THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 22, Issue of June 1, pp. 18836–18842, 2001

Printed in U.S.A.

The Conserved Active Site Motif A of *Escherichia coli* DNA Polymerase I Is Highly Mutable*

Received for publication, December 20, 2000, and in revised form, March 8, 2001
Published, JBC Papers in Press, March 12, 2001, DOI 10.1074/jbc.M011472200

Akeo Shinkai, Premal H. Patel, and Lawrence A. Loeb‡

From the Joseph Gottstein Memorial Cancer Research Laboratory, Department of Pathology, University of Washington, Seattle, Washington 98195-7705

***Escherichia coli* DNA polymerase I participates in DNA replication, DNA repair, and genetic recombination; it is the most extensively studied of all DNA polymerases. Motif A in the polymerase active site has a required role in catalysis and is highly conserved. To assess the tolerance of motif A for amino acid substitutions, we determined the mutability of the 13 constituent amino acids Val^{700}–Arg^{712} by using random mutagenesis and genetic selection. We observed that every residue except the catalytically essential Asp^{705} can be mutated while allowing bacterial growth and preserving wild-type DNA polymerase activity. Hence, the primary structure of motif A is plastic. We present evidence that mutability of motif A has been conserved during evolution, supporting the premise that the tolerance for mutation is adaptive. In addition, our work allows identification of refinements in catalytic function that may contribute to preservation of the wild-type motif A sequence. As an example, we established that the naturally occurring Ile^{709} has a previously undocumented role in supporting sugar discrimination.**

Escherichia coli DNA polymerase I (pol I)[1] is a multifunctional enzyme with roles in DNA replication, DNA repair, and genetic recombination (1). The first recognized and most thoroughly investigated of all DNA polymerases, it is key to our understanding of how DNA polymerases function as protein catalysts and as central enzymes in DNA metabolism. It belongs to one of six families of DNA polymerases, defined on the basis of amino acid sequence comparisons (2–4): family A (*e.g.* *E. coli* pol I, *Thermus aquaticus* (*Taq*) pol I, *Bacillus stearothermophilus* pol I, and T7 DNA polymerase), family B (*e.g.* DNA polymerase α and RB69 DNA polymerase), reverse transcriptase (*e.g.* human immunodeficiency virus reverse transcriptase, and murine leukemia virus reverse transcriptase), family X (*e.g.* DNA polymerase β), the pol III family, and the UmuC/DinB family (*e.g.* DNA polymerase η). Crystal structures of representative enzymes from the first four families have been determined, revealing a common overall architecture that has been likened to a human right hand, with fingers, thumb, and palm subdomains (5–9). Although the structures of the fingers and thumb subdomains vary considerably, the catalytic palm subdomains are all superimposable (10, 11). The palm subdomain includes two conserved sequences, motif A and motif C, each harboring a catalytically essential aspartic acid residue. Essential roles of motif A in catalysis include interaction with the incoming dNTP and coordination with two divalent metal ions that are required for the polymerization reaction (12–15). Motif A begins at an anti-parallel β-strand containing predominantly hydrophobic residues and is followed by a turn and an α-helix. Although there is considerable variation in the amino acid sequence of the anti-parallel β-strand, the sequence of the turn and helix, DYSQIELR, is nearly invariant among known prokaryotic family A polymerases (16).[2]

In a recent study of *Taq* pol I, we observed substantial mutability of motif A (17). This plasticity was surprising when one considers the essentiality of DNA polymerases and the marked conservation of motif A within prokaryotic DNA polymerases. To determine whether or not the plasticity within motif A was a property of DNA polymerases, we examined the mutability of motif A in *E. coli* pol I. We found that *E. coli* pol I also tolerated multiple substitution within motif A. Moreover, the overall pattern and the type of substitutions were similar to those of *Taq* pol I. Our results indicate that the mutability of motif A has been conserved in natural evolution and support the premise that toleration of mutation may be an important feature for the overall fitness of DNA polymerase active sites.

EXPERIMENTAL PROCEDURES

Construction of the pol I Random Mutant Library—The pol I gene (*polA*) of *E. coli* DH5α was amplified by colony polymerase chain reaction with 5′-ATATATATAAGCTTATGGTTCAGATCCCCCAAAATCCACTTATC-3′ and 5′-ATATATATGAATTCTTAGTGCGCCTGATCCCAGTTTTCGCCACT-3′ as primers. The 3-kilobase pair amplified fragment was digested with *Hin*dIII and *Eco*RI and then cloned under the lactose promoter into pHSG576, a low copy number plasmid that has a pol I-independent origin (18), to create pECpol I. Site-directed mutagenesis was performed on pECpol I to introduce silent mutations C to A at position 2,067 and G to C at position 2,214 of the *polA* gene (19) to create *Acc*I and *Eag*I sites, respectively, that flank the sequence encoding motif A. The resulting plasmid was named pECpol IS. To avoid contamination with incompletely cut vectors when preparing the random library, a nonfunctional stuffer vector, pECpoldum, was constructed by replacing the *Acc*I-*Eag*I 130-base pair fragment of *polA* with an oligonucleotide fragment (5′-ATACGATCGATCTGCAGCGATCC-3′ and 5′-GGCCGGATCGCTGCAGATCGATCGT-3′).

The pol I random library was constructed by annealing two single-stranded DNA oligonucleotides containing segments with random sequences: Oligo 1 was a 104-mer corresponding to the sense nucleotides 2,053–2,156, and containing an *Acc*I site for cloning (5′-GAAGGTCGTCGTATACGCCAGGCGTTTATTGCGCCAGAGGATTAT[GTGATTGTCTCAGCGGACTACTCGCAGATTGAACTGCGC]ATTATGGCGCATC-

* This work was supported by National Institutes of Health Grants CA78885, CA39903, and CA74184 (to L. A. L.) and by a grant from Kyowa Hakko Kogyo Co., Ltd. (to A. S.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: The Joseph Gottstein Memorial Cancer Research Lab., Dept. of Pathology, University of Washington, Box 357705, Seattle, WA 98195-7705. Tel.: 206-543-6015; Fax: 206-543-3967; E-mail: laloeb@u.washington.edu.

[1] The abbreviations used are: pol I, DNA polymerase I; *Taq*, *T. aquaticus*; DTT, dithiothreitol; exo$^-$, exonuclease-minus; PAGE, polyacrylamide gel electrophoresis.

