Unnatural base pairs for specific transcription

Takashi Ohtsuki*, Michiko Kimoto†, Masahide Ishikawa‡, Tsuneo Mitsui‡, Ichiro Hirao‡§, and Shigeyuki Yokoyama*†‡§

*Genomic Sciences Center, and ‡Yokoyama CytoLogic Project, Exploratory Research for Advanced Technology (ERATO), Japan Science and Technology Corporation (JST), Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan; and †Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Edited by Leslie Orgel, The Salk Institute for Biological Studies, San Diego, CA, and approved February 22, 2001 (received for review November 7, 2000)

An unnatural base pair of 2-amino-6-(*N*,*N*-dimethylamino)purine (designated as x) and pyridin-2-one (designated as y) has been developed for specific transcription. The ribonucleoside triphosphates of y and a modified y, 5-methylpyridin-2-one, are selectively incorporated into RNA opposite x in the templates by T7 RNA polymerase. In addition, the sequences of the DNA templates containing x can be confirmed by a dideoxynucleotide chain-terminator method supplemented with the deoxynucleoside triphosphate of y. The bulky dimethylamino group of x in the templates effectively eliminates noncognate pairing with the natural bases. These results enable RNA biosynthesis for the specific incorporation of unnatural nucleotides at the desired positions.

Transcription involving specific, unnatural base pairs, in addition to the A·U and G·C pairs, would be a method for the site-directed incorporation of unnatural bases into RNAs, to achieve new functionality as ligands or catalysts (1–5) or novel codon–anticodon interactions between transfer and messenger RNAs in translation (6–9). Compared with the chemical synthesis of RNA, transcription is more efficient to produce and amplify RNAs with long chain lengths from DNA templates, which can be constructed by chemical synthesis in combination with enzymatic ligation. Thus, many efforts have been made to develop unnatural bases that are recognized by RNA polymerases as substrates and are specifically incorporated into RNA opposite the pairing bases in the DNA templates.

Previous efforts to create unnatural base pairs for replication and transcription have relied on nonstandard hydrogen-bonding schemes that differ from those of the Watson–Crick base pairs (2, 10–16). Unnatural pairs of nucleotides, such as isoguanosine·isocytidine and xanthosine·2,4-diaminopyrimidine nucleoside, are complementarily incorporated into DNA and RNA by polymerases with moderate selectivity (11–15). In addition, *in vitro* translation for the site-specific incorporation of an unnatural amino acid into a peptide has been demonstrated by using an extra codon–anticodon interaction between isoguanine and isocytosine (6, 7). These studies indicate that the unnatural, nonstandard hydrogen-bonded base pairs can also function in replication and transcription. However, some misincorporations resulting in noncognate pairings with natural bases cannot be ignored for the practical usage of these unnatural base pairs. For instance, a 14% misincorporation of adenosine into RNA opposite 2,4-diaminopyrimidine is observed at a xanthosine triphosphate-to-ATP ratio of 1:1 in transcription by T7 RNA polymerase (2). Isoguanine tautomerization also causes the misincorporation of T or U opposite this base (11, 12). In addition, the isocytidine and 5-methylisocytidine nucleoside derivatives are chemically unstable, and gradually decompose in solution (11, 16). Furthermore, nucleoside triphosphates of 2-aminopyrimidines, such as isocytosine and 2,4-diaminopyrimidine, are not recognized as substrates by some polymerases (11, 14, 15) because of the lack of the 2-keto group, the presence of which is generally important for the interactions of pyrimidine nucleosides with polymerases (17–20). Therefore, unnatural base pairs that overcome these shortcomings would be applicable to a wide range of nucleic acid biosynthesis *in vitro* and *in vivo*, as well as to the expansion of the genetic alphabet.

On the other hand, the new type of unnatural base pairs, in which pairing is mediated by hydrophobicity and complementarity of shape, has been more quantitatively tested for replication efficiency and fidelity (20–27). The hydrophobic base pairs, such as 4-methylbenzimidazole·difluorotoluene, can replace the natural hydrogen-bonded base pairs in replication and are enzymatically incorporated into DNA. Studies of non-hydrogen-bonded base pairs have shown the importance of the shape complementarity between pairing bases for replication fidelity. However, some polymerases do not recognize the difluorotoluene nucleosides as substrates and nucleosides in templates as well as the 2-aminopyrimidine nucleosides (20). In addition, the hydrophobic bases, such as difluorotoluene, 7-azaindole, and isocarbostyril, exhibit undesirable incorporation self-complementarity (20, 25). Thus, the hydrophobic base pairs are still insufficient for the specific and efficient site-directed incorporation of unnatural bases into DNA and RNA, and are undergoing further development (25–27).

