Molecular Cell, Vol. 8, 417–426, August, 2001, Copyright ©2001 by Cell Press

Structure of the Catalytic Core of *S. cerevisiae* DNA Polymerase η: Implications for Translesion DNA Synthesis

Jose Trincao,[1] Robert E. Johnson,[2] Carlos R. Escalante,[1] Satya Prakash,[2] Louise Prakash,[2] and Aneel K. Aggarwal[1,3]
[1]Structural Biology Program
Department of Physiology and Biophysics
Mount Sinai School of Medicine
New York, New York 10029
[2]Sealy Center for Molecular Science
University of Texas Medical Branch
Galveston, Texas 77555

Summary

DNA polymerase η is unique among eukaryotic polymerases in its proficient ability to replicate through a variety of distorting DNA lesions. We report here the crystal structure of the catalytic core of *S. cerevisiae* DNA polymerase η, determined at 2.25Å resolution. The structure reveals a novel polydactyl right hand-shaped molecule with a unique polymerase-associated domain. We identify the catalytic residues and show that the fingers and thumb domains are unusually small and stubby. In particular, the unexpected absence of helices "O" and "O1" in the fingers domain suggests that openness of the active site is the critical feature which enables DNA polymerase η to replicate through DNA lesions such as a UV-induced *cis-syn* thymine-thymine dimer.

Introduction

The survival of organisms depends critically on the ability to faithfully replicate DNA. However, cellular DNA is continually subjected to damaging agents such as UV and ionizing radiation, as well as oxidation and hydrolysis. A variety of DNA repair pathways has evolved to repair the resulting lesions, but some lesions escape repair and are encountered by the replication machinery (Freidberg et al., 1995). How cells bypass these lesions during DNA replication has been a key question in the areas of DNA replication, mutagenesis, and carcinogenesis.

The clearest answer to this longstanding puzzle has come with the discovery of DNA polymerase η (Polη), the product of the *RAD30* gene in *Saccharomyces cerevisiae* (Johnson et al., 1999b). Unlike classical DNA polymerases that become stalled at a UV-induced *cis-syn* cyclobutane thymine-thymine (T-T) dimer, Polη can efficiently and accurately replicate past this common sunlight-induced lesion (Johnson et al., 1999b). Polη is able to replicate through a variety of other distorting DNA lesions as well (Haracska et al., 2000a, 2000b; Minko et al., 2001). Polη is a member of a new family of DNA polymerases (Johnson et al., 1999c; Goodman and Tippin, 2000) which includes Polθ/Polκ and Polι in humans and DinB (PolIV) and UmuC (PolV) in *E. coli* (Johnson et al., 2000a, 2000b; Ohashi et al., 2000; Reuven et al., 1999; Tang et al., 1999; Wagner et al., 1999; Tissier et al., 2000). The sequence of these DNA polymerases is unrelated to that of classical polymerases (Pol I–III in prokaryotes and Pol α–ε in eukaryotes; Johnson et al., 1999c).

The discovery of Polη has gained added significance with the subsequent finding that mutations in Polη are responsible for an inherited disorder, the variant form of xeroderma pigmentosum (XP-V; Johnson et al., 1999a; Masutani et al., 1999). Xeroderma pigmentosum (XP) patients are hypersensitive to sunlight, and suffer from a high incidence of skin cancers. In most of these patients (belonging to groups XP-A to XP-G), the disease results from defects in one of the seven genes involved in nucleotide excision repair (NER; Freidberg et al., 1995). However, in ~20% of XP patients, the NER pathway is normal but they are defective in their ability to replicate UV-damaged DNA (Lehmann et al., 1975; Cordeiro-Stone et al., 1997). In the majority of cell lines derived from XP-V patients, Polη is severely truncated (Johnson et al., 1999a; Masutani et al., 1999), resulting in a protein with no polymerase activity. Polη, thus, is the first DNA polymerase demonstrated to act as a tumor suppressor in humans.

Yeast and human Polη replicate through a UV-induced T-T dimer with the same efficiency and fidelity as on undamaged DNA. Both polymerases insert A's opposite the two T's of the dimer, and on damaged as well as undamaged DNA, they incorporate wrong nucleotides with the same frequency of $\sim 10^{-2}$–10^{-3} (Washington et al., 1999, 2000; Johnson et al., 2000c). Yeast Polη can also efficiently and accurately replicate DNA containing 7,8-dihydro-8-oxoguanine (8-oxoG) adducts formed by oxidative damage (Haracska et al., 2000b). Eukaryotic replicative DNA polymerases tend to insert an A opposite the lesion, as a consequence of which 8-oxoG is highly mutagenic and causes G:C to T:A transversions. In contrast, yeast Polη inserts a C opposite 8-oxoG (Haracska et al., 2000b).

DNA polymerases with known structures include members of the PolI family in prokaryotes, homologs of Polα in bacteriophage RB69 and archaebacteria (Wang et al., 1997; Hopfner et al., 1999; Zhao et al., 1999; Rodriguez et al., 2000; Hashimoto et al., 2001), and eukaryotic Polβ (Pelletier et al., 1994). Members of the PolI family include the Klenow fragments of *E. coli* and *Bacillus* PolI, *Thermus aquaticus* (Taq) DNA polymerase, and phage T7 DNA polymerase (Ollis et al., 1985; Beese et al., 1993; Kim et al., 1995a; Korolev et al., 1995; Eom et al., 1996; Doublie et al., 1998; Kiefer et al., 1998; Li et al., 1998). All of these DNA polymerases share a similar architectural plan that resembles a partially opened right hand with "thumb," "fingers," and "palm" domains (Steitz, 1999). Currently, there is no structural information on Polη or any other translesion synthesis DNA polymerase. Consequently, many important questions about the architecture and the mechanism of these novel polymerases remain unanswered. Does Polη have the palm, fingers, and thumb geometry of classical polymerases? Which

[3]Correspondence: aggarwal@inka.mssm.edu

Table 1. Data Collection Phasing and Refinement Statistics

Data Collection	Se-edge	Se-peak	Se-remote	Native
Wavelength (Å)	0.97958	0.97941	0.96482	0.96675
Resolution (Å)	2.8	2.8	2.9	2.25
Number of reflections measured	435,361	408,385	383,440	1,335,912
Unique	40,869	40,635	37,112	70,064
Data coverage (%)	99.4 (100)	98.9 (99.8)	99.3 (100)	94.7 (74)
R_{merge} (%)[a,b]	6.6 (26.7)	8.2 (25.5)	7.0 (25.9)	7.4 (19.2)
I/σ	17.6 (5.7)	13.9 (6.3)	16.7 (5.9)	17.8 (5.3)
MAD phasing statistics				
Number of sites	10	10	10	—
FoM (centric/acentric) 3.2Å[c]	0.7510/0.5659			
FoM (DM) 2.25Å[d]	0.8954			
Phasing power	1.7	1.6	1.1	—
Refinement statistics				
Resolution range (Å)				50–2.25
Reflections, $F > 2\sigma$ (F)				68,697
R_{cryst} (%)[e]				22.6
R_{free} (%)[f]				24.9
Nonhydrogen atoms				
Protein				7,879
Water				318
Rms deviations				
Bonds (Å)				0.0066
Angles (°)				1.265
Average B factor (Å^2)				36.4

