Expanding the Genetic Code

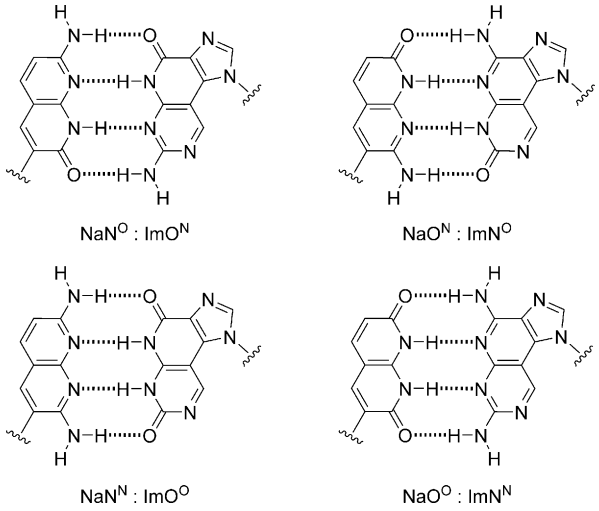
DOI: 10.1002/anie.201406402

Selective Transcription of an Unnatural Naphthyridine:Imidazopyridopyrimidine Base Pair Containing Four Hydrogen Bonds with T7 RNA Polymerase**

*Yusaku Nomura, Satoshi Kashiwagi, Kousuke Sato, and Akira Matsuda**

Abstract: *The naphthyridine:imidazopyridopyrimidine base pair is the first base pair containing four hydrogen bonds that can be replicated selectively and efficiently by the use of DNA polymerases. Herein we describe the synthesis of naphthyridine-C-ribonucleoside 5′-triphosphate (rNaTP) and transcription reactions catalyzed by T7 RNA polymerase with rNaTP and template DNA containing imidazopyridopyrimidine. The transcription reaction was also applied to a longer transcript containing part of the human c-Ha-Ras gene.*

The complementarity of natural base pairs consisting of adenine (A):thymine (T) or uracil (U)/guanine (G):cytosine (C) is an extremely important factor for biological activity in all living organisms according to the "central dogma of molecular biology". The creation of new base pairs—other than the original Watson–Crick base pairs—that are able not only to replicate in DNA and but also to transcribe to RNA selectively would expand the genetic alphabet for the site-specific incorporation of unnatural constituents into DNA, RNA, and proteins with specific functionalities.[1] Since the pioneering studies of Benner and co-workers,[2] the development of new base pairs has been intensely investigated with a view to expanding the genetic alphabet and exploring synthetic biology,[3] especially by Kool and co-workers,[4] Romesberg and co-workers,[5] and Hirao and co-workers.[6] These investigations have suggested that shape complementarity of the purine:pyrimidine base pairs and stacking interactions between the nucleobases as well as hydrogen bonds are critical for selective recognition by DNA and RNA polymerases. We have developed naphthyridine (Na):imidazopyridopyrimidine (Im) base pairs containing four hydrogen bonds.[7] Furthermore, we have demonstrated that DNA duplexes with NaN^{O}:ImO^{N}, NaO^{N}:ImN^{O}, NaN^{N}:ImO^{O}, and NaO^{O}:ImN^{N} pairs (Scheme 1) show high thermal stability resulting from 1) the four noncanonical hydrogen bonds, 2) their high stacking ability, and 3) the shape complementarity of the Na:Im pairs, such as pyrimidine:purine base pairs.



Scheme 1. Structures of naphthyridine:imidazopyridopyrimidine base pairs.

