Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 9670–9675, September 1996
Genetics

Random mutagenesis of *Thermus aquaticus* DNA polymerase I: Concordance of immutable sites *in vivo* with the crystal structure

MOTOSHI SUZUKI*, DALE BASKIN†, LEROY HOOD†, AND LAWRENCE A. LOEB*‡

*The Joseph Gottstein Memorial Cancer Research Laboratory, Department of Pathology, Box 357705, and †Department of Molecular Biotechnology, University of Washington, Seattle, WA 98195-7705

Contributed by Leroy Hood, June 17, 1996

ABSTRACT **Expression of *Thermus aquaticus* (*Taq*) DNA polymerase I (pol I) in *Escherichia coli* complements the growth defect caused by a temperature-sensitive mutation in the host pol I. We replaced the nucleotide sequence encoding amino acids 659–671 of the O-helix of *Taq* DNA pol I, corresponding to the substrate binding site, with an oligonucleotide containing random nucleotides. Functional *Taq* pol I mutants were selected based on colony formation at the nonpermissive temperature. By using a library with 9% random substitutions at each of 39 positions, we identified 61 active *Taq* pol I mutants, each of which contained from one to four amino acid substitutions. Some amino acids, such as alanine-661 and threonine-664, were tolerant of several or even many diverse replacements. In contrast, no replacements or only conservative replacements were identified at arginine-659, lysine-663, and tyrosine-671. By using a library with totally random nucleotides at five different codons (arginine-659, arginine-660, lysine-663, phenylalanine-667, and glycine-668), we confirmed that arginine-659 and lysine-663 were immutable, and observed that only tyrosine substituted for phenylalanine-667. The two immutable residues and the two residues that tolerate only highly conservative replacements lie on the side of O-helix facing the incoming deoxynucleoside triphosphate, as determined by x-ray analysis. Thus, we offer a new approach to assess concordance of the active conformation of an enzyme, as interpreted from the crystal structure, with the active conformation inferred from *in vivo* function.**

Physical and chemical methods have been paramount in delineating the amino acid residues that constitute the active site of an enzyme. X-ray crystallography has provided resolution of structure at the atomic level. In certain situations NMR has been used to confirm that the crystal structure reflects the conformation of the protein in solution. We show here that the selection of active enzymes from molecular libraries containing random nucleotide substitutions may provide a new method to identify the essentiality of amino acid residues within the active site of enzymes and to lend support, based on *in vivo* function, to a particular interpretation of how crystal structure translates into an active conformation.

The crystal structures of diverse polymerases are remarkably similar (1). *Escherichia coli* pol I Klenow fragment, *Thermus aquaticus* (*Taq*) DNA polymerase I (pol I), rat DNA polymerase β, HIV reverse transcriptase, and T7 RNA polymerase have been crystallized and each contains a major structure that characterizes the polymerase domain. It consists of a large cleft, surrounded by finger, palm, and thumb subdomains (2–10). This similarity in structure and function permits a limited level of interchangeability between these enzymes *in vivo*. We have demonstrated that rat DNA polymerase β and HIV reverse transcriptase can each complement the temperature-sensitive phenotype of *E. coli recA718 polA12*, a temperature-sensitive mutant of pol I (11–16). Another example of the interchangeability between DNA polymerases is the substitution of rat DNA polymerase β for DNA polymerase δ in DNA alkylation repair in yeast (17). Such functional complementation provides positive genetic selection that can be used to identify active mutants of DNA polymerases from various sources and to study the essentiality of individual amino acids.

In this study, we first established that a plasmid expressing wild-type *Taq* pol I can complement the growth defect of *E. coli recA718 polA12*, a temperature-sensitive mutant of pol I. We then substituted random nucleotide sequences into a part of the *Taq* pol I gene that encodes the O-helix. By using genetic complementation, we identified new mutants in *Taq* pol I. A comparison of the amino acid replacements in the active *Taq* pol I mutants with the crystal structures of *E. coli* pol I and *Taq* pol I demonstrates that the immutable or nearly immutable residues lie exclusively on the face of the O-helix that interacts with the nucleoside triphosphate substrate.

MATERIALS AND METHODS

Strains, Media, and Oligonucleotides. The *recA718 polA12* strain for complementation refers to the SC18-12 *E. coli* B/r strain, which has the genotype *recA718 polA12 uvrA155 trpE65 lon-11 sulA1* (18). Strains DH5α [*deo*R, *end*A1, *gyr*A96, *phoA*, *hsd*R17($r_k^-mk^+$), *rec*A1, *rel*A1, *sup*E44, *thi*-1, Δ(*lac*ZYA-*arg*FU169), Φ80dlacZM15, F^-, λ^-] and JS295 (*uvrA155 trpE65 lon-11 sulA1*) strains were used for cloning and plasmid constructions. Difco nutrient agar (NA, 23 g/liter) containing NaCl (5 g/liter), chloramphenicol (30 μg/ml), tetracycline (12.5 μg/ml), and isopropyl β-D-thiogalactoside (IPTG, 1 mM) was used for studies on genetic complementation. Nutrient broth (NB, 8 g/liter) containing NaCl (4 g/liter) and tetracycline (12.5 μg/ml) was used for culturing the *recA718 polA12* strain. DNA oligomers were synthesized and purified by Operon Technologies (Alameda, CA), except where otherwise stated. The plasmid pHSG576 is a low-copy-number plasmid containing a pol I-independent pSC101 replication origin and a chloramphenicol-resistance gene (19). The plasmid pFC85, which carries an *Hin*dIII-*Asp*718 fragment of *Taq* Pol I in pBS(+) (20), was provided by the American Type Culture Collection.

