Published online 24 November 2011 *Nucleic Acids Research, 2012, Vol. 40, No. 6* ***2793–2806***
doi:10.1093/nar/gkr1068

Highly specific unnatural base pair systems as a third base pair for PCR amplification

Rie Yamashige[1], Michiko Kimoto[1,2], Yusuke Takezawa[1], Akira Sato[1], Tsuneo Mitsui[2], Shigeyuki Yokoyama[1,3] and Ichiro Hirao[1,2,*]

[1]RIKEN Systems and Structural Biology Center (SSBC), 1-7-22 Suehiro-cho, [2]TagCyx Biotechnologies, 1-6-126 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045 and [3]Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Received September 28, 2011; Revised October 28, 2011; Accepted October 30, 2011

ABSTRACT

Toward the expansion of the genetic alphabet of DNA, we present highly efficient unnatural base pair systems as an artificial third base pair for PCR. Hydrophobic unnatural base pair systems between 7-(2-thienyl)imidazo[4,5-*b*]pyridine (Ds) and 2-nitro-4-propynylpyrrole (Px) were fine-tuned for efficient PCR, by assessing the amplification efficiency and fidelity using different polymerases and template sequence contexts and modified Px bases. Then, we found that some modifications of the Px base reduced the misincorporation rate of the unnatural base substrates opposite the natural bases in templates without reducing the Ds–Px pairing selectivity. Under optimized conditions using Deep Vent DNA polymerase, the misincorporation rate was extremely low (0.005%/bp/replication), which is close to that of the natural base mispairings by the polymerase. DNA fragments with different sequence contexts were amplified $\sim 10^{10}$-fold by 40 cycles of PCR, and the selectivity of the Ds–Px pairing was >99.9%/replication, except for 99.77%/replication for unfavorable purine-Ds-purine motifs. Furthermore, >97% of the Ds–Px pair in DNA survived in the 10^{28}-fold amplified products after 100-cycle PCR (10 cycles repeated 10 times). This highly specific Ds–Px pair system provides a framework for new biotechnology.

INTRODUCTION

Expansion of the genetic alphabet of DNA with artificial extra base pairs (unnatural base pairs) would have great potential for advancing the present genetic recombination systems to next generation biotechnologies. The genetic expansion by unnatural base pair systems could facilitate the site-specific incorporation of extra components into nucleic acids and proteins at desired positions, through the genetic information flow process comprising replication, transcription and translation (1,2). Researchers have chemically synthesized many unnatural base pairs and tested their abilities, and some of them function as a third base pair in *in vitro* biological processes (3–9). In addition, the attachment of functional groups to unnatural bases enables the site-specific functionalization of DNA and RNA molecules by replication and transcription (10–16). The capabilities of several unnatural base pairs have been demonstrated, and they have been utilized in diagnostics for target DNA detection and in basic research for site-specific labeling of nucleic acid molecules (9,17–21). For example, an unnatural base pair between isoguanine and isocytosine (22) is available as Plexor, in a multiplex real-time quantitative PCR detection platform (23–25).

Toward the next generation of biotechnology, the development of unnatural base pairs for specific PCR amplification is a seminal milestone, to allow the large-scale amplification of synthetic DNA fragments containing unnatural base pairs as a supply of DNA templates for use in further biological processes. Since the amplified DNA fragments are reused as templates in the following PCR amplification, high fidelity of the cognate base pairing is essential. If the selectivity of an unnatural base pairing, which is indicated by the incorporation selectivity (%) of the unnatural base substrates opposite their pairing partners in templates, is 99.9%/replication, then the amplified DNA would retain 96% of the unnatural base pair after 40 cycles of PCR ($0.999^{40} = 0.961$) in a theoretical exponential amplification. However, 99.0% selectivity of the unnatural base pairing would reduce the retention to 67% in the amplified products after 40 cycles of PCR ($0.990^{40} = 0.669$). Furthermore, for the high fidelity of unnatural base pairing, extremely low misincorporation rates

*To whom correspondence should be addressed. Tel: +81 45 503 9644; Fax: +81 45 503 9645; Email: ihirao@riken.jp

of unnatural base substrates opposite the natural bases in templates are also required. Even if the misincorporation rate is 0.10%/natural bp/replication, up to 33% of the natural base pairs in a 10-base sequence flanked by PCR primers would be mutated to unnatural base pairs after 40 cycles ($0.33 = 1 - [(1 - 0.10/100)^{10}]^{40}$) in a theoretical exponential amplification. Thus, for the further development of unnatural base pair systems in PCR, the base pairing fidelity of both the high incorporation selectivity in cognate base pairings and the low misincorporation rate of unnatural bases opposite the natural bases, as well as the high amplification efficiency, are required to facilitate the advancement of the new biotechnology.

So far, various unnatural base pairs that function in PCR with 96–99.8% incorporation selectivity per replication or per PCR cycle have been reported (26–35), although the determination of the fidelity of unnatural base pairing depends on each researcher's method. Among them, we developed a hydrophobic unnatural base pair between 7-(2-thienyl)-imidazo[4,5-*b*]pyridine (**Ds**) and 2-nitro-4-propynylpyrrole (**Px**) for efficient PCR amplification (32). The **Ds–Px** pair was designed by strict shape-complementarity of the pairing surface, but both shapes are different from those of the natural bases. In addition, the **Ds–Px** pair lacks notable hydrogen bonds between the pairing bases, and thus the hydrophobic bases effectively exclude the non-cognate pairing with the natural bases. Furthermore, the **Px** base has the nitro group to electrostatically repel the mispairing with A (31). The oxygen of the nitro group of **Px** and the nitrogen of **Ds**, corresponding to the N3 position of the natural purine bases, are also important as proton acceptor residues, to interact with polymerases for efficient replication. By modifying a conventional sequencing method for the unnatural base pair, we determined the incorporation selectivity (>99%/replication) of the **Ds–Px** pairing in PCR with Deep Vent DNA polymerase, possessing a $3' \rightarrow 5'$ exonuclease activity (exo^+). In addition, we identified the efficiencies of the template sequence contexts of the natural bases around the unnatural base in PCR amplification. For example, *in vitro* selection experiments using a DNA library containing the unnatural base pair (32) and primer extension experiments using templates all combinations of 16 sequences of 3′-NDsN-5′ (N = A, G, C or T) revealed that the less efficient sequences contain 5′-Y**Px**Y-3′/3′-R**Ds**R-5′ patterns (Morohashi, Kimoto, and Hirao, unpublished data), such as 5′-GTT**Px**CAT-3′/3′-CAA**Ds**GTA-5′ (32), in PCR amplification. However, the exact differences among these sequence variants, in terms of the efficiency and fidelity of the **Ds–Px** pairing in PCR, still require clarification, and its PCR conditions can be further optimized. Through the process of unnatural base pair development, we have expected to discover other potentially important parameters of the **Ds–Px** pair in PCR amplification.

Here, we present highly efficient, high-fidelity PCR systems involving the **Ds–Px** pair with optimized conditions for a variety of template sequence contexts. For this, we developed accurate determination methods for the fidelity of both the incorporation selectivity of the **Ds–Px** pairing and the misincorporation rates of the natural base replacement with the unnatural bases in DNA templates during PCR, as well as the amplification efficiency, by repeating 10 cycles of PCR to stay within exponential amplification. The **Ds–Px** pair survived through 100 cycles of PCR, and after 100 cycles of PCR (10 cycles of PCR repeated 10 times) under the optimized conditions, the DNA fragment containing the **Ds–Px** pair was amplified 10^{28}-fold, and >97% of the **Ds–Px** pair was retained in the amplified products. Accordingly, the unnatural base pair selectivity was as high as 99.97%/replication. Furthermore, we found that some functional groups attached to the **Px** base reduced the misincorporation rates of the unnatural base substrate opposite the natural bases in templates, without reducing the amplification efficiency and the incorporation selectivity of the unnatural base pairing between **Ds** and modified **Px**. This **Ds–Px** pair system could be practically used for future genetic alphabet expansion technologies.

MATERIALS AND METHODS

General

DNA templates containing **Ds** were synthesized with an Applied Biosystems 392 DNA synthesizer, using CE phosphoramidite reagents for the natural and **Ds** bases (Glen Research). DNA fragments comprising only the natural bases were synthesized or purchased from Invitrogen. The primer labeled with FAM at the 5′-end was purchased from Invitrogen. DNA fragments were purified by gel electrophoresis. The syntheses of d**Ds**TP, NH_2-hx-d**Px**TP and Cy5-hx-d**Px**TP was described previously (30,32,36), and the syntheses of the other modified-d**Px**TPs are described in the Supplementary Data section. The following thermostable DNA polymerases were purchased: Deep Vent (exo^+), Deep Vent (exo^-), Vent (exo^+), N9°, Phusion HF and Taq DNA pols from New England Biolabs; AccuPrime *Pfx*™ and Pfx50 DNA pols from Invitrogen, Pfu DNA pol from Promega; Pwo SY DNA pol from Roche; TITANIUM *Taq* DNA pol from Clontech.

PCR involving the Ds–Px pairing

PCR amplification (25 µl) of the 55-mer DNA fragments (5′-TTTCACACAGGAAACAGCTATGACGG-N_7-CCCTATAGTGAGTCGTATTATC-3′; N_7 = 5′-ATC**Ds**TAT-3′ for **Ds**-temp 1, 5′-CCC**Ds**TTG-3′ for **Ds**-temp 2, 5′-TAC**Ds**GTG-3′ for **Ds**-temp 3, 5′-ATG**Ds**AAC-3′ for **Ds**-temp 4 and 5′-ATCCATT-3′ for Cont-temp) was performed by using 1 µM each 5′-primer (40-mer; 5′-CGTTGTAAAACGACGGCCAGGATAATACGACTCACTATAG-3′) and 3′-primer (24-mer; 5′-TTTCACACAGGAAACAGCTATGAC-3′) and each DNA pol at the manufacturer's recommended concentration (Deep Vent (exo^+), 0.5 U; Deep Vent (exo^-), 0.5 U; AccuPrime *Pfx*, 1.25 U; Pfx50, 2.5 U; Phusion HF, 0.5 U; Pfu, 0.625 U; Pwo SY, 1.25 U; Vent (exo^+), 0.5 U; N9°, 0.5 U; TITANIUM *Taq*, 1×; Taq, 0.5 U), in the 1× reaction buffer accompanying each DNA pol, supplemented with 50 µM each d**Ds**TP and

modified-d**Px**TP, and 300 μM natural dNTPs, unless otherwise indicated. As for PCR with the AccuPrime *Pfx* DNA pol with final concentrations of 400 μM natural dNTPs, 100 μM natural dNTPs and 0.5 mM $MgSO_4$ were further added to the 1× AccuPrime *Pfx* reaction mix, originally containing 300 μM natural dNTPs and 1 mM $MgSO_4$. As for PCR with Deep Vent DNA pol (exo^+) with final concentrations of 500 μM natural dNTPs, 2 mM $MgSO_4$ was further added to the 1× ThermoPol reaction buffer originally containing 2 mM $MgSO_4$. PCR conditions were 30 s at 94°C, 30 s at 45°C and 4 min at 65°C/cycle.