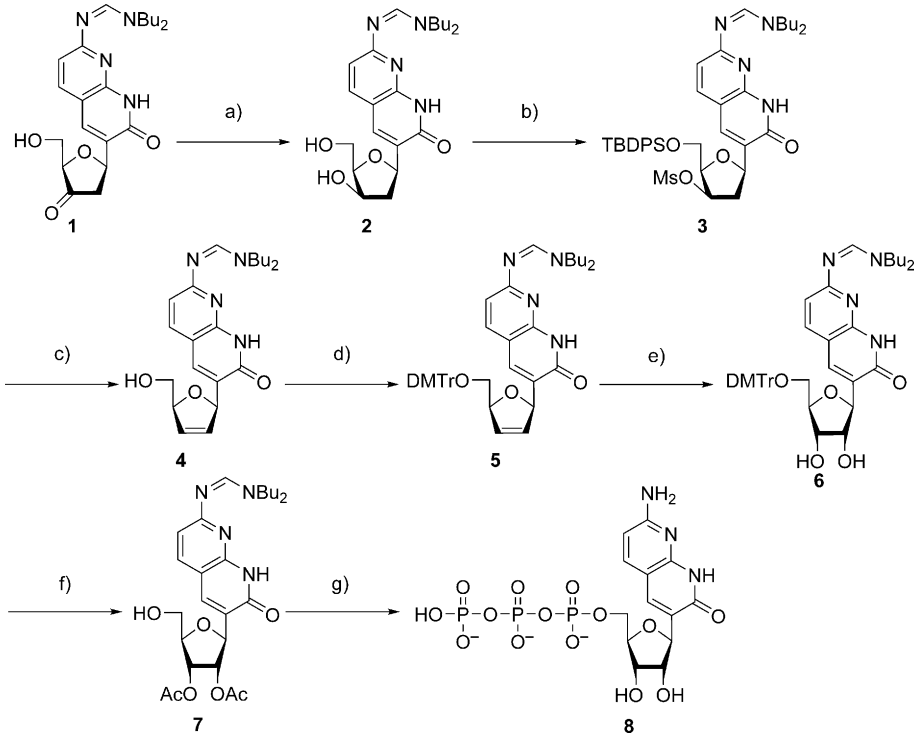
These results prompted us to investigate whether these thermally stable base pairs function according to the central dogma. We previously reported the replication of the Na:Im base pairs by use of the Klenow fragment (KF), which selectively recognizes the NaO^{N}:ImN^{O} pair.[8a] Moreover, Deep Vent DNA polymerase (exo-) recognizes the NaO^{N}:ImN^{O} pair and elongates to give the 30 mer full-length product after its incorporation.[8b] Since the "central dogma" requires that unnatural base pairs should not only replicate in DNA, but also transcribe to RNA with accurate selectivity and efficiency, we studied the transcription of Na:Im base pairs by using T7 RNA polymerase, which would be useful for in vitro transcription.

First, we synthesized the Na-*C*-ribonucleoside $rNaN^{O}$ from the corresponding 2′-deoxyribonucleoside ($dNaN^{O}$).[9] The 3′-keto derivative **1**[7b] of $dNaN^{O}$ was synthesized by a Heck reaction between the 3′-*O*-*tert*-butyldimethylsilyl-protected glycal and the protected 2-amino-7-hydroxy-6-iodo-1,8-naphthyridine and converted into the 3′β-alcohol **2** by stereoselective reduction with K-selectride. The primary alcohol moiety of **2** was protected with a *tert*-butyldiphenylsilyl (TBDPS) group, followed by mesylation of the 3′β-hydroxy group to give **3**. The treatment of **3** with KO*t*Bu in DMF afforded **4**,[10] which was then protected with a 4,4′-dimethoxytrityl (DMTr) group to afford **5**. The desired *C*-ribonucleoside **6** was obtained diastereoselectively (93:7) through oxidation with OsO_4 at −40 °C. After acetylation of the secondary alcohol moieties, followed by removal of the DMTr group, the desired diastereomer **7** was isolated by

[*] Y. Nomura, S. Kashiwagi, Dr. K. Sato, Prof. Dr. A. Matsuda
Faculty of Pharmaceutical Sciences, Hokkaido University
Kita-12, Nishi-6, Kita-ku, Sapporo 060-0812 (Japan)
E-mail: matuda@pharm.hokudai.ac.jp

[**] This research was supported in part by a Grant-in-Aid for Scientific Research (A) (Grant No. 23249008) from the Japan Society for the Promotion of Science (JSPS). We thank S. Oka and A. Tokumitsu (Instrumental Analysis Division, Equipment Management Center, Creative Research Institution, Hokkaido University) for measurement of the mass spectra of synthetic compounds.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201406402.



