

DNA polymerase active site is highly mutable: Evolutionary consequences

Premal H. Patel and Lawrence A. Loeb*

The Joseph Gottstein Memorial Cancer Laboratory, Department of Pathology, University of Washington School of Medicine, Seattle, WA 98195-7705

Communicated by Maynard V. Olson, University of Washington, Seattle, WA, March 9, 2000 (received for review November 4, 1999)

DNA polymerases contain active sites that are structurally superimposable and highly conserved in sequence. To assess the significance of this preservation and to determine the mutational burden that active sites can tolerate, we randomly mutated a stretch of 13 amino acids within the polymerase catalytic site (motif A) of *Thermus aquaticus* DNA polymerase I. After selection, by using genetic complementation, we obtained a library of approximately 8,000 active mutant DNA polymerases, of which 350 were sequenced and analyzed. This is the largest collection of physiologically active polymerase mutants. We find that all residues of motif A, except one (Asp-610), are mutable while preserving wild-type activity. A wide variety of amino acid substitutions were obtained at sites that are evolutionarily maintained, and conservative substitutions predominate at regions that stabilize tertiary structures. Several mutants exhibit unique properties, including DNA polymerase activity higher than the wild-type enzyme or the ability to incorporate ribonucleotide analogs. Bacteria dependent on these mutated polymerases for survival are fit to replicate repetitively. The high mutability of the polymerase active site *in vivo* and the ability to evolve altered enzymes may be required for survival in environments that demand increased mutagenesis. The inherent substitutability of the polymerase active site must be addressed relative to the constancy of nucleotide sequence found in nature.

DNA polymerases are responsible for the replication and maintenance of the genome, a role that is central to accurately transmitting genetic information from generation to generation. DNA polymerases have an active site architecture that specifically configures to and incorporates each of the four deoxynucleoside triphosphates, while taking direction from templates with diverse nucleotide sequences (1). In addition, the active site tends to exclude altered nucleotides produced during cellular metabolism. The overall folding pattern of polymerases resembles the human right hand and contains three distinct subdomains (palm, fingers, and thumb; ref. 2; Fig. 1). Whereas the structures of the fingers and thumb subdomains vary greatly between polymerases that differ in size and in cellular functions, the catalytic palm subdomains are all superimposable. A detailed study of polymerase active site within the conserved palm subdomain is crucial to understanding how polymerases function.

All DNA and RNA polymerases share two conserved regions: motif A and motif C. Both are located within the palm subdomain. Structural data show that amino acids within motif A are in position to interact with the incoming dNTP, coordinate with the two divalent metal cofactors, and interact with the fingers subdomain during the conformation change step after dNTP binding (3, 4). In addition, Asp-610 of motif A is in position to stabilize the transition state during the chemical catalysis step, which leads to phosphodiester bond formation (4). A limited number of mutation studies of the active site have shown that substitutions within motif A can lead to altered dNTP recognition and fidelity (5, 6). The primary amino acid sequences of various DNA polymerase active sites are exceptionally conserved, suggesting that motif A evolved slowly. Motif A retains the sequence DYSQIELR in polymerases from organisms separated by many million years of evolution, including *Thermus*

