Published online February 2, 2004

728–735 Nucleic Acids Research, 2004, Vol. 32, No. 2
DOI: 10.1093/nar/gkh241

PCR amplification of DNA containing non-standard base pairs by variants of reverse transcriptase from Human Immunodeficiency Virus-1

A. Michael Sismour, Stefan Lutz[1], Jeong-Ho Park[2], Michael J. Lutz[3], Paul L. Boyer[4], Stephen H. Hughes[4] and Steven A. Benner*

Departments of Chemistry and Anatomy and Cell Biology, University of Florida, Gainesville, FL 32611-7200, USA, [1]Department of Chemistry, Emory University, Atlanta, GA 30322, USA, [2]Department of Chemical Technology, Hanbat National University, Taejon 305-719, South Korea, [3]Novartis Pharma, Basel CH-4002 Switzerland and [4]HIV Drug Resistance Program, National Cancer Institute, NCI-Frederick, Frederick, MD 21702-1201, USA

Received June 25, 2003; Revised September 22, 2003; Accepted December 22, 2003

ABSTRACT

As the next step towards generating a synthetic biology from artificial genetic information systems, we have examined variants of HIV reverse transcriptase (RT) for their ability to synthesize duplex DNA incorporating the non-standard base pair between 2,4-diaminopyrimidine (pyDAD), a pyrimidine presenting a hydrogen bond 'donor–acceptor–donor' pattern to the complementary base, and xanthine (puADA), a purine presenting a hydrogen bond 'acceptor–donor–acceptor' pattern. This base pair fits the Watson–Crick geometry, but is joined by a pattern of hydrogen bond donor and acceptor groups different from those joining the GC and AT pairs. A variant of HIV-RT where Tyr 188 is replaced by Leu, has emerged from experiments where HIV was challenged to grow in the presence of drugs targeted against the RT, such as L-697639, TIBO and nevirapine. These drugs bind at a site near, but not in, the active site. This variant accepts the pyDAD-puADA base pair significantly better than wild type HIV-RT, and we used this as a starting point. A second mutation, E478Q, was introduced into the Y188L variant, in the event that the residual nuclease activity observed is due to the RT, and not a contaminant. The doubly mutated RT incorporated the non-standard pair with sufficient fidelity that the variant could be used to amplify oligonucleotides containing pyDAD and puADA through several rounds of a polymerase chain reaction (PCR) without losing the non-standard base pair. This is the first time where DNA containing non-standard base pairs with alternative hydrogen bonding patterns has been amplified by a full PCR. This work also illustrates a research strategy that combines *in clinico* pre-evolution of proteins followed by rational design to obtain an enzyme that meets a particular technological specification.

INTRODUCTION

Artificial genetic systems, the targets of organic chemists for more than a decade (1–4), have recently attracted attention in the scientific literature (5–7) and the science press (8,9). One class of these retains the two rules of complementarity that characterize standard Watson–Crick base pairing: (i) size complementarity, whereby a large purine pairs with a small pyrimidine, and (ii) hydrogen bonding complementarity, whereby hydrogen bond donors complement hydrogen bond acceptors (10).

With three hydrogen bonds joining the paired nucleobases, 12 nucleobases and six mutually exclusive hydrogen bonding patterns are possible (Fig. 1). We have prepared these as part of an Artificially Expanded Genetic Information System (AEGIS). Diagnostics products based on AEGIS recently received approval from the Food and Drug Administration for clinical use to monitor viral load in patients living with HIV and Hepatitis C (11).

Incorporating an artificial genetic system of this type into a living cell is now a widely recognized goal (12). To achieve this goal requires, however, a molecular biology for the artificial genetic system that exploits some enzymes of contemporary life forms. These have evolved for billions of years to accept A, T, C and G, not non-standard components of an artificial genetic alphabet.

Many enzymes work well with AEGIS components, including kinases, ligases and even ribosomes (13). Polymerases, in contrast, have proven to accept non-standard components of DNA only inefficiently, judging by rate, processivity, fidelity or some combination of these (14). These inefficiencies need not be apparent with standing start experiments, where the non-standard triphosphate is the first nucleotide to be added to a primer, or a running start experiment, where the polymerase adds standard nucleotides before it is challenged to incorporate a non-standard nucleotide.

*To whom correspondence should be addressed. Tel: +1 352 392 7773; Fax: +1 352 392 7918; Email: benner@chem.ufl.edu

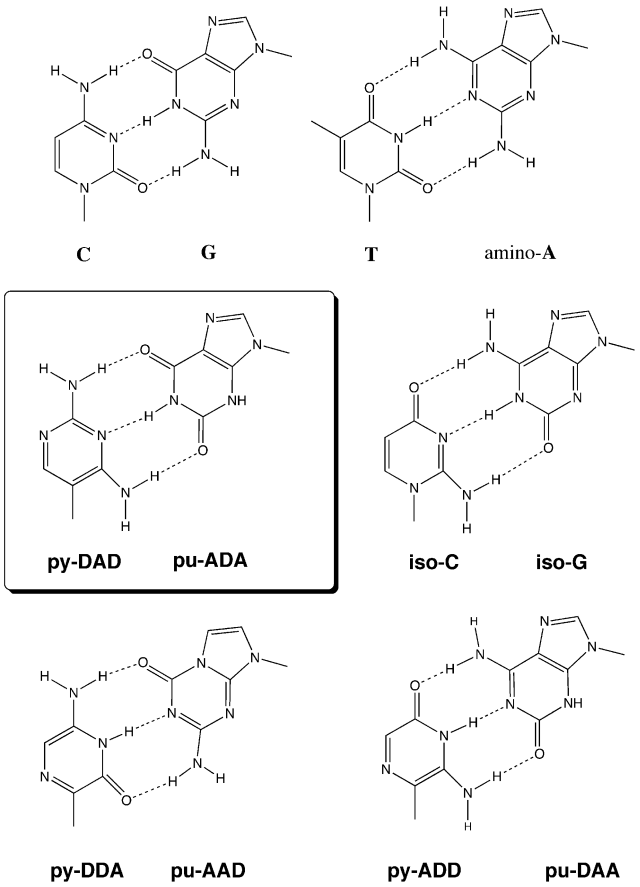

Figure 1. An artificially expanded genetic alphabet.