[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] FoM = Mean figure of merit computed to 3.2Å.
[d] FoM = Overall mean figure of merit at 2.25Å after density modification.
[e] $R_{cryst} = \Sigma ||F_o| - |F_c|| / \Sigma |F_o|$.
[f] R_{free} was calculated using 10% of data excluded from refinement.

are the putative active site residues? How does the enzyme replicate past DNA lesions? To address these questions, we undertook structural analysis of a yeast Polη fragment that retains the DNA polymerase and damage bypass activities of the full-length enzyme. The structure provides an in-depth look at the geometry of this important translesion synthesis DNA polymerase and offers new insights into the mechanism of translesion DNA synthesis.

Results and Discussion

Structure Determination

We have previously shown that yeast Polη containing residues 1–513 has the same DNA polymerizing and damage bypass activities as the full-length enzyme of 632 residues (Kondratick et al., 2001). We chose a deletion from the C terminus because these residues are the most divergent among translesion synthesis DNA polymerases (Johnson et al., 1999c). For the structural work described here, we expressed the C-terminally truncated yeast Polη containing residues 1–513 as a GST fusion and purified the protein from yeast cells. The GST portion was subsequently cleaved off and tetragonal crystals were obtained from solutions containing polyethylene glycol and ammonium acetate, diffracting to 2.25Å resolution with synchrotron radiation. The structure was solved by the multiwavelength anomalous diffraction (MAD) method (Hendrickson, 1991), using selenomethionine-labeled Polη expressed in *E. coli*. The initial MAD phases (3.2Å) were applied to native data, and the phases then extended to 2.25Å with solvent flattening. An electron density map calculated at that resolution (2.25Å) was of excellent quality, allowing the construction of both copies of Polη in the crystallographic asymmetric unit (molecules A and B), without the need for noncrystallographic symmetry averaging. The current model includes residues 1–509 for molecules A and B, and 318 water molecules (Table 1).

Palm, Fingers, Thumb, and PAD

Polη has the shape of a polydactyl right hand, in which a novel polymerase-associated domain (PAD) mimics an extra set of fingers (Figure 1). The DNA binding groove is thus defined by four domains: palm, fingers, and thumb domains characteristic of all known polymerases, and the PAD that packs alongside the fingers (Figure 1). The palm domain is the most similar to other polymerases and carries the active site residues that catalyze the nucleotidyl transfer reaction. The fingers and thumb domains are radically different from those in other DNA polymerases (Figure 2A).

The palm can be divided into large and small subdomains. The large subdomain contains a mixed 6-stranded β sheet (β1, β7, β8, β9, β10, and β11) flanked by two long α helices (αF and αJ) from one side and a short α helix (αK) from the other. The side of the β sheet with the long α helices forms part of the hydrophobic core, while the other side is largely solvent exposed and constitutes the floor of the DNA binding groove (Figure 1). Almost the entire large subdomain can be superimposed on the T7 DNA polymerase (T7 Pol) palm domain

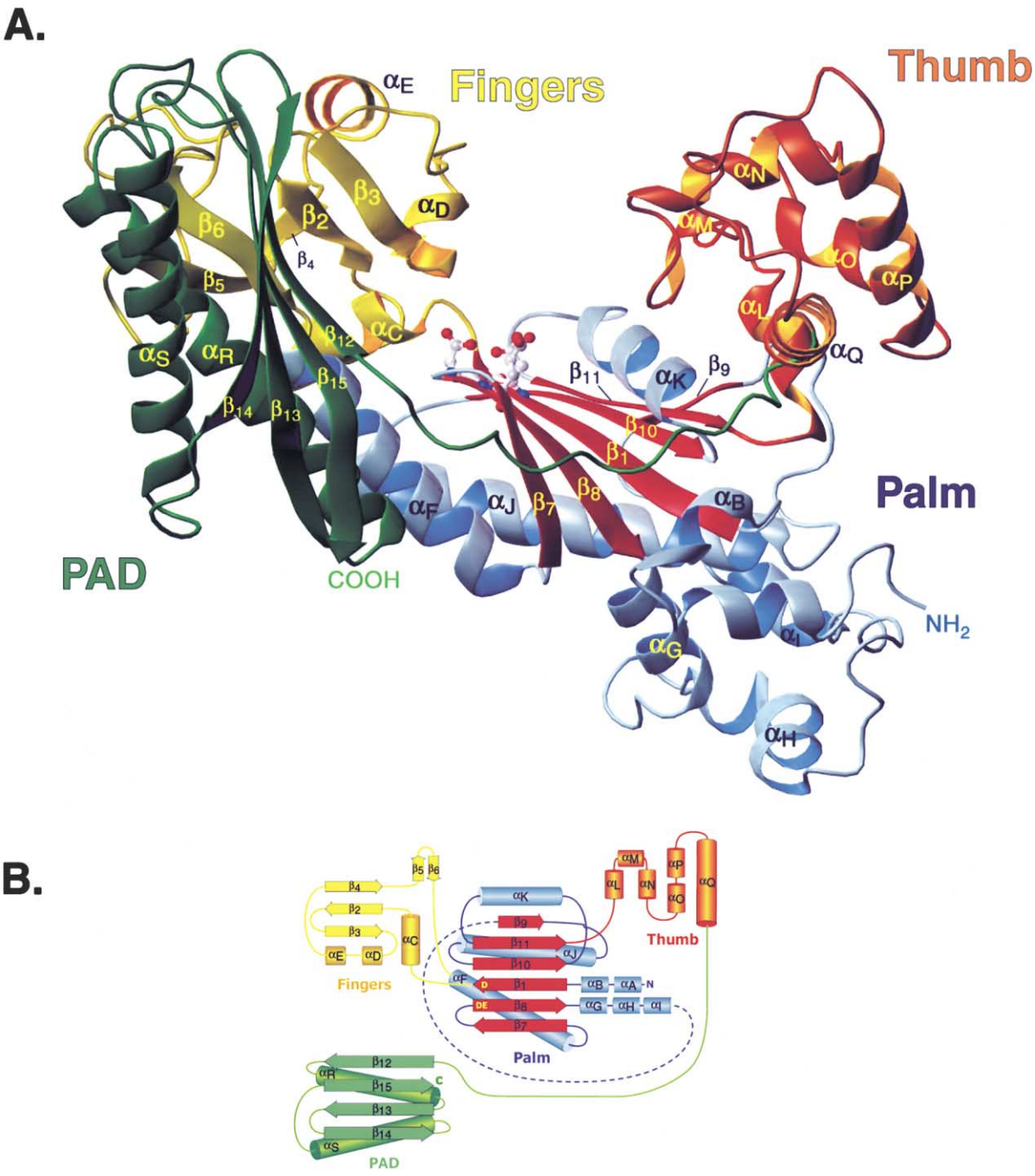

Figure 1. Structure of Polη (Residues 1–513)

(A) A ribbon drawing showing the polydactyl right-hand shape of Polη. Polη is composed of palm (blue and red), fingers (yellow), and thumb (orange) domains, and a unique PAD (green). For clarity, the palm β sheet is drawn in red and the α helices in blue. Also shown are the active site residues (Asp30, Asp155, and Glu156) in a ball-and-stick representation. The α helices (αA to αS) and β strands (β1 to β15) are labeled sequentially from the N to the C terminus.