On the basis of the concepts of the hydrogen-bonding patterns in combination with the shape complementarity, we designed an unnatural base pair of 2-amino-6-(*N*,*N*-dimethylamino)purine (**x**) and pyridin-2-one (**y**) (Fig. 1*a*; ref. 28). The putative interactions of the polymerases with the 2-keto group of **y** and the 3-nitrogen of **x** would facilitate the base pairing in nucleic acid biosynthesis (17–20). In addition, the bulky dimethylamino group of **x** would avoid noncognate pairing with the natural bases. The **x**·T (or **x**·U) and **x**·C pairings may have disadvantages, because the 6-dimethylamino group of **x** clashes with the 4-keto group of T (Fig. 1*b*) and the 4-amino group of C, but not with the corresponding 6-hydrogen of **y**. Although **y** has the potential to pair with A because of their complementarity of shape, the favorable A·T or A·U pairing is expected to competitively exclude the less efficient A·**y** pairing (29) in the enzymatic incorporation. In the present study, we tested the **x**·**y** pairing in transcription and DNA sequencing by using nucleoside triphosphates of **y** (**y**TP and d**y**TP, respectively) and DNA templates containing **x**.

Materials and Methods

Synthesis of Pyridin-2-one Ribonucleoside 5′-Triphosphate (yTP) and 5-Methylpyridin-2-one Ribonucleoside 5′-Triphosphate (m^5yTP). Pyridin-2-one and 5-methylpyridin-2-one ribonucleosides were synthesized according to the literature (30). For the triphosphate synthesis, both ribonucleosides (0.08–0.1 mmol) were dissolved in trimethyl phosphate (250 μl) and were cooled to 0°C. Phosphorus oxychloride (1.2 mol eq) was added into the solution, and the reaction mixture was stirred at 0°C for 6–14 h. Tributylamine (5.0 mol eq) was added, and then a solution of 0.5 M bis(tributylammonium) pyrophosphate (5.0 mol eq) in dimethylformamide was added. After 30 min, the reaction was quenched by the addition of 0.5 M of triethylammonium bicarbonate (500 μl).

This paper was submitted directly (Track II) to the PNAS office.

§To whom reprint requests may be addressed. E-mail: hirao@cytologic.riken.go.jp or yokoyama@y-sun.biochem.s.u-tokyo.ac.jp.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

www.pnas.org/cgi/doi/10.1073/pnas.091532698

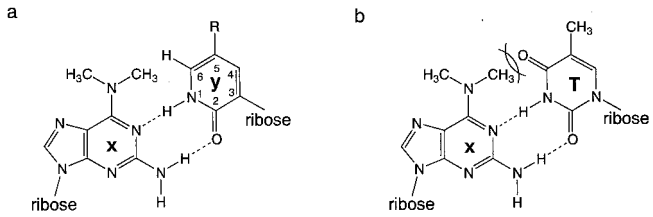

Fig. 1. The unnatural **x**·**y** (y: R = H, m^5y: R = CH_3) pair (*a*) in comparison with a noncognate **x**·**T** pair (*b*).

The ribonucleoside 5′-triphosphates were purified by using a polystyrene column (Amberchrom-CG161m, Supelco), a DEAE Sephadex A-25 column (Amersham Pharmacia), and, finally, a C18-reverse-phased (C18) HPLC column (SynChropak RP-P 1000; Micra Scientific, Suite Darien, IL). The triphosphates were characterized by ^{1}H NMR, ^{31}P NMR, and mass spectrometry. The molar absorption coefficients of **y**TP (7363 at λ_{max} = 300 nm) and m^5yTP (7208 at λ_{max} = 308.5 nm) were determined by quantitative analysis of the phosphorus (31) after dephosphorylation of the triphosphate with calf intestine alkaline phosphatase (Takara Shuzo, Kyoto).

