Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 14 (2004) 2593–2596
Bioorganic & Medicinal Chemistry Letters

Unnatural base pairs mediate the site-specific incorporation of an unnatural hydrophobic component into RNA transcripts

Masayuki Endo,[a] Tsuneo Mitsui,[b] Taeko Okuni,[a] Michiko Kimoto,[a] Ichiro Hirao[a,b,*] and Shigeyuki Yokoyama[a,c,d,*]

[a]*Protein Synthesis Technology Team, Protein Research Group, RIKEN Genomic Sciences Center (GSC), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan*
[b]*Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, 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 27 November 2003; revised 19 February 2004; accepted 19 February 2004

Abstract—Site-specific incorporation of a hydrophobic nucleotide analog into RNA, by T7 transcription mediated by unnatural base pairs, was developed. The nucleotide analog, 5-phenylethynyl-3-(β-D-ribofuranosyl)pyridin-2-one 5-triphosphate (denoted by Ph-**y**TP), was chemically synthesized and then site-specifically incorporated by T7 RNA polymerase into RNA opposite the pairing partner, 2-amino-6-(2-thienyl)purine (denoted by **s**) in DNA templates. The introduction of Ph-**y** into a theophylline-binding RNA aptamer, in which a uridine in the internal loop was replaced by Ph-**y**, raised the thermal stability of the aptamer. Thus, this unnatural nucleotide analog would be useful for stabilizing RNA tertiary structures and complexes between RNA and other molecules.

Unnatural base pairs that can work with the natural A–T and G–C base pairs in transcription enable the site-specific incorporation of extra components into RNA, generating novel RNAs with increased functionality.[1,2] Recently, we developed unnatural base pairs between 2-amino-6-(*N*,*N*-dimethylamino)purine (**x**)[3] and (1*H*)pyridin-2-one (**y**), and 2-amino-6-(2-thienyl)purine (**s**) and **y**[4] (Fig. 1). The nucleoside 5′-triphosphate of **y** (**y**TP) was site-specifically incorporated into RNA opposite **s** or **x** in DNA templates by T7 RNA polymerase. In particular, the specificity of the **s**–**y** pairing was as high as those of the natural base pairings in transcription.[4] In addition, a modified **y**TP, a nucleoside 5′-triphosphate of 5-iodo-(1*H*)pyridine-2-one (I-**y**), was also specifically incorporated into RNA at desired positions.[5] The I-**y** components in RNA molecules have the potential to function as photoactivated crosslinking residues as well as heavy atoms to facilitate X-ray crystallography of large RNA molecules and RNA–protein complexes. In addition to these functions, a series of modified **y** derivatives can be chemically synthesized from the nucleoside of I-**y**.

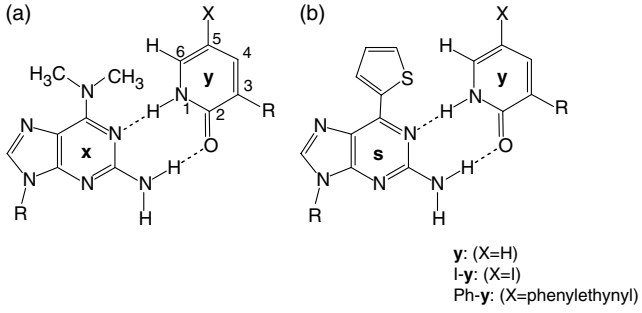


Figure 1. Unnatural **x**–**y** and **s**–**y** base pairs that function in transcription.

Here, we report the synthesis of a nucleoside 5′-triphosphate of a hydrophobic **y** analog, 5-phenylethynyl-3-(β-D-ribofuranosyl)pyridin-2-one 5-triphosphate