Plasmids Carrying Wild-Type *Taq* Pol I. The *Taq* pol I gene was obtained from the bacterial chromosome by cloning in pKK223-3 (Pharmacia). A 3.2-kb fragment containing the *Taq* pol I gene, including the 5′–3′ exonuclease domain and the tac promoter region, was further transferred into the *Sal*I site of pHSG576 (pTacTaq). We sequenced the gene to confirm that

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: *Taq, Thermus aquaticus*; pol I, DNA polymerase I.

*To whom reprint requests should be addressed. e-mail: laloeb@u.washington.edu.

our wild-type construct had no mutations except for lack of the N-terminal three amino acids.

Genetic Complementation. *E. coli recA718 polA12* cells were transformed with plasmids pHSG576 or pTacTaq by electroporation (Bio-Rad Genepulser, 2kV, 25 μFD, 400Ω) (13, 14). Thereafter, 1 ml of NB (8 g/liter) containing NaCl (4 g/liter) and 1 mM IPTG was added and the mixture was incubated for 1 h at 37°C. The transformed cells were plated on NA plates and grown at 30°C overnight. Single colonies were transferred to NB for growth to logarithmic phase at 30°C. Thereafter, $\approx$10 μl (10^4 cells) was introduced at the center of an agar plate, and the inoculation loop was gradually moved from the center to the periphery as the plate was rotated. Duplicate plates were incubated at 30°C or 37°C for 30 h. To determine complementation efficiency by *Taq* pol I and to isolate mutants, cultures of the *recA718 polA12* strain harboring either pHSG576 or *Taq* pol I were diluted with NB medium and plated ($\approx$500 colonies per plate). Duplicate plates were incubated at 30°C or 37°C, and visible colonies were counted after a 30-h incubation.

Nonfunctional Vector. To avoid contamination with incompletely cut vectors, we first constructed a vector containing a nonfunctional insert within the *Taq* pol I gene, and then replaced the nonfunctional insert with an oligonucleotide containing the random sequence. A *Sac*II site was produced using site-directed mutagenesis by changing 2070C to G using a synthetic oligomer, 5′-GGG TCC ACG GCC TCC CGC GGG ACG CCG AAC ATC CAG CTG (SacII-2) and the single-stranded plasmid pFC85 according to Kunkel (21). The *Bst*XI–*Nhe*I fragment that carries the *Sac*II site was substituted for the corresponding fragment in pTacTaq (pTacTaqSac). A *Sac*II–*Nhe*I fragment in pTacTaqSac was further replaced with the synthetic oligomer 5′-GGA CTG CAT ATG ACT G (DUM-U) hybridized with 5′-CTA GCA GTC ATA TGC AGT CCG C (DUM-D) to create the nonfunctional vector (22).

Random Oligonucleotides. Oligonucleotides containing 9% random sequence, in which each nucleotide in brackets was 91% wild-type nucleotide and 3% each of the other three nucleotides, were synthesized by Keystone Laboratories (Menlo Park, CA): O+9 RANDOM is 5′-CGG GAG GCC GTG GAC CCC CTG ATG [CGC CGG GCG GCC AAG ACC ATC AAC TTC GGG GTC CTC TAC] GGC ATG TCG GCC CAC CG; O-0 RANDOM is 5′-TGG CTA GCT CCT GGG AGA GGC GGT GGG CCG ACA TGC C. The 17 nucleotide sequences at the 3′ ends of the two oligonucleotides are complementary. Equimolar amounts of these oligonucleotides (20 pmol) were mixed, hybridized, and extended by five cycles of PCR reaction (94°C for 30 sec, 57°C for 30 sec, and 72°C for 30 sec) in a 100-μl reaction mixture containing 10 mM Tris·HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin, 50 μM dNTPs, and 2.5 units of *Taq* pol I. This PCR product (10 μl) was further amplified for 25 cycles with 20 pmol of O(+)PRIMER (5′-TTC GGC GTC CCG CGG GAG GCC GTG GAC CCC CT) and 20 pmol of O(-)PRIMER (5′-GTA AGG GAT GGC TAG CTC CTG GGA) under the same conditions. The amplified product was purified by phenol/chloroform extraction followed by ethanol precipitation and digestion with the restriction enzymes, *Sac*II and *Nhe*I, at 37°C for 30 min in 50 mM Tris·HCl (pH 7.9), 50 mM NaCl, 10 mM $MgCl_2$, and 1 mM dithiothreitol. The restriction fragment containing the random sequence was purified by phenol/chloroform extraction, ethanol precipitation, and filtration using a Microcon 30 filter (Amicon). For the totally random library five oligonucleotides (80-mers), each having totally random sequence at one of the codons 659, 660, 663, 667, or 668, were combined in equal amounts and hybridized to O-0 RANDOM. After extension and digestion with endonucleases, the combined products were purified and processed as above.

