THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 279, No. 31, Issue of July 30, pp. 32932–32940, 2004
Printed in U.S.A.

Investigating the Role of the Little Finger Domain of Y-family DNA Polymerases in Low Fidelity Synthesis and Translesion Replication*

Received for publication, May 11, 2004
Published, JBC Papers in Press, May 21, 2004, DOI 10.1074/jbc.M405249200

François Boudsocq‡§, Robert J. Kokoska¶, Brian S. Plosky‡, Alexandra Vaisman‡, Hong Ling‖, Thomas A. Kunkel¶, Wei Yang‖, and Roger Woodgate‡ ‡‡**

From the ‡Section on DNA Replication, Repair, and Mutagenesis, Laboratory of Genomic Integrity, NICHD, National Institutes of Health, Bethesda, Maryland 20892-2725, the ¶Laboratory of Structural Biology, NIEHS, National Institutes of Health, Research Triangle Park, North Carolina 27709, and the ‖Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Bethesda, Maryland 20892

Dpo4 and Dbh are Y-family polymerases that originate from two closely related strains of *Sulfolobaceae*. Quite surprisingly, however, the two polymerases exhibit different enzymatic properties *in vitro*. For example, Dpo4 can replicate past a variety of DNA lesions, yet Dbh does so with a much lower efficiency. When replicating undamaged DNA, Dpo4 is prone to make base pair substitutions, whereas Dbh predominantly makes single-base deletions. Overall, the two proteins are 54% identical, but the greatest divergence is found in their respective little finger (LF) domains, which are only 41% identical. To investigate the role of the LF domain in the fidelity and lesion-bypassing abilities of Y-family polymerases, we have generated chimeras of Dpo4 and Dbh in which their LF domains have been interchanged. Interestingly, by replacing the LF domain of Dbh with that of Dpo4, the enzymatic properties of the chimeric enzyme are more Dpo4-like in that the enzyme is more processive, can bypass an abasic site and a thymine-thymine cyclobutane pyrimidine dimer, and predominantly makes base pair substitutions when replicating undamaged DNA. The converse is true for the Dpo4-LF-Dbh chimera, which is more Dbh-like in its processivity and ability to bypass DNA adducts and generate single-base deletion errors. Our studies indicate that the unique but variable LF domain of Y-family polymerases plays a major role in determining the enzymatic and biological properties of each individual Y-family member.

Remarkable progress has been made in the past few years in understanding the molecular mechanisms of damage-induced mutagenesis. It has been suggested that a significant proportion of mutations arises when damaged genomic DNA is replicated in an error prone manner by one or more low fidelity polymerases (1). These polymerases appear to have evolved to specifically facilitate replication of a wide variety of DNA lesions that might otherwise block the high fidelity replication machinery. Most of these specialized polymerases are phylogenetically related to each other and have been collectively termed "Y-family" polymerases (2). The Y-family polymerases are ubiquitous and are found in all three kingdoms of life with many organisms often possessing more than one family member. The latter observations imply that Y-family polymerases play important roles in cellular survival or evolutionary "fitness" (3, 4). Indeed, defects in human Polη[1] result in the sunlight-sensitive and cancer-prone xeroderma pigmentosum variant syndrome (5, 6), whereas mutations in *Escherichia coli dinB* reduce the ability of the cell to undergo adaptive mutagenesis in stationary phase (7, 8).

Although they share little primary amino acid sequence homology with DNA polymerases from other families, structural studies of two archaeal DinB-like polymerases (Dbh and Dpo4) and the catalytic core of *Saccharomyces cerevisiae* Polη reveal that they are topologically similar to classical polymerases in that they resemble a right hand and possess "fingers," "palm," and "thumb" subdomains. In addition they possess a unique domain that has been termed the "little finger" (LF) (9), "wrist" (10), or "PAD (polymerase associated domain)" (11). The thumb and finger domains are smaller than those found in high fidelity polymerases and in the ternary complex of Dpo4 with DNA and an incoming nucleotide; the primer-template is held between the thumb and LF domains and buttresses against the finger domain (see Fig. 1*A*) (9). The importance of the LF-DNA contact is highlighted by the fact that a proteolytic fragment of Dpo4 that retains the fingers, palm, and thumb subdomains (but lacks the LF domain) is much less active than the full-length polymerase (9). Interestingly, the LF domain is the least conserved of the four domains in the Y-family polymerases, and it is hypothesized that such divergence may in part contribute to the assorted biochemical properties reported in the literature for the various Y-family polymerases (9).

To further investigate the role that the LF domain contributes to the overall enzymatic properties of Y-family polymerases, we have taken advantage of the fact that both structural and biochemical data are available for two closely related archaeal DinB-like polymerases, Dbh and Dpo4. Dbh (DinB homolog) was identified and cloned by Kulaeva *et al.* (12) in 1996 using degenerate PCR primers designed against the *E. coli umuC* and *dinB* genes. The genomic DNA used in those studies was from an archaeal strain obtained from the American Type Culture Collection (ATCC, Manassas, VA) that was originally believed to be *Sulfolobus solfataricus* P1. However, the entire genome of *Sulfolobus acidocaldarius* has recently

* This work was funded by the National Institutes of Health Intramural Research Program and by a Fondation pour la Recherche Médicale fellowship (to F. B.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Present address: Genetic Instability and Cancer, IPBS, 205 Route de Narbonne, 31077 Toulouse, Cedex, France.

** Present address: Dept. of Biochemistry, Faculty of Medicine and Dentistry, University of Western Ontario, London, Ontario N6A 5C1, Canada.

‡‡ To whom correspondence should be addressed. Tel.: 301-496-6175; Fax: 301-594-1135; E-mail: woodgate@nih.gov.

[1] The abbreviations used are: Pol, DNA polymerase; LF, little finger; CPD, cyclobutane pyrimidine dimer.

been determined and the ~2.5-kb *dbh*-containing sequence reported by Kulaeva *et al.* (12) matches perfectly with the genomic sequence from *S. acidocaldarius*.[2] Dbh therefore originates from *S. acidocaldarius* and not *S. solfataricus* P1 as was originally thought. Dpo4 (DNA polymerase IV) was identified in the genome of *S. solfataricus* P2 through BLAST searches (13) of the complete P2 genome (14), using the *dbh* gene as a search query (15). Overall, the Dbh and Dpo4 proteins share 54% identity, yet the two polymerases exhibit different enzymatic properties (15–17). Dpo4 is thermostable and exhibits robust polymerase activity. At high enzyme to template ratios Dpo4 can synthesize more than 1 kb of DNA, thereby allowing it to substitute for *Taq* polymerase in PCR assays (15). In addition, the lesion bypass properties of Dpo4 are somewhat like those of the eukaryotic translesion polymerases in that Dpo4 can bypass thymine-thymine cyclobutane pyrimidine dimers (CPDs) (15, 18, 19) and abasic sites (15, 20). In contrast, Dbh is a much more distributive polymerase when replicating undamaged DNA, is unable to incorporate a base opposite a CPD, and bypasses an abasic site with very low efficiency (16, 17, 21).

Structural studies of the two polymerases reveal that in addition to sharing high sequence homology, the fingers, palm, and thumb domains of the proteins are virtually superimposable. This suggests that the different enzymatic properties of the two enzymes might lie more in their sequence-divergent and structurally mobile LF domains. For example, in the Dpo4-DNA complex (9), the linker connecting the thumb and little finger domains interacts only with DNA. In the apo-form of Dbh, however, this linker is hydrogen-bonded to the β-sheets in the palm domain as well as the β-sheets in the little finger domain, thereby pinning the little finger domain to the catalytic core (10). For Dbh to bind substrate, this linker has to peel off from the palm domain to allow the little finger domain to reorient (see Fig. 1*A*).