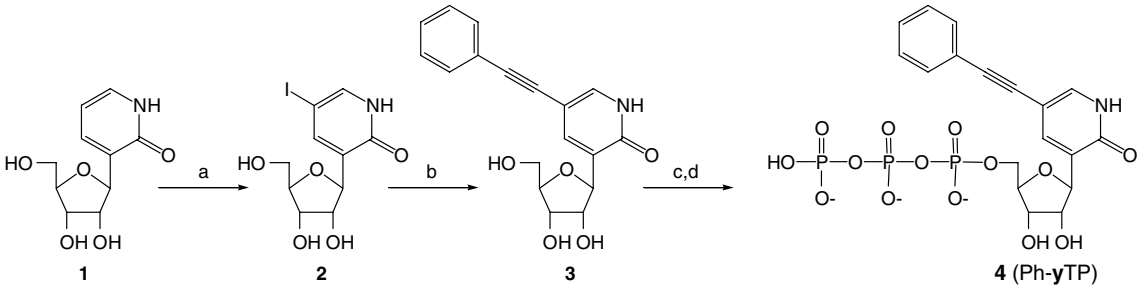
Keywords: Unnatural base pair.
* Corresponding authors. Tel.: +81-45-503-9644; fax: +81-45-503-9645 (I.H.); tel.: +81-3-5841-4413; fax: +81-3-5841-8057 (S.Y.); e-mail addresses: ihirao@postman.riken.jp; hirao@mkomi.rcast.u-tokyo.ac.jp; yokoyama@biochem.s.u-tokyo.ac.jp

0960-894X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2004.02.072

(Ph-**y**TP), and the specificity of the Ph-**y**TP incorporation into RNA fragments by T7 RNA polymerase. Furthermore, Ph-**y** was introduced into an RNA aptamer, in which a uridine residue in the single-stranded region was replaced by Ph-**y**, and the thermal stability of the modified RNA aptamer was examined. The phenyl and ethynyl residues in the **y** analog would be expected to stabilize RNA tertiary structures[6–8] and the interactions between RNA and other molecules, especially between RNA and proteins.[9]

The nucleoside derivatives of Ph-**y** (**3** and **4**) were synthesized, as shown in Scheme 1. The introduction of the phenylethynyl group into the nucleoside of **y** was carried out by coupling the ribonucleoside of I-**y** (**2**)[5] and phenylacetylene with $Pd(Ph_3P)_4$, CuI, and triethylamine,[10] with a 92% yield.[11] The nucleoside **3** was converted to the nucleoside 5′-triphosphate (**4**)[12] according to the literature.[13] The triphosphate was purified by DEAE-Sephadex column chromatography and C18 reversed-phase HPLC. The molar absorption coefficient of Ph-**y**TP (14,000 at 260 nm and 26,000 at λ_{max}: 286 nm) at pH 7.0 was determined by quantitative analysis of the phosphorus in the compound.[14]

The efficiency and specificity of the Ph-**y**TP incorporation into RNA were assessed using a short DNA template (35-mer) containing **s** at a specific position or a template (35-mer) consisting of natural bases as a control, from which 17-mer transcripts were obtained (Fig. 2a). The 35-mer template strands included the promoter sequence for T7 RNA polymerase, followed by a short sequence consisting of C and T, with **s** or A at position +11, and G at position +16. To facilitate the detection and quantification of the mis-incorporation of Ph-**y**TP opposite A and G, the template strands included only one A and/or G. These template strands were annealed with a 21-mer nontemplate strand, and the transcription was carried out by T7 RNA polymerase with 1 mM natural NTPs, 10 mM GMP, and 0.1 μCi/μl [α-^{32}P]ATP in the presence or absence of Ph-**y**TP (1 mM) at 37 °C for 3 h. The internally labeled transcripts obtained by T7 transcription in the presence or absence of Ph-**y**TP were analyzed by gel electrophoresis (Fig. 2b).



Scheme 1. (a) I_2, KI, Na_2CO_3, 100 °C, 4 h, (b) phenylacetylene, $Pd(Ph_3P)_4$, CuI, Et_3N, DMF, rt, 4–6 h, (c) $POCl_3$, 1,8-bis(dimethylamino)naphthalene, $(CH_3O)_3PO$, 0 °C, 2 h, (d) bis(tri-*n*-butylammonium)pyrophosphate, tri-*n*-butylamine, 0 °C, 10 min.