[2] Patel, P. H., Suzuki, M., Adman, E., Shinkai, A., and Loeb, L. A. (2001) *J. Mol. Biol., in press.*

TTTCGCG-3′); Oligo 2 was an 89-mer corresponding to antisense strand nucleotides 2,225–2,137 and containing an *Eag*I site (5′-AACACTTCTGCGGCCGTTGCCCGGTGGATATCTTTTCCTTCCGCGAATGCGGTCAGCAAGCCTTTGTCACGCGAAAGATGCGCCATAAT-3′). The bracketed nucleotides in Oligo 1 were synthesized to contain 88% wild-type nucleotide and 4% each of the other three nucleotides at every position. The 20-base pair complementary regions of hybridization are underlined. Oligo 1 and Oligo 2 were annealed at their nonrandom complementary regions by mixing 250 pmol of each in 20 μl of H_2O and heating to 95 °C for 5 min, followed by cooling for 2 h to room temperature. The partially duplex oligonucleotide was extended by incubation with 50 units of *E. coli* pol I Klenow fragment (New England BioLabs, Beverly, MA) for 2 h at 37 °C in a 0.3-ml reaction mixture containing 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 7.5 mM DTT, and 0.5 mM of all four dNTPs. The resulting DNA was digested with *Acc*I and *Eag*I, purified, and inserted into pECpoldum in place of the stuffer fragment. Plasmids containing the random library were transformed into *E. coli* XLIBlue, and the number of transformed cells was determined by plating an aliquot onto LB agar plates containing 30 μg/ml of chloramphenicol. The remainder of the library was amplified by growing the transformed *E. coli* XLIBlue in 3 liters of 2× YT medium for 16 h at 37 °C, and the random library, pECpolLib, was then purified.

Genetic Selection for Active Mutants—*E. coli* JS200 (*recA718polA12*) (20, 21) was transformed with plasmids pHSG576, pECpol IS, pECpoldum, and pECpolLib. Thereafter, 1 ml of nutrient broth containing 0.4% NaCl was added, and the cells were incubated for 1 h at 37 °C. A small fraction of the mixture was then plated in duplicate onto nutrient agar plates containing 0.4% NaCl, 12.5 μg/ml tetracycline, and 30 μg/ml chloramphenicol; one plate was incubated at 30 °C, and the other was incubated at 37 °C overnight, and the resulting colonies were counted. Only paired samples containing less than 1,500 colonies at 30 °C were analyzed because dense plating of the cells leads to elevated background at 37 °C.

DNA Sequence Analysis of the Motif A Region—Plasmids carrying the mutant pol I gene were prepared, and the 0.6-kilobase pair region covering the motif A region was amplified by polymerase chain reaction with 5′-GATACCATGCTGGAGTCCTACATTC-3′ and 5′-ACGGCGTTGCTCGCTGGTGACGGTT-3′ as primers. Following purification of the polymerase chain reaction product, the *Acc*I-*Eag*I 130-base pair fragment was sequenced by using 5′-TTATCGTCAACCGATCCTAACCTGCA-3′ as a primer.

Preparation of E. coli Cell Extracts—Recombinant *E. coli* JS200 cells were cultured at 30 °C in 4 ml of 2× YT medium supplemented with 12.5 μg/ml tetracycline and 30 μg/ml chloramphenicol. At the exponential growth phase (A_{600} = 0.2–0.5), pol I expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside. After further incubation for 4 h (A_{600} = 2), cells from 1.5 ml of culture were collected, washed with 1 ml of 20 mM sodium phosphate (pH 7.2), suspended in 0.1 ml of the same buffer, and 5 μl of 10 mg/ml lysozyme was added. The cells were disrupted by freezing at −80 °C for 16 h and thawing on ice for 2 h. The cell extract was collected by centrifugation at 15,000 rpm for 15 min.

DNA and RNA Polymerase Assays—DNA polymerase activity was measured at 42 °C for 10 min in 20-μl reaction mixtures containing 0.1 μg of gapped calf thymus DNA (22), 12.5 μM each dNTP, 50 nM [α-^{32}P]dTTP (3,000 Ci/mmol; PerkinElmer Life Sciences), and 2 μl of cell extract in 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 7.5 mM DTT. The reaction was terminated by addition of 0.5 ml of 10% trichloroacetic acid followed by 0.1 ml of 0.1 M sodium pyrophosphate. The ^{32}P-labeled DNA was collected onto glass fiber filters, and radioactivity was measured by using scintillation counter as described (23). The assay for RNA polymerase activity was the same, except 12.5 μM rGTP was substituted for dGTP.

Construction of High Copy Number Vectors for HisKlenow(exo$^-$) Expression—Site-directed mutagenesis was performed on pECpol IS to introduce an A to C transversion at position 1,271, changing the corresponding Asp^{424} to Ala and inactivating the 3′-exonuclease activity (24). Then, with this plasmid as a template and 5′-CAGACGAACATATGCACCATCATCACCATCACATTTCTTATGACAACTACGTCACCATCCTTGAT -3′ and 5′-ATATATATGAATTCTTAGTGCGCCTGATCCCAGTTTTCGCCACT-3′ as primers, polymerase chain reaction was performed to construct the HisKlenow(exo$^-$) gene. The amplified fragment was digested with *Nde*I and *Eco*RI and cloned under the λP_L promoter of pLEX (Invitrogen, Carlsbad, CA). High expression vectors for mutant pol I proteins were constructed by substituting the 1.1-kilobase pair *Sac*I-*Eco*RI fragment of the wild-type pol I gene on the expression vector with the corresponding fragment of the mutant gene.

