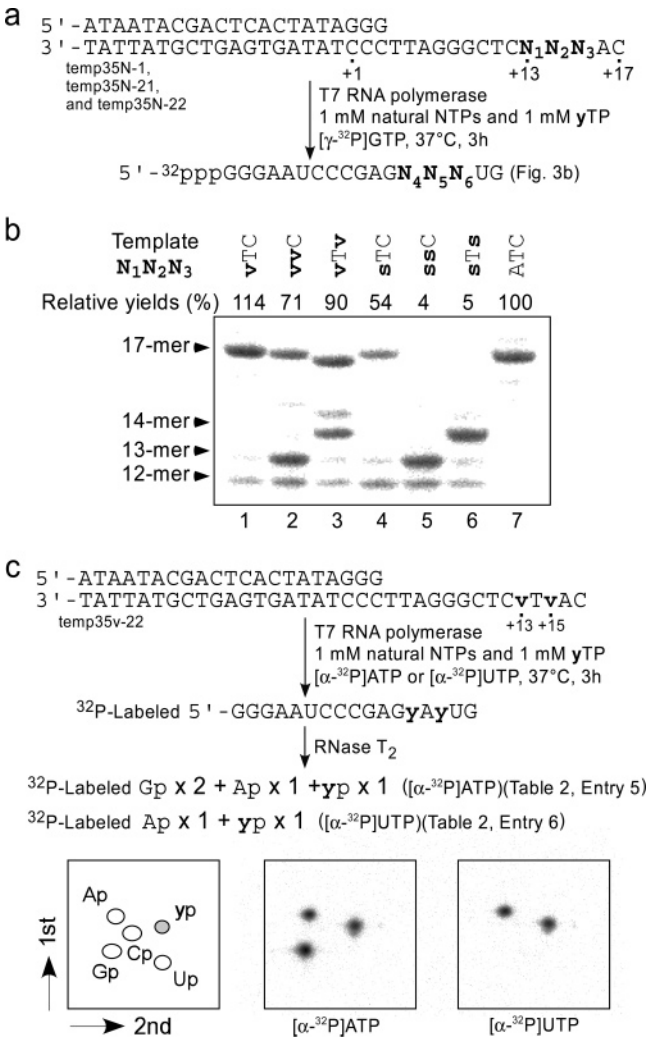


Figure 3. Transcription of the templates containing two **v** or **s** bases. (a) Schemes of the experiments. (b) Gel electrophoresis of transcripts using the templates containing two **v** or **s** bases with the natural NTPs (1 mM) and **y**TP (1 mM). Transcripts were labeled with [γ-^{32}P]GTP. (c) Nucleotide-composition analysis of the transcript containing two **y** bases. The transcripts were internally labeled with [α-^{32}P]ATP or [α-^{32}P]UTP. Each transcript was digested by RNase T_2, and the labeled nucleotides were detected by 2D-TLC.