(B) The secondary structure and domain topology of Polη. The secondary structure elements were defined using PROCHECK (Laskowski et al., 1993). Also indicated are the active site residues Asp30 (as D on strand β1) and the consecutive Asp155 and Glu156 (as DE on strand β8). The coloring scheme is the same as in (A).

(Doublie et al., 1998), with strands β1, β7, β8, and β10 and helices αF and αJ overlapping onto strands β9, β12, β13, and β14 and helices αR and αQ in T7 Pol, respectively (Figure 2B). The root mean square deviation (rmsd) for the superimposed α/β substructures in the two polymerases is 2.1Å (64 Cα's). The palm domains of other polymerases can be similarly superimposed, with rmsd's ranging from ~1.8Å (59 Cα's) for phage

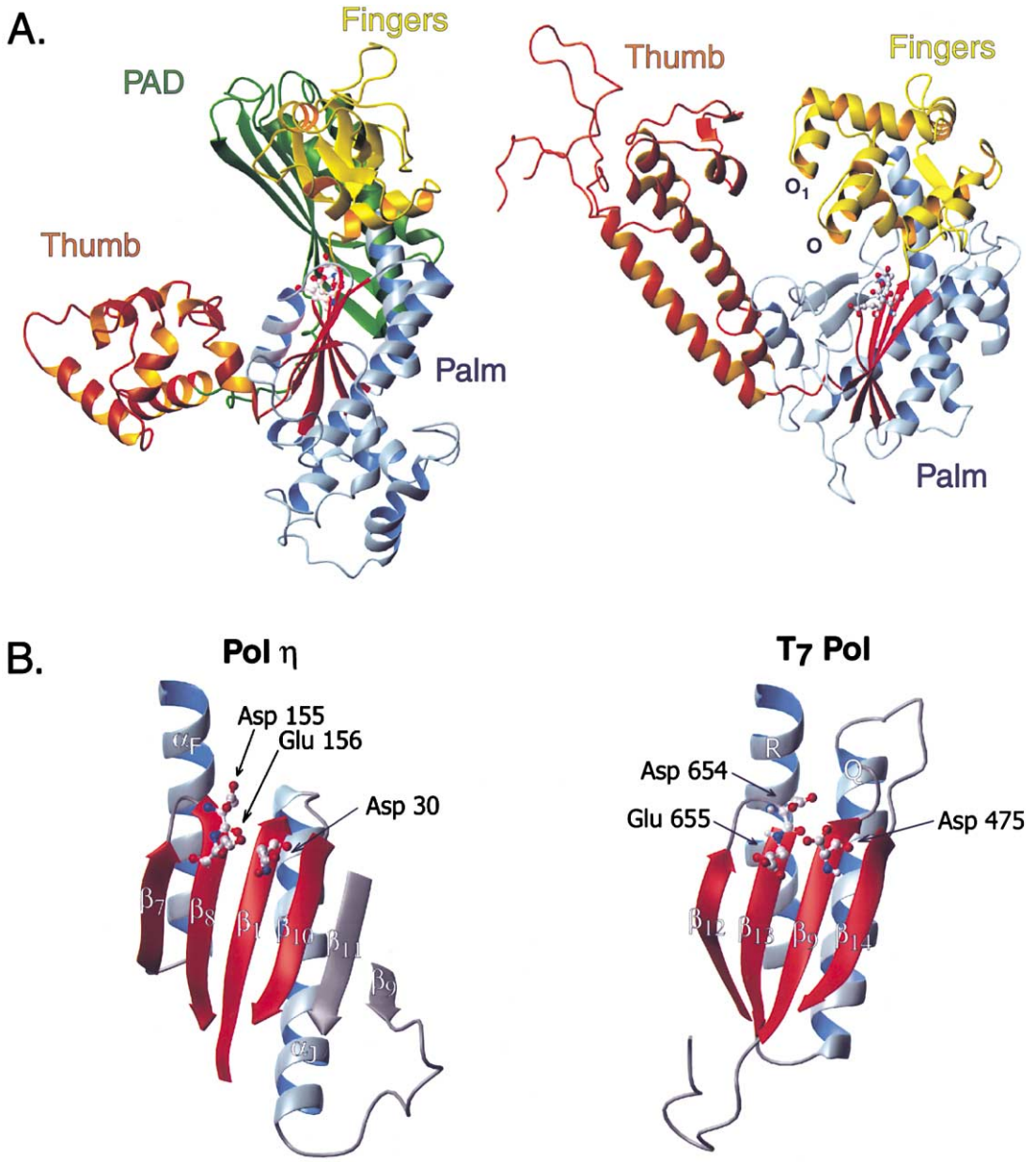

Figure 2. Comparison between Polη and T7 DNA Polymerase

(A) Polη (left) and T7 polymerase (right) are aligned based on a superposition of their palm domains. The view differs from that in Figure 1A by a ~180° rotation about the vertical axis. The protein domains are colored as in Figure 1A. The Polη fingers and thumb domains are smaller than the equivalent domains in T7 polymerase. Note also that Polη fingers domain lacks the equivalent of helices O and O1 (labeled on T7 polymerase).

(B) Comparison between a portion of the palm domain in Polη (left) and T7 polymerase (right). The colored segments (red for β strands and blue for α helices) superimpose with an rmsd of ~2.1Å. Also shown are the active site residues, Asp30, Asp155, and Glu156 in Polη and Asp475, Asp654, and Glu655 in T7 polymerase.

RB69 Polα to ~2.4Å (59 Cα's) for Taq DNA polymerase (Wang et al., 1997; Li et al., 1998). These superpositions establish Asp30, Asp155, and Glu156 as the active site residues in Polη, aligning, for instance, with Asp475, Asp654, and Glu655 in T7 Pol (Doublie et al., 1998). As in T7 Pol, the first carboxylate (Asp30) of this catalytic triad in Polη emanates from a β strand (β1) in the palm domain that leads into the fingers domain, while the second and third carboxylates stem from a neighboring β hairpin (β7 and β8; Figure 2). The small subdomain is a cluster of helices (αA, αB, αG, αH, and αI), whose location at the base of the palm gives the impression of a "wrist" to the yeast Polη hand (Figure 1). Curiously, human Polη appears to lack helix αA of this small subdomain, based on sequence alignment (Figure 3).

