Highly Tolerated Amino Acid Substitutions Increase the Fidelity of *Escherichia coli* DNA Polymerase I*

Ern Loh[1,2], **Juno Choe**[2], and **Lawrence A. Loeb**[3]

From the Joseph Gottstein Memorial Cancer Research Laboratory, Department of Pathology, University of Washington, Seattle, Washington 98195

Fidelity of DNA synthesis, catalyzed by DNA polymerases, is critical for the maintenance of the integrity of the genome. Mutant polymerases with elevated accuracy (antimutators) have been observed, but these mainly involve increased exonuclease proofreading or large decreases in polymerase activity. We have determined the tolerance of DNA polymerase for amino acid substitutions in the active site and in different segments of *E. coli* DNA polymerase I and have determined the effects of these substitutions on the fidelity of DNA synthesis. We established a DNA polymerase I mutant library, with random substitutions throughout the polymerase domain. This random library was first selected for activity. The essentiality of DNA polymerases and their sequence and structural conservation suggests that few amino acid substitutions would be tolerated. However, we report that two-thirds of single base substitutions were tolerated without loss of activity, and plasticity often occurs at evolutionarily conserved regions. We screened 408 members of the active library for alterations in fidelity of DNA synthesis in *Escherichia coli* expressing the mutant polymerases and carrying a second plasmid containing a β-lactamase reporter. Mutation frequencies varied from ¹⁄₁₀₀₀- to 1000-fold greater compared with wild type. Mutations that produced an antimutator phenotype were distributed throughout the polymerase domain, with 12% clustered in the M-helix. We confirmed that a single mutation in this segment results in increased base discrimination. Thus, this work identifies the M-helix as a determinant of fidelity and suggests that polymerases can tolerate many substitutions that alter fidelity without incurring major changes in activity.

DNA polymerases function in DNA replication, repair, and recombination and are essential for maintaining the integrity of the genome. Multiple DNA polymerases have been found in prokaryotes, eukaryotes, and viruses and may have different properties, including variations in the accuracy, or fidelity, of DNA synthesis (1–5). Polymerases that replicate large genomes typically have high fidelity, with error rates on the order of 10^{-5} to 10^{-6} (6–8) to prevent the accumulation of deleterious mutations in the genome. Deficits in DNA polymerase fidelity have been associated with cancer, accelerated aging, and infertility (9–11).

Mutator and antimutator DNA polymerases harbor amino acid substitutions resulting in decreased and increased fidelity, respectively. These variant enzymes provide insights into the structural basis of accurate DNA synthesis (12–18). Many mutators have been identified due to the availability of powerful selection methods (19). The majority of studied mutators harbor amino acid substitutions in well conserved polymerase motifs. Mutator polymerases have also been used extensively in biotechnological applications, such as DNA sequencing and error-prone PCR (20). Far fewer antimutator enzymes have been identified due to the lack of selection methods. The ones identified thus far either exhibit substantially diminished catalytic activity or possess increased 3′–5′ exonuclease proofreading (21–23). Decreased polymerase activity allows more time for removal of misincorporations by a proofreading exonuclease. Recent efforts to identify antimutators in proofreading-deficient constructs through 96-well screening have focused only on 3 residues within Motif C and have produced antimutators with less than 10% of wild-type activity (24).

DNA polymerase I (Pol I)[4] is a high fidelity polymerase belonging to Family A, which includes Taq Pol I, T7 DNA polymerase, and human DNA polymerase γ, θ, and ν (6, 25, 26). In this work, we identified determinants of Pol I fidelity that do not involve alterations in proofreading by the 3′–5′ exonuclease and that do not cripple the polymerase activity. This was attained by conducting a comprehensive analysis of amino acid substitutions in Pol I that result in antimutator or mutator activity. We focused on mutations that directly affect nucleotide selection at the incorporation step. We report on the spectrum of amino acid substitutions that both maintain activity and confer either antimutator or mutator phenotypes. Our results indicate that although the catalytic site is optimized for base selection without loss of activity, it is still feasible to create more accurate DNA polymerases without greatly reducing catalytic activity. Furthermore, many of the mutations that affect fidelity are distant from the active site.

[4] The abbreviation used is: Pol I, polymerase I.

EXPERIMENTAL PROCEDURES

Escherichia coli strain JS200 (*SC-18 recA718 polA12 uvrA155 trpE6 lon-11 sulA1*) was first described as SC18-12. Creation of the plasmids pECpolI-3′exo⁻ and pLA230 have been previously described (27, 28). Custom oligonucleotides and chemicals were purchased from IDT (Coralville, IA) and Sigma unless otherwise specified.

Creation of a Pol I Mutant Library—The *E. coli polA* gene encoded in plasmid pECpolI-3′exo⁻ was mutated to introduce a PstI restriction site at nucleotide 1493. The polymerase domain was amplified from this plasmid using the primers GGACGT-TACGCTGCAGAAG and CGACGGCCAGTGAATTC-TTAG according to the error-prone PCR GeneMorph II kit (Stratagene, La Jolla, CA), which includes error-prone Taq Pol I mutant 53 (29). The PCR product was gel-purified (Qiagen, Valencia, CA), cut using PstI and EcoRI (New England Biolabs, Ipswich, MA), and ligated into a pECpolI-3′exo⁻ dummy vector. The dummy vector had Pol I nucleotides 2219–2646 deleted, resulting in an inactive polymerase. The ligated plasmids were transformed into TOP10 cells (Invitrogen). Library size was calculated by counting a fraction of the surviving transformants.