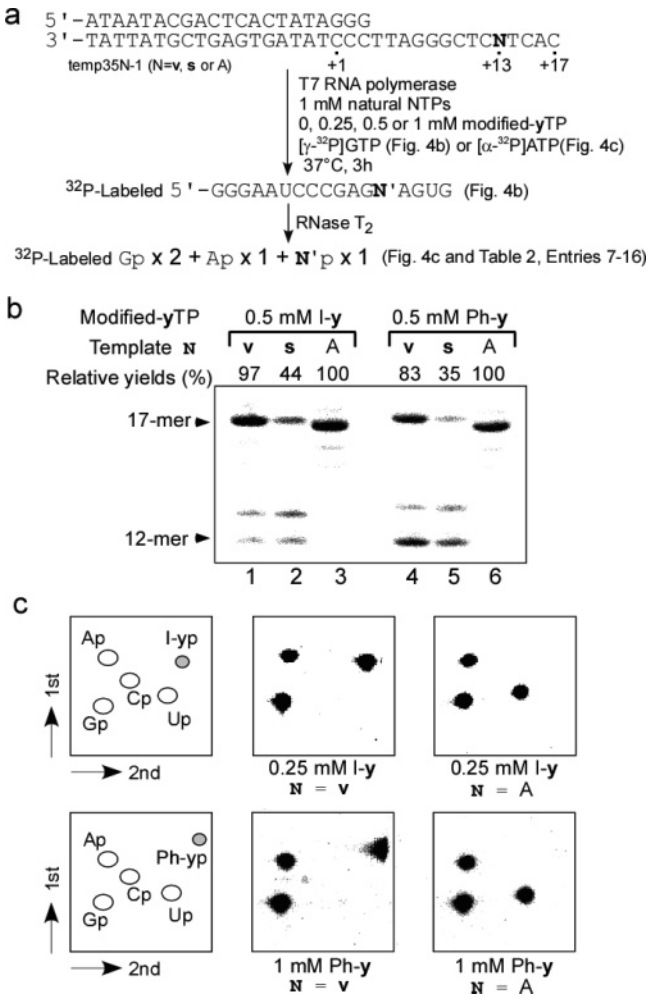


Figure 4. Transcription with modified **y** bases. (a) Schemes of the experiments. (b) Gel electrophoresis of transcripts containing modified **y**, I-**y**, and Ph-**y**, employing the **v**−**y** or **s**−**y** pairing. Transcription was carried out using 1 mM natural NTPs and 0.25, 0.5, or 1 mM I-**y**TP or Ph-**y**TP, and transcripts were labeled at the 5′-termini with [γ-^{32}P]GTP. (c) 2D-TLC analysis of the labeled ribonucleoside 3′-phosphates obtained from the nuclease digestion of the 17-mer RNA fragments transcribed with 0.25 mM I-**y**TP or 1 mM Ph-**y**TP.

polymerase variant.[34] These unnatural base and base-pair systems could overcome the serious limitation of the functionality of RNAs, as compared to that of proteins, and will facilitate advances in the creation of novel RNA molecules as diagnostic and therapeutic agents.

Experimental Section

General. Reagents and solvents were purchased from standard suppliers without further purification. Reactions were monitored by thin-layer chromatography (TLC) using 0.25-mm silica gel 60 plates impregnated with 254-nm fluorescent indicator (Merck). ^{1}H NMR (270 MHz), ^{13}C NMR (68 MHz), and ^{31}P NMR (109 MHz) spectra were recorded on a JEOL EX270 magnetic resonance spectrometer. Purification of nucleosides was performed on a Gilson HPLC system with a preparative C18 column (Waters Microbond Sphere, 150 × 19 mm). The triphosphate derivatives were purified with a DEAE-Sephadex A-25 column (300 × 15 mm) and a C18 column (Synchropak RPP, 250 × 4.6 mm, Eichrom Technologies). High-resolution mass spectra (HRMS) and electrospray ionization mass spectra (ESI-MS) were recorded on a JEOL HX-110 or JM 700 mass spectrometer and a Waters micromass ZMD 4000 equipped with a Waters 2690 LC system, respectively. Fluorescence measurements were made on a FP-6500DS spectrofluorometer (JASCO).

2-Tributylstannylthiazole.[20,21] To a solution of 2-bromothiazole (901 μL, 10 mmol) in ether (50 mL) was added *n*-butyllithium (1.58 M in hexane solution, 6.3 mL, 10 mmol) at −78 °C. The solution was stirred at −78 °C for 30 min. To the solution was added tributyltin chloride (2.7 mL, 10 mmol), and the reaction mixture was stirred at room temperature for 30 min. The mixture was poured into a saturated NaCl solution. The organic phase was washed twice with the saturated NaCl, dried with $MgSO_4$, and evaporated in vacuo. The resulting brown solution (2-tributylstannylthiazole, 10 mmol) was used for the Pd-mediated coupling reaction without further purification.

2-Amino-6-(2-thiazolyl)-9-[2-deoxy-3,5-di-*O*-(*tert*-butyldimethylsilyl)-β-D-ribofuranosyl]purine (2). A mixture of 2-amino-6-*O*-(*p*-toluenesulfonyl)-9-[2-deoxy-3,5-di-*O*-(*tert*-butyldimethylsilyl)-β-D-ribofuranosyl]purine (**1**)[35] (1.3 g, 2.0 mmol), $Pd(PPh_3)_4$ (116 mg, 0.1 mmol), 2-tributylstannylthiazole (3.7 g, 10 mmol), and LiCl (170 mg, 4.0 mmol) in dioxane (20 mL) was refluxed for 3.5 h. The reaction

(33) Jhaveri, S.; Olwin, B.; Ellington, A. D. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2285.
(34) Chelliserrykattil, J.; Ellington, A. D. *Nat. Biotechnol.* **2004**, *22*, 1155.

mixture was evaporated in vacuo, and the product was purified by silica gel column chromatography (1% methanol in CH_2Cl_2) to give **2** (790 mg, 70%).