Scheme 2. Synthesis of rNaN^{O}TP (**8**). Reaction conditions: a) K-selectride, CH_2Cl_2, −40 °C, 72%; b) 1) *tert*-butyldiphenylsilyl chloride, pyridine; 2) methanesulfonyl chloride, Et_3N, pyridine, 0 °C, 81% over 2 steps; c) KO*t*Bu, DMF, 33%; d) 4,4′-dimethoxytrityl chloride, pyridine, 69%; e) OsO_4, *N*-methylmorpholine *N*-oxide, acetone–H_2O (9:1); f) 1) Ac_2O, pyridine; 2) 80% aqueous AcOH, 29% over 3 steps; g) 1) 2-chloro-4*H*-1,3,2-benzodioxaphosphorin-4-one, pyridine-1,4-dioxane (1:3); 2) $P_2O_7^{4-}$, Bu_3N, DMF; 3) I_2, pyridine–H_2O (98:2); 4) concentrated NH_4OH, 36%. DMF = *N*,*N*-dimethylformamide, Ms = methanesulfonyl.

silica-gel column chromatography. rNaN^{O}TP (**8**) was prepared from **7** by a conventional method[11] (Scheme 2). As another example, rNaO^{N}TP was synthesized from the corresponding starting material dNaO^{N}. In this case, however, to introduce the C2′–C3′ double bond, it was unnecessary to change the configuration of the 3′α-hydroxy group. After mesylation of the 3′-hydroxy group, subsequent introduction of the double bond and its oxidation with OsO_4 gave the desired product rNaO^{N}, which was further converted into rNaO^{N}TP (see Scheme S1 in the Supporting Information). On the other hand, rNaO^{O} was synthesized directly by the addition of 2,7-dimethoxy-3-iodo-1,8-naphthyridine to tri-*O*-benzylribonolactone by lithiation of the aglycon (see Scheme S2). The high solubility of the NaOMe^{OMe} aglycon in THF made it possible to synthesize the corresponding compound rNaO^{O}.

The dIm phosphoramidite units were prepared by using our previously reported method.[7a] ODN synthesis was carried out by using a 3400 DNA synthesizer (ABI) with standard reagents and program, except for a 15 min coupling time for the incorporation of dIm. Sequences for the transcription reactions are shown in Figure 1 A.

We first examined transcription reactions by using a template containing one dImO^{N} unit (Scheme 1) with four natural ribonucleoside 5′-triphosphates (NTPs; each 400 μM) as well as rNaN^{O}TP (50–400 μM; Figure 1 B, lanes 1–6). Full-length transcription products (17 mer) formed in an amount similar to that observed in the control experiment (lanes 3–6 versus lane 1), and the initiation fragments were also observed at the bottom of the gel, as in the control experiment. Some of the reactions with a lower concentration of rNaN^{O}TP (lanes 3 and 4) were terminated just one base before the NaN^{O} incorporation site. In the MALDI-TOF mass spectrum of the full-length transcript (lane 6), peaks for the desired product ($m/z = 5792.03$; calcd: 5791.71) were observed, together with those for products with higher masses ($m/z = 5841.08$ and 5889.95; Figure 1 C). The results suggested that one A:NaN^{O} ($m/z = 5841.08$; calcd: 5840.74) and two A:NaN^{O} ($m/z = 5889.95$; calcd: 5889.78) mismatched pairs formed instead of A:U pairs.

