

# An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogs into DNA and RNA

Ichiro Hirao<sup>1,2</sup>, Michiko Kimoto<sup>1</sup>, Tsuneo Mitsui<sup>1,2</sup>, Tsuyoshi Fujiwara<sup>1,5</sup>, Rie Kawai<sup>1</sup>, Akira Sato<sup>1</sup>, Yoko Harada<sup>1</sup> & Shigeyuki Yokoyama<sup>1,3,4</sup>

**Methods for the site-specific incorporation of extra components into nucleic acids can be powerful tools for creating DNA and RNA molecules with increased functionality. We present an unnatural base pair system in which DNA containing an unnatural base pair can be amplified and function as a template for the site-specific incorporation of base analog substrates into RNA via transcription. The unnatural base pair is formed by specific hydrophobic shape complementation between the bases, but lacks hydrogen bonding interactions. In replication, this unnatural base pair exhibits high selectivity in combination with the usual triphosphates and modified triphosphates,  $\gamma$ -amidotriphosphates, as substrates of 3' to 5' exonuclease-proficient DNA polymerases, allowing PCR amplification. In transcription, the unnatural base pair complementarity mediates the incorporation of these base substrates and their analogs, such as a biotinylated substrate, into RNA by T7 RNA polymerase (RNAP). With this system, functional components can be site-specifically incorporated into a large RNA molecule.**

Nucleic acids are important in current biotechnology, as they can be amplified by the complementarity between the A · T(U) and G · C base pairs, and display versatile functionalities. But the limitation of having only four different base components (nucleotides) in standard nucleic acids restricts their functions, as compared to the 20 different amino acids in natural proteins. An unnatural base pair system addresses this problem by expanding the genetic information contained in nucleic acids<sup>1–3</sup>. The specific complementarity of an unnatural base pair allows the site-directed incorporation of extra nucleotide analogs into DNA and RNA by polymerases. DNA fragments containing an unnatural base pair can be prepared by a conventional method combining chemical DNA synthesis and enzymatic ligation. An unnatural base pair that selectively functions in replication and transcription facilitates the PCR amplification of the DNA fragments and synthesis of RNA molecules containing nucleotide analogs at desired positions.

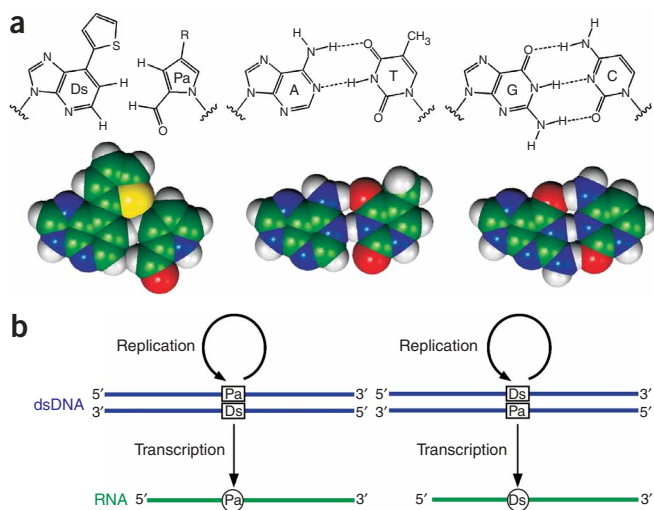
The first unnatural base pairs to be developed were isoguanine and isocytosine (isoG · isoC), and xanthosine and diaminopyrimidine, which had different hydrogen-bonding patterns from those of the natural base pairs<sup>4,5</sup>. DNA fragments containing these unnatural base pairs had been subjected to PCR amplification<sup>6–8</sup>, and a modified isoG had been incorporated into RNA by transcription using isoC-containing templates<sup>9</sup>. Unfortunately, T7 RNAP did not recognize the 2-amino pyrimidine analogs, isoC and diaminopyrimidine, as substrates<sup>5</sup>.

Recently, we developed unnatural base pairs, including 2-amino-6-(2-thienyl)purine (s) and 2-oxypyridine (y)<sup>10,11</sup>. The bulky thienyl group of s effectively prevented noncognate pairings with natural bases; the substrate (nucleoside 5'-triphosphate) of y could be site-specifically incorporated into RNA, opposite s in templates, by T7 RNAP. Although this transcription reaction could be used practically to introduce functional components into RNA<sup>12–14</sup>, the selectivity of the s · y pair in replication was not as high as that in transcription<sup>10</sup>.

Hydrophobic base pairs also have become candidates for the expansion of genetic information. Studies have revealed the importance of the geometrical shape complementarity between the pairing bases, demonstrating that hydrogen bonds are not an absolute requirement for replication<sup>15</sup>. Recently, several hydrophobic base pairs had been developed and tested in replication<sup>16,17</sup>. One problem with these hydrophobic base pairs is nonspecific incorporation between the hydrophobic bases, although this characteristic has led to the development of self-complementary pairs<sup>17</sup>.

Here we report the design and analysis of the hydrophobic base pair, 7-(2-thienyl)-imidazo[4,5-b]pyridine (Ds) and pyrrole-2-carbaldehyde (Pa) (**Fig. 1a**) in *in vitro* replication and transcription (**Fig. 1b**). In replication, this base pair displays robust selectivity in combination with the usual triphosphates and modified triphosphates,  $\gamma$ -amidotriphosphates, as substrates of 3' to 5' exonuclease-proficient DNA polymerases, allowing PCR

<sup>1</sup>Protein Research Group, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. <sup>2</sup>Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan. <sup>3</sup>Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. <sup>4</sup>RIKEN Harima Institute at Spring-8, 1-1-1 Kouto, Mikazuki-cho, Sayo, Hyogo 679-5148, Japan. <sup>5</sup>Present address: Nanotechnology Research Center, Fujitsu Laboratories, 10-1 Morinosato-Wakamiya, Atsugi, Kanagawa 243-0197, Japan. Correspondence should be addressed to I.H. (hirao@riken.jp) or S.Y. (yokoyama@biochem.s.u-tokyo.ac.jp).