Selection of Active Pol I Mutants—The library was transformed into JS200 cells, which were then plated on 2XYT agar and grown overnight at 37 °C. Surviving colonies were inoculated into LB and grown at 30 °C. Individual plasmids from both the unselected and selected library were recovered and sequenced. Analysis of the library mutation spectrum was performed on Phred-Phrap software (30, 31). The index of substitutability at each position was calculated as the percentage of clones containing a mutation at that residue divided by the total number of clones sequenced.

Comparison of Evolutionary Conservation Index with Substitutability Index—An index of evolutionary conservation (Fig. 2B) was calculated based on an alignment of the *E. coli* polymerase domain with orthologues from 30 members of Family A, including representatives from bacteria (*E. coli, Yersina pestis, Vibrio cholera, H. influenzae, Pseudomonas aeruginosa, Bordetella pertussis, Neisseria meningitidis*, Nostoc7120, Taq, *Deinococcus radians*, Bst, *Staphylococcus aureus, Streptococcus pyogenes, Clostridium tetani, Mycobacterium tuberculosis, Chlamydia trachomatis*, and *Helicobacter pylori*), phage (T5, T7, and Spo1), and eukaryotes (*Saccharomyces cerevisiae, Schizosaccharomyces pombe, Homo sapiens* DNA polymerase γ and θ, *Mus musculus, Drosophila melanogaster, Leishmania major, Plasmodium falciparum, Arabidopsis thaliana*, and *Trypanosoma brucei*) (32, 33). The DNA polymerases analyzed were DNA polymerase I (bacteria) or DNA polymerase γ (eukaryotes) unless otherwise noted. DNA polymerase γ sequences were included to further differentiate residues that are absolutely required for catalytic activity. Sequences were obtained from Pubmed (available on the World Wide Web at www.ncbi.nlm.nih.gov/) and aligned with ClustalW version 1.83 (available on the World Wide Web at www.ebi.ac.uk/clustalw/).

Screen for Mutations in Pol I That Alter Fidelity of DNA Synthesis—Plasmid pECpolI was recovered from active clones and transformed into fresh JS200. After recovery, cells were transformed with the second plasmid pLA230 and grown in LB to an A_{600} of 2.0 at 30 °C. The culture was then inoculated to a final dilution of 5×10^{-6} into 1 ml of 2XYT containing chloramphenicol (30 μg/ml) and kanamycin (50 μg/ml). Cultures were then grown with aeration for 20 h at 37 °C and plated on LB agar containing either chloramphenicol and carbenicillin (100 μg/ml) or chloramphenicol and kanamycin at a dilution sufficient for 100–500 colonies per plate. Several controls were included: wild type sequences, D424A,I709N, and D424A, I709N,A759R (34). Assays were performed in duplicate.

Assay for Pol I Polymerase Activity—JS200 cells containing pECpolI mutants were grown in LB at 30 °C with 1 mM isopropyl 1-thio-β-D-galactopyranoside to induce protein expression. Cells were grown to an A_{600} of 2.0, at which time a 1.5-ml aliquot was pelleted, washed with 1 ml of 20 mM sodium phosphate (pH 7.2), and resuspended in 0.1 ml of the same buffer. Cells were lysed with 5 μl of lysozyme (10 mg/ml) and disrupted by freezing (−80 °C) and thawing. The cell extract was collected by centrifugation at 15,000 rpm for 15 min. The polymerase activity of the supernatant was assayed for ability to incorporate radioactive dTMP into activated calf thymus DNA at 37 °C as previously described (35). Assays were performed in triplicate. The background polymerase activity of the cell, measured from cells grown with an empty plasmid vector, was subtracted from all measurements.

Protein Purification—Mutant and wild-type Klenow fragments, encompassing the 3′–5′ exonuclease domain and the polymerase domain, were subcloned into the pLEX vector (Invitrogen), expressed, and purified as previously described (27). This Pol I construct included a six-histidine N terminus tag. In brief, the pLEX expression plasmid was introduced into *E. coli* GI724 (F2, l2, *lac*Iq, *lac*PL8, *amp*C::*P*trp cI, *mcr*A, *mcr*B, INV(*rnn*D-*rnn*E)), and cells were grown at 30 °C in 500 ml of induction medium composed of M9 salts, 0.2% casamino acids, 0.5% glucose, 1 mM MgCl₂, and 100 mg/ml carbenicillin. When the cultures reached an A_{600} of 0.5, tryptophan was added to a final concentration of 100 μg/ml, and the culture was incubated at 37 °C for 4 h. The cells were collected by centrifugation, washed with 40 ml of binding buffer (5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9), and suspended in 4 ml of the same buffer containing 200 μg/ml lysozyme and 0.04 ml Protease Inhibitor Mixture III (Calbiochem). Extracts were prepared by freezing/thawing on ice for 2 h, followed by centrifugation at 15,000 rpm for 15 min, and loaded onto a 4-ml nickel-resin column (Novagen, San Diego, CA). The column was washed with 80 ml of wash buffer (15 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9) and serially with 8 ml of high imidazole wash buffer (30, 60, 120, and 200 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9) and eluted with 8 ml of 400 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9. The eluted sample was mixed with 2.5 ml of 80% glycerol and 12.5 μl of 1 M dithiothreitol and stored at 80 °C. Protein concentrations were determined by the method of Bradford.

To assay the polymerase and 3′–5′ exonuclease activity, a DNA duplex with a 5′ overhang (template strand) was prepared using a template oligonucleotide, CCCGGGAAATTTCCGG-AATTCGATATTGCTAGCGGGAATTCGGCGCG, and the primer oligonucleotide CGCGCCGAATTCCCGCTAGCAA-TAG, the latter of which was ³²P-radiolabeled at the 5′ termi-

nus. The purified enzyme (40 fmol) was incubated with the DNA duplex (20 fmol) in 10 μl of 50 mM Tris-HCl, pH 7.5, 10 mM KCl, 2.5 mM MgCl$_2$, 1 mM dithiothreitol at 37 °C for 30 min. Klenow 3′–5′ exonuclease-proficient enzyme (4 fmol) was used as a control. Reactions were terminated with 20 μl of 98% formamide, 10 mM EDTA; 5 μl of each product was analyzed by electrophoresis through 14% denaturing acrylamide electrophoresis and quantified by phosphorimaging analysis.