To confirm the sequence of the 17 mer full-length transcript, we carried out 2D TLC experiments[12] on the products. After transcription reactions had been performed in the presence of an [α-^{32}P]NTP with cold NTPs, the resulting transcripts were digested with RNase T_2, which cleaved the phosphodiester bond between the 3′-phosphate and 5′-hydroxy groups. We then performed 2D TLC experiments to detect ^{32}P-labeled ribonucleoside 3′-monophosphates. When [α-^{32}P]ATP was used, 5′-neighboring 3′-^{32}P-monophosphates of A in transcribed RNA were detected by the use of 2D TLC, by which the nucleotide composition of the 17 mer was determined to be correct (Table 1, entry 2; see also Figure S1 A). However, when [α-^{32}P]GTP was used, we observed that the amount of uridine 3′-^{32}P-monophosphate decreased (0.30; calcd: 1.0), and that of rNaN^{O} 3′-^{32}P-monophosphate increased (0.61; calcd: 0; Table 1, entry 3; see also Figure S1 B). Therefore, we optimized the concentrations of uridine 5′-triphosphate (UTP) and rNaN^{O}TP (Table 1; see also Figure S1) to prevent the formation of such mismatches. When 800 or 1600 μM UTP and 100 μM rNaN^{O}TP were used, 10 and 8% misincorporation of rNaN^{O}TP occurred, respectively (see Figure S1 C–F). Finally, we found that, when 3200 μM UTP and 100 μM rNaN^{O}TP were used, mismatch formation was minimized to 3% (Table 1, entries 4 and 5; see also Figures S1 G,H).

When rNaO^{O}TP was added together with four natural NTPs to the reaction mixture and a template containing ImN^{N} was used, full-length transcripts were obtained (Figure 2 A, lanes 3 and 4). However, when the concentration of rNaO^{O}TP was increased, the amount of a product shorter than the full-length product increased (Figure 2 A, lanes 5 and 6). To confirm the sequence of the full-length product, 2D TLC

experiments were carried out (see Figure S2 A–C). When 100 μM of rNaO^{O}TP (Figure 2 A, lane 4) was used, the full-length product had the expected nucleotide composition when either [α-^{32}P]ATP or [α-^{32}P]GTP was used (Figure S2 A,B), and the MALDI-TOF mass spectrum corresponded to the full-length product (see Figure S3). However, when a higher concentration of rNaO^{O}TP (200 μM; Figure 2 A, lane 5) was used, the amount of rNaO^{O} 3′-^{32}P-monophosphate increased (0.48; calcd: 0; see Figure S2 C).[2c] On the other hand, rNaO^{N}TP was incorporated only slightly into the RNA 17 mer even when high concentrations of the triphosphate were used, although the template had the complementary dImN^{O} (Figure 2 B, lanes 1–5). The major transcription product appeared to be terminated just one base before the NaO^{N} incorporation site, and the trace amount of the full-length transcript was found to be due to misincorporations during the 2D TLC experiment (see Figure S4).

Table 1: Nucleotide composition of the full-length product formed with rNaN^{O}TP.[a]

Entry	X	Y	UTP [μM]	rNaN^{O}TP [μM]	[α-^{32}P]NTP	Nucleotide composition (calcd) Ap	Gp	Cp	Up	Yp
1	A	U	400	0	ATP	0.97 (1)	2.08 (2)	n.d.	0.95 (1)	–
2	ImO^{N}	NaN^{O}	400	400	ATP	0.98 (1)	2.05 (2)	n.d.	n.d.	0.97 (1)
3	ImO^{N}	NaN^{O}	400	400	GTP	1.82 (2)	1.21 (1)	1.06 (1)	0.30 (1)	0.61 (0)
4	ImO^{N}	NaN^{O}	3200	100	ATP	1.03 (1)	1.98 (2)	n.d.	n.d.	0.99 (1)
5	ImO^{N}	NaN^{O}	3200	100	GTP	2.03 (2)	1.10 (1)	0.96 (1)	0.88 (1)	0.03 (0)

[a] n.d. = not detected.