**Figure 1** | An unnatural base pair system for specific replication and transcription. **(a)** The unnatural Ds-Pa and natural base pairs (Pa: R = H, Pa': R = C≡C-CH<sub>3</sub>). **(b)** The unnatural base pair system that functions in PCR amplification, primer extension, DNA sequencing and T7 transcription. Original DNA templates were prepared by chemical synthesis and ligation, and were amplified by PCR with unmodified dNTPs (dPaTP, dGTP, dCTP and dTTP) and modified dNTPs ( $\gamma$ -amidotriphosphates, dDsTP<sub>N</sub> and dATP<sub>N</sub>). RNA molecules containing Pa or modified Pa bases at specific positions were transcribed from DNA templates containing Ds in the template strands, by T7 RNA polymerase with PaTP (or modified PaTP) and the natural NTPs (left). RNA molecules containing Ds were transcribed from DNA templates containing Pa in the template strands, with DsTP and the natural NTPs (right).

amplification. Furthermore, these unnatural bases and their base analogs were complementarily incorporated into RNA by standard T7 transcription.

## RESULTS

### Design of unnatural base pair systems using Ds-Pa

We designed the Ds-Pa pair (Fig. 1a) with two goals in mind: (i) to increase the pairing selectivity, we used hydrophobic bases with different shapes from those of the natural bases<sup>15,18</sup>, (ii) to allow interaction with polymerases<sup>19,20</sup>, we added proton acceptor groups, including the nitrogen at position 4 of Ds (corresponding to position 3 of adenine and guanine) and the aldehyde group of Pa (corresponding to the 2-keto groups of cytosine and thymine). We initially developed Pa<sup>21</sup> as a specific pairing partner of another unnatural base, 7-methylimidazo[4,5-b]pyridine (Q)<sup>22</sup>. Q, however, is a shape analog of adenine, and it efficiently pairs with both thymine and Pa. Therefore, we replaced the 7-methyl group of Q with a larger thienyl group, which effectively excluded its pairing with the natural bases, but accommodated pairing with the five-membered ring bases well<sup>23</sup>.

One problem with hydrophobic base pairs is their potential for nonshape complementary pairings between hydrophobic bases<sup>17</sup>, such as Ds-Ds pairing. To test for the possible formation of the undesired Ds-Ds pair and the selectivity of the cognate Ds-Pa pair in replication, we determined the kinetic parameters of the cognate and noncognate base pairings between a substrate and a template base by single-nucleotide insertion experiments, using the exonuclease-deficient Klenow fragment of *Escherichia coli* DNA polymerase I (KF exo<sup>-</sup>)<sup>24–26</sup> (Fig. 2a). The substrates (dDsTP and dPaTP) and templates containing Ds or Pa were chemically synthesized (Supplementary Methods online and Supplementary Data online). The kinetic parameters (Fig. 2b,c and Supplementary Table 1 online) indicated that the efficiencies of the cognate dDs-dPa and dA-dT pairings are higher than those of their noncognate pairings (Fig. 2b). But undesired dDsTP incorporation opposite Ds (incorporation efficiency,  $V_{\max}/K_M = 2.0 \times 10^5 \%$  min<sup>-1</sup> M<sup>-1</sup>) was more efficient than the desired dPaTP incorporation opposite Ds ( $V_{\max}/K_M = 6.2 \times 10^4 \%$  min<sup>-1</sup> M<sup>-1</sup>). The incorporation of dDsTP opposite Ds inhibited further extension; the addition of dDsTP (0.5 or 1 mol equivalent of dPaTP) inhibited the extension by the 3' to 5' exonuclease-proficient Klenow fragment (KF exo<sup>+</sup>) after the

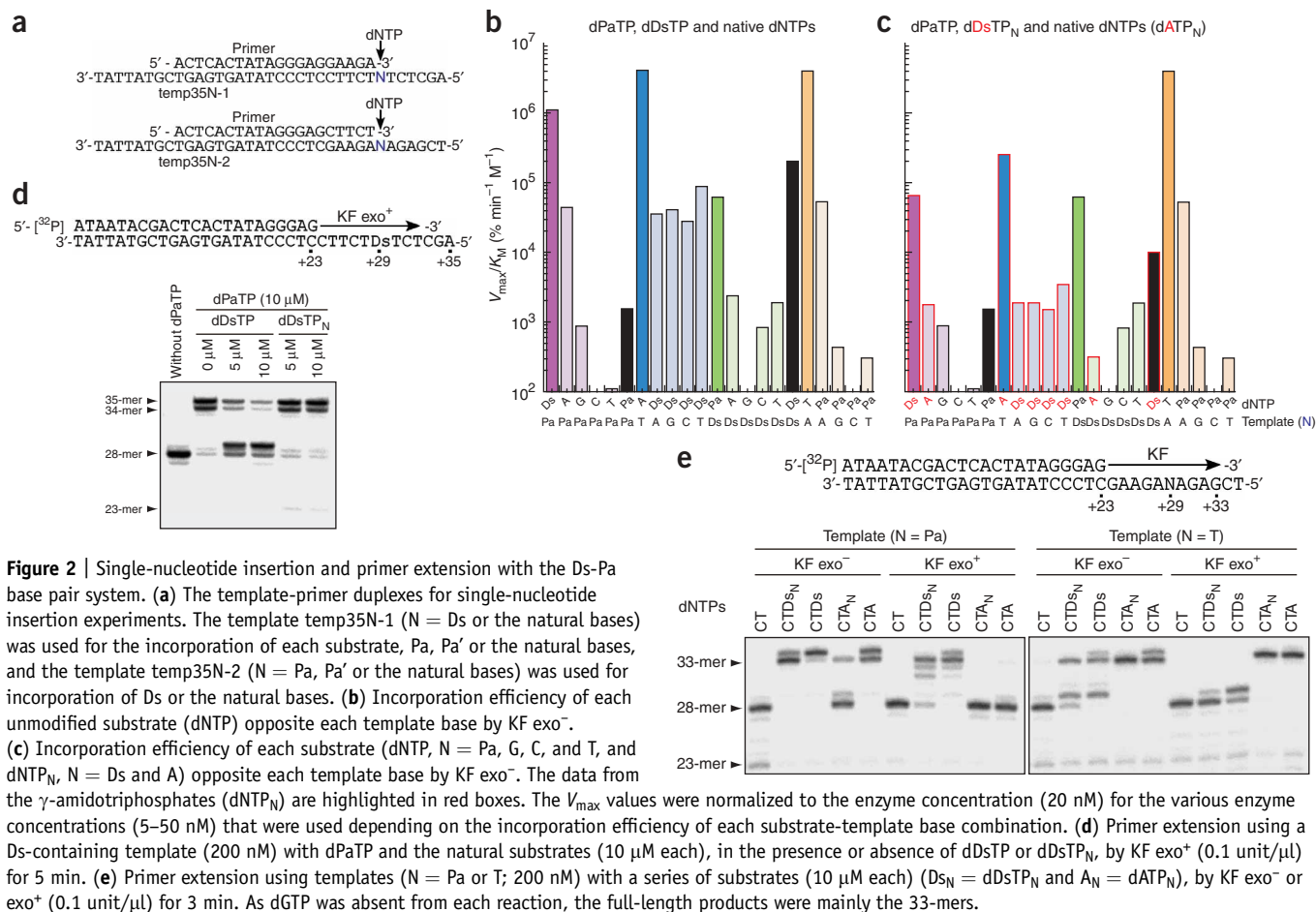
incorporation of dDsTP, instead of dPaTP, opposite Ds at position 29 in the template (Fig. 2d).