To screen DNA pols, we performed a 15-cycle PCR, by using 0.5 nM 55-mer DNA template and FAM-labeled 3′-primer. The amplified PCR products, together with a FAM-labeled 3′-primer (1 μM) as a marker, were analyzed by denaturing 15% polyacrylamide gel electrophoresis (PAGE). The full-length 75-mer DNA fragments on the gel were quantified by a bio-imaging analyzer LAS-4000 (FUJIFILM) in the SYBR-mode, and by comparing the band intensities with that of the marker, the amplification folds were determined. For the sequencing analysis of the amplified fragments to determine the retention rates of the **Ds–Px** pair, as described below, PCR was performed by using non-labeled primers, and the full-length PCR products were purified by denaturing 8% PAGE.

In the 40- and 100-cycle PCR amplifications, we repeated a 10-cycle PCR 4 or 10 times, to maintain exponential amplification conditions. In the first 10-cycle PCR, we used 2 nM DNA template, and in the following 10-cycle PCR amplifications, the amplified products were diluted 80- or 150-fold (5 μl) into a fresh PCR reaction mix (final dilutions: ×750 for PCR with **Ds**-temp 1, **Ds**-temp 3, and Cont-temp and ×400 for PCR with **Ds**-temp 4 for all substrates, except for ×400 for PCR with **Ds**-temp 1 for BPh-d**Px**TP or HBP-hx-d**Px**TP). To determine the fold amplification, the amplified PCR products after each 10-cycle PCR were analyzed by denaturing 15% PAGE, and the full-length 75-mer DNA fragments on the gel were quantified by the LAS-4000 analyzer after SYBR Green II staining. To determine the amplification fidelity, the amplified products after 10, 20, 40, 70 and 100 cycles of PCR were fractionated on an 8% denaturing gel, followed by an analysis of the retention rates of the unnatural base by sequencing (for **Ds**-temps 1, 3 and 4) and the misincorporation rate of the unnatural bases opposite the natural bases (for Cont-temp), as described below.

DNA sequencing

The cycle sequencing reaction (20 μl) was performed with the Cycle Sequencing mix (8 μl) from a BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems), containing 0.3 pmol of DNA fragments and 4 pmol of Sequencing primer (20-mer; 5′-CGTTGTAAAA CGACGGCCAG-3′), in the presence of 4-propynyl-1-(β-D-ribofuranosyl)pyrrole-2-carbaldehyde 5′-triphosphate (d**Pa**′TP) (40 pmol for **Ds**-temp 1 and 1 nmol for **Ds**-temp 2 and **Ds**-temp 3) or 1 nmol dd**Pa**′TP. The sequencing cycling parameters were 25 cycles of 10 s at 96°C, 5 s at 50°C and 4 min at 60°C. The residual dye terminator was removed by passage through a CENTRI-SEP column (Princeton Separations), and the solutions were dried. The residues were resuspended in 4 μl formamide solution and analyzed by an ABI 377 DNA Sequencer, using a 6% polyacrylamide-6 M urea gel. The sequence data were analyzed with the Applied Biosystems PRISM sequencing analysis v3.2 software. The retention rates of the **Ds–Px** pair in the amplified PCR products were determined by comparing the sequencing patterns with the authentic sequencing patterns (see Supplementary Figures S1–S15).

Real-time qPCR

The PCR reaction (25 μl) was performed in 1× ThermoPol reaction buffer with 300 μM each natural dNTP, 50 μM d**Ds**TP and NH_2-hx-d**Px**TP, 1 μM each 5′-primer and 3′-primer, 30 000-fold diluted SYBR Green I (Lonza), 1000-fold diluted ROX (Invitrogen) as a reference dye, 0.5 unit of Deep Vent (exo^+) DNA pol, and DNA template (**Ds**-temp 1, **Ds**-temp 4 or Cont-temp) at different concentrations (final concentrations: 20 pM, 2 pM, 200 fM and 20 fM) on a Mx3005P Real-time system (Agilent Technologies). Data collection was performed with the detection filters for SYBR and ROX, and the data were analyzed with the Mx3005P v4.10 software. The cycle thresholds were automatically set by the software, which gave the cycle threshold values (Ct).

Analysis of misincorporation rates of unnatural bases into natural bases

We first performed a 40- or 100-cycle PCR in the presence of modified-d**Px**TP by using Cont-temp as described earlier (first PCR), and the full-length amplified products after 10, 20, 40, 70 and 100 cycles of PCR were purified by denaturing 8% PAGE. The purified products (0.25 pmol) were subjected to seven extra cycles of PCR with the same concentration of the same DNA polymerase, but with Cy5-hx-d**Px**TP, instead of the modified-d**Px**TP (second PCR), to detect the modified **Px** base positions in the amplified DNA fragments. The conditions used in the second PCR were 1 μM each 5′-primer and FAM-labeled 3′-primer, 50 μM d**Ds**TP, 5 μM Cy5-hx-d**Px**TP and 300 μM natural dNTPs and 2 mM $MgSO_4$ for the Deep Vent DNA pol (exo^+) or 400 μM natural dNTPs and 1.5 mM $MgSO_4$ for the AccuPrime *Pfx* DNA pol. We also performed a 7-cycle PCR directly by using 10 nM (0.25 pmol) **Ds**-temp 1 and Cont-temp, to normalize the correct incorporation of Cy5-hx-d**Px**TP opposite **Ds** in the template during the second PCR and to subtract the background misincorporation of Cy5-hx-d**Px**TP during the second PCR, respectively. Under these conditions, we assumed that almost all of the **Ds–Px** pairs were retained after PCR amplification, since we confirmed the complete retention of the **Ds–Px** pair by a sequencing analysis of the amplified products of the 55-mer DNA template in the second PCR (data not shown). The second PCR products, together with a 75-mer single-stranded DNA (1 μM) as a marker, were analyzed by 15% denaturing PAGE. The fluorescent intensity of the full-length 75-mer DNA fragments on the gel

was quantified with a bio-imager, FLA-7000 (GE-Healthcare) in the Cy5-mode and FAM-mode.

Misincorporations of the unnatural bases into the first PCR-amplified DNAs were calculated as follows. First, the amounts of the full-length amplified products for each sample ($N_{t[sample]}$), as double-stranded DNA, were determined from the FAM intensity. Next, the Cy5 intensity for each sample ($I_{Cy5[sample]}$) was normalized to the amounts of the DNA ($I_{Cy5[sample]}/N_{t[sample]}$). The background misincorporation ratio was then subtracted from the relative Cy5 intensity ($I_{Cy5[sample]}/N_{t[sample]} - I_{Cy5[Cont\text{-}temp]}/N_{t[Cont\text{-}temp]}$). Finally, each sample's misincorporation rate (M%) of the unnatural bases into the total amplified DNAs was calculated by the following formula: $M(\%) = (I_{Cy5[sample]}/N_{t[sample]} - I_{Cy5[Cont\text{-}temp]}/N_{t[Cont\text{-}temp]})/(I_{Cy5[\mathbf{Ds}\text{-}temp\ 1]}/N_{t[Ds\text{-}temp\ 1]} - I_{Cy5[Cont\text{-}temp]}/N_{t[Cont\text{-}temp]}) \times 100$. We performed three independent experiments, and the obtained values were averaged. To calculate the misincorporation rates per bp per replication (F%), the misincorporation rates (M%) were plotted over the replication times [$=\log_2$ (Amplification fold)], and were fitted by linear regression with a high correlation coefficient ($R > 0.99$). The slope (S%) of the regression line corresponds to the averaged misincorporation rate per replication (%), and the misincorporation rate per bp per replication (F%) was calculated by the following formula: $F = [1 - (1 - S/100)^{1/11}] \times 100$.

RESULTS

PCR conditions for the Ds–Px pairing

We first examined the PCR conditions by using several DNA polymerases (DNA pols) and unnatural and natural base substrate combinations of d**Ds**TP, NH_2-hx-d**Px**TP (Figure 1) and natural base dNTPs at different concentrations. In addition, to determine the natural base sequence dependency around the unnatural base pair, we used four single-stranded **Ds**-containing DNA templates (55-mer, **Ds**-temps 1, 2, 3 and 4), as well as a control template (Cont-temp) comprising only the natural bases (Figure 2A). Our previous studies indicated that **Ds**-temp 1 containing 3′-TAT**Ds**CTA-5′, **Ds**-temp 2