Inefficiencies do become apparent, however, when non-standard components are incorporated into a polymerase chain reaction (PCR). Here, infidelity that results in the replacement of a non-standard nucleotide by a standard nucleotide can result in the loss of the non-standard base. This is especially true if a strand containing an AEGIS component is amplified less efficiently by even a small degree.

For these reasons, the next step in the development of artificial genetic systems must improve the interaction between polymerases and unnatural nucleotides. Despite several excellent crystal structures for polymerases (15), neither the resolution of the structure nor chemical theory is adequate to guide the engineering of a polymerase from one that has 95% fidelity with an AEGIS component, for example, to one that has 99% fidelity.

Polymerases in the natural world frequently adapt to a changing environment. For example, reverse transcriptase (RT) from human immunodeficiency virus-1 (HIV-1) is the target of several anti-AIDS drugs that are not nucleotide analogs, but rather bind in a region near the active site. These include L-697639, TIBO and nevirapine (16). When placed in contact with these drugs, the HIV-RT undergoes rapid adaptive evolution, changing amino acid side residues not in direct contact with the substrates, but in contact with amino acids that are. This represents a more delicate modification of the structure than would be had by directed mutation in the active site itself (17,18).

Wild type HIV-RT accepts the pyDAD-puADA nucleobase pair with considerable pausing (19). We reasoned, therefore, that RT variants evolving in response to non-nucleosidic inhibitors should retain high levels of efficiency (the virus survives), but might handle AEGIS components differently from native RT, and that these differences might be advantageous.

To pursue this approach, we examined variants of HIV-1 RT that confer resistance to several RT inhibitors for their ability to incorporate the pu(ADA)-py(DAD) base pair. These include AZT-21, a variant of RT that contains five substitutions (M41L, D67N, K70R, T215Y, K219Q) that arises in HIV-1 selected resistance to AZT (20), Y181I and Y188L, which arise when HIV-1 is selected for resistance to L-697666, TIBO and nevirapine (21), and M184V, which confers resistance to the nucleoside analogs ddI, ddC, β-L-(–)-2′-deoxy-3′-thiacytidine (3TC), and β-(–)-2′-deoxy-5-fluoro-3′-thiacytidine (FTC) (22).

We report here that the Y188L variant, when modified with a second amino acid replacement, replicates DNA containing AEGIS components with high fidelity.

MATERIALS AND METHODS

Synthesis of non-standard nucleosides

2,4-Diamino-5-(β-D-ribofuranosyl)pyrimidine (pyDAD) was synthesized via a route adapted from Chu *et al*. (23), converted to the 2′-deoxygenated nucleoside analog via the route described previously (24). 5′-Dimethoxytrityl-2′-deoxyxanthosine with both heterocyclic ring oxygens protected as p-nitrophenylethyl ethers was prepared by a procedure adapted from Van Aerschot *et al*. (25,26). Both were then converted to the phosphoramidite suitable for automated DNA synthesis (27).

To prepare 2′-deoxyxanthosine-5′-triphosphate d(puADA)TP, 2′-deoxyguanosine triphosphate (sodium-salt, 10 mg, 16.7 mmol) was dissolved in water (220 µl) containing sodium nitrite (10 mg, 80 mmol). A mixture of HCl (8.7 µl, 2 M) and acetic acid (glacial, 25 µl) was added and the sample incubated at room temperature for 3 h. The reaction was quenched with Tris base (400 µl, 1 M). The raw material can be stored at –20°C before being purified by RP-HPLC [Nova Pak C-18 Radial Pak cartridge (Waters), 25 × 100 mm, TEAA (100 mM, pH 7), linear gradient to 10% acetonitrile over 25 min]. The combined product fractions were lyophilized and the residue was dissolved in Tris–HCl (2 ml, 10 mM, pH 7.0). The yield of 16 was determined by UV absorbance (4.2 mg, 42%, 247/277 nm, ε = 10 000/9100 $M^{-1}cm^{-1}$). The purity of the material was >97% as determined by analytical RP-HPLC and anion-exchange HPLC [Macrosphere 300 A WAX 7 U (Alltech, Deerfield IL)] 4.6 × 250 mm; solvent A = water; solvent B = TEA-bicarbonate (0.8 M, pH 7.2); curved gradient (7) from 1 to 50% B in 15 min]. The triphosphate d(pyDAD)TP was synthesized via published procedures from the nucleoside (28).

Oligonucleotide synthesis: primers and templates

The oligonucleotide sequences used in this work are listed in Table 1. Oligonucleotides bearing non-standard bases were prepared by 'trityl off' solid-phase synthesis using an Applied

Table 1. Oligonucleotides used in primer extensions and PCRs

P1-RS	5′-GCG AAT TAA CCC TCA CTA AAG-3′
P2-RS	5′-GCG TAA TAC GAC TCA CTA TAG-3′
P1-SS	5′-GCG AAT TAA CCC TCA CTA AAG AAC G-3′
P2-SS	5′-GCG TAA TAC GAC TCA CTA TAG ACG A-3′
P2-C6	5′-ATGCA-C6C6-GCG TAA TAC GAC TCA CTA TAG-3′ (for PCR)
P2-Rev	5′-GCG AAT TAA CCC TCA CTA AAG-3′
P3	5′-CAG GAA ACA GCT ATG ACG-3′
T1	5′-GCGTAATACGACTCACTATAGACGTTCGTTCTTTAGTGAGGGTTAATTCGC-3′
T2	5′-GCGAATTAACCCTCACTAAAGTACGTTCGTCTATAGTGAGTCGTATTACGC-3′
T1-puADA	5′-GCGTAATACGACTCACTATAGACGT(puADA)CGTTCTTTAGTGAGGGTTAATTCGC-3′
T2-pyDAD	5′-GCGAATTAACCCTCACTAAAGTACG(pyDAD)TCGTCTATAGTGAGTCGTATTACGC-3′
T3	5′-CGTCATAGCTGTTTCCTGGTCC(puADA)CGCATTGCTG-3′

C6 refers to a linker that contains 3 U of polyethyleneglycol, incorporated to permit the separation of the two product strands following PCR. puADA is a nucleotide bearing the xanthine nucleobase. pyDAD is the nucleotide bearing a 2,4-diaminopyrimidine nucleobase.