Some DNA polymerases have been reported to show sequence-independent interactions between the enzyme and the pyrimidine O^2 atom or purine N3 atom of the incoming dNTP.[13] It is not clear if similar interactions are present between T7 RNA polymerase and NTP from its X-ray crystal structure owing to complicated transcription mechanisms, which have two phases: the initiation phase and the elongation phase. However, in T7 RNA polymerase, there is a His784 residue near the O^2 atom of the pyrimidine or N3 atom of the purine in the minor groove in both phases.[14] Since both rNaN^{O}TP and rNaO^{O}TP have a carbonyl group like the O^2 atom of the pyrimidine unit, these triphosphates showed high transcription fidelity of the Na:Im base pair. On the other hand, rNaO^{N}TP may repel hydrogen bonding in the minor groove owing to its amino group. Therefore, the transcription reaction does not occur effectively. Transcription reactions involving several unnatural base pairs, that is, isoG:isoC,[2a,c] s:y,[6a,b] Ds:Pa,[6c] and 5SICS:NaM,[5d] have been reported, and the triphosphates used (isoGTP, sTP, yTP, DsTP, PaTP, 5SICSTP, and NaMTP) have a hydrogen bond acceptor in the minor groove, as described above. Our results together with these reported results suggest that T7 RNA polymerase also participates in sequence-independent interactions with the O^2 atom of the pyrimidine base or N3 atom of the purine base of the incoming NTP in a similar way to DNA polymerases.

Finally, a longer transcription reaction with the NaN^{O}:ImO^{N} pair was carried out under the optimized

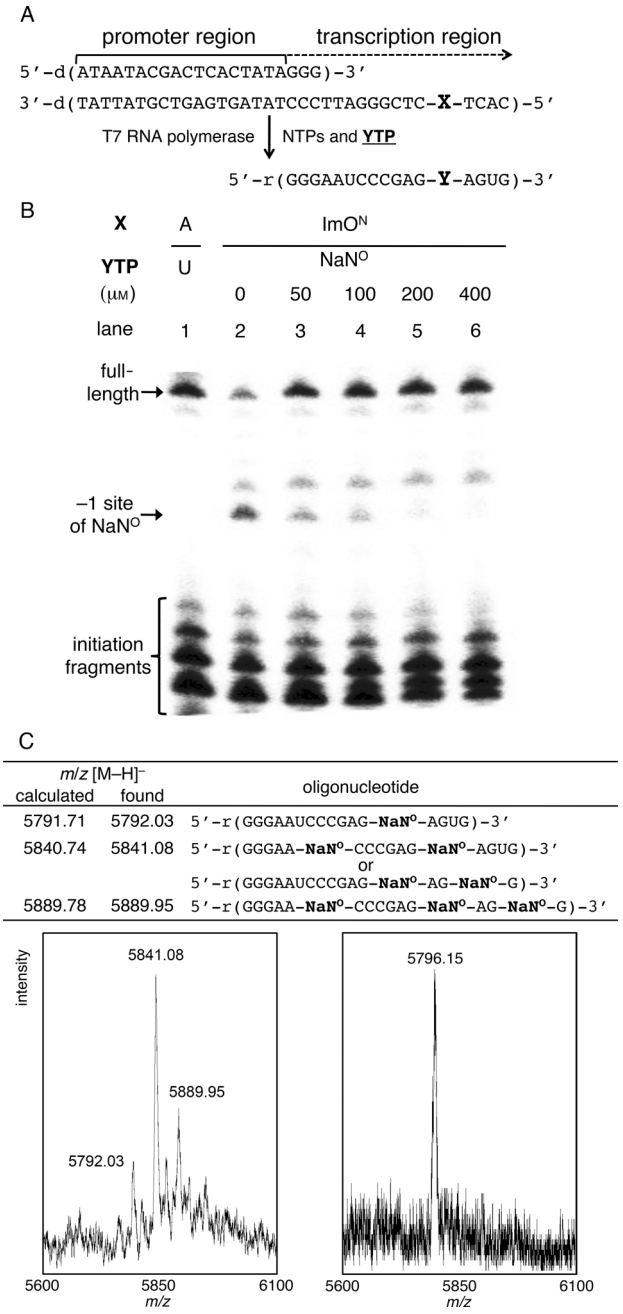


Figure 1. A) Sequence of the 35 mer DNA template containing Im and the 17 mer transcription product containing rNa. B) Gel electrophoresis of the transcription products formed by the use of rNaN^{O}TP, and a 2D TLC experiment for the full-length transcription product at 400 μM rNaN^{O}TP with [α-^{32}P]ATP (lane 6): Lane 1, U:A control; lanes 2–6, various concentrations of rNaN^{O}TP. C) MALDI-TOF mass spectra of the transcription product formed by the use of 400 μM UTP and 400 μM rNaN^{O}TP (left) and 3200 μM UTP and 100 μM rNaN^{O}TP (right). These spectra were obtained from a transcription reaction after desalting with an ODS column.