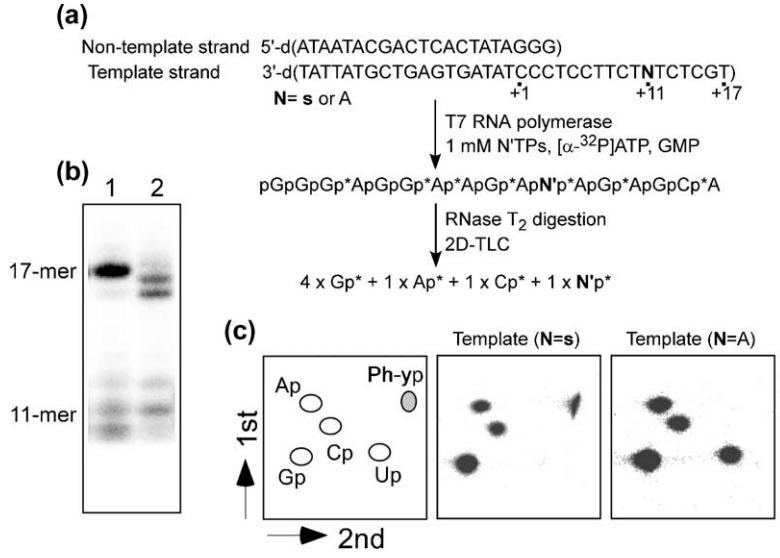


Figure 2. Site-specific incorporation of Ph-**y** into RNA by T7 RNA polymerase. (a) Schemes of the experiments. (b) Gel-electrophoresis of the transcripts. Lane 1: transcription using the template (**N**=**s**) in the presence of Ph-**y**TP, lane 2: transcription using the template (**N**=**s**) in the absence of Ph-**y**TP. (c) Two-dimensional TLC for nucleotide-composition analyses of transcripts using the templates (**N**=**s** or A) in the presence of Ph-**y**TP. The quantitative data are shown in Table 1.

Table 1. Nucleotide composition analysis of T7 transcripts[a]

Entry	Template	Ph-yTP (mM)	Composition of nucleotides incorporated as 5′ neighbor of A				
			Ap[b]	Gp	Cp	Up	Ph-yp
1	**N**=**s**	1	1.06[c] [1][d]	4.01 [4]	0.91 [1]	0.03 [0]	0.98 [1]
2	**N**=A	1	1.07 [1]	4.03 [4]	0.87 [1]	1.03 [1]	0.004 [0]
3	**N**=A	0	1.05 [1]	4.06 [4]	0.86 [1]	1.03 [1]	—[e] [0]

[a] Transcription was carried out using a solution (20 μM) containing 2 μM of each template, 1 mM each natural NTP, 2 μCi of [α-^{32}P]ATP, 10 mM GMP, and 50 units of T7 RNA polymerase at 37 °C for 3 h.

[b] Composition of nucleotides incorporated as 5′ neighbor of p*A, as shown in Figure 2a.

[c] The values were determined using the following formula: (radioactivity of each nucleotide)/(total radioactivity of all nucleotides) × (total number of nucleotides at 5′ neighbor of A).

[d] The theoretical number of each nucleotide is shown in brackets.

[e] Not analyzed.

The full length transcript (17-mer) was generated by the transcription in the presence of Ph-**y**TP (Fig. 2b, lane 1) using the template containing **s**. Although the transcription in the absence of Ph-**y**TP also gave products corresponding to a 17-mer and a 16-mer (Fig. 1b, lane 2), the transcription efficiency was lower than that in the presence of Ph-**y**TP. In addition, the 17-mer transcript consisting of natural ribonucleotides showed slightly higher mobility than that of the 17-mer transcript containing Ph-**y**. These results indicate that Ph-**y**TP was incorporated into RNA by the pairing between the substrate of Ph-**y** and **s** in the template.

To confirm the specificity of the Ph-**y**TP incorporation, we performed a nucleotide composition analysis[4] of the transcripts. In the transcripts, the nucleotides that became the 5′-neighbor of A were labeled at their 3′-phosphates, because [α-^{32}P]ATP was used for internal labeling in transcription (Fig. 2a). Then, the transcripts were digested with RNase T_2, and the resulting labeled nucleoside 3′-monophosphates, including the nucleotide opposite **s**, were analyzed by 2D thin-layer chromatography (2D-TLC) (Fig. 2c and Table 1). On the 2D-TLC, a spot corresponding to the nucleotide of Ph-**y** (Ph-**y**p) appeared only from the transcription using the template containing **s** (Fig. 2c, **N** = **s**). In contrast, the transcription using the control template containing A instead **s** gave a spot corresponding to Up, instead of Ph-**y**p (Fig. 2c, **N** = A). To detect the Ph-**y**TP misincorporation opposite A and G, we added one A and one G before T in the control template, and confirmed that no spot corresponding to Ph-**y** was detected by the analysis of the transcript from the control template (Fig. 2c, **N** = A). Quantification of the spots on the 2D-TLC also confirmed the site-specific incorporation of Ph-**y**TP opposite **s** (Table 1, entry 1); in the transcription in the presence of 1 mM Ph-**y**TP and natural NTPs, Ph-**y**p was incorporated into RNA with 98% specificity. Although Ph-**y**p showed a large mobility shift on the 2D-TLC because of its increased hydrophobicity, it was insignificant for the quantification. In the transcript obtained from the natural template in the presence of Ph-**y**TP, Ph-**y**p was scarcely detected (0.4%) (Table 1, entry 2). In this assay, the composition of Cp was relatively small in comparison to the theoretical number. However, this did not result from the **s**–**y** pairing in transcription, because this tendency also appeared in the natural transcription, using the control template in the absence of Ph-**y**TP (Table 1, entry 3).