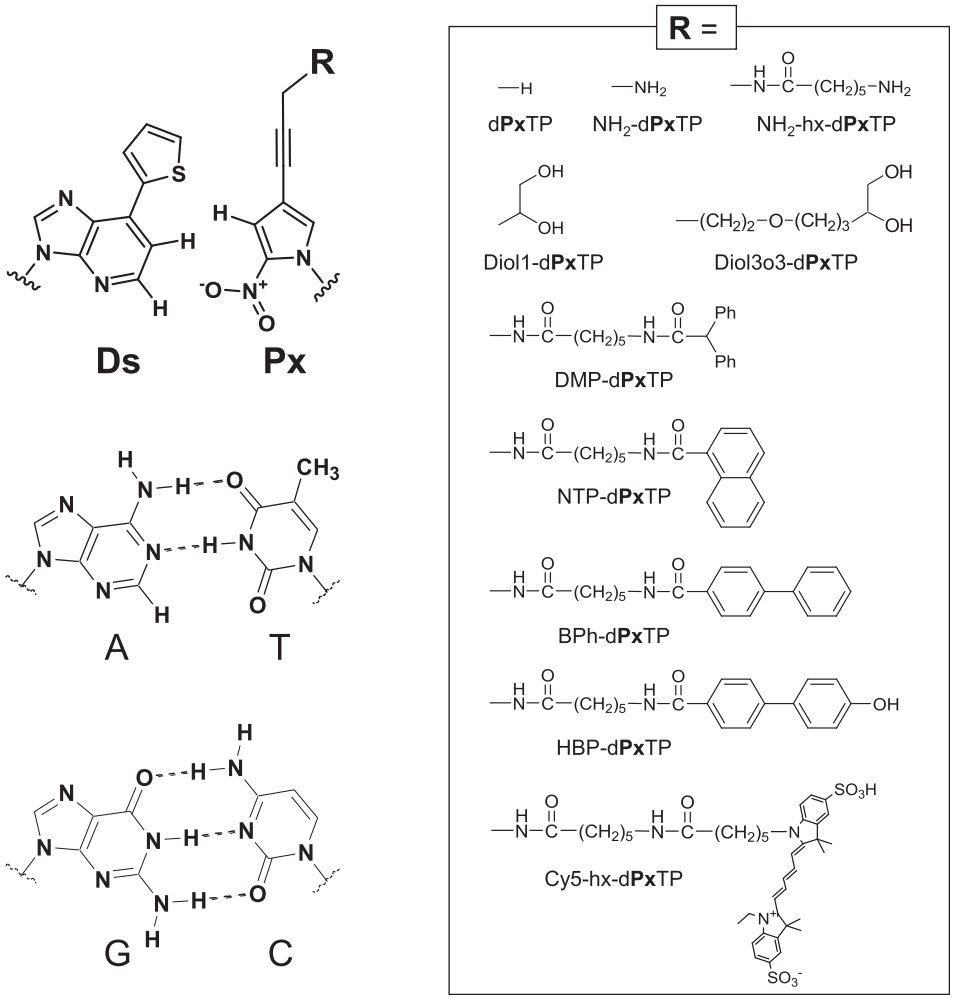


Figure 1. Chemical structures of the unnatural **Ds–Px** pair and the natural A–T and G–C pairs. Functional groups (**R**), attached to the **Px** base via the propynyl linker, used in this study are summarized in the enclosed box.

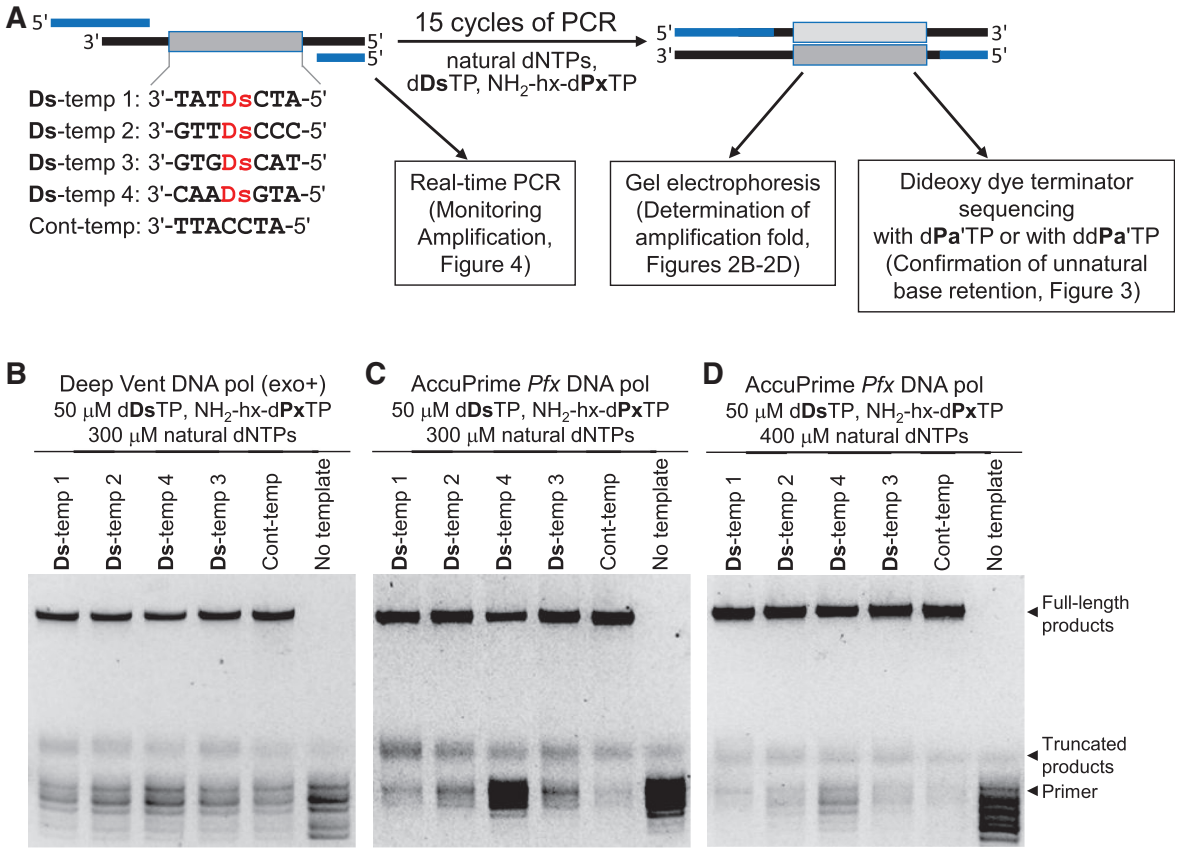


Figure 2. PCR amplification involving the **Ds–Px** pairing with various DNA templates containing Ds for the determination of PCR conditions. (**A**) Scheme for PCR amplification experiments using 55-mer DNA fragments containing one **Ds** base, in the presence of d**Ds**TP, NH_2-hx-d**Px**TP and natural dNTPs. (**B–D**) Analyses by denaturing gel electrophoresis of PCR-amplified products after 15 cycles of PCR under different conditions by Deep Vent DNA pol (exo^+) (B) and AccuPrime *Pfx* DNA pol (C,D), for the determination of the fold amplification at the end point.

containing 3′-GTT**Ds**CCC-5′, and **Ds**-temp 3 containing 3′-GTG**Ds**CAT-5′ are favorable templates, but **Ds**-temp 4 containing 3′-CAA**Ds**GTA-5′ (purine-**Ds**-purine) is an unfavorable one for efficient PCR amplification involving the **Ds–Px** pair by Deep Vent DNA polymerase (exo^+) (32). In this study, we incorporated only one unnatural base into the DNA templates, to focus on the accurate characterization of the **Ds–Px** pairing. To reduce the chemical DNA synthesis error, we utilized 55-mer single-stranded DNA templates, and the total base lengths were extended by PCR using 40-mer and 24-mer PCR primers to facilitate DNA sequencing directly after PCR amplification, which converted the templates into double-stranded 75-mer DNA products. Throughout all of the experiments in the study, we employed a PCR cycle program involving 94°C for 30 s, 45°C for 30 s and 65°C for 4 min. After 15 cycles of PCR, the fold amplification of the PCR products (Supplementary Table S1) was determined by polyacrylamide gel electrophoresis (representative data are shown in Figure 2B–D). Besides the full-length PCR products, truncated products paused after the unnatural base incorporations were also observed on the gel. The unconsumed primers were partially degraded by the exonuclease activity of the polymerases during PCR.

The retention rates of the **Ds–Px** pair in the amplified DNA were determined by two sequencing methods of the **Ds**-containing DNA strands using conventional dideoxy dye-terminator sequencing, but either with a deoxyribonucleoside triphosphate of 4-propynylpyrrole-2-carbaldehyde (d**Pa**′TP), another pairing partner of **Ds**, or with its dideoxyribonucleoside triphosphate (dd**Pa**′TP) (30–32). Since no dye-terminator corresponding to the unnatural base is present in the sequencing, there is a gap at the unnatural base position on the sequencing pattern in the presence of d**Pa**′TP. In another sequencing method in the presence of dd**Pa**′TP, the sequencing reaction is stopped at the unnatural base position by incorporating dd**Pa**′TP opposite **Ds** in DNA templates, and thus the following base peaks disappear. However, if some mutation from an unnatural to natural base occurs at the original unnatural base position during PCR, then the height of each following base peak increases, depending on the mutation rate in the DNA samples. The sequencing with d**Pa**′TP is useful for confirming the unnatural base position, while that with dd**Pa**′TP is appropriate for determining the unnatural base retention rate at the original position in the templates, by comparing the peak heights on the sequencing patterns with those obtained using standard authentic