To reduce the amount of Ds incorporation opposite Ds, we used a modified Ds triphosphate,  $\gamma$ -amidotriphosphate (denoted as dDsTP<sub>N</sub>), as the substrate. The incorporation efficiency of this modified substrate might be lower than that of the usual substrates, so its use in combination with the modified and usual substrates could adjust the selectivity between the cognate and noncognate pairings. We synthesized the  $\gamma$ -amidotriphosphates with a slight modification of the conventional chemical synthesis (Supplementary Methods). As expected, the incorporation efficiency of dDsTP<sub>N</sub> opposite Ds ( $V_{\max}/K_M = 9.9 \times 10^3 \%$  min<sup>-1</sup> M<sup>-1</sup>) was decreased. Although the incorporation efficiency of dDsTP<sub>N</sub> opposite Pa ( $V_{\max}/K_M = 6.7 \times 10^4 \%$  min<sup>-1</sup> M<sup>-1</sup>) was reduced to a level close to that of the dATP incorporation opposite Pa ( $V_{\max}/K_M = 4.3 \times 10^4 \%$  min<sup>-1</sup> M<sup>-1</sup>), we resolved this problem by using the  $\gamma$ -amidotriphosphate of adenine (dATP<sub>N</sub>). The incorporation efficiency of dDsTP<sub>N</sub> opposite Pa was 37-fold higher than that of dATP<sub>N</sub> opposite Pa ( $V_{\max}/K_M = 1.8 \times 10^3 \%$  min<sup>-1</sup> M<sup>-1</sup>). Thus, we achieved highly complementary selectivity by using a combination of the usual triphosphates, dPaTP, dGTP, dCTP and dTTP, and the  $\gamma$ -amidotriphosphates, dDsTP<sub>N</sub> and dATP<sub>N</sub> (Fig. 2c).

Primer extension proceeded after Pa incorporation opposite Ds in the presence of dDsTP<sub>N</sub> (Fig. 2d). Furthermore, extension after the incorporation of dDsTP<sub>N</sub> opposite Pa or dATP<sub>N</sub> opposite thymine also continued efficiently (Fig. 2e). Notably, dATP<sub>N</sub> reduced the extension when adenine was misincorporated opposite Pa (as compared to dATP). Moreover, the 3' to 5' exonuclease activity of KF exo<sup>+</sup> increased the selectivity of the Ds-Pa pair. The noncognate A-Pa and Ds-T pairings were effectively prohibited by KF exo<sup>+</sup>, and the extension involving the noncognate pairings paused around the unnatural base position (Fig. 2e).

### Sequencing of DNA fragments containing Ds

To determine the position of the unnatural base in the DNA, we developed a sequencing method using dideoxynucleotide chain-termination supplemented with the triphosphates of the unnatural bases (Fig. 3). For the sequencing of DNA containing Ds, we used modified Pa, 4-propynylpyrrole-2-carbaldehyde (Pa')<sup>27</sup>. dPa'TP produced clearer sequencing peak patterns around the Ds-Pa site than those generated by dPaTP; dideoxy-terminated fragments next to Pa incorporation yielded larger peaks than those of fragments next to the site of Pa' incorporation (data not shown). For the sequencing experiments, we constructed ten double-stranded DNA fragments, each containing one Ds-Pa pair and a different sequence around the Ds-Pa pair (DNA1 and DNA2, 150-mers and DNA3–DNA10, 174-mers), by primer extension (Fig. 3a) or by



**Figure 2** | Single-nucleotide insertion and primer extension with the Ds-Pa base pair system. **(a)** The template-primer duplexes for single-nucleotide insertion experiments. The template temp35N-1 (N = Ds or the natural bases) was used for the incorporation of each substrate, Pa, Pa' or the natural bases, and the template temp35N-2 (N = Pa, Pa' or the natural bases) was used for incorporation of Ds or the natural bases. **(b)** Incorporation efficiency of each unmodified substrate (dNTP) opposite each template base by KF exo<sup>-</sup>. **(c)** Incorporation efficiency of each substrate (dNTP, N = Pa, G, C, and T, and dNTP<sub>N</sub>, N = Ds and A) opposite each template base by KF exo<sup>-</sup>. The data from the  $\gamma$ -amidotriphosphates (dNTP<sub>N</sub>) are highlighted in red boxes. The  $V_{max}$  values were normalized to the enzyme concentration (20 nM) for the various enzyme concentrations (5–50 nM) that were used depending on the incorporation efficiency of each substrate-template base combination. **(d)** Primer extension using a Ds-containing template (200 nM) with dPaTP and the natural substrates (10 μM each), in the presence or absence of dDsTP or dDsTP<sub>N</sub>, by KF exo<sup>+</sup> (0.1 unit/μl) for 5 min. **(e)** Primer extension using templates (N = Pa or T; 200 nM) with a series of substrates (10 μM each) (Ds<sub>N</sub> = dDsTP<sub>N</sub> and A<sub>N</sub> = dATP<sub>N</sub>), by KF exo<sup>-</sup> or exo<sup>+</sup> (0.1 unit/μl) for 3 min. As dGTP was absent from each reaction, the full-length products were mainly the 33-mers.

ligation using chemically synthesized DNA fragments (**Supplementary Methods**).