Biosystems automated DNA synthesizer from the β-cyanoethyl protected phosphoramidites. They were purified by PAGE (12–20%). Those oligonucleotides containing only standard nucleotides were obtained commercially from Integrated DNA Technologies (Coralville, IA).

Enzyme expression and purification

The HIV-RTs were expressed as p66/p51 heterodimers using a plasmid that coexpresses the p66 coding region of the HIV-1 RT variant, with a hexahistidine tag on the C-terminus, and HIV-1 protease. The expression is induced by the addition of IPTG, and a polycistronic messenger containing both the RT and protease coding region is produced. In *Escherichia coli*, the protease cleaves the p66 homodimer to yield the p66/p51 heterodimer with a hexahistidine tag only on the p66 subunit (29). The enzymes were isolated by the procedure of Boyer *et al*. (29).

Determining the catalytic activity of the enzymes

Enzyme activity was determined by incorporation of [^{3}H]-TTP into a poly(rA)-oligo(dT) template (30–33). All reactions were carried out at 37°C in a water bath. An aliquot of the RT (1–2 μl) was incubated in HIV-RT buffer (Tris–HCl 50 mM, pH 7.2, $MgCl_2$, 5 mM, KCl 100 mM, DTT, 1 mM, EDTA 0.5 mM) in the presence of 5 μg poly(rA)-oligo$(dT)_{12-18}$ (Pharmacia) and [^{3}H]TTP (25 μM, 6000 c.p.m./pmol, Amersham; concentration adjusted with 1 mM TTP). Four aliquots (20 μl) were taken over a 12 min period and quenched with EDTA (10 μl, 0.5 M, pH 8). The quenched reaction mixture (20 μl) was applied to 2.5 cm circles of Whatman DE-81 filter paper. The air-dried filter papers were washed three times with Na_2HPO_4 solution (0.15 M), twice with EtOH, and finally once with Et_2O. The dry filters were counted by liquid scintillation counting in ScintiSafe (30%, 5 ml, Fisher). All experiments were repeated three times and the resulting data averaged. The activity was calculated from the slope of a time versus c.p.m. plot and was expressed as units per ml (U/ml). One unit of enzyme was defined as the amount of polymerase that converts 1 nmol TTP into filter-bound material in 10 min at 37°C.

Enzyme variants had the following specific activities [tested on poly(rA)/oligo(dT)]: HIV-1 RT heterodimer: 8700 U/mg; Variant L74V: 9750 U/mg; Variant K103N: 11 300 U/mg; Variant Y181I: 7500 U/mg; Variant M184V: 10 600 U/mg; Variant Y188L: 5100 U/mg; Variant AZT-21 (M41L, D67N, K70R, T215Y, K219Q): 4150 U/mg; Variant Y188L, E478Q: 5100 U/mg.

Running start experiments

In a typical primer extension experiment, 5′-^{32}P-labeled primer (P1-RS) and template T1, or P2-RS and template T2 (656 nM of the primers, 920 nM of the templates) in HIV-RT buffer were mixed with dATP, dGTP, dCTP and TTP (final concentration 130 μM each) in a total volume of 160 μl. In experiments with non-standard nucleotides, the concentrations for d(pyDAD)TP and d(puADA)TP were also 130 μM. After heating the sample to 95°C for 1 min, the primer/template complex was annealed by cooling to room temperature over 1 h. Primer extension was started by addition of the RT (16 μl). The mixture was then incubated at 37°C. Aliquots (25 μl), taken at various times during the reaction, were quenched by addition of a premixed solution of sodium acetate (2.5 μl, 3 M, pH 5.2), EDTA (1 μl, 0.5 M, pH 8), and ethanol (50 μl). After being stored at –20°C for 20 min, the samples were centrifuged (14 000 r.p.m., 4°C, 20 min) and the pellets dried in the vacuum concentrator. The residues were redissolved in PAGE loading buffer and the samples separated on a 10% PAGE gel (7 M urea). The gel was analyzed using the MolecularImager®.

To improve reproducibility in cases where multiple reactions were run in parallel, a master mixture of primer/template and the dNTPs was prepared by scaling up the listed procedure. Master mixtures were not stored for >24 h at –20°C.

Standing start experiments

Primer P1-SS or P2-SS (15 pmol, 5′-^{32}P-labeled) and the appropriate template (T1 or T2, 21 pmol) were incubated with HIV-RT at a range of pH (8 μl, 3×) and the volume adjusted to 21 μl with water. The DNA was denatured (95°C, 1 min) and cooled to room temperature (1 h). After addition of the appropriate dNTP (1.67 μl, 130 μM final concentration of each) and an aliquot of RT (0.2 U). The mixture was incubated for up to 30 min at 37°C. The reaction was quenched by addition of a premixed solution of sodium acetate (2.5 μl, 3 M, pH 5.2), EDTA (1 μl, 0.5 M, pH 8) and EtOH (50 μl), the DNA was recovered by centrifugation, and the pellet was dried in the vacuum concentrator. The DNA was dissolved in PAGE loading buffer (Bromophenol blue/xylene cyanol mix 0.1 g, water, 1 ml, and formamide, 4 ml) and analyzed using a 10%

PAGE gel (7 M urea). The gel was analyzed with the MolecularImager®.

PCR amplification

To facilitate strand separation, one of the PCR primers (P2-C6) was designed to contain a tetranucleotide appended to the 5′-position via two C6 polyethyleneglycol units. This made the product derived from the primer move slower in a gel electrophoresis experiment than the product derived from the reverse primer (34).