The fingers domain is stubby (25Å × 26Å × 33Å), containing two small β sheets (β2, β3, and β4; β5 and β6) and three short α helices (αC, αD, and αE; Figure

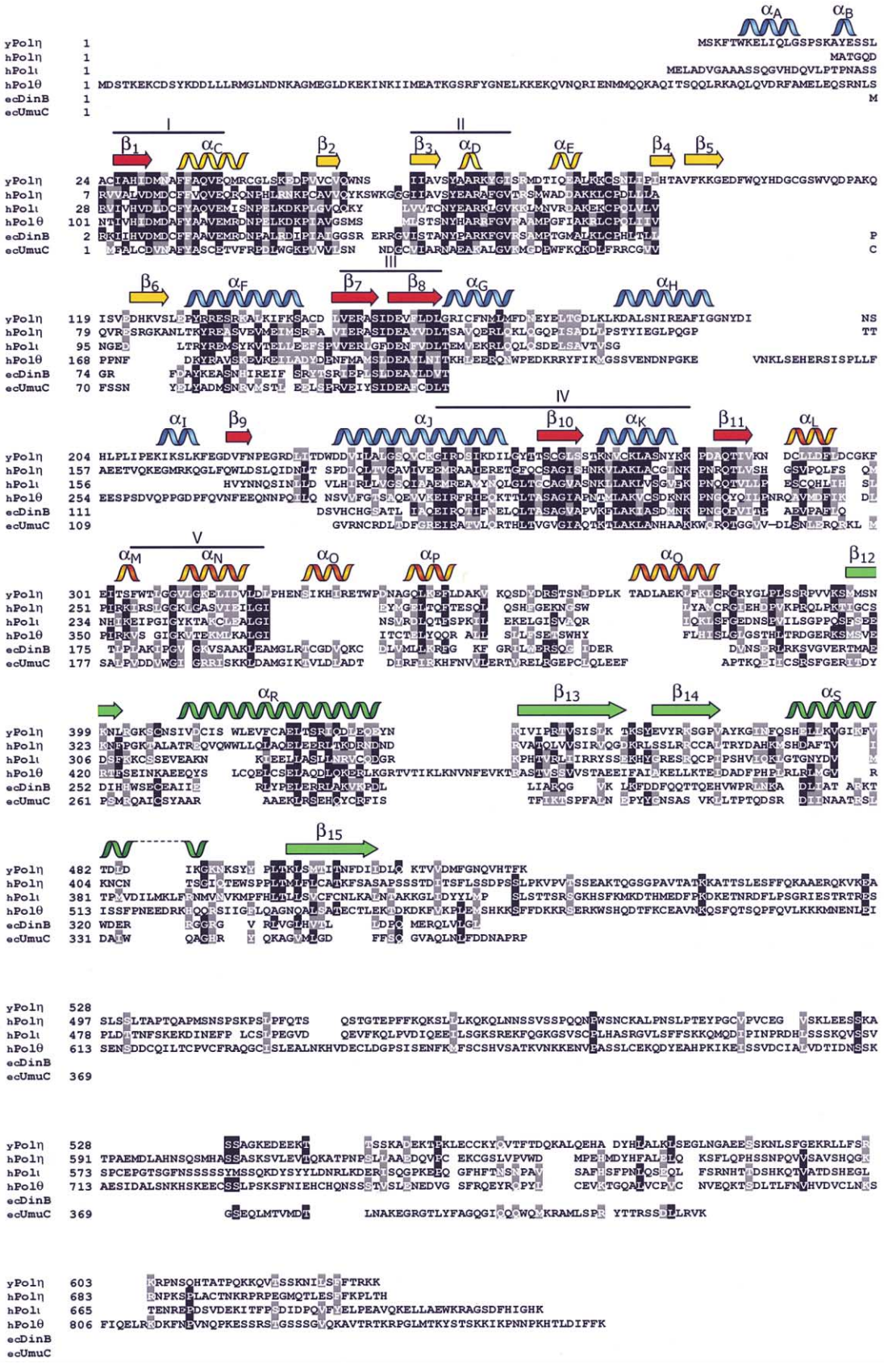

Figure 3. Comparison of Sequences within the Translesion Synthesis DNA Polymerase Family

Included in the comparison are *S. cerevisiae* Polη (yPolη), human Polη (hPolη), human Polι (hPolι), human Polθ (hPolθ), *E. coli* DinB (ecDinB), and *E. coli* UmuC (ecUmuC). Shown above the alignment is the relative location of α helices and β strands in the Polη structure. These secondary structure elements are colored according to which domain they belong to: palm (blue and red), fingers (yellow), thumb (orange), and the PAD (green). Also shown above the alignment are the conserved sequence motifs, designated as I–V (Johnson et al., 1999c).