2-Amino-6-(2-thiazolyl)-9-(2-deoxy-β-D-ribofuranosyl)purine (3). A solution of **2** (790 mg, 1.4 mmol) in tetrahydrofuran (THF) (10 mL) was treated with a solution of 1 M tetrabutylammonium fluoride (TBAF) in THF (4.2 mL, 4.2 mmol) at room temperature for 30 min. The solvent was evaporated in vacuo, and the product was purified by silica gel column chromatography (5% methanol in CH_2Cl_2) and then purified by RP-HPLC (10−50% CH_3CN in H_2O, 15 min) to give **3** (372 mg, 80%) as a yellow solid.

2-*N*-Phenoxyacetyl-6-(2-thiazolyl)-9-(2-deoxy-β-D-ribofuranosyl)-purine (4). A solution of **3** (167 mg, 0.5 mmol) and trimethylsilyl chloride (460 μL, 3.7 mmol) in pyridine (2.5 mL) was stirred at room temperature for 25 min (solution A). A solution of phenoxyacetyl chloride (104 μL, 0.8 mmol) and 1-hydroxybenzotriazole (120 mg, 0.9 mmol) in CH_3CN (240 μL) and pyridine (240 μL) was stirred at 0 °C for 5 min (solution B). Solution A, precooled to 0 °C, was added to solution B at 0 °C, and then the reaction mixture was stirred at room temperature overnight. The solution was cooled to 0 °C and treated with 14% NH_4OH (440 μL) for 10 min. The reaction mixture was separated with EtOAc and H_2O, and the organic phase was dried with Na_2SO_4 and evaporated in vacuo. The product was purified by silica gel column chromatography (5% methanol in CH_2Cl_2) to give **4** (211 mg, 90%) as a white solid.

2-*N*-Phenoxyacetyl-6-(2-thiazolyl)-9-[2-deoxy-5-*O*-(4,4′-dimethoxytrityl)-β-D-ribofuranosyl]purine (5). Compound **4** (211 mg, 0.45 mmol) was coevaporated with dry pyridine three times and was dissolved in pyridine (4.3 mL). To the solution was added 4,4′-dimethoxytrityl chloride (159 mg, 0.47 mmol), and the mixture was stirred at room temperature for 3 h. The reaction mixture was poured into 5% $NaHCO_3$ in H_2O and extracted with EtOAc. The organic phase was washed with saturated NaCl three times, dried with Na_2SO_4, and evaporated in vacuo. The product was purified by silica gel column chromatography (CH_2Cl_2/EtOAc = 1:1, v/v) to give **5** (346 mg, 99%).

2-*N*-Phenoxyacetyl-6-(2-thiazolyl)-9-[2-deoxy-5-*O*-(4,4′-dimethoxytrityl)-β-D-ribofuranosyl]purine 2-cyanoethyl-*N*,*N*-diisopropylphosphoramidite (6). Compound **5** (346 mg, 0.45 mmol) was coevaporated with pyridine and THF three times each and was dissolved in THF (2.3 mL). To the solution was added *N*,*N*-diisopropylethylamine (DIEA) (117 μL, 0.67 mmol) and 2-cyanoethyl-*N*,*N*-diisopropylamino chlorophosphoramidite (110 μL, 0.49 mmol). The mixture was stirred at room temperature for 1.5 h. The reaction was quenched by adding 50 μL of methanol. The solution was diluted with EtOAc/triethylamine (20 mL, 20:1, v/v) and then washed with 5% $NaHCO_3$ and saturated NaCl three times. The organic phase was dried with Na_2SO_4 and evaporated in vacuo. The product was purified by silica gel column chromatography (hexane/CH_2Cl_2 = 3:2, v/v containing 2% triethylamine) to give **6** (360 mg, 83%) as a white foam.

