Prokaryotic DNA Polymerase I: Evolution, Structure, and “Base Flipping” Mechanism for Nucleotide Selection

Premal H. Patel1, Motoshi Suzuki1, Elinor Adman1, Akeo Shinkai1 and Lawrence A. Loeb1*

1The Joseph Gottstein Memorial Cancer Laboratory, Department of Pathology, University of Washington School of Medicine Seattle, WA 98195-7705, USA

Accurate transmission of DNA material from one generation to the next is crucial for prolonged cell survival. Following the discovery of DNA polymerase I in Escherichia coli, the DNA polymerase I class of enzymes has served as the prototype for studies on structural and biochemical mechanisms of DNA replication. Recently, a series of genomic, mutagenesis and structural investigations have provided key insights into how Pol I class of enzymes function and evolve. X-ray crystal structures of at least three Pol I class of enzymes have been solved in the presence of DNA and dNTP, thus allowing a detailed description of a productive replication complex. Rapid-quench stop-flow studies have helped define individual steps during nucleotide incorporation and conformational changes that are rate limiting during catalysis. Studies in our laboratory have generated large libraries of active mutant enzymes (8000) containing a variety of substitutions within the active site, some of which exhibit altered biochemical properties. Extensive genomic information of Pol I has recently become available, as over 50 polA genes from different prokaryotic species have been sequenced. In light of these advancements, we review here the structure-function relationships of Pol I, and we highlight those interactions that are responsible for the high fidelity of DNA synthesis. We present a mechanism for “flipping” of the complementary template base to enhance interactions with the incoming nucleotide substrate during DNA synthesis.

© 2001 Academic Press

Keywords: DNA polymerase; structure-function; replication; active site; recombination

Introduction

Primary functions and sequence of prokaryotic DNA Pol I

Prokaryotic species exist in diverse climates, ranging from radiation-resistant Deinococcus radiodurans (believed to have evolved one billion years ago; Battista, 1997), to thermostable Thermus aquaticus (first isolated from hot springs in Yellowstone National Park; Chien et al., 1976) to enteric bacteria Escherichia coli, to obligate intracellular bacteria Chlamydia trachomatis. Despite this environmental diversity, prokaryotic DNA Pol I sequence has remained remarkably conserved through at least one billion years (Gutman et al., 1993; Joyce et al., 1982). The gene encoding DNA polymerase I (polA) contains approximately 3000 base-pairs that encode approximately 1000 amino acid residues in a simple polypeptide chain. Organisms separated by a billion years of evolution (e.g. Deinococcus-Thermus genera and E. coli) have DNA Pol I groups that exhibit similar activities, retain \( \approx 35\% \) amino acid identity and \( \approx 50\% \) homology. Interestingly, the lethal effects of deleting endogenous Pol I activity from a specific prokaryote can be complemented by Pol I from another organism. For example, Taq Pol I can fully restore viability of E. coli lacking endogenous Pol I activity (Suzuki et al., 1986).
Pol I uses its 5'-3' exonuclease activity to fill in the resulting gap. Okazaki fragments and DNA polymerase I contain an exonuclease domain which functions to excise DNA replication errors. DNA Pol Is from diverse species within the genera Rickettsia and Thermus do not contain motifs ExoI, ExoII, and ExoIII, which are necessary for 3'-5' exonuclease function. Pol I uses its 5' nuclease activity to remove the ribonucleotide portion of newly synthesized Okazaki fragments and DNA polymerase activity to fill in the resulting gap. polA(−) E. coli join Okazaki fragments tenfold slower than wild-type bacteria, and are sensitive to DNA damage by UV damage and alkylating agents (Deluca & Cairns, 1969). During repair, Pol I fills in DNA gaps that result from the removal of a variety of DNA lesions (e.g. UV-induced thymidine dimer, the oxidative lesion 8-oxo guanine, and the alkylation lesion 4-methyl adenine, etc.; for a review of repair mechanisms, see Friedberg et al., 1995).

The Pol I family has served as a prototype for studying DNA polymerase mechanisms. There are at least six known families of DNA polymerases, which are grouped based on amino acid sequence homologies: A, B, X, RT, Pol III, and UmuC/DinB families. Family A polymerases are found primarily in organisms related to prokaryotes and include prokaryotic DNA polymerase I, mitochondrial polymerase γ, and several bacteriophage polymerases including those from odd-numbered phage (T3, T5, and T7). Family B polymerases are present in bacteriophages, viruses, archea and eukarotes. Many of these polymerases function to replicate the host genome, and include those from even-numbered phages T4 and T6, herpes viruses, archeal pol "Vent", and mammalian pols α, δ and ε. Family X, containing mammalian pol β, λ, and μ, function during DNA repair. The reverse transcriptase family contains RTs from retroviruses as well as eukaryotic telomerases (Lingner et al., 1997). The prokaryotic Pol III family encompasses the DNA polymerases that replicate the majority of bacterial genomes. The recently discovered UmuC/DinB family includes pols η, i, κ, and deoxycytidyl transferase (for a review, see Friedberg et al., 2000).