T7 Transcription. Templates (10 μM) were annealed in 10 mM Tris·HCl buffer (pH 7.6) containing 10 mM NaCl by heating at 90°C for 2 min and cooling to 4°C. Transcription was carried out in 20 μl of 40 mM Tris·HCl buffer (pH 8.0) with 24 mM $MgCl_2$/2 mM spermidine/5 mM DTT/0.01% Triton X-100/10 mM GMP/1 mM each NTP/2 μCi of [α-^{32}P]ATP/2 μM template/50 units of T7 RNA polymerase (Takara Shuzo, Kyoto). The addition of GMP reduced the production of the full-length +1 products yielded by the random incorporation of an uncoded extra base, and facilitated the analysis. After an incubation for 3 h at 37°C, the reaction was quenched by the addition of 20 μl of dye solution (10 M urea/0.05% bromphenol blue). This mixture was heated at 75°C for 3 min, and then was loaded onto a 20% polyacrylamide/7 M urea gel. The transcription products were eluted from the gel with water, and were precipitated with ethanol. The transcripts were digested by 0.75 unit of RNase T_2 at 37°C for 14 h in 10 μl of 15 mM sodium acetate buffer (pH 4.5) containing 1.5% glycerol and 0.5 A_{260} unit of *Escherichia coli* tRNA. The digestion products were analyzed by two-dimensional TLC with a Funacel SF plate (100 × 100 mm; Funakoshi, Tokyo) with the following developing solvents: isobutyric acid/ammonia/water (66:1:33, vol/vol/vol) for the first dimension, and 2-propanol/HCl/water (70:15:15, vol/vol/vol) for the second dimension. The products on the gels and the TLC plates were analyzed with a BAS-2500 bio-imaging analyzer (Fuji).

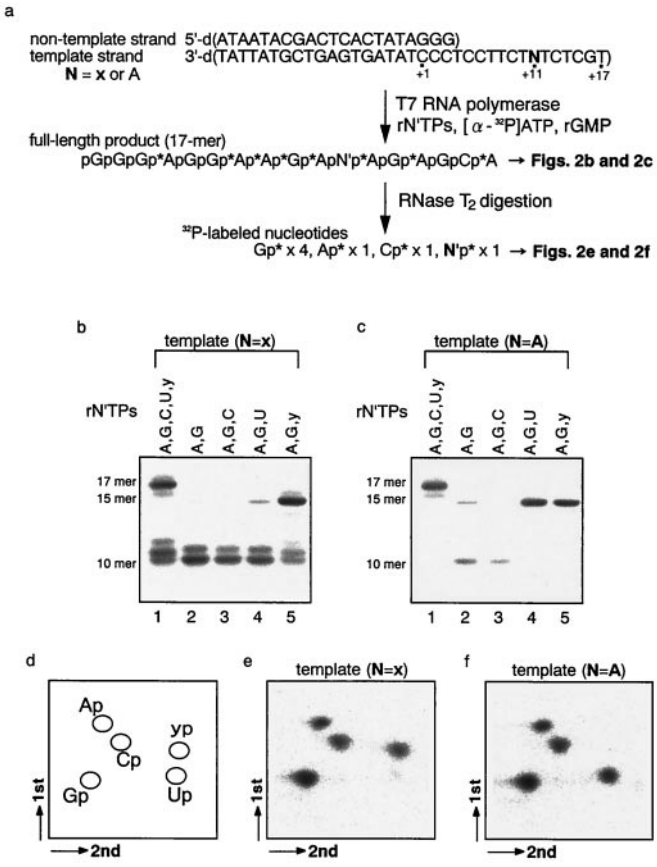

Fig. 2. Unnatural T7 transcription employing **x**·**y** pairing. (*a*) Schemes of the experiments. (*b* and *c*) Gel electrophoresis of transcripts with the templates (**N** = **x** or A) in the presence or absence of the natural and unnatural nucleoside triphosphates. (*d*–*f*) Two-dimensional TLC for nucleotide-composition analyses of transcripts isolated by the gel electrophoresis shown in *b* (lane 1) and *c* (lane 1). The quantitative data are shown in Table 1.