Library Construction and Complementation. The vector containing the nonfunctional insert was digested with *Nhe*I and *Sac*II restriction endonucleases. The large DNA fragment was isolated by electrophoresis in a 0.8% agarose gel and purified by using GenCleanII (Bio101). This large fragment, lacking the nonfunctional insert, was ligated with an oligonucleotide containing randomized sequence by incubating overnight at 16°C with T4 DNA ligase. The ligation mixture was then used to transform DH5α by electroporation according to Bio-Rad. After electroporation, 1 ml of SOC (2% bactotryptone/0.5% yeast extract/10 mM NaCl/2.5 mM KCl/10 mM $MgCl_2$/10 mM $MgSO_4$/20 mM glucose) was added and incubation continued for 1 h at 37°C. An aliquot was plated on 2xYT (16 g/liter tryptone/10 g/liter yeast extract/5 g/liter NaCl, pH 7.3) containing chloramphenicol (30 μg/ml) to determine the total number of transformants, and the remainder was inoculated into 500 ml of 2xYT containing chloramphenicol (30 μg/ml) and cultured at 37°C overnight. Plasmids (random library vector) were purified and used for transformation of *recA718 polA12* strain as described above and illustrated in Fig. 2. Complementation was verified by a second round of electroporation and colony formation at the nonpermissive temperature. Cell-free extracts were prepared from selected colonies obtained at the restrictive temperature and assayed to confirm that they contained a temperature-resistant DNA polymerase activity (20).

DNA Sequencing. The plasmid-borne 92-bp insert containing randomized sequences was sequenced using the *Taq* Dye-Terminator Cycle Sequencing kit, with an Applied Biosystems 373A DNA sequencer or Perkin–Elmer/Applied Biosystems Prizm 377. Sequences were obtained in both orientations using the associated sequence analysis software (Applied Biosystems/Perkin–Elmer). The primers used were O(−)2PRIM, 5′-AGT AGC GCT CAA TGA AGG CCT; O(+)2PRIM, 5′-TGG CTA TTG GTG GCC CTG GAC.

RESULTS

Complementation of *E. coli recA718 polA12*. The *E. coli* strain *recA718 polA12* contains a temperature-sensitive pol I (18). This mutant is unable to grow at 37°C in rich media at low cell density (11, 13, 14, 18). Previous studies from our laboratory demonstrated that the temperature-sensitive phenotype can be complemented by transformation with plasmids that express DNA polymerase β (13) or HIV reverse transcriptase (11).

Complementation by a plasmid that encodes the wild-type *Taq* pol I is shown in Fig. 1. The spiral of colonies was created by inoculating the bacteria in the center of the plate and rotating the plate while gradually moving the inoculating loop to the periphery. At 37°C *E. coli recA718 polA12*, transformed with the parent plasmid pHSG576, forms colonies only at high cell density in the center of the plate. This temperature-sensitive phenotype is complemented by transformation with the plasmid that expresses *Taq* pol I, as indicated by the spiral display of colonies.

The selectivity of this complementation experiment was verified by mixing experiments. *E. coli recA718 polA12* were transformed with a mixture of pTacTaq and the parent plasmid, pHSG576, at ratios of 1:10 or 1:50. Colonies that formed at the nonpermissive temperature were scored for the presence of the pTacTaq plasmid. All 6 colonies from the 1:50 mixture, and all 27 from the 1:10 mixture, contained the pTacTaq plasmid.

Selection for Mutants in 9% Random Library. To evaluate the involvement of different amino acid residues in catalysis by *Taq* pol I, we substituted random sequences for nucleotides encoding a portion of the substrate binding site of *Taq* pol I (O-helix, amino acids Arg-659 through Tyr-671). The substituted stretch was 39 nucleotides long with 9% randomization,