**Three-dimensional structures of the Pol I active site**

High-resolution crystal structures of DNA polymerases in the Pol I family including Taq Pol I (Kim et al., 1995; Li et al., 1998), E. coli Pol I (Klenow) (Beese et al., 1993a; Ollis et al., 1985), Bacillus polymerase (Klenow-like fragment) (Kiefer et al., 1998) and bacteriophage T7 DNA polymerase (Doublié et al., 1998) have been determined (for a review, see Steitz, 1999). The X-ray crystal structures of these polymerases resemble in overall morphology a cupped human right hand, with fingers (which bind an incoming nucleotide and interact with the single-stranded template), palm (which harbors the catalytic amino acid residues and also binds an incoming dNTP) and thumb (which binds double-stranded DNA) subdomains. Distantly related prokaryotic polymerases within the Pol I family such as Taq Pol I and E. coli Pol I exhibit less extensive homologies, nevertheless have three-dimensional structures that are virtually identical, with secondary structural elements, α-helices and β-strands, that are all nearly completely superimposable (Figure 1). These studies show extensive sequence heterogeneity between enzymes can produce identical folding patterns. The structural identity in turn suggests Pol I enzymes from eukaryotic telomerases (Lingner et al., 1997). The prokaryotic Pol III family encompasses the DNA polymerases that replicate the majority of bacterial genomes. The recently discovered UmuC/DinB family includes pols η, i, κ, and deoxycytidyl transferase (for a review, see Friedberg et al., 2000).

**Figure 1.** Superposition of Taq Pol I (Li et al., 1998) and E. coli Pol I (Klenow fragment; Beese et al., 1993a) crystal structures. Taq Pol I (red) and E. coli Pol I (blue) are separated in evolution by one billion years. The structures of these two enzymes are superimposable, such that corresponding α-helices and β-strands adopt nearly identical folding patterns (on average, <1 Å difference). Conformation of the fingers subdomain (magenta and teal) differ slightly, suggesting this region is flexible. Superposition of other related Pol I class of enzymes yielded similar results. All structure illustrations were prepared by E. Adman using MOLSCRIPT (Kraulis, 1991) and Raster3D (Merritt & Murphy, 1994).
Diverse organisms function to incorporate nucleotides by identical mechanisms.

Amino acid sequence alignments of related enzymes can help identify important regions with common catalytic functions (Figure 2). Delarue et al. (1990) compared the C-terminal polymerase domains amongst distantly related members of the Pol I family of enzymes. Following alignment, five conserved regions (numbered 1-5) within diverse DNA Pol Is were delineated; in addition, we identify a sixth conserved region (region 6). Three of these regions (3, 4, and 5) resembled the most highly conserved domains of mammalian DNA pol α in family B (termed motif A, B, and C; Figures 3 and 4). Analysis of high-resolution crystal structures of family A polymerases in complex with DNA and an incoming nucleotide suggests each of these six regions have an important role during DNA synthesis (Figure 2, region 5). Region 1 amino acid residues are located at the tip of the thumb subdomain and form a helix-loop that interacts with the minor groove of the double-stranded DNA during the nucleic acid binding step. Region 2 amino acid residues are located within the palm subdomain and interact with the template strand (along the minor groove) and thus, form the “template grip”. Region 6 amino acid residues interact with the first template base; as a result, this template could be “flipped out” of the helix axis by ≥90°, such that it cannot base-pair with the incoming dNTP (see below). Amino acid residues within regions 3, 4, and 5 correspond to motifs A, B, and C, respectively, and are located near the incoming nucleotide triphosphate and thus form

Figure 2. DNA polymerase I sequence alignments. Sequences of Taq Pol I, T7 Pol and Bst Pol I (accession numbers 70368, GI2126855) were obtained from GenBank and aligned using NCBI BLAST family of programs. Conserved amino acid residues are listed, and conserved motifs are highlighted. Regions of insertion or deletions are marked by (−), and (*) indicates an inserted segment within T7 DNA polymerase that allows binding with an accessory protein and processivity factor, thioridoxin.
portions of the dNTP binding cleft within the DNA polymerase active site.

Motifs A and C are located within the palm subdomain and are conserved in structure within families A, B, RT, and X polymerases, while motif B is located in the fingers subdomain. Structurally, motifs A, B and C of Taq DNA Pol I, E. coli Pol I, Bst Pol I, and T7 DNA Pol are nearly indistinguishable (Figures 3 and 4). Motif A begins at an antiparallel β-strand containing predominantly hydrophobic residues and continues to an α-helix. Amino acid sequences of a large variety of naturally occurring prokaryotic DNA Pol Is (Figures 3 and 4) show many sequences can produce the "antiparallel β-strand" portion of motif A; however, the "turn and α-helix" portions consistently contain the "DYSQIELR" amino acid sequence. Motif B constitutes an α-helix located in the fingers subdomain and is found in both A and B families of DNA polymerases. DNA Pol I sequence alignment indicates motif B contains a consensus sequence RRxKhhNFGhhY, where h represents hydrophobic amino acid and x represents diverse amino acids occupy a particular position.

Figure 3. Pol I sequences of motifs A, B, and C. Sequences of prokaryotic Pol I (in capital letters) and bacteriophage Pols (in lowercase letters) were obtained from GenBank and aligned using NCBI BLAST family of programs. The sequences of conserved motifs A, B, and C are shown. In the consensus sequence (bottom), amino acids that are absolutely conserved are listed; h represents hydrophobic amino acid, and x represents diverse amino acids.