In sequencing DNA1, a double-stranded 150-mer DNA fragment containing a 5'-CDsA-3'/3'-GPaT-5' sequence, the addition of 0.05 mM dPa'TP prevented the incorporation of the dye terminators of natural bases opposite Ds. Only the peak corresponding to the base opposite Ds disappeared, and there was no signal attenuation of the other peaks (**Fig. 3c**). Additionally, in the absence of dPa'TP the sequencing terminated at the position opposite Ds, and the subsequent peaks almost disappeared (**Fig. 3d**). Thus, we could confirm the unnatural base position in DNA fragments by comparing the peak patterns obtained from sequencing in both the presence and absence of dPa'TP (see **Supplementary Methods** and **Supplementary Figs. 1** and **2** online).

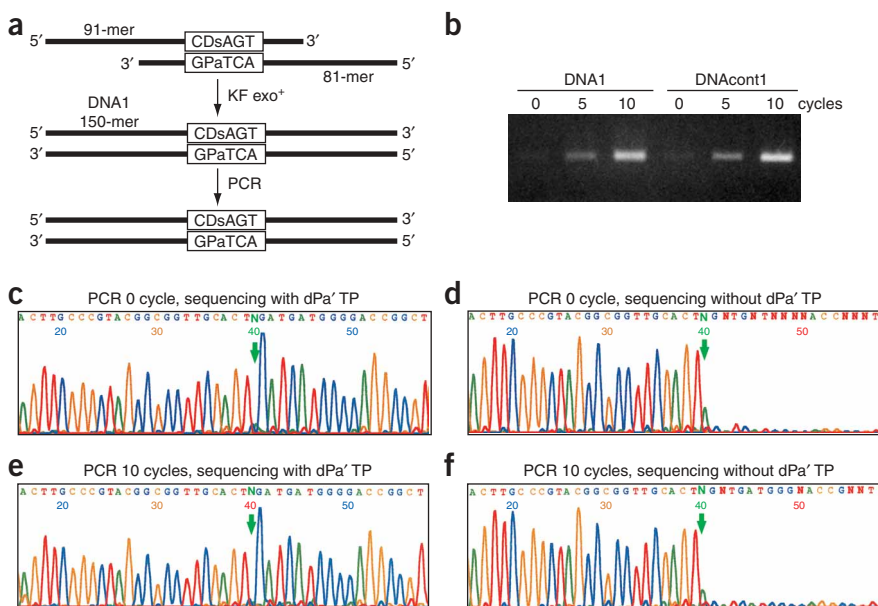
### PCR amplification

We optimized the PCR conditions for the amplification of DNA1–DNA10 templates containing one Ds-Pa pair using a higher concentration of thermophilic DNA polymerase with 3' to 5' exonuclease activity, longer extension time, and a substrate mixture of dDsTP<sub>N</sub>, dPaTP, dATP<sub>N</sub>, dGTP, dCTP and dTTP. By cloning and sequencing the products of a natural DNA fragment, amplified by 10-cycle PCR using the optimized conditions with the substrate mixture, we confirmed that the optimized PCR conditions did not result in increased mutagenesis as compared to conventional PCR conditions (data not shown). The amplification efficiency

of DNA1 under the optimized PCR conditions was ~80% of that of the natural base pair system under conventional PCR conditions (**Fig. 3b**).

The Ds-Pa pair system exhibited high selectivity in PCR amplification. After 10 PCR cycles with DNA1 as the template and using the optimized PCR conditions, we analyzed the products by sequencing (**Fig. 3e,f**). We also performed 10 more cycles of PCR (20 cycles in total) using a portion of the 10-cycle PCR product (**Supplementary Fig. 1**). In the presence of dPa'TP, the sequencing of the 10-cycle and 20-cycle PCR products yielded peak patterns similar to those of the original DNA1. Sequencing without dPa'TP revealed that the read-through peaks resulting from bases following the unnatural base position were slightly increased with greater numbers of PCR cycles. By analyzing the read-through peaks, we estimated the mutation rate of the Ds-Pa pair compared to the rate for the natural base pairs using the peak heights obtained from standard DNA fragments, which contained 1–10% of the A · T pair in place of the Ds-Pa pair (**Supplementary Fig. 1**). The total mutation rates of the Ds-Pa site in DNA1 after 10 cycles and 20 cycles were ~1% and 3–4%, respectively (see **Supplementary Discussion** online).

We also amplified DNA fragments that contained different sequences around the Ds-Pa pair (**Supplementary Fig. 2**) and found two limitations of the system. One limitation was that the amplification efficiency of DNA fragments containing 5'-X(A)<sub>n</sub>-3' sequences (where X = Ds or Pa and n ≥ 2, such as 5'-DsAA-3' and



**Figure 3** | DNA sequencing and PCR amplification of the DNA fragments containing the Ds-Pa pair. **(a)** The double stranded DNA fragment (150-mer, DNA1) was prepared by primer extension using chemically synthesized DNA fragments (91-mer and 81-mer) containing Ds and Pa. **(b)** Agarose-gel analysis of original DNA fragments (0 cycle) and PCR products after 5 and 10 cycles of amplification. For DNA1, PCR was performed with 0.04 unit/ $\mu\text{l}$  Vent DNA polymerase and the following cycle conditions: 0.5 min at 94 °C, 0.5 min at 45 °C and 4 min at 65 °C. For DNAcont1 consisting only of the natural bases, PCR was performed with 0.01 unit/ $\mu\text{l}$  Vent DNA polymerase with the following cycle conditions: 0.5 min at 94 °C, 0.5 min at 45 °C and 1 min at 72 °C. **(c–f)** DNA sequencing, in the presence **(c,e)** or absence **(d,f)** of dPa' TP, of the original DNA1 **(c,d)** and PCR-amplified DNA1 after 10 cycles using the unnatural base pair system **(e,f)**. The sequencing reaction was performed with a BigDye Terminator v1.1 Cycle Sequencing kit, containing a mutated Taq DNA polymerase (**Supplementary Methods**).

5'-PaAA-3' (DNA10)) was substantially reduced. But the presence of A-tract sequences without any unnatural bases in DNA fragments did not affect the amplification efficiency in the presence of dATP<sub>N</sub>. Another limitation was that the amplification efficiency after more than 10 PCR cycles gradually decreased. This problem, however, could be solved by diluting the PCR solution after 10 amplification cycles or by removing the low-molecular-weight products in the solution by filtration. Although both limitations might be caused by the presence of  $\gamma$ -amidotriphosphates and/or the production of amidopyrophosphates, we have not determined the actual reasons. Nevertheless, by keeping in mind these limitations, most DNA fragments containing the Ds-Pa pair can be amplified.