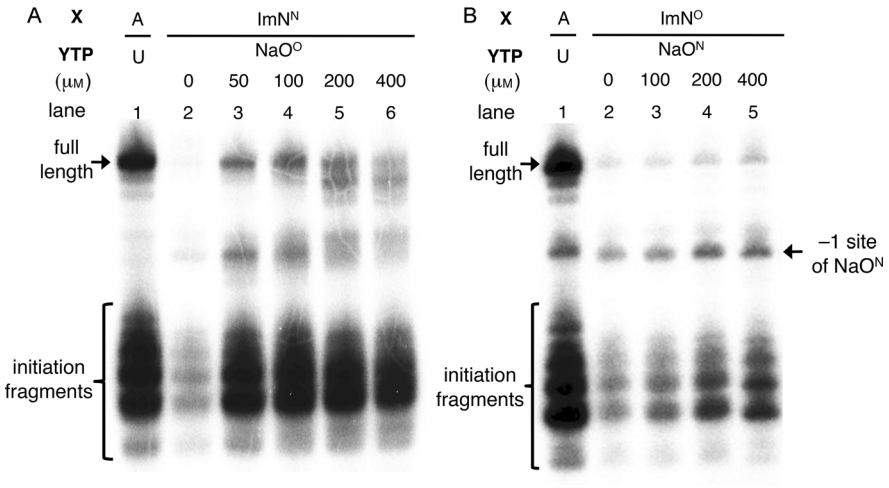
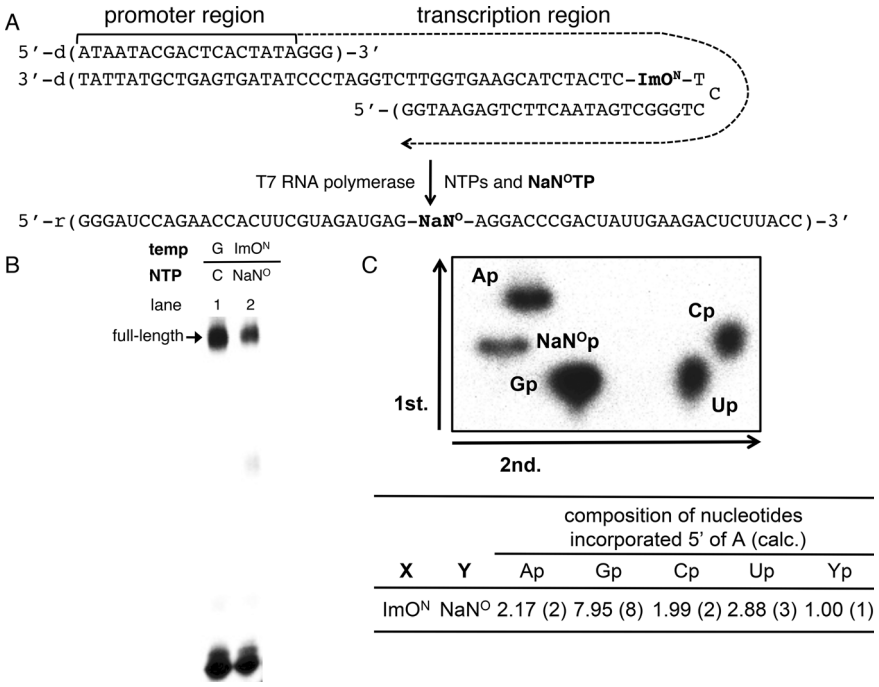


Figure 2. A) Gel electrophoresis of the transcription products formed by the use of rNaO^{O}TP: Lane 1, U:A controls; lanes 2–6, various concentrations of rNaO^{O}TP. B) Gel electrophoresis of the transcription products formed by the use of rNaO^{N}TP: Lane 1, U:A controls; lanes 2–5, various concentrations of rNaO^{N}TP.