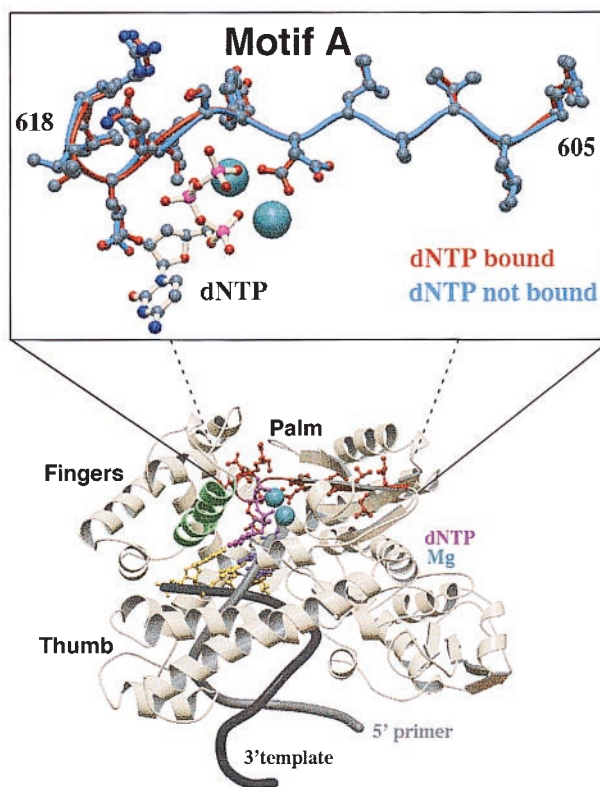


Fig. 1. Structure of *Taq* pol I bound with DNA and incoming dNTP. Evolutionarily conserved motif A (amino acids 605 to 617 highlighted in red) is located within the heart of the polymerase catalytic site. Residues of motif A interact with the incoming dNTP and amino acids in the finger motif during the conformational change step, subsequent to nucleotide binding. Motif A is superimposable in all polymerases with solved structures and begins at a hydrophobic antiparallel β -sheet that proceeds to an α -helix. The orientation of side chains within amino acids of motif A is nearly identical before (in blue) and subsequent to (in red) dNTP binding, with the exception of Asp-610, which rotates around the β carbon while coordinating with the Mg^{2+} -dNTP complex. Motif B (amino acids 656 to 670) is highlighted in green. Coordinate sets 3ktq (ternary complex, closed form) and 4ktq (binary complex) from ref. 4 were obtained from the Protein Data Bank. The coordinates were aligned by using the 1sq_expl command in the software o to generate the inset view.

aquaticus, *Chlamydia trachomatis*, and *Escherichia coli*. Structurally, motif A is superimposable with a mean deviation of 1 Å among mammalian pol α and prokaryotic pol I family DNA polymerases (7) and begins at an antiparallel β -strand containing predominantly hydrophobic residues and continues to an α -helix

Abbreviations: *Taq* pol I, *Thermus aquaticus* DNA polymerase I; WT, wild-type.

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

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.

(Fig. 1). Taken together, these results indicate that polymerases function by similar catalytic mechanisms and that the active site of polymerases may be immutable to ensure the survival of organisms.

To evaluate the degree of plasticity within the polymerase active site *in vivo*, we substituted random DNA sequences in *T. aquaticus* DNA polymerase I (*Taq* pol I) at the region encoding the 13 amino acids of motif A from Leu-605 to Arg-617, and selected active clones by genetic complementation of *E. coli* *recA718 polA12*. This *E. coli* strain, which contains a temperature-sensitive mutation in the *polA* gene, which encodes DNA polymerase I, forms colonies at 30°C, but not at 37°C. After selection, we find DNA polymerase catalytic site (motif A) is in fact highly mutable while preserving wild-type (WT) activity. The high mutability of the active site *in vivo* and the ability to evolve altered enzymes may be required for survival in environments that demand increased mutagenesis.

Methods

Plasmid Construction. The *Taq* pol I gene was cloned into low copy (one to three copies per cell) pHSG576 vector containing an *E. coli* pol I-independent origin of replication, SC101 (8, 9). A silent *Bis*WI site was created in *Taq* pol I by site-directed mutagenesis (C → A) at position 1758 (pTaq). A nonfunctional stuffer vector (pTaqDUM) was constructed by cloning two hybridized oligonucleotides into pTaq between *Bis*WI and *Sac*II sites; these two restriction sites flank the sequence encoding for motif A. A random library (pTaqLIB) was created by preparing a randomized oligonucleotide with a *Bis*WI site in which nucleotides encoding amino acids Leu-605 to Arg-617 contained 88% WT and 4% each of the other three nucleotides. This oligonucleotide was hybridized with an oligonucleotide primer containing *Sac*II site in equimolar proportions, and T7 DNA polymerase (exo⁻) was used to copy the template containing the randomized nucleotides and to make duplex DNA. The double-stranded DNA was digested with *Bis*WI and *Sac*II, purified, and inserted into pTaqDUM between *Bis*WI and *Sac*II restriction sites in place of the stuffer fragment. The reconstructed plasmids were transformed into DH5 α cells by electroporation, and the cells were incubated in 1 ml of 2 \times yeast extract and tryptone (YT) at 37°C for 1 h. The number of clones within the library was determined by plating an aliquot onto 2 \times YT plates containing 30 μ g/ml chloramphenicol. The remainder of the transformation mixture was pooled and incubated in 1 liter of 2 \times YT containing chloramphenicol for 12 h at 37°C. Plasmids were purified (pTaqLIB) by CsCl gradient centrifugation.

Genetic Selection. In complementation studies, *E. coli* *recA718 polA12* cells were transformed with 0.2 μ g each of the plasmids pHSG576, pTaqDUM, pTaq, or pTaqLIB by electroporation, and the cells were allowed to recover in nutrient broth medium for 2 h at 30°C. After recovery, a small fraction of the mixture was plated in duplicate onto nutrient agar plates containing chloramphenicol. One plate was incubated at 30°C and the other at 37°C for 24 h, and the resulting colonies were counted. Only paired samples that contained 200 colonies or fewer at 30°C were analyzed, because dense plating of cells leads to elevated background at 37°C. Complementation experiments with either inactive pHSG576 or pTaqDUM consistently yielded over 100-fold fewer colonies at 37°C relative to 30°C, indicating that the background for our complementation-based selection assay is <1%. Transformation with pTaq consistently yields an equal number of colonies after incubations at 30°C or 37°C, indicating that *Taq* pol I fully restores the growth-defective phenotype at elevated temperatures. Plasmids harboring WT and mutant *Taq* pol Is were isolated by miniprep (Promega) after overnight propagation at 37°C in 2 \times YT, and 200 nt surrounding the randomized region were amplified by PCR and sequenced.