Template T2-pyDAD (50 pmol) was mixed with 5′-radiolabeled primer P2-C6 (750 pmol), primer P2-Rev (750 pmol), dATP, dTTP, dCTP, dGTP, d(puADA)TP, d(pyDAD)TP (final concentration 200 μM each), HIV-RT buffer (333 μl, 3×), and the reaction volume adjusted to 1 ml with water. The mixture was heated to 95°C (10 min) and allowed to cool to ambient temperature (1 h). HIV-RT (Y188L, E478Q) (10 U) was added to the reaction mixture, which was then incubated at 37°C for 24 h. An aliquot (5 μl) was removed and quenched with 20 mM EDTA in formamide (5 μl). The remaining reaction mixture was heated again to 95°C for 10 min and again cooled to ambient temperature over 1 h. Another aliquot of RT was then added. This cycle was repeated four times. The products from each round of PCR amplification were resolved using a 12% PAGE gel (7 M urea). The gel was analyzed using the MolecularImager software. A positive control experiment was run under the same conditions while substituting T2 for T2-pyDAD.

PCR product isolation

The PCR was quenched with EDTA (final concentration 10 mM) and the DNA isolated via ethanol (2.5 ml) precipitation and subsequently washed with 70% ethanol in water. The dry pellet was dissolved in PAGE loading buffer and analyzed by electrophoresis on a 20% PAGE gel (7 M urea). The product generated from full extension of primer P2-C6 was longer, and therefore migrated more slowly, than the product generated from the full extension of P2-Rev. The product from full extension of P2-C6 was cut from the gel and extracted by incubating in a crush and soak buffer [0.1% SDS, 0.5 M NH_4OAc, 10 mM $Mg(OAc)_2$] at 37°C overnight. The solution was filtered through a Millipore filter (0.45 μm pore size) and the DNA recovered by ethanol precipitation. The DNA pellet (T1-puADA-PCR) was dissolved in water to a final concentration of 10 μM.

Paused-extension sequencing

T1-puADA-PCR (2 pmol, presumably generated by the PCR) was mixed with radiolabeled P2-Rev (1 pmol), Thermopol buffer [final concentration 20 mM Tris–HCl, pH 8.8, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100], dATP, dCTP, dTTP, dGTP (final concentration 10 μM each), and the reaction volume was adjusted to 14 μl with water. Heating (95°C, 10 min) and cooling to ambient temperature (1 h) the respective polymerase (reaction 1 with *Taq*, reaction 2 with Vent exo$^+$) was added (1 μl, 2 U/μl) and the mixture was incubated at 72°C for 15 s. The reactions were quenched by addition of PAGE loading buffer containing 20 mM EDTA (15 μl). The samples were subjected to electrophoresis on a 20% PAGE gel and the gel analyzed using the MolecularImager software.

Paused-extension screen

Radiolabeled primer P3 (1 pmol) was mixed with template T3 (2 pmol), Thermopol buffer (1.5 μl, 10×), dATP, dCTP, dTTP, dGTP (final concentration 10 μM each), and water to a final volume of 14 μl. The mixture was heated to 95°C for 10 min and allowed to cool to ambient temperature for 1 h. The respective polymerase (*Taq*, Vent, Deep Vent, Vent exo$^-$ or Deep Vent exo$^-$) was added (1 μl, 2 U/μl) and the reaction incubated at 72°C for 15 s. The reaction was quenched with 20 mM EDTA in PAGE loading buffer (15 μl) and subsequently analyzed by electrophoresis on a 20% PAGE gel. The gel was analyzed using the MolecularImager software.

RESULTS

Primer extension experiments were performed for the four variant forms of HIV-RT (Y181I, Y188L, M184V and AZT-21). Each was challenged to incorporate a single d(puADA) nucleotide opposite a d(pyDAD) at position 26 in the template (T2-pyDAD), five nucleobases upstream of the 3′-end of the primer. Experiments were done in parallel, one containing only standard nucleoside triphosphates (the 'minus' control), the other containing the standard dNTPs plus d(puADA)TP. The progress of the primer extension reaction was followed by PAGE on samples removed at time intervals.

Results of these experiments suggest that HIV-RT variants Y188L (Fig. 2) and Y181I (data not shown) both produced more full-length product in the presence of d(puADA)TP than in its absence, while variants AZT-21 (data not shown) and M184V (Fig. 2) did not. This made variants Y188L and Y181I candidates for further polymerase development.

Primer extension experiments testing the incorporation of d(pyDAD)TP opposite d(puADA) in the template were then performed with these candidates. Variants Y188L and Y181I were incubated (1–480 min) with the standard triphosphates with and without the complementary d(pyDAD)TP. The results (Fig. 3) suggest that Y188L produced a slightly higher ratio of full-length product in the presence of d(pyDAD)TP to full-length product in the absence of d(pyDAD)TP, while variant Y181I produced a lower ratio. This made Y188L the prime candidate for further examination.

Both pu(ADA) and py(DAD) display acid–base chemistry. The heterocycle of pu(ADA) is an acid with a pK_a of ~5.7 (35). In contrast, py(DAD) is protonated, and the conjugate acid was measured to have a $pK_a \approx 6.7$ (36). A series of single nucleotide primer extension experiments were therefore performed with Y188L to identify the nucleotides most likely to compete with the non-standard nucleotides during primer extension, and to assess the impact of pH on incorporation.

To determine the optimal pH for d(puADA)TP incorporation, Y188L was incubated at pH 5.5–7.5. The results (Fig. 4) showed little incorporation at pH <6.5. The optimal pH was between 7.0 and 7.5. While we do not know the exact pK_a of either py(DAD) or pu(ADA) heterocycles when incorporated into an oligonucleotide, they are likely to be higher than 6.7 and 5.7. Therefore, it is possible that at the optimal pH for this reaction, the pu(ADA)-py(DAD) pair is an anion-cation pair.

