Structure, Vol. 12, 1395–1404, August, 2004, ©2004 Elsevier Ltd. All rights reserved. DOI 10.1016/j.str.2004.05.011

Crystal Structure of the Catalytic Core of Human DNA Polymerase Kappa

Sacha N. Uljon,[1] Robert E. Johnson,[2] Thomas A. Edwards,[1] Satya Prakash,[2] Louise Prakash,[2] and Aneel K. Aggarwal[1,*]

[1]Structural Biology Program
Department of Physiology and Biophysics
Mount Sinai School of Medicine
Box 1677
1425 Madison Avenue
New York, New York 10029

[2]Sealy Center for Molecular Science
University of Texas Medical Branch
6.104 Medical Research Building
11th and Mechanic Streets
Galveston, Texas 77555

Summary

We present the crystal structure of the catalytic core of human DNA polymerase kappa (hPolκ), the first structure of a human Y-family polymerase. hPolκ is implicated in the proficient extension of mispaired primer termini on undamaged DNAs, and in the extension step of lesion bypass. The structure reveals a stubby "fingers" subdomain, which despite its small size appears to be tightly restrained with respect to a putative templating base. The structure also reveals a novel "thumb" subdomain that provides a basis for the importance of the N-terminal extension unique to hPolκ. And, most surprisingly, the structure reveals the polymerase-associated domain (PAD) juxtaposed on the dorsal side of the "palm" subdomain, as opposed to the fingers subdomain. Together, these properties suggest that the hPolκ active site is constrained at the site of the templating base and incoming nucleotide, but the polymerase is less constrained following translocation of the lesion.

Introduction

Cellular DNA is continually damaged by external and internal agents, and both eukaryotes and prokaryotes possess DNA polymerases belonging to the Y-family that can replicate through DNA lesions. Humans have four Y-family polymerases—Polκ, Polι, Polη, and Rev1—each with a unique DNA damage bypass and fidelity profile (Goodman, 2002; Prakash and Prakash, 2002). Polη, for example, is unique in its ability to replicate through UV-induced cyclobutane pyrimidine dimers (CPDs) (Stary et al., 2003; Yu et al., 2001), and biochemical studies have shown that it replicates through a *cis-syn* thymine-thymine (T-T) dimer by inserting two As opposite the two Ts of the dimer with the same efficiency and accuracy as opposite undamaged Ts (Johnson et al., 1999b, 2000c; Washington et al., 2000, 2003). Mutations in human Polη cause the variant form of xeroderma pigmentosum (Johnson et al., 1999a; Masutani et al., 1999), characterized by a greatly enhanced predisposition to sun-induced skin cancers. hPolκ, on the other hand, is highly inefficient at replicating through a T-T dimer, and that is because of its inability to incorporate a nucleotide opposite the 3′T of the T-T dimer (Johnson et al., 2000a; Washington et al., 2002). However, hPolκ can efficiently extend from a G nucleotide incorporated opposite the 3′T of the dimer (Washington et al., 2002), suggesting a role for hPolκ in the mutagenic bypass of CPDs. Accordingly, Polκ-deficient mouse and chicken cells exhibit a significant increase in UV sensitivity (Ogi et al., 2002; Okada et al., 2002). On undamaged DNAs, hPolκ is a proficient extender of mispaired termini (Johnson et al., 2000a; Washington et al., 2002), an activity that may contribute to the rescuing of stalled replication fork when mismatches fail to be removed by the exonuclease domain of replicative polymerases during normal DNA replication.

Polκ is the only human Y-family polymerase with homologs in prokaryotes and archaea, including DinB (PolIV) in *Escherichia coli* and Dbh and Dpo4 in *Sufolobus solfataricus*, and it shares with them a tendency to generate frameshift mutations (Kim et al., 1997; Kobayashi et al., 2002; Kokoska et al., 2002; Ogi et al., 1999; Ohashi et al., 2000). However, the mechanism of frameshift mutagenesis differs between hPolκ and its prokaryotic and archeal homologs (Wolfle et al., 2003). PolIV and Dpo4 are also much less efficient at extending mispaired termini than hPolκ (Kobayashi et al., 2002; Trincao et al., 2004). Indeed, Y-family polymerases have proven to be remarkably diverse in their functions and in strategies for replicating through DNA lesions. Polκ, PolIV, and Dpo4 belong to the same subfamily of Y-family polymerases, yet differ in their properties, suggesting key differences in their three-dimensional structures. The amino acid (aa) sequence of Y-family polymerases is unrelated to that of replicative DNA polymerases, and the sequence of hPolκ is set apart from other Y-family members by an extension at the N terminus of approximately 75 amino acids (Figure 3).

Structural information on Y-family DNA polymerases is currently limited to yeast Polη (yPolη) (Trincao et al., 2001) and archaeal Dbh (Silvian et al., 2001; Zhou et al., 2001) and Dpo4, the latter in ternary complex with undamaged and damaged template-primer and an incoming nucleotide (Ling et al., 2001, 2003). Like replicative DNA polymerases, Polη, Dbh, and Dpo4 are right-hand-shaped molecules with palm, fingers, and thumb subdomains; however, they also harbor an additional domain, termed "PAD" by us to signify a polymerase-associated domain but also referred to as "little finger" and "wrist" (Ling et al., 2001; Silvian et al., 2001). In general, Y-family polymerases have less restrictive active sites than do replicative polymerases, and are thus better able to accommodate distortions in the template base. We report here the crystal structure of the catalytic core of the human polymerase kappa. The structure is the first of a human Y-family DNA polymerase and re-