Analysis of Polymerase Activity. Three hundred fifty random colonies that grew on plates at 37°C were isolated and grown in nutrient broth individually overnight at 30°C. Each culture was grown to an OD₅₉₅ of 0.3 at 30°C, and *Taq* pol I expression was induced with 0.5 mM isopropyl β -D-thiogalactoside (IPTG), and the incubations were continued for 4 h. *Taq* pols were purified by using a modified protocol of refs. 10 and 11, which allows efficient (>50%) purification of *Taq* pol I while removing endogenous polymerase and nuclease activities. Polymerase activity was assayed by using a 20- μ l reaction mixture containing 50 mM KCl, 10 mM Tris-HCl (pH 8), 0.1% Triton X-100, 2.5 mM MgCl₂, 0.4 mg of activated calf thymus DNA, 10 μ M each dNTP, 0.25 mCi of [α -³²P]dATP, and 1 μ l of WT or mutant *Taq* pol. Incubations were at 72°C for 5 min, and reactions were stopped with the addition of 100 μ l of 0.1 M sodium pyrophosphate, followed by 0.5 ml of 10% trichloroacetic acid (TCA). Polymerase activity was quantified by collecting precipitated radioactive DNA onto glass filter papers, and the amount of radioactivity was measured by scintillation counting. Activities of all selected active clones were confirmed by conducting primer extension reactions at 55°C in the presence of all four dNTPs as described (12). Extension products from these incubations were analyzed by 14% PAGE. The frequency at which 5'-³²P-labeled primers are fully extended to the 5'-template terminus correlates well with the activity measured on activated calf thymus DNA.

Sequence Analysis. Plasmids harboring either selected or unselected *Taq* pol I were isolated by miniprep (Promega), and a 200-bp fragment containing motif A was sequenced from each isolated plasmid. Natural *E. coli* isolated from geographically diverse environments were provided by Ron Jones from University of Iowa Hospitals and Clinics. The *polA* gene was amplified from a single colony of each strain by PCR with *Pfu* DNA polymerase, and both DNA strands from the 3' terminus of the gene, including the pol I active site, were sequenced. All sequences were compared with the sequence of a laboratory K-12 strain, which we also determined.

Results and Discussion

Random Mutagenesis and Genetic Complementation. To determine the contribution of specific active site residues to proper polymerase function, we randomly mutated 13 residues of the most highly conserved region within DNA polymerases, motif A. Our random mutagenesis protocol allows creation of a large population of mutants in which each amino acid can be altered to potentially any of the other 19 (13, 14). When this protocol is coupled with a stringent selection scheme, we can determine the nature of allowable amino acid substitutions *in vivo* after sequencing selected mutants. We used random sequence mutagenesis to create substitutions within the catalytic site of *Taq* pol I and selected functional mutants in *E. coli* *recA718 polA12*, a strain that contains a temperature-sensitive mutation in the *polA* gene and can be propagated at 30°C, but not at 37°C. *Taq* pol I can fully restore the temperature-sensitive phenotype such that *E. coli* *recA718 polA12* harboring *Taq* pol I exhibits a 100% survival rate at 37°C relative to 30°C. *Taq* pol I shares 38% amino acid identity with, and has a folding pattern almost identical to, *E. coli* DNA pol I (4) and likely functions in the same pathways as *E. coli* pol I. The thermostable nature of this polymerase facilitates efficient purification and subsequent biochemical characterization of all selected mutants. The randomly mutated *Taq* library contains 200,000 independent clones. After transformation of these clones into *E. coli* *recA718 polA12*, 5% formed colonies at 37°C relative to 30°C. After subtracting the background (<1%), we estimate there are 8,000 independent library clones that encode an active *Taq* pol I. This represents the largest collection of physiologically active polymerase mutants of which we are aware. These results suggest that the polymerase catalytic