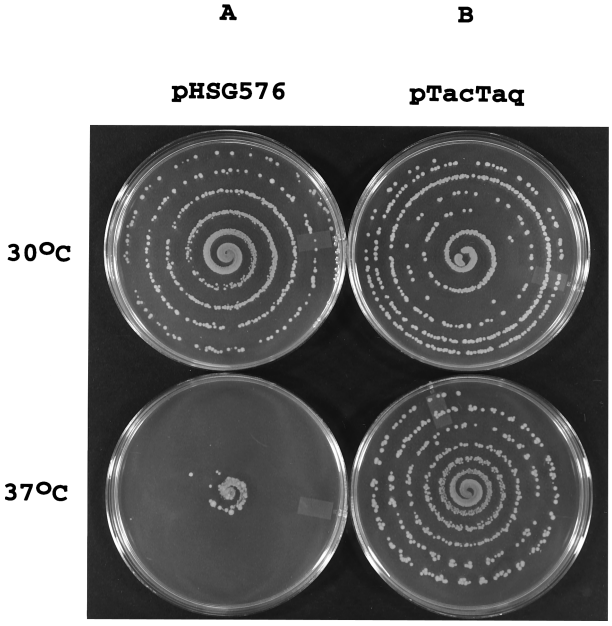

FIG. 1. Functional complementation of *E. coli recA718 polA12* by *Taq* pol I. *E. coli recA718 polA12* was transformed with either pHSG576, a low-copy-number plasmid containing a pol I-independent pSC101 replication origin and a chloramphenicol-resistance gene (19) or with the same plasmid containing the *Taq* pol I gene. Transformed cells were grown to logarithmic phase in NB containing tetracycline (12.5 μg/ml), chloramphenicol (30 μg/ml), and IPTG (1 mM) and then tested for complementation by incubation at 30°C or 37°C for 30 h as described. (*A*) Growth of *E. coli recA718 polA12* with the parent plasmid, pHSG576, at 30°C and 37°C. (*B*) Growth of the *recA718 polA12* strain with pTacTaq at 30°C and 37°C.

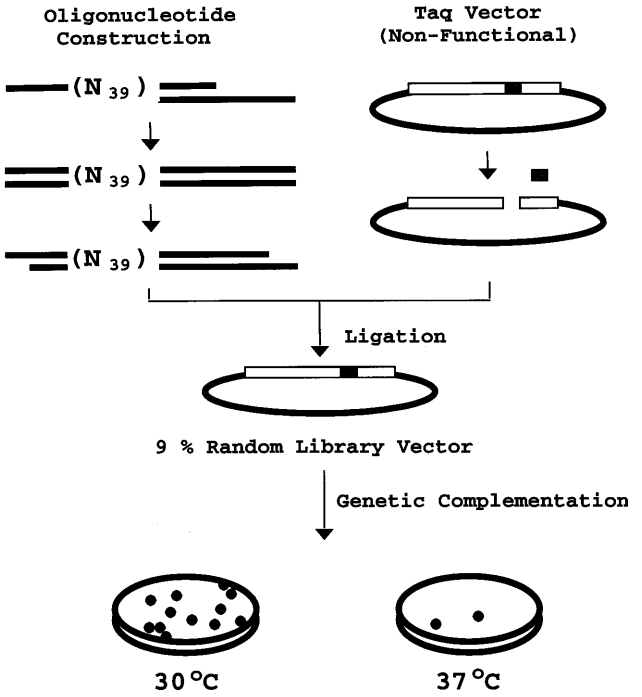

FIG. 2. Selection strategy. An oligomer containing 39 nucleotides of 9% randomized sequence was annealed with a partially complementary oligomer, extended, and amplified by PCR. This fragment was digested with restriction enzymes and substituted in place of the corresponding restriction fragment in the nonfunctional vector. This random library vector was transformed into *E. coli recA718 polA12* and active mutants of *Taq* pol I were selected at 37°C.

i.e., at each position the proportion of the wild-type residue was 91% and the other 3 nucleotides were present in equal amounts. The construction of this library is diagrammed in Fig. 2 and detailed in *Materials and Methods*.

We obtained a library of 50,000 independent mutants. The number of colonies obtained at 37°C was 11.8% of that obtained at 30°C. We sequenced the entire random nucleotide-containing insert from a total of 234 plasmids obtained at 37°C (positively selected), 16 plasmids obtained at 30°C (nonselected), and 29 plasmids obtained at 30°C, which failed to grow at 37°C (negatively selected). All substitutions were in the randomized nucleotides except for 12 clones. Eight of the clones contained silent nucleotide substitutions outside the target sequence, presumably introduced during PCR amplification of the insert. The four clones with amino acid changes outside the target stretch were not taken into account in the data below.

Among the 230 positive plasmids, 168 contained silent mutations in one or more codons. At the amino acid level, 106 encoded the wild-type residue and 124 encoded substitutions, in accord with the expected distribution in the plasmid population (M.S., F. C. Christians, B. Kim, A. Skandalis, M. E. Black, and L.A.L., unpublished work). The number of substituted residues also approximated the theoretical average (Table 1). The number of nucleotide substitutions per positively selected plasmid was 2.0, less than that in plasmids obtained in the absence of selection, i.e., 4.0 (Table 1).

Of the 124 plasmids with amino acid changes, 40 were unique mutants, i.e., they were obtained just once. The remaining 84 plasmids represented 21 different mutants. At least 79% of those encoding the same amino acid substitutions were independently derived since they contained different silent mutations in other codons. In total, we obtained 61 different amino acid sequences that complemented the temperature-sensitive phenotype of the *recA718 polA12* host.

Amino Acid Substitutions. The distribution of single amino acid substitutions among the active mutants was not random (Fig. 3*A*). For example, numerous diverse substitutions were observed at Ala-661 and Thr-664. In contrast, no substitutions were detected at five positions (Arg-659, Arg-660, Lys-663, Phe-667, and Gly-668). This uneven distribution of replacements is unlikely to be the result of a bias in the nucleotide composition of the random insert since sequencing of both the nonselected and negatively selected plasmids revealed multiple nucleotide substitutions at each of the targeted positions, and because silent mutations were detected at each of these positions in the selected clones.