M13 Gapped Assay—We characterized the DNA synthesis fidelity of purified mutant and wild-type polymerase *in vitro* on the M13 gapped forward mutation assay, as previously described (36). Gap fill-in reactions occurred in 10 μl of 50 mM Tris-HCl, pH 7.5, 10 mM KCl, 2.5 mM MgCl$_2$, 1 mM dithiothreitol at 37 °C with all four dNTPs at 50 μM each. The complete fill-in of the M13 gapped substrate was monitored by gel (data not shown) and by the consistency of mutation rates after incubation with increased amounts of the DNA polymerase.

RESULTS

To explore the protein sequence space of Pol I, we generated a random mutation library of the *polA* gene using error-prone PCR. Because we wanted to focus on mutations in the polymerase domain, nucleotides at positions 1493–2784 that encode the polymerase catalytic domain were randomly substituted, and the 3′-5′ exonuclease proofreading activity was inactivated by a D424A substitution (27, 37). For simplicity, mutants henceforth will only be named by their polymerase domain genotype. For example, the D424A mutant is referred to as "wild type," whereas the double mutant D424A,I709N is referred to as only I709N. A library of greater than 1.0×10^4 mutants was established, with a mean amino acid substitution of 4.2 per gene (polymerase domain only). To select for active mutants, the library was transfected into JS200 *E. coli* cells, which contain a temperature-sensitive allele of Pol I; growth at the restrictive temperature requires complementation with a plasmid-borne active Pol I (27, 38). Approximately 10% of the original library formed colonies at the restrictive temperature. From the surviving population, 592 individual clones were sequenced to reveal a spectrum of tolerable substitutions across the polymerase domain (Fig. 1). A mean of 2.8 amino acid mutations were present in the active clones. Across the entire polymerase domain, the probability (30) that a single random amino acid substitution will inactivate the enzyme is $34 \pm 2\%$ (mean ± S.E.).

To compare the tolerance to substitutions at different amino acid positions, an index of substitutability was calculated based on the percentage of active clones that harbored a mutation at that residue (30). These indices were consistent with the prediction that the active site of the polymerase, located in the interior-facing surfaces of the enzyme, harbors the most highly conserved residues, whereas residues located on the outer surfaces and lacking direct contact with the reaction substrates more readily tolerate substitutions (Fig. 2*A* and Table 1). To quantify how well these results match the amino acid conservation among Family A DNA polymerases, we compiled a protein sequence alignment of 30 Family A DNA polymerase species (17 prokaryotic, 10 eukaryotic, and 3 viral) and calculated indices for evolutionary conservation at each position (Fig. 2*B*).

By comparing the substitutability of different amino acid residues in Pol I with the evolutionary conservation among Family A polymerases, we were able to identify regions of discrepancy between the functional and evolutionary conservation (Fig. 2*B*). Residues where the functional tolerance is at least 35% greater than the evolutionary conservation include 516, 656, 720, 723, 726, 738, 757, and 802.

To assess the accuracy of the mutant polymerase enzymes, we screened 408 clones from our active library for their polymerase fidelity using a second plasmid pLA230 as a reporter (28). This reporter contains an opal (TAA) codon at the N terminus of the β-lactamase protein, which terminates its translation. Errors that are produced in this codon during the Pol I-catalyzed replication of pLA230 render the bacteria resistant to the antibiotic carbenicillin. The reversion frequency has been shown to be inversely related to the *in vitro* fidelity of the Pol I enzyme (28). The reversion frequency for cells expressing the wild-type Pol I was $8.9 \pm 1.9 \times 10^{-7}$ (mean ± S.E.). Clones with a reversion frequency at least 10-fold below or above wild-type were classified as antimutators and mutators, respectively (Fig. 3). The lowest and highest reversion frequency were ~1/1000- and 1000-fold that of wild type. Approximately 12% (51) of the clones were antimutators, whereas only 2.4% (10) were mutators.

To determine the catalytic activity of the same 408 clones, we measured their capacity to incorporate radiolabeled dTMP into activated calf thymus DNA *in vitro* (35). The activity of the polymerases varied greatly (Fig. 3), ranging from 15 to 125% of that of clones expressing similar amounts of wild-type enzyme as determined by Western blots. The only clones with activity greater than wild-type were mutators. Both mutators and antimutators were obtained without significant diminution in catalytic activity.

To verify that the results of our reversion and polymerase activity assays were internally valid, we included three wild-type controls that had nucleotide substitutions that did not change the encoded amino acid sequence. In addition, two previously characterized extreme mutators, I709N and the double mutant I709N,A759R, were included (28, 34). Their reversion frequencies are ~200- and 1700-fold, respectively, above wild-type, consistent with prior studies.