Expression and Purification of HisKlenow(exo$^-$) Proteins—Recombinant HisKlenow(exo$^-$) proteins were expressed and purified by using the P_L Expression System (Invitrogen), and His-Bond kits (Novagen, Madison, WI), respectively, essentially according to the manufacturer's directions. The expression plasmid was introduced into *E. coli* GI724 (F$^-$, λ$^-$, *lac*I^q, *lac*PL8, *amp*C::P_{trp} cI, *mcr*A, *mcr*B, INV(*rnn*D-*rnn*E)), and cells were grown at 30 °C in 40 ml of induction medium composed of 1× M9 salts, 0.2% casamino acids, 0.5% glucose, 1 mM $MgCl_2$, and 100 μg/ml ampicillin. When an A_{550} of 0.5 was attained, tryptophan was added to a final concentration of 100 μg/ml, and the culture was incubated at 37 °C for a further 4 h. The cells were collected by centrifugation, washed with 40 ml of phosphate-buffered saline, and suspended in 4 ml of 1× binding buffer (5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9) containing 200 μg/ml lysozyme and 0.5 mM phenylmethylsulfonyl fluoride. Extracts were prepared by freezing at −80 °C for 16 h and thawing on ice for 2 h, followed by centrifugation at 15,000 rpm for 15 min. The extract was then applied to a 1-ml Ni^{2+}-resin column, washed with 10 ml of 1× binding buffer, and eluted with buffer composed of 60 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9. The eluted sample was concentrated to ~1 mg/ml by using a Centricon-30 microconcentrator (Amicon, Beverly, MA), diluted with an equal volume of glycerol containing 2 mM DTT, and stored at −80 °C. Protein concentrations were determined by the method of Bradford (25).

Kinetic Analysis of Nucleotide Incorporation—A steady-state kinetic analysis was performed based on the method of Boosalis *et al*. (26). A 47-mer template (3′-GCGCGGCTTAAGGGCGATCGTTATAGCTTAAGGCCTTTAAAGGGCCC-5′) was hybridized with one of four 5′-^{32}P end-labeled primers: the 23-mer (5′-CGCGCCGAATTCCCGCTAGCAAT-3′), the 24-mer (5′-CGCGCCGAATTCCCGCTAGCAATA-3′), the 25-mer (5′-CGCGCCGAATTCCCGCTAGCAATAT-3′), or the 26 mer (5′-CGCGCCGAATTCCCGCTAGCAATATC-3′). Primer/template (5 nM) was incubated for 5 min at 37 °C in a reaction mixture containing limiting amounts of HisKlenow(exo$^-$) protein and varying concentrations of each dNTP or rNTP in 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 7.5 mM DTT. The ranges of nucleotide substrate concentrations used were 0.5–10 nM for dNTP incorporation, 0.5–17.5 μM for rNTP incorporation by the wild type and E710D mutant, and 2.5–400 nM for rNTP incorporation by the I709F mutant. Following termination of the reaction by addition of 2.5 μl of formamide solution, the products were analyzed by 14% PAGE and quantified by PhosphorImager analysis (27).

RNA Synthesis—The 47-mer template and 5′-^{32}P end-labeled 24-mer primer used for kinetic analysis of rNTP incorporation were hybridized and incubated at 37 °C for 5–60 min in 10-μl reaction mixtures containing 50 nM HisKlenow(exo$^-$) protein and 1 μM each of all four rNTPs in 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 0.5 mM $MnCl_2$, 7.5 mM DTT. After the reaction was terminated by the addition of 2.5 μl of formamide solution, the products were analyzed by 14% PAGE followed by autoradiography.

RESULTS

Creation and Genetic Selection of Motif A Mutants—We used random sequence mutagenesis to create substitutions within the 13 contiguous amino acids comprising motif A of *E. coli* pol I (Val^{700}–Arg^{712}). We then selected functional mutants in *E. coli* JS200 (*recA718 polA12*), a strain that contains a temperature-sensitive mutation in the pol I gene (*polA*) and can be propagated at 30 °C but not at 37 °C (20, 21). Recombinant wild-type *polA* was able to fully complement the temperature-sensitive phenotype, such that *E. coli* JS200 harboring the plasmid pECpol IS exhibited a 100% survival rate at 37 °C relative to 30 °C. The recombinant strain carrying pECpoldum, a nonfunctional stuffer vector, showed a 0.5% survival rate at 37 °C, indicating that the background for our complementation-based selection assay was 0.5%.

The randomly mutated *E. coli* pol I library consisted of 500,000 independent clones. DNA sequencing of 26 unselected clones indicated that the average number of amino acid changes within motif A was three. The unselected library included 10% dummy vectors, and 40% of the clones had a deletion and/or insertion within or outside of motif A, which may have been introduced in the process of library construction. Following transformation of the library into *E. coli* JS200, 8% of the clones formed colonies at 37 °C, relative to 30 °C. After subtracting the background, we estimated that there were 37,500 independent clones encoding active pol I proteins.

Analysis of Selected Motif A Mutants—To establish the spectrum of mutations that restored growth of *E. coli* JS200, we randomly picked 280 colonies that grew at 37 °C, assayed the DNA polymerase activity in cell extracts at 42 °C, isolated the plasmids, and sequenced the 0.15-kilobase pair linker portion containing the 39-base pair randomized region. DNA polymerase activity in extracts of *E. coli* JS200 carrying the parent plasmid pHSG576 or the stuffer vector pECpoldum was 1–4% of that in extracts of cells carrying pECpol IS that expresses wild-type *E. coli* pol I, indicating that the background polymerase activity is 1–4% (data not shown). We found three clones carrying pECpoldum that exhibited higher activity than clones harboring the original pECpoldum, *i.e.* 20–25% of wild-type pol I activity; we did not observe complementation when fresh JS200 cells were retransformed by these three stuffer vectors, indicating that the growth observed at 37 °C was likely due to mutation of the host cells. Twelve clones showed very low activity, less than 10% that of clones carrying pECpol IS. Of the 280 total clones, we estimated that 17–18 should represent background, in that 8% of the total library form colonies at 37 °C and 0.5% of the total library are false positives. Therefore, we attributed the 15 clones just described to background and did not analyze them further. Of the remaining 265 clones, 32 had 1 amino acid substitution and 1 had a deletion of 2 amino acids, all of which were outside of motif A; these clones were also not further investigated. No frameshift mutations were observed among the 280 selected clones.

The remaining 232 selected, active mutants harbored one to five amino acid substitutions within motif A. As illustrated in Fig. 1*A*, the average number of mutations was 1–2. The levels of DNA polymerase activity in extracts of the mutants are shown in Fig. 1*B*, as a function of the number of amino acid changes. 70% of the 232 mutants retained DNA polymerase activity comparable with that of wild type (within 60%-200%), including almost all of the single mutants and even six of the ten mutants with four amino acid replacements. Moreover, 36% of mutants exhibited activity equal to or greater than the wild type. The number of mutants exhibiting moderate (30–60% of wild type) or low (10–30% of wild type) activity followed a Poisson distribution relative to the number of amino acid substitutions, with a median of two or three amino acid substitutions per clone.