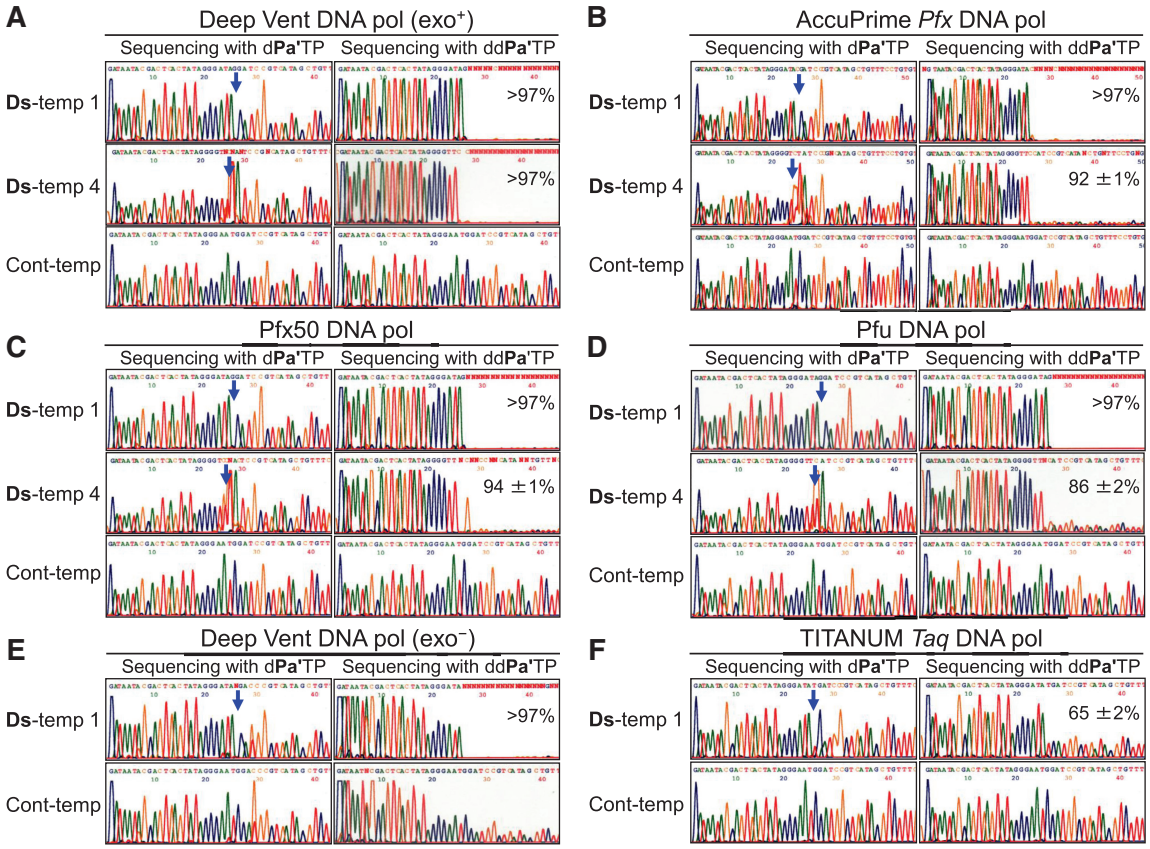


Figure 3. Sequencing analysis of the PCR products after 15 cycles of PCR by each DNA polymerase: (**A**) Deep Vent DNA pol (exo$^+$); (**B**) AccuPrime *Pfx* DNA pol; (**C**) Pfx50 DNA pol; (**D**) Pfu DNA pol; (**E**) Deep Vent DNA pol (exo$^-$); (**F**) TITANIUM *Taq* DNA pol. Sequencing reactions were performed in the presence of d**Pa**′TP (left panels) or dd**Pa**′TP (right panels) in each figure. The blue arrow indicates the original unnatural base position. The percentages indicated in the right panels for **Ds**-temp 1 and **Ds**-temp 4 are the retention rates of the **Ds–Px** pair in the amplified products after 15 cycles of PCR. The details of the calculation method for the retention rates are provided in the Supplementary Data section.

samples. We increased the determination accuracy of the retention rates by improving the peak height quantification (see Supplementary Figures S1 and S2).

Using the unnatural base substrate combination of d**Ds**TP and NH_2-hx-d**Px**TP, we screened 11 DNA polymerases, the Deep Vent (exo$^+$), Deep Vent (exo$^-$), AccuPrime *Pfx*, Pfx50, Phusion HF, Pfu, Pwo SY, Vent (exo$^+$), N9° DNA pols from family B, and the TITANIUM *Taq* and Taq DNA polymerases from family A, by 15 cycles of PCR with 50 μM d**Ds**TP and NH_2-hx-d**Px**TP and 300 μM natural dNTPs. Representative gel and sequencing patterns of the PCR products are shown in Figures 2 and 3, respectively, and all of the data for the amplification folds and the retention rates of the **Ds–Px** pair in the amplified DNA products are summarized in Supplementary Table S1 and Supplementary Figures S3–S15. Although the sequencing patterns in the presence of d**Pa**′TP for **Ds**-temp 4 did not give a clear gap corresponding to the unnatural base position (30), the retention rates of the **Ds–Px** pair in the PCR products were accurately determined from the sequencing patterns in the presence of dd**Pa**′TP (Figure 3). In the PCR amplification of **Ds**-temps 1, 2 and 3 using the family B polymerases, most of the **Ds–Px** pair (>97%) was retained in the PCR products, although the fold amplification depended on the polymerase. In this analysis, >97% of the retention rate is the maximum detection limit. In the PCR amplification of the unfavorable **Ds**-temp 4, the retention rates of the **Ds–Px** pair and the amplification folds varied with the polymerase, but most of them were lower than those of **Ds**-temps 1, 2 and 3. Among them, the Deep Vent (exo$^+$) (Figure 3A), AccuPrime *Pfx* (Figure 3B) and Pfx50 DNA pols (Figure 3C) exhibited high amplification folds and retention rates with all of the templates. In addition, increasing the concentration of the natural base substrates [dNTPs, 500 μM for Deep Vent (exo$^+$) and 400 μM for AccuPrime *Pfx*] improved the amplification folds, especially those of **Ds**-temp 4, without a significant reduction in their retention rates. For example, in a 15-cycle PCR amplification by AccuPrime *Pfx* DNA pol with 50 μM d**Ds**TP and NH_2-hx-d**Px**TP and 400 μM natural dNTPs, the amplification folds were 1774 for **Ds**-temp 1 and 1503 for the unfavorable **Ds**-temp 4, and were only slightly lower than that (1955-fold) for Cont-temp. Under the same conditions, the retention rates of the **Ds–Px** pair in

the amplified products were >97% for **Ds**-temp 1 and 93 ± 2% for **Ds**-temp 4. The dependency of the amplification folds on the natural base substrate concentrations suggests that the natural base incorporation after the unnatural base pairing is one of the rate limiting steps, and thus the amplification efficiency could be improved by increasing the natural base substrate concentrations. Figure 2C and D also support this suggestion, and the amounts of the truncated products were significantly reduced by increasing the concentrations of the natural base substrates.

Next, we examined the effect of the $3' \rightarrow 5'$ exonuclease activity of the DNA polymerases on the unnatural base pair PCR amplification, using Deep Vent DNA pols with (exo^+) and without the exonuclease activity (exo^-). Deep Vent DNA pol (exo^-) performed with relatively high efficiency (424-fold after 15 cycles of PCR) and selectivity (>97% retention rate of the unnatural base pair). However, the sequencing patterns in the presence of dd**Pa**′TP of the PCR products using **Ds**-temp 1 and Cont-temp (Figure 3E) indicated a measurable problem with unnatural base misincorporation opposite the natural bases. As shown in the dd**Pa**′TP sequencing using **Ds**-temp 1 and Cont-temp (Figure 3E), the peak heights corresponding to the natural bases gradually decreased, relative to those corresponding to the exonuclease-proficient pol (exo^+) (Figure 3A). During the sequencing reaction, dd**Pa**′TP may have been incorporated opposite multiple **Ds** positions, which resulted from the unnatural base misincorporations in place of the natural bases during PCR amplification, and the sequencing reaction was paused, depending on the unnatural base misincorporation rate by the dd**Pa**′TP incorporation opposite **Ds**. Therefore, the $3' \rightarrow 5'$ exonuclease activity of polymerases is important to remove the unnatural base substrates misincorporated opposite the natural bases during PCR.

We also examined PCR amplification using family A (Taq and TITANIUM *Taq*) DNA pols. The Taq and TITANIUM *Taq* DNA pols, which have intrinsically no $3' \rightarrow 5'$ exonuclease activity, were ineffective, even in PCRs using **Ds**-temps 1 and 2 (Figure 3F and Supplementary Table S1). The selectivity of the **Ds–Px** pairing was much lower than that using the family B DNA pols. However, based on the sequencing pattern in the presence of dd**Pa**′TP (Figure 3F), no gradual attenuation of each peak was observed, as compared to the pattern from the template obtained using Deep Vent DNA pol (exo^-) (Figure 3E). Thus, the misincorporation rates of the unnatural base substrates opposite the natural bases are not very high when using Taq and TITANIUM *Taq* DNA pols. Indeed, our recent study showed that TITANIUM *Taq* DNA pol exhibited high selectivity and efficiency in a one-way incorporation of modified d**Px**TP opposite **Ds** in templates, and is useful for the unnatural base incorporation in a primer region in PCR (36). These results suggest that the TITANIUM *Taq* and Taq DNA pols efficiently incorporate d**Px**TP into DNA opposite **Ds**, but are not useful to incorporate d**Ds**TP into DNA opposite **Px**.

Sequence dependence of the PCR amplification efficiency

We determined the exact amplification efficiency using the less effective template, **Ds**-temp 4, in comparison to those using **Ds**-temp 1 and Cont-temp, by real-time PCR. Four 10-fold dilutions (20 fM–20 pM) of each template were amplified by Deep Vent DNA pol (exo^+) with 50 μM d**Ds**TP and NH_2-hx-d**Px**TP and 300 μM natural dNTPs, and the amplified products were monitored by the conventional SYBR Green dye detection system with a real-time PCR instrument (Figure 4A–C). Figure 4D shows the standard curves of the PCR amplification using each template, from which the amplification efficiencies ($=[10^{(-1/\text{slope})}]-1$) of the PCRs using **Ds**-temp 1, **Ds**-temp 4 and Cont-temp [0.91, 0.80 and 1.04 (~1.0), respectively] were determined. Although there are some differences among the template sequence contexts with or without the unnatural base, the values (0.80–0.91) of the amplification efficiency using the **Ds**-containing templates would be acceptable for the practical use of PCR involving the **Ds–Px** pair. Indeed, after 40 cycles of PCR starting from single-stranded DNA templates, the DNA fragments (as double strands) would theoretically be amplified 9.0×10^9-fold ($=1.80^{39}$) for 0.80 and 9.1×10^{10}-fold for 0.91, which coincide well with the experimental data, as discussed later (Table 1).

We also found that the PCR amplification with **Ds**-temp 4 took a longer time to reach the exponential phase than those with **Ds**-temp 1 and Cont-temp (Figure 4). The *y*-intercept values in the standard curves were significantly different among the templates (41.73 using **Ds**-temp 1, 46.78 using **Ds**-temp 4 and 37.46 using Cont-temp) (Figure 4D). The Cont-temp value (37.46) is reasonable, relative to that in conventional real-time PCR detection, but the values (41.73 and 46.78) obtained using **Ds**-containing templates, especially **Ds**-temp 4, are relatively high. In general, the high *y*-intercept values result from the degradation of templates during PCR, but we confirmed the stability of the **Ds**-containing DNA templates under the PCR conditions. A plausible possibility is that the complex of the polymerase and the **Ds**-containing template, especially **Ds**-temp 4, may dissociate at low template concentrations. Therefore, increasing the initial concentration of the DNA template, as well as the substrate concentrations as mentioned earlier, might increase the fold amplification when using an unfavorable sequence context, such as **Ds**-temp 4.