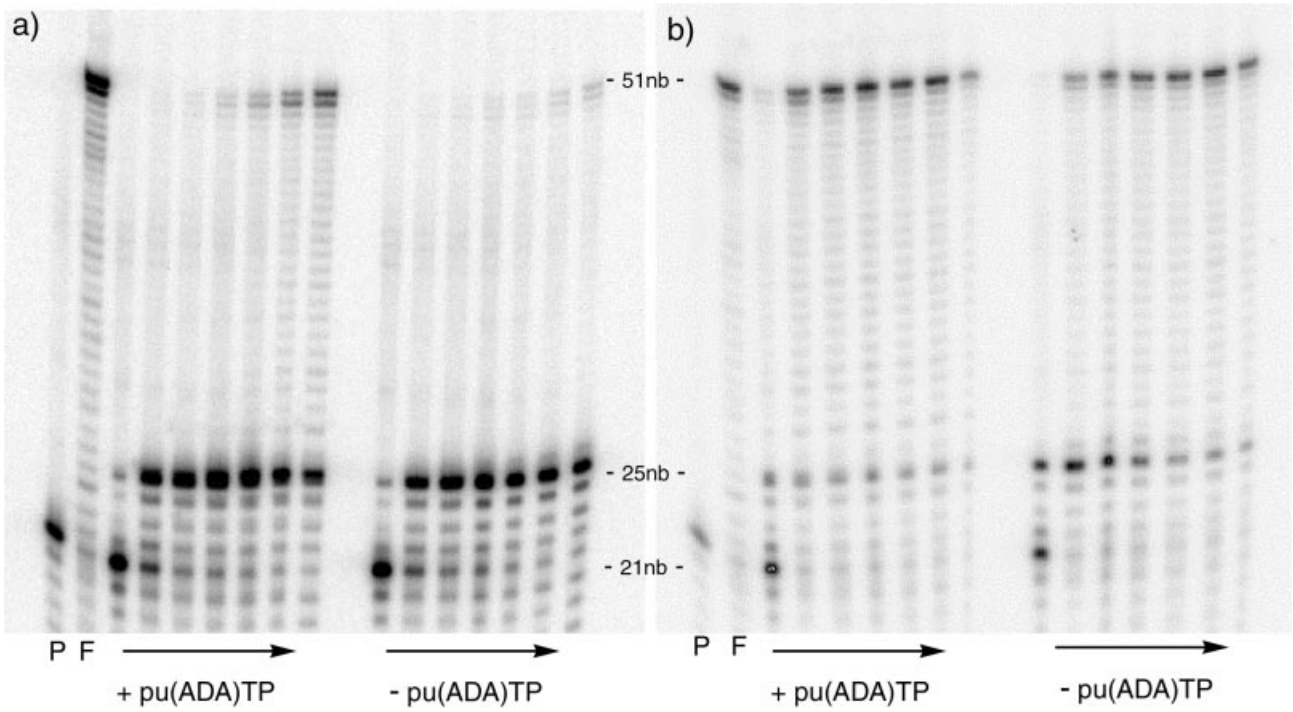

Figure 2. Primer extension experiments with two HIV-RT variants, Y188L (**a**) and M184V (**b**). For each, the variant was incubated for times ranging from 1 to 1440 min [right to left, in direction of arrow, with template T2-pyDAD, radiolabeled primer P2-RS, dATP, dGTP, dCTP and dTTP, and either with the complementary d(puADA)TP (left panel of each gel) or without d(puADA)TP (right panel of each gel)]. Time points are at 1, 30, 60, 120, 240, 480 and 1440 min. P is the unextended primer, 21 nt in length. F is full-length product, following addition of 30 nucleotides to the primer.

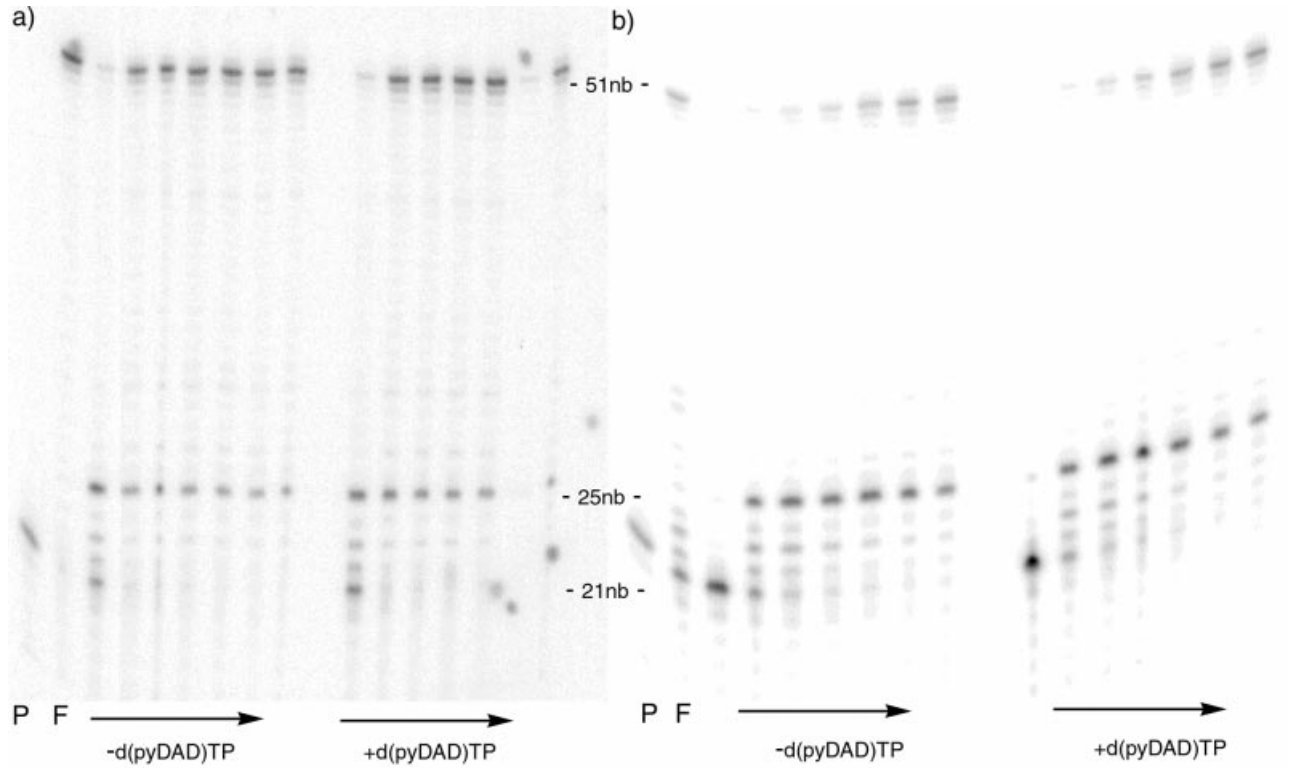

Figure 3. Primer extension experiments with two candidate HIV-RT variants, Y181I (**a**) and Y188L (**b**). For each, the variant was incubated with template T1-puADA, radiolabeled primer P1-RS, dATP, dGTP, dCTP and dTTP, and either with the complementary d(pyDAD)TP (left panel of each gel) or without d(pyDAD)TP (right panel of each gel). Time points are at 1, 15, 30, 60, 120, 240 and 480 min. The data suggest that HIV-RT variant Y188L both produced a slightly higher ratio of full-length product in the presence of d(pyDAD)TP to full-length product in the absence of d(pyDAD)TP, while variant Y181I produced a lower ratio. This made Y188L the prime candidate for further examination.