In Fig. 2 the amino acid substitutions observed in the selected active clones with one (Fig. 2*A*), two (Fig. 2B), or three to five (Fig. 2C) mutations in motif A are shown; the distribution of mutations was similar in the three groups. Six motif A residues (Val^{700}, Val^{702}, Ser^{703}, Ala^{704}, Ser^{707}, and Gln^{708}) tolerated a wide spectrum of substitutions, whereas six others (Ile^{701}, Tyr^{706}, Ile^{709}, Glu^{710}, Leu^{711}, and Arg^{712}) tolerated predominantly conservative substitutions, and only the catalytically essential residue Asp^{705} was immutable. Glu^{710} was substituted solely by Asp, indicating that negative charge at this position may be indispensable for polymerase activity *in vivo*. DNA polymerase activity in extracts of the 53 different mutants with a single amino acid replacement is indicated in Fig. 3. Interestingly, most single mutants with a replacement within the N-terminal 5 amino acids, which form a strand of the structurally conserved anti-parallel β-sheet, exhibited higher than wild-type activity. In contrast, activity tended to be reduced when the C-terminal α-helix region, whose primary structure is practically invariant in the prokaryotic pol I family, was mutated. Thus, even though both the N- and C-terminal portions of motif A are highly mutable, the effects on catalytic activity differ.

Effect of Motif A Mutations on rNTP Discrimination—The foregoing results indicate that the primary structure of motif A in *E. coli* pol I is plastic and that mutations in motif A can be associated with a high degree of biologic and catalytic function. To assess how the highly functioning variant pol I might differ

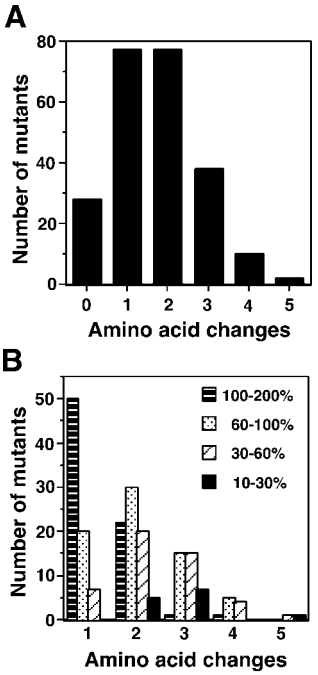

FIG. 1. **Amino acid substitutions in motif A of 232 genetically selected, active *E. coli* pol I mutants.** Mutations were introduced into the 13 codons encoding motif A in the plasmid-borne pol I gene by using random sequence mutagenesis. Active mutants were then isolated by positive genetic selection for variants that complemented the temperature-sensitive growth phenotype of an *E. coli* strain harboring a temperature-sensitive endogenous pol I. *A*, distribution of amino acid substitutions in clones that exhibited >10% of wild-type DNA polymerase activity in an *in vitro* assay, relative to the number of amino acid changes. *B*, distribution of DNA polymerase activity in extracts of cells expressing pol I mutants, relative to the number of the amino acid changes. The *vertical bars* indicate the level of DNA polymerase activity relative to wild type, as follows: *black* with *horizontal stripes*, 100%-200%; *stippled*, 60%-100%; *diagonal stripes*, 30%-60%; *solid*, 10%-30%. Wild-type activity is that observed in extracts of cells expressing wild-type pol I.

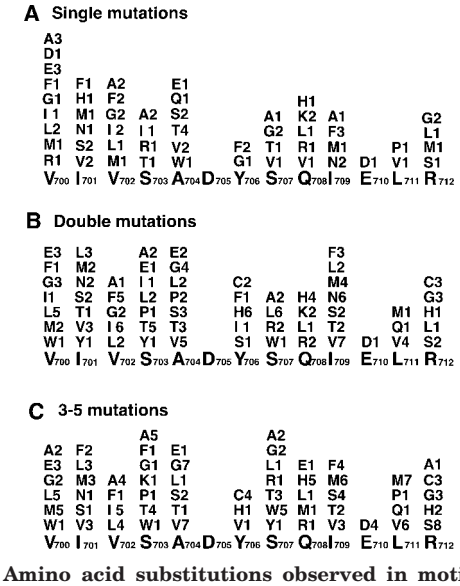

FIG. 2. **Amino acid substitutions observed in motif A of selected, active pol I mutants harboring one (*A*), two (*B*), or three to five (*C*) amino acid changes.** Amino acid substitutions at each residue are listed in alphabetical order from *top* to *bottom*, along with the number of times each substitution was observed.

A1.2
D1.1
E1.0
F1.2 F1.7 A1.3 E1.0
G1.3 H1.0 F1.3 Q1.5 H0.9
I 1.1 M1.6 G1.1 A1.5 S0.8 A1.0 K0.4 A0.7 G0.6
L1.3 N1.7 I 1.1 I 0.9 T1.5 G1.2 L0.7 F0.9 L0.9
M1.2 S1.1 L0.9 R1.1 V1.6 F1.1 T1.2 R0.3 M0.6 P0.9 M0.9
R0.8 V1.2 M1.2 T1.1 W1.5 G0.7 V0.6 V0.6 N0.9 D0.9 V1.1 S0.6
V_{700} I_{701} V_{702} S_{703} A_{704} D_{705} Y_{706} S_{707} Q_{708} I_{709} E_{710} L_{711} R_{712}

FIG. 3. **DNA polymerase activity of pol I mutants with single amino acid substitutions in motif A.** Substitutions at each residue are listed alphabetically from *top* to *bottom*, followed by DNA polymerase activity, relative to wild type. Polymerase activity in cell extracts was assayed at 42 °C by measuring the incorporation of [α-^{32}P]dTTP into gapped calf thymus DNA. Mutant activities are expressed relative to wild type (1.0), *i.e.* the activity observed in extracts of cells expressing recombinant wild-type pol I.