100 cycles of PCR involving the Ds–Px pairing

Using the optimized PCR conditions, we performed 100 cycles of PCR amplification of **Ds**-temp 1 with d**Ds**TP and NH_2-hx-d**Px**TP by the Deep Vent (exo^+) and AccuPrime *Pfx* DNA pols, to examine the abilities and limitations of the **Ds–Px** pair in PCR, as well as to increase the maximum detection limits (>97%) of the unnatural base pair retention rates in the amplified products. To maintain the exponential amplification through 100 cycles of PCR, we repeated the process of a 10-cycle PCR and the dilution of the PCR solution 10 times. **Ds**-temp 1 (50 fmol) was amplified by 10 cycles of PCR with 30 μM d**Ds**TP and NH_2-hx-d**Px**TP and 300 μM natural dNTPs with Deep

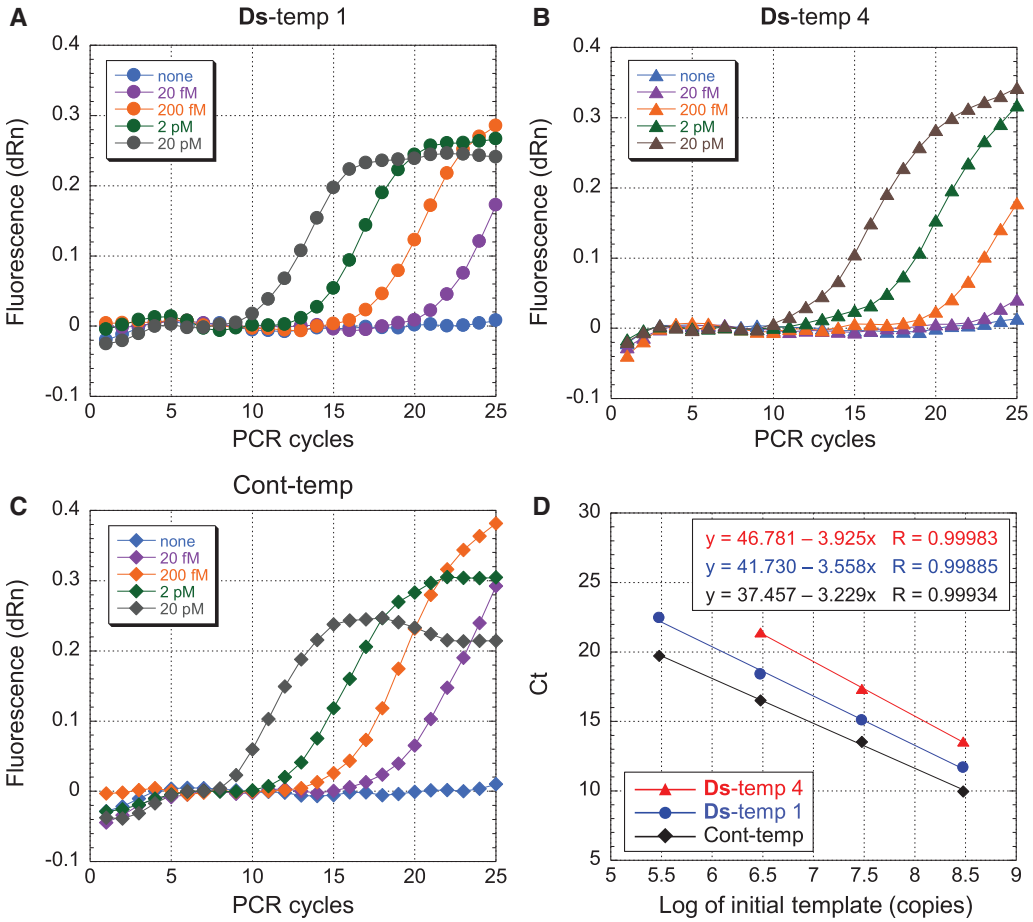


Figure 4. Real-time quantitative PCR amplification plots of each DNA template [(**A**) **Ds**-temp 1, (B) **Ds**-temp 4, (**C**) Cont-temp] and their linear standard curve analysis (**D**). PCR was performed with Deep Vent DNA pol (exo$^+$) in the presence of 50 μM d**Ds**TP and NH_2-hx-d**Px**TP, 300 μM natural dNTPs, and different concentrations of DNA templates (20 fM, 200 fM, 2 pM, 20 pM).

Table 1. Amplification efficiencies and fidelities of the **Ds–Px** pairing in 40 cycles of PCR in different template sequence contexts, using d**Ds**TP and NH_2-hx-d**Px**TP

DNA polymerase	DNA template	Natural substrates[a] (μM)	Unnatural substrates[b] (μM)	Amplification fold[c]	Amplification efficiency	Selectivity[d] (%/replication)	Misincorporation rate[e] (%/bp/replication)
Deep Vent (exo$^+$)	**Ds**-temp 1	300	30	9.6×10^{10}	0.91	99.92	–
			40	9.0×10^{10}	0.91	99.92	–
			50	6.7×10^{10}	0.89	99.92	–
	Ds-temp 3		30	6.6×10^{10}	0.89	99.92	–
			50	4.8×10^{10}	0.88	99.91	–
	Ds-temp 4		30	1.4×10^{10}	0.82	99.61	–
			50	8.4×10^{9}	0.80	99.75	–
	Cont-temp		30	5.7×10^{10}	0.89	–	0.007
			40	9.4×10^{10}	0.91	–	0.010
			50	7.2×10^{10}	0.90	–	0.012
AccuPrime *Pfx*	**Ds**-temp 1	400	30	7.7×10^{10}	0.90	99.74	–
			50	1.0×10^{11}	0.91	99.88	–
	Ds-temp 3		30	1.0×10^{11}	0.91	99.72	–
			50	7.7×10^{10}	0.90	99.89	–
	Ds-temp 4		30	1.7×10^{10}	0.83	98.94	–
			50	1.2×10^{10}	0.81	99.22	–
	Cont-temp		30	1.0×10^{11}	0.91	–	0.011
			40	1.4×10^{11}	0.93	–	0.016
			50	8.0×10^{10}	0.90	–	0.023
		600	50	1.0×10^{11}	0.91	–	0.017

[a]Concentration of each natural base substrate (dNTPs), [b]Concentration of d**Ds**TP and NH_2-hx-d**Px**TP, [c]Amplification folds after 40 cycles of PCR, [d]Selectivity of the unnatural base pairing, [e]Misincorporation rate of the unnatural base substrates opposite the natural bases in templates.

Vent DNA pol (exo$^+$) or with 50 μM d**Ds**TP and NH_2-hx-d**Px**TP and 400 μM natural dNTPs with AccuPrime *Pfx* DNA pol, and the PCR solution was diluted 750-fold for the following amplification (Figure 5A). From the gel electrophoresis of the products after 100 cycles of PCR (Figure 5C), **Ds**-temp 1 was amplified 5×10^{27}-fold by Deep Vent DNA pol (exo$^+$) and 2×10^{28}-fold by AccuPrime *Pfx* DNA pol. The position of the **Ds–Px** pair in the amplified products was confirmed by sequencing in the presence of d**Pa**′TP (left panels in Figure 5D and E). The retention rates of the **Ds–Px** pair in the amplified products were determined by sequencing in the presence of dd**Pa**′TP (right panels in Figure 5D and E), and for the **Ds–Px** pair, >97% for Deep Vent DNA pol (exo$^+$) and 88.4% for AccuPrime *Pfx* DNA pol were retained in the amplified DNA after 100 cycles of PCR. The selectivities of the **Ds–Px** pairing were determined from the amplification folds and the retention rates, and were as high as 99.97 and 99.87%/replication for Deep Vent DNA pol (exo$^+$) and AccuPrime *Pfx* DNA pol, respectively.

Another important consideration is the misincorporation rates of the unnatural base substrates opposite the natural bases in templates during PCR amplification. To determine the misincorporation rates, we developed another method (Figure 5B), in which the misincorporated PCR products were specifically labeled by a second PCR with fluorescent dye-linked d**Px**TP. The natural base template, Cont-temp, was amplified by 100 cycles of PCR in the presence of the examined unnatural base substrates, and the products were again amplified by 7 cycles of the second PCR with another combination of fluorescent dye-attached unnatural base substrates, Cy5-hx-d**Px**TP (36) (Figure 1) and d**Ds**TP, using a fluorescein (FAM)-labeled primer. The amplified products were then analyzed by Cy5 and FAM double detection on the gel. The fluorescence of Cy5 represents the replacement of the misincorporated unnatural bases (see Figure 5G), and the fluorescence of FAM corresponds to the total amplified DNA fragments (see Figure 5F). During the 7-cycle PCR with d**Ds**TP and Cy5-hx-d**Px**TP, these unnatural base substrates are also slightly misincorporated opposite the natural bases, and thus the initial DNA template was also amplified by the second PCR with d**Ds**TP and Cy5-hx-d**Px**TP (see the Cont-temp lane in Figure 5G), to subtract the misincorporation rate of Cy5-hx-**Px** from those after the first and second PCR amplifications.

Figure 5F and G show the gel electrophoreses for determining the misincorporation rates of unnatural base substrates opposite the natural bases after 10-, 20-, 40-, 70- and 100-cycles of the first PCR using 30 μM d**Ds**TP and NH_2-hx-d**Px**TP and 300 μM natural dNTPs by Deep Vent DNA pol (exo$^+$), and the data including the misincorporation rates (% in total amplified products, %/bp in total amplified products, %/bp/replication) are summarized in Figure 5H. After 100 cycles of PCR, Cont-temp was amplified 8×10^{27}-fold. Although 6.85% of the amplified DNA fragments contained the unnatural base pairs, resulting from the unnatural base misincorporation, after 100 cycles of PCR, the misincorporation rate was only 0.0067%/bp/replication. This misincorporation rate corresponds to an error rate of 6.7×10^{-5} error/bp, which is very close to the intrinsic error rate among the natural bases by Deep Vent DNA pol (exo$^+$) ($\sim 2 \times 10^{-5}$ error/bp, from the New England BioLabs information).