In order to confirm that the enhanced fidelity of DNA synthesis measured *in vivo* reflects a greater accuracy by the DNA polymerase, we purified the polymerase from a mutant, K601I,A726V, that exhibited a 10-fold enhanced accuracy without a significant reduction in activity. First, both the mutant and a wild-type control were cloned into the pLEX vector and expressed with a six-histidine N terminus tag. After chromatography on a nickel column, both the purified mutant polymerase and the wild-type control polymerase were tested for 3′–5′ exonuclease activity to ensure that the mutant did not exhibit any proofreading (Fig. 4). The polymerases were incubated with double-stranded DNA with a mismatched G:A (primer-template) at the 3′-primer terminus. Less than 0.001% of the terminal nucleotides were hydrolyzed in 30 min at 37 °C in a reaction containing a 4 nM concentration of the mutant or wild-type DNA polymerase and 2 nM DNA. 3′–5′ exonuclease activity was also absent

FIGURE 2. **Functional and evolutionary conservation of the Pol I polymerase domain.** *A*, after active mutants of Pol I were identified, the number of tolerable mutations at each residue was determined and used to calculate an index of substitutability at that position. Values for these indices are represented as a *color* along a gradient, from *blue* to *red*, representing the gradient from fewest substitutions to the most substitutions. The crystal structure of Pol I Klenow fragment (Protein Data Bank file 1KLN) was then *colored* accordingly at each residue. The two structures depict opposite faces of Pol I, with the duplex DNA (*magenta*) positioned between the 3′-5′ exonuclease domain (*gray*) and the polymerase domain (*colored*). The active site of the polymerase domain is denoted with an *asterisk*. *B*, the functional conservation of Pol I, obtained from genetic complementation, is compared with its evolutionary conservation, obtained from sequence alignment. Amino acid sequences of 30 members of Family A DNA polymerases were aligned to determine conservation at each position. The same *color gradient* used in *A* was used to code the conservation of each residue along the primary sequence of Pol I, represented as *colored boxes*, with *numbers along* the *bottom* denoting the amino acid position.

when the polymerases were incubated with double-stranded DNA with a matched A:T at the 3′-primer terminus (data not shown). The fidelity of the mutant protein and of the wild type were then measured *in vitro* using an M13 gapped forward mutation assay (36), in which the enzymes catalyze the incorporation of four dNTPs into gapped M13 circular DNA. The complete fill-in of the M13 gapped substrate was monitored by gel (data not shown) and by the consistency of mutation rates after incubation with increased amounts of

the DNA polymerase. A 5-fold increase in the amount of enzyme used for fill-in did not alter the mutation frequency significantly.

In two separate experiments, the frequency of mutations in the gapped segment of M13 DNA was 3.1- and 2.6-fold less in reactions catalyzed by K601I,A726V than in reactions catalyzed by the wild-type polymerase (Table 2). 68 and 82% of the mutations were single base substitutions with the wild-type and mutant polymerase, respectively. The largest reduction in error

FIGURE 1. **Tolerated amino acid changes in the Pol I polymerase domain.** The primary sequence of Pol I, numbered as residues 514–928, is depicted in *five rows*, with the wild-type amino acid identity placed *below each number* and the tolerated mutations placed *above each number*. This spectrum of mutations was compiled from 592 sequenced active mutants. Amino acids are *colored* by type. *Boxes* below the wild-type sequence identify absolutely conserved residues and highly conserved motifs of Family A DNA polymerases (53, 54). Identification of α-helices and β-strands, also denoted *below* the wild-type sequence, was obtained from Swiss-PdbViewer version 3.7 (59) using Protein Data Bank file 1KLN (60).

rate, 10-fold, occurred with A to C transversions (Table 3). The mutant exhibited reduced mutation rates for 10 of the possible 12 misinserted bases. Mutation rates for the remaining two, A to G transitions and C to A transversions, were increased less than 2-fold. Many of the error rates for the wild-type Klenow control were comparable with those that have been previously published (13).

The mapping of mutations present in each antimutator and mutator revealed regions where mutations were more frequent. One region, encompassing residues 720–728 and corresponding to the M-helix in the Pol I protein, clearly stood out. Approximately 12% of the mutations among the antimutators occurred in this segment. This high mutation frequency in the M-helix was not due to a bias in the input population, since only 2.6% of the random substitutions in mutants in our screen were located in this region.

All of the antimutators having a mutation in the M-helix also harbored additional mutations elsewhere. To determine whether a single substitution in the M-helix was capable of producing an antimutator, we genetically dissected the dou-

ble mutant, K601I,A726V, into individual mutations and assayed the reversion ability of each (Fig. 5). The double mutant K601I,A726V exhibited a 20-fold decrease in reversion frequency. The K601I mutation and the A726V mutation separately reduced the reversion frequency by 7- and 8-fold, respectively, relative to wild type. This result suggested that each of the substitutions increased base selection and that there was an additive interaction between the two mutations.

DISCUSSION

The centrality of DNA polymerases to the DNA synthetic process, the high degree of evolutionary conservation, and the stringent requirements for selection of four different nucleotide substrates would suggest that these enzymes should have limited tolerance to substitutions within their catalytic domains. Previous studies on random substitutions in Motif A of *E. coli* Pol I and Taq Pol I indicated that this region tolerates many substitutions (27, 39). The results presented in this paper identify many other regions of the polymerase domain of *E. coli* Pol I that were highly tolerant of amino acid replacements, some of which altered the fidelity of DNA synthesis.

Properties of Antimutators and Mutators in the Polymerase Domain—Mutations in Pol I that alter the reversion frequency of *E. coli* by as much as 1000-fold were found to be widely distributed across the polymerase domain and not restricted to specific motifs or domains. Many of the antimutator polymerases exhibit polymerase activity near wild type (Fig. 3). This stands in contrast to all of the characterized antimutators, which manifested more extensive decreases in polymerase activity relative to exonuclease proofreading activity. Mutators, in some cases, had a level of activity that was greater than wild type, which was also unpredicted.