To investigate the role that the LF domain may play in determining the enzymatic properties of Y-family polymerases in general, we have constructed Dbh-Dpo4 chimeras in which the LF domains and the preceding linker have been interchanged (see Fig. 1*B*). Our studies reveal that by replacing the LF domain of Dpo4 with that from Dbh, we make the enzyme more "Dbh-like." Conversely, by replacing Dbh LF with that of Dpo4, the enzyme becomes more "Dpo4-like," indicating that the LF domain is clearly a major factor in determining the physical and enzymatic properties of each polymerase. We discuss our observations in light of the crystal structure of Dbh and of the various Dpo4-DNA complexes that have been reported to date.

EXPERIMENTAL PROCEDURES

Overproduction of S. acidocaldarius Dbh—The *dbh* gene from *S. acidocaldarius* was PCR-amplified from pOS21 (12) with two oligonucleotides: ssdbhbam (5′-CGC GGA TCC TTA AAT GTC GAA GAA ATC AGA TAA ATT TG-3′) and ssdbhbsp (5′-CAT GTC ATG ATA GTG ATA TTC GTT GAT TTT G-3′) containing a BamHI and BspHI restriction enzyme site, respectively (underlined). The ~1050-bp PCR fragment was digested with BamHI and BspHI, and the fragment was gel-purified before cloning into pET16b (Novagen, Madison, WI) digested with NcoI and BamHI. The sequence of the *dbh* gene in the recombinant plasmid, called pJM349, was verified and subsequently introduced into *E. coli* strain RW382, a Δ*umuDC595*::*cat* derivative of BL21(λDE3) (22).

Generation of Dbh-LF-Dpo4 and Dpo4-LF-Dbh Chimera—The first step toward generating Dbh/Dpo4 chimeras was to introduce a unique restriction enzyme site at the junction of the LF domain in Dpo4. This was achieved by site-directed mutagenesis (QuikChange, Stratagene, La Jolla, CA) of Leu-228 (CTA → CTG) and Ala-229 (GCT → GCC) codons to produce a novel BalI restriction enzyme site within the *dpo4* gene. The BalI restriction site was generated in the Dpo4 overexpressing plasmid, p1914 (15), using oligonucleotides P2SWDW (5′-CTC GTC TCT GGC CAG AGA GAT CAA ATA TTT AGC C-3′) and P2SWUP (5′-TTG ATC TCT CTG GCC AGA GAC GAG TAT AAC GAG CC-3′) and gave rise to plasmid p1941. Chimeras were subsequently generated by domain swapping as follows. An ~700-bp NdeI-BalI fragment was amplified using pJM349 (Dbh) DNA as a template with primers P1ndeIup (5′-GGG GGG CAT ATG ATA GTG ATA TTC GTT GAT-3′) and P1bal2dw (5′-GGG GGG ATT CTT GGC CAA CTT TAG TAG ATA TAA GGC TAA GGC-3′) containing NdeI and BalI restriction sites, respectively (underlined). The amplicon was then digested with NdeI and BalI and cloned into the similarly digested plasmid, p1941. The resulting plasmid, called p1947, therefore expresses a chimeric polymerase consisting of the thumb, finger, and palm domains of Dbh and the LF domain of the Dpo4 polymerase (Dbh-LF-Dpo4) (see Fig 1*B*).

A second plasmid, p1946, expressing the thumb, finger, and palm domains of Dpo4 and the LF domain of the Dbh polymerase (Dpo4-LF-Dbh) (see Fig. 1*B*) was obtained by the amplification of a *dbh* fragment from pJM349 with oligonucleotides P1balIup (5′-GGG AAG TTG GCC AGA AAT AAA TAT AGT-3′) and P1bam2dw (5′-CCC CCC GGA TCC TTA AAT GTC GAA GAA ATC AGA-3′) containing BalI and BamHI sites, respectively (underlined). The amplicon was digested with BamHI and BalI and cloned into the similarly digested p1941 plasmid. The sequence of the chimeric *dpo4LFdbh* and *dbhLFdpo4* genes in p1946 and p1947, respectively, were verified, and the plasmids were subsequently introduced into RW382.

Purification of Dpo4, Dbh, Dbh-LF-Dpo4, and Dpo4-LF-Dbh Proteins—The protocol utilized to purify all four polymerases was based upon that described previously for Dpo4 (15) but includes several important modifications. Although all of the recombinant genes are under the control of an isopropyl-1-thio-β-d-galactopyranoside-inducible T7 promoter (in the parental pET vector), we found that there was significant expression of the recombinant proteins in the absence of induction. Furthermore, the *Sulfolobaceae* proteins are very stable in *E. coli*,[3] and significant quantities of the recombinant proteins can be recovered by simply harvesting uninduced stationary phase overnight cultures of RW382 harboring the Dpo4/Dbh-expressing plasmids. Soluble cell extracts were made as described previously. In the heat denaturation step that removes significant quantities of the thermolabile *E. coli* proteins, the temperature was reduced from 85 to 75 °C. Each polymerase was purified to homogeneity in three chromatographic steps using HiTrapQ, hydroxylapatite, and Mono S columns as described previously except that the phosphate buffer used in the HiTrapQ column was replaced by a 20 mM HEPES buffer at pH 7.0 containing 100 mM NaCl, 1 mM dithiothreitol, and 0.1 mM EDTA.

DNA Templates for in Vitro Primer Extension Assays—Most of the synthetic oligonucleotides used in the *in vitro* replication assays were synthesized by Lofstrand Laboratories (Gaithersburg, MD) using standard techniques and were gel-purified prior to use. Where utilized, the synthetic abasic site (dSpacer) was purchased from Glen Research (Sterling, VA) and was incorporated into oligonucleotide templates using standard protocols by Lofstrand Laboratories. The exception was the *cis-syn* cyclobutane pyrimidine dimer-containing oligonucleotide that was synthesized and purified by Phoenix Biotechnologies (Huntsville, AL). Primers were 5′-labeled with [γ-^{32}P]ATP (5000 Ci/mmol; 1 Ci = 37 GBq) (Amersham Biosciences) using T4 polynucleotide kinase (Invitrogen). The sequence of each primer-template is given in the legend of the respective figure in which it was used. Single-stranded M13mp18 DNA was purchased from Invitrogen.

In Vitro Primer Extension Assays—Radiolabeled primer-template DNAs were prepared by annealing the 5′-^{32}P-labeled primer to the unlabeled template DNA at a molar ratio of 1:1.5. Standard 10-μl reactions contained 40 mM Tris·HCl at pH 8.0, 5 mM $MgCl_2$, 100 μM each of ultrapure dNTP (Amersham Biosciences), 10 mM dithiothreitol, 250 μg/ml bovine serum albumin, 2.5% glycerol, and 10 nM primer-template DNA. The concentration of polymerase added varied and is given in Figs. 2–5. After incubation at 37 or 60 °C for various times, reactions were terminated by the addition of 10 μl of 95% formamide, 10 mM EDTA, and the samples were heated to 100 °C for 5 min and were briefly chilled on ice. Reaction mixtures (5 μl) were subjected to polyacrylamide, 8 M urea gel electrophoresis, and replication products were visualized by PhosphorImager analysis.

Forward Mutation Assay—Reaction mixtures (30 μl) contained 1 nM gel-purified M13mp2 gapped DNA substrate, 40 mM Tris·HCl (pH 9.0 at 22 °C), 5 mM $MgCl_2$, 10 mM dithiothreitol, 7.5 μg of bovine serum

[2] R. Garrett, personal communication.

[3] F. Boudsocq, unpublished observations.