DNA Sequencing. Each DNA fragment (0.25 pmol) (109-mer: 3′-GCGGTCCCAAAAGGGTCAGTGCTGCTTAAGTATTAT-

Table 1. Nucleotide-composition analysis of T7 transcripts

The 11th base in template	rN′TP, mM		Composition of nucleotides incorporated as 5′ neighbor of A*				
	Natural (N′ = A, G, C, and U)	Unnatural	Gp	Ap	Cp	Up	Unnatural
A	1	1 (**y**)	3.94 (4)†	1.04 (1)	1.00 (1)	1.03 (1)	ND (0)
A	1	2 (**y**)	3.94	1.04	0.99	1.03	ND
A	2	2 (**y**)	4.04	1.04	0.89	1.01	ND
x	1	1 (**y**)	3.94 (4)	1.07 (1)	1.02 (1)	0.05 (0)	0.92 (1)
x	1	2 (**y**)	3.94	1.05	1.01	0.04	0.97
x	2	2 (**y**)	3.91	1.05	1.01	0.04	0.99
A	1	1 (m^5**y**)	4.04 (4)	1.05 (1)	0.91 (1)	0.99 (1)	ND (0)
A	2	2 (m^5**y**)	4.01	1.03	0.93	1.03	ND
x	1	1 (m^5**y**)	4.01 (4)	1.03 (1)	0.97 (1)	0.04 (0)	0.95 (1)
x	2	2 (m^5**y**)	3.98	1.07	0.98	0.03	0.93

*The values were determined via the following formula: (radioactivity of each nucleotide)/(total radioactivity of all nucleotides) × [(total numbers of nucleotides at 5′ neighbor of A) = 7].
†The theoretical number of each nucleotide is shown in parentheses.
ND, not detected (<0.01).