*Correspondence: aggarwal@inka.mssm.edu

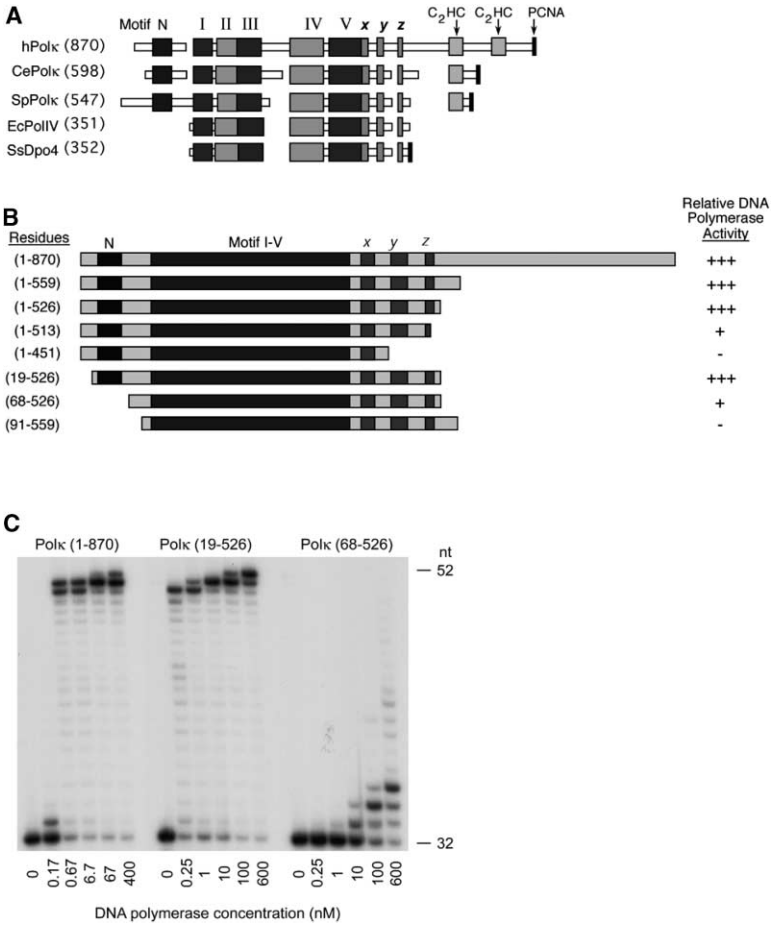

Figure 1. Catalytic Core of hPolκ

(A) Schematic alignment of several members of the DinB subfamily of Y-family polymerases. Larger, shaded boxes represent regions of homology. Thin, white boxes indicate unique sequences. The positions of motifs I–V, identified in all Y-family polymerases, are shown. Motifs *x*, *y*, and *z*, unique to the DinB subfamily, are also indicated. N indicates the region of homology in the N-terminal extension of eukaryotic members of the DinB subfamily. The C-terminal zinc fingers are indicated by C_2HC. h, human; Ce, *Caenorhabditis elegans*, Sp, *Schizosaccharomyces pombe*; Ec, *Eschericia coli*; Ss, *Sulfolobus solfataricus*.

(B) Schematic representation of deletion mutations generated in hPolκ. Amino acid residues contained in each protein are indicated on the left. DNA polymerase activity, in relation to full-length hPolκ is indicated on the right.

(C) DNA polymerase activity of full-length hPolκ versus the hPolκ (19-526) and hPolκ (68-526) truncated proteins. Amount of protein in each assay is indicated on the bottom. Reactions contained 10 nM primer:template DNA substrate and 50 μM dNTPs and were carried for 10 min at 37°C.

veals unique features. In particular, its active site is highly constrained at the site of the templating base and the incoming nucleotide, and unlike the other Y family polymerases, the PAD occupies a position away from the fingers subdomain but near to the palm subdomain. These observations help explain hPolκ's role in lesion bypass.

Results

Defining the Catalytic Core

The human Polκ protein is set apart from other Y-family DNA polymerases by the presence of unique N-terminal and C-terminal regions. In fact, even within the DinB subfamily, the eukaryotic members differ from prokaryotic and archaeal members (Figure 1A). The C terminus of hPolκ contains two zinc finger motifs and the N terminus contains a long extension, both of which are absent in *E. coli* Pol IV and *S. solfataricus* Dpo4. To test the importance of these N-terminal and C-terminal regions, and to identify the minimal catalytic portion of hPolκ, we made a series of N-terminal and C-terminal deletions of hPolκ and tested their DNA polymerase activities. The C terminus was deleted at several positions predicted to be near the end of the PAD region. Deletion of up to 344 amino acids of the C terminus of hPolκ, as in hPolκ (1-526), had no effect on DNA polymerase activity (Figure 1B). Further deletion of C-terminal residues, predicted to form part of the PAD in hPolκ, resulted in loss of DNA polymerase activity (Figure 1B). For instance, hPolκ protein from amino acids 1–513, which lacks the last 357 amino acids of hPolκ, was much reduced in DNA polymerizing activity, and a hPolκ (1-451) protein was found to be completely inactive as a DNA polymerase. Thus we used a protein terminated at amino acid position 526 for use in crystal analysis. To identify the importance of the unique first 100 residues of hPolκ which precedes motif I, we made three N-terminal deletion mutants (Figure 1B). The first was at position 19 of hPolκ, and this protein retains a region of homology found among all eukaryotic Polκ proteins. A second truncation started at position 68, which immediately follows this conserved region. A third truncation was made that mimics the initiation positions in *E. coli* and *S. solfataricus* PolIV and Dpo4, respectively, and is located 10 amino acids N-terminal to motif I. Deletion of the first 19 amino acids had no effect on hPolκ DNA polymerase activity, whereas deletion of the first 68 residues reduced activity (Figure 1C). The hPolκ protein lacking the first 91 amino acids contained no DNA polymerase activity (Figure 1B). Thus, the N-terminal region is indispensable for DNA polymerase activity, and the conserved region between amino acids 19 and 68 is required for complete activity.

Structural Determination

We sought to determine the structure of the catalytic core of hPolκ, but attempts to crystallize the 19-526

Table 1. Crystallographic Parameters

Data Collection Statistics[a]				
		Se-Met Derivative		
	Native	Edge	Peak	Remote
Wavelength (Å)	1.12709	0.97934	0.97920	0.96859
Max. resolution (Å)	2.40	2.90	2.70	3.11
Total no. of reflections	257,587	167,134	206,584	151,068
No. of unique reflections	76,925 (7,574)	23,642 (2,798)	28,784 (3,165)	18,932 (2,007)
R_{merge} (%)[b]	6.4 (43.6)	8.9 (15.6)	11.0 (18.6)	8.6 (13.7)
Completeness (%)	99.7 (99.4)	99.9 (100)	98.7 (99.5)	97.7 (99.3)
$I/\sigma(I)$	11.0 (1.9)	12.94 (6.16)	8.91 (3.62)	8.81 (5.87)
No. of Se sites	N/A	30	30	30
Refinement Statistics				
Resolution range (Å)	50–2.4			
R_{cryst} (%)[c]	24.60			
R_{free} (%)[d]	28.12			
Nonhydrogen atoms				
Protein	6,185			
Water	363			
Rms deviations				
Bonds (Å)	0.011			
Angles (°)	1.88			
Average B factor (Å^2)	40.4			

[a] Values for outermost shell are given in parentheses.
[b] $R_{merge} = \Sigma|I - \langle I \rangle|/\Sigma|$, where I is the integrated intensity of a given reflection.
[c] $R_{cryst} = ||F_o| - |F_c||/\Sigma|F_o|$.
[d] R_{free} was calculated using 5% of data excluded from refinement.

construct were unsuccessful. Fortunately, the 68-526 construct yielded well-diffracting crystals. The construct retains a significant portion of the N-terminal extension crucial to hPolκ activity, though the activity is reduced compared to the 19-526 construct. Orthorhombic crystals were obtained from a mixture of polyethylene glycol and ethylene glycol, containing two molecules per asymmetric unit (AU). For phasing, multiwavelength anomalous dispersion (MAD) data were measured from a selenomethionine (SeMet) derivative (Table 1) that yielded an interpretable electron density map (2.6 Å resolution) that allowed building of both hPolκ molecules (A and B) in the AU (Figure 2). The two molecules were refined independently, without noncrystallographic averaging, through iterative rounds of simulated annealing and rebuilding to the 2.4 Å resolution limit of the native data. The refined model contains residues 71–225, 229–253, 276–407, 411–412, 414–447, 451–472, 474–480, and 483–515 for molecule A, residues 75–224, 281–409, 413–473, and 476–517 for molecule B, and 363 water molecules. R_{cryst} and R_{free} are 24.6% and 28.1%, respectively, for data between 50 and 2.4 Å resolution. The root-mean-square deviations (rmsds) for bonds and angles are 0.0108 Å and 1.878°, respectively.