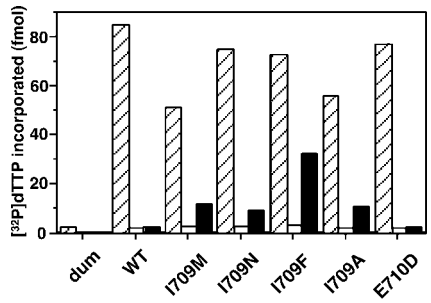

FIG. 4. **Incorporation of ribonucleotides by wild-type and mutant pol I with single amino acid substitutions at I709 or E710.** Polymerase activity in cell extracts was measured in assay mixtures containing gapped calf thymus DNA as a template-primer and either four dNTPs (dG, dA, dC, and dT) (*diagonally striped bars*), three dNTPs (dA, dC, and dT) (*open bars*), or three dNTPs plus rGTP (*solid bars*). The amount of [α-^{32}P]dTTP incorporated into DNA is shown. *dum*, cells expressing the noncomplementing dummy plasmid pECpoldum; *WT*, wild type.

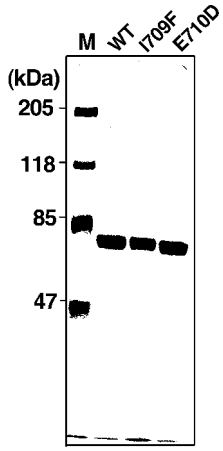

FIG. 5. **SDS-PAGE analysis of purified Klenow fragment derived from wild-type (*WT*), I709F, and E710D pol I.** Exonuclease-minus Klenow fragments were constructed as described under "Experimental Procedures" and expressed in *E. coli* as N-terminal hexahistidine fusions. The proteins were purified on nickel resin columns and analyzed by SDS-PAGE followed by Coomassie Brilliant Blue R-250 staining. *M*, molecular mass markers.

from wild type catalytically, we tested all 53 different single mutants shown in Fig. 3 for altered sugar selectivity. We did this by substituting rGTP for dGTP in the standard DNA polymerase assay. Wild-type pol I exhibited poor incorporation of rGTP in this assay, as did all the single mutants except the four having an Ile^{709} to Met, Asn, Phe, or Ala substitution (Fig. 4 and data not shown). Mutants carrying a I709S substitution, such as I701M/A704G/I709S and V700A/L711V/I709S, also exhibited efficient incorporation of rGTP. In contrast, neither the single mutant E710D nor additional mutants such as I701V/V702I/E710D and I709V/E710D/R712S were effective in incorporating rGTP (data not shown).

We chose the single mutant I709F, which showed the most efficient rNTP incorporation, for detailed analysis of rNTP discrimination. We also analyzed E710D as a reference enzyme; E710D did not exhibit enhanced rNTP incorporation but has displayed modestly reduced discrimination against rNTPs in other assays (28). To eliminate the possibility of proofreading by the 3′-5′ exonuclease activity of *E. coli* pol I, we constructed exonuclease-deficient derivatives of the Klenow fragment (24), containing an intact polymerase domain. The wild-type and mutant Klenow fragments were expressed in *E. coli* as N-terminal hexahistidine fusion proteins (HisKlenow(exo$^-$)) and purified by one-step nickel affinity chromatography. The wild-type, I709F, and E710D preparations each yielded a single band of ~68 kDa in SDS-polyacrylamide gels (Fig. 5); the estimated purity was ≥95% in all cases, and importantly, no bands of 109 kDa (the molecular mass of endogenous pol I) were detected.

We used a steady-state, gel-based assay employing oligonucleotide primer templates (26) to analyze the kinetics of rNTP incorporation by the purified exonuclease-deficient Klenow fragments. The wild-type and mutant proteins showed typical Michaelis-Menten saturation kinetics when initial velocity was plotted against the concentration of either rNTP or dNTP (data not shown). The parameters K_m and V_{max} were derived by hyperbolic curve fitting and were used to calculate k_{cat}, catalytic efficiency, and rNTP/dNTP discrimination factors (Table I). The catalytic constant k_{cat} was obtained by dividing V_{max} by the enzyme concentration. The catalytic efficiency, expressed as k_{cat}/K_m, is a measure of the efficiency of nucleotide incorporation. The discrimination factor dNTP/rNTP, calculated as the ratio of efficiencies for incorporation of dNTP *versus* the corresponding rNTP, is a measure of intrinsic enzymatic specificity for the correct sugar. As indicated in Table I, the discrimination against rNTPs exhibited by the wild-type enzyme ranged from 650- to 53,000-fold; the discrimination was associated exclusively with elevated K_m values for rNTPs, the rate constants for all substrates being essentially the same. The mutant I709F was indistinguishable from wild type with respect to incorporation of dNTPs, in accord with the near wild-type activity in the standard nucleotide incorporation assay employing gapped DNA and all four dNTPs (Fig. 4). However, the ability of the I709F mutant to discriminate against rNTPs is impaired; the observed discrimination factors ranged from 20- to 80-fold less than wild type, virtually entirely because of decreased K_m values for rNTPs relative to the wild-type value. These findings indicate that Ile^{709} in wild-type pol I functions to exclude ribonucleotides from the genome, at least in part via diminished incorporation of rNTPs. As shown below, the I709F mutant also discriminates against extension of incorporated ribonucleotides less efficiently than the wild-type exo$^-$ Klenow fragment. In contrast to the I709F mutation, the E710D substitution had little or no effect on either incorporation of dNTPs or discrimination against rNTPs. These observations are in accord with essentially wild-type activity in both the standard DNA polymerase and ribonucleotide incorporation assays.