Pathway of DNA Synthesis

The rates and steps involved during single nucleotide incorporation reactions have been determined for many polymerases, including those in the Pol I class (e.g. Taq Pol I (Brandis et al., 1996),...
T7 DNA Pol (Patel et al., 1991), and *E. coli* Pol I (Bryant et al., 1983; Kucnta et al., 1988). There emerges a general mechanism for single nucleotide additions:

During DNA synthesis (equation (1)): 1, polymerase (E) binds with template-primer (TP); 2, the appropriate dNTP binds with polymerase-DNA complex; 3, a nucleophilic attack results in phosphodiester bond formation; and 4, pyrophosphate (PPi) is released. Rapid quench kinetic experiments indicate that DNA binding (1) and nucleotide binding (2) occur very rapidly. The rate-limiting step is either phosphodiester bond formation or a conformational change that precedes nucleotide incorporation. The incorporation of complementary nucleotides involves dynamic interactions between the polymerase with its nucleic acid and dNTP substrates. Polymerases are proposed to undergo at least four significant conformational changes: (i) during the DNA binding step; (ii) subsequent to the dNTP binding step and immediately preceding chemical catalysis; (iii) subsequent to nucleotide incorporation during PPI release; and during (iv) translocation towards the new primer 3'-OH terminus. Theoretical studies indicate that multiple conformational changes contribute to the fidelity of catalysis by DNA polymerases (Beckman & Loeb, 1993).

High resolution crystal structures exist within the Pol I family of enzymes with all of these intermediates: (1) without DNA and dNTP (*Taq* Pol I, *E. coli* Pol I (Klenow)); (2) with DNA in the active site but without dNTP (*Taq* Pol I); (3) with DNA and dNTP (*Taq* Pol I and T7 pol); (4) with pyrophosphate (Klenow) (Beese et al., 1993b); and (5) translocated following dNTP addition (*Bst* Pol I). Of these crystal structures, *Taq* Pol I:DNA:dNTP (Li et al., 1998), T7:DNA:dNTP (Doublie et al., 1998) and *Bst* Pol I:DNA:dNTP (Kiefer et al., 1998) are particularly revealing, as each of these high-resolution structures depicts important steps during catalysis.

**Binding of polymerase to DNA**

The first step in polymerization involves the association of the polymerase with the template-primer. Comparisons amongst the crystal structures of *Taq* Pol I, *Bst* Pol I, and *E. coli* Pol I in complex with DNA show that the thumb subdomain changes conformation to nearly completely wrap around the DNA. Two significant enzyme conformational changes occur within *Taq* Pol I thumb subdomain (Li et al., 1998): (1) the thumb subdomain rotates towards the palm subdomain; and (2) the conserved amino acid residues located within the tip of the thumb domain (helices H1 and H2; region 1; Figures 2 and 4) rotate in the opposite direction relative to the rest of the thumb such that the tip is in proximity to the DNA. Together, these changes result in an approximately 30 Å wide cylinder that almost completely engulfs the DNA, such that conserved amino acid residues within the tip of the thumb subdomain (region 1)“grip” the DNA along the minor groove. A conformation change within the thumb subdomain following DNA binding has also been reported for HIV-1 RT (Patel et al., 1995). In each of the diverse polymerase:DNA high-resolution complexes studied, the enzyme interacts primarily with the sugar-phosphate DNA backbone along the minor groove, and these interactions are associated with bending of the DNA such that it adopts an S-shaped conformation. *Taq* Pol I interactions with the thumb subdomain along the minor groove result in bending double-stranded DNA three bases from the primer terminus, causing the DNA near the active site to adopt an A-like conformation (Figures 4 and 5). A second set of interactions, where the single strand template interacts with the palm subdomain (by region 2 or “template grip” amino acid residues), also contributes to this bend. A third set of polymerase:DNA interactions, results in a $\geq 90^\circ$ rotation of the first template base and a $180^\circ$ rotation of the adjacent template bases, such that the template bases are flipped outside the DNA helix axis and away from the dNTP binding site (Figures 4 and 5). This unusual S-shaped DNA conformation, in which the first bend is induced by interactions with the thumb and palm subdomains and the second bend is induced by portions of the active site interacting with the single-strand template, is also seen in high-resolution structures of *Bst* Pol I, T7 pol, and HIV-1 RT bound to DNA (Figure 4). Thus, this DNA conformation, including template base initially flipped out of the helix axis by $\geq 90^\circ$, may have a general relevance to the DNA polymerization mechanism by a wide variety of enzymes.

**Nucleotide binding**

A second conformation change within polymerases occurs during the dNTP-binding step. High-resolution crystal structures of *Taq* Pol I (Li et al., 1998), T7 DNA Pol (Doublie et al., 1998), HIV-1 RT (Huang et al., 1998), and *Bst* Pol I (Kiefer et al., 1998) bound to DNA in the presence of an incoming nucleotide suggest at least three events occur during the dNTP binding step (Figures 5 and...
6). We propose all three of the following steps are central to an “induced-fit” model for nucleotide incorporation.

(1) Structural elements within the fingers domain rotate towards the 3’ primer terminus, resulting in a “closed” structure.

(2) The template base rotates back into the helix axis (by $90^\circ$).

(3) The base portion of the incoming nucleotide forms a Watson-Crick base-pair with the template base and the triphosphate portion forms metal-mediated ionic interactions with amino acid residues of the active site.