Overall Arrangement

The hPolκ catalytic core is composed of palm (amino acids 101–109 and 171–338), fingers (aa 110–170), and

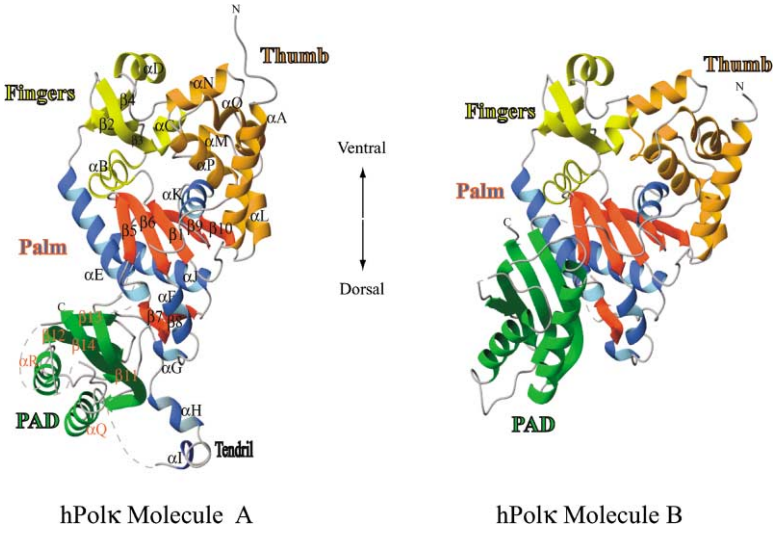

Figure 2. Structure of hPolκ

Molecules A (left) and B (right) of the asymmetric unit are shown. Regions that are unstructured are represented by dashed lines. The palm subdomains are shown in red (for β strands) and blue (for α helices), the thumb subdomains in orange, the fingers subdomains in yellow, and the PAD in green.

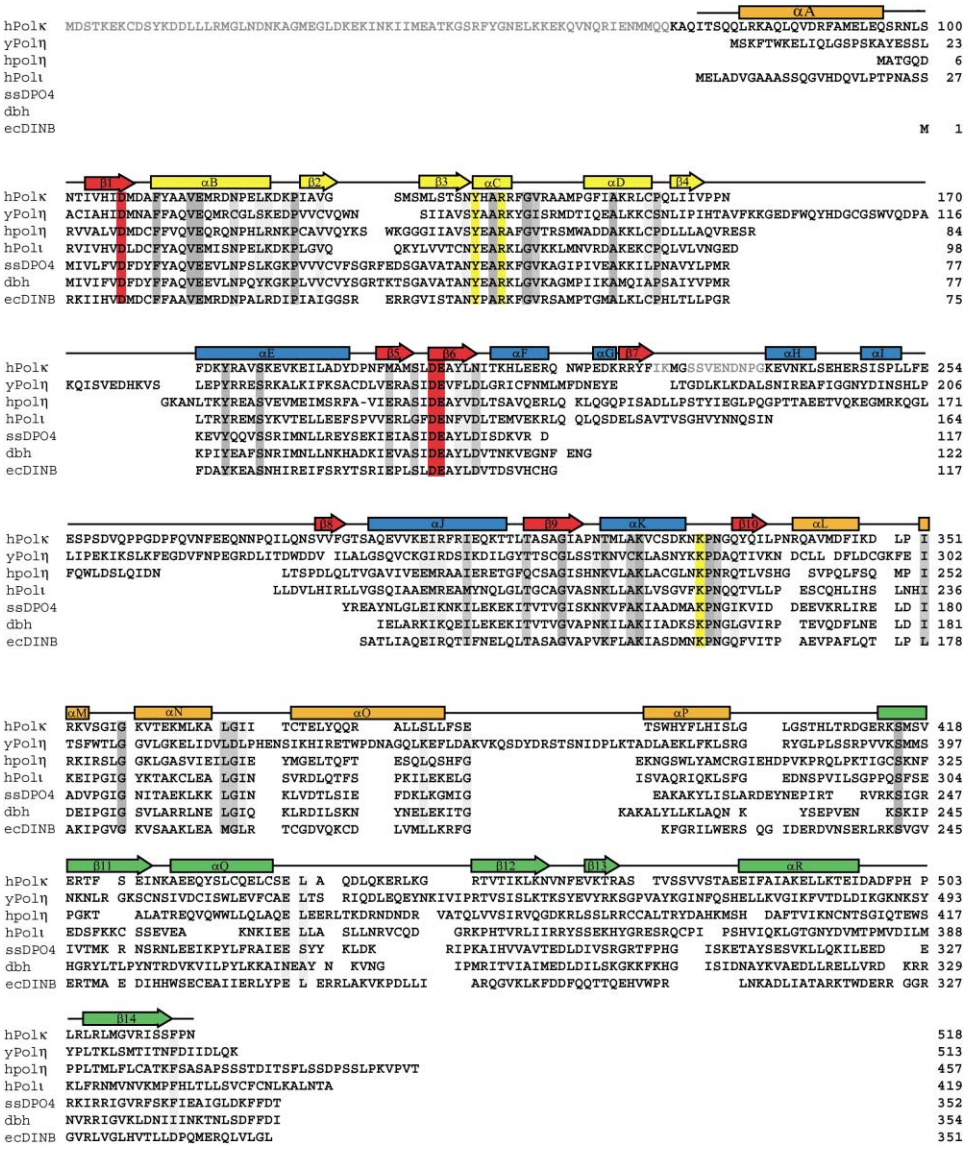

Figure 3. Structure-Based Sequence Alignment of the Human Y-Family Polymerases Polκ, Polη, Polι, *S. cerevisiae* Polη, *S. solfataricus* Dbh, DPO4, and *E. coli* DinB

Secondary structure elements (rectangle for α helix, arrow for β strand) correspond to subdomain colors in Figures 2 and 4.

thumb (aa 79–100 and 339–401) subdomains in the shape of a right hand, as well as the PAD (aa 401–518) unique to Y-family polymerases (Figure 3). The palm subdomain comprises the "floor" of the DNA binding groove and carries the catalytic residues. The fingers subdomain is small and stubby. The thumb subdomain is novel, with its largest secondary structural element derived from the N-terminal extension. However, the most unexpected feature of the structure is the position of the PAD: tucked under the palm subdomain, as opposed to being anchored to the fingers subdomain in other Y-family polymerases (Figure 2).