Figure 4. Single nucleotide primer extension experiments as a function of pH with Y188L. Variant Y188L was incubated for 30 min at the pH indicated, with template T2-pyDAD, radiolabeled primer P2-SS and one of the four standard deoxynucleoside triphosphates or d(puADA)TP. Lane P, primer alone; lane A, with dATP; lane C, with dCTP; lane G, with dGTP; lane T, with TTP; lane ADA, with d(puADA)TP; lane R, positive control with substitution of T2 for T2-pyDAD, thus having dA instead of the non-standard nucleotide. Noticeable in all experiments is the primer band (labeled P) and a degradation band (below the primer band); the degradation of the primer was due to a DNase activity of the RT itself, an activity that was removed by the mutation E478Q (see text). Little incorporation was observed at pH <6.5; the optimal pH was 7.0–7.5. It should be noted that oligonucleotides containing a puADA residue migrate slightly faster due to the negative charge on the heterocycle.

Noticeable in all gels is a band below the primer band. This was not due to a failure in the synthesis of the primer. Rather, the band appeared to arise through degradation caused by a DNase activity present in the reaction mixture. The two possible origins of this activity are contamination, perhaps *E.coli* DNase I, or a residual 3′ exonuclease activity of the RT RNase H domain. The latter activity has not been previously reported. Assuming that the DNase activity was occurring at the ribonuclease site of RT, we replaced Glu 478 by Gln at that site. This generated a double mutant Y188L-E478Q. Preparations of this variant did not degrade the primer.

Although HIV-RT has been tested extensively for DNase activity, the previous investigations did not use a 5-day incubation with single stranded substrate. Therefore, this result may indicate a hitherto undetected trace single stranded 3′-exodeoxyribonuclease activity associated with the enzyme. We cannot, however, rule out the possibility of contamination by *E.coli* DNase I, which could be more easily separated from the double mutant than either the single mutant or native RT.

We then asked whether the variant enzyme (Y188L-E478Q) was able to PCR amplify a DNA duplex containing a py(DAD)-pu(ADA) pair. Because RT is not thermally stable, the amplification was done at 37°C, where additional RT variant was added after each heating-annealing cycle. The amplification was performed over five rounds, with each elongation step lasting 24 h to allow sufficient incorporation and elongation of the NSB. An aliquot (5 µl) was removed from the reaction after each round and examined by PAGE to trace the progress of the reaction. The results (Fig. 5) display the disappearance of primer and the generation of full-length product with amplification.

We then asked whether the PCR-amplified product retained the AEGIS components. To this end, a novel sequencing technique was developed to determine the amount of misincorporation at the NSB site. We found that *Taq* polymerase terminates the elongation of a primer when the polymerase encounters a d(puADA) in the template (Fig. 6). Thus, the PCR generated oligonucleotide containing d(puADA), after isolation from all other PCR products and reactants, was tested for elongation termination using *Taq* polymerase, the natural dNTPs, and the appropriate radio-labeled primer (P1-RS). The results indicate that >95% of the

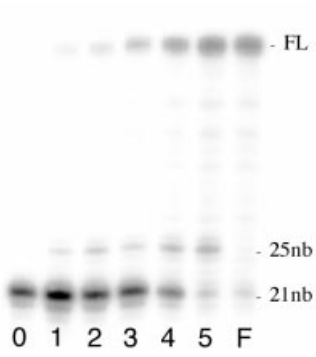

Figure 5. PCR amplification (from left to right, 0 to five rounds, each 24 h) of template T2-py(DAD) using primers P2-C6 and P2-Rev in the presence of dNTPs, d(puADA)TP and d(pyDAD)TP (200 µM each), showing the appearance of full-length product, seen in the positive control labeled F, using the double variant HIV-RT Y188L E478Q. Note the absence of degradation of primer due to the knockout of the nuclease activity via the E478Q mutation. RT was added at each cycle.

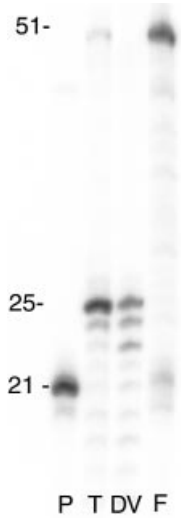

Figure 6. Use of *Taq* and DeepVent (DV) exo$^+$ polymerases as sequencing tools. Both polymerases were incubated with template T3, primer P3 and dNTPs. P, primer alone, 19 nt in length. Both polymerases abort elongation upon encountering dX in the template, but efficiently generate (as expected) full-length products when challenged with templates containing only A, T, G and C. Deep Vent exo appears to be degrading the primer, and perhaps the product as well.

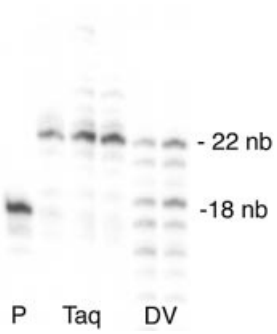

Figure 7. Proof that the PCR product (from round five) contained puADA nucleotide. Primer extension experiment with primer P2-Rev and the product of round five of the PCR experiment, with *Taq* and DeepVent polymerases. Pausing at position 25 demonstrates the presence of neither A, T, G nor C at this position. Generation of <5% full-length product establishes <5% misincorporation of the standard nucleotides after five rounds of PCR amplification.

primer extension stopped at position 25 (Fig. 7), establishing that the non-standard base survived the five rounds of PCR without being replaced by >5%.

DISCUSSION

This work shows that the Y188L-E478Q variant of HIV-RT can be used to PCR-amplify an oligonucleotide containing a single pu(ADA) or a single py(DAD). This represents the first example of an enzyme capable of replicating an artificial genetic system in this way.