Exclusion of rNTPs from DNA involves both discrimination against incorporation of rNTPs and discrimination against extension of DNA chains bearing a 3′-terminal ribonucleotide residue. To assess these two factors concurrently, we examined how well the mutant HisKlenow(exo$^-$) proteins were able to incorporate multiple rNTPs sequentially, *i.e.* to act as RNA polymerases. To do this, we incubated the proteins with a 5′-^{32}P-labeled DNA primer template in the presence of all four rNTPs. As shown in Fig. 6, the wild-type protein added seven

TABLE I
Kinetic analysis of dNTP and rNTP incorporation by wild-type and mutant exo⁻ Klenow fragments

Protein	Nucleotide	dNTP			rNTP			dNTP/rNTP discrimination[a]
		$k_{cat} \times 10^3$	K_m	k_{cat}/K_m	$k_{cat} \times 10^3$	K_m	k_{cat}/K_m	
		s^{-1}	nM	$M^{-1}\ s^{-1}$	s^{-1}	nM	$M^{-1}\ s^{-1}$	
Wild type	G	3.3 ± 0.4	2.5 ± 0.7	1.3×10^6	2.4 ± 0.1	(1.2 ± 0.1) × 10^3	2.0×10^3	650
	A	3.1 ± 0.3	1.6 ± 0.4	1.9×10^6	5.4 ± 1.7	(17 ± 8.8) × 10^3	3.2×10^2	5,900
	C	3.0 ± 0.3	1.9 ± 0.5	1.6×10^6	2.3 ± 0.1	(1.1 ± 0.1) × 10^3	2.1×10^3	760
	T/U	3.1 ± 0.2	1.3 ± 0.3	2.4×10^6	3.7 ± 1.1	(83 ± 29) × 10^3	45	53,000
I709F	G	3.2 ± 0.4	2.0 ± 0.7	1.6×10^6	2.6 ± 0.1	14 ± 1.1	1.9×10^5	8.4
	A	3.1 ± 0.2	1.1 ± 0.3	2.8×10^6	3.8 ± 0.2	(3.9 ± 0.4) × 10^2	9.7×10^3	290
	C	2.9 ± 0.2	1.5 ± 0.3	1.9×10^6	2.4 ± 0.1	13 ± 1.8	1.8×10^5	11
	T/U	2.8 ± 0.2	0.8 ± 0.3	3.5×10^6	2.8 ± 0.5	(1.4 ± 0.3) × 10^3	2.0×10^3	1,800
E710D	G	3.0 ± 0.2	1.6 ± 0.4	1.9×10^6	2.7 ± 0.1	(1.0 ± 0.1) × 10^3	2.7×10^3	700
	A	3.0 ± 0.2	0.9 ± 0.2	3.3×10^6	3.7 ± 0.4	(4.6 ± 1.4) × 10^3	8.0×10^2	4,100
	C	2.9 ± 0.2	1.4 ± 0.3	2.1×10^6	2.4 ± 0.1	(0.8 ± 0.1) × 10^3	3.0×10^3	700
	T/U	3.4 ± 0.2	1.3 ± 0.3	2.6×10^6	4.5 ± 0.8	(35 ± 8.1) × 10^3	1.3×10^2	20,000

[a] dNTP/rNTP discrimination equals efficiency of dNTP incorporation (k_{cat}/K_m) relative to the efficiency of rNTP incorporation.

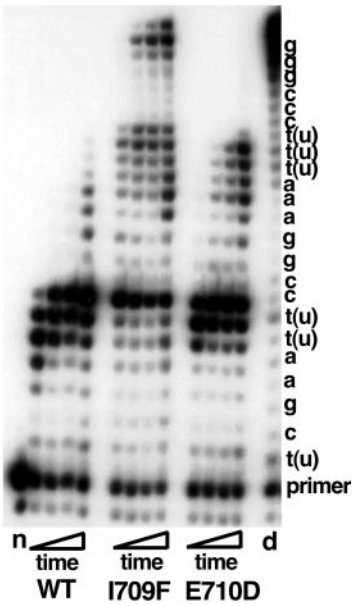

FIG. 6. **RNA synthesis by exonuclease-minus HisKlenow fragments derived from wild-type, I709F, and E710D pol I.** A 5′-^{32}P-labeled DNA primer template was incubated with either wild-type (*WT*), I709F, or E710D HisKlenow(exo⁻) protein (50 nM) and 1 μM each of all four rNTPs at 37 °C for 5, 10, 15, or 60 min. The products were analyzed by 14% PAGE, followed by autoradiography. *Lane n*, incubation was for 60 min in the presence of 50 nM wild-type enzyme and no rNTPs. *Lane d*, incubation was for 15 min in the presence of 50 nM wild-type enzyme and 50 μM of each dNTP. Nucleotides to be incorporated are shown at the *right*.

residues to the primer, stopping upon the addition of the two uracil residues; only at 60 min was addition of further nucleotides observed. In contrast, the I709F mutant was able to overcome the barrier comprised of two uracil residues, as well as the downstream barrier comprised of three uracil residues. We conclude that the I709F substitution permits not only more efficient incorporation of rNTPs, but more efficient extension as well. The E710D mutant extended the primer at a slightly greater rate than the wild-type protein, indicative of modestly reduced discrimination against extension of 3′-ribonucleotide termini.

DISCUSSION

Motif A is shared among DNA polymerases (2), is an essential part of the polymerase active site (12–15, 29), and is highly conserved among prokaryotic DNA polymerase A family members (16).[2] Interestingly, motif A sequences of modern *E. coli* strains from around the world were recently found to be identical (17), even though the bacteria divide more than 100 times each year (30), at mutation rates of 10^{-5}/nucleotide/division in mutators (31–33) to 10^{-9}/nucleotide/division in nonmutators (34). Despite this conservation in nature, we show here that motif A in *E. coli* pol I can tolerate a substantial mutational burden, and most of the 13 constituent amino acids are replaceable, yielding highly competent variant polymerases. *E. coli* strains harboring active mutant pol I that were selected in a genetic complementation system are fit to replicate repetitively, both in liquid broth and on solid agar at 37 °C. These observations are consistent with biochemical data showing that the mutant proteins possess wild type-like DNA polymerase activity *in vitro*. We found that only one residue, the catalytically essential Asp^{705}, was immutable; the corresponding residue in *Taq* pol I coordinates with the two metal ligands required for catalysis (13, 15). Substitution of Glu^{710} was restricted to Asp; in crystals of a closed ternary complex of T7 DNA polymerase complexed with its substrates (13), the glutamate residue equivalent to Glu^{710} is hydrogen-bonded with a tyrosine residue in the O-helix within the fingers subdomain. We conclude that a Asp residue at position 705 and a negative charge at position 710 are indispensable for maintaining polymerase activity; this conclusion is in agreement with the deleterious effects of the D705A and E710A substitutions on catalysis (35). Both the N- and C-terminal parts of motif A tolerated a wide spectrum of substitutions. DNA polymerase activity associated with single amino acid substitutions within the N-terminal 5 amino acid residues was as high or higher than that of the wild-type enzyme. These residues form part of an anti-parallel β-sheet structure that is believed to accommodate the triphosphate moiety of the incoming dNTP and may be a potential target for engineering of pol I derivatives with altered properties. In contrast, amino acid substitutions within the C-terminal five residues tended to be associated with reduced activity.