A nonrandom distribution of substitutions was also observed among active mutants containing multiple substitutions (Fig. 3*B*). Again, Ala-661 and Thr-664 were replaced with a variety of residues. However, no amino acid substitutions were observed in place of Arg-659, Lys-663, and Gly-668, even though different silent nucleotide substitutions were found at each of these positions. A comparison of Fig. 3 *A* and *B* shows that substitutions at Arg-660 and Phe-667 occur only in the presence of substitutions at other positions.

Table 1. Average number of substitutions

Population	No. of clones	Average no. of nucleotide substitutions	Average no. of amino acid substitutions
Entire library*	—	3.5	2.4
Nonselected	16	3.8	2.7
Negatively selected	29	4.0	2.8
Positively selected	230	2.0	0.8

*Numbers are calculated by methods described elsewhere (M.S., F. C. Christians, B. Kim, A. Skandalis, M. E. Black, and L.A.L., unpublished work), assuming that the planned randomness (9%) is introduced.

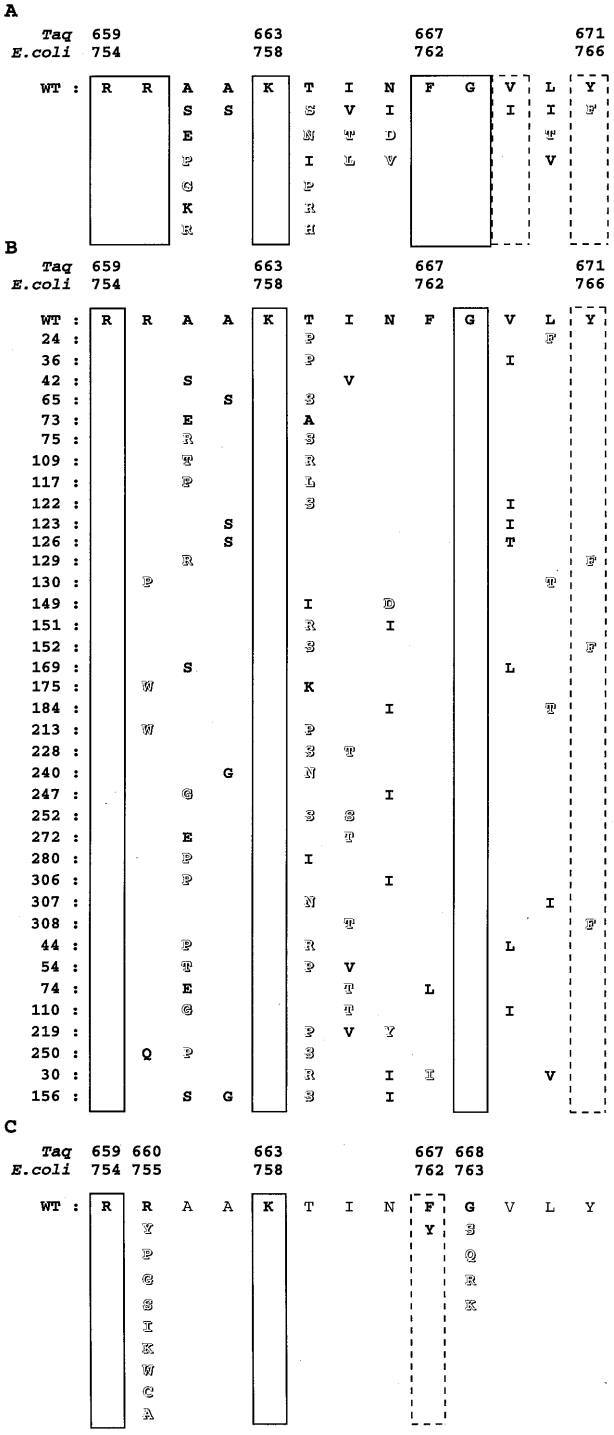

FIG. 3. Compilation of amino acid substitutions. We introduced either 9% random sequence in the codons for amino acids from Arg-659 to Tyr-671 (*A* and *B*) or totally random sequences in each of five codons Arg-659, Arg-660, Lys-663, Phe-667, or Gly-668 (*C*) and selected mutants at the nonpermissive temperature. Amino acid substitutions from the positively selected clones are shown. Solid boxes indicate the amino acid residues for which no substitutions were detected. Dashed boxes mark the amino acids where only conservative substitutions were found. The amino acid positions of *Taq* pol I and corresponding positions of *E. coli* pol I are indicated at the top. WT represents the wild-type sequence and randomized amino acids are written in boldface type. The amino acids that have not been found in the DNA polymerase I family (23) are outlined. (*A*) Single mutations selected from the 9% library are listed under the wild-type amino acids. (*B*) The sequence of each multiply substituted mutant selected from the 9% library is listed. (*C*) Mutations selected from the totally random library are listed.