### T7 transcription of templates with the Ds-Pa pair

The Ds-Pa pair complementarity mediates the site-specific incorporation of PaTP, modified PaTP and DsTP into RNA by T7 RNAP, but does not require  $\gamma$ -amidotriphosphates. We examined transcription of DNA templates (35-mer) containing Ds, Pa or Pa' (see **Supplementary Fig. 3** online). After 3 h of transcription with the triphosphates of unnatural bases, we analyzed the <sup>32</sup>P-labeled transcripts by gel electrophoresis (**Supplementary Fig. 3**). The full-length transcripts (17-mer) containing Pa, Pa' or Ds had 28–91% of the yield relative to that of the transcript containing only natural bases.

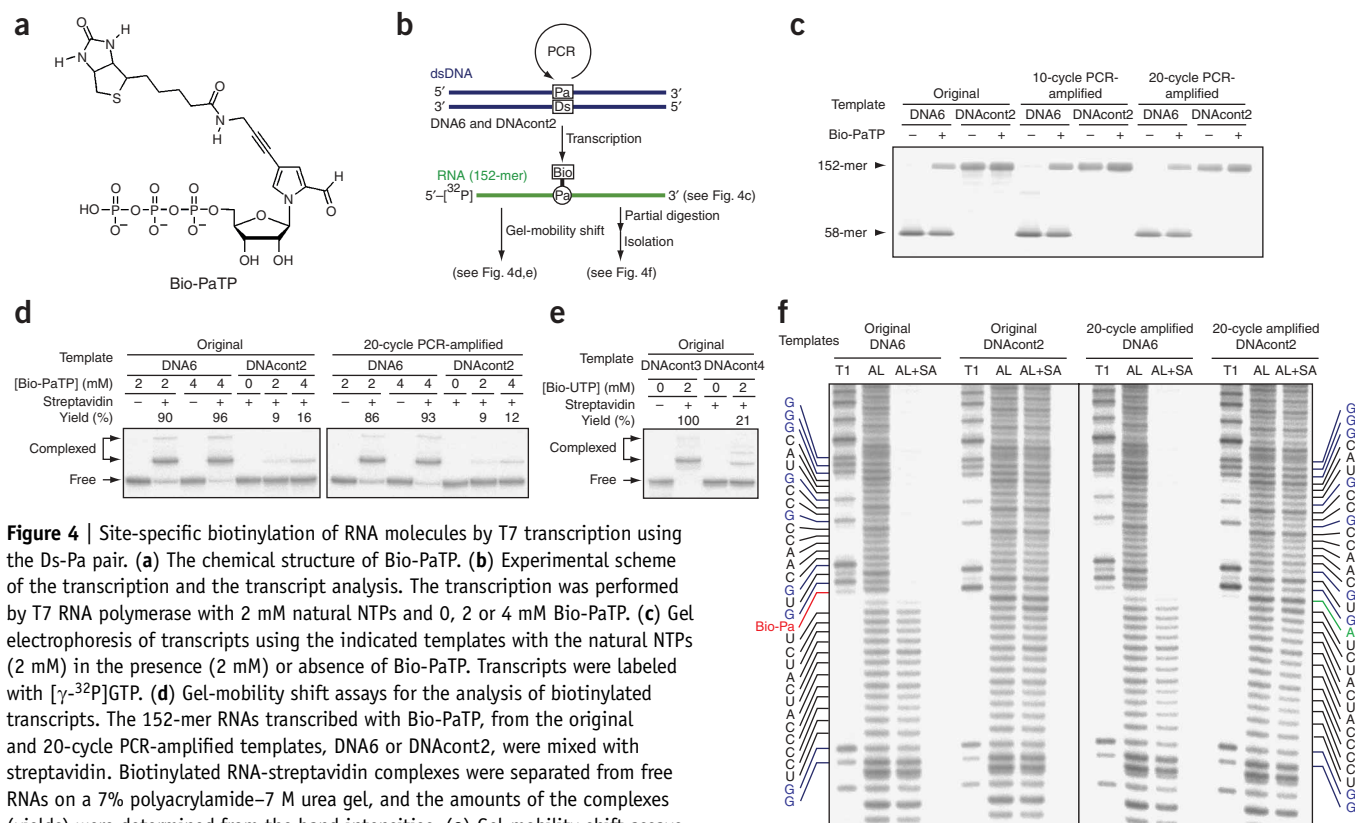
We confirmed the high selectivity of the Ds-Pa and Ds-Pa' pairings in T7 transcription by nucleotide-composition analysis<sup>11</sup> (**Supplementary Fig. 3** and **Supplementary Table 2** online). We observed low yields of the 17-mer transcripts in transcription reactions with templates containing Ds or Pa but without the unnatural base substrates in the reaction (by misincorporation of the natural substrates opposite the unnatural bases). This was especially true for Pa- or Pa'-containing templates. But misincorporation was effectively prevented by the presence of the unnatural base substrates as the incorporation selectivity of the unnatural base substrates opposite their pairing partners was more than 94% (**Supplementary Table 2**). Also, we detected no misincorporation

of the unnatural base substrates opposite the natural bases in this analysis (see **Supplementary Discussion**).

### Site-specific biotinylation of RNA by using Bio-PaTP'

To examine in greater detail the selectivity and potential application of the Ds-Pa pair in both transcription and PCR amplification, we used site-specifically biotinylated 152-mer RNA molecules, which can be detected upon binding to streptavidin. We incorporated biotin-linked Pa (Bio-Pa) into RNA by T7 transcription mediated by Ds-Pa using the original and PCR-amplified templates. We chemically synthesized the substrate, biotinylated-Pa (Bio-PaTP; **Fig. 4a** and **Supplementary Methods**). To incorporate Bio-Pa into an RNA molecule (152-mer), we transcribed DNA6, which contains one Ds-Pa pair, and a natural DNA fragment, DNAcont2, with natural (2 mM) and Bio-PaTP (2 or 4 mM) substrates (**Fig. 4b**). The efficiencies of DNA6 transcription were 47–85% relative to that of DNAcont2 (**Fig. 4c**). In the absence of Bio-PaTP, the transcription of the amplified DNA6 yielded lower amounts of the full-length product (**Fig. 4c**), indicating very little replacement of Ds-Pa with the natural base pairs (3–4% after the 20-cycle amplification) occurred during PCR amplification.