Four high resolution DNA polymerase structures of the ternary enzyme:DNA:dNTP complex, two in an “open conformation” (Bst Pol I and Taq Pol I) and two in a “closed” conformation (T7 pol and Taq Pol I) provide valuable insights into the structural basis for nucleotide incorporation. Two unique conformations of Taq Pol I:DNA:dNTP complex were discovered by Waksman and colleagues, following crystallization experiments with varying amounts of the incoming nucleotide. In the presence of a low concentration of the incoming nucleotide, the polymerase is in an open conformation (Figures 5(a) and 6(a)), and the Taq Pol I structure (Li et al., 1998) is identical to that of Bst Pol I (which depicts the conformation of polymerase following dNTP incorporation and poised for the next round of synthesis (Kiefer et al., 1998)). In both of these structures, the first template base is flipped out $90^\circ$ away from the helix axis, and a highly conserved motif B tyrosine residue (Tyr671 of Taq Pol I and Tyr714 of Bst Pol I; Figure 7) is positioned exactly where the template base would be expected, on top of the first base-pair of the dsDNA. In this state, polymerases adopt a “closed” conformation, with fingers subdomain motif B elements located near the palm subdomain.
Crick base-pairing interactions are not observed initially, as the template base is flipped out.

At a higher concentration of the incoming nucleotide, the polymerase adopts a "closed" state (Li et al., 1998) (Figures 5(b) and 6(b)). In this conformation: (a) the template base is "flipped" $\geq 90^\circ$ into the helix axis; (b) there is a Watson-Crick base-pairing with the incoming nucleotide; (c) Phe667 of O-helix forms stacking interaction with the incoming dNTP; and (d) the fingers subdomain changes conformation to almost completely engulf the incoming nucleotide substrate. This conformational change affects significantly the position of motif B amino acid residues (O-helix) located within the fingers subdomain, and occurs concomitantly with a $40^\circ$ rotation of the O-helix towards the catalytic site amino acid residues Asp610 and Asp785. Following these conformation changes, the base portion of the incoming nucleotide packs against two hydrophobic planar amino acid residues (Tyr671 and Phe667) located in motif B, and the ribose portion of the nucleotide packs against Ile614 and the aliphatic portion of Glu615. Thus, a hydrophobic pocket surrounds the base and ribose portions of the incoming dNTP. The negatively charged triphosphate portion of the dNTP, following the conformational change, interacts with two basic (positive charged) amino acid residues (Lys663 or Arg659) in motif B. Two acidic side-chains, Asp610 and Asp785, in motif A and C, respectively, also participate in metal-mediated interactions with the triphosphate group; these amino acid residues are directly involved in stabilizing the transition state during bond formation.

An induced-fit model for nucleotide incorporation should explain how the following three interactions with the incoming nucleotide are formed during dNTP binding: (1) hydrogen bonding with the template base; (2) stacking interactions with planar ringed amino acid residues; and (3) electrostatic interactions with negatively charged phosphate groups and charged side-chains. The Taq Pol I "open" and "closed" structures suggest that the triphosphate portion of the incoming dNTP initially binds near the catalytic residues Asp610 and Asp785 via metal coordination. We propose protein:dNTP-stacking interaction occurs as the O-helix Phe667 residue interacts with the base position of the incoming nucleotide. Development of this stacking interaction facilitates movement of the O-helix (motif B) towards motif A amino acid residues and displaces Tyr671, which in turn allows the template to rotate $\geq 90^\circ$ back into the helix axis. If the incoming nucleotide is complementary to the base, then hydrogen bonding with the template base would further stabilize the dNTP. Stable stacking and base-pairing interactions would provide time for the tip of the finger (O-helix) to come in close proximity with the palm subdomain. This conformational change results in

Figure 5. Location of conserved regions within Taq Pol I. Six conserved regions within DNA Pol I are shown in "open" (a) and "closed" (b) conformations of Taq Pol I. Region 1 (light green) is located within the thumb subdomain and interacts with the DNA minor groove. Region 2 (orange) is located within the palm subdomain and forms the "template grip" that interacts with the DNA near the minor groove. Motif A (red) interacts with an incoming divalent cation-dNTP complex. Motif B (dark green) also interacts with the incoming dNTP, but is located in the fingers subdomain and changes conformation to adopt a "closed" form during the nucleotide binding step. Motif C (blue), along with motif A, is located in the palm subdomain; both of these motifs harbor the catalytic triad carbonylate containing amino acids. Region 6 (green-yellow) is located in the palm subdomain and interacts with the template strand near the site of "base flipping". The primer strand is shown in yellow, the template strand is in cyan, the flipped out bases are colored in purple, and the incoming ddNTP is colored in magenta.
hydrophobic interactions with the base and ribose portions of the incoming dNTP and hydrophilic interactions with the triphosphate group. Overall, the induced-fit model for nucleotide incorporation involving a conformation change of the O-helix and template rotation of $\geq 90^\circ$ allows establishment of stacking interactions and serves to bring the dNTP $\alpha$-phosphate close to the primer 3'-OH group, thus promoting chemical catalysis. This induced-fit mechanism for nucleotide selection would also restrict conformations and structures of the incoming nucleotide, promoting the efficiency of correct nucleotide incorporation. In summary, we propose that stable stacking interactions with the new base-pair, in conjunction with strong electrostatic interaction brought about by the movement of the tip of the O-helix into close proximity to the palm, results in a conformation that enhances base selection by DNA polymerases.