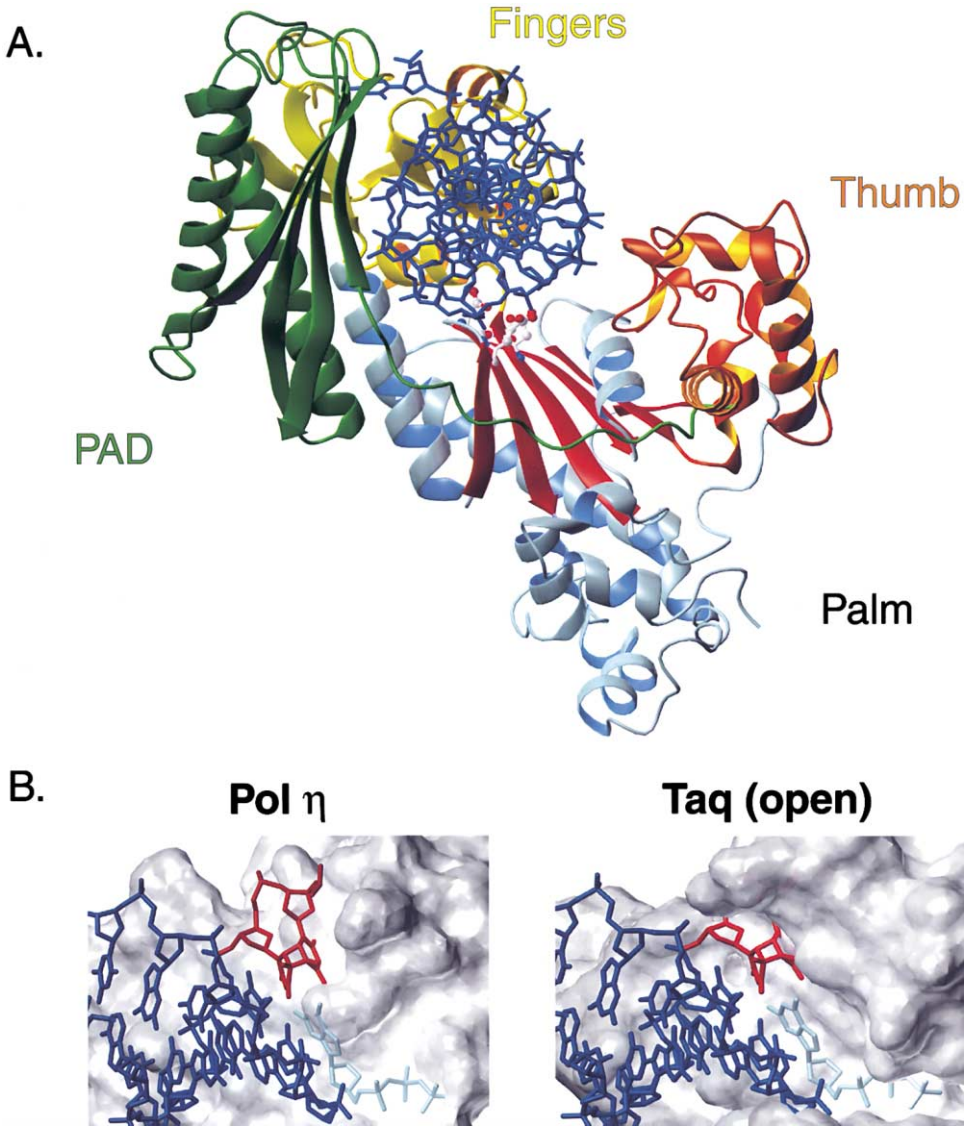

Figure 4. Putative Interactions with Template-Primer

(A) The DNA coordinates (dark blue) were obtained following superposition of the Polη palm domain onto the equivalent domain in the T7 Pol/template-primer/ddGTP ternary complex (cf. Figure 2B; Doublie et al., 1998). Polη is shown in the same orientation as in Figure 1A.
(B) A T-T dimer (red) is modeled in the active sites of Polη (left) and Taq DNA polymerase in the open state (right). The incoming nucleoside triphosphate is drawn in light blue, and the rest of the template and the primer is in dark blue. Polη readily accommodates the 5′ T of the T-T dimer, whereas in Taq DNA polymerase, it faces severe clashes.

1). In contrast, the domain in most other DNA polymerases is larger and composed mostly of α helices. T7 Pol, for instance, has an extended fingers domain (30Å × 32Å × 42Å) that contains eight α helices (Figure 2A; Doublie et al., 1998), while RB69 Polα has a domain characterized by two long α helices that protrude ~50Å from the palm (Wang et al., 1997). However, the most surprising aspect of the Polη fingers domain is the lack of equivalent of helices "O" and "O1" that play a central role in closing off the active site and in the fidelity of Poll DNA polymerases (Figure 2A; Doublie et al., 1998, 1999; Li et al., 1998; Suzuki et al., 2000). Instead, a small loop between helices D and E partially grazes the entrance to the active site in Polη.

The thumb is similarly small and stubby (22Å × 24Å × 25Å), comprised of a 90-residue stump at the palm C terminus (Figure 1). In contrast, the domain in T7 Pol extends >40Å from the base of the palm and, like all Poll polymerases, is encoded as a large insertion within the palm domain (Figure 2A). The Polη thumb is a bundle of six α helices (αL, αM, αN, αO, αP, and αQ) that are structurally unrelated to helices in other DNA polymerases (Figures 1 and 2A). The DNA binding surface area enclosed by the palm, fingers, and thumb domains in Polη (675Å^2) is substantially less than in T7 Pol (1630Å^2) or RB69 Polα (1135Å^2). This could explain why a Polη construct (residues 1–398) containing only the palm, fingers, and thumb domain is unable to bind and polymerize DNA efficiently (Kondratick et al., 2001).

The size of the Polη hand is augmented by an extra

domain, the PAD (residues 393–508). The PAD is joined to the thumb by a flexible tether that traverses the DNA binding groove from the thumb to the fingers side, a distance of over 30Å (Figure 1). The PAD bears uncanny resemblance to the palm in containing a mixed β sheet (β12, β13, β14, and β15) buttressed by two long α helices (αR and αS) from one side. The two β sheets are roughly perpendicular to each other, and are the principal elements defining the floor and the wall of the DNA binding groove (Figure 1). Most importantly, the inclusion of the PAD (13Å × 15Å × 49Å) increases the potential DNA binding surface of Polη from 675Å^2 to 1113Å^2, comparable to that observed in other DNA polymerases (see above).

Conserved Motifs in Translesion Synthesis DNA Polymerases

The Polη sequence is unrelated to that of classical polymerases (Pol I–III in prokaryotes and Pol α–ϵ in eukaryotes), but shows significant homology to Rev1 (a deoxycytidyl transferase) in yeast and DinB (Pol IV) and UmuC (Pol V) in *E. coli* (Johnson et al., 1999c). Other Polη-related proteins have been purified within the last two years and shown to be bona fide polymerases, including products of the human *RAD30B* and *DINB1* genes (Johnson et al., 2000a, 2000b; Ohashi et al., 2000; Tissier et al., 2000) and *E. coli umuC* and *dinB* genes (Reuven et al., 1999; Tang et al., 1999; Wagner et al., 1999). The alignment of these sequences reveals five conserved sequence motifs, designated I–V (Figure 3; Johnson et al., 1999c). The Polη structure defines the basic architecture of these novel DNA polymerases.

Motif I in our structure encodes the β strand (β1) carrying the first catalytic residue (Asp30), while motif III encodes the β hairpin (β7 and β8) carrying the second and third catalytic (Asp155 and Glu156) residues. Motifs I and III are the exact structural analogs of conserved motifs A and C in PolI and Polα DNA polymerases that contain the invariant active site residues (Delarue et al., 1990). Motif II maps to the fingers domain and is characterized by a conserved YxAR sequence (Figure 3). The motif resembles motif B in PolI DNA polymerases (Delarue et al., 1990), with the conserved Tyr and Arg residues mimicking residues in T7 Pol (such as Arg518 and His506) that interact with the incoming nucleoside triphosphate (Doublie et al., 1998). Motif IV is marked by several conserved basic residues, two of which (Arg249 and Lys268) play a structural role in packing helices J and K against the palm β sheet, while another two (Lys272 and Lys279) are in a position to contact the primer DNA strand, analogous to Arg452 and His704 in T7 Pol (Doublie et al., 1998). Motif V maps to a region of the thumb domain facing the DNA binding cleft. The mapping of these conserved motifs to strategic portions of Polη (Figure 3) suggests a similar basic structure for the other related translesion synthesis DNA polymerases. However, because of the differences in their ability to polymerize DNA and to bypass DNA lesions, we expect these polymerases to differ from one another in structural details.