Varied Palm Subdomain

The hPolκ palm subdomain is both similar and different from that of Y-family members, such as yPolη and Dpo4 (Figure 4). The similarity extends to the core of the palm subdomain, comprised of a central mixed β sheet (strands β5, β6, β1, β9, and β10) flanked by three helices: a short αK on the ventral (DNA binding) side and the longer αE and αJ on the dorsal side. These conserved elements superimpose well into the corresponding secondary structures in yPolη (β7, β8, β1, β10, β11 and αJ, αK, and αF) and Dpo4 (β5, β6, β1, β7, β8 and αD, αE, and αF) with rmsds of 2.766 Å (44 carbon alphas) and 1.319 Å (60 carbon alphas), respectively. The less conserved elements of hPolκ, yPolη, and Dpo4 palm subdomains are on the dorsal side: a long α/β extension (αG, αH, αI, β7, and β8) in hPolκ, a compact α-helical subdomain (αA, αB, αG, αH, and αI) in yPolη, and the lack of any such structure in Dpo4 (Figure 4). The long α/β extension in hPolκ gives the appearance of a loosely

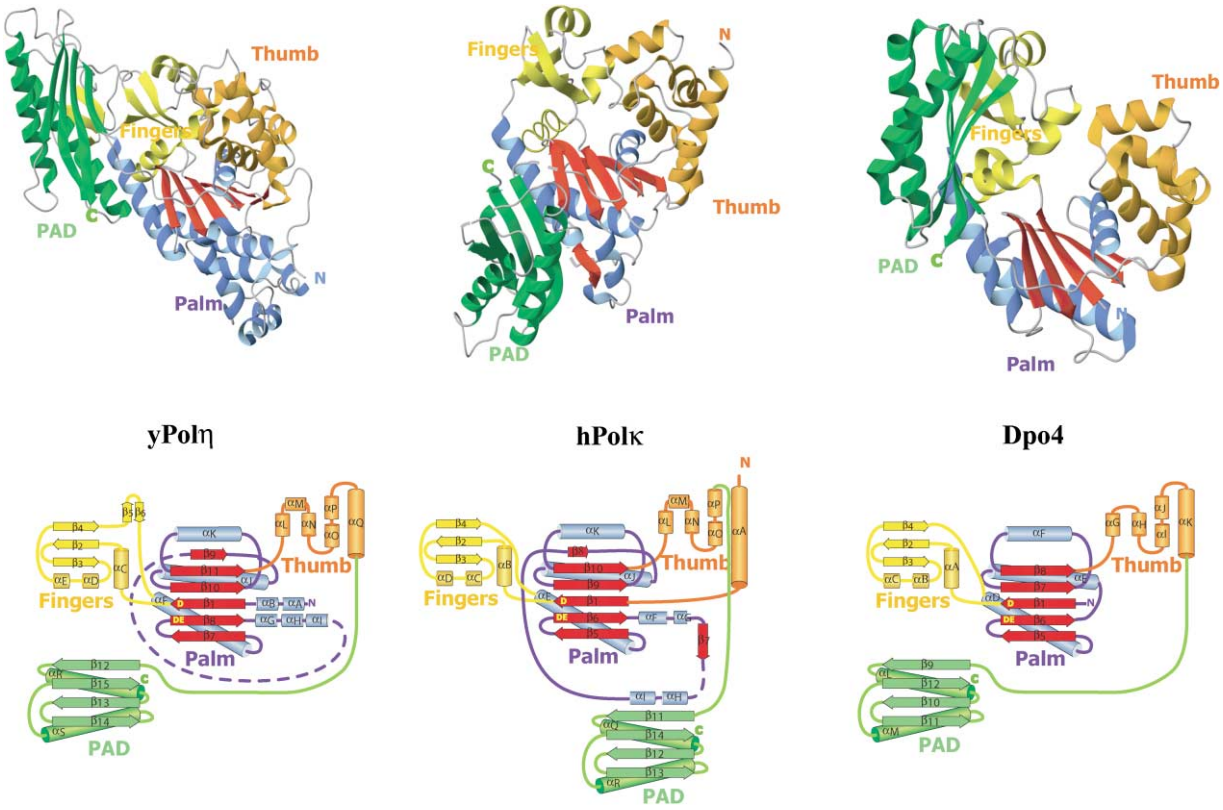

Figure 4. Comparison of hPolκ, yPolη, and Dpo4

Top: structures of hPolκ (middle), yeast Polη (left), and Dpo4 (right), viewed roughly down the DNA axis, as in Figure 2. The structures were aligned by the conserved elements of the palm subdomain. Note the difference in the position of the PAD (green) in the three polymerases, and an N-terminal αA helix in the thumb subdomain of hPolκ (orange). Bottom: secondary structure connectivity plots of hPolκ (middle), yPolη (left), and Dpo4 (right).

swinging "tendril" hanging from the palm subdomain. This dorsal tendril is inherently flexible and could not, for example, be traced in molecule B, and in molecule A it has high B factors (an average of 51 Å^2 for aa 225–280). The configuration of the tendril in molecule A is largely stabilized by intramolecular contacts with the PAD (described below), but even here a 9 amino acid stretch between β7 and αH (aa 227–235) has no defined density. The catalytic residues Asp107, Asp198, and Glu199 are on the ventral side of the palm subdomain. These acidic residues are conserved in all Y-family polymerases, with mutations of the equivalent residues Asp30, Asp155, and Glu156 in Polη inactivating the polymerase (Kondratick et al., 2001) The ventral side of the hPolκ palm subdomain is thus generally similar to that yPolη and Dpo4, while the dorsal side differs in both structure and flexibility (Figures 2 and 4).

Small and Stubby Fingers Subdomain

The hPolκ fingers subdomain is comprised of just three α helices (αB, αC, and αD) and three short β strands (β2, β3, and β4) (Figure 2). The subdomain is smaller than that of yPolη lacking, for example, the strands analogous to β5 and β6 (Figure 4). Nonetheless, the active site cleft at the nexus of fingers and palm subdomain is actually smaller than in yPolη (Figure 5). The narrowness of this active site cleft may be one reason why hPolκ is less effective than Polη in accommodating and bypassing UV-induced T-T dimers. The hPolκ fingers subdomain is also the site of two of the three conserved residues shown to be important in positioning the triphosphate moiety of the incoming nucleotide in Y-family polymerases. These two residues in hPolκ, Tyr141 and Arg144, emanate from the end of helix αC. (The third conserved residue, Lys 328, stems from the loop between helix αK and strand β10 of the palm.) The importance of these residues is highlighted by mutation of the analogous residues in yPolη (Tyr64, Arg67, and Lys269), which reduces the efficiency of correct nucleotide incorporation (Johnson et al., 2003a).

Y-family polymerases in general have smaller fingers subdomains than replicative polymerases and they also lack the equivalent of helices "O" and "O1" that close off the active site in replicative polymerases (Ling et al., 2001; Silvian et al., 2001; Trincao et al., 2001; Zhou et al., 2001). The identity of the catalytic residues and the similarity of the palm subdomain between hPolκ/yPolη and Dpo4 allow both a template-primer and an incoming nucleotide (from Dpo4/DNA/ddADP complex) to be modeled into hPolκ and yPolη DNA binding clefts (Figure 5). Interestingly, although the hPolκ fingers subdomain

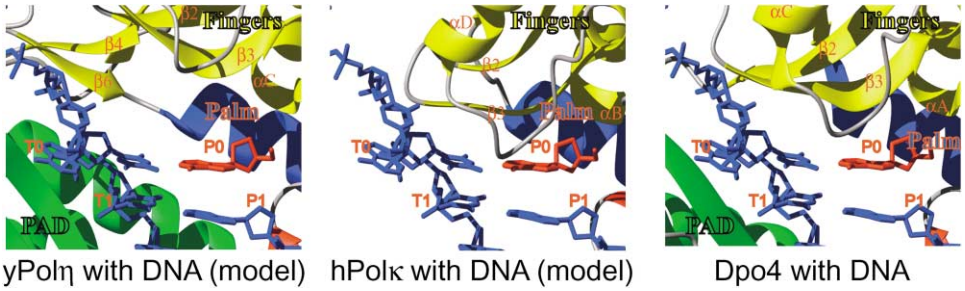

Figure 5. Putative Interactions with Templating Base

Models of hPolκ (middle) and yPolη (left) with DNA based on a least squares fit of their palm subdomain with the Dpo4 palm subdomain (right). Loops in the hPolκ fingers subdomain impinge on the putative templating base (T0), whereas the equivalent loops in yPolη are configured away from the templating base. P0 refers to the incoming nucleotide.