**Role of the “base-flipping” mechanism**

Crystal structures of several enzymes that process either normal or damaged bases (including glycosylases and methyl transferases) function by first rotating the designated base out of the DNA helix axis (Roberts, 1995). This base flipping mechanism involves a rotation around the phosphodiester bonds so that the base is flipped 180° out of helix axis and into the active site of the enzyme. All of the high-resolution DNA polymerase:DNA crystal structures indicate that the first template base is rotated out by $\geq 90^\circ$ and the second template base is rotated by 180° (Figures 4-7). It is currently unknown whether this rotation in diverse DNA metabolizing enzymes involves an active mechanism that rotates the sugar phosphate backbone, or whether this rotation out of the helix axis occurs by passive diffusion and the rotated conformation is stabilized by specific active site amino acid residues. Within diverse Pol I structures complexed with DNA, amino acid residues within helix Q (Region 6 amino acids, specifically in Taq Pol Arg746 to Asp759) interact with the sugar-phosphate backbone near the first template base. The flipped out first template base is initially stabilized by stacking interactions between helices O and O1, and flipped out position of the second template base stacks with His676 of helix O1 within Taq Pol I (and Tyr719 of Bst Pol I).
A mechanism for “base flipping” during DNA polymerization has several interesting implications. Such a mechanism could prevent −1 frame shift errors by hindering base-pairing to upstream templates, and thus enhance the fidelity of DNA replication. During the dNTP-binding step, the “flipping-in” mechanism allows establishment of stable stacking interactions between the new base-pair and planar side-chains within the polymerase active site (Tyr671 and Phe667). These stacking interactions help form a hydrophobic pocket that surrounds the base and ribose portions of the incoming dNTP and the template base. Thus, these base-stacking interactions contribute to the DNA polymerase base-pairing fidelity by allowing the enzyme to form a pocket that can recognize the shape of the incoming new base-pair, prior to the nucleotide incorporation step. Lastly, the torsion created by twisting the single-stranded template strand could be important for translocation (see below).

Nucleophilic attack

The catalytic site of diverse DNA polymerases is composed of three acidic amino acid residues with carboxylate side-chains (in Taq Pol I: Asp610, Asp785, and Glu786). The ternary closed Taq Pol I:DNA:dNTP structure shows that two of the carboxylate side-chains (Asp610 and Asp785) bind two metal ions (Li et al., 1998); the precise role of the third carboxylate amino acid residues is unclear. Each metal (metal A and B; Figure 6(b)) is octahedrally coordinated and contains six ligands, four in one plane and one on each side of the plane. Metal A is coordinated by 3′ OH, the α-phosphate group, oxygen atoms of Asp610 and Asp785 side-chains, and two water molecules. Metal B is coordinated by: α-, β- and γ-phosphate groups, carboxylates of Asp610 and Asp785, and the carbonyl of Tyr611. The two metal sites are 3.8 Å apart, and each ligand is separated by 2.2 Å, a coordination symmetry consistent with octahedral coordination observed in many protein:dNTP structures. A similar two-metal coordination scheme also was observed in Pol β:DNA:dNTP (Pelletier et al., 1994), T7 DNA polymerase: DNA:dNTP (Doublie et al., 1998) and HIV RT:DNA:dNTP (Huang et al., 1998) structures. The interaction of metal A with the α-phosphate is thought to enhance the electropositive character of the phosphorous atom; in addition, metal A interaction with primer 3′-OH is thought to facilitate deprotonation of the 3′-OH group (Brautigam & Steitz, 1998). Metal A coordination geometry brings the two reacting groups (the nucleophile 3′-OH and α-phosphate) in close proximity and facilitates the phosphoryl transfer reaction.

Pyrophosphate release and translocation

The phosphoryl transfer reaction leaves, as products, a primer elongated by a single nucleotide and a pyrophosphate group. Analysis of an “open” E. coli Pol I (Klenow) bound to PPI (Beese et al., 1993b) and several “closed” Pol I structures...
indicates that following nucleotide incorporation, the polymerases adopt an “open” conformation, translocate to next template position, and release PPi prior to beginning the next cycle of nucleotide incorporation. All Pol I class of enzymes not bound to both DNA and dNTP are in an “open” conformation. Within the “open” E. coli Pol I:PPi complex, PPi is complexed with the conserved Lys and Arg (within O-helix) motif B amino acid residues. Within the Taq Pol I:DNA:dNTP and T7 pol:DNA:dNTP “closed” complexes, the βscs and γ-phosphates of the dNTP interact with the O-helix conserved side-chains Arg659 and Lys663 and motif A amino acid residues Asp610 and Asp785, and the base and ribose portions of dNTP are bound in a hydrophobic pocket (Figure 6(b)). Thus, during the transition from a “closed” to “open” conformation, O-helix amino acid residues Arg659 and Lys663 can function to remove PPi away from the catalytic site. What causes the polymerase to adopt an “open” conformation? We propose conformational change into an “open” state is accompanied by a disruption of the stacking interactions following nucleotide incorporation and translocation. Consistent with this hypothesis, the structure of the Bst Pol I:DNA shows, following nucleotide incorporation and translocation, the enzyme is in an “open” conformation and there is no pocket that engulfs the incorporated nucleotide. In this structure, the PPi group has diffused away from the active site (Figure 7(a)).