BIOCHEMISTRY

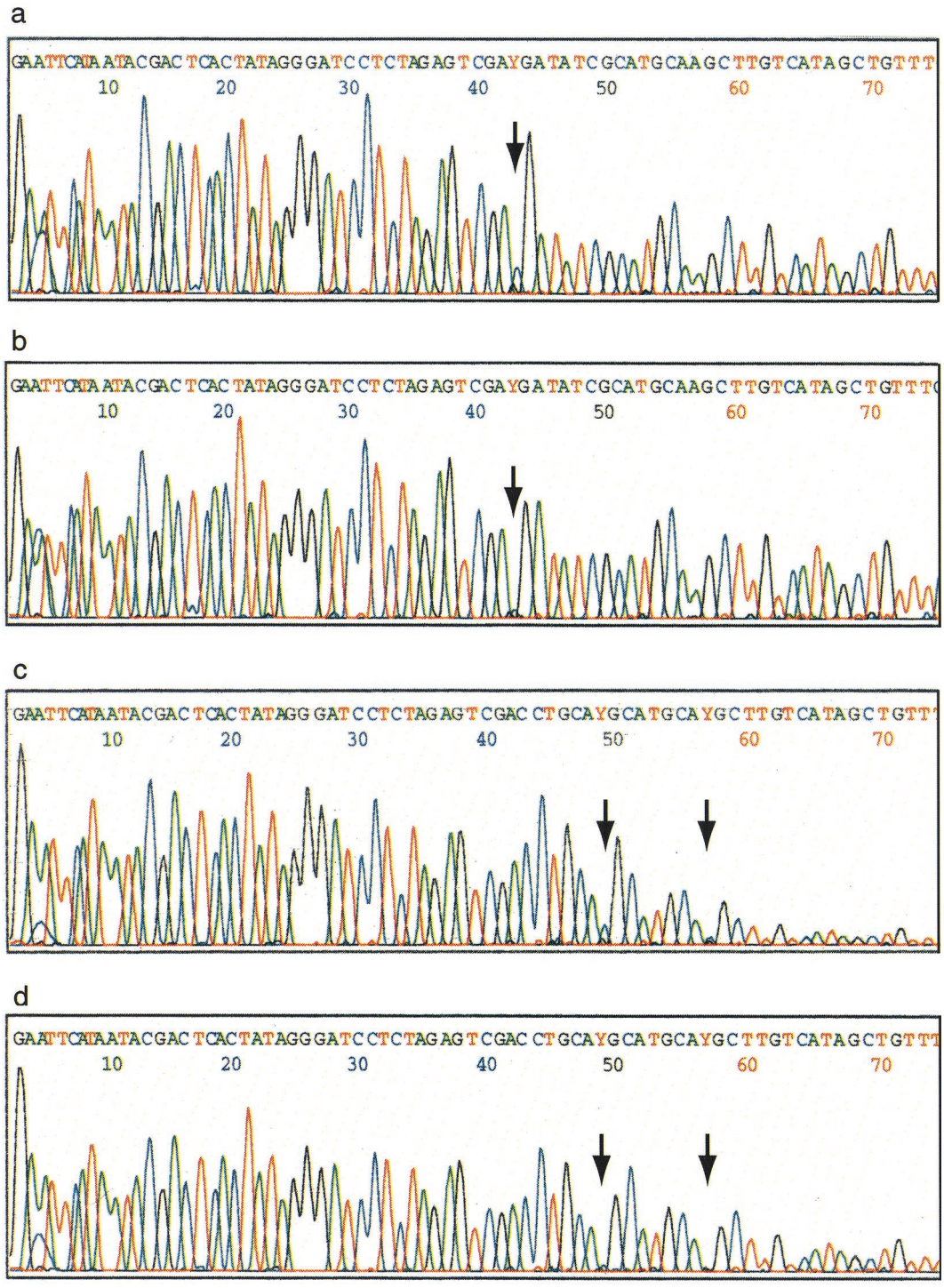

Fig. 3. Sequencing of DNA fragments containing **x**. Peak patterns of sequencing with fragment 1 containing one **x** (*a* and *b*) and fragment 2 containing two **x**s at predetermined positions (*c* and *d*), in the absence of (*a* and *c*) or in the presence of (*b* and *d*) 0.6 mM d**y**TP . The sequences described above the peak pattern show the product sequences. The arrows indicate the positions of **x** in the fragments.

GCTGAGTGATATCCCTAGGAGATCTCAGCT-$\mathbf{N}_{15}$-CGAACAGTATCGACAAAGGACACACTTT; fragment 1: $\mathbf{N}_{15}$ = **x**CTATAGCGTACGTT, and fragment 2: $\mathbf{N}_{15}$ = GGACGT**x**CGTACGT**x**) was mixed with a primer (4 pmol) (5′-CGCCAGGGTTTTCCCAGTCACGAC), d**y**TP (12 pmol), and 8 μl of a sequencing-kit solution (BigDye Terminator Cycle Sequencing FS Ready Reaction Kit; Perkin–Elmer) containing AmpliTaq DNA polymerase (Perkin–Elmer) in a total volume of 20 μl. After 25 cycles of PCR (96°C, 10 sec; 50°C, 5 sec; 60°C, 4 min), the residual dye terminators were removed from the reaction with a spin column (Perkin–Elmer), and the solution was dried *in vacuo* at 55°C. The residue was resuspended in 4 μl of a formamide solution and was heated at 95°C for 2 min. A portion (0.5–1 μl) of the solution was loaded on a DNA sequencer (model 377; Perkin–Elmer). The sequence data were analyzed with the Perkin–Elmer/Applied Biosystems PRISM sequencing analysis V3.2 software).

Results and Discussion

Transcription in the presence of various combinations of NTPs, including **y**TP, was first examined by using T7 RNA polymerase and partially double-stranded DNA templates (32). The template strands (35-mer) included the promoter sequence for T7 RNA polymerase followed by a short sequence consisting of C and T (except for **x** or A at position +11, and G at position +16; Fig. 2*a*). Transcription products were analyzed on a gel by internal labeling with [α-^{32}P]ATP. Transcription of the template (**N** = **x**) gave a full-length product (17-mer) in the presence of all of the natural triphosphates and **y**TP (Fig. 2*b*, lane 1). In contrast, transcription in the presence of only ATP, GTP, and **y**TP yielded a 15-mer product, indicating that **y**TP was incorporated opposite **x**, but not opposite G (Fig. 2*b*, lane 5). Incorporation opposite **x** was negligible for ATP, GTP, or CTP (Fig. 2*b*, lanes 2 and 3), but a slight presence of UTP was detected (Fig. 2*b*, lane 4). Transcription with the control template (**N** = A), consisting of only the natural bases, showed the remarkable misincorporation of **y**TP opposite A in the absence of UTP (Fig. 2*c*, lane 5). Nevertheless, it was still possible that the misincorporations were caused by the noncompetitive conditions.