Active Site

Asp30, Asp155, and Glu156 are three of the four acidic residues identified in a mutational analysis of Polη as important for DNA polymerase and T-T dimer bypass activities (Kondratick et al., 2001). The fourth acidic residue (Glu39) identified in the mutational analysis appears to play more of a structural role in maintaining the integrity of the fingers domain. Residues Asp30, Asp155, and Glu156 are conserved in all Polη-related translesion synthesis DNA polymerases and comprise the active site, aligning with Asp475, Asp654, and Glu655 in T7 Pol (Figure 2B). Based on this structural homology to T7 Pol, Asp30 and Asp155 are expected to coordinate two divalent metal ions in the active site, while Glu156 is expected to play a modest role in catalysis. Accordingly, the E156A mutation in Polη shows a decrease in catalysis but is not completely inactive like the D30A and D155A mutant proteins (Kondratick et al., 2001). Similar results were obtained in a mutagenesis study of the equivalent catalytic residues in the Klenow fragment of *E. coli* PolI, where Asp705 and Asp882 are much more critical for catalysis than Glu883 (Polesky et al., 1990, 1992). Taken together, these structural and biochemical similarities suggest a common metal-assisted mechanism of catalysis among replicative and translesion synthesis DNA polymerases.

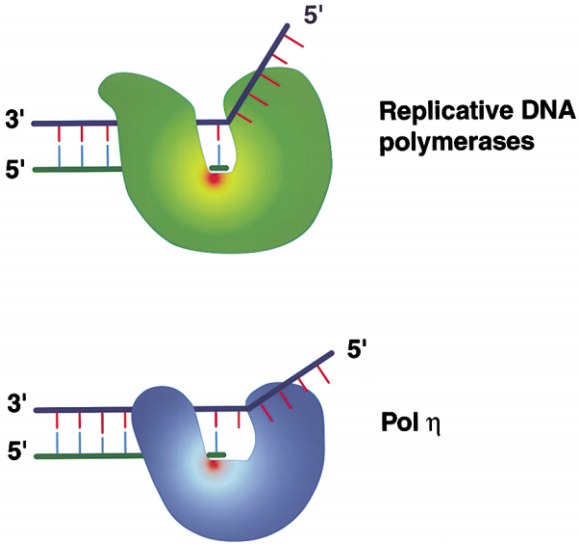

Figure 5. A Model Comparing the Replication Mechanism between Replicative DNA Polymerases and Polη

Replicative DNA polymerases (top) are postulated to contain a tight active site that accommodates only a single template base. Polη (bottom) is shown with a more open active site that can potentially accommodate two template bases.

Putative Interactions with Template-Primer

The similarity between the palm domain of Polη and that of other DNA polymerases allows both a template-primer and an incoming nucleoside triphosphate (NTP) to be modeled into the Polη DNA binding cleft (Figure 4A). Thus, a superposition with the T7 Pol/template-primer/ddGTP ternary complex (Doublie et al., 1998) results in positioning ddGTP in the Polη active site and the primer 3′ end in the joint between the palm and fingers domains. The thumb and the PAD straddle the duplex portion of the modeled template-primer, connected by a long loop that cradles the underside of the

DNA (Figure 4A). Both domains require an inward motion to secure the template-primer, with the thumb making contacts in the minor groove and the PAD interacting in the major groove. The shape of the extended PAD β sheet matches remarkably well to the contour of the major groove surface, compatible with a role for the PAD in stabilizing the Polη/DNA complex.

The role of the PAD may be analogous to that of *E. coli* thioredoxin in T7 DNA replication. T7 Pol recruits thioredoxin to form a tight one-to-one complex that prevents the dissociation of the template-primer during DNA synthesis (Modrich and Richardson, 1975; Huber et al., 1987). Thioredoxin binds an extended, flexible loop within the T7 Pol thumb domain (Doublie et al., 1998) and—like the PAD—it could swing over to the fingers side to encircle the template-primer.

Mechanism for Bypassing DNA Lesions

One of the most intriguing features of Polη to emerge from this DNA modeling is the paucity of putative contacts to the template 5′ end. The unpaired bases of the modeled template 5′ end are relatively unhindered in continuing a helical passage across the Polη fingers domains. In contrast, only a single unpaired template base is held in the active site of T7 polymerase or in Taq or *Bacillus* DNA polymerase I, while the preceding 5′ unpaired template base(s) is directed out of the active site at a 90° angle (Doublie et al., 1998; Kiefer et al., 1998; Li et al., 1998). This steric block comes primarily from helices O and O1 of the fingers domain (Figure 2A) and, in the case of Taq polymerase, it is true for both the closed and open states of the enzyme (Li et al., 1998). (The Taq open state was obtained by soaking out the NTP and has a configuration similar to that of apo enzyme.) Because the 5′ T of a T-T dimer (**T**-T) cannot be flipped out of the active site due to its covalent *cis-syn* cyclobutane linkage to the 3′ T (T-**T**), we suggest that this may be the reason why replicative polymerases such as Taq or T7 become stalled at this common UV-induced lesion. On the other hand, Polη lacks the O and O1 helices (Figure 2A), and its active site is much less restricted in accommodating the 5′ T of the T-T dimer (Figure 4B). NMR studies have shown that the DNA backbone around a T-T dimer is relatively undistorted, and that the thymine bases maintain their parallel stacking as well as their Watson-Crick hydrogen bonding potential (Kemmink et al., 1987; Kim et al., 1995b). Thus, we propose that by accommodating two rather than only a single unpaired template base in the active site, Polη can replicate a T-T dimer without becoming stalled. A tight active site allows replicative DNA polymerases to better sense the geometry of the nascent base pair, and thereby achieve fidelities surpassing those from correct Watson-Crick hydrogen bonding. Polη incorporates wrong nucleotides at a substantially higher error rate (10^{-2}–10^{-3}) than a eukaryotic DNA polymerase such as Polδ (10^{-5}; Washington et al., 1999, 2000; Johnson et al., 2000c; Matsuda et al., 2000). The low fidelity of Polη is consistent with a more open active site, which is less specific but better able to accommodate DNA lesions. Besides a T-T dimer, yeast Polη can also efficiently and accurately replicate DNA containing 8-oxoG adducts (Haracska et al., 2000b). In contrast, eukaryotic replicative DNA polymerases usually insert an A opposite the lesion. This is probably because 8-oxoG, in the absence of conformational restraints, tends to adopt the *syn* conformation, favoring the formation of a Hoogsteen base pair with an adenine. However, it is tempting to speculate that the accommodation of an extra unpaired template base in the Polη active site imposes sufficient backbone constraint or stacking interactions on 8-oxoG to favor *anti* over *syn* conformation, leading to the incorporation of C rather than A. The Polη structure is the first step toward defining the architecture and mechanism of this remarkable DNA polymerase. In particular, the "openness" of the active site appears to be the critical feature which distinguishes Pol η from replicative polymerases, enabling the former to bypass DNA lesions (Figure 5).