Three mechanisms have generally been proposed to account for polymerase antimutators (22, 40). First, the antimutator possesses greater selectivity during nucleotide incorporation. Second, the antimutator phenotype results from a decrease in the ratio of polymerase to exonuclease proofreading activity. Third, the ratio of polymerase to exonuclease activity remains constant, but the nascent DNA is somehow able to translocate more efficiently between the two domains. Since our antimutators lacked an essential residue in the exonuclease domain and did not demonstrate exonucleolytic activity, the most direct explanation is that these antimutators exhibit greater selectivity during nucleotide incorporation. Based on studies with nucleotide analogs of different sizes, Kool and co-workers (41) have suggested that the active site of wild-type Pol I is not optimized for maximum fidelity. Increases in the

TABLE 1

Mean substitutability indices of Pol I regions

The tolerance to mutation of each amino acid was derived by dividing the number of observed substitutions at that residue by the number of sequence reads obtained and expressed as a percentage (×100). A mean index for each region was calculated by taking the mean of all indices for residues belonging to that region. The *t* test was performed on the mean indices for region and nonregion residues to evaluate the significance of the difference, which is indicative of relative importance to protein function.

	Substitutability inside region (mean ± S.E.)	Substitutability outside region (mean ± S.E.)	*t* test (*p* value)
	%	*%*	
Entire protein	0.68 ± 0.03		
Evolutionarily conserved (Motifs A, B, and C)	0.31 ± 0.07	0.71 ± 0.05	2.4*E*-05
Family A conserved (Motifs A, B, C, 1, 2, and 6)	0.37 ± 0.04	0.76 ± 0.05	3.9*E*-08
Helices	0.74 ± 0.05	0.62 ± 0.03	0.16
β strands	0.61 ± 0.09	0.69 ± 0.03	0.46
Hinges	0.48 ± 0.10	0.70 ± 0.04	0.054
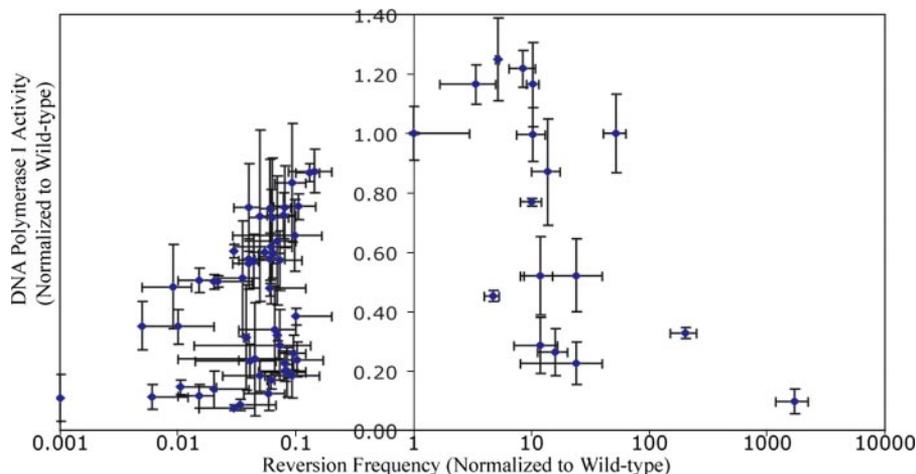
FIGURE 3. **Reversion frequency and polymerase activity of Pol I mutants.** The error frequency of each Pol I mutant was calculated based on reversion at a stop codon (TAA) within the β-lactamase gene encoded on pLA230. The polymerase activity was measured *in vitro* using the incorporation of radiolabeled dTMP into activated calf thymus DNA. The mean reversion frequency and polymerase activity (± S.E.) of each mutant were normalized to wild-type values and graphed as a *single point*. Wild-type controls occupy the position at (1, 1).
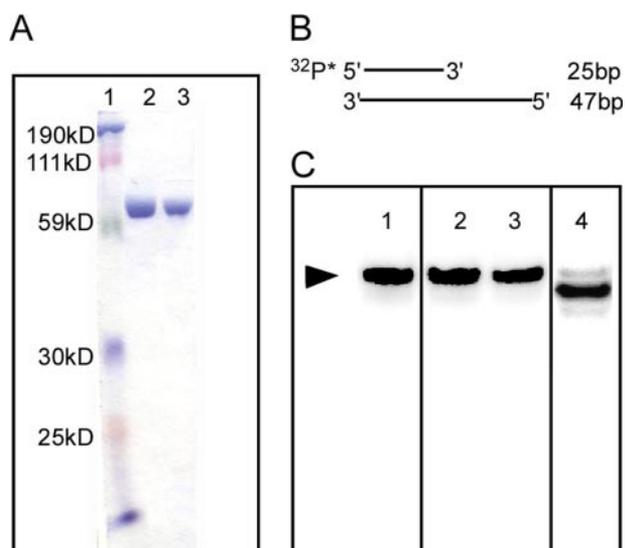
FIGURE 4. **Protein purification of mutator Pol I and test for presence of 3′→5′ exonuclease.** *A*, purified Pol I mutant Klenow K601I,A726V protein (*lane 2*) was loaded on a 10% acrylamide gel along with size standards (*lane 1*) and wild-type Klenow (*lane 3*). *B*, the purified DNA polymerases were assayed for 3′–5′ exonuclease activity by incubation with ^{32}P-radiolabeled duplex DNA with a 5′ overhang. *C*, the DNA was either incubated alone (*lane 1*), with wild-type Pol I Klenow 3′–5′ exo$^-$ (*lane 2*), the mutant K601I,A726V Klenow 3′–5′ exo$^-$ (*lane 3*), or Pol I Klenow. 3′–5′ exo$^+$ (*lane 4*). After incubation for 30 min at 37 °C, the DNA was analyzed by electrophoresis through 14% denaturing acrylamide and visualized by phosphorimaging. The *arrowhead* refers to the position of the 25-base pair oligonucleotide.