To examine whether these misincorporations could be competitively eliminated by the cognate pairings, the nucleotide compositions of the 17-mer products (lane 1, Fig. 2 *b* and *c*) were analyzed. Because [α-^{32}P]ATP was used for internal labeling in transcription, the nucleotides that became the 5′-neighbor of A in the products were labeled at the 3′-phosphates (Fig. 2*a*). The products were then digested with RNase T_2, and the resulting labeled nucleotides, including the nucleotide opposite **x**, were analyzed by two-dimensional TLC (Fig. 2 *e* and *f*, and Table 1). In the transcription of the template (**N** = **x**), the selective incorporation of **y**TP opposite **x** was confirmed, and only a slight incorporation of UTP (<5%) opposite **x** was observed (Fig. 2*e* and Table 1). An increase in the amount of **y**TP tended to reduce the misincorporation of U opposite **x** (Table 1). Interestingly, in the transcription of the control template (**N** = A), no misincorporation of **y**TP was detected, indicating that the U·A pairing completely eliminated the **y**·A mispairing (Fig. 2*d*). The quantitated amount of each nucleotide on the two-dimensional TLC was extremely close to the theoretical values expected from the product sequence (Table 1). Thus, **y** can be site-specifically incorporated into RNA opposite **x** in T7 transcription.

This transcription system also enabled the synthesis of larger RNA fragments (61-mer) containing **y** at specific positions. The nucleotide-composition analysis of the products confirmed the highly selective incorporation of **y** into RNA (data not shown). In this case, the transcription with the partially double-stranded template paired only in the promoter region was not efficient, so transcription with the fully double-stranded templates containing the **y**·**x** pair between the nontemplate and template strands was performed. An efficiency of about 60% (relative to that with the control template consisting of the natural bases) was achieved, as judged by the relative yields of full-length RNA.

Furthermore, modified **y** bases, including functional groups at the 5-position, may also be incorporated into RNA. Actually, the ribonucleoside triphosphate of 5-methylpyridin-2-one [denoted m^5**y**] was incorporated into RNA opposite **x**, with a selectivity and efficiency similar to those of **y**TP (Table 1). The introduction of **y** and m^5**y** endows RNAs with a hydrophobic region on the major groove surface, which may affect the activity and the specificity of interactions with other molecules. In addition to the base modification, the **x**·**y** pairing might mediate the site-directed incorporations of other types of unnatural substrates, such as 2′-modified nucleotides and nucleoside 5′-thiophosphates.

We also showed that the locations of **x** in the templates can be confirmed by the application of the **x**·**y** pairing to DNA sequencing. The sequencing was carried out by using d**y**TP, templates containing **x** at one or two predetermined positions, and a commercially available dye-terminator sequencing reagent premixed with a mutated *Taq* DNA polymerase (BigDye Terminator Cycle Sequencing FS Ready Reaction Kit; Perkin–Elmer). The addition of 0.6 mM d**y**TP into the sequencing reaction prevented the misincorporation of the dye terminators of natural bases opposite **x**, resulting in the disappearance of the peaks only at the positions opposite **x** (Fig. 3 *b* and *d*). The sequencing without d**y**TP mainly gave peaks from the incorporation of the dideoxy-CTP dye terminator opposite **x** (Fig. 3 *a* and *c*). This simple sequencing method will be useful for future applications of the unnatural bases.

Interestingly, although the **x**·**y** pair is selectively and efficiently processed by this RNA polymerase, the same base pair as deoxynucleotides exhibits a lower selectivity in insertion into DNA by the DNA polymerase, the exonuclease-deficient Klenow fragment of *E. coli* DNA polymerase I (data not shown).