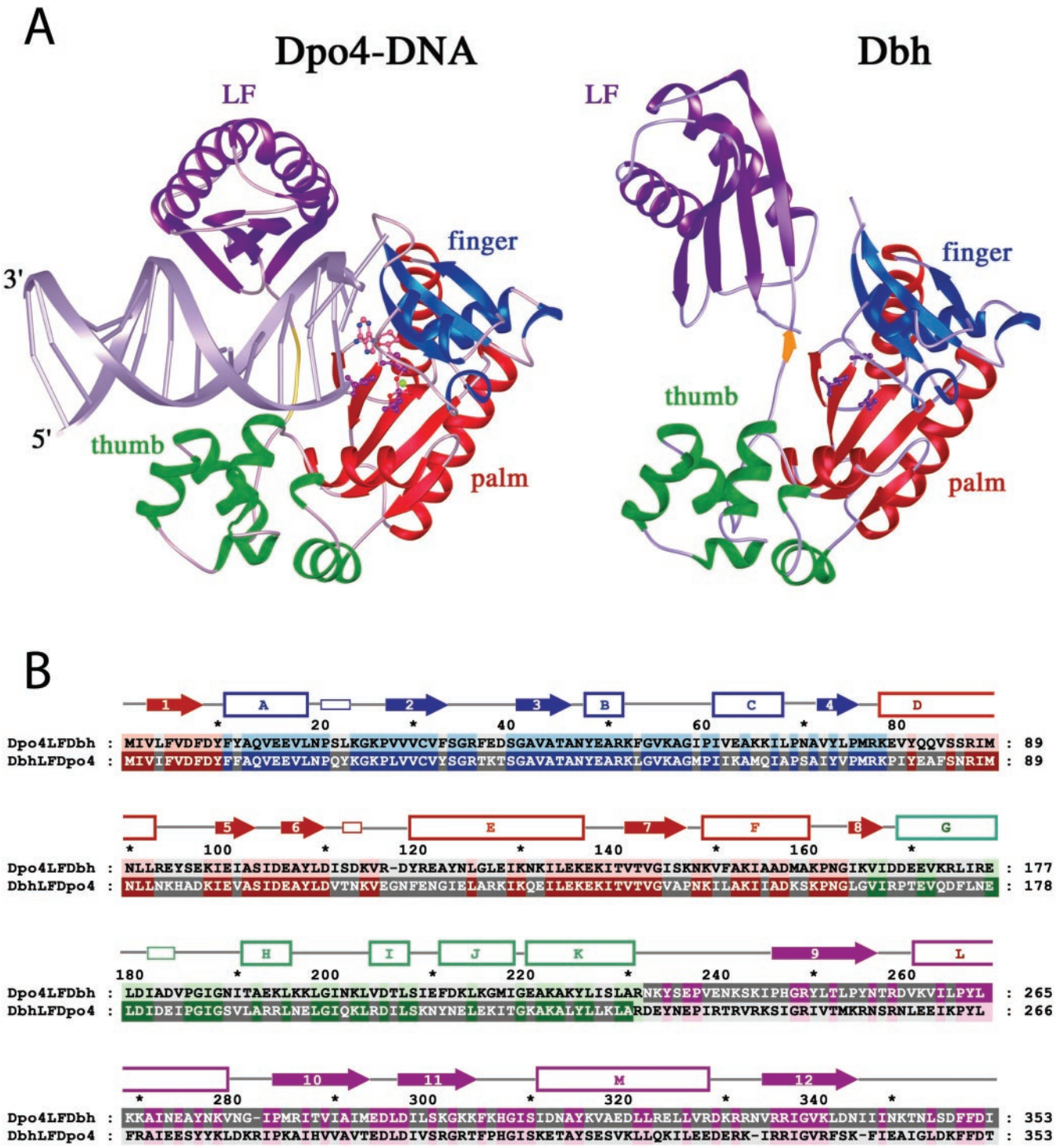

FIG. 1. **Structures of the Dpo4 and Dbh proteins and generation of Dpo4/Dbh chimeras.** *A*, ribbon diagram (34) of Dpo4 complexed with DNA and Dbh in the absence of DNA. The location of the four structural domains in Dpo4 (9) and Dbh (10, 21) are color coded as follows: *red*, palm domain; *green*, thumb domain; *blue*, finger domain; *purple*, little finger domain. A flexible linker tethers the LF and thumb domains of the polymerases. The *orange arrow* in Dbh represents a short β-strand that occurs in the middle of the linker region. This β-strand pins the LF domain to the palm domain by hydrogen bonds through the main chain atoms. The corresponding segment in Dpo4 is an extended loop that traverses the DNA backbone as shown in *yellow*. *B*, alignment of the Dpo4-LF-Dbh and Dbh-LF-Dpo4 chimeras. The finger, thumb, palm, and little finger domains are color coded as described above, and the secondary structures are indicated as *boxes* (α-helices) and *arrows* (β-sheets) above the aligned primary amino acid sequence. Residues that are identical in Dpo4 and Dbh are highlighted. The primary amino acid sequence of Dpo4 is shown in a *black typeface* on a *light gray background*, whereas that of Dbh is shown in a *white typeface* on a *dark gray background*. Dpo4-LF-Dbh consists of Dpo4 finger, palm, and thumb residues and the LF domain from Dbh. Dbh-LF-Dpo4 consists of Dbh finger, palm, and thumb residues and the LF domain of Dpo4. *Asterisk* or number indicates every other tenth residue.

albumin, 2.5% glycerol, and 1 mM concentration each of dATP, dGTP, dCTP, and dTTP. Polymerization reactions were initiated by adding 20 nM Dpo4-LF-Dbh or 1.5 nM Dbh-LF-Dpo4 incubated at 70 °C for 1 h and were terminated by adding EDTA to 15 mM. DNA products were analyzed by agarose gel electrophoresis and assayed for the frequency of *lacZ* mutants as described (23, 24). DNA samples from independent

