J|A|C|S
ARTICLES
Published on Web 04/11/2003

An Unnatural Hydrophobic Base Pair with Shape Complementarity between Pyrrole-2-carbaldehyde and 9-Methylimidazo[(4,5)-b]pyridine

Tsuneo Mitsui,[†,‡,§] Aya Kitamura,[‡,#] Michiko Kimoto,[‡,#] Taiko To,[‡,#] Akira Sato,[†,#] Ichiro Hirao,*[,†,‡,§] and Shigeyuki Yokoyama*[,†,‡,#]

Contribution from the Yokoyama CytoLogic Project, ERATO, JST, c/o RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan, Protein Synthesis Technology Team, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama, Kanagawa 230-0045, Japan, Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan, and Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Received October 4, 2002; E-mail: ihirao@postman.riken.go.jp and yokoyama@biochem.s.u-tokyo.ac.jp

Abstract: An unnatural hydrophobic base, pyrrole-2-carbaldehyde (denoted as **Pa**), was developed as a specific pairing partner of 9-methylimidazo[(4,5)-b]pyridine (**Q**). The **Q** base is known to pair with 2,4-difluorotoluene (**F**) as an isostere of the **A**−**T** pair, and **F** also pairs with **A** efficiently in replication. In contrast, the **Q**−**Pa** pair showed specific selectivity in replication, and the five-membered-ring base **Pa** paired efficiently with **Q** but paired poorly with **A**. In addition, the interaction of **Pa** with DNA polymerases was superior, in comparison to that of **F**. The aldehyde group of **Pa** was recognized well by the Klenow fragment of *Escherichia coli* DNA polymerase I and the reverse transcriptase of Avian myeloblastosis virus. The structural features of the **Q**−**Pa** pair in a DNA duplex were analyzed by NMR, showing the shape complementarity of the **Pa** fitting with **Q**. The structurally unique base **Pa** provides valuable information for the development of unnatural base pairs toward the expansion of the genetic alphabet.

Introduction

Creating unnatural base pairs that have exclusive selectivity in replication, transcription, and translation would lead to great advancements in the bioengineering of nucleic acids and proteins.[1−3] The expansion of the genetic codes permits the site-specific incorporation of amino acid analogues into proteins in vitro and in vivo. In addition, the introduction of unnatural nucleotide components into nucleic acids facilitates their increased functionality.

For the development of unnatural base pairs, Benner and colleagues showed the importance of the hydrogen-bonding patterns between pairing bases by their studies of unnatural base pairs, such as iso**G** and iso**C**.[4−6] Recently, Kool and colleagues synthesized non-hydrogen-bonded base pairs, such as 4-methylbenzimidazole (**Z**) and 2,4-difluorotoluene (**F**), and 9-methylimidazo[(4,5)-b]pyridine (**Q**) and **F**. The nucleotides of **Z** and **F** or **Q** and **F** are enzymatically incorporated into DNA opposite each other.[7−10] From this research, they developed two concepts: the importance of the shape complementarity between pairing bases and the potential of hydrophobic bases as artificial base pairs.[4] By combining the concepts of hydrogen-bonding patterns[4−6] and shape complementarity,[9−11] we developed unnatural base pairs between 2-amino-6-dimethylaminopurine (**x**) and pyridin-2-one (**y**),[12,13] and between 2-amino-6-(2-thienyl)purine (**s**) and **y**.[14,15] These unnatural base pairs showed high specificity in transcription[13,15] and were applied to coupled transcription and translation for the site-specific incorporation of unnatural amino acids into proteins.[15] On the other hand, the concept of a hydrophobic base pair was applied to the creation of unnatural base pairs for the expansion of the genetic

[†] Yokoyama CytoLogic Project, ERATO, JST.
[‡] Protein Synthesis Technology Team, RIKEN Genomic Sciences Center.
[§] Research Center for Advanced Science and Technology, The University of Tokyo.
[#] Department of Biophysics and Biochemistry, The University of Tokyo.

(1) Benner, S. A.; Burgstaller, P.; Battersby, T. R.; Jurczyk, S. *The RNA World*, 2nd ed.; Gesteland, R. F., Cech, T. R., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY, 1999; pp 163−181.
(2) Kool, E. T. *Curr. Opin. Chem. Biol.* **2000**, *4*, 602−608.
(3) Service, R. F. *Science* **2000**, *289*, 232−235.
(4) Switzer, C.; Moroney, S. E.; Benner, S. A. *J. Am. Chem. Soc.* **1989**, *111*, 8322−8323.
(5) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33−37.
(6) Horlacher, J.; Hottiger, M.; Podust, V. N.; Hübscher, U.; Benner, S. A. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6329−6333.
(7) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **1999**, *121*, 2323−2324.
(8) Kool, E. T.; Morales, J. C.; Guckian, K. M. *Angew. Chem., Int. Ed.* **2000**, *39*, 990−1009.
(9) Morales, J. C.; Kool, E. T. *Nat. Struct. Biol.* **1998**, *5*, 950−954.
(10) Matray, T. J.; Kool, E. T. *Nature* **1999**, *399*, 704−708.
(11) Rappaport, H. P. *Biochemistry* **1993**, *32*, 3047−3057.
(12) Ishikawa, M.; Hirao, I.; Yokoyama, S. *Tetrahedron Lett.* **2000**, *41*, 3931−3934.
(13) Ohtsuki, T.; Kimoto, M.; Ishikawa, M.; Mitsui, T.; Hirao, I.; Yokoyama, S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4922−4925.
(14) Fujiwara, T.; Kimoto, M.; Sugiyama, H.; Hirao, I.; Yokoyama, S. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2221−2223.
(15) Hirao, I.; Ohtsuki, T.; Fujiwara, T.; Mitsui, T.; Yokogawa, T.; Okuni, T.; Nakayama, H.; Takio, K.; Yabuki, T.; Kigawa, T.; Kodama, K.; Yokogawa, T.; Nishikawa, K.; Yokoyama, S. *Nat. Biotechnol.* **2002**, *20*, 177−182.

10.1021/ja028806h CCC: $25.00

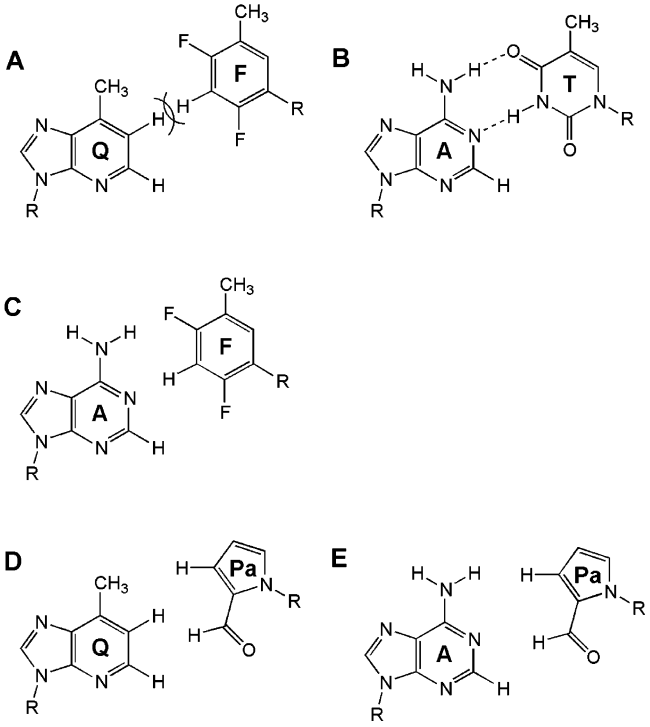


Figure 1. Base-pair structures: (A) the **Q**−**F** base pair, (B) the natural **A**−**T** pair, (C) a noncognate **A**−**F** pair, (D) the novel base **Pa** pairing with **Q**, and (E) a noncognate **A**−**Pa** pair. In these structures, R denotes ribose.

alphabet by the Romesberg and Schultz groups.[16−20] These hydrophobic base pairs are the potential candidates for unnatural base pairs at present.

To expand the development of unnatural base pairs, we focused on the hydrophobic **Z**−**F** and **Q**−**F** pairs. One of the problems of these base pairs is their poor exclusivity, because the shape-fitted **Z**−**F** and **Q**−**F** pairs are isosteres of the **A**−**T** pair, and thus efficient noncognate base pairings, especially between **A** and **F**, appear in replication[7,21,22] (see structures A−C in Figure 1). The shape of **F** mimics that of **T**, and **F** fits well with **A**. The **A**−**F** pair is much more efficient than the other noncognate **Z**−**T** and **Q**−**T** pairs. This may be because the H3 of **T** clashes with the hydrogens of **Z** and **Q** on the pairing surface. Similarly, the hydrogen of **F,** corresponding to the H3 of pyrimidines, may clash with the hydrogens of **Z** and **Q** (structure A in Figure 1). Actually, the geometry of the **Z**−**F** pair is slightly distorted in the DNA duplex, compared to that of the **A**−**T** pair.[23] Thus, the **A**−**F** pair is the most shape-fitted base pair among the cognate and noncognate pairs involving **Z**, **Q**, and **F**. Another problem is that **F** is not recognized by some polymerases, because it lacks the 2-keto group of the natural pyrimidines.[21,22] In this regard, the recognition of **Z** by DNA polymerases was improved by replacing it with **Q**,[7] which has a nitrogen at the site corresponding to position 3 of the natural purine bases that can interact with polymerases. Thus, the development of an unnatural base, which is replaced by **F** and specifically pairs with **Q**, would lead to the creation of an unnatural base pair with exclusive selectivity in replication.