		composition of nucleotides incorporated 5′ of A (calc.)				
X	Y	Ap	Gp	Cp	Up	Yp
ImO^{N}	NaO^{O}	2.17 (2)	7.95 (8)	1.99 (2)	2.88 (3)	1.00 (1)

Figure 3. A) Sequence of the 73 mer DNA template containing ImO^{N} and the 55 mer transcription product containing rNaN^{O}. B) Gel electrophoresis of the transcription products formed by the use of the 73 mer DNA template. The reaction was carried out similarly to that in Figure 1B and then analyzed by 10% PAGE (1000 V, 2.5 h) containing 7 M urea. Lane 1, C:G control; lane 2, rNaN^{O}TP. C) 2D TLC analysis of the full-length transcription product (in lane 2) formed by the use of [α-^{32}P]ATP.

conditions with a 73 mer template DNA molecule containing part of the human c-Ha-Ras gene sequence and the T7 promoter region (Figure 3). A full-length product with nearly the same amount of natural C:G pairs was observed by gel electrophoresis. By using [α-^{32}P]ATP, we determined that the full-length transcription product had the correct nucleotide composition and no A:NaN^{O} mismatches. This result indicates that the NaN^{O}:ImO^{N} pair functions as a new base pair, which can be transcribed from a DNA duplex to the corresponding mRNA.

In conclusion, we investigated transcription reactions with Na:Im base pairs by using T7 RNA polymerase. NaN^{O}:ImO^{N} and NaO^{O}:ImN^{N} pairs in oligodeoxyribonucleotides were transcribed with high fidelity, and the NaN^{O}:ImO^{N} pair provided the transcription product in high yield, similar to that of the control experiment involving only natural nucleotides. Moreover, we optimized the concentrations of UTP and rNaN^{O}TP, and under the optimized conditions, a part of the c-Ha-Ras gene containing one ImO^{N} base instead of G was successfully transcribed to a 55 mer RNA containing one NaN^{O} base without any mismatches. In other words, the NaN^{O}:ImO^{N} pair can be employed in translation studies, which is the final requirement according to the "central dogma".

Received: June 19, 2014
Revised: August 6, 2014
Published online: September 22, 2014

Keywords: hydrogen bonds · unnatural base pairs · nucleobases · nucleotides · RNA recognition

[1] S. A. Benner, *Science* **2004**, *306*, 625–626.
[2] a) C. Switzer, S. E. Moroney, S. A. Benner, *J. Am. Chem. Soc.* **1989**, *111*, 8322–8323; b) J. A. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* **1990**, *343*, 33–37; c) C. Y. Switzer, S. E. Moroney, S. A. Benner, *Biochemistry* **1993**, *32*, 10489–10496; d) H.-J. Kim, N. A. Leal, S. Hoshika, S. A. Benner, *J. Org. Chem.* **2014**, *79*, 3194–3199.
[3] For recent reviews, see: a) A. A. Henry, F. E. Romesberg, *Curr. Opin. Chem. Biol.* **2003**, *7*, 727–733; b) S. A. Benner, A. M. Sismour, *Nat. Rev. Genet.* **2005**, *6*, 533–543; c) I. Hirao, *Curr. Opin. Chem. Biol.* **2006**, *10*, 622–627; d) A. T. Krueger, E. T. Kool, *Curr. Opin. Chem. Biol.* **2007**, *11*, 588–594.
[4] a) S. Moran, R. X.-F. Ren, S. Ramney IV, E. T. Kool, *J. Am. Chem. Soc.* **1997**, *119*, 2056–2057; b) J. C. Morales, E. T. Kool, *J. Am. Chem. Soc.* **1999**, *121*, 2323–2324.
[5] a) D. L. McMinn, A. K. Ogawa, Y. Wu, J. Liu, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **1999**, *121*, 11585–11586; b) A. M. Leconte, G. T. Hwang, S. Matsuda, P. Capek, Y. Hari, F. E. Romesberg, *J. Am. Chem. Soc.* **2008**, *130*, 2336–2343;

c) Y. J. Seo, G. T. Hwang, P. Ordoukhanian, F. E. Romesberg, *J. Am. Chem. Soc.* **2009**, *131*, 3246 – 3252; d) Y. J. Seo, S. Matsuda, F. E. Romesberg, *J. Am. Chem. Soc.* **2009**, *131*, 5046 – 5047.