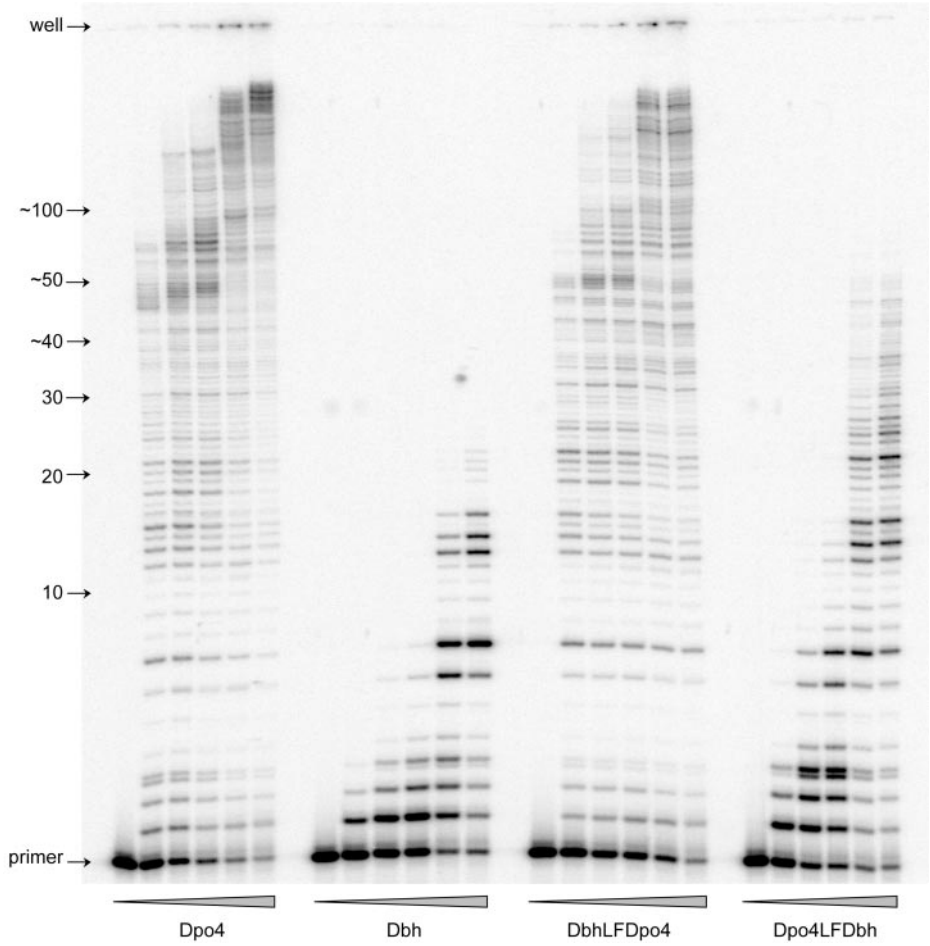

FIG. 2. **Ability of Dpo4, Dbh, and the Dbh-LF-Dpo4 and Dpo4-LF-Dbh chimeras to extend a radiolabeled 16-mer primer (5′-CTT GAA AAC ATA GCG A-3′) annealed to the single-stranded M13mp18 DNA (7.2 kb).** The primer-template was fixed at 10 nM, and elongation of the primer was assayed over a wide range of enzyme concentrations (0, 10 nM, 50 nM, 100 nM, 1 μM, 2 μM). Reactions contained all four dNTPs (100 μM each) and were performed for 5 min at 37 °C. Replication products were separated on a 12%/8 M urea polyacrylamide gel and visualized by PhosphorImager analysis. Size markers are on the *left side* of the figure.

lacZ mutant phage were sequenced to identify the sequence changes generated during gap-filling synthesis. Error rates were calculated as described previously (23, 24).

RESULTS

Generation of Little Finger Domain Chimeras—Native Dbh is a 354-amino acid protein with an estimated pI of 9.37. Dpo4 is two amino acids shorter and has an estimated pI of 9.11. Alignment of the two primary amino acid sequences reveals that although both proteins originate from related *Sulfolobaceae*, they share only 54% identity overall. Interestingly, most identity is found in the fingers, palm, and thumb subdomains of the polymerases, which are 59% identical. In contrast, the LF domain is least conserved, with only 41% primary amino acid sequence identity (Fig. 1*B*).

To investigate the role that the LF domain plays in the enzymatic properties of Y-family polymerases, we constructed chimeric proteins in which the respective LF domains and the flexible linker that tethers the LF domain to the thumb domain were interchanged (Fig. 1*B*). The first step of the process was to introduce a novel BalI restriction site into the *dbh* and *dpo4* genes at the site that corresponds to the very end of the "K" helix in the thumb domain of each polymerase (Fig. 1*B*) (18). In Dpo4 this is at the Ala-229 codon, whereas in Dbh the corresponding site is at Ala-230. The LF domain was then PCR-amplified using primers containing the novel BalI restriction enzyme site. After digestion with BalI, the amplicons were cloned into the appropriately digested parental vectors to make chimeras containing the fingers, palm, and thumb domains of Dpo4 and the LF of Dbh (termed Dpo4-LF-Dbh) or one containing the fingers, palm, and thumb domains of Dbh and the LF or Dpo4 (termed Dbh-LF-Dpo4). The LF domain of Dpo4 is 1 amino acid shorter than the Dbh LF domain, and as a consequence both chimeras are 353 amino acids long (Fig 1*B*).

Size Distribution of Replication Products Synthesized by Native and Chimeric Dpo4 and Dbh Polymerases—In vitro replication reactions with Y-family polymerases have clearly established that they are less processive than high fidelity replicative polymerases. However, the absolute number of nucleotides incorporated per DNA binding event varies considerably among Y-family polymerases. For example, recent studies suggest that archaeal Dpo4 is more processive than human Polη (20). Indeed, when replicating circular M13 DNA at high enzyme to template ratios, Dpo4 synthesizes replication products that are several hundred nucleotides in length (Fig. 2). Under the same assay conditions, Dbh-dependent replication products are much shorter. Moreover, in contrast to Dpo4, adding a large molar excess of Dbh to the reaction does not dramatically change the size distribution of replication products on the circular M13 primer-template. Interestingly, the size distribution of replication products appears to be largely dependent upon the LF domain. Replacing the native LF domain of Dbh with that of Dpo4 leads to a dramatic increase in the size of the overall length of the replication products. Conversely, replacing the native LF domain of Dpo4 with that of Dbh reduces the size distribution of replication products from several hundred nucleotides at a 20-fold molar excess to ~50 nucleotides or less at the same enzyme to template ratio (Fig. 2). Similar results were obtained in experiments performed at 60 °C with a shorter linear DNA template and a large molar excess of substrate over enzyme, which made it possible to more accurately measure the processivity of each enzyme during a single extension reaction (Fig. 3). Under reaction condi-

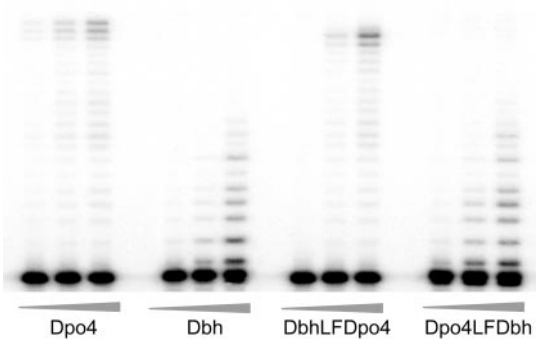

FIG. 3. **Processivity of Dpo4, Dbh, and the Dbh-LF-Dpo4 and Dpo4-LF-Dbh chimeras.** Reactions were performed at 60 °C for 3 min in the presence of all four dNTPs (100 μM each) and contained 10 nM primer-template and limiting amounts of polymerase. The primer for these assays was a radiolabeled 23-mer (5′-GCG GTG TAG AGA CGA GTG CGG AG-3′) that was annealed to a 50-mer template (5′-CTC TCA CAA GCA GCC AGG CAA GCT CCG CAC TCG TCT CTA CAC CGC TCC GC-3′, where the location of the annealed primer is underlined). The concentration of enzyme in these reactions varied considerably and was determined empirically to allow us to compare the size distribution of replication products under conditions where the percentage of primers extended was comparable between the four enzymes. The concentrations of enzyme in the 10-μl reaction were as follows: Dbh, 0.2, 0.8, and 3.3 nM; Dbh-LF-Dpo4, 0.03, 0.17, and 0.83 nM; Dpo4, 0.017, 0.08, and 0.4 nM; Dpo4-LF-Dbh, 5.5, 7.7, and 11 nM. Based upon these assays, one can clearly see that both Dpo4 and Dbh-LF-Dpo4 are considerably more processive than either Dbh or Dpo4-LF-Dbh.

tions where primer usage is minimal, full-length replication products are only observed in the presence of Dpo4 and the chimeric Dbh-LF-Dpo4, whereas those generated by either Dbh or Dpo4-LF-Dbh, are considerably shorter. Based upon these observations, we conclude that the respective LF domain of Dpo4/Dbh is the major factor determining the overall processivity of the two enzymes. Such conclusions are consistent with the crystallized ternary structure of Dpo4-DNA and incoming nucleotide, which revealed that the LF domain of Dpo4 in conjunction with the thumb domain wraps around DNA and helps hold the polymerase on to the primer terminus (Fig. 1*A*) (9).