To address these problems of the **F** base, we considered the stringent shape complementarity for the pairing partner of **Q** and the effective interactions of the partner base with DNA polymerases, and thus the hydrophobic pyrrole-2-carbaldehyde (**Pa**) was newly designed (structure D in Figure 1). The five-membered ring of **Pa** would avoid the steric collision with **Q**, and the aldehyde group of **Pa** could interact with polymerases. In addition, the fitting between **A** and **Pa** might be inferior to that between **A** and **F**. This is because the stacking stability of **Pa** might be weaker than that of **F**, and, thus, the **A**−**Pa** pair would be destabilized by the water solvation of **A**[8] (structure E in Figure 1). We synthesized the nucleotides of **Pa**, examined the efficiency of **Q**−**Pa** pairing in replication, and analyzed the structural features of the **Q**−**Pa** pair in a DNA duplex.

Experimental Section

Sample Preparation. The nucleoside triphosphates and phosphoramidites of **F** and **Q** were synthesized with a slight modification of the method described in the literature.[7] The phosphoramidites of iso**G** and iso**C** (5-methylisocytosine) were purchased from GLEN RESEARCH (Sterling, VA). DNA fragments were synthesized with an Applied Biosystems 392 DNA synthesizer (Applied Biosystems, Foster City, CA) and were purified either by HPLC, using SynChropak RPP (Eichrom Technologies, Darien, IL), or by gel-electrophoresis after deprotection.

Synthesis of 1-(2′-Deoxy-3′,5′-di-O-toluoyl-β-D-ribofuranosyl)-pyrrole-2-carbaldehyde. To a solution of pyrrole-2-carbaldehyde (368 mg, 4.4 mmol) in CH_3CN (56 mL) was added NaH (219 mg, 60% dispersion in mineral oil, 1.4 equiv), and the mixture was stirred at room temperature for 1 h. A quantity of 1-chloro-2-deoxy-3,5-di-O-toluoyl-α-D-erythropentofuranose (1.5 g, 4.4 mmol) was then added to the solution. The reaction mixture was stirred at room temperature for 1 h and was partitioned with EtOAc and H_2O. The organic layer was washed with H_2O, dried with Na_2SO_4, and evaporated in vacuo. The product (1.2 g, 60%) was purified from the residue by silica gel column chromatography (1% MeOH in CH_2Cl_2). ^{1}H NMR (270 MHz, $CDCl_3$) δ 2.33−2.44 (m, 7H, H2′, $CH_3 \times 2$), 2.89 (m, 1H, H2″), 4.55−4.67 (m, 3H, H4′, H5′, H5″), 5.56 (m, 1H, H3′), 6.23 (t, 1H, H1′, J = 3.6 Hz), 6.96−7.01 (m, 2H, pyrrole), 7.23 (dd, 4H, phenyl, J = 8.2, 19.4 Hz), 7.45 (s, 1H, pyrrole), 7.91 (dd, 4H, phenyl, J = 8.2, 19.4 Hz), 9.52 (s, 1H, −CHO); thin-layer chromatography (TLC) R_f = 0.34 (CH_2Cl_2:MeOH = 100:1 v/v). High-resolution mass spectroscopy (HRMS) (FAB, 3-NBA matrix) for $C_{26}H_{26}NO_6$ (M + 1): calcd, 448.1760; found, 448.1751.

Synthesis of 1-(2′-Deoxy-β-D-ribofuranosyl)-pyrrole-2-carbaldehyde. A quantity of 1-(2′-deoxy-3′,5′-di-O-toluoyl-β-D-ribofuranosyl)-pyrrole-2-carbaldehyde (100 mg, 0.22 mmol) was poured into saturated ammonia in methanol (4.0 mL), and the solution was stirred overnight at room temperature. The reaction mixture was evaporated in vacuo and separated with H_2O and EtOAc, and then the water layer was evaporated in vacuo. The product (35 mg, 74%) was purified by reversed-phase HPLC (Waters Microbond Sphere model C18, with a gradient from 5% to 50% (10 min) CH_3CN in H_2O). ^{1}H NMR (270 MHz, DMSO-d_6) δ 2.13 (m, 1H, H2′β), 2.26 (m, 1H, H2″α), 3.54 (m, 2H, H5′, H5″), 3.79 (m, 1H, H4′), 4.24 (m, 1H, H3′), 4.95 (t, 1H, 5′-OH, J = 5.3 Hz, D_2O exchange), 5.23 (d, 1H, 3′-OH, J = 4.1 Hz, D_2O exchange), 6.28 (t, 1H, pyrrole H4), 6.73 (t, 1H, H1′, J = 6.4 Hz), 7.06 (dd, 1H, pyrrole H3, J = 1.6, 3.8 Hz), 7.70 (bs, 1H, pyrrole H5), 9.50 (s, 1H, −CHO). ^{13}C NMR (68 MHz, DMSO-d_6) δ 41.71, 41.83, 41.96, 61.28, 61.34, 61.41, 70.09, 85.74, 85.84, 87.33, 110.02,

(16) McMinn, D. L.; Ogawa, A. K.; Wu, Y.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **1999**, *121*, 11585−11586.
(17) Ogawa, A. K.; Wu, Y.; McMinn, D. L.; Liu, J.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 3274−3287.
(18) Wu, Y.; Ogawa, A. K.; Berger, M.; McMinn, D. L.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 7621−7632.
(19) Ogawa, A. K.; Wu, Y.; Berger, M.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 8803−8804.
(20) Tae, E. L.; Wu, Y.; Xia, G.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2001**, *123*, 7439−7440.
(21) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 1001−1007.
(22) Morales, J. C.; Kool, E. T. *Biochemistry* **2000**, *39*, 12979−12988.
(23) Guckian, K. M.; Krugh, T. R.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 6841−6847.

125.16, 127.67, 127.77, 130.86, 179.21, 179.28. Electrospray ionization−mass spectroscopy (ESI-MS) for $C_{10}H_{13}NO_4$: calcd, 210.08 $[M-H]^-$; found, 210.05 $[M-H]^-$. UV−vis (in EtOH): $\lambda_{max} = 288$ nm ($\epsilon = 11\,230$), 260 nm ($\epsilon = 5610$); $\lambda_{min} = 220$ nm ($\epsilon = 670$). TLC $R_f = 0.21$ (CH_2Cl_2:MeOH = 20:1 v/v). The anomeric configuration of 1-(2′-deoxy-β-D-ribofuranosyl)-pyrrole-2-carbaldehyde was confirmed using nuclear Overhauser effects (NOEs) (see the Supporting Information). HRMS (FAB, 3-NBA matrix) for $C_{10}H_{14}NO_4$ (M + 1): calcd, 212.0923; found, 212.0922.

Synthesis of 5′-O-Dimethoxytrityl-1-(2′-deoxy-β-D-ribofuranosyl)-pyrrole-2-carbaldehyde. A quantity of 1-(2′-deoxy-β-D-ribofuranosyl)-pyrrole-2-carbaldehyde (100 mg, 0.47 mmol) was coevaporated with dry pyridine three times. The residue was dissolved in pyridine (4.7 mL) with 4,4-dimethoxytrityl chloride (160 mg, 0.47 mmol), and the solution was stirred at room temperature for 4 h. Water was then added to the solution and the product was extracted with EtOAc. The organic layer was washed with 5% $NaHCO_3$ three times, dried with Na_2SO_4, and evaporated in vacuo. The product was purified by silica gel column chromatography (1% MeOH in CH_2Cl_2) to give the dimethoxytrityl derivative (240 mg, 99%). ^{1}H NMR (270 MHz, $CDCl_3$) δ 2.27 (m, 1H, H2′), 2.53 (m, 1H, H2″), 3.41 (d, 2H, H5′, H5″, $J = 4.3$ Hz), 3.78 (s, 6H, $-OCH_3 \times 2$), 4.02 (m, 1H, H4′), 4.44 (m, 1H, H3′), 6.15 (m, 1H, pyrrole H4), 6.79−6.84 (m, 5H, DMTr, H1′), 6.95 (m, 1H, pyrrole H3), 7.21−7.46 (m, 10H, pyrrole H5, DMTr), 9.50 (s, 1H, −CHO). TLC $R_f = 0.51$ (CH_2Cl_2:MeOH = 20:1 v/v). HRMS (FAB, 3-NBA matrix) for $C_{31}H_{32}NO_6$ (M + 1): calcd, 514.2230; found, 514.2234.