TABLE 2

Forward mutation rate of K601I,A726V mutant relative to wild type

Mutant K601I, A726V and the wild type were assayed in the M13 gapped forward mutation assay (see "Experimental Procedures") in parallel in two independent experiments. Plaques that were scored as mutant were either white or light blue. Total plaques included mutant plaques and dark blue plaques. The background mutation rates from unfilled M13 gapped plasmid from the first and second experiment were 0.0012 and 0.0010, respectively, which were subtracted from the observed mutation rates.

DNA polymerase	Plaques scored		Mutant frequency	Relative mutation rate
	Total	Mutant		
Wild type	13,326	78	0.00461	1.00
	17,301	90	0.00419	1.00
K601I,A726V	32,512	89	0.00150	0.32
	26,722	71	0.00164	0.39

TABLE 3

Error spectrum of purified mutant K601I,A726V and wild-type Klenow DNA polymerases

Mutation rates for each of the 12 possible base substitutions were calculated as described by Bebenek and Kunkel (36).

Change	Mispair	Wild type (Klenow 3′exo-)	K601I,A726V (Klenow 3′exo-)
A → G	A·dCTP	1/230,000	1/150,000
A → C	A·dGTP	1/93,000	≤1/980,000
A → T	A·dATP	1/340,000	1/890,000
G → A	G·dTTP	1/340,000	1/890,000
G → C	G·dGTP	1/29,000	1/53,000
G → T	G·dATP	1/26,000	1/100,000
C → A	C·dTTP	≤1/420,000	1/400,000
C → G	C·dCTP	1/130,000	1/360,000
C → T	C·dATP	1/23,000	1/46,000
T → A	T·dTTP	1/250,000	≤1/890,000
T → G	T·dCTP	1/360,000	1/530,000
T → C	T·dGTP	1/53,000	1/110,000
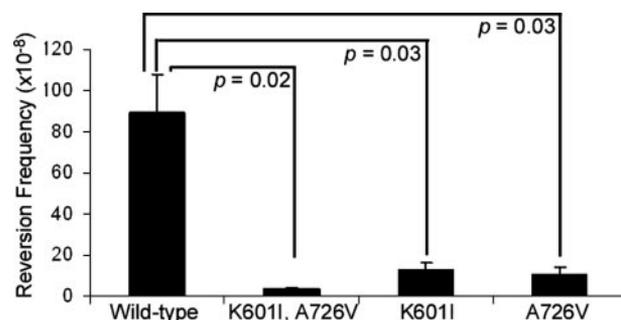
FIGURE 5. **Reversion frequencies of M-helix A726V mutant and other mutants of Pol I.** The reversion frequency of various mutants, as assayed through the β-lactamase locus on pLA230, is reported as mean ± S.E. The *p* value indicating the significance of comparisons (*t* test) between measured values is reported *above* the *bars*.

tution here may disrupt the function of the hinge and allow more time for kinetic proofreading that need not involve the exonuclease (49).

Error Spectrum of K601I,A726V Mutant—In order to verify that the enhanced fidelity measured *in vivo* reflects an intrinsic increase in the accuracy of the DNA polymerase, we compared the *in vitro* fidelity of mutant K601I,A726V and the wild-type polymerase. The error rate of the mutant was 3.1- and 2.6-fold lower than wild type as measured on the M13 gapped forward mutation assay in two independent experiments. The error spectrum of the mutant in this assay appears to be different from other Family A antimutator DNA polymerases reported (50). It is characterized by predominantly C to T transitions and G to C transversions. The K601I mutation is located near Motif 1, which has been postulated to be involved in the binding to and positioning of the DNA duplex (51), whereas the A726V mutation is located on the fingers subdomain, which interacts with the incoming nucleotide and template base as well as the catalytic magnesium ions.

We also compared the shift in mutation spectrum of the K601I,A726V double mutant with the several previously reported TaqDNA Pol I antimutators (50) and found the shifts to be dissimilar. The triple mutant A661E,I665T,F667L and the single mutant F667L (50) harbor mutations exclusively in Motif B of the fingers subdomain and demonstrate 2.8- and 2.7-fold increases in fidelity, respectively, and 3-fold decreases in activity. Their error spectrums have reduced A to T and G to T

tightness of the active site may serve as one mechanism by which the fidelity of the enzyme increases (42).

A Novel Determinant of Fidelity, the M-helix—The M-helix consists of residues 720–728 located at the juncture of the fingers and palm subdomains (43). It is located on the outer surface of the protein and lacks any direct contacts with substrates. Many antimutator mutations (a higher number than expected from random distribution) mapped to the M-helix. Our genetic analysis of a single point mutation in this region, A726V, revealed that a mutation here is sufficient to produce an antimutator phenotype (Fig. 5). Family A DNA polymerases undergo a global conformational change, from an "open" to a "closed" state, during each nucleotide addition reaction (44–46). It has been argued that kinetic proofreading of the incoming deoxynucleotide may occur at this step (47). Recent computational analyses of DNA polymerases have identified amino acid residues, including several in the M-helix, that function as "hinges" during this transformation (48). An amino acid substi-

transversions compared with wild-type Taq. On the other hand, the K601I,A726V mutant exhibits decreased A to C transversions compared with wild-type Pol I, suggesting that some of the mechanisms responsible for its increased fidelity are dissimilar from those in the Taq mutants.

Functional and Evolutionary Conservation of the Polymerase Domain—Using a formula developed by our group (30), we calculate that the overall probability for any single amino acid substitution to inactivate Pol I is 34 ± 2%. This probability is strikingly similar to results obtained for other proteins, including human 3-methyladenine DNA glycosylase (AAG) and the *E. coli lac* repressor (52). The probabilities reported for AAG and the *lac* repressor are 30–39 and 34%, respectively. Given that these three proteins are vastly dissimilar in tertiary structure and in function, this similarity of inactivation probabilities may predominantly reflect overall requirements for global processes, such as protein folding at the secondary structure level (30).