Effect of LF Domain Swapping on Translesion DNA Synthesis of a CPD and Abasic Site—Previous studies have shown that although Dpo4 is phylogenetically located in the DinB branch of the Y-family polymerases, it actually has enzymatic properties that are reminiscent of Polη-like enzymes in that it can bypass *cis-syn* cyclobutane pyrimidine dimers. The efficiency of a Dpo4-dependent bypass of a CPD has recently been estimated to be approximately one-tenth of that of human Polη (19). The reduced ability of Dpo4 to bypass a CPD compared with Polη appears to be largely caused by stearic clashes between the 5′-T of the CPD and Dpo4 when the enzyme attempts to incorporate a nucleotide opposite the covalently linked 3′-T of the CPD (18). Nevertheless, the ability of Dpo4 to bypass a CPD is greater than that of the related PolIV (25), Polκ (26–28), or Dbh polymerases (Fig. 4), which have little ability to incorporate a base opposite the 3′-T of the dimer. Likewise, Dpo4 can bypass a synthetic abasic site (15, 20) (Fig. 4), yet Dbh only does so at high enzyme to template ratios and high levels of dNTPs (17) (Fig. 4). Similar to the results above with undamaged DNAs (Figs. 2 and 3), lesion bypass appears to depend upon the LF domain of the protein. Although Dpo4 bypasses both CPDs and an abasic site, Dbh and the Dpo4-LF-Dbh chimera showed only a limited ability to incorporate a base opposite either lesion even when the two enzymes were present at a 10-fold excess of enzyme to template (Fig. 4). In contrast, replacing the Dbh LF domain with that of Dpo4 allowed the chimera to bypass an abasic site and a CPD, although with

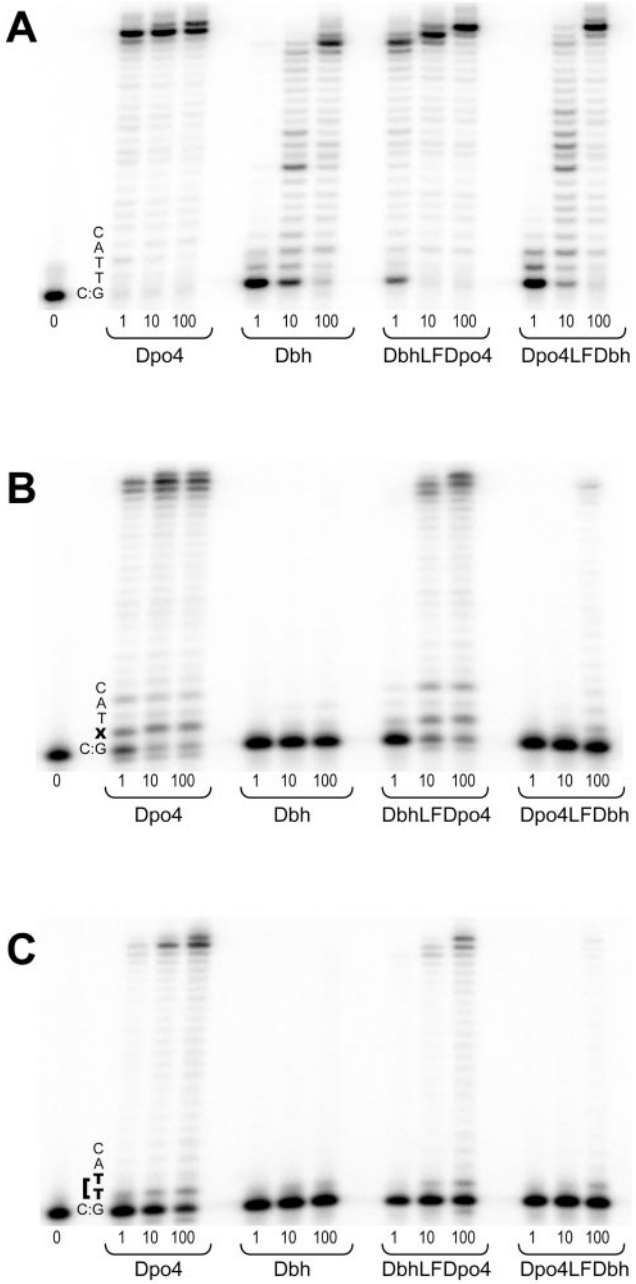

FIG. 4. **Ability of the Dpo4, Dbh, and the Dbh-LF-Dpo4/Dpo4-LF-Dbh chimeras to replicate undamaged DNA and to bypass a synthetic abasic site or a *cis-syn* cyclobutane pyrimidine dimer.** Reactions were performed at 60 °C for 5 min (undamaged DNA) or 10 min (abasic and CPD-templates) in the presence of all four dNTPs (100 μM each) and contained 10 nM primer template and 1, 10, or 100 nM of enzyme. The local sequence context is given at the *left side* of each *panel*. *A*, undamaged DNA; *B*, abasic site-containing DNA; *C*, CPD-containing DNA. The complete sequence of the undamaged template was 5′-CTC TCA CAA GCA GCC AGG CA**T-T**CT CCG CAC TCG TCT CTA CAC CGC TCC GC-3′. The 50-mer *cis-syn* dimer-containing template was identical except that it contained a single CPD located at the adjacent T letters indicated in *bold typeface*. The 50-mer abasic (*X*)-containing template was 5′ CTC TCA CAA GCA GCC AGG CAT ***X***CT CCG CAC TCG TCT CTA CAC CGC TCC GC-3′. All three templates were primed with a radiolabeled 23-mer oligonucleotide with the following sequence: 5′-GCG GTG TAG AGA CGA GTG CGG AG-3′. Replication products were separated on a 12%/8 M urea polyacrylamide gel and replication products were visualized by PhosphorImager analysis.

somewhat lower efficiency than wild-type Dpo4 (Fig. 4). Our observations suggest that like Dpo4, the active site of Dbh can accommodate both adducts but that the efficiency with which

the chimeras bypass these lesions is largely determined by their LF domain.

Effects of LF Domain Swapping on Fidelity—Because the processivity and lesion-bypassing properties of the chimeras appear to be strongly influenced by the LF domain, we were interested in determining whether the LF might also effect the fidelity of the enzymes when replicating undamaged DNA. To examine this possibility, we first analyzed the pattern of insertion of each of the four nucleotides opposite template guanine by Dpo4, Dbh, and the LF chimeras in primer extension assays that were performed at 37 or 60 °C. Similar to the temperature-dependent increase in catalytic activity reported previously for Dbh (17), all four polymerases were more active at 60 °C as compared with 37 °C. Using amounts of each enzyme yielding roughly similar levels of primer extension, we found that at both temperatures, Dbh favors incorporation of correct dCMP rather than any of the three incorrect nucleotides, whereas Dpo4 extends the radiolabeled primer by one or more bases in the presence of either correct dCTP or any of the three incorrect dNTPs (Fig. 5). This difference in insertion specificity largely depends on the LF domain because Dpo4-LF-Dbh gave a pattern similar to wild-type Dbh, whereas Dbh-LF-Dpo4 was more similar to Dpo4 than to Dbh. These qualitative misinsertion assays at a single template nucleotide suggest that the LF domain may influence the fidelity with which Y-family polymerases replicate undamaged DNA.