We assessed the selectivity of Bio-Pa incorporation into RNA by a gel-mobility shift assay of the biotinylated transcripts, in the presence of streptavidin (**Fig. 4d**). In transcription reactions using the original DNA6 template with 2 mM Bio-PaTP (equivalent to the natural substrates) 90% of the transcripts were biotinylated. Increasing the Bio-PaTP concentration to 4 mM improved the incorporation rate to 96%. Using DNAcont2, we also assessed the misincorporation of Bio-PaTP opposite the natural bases. In transcription with 2 and 4 mM Bio-PaTP, the yields of the biotinylated transcripts were 9 and 16%, respectively (**Fig. 4d**). These misincorporations correspond to only 0.06% (for 9% total) and 0.12% (for 16% total) per position in the 152-mer transcript (see **Supplementary Methods**). In transcription reactions using the 20-cycle PCR-amplified template, the selectivity of Bio-Pa incorporation was slightly reduced, by 3–4% (**Fig. 4d**). These values agreed well with the mutation rate of the Ds-Pa site (3–4%)



**Figure 4** | Site-specific biotinylation of RNA molecules by T7 transcription using the Ds-Pa pair. **(a)** The chemical structure of Bio-PaTP. **(b)** Experimental scheme of the transcription and the transcript analysis. The transcription was performed by T7 RNA polymerase with 2 mM natural NTPs and 0, 2 or 4 mM Bio-PaTP. **(c)** Gel electrophoresis of transcripts using the indicated templates with the natural NTPs (2 mM) in the presence (2 mM) or absence of Bio-PaTP. Transcripts were labeled with [ $\gamma$ - $^{32}$ P]GTP. **(d)** Gel-mobility shift assays for the analysis of biotinylated transcripts. The 152-mer RNAs transcribed with Bio-PaTP, from the original and 20-cycle PCR-amplified templates, DNA6 or DNAcont2, were mixed with streptavidin. Biotinylated RNA-streptavidin complexes were separated from free RNAs on a 7% polyacrylamide-7 M urea gel, and the amounts of the complexes (yields) were determined from the band intensities. **(e)** Gel-mobility shift assays for the analysis of biotinylated transcripts using Bio-UTP and the original natural templates, DNAcont3 and DNAcont4. The 152-mer transcripts were mixed with streptavidin. Biotinylated RNA-streptavidin complexes were analyzed on the gel. **(f)** Sequence analysis of the 152-mer transcripts containing Bio-Pa or adenine at position 59. The 5'  $^{32}$ P-labeled transcripts were partially digested with RNase T1 (T1) or with alkali (AL). A portion of the partially alkali-digested transcripts was treated with streptavidin magnetic beads to capture RNA fragments containing Bio-Pa (AL+SA). Each digested fragment was analyzed on a 10% polyacrylamide-7 M urea gel.

estimated by the sequencing analysis (Supplementary Fig. 1). Notably, Bio-Pa misincorporation opposite the natural bases did not increase in transcription reactions using the 20-cycle PCR-amplified DNAcont2 template (9 and 12% in the presence of 2 and 4 mM Bio-Pa, respectively; Fig. 4d). Therefore, the misincorporation of Ds and Pa opposite the natural bases is extremely low during PCR amplification.

We compared the selectivity of the Ds-Pa pair in T7 transcription with that of the natural A·T(U) pair, and found that the misincorporation percentages of Bio-PaTP opposite the natural bases (9 and 16% in Fig. 4d, respectively) were lower than those of biotin-linked UTP (Bio-UTP) opposite the natural bases (Fig. 4e). To assess the misincorporation of Bio-UTP opposite G, C and T, we synthesized control DNA templates containing only one adenine (DNAcont3) or no adenines (DNAcont4) in the coding regions of the template strands. We transcribed DNAcont3 or DNAcont4 under the same conditions as those for the Ds-Pa-containing templates, but with 0 or 2 mM of Bio-UTP and the other natural substrates (2 mM ATP, GTP and CTP). The transcripts from DNAcont3 had 100% incorporation of Bio-U opposite A, but 21% of transcripts from DNAcont4 were biotinylated by Bio-U misincorporation opposite G, C and T (Fig. 4e). This misincorporation rate was higher than that of Bio-Pa opposite the natural bases, even in the presence of 2:1 ratio of Bio-PaTP (4 mM) to the natural substrates (2 mM). Thus, in T7 transcription, the selective

exclusion of Bio-Pa misincorporation at the natural base sites was higher than that of Bio-U misincorporation opposite G, C and T, although the incorporation of Bio-Pa opposite Ds was slightly less efficient than that of Bio-U opposite adenine.

To determine the incorporation site of Bio-Pa in the transcripts, we sequenced the biotinylated transcripts (Fig. 4f). If Bio-Pa was precisely incorporated into the transcripts opposite Ds, then position 59 in the transcripts would be biotinylated. We partially digested the 5'  $^{32}$ P-labeled transcripts, by alkali or RNase T1. For the transcripts containing Bio-Pa the bands corresponding to fragments larger than a 59-mer were shifted on the sequencing ladders. We treated the alkali-digested fragments with streptavidin to capture the biotinylated fragments. As a result, fragments larger than a 59-mer almost disappeared, confirming that Bio-Pa was site-specifically incorporated at position 59 in the transcripts. Furthermore, the transcripts generated from the 20-cycle amplified templates had the same patterns as those obtained with the original templates.

## DISCUSSION

We have developed an expanded genetic system using an unnatural hydrophobic base pair for replication and transcription. In replication, we found that the  $\gamma$ -amidotriphosphates have slower incorporation rates, especially in noncognate pairings. For example, in the single-nucleotide insertion experiments, the  $V_{max}$  value

(2.2% min<sup>-1</sup>) of dATP<sub>N</sub> opposite Pa was 9.5-fold smaller than that (21% min<sup>-1</sup>) of dATP opposite Pa. In contrast, the  $V_{\max}$  value (12% min<sup>-1</sup>) of dDsTP<sub>N</sub> opposite Pa was 2.3-fold smaller than that (28% min<sup>-1</sup>) of dDsTP opposite Pa (**Supplementary Table 1**). The slower reaction rates of the noncognate pairings between  $\gamma$ -amidotriphosphates and template bases might improve the proofreading by the exonuclease activity of the polymerases, thus facilitating the high selectivity of this unnatural base pair in PCR amplification. Thus, the  $\gamma$ -modified triphosphates might be useful as DNA polymerase substrates for recognizing the correct shape-complementarity between pairing bases. Recently, other  $\gamma$ -modified triphosphates, including  $\gamma$ -P-aminonaphthalene-5-sulfonate triphosphates, had been reported to improve the fidelity of a reverse transcriptase<sup>28</sup>. These findings indicate that the recognition of the  $\gamma$ -modified triphosphates by DNA polymerases requires more accurate geometrical fitting between the pairing bases than that of the usual triphosphates. The combination of the  $\gamma$ -amidotriphosphates and the polymerases with exonuclease activity would be applicable to other unnatural base pairs. In this regard, the recent evolution of polymerases could also reinforce the unnatural base pair system<sup>29,30</sup>.