Immutable Amino Acids. The partially substituted library (9%) does not provide a vigorous test of the immutability of specific codons. Only 0.07% of sequences at each codon would be expected to contain nucleotide substitutions at all three positions. To further probe the mutability of specific amino acid residues we constructed a second library that contained totally random substitutions at a limited number of designated codons. In this library, nucleotides encoding each of the five amino acids, Arg-659, Arg-660, Lys-663, Phe-667, and Gly-668 were randomized, i.e., the positions that did not yield single substitutions in the 9% random library (Fig. 3*A*). We screened ≈1300 transformants, which is 4 times more than the number required for each possible substitution at each of the target codons. We obtained 113 colonies at the nonpermissive temperature, 84 of which contained codons that encoded the wild-type amino acid sequence. Most of the amino acid substitutions occurred in place of Arg-660 or Gly-668.

Again, Arg-659 and Lys-663 were completely conserved. We scored 16 and 5 silent mutations at these codons, respectively. We expected 21 and 4.2 silent mutations, respectively, assuming that the 5 randomized oligomers that comprised the library were mixed in equimolar proportions. These numbers show that the oligomers were roughly equally represented in the library and that we sampled sufficient mutants to conclude that Arg-659 and Lys-663 are immutable in these genetic complementation experiments ($P < 0.05$ for Met and Trp, $P < 0.01$ for all other substitutions). Only tyrosine substituted for Phe at position 667 (Fig. 3*C*), and we scored six silent mutations for this codon.

DISCUSSION

DNA polymerases and reverse transcriptases take direction from a second molecule, a polynucleotide template (24, 25). Kinetic studies indicate that these enzymes change their conformation during catalysis (24, 26, 27), and that the binary complex of Klenow fragment and a dNTP is not in a catalytically active conformation (24). Crystal structures of the ternary complex are not available yet, except for DNA polymerase β (7). In this context, the interpretation of crystal structures of binary complexes, especially Klenow fragment and substrates, have been discussed very carefully (1, 28). Functional studies involving site-directed mutagenesis have been carried out to gain information about the conformation of polymerases in the presence of a template. However, this approach is limited in that, in practice, one can only test a small subset of single amino acid substitutions at different positions. Furthermore, most of the mutations in DNA polymerases that have been tested so far using site-directed mutagenesis result in a diminution in activity.

Here we demonstrate that *Taq* pol I can complement the temperature-sensitive phenotype of *E. coli*, *recA718 polA12.* Although the temperature stability of *Taq* pol I and *E. coli* pol I are different, *Taq* pol I is classified in the same family as *E. coli* pol I based on sequence homology (23, 29, 30). Both polymerases have a 5′–3′ exonuclease domain and are 38% identical in amino acid sequence (20, 31). Furthermore, the crystal structures of the polymerase domains in both enzymes are nearly identical (3, 9, 10). The lack of a 3′–5′ exonuclease activity in *Taq* pol I does not preclude complementation; both DNA polymerase β (13, 14) and HIV reverse transcriptase lack this activity (11), and can also complement the temperature-sensitive strain of pol I.

We further exploited this system to recover active mutants of *Taq* pol I from libraries of plasmids containing random nucleotide substitutions in the O-helix. This region of *E. coli* pol I has been demonstrated to coordinate with the nucleoside triphosphate substrates as determined by NMR (32), and by crosslinking (33, 34). A peptide fragment of 50 residues

corresponding to the O-helix binds deoxynucleoside triphosphates (35).

Random sequence mutagenesis coupled with positive genetic selection has allowed us to examine multiple amino acid substitutions and to determine which of these yield functionally active enzymes. It provides a stringent test of the functionality of amino acid side chains within the active site of an enzyme. Based on the results using both the 9% and totally random libraries, we can subdivide the amino acid residues that we analyzed into three groups: those that are immutable, those that tolerate only a limited number of conservative substitutions, and those that tolerate many substitutions with varying sizes and charges.

No substitutions were observed for Arg-659 and Lys-663 among the active mutants selected from either the 9% or the totally random libraries. In the studies with the totally random library, there were more than 200 chances for substitution at each of these positions. These immutable amino acids may be required for maintaining the structure of the protein or for direct interactions with one of the substrates. Protein-folding studies indicated that charged residues are not easily accommodated in the interior of proteins, and that surface amino acids are more likely to tolerate substitutions with differing side chains (refs. 36 and 37, and references therein). Thus, our finding that these charged residues, Arg-659 and Lys-663, are immutable suggests that the two amino acids are essential for catalytic activity by directly contacting acidic molecules. Arg-659 and Lys-663 in *Taq* pol I correspond to Arg-754 and Lys-758 in *E. coli* pol I. Based on the crystal structures of the large fragment of *E. coli* pol I complexed with dNTP or with pyrophosphate, it is hypothesized that these residues coordinate with the β and γ phosphate residues on the incoming dNTP (1, 28). Site-directed mutagenesis and kinetic studies indicate that these residues are important in catalysis. Substitution of Ala or Arg at Lys-758 in *E. coli* pol I reduced k_{cat} 300- to 1000-fold with poly(dA)·$(dT)_{15}$ as template-primer (38, 39). Similarly, in T7 RNA polymerase, replacement of the corresponding Lys-631 with Met reduced k_{cat} (40). Thus, our data on the immutability of Arg-659 and Lys-663 in *Taq* pol I are in accord with the *in vitro* results obtained with other polymerases.