Synthesis of 5′-O-Dimethoxytrityl-1-(2′-deoxy-β-D-ribofuranosyl)-pyrrole-2-carbaldehyde-3′-phosphoramidite. A quantity of 5′-O-dimethoxytrityl-1-(2′-deoxy-β-D-ribofuranosyl)-pyrrole-2-carbaldehyde (103 mg, 0.20 mmol) was coevaporated with pyridine and tetrahydrofuran (THF) three times each and was dissolved in THF (1.0 mL) with diisopropylethylamine (39 μL, 1.1 equiv) and 2-cyanoethyl-*N*,*N*-diisopropylamino-chlorophosphoramidite (50 μL, 1.1 equiv). The reaction mixture was stirred at room temperature for 2 h, and then MeOH (50 μL) was added. The solution was diluted with EtOAc/triethylamine (EtOAc/TEA) (10 mL, 20:1 v/v) and then was washed with 5% $NaHCO_3$ and saturated NaCl three times each. The organic layer was dried with Na_2SO_4 and evaporated in vacuo. The residue was purified by silica gel column chromatography (3:2 (v/v) hexane:CH_2Cl_2 ratio, containing 2% TEA) to give the amidite (147 mg, 100%). ^{1}H NMR (270 MHz, $CDCl_3$) δ 1.01−1.16 (m, 12H, $-CH(CH_3)_2 \times 2$), 2.23 (m, 1H, H2′), 2.39 (t, 1H, $-CH_2-$), 2.58−2.63 (m, 2H, H2″, $-CH_2-$), 3.25−3.61 (m, 5H, H5′, H5″, $-CH(CH_3)_2 \times 2$, $-CH_2-$), 3.77 (s, 7H, $-OCH_3 \times 2$, $-CH_2-$), 4.16 (m, 1H, H4′), 4.54 (m, 1H, H3′), 6.14 (m, 1H, pyrrole H4), 6.78−6.86 (m, 5H, DMTr, H1′), 6.94 (m, 1H, pyrrole H3), 7.16−7.44 (m, 9H, DMTr), 7.48 and 7.56 (dbs, 1H, pyrrole H5), 9.51 (s, 1H, −CHO). ^{31}P NMR (109 MHz, $CDCl_3$) δ 148.90, 149.63 (diastereoisomers). TLC $R_f = 0.32$ (hexane:CH_2Cl_2 = 2:3 v/v, containing 2% TEA). HRMS (FAB, 3-NBA matrix) for $C_{40}H_{49}N_3O_7P$ (M + 1): calcd, 714.3308; found, 714.3298.

Synthesis of 1-(2′-Deoxy-β-D-ribofuranosyl)-pyrrole-2-carbaldehyde-5′-triphosphate. To a solution of 1-(2′-deoxy-β-D-ribofuranosyl)-pyrrole-2-carbaldehyde (22 mg, 0.1 mmol) and a proton sponge (33 mg, 0.15 mmol) in trimethyl phosphate (500 μL) was added $POCl_3$ (13 μL, 1.3 equiv) at 0 °C.[24] The reaction mixture was stirred at 0 °C for 1.5 h. Tri-*n*-butylamine (119 μL, 5.0 equiv) was added to the reaction mixture, followed by 0.5 M bis(tributylammonium)pyrophosphate in a DMF solution (1.0 mL, 5.0 equiv). After 30 min, the reaction was quenched by the addition of 0.5 M triethylammonium bicarbonate (TEAB, 500 μL). The resulting crude solution was purified by DEAE Sephadex A-25 column chromatography (1.5 cm × 30 cm, eluted by a linear gradient of 50 mM to 1 M TEAB), and then by C18-HPLC (Synchropak RPP, Eichrom Technologies, eluted by a gradient of 0%−30% CH_3CN in 100 mM triethylammonium acetate). ^{1}H NMR (270 MHz, D_2O) δ 1.10 (t, 27H, $-CH_2CH_3$, $J = 7.3$ Hz), 2.34 (m, 2H, H2′, H2″), 3.02 (q, 18H, $-CH_2CH_3$, $J = 7.3$ Hz), 3.96−4.04 (m, 3H, H4′, H5′, H5″), 4.51 (m, 1H, H3′), 6.28 (dd, 1H, pyrrole H4, $J = 2.8$ Hz, 4.0 Hz), 6.77 (t, 1H, H1′, $J = 6.6$ Hz), 7.05 (dd, 1H, pyrrole H3, $J = 1.7$ Hz, 4.0 Hz), 7.59 (s, 1H, pyrrole H5), 9.25 (s, 1H, −CHO). ^{31}P NMR (109 MHz, D_2O) δ −22.91 (t, 1H, $J = 20.1$ Hz), −10.96 (d, 1H, $J = 20.1$ Hz), −10.40 (d, 1H, $J = 19.5$ Hz); ESI-MS for $C_{10}H_{16}NO_{13}P_3$: calcd, 449.98 $[M-H]^-$; found, 449.80 $[M-H]^-$.

Thermal Denaturation. The concentration of DNA fragments was determined from each hypochromicity, obtained by the complete degradation of the fragments with nuclease P1. The absorbance at 260 nm of the DNA fragments was monitored as a function of temperature (15−65 °C) on a Beckman model DU650 spectrophotometer. The duplexes of 5′-GGTAACNATGCG and 5′-CGCATN′GTTACC (N = **Q**, **A**, or **G**, and N′ = **Pa** or **T**) were dissolved in 10 mM sodium phosphate (pH 7.0), 100 mM NaCl, and 0.1 mM EDTA to give a duplex concentration of 5 mM. Melting temperature (T_m) values were calculated by the first derivatives of the melting curves.

Steady-State Kinetics. Steady-state kinetics for single-nucleotide insertions were performed according to the literature.[25,26] Primers were 5′-labeled, using [γ-^{32}P]ATP and T4 polynucleotide kinase. Primer−template duplexes (10 μM) were annealed in a buffer containing 100 mM tris-HCl (pH 7.5), 20 mM $MgCl_2$, 2 mM DTT, and 0.1 mg/mL bovine serum albumin by heating at 95 °C and slow cooling to 4 °C. The duplex solution (5 μL) was mixed with 2 μL of a solution containing the exonuclease-deficient Klenow fragment (Amersham USB, Cleveland, OH) diluted in a buffer containing 50 mM potassium phosphate (pH 7.0), 1 mM DTT, and 50% glycerol, and was incubated at 37 °C for more than 2 min. Reactions were initiated by adding 3 μL of a dNTP solution to the DNA−enzyme mixture at 37 °C. The amount of polymerase used (3−50 nM), the reaction time (1−22 min), and the gradient concentration of dNTP (0.6−2100 μM) were adjusted to give reaction extents of 25% or less. Reactions were quenched by adding 10 μL of a dye solution containing 89 mM trisborate, 2 mM EDTA, and 10 M urea, and the mixtures were immediately heated at 75 °C for 3 min. The products were analyzed on a 15%−20% polyacrylamide gel containing 7 M urea. The reaction extents were measured with a bio-imaging analyzer (Fuji model BAS 2500). Relative velocities (v_0) were calculated as the extents of the reaction divided by the reaction time and were normalized to the enzyme concentration (20 nM) for the various enzyme concentrations used. The kinetic parameters (K_m and V_{max}) were obtained from Hanes−Woolf plots of [dNTP]/v_0 against [dNTP]. Each parameter was averaged from 3−12 data sets.

Primer Extension Reaction. The 5′-labeled primer was annealed to the template in an annealing buffer, by heating at 95 °C and slow cooling to 4 °C. The duplex solution (5 μL) was mixed with 2 μL of solution A containing dNTP substrates, and polymerase reactions were started by adding 3 μL of solution B containing the polymerase. The reaction mixture was incubated at 37 °C and was terminated by adding 10 μL of the dye solution and heating at 75 °C for 3 min. The products were analyzed on a 15% polyacrylamide gel containing 7 M urea.

(A) KF Reactions. The annealing buffer contained 20 mM tris-HCl (pH 7.5), 14 mM $MgCl_2$, and 0.2 mM DTT. Solution A contained 50 μM dNTPs, and solution B contained the exonuclease-proficient Klenow fragment (TaKaRa, Tokyo) diluted in distilled water. The concentrations used for primer extension were as follows: primer−template, 200 nM; KF, 0.1 unit/μL; and dNTPs, 10 μM each; the incubation time for each was 5 min. Unit definition: One unit incorporates 10 nmol of total nucleotides into acid-insoluble products in 30 min at 37 °C using poly d(**A**−**T**) as the template-primer.

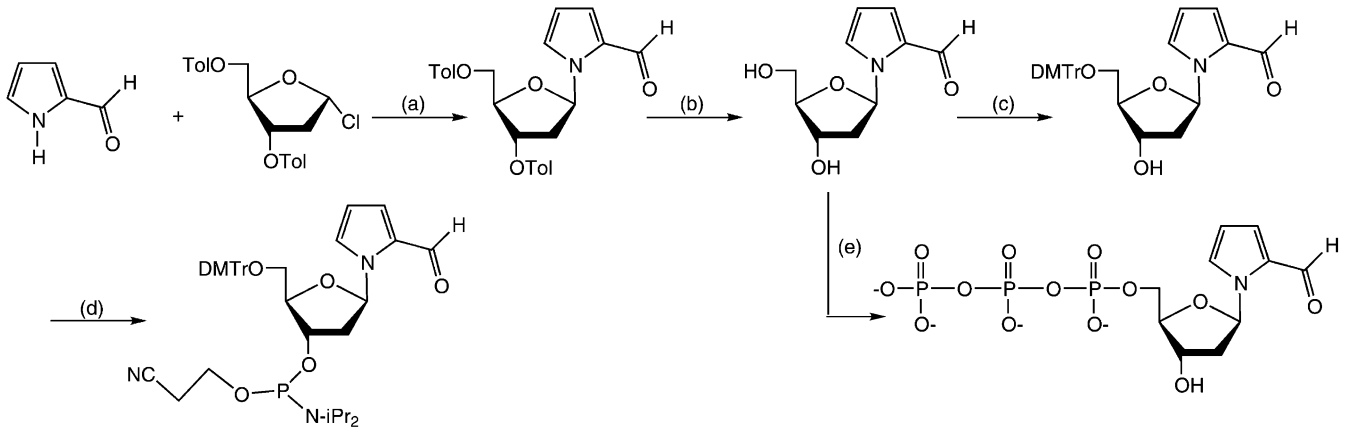
(B) Reverse Transcriptase of Avian Myeloblastosis Virus (AMV-RT) Reactions. Solution A contained 500 μM dNTPs diluted in an annealing buffer [50 mM tris-HCl (pH 8.3), 10 mM $MgCl_2$, 100 mM KCl, and 4 mM DTT], and solution B contained the reverse transcriptase

(24) Kovács, T.; Ötvös, L. *Tetrahedron Lett.* **1988**, *29*, 4525−4528.