is smaller than that of yPolη, the active site cleft appears from the modeling to be more tightly restrained with respect to a putative templating base (Figure 5). For example, compared to yPolη, the hPolκ fingers subdomain is shifted toward the templating base (by ~5.5 Å) and residues Met135 and Ala151 sterically overlap with the base (Figure 5). hPolκ is the most faithful of all Y-family DNA polymerases; incorporating nucleotides with a frequency of 10^{-3} to 10^{-4}, as compared to 10^{-2} to 10^{-3} for Polη (Johnson et al. 2000a, 2000c; Washington et al. 1999). The restrictive active site cleft of hPolκ may underlie its better fidelity on undamaged DNA

Novel Thumb Subdomain

The hPolκ thumb subdomain is topologically different than that of other Y family polymerases. In yPolη, for example, the thumb subdomain is comprised of contiguous amino acids between the palm subdomain and the PAD that fold into six consecutive α helices (Figures 3 and 4). The analogous segment in hPolκ (aa 339–401) forms five α helices (αL–αP) of the thumb subdomain, but the sixth and longest helix, αA (aa 79–95), derives from the N-terminal sequence unique to hPolκ (Figures 3 and 4). Interestingly, the top half of this N-terminal helix (αA) can be roughly structurally aligned to helix αQ of yPolη thumb subdomain, but it has an opposite polarity.

An intriguing question is the role of residues 19–67 that precede our construct and are required for full hPolκ activity. Given the polarity of αA, these residues will extend beyond the top of the thumb subdomain and could potentially reach over and interact with the template-primer. Correspondingly, the deletion of residues 19–67 diminishes considerably the ability of hPolκ to bind DNA (R.E.J., S.P., and L.P., unpublished data). In all, the hPolκ structure provides a basis for the importance of the N-terminal amino acids in its function: revealing that a portion of N-terminal extension (aa 79–95) is integral to the folding of the thumb subdomain (Figures 2 and 4), while another portion (aa 19–67) can potentially reach over and interact with the DNA.

Conformational Freedom of the PAD

Perhaps the most unexpected feature of our structure is the position of the PAD. The PAD occupies two different positions, both far removed (>50 Å) from the positions in yPolη, Dpo4, and Dbh (Ling et al., 2001; Silvian et al., 2001; Trincao et al., 2001; Zhou et al., 2001). In hPolκ molecule A, the PAD is tucked under and behind the palm subdomain—distant from the DNA binding surface (Figure 2). The curved β sheet of the PAD cradles the dorsal helices of the palm subdomain (αE and αJ), making mostly water-mediated interactions. The most intimate contacts (nonpolar and polar) occur between the β12 -β13 loop of the PAD and the α/β tendril of the palm subdomain. The total buried surface area between the PAD and the palm subdomain in molecule A is about 1200 Å^2. The PAD in molecule B is again juxtaposed on the dorsal side of the palm subdomain but differently than in molecule A (Figure 2). Relative to the PAD in A, the PAD in B is shifted ~18 Å and rotated 56° more toward the thumb subdomain. Because the α/β tendril is unstructured in molecule B, the PAD-palm contacts are limited to water-mediated interactions between strands β11, β12, and β13 with helices αE and αJ. Consequently, only 600 Å^2 of buried surface area anchors the PAD to the palm subdomain in molecule B.

The difference in PAD positions between hPolκ and other Y-family polymerases is striking. In yPolη, the PAD is anchored to the fingers subdomain with >2000 Å^2 of buried surface area (Figure 4). The majority of the resulting intramolecular interactions are electrostatic and occur between helix αR of the PAD and the loop between strands β5 and β6 of the fingers subdomain. Interestingly, the hPolκ fingers subdomain lacks the equivalent of strands β5 and β6 and this may be one reason why the PAD in hPolκ packs on the dorsal side of the palm subdomain, bolstered by interactions with the α/β tendril. In the Dpo4-DNA complex, contacts between the PAD and fingers subdomain are limited to few residues in the loop between β2 and β3 of the fingers subdomain and strand β9 of the PAD, with only ~600 Å^2 of buried surface area. Compared to yPolη, the Dpo4 PAD is rotated toward the DNA major groove by ~50° and interactions with the DNA and the palm subdomain bury an additional ~1000 Å^2 of surface area (Figure 4). In all, the PAD moves ~10 Å between apo-yPolη and DNA-bound Dpo4, with a similar shift of the PAD between apo-Dbh and Dpo4 (Ling et al., 2001; Silvian et al., 2001; Trincao et al., 2001; Zhou et al., 2001). In contrast, the PAD in molecule A would have to move >50 Å and twist >100° in order to bind the DNA major groove (Figure 6). In the

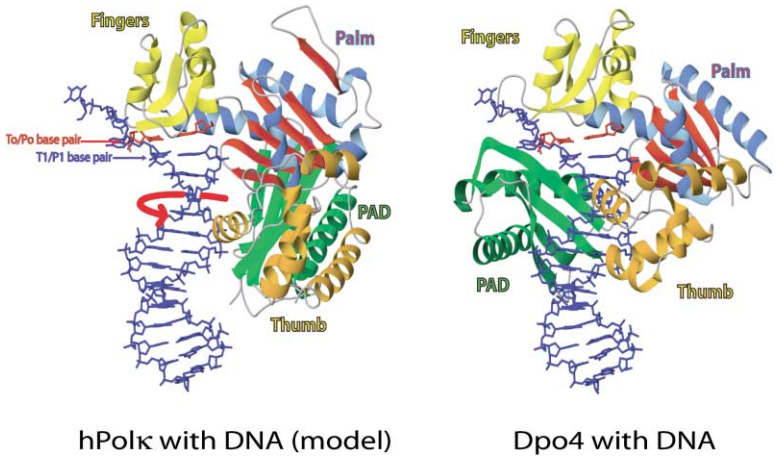

Figure 6. Comparison to Dpo4/DNA/ddADP Complex

Model of hPolκ with DNA (left) based on a least squares fit of its palm subdomain with the Dpo4 palm subdomain (right). The molecules are oriented approximately perpendicular to the DNA axis with the templating strand on the left, the fingers subdomain (yellow) above the templating base, the thumb subdomain (orange) in the minor groove, and, in the case of Dpo4, the PAD contacting the major groove of the DNA. The red arrow shows the direction the hPolκ PAD must move in order to bind DNA in a manner similar to Dpo4. The site of nucleotide insertion (T0/P0) is highlighted in red. The downstream base pair (T1/P1) is also indicated.

case of molecule B, the shift would be ~40 Å and the rotation ~145°. Altogether, the magnitude of motion for hPolκ PAD from unbound to DNA-bound state appears to be much bigger than for yPolη or Dpo4 PAD.