Many of the Pol I:DNA interactions by region 1 and 2 amino acid residues involve the sugar-phosphate backbone along the minor groove. These non-specific interactions allow polymerase (or the DNA) to translocate freely in a spiral motion. Following nucleotide incorporation and translocation, the same interactions of the polymerase active site would need to be reforming at the new primer terminus. Thus, translocation needs to be halted when the primer 3’OH group is near the active site. Tyr714 within the translocated Bst Pol I:DNA complex (Tyr671 of Taq Pol I) projects into the DNA and stacks over the template base of the newly formed base-pair (Figure 7(a)). Following translocation, this interaction positions the primer:template at the catalytic site. The strain created by the DNA helix distortion (i.e. A-form DNA at the active site and 90-180° base flipping of the template strand) may help guide the direction of translocation (5’-3’) with respect to the primer. One can speculate that other DNA metabolizing enzymes, including helicases, may adopt a base flipping mechanism to simultaneously eliminate base-pairing and stacking interactions within the double helix during DNA unwinding. In addition, the distortion created by rotating the sugar phosphate backbone may guide the direction of DNA unwinding.

Function of Individual Amino Acid Residues in Polymerization

Structural data indicate that fewer than 10 out of over 50 amino acid residues in conserved motifs have direct interactions with the incoming dNTP substrate. For example, essential putative functions can be ascribed to amino acid residues Asp610 and Glu615 (of motif A), Tyr671, Phe667, Lys663, and Arg659 (of motif B), Asp785 and Glu786 (of motif C); however, the precise roles of all other amino acid residues within these conserved motifs are unclear. The function of different amino acid residues and side-chains in catalysis can be gleaned from a combination of kinetic measurements, site-specific mutagenesis, and random mutagenesis coupled with genetic selection. Site-specific mutagenesis has the advantage that one can assign a particular change in kinetics or substrate specificity to a distinct amino acid substitution. However, this approach is frequently limited in that catalytic efficiency of the enzyme is compromised, and the effects observed might reflect a global change in enzyme structure. Using random mutagenesis with positive genetic selection, one obtains mutant enzymes that exhibit catalytic efficiencies similar to the wild-type. However, many of the enzymes contain multiple substitutions that militate against the direct assignment of altered catalytic properties to specific substitutions. As a result, a combination of these approaches offers many advantages in structure-function studies. In considering the results so far obtained, we have chosen to analyze substitutions that occur in different motifs with the understanding that this simplification does not adequately consider distant interactions of amino acid substitutions.

Extensive number of site-directed mutagenesis studies, yielding enzymes with single amino acid substitutions have been done with the Klenow fragment, Pol β, and HIV RT. Kinetic analysis of these polymerases, which have a conserved amino acid residue substituted by a neutral residue, show these mutants frequently have low activity, and some mutants have a unique substrate specificity. For example, Joyce and colleagues (Minnick et al., 1999) replaced 28 conserved side-chains individually within the E. coli Pol I (Klenow fragment) active site to alanine (or, in one example, to leucine), and tested each of the mutants for activity and fidelity. In these experiments, the majority of mutants had a 5- to 100-fold reduction in activity relative to wild-type using gapped DNA as a template. Only four of the 28 mutants tested exhibited low fidelity. Specifically, E710A within motif A (equivalent to E615A in Taq Pol I), Y766A (Y671A) within motif B, Arg668 (Arg573) and Arg682 (Arg587) within region 2, and N845A (Asn750) of region 6 have low fidelity. These results suggest fidelity is determined by very few conserved amino acid interactions. In addition, a complete deletion of region 1 amino acid residues (located at the tip of the thumb subdomain and binds within
the minor groove of the DNA duplex) results in polymerases with low activity, processivity, and high propensity for +1 frame-shift errors (Minnick et al., 1996). Together, these site-directed mutagenesis and deletion mutagenesis studies have verified that many substitutions of conserved amino acid residues lower enzyme activity, and in some cases, alter substrate specificity.

**Applied molecular evolution of the polymerase active site**

To determine the contribution of specific active site residues to polymerase function and to evaluate the degree of plasticity within the polymerase active site in vivo, we randomly mutated 13 residues each within the two most highly conserved regions within DNA polymerases: motif A (amino acid residues 605-617 in Taq Pol I (Patel & Loeb, 2000a)) and motif B (amino acid residues 659-671 in Taq Pol I (Suzuki et al., 1996b)). Briefly, these experiments involved replacement of nucleotide sequence encoding amino acid residues with a partially randomized sequence, such that each amino acid can be altered to potentially any of the other 19. When coupled with a stringent selection scheme, one can determine the nature of allowable amino acid substitutions in vivo after sequencing selected mutants. Functional mutants were selected by complementation of 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 (Sweasy & Loeb, 1992; Uyemura & Lehman, 1976). Taq Pol I can fully restore the temperature-sensitive phenotype such that E. coli recA718 polA12 harboring Taq Pol I exhibits 100% survival rate at 37°C relative to 30°C (Patel & Loeb, 2000a).

Following transformation of the motif A library into Pol I deficient E. coli, we obtained 8000 active mutants, of which 291 were characterized for activity, fidelity and ability to incorporate ribonucleotides (Patel & Loeb, 2000a). Sequence analysis of all 291 selected active Taq Pol I clones indicated that 8 out of 13 motif A amino acid residues tolerate a wide spectrum of substitutions, four amino acid residues tolerate conservative substitutions and only one amino acid residue (Asp610) is immutable. Comparison of a high-resolution Taq Pol I bound to DNA and dNTP complex shows that the immutable residue (Asp610) functions to coordinate the metal-mediated catalysis reaction, leading to the incorporation of the incoming nucleotide. Those residues that tolerate predominantly conservative substitutions stabilize tertiary structure. For example, Glu615, which hydrogen bonds with O-helix side-chain Tyr617, can only be substituted by an Asp, and Tyr611, which projects into a hydrophobic pocket, can only be replaced by other planar amino acid residues. Most other motif A amino acid residues can be substituted to a wide range of amino acids that differ in size, shape and charge, frequently without compromising activity. Interestingly, the conserved amino acid residues from our random mutagenesis experiments correlate well with motif A consensus sequence found in nature (Figure 3).