(25) Petruska, J.; Goodman, M. F.; Boosalis, M. S.; Sowers, L. C.; Cheong, C.; Tinoco, I. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6252−6256.

(26) Goodman, M. F.; Creighton, S.; Bloom, L. B.; Petruska, J. *Crit. Rev. Biochem. Mol. Biol.* **1993**, *28*, 83−126.

Scheme 1 [a]



[a] Conditions: (a), NaH, CH_3CN, room temperature, 60%; (b), NH_3-MeOH, room temperature, 74%; (c), DMTr-Cl, pyridine, room temperature, 99%; (d), $ClP(N\text{-}iPr_2)(OCH_2CH_2CN)$, iPr_2EtN, THF, room temperature, quant.; (e), $POCl_3$, proton sponge, $(CH_3O)_3PO$, 0 °C, then tri-*n*-butylamine, bis(tributylammonium)pyrophosphate. Abbreviations: Tol, toluoyl; DMTr, dimethoxytrityl; iPr, isopropyl.

Table 1. Steady-State Kinetic Parameters for Insertion of Single Nucleotides into a Template−Primer Duplex by the Exonuclease-Deficient Klenow Fragment[a]

Primer 1, 5′-ACTCACTATAGGGAGGAAGA
Template 1, 3′-TATTATGCTGAGTGATATCCCTCCTTCT**N**TCTCGA

template (**N**)	nucleoside triphosphate	K_m[b] (μM)	V_{max}[d] (%·min^{-1})	efficiency, V_{max}/K_m (%·min^{-1}·M^{-1})
Q	**Pa**	240 (80)	30 (7)	1.3×10^5
A	**Pa**	520 (210)	29 (11)	5.6×10^4
G	**Pa**	460 (160)	0.23 (0.06)	5.0×10^2
C	**Pa**	n.d.[c]	n.d.[c]	
T	**Pa**	430 (270)	0.12 (0.05)	2.8×10^2
Q	**A**	91 (39)	1.2 (0.5)	1.3×10^4
Q	**G**	190 (80)	0.21 (0.06)	1.1×10^3
Q	**C**	690 (130)	7.7 (2.5)	1.1×10^4
Q	**T**	370 (110)	3.3 (0.9)	8.9×10^3
Q	**F**	70 (11)	15 (2)	2.1×10^5
A	**F**	170 (40)	24 (5)	1.4×10^5
G	**F**	1800 (1200)	0.52 (0.27)	2.9×10^2
iso**G**	**Pa**	n.d.[c]	n.d.[c]	
iso**G**	**C**	390 (120)	0.93 (0.23)	2.4×10^3
iso**G**	**T**	340 (100)	69 (10)	2.0×10^5
A	**T**	1.6 (0.7)	2.1 (0.7)	1.3×10^6

[a] Assays were performed at 37 °C for 1−20 min using 5 μM template−primer duplex, 5−50 nM enzyme, and 0.6−2100 μM nucleoside triphosphate in a solution (10 μL) containing 50 mM tris-HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM DTT, and 0.05 mg/mL bovine serum albumin. [b] Standard deviations are given in parentheses. [c] No inserted products were detected after an incubation of 20 min with 1500 or 2100 μM nucleoside triphosphate and 50 nM enzyme. [d] Values were normalized to the enzymatic concentration (20 nM) for the various enzyme concentrations used. Standard deviations are given in parentheses.

of Avian myeloblastosis virus (AMV-RT) (Life Sciences Inc., St. Petersburg, FL) diluted in the annealing buffer. The concentrations used for primer extension were as follows: primer−template, 150 nM; AMV-RT, 0.1 unit/μL; and dNTPs, 100 μM each; the incubation time for each was 20 min. Unit definition: One unit catalyzes the incorporation of 1 nmol of dTMP into an acid-insoluble product in 10 min at 37 °C using poly[(rA)·oligo(dT)] as a template−primer.

NMR Spectra. NMR experiments were performed on a Bruker model DRX600 spectrometer. Proton assignments were made using standard two-dimensional (2D) techniques, including NOESY (mixing times of 75, 150, 225, 300, and 375 ms), DQF−COSY, and NOESY with jump and return water suppression (mixing time of 150 ms). Data processing was done using Felix950 software (BIOSYM/Molecular Simulations). All structural restraints involving nonexchangeable protons were derived from NOESY data acquired at 15 °C with a mixing time of 75 ms. Restraints for exchangeable protons were derived from NOESY with jump and return water suppression data acquired at 3 °C with a mixing time of 150 ms. The backbone conformation and the sugar pucker were investigated at 15 °C using one-dimensional (1D) ^{31}P and DQF−COSY experiments, respectively.

Restraint Generation. Distance constraints were generated from the 75-ms-mixing-time NOE volumes with the pyrimidine H5/H6 cross peaks as a reference at 2.45 Å. All distances were calculated from the two-spin cross-relaxation approximation, with the usual r^{-6} dependence on the proton−proton distance. Limits of ±30% were placed about the calculated interproton distances.[27] A total of 232 restraints involving nonexchangeable protons were derived. Sixty exchangeable proton restraints were given an upper bound of 5 Å and a lower bound of 1.5 Å. Dihedral angle restraints were used to preserve a right-handed DNA helix.[28] Pseudorotational phase angle restraints were derived on the basis of the H1′−H2′/H2″ and H3′−H4′ correlations in the COSY spectrum. A total of 292 NOE distance restraints and 132 dihedral angle restraints were used for the calculations. Watson−Crick base pairs were identified using two criteria: the observation of an NH or NH_2 proton resonance significantly downfield-shifted and the observation of strong **G**−**C** NH−NH_2 or **A**−**T** H2−NH NOEs. In addition, 56 restraints were included to maintain the Watson−Crick pairing for base pairs 1−5 and 7−12.[23,29]

Molecular Dynamics. All structural calculations and analyses were conducted using the AMBER 6.0 program. The calculations used the AMBER force field in addition to the flat-bottom restraint potentials, with force constants of 50 kcal·mol^{-1}·Å^{-2} for the distance and 50

(27) Burkard, M. E.; Turner, D. H. *Biochemistry* **2000**, *39*, 11748−11762.
(28) Clore, G. M.; Oschkinat, H.; McLaughlin, L. W.; Benseler, F.; Happ, C. S.; Happ, E.; Gronenborn, A. M. *Biochemistry* **1988**, *27*, 4185−4197.
(29) Schmitz, U.; James, T. L. *Methods Enzymol.* **1995**, *261*, 3−44.

Table 2. Steady-State Kinetic Parameters for Insertion of Single Nucleotides into a Template−Primer Duplex by the Exonuclease-Deficient Klenow Fragment[a]

Primer 2, 5′-ACTCACTATAGGGAGCTTCT
Template 2, 3′-TATTATGCTGAGTGATATCCCTCGAAGAN̲AGAGCT

template (N̲)	nucleoside triphosphate	K_m[b] (μM)	V_{max}[c] (%· min^{-1})	efficiency, V_{max}/K_m (%·min^{-1}·M^{-1})
Pa	**Q**	140 (20)	110 (10)	7.9×10^5
Pa	**A**	370 (110)	53 (3)	1.4×10^5
Pa	**G**	330 (100)	1.9 (0.4)	5.8×10^3
Pa	**C**	1600 (700)	0.27 (0.07)	1.7×10^2
Pa	**T**	970 (140)	0.83 (0.10)	8.6×10^2
A	**Q**	200 (30)	20 (3)	1.0×10^5
G	**Q**	190 (10)	13 (1)	6.8×10^4
C	**Q**	280 (40)	17 (3)	6.1×10^4
T	**Q**	220 (60)	20 (5)	9.1×10^4
F	**Q**	50 (6)	86 (16)	1.7×10^6
F	**A**	76 (3)	91 (28)	1.2×10^6
F	**G**	530 (240)	0.18 (0.06)	3.4×10^2
Pa	iso**G**	1200 (500)	1.9 (0.6)	1.6×10^3
C	iso**G**	360 (60)	0.41 (0.09)	1.1×10^3
T	iso**G**	610 (110)	88 (11)	1.4×10^5
iso**C**	iso**G**	530 (80)	37 (8)	7.0×10^4
iso**C**	**Q**	320 (110)	5.7 (0.5)	1.8×10^4
iso**C**	**A**	740 (170)	7.4 (1.0)	1.0×10^4
iso**C**	**G**	490 (130)	0.66 (0.09)	1.3×10^3
T	**A**	3.1 (1.6)	8.7 (2.6)	2.8×10^6

[a] Assays were performed at 37 °C for 1−22 min using 5 μM template−primer duplex, 3−45 nM enzyme, and 0.6−2100 μM nucleoside triphosphate in a solution (10 μL) containing 50 mM tris·HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM DTT, and 0.05 mg/mL bovine serum albumin. [b] Standard deviations are given in parentheses. [c] The values were normalized to the enzymatic concentration (20 nM) for the various enzyme concentrations used. Standard deviations are given in parentheses.