The mutational tolerance of the Pol I polymerase domain parallels the extent of its evolutionary conservation (Fig. 2*B* and Table 1). The highly conserved Motifs A, B, and C, which harbor catalytically important residues, are the least tolerant of substitutions. The second most conserved regions are Motifs 1, 2, and 6, which interact with the polymerase primer or template and are found across all Family A DNA polymerases (53, 54). Hinge region residues (48) also exhibit an intermediate level of tolerance for substitutions compared with Motifs A, B, and C (Table 1). Several of these residues, 794, 822, 826, 845, 906, and 912, are well conserved evolutionarily but mutable nevertheless (Fig. 2*B*). This quality may make them excellent targets for altering the properties of DNA polymerases without abolishing catalytic activity (55).

One residue that stood out in our analysis of evolutionary conservation is Asp[854] (Fig. 2*B*). This residue is highly conserved among bacterial Pol I proteins and interacts with Arg[821] (56). Our alignment of eukaryotic and viral Family A DNA polymerases reveals that this conservation extends across even more diverse phyla. Asp[854] has 90% identity and 97% similarity across prokaryotic, eukaryotic, and viral evolution, which is the highest for any non-substrate-binding residue. Functionally, it is also well conserved (Fig. 2*B*). Asp[854] is positioned on the T-helix, which spans both the fingers and palm subdomains of the polymerase. The residue appears to anchor this helix to the rest of the fingers through its hydrogen-bonding partners, Arg[821] (56) and the previously unreported Thr[817], which are located in the loop between β-strands 10 and 11. This stabilizing interaction is likely to be critical for polymerase function, because the T-helix contains the evolutionarily immutable residue Gln[849], which binds the base that is 3′ to the template base (57).

In summary, we have identified novel antimutator DNA polymerases with high catalytic activity, described the activity of polymerases across a broad range of fidelity mutants, pinpointed the M-helix as a novel determinant of fidelity, and presented the most comprehensive analysis of Pol I protein conservation in *E. coli* and across diverse evolutionary species. Viewed together, our results indicate that the catalytic site of *E. coli* Pol I is highly conserved and yet is able to accept multiple

substitutions. Both mutator and antimutator mutations can be accommodated. *In vivo*, many of these mutations can alter base selection at this site without severely compromising catalysis. Considering the structural and functional conservation among different DNA polymerases, it is likely that other DNA polymerases are similarly positioned. Error-prone mutations may have a role in generation of random mutations in malignancies, whereas enhanced accuracy may be of importance in delaying the onset of diseases that are associated with mutation accumulation (58).

REFERENCES

1. Bessman, M. J., Kornberg, A., Lehman, I. R., and Simms, E. S. (1956) *Biochim. Biophys. Acta* **21,** 197–198
2. Burgers, P. M. J., Koonin, E. V., Bruford, E., Blanco, L., Burtis, K. C., Christman, M. F., Copeland, W. C., Friedberg, E. C., Hanaoka, F., Hinkle, D. C., Lawrence, C. W., Nakanishi, M., Ohmori, H., Prakash, L., Prakash, S., Reynaud, C.-A., Sugino, A., Todo, T., Wang, Z., Weill, J.-C., and Woodgate, R. (2001) *J. Biol. Chem.* **276,** 43487–43490
3. Matsuda, T., Bebenek, K., Masutani, C., Hanaoka, F., and Kunkel, T. A. (2000) *Nature* **404,** 1011–1013
4. Beard, W. A., and Wilson, S. H. (2003) *Structure* **11,** 489–496
5. Plosky, B. S., and Woodgate, R. (2004) *Curr. Opin. Genet. Dev.* **14,** 113–119
6. Rattray, A. J., and Strathern, J. N. (2003) *Annu. Rev. Genet.* **37,** 31–66
7. Thomas, D. C., Roberts, J. D., Sabatino, R. D., Myers, T. W., Tan, C. K., Downey, K. M., So, A. G., Bambara, R. A., and Kunkel, T. A. (1991) *Biochemistry* **30,** 11751–11759
8. Hubscher, U., Maga, G., and Spadari, S. (2002) *Annu. Rev. Biochem.* **71,** 133–163
9. Kujoth, G. C., Hiona, A., Pugh, T. D., Someya, S., Panzer, K., Wohlgemuth, S. E., Hofer, T., Seo, A. Y., Sullivan, R., Jobling, W. A., Morrow, J. D., Van Remmen, H., Sedivy, J. M., Yamasoba, T., Tanokura, M., Weindruch, R., Leeuwenburgh, C., and Prolla, T. A. (2005) *Science* **309,** 481–484
10. Trifunovic, A., Wredenberg, A., Falkenberg, M., Spelbrink, J. N., Rovio, A. T., Bruder, C. E., Bohlooly, Y. M., Gidlof, S., Oldfors, A., Wibom, R., Tornell, J., Jacobs, H. T., and Larsson, N. G. (2004) *Nature* **429,** 417–423
11. Goldsby, R. E., Lawrence, N. A., Hays, L. E., Olmsted, E. A., Chen, X., Singh, M., and Preston, B. D. (2001) *Nat. Med.* **7,** 638–639
12. Eger, B. T., Kuchta, R. D., Carroll, S. S., Benkovic, P. A., Dahlberg, M. E., Joyce, C. M., and Benkovic, S. J. (1991) *Biochemistry* **30,** 1441–1448
13. Bebenek, K., Joyce, C. M., Fitzgerald, M. P., and Kunkel, T. A. (1990) *J. Biol. Chem.* **265,** 13878–13887
14. Kunkel, T. A., and Bebenek, K. (2000) *Annu. Rev. Biochem.* **69,** 497–529
15. Carroll, S. S., Cowart, M., and Benkovic, S. J. (1991) *Biochemistry* **30,** 804–813
16. Minnick, D. T., Bebenek, K., Osheroff, W. P., Turner, R. M. J., Astatke, M., Liu, L., Kunkel, T. A., and Joyce, C. M. (1999) *J. Biol. Chem.* **274,** 3067–3075
17. Donlin, M. J., and Johnson, K. A. (1994) *Biochemistry* **33,** 14908–14917
18. Bell, J. B., Eckert, K. A., Joyce, C. M., and Kunkel, T. A. (1997) *J. Biol. Chem.* **272,** 7345–7351
19. Miller, J. H., Suthar, A., Tai, J., Yeung, A., Truong, C., and Stewart, J. L. (1999) *J. Bacteriol.* **181,** 1576–1584
20. Tabor, S., and Richardson, C. C. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 6339–6343
21. Muzyczka, N., Poland, R. L., and Bessman, M. J. (1972) *J. Biol. Chem.* **247,**