Discussion

Organisms have evolved a variety of repair mechanisms to remove DNA lesions, but some lesions escape repair and block the replication machinery. The recently discovered Y-family DNA polymerases promote continuity of the replication fork by allowing replication through DNA lesions. Interestingly, there is growing evidence that lesion bypass often requires the sequential action of two polymerases, an "inserter" and an "extender" (Prakash and Prakash, 2002). The inserter is efficient at insertion of an incoming nucleotide across from the lesion and the extender is recruited to add bases downstream of the lesion (Prakash and Prakash, 2002). In eukaryotes, Polκ and Polζ, a B-family polymerase, are specialized for the extension step of lesion bypass (Haracska et al., 2002; Johnson et al., 2000b, 2003b; Prakash and Prakash, 2002; Washington et al., 2002). On undamaged DNAs, for example, hPolκ misincorporates nucleotides with a frequency of ~10^{-3} to 10^{-4}, whereas it extends mispaired termini almost two orders of magnitude more efficiently, with a frequency of ~10^{-1} to 10^{-2} (Washington et al., 2002). Also, Polκ is unable to insert nucleotides opposite the 3′T of a *cis-syn* T-T dimer, but it can efficiently extend from a nucleotide inserted opposite the 3′T of the dimer by another DNA polymerase (Washington et al., 2002). In addition to UV sensitivity, Polκ-deficient mouse cells display increased sensitivity to benzo[a]pyrene (BaP) (Ogi et al., 2002). BaP introduces bulky adducts at the N^2 position of dG and less frequently at the N^6 position of dA. Both these adducts present a strong block to nucleotide incorporation by Polκ; the extension step, however, is not as severely affected (Rechkoblit et al. 2002; Suzuki et al., 2002; Frank et al., 2002). The reaction of acrolein, an α,β-unsaturated aldehyde, with the N^2 of dG followed by ring closure at N1 leads to the formation of the cyclic adduct γ-hydroxy-1, N^2-propano-2′deoxygunaosine (γ-HOPdG). This adduct, too, is a strong block to nucleotide incorporation by Polκ, but the polymerase carries out efficient extension from a C nucleotide incorporated opposite γ-HOPdG by another polymerase, such as Polι (Washington et al., 2004). Together, these properties suggest that the hPolκ active site is constrained at site of templating base and incoming nucleotide, but the polymerase is less constrained following translocation of the lesion.

To begin to understand how the structure of hPolκ differs from that of other Y-family polymerases, we determined the crystal structure of the catalytic core of hPolκ at 2.4 Å resolution. The structure reveals a fingers subdomain that is smaller than in yPolη. However, despite its small size, the fingers subdomain appears to be more tightly restrained with respect to a template base (Figure 5; see below). The structure also reveals a novel thumb subdomain that provides a basis for the importance of the N-terminal extension in the hPolκ primary sequence: revealing that the N-terminal extension is integral to both the folding of the thumb subdomain and that it can potentially interact with the DNA. And, most surprisingly, the structure reveals a PAD juxtaposed on the dorsal side of the palm subdomain, as opposed to the fingers subdomain in yPolη, Dpo4, and Dbh.

Does the structure offer any clues as to the specific role of hPolκ in the extension step of DNA synthesis? The modeling of template-primer and an incoming nucleotide (from Dpo4/DNA/ddADP complex) reveals a particularly tight active site in hPolκ (Figures 5 and 6). Although the fingers subdomain will likely move on actual DNA binding, the comparison with yPolη suggests that the inability of hPolκ to insert a nucleotide opposite the 3′T of a *cis-syn* T-T dimer stems from a constrained active site cleft that cannot accommodate both Ts (connected by a covalent cyclobutane linkage) of the T-T dimer. The strong block to nucleotide incorporation opposite BaP and γ-HOPdG adducts may be a further reflection of the constrained active site cleft in the Polκ structure. On undamaged DNAs, Polκ is the most faithful of all Y-family DNA polymerases; incorporating nucleotides with a frequency of 10^{-3} to 10^{-4}, as compared to 10^{-2} to 10^{-3} for Polη (Johnson et al. 2000a, 2000c; Washington et al. 1999). From the structure, the restrictive active site cleft of hPolκ again appears to underlie its better fidelity on undamaged DNA.

Extension of the primer terminus opposite from a lesion or a mismatch poses a different structural challenge than insertion. Following the insertion of a nucleotide opposite a lesion, the lesion base pair is translocated along the template-primer from T_0-P_0 to T_1-P_1 position, where T and P refer to template and primer strands, respectively, and the subscripts refer to number of base pairs from the templating base position (Figures 5 and 6). The distorted DNA backbone geometry of a lesion (or a mismatch) at T_1-P_1 will impact the position of the primer 3′-OH in the active site, and thereby affect the nucleophilic attack on the incoming nucleotide (Trincao et al., 2004). One could therefore envisage a unique residue (or a set of residues) in the hPolκ active site that optimally aligns the primer 3′-OH for the nucleophilic attack. However, the identity of hPolκ residues in the vicinity of the modeled primer 3′-OH (Figure 5) is nearly identical to that in other Y-family polymerases. Alternatively, the dexterity of the PAD in absorbing the "shock" of a DNA lesion at T_1-P_1 could also influence the efficiency of a Y-family polymerase in extending past the lesion. In the Dpo4 ternary complex, residues Arg247, Ser250, and Arg332 of the PAD make extensive hydrogen bonds with the sugar-phosphate backbone of the T_1 base. The hPolκ structure reveals a remarkably flexible PAD: more configurationally dynamic than anticipated from a comparison between yPolη, Dpo4, and Dbh structures (Ling et al., 2001; Silvian et al., 2001; Trincao et al., 2001; Zhou et al., 2001). The PAD occupies two different positions in hPolκ molecules A and B, and may move by as much as ~50 Å on DNA binding (Figures 2 and 6). It is tempting to think that the hPolκ PAD on DNA binding is perhaps less constrained in its interaction with template-primer and thus a better "shock absorber" than in other Y-family polymerases. Also, any weakening of the DNA binding affinity from a less constrained PAD may be compensated by the N-terminal extension (unique to hPolκ) wrapping around the DNA.

Experimental Procedures

Construction of Polκ Truncations

Truncations of hPolκ were derived from pBJ733 (Johnson et al., 2000a), which contains the entire hPolk open reading frame (ORF) in plasmid YIpLac211. The hPolκ (1-451) protein was generated by digestion of pBJ733 with AflII, followed by generation of blunt ends using dNTPs and T4 DNA polymerase. The religated vector, pBJ830, generates a frameshift mutation which causes the hPolκ protein to be truncated at residue 451. hPolκ (1-513) was generated by digestion of pBJ733 with EcoRV and SpeI, which cleaves in the 3′ region of the hPolκ gene, and similarly blunt ended and religated. The resulting plasmid, pBJ845, contains a stop codon at position after amino acid codon 513 of the hPolκ gene. The hPolκ (1-559) protein was generated by digestion of pBJ733 with XbaI and SpeI followed by religation, generating plasmid pBJ828 which contains a frameshift mutation that truncates the hPolκ gene after codon 559. hPolκ proteins prematurely terminated at residue 526 were generated by PCR using an oligonucleotide containing the termination codon TAG at position 527 of the hPolκ gene. N-terminal deletions of hPolκ were generated by PCR, using primers to amplify regions corresponding to the residues indicated. All PCR products were verified by sequencing. All hPolκ ORFs were subsequently cloned into pBJ842, to generate yeast expressible GST-fusion proteins.