Table 3. Proton Chemical Shifts for the **Pa**−**Q** Duplex[a]

	H8/H6	H5/Met/H2			H1′	H2′	H2″	H3′	H4′	amino	imino
C1	7.75	6.00	na	na	5.86	2.10	2.53	4.82	4.18	8.12/7.02	na
G2	8.08	na	na	na	6.03	2.80	2.88	5.10	4.48		13.00
C3	7.51	7.74	na	na	5.74	2.20	2.54	4.98	4.32	8.31/6.50	na
A4	8.47	1.63	na	na	6.36	2.81	3.01	5.13	4.54	7.65/6.37	na
T5	7.18	5.15	na	na	5.90	1.61	2.18	4.85	4.15	na	13.32
Pa6	7.60 (H5)	6.06 (H4)	6.32 (H3)	8.85 (H6)	6.53	2.69	2.69	5.04	4.41	na	na
G7	7.93	na	na	na	6.06	2.69	2.86	5.04	4.50		12.03
T8	7.38	1.35	na	na	6.08	2.21	2.64	4.96	4.25	na	13.80
T9	7.48	1.75	na	na	5.83	2.21	2.57	4.99	4.23	na	13.62
A10	8.43	7.59	na	na	6.30	2.83	2.97	5.14	4.54	7.76/6.21	na
C11	7.46	5.45	na	na	6.03	2.16	2.51	4.86	4.27	8.15/6.67	na
C12	7.71	5.77	na	na	6.29	2.35	2.35	4.64	4.12	8.18/6.89	na
G13	7.99	na	na	na	5.80	2.69	2.81	4.92	4.29		12.76
G14	7.95	na	na	na	6.11	2.74	2.90	5.09	4.51		na
T15	7.37	1.55	na	na	5.72	2.19	2.51	4.99	4.31	na	13.44
A16	8.37	7.09	na	na	6.04	2.86	3.01	5.18	4.53	7.61/6.27	na
A17	8.22	7.63	na	na	6.17	2.65	2.93	5.13	4.54	7.59/5.94	na
C18	7.29	5.33	na	na	5.56	1.88	2.26	4.87	4.17	7.95/6.81	na
Q19	8.43 (H2)	2.06 (Met)	5.89 (H8)	7.24 (H7)	6.16	2.78	2.93	5.12	4.46	na	na
A20	8.21	7.71	na	na	6.17	2.65	2.95	5.08	4.54	7.15/6.17	na
T21	7.16	1.39	na	na	5.80	2.07	2.46	4.95	4.23	na	13.42
G22	7.96	na	na	na	5.94	2.71	2.77	5.08	4.46		12.59
C23	7.46	5.51	na	na	5.87	2.03	2.45	4.93	4.29	8.35/6.59	na
G24	8.06	na	na	na	6.27	2.73	2.47	4.79	4.29		na

[a] Non-exchangeable proton chemical shifts were measured at 15 °C and referenced to the HDO signal at 4.88 ppm, and exchangeable proton chemical shifts were measured at 3 °C and referenced to the HDO signal at 5.00 ppm. Throughout the table, "na" denotes not applicable.

kcal·mol^{-1}·rad^{-2} for the torsion angle restraints. The starting structure generation and the visualization of the calculated structures were done using InsightII 98 software (BIOSYM/Molecular Simulations). The partial charges for **Pa** and **Q** were assigned by ab initio methods at the RHF/6-31G(d) level, using Gaussian 94 (Gaussian, Inc.). The partial charges used in the molecular dynamics simulation for the **Pa** base are the following: N1, −0.63; C2, 0.20; C3, −0.24; H3, 0.22; C4, −0.30; H4, 0.21; C5, 0.07; H5, 0.22; C6, 0.33; O6, −0.54; and H6, 0.16. The partial charges for the **Q** base are the following: N1, −0.70; C2, 0.31; H2, 0.22; N3, −0.58; C4, 0.12; C5, 0.63; N6, −0.61; C7, 0.06; H7, 0.20; C8, −0.32; H8, 0.20; C9, 0.09; C10Me, −0.49; and C7MeH, 0.19. Structures were determined by restrained molecular dynamics simulations and energy minimization. Fifty calculations were performed using B-form DNA as the starting structure, and a second set of 50 calculations was performed with an A-form starting structure. The two sets of 50 structures were averaged and subjected to an additional 100 iterations of steepest-descents energy minimization with full restraints to generate the final structures.

Results and Discussion

Synthesis of the 2′-Deoxyribonucleotide Derivatives of Pyrrole-2-carbaldehyde (Pa) and Thermal Stability of the DNA Duplexes Containing Pa. The synthesis of the ribonucleosides of pyrrole-2-carbaldehyde via pyrrole-2-carboxylic

acid derivatives was previously reported.[30] To facilitate the synthesis, we directly synthesized the 2′-deoxyribonucleoside of **Pa**, by the coupling reaction of pyrrole-2-carbaldehyde with 1-chloro-2-deoxy-3,5-di-O-toluoyl-α-D-erythropentofuranose, which was obtained as a single product in a 44% yield after deprotection and purification (Scheme 1). This type of direct glycosylation, via the sodium salt of pyrrole derivatives with the α-configuration of the 1-chloro-sugar, reportedly gives 2′-deoxy-β-nucleosides specifically. The nucleoside of **Pa** was characterized by ^{1}H NMR and ^{13}C NMR spectra and by mass spectrometry, and the production of the 2′-deoxy-β-nucleoside was confirmed (see the Supporting Information). The pattern of the anomeric proton signal was identical with that published for the 2′-deoxy-β-nucleosides of pyrrole derivatives, such as pyrrole-2-carbonitrile,[31] 3-nitropyrrole,[32] and pyrrolopyrimidines.[33,34] The ^{1}H NMR spectrum of the 2′-deoxyribonucleoside of **Pa** showed the characteristic triplet for H1′ at 6.73 ppm and the narrow multiplet for H2′ and H2″, which are patterns similar to those of other β-isomers of pyrrole derivatives.[31−34] Furthermore, NOEs between H1′ and H4′ and between H1′ and H2″ were observed, suggesting the β-isomer of the product. The nucleoside of **Pa** was converted to the amidite and to the triphosphate by conventional methods. The molar absorption coefficient of the triphosphate (14 000 at λ_{max} = 291 nm) was determined by a quantitative analysis of the phosphorus, after dephosphorylation of the triphosphate with calf intestine alkaline phosphatase (TaKaRa, Tokyo).

We first synthesized a trimer d(TPaT) to examine the side reactions of the **Pa**−amidite during DNA chemical synthesis. The coupling efficiency of the amidite of **Pa** was >98% on a DNA synthesizer (Applied Biosystems), and the production of the trimer was confirmed by mass spectrometry (see the Supporting Information). After synthesis and deprotection, the trimer was analyzed by HPLC, and no byproduct was observed after treatments with 3.5% dichloroacetic acid in dichloromethane at room temperature for 1 h, followed by concentrated ammonia at 60 °C for 7 h, or with 0.02 M iodine in a solution containing 20% pyridine, 70% THF, and 10% water at room temperature for 1 h, followed by concentrated ammonia at 60 °C for 7 h. The DNA fragments (12-mer and 35-mer) containing **Pa** were then synthesized.

The **Q**-**Pa** pair in the DNA duplex showed selective thermal stability. The stability of the DNA duplex (5 mM) of 5′-GGTAACQATGCG and 5′-CGCATPaGTTACC (T_m = 39.6 °C) was higher than that of a duplex containing an **A**−**Pa** pair (T_m = 34.2 °C) or a **G**−**Pa** pair (T_m = 36.0 °C), although the **Q**−**Pa** duplex was less stable than the natural **A**−**T** duplex (T_m = 49.7 °C). The thermal stability of the **Q**−**Pa** duplex was compared to that of the **Q**−**F**, **A**−**F**, and **G**−**F** duplexes. The T_m values of duplexes containing each of the **Q**−**F**, **A**−**F**, and **G**−**F** pairs are 41.1, 37.6, and 37.6 °C, respectively. The lower stability of the **Q**−**Pa** duplex, as compared to the **Q**−**F** duplex, may be because the stacking interaction of a five-membered ring, such as **Pa**, with neighboring bases is weaker than that of a six-membered ring, such as that observed in natural pyrimidine