7116–7122

22. Gillin, F. D., and Nossal, N. G. (1976) *J. Biol. Chem.* **251,** 5225–5232
23. Reha-Krantz, L. J., and Bessman, M. J. (1977) *J. Mol. Biol.* **116,** 99–113
24. Summerer, D., Rudinger, N. Z., Detmer, I., and Marx, A. (2005) *Angew. Chem. Int. Ed. Engl.* **44,** 4712–4715
25. Sharief, F. S., Vojta, P. J., Ropp, P. A., and Copeland, W. C. (1999) *Genomics* **59,** 90–96
26. Marini, F., Kim, N., Schuffert, A., and Wood, R. D. (2003) *J. Biol. Chem.* **278,** 32014–32019
27. Shinkai, A., Patel, P. H., and Loeb, L. A. (2001) *J. Biol. Chem.* **276,** 18836–18842
28. Shinkai, A., and Loeb, L. A. (2001) *J. Biol. Chem.* **276,** 46759–46764
29. Patel, P. J., Kawate, H., Adman, E., Ashbach, M., and Loeb, L. A. (2001) *J. Biol. Chem.* **276,** 5044–5051
30. Guo, H. H., Choe, J., and Loeb, L. A. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101,** 9205–9210
31. Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998) *Genome Res.* **8,** 175–185
32. Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B., and Bork, P. (2006) *Science* **311,** 1283–1287
33. Braithwaite, D. K., and Ito, J. (1993) *Nucleic Acids Res.* **21,** 787–802
34. Camps, M., Naukkarinen, J., Johnson, B. P., and Loeb, L. A. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 9727–9732
35. Glick, E., Anderson, J. P., and Loeb, L. A. (2002) *BioTechniques* **33,** 1136–1144
36. Bebenek, K., and Kunkel, T. A. (1995) *Methods Enzymol.* **262,** 217–232
37. Derbyshire, V., Freemont, P. S., Sanderson, M. R., Beese, L., Friedman, J. M., Joyce, C. M., and Steitz, T. A. (1988) *Science* **240,** 199–201
38. Witkin, E. M., and Roegner, M. V. (1992) *J. Bacteriol.* **174,** 4166–4168
39. Patel, P. H., and Loeb, L. A. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 5095–5100
40. Fijalkowska, I. J., Dunn, R. L., and Schaaper, R. M. (1993) *Genetics* **134,** 1023–1030
41. Kim, T. W., Delaney, J. C., Essigmann, J. M., and Kool, E. T. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15803–15808
42. Petruska, J., Sowers, L. C., and Goodman, M. F. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 1559–1562
43. Kornberg, A., and Baker, T. (1992) *DNA Replication*, W.H. Freeman and Co., New York
44. Johnson, S. J., Taylor, J. S., and Beese, L. S. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 3895–3900
45. Doublie, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) *Nature* **391,** 251–258
46. Li, Y., Korolev, S., and Waksman, G. (1998) *EMBO J.* **17,** 7514–7525
47. Joyce, C. M., and Benkovic, S. J. (2004) *Biochemistry* **43,** 14317–14324
48. Zheng, W., Brooks, B. R., Doniach, S., and Thirumalai, D. (2005) *Structure* **13,** 565–577
49. Hopfield, J. J. (1974) *Proc. Natl. Acad. Sci. U. S. A.* **71,** 4135–4139
50. Suzuki, M., Yoshida, S., Adman, E. T., Blank, A., and Loeb, L. A. (2000) *J. Biol. Chem.* **275,** 32728–32735
51. Loh, E., and Loeb, L. A. (2005) *DNA Repair* (*Amst.*) **4,** 1390–1398
52. Markiewicz, P., Kleina, L. G., Cruz, C., Ehret, S., and Miller, J. H. (1994) *J. Mol. Biol.* **240,** 421–433
53. Delarue, M., Poch, O., Tordo, N., Moras, D., and Argos, P. (1990) *Protein Eng.* **3,** 461–467
54. Patel, P. H., Suzuki, M., Adman, E., Shinkai, A., and Loeb, L. A. (2001) *J. Mol. Biol.* **308,** 823–837
55. Yoshikuni, Y., Ferrin, T. E., and Keasling, J. D. (2006) *Nature* **440,** 1078–1082
56. Turner, R. M., Jr., Grindley, N. D., and Joyce, C. M. (2003) *Biochemistry* **42,** 2373–2385
57. Singh, K., and Modak, M. J. (2003) *J. Biol. Chem.* **278,** 11289–11302
58. Loeb, L. A., Loeb, K. R., and Anderson, J. P. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 776–781
59. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* **18,** 2714–2723
60. Beese, L. S., Derbyshire, V., and Steitz, T. A. (1993) *Science* **260,** 352–355