GST-fusion plasmids were transformed into yeast strain BJ5464, and proteins were expressed and purified using glutathione Sepharose beads as previously described (Trincao et al., 2001). All proteins were cleaved from their GST tags by treatment with PreScission protease (Amersham-Pharmacia). For crystal formation, hPolκ (68-526) was purified from yeast similar to that described for yPolη (Trincao et al., 2001), except that protein extracts were treated with 0.208 g/ml (35%) ammonium sulfate.

DNA Polymerase Assays

DNA polymerase assays were carried out in 5 μl reactions containing 25 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 100 μg/ml BSA, and 10% glycerol, and contained 50 μM each of dATP, dGTP, dTTP, and dCTP. DNA substrate was composed of the oligodeoxynucleotide primer (32-mer), 5′ GTTTTCCCAG TCAC GACGAT GCTCCGGTAC TC 3′, annealed to a 52-mer template 5′ TTCGTATAAT GCCTACACTG GAGTACCGGA GCATCGTCGT GAC TGGGAAAAC 3′. Reactions were carried out for 10 min at 37°C, stopped with loading buffer (95% formamide, 0.03% bromophenol blue, 0.03% xylene cyanol), and products were separated on 15% polyacrylamide gels containing 8 M urea.

Purification of hPolκ (68-526)

The 0%–35% pellets were dialyzed to remove any remaining ammonium sulfate and were loaded onto a column of Glutathione Sepharose 4B resin (Pharmacia). The GST-fusion protein was bound for 4 hr at 4°C. The GST tag was cleaved on the column and the protein was further purified using an SP Sepharose resin (Pharmacia). Purity was verified by electrophoresis and by MALDI-MS using standard methods (Cohen and Chait, 2001). To prepare SeMet hPolκ, GST-hPolκ (68-526) was expressed in *E. coli* by transforming BL21 codon+ strain with the T7 plasmid pBJ1075. Cells were grown in M9 minimal media and SeMet incorporated by inhibition of the methionine synthesis pathway as previously described (Carter and Sweet, 1997). The SeMet protein was purified as described for the native, except with a higher concentration of DTT (5 mM) to prevent oxidation of SeMet. SeMet incorporation was verified by MALDI-MS (Cohen and Chait, 2001).

Crystallization

Crystals of native hPolκ (68-526) were grown using the vapor diffusion method with a well solution of 8% PEG 8K, 8% ethylene glycol, and 100 mM HEPES (pH 7.75). Crystals took approximately 2 weeks to appear and were grown for more than a month before data collection. The crystals were soaked briefly (~1 min) in a cryoprotectant of well solution doped with 40% ethylene glycol. The crystals were frozen by plunging into liquid nitrogen and were mounted at a later time. Crystals of SeMet hPolκ were grown using the sitting drop method with a well solution of 4% PEG 5K MME, 2% ethylene glycol, and 100 mM HEPES (pH 7.0). Cryoprotectant in this case was the well solution doped with 35% ethylene glycol. Crystals used for data collection were grown by microseeding from smaller crystals. The native and SeMet crystals belong to space group $P2_12_12_1$ with unit cell dimensions of a = 85.209 Å, b = 109.46 Å, c = 111.205 Å. The native crystals diffract to 2.4 Å with synchrotron radiation (Advanced Photon Source) and there are two molecules in the asymmetric unit (Table 1). SeMet crystals diffract to 2.7 Å.

Data Collection and Structure Determination

The native data were measured at the Advanced Photon Source (APS, beamline 14 ID). The MAD data were also measured at APS (beamline 19 ID) at wavelengths corresponding to the edge and peak of the selenium K edge absorption profile, plus one remote wavelength (Table 1). The positions of 24 of the 30 selenium atoms in the asymmetric unit were determined using SOLVE (Terwilliger and Berendzen, 1999). The initial experimental phases (2.7 Å) were applied to the native data measured at APS (beamline 14 ID), and extended to the resolution of the native data (2.4 Å). After solvent flattening, this yielded a readily interpretable electron density map. Initial models of molecules A and B in the AU were built separately without noncrystallographic symmetry (NCS) averaging, yielding an R_{free} of 42%. Successive rounds of building with program O (Jones et al., 1991), in composite omit maps calculated with CNS (Brunger et al., 1998), and refinement lowered the R_{free} to 30.14%. The picking of waters and additional rounds of building and refinement resulted in R_{cryst} and R_{free} of 24.6% and 28.1%, respectively, for native data between 50 and 2.4 Å resolution.

Acknowledgments

We thank the staff at APS (beamlines 14ID and 19ID) for facilitating X-ray data collection. We thank Rong Wang for help with mass spectrometry. This work was supported by NIH grant CA094006 (A.K.A and L.P.). S.N.U. is supported by an NIH fellowship ES012116.

Received: April 1, 2004
Revised: May 13, 2004
Accepted: May 20, 2004
Published: August 10, 2004

References

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Carter, C.W., and Sweet, R.M. eds. (1997). Macromolecular Crystallography, Part A (New York: Academic Press, Inc.).

Cohen, S.L., and Chait, B.T. (2001). Mass spectrometry as a tool for protein crystallography. Annu. Rev. Biophys. Biomol. Struct. *30*, 67–85.

Frank, E.G., Sayer, J.M., Kroth, H., Ohashi, E., Ohmori, H., Jerina, D.M., and Woodgate, R. (2002). Translesion replication of benzo[*a*]pyrene and benzo[*c*]phenanthrene diol epoxide adducts of deoxyadenosine and deoxyguanosine by human DNA polymerase ι. Nucleic Acids Res. *30*, 5284–5292.

Goodman, M.F. (2002). Error-prone repair DNA polymerases in prokaryotes and eukaryotes. Annu. Rev. Biochem. *71*, 17–50.

Haracska, L., Prakash, L., and Prakash, S. (2002). Role of human DNA polymerase kappa as an extender in translesion synthesis. Proc. Natl. Acad. Sci. USA *99*, 16000–16005.

Johnson, R.E., Kondratick, C.M., Prakash, S., and Prakash, L. (1999a). hRAD30 mutations in the variant form of xeroderma pigmentosum. Science *285*, 263–265.

Johnson, R.E., Prakash, S., and Prakash, L. (1999b). Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Poleta. Science *283*, 1001–1004.

Johnson, R.E., Prakash, S., and Prakash, L. (2000a). The human DINB1 gene encodes the DNA polymerase Poltheta. Proc. Natl. Acad. Sci. USA *97*, 3838–3843.

Johnson, R.E., Washington, M.T., Haracska, L., Prakash, S., and Prakash, L. (2000b). Eukaryotic polymerases iota and zeta act sequentially to bypass DNA lesions. Nature *406*, 1015–1019.

Johnson, R.E., Washington, M.T., Prakash, S., and Prakash, L. (2000c). Fidelity of human DNA polymerase eta. J. Biol. Chem. *275*, 7447–7450.