Transformation into E. coli recA718 polA12 of motif B library containing random substitutions at amino acid residues 659-671 also yielded an extensive library of highly active DNA polymerases that contained diverse amino acid substitutions (Suzuki et al., 1996b). Sequence analysis of the selected active clones showed only two of the 13 amino acid residues are immutable (Arg659 and Lys663), two tolerate conservative substitutions (Phe667 and Tyr671), and all others tolerate a wide range of substitutions. Structural analysis shows the immutable and conservatively substitutable amino acid residues interact with the incoming dNTP. These results are largely consistent with sequence alignment data, which show within prokaryotic and bacteriophage Pol I polymerases only motif B amino acid residues Arg659, Lys663, Gly668, and Tyr671 are conserved, and Phe667 tolerates only conservative substitutions. The random mutagenesis data suggest only 3 out of 26 active site (motif A and B) residues analyzed are required for function in vivo, six can tolerate only conservative substitutions, and all other residues are highly substitutable. This inherent plasticity of the Taq Pol I active site is consistent with Pol I consensus sequence found in nature (Figure 3). Taken together, these data indicate that the DNA polymerase active site is highly mutable.

**Altered substrate specificity**

DNA polymerases have an active site architecture that must specifically configure to and incorporate each of the four deoxynucleoside triphosphates, while taking direction from templates with diverse nucleotide sequences. In addition, the active site must exclude altered nucleotides produced during cellular metabolism. The amino acid residues of motifs A and B are positioned to have a potentially important contribution toward DNA polymerase fidelity and substrate specificity. The effect of amino acid substitutions within the active site of Taq DNA polymerase was analyzed using over 300 active mutants. Mutant libraries were tested for base-pair fidelity and for the incorporation of ribonucleotides.

**Base-pairing fidelity**

The conformational changes at the catalytic site of DNA polymerases that guarantee the high fidelity of nucleotide incorporation have not been adequately established. These changes must accommodate each of the complementary, yet structurally different base-pairings. One model that addresses the fidelity requirements is that during nucleotide incorporation, stable base stacking interactions and base-pairing allow time for a confor-
natioal change, which brings the $\alpha$-phosphate towards the primer 3'-OH (Patel et al., 1995). If a non-complementary nucleotide were to bind, then the base-pairing and stacking interacting would be disrupted, and as a result, the conformational change would be retarded. In this model, low fidelity can be conferred by mutations that interfere with the transitions between “open” and “closed” conformations. Two types of mutations reduce fidelity: (1) mutations that stabilize the closed conformation, even in the absence of proper base-pairing and stacking; and (2) mutations that “widen” the hydrophobic pocket to accommodate binding of non-complementary nucleotides. Substitutions within the O-helix that confer low fidelity include Ala661Glu and Thr664Arg (Suzuki et al., 1997, 2000). These residues are located near the distal portion of the fingers subdomain, adjacent to amino acid residues that interact with the triphosphate group (Arg659 and Lys663) and away from the base stacking residues (Tyr671 and Phe667) positioned near the palm subdomain. These amino acid substitutions confer low fidelity by stabilizing the closed conformation. Substitutions within motif A that confer low fidelity include hydrophilic substitutions at position 614 (Patel et al., 2001). These mutants exhibit a tenfold higher efficiency of misinserting bases, as well as at least tenfold higher efficiency in extending mismatches. Following PCR, Taq Pol I mutants containing hydrophilic residue at position 614 exhibit up to 100-fold higher error rates relative to the wild-type enzyme and efficiently catalyze both transition and transversion errors. In addition to conferring low base-pairing fidelity, hydrophilic substitutions for Ile614 also allow mutant Taq Pol I to bypass blocking template lesions such as abasic site and vinyl chloride alkylation product ethenoA. Hydrophilic substitutions for Ile614 presumably confer a low fidelity by “widening” the dNTP-binding pocket to accommodate unusual template base and incoming nucleotide structures (for further discussion, see Patel et al., 2001).