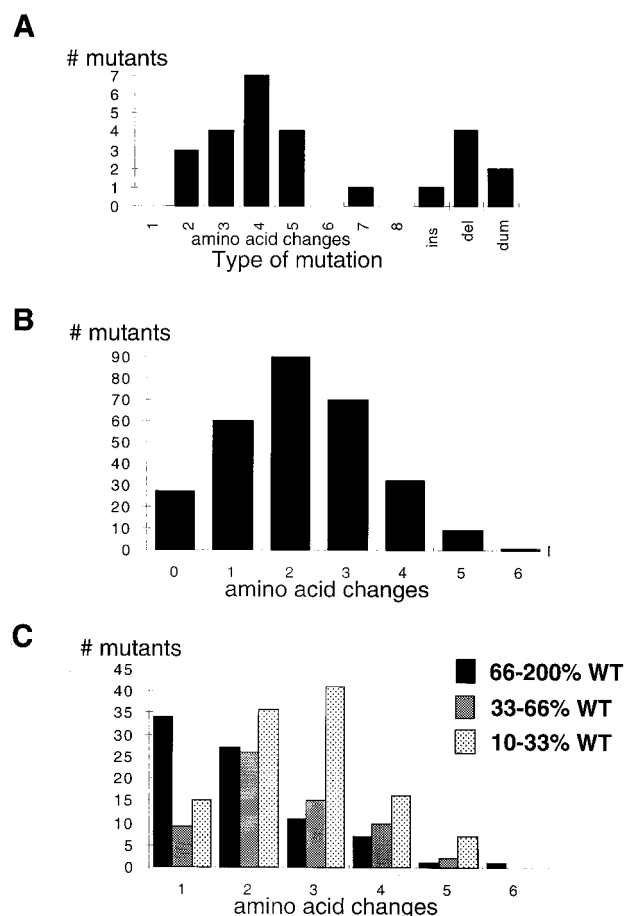


Fig. 2. Distribution of amino acid substitutions. (A) *Taq* pol I random library contains 200,000 individual clones; sequence analysis shows that the average number of amino acid substitutions within unselected library members is four; in addition, the unselected library contains insertion, deletion, and dummy fragments. (B) There are 8,000 independent clones within the selected, active population; sequence analysis shows that the average number of amino acid changes is two, with no clones containing insertion or deletion errors. (C) Distribution of high, moderate, and low activity mutants relative to number of amino acid changes. Most mutants containing a single amino acid substitution exhibit WT-like DNA polymerase activity. Mutants with moderate or low activity contain on average two or three substitutions, respectively. Mutants with high activity generally produce greater yields during PCR and presumably have higher processivity.

site can potentially accommodate a surprisingly large number of amino acid substitutions *in vivo*.

To establish the spectrum of mutations that restored growth of *E. coli recA718 polA12*, we sequenced the randomized insert from both unselected clones (30°C) and from selected clones (37°C). Analysis of sequences (Fig. 2A) from unselected plasmids, which reflects the distribution of mutants found in the random library before selection, shows that the average number of amino acid substitutions is four. In addition, several clones contained insertions, deletions, and stuffer fragments that arose during library construction. After selection, we randomly picked 350 colonies that grew at 37°C, measured *Taq* DNA polymerase activity, isolated the plasmids, and sequenced 200 nt encompassing the substituted random sequence. Of the 350 clones, 20 were inactive (<2% DNA polymerase activity relative to WT); 39 clones had low activity at 72°C (2 to 10%); whereas 291 were active (>10 to 200% WT activity). The 291 independent active clones had on average two amino acid changes, ranging from none (27 clones) to 1 clone containing six amino acid changes;

no selected active clone contained insertions or deletions (Fig. 2B).

A large number of these selected mutants retain DNA polymerase activity comparable to WT (66–200% WT). Of the 264 active mutants, approximately one-third of the mutants exhibit WT DNA polymerase activity (Fig. 2C). Most mutants containing a single amino acid substitution possess WT activity. Twenty percent (7 of 33) of mutants with four amino acid changes maintain normal activity, and a clone containing six amino acid substitutions also exhibits WT activity. The number of mutants exhibiting moderate activity (33–66% WT) or low activity (10–33% WT) follows a Poisson distribution relative to number of amino acid changes, with a median of two or three amino acid substitutions per clone, respectively. Plasmids (27 clones) that encode WT enzyme at the amino acid sequence contain silent nucleotide substitutions; these enzymes have activities similar to those of WT controls. These data indicate that, even in cases of especially pronounced mutation burden with many amino acids substituted within an evolutionarily conserved motif, a large number of mutants can exhibit high activity.

Nature of Allowable Substitutions. Sequence analysis of all 291 selected active clones (10–200% WT activity; Fig. 3A), including the 87 most active clones (>66–200% WT activity; Fig. 3B), showed that most motif A residues tolerate a wide spectrum of substitutions (Leu-605, Leu-606, Val-607, Ala-608, Leu-609, Ser-612, Ile-614, and Arg-617), some residues tolerate predominantly conservative substitutions (Tyr-611, Gln-613, Glu-615, and Leu-616), and only one residue is immutable (Asp-610). Of the highly mutable residues, Ser-612, which is present in nearly all eukaryotic and prokaryotic DNA polymerases sequenced, tolerates substitutions that are diverse in size and hydrophilicity while often preserving WT-like activity. Consistent with these mutability data, structural analysis shows that Ser-612 projects away from the catalytic site and does not appear to maintain significant interactions. Of the other highly mutable amino acids, hydrophobic residues Leu-605 to Leu-609 form a strand of the structurally conserved antiparallel β sheet that accommodates the triphosphate portion of the incoming dNTP. That these hydrophobic residues can be replaced indicates none are essential for WT activity; in fact, a mutant with six substitutions (Leu-605 \rightarrow Arg, Leu-606 \rightarrow Met, Val-607 \rightarrow Lys, Ala-608 \rightarrow Ser, Leu-609 \rightarrow Ile, and Ser-612 \rightarrow Arg) exhibits WT DNA polymerase activity.