[6] a) T. Ohtsuki, M. Kimoto, M. Ishikawa, T. Mitsui, I. Hirao, S. Yokoyama, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4922 – 4925; b) I. Hirao, T. Ohtsuki, T. Fujiwara, T. Mitsui, T. Yokogawa, T. Okuni, H. Nakayama, K. Takio, T. Yabuki, T. Kigawa, K. Kodama, T. Yokogawa, K. Nishikawa, S. Yokoyama, *Nat. Biotechnol.* **2002**, *20*, 177 – 182; c) I. Hirao, M. Kimoto, T. Mitsui, T. Fujiwara, R. Kawai, A. Sato, Y. Harada, S. Yokoyama, *Nat. Methods* **2006**, *3*, 729 – 735; d) N. Morohashi, M. Kimoto, A. Sato, R. Kawai, I. Hirao, *Molecules* **2012**, *17*, 2855 – 2876.

[7] a) N. Minakawa, N. Kojima, S. Hikishima, T. Sasaki, A. Kiyosue, N. Atsumi, Y. Ueno, A. Matsuda, *J. Am. Chem. Soc.* **2003**, *125*, 9970 – 9982; b) S. Hikishima, N. Minakawa, K. Kuramoto, Y. Fujisawa, M. Ogawa, A. Matsuda, *Angew. Chem. Int. Ed.* **2005**, *44*, 596 – 598; *Angew. Chem.* **2005**, *117*, 602 – 604; c) S. Hikishima, N. Minakawa, K. Kuramoto, S. Ogata, A. Matsuda, *ChemBioChem* **2006**, *7*, 1970 – 1975; d) K. Kuramoto, N. Tarashima, Y. Hirama, Y. Kikuchi, N. Minakawa, A. Matsuda, *Chem. Commun.* **2011**, *47*, 10818 – 10820.

[8] a) N. Minakawa, S. Ogata, M. Takahashi, A. Matsuda, *J. Am. Chem. Soc.* **2009**, *131*, 1644 – 1645; b) S. Ogata, M. Takahashi, N. Minakawa, A. Matsuda, *Nucleic Acids Res.* **2009**, *37*, 5602 – 5609.

[9] The direct coupling of NaN^{O} aglycon to a ribosugar was not possible owing to the aglycon's insolubility in THF.

[10] When known the 3′α-hydroxy analogue was used for this synthetic strategy, cyclization between O^{7} and 3′C occurred instead of the formation of the desired elimination product.

[11] J. Ludwig, F. Eckstein, *J. Org. Chem.* **1989**, *54*, 631 – 635.

[12] M. Silberklang, A. Prochiantz, A.-L. Haenni, U. L. Rajbhandary, *Eur. J. Biochem.* **1977**, *72*, 465 – 478.

[13] a) S. Doublié, S. Tabor, A. M. Long, C. C. Richardson, T. Ellenberger, *Nature* **1998**, *391*, 251 – 258; b) Y. Li, S. Korolev, G. Waksman, *EMBO J.* **1998**, *17*, 7514 – 7525; c) J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, *Nature* **1998**, *391*, 304 – 307; d) Y. Li, G. Waksman, *Protein Sci.* **2001**, *10*, 1225 – 1233.

[14] a) Y. W. Yin, T. A. Steitz, *Cell* **2004**, *116*, 393 – 404; b) W. P. Kennedy, J. R. Momand, Y. W. Yin, *J. Mol. Biol.* **2007**, *370*, 256 – 268; c) D. Temiakov, V. Patlan, M. Anikin, W. T. McAllister, S. Yokoyama, D. G. Vassylyev, *Cell* **2004**, *116*, 381 – 391; d) G. M. T. Cheetham, T. A. Steitz, *Science* **1999**, *286*, 2305 – 2309.