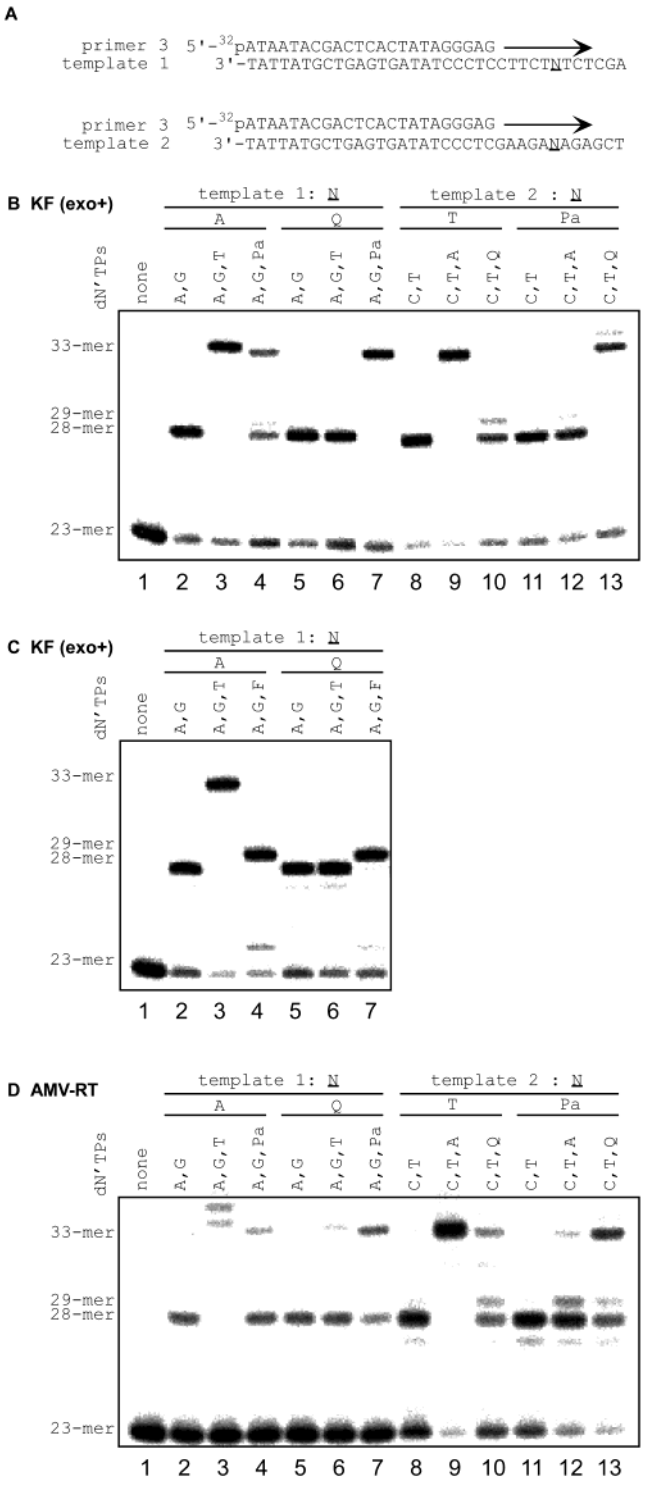


Figure 2. Primer extensions using natural and unnatural nucleotides in the templates and as substrates. (A) Sequences of the 23-mer primer and the 35-mer templates, templates 1 and 2, used in the reactions. (B and C) Autoradiograms of a denaturing polyacrylamide gel, showing primer extensions with the 3′ → 5′ exonuclease-proficient Klenow fragment [KF (exo$^+$)]; data were obtained at 37 °C, using 0.1 unit/μL enzyme, 200 nM primer−template duplex, 10 μM dNTP, and a reaction time of 5 min. (D) Autoradiogram of a denaturing polyacrylamide gel, showing primer extensions with Avian myeloblastosis virus reverse transcriptase (AMV-RT); data were obtained at 37 °C using 0.1 unit/μL enzyme, 150 nM primer−template duplex, 100 μM dNTP, and a reaction time of 20 min.

(30) Meade, E. A.; Wotring, L. L.; Drach, J. C.; Townsend, L. B. *J. Med. Chem.* **1992**, *35*, 526−533.
(31) Ramasamy, K.; Robins, R. K.; Revankar, G. R. *Tetrahedron* **1986**, *42*, 5869−5878.
(32) Bergstrom, D. E.; Zhang, P.; Toma, P. H.; Andrews, P. C.; Nichols, R. *J. Am. Chem. Soc.* **1995**, *117*, 1201−1209.
(33) Kazimierczuk, Z.; Cottam, H. B.; Revankar, G. R.; Robins, R. K. *J. Am. Chem. Soc.* **1984**, *106*, 6379−6382.
(34) Robins, M. J.; Robins, R. K. *J. Am. Chem. Soc.* **1965**, *87*, 4934−4940.

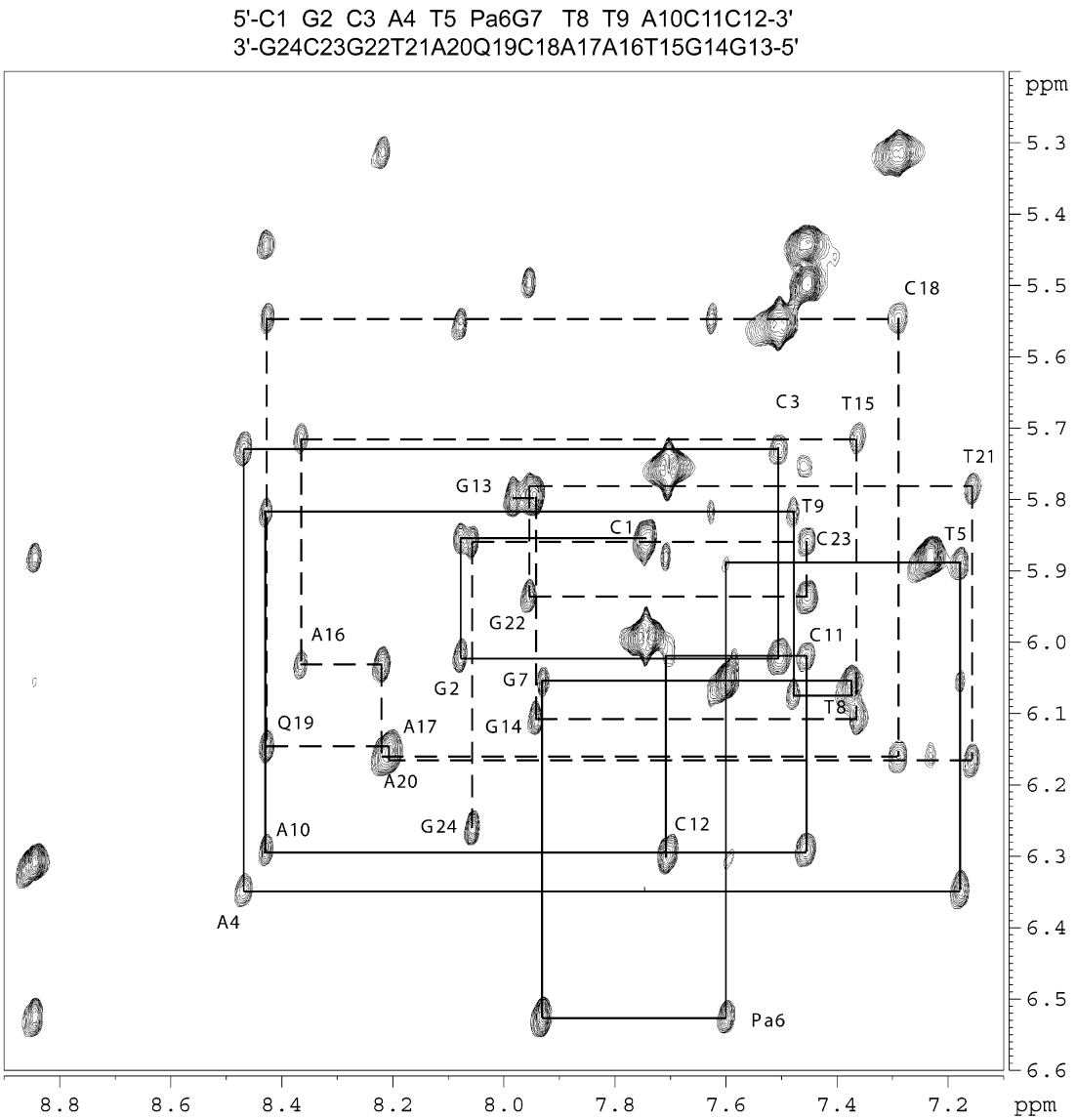


Figure 3. Base-to-H1′ region of a 75-ms NOESY spectrum obtained at 15 °C.

bases.[35] This instability is also present in the other base pairs with five-membered rings. For example, a DNA duplex (13-mer) containing an unnatural self-pair between 2-methylthiophenes (T_m = 52.1 °C) is less stable than that containing the **A**−**T** pair (T_m = 59.2 °C).[32] Interestingly, relative to the **A**−**F** duplex, the **A**−**Pa** duplex was destabilized by 3.4 °C, showing the poor fitting between **A** and **Pa**, in comparison to that between **A** and **F**.

Single-Nucleotide Insertion Experiments. The efficiency and selectivity of the **Q**−**Pa** pairing in replication were assessed by the steady-state kinetics of single-nucleotide insertion experiments.[25,26] The exonuclease-deficient Klenow fragment was used with combinations of a nucleoside triphosphate and a partially double-stranded template (35-mer) with a ^{32}P-labeled primer (20-mer), in which various bases in the template were adjacent to the 3′ end of the primer. The insertion of each substrate opposite each base in the template was analyzed by gel electrophoresis, and the kinetic parameters of each base pairing were determined (Tables 1 and 2). The substrate of **Pa** was efficiently inserted into DNA opposite **Q** in the template. The incorporation efficiency of **Pa** into DNA opposite **Q** (V_{max}/K_M = 1.3 × 10^5 %·min^{-1}·M^{-1}) was higher than that of **T** opposite **Q** (V_{max}/K_M = 8.9 × 10^3 %·min^{-1}·M^{-1}) as well as that of the previously developed **s**−**y** pair (the incorporation of **y** opposite **s**; V_{max}/K_M = 5.9 × 10^4 %·min^{-1}·M^{-1}).