Those residues that tolerate predominantly conservative changes (Tyr-611, Gln-613, Glu-615, and Leu-616) are believed to function in dNTP binding and/or for protein stability (4). Tyr-611 projects into a large hydrophobic pocket, anchors motif A, and is replaced only by a planar ringed amino acid. Gln-613 projects out into a small pocket near the finger motif, limits the extent of the conformational change of the finger motif during the dNTP binding step (Fig. 1), and is replaced only by amino acids that can fit into this pocket. Glu-615 hydrogen bonds with Tyr-671 (a residue located in helix O within the finger motif and stacks with the base portion of the incoming dNTP) and can be replaced only by Asp, thus retaining the hydrogen bonding potential. Leu-616 interacts with the finger motif during the dNTP binding step and is predominantly replaced by other hydrophobic amino acids. The only immutable residue is Asp-610, which even in the context of other mutations (Fig. 3A and B), cannot be replaced even by glutamic acid. Asp-610 functions to coordinate the metal-mediated catalysis reaction, leading to the incorporation of the incoming nucleotide.

Analysis of 60 mutants with a single amino acid change yields a similar distribution of mutability and allows us to determine the effect on activity conferred by specific amino acid substitutions (Fig. 3C). Residues of the antiparallel β -strand can accommodate numerous single amino acid substitutions without signifi-

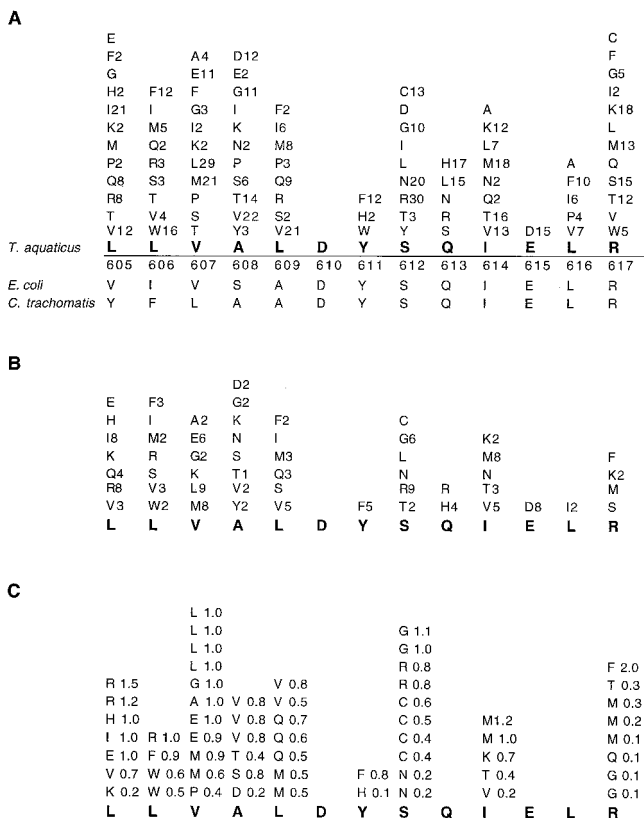


Fig. 3. High mutability of motif A. To test the importance of this conservation, Residues L⁶⁰⁵ to R⁶¹⁷ (D⁶¹⁰YSQIELR⁶¹⁷) were randomly mutated such that each contiguous amino acid can be replaced by potentially any of the other 19. (A) The degree of mutability of each amino acid within motif A from all active clones (>10% to 200% activity relative to WT) complementing an *E. coli* DNA polymerase I temperature-sensitive strain. Amino acid substitutions at the locus are listed, along with the number of times each substitution is observed. (B) Mutations in clones exhibiting high activity (66% to 200% WT). (C) Mutations in clones containing a single amino acid substitution followed by activity relative to WT. Clones containing the same amino acid substitution contain unique silent changes at other loci; thus, they represent independent sequences.

cantly compromising activity. Some of these substitutions, including Leu-605 \rightarrow Arg, confer greater polymerase activity relative to WT *Taq* pol I; all selected Leu-605 \rightarrow Arg mutants occurring in context of multiple mutations also exhibit high activity (Fig. 3A and B). The single substitution Arg-617 \rightarrow Phe confers twice the activity of WT *Taq* Pol I, although other substitutions at this locus lower *Taq* pol I activity.