Incorporations of nucleotide analogs containing a hydrophobic propynyl or phenyl group into nucleic acids increase the thermal stabilities of their duplex structures.[6–8] In RNA–protein complexes, stacking interactions between the aromatic groups on the β-sheets of proteins and the base moieties of RNAs are a common structural feature.[9] Thus, the introduction of Ph-**y** to appropriate sites in RNAs would strengthen the stacking interactions and stabilize the tertiary structures of RNA and RNA–other molecule complexes.

To assess the utility of the specific Ph-**y** incorporation into RNA, we introduced Ph-**y** or **y** into a theophylline-binding RNA aptamer,[15–17] in place of the uridine at position 24 in an internal loop in the aptamer, and compared the thermal stability of these aptamers with the original aptamer. The sequence of the original aptamer is shown in Figure 3, and the uridine at position 24 was replaced by Ph-**y** or **y**. Transcription was carried out by T7 RNA polymerase, with 1 mM natural NTPs, Ph-**y**TP (1 mM) or **y**TP (1 mM), and a double-stranded DNA template containing **s**, at 37 °C for 6 h. The thermal stability of each RNA aptamer (1.3–1.6 μM) was examined in 10 mM sodium phosphate (pH 7.0), 50 mM NaCl, and 0.1 mM EDTA using a SHIMADZU spectrophotometer, UV-2450, at a heating rate of 0.5 °C/min. The stability of the aptamer containing Ph-**y** at position 24 (Ph-**y**24: Tm = 67.0 °C) was greatly improved in comparison to that of the original aptamer (U24: Tm = 63.3 °C). In contrast, the stability of the aptamer containing **y** at position 24 (**y**24: Tm = 63.3 °C) was as low as that of the original aptamer. These results show the importance of the phenylethynyl group of **y** for the stabilization of the RNA aptamer. The binding efficiency of this Ph-**y** modified aptamer with theophilline is being investigated.

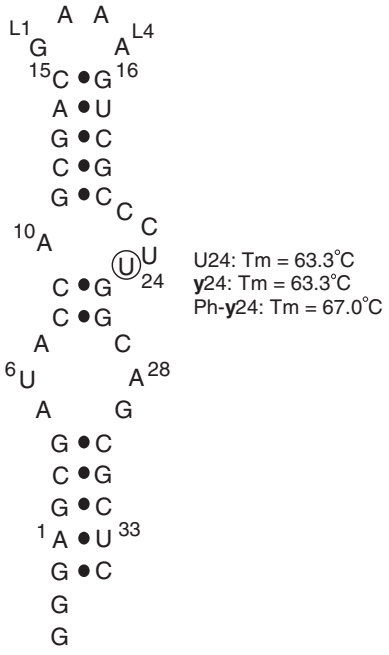


Figure 3. Site-specific incorporation of Ph-**y** into a theophylline-binding RNA aptamer. Unnatural components, **y** and Ph-**y**, were introduced at position 24, in place of the uridine of the original aptamer (U24). The numbering of the residues corresponds to the original duplex aptamer.[16,17]

By the specific transcription involving the unnatural **s**–**y** pair, the hydrophobic Ph-**y** can be incorporated into large RNA fragments at desired positions. The DNA templates containing **s** are easily prepared by PCR amplification with a 3′-primer containing **s**.[5] Thus, this specific transcription mediated by unnatural base pairs is more useful than chemical RNA synthesis especially for the preparation of long RNA fragments. Efficient sites in RNA molecules for the incorporation of unnatural components can be predicted from the 3D structures of the RNA molecules. Uridine residues in single-stranded regions, G–U pairs, and certain base triplets[18] in the RNA molecules might be suitable Ph-**y** incorporation sites for the thermal stabilization of RNA molecules.