Fidelity of the Ds–Px pairing in 40 cycles of PCR

Next, we determined the fidelity of both the selectivity of the **Ds–Px** pairing and the misincorporation rate of the unnatural base substrates opposite the natural bases. We performed 40 cycles of PCR (10 cycles repeated four times) using various templates, **Ds**-temps 1, 3 and 4 and Cont-temp, with different concentrations of d**Ds**TP, NH_2-hx-d**Px**TP, the natural dNTPs, and the Deep Vent (exo$^+$) and AccuPrime *Pfx* DNA pols. The amplification folds and efficiencies, the selectivities of the **Ds–Px** pairing, and the misincorporation rates under several PCR conditions are summarized in Table 1. When using 50 μM d**Ds**TP and NH_2-hx-d**Px**TP and 300 μM natural dNTPs, the PCR amplification folds of each template by Deep Vent DNA pol (exo$^+$) were relatively high (6.7×10^{10}-fold for **Ds**-temp 1, 4.8×10^{10}-fold for **Ds**-temp 3 and 8.4×10^{9}-fold for **Ds**-temp 4), and from the amplification folds, the amplification efficiencies [amplification fold = $(1 + \text{amplification efficiency})^n$, where n = PCR cycles − 1 for single-stranded templates] were estimated as 0.89 for **Ds**-temp 1, 0.88 for **Ds**-temp 3 and 0.80 for **Ds**-temp 4. These efficiencies closely agreed with those (0.91 for **Ds**-temp 1 and 0.80 for **Ds**-temp 4) obtained by the real-time PCR experiments as mentioned earlier, confirming the accuracy of our determination methods. Furthermore, reducing the concentration of the unnatural base substrates, from 50 to 30 μM each, improved both the amplification folds and misincorporation rates without any significant loss of the **Ds–Px** pairing selectivity. The improved amplification efficiency might be a consequence of the reduced misincorporation rates opposite the natural bases. In PCR with 30 μM d**Ds**TP, NH_2-hx-d**Px**TP and 300 μM natural dNTPs for all template sequence contexts, the amplification efficiency, the selectivity and the misincorporation rate were 0.82–0.91 and 99.61–99.92%/replication, and 0.007%/bp/replication, respectively. The PCR amplification by the AccuPrime *Pfx* DNA pol was less selective, as compared to that by the Deep Vent DNA pol. However, it exhibited the highest amplification efficiency among the examined polymerases, and thus decreasing the number of PCR cycles could compensate for the selectivity loss when using the AccuPrime *Pfx* DNA pol.

Side-chain effects of the Px base on the Ds–Px pair fidelity

During the extensive experiments of PCR amplification using modified **Px** base substrates for generating new functional DNA molecules, we found that modifications of the **Px** base positively affected the fidelity of the **Ds–Px** pairing in PCR. One of the advantages of the **Ds–Px** pair is the capability of the modified **Px** base, as we used

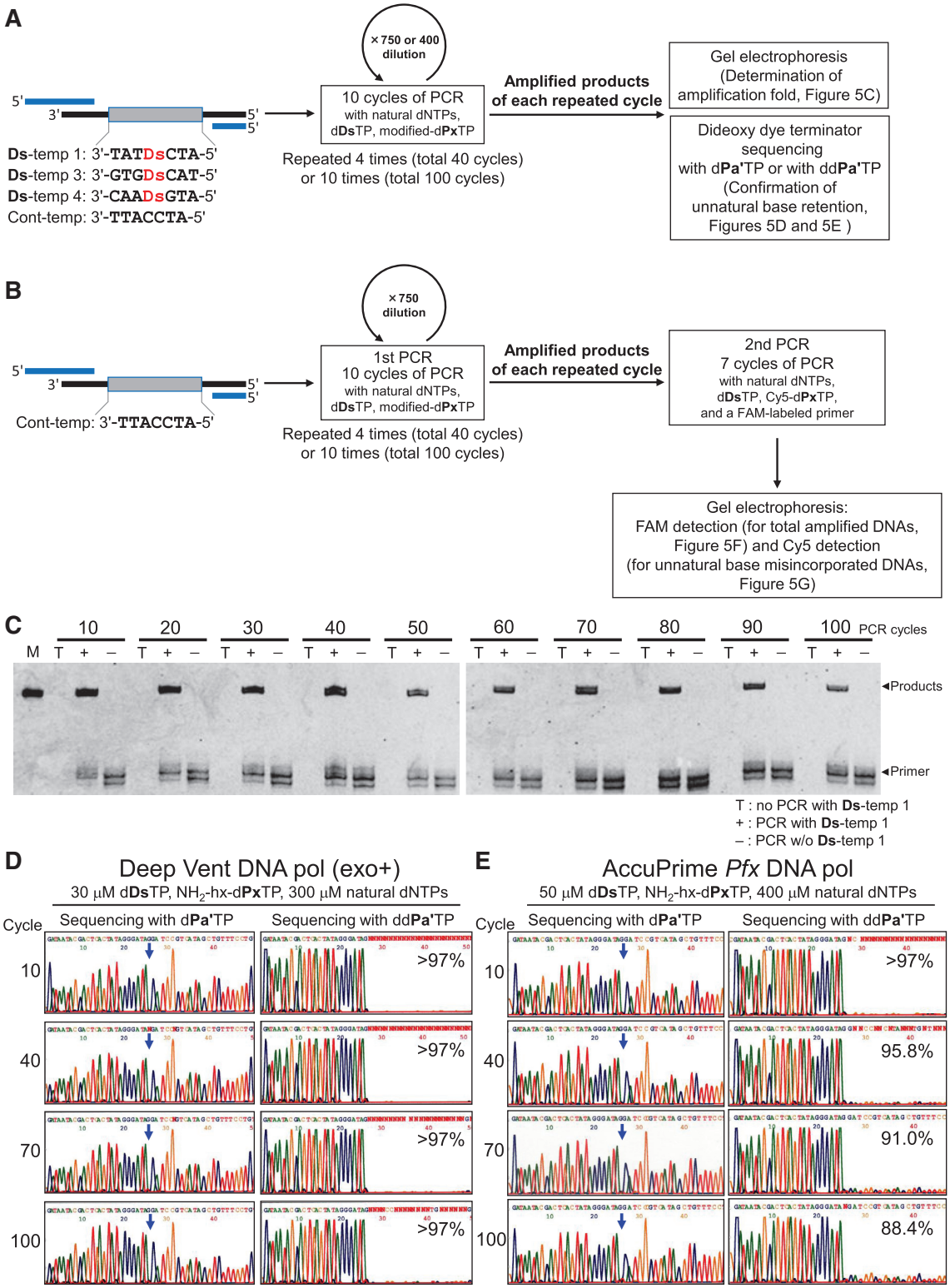
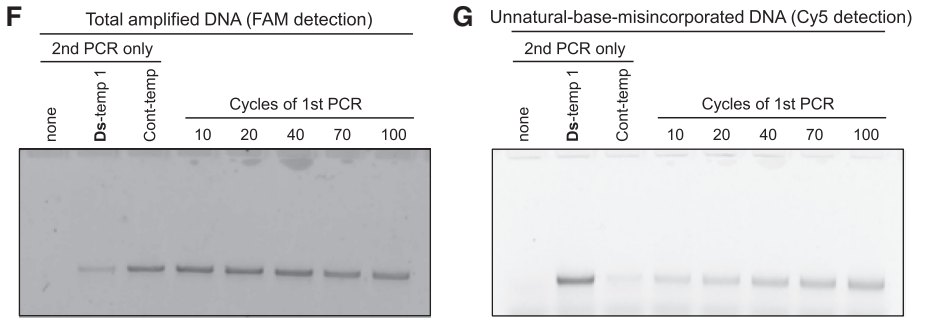


Figure 5. Amplification efficiency and fidelity assessments of the **Ds–Px** pairing by 40-cycle and 100-cycle PCR amplifications (repeated 10-cycle PCR). (**A**) Scheme of the 40- or 100-cycle PCR amplifications of the **Ds**-containing DNA templates using d**Ds**TP and modified-d**Px**TP, for the determination of the fold amplification, efficiency and selectivity of the **Ds–Px** pairing in PCR. (**B**) Scheme of the 40- or 100-cycle PCR amplification of the DNA template comprising only the natural bases using d**Ds**TP and modified-d**Px**TP, for the determination of the misincorporation rates of the unnatural base substrates opposite the natural bases. The results are summarized in Tables 1 and 2. (**C**) Gel electrophoresis of the amplification products after each 10-cycle PCR (total 100 cycles of PCR) of **Ds**-temp 1 by the Deep Vent DNA pol (exo$^+$) using 1 μM each primer, 30 μM d**Ds**TP and NH_2-hx-d**Px**TP and 300 μM natural dNTPs. A portion of each 10-PCR solution before (lanes T) and after each 10-cycle PCR (lanes +) was analyzed by 15% polyacrylamide denaturing gel electrophoresis, followed by SYBR Green II staining. M: marker (75-mer single-stranded DNA). (**D** and **E**) Sequencing analysis of the PCR products after 10, 40, 70 and 100 cycles of PCR by Deep Vent DNA pol (exo$^+$) (D) and AccuPrime *Pfx*

(continued)



H

PCR cycles	Amplification folds	Ratio of the unnatural base misincorporated products in the total PCR products [% in total amplified DNAs]	Misincorporation rates of unnatural opposite the natural bases [%/bp in total PCR]	Misincorporation rates of unnatural opposite the natural bases [%/bp/replication]
10	3.2×10^2	0.87 (0.05)*	0.079	0.0095
20	2.1×10^5	1.49 (0.08)	0.137	0.0077
40	5.7×10^{10}	2.89 (0.12)	0.266	0.0074
70	1.4×10^{19}	5.05 (0.72)	0.470	0.0072
100	8.0×10^{27}	6.85 (0.68)	0.643	0.0067

*Standard deviations are shown in parentheses.

Figure 5. Continued
DNA pol (E). Sequencing reactions were performed in the presence of d**Pa**′TP (left panels) or dd**Pa**′TP (right panels). The blue arrow indicates the original unnatural base position. The percentage indicated in the right panels is the retention rate of the **Ds–Px** pair in the amplified products. The method for the calculation of the retention rates is provided in the Supplementary Data section. (**F–H**) Determination of misincorporation rates of the unnatural base substrates opposite the natural bases in Cont-temp after 10, 20, 40, 70 and 100 cycles of PCR by the Deep Vent DNA pol (exo$^+$) using 30 μM d**Ds**TP and NH_2-hx-d**Px**TP and 300 μM natural dNTPs (first PCR). The second PCR products with the FAM-labeled 3′-primer and Cy5-hx-d**Px**TP were analyzed by 15% denaturing PAGE, and the DNA fragments were detected with FAM fluorescence (F) for the quantification of the total PCR products and with Cy5 fluorescence (G) for the quantification of the unnatural base misincorporated products. (H) Misincorporation rates calculated from the data are summarized in the panel.