To further assess variation in catalytic properties among the selected active mutants, we screened them for incorporation of ribonucleotides. Among the 53 amino acid substitutions analyzed, we found that certain substitutions at Ile^{709} (Phe, Met, Ala and Asn) permit more efficient utilization of rNTPs and that the phenylalanine substitution permitted the most extensive incorporation of rNTPs. The corresponding isoleucine to phenylalanine substitution has so far not been found among random sequence substitutions in *Taq* pol I (17, 27). The I709F substitution essentially converts *E. coli* pol I from a DNA-dependent DNA polymerase to an enzyme that can effectively use both DNA and RNA substrates. We conclude that isoleucine at position 709 contributes to sugar discrimination by wild-type pol I and that this function may promote conservation of the wild-type motif A sequence. Based on analysis of a structural

A *E. coli* **pol I**

			A9										
			E1										
A5			F1										
D1	F3		G1	E4									
E9	H1		I2	G11			A5		A1				
F2	L6		K1	L3		C6	G4	E1	F10				A1
G6	M6	A7	L2	P2		F3	L7	H10	L2				C6
I2	N4	F8	P2	Q1		G1	R3	K4	M11				G8
L12	S5	G4	R1	S7		H7	T4	L3	N8		M8		H3
M8	T1	I13	T10	T8		I1	V1	M1	S6		P2		L2
R1	V8	L7	W1	V14		S1	W8	R4	T4		Q2		M1
W2	Y1	M1	Y1	W1		V1	Y1	V1	V10	D6	V11		S11
V_{700}	I_{701}		V_{702}	S_{703}	A_{704}	D_{705}	Y_{706}	S_{707}	Q_{708}	I_{709}	E_{710}	L_{711}	R_{712}
HP 31	23	36	15	18	0	5	21	5	34	0	19	4	
Po 6	12	4	14	29	0	15	9	10	18	0	4	28	
(-) 10	0	0	1	4	0	0	0	1	0	6	0	0	
(+) 1	0	0	2	0	0	0	3	8	0	0	0	0	
T 48	35	40	32	51	0	20	33	24	52	6	23	32	

B *Taq* **pol I**

E1													C1
F2		A4	D12										F1
G1		E11	E2										G5
H2	F12	F1	G11				C13						I2
I21	I1	G3	I1	F2			D1		A1				K18
K2	M5	I2	K1	I6			G10		K12				L1
M1	Q2	K2	N2	M8			I1		L7				M13
P2	R3	L29	P1	P3			L1	H17	M18		A1		Q1
Q8	S3	M21	S6	Q9			N20	L15	N2		F10		S15
R8	T1	P1	T14	R1		F12	R30	N1	Q2		I6		T12
T1	V4	S1	V22	S2		H2	T3	R1	T16		P4		V1
V12	W16	T1	Y3	V21		W1	Y1	S1	V13	D15	V7		W5
L_{605}	L_{606}	V_{607}	A_{608}		L_{609}	D_{610}	Y_{611}	S_{612}	Q_{613}	I_{614}	E_{615}	L_{616}	R_{617}
HP 38	38	58	24	40	0	13	2	15	39	0	28	23	
Po 12	6	5	36	11	0	2	47	19	20	0	0	34	
(-) 1	0	11	14	0	0	0	1	0	0	15	0	0	
(+) 8	3	0	1	1	0	0	30	1	12	0	0	18	
T 59	47	74	75	52	0	15	80	35	71	15	28	75	

FIG. 7. **Similarity of amino acid replacements observed in motif A of active mutants of *E. coli* pol I (*A*) and *Taq* pol I (*B*).** Mutants of each polymerase were selected by using the same functional complementation protocol and bear mutations only within motif A. Substitutions at each locus are listed alphabetically from *top* to *bottom*, along with the number of times each mutation was observed. The *Taq* pol I mutations are those found by Patel and Loeb (17). *HP*, *Po*, *(+)*, and *(−)* denote the number of hydrophobic, polar, positively charged, and negatively charged amino acid residues, respectively. *T* represents the total number of substitutions observed.

model of *Taq* pol I bound with DNA and an rNTP (27), we infer that the ribose ring of the rNTP interacts with Ile^{709} (specifically, with the methyl group at the β-carbon). Substitution of Ile^{709}, which is a highly mutable residue, may allow rNTP binding and incorporation by directly altering the position of the incoming nucleotide and/or by indirect effects that introduce a local conformational change in the protein. The neighboring residue, Glu^{710}, has been proposed to function as the "steric gate" that excludes the 2′-hydroxyl group of an incoming rNTP (27, 28). Therefore, alteration of Ile^{709} might allow repositioning of the Glu^{710} residue in the chain so that steric exclusion of rNTP is no longer as effective. Interestingly, the E710D substitution previously found to incorporate ribonucleotides (28) had little or no effect on the kinetics of rNTP incorporation in our assay but did confer an increased ability to add multiple rNTP residues to a growing oligomer.

We have shown that motif A in *Taq* pol I is highly mutable (17) and undertook the present work to examine the extent to which this mutability might be conserved in evolution. As summarized in Fig. 7, the amino acid substitutions observed at each position in *E. coli* pol I and *Taq* pol I are quite similar. Amino acids 700–704 in *E. coli* pol I and the corresponding residues 605–609 in *Taq* pol I tolerate predominantly hydrophobic substitutions, *i.e.* 60 and 64% of total substitutions, respectively. Asp^{705} in *E. coli* and Asp^{610} in *Taq* pol I are nonsubstitutable. Residues 706–709 and 712 in *E. coli*, and the corresponding residues in *Taq* pol I tolerate a large number of polar and charged substitutions, *i.e.* 57 and 67%, respectively. Of the 37 substitutions found in *E. coli* at these residues, 21 were observed in *Taq* pol I. Both Ile^{709} in *E. coli* pol I and the corresponding isoleucine in *Taq* pol I contribute to discrimination of the sugar moiety of the incoming nucleotide. Lastly, Glu^{710} in *E. coli* and the corresponding Glu^{615} in *Taq* pol I are substitutable only by Asp. These results suggest that 1) the structural requirements for nearly wild-type motif A function *in vivo* are similar and 2) that the observed tolerance for substitutions within motif A is intrinsic and is evolutionarily conserved. We infer that motif A mutability may be adaptive and may promote survival by permitting toleration of a mutational burden at the polymerase active site without major loss of ability to function in replication.