Evolution of the Polymerase Active Site. Our genetic selection protocol allows isolation of mutant polymerases that retain a high DNA polymerase activity. Bacteria dependent on these approximately 300 mutant active polymerases can be grown under logarithmic growth conditions in liquid broth at 37°C (before plasmid isolation and protein purification) or as colonies in solid agar at 37°C (>50 generations) without significant variations in growth kinetics. Thus, bacteria dependent on mutant enzymes for survival are fit to replicate repetitively. In addition, most enzymes containing substitutions within motif A exhibit WT-like biochemical properties; however, a subset of enzymes (approximately 10%) exhibit diverse base pairing fidelity or altered dNTP/rNTP specificity relative to the WT enzyme (unpublished data). This inherent plasticity of the active site amino acid sequence is in contrast with the constancy of

motif A found in nature. GenBank BLAST search and sequence alignment of *polA* genes from 34 individual prokaryotes and eubacteria showed that 27 organisms have maintained the DYSQIELR motif A sequence. The few naturally occurring amino acid substitutions within motif A include Leu \rightarrow Met in species within the *Mycobacterium* genus, Tyr \rightarrow Phe in *Aquifex aeolicus*, Ser \rightarrow Val in *Rhodothermus obamensis*, Arg \rightarrow Ala in *Borrelia burgdorferi*, and Ser \rightarrow Thr and Arg \rightarrow Val in *Treponema pallidum*. These data indicate that the *polA* gene, and specifically the active site, evolves very slowly.

To further characterize the evolution of the *polA* gene, we collected strains of *E. coli* from geographically diverse regions and sequenced a portion of the *polA* gene. We assume that this geographical diversity can be equated with multiple rounds of division. The inherent mutation rate of *E. coli* (10^{-5} per nucleotide per division in mutators to 10^{-9} per nucleotide per division in nonmutators; ref. 15) predicts that, after over 10^8 years of *E. coli* and archaeal evolution (16), their *polA* genes could potentially contain silent or neutral mutations at each codon position. However, analysis of *E. coli* strains isolated from geographically distinct locations shows that each has a nearly identical amino acid and nucleotide sequence (>99% identity); nucleotides encoding motif A amino acids are 100% identical in all of these distinct *E. coli* strains (Fig. 4). Most mutations cluster in geographically distinct loci, suggesting that these alterations occurred in an early founder population and have been maintained with time. Whereas these data show that some mutations do occur in the *polA* gene, the degree of mutability is much less than that expected from our random mutagenesis study and from that predicted by the normal mutation rate of *E. coli*. Consistent with the *E. coli* sequence data, two geographically distinct strains of *Rickettsia prowazekii* (strains B and Madrid E) have identical amino acid and nucleotide sequences for the entire *polA* gene (Fig. 5 and data not shown). Although these data show high sequence identity in the *polA* gene of diverse *E. coli*, they do not allow us to determine whether or not motif A nucleotide sequence is uniquely conserved.

To determine the level of nucleotide conservation within the active sites of polymerases from other organisms, we aligned both the amino acid and nucleotide sequence of *polA* genes from diverse prokaryote and eubacteria species (Fig. 5). Only if both sequences contained an identical amino acid at a locus, was the nucleotide codon then compared. Sequence alignments of related species within the same genus (e.g., *Rickettsia* genus, *Mycobacterium* genus, or *Thermus* genus), which branched many million years ago, show nearly complete nucleotide conservation of motif A within each genus (Fig. 5). Comparison of *polA* sequences from individual species within the *Rickettsia* genus showed that the greatest degree of nucleotide sequence identity (as judged by the presence of synonymous codons) occurs within the DNA polymerase and 5' nuclease catalytic sites. For example, 24 amino acid residues within and flanking motif A are encoded by the identical codons in at least seven *Rickettsia* species. The frequency for this level of conservation for any one codon position within the entire *polA* gene of *Rickettsia* is 0.8. Thus, the probability of 24 contiguous amino acids being encoded by the same codon by random chance is <0.005 ($= 0.8^{24}$). The sequence homogeneity of the active site is unlikely due to codon selection bias through the use of a specific set of tRNAs, because multiple codons encoding the same amino acids are used randomly elsewhere in the *polA* gene within individual species of the *Rickettsia*, *Mycobacterium*, *Thermus*, and *Escherichia* genera.

Rapid Evolution in Laboratory vs. Slow Evolution in Nature. We find, after random sequence mutagenesis and selection by genetic complementation, that amino acids of the polymerase catalytic site are highly mutable. This high plasticity is a property not only of the thermostable *Taq* pol I, because regions just outside of the

	Motif A	2184	2205	2208	2247	2262	2268	2271	2340	2427	2451	2492	2499	2514	2599	2616	2619	2622
Argentina	-	G->C				C->T			A->G						T->C		T->C	A->T
Brazil	-	G->C	G->T						A->G						T->C			
Chile	-		G->T			C->T	C->T	G->A	A->G	G->T					T->C			
Columbia	-			A->T							G->A				T->C			A->G
Venezuela	-	G->C	G->T			C->T			A->G						T->C			
France	-																	
Greece	-										G->T			C->T				
Spain	-				C->A								T->G		T->C			
Switzerland	-			A->T	C->A								T->G		T->C			
Turkey	-														T->C			
Israel	-														T->C			
Mexico	-											G->C						
Canada	-								A->G						T->C			A->G
USA	-															G->A		

Fig. 4. Sequence of the *polA* gene from natural *E. coli* isolates. The sequence of both strands of the *polA* gene from nucleotide 2113 to 2642 encoding residues of motif A and the C terminus portion of DNA pol I was determined. The base change and position is shown for each isolate; nucleotides encoding motif A show no change. Position 2599 contains a cytosine in most isolates, and thus could represent the global WT *polA* sequence. An isolate from France has the identical nucleotide sequence as the K-12 strain.

catalytic site of other polymerases, including mammalian DNA pol β (17) and HIV-1 reverse transcriptase (18) and other proteins (e.g., β -lactamase, herpes simplex virus thymidine kinase, human thymidylate synthetase, and human alkyl transferase) from diverse organisms, are also highly mutable (13, 19–21).