The use of nature to generate variant enzymes capable of altered functionalities facilitates the finding of new biocatalysts. While the directed (*in vitro*) evolution of proteins is often useful in developing variants with new or altered catalytic properties (37), it suffers from a major disadvantage in that many (if not most) of the variant enzymes do not retain any functional behavior at all. Further, it is difficult to screen for subtle properties of a variant (such as high fidelity and high processivity, as opposed to simple fidelity and simple processivity) that may be critical to the value of the polymerase. Thus, combining directed evolution techniques with those described here, where natural evolution is exploited, may in special cases speed the development of useful biocatalysts.

From a scientific perspective, it is interesting to note that only two amino acids must be substituted in a natural polymerase optimized for the four standard nucleotides to create one that supports repeated PCR cycles for the amplification of an expanded genetic system. We did not expect a useful polymerase to be so close in 'sequence space' to that of the wild type polymerase.

From a technological perspective, since mutation of the polymerases of pathogens is a common process by which pathogens develop resistance to drugs targeting the polymerase (38–40), the study of polymerase variants is becoming an important tool for understanding the development of drug resistance, one of the most significant emerging challenges in human therapy.

Loeb and coworkers have shown that HIV-RT complements the polymerase deficiency in a strain of *E.coli* originally developed by Witkin and coworkers (41,42). We have now shown that Y188L complements this defect as well (J.-H. Park, unpublished). Thus, the work reported here takes the next step towards implementing an artificial genetic system in *E.coli*.

ACKNOWLEDGEMENTS

We are indebted to the National Institutes of Health (GM 54048) for partial support of this work. Work in S.H.H.'s laboratory was supported by the National Cancer Institute (NCI training grant CA-09229) and NIGMS. We are also indebted to Dr J. Michael Thomson for helpful discussion.

REFERENCES

1. Switzer,C.Y., 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.
2. Piccirilli,J.A., Krauch,T., Moroney,S.E. and Benner,S.A. (1990) Extending the genetic alphabet: enzymatic incorporation of a new base pair into DNA and RNA. *Nature*, **343**, 33–37.
3. Strazewski,P. and Tamm,C. (1990) Replication experiments with nucleotide base analogs. *Int. Ed. Engl.*, **29**, 36–57.
4. Rappaport,H.P. (1993) Replication of the base pair 6-thioguanine/5-methyl-2-pyrimidinone with the large Klenow fragment of *Escherichia coli* DNA-polymerase-I. *Biochemistry*, **32**, 3047–3057.
5. Kool,E.T. (2001) Hydrogen bonding, base stacking and steric effects in DNA replication. *Annu. Rev. Biophys. Biomol. Struct.*, **30**, 1–22.
6. Wu,Y.Q., Ogawa,A.K., Berger,M., McMinn,D.L., Schultz,P.G. and Romesberg,F.E. (2000) Efforts toward expansion of the genetic alphabet: optimization of interbase hydrophobic interactions. *J. Am. Chem. Soc.*, **122**, 7621–7632.
7. Ohtsuki,T., Kimoto,M., Ishikawa,M., Mitsui,T., Hirao,I. and Yokoyama,S. (2001) Unnatural base pairs for specific transcription. *Proc. Natl Acad. Sci. USA*, **98**, 4922–4925.
8. Goodman,M.F. (1999) On the wagon. DNA polymerase joins 'H-bonds anonymous'. *Nat. Biotechnol.*, **17**, 640–641.
9. Service,R. (2000) Creation's Seventh Day. *Science*, **289**, 232–235.
10. Watson,J.D. and Crick,F.H.C. (1953) Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature*, **171**, 964–967.
11. Collings,M.L, Irvine,B., Tyner,D., Fine,E., Zayati,C., Chang,C., Horn,T., Ahle,D., Detmer,J., Shen,L., Kolberg,J., Bushnell,S., Urdea,M.S. and Ho,D.D. (1997) A branched DNA signal amplification assay for the quantification of nucleic acid targets below 100 molecules/ml. *Nucleic Acids Res.*, **25**, 2979–2984.
12. Yu,C.Z., Henry,A.A., Romesberg,F.E. and Schultz,P.G. (2002) Polymerase recognition of unnatural base pairs. *Int. Ed. Engl.*, **41**, 3841–3844.
13. Bain,J.D., Chamberlin,A.R., Switzer,C.Y. and Benner,S.A. (1992) Ribosome-mediated incorporation of non-standard amino acids into a peptide through expansion of the genetic code. *Nature*, **356**, 537–539.
14. Horlacher,J., Hottiger,M., Podust,V.N., Hübscher,U. and Benner,S.A. (1995) Expanding the genetic alphabet: recognition by viral and cellular DNA polymerases of nucleosides bearing bases with non-standard hydrogen bonding patterns. *Proc. Natl Acad. Sci. USA*, **92**, 6329–6333.
15. Kiefer,J.R., Mao,C., Braman,J.C. and Beese,L.S. (1998) Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature*, **391**, 304–307.
16. Hingorani,M.M. and O'Donnell,M. (2000) DNA polymerase structure and mechanisms of action. *Current Org. Chem.*, **4**, 887–913.
17. Das,K., Ding,J., Hsiou,Y., Clark,A.D.,Jr, Moereels,H., Koymans,L., Andries,K., Pauwels,R., Janssen,P.A.J., Boyer,P.L., Clark,P., Smith,R.H.,Jr, Kroeger-Smith,M.B., Michejda,C.J., Hughes,S.H. and Arnold,E. (1996) Crystal structure of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181 Cys HIV-1 RT drug-resistant mutant. *J. Mol. Biol.*, **264**, 1085–1100.