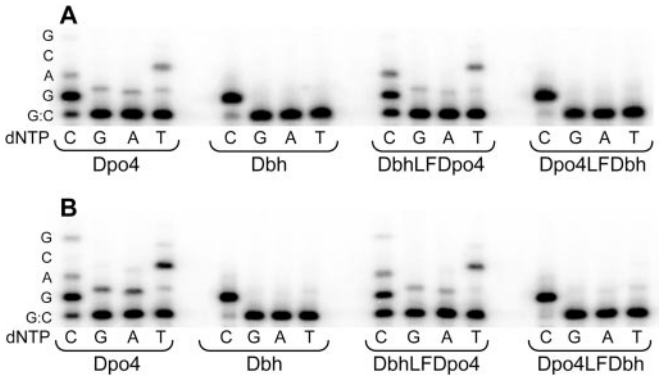

FIG. 5. **Specificity of Dpo4-, Dbh-, Dbh-LF-Dpo4-, and Dpo4-LF-Dbh-dependent nucleotide incorporation on an undamaged template.** Standard 10-μl reactions were performed at 37 or 60 °C for 2 min and contained a 10 nM concentration of radiolabeled primer-template (primer, 5′-GTG TCG GGG CGA GTG CGC CG-3′; template, 5′-CTC TCA CAA GCA GCT AAG CAG CGG CGC ACT CGC CCC GAC ACC GC-3′, with the position of the annealed primer underlined) and various amounts of polymerase. Reactions performed at 37 °C (*A*) contained 30 nM Dbh, 75 nM Dpo4-LF-Dbh, 5 nM Dpo4, and 10 nM Dbh-LF-Dpo4. Those performed at 60 °C (*B*) contained 10 nM Dbh, 25 nM Dpo4-LF-Dbh, 2.5 nM Dpo4, or 5 nM Dbh-LF-Dpo4. Products were resolved by denaturing polyacrylamide gel electrophoresis (8 M urea, 15% acrylamide) and subsequently visualized using an Amersham Biosciences PhosphorImager.

To test this hypothesis quantitatively and at a large number of template positions, we determined the effects of LF domain swapping on fidelity during synthesis of a 407-base single-stranded gap in M13mp2 DNA that contains the *lacZ* α-complementation gene sequence. We compared base substitution, addition, and deletion error rates for the two chimeric polymerases to those reported previously for Dpo4 using the same assay (24). Because we were unable to fill the *lacZ* gapped substrate with Dbh under any condition examined, for comparison to the other three enzymes we included the error rates for Dbh obtained using a 203-base substrate containing the herpes simplex virus-*tk* target gene (17). Although the *lacZ* and herpes simplex virus-*tk* gene sequences are not identical, both mutational targets score many different types of errors in numerous sequence contexts such that overall average error rates are likely to be representative when considering major classes of events, as described below.

Table I displays the results of the fidelity assays for all four polymerases. As reported earlier (24), Dpo4 has low fidelity and an average error rate for single-base deletions about 3-fold lower than for single base substitutions (Fig. 6*A*). When averaged for all 12 mispairs at many different template locations, the overall single-base substitution error rate of Dbh is similar to that of Dpo4 (Table I, far right column, from Ref. 17). However, Dbh is 13-fold less accurate than Dpo4 for single-base deletions (Table I) such that the ratio of single-base deletions to single-base substitutions (Fig. 6*A*) is much higher for Dbh (5.5 to 1) than for Dpo4 (0.35:1). Moreover, Dbh generates single-base deletions within repetitive sequence tracts at a rate that is about 8-fold higher than for single-base deletions of non-repeated nucleotides (Fig. 6*B*). In contrast, the Dpo4 rates for deleting iterated and non-iterated nucleotides differ by less than 2-fold. These results suggest that despite their sequence homology and structural similarities Dpo4 and Dbh differ in single-base deletion fidelity, with Dbh being particularly prone to deleting nucleotides in mononucleotide runs.

In this analysis, swapping the LF domains had little effect on the overall average single-base substitution error rates of Dpo4 and Dbh (Table I, second and third columns). This does not exclude an effect of the LF on error rates for specific base substitutions at specific sites (as suggested by the data shown in Fig. 5), but testing for differences in error rates for individual mispairs would require sequencing much larger collections of *lacZ* mutants or performing more sensitive base substitution reversion assays or extensive kinetic analysis. Interestingly, however, when the Dpo4 LF is swapped into Dbh (second column in Table I), the ratio of the deletion to substitution rate is more similar to that of Dpo4 than Dbh (Fig. 6*A*), as are the

TABLE I
Fidelity of Dpo4, Dbh-LF-Dpo4, Dpo4-LF-Dbh, and Dbh polymerases

	Dpo4	Dbh-LF-Dpo4	Dpo4-LF-Dbh	Dbh (17)
Total plaques	6253	1782	4574	
Total mutants	975	233	2077	
Frequency	1.56×10^{-1}	1.31×10^{-1}	4.54×10^{-1}	1.7×10^{-1}
Mutants sequenced	182	78	68	46
Total bases sequenced	50,050	21,450	18,700	7682
Total sequence changes	476	154	140	290
Changes/mutant	2.6	2.0	2.1	6.3
Number of substitutions	326	102	37	42
Number of 1-bp deletions	116	41	91	228
Number of 1-bp additions	9	5	1	3
Other	25	6	11	7
Rate of substitutions	6.5×10^{-3}	4.8×10^{-3}	2.0×10^{-3}	5.5×10^{-3}
Rate of 1-bp deletions	2.3×10^{-3}	1.9×10^{-3}	4.9×10^{-3}	3.0×10^{-2}

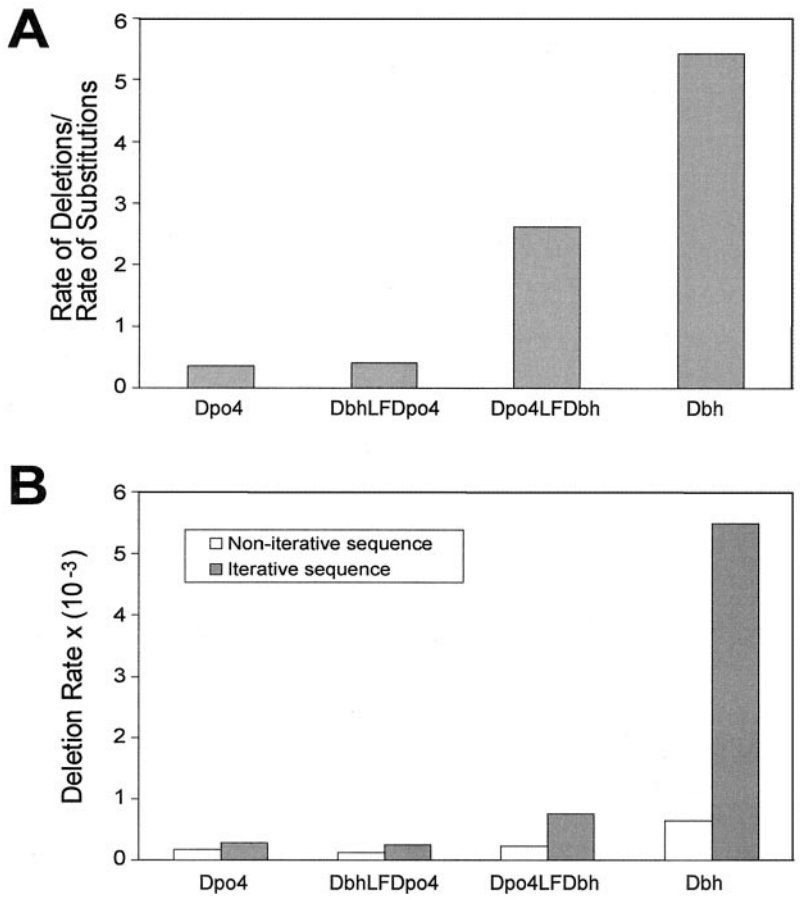

FIG. 6. **Fidelity characteristics of Dpo4, Dbh-LF-Dpo4, Dpo4-LF-Dbh, and Dbh.** *A*, ratio of single-base deletion to single-base substitution error rates for each polymerase (from Table I). *B*, comparison of deletion rates within repetitive *versus* non-repetitive sequences for each of the polymerases. Error rates are given as deletions per nucleotide copied and are calculated as described (23) to correct for differences in the number of repetitive *versus* non-repetitive nucleotides in the targets.