Steady-State Kinetics for the Single-Nucleotide Insertion Experiments with the Klenow Fragment. Steady-state kinetic analyses for single-nucleotide insertions were performed according to the literature.[26] A primer (primer 1, 20-mer) labeled with 6-carboxyfluorescein at the 5′-end was annealed with templates (template 1 or template 2, 35-mer) in a buffer containing 100 mM Tris-HCl (pH 7.5), 20 mM $MgCl_2$, 2 mM DTT, and 0.1 mg/mL bovine serum albumin, by heating at 95 °C and slow cooling to 4 °C. The primer-template duplex solution (10 μM, 5 μL) was mixed with 2 μL of a solution containing the exonuclease-deficient Klenow fragment (Amersham USB, Cleveland, OH), which was diluted in a buffer containing 50 mM potassium phosphate (pH 7.0), 1 mM DTT, and 50% glycerol. The mixture was incubated for more than 2 min. Reactions were initiated by adding 3 μL of each dNTP solution to the duplex-enzyme mixture at 37 °C. The amount of enzyme used (3−66 nM), the reaction time (1−20 min), and the gradient concentration of dNTP (0.6−3000 μM) were adjusted to give reaction extents of 25% or less. Reactions were quenched by adding 10 μL of a stop solution containing 95% formamide and 20 mM EDTA, and the mixtures were immediately heated at 75 °C for 3 min. The diluted products were analyzed on an automated ABI 377 DNA sequencer equipped with the *GeneScan* software (version 3.0). Relative velocities (v_0) were calculated as the extents of the reaction divided by the reaction time and were normalized to the enzyme concentration (20 nM) for the various enzyme concentrations used. The kinetic parameters (K_M and V_{max}) were obtained from Hanes−Woolf plots of [dNTP]/v_0 against [dNTP]. Each parameter was averaged from three to six data sets.

T7 Transcription. Templates (10 μM of a 35-mer coding DNA and a 21-mer noncoding DNA) were annealed in a buffer containing 10 mM Tris-HCl (pH 7.6) and 10 mM NaCl by heating at 95 °C and slow cooling to 4 °C. Transcription was carried out in a buffer (20 μL) containing 40 mM Tris-HCl (pH 8.0), 24 mM $MgCl_2$, 2 mM spermidine, 5 mM DTT, 0.01% Triton X-100, 1 mM natural NTPs, 0, 0.5 or 1 mM modified or unmodified **y**TP, 2 μCi [γ-^{32}P]GTP, 2 μM template, and 50 units of T7 RNA polymerase (Takara, Kyoto). By the use of [γ-^{32}P]-GTP, the transcripts were labeled only at the 5′-end, which facilitated the analyses of the yields. After an incubation at 37 °C for 3 h, the reaction was quenched by the addition of a dye solution (20 μL) containing 10 M urea and 0.05% BPB. The mixture was heated at 75 °C for 3 min, and the products were analyzed on a 20% polyacrylamide-7M urea gel.

Nucleotide-Composition Analysis in T7 Transcription.[15,16] Templates (10 μM) were annealed in 10 mM Tris-HCl buffer (pH 7.6) containing 10 mM NaCl by heating at 95 °C for 3 min and cooling to 4 °C. Transcription was carried out in 40 mM Tris-HCl buffer (pH 8.0, 20 μL) containing 24 mM $MgCl_2$, 2 mM spermidine, 5 mM DTT, 0.01% Triton X-100, 10 mM GMP, 1 mM natural NTPs, 0−1 mM **y**TP or modified **y**TP, 2 μCi [α-^{32}P]ATP or [α-^{32}P]UTP (Amersham), 2 μM template, and 50 units of T7 RNA polymerase (Takara). After an incubation for 3 h at 37 °C, the reaction was quenched by the addition of 20 μL of the dye solution. This mixture was heated at 75 °C for 3 min and then was loaded onto a 15% polyacrylamide-7 M urea gel. The full-length products were eluted from the gel with water and were precipitated with ethanol and 0.05 A_{260} units of *E. coli* tRNA. The transcripts were digested by 0.75 units of RNase T_2 at 37 °C for 70−120 min, in 10 μL of 15 mM sodium acetate buffer (pH 4.5). The digestion products were analyzed by 2D-TLC using a Merck HPTLC plate (100 × 100 mm) (Merck, Darmstadt, Germany) with the following developing solvents: isobutyric acid/NH_4OH/H_2O (66:1:33 v/v/v) for the first dimension, and isopropyl alcohol/HCl/H_2O (70:15:15 v/v/v) for the second dimension. The products on the gels and the TLC plates were analyzed by the Bio-imaging analyzer. The quantification of each spot was averaged from three to five data sets.

Acknowledgment. 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 a Grant-in-Aid for Scientific Research (KAKENHI 15350097) from the Ministry of Education, Culture, Sports, Science, and Technology. We are grateful to Dr. Makoto Komiyama, The University of Tokyo, for his support and helpful discussions.

Supporting Information Available: NMR and MS data for the nucleoside derivatives of **v** and MS data for the d(T**v**T) trimer. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0425280

(35) Nagatsugi, F.; Uemura, K.; Nakashima, S.; Maeda, M.; Sasaki, S. *Tetrahedron* **1997**, *53*, 3035.