We found that the shape-complementary, hydrophobic Ds-Pa pair also functions in transcription. The site-specific biotinylation made possible by the Ds-Pa system would be useful for the immobilization of RNA molecules without any loss of their activity<sup>13</sup>. Since such a large modified-Pa can be efficiently incorporated into RNA, other functional, 4-modified Pa bases, such as a fluorophore-linked Pa, could also be introduced into RNA. The DNA templates containing the Ds-Pa pair can be amplified by PCR, facilitating large-scale transcription. Thus, the Ds-Pa pair system is a powerful tool for creating nucleic acids bearing extra components at desired positions. Furthermore, the hydrophobic Ds-Pa pair could be used in combination with unnatural hydrogen-bonding base pairs, such as isoG · isoC<sup>6,9</sup> and s ·  $\gamma$  pairs<sup>12–14</sup>. Such a system with multiple unnatural base pairs could be used to introduce several different components into nucleic acids.

## METHODS

**PCR amplification.** We performed PCR (optimized conditions) in 20 mM Tris-HCl buffer (pH 8.8), with 10 mM KCl, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 2 mM MgSO<sub>4</sub>, 0.1% Triton X-100, 0.3 mM each dNTP (N = Pa, G, C and T) and dNTP<sub>N</sub> (N = Ds and A), 1  $\mu$ M each Primer1 (5'-GAAATTAATACGACTCACTATAGGG-3') and Primer2 (5'-TTTCACACAGGAAACAGCTATGAC-3'), 2.3–2.4 nM double-stranded DNA fragment and 0.04 unit/ $\mu$ l Vent DNA polymerase (NEB). For the conventional conditions, we instead used 0.2 mM of each natural dNTP and 0.01 unit/ $\mu$ l of polymerase. For the sequencing analysis, we purified the PCR products by gel electrophoresis or filtration using Microcon YM-30 and Micropure-EZ (Millipore). For DNA template preparation, we carried out primer extension in 10 mM Tris-HCl (pH 7.5) buffer, containing 7 mM MgCl<sub>2</sub> and 0.1 mM DTT.

**Site-specific biotinylation of RNA (152-mer).** We performed transcription (20- $\mu$ l reactions) for 6 h at 37 °C, using 3  $\mu$ Ci [ $\gamma$ -<sup>32</sup>P]GTP, 2 mM each natural NTP, 0–4 mM Bio-PaTP and 30 nM template (the original and PCR-amplified DNA6 and DNAcont2), in 40 mM Tris-HCl buffer (pH 8.0) containing 24 mM MgCl<sub>2</sub>, 2 mM spermidine, 5 mM DTT and 0.01% Triton

X-100. For the control reactions using the original DNAcont3 and DNAcont4, we carried out transcription only with 2 mM ATP, GTP and CTP in the presence or absence of 2 mM biotin-16-UTP (Roche). We analyzed and purified the products on a gel. We detected the biotinylated transcripts by the gel mobility-shift assay using streptavidin (Promega). We incubated the mixture (10- $\mu$ l volume) of 2 pmol <sup>32</sup>P-labeled transcripts and 100 pmol streptavidin for 1 h at 20 °C, in 10 mM Tris-HCl buffer (pH 7.6) containing 50 mM NaCl and 10 mM EDTA. We analyzed the biotinylated RNA–streptavidin complexes by gel electrophoresis. To determine the Bio-Pa insertion position, we labeled each RNA (152-mer) with [ $\gamma$ -<sup>32</sup>P]ATP (Perkin Elmer) after 5'-dephosphorylation by calf intestinal alkaline phosphatase (CIAP; Takara). We partially digested the labeled RNAs by RNase T1 (0.17 U/ $\mu$ l) for 12 min at 55 °C, in 13.3 mM sodium citrate buffer (pH 5.0) containing 4.7 M urea, 0.7 mM EDTA and 0.17 mg/ml *E. coli* tRNA, and by alkali for 15 min at 90 °C, in 32 mM sodium carbonate buffer (pH 9.1) containing 0.6 mM EDTA. We mixed 9  $\mu$ l of the alkali-digested RNA solution with 11  $\mu$ l of 20 mM Tris-HCl buffer (pH 7.6) containing 150 mM NaCl, and incubated the 10- $\mu$ l mixtures with 0.4 mg of Streptavidin magnetic beads (NEB) for 5 min at room temperature (18–25 °C). We analyzed 5- $\mu$ l aliquots of the supernatant and the other digested samples on a 10% polyacrylamide–7 M urea gel.

**Additional information.** Chemical synthesis, single-nucleotide insertion, DNA sequencing, PCR amplification and transcription are described in **Supplementary Methods**.

*Note: Supplementary information is available on the Nature Methods website.*

## ACKNOWLEDGMENTS

This work was funded by the RIKEN Structural Genomics/Proteomics Initiative, the National Project on Protein Structural and Functional Analyses, Ministry of Education, Culture, Sports, Science and Technology of Japan, by a Grant-in-Aid for Scientific Research (KAKENHI 15350097), and by the Program of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NIBIO).

## AUTHOR CONTRIBUTIONS

I.H. conceived and designed the study, supervised the work and prepared samples; M.K. contributed to the biological idea and performed experiments; T.M., T.F. and A.S. performed chemical synthesis; R.K. and Y.H. performed biological experiments; S.Y. conceived the study and supervised the work.

## COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Methods* website for details).

Published online at <http://www.nature.com/naturemethods/>  
Reprints and permissions information is available online at  
<http://npg.nature.com/reprintsandpermissions/>

1. Benner, S.A., Burgstaller, P., Battersby, T.R. & Jurczyk, S. Did the RNA world exploit an expanded genetic alphabet? in *The RNA World* (eds. Gesteland, R.F., Cech, T.R. & Atkins, J.F.) 163–181 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999).
2. Henry, A.A. & Romesberg, F.E. Beyond A, C, G and T: augmenting nature's alphabet. *Curr. Opin. Chem. Biol.* **7**, 727–733 (2003).
3. Moser, M.J. & Prudent, J.R. Enzymatic repair of an expanded genetic information system. *Nucleic Acids Res.* **31**, 5048–5053 (2003).
4. Piccirilli, J.A., Krauch, T., Moroney, S.E. & Benner, S.A. Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature* **343**, 33–37 (1990).
5. Switzer, C.Y., Moroney, S.E. & Benner, S.A. Enzymatic recognition of the base pair between isocytidine and isoguanosine. *Biochemistry* **32**, 10489–10496 (1993).

6. Sismour, A.M. *et al.* PCR amplification of DNA containing non-standard base pairs by variants of reverse transcriptase from Human Immunodeficiency Virus-1. *Nucleic Acids Res.* **32**, 728–735 (2004).
7. Johnson, S.C., Sherrill, C.B., Marshall, D.J., Moser, M.J. & Prudent, J.R. A third base pair for the polymerase chain reaction: inserting isoC and isoG. *Nucleic Acids Res.* **32**, 1937–1941 (2004).
8. Sismour, A.M. & Benner, S.A. The use of thymidine analogs to improve the replication of an extra DNA base pair: a synthetic biological system. *Nucleic Acids Res.* **33**, 5640–5646 (2005).
9. Tor, Y. & Dervan, P.B. Site-specific enzymatic incorporation of an unnatural base, N<sup>6</sup>-(6-aminoethyl)isoguanosine, into RNA. *J. Am. Chem. Soc.* **115**, 4461–4467 (1993).
10. Fujiwara, T., Kimoto, M., Sugiyama, H., Hirao, I. & Yokoyama, S. Synthesis of 6-(2-thienyl)purine nucleoside derivatives that form unnatural base pairs with pyridin-2-one nucleosides. *Bioorg. Med. Chem. Lett.* **11**, 2221–2223 (2001).
11. Hirao, I. *et al.* An unnatural base pair for incorporating amino acid analogs into proteins. *Nat. Biotechnol.* **20**, 177–182 (2002).
12. Kimoto, M. *et al.* Site-specific incorporation of a photo-crosslinking component into RNA by T7 transcription mediated by unnatural base pairs. *Chem. Biol.* **11**, 47–55 (2004).
13. Moriyama, K., Kimoto, M., Mitsui, T., Yokoyama, S. & Hirao, I. Site-specific biotinylation of RNA molecules by transcription using unnatural base pairs. *Nucleic Acids Res.* **33**, e129 (2005).
14. Kawai, R. *et al.* Site-specific fluorescent labeling of RNA molecules by specific transcription using unnatural base pairs. *J. Am. Chem. Soc.* **127**, 17286–17295 (2005).
15. Morales, J.C. & Kool, E.T. Efficient replication between non-hydrogen-bonded nucleoside shape analogs. *Nat. Struct. Biol.* **5**, 950–954 (1998).
16. McMinn, D.L. *et al.* Efforts toward expansion of the genetic alphabet: DNA polymerase recognition of a highly stable, self-pairing hydrophobic base. *J. Am. Chem. Soc.* **121**, 11585–11586 (1999).
17. Wu, Y. *et al.* Efforts toward expansion of the genetic alphabet: optimization of interbase hydrophobic interactions. *J. Am. Chem. Soc.* **122**, 7621–7632 (2000).
18. Matray, T.J. & Kool, E.T. A specific partner for abasic damage in DNA. *Nature* **399**, 704–708 (1999).
19. Doublé, S., Tabor, S., Long, A.M., Richardson, C.C. & Elenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, 251–258 (1998).
20. Kiefer, J.R., Mao, C., Braman, J.C. & Beese, L.S. Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature* **391**, 304–307 (1998).
21. Mitsui, T. *et al.* An unnatural hydrophobic base pair with shape complementarity between pyrrole-2-carbaldehyde and 9-methylimidazo[(4,5)-b]pyridine. *J. Am. Chem. Soc.* **125**, 5298–5307 (2003).
22. Morales, J.C. & Kool, E.T. Minor groove interactions between polymerase and DNA: More essential to replication than Watson-Crick hydrogen bonds? *J. Am. Chem. Soc.* **121**, 2323–2324 (1999).
23. Hirao, I. *et al.* A two-unnatural-base-pair system toward the expansion of the genetic code. *J. Am. Chem. Soc.* **126**, 13298–13305 (2004).
24. Petruska, J. *et al.* Comparison between DNA melting thermodynamics and DNA polymerase fidelity. *Proc. Natl. Acad. Sci. USA* **85**, 6252–6256 (1988).
25. Goodman, M.F., Creighton, S., Bloom, L.B. & Petruska, J. Biochemical basis of DNA replication fidelity. *Crit. Rev. Biochem. Mol. Biol.* **28**, 83–126 (1993).
26. Kimoto, M., Yokoyama, S. & Hirao, I. A quantitative, non-radioactive single-nucleotide insertion assay for analysis of DNA replication fidelity by using an automated DNA sequencer. *Biotechnol. Lett.* **26**, 999–1005 (2004).
27. Mitsui, T., Kimoto, M., Sato, A., Yokoyama, S. & Hirao, I. An unnatural hydrophobic base, 4-propynylpyrrole-2-carbaldehyde, as an efficient pairing partner of 9-methylimidazo[(4,5)-b]pyridine. *Bioorg. Med. Chem. Lett.* **13**, 4515–4518 (2003).
28. Mulder, B.A. *et al.* Nucleotide modification at the  $\gamma$ -phosphate leads to the improved fidelity of HIV-1 reverse transcriptase. *Nucleic Acids Res.* **33**, 4865–4873 (2005).
29. Leconte, A.M., Chen, L. & Romesberg, F.E. Polymerase evolution: Efforts toward expansion of the genetic code. *J. Am. Chem. Soc.* **127**, 12470–12471 (2005).
30. Chelliserykattil, J. & Ellington, A.D. Evolution of a T7 RNAP variant that transcribes 2'-O-methyl RNA. *Nat. Biotechnol.* **22**, 1155–1160 (2004).