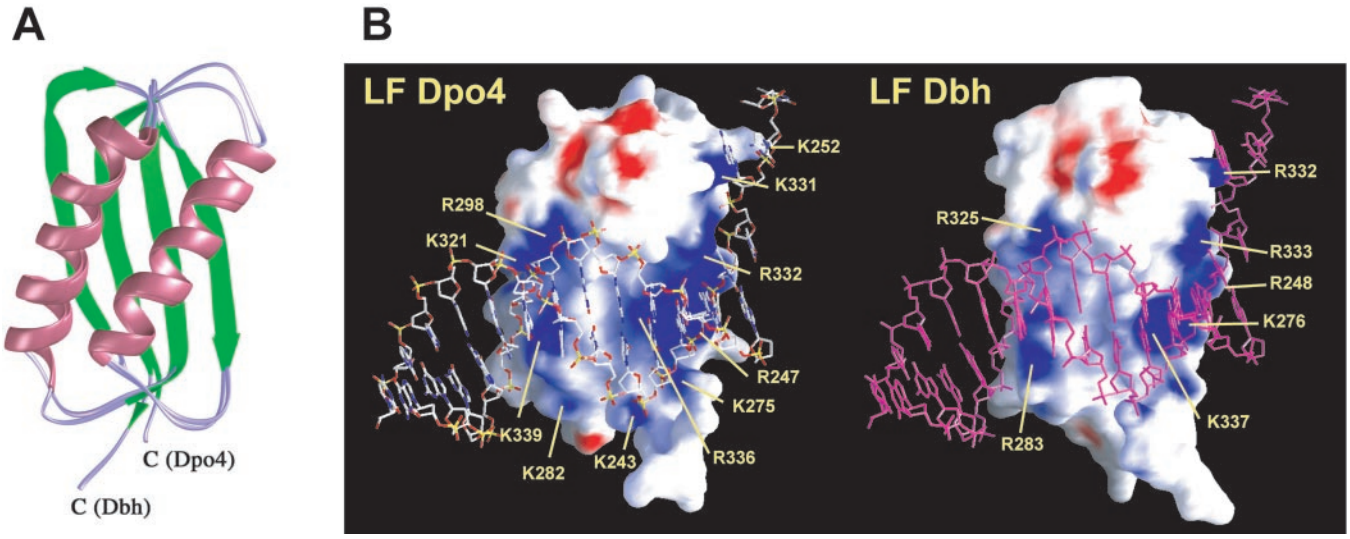

FIG. 7. **Sructural comparison of the Dpo4 and Dbh LF domains.** *A*, superposition of the Dpo4 and Dbh LF domains. Although the greatest primary amino acid sequence variation occurs in their respective LF domains, after superposition of the catalytic cores of Dpo4 and Dbh polymerases shown in Fig. 1*A*, the two LF domains can be superimposed by rotating one LF domain by 62.7 Å and translating it by 6.2 Å relative to the other. This figure was generated using RIBBONS (34). *B*, electrostatic surface representation of the isolated Dpo4 and Dbh LF domains. The two LF domains are superimposed as depicted in *A*. The LF is shown in a molecular surface representation with surface potential displayed in *blue* (positively charged) and *red* (negatively charged). The DNA substrate from the type-I Dpo4-DNA complex (9) is shown in a *stick model*. The DNA associated with Dbh is modeled after the Dpo4-DNA complex and shown in *dark pink*. Although the α-carbon backbone of the two LF domains is essentially superimposable, the proteins differ in their surface potential with Dpo4 being slightly more positively charged on the surface that contacts DNA. The positively charged residues at the protein-DNA interface are labeled. In addition, the curvature of the molecular surface also differs between the two LF domains. This figure was generated using GRASP (35).

error rates for deleting iterated and non-iterated bases (Fig. 6*B*). In contrast, when the Dbh LF is swapped into Dpo4 (third column in Table I) the ratio of deletion to substitution rate is higher (Fig. 6*A*), and the rate of deleting iterated bases is about 3-fold higher than the rate of deletion of non-iterated bases (Fig. 6*B*). In other words, the LF domains of Dpo4 and Dbh appear to differentially influence error rates, primarily for the deletion of single nucleotides within repetitive sequences.

DISCUSSION

In 2001, Ohmori *et al.* (2) presented a phylogenetic tree of the 52 available Y-family polymerases. The number of proteins related to the Y-family polymerases has now increased to well over 200 orthologs.[4] Although only a handful of these orthologs have been purified and characterized at the biochemical level, the picture emerging from these studies is that despite being closely related at the phylogenetic level, each polymerase exhibits a unique set of properties, *e.g.* processivity, fidelity, and the ability to bypass certain types of DNA lesions. These properties most likely reflect a combination of structural and kinetic features. One of these features is undoubtedly the structure of the catalytic core of each enzyme. Indeed, it is clear that single amino acid substitutions within the active site, palm, or finger domains can have a profound effect on the ability of the enzyme to perform translesion synthesis (29). A good example is the single S62G substitution in the finger domain of human Polη identified by Glick *et al.* (30), which allows the mutant enzyme to bypass a variety of lesions more efficiently than the wild-type polymerase.

Our current study reveals that the LF domain of the polymerase also has a significant effect on the processivity, fidelity, and lesion-bypassing potential of the Y-family polymerases. By generating chimeras in which the LF domain of the Dpo4 and Dbh polymerases were interchanged, we demonstrate that certain biochemical characteristics of the recombinant chimeras are similar to the polymerase from which the LF originated. For example, Dbh-LF-Dpo4 is similar to native Dpo4 in its processivity and ability to bypass a CPD or an abasic site. Conversely, swapping the LF domain of Dpo4 with that of Dbh (Dpo4-LF-Dbh) reduces the catalytic activity of the enzyme on undamaged DNA and its ability to bypass a CPD and abasic site (Figs. 2–4). The fidelity results with undamaged DNA templates (Fig. 6; Table I) indicate that error rates for single nucleotide deletions within repetitive sequences are differentially influenced by the identity of the LF domain. Interestingly, Dpo4-LF-Dbh, which deletes single nucleotides within repetitive sequences at higher rates than Dpo4, is quite distributive (Fig. 3) and is unable to generate long DNA chains in reactions involving multiple cycles of DNA binding-synthesis-dissociation (Fig. 2). This correlation is noteworthy in light of extensive evidence with polymerases in other families (31), suggesting that single-base deletion intermediates in repetitive sequences arise as the polymerase dissociates and/or reassociates with the primer-template. This implies that for Y-family polymerases with generally low fidelity for single-base deletion errors (Ref. 24 and references therein), LF domain-specific interactions with the duplex primer-template region that control enzyme processivity may also influence DNA strand alignment.

Structural Aspects of the LF Domain That Lead to DNA Binding and Translocation—Alignment of the primary amino acid sequence of the Dpo4 and Dbh LF domains reveals that they are less conserved than their catalytic cores (41 *versus* 59%), yet superposition of the LF domains in the Dpo4-DNA and apo-Dbh structures shows that the polypeptide backbone of the two domains is nearly indistinguishable (Fig. 7*A*). However, despite their overall structural similarity, the LF domains of Dbh and Dpo4 differ in their surface curvature and electropotential (Fig. 7*B*). In particular, the LF domain of Dpo4 is more positively charged, especially at key residues shown to contact the DNA backbone in the Dpo4 ternary complex structures. We suspect that these subtle changes, combined with the slightly different curvature of the surface, may alter the ability of the domain to interact with DNA. In addition, the linker region (residues 232–245) that connects the LF and thumb domains of each polymerase is highly variable and has a very different electrostatic potential. The 14-amino acid Dbh linker has an estimated pI of 8.6, but the corresponding region in Dpo4 is much more basic (estimated pI of 10), which may alter the interactions with a DNA substrate. Furthermore, the Dbh linker appears to be sandwiched between the β-sheets of both the palm and LF domains (Fig. 1*A*). This particular conformation would make it difficult for the LF to dissociate from the catalytic core of Dbh and to subsequently interact with DNA. At the present time, we do not have a crystal structure of apo-Dpo4; so we cannot exclude the possibility that in the absence of a DNA substrate, Dpo4 may also retract its LF domain in the same manner as Dbh.