In each case, this inherently high plasticity facilitated isolation of highly active enzymes with altered substrate specificity.

Preservation of a plastic, mutable active site could facilitate the generation of beneficial mutants under specific selective forces, such as mutant polymerases with enhanced activity or

	D	Y	S	O	I	E	L	R
	gat	tat	tct	caa	att	gaa	ctt	cgt
	gac	tac	tcc	cag	atc	gag	ctc	cgc
			tca		ata		cta	cga
			tcg				ctg	cgg
			agt				tta	aga
			agc				ttg	agg
<hr/>								
<i>Thermus aquaticus</i>	gac	tat	agc	cag	ata	gag	ctc	agg
<i>Thermus thermophilus</i>	gac	tat	agc	cag	ata	gag	ctc	cgc
<i>Thermus caldophilus</i>	gac	tat	agc	cag	ata	gag	ctc	cgc
<hr/>								
<i>Rickettsia felis</i>	gat	tat	tcc	caa	att	gag	ctt	aga
<i>Rickettsia helvetica</i>	gat	tat	tct	caa	att	gag	ctt	aga
<i>Rickettsia rhipicephali</i>	gat	tat	tct	caa	att	gag	ctt	aga
<i>Rickettsia montanensis</i>	gat	tat	tct	caa	att	gag	ctt	aga
<i>Rickettsia sibirica</i>	gat	tat	tct	caa	att	gag	ctt	aga
<i>Rickettsia rickettsii</i> (84-21C)	gat	tat	tct	caa	att	gag	ctt	aga
<i>Rickettsia typhi</i> (Wilmington)	gat	tat	tct	caa	att	gag	ctt	aga
<i>Rickettsia prowazekii</i> (B)	gat	tat	tct	caa	att	gag	ctc	aga
<i>Rickettsia prowazekii</i> (Madrid E)	gat	tat	tct	caa	att	gag	ctc	aga
<hr/>								
<i>Mycobacterium tuberculosis</i>	gac	tac	agc	cag	atc	gag	atg	cgg
<i>Mycobacterium smegmatis</i>	gac	tac	agc	cag	atc	gag	atg	cgg

Fig. 5. Nucleotide sequence conservation of the pol I catalytic site within individual genus. Most prokaryote and eubacteria *polA* genes sequenced retain the DYSQIELR motif A sequence; however, a few organisms, including those in the *Mycobacterium* genus, contain amino acid substitutions (e.g., Leu \rightarrow Met). To further characterize the apparently slow evolution of the polymerase active site, nucleotide sequences obtained from GenBank BLAST searches were compared. These results show nearly 0% base sequence change of motif A residues within each genus. Outside of motif A, 10–50% of identical amino acids at specific loci are encoded by alternate codons. Universal codons from the DNA (+) strand encoding for motif A amino acids are shown on *Top*.

those unable to incorporate harmful nucleotide analogs during chemotherapy. In addition, the plastic nature of active sites may allow proteins to tolerate high mutation burdens. It has been demonstrated that as few as three successive selection steps yielded a population of *E. coli* cells that mutated at elevated rates (22), and 1–5% of pathogenic *E. coli* and *Salmonella enterica* are mutators (23). These data predict that, after 100 million years of dividing naturally at >100 generations per year (24) and at a mutation rate of at least 10^{-9} per base per generation (15), geographically isolated prokaryotic organisms should have had a chance to mutate at each nucleotide position and exhibit markedly heterogeneous genotypes. The sequence divergence would be even greater for progeny of cells that existed as transient mutators and divided at a mutation rate up to 10^{-5} per base per generation (22, 25). However, we find that strains of identical species isolated from geographically diverse areas have nearly the same nucleotide sequence within the *polA* gene, and species within the same genus have nearly identical active site nucleotide sequence, suggesting that the *polA* gene evolves at a very slow rate. Other groups have also suggested that the entire prokaryotic and archaeal genomes evolve at rates significantly slower than that predicted by laboratory mutation rates (26).