Cy5-hx-d**Px**TP to site-specifically incorporate it into DNA opposite **Ds**. A variety of functional groups can be attached to the **Px** base via the propynyl linker, and these modified d**Px**TPs are also site-specifically incorporated opposite **Ds** in replication. We synthesized several modified **Px** substrates by attaching amino (NH_2-C1- and NH_2-hx-d**Px**TPs), diol (Diol1- and Diol3o3-d**Px**TPs), and aromatic groups (DMP-, NTP-, BPh- and HBP-d**Px**TPs) with different linkers, as well as the non-modified d**Px**TP (Figure 1, see Supplementary Data section for the syntheses). The amplification efficiency and fidelity of the **Ds–Px** pairing using these modified d**Px**TP were assessed by 40 cycles of PCR with 50 μM d**Ds**TP and modified d**Px**TP, 300 μM natural dNTPs, DNA templates (**Ds**-temps 1 and 4 and Cont-temp) and the Deep Vent DNA pol (exo$^+$).

Table 2 shows the amplification fold and efficiency, the selectivity of the **Ds–Px** pairing, and the misincorporation rate for PCR amplification using each unnatural base pair combination of modified d**Px**TP and d**Ds**TP. Most of the modified d**Px**TPs, except for the biphenyl modifications (BPh- and HBP-d**Px**TPs), exhibited relatively high efficiency (0.79–0.91, around 10^{10}-fold amplification after 40 cycles of PCR) and high selectivity (>99.9%/replication) when using **Ds**-temp 1. In the PCR amplification of **Ds**-temp 4, the efficiency and selectivity with all of the modified d**Px**TPs were decreased but were still as high as 0.75–0.80 and 99.22–99.87%/replication, respectively.

The biphenyl modifications (BPh- and HBP-d**Px**TPs) showed a unique sequence context dependency (Table 2). Although the amplification efficiency and selectivity were lower than those of the other modified **Px** substrates, these modifications worked better with **Ds**-temp 4 than **Ds**-temp 1 in PCR, even by the Deep Vent DNA pol (exo$^+$). For example, in PCR using HBP-d**Px**TPs and d**Ds**TP, the amplification efficiencies were 0.68 for **Ds**-temp 1 and 0.71 for **Ds**-temp 4, and the selectivities of the **Ds–Px** pairing were 98.97% for **Ds**-temp 1 and 99.83% for **Ds**-temp 4, with a relatively low misincorporation rate opposite the natural bases (0.008%/bp/replication).

We also found that **Px** modifications improved the fidelity of the **Ds–Px** pairing in PCR. As compared with the misincorporation rate using d**Px**TP (0.019%/bp/replication), all of the **Px** modifications reduced the misincorporation rates (0.003–0.016%/bp/replication). Among the modified d**Px**TPs, the selectivity of the **Ds–Px** pairing was less correlated with the misincorporation rate, and the diol-modified d**Px**TPs showed the lowest misincorporation rate (0.003–0.005%/bp/replication) even when using high concentrations (50 μM) of the

Table 2. Amplification efficiencies and fidelities of the **Ds–Px** pairing in 40 cycles of PCR in different template sequence contexts using several combinations of d**Ds**TP and modified d**Px**TP by Deep Vent DNA pol (exo^+)

Modified dPxTP	DNA template	Amplification fold[a]	Amplification efficiency	Selectivity[b] (%/replication)	Misincorporation rate[c] (%/bp/replication)
d**Px**TP	**Ds**-temp 1	4.3×10^{10}	0.87	99.91	–
	Ds-temp 4	4.3×10^{9}	0.77	99.88	–
	Cont-temp	4.2×10^{10}	0.87	–	0.019
NH_2-C1-d**Px**TP	**Ds**-temp 1	3.4×10^{10}	0.86	99.91	–
	Ds-temp 4	2.9×10^{9}	0.75	99.22	–
	Cont-temp	5.0×10^{10}	0.88	–	0.006
NH_2-hx-d**Px**TP	**Ds**-temp 1	6.7×10^{10}	0.89	99.92	–
	Ds-temp 4	8.4×10^{9}	0.80	99.75	–
	Cont-temp	7.2×10^{10}	0.90	–	0.012
Diol1-d**Px**TP	**Ds**-temp 1	6.2×10^{10}	0.89	99.92	–
	Ds-temp 4	7.6×10^{9}	0.79	99.77	–
	Cont-temp	6.1×10^{10}	0.89	–	0.005
Diol3o3-d**Px**TP	**Ds**-temp 1	9.0×10^{10}	0.91	99.92	–
	Ds-temp 4	5.3×10^{9}	0.78	99.64	–
	Cont-temp	1.2×10^{11}	0.92	–	0.003
DMP-hx-d**Px**TP	**Ds**-temp 1	1.7×10^{10}	0.83	99.91	–
	Ds-temp 4	3.3×10^{9}	0.75	99.77	–
	Cont-temp	5.8×10^{10}	0.89	–	0.011
NTP-hx-d**Px**TP	**Ds**-temp 1	6.9×10^{9}	0.79	99.91	–
	Ds-temp 4	3.3×10^{9}	0.75	99.87	–
	Cont-temp	4.4×10^{10}	0.87	–	0.016
BPh-hx-d**Px**TP	**Ds**-temp 1	3.1×10^{8}	0.65	98.66	–
	Ds-temp 4	8.2×10^{8}	0.69	99.76	–
	Cont-temp	2.2×10^{10}	0.84	–	0.011
HBP-hx-d**Px**TP	**Ds**-temp 1	5.5×10^{8}	0.68	98.97	–
	Ds-temp 4	1.2×10^{9}	0.71	99.83	–
	Cont-temp	5.2×10^{10}	0.88	–	0.008

[a]Amplification folds after 40 cycles of PCR using 50 μM d**Ds**TP, modified-d**Px**TP and 300 μM natural base dNTPs, [b]Selectivity of the unnatural base pairing, [c]Misincorporation rate of the unnatural base substrates opposite the natural bases in templates.

unnatural base substrates, without any significant reductions in the amplification fold and the selectivity of the **Ds–Px** pairing. From the overall evaluation among these modified d**Px**TPs, the combination of Diol1-d**Px**TP and d**Ds**TP was the best for PCR amplification, and exhibited the highest amplification efficiency (0.79–0.89) and selectivity (99.77–99.92%/replication) among all of the templates, and the lowest misincorporation rate (0.005%/bp/replication $\rightarrow 5 \times 10^{-5}$ error/bp of error rate) opposite the natural base templates. These results suggest that the modification of unnatural bases protruding into the major groove side also affects the fidelity of the unnatural base pairing in replication.

DISCUSSION

We fine-turned the PCR conditions involving the **Ds–Px** pair as a third base pair and developed methods for assessing the amplification efficiency and fidelity of the unnatural base pairing. In the methods, we increased the accuracy of the determination of the PCR amplification efficiency and the fidelity of the **Ds–Px** pair by maintaining exponential amplification conditions, by repeating 10-cycle PCRs. The best combination for the **Ds–Px** pair PCR system is d**Ds**TP and Diol1-d**Px**TP. Under the optimized conditions using 50 μM d**Ds**TP and Diol1-d**Px**TP and 300 μM natural dNTPs by Deep Vent (exo^+) pol, DNA fragments with different sequence contexts, except for purine-**Ds**-purine motifs, were amplified $\sim 10^{10}$-fold (amplification efficiency = 0.89) by 40 cycles of PCR and the selectivity of the cognate **Ds–Px** pairing was >99.9%/replication. Even in PCR with unfavorable sequence contexts containing purine-**Ds**-purine, the unnatural base pair selectivity was 99.77%/replication, with high amplification efficiency (0.79) similar to those of other sequence contexts. The diol modification of the **Px** base reduced the misincorporation rate of the unnatural base substrates opposite the natural bases in template, which was 0.005%/bp/replication. The low misincorporation rate of the **Ds** and Diol1-**Px** substrates (0.005%/bp/replication = 5×10^{-5} error/bp) is very close to the intrinsic mispairing error rate among the natural bases ($\sim 2 \times 10^{-5}$ error/bp) of Deep Vent DNA pol (exo^+). Therefore, the ability of the unnatural base pairing is approaching those of the natural A–T and G–C pairings.

The **Ds–Px** pair in combination with d**Ds**TP and NH_2-hx-d**Px**TP also exhibited high efficiency and fidelity in PCR amplification using the Deep Vent (exo^+) and AccuPrime *Pfx* DNA pols. The selectivities in PCR with **Ds**-temp 1 using 30 μM d**Ds**TP and NH_2-hx-d**Px**TP and 300 μM natural dNTPs were 99.92%/replication for Deep Vent (exo^+) pol and 99.74% for AccuPrime *Pfx* DNA pol. Under the same conditions, the misincorporation rates

were 0.007%/bp/replication for Deep Vent (exo^+) pol and 0.011%/bp/replication for AccuPrime *Pfx* DNA pol. We demonstrated the great potential of the **Ds–Px** pair by performing 100 cycles of PCR using d**Ds**TP and NH_2-hx-d**Px**TP. Even in PCR using the unfavorable template sequence context by Deep Vent (exo^+) pol, the amplification efficiency (0.82 for **Ds**-temp 4) and the selectivity (99.6%/replication for **Ds**-temp 4) adequately reached practical levels for PCR. Although the PCR amplification by the AccuPrime *Pfx* DNA pol was less selective than that by Deep Vent (exo^+), the higher amplification efficiency is compensated by the shorter PCR cycles. In addition, the AccuPrime *Pfx* DNA pol might be useful for PCR amplification of DNA templates containing multiple unnatural bases. Furthermore, a variety of modified **Px** substrates can also be used for the **Ds–Px** pair PCR system, enabling the site-specific functionalization of DNA molecules.