Experimental Procedures

Protein Expression and Purification

The GST-Polη (residues 1–513) fusion protein (Kondratick et al., 2001) was expressed in yeast from plasmid pBJ847. This fusion protein contains a PreScission protease recognition sequence, LEVLFQGP, which is cleaved specifically between the glutamine and glycine residues and is located 7 amino acids N-terminal to the first methionine of Polη. Yeast strain BJ5464 harboring plasmid pBJ847 was grown in synthetic complete medium lacking leucine and induced with galactose as described (Johnson et al., 2000c). GST-Polη(1–513) protein was purified as described previously for the full-length protein (Johnson et al., 1999b) with the following modifications: prior to affinity purification on glutathione-Sepharose, protein was precipitated from yeast cell extract using 35%–50% ammonium sulfate. The pellets were then solubilized and passed over a glutathione-Sepharose column. The Polη(1–513) protein lacking the GST tag was eluted from the column by treatment with PreScission protease (Amersham Pharmacia) and was further purified through a Mono Q column, concentrated, and used to obtain native crystals.

To express the GST-Polη(1–513) protein in *E. coli*, the EcoNI/SalI fragment from pBJ847, containing the GST-Polη(1–513) fusion, was used to replace the *GST* gene in plasmid pGEX-6P-3 (Amersham Pharmacia), generating plasmid pBJ875. To prepare selenomethionine-labeled GST-Polη(1–513) protein, plasmid pBJ875 was transformed into an *E. coli* B834 methionine auxotrophic strain, and cells were grown in M9 minimal medium supplemented with all amino acids, except that selenomethionine replaced methionine. Se-Met-labeled protein was purified in a manner similar to the one used for purification from yeast, involving affinity purification over a glutathione-Sepharose column and proteolysis with PreScission protease, followed by a Mono Q column.

Crystallization

Yeast Polη crystallizes in two crystal forms: orthorhombic and tetragonal. We first obtained the orthorhombic crystals from solutions containing 8% PEG 4K and 700 mM ammonium acetate (pH 6.5), at 20°C. The crystals belong to space group $P2_12_12_1$, with unit cell dimensions of $a = 86.3$Å, $b = 106.0$Å, $c = 167.6$Å, and $\alpha = \beta = \gamma = 90°$. Although these crystals are fairly large (up to 1.5 × 0.2 × 0.2 mm), they are hollow and have a diffraction limit of 2.8Å at home. The tetragonal crystals were obtained from solutions containing 6% PEG 20K and 600 mM ammonium acetate, at 4°C. The crystals belong to space group $P4_12_12$ with unit cell dimensions of $a = b = 104.8$Å, $c = 292.3$Å, and $\alpha = \beta = \gamma = 90°$. These crystals are smaller (usually 0.2 × 0.2 × 0.05 mm) than the orthogonal form, but they diffract better and were used for the subsequent structure determination.

Data Collection, Structure Determination, and Refinement

The MAD data were measured at the Advanced Photon Source (APS, beamline 31-ID), at wavelengths corresponding to the edge and peak of the selenium *K* edge absorption profile plus at two

remote points (Table 1). The positions of the selenium atoms and the experimental phases were computed with CNS (Brunger et al., 1998). The initial experimental phases (3.2Å) were applied to native data measured at the National Synchrotron Light Source (beamline X4A), and the phases were then extended to 2.25Å with solvent flattening. This yielded an experimental electron density map that was readily interpretable without the need for noncrystallographic averaging. The model for both Polη molecules (A and B) was built into this map. The initial model had an R factor of 42.5% (R_{free} = 42%), which quickly converged to 22.6% (R_{free} = 24.9%) after iterative rounds of refinement with CNS, model building with O (Jones et al., 1991), and water picking. The final model includes residues 1–509 for molecules A and B, and 318 water molecules (Table 1). The model has good stereochemistry (Table 1), with 87.6% of the residues in the most favored conformation in a Ramachandran plot and only 0.3% in the disallowed regions.

Acknowledgments

We are grateful to K. D'Amico and C. Ogata for facilitating X-ray data collection at APS and NSLS, respectively. We thank L. Shapiro for help with data collection and comments on the manuscript. This work was supported by NIH grants GM44006 (A.K.A.) and GM19261 (L.P.) and institutional funds (A.K.A.). J.T. is supported by Praxis XXI fellowship BD/13594/97 from Fundacao para a Ciencia e Tecnologia.

Received June 8, 2001; revised July 3, 2001.

References

Beese, L.S., Derbyshire, V., and Steitz, T.A. (1993). Structure of DNA polymerase I Klenow fragment bound to duplex DNA. Science *260*, 352–355.

Brunger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R., Jiang, W., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a software suite for macromolecular structure determination. Acta Crystallogr. *D54*, 905.

Cordeiro-Stone, M., Zaritskaya, L.S., Price, L.K., and Kaufmann, W.K. (1997). Replication fork bypass of a pyrimidine dimer blocking leading strand DNA synthesis. J. Biol. Chem. *272*, 13945–13954.

Delarue, M., Poch, O., Tordo, N., Moras, D., and Argos, P. (1990). An attempt to unify the structure of polymerases. Protein Eng. *3*, 461–467.

Doublie, S., Tabor, S., Long, A.M., Richardson, C.C., and Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. Nature *391*, 251–258.

Doublie, S., Sawaya, M.R., and Ellenberger, T. (1999). An open and closed case for all polymerases. Structure *7*, R31–R35.

Eom, S.H., Wang, J., and Steitz, T.A. (1996). Structure of Taq polymerase with DNA at the polymerase active site. Nature *382*, 278–281.

Freidberg, E.C., Walker, G.C., and Siede, W. (1995). DNA Repair and Mutagenesis (Washington, DC: American Society for Microbiology).

Goodman, M.F., and Tippin, B. (2000). Sloppier copier DNA polymerases involved in genome repair. Curr. Opin. Genet. Dev. *10*, 162–168.

Haracska, L., Prakash, S., and Prakash, L. (2000a). Replication past O^6-methylguanine by yeast and human DNA polymerase η. Mol. Cell. Biol. *20*, 8001–8007.

Haracska, L., Yu, S.L., Johnson, R.E., Prakash, L., and Prakash, S. (2000b). Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase η. Nat. Genet. *25*, 458–461.

Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T., and Kai, Y. (2001). Crystal structure of DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. J. Mol. Biol. *306*, 469–477.

Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science *254*, 51–58.

Hopfner, K.P., Eichinger, A., Engh, R.A., Laue, F., Ankenbauer, W., Huber, R., and Angerer, B. (1999). Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. Proc. Natl. Acad. Sci. USA *96*, 3600–3605.