**Incorporation of ribonucleotides**

In cells, the concentration of ribonucleotides is orders of magnitude greater than that of the corresponding deoxynucleotides. As a result, DNA polymerases evolved mechanisms to prevent the incorporation of ribonucleotides in DNA (Joyce, 1997). We analyzed this property by testing 291 “motif A mutant enzymes” for the ability to synthesize RNA. Twenty-three different mutant polymerases containing substitutions in one of two amino acid residues were identified that incorporated ribonucleotides at a rate approaching 10-fold greater than that of wild-type Taq DNA polymerase (Patel & Loeb, 2000b). Many of the 23 ribonucleotide incorporating mutants contained multiple substitutions but could be divided into two major classes: (1) Those encoding a hydrophilic substitution at Ile614; and (2) those that encode a Glu615Asp substitution. Kinetic analysis of some of the mutants shows each incorporates ribonucleotides at an efficiency ($k_{cat}/K_m$) 1000 times higher relative to the wild-type enzyme. Analysis of the Taq Pol I structure model bound with DNA and rNTP shows the ribose ring of the rNTP interacts with Ile614 (specifically, with methyl group on the $\beta$-carbon). In addition, Ile614 amino acid residues in an important region of the protein located between an $\alpha$-helix and a $\beta$-strand. Substitutions of Ile614, which is a highly mutable residue, may allow rNTP binding and incorporation by disrupting the shape of the hydrophobic pocket that interacts with the base and ribose portions of the incoming dNTP. The aliphatic portion of Glu615 also forms a portion of this hydrophobic pocket. Thus, Glu615Asp substitution may similarly disrupt the overall conformation of the pocket to allow ribonucleotides incorporation; however, this disruption is not significant enough to alter base-pairing fidelity of Taq Pol I. Taken together, these data indicate that 3 out of the 13 motif A amino acid residues (Ile614, Glu615, and Arg617), when substituted, altered biochemical properties of Taq Pol I, while many other substitutions at the other amino acid residues appear to be neutral in biochemical properties.

**Evolution of the DNA Polymerase Active Site**

Our genetic selection protocol allows isolation of mutant polymerases that retain a high DNA polymerase activity. Bacteria dependent on these over 300 mutant active polymerases can be grown under logarithmic conditions in liquid broth at 37°C (prior to 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, many enzymes containing substitutions within motif A exhibit wild-type-like DNA-dependant DNA polymerase biochemical activity. The random mutagenesis data suggest only 3 out of 26 active site (motif A and B) residues analyzed are required for function in vivo, and another three or four amino acid residues, when substituted, alter the substrate specificity of the enzyme. This high level of substitutability of evolutionary conserved amino acid residues in the polymerase active site indicates multiple sequences can encode functioning polymerases.

This inherent plasticity of the DNA polymerase active site suggests DNA polymerases from geographically isolated populations of *E. coli* should contain diverse amino acid substitutions. Specifically, we expect after $10^8$ years of evolution (Ochman & Wilson, 1988), dividing >100 times each year (Gibbons & Kapsimalis, 1967) at a mutation rate of $10^{-5}$ per nucleotide/division.
mutators; Mao et al., 1997; Oliver et al., 2000; Taddei et al., 1997) to $10^{-9}$ per nucleotide/division (in non-mutators) (Drake, 1991). *E. coli* polA genes might contain silent or neutral mutations at each codon position. However, analysis of *E. coli* strains isolated from 19 countries in four continents shows that each has a nearly identical amino acid and nucleotide sequence (>99% identity); nucleotides encoding motif A amino acid residues are 100% identical in all of these distinct *E. coli* strains. 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 (Andersson & Andersson, 1999). These data show that there is a high sequence identity in the polA gene of diverse *E. coli*. Comparison of polA sequences from individual species within the *Rickettsia* genus showed 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 acid residues being encoded by the same codon by random chance is $<0.005 (=0.824)^{24}$. 

In summary, we find 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 the polA gene evolves at a very slow rate. 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. It is always conceivable that the wild-type amino acid sequence has a selective advantage over prolonged evolution; however, it seems improbable that selection would also prevail at the nucleotide sequence level. As an alternative, we propose that genetic transfer mechanisms maintain homogenous sequences. Specifically, we propose the inherent plasticity of DNA polymerase enables tolerance of the high mutation burden during adverse conditions characterized by selection of mutators (Mao et al., 1997) and facilitates the generation of beneficial mutations with a short-term selective advantage. Following successful survival through periods of adverse conditions, the wild-type sequence (one that is fit and the most prevalent) can be generated through genetic transfer, a process involving non-reciprocal recombination by substitution with wild-type sequences. Selection for a fit sequence could potentially be the driving force for genetic transfer. This model suggests that genetic transfer can function to maintain sequence homogeneity of DNA polymerase I and presumably other enzymes and could account for the dichotomy between inherent amino acid substitutability within diverse enzymes and the constancy of the nucleotide sequence found in nature. At the very least, the plasticity of the active site of DNA polymerase I, in contrast to the exceptional conservation of this region during evolution, suggests that there are mechanisms for maintaining nucleotide sequences during evolution that need to be explored.

### Conclusions

Diverse prokaryotic species contain DNA polymerase I that are structurally identical and similar in sequence and in function. Furthermore, structural data indicate that very few (<10) amino acid residues within the highly conserved motifs A, B, and C have a direct role during nucleotide binding and incorporation. Random mutagenesis data indicate that only those residues that are important during catalysis and/or for protein folding need to be maintained; all other residues are mutable. While some substitutions alter the substrate specificity of the polymerase, many do not. Bacteria dependent on these mutated polymerases for survival are fit to replicate repetitively without significant variations in growth kinetics; however, the long-term fitness of bacteria harboring these mutant enzymes is unknown. Sequence analysis of the gene encoding Pol I shows 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. These combined data provide the underpinning for the use of *E. coli* DNA polymerase I or *Taq* DNA polymerase I as a prototype for studying mechanisms and functions of DNA polymerases in diverse organisms.

### References


Brandis, J. W., Edwards, S. G. & Johnson, K. A. (1996). Slow rate of phosphodiester bond formation accounts for the strong bias that *Taq* DNA polymerase...
ase shows against 2,3'-dideoxynucleotide terminators. Biochemistry, 35, 2189-2200.


*Edited by J. H. Miller*

(Received 1 November 2000; received in revised form 1 February 2001; accepted 14 March 2001)