BIOCHEMISTRY

1. Benner, S. A., Burgstaller, P., Battersby, T. R. & Jurczyk, S. (1999) in *The RNA World*, eds. Gesteland, R. F., Cech, T. R. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 163–181.
2. Piccirilli, J. A., Krauch, T., Moroney, S. E. & Benner, S. A. (1990) *Nature (London)* **343,** 33–37.
3. Verma, S. & Eckstein, F. (1998) *Annu. Rev. Biochem.* **67,** 99–134.
4. Battersby, T. R., Ang, D. N., Burgstaller, P., Jurczyk, S. C., Bowser, M. T., Buchanan, D. D., Kennedy, R. T. & Benner, S. (1999) *J. Am. Chem. Soc.* **121,** 9781–9789.
5. Jensen, K. B., Atkinson, B. L., Willis, M. C., Koch, T. H. & Gold, L. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 12220–12224.
6. Benner, S. A. (1994) *Trends Biotechnol.* **12,** 158–163.
7. Bain, J. D., Switzer, C., Chamberlin, A. R. & Benner, S. A. (1992) *Nature (London)* **356,** 537–539.
8. Ogawa, A. K., Wu, Y., McMinn, D. L., Liu, J., Schultz, P. G. & Romesberg, F. E. (2000) *J. Am. Chem. Soc.* **122,** 3274–3287.
9. Liu, D. R. & Schultz, P. G. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 4780–4785.
10. Piccirilli, J. A., Moroney, S. E. & Benner, S. A. (1991) *Biochemistry* **30,** 10350–10356.
11. Switzer, C. Y., Moroney, S. E. & Benner, S. A. (1993) *Biochemistry* **32,** 10489–10496.
12. Switzer, C., Moroney, S. E. & Benner, S. A. (1989) *J. Am. Chem. Soc.* **111,** 8322–8323.
13. Tor, Y. & Dervan, P. B. (1993) *J. Am. Chem. Soc.* **115,** 4461–4467.
14. Lutz, M. J., Horlacher, J. & Benner, S. A. (1998) *Bioorg. Med. Chem. Lett.* **8,** 1149–1152.
15. Horlacher, J., Hottiger, M., Podust, V. N., Hübscher, U. & Benner, S. A. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 6329–6333.
16. Voegel, J. J. & Benner, S. A. (1996) *Helv. Chim. Acta* **76,** 1863–1880.
17. Eom, S. H., Wang, J. & Steitz, T. A. (1996) *Nature (London)* **382,** 278–281.
18. Kiefer, J. R., Mao, C., Braman, J. C. & Beese, L. S. (1998) *Nature (London)* **391,** 304–307.
19. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998) *Nature (London)* **391,** 251–258.
20. Morales, J. C. & Kool, E. T. (2000) *J. Am. Chem. Soc.* **122,** 1001–1007.
21. Kool, E. T., Morales, J. C. & Guckian, K. M. (2000) *Angew. Chem. Int. Ed. Engl.* **39,** 990–1009.
22. Matray, T. J. & Kool, E. T. (1999) *Nature (London)* **399,** 704–708.
23. Morales, J. C. & Kool, E. T. (1998) *Nat. Struct. Biol.* **5,** 950–954.
24. Moran, S., Ren, R. X.-F. & Kool, E. T. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 10506–10511.
25. Wu, Y., Ogawa, A. K., Berger, M., McMinn, D. L., Schultz, P. G. & Romesberg, F. E. (2000) *J. Am. Chem. Soc.* **122,** 7621–7632.
26. McMinn, D. L., Ogawa, A. K., Wu, Y., Liu, J., Schultz, P. G. & Romesberg, F. E. (1999) *J. Am. Chem. Soc.* **121,** 11585–11586.
27. Hirao, I., Ohtsuki, T., Mitsui, T. & Yokoyama, S. (2000) *J. Am. Chem. Soc.* **122,** 6118–6119.
28. Ishikawa, M., Hirao, I. & Yokoyama, S. (2000) *Tetrahedron Lett.* **41,** 3931–3934.
29. Solomon, M. S. & Hopkins, P. B. (1993) *J. Org. Chem.* **58,** 2232–2243.
30. Matulic-Adamic, J. & Beigelman, L. (1997) *Tetrahedron Lett.* **38,** 203–206.
31. Chen, P. S., Toribara, T. Y. & Warner, H. (1956) *Anal. Chem.* **28,** 1756–1758.
32. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. (1987) *Nucleic Acids Res.* **15,** 8783–8798.