Huber, H.E., Tabor, S., and Richardson, C.C. (1987). *Escherichia coli* thioredoxin stabilizes complexes of bacteriophage T7 DNA polymerase and primed templates. J. Biol. Chem. *262*, 16224–16232.

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, Polη. Science *283*, 1001–1004.

Johnson, R.E., Washington, M.T., Prakash, S., and Prakash, L. (1999c). Bridging the gap: a family of novel DNA polymerases that replicate faulty DNA. Proc. Natl. Acad. Sci. USA *96*, 12224–12226.

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

Johnson, R.E., Washington, M.T., Haracska, L., Prakash, S., and Prakash, L. (2000b). Eukaryotic polymerases ι and ζ 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 η. J. Biol. Chem. *275*, 7447–7450.

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

Kemmink, J., Boelens, R., Koning, T., van der Marel, G.A., van Boom, J.H., and Kaptein, R. (1987). ^{1}H NMR study of the exchangeable protons of the duplex d(GCGTTGCG).d(CGCAACGC) containing a thymine photodimer. Nucleic Acids Res. *15*, 4645–4653.

Kiefer, J.R., Mao, C., Braman, J.C., and Beese, L.S. (1998). Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. Nature *391*, 304–307.

Kim, Y., Eom, S.H., Wang, J., Lee, D.S., Suh, S.W., and Steitz, T.A. (1995a). Crystal structure of *Thermus aquaticus* DNA polymerase. Nature *376*, 612–616.

Kim, J.K., Patel, D., and Choi, B.S. (1995b). Contrasting structural impacts induced by cis-syn cyclobutane dimer and (6–4) adduct in DNA duplex decamers: implication in mutagenesis and repair activity. Photochem. Photobiol. *62*, 44–50.

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 η. Mol. Cell. Biol. *21*, 2018–2025.

Korolev, S., Nayal, M., Barnes, W.M., Di Cera, E., and Waksman, G. (1995). Crystal structure of the large fragment of *Thermus aquaticus* DNA polymerase I at 2.5-A resolution: structural basis for thermostability. Proc. Natl. Acad. Sci. USA *92*, 9264–9268.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *A47*, 110–119.

Lehmann, A.R., Kirl-Bell, S., Arlett, C.F., Paterson, M.C., Lohman, P.H.M., de Weerd-Kastelein, E.A., and Bootsma, D. (1975). Xeroderma pigmentosum cells with normal levels of excision repair have a defect in DNA synthesis after UV-irradiation. Proc. Natl. Acad. Sci. USA *72*, 219–223.

Li, Y., Korolev, S., and Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. EMBO J. *17*, 7514–7525.

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 η. Nature *399*, 700–704.

Matsuda, T., Bebenek, K., Masutani, C., Hanaoka, F., and Kunkel, T.A. (2000). Low fidelity DNA synthesis by human DNA polymerase η. Nature *404*, 1011–1013.

Minko, I.G., Washington, M.T., Prakash, L., Prakash, S., and Lloyd, R.S. (2001). Translesion DNA synthesis by yeast DNA polymerase

η on templates containing N^2-guanine adducts of 1,3-butadiene metabolites. J. Biol. Chem. *276*, 2517–2522.

Modrich, P., and Richardson, C.C. (1975). Bacteriophage T7 deoxyribonucleic acid replication in vitro. A protein of *Escherichia coli* required for bacteriophage T7 DNA polymerase activity. J. Biol. Chem. *250*, 5508–5514.

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 κ. Genes Dev. *14*, 1589–1594.

Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G., and Steitz, T.A. (1985). Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. Nature *313*, 762–766.

Pelletier, H., Sawaya, M.R., Kumar, A., Wilson, S.H., and Kraut, J. (1994). Structures of ternary complexes of rat DNA polymerase β, a DNA template-primer, and ddCTP. Science *264*, 1891–1903.

Polesky, A.H., Steitz, T.A., Grindley, N.D., and Joyce, C.M. (1990). Identification of residues critical for the polymerase activity of the Klenow fragment of DNA polymerase I from *Escherichia coli*. J. Biol. Chem. *265*, 14579–14591.

Polesky, A.H., Dahlberg, M.E., Benkovic, S.J., Grindley, N.D., and Joyce, C.M. (1992). Side chains involved in catalysis of the polymerase reaction of DNA polymerase I from *Escherichia coli*. J. Biol. Chem. *267*, 8417–8428.

Reuven, N.B., Arad, G., Maor-Shoshani, A., and Livneh, Z. (1999). The mutagenesis protein UmuC is a DNA polymerase activated by UmuD′, RecA, and SSB and is specialized for translesion replication. J. Biol. Chem. *274*, 31763–31766.

Rodriguez, A.C., Park, H.W., Mao, C., and Beese, L.S. (2000). Crystal structure of a pol α family DNA polymerase from the hyperthermophilic archaeon *Thermococcus* sp. 9 degrees N-7. J. Mol. Biol. *299*, 447–462.

Steitz, T.A. (1999). DNA polymerases: structural diversity and common mechanisms. J. Biol. Chem. *274*, 17395–17398.

Suzuki, M., Yoshida, S., Adman, E.T., Blank, A., and Loeb, L.A. (2000). *Thermus aquaticus* DNA polymerase I mutants with altered fidelity. J. Biol. Chem. *275*, 32728–32735.

Tang, M., Shen, X., Frank, E.G., O'Donnell, M., Woodgate, R., and Goodman, M.F. (1999). $UmuD'_2C$ is an error-prone DNA polymerase, *Escherichia coli* polV. Proc. Nat. Acad. Sci. USA *96*, 8919–8924.

Tissier, A., McDonald, J.P., Frank, E.G., and Woodgate, R. (2000). polι, a remarkably error-prone human DNA polymerase. Genes Dev. *14*, 1642–1650.

Wagner, J., Gruz, P., Kim, S.-R., Yamada, M., Matsui, K., Fuchs, R.P.P., and Nohmi, T. (1999). The *dinB* gene encodes a novel *E. coli* DNA polymerase, DNA Pol IV, involved in mutagenesis. Mol. Cell *4*, 281–286.

Wang, J., Sattar, A.K., Wang, C.C., Karam, J.D., Konigsberg, W.H., and Steitz, T.A. (1997). Crystal structure of a pol α family replication DNA polymerase from bacteriophage RB69. Cell *89*, 1087–1099.

Washington, M.T., Johnson, R.E., Prakash, S., and Prakash, L. (1999). Fidelity and processivity of *Saccharomyces cerevisiae* DNA polymerase η. 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 η. Proc. Natl. Acad. Sci. USA *97*, 3094–3099.

Zhao, Y., Jeruzalmi, D., Moarefi, I., Leighton, L., Lasken, R., and Kuriyan, J. (1999). Crystal structure of an archaebacterial DNA polymerase. Structure *7*, 1189–1199.

Accession Numbers

The structure has been deposited in the Protein Data Bank with the accession number 1JIH.