Several structures of Dpo4 in a ternary complex with undamaged (9) and lesion-containing DNA (18, 32, 33) have been solved in the past few years. Comparison of these structures reveals that tertiary structures are very similar in nature with only modest variations in their finger, palm, and thumb domains. Quite strikingly, there is, however, substantial movement of the LF domain to accommodate the various types of DNA lesions encountered. Interestingly, one of these structures (Ab-5) (32), which depicts an abortive complex of Dpo4 attempting to replicate an abasic lesion, reveals that the Dpo4 LF undergoes a rotation of 54° and a 4.1-Å translation, which is similar in magnitude to the movement that would be required for the Dbh LF to dissociate from the catalytic core and to bind substrate. In the Ab-5 structure, the DNA template makes a "U-turn," and the base 3′ to the abasic lesion becomes the template for replication. These observations imply that the LF domain not only helps the polymerase bind to DNA but physical movement of the domain also facilitates DNA translocation along the primer-template. Such observations are therefore in good agreement with our findings that the LF domain plays an important role in determining the processivity of the polymerase (Fig. 4).

By making chimeras in which the LF domains of two closely related archaeal Y-family polymerases were interchanged, we have uncovered a pivotal role of the LF domain in determining the enzymatic properties of the enzymes. This includes effects on their processivity, ability to bypass template lesions, and capacity to generate base pair substitutions *versus* single-base deletions during low fidelity DNA synthesis of undamaged DNA. The LF domain of Y-family polymerases is the least conserved among their four structural domains, but like the two diverged LF domains from Dpo4 and Dbh, they will undoubtedly share common structural features. It therefore seems reasonable to hypothesize that just like Dpo4 or Dbh polymerases, subtle amino acid variations in the LF domains of the various polymerases probably contribute to a wide variety of enzymatic properties previously attributed to Y-family polymerases.

Acknowledgment—We thank John McDonald for the Dbh overexpressing plasmid, pJM349.

[4] B. S. Plosky, unpublished observations.

REFERENCES

1. Goodman, M. F. (2002) *Annu. Rev. Biochem.* **71,** 17–50
2. Ohmori, H., Friedberg, E. C., Fuchs, R. P. P., Goodman, M. F., Hanaoka, F., Hinkle, D., Kunkel, T. A., Lawrence, C. W., Livneh, Z., Nohmi, T., Prakash, L., Prakash, S., Todo, T., Walker, G. C., Wang, Z., and Woodgate, R. (2001) *Mol. Cell* **8,** 7–8
3. Friedberg, E. C., Wagner, R., and Radman, M. (2002) *Science* **296,** 1627–1630
4. Yeiser, B., Pepper, E. D., Goodman, M. F., and Finkel, S. E. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 8737–8741
5. Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999) *Nature* **399,** 700–704
6. Johnson, R. E., Kondratick, C. M., Prakash, S., and Prakash, L. (1999) *Science* **285,** 263–265
7. McKenzie, G. J., Lee, P. L., Lombardo, M. J., Hastings, P. J., and Rosenberg,

S. M. (2001) *Mol. Cell* **7,** 571–579
8. Tompkins, J. D., Nelson, J. L., Hazel, J. C., Leugers, S. L., Stumpf, J. D., and Foster, P. L. (2003) *J. Bacteriol.* **185,** 3469–3472
9. Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2001) *Cell* **107,** 91–102
10. Silvian, L. F., Toth, E. A., Pham, P., Goodman, M. F., and Ellenberger, T. (2001) *Nat. Struct. Biol.* **8,** 984–989
11. Trincao, J., Johnson, R. E., Escalante, C. R., Prakash, S., Prakash, L., and Aggarwal, A. K. (2001) *Mol. Cell* **8,** 417–426
12. Kulaeva, O. I., Koonin, E. V., McDonald, J. P., Randall, S. K., Rabinovich, N., Connaughton, J. F., Levine, A. S., and Woodgate, R. (1996) *Mutat. Res.* **357,** 245–253
13. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410
14. She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P. M., Heikamp-De Jong, I., Jeffries, A. C., Kozera, C. J., Medina, N., Peng, X., Thi-Ngoc, H. P., Redder, P., Schenk, M. E., Theriault, C., Tolstrup, N., Charlebois, R. L., Doolittle, W. F., Duguet, M., Gaasterland, T., Garrett, R. A., Ragan, M. A., Sensen, C. W., and Van Der Oost, J. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 7835–7840
15. Boudsocq, F., Iwai, S., Hanaoka, F., and Woodgate, R. (2001) *Nucleic Acids Res.* **29,** 4607–4616
16. Gruz, P., Pisani, F. M., Shimizu, M., Yamada, M., Hayashi, I., Morikawa, K., and Nohmi, T. (2001) *J. Biol. Chem.* **276,** 47394–47401
17. Potapova, O., Grindley, N. D., and Joyce, C. M. (2002) *J. Biol. Chem.* **277,** 28157–28166
18. Ling, H., Boudsocq, F., Plosky, B. S., Woodgate, R., and Yang, W. (2003) *Nature* **424,** 1083–1087
19. McCulloch, S. D., Kokoska, R. J., Masutani, C., Iwai, S., Hanaoka, F., and Kunkel, T. A. (2004) *Nature* **428,** 97–100
20. Kokoska, R. J., McCulloch, S. D., and Kunkel, T. A. (2003) *J. Biol. Chem.* **278,** 50537–50545
21. Zhou, B., Pata, J. D., and Steitz, T. A. (2001) *Mol. Cell* **8,** 427–437
22. McDonald, J. P., Frank, E. G., Levine, A. S., and Woodgate, R. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 1478–1483
23. Bebenek, K., and Kunkel, T. A. (1995) *Methods Enzymol.* **262,** 217–232
24. Kokoska, R. J., Bebenek, K., Boudsocq, F., Woodgate, R., and Kunkel, T. A. (2002) *J. Biol. Chem.* **277,** 19633–19638
25. Tang, M., Pham, P., Shen, X., Taylor, J.-S., O'Donnell, M., Woodgate, R., and Goodman, M. (2000) *Nature* **404,** 1014–1018
26. Johnson, R. E., Prakash, S., and Prakash, L. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 3838–3843
27. Ohashi, E., Ogi, T., Kusumoto, R., Iwai, S., Masutani, C., Hanaoka, F., and Ohmori, H. (2000) *Genes Dev.* **14,** 1589–1594
28. Zhang, Y., Yuan, F., Wu, X., Wang, M., Rechkoblit, O., Taylor, J. S., Geacintov, N. E., and Wang, Z. (2000) *Nucleic Acids Res.* **28,** 4138–4146
29. Boudsocq, F., Ling, H., Yang, W., and Woodgate, R. (2002) *DNA Repair* (*Amst.*) **1,** 343–358
30. Glick, E., Vigna, K. L., and Loeb, L. A. (2001) *EMBO J.* **20,** 7303–7312
31. Bebenek, K., and Kunkel, T. A. (2000) *Cold Spring Harbor Symp. Quant. Biol.* **65,** 81–91
32. Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2004) *Mol. Cell* **13,** 751–762
33. Ling, H., Sayer, J. M., Plosky, B. S., Yagi, H., Boudsocq, F., Woodgate, R., Jerina, D. M., and Yang, W. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101,** 2265–2269
34. Carson, M. (1987) *J. Mol. Graphics* **5,** 103–106
35. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins* **11,** 281–296