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. Benner, S. A.; Burgstaller, P.; Battersby, T. R.; Jurczyk, S. In *The RNA World*; Gesteland, R. F., Cech, T. R., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory: Cold Spring Harbor, New York, 1999; pp 163–181.
2. Tor, Y.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 4461.
3. Ohtsuki, T.; Kimoto, M.; Ishikawa, M.; Mitsui, T.; Hirao, I.; Yokoyama, S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4992.
4. 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.
5. Kimoto, M.; Endo, M.; Mitsui, T.; Okuni, T.; Hirao, I.; Yokoyama, S. *Chem. Biol.* **2003**, *11*, 47.
6. McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **1999**, *121*, 11585.
7. Mitsui, T.; Kimoto, M.; Sato, A.; Yokoyama, S.; Hirao, I. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4515.
8. Nakano, S.; Uotani, Y.; Nakashima, S.; Anno, Y.; Fujii, M.; Sugimoto, N. *J. Am. Chem. Soc.* **2003**, *125*, 8086.
9. Mattaj, I. W.; Nagai, K. *Nat. Struct. Biol.* **1995**, *2*, 518.
10. Brown, L. J.; May, J. P.; Brown, T. *Tetrahedron Lett.* **2001**, *42*, 2587.
11. 5-Phenylethynyl-3-(β-D-ribofuranosyl)pyridin-2-one (**3**): ^{1}H NMR (270 MHz, DMSO-d_6) δ 3.50 (m, 1H), 3.65 (m, 1H), 3.83 (m, 3H), 4.68 (d, 1H, $J = 4.1$ Hz), 4.73 (d, 1H, $J = 5.1$ Hz), 4.96 (t, 1H, $J = 6.1$ Hz), 5.11 (br s, 1H), 7.39 (m, 3H), 7.48 (m, 2H), 7.70 (s, 1H), 7.72 (s, 1H), 12.11 (s, 1H); ^{13}C NMR (68 MHz, DMSO-d_6) δ 61.17, 70.57, 74.57, 80.37, 83.47, 86.10, 88.71, 100.17, 122.32, 128.30, 128.54, 130.89, 131.16, 137.68, 138.49, 160.24. HR-MS (FAB, 3-NBA matrix) calcd for $C_{18}H_{18}NO_5$ (M+1) 328.1185, found: 328.1188.
12. 5-Phenylethynyl-3-(β-D-ribofuranosyl)pyridin-2-one 5-triphosphate (**4**): ^{1}H NMR (270 MHz, D_2O) δ 1.11 (t, 27H, $J = 7.3$ Hz), 3.03 (q, 18H, $J = 7.3$ Hz), 4.09 (m, 5H), 4.86 (d, 1H, $J = 3.7$ Hz), 7.27 (m, 3H), 7.43 (m, 2H), 7.60 (s, 1H), 7.81 (s, 1H). ^{31}P NMR (109 MHz, D_2O) δ −22.83 (t, 1H, $J = 19.5$ Hz, 21.4 Hz), −10.72 (d, 1H, $J = 19.5$ Hz), −10.15 (d, 1H, $J = 21.4$ Hz). ESI-MS calcd for $C_{18}H_{19}NO_{14}P_3$ (M−1) 566.00, found: 565.50.
13. Kovács, T.; Ötvös, L. *Tetrahedron Lett.* **1988**, *29*, 4525.
14. Mitsui, T.; Kitamura, A.; Kimoto, M.; To, T.; Sato, A.; Hirao, I.; Yokoyama, S. *J. Am. Chem. Soc.* **2003**, *125*, 5298.
15. Jenison, R. D.; Gill, S. C.; Pardi, A.; Polinsky, B. *Science* **1994**, *263*, 1425.
16. Zimmermann, G. R.; Shields, T. P.; Jenison, R. D.; Wick, C. L.; Pardi, A. *Biochemistry* **1998**, *37*, 9186.
17. Zimmermann, G. R.; Wick, C. L.; Shields, T. P.; Jenison, R. D.; Pardi, A. *RNA* **2000**, *6*, 659.
18. Walberer, B. J.; Cheng, A. C.; Frankel, A. D. *J. Mol. Biol.* **2003**, *327*, 767.