We compared the incorporation efficiency and selectivity of the **Q**−**Pa** pairing with the **Q**−**F** pairing (Tables 1 and 2). The **T** analogue, **F**, efficiently paired with both **Q** and **A**. For example, the incorporation efficiencies of **F** opposite **Q** and **A** were V_{max}/K_M = 2.1 × 10^5 and 1.4 × 10^5 %·min^{-1}·M^{-1}, respectively. In contrast, the **Q**−**Pa** pairings in the insertion experiments, especially the incorporation of **Q** opposite **Pa** (V_{max}/K_M = 7.9 × 10^5 %·min^{-1}·M^{-1}), were more efficient than the noncognate pairing of the substrate **A** opposite **Pa** (V_{max}/K_M = 1.4 × 10^5 %·min^{-1}·M^{-1}). This suggests that the fitting of the five-membered-ring base **Pa** with the hydrophilic natural purine bases is inferior, in compairson to that of the six-membered-ring bases, such as **F** and the natural pyrimidines, **C** and **T**.

The **Pa**−**Q** pairing in replication showed the highest selectivity, relative to the noncognate pairings with the natural bases.

(35) Kool, E. T. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 1−22.

Table 4. Structural Statistics for the **Pa**−**Q** Duplex

Number of Restraints		
	total	480
distance restraints	intraresidue	176
	interresidue	116
	exchangeable	60
	non-exchangeable	232
	hydrogen bonding	56
dihedral angle restraints		132
Violations of Experimental Restraints in the Final Structure When:[a]		
A-form DNA was used as the starting structure	distance violations (0.0.1 Å)	⟨SA⟩[b]
	total	0
	intraresidue	0
	interresidue	0
	dihedral violations (>2°)	0
B-form DNA was used as the starting structure	distance violations (0.0.1 Å)	⟨SA⟩[b]
	total	0
	intraresidue	0
	interresidue	0
	dihedral violations (>2°)	0
Atomic rms Differences (Å)[c]		
final(A-form starting) and final(B-form starting)		2.44
final(A-form starting) and final(B-form starting) (residues 6 and 19)		0.18
final(B-form starting) and B-form starting structure		4.23

[a] Not including the residues of the 3′- or 5′-terminus (residues 1, 12, 13, and 24). [b] Energy-minimized average structure from 50 calculations. [c] Average pairwise rms differences were calculated using energy-minimized average structures.

The efficiency of **Pa** pairing opposite **Q** was 10−118 times higher than those of the natural bases opposite **Q**. Similarly, the incorporation efficiency of **Q** opposite **Pa** was 6−4600 times higher than those of the natural bases opposite **Pa**. In contrast, the incorporation efficiencies of **Pa** or **Q** opposite the natural bases were lower (>23 times) than those of both **T** opposite **A** and **A** opposite **T**. In this context, the incorporation of **Pa** opposite **A** was somewhat high ($V_{max}/K_M = 5.6 \times 10^4$ %·min^{-1}·M^{-1}), but the efficiency was much lower than that of the incorporation of **T** opposite **A** ($V_{max}/K_M = 1.3 \times 10^6$ %·min^{-1}·M^{-1}), suggesting that **T** might be predominantly incorporated opposite **A**. Thus, the **Pa**−**Q** pairing exhibited specificity and may work with the natural base pairs in replication.

To determine the utility of the **Q**−**Pa** pair with a combination of other unnatural base pairs as the third and fourth pairs, we performed the cross comparisons of the incorporation selectivity between the **Q**−**Pa** and the hydrophilic iso**G**−iso**C** pairs. The incorporation by all combinations among **Q**, **Pa**, iso**G**, iso**C** (5-methylisocytosine), and natural bases as substrates and template bases was conducted using the Klenow fragment (Tables 1 and 2), except for the incorporation of d(iso**C**)TP, because of the chemical instability in the assay system. The iso**G**−**Pa** pairing was inferior to the **Q**−**Pa** and iso**G**−iso**C** pairings, and in particular, the incorporation of **Pa** (ranging in concentration from 0.06 mM to 1.5 mM) opposite iso**G** was not observed. Although the incorporation of **Q** opposite iso**C** showed relatively high efficiency ($V_{max}/K_M = 1.8 \times 10^4$ %·min^{-1}·M^{-1}), the incorporation was less efficient than that of iso**G** opposite iso**C** ($V_{max}/K_M = 7.0 \times 10^4$ %·min^{-1}·M^{-1}). This high efficiency of the **Q**−iso**C** pairing may be due to the favorable shape complementarity between **Q** and iso**C**. Similarly, the incorporation of iso**G** opposite **T** was very high ($V_{max}/K_M = 1.4 \times 10^5$ %·min^{-1}·M^{-1}), but this incorporation was less effective than that of **A** opposite **T** ($V_{max}/K_M = 2.8 \times 10^6$ %·min^{-1}·M^{-1}). Thus, for practical use, the selective incorporation of iso**G** into DNA opposite iso**C** can be combined with the **Q**−**Pa** pairing.

Primer Extension. Next, we examined the extension after the incorporation of **Pa** or **Q** into the primers by the exonuclease-proficient (exo$^+$) Klenow fragment and AMV-RT. The primer extension reactions with combinations of the several substrates and templates were conducted using the duplex DNA between a 23-mer primer and 35-mer templates, template 1 for the incorporation of **Pa**, and template 2 for the incorporation of **Q** (Figure 2A). The extension was analyzed by gel electrophoresis, as shown in Figure 2B−D. The full-extension products were observed as 33-mer fragments, because of the absence of a substrate (dCTP for template 1 and dGTP for template 2) opposite the bases at position 34 in the templates.

The extension including the **Q**−**Pa** pairing at position 28 by the Klenow fragment (exo$^+$) was observed with high efficiency and selectivity (Figure 2B). There were no terminated products after the incorporations of **Pa** opposite **Q** and of **Q** opposite **Pa**, and the 33-mer products were obtained (Figure 2B, lanes 7 and 13). In contrast, the extension after the incorporation of **Pa** opposite **A** was much less efficient than that of the **T**−**A** and **Pa**−**Q** pairings (see Figure 2B, lanes 3, 4, 7, and 13), and the extensions involving the incorporations of **T** opposite **Q**, **Q** opposite **T**, and **A** opposite **Pa** were paused at positions 28 and/or 29 (see Figure 2B, lanes 6, 10, and 12).

Interestingly, the extension involving the **Q**−**F** pairing by the Klenow fragment (exo$^+$) was much less effective (Figure 2C). The substrate of **F** was efficiently incorporated opposite **Q** at position 28 in the template, but the extension after the incorporation did not proceed (Figure 2C, lanes 4 and 7). It is known that the minor groove interaction between **F** in the primer and the Klenow fragment is too small to process the subsequent extension effectively, although the **Q**−**F** pair is relatively stable against the exonuclease activity of the Klenow fragment.[7,21,22] Thus, the **Q**−**Pa** pair has more efficient processivity than that of the **Q**−**F** pair.

The substrate of **F** was not recognized by AMV-RT and was not incorporated into DNA.[21] In contrast, the substrate of **Pa** was incorporated into DNA opposite **Q** by AMV-RT, and the extension proceeded after the incorporation of **Pa** into the primer