Johnson, R.E., Trincao, J., Aggarwal, A.K., Prakash, S., and Prakash, L. (2003a). Deoxynucleotide triphosphate binding mode conserved in Y family DNA polymerases. Mol. Cell. Biol. *23*, 3008–3012.

Johnson, R.E., Yu, S.L., Prakash, S., and Prakash, L. (2003b). Yeast DNA polymerase zeta (zeta) is essential for error-free replication past thymine glycol. Genes Dev. *17*, 77–87.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47 (Pt 2)*, 110–119.

Kim, S.R., Maenhaut-Michel, G., Yamada, M., Yamamoto, Y., Matsui, K., Sofuni, T., Nohmi, T., and Ohmori, H. (1997). Multiple pathways for SOS-induced mutagenesis in Escherichia coli: an overexpression of dinB/dinP results in strongly enhancing mutagenesis in the absence of any exogenous treatment to damage DNA. Proc. Natl. Acad. Sci. USA *94*, 13792–13797.

Kobayashi, S., Valentine, M.R., Pham, P., O'Donnell, M., and Goodman, M.F. (2002). Fidelity of Escherichia coli DNA polymerase IV. Preferential generation of small deletion mutations by dNTP-stabilized misalignment. J. Biol. Chem. *277*, 34198–34207.

Kokoska, R.J., Bebenek, K., Boudsocq, F., Woodgate, R., and Kunkel, T.A. (2002). Low fidelity DNA synthesis by a y family DNA polymerase due to misalignment in the active site. J. Biol. Chem. *277*, 19633–19638.

Kondratick, C.M., Washington, M.T., Prakash, S., and Prakash, L. (2001). Acidic residues critical for the activity and biological function of yeast DNA polymerase eta. Mol. Cell. Biol. *21*, 2018–2025.

Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2001). Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. Cell *107*, 91–102.

Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2003). Replication of a cis-syn thymine dimer at atomic resolution. Nature *424*, 1083–1087.

Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999). The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. Nature *399*, 700–704.

Ogi, T., Kato, T., Jr., Kato, T., and Ohmori, H. (1999). Mutation enhancement by DINB1, a mammalian homologue of the Escherichia coli mutagenesis protein dinB. Genes Cells *4*, 607–618.

Ogi, T., Shinkai, Y., Tanaka, K., and Ohmori, H. (2002). Polkappa protects mammalian cells against the lethal and mutagenic effects of benzo[a]pyrene. Proc. Natl. Acad. Sci. USA *99*, 15548–15553.

Ohashi, E., Ogi, T., Kusumoto, R., Iwai, S., Masutani, C., Hanaoka, F., and Ohmori, H. (2000). Error-prone bypass of certain DNA lesions by the human DNA polymerase kappa. Genes Dev. *14*, 1589–1594.

Okada, T., Sonoda, E., Yamashita, Y.M., Koyoshi, S., Tateishi, S., Yamaizumi, M., Takata, M., Ogawa, O., and Takeda, S. (2002). Involvement of vertebrate polkappa in Rad18-independent postreplication repair of UV damage. J. Biol. Chem. *277*, 48690–48695.

Prakash, S., and Prakash, L. (2002). Translesion DNA synthesis in eukaryotes: a one- or two-polymerase affair. Genes Dev. *16*, 1872–1883.

Rechkoblit, O., Zhang, Y., Guo, D., Wang, Z., Amin, S., Krzeminsky, J., Louneva, M., and Geacintov N.E. (2002). *trans*-lesion synthesis past bulky benzo[*a*]pyrene diol epoxide N^2-dG and N^6-dA lesions catalyzed by DNA bypass plymerases. J. Biol. Chem. *277*, 30488–30494.

Silvian, L.F., Toth, E.A., Pham, P., Goodman, M.F., and Ellenberger, T. (2001). Crystal structure of a DinB family error-prone DNA polymerase from Sulfolobus solfataricus. Nat. Struct. Biol. *8*, 984–989.

Stary, A., Kannouche, P., Lehmann, A.R., and Sarasin, A. (2003). Role of DNA polymerase eta in the UV mutation spectrum in human cells. J. Biol. Chem. *278*, 18767–18775.

Suzuki, N., Ohashi, E., Kolbanovskiy, A., Geacintov, M.E., Grollman, A.P., Ohmore, H., and Shibutani, S. (2002). Translesion synthesis by human DNA polymerase κ on a DNA template containing a single stereoisomer of dG-(+)- or dG-(-)-*anti*-N^2-BPDE (7,8-dihydroxy-*anti*-9,10-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene). Biochemistry *41*, 6100–6106.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. *55*, 849–861.

Trincao, J., Johnson, R.E., Escalante, C.R., Prakash, S., Prakash, L., and Aggarwal, A.K. (2001). Structure of the catalytic core of S. cerevisiae DNA polymerase eta: implications for translesion DNA synthesis. Mol. Cell *8*, 417–426.

Trincao, J., Johnson, R.E., Wolfle, W.T., Escalante, C.R., Prakash, S., Prakash, L., and Aggarwal, A.K. (2004). Dpo4 is hindered in extending a G·T mismatch by a reverse wobble. Nat. Struct. Mol. Biol. *11*, 457–462.

Washington, M.T., Johnson, R.E., Prakash, S., and Prakash, L. (1999). Fidelity and processivity of Saccharomyces cerevisiae DNA polymerase eta. J. Biol. Chem. *274*, 36835–36838.

Washington, M.T., Johnson, R.E., Prakash, S., and Prakash, L. (2000). Accuracy of thymine-thymine dimer bypass by Saccharomyces cerevisiae DNA polymerase eta. Proc. Natl. Acad. Sci. USA *97*, 3094–3099.

Washington, M.T., Johnson, R.E., Prakash, L., and Prakash, S. (2002). Human DINB1-encoded DNA polymerase kappa is a promis-

cuous extender of mispaired primer termini. Proc. Natl. Acad. Sci. USA *99*, 1910–1914.

Washington, M.T., Prakash, L., and Prakash, S. (2003). Mechanism of nucleotide incorporation opposite a thymine-thymine dimer by yeast DNA polymerase eta. Proc. Natl. Acad. Sci. USA *100*, 12093–12098.

Washington, M.T., Minko, I.G., Johnson, R.E., Wolfle, W.T., Haris, T.J., Lloyd, R.S., Prakash, S., and Prakash, L. (2004). Efficient and error free replication past a minor groove DNA adduct by the sequential action of human DNA polymerases iota and kappa. Mol. Cell. Biol. *24*, 5687–5693.

Wolfle, W.T., Washington, M.T., Prakash, L., and Prakash, S. (2003). Human DNA polymerase kappa uses template-primer misalignment as a novel means for extending mispaired termini and for generating single-base deletions. Genes Dev. *17*, 2191–2199.

Yu, S.L., Johnson, R.E., Prakash, S., and Prakash, L. (2001). Requirement of DNA polymerase eta for error-free bypass of UV-induced CC and TC photoproducts. Mol. Cell. Biol. *21*, 185–188.

Zhou, B.L., Pata, J.D., and Steitz, T.A. (2001). Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain. Mol. Cell *8*, 427–437.

Accession Numbers

The coordinates have been deposited in the PDB under ID code 1T94.