Phe-667 and Tyr-671 were tolerant only of conservative substitutions. Phe-667 was mutated exclusively to Tyr in the totally random library. Tabor and Richardson (41) demonstrated that Phe-667 is critical for substrate discrimination. They found that substitution by Tyr yielded a mutant *Taq* pol I that exhibits a 1000-fold enhanced incorporation of dideoxynucleoside triphosphates relative to dNTPs, a property that will facilitate DNA sequencing by the chain termination method (41). Phe-667 was also substituted by Leu or Ile using the 9% library but only in the presence of other amino acid changes. In separate studies using site-directed mutagenesis, we demonstrated that the single substitution Phe-667 to Leu did not yield an active polymerase as defined by complementation. Thus, Phe-667 falls into a subset of residues that are subject to conservative single amino acid substitutions. Although we did not analyze Tyr-671 mutants in the totally random library, Tyr-671 might be classified in this group since it was replaced exclusively by Phe, which was selected as many as 12 times, in the 9% library. In *E. coli* pol I, substitution of Phe for the corresponding Tyr-766 does not alter the K_m or K_{cat} for dNTPs, whereas substitution by Ser yields a 3- to 4-fold decrease in activity and a 2- to 3-fold increase in K_m for dNTPs (42). Similarly, the corresponding Y639F mutation in T7 RNA polymerase is as active as wild type but shows modified rNTP/dNTP discrimination (43), whereas the Y639S substitution results in a ≈2000-fold reduction in activity (40, 44).

Nine of the other 13 amino acid residues that we studied could be substituted by a variety of amino acids. Amino acids such as Ala-661 and Thr-664 accepted a variety of replacements, and thus may not be essential for activity. Substitution at six other positions in the C-terminal region of the O-helix (Ile-665 through Tyr-671) was predominantly by hydrophobic residues. Some of these amino acids could be involved in packing in the interior of the protein, and some of the hydrophobicity of this stretch may be required for complementation of the pol I^{ts} growth defect in *E. coli*, and in particular, for binding deoxynucleoside triphosphates (27), or for interacting with the template-primer (35, 39, 45). From the results of available studies, we suggest that amino acids Ile-665 through Tyr-671 constitute a portion of the substrate binding pocket, but do not necessarily interact with the incoming deoxynucleoside triphosphate directly.

A comparison between the number of amino acid substitutions obtained from each library and the orientation of individual amino acids in the O-helix is presented in Fig. 4. The two immutable residues and the two residues that tolerate only highly conservative replacements are found on the side of the O-helix facing the incoming deoxynucleotide triphosphate (Fig. 4). Because the mutagenesis studies test functionality *in*

FIG. 4. Comparison of mutability/essentiality and orientation of residues in the O-helix. The number of permissible amino acid replacements at each position in the target site is plotted for both libraries. (*A*) Results from the 9% random library. Both single and multiple mutants (compiled in Fig. 3 *A* and *B*) are included. (*B*) The results from the totally random library. (*C*). Representation of the O-helix in *Taq* pol I.

vivo, they provide evidence that the O-helix interacts with incoming substrates in catalysis, which has been suggested by other *in vitro* studies, including crystallography. This approach permits assessment of the essentiality of individual amino acids within the active site. It also permits comparison of the active conformation derived from the crystal structure with the active conformation deduced from functionality in living cells.

Finally, it should be noted that 34 of the 51 amino acid substitutions we observed (Fig. 3) have not been found among naturally occurring members of the DNA polymerase I family studied to date. Thus, our work provides new active *Taq* pol I mutants, which may possess useful catalytic properties.

We are grateful to Ann Blank and Adonis Skandalis for helpful discussions. This study was supported by National Institutes of Health Grant OIG-R35-CA-39903 to L.A.L. and National Science Foundation Science and Technology Center for Molecular Biotechnology Grant BIR9214821AM03 to L.H.