The main difference in mutability between *E. coli* and *Taq* pol I involves a tyrosine residue. In the case of *Taq* pol I, Tyr^{611} was replaced only by the planar-ringed amino acids, Phe, His, and Trp, whereas substitutions at the corresponding Tyr^{706} of *E. coli* pol I were not similarly restricted. In the closed ternary complex of *Taq* pol I (15), the side chain of Tyr^{611} projects into a large hydrophobic pocket. We surmise that the side chains of Phe, His, or Trp may perform a space-filling function in a manner comparable with that of Tyr^{611}, thus permitting replacement; such a function might be less important in *E. coli* pol I, which tolerates other replacements. Another difference in mutability is that the number of positively charged residues, especially at Ser^{612}, Ile^{614}, and Arg^{617}, is greater among the *Taq* than the *E. coli* mutants. These residues may facilitate proper folding of *Taq* pol I while also maintaining polymerase activity at elevated temperatures (17). The optimum growth temperature for *T. aquaticus* is far higher than for *E. coli*; hence, the structure of proteins from *T. aquaticus* would presumably be restricted to ensure thermostability. Such structural constraints might be reflected in the restricted mutability of Tyr^{611} or the greater prevalence of positively charged replacements.

In conclusion, we reiterate that all DNA polymerases thus far examined appear to share a common overall architecture with superimposable catalytic palm subdomains and a common polymerase mechanism. In view of this conservation of structure and mechanism, we speculate that high mutability of motif A has also been retained throughout evolution so as to promote tolerance of a mutational burden at the polymerase active site with minimal loss of replicative capacity under conditions of changing environmental stresses.

Acknowledgments—We thank Dr. Elinor Adman for helpful discussions and Dr. Ann Blank for critical reading of the manuscript.

REFERENCES

1. Kornberg, A., and Baker, T. (1992) *DNA Replication*, 2nd Ed., W. H. Freeman and Co., New York
2. Delarue, M., Poch, O., Tordo, N., Moras, D., and Argos, P. (1990) *Protein Eng.* **3,** 461–467
3. Blanco, L., Bernad, A., Blasco, M. A., and Salas, M. (1991) *Gene* (*Amst.*) **100,** 27–38
4. Friedberg, E. C., Feaver, W. J., and Gerlach, V. L. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 5681–5683
5. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., and Steitz, T. A. (1985) *Nature* **313,** 762–766
6. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., and Steitz, T. A. (1992) *Science* **256,** 1783–1790
7. Sawaya, M. R., Pelletier, H., Kumar, A., Wilson, S. H., and Kraut, T. (1994) *Science* **264,** 1930–1935
8. Wang, J., Sattar, A. K. M. A., Wang, C. C., Karam, J. D., Konigsberg, W. H., and Steitz, T. A. (1997) *Cell* **89,** 1087–1089
9. Steitz, T. A. (1999) *J. Biol. Chem.* **274,** 17395–17398
10. Kunkel, T. A., and Wilson, S. H. (1998) *Nat. Struct. Biol.* **5,** 95–99
11. Kunkel, T. A., and Bebenek, K. (2000) *Annu. Rev. Biochem.* **69,** 497–529
12. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H., and Kraut, J. (1994) *Science* **264,** 1891–1903
13. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) *Nature* **391,** 251–258
14. Huang, H., Chopra, R., Verdine, G. L., and Harrison, S. C. (1998) *Science* **282,** 1669–1675
15. Li, Y., Korolev, S., and Waksman, G. (1998) *EMBO J.* **17,** 7514–7525
16. Astatke, M., Grindley, N. D. F., and Joyce, C. M. (1998) *J. Mol. Biol.* **278,** 147–165
17. Patel, P. H., and Loeb, L. A. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 5095–5100
18. Takeshita, S., Sato, M., Toba, M., Masahashi, W., Hashimoto-Gotoh, T. (1987) *Gene* (*Amst.*) **61,** 63–74

19. Joyce, C. M., Kelley, W. S., and Grindley, N. D. F. (1982) *J. Biol. Chem.* **257,** 1958–1964
20. Sweasy, J. B., and Loeb, L. A. (1992) *J. Biol. Chem.* **267,** 1407–1410
21. Witkin, E. M., and Boegner-Maniscalco, V. (1992) *J. Bacteriol.* **174,** 4166–4168
22. Spanos, A., Sedgwick, S. G., Yarranton, G. T., Hübscher, U., and Banks, G. R. (1981) *Nucleic Acids Res.* **9,** 1825–1839
23. Battula, N., and Loeb, L. A. (1974) *J. Biol. Chem.* **249,** 4086–4093
24. 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
25. Bradford, M. M. (1976) *Anal. Biochem.* **72,** 248–254
26. Boosalis, M. S., Petruska, J., and Goodman, M. F. (1987) *J. Biol. Chem.* **262,** 14689–14696
27. Patel, P. H., and Loeb, L. A. (2000) *J. Biol. Chem.* **275,** 40266–40272
28. Astatke, M., Ng, K., Grindley, N. D. F., and Joyce, C. M. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 3402–3407
29. Brautigam, C. A., and Steitz, T. A. (1998) *Curr. Opin. Struct. Biol.* **8,** 54–63
30. Gibbons, R. J., and Kapsimalis, B. (1967) *J. Bacteriol.* **93,** 510–512
31. Mao, E. F., Lane, L., Lee, J., and Miller, J. H. (1997) *J. Bacteriol.* **179,** 417–422
32. Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P. H., and Goddelle, B. (1997) *Nature* **387,** 700–702
33. Oliver, A., Cantón, R., Campo, P., Baquero, F., and Blázquez, J. (2000) *Science* **288,** 1251–1254
34. Drake, J. W. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 7160–7164
35. Polesky, A. H., Dahlberg, M. E., Benkovic, S. J., Grindley, N. D. F., and Joyce, C. M. (1992) *J. Biol. Chem.* **267,** 8417–8428