Thus, we are presented with two surprising and seemingly contradictory findings: DNA pol I is highly plastic and allows multiple amino acid substitutions within its catalytic site, yet the *polA* gene is highly stable in nucleotide sequence within species. There are at least three possible mechanisms that could account for this dichotomy. (i) *Taq* pol I contains a highly plastic active site, although other polymerases and enzymes are not so plastic. We doubt this hypothesis is the case, because we have shown that other polymerases and enzymes are also highly mutable (9, 13, 17–21). (ii) Subtle selection pressures maintain the WT active

site. Although *E. coli* dependent on these mutant enzymes are fit to replicate repetitively (>50 generations), it is difficult to fully rule out the effects of long-term selection. However, these pressures must be stringent to maintain WT nucleotide sequences, while allowing almost no third codon position changes after prolonged evolution. At the very least, the data suggest that there are aspects of natural evolution (relative to evolution in the laboratory) that we do not yet comprehend. (iii) Genetic transfer mechanisms maintain homogeneous sequences. Specifically, we propose that the inherent plasticity of proteins we describe here enables tolerance of the high mutation burden during adverse conditions characterized by selection of mutators (22) and facilitates the generation of beneficial mutations with a selective advantage; after successful survival through periods of adverse conditions, WT sequence (one that is fit and the most prevalent) can be generated through genetic transfer. Selection for a fit sequence could potentially be the driving force for genetic transfer. This model is consistent with at least 50-fold higher rate of lateral transfer relative to the mutation rate observed in natural *E. coli* isolates (27). This model suggests that genetic transfer can function to maintain sequence homogeneity and could account for the dichotomy between inherent amino acid substitutability within diverse enzymes and the constancy of the nucleotide sequence found in nature.

We thank Dr. Ellie Adman for *Taq* pol I structure illustrations and discussions, Drs. Brad Preston and Motoshi Suzuki for helpful discussions, Matt Ashbach for technical assistance, and Drs. Phil Green, Barry Hall, Mickey Fry, Lance Encell, and Al Mildvan for critical analysis of this manuscript. This work was supported by the Medical Scientist Training Program (National Institutes of Health, National Institute of General Medical Sciences, 5T3207266) to P.H.P. and grants from the National Cancer Institute to L.A.L. (R35 CA39903 and CA78885).

- Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. (1994) *Science* **264**, 1891–1903.
- Beese, L. S., Derbyshire, V. & Steitz, T. A. (1993) *Science* **260**, 352–355.
- 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.
- Li, Y., Korolev, S. & Waksman, G. (1998) *EMBO J.* **17**, 7514–7525.
- Reha-Krantz, L. J. & Nonay, R. L. (1994) *J. Biol. Chem.* **269**, 5635–5643.
- Astatke, M., Grindley, N. D. & Joyce, C. M. (1998) *J. Mol. Biol.* **278**, 147–165.
- Wang, J., Sattar, A. K., Wang, C. C., Karam, J. D., Konigsberg, W. H. & Steitz, T. A. (1997) *Cell* **89**, 1087–1099.
- Takehsita, S., Sato, M., Toba, M., Masahashi, W. & Hashimoto-Gotoh, T. (1987) *Gene* **61**, 63–74.
- Suzuki, M., Baskin, D., Hood, L. E. & Loeb, L. A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 9670–9675.
- Grimm, E. & Arbutnot, P. (1995) *Nucleic Acids Res.* **23**, 4518–4519.
- Desai, U. J. & Pfaffle, P. K. (1995) *BioTechniques* **19**, 780–782,784.
- Patel, P. H. & Preston, B. D. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 549–553.
- Dube, D. K. & Loeb, L. A. (1989) *Biochemistry* **28**, 5703–5707.
- Encell, L. P., Landis, D. M. & Loeb, L. A. (1999) *Nat. Biotech.* **17**, 143–147.
- Drake, J. W. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7161–7164.
- Ochman, H. & Wilson, A. C. (1988) *J. Mol. Evol.* **26**, 74–86.
- Sweasy, J. B. & Loeb, L. A. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4626–4630.
- Kim, B., Hathaway, T. R. & Loeb, L. A. (1996) *J. Biol. Chem.* **271**, 4872–4878.
- Black, M. E., Newcomb, T. G., Wilson, H.-M. P. & Loeb, L. A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 3525–3529.
- Landis, D. M. & Loeb, L. A. (1998) *J. Biol. Chem.* **273**, 25809–25817.
- Christians, F. C. & Loeb, L. A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 6124–6128.
- Mao, E. F., Lane, L., Lee, J. & Miller, J. H. (1997) *J. Bacteriol.* **179**, 417–422.
- LeClerc, J. E., Li, B., Payne, W. L. & Cebula, T. A. (1996) *Science* **274**, 1208–1211.
- Gibbons, R. J. & Kapsimalis, B. (1967) *J. Bacteriol.* **93**, 510–512.
- Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P. H. & Godelle, B. (1997) *Nature (London)* **387**, 700–702.
- Ochman, H., Elwyn, S. & Moran, N. A. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 12638–12643.
- Guttman, D. S. & Dykhuizen, D. E. (1994) *Science* **266**, 1380–1383.