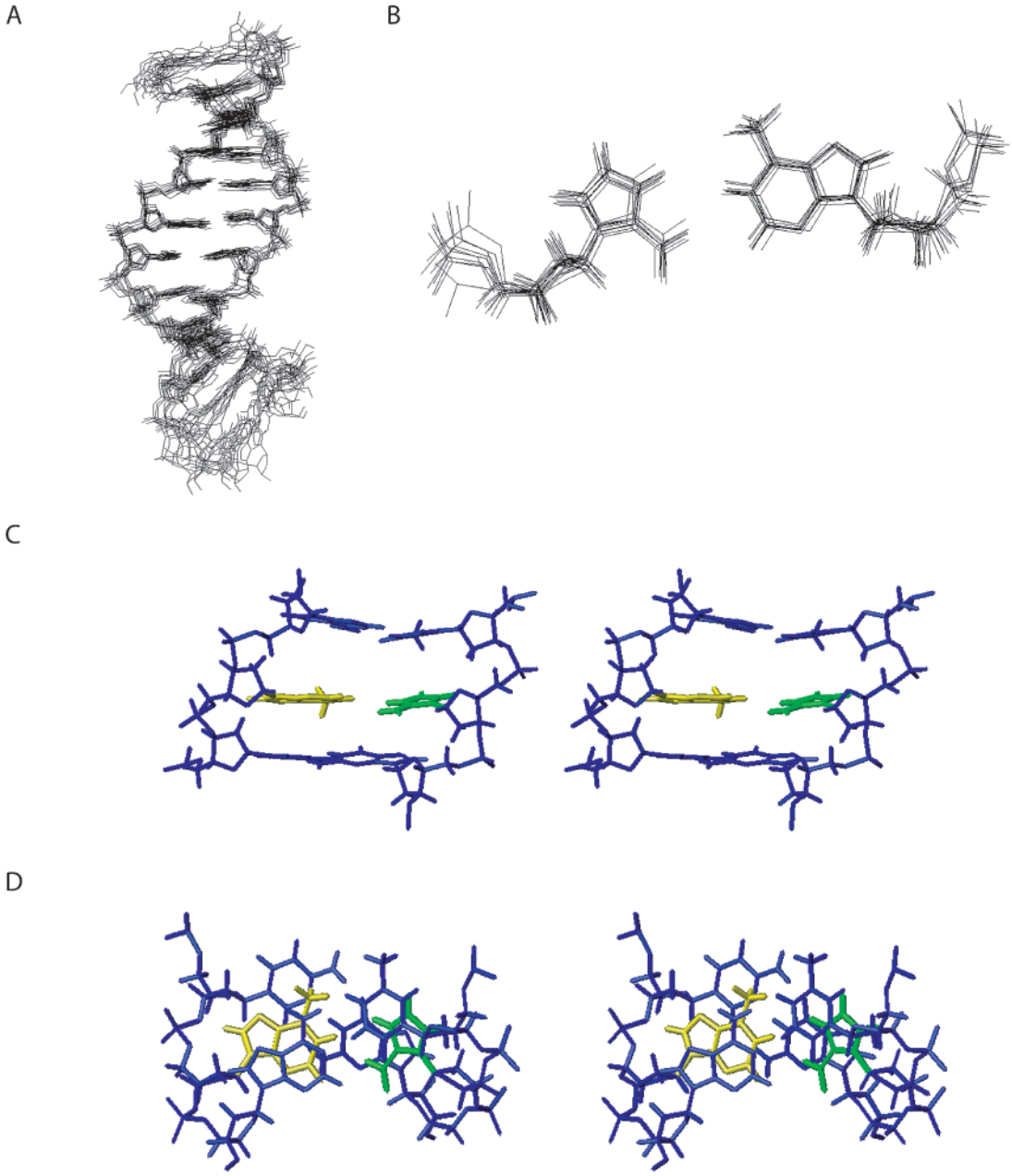


Figure 4. Three-dimensional structures of the **Q**−**Pa** duplex. Depiction A shows the superimposition of 10 randomly selected final structures after restrained molecular dynamics and energy minimization; five started from B-form DNA, whereas the other five started from A-form DNA. Depiction B shows the superimposition of the **Pa**−**Q** pairs in the 10 structures shown in depiction A. Depiction C shows a stereo view from the major groove of the central three base pairs for the **Pa**−**Q** duplex, showing the Pa6 (green) and Q19 (yellow) pair, as well as their immediate neighbors (blue). Depiction D shows a cutaway view along the helical axis.

(Figure 2D, lane 7). Similarly, the extension proceeded after the incorporation of **Q** opposite **Pa** (Figure 2D, lane 13). These results of the extensions by AMV-RT, as well as the Klenow fragment (exo$^+$), show that the aldehyde group of **Pa** was well recognized by the DNA polymerases, in comparison with the fluoro group of **F**.

Structural Analysis. To assess the steric fitting of the bases in the duplex, we determined the geometry of the **Q**−**Pa** pair in a DNA duplex by NMR. Structural studies were performed at 600 MHz in an aqueous solution using the **Q**−**Pa** containing duplex, 5′-C1G2C3A4T5Pa6G7T8T9A10C11C12 and 5′-G13G14T15A16A17C18Q19A20T21G22C23G24. The same sequences of the **Z**−**F** duplex used for the structural analysis were employed to compare the base-pair fitting of the **Q**−**Pa** pair with the **Z**−**F** pair.[23] Nonexchangeable proton signals were assigned by standard two-dimensional techniques, including DQF−COSY and NOESY (Table 3). The chemical shifts of the imino protons for the **G**−**C** and **A**−**T** base pairs were consistent with the formation of Watson−Crick base pairs. The cross peaks in a NOESY spectrum recorded in an H_2O solution also confirmed the formation of Watson−Crick base pairs in the duplex. The NOE connectivities of the base to H1′ (Figure 3) to H2′/2″, and to H3′, were observed throughout both strands, showing the continuous stacking between the bases in the duplex. The observed base to H1′ intensities indicated that all the bases adopted an anti conformation. In addition, the resolvable DQF−COSY cross-peak patterns indicated an S (C2′-endo) sugar pucker. The ^{31}P NMR spectrum showed that all

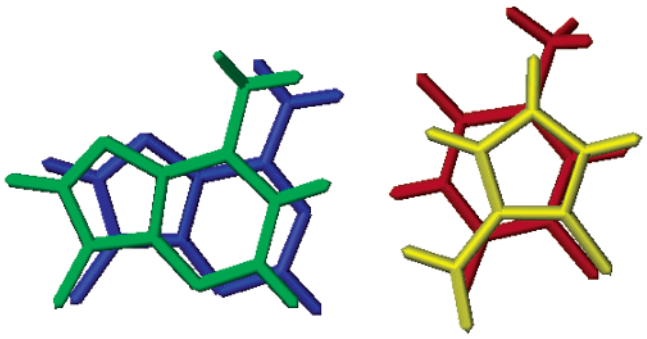


Figure 5. Superimposition of the **Q**−**Pa** pair (green−yellow) on the **A**−**T** pair (blue−red) in the canonical B-form conformation.

the phosphorus resonances resided within a 1 ppm range for the duplex, which is consistent with a B-form DNA structure.

Several NOEs were observed among the **Q**−**Pa** pair and the neighboring bases. The A20(H2) and T5(H6) protons exhibit NOEs to the Pa6(H6) and Pa6(H4) protons, respectively. In addition, the T5(H3) proton exhibits NOEs to the Pa6(H6,H4), and Q19(H8) protons, and the G7(H1) proton exhibits NOEs to the Pa6(H6,H3) and Q19(H7,H8) protons. Similarly, NOEs between Q19(H7) and Pa6(H6,H3), between Q19(H8) and Pa6-(H6, H3), and between Q19(Me) and Pa6(H6,H3,H4) were observed. These NOE connectivities are consistent with a predominant conformation, in which both Pa6 and Q19 are stacked within the DNA helix.

The structural determination of the duplex containing the **Q**−**Pa** pair was started from both A-form and B-form 12-mer duplexes; a series of simulated annealing calculations was performed using the AMBER force field with 292 distance restraints and 132 dihedral-angle restraints (Table 4). The convergence of the structures is illustrated in Figure 4A and B, and the stereoviews of the central three base pairs for the **Q**−**Pa** duplex that were averaged and subjected from the 100 structures are shown in Figure 4C and D. The duplex adopts the B-form structure. The sugar puckers of Pa6 and Q19 are both of the S-type, and both of the bases adopt anti glycosidic orientations, stacking with neighboring bases. The duplex seems to bend as an entire unit, and both the Q19 and Pa6 bases are slightly tilted to each opposite direction. However, the NMR signals derived from the **Q**−**Pa** pair are relatively broad, suggesting that the base pair is flexible in solution. As shown in Figure 4A, the fluctuations of each structure are relatively large, especially at both termini, and are larger than those of the related **Z**−**F** duplex obtained by NMR.[23] These may be due to the weak stacking stability of **Pa**, which renders the structural flexibility in a solution.

The cutaway view along the helical axis shows that the actual geometry of the **Q**−**Pa** pair (Figure 4B and D) is similar to our expected structure, as shown in structure D in Figure 1. The conformation of the **Q**−**Pa** pair closely resembles that of the Watson−Crick base pairs (Figure 5). The oxygen of the aldehyde group of **Pa** is located in a slightly different place than in the 2-keto group of **T**. The polymerases, such as the Klenow fragment and AMV-RT, must tolerate this difference for the minor groove interaction, or the flexibility of the **Pa** base may be favorable for the recognition.

Conclusion

As a substrate and a base in templates, the hydrophobic base combined with the aldehyde group, and the five-membered ring, can be recognized by DNA polymerases and function in replication. In addition, the unique specificity of each of the **Q**−**Pa** and **Q**−**F** pairings in replication (where **Q** is 9-methylimidazo[(4,5)-b]pyridine, **Pa** is pyrrole-2-carbaldehyde, and **F** is 2,4-difluorotoluene) shows that the selectivity of an unnatural base pair can be increased by fine-tuning the shape complementarity. In contrast to the specific **Q**−**F** and **A**−**F** pairing, our biological and structural studies show that the **Q**−**Pa** pairing is favored over the **A**−**Pa** pairing, because of the better shape-fitting between **Q** and **Pa** and the weak stacking ability of **Pa**, which destabilizes the pairing with the water-solvated **A**. The framework of the **Pa** base, as well as that of other five-membered-ring bases,[36−38] would be useful for the further development of unnatural base pairs, and studies of practical uses of the **Q**−**Pa** pairing are in progress.

Supporting Information Available: NMR, NOE, and MS data for the characterization of the nucleoside of **Pa**; MS data of the trimer of d(TPaT) (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA028806H

(36) Berger, M.; Luzzi, S. D.; Henry, A. A.; Romesberg, F. E. *J. Am. Chem. Soc.* **2002**, *124*, 1222−1226.
(37) Hirao, I.; Mitsui, T.; Fujiwara, T.; Kimoto, M.; To, T.; Okuni, T.; Sato, A.; Harada, Y.; Yokoyama, S. *Nucleic Acids Res. Suppl.* **2001**, *1*, 17−18.
(38) Jiang, X.-J.; Kalman, T. I. *Nucleosides Nucleotides* **1994**, *13*, 379−388.