18. Kohlstaedt,L.A., Wang,J., Friedman,J.M., Rice,P.A. and Steitz,T.A. (1992). Crystal structure at 3.5Å Resolution of HIV-1 Reverse Transcriptase Complexed with an Inhibitor. *Science*, **256**, 1783–1790.
19. Lutz,M.J., Held,H.A., Hottinger,M., Hübscher,U. and Benner,S.A. (1996) Differential discrimination of DNA polymerases for variants of the non-standard nucleobase pair between xanthosine and 2,4-diaminopyrimidine, two components of an expanded genetic alphabet. *Nucleic Acids Res.*, **24**, 1308–1313.
20. Larder,B.A. and Kemp,S.D. (1989). Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science*, **246**, 1155–1158.
21. Tantillo,C., Ding,J.P., Jacobomolina,A., Nanni,R.G., Boyer,P.L., Hughes,S.H., Pauwels,R., Andries,K., Janssen,P.A.J. and Arnold,E. (1994) Locations of anti-AIDS drug-binding sites and resistance mutations in the 3-dimensional structure of HIV-1 reverse-transcriptase. Implications for mechanisms of drug-inhibition and resistance. *J. Mol. Biol.*, **243**, 369–387.
22. Gu,Z.X., Qing,G., Li,X.G., Parniak,M.A. and Wainberg,M.A. (1992) Novel mutation in the human-immunodeficiency-virus type-1 reverse-transcriptase gene that encodes cross-resistance to 2′,3′-dideoxyinosine and 2′,3′-dideoxycytidine. *J. Virol.*, **66**, 7128–7135.
23. Chu,C.K., Wempen,I., Watanabe,K.A. and Fox,J.J. (1976) Nucleosides. 100. General synthesis of pyrimidine C-5 nucleosides related to pseudouridine. Synthesis of 5-(β-D-ribofuranosyl)isocytosine (pseudoisocytosine), 5-(β-D-ribofuranosyl)-2-thiouracil (2-thiopseudouridine) and 5-(β-D-ribofuranosyl)uracil (pseudouridine). *J. Org. Chem.*, **41**, 2793–2797.
24. Lutz,M.J., Horlacher,J. and Benner,S.A. (1998) Recognition of a non-standard base pair by thermostable DNA polymerases. *Bioorg. Med. Chem. Lett.*, **8**, 1149–1152.
25. Jurczyk,S.C., Horlacher,J., Devine,K.G., Benner,S.A. and Battersby,T.R. (2000) Synthesis and characterization of oligonucleotides containing 2′-deoxyxanthosine using phosphoramidite chemistry. *Helv. Chim. Acta*, **83**, 1517–1524.
26. VanAerschot,A., Mag,M., Herdewijn,P. and Vanderhaeghe,H. (1989) Double protection of the heterocyclic base of xanthosine and 2′-deoxyxanthosine. *Nucl. Nucl.*, **8**, 159–178.
27. Sinha,N.D., Biernat,J. and Köster,H. (1983) Beta-cyanoethyl n,n-dialkylamino/n-morpholinomonochloro phosphoamidites, new phosphitylating agents facilitating ease of deprotection and work-up of synthesized oligonucleotides. *Tetrahedron Lett.*, **24**, 5843–5846.
28. Ludwig,J. and Eckstein,F. (1989) Rapid and efficient synthesis of nucleoside 5′-o-(1-thiotriphosphates), 5′-triphosphates and 2′,3′-cyclophosphorothioates using 2-chloro-4H-1,,2-benzodioxaphosphorin-4-one. *J. Org. Chem.*, **54**, 631–635.
29. Boyer,P.L., Tantillo,C., Jacobo-Molina,A., Nanni,R.G., Ding,J.P., Arnold,E. and Hughes,S.L. (1994) Sensitivity of wild-type human-immunodeficiency-virus type-1 reverse-transcriptase to dideoxynucleotides depends on template length - the sensitivity of drug-resistant mutants does not. *Proc. Natl Acad. Sci. USA*, **91**, 4882–4886.
30. Bryant,F.R., Johnson,K.A. and Benkovic,S.J. (1983). Elementary steps in the DNA polymerase I reaction pathway. *Biochemistry*, **22**, 3537–3546.
31. Kuchta,R.D. (1996). Isotopic assays of viral polymerases and related proteins. *Methods Enzymol.*, **275**, 241–257.
32. Reardon,J.E. and Miller,W.H. (1990). Human immunodeficiency virus reverse transcriptase. *J. Biol. Chem.*, **265**, 20302–20307.
33. Stahlhut,M.W. and Olsen,D.B. (1996). Expression and purification of retroviral HIV-1 Reverse Transcriptase. *Methods Enzymol.*, **275**, 122–132.
34. Williams,K.P. and Bartel,D.P. (1995). PCR products with strands of unequal length. *Nucleic Acids Res.*, **23**, 4220–4221.
35. Roy,K.B. and Miles,H.T. (1983) Tautomerism and ionization of xanthosine. *Nucl. Nucl.*, **2**, 231–242.
36. Krauch,T. (1989) Ein neues Basenpaar Fur den genetischen Code. PhD Dissertation, ETH No. 8940, Swiss Federal Institute of Technology, Zurich, Switzerland.
37. Arnold,F.H. (1998). Design by directed evolution. *Acc. Chem. Res.*, **31**, 125–131.
38. Kinchington,P.R., Araullo-Cruz,T., Vergnes,J.P., Yates,K. and Gordon,Y.J. (2002) Sequence changes in the human adenovirus type 5 DNA polymerase associated with resistance to the broad spectrum antiviral cidofovir. *Antivir. Res.*, **56**, 73–84.
39. Seigneres,B., Pichoud,C., Ahmed,S.S., Hantz,O., Trepo,C. and Zoulim,F. (2000) Evolution of hepatitis B virus polymerase gene sequence during famciclovir therapy for chronic hepatitis B. *J. Infect. Dis.*, **184**, 1221–1233.
40. Chou,S.W., Miner,R.C. and Drew,W.L. (2000) A deletion mutation in region V of the cytomegalovirus DNA polymerase sequence confers multidrug resistance. *J. Infect. Dis.*, **182**, 1765–1768.
41. Witkin,E.M. and Roegner-Maniscalco,V. (1992). Overproduction of DnaE Protein (a-subunit of DNA polymerase III) restores viability in a conditional inviable *Escherichia coli* strain deficient in DNA polymerase I. *J. Bacteriol.*, **174**, 4166–4168.
42. Kim,B. and Loeb,L. (1995) Human immunodeficiency virus reverse transcriptase substitutes for DNA polymerase I in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **92**, 684–688.