1. Joyce, C. M. & Steitz, T. A. (1994) *Annu. Rev. Biochem.* **63,** 777–822.
2. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1992) *Science* **256,** 1783–1790.
3. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. (1985) *Nature (London)* **313,** 762–766.
4. Sousa, R., Chung, Y. J., Rose, J. P. & Wang, B.-C. (1993) *Nature (London)* **364,** 593–599.
5. Sawaya, M. R., Pelletier, H., Kumar, A., Wilson, S. H. & Kraut, J. (1994) *Science* **264,** 1930–1935.
6. Davies, J. F., II, Almassy, R. J., Hostomska, Z., Ferre, R. A. & Hostomsky, Z. (1994) *Cell* **76,** 1123–1133.
7. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. (1994) *Science* **264,** 1891–1903.
8. Arnold, E., Jacobo-Molina, A., Nanni, R. G., Williams, R. L., Lu, X., Ding, J., Clark, A. D., Jr., Zhang, A., Ferris, A. L., Clark, P., Hizi, A. & Hughes, S. H. (1992) *Nature (London)* **357,** 85–89.
9. Korolev, S., Nayal, M., Barnes, W. M., Di Cera, E. & Waksman, G. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 9264–9268.
10. Kim, Y., Eom, S. H., Wang, J., Lee, D.-S., Suh, S. W. & Steitz, T. A. (1995) *Nature (London)* **376,** 612–616.
11. Kim, B. & Loeb, L. A. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 684–688.
12. Kim, B. & Loeb, L. A. (1995) *J. Virol.* **69,** 6563–6566.
13. Sweasy, J. B. & Loeb, L. A. (1992) *J. Biol. Chem.* **267,** 1407–1410.
14. Sweasy, J. B. & Loeb, L. A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 4626–4630.
15. Sweasy, J. B., Chen, M. & Loeb, L. A. (1995) *J. Bacteriol.* **177,** 2923–2925.
16. Kim, B., Hathaway, T. R. & Loeb, L. A. (1996) *J. Biol. Chem.* **271,** 4872–4878.
17. Blank, A., Kim, B. & Loeb, L. A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 9047–9051.
18. Witkin, E. M. & Roegner-Maniscalo, V. (1992) *J. Bacteriol.* **174,** 4166–4168.
19. Takeshita, S., Sato, M., Toba, M., Masahashi, W. & Hashimoto-Gotoh, T. (1987) *Gene* **61,** 63–74.
20. Lawyer, F. C., Stoffel, S., Saiki, R. K., Myambo, K., Drummond, R. & Gelfand, D. H. (1989) *J. Biol. Chem.* **264,** 6427–6437.
21. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 488–492.
22. Dube, D. K., Parker, J. D., French, D. C., Cahill, D. S., Dube, S., Horwitz, M. S. Z., Munir, K. M. & Loeb, L. A. (1991) *Biochemistry* **30,** 11760–11767.
23. Braithwaite, D. K. & Ito, J. (1993) *Nucleic Acids Res.* **21,** 787–802.
24. Bryant, F. R., Johnson, K. A. & Benkovic, S. J. (1983) *Biochemistry* **22,** 3537–3546.
25. Patel, P. H., Jacobo-Molina, A., Ding, J., Tantillo, C., Clark, A. D., Jr., Raag, R., Nanni, R. G., Hughes, S. H. & Arnold, E. (1995) *Biochemistry* **34,** 5351–5363.
26. Mizrahi, V., Henrie, R. N., Marlier, J. F., Johnson, K. A. & Benkovic, S. J. (1985) *Biochemistry* **24,** 4010–4018.
27. Kornberg, A. & Baker, T. A. (1992) *DNA Replication* (Freeman, New York).
28. Beese, L. S., Friedman, J. M. & Steitz, T. A. (1993) *Biochemistry* **32,** 14095–14101.
29. Blanco, L., Bernad, A., Blasco, M. A. & Salas, M. (1991) *Gene* **100,** 27–38.
30. Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. (1990) *Protein Eng.* **3,** 461–467.
31. Joyce, C. M., Kelley, W. S. & Grindley, N. D. F. (1982) *J. Biol. Chem.* **257,** 1958–1964.
32. Ferrin, L. J. & Mildvan, A. S. (1985) *Biochemistry* **24,** 6904–6913.
33. Basu, A. & Modak, M. J. (1987) *Biochemistry* **26,** 1704–1709.
34. Rush, J. & Konigsberg, W. H. (1990) *J. Biol. Chem.* **265,** 4821–4827.
35. Mullen, G. P., Shenbagamurthi, P. & Mildvan, A. S. (1989) *J. Biol. Chem.* **264,** 19637–19647.
36. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A. & Sauer, R. T. (1990) *Science* **247,** 1306–1310.
37. Rennell, D., Bouvier, S. E., Hardy, L. W. & Poteete, A. R. (1991) *J. Mol. Biol.* **222,** 67–88.
38. Pandey, V. N., Kaushik, N. & Modak, M. J. (1994) *J. Biol. Chem.* **269,** 13259–13265.
39. Astatke, M., Grindley, N. D. F. & Joyce, C. M. (1995) *J. Biol. Chem.* **270,** 1945–1954.
40. Osumi-Davis, P. A., de-Aguilera, M. C., Woody, R. W. & Woody, A.-Y. M. (1992) *J. Mol. Biol.* **226,** 37–45.
41. Tabor, S. & Richardson, C. C. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 6339–6343.
42. Carroll, S. S., Cowart, M. & Benkovic, S. J. (1991) *Biochemistry* **30,** 804–813.
43. Sousa, R. & Padilla, R. (1995) *EMBO J.* **14,** 4609–4621.
44. Bonner, G., Patra, D., Lafer, E. M. & Sousa, R. (1992) *EMBO J.* **11,** 3767–3775.
45. Pandey, V. N., Kaushik, N. & Modak, M. J. (1994) *J. Biol. Chem.* **269,** 21828–21834.