The present analyses of PCR involving the **Ds–Px** pair suggest several points for the further development of unnatural base pair systems. Some of the points might also offer suggestions for the mechanisms of the natural base pairing in replication and the development of natural base pair PCR systems with high efficiency and fidelity. (i) The main role of the $3' \rightarrow 5'$ exonuclease activity of Deep Vent DNA pol is to prevent the misincorporation of the unnatural base substrates opposite the natural bases. Thus, the polymerase proofreading mechanisms function for the recognition of non-cognate pairs between the natural and unnatural bases. In addition, mixed polymerases with and without exonuclease activity might improve the unnatural base pair efficiency and selectivity. Indeed, AccuPrime *Pfx* DNA pol contains two types of polymerases. (ii) One of the rate limiting steps is the natural base incorporation opposite its complementary base after the unnatural base pair, as well as the unnatural base pairing. Thus, increasing the concentration of the natural base substrates could improve the amplification efficiency without any loss of fidelity in PCR. (iii) The real-time PCR analyses suggested that the amplification fold is also affected by the initial template concentration, and a higher template concentration could increase the fold amplification with fewer PCR cycles. The less effective purine-**Ds**-purine sequences might reduce the stability of the polymerase and substrate complex, and thus increasing the concentration of each component is important for efficient PCR amplification for all template sequence contexts and multiple incorporations of the unnatural base pairs. (iv) Extra side chains attached to unnatural bases affect the fidelity of the unnatural base pairing. Among the side chain functional groups, some of them increase the unnatural base pair fidelity and others change the polymerase's preference for the sequence contexts.

Our studies of the **Ds–Px** pair in PCR have confirmed the utility of a third base pair as a replicative component with high efficiency and fidelity, besides the natural A–T and G–C base pairs, and further development of the unnatural base pair systems and applications to new biotechnologies are expected in the future.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Methods, Supplementary Table 1 and Supplementary Figures 1–15.

FUNDING

Funding for open access charge: Targeted Proteins Research Program and RIKEN Structural Genomics/Proteomics Initiative; National Project on Protein Structural and Functional Analyses, Ministry of Education, Culture, Sports, Science and Technology of Japan; and Grants-in-Aid for Scientific Research (KAKENHI 19201046 to I.H. and 20710176 to M.K.) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Conflict of interest statement. None declared.

REFERENCES

1. Service,R.F. (2000) Molecular biology. Creation's seventh day. *Science*, **289**, 232–235.
2. Benner,S.A. and Sismour,A.M. (2005) Synthetic biology. *Nat. Rev. Genet.*, **6**, 533–543.
3. Hunziker,J. and Mathis,G. (2005) DNA with artificial base pairs. *Chimia*, **59**, 780–784.
4. Bergstrom,D.E. (2009) Unnatural nucleosides with unusual base pairing properties. *Curr. Protoc. Nucleic Acid Chem.*, **Chapter 1**, Unit 1.4.
5. Krueger,A.T. and Kool,E.T. (2009) Redesigning the architecture of the base pair: toward biochemical and biological function of new genetic sets. *Chem. Biol.*, **16**, 242–248.
6. Kool,E.T. (2000) Synthetically modified DNAs as substrates for polymerases. *Curr. Opin. Chem. Biol.*, **4**, 602–608.
7. Henry,A.A. and Romesberg,F.E. (2003) Beyond A, C, G and T: augmenting nature's alphabet. *Curr. Opin. Chem. Biol.*, **7**, 727–733.
8. Hirao,I. (2006) Unnatural base pair systems for DNA/RNA-based biotechnology. *Curr. Opin. Chem. Biol.*, **10**, 622–627.
9. Kimoto,M., Cox,R.S. III and Hirao,I. (2011) Unnatural base pair systems for sensing and diagnostic applications. *Expert Rev. Mol. Diagn.*, **11**, 321–331.
10. Tor,Y. and Dervan,P.B. (1993) Site-specific enzymatic incor poration of an unnatural base, N6-(6-aminohexyl)isoguanosine, into RNA. *J. Am. Chem. Soc.*, **115**, 4461–4467.
11. Hirao,I. (2006) Placing extra components into RNA by specific transcription using unnatural base pair systems. *Biotechniques*, **40**, 711–715.
12. Kimoto,M., Endo,M., Mitsui,T., Okuni,T., Hirao,I. and Yokoyama,S. (2004) Site-specific incorporation of a photo-crosslinking component into RNA by T7 transcription mediated by unnatural base pairs. *Chem. Biol.*, **11**, 47–55.
13. Kawai,R., Kimoto,M., Ikeda,S., Mitsui,T., Endo,M., Yokoyama,S. and Hirao,I. (2005) Site-specific fluorescent labeling of RNA molecules by specific transcription using unnatural base pairs. *J. Am. Chem. Soc.*, **127**, 17286–17295.
14. Moriyama,K., Kimoto,M., Mitsui,T., Yokoyama,S. and Hirao,I. (2005) Site-specific biotinylation of RNA molecules by transcription using unnatural base pairs. *Nucleic Acids Res.*, **33**, e129.
15. Kimoto,M., Mitsui,T., Harada,Y., Sato,A., Yokoyama,S. and Hirao,I. (2007) Fluorescent probing for RNA molecules by an unnatural base-pair system. *Nucleic Acids Res.*, **35**, 5360–5369.
16. Hikida,Y., Kimoto,M., Yokoyama,S. and Hirao,I. (2010) Site-specific fluorescent probing of RNA molecules by unnatural base-pair transcription for local structural conformation analysis. *Nat. Protoc.*, **5**, 1312–1323.

17. Moser,M.J., Marshall,D.J., Grenier,J.K., Kieffer,C.D., Killeen,A.A., Ptacin,J.L., Richmond,C.S., Roesch,E.B., Scherrer,C.W., Sherrill,C.B. *et al.* (2003) Exploiting the enzymatic recognition of an unnatural base pair to develop a universal genetic analysis system. *Clin. Chem*, **49**, 407–414.
18. Sherrill,C.B., Marshall,D.J., Moser,M.J., Larsen,C.A., Daudé-Snow,L., Jurczyk,S., Shapiro,G. and Prudent,J.R. (2004) Nucleic acid analysis using an expanded genetic alphabet to quench fluorescence. *J. Am. Chem. Soc.*, **126**, 4550–4556.
19. Prudent,J.R. (2006) Using expanded genetic alphabets to simplify high-throughput genetic testing. *Expert Rev. Mol. Diagn.*, **6**, 245–252.
20. Moser,M.J., Ruckstuhl,M., Larsen,C.A., Swearingen,A.J., Kozlowski,M., Bassit,L., Sharma,P.L., Schinazi,R.F. and Prudent,J.R. (2005) Quantifying mixed populations of drug-resistant human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.*, **49**, 3334–3340.
21. Kimoto,M., Mitsui,T., Yamashige,R., Sato,A., Yokoyama,S. and Hirao,I. (2010) A new unnatural base pair system between fluorophore and quencher base analogues for nucleic acid-based imaging technology. *J. Am. Chem. Soc.*, **132**, 15418–15426.
22. Switzer,C., Moroney,S.E. and Benner,S.A. (1989) Enzymatic incorporation of a new base pair into DNA and RNA. *J. Am. Chem. Soc.*, **111**, 8322–8323.
23. Moser,M.J., Christensen,D.R., Norwood,D. and Prudent,J.R. (2006) Multiplexed detection of anthrax-related toxin genes. *J. Mol. Diagn.*, **8**, 89–96.
24. Marshall,D.J., Reisdorf,E., Harms,G., Beaty,E., Moser,M.J., Lee,W.M., Gern,J.E., Nolte,F.S., Shult,P. and Prudent,J.R. (2007) Evaluation of a multiplexed PCR assay for detection of respiratory viral pathogens in a public health laboratory setting. *J. Clin. Microbiol.*, **45**, 3875–3882.
25. Nolte,F.S., Marshall,D.J., Rasberry,C., Schievelbein,S., Banks,G.G., Storch,G.A., Arens,M.Q., Buller,R.S. and Prudent,J.R. (2007) MultiCode-PLx system for multiplexed detection of seventeen respiratory viruses. *J. Clin. Microbiol.*, **45**, 2779–2786.
26. Johnson,S.C., Sherrill,C.B., Marshall,D.J., Moser,M.J. and Prudent,J.R. (2004) A third base pair for the polymerase chain reaction: inserting isoC and isoG. *Nucleic Acids Res.*, **32**, 1937–1941.
27. Yang,Z., Chen,F., Chamberlin,S.G. and Benner,S.A. (2010) Expanded genetic alphabets in the polymerase chain reaction. *Angew. Chem. Int. Ed. Engl.*, **49**, 177–180.
28. Yang,Z., Sismour,A.M., Sheng,P., Puskar,N.L. and Benner,S.A. (2007) Enzymatic incorporation of a third nucleobase pair. *Nucleic Acids Res.*, **35**, 4238–4249.
29. Sismour,A.M., Lutz,S., Park,J.H., Lutz,M.J., Boyer,P.L., Hughes,S.H. and Benner,S.A. (2004) 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.
30. Hirao,I., Kimoto,M., Mitsui,T., Fujiwara,T., Kawai,R., Sato,A., Harada,Y. and Yokoyama,S. (2006) An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogs into DNA and RNA. *Nat. Methods*, **3**, 729–735.
31. Hirao,I., Mitsui,T., Kimoto,M. and Yokoyama,S. (2007) An efficient unnatural base pair for PCR amplification. *J. Am. Chem. Soc.*, **129**, 15549–15555.
32. Kimoto,M., Kawai,R., Mitsui,T., Yokoyama,S. and Hirao,I. (2009) An unnatural base pair system for efficient PCR amplification and functionalization of DNA molecules. *Nucleic Acids Res.*, **37**, e14.
33. Malyshev,D.A., Seo,Y.J., Ordoukhanian,P. and Romesberg,F.E. (2009) PCR with an expanded genetic alphabet. *J. Am. Chem. Soc.*, **131**, 14620–14621.
34. Malyshev,D.A., Pfaff,D.A., Ippoliti,S.I., Hwang,G.T., Dwyer,T.J. and Romesberg,F.E. (2010) Solution structure, mechanism of replication, and optimization of an unnatural base pair. *Chemistry*, **16**, 12650–12659.
35. Yang,Z., Chen,F., Alvarado,J.B. and Benner,S.A. (2011) Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J. Am. Chem. Soc.*, **133**, 15105–15112.
36. Yamashige,R., Kimoto,M., Mitsui,T., Yokoyama,S. and Hirao,I. (2011) Monitoring the site-specific incorporation of dual fluorophore-quencher base analogues for target DNA detection by an unnatural base pair system. *Org. Biomol. Chem.*, **9**, 7504–7509.